DEVELOPMENT OF DATA-DRIVEN MODELS FOR MEMBRANE

FOULING PREDICTION AT WASTEWATER TREATMENT PLANTS

By DAVID KOVACS, B.ENG.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the

Requirements for the Degree Master of Applied Science

Master of Applied Science (2022)   McMaster University

(Civil Engineering)   Hamilton, Ontario

TITLE: Development of Data-Driven Models for Membrane Fouling Prediction at Wastewater Treatment Plants

AUTHOR: David Kovacs

                                          B.Eng. (McMaster University)

CO-SUPERVISORS: Dr. Zhong (Zoe) Li, P.Eng.

                                          Dr. Brian Baetz, P.Eng.

NUMBER OF PAGES: xii, 61

ABSTRACT

Membrane bioreactors (MBRs) have proven to be an extremely effective wastewater treatment process combining ultrafiltration with biological processes to produce high-quality effluent. However, one of the major drawbacks to this technology is membrane fouling – an inevitable process that reduces permeate production and increases operating costs. The prediction of membrane fouling in MBRs is important because it can provide decision support to wastewater treatment plant (WWTP) operators. Currently, mechanistic models are often used to estimate transmembrane pressure (TMP), which is an indicator of membrane fouling, but their performance is not always satisfactory. In this research, existing mechanistic and data-driven models used for membrane fouling are investigated. Data-driven machine learning techniques consisting of random forest (RF), artificial neural network (ANN), and long-short term memory network (LSTM) are used to build models to predict transmembrane pressure (TMP) at various stages of the MBR production cycle. The models are built with 4 years of high-resolution data from a confidential full-scale municipal WWTP. The model performances are examined using statistical measures such as coefficient of determination (R^2), root mean squared error, mean absolute percentage error, and mean squared error. The results show that all models provide reliable predictions while the RF models have the best predictive accuracy when compared to the ANN and LSTM models. The corresponding R^2 values for RF when predicting before, during, and after back pulse TMP are 0.996, 0.927, and 0.996, respectively. Model uncertainty (including hyperparameter and algorithm uncertainty) is quantified to determine the impact of hyperparameter tuning and the variance of extreme predictions caused by algorithm choice. The ANN models are most impacted by hyperparameter tuning and have the highest variability when predicting extreme values within

each model's respective hyperparameter range. The proposed models can be useful tools in providing decision support to WWTP operators employing fouling mitigation strategies, which can potentially lead to better operation of WWTPs and reduced costs.

ACKNOWLEDGMENTS

I would like to express my sincerest thanks and gratitude to my supervisor, Dr. Zoe Li. Her advice throughout the last two years has been invaluable in both an academic and personal sense. Without her patience and guidance throughout the entire academic process including selecting applicable courses, the model development process, and countless other areas, this work would not have been possible. She has been there every step of the way, through questions big and small; I am grateful for that. I also appreciate the support and feedback from my fellow research group members, especially Pengxiao Zhou. Thank you, Dr. Brian Baetz, for co-supervising my thesis, and thank you Drs. Georgios Balomenos and Younggy Kim for serving on my thesis committee.

I would also like to thank those at SUEZ whose advice, expertise, and input made all the difference in the usefulness of this work. This includes Howard, Youngseck, Sylvain, and the rest of the team – I appreciate all the *InSightful* questions and feedback. I thank the Faculty of Engineering at McMaster whose graduate courses enabled me to apply my knowledge to make an impact throughout my academic research.

Finally, I would like to thank my family and friends for the love and support. Without the love, guidance, and support from my parents, I would not be here today. Special thanks to my girlfriend, Nola, whose unwavering support has allowed me to succeed and grow throughout the years.

PUBLICATION AND CO-AUTHORSHIP

This M.A.Sc thesis is organized in accordance with the guidelines provided by the School of Graduate Studies at McMaster University. The work done in this thesis will be submitted for journal publication, and will be named as follows:

1. David J. Kovacs, Zhong Li, Brian W. Baetz, Youngseck Hong, Huihuang Ding, Sylvain Donnaz, Xiaokun Zhao, Pengxiao Zhao, "Membrane Fouling Prediction and Uncertainty Analysis Using Machine Learning: A Wastewater Treatment Plant Case Study".

David Kovacs, in consultation with Dr. Zoe Li, processed the data, designed the RF and ANN models, performed computations and analysis, designed the figures and tables, and drafted the manuscript. The manuscript was edited and reviewed by both Dr. Zoe Li and Dr. Brian Baetz. Valuable expert insight was provided by Youngseck Hong, Huihuang Ding, and Sylvain Donnaz. Xiaokun Zhao designed the LSTM model. Pengxiao Zhao provided support throughout the process. David Kovacs has made a significant original contribution to this paper and is the first author.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABP	After back pulse
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASM	Active Sludge Model
ASP	Activated Sludge Process
BBP	Before back pulse
BFGS	Broyden-Fletcher-Goldfarb-Shanno (algorithm)
COD	Chemical oxygen demand
CP	Columnwise Pairwise
DBP	During back pulse
DT	Decision Tree
GA	Genetic Algorithm
LHS	Latin Hypercube Sampling
LMH	Litres per square metre hour
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
MBR	Membrane bioreactor
MC	Maintenance clean
MCS	Monte Carlo Simulation
ML	Machine learning

MLP	Multi-Layer Perceptron
MLSS	Mixed liquor suspended solids
MSE	Mean Squared Error
MT	Model Tree
OOB	Out-of-bag
PDP	Partial dependence plot
PI	Prediction interval
R^2	Coefficient of Determination
RBF	Radial basis function
RC	Recovery clean
RF	Random Forest
RIS	Resistance-in-series
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SGD	Stochastic gradient descent
SLP	Single-layer Perceptron
SRT	Sludge retention time
SVM	Support Vector Machine
SMBR	Submerged membrane bioreactor
TMP	Transmembrane pressure
WWTP	Wastewater treatment plant

Chapter 1 – Introduction

1.1   Background

Often taken for granted, water is one of the most important natural resources and is the foundation for life to exist. Water covers approximately 71% of the Earth's surface; however, only about 2.5% is freshwater, and the amount of freshwater easily accessed by humans is less than 1% of all water (Stephens et al., 2020). Although restored through the water cycle, the increase in global population has put an increasing strain on water supply and many areas have water shortages. Clean water and sanitation is recognized as one of the 17 Sustainable Development Goals (SDGs) by the United Nations (UN) (United Nations, 2020a). Despite this recognition, billions of people still lack access to safe drinking water. Climate change will only make the matter worse and is "projected to increase the number of water-stressed regions and exacerbate shortages in already water-stressed regions" (United Nations, 2020b). Therefore, it is essential to protect and repurpose the water that we use. Wastewater reuse dates to 3200 BC when civilizations used it for irrigation and aquaculture (Angelakis et al., 2018). As technology and knowledge evolved, wastewater reuse turned into wastewater treatment, where wastewater was effectively treated in wastewater treatment plants (WWTPs) with modern techniques such as the activated sludge process (ASP) that uses microorganisms for eliminating organic pollutants (Chang et al., 2002). Membrane bioreactors (MBRs) integrate ASP WWTP technology and are becoming one of the most promising water treatment technologies in the 21st century. MBRs are popular due to numerous benefits including extremely high-quality effluent, small footprint, and low operational and capital costs (Hamedi et al., 2019). MBRs combine both biological treatment and filtration into one process to treat wastewater. However, a major drawback to this technology is membrane fouling, an impediment to the membrane process that leads to higher operating costs and reduced

effluent quality (Guo et al., 2012; Iorhemen et al., 2016; Meng et al., 2009). This process is exhibited when an MBR operates at a constant flux and has a rise in transmembrane pressure (TMP), or a decrease in flux when operated at a constant TMP. To allow MBRs to be more widely implemented, accurate models that provide decision support for membrane fouling prediction are required.

The modeling of membrane fouling in MBR applications has been investigated since the mid-1980s. This investigation includes mechanistic membrane fouling models, with data-driven membrane fouling models becoming more popular in the early 2000s (Shi et al., 2021; Stephenson et al., 2000). Specifically, mechanistic membrane fouling models incorporate three main mechanisms associated to membrane fouling: the deposition of solids on the membrane surface, adsorption of solids onto pore surfaces, or complete pore-blocking (Stephenson et al., 2000). The problem is that membrane fouling is extremely complex and is influenced by many factors and non-linear relationships between biological, chemical, and physical variables. Additionally, each MBR system has a different process for which variables and calibration procedures will change (Dalmau et al., 2015; Drews, 2010). These models face challenges relating to complexities and assumptions about underlying variables and behaviour; they also require extensive calibration of model parameters.

In comparison, data-driven models seek to capture the underlying processes between input and output variables in datasets. The application of data-driven modeling techniques come with some risk because the generated models are not based directly on mechanical processes. To combat this, careful input variable selection and expert consultation should be employed. By nature, data-driven models give a single-point prediction for a single data observation. It is important to quantify the uncertainties associated to these models to provide more robust decision support that

includes reliable prediction intervals (PIs) (Gönder et al., 2011; Nourani et al., 2022). Additionally, many data-driven models found in literature generally use data from MBR pilot-scale plants and there is a lack of long-term data from full-scale municipal WWTPs. With the improvement of technology and data collection techniques, datasets pertaining to WWTPs (including operational data and influent/effluent characteristics) are becoming more readily available and of higher quality. This provides the opportunity to utilize these datasets for the prediction of variables directly relating to membrane fouling, such as transmembrane pressure (TMP).

1.2   Objective and Scope

The main objective and scope of this thesis is to create data-driven models based on machine learning algorithms, including RF, ANN, and LSTM, to predict TMP at the various stages of an MBR production cycle. These proposed models will be applied to a confidential full-scale WWTP to test prediction accuracy utilizing performance metrics. This study entails the following primary objectives: (1) develop data-driven models to predict TMP at various stages of the MBR production cycle using the three algorithms; (2) evaluate the performance of the developed models and compare model accuracies; and (3) conduct an uncertainty analysis on the developed models to quantify hyperparameter and algorithm uncertainties. The work proposed in this thesis will provide useful tools for WWTP operators and provide insight into membrane fouling control.

1.3   Thesis Outline

Chapter 2 explores the literature of membrane bioreactor technology and associated membrane fouling phenomenon. Further, existing mechanistic and data-driven models used for membrane fouling will be reviewed and discussed; this includes advantages and disadvantages of both types of models and how they apply.

Chapter 3 explores the case study application. This includes general MBR information, information pertaining to the full-scale municipal WWTP in this case study, and how operational data from the plant is gathered and collected.

Chapter 4 explores the methodology, implementation of, and uncertainties associated to the data-driven techniques that are used to model membrane fouling in this MBR case study. These techniques include random forest, artificial neural network, and long short-term memory network; how they operate will be discussed in detail. The implementation of these techniques corresponds to a careful input variable selection process and associated data preprocessing. Thereafter, the models will be compared using popular performance metrics to determine which perform the best. The uncertainty analysis procedure for both hyperparameter and model structure uncertainty will be discussed in detail.

Chapter 5 explores model performance pertaining to the implemented data-driven techniques. Further, it explores the uncertainty analysis results for hyperparameter uncertainty and model structure uncertainty (characterized by extreme point prediction).

Chapter 6 summarizes the research conducted in this work and provides recommendations for future work.

Chapter 2 – Literature Review

2.1   MBR and Membrane Fouling

MBR technology combines membrane filtration with the ASP. In an MBR, the separation of solids from wastewater is conducted by membrane filtration through microfiltration and/or ultrafiltration (Hamedi et al., 2019; Hermanowicz et al., 2006) as opposed to gravity-driven separation in a secondary clarifier. There are two different configurations of an MBR, an internal/submerged MBR or an external/side stream MBR. The submerged MBR utilizes membranes immersed in the bioreactor to pull wastewater through membrane pores with the assistance of a vacuum pump (Melin et al., 2006). These membranes will either be flat-sheet or hollow-fibre, both of which have their advantages and disadvantages (Hai & Yamamoto, 2011). For side stream MBRs, the membrane module is separate from the bioreactor and the mixed liquor wastewater is pumped through this module, employing cross filtration. In municipal wastewater applications, submerged MBRs are often used. Some of the main advantages of MBRs include high quality effluent, control over sludge retention time (SRT) and hydraulic retention time (HRT), a smaller capital footprint, and the ability to operate at low dissolved oxygen concentrations (Hamedi et al., 2019). However, a major drawback to MBR technology is membrane fouling.

Within the MBR, membrane fouling is shown through the drop in permeate flux when the MBR is operated at a constant TMP, or an increase in TMP when the MBR is operated at a constant flux, thereby reducing the amount of permeate being produced. This increase in TMP, or TMP jump, has been described in a three stage process: Stage 1 represents initial fouling caused by initial pore blocking and adsorption of solutes; Stage 2 represents the weak gradual rise of TMP over time due to biofilm formation and additional pore blocking; Stage 3 represents a rapid change in the rate of TMP increase (Hai & Yamamoto, 2011; Hwang et al., 2008; Iorhemen et al., 2016).

The main mechanisms of membrane fouling are shown in Fig. 1 and include: cake layer formation, the narrowing of membrane pores, and the blocking of membrane pores (Chang et al., 2002; Hai & Yamamoto, 2011; Hamedi et al., 2019). Some additional mechanisms that impact membrane fouling are biomass characteristics (including mixed liquor suspended solids concentration and soluble microbial product concentration), membrane characteristics (membrane material and pore size), and operating conditions (Brauns et al., 2002; Hong et al., 2009; Jarusutthirak & Amy, 2006).

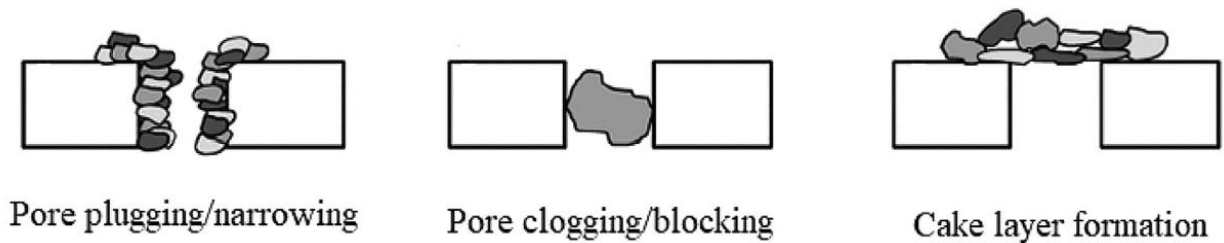Pore plugging/narrowing   Pore clogging/blocking   Cake layer formation

Fig. 1. Three main mechanisms of membrane fouling (Hamedi et al., 2019)

In general, membrane fouling occurs in different severities and can be split into both *reversible* and *irreversible* fouling. *Reversible* fouling can be managed through physical fouling mitigation processes like air scouring and membrane backflushing. *Irreversible* fouling is more serious and requires chemical cleaning procedures (Tsuyuhara et al., 2010). Membrane fouling leads to increased operational costs, material costs, and reduced membrane lifetime. Due to the complexity of MBR technology, membrane fouling, and nonlinear relationships between variables, it is difficult to accurately predict and/or model membrane fouling.

2.2   Existing Mechanistic Models for Membrane Fouling

Since membrane fouling is one of the major challenges for the widespread implementation of MBR technology, many mechanistic models were created to investigate this issue. Belfort et al. (1994) outlined the generic mass transfer approach that is commonly used to model membrane fouling in MBRs. In this approach, the continuity equation along with the Navier-Stokes equation and

boundary conditions are used to determine membrane fouling for a specific channel geometry (Belfort et al., 1994). This type of model seeks to identify membrane fouling through Brownian diffusion, shear-induced diffusion, inertial lift, and the mass transport of particles. Initial mass transport models were built on non-realistic assumptions such as steady-state operation, laminar flow, uniform permeation along the channel geometry, no external forces and constant influence characteristics such as viscosity and density (Belfort et al., 1994; Berman, 1953; Terrill & Thomas, 1969). This work helped establish a base for future mechanistic models.

An empirical model based on hydrodynamic interpretation was introduced by Liu et al. (2002) and investigated the impact of cross-flow velocity, membrane flux, and sludge concentration on the membrane fouling rate. The model uses several mathematical equations and expressions to simplify the rate of filtration resistance, filtration resistance, fluid viscosity, and the relationship between various cross flow velocities – the rate of filtration resistance and filtration resistance are directly related to TMP. The model is easy to use and produces accurate results for a small experimental dataset under specific operating parameters outlined in the study ($R^2 = 0.93$) (Liu et al., 2003). However, the membrane fouling process is too complex to make numerous assumptions about underlying nonlinear relationships; additionally, a wide variety of operating conditions must be considered to provide useful decision support.

The resistance-in-series (RIS) model was proposed by Field et al. (1995) and used mathematical equations and assumptions to characterize membrane fouling; many membrane fouling models in literature are based on the RIS model. In the model, dead-end filtration was assumed, thus the equation for permeation flux was based on Darcy's Law. Permeate flux was considered as a function of TMP, viscosity (temperature-dependent), and three constant resistance values including the intrinsic membrane resistance, cake layer resistance, and irreversible fouling

resistance (including pore narrowing/blocking). Further, the idea of a *critical flux* was proposed, where there exists a flux value for an MBR where any operating flux above it will cause fouling and any operating flux below it will not. In such a model, it is crucial that the correct initial TMP is chosen as to not exceed the critical flux, thereby greatly reducing the rate of fouling (Chang et al., 2002; Field et al., 1995). Since it was first introduced, the RIS model has been used extensively to analyze membrane fouling within MBRs. The RIS model has been extended to conclude that the resistance of the activated sludge suspension in an MBR is equal to the sum of the resistances of the suspended solids, colloids, and solutes (Meng & Yang, 2007). Ludwig et al. (2012) utilized the RIS model within a dynamic mechanistic simulation model incorporating suction pressure, water head pressure, and crossflow aeration from a full-scale municipal WWTP. The model was calibrated and validated using a constrained linear least-squares optimization procedure to limit the root mean squared error (RMSE) between predicted and observed values. After testing, the model predicts TMP accurately (86% - 95% accuracy) in both hollow-fibre and flat-sheet MBR configurations (Ludwig et al., 2012) and can be integrated into existing active sludge simulation models (ASM). Further, Liang et al. (2006) integrated the RIS model using parameters for mixed liquor suspended solids (MLSS), dissolved organic matter, and varying resistances over time. The model was trained and validated with experimental data from a pilot submerged MBR plant yielding good accuracy between observed and predicted data. Mannina & Di Bella (2012) created an integrated model utilizing biological equations and the resistance-in-series method for membrane fouling characterization.

While promising, the RIS model is built on the assumption that resistances behave the same individually as they do in combination; it has been proven that the sum of individual components are different from that of the mixture (Chang et al., 2009) and it is not clear how the change in one

resistance will affect another. Similarly, they do not consider concentration polarization (Belfort et al., 1994) of contaminants at the membrane surface. Due to their limitations, mechanistic and empirical models should be used with caution when quantifying membrane fouling in full-scale MBR applications.

2.3   Existing Data-Driven Models for Membrane Fouling

In recent times, data-driven techniques have become increasingly more popular for modeling membrane fouling. Data-driven models, unlike mechanistic models, do not rely on underlying mechanistic or mathematical principles, they rely on information from datasets. The assumptions made in mechanistic models can sometimes make them ineffective for real-world applications and cannot investigate complex underlying relationships between variables. In comparison, data-driven techniques can effectively use information in datasets to model nonlinear relationships between input and output variables. There are many different algorithms available for data-driven techniques. A commonly used algorithm in the formulation of data-driven models is the artificial neural network (ANN) and its derivatives. For instance, Dornier et al. (1995) proposed a back-propagation ANN to predict membrane fouling by modeling hydraulic resistance defined by Darcy's Law. The ANN consisted of one input layer, two hidden layers, and one output layer using a sigmoidal activation function. Since there are one or more hidden layers, this type of model would be considered a multilayer perceptron (MLP) ANN. The input variables for this model were time, TMP, and cross flow velocity; the output variable for the model was hydraulic resistance. The model was trained and tested on 6 different short-term experimental datasets in which the coefficient of determination (R^2) value for constant and variable operating conditions was 0.97 and 0.90, respectively. This model showed positive results and led to further investigation into data-driven models for membrane fouling prediction.

Liu et al. (2009) created an MLP ANN model using permeate flux, raw water turbidity, UV_{254} measurement (for organic content detection), operating time, and backwash type (binary) as input variables; the output variable for the model was TMP. The set of 990 data was normalized and split roughly 60:40 for training and testing purposes. To train the model, the Levenburg-Marquardt (LM) technique was used for back propagation and used more than 120 epochs. This technique searches for the minimum mean squared error (MSE) between observed and predicted values. The results of the model showed a good result between observed and predicted values with correlation coefficient (R) of 0.85. The authors determined that water quality parameters including (turbidity and UV_{254}) were just as important as operational parameters (flux, backwash frequency, and time) for the prediction of TMP.

Mirbagheri et al. (2015) created both MLP and radial basis function (RBF) ANNs to quantify membrane fouling. In RBF ANNs, a radial basis function is applied between the input and hidden layer, and a linear activation function is used between the hidden and output layers. In the MLP ANNs, there is no function between the input and hidden layer, and the sigmoidal function is commonly used between the hidden and output layers (Bayram et al., 2016). For both models, the input variables were time, total suspended solids (TSS), chemical oxygen demand (COD_{in}), solids retention time (SRT), and MLSS; the output variables for the models were TMP and permeability. The dataset used in this study comes from a submerged MBR pilot plant fed with municipal wastewater. The dataset was comprised of 60 days of experimental data with a daily temporal resolution. The data was normalized and split 70:15:15 for training, testing, and validation, respectively. To train the model, the LM technique was used for back propagation with a maximum of 1000 epochs. The trial-and-error approach and genetic algorithm (GA) approach were both used to calibrate the models, where the GA approach produced better parameters for the

model. The authors concluded that the MLP ANN model performed better than the RBF ANN model with R^2 values of 0.99 and 0.98, respectively.

Hazrati et al. (2017) created an artificial neural network (ANN) with a 4-month dataset on a pilot-sized MBR utilizing mixed liquor suspended solids (MLSS) concentration, hydraulic retention time (HRT), and time as input variables to predict chemical oxygen demand (COD) removal and TMP with great accuracy. The results for these models showed excellent accuracy when predicting on both training and testing datasets. Li et al. (2020) created a random forest (RF) model on the Hadoop big data platform to determine membrane fouling by predicting membrane flux. ANN and support vector machine (SVM) models were created to evaluate the RF model's prediction accuracy and the results showed that the RF model had the best prediction accuracy.

Many data-driven models with various algorithms have been applied to predict membrane fouling; however, many of these models have drawbacks. Common drawbacks in the models include small datasets with coarse temporal resolution, datasets from pilot MBR plants, and non-calibrated models. Additionally, these models lack uncertainty analyses to characterize both hyperparameter uncertainty and uncertainty due to model structure. The incorporation of prediction intervals for output prediction can be extremely useful to provide robust decision support.

In summary, data-driven models have produced promising results, but are not yet ready to be implemented for full-scale MBR decision support. The development of more robust data-driven models for the prediction of membrane fouling is required.

Chapter 3 – Study Area and Data Collection

3.1   MBR

MBR technology is a wastewater treatment technology that is increasingly used for municipal and commercial applications across the world. MBRs produce extremely high effluent quality with a small capital footprint (Hai & Yamamoto, 2011). Membrane fouling is one of the major challenges to wide-scale MBR implementation for wastewater treatment plants. This occurs over time as solids build up on the membrane surface and within membrane pores, increasing transmembrane pressure (TMP) between the inside and outside of the membrane. Fouling mainly occurs due to pore narrowing, pore clogging, and cake formation (Iorhemen et al., 2016). Membrane fouling can be split up into two main types of fouling, *reversible* and *irreversible*.

Reversible membrane fouling is the deposition of solids on the membrane surface that are readily removable. This is primarily controlled through air scouring, membrane back pulsing (backflushing), and maintenance cleaning (Chang et al., 2002). Air scouring utilizes coarse bubbles from an aeration device to produce shear forces across the membrane surface, removing solids that are deposited on the membrane. The crossflow velocity of air bubbles is produced by bubbling air with an aeration device located underneath the membrane (Sofia et al., 2004). Membrane back pulsing is used to reverse the flow of permeate through the membrane. This process has been successful in removing most of the reversible fouling due to pore blocking, transporting the solids back into the bioreactor and partially dislodging loosely attached sludge cake from the membrane surface (Bouhabila et al., 2001; Hai & Yamamoto, 2011). Maintenance cleaning procedures with moderate chemicals are generally conducted on a weekly basis and helps reduce the need for more intense recovery cleaning procedures.

Irreversible fouling occurs over time where TMP increases to a point where further permeate production is no longer sustainable. Irreversible membrane fouling is controlled by recovery cleaning through means of chemical cleaning. Two commonly used chemicals for this type of cleaning are sodium hypochlorite and citric acid used for organic and inorganic foulants, respectively (Hai & Yamamoto, 2011). Both chemicals are used for either maintenance or recovery cleaning procedures, however the concentration used will be higher in recovery cleaning procedures.

Membrane bioreactors are operated at either constant permeate flux with a changing TMP or a changing permeate flux with a constant TMP (Guo et al., 2012). Membrane fouling occurs as TMP increases to maintain a constant permeate flux or when permeate flux decreases while maintaining constant TMP. Membrane fouling and the subsequent rise of TMP (at a constant flux) is determined by physical and chemical interactions. These interactions and resulting membrane fouling are controlled by foulant characteristics, influent chemical composition, membrane properties, and hydrodynamic conditions (Li & Elimelech, 2004). Further, the length of the operation cycle and the frequency and length of the back pulse is important in controlling membrane fouling.

3.2   Wastewater Treatment Plant

The WWTP data used in this study is from a confidential municipal wastewater facility using the submerged membrane bioreactor (SMBR) technology equipped with hollow fiber membranes. This confidential WWTP is designed with an average influent capacity of 60,000 m^3/d with a peak flow of 7,020 m^3/d. The effective membrane surface area across all membrane trains is 412 m^2. The plant utilizes a pretreatment phase consisting of 10 mm screening, equalization with a buffer tank, 1 mm fine screening, and grease removal/cooling. The primary treatment phase consists of

four aeration tanks using activated sludge followed by eight SMBR tanks to produce treated effluent. The tertiary treatment involves chlorination and sludge treatment through gravity belt thickening, aerobic digestion, and dewatering using filter presses. A graphical representation of the plant is shown below in Fig. 2. At this WWTP, membrane back pulsing occurs every 11 minutes. Maintenance cleaning cycles generally occur 1 to 2 times per week and use sodium hypochlorite or citric acid as cleaning chemicals. Recovery cleaning cycles generally occur 2 to 4 times per year utilizing higher concentrations of sodium hypochlorite or citric acid and a different cleaning procedure.
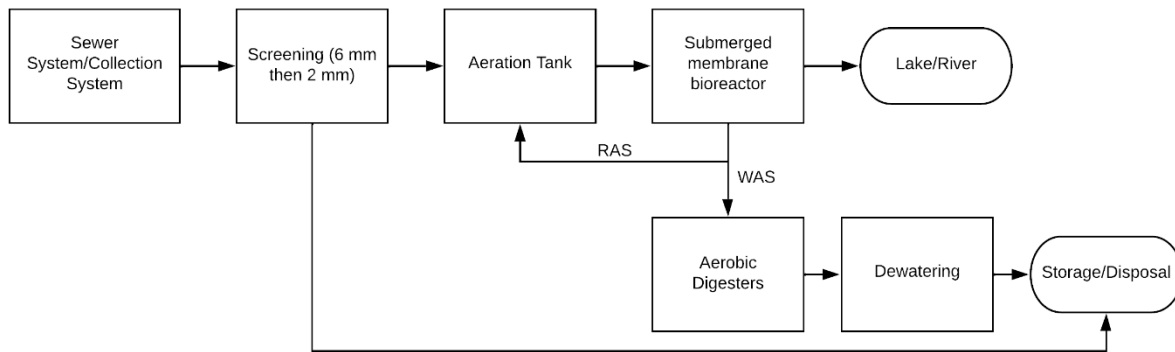

Fig. 2. A simplified process diagram for the confidential WWTP

3.3   Data Collection

The timeseries data used in this study were collected through an automated asset performance management system, InSight. The InSight system collects data every other back pulsing cycle, thus the MBR operational data is available at a 22-minute temporal resolution. The before back pulse (BBP), during back pulse (DBP), and after back pulse (ABP) operational data measured between January 1, 2017, to December 31, 2020, were obtained and used for developing unique

models for BBP, DBP, and ABP TMP. The total dataset consisted of 83,982 instances. Operational data from 2015 to 2016 was available but omitted due to many missing data observations. Due to the cyclical nature of the data with local maximum and minimum values, common outlier determination methods like the three-standard deviation (3σ) method were not used. Instead, expert elicitation was used to manually identify and remove 159 outlier instances resulting in 83,823 instances to be used to create the models. Continuous before, during, and after back pulse data are available every 22 minutes for TMP (kPa), flow (m^3/d), flux (LMH), and permeability (LMH/kPa). The TMP data from different stages (i.e., before, during, and after back pulse) of the membrane cycle is shown in Fig. 3. Fig. 3 shows that BBP and ABP TMP data are negative, whereas DBP TMP data are positive. The positive and negative represents the direction of flow through the membrane. When TMP is negative, there is production in the MBR where water from the membrane tank is being pulled through the membrane, leaving as permeate. When TMP is positive, there is a back pulse occurring where permeate is being pushed back through the membrane and into the membrane tank. Additional after back pulse variables include permeate temperature (°C) and tank level (m) measured after each back pulse. Binary condition data are available for maintenance clean, recovery clean, production, and standby status. MLSS concentration (g/L) data is available at a temporal resolution of 1 minute.

The collected data were preprocessed to prepare a dataset with a temporal resolution of 22 minutes for model development. To deal with the discrepancy in temporal resolution, the most recent MLSS concentration relative to cycle data is used because the composition of the pretreated water does not change significantly over short periods of time (Wang et al., 2021). A full list of input variables for model development is shown in Table 1.

Table 1. Raw data collected from InSight

Stage	Variable	Unit	Frequency
BBP	BBP TMP	kPa	22-minute
	BBP Flow	m³/d	22-minute
	BBP Flux	LMH	22-minute
	BBP Permeability	LMH/kPa	22-minute
DBP	DBP TMP	kPa	22-minute
	DBP Flow	m³/d	22-minute
	DBP Flux	LMH	22-minute
	DBP Permeability	LMH/kPa	22-minute
ABP	ABP TMP	kPa	22-minute
	ABP Flow	m³/d	22-minute
	ABP Flux	LMH	22-minute
	ABP Permeability	LMH/kPa	22-minute
	ABP Permeate Temperature	°C	22-minute
	ABP Tank Level	Metre	22-minute
Miscellaneous	Date	MM/DD/YYYY	Daily
	Days	Numeric	Daily
	Time	Minutes	22-minute
	MLSS	g/L	1-minute
	MC Status	Binary	1-minute
	RC Status	Binary	1-minute
	Cumulative Flow	m³/d	22-minute
	Cycles Since RC	Numeric	22-minute

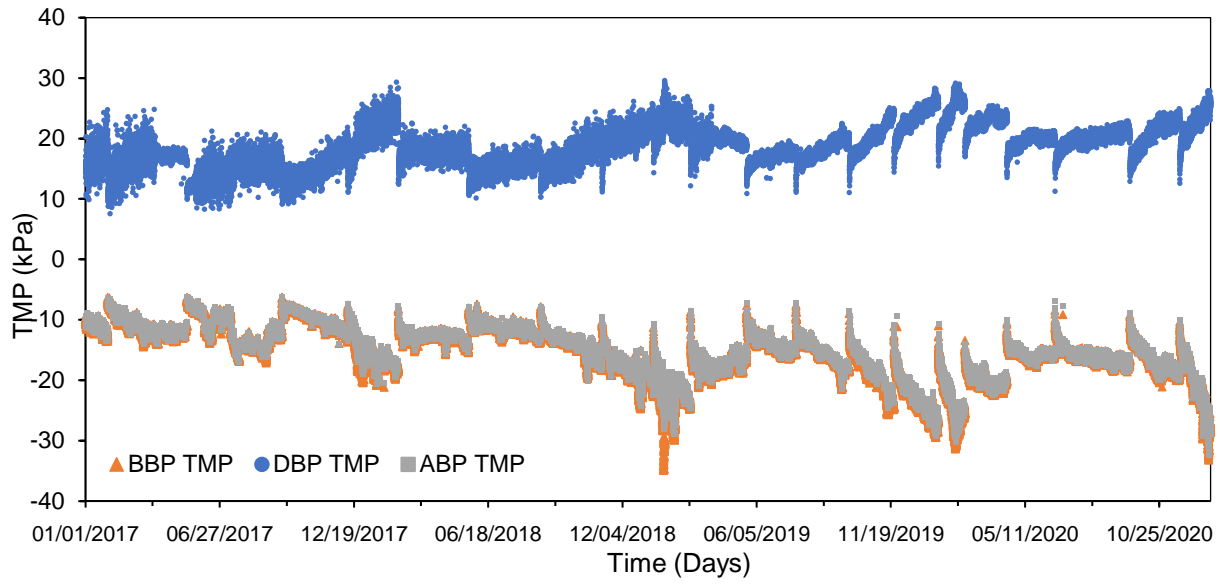Note: BBP = Before back pulse, DBP = During back pulse, ABP = After back pulse

Fig. 3. Timeseries data for BBP, DBP, and ABP TMP from 2017 to 2020

Chapter 4 – Methodology

4.1   Machine Learning

For this research, RF, ANN, and LSTM algorithms were chosen. The random forest has proven to be an effective algorithm for use in wastewater applications (Li et al., 2020; Zhou et al., 2019) and is useful for investigating variable importance by reduction of MSE; this process is well-outlined by Grömping (2009). The ANN has proven to be very effective for membrane fouling prediction applications (Dornier et al., 1995; Hazrati et al., 2017; Liu et al., 2009; Mirbagheri et al., 2015) and has been used to check the accuracy of other algorithms (Li et al., 2020). LSTM has not been used for the prediction of membrane fouling but has been proven useful in wastewater effluent quality prediction (Farhi et al., 2021) due to its ability to recognize patterns and sequences in timeseries data.

4.1.1   Random Forest

The RF method is an ensemble method where the combination of decision tree or model tree predictors are used to make predictions. This method was introduced by Breiman (2001) influenced by previous work (Amit & Geman, 1997; Dietterich, 2000). It combines bootstrapping and random split selection to create independent decision trees whose outcomes are aggregated together to make a prediction (Breiman, 2001). The common element used in all procedures is that for the kth tree, a random vector Θ_k is generated that is independent of the past random vectors $\Theta_{1,...,}\ \Theta_{k-1}$ but with the same distribution; and a tree is grown using the training set and random vector Θ_k, resulting in a classifier $h\ (\mathbf{x}, \Theta_k)$ where $\mathbf{x}$ is an input vector (Breiman, 2001). More trees are generated independently until there are many trees in the forest. Then, a majority voting process happens, and a decision or prediction is made. Since bootstrapping is used, there are samples 'in-the-bag' which are used as the training datasets and samples 'out-of-bag' used as the testing

datasets. Thus, out-of-bag (OOB) Mean Squared Error (MSE) can be calculated without additional computational resources by comparing the majority vote of the value for OOB data and comparing it to the true OOB data.

Advantages of the random forest method include the robustness to changes in hyperparameters, that they are less likely to suffer from overfitting compared to other methods, and the ability to see variable importance through regression impurity (Breiman, 2001; Segal, 2003; Wang et al., 2021). In this study, the *randomForest v 4.6-14* package within the open-source programming language R is used to create the RF models. This package uses Breiman and Cutler's random forest algorithm for classification and regression (Breiman, 2018).

4.1.2   Artificial Neural Network

Artificial Neural Network (ANN) models relate the input and output of a system to map nonlinear input-output relationships. These methods have been widely successful in wastewater applications, including influent flow prediction and membrane fouling through the prediction of TMP (Boyd et al., 2019; Mirbagheri et al., 2015; Zhang et al., 2019; Zhou et al., 2019). An ANN with one input layer, one output layer, and one or more than one hidden layer(s) is a multi-layer perceptron (MLP). The general idea behind this approach is that the information from the input layer is fed forward to a hidden layer with neurons that will process the information and pass it forward through an activation function to create a model output. The following equation shows how one neuron processes information from the previous layer to the next layer.

$$x_{j+1} = f\left(\sum_i w_{i,j+1} x_{i,j} + b_{j+1}\right) \tag{1}$$

where x is a neuron, j is the layer index, i is the neuron index for layer j, w is the weight assigned between two layers, b is the bias weight term, and $f(\bullet)$ is the activation function (Wang et al.,

2021). In this study, the sigmoidal activation function is used between the hidden and output layers. The *caret v 6.0-89* package (Kuhn, 2021) within R is used with model selection *'nnet'* to produce regression based MLP ANNs to create before, during, and after back pulse TMP predictive models. The ANN utilizes the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to optimize model parameters, changing neuron weights to minimize error over iterations.

4.1.3   Long Short-Term Memory Network

Long short-term memory (LSTM) models are a type of recurrent neural network (RNN) introduced by Hochreiter and Schmidhuber (1997). LSTM units are now commonly comprised of a forget gate, input gate, and an output gate. In the forget gate, the model can decide what information in the time series will be kept and what will be thrown away. In the input gate, new information is fed into the model and the gate decides what new information will be kept in the cell state. The output gate decides what information can be used as the output based on the cell state (Yu et al., 2019). The architecture of this type of LSTM unit is shown in Fig. 4.
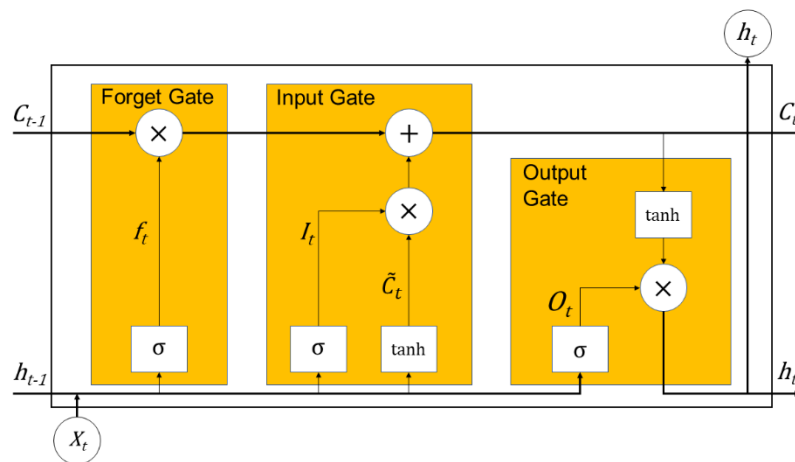

Fig. 4. Architecture of an LSTM unit

The forget gate is represented by f_t where it has a value between 0 or 1 due to the sigmoid function applied to it. If the value is 1, all information will be saved inside the network. If the value is 0, all information will be discarded.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

The input gate is represented by I_t where it also has a value between 0 or 1. If the value is 1, the new information will be allowed to be moved to the cell state; whereas if the value is 0, it will not be allowed.

$$I_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

The cell state is represented by C_t and denotes information carried from one LSTM unit to the next.

$$C_t = f_t \cdot C_{t-1} + I_t \cdot \tilde{C}_t \tag{4}$$

The new information from input variables is denoted by $\tilde{C}_t$ and is transformed by the hyperbolic tangent function to obtain a value between -1 and 1, where -1 indicates subtraction of information from the cell state and 1 indicates addition of information.

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_{\tilde{c}}) \tag{5}$$

The output gate is represented by O_t with a value between 0 or 1 where 1 allows information transfer and 0 does not. This is then be multiplied with the cell state information after it is transformed by the hyperbolic tangent function; the product of the multiplication is then used as the output of the unit, h_t.

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = O_t \cdot \tanh(C_t) \tag{7}$$

4.2   Model Development and Performance Evaluation

4.2.1   Input Variable Selection

Input variable selection was conducted to determine the best variables to be used in each of the models. The models are carefully designed to limit correlated input variables to investigate the impact that other input variables have on the output of the model. A manual input variable omission process was used to determine the decrease in model accuracy after removing input variables; the final models have a small drop in model accuracy compared to preliminary models and it is easier to see which input variables impact model accuracy in the final models. Additionally, less input variables used results in less computational resources required to train and tune the model, making it easier to apply to other WWTPs.

Table 2 shows the selected input variables for models at each stage, statistical measures, and the correlation of that variable to other variables. Since TMP is directly related to permeability where permeability = flux/TMP, permeability is excluded. It is worth mentioning that in cases where it is not important to limit correlated input variables, permeability can be included to predict TMP, and has been shown to increase predictive accuracy in preliminary models. To model the effect of recovery cleaning, a new input variable "Cycles since RC" was introduced to count the number of membrane cycles between recovery cleaning events. This is important because the cyclical behaviour shown in Fig. 3 is caused by recovery cleaning cycles where a drastic change in TMP is shown. To incorporate the general trend that TMP increases as membrane life progresses, cumulative flow (m^3/d) was also introduced as a new input variable calculated by summing before back pulse flow from each membrane cycle between the beginning of 2017 and the end of 2020.

Once the final input variables were selected for each of the three models (i.e., BBP, DBP, and ABP), a time lag based on chronological order was introduced to certain input variables depending on the point in the membrane cycle. For instance, for the BBP and DBP models, the ABP permeate temperature and tank level values were taken from the previous cycle. For the ABP models, the ABP permeate temperature and tank levels were taken from the current cycle. All other input variables were taken from the same cycle.

Table 2. List of input variables used in the models, statistical measures, and relationship to other variables

Stage	Variable	Use in model	Min	Mean	Max	Standard Deviation	Coefficient of Variation	Relationship to other variables
BBP	BBP TMP	Y (Target, not used in DBP or ABP models)	-34.96	-15.35	-6.22	4.36	-0.284	BBP Permeability
	BBP Flow	Y (Not used in DBP or ABP models)	198.3	300.9	397.5	17.1	0.057	BBP Flux
DBP	DBP TMP	Y (Target, not used in BBP or ABP models)	7.61	18.64	29.61	3.34	0.179	DBP Permeability
	DBP Flow	Y (Not used in BBP or ABP models)	300.1	405.8	500.0	6.50	0.016	DBP Flux
ABP	ABP TMP	Y (Target, not used in	-32.41	-15.13	-6.16	4.21	-0.278	ABP Permeability

		BBP or DBP models)						
	ABP Flow	Y (Not used in BBP or DBP models)	170.1	301.0	399.2	17.0	0.056	ABP Flux
	ABP Permeate Temperature	Y	24.2	32.3	38.3	3.50	0.108	
	ABP Tank Level	Y	2.999	3.189	3.259	0.014	0.004	
	Days	Y	-	-	-	-	-	
	MLSS	Y	0.702	5.855	8.854	0.588	0.100	
	Cumulative Flow	Y	-	-	-	-	-	BBP Flow
	Cycles Since RC	Y	-	-	-	-	-	

4.2.2   ML models for TMP prediction

In this study, RF, ANN, and LSTM were used to create nine different models that predict TMP (kPa) before, during, and after membrane back pulsing. Nine different models, three for each algorithm, were created because the output values for TMP varies between stages of the MBR production cycle. Creating separate models provides the best solution for accurate prediction of TMP at any point in the membrane cycle because the training and testing target data are exclusive to that part of the cycle (before, during, or after membrane back pulsing). The dataset was split randomly 75:25 for training and testing, respectively. There are 62,868 observations used for

training and 20,955 observations used for testing. The variables used in each model vary depending on the specific model and are shown in Table 1.

To obtain an optimal RF model for each stage, a grid search was used to determine the optimal hyperparameters. The most optimal BBP RF model had 2,407 trees (*ntree*) and the number of variables tried at each split (*mtry*) was equal to 4. Since the random forest model already uses bootstrapping to resample data, cross validation was not used. The optimal DBP RF model had *ntree* equal to 1,598 and *mtry* equal to 2. The optimal ABP RF model had *ntree* equal to 1,576 and *mtry* equal to 4.

To obtain the optimal ANN model for each stage, 5-fold cross validation was used on the training set while the testing set was used as a hold-out set for a maximum of 1,000 iterations per hyperparameter combination. The performance differences were negligible between models trained with 5-fold cross validation and models trained with 10-fold cross validation. Both training and testing datasets were centered and scaled using the mean and standard deviation of the training dataset. A grid search was used to determine the optimal hyperparameters available for *decay* and *size*. For the BBP ANN model, the *decay* and *size* were 0.001 and 30, respectively. The resulting ANN had a 7-30-1 network architecture. For the DBP ANN model, the *decay* and *size* were 0.01 and 30, respectively. For the ABP ANN model, the *decay* and *size* were 0.001 and 29, respectively.

To obtain the optimal LSTM model for each stage, a stochastic gradient descent (SGD) optimizer from the *PyTorch* package (Paszke et al., 2019) in *Python 3.8.8* is used to optimize hyperparameters. The number of epochs used is equivalent to twice the number of observations in the training dataset. In each LSTM model, the batch size is 20, a parameter representing the number of previous observations used for training each training observation. The tuning for *decay* and *size*

was done in conjunction with the uncertainty analysis and was determined to be 0 and 14, respectively, for BBP, DBP, and ABP models.

4.2.3 Performance Evaluation

Four model performance measures were used to compare the RF, ANN, and LSTM models developed in this study. These statistical measures include the Coefficient of Determination (R^2), MSE, Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). R^2 is a useful index in regression analyses, providing a good sense of how well data fit to a regression line. In this case, R^2 is used to determine the correlation between model predictions and observed target variables from the testing datasets. R^2 is calculated using Eq. (8):

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2} \tag{8}$$

where n is the number of samples, y_i is the observed value, and $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean value of a sample. There is no universally accepted value range for R^2, because a good value is relative to the problem. However, the closer to 1 the higher the correlation between observed and predicted data.

MSE is the average squared error between predicted and actual values. The term is squared so any negative value is simply taken as a magnitude and does not negatively impact the accuracy of the term. MSE is calculated using Eq. (9).

$$MSE = \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{9}$$

MAPE is the average error between predicted and observed data expressed as a percentage. It is calculated using Eq. (10).

26

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \times 100 \tag{10}$$

RMSE is the squared root of the MSE. The advantage to having this metric is that the error

is easy to understand because it will have the same units as the output variable (i.e., kPa). However,

a disadvantage is that RMSE is difficult to use to compare different models because of varying

system elements and different system operation parameters. It is calculated using Eq. (11).

$$RMSE = \sqrt{MSE} = \sqrt{\sum_{i=1}^{n}\frac{(y_i-\hat{y}_i)^2}{n}} \tag{11}$$

4.3   Uncertainty Analysis

An uncertainty analysis was conducted to determine both hyperparameter uncertainty and model

structure uncertainty. Latin Hypercube Sampling (LHS) was used to randomly generate

hyperparameter values for the quantification of parameter uncertainty. LHS is a method of

sampling that can be used to produce hyperparameter values to be fed into the developed models

for uncertainty quantification. In this case, the LHS process utilized sampling from a uniformly-

distributed square grid with two dimensions, selecting only one sample per combination of row

and column (Stein, 1987). This provides a better distribution of hyperparameters across the

sampling space when compared to using a random sampling method such as Monte Carlo

simulation. To create each model's Latin Hypercube (LH), uniform distributions of

hyperparameters unique to each algorithm were created. Then, the *LHS v 1.1.1* package in *R*

(Carnell, 2021) was used to create each LH. This function utilizes a Columnwise Pairwise (CP)

algorithm to generate an optimal grid to sample from.

For each model, a LH was created containing 100 unique combinations of hyperparameters

for each of the 3 algorithms. Important hyperparameters for each of the three algorithms (i.e., RF,

ANN, and LSTM) were identified based on the literature for uncertainty (Alvarez & Salzmann, 2016; Probst et al., 2019; Zhang et al., 2019). The hyperparameters selected were *size* and *decay* for the ANN models, *ntree* and *mtry* for the RF models, and *size* and *decay* for the LSTM models. One hundred iterations of the models were run, recording each unique timeseries and corresponding performance metrics when predicting on the testing dataset. The timeseries were used to create prediction intervals (PIs), thereby quantifying a range of prediction instead of a single deterministic value.

To quantify model structure uncertainty, the impact of algorithm selection on the prediction of the most extreme minimum and maximum TMP values were investigated. For each of the nine models, there were 100 associated predictions for the two most extreme values. The minimum, maximum, and average predictions were investigated to determine how the models perform at extreme prediction.

Chapter 5 – Results and Discussion

5.1   Model Performance

RF, ANN, and LSTM models were created for each stage of the membrane cycle. These models were tuned to optimize performance while minimizing the risk of overfitting. The resulting performance metrics for each model and stage are shown in Table 3. The best-performing algorithm was the random forest with the highest R^2 and lowest RMSE for each model for all stages of the MBR cycle. For prediction of testing data, the metrics range between 0.927 to 0.996 and 0.264 kPa to 0.904 kPa for R^2 and RMSE, respectively. Comparatively, for ANN prediction of testing data, the metrics range between 0.910 to 0.973 and 0.692 kPa to 1.00 kPa for R^2 and RMSE, respectively. For LSTM prediction on testing data, the metrics range between 0.878 and 0.915 and 1.170 kPa to 1.397 kPa for R^2 and RMSE, respectively.

While the differences in metrics seem small, the BBP RF model had a RMSE value that is roughly 3 times smaller than the BBP ANN model's RMSE value, showing a significant difference in accuracy. Model accuracies for BBP and ABP were similar because BBP and ABP TMP data are closely related – both gradually increase over time as membrane life goes on. The difference in performance metrics between BBP/ABP RF and ANN models were relatively similar in magnitude with the random forest models being more accurate in training and testing. However, DBP RF and ANN models have performance metrics much closer in magnitude when comparing performance metrics for the testing dataset – the difference in RMSE is 0.1 kPa. The LSTM models have satisfactory performance but were less accurate than the RF and ANN models. However, the optimal LSTM models had less variation in performance metrics across the BBP/DBP/ABP models in comparison to the RF and ANN models.

Table 3. Performance metrics for each model, by algorithm and cycle stage

Model	Cycle Stage and Performance Metric											
	BBP				DBP				ABP			
	R^2	MAPE	RMSE	MSE	R^2	MAPE	RMSE	MSE	R^2	MAPE	RMSE	MSE
RF (Train)	0.999	0.47%	0.116	0.013	0.985	1.61%	0.415	0.172	0.999	0.49%	0.121	0.015
RF (Test)	0.996	1.08%	0.264	0.070	0.927	3.55%	0.904	0.816	0.996	1.14%	0.277	0.077
ANN (Train)	0.973	3.51%	0.721	0.520	0.910	4.17%	1.00	1.01	0.974	3.46%	0.684	0.467
ANN (Test)	0.973	3.52%	0.725	0.525	0.910	4.19%	1.00	1.01	0.974	3.47%	0.692	0.478
LSTM (Train)	0.933	5.48%	1.125	1.266	0.882	4.81%	1.147	1.317	0.936	5.38%	1.060	1.123
LSTM (Test)	0.932	5.54%	1.15	1.322	0.883	4.78%	1.145	1.310	0.935	5.42%	1.082	1.171

It is worth mentioning that the predictive performance for all models decreases when attempting to predict extreme values. In machine learning models, extreme value prediction poses a problem to a trained model and there is often a trade-off between model accuracy and overfitting (Li et al., 2020; Zhou et al., 2019). In the BBP scatterplot shown in Fig. 5 (a), many of the outliers are toward the bottom-left quadrant and above the line indicating where predicted data is equal to observed data. This indicates that on average, BBP RF, ANN and LSTM models were predicting more conservatively with a more extreme predicted value (compared to the observed value), with the LSTM model predicting the furthest from the observed value. The ANN model made a higher number of extreme predictions when compared to the RF model. In the DBP scatterplot shown in Fig. 5 (b), the differences between RF, ANN, and LSTM models were much less apparent. Extreme predictions were better distributed across the prediction range when compared to the BBP scatterplot where extreme predictions are more frequent towards either end of the prediction range. In the ABP scatterplot shown in Fig. 5 (c), the results for RF, ANN, and LSTM models were like

those discussed in the BBP scatterplot. However, extreme predictions were better distributed throughout the prediction range compared to the skewed behaviour to the lower left quadrant seen in the BBP scatterplot. Most outliers in the top-right quadrant were from the LSTM model, predicting a magnitude of TMP that was lower than the observed value, and resulted in a less conservative prediction.
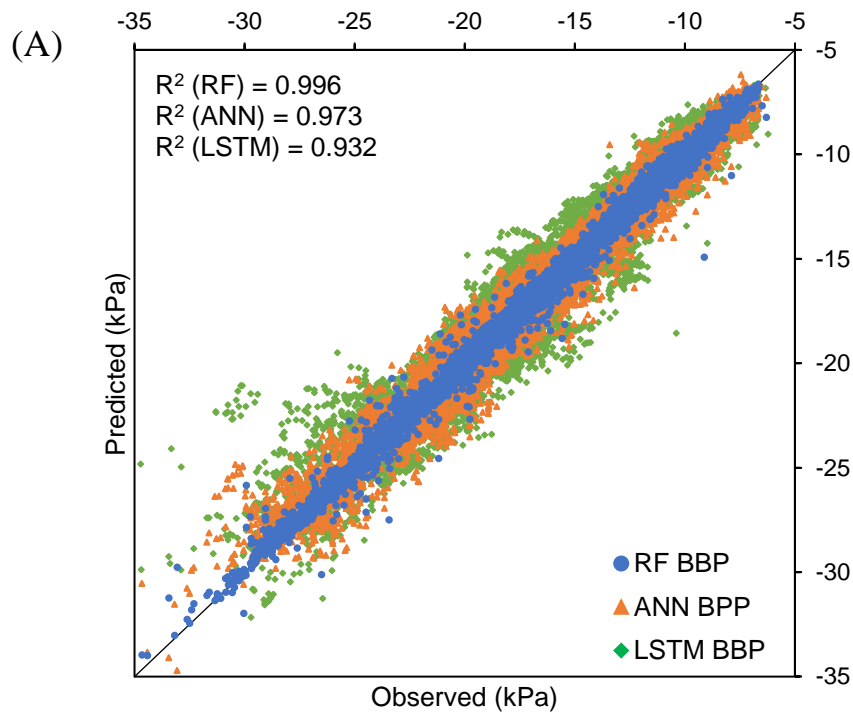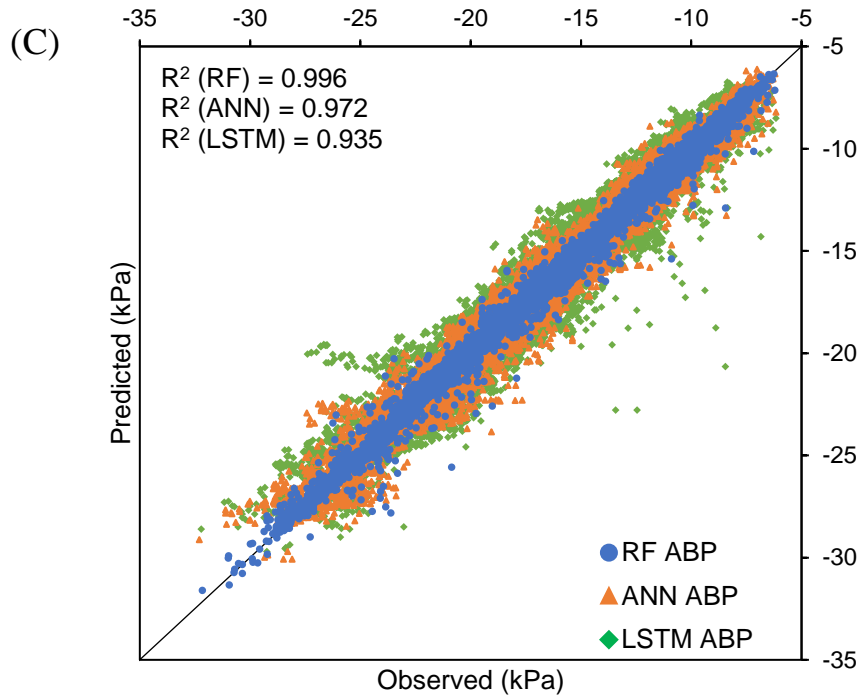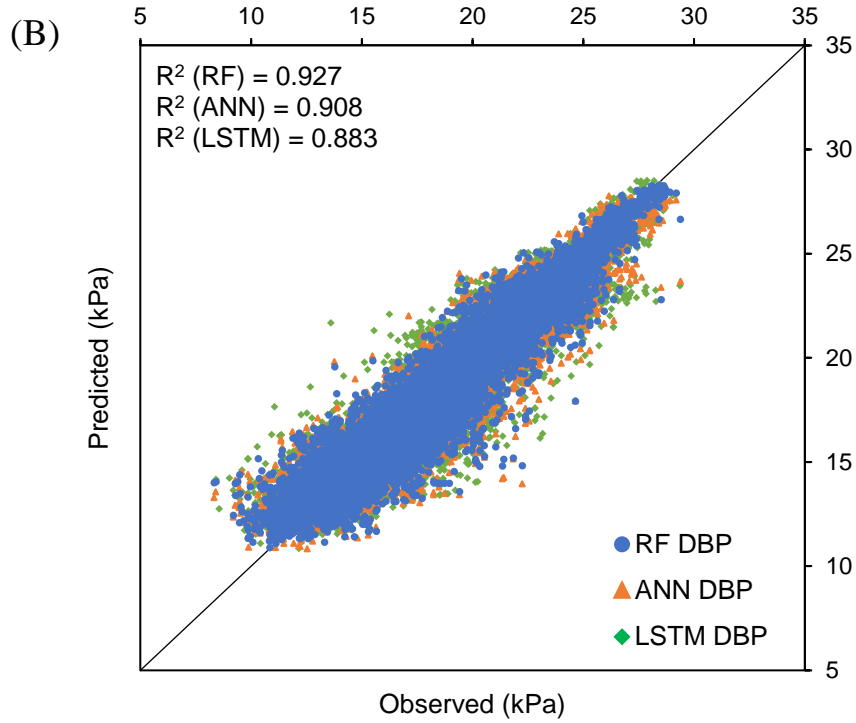
Fig. 5. Scatter plots for predictions from RF, ANN, and LSTM: (a) testing dataset for BBP models, (b) testing dataset for the DBP models, (c) testing dataset for the ABP models

Since RF had the highest overall accuracy, it was selected for further analysis. In the BBP RF timeseries plot shown in Fig. 6 (a), very good results were shown with respect to prediction accuracy. The model could identify the cyclical nature of TMP caused by recovery cleaning cycles despite the random selection of testing data. The predictive accuracy of the model decreases at extreme values, but the model still provides a good estimate of TMP. It is important to note that the extreme values are often single observed values and do not represent the general trend of TMP. Shown in Fig. 6 (b) is the DBP RF timeseries plot. The predictive accuracy was relatively good but not as good as the BBP and ABP RF models. The model had a tougher time recognizing the local max-min values in the cycle. However, it is worth noting that between 2017 and mid-2019 the DBP TMP observed values contained many more outliers or extreme values. From mid-2019 to the end of 2020, the predictive accuracy increased, `and the model better predicted the observed value, even at extreme values. Fig. 6 (c) represents the ABP RF timeseries plot. The results shown were very similar to those shown by the BBP RF timeseries plot in Fig. 6 (a). Observed values between BBP and ABP were similar because the ABP values from one cycle essentially become the BBP values for the next cycle. However, the negative trough and global minimum shown in BBP TMP data in mid-2019 is not present in the ABP TMP data. It is noteworthy that the ANN and LSTM timeseries plots have predictions with a higher variation between observed and predicted values. The timeseries plots for the ANN and LSTM models can be found in the Appendix as Figures A1-A3 and A4-A6, respectively.
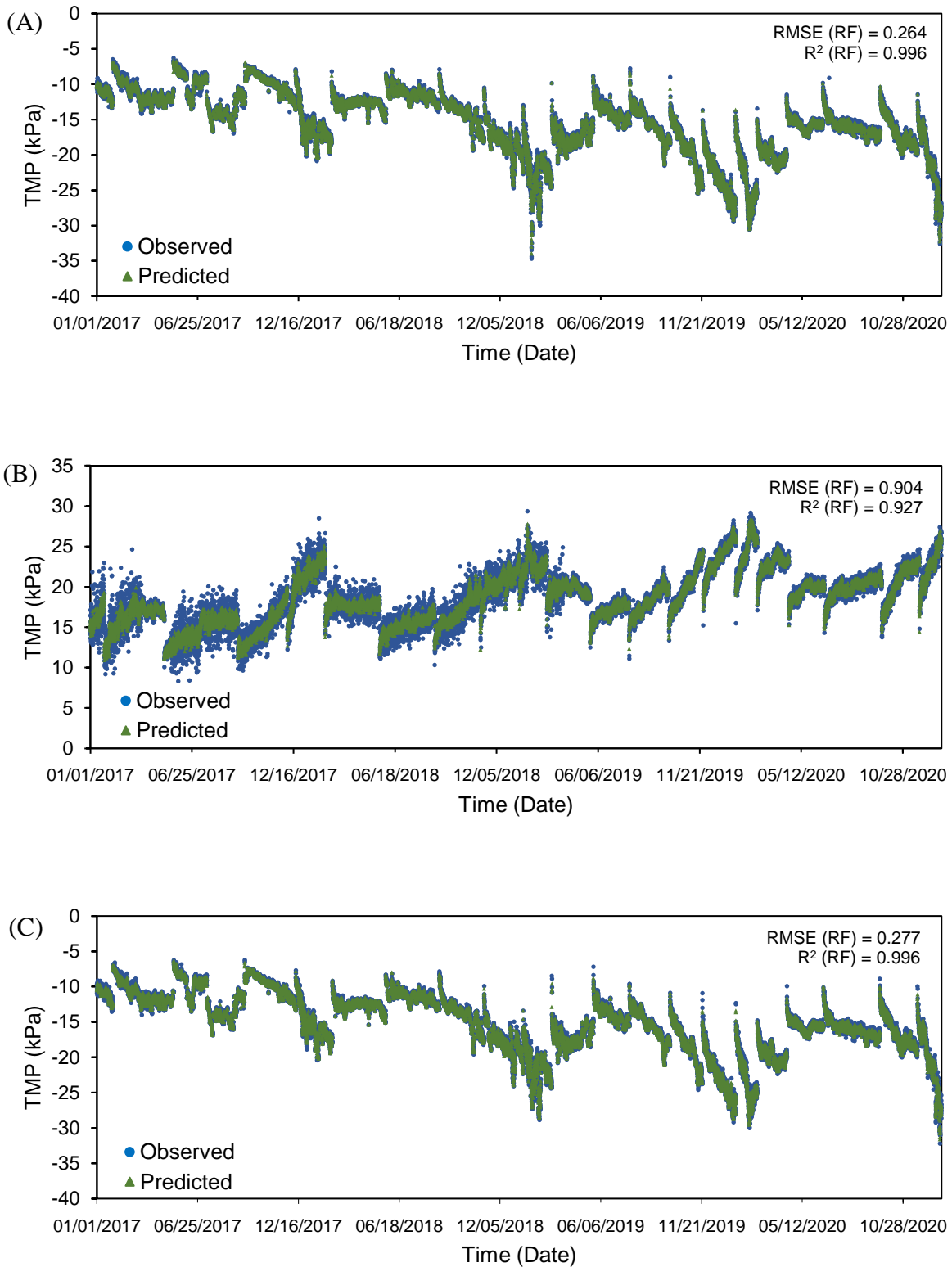
Fig. 6. Timeseries plots for RF: (a) testing dataset for the BBP model, (b) testing dataset for the DBP model,

(c) testing dataset for the ABP model

The results achieved for the developed models were comparable to or better than some existing empirical, semi-empirical, and ML models found in literature that also use TMP prediction to characterize membrane fouling (Liang et al., 2006; Liu et al., 2009; Mannina & Di Bella, 2012; Mirbagheri et al., 2015; Schmitt et al., 2018; Zuthi et al., 2017). However, pilot-scale MBRs and small datasets were used to build the models found in literature. The models built for this paper utilized high-resolution operational data over a 4-year period from a full-scale MBR municipal WWTP. The comparison of performance metrics between the models made in this study and literature can be found below in Table 4. There were no comparisons from literature for how well the RF algorithm can perform when predicting TMP from a MBR system. For the ANN-based models, the accuracies found here are slightly worse than those found in literature. It can be difficult to model the weak gradual rise in TMP over time due to biofilm formation and pore blocking if the datasets are from short experimental time periods (Hwang et al., 2008; Iorhemen et al., 2016). However, it is worth noting that model input variables, model structure, dataset size, MBR plant size/configuration, and influent characteristics are vastly different for those found in literature.

Table 4. Performance metrics for TMP models from literature

Model	Output variable(s)	Performance (Testing)	Reference	Comments/Limitations
ANN	TMP	$R^2 = 0.72$	(Liu et al., 2009)	938 observed values, very small membrane area (17 cm^2)

| Model tree (MT), Deterministic	TMP	RMSE = 2.00 kPa, 1.93 kPa	(Dalmau et al., 2015)	Dataset spanning 462 days with a 10-second resolution averaged to a daily resolution, pilot scale MBR system (2260 L, 8 m^2 of membrane area), data-driven model was not calibrated
RBFANN-GA, MLP ANN-GA	TMP & Permeability	R^2 = 0.98, 0.99	(Mirbagheri et al., 2015)	Pilot scale MBR system (8 m^2 of membrane area), dataset spans 60 days with varying HRT, optimized with a genetic algorithm (GA)
ANN	TMP	R^2 = 0.99	(Yusuf et al., 2015)	Pilot scale MBR system (20 L), entire dataset spanning less than 1 hour, 50:50 train/test split
ANN	TMP	R^2 = 0.99	(Hazrati et al., 2017)	Pilot scale MBR system (7 L), dataset spans 125 days with varying HRT
ANN	TMP	R^2 = 0.97	(Schmitt et al., 2018)	Pilot scale MBR system (75 L) operating for 180 days, trial-and-error-based hyperparameter tuning, small dataset
ANN, SVM, RF	Flux	R^2 = 0.90, 0.92, 0.95	(Li et al., 2020)	Information regarding the datasets used is unavailable, model was not calibrated

5.2   Uncertainty Analysis

The impacts of hyperparameter uncertainty on model output were analyzed using the LHS approach described in Section 3.3. In general, the ANN models were most affected by changing hyperparameters, with an R^2 ranging between 0.847 to 0.972 for the BBP model, 0.797 to 0.908

for the DBP model, and 0.862 to 0.971 for the ABP model. For the RF models, R^2 was ranged between 0.993 to 0.996 for the BBP model, 0.925 to 0.927 for the DBP model, and 0.992 to 0.996 for the ABP model. For the LSTM models, R^2 ranged between 0.160 to 0.895 for the BBP model, 0.451 to 0.861 for the DBP model, and 0.192 to 0.907 for the ABP model.

For the RF models, accuracy generally increased with increasing *ntree* and the best *mtry* values were 4 and 5. However, changing *mtry* had a bigger impact on predictive accuracy than *ntree*, and randomly selecting hyperparameters for RF did not lead to unsatisfactory results. For the ANN models, accuracy increased with increasing *size*, or number of hidden layer neurons, and a *decay* parameter closer to 0. The *decay* parameter represents a regularization tool that helps the optimization process and aims to avoid overfitting (Venables & Ripley, 2002). While the predictive accuracy increased with a smaller *decay* value, there is less variation between training and testing performance metrics shown in Table 3. It is suspected that larger datasets require smaller decay values because complex data provides regularization and other regularization methods are not as necessary for tuning (Smith, 2018). For the LSTM models, the performance metrics varied significantly across all BBP, DBP, and ABP models. Like the ANN models, a *decay* parameter closer to 0 produced a much better result, where the best-performing LSTM models had values close to 0. Models with similar *decay* values and vastly different *size* values had close to the same predictive accuracy, so *decay* has a greater impact on model accuracy.

Because ANN had the highest variation at extreme observed values between maximum and minimum predicted values, it was selected for further uncertainty analysis. All predicted values were between the maximum-minimum band which encloses the prediction interval (PI) for the model. The band graphs shown in Fig. 7 display the predictions from the 100 model iterations conducted for uncertainty analysis on the ANN models. The "Average" line represents the average

prediction value from 100 model iterations. The "Maximum" and "Minimum" bands show the highest and lowest predicted value for each observed value from the 100 timeseries, respectively. As shown in Fig. 7, the maximum and minimum bands had more severe jumps in TMP at drastic changes in TMP, exacerbating the behaviour shown in the observed TMP values and exhibited the model's inaccuracy in predicting extreme values. The largest difference in R^2 between the average prediction and observation is shown with the BBP model in Fig. 7 (a). The timeseries uncertainty band graphs for the RF and LSTM models can be found in the Appendix as Figures A7 – A9 and Figures A10 – A12, respectively.
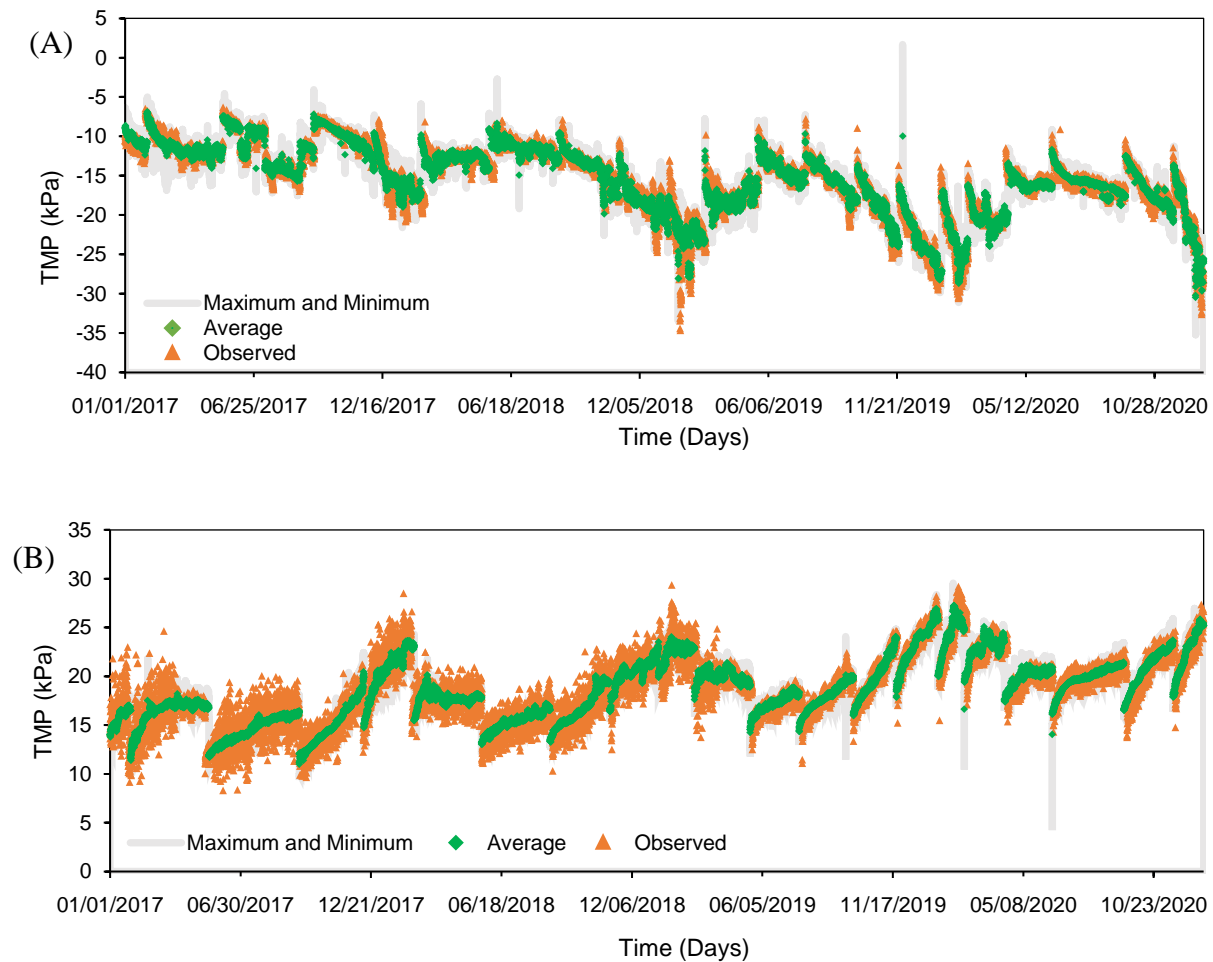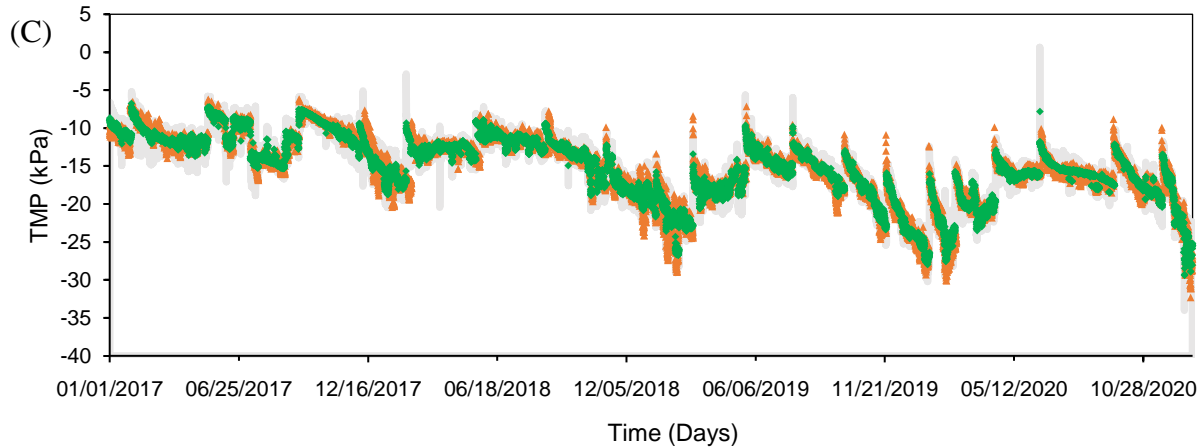
Fig. 7. Timeseries uncertainty band graphs displaying the average, maximum, and minimum values from 100 iterations for: (a) BBP ANN model, (b) DBP ANN model, (c) ABP ANN model

To further analyze the uncertainty due to model structure, the most extreme values in the observed testing dataset (maximum and minimum TMP) were investigated. Table 5 shows a summary of the statistical measures associated to the predictions of the extreme maximum and minimum observed values. The ranges of prediction for the RF models were significantly smaller than the ranges shown by the ANN and LSTM models. Further, the RF models had mean prediction values closest to the minimum or maximum observed value in most cases.

Fig. 8 represents the 100 predicted TMP values and the corresponding TMP range it fits into. For instance, Fig. 8 (a) represents the predictions of the minimum value in the BBP testing dataset and the associated predictions from the 100 iterations of each model – this resulted in 300 predicted values from ANN, RF, and LSTM BBP models. The observed minimum value is -34.7 kPa. Thus, all 100 predictions from the RF model iterations lie between the range of -36 kPa to -32 kPa, with an average of -33.8 kPa. However, the 100 predictions from the ANN models lie between the range of -30 kPa to -20 kPa, with an average of -25.5 kPa. The predictions from the LSTM models lie between the range of -24 kPa to -15 kPa with an average of -19.3 kPa. The

uncertainties shown within the RF models were significantly less than the uncertainties shown within both ANN and LSTM models within their respective defined hyperparameter ranges. The RF model showed low variance with low bias while the ANN model showed high variance with high bias. Most LSTM predictions were higher than -20 kPa, which is much further than the observed minimum value of -34.7 kPa. Likewise, Fig. 8 (b) shows 300 values for the maximum value in the BBP testing dataset, with an observed maximum value of -6.3 kPa. All 100 predictions from the RF model iterations lie between the range of -8.5 kPa to -7.5 kPa, with an average of -8.1 kPa. In comparison, the 100 predictions from the ANN model iterations lie between the range of -6.0 kPa to -9.0 kPa, with an average of -7.5 kPa. The LSTM predictions lie between the range of -11.6 kPa and -6.5 kPa with an average of -9.4 kPa. The RF model exhibited behaviour that is consistent with low variance and high bias, whereas the ANN model exhibited behaviour more consistent with high variance and high bias. In this case, the ANN predicted closer to the test value than the RF. The LSTM model predicted the least accurately out of the three algorithm choices within their respective defined hyperparameter ranges.

For the DBP testing dataset, the observed minimum and maximum TMP values were 8.34 kPa and 29.38 kPa, respectively. The average predictions for the observed minimum value for the RF, ANN, and LSTM models are 14.1 kPa, 13.5 kPa, and 14.1 kPa, respectively. The average predictions for the observed maximum value for the RF, ANN, and LSTM models are 26.5 kPa, 22.2 kPa, and 20.7 kPa, respectively. As shown in Fig. 8 (c) and Fig. 8 (d), the ANN model predicted the minimum value more accurately, whereas the RF predicted the maximum value more accurately with greater frequency. The LSTM is performing more competitively for DBP TMP prediction when compared to the RF and ANN models but is still performing the least accurately.

For the ABP testing dataset, the observed minimum and maximum TMP values are -32.2 kPa and -6.2 kPa, respectively. The average predictions for the observed minimum value for the RF, ANN, and LSTM models are -31.3 kPa, -29.0 kPa, and -24.5 kPa, respectively. The average predictions for the observed maximum value for the RF, ANN, and LSTM models are -7.3 kPa, -7.8 kPa, and -10.3 kPa, respectively. As shown in Fig. 8 (e) and Fig. 8 (f), the RF model predicted both minimum and maximum values more accurately when compared to the ANN and LSTM models. Through BBP/DBP/ABP model creation, it is shown that the LSTM algorithm is less sensitive to drastic changes in observed TMP values and cannot recognize extreme predictions effectively.
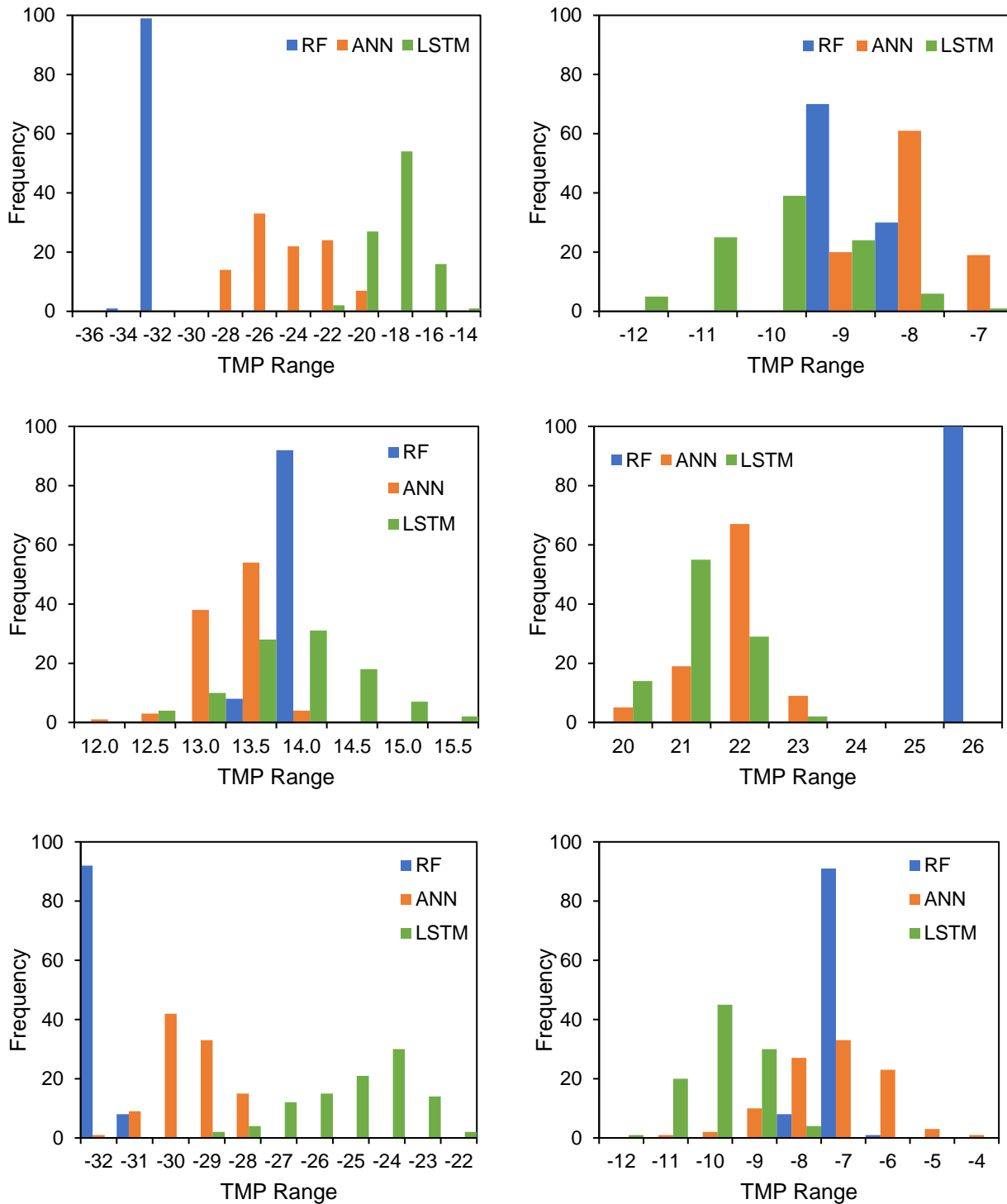
Fig. 8. Histograms depicting frequency for the most extreme test dataset observations: (a) BBP RF, ANN, LSTM for min-value, (b) BBP RF, ANN, LSTM for max-value, (c) DBP RF, ANN, LSTM for min-value, (d) DBP RF, ANN, LSTM for max-value, (e) ABP RF, ANN, LSTM for min-value, (f) ABP RF, ANN, LSTM for max-value

Table 5. Summary of model structure uncertainty for extreme value prediction

| | BBP | | | | | | DBP | | | | | | ABP | | | | | |
| | Min. value | | | Max. value | | | Min. value | | | Max. value | | | Min. value | | | Max. value | | |
	RF	ANN	LSTM	RF	ANN	LSTM	RF	ANN	LSTM	RF	ANN	LSTM	RF	ANN	LSTM	RF	ANN	LSTM
Minimum	-34.0	-29.9	-22.8	-8.4	-8.9	-11.6	14.0	12.5	12.7	26.1	20.2	19.1	-31.6	-31.2	-28.9	-8.2	-11.2	-12.2
Maximum	-32.8	-20.4	-15.1	-7.7	-6.4	-6.5	14.2	14.5	15.6	26.8	23.6	22.4	-30.1	-27.3	-21.4	-7.0	-4.7	-8.8
Mean	-33.8	-25.5	-19.3	-8.1	-7.5	-9.4	14.1	13.5	14.1	26.5	22.2	20.7	-31.3	-29.0	-24.5	-7.3	-7.8	-10.3
Median	-33.9	-25.5	-19.2	-8.1	-7.4	-9.5	14.1	13.6	14.1	26.5	22.3	20.7	-31.4	-29.0	-24.3	-7.1	-7.6	-10.3
Standard Deviation	0.3	2.2	1.3	0.2	0.6	1.0	0.0	0.3	0.6	0.2	0.6	0.6	0.3	0.8	1.5	0.3	1.2	0.8
Range	1.22	9.47	7.68	0.8	2.51	5.13	0.23	1.99	2.91	0.67	3.39	3.28	1.58	3.91	7.45	1.23	6.52	3.47
Test value	-34.67 kPa			-6.31 kPa			8.34 kPa			29.38 kPa			-32.17 kPa			-6.22 kPa		

Chapter 6 – Conclusions

In this research, data-driven models utilizing RF, ANN, and LSTM algorithms were developed to characterize membrane fouling through the prediction of TMP at various points of the MBR production cycle. These models were developed utilizing 4 years of high-resolution operational data from a full-scale municipal WWTP. After a careful input variable selection process, the models were successfully applied to predict BBP, DBP, and ABP TMP showing excellent predictive accuracy. In addition, the uncertainty analysis conducted provides PIs for hyperparameter uncertainty and insight into model structure uncertainty associated to extreme predictions. These proposed models can provide decision support to WWTP operators to assist in the reduction of membrane fouling.

The RF models perform the best in terms of the statistical measures R^2, MAPE, RMSE, and MSE when compared to both ANN and LSTM models. The R^2 values for the testing of the BBP, DBP, and ABP RF models were 0.996, 0.927, and 0.996, respectively; the values for the testing of the ANN models were 0.973, 0.910, and 0.974, respectively; and the values for the testing of the LSTM models were 0.932, 0.883, and 0.935, respectively. The results of the model testing show that these models can be a useful tool when predicting TMP, the RF models being the most robust and reliable within the selected hyperparameter range. When compared to the RF and ANN models, the LSTM models had the highest degree of overall hyperparameter and model structure uncertainty. However, the ANN had the highest degree of uncertainty with extreme value prediction. The ANN models had difficulty maintaining consistency when predicting extreme TMP values, whereas the LSTM models did not recognize extreme observed TMP values.

In future studies, these models can be applied to other WWTPs with MBR systems. Additionally, variable importance can be derived from data-driven models to obtain information

on the model input variables that have the greatest impact on membrane fouling. For example, the RF algorithm can utilize out-of-bag MSE to determine the change in error if a variable is to be removed. Further, partial dependence plots can be used to determine the approximate change in the output variable based on a change in an input variable. The predictive models presented in this study provide promising results, however, they are limited to predicting TMP based on data at a present point in time. Implementing models to forecast TMP could provide better decision-support for operators to adjust for membrane fouling before it occurs.

References

Alvarez, J. M., & Salzmann, M. (2016). Learning the number of neurons in deep networks. *Advances in Neural Information Processing Systems*, *Nips*, 2270–2278.

Amit, Y., & Geman, D. (1997). Shape Quantization and Recognition with Randomized Trees. *Neural Computation*, *9*(7), 1545–1588. https://doi.org/10.1162/NECO.1997.9.7.1545

Angelakis, A. N., Asano, T., Bahri, A., Jimenez, B. E., & Tchobanoglous, G. (2018). Water reuse: From ancient to modern times and the future. *Frontiers in Environmental Science*, *6*(MAY). https://doi.org/10.3389/fenvs.2018.00026

Bayram, S., Ocal, M. E., Laptali Oral, E., & Atis, C. D. (2016). Comparison of multi layer perceptron (MLP) and radial basis function (RBF) for construction cost estimation: the case of Turkey. *Journal of Civil Engineering and Management*, *22*(4), 480–490. https://doi.org/10.3846/13923730.2014.897988

Belfort, G., Davis, R. H., & Zydney, A. L. (1994). The behaviour of suspentions and macromolecular solutionsfiltration. *Journal of Membrane Science*, *96*(1–2), 1–58.

Beril Gönder, Z., Arayici, S., & Barlas, H. (2011). Advanced treatment of pulp and paper mill wastewater by nanofiltration process: Effects of operating conditions on membrane fouling. *Separation and Purification Technology*, *76*(3), 292–302. https://doi.org/10.1016/J.SEPPUR.2010.10.018

Berman, A. S. (1953). Laminar flow in channels with porous walls. *Journal of Applied Physics*, *24*(9), 1232–1235. https://doi.org/10.1063/1.1721476

Bouhabila, E. H., Ben Aïm, R., & Buisson, H. (2001). Fouling characterisation in membrane

bioreactors. *Separation and Purification Technology*, *22–23*, 123–132. https://doi.org/10.1016/S1383-5866(00)00156-8

Boyd, G., Na, D., Li, Z., Snowling, S., Zhang, Q., & Zhou, P. (2019). Influent forecasting for wastewater treatment plants in North America. *Sustainability (Switzerland)*, *11*(6), 1–14. https://doi.org/10.3390/su11061764

Brauns, E., Van Hoof, E., Molenberghs, B., Dotremont, C., Doyen, W., & Leysen, R. (2002). A new method of measuring and presenting the membrane fouling potential. *Desalination*, *150*(1), 31–43. https://doi.org/10.1016/S0011-9164(02)00927-X

Breiman, L. (2001). Random Forests. *Machine Learning 2001 45:1*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L. C. L. W. (2018). *Package "randomForest": Breiman and Cutler's Random Forests for Classification and Regression*. https://doi.org/10.1023/A:1010933404324

Carnell, R. (2021). *lhs: Latin Hypercube Samples. R package version 1.1.3*. https://cran.r-project.org/package=lhs

Chang, I.-S., Le Clech, P., Jefferson, B., & Judd, S. (2002). Membrane Fouling in Membrane Bioreactors for Wastewater Treatment. *Journal of Environmental Engineering*, *128*(11), 1018–1029. https://doi.org/10.1061/(asce)0733-9372(2002)128:11(1018)

Chang, I. S., Field, R., & Cui, Z. (2009). Limitations of resistance-in-series model for fouling analysis in membrane bioreactors: A cautionary note. *Desalination and Water Treatment*, *8*(1–3), 31–36. https://doi.org/10.5004/dwt.2009.687

Dalmau, M., Atanasova, N., Gabarrón, S., Rodriguez-Roda, I., & Comas, J. (2015). Comparison

of a deterministic and a data driven model to describe MBR fouling. *Chemical Engineering Journal*, *260*, 300–308. https://doi.org/10.1016/j.cej.2014.09.003

Dietterich, T. G. (2000). Experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Machine Learning*, *40*(2), 139–157. https://doi.org/10.1023/A:1007607513941

Dornier, M., Decloux, M., Trystram, G., & Lebert, A. (1995). Dynamic modeling of crossflow microfiltration using neural networks. *Journal of Membrane Science*, *98*(3), 263–273. https://doi.org/10.1016/0376-7388(94)00195-5

Drews, A. (2010). Membrane fouling in membrane bioreactors-Characterisation, contradictions, cause and cures. *Journal of Membrane Science*, *363*(1–2), 1–28. https://doi.org/10.1016/j.memsci.2010.06.046

Farhi, N., Kohen, E., Mamane, H., & Shavitt, Y. (2021). Prediction of wastewater treatment quality using LSTM neural network. *Environmental Technology and Innovation*, *23*, 101632. https://doi.org/10.1016/j.eti.2021.101632

Field, R. ., Wu, D., Howell, J. ., & Gupta, B. . (1995). Critical flux concept for microfiltration fouling. *Journal of Membrane Science*, 259–272.

Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *American Statistician*, *63*(4), 308–319. https://doi.org/10.1198/tast.2009.08199

Guo, W., Ngo, H. H., & Li, J. (2012). A mini-review on membrane fouling. *Bioresource Technology*, *122*, 27–34. https://doi.org/10.1016/j.biotech.2012.04.089

Hai, F. I., & Yamamoto, K. (2011). Membrane Biological Reactors. *Treatise on Water Science*, *4*, 571–613. https://doi.org/10.1016/B978-0-444-53199-5.00096-8

Hamedi, H., Ehteshami, M., Mirbagheri, S. A., Rasouli, S. A., & Zendehboudi, S. (2019). Current Status and Future Prospects of Membrane Bioreactors (MBRs) and Fouling Phenomena: A Systematic Review. *Canadian Journal of Chemical Engineering*, *97*(1), 32–58. https://doi.org/10.1002/cjce.23345

Hazrati, H., Moghaddam, A. H., & Rostamizadeh, M. (2017). The influence of hydraulic retention time on cake layer specifications in the membrane bioreactor: Experimental and artificial neural network modeling. *Journal of Environmental Chemical Engineering*, *5*(3), 3005–3013. https://doi.org/10.1016/j.jece.2017.05.050

Hermanowicz, S. W., Jenkins, D., Merlo, R. P., & Trussell, R. S. (2006). *Effects of Biomass Properties on Submerged Membrane Bioreactor (SMBR) Performance and Solids Processing*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1007/978-1-4757-5388-2_2

Hong, K., Lee, S., Choi, S., Yu, Y., Hong, S., Moon, J., Sohn, J., & Yang, J. (2009). Assessment of various membrane fouling indexes under seawater conditions. *Desalination*, *247*(1–3), 247–259. https://doi.org/10.1016/j.desal.2008.12.029

Hwang, B. K., Lee, W. N., Yeon, K. M., Park, P. K., Lee, C. H., Chang, I. S., Drews, A., & Kraume, M. (2008). Correlating TMP increases with microbial characteristics in the bio-cake on the membrane surface in a membrane bioreactor. *Environmental Science and Technology*, *42*(11), 3963–3968. https://doi.org/10.1021/es7029784

Iorhemen, O. T., Hamza, R. A., & Tay, J. H. (2016). Membrane bioreactor (Mbr) technology for wastewater treatment and reclamation: Membrane fouling. *Membranes*, *6*(2), 13–16. https://doi.org/10.3390/membranes6020033

Jarusutthirak, C., & Amy, G. (2006). Role of soluble microbial products (SMP) in membrane fouling and flux decline. *Environmental Science and Technology*, *40*(3), 969–974. https://doi.org/10.1021/es050987a

Kuhn, M. (2021). *caret: Classification and Regression Training*. https://cran.r-project.org/package=caret

Li, Q., & Elimelech, M. (2004). Organic fouling and chemical cleaning of nanofiltration membranes: Measurements and mechanisms. *Environmental Science and Technology*, *38*(17), 4683–4693. https://doi.org/10.1021/es0354162

Li, W., Li, C., & Wang, T. (2020). Application of machine learning algorithms in mbr simulation under big data platform. *Water Practice and Technology*, *15*(4), 1238–1247. https://doi.org/10.2166/wpt.2020.095

Li, X., Li, Z., Huang, W., & Zhou, P. (2020). Performance of statistical and machine learning ensembles for daily temperature downscaling. *Theoretical and Applied Climatology*, *140*(1–2), 571–588. https://doi.org/10.1007/s00704-020-03098-3

Liang, S., Song, L., Tao, G., Kekre, K. A., & Seah, H. (2006). A Modeling Study of Fouling Development in Membrane Bioreactors for Wastewater Treatment. *Water Environment Research*, *78*(8), 857–864. https://doi.org/10.2175/106143005x73028

Liu, Q. F., Kim, S. H., & Lee, S. (2009). Prediction of microfiltration membrane fouling using

artificial neural network models. *Separation and Purification Technology*, *70*(1), 96–102. https://doi.org/10.1016/j.seppur.2009.08.017

Liu, R., Huang, X., Sun, Y. F., & Qian, Y. (2003). Hydrodynamic effect on sludge accumulation over membrane surfaces in a submerged membrane bioreactor. *Process Biochemistry*, *39*(2), 157–163. https://doi.org/10.1016/S0032-9592(03)00022-0

Ludwig, T., Gaida, D., Keysers, C., Pinnekamp, J., Bongards, M., Kern, P., Wolf, C., & Sousa Brito, A. L. (2012). An advanced simulation model for membrane bioreactors: Development, calibration and validation. *Water Science and Technology*, *66*(7), 1384–1391. https://doi.org/10.2166/wst.2012.249

Mannina, G., & Di Bella, G. (2012). Comparing two start-up strategies for MBRs: Experimental study and mathematical modelling. *Biochemical Engineering Journal*, *68*, 91–103. https://doi.org/10.1016/j.bej.2012.07.011

Melin, T., Jefferson, B., Bixio, D., Thoeye, C., De Wilde, W., De Koning, J., van der Graaf, J., & Wintgens, T. (2006). Membrane bioreactor technology for wastewater treatment and reuse. *Desalination*, *187*(1–3), 271–282. https://doi.org/10.1016/j.desal.2005.04.086

Meng, F., Chae, S. R., Drews, A., Kraume, M., Shin, H. S., & Yang, F. (2009). Recent advances in membrane bioreactors (MBRs): Membrane fouling and membrane material. *Water Research*, *43*(6), 1489–1512. https://doi.org/10.1016/j.watres.2008.12.044

Meng, F., & Yang, F. (2007). Fouling mechanisms of deflocculated sludge, normal sludge, and bulking sludge in membrane bioreactor. *Journal of Membrane Science*, *305*(1–2), 48–56. https://doi.org/10.1016/j.memsci.2007.07.038

Mirbagheri, S. A., Bagheri, M., Bagheri, Z., & Kamarkhani, A. M. (2015). Evaluation and prediction of membrane fouling in a submerged membrane bioreactor with simultaneous upward and downward aeration using artificial neural network-genetic algorithm. *Process Safety and Environmental Protection*, *96*, 111–124. https://doi.org/10.1016/j.psep.2015.03.015

Nourani, V., Khodkar, K., Paknezhad, N. J., & Laux, P. (2022). Deep learning-based uncertainty quantification of groundwater level predictions. *Stochastic Environmental Research and Risk Assessment*, *7*. https://doi.org/10.1007/s00477-022-02181-7

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., & Gimelshein, N. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 8024--8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3). https://doi.org/10.1002/widm.1301

Schmitt, F., Banu, R., Yeom, I. T., & Do, K. U. (2018). Development of artificial neural networks to predict membrane fouling in an anoxic-aerobic membrane bioreactor treating domestic wastewater. *Biochemical Engineering Journal*, *133*, 47–58. https://doi.org/10.1016/j.bej.2018.02.001

Segal, M. R. (2003). *Machine Learning Benchmarks and Random Forest Regression*.

Shi, Y., Wang, Z., Du, X., Gong, B., Jegatheesan, V., & Haq, I. U. (2021). Recent advances in the prediction of fouling in membrane bioreactors. *Membranes*, *11*(6).

https://doi.org/10.3390/membranes11060381

Smith, L. N. (2018). *A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay*. 1–21. http://arxiv.org/abs/1803.09820

Sofia, A., Ng, W. J., & Ong, S. L. (2004). Engineering design approaches for minimum fouling in submerged MBR. *Desalination*, *160*(1), 67–74. https://doi.org/10.1016/S0011-9164(04)90018-5

Stein, M. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, *29*(2), 143–151. https://doi.org/10.1080/00401706.1987.10488205

Stephens, G. L., Slingo, J. M., Rignot, E., Reager, J. T., Hakuba, M. Z., Durack, P. J., Worden, J., & Rocca, R. (2020). Earth's water reservoirs in a changing climate. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *476*(2236). https://doi.org/10.1098/rspa.2019.0458

Stephenson, T., Judd, S., Jefferson, B., & Brindle, K. (2000). Membrane bioreactors for wastewater treatment. *IWA Publishing*.

Terrill, R. M., & Thomas, P. W. (1969). *On laminar flow through a uniformly porous pipe*. 21–67.

Tsuyuhara, T., Hanamoto, Y., Miyoshi, T., Kimura, K., & Watanabe, Y. (2010). Influence of membrane properties on physically reversible and irreversible fouling in membrane bioreactors. *Water Science and Technology*, *61*(9), 2235–2240. https://doi.org/10.2166/wst.2010.122

United Nations. (2020a). The Sustainable Development Goals Report 2020. *United Nations Publication Issued by the Department of Economic and Social Affairs*, 1–64.

United Nations. (2020b). *Water and Climate Change*. https://www.unwater.org/water-facts/climate-change/

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*.

Wang, D., Thunéll, S., Lindberg, U., Jiang, L., Trygg, J., Tysklind, M., & Souihi, N. (2021). A machine learning framework to improve effluent quality control in wastewater treatment plants. *Science of the Total Environment*, *784*, 147138. https://doi.org/10.1016/j.scitotenv.2021.147138

Xu, Y., Hu, C., Wu, Q., Jian, S., Li, Z., Chen, Y., Zhang, G., Zhang, Z., & Wang, S. (2022). Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. *Journal of Hydrology*, *608*(August 2020), 127553. https://doi.org/10.1016/j.jhydrol.2022.127553

Yu, Y., Xiaosheng, S., Changhua, H., & Jianxun, Z. (2019). A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*, *31*(7), 1235–1270. https://doi.org/10.1162/NECO

Yusuf, Z., Wahab, N. A., & Sahlan, S. (2015). Modeling of submerged membrane bioreactor filtration process using NARX-ANFIS model. *2015 10th Asian Control Conference: Emerging Control Techniques for a Sustainable World, ASCC 2015*. https://doi.org/10.1109/ASCC.2015.7244710

Zhang, G., Wang, C., Xu, B., & Grosse, R. (2019). Three mechanisms of weight decay

regularization. *7th International Conference on Learning Representations, ICLR 2019*, 1–16.

Zhang, Q., Li, Z., Snowling, S., Siam, A., & El-Dakhakhni, W. (2019). Predictive models for wastewater flow forecasting based on time series analysis and artificial neural network. *Water Science and Technology*, *80*(2), 243–253. https://doi.org/10.2166/wst.2019.263

Zhou, P., Li, Z., Snowling, S., Baetz, B. W., Na, D., & Boyd, G. (2019). A random forest model for inflow prediction at wastewater treatment plants. *Stochastic Environmental Research and Risk Assessment*, *33*(10), 1781–1792. https://doi.org/10.1007/s00477-019-01732-9

Zuthi, M. F. R., Guo, W., Ngo, H. H., Nghiem, D. L., Hai, F. I., Xia, S., Li, J., Li, J., & Liu, Y. (2017). New and practical mathematical model of membrane fouling in an aerobic submerged membrane bioreactor. *Bioresource Technology*, *238*, 86–94. https://doi.org/10.1016/j.biortech.2017.04.006

Appendix
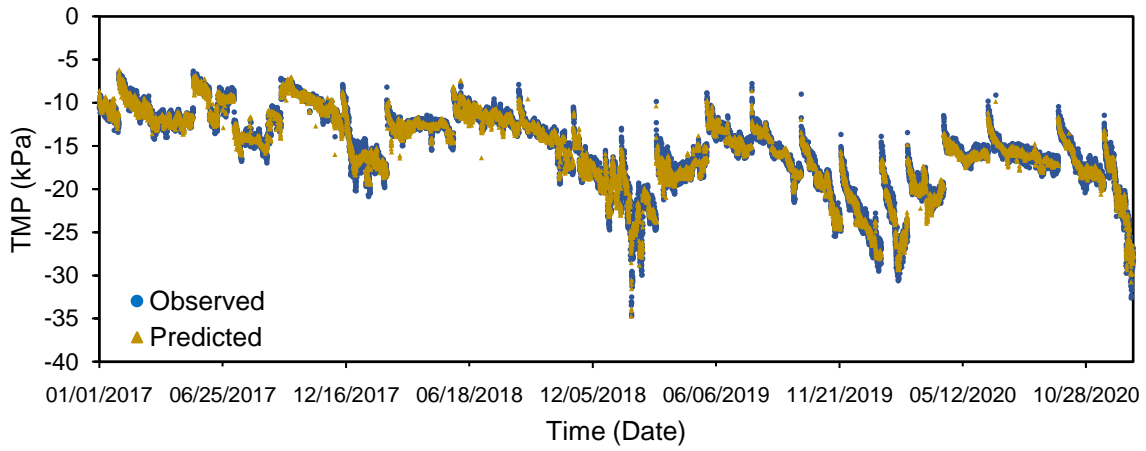
Figure A1. Timeseries for optimal BBP ANN Model
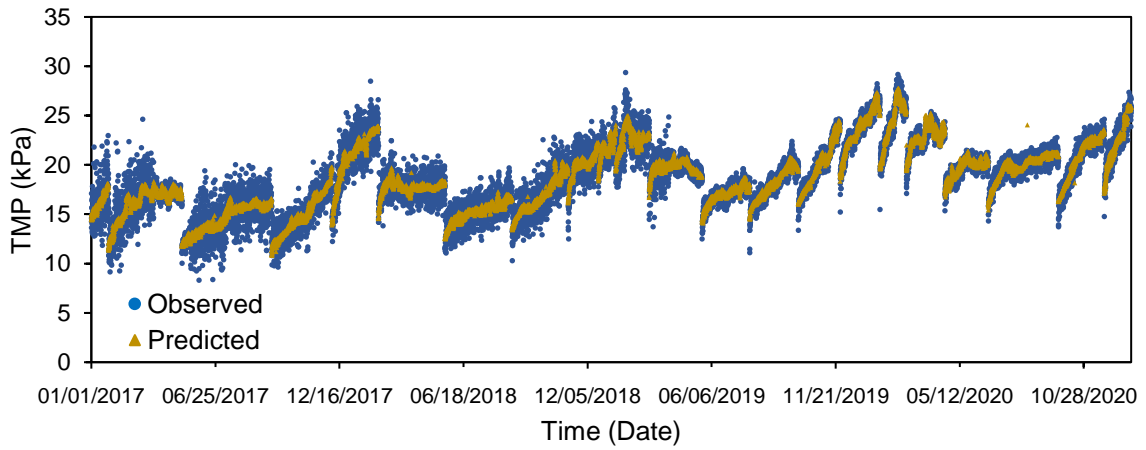
Figure A2. Timeseries for optimal DBP ANN Model
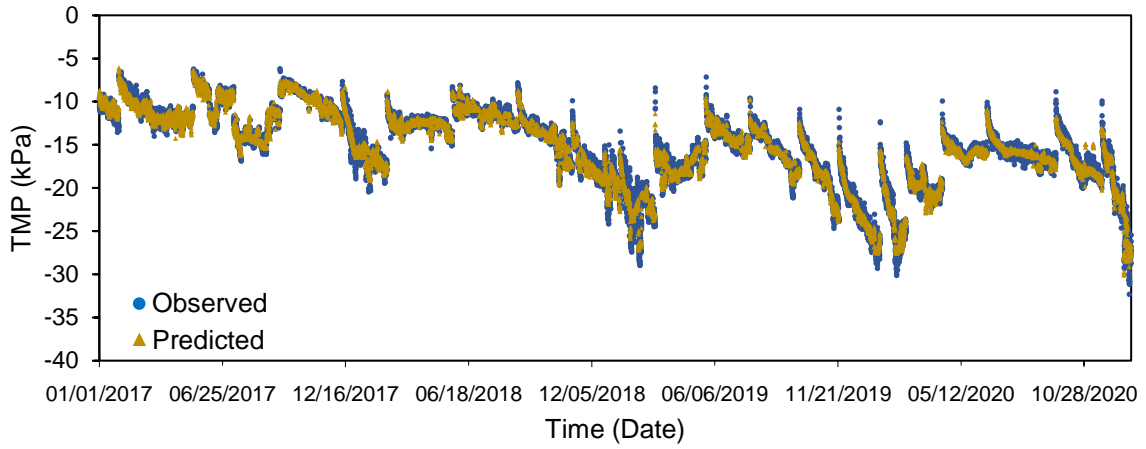
Figure A3. Timeseries for optimal ABP ANN Model
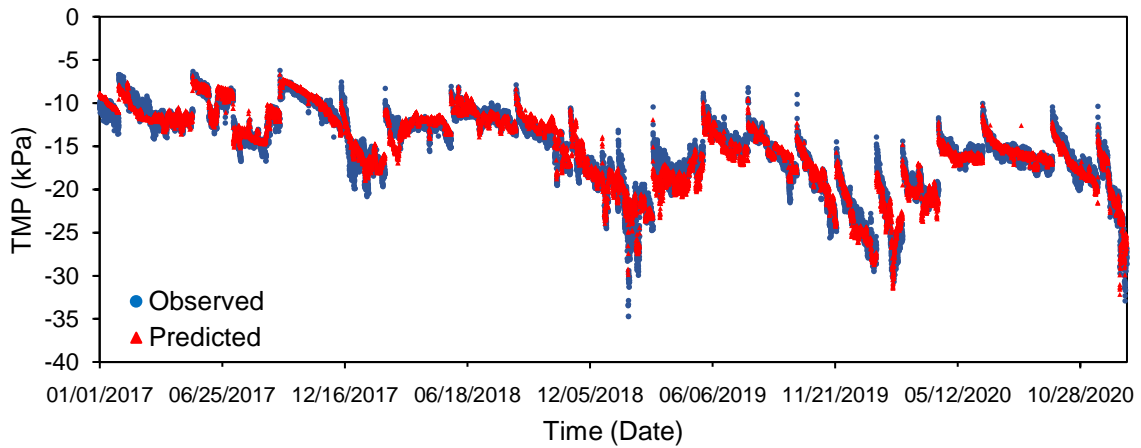
Figure A4. Timeseries for optimal BBP LSTM Model
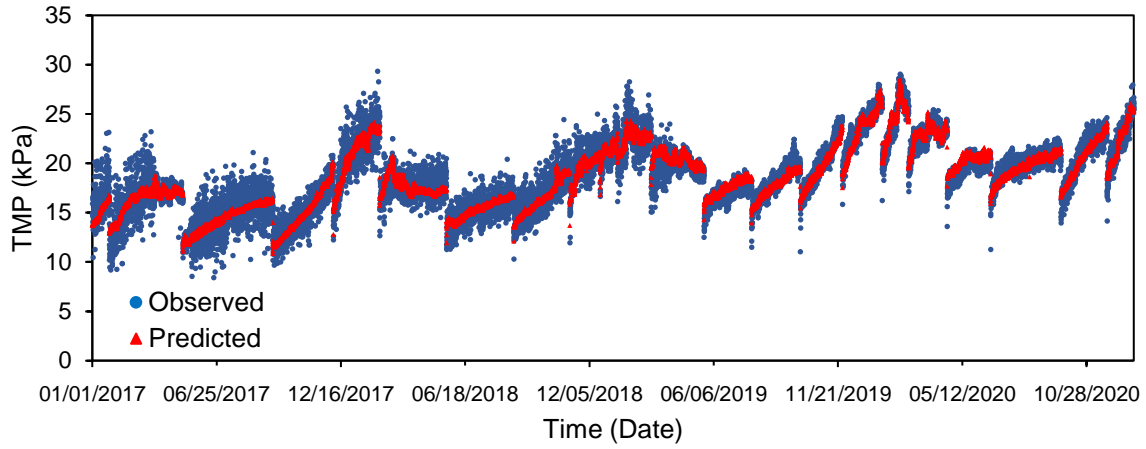
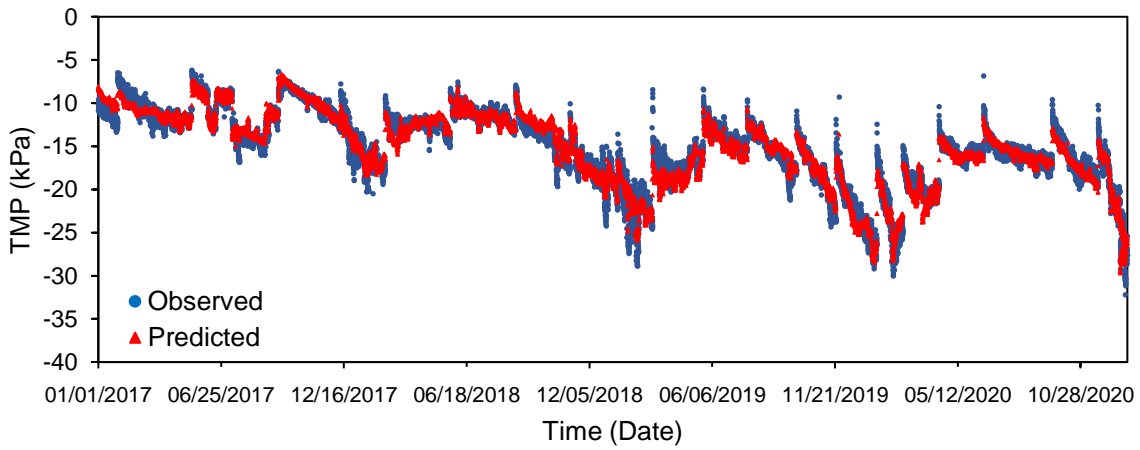Figure A5. Timeseries for optimal DBP LSTM Model
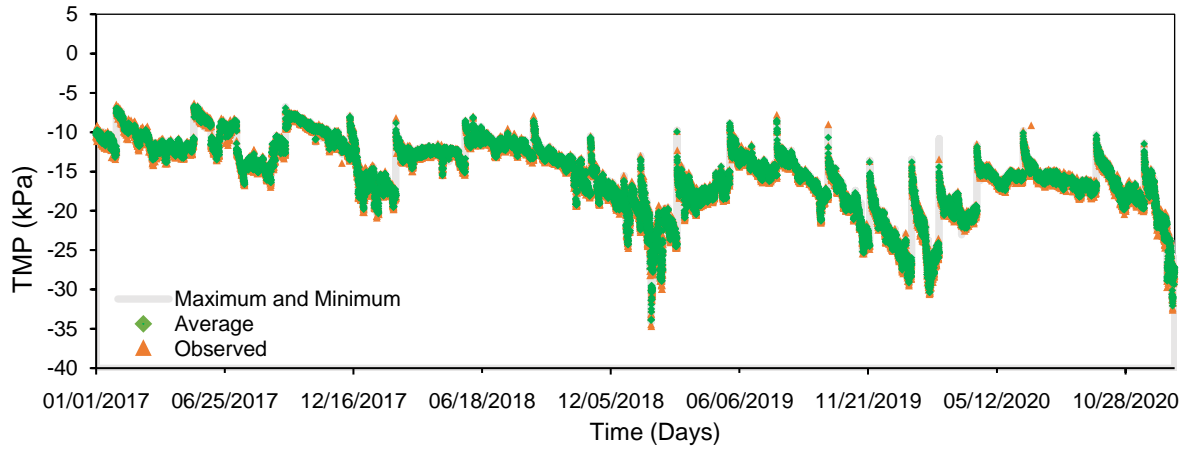

Figure A6. Timeseries for optimal ABP LSTM Model

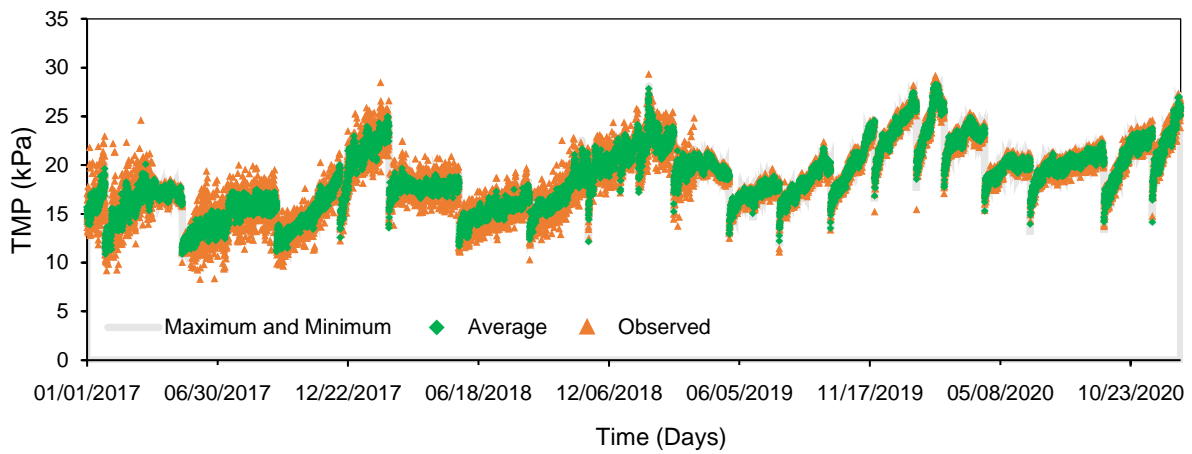Figure A7. Timeseries uncertainty band graph for BBP RF Model

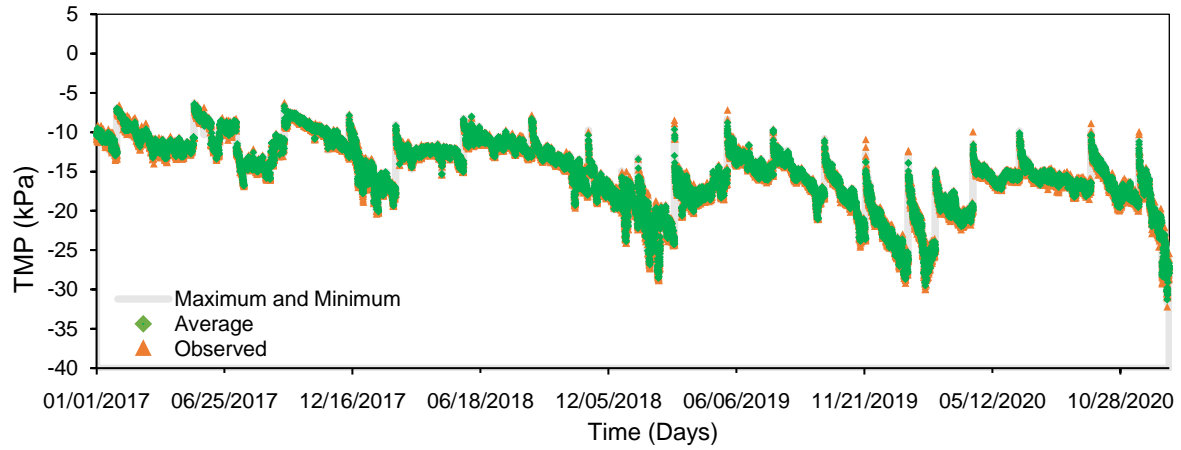Figure A8. Timeseries uncertainty band graph for DBP RF Model

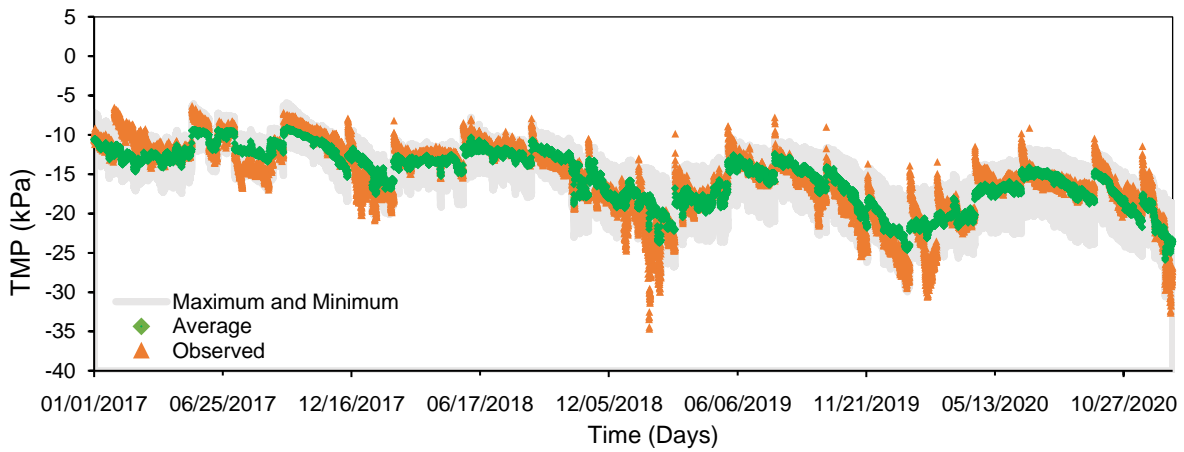Figure A9. Timeseries uncertainty band graph for ABP RF Model

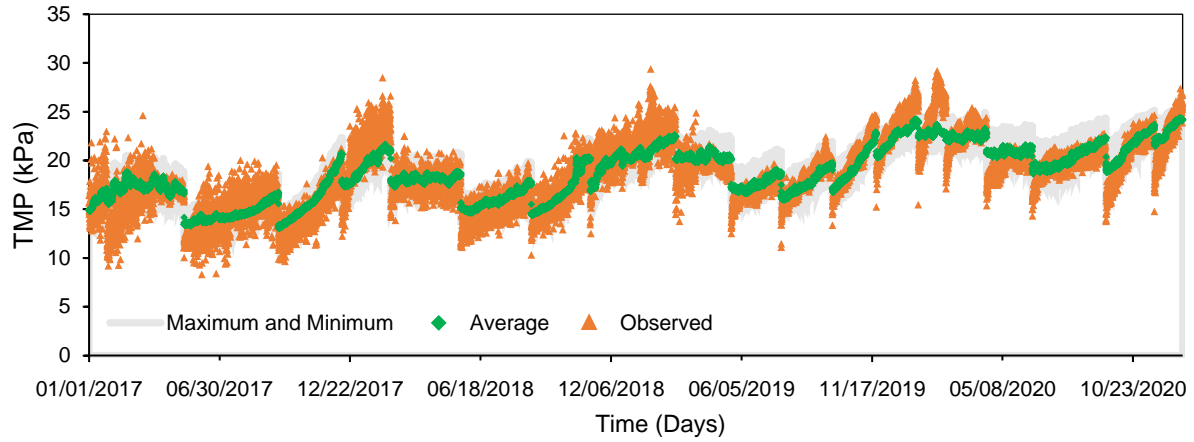Figure A10. Timeseries uncertainty band graph for BBP LSTM Model

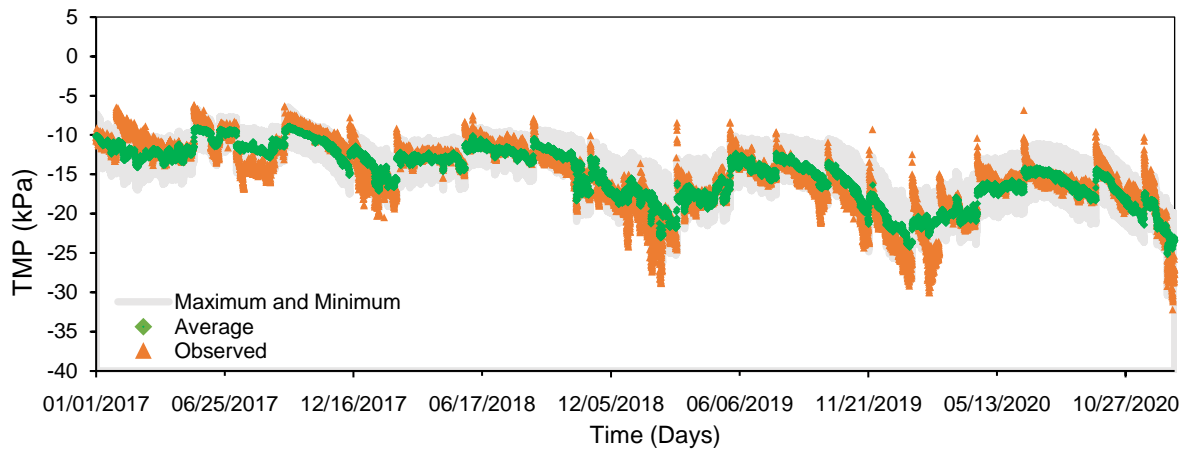Figure A11. Timeseries uncertainty band graph for DBP LSTM Model

Figure A12. Timeseries uncertainty band graph for ABP LSTM Model