APPLICATION OF ARTIFICIAL INTELLIGENCE TO WASTEWATER TREATMENT PLANT OPERATION

by

Praewa Wongburi

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Civil and Environmental Engineering)

at the

UNIVERSITY OF WISCONSIN-MADISON

2021

Date of final oral examination: 07/28/2021

The dissertation is approved by the following members of the Final Oral Committee:

Jae Park. Faculty, Professor, Civil and Environmental Engineering

Greg Harrington. Faculty, Professor, Civil and Environmental Engineering

Chin Wu. Faculty, Professor, Civil and Environmental Engineering

Andrea Hicks. Faculty, Assistant Professor, Civil and Environmental Engineering

Miaoyan Wang. Faculty, Assistant Professor, Statistics

ABSTRACT

In a wastewater treatment plant (WWTP), big data is collected from sensors installed in various unit processes, but limited data is used for operation and regulatory permit requirements. With the advancement in information technology, the data size in wastewater treatment systems has increased significantly. However, WWTPs have not used big data systematically to aid the operation and detect potential operational issues due to the lack of specialized analytical tools.

The objectives of the study were to: (1) develop analytics methods suitable for the management of big data generated in WWTPs, (2) interpret analytics results for extracting meaningful information, (3) implement a recurrent neural network (RNN) and Long Short-Term Memory (LSTM) to predict effluent water quality parameters and Sludge Volume Index (SVI), (4) apply an Explainable Artificial Intelligence (AI) algorithm to determine causes of predicted values, and (5) propose a real-time control using a predictive model to monitor and optimize the operation of WWTPs.

The predictive AI models in WWTPs were developed by applying big data analytics, statistical analysis, and RNN algorithms with an Explainable AI algorithm. The models successfully and accurately predicted the effluent water quality data and a key operational parameter, SVI. Furthermore, the Explainable AI algorithm provided insight into which influent parameters affected higher predicted effluent concentrations and SVI on a specific day, allowing operators to take corrective actions.

From a WWTP's operational data analysis, the RNN model successfully predicted the effluent concentrations of BOD_5, total nitrogen (TN) and total phosphorus (TP), and SVI. Furthermore, the

Explainable AI analysis found that higher influent NH_3N values lead to higher effluent BOD_5, and higher influent total suspended solids (TSS) and TP values resulted in lower effluent BOD_5, implying the importance of controlling dissolved oxygen (DO) in aeration basins. Since aeration is one of the major energy consumption sources in WWTPs, real-time prediction of the effluent water quality using the self-learning AI system developed in this study can be adopted to lower the energy cost significantly while improving effluent water quality. WWTPs must develop control methods based on the RNN prediction and Explainable AI analysis due to different operational conditions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Background and Research Motivation

Due to the demands for lower operation and maintenance cost, energy cost, stringent compliance requirements of water quality parameters, and lower greenhouse gas emission, WWTPs (WWTPs) must be in a smart management mode. Due to the extensive use of water quality sensors, big data is generated every day or even every second. In the wastewater treatment sector and many other fields such as finance, marketing, stocks, health care, and so on, big data plays an essential role. Data generation has been tremendously increasing since 2010, and 90% of the world's data has been created in the past two years (Figure 1.1).

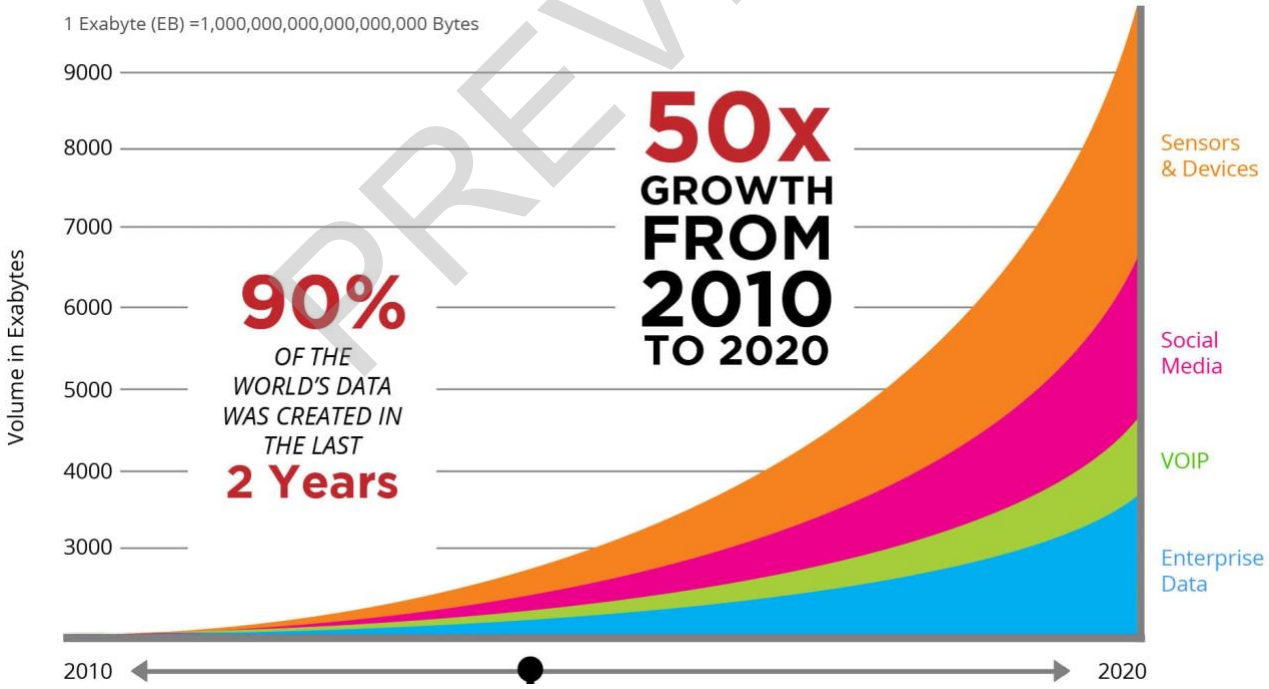

Figure 1.1 The growth of data from 2010 to 2020. (Source: Roser et al., 2015)

Big data is currently most preferably exploited in organizations, companies, and businesses. However, a massive amount of data results in the difficulties of storing, monitoring, analyzing, and visualizing data for further processing. Therefore, data is often stored and then underutilized.

A good understanding of the data dynamics in WWTPs is vital for reliable monitoring and control activities. However, the dynamical behavior of the data is usually complicated and uncertain due to nonlinearity, variations from the environmental conditions, strong interactions between the process variables involved, and changes in the flow rate and concentration of the composition of the influent (Harrou et al., 2018). Finding insight from historical and real-time data can directly improve traditional operational systems in a better direction.

A series of wastewater treatment processes remove pollutants from wastewater to be safely reused or discharged into natural water resources. Treated water can be recycled and redistributed for agricultural, industrial, and other purposes or safely released back into the natural resources without causing any adverse effects (Grant et al., 2012). The effluent from a WWTP must meet the National Pollutant Discharge Elimination System (NPEDES) permit to protect the environment and public health (Siegrist, 2017). Lack of access to safe water leads to a risk factor for infectious diseases such as cholera, diarrhea, and dysentery. According to the Global Burden of Disease study, 1.2 people died prematurely in 2017 due to unsafe water (Figure1.2). This number was three times the number of homicides in 2017 and approximately equal to the amount that died in road accidents globally. Besides, only 71% of the world population has access to safe drinking water, which means that 29% of the world population does not have access to safe water. It equates to 2.1 billion people globally (Figure 1.3).

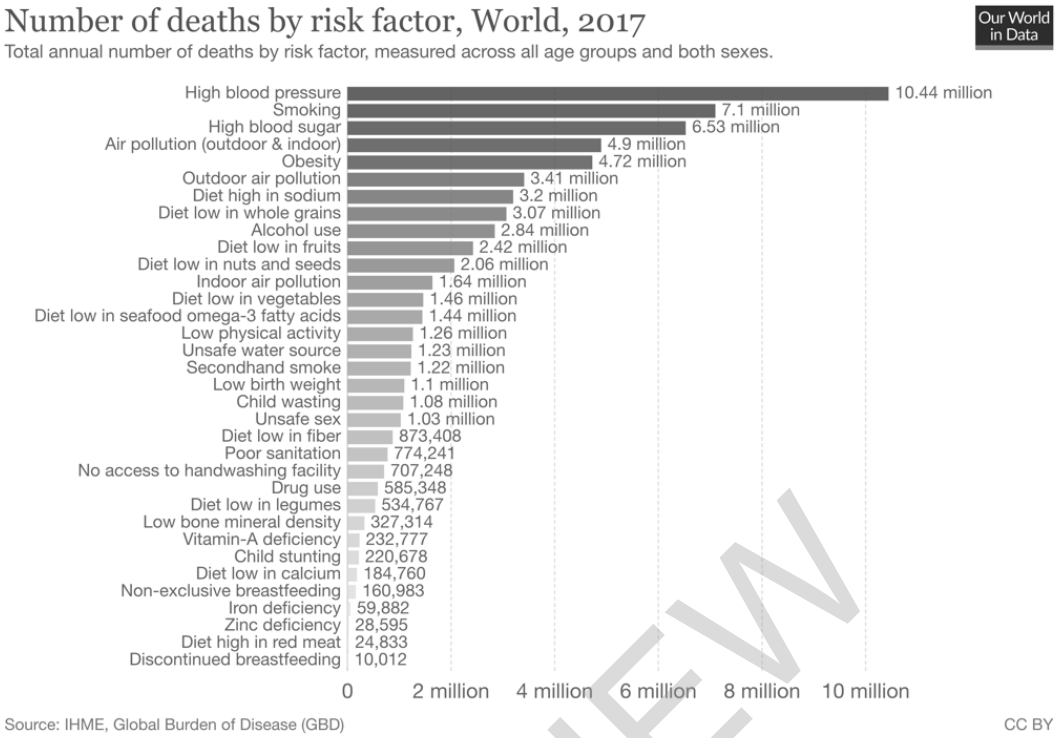

Figure 1.2 Number of deaths by risk factor. (Source: Ritchie & Roser, 2018)

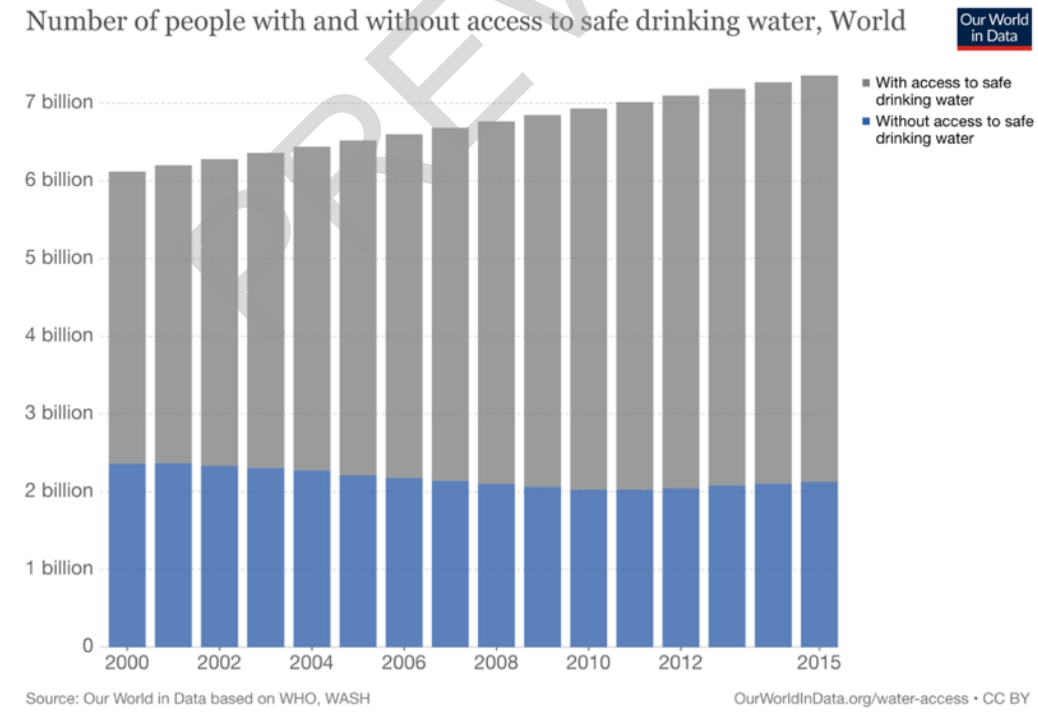

Figure 1.3 Number of people with and without access to safe drinking water.

(Source: Ritchie & Roser, 2018)

Even though there is an increasing amount of historical data in WWTPs, most information in the data will remain unexploited. According to the operator's point of view, the main reason for this is the high dimensionality of the data, where traditional analysis tools cannot be used (Durrenmatt, 2011). As a result, a large amount of data is lost. Various methods and tools such as big data analytics, statistical analysis, deep learning, and artificial intelligence (AI) applications extract information hidden in the data. Furthermore, this would assist the operator in further optimization of the WWTP, improve the effluent quality, reduce the human errors in operating processes, foster the operator's knowledge of the plant processes, and provide other supporting information. Finally, the main goal of a WWTP is to ensure the efficiency of wastewater treatment operation, which is tremendously essential for community health and the environment. Advanced tools and techniques to improve wastewater treatment operations were introduced in this study for analyzing, modeling, optimizing, and forecasting wastewater treatment quality.

1.2 Problem Statement

A literature review on big data management in wastewater treatment operation indicates four principle wastewater management problems that need to be solved in this study: (1) lack of research on big data management from WWTP, (2) lack of comprehensive statistical analysis of the data, and (3) lack of real-time predictive models to forecast effluent quality.

1.2.1 Lack of Research on Big Data Management from WWTPs

After researching publications in big data management in wastewater treatment facilities from various resources such as Google Scholar®, Scopus®, and Web of Science®, there have been limited studies (Ghernaout et al., 2018; Romero et al., 2017). A large amount of data is generated from various wastewater treatment operations every day. Big data should be exploited to enhance the operating systems. Unfortunately, due to the lack of specialized tools, operators and engineers

cannot extract meaningful and valuable information from the massive amount of high-dimensional data.

1.2.2 Lack of Comprehensive Statistical Analysis of Wastewater Treatment Data

Several studies have analyzed energy flow and influent in wastewater treatment facilities to monitor, assess, and model the WWTPs (Martin & Vanrolleghem, 2014). In addition, many publications have illustrated the usefulness of statistical analysis models for WWTP optimization (Cheng et al., 2019; Newhart et al., 2019); operation (Garbowski et al., 2018); analysis (Pantsar-Kallio et al., 1999; Taheriyoun & Moradinejad, 2015; Zhang et al., 2019) and control (Harrou et al., 2018; Maiza et al., 2013). However, few studies have been performed on finding patterns, determining the relationship of each parameter, and selecting meaningful information from big data in wastewater treatment operations with the use of advanced analytics. As a result, big data is underutilized.

1.2.3 Lack of Predictive Models to Forecast Effluent Quality

The water quality predictions in WWTPs were attempted using advanced machine learning tools and techniques. Nonetheless, without the pretreatment of big data, the forecast may not be accurate. Also, previous studies focused on traditional deterministic modeling methodology (Boyd et al., 2019; Huang et al., 2016; Khademikia et al., 2016; Pisa et al., 2019). The novel deep learning method, an RNN, is rarely applied. The conventional approaches might be accurate in predicting if the system is fixed and all the parameters are determined. Furthermore, it is time-consuming, intrusive, and limited by small data analysis.

1.2.4 No Logistics for a Real-Time Model for WWTPs

Real-time data reflects the status of an operational system. The common characteristic of the data is the strict time constraint (Wu et al., 2006). Real-time information is associated with a timestamp and life cycle, and they are only valid for the responding sampling time. Deep learning algorithms with real-time modeling would be a reliable tool for early warning of potential operational upsets and subsequent effluent quality permit violations. There is no study on the development of logistics for real-time modeling using RNNs to help operators for monitoring, detecting fault operation systems in WWTPs.

1.3 Research Objectives

The main objective of this paper is to enhance the operation and performance of WWTPs through big data management and model prediction with AI applications. The study has the following four specific objectives:

(1) To develop analytics for big data from a WWTP;

(2) To interpret analytics results for extracting meaningful information;

(3) To develop deep learning models for forecasting effluent quality from historical wastewater treatment data, which include train and evaluate the models to acquire the most effective algorithms; and

(4) To establish logistics for a real-time self-learning AI system for monitoring and detecting problems during WWTP operation.

1.4 Research Scope and Methodology

Four principal tasks in this study are discussed below.

Task 1: Big data analytics frameworks in wastewater treatment operation.

This task involves collecting data, understanding the processes in wastewater treatment, visualizing data, selecting meaningful information, and developing big data analytics procedures.

Task 2: Statistical analysis techniques to obtain a pattern and meaningful information.

This task applies various statistical methods such as descriptive statistical analysis, correlation coefficient, box plot, normal distribution, and hypothesis testing to find a relationship between parameters, a pattern of data, and insight of information.

Task 3: DNNs model development for prediction.

This task is to develop a predictive model by implementing advanced modeling techniques, Recurrent Neural Networks (RNNs). The methods include preparing the data, selecting the train and test dataset, build a simple RNNs model and the Long Short-Term Memory (LSTM) model with a different number of hidden layers, train the models to predict an output result, and compare and evaluate the models. The parameters to be predicted will be Total Phosphorus (TP), Total Suspended Solids (TSS), and NH_3/Total Nitrogen (TN) in addition to BOD_5 (Biochemical Oxygen Demand) and lastly, Sludge Volume Index (SVI). Control of dissolved oxygen (DO), sludge and energy management logistics may be attempted from the big data.

Task 4: Propose logistics for a real-time prediction model to detect fault errors in wastewater treatment operations.

Unless a WWTP allows access to real-time big data and incorporates the AI model into their Supervisory Control and Data Acquisition (SCADA) system, it is impossible to implement a real-time prediction model to the existing system. Therefore, this task proposes logistics for

real-time model development in WWTPs to help a WWTP's operator for failure detections beforehand.

1.5 Organization of the Dissertation Proposal

The thesis proposal is organized as follows:

Chapter 2 presents a comprehensive review of previous studies on big data in the wastewater treatment sector, statistical analysis in wastewater treatment systems, and the development of prediction models in WWTPs.

Chapter 3 presents big data management with statistical analysis. Big data analytics processes include data collection, data understanding, data preparation, data mining, evaluation, and deployment. Data was collected from the Nine Springs WWTP in Madison Metropolitan Sewerage District (MMSD), Wisconsin. This data is comprehensively studied through data preprocessing techniques contain data cleaning, data integration, data transformation, and data reduction. Finally, statistical analysis techniques extracted meaningful patterns and information to obtain the appropriate dataset.

Chapter 4 presents the development of model predictions using RNNs. After the big data is extracted to a manageable size and the statistical preprocessing technique is implemented to select meaningful information, the prediction efficiency of wastewater effluent quality will improve significantly. The results of traditional RNN models and RNN-LSTM models were compared and evaluated to choose an optimization algorithm for implementation.

Chapter 5 proposes the subsequent work plan to develop an RNN model using different parameters, develop logistics for monitoring, detecting, and proactive maintenance, assist better decision making, and optimize wastewater treatment facilities.