

# **Forecasting the Color and Ammonia Concentration in the Reclaimed Water using Deep Learning**

by

**Ting Hsi LEE**

A Thesis Submitted to  
The Hong Kong University of Science and Technology  
in Partial Fulfillment of the Requirements for  
the Degree of Master of Philosophy  
in the Department of Civil and Environmental Engineering

July 2022, Hong Kong

## **Authorization**

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

---

Ting Hsi LEE

July 2022

# Forecasting the Color and Ammonia Concentration in the Reclaimed Water using Deep Learning

by

Ting Hsi LEE

This is to certify that I have examined the above MPhil thesis  
and have found that it is complete and satisfactory in all respects,  
and that any and all revisions required by  
the thesis examination committee have been made.

---

Prof. Chii SHANG, Thesis Supervisor

---

Prof. Meimei Han, Head of Department

Department of Civil and Environmental Engineering  
July 2022

# Acknowledgments

First of all, I am truly grateful for being one of the first PhD students supervised by Prof. Li. He was full of passion and patience when helping me build the know-how for this degree. It has been a great pleasure for me to be part of this team and grow together with the lab during the last four years. Furthermore, I would like to thank all of the members of the thesis examination committee for their careful examination of my thesis.

Finally, I would not stand at this current point without their endless love and unconditional support for all these years.

# TABLE OF CONTENTS

Title Page	i
Authorization Page	ii
Signature Page	iii
Acknowledgments	iv
Table of Contents	v
List of Figures	vi
List of Tables	vii
Abstract	viii
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Background	1
1.2 Objectives	2
1.3 Organization of the thesis	2
<b>Chapter 2 Literature Review</b>	<b>3</b>
2.1 The application of machine learning techniques for water quality control	3
2.1.1 Water quality control in drinking water treatment plants	3
2.1.1.1 Membrane fouling	3
2.1.1.2 Analysis of precursors of DBPs	3
2.1.1.3 Disinfection	3
2.1.1.4 Prediction of the source water contaminants	3
2.1.1.5 Coagulation	3
2.1.2 Water quality control in wastewater treatment plants	4
2.1.3 Water quality control in reclaimed water system	4
2.2 Recent advances in time series models for water quality forecasting	4
2.2.1 Machine learning models	4
2.2.2 Deep learning models	4
2.2.3 Comparison of the artificial intelligence model and traditional model in drinking water treatment	4
2.2.3.1 Traditional modeling methods	4
2.3 Different techniques for enhancing the performance of forecasting models	4
2.3.1 Data preprocessing	4
2.3.2 Feature engineering	4
<b>References</b>	<b>5</b>

## LIST OF FIGURES

# LIST OF TABLES

# Forecasting the Color and Ammonia Concentration in the Reclaimed Water using Deep Learning

by Ting Hsi LEE

Department of Civil and Environmental Engineering

The Hong Kong University of Science and Technology

## Abstract

Water scarcity is a global challenge. One of the promising ways to mitigate the water resource crisis is via wastewater reclamation. Chlorine is commonly used for reclaimed water disinfection and requires precise dosing to satisfy endorsed quality standards. Ammoniacal nitrogen ( $\text{NH}_3\text{N}$ ) and colour exist in the reclaimed water at concentrations between 0.23 – 5.44 mg N/L and 80 – 150 Hazen units, respectively, and can affect the chlorine demand. Forecasting the reclaimed water quality enables a feedback control system over the disinfection process by predicting the exact chlorine dose required which secures sufficient time to respond to sudden surges in color and ammonia levels. This study developed time-variant models based on machine learning to predict the  $\text{NH}_3\text{N}$  concentration and colour three hours into the future in the reclaimed water. The  $\text{NH}_3\text{N}$  data was collected by an online analyzer, and colour data was collected by a customized auto-sampling spectrophotometer, both are installed in the reclaimed water treatment plant in Hong Kong. Long Short-Term Memory (LSTM) was found to be the most effective architecture for training  $\text{NH}_3\text{N}$  and colour forecasting models. In the training processes, we applied data pre-processing methods and feature engineering, a technique to select or create relevant variables in raw data to enhance predictive model performance. From feature engineering, we discovered that the daily fluctuation in  $\text{NH}_3\text{N}$  and colour has correlations with the urban water consumption patterns. This finding further enhanced the  $\text{NH}_3\text{N}$  and colour forecasting model performance by 4.9% and 5.4% compared to baseline models. This research work offers novel methods and feature engineering pro-



cesses for  $\text{NH}_3\text{N}$  concentration and colour forecasting in reclaimed water for treatment optimization.

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

Forecasting models play an important roles in water quality control in drinking water treatment plants (DTPs) and wastewater treatment plants (WWTPs). The need of using forecasting models are becuase the unpredictable nature of water quality, and the treatment operations are subjected to the change of water quality to prodcue effluent complied the government regulation [1].

Forecasting models can also be called time series model becuase the data is consisted of the values and the time (need to be further revised). For the well-know time series models are for example, RNN, ... These are used to replace the theory-based models, for example Activated Sludge Model (ASM). The difference between these two models are, machine learning based models require to learn from historic data, while the thoery-based models only need to enter the basic operational parameters (e.g., influent flow, tempearture, and pH, etc).

Despite the promising usage and performance of machine learning models, the collection of the data became the most difficult tasks. Many small scale or old treatment plants do not have the capital or the available environment for the set-ups of the online sensors to collect data. Although these are the major issues, it's still possible to train a forecasting model with one input, which is also called a self-prediction model. Although the accuracy or stability compared to multi-input models, the forecasted results can be used at some cases. To increase the model performance, there are several ways. Paper included weather data, or perform data-preprocessing methods to improve the model performance.

These solutions (data preprocessing, feature engineering) are not well discussed in this field, also the potential of using univariate models are under estimated.

## **1.2 Objectives**

The specific objectives of this thesis work are:

- (1) To build baseline univariate forecasting models using machine learning and deep learning models.
- (2) To develop data preprocessing methods for enhancing model forecasting performance.
- (3) To extract features and hidden relations of water parameters in MBR effluent by analyzing the wastewater collected upstream of the WWTPs.
- (4) To develop methods for improving performance of forecasting models using the hidden features and relations of the water parameters.

## **1.3 Organization of the thesis**

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 The application of machine learning techniques for water quality control**

Machine learning is a subset of artificial intelligence, and deep learning is a subset of machine learning. In artificial intelligence can be used to solve four types of problems: classification, regression, dimensionality reduction and clustering.

##### **2.1.1 Water quality control in drinking water treatment plants**

###### **2.1.1.1 Membrane fouling**

Madfs

###### **2.1.1.2 Analysis of precursors of DBPs**

###### **2.1.1.3 Disinfection**

###### **2.1.1.4 Prediction of the source water contaminants**

###### **2.1.1.5 Coagulation**

Traditional modelling methods mainly use numerical simulations or physical formulas to model target prediction objects from a microscopic perspective . For example, the advantage of particle coagulation dynamics simulation is that it can explain the behaviour evolution mechanism of particles in the water treatment process in a very specific way because it is usually based on the collision mechanism with physical meaning and mathematical description

### **2.1.2 Water quality control in wastewater treatment plants**

### **2.1.3 Water quality control in reclaimed water system**

In this study the new control objectives for the reclaimed water system in Shek Wu Hui Effluent Polish Plant have been established: to monitor color and ammonia concentration in the MBR effluent and at the same time provide a predictive model to assist the disinfection control strategy for disinfecting the MBR effluent to meet the endorsed reclaimed water standard.

## **2.2 Recent advances in time series models for water quality forecasting**

### **2.2.1 Machine learning models**

### **2.2.2 Deep learning models**

### **2.2.3 Comparison of the artificial intelligence model and traditional model in drinking water treatment**

#### **2.2.3.1 Traditional modeling methods**

In traditional modeling methods, numerical simulations or physical formulas to model target prediction objects

## **2.3 Different techniques for enhancing the performance of forecasting models**

### **2.3.1 Data preprocessing**

### **2.3.2 Feature engineering**

## REFERENCES

- [1] J.C. Chen, N.B. Chang, and W.K. Shieh. Assessing wastewater reclamation potential by neural network model. *Engineering Applications of Artificial Intelligence*, 16(2):149–157, March 2003.