A Surrogate Modeling Framework for Compound Flood Risk and Optimization Analysis

Investigating the application of LSTM models for assessing compound flood mitigation designs at Clear Lake, Texas

Master Thesis Civil Engineering

Auke Sebastiaan Dijkstra

A Surrogate Modeling Framework for Compound Flood Risk and Optimization Analysis

Investigating the application of LSTM models for assessing compound flood mitigation designs at Clear Lake, Texas

by

Auke Sebastiaan Dijkstra

Student Name	Student Number
Auke	5166071

Image source: NASA Earth Observatory, ”Hurricane Harvey” (NASA Earth Observatory, 2017)

Supervisors:	Juan Anguilar Lopez (TU Delft) and Yoeri Jongerius (RPS)
Committee Members:	José A. Antolínez and Bas Jonkman (TU Delft)
Advisors:	Andrew Juan (A&M University), Saúl Buitrago Díaz (TU Delft) and Mieke van Arkel (RPS)
Project Duration:	Feb-Nov 2025
Faculty:	Faculty of Civil Engineering and Geo-sciences, TU Delft

TU Delft

TETRA TECH | rps A TETRA TECH COMPANY

Acknowledgements

This master's thesis marks the completion of my MSc degree in Civil Engineering at TU Delft. Looking back on the past years, I feel grateful and fulfilled. I have learned so many things, not only academically, but also personally. My studies have broadened my interests, given me opportunities to meet new people, to live abroad, and to explore new places and cultures. This thesis, in particular, brings together everything I am passionate about: hydraulic engineering, flood mitigation, advanced modeling techniques, and an international project. I am deeply grateful for the opportunity to carry out this work in collaboration between TU Delft and Tetra Tech/RPS.

There are several people I would like to thank specifically: Prof. Dr. Ir. S.N. Jonkman, for his guidance and for making my research stay in the United States possible. This experience has been truly inspiring and valuable. Dr. Ir. José A. A. Antolínez, for his concise, accurate, and constructive feedback throughout the process. Dr. Ir. Andrew Juan and his colleagues from IDRT, for their warm hospitality and for welcoming me into the IDRT team during my stay in Houston. Besides, Andrew has always taken the time to help me navigate all HEC-RAS questions I came up with, even online across the time difference.

Of course, Dr. Ir. Juan Pablo Aguilar Lopez, for always taking the time to schedule personal meetings, even with a busy agenda. Your effort and time were greatly appreciated. Ir. Saúl Buitrago Díaz, for all the time he invested in me. I always felt welcome to stop by your PhD office and truly appreciate the time you made to help me. I would also like to thank Ir. Yoeri Jongerius for helping me set up this topic and for his feedback, particularly in keeping an overview. Ir. Mieke van Arkel, thank you for your flexibility and for your help in arranging the opportunity to visit the US for this work.

A special thanks to everyone who took the time to talk with me during my stay in the US: Nefi Garza, Ryan Hostak, James Costello, and Mark Williams from Tetra Tech, Robert Thomas and Himangshu Das from USACE, and Dr. Ir. Jens Figlus from Texas A&M University. Your insights were very valuable.

Finally, I want to, of course, thank my friends and family, who have always supported me proudly and unconditionally, both directly and indirectly. Without them, this work would not have been possible.

Auke Sebastiaan Dijkstra
Delft, November 2025

Abstract

This research develops a surrogate modeling framework to efficiently analyze and optimize a proposed pump and gate system designed to mitigate compound flooding in the Clear Lake region, Texas. Traditional numerical hydraulic models are often computationally expensive for large number of simulations. As probabilistic assessments can require $O(10^3)$ to $O(10^5)$ runs, a cheaper alternative would be necessary for robust probabilistic assessment. This study addresses this limitation by developing a deep-learning surrogate to approximate the complex hydrodynamic behavior.

The methodology involved three main stages. First, a 1D HEC-RAS model of the Clear Lake system was adapted to serve as the physics-based "ground truth" generator. Second, this model was used to generate a training dataset of 2,400 simulations. This was achieved by systematically sampling key infrastructure design parameters (gate width W_g, number of pumps n_p, and activation levels h_{on}) alongside a wide range of synthetic compound flood forcings (inflow hydrographs and downstream storm surge boundaries).

Third, three distinct Long Short-Term Memory (LSTM) network architectures (Models A, B, and C) were developed to compare different data encoding strategies. Model A, a direct sequence-to-sequence (seq2seq) model, was provided with all dynamic inputs, including the known pump discharge time series (Q_{pump}). Model B tested the model's ability to infer dynamics by replacing the Q_{pump} time series with static design parameters (n_p, h_{on}). Model C used an autoregressive structure, feeding its own past water level predictions back as inputs to dynamically infer the pump response.

The results demonstrate that the fully-informed LSTM (Model A) can successfully learn and reproduce the governing hydrodynamic processes with very high accuracy. However, models that attempted to infer dynamic behavior from static design parameters (Models B and C) show reduced performance. These models particularly struggled to capture the sharp, transient effects of pump (de)activation, leading to overly smoothed predictions. This study concludes that while LSTMs are capable of learning the physical patterns of the system. The main challenge lies in feature encoding, specifically, enabling the model to capture complex, dynamic responses from static inputs. The framework demonstrates the potential of LSTMs, but emphasizes that how the data is represented is the key factor in developing a surrogate model suitable for design optimization.

Contents

Acknowledgements i

Abstract ii

1 Introduction 1

2 Clear Lake Case Study 5
2.1 Socio-Geographic Background . . . 5
2.2 Hydrodynamic System . . . 6
2.3 Compound Flooding Clear Lake . . . 6
2.4 Performed (Modeling) Studies . . . 8
2.5 Mitigation Strategies . . . 9
2.5.1 The Galveston Plan . . . 9
2.5.2 Preliminary Design Analysis . . . 9

3 Theoretical Background 13
3.1 From Physics-based to Data-Driven Models . . . 13
3.2 Introduction to Deep Learning . . . 13
3.2.1 Machine Learning vs. Deep Learning . . . 13
3.2.2 Model Training and Generalization . . . 14
3.3 Sequential Data-Driven Modeling . . . 15
3.4 Long Short-Term Memory (LSTM) Models . . . 15

4 Methodological Overview 17

5 Hydrodynamic Modeling 19
5.1 HEC-RAS Model . . . 19
5.2 Model Adjustment . . . 20
5.3 Model Validation and Justification . . . 22
5.3.1 Validation USACE Simulations . . . 22
5.3.2 Validation Storm Events . . . 23
5.4 Implementation of Infrastructure . . . 25
5.4.1 Floodwall and Gate . . . 26
5.4.2 Pump Station . . . 27
5.4.3 Primary Design Variables . . . 27
5.4.4 Stability Enhancement . . . 28
5.5 Results and Discussion . . . 28
5.5.1 Hydrodynamic Model Validation Results . . . 28
5.5.2 Model Justification . . . 30
5.5.3 Final System Overview . . . 30

6 Training Data Construction 31
6.1 Design Configurations . . . 32
6.1.1 Pump Configurations (Sampling n_p and h_{on}) . . . 32
6.1.2 Gate Width (Sampling W_g) . . . 32
6.1.3 Total Design Configurations . . . 33
6.2 Forcing Scenarios . . . 33
6.2.1 Inflow Hydrographs . . . 33
6.2.2 Downstream Boundary Condition . . . 34
6.2.3 Parameter Overview and Sampling . . . 35
6.2.4 Automation of HEC-RAS Model Runs . . . 36
6.3 Results and Discussion . . . 36

6.3.1 Forcing Parameter Distributions . . . 36
6.3.2 Generated Forcing Scenarios . . . 36
6.3.3 Final HEC-RAS Simulation Outputs . . . 37

7 LSTM Modeling **38**
7.1 Model-A: Baseline System Replication . . . 38
7.2 Model-B: Static Design Evaluator (HEC-RAS Replacement) . . . 40
7.3 Model C: Operational Forecaster (Autoregressive) . . . 41
7.4 Splitting and Normalization . . . 43
7.5 Overview LSTM Models . . . 43
7.5.1 Model Comparison and Evaluation . . . 44

8 Results **46**
8.1 LSTM Model Results . . . 46
8.1.1 Training and Validation Performance . . . 46
8.1.2 Computational Efficiency: HEC-RAS vs LSTM . . . 47
8.1.3 Model Evaluation . . . 47

9 Discussion and Recommendation **55**
9.1 General Interpretation of Results . . . 55
9.2 Limitations and Recommendations . . . 56
9.2.1 Hydrodynamic Model . . . 56
9.2.2 Training Data Generation . . . 57
9.2.3 Choice for LSTM Architecture . . . 59
9.2.4 LSTM-specific Considerations . . . 60
9.2.5 Further Exploration: Improvement of Model B . . . 61
9.2.6 Limitations and Recommendations Overview . . . 62

10 Conclusion **63**

A System Representation **69**
A.1 Additional Gate Parameters . . . 69
A.2 Models USED Mott MacDonald (2021) . . . 69

B Database construction **70**
B.1 Inflow Distribution Fitting . . . 71
B.2 Pump Configurations . . . 72

C Tidal Analysis Eagle Point **73**

D Feature Engineering and Reprocessing **75**
D.0.1 Lagged Correlation Analysis (LCA) . . . 75
D.0.2 Cross Correlation Analysis (Cov Matrix) . . . 76

E Sensitivity Analysis of Model C **78**
E.1 Sensitivity to Design Variables . . . 78
E.2 Sensitivity HMS Discharges . . . 81

F Results Model D **83**
F.1 Quantitative Metrics . . . 83
F.2 Visual Assessment . . . 83

List of Figures

1.1 Overview of proposed infrastructure and its location (USACE, 2024) . . . 3
1.2 General workflow . . . 4

2.1 Watershed Clear Creek and Armand Bayou with measurement stations (USGS, 2025; HCFCD, 2025d) . . . 5
2.2 Clear Lake (and Armand Bayou) Watershed (Bass and Bedient, 2018) . . . 7
2.3 Multiple lines of defense at Galveston Bay (USACE and GLO, 2021a) . . . 9
2.4 Location of proposed infrastructure at the Clear Lake inlet . . . 10
2.5 Technical drawing pump station (USACE, 2021) . . . 10
2.6 Side view of infrastructure, (MacDonald and Tetra Tech, 2021) . . . 11
2.7 Elevation map of area around flood wall . . . 12

3.1 Fully connected neural network . . . 14
3.2 Training process in deep learning algorithms. The learning algorithm (e.g., stochastic gradient descent) updates the network parameters through *backpropagation*, which occurs within the model's computational graph, even though it is illustrated here as a separate process. (Rivas-Blanco et al., 2021) . . . 14
3.3 LSTM Gates within LSTM cell (Figure obtained from DSAIE Teaching Team (2024)) . . 15
3.4 LSTM overview (Figure obtained from DSAIE Teaching Team (2024)) . . . 16

4.1 Flow Diagram Methodology . . . 17

5.1 Location of HEC-RAS model truncation marked by red dotted line. The model used in this study is the part to the right of this line. . . . 20
5.2 Clear Lake model domain and required hydrodynamic boundary conditions. Q_{1-4} reflect the inflow hydrograph boundary conditions. The gate operations for the existing gate are visualized on the map ($g_{exg}(t)$). The downstream boundary condition is h_{bay} . . . 21
5.3 Inflow boundary conditions (+30% rainfall) generated by HEC-HMS. . . . 23
5.4 Rainfall time series obtained from HCFCD (2025c) for (a) Hurricane Harvey and (b) Tropical Storm Beta. . . . 24
5.5 HEC-HMS–simulated vs. USGS–measured discharge for (a) Hurricane Harvey and (b) Tropical Storm Beta at the Friendswood station. . . . 24
5.6 HEC-HMS–computed inflow hydrographs for (a) Hurricane Harvey and (b) Tropical Storm Beta, used as upstream boundary conditions in HEC-RAS. . . . 25
5.7 Measured downstream water levels at Eagle Point (NOAA) (for location see Figure 2.1) for (a) Hurricane Harvey and (b) Tropical Storm Beta, applied as downstream boundary conditions. . . . 25
5.8 Implementation infrastructure in HEC-RAS model . . . 26
5.9 Inline flood structure at HEC-RAS river section 3823 . . . 26
5.10 Maximum water levels from the modified HEC-RAS model (HMS+30% rainfall). . . . 29
5.11 Modeled vs. Measured water levels at Clear Lake for Harvey (panel a) and Beta (panel b). "Modeled" refers to the "Q Model", while "Modeled (HS)" refers to the "h Model". . . . 29
5.12 Clear Lake system representation showing the location of the boundary conditions, and the design parameters of the infrastructure . . . 30

6.1 Flow Diagram Training Data Generation . . . 31
6.2 From system forcings to synthetic data . . . 33
6.3 Example of a Gamma distribution fitted to a normalized inflow hydrograph. Where q_p and t_p refer to the peak discharge and its offset respectively. . . . 33

6.4 Example of a lognormal fit (red line) to the peak discharge (q_{peak}, blue dots) of tributary Q_1. 34
6.5 Parameter distribution fits for synthetic storm surge generation. (a) Generalized Pareto Distribution (GPD) fitted to surge height exceedances over the 0.9 m threshold. (b) Normal distribution fitted to the observed surge asymmetry coefficient. 35
6.6 Synthetic generated Q1, Q3 and Downstream boundary from sampled parameters . . . 37
6.7 HEC-RAS model inputs and outputs 37

7.1 LSTM Architecture Schematization (Figure from Compare, Ye, and Meyer-Baese (2023) 39
7.2 General training loop schematic 40
7.3 Schematization teacher forcing (a) vs autoregressive (b) training 42

8.1 Computed model loss curves over epochs for Model A (blue), Model B (orange), and Model C (green). 46
8.2 Scatter plots of predicted versus observed water levels for Models A, B, and C. Model A shows the tightest clustering around the 1:1 line, indicating the highest predictive accuracy, while Model C exhibits larger scatter, particularly at higher water levels. Deviations around 1–1.5 m suggest challenges in capturing the dynamics of pump activation. . . . 48
8.3 Comparison of LSTM model predictions with observed water levels for selected test set samples. Each subplot shows the input inflows and downstream boundary conditions (top) and the predicted and measured water levels (bottom). In light grey, the pump response is shown, based on the pump variables n_p, h_{on}, and the target variable h_{lake}. . 50
8.4 Sequential inputs for test sample 65 52
8.5 Sensitivity of Model C predictions to varying gate widths. Inputs are based on test sample 65. 52
8.6 Sensitivity of Model C predictions to varying numbers of pumps. Inputs are based on test sample 65. 52
8.7 Sensitivity of Model C predictions to different pump activation patterns (see Table 8.4). Inputs are based on test sample 65. 53
8.8 Sensitivity of Model C predictions to different, unseen discharge data. 54

9.1 Input space for (a) pump capacity over time, and (b) n_p and & h_{on} 59

A.1 Hydrodynamic and hydrologic model of Clear Creek watershed (HCFCD, 2022) 69

B.1 Gamma distribution fit to available hydrographs 71

C.1 Measured tidal signal at Eagle Point over the analysis period. 73
C.2 Detected tidal events using a 0.5 m threshold. 73
C.3 Detected tidal events using a 0.9 m threshold. 74

D.1 Lagged-Correlation-Analysis 76
D.2 Cross-correlation matrix across model simulations 77

E.1 Effect of gate width on Model C performance. 78
E.2 Impact of the number of pumps on Model C results. 79
E.3 Sensitivity of Model C to the pump activation stagger. 80
E.4 Overview of the input configurations corresponding to the sensitivity analysis. 81
E.5 Model C, HMS Inflows 81
E.6 Model C, HMS Inflows 82
E.7 Model C, HMS Inflows 82
E.8 Model C, HMS Inflows 82

F.1 Scatter plots of predicted versus observed water levels for Models B and D. 83

List of Tables

2.1 Key hydrodynamic parameters for the Clear Lake system. . . . 6
2.2 Major storm events causing flooding in the Clear Lake area over the past 20 years. . . . 7
2.3 Overview of available modeling studies for compound flooding at Clear Lake. [1] *The report referenced in the final row was released during the construction of this thesis. A short reflection on its implications is added in the Limitations–Recommendations section.* . . . 8

5.1 Model characteristics . . . 21
5.2 Overview of model inputs and boundary conditions. . . . 21
5.3 Rainfall intensity (in millimeters) for various return periods and durations. . . . 23
5.4 Peak water levels in Clear Lake for various return periods based on USACE simulations and the modified HEC-RAS model results (in meters, NAVD88). . . . 28
5.5 Summary of variables used in the framework. . . . 30

6.1 Overview of inflow parameters, forcing types, and probability distributions. . . . 34
6.2 Overview of parameters, forcing types, and probability distributions. . . . 36

7.1 Overview of LSTM model setups and their objectives . . . 38
7.2 Comparison of input features across the three LSTM dataset variants. . . . 44
7.3 Summary of LSTM model architectures, trainable parameters, and overfitting considerations. . . . 44

8.1 Training duration and efficiency metrics for Models A, B, and C. The epoch of convergence is defined as the first epoch after which the validation loss remains within 2% of the final loss. . . . 47
8.2 Computational efficiency of HEC-RAS and LSTM surrogate models. . . . 47
8.3 Performance metrics for the LSTM models. . . . 47
8.4 Overview of pump activation patterns used in the analysis. Each pattern defines the activation head (m) for all $n_{\text{pumps}} = 8$. . . . 51

9.1 Summary of key limitations, their implications, and recommendations. . . . 62

A.1 Additional gate parameters . . . 69

B.1 Overview of pump configurations for varying numbers of pumps, with rounded h_{on} and h_{off} levels. . . . 72

F.1 Performance metrics for the LSTM models. . . . 83

1

Introduction

Research Context

Compound flooding, where intense rainfall, river discharge and storm surges combine to create severe flood hazards, forms a serious and increasing threat to many low-lying coastal watersheds in the world (Leonard et al., 2014; Antonia Sebastian, 2022). The eastern United States coastal region is prone to these compound flooding events. Here, hurricanes and tropical storms can generate significant wind-driven storm surges, and at the same time, release huge amounts of rain (Wahl et al., 2015), (Antonia Sebastian, 2022), and (Bevacqua et al., 2019).

In this region, the trend for increased compound flooding events has also been evident. Climate change causes storms to be more extreme, developing faster, generating higher wind speeds, and even larger rainfall amounts (Wahl et al., 2015; Nasr et al., 2023). Only in recent years (2012-2019), this area had to deal with 12 flood events.

One of the most illustrative examples of recent compound flooding in this region, is the flood event after Hurricane Harvey in 2017. Harvey brought record-breaking rainfall in this area, while at the same time generating significant storm surge. The event, that hit this densely populated area with a lot of critical infrastructure for the rest of the country, led to the death of at least 88 people. 13 million people in the Southeast region of the U.S. were affected by the hurricane and the total damage caused was estimated at 125 billion US dollars. Making it one of the costliest natural disasters in U.S. history (Amadeo, 2018; NOAA, 2018).

To mitigate compound floods, storm surge barriers combined with pumping stations are increasingly proposed to protect low-lying coastal areas from both marine and pluvial flooding (Kharoubi et al., 2024). These large systems mitigate flood disasters by closing during storm surges while allowing tidal movements and navigation under normal conditions (Mooyaart et al., 2014). Given their scale and the increasing frequency of floods due to climate change, an effective and robust design of such systems is crucial.

In modern coastal engineering, probabilistically-informed design is becoming increasingly important. This approach enables uncertainty quantification and robustness assessments (Federal Energy Regulatory Commission, 2020). A probabilistic flood framework typically consists of several statistical steps. First, boundary conditions or forcings, such as river discharges or storm surges, are analyzed to identify extreme values (for example, using the Peaks-Over-Threshold method). Next, a probability distribution is fitted to these data, allowing extrapolation to rare return periods (e.g. 100- or 500-year events). This process yields a probability of exceedance, which forms the basis of classical flood frequency analysis (FEMA, 2023; Federal Energy Regulatory Commission, 2020).

However, this approach faces limitations when dealing with non-stationary, compound events or interactions. Later developments such as joint probability methods, copulas, and multivariate statistical approaches address some of these challenges (Naseri et al., 2022), (Song et al., 2024), (Wang, 2023), (Ai et al., 2018), but translating combinations of forcings and infrastructural settings into probabilities

of exceedance remains difficult.

A more robust way to assess how coastal infrastructure affects flood probabilities is through Monte Carlo sampling of model simulations under a wide range of possible boundary conditions and system parameters. Such analyses often require $\mathcal{O}(10^3$–$10^5)$ simulations. The error of a simple Monte Carlo estimator is given by:

$$e = \frac{\sigma}{\sqrt{N}},$$

where σ is the standard deviation and N is the number of samples. In flood risk applications, exceedance probabilities in the order of 10^{-2} to 10^{-3} (or even smaller) are common. Achieving a stable estimate therefore requires approximately $N \approx 100/P$ simulations, for example, $10{,}000$ runs for $P = 0.01$ and $100{,}000$ runs for $P = 0.001$ (Nathan, Weinmann, and Hill, 2003; Ziya and Safaie, 2023).

It should be noted, however, that the standard error formula relates to the standard deviation of the estimator around the mean and does not fully capture uncertainty in the tails of the distribution. For rare events (small exceedance probabilities), the actual error in the estimated probability can be much larger than suggested by $e = \sigma/\sqrt{N}$, especially if the sampled events are not sufficiently representative of extreme conditions. Thus, care must be taken when interpreting Monte Carlo estimates for very low-probability, high-impact events.

Research Problem

Traditional physic-based models are powerful and intuitive but can be computationally expensive. When thousands of scenarios must be simulated, these models become impractical. This limits the ability to perform a Monte Carlo analysis, and hence, quantify realistic probabilistic risks and uncertainties when pump or gate systems are implemented.

Recent advances in deep learning offer a promising alternative. Data-driven surrogate models can approximate hydrodynamic behavior at lower computational costs (Haces-Garcia et al., 2025; Bass and Bedient, 2018; Dang et al., 2024). Several studies have demonstrated the ability of neural networks to reproduce water levels and flow dynamics from traditional models (A. Li et al., 2024; Ziya and Safaie, 2023; Bass and Bedient, 2018).

However, the use of alternative models for optimizing hydraulic infrastructure, particularly the primary design and operations of pumps and gates, remains largely unexplored. This creates a knowledge gap in how data-driven techniques can efficiently approximate the behavior of complex coastal systems, where these flood mitigation measures are implemented.

Case study: Clear Lake, Texas

To address this research gap, this study focuses on the Clear Lake area along the Texas coast, one of the regions prone to compound flooding from rainfall and storm surge (Bass and Bedient, 2018; Torres Dueñas, 2018).

As part of the Coastal Texas Protection and Restoration Feasibility Study (Coastal Texas Study) by the U.S. Army Corps of Engineers (USACE) and the Texas General Land Office (GLO), a storm surge barrier with an integrated pump station has been proposed to protect this area (USACE and GLO, 2021a). However, the capacity, configuration, and operation of the proposed pump system have not yet been optimized, making Clear Lake an ideal case for developing and testing a surrogate-based optimization framework.

Figure 1.1 shows the location of the Clear Lake proposed flood mitigation plan.

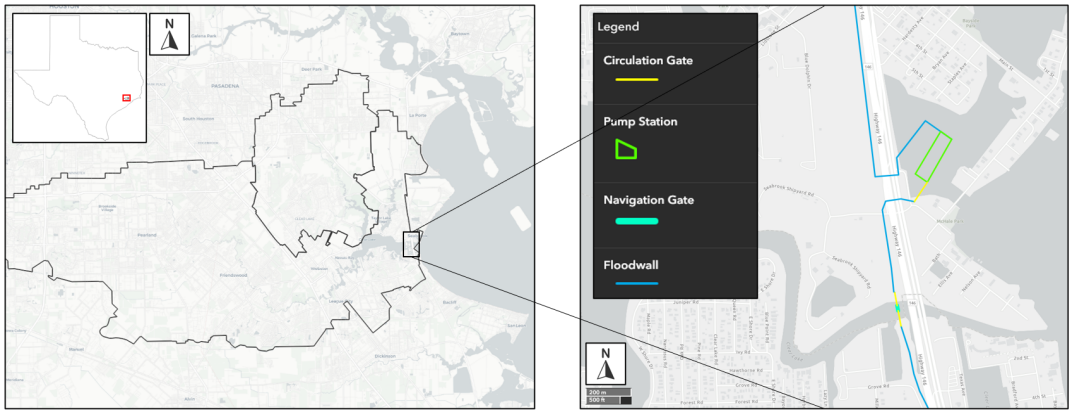


Figure 1.1: Overview of proposed infrastructure and its location (USACE, 2024)

The findings from this case are intended to support the ongoing Coastal Texas Study and to provide insights relevant to other low-lying coastal systems worldwide.

Research Objectives

This research aims to investigate how to develop a surrogate deep-learning model that enables computationally efficient simulations of the proposed pump and gate system at the Clear Lake inlet, supporting robust and risk-informed hydraulic design. Therefore, the central research question is:

> How can deep-learning models be applied to efficiently represent and simulate the hydraulic behavior of a pump and gate system under compound flooding conditions for the Clear Lake area (Texas)?

To answer this question, a threefold workflow is presented. Based on the time-dependent hydraulic behavior of the system, the Long Short-Term Memory (LSTM) network was selected as the specific deep-learning architecture for this study. LSTM is a recurrent neural network architecture known for effectively capturing complex, delayed relationships in time-series data (Goodfellow, Bengio, and Courville, 2016). A detailed justification for the selection and theoretical background of the LSTM model is provided in Chapter 3. This workflow defines the following sub-questions:

1. **System Representation:** How can the Clear Lake system be represented in a hydrodynamic model that accurately captures system responses to proposed infrastructure interventions?
2. **Training Data Generation:** How can this physics-based representation be used to generate a representative and computationally-efficient training dataset that captures the system's response to both a wide range of compound flood forcings and key infrastructure design parameters?
3. **Surrogate Encoding and Training:** How can the simulation dataset be used to develop an efficient LSTM model that is capable of capturing response in waters level to pump and gate variations and operations?

Research Scope

The primary focus of this thesis is on developing a valid surrogate modeling framework, capable of accurately reproducing the hydrodynamic behavior of the Clear Lake system. The study therefore concentrates on how to build such a data-driven surrogate, including the data generation and the model design.

To support this objective, the Clear Lake system is represented through a simplified hydrodynamic model. Because a modular workflow is presented, this system representation can in future work be replaced by a more detailed model, without changing the main workflow.

While the developed surrogate model is designed to enable fast and flexible analysis and optimization of pump and gate configurations, these applications are not performed within this thesis. The methodology forms a foundation that can be extended in future research toward full optimization and probabilistic assessment of coastal protection systems.

Research Methods and Outline

Chapter 2 and Chapter 3 describe the theoretical background. Where Chapter 2 provides background information about the Clear Lake system and Chapter 3 introduces key concepts of data-driven deep-learning methods that form the foundation of this study's methodology.

Chapter 4 presents an overview of the methodology, which follows a structured workflow. First, the Clear Lake system is represented using a traditional hydrodynamic model, described in Chapter 5. Chapter 6 explains how this model is used to generate training data, which serve as inputs for the LSTM surrogate models developed in Chapter 7. Figure 1.2 illustrates the overall workflow of the study.

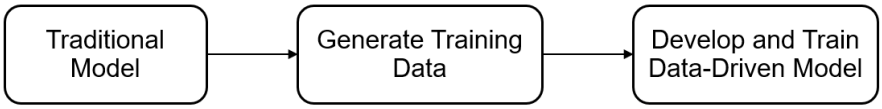


Figure 1.2: General workflow

Chapters 5 and 6 also present intermediate results, which inform the subsequent steps of the workflow. The results of the surrogate LSTM model are presented in Chapter 8. Chapter 9 evaluates these results and discusses the study's limitations, followed by corresponding recommendations. Finally, Chapter 10 concludes the study, linking the main findings, interpretations, and limitations to the overarching research objective.

2

Clear Lake Case Study

This chapter provides the theoretical background on the Clear Lake region, outlining its socio-geographic characteristics, hydrodynamic system, and history of compound flooding, followed by proposed mitigation strategies and an overview of previous modeling studies.

2.1. Socio-Geographic Background

Clear Lake, located in southeastern Texas near Houston (US) (Figure 2.1), is a brackish estuary connected to Galveston Bay. The total Clear Creek watershed (including Armand Watershed, whose streams also end in Clear Lake), spans approximately 660 km2 and lies within the counties Brazoria, Fort Bend, Galveston and Harris (Texas Stream Team, 2024). The total watershed area holds a population of around 320.000 people (estimates of 2020) (HCFCD, n.d.[b]; HCFCD, n.d.[a]). The land-cover/land use in the area is roughly 50% residential/commercial, 20% industrial (including NASA's Johnson Space Centre) and 30% green space or wetlands. The region has a humid subtropical climate, with a mean annual precipitation around 1330 mm and a mean evaporation of roughly 50% (Texas Stream Team, 2024).

Located along the Gulf Coast of the United States, the Clear Lake system is part of a densely populated and economically important region, prone to tropical storms, hurricanes and sea-level rise. Its location makes this watershed a representative study case of many coastal urban watersheds where limited elevation, rapid historical development and climate change interact to increase flood risk (Thatcher, Brock, and Pendleton, 2013; USACE and GLO, 2021a).

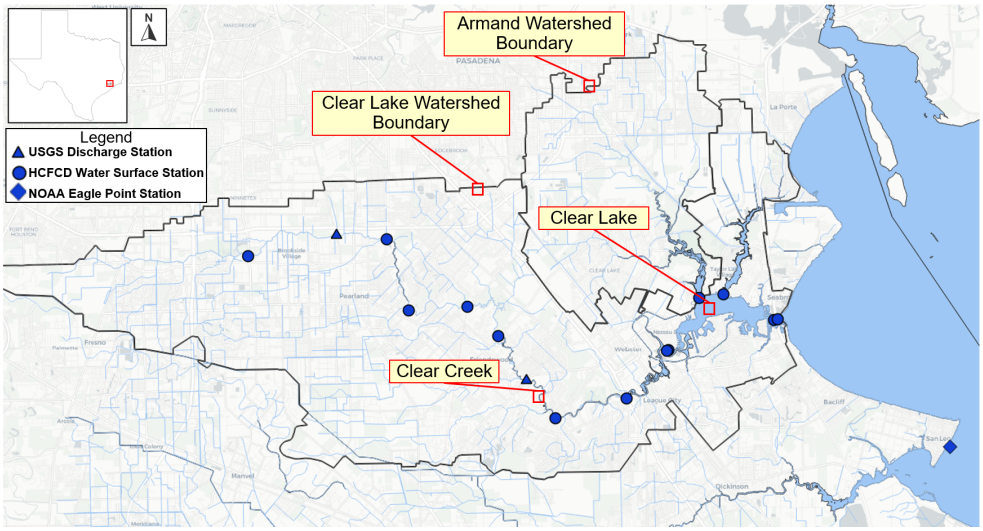


Figure 2.1: Watershed Clear Creek and Armand Bayou with measurement stations (USGS, 2025; HCFCD, 2025d)

2.2. Hydrodynamic System

Figure 2.1 visualizes how a network of tributaries and drainage channels converges on the Clear Lake, which connects to the Galveston bay through two channels; a gated channel in the north, and an ungated main navigation channel in the south.

Under normal flow conditions, the system's water level is strongly influenced by the downstream boundary, the tidal fluctuations of Galveston Bay water surface. The water level in Clear Lake on average fluctuates between -0.1 m and +0.3 m NAVD88. However, wind-driven setup can cause deviations from the tidal signal. This becomes evident looking at the difference between predicted water level and measured water level at HCFCD (2025d) and NOAA (2025). During rainfall events, most water is discharged in Clear Lake from the Clear Creek and Armand Bayou (and their tributaries) .

The watershed is a low-lying, flat, wetland area, with elevations ranging from approximately 23 m NAVD88 upstream to 1.5 m–2 m near Clear Lake. The area has a mild slope of roughly 0.03% (HCFCD, 2022). Studies have shown that the backwater curves of a storm surge, indicating tidal influence, can reach up to 43 km inland from the Clear Lake inlet at Galveston Bay (Bass and Bedient, 2018).

According to HCFCD (2025b), there is a probability on flooding from 1.43 m NAVD88. From 1.74 m NAVD88, flooding is likely.

Table 2.1 summarizes key hydrodynamic parameters relevant to the Clear Lake system:

Table 2.1: Key hydrodynamic parameters for the Clear Lake system.

Parameter	Value / Description
General Parameters	
Average slope	0.03%.
Water levels	Typically range between -0.1 and 0.3 m (NAVD88), but can be higher due to wind set-up.
Storm surge backpropagation	Approximately 43 km, governed by wind-driven setup, bay-water level rise, and tidal interactions.
Roughness (Manning's n)	Depth-varying: $n = 0.024{-}0.0678$ s/m$^{1/3}$ (Torres Dueñas, 2018).
Flooding possible	1.43 m (NAVD88).
Flooding likely	1.73 m (NAVD88).
Flood Frequency Levels	
10% (10-year)	1.83 m (NAVD88).
2% (50-year)	3.02 m (NAVD88).
1% (100-year)	3.54 m (NAVD88).
0.2% (500-year)	4.57 m (NAVD88).

Measurement Stations

Figure 2.1 shows the availability of HCFCD and USGS measurement stations. The HCFCD network consists primarily of surface water level elevation and rainfall gauges (HCFCD, 2025a). For discharge data, a key location is the USGS station 08077600 near Friendswood (eastern blue triangle), situated approximately 20 km upstream of Clear Lake (USGS, 2025). This measurement station only measures discharge during high flow events, so normal conditions are not captured. The closest NOAA measurement station is at Eagle Point (NOAA, 2025).

2.3. Compound Flooding Clear Lake

Historical events have demonstrated the regions high vulnerability to flooding. Major events topically occur during hurricanes or tropical storms that cause compound flooding. According to Leonard et al. (2014), a compound flood event occurs when multiple drivers, such as storm surge and high river discharge, interact to produce extreme impacts that would not occur if these drivers acted independently. This study focuses on compound flood events that occur when large runoff from heavy precipitation

combine with a storm surge in the Galveston Bay. Figure 2.2 schematizes a compound effect on the watershed.

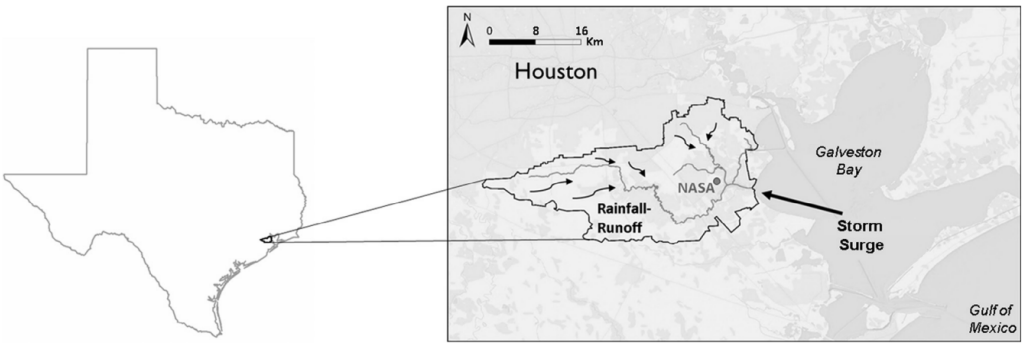


Fig. 2. Southeast Houston study area with respect to Texas.

Figure 2.2: Clear Lake (and Armand Bayou) Watershed (Bass and Bedient, 2018)

In this region compound floodings are often triggered by storm events because tropical storms and hurricanes can generate wind-driven storm surges while at the same time release significant rainfall (Wahl et al., 2015; Antonia Sebastian, 2022; Bevacqua et al., 2019). Recent major storm events, such as hurricanes Ike and Harvey, have caused devastating damages, including significant economic losses and loss of life. Hurricane Ike (2008), with its storm surge exceeding the estimated 100-year level, resulted in billions of dollars in damages and numerous fatalities. In 2017, Hurricane Harvey brought massive amounts of rainfall to the region, again leading to catastrophic flooding with severe economic and human impacts (Harris County Flood Control District, 2025).

In addition to these events, several other storms over the past two decades have produced combined storm surge and rainfall impacts that resulted in extensive flooding across the area. Recent studies have indicated a further increase in both the intensity and frequency of hurricanes due to climate change (Bevacqua et al., 2019; Antonia Sebastian, 2022). Table 2.2 provides an overview of the major flooding events that have occurred in this region.

Table 2.2: Major storm events causing flooding in the Clear Lake area over the past 20 years.

Storm	Year	Rainfall (approx.)	Description
Tropical Storm Erin	2007	200–250 mm	Localized pluvial flooding along Clear Creek and its tributaries Wikipedia contributors, 2025.
Hurricane Ike	2008	150–200 mm[a]	Severe surge-driven flooding and backwater effects within Clear Lake and adjacent bayous Houston Chronicle, 2008.
May 2015 Floods	2015	250–300 mm	Extreme short-duration rainfall causing extensive urban and fluvial flooding in the Clear Creek watershed Texas A&M University (2015).
Hurricane Harvey	2017	1000–1300 mm	Catastrophic rainfall event leading to prolonged inundation across the Clear Creek and Clear Lake catchment Harris County WCID 50, 2017; NOAA National Centers for Environmental Information, 2024.

[a] Most flooding resulted from storm surge rather than rainfall.

Sources: Texas A&M University (2015); Harris County WCID 50 (2017); NOAA National Centers for Environmental Information (2024); Houston Chronicle (2008); Wikipedia contributors (2025).

2.4. Performed (Modeling) Studies

Several studies have analyzed compound flooding in the Houston–Galveston region, focusing on the integrated effects of coastal, fluvial, and pluvial drivers to improve flood predictions (Maymandi, Hummel, and Zhang, 2022; Torres et al., 2015; Couasnon, A. Sebastian, and Morales Napoles, 2018; Bass and Bedient, 2018; Weathington, 2023; Torres & Associates, 2025). Coupled modeling frameworks, such as ADCIRC+SWAN and HEC-RAS/HEC-HMS, have demonstrated that integrated simulations provide more accurate flood representations compared to separate analyses as they better capture the physical interaction between flood drivers (Torres et al., 2015). Some studies have also implemented surrogate models to enable rapid probabilistic assessments.

Table 2.3 below provides an overview of the modeling studies performed to the Clear Creek watershed specific.

Study / Source	Model Type	Scope / Application	Notes
Weathington (2023)	2D HEC-RAS	Spatial inundation assessment	One-way coupled with ADCIRC storm surge data
Bass and Bedient (2018)	Surrogate model	Rapid probabilistic compound flooding	Combines ADCIRC+SWAN + 1D HEC-RAS + HEC-HMS
FEMA / USGS (EstBFE Viewer)	2D HEC-RAS	High-resolution floodplain mapping	Developed by AECOM under TWDB/FEMA CTP program
HCFCD (2025) M3 system	1D HEC-RAS + HEC-HMS	Stream-specific flooding	Official watershed models; includes Clear Lake and Armand Bayou
Mott MacDonald (2021)	1D HEC-RAS	Preliminary pump and gate design analysis	Uses HCFCD 1D model as input
Torres & Associates (2025)	2D HEC-RAS (Rain-on-Mesh)	Baseline risk & hotspot analysis	Bivariate rainfall–surge analysis.[1]

Table 2.3: Overview of available modeling studies for compound flooding at Clear Lake.

[1] *The report referenced in the final row was released during the construction of this thesis. A short reflection on its implications is added in the Limitations–Recommendations section.*

Region-specific studies at Clear Lake include a one-way coupling of ADCIRC storm surge outputs into a 2D HEC-RAS model to evaluate proposed flood protection measures (Weathington, 2023), as well as the combination of ADCIRC+SWAN outputs with 1D HEC-RAS and HEC-HMS for surrogate-based joint flooding estimation (Bass and Bedient, 2018). High-resolution 2D HEC-RAS models of Clear Lake and Armand Bayou are available via FEMA's "EstBFE Viewer," developed by AECOM under the TWDB/FEMA Cooperative Technical Partners (CTP) program (FEMA, n.d.).

Additionally, FEMA has produced 1D HEC-RAS models for multiple streams in the watershed, available through HCFCD's M3 webtool (HCFCD, 2022). The Mott MacDonald (2021) study utilized these official 1D models in preliminary design analyses for the proposed pump and gate system at the Clear Lake inlet.

Recently, the GGFCD (Galveston County Consolidated Drainage District) and IDRT (Institute for a Disaster Resilient Texas) have partnered to launch a new research study to the Clear Lake region. This study aims to provide new insights into flood management through data collection and numerical modeling (IDRT, 2025).

Torres Dueñas (2018) compared FEMA-computed floodplains with results from compound flood modeling using SFINCS, a semi-advanced 2D model designed to capture all relevant processes in coastal catchments efficiently (Leijnse, 2018).

However, apart from the preliminary design analysis of Mott MacDonald (2021), none of these studies explicitly incorporate the proposed Clear Lake inlet pump and gate system in their simulations because

of the early design stage. This gap is addressed in the present work, using the 1D HEC-RAS Clear Lake model from HCFCD (2007) as the basis for developing a surrogate model.

2.5. Mitigation Strategies

2.5.1. The Galveston Plan

Because of the high vulnerability to tropical storms and hurricanes of the Texas coastline, the United States Army Corps of Engineers (USACE) an the State of Texas General Land Office (GLO) have partnered to identify and recommend solutions, containing of multiple defense lines, to reduce the impacts from coastal hazards (USACE and GLO, 2021a). Their final feasibility report provides a recommended plan for the coast of Texas. This includes multiple measures along the whole Texas Gulf Coast to reduce flood risk. Examples are the Bolivar Road Gate System, that should keep out significant storm surge during hurricanes from the Gulf of Mexico. Figure 2.3 below provides an overview of the recommended plans at the Galveston Bay.

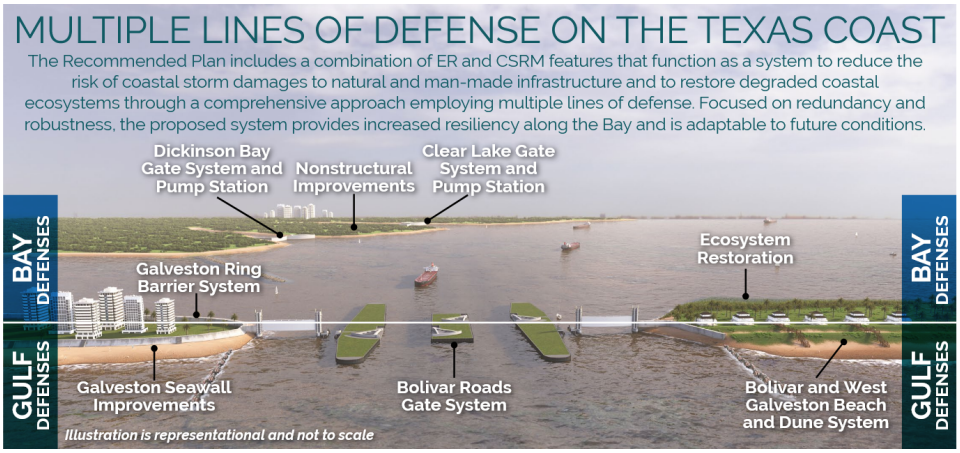


Figure 2.3: Multiple lines of defense at Galveston Bay (USACE and GLO, 2021a)

Given the size and complexity of the project (estimated to cost over $40 billion) numerous local and international research institutes and engineering firms are involved in its design and execution. The overall project timeline is extensive, meaning that full implementation will take many years (USACE and GLO, 2021a). Because of this long planning and construction horizon, smaller-scale protection measures, such as the proposed pump and gate systems, could potentially be implemented much earlier than the Bolivar Gate System. Therefore, this study considers a scenario in which the Bolivar Gates are not yet in place. Furthermore, USACE and GLO (2021b) note that even with the Bolivar Gate in place, significant storm surge can still develop within Galveston Bay itself due to wind-driven effects.

2.5.2. Preliminary Design Analysis

The Texas Coast Project includes the implementation of a floodgate system at the Clear Lake inlet that should eliminate wind driven surges fromm the Galveston Bay. In Appendix D, Annex 2 of the Coastal Texas Study, a feasibility analysis is performed by Mott MacDonald (2021), for this preliminary design.

The design, as described in the feasibility study, includes a 22.86-meter (75 ft) wide sector gate, with a sill elevation of -3.66 m (-12 ft) to match the authorized width and depth of the main channel (USACE and GLO, 2021a; Mott MacDonald, 2021). This sector gate is placed in the main channel within a large, concrete, flood wall, with an elevation of +5.18 meter (+ 17 ft) NAVD88. This structure should be able to withstand storm surge events with a 100-year return period from the Galveston Bay. To be able to discharge water from Clear Lake to the Galveston Bay, a pump station is proposed. In the final report of USACE and GLO (2021a) and MacDonald and Tetra Tech (2021), a design capacity of 566 m3/s (20,000 cfs) is described. While the feasibility study of Mott MacDonald (2021) (Appendix D, Annex 2) advised a pump capacity of more than twice that; namely 1,246 m3/s (44,000 cfs). In comparison, the pump station at IJmuiden, currently the largest in Europe, has a capacity of only 260 m3/s.

Figure 5.8 below visualizes the location of the proposed infrastructure. The marked existing gate is a gate structure present in the second outlet channel. This gate was not designed to provide protection from tides, hurricanes, or storm surges from the Galveston Bay, but rather to stimulate extra outflow for large rainfall events (HCFCD, n.d.[b]).

More detailed technical drawings from USACE (2021) of the proposed infrastructure are shown in Figures 2.5 and 2.6.

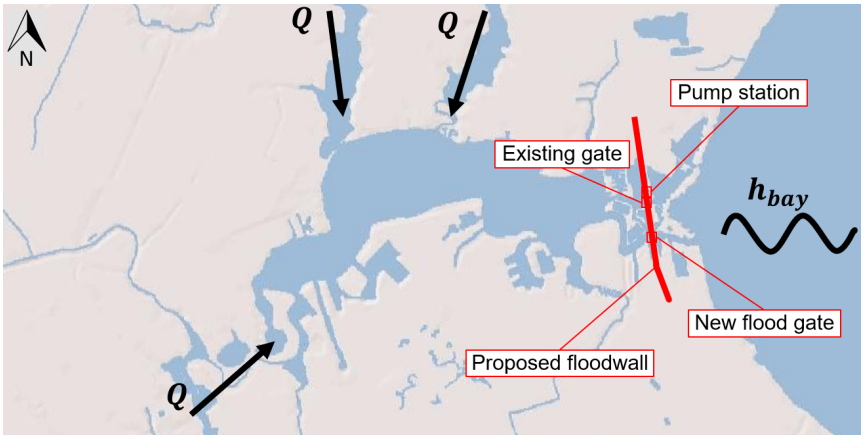


Figure 2.4: Location of proposed infrastructure at the Clear Lake inlet

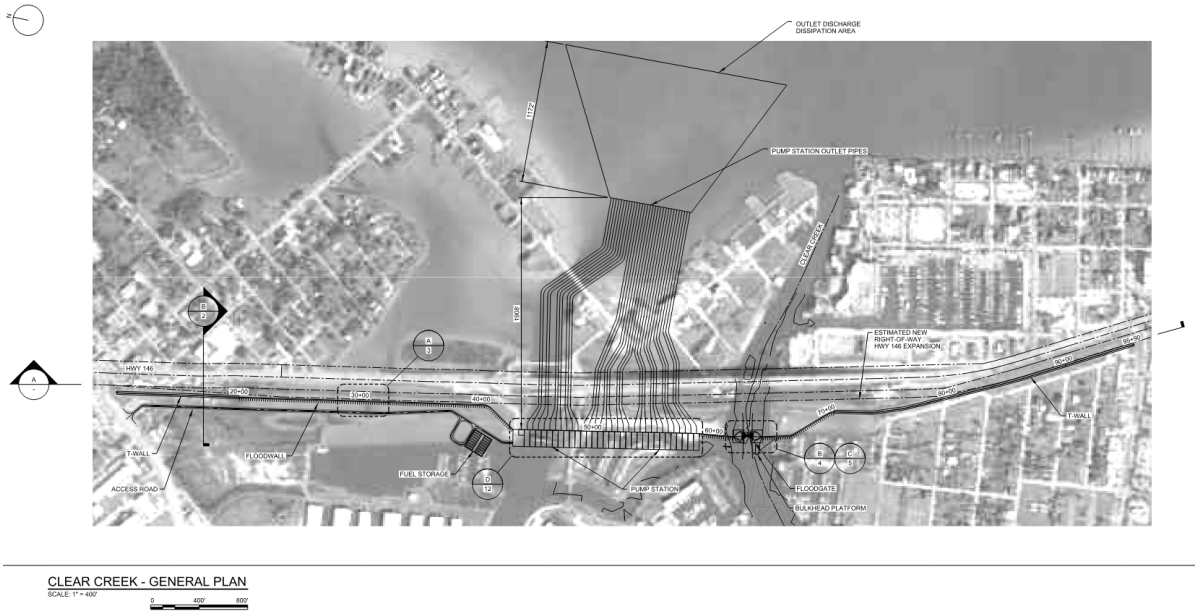


Figure 2.5: Technical drawing pump station (USACE, 2021)

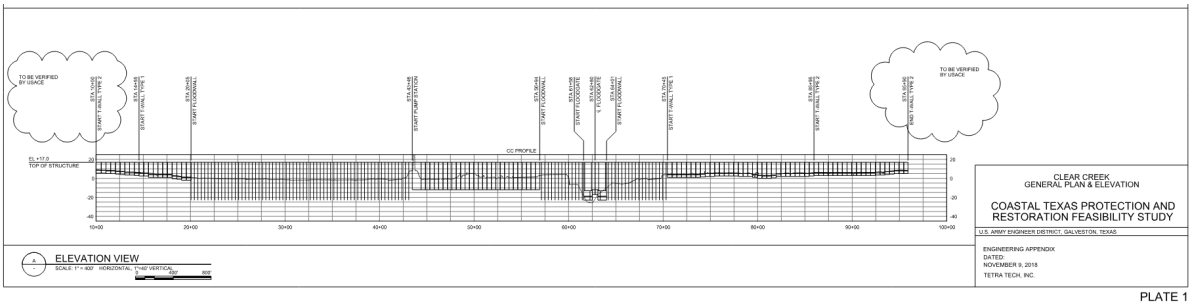


Figure 2.6: Side view of infrastructure, (MacDonald and Tetra Tech, 2021)

The hydrologic and hydrodynamic analyses summarized in Annex 2 (Mott MacDonald, 2021) were performed using the HEC-HMS hydrology model combined with the one-dimensional HEC-RAS hydrodynamic model previously developed for the Clear Lake watershed by HCFCD (2022). The objective of the this study was to assess the system's ability to protect the area against a 100-year storm surge combined with a 25-year rainfall event. The design principle was that the flood wall and gate system should prevent flooding for this case, while not worsening flood conditions during other event combinations compared to the current conditions.

Modeling Approach

Several rainfall-runoff simulations were conducted for the 10-, 25, 50, 100-, and 500-year rainfall events with a 24 hour duration. Because the analyses was performed before the release of updated NOAA Atlas 14 precipitation data USACE applied a 30 percent increase to rainfall intensities. Each simulation performed in the preliminary analysis covered a three day simulation, using a 30-min time step.

Before each simulated event, the water level in Clear Lake was assumed to be pre-pumped to 0.31 m MLLW (-0.38 ft NAVD88) and allowed to rise to the target water level of MHW (+0.25 ft NAVD88). If the gates where closed, the they remained closed throughout the whole simulation period, and wen open, they remained open. The downstream boundary condition at Galveston Bay was always set to +0.25 ft NAVD88. The total overviews of performed simulations by Mott MacDonald (2021) is presented in Appendix A.

Key Takeaways

The pump capacity determined in Annex 2 was based on maintaining the water level in Clear Lake at MHW (+0.25 m NAVD88. This represents a conservative assumption. Earlier analysis by USACE applied higher inland target water levels of +1.5 m to +1.73 m NAVD88 (Mott MacDonald, 2021), which are more consistent with the observed flood threshold. The Harris County Flood Control District (HCFCD) states that flooding at Clear Lake is likely to occur from 1.73 m NAVD88.

Secondly, the proposed 5.18 m (17-foot) flood wall must be integrated to a broader regional flood defense line to ensure storm surge water can not reach the hinterland. The surrounding terrain generally has an elevation much lower than 5 meter. As currently planned, the structure should tie into the nearby highway embankment, which indeed has a higher elevation than its surroundings, but still below 5 m NAVD88.

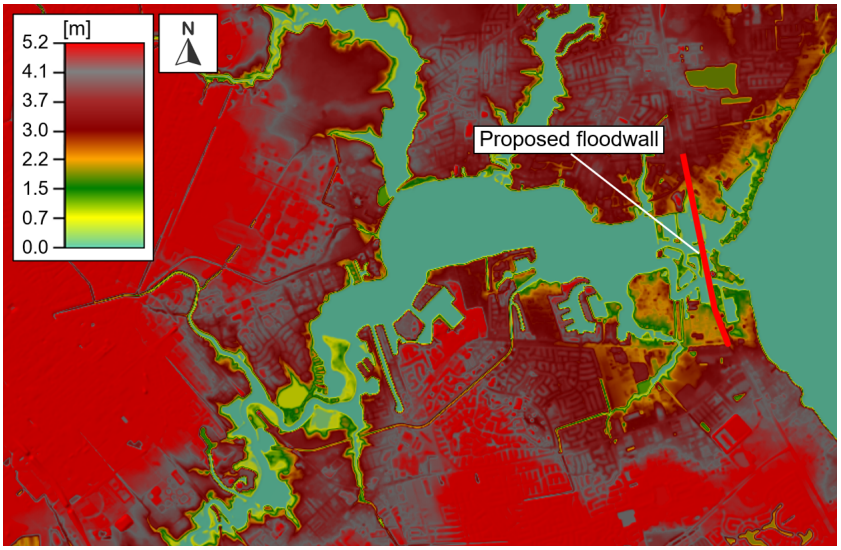


Figure 2.7: Elevation map of area around flood wall

Finally, the preliminary analysis assumed that the gates would remain closed for the entire three-day simulation period. This assumption is also conservative, since storm surges rarely persist for more than 24 hours. Chapter 5 of this study provides a more detailed analysis of storm surge duration and dynamic specific to this area. Maintaining the gate closed throughout the whole simulation can lead to unnecessary water accumulation upstream of the structure, requiring more pump capacity than needed.

It is important to keep in mind that the main objective of the USACE study was not to eliminate all forms of flooding or to maximize flood reduction, but rather to mitigate storm surge impacts while ensuring that fluvial flooding does not worsen as a result of the proposed infrastructure.

3

Theoretical Background

This chapter provides a brief theoretical background about the key concepts of data-driven modeling and deep learning, especially LSTM networks, needed to understand the methods used in this thesis.

3.1. From Physics-based to Data-Driven Models

Hydrodynamic systems are traditionally represented by physics-based models, such as HEC-RAS, which solve the shallow water equations based on physical laws and boundary conditions. These models provide interpretable results but are computationally expensive, particularly when used for large ensembles, long-term forecasting, or optimization under uncertainty.

An alternative is data-driven modeling, where system dynamics are approximated directly from data. Once trained, such models can reproduce input–output relationships at a fraction of the computational cost, making them attractive for real-time applications and scenario exploration (Shen, 2018; Sit et al., 2020). In this study, physics-based simulations are used to generate training data, enabling the construction of a surrogate model that combines the robustness of physical modeling with the efficiency of machine learning.

3.2. Introduction to Deep Learning

This section introduces the machine-learning and deep-learning models that will be used in this study. It briefly explains important definitions, keywords and the basic structures for those who are less familiar with this field of research. The section is not meant to reinvent the wheel, rather to give an overview of the most important factors to properly understand this work.

3.2.1. Machine Learning vs. Deep Learning

Deep learning is a subset of machine learning that uses neural networks composed of multiple layers to learn hierarchical feature representation directly from raw data. This depth allows the model to capture complex spatial and temporal dependencies and highly non-linear relationships between inputs and outputs (Goodfellow, Bengio, and Courville, 2016).

At the core of deep learning are artificial neural networks, inspired by the biological neural networks of the human brain. These networks consist of interconnected nodes called neurons organized into layers: an input layer, one or more hidden layers, and an output layer. Most layers of these networks have trainable weights that are updated during training to make the model better fit the target output. Each neuron receives inputs, applies a weighted sum followed by a non-linear activation function, and passes the result to neurons in the next layer. The depth and complexity of these layers enable deep learning models to capture intricate patterns and relationships within data, making them highly effective for complex tasks such as image recognition, natural language processing, and speech recognition (Goodfellow, Bengio, and Courville, 2016; Kratzert, Klotz, Brenner, et al., 2018).

Figure 3.1 below shows an example of a fully connected neural network (all the nodes of the input layer

are connected to all the nodes of the following hidden layers).

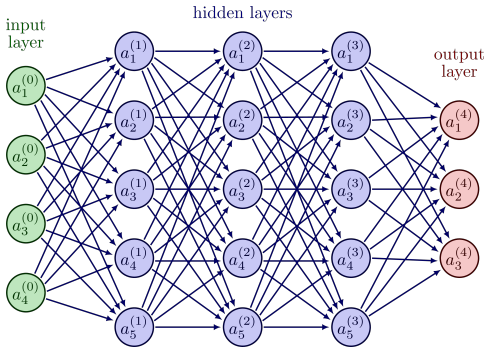


Figure 3.1: Fully connected neural network

Alternative neural networks introduce inductive biases. These built in assumptions help constrain the learning process to overcome the curse of dimensionality (where the number of parameters in the model grows rapidly as input dimensions increase). Examples of these alternative neural networks are CNNs (Convolutional Neural Networks) that apply spatial inductive biases to efficiently learn local spatial representations across input data, and RNNs (Recurrent Neural Networks), that apply sequential inductive biases to learn how elements in sequence relate to each other. The choice of architecture and arrangement of layers depends on the specific problem and data type (Shen, 2018).

3.2.2. Model Training and Generalization

Training a deep learning model involves adjusting its parameters (weights and biases) to minimize the difference between its predictions $\hat{Y}$ and the actual target values Y. This difference is quantified by a "loss function" (e.g., mean squared or absolute error). The gradients of the weights of the network with respect to the loss function can be calculated using the chain rule for differentiation. This is called a backpropagation step. Gradient descent is the process of adjusting these weights to minimize the loss function. The *learning rate* is a crucial hyperparameter that controls the step size during gradient descent updates. Setting it too high can cause the model to overshoot minima, while too low a rate leads to slow convergence (Goodfellow, Bengio, and Courville, 2016). The training process is visualized below in figure 3.2

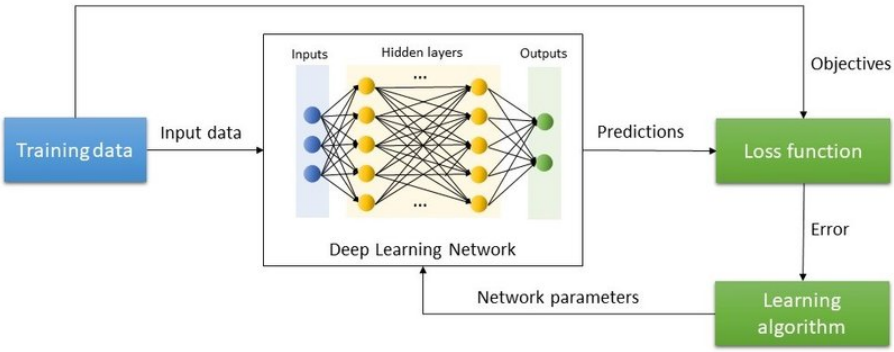


Figure 3.2: Training process in deep learning algorithms. The learning algorithm (e.g., stochastic gradient descent) updates the network parameters through *backpropagation*, which occurs within the model's computational graph, even though it is illustrated here as a separate process. (Rivas-Blanco et al., 2021)

After model training, the model can be used at the interface (or in the *inference stage*). Model parameters remain fixed, and it generated predictions for unseen input data. The distinction between training and inference is crucial for understanding how neural networks generalize from training data to application.

3.3. Sequential Data-Driven Modeling

This thesis addresses a time-dependent prediction problem: estimating water levels in Clear Lake from sequential inflow hydrographs and pump/gate operations. In such tasks, the system state at the current timestep depends not only on the immediately preceding state but also on delayed cause-effect relationships that can span long durations.

Given the nature of this problem, several data-driven architectures are suited. For example, Multilayer Perceptrons (MLPs) can approximate nonlinear relationships but lack temporal awareness. Recurrent Neural Networks (RNNs) were developed specifically to address this by introducing sequential inductive biases. However, standard RNNs struggle with long-term dependencies due to the ”vanishing gradient” problem. During the backpropagation process, the signal required to update weights often diminishes or explodes as it propagates back through long sequences (Goodfellow, Bengio, and Courville, 2016).

To address the limitations of traditional recurrent neural networks more advanced architectures have been developed. Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) networks overcome the limitations of standard RNNs by employing gating mechanisms. These mechanisms facilitate a ”constant error flow” during training by truncating the gradient where it ”does no harm” (Hochreiter and Schmidhuber, 1997). More recent Transformer architectures capture long-range dependencies using attention mechanisms instead of recurrence, and are increasingly used for time series forecasting (Vaswani et al., 2017; Shen, 2018).

For this research, the Long Short-Term Memory (LSTM) network was selected. While Transformers are a promising alternative, LSTMs represent a robust and well-established ”state-of-the-art” in hydrological and hydrodynamical modeling. Their ability to model sequential dependencies has been proven in numerous studies, being effective at capturing delayed, non-linear relations, common in these systems (Kratzert, Klotz, Brenner, et al., 2018; Kratzert, Klotz, Shalev, et al., 2019; X. Li, Zhou, Hou, et al., 2024; An and Ouyang, 2025).

3.4. Long Short-Term Memory (LSTM) Models

LSTMs, first introduced by Hochreiter and Schmidhuber (1997), overcome the limitations of simple RNNs, particularly the vanishing and exploding gradient problem, by introducing gating mechanisms that explicitly regulate the flow of information. The core innovation of the LSTM is the cell state, which acts as a ”memory” path, allowing information to be preserved over long sequences with minimal decay. Unlike the hidden state in standard RNNs, which is typically modified by matrix multiplications at every step, the cell state interacts with the network primarily through linear addition operations. This additive structure is crucial as it ensures that gradients can flow backward through time without vanishing, enabling the model to preserve relevant information (Hochreiter and Schmidhuber, 1997; Goodfellow, Bengio, and Courville, 2016).

Each LSTM unit contains three distinct *gates*, visualized in Figure 3.3. These gates consist of sigmoid neural network layers that output values between 0 and 1, effectively determining how much information is allowed to pass through the cell.

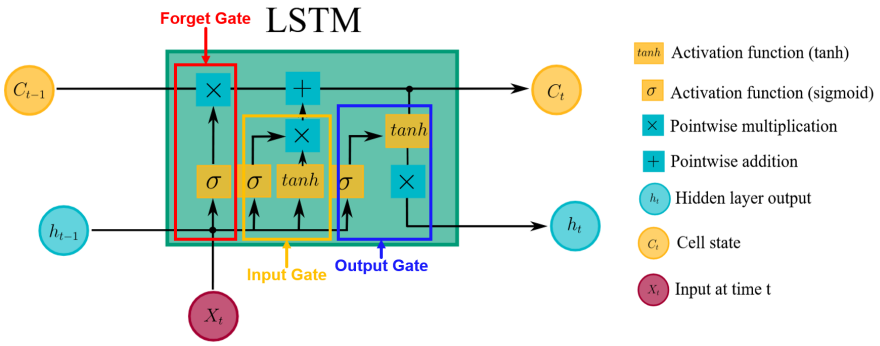


Figure 3.3: LSTM Gates within LSTM cell (Figure obtained from DSAIE Teaching Team (2024))

Figure 3.4 shows three consecutive LSTM cells at corresponding time steps.

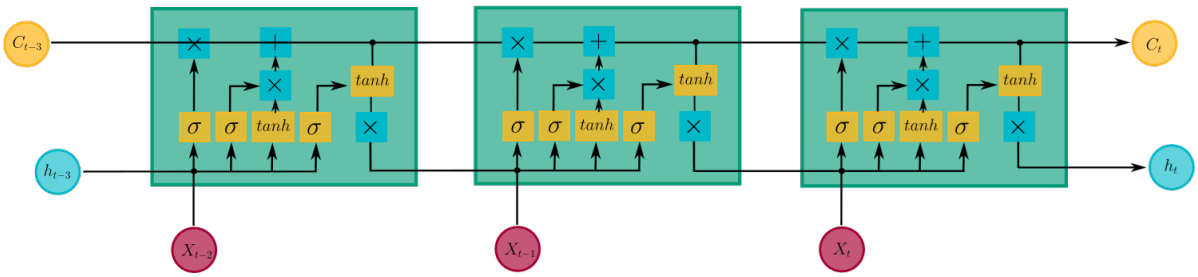


Figure 3.4: LSTM overview (Figure obtained from DSAIE Teaching Team (2024))

The flow through an LSTM network is as follows (Goodfellow, Bengio, and Courville, 2016)

1. First, the **forget gate** f_t decides which information from the previous cell state $\mathbf{C}_{t-1}$ is no longer relevant. It evaluates both the previous hidden state h_{t-1} and the current input x_t, producing a value between 0 (completely forget) and 1 (completely retain).
2. Second, the **input gate** i_t determines what new information from the current time step should be stored in the cell state. This step includes two operations: the input gate i_t decides which components of the cell state will be updated, while a $\tanh$ layer produces a vector of new candidate values $\tilde{\mathbf{C}}_t$ that may be added to the state.
3. Third, the previous cell state is updated to form the new cell state $\mathbf{C}_t$, highlighting the additive nature of LSTMs:
$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t,$$
where $\odot$ denotes element-wise multiplication. Here, $\mathbf{f}_t$ scales the previous state (forgetting what is no longer needed), while $\mathbf{i}_t$ determines how much of the candidate information $\tilde{\mathbf{C}}_t$ is incorporated.
4. Finally, the **output gate** o_t generates the hidden state h_t. The updated cell state is first passed through a $\tanh$ activation to constrain values to $[-1, 1]$, and then multiplied by the output gate activation. This determines which parts of the cell state are relevant for producing the output at the current time step.

This architecture allows the LSTM to combine external inputs x_t and previous hidden states h_{t-1} while carrying forward its internal memory $\mathbf{s}_{t-1}$ with minimal degradation. By separating the long-term memory (cell state) from the short-term output (hidden state), the model can preserve slowly varying hydrodynamic base states over many time steps while still responding to immediate fluctuations such as rainfall or pump activation.

4

Methodological Overview

This chapter provides a general overview of the adopted methodology. It explains the threefold method structure, what the reader can expect in every subsequent section, and how the information flows between them.

The methodology is structured around three general steps:

1. **Represent** the Clear Lake hydrodynamic system in a physics-based model.
2. Simulate this model N times **to generate data** to train an alternative, data-driven model.
3. **Develop and train** the surrogate model using the constructed database.

The overall methodology is visualized in Figure 4.1

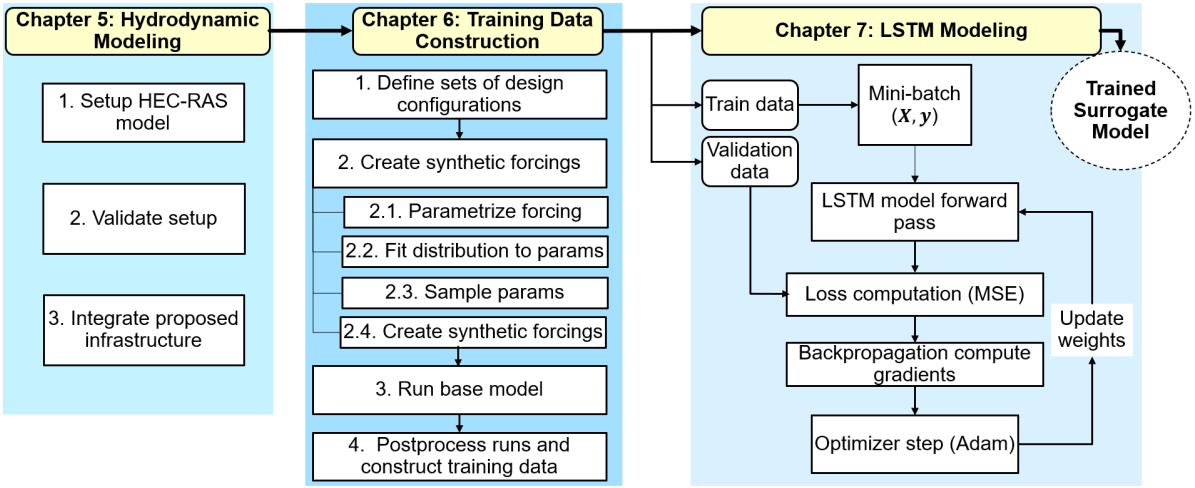


Figure 4.1: Flow Diagram Methodology

Chapter 5 outlines the first part of the methodology: the representation of the Clear Lake system in a hydrodynamic model. It defines the required model inputs, the implementation of the proposed infrastructure and the primary design variables of this infrastructure. The model will be validated and compared using performed studies and historical events. The result is a physical based model, representing the hydrodynamic system of Clear Lake, directly addresses the first research question.

Chapter 6 presents how such a physical based model can be used to generate efficient and computationally feasible training data.

Chapter 7 shows how a dataset of one-dimensional HEC-RAS simulations can be used to develop and train an efficient LSTM model that is capable of accurately capturing response in water level to pump and gate design and operation variations. To achieve this, three LSTM Model setups were developed and analyzed.

Chapters 5 and 7 each produce intermediate results that are essential for the subsequent methodological steps. For this reason, the intermediate findings of those chapters are presented directly within those chapters, as they form the necessary input for the next step of the workflow. Chapter 7 builds upon these results and provides the trained surrogate models, forming the basis of the third research questions as well as the overall objective of this thesis. Therefore, the final results are presented in a dedicated chapter *"Results"* following these three methodological sections.

The trained surrogate LSTM models can be used in the inference stage to efficiently simulate a large number of potential flow scenarios across different pump/gate designs. This computationally cheap alternative to HEC-RAS enables Monte-Carlo simulations to explore the performance of various pump and gate designs and operation settings, supporting a probabilistic assessment of system behavior. Although optimization itself is beyond the scope of this study, the methodology establishes a framework for future optimization studies.

5

Hydrodynamic Modeling

This chapter outlines the first part of the methodology: how the Clear Lake system was represented in a hydrodynamic model. The objective is to develop a physics-based HEC-RAS model to serve as 'ground truth' for the subsequent methodological steps. The process was defined as follows: First, the modification of an official HEC-RAS model is explained. This modified model is then validated against performed studies and historical events. After validation, the proposed pump and gate infrastructure is implemented. The results of the model validations are presented in the last section of this chapter, directly leading to a system overview that answers the first research question.

5.1. HEC-RAS Model

The first step was to set up a hydrodynamic model of the Clear Creek river reach. The model used in this study was based on the official 1D Clear Lake HEC-RAS model, obtained from HCFCD. This model was originally developed in 2007 and last updated in 2022 (HCFCD, 2022). Although it is a one-dimensional model, it is suitable for this research for several reasons.

First, the model is publicly available and already validated. Secondly, since the study aims to develop a surrogate, rather than to build a full hydrodynamic model from scratch, the use of an existing model significantly reduces development time and complexity. Moreover, the area of interest for this study, which is the Clear Lake area, has a relatively constant water level, making a 1D model sufficient.

The model discussed was also used in the preliminary design analysis of the proposed infrastructure by Mott MacDonald (2021), performed for USACE. This is a valuable reference, as it allows for assessment of how the floodwall, the gate, and the pump station were implemented in their professional design study. Moreover, it enables a comparison of simulated results, which can be used to support the validation of the modified HEC-RAS model.

HEC-RAS

HEC-RAS 1D solves the full dynamic-wave form of the Saint-Venant equations, namely the continuity and momentum balances (Brunner, 2021). In conservative form, these read:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = q_l(x,t) \tag{5.1}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}\left(\beta \frac{Q^2}{A}\right) + gA\frac{\partial h}{\partial x} + g\frac{Q|Q|}{CAR^{4/3}} = 0 \tag{5.2}$$

Here, A is flow area, Q discharge, h water-surface elevation, q_l lateral inflow, g gravity, $R = A/P$ hydraulic radius, β the momentum-correction coefficient, and C the conveyance related to Manning's n. HEC-RAS employs an implicit finite-difference scheme, specifically a four-point box discretization in

space and time, and solves the resulting nonlinear algebraic system simultaneously at all cross sections using Newton–Raphson iteration at each time step (Brunner, 2021).

Limitations and Assumptions
The model is one-dimensional and steady state by default, meaning modifications are required to simulate unsteady flow. To make this model unsteady, boundary conditions produced with the accompanying HEC-HMS model were used. These HMS computed boundary conditions can be coupled using the DSS file. Fully coupling these models is outside of the scope of this research, as it is too time-consuming.

The limitations of 1D modeling must be kept in mind: although the river reach is largely 1D, significant tributaries, side channels, and other 2D aspects in the Clear Lake area will not be properly represented.

Additionally, the original model was developed in an outdated version of HEC-RAS; version 3.0. This limits the ability to use more recent features such as RAS-Mapper or advanced structure implementations. Therefore, the model was updated and made compatible with the most up-to-date version, HEC-RAS 6.6.

It was assumed that the geometric and bathymetry data in the original model is still valid. This includes cross section layouts, roughness coefficients, inactive flow paths, and other default parameters.

5.2. Model Adjustment

For this research, the full official 1D model was modified. The main adjustment is the 'cut-off' of the model domain to the specific area of interest of this study. The model was truncated at cross section 57171.49, ensuring two cross sections are present upstream of the bridge section at B55625 (the bridge of FM-270 highway), which is required for unsteady flow in HEC-RAS 6.6.

The domain reduction was done for multiple reasons. First, the preliminary assessment of Mott MacDonald (2021) showed that the flood infrastructure primarily affects the downstream part of the system, specifically the Clear Lake area. To capture the hydrodynamic behavior of the system, it was advisory to include the point where the channel widens and transitions into the Lake. To ensure the inclusion of three important HEC-RAS control structures (I-45, Old Galveston Road, and FM-270), which significantly influence flow patterns into Clear Lake, the model domain was extended slightly upstream of the channel–lake transition point (see Figure 5.1).

By reducing the model domain, time and effort were saved transferring the model from HEC-RAS 3.0 into HEC-RAS 6.6, particularly in updating cross sections. It also reduced the number of tributaries and sub-basins that needed to be manually inserted from the corresponding HMS model, simplifying boundary condition setup.

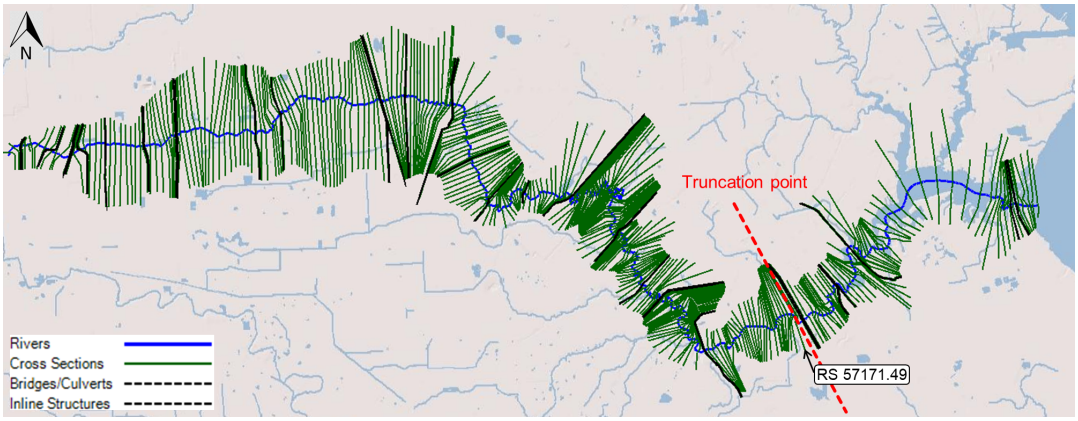


Figure 5.1: Location of HEC-RAS model truncation marked by red dotted line. The model used in this study is the part to the right of this line.

The model truncation resulted in a "modified HEC-RAS model" (hereafter also referred to as such).

Table 5.1 presents the final extent of the modified model, and Figure 5.2 illustrates its domain and boundary conditions.

Table 5.1: Model characteristics

Parameter	Value	Unit
Number of cross sections	76	–
Of which inline structures	6	–
Total river length	17.43	km

Boundary Conditions

For the official developed 1D model of HCFCD (2022), the upstream boundary conditions are hydrographs, obtained from the HMS Hydrologic model. All the hydrographs produced for the reaches and sub-basins are implemented at the corresponding river sections in the hydrodynamic model.

The truncation of the 1D model reduced the number of required upstream boundary conditions. The primary model boundary conditions are presented in Table 5.2. Figure 5.2 shows a visualization of the final model domain, with the required hydrodynamic boundary conditions.

Table 5.2: Overview of model inputs and boundary conditions.

Label	Source / Location	Description
Q_1	RS 57171.49 / HEC-HMS stream 0566_J	Upstream inflow hydrograph representing the main Clear Creek inflow.
Q_2	RS 36556 / HEC-HMS stream 0022_J	Inflow hydrograph from Cow Bayou tributary.
Q_3	RS 18407.60 / HEC-HMS stream 0022_J	Hydrograph representing inflow from Armand Bayou tributary.
Q_4	RS 14000.00 / HEC-HMS stream 0052_J	Hydrograph representing inflow from Taylor Bayou tributary.
h_{bay}	Downstream boundary (Galveston Bay) RS 0	Stage–hydrograph defining downstream water level. For storm surge simulations, a synthetic time series is applied; for historical events, tidal data are obtained from NOAA (2025)
$g_{exg}(t)$	RS 3858	Existing lift gate operations in secondary (Nortern) outlet channel.

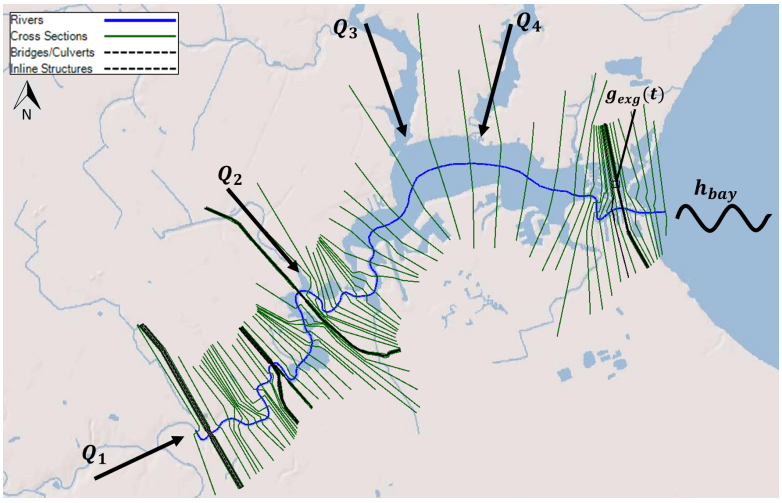


Figure 5.2: Clear Lake model domain and required hydrodynamic boundary conditions. Q_{1-4} reflect the inflow hydrograph boundary conditions. The gate operations for the existing gate are visualized on the map ($g_{exg}(t)$). The downstream boundary condition is h_{bay}

.

It was chosen to neglect the sub-basin inflow to the river reach. The cumulative flow of the sub-basins only contributes a few percent of that of the tributaries. Furthermore, including them would complicate the input structure without a significant benefit for the surrogate model training.

Limitations and Assumptions
This truncation introduced the following limitations: Although the impact of the proposed infrastructure is limited upstream of the cutoff point, it is not negligible. Additionally, the new upstream boundary lies within the tidally influenced domain, meaning that defining a clear discharge–water level relationship is challenging. This limitation was accepted within the context of this study, as the primary goal is to develop a surrogate model that approximates the behavior of the modified HEC-RAS model. Some inaccuracies in this underlying physics-based model are tolerable as long as the surrogate can replicate them reliably. No recalibration of model parameters (e.g., Manning's n) was performed after the model domain was truncated.

5.3. Model Validation and Justification

The modified one-dimensional HEC-RAS model was validated by comparing its maximum water level outputs with the study of Mott MacDonald (2021) and evaluating its performance for two historical (compound) storm events: Hurricane Harvey (2017) and Tropical Storm Beta (2020). The goal is to assess whether the simplified, modified version of the 1D model from HCFCD (2022) can reproduce the hydrodynamic behavior consistently with earlier performed studies and measurements

The required inflow boundary conditions (Q_{1-4}) were obtained using the accompanying HEC-HMS hydrologic model, which uses precipitation as input. Since this thesis already involves the implementation of the hydrodynamic model and the development of a data-driven LSTM-based surrogate model, the scope of hydrologic validation (HEC-HMS) was limited. Hence, the model was validated for the limited number of two design events only.

5.3.1. Validation USACE Simulations

The same conditions were replicated in the modified HEC-RAS model as Mott MacDonald (2021) performed in their existing condition simulations. These were:

- Simulation duration was set to three days (72 hours).
- The downstream boundary condition was fixed at 0.24 m NAVD88 throughout the entire simulation.
- The six existing gates were kept open throughout the simulation.
- The inflow hydrographs were derived from the accompanying hydrologic model that used the 10-, 50-, 100-, and 500Y +30% as input.

At the time of that study, the updated Atlas 14 HCFCD rainfall return period data were not yet available. To compensate for the outdated 2007 values, the USACE applied a 30% increase to the rainfall intensities (Mott MacDonald, 2021). For consistency, the same adjustment is adopted in this study. The 25-year rainfall event is not included, as the HEC-HMS model obtained from the FEMA M3 modeling website (HCFCD, 2022) does not contain a corresponding basin setup for this event. The rainfall return periods used for this study are presented in Table 5.3.

Table 5.3: Rainfall intensity (in millimeters) for various return periods and durations.

Duration	10Y	50Y	100Y	500Y
5 min	33.0	35.6	40.6	45.7
10 min	50.8	50.8	58.4	68.6
15 min	76.2	63.5	68.6	83.8
30 min	96.5	94.0	104.1	132.1
1 hr	109.2	124.5	142.2	182.9
2 hr	121.9	165.1	188.0	254.0
3 hr	139.7	195.6	223.5	309.9
6 hr	175.3	254.0	299.7	431.8
12 hr	210.8	315.0	365.8	525.8
1 day	256.5	383.5	447.0	637.5

Using these rainfall inputs, HEC-HMS simulations were performed for 24-hour storm events. Default model parameters and loss–transform–routing methods from HCFCD (2022) were retained. For the 500-year return period event, several storage–discharge and inflow–diversion tables required extrapolation, which was carried out using linear or polynomial fits.

To improve numerical stability at the onset of the simulation, an additional 24-hour warm-up period with very low base-flow was applied to all runs. The resulting discharge hydrographs from HEC-HMS were then used as upstream boundary conditions for the HEC-RAS hydrodynamic model. These inflow boundary hydrographs are shown in Figure 5.3.

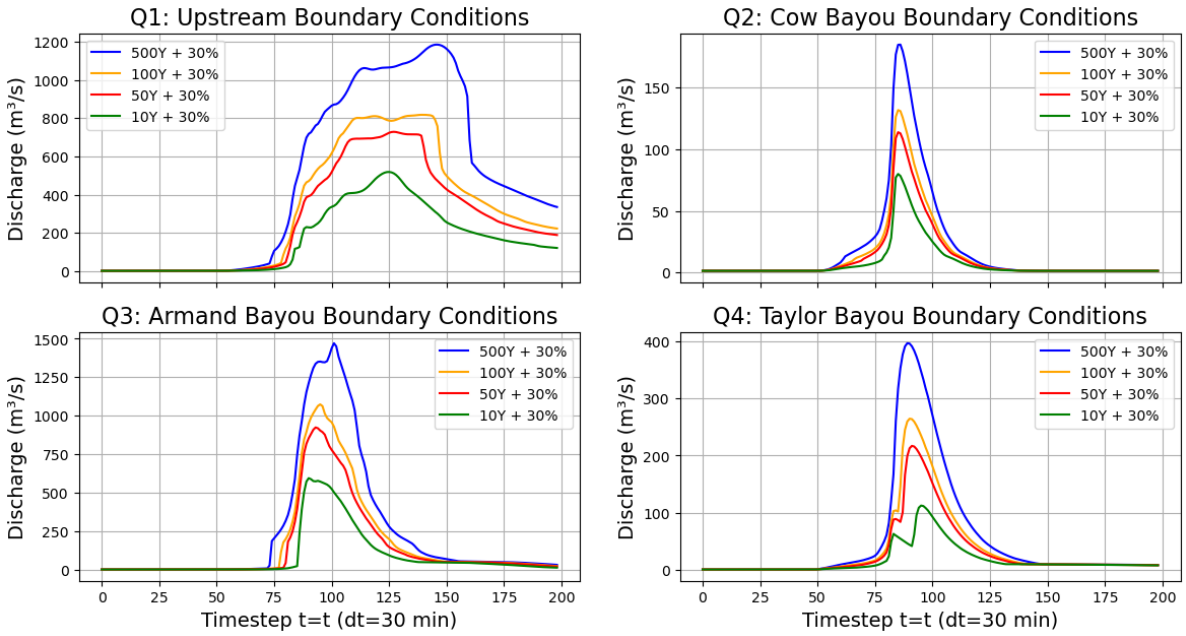


Figure 5.3: Inflow boundary conditions (+30% rainfall) generated by HEC-HMS.

5.3.2. Validation Storm Events

To further evaluate the model setup under realistic conditions, two historical storm events were selected: Hurricane Harvey (August 2017) and Tropical Storm Beta (September 2020). Rainfall data for both events were obtained from the Harris County Flood Warning System (station 140) (HCFCD, 2025c), chosen for its central position within the watershed. The rainfall was assumed to be spatially uniform across the catchment. This assumption introduced uncertainty due to the spatial variability of the storms, but was adapted to reduce computational and data demands. This simplification is acknowledged to potentially influence the results.

The rainfall inputs applied in the HEC-HMS simulations are shown in Figure 5.4. Each simulation was run for the duration corresponding to the plotted time series. For the base configuration, the 100-year precipitation model from HCFCD (2022) was used. For Hurricane Harvey, several storage–discharge and inflow–diversion tables required extrapolation to accommodate the extreme event magnitudes.

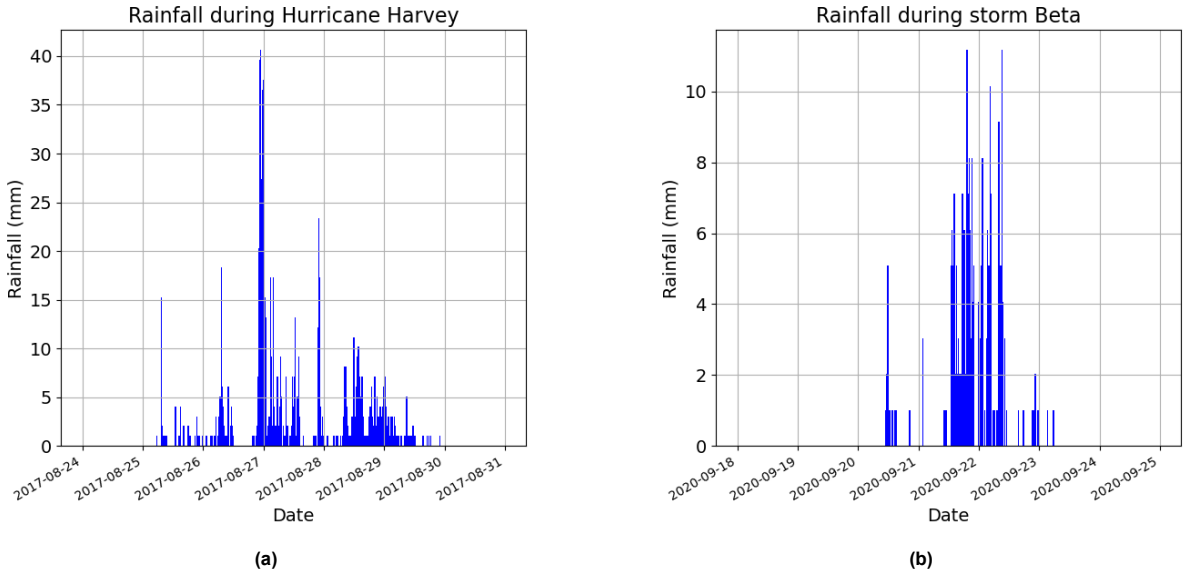


Figure 5.4: Rainfall time series obtained from HCFCD (2025c) for (a) Hurricane Harvey and (b) Tropical Storm Beta.

The runoff hydrographs computed by HEC-HMS were subsequently compared to measured discharges from the USGS Friendswood station (*Clear Lake, Texas*; ID 08077600) to evaluate input accuracy (Figure 5.5). This step served to identify potential errors in the hydrologic input rather than to assess model performance.

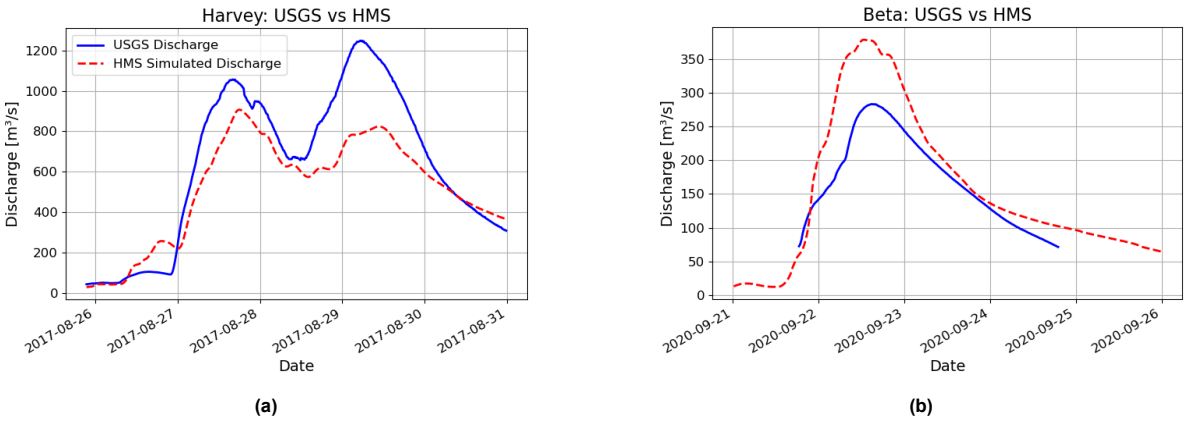


Figure 5.5: HEC-HMS–simulated vs. USGS–measured discharge for (a) Hurricane Harvey and (b) Tropical Storm Beta at the Friendswood station.

For both cases, although the overall temporal behavior of the hydrographs is well captured, there is a significant mismatch in discharge magnitudes. This indicates that the errors occurring in the modified HEC-RAS model results are not solely due to model limitations or applied modifications, but also originate from inaccuracies in the input hydrographs generated by HEC-HMS. Consequently, distinguishing the source of these errors becomes challenging, making the interpretation of the modified HEC-RAS model outcomes more complex.

The four HEC-HMS hydrographs (main channel and three tributaries) were then used as inflow boundary conditions in the modified HEC-RAS model (Figure 5.6). Downstream boundary conditions were

defined using water levels from the NOAA Eagle Point gauge (NOAA, 2025), implemented as stage hydrographs (Figure 5.7). Water levels were converted from MLLW to NAVD88 by subtracting 0.07 m, based on the tidal datum conversion provided by National Geodetic Survey (2025).

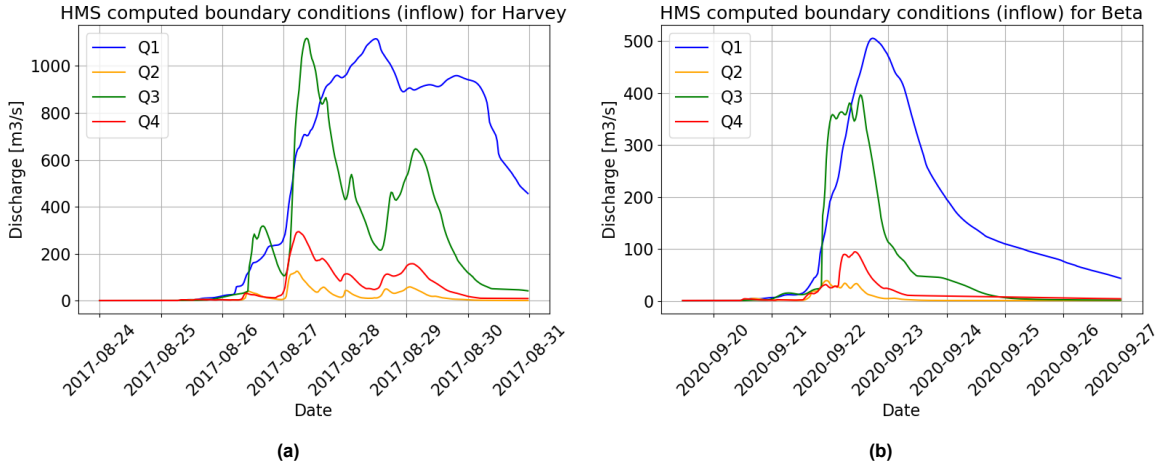


Figure 5.6: HEC-HMS–computed inflow hydrographs for (a) Hurricane Harvey and (b) Tropical Storm Beta, used as upstream boundary conditions in HEC-RAS.

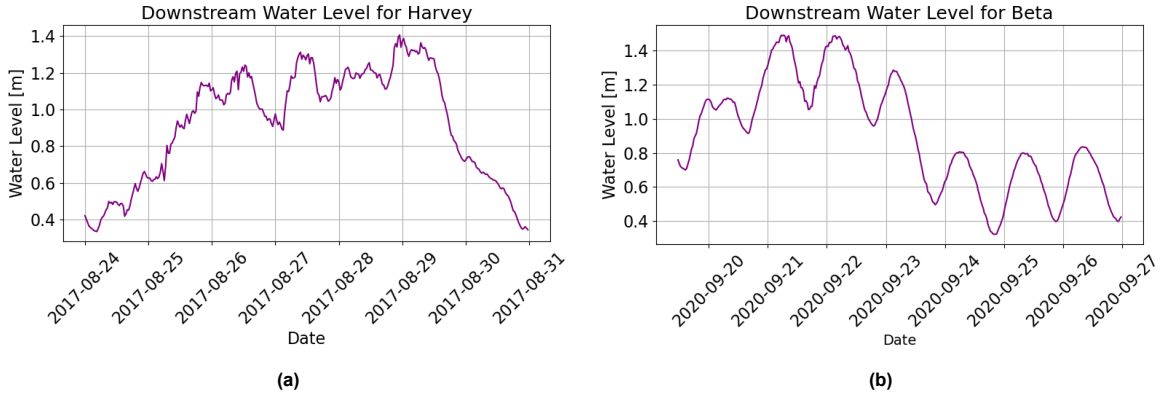


Figure 5.7: Measured downstream water levels at Eagle Point (NOAA) (for location see Figure 2.1) for (a) Hurricane Harvey and (b) Tropical Storm Beta, applied as downstream boundary conditions.

To assess the model's sensitivity to the upstream boundary definition in a tidally influenced domain, two configurations were tested:

1. **Discharge-driven configuration ("Q Model")**: Uses the HEC-HMS–computed discharge hydrographs (Q_1–Q_4) at all inflow locations.
2. **Stage-driven configuration ("h Model")**: Replaces the main upstream inflow (Q_1) with a stage hydrograph derived from the HCFCD upstream gauge (HCFCD, 2025d), while the tributaries (Q_2–Q_4) remain discharge-driven.

5.4. Implementation of Infrastructure

The implementation of the proposed pump and gate infrastructure was based on the preliminary analysis by Mott MacDonald (2021). This section explains how each component was represented within the modified HEC-RAS model. Figure 5.8 shows the location of the infrastructure within the model domain.

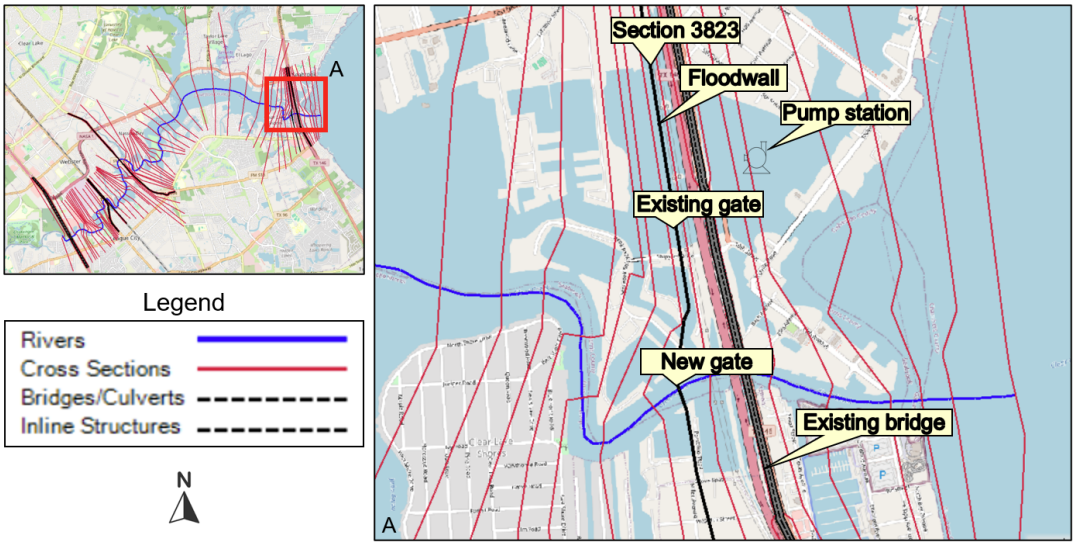


Figure 5.8: Implementation infrastructure in HEC-RAS model

5.4.1. Floodwall and Gate

The floodwall was implemented at river section 3823, consistent with the implementations of Mott MacDonald (2021). The crest elevation of the floodwall was set to +5.18 m NAVD88. The wall was represented as a concrete inline structure in the HEC-RAS model. Figure 5.9 illustrates the configuration in HEC-RAS, showing the existing sill gate on the left (North) side, and the new flood gate, visible as a gap within the floodwall. The center of the gate is positioned approximately at the midpoint of the main channel.

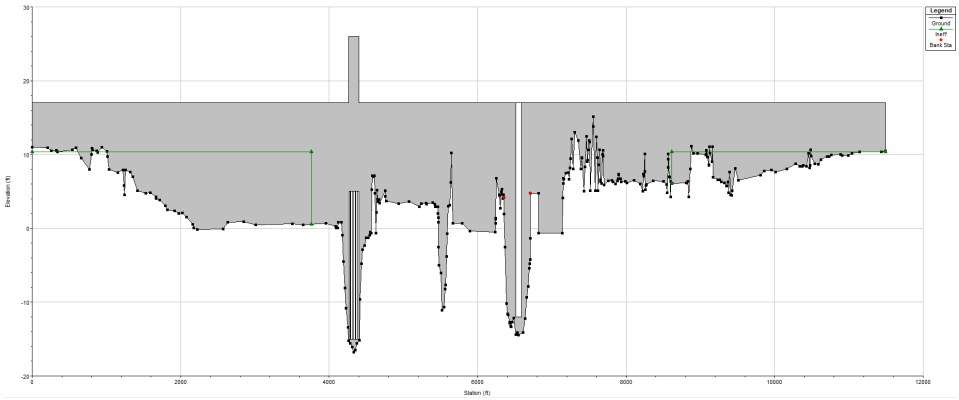


Figure 5.9: Inline flood structure at HEC-RAS river section 3823

An initial gate width of 22.86 m was implemented; however, this dimension is treated as a design variable and is varied in subsequent analyses. The sill elevation was fixed at –3.66 m NAVD88. The gate type was defined as radial that opens vertically, and default HEC-RAS parameters were adopted for the required gate flow characteristics. The specific parameters and their assigned values are summarized in Table A.1 in Appendix A. The

Operations

Gate operation was controlled through time series data that specifies the opening height (vertical gate) at each time step. The total height of the gate extends from -3.66 m to +5.18 m NAVD88, corresponding to a total height of 8.84 meters (29 feet). In this study, a simplified binary operation scheme was applied: the gate is either fully open or fully closed, with transitions occurring within a single time step. Although instant closure or opening is unrealistic in practice, this assumption reduces the number of

input parameters required in the subsequent simulations and simplifies the modeling process. This signal also allows the operational input to be interpreted as either an opening depth or width, which is advantageous for simulations described in Chapter 6, where the gate width is given as input variable.

The currently existing gate will lose its function when the new infrastructure is implemented. Therefore, in all simulations including the proposed design, the existing gate remains closed (HCFCD, n.d.[b]).

5.4.2. Pump Station

The pump station is located at the same cross section as the floodwall and gate. Because the pumped discharge is directed towards the Galveston Bay, a large downstream storage area, it is assumed that the pumped discharge does not significantly influence the downstream water level. As a result, the pumping process can be represented as a “pump-out” of the system. This approach simplifies the required model adjustments, as HEC-RAS does not allow direct pumping from one cross section (upstream of the floodwall) to another (downstream) (Brunner, 2021). Under this assumption, the pumped discharge is effectively removed from the system.

Water volumes were extracted from river section 4937, as HEC-RAS requires at least two cross sections between the implementation point and the extraction location. This section is the nearest feasible position to the intended pump station site. Appendix A presents the technical drawings of the proposed pump station and gate, where their spatial configuration is shown in detail.

Two simplifying assumptions were made for the pump modeling:

- Startup and shutdown times were not included
- Pump efficiency curves were not considered. Each pump is assumed to operate at its maximum rated capacity, regardless of head differences.

While this approach is simplified, it provides a practical first-order estimation of the required pump capacity and operational plan.

Operations

Pump operations were governed by activation and deactivation water levels. The activation level corresponds to the water level at the extraction point (RS 4937). The deactivation level was kept at a fixed offset of 0.15 m below the activation water level for each pump. Based on the provided pump capacity, the number of pumps, their activation and de-activation water level, HEC-RAS computes a ’pump plan’, which is the pumped capacity over time.

5.4.3. Primary Design Variables

The previous sections described the implementation of an infrastructure configuration based on the preliminary Mott MacDonald (2021) design. However, the primary objective of this thesis is not to evaluate this single design, but to develop a methodological framework that can efficiently explore a wide range of design alternatives and support future optimization.

To achieve this, the validated HEC-RAS model will be used to generate a large dataset for surrogate modeling. This dataset is created by systematically varying the key parameters that define the flood mitigation system. These **Primary Design Variables** are:

- **Gate Width (W_g)**: The horizontal width (in meters) of the new flood gate. This directly controls the hydraulic conveyance during open-gate conditions.
- **Total Pump Capacity ($Q_{p,total}$)**: The total combined discharge (in m^3/s) of the pump station when all pumps are active.
- **Number of Pumps (n_p)**: The total number of individual, equally-sized pumps. This variable, combined with $Q_{p,total}$, defines the capacity of each individual pump ($Q_p = Q_{p,total}/n_p$) and dictates the staging of operations.
- **Pump Activation Water Levels ($h_{on,i}$)**: The sequence of water levels (in m NAVD88) at RS 4937 that trigger each of the n_p pumps to activate.

These variables were selected as they represent the most significant engineering decisions influencing both the construction cost and the hydraulic performance of the system.

5.4.4. Stability Enhancement

Time span

Each simulation runs over a time span of three days with a 30-minute time step. While the storms themselves typically last 24 hours, the watershed's time of concentration extends to about three days, after which flows begin to recede and drainage returns to normal (Mott MacDonald, 2021). See also Figure 5.3. Despite the relatively short model domain and the upstream locations where the inflow hydrographs are applied, the flood-wave travel time (approximately 2–5 hours) is still fully captured within this three-day window.

A fixed simulation period and time step was used across all runs. This simplification ensures consistency in the input-output structure for the surrogate model and avoids added complexity due to varying simulation configurations.

Warm up time

To improve model stability, a 24-hour warm-up period was added prior to the onset of flood hydrographs. During this period, a low base flow of 1.42 m3/s (50 cfs) was applied. This low discharge flow helps the model settle and avoids numerical problems and instabilities, caused by a sudden rise in discharge at the start of the simulation.

Because the hydrographs from HMS start at zero flow and rise steeply, the warm-up period ensures a smooth transition and helps define appropriate initial conditions. The base flow quantity is used to define initial conditions of the model because the steady-state simulation gives unrealistic results for low flows.

5.5. Results and Discussion

This section presents the results of the validation of the modified model, after which a total system overview is presented.

5.5.1. Hydrodynamic Model Validation Results

Validation USACE Simulations

The modified HEC-RAS model was validated by comparing its maximum water level outputs with the study of Mott MacDonald (2021).

Table 5.4 compares the maximum water levels in Clear Lake predicted by the modified HEC-RAS model and those reported by Mott MacDonald (2021). The longitudinal distribution of maximum water levels is shown in Figure 5.10.

Table 5.4: Peak water levels in Clear Lake for various return periods based on USACE simulations and the modified HEC-RAS model results (in meters, NAVD88).

Return Period	USACE [m NAVD88]	Modified HEC-RAS [m NAVD88]
10Y	1.51	1.43
50Y	1.91	1.84
100Y	2.08	1.98
500Y	2.60	2.43

The results show a consistent underestimation of water levels compared to the reference simulations. This behavior was expected because sub-basin runoffs in the modified setup were neglected, and the modified HEC-RAS model only considers inflow from the main tributaries ($Q_1 - Q_4$). Consequently, the total discharge entering the system is slightly underestimated.

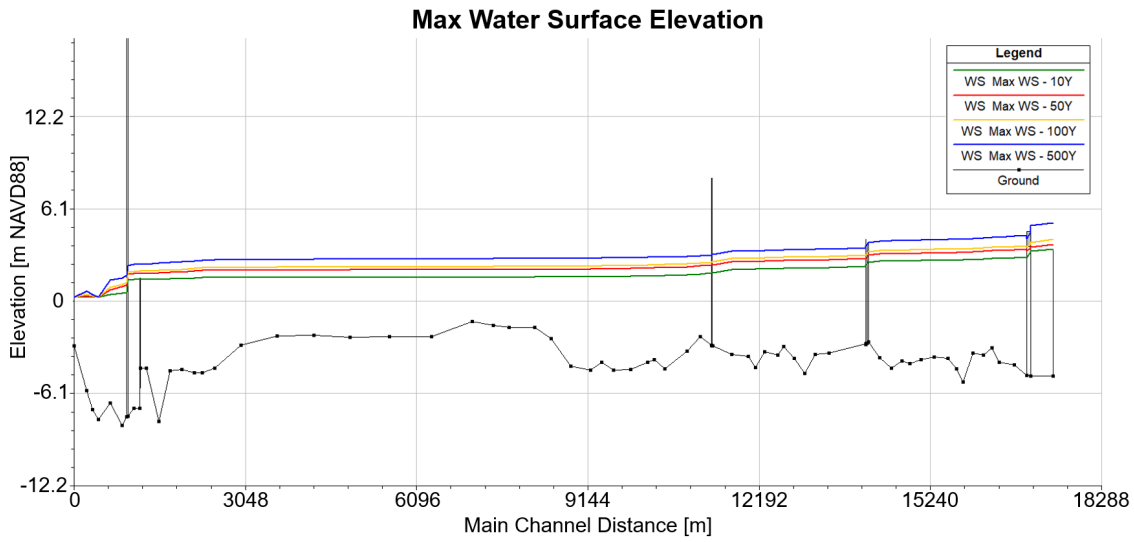


Figure 5.10: Maximum water levels from the modified HEC-RAS model (HMS+30% rainfall).

The difference between the modeled peak water levels presented in Table 5.4 and the flood frequency water levels obtained from HCFCD (2025b) (Table 2.1) directly illustrates the influence of storm surge. The modeled peak water levels represent only the probabilistic contribution of precipitation, as the downstream boundary was fixed at mean sea level. In contrast, the HCFCD values are derived from observed water level records and therefore include both precipitation and storm surge effects, resulting in higher water level values.

Validation Historical Storm Events

The simulated water levels for both configurations are compared against measured data in Figure 5.11.

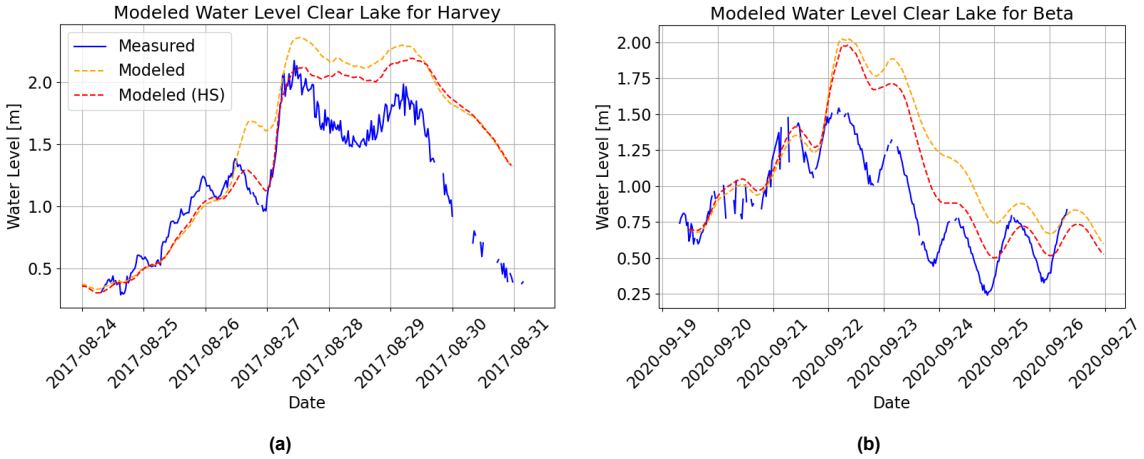


Figure 5.11: Modeled vs. Measured water levels at Clear Lake for Harvey (panel a) and Beta (panel b). ”Modeled” refers to the ”Q Model”, while ”Modeled (HS)” refers to the ”h Model”.

The results show that both models, the ”Q Model” and the ”h Model” show slightly different results, but both have inaccuracies. This implies that errors originate from two sources: the simplified boundary conditions provided by the HMS model (based on uniform rainfall data) and the implementation of the upstream boundary and tributaries in tidally influenced locations. The slightly better performance of the ”h Model,” which uses a stage hydrograph as input, indicates that both types of errors are present.

5.5.2. Model Justification

Despite the mismatches shown in Figure 5.11, the modified HEC-RAS model will be used for the subsequent analysis. Because further calibration is outside the project scope and the model provides a reasonable representation (especially the main temporal pattern), the modified HEC-RAS model is considered appropriate for this study. Furthermore, in the subsequent simulations where the flood gate is implemented, the estuary is closed off, which will limit the influence of these complex boundary dynamics. The modular threefold method described in Chapter 4 allows to replace this model by one with better performance in future work.

5.5.3. Final System Overview

Figure 5.12 shows the resulting total system overview. The variables are summarized and explained in Table 5.5.

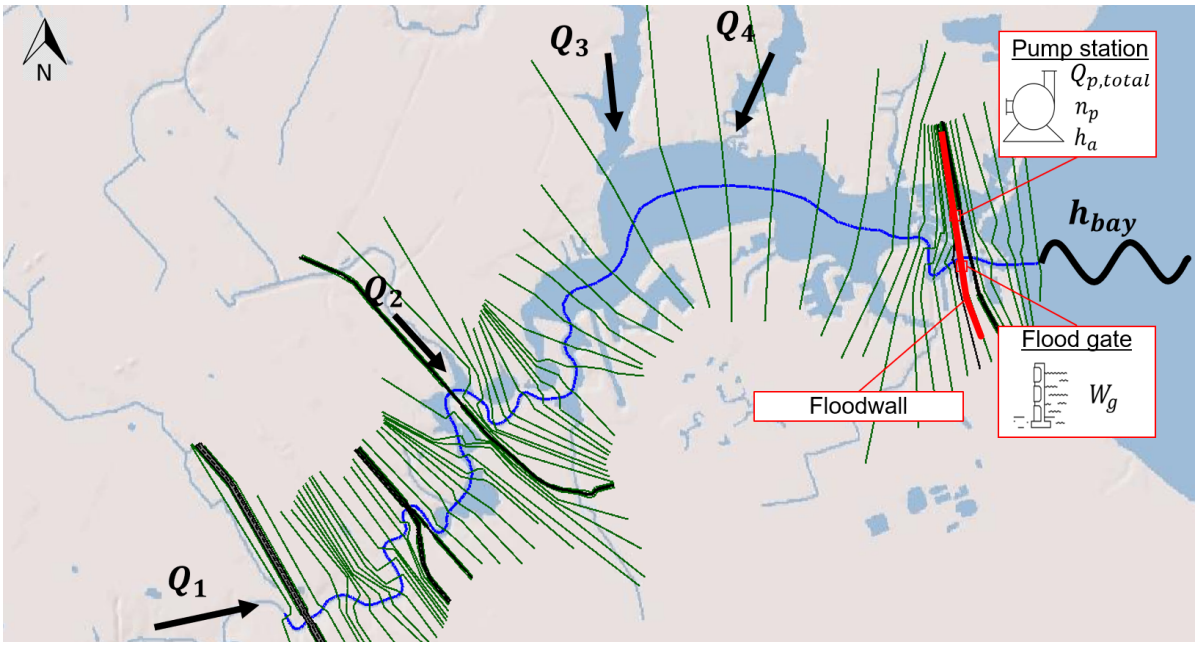


Figure 5.12: Clear Lake system representation showing the location of the boundary conditions, and the design parameters of the infrastructure

Table 5.5: Summary of variables used in the framework.

Symbol	Variable	Type	Unit
$Q_1 - Q_4$	Inflow hydrographs	Forcing (upstream B.C.)	$[\mathrm{m}^3/\mathrm{s}]$
h_{bay}	Water level Galveston Bay	Forcing (downstream B.C.)	$[\mathrm{m}]$
$Q_{p,\mathrm{total}}$	Total Pump Capacity	Design variable	$[\mathrm{m}^3/\mathrm{s}]$
n_p	Number of pumps	Design variable	$[-]$
$h_{a,n}$	Activation water level for pump n	Design variable	$[\mathrm{m}]$
W_{gate}	Width of the flood gate	Design variable	$[\mathrm{m}]$

This system is defined by the upstream and downstream forcings, as well as the design variables controlling the pumps and gates. By systematically varying and sampling these inputs, a comprehensive dataset of system responses can be generated. The methodology for constructing this dataset is presented in the next chapter.

6

Training Data Construction

The development of a reliable deep-learning surrogate model requires a training dataset that is both representative of the underlying hydrodynamic processes and computationally feasible to generate. Since HEC-RAS simulations can be computationally expensive, the construction of the database must be performed strategically. The sampling of input conditions should maximize the information content of the dataset (entropy) while limiting the number of required model runs. In this section, the methodology for constructing such a dataset is presented. The methodology aligns with the framework presented in Chapter 4 and is visualized in Figure 6.1.

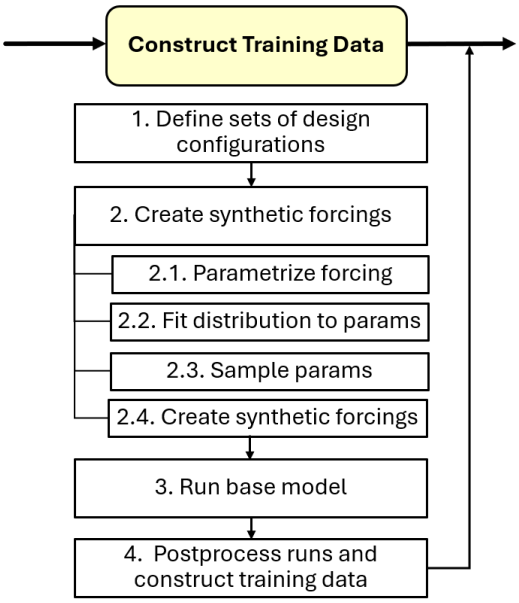


Figure 6.1: Flow Diagram Training Data Generation

This methodology consists of two main components:

- **Defining the Design Configurations:** A 'configuration' is a unique combination of these variables, such as the number of pumps (n_p), their activation water levels (h_{on}), and the gate width (W_g). These are derived from the **Primary Design Variables (PDVs)** identified in Section 4.7. The parameters are sampled to generate construct set of design configurations.
- **Creating the Forcing Scenarios:** Generating a set of external hydrodynamic conditions imposed on the system. A 'scenario' is defined as a set of boundary conditions, namely the four inflow hydrographs and the downstream water level (tidal signal plus storm surge).

Each design configuration was tested under multiple forcing scenarios. By combining the selected configurations with the generated synthetic hydraulic scenarios, a comprehensive set of HEC-RAS simulations can be computed. Each simulation produces time series of the water level of Clear Lake, considered the target variable. The HEC-RAS simulations were performed in an automated framework to ensure efficient execution. The model inputs and outputs were then systematically stored and post-processed into a structured database, forming the foundation for training the deep-learning surrogate model.

6.1. Design Configurations

As established in Section 4.7, the flood mitigation system is defined by the Primary Design Variables (PDVs): Gate Width (W_g), Total Pump Capacity ($Q_{p,total}$), Number of Pumps (n_p), and Pump Activation Water Levels ($h_{on,i}$). To create a feasible and efficient dataset, several simplifying assumptions were made:

- The **total pump capacity ($Q_{p,total}$)** was not sampled directly. Instead, the individual pump capacity was fixed at $Q_p = 56.7\ m^3/s$. The total capacity is then a function of the number of pumps ($Q_{p,total} = n_p \times Q_p$). This simplification was performed to reduce the number of pump input parameters and enhance efficient variations of design configurations. However, the reduction in flexibility should also be acknowledged.
- The pump **deactivation levels (h_{off})** were fixed at 0.15 m below their corresponding activation level.

This leaves three PDVs to be sampled: **gate width (W_g)**, **number of pumps (n_p)**, and **pump activation water levels ($h_{on,i}$)**. The following sections describe how these variables were sampled.

6.1.1. Pump Configurations (Sampling n_p and h_{on})

A representative set of pump configurations was constructed by varying the number of pumps and their activation rules:

- **Number of Pumps (n_p):** Chosen to vary from $\{4, 6, 8, 10\}$. This range accounts for redundancy and the maximum number of pumps per station in HEC-RAS.
- **Activation Staggering (h_{on}):** Three different staggering patterns were considered:
  - **Uniform:** all pumps activated at the same lake level,
  - **Moderate stagger:** activation levels evenly distributed between 0.91 m and 2.06 m,
  - **Strong stagger:** larger spread of activation levels with incremental offsets.

In practice, pump stations often consist of multiple pumps with different capacities and efficiency curves. This is neglected for this study, as we are interested in the primary design capacities (i.e., total capacity required, first estimation of number of pumps, etc.).

This approach balances maximizing entropy with keeping the data generation manageable.

6.1.2. Gate Width (Sampling W_g)

For the gate width (W_g), two representative values were selected:

- $W_g = 22.86$ m, corresponding to the minimum navigation width required for the main channel (USACE and GLO, 2021b).
- $W_g = 60.0$ m, representing the maximum feasible width without substantial dredging.

For the gate operation signal, the gate was modeled as either fully open or fully closed, transitioning from open to closed within a single time step (as described in the previous chapter). To prevent the surrogate model from implicitly learning that the gate always closes during high water levels, 2% of the simulations were run with the gate kept open throughout the entire event. This ensured that the model was exposed to a broader range of hydraulic responses.

6.1.3. Total Design Configurations

By combining the sampled pump and gate variables, a total set of K design configurations was designed.

$$K = (\text{12 Pump Configurations}) \times (\text{2 Gate Widths})$$

Resulting in a total of 24 Design Configurations

6.2. Forcing Scenarios

Synthetic forcings were generated to evaluate each configuration under a range of flow and storm conditions. The process is visualized in Figure 6.2.

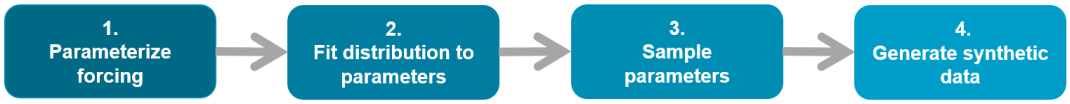


Figure 6.2: From system forcings to synthetic data

6.2.1. Inflow Hydrographs

The inflow hydrographs were represented using a three-parameter Gamma distribution because a quick analysis revealed a good fit of the simulated HMS inflows from previous chapter (Appendix B. This allowed the characterization of each inflow hydrograph (Q_1–Q_4) by three parameters (Figure 6.3):

- q_{peak}: peak discharge,
- α: rising limb shape parameter,
- β: falling limb (decay) parameter.

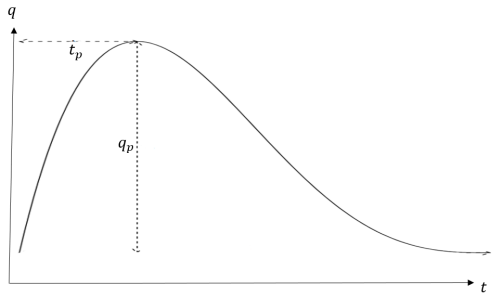


Figure 6.3: Example of a Gamma distribution fitted to a normalized inflow hydrograph. Where q_p and t_p refer to the peak discharge and its offset respectively.

The base hydrograph shape was generated using the Gamma probability density function (Casella and Berger, 2002):

$$g(t;\,\alpha,\beta) = \frac{1}{\Gamma(\alpha)\,\beta^{\alpha}}\,t^{\alpha-1}e^{-t/\beta}, \qquad t \geq 0,$$

The Gamma curve was normalized and scaled using q_{peak}, as previously done by Nadarajah (2007):

$$Q(t) = q_{\text{peak}}\frac{g(t;\alpha,\beta)}{\max\limits_{t} g(t;\alpha,\beta)}.$$

By maintaining a consistent simulation duration, the *time to peak* of each hydrograph can be expressed as a function of the Gamma parameters:

$$t_{\text{peak}} = (\alpha - 1)\,\beta$$

This relation allows the gamma distribution to implicitly capture the timing and offset of the hydrograph peak, eliminating the need to explicitly define the time lag between inflows by an additional parameter. Hence, the three parameters (q_{peak}, α, and β) jointly describe both the discharge magnitude and the temporal characteristics of each tributary inflow.

Parameter Distributions
For each tributary (Q_1–Q_4), multiple HEC-HMS simulations were available, including:

1. 10-, 50-, 100-, and 500-year events based on historical precipitation return periods,
2. the same events with +30% discharge, to align with Mott MacDonald (2021).
3. simulations based on Hurricane Harvey rainfall.
4. simulations based on Tropical Storm Beta rainfall.

Gamma distributions were fitted to these hydrographs to obtain representative values for q_{peak}, α, and β for each tributary. To link the return periods to probabilistic peak discharges, an additional HEC-HMS simulation was performed using the updated HCFCD Atlas 14 precipitation statistics (HCFCD, 2025c).

The statistical distributions adopted for each inflow parameter are summarized in Table 6.1.

Parameter	Forcing Type	Distribution
q_{peak,Q_n}	Inflow	Lognormal
α_{Q_n}	Inflow	Uniform
β_{Q_n}	Inflow	Uniform

Table 6.1: Overview of inflow parameters, forcing types, and probability distributions.

The lognormal distribution for q_{peak} was fitted to the available return period data, representing the probabilistic variability in peak discharge (Figure 6.4). The shape (α) and scale (β) parameters were assumed to follow a uniform distribution. Their lower and upper bounds were directly obtained from the fitted Gamma distributions of the available hydrographs for each tributary.

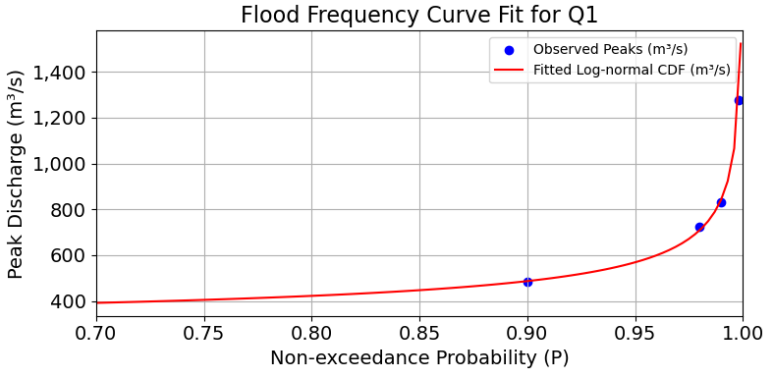


Figure 6.4: Example of a lognormal fit (red line) to the peak discharge (q_{peak}, blue dots) of tributary Q_1.

6.2.2. Downstream Boundary Condition

The downstream boundary condition at the Clear Lake outlet was parameterized to represent realistic storm-surge events originating from Galveston Bay. First an analysis was performed on tong-term water level observations from NOAA Station 8771013 (Eagle Point, TX), covering data from 1995-2025. The first step was to extract these data from NOAA database and convert the water levels and tidal predictions from MSL to NAVDD88 using a vertical offset of -0.107 m (National Geodetic Survey, 2025).

From this data, the non-tidal residuals (NTR), representing the storm surge, was computed as:

$$\mathrm{NTR}(t) = \mathrm{WL}(t) - \mathrm{Tide}(t),$$

where $\mathrm{WL}(t)$ and $\mathrm{Tide}(t)$ are the observed water level and tidal prediction, respectively.

Surge events were identified through threshold exceedance of the NTR. An event was defined when the residual exceeded 0.9 m above the tidal prediction for a minimum duration of 24 hours. The threshold was calibrated by comparing observed surge peaks to return periods reported by Mott MacDonald (2021).

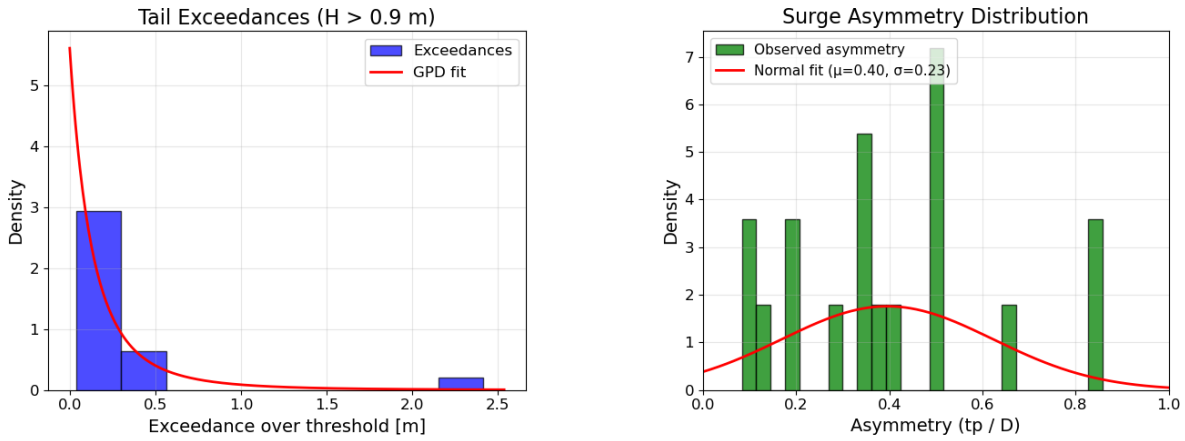
For each detected surge event, the following parameters were extracted and analyzed:

- Peak surge height $h_{\mathrm{peak}} = \max(\mathrm{NTR})$,
- Asymmetry coefficient $h_{\mathrm{asym}} = t_{\mathrm{rise}}/(t_{\mathrm{rise}} + t_{\mathrm{fall}})$.

To be consistent with the sampled hydrographs, the duration (D) of all synthetic storm events was fixed at 24 hours. The surge can then simply be characterized by the peak surge height (h_{peak}) and the asymmetry coefficient (h_{asym}).

Parameter Distributions
The statistical properties of these parameters were then quantified as follows:

- **Peak height (h_{peak})**: modeled using a Generalized Pareto Distribution (GPD) fitted to exceedance above the 0.9 m threshold:
$$F_X(x) = 1 - \left(1 + \xi\frac{x}{\sigma}\right)^{-1/\xi},$$
where ξ and σ denote the shape and scale parameters, respectively. The fitted GPD was used to estimate return-level water heights for 10-, 50-, 100-, and 500-year events.
- **Asymmetry (h_{asym})**: approximated by a normal distribution with mean $\bar{a}$ and standard deviation s_a, truncated to the physical range $(0, 1)$.



(a) GPD fit to surge exceedances

(b) Normal fit to surge asymmetry

Figure 6.5: Parameter distribution fits for synthetic storm surge generation. (a) Generalized Pareto Distribution (GPD) fitted to surge height exceedances over the 0.9 m threshold. (b) Normal distribution fitted to the observed surge asymmetry coefficient.

6.2.3. Parameter Overview and Sampling

To efficiently explore the parameter space, Latin Hypercube Sampling (LHS) was applied. A total of $n = 100$ forcing scenarios were generated across 14 dimensions (12 inflow parameters and 2 downstream parameters). These sampled values were scaled according to their fitted distributions and combined into a dataframe of forcing scenarios.

Parameter	Forcing Type	Distribution
α_{Q_n}	Inflow	Uniform
β_{Q_n}	Inflow	Uniform
q_{peak,Q_n}	Inflow	Lognormal
h_{peak}	Downstream water level	Generalized Pareto
h_{asym}	Downstream water level	Normal

Table 6.2: Overview of parameters, forcing types, and probability distributions.

6.2.4. Automation of HEC-RAS Model Runs

To automate the 2,400 HEC-RAS simulations, the base model was duplicated into run-specific folders. Upstream inflows were generated from the sampled gamma-distribution parameters. To improve stability, a small base flow of 50 cfs was added to each inflow (see Chapter 4).

These inflows, together with the downstream water level time series and the gate operation schedule, were implemented in each run by modifying the .u01 HEC-RAS unsteady flow file. Structural parameters, including the number of pumps, their activation and deactivation water levels, and the width of the flood gate, were added to each run by adjusting the HEC-RAS geometry file.

Both file modifications were fully automated using Python scripts, which used the base model files containing placeholders for each parameter. Each customized model run was then executed using the HEC-RAS 6.6 Controller (COM interface) in silent mode, enabling full automation. This approach ensures consistent input formatting, reproducibility, and efficient handling of a large number of simulations.

6.3. Results and Discussion

The training dataset was constructed by combining 24 unique design configurations with 100 synthetically generated forcing scenarios. This section presents the results of that data generation process, which forms the basis for training the LSTM models.

6.3.1. Forcing Parameter Distributions

The synthetic forcings were generated by sampling from statistical distributions fitted to data from HEC-HMS simulations and historical observations. The methodology and statistical fits for this process were detailed in Section 6.2.

For the inflow hydrographs, peak discharge (q_{peak}) was modeled using a lognormal distribution, as shown in Figure 6.4 (in the Methods section) for the main Q_1 tributary. For the downstream boundary, storm surge height (h_{peak}) was modeled using a Generalized Pareto Distribution (GPD), and surge asymmetry (h_{asym}) was approximated by a normal distribution (see Figure 6.5).

6.3.2. Generated Forcing Scenarios

Using Latin Hypercube Sampling (LHS) on the parameter distributions described above, 100 unique forcing scenarios were generated. Figure 6.6 provides an illustrative example of the resulting synthetic time series for the dominant inflows (Q_1 and Q_3) and the downstream boundary condition, demonstrating the variability in the generated events.

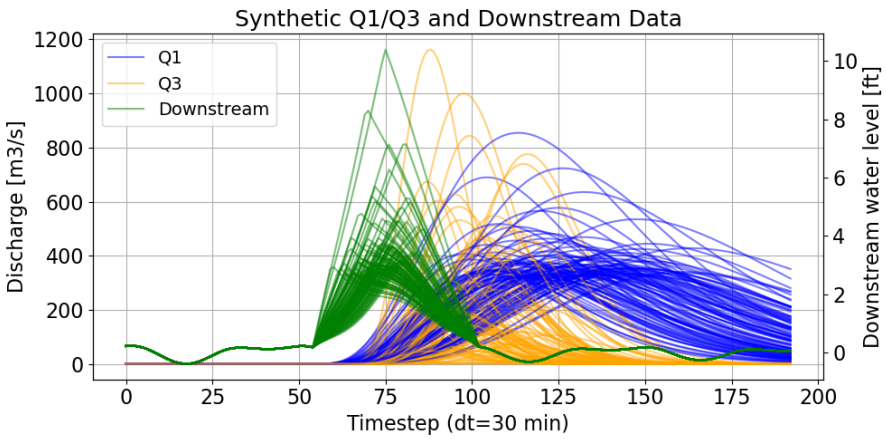


Figure 6.6: Synthetic generated Q1, Q3 and Downstream boundary from sampled parameters

6.3.3. Final HEC-RAS Simulation Outputs

The 24 design configurations were run against the 100 forcing scenarios, resulting in a total dataset of 2,400 HEC-RAS simulations. These automated runs took approximately 8 hours to complete. The final dataset consists of the time-series inputs (forcings, gate operations) and the corresponding HEC-RAS outputs. As visualized in Figure 6.7, the key outputs of interest for surrogate modeling were the Clear Lake water level time series (h_{lake}) and the resulting pump discharge time series (Q_{pump}).

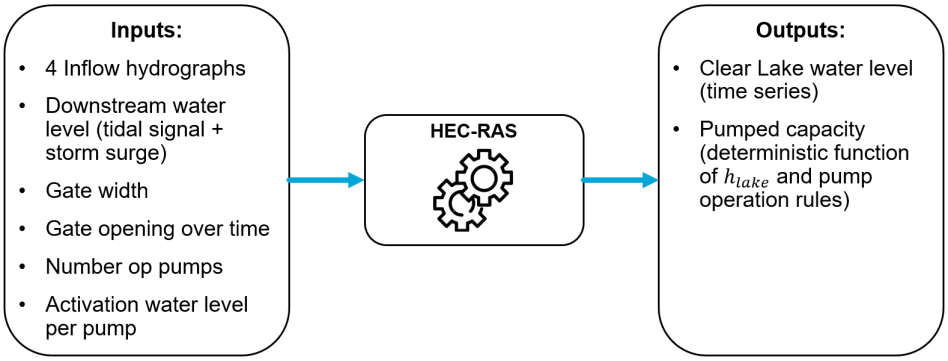


Figure 6.7: HEC-RAS model inputs and outputs

7

LSTM Modeling

This chapter shows the methodology for the development of the surrogate LSTM models. The objective is directly linked to the research sub-question: "How can the simulation dataset be used to best develop an efficient LSTM model that is capable of capturing response in water level to pump and gate variations and operations?"

To answer this, a comparative approach was employed, involving the development of three distinct LSTM Models: A, B and C, representing an information-rich baseline to a more realistic setup (Table 7.1. All models use a **single-task-learning framework**: The target remains the the same for all three approaches: a time series of the Clear Lake water level h_{lake}.

Table 7.1: Overview of LSTM model setups and their objectives

Model	Description	Main Inputs
A	**Baseline model** designed to test the LSTM's ability to learn the system's physics when provided with full information, including the dynamic pump discharge time series. It evaluates whether the model can accurately reproduce the hydrodynamic model output under ideal conditions.	$(Q_{1\text{–}4},\, h_{\text{bay}},\, g(t),\, Q_{\text{pump}})$
B	Represents a **realistic design-oriented** scenario. Instead of the pump discharge time series, the model receives only static pump parameters such as the number of pumps and the pump activation water levels. It predicts water levels directly from the design setup, serving as a surrogate for HEC-RAS in design evaluation.	$(Q_{1\text{–}4},\, h_{\text{bay}},\, g(t),\, \mathbf{p})$
C	**Autoregressive forecaster** that augments inputs with lagged water levels and computes pump discharge from predicted states, trained with scheduled sampling to transition from teacher forcing to self-fed predictions.	$(Q_{1\text{–}4},\, h_{\text{bay}},\, g(t),\, \mathbf{p}, Q_{pump},\, h_{\text{lake},\, t-1:t-n_{\text{lag}}})$

7.1. Model-A: Baseline System Replication

The primary objective of Model-A is to establish a *best-case* performance baseline. By providing the model with complete sequential information about all forcings and operational conditions, its fundamental capacity to replicate the hydrodynamic system is evaluated. If Model-A successfully learns the system dynamics under these idealized conditions, it validates the suitability of the LSTM architecture for this application.

Data Preparation and Features

Hydrodynamic forcings. The four upstream inflows (Q_1–Q_4) and the downstream boundary condition (h_{bay}) were used directly as input features for the LSTM model.

Gate operation. The gate opening was represented as a time-dependent signal $g(t)$ representing the effective gate width at each timestep. The operation schedule was derived deterministically from the sampled gate width W and prescribed opening/closing times t_{open} and t_{close}:

$$g(t) = \begin{cases} 0, & t < t_{\text{open}} \text{ or } t > t_{\text{close}}, \\ W, & t_{\text{open}} \leq t \leq t_{\text{close}}. \end{cases} \tag{7.1}$$

This step function allows the model to learn the impact of discrete gate operations combined with continuous hydrodynamic forcings. The sequence was added to the input feature matrix as sixth feature.

Pump discharge. The HEC-RAS output time series Q_{pump} was extracted and used as the seventh input feature.

Final input matrix. The final input matrix for each run is $X \in \mathbb{R}^{T \times 7}$.

$$X = \left[Q_1, Q_2, Q_3, Q_4, h_{\text{bay}}, g(t), Q_{\text{pump}}\right] \tag{7.2}$$

Target ($\mathbf{y}$): The target variable is the time series of the lake water levels:

$$y = h_{\text{lake}} \in \mathbb{R}^{T \times 1}. \tag{7.3}$$

Model Architecture

The core architecture was based on the common default settings used in hydrology/hydrodynamic modeling and was not extensively optimized in this study. Two stacked LSTM layers with 64 hidden units per layer were used. This configuration was selected as preliminary testing indicated no significant performance gain with a larger hidden size. The output from the last LSTM layer at each timestep was passed through a fully connected (linear) layer, mapping the hidden representation to a single scalar water level value per time step. This resulted in a sequence-to-sequence prediction of the same length as the input sequence.

A dropout rate of 0.2 was used to reduce overfitting. The input dimension was set to match the number of features in the dataset variant used (e.g., F = 7). The output dimension is always equal to 1. Figure 7.1 shows a schematic representation of the architecture.

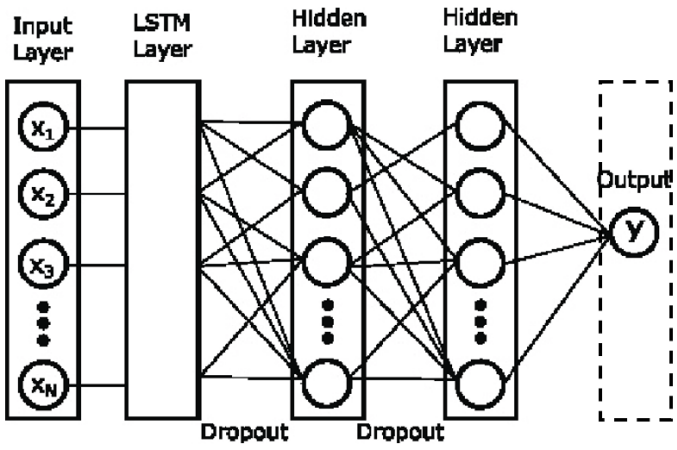


Figure 7.1: LSTM Architecture Schematization (Figure from Compare, Ye, and Meyer-Baese (2023)

Training and Evaluation

Training was performed using mini-batches of input–target pairs $(\mathbf{X}, \mathbf{y})$, drawn from the preprocessed and normalized datasets. Each batch contained $B = 32$ sequences of length T, with F input features.

The models were optimized using the Adam optimizer with a learning rate of 10^{-3}, which was found to provide convergence without excessive tuning. The loss function was the Mean Squared Error (MSE), penalizing deviations between predicted and observed lake levels:

$$\mathcal{L}_{MSE} = \frac{1}{BT} \sum_{b=1}^{B} \sum_{t=1}^{T} \left(\hat{y}_{b,t} - y_{b,t} \right)^2.$$

The Mean Squared Error (MSE) provides a smooth, differentiable measure of the average prediction error, which is well-suited for continuous regression problems such as water level prediction. MSE penalizes larger errors more strongly due to the squared term, making the model sensitive to outliers and encouraging accurate prediction of extreme water levels. At the same time, this also means that large deviations can dominate the loss, potentially biasing the model towards reproducing peak rather than average conditions (Goodfellow, Bengio, and Courville, 2016).

The following training loop structure (Figure 7.2) was followed for this sequence-to-sequence LSTM model.

1. Sample a mini-batch $(\mathbf{X}, \mathbf{y})$ from the training set.
2. Forward pass: propagate $\mathbf{X}$ through the LSTM and fully connected output layer, obtaining $\hat{\mathbf{y}} \in \mathbb{R}^{B \times T \times 1}$.
3. Compute the loss $\mathcal{L}_{\text{MSE}}$ between $\hat{\mathbf{y}}$ and $\mathbf{y}$.
4. Backpropagates gradients through time (BPTT) and update parameters using Adam.
5. Evaluate periodically on the validation set using teacher forcing for consistency.

This setup ensures that the model learns to minimize sequence-wide error while being exposed to realistic time dependencies.

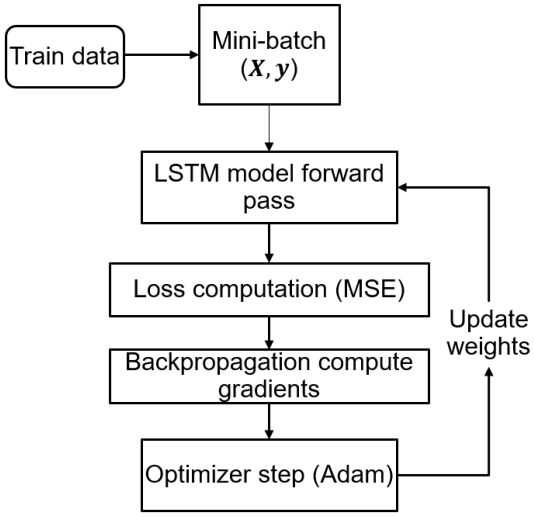


Figure 7.2: General training loop schematic

7.2. Model-B: Static Design Evaluator (HEC-RAS Replacement)

Model B replaces the Q_{pump} time series with static parameters to avoid data leakage, as the pump discharge time series would not be available at inference time.

Dynamic features. Includes the inflows (Q_1–Q_4), downstream BC (h_{bay}), and gate signal $g(t)$ as in Model A.

Static pump parameters (p). The pump operation rules (n_p and $h_{\text{on},i}$) were encoded as a fixed-length static vector. A maximum of 10 pumps was defined, so the vector has 11 entries (1 for n_p, 10 for $h_{\text{on},i}$), with unused slots padded with -999. This vector $\mathbf{p}$ was repeated at every timestep.

$$\mathbf{p} = [n_p, h_{\text{on},1}, h_{\text{on},2}, \ldots, h_{\text{on},10}] \tag{7.4}$$

Final input matrix. Concatenating these features resulted in $X \in \mathbb{R}^{T \times 17}$.

$$X = \left[Q_1, Q_2, Q_3, Q_4, h_{\text{bay}}, g(t), \mathbf{p} \right] \tag{7.5}$$

The target remains $y = h_{\text{lake}}$.

Model Architecture

The network architecture is identical to Model-A, except for the reduced input dimensionality resulting from the removal of $Q_{\text{pump}}(t)$. The model comprises two stacked LSTM layers with 64 hidden units per layer and a dropout rate of 0.2. Static design features are concatenated to the dynamic forcings at each time step, allowing the model to infer the hydraulic effects of different pump configurations over time. The output is a sequence-to-sequence prediction of water levels with one output per timestep.

Training and Evaluation

Training followed the same procedure as in Model-A, minimizing the Mean Squared Error (MSE) between the predicted and reference water levels. Optimization was again performed using the Adam optimizer with a learning rate of 10^{-3} and mini-batches of size $B = 32$.

The performance of Model-B was expected to be lower than that of Model-A, as it must infer operational effects indirectly through the static parameters. The key metric of interest is the performance gap between both models: A small gap would indicate that the static design variables alone are sufficient for the LSTM to capture system behavior accurately, confirming its value as a computationally efficient surrogate for HEC-RAS.

7.3. Model C: Operational Forecaster (Autoregressive)

Model C modifies Model B by adding autoregressive features (past target water levels) and reintroducing Q_{pump} as a dynamic, computed input. This leverages the LSTM's memory capabilities for forecasting, where past water levels are strong predictors of future values.

Base features. Includes all dynamic inputs from Model B (inflows, h_{bay}, $g(t)$), the static pump vector $\mathbf{p}$, and the pump discharge Q_{pump}.

Autoregressive lag features. The input vector at each timestep t was augmented with the previous n_{lag} observed lake levels. A lag length of $n_{\text{lag}} = 20$ was selected based on iterative testing. For a starting point, a Lagged Correlation Analysis was performed on the main hydrodynamic drivers (Q_{1-4} and h_{bay}), included in Appendix D. This analysis indicated that Q_3 had a strong positive lagged correlation with the target at a lag of 13 timesteps. Further testing showed that using $n_{\text{lag}} = 20$ provided the best forecasting performance.

$$\tilde{x}_t = \left[x_t, h_{\text{lake},t-1}, \ldots, h_{\text{lake},t-n_{\text{lag}}}\right] \tag{7.6}$$

This expands the feature dimension to 38 (from $D = 18$ base features + $n_{\text{lag}} = 20$ lag features). The first n_{lag} timesteps are discarded to accommodate the lagged features, resulting in a shorter sequence length of $(T - n_{\text{lag}})$ or 173 timesteps.

Model Architecture

The architecture builds on the sequence-to-sequence LSTM structure used in previous models but accommodates multiple outputs and autoregressive feedback:

- **Input layer:** concatenates dynamic forcings, static parameters, and lagged water levels.
- **LSTM stack:** a set of 2 layers with 64 hidden units, capturing nonlinear temporal dependencies.
- **Output layer:** A fully connected (linear) layer, mapping the hidden representation to a single scalar water level value per time step. This results in a prediction of the target of the same length as the input sequences, which is now reduced by 20 (173 time steps in total).

Training Procedure: Teacher Forcing and Autoregression

Model C required a modification of the general training procedure due to its explicit autoregressive character. Instead of always providing the true water levels to the model, predicted values were progressively fed back as inputs during training. This mechanism required two additional concepts:

- **Teacher forcing:** during early training, the true lake level $h_{\text{lake},t-1}$ was used as input for the next timestep, ensuring stable learning and faster convergence.

- **Autoregression:** later in training, the model's own prediction $\hat{h}_{\text{lake},t-1}$ replaced the true value, simulating deployment conditions.

During teacher forcing, at each time step the model predicts the target variable h_{lake} based on the input feature vector $\mathbf{X}_t$. This vector includes the sequential drivers, static pump features, and the past n_{lag} target values. The predicted $\mathbf{y}_t$ is used as the output for that time step but not as input for the next one. Instead, the true water level is again provided for the following step.

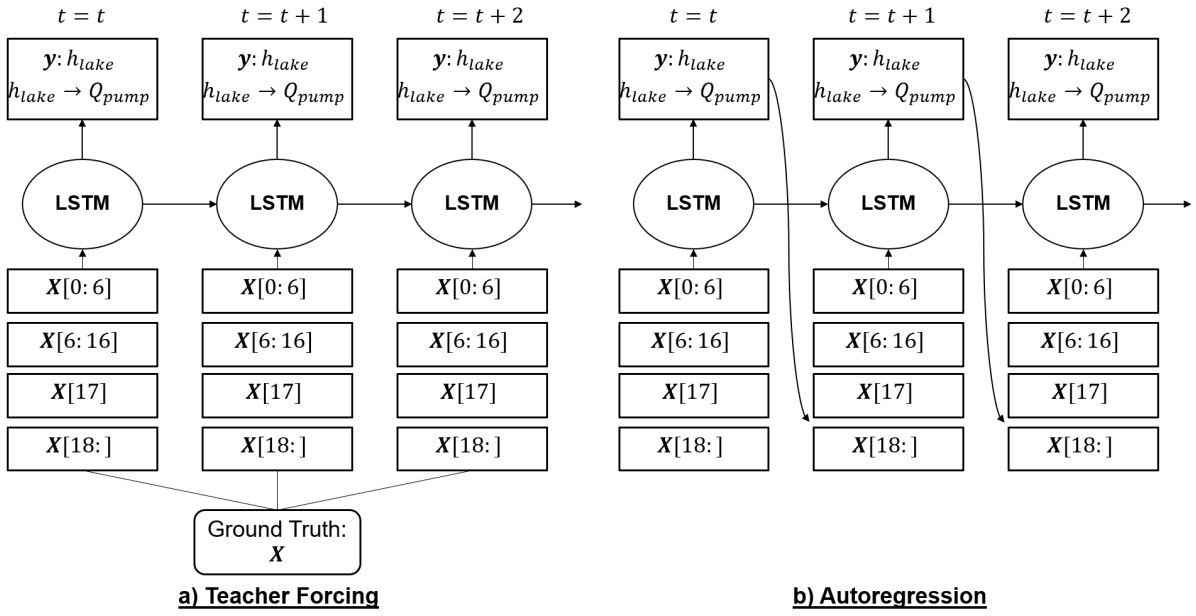


Figure 7.3: Schematization teacher forcing (a) vs autoregressive (b) training

At the inference stage, however, the true target water levels are not available. Therefore, the model must rely on its own predictions: the value predicted at t becomes the input for $t + 1$. This process represents the essence of **autoregression** and is visualized in Figure 7.3

To provide the LSTM with additional information about system response, the predicted water level ($\hat{h}_{\text{lake}}$) was also used to compute the corresponding pump discharge $Q_{\text{pump}}(t)$. During teacher forcing, the true Q_{pump} values were provided (as in Model A), while during autoregression, $Q_{\text{pump}}(t)$ was computed dynamically using the predicted water level and the pump operation parameters stored in $\mathbf{X}_{6:16}$. Figure 7.3 also illustrates this feedback loop.

It is important to note that while Model C incorporates a feedback loop were predicted water level is used to compute the pump discharge (Q_{pump}), it remains a single-task-learning model. The Q_{pump} value is not predicted by the neural network. Instead, it is deterministically computed using the predicted water level h_{lake} and the static pump parameters.This approach contrasts with a multi-task-learning framework, which was not explored in this research. Such an approach would involve training a model to predict multiple targets (e.g., h_{lake} and Q_{pump}).

During training, the model gradually goes from teacher forcing towards autoregression. The process called **scheduled sampling** was applied: a probability p_{TF} determined whether teacher forcing or autoregression was used at each step. This probability decreased linearly over epochs:

$$p_{TF}(e) = p_{start} - \frac{e}{E-1}\big(p_{start} - p_{end}\big),$$

where E is the total number of epochs.

This setup allowed the model to learn the system dynamics and operational feedback based on, during inference, available data. By feeding back past lake water levels and the computed pump capacity per time-step. Valuable extra predictors are added to the input features to allow consistent and accurate predictions that are capable of capturing system response to pump operations.

Loss Function and Evaluation

Training followed the same procedure as in Model-A, minimizing the Mean Squared Error (MSE) between the predicted and reference water levels. Optimization was again performed using the Adam optimizer with a learning rate of 10−3 and mini-batches of size B = 32.

7.4. Splitting and Normalization

Splitting

Before training, the training dataset was split into a training (70%), validation (20%), and test (10%) set:

$$N_{\text{train}} = 0.7N, \quad N_{\text{val}} = 0.2N, \quad N_{\text{test}} = 0.1N.$$

Where N is the total number of simulations (input-output sets). Indices were randomly assigned and saved to ensure same splitting for the distinct models. The training set was used to optimize the model parameters, the validation set to monitor overfitting and tune hyperparameters, and the held-out test set to evaluate the generalization performance. Randomized splitting ensured that the subsets were disjoint.

Scaling and Normalization

To ensure numerical stability and improved model convergence, all datasets were normalized before training. The input features and target variable were first extracted from the preprocessed PyTorch tensors and converted to NumPy arrays. To apply normalization consistently across all samples and timesteps, the three-dimensional tensors $\mathbf{X} \in \mathbb{R}^{N\times T\times F}$ and $\mathbf{y} \in \mathbb{R}^{N\times T\times 1}$ (with N the number of sequences, T the sequence length, and F the number of features) were reshaped into two-dimensional arrays.

Normalization was then performed using a Standard scaling transformation fitted only on the training data:

$$x' = \frac{x - \mu}{\sigma}$$

Where:

- μ is the mean of the feature
- σ is the standard deviation of the feature
- x is the original feature value
- x' is the scaled feature value

Separate scalers were fitted for the inputs and the target, after which the normalized data were reshaped back to the original three-dimensional format required by the recurrent network:

$$\mathbf{X}_{scaled} \in \mathbb{R}^{N\times T\times F}, \qquad \mathbf{y}_{scaled} \in \mathbb{R}^{N\times T\times 1}.$$

For Model C, the lagged target values were scaled using the same standardization parameters as the target variable **y** These scaled lagged water levels were then appended to the input feature matrix **X**.

Finally, the NumPy arrays were converted to PyTorch tensors and wrapped in `TensorDataset` objects for training, validation, and testing. DataLoaders were created with batch sizes of 32 and shuffling for training, ensuring efficient feeding of sequences to the LSTM.

7.5. Overview LSTM Models

Table 7.2 shows a summary of which inputs features are used per model set-up.

Models A and B were trained for a predefined number of 50 epochs, while Model C was trained for 30 epochs. The values were based on preliminary model runs and experiments in which the validation loss was used to identify the point of convergence. Convergence was defined where the validation loss remained within 2% of the final loss, chosen as a practical threshold indicating that the loss had

Table 7.2: Comparison of input features across the three LSTM dataset variants.

Feature	Model A	Model B	Model C
Upstream inflows Q_1–Q_4	✓	✓	✓
Downstream BC h_{bay}	✓	✓	✓
Gate width $g(t)$	✓	✓	✓
Pump discharge Q_{pump}	✓	×	✓
Static pump config $\mathbf{p}$	×	✓	✓
Lagged water levels $h_{\text{lake},t-1:t-n_{\text{lag}}}$	×	×	✓
Input dimension per timestep	7	17	38
Sequence length	193	193	173

effectively stabilized. The training and validation losses were monitored at each epoch to assess convergence and potential overfitting. Overfitting was considered minimal if the validation loss remained close to the training loss and did not exhibit sustained increases.

LSTM Model Parameterization

All LSTMs used 2 stacked layers with a hidden dimension of 64 and a dropout of 0.2 between layers. The final output layer maps the hidden state(s) to a single predicted value per time step. The total number of trainable parameters for each model was calculated using PyTorch's `model.parameters()` function, which accounts for all weights and biases.

The available training data contained 1,680 sequences, each with 193 time steps, resulting in 324,240 target values per model. Comparing the number of trainable parameters to the number of labels (target values per model for training) can provide a rough indication of overfitting risk. While this relation is merely a rough guideline, the underlying principle is well known in machine learning: Models with a high capacity compared to the dataset size are more prone to overfitting (Goodfellow, Bengio, and Courville, 2016; Hastie, Tibshirani, and Friedman, 2009). Table 7.3 summarizes the model configurations, total parameters and the label-to-parameter ratio. As the three models show a label-to-parametric ratio of roughly 5-6, there is no immediate sign that the model complexity is too large for the available data. However, this ratio can only serve as indicative check and does not serve as a reliable predictor whether a model will generalize well.

Table 7.3: Summary of LSTM model architectures, trainable parameters, and overfitting considerations.

Model	Input dim	Hidden dim	Layers	Total Params	Label/Param Ratio
A	7	64	2	52,289	6.2
B	17	64	2	54,081	6.0
C	38	64	2	59,457	5.5

7.5.1. Model Comparison and Evaluation

To systematically evaluate the performance of Models A, B, and C, a comparison was conducted. The effectiveness of the trained models was assessed in four main ways:

- **Training and Validation Performance:** The training and validation loss curves for each model were compared to evaluate convergence speed, stability, and any evidence of overfitting.
- **Quantitative Metrics:** The models' predictive accuracy was measured on the held-out test set using standard regression metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Nash-Sutcliffe Efficiency (NSE).

  RMSE penalizes larger errors more strongly than smaller ones, making it sensitive to occasional large deviations. MAE measures the average magnitude of errors without considering their direction, providing a more intuitive sense of overall accuracy. The Nash–Sutcliffe Efficiency (NSE) assesses how well the predicted values follow the observed variance.

  These metrics were adopted in this research as they together can indicate a good overall performance, where RMSE highlights large errors, MAE gives an average error magnitude, and NSE/R^2

indicate overall goodness of fit. These metrics are commonly used in hydrodynamic/hydrology modeling, and have been employed in comparable studies of surrogate models (A. Li et al., 2024).

- **Visual Assessment:** Model performance was visually inspected by comparing their predicted water level time series against the HEC-RAS ground truth for several representative test scenarios.
- **Robustness and Sensitivity Analysis:** A sensitivity analysis was performed on Model C to assess its robustness and ability to generalize to input parameters outside of the original training distribution. However, this analysis serves purely as indication, as ground truths of the target were not simulated.

8 Results

This chapter presents the results of the LSTM models developed in the previous chapter. It covers the training and validation performance, evaluates the models on the test dataset, and includes a visual assessment of their predictions. Finally, a sensitivity and robustness analysis is presented for Model C.

8.1. LSTM Model Results

8.1.1. Training and Validation Performance

The training-validation loss is presented in Figure 8.1. Both Model A and B reach stable validation losses after approximately 40 epochs, indicating good convergence without severe overfitting. Model C, trained for 30 epochs instead of 50. Hence, the shorter green line shows faster convergence, but also oscillates after that point, which might indicate instability.

Training and validation loss curves

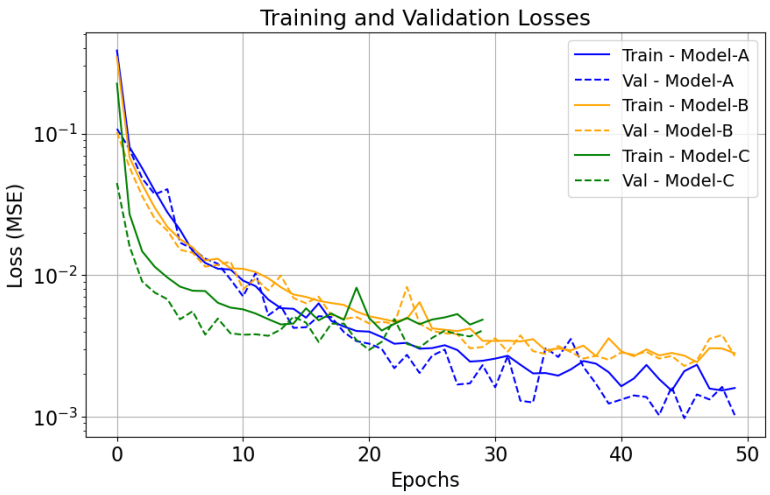


Figure 8.1: Computed model loss curves over epochs for Model A (blue), Model B (orange), and Model C (green).

Training Duration

The training efficiency is summarized in Table 8.1. Model C requires much longer training due to its autoregressive character. The nearly identical computational efficiency of Models A and B suggests that replacing one sequential feature with static values has a negligible effect on training cost.

Generally, all models successfully learn the underlying temporal patterns, but the validation loss behavior of Model C hints that the increased complexity does not lead to improved generalization but likely

Table 8.1: Training duration and efficiency metrics for Models A, B, and C. The epoch of convergence is defined as the first epoch after which the validation loss remains within 2% of the final loss.

Model	Epochs	Total Time (h:min:s)	Avg / Epoch (s)	Epoch of Convergence
Model A	50	00:02:07.5	2.55	44
Model B	50	00:02:09.0	2.58	38
Model C	30	01:37:55.2	195.84 (3.26 min)	8

the opposite.

8.1.2. Computational Efficiency: HEC-RAS vs LSTM

The full training dataset required 2,400 HEC-RAS simulations, resulting in a total runtime of 6 h 15 m 20 s, corresponding to an average of 9.38 s per simulation. This is already computational cheap due to the very small spatial and temporal extent of the used hydrodynamic model.

Once trained, the LSTM surrogate models still provide a strong reduction in inference time. A single forward pass (reflecting 1 simulation) requires only 4.14 ms for Model A, 4.05 ms for Model B, and 313 ms for the fully autoregressive Model C. Even the slowest surrogate (Model C) is approximately 30$\times$ faster than a full HEC-RAS simulation, while Models A and B achieve speed-ups of more than 2,000$\times$.

Table 8.2: Computational efficiency of HEC-RAS and LSTM surrogate models.

Metric	Value
HEC-RAS runtime (2400 simulations)	6 h 15 m 20 s
Average HEC-RAS runtime per simulation	9.38 s
LSTM runtime per forward pass (Model A)	4.1408 ms
LSTM runtime per forward pass (Model B)	4.0536 ms
LSTM runtime per forward pass (Model C)	313.1254 ms
Speed-up Model A vs HEC-RAS	$\sim$2260$\times$
Speed-up Model B vs HEC-RAS	$\sim$2316$\times$
Speed-up Model C vs HEC-RAS	$\sim$30$\times$

8.1.3. Model Evaluation

Quantitative Metrics

The performance of the three surrogates was evaluated on the test data (unseen) using four commonly applied regression metrics: the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), the coefficient of determination (R^2), and the Nash–Sutcliffe Efficiency (NSE). Table 8.3 summarizes the obtained results.

Table 8.3: Performance metrics for the LSTM models.

Metric	Model A	Model B	Model C
RMSE (m)	0.0520	0.0756	0.1234
MAE (m)	0.0343	0.0484	0.0878
NSE (-)	0.9990	0.9979	0.9944

Across all metrics, Model A demonstrates the best overall performance, with the lowest RMSE (0.0520), lowest MAE (0.0343), and highest NSE (0.9990). This indicates that the model reproduces the system dynamics with very high accuracy and minimal error.

Model B shows a small increase in predictive error and a slightly lower NSE, but the values still indicate good capability in capturing overall hydrodynamic responses. The slight decrease in performance suggests that the model may be limited by less sequential information compared to Model A.

Model C exhibits notably larger errors. The larger increase of the RMSE compared to the MAE error indicates the model especially struggles with the higher water level values. Given that most water levels

range from 0 to 2.7 m, the errors (RMSE = 0.123 m, MAE = 0.09 m) are relatively large compared to the target values, likely exceeding what would be considered ideal. However, they are not so large as to make the predictions completely unreliable or meaningless, suggesting that the model can still reproduce water levels in a generally reasonable way for inputs within the training dataset.

The decreased performance is likely due to error accumulation inherent in the autoregressive setup. Despite receiving the same inputs as Model B, along with additional historical water levels and pumped capacities, Model C's predictions show more error than expected. This suggests that in the autoregressive configuration, the gradual accumulation of small errors over time outweighs the potential benefit of including additional input information.

Overall, these results highlight the trade-off between model complexity and sequential prediction stability, emphasizing that Model A is the most robust for accurate surrogate modeling of the hydrodynamic system.

The scatter plot (Figure 8.2) between the predictions and targets for each model strengthens these statements.

Visual Assessment

Figure 8.2 further supports the observations described above. For Model A, the points are tightly clustered around the 1:1 line, indicating an excellent match between predicted and observed values. Model B also shows strong alignment, although the scatter is slightly larger, consistent with its marginally higher RMSE and lower NSE. In contrast, Model C exhibits noticeably greater scatter around the 1:1 line, particularly at higher values, confirming that while the model captures the overall trends, its predictions are less precise compared to Models A and B.

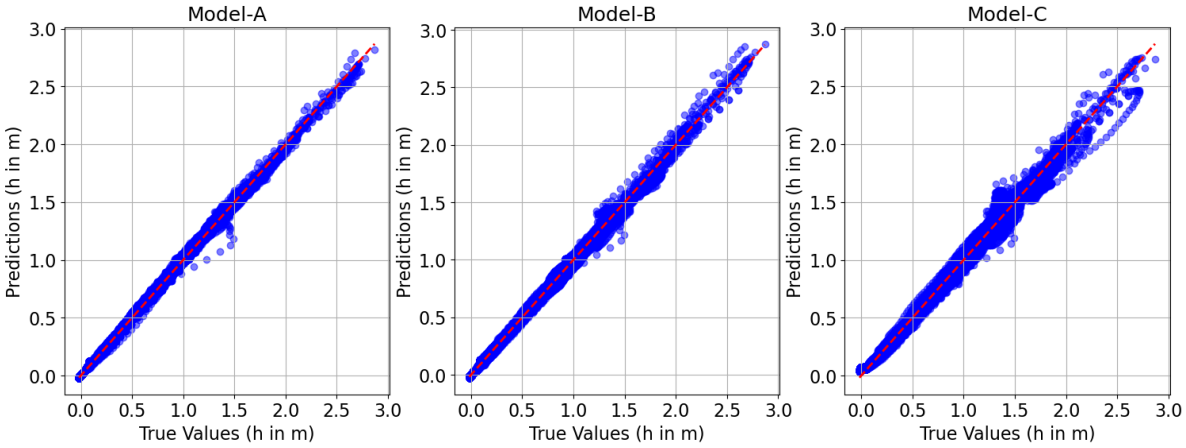


Figure 8.2: Scatter plots of predicted versus observed water levels for Models A, B, and C. Model A shows the tightest clustering around the 1:1 line, indicating the highest predictive accuracy, while Model C exhibits larger scatter, particularly at higher water levels. Deviations around 1–1.5 m suggest challenges in capturing the dynamics of pump activation.

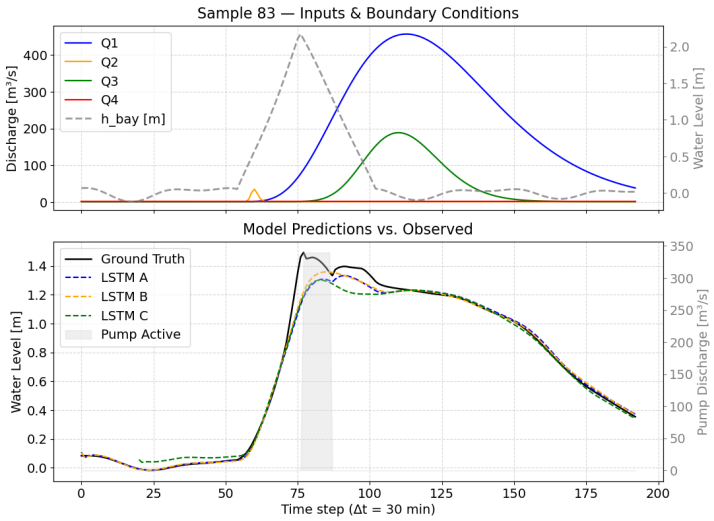
An interesting feature is observed in the central region of the scatter plots (between 1 and 1.5 m for both predicted and observed values). Here, a small deviation from the 1:1 line appears, most pronounced in Model A because of its tighter clustering. This region corresponds to the period when pump activation is initiated in most test runs, suggesting that all models experience some difficulty capturing the precise dynamics of pump operations.

Because evaluating model performance solely using metrics (e.g., RMSE, NSE) can be limiting, a visual assessment of model predictions against ground truth water levels was also conducted. The ground truth corresponds to water levels computed by HEC-RAS for the test set simulations. The test set comprised a total of 240 simulations.

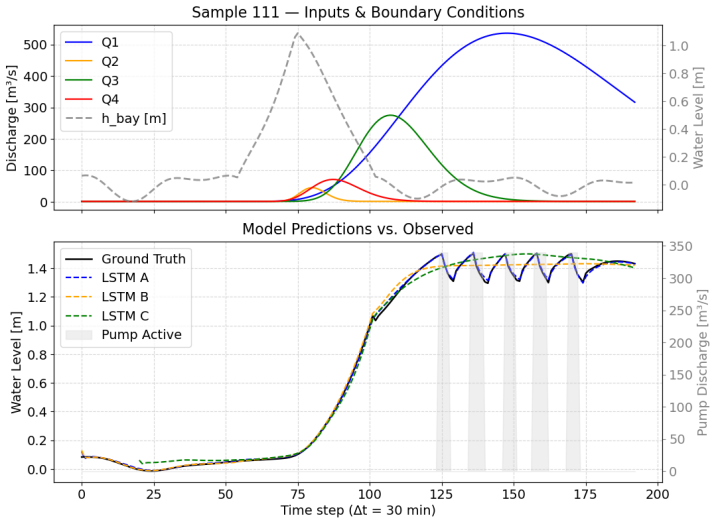
Here, three representative samples are presented to illustrate model performance under different flow scenarios and gate configurations:

- Test sample 83: The gate did not close, allowing a storm surge to enter the system.
- Sample 111: Pump activation settings caused significant variations in water levels, highlighting the models' response to dynamic pump operations.
- Sample 169: A small timing offset between the surge and discharges, combined with high flows, led to very high lake water levels and triggered all pumps.

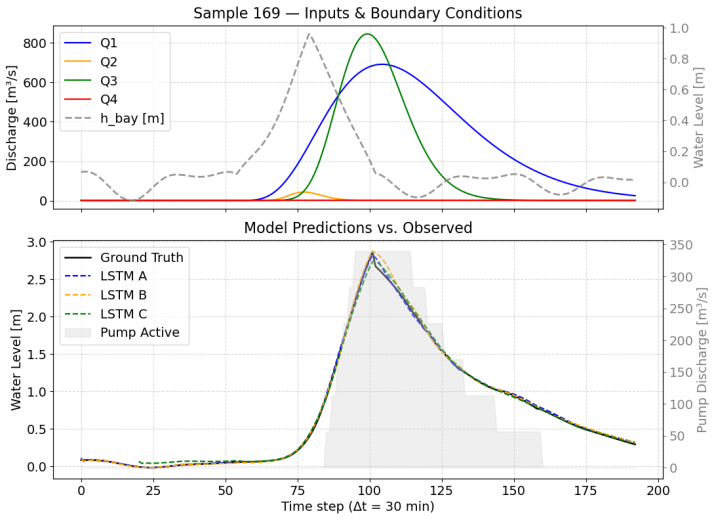
Figure 8.3 shows the predicted water levels for each model compared to the target (black line), along with the corresponding hydraulic forcings (inputs).



(a) Sample 83: Model predictions and corresponding input signals. $W_g = 60.05$ m; 6 pumps activate at $H_{\text{on}} = 1.49$ m.



(b) Sample 111: Model predictions and corresponding input signals. $W_g = 22.86$ m; 6 pumps activate at $H_{\text{on}} = 1.49$ m.



(c) Sample 169: Model predictions and corresponding input signals. $W_g = 60.05$ m; 6 pumps activate between 0.91 m and 2.21 m.

Figure 8.3: Comparison of LSTM model predictions with observed water levels for selected test set samples. Each subplot shows the input inflows and downstream boundary conditions (top) and the predicted and measured water levels (bottom). In light grey, the pump response is shown, based on the pump variables n_p, h_{on}, and the target variable h_{lake}.

Sensitivity and Robustness Model - C

Until now, the LSTM models have only been evaluated on the test set, which lies entirely within the solution space of the training data. Model performance outside of this space has not yet been assessed. To provide a first estimate of Model C's capability to generalize beyond the training data, a series of sensitivity and robustness analyses were conducted. Despite its worse performance, Model C was chosen because it was intended to be used best in an operational inference setting. Testing this model specifically investigates the model robustness if it would be used for inference.

It should be noted that this assessment is purely indicative: ground truth water levels for these scenarios were not generated in HEC-RAS, and the goal is to evaluate whether the LSTM captures the relationships encoded in the training dataset, particularly with respect to feature selection and implementation.

Sensitivity Primary Design Variables

The base case used for visualization was Sample 65, because it average forcings conditions in the test dataset. The hydrodynamic forcings for Sample 65 can be found in figure 8.4. However, the sensitivity analyses were also repeated for multiple base forcings to ensure a more comprehensive evaluation. The results of these can be found in Appendix E.

Gate Width Sensitivity The first parameter tested was gate width. During training, the LSTM was exposed only to two gate widths (22.86 m and 60.96 m). To examine whether Model C can interpolate or extrapolate between these values, simulations were conducted with all other inputs held constant (8 pumps activate between 0.91 m and 2.27 m) while varying gate widths across a wider range. This analysis evaluates whether the model captures the physical response of the system to different gate sizes.

Number of Pumps Next, the number of pumps was varied. The base inflows and other operational parameters were kept the same as in the gate width analysis. Pumps were initially activated uniformly at a water surface elevation of 0.91 m. Sensitivity tests included values both within the training input space (4, 6, 8, 10 pumps) and interpolated or extrapolated scenarios (5, 7 and 3, 12 pumps) to assess the model's ability to generalize.

Pump Activation Levels With gate width and number of pumps fixed (W_g = 22.86 m and 8 pumps), the pump activation levels were varied. This analysis tested whether Model C can reproduce system dynamics under different operational rules, while keeping the hydraulic forcing consistent with previous scenarios. The varying activation levels tested for this analysis are presented in Table 8.4 below.

Table 8.4: Overview of pump activation patterns used in the analysis. Each pattern defines the activation head (m) for all $n_{\text{pumps}} = 8$.

Pattern name	Type	Value range [m]	Example (first–last)
uniform_low	Uniform	2.5 – 2.5	2.5, 2.5, …, 2.5
uniform_mid	Uniform	3.5 – 3.5	3.5, 3.5, …, 3.5
uniform_high	Uniform	4.5 – 4.5	4.5, 4.5, …, 4.5
moderate_low	Linear	2.5 – 3.5	2.5, 2.6, …, 3.5
moderate_mid	Linear	3.0 – 4.0	3.0, 3.1, …, 4.0
moderate_high	Linear	4.5 – 5.5	4.5, 4.6, …, 5.5
strong_low	Linear	2.0 – 4.5	2.0, 2.4, …, 4.5
strong_mid	Linear	3.0 – 5.5	3.0, 3.4, …, 5.5
strong_high	Linear	4.5 – 6.5	4.5, 4.8, …, 6.5

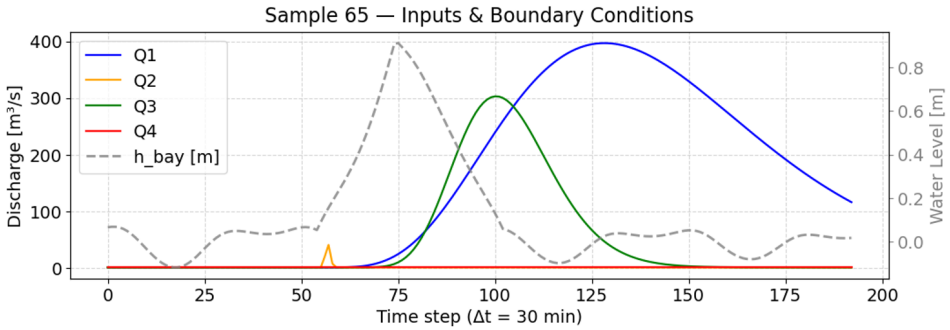


Figure 8.4: Sequential inputs for test sample 65

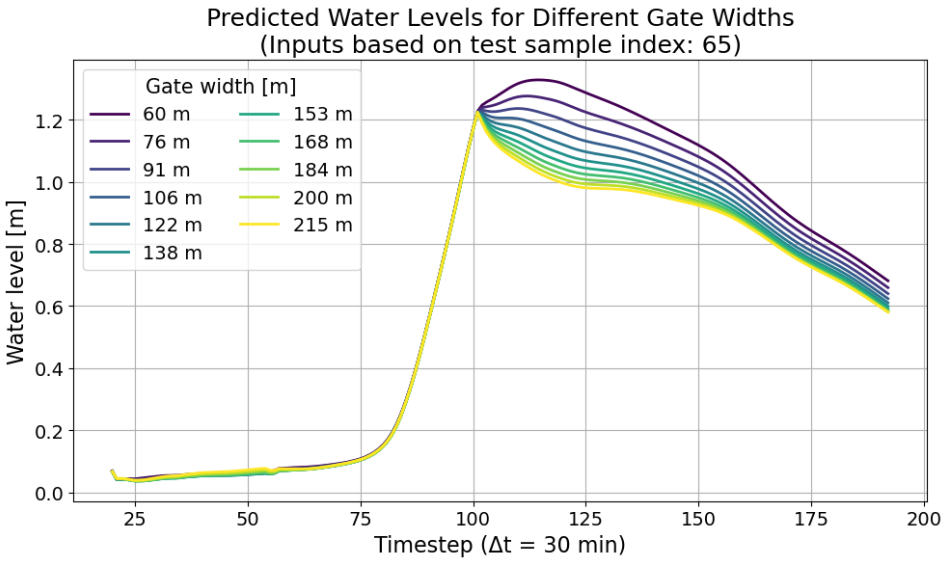


Figure 8.5: Sensitivity of Model C predictions to varying gate widths. Inputs are based on test sample 65.

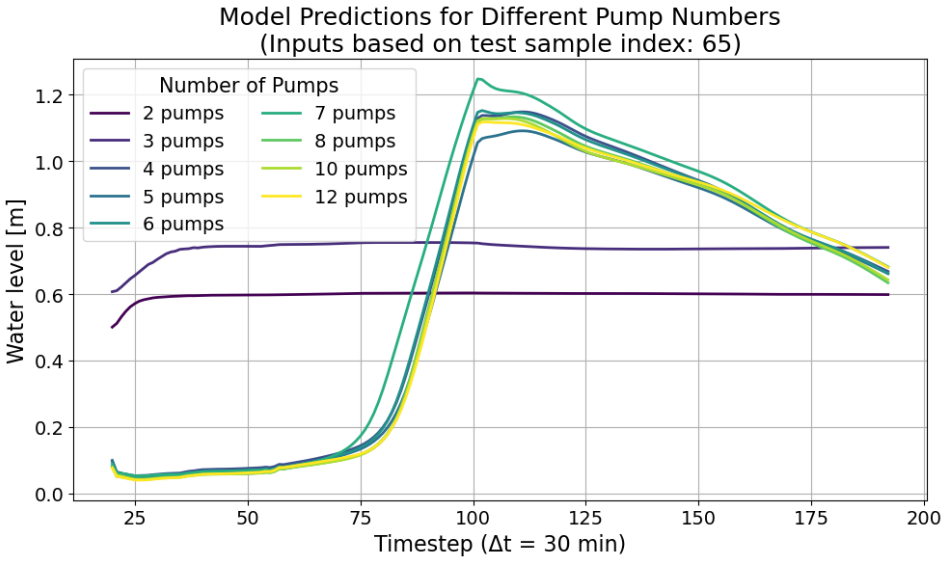


Figure 8.6: Sensitivity of Model C predictions to varying numbers of pumps. Inputs are based on test sample 65.

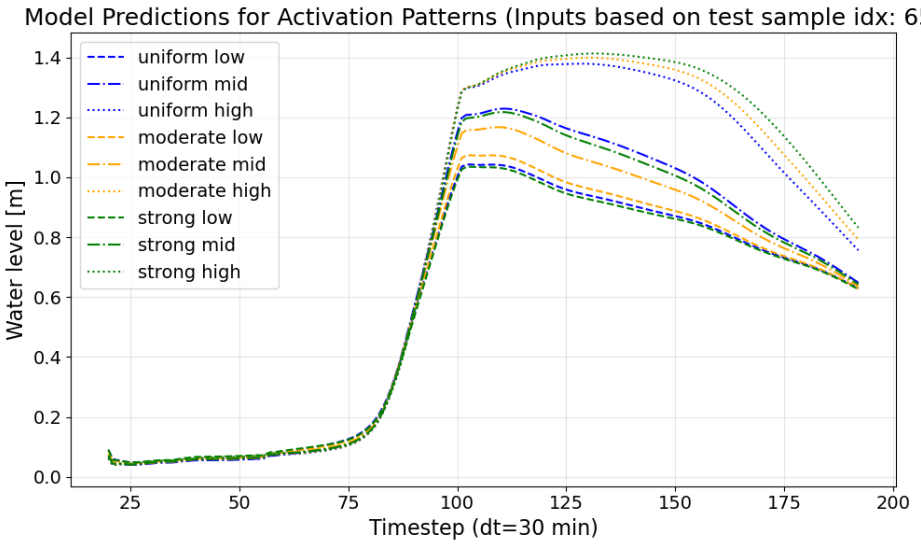


Figure 8.7: Sensitivity of Model C predictions to different pump activation patterns (see Table 8.4). Inputs are based on test sample 65.

Figure 8.5 shows the effect of varying gate width. The model can interpolate and even extrapolate beyond the training values. The predicted water levels remain within physically reasonable ranges, suggesting that the model captures the effect of gate width well. No ground truth is available, so these results are indicative only.

Figure 8.6 presents the model response to changes in the number of pumps. When extrapolating below the training range, the model is unstable. Within the training range, it captures the general trends but does not fully reproduce the expected decrease in water level for more pumps. This indicates that the static pump rules are not perfectly encoded in the model and the model is unable to learn the underlying hydrodynamic patterns.

Figure 8.7 shows results for different pump activation levels and staggering patterns. The model correctly reflects higher water levels when pumps turn on later, but it does not clearly capture differences between staggered and uniform pump operation. Higher activation levels also appear to flatten the water level peaks. Whether this behavior is physically realistic should be compared with HEC-RAS results.

Sensitivity Inflow Discharges

Finally, the model was also tested by using HMS-derived discharges for the 10-, 50-, 100-, and 500-year rainfall events. Several primary design variables were analyzed. Figure 8.8 presents the results for one specific set of design configurations, while the appendix (Appendix D) shows the outcomes for additional design sets.

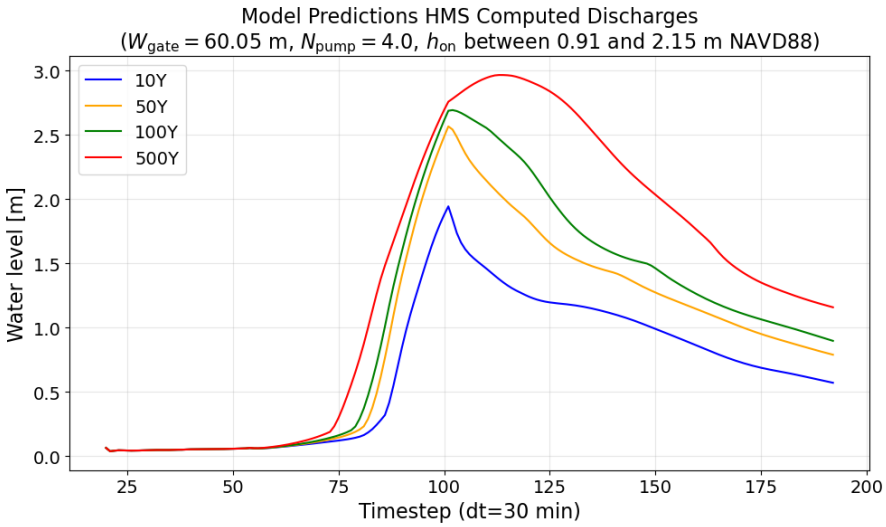


Figure 8.8: Sensitivity of Model C predictions to different, unseen discharge data.

Since no ground truth data are available, these results do not reflect the model's actual predictive performance. However, they indicate that the model remains stable when exposed to previously unseen inflow data.

In summary, the sensitivity analysis suggests Model C learned a physically reasonable relationship for gate width, allowing it to interpolate and even extrapolate reasonably. Although a definitive conclusion cannot be made due to the absence of a ground truth. However, its understanding of pump operations is less robust. It struggles to generalize to pump numbers outside the training range and, while it correctly identifies that higher activation levels lead to higher lake levels, it fails to capture the nuances between different staggering patterns. This indicates that the static encoding of pump rules was not learned successfully.

9

Discussion and Recommendation

This chapter discusses the main outcomes of the study. The first section provides a general interpretation of the results presented in the previous chapter, linking them to broader insights and the overall research objectives. Section 8.2 outlines the main limitations of the study and offers recommendations for future research.

9.1. General Interpretation of Results

The systematic design of the three model setups allows to isolate key aspects of data-driven hydrodynamic modeling. **Model A** demonstrates that the LSTM can successfully learn the governing water balance, when all essential drivers are known (including the pump capacity time series). Although the model does not use explicit physical equation, it effectively captures relations between inflow, outflow, and storage by adjusting the weights in its hidden layers.

However, Model A must be interpreted with care. It requires the pump plan ($Q_{pump}(t)$, which is an output of the HEC-RAS model, as an input. This ”data leakage” means Model A is not a direct replacement for HEC-RAS and cannot be used for a one-to-one comparison. The primary role of Model A is therefore not the practical application. Its good performance simply validates that an LSTM architecture is capable of learning the complex relations of the Clear Lake water balance when all information is provided.

However, the setup for Model A does open a path for a different use case, where the trained model could be used to test the water level impact of any arbitrary pump plan. The challenge is that the HEC-RAS setup should be changed to accept the pump plan as input time series instead of its internal rules. This makes the validation of the LSTM model with the ground truth model more complex.

For future work, this suggests an interesting optimization loop where Model A could be used to find a theoretically ’optimal’ pump plan, and then work backward to determine the design parameters (e.g., h_{on}, n_p) that would be required to reproduce that optimal plan in practice.

Model B highlights the importance of complete sequential dynamic driver features. By replacing the time-varying pump capacity with static pump features, the model loses access to temporal coupling information (on/off behavior) and compensates by interpolating the average relationship between the resulting water level (figure 8.3b). The resulting smoother predictions indicate that the network struggles to capture the system’s response to sharp pump (de)activations. This confirms that data-driven models struggle without complete, high-resolution temporal information (Mohammed et al., 2025).

Model C was designed to overcome the limitation of lacking pump responses by reintroducing feedback through an autoregressive target. The idea was to improve on Model B by: 1) providing extra information in the form past target water levels, and 2) using the predicted target water level at time t to predict pump capacity at the same time. This approach aimed to re-implement the information that Model A had, but Model B missed.

However, Model C performed worse than Model B. The increased complexity did not improve performance. Error accumulation caused the model to misinterpret trigger points that Model A correctly identifies. Because of the autoregressive character, errors propagate over time, reducing overall accuracy. The model does not capture the timing of pump operations and instead tends to smooth out fluctuations by interpolating expected water levels between (de)activation events.

In summary, however the results for Model C were not as expected, this staged framework: Model A $\rightarrow$ B $\rightarrow$ C provides a clear view of how LSTM models handle temporal, and feedback information. The comparison between dynamic and static pump features, confirms that feature encoding is a key aspect of model behavior. While replacing time series of pumped capacity with static attributes improved the presentability and reduced data requirements at interface, it removed the model's ability to directly perceive on-and off triggers. This also demonstrates that for this hydrodynamic system, increasing complexity does not necessarily improve predictive accuracy. Instead, the quality and structure of the input data have a much stronger influence on performance.

Practical Implications

Even with the relatively small and computationally inexpensive hydrodynamic model used in this study, all LSTM surrogate models show a strong increase in computational efficiency. The HEC-RAS model was kept simple to facilitate the generation of training data and to allow straightforward testing of the LSTM training procedure. Despite this simplicity, the LSTM models achieve forward-pass times on the order of milliseconds, representing speed-ups of several orders of magnitude compared to the physics-based simulations.

While for this specific case the HEC-RAS model would still be feasible for Monte-Carlo analyses, the computational gains observed for the surrogate models highlight the potential efficiency benefits for larger, more complex hydrodynamic models. For such expensive simulations, replacing the solver with an LSTM surrogate could reduce computation time from hours or days to seconds, enabling rapid exploration of design and operational scenarios that would otherwise be unreasonable.

9.2. Limitations and Recommendations

As this study aims to develop a comprehensive framework to replace a hydrodynamic model with a data-driven alternative, the work consists of multiple connected steps. Throughout these steps, several assumptions, simplifications, and methodological choices were made. Some of these proved realistic and well-justified, while others turned out to be less effective. As a result a number of aspects remain open for discussion. This section presents the main discussion points and limitations of the thesis and links them to recommendations for future work.

9.2.1. Hydrodynamic Model

1. The first aspect to discuss is the justified choice of the hydrodynamic model. For future design studies, the hydrodynamic model can be replaced by one with improved performance while maintaining the same framework. However, the nature of hydrodynamic forcings (boundary conditions) determine the foundation of the entire framework.

   Although a full coupling between HEC-HMS and HEC-RAS was not feasible for this study, exploring this approach would be highly valuable. A fully coupled setup would be only driven by storm parameters. This would completely change the boundary conditions (from inflow discharges to wind speeds and pressure fields). This would require significant adjustment of the surrogate framework. The sampling process would need to be performed on storm parameters (e.g., wind and pressure fields), which would then be used to generate storm surge and rainfall via hydrodynamic wave, and rainfall-runoff models respectively. Calibration would shift toward matching simulated water levels against the storm-driven forcing conditions.

   Although this approach has been shown to be accurate (for example by Bass and Bedient (2018)), it introduces a trade-off in uncertainty: Overall predictive performance may improve due to a more physically consistent representation of the system, but errors can propagate through the entire model flow and therefore, compound. As a result, it becomes more difficult attribute uncertainty to a specific source compared to separate model frameworks.

In this work, a simplified coupling with a HEC-HMS model was implemented and the downstream boundary condition was imposed separately. This approach does increase the variability of possible scenarios, which potentially improves surrogate model training because of larger entropy. Nonetheless, literature indicates that fully coupled models generally provide more accurate representation of compound flooding (Maymandi, Hummel, and Zhang, 2022; Torres et al., 2015; Bass and Bedient, 2018).

Therefore, it is recommended that future work will investigate a surrogate of a fully coupled modeling approach. Often these models are computationally also more expensive, so it would benefit even more from a faster alternative model

2. The short computational domain of the modified HEC-RAS lies entirely within the tidal range and likely also in the range of backwater curves induced by the pump and gate. As a result, backwater effects are not fully covered within the model boundaries, which limits the accuracy of simulated water level dynamics. For a real assessment of the pump/gate system it is therefore recommended to replace the used, simplified HEC-RAS Model by a model with a large spatial domain. Or to use the full official 1D model (of HCFCD (2022)).
3. Another limitation of this model concerns the representation of storage areas north of Clear Lake, where Armand and Taylor Bayou connect to the lake. In their setup, Mott MacDonald (2021) explicitly included these river sections (B100-00-00 and A104-00-00 for Armand Bayou and Taylor Bayou respectively) because of their storage capacity (HCFCD, 2022).

   This study only represents the inflows from these tributaries (Q3 and Q4). This simplification neglects potential reverse flows from Clear Lake back into the upstream storage areas during high water events. Since rainfall discharges were derived from HEC-HMS (which did not compute backflows) and the downstream boundary condition was mostly cut-off by the flood gat, the effects of this backflow was expected to be limited. However, including these river reaches in future models would allow a more complete representation. Especially under the flow dynamics imposed by the pump/gate infrastructure.
4. Another limitation concerns the simplified representation of the flood gate and pump station. The assumption that the gate can close within a single 30-minute time step is unrealistic in practice. As is the use of a constant pump efficiency or capacity curve. These simplifications are acceptable only in the early design stages but should be replaced by more detailed formulation. Furthermore, the default parameter values described in Chapter 4 and Appendix A should be calibrated and optimized in future studies.
5. The use of a constant time step represents another limitation. Fixing this value at 30 min restricts the model's ability to capture important short-term fluctuations in flow and water levels, particularly those induced by pump and gate operations. Therefore, a smaller time step could also improve the overall model accuracy, especially near the infrastructure.
6. A final limitation and recommendation concerns the recent study by Torres & Associates (2025), which was published during the development of this thesis and could therefore not be fully incorporated into the modeling approach. Their coupled 2D Rain-on-Mesh framework allows rainfall and coastal surge to be applied as boundary conditions, directly addressing the first limitation identified in this section regarding the absence of a fully coupled hydrodynamic setup.

   In addition, the study provides a validated model for the Clear Creek watershed, which could potentially serve as a benchmark for the inflow boundary conditions (Q_{1-4} used in this work. Especially given the absence of streamflow gauges at these locations.

9.2.2. Training Data Generation

The setup of the training database contains several important limitations that influence the diversity of model input data, and hence, the performance. The most important limitations are summarized below;

1. The most significant limitation concerns the implementation of the downstream boundary condition and the gate operations. During the database construction, the timing of the gate closure is kept constant and coinciding with the timing of the storm surge. As a result, the gate always

”cuts-off” the full storm surge, except for the simulations where the gate remains open for the entire simulation. This strongly restricts the patterns the LSTM can learn in two ways:

(a) The timing/offset between the storm surge and the peak discharges can vary. However, the training dataset does not capture the full range of this variability as only the discharge offset was (inertly in the gamma parameters) varied and the start of the storm surge was fixed over all simulations. This means the model learns only a small subset of possible system responses.

(b) Secondly, in most simulations, the storm surge is completely cut off, which does not represent a realistic scenario. Ideally, the gate should open when the inland water level exceeds that of Galveston Bay, allowing for gravitational outflow. A better approach would be to implement this operational strategies in the HEC-RAS model.

2. Another limitation lies in the representation of the downstream boundary condition itself. The simplified triangular shape used for the storm surge may be acceptable for first representation, but in this case, extensive modeling studies and datasets of historical surge events already exist for Galveston Bay.

   In this study, surge events were characterized by calibrating a threshold on peak water levels corresponding to specific return periods from these studies. This approach is debatable since other parameters such as duration and asymmetry also influence the probability of occurrence. Moreover, the limited number of characterized events makes it difficult to reliably relate peak surge levels to return periods. This also becomes evident when comparing the performed analysis to the described literature of the storm surges. Where especially the storm duration was underestimated by the performed analysis compared to the studies of USACE and GLO (2021b).

   These limitation leads directly to the following recommendations. For this specific case, it would be more appropriate to sample the downstream storm surge boundary condition, based on the synthetic storm modeling by USACE and GLO (2021b). This would not only yield more realistic surge scenarios but also allow for more consistent translation between storm surge parameters and their return periods. If the synthetic surge events also account for relative timing to the riverine discharges, a robust and realistic range of hydrodynamic forcing conditions can be generated, leading to more robust models.

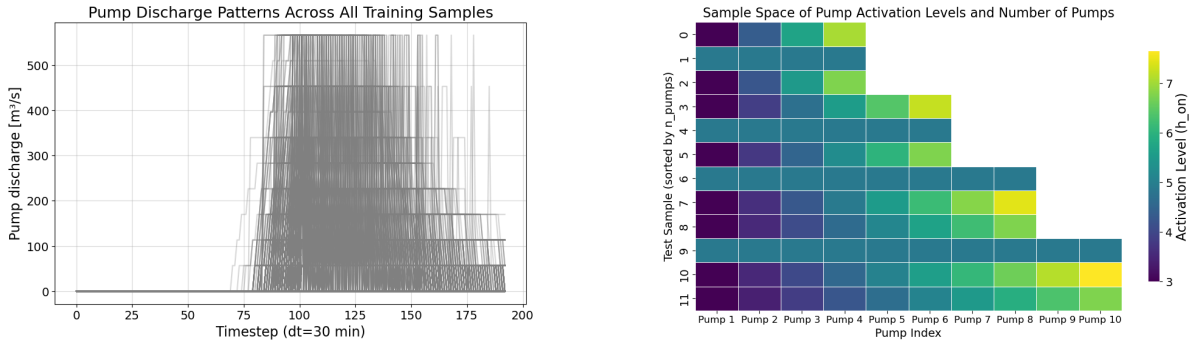
3. A key limitation observed in this framework lies in the encoding of the pump operations for Model B and Model C. In HEC-RAS, a pump plan, meaning the pumped capacity over time, is generated based on the described inputs in Chapter 5.1; the number of pumps and their activation levels. This produces a time series signal describing exactly when each pump turns on or off and hence, the total pump flow over time.

   - Model A uses the full pump time series as input. A sampling strategy of the pump parameters (n_p and h_{on}) was applied to maximize the coverage of the input space (the pump time series), allowing the LSTM to learn the full temporal behavior and total effects of the pumps.
   - For Model B and Model C, this dynamic pump feature was replaced by the sampled static parameters (n_p and h_{on}).

   The problem is that the LSTM no longer directly observes when pumps are turned on or off, but instead has to infer this behavior from the static parameters, which is inherently more difficult.

   Furthermore, the pump time series contains more information, such as the total or maximum pump capacity and the total pumped volume, information that is not explicitly calculated by the LSTM model but contributes valuable input for updating the hidden state weights.

   Finally, the sampling of the parameters n_p and h_{on} was designed to maximize the entropy of the resulting pump time series $Q_p(t)$, ensuring a wide coverage of possible pump operation sequences. This can be seen in Figure 9.1a, where the generated pump signals span a broad range of temporal patterns.



(a) Input Space Pump Plan (Capacity over Time)

(b) Input Space Pump Variables

Figure 9.1: Input space for (a) pump capacity over time, and (b) n_p and & h_{on}

While the time series appear diverse, the actual range of parameter values is relatively narrow and does not fully reflect realistic pump operation scenarios (Figure 9.1b). Therefore, if an LSTM model is used with a parameter-sampling approach, it is important that the static features are sampled in a physically meaningful and realistic way. For this specific case, for example, the inclusion of three different staggering patterns introduced unnecessary complexity in the encoding while providing limited additional information.

A more effective approach would be to fix the staggering pattern and vary only the activation level (h_{on}) and pump capacity. This would enable the model to focus on learning the key indicators, namely; the timing of activation and the maximum pumping capacity, while requiring fewer data samples. Furthermore, despite the limited input space of activation levels, in the current approach the model already seems to learn the pattern of an increased/decreased activation water level (visible in Figure 7.6) in section 7.4).

This consideration also forms the main recommendation for future work: to investigate how static pump design parameters can be best represented to train an LSTM model efficiently. An illustrative example is the gate width, which in this study is included in the time series representation. Although the gate width is not ideally encoded here (as previously discussed), the model was still capable of interpolating between widths, despite being trained on only two distinct configurations.

4. Finally, the report of Torres & Associates (2025) also includes a bivariate analysis of coastal rainfall–surge interactions. This information could be used to construct a more realistic, high-risk training dataset for future surrogate models.

   Sampling boundary conditions directly from the joint distribution presented in their study would further support robust scenario generation at the inference stage. Overall, integrating these findings represents a valuable direction for future work.

9.2.3. Choice for LSTM Architecture

The Long Short-Term Memory (LSTM) network was selected for this study due to its ability to capture temporal dependencies and retain information over extended time periods, making it well suited for modeling sequential hydrodynamic processes (Hochreiter and Schmidhuber, 1997; Kratzert, Klotz, Brenner, et al., 2018). The choice for the LSTM model was primarily guided by examples in literature, as well as by the temporal nature of the data. Hence, alternative data-driven approaches and model architectures that also may be effective for this type of system, were not analyzed. This subsection discusses key considerations and limitations related to the adopted LSTM architecture and suggests directions for future work.

1. **Model capacity.** While the LSTM architecture offers strong temporal learning capabilities, its full representational power may exceed what is necessary for this relatively low-dimensional system. Simpler dynamics may be sufficiently captured by less complex architectures. The LSTM becomes more advantageous when modeling longer sequences, finer temporal resolutions, or systems with pronounced memory effects and delayed responses (Goodfellow, Bengio,

and Courville, 2016).

2. **Hybrid Modeling for Pump Operations.** This study attempted to include all variables within a single LSTM structure. An alternative and potentially more robust approach could be a hybrid modeler setup. The LSTM could focus on predicting the hydrodynamic response to variables/feature it models well (inflows Q_{1-4} and gate $g(t)$). The pump operations, that where poorly represented in Model B and Model C, could be handled by a separate, external model/algorithm. This could be a simple rule based algorithm or small physical-based model. Exploring this separation/hybrid approach in future work could improve interpretability and stability.
3. **Alternative architectures.** Given the high accuracy achieved by some of the current LSTM configurations, future studies could examine simpler neural models such as fully connected Multi-Layer Perceptrons (MLPs), Gated Recurrent Units (GRUs) (Cho et al., 2014), or one-dimensional Convolutional Neural Networks (1D CNNs) (Bai, Kolter, and Koltun, 2018). Particularly for Model A, where the input structure is straightforward and the temporal dependencies are limited, an MLP or GRU could offer comparable performance with reduced computational cost and training complexity.
4. **Advanced architectures.** Conversely, more sophisticated architectures such as Temporal Convolutional Networks (TCNs) (Bai, Kolter, and Koltun, 2018) or transformer-based models (Vaswani et al., 2017) have shown strong performance in capturing long-range dependencies and spatial–temporal patterns. The self-attention mechanism at the core of the Transformer allows the model to look at all timesteps in the input sequence simultaneously and learn direct, non-local relationships between inputs and outputs (Vaswani et al., 2017). This may be highly effective for this problem, as a pump's activation depends on a complex interaction of recent inflows and gate operations (that determine the Clear Lake water level linked to pump triggering). Therefore, a recommendation for future work is to research this type of model, specifically Transformers, for this application.

   An intermediate step could also be to implement an attention mechanism into the standard LSTM used here. Such an "Attention-LSTM" model, has shown successes in other hydrological sequence-to-sequence modeling tasks (Dai et al., 2023)

9.2.4. LSTM-specific Considerations

1. The LSTM models were not extensively **tuned or optimized**. Hyperparameter selection was restricted to a limited set of key parameters in order to maintain a balance between model accuracy, computational efficiency, and training time. While this pragmatic approach ensured stable model convergence, it likely constrained the achievable performance. Future work could apply more systematic optimization techniques such as grid search, Bayesian optimization, or evolutionary algorithms to better explore the hyperparameter space.
2. **Smooth pump activation representation.** Models B and C exhibited difficulties in accurately reproducing the sharp activation and deactivation moments of the pumps, resulting in overly smooth transitions in the predicted water level time series. This effect is partly attributed to the inherently smooth nature of the LSTM's activation functions (typically sigmoid or tanh), which tend to blur abrupt changes in the input signal. Future research could explore whether alternative formulations—such as step-like or threshold-based activation functions, or hybrid physical–data-driven approaches—can improve the representation of such discrete events. Implementing these custom activations in PyTorch, however, remains non-trivial and may require custom gradient definitions.
3. **Evaluation and generalization.** A comprehensive evaluation of the LSTM models on independent, unseen datasets was not conducted within this study. Consequently, the generalization capability of the trained models remains uncertain. Future work should incorporate a more structured and extensive validation procedure, ideally supported by a consistent and representative database that captures a wide range of operational and hydrological conditions.
4. A key limitation of Model C is the accumulation of prediction errors during the autoregressive forecasting. This study did not incorporate error-encapsulation. A technique where a second model is trained to predict the residuals of the primary LSTM to correct its output. These residual

models could learn systematic biases of the primary LSTM and can therefore reduce the drift of error accumulation (e.g., study of Wenchuan et al. (2024)). Implementing a small residual predictor is therefor a promising direction for improving the robustness and stability of Model C.

5. Future work could explore Physics-Informed Neural Networks (PINNs), specifically by modifying the LSTM architecture itself. For instance, the LSTM cell state c_t, which is responsible for the preserved long-term memory (Chapter 3), could be adjusted or constrained to carry a physical state variable (e.g., Hoedt et al. (2021)), such as a flag for the pumps to be ”on” or ”off”. This would force the LSTM model to respect the operational logic rather than learn a statistical approximation. This combines the learning capabilities of an LSTM with hard constrains of the physical system, potentially improving results. Such implementation was not researched in this study, but is definitely interesting to investigate further in future work.

9.2.5. Further Exploration: Improvement of Model B

While Model C attempted to improve Model B by feeding the predicted pump discharge back into the model, its performance remained limited. As discussed in Section 9.2.4, alternative approaches exist to enhance Model B. To explore this, an additional strategy was developed as an iterative improvement of Model B, referred to as Model D.

Model D aimed to better preserve the system’s response to pump (de)activation by explicitly forcing the water-level predictions to reflect the correct triggering behavior. The model structure remained a single-task framework, predicting only the water level, but incorporated the following changes compared to Model B:

- The static pump variables n_p and $h_{on,i}$ were provided once as input conditions instead of being repeated at each time step as in Model B.
- Trigger events for both the true and predicted water-level time series were determined and translated into a binary signal, where a value of 1 indicates that the pumps are triggered (activated or deactivated).
- This binary signal was smoothed using a Gaussian window centered around each triggering point (value 1). This smoothing avoids the instability of hard binary labels and enables gradient-based optimization using Adam (Janocha and Czarnecki, 2017).
- The loss function was extended to a hybrid form, combining the standard regression MSE with a binary cross-entropy term applied to the smoothed trigger signals. A weighting parameter controlled the balance between the two parts, with a value of 0.1 yielding the best performance.

$$\mathcal{L} = \alpha\, \mathcal{L}_{\text{MSE}} + (1-\alpha)\, \mathcal{L}_{\text{BCE}}, \tag{9.1}$$

where $\alpha \in [0,1]$ controls the weighting between the two components. The value of $\alpha = 0.1$ was calibrated interactively. The MSE term $\mathcal{L}_{\text{MSE}}$ was described in Section 7.1.

The binary cross-entropy on the smoothed trigger signals, where $\tilde{s}_{t,p}$ and $\tilde{\hat{s}}_{t,p}$ denote the Gaussian-smoothed true and predicted trigger signals for pump p, is given by:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{TP} \sum_{t=1}^{T} \sum_{p=1}^{P} \left[\tilde{s}_{t,p}\, \log\Big(\tilde{\hat{s}}_{t,p}\Big) + \big(1-\tilde{s}_{t,p}\big)\, \log\Big(1-\tilde{\hat{s}}_{t,p}\Big) \right]. \tag{9.2}$$

Despite the theoretical promise of incorporating explicit trigger information, Model D did not produce substantial improvements compared to Model B when evaluating the performance metrics.

Further gains may be achievable through a more flexible design. For example, treating the binary trigger sequence as an additional predicted output rather than computing it deterministically. Such a multivariate formulation would allow the trigger head to be passed through a dedicated prediction layer and trained with its own loss function, likely improving consistency between predicted water levels and operational behavior.

A more detailed presentation of the experimental results for Model D is provided in Appendix F.

9.2.6. Limitations and Recommendations Overview

In summary, the LSTM architecture provides a robust and interpretable framework for sequence modeling in this context. Nonetheless, exploring both simpler and more advanced model types could offer valuable insights into the trade-offs between model complexity, interpretability, and computational efficiency.

The limitations and their coupled recommendations are summarized in the table 9.1

Table 9.1: Summary of key limitations, their implications, and recommendations.

Limitation	Implication	Recommendation
Hydrodynamic model not fully coupled	Reduced realism of boundary conditions; uncertainty harder to attribute	Explore fully coupled HEC-HMS-HEC-RAS so precipitation can replace simplified inflows
Short HEC-RAS domain	Backwater effects partially missed; limits water level accuracy	Use larger spatial domain like full official 1D model
Simplified storage areas (Armand & Taylor Bayou)	Potential reverse flows ignored; incomplete representation under high flows	Include upstream storage areas and river reaches in future models
Simplified flood gate and pump representation	Unrealistic gate closure; constant pump efficiency; limits temporal accuracy	Use more detailed pump/gate modeling; calibrate parameters
Constant 30-min time step	Misses short-term fluctuations, especially near infrastructure	Reduce time step for improved resolution and accuracy
Downstream boundary condition fixed (storm surge timing)	LSTM sees limited patterns; may not capture realistic variability	Sample surge events based on historical/synthetic data with variable timing relative to river inflows
Simplified storm surge shape	Duration and asymmetry of surge misrepresented	Use synthetic storm datasets to better represent surge characteristics
Pump operations represented by static parameters (Models B & C)	Temporal dynamics not directly observed; harder for LSTM to learn system response	Represent key pump indicators dynamically or sample static parameters realistically; limit unnecessary complexity
Error accumulation in autoregressive forecasts (Model C)	Reduced robustness and stability	Implement residual correction models to mitigate systematic bias
Single-task LSTM	Limits learning for features like pump triggers	Explore multi-output LSTM or hybrid architecture with separate pump predictor
LSTM architecture not optimized	Potential sub-optimal performance	Apply systematic hyperparameter tuning
Smooth LSTM activations	Sharp pump events blurred in predictions	Explore alternative activations or hybrid physical–data-driven approaches
Limited evaluation on unseen data	Generalization uncertain	Use extensive, independent datasets for validation
Recent studies not incorporated (Torres & Associates (2025))	Missing benchmark and improved boundary conditions	Integrate new rainfall-surge coupling and inflow datasets

10

Conclusion

This master's thesis aimed to develop a deep-learning surrogate modeling framework capable of efficiently simulating the hydrodynamic response under different design configurations of the proposed pump and gate system at Clear Lake inlet.

The first objective, to represent this system, was addressed by simplifying an existing 1D HEC-RAS model. Although the validation of this modified model indicated limited performance, it was considered sufficiently accurate to serve as a ground truth for generating the training data required by the surrogate model. The modular approach of the framework ensures that the hydrodynamic model can be replaced by one with better performance.

The second objective addressed how this physical-based model could be used to generate representative and efficient training data. A total of 2400 simulations were produced by sampling key infrastructure parameters and compound flood forcings. The models performed well within sampled parameter space but showed reduced accuracy when applied to unseen conditions, although extensive testing outside of training data was beyond the scope of this study.

Equally important is the choice of sampling parameters itself. In this study, the gate operation schedule was kept constant, and only a limited offset between surge and river peaks was explored. These limitations restricted the analysis of interactions between hydraulic forcings and gate operation controls even though such interactions strongly influence the system behavior. Future work should therefore broaden the sampling of these boundary conditions.

The final objective was to use the simulation dataset for training a surrogate capable of predicting dynamic system behavior from static design parameters. Two key findings were found. First, LSTM networks can successfully capture the governing hydrodynamic processes when provided with all relevant dynamic inputs. Model A, which included the pump discharge as sequential input, successfully reproduced complex water balance dynamics. Second, LSTMs struggle to infer such dynamics from static inputs. When the dynamic pump discharge was replaced by static operational parameters (Model B), performance dropped. A more complex, autoregressive version (Model C) decreased performance due to error amplification. These findings underline that feature encoding is a key component of the framework.

To conclude, LSTM-based surrogate models can emulate hydrodynamic behavior governed by pumps and gates when trained with well-structured, dynamic data. The main challenge lies in teaching the model to predict dynamic behavior based on static design input. A successful framework should therefore either (a) use an LSTM model like Model A, where pump operation logic is given as a time series input or (b) adopt a simpler static surrogate structure with more direct relationships. Future research should prioritize incorporating realistic gate and boundary conditions and testing model robustness under scenarios beyond the training domain before applying the framework for design optimization.

Bibliography

Ai, P. et al. (2018). “Copula-Based Joint Probability Analysis of Compound Flood Events”. In: *Sustainability* 10.7, p. 2232. DOI: 10.3390/su10072232. URL: https://doi.org/10.3390/su10072232.

Amadeo, Kimberly (2018). “Hurricane Harvey facts, damage and costs”. In: *The Balance*.

An, Huicong and Chaojun Ouyang (2025). “A hybrid framework for real-time flash flood forecasting in small ungauged catchments: integrating hydrodynamic simulations with LSTM networks”. In: *Journal of Hydrology* 661, p. 133688. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2025.133688. URL: https://www.sciencedirect.com/science/article/pii/S0022169425010261.

Bai, Shaojie, J. Zico Kolter, and Vladlen Koltun (2018). “An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling”. In: *arXiv preprint arXiv:1803.01271*.

Bass, Benjamin and Philip Bedient (2018). “Surrogate modeling of joint flood risk across coastal watersheds”. In: *Journal of Hydrology* 558, pp. 159–173. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2018.01.014. URL: https://www.sciencedirect.com/science/article/pii/S0022169418300143.

Bevacqua, Emanuele et al. (2019). “Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change”. In: *Science advances* 5.9, eaaw5531.

Brunner, Gary W. (Feb. 2021). *HEC-RAS River Analysis System, Hydraulic Reference Manual*. Tech. rep. Version 6.0. Davis, CA: US Army Corps of Engineers, Hydrologic Engineering Center (HEC).

Casella, George and Roger L. Berger (2002). *Statistical Inference*. Duxbury.

Cho, Kyunghyun et al. (2014). “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation”. In: *arXiv preprint arXiv:1406.1078*.

Compare, Kyle, Ming Ye, and Anke Meyer-Baese (2023). “Using LSTM Neural Networks to Simulate Stage of Wakulla Springs in Northwest Florida”. In: *Proceedings of the 17th Multidisciplinary Conference on Sinkholes and the Engineering and Environmental Impacts of Karst*, pp. 132–140. URL: https://www.researchgate.net/publication/370112068_Using_LSTM_neural_networks_to_simulate_stage_of_Wakulla_Springs_in_northwest_Florida.

Couasnon, A., A. Sebastian, and O. Morales Napoles (2018). “A Copula-Based Bayesian Network for Modeling Compound Flood Hazard from Riverine and Coastal Interactions at the Catchment Scale: An Application to the Houston Ship Channel, Texas”. In: *Water* 10.9, p. 1190. DOI: 10.3390/w10091190. URL: https://doi.org/10.3390/w10091190.

Dai, Zhihui et al. (2023). “A Hydrological Data Prediction Model Based on LSTM with Attention Mechanism”. In: *Water* 15.4. ISSN: 2073-4441. DOI: 10.3390/w15040670. URL: https://www.mdpi.com/2073-4441/15/4/670.

Dang, Thanh Quang et al. (2024). “Application of machine learning-based surrogate models for urban flood depth modeling in Ho Chi Minh City, Vietnam Image 1”. In: *Applied Soft Computing* 150, p. 111031. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2023.111031. URL: https://www.sciencedirect.com/science/article/pii/S1568494623010499.

DSAIE Teaching Team (2024). *Data Science and Artificial Intelligence for Engineers*. Course book for the module *Data Science and Artificial Intelligence for Engineers* at TU Delft.

Federal Emergency Management Agency (2023). *Coastal Flood Frequency and Extreme Value Analysis*. Accessed: 2025-10-18. URL: https://www.fema.gov/sites/default/files/documents/Coastal_Flood_Frequency_and_Extreme_Value_Analysis_Guidance_Nov_2023.pdf.

Federal Emergency Management Agency (n.d.). *Estimated Base Flood Elevation (estBFE) Viewer*. Accessed: 2025-10-12. U.S. Geological Survey. URL: https://webapps.usgs.gov/infrm/estBFE/.

Federal Energy Regulatory Commission (2020). *Chapter R19: Probabilistic Flood Hazard Analysis*. Tech. rep. Engineering Guidelines, Risk-Informed Decision Making. FERC. URL: https://www.ferc.gov/sites/default/files/2020-04/chapter-R19.pdf.

Galveston County Consolidated Drainage District and Institute for a Disaster Resilient Texas (May 2025). *IDRT Study 1: Sedimentation*. URL: https://gccdd.dst.tx.us/downloads/Roadmap_ClearCreekLake.pdf (visited on 10/21/2025).

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. MIT Press. URL: http://www.deeplearningbook.org.

Haces-Garcia, Francisco et al. (2025). "Rapid 2D hydrodynamic flood modeling using deep learning surrogates". In: *Journal of Hydrology* 651, p. 132561. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2024.132561. URL: https://www.sciencedirect.com/science/article/pii/S0022169424019577.

Harris County Flood Control District (2025a). *Harris County Flood Warning System*. URL: https://www.harriscountyfws.org/ (visited on 10/20/2025).

— (2025). *Harris County's Flooding History*. Accessed: 20 October 2025. Harris County Flood Control District. URL: https://www.hcfcd.org/About/Harris-Countys-Flooding-History (visited on 10/20/2025).

— (2025b). *Taylor Lake @ NASA Rd 1 (Gauge #200) real-time water level data*. Accessed: 2025-11-06. Harris County Flood Warning System. URL: https://www.harriscountyfws.org/GageDetail/Index/200.

— (n.d.[a]). *Armand Bayou*. Accessed: 2025-04-12. URL: https://www.hcfcd.org/Activity/Projects/Clear-Creek.

— (n.d.[b]). *Clear Creek*. Accessed: 2025-04-12. URL: https://www.hcfcd.org/Activity/Projects/Clear-Creek.

Harris County Flood Control District (HCFCD) (2022). *Model and Map Management (M3) System*. Accessed: 2025-04-03. URL: https://www.hcfcd.org/interactive-mapping-tools/model-and-map-management-m3-system/.

Harris County WCID 50 (2017). *Hurricane Harvey Flood Report*. Accessed: October 2025. Harris County Water Control and Improvement District No. 50. URL: https://www.harriscountywcid50.org/.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. Springer. ISBN: 978-0-387-84857-0.

HCFCD (2025c). *Rainfall Data – Friendswood Gauge*. Accessed: 2025-07-13. Harris County Flood Control District. URL: https://www.harriscountyfws.org/GageDetail/Index/140?From=7/13/2025%2010:24%20AM&span=24%20Hours&r=1&v=rainfall&selIdx=1.

— (2025d). *Water Level Data for Gauge 100 – Harris County Flood Warning System*. Accessed: 2025-07-12. Includes data from July 12, 2025 over a 24-hour period. Harris County Flood Warning System (FWS). URL: https://www.harriscountyfws.org/GageDetail/Index/100.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long Short-Term Memory". In: *Neural Computation* 9.8, pp. 1735–1780.

Hoedt, Pieter-Jan et al. (18–24 Jul 2021). "MC-LSTM: Mass-Conserving LSTM". In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, pp. 4275–4286. URL: https://proceedings.mlr.press/v139/hoedt21a.html.

Houston Chronicle (2008). *Hurricane Ike Flooding Impacts in Clear Lake Area*. Accessed: October 2025. URL: https://www.chron.com/news/houston-texas/article/Hurricane-Ike-flooding-2008-1764209.php.

Janocha, Katarzyna and Wojciech Czarnecki (Feb. 2017). "On Loss Functions for Deep Neural Networks in Classification". In: *Schedae Informaticae* 25. DOI: 10.4467/20838476SI.16.004.6185.

Kharoubi, Yara et al. (2024). "Asset management for storm surge barriers: how and why?" In: *Structure and Infrastructure Engineering* 0.0, pp. 1–15. DOI: 10.1080/15732479.2024.2391034. eprint: https://doi.org/10.1080/15732479.2024.2391034. URL: https://doi.org/10.1080/15732479.2024.2391034.

Kratzert, Frederik, Daniel Klotz, Christoph Brenner, et al. (2018). "Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks". In: *Hydrology and Earth System Sciences* 22.11, pp. 6005–6022.

Kratzert, Frederik, Daniel Klotz, Guy Shalev, et al. (2019). "Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets". In: *Hydrology and Earth System Sciences* 23.12, pp. 5089–5110.

Leijnse, T. (2018). "Computationally Efficient Modelling of Compound Flooding due to Tropical Cyclones with the Explicit Inclusion of Wave-Driven Processes - Research into the required processes and the implementation within the SFINCS model". PhD thesis. Delft University of Technology.

Leonard, Michael et al. (Jan. 2014). "A compound event framework for understanding extreme impacts". In: *Wiley Interdisciplinary Reviews: Climate Change* 5. DOI: 10.1002/wcc.252.

Li, Ao et al. (2024). "A Deep U-Net-ConvLSTM Framework with Hydrodynamic Model for Basin-Scale Hydrodynamic Prediction". In: *Water* 16.5. ISSN: 2073-4441. DOI: 10.3390/w16050625. URL: https://www.mdpi.com/2073-4441/16/5/625.

Li, X., X. Zhou, J. Hou, et al. (2024). "A Hydrodynamic Model and Data-Driven Evolutionary Multi-Objective Optimization Algorithm Based Optimal Operation Method for Multi-barrage Flood Control". In: *Water Resources Management* 38, pp. 4323–4341. DOI: 10.1007/s11269-024-03867-z.

MacDonald and Tetra Tech (2021). *Design Criteria for Pump Stations (Appendix D, Annex 8)*. Tech. rep. Accessed: 2025-03-13. U.S. Army Corps of Engineers, Galveston District. URL: https://www.swg.usace.army.mil/Portals/26/CTX_MR_AppendixD_Annex8%20%28Design%20Criteria%20Pump%20Stations%29_Final_1.pdf.

Maymandi, Nahal, Michelle A. Hummel, and Yu Zhang (2022). "Compound coastal, fluvial, and pluvial flooding during historical hurricane events in the Sabine–Neches Estuary, Texas". In: *Water Resources Research* 58.12, e2022WR033144.

Mohammed, Sedir et al. (2025). "The effects of data quality on machine learning performance on tabular data". In: *Information Systems* 132, p. 102549. ISSN: 0306-4379. DOI: https://doi.org/10.1016/j.is.2025.102549. URL: https://www.sciencedirect.com/science/article/pii/S0306437925000341.

Mooyaart, Leslie et al. (Jan. 2014). "STORM SURGE BARRIER: OVERVIEW AND DESIGN CONSIDERATIONS". In: *Coastal Engineering Proceedings* 1, p. 45. DOI: 10.9753/icce.v34.structures.45.

Mott MacDonald (Aug. 2021). *Coastal Texas Hydrology and Hydraulics Report*. Prepared for the U.S. Army Corps of Engineers and Texas General Land Office. Appendix D – Annex 2 of the *Coastal Texas Protection and Restoration Feasibility Study: Final Feasibility Report*.

Nadarajah, Saralees (2007). "Probability models for unit hydrograph derivation". In: *Journal of Hydrology* 344.3, pp. 185–189. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2007.07.004. URL: https://www.sciencedirect.com/science/article/pii/S0022169407004040.

NASA Earth Observatory (2017). *Hurricane Harvey: A Slow-Moving Storm Causes Record Flooding*. Accessed October 21, 2025. URL: https://earthobservatory.nasa.gov/images/event/90819/hurricane-harvey (visited on 10/21/2025).

Naseri, K. et al. (2022). “A Bayesian Copula-Based Nonstationary Framework for Compound Flood Risk Assessment”. In: *Journal of Hydrology* 603, p. 127043. DOI: 10.1016/j.jhydrol.2021.127043. URL: https://doi.org/10.1016/j.jhydrol.2021.127043.

Nasr, Ahmed A. et al. (2023). “Temporal changes in dependence between compound coastal and inland flooding drivers around the contiguous United States coastline”. In: *Weather and Climate Extremes* 41, p. 100594. ISSN: 2212-0947. DOI: https://doi.org/10.1016/j.wace.2023.100594. URL: https://www.sciencedirect.com/science/article/pii/S2212094723000476.

Nathan, R. J., E. M. Weinmann, and P. Hill (2003). “Use of Monte Carlo simulation to estimate the expected probability of large to extreme floods”. In: *28th International Hydrology and Water Resources Symposium: About Water; Symposium Proceedings*. Ed. by M. J. Boyd et al. Barton, A.C.T., Australia: Institution of Engineers, Australia, pp. 1.105–1.112. URL: https://search.informit.com.au/documentSummary;dn=350910043129567;res=IELENG.

National Geodetic Survey (2025). *NCAT: NGS Coordinate Conversion and Transformation Tool*. https://www.ngs.noaa.gov/NCAT/. Accessed: YYYY-MM-DD.

National Oceanic and Atmospheric Administration (NOAA) (2018). *2017 U.S. Billion-Dollar Weather and Climate Disasters: A Historic Year in Context*. Accessed: 2025-10-18. URL: https://www.climate.gov/news-features/blogs/beyond-data/2017-us-billion-dollar-weather-and-climate-disasters-historic-year?.

NOAA Center for Operational Oceanographic Products and Services (2025). *Water Level Observations at Station 8771013 (Eagle Point, TX)*. Accessed: 2025-07-12. URL: https://tidesandcurrents.noaa.gov/stationhome.html?id=8771013.

NOAA National Centers for Environmental Information (2024). *Storm Events Database*. Accessed: October 2025. URL: https://www.ncei.noaa.gov/maps/hazards/.

Rivas-Blanco, Irene et al. (Mar. 2021). “A Review on Deep Learning in Minimally Invasive Surgery”. In: *IEEE Access* PP, pp. 1–1. DOI: 10.1109/ACCESS.2021.3068852.

Sebastian, Antonia (2022). “Compound flooding”. In: *Coastal flood risk reduction*. Elsevier, pp. 77–88.

Shen, Chaopeng (2018). “A transdisciplinary review of deep learning research and its relevance for water resources scientists”. In: *Water Resources Research* 54.11, pp. 8558–8593.

Sit, Manuela et al. (2020). “A comprehensive review of deep learning applications in hydrology and water resources”. In: *Water* 12.9, p. 2711.

Song, M. et al. (2024). “Time-Varying Copula-Based Compound Flood Risk Analysis”. In: *Journal of Flood Risk Management* 17.2, e13032. DOI: 10.1111/jfr3.13032. URL: https://doi.org/10.1111/jfr3.13032.

Texas A&M University (2015). *Texas Climate Summary 2015*. Accessed: October 2025. Texas A&M University, Office of the State Climatologist. URL: https://climatexas.tamu.edu/.

Texas Stream Team (2024). *Clear Creek Watershed Data Summary Report*. Texas State University. URL: https://docs.gato.txst.edu/696970/2024ClearCreek.pdf.

Thatcher, Cindy A., John Brock, and Elizabeth A. Pendleton (2013). “Economic vulnerability to sea-level rise along the northern U.S. Gulf Coast”. In: *Journal of Coastal Research* SI63. Spring 2013, pp. 234–243. DOI: 10.2112/SI63-017.1. URL: https://pubs.usgs.gov/publication/70121472.

Torres, Jacob M. et al. (2015). “Characterizing the hydraulic interactions of hurricane storm surge and rainfall–runoff for the Houston–Galveston region”. In: *Coastal Engineering* 106, pp. 7–19. ISSN: 0378-3839. DOI: https://doi.org/10.1016/j.coastaleng.2015.09.004. URL: https://www.sciencedirect.com/science/article/pii/S0378383915001520.

Torres & Associates (Apr. 2025). *Clear Creek-Frontal Galveston Bay Baseline Conditions*. GLO River Basin Flood Study (RBFS) - Central Region Deliverable #47 / Contract No. 20-094-002. Texas General Land Office. URL: https://dmqt.dev.cloud.tdis.io/ (visited on 11/16/2025).

Torres Dueñas, L.F. (2018). “Flood Risk Assessment of the Clear Creek Watershed considering compound events”. Master Thesis; Mentors: Bas Jonkman, Jeremy D. Bricker, Toni Sebastian; Graduation committee: K.M. De Bruijn, M. Hrachowitz. MA thesis. TU Delft, Civil Engineering & Geosciences. URL: https://resolver.tudelft.nl/uuid:3d2791c1-902f-4107-a5ed-2bb41d4de592.

U.S. Army Corps of Engineers Galveston District (2024). *Clear Lake and Dickinson Bay*. Accessed: 2025-10-18. URL: https://storymaps.arcgis.com/stories/bacf6de01b0c43a5825c8ea4705c83e7.

U.S. Army Corps of Engineers, Southwestern Division (2021). *CTX MR Appendix D Annex 7 (Pump Station Floodgate Plates)*. Technical Report. Accessed: October 2025. U.S. Army Corps of Engineers, Southwestern Division. URL: https://www.swg.usace.army.mil/Portals/26/CTX_MR_AppendixD_Annex7%20%28Pump%20Station%20%26%20Floodgate%20Plates%29.pdf.

U.S. Geological Survey (2025). *USGS 08077600 Clear Ck nr Friendswood, TX*. URL: https://waterdata.usgs.gov/monitoring-location/USGS-08077600/ (visited on 10/20/2025).

USACE and GLO (Aug. 2021a). *Coastal Texas Protection and Restoration Feasibility Report*. Tech. rep. FINAL. Accessed: 2025-02-24. U.S. Army Corps of Engineers and Texas General Land Office. URL: file:///C:/Users/aukes/Downloads/Coastal%20TX%20Protection%20and%20Restoration%20FINAL%20Feasibility%20Report_20210827(1).pdf.

— (2021b). *Coastal Texas Protection and Restoration Study: Appendix D - Engineering*. Accessed: 2025-03-04. Galveston, TX: U.S. Army Corps of Engineers, Galveston District and Texas General Land Office. URL: https://www.swg.usace.army.mil/Portals/26/CTX_MR_Appendix%20D_Engineering_2020823_1.pdf.

Vaswani, Ashish et al. (2017). “Attention is All You Need”. In: *Advances in Neural Information Processing Systems* 30.

Wahl, Thomas et al. (2015). “Increasing Risk of Compound Flooding from Storm Surge and Rainfall for Major US Cities”. In: *Nature Climate Change* 5.12, pp. 1093–1097. ISSN: 1758-6798. DOI: 10.1038/nclimate2736. URL: https://doi.org/10.1038/nclimate2736.

Wang, H. (2023). “A Nonstationary Multivariate Framework for Modeling Compound Flooding”. In: *Environmental Modelling Software* 160, p. 105525. DOI: 10.1016/j.envsoft.2023.105525. URL: https://doi.org/10.1016/j.envsoft.2023.105525.

Weathington, Briggs Tomás (2023). “Coupled Ocean-Hydraulic Modeling for Socially Informed Compound Infrastructure Analysis in Clear Lake, TX”. MA thesis. Rice University.

Wenchuan, Wang et al. (2024). “Error correction method based on deep learning for improving the accuracy of conceptual rainfall-runoff model”. In: *Journal of Hydrology* 643, p. 131992. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2024.131992. URL: https://www.sciencedirect.com/science/article/pii/S002216942401388X.

Wikipedia contributors (2025). *Tropical Storm Erin (2007)*. Accessed: October 2025. URL: https://en.wikipedia.org/wiki/Tropical_Storm_Erin_(2007).

Ziya, Oveys and Ammar Safaie (2023). “Probabilistic modeling framework for flood risk assessment: A case study of Poldokhtar city”. In: *Journal of Hydrology: Regional Studies* 47, p. 101393. ISSN: 2214-5818. DOI: https://doi.org/10.1016/j.ejrh.2023.101393. URL: https://www.sciencedirect.com/science/article/pii/S2214581823000800.

A

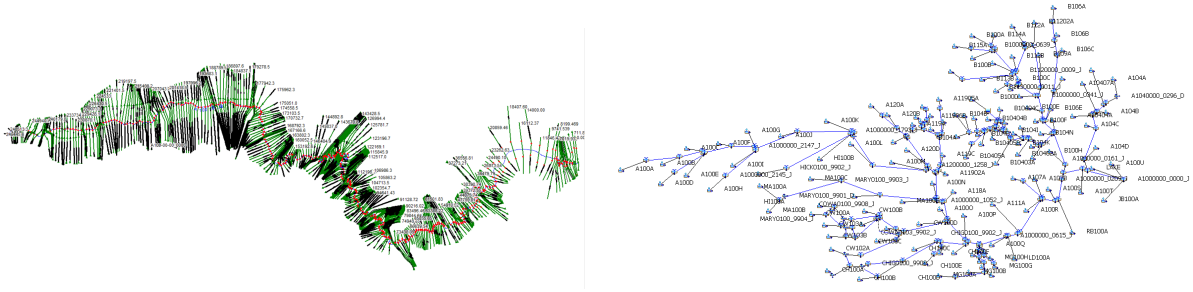
System Representation

A.1. Additional Gate Parameters

Parameter	Value
Gate type	Radial
Radial Discharge Coefficient	0.7
Trunnion Exponent	1
Opening Exponent	0.5
Head Exponent	0.5
Trunnion Height	-11
Orifice Coefficient	0.8
Head reference	Sill (Invert)
Weir Shape	Broad Crested
Weir Coefficient	1

Table A.1: Additional gate parameters

A.2. Models USED Mott MacDonald (2021)



(a) HEC-RAS model Clear Creek

(b) HMS Model Clear Creek Watershed

Figure A.1: Hydrodynamic and hydrologic model of Clear Creek watershed (HCFCD, 2022)

B

Database construction

B.1. Inflow Distribution Fitting

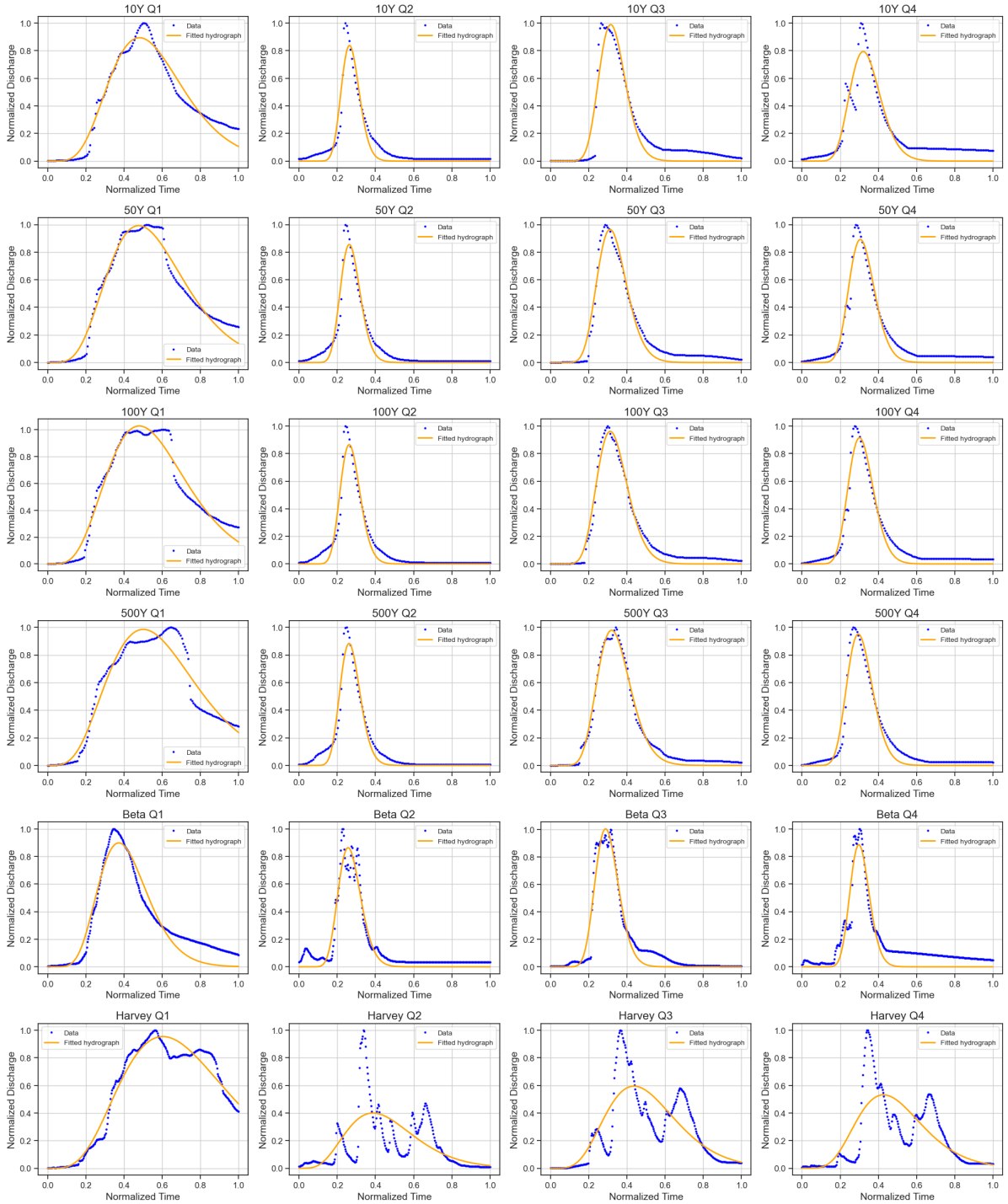


Figure B.1: Gamma distribution fit to available hydrographs

B.2. Pump Configurations

# Pumps	Q_{pump} (m^3/h)	h_{on} (m)	h_{off} (m)	Pattern	Gate Width (m)
4	2000	[4.88, 4.88, 4.88, 4.88]	[4.38, 4.38, 4.38, 4.38]	uniform	75
6	2000	[3.00, 3.75, 4.50, 5.25, 6.00, 6.75]	[2.50, 3.25, 4.00, 4.75, 5.50, 6.25]	moderate	75
8	2000	[3.00, 3.64, 4.27, 4.91, 5.54, 6.18, 6.81, 7.45]	[2.50, 3.14, 3.77, 4.41, 5.04, 5.68, 6.31, 6.95]	strong	75
10	2000	[4.88, 4.88, 4.88, 4.88, 4.88, 4.88, 4.88, 4.88, 4.88, 4.88]	[4.38, 4.38, 4.38, 4.38, 4.38, 4.38, 4.38, 4.38, 4.38, 4.38]	uniform	75
4	2000	[3.00, 4.25, 5.50, 6.75]	[2.50, 3.75, 5.00, 6.25]	moderate	75
6	2000	[3.00, 3.85, 4.70, 5.55, 6.40, 7.25]	[2.50, 3.35, 4.20, 5.05, 5.90, 6.75]	strong	75
8	2000	[4.88, 4.88, 4.88, 4.88, 4.88, 4.88, 4.88, 4.88]	[4.38, 4.38, 4.38, 4.38, 4.38, 4.38, 4.38, 4.38]	uniform	75
10	2000	[3.00, 3.42, 3.83, 4.25, 4.67, 5.08, 5.50, 5.92, 6.33, 6.75]	[2.50, 2.92, 3.33, 3.75, 4.17, 4.58, 5.00, 5.42, 5.83, 6.25]	moderate	75
4	2000	[3.00, 4.35, 5.70, 7.05]	[2.50, 3.85, 5.20, 6.55]	strong	75
6	2000	[4.88, 4.88, 4.88, 4.88, 4.88, 4.88]	[4.38, 4.38, 4.38, 4.38, 4.38, 4.38]	uniform	75
8	2000	[3.00, 3.54, 4.07, 4.61, 5.14, 5.68, 6.21, 6.75]	[2.50, 3.04, 3.57, 4.11, 4.64, 5.18, 5.71, 6.25]	moderate	75
10	2000	[3.00, 3.52, 4.03, 4.55, 5.07, 5.58, 6.10, 6.62, 7.13, 7.65]	[2.50, 3.02, 3.53, 4.05, 4.57, 5.08, 5.60, 6.12, 6.63, 7.15]	strong	75

Table B.1: Overview of pump configurations for varying numbers of pumps, with rounded h_{on} and h_{off} levels.

C

Tidal Analysis Eagle Point

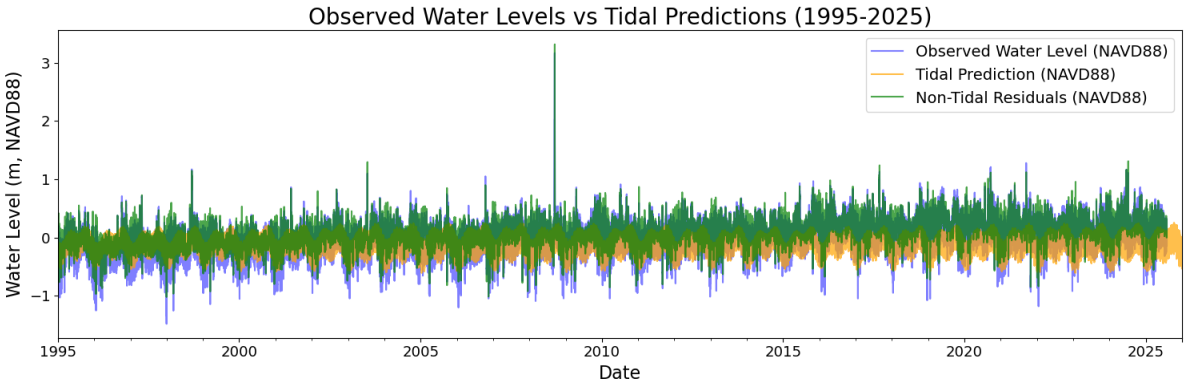


Figure C.1: Measured tidal signal at Eagle Point over the analysis period.

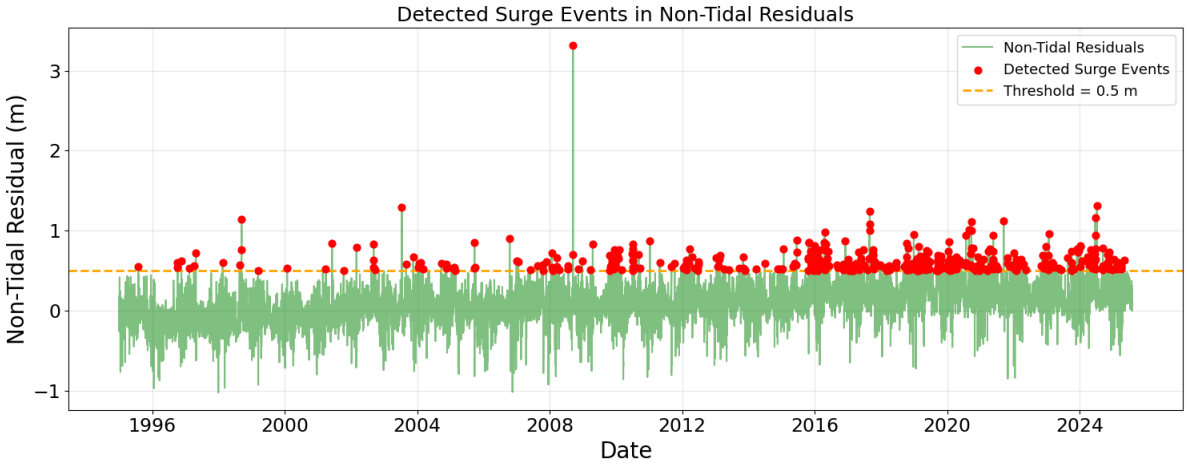


Figure C.2: Detected tidal events using a 0.5 m threshold.

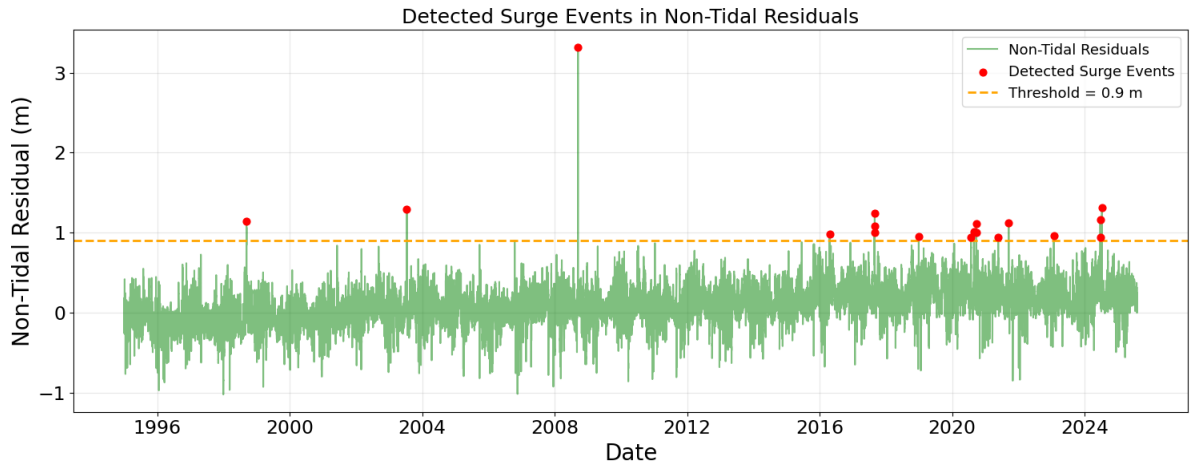


Figure C.3: Detected tidal events using a 0.9 m threshold.

D

Feature Engineering and Reprocessing

Before training the surrogate model, additional steps were applied to the input-output database to gain understanding between de relations between feature and target. By getting better understanding of the relations and correlations, more insight is gained and the training dataset can potentially be structured in a more efficient way.

Two complementary analyses were conducted: First, a lag correlation analysis (LCA to quantify time shifted dependencies between predictors and the target water level. Secondly, a cross-correlation analysis to asses mutual dependencies between inputs themselves and the target. These analyses were performed on the **dynamic forcing features**, meaning the 5 hydrodynamic input forcings (inflows and downstream water level), as the gate and pump signal as they are also sequences and impact the behavior of the target water level. These steps serve to indicate leading/lagging relationships in the time series and to identify potentially collinear predictors that can than be removed from the training set.

D.0.1. Lagged Correlation Analysis (LCA)

The LCA quantifies the strength of correlation between each input feature $x_f(t)$ and the target water level $y(t)$ at different temporal lags. To efficiently compute correlations over a range of time lags, a Fast Fourier Transform (FFT)-based cross-correlation method was employed. For each run r and feature f, the cross-correlation function was evaluated as

$$\rho_{f,r}(\tau) = \frac{1}{\sigma_{x_f}\,\sigma_y} \sum_{t=1}^{T} \Big(x_{f,r}(t) - \mu_{x_f}\Big)\,\Big(y_r(t+\tau) - \mu_y\Big), \tag{D.1}$$

where $\tau \in [-\tau_{\max}, \tau_{\max}]$ is the considered lag window, μ and σ denote mean and standard deviation, and T is the sequence length. By convention, a positive lag $\tau > 0$ indicates that the input leads the target. The resulting lag-correlation curves $\rho_f(\tau)$ were then averaged across all runs to obtain robust estimates of temporal dependencies, with ± 1 standard deviation envelopes used to indicate variability.

Figure D.1 below shows the results of the mean lagged correlation over the runs with their standard deviation per time lag for a maximum time lag $\tau_{\max} = 60$ timestep of 30 minutes.

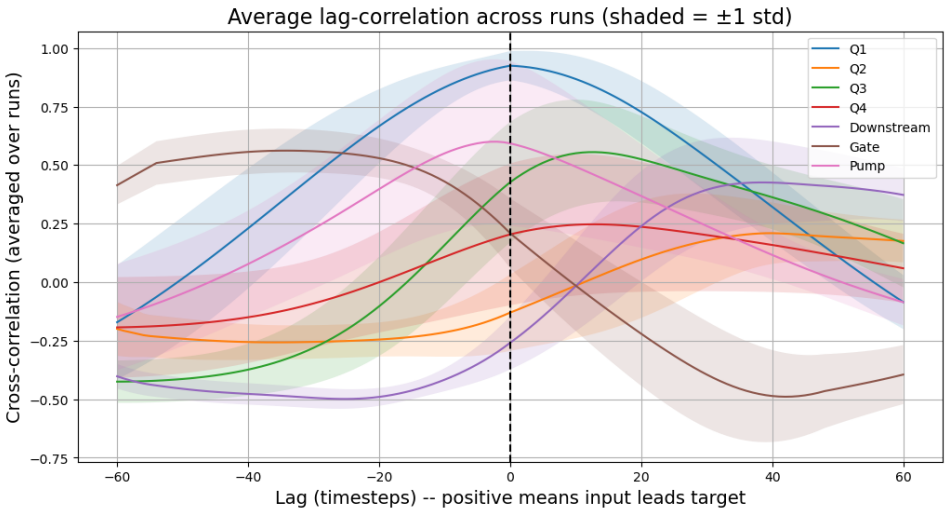


Figure D.1: Lagged-Correlation-Analysis

The results of the analysis in Figure D.1 shows that Q1 is clearly the strongest predictor of lake water levels, with a maximum correlation near 1 and no noticeable lag, likely because the timing of inflows from faster tributaries closely matches the lake response. Q3 also provides useful forecasting information, with a moderate correlation of 0.55 leading the target by about 13 time steps (≈6.5 hours), reflecting its smaller watershed and quicker contribution. In contrast, Q2 and Q4 show weaker correlations, peaking at 38 and 13 time steps respectively, since they contribute smaller volumes of water and their flood waves reach the lake at different times. The gate and downstream water level show a strong but inverted correlation with the lake, peaking around 40 time steps (≈20 hours). The gate tends to be open when the lake level is low and closed during storm surges, while the downstream level rises earlier due to the storm surge. Finally, pumped capacity, which is controlled by the lake water level in HEC-RAS, shows a strong correlation at a small negative lag because it essentially follows the target.

D.0.2. Cross Correlation Analysis (Cov Matrix)

To evaluate dependencies and redundancies among all variables, correlation and covariance matrices were computed over the concatenated dataset. For each feature f, the time series were flattened over all runs and timesteps into a single vector x_f.

The covariance matrix was computed as

$$\mathbf{C} = \mathrm{cov}(X) \in \mathbb{R}^{N_f \times N_f}, \tag{D.2}$$

and normalized to obtain the correlation matrix

$$\mathbf{R}_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}\, C_{jj}}}. \tag{D.3}$$

The correlation matrix is visible in D.2.

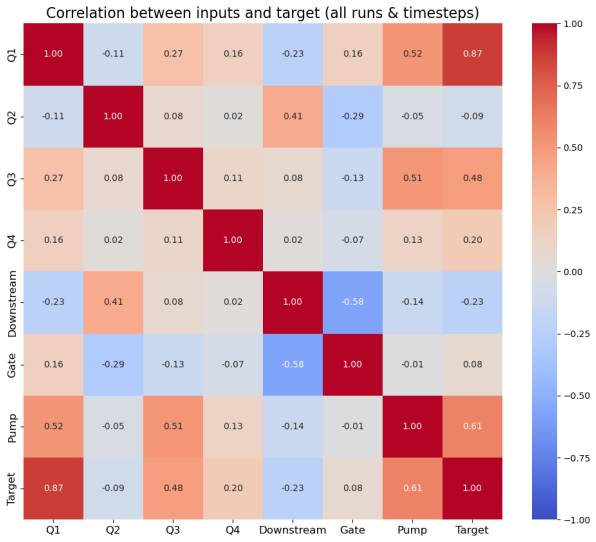


Figure D.2: Cross-correlation matrix across model simulations

The correlation matrix (Figure D.2 confirms that the target water level is strongly related to Q1 and the pumped capacity, which is expected since the pump operation is determined by the target. Q3 also shows a notable correlation with the target and the pump. In contrast, Q2 and Q4 exhibit very weak correlations with the target and other features, with the exception of a small apparent correlation between Q2 and the downstream water level and gate, which is likely coincidental. Q1 and Q3 are moderately correlated, reflecting the similar contributions of their watersheds.

Based on these observations, a few improvements to the dataset were proposed: removing Q2 and Q4, as they contain limited information for predicting the target; shifting Q3 by 13 time steps to better align with the target response; and applying a time shift of approximately 38 steps for the gate opening and downstream water level. Implementing these adjustments would create a more efficient and informative training dataset. However, these changes were not applied because it showed a small decrease in performance and did not result in faster convergence. Hence, it was chosen to keep all input features.

For the development of Model C, the analysis was used to obtain an initial estimate of the time series lag parameter, n_{lag}. This estimate was compared to the observed lags of 13 and 38 time steps, and ultimately a value of 20 was selected, as it yielded the best results.

E

Sensitivity Analysis of Model C

E.1. Sensitivity to Design Variables

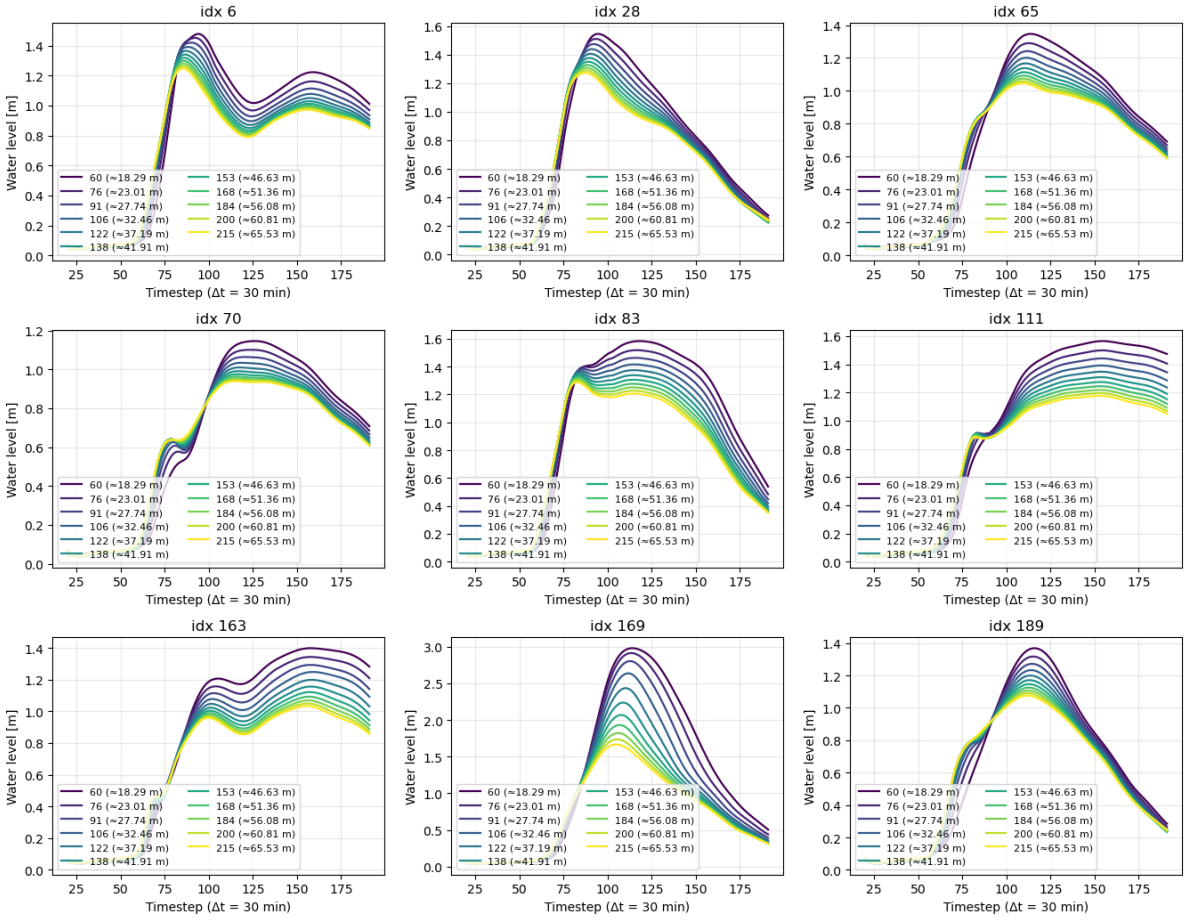


Figure E.1: Effect of gate width on Model C performance.

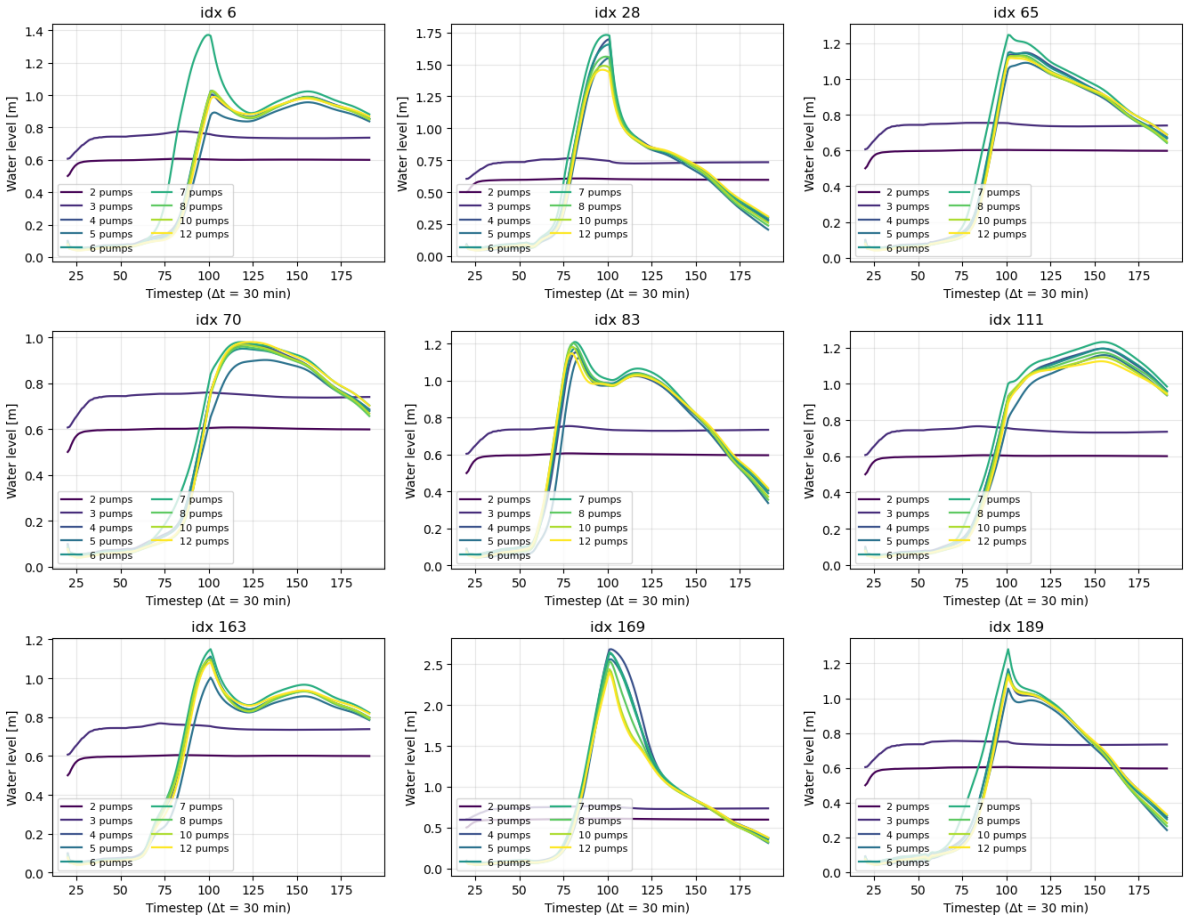


Figure E.2: Impact of the number of pumps on Model C results.

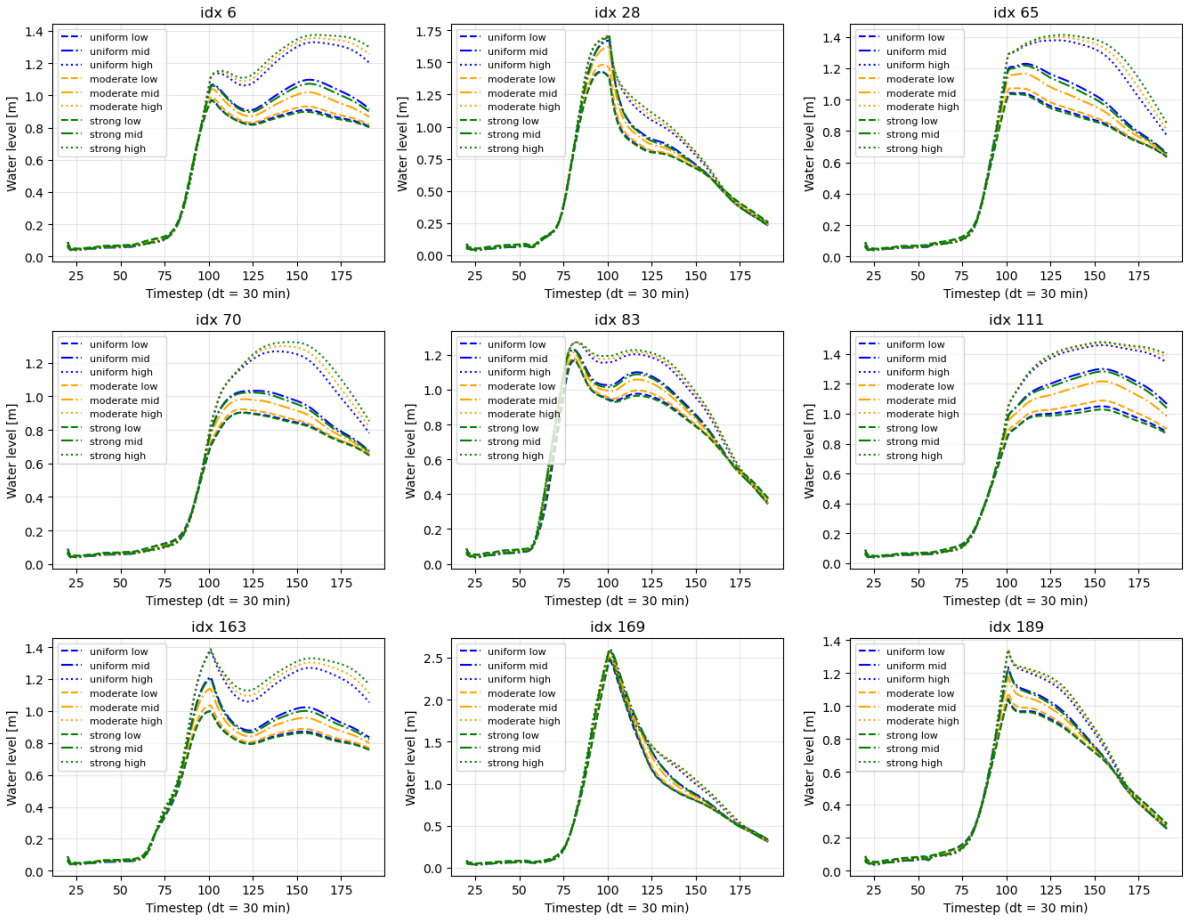


Figure E.3: Sensitivity of Model C to the pump activation stagger.

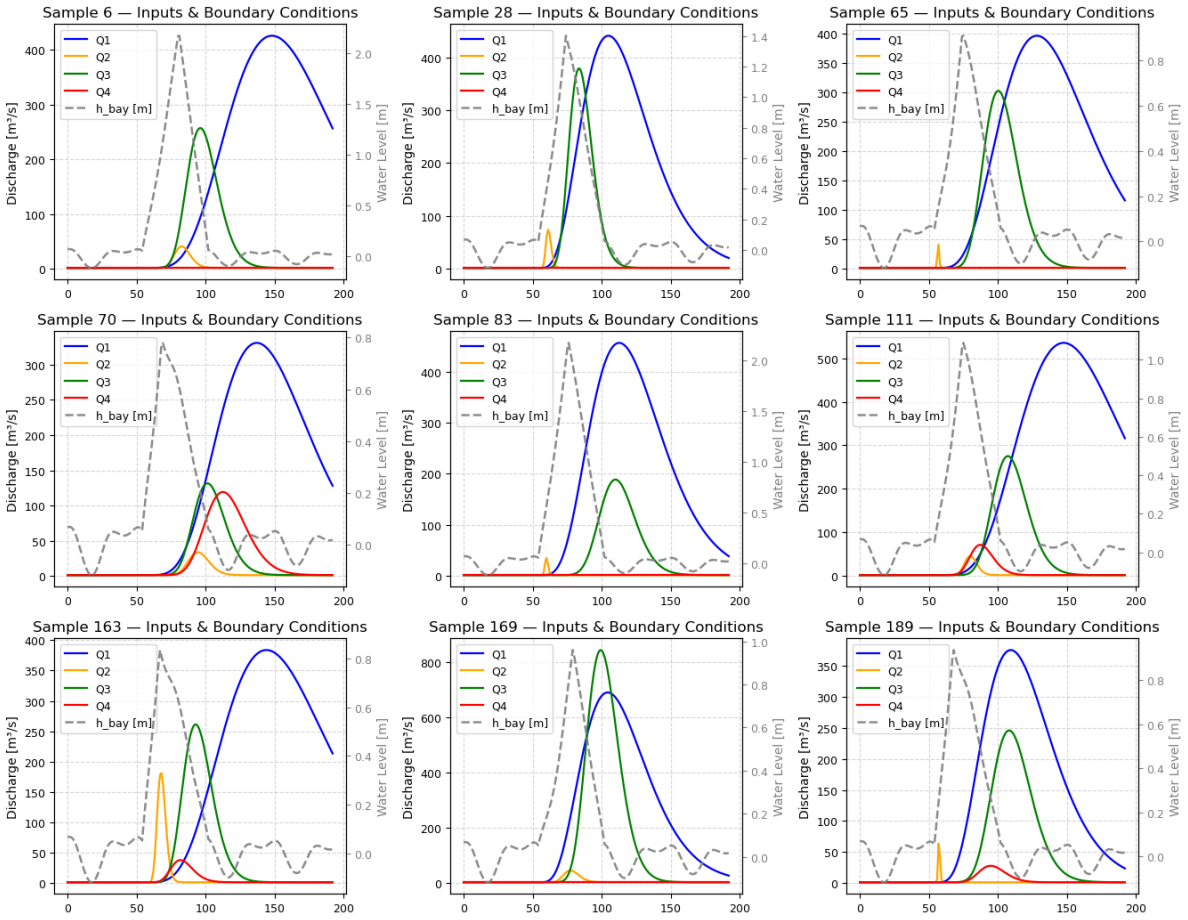


Figure E.4: Overview of the input configurations corresponding to the sensitivity analysis.

E.2. Sensitivity HMS Discharges

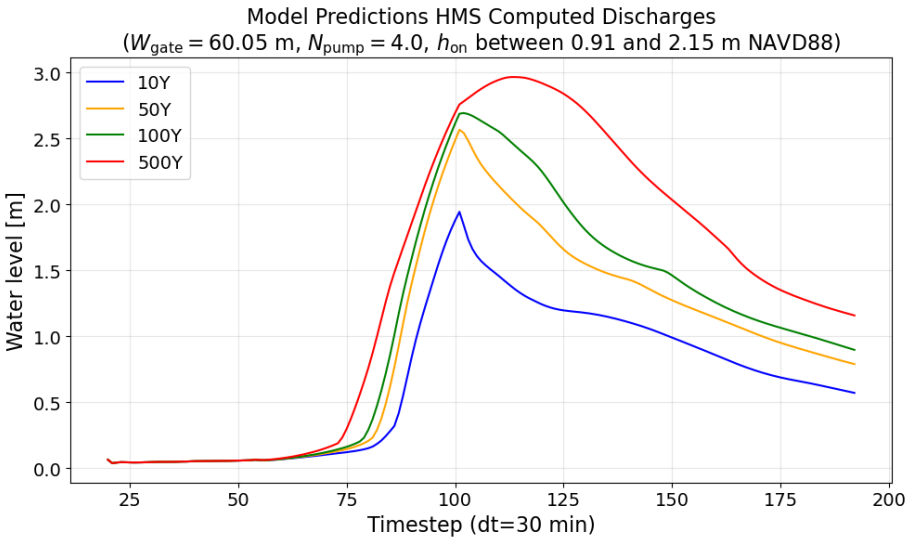


Figure E.5: Model C, HMS Inflows

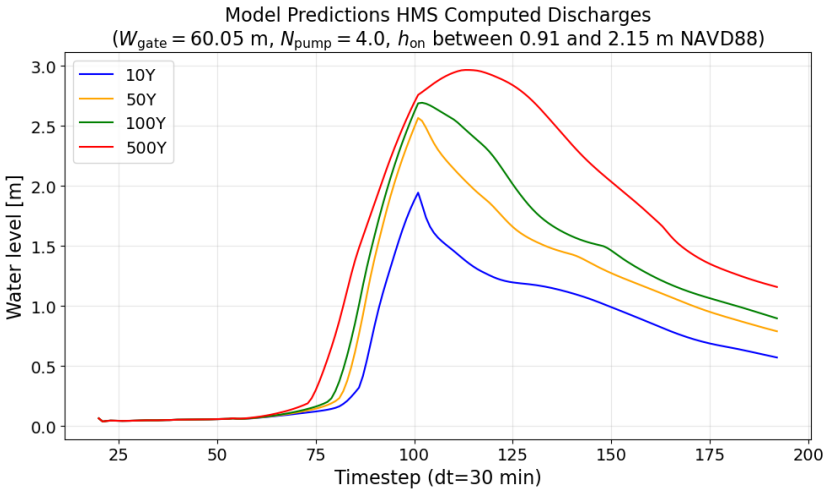


Figure E.6: Model C, HMS Inflows

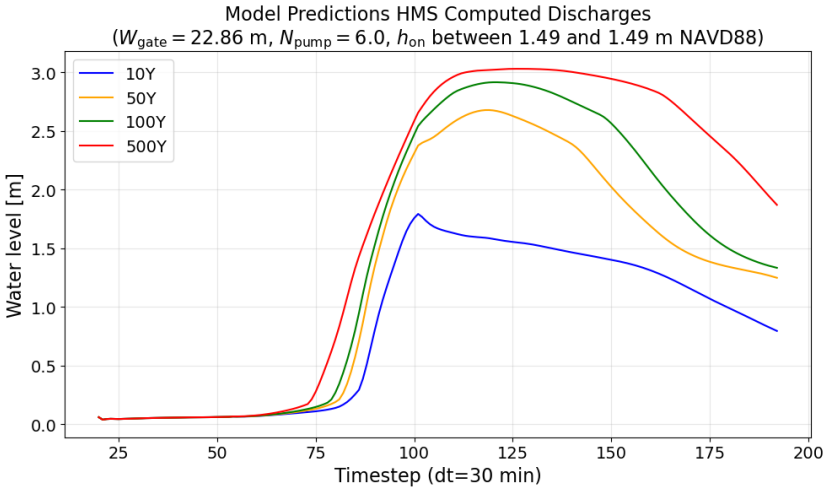


Figure E.7: Model C, HMS Inflows

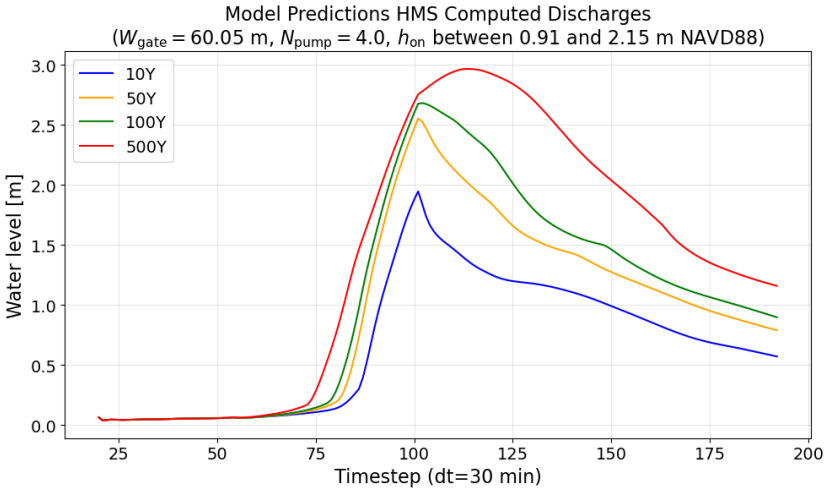


Figure E.8: Model C, HMS Inflows

F

Results Model D

F.1. Quantitative Metrics

Table F.1 presents the performance metrics of all models as described in 8.3, but adds the values of later developed Model D.

Table F.1: Performance metrics for the LSTM models.

Metric	Model B	Model D
RMSE (m)	0.0756	0.0734
MAE (m)	0.0484	0.0476
NSE (-)	0.9979	0.9980

As described, does the model, based on these metrics, not show increased performance compared to Model B. Although, the metrics indicate better fit compared to Model C.

F.2. Visual Assessment

A scatter plot where this new Model D is compared with Model B is visualized below:

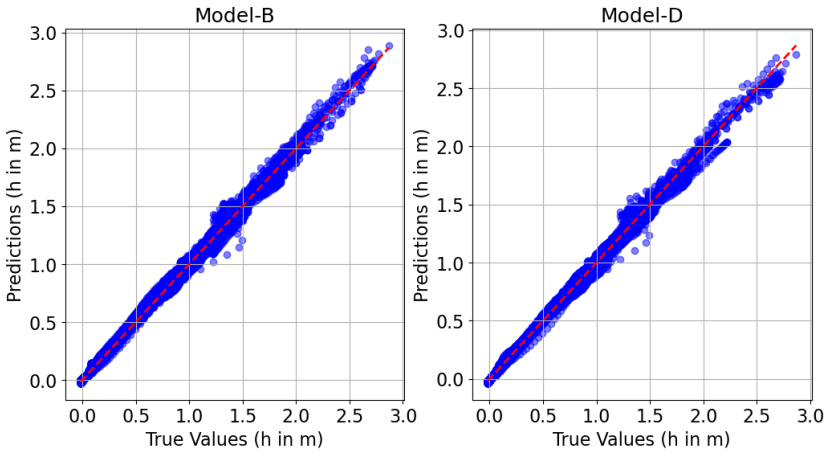


Figure F.1: Scatter plots of predicted versus observed water levels for Models B and D.

From this, the slightly improved performance of Model D compared to B is also visible, aligning with the values found in Table F.1