PREDICTING DISCHARGE VARIATIONS IN THE OTTAWA RIVER INTEGRATING HEC-RAS WITH ARTIFICIAL INTELLIGENCE IN THE CONTEXT OF CLIMATE CHANGE

MOHAMMAD UZAIR ANWAR QURESHI

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
Master of Applied Science degree in Civil Engineering

Department of Civil Engineering
Faculty of Engineering
University of Ottawa

Abstract

The importance of accurate flood prediction and effective flood management has become increasingly critical in the face of climate change, particularly in regions such as the Ottawa River, which has experienced devastating flood events (e.g., the 2017 and 2019 floods) that exceeded 100-year flood thresholds. These events underscore the urgent need for more advanced methods to predict discharge variations and improve disaster preparedness. This thesis develops an innovative framework by combining Hydrologic Engineering Center - River Analysis System (HEC-RAS) modeling with the New-Generation Group Method of Data Handling (New-Gen GMDH), a machine learning model, to predict discharge variations under multiple climate change scenarios. The study integrates 1D and 2D HEC-RAS simulations to generate a wide range of potential future flood scenarios, which are validated against historical data and satellite imagery. The generated datasets are fed into the New-Gen GMDH model, incorporating meteorological variables such as temperature and precipitation. This model improves prediction accuracy by identifying complex, nonlinear relationships between meteorological data and discharge outputs. The research provides a comprehensive, data-driven methodology for predicting flood risk and discharge variations, bridging the gap between traditional hydraulic models and advanced machine learning techniques. The hybrid framework provides policymakers and flood management authorities with practical tools, generating user-friendly equations that enable real-time decision-making in flood risk management. Through a comparative analysis of three modeling scenarios, the study demonstrates that Scenario 2 (SC2), which includes lagged variables, outperforms the other scenarios in predictive performance, with higher R^2 values (0.88 for training, 0.78 for testing) and Nash-Sutcliffe Efficiency (NSE) (0.75 for training, 0.57 for testing).

The findings underscore the need to improve flood forecasting techniques and demonstrate the value of integrating traditional hydraulic modeling with advanced machine learning to provide more reliable and scalable flood risk assessments. This integrated approach applies to other river systems facing similar climate-related challenges, contributing to climate change adaptation and enhancing resilience building.

Key Words: HEC-RAS, Artificial Intelligence (AI), Machine Learning (ML), New-Gen Group Method of Data Handling (GMDH), Ottawa River

Acknowledgements

I would like to first express my deepest gratitude to my thesis supervisor, Dr. Hossein Bonakdari at the University of Ottawa, for his invaluable guidance and unwavering support throughout my academic journey. I am also sincerely thankful to Isa Ebtehaj from the University of Laval for his patience and assistance with the Machine Learning component of my thesis. My appreciation extends to Mr. Juraj Cunderlik and the entire team at the Mississippi Valley Conservation Authority for their support in providing the geometric data essential for developing the 2D numerical model. Lastly, I am profoundly grateful to my wife and daughter—without their constant encouragement, support, and patience, this work would not have been possible.

Table of Contents

List of Figures

List of Tables

Legend & Abbreviations

Machine Learning (ML)

HEC-RAS (Hydraulic Engineering Center's River Analysis System)

Group Method of Data Handling (GMDH)

One-dimensional (1D)

Two-Dimensional (2D)

Geographic Information Systems (GIS)

AI (Artificial Intelligence)

Light Detection and Ranging (LiDAR)

Digital Elevation Model (DEM)

High-Resolution Digital Elevation Model (HRDEM)

Mississippi Valley Conservation Authority (MVCA)

Water Surface Elevation Level (WSEL)

Flood Frequency Analysis (FFA)

Kolmogorov-Smirnov (KS)

Anderson-Darling (AD)

Chi-Squared (CS)

Cumulative Distribution Function (CDF)

Coefficient of Determination (R^2)

Variance Accounted For (VAF)

Root Mean Square Error (RMSE)

Mean Absolute Error (MAE)

Scatter Index (SI)

Mean Absolute Relative Error (MARE)

Root Mean Square Relative Error (RMSRE)

Akaike Information Criterion (AIC)

Nash-Sutcliffe Efficiency (NSE)

Log-Pearson Type III (LP3)

Chapter 1

Introduction

1.1 - Background:

Floods are among the most common and costly natural disasters worldwide. From 1900 to 2019, recorded floods had a devastating global impact, leading to over 7 million deaths and impacting more than 4 billion people worldwide. These natural disasters have inflicted direct economic losses of approximately $1.2 trillion (EMDAT, 2024), underscoring their profound effect on human life, the economy, and infrastructure. As one of the globe's most catastrophic natural phenomena, flooding is recognized as a preeminent natural hazard. It constitutes 44% of all recorded disaster occurrences between 2000 and 2019 (CRED and UNDRR, 2020). Floods exert considerable ecological repercussions, transforming landscapes and driving species towards extinction. Consequently, these events have an adverse impact on socio-economic conditions, degrade public health, increase unemployment, harm ecosystems, and result in significant loss of human life. (Manzoor et al., 2022). The impacts of fluvial floods are projected to intensify in the future, primarily due to the expansion of population and economic activities within areas susceptible to flooding (Jongman et al., 2012). The acceleration of the global hydrological cycle due to climate change is poised to further amplify future flood risks (Alfieri et al., 2017), through heightened rainfall intensity and elevated river levels, which could markedly expand the number of people at risk of flooding unless preventive actions are taken (Noori and Bonakdari, 2023).

Flood damages are categorized into direct and indirect impacts, with delineation based on the interplay between climatic variations and the socio-economic conditions of the affected region. Direct damage refers to the immediate physical impact of flooding, such as structural damage to buildings, infrastructure, and agricultural lands. Indirect damage encompasses the broader

economic and social consequences, including business interruptions, loss of income, health impacts, and disruption of services (Koks and Thissen, 2016), which are influenced by both the severity of the flood event and the resilience or vulnerability of the socio-economic fabric of the area. Indirect damage may persist long after the flood event has subsided, capturing the complete temporal scope of the impact, and can both negatively and positively influence areas beyond the initial site of the disaster (Carrera et al., 2015). Due to these dynamics, indirect losses can be substantial, potentially exceeding direct damages in magnitude (Koks et al., 2015).

Canada has a well-documented history of flood events, which are the most common and costliest natural disasters in the nation (Sandink et al., 2016). Canada experiences floods throughout the year; however, snowmelt flooding has proven to be more significant in the spring and early summer seasons. Flash flooding is also a concern due to the generation of intense rainfall over the summer season.

The Ottawa River's history reflects its complex hydrological dynamics and sensitivity to extreme hydro-meteorological events, resulting in significant flood occurrences. Since 1948, the regions of Ontario and Quebec have experienced a rise in winter precipitation, accompanied by warmer and more humid springs (Bush et al., 2019). As recent as 2017, destructive springtime floods in the provinces of Ontario and Quebec caused insured damages exceeding 220 million CAD (Insurance Bureau of Canada, 2017). This flood affected 286 municipalities in Quebec, causing over 5,300 houses to submerge underwater and forcing over 4,000 people into emergency evacuations (Quebec Government, 2017). Figure (1.1) shows the impact of flooding in Gatineau Point during the 2017 flood.

Figure 1.1 – Gatineau Point during the 2017 flood in the Ottawa River (Ottawa Citizen, 2017).

In the spring of 2019, the Ottawa river saw another record-setting flood. The river crested on the 1st of May, but the state of emergency in Ottawa lasted until the second week of June (Kirchmeier-Young et al., 2021). The flood impacted thousands causing widespread evacuations, an extended state of emergency, and about 200 million CAD in insured losses (Insurance Bureau of Canada, 2019). These flood events occurred despite the Ottawa River's comprehensive river discharge management system, which includes 50 major dams and 13 principal reservoirs (Ottawa River Keeper, 2019; Ottawa River Regulation Planning Board, 2019). Figure (1.2) is an aerial photograph of the City of Ottawa during the 2019 flood. Both these events surpassed the anticipated threshold of a 100-year flood event with discharge reaching almost 6000 m^3/s. Such severe flooding episodes are linked to a range of hydrological elements, including heavy downpours, rapid snowmelt, ice blockages, and substantial spring thaws, all of which contribute

to an increased likelihood of overflow throughout the river's vast basin (Kirchmeier-Young et al., 2021). Due to its extensive drainage basin and the hydrological interconnectivity of its tributaries, the Ottawa River system exhibits heightened susceptibility to compound flooding dynamics (Bhuiyan et al., 2022). Particularly under climate change scenarios, projections suggest a higher frequency of extreme precipitation events and snowmelt-driven flooding, exacerbated by rising global temperatures. (Kirchmeier-Young et al., 2021, Letessier et al., 2023, Szeto et al., 2015, Mohanty et al., 2021). These shifts in climate could lead to higher frequencies of flood-inducing rainfall events, posing heightened risks to communities and infrastructure along the river. Therefore, it's crucial to implement adaptive approaches and strong flood risk management systems to lessen the impact of these potential threats in the future (Cai et al., 2024).

Figure 1.2 – City of Ottawa, May 2019 (Aerial Photography, 2019).

1.2 - Purpose of Study:

Traditional river modeling has largely been based on hydraulic simulations and machine learning (ML) techniques. Over the years, these methods have been continuously refined, leading to more accurate and reliable predictions. Widely used numerical modeling tools such as HEC-RAS (Hydraulic Engineering Center's River Analysis System) (Mahapatra et al., 2022, Marimin et al., 2018, Yang et al., 2006), Delft3D (Parsapour-Moghaddam et al., 2018, Parsapour-Moghaddam et al., 2023) and SWAT (Soil and Water Assessment Tool) (Yang et al., 2016) play a crucial role in simulating river flow, predicting discharge variations, and estimating water surface elevations. These models incorporate key hydrological and hydraulic factors, including riverbed topography, channel roughness, and stage-discharge relationships. By integrating these variables, numerical simulations can provide valuable insights into flood dynamics, improve risk assessments, and enhance flood forecasting capabilities.

In recent years, there has been a growing emphasis on integrating AI (Artificial Intelligence)-driven tools into flood management and mitigation strategies. Researchers have explored both conventional ML algorithms and advanced deep learning networks to enhance flood prediction and response efforts. Unlike traditional numerical modeling, which relies heavily on detailed hydrological parameters, AI and ML techniques can learn directly from data, minimizing the need for extensive input variables (Roh et al., 2019). These models can develop predictive algorithms without relying on predefined physical laws or geometrical data, making them highly adaptable to evolving hydrological conditions. Classical AI models, such as linear regression, decision trees, logistic regression, support vector machines (SVM), and clustering techniques, offer advantages in terms of interpretability, lower computational costs, and effectiveness in capturing complex flood dynamics (Liu et al., 2024). In contrast, deep learning architectures—

including long short-term memory (LSTM) networks, artificial neural networks (ANN), and convolutional neural networks (CNN)—have demonstrated superior performance in handling vast datasets and extracting intricate patterns, further advancing the accuracy of flood forecasting models (Liu et al., 2022).

While prior studies have established that various ML algorithms exhibit substantial predictive capabilities in flood forecasting, their accuracy is predominantly reliant on the quality and comprehensiveness of historical datasets used during the training phase (Ghorpade et al., 2021). A key challenge for these models is their difficulty in accurately capturing abrupt shifts in extreme flood events, particularly those influenced by evolving climatic patterns. To address this limitation, this research introduces a hybrid methodology that integrates numerical simulations with ML-based predictive modeling.

Observed hydrometric data have been employed as reference benchmarks for HEC-RAS simulations, ensuring high-fidelity hydraulic modeling. To account for meteorological variability, satellite-derived observations have been incorporated due to their extensive spatial resolution and consistent temporal datasets. The HEC-RAS framework will produce a detailed dataset comprising river discharge rates, water surface elevations, channel widths, and floodplain morphology across four flood periods: 50-year, 100-year, 1,000-year, and 10,000-year flood events. This dataset will serve as input for ML models, facilitating the derivation of user-oriented explicit equations utilizing the New-Gen Group Method of Data Handling (GMDH) ML algorithm tailored for policymakers and flood management authorities. These equations will enable direct estimations of critical hydrodynamic variables for the Ottawa River, including floodplain expansion, and discharge variations, thereby enhancing predictive robustness and operational usability. By merging numerical modeling with ML methodologies, this work aims to refine flood forecasting

accuracy while equipping stakeholders with data-driven tools for proactive flood mitigation. Additionally, this integrated framework accommodates the evolving nature of flood regimes under climate change, establishing a strategic foundation for forecasting and minimizing the repercussions of future catastrophic flood events on the Ottawa River and its surrounding communities.

To evaluate the influence of various meteorological and hydrological drivers on river discharge, the study developed three modeling scenarios (SC1, SC2, and SC3) using the New-Gen GMDH framework. Scenario 1 (SC1) included four core meteorological variables: Maximum Temperature, Minimum Temperature, Mean Temperature, and Total Precipitation. These variables were selected due to their strong physical relationship with snowmelt rates, rainfall-runoff generation, and seasonal temperature patterns that drive discharge dynamics in cold-region basins like the Ottawa River. Scenario 2 (SC2) extended SC1 by incorporating lagged versions of these variables — lag1, lag2, and lag12 — based on autocorrelation analysis of the dataset. These lag inputs account for the system's memory, capturing delayed hydrologic responses caused by prior meteorological conditions (e.g., consecutive warm days that accelerate melt or rain-on-snow events). Scenario 3 (SC3) tested the inclusion of discharge-based statistics (average minimum, average mean, and average maximum discharge) alongside meteorological variables. Although SC3 added hydrological realism, SC2 outperformed both scenarios in predictive accuracy. The lagged meteorological structure in SC2 allowed the model to track evolving climate-driven discharge patterns better, making it most effective for forecasting flood risks under changing conditions.

This thesis represents the culmination of almost two years of dedicated research, involving the application of HEC-RAS numerical modeling integrated with machine learning techniques. The author has presented and published related work at various conferences and in an academic journal, as listed below:

- Qureshi, M. U. A.; Dussel, A.; Ebtehaj, I.; Bonakdari, H. (2023). Exploring Flood Zone Width Prediction Using Multi-Objective Evolutionary Polynomial Regression: a Rideau Canal Analysis. Proceedings of the 4th IAHR International Young Professionals Congress. https://www.iahr.org/library/infor?pid=29801
- Qureshi, M. U. A.; Seidou O.; Motiyani. H. H.; Bonakdari, H. (2024). Developing a Comprehensive Water Management Plan for the Niger Basin. Proceedings of the MDPI 8th Electronic Conference on Water Sciences' Water Resources Management, Flood & Risk Mitigation Session. https://sciforum.net/paper/view/19146
- Qureshi, M. U. A.; Shaheen, M.; Seidou, O; Cunderlik, J; Bonakdari, H. (2024). Assessing HEC-RAS and AI Integration for Hydraulic Behavior Analysis of the Ottawa River. Proceedings of the 5th IAHR International Young Professionals Congress. https://www.iahr.org/library/infor?pid=31062
- Qureshi, M. U. A.; Amiri, A.; Ebtehaj, I.; Guimere, S. J.; Cunderlik, J.; Bonakdari, H. (2025). Coupling HEC-RAS and AI for River Morphodynamics Assessment Under Changing Flow Regimes: Enhancing Disaster Preparedness for the Ottawa River. Hydrology, 12, 2. https://doi.org/10.3390/hydrology12020025
- 2nd Place Prize (Sustainable and Resilient Infrastructure Category) in the Engineering Research Day Competition at the University of Ottawa with Poster titled "Enhancing Ottawa River Flood and Morphodynamics Assessment Incorporating Socio-Economic and Sustainability Aspects through HEC-RAS and AI Integration"

Chapter 2

Literature Review

The first step in the Literature Review process involved developing a comprehensive search strategy to identify the most relevant, credible, and high-quality data sources. Key databases included PubMed, Scopus, Web of Science, Google Scholar, IEEE Xplore, and the Engineering Village (Compendex), among others. A critical aspect of the strategy was the strategic use of high-impact keywords—such as "AI, machine learning, neural networks," "HEC-RAS, hydrodynamic modeling, flood modeling," and "Ottawa River, climate change, flooding"—combined with Boolean operators to refine search results and ensure relevance. To maintain academic rigor, only peer-reviewed publications were considered, with a strict focus on AI, ML, hydrodynamic modelling, climate change and flood modelling. Beyond primary databases, the review process extended to reference lists and citation tracking services, such as Google Scholar, to capture additional relevant literature.

The Literature Review section includes recent research articles, conference papers, and studies encompassing this study's three key topics. To ensure relevance and up-to-date information, only studies from the past few years were selected. The discussion follows a chronological order, providing a structured overview of advancements in the field. The first subsection focuses on Numerical Modeling, highlighting studies on the application of HEC-RAS, HEC-HMS, and general numerical modeling approaches. It also explores research that integrates numerical modeling with other simulation techniques. The second subsection reviews the application of AI in hydrology, covering studies that provide foundational knowledge on AI techniques in this domain. This serves as a basis for understanding AI-driven approaches before delving into the specific GMDH model. The final subsection examines the use of GMDH in

hydrological modeling, presenting relevant studies that demonstrate its implementation, effectiveness, and potential advantages in this field.

2.2 - Numerical Modeling:

HEC-RAS, developed by the Hydrologic Engineering Center in the early 1990s, was introduced as a replacement for HEC-2, providing enhanced capabilities for water surface profile computations. The initial versions were limited to one-dimensional (1D) steady-state analysis, but with the release of Version 3.0 in 2001, the software was upgraded to include 1D unsteady flow modeling (Dewberry, 2016). Over time, its functionality has expanded significantly, culminating in the release of Version 5.0 in 2016, which introduced two-dimensional (2D) unsteady flow simulations.

Numerical models like the HEC-RAS play a crucial role in simulating river hydraulics, providing critical insights into flood extents, depths, and velocities under various scenarios and predicting flood behavior in rivers and streams (Ehsan et al., 2024). Numerical models leverage advanced mathematical and physical equations to simulate water movement within river channels and floodplains, delivering comprehensive analyses of hydraulic dynamics and flood risks. The software offers a comprehensive suite of functionalities, including 1D steady flow analysis, 1D and 2D unsteady flow simulations, sediment transport modeling, and water quality assessments. These capabilities make it a highly versatile tool for water resource management and hydraulic modeling (Zainal et al., 2024). When combined with geospatial techniques such as remote sensing and Geographic Information Systems (GIS), HEC-RAS facilitates the creation of detailed flood hazard maps and inundation extent predictions, which are essential for effective flood monitoring, risk assessment, and management strategies (Singh et al., 2021).

Villazon et al., (2009) developed an unsteady flow model using HEC-RAS to simulate flood events in a 90 km stretch of the Piraí River in Santa Cruz, Bolivia, focusing on improving flood prediction and management. The model was calibrated using historical flood events, including the severe flooding of 1983 and 2006, with peak discharges reaching approximately 4500 m^3/s and 3200 m^3/s, respectively. The study area of the Piraí River included three main gauging stations that were used to calibrate the initial boundary conditions of the model run. These gauging stations were located at Angostura and Espejos. The model performed well for low flow scenarios using a variable roughness coefficient scenario. Their findings showed that flow velocities could exceed 4.50 m/s during extreme flood conditions, increasing erosion risks along riverbanks and infrastructure. Sensitivity analyses indicated that flood extent predictions varied significantly based on roughness coefficients, with Manning's n values ranging from 0.025 to 0.045 affecting water surface elevations by up to 0.80 meters. These results underscored the necessity for improved floodplain zoning and reinforced infrastructure to mitigate flood impacts in Santa Cruz. Additionally, the study emphasized the need for continuous hydrometric monitoring, as the lack of real-time flow data limited the accuracy of flood forecasting models (Villazon et al., 2009).

Agrawal et al., (2016) conducted an unsteady flow analysis of the Lower Dudhana River in India located between 19°60 'N to 19°45 'N and 76.51° 'E 76.51° 'E using HEC-RAS to assess flood behavior and water surface elevation variations. The study's main purpose was to calculate head difference on the downstream end of the Lower Dudhana River Dam. The river reach included cross-sections 50 m apart for a stretch of 1100 m. The model was run with a discharge of 300 m^3/s to 1300 m^3/s. Their results indicated flood depths could reach up to 3.80 m under peak discharge, with velocities exceeding 2.50 m/s in certain sections, increasing the risk of bank

erosion and infrastructure damage. Using the simulations, they concluded that the cross-section at the dam needed to be properly embanked and raised around 412 m in order to protect the surrounding village from extreme inundation with emphasis on the right overbanks of the dam. The study emphasized the effectiveness of unsteady flow modeling in predicting flood risks and improving floodplain management, particularly for the protecting villages and isolated settlements (Agrawal et al., 2016).

Additionally, Stoleriu et al. (2019) utilized a combination of Light Detection and Ranging (LiDAR) data and HEC-RAS to improve the accuracy of flood risk maps in north-eastern Romania. Three tests (T1, T2, and T3) with varying resolutions of Digital Elevation Models were run to compare the difference in flooding extents. The study area was located in the Başeu River Basin in northeastern Romania, which included 70 settlements with populations ranging from 41 to almost 7,000 residents. The results showed that very large differences were observed regarding the built-up areas potential to flood risks. The SRTM-based T1 result underestimated the flooding extents, while the LiDAR-based T2 and T3 performed very well. This was attributed to the coarse nature of the Digital Elevation Model (DEM) utilized in the T1 test. The study evaluated flood risk using three models: T1 [1%] (SRTM-based), T2 [1%] (LiDAR-based), and T3 [3%; 1%; 0.1%] (LiDAR-based), demonstrating that LiDAR data significantly improve flood hazard assessment. For this reason, the T3 scenario incorporating high-resolution LiDAR and HEC-RAS simulations were utilized to study the flood hazards for 100-year, 1000-year and 33.3-year floods. The analysis showed that floods with 3% (33.3-year), 1% (100-year), and 0.1% (1,000-year) recurrence intervals would affect 112 ha, 140 ha, and ~170 ha, respectively, with arable land being the most impacted. The 1% flood extent indicates that over 500 buildings and more than 4000 people are vulnerable to flooding. Overall, HEC-RAS flow models combined with LiDAR DEMs provide

superior accuracy for flood hazard mapping, making SRTM data inadequate for small-scale assessments (Stoleriu et al., 2019).

Furthermore, Faudzi et al., (2019) implemented the 2D modelling tool of HEC-RAS to develop a highly detailed flood hazard map of the Upper Sungai Langat River in Malaysia. The Sungai Langat basin included major cities such as Kajang, Bangi, Seri Kambangan, and Dengil, which experienced flooding due to their location on the downstream end of the river. The flooding was attributed to rainfall. Using ArcGIS, they delineated a DEM layer for the basin. To simulate discharge, rainfall events were simulated using the HEC-HMS software, allowing meteorological data to be incorporated into the study, with a focus on simulating a 100-year flood event. The model was calibrated and validated using observed discharge values obtained from the Department of Irrigation and Drainage Malaysia. Their calibration result yielded an R^2 value of 0.85, and the validation yielded an R^2 of 0.81. Comparing precipitation values with corresponding discharge yielded maps of the region with varying susceptibility of flood hazard scenarios. The 100-year flood event proved to be the worst-case scenario in the upper Sungai Langat area with a peak discharge of 426 m^3/s and an inundation area of 93 km^2 as represented in Figure (2.1). The maximum inundation depth was 7.14 m, showcasing the dire situation of the inhabitants in the basin. The results were hoped to provide clarity to the residents of the region when comparing rainfall scenarios for better preparations to flooding events (Faudzi et al., 2019).

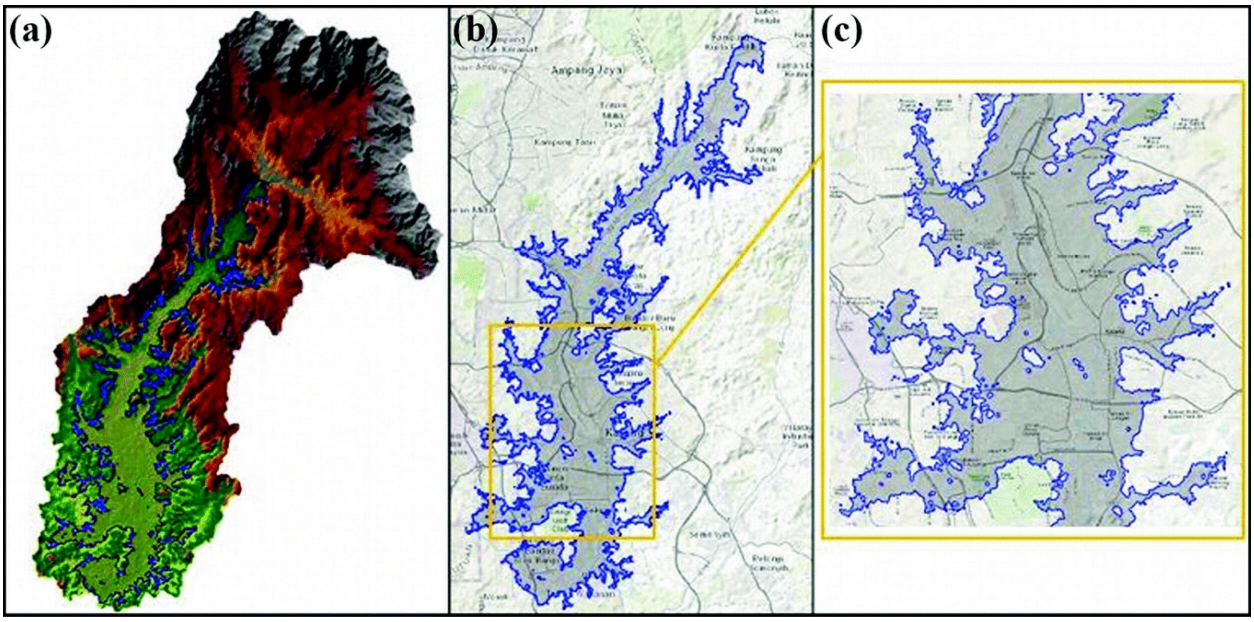


Figure 2.1 – Flood Hazard Map of Upper Sungai Langat on a (a) DEM; (b) Google Map; and (c) Kajang area based on 100-year Average Recurrence Interval (ARI) (Faudzi et al., 2019).

Filianoti et al., (2020) developed a novel performance matrix to quantitatively assess flood prediction accuracy across multiple hydrological and hydraulic models, integrating stakeholder-driven evaluation metrics and multi-parameter assessment criteria. Additionally, they conducted a comparative analysis of software usability, examining the computational efficiency and functional limitations of different modeling frameworks. The study systematically evaluated four conceptually based physically driven flood models applied to a mid-sized Mediterranean watershed in southern Italy, considering their capability to simulate watershed hydrodynamics under varying boundary conditions. Results identified HEC-HMS and MIKE 11 as the most computationally efficient and robust models, exhibiting optimized numerical stability, low computational overhead, and high predictive accuracy. These models demonstrated superior performance due to their streamlined data processing, reduced algorithmic complexity, and enhanced hydrological-hydraulic coupling, making them well-suited for flood forecasting applications in data-limited and operationally constrained environments (Filianoti et al., 2020).

Khan et al., (2020) used the steady flow modelling feature in HEC-RAS to simulate the Naray-Khwar canal in Peshawar, Pakistan. Geometric data for the study was collected by conducting total surveys in sixteen different locations along the canal. Due to the unavailability of gauge stations, flow data was simulated using daily precipitation data using the SCS curve number method and Win TR20. Log Pearson type III method was applied to the peak precipitation data to simulate discharge scenarios for a 40-year period. Scenarios for 2, 5, 10, 25, 50, and 100-year flood events calculated with the Log Pearson type III method yielded discharge values of 9.60 m^3/s, 43.20 m^3/s, 91.50 m^3/s, 140.50 m^3/s, 190.90 m^3/s, and 299.70 m^3/s. The study provided insights into the different cross-sections and their susceptibility to flooding. 8 out of the 10 cross-sections proved to be safe from flooding during a 100-year flood event during long linear stretches of the canal. However, for a specific location near the Nasir Bagh Bridge which was observed to be a junction into the Naray-Khwar Lower, overtopping was observed at all cross-sections during the 100-year flood scenario. Additionally, extreme flooding was only observed during the monsoon season in the region with factors such as poor waste management, illegal encroachments by the local population causing additional inundation due to the adverse effect on the canal. The study further aimed to develop a 2D HEC-RAS model to simulate floodplain dynamics along with recommendations for the construction of sewage treatment plants and retaining walls along the lower banks of the Naray-Khwar (Khan et al., 2020).

Similarly, Daham et al. (2020) conducted a hydraulic simulation of a reach in the Al-Gharraf River, Iraq, using HEC-RAS's 1D and 2D modeling capabilities to assess flood dynamics and river hydraulics. The study incorporated field-surveyed cross-sections and a high-resolution DEM to define the river's geometry, enabling a comparison between 1D and 2D modeling approaches. The river reach was over 58,000 m long. Simulations were performed for discharge

values ranging from 100 m^3/s to 350 m^3/s, representing both normal and extreme flow conditions. Multi-scenarios the Hai Regulator gate openings were applied to study the impacts of flood prevention measures in the channel with opening ranges of 60 cm to fully open. The results demonstrated that the accurate assumption of Manning's n value of 0.025 was suitable for the unsteady flow model run. Average velocities for the 1D model ranged between 0.36 m/s to 0.5 m/s with corresponding water surface elevations of 15.14 m and 17.84 m. The 2D model runs provided velocities of 0.25 m/s to 0.44 m/s with corresponding stage of 14.13 m and 18.82 m. The results of the 1D and 2D model runs were compared to measured values in the Al-Gharraf river, indicating that the 2D model run provided better results based on the Root Mean Square Error and determination coefficients (Daham et al., 2020).

Afzal et al., (2022) utilized a combination of HEC-RAS one-dimensional modelling with GIS based satellite imagery to develop a model that could benefit early flood warning systems in the Indus River Basin in Pakistan. The model was calibrated and validated for the 2010 and 2015 flood events in the Indus River Basin. The Landsat 5 Thematic Mapper (TM), Landsat 8 Operational Land Imager (OLI) and the Moderate Resolution Imaging Spectroradiometer (MODIS) MOD09A1, and MOD09GA were used in the analysis of the flood events. Delineation of the water bodies was done utilizing the Normalized Difference Water Index (NDWI) and the Modified NDW11 (MNDW12). The study was done to pinpoint an accurate value of Manning's n value for the river and floodplain to represent the flood event conditions in the basin as shown in Figure (2.2). The study's findings demonstrate a comparison of land use maps from 2010 and 2015, with accuracies of 64.06% and 61.64%, respectively. Accuracy of the model simulation "n" values was successful as the DSC, F1, and F2 values were 0.77, 0.64 and 0.97 for the calibration period while 0.71, 0.66, and 0.86 for the validation period, respectively. The study concluded that HEC-

RAS provides a powerful tool to aid early flood warning systems in areas with extreme hydrometeorological conditions and complex topography (Afzal et al., 2022).

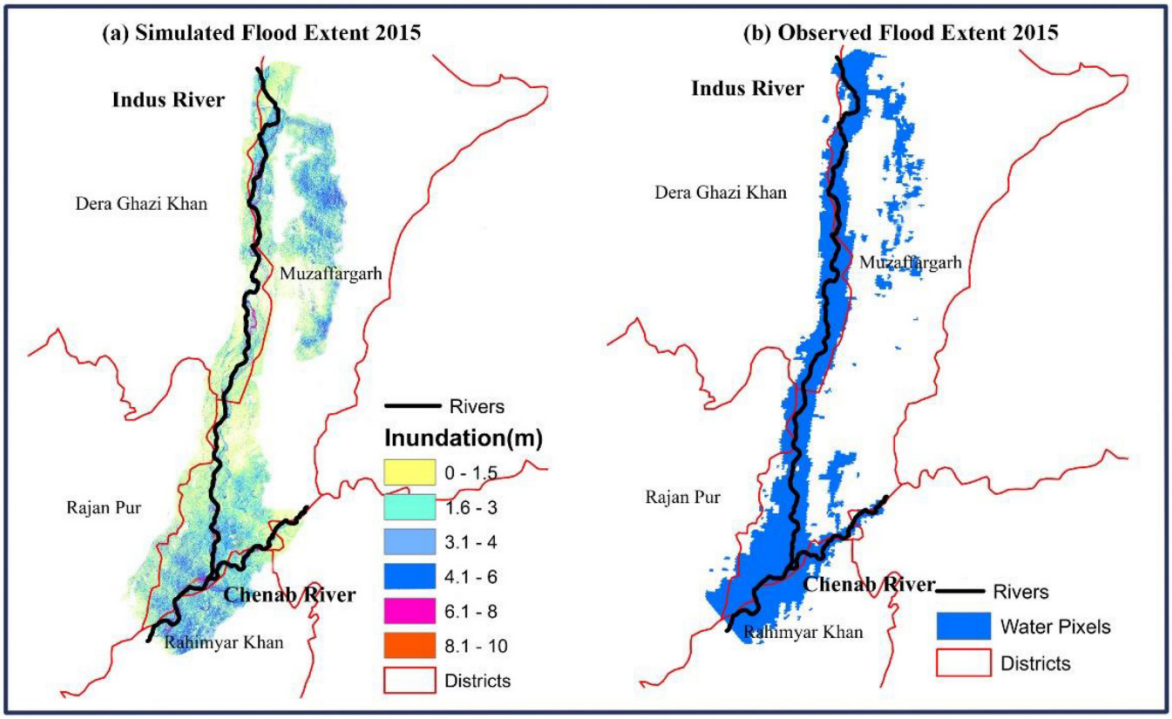


Figure 2.2 – (a) HEC–RAS-based simulated, (b) satellite-based flood area for the 2015 flood. (Afzal et al., 2022).

Aryal et al., (2022) utilized HEC-RAS to develop a model that compares its results with Federal Emergency Management Agency (FEMA)'s own flood inundation map in the Rock River Watershed in the United States of America. The study aimed to provide insight into the impact of future projected streamflow on flood inundation extent by comparing floodplain changes under historical and future simulated scenarios. The Atmosphere/Ocean General Circulation Model (AOGCM) developed by the Coupled Model Intercomparison Project Phase 6 (CMIP6) was used to illustrate behaviour of the flooding extents in future climate change scenarios. The study found that climate change could cause devastating effects in the region with both 100-year as shown in

Figure (2.3) and 500-year flood extents being increased by 20% compared to the current estimations (Aryal et al., 2022).

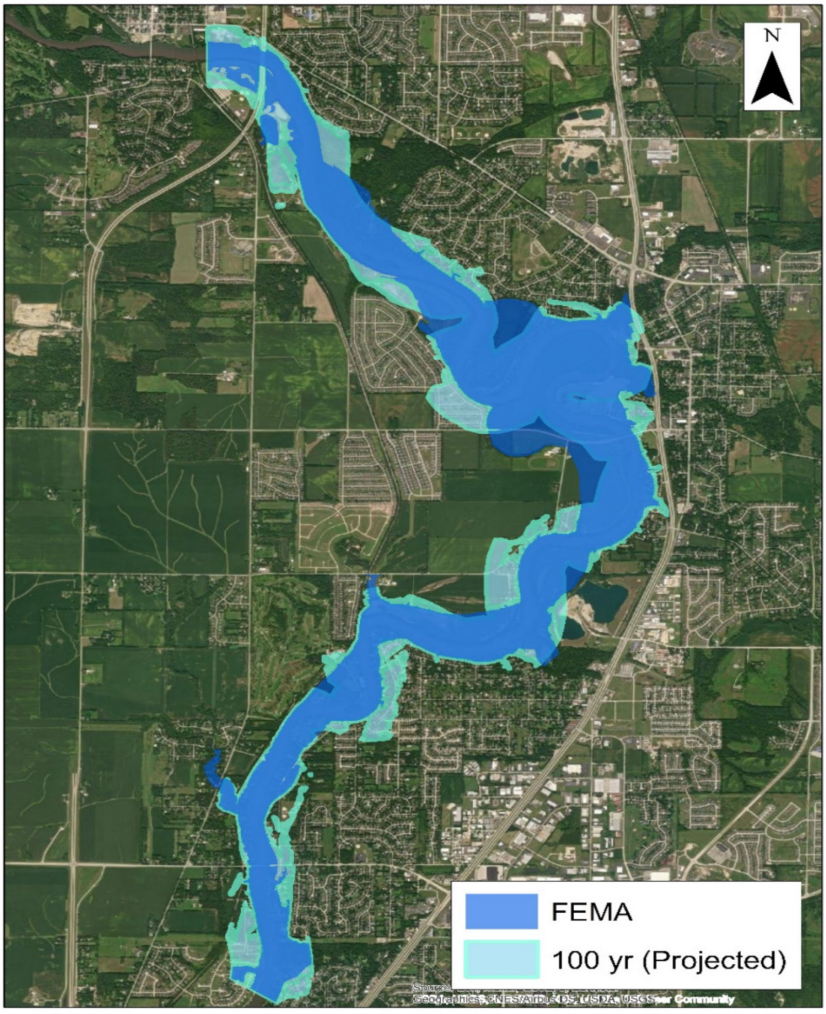


Figure 2.3 – Comparison of inundation map between FEMA 100-yr and future projected 100-year flow (Aryal et al., 2022).

Additionally, Bruno et al., (2022) applied HEC-RAS and HEC-HMS to study linked modeling in a small urban waterway in Campo Grande, Mato Grosso do Sul, Brazil, using detailed data to simulate flood events. The models were calibrated with historical data (2015–2018), incorporating flow sensors, water levels, rain gauges, land cover, land use, and topography. Flood scenarios were generated using based on synthetic rainfall with 5, 10, 50, and 100-year return

periods for a specific basin, resulting in highly accurate models. HEC-HMS calibration and validation yielded R^2 values of 0.93, RMSE of 1.29, and NSE of 0.92. The study highlighted that certain channel sections are naturally prone to flooding, worsened by local conditions and land-use changes (Bruno et al., 2022).

Moreover, Namara et al., (2022) developed a combined HEC-RAS and HEC-GeoRAS model to map the flood inundation in the Upper Awash River Basin in Ethiopia. The annual peak flood frequency analysis for recurrence intervals of 2, 5, 10, 25, 50, and 100 years was conducted using the calibrated and validated HEC-HMS model. This analysis was based on 25 years (1990–2015) of hydro-meteorological data obtained from the Ministry of Water Resources, Irrigation, and Electricity, as well as the National Meteorological Agency in Ethiopia. HEC-GeoRAS was utilized to extract the cross-sectional bathymetric data in the river channel. Extreme flood values with discharges ranging from 526 to 1461.3 m^3/s were simulated using the provided geometric and hydraulic data. The HEC-HMS model outputs were compared with flood extreme values derived from the Log-Pearson Type III, General Pareto, and Gumbel extreme value distribution functions to enhance reliability. The results indicated that the HEC-HMS model produced higher values for every recurrence interval than the extreme value distribution function outputs. The model provided flood inundation areas ranging from 70 to almost 110 km^2 suggesting that the whole Awash Bello Basin was susceptible to flooding caused by intense rainfall events (Namara et al., 2022).

Focused on the Prince Edward Island in Canada, Dau et al., (2024) developed 2D HEC-RAS model to simulate compound flooding in the island. Flood scenarios with return periods of 10, 25, 50, and 100-year were implemented under current and future climate conditions. Compound flooding utilizes the combined effects of high-tide levels, storm surge, and intense rainfall events within the inland watershed, leading to severe flooding compared to their individual

representation (Yang et al., 2023). The model utilized a High-Resolution Digital Elevation Model (HRDEM) from the National Resources Canada (NRCan) with a resolution of 1 m x 1m grid cells which was derived from airborne LiDAR data. Additionally, two approaches were used to represent hydraulic structures in the HEC-RAS 2D models. The first involved municipal models explicitly included individual culverts and bridges with detailed dimensional data, while island-wide models treated all structures as breached for hydro-connection. The model was calibrated using the September 2021 flood event due to the availability of flood data. A combination of social media video footage and field trips to the region provided a focus on two specific locations in downtown Charlottetown as being susceptible to flooding, with watermarks from the 2021 flood providing valuable insights into the area's inundation history. The study developed municipal and island-wide flood models for PEI as shown in Figure (2.4), highlighting a potential 20% increase in flood extent due to climate change and the need for adaptation measures. While automated DEM breaching helped address hydro-structural gaps, limitations in urban stormwater system representation and hydrometric station availability suggest that future studies should incorporate more advanced flood models like PCSWMM for improved accuracy (Dau et al., 2024).

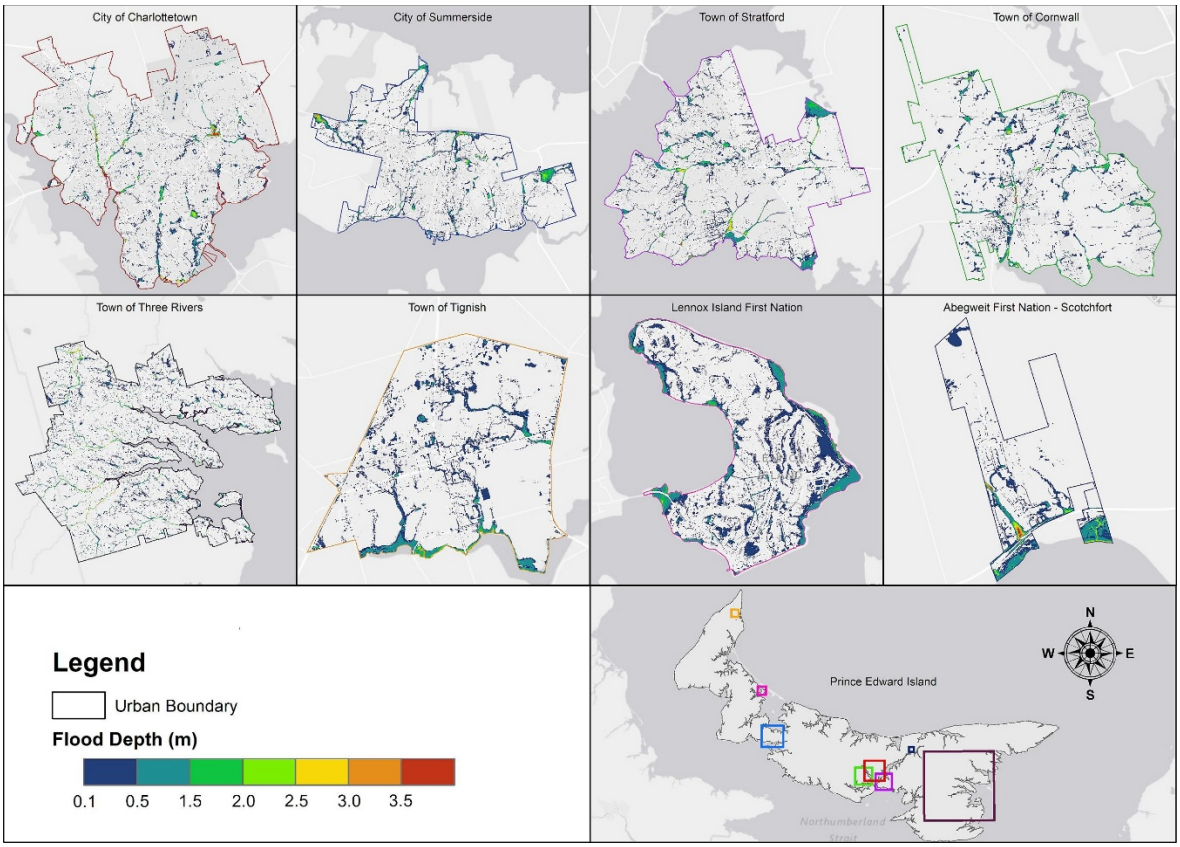


Figure 2.4 – Flood hazards mapping for major municipalities in PEI under 100-year return period current climate condition. (Dau et al., 2024).

As shown above, recent research has emphasized the effectiveness of 2D hydraulic modeling using HEC-RAS in generating high-resolution flood discharge scenarios and enhancing the accuracy of flood hazard mapping (Priya et al., 2020). When combined with GIS, HEC-RAS gains advanced capabilities for flood risk visualization and assessment, making it a powerful tool for urban planning, disaster preparedness, and floodplain management (Azouagh et al., 2018). However, its implementation in real-time flood forecasting presents both advantages and challenges. In its 2D configuration, HEC-RAS is computationally intensive, which can hinder real-time predictions that demand high-speed data processing and rapid model execution (Lago et al., 2023). Furthermore, the high computational demand and lengthy processing time of HEC-RAS simulations pose a significant limitation for real-time flood forecasting, where rapid analysis and

timely decision-making are critical (Idfi et al., 2019). These constraints have driven interest in alternative and complementary technologies, such as AI techniques, which offer faster and more adaptable solutions for improving flood prediction and floodplain management. Table (2.1) provides a summary for all the literature discussed in the section along with its key findings and date and author of the publication.

Table 2.1 - Summary of Numerical Modelling Section of Literature with Key Findings.

Study	Technique	Key Findings
Villazon et al., (2009)	HEC-RAS (Unsteady)	Modeled Piraí River floods; peak discharges: 4,500 m^3/s (1983), 3,200 m^3/s (2006). Roughness variations impacted flood extent (0.8m WSE change). Highlighted need for real-time monitoring.
Agrawal et al., (2016)	HEC-RAS (Unsteady)	Assessed the Lower Dudhana River floods. Peak flood depth: 3.8m; velocity: 2.5 m/s. Proposed embankment raising (412m) to protect nearby villages.
Stoleriu et al., (2019)	HEC-RAS + LiDAR	Compared DEM resolutions for flood mapping. LiDAR-based models (T2, T3) were more accurate. 1% flood impacted 500+ buildings, 4,000+ people.
Faudzi et al., (2019)	HEC-RAS (2D) + HEC-HMS	Simulated 100-year flood in Sungai Langat River. Peak discharge: 426 m^3/s, max inundation: 7.14m, covering 93 km^2. Provided hazard maps for flood preparedness.
Filianoti et al., (2020)	Multi-model comparison	Evaluated hydrologic-hydraulic models in Italy. HEC-HMS & MIKE 11 were most efficient for forecasting, balancing accuracy and computational demand.
Khan et al., (2020)	HEC-RAS (Steady Flow)	Modeled Naray-Khwar canal. 100-year flood risk at Nasir Bagh Bridge due to poor waste management and encroachments. Suggested retaining walls & sewage treatment.
Daham et al., (2020)	HEC-RAS (1D & 2D)	Simulated Al-Gharraf River. 2D model outperformed 1D in accuracy (RMSE & R^2). 1D: WSE 15.14-17.84m, 2D: WSE 14.12-18.82m. Recommended 2D for flood studies.
Afzal et al., (2022)	HEC-RAS + GIS	Modeled Indus River floods (2010, 2015) using satellite data. Accuracy: 64.06%-61.64%. Validated Manning's n values, aiding early flood warnings.
Aryal et al., (2022)	HEC-RAS + FEMA maps	Assessed Rock River climate change impacts. Projected 100-year & 500-year flood extents to increase by 20%.
Bruno et al., (2022)	HEC-RAS + HEC-HMS	Simulated Campo Grande urban floods. Highly accurate models (R^2 = 0.93, RMSE = 1.29, NSE = 0.92). Identified land-use impacts on flooding.
Namara et al., (2022)	HEC-RAS + HEC-GeoRAS	Mapped Upper Awash River floods (2-100 yr). HEC-HMS overestimated flood discharge compared to statistical methods. Flood extent: 70-110 km^2.
Dau et al., (2024)	HEC-RAS (2D)	Simulated compound flooding in PEI under climate change. 10-100 yr floods showed 20% increase in inundation. HRDEM improved accuracy.

2.3 - Artificial Intelligence (AI) and Machine Learning (ML):

Recently, there has been a notable surge in interest toward ML -based models for river flow computations, driven by their ability to handle nonlinear hydrodynamic relationships and complex pattern recognition. These models offer enhanced adaptability to dynamic environmental conditions, seamless integration of heterogeneous data sources, and the capability to learn from historical flood events (Soltani et al., 2021), meteorological data (Grégoire et al., 2023), and other relevant inputs. The application of AI techniques offers tremendous potential in flood prediction and floodplain management, delivering notable advancements in both accuracy and computational efficiency compared to traditional methods (Ebtehaj and Bonakdari, 2022). The application of ML algorithms in flood modeling has yielded significant improvements in predictive accuracy, computational efficiency, and scalability (Gholami et al., 2017).

The effectiveness of ML models in flood modeling is highly contingent on several critical factors, presenting notable challenges. A primary challenge is the necessity for extensive data preprocessing, including data cleaning, normalization, and feature engineering (Ebtehaj et al., 2016), to ensure the integrity, reliability, and quality of the dataset. Furthermore, the selection of relevant parameters and the identification of optimal input combinations remain critical challenges in the application of ML techniques for flood modeling (Ebtehaj and Bonakdari, 2021, Bonakdari and Ebtehaj, 2021, Ebtehaj and Bonakdari, 2021, Ebtehaj et al., 2020). An additional challenge pertains to model selection and hyperparameter optimization, as various ML algorithms can be employed for flood prediction, each exhibiting performance variations based on dataset characteristics and hydrological complexities. The selection of an optimal algorithm necessitates an in-depth understanding of data distributions, spatial-temporal dependencies, and predictive capabilities. Furthermore, hyperparameter tuning, encompassing optimization strategies and

regularization techniques, is crucial for enhancing model generalization while mitigating the risks of overfitting or underfitting. Another critical aspect is model interpretability, as deriving actionable insights from high-dimensional, non-linear ML models remains challenging, particularly in the context of data-driven flood risk assessments and decision-making processes.

Hosseiny et al., (2020) developed a hybrid modeling framework to efficiently predict flood extents and depths over large-scale regions by integrating hydraulic simulations with ML techniques. The International River Interface Cooperative (iRIC) two-dimensional model was first calibrated using measured water surface elevation data to simulate river depths across the domain for various discharge scenarios. These simulation results served as training data for two ML models: a Random Forest (RF) classifier and a Multilayer Perceptron (MLP) regressor. The RF classifier was designed to distinguish between wet and dry nodes within the domain, achieving an overall accuracy of 98.5% on test data. Subsequently, the MLP regressor estimated river depths at the identified wet nodes, yielding a regression coefficient R^2 of 0.88, indicating strong predictive performance. The hybrid approach demonstrated that coupling hydraulic models with ML techniques can significantly reduce computational demands, facilitating real-time, large-scale flood simulations as shown in Figure (2.5) that would be computationally prohibitive using traditional hydraulic modeling methods alone (Hosseiny et al., 2020).

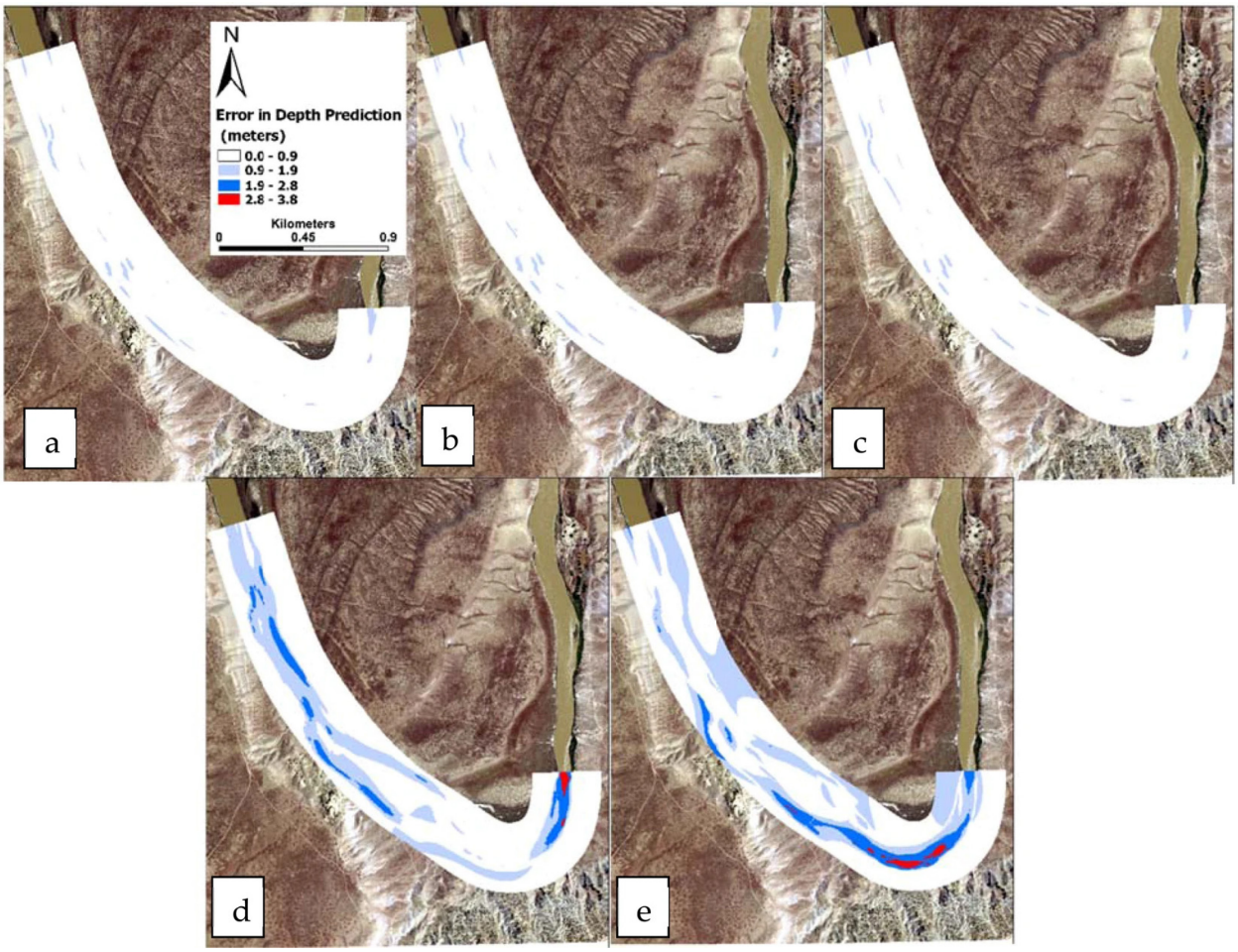


Figure 2.5 – The difference between predicted depth by the ANN and the iRIC (defined as the error of predicted depth by ANN) is depicted for different flow discharges of (a) 20, (b) 30, (c) 45, (d) 225, and (e) 350 m3/s (Hosseiny et al., 2020).

Hao et al., (2021) utilized pre- and post-flood Synthetic Aperture Radar (SAR) imagery from Sentinel-1 satellites, along with the Multi-Error-Removed Improved-Terrain (MERIT) DEM to rapidly assess the extent and depth of flood inundation in a southern region of India that experienced heavy rainfall in August 2019. The study further employed a Random Forest (RF) ML algorithm to predict flood susceptibility in adjacent basins. Due to the limited coverage of the Sentinel-1 satellite data during the flood event, an ML model which employed DEM-derived variables and training data from available SAR images, achieving accuracy rates of 90% for training data and 86% for validation data. The methodology underscored the effectiveness of

combining SAR imagery, DEM data, and ML techniques for rapid and reliable flood mapping, even in regions with incomplete satellite coverage (Hao et al., 2021).

Hosseiny (2021) utilized the U-NetRiver deep learning framework, combined with the iRIC 2D numerical model, to develop a model that enhances the prediction of river flood extents and depths. U-NetRiver is a modified convolutional neural network (CNN) that processes composite images comprising ground elevation and flood discharge data to output water depth predictions. Trained with data from the iRIC modelling software, the U-NetRiver demonstrated the capability to delineate river geometries and inundation areas accurately. Notably, the maximum discrepancy between U-NetRiver's predicted river depths and those obtained from traditional hydraulic models was found to be 2.7 meters, indicating a 29% improvement in predicting maximum flood depths. Similarly, another deep learning framework, U-NETwet, was developed to predict river depth and floodplain inundation. This model utilized high-resolution digital surface and river elevation data, combined with streamflow information, to forecast water depths across river channels and floodplains. Evaluations showed that U-NETwet efficiently identified and predicted river shapes, depths, and wetted areas, with a maximum difference of 1.70 meters compared to traditional hydrological models. These advancements highlighted the potential of deep learning models to provide rapid and accurate flood predictions, offering valuable tools for flood risk assessment and management (Hosseiny, 2021).

Kabir et al., (2021) introduced a data-driven modeling framework that leveraged ML algorithms to forecast probabilistic flood inundation maps with a 3-hour lead time. To classify wet/dry cells, the framework integrated rainfall–discharge models based on Random Forest (RF) techniques with classifiers utilizing multi-layer perceptron networks to effectively emulate the outcomes of hydrodynamic models. Tested during a fluvial flood event in a small town in the

United Kingdom, the model demonstrated a mean flood arrival time difference of 1 hour and 53 minutes compared to the Flood Modeller (FM) hydrodynamic numerical model. Additionally, it achieved a classification accuracy of 88.22% when measured against synthetic aperture radar imagery, slightly outperforming the 86.58% accuracy of the conventional FM numerical model. Notably, this ML-based approach simplified implementation, enhanced flooded area detection, and significantly reduced computational time, offering a practical tool for real-time flood forecasting (Kabir et al., 2021).

Tamiru and Dinka, (2021) presented an Artificial Neural Networks (ANN) model coupled with 2D HEC-RAS numerical modelling to map the flood inundation in lower Baro Akobo Basin River, Ethiopia. The study utilized data from 14 meteorological stations daily from periods of 1999-2005 and 2006-2008. Topographical Wetness Index (TWI) was utilized to train and test the model. The ANN's run-off time series was fed into the HEC-RAS numerical model to generate flood inundation depths. The flood inundation was calibrated and validated using the Normal Difference Water Index (NDWI). The model provided R^2 values of between 0.89 and 0.96 during the training and testing phases. They also utilized Nash-Sutcliffe Efficiency (NSE) and Percent Bias (PBIAS) with values of NSE between 0.86 and 0.89 and PBIAS values between 8.20% and 8.50%. Furthermore, the runoff values provided by the ANN model yielded flood inundation depth up to 250 cm. The results of the integrated ANN and HEC-RAS model successfully improved the quality of flood inundation mapping and enhanced the accuracy of flood predictions in flood forecasting methods (Tamiru and Dinka, 2021).

Rozos et al., (2022) explored the application of ML techniques to develop stage–discharge rating curves, which are essential for estimating river discharge based on water level measurements. Initially, the Density-Based Spatial Clustering of Applications with Noise

(DBSCAN) algorithm was applied to partition the dataset and fit power–law trend lines to each cluster. However, this approach yielded suboptimal performance compared to earlier PINAX statistical methods. Subsequently, DBSCAN was evaluated solely for outlier detection before training a Multilayer Perceptron (MLP) neural network. This allowed the model to account for temporal variations in the stage–discharge relationship, potentially due to changes in the river cross-section over time. The MLP's ability to approximate these temporal dynamics was validated across three real-world case studies (Sakoulevas River, Trikeriotis River, Agrafiotis River, Greece) where the model's schematization of cross-sectional changes aligned with expectations. They concluded that incorporating temporal data as an input in ML models provides a practical and effective alternative to traditional methods for developing stage–discharge rating curves, enabling the capture of temporal variations in river dynamics (Rozos et al., 2022).

Li et al., (2022) employed ML combined with a Hexagonal Discrete Global Grid System (DGGS) to evaluate flood risks in southern New Brunswick, Canada under climate change scenarios. The use of an Icosahedral Snyder Equal Area Aperture 3 Hexagonal Discrete Global Grid System (ISEA3H DGGS) provided a scalable, standard spatial framework for computation, integration, and analysis of multi-source geospatial data. The study utilized a variety of explanatory factors including hydrological, meteorological factors etc. The results indicated that low-lying topography and proximity to permanent water bodies were the primary determinants of flood occurrence, while increasing spring temperatures were found to elevate flood risk by exacerbating hydrological processes such as snowmelt runoff and soil saturation. Their analysis projected a 135–203% increase in flood extents compared to the 2019 flood area by 2100 with detailed flood inundation maps shown in Figure (2.6), highlighting DGGS’s effectiveness in scalable data integration and spatial analysis for flood risk assessment (Li et al., 2022).

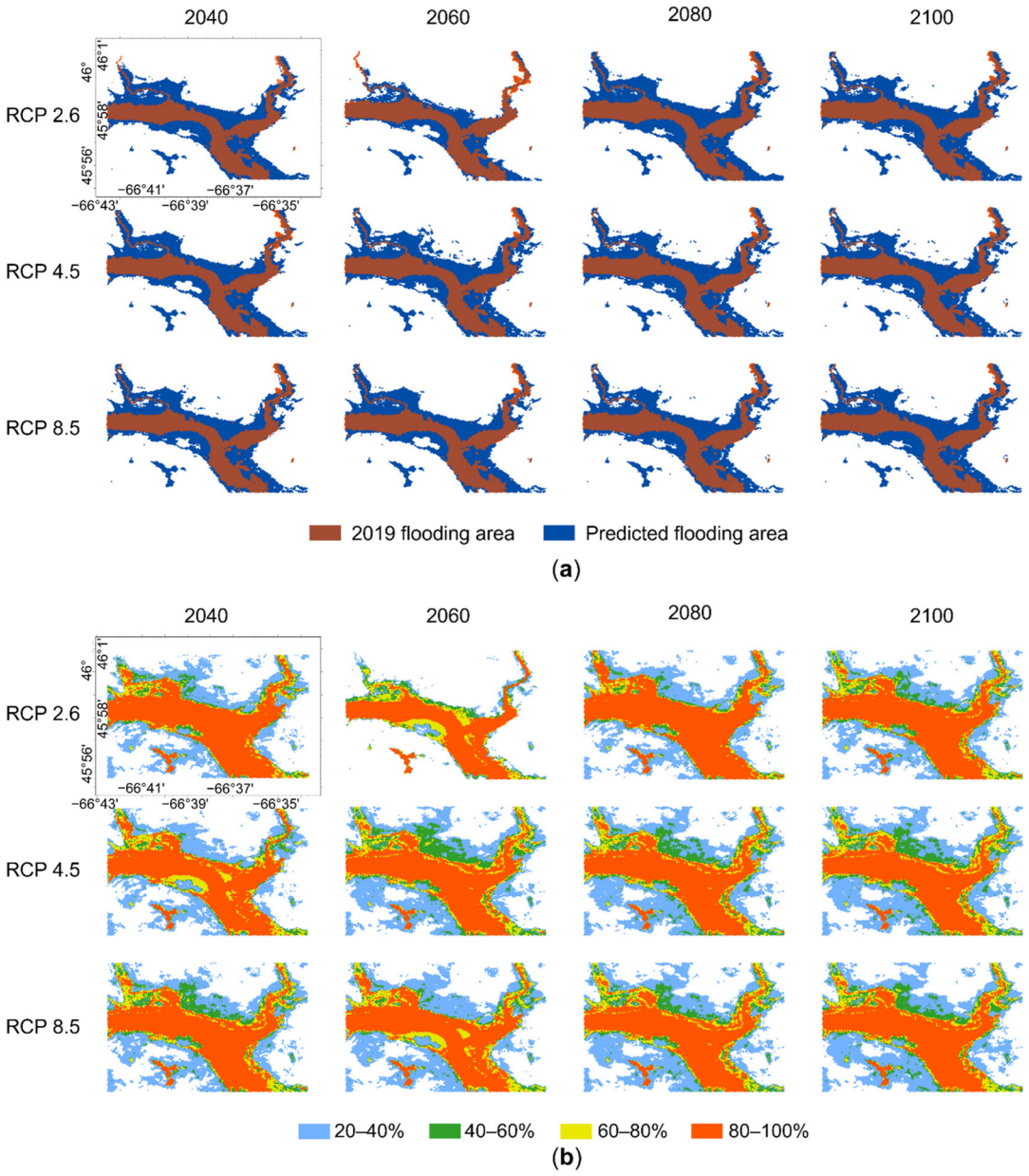


Figure 2.6 – Visualization of (a) flooding extent; and (b) probabilities of flooding predicted by the ensemble model under three climate change scenarios around Fredericton, New Brunswick, in the years 2040, 2060, 2080, and 2100 at level 23 (Li et al., 2022).

Kumar et al., (2023) investigated the application of ML techniques to accurately predict flow velocity in vegetated alluvial channels. Vegetation in water bodies significantly influences flow velocity in open channels due to the resistance it introduces. This resistance impacts local hydrodynamics, affecting various morphodynamic and biophysical processes in rivers, estuaries,

and coastal areas. The study investigated the performance of multiple standalone and hybrid ML techniques to predict flow velocity (U) in vegetated alluvial channels. The study utilized an extensive dataset from existing literature, encompassing variables such as the number of cylinders per unit area (m), flow depth (h), channel slope (i), vegetation height (k), vegetation diameter (D), and the non-dimensional drag coefficient (Cd). Standalone methods like M5Prime and Random Tree (RT) were employed, alongside hybrid approaches such as Additive Regressor (AR) and Bagging (BA). Among the six ML methods explored—M5P, AR-M5P, BA-M5P, RT, BA-RT, and AR-RT—the AR-M5P hybrid model demonstrated superior predictive performance, achieving an R^2 of 0.95, correlation coefficient (CC) of 0.98, Nash-Sutcliffe efficiency (NSE) of 0.95, mean absolute error (MAE) of 0.042, mean squared error (MSE) of 0.003, and PBIAS of 1.47 with sensitivity analysis identifying vegetation height as the most influential variable when predicting flow velocity. These findings underscore the potential of hybrid ML models, particularly AR-M5P, in effectively predicting flow velocity in vegetated alluvial channels, offering a promising alternative to traditional empirical methods (Kumar et al., 2023).

Khoshkonesh et al., (2024) utilized an integrated approach that combined HEC-RAS and ML to improve urban flood prediction and hazard analysis. High-resolution LiDAR data and sophisticated modeling were used to simulate various flood scenarios in the River Thames, West London, England. The study also incorporated socio-economic data to map vulnerable zones in the region. Findings indicated that 2D hydrodynamic models, despite high computational demands, provided superior predictive accuracy for flood propagation, while 1D models exhibited limitations in estimating inundation depth and spatial extent in adjacent urban areas. Utilizing Extra-Trees-Principal Component Analysis (ET-PCA), they received R^2 values of 0.99 showcasing the capability of the model in providing a benchmark in flood risk analysis.

Furthermore, the integration of socioeconomic data further emphasized urban vulnerability and underscored the necessity of targeted flood mitigation strategies (Khoshkonesh et al., 2024).

Sorboni et al., (2024) introduced a Convolutional Siamese Network (CSN) for urban flood mapping in Ottawa, Gatineau, Abbotsford, and Leverkusen, utilizing Sentinel-1 and RADARSAT Constellation Mission (RCM) datasets. The C-band SAR data (10 m and 5 m resolutions) were analyzed against other deep learning (DL)-based segmentation algorithms, including Unet, Unet++, DeepLabV3+, and Siamese-Unet, confirming the robustness of CSN. While Sentinel-1 data had limitations due to medium resolution, RCM data achieved a 0.79 precision score in non-urban settings. Their approach demonstrated reliable performance across multiple test locations, with RCM data yielding higher precision in non-urban areas. However, the medium resolution of Sentinel-1 posed limitations in accurately mapping urban floods (Sorboni et al., 2024).

Moreover, Palash et al., (2024) highlight the urgent need to improve flood preparedness through early warning systems. They introduce the Requisitely Simple (ReqSim) flood forecasting system as shown in Figure (2.7), which leverages data-driven techniques to seamlessly integrate key hydrological and meteorological variables, enhancing prediction accuracy and decision-making efficiency. ReqSim's streamlined structure allowed customization and implementation in medium to large rain-fed river basins worldwide, requiring only water level or discharge measurements at forecast locations. ReqSim demonstrated efficacy by delivering 3–10-day forecasts for various river basins, including the Ganges, Brahmaputra, Amur, Yangtze, Niger, Congo, Zambezi, Danube, Parana, Mississippi, Missouri, Ohio, and Arkansas rivers. Its simplicity, minimal data requirements, and operational accuracy made ReqSim an attractive option for real-time flood forecasting globally (Palash et al., 2024).

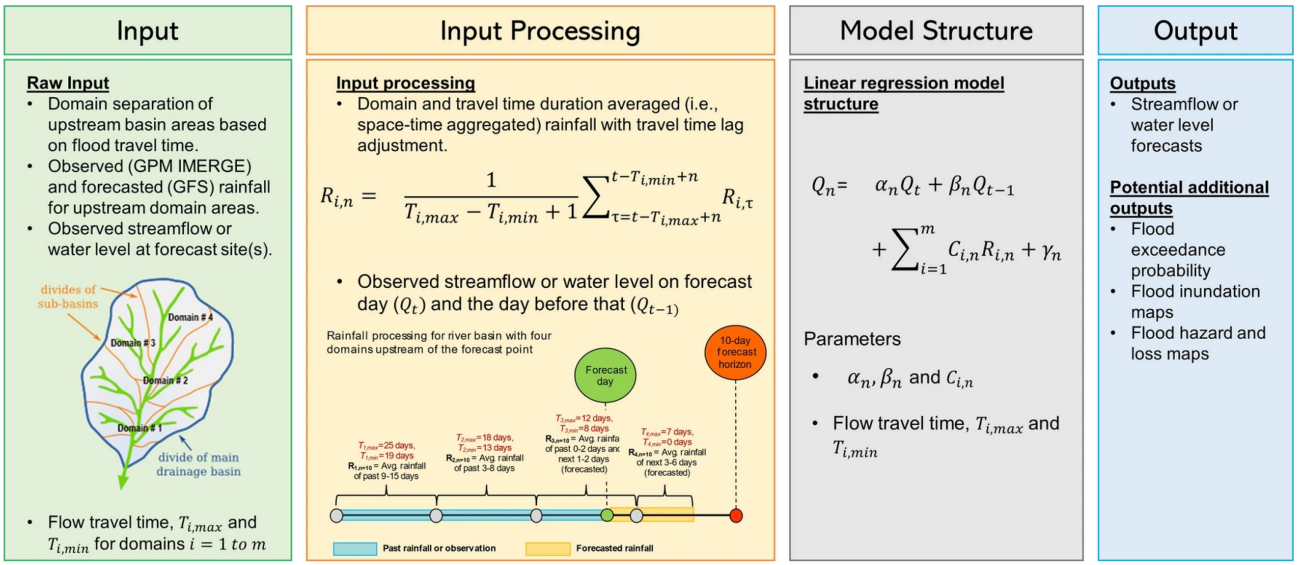


Figure 2.7 – ReqSim flood forecast system's model structure (Palash et al., 2024).

As discussed above, recent and past research has emphasized the applicability and effectiveness of various ML techniques in predicting and modeling various hydraulic scenarios in rivers. AI techniques have significantly advanced flood prediction and floodplain management, offering notable improvements in accuracy and computational efficiency over traditional methods. By analyzing extensive historical data, AI models can swiftly process complex datasets, enabling rapid and precise forecasts. (Ebtehaj and Bonakdari, 2022). The application of ML algorithms in flood modeling has resulted in significant improvements in predictive accuracy, computational efficiency, and scalability. (Gholami et al., 2017). GMDH (Ivakhnenko, 1978) is a notable and recognized ML technique (Safari et al., 2019. Bonakdari et al., 2022) that offers several advantages compared to other ML techniques discussed above. It has been utilized as valuable tool in various hydrological domains, particularly when dealing with complex datasets. It can generate use-friendly explicit equations for decision-makers, enabling direct simulation and prediction of floodplain scenarios and river modeling. Table (2.2) provides a summary for all the literature discussed in the section along with its key findings and date and author of publication.

Table 2.2 - Summary of AI and ML Section of Literature with Key Findings.

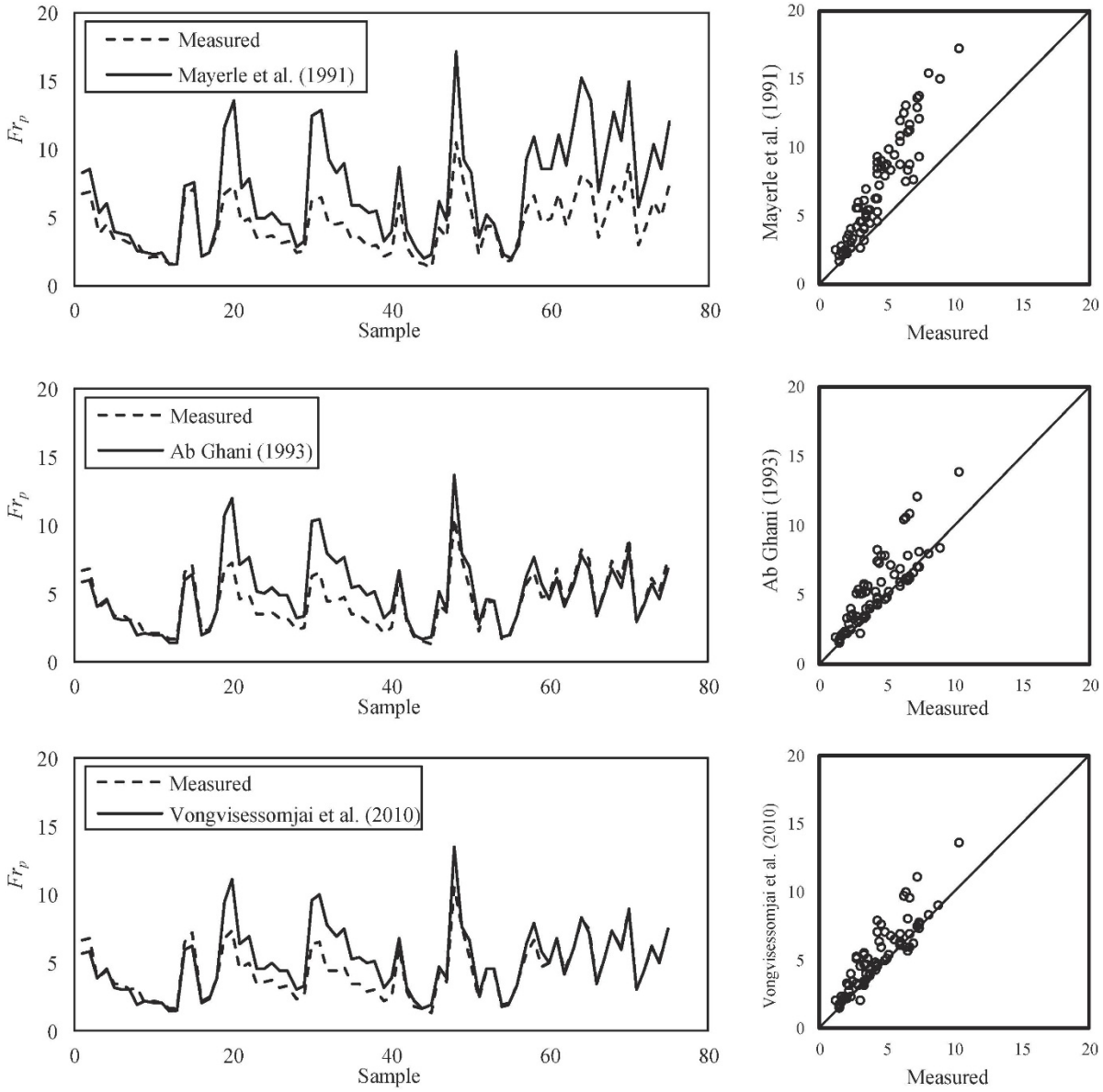
Study	Technique	Key Findings
Hosseiny et al., (2020)	iRIC (2D) + ML	Hybrid model integrated hydraulic simulations with ML (RF & MLP) to predict floods. RF achieved 98.5% accuracy, MLP $R^2 = 0.88$. Reduced computational demands.
Hao et al., (2021)	SAR + RF	Used Sentinel-1 SAR & MERIT DEM for flood mapping in India (2019). RF predicted flood susceptibility with 90% training & 86% validation accuracy.
Hosseiny (2021))	U-NetRiver (DL) + iRIC	U-NetRiver (CNN-based) predicted river depths with 29% improved accuracy. U-NETwet refined floodplain predictions (1.7m max depth error).
Kabir et al., (2021)	RF + MLP	ML-based flood forecasting model (3-hour lead time) in the UK. Achieved 88.22% accuracy, outperforming traditional hydrodynamic models.
Tamiru and Dinka (2021)	ANN + HEC-RAS	Integrated ANN with HEC-RAS for flood mapping in Ethiopia. R^2: 0.89–0.96, NSE: 0.86–0.89. Maximum flood depth: 250 cm.
Rozos et al., (2022)	DBSCAN + MLP	ML-based stage-discharge rating curves for rivers in Greece. MLP captured temporal changes, improving discharge estimates.
Li et al., (2022)	ML + DGGS	Evaluated flood risks in New Brunswick under climate change. Predicted 135–203% flood extent increase by 2100.
Kumar et al., (2023)	ML (AR-M5P)	Predicted flow velocity in vegetated alluvial channels. AR-M5P model had $R^2 = 0.954$, NSE = 0.954, with vegetation height as key factor.
Khoshkonesh et al., (2024)	HEC-RAS + ML (ET-PCA)	Simulated River Thames floods using LiDAR & ML. ET-PCA achieved $R^2 = 0.99$. Identified urban flood vulnerability.
Sorboni et al., (2024)	CSN + SAR	Used Sentinel-1 & RCM SAR data for urban flood mapping in Canada & Germany. RCM had 0.79 precision in non-urban areas.
Palash et al., (2024)	ReqSim (Data-driven)	Developed ReqSim for global flood forecasting (3–10 days). Applied to major rivers, requiring minimal input data.

2.4 – Group Method of Data Handling (GMDH):

GMDH, developed by Ukrainian scientist Alexey Ivakhnenko (Ivakhnenko, 1978) is a unique approach to ML. It enables computers to autonomously generate and select optimal models based on data inputs, minimizing human intervention. The GMDH distinguishes itself from other ML techniques through several key advantages. It features automatic feature selection (Ebtehaj and Bonakdari, 2022), allowing the model to identify relevant variables without manual intervention. Additionally, GMDH employs a self-organizing algorithm that optimizes the model's structure and complexity (Walton et al., 2019), enhancing its adaptability to various datasets. The method's design facilitates interpretability by generating simple and practical models (Soltani et al., 2021), making it accessible to users seeking clarity in their predictive tools. Moreover, GMDH effectively handles non-linear relationships within data (Pazuki et al., 2023), further broadening its applicability across diverse domains (Huang et al., 2020). These characteristics collectively contribute to GMDH's robustness and versatility in modeling complex systems. These advantages make GMDH a valuable tool in various domains, particularly when dealing with complex datasets.

Safari et al., (2021) applied four ML techniques—Gene Expression Programming (GEP), Extreme Learning Machine (ELM), Generalized Structure Group Method of Data Handling (GS-GMDH), and Fuzzy c-means based Adaptive Neuro-Fuzzy Inference System (FCM-ANFIS)—to model sediment transport in open channels. Utilizing four datasets from existing literature, encompassing a wide range of variables such as pipe size, sediment size, sediment volumetric concentration, channel bed slope, and flow depth, the models were developed and compared against conventional regression models using various statistical performance indices. The results demonstrated that these ML techniques outperformed traditional multiple non-linear regression models as shown in Figure (2.8). Among them, GS-GMDH exhibited slightly superior

performance, likely due to its generalized structure. To facilitate practical engineering applications, a MATLAB code was provided for calculating sediment transport in open channels (Safari et al., 2021).



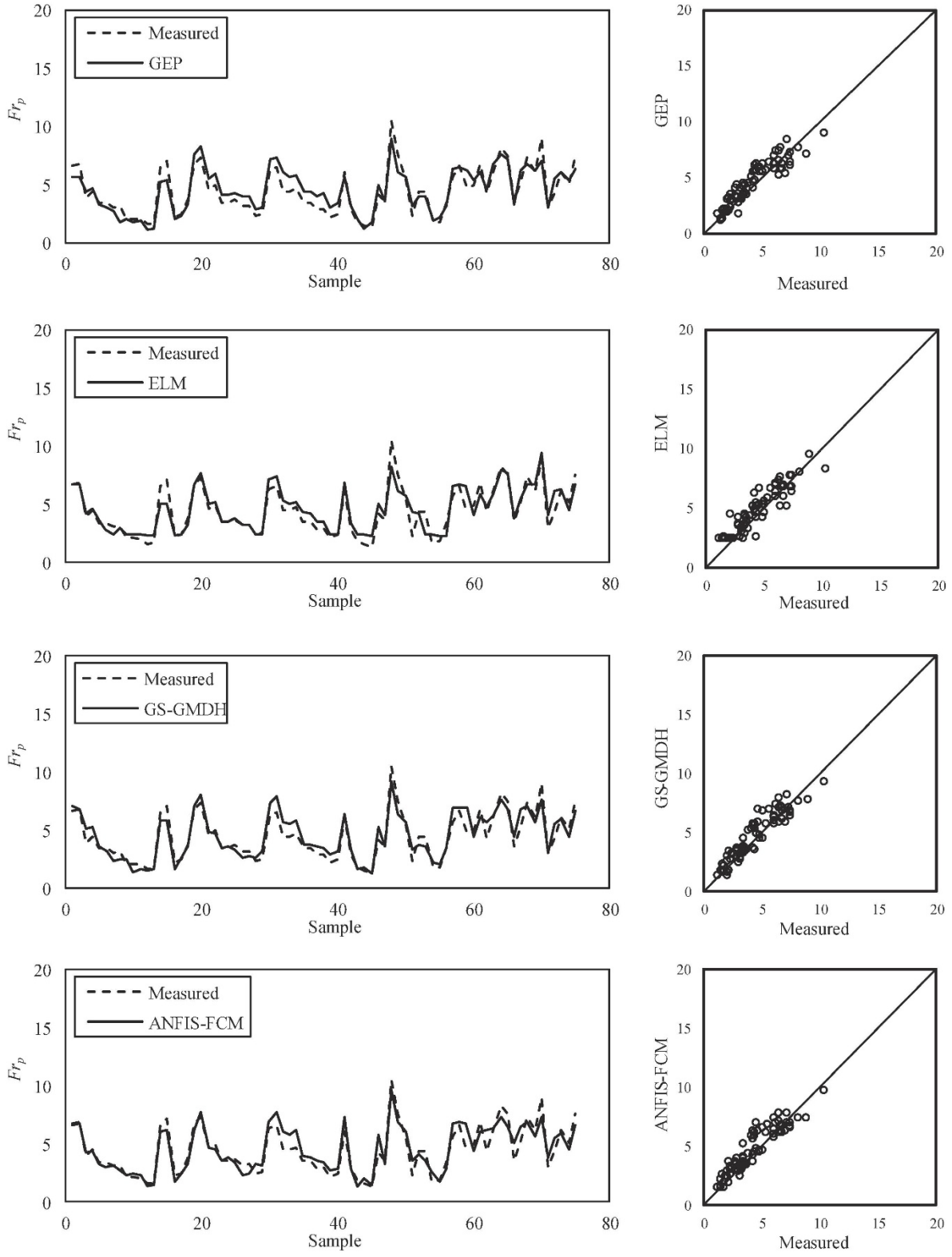


Figure 2.8 – Performance of the machine learning and conventional regression models on testing data, comparison of measured and calculated particle Froude numbers (Fr_p). (Safari et al., 2021).

Elkurdy et al. (2022) utilized a Generalized Structure Group Method of Data Handling (GS-GMDH), a polynomial network approach, to predict river flows. In a study focusing on the Bow River in Alberta, Canada, the GS-GMDH model was trained and tested using daily and hourly flow data from 1990 to 2018 and 1996 to 2018, respectively. For daily flow predictions over a nine-year period, the model achieved an R^2 of 0.64, a Root Mean Square Error (RMSE) of 46.89, a Mean Absolute Error (MAE) of 6.70, a Bias of 1.80, and a Nash-Sutcliffe Efficiency (NSE) of 0.64. When applied to hourly flow data, the GS-GMDH model demonstrated higher accuracy, with an R^2 of 0.99, an RMSE of 3.32, an MAE of 0.99, a Bias of 0.0044, and an NSE of 0.99. The comparison of the results is shown in Figure (2.9). Notably, the model maintained an R^2 greater than 0.90 for predictions up to 17 hours in advance, indicating its potential for short-term river flow forecasting (Elkurdy et al., 2022).

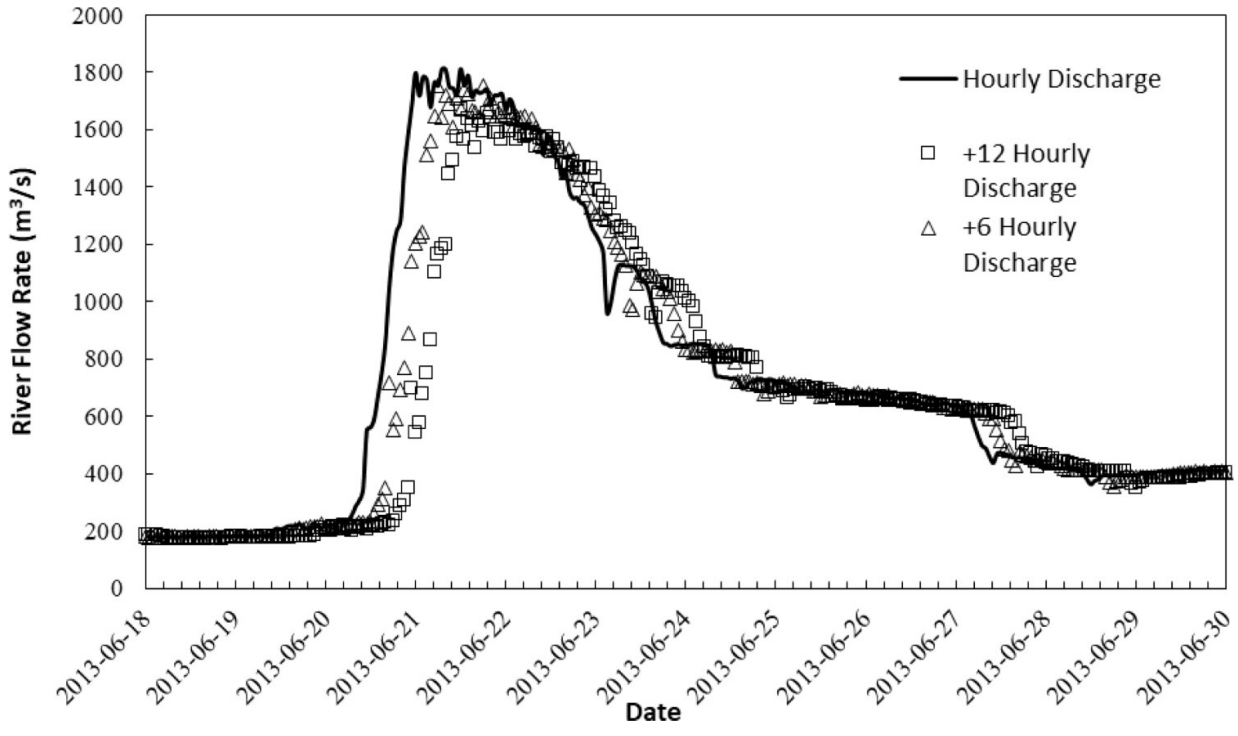


Figure 2.9 – +6-hour and +12-hour prediction of the 2013 flood hydrograph (Elkurdy et al., 2022).

Bonakdari et al., (2022) utilized an Improved Architecture of the Group Method of Data Handling (IAGMDH) which addressed key limitations of the classical GMDH model by incorporating higher-order polynomials, allowing more than two inputs per neuron, and utilizing neurons from non-adjacent layers. In a study applying IAGMDH to estimate bank profile specifications of stable channels, flow discharge (Q) and transverse distance (x) from the centerline were used as input parameters, while the vertical boundary level (y) served as the output parameter. When compared to seven previous models, a GMDH optimized with a genetic algorithm (GMDH-GA), and a Non-Linear Regression (NLR) model, IAGMDH demonstrated superior performance as shown in Figure (2.10), achieving a Mean Absolute Relative Error (MARE) of 0.51, Root Mean Square Error (RMSE) of 0.052, and a correlation coefficient (R) of 0.98 in testing mode. The Vigilar and Diplas Model (VDM) demonstrated better performance compared to previous models, with an RMSE of 0.29. The polynomial curve shape suggested by the GMDH model aligns closely with experimental values, indicating its suitability for predicting cross-sectional dimensions of various channels. Consequently, the IAGMDH model offers a robust and straightforward approach for designing, constructing, and operating artificial channels and rivers (Bonakdari et al., 2022).

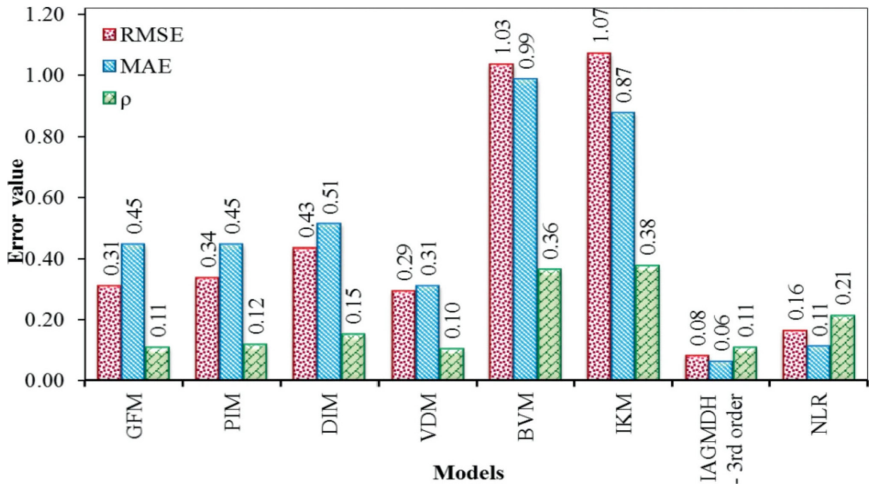


Figure 2.10 – Various index error values in the calculation of bank profile characteristics using different models compared to the corresponding experimental values (Bonakdari et al., 2022).

In summary, the application of GMDH in hydrological modeling has demonstrated significant potential in enhancing predictive accuracy and model interpretability. Studies have shown that GMDH-based approaches, including GS-GMDH and IAGMDH, outperform traditional regression models and even other ML techniques in various hydrological applications, such as sediment transport prediction, river flow forecasting, and channel stability assessment. The adaptability of GMDH, coupled with its ability to handle complex, nonlinear relationships, makes it a valuable tool for hydrological studies. As research continues to refine and expand its applications, GMDH stands as a promising methodology for addressing critical challenges in water resources management and environmental modeling.

Chapter 3

Material and Methods

3.1 - Study Area:

The study area is located upstream of Ottawa, near Morris Island, along a tributary of the Ottawa River that runs parallel to the Mississippi Valley Watershed Figure (3.1). Geographically, it spans approximately 14 kilometers within the coordinates 44°30'N to 45°30'N latitude and 77°00'W to 76°00'W longitude in Ontario, about 50 kilometers northwest of Ottawa. The dataset comprises 22 cross-sectional surveys, compiled using high-resolution LiDAR and river bathymetry data provided by the Mississippi Valley Conservation Authority. The riverbed and adjacent floodplain elevations range from 50 to 60 meters above sea level, with the surrounding floodplains exhibiting undulating topography. The landscape is primarily composed of croplands and herbaceous grasslands, shaping the region's hydrological characteristics.

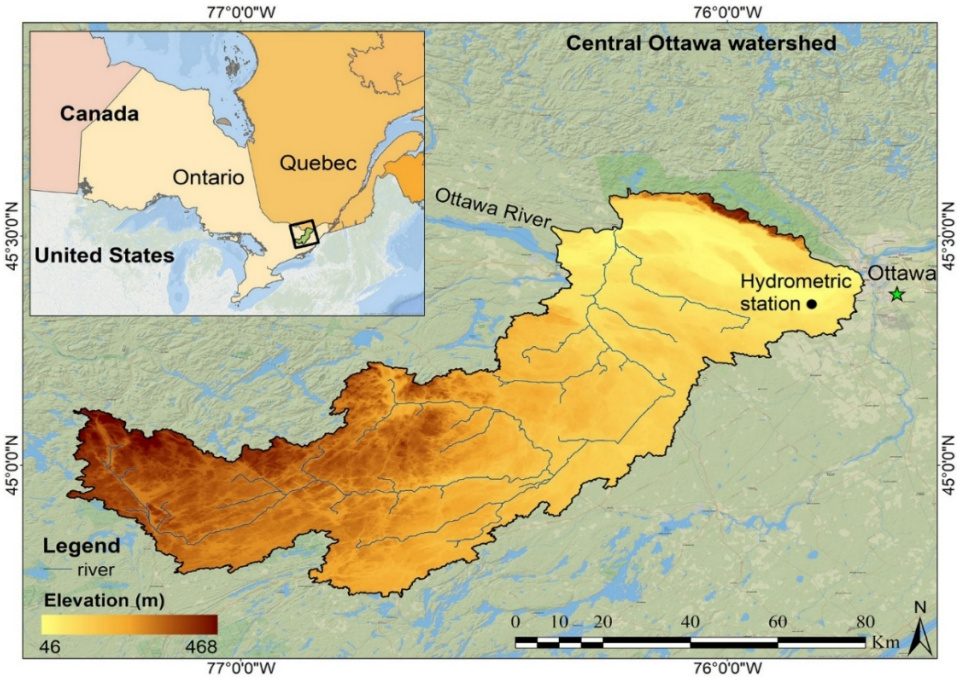


Figure 3.1 - Study Area.

Within this section, a significant floodplain near the town of Quyon has been pinpointed as prone to inundation. This floodplain has seen repeated flooding in recent years, mainly due to spring snowmelt and increased precipitation patterns linked to climate change. In Early May of 2017, the flood in the Ottawa River drove almost 2000 residents from their home in the town of Quyon, this disaster forced the deployment of Canadian soldiers in the Pontiac region of Quebec to aid in the mitigation efforts (The Oklahoman, 2017). In May 2019, severe flooding in Quyon necessitated the evacuation of more than 100 residents because of fears about the dike's structural integrity, which safeguards the region from the Ottawa and Quyon rivers (Canadian Broadcasting Corporation, 2019). This flood also necessitated the temporary closure of the Quyon ferry service that operates between Quyon and Fitzroy Harbor in Ottawa (La Radio Du Pontiac, 2019). As recently as May of 2023, the Municipalité Régionale de Comté (MRC) in coordination with the Ottawa River Regulation Committee warned that the levels of the Ottawa River would rise amid concerns of rainfall leading to minor level floods in the region including Quyon (Pontiac Journal, 2023). These flood incidents underscore the crucial need for comprehensive hydrological and hydraulic studies to enhance flood risk assessment and management strategies.

3.2 - Meteorological Characteristics of the Study Area:

Figure (3.2) displays the historical average, maximum, and minimum temperatures along with average precipitation for the central Ottawa watershed from the period of 1981 to 2023. The temperature distribution across the Central Ottawa watershed reveals distinct spatial patterns. These historical data were obtained from the ERA5_LAND dataset (Muñoz-Sabater et al., 2023), a high-resolution climate reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). This dataset was accessed via the Google Earth Engine platform, utilizing a custom-developed JavaScript code specifically designed to extract the relevant climatic

parameters for the study area. Once the data were retrieved, it was imported into ArcMap 10.8.2 for further processing, analysis, and visualization. In ArcMap, the raster layers were first converted into point format to enable more detailed spatial analysis and facilitate the creation of accurate maps. The radial basis function interpolation meth-od was applied to generate continuous surface maps from the point data.

Average temperatures range between 5.36°C and 6.99°C across the watershed. Lower temperatures dominate the western and northern regions, while the eastern and southeastern areas experience the highest average temperatures. This gradient in temperature distribution reflects the influence of elevation and latitude on the region's microclimate. Higher elevations and northern latitudes contribute to cooler temperatures, while lower elevations in the southeast create favorable conditions for warmer temperatures. The temperature transition is visually evident, shifting from cool blue and purple hues in the southwest (approximately 5.36°C) to warmer yellow and orange tones in the northeast (up to 6.89°C). This pattern highlights the urban heat influence near Ottawa, where developed areas retain more heat than rural and forested regions.

The average minimum temperature varies across the region, with colder areas concentrated in the southwest, depicted by deep purples and pinks (-7.52°C). In contrast, the northeast exhibits warmer orange and yellow tones (-5.39°C). This distribution reflects the influence of elevation, latitude, and urbanization, where lower-lying and rural areas tend to experience more significant temperature drops compared to urbanized zones.

The average maximum temperature map shifts from pink and purple shades in the southwest (19.1°C) to green, yellow, and vibrant orange tones in the northeast (20.40°C), highlighting increased daytime heat retention near Ottawa. These patterns consistently show

warmer conditions in the northeast, influenced by urbanization and lower elevation, while the southwest remains cooler, shaped by higher elevations and less urban development.

Similarly, for precipitation, the western part of the watershed averages around 936 mm annually, indicating relatively drier conditions. Moving eastward towards Ottawa, the precipitation increases significantly, reaching up to 1074mm, suggesting wetter conditions in this area.

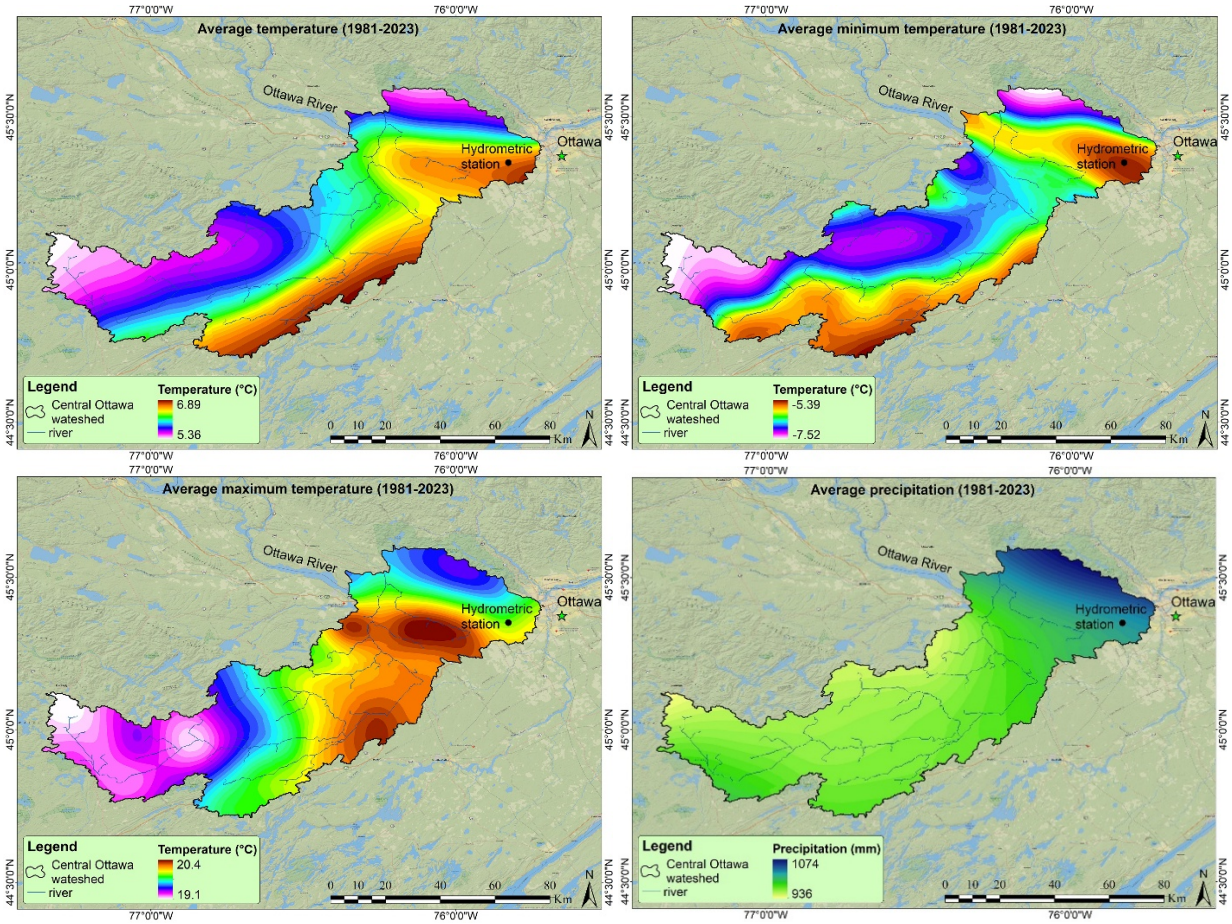


Figure 3.2 - Spatial distribution of historical (1981–2023) average, maximum, and minimum temperatures, along with average precipitation across the central Ottawa watershed.

3.3.1 - One-Dimensional (1D) HEC-RAS Governing Equations:

As outlined in the literature review section of this thesis, HEC-RAS offers the capability to perform 1D modeling of river systems. In 1D mode, the software performs hydraulic computations by generating water surface profile plots for steady-state flow conditions using the

Energy Equation (3.1). This process is executed using the standard step method, which iteratively solves the equation from one cross-section to the next to determine water surface elevations along the river reach.

$$Z_2 + Y_2 + \frac{\alpha_2 V_2^2}{2g} = Z_1 + Y_1 + \frac{\alpha_1 V_1^2}{2g} + h_e \quad (3.1)$$

where:

Z_1, Z_2 are the main channel invert elevations

Y_1, Y_2 are flow depths at each cross-section

V_1, V_2 are the depth averaged flow velocities at each cross-section

α_1, α_2 are the velocity weighting coefficients

g is the gravitational acceleration constant

h_e is the energy head loss

Beyond steady-state computations, HEC-RAS can perform 1D unsteady flow routing by applying the principles of conservation of mass and momentum. This unsteady flow modeling better represents natural river dynamics, as steady-state conditions are rarely observed in real-world riverine systems. The conservation of mass principle dictates that the total inflow to a system must equal the sum of the outflow and the change in storage, which is mathematically expressed through the Continuity Equation (3.2).

$$\frac{\partial A_t}{\partial t} + \frac{\partial Q}{\partial x} - q_l = 0 \quad (3.2)$$

where:

A is the cross-sectional area

t is time

Q is the flowrate

x is the length along the channel

q_l is the lateral inflow per unit length

The conservation of momentum principle asserts that the total momentum entering a control volume, along with the influence of external forces such as pressure, gravity, and friction, must be equal to the rate of momentum accumulation within that volume. This relationship is mathematically expressed by the momentum Equation (3.3).

$$\frac{\partial_Q}{\partial t} + \frac{\partial_{QV}}{\partial x} + gA\left(\frac{\partial z}{\partial x} + S_f\right) = 0 \tag{3.3}$$

where:

V is velocity

$\frac{\partial z}{\partial x}$ is the water surface slope

S_f is the friction slope

By utilizing these equations, 2D flow fields can be simplified into a 1D representation by treating the main channel and its floodplain as separate flow paths. This approach allows for the computation of unsteady flow rate solutions in HEC-RAS, as formulated in Equations (3.4) and (3.5) (Brunner, 2024).

$$\frac{\partial A}{\partial t} + \frac{\partial(\phi Q)}{\partial x_c} + \frac{\partial[(1-\phi)Q]}{\partial x_f} = 0 \tag{3.4}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial\left(\frac{\phi^2 Q^2}{A_c}\right)}{\partial x_c} + \frac{\partial\left[\frac{(1-\phi)^2 Q^2}{A_f}\right]}{\partial x_f} + gA_c\left[\frac{\partial z}{\partial x_c} + S_{fc}\right] + gA_f\left[\frac{\partial z}{\partial x_f} + S_{ff}\right] = 0 \tag{3.5}$$

where:

Q is total flow

$\phi = \frac{K_c}{K_c + K_f}$

K is the conveyance

The 1D model is most effective for rivers with uniform flow patterns and well-defined, compact streamlines (Büchele et al., 2006). It is relatively simple to develop and requires significantly less computational time than 2D models. While 1D models can accurately estimate floodwater levels and delineate inundation zones, they may be less reliable in expansive floodplains, where 2D modeling provides better accuracy (Büchele et al., 2006). Key limitations of 1D modeling include its inability to simulate lateral flood wave diffusion, the discrete representation of floodplains as cross-sections rather than a continuous surface, and the arbitrary orientation of cross-sections, which may not always align with natural flow paths (Teng et al., 2017).

3.3.2 - Two-Dimensional (2D) HEC-RAS Governing Equations:

The 2D modeling capabilities in HEC-RAS enable its integration with GIS tools to perform discrete spatial computations, generating simulated spatiotemporal representations of flood inundation or other significant hydrodynamic events (Brunner, 2024). HEC-RAS solves the Saint-Venant equations, incorporating the effects of turbulence and the Coriolis force to improve the accuracy of hydrodynamic simulations in complex flow conditions. It is designed to model dynamic flood waves, simulating rapidly varying flow scenarios, such as dam breaks and flash floods. Additionally, HEC-RAS supports the analysis of mixed flow regimes, seamlessly transitioning between subcritical and supercritical flow conditions. Its capabilities also extend to tidal environments, enabling the assessment of coastal flooding, estuarine dynamics, and storm

surge impacts (Dhungel et al., 2019, Afzal et al., 2022). The 2D Conservation of Mass Equation (3.6) and Conservation of Momentum Equation (3.7-3.8) are shown below.

$$\frac{\partial H}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} + q = 0 \tag{3.6}$$

where:

H is the Water Surface Elevation

h is the flow depth

u is the velocity x-component

v is the velocity y-component

q is a source/sink flux term

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -g\frac{\partial H}{\partial x} + v_t\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) - c_f u + fv \tag{3.7}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -g\frac{\partial H}{\partial y} + v_t\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) - c_f v + fu \tag{3.8}$$

where:

u is the velocity x-component

v is the velocity y-component

v_t is the horizontal eddy viscosity coefficient

c_f is the bottom friction coefficient

f is the Coriolis parameter

Since overland flow is considered shallow water, the diffusion wave approximation is applied to account for gravity and bottom friction effects while neglecting other less significant terms (Brunner, 2024). This approach enhances computational efficiency and stability, allowing

the model to run faster while maintaining accuracy in flood extent predictions. The diffusion wave approximation simplifies and expresses the momentum Equation (3.9).

$$V = \frac{-(R(H))^{2/3}}{n} \frac{\nabla H}{|\nabla H|^{1/2}} \tag{3.9}$$

where:

V is the velocity vector

R is the hydraulic radius

∇H is the surface elevation gradient

n is the Manning's roughness coefficient

Therefore, the mass conservation equation for unsteady flow is expressed mathematically in Equation (3.10).

$$\frac{\partial H}{\partial t} - \nabla \cdot \frac{(R(H))^{5/3}}{n|\nabla H|^{1/2}} \nabla H + q = 0 \tag{3.10}$$

A 2D computational model is well-suited for spatially varied hydraulic analysis in complex river geometries and flow patterns, particularly when detailed parameters such as flow direction and velocity distributions are required (Büchele et al., 2006). Additionally, its ability to incorporate a wide range of input boundary conditions—including precipitation data, flow and stage hydrographs, and lateral inflows—allows for the simulation of diverse scenarios and the representation of complex hydraulic conditions with greater accuracy.

3.3.3 – Digital Elevation Models (DEM):

For hydraulic analysis within HEC-RAS, integrating geometric models of river systems with high-resolution elevation data is a fundamental step in achieving accurate hydrodynamic simulations. Typically, these geometric models are combined with DEMs, which are commonly

derived from LiDAR technology. LiDAR-based DEMs are often sourced from local, regional, or national agencies and are provided as raster datasets with varying grid resolutions, depending on the availability and quality of site-specific data.

Once acquired, the DEM is imported into HEC-RAS as a terrain file, serving as the foundation for modeling flood inundation, channel morphology, and other key hydraulic processes. The selection of an appropriate DEM resolution plays a critical role in determining the accuracy of the model outputs. Higher-resolution DEMs can capture finer topographic details, improving model precision, particularly in areas with complex terrain, floodplains, or urban landscapes. However, finer resolutions also require greater computational resources and processing time. Conversely, coarser DEMs, while computationally efficient, may overlook important small-scale features, potentially reducing model accuracy.

Therefore, understanding the impact of DEM resolution on hydraulic model precision is essential for ensuring reliable flood simulations, accurate water surface elevation predictions, and effective flood risk assessments.

3.4 - Numerical HEC-RAS Model Development:

The methodology for generating a coupled 1D/2D HEC-RAS hydrodynamic model output involves the integration of geometric, hydraulic, and hydrological inputs to simulate riverine and floodplain dynamics with high spatial and temporal accuracy. The geometric dataset consists of high-resolution DEMs, overbank and channel bed elevation profiles, and topographic data sourced from LiDAR surveys or other remote sensing technologies. These datasets are essential for defining terrain morphology and flow pathways within the computational domain.

Hydraulic parameters incorporated into the model include Manning's roughness coefficients (n-values) for channel and floodplain surfaces, as well as initial and boundary conditions, which govern inflow and outflow characteristics. Additionally, hydrological data layers such as streamflow time series, stage hydrographs, and hydraulic properties derived from gauging stations further enhance the model's predictive accuracy. These inputs provide critical insights into discharge variations, flood wave propagation, and transient flow conditions across the study area.

The modeling workflow involves preprocessing and parameterization, followed by computational simulations to produce key hydraulic outputs such as water surface elevation profiles, velocity distributions, and floodplain flow dynamics. The model undergoes rigorous calibration and validation phases, where simulated water surface elevations, velocity fields, and inundation extents are compared against observed data from gauging stations, river cross-sectional surveys, and historical flood records. The calibration process involves iterative refinement of Manning's coefficients and boundary conditions to minimize discrepancies between observed and simulated results.

Once validated, the model is configured to run unsteady flow simulations to capture temporal variations in flood dynamics across different return periods, including 50-year, 100-year, 1000-year, and extreme 10,000-year flood events. These simulations provide critical flood hazard assessments by delineating inundation extents through water surface elevation profiles, mapping velocity gradients to understand flow energy distribution, and visualizing floodplain flow interactions.

Floodplain maps generated for each return period serve as key decision-support tools, offering spatial representations of flood risks and identifying vulnerable regions. The accuracy of

these model outputs is further validated through comparative analysis with satellite-based inundation maps and historical remote sensing data. Discrepancies between simulated and observed flood extents are systematically analyzed, and necessary model refinements are implemented to improve predictive reliability.

The final stage of the modeling process involves a comprehensive comparative assessment between satellite-derived flood extents and model-simulated inundation areas. This validation ensures the robustness of the HEC-RAS model in predicting flood hazards under various hydrological scenarios. The iterative modeling framework enhances flood risk assessment, supports hydrological scenario analysis, and informs the development of mitigation strategies for adaptive floodplain management under changing climatic and hydrological conditions. The overall methodology framework is illustrated in Figure (3.3).

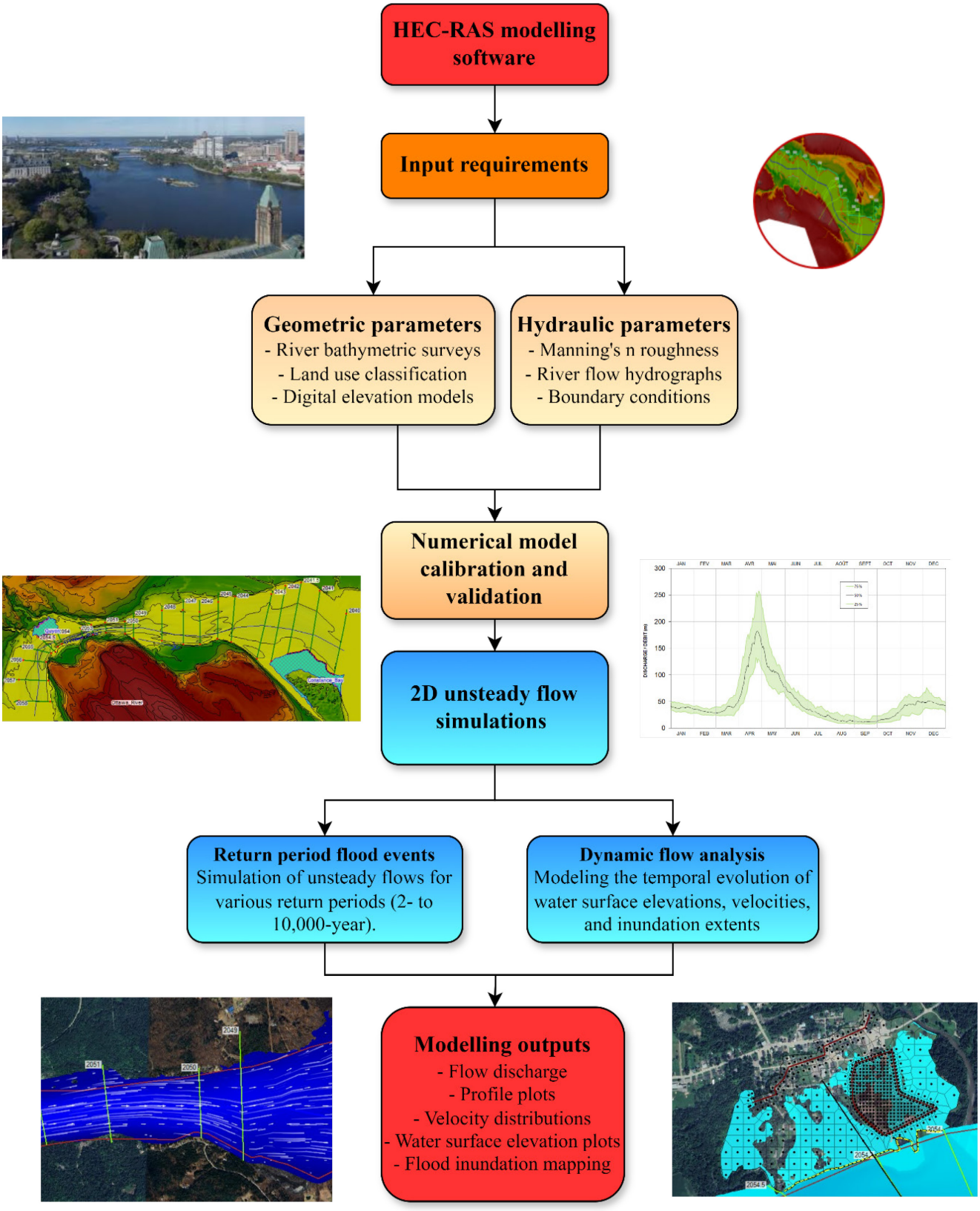


Figure 3.3 - Methodology flowchart for HEC-RAS Numerical modelling.

3.5 – Data Preparation and Input Requirements

3.5.1 – River Geometry (Bathymetry)

The riverbed topography was modeled using a series of cross-sectional profiles derived from high-resolution LiDAR data, integrated with detailed riverbed bathymetric surveys to enhance the accuracy of subaqueous terrain representation. These cross-sectional profiles, referred to as river stations in HEC-RAS, are systematically arranged along the river channel in a sequential manner, with station numbers increasing in the upstream direction.

To accurately represent the spatial configuration of the river system, the spacing between successive cross-sections is defined using downstream reach lengths. These reach lengths are further decomposed into three distinct components—left overbank, main channel, and right overbank reach lengths—allowing for a precise characterization of channel meandering and variable flow pathways across the floodplain. This segmentation ensures that overbank flow interactions and lateral flow redistribution are appropriately captured within the model. Each cross-section is georeferenced within a Cartesian coordinate system, where the lateral extent is measured from the left bank toward the right bank across the channel width.

The elevation profile at each cross-section is extracted from the LiDAR and bathymetric datasets, providing a detailed representation of terrain morphology, channel bed undulations, and floodplain topography. Additionally, each cross-section is anchored to a one-dimensional centerline, which serves as the primary reference for flow computations within the HEC-RAS model.

The hydrodynamic model's accuracy was further enhanced by incorporating data provided by the Mississippi Valley Conservation Authority (MVCA) as part of their floodplain mapping

initiative. The accuracy of the point cloud on a flat ground surface without vegetation (Fundamental Vertical Accuracy) was 13 cm. The Supplemental Vertical Accuracy considering crop/pasture, forested/wooded, and thicket/shrub land cover was 25.50 cm, and the Consolidated Vertical Accuracy, merging all land cover types with open flat surfaces, was 30.30 cm. These datasets, consisting of surveyed elevation points and corresponding coordinate references, were directly imported into the model, ensuring precise alignment with regional hydraulic characteristics and topographic conditions.

Example plots of the Cross-sections at the upstream, mid-stream and downstream end of the Ottawa River Channel are shown below in Figures (3.4), (3.5) and (3.6).

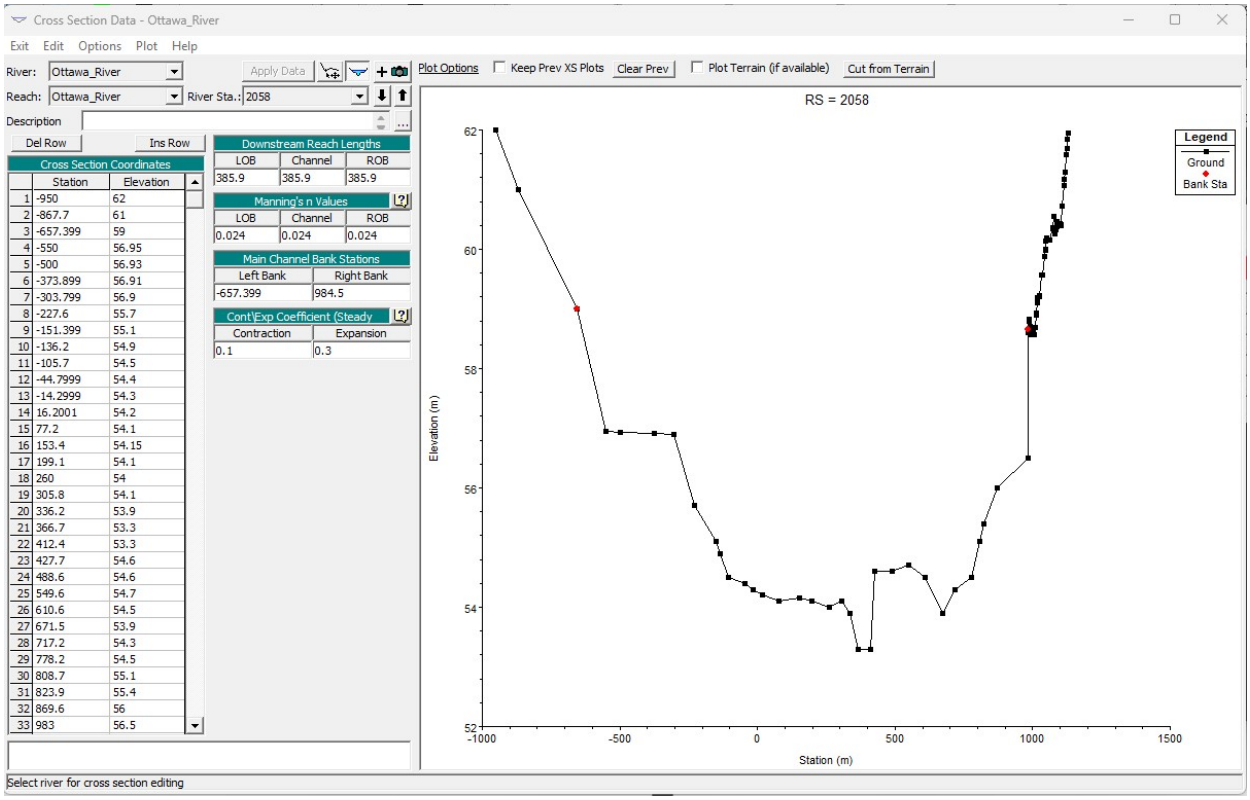


Figure 3.4 - Bathymetry extracted from HEC-RAS of Cross-section 2058 at the Upstream end the River Channel.

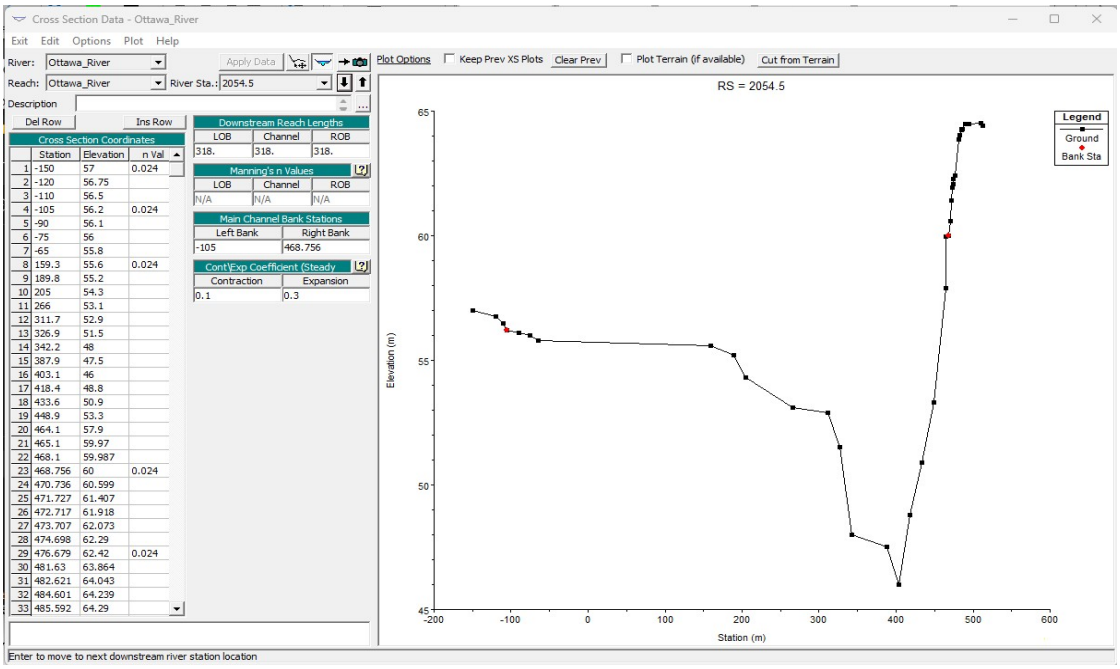


Figure 3.5 - Bathymetry extracted from HEC-RAS of Cross-section 2054.5 located mid-stream in the River Channel.

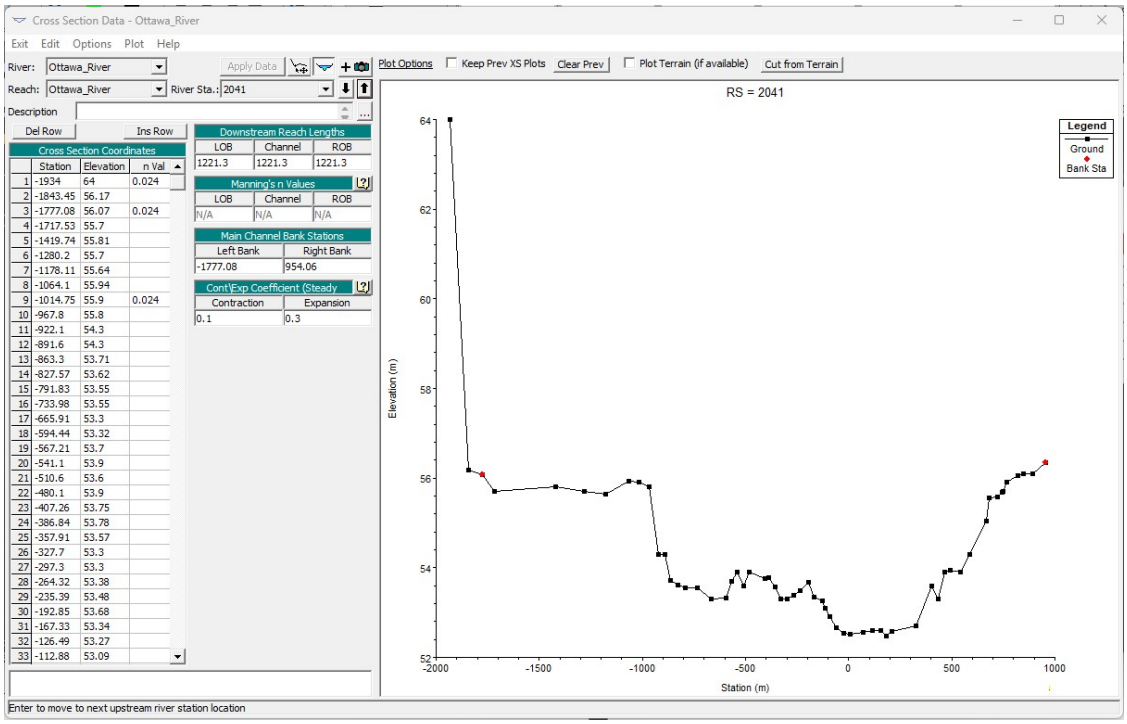


Figure 3.6 - Bathymetry extracted from HEC-RAS of Cross-section 2041 located at the downstream end of the River Channel.

3.5.2 – Floodplain Geometry (Terrain Layer)

A fundamental geometric component of a 2D HEC-RAS model is the incorporation of a HRDEM to precisely represent the floodplain, channel morphology, and surrounding terrain features. In this study, the HRDEM was sourced from the CanElevation Series, an authoritative dataset provided by the Government of Canada, ensuring data consistency, accuracy, and spatial reliability. The HRDEM was imported into the modeling framework at a 1m × 1m grid cell resolution, providing a fine-scale representation of elevation variations across the study area.

This high-resolution topographic dataset is critical for capturing localized terrain undulations, embankment structures, and subtle variations in floodplain morphology, all of which significantly influence hydrodynamic flow patterns, overbank inundation, and channel-floodplain interactions. The detailed resolution enhances the model's ability to resolve hydraulic gradients and accurately predict flow propagation under different flood scenarios.

The model's spatial reference framework was established using the NAD_1983_CSRS_UTM_Zone_18N projected coordinate system, ensuring geospatial alignment with regional mapping standards and facilitating seamless integration of data across different hydrological and geographic datasets. This standardized projection system enhances compatibility with external GIS datasets and remote sensing analyses, further improving the accuracy of flood risk assessments and terrain-based hydraulic modeling. The terrain layer incorporated in the hydraulic model is shown in Figure (3.7).

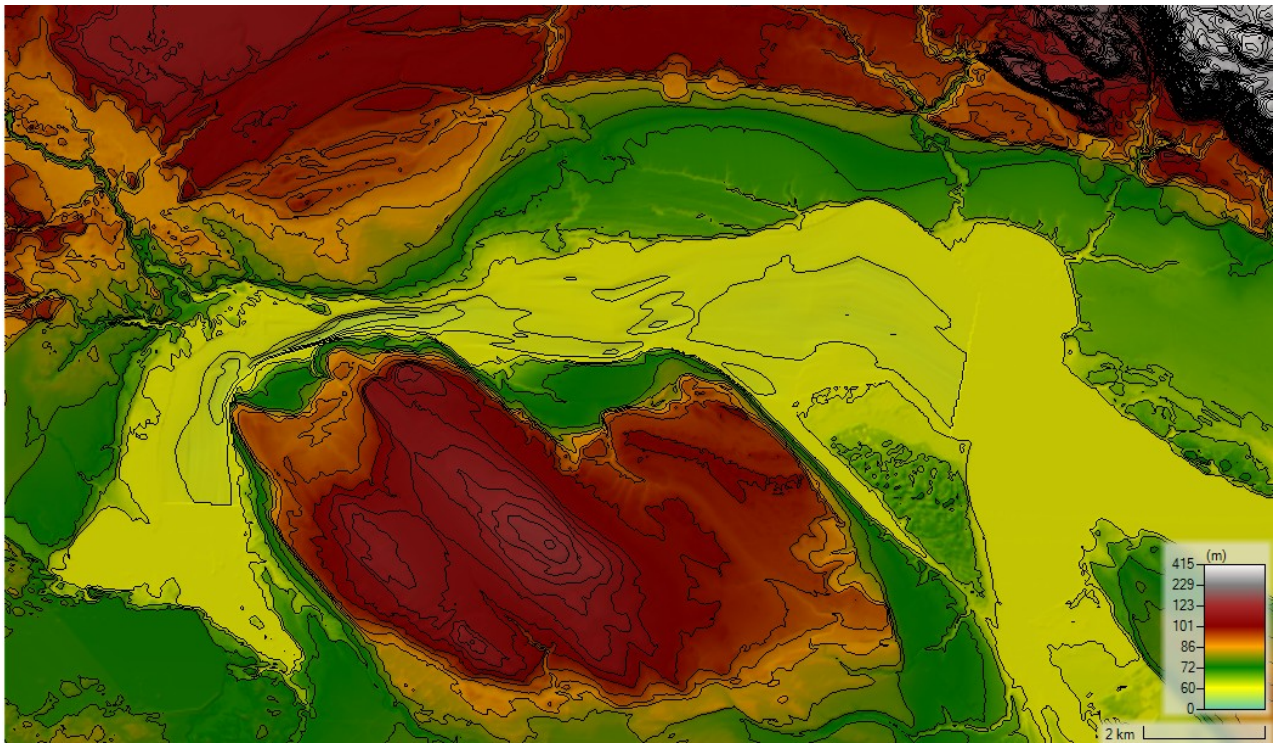


Figure 3.7 - Terrain Layer of the surrounding floodplains of the channel network.

3.5.3 – Coupled 1D/2D Terrain of River Channel & Floodplain

To establish a comprehensive and cohesive representation of the river system, bathymetric cross-sectional data of the river channel were seamlessly integrated with the HRDEM of the surrounding floodplains as shown in Figure (3.8). This integration was essential for accurately capturing the complex interactions between in-channel flow and overbank inundation processes. By merging these datasets, a smooth and continuous transition was achieved between the floodplain's high-resolution topographic data and the riverbed's detailed bathymetric data, ensuring that critical variations in elevation and channel morphology were well represented. This approach allowed for the precise modeling of hydraulic interactions not only within the main channel but also across the adjoining floodplain, where variations in terrain elevation influence the extent and dynamics of floodwater movement.

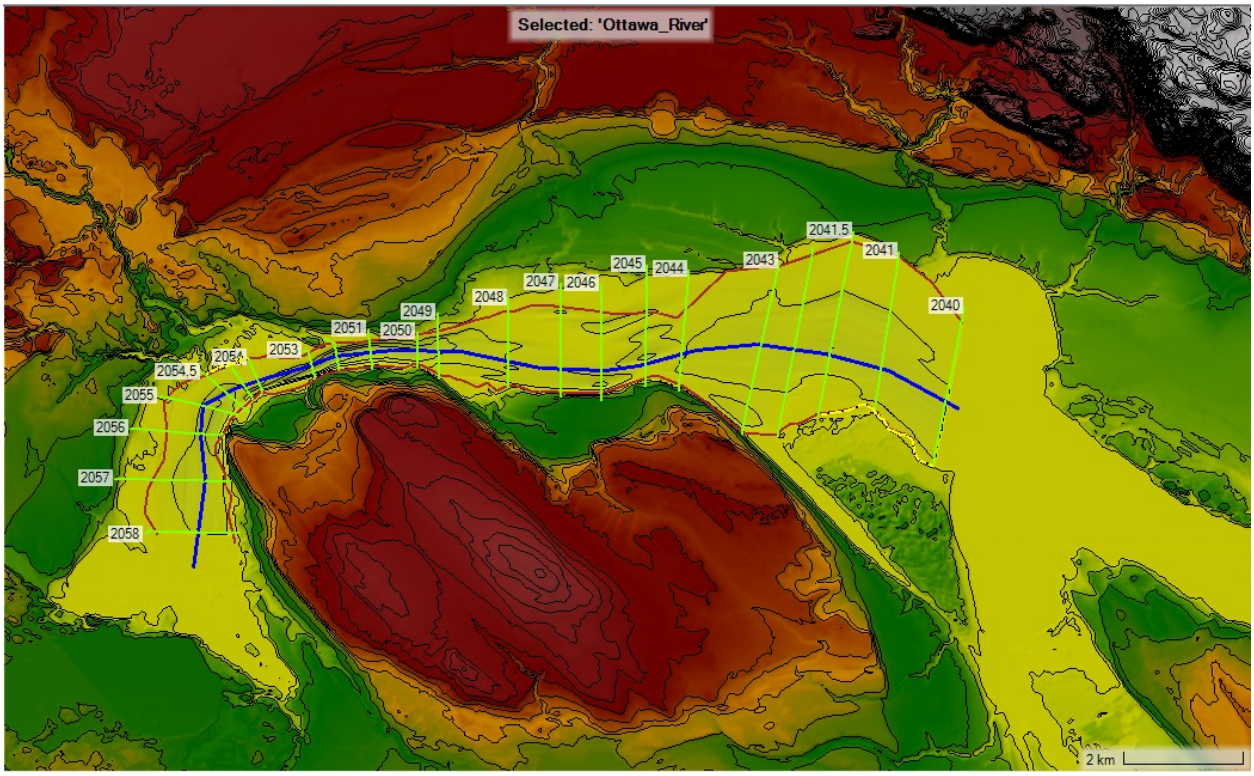


Figure 3.8 - Coupled 1D/2D Terrain Layer of the River Channel and surrounding Floodplains.

3.5.4 - Manning's Roughness for River Channel and Floodplains:

In addition to geometry data, Manning's "n" roughness coefficients were assigned to each cross-section to accurately represent surface resistance and flow conveyance characteristics. The channel roughness coefficient was fixed at 0.024, as specified by the MVCA based on their floodplain management initiative. For the right and left overbank areas, Manning's roughness coefficients varied depending on the land cover characteristics surrounding the river stations. These values ranged from 0.035 for herbaceous grasslands to as high as 0.40 for densely vegetated grasslands, reflecting the variability in surface roughness. The spatial distribution of these roughness values was determined by importing and processing the '2020 Land Cover of Canada' raster dataset, sourced from the Government of Canada's Open Canada Maps repository. This dataset provided a high-resolution classification of land cover types, enabling precise allocation of roughness values across the floodplain. Figure (3.9) shows the land cover layer used in the model.

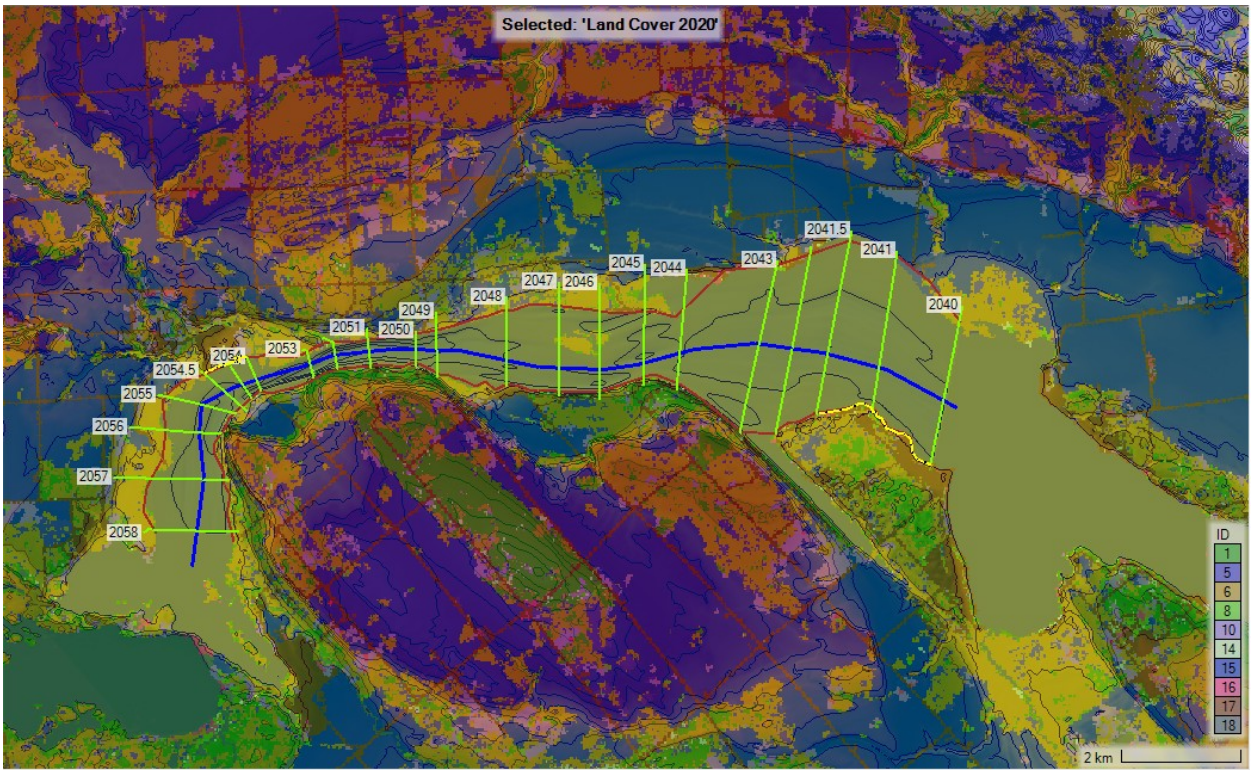


Figure 3.9 - Land Cover Layer imported from the "2020 Land Cover of Canada' raster dataset.

To ensure accuracy and alignment with established hydraulic modeling standards, the selected Manning's 'n' values for floodplain areas were cross-referenced with empirical data and recommendations outlined in the HEC-RAS 2D User's Manual (Brunner, 2024). This validation step helped refine the roughness parameterization, ensuring that the model accurately captured the interaction between surface resistance, flow velocity distribution, and floodplain inundation dynamics. A comprehensive list of Manning's roughness coefficients assigned to different land cover types is presented in Table (3.1).

Table 3.1 - Manning's Roughness for River Channel and Floodplain.

Terrain ID	Terrain Type	Manning's n Value
1	Grassland	0.40
5	Developed - Open Space	0.05
6	Cultivated Crops	0.04
8	Grassland-Herbaceous	0.035
10	Evergreen Forest	0.11
14	Palustrine Scrub-Shrub Wetland	0.05
15	Palustrine Emergent Wetland	0.06
16	Estuarine Forested Wetland	0.03
17	Estuarine Scrub-Shrub Wetland	0.048
18	River Channel	0.024

In addition to the literature-based assignment and land cover mapping, the selected Manning's "n" values for the floodplains and surrounding areas were further evaluated through calibration against observed flood behavior during the 2019 flood event. Specifically, the HEC-RAS model outputs (e.g., water surface elevation and flood extent) were compared to hydrometric records at Britannia station and available aerial flood imagery. Iterative adjustments were made to refine "n" values in floodplain areas where discrepancies between observed and simulated flood extents were noted. This calibration ensured that roughness values not only reflected realistic surface resistance based on land cover but also produced hydraulically consistent model outputs. This dual-validation approach utilizing empirical assignment supported by observation-based refinement enhanced the credibility and accuracy of the floodplain modeling framework.

3.5.5 – Flood Data (Hydrograph and Boundary Condition Setup):

Beyond establishing the geometric framework and selecting appropriate Manning's roughness coefficients, careful definition of both boundary and initial conditions is critical to ensure the accurate performance of a HEC-RAS model simulation. These conditions serve as

critical inputs that define the flow regime entering and leaving the model domain, as well as the starting hydraulic state within the system. For this study, we utilized historical monthly flow data obtained from Environment and Climate Change Canada's Hydrometric Station 02KF005, situated at Britannia, Ontario, at coordinates 45°21′04″ N, 75°49′35″ W. This hydrometric station has maintained a continuous and extensive record of both discharge and Water Surface Elevation Level (WSEL) data for over 110 years, making it one of the most reliable long-term monitoring sites along the Ottawa River.

The availability of such a robust and temporally extensive dataset was instrumental in ensuring the accuracy and reliability of our model calibration and validation processes. By incorporating flow data spanning multiple decades, the model was able to account for variations in hydrological patterns, including seasonal fluctuations, long-term trends, and extreme flow events. This comprehensive approach allowed for the development of a model that is not only well-calibrated to historical observations but also capable of providing reliable predictions under a range of potential future conditions.

To illustrate the temporal variability and long-term trends in river discharge, Figure (3.10) presents the historical flow discharge data recorded at Station 02KF005 from 1960 to 2023. This period captures a wide range of hydrological events, from low-flow conditions to significant flood events, all of which played a crucial role in validating the model's ability to simulate realistic hydraulic responses within the river system. The use of such an extensive and high-quality data set strengthens confidence in the model's performance, ensuring that it can be used effectively for future flood risk assessments, scenario analysis, and water management planning within the Ottawa River watershed.

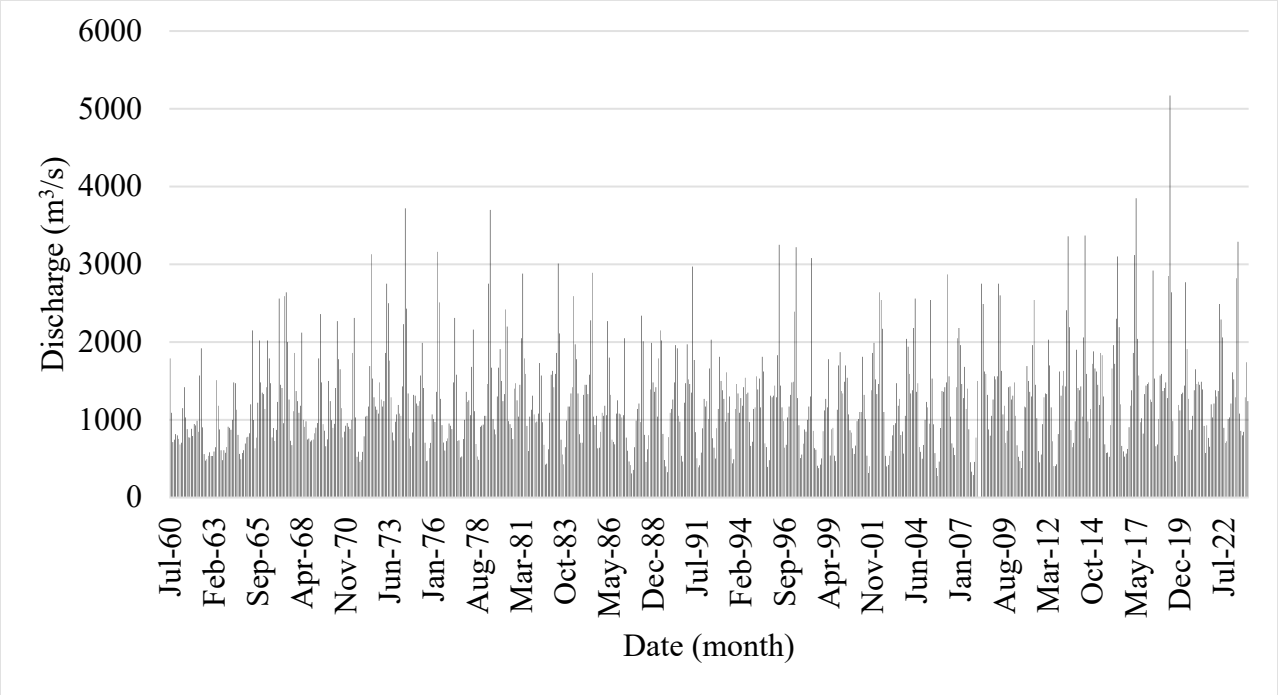


Figure 3.10 - Monthly discharge for flow at the hydrometric station at Britania (02KF005).

To define the boundary conditions for the simulation, a triangular flow hydrograph was developed using historical discharge data collected from Environment and Climate Change Canada's Britannia hydrometric station to represent each flood event. This dataset provided a reliable basis for establishing realistic flow conditions at the upstream boundary of the model domain. The simulation period was selected to capture a timeframe characterized by significant discharge variability, particularly during the snowmelt and spring thaw seasons, when flows are typically at their highest. The flow hydrograph included discharge values corresponding to return periods of 2 years, 10 years, 25 years, 50 years, 100 years, 1000 years, and 10,000 years, providing a wide range of hydrological conditions for analysis. To further refine the model's sensitivity to different flow rates, a discharge increment of 100 m^3/s was applied between successive simulations, resulting in a total of 51 individual model runs. This incremental approach allowed for detailed analysis of how the river's hydraulic behavior evolves under varying flow scenarios.

To define inflow boundary conditions for each flood event (2-, 10-, 25-, 50-, 100-, 1,000-, and 10,000-year), a triangular-shaped hydrograph was developed. This simplified hydrograph form is commonly used in unsteady hydraulic simulations where the goal is to analyze the system's response to a specified peak discharge over a generalized flood duration, rather than replicating a specific historical hydrograph. As described in the HEC-RAS 2D Modeling User's Manual, simplified synthetic hydrographs are suitable for design flood simulations, particularly when return-period-based discharge estimates are used and temporal rainfall-runoff distribution is not explicitly available (Brunner, 2024).

While this method does not account for detailed hydro-meteorological inputs (e.g., rainfall intensity, melt rate curves, or antecedent conditions), it provides a hydraulically defensible, temporally structured inflow scenario for assessing floodplain behavior under extreme discharge conditions. Furthermore, it allows for uniform comparison of model outputs across a range of return periods, isolating the effects of peak flow magnitude on hydraulic variables such as water surface elevation, flood extent, and flow velocity.

At the downstream boundary, the Normal Depth method was applied to define the WSEL at the model's exit point located at Cross Section 2040. The normal depth was calculated using the average channel bed slope, derived from surveyed cross-sectional data along the modeled reach. This approach provided a reasonable and physically based estimate of downstream water levels, consistent with the natural slope and flow conditions of the Ottawa River. In the context of unsteady flow modeling, a normal depth boundary condition is a computational assumption applied at the downstream boundary to define how water exits the model domain. It assumes that flow is uniform and steady near the boundary, such that the bed slope, water surface slope, and friction slope are approximately equal. HEC-RAS applies this boundary condition using

Manning's equation, with the specified slope and Manning's n value, allowing the model to dynamically compute the water surface elevation at each time step based on the instantaneous discharge.

Although unsteady flow includes time-variable acceleration, the normal depth condition remains valid when the flow near the boundary is gradually varied and subcritical, and the downstream hydraulic influence is minimal. According to the HEC-RAS 2D Modeling User's Manual, a normal depth boundary condition is commonly used for downstream boundaries of unsteady flow models, especially when the reach extends far enough downstream that the flow is expected to be close to uniform (Brunner, 2024). In this study, the downstream boundary condition of the modelled reach satisfies this criterion, as the downstream river station 2040 is located approximately 19 km from the study area at the town of Quyon with minimal variation in channel bed elevation. Therefore, the normal depth assumption provides a numerically stable, hydraulically defensible method for defining outflow behavior in the model.

3.5.6 – Establishment of 2D Flow Area at the town of Quyon:

To Establish the 2D Flow Area of the model, a perimeter was drawn located at the town of Quyon. The RAS Mapper 2D Flow Area Perimeter generation tool was utilized to generate a manual hand-drawn perimeter of the flow area ensuring the location of the boundary of the town as shown in Figure (3.11). For the 2D flow area in the model, a computational mesh was generated with a uniform spatial resolution of 50 m in both the x and y directions. Key computational parameters included a Cell Elevation, Face Profile, and Face Elevation-Area tolerance of 0.003, alongside a Face Conveyance Tolerance Ratio of 0.02 (2%). The Quyon 2D flow area consisted of 592 cells, with an average face length of 29 m.

Breaklines were incorporated within the Quyon 2D flow area to enforce cell faces along linear features such as high ground, ensuring that water flow was appropriately directed through the domain (Brunner, 2024). These breaklines were strategically placed to account for elevated terrain features within the domain, such as roads and areas of varying terrain, which helped redistribute water flow within lower-lying areas. Additionally, refinement regions were introduced in areas with lower-than-average elevation to enhance computational resolution by increasing the point density, allowing for more detailed hydraulic results (Brunner, 2024). This refinement was particularly important for identifying flood-prone areas and evaluating potential property loss risks.

Figure 3.11 - 2D flow area of the HEC-RAS model located at the town of Quyon.

The Quyon 2D flow area was initialized with dry conditions for each simulation run, and the external boundary of the 2D domain was treated as a Neumann boundary condition. Prior to

running hydraulic computations, Hydraulic Property Tables were generated for the Quyon 2D flow area using the Geometry Pre-processing tool in HEC-RAS. These tables included elevation-volume curves, which describe how water surface elevations are computed within each cell as water fills up, based on the underlying terrain data for each cell. Examples of these elevation-volume curves, along with cell face profiles, are provided in Figure (3.12).

Cell face profiles, which are essentially detailed cross-sections at each cell face, were also generated to represent the geometry governing flow conveyance between adjacent cells. Therefore, the number and spacing of cells and cell face profiles must be sufficient to accurately represent the profile of the terrain and account for changes in water surface slopes and velocity. If extensive changes in water surface slopes and velocities are expected, the computational point spacing must be reduced to account for those changes (Brunner, 2024).

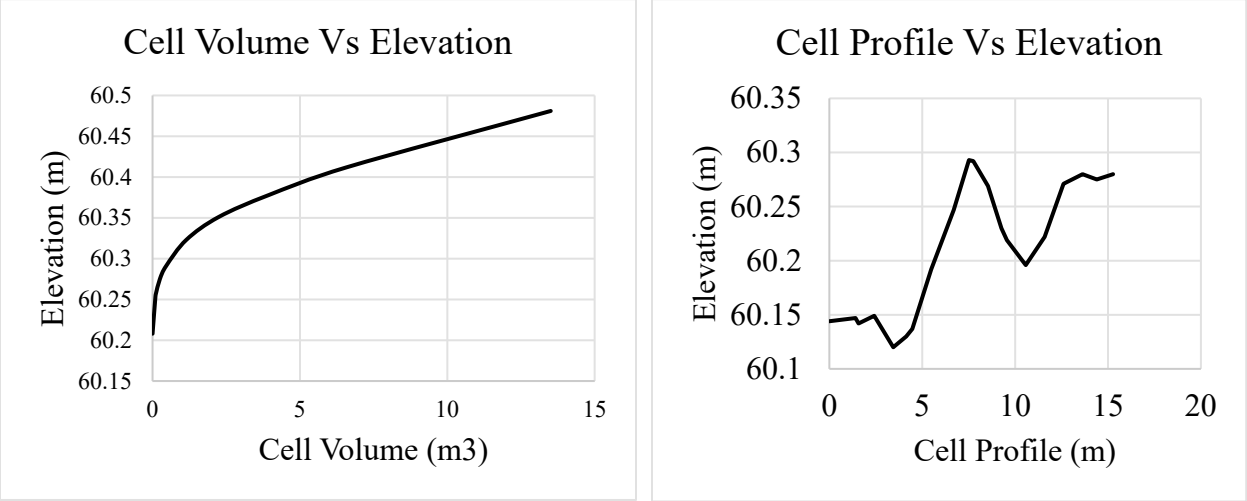


Figure 3.12 - 2D geometric preprocessor plots for cell volume & and profile versus elevation.

3.5.7 – Lateral Structure at the town of Quyon:

The 1D river channel was connected to the Quyon 2D flow area through a carefully positioned lateral structure, aligned along the natural high ground on the riverbank. This lateral

structure served as a hydraulic link between the 1D main channel and the adjacent 2D floodplain, allowing water to flow between the two modeling domains during overbank flow events. Specifically, the lateral structure at Quyon was placed between cross-sections 2054.5 and 2054, covering a section of the overbank area where flow was expected to spill from the main river channel into the adjacent floodplain.

Positioning the lateral structure in the overbank region, rather than directly within the active river channel, ensured that there was no spatial overlap between the 1D river geometry and the 2D flow area. This separation was critical for maintaining computational stability and ensuring that each modeling domain accurately represented the hydraulic conditions specific to its portion of the system. By defining the lateral structure in this way, the model could more effectively capture the transition of flow from the confined channel to the broader floodplain under high flow conditions.

The lateral structure was modeled as a broad-crested weir, allowing overbank flows to be simulated using the standard weir Equation (3.11). The broad-crested weir formulation provided flexibility in defining flow parameters along the structure, such as crest elevation and flow coefficients, which could be adjusted to reflect local topography and observed flow behavior. The weir was assigned a uniform width of 2 m along its length, which balanced computational efficiency with the need to capture variations in terrain and flow along the riverbank.

$$Q = CLH^{(3/2)} \tag{3.11}$$

where:

C is the weir coefficient

L is the length of the spillway crest

H is the upstream energy head

The weir coefficient was set at 0.25 since a weir coefficient of 0.5 is the maximum coefficient suggested in the HEC-RAS 2D Modeling User's Manual (Brunner, 2024). The lateral structure at Quyon was assigned a zero height to allow smooth water transfer from the overbank into the adjacent floodplain without any artificial obstructions. This configuration ensured that once water levels in the river channel exceeded the bank elevation, flow could seamlessly spill over into the 2D flow area, accurately representing natural overbank flooding conditions. The geometry of the Quyon lateral structure was defined using RAS Mapper, where the structure's centerline was carefully drawn along the left overbank of the river to align with the natural topographic high points along the bank. This placement allowed the lateral structure to accurately capture overbank flows without interfering with the main channel hydraulics. The elevation profile underlying the lateral structure was directly derived from the terrain data, ensuring that the weir profile along the lateral structure reflected the actual topography. This profile provided the necessary data for HEC-RAS to compute overbank flow based on the local ground elevations and water surface levels. The terrain elevation profile was also used to establish the weir crest elevations, which dictated when and where flow would spill into the Quyon 2D flow area. The plan view of the Quyon lateral structure is provided in Figure (3.13). The headwater position for the Quyon lateral structure was automatically set to the overbank region of the 1D river channel, ensuring that the upstream control point for flow exchange was appropriately located at the interface between the main channel and the floodplain. This headwater positioning was automatically determined using the default computed weir stationing within HEC-RAS, allowing the model to align the structure with the riverbank profile appropriately. All tailwater connections were set to direct flow into the Quyon 2D flow area, ensuring accurate representation of floodplain hydraulics.

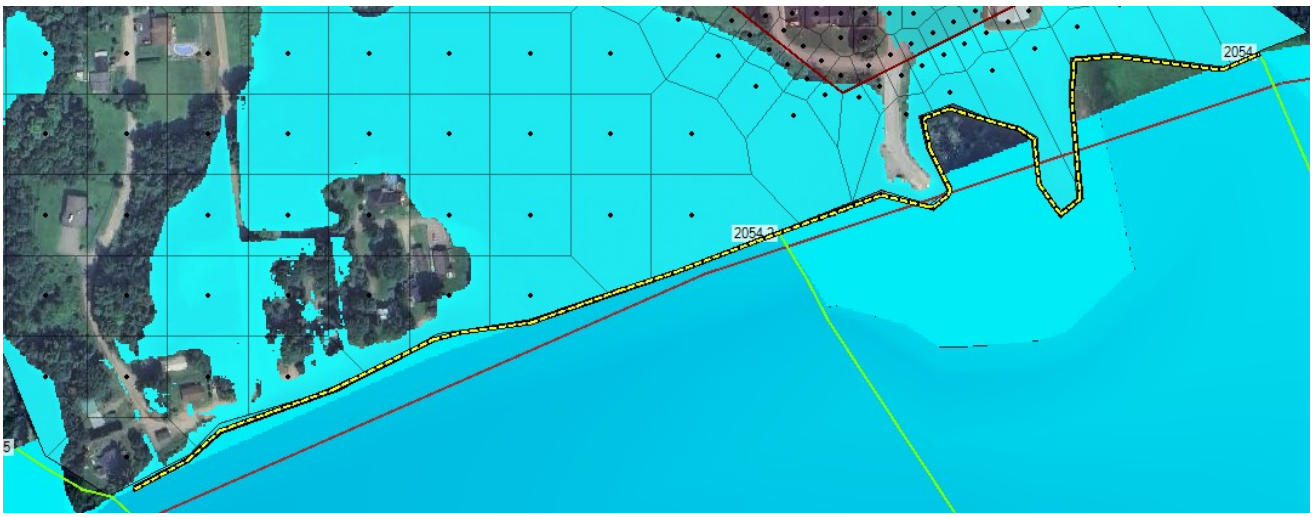


Figure 3.13 - Plan View of Lateral Structure located at town of Quyon.

3.5.8 – Unsteady Flow Analysis Equations and Computation Settings:

To ensure the HEC-RAS model produced accurate and reliable results, a comprehensive set of computation settings and solver options were specified to optimize model performance and output quality. The model underwent a rigorous process of iterative calibration, optimization, and re-running, with over 90 individual simulations conducted to achieve an optimal balance between computational accuracy, numerical stability, and reasonable processing times. The model analysis was carried out using a structured sequence of HEC-RAS tools, including the geometry preprocessor, the unsteady flow simulation engine, and the results post-processor. Each of these components played a key role in ensuring that both the physical processes and numerical approximations were appropriately captured across the modeled reach. Key computational parameters were configured to maintain numerical stability while capturing critical hydrodynamic processes in the model domain's 1D and 2D portions. The computational time step was set to 5 seconds, providing a sufficiently fine temporal resolution to capture rapid changes in flow conditions during high flow events. Hydrograph output intervals were set at 6 hours, allowing for a balanced level of temporal detail in the recorded flow series. Mapping output intervals were set

at 3 hours, which ensured the generation of sufficiently frequent spatial outputs to capture the evolving extent and depth of flood inundation across the floodplain. Additionally, detailed output intervals were set at 12 hours to preserve key hydraulic variables for in-depth review and post-analysis.

For the 1D unsteady flow options, the Theta parameter, also known as the implicit weighting factor, was set at 0.60, which falls within the recommended range of 0.60 to 1.0 specified by HEC-RAS. This parameter controls the balance between explicit and implicit numerical methods, contributing to model stability during unsteady flow simulations. The water surface calculation tolerance was set at 0.006 m, with the maximum allowable tolerance, as recommended by HEC-RAS, specified at 0.06 m. To further control numerical behavior, the maximum allowable error in the water surface elevation solution was set at 30 m which acts as a safety hard limit to prevent unstable or meaningless results, and the maximum number of iterations was limited to 20. If the model failed to converge within this limit, it would automatically abort the calculation for that time step.

For 1D/2D unsteady flow interactions, the model included 20,000 warm-up steps to allow sufficient time for the hydraulic conditions to stabilize before initiating the formal simulation period. This extended warm-up period ensured that initial transients did not distort the results during the primary simulation period. The maximum number of time slices per computational time step was set at 20, providing finer resolution in instances where rapid changes in flow or water surface elevation occurred. Lateral structure flow stability was assigned a value of 2, while inline structure flow stability was set at 1, aligning with HEC-RAS recommendations to improve computational robustness at these structures.

For the 2D flow options, the water surface elevation tolerance within the 2D flow areas was set at 0.003 m, and the flow tolerance was set at 0.1%. These tolerances ensured that the iterative solution process within the 2D floodplain converged accurately without excessive computational cycles. The shallow water Eulerian-Lagrangian method was employed to solve the flow equations in the 2D floodplain areas, while the diffusion wave equations were applied within the 1D river channel. This hybrid approach provided the necessary flexibility to handle differing hydraulic conditions between the main channel and floodplain.

The Boundary Condition Volume Check option was enabled for the 2D flow areas, allowing the model to track volume continuity at the external boundaries, which aided in ensuring mass balance throughout the simulation. The 1D mixed flow regime option was also activated, allowing the model to simulate both subcritical and supercritical flow conditions within the 1D river channel and its transitions to the 2D floodplain. To further enhance numerical stability in transitional flow conditions, the Froude number reduction factor exponent was set at 4, and the Froude number threshold for eliminating acceleration terms was set at 0.80.

In selecting the governing equations and physical processes to include in the unsteady flow analysis, a balance was struck between physical realism, computational efficiency, and relevance to the study domain. The diffusion wave approximation was applied to the 1D river channel to reduce computational demand while maintaining an adequate representation of varied flow conditions, which dominate in the Ottawa River under flood scenarios. In contrast, the 2D flow areas used the full shallow water equations (Saint-Venant equations) with the Eulerian-Lagrangian solution scheme, allowing the model to resolve local acceleration, pressure gradients, and dynamic flow behavior across complex terrain.

Certain physical processes, such as the Coriolis effect, were excluded from the model because of their negligible influence on the spatial and temporal scales of the study. The Coriolis force primarily affects large-scale geophysical flows, such as in oceanic or estuarine environments, and has minimal impact on flood routing in mid-sized river systems like the Ottawa River. Similarly, wind stress and groundwater interaction were omitted due to their limited influence on short-duration flood events and lack of spatially detailed input data. These modeling decisions ensured that computational effort was focused on simulating the dominant hydraulic processes including unsteady flow dynamics, floodplain conveyance, and surface resistance that most strongly influence flood behavior in the study area.

3.6 – Flood Frequency Analysis (FFA):

Flood Frequency Analysis (FFA) is a critical process in hydrology used to estimate the probability of floods of varying magnitudes at specific locations, often at hydrometric stations or river gauge sites. By analyzing historical peak flow data, FFA enables the estimation of recurrence intervals and exceedance probabilities for floods, forming the basis for flood risk management, infrastructure design, and floodplain mapping (Zaghloul et al., 2024). The central objective of FFA is to fit an appropriate statistical distribution to the observed flood data, allowing for extrapolation beyond the observed record to estimate the magnitude of rare floods, such as the 100-year or 1000-year flood (Chung et al., 2013). These estimates are essential for designing flood control structures, developing emergency preparedness plans, establishing flood insurance programs, and guiding sustainable land use planning in flood-prone areas (Blöschl et al., 2019). Given the often-limited length of available hydrometric records, combined with the rarity of extreme events, selecting and appropriate distribution and applying goodness-of-fit tests are key components of Flood Frequency Analysis (Haktanir, 1992). Modern FFA approaches also incorporate non-stationarity techniques

to account for changing climate and land use conditions, recognizing that flood frequency characteristics are increasingly influenced by anthropogenic changes (Perez et al., 2024). While several probability distributions are used in FFA, the Log-Pearson Type III (LP3) distribution remains one of North America's most widely applied methods, particularly in regulatory flood risk assessments and hydrologic design (United States Geological Survey, 2023).

The LP3 distribution is one of the most used probability distributions for flood frequency analysis (FFA), particularly in North America. It is recommended by several regulatory agencies, including the U.S. Geological Survey (USGS) and the Federal Emergency Management Agency (FEMA), for estimating design floods at gauged locations (United States Geological Survey, 2019). The LP3 distribution is particularly suited for modeling positively skewed flood data, where extreme floods are far larger than average events (United States Geological Survey, 2023). The method involves applying a logarithmic transformation to the observed annual peak flows, after which the transformed data are assumed to follow a LP3 distribution, which can accommodate skewness in the data (Griffis and Stedinger, 2007; Haktanir, 1992). This flexibility allows the LP3 to better fit the upper tail of flood distributions, which is critical when estimating rare and extreme flood events (Zaghloul et al., 2024).

In practical application, the method of moments is commonly used to estimate the mean, standard deviation, and skew coefficient of the logarithms of the observed peak flows (United States Geological Survey, 2023). The estimated parameters are then used to calculate design floods for various return periods, such as the 2-year, 10-year, 100-year, and 1000-year floods. One advantage of LP3 is its adaptability to a wide range of hydrologic conditions, from small mountainous catchments to large river basins, making it a robust choice for regional flood frequency analysis (Haktanir, 1992). LP3 remains a cornerstone distribution in FFA, particularly

when working with datasets from government hydrometric networks, and continues to be widely used in regulatory and academic flood risk assessments (United States Geological Survey, 2019). The probability density function of the LP3 distribution is given in Equation (3.12).

$$f(x) = \frac{1}{x|\beta|\Gamma(\alpha)}\left(\frac{ln(x)-\gamma}{\beta}\right)^{\alpha-1} exp\left(-\frac{ln(x)-\gamma}{\beta}\right) \quad (3.12)$$

where:

α is the shape parameter ($\alpha > 0$)

β is the scale parameter ($\beta \neq 0$)

γ is the location parameter

$ln(x)$ is the natural logarithm of the variable x

$\Gamma(\alpha)$ is the gamma function of the shape parameter α

The cumulative distribution function is expressed in Equation (3.13) as:

$$F(x) = \frac{\Gamma\left(\frac{ln(x)-\gamma}{\beta},\alpha\right)}{\Gamma(\alpha)} \quad (3.13)$$

where:

$\Gamma(\alpha)$ is the incomplete gamma function

$F(x)$ is the probability that the variable x does not exceed a given value.

The goodness-of-fit of the LP3 distribution used in this study was evaluated through the application of several statistical tests, including the Kolmogorov-Smirnov (KS) test, the Anderson-Darling (AD) test, and the Chi-Squared (CS) test. These tests were applied to quantitatively assess how well the LP3 distribution fits the observed hydrological dataset, ensuring the selected distribution provides an appropriate and reliable representation of the underlying flood frequency characteristics.

3.6.1 – Kolmogorov–Smirnov (KS) Test:

The KS test is a non-parametric test used to assess the goodness-of-fit between the empirical distribution function (EDF) of observed data and a specified theoretical distribution, such as the LP3. It measures the maximum absolute difference between the observed and theoretical cumulative distribution functions, with smaller differences indicating better fit (Massey, 1951; Evans et al., 2008). The KS test is widely used in flood frequency analysis due to its flexibility, as it does not require the data to follow a normal distribution. However, the test is more sensitive to the median and less effective at detecting differences in the tails, which can be a limitation when modeling extreme flood events (Razali and Yap, 2011, Karson, 1968, Samantaray and Sahoo, 2020). Despite these limitations, it remains an effective initial screening tool for evaluating distributional fit in hydrologic datasets (Chen et al., 2017).

The KS test evaluates how closely the observed dataset aligns with a specified theoretical distribution by calculating the largest vertical difference between their cumulative distribution functions (CDFs). The empirical CDF, which represents the cumulative probability derived directly from the sample data, is expressed in Equation (3.14).

$$F_n(x) = \frac{1}{n}[Number\ of\ observations \leq x] \tag{3.14}$$

where:

n is the sample size

The Kolmogorov–Smirnov statistic D is defined in Equation (3.15) as:

$$D = \max_{1 \leq i \leq n} |F(x_i) - F_n(x_i)| \tag{3.15}$$

where:

$F(x_i)$ represents the theoretical CDF of the hypothesized distribution

$F_n(x_i)$ denotes the empirical CDF based on the sample data

3.6.2 – Anderson-Darling (AD) Test:

The AD test is a modification of the Cramér-von Mises test that measures the goodness-of-fit between an empirical CDF and a specified theoretical distribution. The AD test assigns greater weight to the tails of the distribution, making it particularly valuable in flood frequency analysis, where extreme flood events are of primary concern (Stephens, 1974, Evans et al., 2008, Razali and Yap, 2011). This focus on distribution tails improves its sensitivity to deviations in low-probability, high-magnitude events, which are crucial for flood risk estimation and infrastructure design (Samantaray and Sahoo, 2020, Chen et al., 2017). The test statistics, computed as a weighted sum of squared differences between the observed and theoretical CDFs, provides a quantitative measure of fit, with smaller values indicating better alignment (Ghaderi et al., 2019). Due to its robustness across sample sizes and distributions, the AD test is widely used in hydrology and environmental risk assessments (Ghaderi et al., 2019).

The AD test compares the observed data to the Log-Pearson Type III distribution, emphasizing the tails to ensure the suitability of the distribution for modeling extreme discharge values and flood events. The AD test is defined in Equation (3.16).

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}[(2i-1)\,ln(\,F(X_i) + ln(\,1 - F(X_{n-i+1}))] \tag{3.16}$$

where:

$F(x_i)$ represents the theoretical CDF of the distribution

A^2 is the Anderson-Darling Statistic

n is the sample size

3.6.3 – Chi-Squared (CS) Test:

The CS test is a widely used goodness-of-fit test that evaluates how well observed frequency data matches the expected frequency data derived from a theoretical distribution. The test works by dividing the data range into discrete intervals (bins) and comparing the observed count of data points in each bin to the expected count predicted by the fitted distribution (Greenwood and Nikulin, 1996, Razali and Yap, 2011). The test statistic, calculated as the sum of the squared differences between observed and expected frequencies divided by the expected frequency, follows a CS distribution with degrees of freedom dependent on the number of bins and estimated parameters (McHugh, 2013, Chen et al., 2017). The CS test is particularly useful for identifying large-scale deviations between the data and the theoretical model, though it is generally less sensitive to differences in the tails of the distribution compared to the Anderson-Darling test (Ghaderi et al., 2019, Samantaray and Sahoo, 2020). Despite its limitations, it remains a simple and interpretable test often used in hydrological studies to support the evaluation of flood frequency distributions (Ghaderi et al., 2019).

This test is ideal for validating the distribution when data are grouped into intervals, such as discharge ranges, and provides a measure of how well the LP3 distribution represents the observed data. The Chi-Squared test is defined in Equation (3.17).

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{3.17}$$

where:

χ^2 is the Chi-Squared statistic

O_i denotes the observed frequency for bin i

E_i is the expected frequency for bin i

E is calculated in Equation (3.18) as:

$$E_i = F(x_2) - F(x_1) \tag{3.18}$$

where:

F is the CDF of the hypothesized distribution

x_1 & x_2 are the limits for bin i

3.7 – New-Generation Group Method of Data Handling (GMDH) Model:

3.7.1 – New-Gen GMDH Model Structure and Equations:

Pioneered by Ivakhnenko (1978), the GMDH is a robust computational approach designed to model complex nonlinear systems with highly interdependent input variables. This advanced technique systematically constructs neural layers by iteratively pairing two input neurons from the preceding layer and linking them through quadratic polynomial functions (Safari et al., 2019). Over the past few decades, GMDH has been extensively validated across multiple disciplines for its ability to approximate both linear and nonlinear phenomena (Walton et al., 2019) even when the underlying system dynamics remain partially or entirely unknown. Extensive research has reinforced its effectiveness, highlighting its reliability and precision across diverse applications.

GMDH exhibits behaviors of a self-organizing algorithm (Dodangeh et al., 2020), it does not rely on prior information fed into the algorithm. GMDH enables computers to identify the most suitable model based on predefined criteria autonomously, reducing the need for extensive human intervention. By incorporating a statistical dataset and specifying selection metrics, this method automates the identification of optimal model structures. The training phase plays a crucial role in dynamically determining the ideal number of neurons and layers, thereby enhancing model

efficiency. Additionally, GMDH employs principles of natural selection to optimize network size, complexity, and predictive accuracy.

GMDH systematically evaluates all possible input variable combinations to identify strong correlations while eliminating less significant parameters. Only neurons that satisfy predefined performance thresholds progress to subsequent layers, whereas underperforming ones are discarded (Ebtehaj et al., 2015). This inductive learning approach autonomously optimizes the network architecture, minimizing the need for manual adjustments. The fundamental objective is to derive a mathematical function that accurately maps input variables to their corresponding output values.

The classical GMDH model utilizes a distinctive approach by constructing a model as a network of interconnected neurons. Each neuron in a layer is formed by pairing different neurons from the previous layer and connecting them using a quadratic polynomial. This iterative process generates new neurons in subsequent layers, allowing the model to capture complex relationships between inputs and outputs. The flexible structure of GMDH enables it to effectively map inputs to one or more outputs, making it a powerful tool for modeling nonlinear dynamics in various prediction and analysis tasks (Walton et al., 2019). The GMDH algorithm is designed to minimize the computational error between the observed output variable y_i and the predicted output $\hat{y}_i$. It accomplishes this by constructing a model $\hat{f}_i$ that maps the input variables $X = (x_1, x_2, x_3, \dots, x_n)$, where n is the number of inputs is to the predicted output $\hat{y}_i$. Utilizing its self-organizing capability, GMDH builds a network of interconnected neurons, each representing a second-order polynomial function that combines different input variables. By optimizing both the structure and parameters of this network, GMDH aims to develop a model $\hat{f}_i$ that closely approximates the actual output y_i based on the given inputs $(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$. As such, GMDH operates by drastically

reducing the computational error between the recorded observed and predicted modelled output, resulting in an enhanced modeling approach, as shown in Figure (3.14) (Letessier et al., 2023).

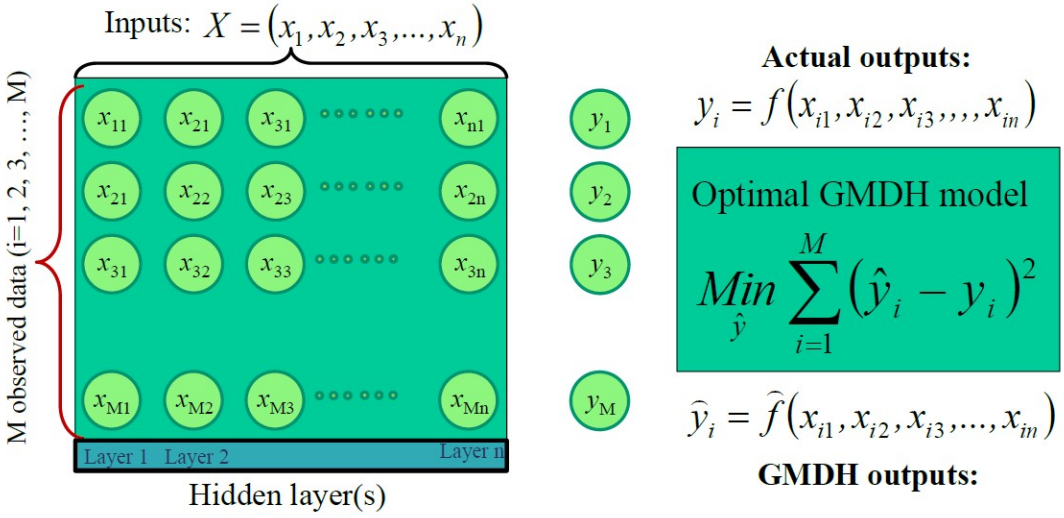


Figure 3.14 - The schematic structure of GMDH for a provided set of input data (Letessier et al., 2023).

By determining an optimal function, the GMDH-type neural network seeks to minimize the squared difference between the actual output y_i and the predicted output $\hat{y}_i$. This is represented in Equation (3.19).

$$OF = \underset{\hat{y}}{\text{Min}} \sum_{i=1}^{M} (\hat{y}_i - y_i)^2 \tag{3.19}$$

where:

OF is the objective function.

M is the total number of data samples.

$y_i, \hat{y}_i$ are the ith samples of the actual and predicted output.

The relationship between input variables and the predicted output variable can be represented using a complex discrete form of the Volterra functional series, commonly referred to

as the Kolmogorov-Gabor polynomial (Letessier et al., 2023). This polynomial is expressed in Equation (3.20) as:

$$\hat{y} = q_0 + \sum_{i=1}^{M} q_i x_i + \sum_{i=1}^{M} \sum_{j=1}^{M} q_{ij} x_i x_j + \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{M} q_{ijk} x_i x_j x_k + \cdots \quad (3.20)$$

where:

$\hat{y}_i$ is the predicted output value.

$(q_0 q_i q_{ij} q_{ijk})$ are unknown coefficients.

$(x_i,\ x_j, and\ x_k)$ are the ith, jth, and kth input variables (respectively).

M is the total number of data samples

The Kolmogorov-Gabor polynomial serves as a flexible and expressive representation of the nonlinear relationships between input and output variables, capturing intricate patterns and interactions within the data. Each term in the polynomial reflects the contribution of different combinations of input variables, while the unknown coefficients $q_0, q_i, q_{ij}, q_{ijk}, \ldots$ determine the impact of each term on the overall relationship. In this context, the input vector $X = (x_1, x_2, \ldots, x_n)$ represents a collection of input variables, and the unknown coefficients q_i correspond to weight vectors or polynomial coefficients that influence the interaction between inputs and the output variable. By adjusting these coefficients, the model can effectively learn complex dependencies, enabling accurate modeling and prediction. A simplified mathematical description of this polynomial can be expressed as a partial quadratic polynomial system involving only two variables (neurons). This reduced form is given in Equation (3.21) as:

$$\hat{y} = G(x_i, x_j) = q_1 + q_2 x_i + q_3 x_j + q_4 x_i^2 + q_5 x_j^2 + q_6 x_i x_j \quad (3.21)$$

where:

$x_i \; and \; y_i$ represent the input variables.

$q_0, q_1, q_2, q_3, q_4 \; and \; q_5$ are the unknown coefficients or weights associated with the polynomial terms.

This quadratic representation forms the basis for GMDH network neurons, allowing the model to systematically construct higher-order polynomials and capture nonlinear dependencies in the data. Each term in this equation represents the influence of different combinations of input variables and their quadratic forms. By adjusting the coefficients, the model can effectively capture complex interactions and patterns among the input variables, enhancing its ability to model and predict outcomes accurately. In the fundamental structure of the GMDH algorithm, all possible pairs of two separate variables from a total of n input variables $(x_i, i = 1, 2, 3, \dots, M)$ are considered to construct a regression polynomial in the form of quadratic polynomials with only two variables that best fit the dependent observations $(y_i, i = 1, 2, 3, \dots, M)$ in a least-squares sense (Letessier et al., 2023). The quantity of neurons in the first subsequent hidden layer is calculated by the number of these combinations, which is given by $(n, 2) = n(n-1)/2$, (where n is the number of input variables). As such, $n(n-1)/2$ neurons are built in the first hidden layer based on the observed data. Each neuron in the hidden layer corresponds to a unique combination of two input variables. By systematically evaluating all possible pairs, the GMDH algorithm constructs a model that effectively captures the underlying relationships and patterns in the data. To ensure a precise and reliable representation, the algorithm employs the least-squares method to optimize the fit between the model's predictions and the actual observations (Letessier et al., 2023). Each neuron represents a combination of variables (x_{ia}, x_{in}) from the input vector $\{(y_i, x_{ia}, x_{ib}) ; (i =$

$1,2,3,\dots,M)(a,b\epsilon\{1,2,3,\dots,n\})\}$. The model is defined as follows in Equation (3.22) (Letessier et al., 2023).

$$\begin{cases} y_1 = q_0 + q_1x_{1a} + q_2x_{1b} + q_3x_{1a}x_{1b} + q_4x_{1a}^2 + q_5x_{1b}^2 \\ y_2 = q_0 + q_1x_{2a} + q_2x_{2b} + q_3x_{2a}x_{2b} + q_4x_{2a}^2 + q_5x_{2b}^2 \\ ---------- \\ y_M = q_0 + q_1x_{Ma} + q_2x_{Mb} + q_3x_{Ma}x_{Mb} + q_4x_{Ma}^2 + q_5x_{Mb}^2 \end{cases} \quad (3.22)$$

where:

y_i represents the observed output variable for the ith observation.

$Q = \{q_0, q_1, q_2, q_3, q_4, q_5\}$ is the vector of polynomial coefficients.

The y values can be represented by the equation $y = Aa$. The matrix A represents the input variables as shown in Equaiton (3.23).

$$A = \begin{bmatrix} 1 & x_{1a} & x_{1b} & x_{1a}x_{1b} & x_{1a}^2 & x_{1b}^2 \\ 1 & x_{2a} & x_{2b} & x_{2a}x_{2b} & x_{2a}^2 & x_{2b}^2 \\ . & . & . & .. & . & . \\ . & . & . & .. & . & . \\ 1 & x_{Ma} & x_{Mb} & x_{Ma}x_{Mb} & x_{Ma}^2 & x_{Mb}^2 \end{bmatrix} \quad (3.23)$$

where:

y is the vector of observed output variables $y = [y_1, y_2, y_3, \dots, y_M]$.

The goal of the algorithm is to find the optimal coefficients $Q = \{q_0, q_1, q_2, q_3, q_4, q_5\}$ that minimizes the difference between the predicted and observed output values based on a set of criteria that are specified in the GMDH model (Letessier et al., 2023).

As specified by Letessier at al., (2023), the classical GMDH model has several limitations when applied to daily river flow predictions, including: (i) its restriction to a 2nd-order polynomial

structure, (ii) the limitation of using only two input variables per neuron, and (iii) the restriction of neurons being formed only from adjacent layers. To address these issues, a new approach is introduced to evaluate the impact of more complex models on improving the predictive performance of GMDH. This modified GMDH model enables the generation of 2nd and 3rd-order polynomials with two and three inputs, respectively, within each polynomial. As a result, three additional polynomial forms, beyond the previously mentioned 2nd-order polynomial in Equation (3.21) are defined in Equation (3.24, 3.25, and 3.26), as follows to incorporate the additional terms:

$$\hat{y} = G(x_i, x_j, x_k) = q_0 + q_1 x_i + q_2 x_j + q_3 x_k + q_4 x_i^2 + q_5 x_j^2 + q_6 x_k^2 + q_7 x_i x_j + q_8 x_1 x_k + q_9 x_j x_k \quad (3.24)$$

$$\hat{y} = G(x_i x_j) = q_0 + q_1 x_i + q_2 x_j + q_3 x_i^2 + q_4 x_j^2 + q_5 x_i x_j + q_6 x_i x_j^2 + q_7 x_i^2 x_j + q_8 x_i^3 + q_9 x_j^3 \quad (3.25)$$

$$\hat{y} = G(x_i, x_j, x_k) = q_0 + q_1 x_1 + q_2 x_2 + q_3 x_3 + q_4 x_1^2 + q_5 x_2^2 + q_6 x_3^2 + q_7 x_1 x_2 + q_8 x_1 x_3 + q_9 x_2 x_3 + q_{10} x_1^3 + q_{11} x_2^3 + q_{12} x_3^3 + q_{13} x_1^2 x_2 + q_{14} x_1^2 x_3 + q_{15} x_2^2 x_1 + q_{16} x_2^2 x_3 + q_{17} x_3^2 x_1 + q_{18} x_3^2 x_2 + q_{19} x_1 x_2 x_3 \quad (3.26)$$

From these equations, it is clear that both the 2nd-order polynomial with three inputs (Equation (3.24)) and the 3rd-order polynomial with two inputs (Equation (3.25)) each involve ten unknown weight coefficients. However, the 3rd-order polynomial with three inputs (Equation (3.26)) contains twenty unknown weight coefficients. This new approach of GMDH is referred to as the New-Generation of GMDH (Qureshi et al., 2025), which can dynamically adjust and optimize based on available data or changing input patterns. This adaptability allows the model to effectively capture and represent complex relationships and patterns within the data. By incorporating an adaptive structure into the GMDH method, the model enhances its ability to

handle diverse datasets, optimize performance, and improve predictive accuracy (Qureshi et al., 2025). One notable improvement in the New-Gen GMDH is the automatic selection of input variables. Not all input variables are necessarily included in the final model, which is composed of one or more polynomials. Essentially, this method autonomously performs feature selection during the process of generating optimal polynomials. This ability to select relevant features is valuable because it enables the model to identify and incorporate only the most significant input variables, reducing dimensionality and potentially improving predictive performance. By automatically selecting the inputs that contribute most to the model's accuracy, the New-Gen GMDH optimizes the representation of the underlying relationships in the dataset (Qureshi et al., 2025).

3.7.2 – New-Gen GMDH Model Development:

In our model, input variables were selected from climate station data obtained from the Climate Station Ottawa CDA, located at 45°23′N and 75°43′W, with an elevation of 79.20 m and a climate ID of 6105976. The dataset was sourced from the Government of Canada's website, comprising monthly records spanning from July 1960 to August 2024. The model is written in a structured approach that enhances the input dataset before it is processed and run through the New-Gen GMDH equations. The first scenario (SC1) provides five different input variables to the model, organizing them as input and output variables. The second scenario (SC2) introduces lags in the model by incorporating the autocorrelation function (ACF) to find trends and dependencies in the input data. The third Scenario (SC3) introduces three further input variables, that enhance the results and provide more training options to the algorithm. All three scenarios also include Sensitivity Analysis options to test the impact of each variable in comparison to the others.

In the first stage of the model, designated as Scenario 1 (SC1), the input variables incorporated into the ML model included the average maximum temperature, average minimum temperature, average mean temperature, and average total precipitation. This dataset was used to train the model against the output variable, daily mean discharge, obtained from the Government of Canada's Hydrometric Station 02KF005, located in Britannia, Ontario, at coordinates 45°21′04″N, 75°49′35″W. The dataset was validated using the Log-Pearson Type III distribution through the previously mentioned FFA technique. Different model runs were conducted for sensitivity analysis by systematically varying input combinations, each excluding one variable to assess its impact on model performance. The best output from these were incorporated into the results.

In the second stage of the model, designated as Scenario 2 (SC2), an ACF was applied in MATLAB to analyze the influence of historical values on predicted values within the previously incorporated monthly time-series data. This analysis generated multiple time lags, which aided in identifying patterns, seasonal variations, and temporal dependencies within the dataset. The autocorrelation function plot is presented in Figure (3.15). The dataset exhibited strong autocorrelation at lags of 1, 2, and 12, with ACF values exceeding 0.50, indicating significant temporal dependencies. To account for these dependencies in the input and output functions, missing values (NaNs) were strategically introduced within the algorithm.

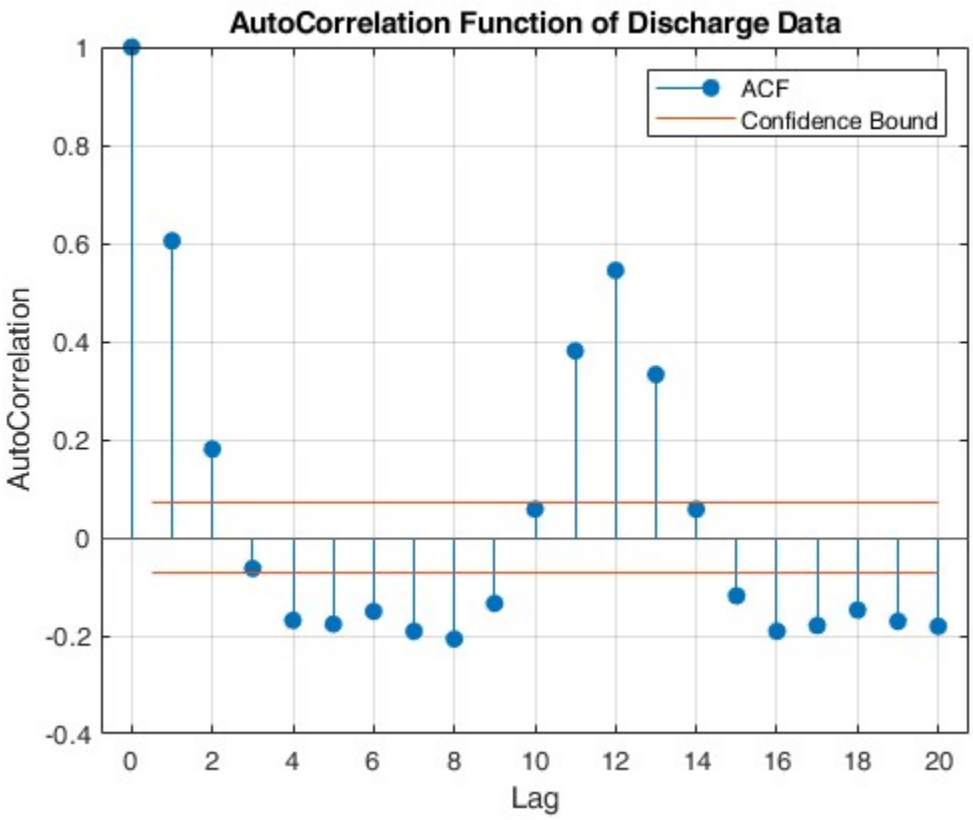


Figure 3.15 - The plot for the autocorrelation function in the dataset showing lag1, lag2 and lag12 demonstrating high ACF values.

These time lags were integrated into the model to establish a sensitivity analysis scenario, enabling the evaluation of model behavior under different lag configurations for comparative analysis. Each lag variable systematically shifted the dataset by a predefined number of steps, corresponding to the activated scenario, to assess its impact on predictive performance with the best results chosen.

Scenario 3 (SC3) of the model introduced the average of maximum, minimum, and mean discharge values into the input dataset. These discharge values were obtained from the HEC-RAS numerical model and supplemented with observed discharge data from the Hydrometric Station at Britannia. The inclusion of these variables expanded the input dataset to a total of seven different input variables, allowing for a more comprehensive assessment of the model's predictive

capabilities. A sensitivity analysis was conducted by testing different input combinations, systematically incorporating one or more of the newly introduced discharge variables to evaluate their impact on model performance.

To systematically assess the impact of different meteorological and hydrological variables on discharge prediction, the three modeling scenarios (SC1, SC2, and SC3) were formulated based on domain knowledge and statistical analysis of input behavior. All scenarios utilized core meteorological inputs such as maximum temperature, minimum temperature, mean temperature, and total precipitation due to their direct influence on snowmelt processes, runoff generation, and seasonal hydrological behavior. SC2 extended this input set by introducing lagged variables (lag1, lag2, lag12) based on high autocorrelation values identified in the data, capturing short- and medium-term dependencies in the hydrologic response. This structure enabled the New-Gen GMDH model to learn temporal patterns linked to consecutive weather events or cumulative climatic effects. SC3 integrated discharge-based statistics such as average minimum, mean, and maximum discharge to test whether historical flow characteristics improve prediction.

The three scenarios in the model were designed to assess the impact of different input variables on predictive performance. SC1 utilized climate variables, including average maximum, minimum, and mean temperatures, along with total precipitation, to model daily mean discharge. SC2 introduced autocorrelation analysis to incorporate time lags, identifying strong temporal dependencies at lags of 1, 2, and 12, which were integrated into the model for sensitivity analysis. SC3 expanded the input dataset by incorporating discharge values—average maximum, minimum, and mean—extracted from the HEC-RAS numerical model and observed data from the Britannia hydrometric station. Each scenario included a sensitivity analysis to evaluate model behavior under different input combinations, ensuring a comprehensive assessment of its predictive capability.

After setting the input variables and providing a detailed set of varying scenarios due to the incorporation of SC1, SC2, and SC3, the next step of the model was to break down the dataset into training and testing data. The data was allocated by incorporating 30% of the data into the testing phase and 70% of the data into the training phase. The training phase of the New-Gen GMDH network focused on configuring and optimizing the model to minimize the objective function, which is defined as the squared difference between the actual and predicted outputs. This objective function quantifies the accuracy of the model's approximated mapping of input variables to the output, ensuring optimal performance.

The modelling phase utilizes a non-linear regression approach with varying set of input variables and functions that are used to build the model. The input parameters of the model are highlighted in Table (3.2) with their descriptions. The New-Gen GMDH model uses several key criteria to control the selection and evaluation of neurons during the training process. The maximum number of inputs per neuron defines how many input variables a neuron can use when forming its polynomial relationships. The input selection strategy determines whether neurons can take inputs from only the previous layer or the original input variables, influencing the model's flexibility. The maximum number of neurons per layer restricts the model's complexity by limiting how many neurons can be created at each stage. Additionally, the reduction rate of neurons per layer controls how the number of neurons decreases as the model deepens. The polynomial degree sets whether the model includes quadratic or cubic terms, impacting the nonlinearity captured in the data relationships. A selection criterion is used to evaluate the quality of neurons, which determines how neurons are retained or discarded based on performance metrics. Finally, a threshold for improvement dictates when new layers are added, ensuring that additional complexity is only introduced if it leads to better model performance. The validation dataset

provides an external check to prevent overfitting, ensuring the model generalizes well beyond the training data.

Table 3.2 - New-Gen GMDH Model Configuration Parameters.

Variable Name	Description
Xtr	Matrix of input variables for training.
Ytr	Vector of corresponding output values.
maxNumInputs	Maximum number of inputs per neuron. Neurons can consider both 2-input and 3-input polynomials.
inputsMore	Neurons can take inputs from both previous layers and original input variables.
maxNumNeurons	Maximum number of neurons per layer.
decNumNeurons	Reduction in the number of neurons per layer (stays at a fixed number of neurons until convergence).
P	Polynomial degree for neurons, 3 (implying cubic terms are included).
critNum	Evaluation criterion is regularity criterion using validation data (Xts, Yts) and training data.
Delta	New layers are added if the evaluation criterion improves.
Xts	Matrix of input variables for validation/testing.
Yts	Vector of corresponding output values for validation/testing.

3.8 – Evaluation Metrics for Predictive Model Performance of GMDH:

When evaluating the performance of a predictive model, various statistical indices are used to assess its accuracy, reliability, and overall effectiveness. These metrics help determine how well the model captures patterns in the data, measures the magnitude of errors, and avoids overfitting

or underfitting. Some indices, such as the Coefficient of Determination (R^2) and Variance Accounted For (VAF), focus on explaining the proportion of variance captured by the model. Others, like Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), quantify the magnitude of prediction errors. Additionally, relative error measures such as Scatter Index (SI), Mean Absolute Relative Error (MARE), and Root Mean Square Relative Error (RMSRE) provide insight into the model's performance regardless of scale. More advanced criteria, including the Akaike Information Criterion (AIC) and Nash-Sutcliffe Efficiency (NSE), evaluate model complexity and efficiency, particularly in hydrological applications. Below is a breakdown of the key statistical metrics used to assess model performance.

The R^2 measures how well the model explains variability in the data, it is expressed in Equation (3.27) with a range of 0 (poor variability) and 1 (perfect fit).

$$R^2 = 1 - \frac{\sum(Y_{actual} - Y_{predicted})^2}{\sum(Y_{actual} - \bar{Y})^2} \tag{3.27}$$

The VAF indicated how much variance in the data the model can capture as expressed in Equation (3.28). A 0% range indicates poor model performance while a 100% indicates a perfect model.

$$VAF = \left(1 - \frac{Var(Y_{actual} - Y_{predicted})}{Var(Y_{actual})}\right) \times 100 \tag{3.28}$$

The RMSE measures the average error magnitude, giving more weight to large errors. It is expressed in Equation (3.29), a lower RMSE value is desired for better model performance.

$$RMSE = \sqrt{\frac{1}{n}\sum\left(Y_{actual} - Y_{predicted}\right)^2} \tag{3.29}$$

The SI measures the relative error in the model based on the standard deviation of actual values. This is expressed in Equation (3.30). A lower SI value indicates higher model accuracy.

$$SI = \frac{RMSE}{\bar{Y}} \tag{3.30}$$

The MAE measures the absolute difference between actual ad predicted values as expressed in Equation (3.31). A lower MAE indicates accurate model performance.

$$MAE = \frac{1}{n}\sum\left|Y_{actual} - Y_{predicted}\right| \tag{3.31}$$

The MARE measures the relative prediction error in the model. It is expressed in Equation (3.32) with a lower MARE value indicating better accuracy in the modelling technique.

$$MARE = \frac{1}{n}\sum\left|\frac{Y_{actual} - Y_{predicted}}{Y_{predicted}}\right| \tag{3.32}$$

The RMSRE is similar to the RMSE discussed above but is relative to the actual values used in the model. It is expressed in Equation (3.33).

$$RMSRE = \sqrt{\frac{1}{n}\sum\left(\frac{Y_{actual} - Y_{predicted}}{Y_{actual}}\right)^2} \tag{3.33}$$

The AIC is used to evaluate the complexity of a model versus its accuracy. A lower AIC value indicates better model performance. This metric is commonly used in model selection to reduce overfitting. The NSE measures model efficiency for hydrological predictions expressed in Equation (3.34).

$$NSE = 1 - \frac{\sum(Y_{actual} - Y_{predicted})^2}{\sum(Y_{actual} - \bar{Y})^2} \tag{3.34}$$

The range of NSE is from -1 to 1, with an NSE of 1 being the perfect value for the index. The condition of an NSE of 1 is achieved when the estimation error variance is zero.

These statistical metrics collectively provide a comprehensive evaluation of the GMDH model's performance. RMSE and MAE quantify the absolute and directional errors, offering insights into prediction accuracy and bias. R^2, NSE, and VAF assess the model's explanatory power and efficiency, indicating how well it captures variations in observed data. SI, MARE, and RMSRE normalize errors relative to observed values, making them useful for comparing different datasets. Finally, AIC provides a balance between model complexity and goodness of fit, aiding in model selection. Together, these metrics enable a robust assessment of the GMDH model's predictive reliability and effectiveness.

Chapter 4

Results

4.1 – FFA-Based Results with Discharge-Return Period Relationship for Model Simulations:

As published in our research paper (Qureshi et al., 2025), The LP3 distribution for the historical dataset was characterized by the following parameters: shape parameter (α) = 2378.80, scale parameter (β) = 0.0053, and location parameter (γ) = 20.58. To assess the goodness-of-fit of this distribution, three statistical tests—KS, AD, and CS—were conducted at significance levels of 0.2, 0.1, 0.05, 0.02, and 0.01 (Qureshi et al., 2025).

The KS test yielded a test statistic of 0.044, which was lower than the critical values at all significance levels. As a result, the null hypothesis that the data follow the LP3 distribution could not be rejected at any significance level, indicating a good overall fit. Similarly, the AD test produced a test statistic of 0.14, also below the critical values across all significance levels. Given that the AD test places greater emphasis on distribution's tails, these results confirm that the LP3 distribution effectively captures extreme values, such as high-magnitude flood events. The CS test returned a test statistic of 3.08, significantly lower than the critical values at all significance levels, suggesting that the observed frequency distribution closely aligns with the expected frequencies derived from the LP3 model. Overall, the findings from all three tests consistently validate the suitability of the LP3 distribution for modeling the given historical dataset (Qureshi et al., 2025).

After validating the adequacy of the fitted LP3 distribution using the KS, AD, and CS tests, discharge values corresponding to various return periods were computed (Qureshi et al., 2025). The development of the coupled HEC-RAS 1D/2D model required the assignment of flow values

at the upstream boundary to simulate flood conditions of varying magnitudes. These flows, shown in Table (4.1), represent peak discharge values associated with different return periods, ranging from frequent, lower magnitude floods (2-year and 10-year events) to rare, high-magnitude floods (1000-year and 10,000-year events). Each return period corresponds to the likelihood of a specific flood occurring within any given year, with longer return periods indicating less frequent but more severe floods.

Table 4.1 - Discharge-Return Period Relationship

Return Period (Year)	**Q (m^3/s)**
2	3000
10	4500
25	5000
50	5500
100	6000
1000	7000
10,000	8000

These discharge values were applied at River Station 2058, the upstream boundary of the model domain. By simulating a range of flows associated with different return periods, the model was able to capture the hydraulic response of the river and floodplain under various flood scenarios. This ensured that both routine flood events and extreme, low-probability events were included in the analysis.

The simulated flows were used to assess the extent of overbank flooding, floodplain inundation patterns, and hydraulic connectivity between the river channel and adjacent low-lying areas. This information supports the identification of areas most vulnerable to flooding and helps

evaluate how water moves across the floodplain under different flow conditions. By incorporating a wide range of discharge scenarios, the model was able to represent both typical seasonal flooding and extreme flood events, providing a comprehensive understanding of the river system's behavior across different hydrological conditions.

4.2 – 1D HEC-RAS Modelling Results:

Using the HEC-RAS software, a range of 1D results were produced based on the model development process and the input parameters discussed in the previous section. 1D results of the HEC-RAS software present graphical representations of cross-sections for different flow scenarios, illustrating how water depths and flow patterns change under various discharge conditions. Each cross-section plot displays the riverbed profile, bank station locations, and the maximum water surface elevations associated with flood events of different return periods. The model was run for different flood events with increments of 100 m^3/s at the Upstream boundary condition. The first model run had an initial discharge value of 3,000 m^3/s, representing a 2-year flood event based on the fitted LP3 distribution, which increased up to a discharge value of 10,000 m^3/s, representing a 10,000-year flood event. With these increments in mind, the model was run over 65 times with different discharge scenarios to provide a deep understanding of the hydrodynamic behavior of the river channel. Results from each model run were compared to the corresponding stage-flow hydrograph provided at the Hydrometric station 02KF005.

The focus of the study was on cross-sections 2054.5, 2054.2 & 2054 which were located adjacent to the town of Quyon. These cross-sections exhibited steep banks on the right overbank due to a drastic difference in elevation between the channel overbank and surrounding floodplain with a difference in elevation observed to be 10 m in a cross-stream distance of just 100 m. As

such, most of the study was focused on the left overbank which was connected directly to the 2D floodplain in the town of Quyon through a lateral structure discussed in the previous section.

Figure (4.1) represents cross-section 2054.5 with an initial model discharge value of 3000 m^3/s (2-year flood event). The cross-section profile displays the ground surface, shown as a black line, which highlights the variation in terrain elevation across the section. The bank stations, indicated by red dots, mark the left and right channel banks, defining the active channel area. The Maximum Water Surface Elevation (WS Max WS), shown as a blue line, represents the computed water level during a high flow event or flood scenario. The shaded cyan area depicts the portion of the cross-section submerged under these conditions, providing insight into flow capacity, channel confinement, and potential flooding extends beyond the main channel.

As per the river bathymetric surveys, the left overbank of the channel is situated at an elevation of 56.20 m. The 2-year flood event resulted in minimal overtopping of the river channel at the left overbank. The maximum WSEL for this model run reached an elevation of 57.13 m. This resulted in overtopping of the left overbank with an inundation depth of 0.93 m.

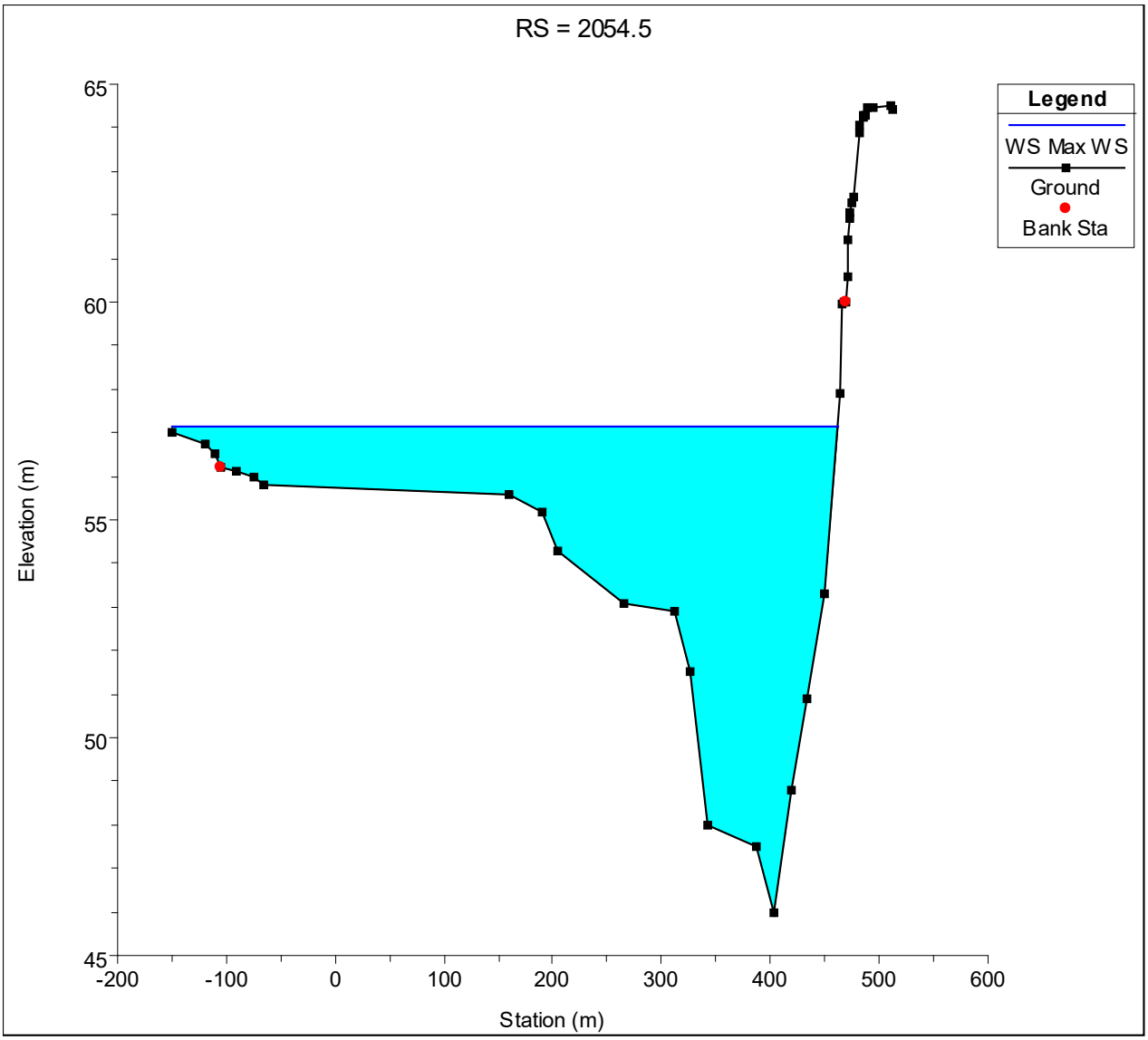


Figure 4.1 - Cross-section 2054.5 maximum water surface profile for 2-year flood event.

Figure (4.2) represents cross-section 2054.5 at a 10-year flood event with initial model discharge representing a value of 4500 m^3/s. The 10-year flood event resulted in the maximum WSEL of the channel reaching 58.29m. This is consistent with the stage-discharge relationship observed during the 2017 & 2019 flood event when discharge in the Ottawa River on 4th May 2017 and 25th April 2019. Overtopping of the river channel was at an inundation depth of 2.09 m.

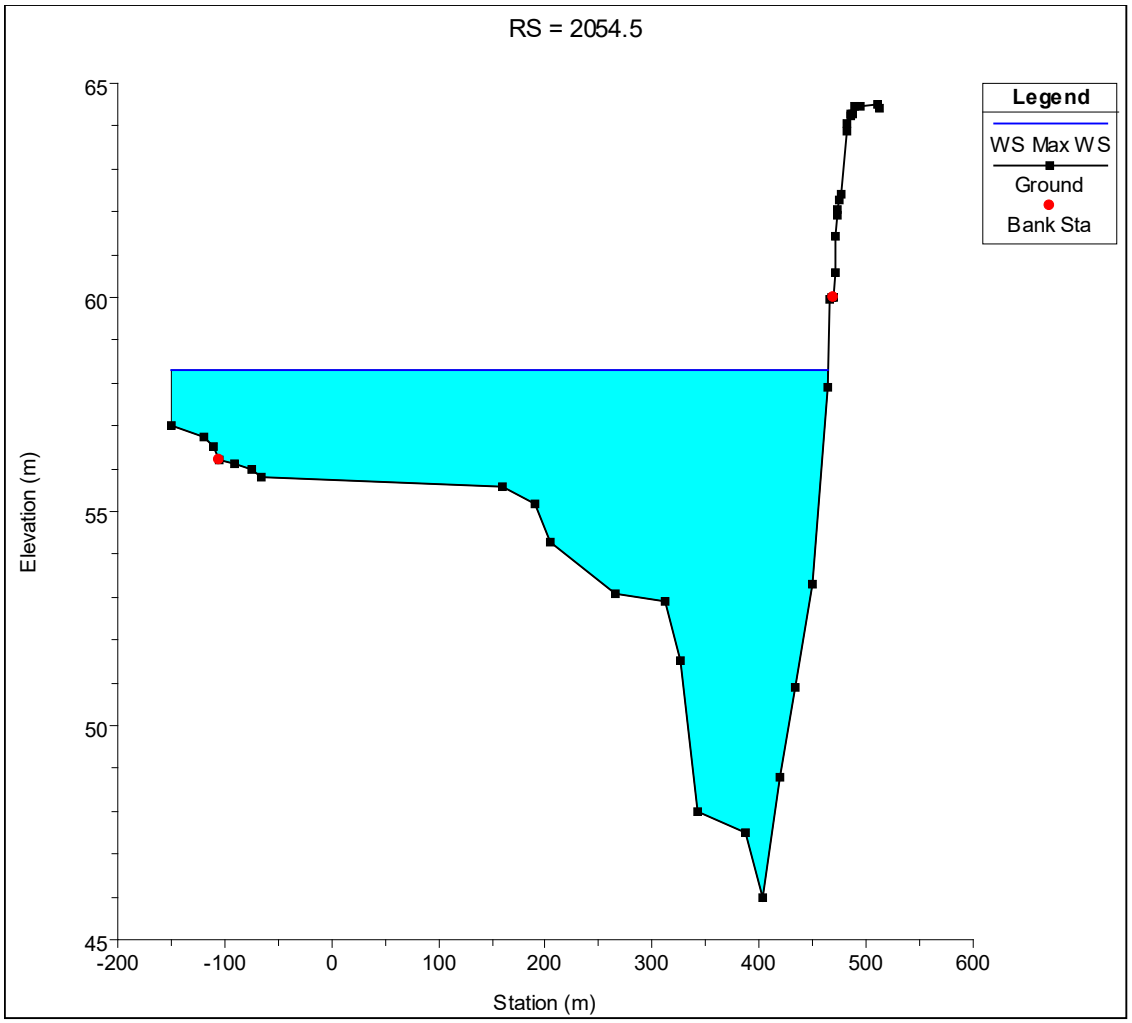


Figure 4.2 - Cross-section 2054.5 maximum water surface profile for 10-year flood event.

Similarly, Figure (4.3) represents cross-section 2054.5 at a 25-year flood event with an initial discharge peak taken as 5000 m^3/s. The 25-year flood event resulted in significant overtopping of the river channel at the left overbank. The maximum WSEL of the river channel at the cross-section was 58.66 m. This corresponded to the flooding in the Ottawa River on the 6th of May 2017 where the peak flood discharge was at 5110 m^3/s. For the 2019 flood, this discharge was observed on the 28th of April 2019 at a value of 5420 m^3/s. The inundation depth was 2.46 m showcasing an increase of almost 0.40 m from the 10-year event and over 1.50 m from the 2-year event.

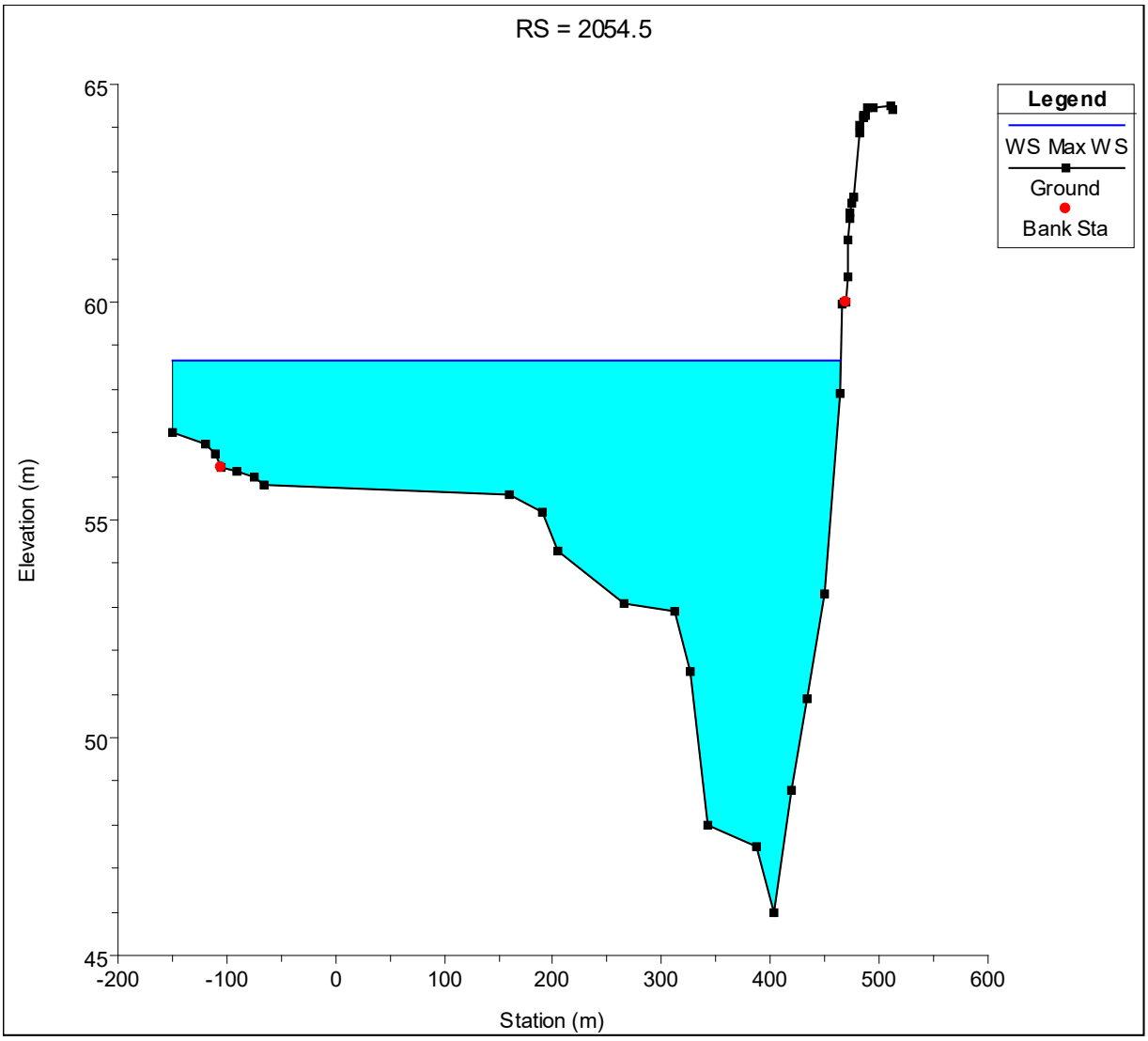


Figure 4.3 - Cross-section 2054.5 maximum water surface profile for 25-year flood event.

Figure (4.4) represents the 50-year flood event with a discharge value of 5500 m^3/s. This value was cross-referenced with the hydrometric station hydrograph when the peak reached the maximum value of discharge observed in the Ottawa River during the 2017 flood. The discharge was observed on 7 May 2017 when the peak flow rate was 5370 m^3/s. For the 2019 flood, a value of 5420 m^3/s was recorded on the 28th of April 2019. The maximum water surface elevation calculated by HEC-RAS was 59.03 m. This represented an inundation depth of 2.83 m.

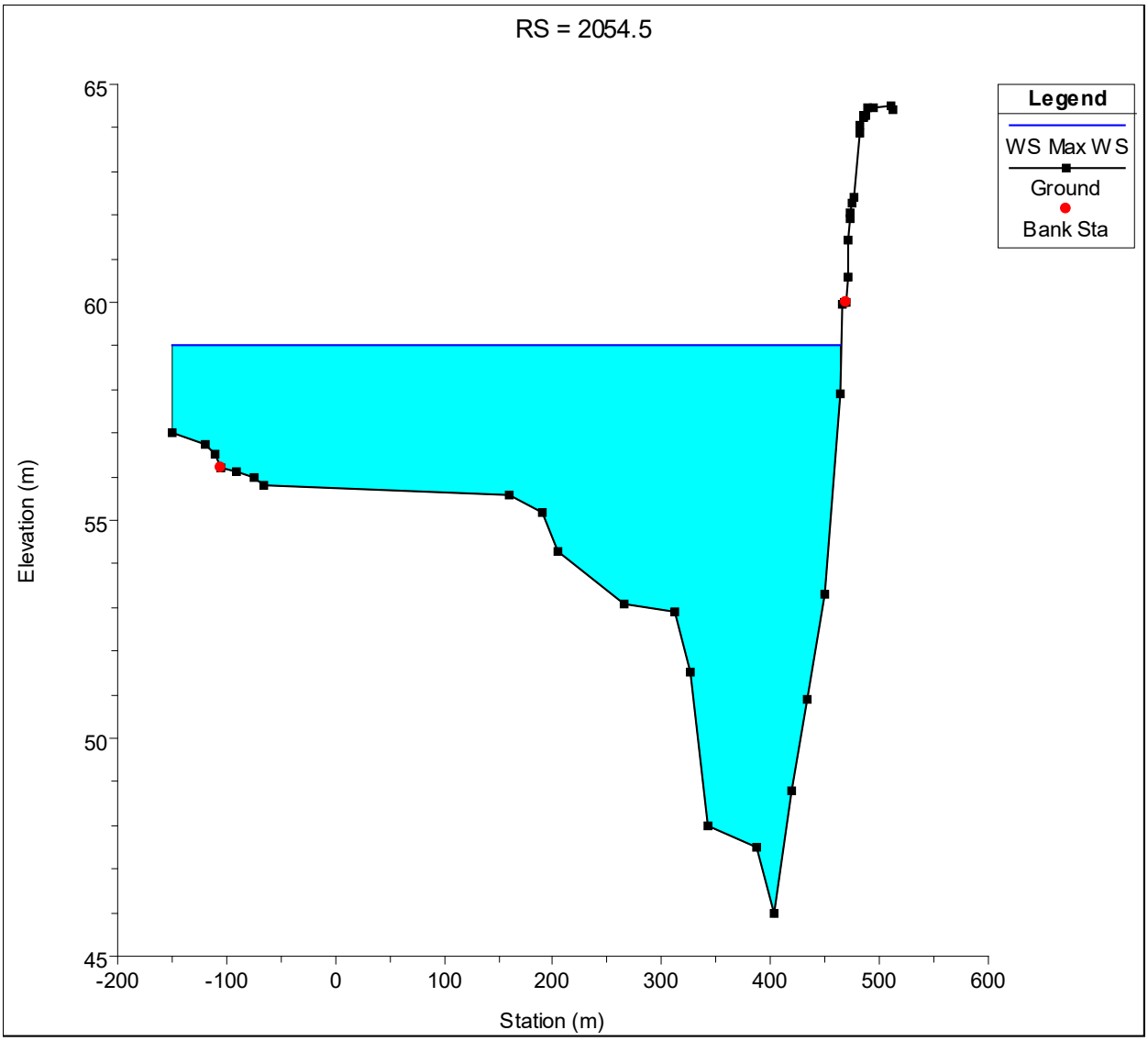


Figure 4.4 - Cross-section 2054.5 maximum water surface profile for 50-year flood event.

Figure (4.5) represents the 100-year flood event with a corresponding discharge value of 6000 m^3/s. This value was not observed during the 2017 flood as the maximum discharge recorded for the flood was 5370 m^3/s. For the 2019 flood, a value of 5970 m^3/s was recorded on the 1st of May 2019 representing the peak discharge recorded by the hydrometric station for the flood. As per the record, this value exceeded the 100-year flood event threshold. The maximum water surface elevation calculated by HEC-RAS was 59.33 m. representing an inundation depth of 3.13 m at the cross-section.

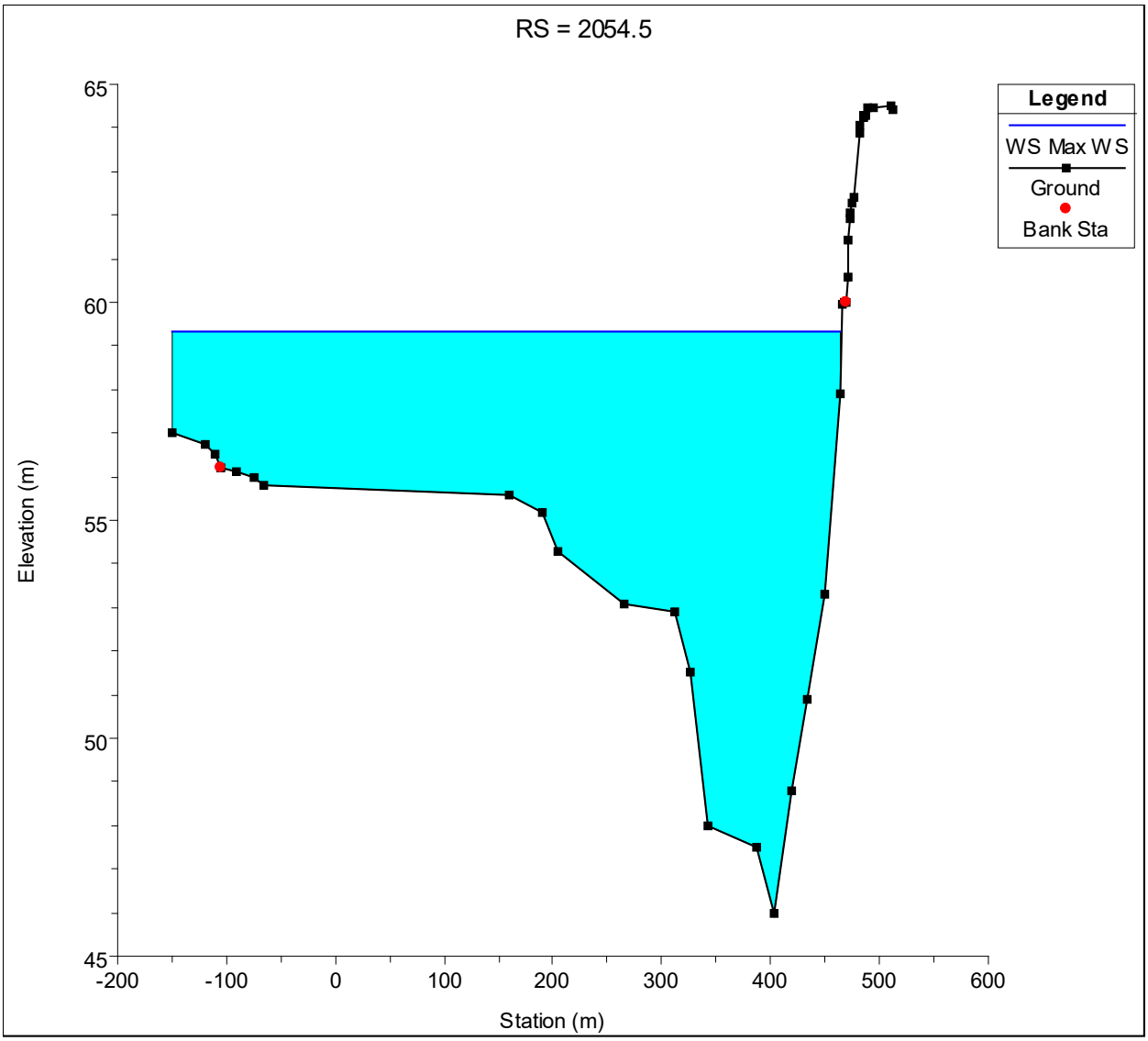


Figure 4.5 - Cross-section 2054.5 maximum water surface profile for 100-year flood event.

As part of the HEC-RAS model simulation, moving from the 100-year flood event to the 1000-year flood event was a key aspect that was taken into consideration. As data unavailable from hydrometric stations for the discharge peak of 7000 m^3/s, we moved into the territory of climate change scenario predictions. The results from the model were scrutinized and verified using the FFA techniques mentioned in the previous section. The model provided consistent results compared to simulations of 2-year, 10-year, 25-year, 50-year and 100-year flood events with inundation depth significantly increasing with the increase in initial boundary conditions. Figure

(4.6) represents the cross-section during the 1000-year flood event. Maximum water surface elevation observed was 59.81 m providing a substantial increase in inundation depth reaching 3.61 m.

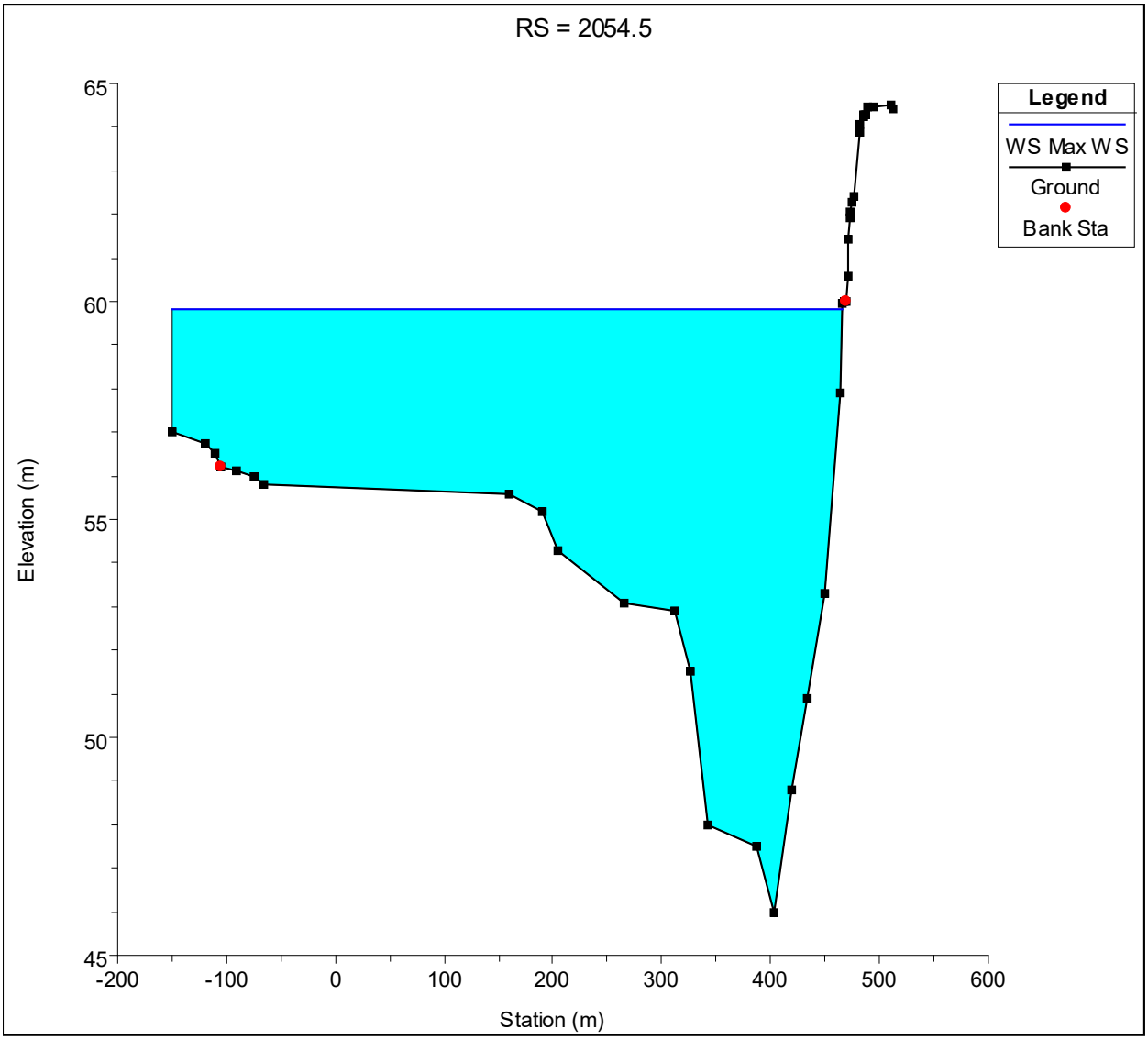


Figure 4.6 - Cross-section 2054.5 maximum water surface profile for 1000-year flood event.

The final model was run was done at a discharge peak of 7000 m^3/s. This represented a 10,000-year flood event scenario. Figure (4.7) shows cross-section 2054.4 during the model run

where the maximum water surface elevation observed was 60.22 m representing an inundation depth at the left overbank of 4.02 m.

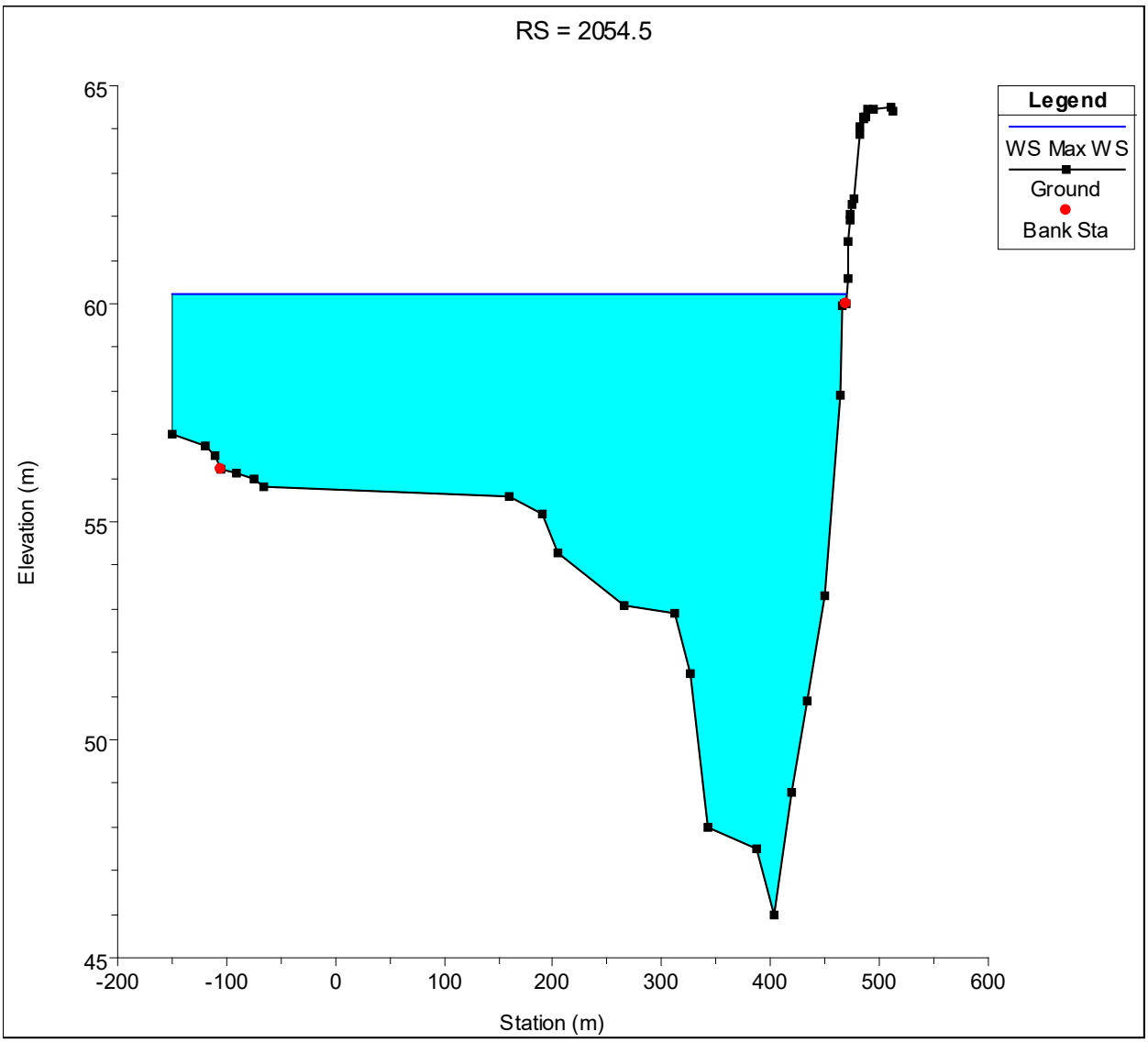


Figure 4.7 - Cross-section 2054.5 maximum water surface profile for 10,000-year flood event.

This comprehensive representation of cross-section 2054.5 provides a deep understanding of the river channel's relationship between discharge and maximum water surface elevation. Similar results were obtained for cross-section 2054.2 and 2054 which were downstream of 2054.5. The close examination and observations of these cross-sections was of paramount

importance to understand the relationship of the river channel with the 2D flood plain located in the town of Quyon.

Figure (4.8) represents cross-section 2054.2 with maximum water surface elevations shown for 2-year, 10-year, 25-year, 50-year, 100-year, 1000-year and 10,000-year flood events. Due to the variation in the bathymetric data between the cross-sections, the river channel's behaviour is somewhat different when compared to cross-section 2054.2.

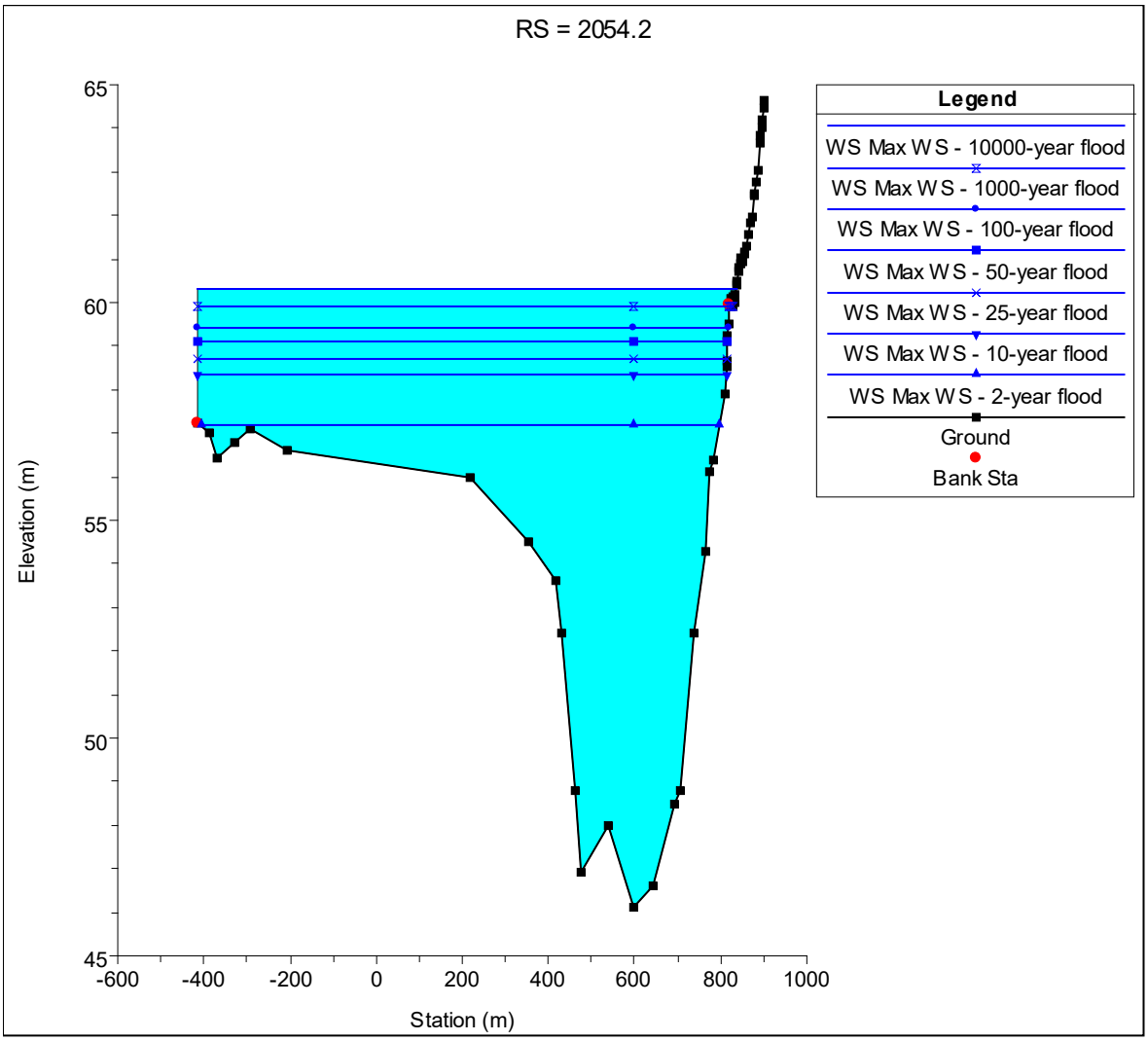


Figure 4.8 - Cross-section 2054.5 maximum water surface profile for 2-year, 10-year, 25-year, 50-year, 100-year, 1000-year and 10,000-year flood event.

Cross-section 2054.2 was located 318 m downstream of cross-section 2054.2. The right overbank was located at an elevation of 60 m while the left overbank connected to the 2D flow area was at an elevation of 57.25 m. The channel width was 573.76 m in the cross-stream direction between the left and right overbanks. The 2-year flood event resulted in a maximum WSEL of 57.17 m just below the river's left overbank, while the 10-year flood event resulted in a maximum WSEL of 58.35 m causing inundation at a depth of 1.1 m. The 25-year, 50-year and 100-year flood events resulted in maximum WSEL of 58.73 m, 59.10 m and 59.41 m. Inundation due to these events was 1.48 m, 1.85 m and 2.16 m respectively. Moving into climate change scenario predictions, the 1000-year and 10,000-year flood events resulted in maximum WSEL modelled at 59.90 m and 60.32 m with inundation depths of 2.65 m and 3.07 m.

Similar to the previous figure, Figure (4.9) represents cross-section 2054 which was located at the downstream end of the 2D flow area located in the town of Quyon. Cross-section 2054 was located 632.4 m downstream of cross-section 2054.2. The left overbank was elevated at 58 m and the right overbank at 59.97 m. The channel width was 1233.842 m representing a widening of the river channel. The 2-year, 10-year, 25-year, 50-year, 100-year and 10,000-year flood events resulted in maximum WSEL of 57.16 m, 58.33 m, 58.71 m, 59.09 m, 59.40 m, 59.88 m and 60.31 m respectively. The 2-year flood event did not have an inundation depth as the maximum WSEL was 0.84 m below the channel left overbank. The inundation depth for the 10-year event was 0.33m, 25-year at 0.71 m, 50-year at 1.09 m, 100-year at 1.40 m, 1000-year at 1.88 m and 10,000-year at 2.31 m. This information is critical for understanding the river's capacity to convey floodwaters and identifying the potential extent of overbank flooding during various return period events. It helps in assessing flood risks to nearby infrastructure and communities in the town of Quyon, supporting effective floodplain management and mitigation planning.

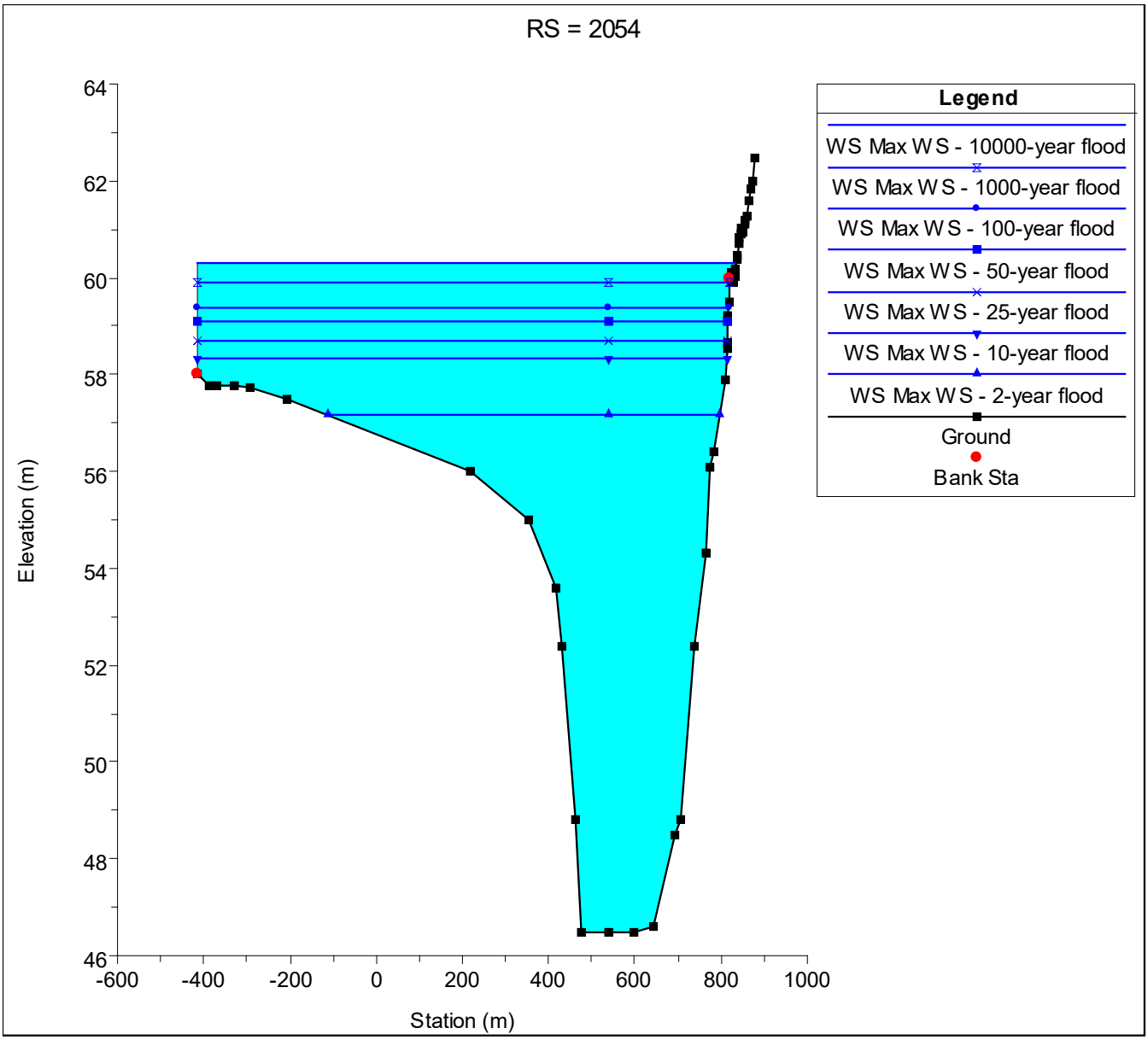


Figure 4.9 - Cross-section 2054 maximum water surface profile for 2-year, 10-year, 25-year, 50-year, 100-year, 1000-year and 10,000-year flood event.

The plan view and location of the three cross-sections is provided in Figure (4.10) to provide an idea of the location of the river channel and surrounding 2D floodplain. Cross-section 2054.5 is upstream of the 2D area, cross-section 2054.2 is in the middle and cross-section 2054 is at the downstream end of the town of Quyon.

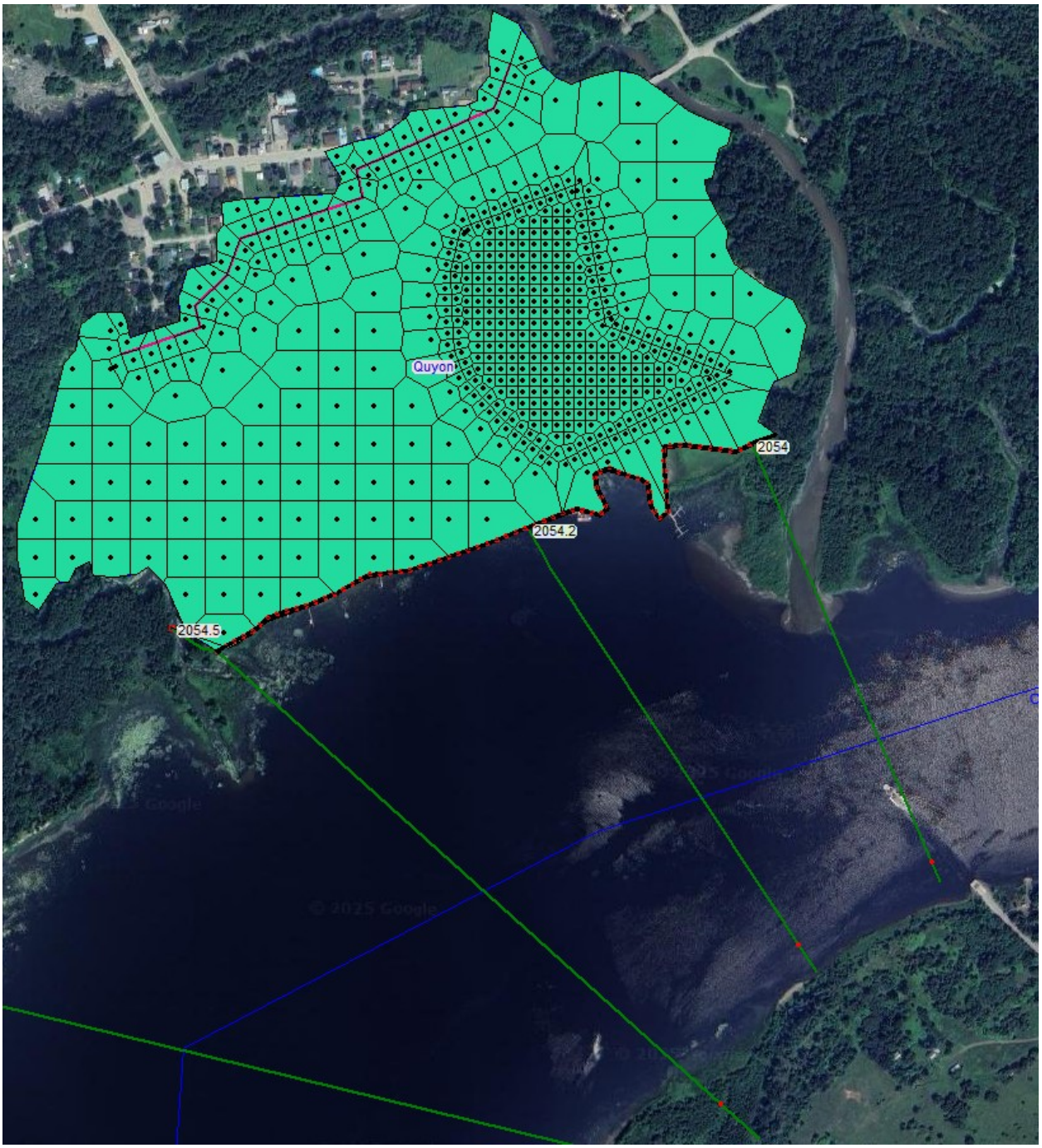


Figure 4.10 - Plan view of Cross-sections 2054.5, 2054.2 and 2054 with the 2D flow area situated in the town of Quyon, Quebec.

Along with Cross-section results, the 1D model run in HEC-RAS also provides general profile plots that show the Maximum Water Surface Elevations in the river channel along with a plethora of other variables. Figure (4.11) shows the general profile plot of the HEC-RAS model

run with maximum Water Surface elevations represented in dark blue for all the different model runs. The Lateral structure located at the town of Quyon is represented in grey with the right and left overbanks of the channel represented in dotted green and red lines respectively. Cross-sections are located plot's x-axis, moving from the left to right in the upstream to downstream direction. The y-axis represents the elevations of the channel at separate cross-sections and the corresponding maximum water surface elevations for the 2-year, 10-year, 25-year, 50-year, 100-year, 1000-year and 10,000-year return period flood events.

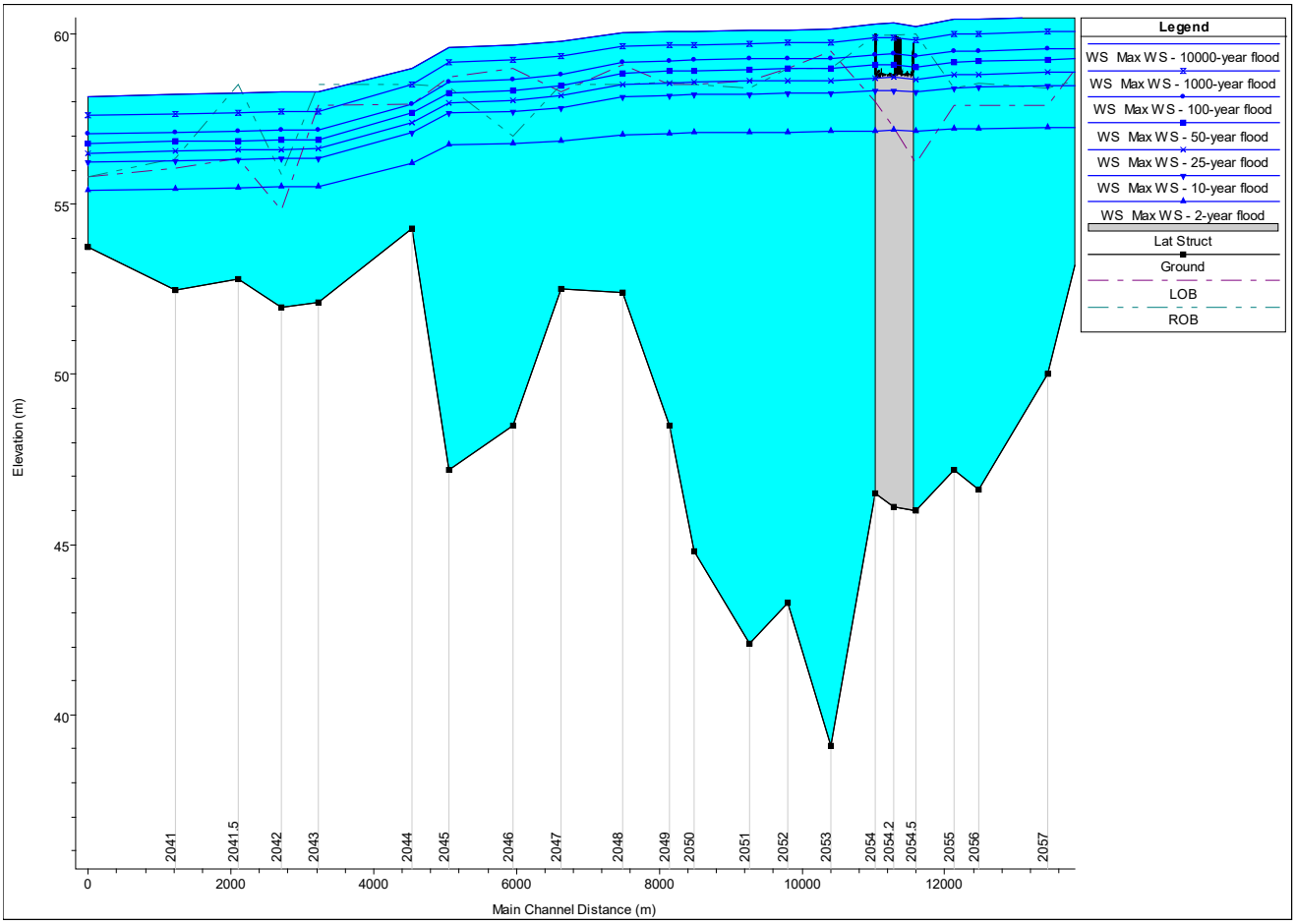


Figure 4.11 - General Profile Plot of the River Channel network displaying cross-sections, maximum WSEL, lateral structures and corresponding elevations.

4.3 – 2D HEC-RAS Modelling Results:

Apart from the 1D results shown in Section 4.2, the HEC-RAS numerical modeling software also has the capability to provide 2D results. These are displayed in the form of 2D inundation maps commonly seen in QGIS or ArcGIS flood extent maps (Rincón et al., 2018, Suthakaran et al., 2022). In the HEC-RAS model, 2D floodplain maps were generated for the town of Quyon corresponding to flood events of varying return period. As the inundation was mainly considered relevant for floods of return periods exceeding 25-year scenarios, these are presented in the section along with detailed analysis of flood inundation extents and depth across the town.

Figure (4.12) shows the inundation in the town of Quyon corresponding to a 50-year flood event. The satellite imagery is extracted from the QGIS "Google Satellite" layer with dark blue shade representing the extent of flooding.

Figure 4.12 - Flood Inundation Map for the 50-year flood event in the town of Quyon, Quebec.

During the 50-year flood event, floodwaters extended up to 356.11 m into the town from the left bank of the river, with the deepest inundation within the town center reaching 0.35 m, approximately 134.11 m from the left bank. The highest water depths were concentrated within the river channel itself, while the largest flood extents were observed near cross-section 2054.2. Minimal inundation occurred at the eastern and western edges of the town, submerging approximately 12% of the town's area.

Figure 4.13 - Flood Inundation Map for the 100-year flood event in the town of Quyon, Quebec.

Similar to the previous figure, Figure (4.13) shows the inundation in the town of Quyon corresponding to a 100-year flood event. The 100-year flood event extended further, with floodwaters reaching 364.4 meters from the left bank near cross-section 2054.2. The town center experienced the highest inundation depth of 0.65 m, while flooding at the eastern edge reached 0.20 m at cross-section 2054, extending 174.11 m into the town. This event submerged 24% of

Quyon, demonstrating a significant increase in both flood extent and depth compared to the 50-year flood.

For the 1000-year flood event shown in Figure (4.14), the maximum inundation extent expanded to 405.4 m, primarily at the eastern edge of the town. The deepest water, measuring 1.20 m, was observed near the town center, 132.54 m from cross-section 2054.2. This flood event caused widespread inundation across the east, west, and central areas, with 56% of the town submerged, reflecting the impact of higher river discharges and broader floodplain coverage.

Figure 4.14 - Flood Inundation Map for the 1000-year flood event in the town of Quyon, Quebec.

The 10,000-year flood event resulted in extensive and widespread flooding throughout Quyon as shown in Figure (4.15). The maximum flood extent reached 464.4 m at the eastern end

of the town, with the deepest water recorded near the town center at 1.74 m. Over 74% of the town was submerged, highlighting the vast reach and devastating potential of extreme flood events.

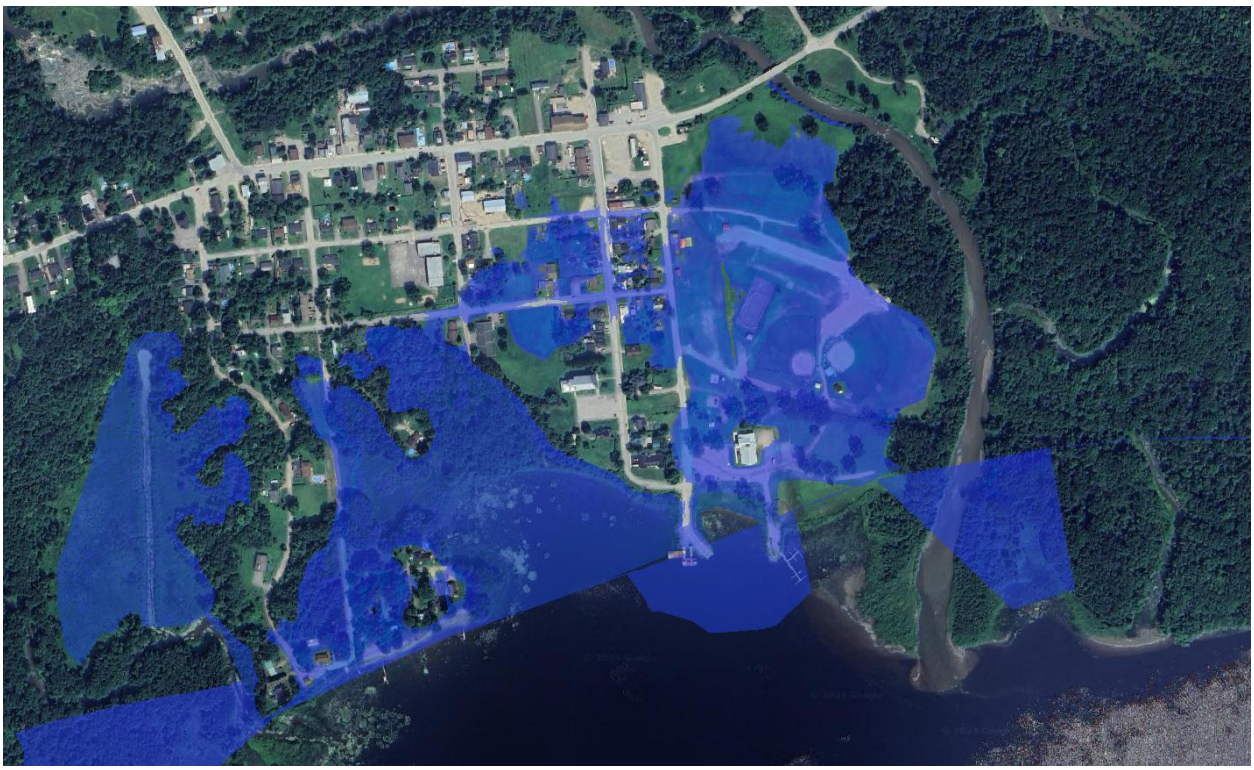


Figure 4.15 - Flood Inundation Map for the 10,000-year flood event in the town of Quyon, Quebec.

These HEC-RAS 2D results play a crucial role in floodplain delineation by offering detailed insights into both the extent of flooding and the associated inundation depths. These results effectively capture the complex hydrodynamic behavior of the river channel, including flow distribution, velocity patterns, and flood propagation, extending from the main channel into the adjacent floodplain areas. This comprehensive understanding supports accurate flood mapping, risk assessment, and the development of effective flood management strategies. As larger and less frequent flood events occur, more of the town's infrastructure, residential areas, and critical services are placed at risk, underscoring the need for proactive and adaptive flood management strategies.

4.4 – Comparative Analysis of Flood Extents with Satellite Imagery:

4.4.1 – 2017 Flood Event:

The May 2017 flood in Ottawa was primarily driven by a combination of unusual weather conditions and significant changes in snow cover, which contributed to elevated river discharge and water levels. During this month, the average temperature was 12.80°C, which was above the typical seasonal range, with a maximum temperature of 17.40°C and a minimum of 8.20°C. These warmer-than-usual temperatures, combined with an unusually high amount of rainfall—172.40 mm—led to significant runoff into the Ottawa River. The Ottawa River basin experienced an exceptional decrease in snow cover as shown in Figure (4.16). The average snow cover area in May 2017 was 29,792 km², which was 26,349 km² below the long-term average since 2000, marking the lowest snow cover extent recorded in recent decades. This lack of snow meant less snowmelt contributed to the river's flow, but the large rainfall overwhelmed the system, leading to rapid increases in river discharge. As a result, the river's discharge peaked at 3854.83 m³/s, and the water level rose to 59.74 meters, which caused severe flooding in the Ottawa region as shown in Figure (4.13). The combination of these factors—a warm month, high precipitation, and low snow cover—created ideal conditions for the river's capacity to be exceeded, resulting in widespread flooding that impacted communities and infrastructure along the river.

This region's snow cover has fluctuated significantly in recent years. In some years, the amount of snow cover has been significantly higher than average, and in others it has been significantly lower. These changes are likely due to climate change and rising global temperatures. Reduced snow cover can have serious consequences for the environment and humans, including reduced freshwater resources, increased flooding risks due to faster snowmelt, disruption of ecosystems dependent on seasonal snowpacks, and negative impacts on agriculture and water

quality. These changes can also exacerbate the effects of climate change, leading to more extreme weather events and altering local hydrological cycles.

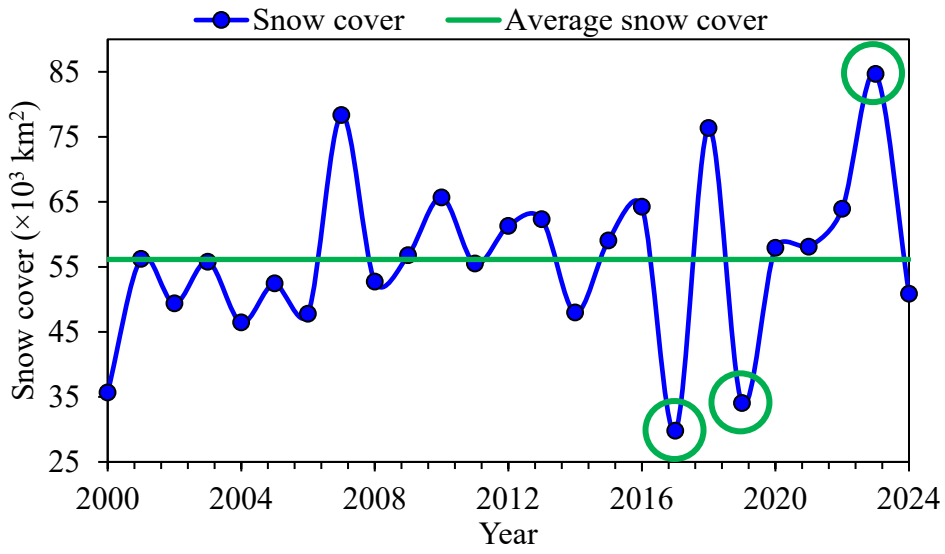


Figure 4.16 - Changes in snow cover in Ottawa River watershed from 2000 to 2024. These data were extracted from MODIS MOD10A1 product.

Floodplain extraction was performed using Sentinel-1 radar images. Sentinel-1 radar images operate at a wavelength of 5.60 cm, which allows them to penetrate clouds and dust, providing clear imagery of the Earth's surface. Images with VV polarization and an ascending pass with 10-meter resolution were used. Satellite images from 2022 were employed as the base image to represent the pre-flood condition. The identification of differences between flooded and non-flooded water areas was developed using JavaScript code within the Google Earth Engine platform. By determining the appropriate threshold, the water surface was detected and separated from other features in the satellite image (Figures 4.17–4.19).

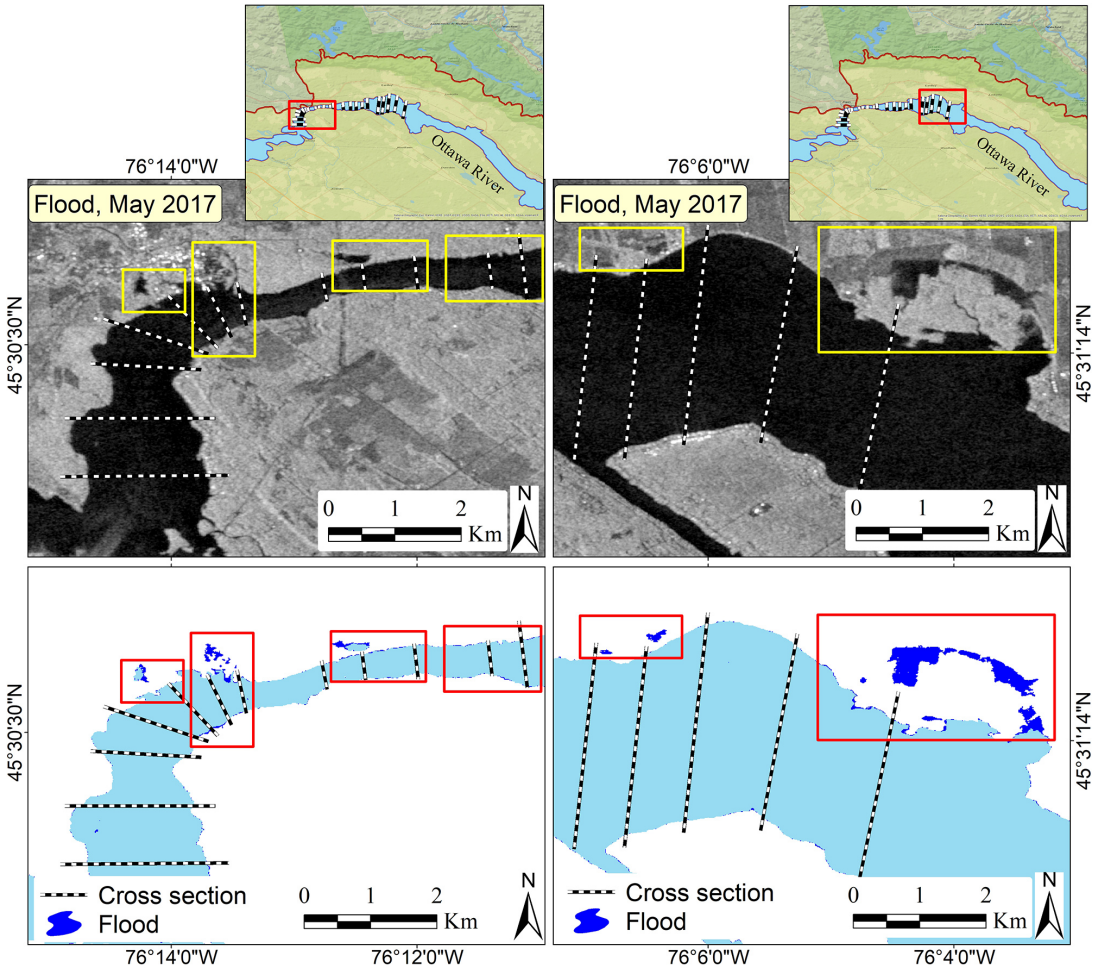


Figure 4.17 - The flood extent in May 2017 is shown, with a Sentinel-1 satellite image using VV polarization and ascending pass.

4.4.2 – 2019 Flood Event:

The May 2019 flood in Ottawa was a result of a combination of moderate temperatures, significant rainfall, and a marked reduction in snow cover, which led to increased river discharge and elevated water levels. In May 2019, the average temperature was 11.70°C, with a maximum of 16.50°C and a minimum of 6.82°C, which, although not extreme, were sufficient to melt a significant portion of the snowpack. The region received 102.60 mm of rainfall, which was substantial and contributed to increased runoff. The snow cover in the Ottawa River basin during May was 34,066 km², which was notably lower than the long-term average by 22,076 km²—

making it the second-lowest snow cover observed since 2000, following the record low in May 2017. The reduced snowpack meant that much of the snowmelt had already occurred earlier in the season, but the heavy rainfall compounded the runoff and further saturated the ground. This combination of factors led to a dramatic increase in river discharge, which peaked at 5173.80 m^3/s. The water level of the Ottawa River reached 60.35 meters, triggering widespread flooding in the region. Despite the lower snow cover compared to previous years, the rainfall in combination with residual snowmelt and saturated soils led to an exceptionally high river flow, making May 2019 another severe flooding event in the region as shown in Figure (4.18).

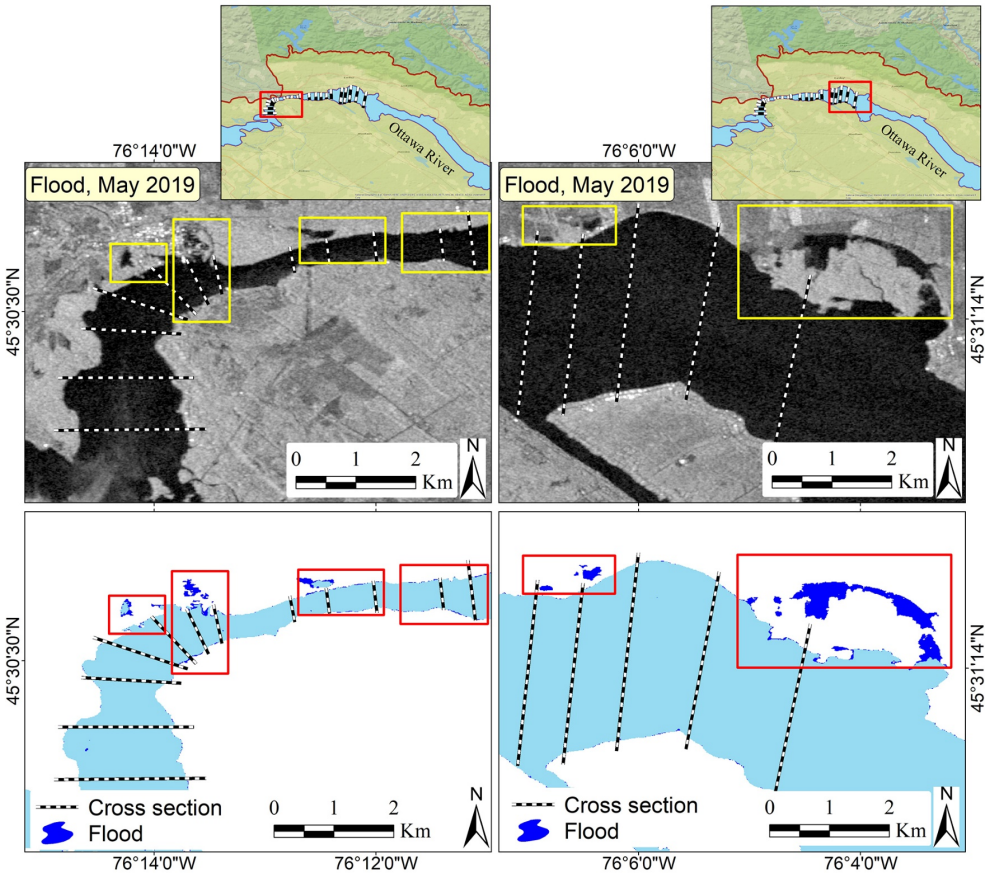


Figure 4.18 - The extent of flooding in May 2019 is depicted using a Sentinel-1 satellite image with VV polarization and an ascending orbit pass. (Qureshi et al., 2025).

4.4.3 – 2023 Flood Event:

The May 2023 flood in Ottawa was driven by a combination of relatively warm temperatures, moderate rainfall, and unusually high snow cover in the Ottawa River basin, which contributed to substantial river discharge and elevated water levels. During May 2023, the average temperature was 13.82°C, with a maximum of 20.25°C and a minimum of 7.35°C, which were above seasonal norms and likely accelerated snowmelt. Despite receiving only 51.60 mm of rainfall, the region experienced significant runoff due to an unusually large snowpack. The snow cover in the Ottawa River basin during May 2023 was 84,676 km², which was 28,534 km² above the long-term average since 2000, marking the highest snow cover recorded in the region since the beginning of the century as seen in Figure (4.16). This extensive snow cover meant that a considerable amount of snowmelt was still contributing to the river's flow, further exacerbating the effects of the rainfall. As a result, the river discharge peaked at 3290.96 m³/s, and the water level reached 59.46 meters, leading to localized flooding. The combination of a large snowpack, accelerated melting due to above-average temperatures, and moderate rainfall resulted in significant hydrological stress on the Ottawa River system, causing the 2023 flood to be one of the most notable in recent years as shown in Figure (4.19).

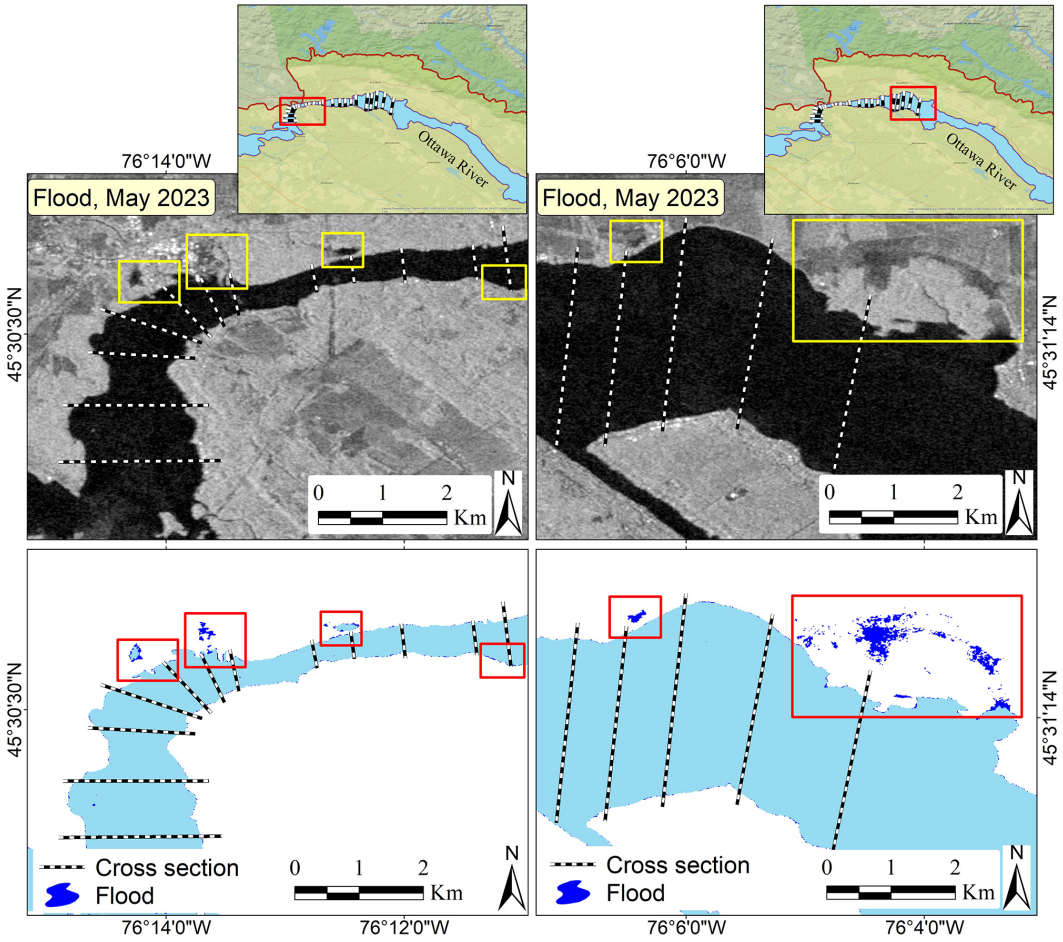


Figure 4.19 - The flood extent in May 2023, highlighted in dark blue, is depicted using a Sentinel-1 satellite image with VV polarization and an ascending orbit pass.

4.4.4 – Insights into the Flood Events and HEC-RAS

The floods of May 2017, 2019, and 2023 in Ottawa were all marked by varying climatic and hydrological conditions, each contributing to significant river discharge and elevated water levels. In 2017, the region experienced exceptionally high rainfall (172.40 mm) combined with a notably low snow cover of 29,792 km², which was the smallest since 2000. The minimal snowpack, alongside intense rainfall, led to rapid runoff and caused the Ottawa River's discharge to peak at 3854.83 m³/s, with a water level reaching 59.74 meters.

In contrast, 2019 saw more moderate conditions, with 102.60 mm of rainfall and a snow cover of 34,066 km², still below the long-term average but higher than in 2017. Despite the increased snowfall compared to 2017, the river's discharge was higher at 5173.80 m³/s, and the water level reached 60.35 meters, due in part to the significant rainfall combined with snowmelt.

By 2023, conditions had shifted again, with the snow cover reaching an unprecedented 84,676 km², the highest since 2000, marking a significant increase compared to both 2017 and 2019. The snowfall anomaly was 28,534 km² above the long-term average, contributing to substantial snowmelt in addition to 51.60 mm of rainfall. While the rainfall was lower compared to previous years, the large snowpack resulted in a higher river discharge of 3290.96 m³/s and a water level of 59.46 meters, slightly lower than in 2019. This comparison highlights how variations in snow cover, rainfall, and temperatures played critical roles in shaping the flood dynamics in each year, with 2017 being influenced more by heavy rainfall, 2019 by a combination of moderate snowmelt and rain, and 2023 by the unusual combination of a large snowpack and moderate rainfall leading to significant runoff.

By comparing floods for May 2017, 2019, and 2023, we can observe how the flood extent, surface water distribution, and inundated areas varied across these years. It’s likely that 2017 and 2019 (years with higher precipitation) show more extensive flooded areas, while 2023, with lower precipitation and discharge, would show smaller inundated zones, which aligns with the lower river discharge and water levels in that year. By analyzing flood and non-flood conditions through remote sensing, we can identify how flood-prone areas shift in response to varying precipitation patterns.

To summarize, Figures (4.17-4.19) visually confirm the extent of flooding in the region, complementing the findings of our HEC-RAS analysis and providing additional context on the

link between climatic factors and flood risk. The maps present Sentinel-1 radar satellite imagery, which captures data in the C-band (wavelength of 5.60 cm), depicting the flooding events of May 2017, 2019, and 2023. They highlight the areas impacted by the flood, offering valuable ground truth data to validate the study's hydrological analysis and flood risk assessment. The observed flood extent along the left riverbank closely aligns with the results generated in this study, further reinforcing the accuracy of the modeled floodplain.

4.5 – Flood Hazard Analysis and Strategic Management:

Figure (4.20) presents the number of households affected by flood events, along with a comprehensive flood risk analysis for the town of Quyon, based on advanced 2D modeling conducted using HEC-RAS. This analysis provides essential insights for prioritizing evacuation planning and flood control measures. Demographic data, including population distribution, were obtained from the Open Building Population Layer (Maxim Fortin, 2021), allowing for precise identification of affected areas and populations. The 50-year flood zone (blue) representing a 2% annual probability — affects approximately 1.54% of the population. Although relatively limited to spatial extent, its higher probability of occurrence makes it a key area for frequent flood risk management. The 100-year flood zone (green), with a 1% annual probability, affects 2.06% of residents, extending slightly into additional areas. The 1000-year flood zone (yellow), corresponding to a 0.10% annual probability, impacts 46.90% of the population, highlighting a significant increase in both flood extent and severity. The most severe scenario, the 10,000-year flood zone (red), with a 0.01% annual probability, threatens 78.30% of the population, underscoring the catastrophic consequences of rare, extreme flood events and their potential to inundate large portions of the town, including critical infrastructure and residential areas.

The results highlight the need for a hierarchical and strategic approach to flood management in Quyon, addressing immediate, medium-term, and long-term priorities to mitigate risks and enhance community resilience.

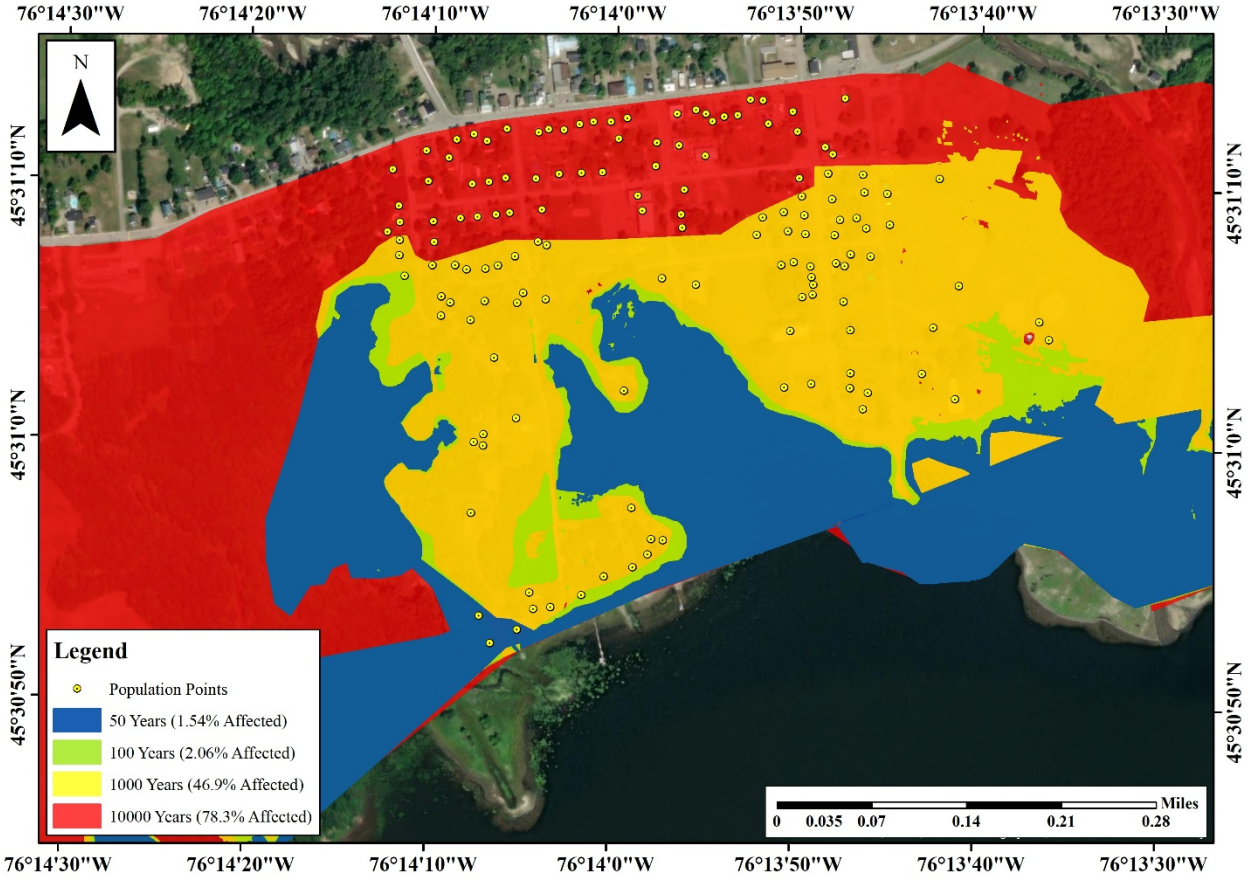


Figure 4.20 - Flood hazard map of Quyon showing the spatial distribution of flood zones based on different return periods.

For high-frequency flood zones (Blue and Green), immediate measures should focus on evacuation planning, infrastructure upgrades, and early warning systems. Detailed evacuation routes must be developed and maintained, ensuring accessibility for seniors, individuals with disabilities, and those with limited mobility, while regular community drills enhance preparedness. Flood control infrastructure, including levees, culverts, and drainage systems, should be upgraded to effectively manage frequent floodwaters, with permeable pavements incorporated in urban areas to improve water absorption. Additionally, real-time monitoring systems for river levels, ice jam

movements, and rainfall should be installed to provide timely alerts for residents and emergency services, ensuring a swift and coordinated response.

For moderate-risk areas (Yellow), medium-term strategies should focus on infrastructure resilience, community engagement, and natural flood mitigation. Flood-resistant construction techniques, such as elevating homes and critical infrastructure and using water-resistant materials, should be implemented, with a priority on retrofitting public buildings like schools and hospitals. Community engagement initiatives should include educational programs to enhance flood awareness and preparedness, training residents on flood-proofing homes and emergency response protocols. Additionally, the establishment of vegetative buffers along riverbanks can help reduce runoff velocity, stabilize soil, and mitigate erosion during high flows, enhancing overall flood resilience.

Long-term planning for extreme event zones (Red) should prioritize sustainable land use, adaptive infrastructure, and climate resilience. Stricter zoning laws should be enforced to restrict development in high-risk flood zones, encouraging the relocation of critical infrastructure and densely populated areas. Large-scale flood diversion projects should be developed to manage peak flows during extreme events, including flood channels, reservoirs, and retention basins. Climate adaptation strategies must incorporate future projections of precipitation, temperature changes, and altered snowmelt patterns, alongside natural floodplain restoration to enhance water absorption. An integrated watershed management approach should be adopted, fostering collaboration across municipalities to regulate upstream and downstream flows, ensuring consistent and equitable flood protection.

4.6 - New-Generation Group Method of Data Handling (GMDH) Results:

When compared to the standard GMDH model (Ivakhnenko, 1978), the New-Gen GMDH model adopts a flexible design, this is achieved by using both adjacent and non-adjacent layers containing neurons. This method ensures a balance between simplicity and the necessary complexity to capture critical dependencies. By establishing direct connections from the initial layer to the final output layer, the New-Gen GMDH reduces reliance on intermediate layers while maintaining high predictive accuracy. Additionally, neurons in the intermediate layers process inputs from fewer sources, simplifying the model structure and enhancing computational efficiency (Qureshi et al., 2025).

As discussed in the model development section of the report, the study considered an initial five different inputs. These inputs included real-time meteorological data obtained from the Climate Station Ottawa CDA, provided by the Government of Canada, which utilized air temperatures in the region (minimum, maximum, and mean) and precipitation. The final variable was the river discharge obtained from the Hydrometric Station located at Britania (02KF005), which served as a historical record summarizing the behavior of the Ottawa River. For climate change scenario discharge data, outputs from the HEC-RAS model were utilized to enhance the predictive capabilities of the ML model.

The input variables within our ML model were characterized based on an alpha-numeric system in MATLAB. The breakdown of each input variable and its corresponding input variable is provided in Table (4.2) to better explain the relationship between values in our dataset. SC1 incorporated variables from $x1 - x4$, SC2 was run utilizing input variables from $x1 - lag12$ and SC3 included variables from $x1\ to\ x7$ without the incorporation of lags.

Table 4.2 - Input Variable Classification

Input Variable in the New-Gen GMDH Model	Classification within dataset
$x1$	Maximum Temperature
$x2$	Minimum Temperature
$x3$	Mean Temperature
$x4$	Total Precipitation
lag1	Lag at ACF of 1
lag2	Lag at ACF of 0.6
lag12	Lag at ACF of 0.55
$x5$	Minimum Discharge
$x6$	Mean Discharge
$x7$	Maximum Discharge

The outputs produced by the incorporation of input variables from $x1 - x4$ is mentioned as SC1 (Mean Temperature, Maximum Temperature, Minimum Temperature, and Total Precipitation as Input Variables and Observed Discharge as Output Variable). SC2 included the inputs from SC1 along with Lags (lag 1, lag 2, lag 12) taking the total number of input variables to seven. A wide set of criteria were modified to test the results of the model utilizing time lags and the introduction/removal of separate variables of the dataset for Sensitivity Analysis. SC3, on the other hand, utilized input variables from SC1 along with three different input variables $x5, x6, x7$ (Minimum Discharge, Mean Discharge, and Maximum Discharge) bringing the total set of input variables to seven.

The representation of the different variables used in SC1 is provided in Figure (4.21). A polynomial equation of 3 was applied to the model. The model utilized two input variables to generate Layer 1, the function $x5 = f(x3, x4)$ is a relationship developed between the third and fourth variables to generate a new variable termed $x5$. For the Output layer, the final equation

utilized $x5$ from Layer 1 along with Input variable $x1$ to generate the final equation $y = f(x1, x3, x5)$.

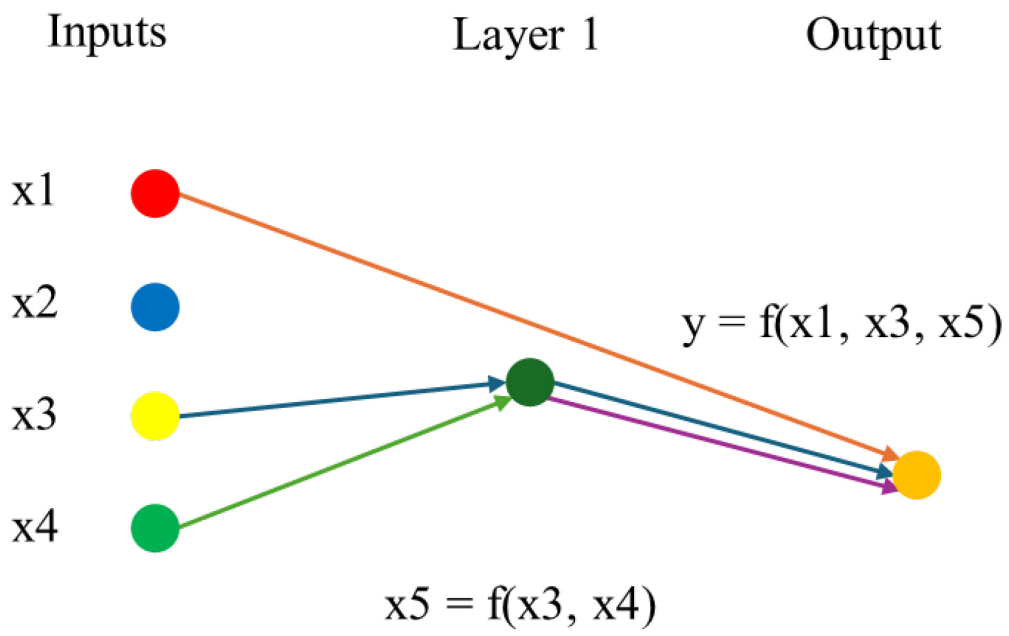


Figure 4.21 - Structure of the model run utilizing SC1 (Input Variables utilized are Maximum Temperature, Minimum Temperature, Mean Temperature, and Total Precipitation).

Equation (4.1) shows the final numeric representation generated by the model when predicting discharge based on the structure provided in Figure (4.21). The direct contribution of each input variable is represented in the equation with values that include single variables such as $2.34 \times x5$. Quadratic terms such as $1.69 \times x3 \times x5$ account for pairwise interactions between variables and their squared effects, allowing the model to capture curvature in the response. Cubic terms such as $27.93 \times x3^3$ represent higher-order interactions where three variables interact simultaneously.

$$y = 5109.51 + 2.34 \times x5 + 2098.02 \times x3 - 2030.65 \times x1 + 1.69 \times x3 \times x5 - 1.55 \times x1 \times x5 - 1405.68 \times x1 \times x3 - 0.0051 \times x5^2 + 753.32 \times x3^2 + 654.69 \times x1^2 +$$

$$0.52 \times x1 \times x3 \times x5 - 0.0020 \times x3 \times x5^2 - 0.30 \times x3^2 \times x5 + 0.0015 \times x1 \times x5^2 - 80.03 \times x1 \times x3^2 - 0.22 \times x1^2 \times x5 + 77.62 \times x1^2 \times x3 + 1.36 \times 10^{-6} \times x5^3 + 27.93 \times x3^3 - 25.39 \times x1^3 \quad (4.1)$$

The visual representation of the input variables of SC2 are provided in Figure (4.22). A similar polynomial equation of 3 was applied to the SC2 model as used in SC1. With the inclusion of lags (lag1, lag2, lag12), the model utilized 3 input variables to generate Layer 1, the function $x8 = f(x1, x4, lag\ 1)$ is a relationship that the New-Gen GMDH model developed between the three input variables to generate a new term $x8$. Layer 2 of the model utilized the input variable $x8$ from Layer 1 along with $x2$ and $lag1$ from the original set of input variables giving the function $x15 = f(x2, lag1, x8)$. The Output Equation (4.2) of the model utilized $lag12$ and $x4$ from the original input variables along with $x15$ generated in Layer 2 to give the function $y = f(x4, lag12, x15)$. The interaction between the variables is similar to the one explained in SC1's Equation.

$$y = 213.96 + 0.58 \times x15 - 0.24 \times lag12 - 29.12 \times x4 + 0.00040 \times lag12 \times x15 - 0.050 \times x4 \times x15 + 0.11 \times x4 \times \text{lag12} + 6.0082 \times 10^{-5} \times x15^2 + 4.38 \times 10^{-5} \times lag12^2 + 13.0065 \times x4^2 - 4.42 \times 10^{-5} \times x4 \times lag12 \times x15 + 8.94 \times 10^{-8} \times lag12 \times x15^2 - 1.16 \times 10^{-7} \times lag12^2 \times x15 + 3.52 \times 10^{-5} \times x4 \times x15^2 + 4.71 \times 10^{-6} \times x4 \times lag12^2 + 0.0091 \times x4^2 \times x15 - 0.025 \times x4^2 \times lag12 - 7.0021 \times 10^{-8} \times x15^3 + 1.27 \times 10^{-8} \times lag12^3 - 0.06 \times x4^3 \quad (4.2)$$

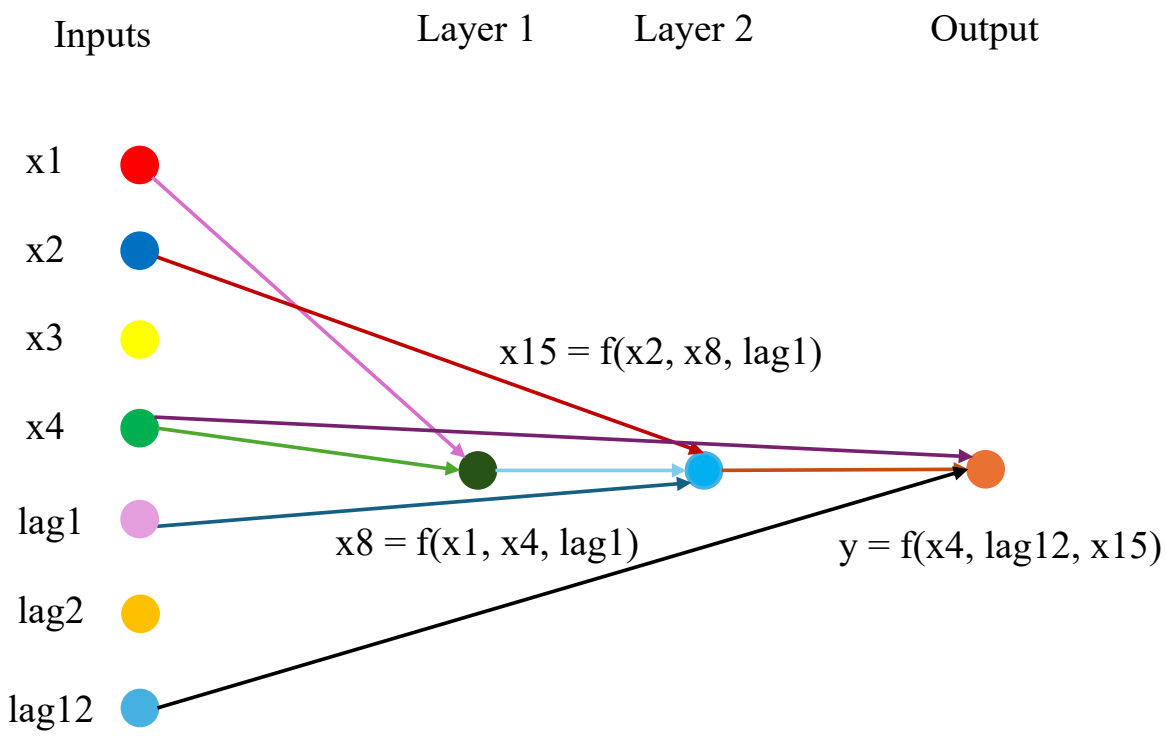


Figure 4.22 - Structure of the model run utilizing SC2 (Input Variables utilized are Maximum Temperature, Minimum Temperature, Mean Temperature, Total Precipitation, lag1, lag2, lag12).

SC3 was developed with the incorporation of the average mean, maximum, and minimum discharge into the input variables, the model was able to predict the discharge in the Ottawa River with enhanced performance. Figure (4.23) represents the combinations of input variables used in the scenario for our New-Gen GMDH ML model. With the inclusion of the above-mentioned variables, the model utilized 7 input variables for the initial phase. The polynomial degree of 3 was applied to the model. The selection from a wider set of input variables and utilizing a polynomial equation of 3 provided more versatility to the model. It enhanced the ability of the model to capture complex relationships between adjacent and previous layers.

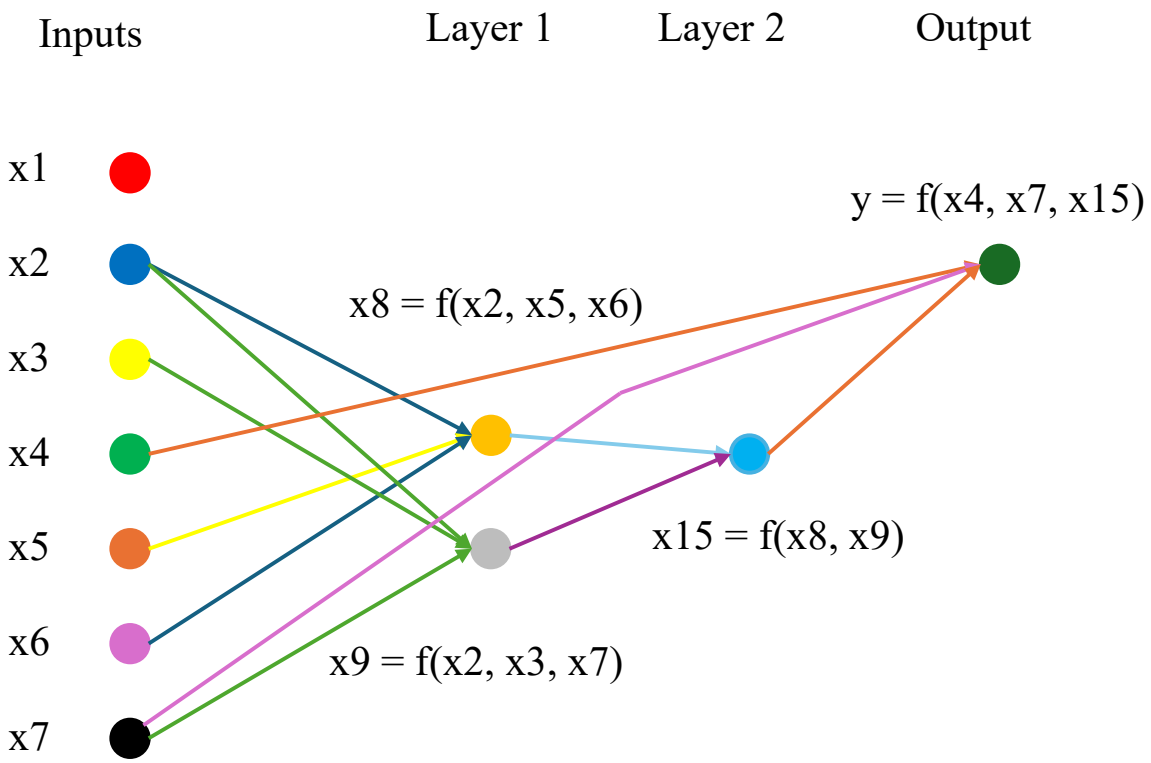


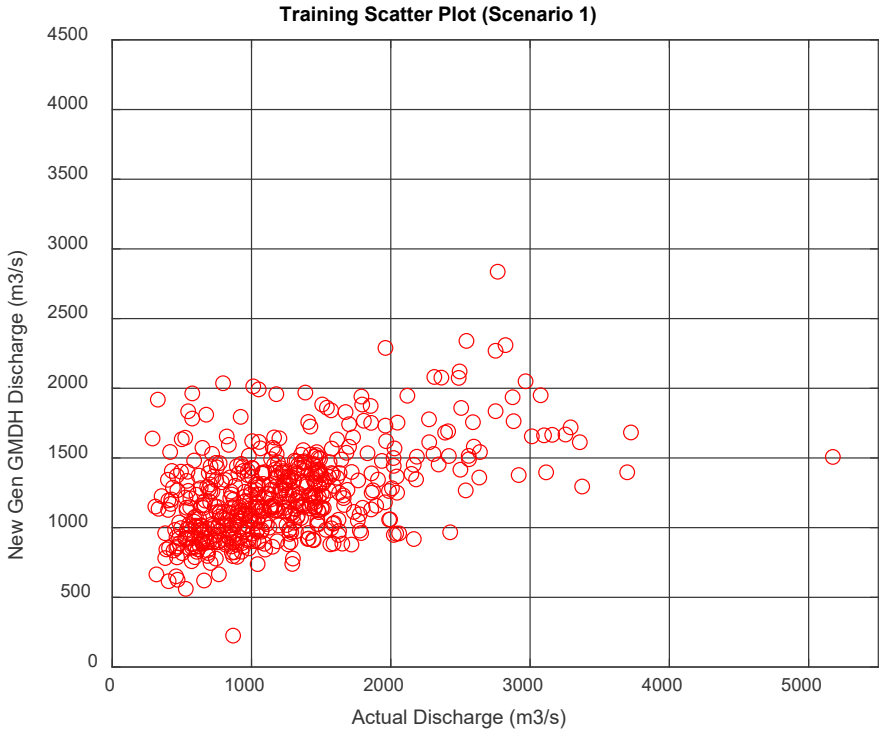
Figure 4.23 - Structure of the model run utilizing SC3 (Input Variables utilized are Maximum Temperature, Minimum Temperature, Mean Temperature, Total Precipitation, Average Minimum Discharge, Average Maximum Discharge, and Average Minimum Discharge).

Layer 1 of the model generated an equation that combined three input variables. Functions $x8 = f(x2, x5, x6)$ and $x9 = f(x2, x3, x7)$ are created with a combination of input variables to capture complex relationships between variables in the provided dataset. Layer 2 of the model generated a new function $x15 = f(x8, x9)$ which combined the two generated inputs from Layer 1. The final Output equation for discharge in the second modeling scenario $y = f(x4, x7, x15)$ showcases the advantage of having selection criterion utilizing variables from the previous layer and the original dataset, resulting in an equation that extracted data from Layer 2 and the original dataset of input variables to give the most accurate result.

Equation (4.3) represents the final numeric representation generated by the model when predicting discharge based on the structure provided in Figure (4.23). The interaction between the variables is similar to the one observed in the explanation of Equation (4.3).

$$y = 91.65 + 1.30 \times x15 - 0.34 \times x7 + 28.89 \times x4 + 0.0015 \times x7 \times x15 - 0.44 \times x4 \times x15 + 0.47 \times x4 \times \mathrm{x7} - 0.00081 \times x15^2 - 0.00044 \times x7^2 - 90.94 \times x4^2 - 0.00027 \times x4 \times x7 \times x15 + 5.17 \times 10^{-7} \times x7 \times x15^2 - 3.10 \times 10^{-7} \times x7^2 \times x15 + 0.00017 \times x4 \times x15^2 + 4.88 \times 10^{-5} \times x4 \times x7^2 + 0.12 \times x4^2 \times x15 - 0.063 \times x4^2 \times x7 - 2.61 \times 10^{-7} \times x15^3 + 8.04 \times 10^{-8} \times x7^3 + 8.75 \times x4^3 \quad (4.3)$$

Figures (51, 52) show the scatter plots generated for the group of Samples for SC1, SC2, and SC3 in the training and testing phases of the model runs. The scatter plots illustrate the predictive accuracy of each method compared to actual values, offering insights into their respective strengths and limitations.



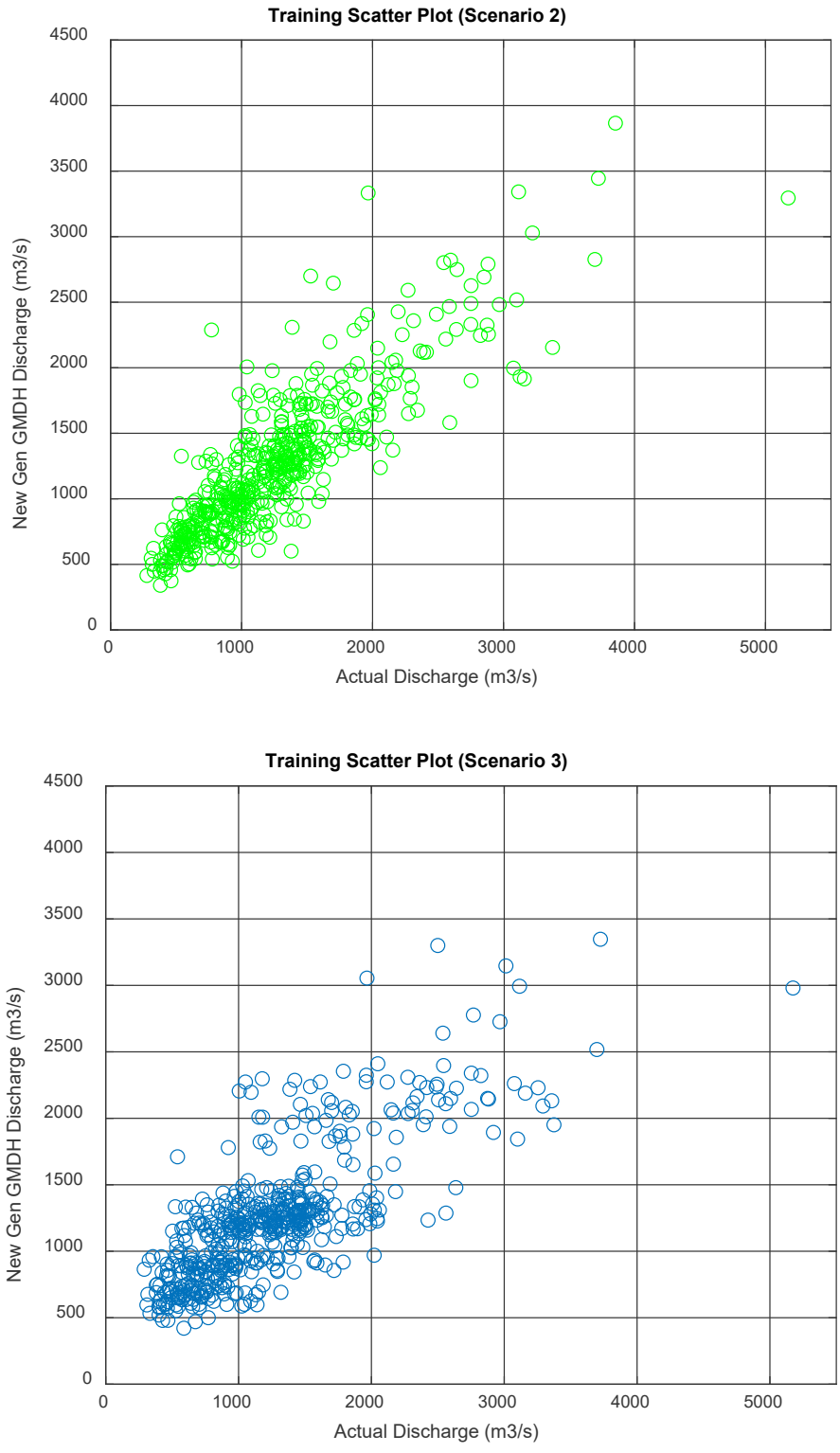
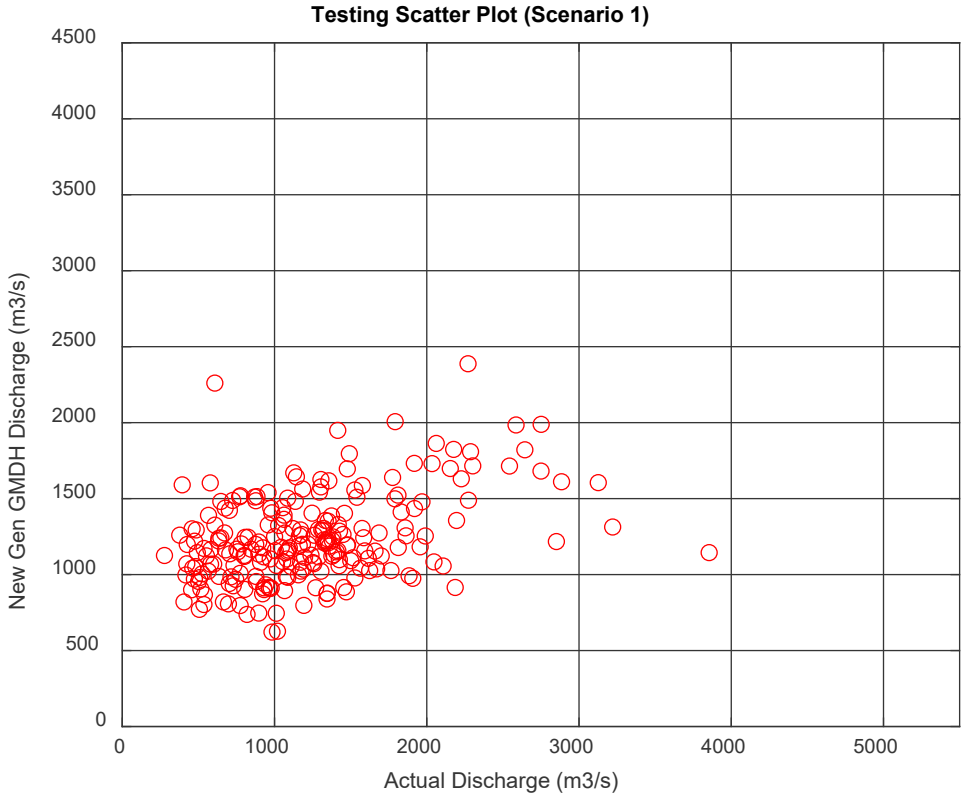
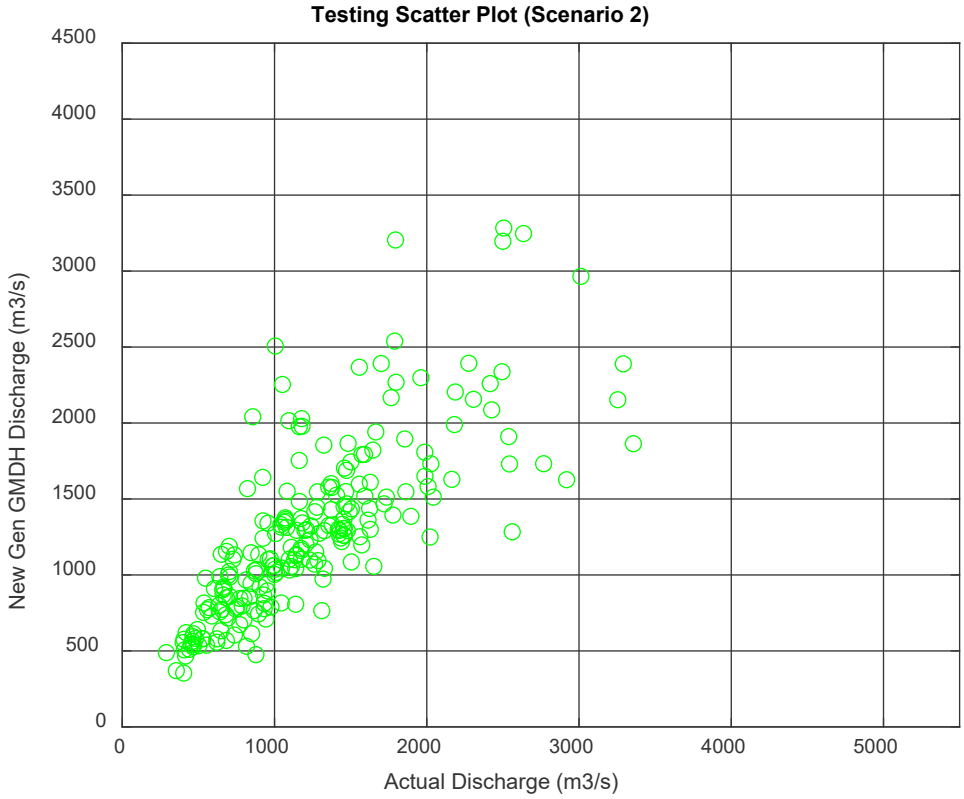


Figure 4.24 - Scatter Plots Comparing Performance of the New-Gen GMDH Model between SC1, SC2, and SC3 Training Phase.


Testing Scatter Plot (Scenario 1)
New Gen GMDH Discharge (m3/s)
Actual Discharge (m3/s)
0
500
1000
1500
2000
2500
3000
3500
4000
4500
0
1000
2000
3000
4000
5000


Testing Scatter Plot (Scenario 2)
New Gen GMDH Discharge (m3/s)
Actual Discharge (m3/s)
0
500
1000
1500
2000
2500
3000
3500
4000
4500
0
1000
2000
3000
4000
5000

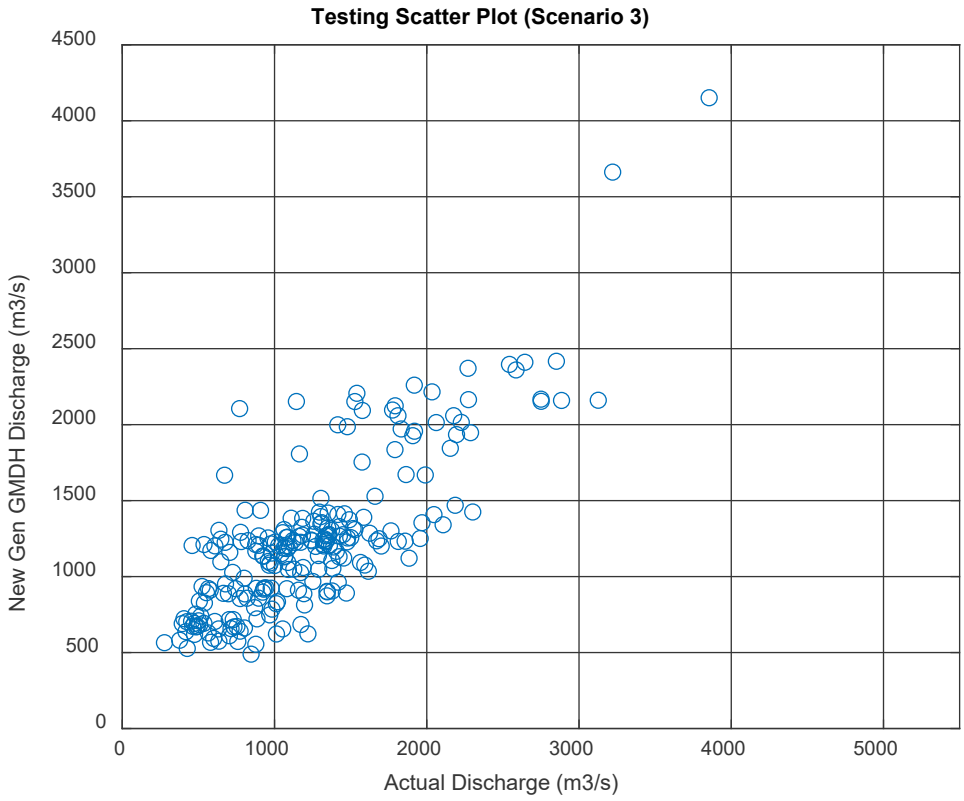
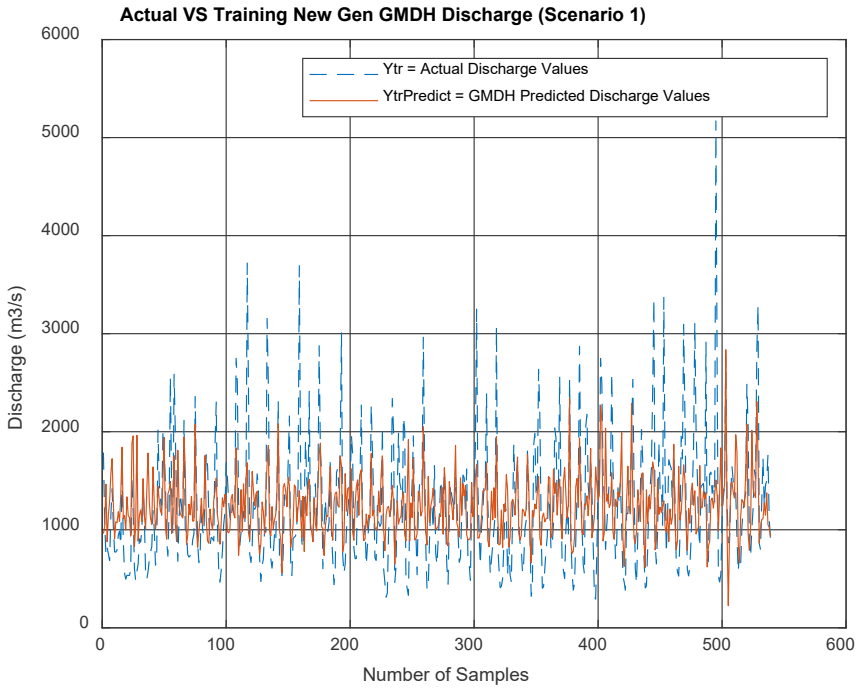


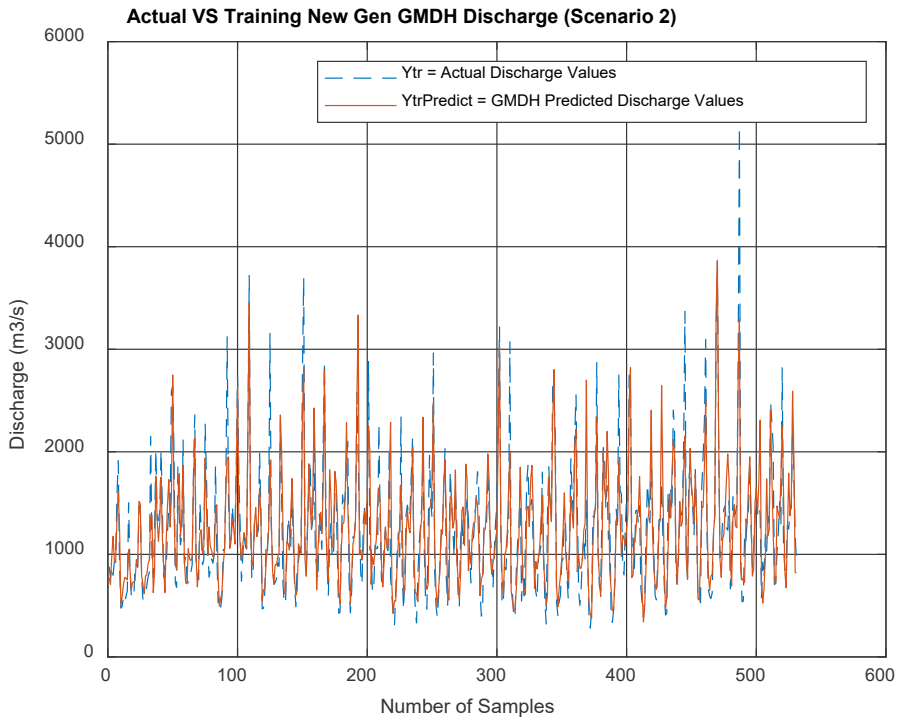
Figure 4.25 - Scatter Plots Comparing Performance of the New-Gen GMDH Model between SC1, SC2, and SC3 Testing Phase.

In the training phase, SC2 with green points clearly outperforms SC1 and SC3, as its scatter plot shows a tighter clustering along the diagonal line y = x, indicating that the predicted "New Gen GMDH Discharge (m^3/s)" values align closely with the "Actual Discharge (m^3/s)" across a wide range from 0 to 4500 m^3/s. SC2 exhibits a more linear trend with a narrower spread and fewer outliers, suggesting consistent and accurate predictions even at higher discharge values, though the density remains higher at lower values (0 to 1500 m^3/s). In contrast, SC1 (red points) and SC3 (blue points) display similar patterns, with heavy clustering at lower discharge values (0 to 1500 m^3/s) and a dense concentration around 500–1000 m^3/s, but both show a significant spread as discharge increases, with many points deviating far from the diagonal, especially above 2000 m^3/s,

indicating under- or over-predictions and a higher number of outliers (e.g., predicted values around 3500 m^3/s for actual discharges near 1000 m^3/s).

Moving to the testing phase, SC2 again demonstrates superior performance, maintaining its tight alignment along the diagonal with a consistent spread across the range, suggesting robust generalization to unseen data. SC1 in the testing phase mirrors its training behavior, with dense clustering at lower values (0 to 1500 m^3/s) but a wider spread and more outliers at higher discharges, indicating persistent inconsistency. SC3 in the testing phase shows a slight improvement over its training performance, with a somewhat tighter spread, but it still struggles with significant deviations from the diagonal, particularly for discharges above 1500 m^3/s, and remains less reliable than SC2. Overall, SC2 is the best-performing scenario in both training and testing phases, offering the most accurate and consistent predictions across the entire range of discharge values, while SC1 and SC3 lag behind, particularly in handling higher discharge values, with SC3 showing marginal improvement in testing but still falling short of SC2's performance.





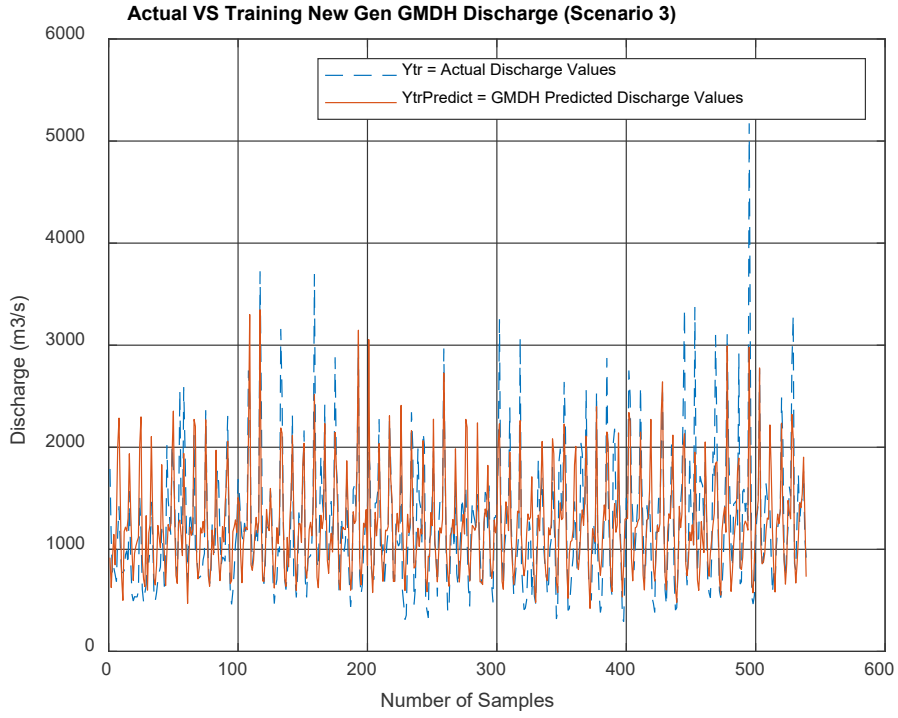
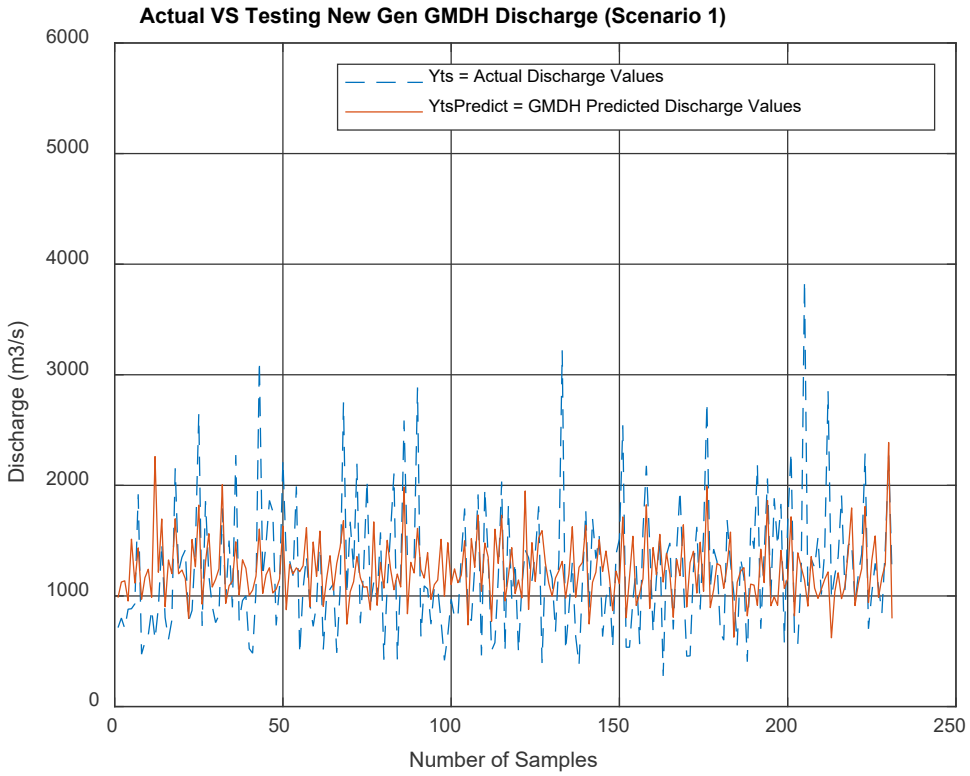
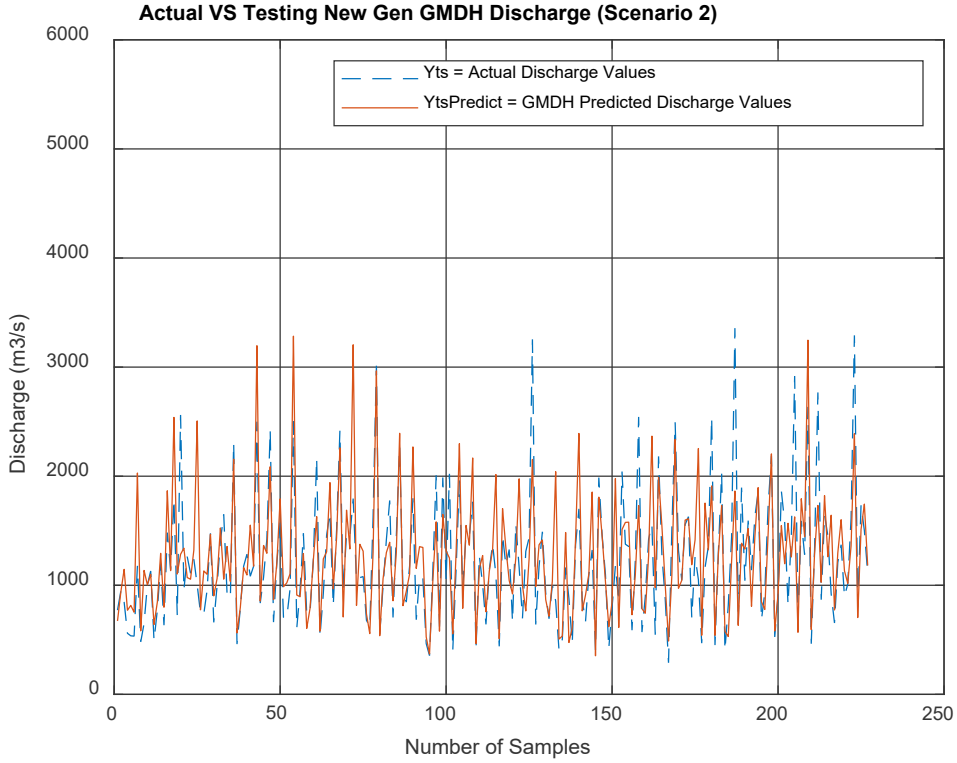


Figure 4.26 - Discharge Prediction Comparing Performance of the New-Gen GMDH Model between SC1, SC2 and SC3 Training Phase.


Actual VS Testing New Gen GMDH Discharge (Scenario 1)
Yts = Actual Discharge Values
YtsPredict = GMDH Predicted Discharge Values
Discharge (m3/s)
Number of Samples
0
1000
2000
3000
4000
5000
6000
0
50
100
150
200
250


Actual VS Testing New Gen GMDH Discharge (Scenario 2)
Yts = Actual Discharge Values
YtsPredict = GMDH Predicted Discharge Values
Discharge (m3/s)
Number of Samples
0
1000
2000
3000
4000
5000
6000
0
50
100
150
200
250

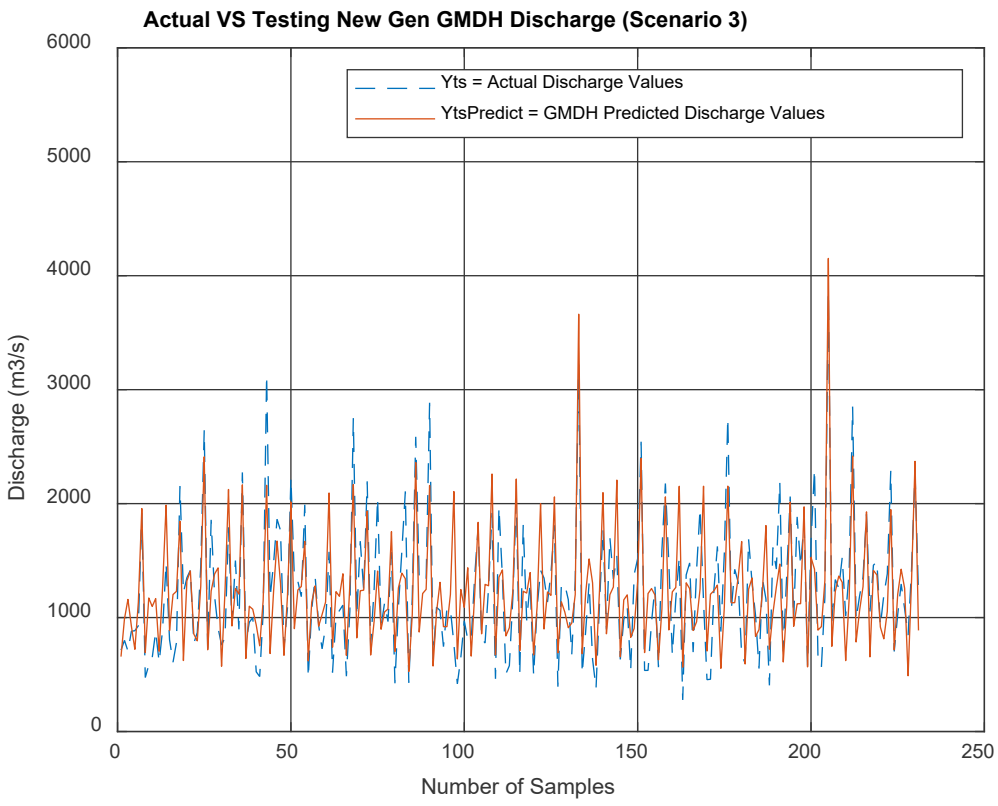


Figure 4.27 - Discharge Prediction Comparing Performance of the New-Gen GMDH Model between SC1, SC2, and SC3 Testing Phase.

Figures (4.26, 4.27) depict the discharge predictions of the model runs compared to the actual discharge values for the Ottawa River. 30% of the samples were allocated to the testing phase and 70% to the training phase. In SC1, the model consistently underestimates peak discharge events, with actual values reaching up to 3800-4000 m^3/s while predictions rarely exceed 2000-2300 m^3/s, representing approximately 40-45% underestimation during critical high-flow periods in the training phase. The testing dataset (~225 samples) shows this pattern consistently, with prediction values averaging around 1000-1400 m^3/s compared to actual values that fluctuate more dramatically between 400-3800 m^3/s. This can be attributed to the lower number of input variables provided to the New-Gen GMDH model when compared to other scenarios. SC2 demonstrates markedly improved performance with better alignment between predicted and actual values,

capturing approximately 75-85% of discharge variation amplitude compared to 55-65% in SC1. In this scenario, the model maintains predictions mostly within ±500 m^3/s of actual measurements, with baseline flows consistently around 1100-1300 m^3/s for both predicted and actual values. The training dataset for SC2 (~550 samples) shows improved handling of discharge events up to 3000-3500 m^3/s compared to SC1. Most notably, SC3 exhibits concerning anomalies with severe overprediction spikes reaching 3500-4000 m^3/s at sample points 150 and 210 in the testing dataset when actual values were only around 1500-2000 m^3/s, representing 100-167% overestimation. Across all scenarios, the training datasets reveal similar pattern recognition capabilities, but SC2 maintains the most balanced error distribution with mean absolute errors visibly lower (estimated 300-400 m^3/s) than both SC1 (estimated 500-700 m^3/s during normal flows, 1000-1500 m^3/s during peaks) and SC3 (highly variable, ranging from 300-1500+ m^3/s with unpredictable error distribution). The New-Gen GMDH model's response to extreme discharge events above 3000 m^3/s shows variance: SC1 consistently underestimates by 1000-1500 m^3/s, SC2 underestimates by a more moderate 500-800 m^3/s, while SC3 occasionally overestimates by 1000-2000 m^3/s during specific sample periods. Based on the Discharge results, SC2 emerges as the clearly superior model configuration due to the addition of Lags that help identify key patterns when including input variables, offering the optimal balance between prediction accuracy across the full range of discharge conditions and reliability during extreme high-flow events that are critical for effective hydrological management and flood prevention.

After evaluating the performance of both models qualitatively through scatter plots and discharge relationships, it is essential to assess the models' performances statistically to provide a more comprehensive evaluation of predictive accuracy and reliability. Figure (4.28) represents R^2

and VAF radar plots of the New-Gen GMDH's performance comparing SC1, SC2, and SC3 in the training and testing phases.

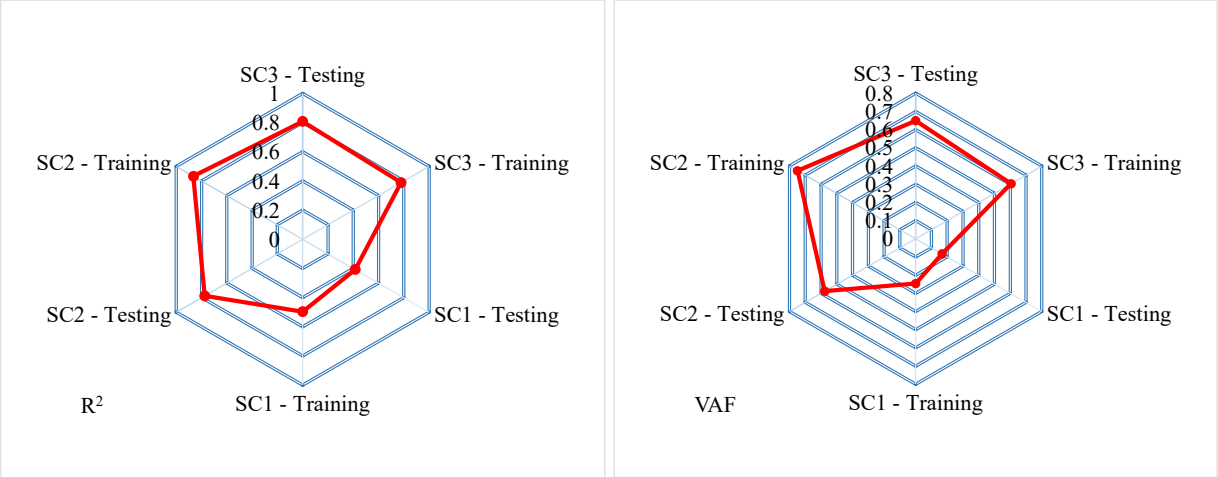


Figure 4.28 - Statistical Indices (R2) and (VAF) of the New-Gen GMDH Model between SC1, SC2 and SC3.

The comparison of SC1, SC2, and SC3 based on R^2 and VAF highlights clear differences in model performance. R^2 measures the proportion of variance in the observed data explained by the model, with values closer to 1 indicating stronger predictive power. SC1 performs the weakest, with R^2 values of 0.49 (training) and 0.42 (testing), reflecting limited model accuracy. SC3 shows improvement, especially in testing (0.81), however with a training R^2 value of 0.78 suggesting good generalization. However, SC2 outperforms both, achieving the highest R^2 in training (0.87) and strong testing performance (0.78), indicating excellent model fit and predictive reliability. Similarly, VAF—another measure of how much variance in the data is accounted for by the model—further supports SC2's superiority. SC1 yields the lowest VAF (0.25 training and 0.17 testing), while SC3 improves with 0.61 (training) and 0.65 (testing). SC2 again leads with VAF values of 0.75 in training and 0.58 in testing, demonstrating its strong ability to capture data variance across both phases. Figure (4.29) displays the RMSE and RMSRE radar plots for the New-Gen GMDH model runs.

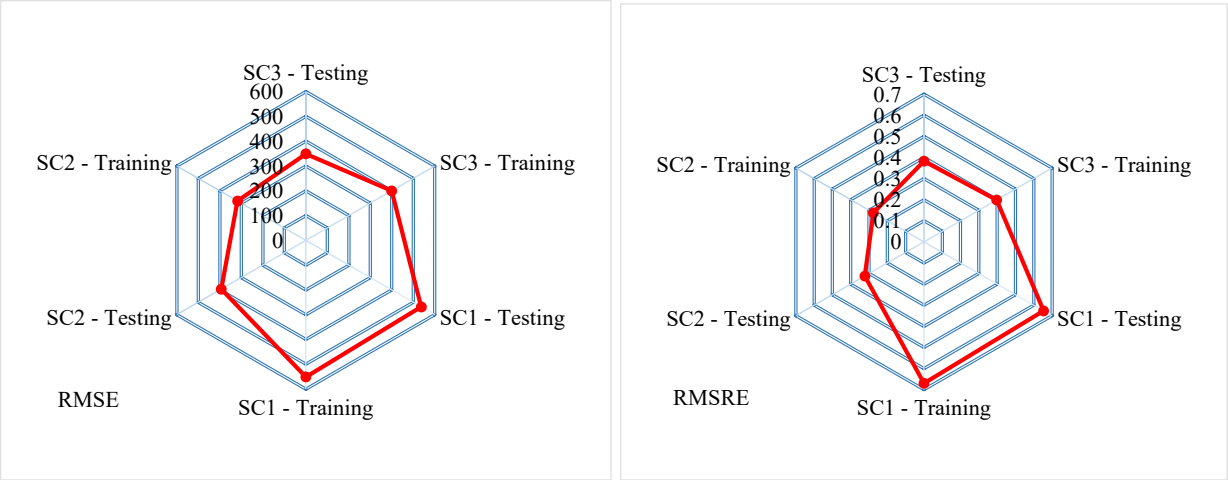


Figure 4.29 - Statistical Indices (RMSE) and (RMSRE) of the New-Gen GMDH Model between SC1, SC2 and SC3.

The comparison of SC1, SC2, and SC3 based on RMSE and RMSRE highlights the varying predictive accuracy of the models. RMSE measures the average magnitude of prediction errors, where lower values indicate higher accuracy. SC1 shows the weakest performance with the highest RMSE values (551.27 training, 537.25 testing), while SC3 performs better (399.05 training, 348.71 testing). SC2, however, achieves the lowest RMSE during training (317.97), and maintains competitive performance in testing (393.51), reflecting strong learning capability and good generalization. RMSRE, a normalized version of RMSE, evaluates prediction errors relative to actual values. SC1 again ranks lowest with the highest RMSRE values (0.67 training, 0.66 testing), while SC3 shows improved consistency (0.40 training, 0.38 testing). SC2 outperforms both with the lowest RMSRE during training (0.28), and a solid testing value (0.32), indicating better model stability and efficiency across different data phases. Similarly, Figure (4.30) displays the MAE and Mean MARE radar plots for the New-Gen GMDH model runs.

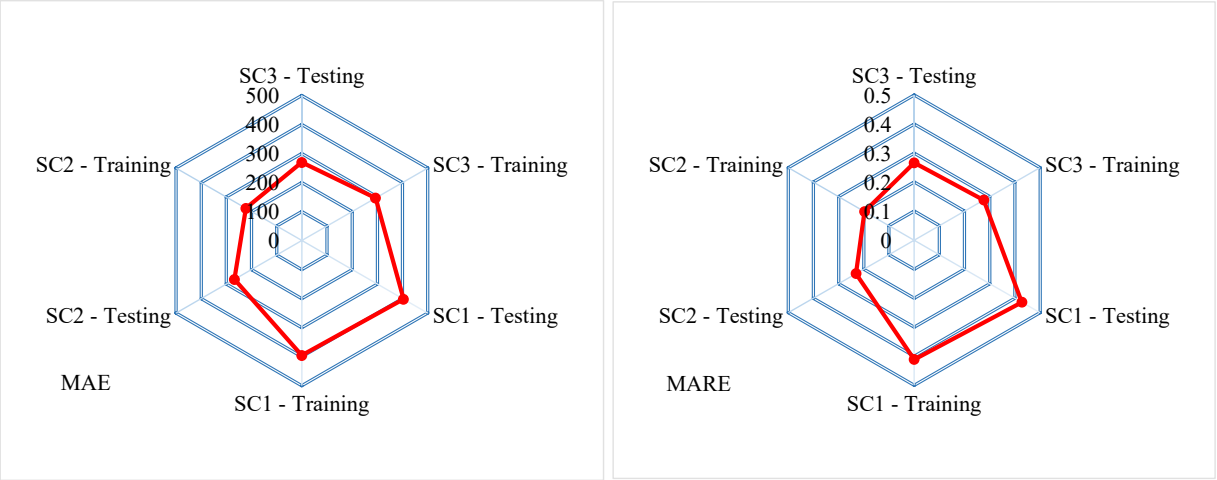


Figure 4.30 - Statistical Indices (MAE) and (MARE) of the New-Gen GMDH Model between SC1, SC2 and SC3.

The comparison of SC1, SC2, and SC3 using MAE and MARE significant differences in model accuracy. MAE measures the average absolute difference between predicted and observed values, with lower values indicating better performance. SC1 performs the weakest, with high MAE values (395.80 training, 404.18 testing). SC3 improves on this (292.83 training, 269.08 testing), but SC2 shows the best results, achieving the lowest MAE in both training (222.09) and testing (267.85), indicating more precise predictions overall. MARE provides a relative measure of prediction error by comparing the absolute error to the actual value, making it useful for assessing performance across different scales. SC1 again ranks lowest with the highest MARE values (0.41 training, 0.43 testing), while SC3 shows moderate improvement (0.28 training, 0.27 testing). SC2 stands out with the lowest MARE values (0.20 training, 0.23 testing), demonstrating strong model efficiency and stability. Figure (4.31) displays the NSE and SI radar plots for the New-Gen GMDH model runs.

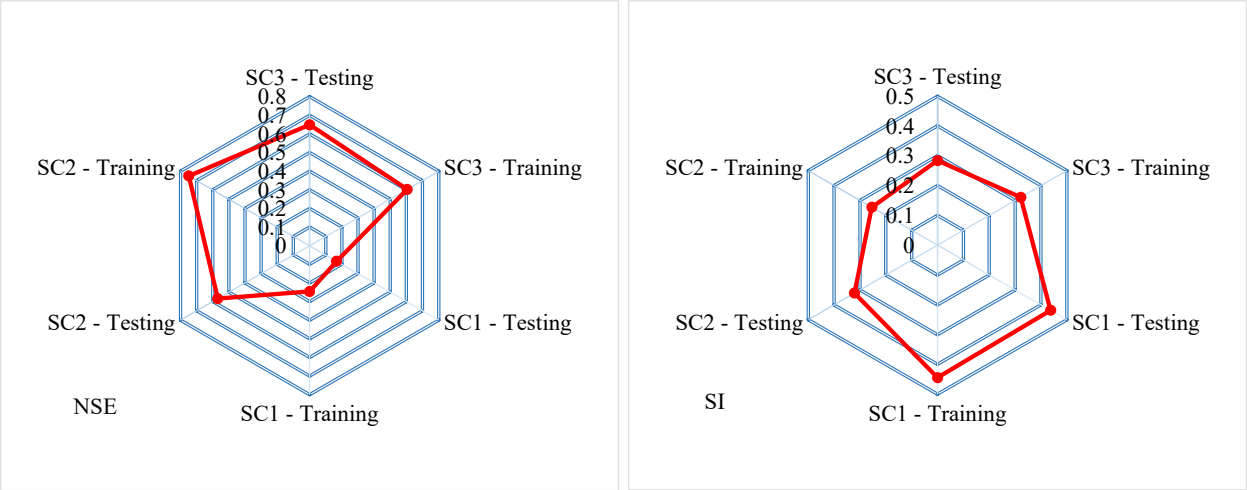


Figure 4.31 - Statistical Indices (NSE) and (SI) of the New-Gen GMDH Model between SC1, SC2 and SC3.

The comparison of SC1, SC2, and SC3 using NSE and SI highlights each model's predictive effectiveness. NSE measures how well the model predicts relative to the mean of observed values, where values closer to 1 indicate better performance. SC1 shows the weakest performance with low NSE values (0.25 training, 0.17 testing), while SC3 performs better (0.61 training, 0.65 testing). SC2 demonstrates the highest NSE in training (0.75) and solid testing performance (0.57), indicating strong model fit and good predictive skill. SI, a normalized error index computed as the ratio of RMSE to the mean of observed values, favors lower values for better accuracy. SC1 records the highest SI values (0.44 training, 0.44 testing), reflecting high prediction dispersion. SC3 improves significantly (0.32 training, 0.28 testing), but SC2 achieves the best SI during training (0.25), with competitive testing performance (0.32), showing its ability to maintain low error while fitting the data well. Finally, Table (4.3) presents the AIC values for the New-Gen GMDH model, comparing SC1, SC2, and SC3. The variation in AIC values highlights the performance differences between the model runs, where lower AIC values indicate a more optimal balance between model simplicity and predictive accuracy.

Table 4.3 - AIC values for the New-Gen GMDH model when predicting discharge variations in the Ottawa River between SC1, SC2 and SC3 for Training and Testing Phases.

Scenario - Phase	AIC Values
SC1 – Training	6885.64
SC1 – Testing	2985.33
SC2 – Training	6240.21
SC2 – Testing	2829.64
SC3 – Training	6617.23
SC3 – Testing	2865.66

The comparison of SC1, SC2, and SC3 based on AIC (Akaike Information Criterion) highlights the trade-off between model accuracy and complexity. AIC estimates the relative quality of statistical models, where lower values indicate a better balance between goodness of fit and model simplicity. SC1 shows the highest AIC values (6885.64 training, 2985.33 testing), suggesting overfitting and lower efficiency. SC3 improves on this with reduced AIC values (6617.23 training, 2865.66 testing). SC2 outperforms both, achieving the lowest AIC in both training (6240.21) and testing (2829.64), indicating it is the most efficient and well-balanced model among the three.

In conclusion, the comprehensive evaluation of SC1, SC2, and SC3 across multiple statistical indicators clearly positions SC2 as the superior model. For error-based metrics, SC2 consistently outperforms with the lowest RMSE (317.97 training, 393.51 testing), RMSRE (0.28 training, 0.32 testing), MAE (222.09 training, 267.85 testing), and MARE (0.20 training, 0.23 testing), demonstrating exceptional predictive accuracy and stability compared to SC1's weakest performance (e.g., RMSE 551.27 training, 537.25 testing) and SC3's moderate results (e.g., RMSE 399.05 training, 348.71 testing). In terms of goodness-of-fit measures, SC2 achieves the highest

R^2 (0.87 training, 0.78 testing), VAF (0.75 training, 0.58 testing), and NSE (0.75 training, 0.57 testing), far surpassing SC1's poor fit (e.g., R^2 0.50 training, 0.42 testing) and SC3's solid but lesser performance (e.g., R^2 0.78 training, 0.81 testing). Additionally, efficiency metrics further favor SC2, with the lowest SI (0.25 training, 0.32 testing) and AIC (6240.21 training, 2829.64 testing), contrasting SC1's inefficiency (e.g., AIC 6885.64 training, 2985.33 testing) and SC3's intermediate balance (e.g., AIC 6617.23 training, 2865.66 testing). While SC3 shows competitive generalization and SC1 consistently underperforms, SC2 stands out as the most accurate, reliable, and efficient model across all analysis techniques.

Chapter 5

Conclusion And Recommendations

5.1 – Conclusion

This study successfully developed and evaluated an integrated framework combining the HEC-RAS with a New-Gen GMDH ML model to predict discharge variations in the Ottawa River under the influence of climate change. The research addressed the pressing need for enhanced flood forecasting and management strategies in a region historically prone to severe flooding, as evidenced by the devastating events of 2017 and 2019, which surpassed 100-year flood thresholds and caused significant socio-economic impacts. By leveraging HEC-RAS's robust hydraulic simulation capabilities and the adaptive, data-driven strengths of the New-Gen GMDH model, this work provided a comprehensive approach to modeling river morphodynamics and flood hazards, offering actionable insights for disaster preparedness and mitigation.

The HEC-RAS simulations, incorporating both 1D and 2D modeling, effectively captured flood extents and water surface elevations across various return periods (50-year, 100-year, 1,000-year, and 10,000-year) and were validated against historical flood events and satellite imagery from 2017, 2019, and 2023. The 2D modeling, in particular, proved superior in delineating flood inundation zones, such as those affecting the town of Quyon, where the 10,000-year flood scenario threatened 78.3% of the population. These results underscored the escalating flood risk under climate change scenarios, driven by increased precipitation intensity and snowmelt, and highlighted the necessity of strategic flood management tailored to different risk levels.

The New-Gen GMDH model further enhanced predictive accuracy by integrating meteorological data (temperature and precipitation) and HEC-RAS-derived discharge outputs.

Comparative analysis of three scenarios revealed that SC2, which incorporated lags along with 4 input variables (Minimum Temperature, Mean Temperature, Maximum Temperature, Total Precipitation, lag 1, lag 2, lag 12) based on ACF values significantly outperformed SC1 which incorporated only five input variables (Minimum Temperature, Mean Temperature, Maximum Temperature, and Total Precipitation) and SC3 which incorporated seven input variables (Minimum Temperature, Mean Temperature, Maximum Temperature, Total Precipitation, Mean Discharge, Maximum Discharge, Minimum Discharge). Statistical indices demonstrated SC2's superiority, with R^2 values of 0.87 (training) and 0.78 (testing), NSE values of 0.75 (training) and 0.57 (testing), and lower error metrics (e.g., RMSE: 317.97 and 393.51; MAE: 222.09 and 267.85) compared to SC1's (e.g., R^2: 0.50 and 0.42; RMSE: 551.27 and 537.25) and SC3 (e.g., R^2: 0.78 and 0.81; RMSE: 399.05 and 348.71). Additionally, SC2's lower AIC values (6240.21 for training and 2829.64 for testing) confirmed its optimal balance of model complexity and predictive power, making it a more reliable tool for discharge forecasting.

This hybrid approach not only improved the accuracy and efficiency of flood predictions but also generated user-friendly polynomial equations, empowering policymakers and flood management authorities with practical tools for decision-making. It further enhances adaptive flood risk management strategies by accounting for uncertainties in climate-driven hydrological variations. By improving the accuracy of flood predictions across different return periods and validating them against historical data, the approach ensures reliability in practical, real-world applications. The study's findings align with the growing need to adapt to climate-induced hydrological shifts, providing a scalable methodology that can be applied to other river systems facing similar challenges. By bridging traditional hydraulic modeling with advanced ML, this

research contributes to the evolving field of flood risk assessment and establishes a foundation for proactive, data-driven flood mitigation strategies along the Ottawa River.

5.2 – Recommendations

To further refine and operationalize the integrated HEC-RAS and New-Gen GMDH framework while strengthening flood management along the Ottawa River, several expanded recommendations are proposed. First, expanding the range of climate change scenarios by incorporating outputs from multiple CMIP6 ensemble models—accounting for diverse Representative Concentration Pathways (RCPs) such as RCP 2.6, 4.5, and 8.5—would better capture the uncertainty and variability of future flood regimes, allowing the model to simulate a broader spectrum of extreme precipitation, temperature shifts, and snowmelt-driven events projected through 2100. Second, further calibration of the New-Gen GMDH model should leverage longer historical datasets (e.g., spanning 100+ years) and additional variables such as soil moisture, snowpack depth, upstream reservoir releases, and ice jam dynamics, which could enhance the model's sensitivity to extreme discharge peaks and short-term fluctuations that were occasionally underestimated in this study; sensitivity analyses could pinpoint the most influential parameters for optimization. Fourth, operationalizing the New-Gen GMDH-derived polynomial equations into a real-time decision-support tool—potentially a web-based or mobile application—would allow stakeholders, including municipal planners and emergency responders, to input live data and receive immediate discharge predictions, facilitating rapid response planning and resource allocation during flood events; this tool could be paired with visualization dashboards to improve accessibility for non-technical users. Additionally, validating the model with emerging climate scenarios from advanced projections like those of the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Dey et al., 2022) would strengthen its relevance for long-term planning

by simulating future shifts in precipitation and temperature patterns. Comparing the New-Gen GMDH's performance against other ML techniques, such as Long Short-Term Memory (LSTM) networks (Zhang et al., 2018), Random Forests (Wang et al., 2015), or Convolutional Neural Networks (Kabir et al., 2020), could further refine the approach, potentially leading to hybrid models that leverage the strengths of multiple algorithms for different flood prediction tasks. Fifth, implementing targeted flood mitigation strategies in high-risk areas like Quyon should prioritize infrastructure upgrades—such as reinforced levees, expanded culverts, and large-scale retention basins—alongside stricter zoning laws to limit development in vulnerable floodplains, while fostering collaborative watershed management across the Ottawa River basin through agreements with upstream municipalities and the Ottawa River Regulation Planning Board to regulate flows and reduce downstream impacts; community-level initiatives, like flood-proofing workshops and vegetative buffer restoration, could further bolster resilience.

Beyond technical enhancements, integrating socio-economic data and urban stormwater dynamics—perhaps through tools like PCSWMM (Manchikatla et al., 2022)—would enrich the flood hazard analysis, offering a more comprehensive view of economic losses and infrastructure vulnerability, especially in urban settings with complex drainage systems. Engaging local stakeholders, including municipal authorities and residents, is equally critical; developing user-friendly decision-support tools based on the GMDH outputs could ensure that the research translates into actionable flood management strategies tailored to community needs. By pursuing these avenues, future research can build on this study's contributions, advancing the synergy of numerical modeling and artificial intelligence to create a more robust, scalable, and community-focused approach to flood risk management in the Ottawa River basin and beyond.

Finally, validating this hybrid methodology across diverse river systems—such as those with contrasting hydrological regimes (e.g., arid, tropical, or glacial-fed basins)—would test its generalizability and identify context-specific adjustments, ensuring its scalability and robustness as a global flood prediction framework; pilot studies in regions like the Indus River Basin or the Mississippi River could provide valuable comparative insights. By pursuing these recommendations, future research and practice can enhance the precision, practicality, and adaptability of this integrated approach, ultimately supporting climate-resilient flood management and safeguarding communities along the Ottawa River and beyond against the escalating risks of climate-driven flooding.

Chapter 6

References

EM-DAT. (2024). EM-DAT: the emergency events database - Université Catholique de Louvain (UCL) – CRED, Brussels, Belgium. Available Online: www.emdat.be (Accessed 7 Jan 2025)

CRED, UNDRR. (2020). The Human Cost of Disasters - An overview of the last 20 years 2000-2019. Available Online: https://reliefweb.int/report/world/human-cost-disasters-overview-last-20-years-2000-2019 (Accessed 7 Jan 2025)

Manzoor, Z.; Ehsan, M.; Khan; M. B.; Manzoor, A.; Akhter, M. M.; Sohail, M. T.; Hussain, A.; Shafi, A.; Abu-Alam, T.; Abioui, M. (2022). Floods and flood management and its socio-economic impact on Pakistan: A review of the empirical literature. *Frontiers in Environmental Science, 10,* 1021862. https://doi.org/10.3389/fenvs.2022.1021862

Jongman, B.; Ward, P. J.; Aerts, J. C. J. H. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change, 22*, 823–835. https://doi.org/10.1016/j.gloenvcha.2012.07.004

Alfieri, L.; Bisselink, B.; Dottori, F.; Naumann, G.; De Roo, A.; Salamon, P.; Wyser, K.; Feyen, L. (2017). Global projections of river flood risk in a warmer world. *Earth's Future*, *5*, 171–182. https://doi.org/10.1002/2016EF000485

Noori, A.; Bonakdari, H. (2023). A GIS-Based Fuzzy Hierarchical Modeling for Flood Susceptibility Mapping: A Case Study in Ontario, Eastern Canada. *Environmental Sciences Proceedings, 25(1)*, 62. https://doi.org/10.3390/ECWS-7-14242

Koks, E. E.; Thissen, M. (2016). A Multiregional Impact Assessment Model for disaster analysis. *Economic Systems Research, 28*, 429–449. https://doi.org/10.1080/09535314.2016.1232701

Carrera, L.; Standardi, G.; Bosello, F.; Mysiak, J. (2015). Assessing direct and indirect economic impacts of a flood event through the integration of spatial and computable general equilibrium modelling. *Environmental Modelling & Software*, 63, 109–122. https://doi.org/10.1016/j.envsoft.2014.09.016

Koks, E. E.; Bočkarjova, M.; De Moel, H.; Aerts, J. C. J. H. (2015). Integrated Direct and Indirect Flood Risk Modeling: Development and Sensitivity Analysis. *Risk Analysis, 35,* 882–900. https://doi.org/10.1111/risa.12300

Sandink, D.; Kovacs, P.; Oulahen, G.; Shrubsole, D. (2016). Public relief and insurance for residential flood losses in Canada: Current status and commentary. *Canadian Water Resources Journal, 41,* 220-237. https://doi.org/10.1080/07011784.2015.1040458

Government of Canada. (2019). Canada's changing climate report. Available Online: https://publications.gc.ca/site/eng/9.870430/publication.html (Accessed 15 Jan 2025)

Insurance Bureau of Canada. (2017). Spring flooding in Ontario and Quebec caused more than $223 million in insured damage. Available Online: http://www.ibc.ca/on/resources/media-centre/media-releases/spring-flooding-in-ontario-and-quebec-caused-morethan-223-million-in-insured-damage (Accessed 8 January 2025)

Québec Gouvernement. (2017). Forum Inondations 2017 : Ensemble planifions l'avenir autrement. Available Online: http://www.mddelcc.gouv.qc.ca/infuseur/communiqueen.asp?no=3856 (Accessed 8 January 2025)

Ottawa Citizen. (2017). Photos: The view from the sky of the massive Ottawa-Gatineau flood. Available Online: https://ottawacitizen.com/news/local-news/photos-the-view-from-the-sky-of-the-massive-ottawa-gatineau-flood (Accessed 28 Feb 2025)

Insurance Bureau of Canada. (2019). Eastern Canada spring flooding caused close to $208 million in Insured Damage. Available Online: http://www.ibc.ca/on/resources/media-centre/media-releases/eastern-canada-spring-flooding-caused-close-to-208-million-in-insured-damage (Accessed 8 January 2025)

Ottawa Riverkeeper. (2025). Available online: https://www.ottawariverkeeper.ca/home/explore-the-river/dams/ (Accessed 8 January 2025)

Ottawa River Regulation Planning Board. (2019). Spring Flood—Questions and Answers; Gatineau, QC, Canada, 1–18. Available Online: https://ottawariver.ca/information/publications/ (Accessed 8 January 2025)

Kirchmeier-Young, M. C.; Wan, H.; Zhang, X. (2021). Anthropogenic Contribution to the Rainfall Associated with the 2019 Ottawa River Flood. *Bulletin of the American Meteorological Society, 102,* S33–S38. https://doi.org/10.1175/BAMS-D-20-0191.1

Cai, J.; Wei, Y.; Yang, J.; Ji, C.; Liu, M.; Fang, W.; Ma, Z.; Bi, J. (2024). A two-stage robust decision-making framework (2S-RDM) for flood risk adaptation under deep uncertainty. *Fundamental Research.* https://doi.org/10.1016/j.fmre.2024.05.005

Aerial Photography. (2019). Available Online: https://aerialphotographs.ca/flood2019/content/IMG_1694_large.html (Accessed 8 Feb 2025).

Bhuiyan, S. A.; Bataille, C. P.; McGrath, H. (2022). Harmonizing and Extending Fragmented 100 Year Flood Hazard Maps in Canada's Capital Region Using Random Forest Classification. *Water, 14(23),* 3801. https://doi.org/10.3390/w14233801

Letessier, C.; Cardi, J.; Dussel, A.; Ebtehaj, I.; Bonakdari, H. (2024). Enhancing Flood Prediction Accuracy through Integration of Meteorological Parameters in River Flow Observations: A Case Study Ottawa River. *Hydrology, 10(8),* 164. https://doi.org/10.3390/hydrology10080164

Szeto, K.; Gysbers, P.; Brimelow, J.; Stewart, R. (2015). The 2014 Extreme Flood on the Southeastern Canadian Prairies. *Bulletin of the American Meteorological Society*, *96*, S20–S24. https://doi.org/10.1175/BAMS-D-15-00110.1

Mohanty, M. P.; Simonovic, S. P. (2021). Changes in floodplain regimes over Canada due to climate change impacts: Observations from CMIP6 models. *Science of The Total Environment, 792,* 148323. https://doi.org/10.1016/j.scitotenv.2021.148323

Mahapatra, A.; Mahammood, V.; Venkatesh, K. (2022). Unsteady flow analysis using hydrological and hydraulic models for real-time flood forecasting in the Vamsadhara river basin. *Journal of Hydroinformatics, 24(6),* 1207–1233. https://doi.org/10.2166/hydro.2022.065

Marimin, N. A.; Razi, M. A. M.; Mustaffa, A. A.; Adnan, M. S.; Rahmat, S. N. (2018). HEC-RAS hydraulic model for floodplain area in Sembrong River. *International Journal of Integrated Engineering, 10(2),* 151–157. https://doi.org/10.30880/ijie.2018.10.02.029

Yang, J.; Townsend, R. D.; Daneshfar, B. (2006). Applying the HEC-RAS model and GIS techniques in river network floodplain delineation. *Canadian Journal of Civil Engineering, 33,* 19–28. https://doi.org/10.1139/l05-102

Parsapour-Moghaddam, P.; Rennie, C. D.; Slaney, J. (2018). Hydrodynamic Simulation of an Irregularly Meandering Gravel-Bed River: Comparison of MIKE 21 FM and Delft3D Flow Models. *Proceedings of the E3S Web Conference, 40,* 02004. https://doi.org/10.1051/e3sconf/20184002004

Parsapour-Moghaddam, P.; Rennie, C. D.; Slaney, J.; Platzek, F.; Shirkhani, H.; Jamieson, E.; Mosselman, E.; Measures, R. (2023). Implementation of a New Bank Erosion Model in Delft3D. *Journal of Hydraulic Engineering*, *149(10),* 04023038. https://doi.org/10.1061/JHEND8.HYENG-13206

Yang, X.; Liu, Q.; He, Y.; Luo, X.; Zhang, X. (2015). Comparison of daily and sub-daily SWAT models for daily streamflow simulation in the Upper Huai River Basin of China. *Stochastic Environmental Research and Risk Assessment, 30,* 959–972. https://doi.org/10.1007/s00477-015-1099-0

Roh, Y.; Heo, G.; Whang, S. E. (2019). A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective. *IEEE Transactions on Knowledge and Data Engineering, 33(4),* 1328–1347. https://doi.org/10.1109/TKDE.2019.2946162

Liu, Z.; Felton, T.; Mostafavi, A. (2024). Interpretable machine learning for predicting urban flash flood hotspots using intertwined land and built-environment features. *Computers, Environment and Urban Systems, 110,* 102096. https://doi.org/10.1016/j.compenvurbsys.2024.102096

Liu, Z.; Shi, W.; Yu, Y.; Chen, P.; Chen, B. Y. (2022). A LSTM-based approach for modelling the movement uncertainty of indoor trajectories with mobile sensing data. *International Journal of Applied Earth Observation and Geoinformation, 108*, 102758. https://doi.org/10.1016/j.jag.2022.102758

Federal Emergency Management Agency. (2002). HEC-RAS Procedures for HEC-2 Modelers. Available Online: https://www.fema.gov/sites/default/files/documents/fema_floodplain-modeling-manual_hec-ras-procedures-hec-2-modelers_4-2002.pdf (Accessed 15 January 2025)

Ghorpade, P.; Gadge, A.; Lende, A.; Chordiya, H.; Gosavi, G.; Mishra, A.; Hooli, B.; Ingle, Y. S.; Shaikh, N. (2021). Flood Forecasting Using Machine Learning: A Review. *In Proceedings of the 8th International Conference on Smart Computing and Communications (ICSCC),* pp. 32–36. https://doi.org/10.1109/ICSCC51209.2021.9528099

Ehsan, S.; Yasir, Q. M.; Awais, R. M.; Sattar, T.; Mustafa, G.; Fida, U. (2024). Modelling of Riverine Flooding Using HEC-RAS-A Case Study of Badri Khawar River in KPK, Pakistan. *Jurnal Kejuruteraan, 36(1),* 169–177. https://doi.org/10.17576/jkukm-2024-36(1)-16

Zainal, N. N.; Talib, S. H. A. (2024). Review paper on applications of the HEC-RAS model for flooding, agriculture, and water quality simulation. *Water Practice & Technology, 19(7),* 2883–2900. https://doi.org/10.2166/wpt.2024.173

Singh, S. K.; Kanga, S.; Đurin, B.; Kranjčić, N.; Chaurasia, R.; Markovinović, D. (2021). FLOOD RISK MODELING USING HEC-RAS AND GEOSPATIAL TECHNIQUES. *e-Zbornik Electronic Collection of Papers of the faculty of Civil Engineering, 11(22)*, 20–36. https://doi.org/10.47960/2232-9080.2021.22.11.20

Villazon, M. F.; Aranibar, C., Willems, P. (2009). Hydrodynamics unsteady flow model applied to river Piraí for flood events estimation in Santa Cruz – Bolivia. *In Proceedings of the 1st International Congress of Hydroclimatology, 1.* https://lirias.kuleuven.be/retrieve/96213

Agrawal, R. P.; Regulwar, D. G. (2016). Unsteady Flow Analysis of Lower Dudhana River using HEC-RAS. *In Proceedings of the 3rd National Conference "Sustainable Water Resources Development and Management SWARDAM-2016", 3.* https://www.academia.edu/27636431/Unsteady_Flow_Analysis_of_Lower_Dudhana_River_using_HEC_RAS

Stoleriu, C. C.; Urzica, A.; Mihu-Pintilie, A. (2019). Improving flood risk map accuracy using high-density LiDAR data and the HEC-RAS river analysis system: A case study from north-eastern Romania. *Journal of Flood Risk Management, 13(1),* e12572. https://doi.org/10.1111/jfr3.12572

Faudzi, S. M. M.; Abustan, I.; Sabtu, N.; Wahab, M. K. A. (2019). Flood Hazard Map of Upper Sungai Langat Using HEC-RAS. *In Proceedings of the AICCE'19. Lecture Notes in Civil Engineering, 53.* https://doi-org.proxy.bib.uottawa.ca/10.1007/978-3-030-32816-0_125

Filianoti, P.; Gurnari, L.; Zema, D. A.; Bombino, G.; Sinagra, M.; Tucciarelli, T. (2020). *An Evaluation Matrix to Compare Computer Hydrological Models for Flood Predictions. Hydrology, 7,* 42. https://doi.org/10.3390/hydrology7030042

Khan, M.; Uzair, A.; Ali, U.; Khan, W. (2020). Flood modeling of Naray-Khwar using HEC-RAS. *Journal of Himalayan Earth Sciences, 53(1),* 1-11. https://doi.org/10.13140/RG.2.2.21003.92964

Daham, M. H.; Abed, B. S. (2020). One and Two-Dimensional Hydraulic Simulation of a Reach in Al-Gharraf River. *Journal of Engineering. 26(7),* 28-44. https://doi.org/10.31026/j.eng.2020.07.03

Brunner, G.W. (2024). HEC-RAS Hydraulic Reference Manual Version 6.6. Hydraulic Engineering Centre, Davis, CA. Available Online: https://www.hec.usace.army.mil/confluence/rasdocs/ras1dtechref/latest (Accessed 15 January 2025)

Afzal, M. A.; Ali, S.; Nazeer, A.; Khan, M. I.; Waqas, M. M.; Aslam, R. A.; Cheema, M. J. M.; Nadeem, M.; Saddique, N.; Muzammil, M; Shah, A. N. (2022). Flood Inundation Modeling by Integrating HEC–RAS and Satellite Imagery: A Case Study of the Indus River Basin. *Water, 14(19),* 2984; https://doi.org/10.3390/w14192984

Aryal, A.; Acharya, A; Kalra, A. (2022). Assessing the Implication of Climate Change to Forecast Future Flood Using CMIP6 Climate Projections and HEC-RAS Modeling. *Forecasting, 4(3),* 582-603. https://doi.org/10.3390/forecast4030032

Bruno, L. S.; Mattos, T. S.; Oliveira, P. T. S.; Almagro, A.; Rodrigues, D. B. B. (2022). Hydrological and Hydraulic Modeling Applied to Flash Flood Events in a Small Urban Stream. *Hydrology, 9,* 223. https://doi.org/10.3390/hydrology9120223

Namara, W. G.; Damisse, T. A.; Tufa, F. G. (2022). Application of HEC-RAS and HEC-GeoRAS model for Flood Inundation Mapping, the case of Awash Bello Flood Plain, Upper Awash River Basin, Oromiya Regional State, Ethiopia. *Modeling Earth Systems and Environment, 8,* 1449–1460. https://doi.org/10.1007/s40808-021-01166-9

Dau, Q. V.; Wang, X.; Aziz, F.; Nawaz, R. A.; Pang, T.; Mahmood, M. Q.; Fortin, M. (2024). Pluvial flood modeling for coastal areas under future climate change – A case study for Prince Edward Island, Canada. *Journal of Hydrology, 641,* 131769. https://doi.org/10.1016/j.jhydrol.2024.131769

Yang, P.; Law, A. W. K.; Xu, S.; Sim, S. T. V.; Chan, H.; Chitwatkulsiri, D.; Irvine, K. N. (2023). Assessment of compound flooding through seamless linkage of coastal hydrodynamic and inland catchment models. *Journal of Hydro-environment Research, 46,* 31–43. https://doi.org/10.1016/j.jher.2022.11.001

Priya, R.; Kumar, R. S.; Naiklal, S. K.; Kumar, A. (2020). Flood Inundation Modelling for Tungabhadra Basin using HEC-RAS: A Case Study for Haralahalli Discharge Site. *Geographical Analysis, 9(2),* 33–41. https://doi.org/10.53989/bu.ga.v9i2.8

Azouagh, A.; Bardai, R. E.; Hilal, I.; Messari, J. S. E. (2018). Integration of GIS and HEC-RAS in Floods Modeling of Martil River (Northern Morocco). *European Scientific Journal, 14(12),* 130. https://doi.org/10.19044/esj.2018.v14n12p130

Lago, C. A. F. D.; Giacomoni, M. H.; Bentivoglio, R.; Taormina, R.; Junior, M. N. G.; Mendiondo, E. M. (2023). Generalizing rapid flood predictions to unseen urban catchments with conditional generative adversarial networks. *Journal of Hydrology, 618,* 129276. https://doi.org/10.1016/j.jhydrol.2023.129276

Idfi, G.; Wahyono, I. D.; Yulistyorini, A.; Khomsiati, N. L. (2019). The comparative study of flood modelling with the unsteady and the steady flow on Ngotok river. *In Proceedings of the IOP*

Conference Series: Materials Science and Engineering, 669(1), 012018. https://doi.org/10.1088/1757-899x/669/1/012018

Soltani, K.; Ebtehaj, I.; Amiri, A.; Azari, A.; Gharabaghi, B.; Bonakdari, H. (2021). Mapping the spatial and temporal variability of flood susceptibility using remotely sensed normalized difference vegetation index and the forecasted changes in the future. *Science of the Total Environment, 770,* 145288. https://doi.org/10.1016/j.scitotenv.2021.145288

Grégoire, G.; Fortin, J.; Ebtehaj, I.; Bonakdari, H. (2023). Forecasting Pesticide Use on Golf Courses by Integration of Deep Learning and Decision Tree Techniques. *Agriculture, 13(6),* 1163. https://doi.org/10.3390/agriculture13061163

Ebtehaj, I.; Bonakdari, H. (2022). A reliable hybrid outlier robust non-tuned rapid machine learning model for multi-step ahead flood forecasting in Quebec, Canada. *Journal of Hydrology, 614,* 128592. https://doi.org/10.1016/j.jhydrol.2022.128592

Gholami, A.; Bonakdari, H.; Ebtehaj, I.; Shaghaghi, S.; Khoshbin, F. (2016). Developing an expert group method of data handling system for predicting the geometry of a stable channel with a gravel bed. *Earth Surface Processes and Landforms, 42,* 1460–1471. https://doi.org/10.1002/esp.4104

Ebtehaj, I.; Bonakdari, H. (2016). Bed load sediment transport estimation in a clean pipe using multilayer perceptron with different training algorithms. *KSCE Journal of Civil Engineering, 20,* 581–589. https://doi.org/10.1007/s12205-015-0630-7

Ebtehaj, I.; Bonakdari, H. (2021). Discussion of "Model Development for Estimation of Sediment Removal Efficiency of Settling Basins Using Group Methods of Data Handling" by Faisal Ahmad, Mujib Ahmad Ansari, Ajmal Hussain, and Jahangeer Jahangeer. *Journal of Irrigation and Drainage Engineering, 147,* 07021021. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001638

Bonakdari, H.; Ebtehaj, I. (2021). Discussion of "Time-Series Prediction of Streamflows of Malaysian Rivers Using Data-Driven Techniques" by Siraj Muhammed Pandhiani, Parveen Sihag, Ani Bin Shabri, Balraj Singh, and Quoc Bao Pham. *Journal of Irrigation and Drainage Engineering, 147,* 07021014. https://doi.org/10.1061/(ASCE)IR.1943-4774.00016

Ebtehaj, I.; Bonakdari, H. (2021). Discussion of "Comparative Study of Time Series Models, Support Vector Machines, and GMDH in Forecasting Long-Term Evapotranspiration Rates in Northern Iran" by Afshin Ashrafzadeh, Ozgur Kişi, Pouya Aghelpour, Seyed Mostafa Biazar, and Mohammadreza Askarizad Masouleh. *Journal of Irrigation and Drainage Engineering, 147,* 07021005. https://doi.org/10.1061/(ASCE)IR.1943-4774.000156

Ebtehaj, I.; Zeynoddin, M.; Bonakdari, H. (2020). Discussion of "Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: A local and external data analysis approach" by Saeid Mehdizadeh, Farshad Fathian, Mir Jafar Sadegh Safari and Jan F. Adamowski. *Journal of Hydrology, 583,* 124614. https://doi.org/10.1016/j.jhydrol.2020.124614

Hosseiny, H.; Nazari, F.; Smith, V.; Nataraj, C. (2020). A Framework for Modeling Flood Depth Using a Hybrid of Hydraulics and Machine Learning. *Scientific Reports, 10,* 8222. https://doi.org/10.1038/s41598-020-65232-5

Hao, C.; Yunus, A. P.; Subramanian, S. S.; Avtar, R. (2021). Basin-wide flood depth and exposure mapping from SAR images and machine learning models. *Journal of Environmental Management, 297,* 113367. https://doi.org/10.1016/j.jenvman.2021.113367

Hosseiny, H. (2021). A deep learning model for predicting river flood depth and extent. *Environmental Modelling & Software, 145,* 105186. https://doi.org/10.1016/j.envsoft.2021.105186

Kabir, S.; Patidar, S.; Pender, G. (2021). A Machine Learning Approach for Forecasting and Visualising Flood Inundation Information. *In Proceedings of the Institution of Civil Engineers - Water Management*, *174(1),* 27-41. https://doi.org/10.1680/jwama.20.00002

Tamiru, H.; Dinka, M. O. (2021). Application of ANN and HEC-RAS model for flood inundation mapping in lower Baro Akobo River Basin, Ethiopia. *Journal of Hydrology: Regional Studies, 36,* 100855. https://doi.org/10.1016/j.ejrh.2021.100855

Rozos, E.; Leandro, J.; Koutsoyiannis, D. (2022). Development of Rating Curves: Machine Learning vs. Statistical Methods. *Hydrology 2022, 9(10),* 166. https://doi.org/10.3390/hydrology9100166

Li, M.; McGrath, H.; Stefanakis, E. (2022). Multi-Scale Flood Mapping under Climate Change Scenarios in Hexagonal Discrete Global Grids. *ISPRS International Journal of Geo-Informatics, 11*, 627. https://doi.org/10.3390/ijgi11120627

Kumar, S.; Kumar, B.; Deshpande, V.; Agarwal, M. (2023). Predicting flow velocity in a vegetative alluvial channel using standalone and hybrid machine learning techniques. *Expert Systems with Applications, 232,* 120885. https://doi.org/10.1016/j.eswa.2023.120885

Khoshkonesh, A.; Nazari, R.; Nikoo, M. R.; Karimi, M. (2024). Enhancing flood risk assessment in urban areas by integrating hydrodynamic models and machine learning techniques. *Science of The Total Environment, 965,* 175859. https://doi.org/10.1016/j.scitotenv.2024.175859

Sorboni, N. G.; Wang, J.; Najafi, M.R. (2024). Urban flood mapping using Sentinel-1 and RADARSAT Constellation Mission image and convolutional Siamese network. *Natural Hazards 120,* 5711–5742. https://doi.org/10.1007/s11069-024-06434-2

Palash, W.; Akanda, A. S.; Islam, S. (2024). A data-driven global flood forecasting system for medium to large rivers. *Scientific Reports, 14,* 8979. https://doi.org/10.1038/s41598-024-59145-w

Ivakhnenko, A. (1978). The group method of data handling in long-range forecasting. *Technological Forecasting and Social Change, 12(2-3)*, 213–227. https://doi.org/10.1016/0040-1625(78)90057-4

Ebtehaj, I.; Bonakdari, H. (2022). Early Detection of River Flooding Using Machine Learning for the Sain-Charles River, Quebec, Canada. In Proceedings of the 39th IAHR World Congress. https://doi.org/10.3850/IAHR-39WC252171192022627

Walton, R.; Binns, A.; Bonakdari, H.; Ebtehaj, I.; Gharabaghi, B. (2019). Estimating 2-year flood flows using the generalized structure of the Group Method of Data Handling. *Journal of Hydrology, 575,* 671–689. https://doi.org/10.1016/j.jhydrol.2019.05.068

Pazuki, G.; Kakhki, S. S. (2013). A hybrid GMDH neural network to investigate partition coefficients of Penicillin G Acylase in polymer–salt aqueous two-phase systems. *Journal of Molecular Liquids, 188,* 131–135. https://doi.org/10.1016/j.molliq.2013.10.001

Huang, W.; Du, Y.; Ren, H.; Guo, J.; Wang, R.; Wang, Z.; Zhao, L.; Hao, Z. (2020). Application of modified GMDH network for CO_2-oil minimum miscibility pressure prediction. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 42(16),* 2049–2062. https://doi.org/10.1080/15567036.2019.1604908

Safari, M. J. S.; Ebtehaj, I.; Bonakdari, H.; Es-haghi, M. S. (2019). Sediment transport modeling in rigid boundary open channels using generalize structure of group method of data handling. *Journal of Hydrology, 577,* 123951. https://doi.org/10.1016/j.jhydrol.2019.123951

Elkurdy, M.; Binns, A. D.; Bonakdari, H.; Gharabaghi, B.; McBean, E. (2022). Early detection of riverine flooding events using the group method of data handling for the Bow River, Alberta, Canada. *International Journal of River Basin Management, 20(4),* 533–544. https://doi.org/10.1080/15715124.2021.1906261

Bonakdari, H.; Gholami, A.; Ebtehaj, I.; Gharebaghi, B. (2022). An Improved Architecture of Group Method of Data Handling for Stability Evaluation of Cross-sectional Bank on Alluvial Threshold Channels. *In Proceedings of the Intelligent Computing Conference, Lecture Notes in Networks and Systems, 506.* https://doi.org/10.1007/978-3-031-10461-9_53

The Oaklahoman. (2017). Floods drive nearly 1,900 from homes in Canada's Quebec. Available Online: https://www.oklahoman.com/picture-gallery/weather/2017/05/08/floods-drive-nearly-1900-from-homes-in-canadas-quebec/5858471007/ (Accessed 22 January 2025)

Canadian Broadcasting Corporation. (2019). Quyon, Que., residents return home after forced evacuation. Available Online: https://www.cbc.ca/news/canada/ottawa/quyon-residents-return-home-after-mandatory-flood-evacuation-1.5123796 (Accessed 22 January 2025)

La Radio Du Pontiac. (2019). Flood in Quyon: temporary closure of the ferry. Available Online https://chipfm.com/en/flood-in-quyon-temporary-closure-of-the-ferry (Accessed 22 January 2025)

Pontiac Journal. (2023). Flooding concerns continue. Available Online: https://pontiacjournal.com/flooding-concerns-continue/ (Accessed 22 January 2025)

Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; Martens, B.; Miralles, D. G.; Piles, M.; Rodríguez-Fernández, N. J.; Zsoter, E.; Buontempo, C.; Thépaut, J. N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data, 13(9),* 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Büchele, B.; Kreibich, H.; Kron, A.; Thieken, A.; Ihringer, J.; Oberle, P.; Merz, B.; Nestmann, F. (2006). Flood-risk mapping: contributions towards an enhanced assessment of extreme events and associated risks. *Natural Hazards and Earth System Sciences, 6(4),* 485-503. https://doi.org/10.5194/nhess-6-485-2006

Teng, J.; Jakeman, A. J.; Vaze, J.; Croke, B. F. W.; Dutta, D.; Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software, 90,* 201–216. https://doi.org/10.1016/j.envsoft.2017.01.006

Dhungel, S.; Barber, M. E.; Mahler, R. L. (2019). Comparison of one- and two-dimensional flood modeling in urban environments. *International Journal of Sustainable Development and Planning, 14(4),* 356–366. https://doi.org/10.2495/SDP-V14-N4-356-366

Zaghloul, M. A.; Elshorbagy, A.; Papalexiou, A. M. (2024). Enhancing regional flood frequency analysis by integrating site-similarity measures with watershed modeling. *Journal of Hydrology, 641,* 131754. https://doi.org/10.1016/j.jhydrol.2024.131754

Chung, E. S.; Kim, S. U. (2013). Bayesian rainfall frequency analysis with extreme value using the informative prior distribution. *KSCE Journal of Civil Engineering, 17(6),* 1502-1514. https://doi.org/10.1007/s12205-013-0189-0

Blöschl, G.; Hall, J.; Parajka, J.; Perdigão, R. A. P.; Merz, B.; Arheimer, B.; Aronica, G. T.; Bilibashi, A.; Bonacci, O.; Borga, M.; Čanjevac, I.; Castellarin, A.; Chirico, G. B.; Claps, P.; Fiala, K.; Frolova, N.; Gorbachova, L.; Gül, A; Hannaford, J.; Harrigan, S.; Kireeva, M.; Kiss, A.; Kjeldsen, T. R.; Kohnová, S.; Koskela, J. J.; Ledvinka, O.; Macdonald, N.; Mavrova-Guirguinova, M.; Mediero, L.; Merz, R.; Molnar, P.; Montanari, A.; Murphy, C.; Osuch, M.; Ovcharuk, V.; Radevski, I.; Rogger, M.; Salinas, J. L.; Sauquet, E.; Šraj, M.; Szolgay, J.; Viglione, A.; Volpi, E.; Wilson, D.; Zaimi, K.; Živković, N. (2017). Changing climate shifts timing of European floods. *Science, 357(6351),* 588-590. https://doi.org/10.1126/science.aan2506

Haktanir, T. (1992). Comparison of various flood frequency distributions using annual flood peaks data of rivers in Anatolia. *Journal of Hydrology, 136(1-4),* 1-31. https://doi.org/10.1016/0022-1694(92)90002-D

Perez, G.; Coon, E. T.; Rathore, S. S.; Le, P. V. V. (2024). Advancing process-based flood frequency analysis for assessing flood hazard and population flood exposure. *Journal of Hydrology, 639,* 131620. https://doi.org/10.1016/j.jhydrol.2024.131620

United States Geological Survey. (2023). Flood-frequency analysis in the Midwest: Addressing potential nonstationarity of annual peak-flow records. Available Online: https://www.usgs.gov/publications/flood-frequency-analysis-midwest-addressing-potential-nonstationarity-annual-peak-flow (Accessed 23 January 2025)

United States Geological Survey. (2023). Guidelines for Determining Flood Flow Frequency Bulletin 17C. Available Online: https://pubs.usgs.gov/tm/04/b05/tm4b5.pdf (Accessed 23 January 2025)

Griffis, V. W.; Stedinger, J. R. (2007). Log-Pearson Type 3 Distribution and Its Application in Flood Frequency Analysis. I: Distribution Characteristics. *Journal of Hydrologic Engineering, 12(5),* 482-491. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(482)

Massey Jr., F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association, 46(253),* 68-78. https://doi.org/10.1080/01621459.1951.10500769

Evans, D. L.; Drew, J. H.; Leemis, L. M. (2008). The Distribution of the Kolmogorov–Smirnov, Cramer–von Mises, and Anderson–Darling Test Statistics for Exponential Populations with Estimated Parameters. Communication in Statistics- Simulation and Computation, 37(7), 1396-1421. https://doi.org/10.1080/03610910801983160

Razali, N. M.; Yap B. W. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics, 2(1),* 21-33. Available Online: https://www.nrc.gov/docs/ml1714/ml17143a100.pdf (Accessed 25 January 2025)

Karson, M. (1968). Handbook of Methods of Applied Statistics. Volume I: Techniques of Computation Descriptive Methods, and Statistical Inference. Volume II: Planning of Surveys and Experiments. I. M. Chakravarti, R. G. Laha, and J. Roy, New York, John Wiley; 1967, $9.00. *Journal of the American Statistical Association,63(323),* 1047-1049. https://doi.org/10.1080/01621459.1968.11009335

Samantaray, S.; Sahoo, A. (2020). Estimation of flood frequency using statistical method: Mahanadi River basin, India. *H2Open Journal, 3(1),* 189–207. https://doi.org/10.2166/h2oj.2020.004

Chen, X.; Shao, Q.; Xu, C. Y.; Zhang, J.; Zhang, L.; Ye, C. (2017). Comparative Study on the Selection Criteria for Fitting Flood Frequency Distribution Models with Emphasis on Upper-Tail Behavior. *Water, 9(5).* 320. https://doi.org/10.3390/w9050320

Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association, 69(347),* 730-737. https://doi.org/10.1080/01621459.1974.10480196

Ghaderi, K.; Motamedvaziri, B.; Vafakhah, M.; Dehghani, A. A. (2019). Regional flood frequency modeling: a comparative study among several data-driven models. *Arabian Journal of Geosciences, 12,* 588. https://doi.org/10.1007/s12517-019-4756-7

Greenwood, P. E.; Nikulin, M. S. (1996). A Guide to Chi-Squared Testing. Wiley, New York. ISBN: 978-0-471-55779-1. https://www.wiley.com/en-br/A+Guide+to+Chi-Squared+Testing-p-9780471557791

McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica, 23(2),* 143-149. https://doi.org/10.11613/BM.2013.018

Dodangeh, E.; Panahi, M.; Rezaie, F.; Lee, S.; Bui, D. T.; Lee, C. W.; Pradhan, B. (2020). Novel hybrid intelligence models for flood-susceptibility prediction: Meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search. *Journal of Hydrology, 590,* 125423. https://doi.org/10.1016/j.jhydrol.2020.125423

Ebtehaj, I.; Bonakdari, H.; Zaji, A. H.; Azimi, H.; Khoshbin, F. (2015). GMDH-type neural network approach for modeling the discharge coefficient of rectangular sharp-crested side weirs. *Engineering Science and Technology, an International Journal, 18(4),* 746–757. https://doi.org/10.1016/j.jestch.2015.04.012

Rincón, D.; Khan, U. T.; Armenakis, C. (2018). Flood Risk Mapping Using GIS and Multi-Criteria Analysis: A Greater Toronto Area Case Study. *Geosciences, 8(8), 275*. https://doi.org/10.3390/geosciences8080275

Suthakaran, S.; Jayakody, S.; Subasinghe, S.; Seneviratne, N.; Alahakoon, R. (2022). Mapping the Flood Risk Exposure Using Open-Source Geospatial Tools and Techniques: A Case of Gampaha Divisional Secretariat Division, Sri Lanka. *Journal of Geoscience and Environment Protection, 10,* 18-31. https://doi.org/10.4236/gep.2022.1010002

Qureshi, M. U. A.; Amiri, A.; Ebtehaj, I.; Guimere, S. J.; Cunderlik, J.; Bonakdari, H. (2025). Coupling HEC-RAS and AI for River Morphodynamics Assessment Under Changing Flow Regimes: Enhancing Disaster Preparedness for the Ottawa River. *Hydrology, 12(2),* 25. https://doi.org/10.3390/hydrology12020025

Maxim Fortin. (2021). Open Building Population Layer (Canada) - 2021. Available Online: https://maximfortin.com/project/obpl-ca-2021/#:~:text=The%20Open%20Building%20Population%20Layer,completeness%20of%20the%20information%20provided (Accessed 21 February 2025)

Dey, A.; Sahoo, D. P.; Kumar, R.; Remesan, R. (2022). A multimodel ensemble machine learning approach for CMIP6 climate model projections in an Indian River basin. *International Journal of Climatology, 42(16),* 9215-9236. https://doi.org/10.1002/joc.7813

Zhang, D.; Hølland, E. S.; Lindholm, G.; Ratnaweera, H. (2018). Hydraulic modeling and deep learning based flow forecasting for optimizing inter catchment wastewater transfer. *Journal of Hydrology, 567,* 792-802. https://doi.org/10.1016/j.jhydrol.2017.11.029

Wang, Z.; Lai, C.; Chen, X.; Yang, B.; Zhao, S.; Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology, 527,* 1130-1141. https://doi.org/10.1016/j.jhydrol.2015.06.008

Kabir, S.; Patidar, S.; Xia, X.; Liang, Q.; Neal, J.; Pender, G. (2020). A deep convolutional neural network model for rapid prediction of fluvial flood inundation. *Journal of Hydrology, 590,* 125481. https://doi.org/10.1016/j.jhydrol.2020.125481

Manchikatla, S. K.; Umamahesh, N. V. (2022). Simulation of flood hazard, prioritization of critical sub-catchments, and resilience study in an urban setting using PCSWMM: a case study. *Water Policy, 24(8),* 1247-1268. https://doi.org/10.2166/wp.2022.291