

**IDENTIFY POTENTIAL BMP TOOLS TO REDUCE BACTERIA LOADING IN
NECHES RIVER**

Final Report

GLO Contract No. 23-020-007-D601

September 2024

Prepared by

Qin Qian, Principal Investigator, Lamar University

Yu Zhang, Co-Principal Investigator, UTA

Anish R. Jantrania, Co- Principal Investigator, Texas A&M University

Submitted to:

Texas General Land Office

1700 Congress Ave. Austin, TX 78701-1495

This report was funded in part by a Texas Coastal Management Program grant approved by the Texas Land Commissioner, providing financial assistance under the Coastal Zone Management Act of 1972, as amended, awarded by the National Oceanic and Atmospheric Administration (NOAA), Office for Coastal Management, pursuant to NOAA Award No. NA22NOS4190148. The views expressed herein are those of the author(s) and do not necessarily reflect the views of NOAA, the U.S. Department of Commerce, or any of their subagencies



TABLE OF CONTENTS

Chapter I Introduction.....	1
1 Background Information.....	1
2 Study Areas	2
Chapter II Methods	7
1 Field Water Sampling.....	7
2 Machine Learning Models with YSI Data	10
2.2 Machine Learning Model to Predict Dissolved Oxygen (DO).....	12
2.3 Machine Learning Model to Predict Turbidity	14
3 SWMM & MOPUS Model	15
Chapter III Results and Discussion.....	19
1 Results of Field Water Sampling Measurements	19
2 Results of Machine Learning (ML) Models	25
2.1 Results of ML DO Models.....	26
2.1.1. Current Hourly DO Prediction	26
2.1.2. 14-day Hourly DO Forecast	28
2.2 Results of ML Turbidity Model	30
3 Results of SWMM & MOPUS model	33
Chapter IV Summary	40

EXECUTIVE SUMMARY

The goal of this study is to determine the origin and transport pathways of the fecal bacteria in the region and assess BMPs as countermeasures to alleviate the excess loading. The outcome of the project is to reduce bacteria loading, and improved water quality. Cooperated the available OSSF data with local agencies suggestion, we conducted a field survey to identify 14 critical water sampling locations. Five runs of water sampling and water quality measurements demonstrate the elevated nutrient and *E.coli* were observed in the system. The bacteria source tracking showed that the Human qPCR marker at sites 5, 7, 13 were higher than the criteria, which indicated the potential contribution of human fecal contaminants from the OSSFs. The elevated enterococci at five TCEQ monitoring sites response to the nonpoint source we found in the drainage system accordingly. Therefore, the failing OSSFs is one of the bacteria pollution origins. We suggest that fix OSSFs at site 5, 7 and 13 is one of the BMPs to decrease the bacteria.

A coupled EPA-SWMM/MOPUS system was established to help assess the potential of infiltration BMPs, specifically bioretention cells, in alleviating the bacteria loading problems if installed in the upstream of watersheds with high concentration of OSSFs. Using a default configuration of bioretention cell leads to 1-88% reductions in peak simulated runoff, and 0-26% reductions in peak bacteria loading. Sensitivity experiments suggest that increasing capacity of the BMP leads to larger reductions in peak runoff but relatively modest reductions in bacteria loading.

LIST OF FIGURES

Figure 1 . Study area with OSSF inventory map as well as updated parcels and wastewater outfalls.....	4
Figure 2 . Map of 100-year floodplain, SSURGO soil maps and OSSF inventory.	5
Figure 3 . Map of seven areas shown in the insert in top right with purple boundary and 14 sampling points for water quality sampling analysis.....	6
Figure 4 . Sampling schedule based on rain fall events.....	8
Figure 5 . Sample of "Wastewater Sampling Form" on 07/06/2023.....	9
Figure 6 . Locations of YSI water quality sensors and USGS stations.....	11
Figure 7 . The framework of the DO prediction model	13
Figure 8 . The sequential-to-sequential forecasting framework via a temporal attention-based encoder & decoder model.	16
Figure 9 . Locations where hypothetical BMPs (bioretention cells) are implemented in SWMM.	17
Figure 10 . Schematic of coupled modeling system comprised of SWMM and MOPUS.	18
Figure 11 . Configuration of a bioretention cell.....	18
Figure 12 . Comparison of models for current hourly DO predictions.....	27
Figure 13 . Comparison of models for 14-day hourly DO forecasts.	30
Figure 14 . Visualization of the next time-step (1 hour) prediction of the test set.....	32

Figure 15 . Visualization of 5-day forecasting results of the training set, valid set, and test set.	32
Figure 16 . Watershed delineated using HEC-HMS (left) and after manual editing (right).	33
Figure 17 . Schematization of watersheds in EPA-SWMM for sites 12-14.	34
Figure 18 . Example time series of rainfall from the CocoHaRS station and runoff simulations by EPA-SWMM.	35
Figure 19 . MOPUS Model with calibrated coefficients based on observed bacterial concentrations.	36

LIST OF TABLES

Table 1 . Lab and field measurements on 05/19/2023.	19
Table 2 . Lab and field measurements on 07/06/2023.	20
Table 3 . Lab and field measurements on 12/06/2023.	21
Table 4 . Lab and field measurements on 01/30/2024.	22
Table 5 . Lab and field measurements on 03/22/2024.	23
Table 6 . qPCR analysis at different sites.....	24
Table 7 . Performance of the DO models for 1-hour prediction task.....	28
Table 8 . Performance of the DO models for 14-day hourly forecast.....	29
Table 9 . A comparison of the errors of the training set, the valid set, and the test set.	31
Table 10 . Reduction in runoff peaks after BMP implementation for 2023-2024, with default BMP configuration.....	36
Table 11 . Reduction in bacteria concentrations after BMP implementation for 2023-2024, with default BMP configuration.	37
Table 12 . Three BMP configuration used in sensitivity analysis.....	38
Table 13 . Percentage reduction in runoff and bacteria concentration with each hypothetical BMP configuration.....	38

Chapter I

Introduction

1 Background Information

The Neches River in Southeast Texas has been reported (TCEQ, 2010) that the bacteria indicator, namely *Escherichia coli* (*E. coli*) in freshwater or Enterococci in saltwater exceeded the criterion of 126 mpn/100ml or 35 mpn/100ml, respectively. Recent flooding caused the increase of the fecal bacteria loading along the Lower Neches River Tidal is a major water quality concern of the region. This increase can be explained by an increase in failures of on-site sewage facilities (OSSFs) that are an aggregate result of aging facilities and more frequent occurrences of extreme flooding in recent years. A total maximum daily load (TMDL) and implementation plan (I-Plan) to reduce bacteria and protect recreational safety in the Neches River Tidal are being developed by the stakeholders and Texas Commission on Environmental Quality (TCEQ, <https://www.tceq.texas.gov/waterquality/tmdl/nav/118-nechestidal-bacteria>). The study (TCEQ, 2022) has illustrated that bacteria load contributions from regulated and unregulated stormwater sources follows a pattern of higher concentrations in the water body as the first flush of storm runoff enters the receiving stream and declines as runoff washes bacteria from the land surface. The bacteria released from OSSFs can travel considerable distances through saturated soils and contaminate groundwater, terrestrial runoff, and coastal waters. If left unmitigated, the rising presence of bacteria along Neches River and in coastal waters can pose a serious public health risk, and this risk may amplify

with increasing frequency and intensity of flood events, and by sea-level rise. Therefore, there is an urgent need to determine the origin and transport pathways of the bacteria in the region and assess BMPs as countermeasures to alleviate the excess loading.

The project was a joint effort among Lamar University, UT Arlington, and Texas A & M, seeking answers to this need under Coastal Hazards and Resiliency Planning category. In this project, we focus on the Neches River Tidal watershed to a) conduct a field survey on characteristics of OSSFs to identify the criteria that help us find bacteria loading hot spots; b) assess the bacteria impact by analyzing the water quality data collected at the Salt Water Barrier using YSI sensors using innovative machine learning technologies, running water samples at locations where the most pollution source origin are discharged and, and applying innovative lab technology to track fecal bacteria sources from the selected water samples, c) couple the established SWMM model with an offline bacteria life cycle-MOPUS model for understanding bacteria transport pathways and assessing BMPs to alleviate excess loadings. Three universities incorporated in parallel to conduct three goals as three project tasks.

2 Study Areas

The research team from Texas A&M used the OSSF permit data collected from the TCEQ OARS database (data as of December 2021) to develop a map showing OSSF inventory and updated parcels as shown in Chapter I2Figure 1. The map included circles with six distinct colors indicating the age of the OSSFs when available from the permit or

building records as well as sewer area and location of TCEQ permitted wastewater outfalls (discharge points). Figure 2 show a map of 100-year floodplain data from the FEMA National Flood Hazard Layer (NFHL) digital database, the Soil Survey Geographic Database (SSURGO) soil data for the project area, and the OSSF inventory.

After consulted with the representatives of TOWA (Texas Onsite Wastewater Association) and the Orange County Health Department to determine areas where OSSFs may be “failing” due to age and/or limiting soil conditions (shallow depth to water table and clay soil) with the 100-year floodplain. Seven “hot” areas were identified for conducting walk-through site visits on March 14 and 15, 2023. The field survey evaluated the operating conditions of the OSSFs and followed the water flow through the creek to determine adequate and safe locations for water quality sampling. Field observations did not identify any major failures of OSSF (no sewage on the ground or sewer odor in the area), however in two locations (sampling points #2 and #13), presence of algae was noticed indicating potential contamination from the subsurface movement of partially treated wastewater (OSSF discharges). Students from Lamar were trained for collecting water samples using sampling bottles and sampling locations (1 – 14) were marked on the map. Figure 3 shows the seven areas for site visits and 14 locations that were selected for water quality sampling events.

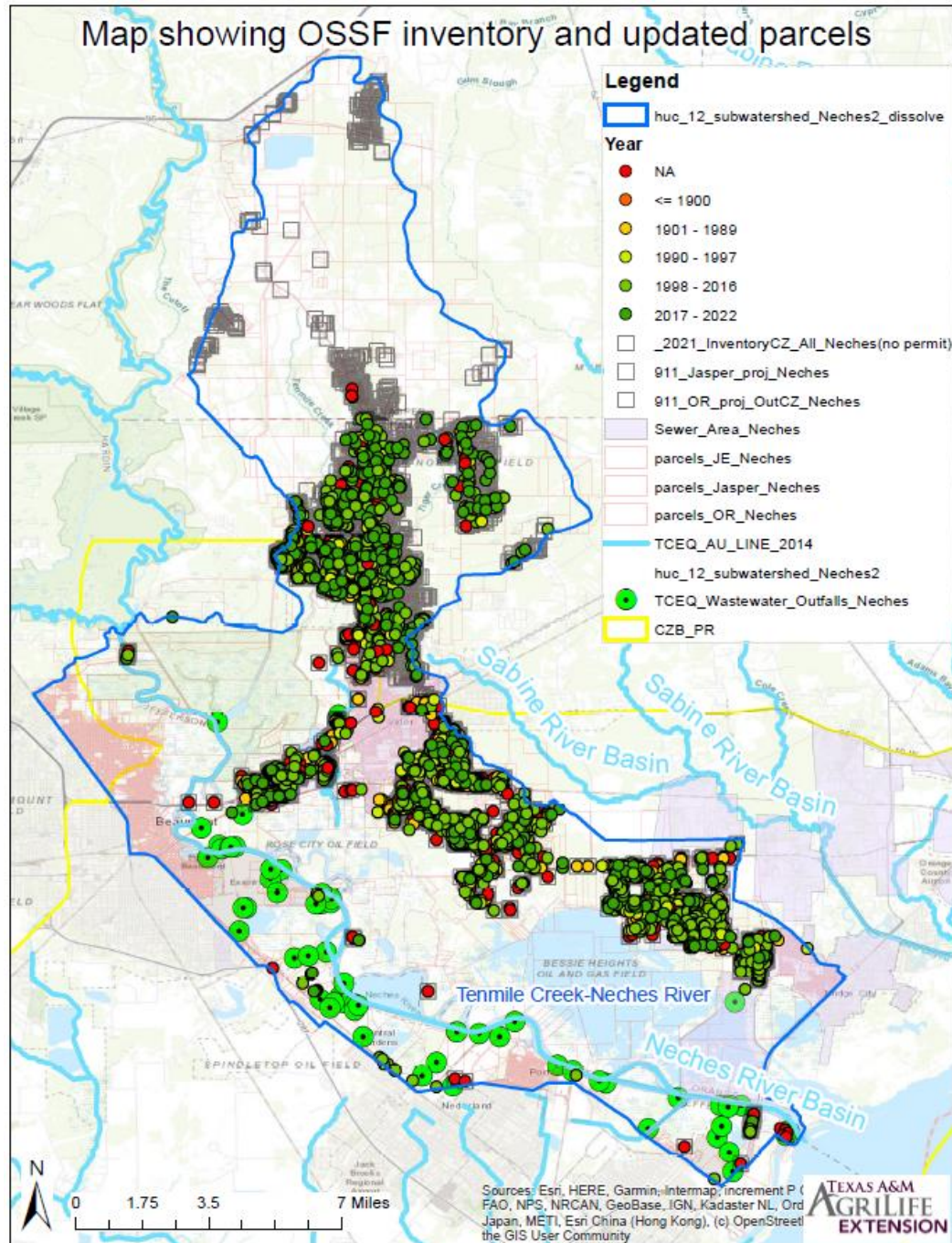


Figure 1. Study area with OSSF inventory map as well as updated parcels and wastewater outfalls.

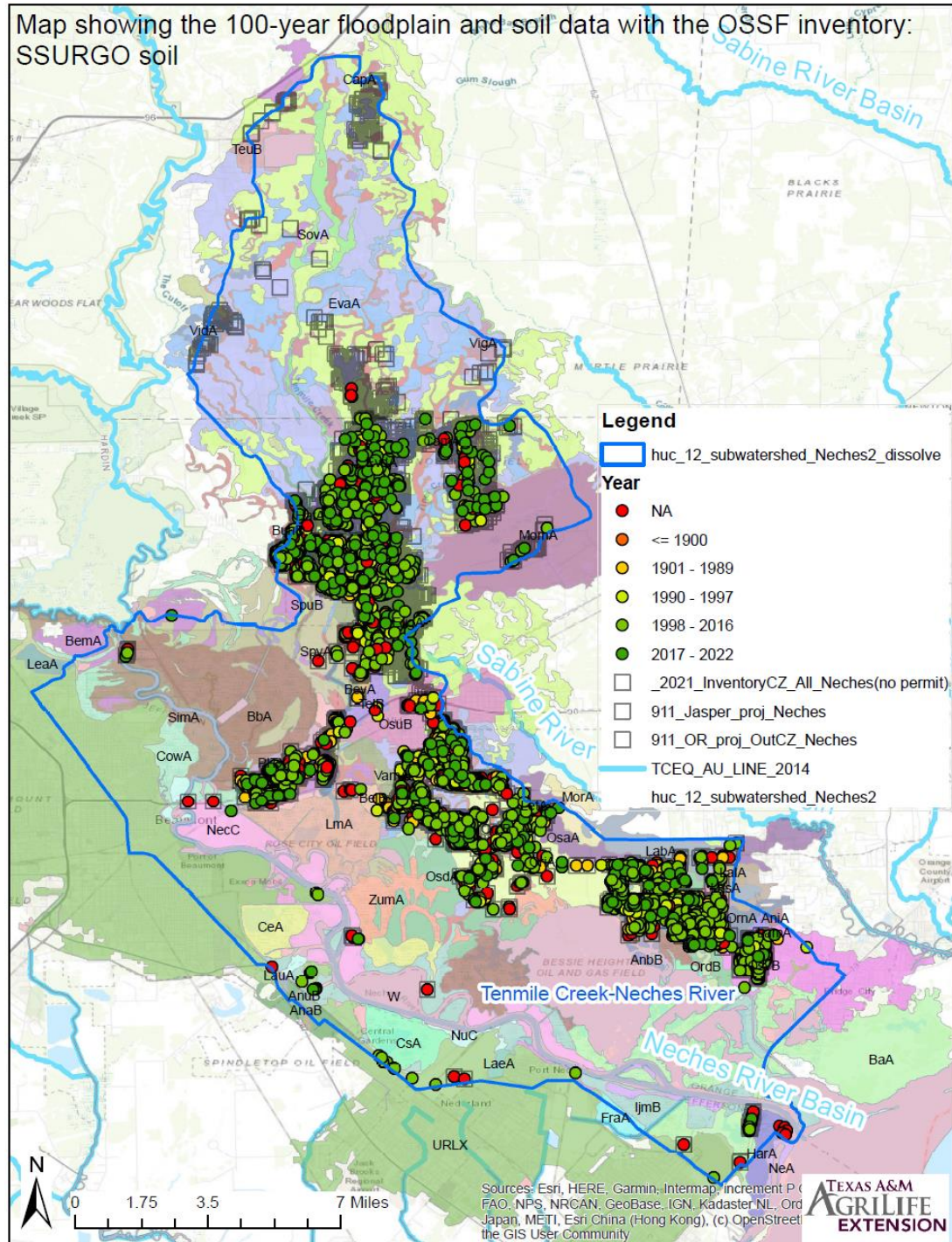


Figure 2. Map of 100-year floodplain, SSURGO soil maps and OSSF inventory.

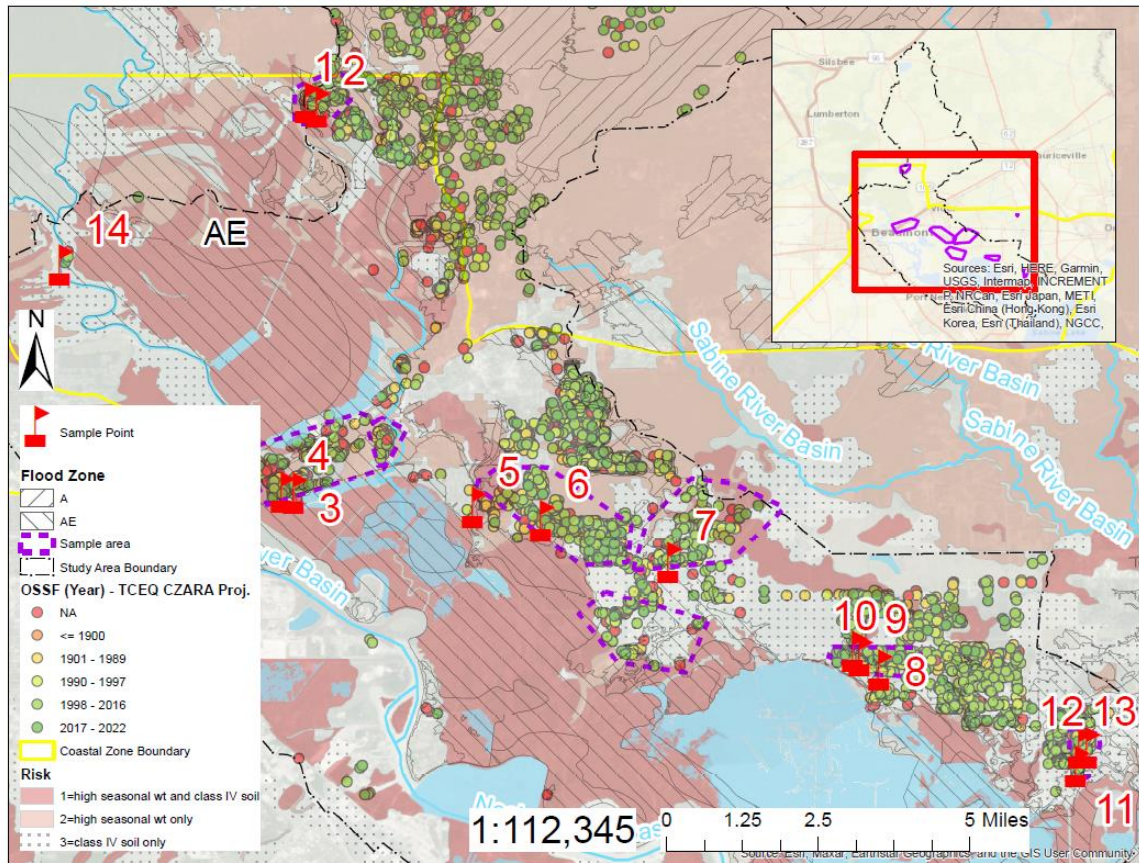


Figure 3. Map of seven areas shown in the insert in top right with purple boundary and 14 sampling points for water quality sampling analysis.

Chapter II

Methods

1 Field Water Sampling

After major rain events as shown in Figure 4, the field sampling was conducted to collect the 1000-ml water samples and stored in the cooler immediately. At each site, the YSI-ProDSS system was used to measure the field temperature, DO, pH, conductivity, Turbidity, Ammonia-Nitrogen, and Nitrate-Nitrogen. The “wastewater Sampling Form” was filled to record the site condition and measured data. Figure 5 shows a sample sheet on 07/06/2023. The water samples were sent back to laboratory in less than 6-hours to conduct the E. coli measurement.

Following the “Pre-processing of water samples for quantitative PCR” provided by Dr. Terry Gentry from Texas A&M University Soil & Aquatic Microbiology Laboratory, the sample was pre-processing with disposable membrane filtration units (filter base, polycarbonate filter [0.4 μm] with 47 mm diameter, and 100 mL capacity funnel) in the lab and stored the folded filters in the storage tube at -80°C until shipment to lab for DNA extraction and purification. In addition, the Ammonia-Nitrogen, Nitrite-Nitrogen, Nitrate-Nitrogen, Total Nitrogen, Phosphate, COD were also measured in the laboratory to understand the water chemistry.

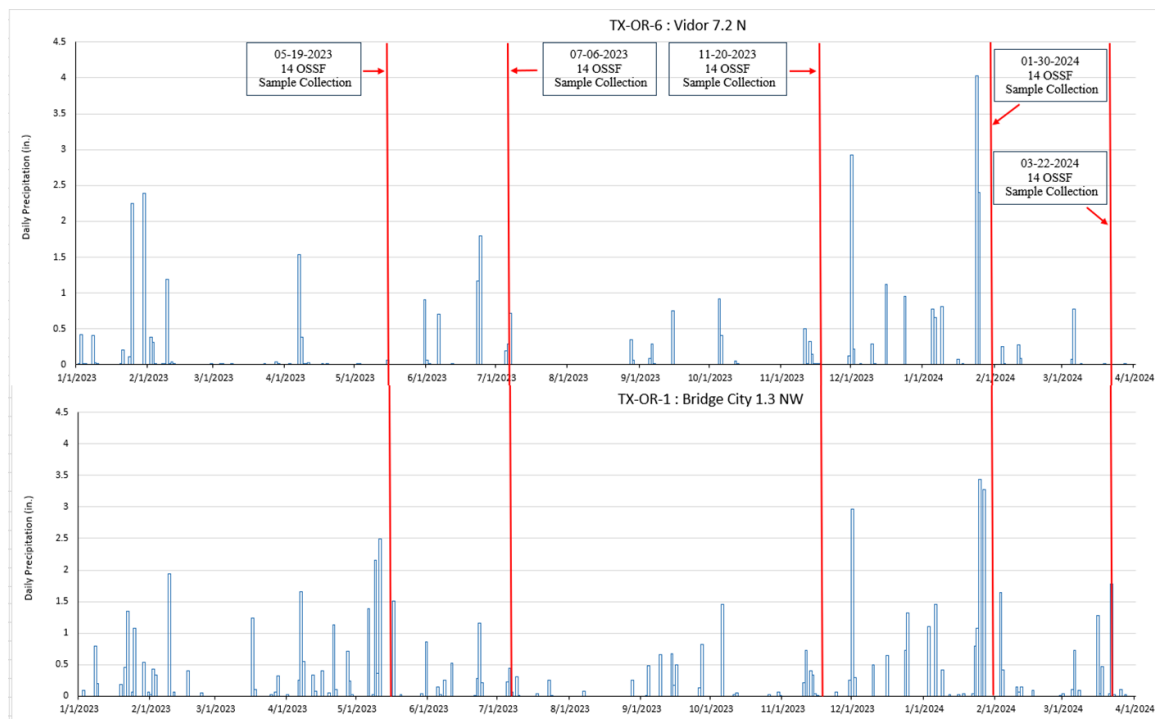


Figure 4. Sampling schedule based on rain fall events.

Wastewater Sampling Form						
Sample	SR					
Collector:	Jian Fang, Ameer Chaulagain					
Date:	July 6, 7/6/2023					
Time:	1:55 pm					
Recent Weather Conditions:	Rainy, Cloudy					
Site Name:	Easy St.					
Description of the site surrounding and pictures: No vegetation, No slow Blackish color						
Field measurements:						
Temp. (°F)	DO (mg/L)	pH	Conductivity (Microseimens/cm)	Ammonium (mg/L)	Turbidity (NTU)	Nitrate NO ₃ ⁻ (mg/L)
80.9	2.5	7.36	220.5	0.60	64.9	0.89
81.0	3.1	7.33	198.4	0.84	80.9	1.57
Quantitative PCR (mL)		100mL				
E. Coli (MPN/100 mL)		1413.6				
Signature <i>Dim</i>						

Figure 5. Sample of "Wastewater Sampling Form" on 07/06/2023.

2 Machine Learning Models with YSI Data

The Neches River shown in Figure 6 rises east of Colfax (at 32°30' N, 95°45' W) and flows southeast for 669.49 km to its mouth on Sabine Lake (at 29°58' N, 93°51' W). The total of 25928.37 km² drainage area with abundant rainfall results in a flow of 7.4 km³ per year (TSHA, 1995a). Saltwater Barrier (SWB) was constructed on the Lower Neches River, downstream of Beaumont's drinking water pump station to prevent the intrusion of the saltwater wedge from the Gulf of Mexico (Pizano-Torres et al., 2017). A complex EXO2 YSI sensor system was installed at the SWB co-location with USGS station of 08041780 to monitor the water quality, including water temperature, sample depth, conductivity, turbidity, total dissolved solids (TDS), water pH, chlorophyll, nitrate (NO₃-N) and DO (Qian et al., 2019, 2024). Pine Island Bayou (PIB), the major tributary of Lower Neches River, rises two miles south of Fuqua (at 30°25' N, 94°44' W) and reaches its mouth 9.66 km north of downtown Beaumont (at 30°10' N, 94°07' W), with approximately 122.31 km long (TSHA, 1995b). A water quality sensor installed at the same USGS station 08041749 monitors real-time water quality measuring water temperature, sample depth, conductivity, turbidity, TDS, water pH and DO since June 11, 2008 (TCEQ, 2018a). Discharges obtained from USGS stations of 08041000, 08041749 and 08041780 (<https://waterdata.usgs.gov/>) are important hydrological characteristics to illustrate the local hydro-meteorological conditions.

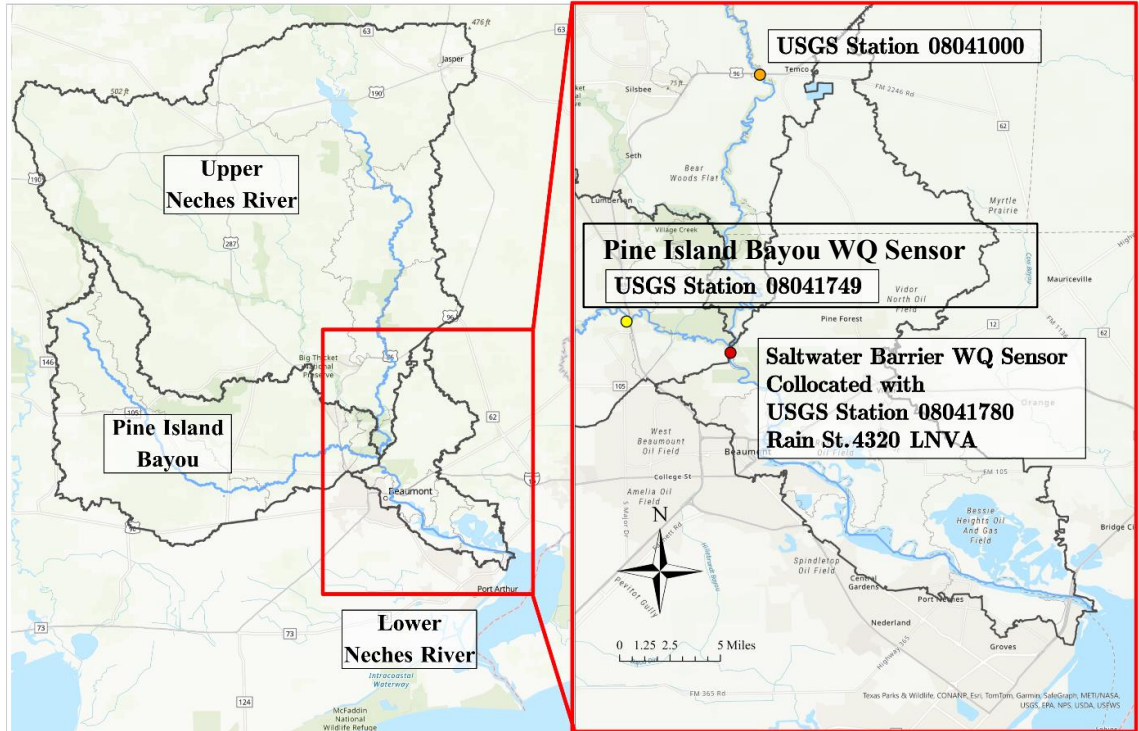


Figure 6. Locations of YSI water quality sensors and USGS stations

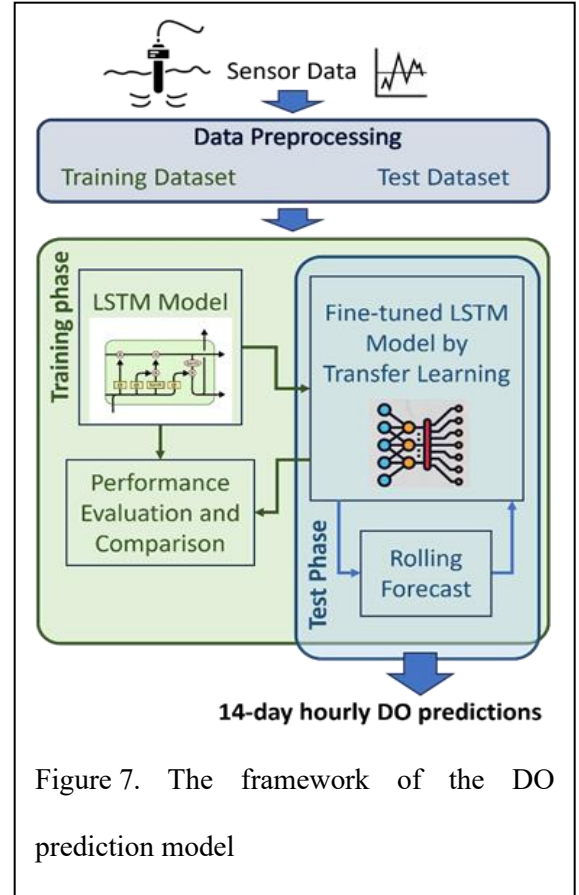
As shown in Figure 6, 15-minute interval gage height and discharge were obtained from USGS gage station 08041780 (<https://waterdata.usgs.gov/>). USGS station 08041000 (at 30° 21' 20.75" N, 94° 5' 35.65" W) is monitoring water discharge for the upper Neches River, while USGS station 08041749 (at 30° 10' 43.76" N, 94° 11' 19.66"W) is associated with the Pine Island Bayou, the main tributary of the Neches River. Hourly rain increment data at the SWB was collected from St.4320 LNVA Saltwater Barrier (<https://dd6.onerain.com/>).

2.2 Machine Learning Model to Predict Dissolved Oxygen (DO)

Recently, approaches utilizing advanced machine learning techniques to predict water quality with wireless sensor measurements demonstrated high performance (Kim and Ahn, 2022; Jiang et al., 2023). The machine learning algorithms using sensor data to predict water quality DO are Support Vector Machine (SVM) (Li et al., 2020; Nong et al., 2023), Random Forest (RF) (Tiyasha et al., 2021; Ayesha Jasmin et al., 2022), Artificial Neural Network (ANN) (F. Yang et al., 2021; Azma et al., 2023), Recurrent Neural Network (RNN) (Y. Liu et al., 2019) and Long Short-Term Memory (LSTM) (Zhi et al., 2021). It is found that machine learning models outperformed traditional Multiple Linear Regression (MLR) models due to their capabilities of exploring nonlinear relationships between target and input features (Csábrági et al., 2017).

RNN and LSTM are widely used deep learning models to solve the prediction problems with sequence data due to the effective ability of memorizing the previous data (Greff et al., 2017). LSTM is designed RNN to remember sequences with a data length of 10 or more, avoiding the weakness of the classical RNN in long-term memory ability (Abba et al., 2020). The special capability comes from three kinds of gates in the LSTM memory cell, namely forget gate, input gate, and output gate (Huan et al., 2020; P. Liu et al., 2019). Thus, the LSTM is suitable to develop longer period DO prediction with time series measurements. To generate a consistent long period DO forecasting, rolling forecast procedure is employed through continuously repeating the prediction process with updated future data to provide real-time and dynamic forecast in unstable environment, (Zeller and

Metzger, 2013). In general, LSTM needs a large quantity of training data to mine the potential relationships with nonlinear inputs, which limits their applications in many instances due to under-fitting problem (Espejo-Garcia et al., 2020). Transfer learning method is an ideal and commonly used method to solve the insufficient data issue by transferring knowledge across similar domains, tasks, and distributions (Lumini and Nanni, 2019). Zhu et al. (2021) developed a pre-training DO model based on the bidirectional LSTM (BiLSTM) with a



large dataset of the Lake Taihu in China, and used transfer learning to fine-tune the model with the target dataset of another lake to increase coefficient of determination (R^2) from 0.381 to 0.793.

To develop a LSTM deep machine model and provide a 14-day forecast for local agencies to make appreciative decisions on water resource planning and exploit relationships between DO and other sensor measurements under different hydro-meteorological conditions, the main steps as illustrated in include 1) building the LSTM model for current hourly DO estimations with different input feature scenarios after

the comprehensive sensor data analysis, and comparing with a baseline model using traditional MLR, 2) combining transfer learning to improve the performance of the LSTM model with insufficient dataset, 3) generating consistent 14-day hourly DO forecasts with a rolling forecast procedure.

2.3 Machine Learning Model to Predict Turbidity

Machine learning has become a valuable tool for analyzing turbidity due to the rapid increase in sensor measurements in the aquatic environment. For example, Support Vector Machine (SVM), Fuzzy Inference Systems (FIS), group method of data handling (GMDH), Genetic programming (GP), Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) models have been applied to predict turbidity in reservoirs, rivers, and coastal bays, and proved their desirable accuracy (Teixeira et al., 2020; Tsai and Yen, 2017; Wang et al., 2021). Those models are commonly built with rainfall, discharge, and water level, because heavy rainfall affects turbidity via erosion and subsequent runoff, increasing sediment loads in water bodies (Leigh et al., 2019). However, the kinds of literature only predict the current turbidity with other current or previous measurements. To our best knowledge, no model is built for forecasting a period of turbidity, which is more useful for local agencies by providing enough time and flexibility for better water resource management decisions. To predict hourly turbidity in the following days with hydro-meteorological measurements, discharge, gage height, and precipitation, a multivariate time series multi-step forecasting framework via an attention-based encoder & decoder

structure is applied. The encoder and decoder are well-designed using the Gated Recurrent Unit (GRU), an advancement of the standard Recurrent Neural Network (RNN) introduced by Cho et al. (Cho et al., 2014). It uses special gates (update gate and reset gate) to control the flow of information, which can solve the limitations of standard RNNs on long-term memory. It is motivated by the Long Short-Term Memory (LSTM) unit but is much simpler to compute and implement. It has three components: an encoder component, a decoder component, and a temporal attention layer as an attention component (Bahdanau et al., 2016; Cho et al., 2014; Du et al., 2020; Zhang et al., 2017). The encoder learns the hidden representation of input data with arbitrary lengths, which extracts the deep temporal dependency features from the multivariate time series and then uses the temporal attention layer to construct latent space variables (temporal attention context vectors). The decoder generates latent space variables for forecasting future time series values. Figure 8 shows the graphical illustration of the framework.

3 SWMM & MOPUS Model

EPA's Stormwater Management Model (EPA-SWMM) is a useful tool in planning local, regional and national water resources solution by green infrastructures. EPA-SWMM can determine the reduction of runoff by infiltration and retention over the watershed. Low impact development, such as green/grey/hybrid infrastructure can be added as a fraction of impervious area of a specific catchment into EPA-SWMM model by LID control tool.

MOPUS (McCarthy et al., 2011) is a conceptually based model to calculate wash off and storage of microorganisms for a catchment.

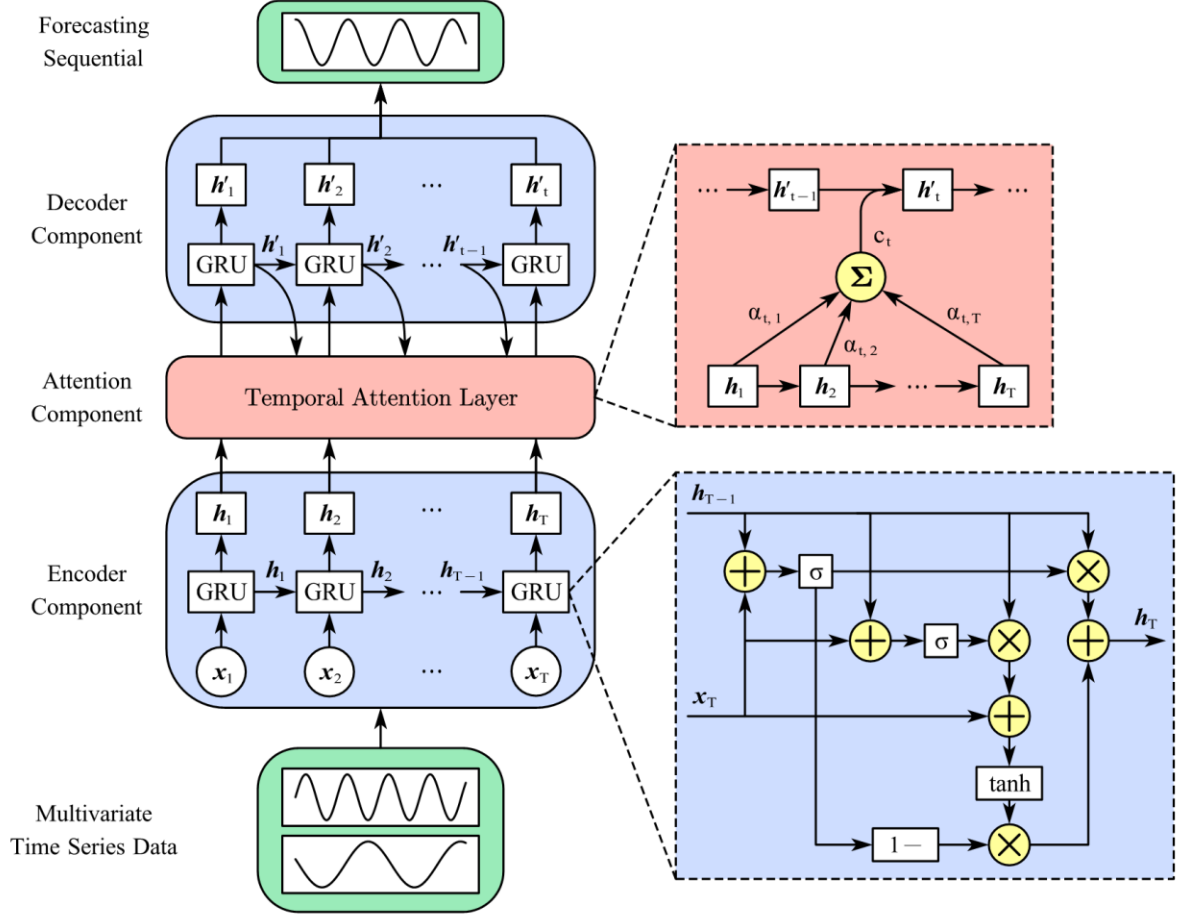


Figure 8. The sequential-to-sequential forecasting framework via a temporal attention-based encoder & decoder model.

For each site, EPA-SWMM model is developed for the upstream drainage to produce runoff simulations under current conditions and with BMPs. Then coupled with MOPUS model at 14 sites (Figure 9) to assess the contribution of runoff to bacterial loading, and to determine the potential ability of BMPs (bioretention) to reduce runoff and bacterial

loading. These BMP sites were selected to be within areas with a large number of septic tanks.

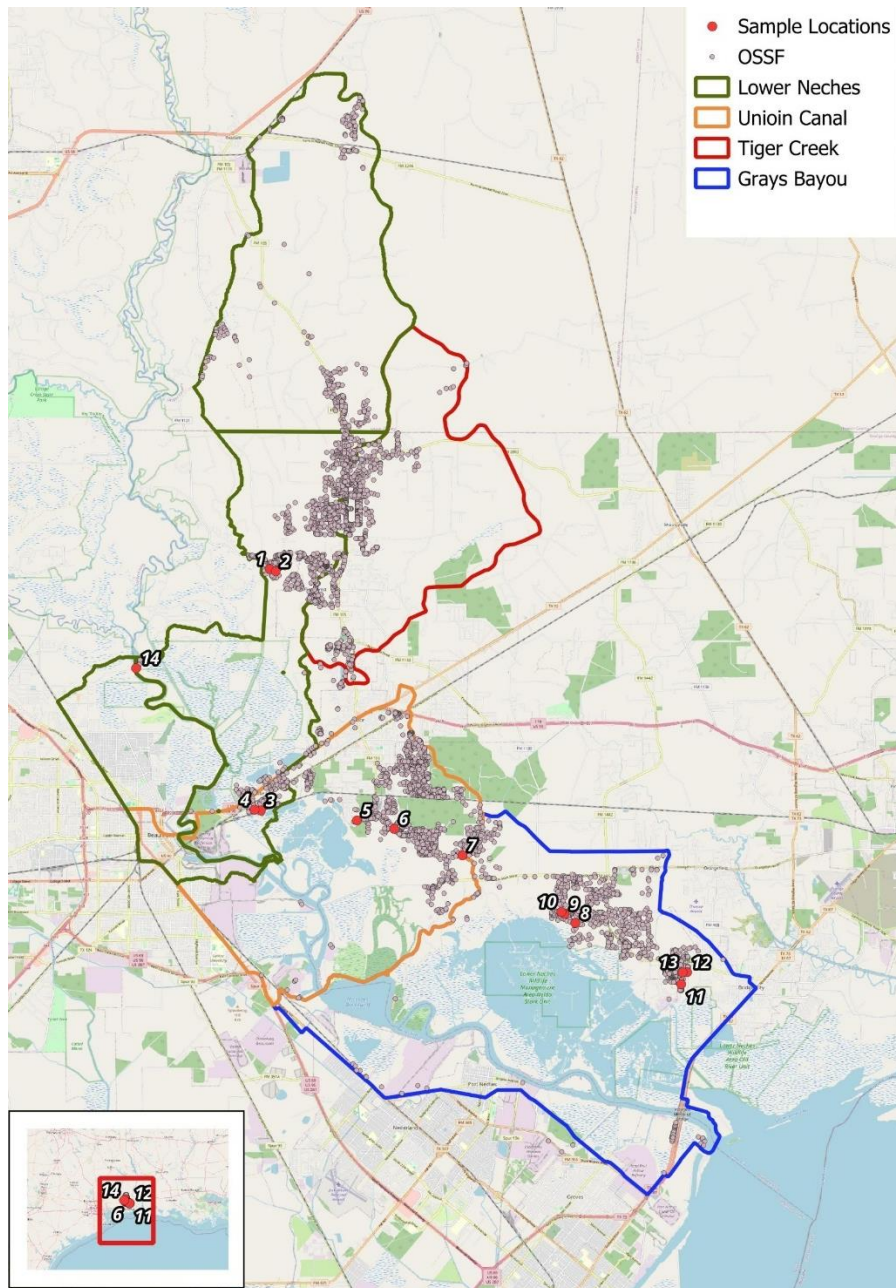


Figure 9. Locations where hypothetical BMPs (bioretention cells) are implemented in SWMM.

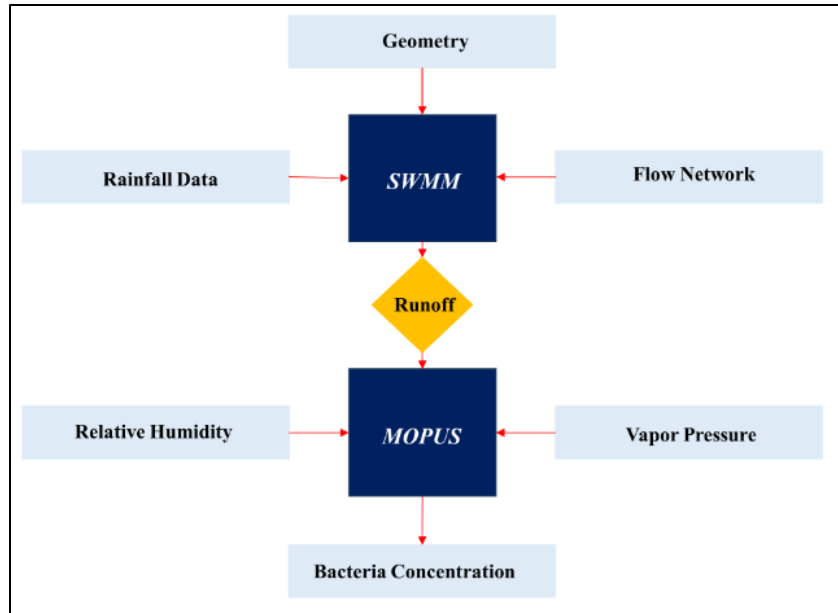


Figure 10. Schematic of coupled modeling system comprised of SWMM and MOPUS.

The coupled modeling system is illustrated in Figure 10. In this system, EPA-SWMM produces runoff simulations which are ingested into the MOPUS model to produce time series of bacterial concentration. To determine the impacts of BMPs, various configurations of BMPs (bioretention cells, Figure 11) were implemented within EPA-SWMM.

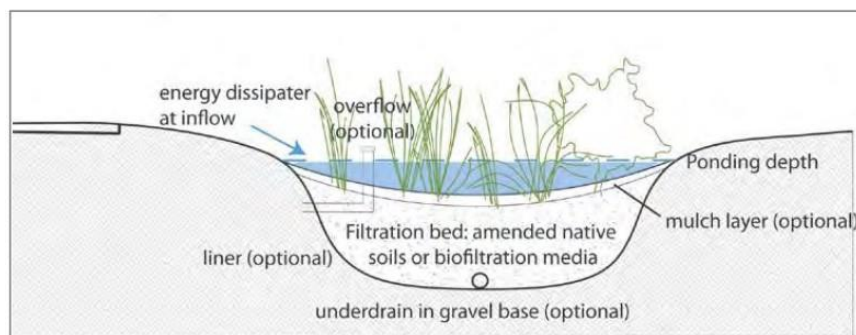


Figure 11. Configuration of a bioretention cell.

(Source: <https://www.cleancoast.texas.gov/documents/5-6-sw-manual.pdf>)

Chapter III

Results and Discussion

1 Results of Field Water Sampling Measurements

Table 1. Lab and field measurements on 05/19/2023.

Measurements from Dr. Thinesh's lab							Field sensor readings							CAWAQ
Parameters	Ammonia Nitrogen (NH ₃ -N) <0.46mg/L	Nitrite Nitrogen (NO ₂ -N) <1.1mg/L	Nitrate Nitrogen (NO ₃ ⁻ -N) <1.1mg/L	Phosphate (PO ₄ ³⁻) <0.66mg/L	Total Nitrogen (N) < 10mg/L	COD (mg/L)	Temp <95 °F	DO >3 mg/L	pH 6.0~8.5	Conductivity <2500 µS/cm	Ammonia mg/L	Turbidity <75 NTU	Nitrate mg/L	E. coli <126 MPN/100ml
Name of test reagent	TNT 832	TNT 841	TNT 835	TNT 826	PhosVer Rea.	TNT 821								
Measurement range	2.0-47.0	2.0-90.0	0.23-13.50	1.0-16.0	0.0-2.5	3.0-150.0								
Sample 1	0.07	0.43	1.01	2.03	0.1	87.8	72.7	1.9	7.75	110.7	0.53	76.1	5.64	1
Sample 2	0.26	0.12	1.18	1.96	0.34	86.7	73.2	0.9	6.79	98.3	0.86	62.8	7.70	1
Sample 3	0	0.06	0.31	0.61	0	24.4	76.0	1.4	7.11	195.5	xxx	7.5	0.91	8.5
Sample 4	0	0.37	0.66	1.56	0.45	56.1	72.8	1.2	7.12	439.4	1.77	21.0	1.80	9.4
Sample 5	0	0.23	0.84	1.34	0	53.7	78.3	3.0	7.07	225.7	0.24	17.5	5.78	<1.0
Sample 6	0	0.64	0.69	1.04	0	71.7	76.1	1.1	7.10	152.0	0.25	12.2	2.64	1
Sample 7	0	0.57	0.58	0.87	0	70.1	77.5	2.9	6.97	273.4	0.16	2.8	1.69	<1.0
Sample 8	0	0.30	0.91	1.15	0	70.1	77.5	0.8	6.79	150.4	0.20	19.6	3.24	2
Sample 9	0	0.15	0.91	1.33	0	65.8	76.8	1.9	6.98	199.4	0.39	21.5	3.42	<1.0
Sample 10	0	0.15	1.04	1.20	0	77.8	76.1	1.6	6.80	131.2	0.20	18.6	2.69	<1.0
Sample 11	0.21	0.45	0.89	1.68	0.06	73.9	75.7	0.6	7.05	246.9	0.95	21.5	4.17	<1.0
Sample 12	0	0.44	0.81	1.44	0.25	64.5	76.3	4.4	7.54	395.2	0.29	12.1	3.32	<1.0
Sample 13*	6.11	0.54	0.71	8.36	0	64.8	89.4	0.5	7.98	835.0	5.47	189.7	5.50	>2419.6
Sample 14	0	0.43	0.63	0.77	0	42.9	79.9	6.1	6.82	124.0	0.28	28.5	6.82	<1.0
Note:	*The values provided here are the average of two independent measurements of each sample.													

Table 2. Lab and field measurements on 07/06/2023.

Measurements from Dr. Thinesh's lab*							Field sensor readings *							CAWAQ
Parameters	Ammonia Nitrogen (NH ₃ -N) <0.46mg/L	Nitrite Nitrogen (NO ₂ -N) <1.1mg/L	Nitrate Nitrogen (NO ₃ ⁻ -N) <1.1mg/L	Phosphate (PO ₄ ³⁻) <0.66mg/L	Total Nitrogen (N) < 10mg/L	COD (mg/L)	Temp <95 °F	DO >3 mg/L	pH 6.0~8.5	Conductivity <2500 µS/cm	Ammonia mg/L	Turbidity <75 NTU	Nitrate mg/L	E. coli <126 MPN/100ml
Name of test reagent	TNT 832	TNT 841	TNT 835	TNT 826	PhosVer Rea.	TNT 821								
Measurement range	2.0-47.0	2.0-90.0	0.23-13.50	1.0-16.0	0.0-2.5	3.0-150.0								
Sample 1	1.28	0.52	0.00	3.48	0.14	64.3	82.3	0.6	10.30	623.0	8.73	242.6	1.54	4.1
Sample 2	0.00	0.43	0.37	1.51	0.07	0.0	80.9	2.8	7.45	209.5	0.72	72.9	1.23	1413.6
Sample 3	0.08	0.41	0.33	1.10	0.10	0.0	79.7	1.1	7.21	155.2	0.14	8.0	0.60	129.6
Sample 4	0.62	0.6	0.65	2.62	3.54	0.0	79.4	0.4	7.10	182.5	0.64	40.4	1.27	>2419.6
Sample 5	0.52	0.42	0.29	1.66	0.43	0.0	83.3	0.2	7.96	179.4	1.11	12.4	0.33	>2419.6
Sample 6	0.03	0.35	0.28	0.78	0.17	0.0	82.6	0.3	7.63	204.9	0.21	6.8	0.70	5.2
Sample 7	0.00	0.36	0.38	0.90	0.34	0.0	83.6	1.0	7.47	297.3	0.21	21.5	1.03	68.2
Sample 8	0.00	0.33	0.30	0.86	0.16	11.3	84.9	0.2	7.32	207.1	1.86	7.3	2.47	365.4
Sample 9	0.10	0.39	0.19	0.46	1.33	0.0	82.1	2.1	7.77	756.0	0.61	16.0	1.13	48
Sample 10	0.12	0.45	0.31	1.12	0.18	1.2	82.8	3.0	7.38	1173.5	0.87	466.0	2.22	10.9
Sample 11	0.40	0.44	0.59	1.51	0.37	2.8	82.0	0.9	7.35	348.8	4.54	11.6	1.21	>2419.6
Sample 12	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sample 13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sample 14	0.06	0.49	0.18	0.54	0.10	9.4	88.8	4.8	7.41	173.3	1.23	24.1	1.39	10.8
Note:	* The values provided here are the average of two independent measurements of each sample.													

Table 3. Lab and field measurements on 12/06/2023.

Measurements from Dr. Thinesh's lab*							Field sensor readings *							CAWAQ
Parameters	Ammonia Nitrogen (NH ₃ -N) <0.46mg/L	Nitrite Nitrogen (NO ₂ -N) <1.1mg/L	Nitrate Nitrogen (NO ₃ ⁻ -N) <1.1mg/L	Phosphate (PO ₄ ³⁻) <0.66mg/L	Total Nitrogen (N) < 10mg/L	COD (mg/L)	Temp <95 °F	DO >3 mg/L	pH 6.0~8.5	Conductivity <2500 µS/cm	Ammonia mg/L	Turbidity <75 NTU	Nitrate mg/L	E. coli <126 MPN/100ml
Name of test reagent	TNT 832	TNT 841	TNT 835	TNT 826	PhosVer Rea.	TNT 821								
Measurement range	2.0-47.0	2.0-90.0	0.23-13.50	1.0-16.0	0.0-2.5	3.0-150.0								
Sample 1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sample 2	0.08	0.63	0.50	2.02	0.39	52.2	67.1	2.4	7.69	542.5	0.34	6.2	1.53	126.7
Sample 3	0.14	0.52	0.32	1.93	0.13	35.8	64.1	1.1	6.14	4440.5	7.35	6.3	2.76	14.2
Sample 4	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sample 5	0.03	0.48	0.31	1.65	0.13	43.9	67.1	0.6	6.75	14361.0	9.79	14.0	3.11	56
Sample 6	0.49	0.22	0.21	1.00	0.07	17.2	64.3	1.3	7.24	347.7	0.28	12.6	1.13	30.9
Sample 7	0.42	0.60	0.34	2.01	0.28	30.4	66.7	1.3	7.06	13983.5	8.96	20.7	2.45	1732.9
Sample 8	0.71	0.60	0.31	1.74	0.26	62.4	65.7	1.1	6.81	26930.0	19.37	8.0	3.64	1119.9
Sample 9	0.12	0.57	0.21	0.67	0.99	13.7	65.0	0.9	7.34	1208.5	0.26	26.3	0.81	1413.6
Sample 10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sample 11	0.16	0.59	0.45	1.22	0.58	29.8	63.7	0.9	7.34	452.5	0.35	14.6	0.84	23.1
Sample 12	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sample 13	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Sample 14	0.11	0.62	0.31	0.55	0.19	17.8	67.0	5.4	7.36	5211.5	2.98	17.8	1.50	27.2
Note:	* The values provided here are the average of two independent measurements of each sample.													

Table 4. Lab and field measurements on 01/30/2024.

Measurements from Dr. Thinesh's lab*							Field sensor readings *							CAWAQ
Parameters	Ammonia Nitrogen (NH ₃ -N) <0.46mg/L	Nitrite Nitrogen (NO ₂ -N) <1.1mg/L	Nitrate Nitrogen (NO ₃ ⁻ -N) <1.1mg/L	Phosphate (PO ₄ ³⁻) <0.66mg/L	Total Nitrogen (N) < 10mg/L	COD (mg/L)	Temp <95 °F	DO >3 mg/L	pH 6.0~8.5	Conductivity <2500 µS/cm	Ammonia mg/L	Turbidity <75 NTU	Nitrate mg/L	E. coli <126 MPN/100ml
Name of test reagent	TNT 832	TNT 841	TNT 835	TNT 826	PhosVer Rea.	TNT 821								
Measurement range	2.0-47.0	2.0-90.0	0.23-13.50	1.0-16.0	0.0-2.5	3.0-150.0								
Sample 1	0.00	0.74	0.64	0.97	0.38	66.7	52.7	6.2	5.08	10.2	0.36	62.5	3.29	53.8
Sample 2	0.00	0.76	0.59	1.02	0.36	65.9	53.2	6.7	5.15	10.8	0.38	60.3	5.72	61.3
Sample 3	0.00	0.23	0.53	0.56	0.10	21.4	51.8	5.5	6.48	5.9	0.57	195.1	1.94	96.0
Sample 4	0.03	0.40	0.63	1.43	1.10	36.4	54.2	2.1	6.38	24.7	1.43	442.6	1.79	178.9
Sample 5	0.00	0.75	0.59	1.07	0.51	47.9	54.5	3.7	6.17	23.0	0.88	401.6	2.78	488.4
Sample 6	0.01	0.46	0.43	1.44	0.23	33.0	53.6	4.8	6.24	53.2	0.50	96.5	9.75	25.9
Sample 7	0.00	0.76	0.69	1.23	0.40	50.6	53.0	7.8	6.33	30.5	0.79	121.6	6.81	816.4
Sample 8	0.00	0.92	0.61	1.28	0.45	69.5	53.6	5.4	5.59	48.3	0.71	112.0	4.87	52.9
Sample 9	0.00	0.90	0.62	1.22	0.38	72.7	52.3	6.9	5.54	18.5	0.52	73.8	6.22	67.0
Sample 10	0.00	1.02	0.58	1.43	0.51	71.2	53.2	6.8	5.64	18.2	0.9	71.4	7.39	33.1
Sample 11	0.56	1.06	0.62	2.34	0.72	52.8	53.2	3.1	6.93	29.3	4.48	379.0	5.85	228.2
Sample 12	0.00	0.49	0.53	1.65	0.62	46.3	55.5	7.8	7.34	7.7	1.34	948.0	1.81	77.1
Sample 13	0.00	0.97	0.69	2.68	2.22	81.1	52.9	4.4	7.65	47.5	1.07	333.5	5.21	770.1
Sample 14	0.00	0.47	0.72	0.81	0.27	32.05	56.7	8.7	5.79	35.4	0.57	91.8	5.86	156.5
Note:	* The values provided here are the average of two independent measurements of each sample.													

Table 5. Lab and field measurements on 03/22/2024.

Measurements from Dr. Thinesh's lab*							Field sensor readings *							CAWAQ
Parameters	Ammonia Nitrogen (NH ₃ -N) <0.46mg/L	Nitrite Nitrogen (NO ₂ -N) <1.1mg/L	Nitrate Nitrogen (NO ₃ ⁻ -N) <1.1mg/L	Phosphate (PO ₄ ³⁻) <0.66mg/L	Total Nitrogen (N) < 10mg/L	COD (mg/L)	Temp <95 °F	DO >3 mg/L	pH 6.0~8.5	Conductivity <2500 µS/cm	Ammonia mg/L	Turbidity <75 NTU	Nitrate mg/L	E. coli <126 MPN/100ml
Name of test reagent	TNT 832	TNT 841	TNT 835	TNT 826	PhosVer Rea.	TNT 821								
Measurement range	2.0-47.0	2.0-90.0	0.23-13.50	1.0-16.0	0.0-2.5	3.0-150.0								
Sample 1	0.08	0.38	0.74	1.04	0.28	55.5	59.3	5.9	5.99	72.4	0.36	29.9	1.07	2419.6
Sample 2	0.16	0.25	0.74	1.25	0.36	54.7	59.3	6.4	5.37	52.7	0.50	33.3	1.79	>2419.6
Sample 3	0.32	0.32	0.49	0.90	0.48	32.3	59.3	5.5	6.24	189.6	0.46	25.7	0.41	>2419.6
Sample 4	0.27	0.62	0.81	1.90	2.20	43.5	59.0	4.6	6.15	117.5	1.06	56.6	0.91	>2419.6
Sample 5	0.16	0.55	0.75	1.22	0.89	40.7	60.0	7.3	5.91	98.4	0.58	43.0	1.31	>2419.6
Sample 6	0.06	0.21	0.39	0.90	0.46	30.3	59.7	6.3	5.77	90.8	0.63	55.3	1.00	>2419.6
Sample 7	0.25	0.81	0.88	1.92	1.10	58.5	59.9	6.7	5.75	99.9	0.97	75.4	2.30	>2419.6
Sample 8	0.18	0.82	1.02	1.97	0.90	51.3	59.3	6.9	5.68	89.0	0.85	65.4	2.29	>2419.6
Sample 9	0.32	0.64	0.89	1.82	0.45	53.2	59.1	7.4	6.34	70.1	0.59	59.7	2.58	>2419.6
Sample 10	0.16	0.34	1.10	1.86	0.87	52.6	59.1	6.0	6.77	128.9	0.65	38.8	2.89	>2419.6
Sample 11	0.21	0.42	0.85	1.80	0.72	53.1	59.5	5.8	6.53	122.2	1.04	31.3	1.96	>2419.6
Sample 12	0.25	0.39	0.97	2.38	0.85	62.6	59.7	5.1	6.78	118.1	1.36	40.5	2.30	>2419.6
Sample 13	0.53	0.45	1.08	2.72	1.40	68.6	61.4	6.4	7.03	143.1	1.19	20.7	2.62	>2419.6
Sample 14	0.07	0.40	0.24	1.26	0.10	21.6	62.8	8.4	6.37	140.1	1.60	17.4	2.48	129.1
Note:	* The values provided here are the average of two independent measurements of each sample.													

As shown in Table 1, Table 2, Table 3, Table 4, & Table 5, the measurements over the criteria are highlight in red. The elevated Phosphate (PO_4^{3-}) and *E. coli* were found in water samples collected in December, January, and March 2023. Due to extreme drought and heat during August to November 2023, the abnormal conductivities were found in December 2023. The lower DO occurs when the flow rate is small.

Water samples with both high and low *E. coli* were selected to perform Human qPCR marker analysis, results of the analysis are shown in Table 6. Sites 5, 7 and 13 showed higher human qPCR marker (levels > 525 copies/100 ml), which indicated the potential contribution of human fecal contaminants from the OSSFs operating in those areas. Further analysis at site 7 showed that qPCR remained at higher level although *E. Coli* at site 7 on 05/19 and 07/06 were below the level of concern (levels <126 MPN/100 mL). Therefore, we identified that the site 7 is a potential hotspot with failing OSSFs and should be considered for fixing the existing OSSFs as a BMP approach.

Table 6. qPCR analysis at different sites.

Date	Site	Copies/ 100 ml filter	E. Coli MPN/100 mL
7/6/2023	Sample 2	380	1413.6
11/20/2023	Sample 2	615	126.7
3/22/2024	Sample 2	2152	>2419.6
7/6/2023	Sample 4	902	>2419.6
1/30/2024	Sample 4	532	178.9
3/22/2024	Sample 4	1346	>2419.6
7/6/2023	Sample 5	4881	>2419.6
1/30/2024	Sample 5	2514	56

3/22/2024	Sample 5	78560	>2419.6
5/19/2023	Sample 7	18730	<1.0
7/6/2023	Sample 7	951	68.2
11/20/2023	Sample 7	105723	1732.9
1/30/2024	Sample 7	1825	816.4
3/22/2024	Sample 7	13511	>2419.6
7/6/2023	Sample 8	880	365.4
11/20/2023	Sample 8	687	1119.9
3/22/2024	Sample 8	2477	>2419.6
5/19/2023	Sample 13	15778	>2419.6
1/30/2024	Sample 13	1233	770.1
3/22/2024	Sample 13	17800	>2419.6

2 Results of Machine Learning (ML) Models

Models and algorithms used in this study were implemented in Python 3.7 programming language. The LSTM model was built with Pytorch 2.1 (<https://pytorch.org/>) while IG and GSHAP algorithms were implemented using Captum (<https://captum.ai>).

Three metrics, namely coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE), defined by equations **Error! Reference source not found.**, **Error! Reference source not found.** and **Error! Reference source not found.** were applied to quantitatively evaluate the performance of the proposed models. R^2 is much informative in regression analysis evaluation (Chicco et al., 2021). The value of R^2 of a best model can achieve 1, representing the predicted values exactly match the true values. The negative R^2 appears when the models have worse predictions than the baseline model

which always predicts the mean of the targets. RMSE is sensitive to outliers while MAE describes only the average magnitude of the errors with a linear rule, ignoring their direction (Willmott and Matsuura, 2006). According to their characteristics, the performance of the applied models can be sufficiently assessed by combining these metrics.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (3)$$

where, x_i is the predicted value at the i -th time step, y_i is the corresponding observation, $\bar{y}_i$ is mean of the observed values and n is number of time steps.

2.1 Results of ML DO Models

2.1.1. Current Hourly DO Prediction

As shown in Table 7, the performance of the LSTM and TL-LSTM models was superior to the MLR models for both datasets. Figure 12 demonstrates the comparison of 7 features models at both locations. The best performance (Figure 12 (b)) was achieved by the LSTM model with 7 features of the PIB dataset, and have the R^2 , RMSE and MAE were 0.982, 0.263 and 0.149, respectively. Although the R^2 values of the MLR models at PIB were acceptable ($R^2 = 0.795$), the predictions in Figure 12 (a) were not agree well by simply repeating the previous DO observations. For the SWB dataset, the R^2 values of the TL-LSTM models were 0.043 and 0.128 higher than those of the LSTM models, and 0.678

and 0.617 higher than those of the MLR models with 2 or 7 input features, respectively. This indicated that the LSTM model learned about the spatial-temporal dynamics of DO and can predict hourly DO with TL-LSTM in the Neches River.

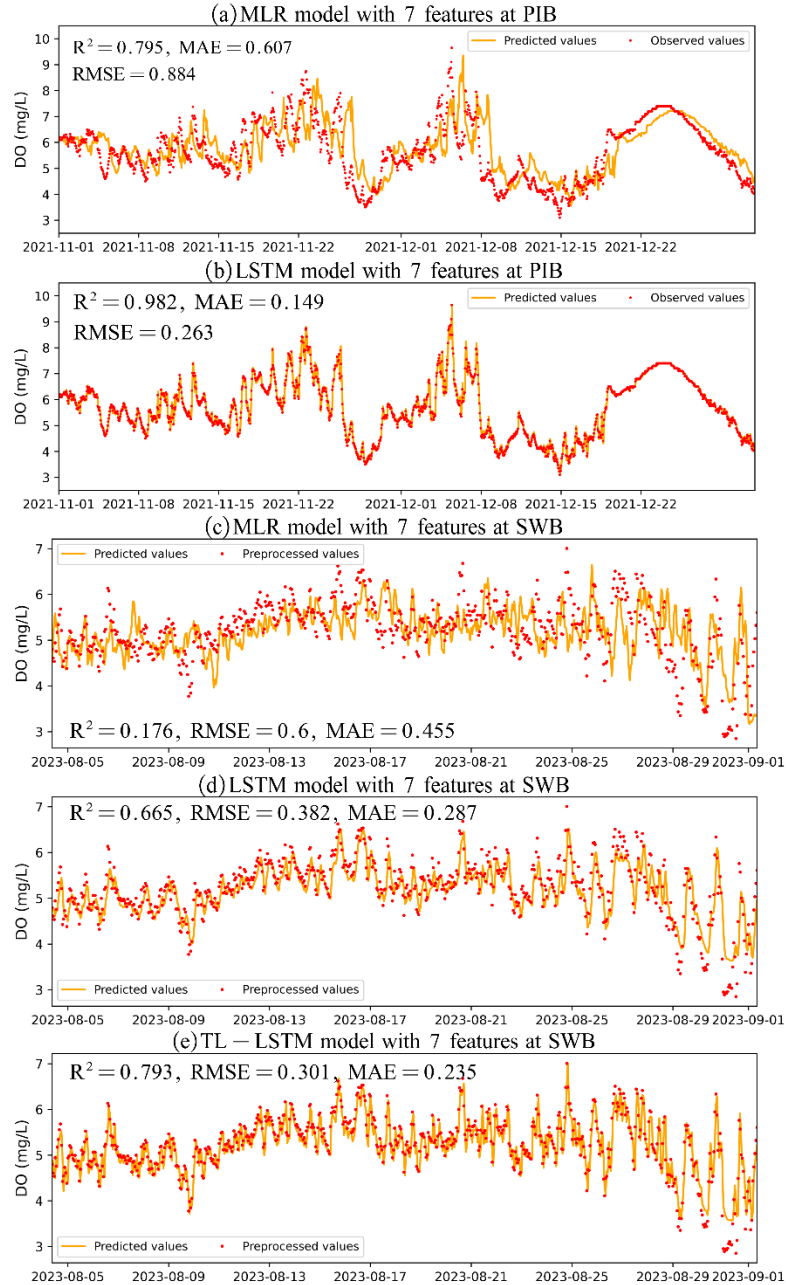


Figure 12. Comparison of models for current hourly DO predictions.

Table 7. Performance of the DO models for 1-hour prediction task.

Dataset	Model	Features	R ²	RMSE (ml/L)	MAE (ml/L)
PIB	MLR	2	0.795	0.883	0.603
		7	0.795	0.884	0.607
	LSTM	2	0.981	0.267	0.154
		7	0.982	0.263	0.149
SWB	MLR	2	0.149	0.609	0.462
		7	0.176	0.600	0.455
	LSTM	2	0.774	0.314	0.234
		7	0.665	0.382	0.287
	TL-LSTM	2	0.817	0.283	0.222
		7	0.793	0.301	0.235

2.1.2. 14-day Hourly DO Forecast

The visual comparisons of rolling forecasts with both the PIB and SWB datasets were shown in Figure 13. The outcomes summarized in Table 8 revealed that the LSTM and TL-LSTM models outperformed the MLR models with higher accuracy for the 14-day prediction task. The best models for the PIB and SWB were the LSTM model ($R^2=0.891$) and the TL-LSTM model ($R^2=0.502$), respectively. In particular, the R^2 values of the TL-LSTM models with 2 and 7 features at SWB increased by 42.5% and 23.3% compared to those of the LSTM models. These results indicated that the general knowledge of DO dynamics learned from the PIB dataset were useful for building DO models for the SWB dataset. In addition, the forecasts generated by the LSTM and TL-LSTM models with 2 input features were capable to follow the general trend of DO observations (Figure 13 (a)). Compared to 2 features, the LSTM and TL-LSTM models employing 7 features obtained

slightly higher R^2 and smaller RMSE and MAE. Moreover, more sharp changes in DO peaks and troughs were captured, particularly during the period from 03/03/2022 to 03/15/2022 at the PIB (Figure 13 (b)). Furthermore, the LSTM and TL-LSTM models maintain a satisfactory performance during the 14-day forecast applying the predicted DO as input. It demonstrates models could understand the potential relationship between DO changes and input features. On the contrary, the forecasts generated by the MLR models only oscillated around the mean of the initial 24-hour DO inputs, and its shape resembled that of the initial inputs (Figure 13 (c)). The deficient performance of the MLR model demonstrated the limitation of the MLR models on the longer period prediction.

Table 8. Performance of the DO models for 14-day hourly forecast.

Dataset	Model	Features	R^2	RMSE (ml/L)	MAE (ml/L)
PIB	MLR	2	0.238	1.510	1.133
		7	0.288	1.459	1.111
	LSTM	2	0.802	0.769	0.592
		7	0.891	0.572	0.327
SWB	MLR	2	-0.217	0.670	0.513
		7	-0.109	0.639	0.487
	LSTM	2	0.334	0.496	0.387
		7	0.407	0.468	0.363
	TL-LSTM	2	0.476	0.440	0.346
		7	0.502	0.428	0.337

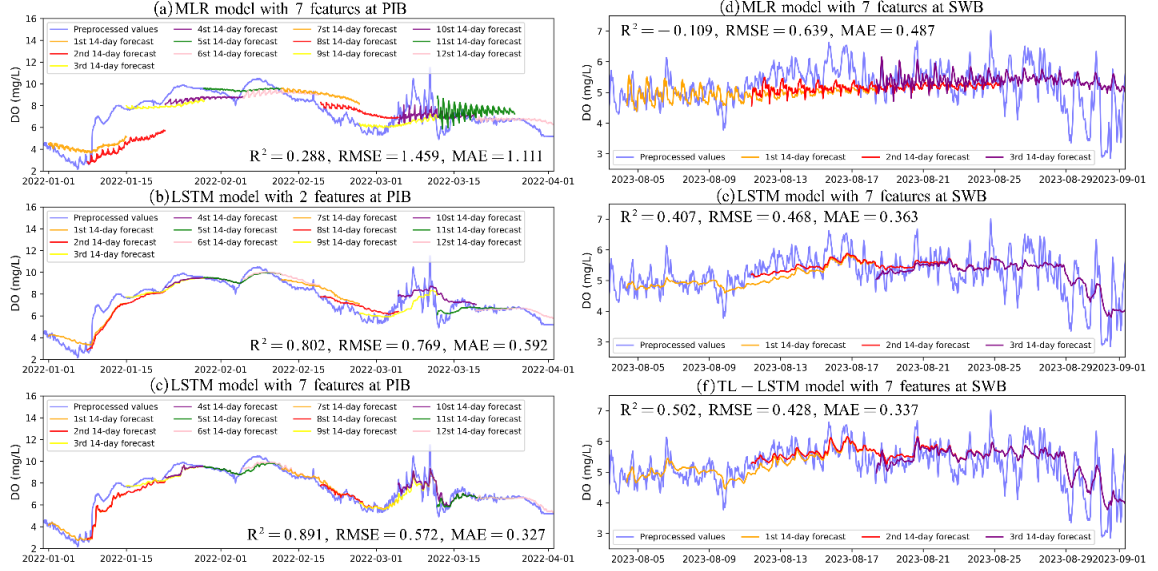


Figure 13. Comparison of models for 14-day hourly DO forecasts.

2.2 Results of ML Turbidity Model

Table 9 shows the performance of the Attention-based Encoder & Decoder model for the three split datasets. The three errors were calculated with predicted values of the next time-step (1 hour) and the next 120 time-steps (5 days), respectively. The test set obtained the best MAE, MAPE, and RMSE equal to 0.887, 3.74%, and 1.175 for the next time-step, while 2.133, 8.837%, and 7.853 for total 5-day predicted values, respectively. Considering the shortest length of the test set, the model performance was also evaluated on the training and validation sets and the results still showed acceptable error levels. The predictions of the next step for the test set are visualized in Figure 14. **Error! Reference source not found..** The predictions were close to the historical values and captured the fluctuation caused by tides. In order to show our proposed methods more intuitively and

clearly, some of 5-day predicted results for the three datasets are plotted in Figure 15. Although the predicted sequential values did not manage the sharp changes in hourly data, they captured the general trend of turbidity that could give more suggestions than the next time-step prediction for the public.

Table 9. A comparison of the errors of the training set, the valid set, and the test set.

Dataset	Calculated prediction length	Metrics		
		MAE (NTU)	MAPE (%)	RMSE (NTU)
Training set	Next time-step	1.707	5.064	7.357
	All 120 time-steps	2.363	7.146	22.733
Valid set	Next time-step	1.782	6.929	6.181
	All 120 time-steps	3.055	11.410	15.293
Test set	Next time-step	0.887	3.740	1.175
	All 120 time-steps	2.133	8.837	7.853

According to earlier research, most studies focus on the one-time-step turbidity prediction with machine learning models. Wang et al. (Wang et al., 2021) used tidal average significant wave height and tidal range to predict hourly turbidity by Artificial Neural Network (ANN), Support Vector Machine (SVM), and Genetic Programming (GP) at a coastal bay. The best RMSE is equal to 10.83 obtained by the ANN model. Zhang et al. (Zhang et al., 2021) used wind field, air temperature, and rainfall data to predict hourly turbidity in a lake by the Random Forest (RF) model, achieving 39.69% MAPE. Rele et al (Rele et al., 2023) built an Autoregressive integrated moving average (ARIMA), Long Short-Term Memory (LSTM), and Generalized additive model (GAM) to predict daily

turbidity in a river according to the previous 30-day turbidity, rainfall, water level, temperature, and total global solar exposure data. The MAE and RMSE of the three models ranged from 5.91 to 11.5 and 9.46 to 17.27, respectively. Then they proposed a meta-model, integrated ARIMA, LSTM, and GAM, which can auto select the optimal model at each time step and improve the performance to 1.67 MAE and 2.31 RMSE. Compared with their results, the errors of the next time-step in this study were relatively small. The errors of all time-steps were much higher than those of the one-time-step prediction, but they were still compatible and useful to provide a 5-day trend of turbidity changes for practical application.

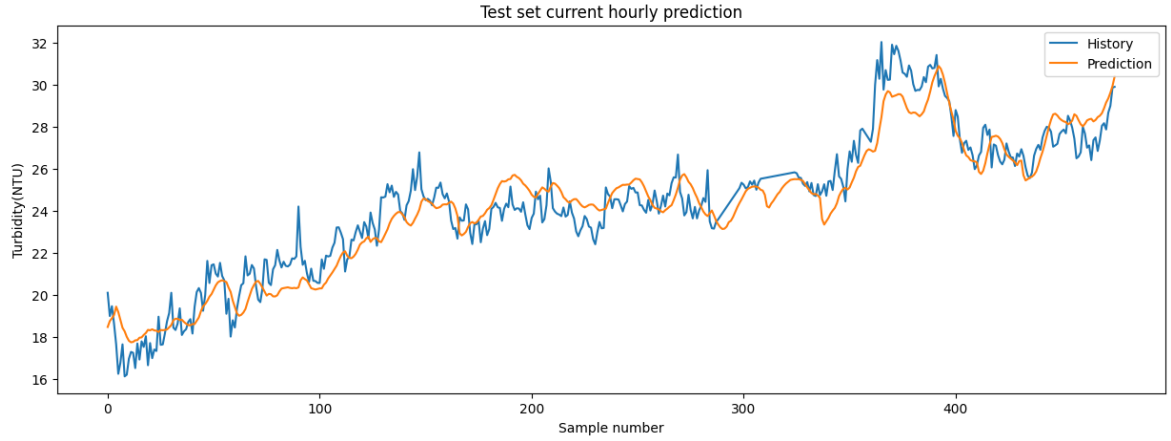


Figure 14. Visualization of the next time-step (1 hour) prediction of the test set.

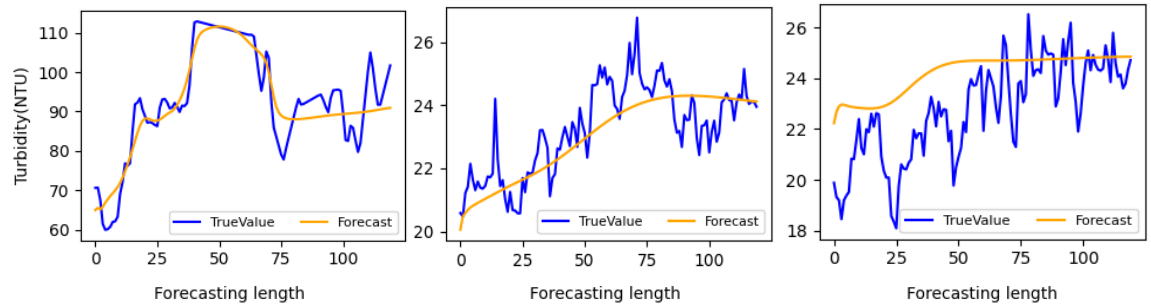


Figure 15. Visualization of 5-day forecasting results of the training set, valid set, and test set.

3 Results of SWMM & MOPUS model

The watersheds modeled by SWMM were delineated from 1-m Digital Elevation Model (DEM) using tools in the HEC-HMS modeling suite. The drainages thus derived then underwent manual adjustments using Google Earth Pro. (Figure 16)

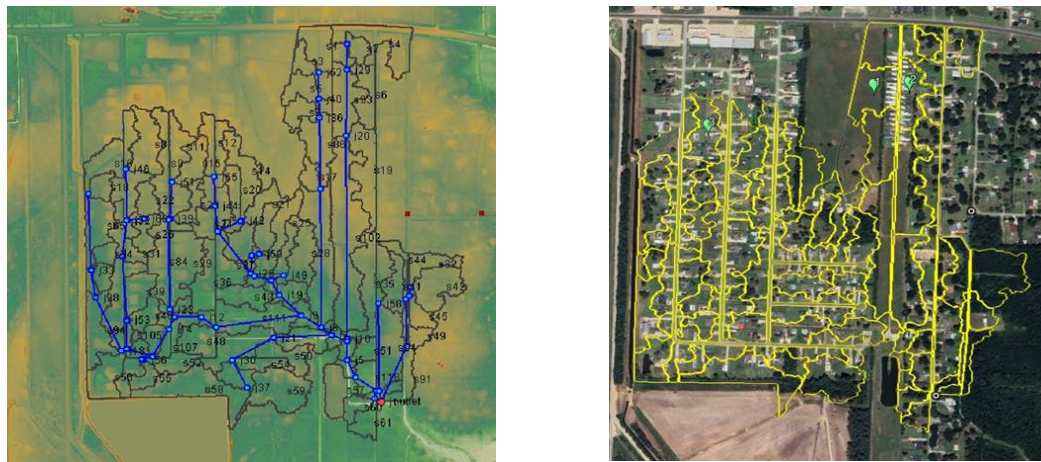


Figure 16. Watershed delineated using HEC-HMS (left) and after manual editing (right).

The final watershed boundaries and flow network were integrated into EPA-SWMM to create schematics of drainage (see an example in Figure 17). The area and width of the subcatchments were determined from Google Earth Pro. For simplicity, each watershed was assumed to comprise 10% impervious land and 5% zero impervious land. Default values were used for the rest of the parameters. The schematization of subcatchments at each site can be found in Appendix A.

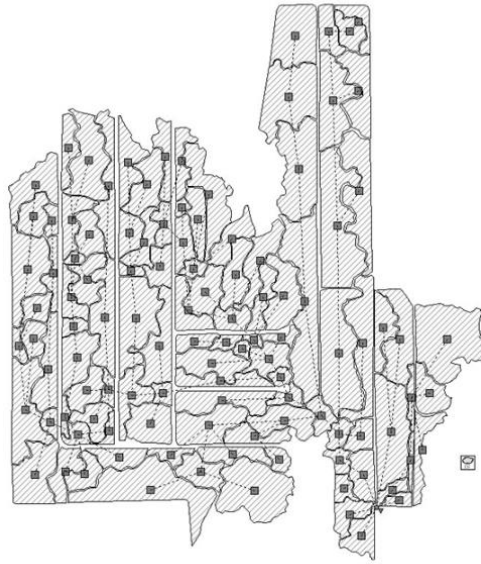


Figure 17. Schematization of watersheds in EPA-SWMM for sites 12-14.

The models were run to produce runoff simulations for over 1 year (1 January 2023 – 31 March 2024). The baseline simulations were done using rainfall data collected from CoCoHaRS website (Community Collaborative Rain, Hail and Snow Network, Figure 18), and the models underwent light calibration by adjusting Manning's n . In the scenario analysis, design rainfall was generated for return periods 25-year, 50-year and 100-year following Kiprich's method. The model is simulated for these design storms and corresponding runoffs are generated.

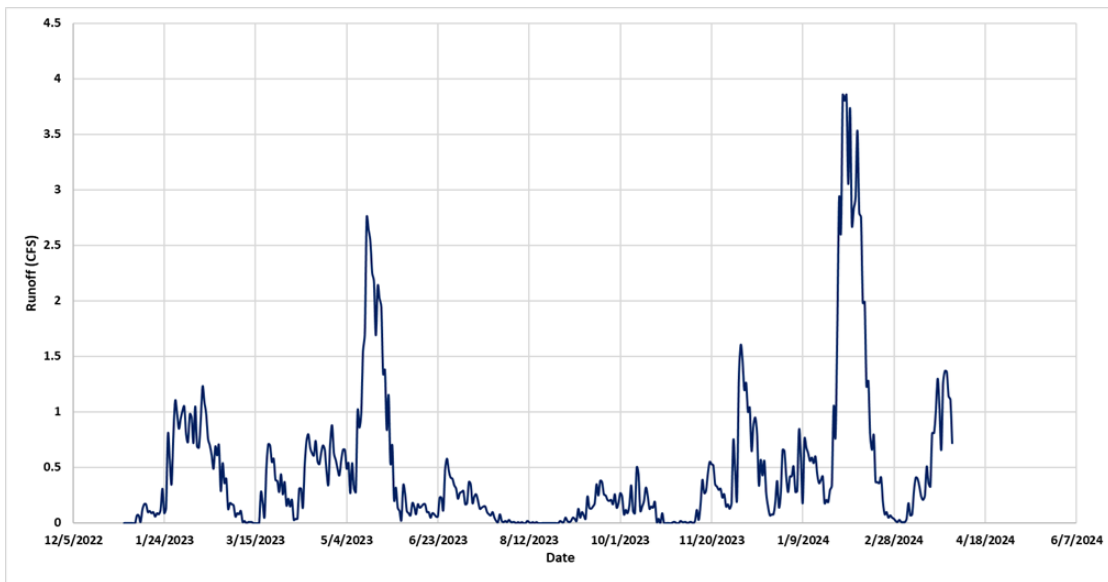
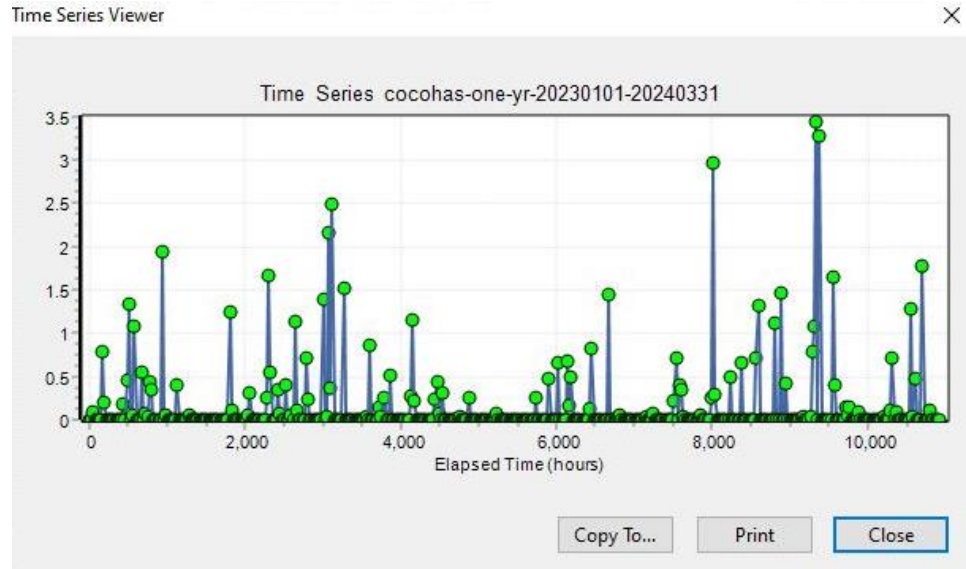


Figure 18. Example time series of rainfall from the CocoHaRS station and runoff simulations by EPA-SWMM.

MOPUS model is based on two conceptual equations shown in Figure 19. These equations incorporate two empirical coefficients. In this study, these coefficients were

adjusted to allow the maximum bacteria concentration from model simulations to match the observations collected by Lamar University (Figure 19, lower right panel).

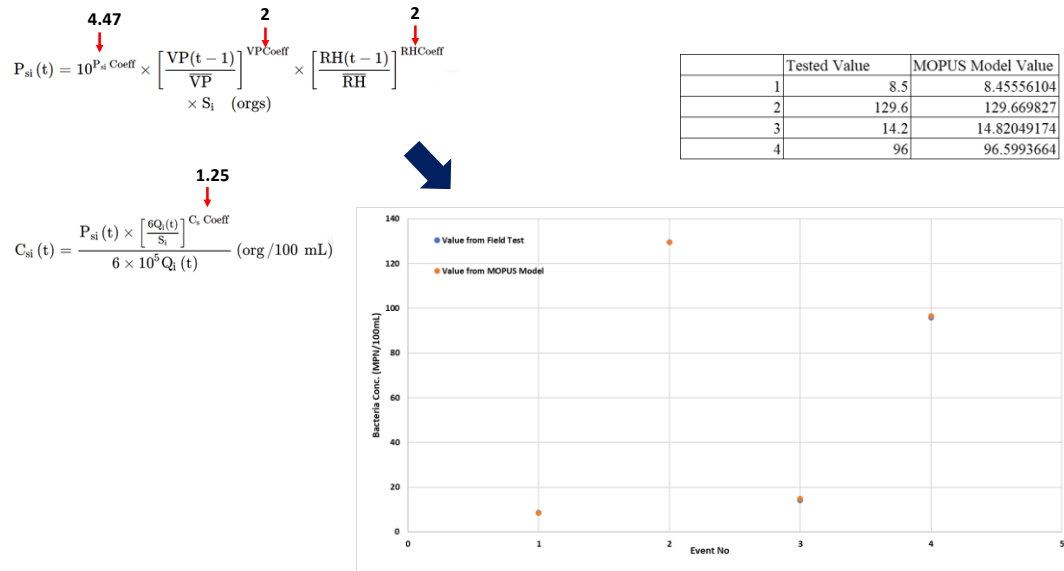


Figure 19. MOPUS Model with calibrated coefficients based on observed bacterial concentrations.

The impacts of BMPs on simulated peak bacterial concentrations are summarized in Table 10, where aggregate peak bacterial simulations by the coupled modeling system without and with BMPs are compared for each of the 14 sites. The percentage reduction in runoff ranges from 0% at site 14 to 26% at site 6.

Table 10. Reduction in runoff peaks after BMP implementation for 2023-2024, with default BMP configuration.

Sites #	Runoff-Baseline [cfs]	Runoff – BMP [cfs]	Reduction [%]
1	3.60	3.40	5.56
2	2.17	2.00	7.83
3	0.81	0.80	1.23

4	10.71	10.30	3.83
5	2.22	2.10	5.41
6	1.50	1.40	6.67
7	12.94	12.40	4.17
8	1.51	1.40	7.28
9	19.01	18.20	4.26
10	4.44	4.20	5.41

The impacts of BMPs on simulated peak bacterial concentrations are summarized in Table 11, where aggregate peak bacterial simulations by the coupled modeling system without and with BMPs are compared for each of the 14 sites. The percentage reduction in runoff ranges from 0% at site 14 to 26% at site 6.

Table 11. Reduction in bacteria concentrations after BMP implementation for 2023-2024, with default BMP configuration.

Site #	Max Bacteria Conc.- Baseline	Max Bacteria Conc.-BMP	Maximum % Reduction
1	27	25	6
2	309	294	5
3	367	365	1
4	400	366	8
5	302	295	2
6	61	45	26
7	926	903	3
8	2381	2271	5
9	1542	1255	19
10	67	63	6
11	356	339	5
12	139	133	4
13	19667	15885	19
14	99	98	0

In order to assess the dependence of efficacy of BMPs on their size and infiltration properties, three hypothetical BMP configurations were created with different Berm height and soil permeability (Table 12).

Table 12. Three BMP configuration used in sensitivity analysis.

		BMP 1	BMP 2	BMP 3
Berm height		100	100	1000
Biofiltration Media Thickness		650	650	650
Soil Mix Property	Porosity	0.5	0.5	0.6
	Field capacity	0.2	0.4	0.4
Underdrain/ gravel base height		650	650	650

The outcomes from the sensitivity analysis are shown in Table 13. Increasing capacity of the BMP results in increased percentage reduction in runoff, and smaller changes in the reduction in bacteria loading.

Table 13. Percentage reduction in runoff and bacteria concentration with each hypothetical BMP configuration.

BMP-1		BMP-2		BMP-3	
% reduction		% reduction		% reduction	
Runoff	Bacteria Conc	Runoff	Bacteria Conc	Runoff	Bacteria Conc
24	26	25	26	24	26
1	26	1	26	1	26
74	29	45	14	74	29
54	20	48	16	54	20
0	0	0	0	0	0
44	10	11	2	11	2
42	7	26	7	34	7
48	49	48	28	48	49
35	12	18	6	24	8
0	17	0	17	0	17
11	3	6	2	11	3

17	5	17	5	17	5
39	28	39	27	39	28
10	3	10	3	10	3

Chapter IV

Summary

Five runs of water sampling and water quality measurements have been conducted to assess the bacteria pollution in the drainage system after the main rain events. The results demonstrate the elevated nutrient and *E.coli* were observed in the system. The bacteria source tracking showed that the Human qPCR marker at sites 5, 7, 13 were higher than the criteria, which indicated the potential contribution of human fecal contaminants from the OSSFs. To understand how the bacteria from OSSF wastewater affect the water quality in Neches River, the outfalls from sample sites were identified. Along the Neches River, site 14 outfall is upstream of TCEQ 20774; sample sites 1&2 outfalls are upstream at TCEQ 10575, sample sites 3,4,5&6 outfalls are upstream of TCEQ 10570, and site 7 outfall is upstream of TCEQ 10566. The samples sites 8&9&10 discharges to the Bessie Heights Oil and Gas Field before it reaches the TCEQ 10563. The sample sites 11&12&13 discharge to the Old River Cove before it reaches the Sabine Lake. The elevated enterococci at all monitoring sites in January 2023, and at TCEQ 10575 in year 2023, at TCEQ 10575 in January and October. The water quality in Neches River response to the nonpoint source we found in the drainage system accordingly. Therefore, the failing OSSFs is one of the bacteria pollution origins.

Analyzing the YSI water quality data collected at the SWB with innovative machine learning technologies, we developed the LSTM models with and without transfer learning method to predict hourly DO for current and consistent 14-day in rivers to support water quality management. The contributions of this study were concluded as following: (1) This was the first study to investigate the ability of the LSTM model to predict hourly DO concentrations up to 14 days under different hydro-meteorological conditions, The LSTM model performed much better than the traditional MLR model, where its R^2 achieved up to 0.98 and 0.89 for current and 14-day hourly DO forecasts, respectively; (2) The efficacy of transfer learning algorithm was proved in solving the problem of insufficient data volume for the LSTM model. The relationship between DO and input features obtained by the large dataset can enhance the performance of the TL-LSTM model with the small target dataset although the improvement had limits due to different environmental conditions at two locations; (3) Previous DO was identified as the most important feature at both locations and water temperature was also critical for forecasting the general trend of DO changes within 14 days. The Attention-based Encoder & Decoder model associated with the GRU was applied to predict hourly turbidity in the following 5 days with hydro-meteorology measurements. Compared with earlier research, the MAE, MAPE, and RMSE of the next time-step predictions are acceptable. This model also can provide the general trend of turbidity changes within 5 days. The capability of forecasting 5-day hourly turbidity can be applied to local water management systems with 5-day gage height and rainfall forecasting data from NOAA.

A coupled EPA-SWMM/MOPUS system was established to help assess the potential of infiltration BMPs, specifically bioretention cells, in alleviating the bacteria loading problems if installed in the upstream of watersheds with high concentration of OSSFs. The EPA-SWMM model was implemented for 14 sites in different tributaries of the Neches River. For each site, the model was used to simulate runoff using observed rainfall over a 1-year window, and the resulting runoff time series was ingested into the MOPUS system to create time series of bacteria loading. Using a default configuration of bioretention cell leads to 1-88% reductions in peak simulated runoff, and 0-26% reductions in peak bacteria loading. Sensitivity experiments suggest that increasing capacity of the BMP leads to larger reductions in peak runoff but relatively modest reductions in bacteria loading.

REFERENCE

- Abba, S.I., Linh, N.T.T., Abdullahi, J., Ali, S.I.A., Pham, Q.B., Abdulkadir, R.A., Costache, R., Nam, V.T., Anh, D.T., 2020. Hybrid Machine Learning Ensemble Techniques for Modeling Dissolved Oxygen Concentration. *IEEE Access* 8, 157218–157237. <https://doi.org/10.1109/ACCESS.2020.3017743>
- Ayesha Jasmin, S., Ramesh, P., Tanveer, M., 2022. An intelligent framework for prediction and forecasting of dissolved oxygen level and biofloc amount in a shrimp culture system using machine learning techniques. *Expert Systems with Applications* 199, 117160. <https://doi.org/10.1016/j.eswa.2022.117160>
- Azma, A., Liu, Y., Azma, M., Saadat, M., Zhang, D., Cho, J., Rezania, S., 2023. Hybrid machine learning models for prediction of daily dissolved oxygen. *Journal of Water Process Engineering* 54, 103957. <https://doi.org/10.1016/j.jwpe.2023.103957>
- Bahdanau, D., Cho, K., Bengio, Y., 2016. Neural Machine Translation by Jointly Learning to Align and Translate.
- Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science* 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, in: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Presented at the *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, Doha, Qatar, pp. 1724–1734. <https://doi.org/10.3115/v1/D14-1179>
- Csábrági, A., Molnár, S., Tanos, P., Kovács, J., 2017. Application of artificial neural networks to the forecasting of dissolved oxygen content in the Hungarian section of the river Danube. *Ecological Engineering* 100, 63–72. <https://doi.org/10.1016/j.ecoleng.2016.12.027>
- Du, S., Li, T., Yang, Y., Horng, S.-J., 2020. Multivariate time series forecasting via attention-based encoder–decoder framework. *Neurocomputing* 388, 269–279. <https://doi.org/10.1016/j.neucom.2019.12.118>
- Espejo-Garcia, B., Mylonas, N., Athanasakos, L., Fountas, S., Vasilakoglou, I., 2020. Towards weeds identification assistance through transfer learning. *Computers and Electronics in Agriculture* 171, 105306. <https://doi.org/10.1016/j.compag.2020.105306>
- Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J., 2017. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learning Syst.* 28, 2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>

- Huan, J., Li, H., Li, M., Chen, B., 2020. Prediction of dissolved oxygen in aquaculture based on gradient boosting decision tree and long short-term memory network: A study of Chang Zhou fishery demonstration base, China. *Computers and Electronics in Agriculture* 175, 105530. <https://doi.org/10.1016/j.compag.2020.105530>
- Jiang, J., Men, Y., Pang, T., Tang, S., Hou, Z., Luo, M., Sun, X., Wu, J., Yadav, S., Xiong, Y., Liu, C., Zheng, Y., 2023. An integrated supervision framework to safeguard the urban river water quality supported by ICT and models. *Journal of Environmental Management* 331, 117245. <https://doi.org/10.1016/j.jenvman.2023.117245>
- Kim, K.-M., Ahn, J.-H., 2022. Machine learning predictions of chlorophyll-a in the Han river basin, Korea. *Journal of Environmental Management* 318, 115636. <https://doi.org/10.1016/j.jenvman.2022.115636>
- Leigh, C., Kandanaarachchi, S., McGree, J.M., Hyndman, R.J., Alsibai, O., Mengersen, K., Peterson, E.E., 2019. Predicting sediment and nutrient concentrations from high-frequency water-quality data. *PLOS ONE* 14, e0215503. <https://doi.org/10.1371/journal.pone.0215503>
- Li, W., Fang, H., Qin, G., Tan, X., Huang, Z., Zeng, F., Du, H., Li, S., 2020. Concentration estimation of dissolved oxygen in Pearl River Basin using input variable selection and machine learning techniques. *Science of The Total Environment* 731, 139099. <https://doi.org/10.1016/j.scitotenv.2020.139099>
- Liu, P., Wang, J., Sangaiah, A., Xie, Y., Yin, X., 2019. Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment. *Sustainability* 11, 2058. <https://doi.org/10.3390/su11072058>
- Liu, Y., Zhang, Q., Song, L., Chen, Y., 2019. Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction. *Computers and Electronics in Agriculture* 165, 104964. <https://doi.org/10.1016/j.compag.2019.104964>
- Lumini, A., Nanni, L., 2019. Deep learning and transfer learning features for plankton classification. *Ecological Informatics* 51, 33–43. <https://doi.org/10.1016/j.ecoinf.2019.02.007>
- McCarthy, D.T., Deletic, A., Mitchell, V.G., Diaper, C., 2011. Development and testing of a model for Micro-Organism Prediction in Urban Stormwater (MOPUS). *Journal of Hydrology* 409, 236–247. <https://doi.org/10.1016/j.jhydrol.2011.08.023>
- Nong, X., Lai, C., Chen, L., Shao, D., Zhang, C., Liang, J., 2023. Prediction modelling framework comparative analysis of dissolved oxygen concentration variations using support vector regression coupled with multiple feature engineering and optimization methods: A case study in China. *Ecological Indicators* 146, 109845. <https://doi.org/10.1016/j.ecolind.2022.109845>
- Pizano-Torres, R.I., Roach, K.A., Winemiller, K.O., 2017. Response of the fish assemblage to a saltwater barrier and paper mill effluent in the Lower Neches River (Texas)

- during drought. *Journal of Freshwater Ecology* 32, 147–162. <https://doi.org/10.1080/02705060.2016.1253622>
- Qian, Q., He, M., Sun, F., Liu, X., 2024. Monitoring and evaluation of the water quality of the Lower Neches River, Texas, USA. *Water Science and Engineering* 17, 21–32. <https://doi.org/10.1016/j.wse.2023.10.002>
- Qian, Q., Sun, B., Li, X., Sun, F., Lin, C.-J., Jiang, L., 2019. Water Quality Evaluation on an Urban Stormwater Retention Pond Using Wireless Sensor Networks and Hydrodynamic Modeling. *J. Irrig. Drain Eng.* 145, 05019011. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0001434](https://doi.org/10.1061/(ASCE)IR.1943-4774.0001434)
- Rele, B., Hogan, C., Kandanaarachchi, S., Leigh, C., 2023. Short-term prediction of stream turbidity using surrogate data and a meta-model approach: A case study. *Hydrological Processes* 37, e14857. <https://doi.org/10.1002/hyp.14857>
- TCEQ, 2022. Quality Assurance Project Plan for the Continuous Water Quality Monitoring Network Program [WWW Document]. Texas Commission on Environmental Quality. URL <https://www.tceq.texas.gov/downloads/water-quality/monitoring/quality-assurance/cwqmn-qapp.pdf> (accessed 11.14.23).
- TCEQ, 2018. Pine Island Bayou C749 [WWW Document]. Texas Commission on Environmental Quality. URL https://www.tceq.texas.gov/cgi-bin/compliance/monops/water_site_photo.pl?cams=749 (accessed 11.14.23).
- TCEQ, 2010. Texas Administrative Code [WWW Document]. URL [https://texreg.sos.state.tx.us/public/readtac\\$ext.TacPage?sl=T&app=9&p_dir=N&p_rloc=210363&p_tloc=9995&p_ploc=1&pg=9&p_tac=&ti=30&pt=1&ch=307&rl=1](https://texreg.sos.state.tx.us/public/readtac$ext.TacPage?sl=T&app=9&p_dir=N&p_rloc=210363&p_tloc=9995&p_ploc=1&pg=9&p_tac=&ti=30&pt=1&ch=307&rl=1) (accessed 6.14.24).
- Teixeira, L.C., Mariani, P.P., Pedrollo, O.C., Dos Reis Castro, N.M., Sari, V., 2020. Artificial Neural Network and Fuzzy Inference System Models for Forecasting Suspended Sediment and Turbidity in Basins at Different Scales. *Water Resour Manage* 34, 3709–3723. <https://doi.org/10.1007/s11269-020-02647-9>
- Tiyasha, T., Tung, T.M., Bhagat, S.K., Tan, M.L., Jawad, A.H., Mohtar, W.H.M.W., Yaseen, Z.M., 2021. Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models. *Marine Pollution Bulletin* 170, 112639. <https://doi.org/10.1016/j.marpolbul.2021.112639>
- Tsai, T.-M., Yen, P.-H., 2017. GMDH algorithms applied to turbidity forecasting. *Appl Water Sci* 7, 1151–1160. <https://doi.org/10.1007/s13201-016-0458-4>
- TSHA, 1995a. Neches River [WWW Document]. Texas State Historical Association. URL <https://www.tshaonline.org/handbook/entries/neches-river> (accessed 11.14.23).
- TSHA, 1995b. Pine Island Bayou [WWW Document]. Texas State Historical Association. URL <https://www.tshaonline.org/handbook/entries/pine-island-bayou> (accessed 11.14.23).

- Wang, Y., Chen, J., Cai, H., Yu, Q., Zhou, Z., 2021. Predicting water turbidity in a macro-tidal coastal bay using machine learning approaches. *Estuarine, Coastal and Shelf Science* 252, 107276. <https://doi.org/10.1016/j.ecss.2021.107276>
- Willmott, C.J., Matsuura, K., 2006. On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *International Journal of Geographical Information Science* 20, 89–102. <https://doi.org/10.1080/13658810500286976>
- Yang, F., Moayedi, H., Mosavi, A., 2021. Predicting the Degree of Dissolved Oxygen Using Three Types of Multi-Layer Perceptron-Based Artificial Neural Networks. *Sustainability* 13, 9898. <https://doi.org/10.3390/su13179898>
- Zeller, T.L., Metzger, L.M., 2013. Good Bye Traditional Budgeting, Hello Rolling Forecast: Has The Time Come? *AJBE* 6, 299–310. <https://doi.org/10.19030/ajbe.v6i3.7810>
- Zhang, J., Du, J., Dai, L., 2017. A GRU-Based Encoder-Decoder Approach with Attention for Online Handwritten Mathematical Expression Recognition, in: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Presented at the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), IEEE, Kyoto, pp. 902–907. <https://doi.org/10.1109/ICDAR.2017.152>
- Zhang, Y., Yao, X., Wu, Q., Huang, Y., Zhou, Z., Yang, J., Liu, X., 2021. Turbidity prediction of lake-type raw water using random forest model based on meteorological data: A case study of Tai lake, China. *Journal of Environmental Management* 290, 112657. <https://doi.org/10.1016/j.jenvman.2021.112657>
- Zhi, W., Feng, D., Tsai, W.-P., Sterle, G., Harpold, A., Shen, C., Li, L., 2021. From Hydrometeorology to River Water Quality: Can a Deep Learning Model Predict Dissolved Oxygen at the Continental Scale? *Environ. Sci. Technol.* 55, 2357–2368. <https://doi.org/10.1021/acs.est.0c06783>
- Zhu, N., Ji, X., Tan, J., Jiang, Y., Guo, Y., 2021. Prediction of dissolved oxygen concentration in aquatic systems based on transfer learning. *Computers and Electronics in Agriculture* 180, 105888. <https://doi.org/10.1016/j.compag.2020.105888>

APPENDIX 1: Schematics of watersheds in EPA-SWMM

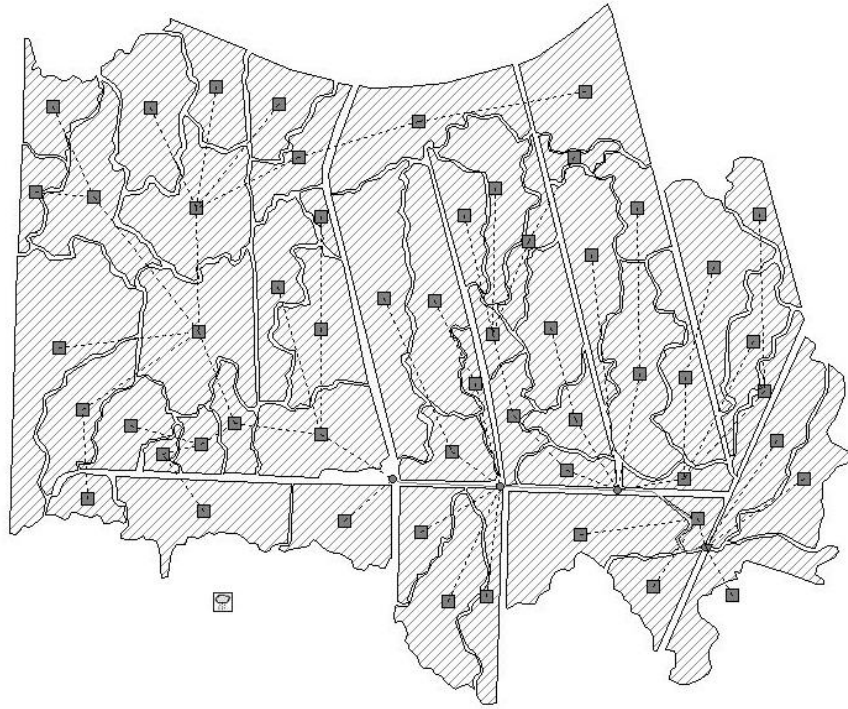


Figure A-1: Schematization of watersheds in EPA-SWMM for sites 1 and 2.

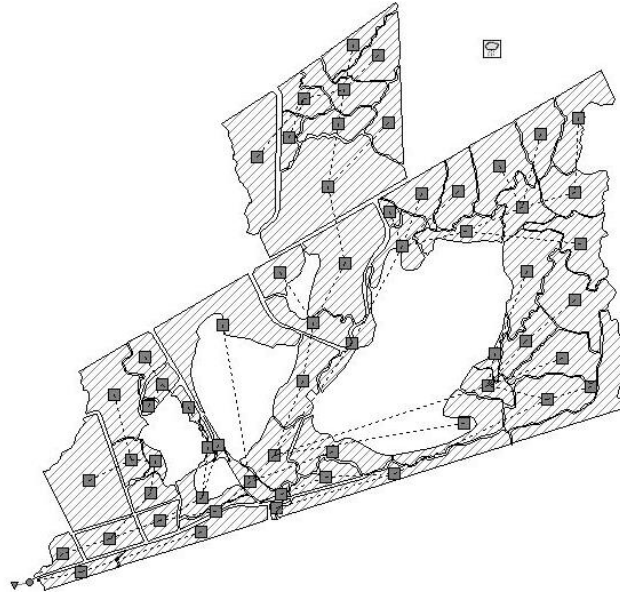


Figure A-2: Schematization of watersheds in EPA-SWMM for site 3.

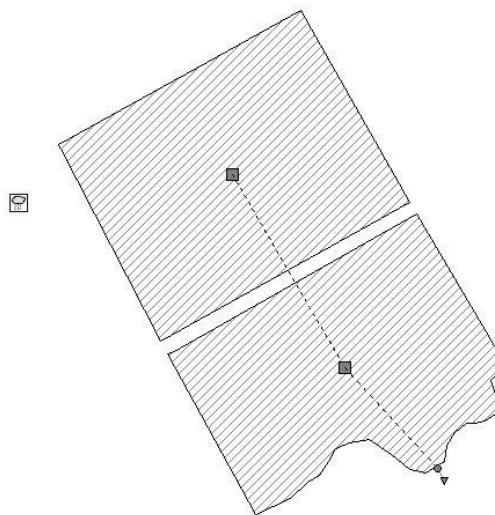


Figure A-3: Schematization of watersheds in EPA-SWMM for site 4.

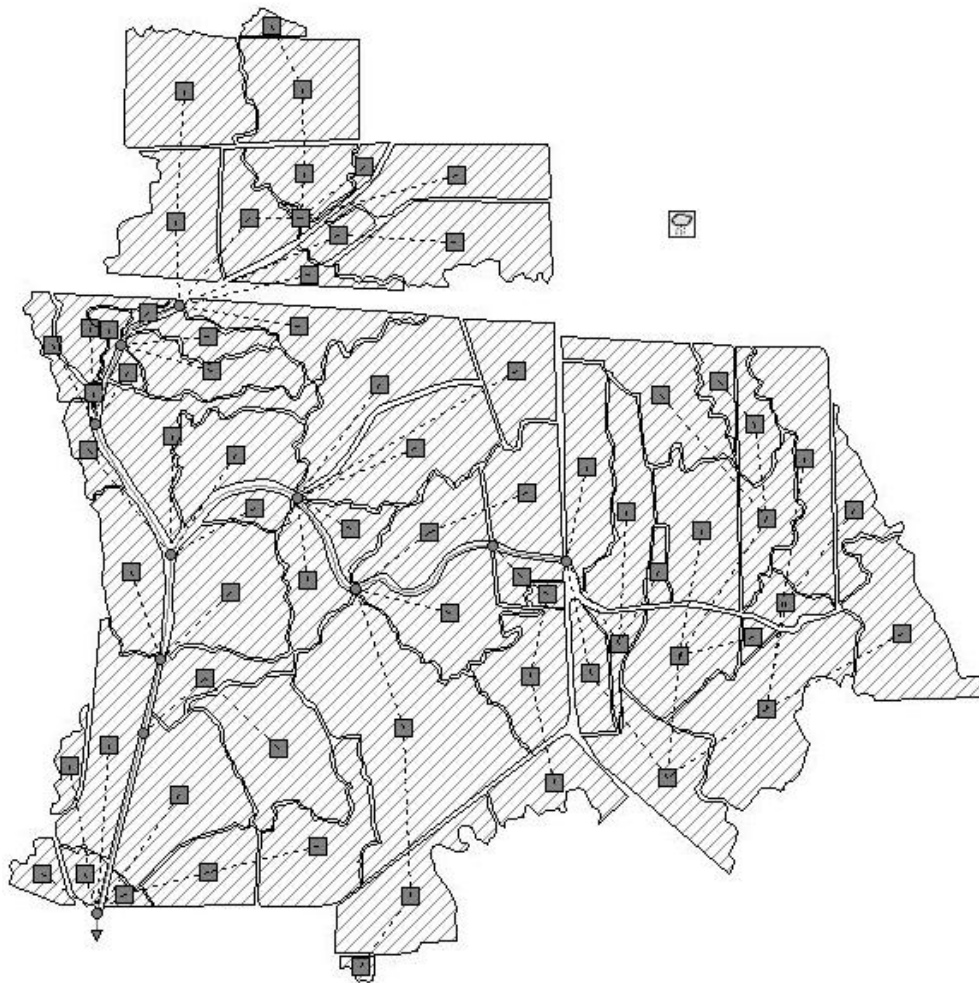


Figure A-4: Schematization of watersheds in EPA-SWMM for site 5.



Figure A-5: Schematization of watersheds in EPA-SWMM for site 6.

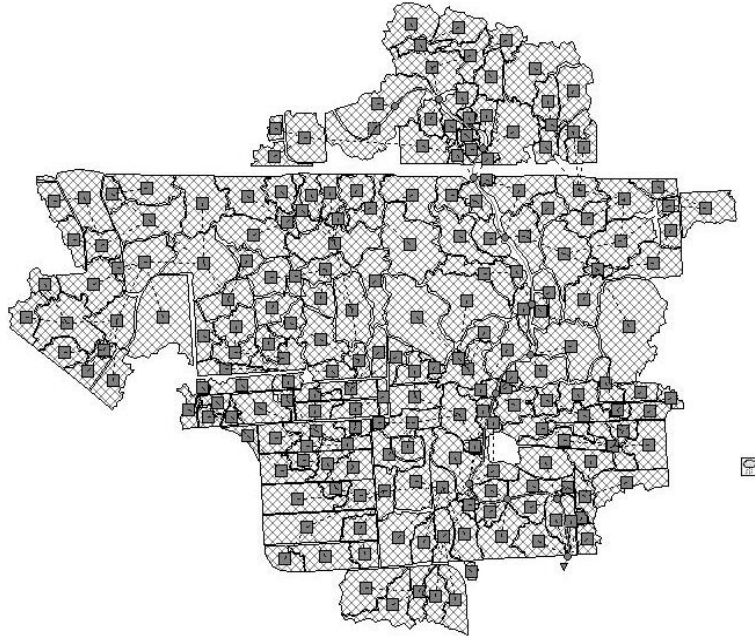


Figure A-6: Schematization of watersheds in EPA-SWMM for site 7.

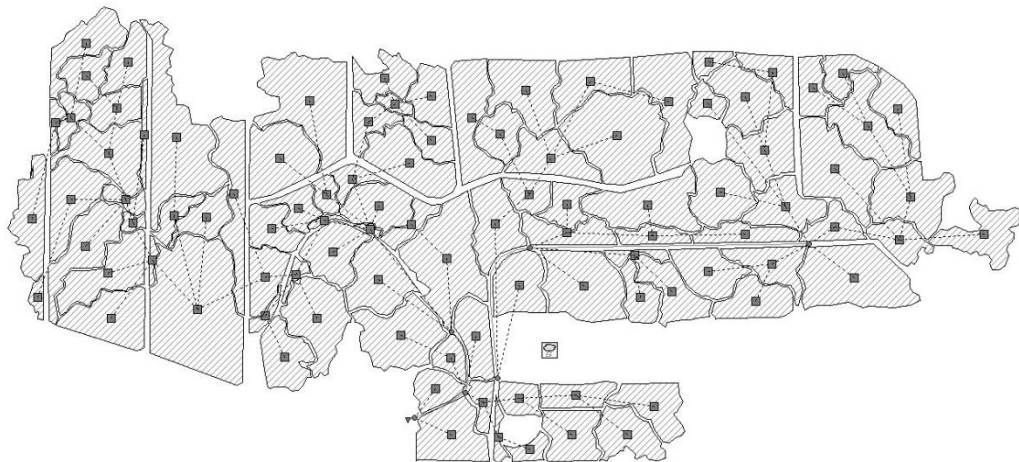


Figure A-7: Schematization of watersheds in EPA-SWMM for site 8.

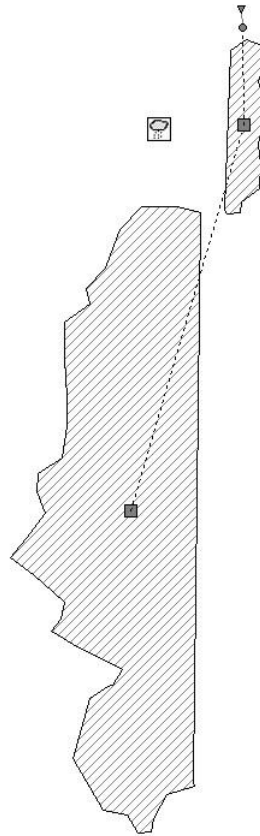


Figure A-8: Schematization of watersheds in EPA-SWMM for site 9.

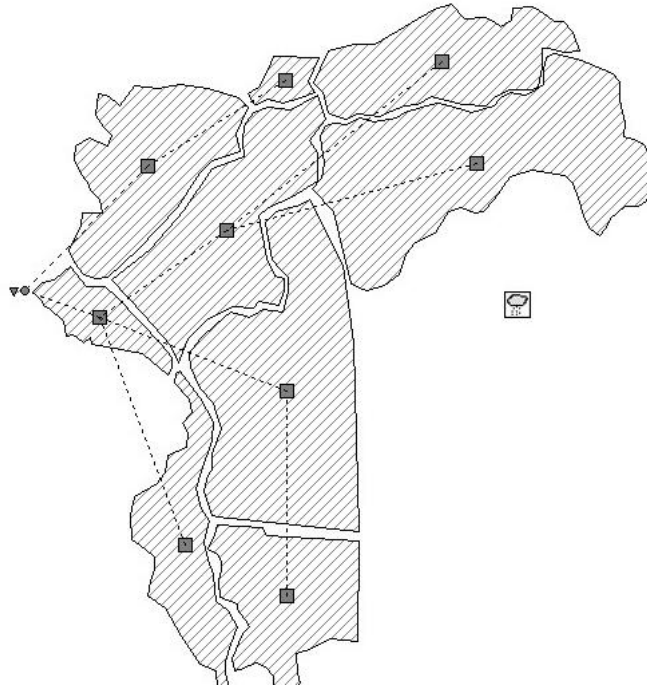


Figure A-9: Schematization of watersheds in EPA-SWMM for site 10.

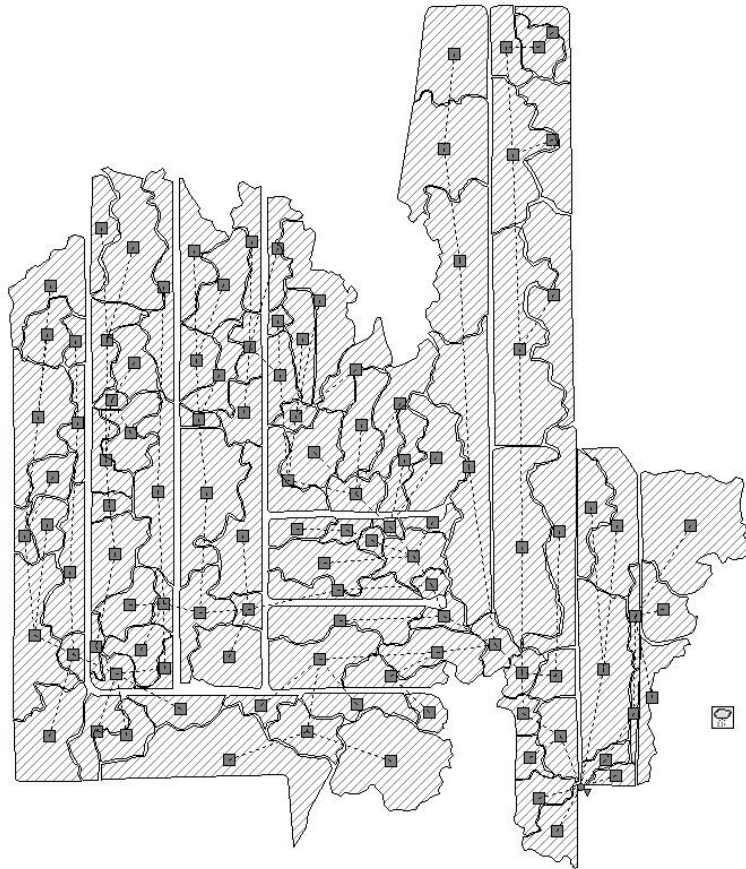


Figure A-10: Schematization of watersheds in EPA-SWMM for site 11.

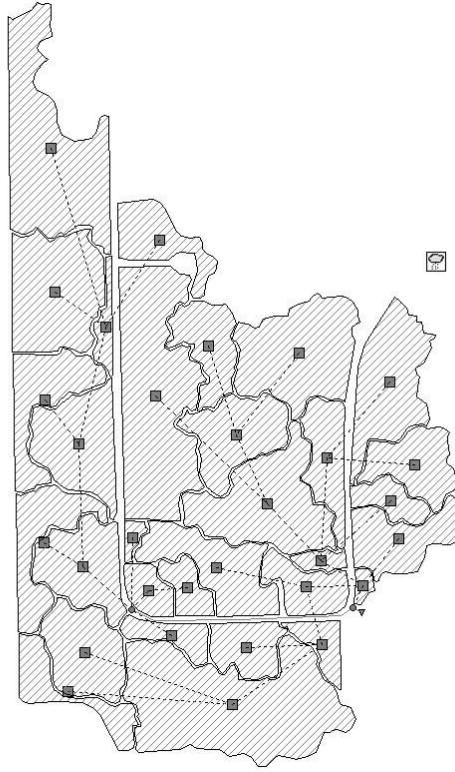


Figure A-11: Schematization of watersheds in EPA-SWMM for sites 12 and 13.

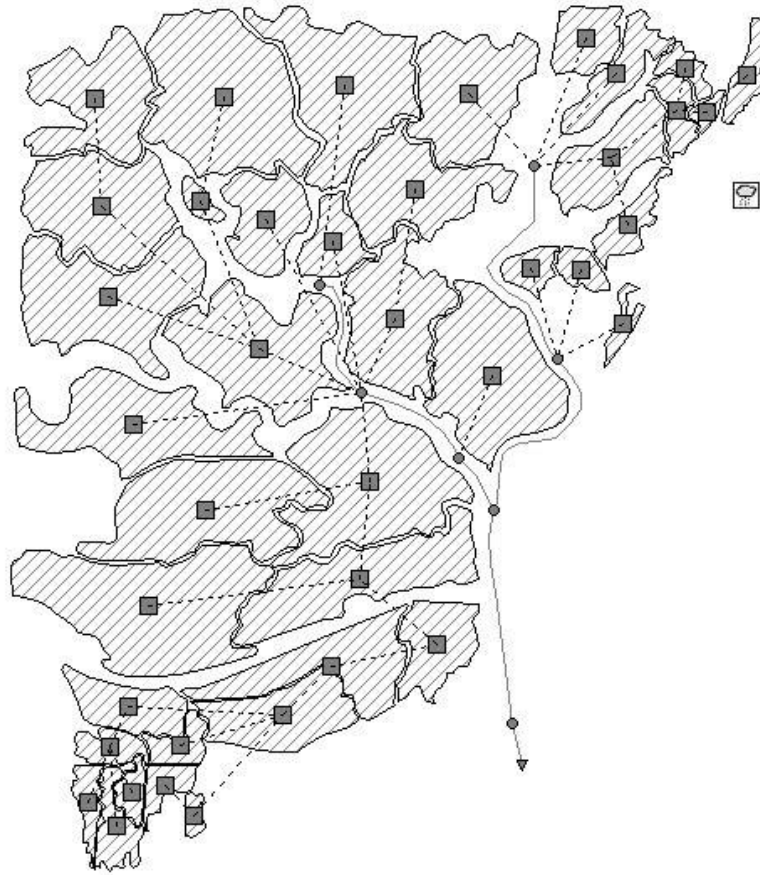


Figure A-12: Schematization of watersheds in EPA-SWMM for site 14.

APPENDIX B: Pictures for outreach and education activities



Fig. B_1: Research team met the local groups to discuss the survey locations (March 2023)



Fig. B_2: The student collected the water sample in the drainage ditch.



Fig. B_3: TOWA representative (Randy Chelette) demonstrated to students how to collect samples from the creek.



Fig. B_4: Presentation at ASCE EWRI 2024 conference at, *Milwaukee, Wisconsin in May 2024*



Fig B_4: Presented our field measurement to high school summer camp students in June 2024.



Fig B_5: high school summer camp students listened to our presentation in June 2024