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

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 (NH_3N) 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 NH_3N concentration and colour three hours into the future in the reclaimed water. The NH_3N 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 NH_3N 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 NH_3N and colour has correlations with the urban water consumption patterns. This finding further enhanced the NH_3N 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 NH_3N concentration and colour forecasting in reclaimed water for treatment optimization.

CHAPTER 1

INTRODUCTION

1.1 Background

AI technologies have been successfully applied to different DWT processes, such as the prediction of the coagulant dosage, discrimination of the DBP formation potential, advanced control of membrane fouling, membrane preparation and optimization, and water quality prediction. Li et al. (2021)

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 Chen et al. (2003)

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 Sludege 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 Introduction to water quality control

2.1.1 Automated control system for water quality control

Programmable logic controller (PLC) is an industrial computer system designed for any process requiring a series of devices and equipment operates cohesively to achieve multiple purposes in manufacturing or treatment processes. The main components of PLC include a center process unit (CPU), input modules and output modules (I/O). CPU is responsible to process digital signals from input modules and send commands through output modules based on the control logics programmed on the PLC. For chemical dosing control in water treatment plants (WTPs), PLC system receives readings from turbidity and pH sensors and uses pumps to dose aluminum solution automatically (Andhare and Palkar, 2014). The PLC system with the capability of producing real-time output commands in response to the input sigals also makes it widely used in the wastewater treatment plants (WWTPs). For oxygen concentration control in the aeration tank, PLC system receives signals of dissovled oxygen (DO) detectors and transmits signals to open or close the electric butterfly valves to further alter the DO concentration (Zhu and Qiu, 2017). Although PLC systems are the most used system across industries for its easy programming and reliable control, PLC system is merely a device that can be programmed to control operative devices with on-off logic (i.e., a logic control with two states) and the capability of complex control is compromised. In reality, many WTPs or WWTPs have the need of precise control of the treatment processes. Being awared of the limitations of the PLC systems, a more advanced controller called proportional–interal–derivative (PID) controller for receiving analog signals was developed to obtain more sophisticated controls over the operative devices.

To react to rapidly-changing process conditions, a PID controller generates an output value based on continuous calculation of an error value e(t) as the difference between a desired setpoint (SP) and a measured process variable and applies a correction based on proportional, integral, and derivative terms. The use of the "P", "I", and "D" allows the system to quickly reach steady state with a feedback control system (i.e., the system output is returned to the system input which is included in the decision making process in PID controller). Generally speaking, a PID controller is a technology (i.e., a specialist algorithm) for contorlling a single device with more complex logics, while a PLC system is a physical system consists of different modules and capable of controlling dozens of devices only with two-state logic. In addition, A PID controller can be designed to operate on PLC device and provide a more precise control strategy to a designated device. In WWTPs, a single-variable feedback analog control loop in PID can be used to control the temperature in the activated sludge treatment by stablizing the system temperature in a shorter time (Bados and Morejon, 2020). The feedback control scheme is also applied in WTPs to adjust the addition of chlorine dosage (i.e., also known as the disinfection process, chlorination, or postchlorination) to reach the target concentration of free chlorine residual (FRC) (Wang and Xiang, 2019). Disinfection process is carried out in a chlorine contact tank which provides sufficient time for chlorine to disinfect pollutants. Since the chlorine added by the dosing device requires time to travel from the entry to the exit, the system output can only reflect the changes of water quality in a delayed time of 30 minutes (i.e., the designed time for water to travel in chlorine contact tank is usually 30 minutes or longer). In the case of chlorination, the lag of time makes feedback control difficult (Kobylinski et al., 2006) as the system is delayed in responding to any sudden surge of the pollutants when it can only receive output at the end of the disinfection process. PID controllers in WWTPs also encounter similar challenges as the increasing complexity of water quality and stricter regulations on the discharged water quality.

To tackle the difficulties encountered in process control system, many control strategies are proposed, such as feed forward-feedback control, linearized and optimal control, model-predictive control, and fuzzy control, etc (Demir and Woo, 2014). Among the algorithms used in control strategies, Artificial Intelligence (AI) modeling has gained the most attentions in recent years compared to modeling based on mathematical models or empirical formulas. In WTPs or WWTPs, to fully understand the physical, biological, and chemical interactions in the treatment plants is very difficult. The unpredictable behaviors during the water treatment can be the significant changes of influent flow rate, flucutations of water quality, the complexity of biological treatment process, and the large time delay exists between this control variable and the process input, etc. Therefore, AI modeling shows a great potential in dealing with the highly complex conditions in the treatment process Li et al. (2021). In the next sections, the applications of different AI modeling methods will be discussed.

2.1.2 Artificial Intelligence

Artificial intelligence (AI) can perform cognitive tasks with the development of computational solutions. The concepts of AI are usually confused, in fact, AI is a very broad term and any kind of algorithms or models which involved in decision-making with computation fall in the domain of AI. For example, fuzzy logic and optimization algorithm are formulated with human design and computer decision making process. There are another subset of AI called machine learning (ML), but the process of generating a ML model is different to generating a fuzzy logic model. ML uses learning algorithms to generate a model via learning from historical or large amount of data without being explicitly programmed. ML algorithms can be classified into three categories, which are Supervised, Unsupervised, and Reinforcement learning. In the training process of supervised learning, input variable (x) and output variable(Y) we will provided, and model will learn from the provided dataset to map the x to the Y. A trained supervised model can generate a prediction for the response to the new data (i.e., also called the unseen data). Unsupervised learning is when the dataset is not labelled, the model can learn to infer patterns in the dataset without reference to the known outputs. This type of algorithm can find similarities and differences in the data. In reinforcement learning, models are designed to constantly interact with the environment in a try-and-error way and recieved rewards and punishments based on the purpose of the tasks. Generating a optimal action to achieve lowest penalties is the main function of a reinforcement learning model. In process control, supervised learning are frequently used in many senarios.

Regression is a supervised machine learning technique used to predict continuous values. A regression model can estimate the relationship between the input variables in the system and the output target from a given dataset, and then use the nonlinear relationship to map the unseen input data to a predicted output data. This type of application is sutiable for water quality prediction (Librantz et al., 2018), and sensor fault detection (Cecconi and Rosso, 2021), etc.

2.1.3 AI modelings for water quality control

In machine learning, there are some popular models used by the researchers. Supporting vector machine (SVM), random forest (RF), and Artificial Neural Networks (ANN) are used to train predictive model. Librantz et al. (2018) trained a RF model to predict the free residual chlorine concentration (FRC) in a WTP, and Xu et al. (2021) built a RF-based model to predict total nitrogen concentration in water bodies. Guo et al. (2015) compared the reliability and accuracy of an ANN model and a SVM model in predicting 1-day interval T-N concentration in a WWTP, and the results showed that RF model has higher accuracy while ANN model is more reliable for assisting decision-making process. Fuzzy logic approach

RF is commonly used for building prediction models, for instance, Xu et al. (2021) uses RF-based model to predict total nitrogen concentrations in water bodies, and Bulacan State University, City of Malolos, Bulacan, Philippines. He is also with AMA University Quezon City, Philippines et al. (2020) used the RF model to predict the water pollution level. ANN can also perform predicting tasks, such as predicting the free residual chlorine concentration (FRC) in a WTP (Librantz et al., 2018).

As the the computing power doubled every 18 months according to Moore's law. A subset of ML, Deep Learning (DL) becomes more accessible for sovling everyday issues. In simplicity, DL models can be defined as neural networks with more than two hidden layers (i.e., the model complexity increased and required more computing power to calculate). In DL, there are various types of architectures designed based on the type of problems. For image processing, Convolutional Neural Network (CNN) is designed to extract important features from the image vectors. Another popular DL architecture is Recurrent Neural Network (RNN), which is powerful in solving time series-related applications and Natural Language Processing (NLP) tasks (Li et al., 2018). Although each architecture has their strengh in tackling different types of problems, both architectures can be used for a single task Li et al. (2022) built a regression CNN-RNN model for rainfall-runoff prediction. DL can be extremely powerful when multiple architectures ared fused into a single model to perform a specific task, which cannot be realized by machine leraning models. That being said, DL can achieve higher model performance in terms of the prediction accuracy compared to ML.

2.2 Water quality control with the use of machine learning modeling

2.2.1 Drinking water treatment plants

Disinfection is the last step of water treatment processes in drinking water treatment plants (DWTPs) to generate safe potable water. In this step, one or more chemical disinfectants like chlorine, chloramine, or chlorine dioxide are added into the water to inactivate any remaining pathogenic microorganisms. The residual disinfectant concentration in disinfected water must contains low levels of the chemical disinfectant to stop nuisance growths in the water distribution pipes, storage facilites and conduits. Nowadays, the widely used disinfectant in disinfection process is chlorine as gas or hypochlorite (i.e., in form of liquid solution), and the treatment process is known as "chlorination". According to World Health Organization's Guidelines for Drinking-water Quality (WHO Guidelines), the maximum allowable value for free chlorine residual in drinking water is 5 mg/L, and the minimum recommended valus is 0.2 mg/L.

Current analysis proposes a multivariable control for post-chlorination dosage system in a WTP using artificial neural networks applied to the disinfection process to reduce free residual chlorine variations of treated water in the water tank and, consequently, in the main water distribution (Librantz et al., 2018).

Despite the benefit brought by dosing chlorine to the water, negative impacts also come along. In the real world case, the influent water quality and the efficiency of the drinking water treatment processes are not always stable, and the invariability of the treated water quality becomes an big issue for disinfection. For instance, chlorine dose can be excessive dosed when the treated water contains less pollutants (e.g., non-organic matters and ammonia nitrogen). Although the quality of disinfected water fulfill the regulation standard, it increases the costs and can potentially generate undesired disinfection by-products (e.g., trihalomethanes, which are carcinogenic to human) due to the chemcial reaction between pollutants and overly dosed chlorine. On the flip side, insufficient dosing of chlorine cuases the concentration of residual chlorine lower than the legal regulation. To prevent both senarios occur, a water quality control strategy is required to produce drinking water with satisfactory quality.

Up untill present, there are several ways to perform disinfected water quality. In the earliest time, feed-back.... PI... feed-foward...

2.2.1.1 Membrane fouling

Madfs

2.2.1.2 Analysis of precursors of DBPs

2.2.1.3 Disinfection

2.2.1.4 Prediction of the source water contaminants

2.2.1.5 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 ex plain 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

2.2.2 Wastewater treatment plants

2.2.3 Reclaimed water system and water body

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.

2.3 Tools and techniques for enchancing the performance of machine learning modeling

2.3.1 Programming languages

2.3.2 Data preprocessing

2.3.3 Feature engineering

CHAPTER 3

METHODS AND MATERIALS

3.1 Wastewater treatment plant description

3.1.1 Treatment processes

3.1.2 Reclaimed water standard

3.2 Data collection and preparation

3.2.1 Ammonia data monitoring and collection

3.2.2 Color data monitoring and collection

3.2.3 Data cleaning and pre-processing

3.2.3.1 Data smoothing with Savitzky-Golay filter

3.2.3.2 Exponentially Weighted Moving Average

3.2.3.3 Outlier Removal

3.2.4 Data transformation

Split of Train/valid/test dataset

3.3 Architecture design of the selected baseline models

3.3.1 LSTM

3.3.2 RNN

3.3.3 GRU

3.4 Implementation of regularization

3.4.1 Scheduler

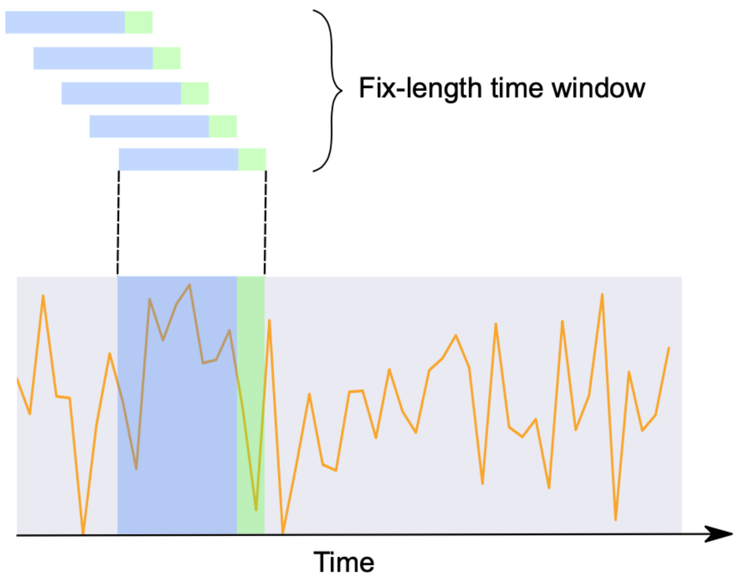

Figure 5. Using sliding window to construct supervised learning examples from time series data.

Figure 4.1: The network structure for the actor-evaluation estimation. It is a combination of convolutional networks for feature extraction and fullyconnected layers for policy learning. They have been separately proven to be effective in our previous works.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Baseline performance

4.2 Pre-processing

4.3 Feature engineering

4.4 Architecture desing

CHAPTER 5

CONCLUSION

Bibliography

Sunil L Andhare and Prasad J Palkar. SCADA a tool to increase efficiency of water treatment plant. *Asian Journal of Engineering and Technology Innovation*, page 8, 2014.

Jhon Stalin Figueroa Bados and Iralmy Yipsy Platero Morejon. Design of a PID Control System for a Wastewater Treatment Plant. In *2020 3rd International Conference on Robotics, Control and Automation Engineering (RCAE)*, pages 31–35, Chongqing, China, November 2020. IEEE. ISBN 978-1-72818-638-2. doi: 10.1109/RCAE51546.2020.9294199.

Bulacan State University, City of Malolos, Bulacan, Philippines. He is also with AMA University Quezon City, Philippines, Jayson M. Victoriano, Luisito L. Lacatan, and Albert A. Vinluan. Predicting River Pollution Using Random Forest Decision Tree with GIS Model: A Case Study of MMORS, Philippines. *International Journal of Environmental Science and Development*, 11(1):36–42, 2020. ISSN 20100264. doi: 10.18178/ijesd.2020.11.1.1222.

Francesca Cecconi and Diego Rosso. Soft Sensing for On-Line Fault Detection of Ammonium Sensors in Water Resource Recovery Facilities. *Environmental Science: Water Research and Technology*, 2021. doi: 10.1021/acs.est.0c06111.

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. ISSN 09521976. doi: 10.1016/S0952-1976(03)00056-3.

Feridun Demir and Wilbur W. Woo. Feedback control over the chlorine disinfection process at a wastewater treatment plant using a Smith predictor, a method of characteristics and odometric transformation. *Journal of Environmental Chemical Engineering*, 2(2):1088–1097, June 2014. ISSN 22133437. doi: 10.1016/j.jece.2014.04.006.

Hong Guo, Kwanho Jeong, Jiyeon Lim, Jeongwon Jo, Young Mo Kim, Jong pyo Park, Joon Ha Kim, and Kyung Hwa Cho. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *Journal of Environmental Sciences (China)*, 32:90–101, 2015. doi: 10.1016/j.jes.2015.01.007.

Edmund A. Kobylinski, Gary L. Hunter, and Andrew R. Shaw. On Line Control Strategies for Disinfection Systems: Success and Failure. *Proceedings of the Water Environment Federation*, 2006(5):6371–6394, January 2006. ISSN 1938-6478. doi: 10.2175/193864706783761716.

Lei Li, Shuming Rong, Rui Wang, and Shuili Yu. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. *Chemical Engineering Journal*, 405:126673, February 2021. ISSN 13858947. doi: 10.1016/j.cej.2020.126673.

Peifeng Li, Jin Zhang, and Peter Krebs. Prediction of Flow Based on a CNN-LSTM Combined Deep Learning Approach. *Water*, 14(6):993, March 2022. ISSN 2073-4441. doi: 10.3390/w14060993.

Zhe Li, Caiwen Ding, Siyue Wang, Wujie Wen, Youwei Zhuo, Chang Liu, Qinru Qiu, Wenyao Xu, Xue Lin, Xuehai Qian, and Yanzhi Wang. E-RNN: Design Optimization for Efficient Recurrent Neural Networks in FPGAs, December 2018.

André Felipe Librantz, Fábio Cosme Rodrigues dos Santos, and Cleber Gustavo Dias. Artificial neural networks to control chlorine dosing in a water treatment plant. *Acta Scientiarum. Technology*, 40(1):37275, September 2018. ISSN 1807-8664, 1806-2563. doi: 10.4025/actascitechnol.v40i1.37275.

Dongsheng Wang and Hao Xiang. Composite Control of Post-Chlorine Dosage During Drinking Water Treatment. *IEEE Access*, 7:27893–27898, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2901059.

Jianlong Xu, Zhuo Xu, Jianjun Kuang, Che Lin, Lianghong Xiao, Xingshan Huang, and Yufeng Zhang. An Alternative to Laboratory Testing: Random Forest-Based Water Quality Prediction Framework for Inland and Nearshore Water Bodies. *Water*, 13(22): 3262, November 2021. ISSN 2073-4441. doi: 10.3390/w13223262.

Huijun Zhu and Xinglei Qiu. The Application of PLC in Sewage Treatment. *Journal of Water Resource and Protection*, 09(07):841–850, 2017. ISSN 1945-3094, 1945-3108. doi: 10.4236/jwarp.2017.97056.