Study 1: Algorithm Selections and Data Pre-processing Methodologies in Developing Machine Learning Models for NH3N Forecasting.  
Study 2: To incoporate the use of domain knowledge in water/wastewater treatment and established methodology to construct a optimized model for NHN forecasting.

# 1 Introduction

## 1.1 Background

## 1.2 Objectives

### 1.2.1 To evaluate baseline model performance in forecasting NH by developing models with traditional and deep learning algorithms.

### 1.2.2 To develop raw data cleaning methodologies for improved machine learning model performance of forecasting NHN.

### 1.2.3 To explore feature engineering with the use of domain knowledge in wastewater treatment to creating new variables for optimizing model performance of forecasting NHN.

### 1.2.4 To construct an optimized procedure of training a machine learning model for NHN forecasting.

## 1.3 Organization of the thesis

# 2 Literature Review

## 2.1 Water quality forecasting in wastewater treatment plant

### 2.1.1 Tools and technologies for parameter forecasting in wastewater treatment plant

## 2.2 Machine learning models for water quality forecasting

### 2.2.1 Introduction to time-series data

### 2.2.2 Machine learning models and comparison

### 2.2.3 Review of existing cases of applying machine learning for water quality forecasting

## 2.3 Techniques for improving model forecasting performance

### 2.3.1 Data pre-processing with smoothing and outlier removal

### 2.3.2 Implementation of weight regularization to avoid model overfittings

### 2.3.3 Other regularization methods to avoid model overfittings

# 3 Materials and methods

## 3.1 Wastewater treatment plant description

### 3.1.1 Treatment processes

### 3.1.2 Historical water quality data

### 3.1.3 Reclaimed water standard

## 3.2 Data collection and preparation

### 3.2.1 NH3-N data monitoring and collection

### 3.2.2 Data cleaning and pre-processing

#### 3.2.2.1 Data smoothing with Savitzky-Golay filter

#### 3.2.2.2 Data smoothing with EMCA filter

#### 3.2.2.3 Outlier detection and removal

### 3.2.3 Data transformation

#### 3.2.3.1 Split of Train/valid/test dataset

## 3.3 Architecture design of the selected baseline models

### 3.3.1 Model A (LSTM)

### 3.3.2 Model B (DNN)

### 3.3.3 ……

## 3.4 Implementation of regularization on machine learning models

### 3.4.1 Early-stopping

### 3.4.2 Dropout

### 3.4.3 Weight regularization

## 3.5 Proposed time series forecasting workflow

# 4 Results and Discussion

## 4.1 Comparisons of forecast accuracy in statistical and machine learning models

## 4.2 The effect of data cleaning on forecast accuracy

### 4.2.1 Data smoothing

### 4.2.2 Outlier removal

## 4.3 The effect of regularization techniques on forecast accuracy

### 4.3.1 Early-stopping

### 4.3.2 Dropout

### 4.3.3 Weight regularization

## 4.4 The effect of input training datasets on the stability of forecast models

### 4.4.1 Selection of the data training size

### 4.4.2 Update input training dataset with up-to-date data

### 4.4.3 Cross-validation

# 5 Conclusions and recommendations