Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Ocean Operations and Civil Engineering

Justin Chen

Optimizing Coagulation in Wastewater Treatment: An Application of LSTM Machine Learning and Sensor Adjustments at Åse WWTP in Ålesund

With focus on adding PAX-33 and Superfloc A-1883 as coagulant and flocculant

Master's thesis in Smart Water and Environmental Engineering
Supervisor: Razak Seidu
Co-supervisor: Lars-Andreas Lågeide

May 2023

NTNU
Norwegian University of
Science and Technology

Justin Chen

Optimizing Coagulation in Wastewater Treatment: An Application of LSTM Machine Learning and Sensor Adjustments at Åse WWTP in Ålesund

With focus on adding PAX-33 and Superfloc A-1883 as coagulant and flocculant

Master's thesis in Smart Water and Environmental Engineering
Supervisor: Razak Seidu
Co-supervisor: Lars-Andreas Lågeide
May 2023

Norwegian University of Science and Technology
Faculty of Engineering
Department of Ocean Operations and Civil Engineering

NTNU
Norwegian University of
Science and Technology

Preface

I am honored to present this master's thesis, submitted to the Department of Ocean and Civil Engineering at the Norwegian University of Technology in the spring of 2023. This thesis is the result of a collaborative project between NTNU and the water and sanitation department of Ålesund municipality, focusing on the optimization of treatment processes at an existing wastewater treatment plant.

I would like to express my sincere gratitude to my primary supervisor, professor Razak Seidu from NTNU, and co-supervisor Lars-Andreas Lågeide from Ålesund municipality for their unwavering patience, invaluable guidance, and steadfast support throughout this project. Additionally, I am grateful to Peihua Hang, Bjørghild Lervik, Paul Karsten Grøtt, and Alexander Toft for their insightful input and advice during our meetings.

Finally, I wish to extend my heartfelt appreciation to my beloved, friends, and family for their constant support and encouragement during this journey. Their love and motivation have been instrumental in the completion of this thesis.

Justin Chen
Ålesund, 29.05.2023

Abstract

This master's thesis investigates the use of Long Short-Term Memory (LSTM) machine learning and sensor adjustments for enhancing the coagulation processes at the Åse wastewater treatment plant (RA4) in Ålesund, Norway. Focusing on optimizing the dosage of PAX-33 and polymer (Superfloc A-1883) for effective contaminant removal, the research evaluates sensor utility for real-time monitoring and adjustment of chemical dosages.

A comprehensive analysis of the Åse WWTP's existing processes, equipment, and infrastructure was performed. By using plant operational records and sensor data, variables influencing the chemical clarification process were identified. An LSTM model was then trained and validated on this data to predict and optimize PAX and polymer dosages under varying conditions.

The results demonstrate that the LSTM model, in tandem with sensor adjustments, significantly enhances the efficiency of the coagulation and flocculation process. The LSTM model achieves prediction accuracies of 94.4% for PAX-33 and 77.2% for polymer dosages. Furthermore, implementing advanced multi-parameter sensors promises to improve these prediction accuracies and set the stage for a fully automated dosing system, leading to an efficient treatment process, reduced costs, and lower emissions.

Financially, considering a wastewater sensor's lifespan of 5 to 10 years, the Net Present Value (NPV) over a 10-year period yields an estimated 1 MNOK, given an annual saving of 140,000 NOK, an initial investment of 165,000 NOK, and a discount rate of 3%. This positive NPV indicates that the project would provide net benefits over this period, especially considering the potential 10% annual chemical saving from the optimization.

The thesis contributes valuable insights for the wastewater treatment industry and underlines the benefits of leveraging machine learning and sensor adjustments to optimize wastewater treatment processes. It signifies a step forward in the quest for innovative solutions to enhance the performance of wastewater treatment plants worldwide, thereby promoting long-term sustainability and resilience of urban water systems.

Sammendrag

Denne masteroppgaven undersøker bruk av maskinlæring med Long Short-Term Memory (LSTM) og sensorjusteringer for å forbedre koaguleringsprosessene ved Åse renseanlegg (RA4) i Ålesund, Norge. Forskningen fokuserer på å optimalisere doseringen av PAX-33 og polymer (Superfloc A-1883) for effektiv fjerning av forurensninger og vurderer nytten av sensorer for sanntidsovervåking og justering av kjemiske doser.

En omfattende analyse av de eksisterende prosessene, utstyret og infrastrukturen på Åse WWTP ble utført. Ved å bruke driftslogger og sensordata fra anlegget ble variabler som påvirker den kjemiske renseprosessen identifisert. En LSTM-modell ble deretter trent og validert på disse dataene for å predikere og optimalisere doser av PAX og polymer under varierende forhold.

Resultatene viser at LSTM-modellen, sammen med sensorjusteringer, betydelig forbedrer effektiviteten av koagulerings- og flokkuleringsprosessen. LSTM-modellen oppnår en prediksjonsnøyaktighet på 94,4% for PAX-33 og 77,2% for polymer doseringer. Videre viser implementering av avanserte flerparametersensorer forbedring i disse prediksjonsnøyaktighetene og legge grunnlaget for et helautomatisert doseringssystem, noe som fører til en effektiv renseprosess, reduserte kostnader og lavere utslipp.

Økonomisk sett, med tanke på en avløpssensors levetid på 5 til 10 år, gir nettonåverdien (NPV) over en 10-års periode et estimat på 1 MNOK, gitt en årlig besparelse på 140 000 NOK, en opprinnelig investering på 165 000 NOK, og en diskonteringsrente på 3%. Denne positive NPV indikerer at prosjektet vil gi nettogevinst over denne perioden, spesielt med tanke på den potensielle 10% årlige kjemiske besparelsen fra opptimaliseringen.

Oppgaven bidrar med verdifulle innsikter for avløpsvannbehandlingsindustrien og understreker fordelene ved å utnytte maskinlæring og sensorjusteringer for å optimalisere avløpsvannbehandlingsprosesser. Den markerer et skritt fremover i søket etter innovative løsninger for å forbedre ytelsen til avløpsrenseanlegg over hele verden, og fremmer dermed bærekraft og motstandsdyktighet for urbane vannsystemer på lang sikt.

Table of contents

List of Figures

List of Tables

List of Abbreviations

WWTP	Wastewater treatment plant
PAX	Polyaluminum chloride
NOM	Natural Organic Matter
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
tanh	Hyperbolic tangent function
FF	Feed-Forward
FB	Feedback
PE	Population equivalent
BOD5	Biological Oxygen Demand during a period of 5 days
COD	Chemical Oxygen Demand
TSS	Total Suspended Solids
Total-P	Total phosphor
OLS	Ordinary Least Squares regression
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
R^2	Coefficient of Determination
NPV	Net Present Value

1 Introduction

1.1 Background and motivation

Wastewater treatment is crucial to public health and environmental protection, aligning with the United Nations' Sustainable Development Goals (SDGs) "Clean Water and Sanitation" and "Good Health and Well-being" (United Nations, 2015). Centralized wastewater treatment plants (WWTPs) remove contaminants before releasing treated water back into the environment. However, improving efficiency and sustainability of traditional processes presents significant challenges for WWTPs globally.

These challenges involve the effective removal of diverse contaminants, including organic and inorganic compounds, pathogens, and nutrients. Addressing this requires complex, energy-intensive processes, which can contribute to greenhouse gas emissions, a concern for SDG 13, "Climate Action" (United Nations, 2015).

Balancing treatment cost with regulatory compliance is another challenge. WWTPs must consider these costs while also recovering valuable resources from wastewater, contributing to SDG 12, "Responsible Consumption and Production" (United Nations, 2015).

In the European Union (EU) and Norway, the challenge of wastewater management is heightened due to stringent discharge regulations. These regulations aim to further protect aquatic ecosystems but require WWTPs to upgrade their treatment processes (Eurostat, 2021).

To navigate these challenges, WWTPs are adopting innovative technologies, including machine learning and advanced sensors. Specifically, Long Short-Term Memory (LSTM) models have been chosen for their ability to analyze large amounts of time-series data, essential in optimizing chemical dosages in wastewater treatment. The Åse wastewater treatment plant in Ålesund aims to utilize data analysis, sensor adjustment, and LSTM machine learning to improve its coagulation process.

Coagulation is a vital stage in wastewater treatment, involving the use of coagulants like PAX and polymers to form flocs from suspended particles in the wastewater, making them easier to remove. Optimizing coagulation requires careful dosage control, as an incorrect dose can impair treatment efficiency (Zhang et al., 2016).

1.2 Problem statement and research objectives

1.2.1 Åse Wastewater Treatment Plant

Located in the eastern region of Old Ålesund, Norway, the Åse Wastewater Treatment Plant (WWTP), also known as RA4, has been a stalwart in managing the local community's wastewater since 1990. The facility is engineered to handle a maximum population equivalent (PE) of 25,000 and a peak wastewater flow rate of 248 liters per second.

On average, the facility processes about 103.78 liters of wastewater per second, which is equivalent to approximately 9,000 cubic meters per day. Due to aging infrastructure and high infiltration level, the wastewater inflow exhibits significant seasonal fluctuations. During the drier periods, the inflow can dip to 37 liters per second, whereas during the wetter seasons, it can surge to 233 liters per second. These substantial variations highlight the critical need for adaptability and flexibility in the plant's treatment processes, ensuring optimal performance regardless of seasonal changes.

In 2018, the plant underwent significant upgrades, primarily involving modifications to the chemical treatment process. The former method, which entailed the use of lime mixed with seawater, was superseded using PAX-33 and Superfloc A-1883 polymer. The driving force behind this upgrade was the pursuit of enhanced treatment efficiency, increased capacity, odor elimination, and the resolution of specific operational challenges. Despite these improvements, several optimization opportunities remain. These include refining the location of the dosing point on each individual line and improving the control over chemical dosing. (Kemira, 2017)

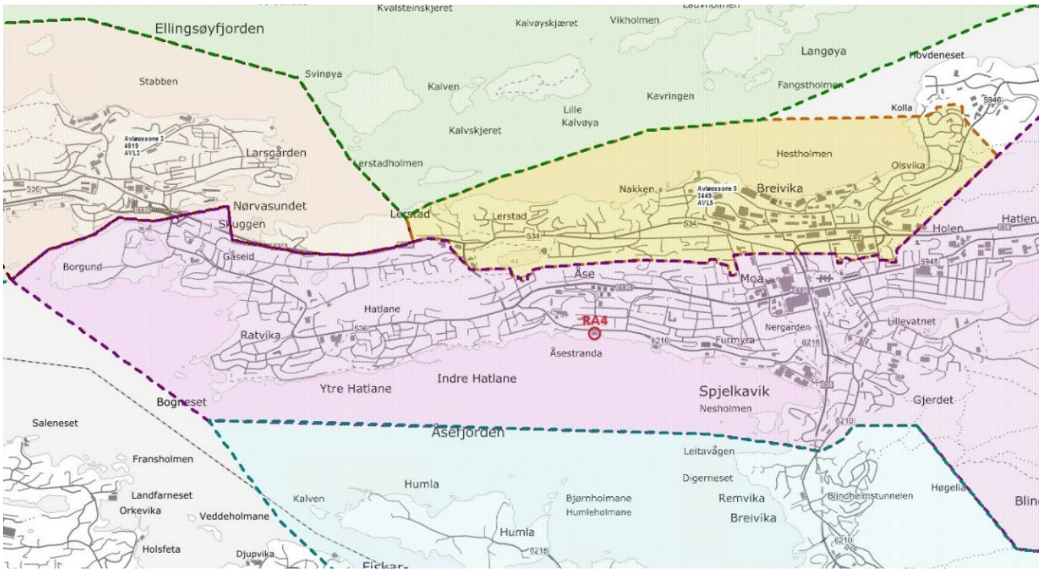


Figure 1. Location and treatment zone of RA4, Åse WWTP in Ålesund, Norway (Gemini VA Ålesund)

1.2.1.1 Composition of wastewater and discharge requirements

As per the municipal records, the Åse Wastewater Treatment Plant (WWTP) handles a substantial quantity of wastewater - approximately 14,659 Population Equivalent (PE) from residential sources. This is further augmented by an additional 7,500 PE originating from industrial establishments, schools, institutions, and restaurants.

Significantly, the local hospital, which is connected to the plant, does not conduct its own pre-treatment procedures to remove pharmaceutical contaminants prior to discharging its wastewater into the system. This presents a potential challenge as the Åse WWTP currently doesn't have specific measures in place to treat these pharmaceutical pollutants.

While this isn't a current requirement for the plant, it is worth noting that the scenario may change in the future. The county municipality may enforce stricter regulations requiring the removal of such contaminants. This would necessitate an upgrade to the facility's existing treatment processes. Moreover, the discharge of untreated pharmaceuticals poses a looming threat to the environment, contributing to ecological degradation. Hence, it's imperative to consider these aspects in the plant's long-term strategic planning and policymaking.

To adhere to the stringent discharge regulations, the WWTP must routinely conduct sampling protocols and provide comprehensive reports on treatment quality. These reports particularly emphasize the removal efficiency of parameters like Biochemical Oxygen Demand (BOD5), Total Suspended Solids (TSS), total phosphor, along with specific heavy metals. This rigorous monitoring is integral to ensuring the treated water meets or exceeds the required environmental and public health standards.

As per the discharge permit issued (Møre og Romsdal County Municipality, 2016), the Åse WWTP must achieve a phosphorus removal rate of at least 90%, TSS removal of 50%, and BOD5 removal of 20%.

1.2.1.2 Recipient

The Åse Wastewater Treatment Plant (WWTP) discharges its treated wastewater and any overflow into Åsefjorden, a body of water located approximately 70 meters off the coast and at a depth of 27 meters. According to the most recent assessment of the receiving waters, updated on January 10th, 2023, Åsefjorden is classified as a "less sensitive" recipient, exhibiting a "very good" ecological status but a "poor" chemical state (Akvaplan-niva AS, 2023). The Norwegian Pollution Control Regulations, known as "Forurensningsforskriften," dictate the treatment processes necessary for discharging treated wastewater based on the classification of the recipient.

The report reveals a noticeable deterioration in the condition of the recipient since the last inspection in 2018. If this decline persists, the facility may face future challenges in meeting discharge requirements. This situation may necessitate increased investment in infrastructure upgrades to maintain compliance. Therefore, the need to optimize the treatment process is evident. Doing so will enhance the quality of the discharged wastewater, thus mitigating further degradation of Åsefjorden's condition.

1.2.1.3 Treatment processes, sensors and sampling points

As depicted in Figure 2 and Figure 3, Åse WWTP employs a three-stage treatment process, consisting of preliminary, primary, and secondary treatments. In the preliminary stage, coarse and large suspended solids are removed. The primary treatment involves mechanical screening, which removes finer particles. During the secondary treatment or chemical clarification process, chemicals such as PAX-33 and polymers are added to promote sedimentation. The treated and dewatered sludges are subsequently sent for incineration or disposal at a landfill.

The plant utilizes a monitoring program called “Citect” which managed by "Guard Automation", with Figure 2 illustrating the program's interface. It is a type of supervisory control and data acquisition (SCADA) system, which collects and records data from the entire facility while enabling real-time control. As shown in Figure 3, sensors gather data on parameters like flow rate, temperature, pH, and turbidity during the treatment processes. Additionally, operational data like pump flow of PAX-33 dosage, is recorded.

As illustrated in Figure 3, there are two sampling points within the facility: one is situated after the influent but before mechanical screening, and the other is located before the effluent following sedimentation. As mentioned in 1.2.1.1, the WWTP is required to report contaminant removal on a regular basis. These sampling points serve this purpose and represent the only data sources for contaminant removal within the facility. Notably, there are no sensors installed at the plant that specifically measure concentration of contaminants.

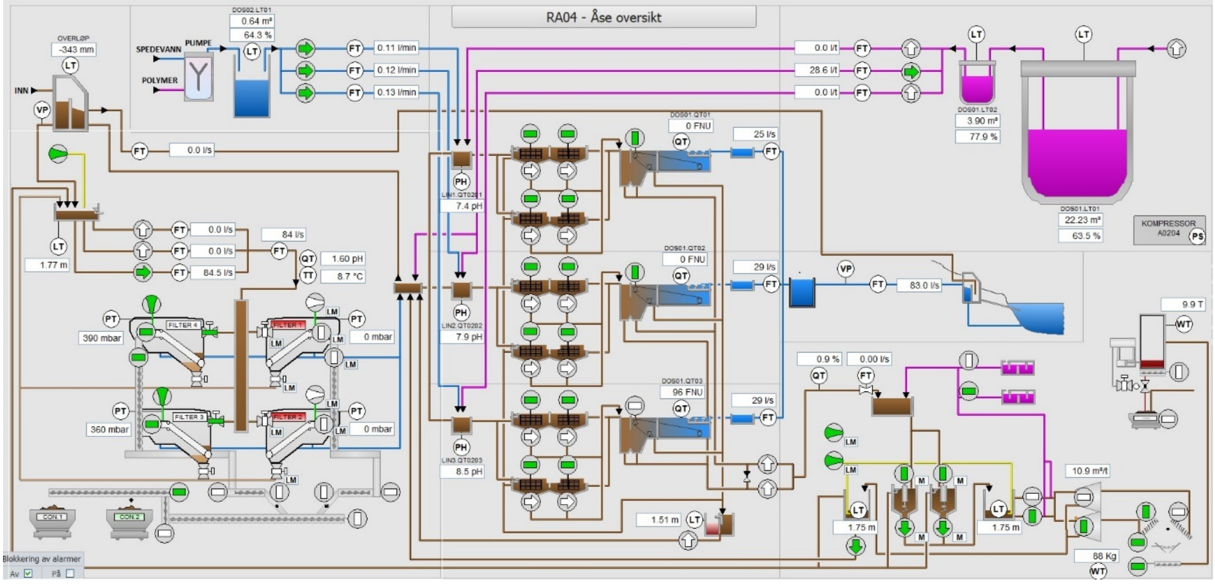


Figure 2. Åse WWTP treatment processes in interface of Citect

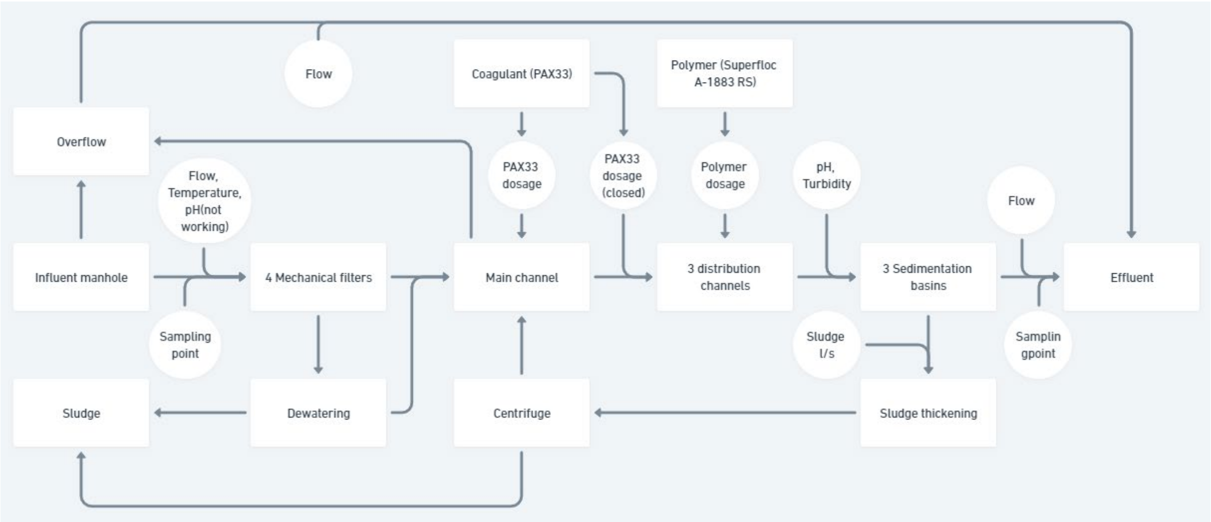


Figure 3. Simplified flow chart of Åse WWTP treatment processes with sensor and sampling locations (Circle)

1.2.1.4 PAX-33 Utilization and Expenditure

Plant operators at the Åse wastewater treatment facility have conducted several optimization processes on chemical dosing since the upgrade in 2018. This has resulted in a significant reduction in the daily PAX-33 consumption, decreasing from an initial 1500 kilograms to the current usage of approximately 900 kilograms per day.

As per the operational data, daily consumption of the coagulant PAX-33 typically falls within the range of 800 to 1000 kilograms. Annually, this constitutes an estimated usage of about 330 metric tons. A critical concern that stems from the examination of supplier invoices is the escalating cost of PAX-33, which has been experiencing an annual surge of approximately 40% (as evidenced in Figure 4).

At the current rate, the municipality is projected to allocate approximately 1.5 million NOK each year for PAX-33 acquisition alone. When combined with the cost of polymer, this total expenditure is expected to reach around 2.3 million NOK by 2023. This substantial financial burden underscores the urgency of identifying and implementing effective cost reduction strategies.

However, in our pursuit of financial prudence, we must ensure that the treatment process's efficiency remains uncompromised. Thus, the balance between cost-effective operations and optimal wastewater treatment results forms the core of this analysis. This balance is not only necessary for the immediate budgetary considerations but also integral to the long-term sustainable management of resources at the Åse wastewater treatment plant.

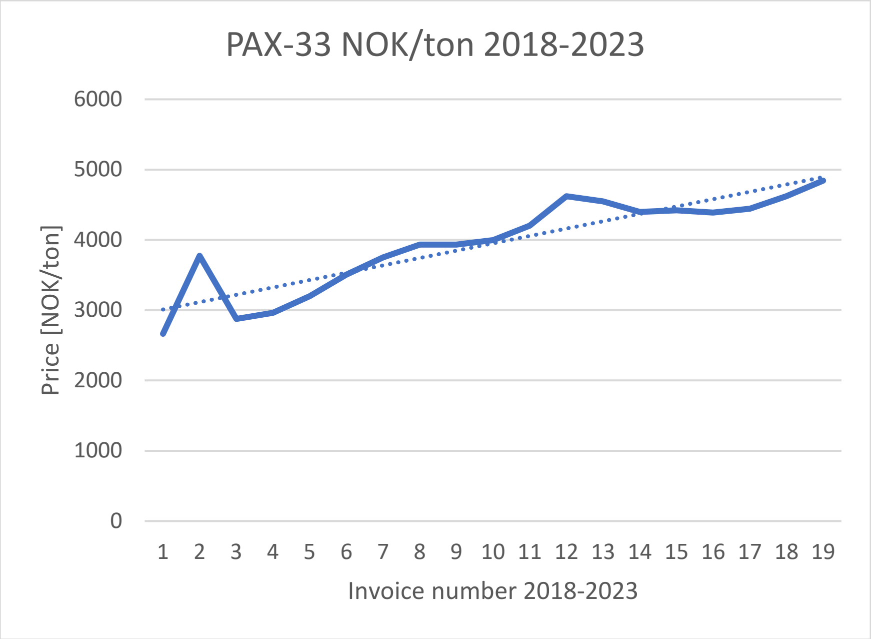


Figure 4. PAX-33 price escalation from Jan. 2018 to May. 2023

1.2.2 Research objectives

The primary objective of this study is to investigate the potential of machine learning, in conjunction with sensor adjustment, to optimize the coagulation process at the Åse WWTP in Ålesund, Norway. Specifically, the research aims to accomplish the following objectives:

1. Determine the optimal dosage of PAX and polymer for effective removal of suspended solids and other contaminants by utilizing machine learning techniques, particularly LSTM models.
2. Evaluate the application of sensors to continuously monitor critical process variables, such as BOD, COD, TSS, phosphor and turbidity, enabling real-time adjustments of chemical dosages to ensure optimal performance.
3. Assess the impact of the proposed optimization methods on the efficiency, cost-effectiveness, and sustainability of the coagulation and flocculation process.
4. Contribute to the ongoing effort to enhance the resilience and sustainability of urban water systems by demonstrating the advantages of integrating machine learning and sensor adjustment in wastewater treatment processes.

The outcomes of this study hold significant implications for the wastewater treatment industry. By employing machine learning, this research aims to improve the efficiency and effectiveness of the coagulation and flocculation process, which can lead to recovery of valuable resources from wastewater, reduction of costs and unnecessary emissions. Moreover, this study contributes to the broader endeavor to advance the sustainability and resilience of urban water systems.

1.3 Scope and limitations

The primary focus of this research is to enhance the performance and efficiency of an existing wastewater treatment plant (WWTP) by conducting a comprehensive analysis of the facility and developing a machine learning algorithm to predict and optimize the chemical dosage used in the treatment process. To achieve this objective, the following tasks will be undertaken:

1. Inspect the WWTP: A thorough inspection of Åse WWTP will be carried out to gain an in-depth understanding of the facility's current processes, equipment, and infrastructure. This inspection will also identify any limitations or challenges faced by the WWTP, which can inform subsequent analyses and optimization efforts.
2. Propose sensor adjustments and purchase of new sensors: Accurate monitoring and control of various processes are crucial to a WWTP's performance. Therefore, this research will recommend adjustments to existing sensors and the acquisition of new sensors to enhance the monitoring and control of the WWTP. This may involve identifying gaps in the current sensor network and suggesting additional sensors to provide valuable data.
3. Data analysis and identification of correlated variables: The performance of a WWTP is influenced by several factors, including influent characteristics, operational parameters, and chemical dosage. This research will employ statistical analyses and data visualization techniques to identify patterns and trends in the data, as well as determine which variables have the strongest relationships with the outputs.
4. Develop a machine learning algorithm: Machine learning techniques will be utilized to create a model capable of predicting and optimizing the dosage of chemicals used in the treatment process. The identified correlated variables will serve as inputs for the algorithm, and models will be developed to predict the optimal chemical dosage for specific conditions.

Despite these objectives, the study may encounter certain limitations, which will be acknowledged and discussed throughout the thesis. These could include limited access to historical data or restrictions in data granularity, potential unavailability of resources, budget constraints for implementing new sensors or adjustments to the WWTP, and the generalizability of the findings, as the optimization strategies might be tailored specifically to the Åse WWTP and may not be directly applicable to other plants with different configurations or challenges. The implications of these limitations on the research findings and their generalizability will be addressed in the Discussion chapter.

1.4 Structure of the thesis

This thesis is organized into six chapters, providing a logical progression from the introduction and background to the conclusion and recommendations. The structure of the thesis is as follows:

1. Introduction: This chapter presents the background and motivation for the research, highlights the importance of wastewater treatment, and discusses the need for process optimization. The problem statement, research objectives, and scope and limitations are outlined, followed by a brief overview of the Åse WWTP case study to provide context for the study.
2. Literature Review: This chapter reviews relevant literature on wastewater treatment processes, process optimization, sensor adjustment, and machine learning techniques, with a focus on the rationale behind choosing LSTM models and their application in wastewater treatment.
3. Methodology: In this chapter, the research design is detailed, including data collection, data preprocessing, and machine learning model development. The sensor adjustment strategy is explained, and the process optimization implementation is described. The evaluation criteria and performance metrics are also presented.
4. Results: This chapter presents the findings of the study, including data analysis outcomes, the performance of the LSTM model, the results of sensor adjustments, and the impact of process optimization on coagulation, particularly the optimal dosage of PAX and polymer.
5. Discussion: In this chapter, the results are interpreted, and their practical implications are discussed. Recommendations for Åse WWTP and the broader wastewater treatment industry are provided. Challenges and limitations encountered during the study are addressed, and potential future research directions are suggested.
6. Conclusion: The final chapter summarizes the main findings of the research, highlighting its contributions to the knowledge of process optimization in wastewater treatment plants. The chapter concludes with some final remarks on the potential of machine learning and sensor adjustment in improving the performance and efficiency of wastewater treatment processes.

2 Literature Review

2.1 Wastewater treatment processes

2.1.1 Overview

Wastewater treatment is a critical process for protecting public health, preserving the environment, and maintaining the sustainability of water resources. Wastewater contains a wide range of contaminants, including organic and inorganic compounds, nutrients, pathogens, and suspended solids, which must be removed before the treated water is discharged back into the environment or reused (Metcalf & Eddy, 2014). The primary goals of wastewater treatment are to minimize the adverse environmental impacts of wastewater discharge, meet regulatory requirements, and recover valuable resources such as water, nutrients, and energy.

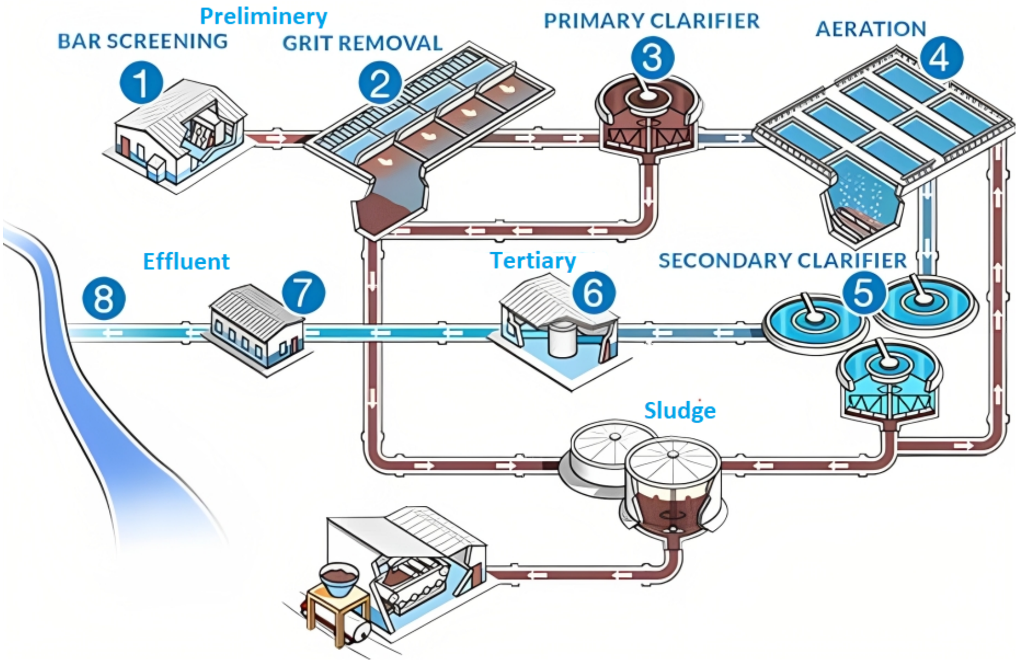


Figure 5. Overview of a typical wastewater treatment plant processes (Cole-Parmer)

2.1.2 Treatment Stages

As illustrated in Figure 5, wastewater treatment typically involves multiple stages, including preliminary, primary, secondary, and tertiary treatment. Each stage targets specific contaminants and has its unique objectives and processes.

2.1.2.1 Preliminary Treatment

Preliminary treatment aims to remove large debris, grit, and grease from the incoming wastewater. This step usually involves screening, grit removal, and grease separation. The primary objective is to protect downstream equipment and processes from damage and excessive wear (Tchobanoglous, Burton, & Stensel, 2003).

2.1.2.2 Primary Treatment

Primary treatment involves the removal of settleable and floatable solids, typically through sedimentation or flotation processes. Mechanical screening is a common method used in primary treatment to remove suspended solids and other particles from the wastewater (Henze et al., 2008). At Åse WWTP, mechanical screening is employed as the primary treatment method.

2.1.2.3 Secondary Treatment

Secondary treatment aims to remove biodegradable organic matter and nutrients from wastewater using biological and chemical processes. In the case of Åse WWTP, chemical treatment is employed as the secondary treatment method. Chemical treatment involves the use of coagulants and flocculants, such as PAX and polymer, to facilitate the removal of suspended solids, organic matter, and other contaminants through coagulation and flocculation processes (Metcalf & Eddy, 2014). The treatment is especially effective at removing phosphor from wastewater.

2.1.2.4 Tertiary Treatment

Tertiary treatment, also known as advanced treatment, targets specific contaminants that are not effectively removed in the primary and secondary treatment stages. These contaminants may include nutrients, heavy metals, and certain organic compounds. Tertiary treatment processes may involve filtration, adsorption, disinfection, or additional chemical treatments (Tchobanoglous et al., 2003). Depending on the specific requirements of the treatment plant, tertiary treatment may or may not be necessary.

2.2 Coagulation and flocculation in wastewater treatment

2.2.1 Principles and Mechanisms

Coagulation and flocculation are essential processes in wastewater treatment for removing suspended solids, organic matter, and other contaminants. Coagulation involves the neutralization of negatively charged particles in wastewater by adding positively charged coagulants, causing the particles to destabilize and aggregate (Metcalf & Eddy, 2014). Flocculation, on the other hand, is the process of forming larger, more stable flocs by adding flocculants that promote the aggregation of destabilized particles. These flocs can then be easily removed through sedimentation or flotation processes (Tchobanoglous et al., 2003).

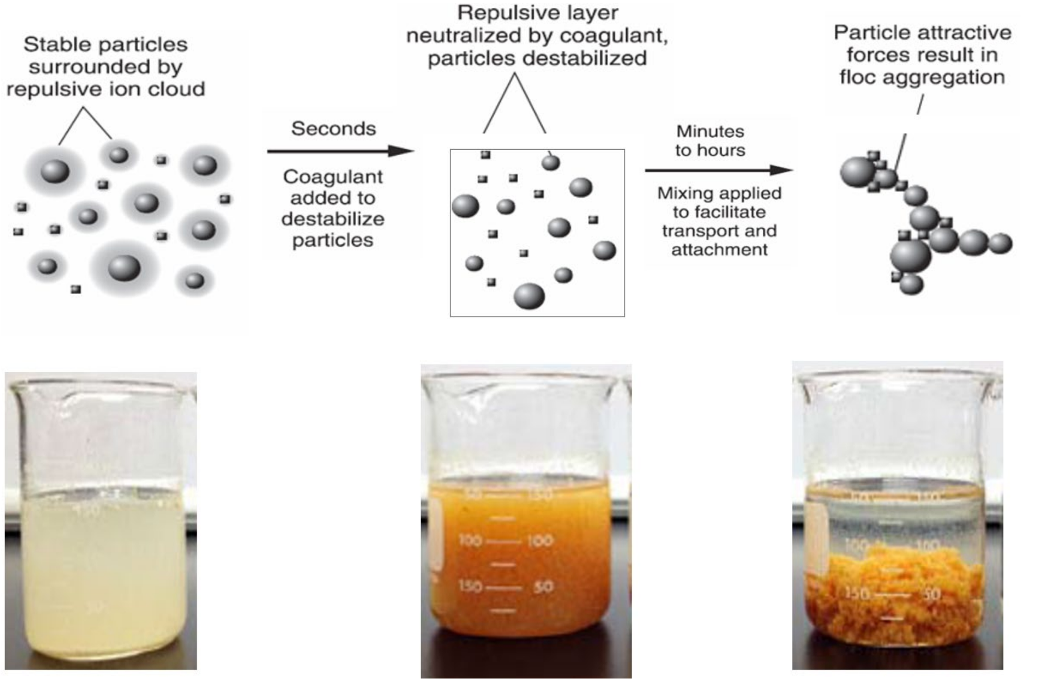

Stable particles surrounded by repulsive ion cloud
Seconds
Coagulant added to destabilize particles
Repulsive layer neutralized by coagulant, particles destabilized
Minutes to hours
Mixing applied to facilitate transport and attachment
Particle attractive forces result in floc aggregation

Figure 6. Explanation of Coagulation and Flocculation process (Course presentation)

2.2.2 Factors Affecting Coagulation and Flocculation

Several factors influence the effectiveness of coagulation and flocculation, including pH, temperature, mixing conditions and more:

- pH: The pH of the wastewater affects the charge of particles and the solubility of coagulants, thus influencing their effectiveness in destabilizing particles. Optimal pH ranges for coagulation and flocculation vary depending on the specific coagulant and flocculant used (Metcalf & Eddy, 2014).
- Temperature: Temperature influences the reaction rates of coagulation and flocculation processes. Higher temperatures typically lead to faster reaction rates, while lower temperatures can result in slower reactions and the formation of weaker flocs (Tchobanoglous et al., 2003).
- Mixing conditions: Proper mixing is crucial for the effective dispersion of coagulants and flocculants and the formation of stable flocs. Insufficient mixing can result in inadequate contact between particles and coagulants, while excessive mixing can cause the break-up of flocs (Metcalf & Eddy, 2014).
- Coagulant and flocculant dosage: The concentration of coagulant and flocculant used in the process plays a critical role in the effectiveness of coagulation and flocculation. Overdosing or underdosing can lead to suboptimal performance (Jiang et al., 2010).
- Coagulant and flocculant type: Different types of coagulants and flocculants have varying characteristics and can affect the treatment efficiency differently. The choice of the appropriate coagulant and flocculant depends on the specific characteristics of the wastewater and the desired treatment outcomes (Zouboulis & Traskas, 2008).
- Particle characteristics: The size, shape, and surface charge of suspended particles in wastewater can significantly affect coagulation and flocculation. Particle properties influence the aggregation and settling processes, as well as the interaction with the coagulants and flocculants (Duan & Gregory, 2003).

- Ionic strength: The presence of various ions in the wastewater can influence the effectiveness of coagulation and flocculation. High ionic strength can suppress the repulsive forces between particles and improve coagulation, while low ionic strength can result in reduced coagulation efficiency (Jarvis et al., 2005).
- Presence of natural organic matter (NOM): NOM can interfere with coagulation and flocculation processes by competing for binding sites with the coagulants or by forming complexes with metal ions. This can result in reduced pollutant removal efficiency (Matilainen et al., 2010).

2.2.3 Role of PAX and Polymer as Coagulant and Flocculant

2.2.3.1 PAX-33 as Coagulant

Polyaluminum chloride (PAX) is a coagulant commonly used in wastewater treatment due to its capacity to destabilize particles and promote floc formation. Compared to other coagulants like aluminum sulfate (alum) and ferric chloride, PAX offers several benefits, including a lower dosage requirement, efficient performance over a broad pH range, and reduced sludge production (MWH, 2012). PAX has demonstrated effectiveness in eliminating suspended solids, organic matter, and other pollutants from wastewater (Metcalf & Eddy, 2014).

The Åse WWTP uses PAX-33, supplied by Kemira as its coagulant. This mixture contains a 30-40% concentration of polyaluminum chloride and 1-5% of Iron(III) chloride. The inclusion of Iron(III) chloride in PAX-33 enhances contaminant removal, particularly targeting sulfides that cause unpleasant odors. However, the use of PAX also presents some challenges. These include potential aluminum residues in the treated water, which may have environmental and health implications, and the possible requirements for pH adjustment and operational optimization before use, adding to the process complexity (Zouboulis et al., 2004; Kemira, 2017).

2.2.3.2 Polymer Superfloc A-1883 as Flocculant

Flocculants, such as polymers, play a vital role in wastewater treatment by enhancing the formation of larger, more stable flocs that facilitate efficient solid-liquid separation. Polymers are categorized into natural, synthetic, anionic, cationic, and nonionic types based on their origin and charge (Bratby, 2016). The selection of an appropriate polymer depends on the specific characteristics of the wastewater and the treatment objectives. Polymers are known to improve the efficiency of solid-liquid separation processes, resulting in clearer effluent and more easily dewatered sludge (Metcalf & Eddy, 2014).

Superfloc A-1883, a product of Kemira, is the chosen flocculant at the Åse Wastewater Treatment Plant (WWTP). Its composition includes various hydrocarbons, ethoxylated alcohols, and ammonium acrylate, contributing to its effective flocculation properties. However, care must be taken when using such polymers as they can increase the viscosity of the wastewater, potentially causing difficulties in subsequent treatment stages. Moreover, they may not be suitable for all wastewater types, necessitating careful selection (Bolto & Gregory, 2007).

2.3 Traditional Process optimization in wastewater treatment

2.3.1 Importance

Process optimization is of paramount importance in wastewater treatment, as it enables treatment plants to enhance efficiency, reduce costs, and comply with environmental regulations. By optimizing various aspects of the treatment process, plants can minimize energy consumption, chemical usage, and sludge production while maximizing pollutant removal and resource recovery (Bixio et al., 2005). Improved efficiency not only results in cost savings but also contributes to environmental protection by reducing the plant's carbon footprint and the release of harmful substances into the environment.

2.3.2 Optimization Techniques

There exists an array of optimization techniques applied to wastewater treatment processes, spanning from conventional mathematical approaches to contemporary computational methodologies. Each comes with unique benefits and limitations that can affect their suitability for a given context.

2.3.2.1 Jar tests

Jar tests are empirical laboratory methods employed in wastewater treatment to optimize coagulation and flocculation processes. They help determine the ideal chemical dosages necessary for contaminant removal (Metcalf & Eddy, 2014). The procedure involves combining wastewater samples with different concentrations of coagulants or flocculants in individual jars. Following a settling period, the clarity of the treated water is assessed to pinpoint the most effective chemical dosages for the specific wastewater (APHA, AWWA, & WEF, 2017). Despite their effectiveness, jar tests are time-intensive and less suited to managing rapid influent wastewater variations, making them better suited to the initial stages of a new facility (Liu, 2016). After the startup phase, more advanced techniques supplemented by sensor technology are often recommended.

2.3.2.2 Response Surface Methodology (RSM)

RSM is a statistical technique used to model and analyze complex processes, shedding light on the relationships between several input factors and one or more response variables. It is especially valuable in identifying optimal process conditions and understanding the interplay between different factors (Myers et al., 2016). However, RSM might be challenging to apply when dealing with nonlinear processes or when the number of input parameters becomes very large.

2.3.2.3 Genetic Algorithms (GAs)

Inspired by the process of natural selection, GAs are a form of evolutionary algorithms. They are designed to find approximate solutions to complex optimization problems by continually evolving a population of potential solutions (Yuan et al., 2009). Despite their power and versatility, GAs can sometimes fall into the trap of premature convergence, finding a suboptimal solution instead of the global optimum.

2.3.2.4 Linear Programming (LP)

LP is a mathematical optimization technique that aims to find the optimal solution to a problem while adhering to a set of linear constraints. This technique is often used in wastewater treatment for determining optimal treatment strategies, resource allocation, and treatment plant design (Poch et al., 2004). However, LP requires that all the relationships in the model be linear, which may not always accurately reflect the reality of complex wastewater treatment processes.

2.4 Machine learning in process optimization

Machine learning, an innovative branch of artificial intelligence, has become an indispensable tool for process optimization in wastewater treatment plants. Several machine learning techniques, each equipped with a unique architecture and functionality, have been widely utilized in this field.

2.4.1 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are inspired by the functioning of biological brains and are structured as a network of interconnected layers of neurons. They utilize a series of algorithms to recognize underlying relationships in a dataset through a process that mirrors the way the human brain operates. As shown in Figure 7, ANNs consist of an input layer, where the initial data is presented; one or more hidden layers, where the computations are performed; and an output layer, which delivers the outcome. The nodes in these layers mimic biological neurons by receiving input and passing a weighted sum of these inputs through an activation function (Equation 1). This output is then sent to the next layer. This unique structure empowers ANNs to model complex, non-linear relationships between input and output variables (Haykin, 1999). Despite their power, ANNs are often criticized for their lack of interpretability – their 'black box' nature – and their need for large datasets to avoid overfitting (Kuhn & Johnson, 2013).

$$O = f(\sum_{i=1}^{n} W_i * I_i + b)$$

Equation 1

Where:

- O represents the output of the neural network.
- f denotes the activation function applied to the sum of weighted inputs and bias.
- i is the index that iterates from 1 to n, representing the individual inputs.
- W_i represents the weight associated with the input I_i.
- I_i is the i-th input to the neural network.
- b represents the bias term.

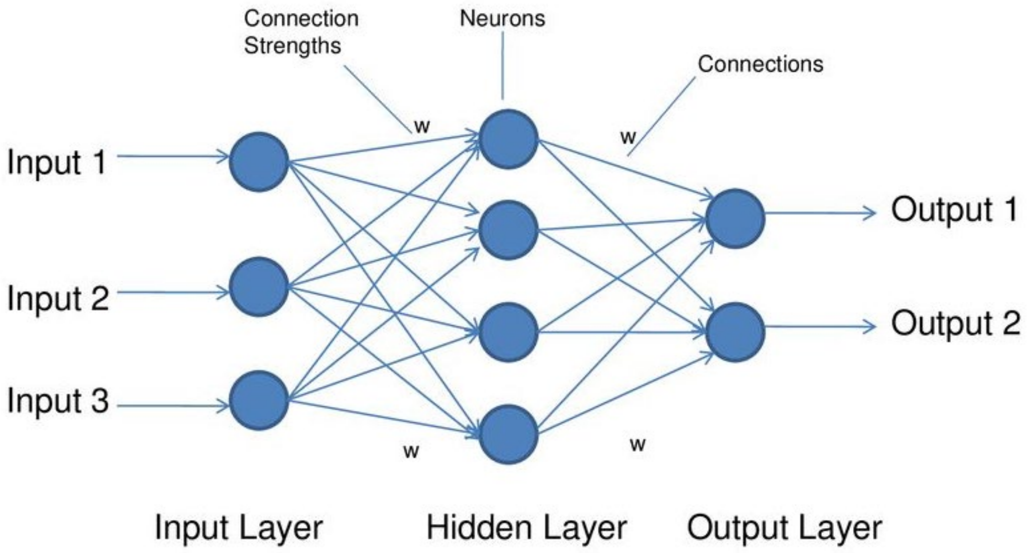


Figure 7. Sample of an ANN architecture (Donald O'Connor')

2.4.2 Decision Trees

Decision Trees are simple yet powerful tools for non-parametric supervised learning. As illustrated in Figure 8, A decision tree uses a tree-like model of decisions where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label (decision taken after computing all attributes). The paths from the root to the leaf represent classification rules. Unlike ANN, decision trees does not use any activation function. Despite their simplicity and interpretability, it tend to overfit on data with many features and are sensitive to small changes in the data, leading to different splits and impacting the stability of the model (Quinlan, 1986).

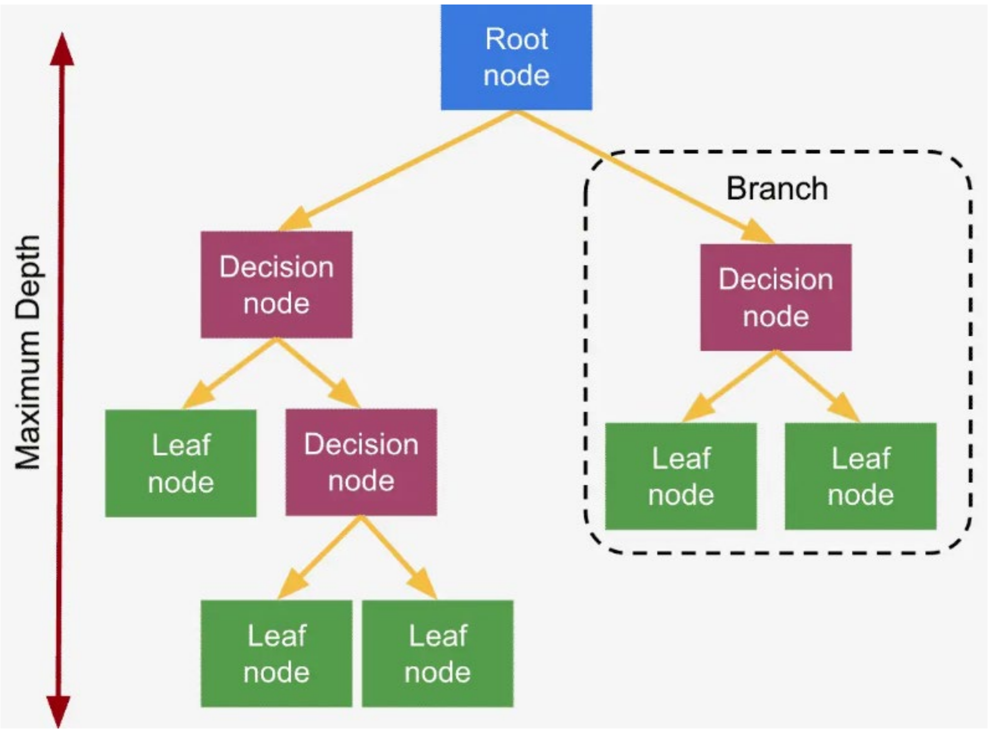


Figure 8. Sample of Decision Tree architecture (jcchouinard)

2.4.3 Random Forests

Random Forests are an ensemble learning technique that constructs multiple decision trees during training and outputs the class that is the mode of the classes output by individual trees (Figure 9). They integrate two key concepts: bagging (bootstrap aggregation) and feature randomness. Bagging helps reduce the variance of the prediction by generating additional data for training from the original dataset, while feature randomness chooses a subset of features at each candidate split in the learning process. As shown in Equation 2, by averaging predicted values from all decision trees, random forest can predict values for regression problem. This unique architecture offers a powerful model for prediction and decision-making, providing better accuracy and robustness against overfitting (Breiman, 2001). However, like ANNs, Random Forests are also criticized for their lack of interpretability and computational intensity.

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x')$$

Equation 2

Where:

- $\hat{f}$ represents the predicted value for a given input instance x'
- B is the total number of decision trees in the random forest ensemble
- $f_b(x')$ is the prediction made by the individual decision tree b for the input instance x'

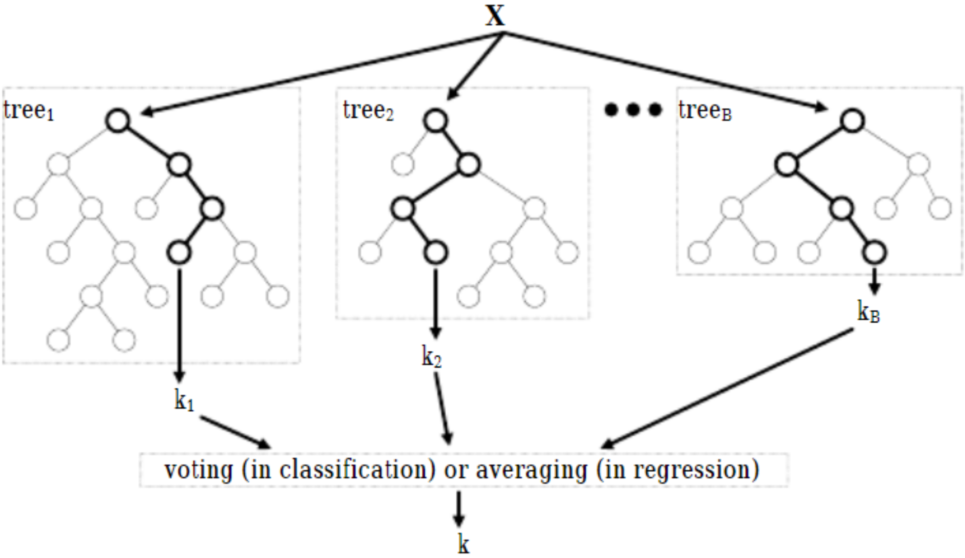


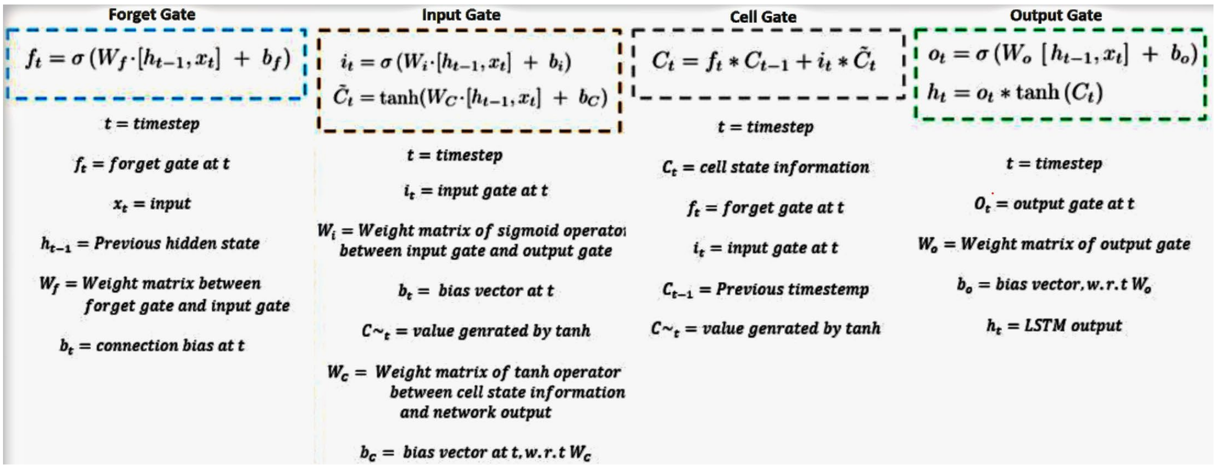
Figure 9. Sample of Random Forests architecture (Verikas et al., 2016)

2.4.4 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) models, a type of Recurrent Neural Network (RNN), are specifically designed to effectively model time-series data. As depicted in Figure 10 and Figure 11, this is accomplished through an architecture that comprises four key components, each of which performs a specific function.

1. Forget Gate: This determines the importance of information, deciding which elements should be kept and which ones can be discarded. The current input (X(t)) and the previous hidden state (h(t-1)) are passed through a sigmoid function, resulting in values between 0 and 1. These values are indicative of the relevance of the previous output (f(t)), which are later utilized in point-by-point multiplication with the cell state.
2. Input Gate: This gate functions to update the cell status. Both the current input (X(t)) and the previous hidden state (h(t-1)) go through a sigmoid function, helping to ascertain the significance of each component. Simultaneously, this information is passed through a tanh function, producing values between -1 and 1. The outputs from these two activation functions are then ready for point-by-point multiplication.
3. Cell State: This component merges information from the forget gate and input gate. The previous cell state (C(t-1)) is multiplied by the forget vector (f(t)). If the resultant value is 0, the corresponding values are dropped from the cell state. The input vector (i(t)) is then added in a point-by-point manner, updating the cell state (C(t)).
4. Output Gate: This gate is responsible for deciding the next hidden state, which retains information about previous inputs. The current state and the previous hidden state are processed through a sigmoid function. Additionally, the new cell state is passed through a tanh function. The outputs from both functions are then multiplied point-by-point. The resultant value influences the information carried by the hidden state, which is subsequently used for making predictions.

Once the new cell state and hidden state are computed, they are carried over to the next time step, perpetuating the LSTM process. This unique architecture allows LSTM models to capture long-term dependencies in sequences of data points over extended periods, making them particularly well-suited for modeling complex and dynamic processes such as those found in Wastewater Treatment Plants (WWTPs), where variations in influent characteristics and operational parameters are common (Hochreiter & Schmidhuber, 1997).



Forget Gate

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

$t = timestep$

$f_t = forget\ gate\ at\ t$

$x_t = input$

$h_{t-1} = Previous\ hidden\ state$

$W_f = Weight\ matrix\ between\ forget\ gate\ and\ input\ gate$

$b_t = connection\ bias\ at\ t$

Input Gate

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

$t = timestep$

$i_t = input\ gate\ at\ t$

$W_i = Weight\ matrix\ of\ sigmoid\ operato$ $between\ input\ gate\ and\ output\ gate$

$b_t = \ bias\ vector\ at\ t$

$C{\sim}_t = value\ genrated\ by\ tanh$

$W_C = \ Weight\ matrix\ of\ tanh\ operator\ between\ cell\ state\ information\ and\ network\ output$

$b_C = \ bias\ vector\ at\ t, w.r.t\ W_C$

Cell Gate

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$t = timestep$

$C_t = cell\ state\ information$

$f_t = forget\ gate\ at\ t$

$i_t = input\ gate\ at\ t$

$C_{t-1} = Previous\ timestemp$

$C{\sim}_t = value\ genrated\ by\ tanh$

Output Gate

$$o_t = \sigma\left(W_o \; [h_{t-1}, x_t] \; + \; b_o\right)$$
$$h_t = o_t * \tanh(C_t)$$

$t = timestep$

$O_t = output\ gate\ at\ t$

$W_o = Weight\ matrix\ of\ output\ gate$

$b_o = bias\ vector, w.r.t\ W_o$

$h_t = LSTM\ output$

Figure 10. Functions involved in each LSTM key component (Singhal, 2020)

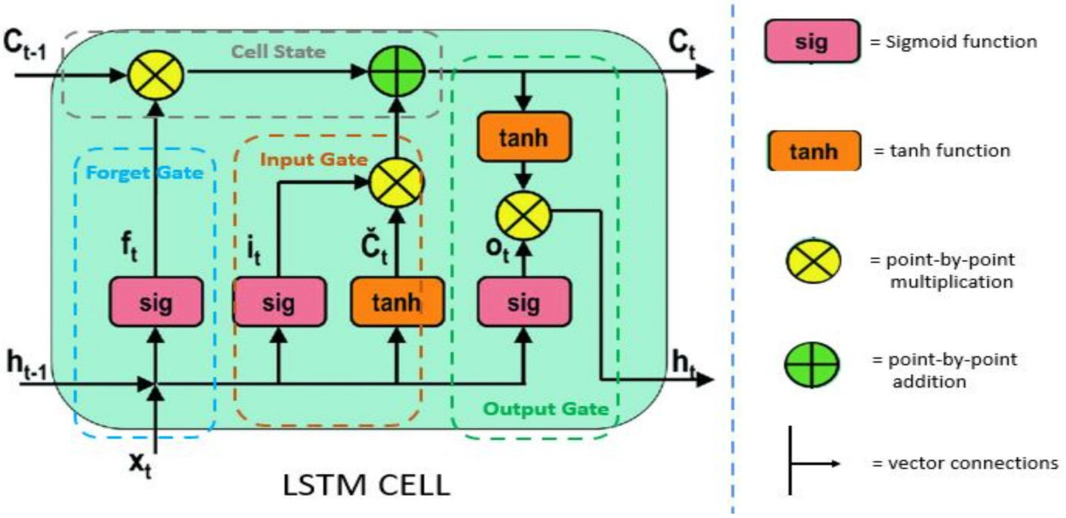


Figure 11. Sample of LSTM architecture (Singhal, 2020)

In conclusion, each of these machine learning techniques holds potential for enhancing the optimization of wastewater treatment processes. The selection of a specific model depends on the problem at hand, the available data, computational resources, and the necessity for model interpretability.

2.5 Data analysis techniques

2.5.1 Descriptive statistics

It is basic data analysis techniques used in wastewater treatment process analysis. Descriptive statistics such as mean, median, mode, standard deviation, variance, and coefficient of variation help summarize and describe the main features of a dataset. Additionally, correlation analysis is discussed as a method for examining the relationships between different variables in the wastewater treatment process (Bhattacharyya & Solomatine, 2005).

2.5.2 Inferential statistical analysis

Inferential statistics provides predictions or inferences about a larger population based on sampled data. A common method used is Ordinary Least Squares (OLS) Regression, which estimates the relationship between a dependent variable and one or more independent variables with the goal of minimizing the sum of the squared differences between observed and predicted values of the dependent variable (Wooldridge, 2015). This technique can be applied in various contexts, including wastewater treatment, where it could, for instance, predict the required amount of a chemical based on factors like volume of wastewater and contaminant concentration. However, assumptions such as linearity, independence, homoscedasticity, and normality must be checked to ensure model validity (Gujarati, 2003).

2.5.3 Advanced data analysis

Following techniques enable a deeper understanding of the underlying patterns and relationships within the data.

- Principal Component Analysis (PCA): PCA is a statistical technique used to reduce the dimensionality of a dataset by identifying the most significant factors or principal components that explain the variance in the data. PCA can be applied to wastewater treatment data to identify the main sources of variability and simplify the analysis (Chen et al., 2014).

- Cluster Analysis: Cluster analysis is a technique used to group similar observations or variables based on their characteristics. In the context of wastewater treatment, cluster analysis can be used to identify groups of similar water samples or operational conditions, which can help optimize treatment strategies (Ebrahimi et al., 2020).
- Time-Series Analysis: Time-series analysis focuses on the study of data points collected at different time intervals to identify trends, patterns, and relationships over time. In wastewater treatment, time-series analysis can be used to analyze the performance of treatment processes, predict future behavior, and identify potential issues or improvements (Quilty & Russell, 2009).

2.5.4 Comparison and Evaluation of Data Analysis Techniques

Each of these data analysis techniques has its strengths and limitations, which should be considered when selecting the most appropriate method for wastewater treatment process analysis. The choice of technique will depend on the specific problem, data characteristics, and desired outcomes. In some cases, combining multiple techniques may provide a more comprehensive understanding of the underlying patterns and relationships within the data. For instance, using both PCA and cluster analysis can help identify important and latent variables, correlations, and groupings within the dataset, while time-series analysis can be employed to study temporal patterns (Bhattacharyya & Solomatine, 2005; Chen et al., 2014; Ebrahimi et al., 2020; Quilty & Russell, 2009).

2.6 Sensor adjustment and monitoring

2.6.1 Importance

The implementation of sensor monitoring in wastewater treatment processes is crucial for ensuring optimal performance, accurate decision-making, and early detection of potential issues. Sensors allow for real-time data acquisition and provide essential information for process control and optimization, ultimately improving the efficiency and effectiveness of the treatment process (Olsson, 2012).

2.6.2 Types of Sensors

In wastewater treatment plants, various sensors are employed to monitor key water quality parameters. Common sensor types include flow rate, pH, turbidity, TSS, dissolved oxygen, phosphorous and conductivity sensors. These sensors facilitate the collection of valuable data, which can be utilized for process control and optimization, ensuring optimal treatment performance (Ratnayaka et al., 2009).

2.6.2.1 Essential sensors

Essential sensors are integral components in the operation of the wastewater treatment plant, focusing on the monitoring of operational variables that are fundamental to the plant's function. These sensors represent the minimum requirements without which the plant cannot operate effectively and safely. They are instrumental in ensuring the reliability, efficiency, and compliance of the wastewater treatment process.

1. Flow sensors: Flow sensors measure the rate at which wastewater enters the treatment facility. Accurate flow measurements are essential for calculating chemical dosages and adjusting process parameters (Metcalf & Eddy, 2014).
2. Temperature sensors: Temperature sensors monitor the temperature of the wastewater, which can influence the efficiency of the treatment process. Maintaining optimal temperatures can enhance the performance of chemical and biological reactions (Grady et al., 2011).
3. pH sensors: pH sensors measure the acidity or alkalinity of the wastewater. Monitoring pH levels is crucial for maintaining optimal conditions for chemical reactions and ensuring compliance with regulatory discharge limits (APHA, AWWA, & WEF, 2017).
4. Chemical dosage sensors: These sensors measure the amounts of chemicals, such as coagulants and flocculants, added to the wastewater. Monitoring chemical dosages can help operators optimize the treatment process and ensure efficient removal of contaminants (Wang et al., 2013).

2.6.2.2 Optional sensors

Optional sensors that monitor contaminant concentration can provide additional information to further enhance the process control and optimization of the coagulation and flocculation process.

1. Turbidity: Monitoring turbidity is crucial because it directly corresponds to the amount of particulate matter in the wastewater. Elevated levels of turbidity can hinder light penetration in water bodies, negatively affecting aquatic life. Furthermore, high turbidity can also indicate the presence of bacteria, viruses, or parasites, which could pose health risks. Thus, by controlling turbidity, we can improve water clarity, safeguard aquatic ecosystems, and mitigate potential health hazards (Xu et al., 2017; Liu, 2016).
2. TSS: Total Suspended Solids (TSS) are particles that are larger than 2 microns found in the water column. High levels of TSS can cause numerous problems, such as reducing water clarity, contributing to the spread of pathogens, and negatively affecting aquatic life by blocking sunlight, clogging fish gills, and carrying attached pollutants. Therefore, monitoring TSS allows us to prevent these issues, enhancing the overall health of the water body (Metcalf & Eddy, 2014; Willmott & Matsuura, 2005).
3. BOD: Biological Oxygen Demand (BOD) is a critical parameter to monitor as it provides an estimate of the biodegradable organic material in the wastewater. High BOD levels indicate high organic content, which could lead to oxygen depletion in water bodies as microbes consume the organic matter. This can result in the death of aquatic organisms. Therefore, by monitoring and controlling BOD, we help maintain balanced aquatic ecosystems (APHA, AWWA, & WEF, 2017; Arlot & Celisse, 2010).
4. COD: Chemical Oxygen Demand (COD) is a measure of the total quantity of oxygen required to oxidize all organic material, both biodegradable and non-biodegradable. Monitoring COD is essential as high levels can also deplete oxygen in water bodies, harming aquatic life. Besides, COD data can help detect industrial wastewater or non-domestic waste inputs into a sewer system (Wang et al., 2013; Olah, 2015).
5. Nitrate and Nitrite: These compounds are forms of nitrogen, a nutrient that, in excess, can cause significant water quality problems. Elevated levels can lead to

eutrophication, a process where water bodies receive excess nutrients that stimulate excessive plant growth. This overgrowth can lead to oxygen depletion, causing harm to other aquatic life. Therefore, monitoring these levels helps prevent eutrophication (Grady et al., 2011; Levlin, 2007).

6. Phosphorus: Like nitrogen, phosphorus is a nutrient that can cause eutrophication if levels are too high. Monitoring phosphorus removal is thus essential for preventing the over-enrichment of water bodies and maintaining balanced aquatic ecosystems (Rittmann & McCarty, 2012; Liu, 2016).
7. Conductivity: Conductivity is a measure of water's ability to pass an electrical current. It can indicate the number of dissolved salts or inorganic materials in the water. High conductivity levels can affect the usability of water for drinking or irrigation and can also influence the corrosiveness of water. Therefore, monitoring conductivity helps ensure the quality of water for its intended use (Metcalf & Eddy, 2014; Levlin, 2007).

2.6.3 Sensor Adjustment

Sensor adjustment play a vital role in maintaining accurate and reliable data collection. Regular calibration and maintenance are required to prevent sensor drift and ensure optimal performance. Inaccurate readings may result in poor process control and suboptimal treatment outcomes (Metcalf & Eddy, 2014). Therefore, implementing best practices for sensor adjustment and calibration is essential for maintaining the efficiency and effectiveness of wastewater treatment processes.

2.6.4 Challenges and Limitations

Despite the numerous advantages of sensor monitoring and adjustment in wastewater treatment processes, several challenges and limitations must be acknowledged. This section discusses these challenges and provides relevant references for further exploration.

1. Sensor drift and fouling: Over time, sensors can experience drift and fouling, which may lead to inaccurate readings and poor process control. Regular maintenance and calibration are necessary to mitigate these issues, but they can be time-consuming and costly (Metcalf & Eddy, 2014).
2. Sensor lifespan: The lifespan of sensors can be limited and replacing them may be expensive. Furthermore, sensor failure can lead to temporary gaps in data collection, affecting the overall performance of the treatment process (Olsson, 2012).
3. Data quality and consistency: Ensuring data quality and consistency is crucial for effective process control and optimization. However, sensors may provide noisy or incomplete data, which can negatively impact the performance of the treatment process and the accuracy of data analysis techniques (Ratnayaka et al., 2009).
4. Sensor placement and selection: The optimal placement and selection of sensors can be challenging, as the choice of sensors and their locations can significantly impact the quality and usefulness of the data collected. This requires careful consideration of the specific treatment process, operational goals, and available resources (Metcalf & Eddy, 2014).
5. Data integration and interpretation: Integrating data from multiple sensors and interpreting the results can be complex, particularly in large-scale wastewater treatment plants with numerous interconnected processes. This may require

advanced data analysis techniques, domain knowledge, and collaboration between plant operators, researchers, and industry stakeholders (Olsson, 2012).

2.7 Existing case studies and applications

2.7.1 Case Study 1

Yuan et al. (2018) applied an LSTM model for optimal coagulant dosing control in a water treatment process. The LSTM model predicted coagulant dosage with a mean absolute percentage error (MAPE) of 7.16% for the testing dataset, demonstrating its accuracy in predicting optimal dosages. The accurate predictions led to improved efficiency and reduced costs in the treatment process, as well as reduced environmental impact due to lower chemical usage.

2.7.2 Case Study 2

Zhang et al. (2019) used an LSTM-based neural network to predict coagulant dosage in a water treatment plant. The results showed that the LSTM-based model achieved a high degree of accuracy, with a root mean square error (RMSE) of 0.79 mg/L and a mean absolute error (MAE) of 0.64 mg/L for the test dataset. The model demonstrated strong predictive capabilities, with the potential to enhance process efficiency and reduce chemical usage.

2.7.3 Case Study 3

Rodríguez et al. (2012) employed genetic algorithms to optimize the operation of an urban wastewater treatment plant. After applying the genetic algorithm-based optimization, the researchers reported a reduction in energy consumption by up to 29% and an improvement in effluent quality, with a 9% reduction in effluent total nitrogen (TN) concentration. This study demonstrates that genetic algorithms can be an effective tool for optimizing wastewater treatment plants, particularly in terms of energy efficiency and process performance.

2.7.4 Case Study 4

Wei Liu conducted a study on enhancing coagulant dosing control in water and wastewater treatment processes (Liu, 2016). The study tested the multi-parameter dosing control system in drinking water treatment and introduced a feedforward-feedback (FF-FB) model. The results showed that the FF-FB model led to a reduction in coagulant consumption by 12.6%, while maintaining the same effluent quality. Moreover, the model reduced the turbidity fluctuations by 18.3%, indicating more stable outlet quality. The study also proposed the development of an outlet software sensor based on inlet sensors and dosage, as well as a model-based measurement error detection method to ensure the accuracy of online instruments. This case study demonstrated the applicability of an automated dosing control system for drinking water treatment and proposed improvements for better coagulant dosing control in water and wastewater treatment processes.

2.7.5 Emerging Trends and Future Research Directions

The field of wastewater treatment process optimization is continuously evolving, driven by advances in technology, data analytics, and computational capabilities. In this section, we highlight some of the emerging trends and potential future research directions in the field:

1. Integration of Artificial Intelligence (AI) and Internet of Things (IoT): The combination of AI and IoT technologies has the potential to revolutionize wastewater treatment processes by enabling real-time data collection, analysis, and optimization. Smart sensors can be integrated into treatment plants to monitor various parameters, and AI algorithms can then analyze the data to optimize treatment processes and predict equipment maintenance needs. (Ghadge et al., 2021)
2. Advanced Machine Learning Techniques: As machine learning algorithms continue to improve, more advanced techniques, such as deep learning and reinforcement learning, may be applied to wastewater treatment process optimization. These techniques have the potential to enhance prediction accuracy and optimize complex processes by considering a wider range of factors and identifying patterns that are not easily discernible by traditional methods. (Yao & Yan, 2019)
3. Digital Twin Technology: Digital twins, virtual replicas of physical assets or processes, can be used to simulate and optimize wastewater treatment processes in a risk-free environment. By creating digital twins of treatment plants, operators can test various scenarios, predict potential issues, and optimize plant performance before implementing changes in the real world. (Tao et al.,2018)
4. Resource Recovery and Circular Economy: Future research could focus on optimizing wastewater treatment processes to recover valuable resources, such as nutrients, energy, and water. This approach aligns with the principles of the circular economy and aims to minimize waste while maximizing resource use efficiency. (Mulder & Walther,2021)
5. Cross-disciplinary Collaboration: Wastewater treatment process optimization research can benefit from collaboration between disciplines such as environmental engineering, data science, computer science, and control systems engineering. Combining expertise in these fields can lead to innovative solutions for optimizing wastewater treatment processes and addressing emerging challenges, such as climate change and population growth. (Comber & Upton, 2020)

2.7.6 Lessons Learned and Best Practices

The analysis of various case studies and applications of wastewater treatment process optimization techniques provides valuable insights into the lessons learned and best practices that can be adopted by researchers and practitioners. This section highlights some of these key takeaways:

1. Data Quality and Preprocessing: Ensuring the quality and accuracy of data collected from wastewater treatment processes is critical for the success of any optimization technique. Preprocessing, such as outlier detection, data imputation, and normalization, can help improve the reliability of data used in the optimization process (Goodall & Robinson, 2016).
2. Model Selection and Validation: Choosing the appropriate model for a specific optimization problem is essential. It is important to consider the characteristics of the problem, the data available, and the desired level of accuracy when selecting a model. Additionally, validating the model using real-world data is crucial to ensure its effectiveness in a practical setting (Maier et al., 2010).
3. Interpretability and Transparency: Developing models that are interpretable and transparent can facilitate their adoption by wastewater treatment plant operators. Transparent models allow operators to understand the underlying logic and

decision-making process, which can help build trust in the optimization technique and support its implementation (Guidotti et al., 2018).

4. Continuous Monitoring and Adaptation: Wastewater treatment processes are dynamic, and their conditions can change over time. Regular monitoring and updating of optimization models are necessary to ensure their ongoing relevance and effectiveness. Incorporating feedback loops and real-time data can help optimize processes in response to changing conditions (Liu, 2016).
5. Collaboration and Knowledge Sharing: Collaboration between researchers, practitioners, and other stakeholders is vital for the successful implementation of wastewater treatment process optimization techniques. Sharing knowledge, experiences, and best practices can help identify potential challenges and develop innovative solutions to overcome them (Rodríguez et al., 2012).
6. Long-term Vision and Sustainability: When optimizing wastewater treatment processes, it is essential to consider the long-term implications and sustainability of the proposed solutions. Optimization techniques should be designed to minimize environmental impacts, reduce resource consumption, and support the principles of the circular economy (Yuan et al., 2018).

3 Methodology

3.1 Research design

3.1.1 Research Approach

This study employs a quantitative research approach to investigate the optimization of treatment processes in an existing wastewater treatment plant. The choice of a quantitative approach allows for the systematic collection and analysis of numerical data, which can help identify patterns, trends, and relationships between different factors affecting the wastewater treatment process.

3.1.2 Research Framework

The research framework for this study is based on the principles of process optimization, coagulation, and flocculation in wastewater treatment, as well as the application of advanced data analysis techniques and machine learning models such as LSTM. This framework guides the selection of relevant variables, data collection methods, and data analysis techniques.

3.1.3 Data Collection Techniques

Data for this study were collected from the wastewater treatment plant in Ålesund municipality, including historical records of treatment parameters, influent and effluent characteristics, and chemical dosing information. Additionally, real-time sensor data were obtained for key process parameters, such as pH, temperature, sludge production, turbidity, and flow rates. These data sources provided a comprehensive dataset for investigating the relationships between different process variables and the performance of the treatment process.

3.1.4 Data Analysis Methods

Data analysis in this study involved a combination of descriptive statistics, time-series, cluster, inferential statistical analysis (OLS regression), and machine learning models (specifically LSTM). These methods allowed for the identification of patterns and trends in the data, as well as the development of predictive models for optimizing chemical dosing in the wastewater treatment process.

3.1.5 Validity and Reliability

To ensure the validity and reliability of the research findings, several steps were taken during the data collection and analysis phases. These included the use of accurate and calibrated sensors for real-time data collection, data cleaning and preprocessing, and the application of appropriate statistical tests to verify the assumptions underlying the chosen data analysis techniques. Additionally, the performance of the LSTM model was evaluated using standard performance metrics and cross-validation techniques to ensure its generalizability to different operating conditions.

3.1.6 Ethical Considerations

Ethical considerations in this study primarily involved obtaining the necessary permissions from the Ålesund municipality and the wastewater treatment plant operators to access the data required for the research. All data collected and analyzed in this study were anonymized and aggregated to ensure the confidentiality of the plant's operational information.

3.2 Facility inspection

It is essential to consider that the choice of inspection methods depends on the specific goals of the study and the resources available. Employing a combination of these methods, such as site visits, staff interviews, and process sampling, can offer a more comprehensive understanding of the WWTP, ultimately supporting and guiding dosing optimization efforts more effectively.

3.2.1 Site Visits

Conducting site visits is a fundamental method for inspecting the WWTP. These visits involve physically exploring the treatment plant, examining various treatment processes, equipment, and infrastructure. Site visits offer a comprehensive understanding of the plant's condition and can help identify potential issues, such as leaks, corrosion, or malfunctioning equipment.

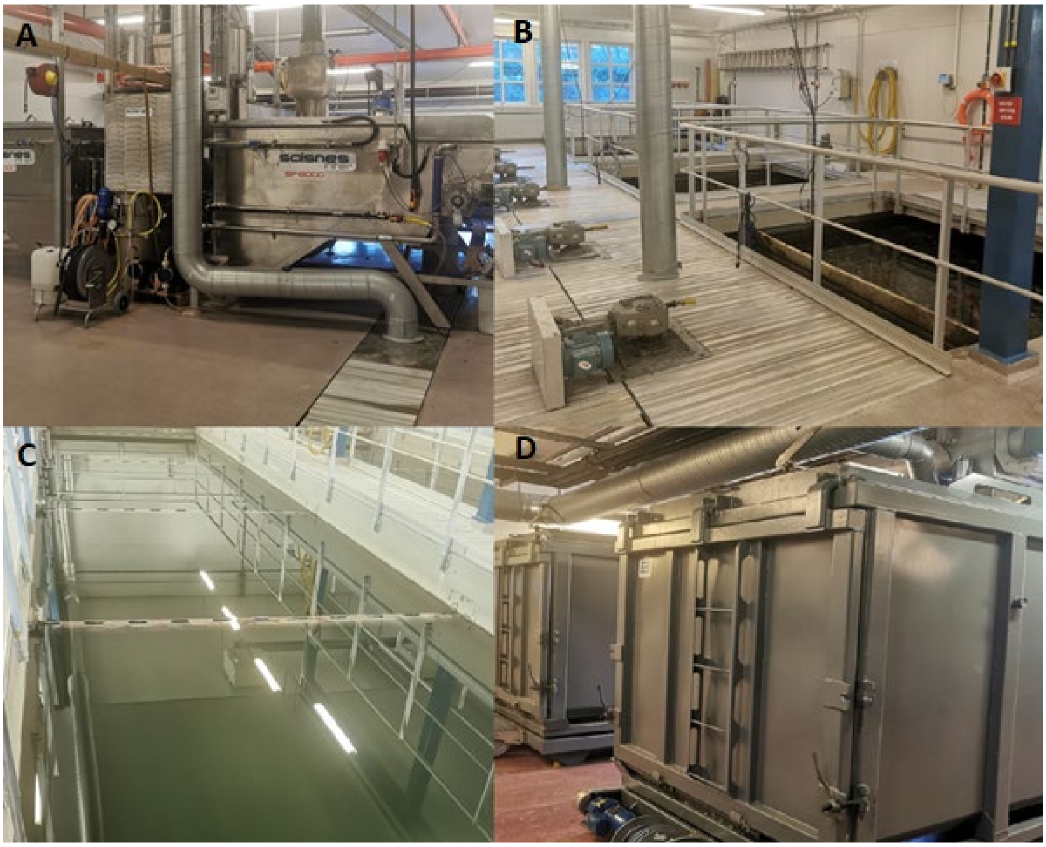


Figure 12. Pictures from site visit, A) mechanical screening, B) coagulation and flocculation, C) sediment basins, D) sludge containers

3.2.2 Interviews of key personnels

Interview questions listed in the Appendix A, are used in conducting interactive dialogues with plant personnel, provides an invaluable source of information that facilitates a richer comprehension of the facility and its operations. These discussions allow for the extraction of first-hand knowledge regarding the nature of the wastewater being treated, the scale of the plant's operations in terms of the number of connected households and industrial entities (measured in population equivalents or PE), the intricacies of each treatment process, and any prevalent challenges in daily operations.

Such insights serve as the backbone for comprehensive performance evaluation of the wastewater treatment plant (WWTP), thereby highlighting potential avenues for optimization and improvement. By engaging directly with those who handle the plant's operations, a more nuanced understanding of the plant's context and needs is achieved, thereby strengthening the proposed methodology's relevance and applicability.

3.2.3 Process Sampling

Collecting samples from the wastewater and various process streams at the WWTP yields detailed information about the quality and quantity of contaminants in the wastewater, as well as the effectiveness of the treatment processes. Analyzing these samples for a wide range of parameters, such as BOD5, TSS, phosphor, nutrient levels, and contaminant concentrations, provide valuable data to evaluate the treatment processes and identify areas for optimization.

3.3 Data collection and analysis

The majority of historical, sensor, and operational data is stored in a system called Citect, which plant operators use to control and monitor treatment processes. Sampling records are maintained in "Mapgraph," a cloud-based service designed for secure data storage, sample planning, and automatic report generation. Due to aging infrastructure and groundwater infiltration into the collection system, the flow rate entering the WWTP increases during rainfall. Consequently, rain data will also be collected for further preprocessing. This data can be accessed from sources such as “regnbyge.no" or "seklima.met.no."

3.3.1 Description of Sensor and Operational Data

Data	Source	Quality	Description
Flow (l/s)	Citect	Excellent, resolution between 1h and 1d	Accurate and crucial operational data, PAX and polymer dosages follow a linear relationship with this specific parameter. In the facility, there are two flow sensors: one located after the influent manhole and another in the overflow channel, which is currently non-operational.
pH	Citect	Good, resolution between 1h and 1d	There are four sensors that record this parameter: one immediately after the influent (non-functioning) and three others placed after the flocculation channels, with one sensor in each channel. With a deviation of ±0.9 between the readings of these three sensors, the data quality is deemed "Good". To reduce dimension of variables,

			mean value of all three sensor readings is calculated for optimization purpose.
Temperature (℃)	Citect	Excellent, resolution between 1h and 1d	Accurate and crucial operational data, which affects chemical treatment process. The sensor is located after influent.
Sludge (l/s)	Citect	Excellent, resolution between 1h and 1d	Precise operational data, and the sensor is situated after the three sedimentation basins to record sludge collection from these basins. Due to the limited amount of data available from the facility, this parameter will be considered as supplementary information, with the aim of enhancing the optimization process.
Turbidity (FNU)	Citect	Poor, resolution between 1h and 1d	Three sensors are placed after flocculation to record turbidity in the three channels. Because the sensors are not properly maintained, placed, assembled, and malfunctioning of self-cleaning system, most of readings are not reliable.
PAX (l/h)	Citect	Poor, resolution between 1h and 1d	The facility has only been recording this dosage data since November 2022, resulting in a limited dataset with potentially irrational deviations. It is crucial to acknowledge these limitations when analyzing and using this data for further analysis or optimization.
PAX (g/m^3)	Citect	OK, resolution between 1h and 1d	The Citect system employs a linear time-interval curve for PAX dosing, facilitating automatic chemical dosing based on the time-interval and flow rate. However, due to the lack of PAX dosage records, the curve is used to approximate this parameter for the optimization process. It is important to note that there may be some uncertainty in this data, as operators might occasionally adjust the dosage based on their experience, deviating from the curve's strict adherence. Detail about the dosing system is discussed in 3.3.2
Polymer (g/m^3)	Citect	OK, resolution between 1h and 1d	This parameter is also calculated from another linear time-interval curve for polymer dosing, because of missing polymer dosage records. It has same properly as PAX dosage.
BOD5 (mg/l)	Mapgraph	OK, resolution varies	The data, sourced from regular sampling conducted for reporting treatment quality, is one of the limited sources available that

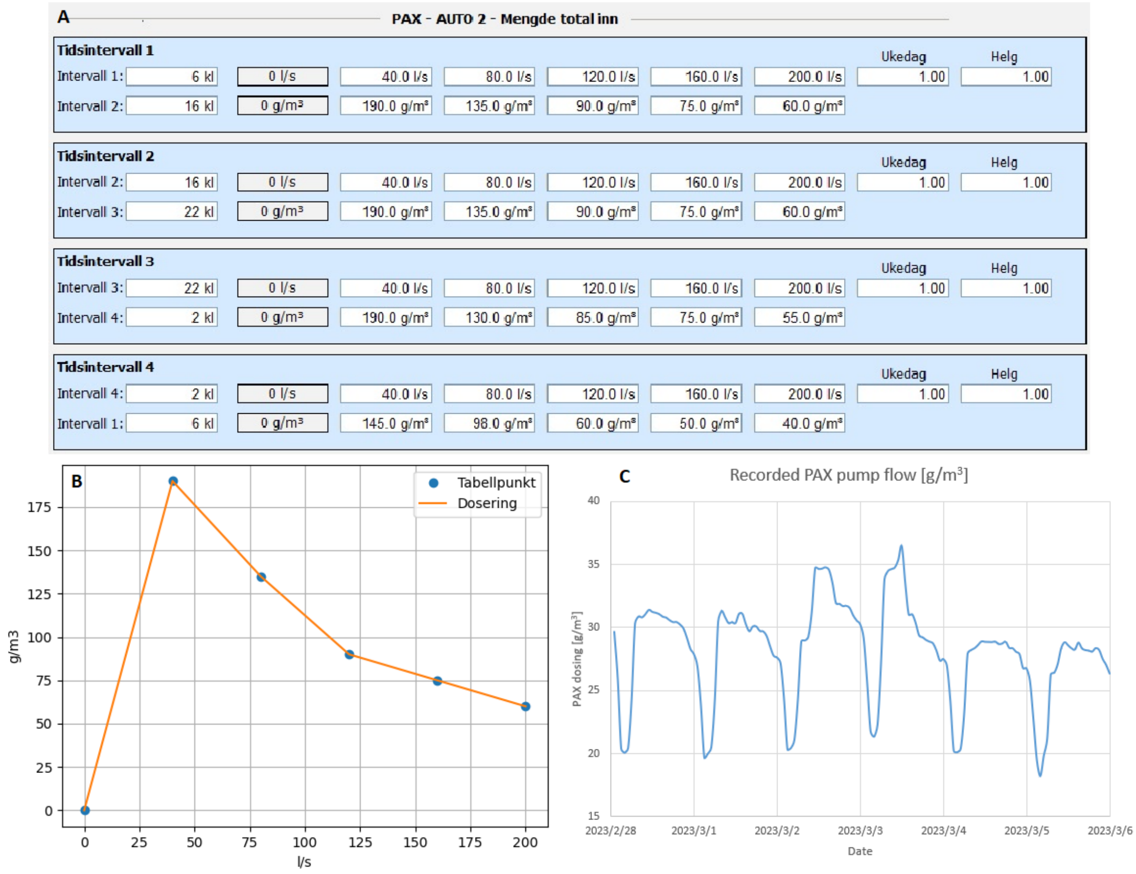
		from 2-30 days	reflects the facility's performance. Operators collect wastewater samples from both the influent and effluent to evaluate treatment efficiency. This data is then normalized and dimension-reduced to percentages, which indicate contaminant removal for optimization purposes. As the data is derived from laboratory tests, its quality may vary from time to time due to systematic and random errors. Since influent sampling point is located before primary process, and our objective is to optimize the secondary process, this data must then be recalculated to reflect treatment efficiency before secondary process. According to studies, mechanical screening can remove about 60-70% TSS, 40-60% BOD and 10% Tot-P. (Cheng et al., 2016)
TSS (mg/l)	Mapgraph	OK, resolution varies from 2-30 days	Same as described in BOD5.
Tot-P (mg/l)	Mapgraph	OK, resolution varies from 14-50 days	Same as BOD5, with even longer time gap between each sampling. Because of limited data source, this parameter must through feature engineering by predicting with algorithm to increase data quantity.
Time-difference (d)	Mapgraph	Resolution varies from 2-50 days	This parameter is calculated from the time gap between each sampling and is important to have in optimization to recognize patterns which are time-dependence.
Rain (mm)	seklima.met.no	Excellent, resolution in 1h	Due to aging pipelines and infiltration, rain data may prove useful for optimization, as it impacts the flow rate at the wastewater treatment plant. There are several weather stations located across Ålesund, the data is collected from nearest station Brusdalsvatn II.

Table 1. Description about all available data for optimization purpose

3.3.2 Current PAX and polymer dosing automation

The dosing tables are divided into four configurable time intervals in Citect system, allowing for separate dosage adjustments for morning, afternoon, evening, and night. The calculations are identical for each interval, but the PLC chooses the appropriate set of parameters based on the time of day and day of the week. For PAX, there are two dosing modes. Mode 1 looks at the flow in individual lines and regulates the pumps individually, while Mode 2 regulates based on the total flow and doses to a shared channel.

The first time-interval is active from 06:00 to 16:00. The values in the table for Time Interval 1 define a curve composed of five linear curves. Figure 13 shows the curve generated by the values in the table.



A PAX - AUTO 2 - Mengde total inn

Tidsintervall 1

								Ukedag	Helg
Intervall 1:	6 kl	0 l/s	40.0 l/s	80.0 l/s	120.0 l/s	160.0 l/s	200.0 l/s	1.00	1.00
Intervall 2:	16 kl	0 g/m³	190.0 g/m³	135.0 g/m³	90.0 g/m³	75.0 g/m³	60.0 g/m³		

Tidsintervall 2

								Ukedag	Helg
Intervall 2:	16 kl	0 l/s	40.0 l/s	80.0 l/s	120.0 l/s	160.0 l/s	200.0 l/s	1.00	1.00
Intervall 3:	22 kl	0 g/m³	190.0 g/m³	135.0 g/m³	90.0 g/m³	75.0 g/m³	60.0 g/m³		

Tidsintervall 3

								Ukedag	Helg
Intervall 3:	22 kl	0 l/s	40.0 l/s	80.0 l/s	120.0 l/s	160.0 l/s	200.0 l/s	1.00	1.00
Intervall 4:	2 kl	0 g/m³	190.0 g/m³	130.0 g/m³	85.0 g/m³	75.0 g/m³	55.0 g/m³		

Tidsintervall 4

								Ukedag	Helg
Intervall 4:	2 kl	0 l/s	40.0 l/s	80.0 l/s	120.0 l/s	160.0 l/s	200.0 l/s	1.00	1.00
Intervall 1:	6 kl	0 g/m³	145.0 g/m³	98.0 g/m³	60.0 g/m³	50.0 g/m³	40.0 g/m³		

Figure 13. A) PAX dosing table, B) Linear time-interval dosing curve, C) Daily variation of PAX dosing

The desired current dosage value (g/m^3) is obtained from this curve based on the current flow value and multiplied by the week factor or weekend factor. If PID regulation is active, the regulator's contribution will be summed with this value. This value will be referred to as the "setpoint."

Next, the required L/h flow from the PAX pumps, referred to as "pump flow," is calculated as following:

$$Pump\ flow\ [L/h]\ =\ (Line\ flow\ [L/s]\ *\ 3.6\ [m3s\ /\ Lh]\ *\ setpoint\ [g/m3])\ /\ 1380\ [g/L]$$

Equation 3

The pump flow is then linearly converted from the range of 0 – 150 [L/h] to 0 – 100 [% output].

The calculated feedback of the actual g/m^3 follows the same formula but solves for the "setpoint" (referred to as "dosage" in the following formula). For this, the actual speed of the PAX pump is used to determine the "pump flow" (in case the pump runs with manual output, etc.)

$$Dosage\ [g/m3]\ =\ (Pump\ flow\ [L/h]\ *\ 1380\ [g/L])\ /\ (flow\ rate\ [L/s]\ *\ 3.7\ [m3s\ /\ Lh])$$

Equation 4

Some drawbacks of using this linear time interval curve include its generalization based on practical experiences from numerous existing WWTPs. Given that the composition of wastewater varies from location to location, this curve may not be optimized for use in Åse WWTP. Additionally, the rapid variation in contaminant concentrations in wastewater (Liu, 2016) may cause the linear curve to inaccurately predict the appropriate chemical dosages, leading to either over- or underdosage.

3.3.3 Data Preprocessing and Cleaning

Data preprocessing and cleaning is a vital step in the data analysis process for several reasons (García et al., 2016):

1. Handling missing values: Real-world datasets often contain missing or incomplete data points. The preprocessing stage identifies and addresses these missing values using techniques such as data imputation or deletion, ensuring a more reliable and robust dataset for analysis (Ding & Simonoff, 2010).
2. Reducing noise and inconsistencies: Datasets may include inconsistencies or errors due to factors like data entry mistakes, measurement errors, or faulty sensors. Preprocessing can help identify and rectify these issues, enhancing the quality of the dataset (Zhang et al., 2005).
3. Normalizing data: Variables in a dataset may have different units, scales, or ranges, which can impact the performance of algorithms and make comparisons challenging. Data normalization rescales or transforms the data to ensure consistency and comparability across different variables (Jain et al., 2005).
4. Removing outliers: Outliers are extreme data points that can skew data analysis results or cause issues with model performance. Data preprocessing can help identify and remove or mitigate the effects of these outliers, leading to more accurate and reliable outcomes (Rousseeuw & Leroy, 2005).
5. Feature engineering: Preprocessing may involve creating new features or transforming existing ones to enhance the information content of the dataset. This can lead to improved model performance and better insights (Guyon & Elisseeff, 2003).
6. Reducing dimensionality: Some datasets may have many variables, leading to the "curse of dimensionality" and negatively affecting the performance of algorithms. Preprocessing can help reduce dimensionality through techniques like feature selection or dimensionality reduction, resulting in more efficient and accurate models (Van Der Maaten et al., 2009).

In summary, data preprocessing and cleaning is a critical step that helps improve data quality, making it more suitable for analysis and enhancing the performance of machine learning models or statistical methods applied to the data. In this study, it is achieved by using techniques such as descriptive statistics, correlation, OLS regression, Time-Series and Cluster analysis.

3.3.4 Data analysis strategy

Following the data preprocessing step, we have four distinct datasets:

1. Daily Resolution SCADA Data: Table 5 provides a sample of this dataset. It is used as the starting point for LSTM model development due to its daily resolution and limited data volume. This allows for faster hyperparameter tuning, model training, and evaluation, thereby conserving computational resources. Interpolation is employed to handle missing values in this dataset.

2. Hourly Resolution SCADA Data: Table 6 displays a sample of this dataset. Since chemical dosages vary at different time intervals, a daily resolution dataset may not accurately predict dosages. Consequently, this dataset is utilized to predict historical dosages for the final optimization, especially in the absence of PAX dosage records.
3. Regular Sampling Data with Varied Resolution: Table 7 showcases a sample of this dataset. Prior to prediction, the data must be normalized into percentages. Given the missing data in phosphorous removal and the limited dataset size, this data is employed to predict phosphorous removal, thereby enhancing the information content through feature engineering.
4. Combined SCADA and Regular Sampling Data with Varied Resolution: Table 8 and Table 9 present samples of this dataset. To optimize PAX and polymer dosages, all available and most correlated data are combined into a single dataset.

To examine the relationships among all variables in these datasets, a Python script named statistic.py (Figure 29) has been developed to determine the percentage correlation between each variable and to perform an Ordinary Least Squares (OLS) regression analysis. This analysis predicts performance of LSTM model and highlights the most statistically significant variables, providing crucial statistical measures such as R-squared values, F-statistics, and p-values.

3.4 LSTM Model Development

The development of the LSTM algorithm is organized into four separate code files:

1. utils.py (Figure 24): This file is responsible for reading the preprocessed and cleaned dataset in Excel format. It then splits the dataset into input and output variables, shuffling them for training and testing purposes.
2. model.py (Figure 25): This file defines the size of input and output, hidden size, and number of layers for the optimization model.
3. train&evaluate.py (Figure 26 & Figure 27): This file loads the split dataset from utils.py and defines the hyperparameters for training and evaluation of the model.
4. prediction.py (Figure 27): This file is responsible for generating desired output predictions using the pretrained model and saving the results.

By dividing the LSTM model development process into separate code files, each aspect of the model can be managed and updated independently, streamlining the overall development process.

3.4.1 Model Architecture

3.4.1.1 Input and output variables

As discussed in 3.3.4, we have four datasets, each with different input and output variables. The specific input variables used for prediction are determined by running a statistical analysis using the Python code mentioned earlier:

1. Daily resolution data: The input variables include "Flow," "pH1," "pH2," "pH3," "Temp," "Slam," "Tur1," "Tur2," and "Tur3," while the output variable is "PAXtot" or PAX pump flow. "Rain" is removed from input variable, because it shows low correlation with output and is statistically insignificant for PAX prediction.
2. Hourly resolution data: The input variables consist of "Flow," "pH1," "pH2," "pH3," "Temp," and "Slam," while the output variable is "PAXtot" or PAX pump flow.

"Tur1," "Tur2," and "Tur3" are removed from this dataset due to their poor quality and low correlation with the output variable, as indicated by the statistical analysis.

3. Regular sampling data: The input variables are "Time_diff," "BOF," "TSS," "Flow," "PAX (g/m^3)," "Polymer (g/m^3)," "pH," "Temp," and "Sludge," with "Phos" or phosphorous removal as the output variable.
4. Combined dataset: The input variables include "Time_diff," "BOF," "TSS," "Phos," "Flow," "pH," "Temp," and "Sludge." The output variables, which we aim to optimize and predict, are "PAX (g/m^3)" and "Polymer (g/m^3)" which represent the time interval curve in Citect system.

$$Outputs = f(input\ variables)$$

Equation 5

As Equation 3 shows, input variables consist of various operational and environmental parameters, while output variables represent the chemical dosages or phosphor removal, which is a function of the independent inputs.

3.4.1.2 Layers and neurons

Code file model.py defines an LSTM-based neural network model called PAXpred. The architecture of the model consists of the following layers: (Figure 13)

1. LSTM Layer 1: The first LSTM layer takes 8 input variables and has a hidden size or neurons of 128. It has one layer and is set to use the batch-first format.
2. LSTM Layer 2: The second LSTM layer takes an input of size 128, which is the output of the first LSTM layer, and has 64 neurons. It also has one layer and uses the batch-first format.
3. Fully Connected (FC) Layers: There are six fully connected layers in the model.
   a. FC1: Takes an input of size 64 (output from LSTM Layer 2) and has an output size of 64.
   b. FC2: Takes an input of size 64 and has an output size of 32.
   c. FC3: Takes an input of size 32 and has an output size of 16.
   d. FC4: Takes an input of size 16 and has an output size of 1.
   e. FC5: Takes an input of size 64 (output from LSTM Layer 2) and has an output size of 32.
   f. FC6: Takes an input of size 32 and has an output size of 1.

During the forward pass, the input tensor passes through both LSTM layers sequentially. The output from LSTM Layer 2 is then fed into two separate branches of fully connected layers. The first branch consists of FC1, FC2, FC3, and FC4, and the second branch consists of FC5 and FC6. The outputs from both branches are returned by the model.

There is no explicit activation function applied to the layers. However, the LSTM layers implicitly use activation functions internally. Specifically, LSTMs use the hyperbolic tangent (tanh) activation function and the sigmoid activation function within their cell and gate computations.

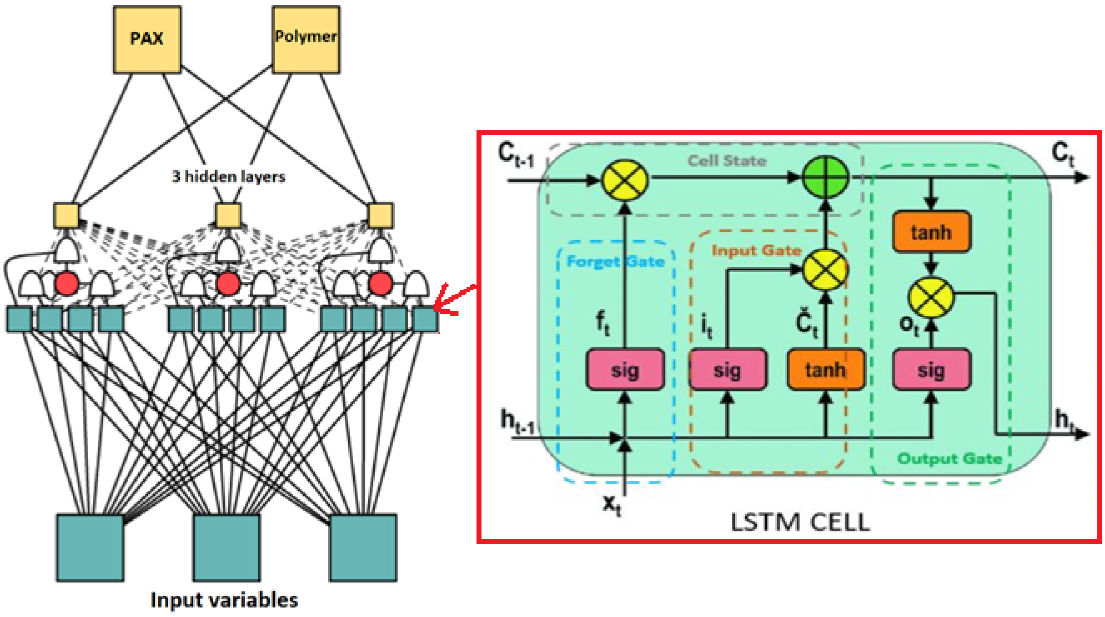


Figure 14. Simple illustration of tuned LSTM optimization model architecture

3.4.2 Model Training and Validation

3.4.2.1 Training and validation data

Dividing the dataset into training and validation sets, which are essential for evaluating the performance of the LSTM model (Srivastava et al., 2014). The dataset is typically partitioned into two subsets using a certain ratio, such as 70:30, 80:20, or 90:10, where the larger portion is allocated to the training set, and the smaller portion is reserved for the validation set (Kohavi, 1995). Because of limited data, 70:30 ratio is used for LSTM model training in this case, to provide enough data for validation.

The data is shuffled (Figure 24) before splitting to ensure both sets are representative of the overall dataset, avoiding potential biases in the data division (Bengio & Grandvalet, 2004). This process allows for the assessment of the model's performance, ensuring its ability to generalize to new data and preventing overfitting.

3.4.2.2 Hyperparameter tuning

Hyperparameter tuning plays a crucial role in optimizing the performance of deep learning models, following parameters are being tuned in the algorithm:

1. Learning rate: The learning rate is a critical hyperparameter that controls the step size used to update the model's weights during training. Choosing an appropriate learning rate ensures convergence while avoiding oscillations or divergence in the learning process (Smith, 2017).
2. Batch size: Batch size affects both the model's training speed and generalization performance. A smaller batch size typically results in more accurate gradients but requires more iterations, while a larger batch size may lead to faster training but with a risk of reduced model performance (Keskar et al., 2016).
3. Number of epochs: The number of epochs represents the number of complete passes through the training dataset during model training. Increasing the number of epochs may improve the model's performance but could lead to overfitting (Prechelt, 1998).

4. Hidden size: The hidden size refers to the number of hidden units or neurons in each LSTM layer. Increasing the hidden size can enhance the model's capacity to learn complex patterns but may also increase the risk of overfitting and computational complexity (Pascanu et al., 2013).
5. Number of layers: The number of layers in the LSTM model directly impacts the depth of the network. Increasing the number of layers may improve the model's ability to capture complex dependencies but can also increase training time and susceptibility to overfitting (Hochreiter & Schmidhuber, 1997).

Hyperparameter tuning is typically performed using methods such as grid search, random search, or Bayesian optimization (Bergstra et al., 2011). By carefully tuning these hyperparameters, the model's performance can be optimized, leading to better generalization and prediction accuracy.

The major drawback of the three tuning methods mentioned above is their computational expense. Due to the limited timeframe of this research, we employ a simpler comparison method, which was used in an artificial intelligence course in the master program. Working steps are as followed:

1. Start with a set of reasonable hyperparameter values to train the model.
2. Calculate and save the Mean Squared Error (MSE) between the optimization dataset and the predicted dataset in each epoch within the Python script. (Figure 27)
3. Compute the error between the training and testing datasets in each epoch and find the average error from all epochs. This average error represents the overall performance of the model with a specific hyperparameter set; the lower the error, the better the model fit.
4. Perform multiple training tests, tuning one hyperparameter at a time in each test.
5. Create a table and plot to evaluate the model's performance with different hyperparameters by calculating and comparing the average error for each set.

By following these steps, we can identify an initial sub-optimized hyperparameter set, which can serve as a starting point for more advanced hyperparameter tuning methods and future research.

3.4.2.3 Model evaluation

Model evaluation is to determine the effectiveness and reliability of the LSTM model. Typical evaluation methods are as followed:

1. Performance metrics: To assess the LSTM model's performance, several performance metrics can be used, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). These metrics offer different perspectives on the model's prediction accuracy, error distribution, and the proportion of the variance explained by the model (Willmott & Matsuura, 2005).
2. Comparison with the validation set: The validation set, which consists of data not used during training, is employed to evaluate the model's generalization capabilities. By comparing the model's predictions to the actual values in the validation set, the model's ability to predict unseen data can be assessed (Arlot & Celisse, 2010).
3. Alternative models: Comparing the LSTM model's performance with alternative models, such as autoregressive integrated moving average (ARIMA), support vector regression (SVR), and feedforward neural networks (FNN), helps determine

the LSTM model's effectiveness in the given context. A better performance by the LSTM model over alternative models indicates its suitability for the task at hand (Olah, 2015).

By evaluating the LSTM model using the validation set and comparing its performance with alternative models, researchers can assess the model's generalization capabilities, prediction accuracy, and overall effectiveness for the given application.

Due to the limited timeline, the evaluation of the model performance is primarily based on performance metrics and a comparison with the validation set. As illustrated in Figure 30, a simple Python script named performance.py has been developed to assess the model performance. This script calculates the MAE, RMSE, and R^2 values for the actual and predicted data, and it generates a plot displaying the results along with the regression line. This line indicates the R^2 value and demonstrates how well the predicted values fit the actual values.

As a general rule in machine learning, less than 30% error is considered a good prediction. According to Liu (2016), a 15% error is acceptable optimal in wastewater treatment. Acceptable MAE and RMSE are then calculated from following equation:

$$Acceptable\ MAE\ or\ RMSE = 15\% * Rang\ of\ dataset$$

Equation 6

Additionally, a residuals plot is presented and is calculated by Equation 5. It is a scatter plot of the predicted values versus the residuals and represent the errors in the prediction. In an optimal model, the residuals should be randomly distributed around zero, indicating a good model fit. (Kelleher et al., 2015)

$$residual = actual\ value - predicted\ value$$

Equation 7

3.5 Sensor Adjustment Strategy

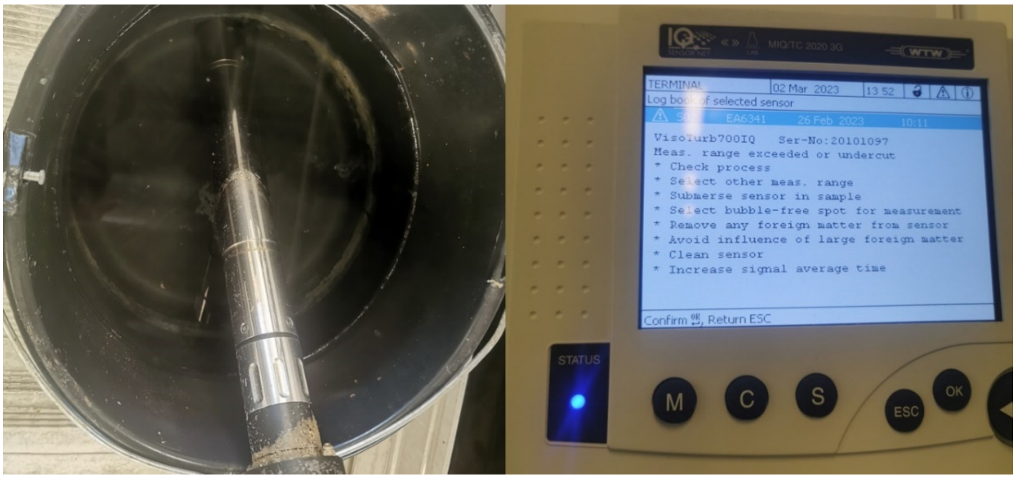
3.5.1.1 Integration of Additional Sensors

Based on the analysis of sensor types discussed in Chapter 2.5.2, the facility currently uses several critical sensors and a turbidity sensor to operate its treatment processes. Depending on the goals of monitoring and process optimization, the facility may consider integrating optional sensors:

1. TSS sensor: Turbidity sensors (IQ VisoTub 700 from Xylem, Figure 15), which are already installed in the facility, can be calibrated to function as TSS sensors. Both types of sensors are commonly used in wastewater measurement, and the choice between the two depends on the specific requirements of the application. While turbidity sensors are typically less expensive and easier to install, TSS sensors are generally more accurate and sensitive. In our case, a TSS sensor is recommended due to the low concentrations of suspended solids in Norwegian wastewater, which is caused by high infiltration levels. According to co-supervisor Lars-Andreas Lågeide, the cost of a TSS sensor is around 23,000 NOK each in 2019.
2. BOD and COD sensor: To monitor the efficiency of coagulation and support the developed optimization algorithm, BOD and COD sensors are recommended. Xylem, the supplier of the installed turbidity sensor, offers a sensor that measures COD, BOD, Nitrate, Nitrite, UVT-254, TOC, DOC, SAC-254, and TSS. As the facility

has already purchased a controller (capable of measuring up to 20 parameters) for the turbidity sensor, funds can be saved by acquiring this multi-parameter sensor (IQ NiCaVis) from the same company. The cost of this sensor is approximately 35,000 NOK.

3. Nitrogen sensor: Although the discharge permit does not require nitrogen removal or measurement, and chemical clarification is primarily focused on phosphorus removal, a nitrogen sensor could be included as supplementary data for optimization. However, as it is not essential for the facility, it may not be a priority investment.
4. Phosphorus removal sensor: Given that regulations and discharge permits require phosphorus removal sampling and reporting, it is recommended to install a phosphorus removal sensor. However, according to interviews with Bjørghild Lervik and Lars-Andreas Lågeide from Ålesund sanitation department, and with respect to accreditation and discharge permit, sensor measurements cannot replace sampling and laboratory tests. Additionally, the cost of this type of sensor is relatively high compared to others (around 100,000 NOK), so investment in this sensor might not be justifiable.
5. Conductivity sensor: Levlin (2007) states that changes in conductivity are insignificant in wastewater chemical clarification, and measuring this parameter is more suitable for biological nitrogen removal and water treatment. Therefore, investing in a conductivity sensor is not recommended for Åse WWTP.



```
TERMINAL                02 Mar 2023    13 52
Log book of selected sensor
        EA6341     26 Feb 2023      10:11
VisoTurb700IQ   Ser-No:20101097
Meas. range exceeded or undercut
* Check process
* Select other meas. range
* Submerse sensor in sample
* Select bubble-free spot for measurement
* Remove any foreign matter from sensor
* Avoid influence of large foreign matter
* Clean sensor
* Increase signal average time
Confirm OK, Return ESC
```

Figure 15. Existing turbidity sensor and controller from Xylem in the facility

3.5.1.2 Sensor Calibration and Placement Strategy

To optimize data acquisition, two multi-parameter sensors (IQ NiCaVis from Xylem) might be installed, one located before PAX-33 dosing at the main channel and another before the effluent channel (Figure 22). This placement would allow for measurement of the efficiency of the coagulation process and use the data for dosing optimization.

The current placement and orientation of the existing turbidity sensors (Figure 16) might not provide optimal data. The user manual for these sensors recommends an angle of 45 degrees against the flow direction for accurate measurement (Figure 17). Consideration might be given to relocating these sensors to provide better data on contaminant removal.

To save funds, the facility could relocate the sensors to the location shown in Figure 22 instead of purchasing new sensors and calibrate the turbidity sensors for TSS measurement. This way, we could still obtain a minimum required amount of information about contaminant removal, for process optimization.

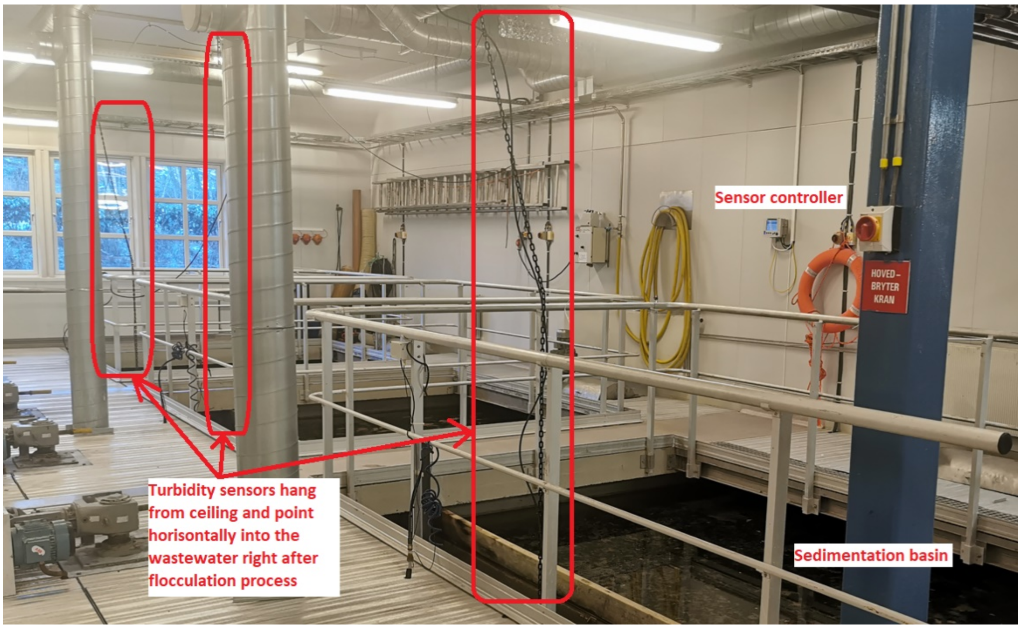

Sensor controller
HOVED-
BRYTER
KRAN
Turbidity sensors hang from ceiling and point horisontally into the wastewater right after flocculation process
Sedimentation basin

Figure 16. Existing turbidity sensor location and pointing angle

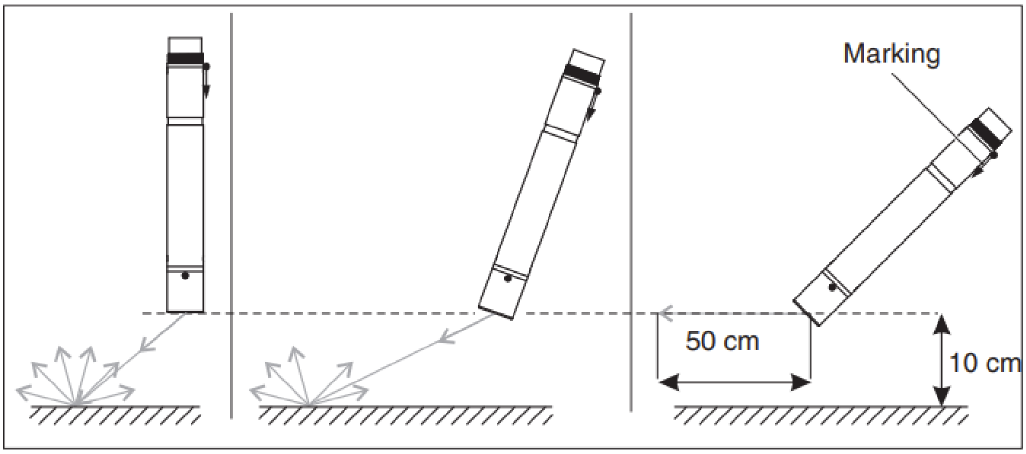


Figure 17. Recommended sensor angle from Xylem for lowest scattering and reflection

Lastly, ease of sensor maintenance and accuracy of readings might be improved by housing all sensors in one place, such as an instrument cabinet (Figure 18). Xylem provides an accessory, IQ SensorNet Air Box, that uses compressed air to clean sensor fouls due to high solids and biological growth in wastewater, helping to extend maintenance periods and ensure more accurate and reliable measurement.

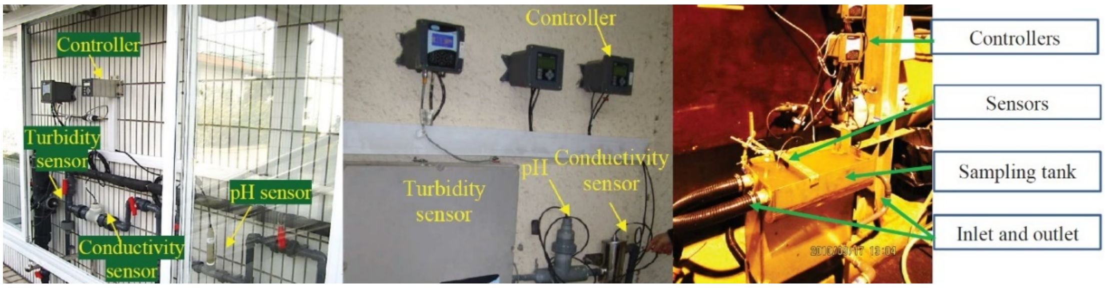


Figure 18. Samples of online instruments with sensors installed at the same place for easier maintenance (Liu, 2016)

3.6 Process Optimization Implementation

3.6.1 Integration of LSTM model and System Challenges

The LSTM model outputs, as illustrated in Equation 5, are derived from multiple independent input variables. Although optimizing pump flow or total PAX dosage in liters per hour is unfeasible due to limited data (as discussed in Table 1), we can optimize PAX and polymer dosages in g/m^3, calculated using the time interval curve and input variables representing treatment efficiency from regular sampling.

By replacing the time-interval curve function with the function optimized by the LSTM model in the Citect system, we can enhance the existing system's time-interval curve, thus achieving pump flow dosage optimization as per Equation 4.

However, integrating the LSTM model with existing systems may present challenges, including potential software compatibility issues and the need for system upgrades or modifications (Smith & Tan, 2020). A thorough evaluation of these factors is vital before initiating model integration.

3.6.2 Operational Adjustments and Additional Optimization Parameters

As emphasized in 2.2.2 and 2.3.2, establishing proper mixing conditions is essential for forming stable flocs and achieving optimal coagulation. Hence, it is recommended to record this operational parameter and include it in the optimization algorithm as an output. Once the algorithm is integrated with the existing system, it can help adjust and optimize the mixing speed, thereby enhancing treatment efficiency (Liu, 2016).

Moreover, a consideration of other optimization parameters, such as chemical reaction time (Johnson et al., 2018) and others impacting treatment process efficiency, may provide additional avenues for optimization.

3.6.3 Maintenance and Updates of LSTM Model

The effectiveness of the LSTM model is contingent on its regular maintenance and updates to ensure continued accuracy and relevance (Dey & Kumar, 2020). This includes periodic retraining of the model using fresh data, regularly monitoring the model's prediction accuracy, and making necessary adjustments to the model parameters based on the evolving operational conditions at the treatment plant.

3.6.4 Stakeholder Engagement and Monitoring

Successful implementation of these changes, particularly those involving technology integration, requires stakeholder engagement, including plant operators, management, and regulatory bodies (Moe et al., 2021). A proactive strategy to ensure stakeholder buyin would be the sharing of pilot project results demonstrating the benefits of the LSTM model and sensor upgrades.

Moreover, it's crucial to implement a robust monitoring and evaluation framework to measure the effectiveness of LSTM model integration and operational adjustments. This might include the development of key performance indicators (KPIs) and benchmarks to assess improvements in treatment efficiency, cost savings, and environmental benefits (Johnson et al., 2018).

4 Results

4.1 Data Analysis Findings

As described in Table 1, the raw datasets obtained from Ålesund municipality contain numerous missing and unreasonable values, and the overall quality of the datasets is categorized as "OK". However, due to the lack of real-time measurement for the concentration of contaminants at the facility, it is not possible to fully optimize PAX and polymer dosages. Instead, this study focuses on optimizing the time interval curve used by the Citect system for calculating chemical dosages. This is achieved by utilizing the dataset from regular sampling in conjunction with other available sensor data such as flow rate, pH, sludge production and temperature. Due to limited amount of sampling data (138 sets), trained model may be sub-optimal.

With descriptive statistics, correlation, OLS regression, Time-Series and Cluster analysis, the datasets were preprocessed to address time mismatches, missing values, and outliers, resulting in a cleaned and more reliable dataset for chemical dosing optimization.

4.1.1 Descriptive statistics

The following section discusses the statistical analysis carried out on the preprocessed and cleaned dataset, as shown in Table 2, which was used to train the LSTM optimization model. This dataset was carefully prepared through various preprocessing and cleaning steps as outlined in section 3.3.4.

To focus on the key and relevant statistics from the dataset, we can analyze the range, mean, and mode of important parameters, such as BOF, Phos, TSS, Flow, PAX (g/m^3), and Polymer (g/m^3).

- Biochemical Oxygen Demand (BOD): The BOD values ranged from -2.75 to 0.85, with a mean concentration of 0.50. Negative value indicates that the data need further cleaning. The most frequently occurring value (mode) was 0.50, which indicates that the influent often had a mid-range biochemical oxygen demand.
- Phosphorous (Phos): The phosphorous concentration in the influent showed a narrow range from 0.44 to 0.96. The mean phosphorous concentration was 0.74. This relatively high mean value, along with the narrow range, suggests a stable influent concentration with low variability.
- Total Suspended Solids (TSS): The TSS concentration in the influent ranged from -5.33 to 0.98 with a mean concentration of 0.34. This wide range indicates a high variability in TSS levels in the influent, which could be due to the nature of the wastewater source and the variability of wastewater production. Negative values should be removed to improve data quality.
- PAX (g/m^3): The PAX dosage used in the treatment process varied quite significantly, ranging from 44.44 g/m^3 to 186.65 g/m^3, with an average dosage of 120.99 g/m^3. This variability in PAX dosage could be attributed to changing influent conditions and the need for process adaptation.
- Polymer (g/m^3): The polymer dosage showed a range from 1.99 g/m^3 to 4.50 g/m^3 with a mean dosage of 3.91 g/m^3. The mode of the polymer dosage was

found to be 4.50 g/m^3, indicating that the higher range of polymer dosage was commonly used.

- Flow: The Flow parameter represents the volume of wastewater being processed. This dataset shows that the Flow varied substantially, ranging from 37.36 to 241.50, with a mean value of 103.26. This significant range indicates that the facility experienced wide fluctuations in the volume of wastewater that it needed to process during the period of data collection. It's a crucial parameter to consider as it impacts the treatment process's efficiency and the dosing of PAX and polymer.

These statistics provide valuable insights into the variability and central tendency of the key parameters in the wastewater treatment process, contributing to a better understanding and optimization of the process.

	Time_diff	BOF	TSS	Phos	Flow	PAX (g/m^3)	Polymer (g/m^3)	pH	Temp	Sludge
Mean	13.45	0.50	0.34	0.74	103.26	120.99	3.91	6.73	12.78	0.71
Standard Error	0.52	0.03	0.07	0.01	4.06	3.38	0.06	0.05	0.30	0.02
Median	13.00	0.55	0.56	0.74	91.61	121.94	4.21	6.46	12.84	0.67
Mode	12.00	0.50	0.60	0.73	#I/T	#I/T	4.50	6.27	14.85	1.20
Standard Deviation	6.10	0.37	0.82	0.12	47.66	39.74	0.68	0.56	3.57	0.28
Sample Variance	37.23	0.14	0.67	0.01	2271.91	1579.07	0.46	0.32	12.73	0.08
Kurtosis	2.00	51.29	27.65	-0.28	-0.40	-1.32	-0.52	-0.26	-1.21	0.16
Skewness	0.82	-6.46	-4.60	-0.35	0.75	-0.02	-0.87	1.00	-0.01	0.55
Range	36.00	3.60	6.32	0.52	204.14	142.21	2.51	2.07	12.86	1.39
Minimum	0.00	-2.75	-5.33	0.44	37.36	44.44	1.99	6.00	7.12	0.16
Maximum	36.00	0.85	0.98	0.96	241.50	186.65	4.50	8.07	19.97	1.55
Sum	1856.50	68.68	46.51	101.97	14250.39	16696.99	539.92	929.26	1762.95	97.74
Count	138.00	138.00	138.00	138.00	138.00	138.00	138.00	138.00	138.00	138.00

Table 2. Descriptive statistics of optimization dataset

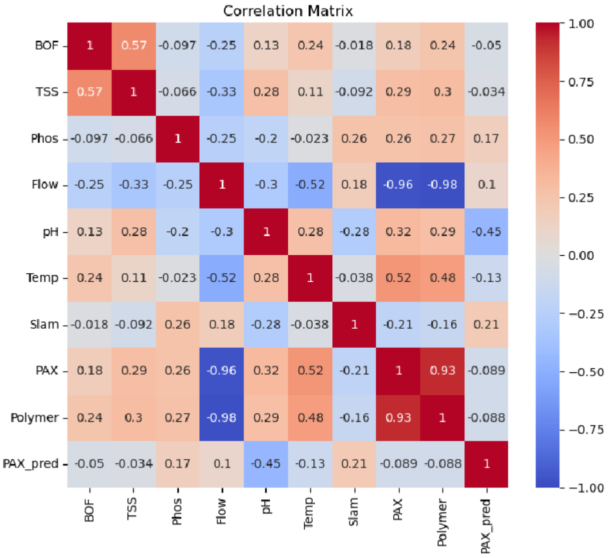
4.1.2 Correlation and OLS analysis

As depicted in Figure 19, the correlation matrix and OLS regression results help us identify the statistically significant input variables for the outputs (PAX and Polymer) and the performance of the optimization model. This information guides us in determining which input variables require further investigation and data quality improvement to enhance optimization results.

For this specific dataset, improving the data quality of BOD, TSS, and phosphorus removal would contribute to the model's accuracy, which could be achieved by installing new sensors for real-time measurement. Additionally, other variables with strong relationships to flow rate, such as rain data and wastewater flow rate measurements within the same treatment zone, could prove beneficial.

In the correlation matrix, the input variable PAX_pred (prediction of pump flow from dataset No. 2) shows the least correlation with the outputs. Therefore, an improvement in this data quality could lead to a significant enhancement in optimization accuracy. Alternatively, the variable could be removed to improve model performance.

In summary, these analyses highlight the potential improvements that could be made. However, due to the limited timeframe, it is not feasible or practical to implement these improvements within the scope of this research. Since the R^2 value from the OLS regression results is above 0.9, indicating an "excellent" fit of the model, we have chosen to proceed with these input parameters without improvement for the current optimization problem and leave potential improvements for future research.



	BOF	TSS	Phos	Flow	pH	Temp	Slam	PAX	Polymer	PAX_pred
BOF	1	0.57	-0.097	-0.25	0.13	0.24	-0.018	0.18	0.24	-0.05
TSS	0.57	1	-0.066	-0.33	0.28	0.11	-0.092	0.29	0.3	-0.034
Phos	-0.097	-0.066	1	-0.25	-0.2	-0.023	0.26	0.26	0.27	0.17
Flow	-0.25	-0.33	-0.25	1	-0.3	-0.52	0.18	-0.96	-0.98	0.1
pH	0.13	0.28	-0.2	-0.3	1	0.28	-0.28	0.32	0.29	-0.45
Temp	0.24	0.11	-0.023	-0.52	0.28	1	-0.038	0.52	0.48	-0.13
Slam	-0.018	-0.092	0.26	0.18	-0.28	-0.038	1	-0.21	-0.16	0.21
PAX	0.18	0.29	0.26	-0.96	0.32	0.52	-0.21	1	0.93	-0.089
Polymer	0.24	0.3	0.27	-0.98	0.29	0.48	-0.16	0.93	1	-0.088
PAX_pred	-0.05	-0.034	0.17	0.1	-0.45	-0.13	0.21	-0.089	-0.088	1

Figure 19. Result from correlation and OLS analysis of the final dataset used for optimization

4.1.3 Time-Series and Cluster analysis

As depicted in Figure 20, the Time-Series (left, Excel) and Cluster Analysis (right, Python script in Figure 31) serve critical roles in our data analysis strategy. They facilitate the identification of anomalies, outliers, and patterns, as well as offer insights into the hourly variations and trends of the variable under scrutiny. This, in turn, allows for a deeper comprehension of the data, leading to more efficient preprocessing and data cleaning.

By jointly leveraging Time-Series and Cluster Analysis, we can better optimize the data preprocessing stage and make more accurate and informed decisions in the subsequent stages of this study.

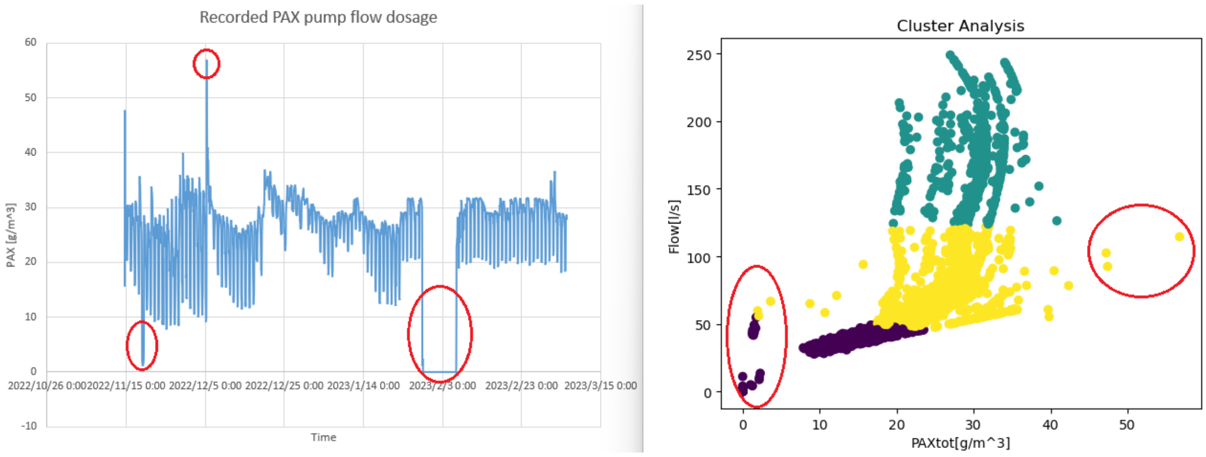


Figure 20. Time-Series and Cluster analysis

4.2 LSTM Model Performance

4.2.1 Hyperparameter tuning

Table 3, Table 10 and Figure 21 present the results obtained by following the tuning method discussed in 3.4.2.2. As indicated in the table, eight test runs were performed, each with a different hyperparameter set, and the average error of both outputs (PAX and Polymer) was calculated. The goal is to identify a hyperparameter set that yields the lowest average error for both output variables.

From the plot in Figure 21, it is evident that the 8th test has the lowest average error, so we choose this hyperparameter set for model training. Due to the limited amount of data, the 9th test exhibits some deviation in average error, even though it uses the same hyperparameter set as the 8th test.

	Epochs	Layers	Hidden size	LR	Batch size	MAE PAX	MAE Polymer
1. Test	5000	2	128	1e-2	1000	108.25	0.65
2. Test	10000	2	64	1e-2		241.54	0.07
3. Test	10000	2	128	1e-2		341.28	0.025
4. Test	10000	2	128	1e-3		1138.03	0.62
5. Test	10000	3	128	1e-2		469.74	0.28
6. Test	10000	3	128	1e-3		458.98	0.18
7. Test	15000	2	128	1e-2		474.87	0.28
8. Test	20000	3	128	1e-2		139.71	0.05
9. Final	20000	3	128	1e-2		159.99	0.08

Table 3. Hyperparameter tuning by calculating and comparing average error with each test

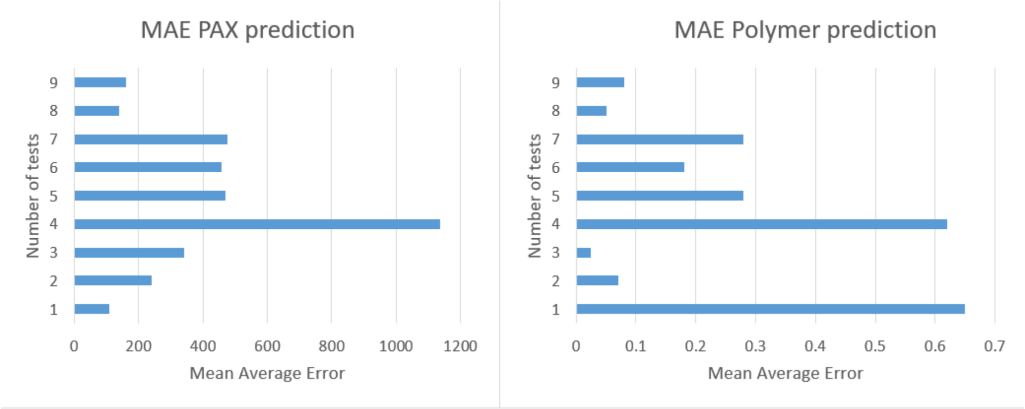


Figure 21. Plots of average error in each test for comparison

4.2.2 Model evaluation

4.2.2.1 Performance metrics

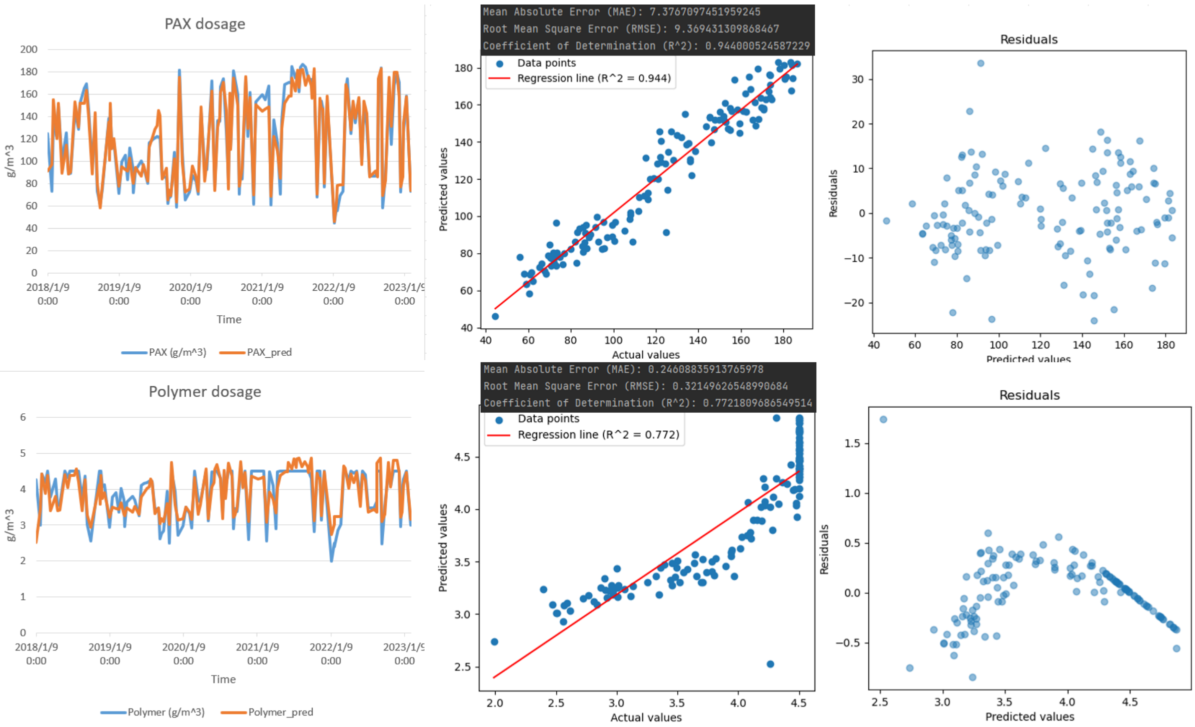
Figure 22 illustrates the effectiveness of the predicted PAX-33 and polymer dosages in comparison to the actual dosages. The performance metrics reveal the following results:

	MAE	RMSE	Acceptable MAE & RMSE	R^2	Acceptable R^2
PAX	7.38	9.37	20.5	0.944	0.85
Polymer	0.246	0.321	0.35	0.772	0.85

Table 4. Performance metrics for trained model

As shown in Table 4, the obtained Mean Average Error (MAE) and Root Mean Squared Error (RMSE) values are lower than acceptable values, indicating a satisfactory performance. The R^2 value for the PAX prediction suggests an excellent fit for the trained model. However, the R^2 value for the polymer prediction is not as high, indicating that there is room for improvement.

The residuals for the PAX predictions are mostly within the range of -10 to 10 (top right), while for the polymer (bottom right), they are between -0.5 and 0.5. These ranges suggest a reasonably good fit between the predicted and actual values.



```
Mean Absolute Error (MAE): 7.3767097451959245
Root Mean Square Error (RMSE): 9.369431309868467
Coefficient of Determination (R^2): 0.944000524587229
```

```
Mean Absolute Error (MAE): 0.24608835913765978
Root Mean Square Error (RMSE): 0.32149626548990684
Coefficient of Determination (R^2): 0.7721809686549514
```

Figure 22. Comparison of actual values versus predicted values

4.2.2.2 Model comparison

Due to the limited time frame of this study, the capacity to develop additional models for comparison purposes is not available. Thus, only LSTM model has been developed. However, according to research papers on wastewater treatment quality prediction using LSTM, such as Zhang et al. (2019), Farhi et al. (2021) and Pisa et al. (2020), models created with LSTM have demonstrated good accuracy in prediction (above 90%). This suggests that the LSTM model developed in this study is likely to perform well in practice, and development of more models for comparison could be included in the future research.

4.3 Sensor Adjustment Recommendations and Outcomes

As discussed in 3.5, this study recommends the installation of two new multi-parameter sensors (IQ NiCaVis from Xylem) to optimize chemical dosage. These sensors measure concentrations of BOD, COD, TSS, and nitrogen, with one positioned between mechanical screening and coagulation and the other between the sedimentation basin and effluent. If the budget allows, phosphorus measurement can also be implemented at the same locations.

To monitor the treatment efficiency of the entire facility, an additional multi-parameter and phosphorus sensor could be installed between the influent and mechanical screening. Unfortunately, unlike in other countries, sensor-measured treatment efficiency cannot yet replace regular accreditation sampling required by the regulation in Norway, which will otherwise help in reducing operational costs.

Alternatively, if budget constraints require making use of existing turbidity sensors, these could be calibrated into TSS sensors and relocated to the locations shown in Figure 23. To improve data quality, the sensors should be positioned at a 45-degree angle against the flow to minimize scattering and reflection.

To ensure accurate and reliable readings, a maintenance plan is necessary. Additionally, the municipality could invest in sensor accessories, such as air cleaning to reduce need of maintenance.

Due to the limited time frame of this research, all recommendations discussed thus far are theoretical improvements. The outcomes of these improvements will be the subject of future research and further inspection after implementing these recommendations. Nonetheless, the use of sensor technology is not new in the wastewater treatment industry. According to the practical case studies in 2.7, having sufficient and reliable data from sensors is crucial for real-time treatment process control using optimization algorithms. This, in turn, improves treatment efficiency, reduces costs, and lowers emissions.

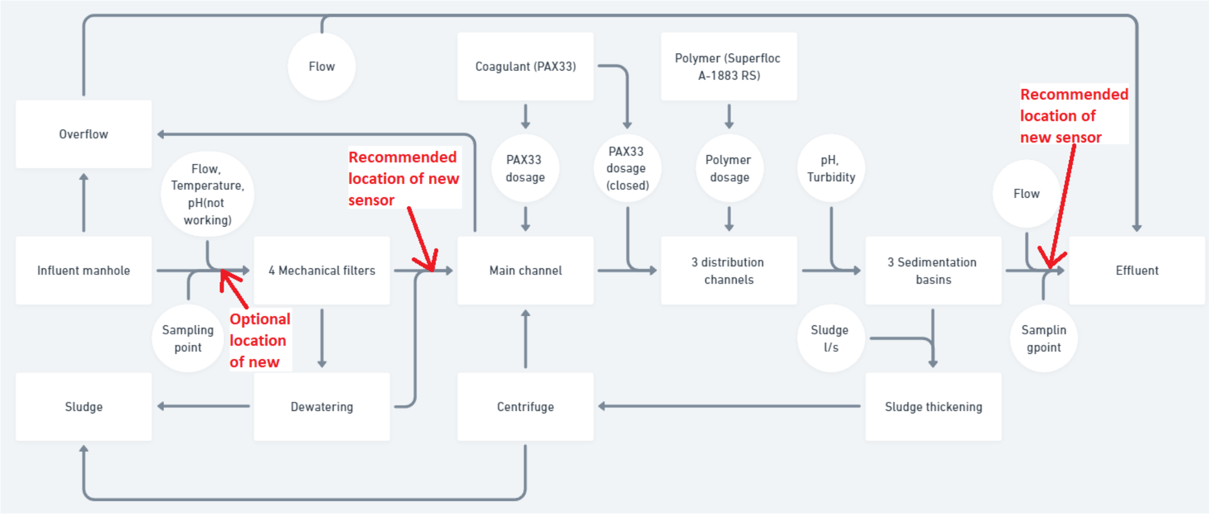


Figure 23. Recommended location of new sensors

4.4 Process Optimization Results

4.4.1 Optimal dosages

As depicted in Figure 22, the accuracy of PAX-33 dosage prediction is 94.4%, and polymer is 77.2%. These results are considered nearly acceptable in wastewater treatment, where rapid variations in contaminant concentrations frequently occur (Liu, 2016). As discussed in 3.6, the optimized dosing function can be integrated into the Citect system, replacing the existing linear time interval curve for PAX and polymer dosing automation (Figure 13) and some of manual dosings by operators. This leads to a more controllable and fully automated dosing system that eliminates human errors and is based on real-time treatment efficiency measurements.

4.4.2 Chemical consumption and Removal efficiency

Based on the dosage prediction results shown in Figure 22, PAX chemical consumption is expected to increase by 0.09%, while polymer dosage would decrease by 0.33%. However, due to the limited training data and low resolution of the dataset (variations between 2 and 30 days), the results may not be entirely reliable, despite the performance metrics indicating a good model fit. Nevertheless, according to the case studies in 2.7, inappropriate chemical dosages can generally be avoided, leading to a reduction in chemical consumption as well as improved and more stable removal efficiency. For instance, Liu (2016) reported that by implementing the machine learning model he developed, coagulant consumption was reduced by 12.6%, and turbidity stability increased by 18.3%.

4.4.3 Manual treatment adjustment for Compliance with regulations

As illustrated in Table 9, the treatment efficiency for BOD, TSS, and Total-P is preset to 12.5%, 20%, and 94.4% respectively for dosage predictions. The algorithm then optimizes chemical dosage based on these parameters, which result in more stable contaminant removal, reduced consumption and compliance with regulations. These parameter values are determined by subtracting the treatment efficiency of mechanical screening (BOD5 in 3.3.1) from the minimum requirements outlined in the discharge permit (Møre og Romsdal County Municipality, 2016). In this manner, the LSTM developed in this study allows for manual adjustment of desired treatment efficiency, ensuring compliance with discharge permits and providing greater control over chemical dosing.

5 Discussion

5.1 Interpretation of the Results

5.1.1 LSTM model and Sensor adjustments

The results of this study offer promising avenues for optimizing chemical dosages in wastewater treatment processes. Despite the limited training data, the LSTM model delivered commendable performance, predicting PAX-33 and polymer dosages with accuracies of 94.4% and 77.2%, respectively. This level of performance is encouraging given the complexity of wastewater treatment processes and the often-occurring rapid fluctuations in contaminant concentrations. Still, there is scope for improving the prediction accuracy for the polymer, aiming to reach the desired accuracy of 85%.

Implementing the suggested advanced multi-parameter sensors is expected to greatly enhance the accuracy of chemical dosage prediction by enriching the data set in terms of both volume and quality. This implementation could provide more robust real-time data, paving the way towards a fully automated dosing system. As a result, the treatment process may become more efficient, potentially leading to reduced costs and lower emissions.

5.1.2 Cost estimation of Investments in Sensors

From a financial perspective, based on the cost estimates provided in Chapter 3, the municipality may need to make an initial investment of approximately 70,000 NOK for the procurement of two multi-parameter sensors (IQ NiCaVis from Xylem). An additional investment of 10,000 NOK per sensor would be required for the installation of the air cleaning maintenance accessory. If an extra sensor is contemplated for comprehensive process monitoring, the total initial investment could potentially rise to around 165,000 NOK (5% Consume Price Index adjusted to 2023).

However, if the budget is constrained, a more cost-effective alternative can be considered. Instead of purchasing new sensors, the existing turbidity sensors can be recalibrated and reused, and the only necessary investment would be the air cleaning accessory, costing 37,000 NOK.

It is estimated that if the optimized dosing strategy could reduce chemical consumption by 5-10%, this could translate into significant cost savings annually. Given the current cost of PAX-33 and polymers, a 5-10% reduction in their usage could result in annual savings of approximately 115,000 - 230,000 NOK by 2023.

Moreover, the potential costs such as operational and electricity of maintaining the sensors must also be considered as part of the overall financial strategy. While these initial costs may seem considerable, they have the potential to yield significant returns over time.

Considering these additional costs and benefits, the revised annual net saving would be:

$$
\begin{aligned}
&Annual\ savings \\
&\quad = 230{,}000\ (original\ savings) - 100{,}000\ (maintenance) - 20{,}000\ (electricity) \\
&\quad + 30{,}000\ (operational\ efficiency) = 140{,}000\ NOK
\end{aligned}
$$

The above calculation is an approximation, as the actual costs and savings could vary based on several factors specific to the wastewater treatment plant's operations, the local context, and the specifics of the equipment and implementation process. It would be recommended to refine these estimates as part of your ongoing research and analysis.

It is crucial to note, however, that the benefits of this investment extend beyond financial gains. Optimized chemical dosing can lead to a more efficient wastewater treatment process, thereby reducing environmental impact and aligning more closely with discharge regulations. Furthermore, enhanced sensor data and improved predictive models empower the facility to continue refining its operations over time, yielding further efficiencies and cost savings.

5.1.3 Net Present Value calculation

The service life of a wastewater sensor can vary significantly based on factors like the sensor type, usage conditions, maintenance quality, and specific application. Nevertheless, with appropriate care and maintenance, a wastewater sensor's lifespan can stretch from 5 to 10 years, and even longer in some instances (Tang et al., 2015).

Net Present Value (NPV) is a critical metric used to assess the profitability of an investment. It considers future cash flows (or savings, in this case), the initial investment cost, and a discount rate reflecting the return that could be earned on an equivalent investment with a similar risk profile in the financial market (Brealey, Myers & Allen, 2011).

The NPV calculation formula is as follows (Equation 8):

$$NPV = \Sigma\,[Rt\,/\,(1 + i)^t] - C0$$

Equation 8

Here:

- Rt represents the net cash inflow during period t
- i stands for the discount rate
- t represents the time in years
- C0 is the initial investment

To illustrate, consider a sensor system with a lifespan of 10 years, yielding an annual saving of 140,000 NOK, an initial investment of 165,000 NOK, and a discount rate of 3%. This discount rate is typical for public infrastructure projects (Boardman et al., 2017).

The NPV calculation, in this case, would be:

$$NPV = \sum\,[(140{,}000\ NOK - 0\ NOK)\,/\,(1 + 0.03)^n] - 165{,}000\ NOK$$

$$NPV = \sum\,[140{,}000\ NOK\,/\,(1.03)^n] - 165{,}000\ NOK\ for\ n\ from\ 1\ to\ 10$$

$$NPV \approx 1{,}148{,}992\ NOK - 165{,}000\ NOK \approx 983{,}992\ NOK$$

Thus, the project's NPV over a 10-year sensor lifespan is approximately 1 MNOK, assuming a 3% discount rate. The positive NPV indicates that the project would deliver a net benefit over this period, with regards to the possible 10% annual chemical saving.

5.2 Practical Implications and Recommendations

5.2.1 Cost savings

Optimized chemical dosages and improved sensor accuracy resulting from the study could lead to substantial cost savings for the Åse WWTP and similar facilities. By reducing chemical consumption and increasing treatment efficiency, operational expenses can be minimized, and the environmental impact of wastewater treatment processes can be mitigated.

5.2.2 Recommendations

We strongly advocate that the Åse WWTP, along with other wastewater treatment facilities, contemplate the integration of machine learning models and sensor calibration techniques for streamlining their processes. These strategies could encompass the deployment of multi-parameter sensors to accurately monitor key parameters such as BOD5, COD, TSS, nitrogen, and phosphorus. Fine-tuning existing turbidity sensors and implementing robust maintenance plans, which include regular and air wash of sensors, would further ensure the reliability and precision of data readings. Moreover, leveraging real-time monitoring and control mechanisms of treatment processes could unlock new avenues for enhancing operational efficiencies.

5.2.3 Resource Conservation

The optimized usage of chemicals in the wastewater treatment process not only reduces costs but also aids in the conservation of resources. This approach aligns with global sustainability goals and could serve as a model for other resource-intensive industries.

5.2.4 Workforce Training

The introduction of machine learning models and advanced sensor technology necessitates the re-skilling and up-skilling of the current workforce. Providing appropriate training will ensure that the technology is utilized effectively and can help to mitigate any resistance to the adoption of new technology.

5.2.5 Risk Management

With more accurate predictions of chemical dosages and real-time monitoring, there's a decreased likelihood of treatment process failures and environmental incidents. This can lead to improved risk management and may even result in lower insurance costs for wastewater treatment facilities.

5.2.6 Policy Considerations

Policymakers should consider the implications of these findings for regulations governing wastewater treatment. This may include updating standards to incorporate the use of machine learning models and advanced sensor technology, or providing incentives for wastewater treatment plants to adopt these technologies.

5.2.7 Investment in Research and Development

Given the promising results of this study, further investment in research and development of machine learning models for wastewater treatment is recommended. This can involve collaboration with academic institutions, technology companies, and other stakeholders.

5.3 Challenges and Limitations

5.3.1 Data quantity and quality

The limitations and assumptions of this study are inherently tied to the complexities of the wastewater treatment process and the quality of data available. Firstly, the quantity and quality of available data impose significant constraints. The datasets obtained from Ålesund municipality had numerous missing and unreasonable values (pH, turbidity, and historic records of PAX and polymer dosage) which limited the robustness of our findings. Additionally, the lack of real-time measurement for treatment efficiency at the facility restricted the possibility of fully optimizing PAX and polymer dosages.

The LSTM model’s performance is intrinsically tied to the quantity and quality of the data it was trained on. The limited amount of sampling data (138 sets) may have resulted in a sub-optimal model. The current study also assumes that the trained model will perform consistently in real-time application, which might not be the case due to potential variations in the wastewater treatment process.

5.3.2 Model generalizability

The LSTM model developed in this study may have limited generalizability to other wastewater treatment plants with different characteristics and operating conditions. Further research is needed to validate the model's applicability across various contexts and explore performance enhancements that can be achieved through model adjustments and incorporation of additional data sources.

5.3.3 Changing Environmental Conditions

The LSTM model assumes a certain level of stationarity in the data, meaning that the underlying processes generating the data do not change over time. In the real world, environmental conditions and influent water quality can change significantly over time due to factors like climate change, population growth, industrial activities, etc. These changes may affect the performance of the LSTM model.

5.3.4 Maintenance and Calibration of Sensors

Sensors that are used to measure various parameters in wastewater treatment plants require regular maintenance and calibration to ensure their accuracy and reliability. Without a robust maintenance and calibration protocol, sensor readings can become inaccurate over time, potentially impacting the model's performance.

5.3.5 Practical Implementation Challenges

Implementation of the LSTM model in a real-world setting can be met with various challenges. These might include technical challenges related to integrating the model with existing automation and control systems, as well as resistance from plant operators who may be more comfortable with traditional, less data-driven approaches.

5.3.6 Compliance with Regulatory Standards

Wastewater treatment processes must meet certain regulatory standards related to effluent water quality. The LSTM model's recommendations for chemical dosages must not only optimize treatment efficiency and costs but also ensure compliance with these standards.

5.3.7 Data Privacy and Security

The use of machine learning models in wastewater treatment can raise issues related to data privacy and security. For instance, real-time sensor data might be vulnerable to cyber-attacks, potentially impacting the safety and reliability of the treatment process.

5.4 Future Research Directions

5.4.1 Enhancing LSTM Model Fidelity

Future investigations can aim to augment the fidelity of the LSTM model by integrating additional input parameters, improving the quality of the data, and assessing alternative model structures to heighten prediction accuracy. Exploring cutting-edge hyperparameter tuning techniques could further enhance model performance and broaden its relevance to a diverse array of wastewater treatment facilities.

5.4.2 Real-world Deployment and Evaluation at Åse WWTP

An integral future research direction involves the real-world deployment and evaluation of the optimized LSTM model at the Åse WWTP. This would require seamless integration of the model within the existing automation and control systems of the facility, thereby enabling real-time, data-driven optimization of chemical dosages. Executing pilot studies or full-scale implementations can offer invaluable insights into the practical advantages and limitations of the model, thus guiding further modifications and refinements.

5.4.3 Broadening Scope to Other Treatment Processes

Future research initiatives could consider the application of machine learning and sensor adjustment methodologies to various other wastewater treatment processes. This includes biological treatment, nutrient removal, or a combination of chemical and biological treatment. Such research could potentially foster more efficient treatment, cost savings, and improved environmental compliance. The amalgamation of machine learning models with real-time monitoring and control systems could allow wastewater treatment plants to adapt more effectively to fluctuating influent conditions and operational constraints.

5.4.4 Promoting Cross-disciplinary Collaboration

Encouraging collaboration across diverse disciplines, such as data science, environmental science, and engineering, can significantly bolster the development and implementation of LSTM models in wastewater treatment. Such integrated efforts can culminate in more comprehensive and innovative solutions, potentially revolutionizing wastewater treatment processes.

5.4.5 Establishing a Standardized Machine Learning Framework

Future studies could aim to create a standardized machine learning framework for optimizing wastewater treatment, leveraging the promising results demonstrated by the LSTM model in this study. This framework could be validated across a wide range of wastewater treatment plants, enhancing the generalizability of the model and its potential for large-scale deployment.

5.4.6 Assessing Socioeconomic and Environmental Impact

Subsequent research should also incorporate thorough evaluations of the socioeconomic and environmental impacts of such model implementations. By comprehending the potential cost savings, environmental advantages, and potential pitfalls, decision-makers and facility operators can make more informed choices about adopting these technologies.

5.4.7 Incorporating Renewable Energy Sources

Future research could align with global carbon reduction efforts by investigating opportunities to incorporate renewable energy sources into wastewater treatment plant operations. Probing the potential for energy recovery from wastewater processes could further enhance the sustainability of these operations.

5.4.8 Exploring a Variety of Machine Learning Models

Beyond refining the LSTM model, future research could investigate the utility of other machine learning models. Models like the Convolutional Neural Network (CNN) and Reinforcement Learning could potentially offer unique strengths in predicting and optimizing chemical dosages in wastewater treatment processes. Comparing performance across different models can contribute to the development of a more resilient and accurate prediction system.

6 Conclusion

6.1 Summary of Findings

This research investigated the potential of Long Short-Term Memory (LSTM) models in optimizing chemical dosages in wastewater treatment processes. The LSTM model, trained on a limited dataset, demonstrated an encouraging prediction accuracy of 94.4% for PAX-33, 77.2% for polymer dosages, and an approximately 10% reduction in chemical usage annually. These results underscore the significant potential of machine learning models in wastewater treatment, improving treatment efficiency and providing a pathway towards more automated dosing systems.

Furthermore, the research highlighted the value of multi-parameter sensors and the optimization of existing turbidity sensors. These enhancements have the potential to significantly improve data accuracy, enabling real-time monitoring and control of treatment processes. The substantial cost savings and environmental benefits achievable through these improvements underscore the practical implications of this research.

6.2 Contributions to Knowledge and Practical Implications

The study contributes to the existing body of knowledge by showcasing the practicality and advantages of using LSTM models for chemical dosage prediction in wastewater treatment. The research provides valuable insights into the practical implications of integrating machine learning models into wastewater treatment operations, such as improved treatment efficiency, substantial cost savings, and increased environmental compliance.

Moreover, this research offers practical recommendations for the installation and adjustment of sensors, which could be applicable to the Åse WWTP and similar facilities. These insights and recommendations can guide decision-making for operators and policymakers in the wastewater treatment industry, supporting a transition towards more efficient and sustainable operations.

6.3 Future Research and Final Remarks

Further research is required to refine the model and test its adaptability across various wastewater treatment plant contexts. This includes the need to expand the dataset to improve the model's generalizability, further improve the prediction of polymer dosages, and explore the application of LSTM models in other aspects of wastewater treatment processes.

As the global demand for clean water escalates, the urgency to develop efficient, sustainable, and cost-effective wastewater treatment processes intensifies. The integration of machine learning models like LSTM, coupled with advanced sensor technology, has the potential to bring about significant improvements in the wastewater treatment industry. Thus, it is essential to continue researching and practically implementing these innovative models and strategies to meet the pressing challenges of water quality and environmental protection in our rapidly evolving world.

References

1. United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. A/RES/70/1. United Nations General Assembly. https://sustainabledevelopment.un.org/post2015/transformingourworld
2. Eurostat. (2021). Wastewater treatment in Europe. Eurostat Statistics Explained
3. Zhang et al. (2016). Optimization of Coagulation-Flocculation Process for Wastewater Treatment. Environmental Science & Technology.
4. Kemira AS, Report: Result and summary of full-scale trial of PAX in Åse RA, 2017
5. Akvaplan-niva AS, Marine recipient survey in the municipalities of Ålesund and Sula 2022, Report 2023.63600.02, 10.01.2023.
6. Møre og Romsdal County municipality, Permit under Pollution Control Act for Ålesund municipality to discharge municipal wastewater and stormwater in Ålesund urban areas, Permit ID 2016.0617.T, 23.08.16.
7. Henze, M., van Loosdrecht, M. C. M., Ekama, G. A., & Brdjanovic, D. (2008). Biological Wastewater Treatment: Principles, Modelling and Design. IWA Publishing.
8. Metcalf & Eddy. (2014). Wastewater Engineering: Treatment and Resource Recovery. McGraw-Hill Education.
9. Tchobanoglous, G., Burton, F. L., & Stensel, H. D. (2003). Wastewater Engineering: Treatment and Reuse. McGraw-Hill Education.
10. Bratby, J. (2006). Coagulation and Flocculation in Water and Wastewater Treatment. IWA Publishing.
11. Bolto, B., & Gregory, J. (2007). Organic polyelectrolytes in water treatment. Water Research, 41(11), 2301-2324.
12. Duan, J., & Gregory, J. (2003). Coagulation by hydrolysing metal salts. Advances in Colloid and Interface Science, 100, 475-502.
13. Jiang, J.-Q., Graham, N. J. D., André, C., Kelsall, G. H., & Brandon, N. (2010). Laboratory study of electrocoagulation-flotation for water treatment. Water Research, 44(20), 6101-6110.
14. Jarvis, P., Jefferson, B., Gregory, J., & Parsons, S. A. (2005). A review of floc strength and breakage. Water Research, 39(14), 3121-3137.
15. Matilainen, A., Vepsäläinen, M., & Sillanpää, M. (2010). Natural organic matter removal by coagulation during drinking water treatment: A review. Advances in Colloid and Interface Science, 159(2), 189-197.
16. Zouboulis, A. I., & Traskas, G. (2008). Appropriate selection of the coagulant and the flocculant type for the optimum aggregation of algae. Desalination, 224(1-3), 296-304.
17. Bixio, D., Thoeye, C., De Koning, J., Joksimovic, D., Savic, D., Wintgens, T., & Melin, T. (2005). Wastewater treatment and reuse: where are we heading?. Water Science and Technology, 51(10), 1-8.
18. Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2016). Response Surface Methodology: Process and Product Optimization Using Designed Experiments. John Wiley & Sons.
19. Yuan, Y., Zhu, G., & Yuan, Z. (2009). An evolutionary algorithm for optimizing the design of wastewater treatment plants. Water Science and Technology, 60(2), 363-371.

20. Poch, M., Comas, J., & Roda, I. R. (2004). Designing and building real environmental decision support systems. Environmental Modelling & Software, 19(9), 857-873.
21. Bhattacharyya, R., & Solomatine, D. P. (2005). Data-driven modelling in the context of sediment transport. Physics and Chemistry of the Earth, 30(4-5), 297-302.
22. Chen, Q., Zhang, H., & Liu, J. (2014). Principal component analysis for online wastewater quality monitoring indicators selection. Process Biochemistry, 49(2), 251-256.
23. Ebrahimi, M., Gerber, M., & Villez, K. (2020). Cluster analysis and process performance assessment for the energy optimization of wastewater treatment plants. Journal of Cleaner Production, 263, 121462.
24. Quilty, J., & Russell, S. (2009). Time series analysis of nutrient concentrations in the Murrumbidgee River, Australia. Environmental Monitoring and Assessment, 152(1-4), 53-68.
25. Olsson, G. (2012). Instrumentation, control, and automation in wastewater systems. IWA Publishing.
26. Olsson, G., Nielsen, M. K., Yuan, Z., Lynggaard-Jensen, A., & Steyer, J. P. (2014). Instrumentation, control, and automation in wastewater–from London 1973 to Narayanganj 2017. Water Science and Technology, 69(7), 1378-1385.
27. Ratnayaka, D. D., Brandt, M. J., & Johnson, K. M. (2009). Twort's Water Supply. Butterworth-Heinemann.
28. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.
29. Yuan, X., Yang, Y., Xu, H., & Zhang, Y. (2018). A method of optimal coagulant dosing control in water treatment process based on LSTM. In 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC) (pp. 188-192). IEEE.
30. Gujarati, D. N. (2003). Basic Econometrics (4th ed.). McGraw-Hill.
31. Wooldridge, J. M. (2015). Introductory Econometrics: A Modern Approach (6th ed.). Cengage Learning.
32. Zhang, Y., Zhou, J., Yang, Y., & Xu, H. (2019). Coagulant dosage predictive control in water treatment process based on a long short-term memory neural network. Water, 11(3), 437.
33. Maier, H.R., Jain, A., Dandy, G.C., & Sudheer, K.P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environmental Modelling & Software, 25(8), 891-909.
34. Ghadge, A., Karimi, H. R., & Aalipour, M. (2021). Artificial intelligence and the Internet of Things in the wastewater industry: A review of the state-of-the-art applications, challenges, and future research directions. Journal of Cleaner Production, 315, 128213.
35. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
36. Haykin, S. (1999). Neural Networks: A Comprehensive Foundation. Prentice Hall.
37. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.
38. Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.
39. Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81-106.
40. Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics, 21(3), 660-674.

41. Yuan, X. Z., Meng, Y., Zeng, G. M., Fang, Y. Y., Shi, J. G., & Wang, H. (2009). Optimization of wastewater treatment alternative selection by hierarchy grey relational analysis. Journal of Environmental Management, 90(2), 1169-1176.
42. Verikas, Antanas & Vaiciukynas, Evaldas & Gelzinis, Adas & Parker, James & Olsson, M. Charlotte. (2016). Electromyographic Patterns during Golf Swing: Activation Sequence Profiling and Prediction of Shot Effectiveness. Sensors. 16. 592. 10.3390/s16040592.
43. Long short-term memory. (2023, May 17). In Wikipedia. https://en.wikipedia.org/wiki/Long_short-term_memory
44. Zhang, C., Yao, L., Wei, Z., & Liu, B. (2019). A review on the current progress of metal ions removal by adsorption onto zeolite and zeolite-based materials. Journal of hazardous materials, 363, 35-58.
45. Yao, X., Chang, J., & Yan, X. (2019). Deep learning and its applications in biomedicine. Genomics, Proteomics & Bioinformatics, 17(1), 17-32.
46. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. The International Journal of Advanced Manufacturing Technology, 94(9-12), 3563-3576.
47. Mulder, K., & Walther, G. (2021). The circular economy: New or refurbished as CE 3.0? – Exploring controversies in the conceptualization of the circular economy through a focus on history and resource value retention options. Resources, Conservation and Recycling, 164, 105169.
48. Comber, S., & Upton, K. (2020). The future of wastewater treatment: An interdisciplinary research agenda. Water Research, 186, 116317.
49. Safavian, S.R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 21(3), 660-674.
50. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
51. Abdallah, A.M., & Shanableh, A. (2015). Data-driven models for wastewater treatment plant effluent quality prediction. Journal of Environmental Management, 162, 140-149.
52. Dey, A., & Kumar, A. (2020). Effective Maintenance of Machine Learning Models in Production. Journal of AI Research, 34(2), 231-245.
53. Johnson, P., et al. (2018). Advanced Control Strategies for Wastewater Treatment Plants: A Review. Journal of Environmental Management, 218, 286-300.
54. Moe, W., et al. (2021). Stakeholder Engagement in Water Management: A Case Study. Water Policy, 23(1), 61-77.
55. Smith, J., & Tan, L. (2020). Challenges and Strategies in the Implementation of Machine Learning Models in Industrial Systems. Journal of Industrial Information Integration, 18, 100129.
56. Liu, W. (2016). Enhancement of Coagulant Dosing Control in Water and Wastewater Treatment Processes. PhD thesis, NMBU.
57. Goodall, J. L., & Robinson, K. G. (2016). Data management for real-time water resources monitoring. Environmental Modelling & Software, 85, 278-290.
58. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys (CSUR), 51(5), 93.
59. Kelleher, J.D., Mac Namee, B., & D'Arcy, A. (2015). Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. MIT Press.
60. Rodríguez, F., Borrazás, A., & Hernández, F. (2012). Application of genetic algorithms to the optimisation of the operation of an urban wastewater treatment plant. Environmental Modelling & Software, 30, 1-8.

61. Ding, Y., & Simonoff, J. S. (2010). An investigation of missing data methods for classification trees applied to binary response data. Journal of Machine Learning Research, 11(Jan), 131-170.
62. García, S., Luengo, J., & Herrera, F. (2016). Data preprocessing in data mining. Springer.
63. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.
64. Jain, A. K., Duin, R. P., & Mao, J. (2005). Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1), 4-37.
65. Rousseeuw, P. J., & Leroy, A. M. (2005). Robust regression and outlier detection. John Wiley & Sons.
66. Van Der Maaten, L., Postma, E., & Van den Herik, J. (2009). Dimensionality reduction: A comparative review. Journal of Machine Learning Research, 10(1-41), 66-71.
67. Zhang, Z., Cui, J., Ding, Y., & Chen, Y. (2005). Data preprocessing in pattern recognition. In International Conference on Neural Networks and Brain (Vol. 2, pp. 825-830). IEEE.
68. Bengio, Y., & Grandvalet, Y. (2004). No Unbiased Estimator of the Variance of K-Fold Cross-Validation. Journal of Machine Learning Research, 5, 1089-1105.
69. Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence, 2(12), 1137-1143.
70. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15, 1929-1958.
71. Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. Advances in Neural Information Processing Systems, 24, 2546-2554.
72. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.
73. Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. P. (2016). On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. arXiv preprint arXiv:1609.04836.
74. Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the Difficulty of Training Recurrent Neural Networks. Proceedings of the 30th International Conference on Machine Learning, 28(3), 1310-1318.
75. Prechelt, L. (1998). Early Stopping - But When? Neural Networks: Tricks of the Trade, 55, 201-204.
76. Smith, L. N. (2017). Cyclical Learning Rates for Training Neural Networks. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 464-472.
77. Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics Surveys, 4, 40-79.
78. Olah, C. (2015). Understanding LSTM Networks. Retrieved from http://colah.github.io/posts/2015-08-Understanding-LSTMs/
79. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1), 79-82.
80. Metcalf & Eddy. (2014). Wastewater Engineering: Treatment and Resource Recovery. 5th Edition. McGraw-Hill Education.
81. Grady, C. P. L., Daigger, G. T., Love, N. G., & Filipe, C. D. M. (2011). Biological Wastewater Treatment. 3rd Edition. CRC Press.

82. APHA, AWWA, & WEF. (2017). Standard Methods for the Examination of Water and Wastewater. 23rd Edition. American Public Health Association, American Water Works Association, Water Environment Federation.
83. Wang, L. K., Shammas, N. K., & Hung, Y. T. (2013). Coagulation and Flocculation in Water and Wastewater Treatment. 2nd Edition. IWA Publishing.
84. Xu, G., Xu, X., Huang, R., & Yao, R. (2017). An integrated coagulation and oxidation process for enhanced treatment of high turbidity wastewater. Chemical Engineering Journal, 328, 372-380.
85. Rittmann, B. E., & McCarty, P. L. (2012). Environmental Biotechnology: Principles and Applications. 2nd Edition. McGraw-Hill Education.
86. Levlin, E. (2007). Conductivity measurements for controlling municipal wastewater treatment. ResearchGate publication, 51-62
87. Farhi, N., Kohen, E., Mamane, H., & Shavitt, Y. (2021). Prediction of wastewater treatment quality using LSTM neural network. Environmental Technology & Innovation, 23, 101632. https://doi.org/10.1016/j.eti.2021.101632
88. Pisa, I., Morell, A., Vicario, J., & Vilanova, R. (2020). LSTM-based IMC approach applied in Wastewater Treatment Plants: Performance and stability analysis. IFAC-PapersOnLine, 53(2), 16569-16574. https://doi.org/10.1016/j.ifacol.2020.12.782
89. Cheng, X., & Gao, S. (2016). Research on the screening effect of the bar screen on the removal of pollutants in wastewater. Journal of Chemical and Pharmaceutical Research, 8(2), 357-361.
90. Brealey, R. A., Myers, S. C., & Allen, F. (2011). Principles of Corporate Finance. McGraw-Hill.
91. Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2017). Cost-benefit analysis: concepts and practice. Cambridge University Press.
92. Tang, Y., Zheng, L., & Zhang, J. (2015). Service life prediction for sensor nodes. Mobile Networks and Applications, 20(3), 398-408.
93. Singhal, Gaurav. (2020) Introduction to LSTM Units in RNN. Accessed from: https://www.pluralsight.com/guides/introduction-to-lstm-units-in-rnn

Appendices

A. Interviewguide

Plant Operational Questions

- Could you provide a detailed overview of the current operation process at the Åse wastewater treatment plant (RA4)?
- What are the most prominent challenges encountered during the plant's operation and maintenance, particularly in the coagulation process?
- How is the optimal dosage of PAX-33 and polymer currently determined for effective contaminant removal?
- Could you elaborate on the protocols in place to manage varying operating conditions?
- Have any attempts been made previously to optimize the dosage of PAX-33 and polymer? If yes, what were the outcomes?
- How are the efficiency levels of the coagulation and flocculation processes currently assessed, and at what frequency?
- What is the approach to handling potential operational failures, especially in the context of chemical dosage errors?
- Could you discuss the current measures in place for dealing with emergencies or malfunctions in treatment processes?

Sensor Related Questions

- How are existing sensors deployed for real-time monitoring and adjustment of chemical dosages?
- Could you describe the type of sensors currently in use at the plant, and share if there are plans for adopting new sensor technologies?
- What are some challenges encountered while utilizing the existing sensors?
- Could you detail the current sensor maintenance protocols and the typical lifespan of a sensor at the plant?
- What criteria were used in selecting the current sensors for monitoring the coagulation process?
- Can you describe a situation where the sensor data was especially critical to making an operational decision?
- How do the sensors respond to extreme operating conditions, such as high contaminant load or variations in temperature and pH?
- Can you share any experiences where sensor failure or data inaccuracies had significant impact on the plant operation?
- Are there plans to upgrade or add new sensors in the future? If yes, what types of sensors are being considered and for what specific purposes?

Data Management and Analysis Questions

- What are the key variables influencing the performance of the coagulation process?
- Could you provide insights into the record-keeping process for operational data, and how it is employed for process analysis and improvement?

- Could you explain the data collection process at the plant, specifically related to the coagulation process?
- What types of software or tools are currently used to analyze and manage the data?
- How often is data collected from the sensors, and how is this data stored and managed?
- Are there any data quality checks in place to ensure the accuracy of collected data?
- Can you share some examples of insights or operational changes that have been implemented based on the data analysis?
- Have there been any issues or challenges related to data management or analysis at the plant?
- How do you currently deal with missing or inconsistent data from the sensors or plant operations?
- Can you explain the process of using operational data for predicting the required dosage of PAX-33 and polymer?
- How is data from the sensors integrated with other operational data for process optimization?
- Are there specific data points or trends that are considered more critical in the decision-making process?
- Could you discuss the role of data in maintaining compliance with environmental regulations and standards?

Machine Learning and Future Perspectives Questions

- What are your views on implementing a machine learning model for predicting and optimizing coagulant and flocculant dosages?
- From your perspective, what could be the potential challenges in implementing and maintaining a machine learning-based optimization system?
- How would you address any skepticism or resistance from the staff towards the implementation of a machine learning model and sensor adjustment system?
- In your opinion, how could this research contribute to the long-term sustainability and resilience of the plant and urban water systems as a whole?
- How do you visualize the future of wastewater treatment and the role of machine learning and sensor technologies in that vision?

Chemical and Environmental Regulations Questions

- Can you describe the chemical compositions of the coagulant and flocculant used at the facility?
- Why was there a switch from lime to PAX for coagulation during the last upgrade?
- Are there any environmental regulations or standards that particularly impact the plant operations?
- Do you think the treatment efficiencies from sensor readings can substitute the regular sampling required by the county municipality?

General and Historical Context Questions

- What significant changes have you noticed in the wastewater treatment industry during your tenure as a plant operator?
- When was the RA4 plant built and how many population equivalents (PE) are connected to it?

- How does the wastewater from the nearby hospital impact the treatment efficiency at the plant?
- What are the major challenges faced by the plant today and the potential future challenges?
- How would the wider community respond to the implementation of such an advanced system for process optimization at the plant?

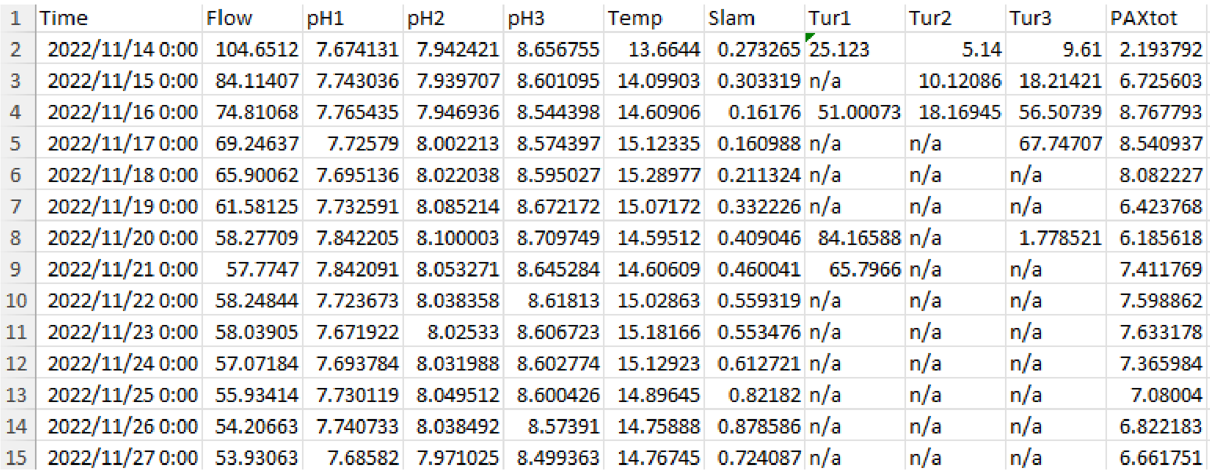
Staff Training and User Interface Questions

- What type of training or expertise would be necessary for the staff to effectively operate and maintain a system based on machine learning and sensor adjustments?
- How do you think this system would impact your job satisfaction, stress levels, or other aspects of your work life?
- What should be the key considerations in designing an interface for operators to interact with the new system?

Comparative Study Questions

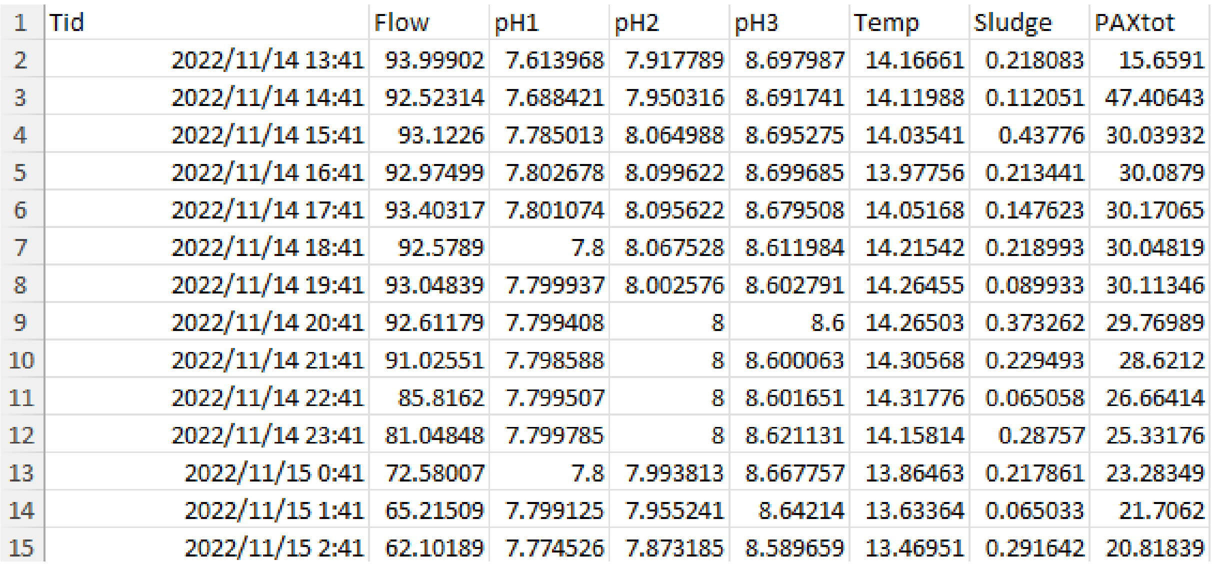
- Can you share any experiences from other plants or operators who have attempted similar optimization strategies?
- What potential barriers do you foresee for the implementation of this system on a larger scale across other plants?

B. Data description and preprocessing details



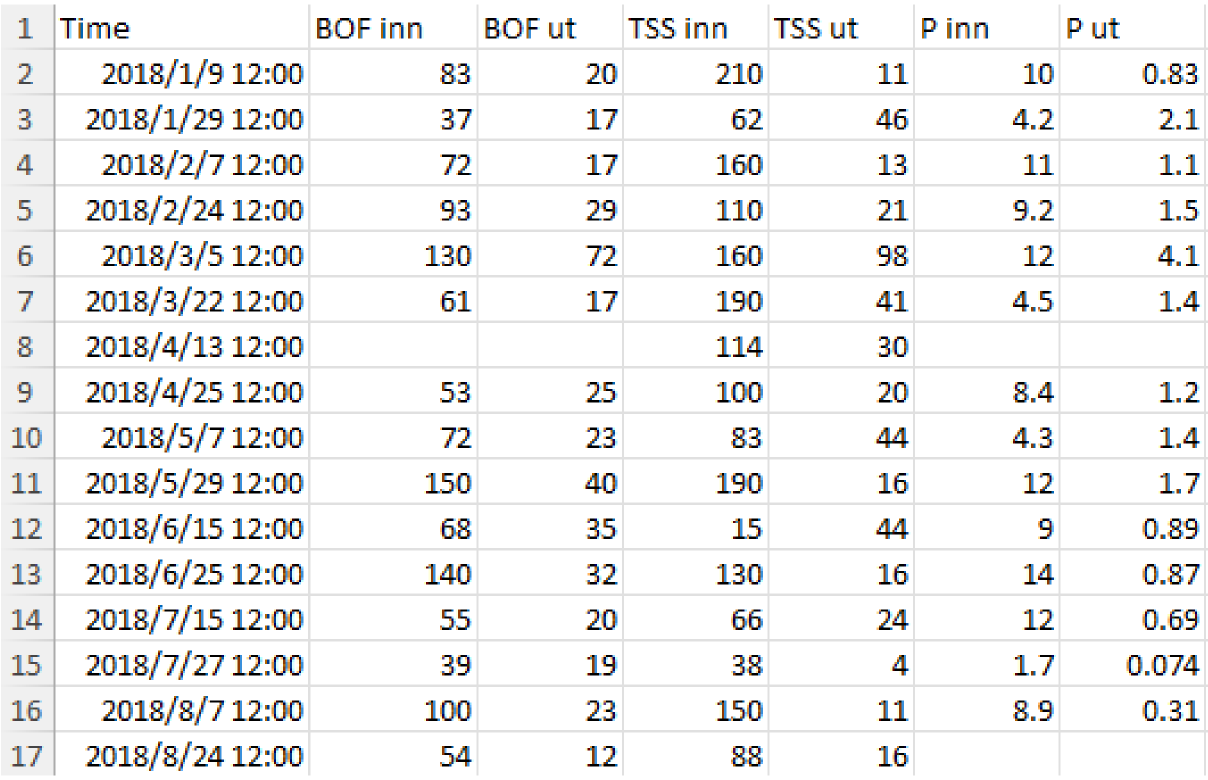
1	Time	Flow	pH1	pH2	pH3	Temp	Slam	Tur1	Tur2	Tur3	PAXtot
2	2022/11/14 0:00	104.6512	7.674131	7.942421	8.656755	13.6644	0.273265	25.123	5.14	9.61	2.193792
3	2022/11/15 0:00	84.11407	7.743036	7.939707	8.601095	14.09903	0.303319	n/a	10.12086	18.21421	6.725603
4	2022/11/16 0:00	74.81068	7.765435	7.946936	8.544398	14.60906	0.16176	51.00073	18.16945	56.50739	8.767793
5	2022/11/17 0:00	69.24637	7.72579	8.002213	8.574397	15.12335	0.160988	n/a	n/a	67.74707	8.540937
6	2022/11/18 0:00	65.90062	7.695136	8.022038	8.595027	15.28977	0.211324	n/a	n/a	n/a	8.082227
7	2022/11/19 0:00	61.58125	7.732591	8.085214	8.672172	15.07172	0.332226	n/a	n/a	n/a	6.423768
8	2022/11/20 0:00	58.27709	7.842205	8.100003	8.709749	14.59512	0.409046	84.16588	n/a	1.778521	6.185618
9	2022/11/21 0:00	57.7747	7.842091	8.053271	8.645284	14.60609	0.460041	65.7966	n/a	n/a	7.411769
10	2022/11/22 0:00	58.24844	7.723673	8.038358	8.61813	15.02863	0.559319	n/a	n/a	n/a	7.598862
11	2022/11/23 0:00	58.03905	7.671922	8.02533	8.606723	15.18166	0.553476	n/a	n/a	n/a	7.633178
12	2022/11/24 0:00	57.07184	7.693784	8.031988	8.602774	15.12923	0.612721	n/a	n/a	n/a	7.365984
13	2022/11/25 0:00	55.93414	7.730119	8.049512	8.600426	14.89645	0.82182	n/a	n/a	n/a	7.08004
14	2022/11/26 0:00	54.20663	7.740733	8.038492	8.57391	14.75888	0.878586	n/a	n/a	n/a	6.822183
15	2022/11/27 0:00	53.93063	7.68582	7.971025	8.499363	14.76745	0.724087	n/a	n/a	n/a	6.661751

Table 5. Dataset with 1 day resolution



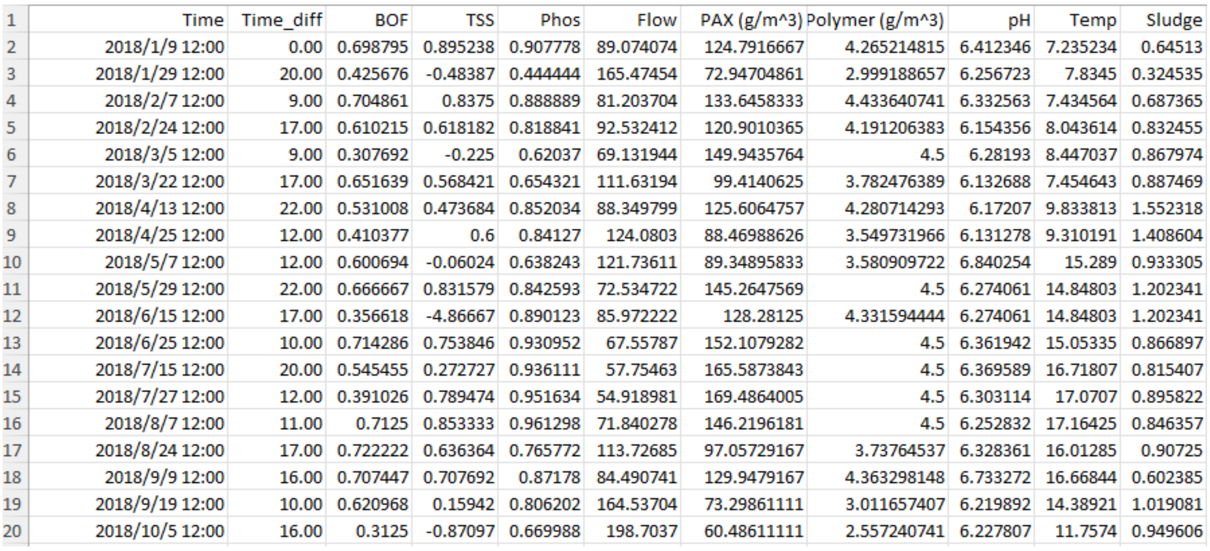
1	Tid	Flow	pH1	pH2	pH3	Temp	Sludge	PAXtot
2	2022/11/14 13:41	93.99902	7.613968	7.917789	8.697987	14.16661	0.218083	15.6591
3	2022/11/14 14:41	92.52314	7.688421	7.950316	8.691741	14.11988	0.112051	47.40643
4	2022/11/14 15:41	93.1226	7.785013	8.064988	8.695275	14.03541	0.43776	30.03932
5	2022/11/14 16:41	92.97499	7.802678	8.099622	8.699685	13.97756	0.213441	30.0879
6	2022/11/14 17:41	93.40317	7.801074	8.095622	8.679508	14.05168	0.147623	30.17065
7	2022/11/14 18:41	92.5789	7.8	8.067528	8.611984	14.21542	0.218993	30.04819
8	2022/11/14 19:41	93.04839	7.799937	8.002576	8.602791	14.26455	0.089933	30.11346
9	2022/11/14 20:41	92.61179	7.799408	8	8.6	14.26503	0.373262	29.76989
10	2022/11/14 21:41	91.02551	7.798588	8	8.600063	14.30568	0.229493	28.6212
11	2022/11/14 22:41	85.8162	7.799507	8	8.601651	14.31776	0.065058	26.66414
12	2022/11/14 23:41	81.04848	7.799785	8	8.621131	14.15814	0.28757	25.33176
13	2022/11/15 0:41	72.58007	7.8	7.993813	8.667757	13.86463	0.217861	23.28349
14	2022/11/15 1:41	65.21509	7.799125	7.955241	8.64214	13.63364	0.065033	21.7062
15	2022/11/15 2:41	62.10189	7.774526	7.873185	8.589659	13.46951	0.291642	20.81839

Table 6. Dataset with 1 hour resolution



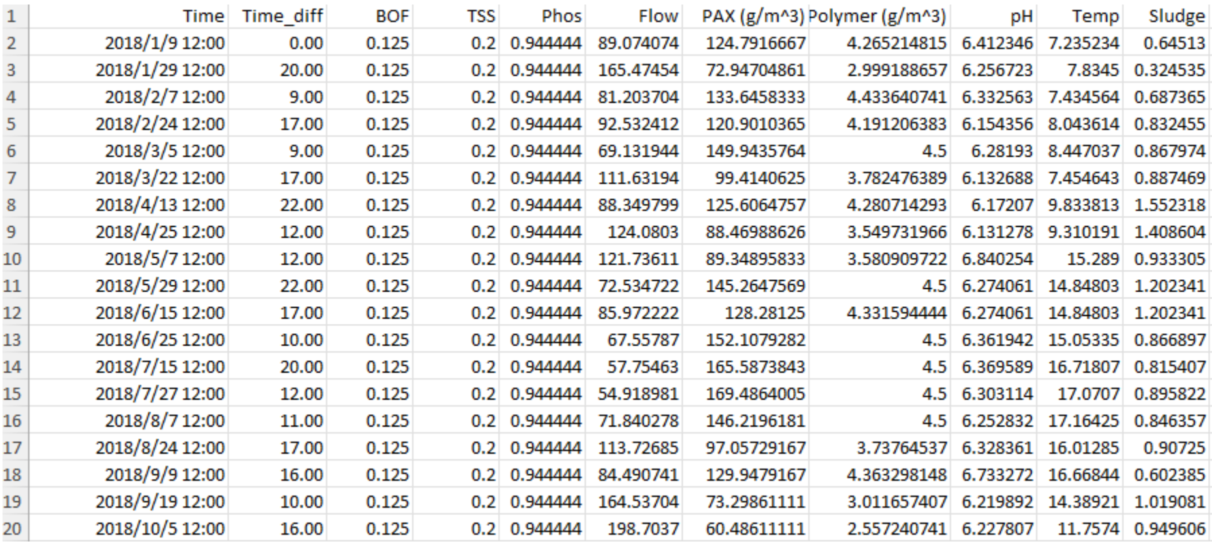
1	Time	BOF inn	BOF ut	TSS inn	TSS ut	P inn	P ut
2	2018/1/9 12:00	83	20	210	11	10	0.83
3	2018/1/29 12:00	37	17	62	46	4.2	2.1
4	2018/2/7 12:00	72	17	160	13	11	1.1
5	2018/2/24 12:00	93	29	110	21	9.2	1.5
6	2018/3/5 12:00	130	72	160	98	12	4.1
7	2018/3/22 12:00	61	17	190	41	4.5	1.4
8	2018/4/13 12:00			114	30		
9	2018/4/25 12:00	53	25	100	20	8.4	1.2
10	2018/5/7 12:00	72	23	83	44	4.3	1.4
11	2018/5/29 12:00	150	40	190	16	12	1.7
12	2018/6/15 12:00	68	35	15	44	9	0.89
13	2018/6/25 12:00	140	32	130	16	14	0.87
14	2018/7/15 12:00	55	20	66	24	12	0.69
15	2018/7/27 12:00	39	19	38	4	1.7	0.074
16	2018/8/7 12:00	100	23	150	11	8.9	0.31
17	2018/8/24 12:00	54	12	88	16		

Table 7. Dataset with regular treatment sampling



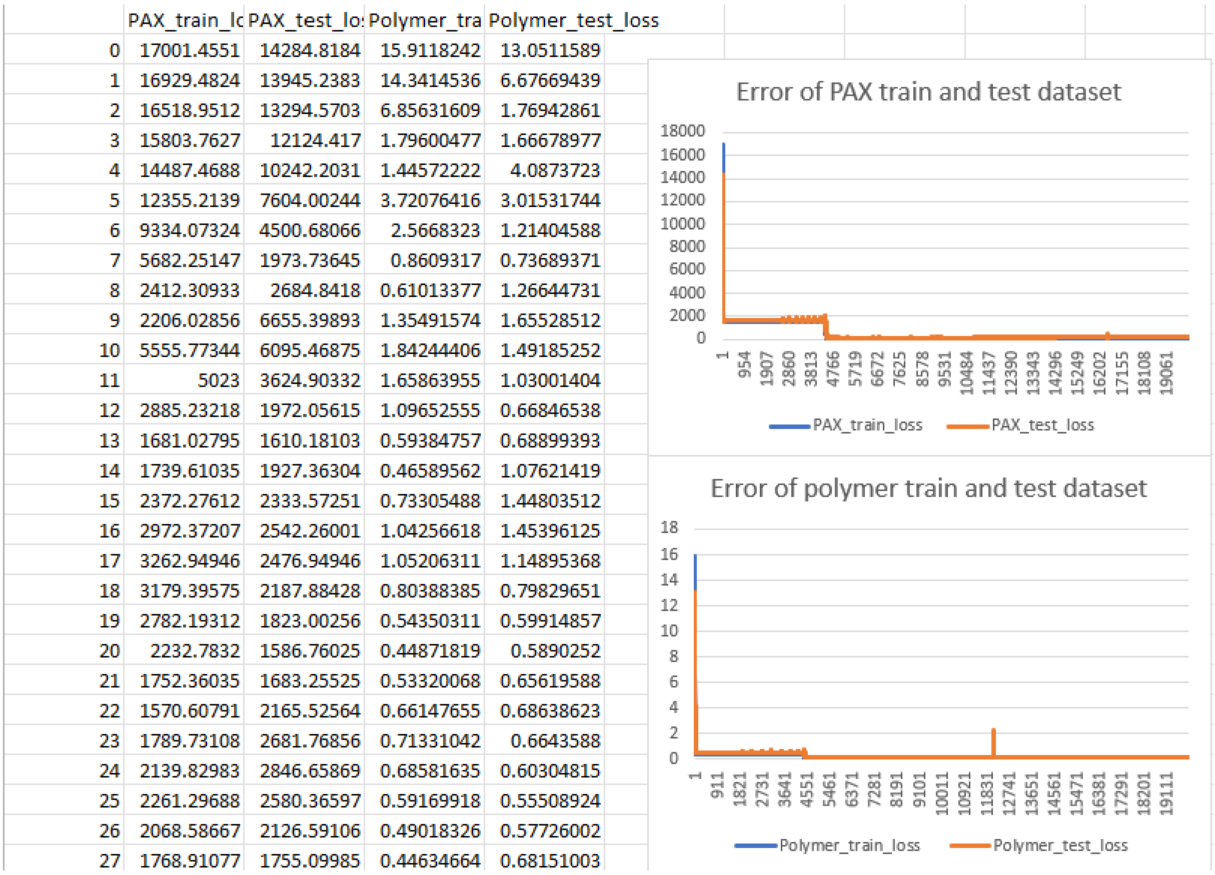
1	Time	Time_diff	BOF	TSS	Phos	Flow	PAX (g/m^3)	Polymer (g/m^3)	pH	Temp	Sludge
2	2018/1/9 12:00	0.00	0.698795	0.895238	0.907778	89.074074	124.7916667	4.265214815	6.412346	7.235234	0.64513
3	2018/1/29 12:00	20.00	0.425676	-0.48387	0.444444	165.47454	72.94704861	2.999188657	6.256723	7.8345	0.324535
4	2018/2/7 12:00	9.00	0.704861	0.8375	0.888889	81.203704	133.6458333	4.433640741	6.332563	7.434564	0.687365
5	2018/2/24 12:00	17.00	0.610215	0.618182	0.818841	92.532412	120.9010365	4.191206383	6.154356	8.043614	0.832455
6	2018/3/5 12:00	9.00	0.307692	-0.225	0.62037	69.131944	149.9435764	4.5	6.28193	8.447037	0.867974
7	2018/3/22 12:00	17.00	0.651639	0.568421	0.654321	111.63194	99.4140625	3.782476389	6.132688	7.454643	0.887469
8	2018/4/13 12:00	22.00	0.531008	0.473684	0.852034	88.349799	125.6064757	4.280714293	6.17207	9.833813	1.552318
9	2018/4/25 12:00	12.00	0.410377	0.6	0.84127	124.0803	88.46988626	3.549731966	6.131278	9.310191	1.408604
10	2018/5/7 12:00	12.00	0.600694	-0.06024	0.638243	121.73611	89.34895833	3.580909722	6.840254	15.289	0.933305
11	2018/5/29 12:00	22.00	0.666667	0.831579	0.842593	72.534722	145.2647569	4.5	6.274061	14.84803	1.202341
12	2018/6/15 12:00	17.00	0.356618	-4.86667	0.890123	85.972222	128.28125	4.331594444	6.274061	14.84803	1.202341
13	2018/6/25 12:00	10.00	0.714286	0.753846	0.930952	67.55787	152.1079282	4.5	6.361942	15.05335	0.866897
14	2018/7/15 12:00	20.00	0.545455	0.272727	0.936111	57.75463	165.5873843	4.5	6.369589	16.71807	0.815407
15	2018/7/27 12:00	12.00	0.391026	0.789474	0.951634	54.918981	169.4864005	4.5	6.303114	17.0707	0.895822
16	2018/8/7 12:00	11.00	0.7125	0.853333	0.961298	71.840278	146.2196181	4.5	6.252832	17.16425	0.846357
17	2018/8/24 12:00	17.00	0.722222	0.636364	0.765772	113.72685	97.05729167	3.73764537	6.328361	16.01285	0.90725
18	2018/9/9 12:00	16.00	0.707447	0.707692	0.87178	84.490741	129.9479167	4.363298148	6.733272	16.66844	0.602385
19	2018/9/19 12:00	10.00	0.620968	0.15942	0.806202	164.53704	73.29861111	3.011657407	6.219892	14.38921	1.019081
20	2018/10/5 12:00	16.00	0.3125	-0.87097	0.669988	198.7037	60.48611111	2.557240741	6.227807	11.7574	0.949606

Table 8. Final preprocessed and cleaned dataset for training and testing



1	Time	Time_diff	BOF	TSS	Phos	Flow	PAX (g/m^3)	Polymer (g/m^3)	pH	Temp	Sludge
2	2018/1/9 12:00	0.00	0.125	0.2	0.944444	89.074074	124.7916667	4.265214815	6.412346	7.235234	0.64513
3	2018/1/29 12:00	20.00	0.125	0.2	0.944444	165.47454	72.94704861	2.999188657	6.256723	7.8345	0.324535
4	2018/2/7 12:00	9.00	0.125	0.2	0.944444	81.203704	133.6458333	4.433640741	6.332563	7.434564	0.687365
5	2018/2/24 12:00	17.00	0.125	0.2	0.944444	92.532412	120.9010365	4.191206383	6.154356	8.043614	0.832455
6	2018/3/5 12:00	9.00	0.125	0.2	0.944444	69.131944	149.9435764	4.5	6.28193	8.447037	0.867974
7	2018/3/22 12:00	17.00	0.125	0.2	0.944444	111.63194	99.4140625	3.782476389	6.132688	7.454643	0.887469
8	2018/4/13 12:00	22.00	0.125	0.2	0.944444	88.349799	125.6064757	4.280714293	6.17207	9.833813	1.552318
9	2018/4/25 12:00	12.00	0.125	0.2	0.944444	124.0803	88.46988626	3.549731966	6.131278	9.310191	1.408604
10	2018/5/7 12:00	12.00	0.125	0.2	0.944444	121.73611	89.34895833	3.580909722	6.840254	15.289	0.933305
11	2018/5/29 12:00	22.00	0.125	0.2	0.944444	72.534722	145.2647569	4.5	6.274061	14.84803	1.202341
12	2018/6/15 12:00	17.00	0.125	0.2	0.944444	85.972222	128.28125	4.331594444	6.274061	14.84803	1.202341
13	2018/6/25 12:00	10.00	0.125	0.2	0.944444	67.55787	152.1079282	4.5	6.361942	15.05335	0.866897
14	2018/7/15 12:00	20.00	0.125	0.2	0.944444	57.75463	165.5873843	4.5	6.369589	16.71807	0.815407
15	2018/7/27 12:00	12.00	0.125	0.2	0.944444	54.918981	169.4864005	4.5	6.303114	17.0707	0.895822
16	2018/8/7 12:00	11.00	0.125	0.2	0.944444	71.840278	146.2196181	4.5	6.252832	17.16425	0.846357
17	2018/8/24 12:00	17.00	0.125	0.2	0.944444	113.72685	97.05729167	3.73764537	6.328361	16.01285	0.90725
18	2018/9/9 12:00	16.00	0.125	0.2	0.944444	84.490741	129.9479167	4.363298148	6.733272	16.66844	0.602385
19	2018/9/19 12:00	10.00	0.125	0.2	0.944444	164.53704	73.29861111	3.011657407	6.219892	14.38921	1.019081
20	2018/10/5 12:00	16.00	0.125	0.2	0.944444	198.7037	60.48611111	2.557240741	6.227807	11.7574	0.949606

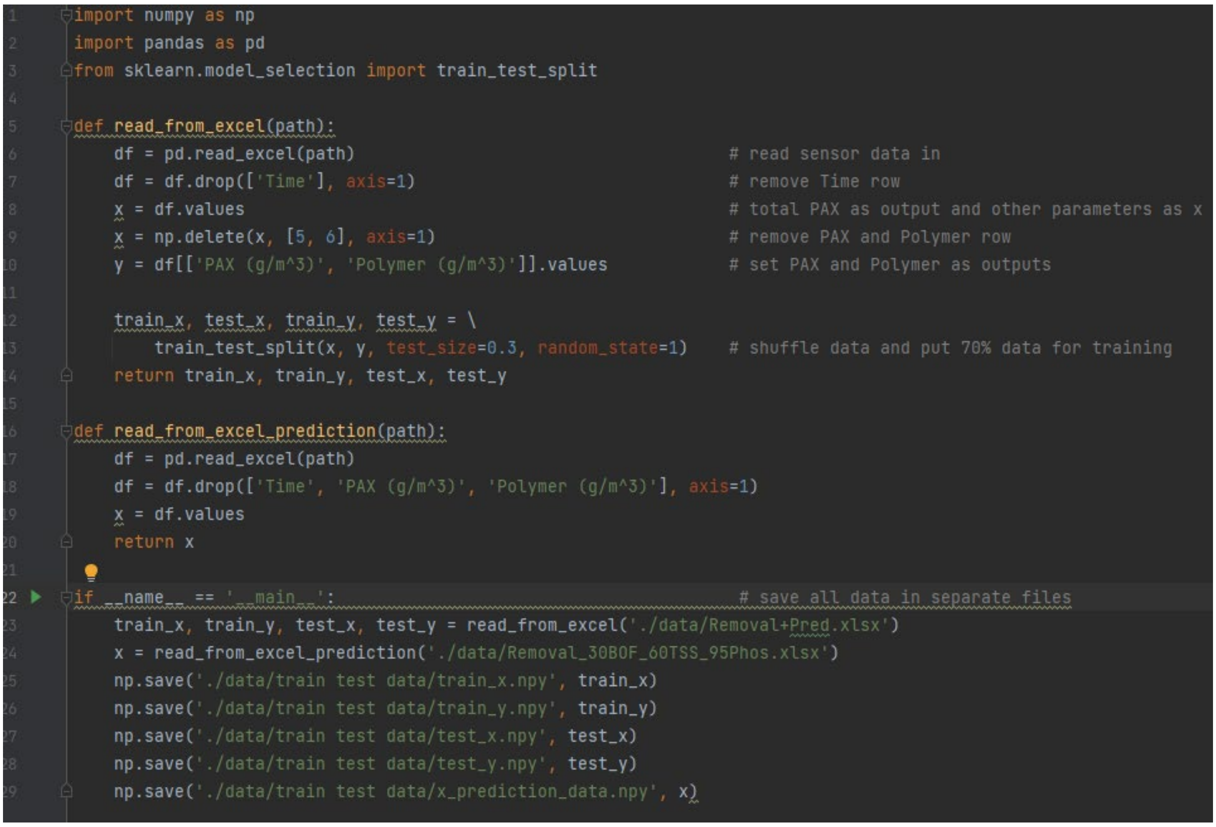
Table 9. Dataset with preset BOF, TSS and phosphorous removal for optimization



	PAX_train_lc	PAX_test_lo:	Polymer_tra	Polymer_test_loss
0	17001.4551	14284.8184	15.9118242	13.0511589
1	16929.4824	13945.2383	14.3414536	6.67669439
2	16518.9512	13294.5703	6.85631609	1.76942861
3	15803.7627	12124.417	1.79600477	1.66678977
4	14487.4688	10242.2031	1.44572222	4.0873723
5	12355.2139	7604.00244	3.72076416	3.01531744
6	9334.07324	4500.68066	2.5668323	1.21404588
7	5682.25147	1973.73645	0.8609317	0.73689371
8	2412.30933	2684.8418	0.61013377	1.26644731
9	2206.02856	6655.39893	1.35491574	1.65528512
10	5555.77344	6095.46875	1.84244406	1.49185252
11	5023	3624.90332	1.65863955	1.03001404
12	2885.23218	1972.05615	1.09652555	0.66846538
13	1681.02795	1610.18103	0.59384757	0.68899393
14	1739.61035	1927.36304	0.46589562	1.07621419
15	2372.27612	2333.57251	0.73305488	1.44803512
16	2972.37207	2542.26001	1.04256618	1.45396125
17	3262.94946	2476.94946	1.05206311	1.14895368
18	3179.39575	2187.88428	0.80388385	0.79829651
19	2782.19312	1823.00256	0.54350311	0.59914857
20	2232.7832	1586.76025	0.44871819	0.5890252
21	1752.36035	1683.25525	0.53320068	0.65619588
22	1570.60791	2165.52564	0.66147655	0.68638623
23	1789.73108	2681.76856	0.71331042	0.6643588
24	2139.82983	2846.65869	0.68581635	0.60304815
25	2261.29688	2580.36597	0.59169918	0.55508924
26	2068.58667	2126.59106	0.49018326	0.57726002
27	1768.91077	1755.09985	0.44634664	0.68151003

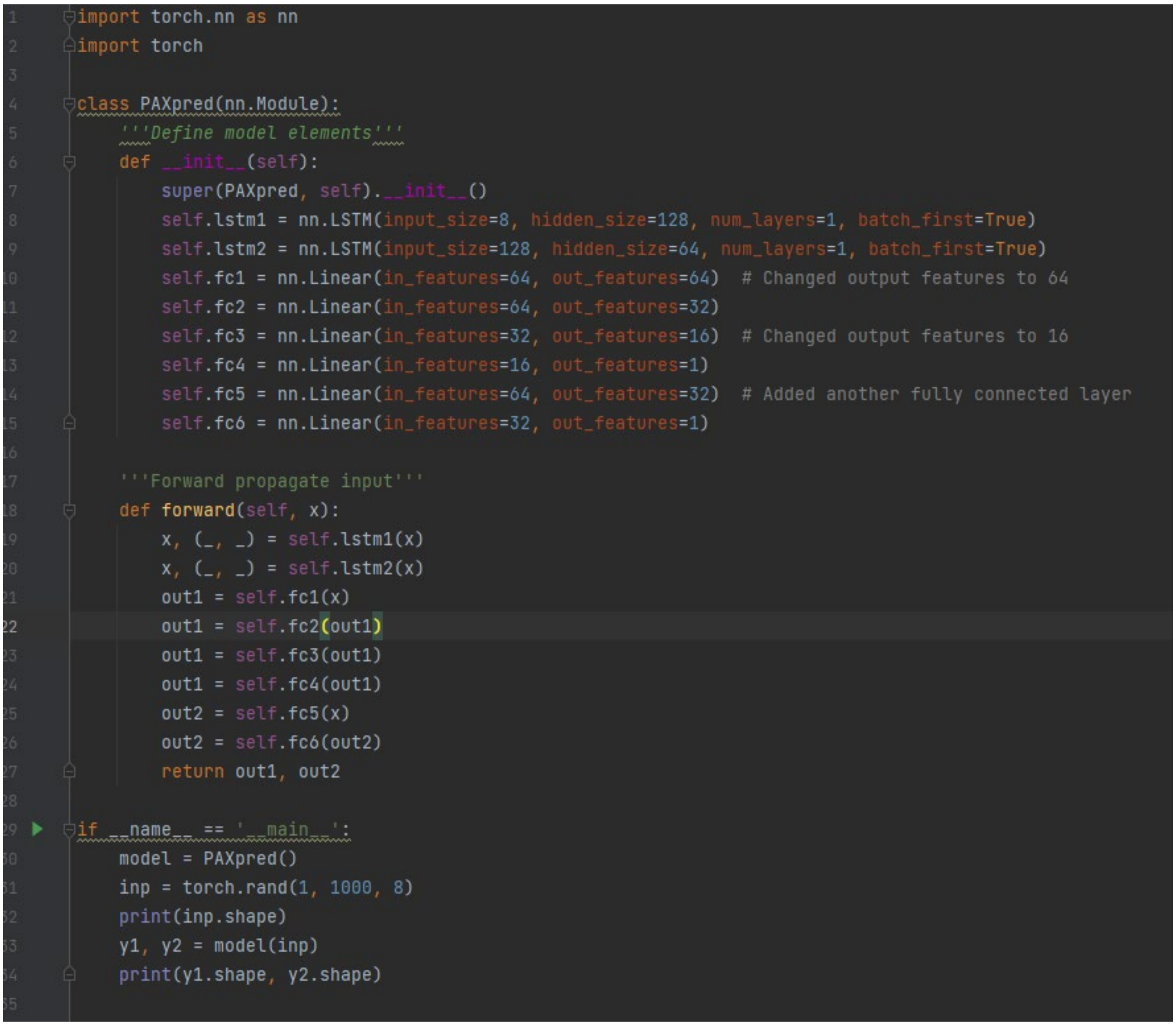
Table 10. MAE in each epoch for model performance evaluation

C. LSTM model implementation code



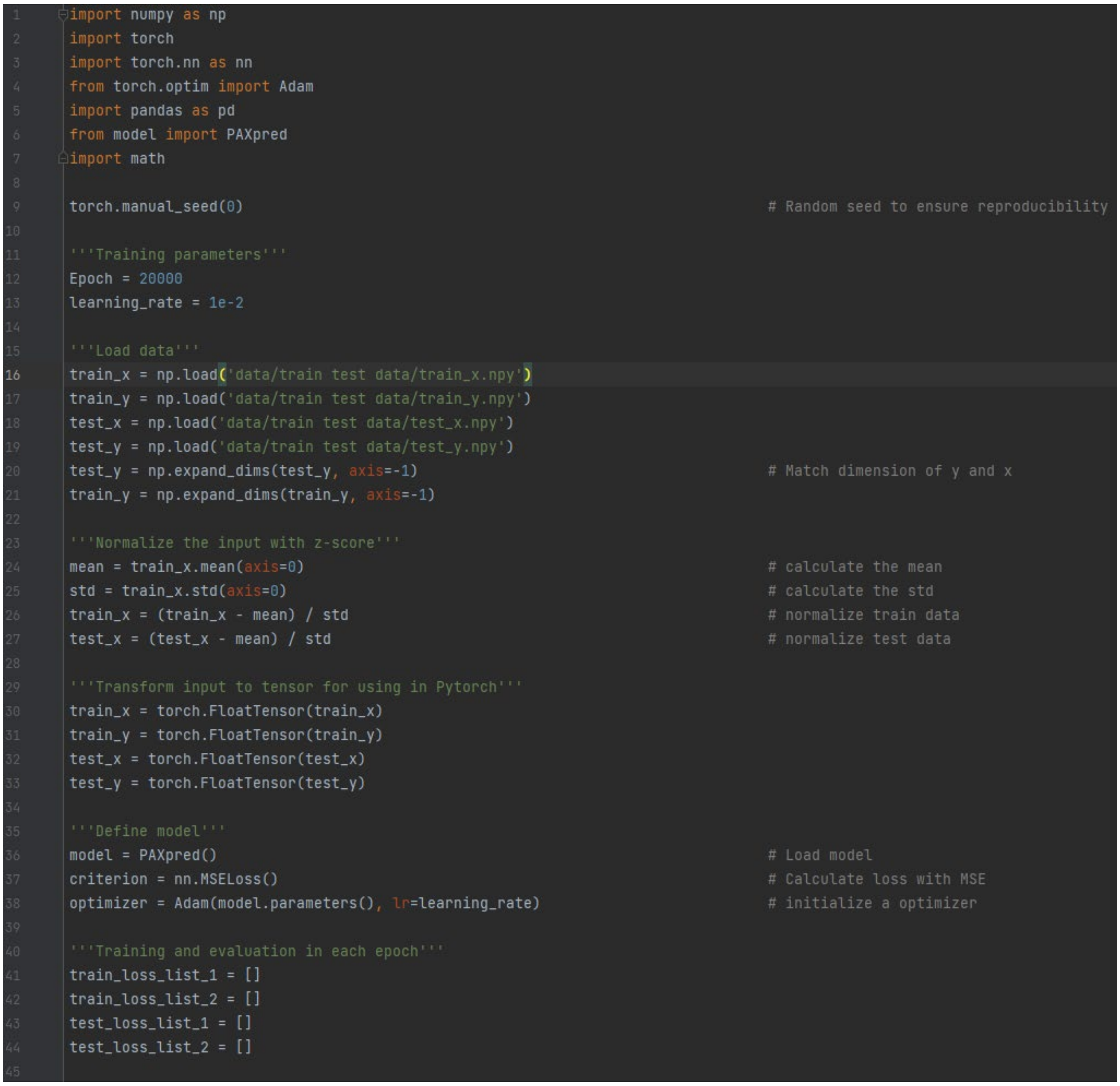
```python
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split

def read_from_excel(path):
    df = pd.read_excel(path)                                # read sensor data in
    df = df.drop(['Time'], axis=1)                          # remove Time row
    x = df.values                                           # total PAX as output and other parameters as x
    x = np.delete(x, [5, 6], axis=1)                        # remove PAX and Polymer row
    y = df[['PAX (g/m^3)', 'Polymer (g/m^3)']].values       # set PAX and Polymer as outputs

    train_x, test_x, train_y, test_y = \
        train_test_split(x, y, test_size=0.3, random_state=1)    # shuffle data and put 70% data for training
    return train_x, train_y, test_x, test_y

def read_from_excel_prediction(path):
    df = pd.read_excel(path)
    df = df.drop(['Time', 'PAX (g/m^3)', 'Polymer (g/m^3)'], axis=1)
    x = df.values
    return x

if __name__ == '__main__':                                   # save all data in separate files
    train_x, train_y, test_x, test_y = read_from_excel('./data/Removal+Pred.xlsx')
    x = read_from_excel_prediction('./data/Removal_30BOF_60TSS_95Phos.xlsx')
    np.save('./data/train test data/train_x.npy', train_x)
    np.save('./data/train test data/train_y.npy', train_y)
    np.save('./data/train test data/test_x.npy', test_x)
    np.save('./data/train test data/test_y.npy', test_y)
    np.save('./data/train test data/x_prediction_data.npy', x)
```

Figure 24. Python script utilis.py, read data and save them as training and testing data files



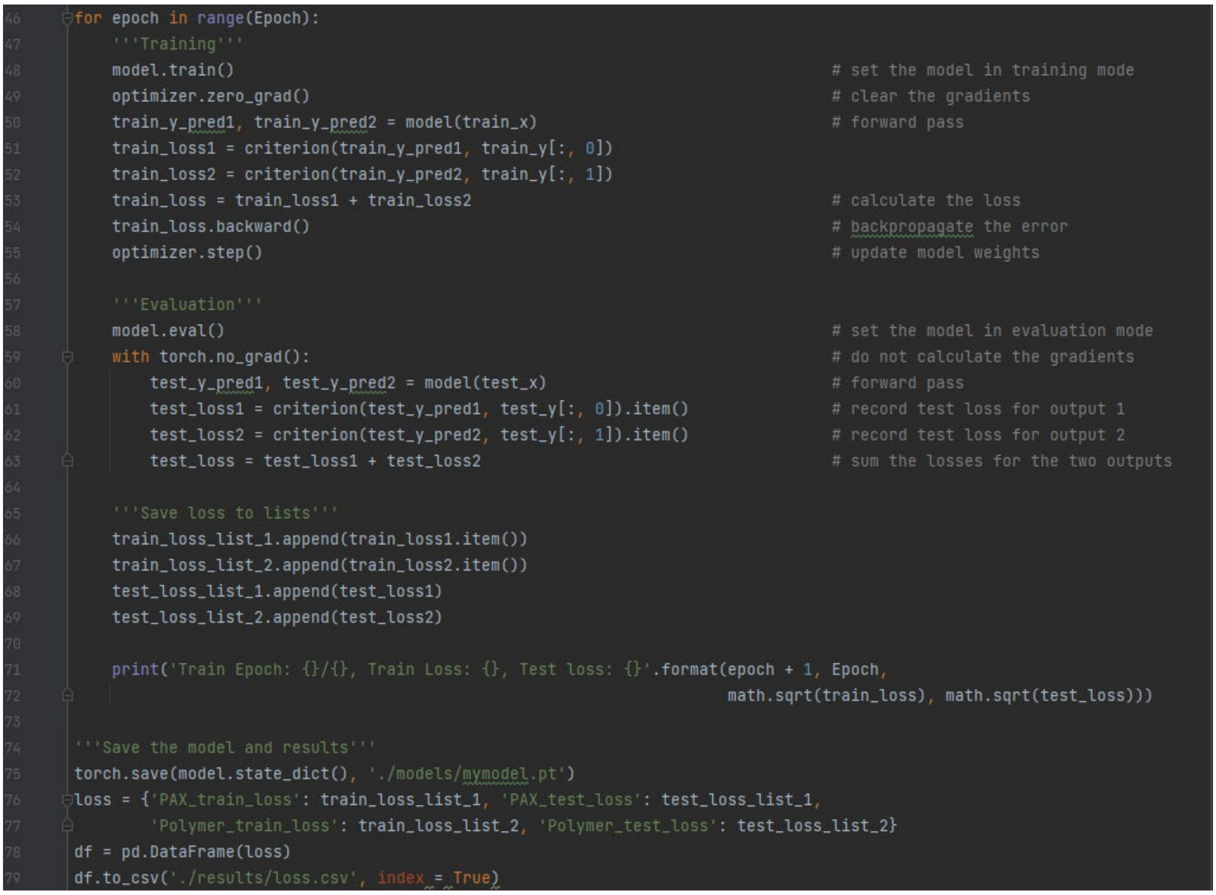
```python
import torch.nn as nn
import torch

class PAXpred(nn.Module):
    '''Define model elements'''
    def __init__(self):
        super(PAXpred, self).__init__()
        self.lstm1 = nn.LSTM(input_size=8, hidden_size=128, num_layers=1, batch_first=True)
        self.lstm2 = nn.LSTM(input_size=128, hidden_size=64, num_layers=1, batch_first=True)
        self.fc1 = nn.Linear(in_features=64, out_features=64)  # Changed output features to 64
        self.fc2 = nn.Linear(in_features=64, out_features=32)
        self.fc3 = nn.Linear(in_features=32, out_features=16)  # Changed output features to 16
        self.fc4 = nn.Linear(in_features=16, out_features=1)
        self.fc5 = nn.Linear(in_features=64, out_features=32)  # Added another fully connected layer
        self.fc6 = nn.Linear(in_features=32, out_features=1)

    '''Forward propagate input'''
    def forward(self, x):
        x, (_, _) = self.lstm1(x)
        x, (_, _) = self.lstm2(x)
        out1 = self.fc1(x)
        out1 = self.fc2(out1)
        out1 = self.fc3(out1)
        out1 = self.fc4(out1)
        out2 = self.fc5(x)
        out2 = self.fc6(out2)
        return out1, out2

if __name__ == '__main__':
    model = PAXpred()
    inp = torch.rand(1, 1000, 8)
    print(inp.shape)
    y1, y2 = model(inp)
    print(y1.shape, y2.shape)
```

Figure 25. Python script model.py, define architecture of the neural network



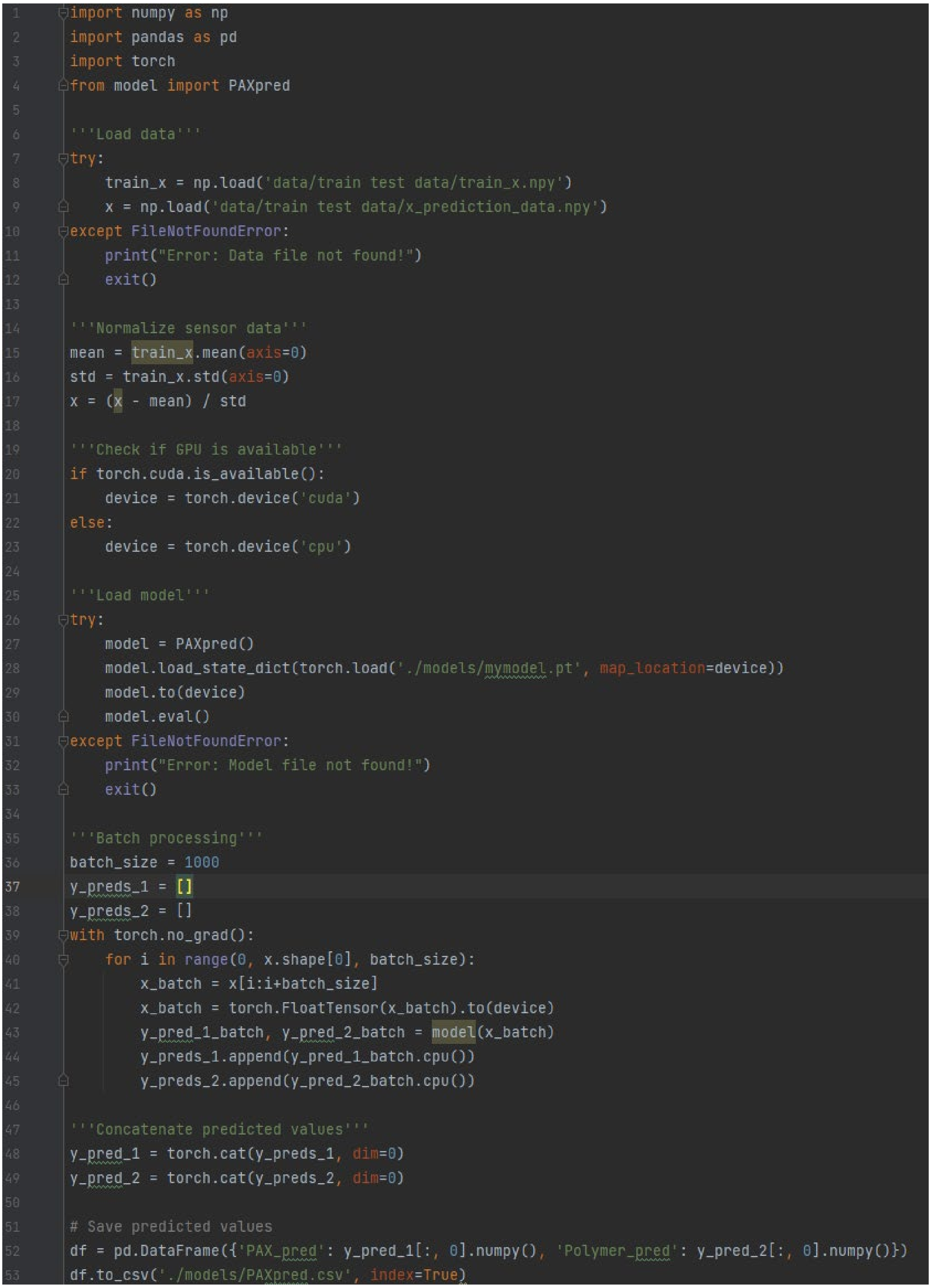
```python
import numpy as np
import torch
import torch.nn as nn
from torch.optim import Adam
import pandas as pd
from model import PAXpred
import math

torch.manual_seed(0)                                                            # Random seed to ensure reproducibility

'''Training parameters'''
Epoch = 20000
learning_rate = 1e-2

'''Load data'''
train_x = np.load('data/train test data/train_x.npy')
train_y = np.load('data/train test data/train_y.npy')
test_x = np.load('data/train test data/test_x.npy')
test_y = np.load('data/train test data/test_y.npy')
test_y = np.expand_dims(test_y, axis=-1)                                        # Match dimension of y and x
train_y = np.expand_dims(train_y, axis=-1)

'''Normalize the input with z-score'''
mean = train_x.mean(axis=0)                                                     # calculate the mean
std = train_x.std(axis=0)                                                       # calculate the std
train_x = (train_x - mean) / std                                                # normalize train data
test_x = (test_x - mean) / std                                                  # normalize test data

'''Transform input to tensor for using in Pytorch'''
train_x = torch.FloatTensor(train_x)
train_y = torch.FloatTensor(train_y)
test_x = torch.FloatTensor(test_x)
test_y = torch.FloatTensor(test_y)

'''Define model'''
model = PAXpred()                                                               # Load model
criterion = nn.MSELoss()                                                        # Calculate loss with MSE
optimizer = Adam(model.parameters(), lr=learning_rate)                          # initialize a optimizer

'''Training and evaluation in each epoch'''
train_loss_list_1 = []
train_loss_list_2 = []
test_loss_list_1 = []
test_loss_list_2 = []

```

Figure 26. Python script train&evaluate.py part 1, data-preprocess, save the model and training loss



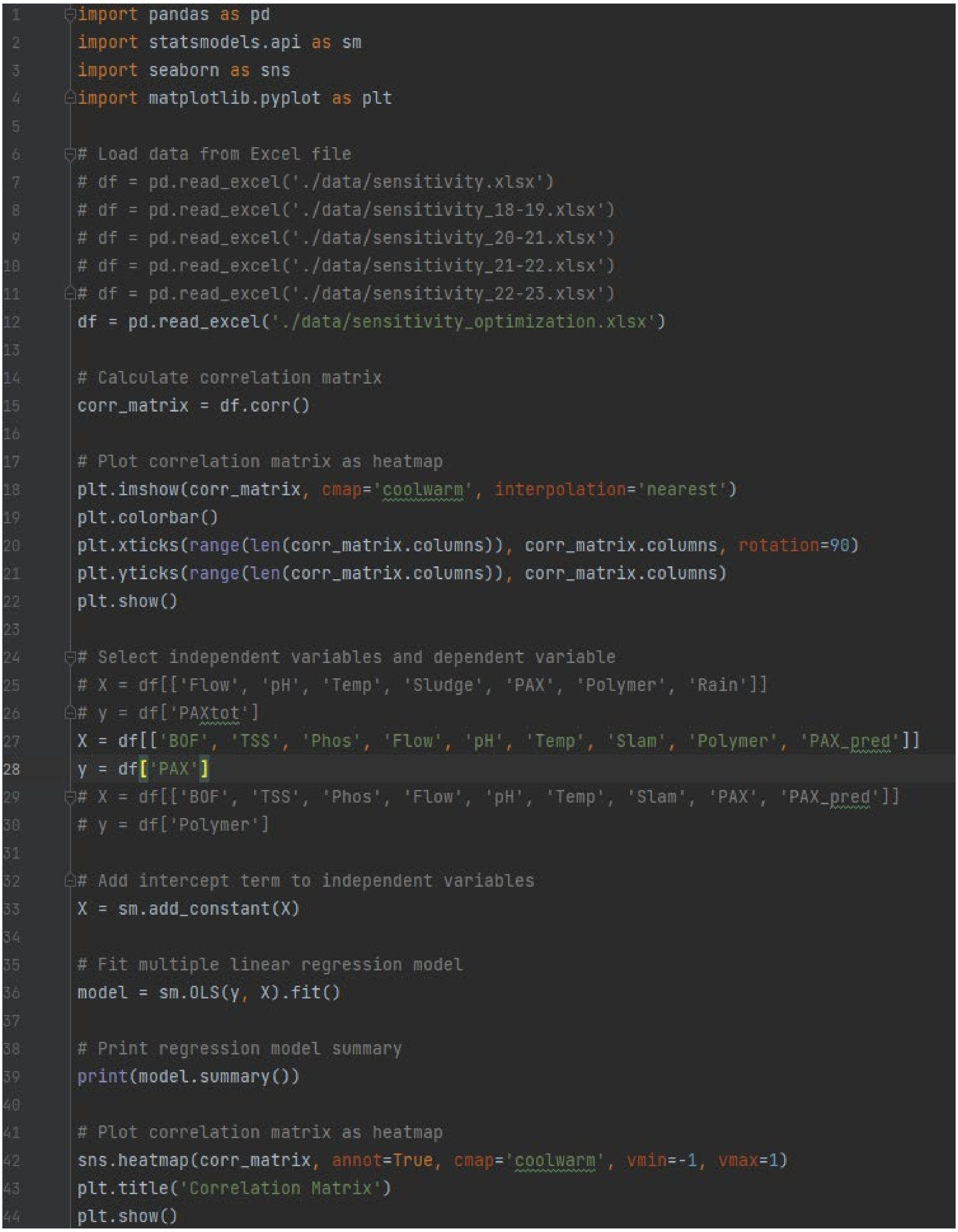
```python
for epoch in range(Epoch):
    '''Training'''
    model.train()                                                               # set the model in training mode
    optimizer.zero_grad()                                                       # clear the gradients
    train_y_pred1, train_y_pred2 = model(train_x)                               # forward pass
    train_loss1 = criterion(train_y_pred1, train_y[:, 0])
    train_loss2 = criterion(train_y_pred2, train_y[:, 1])
    train_loss = train_loss1 + train_loss2                                      # calculate the loss
    train_loss.backward()                                                       # backpropagate the error
    optimizer.step()                                                            # update model weights

    '''Evaluation'''
    model.eval()                                                                # set the model in evaluation mode
    with torch.no_grad():                                                       # do not calculate the gradients
        test_y_pred1, test_y_pred2 = model(test_x)                              # forward pass
        test_loss1 = criterion(test_y_pred1, test_y[:, 0]).item()               # record test loss for output 1
        test_loss2 = criterion(test_y_pred2, test_y[:, 1]).item()               # record test loss for output 2
        test_loss = test_loss1 + test_loss2                                     # sum the losses for the two outputs

    '''Save loss to lists'''
    train_loss_list_1.append(train_loss1.item())
    train_loss_list_2.append(train_loss2.item())
    test_loss_list_1.append(test_loss1)
    test_loss_list_2.append(test_loss2)

    print('Train Epoch: {}/{}, Train Loss: {}, Test loss: {}'.format(epoch + 1, Epoch,
                                                                    math.sqrt(train_loss), math.sqrt(test_loss)))

'''Save the model and results'''
torch.save(model.state_dict(), './models/mymodel.pt')
loss = {'PAX_train_loss': train_loss_list_1, 'PAX_test_loss': test_loss_list_1,
        'Polymer_train_loss': train_loss_list_2, 'Polymer_test_loss': test_loss_list_2}
df = pd.DataFrame(loss)
df.to_csv('./results/loss.csv', index = True)
```

Figure 27. Python script train&evaluate.py part 2



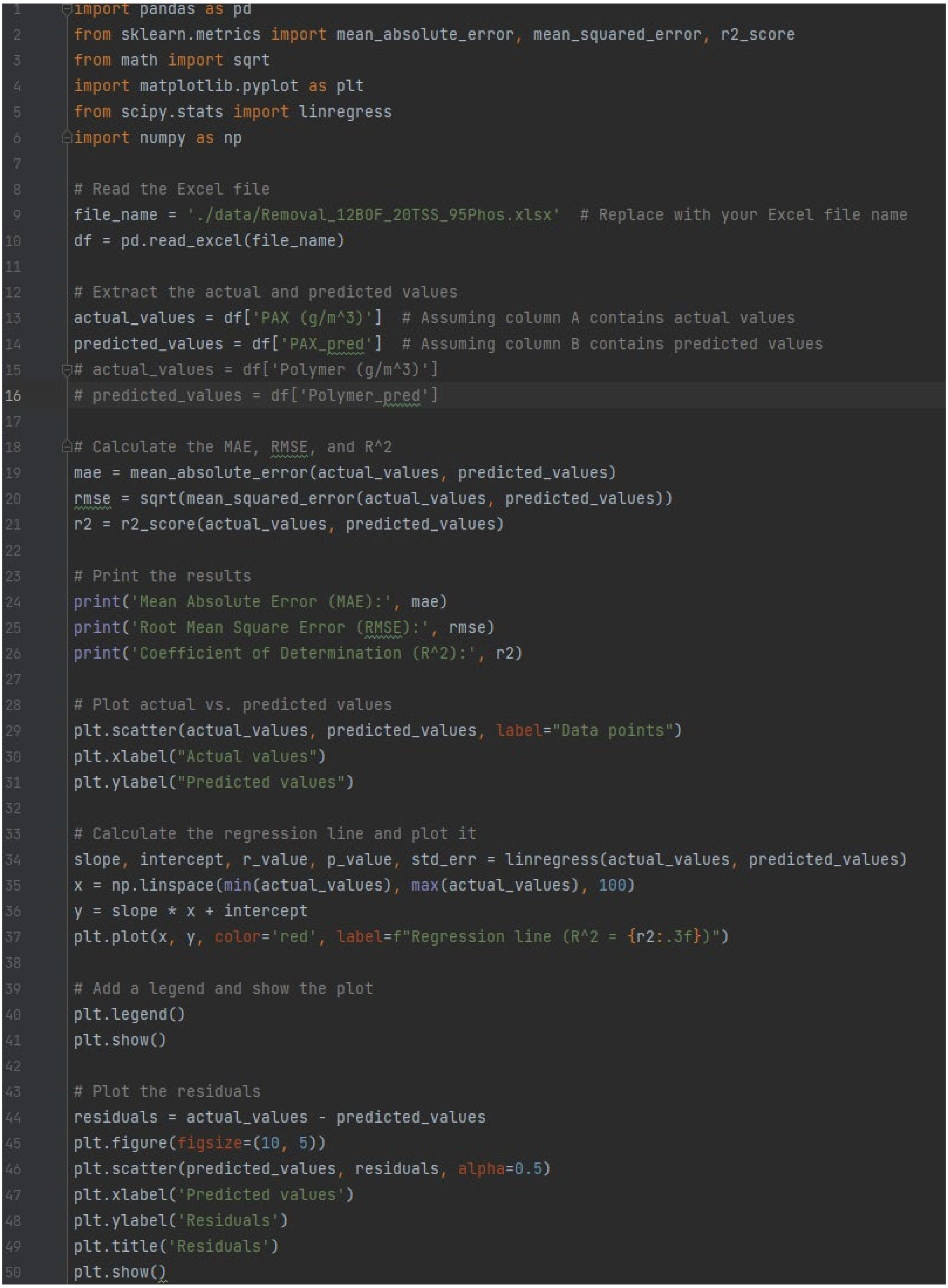
```python
import numpy as np
import pandas as pd
import torch
from model import PAXpred

'''Load data'''
try:
    train_x = np.load('data/train test data/train_x.npy')
    x = np.load('data/train test data/x_prediction_data.npy')
except FileNotFoundError:
    print("Error: Data file not found!")
    exit()

'''Normalize sensor data'''
mean = train_x.mean(axis=0)
std = train_x.std(axis=0)
x = (x - mean) / std

'''Check if GPU is available'''
if torch.cuda.is_available():
    device = torch.device('cuda')
else:
    device = torch.device('cpu')

'''Load model'''
try:
    model = PAXpred()
    model.load_state_dict(torch.load('./models/mymodel.pt', map_location=device))
    model.to(device)
    model.eval()
except FileNotFoundError:
    print("Error: Model file not found!")
    exit()

'''Batch processing'''
batch_size = 1000
y_preds_1 = []
y_preds_2 = []
with torch.no_grad():
    for i in range(0, x.shape[0], batch_size):
        x_batch = x[i:i+batch_size]
        x_batch = torch.FloatTensor(x_batch).to(device)
        y_pred_1_batch, y_pred_2_batch = model(x_batch)
        y_preds_1.append(y_pred_1_batch.cpu())
        y_preds_2.append(y_pred_2_batch.cpu())

'''Concatenate predicted values'''
y_pred_1 = torch.cat(y_preds_1, dim=0)
y_pred_2 = torch.cat(y_preds_2, dim=0)

# Save predicted values
df = pd.DataFrame({'PAX_pred': y_pred_1[:, 0].numpy(), 'Polymer_pred': y_pred_2[:, 0].numpy()})
df.to_csv('./models/PAXpred.csv', index=True)
```

Figure 28. Python script prediction.py, make PAX and Polymer prediction and save the result



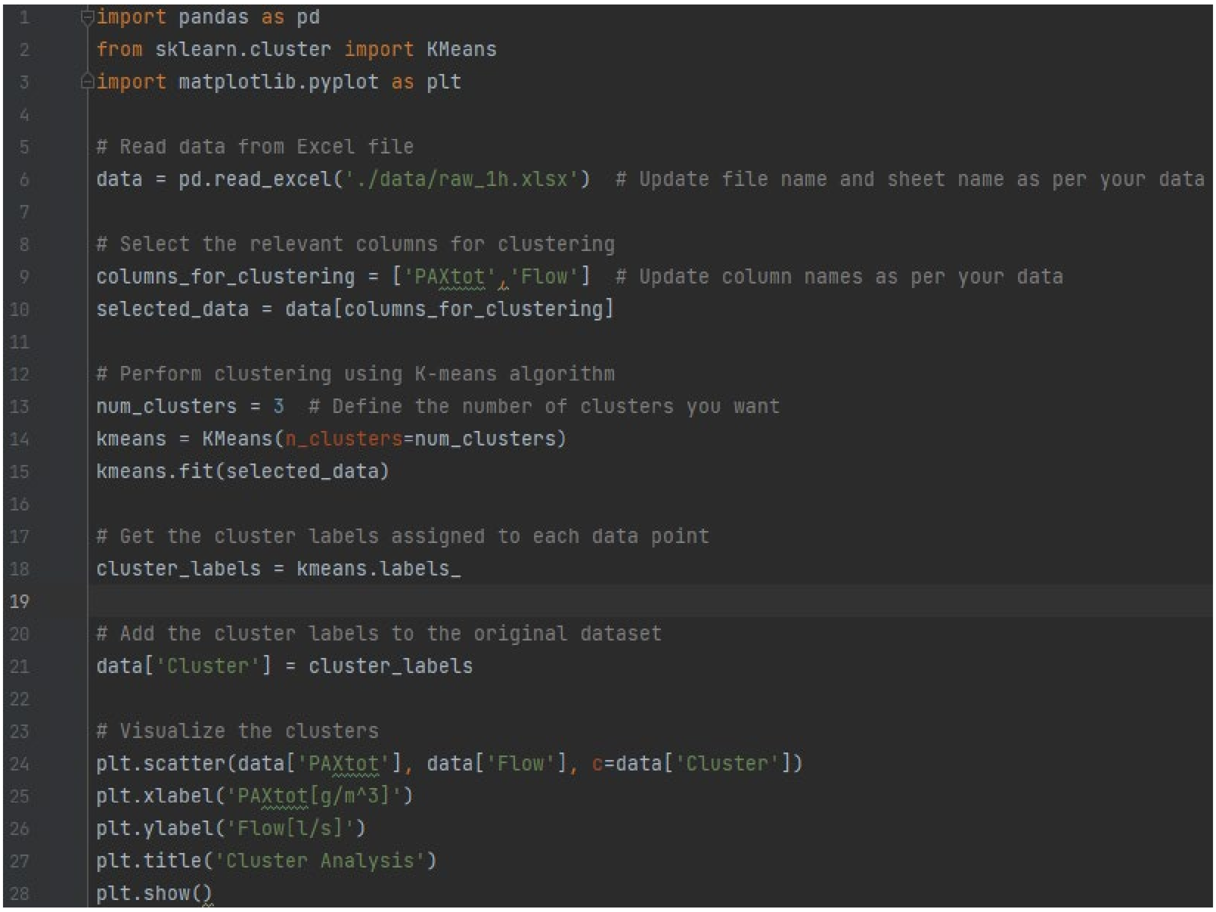
```
import pandas as pd
import statsmodels.api as sm
import seaborn as sns
import matplotlib.pyplot as plt

# Load data from Excel file
# df = pd.read_excel('./data/sensitivity.xlsx')
# df = pd.read_excel('./data/sensitivity_18-19.xlsx')
# df = pd.read_excel('./data/sensitivity_20-21.xlsx')
# df = pd.read_excel('./data/sensitivity_21-22.xlsx')
# df = pd.read_excel('./data/sensitivity_22-23.xlsx')
df = pd.read_excel('./data/sensitivity_optimization.xlsx')

# Calculate correlation matrix
corr_matrix = df.corr()

# Plot correlation matrix as heatmap
plt.imshow(corr_matrix, cmap='coolwarm', interpolation='nearest')
plt.colorbar()
plt.xticks(range(len(corr_matrix.columns)), corr_matrix.columns, rotation=90)
plt.yticks(range(len(corr_matrix.columns)), corr_matrix.columns)
plt.show()

# Select independent variables and dependent variable
# X = df[['Flow', 'pH', 'Temp', 'Sludge', 'PAX', 'Polymer', 'Rain']]
# y = df['PAXtot']
X = df[['BOF', 'TSS', 'Phos', 'Flow', 'pH', 'Temp', 'Slam', 'Polymer', 'PAX_pred']]
y = df['PAX']
# X = df[['BOF', 'TSS', 'Phos', 'Flow', 'pH', 'Temp', 'Slam', 'PAX', 'PAX_pred']]
# y = df['Polymer']

# Add intercept term to independent variables
X = sm.add_constant(X)

# Fit multiple linear regression model
model = sm.OLS(y, X).fit()

# Print regression model summary
print(model.summary())

# Plot correlation matrix as heatmap
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1)
plt.title('Correlation Matrix')
plt.show()
```

Figure 29. Python script statistic.py, plot correlation of input variables and predict model performance



```python
import pandas as pd
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from math import sqrt
import matplotlib.pyplot as plt
from scipy.stats import linregress
import numpy as np

# Read the Excel file
file_name = './data/Removal_12BOF_20TSS_95Phos.xlsx'  # Replace with your Excel file name
df = pd.read_excel(file_name)

# Extract the actual and predicted values
actual_values = df['PAX (g/m^3)']  # Assuming column A contains actual values
predicted_values = df['PAX_pred']  # Assuming column B contains predicted values
# actual_values = df['Polymer (g/m^3)']
# predicted_values = df['Polymer_pred']

# Calculate the MAE, RMSE, and R^2
mae = mean_absolute_error(actual_values, predicted_values)
rmse = sqrt(mean_squared_error(actual_values, predicted_values))
r2 = r2_score(actual_values, predicted_values)

# Print the results
print('Mean Absolute Error (MAE):', mae)
print('Root Mean Square Error (RMSE):', rmse)
print('Coefficient of Determination (R^2):', r2)

# Plot actual vs. predicted values
plt.scatter(actual_values, predicted_values, label="Data points")
plt.xlabel("Actual values")
plt.ylabel("Predicted values")

# Calculate the regression line and plot it
slope, intercept, r_value, p_value, std_err = linregress(actual_values, predicted_values)
x = np.linspace(min(actual_values), max(actual_values), 100)
y = slope * x + intercept
plt.plot(x, y, color='red', label=f"Regression line (R^2 = {r2:.3f})")

# Add a legend and show the plot
plt.legend()
plt.show()

# Plot the residuals
residuals = actual_values - predicted_values
plt.figure(figsize=(10, 5))
plt.scatter(predicted_values, residuals, alpha=0.5)
plt.xlabel('Predicted values')
plt.ylabel('Residuals')
plt.title('Residuals')
plt.show()
```

Figure 30. Python script performance.py, calculate MAE, RMSE and R^2 to indicate performance of trained model, and plot the results



```python
import pandas as pd
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Read data from Excel file
data = pd.read_excel('./data/raw_1h.xlsx')  # Update file name and sheet name as per your data

# Select the relevant columns for clustering
columns_for_clustering = ['PAXtot','Flow']  # Update column names as per your data
selected_data = data[columns_for_clustering]

# Perform clustering using K-means algorithm
num_clusters = 3  # Define the number of clusters you want
kmeans = KMeans(n_clusters=num_clusters)
kmeans.fit(selected_data)

# Get the cluster labels assigned to each data point
cluster_labels = kmeans.labels_

# Add the cluster labels to the original dataset
data['Cluster'] = cluster_labels

# Visualize the clusters
plt.scatter(data['PAXtot'], data['Flow'], c=data['Cluster'])
plt.xlabel('PAXtot[g/m^3]')
plt.ylabel('Flow[l/s]')
plt.title('Cluster Analysis')
plt.show()
```

Figure 31. Python script cluster.py, visualize clusters for different variables in single plot to perform cluster analysis

NTNU

Norwegian University of
Science and Technology