TECHNICAL REPORT 19

July 2023

OWP | OFFICE OF WATER PREDICTION

CUAHSI
allied for water science
www.cuahsi.org

National Water Center Innovators Program
Summer Institute Report 2023

National Water Center Innovators Program
Summer Institute Report 2023

Editors:

Mark Wang

Ebrahim Hamidi

Deanna McCay

Prepared in cooperation with the Consortium of Universities for the Advancement of Hydrologic Science, Inc., and the National Water Center.

CUAHSI Technical Report No. 19

Version 1.0

July 2023

Suggested Citation:

Wang, M., E. Hamidi, et al. (2023). National Water Center Innovators Program - Summer Institute, CUAHSI Technical Report, HydroShare.

Contents

Preface

This report summarizes the research conducted during the 2023 National Water Center Innovators Program Summer Institute (NWC-SI). The NWC-SI is the result of a partnership between the National Weather Service's Office of Water Prediction (OWP) and the Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI). OWP is a part of the National Oceanic and Atmospheric Administration (NOAA) while CUAHSI is a 501(c)(3) nonprofit organization with a mission "to empower the water community and advance science through collaboration, infrastructure, and education." The partnership between OWP and CUAHSI is facilitated by the Cooperative Institute for Research to Operations in Hydrology (CIROH).

The NWC-SI has been held annually at the National Water Center (NWC) on the campus of University of Alabama since 2015. Throughout these Summer Institutes, the brightest minds in the flooding space—graduate student fellows, senior academic faculty, federal scientists, staff and contractors—have converged in Tuscaloosa to tackle ambitious scientific projects aimed at enhancing the United States' water resources modeling, science, and services. For the past two years, NWC-SI projects have focused on supporting the Next Generation National Water Model (NextGen) framework, which is a major update to the National Water Model (NWM). The NextGen framework refactors the NWM to accept different physical models for different watersheds, simply put, allowing for the most appropriate model to be selected on a watershed-by-watershed basis at the continental scale.

Synergy has emerged from this year's NWC-SI through the concerted efforts of six motivated groups of fellows, dedicated theme leaders, and collaborators across multiple organizations. Over the years, this collaborative spirit has fostered not only groundbreaking research but also enduring friendships and professional networks. This year, each project fell within one of four themes: (1) Hydro-Data Science for NextGen: Methods to improve streamflow forecasting accuracy, (2) Urban Flooding Under Climate Change, (3) Real-Time Urban Flooding Awareness, and (4) Channel Flow Routing and Flood Inundation Mapping. Each theme is carefully defined during the months leading up to the NWC-SI to align with NWC goals and the expertise of faculty mentors (theme leads).

During the program's first two weeks, fellows were introduced to the themes, CUAHSI, and the NWC. They received training on the Nextgen framework, collaborative resources, software version control, data science, and project management. Fellows got to know each other and the theme leads, and organized their own teams by the end of week 1. The fellows also connected with a diverse array of stakeholders: Holden Smith of USGS led a field visit to measure discharge at the Cahaba River, and Pete Cichetti, a director of the NYS Office of Emergency Management, spoke about the importance of flood inundation mapping (FIM). Throughout the seven-week program, the fellows' dedication and hard work led to remarkable progress and innovative findings. Their capstone presentations showcased their accomplishments and highlighted the strides made in the understanding of streamflow forecasting, high-resolution urban flood mapping, and the Height Above Nearest Drainage (HAND) method that underpins the FIM generated with the NWM.

Fellows

The eighth NWC-SI cohort consists of 23 students pursuing master's and doctoral degrees from 15 universities across the United States. The fellows hail from various academic departments including civil engineering, geography, Earth sciences, and biological engineering. The diversity of fellows encourages robust and unique projects and is an integral aspect of the NWC-SI.

Themes and Theme Leads

The NWC-SI 2023 themes and theme leads were:

- The "Hydro-Data science for NextGen" theme, led by Jonathan Frame (Floodbase). Additional technical support was provided by Fred Ogden (NOAA), James Halgren (CIROH), Mike Johnson (NOAA), Tony Castronova (CUAHSI), Irene Garousi-Nejad (CUAHSI), Nels Frazier (NOAA), Luciana Kindl da Cunha (NOAA), Zach Wills (CIROH), Arpita Patel (CIROH), Peter La Follette (NOAA), Ahmad Jan (NOAA), and Hamid Moradkhani (University of Alabama).

- The "Urban Flooding Under Climate Change" theme was led by Kyle Mandli (Columbia University). Additional technical support was provided by Fred Ogden (NOAA), Tony Castronova (CUAHSI), Irene Garousi-Nejad (CUAHSI), and Ebrahim Hamidi (University of Alabama).

- The "Real-Time Urban Flooding Awareness" theme was led by Barbara Minsker (Southern Methodist University). Additional technical support was provided by Sammy Rivera Aparicio (Oregon State University), Bradford Bates (NOAA), Tyler Schrag (NOAA), Jonathan Frame (Floodbase Inc.), Anthony Castronova (CUAHSI), Tadd Bindas (Penn State University), and Jeremy Rapp (Michigan State University).

- The "Channel Flow Routing and Flood Inundation Modeling" theme was led by Sagy Cohen (University of Alabama) and Ehab Meselhe (Tulane University). Additional technical support was provided by David Weiss (CIROH), Brad Bates (NOAA), Robert Hanna (NOAA), and Anupul Baruah (University of Alabama).

Project Summaries

The 2023 NWC-SI projects are summarized below. Chapters 1-6 present the complete reports.

1. Projects within the Hydro-Data Science for NextGen theme:

"Data Assimilation for Improving Forecast Accuracy and Streamflow Prediction in Ungauged Basins" (Ch. 1) incorporates data assimilation techniques into the NextGen framework to improve the accuracy of streamflow predictions. Predictive capabilities of the Conceptual Functional Equivalent (CFE) model were enhanced by utilizing an ensemble Kalman filter and Muskingum-Cunge routing scheme in California's Carmel watershed.

"On numerical methods and differentiable modeling for soil process representations in the NextGen Framework in arid region" (Ch. 2) addresses the representation of physical processes in Nextgen framework compatible streamflow models. An ordinary differential equation representation of soil fluxes was developed within the CFE model and compared to the existing framework using observed streamflow from 498 CAMELS basins. In addition, model parameters were determined with differential programming approaches for CFE and the Layered Green & Ampt with Redistribution model.

2. Projects within the Urban Flooding Under Climate Change theme:

"Analysis of Flood Drivers Contributions to Compound Flooding Using Coupled Modeling and Machine Learning" (Ch. 3) explores the factors contributing to compound flooding in New York City (NYC) through advanced hydrological and hydrodynamic modeling techniques and machine learning methods. River discharge data from the NWC's CFE model was integrated into the hydrodynamic model GeoClaw. Then, to assess the relative contributions of precipitation, storm surge, and river discharge in compound flooding, four machine learning models were trained and validated, including Random Forest, Support Vector Machine, LSTM, and Multi-Layer Perceptron. This approach provided insight into compound flooding drivers and contributed to the understanding of how these factors interact in NYC.

3. Projects within the Real-Time Urban Flooding Awareness theme:

"Predicting Flood Inundation Susceptibility Using HAND FIM, Crowd-Sourced and Satellite Data with Machine Learning" (Ch. 4) introduces a hybrid modeling approach integrating crowd-sourced data, hydrological information, and NWC-generated FIM in order to address the challenges of FIM in urban environments. This group enhanced modeled FIM with citizen-provided flood depth information. A random forest algorithm was used to classify flood inundation, incorporating geospatial features such as topography, flow accumulation and direction, rainfall, crowd-sourced data, damages, and FEMA Flood Insurance Program maps.

4. Projects within the Channel Flow Routing and Flood Inundation Modeling theme:

"Quantifying the Sources of Uncertainty in OWP HAND-FIM Predictions" (Ch. 5) evaluates OWP HAND-FIM performance and isolates uncertainties in Synthetic Rating Curves (SRC) and NWM-predicted discharge on flood extent. The Amite River Basin (ARB) was chosen as the study area due to its susceptibility to flooding and the availability of HEC-RAS modeled data for historical events. SRCs were compared at USGS gauge locations with HEC-RAS rating curves. They found that while OWP HAND-FIM had a 44% agreement with observed data, the ARB HEC-RAS model had a much higher agreement of 93%. Disparities between SRCs and USGS rating curves were studied in order to explain uncertainties in OWP HAND-FIM.

"Improving the Fidelity and Performance of OWP HAND-FIM Using a Surrogate Model Technique (SMT)" (Ch. 6) aims to improve flood prediction accuracy and overcome limitations of the current HAND-FIM method. OWP HAND-FIM's fidelity and performance was enhanced by implementing a machine learning surrogate model technique (SMT). The SMT was designed to emulate the hydrodynamic properties of a high-fidelity HEC-RAS model, allowing seamless integration with the lower-fidelity HAND-FIM approach for more reliable predictions.

Acknowledgements

The successful execution of this program would not have been possible without the tireless efforts and dedication of numerous individuals. We extend our sincere gratitude to all those who contributed to the NWC-SI.

First and foremost, we would like to thank those from CUAHSI who played pivotal roles in making this program a reality: Deanna McCay, Julia Masterman, Jordan Read, Irene Garousi-Nejad, and Tony Castronova. Lauren Stewart, Ed Clark, Fred Ogden, Matt Williamson, Christy Westcott, and Trey Flowers of the NWC provided technical guidance and administrative support. Special thanks to CIROH, specifically Sagy Cohen and Lanna Nations, who managed the logistics of living on the University of Alabama campus, including travel arrangements, accommodations, and institutional relationships. Technical training provided by Irene Garousi-Nejad and Tony Castronova (CUAHSI), Dylan Lee, Brad Bates, and Mike Johnson (NWC), Jonathan Frame (Floodbase), Barbara Minsker (SMU), and Andy Carter (UT Austin) introduced fellows to critical computational tools and data products. Course Coordinators Ebrahim Hamidi and Mark Wang helped plan training sessions and research deliverables, and lived and worked with the fellows for the duration of the NWC-SI's seven weeks. To all those whose names may not be mentioned here but have played a part in making this program a resounding success, we offer our heartfelt thanks. Your contributions are deeply valued and have left a lasting impact on the NWC-SI.

Since the first NWC-SI in 2015, hundreds of graduate student fellows have collaborated to bring research projects to fruition. The friendships and professional connections they formed in Tuscaloosa are of equal importance to the technical progress made. This is an invaluable experience and we express our appreciation to the NWC and NOAA OWP for hosting this unique activity, and for the opportunity to contribute to the enhancement of water prediction for our nation.

Mark Wang
Student Course Coordinator, NWC Innovators Program - Summer Institute 2023
Ph.D. Candidate, University of Texas

Ebrahim Hamidi
Student Course Coordinator, NWC Innovators Program - Summer Institute 2023
Ph.D. Candidate, University of Alabama

Chapter 1

Data Assimilation for Improving Forecast Accuracy and Streamflow Prediction in Ungauged Basins

Ehsan Foroumandi[1], Jeil Oh[2], Parnian Ghaneei[3], and Sujana Timilsina[4]

[1]The University of Alabama; *eforoumandi@crimson.ua.edu*

[2]The University of Texas at Austin; *jeoh@utexas.edu*

[3]The University of Alabama; *pghaneei@crimson.ua.edu*

[4]The University of Texas at Austin; *sujanatimilsina@utexas.edu*

Academic Advisors: Hamid Moradkhani, *The University of Alabama*; Matt Bartos, *University of Texas at Austin*; Mukesh Kumar , *The University of Alabama*; Paola Passalacqua, *University of Texas at Austin*

Summer Institute Theme Advisors: Jonathan M. Frame, Floodbase

Abstract: Improving the predictive accuracy of hydrological models remains a daunting challenge due to the numerous sources of uncertainties. Data Assimilation (DA) techniques offer a potential solution to mitigate these uncertainties and improve streamflow predictions. The primary objective of this research is to integrate the data assimilation methods within the Next Generation Water Resources Modeling Framework (NextGen), thereby enhancing streamflow prediction accuracy. This study focuses on enhancing the predictive performance of the Conceptual Functional Equivalent (CFE) model by employing various DA techniques in the Carmel watershed, California. First, an Ensemble Kalman Filter (EnKF) method was developed to assimilate streamflow data into the CFE model and estimate the soil moisture storage state. It is then compatible with the NextGen using a Basic Model Interface (BMI)-enabled Python module. Furthermore, a state-space Muskingum-Cunge channel routing scheme, integrated with a Kalman filter, was designed to route the outflows from the CFE model within the watershed to predict the streamflow at ungauged basins. The findings of this study indicate that the streamflow forecasts of the CFE model and the predictions in ungauged basins can be enhanced by assimilating streamflow observations through multiple data assimilation methods.

1. Motivation

A hydrologic model simulates fluctuations in water transport and storage over time across one or more components of the natural hydrologic cycle, thus providing a means to project future situations. These predictions depend on both the initial conditions of the model and the external forces acting upon it during the prediction period. However, identifying these state estimates and parameters has long been a complex challenge in the field of hydrology [1]. Data assimilation (DA) has been widely used to address this challenge [2]. DA identifies uncertainties in the hydrological model and the observation data. It then estimates the hydrological model states to

bring the model's simulations in line with observations, which can improve the future prediction by reflecting the real-world data to the baseline of the model. The application of data assimilation can enhance state estimates, consequently improving the accuracy of streamflow forecasts. In distributed hydrological models, DA can enhance state estimates at specific locations within a gauged river basin and also leverage state estimates in ungauged basins [3]. Despite the promise of DA, the integration of data assimilation methodologies into the National Water Model (NWM) remains relatively limited. Currently, the NWM uses a nudging method that directly incorporates USGS streamflow observations [4]. While this can enhance the accuracy of the simulation results, it fails to account for the inherent uncertainties present in both the model and the observations.

The Office of Water Prediction (OWP) is currently developing the Next Generation Water Resources Modeling Framework (NextGen). Being model-agnostic, NextGen can utilize specific model formulas that are regionally custom-designed to account for streamflow generation processes [5]. It has the capability to operate either a single hydrological model or multiple ones, regardless of their programming language, by using the open-access methods defined by the Basic Model Interface (BMI) standard. In light of this, there is a need to develop a DA module that can be integrated into the NextGen framework to enhance the accuracy of streamflow predictions.

2. Objectives and Scope

The central goal of this research is the advancement of data assimilation methods within the NextGen framework to enhance streamflow predictions. The focus falls particularly on the Conceptual Functional Equivalent (CFE) model, which is one among a variety of models (CFE, LSTM, NOAH-MP) incorporated into the NextGen framework. Functioning at the catchment-scale, the CFE model replicates the WRF-Hydro configuration utilized in the National Water Model [6]. The paramount objective of this research is twofold: (1) to improve the predictive accuracy of the CFE model and (2) to enhance streamflow predictions in ungauged basins using channel routing. The first part involves the creation of a DA module that aligns with the BMI for direct integration into the NextGen framework. This module is designed to assimilate sensor gauge data into the CFE model and estimate the state of soil moisture at the catchment level, employing the Ensemble Kalman Filter (EnKF). The second part of our research goal demands the implementation of a Muskingum-Cunge routing scheme with DA, which can be potentially incorporated into the T-route in the NextGen. Combined with DA via the application of the Kalman Filter, this scheme is devised to provide improved streamflow predictions in ungauged locations.

3. Previous Studies

Data Assimilation has been applied to various hydrological models using different methodologies. These methodologies include the variational method [7,8], the Extended Kalman filter [9], the Particle Filter [10, 11], and the EnKF [1,3]. While the National Water Model does not actively utilize data assimilation techniques beyond nudging in real-world operations, there

have been studies aiming to enhance the National Water Model using data assimilation [12]. For instance, researchers have applied EnKF to integrate stream gauge data and estimate soil moisture [13].

Regarding routing, the NWM adopts the Muskingum-Cunge channel routing method, a widely accepted approach for channel routing [14]. Studies indicate that the performance of streamflow prediction models improves when data assimilation is used in conjunction with the routing method [15]. Earlier studies also aimed to predict downstream flow by assimilating upstream flow information. This was done by employing a routing procedure in conjunction with a Kalman filter. Specifically, a state space equation was developed based on the Muskingum-Cunge method [16]. This approach was subsequently used to tackle a watershed-scale issue, where the influence of sensor location was closely examined [17].

4. Methodology

4.1. Study Area

We focus our study on the Carmel River watershed in coastal Monterey County, California, which spans approximately 660 square kilometers. Influenced by a Mediterranean climate, the watershed experiences significant rainfall from November to March. This precipitation primarily originates from storms in the Gulf of Alaska and atmospheric rivers in the Central Pacific [18]. The watershed consists of 90 subcatchments derived from the Hydro Fabric system. Two gauges, operated by the United States Geological Survey (USGS), are installed along the river. USGS gauge 11143200 is located at the upstream location (at Esquiline Road and) whereas gauge 11143250 is located at the downstream location (Via Mallorca Road) respectively [19].

4.2. Conceptual Functional Equivalent (CFE)

CFE is a conceptual hydrological model, operating within the Nextgen Framework that simplifies the runoff schemes and offers functionalities comparable to the current version of the NWM that is based on WRF-Hydro [6,20]. The rainfall and potential evapotranspiration (PET) are the forcing data for the CFE model. In this study, we employed the Analysis of Record for Calibration (AORC) forcing dataset, which provides a gridded representation of near-surface weather conditions. This dataset is defined on a spatial grid with a mesh length of approximately 800 meters (30 arc seconds) and a temporal resolution of one hour. We obtained the AORC dataset for 90 subcatchments within the Carmel River watershed, employing the NextGen hydrofabric tools. The NextGen hydrofabric represents and discretizes the hydrologic landscape and drainage network through a comprehensive three-part data product including features pertaining to catchments and flowpaths, their interconnectivity, and attribute sets essential for model execution. We also developed a Python BMI module to compute PET using the Penman-Monteith equation and the AORC dataset [21]. With calculated PET, precipitation from the AORC forcing dataset and attributes from hydrofabric, we run the CFE model which provides the output in the form of total runoff depth and total runoff volume flux.

4.3. Hydrologic Routing

NWM employs the Muskingum-Cunge (MC) method for channel routing [15]. The MC method in the NWM is implemented as given by the equation (1).

$$Q_{j+1}^{t+1} = C_1 Q_j^{t+1} + C_2 Q_j^t + C_3 Q_{j+1}^t + \frac{q_l dt}{D} \quad (1)$$

where Q represents the estimated flow of a segment at both the upstream (j) and downstream (j+1) locations, calculated at the previous time step (t) and the current time step (t+1). The parameters, C_1, C_2, C_3, and D, can be found in [15]. The MC method models diffusion using parameters that are derived from the geometry of the channel. To align the approach with the current NWM, the channel geometry parameters for this study are based on the channel geometry equations given by [15] and the hydro fabric outputs.

4.4. Data Assimilation

4.4.1. Kalman Filter (KF)

Kalman Filter is a Bayesian recursive estimator that initially predicts a prior estimate of system states based on a given dynamical system. This prior estimate is then updated using observational data to produce a posterior estimate [22]. It is also considered an optimal estimator as it combines two sources of information—the predicted states and the noisy measurements—in such a way that it minimizes the variance in the estimated states. For this study, KF is used for the data assimilation in the channel routing with the Muskingum-Cunge method. The network topology of reaches allows us to express the routing equation (1) as a set of difference equations which can be arranged as a state-space formulation for real-time data assimilation for channel routing. As the routing process advances forward in time with the streamflow output from the CFE for each catchment, KF is subsequently applied at each time step to integrate the USGS sensor data into the router. This generates an updated, optimal posterior estimate of discharge across the entire watershed.

4.4.2. Ensemble Kalman Filter (EnKF)

Sequential data assimilation involves estimating the state of a model at every point of observation, which only relies on the previously observed data. The previously-mentioned KF method provides an optimal solution for this task in linear situations as it is a recursive data-processing algorithm. However, when dealing with nonlinear dynamics, it is not possible to utilize the KF method; therefore, an alternative approach is taken by employing one of the extensions of KF, the so-called ensemble KF (EnKF) [1]. The EnKF relies on generating ensembles through Monte Carlo methods, where the forecasted state error covariance matrix is estimated by propagating a group of model states using the updated states (in this study, soil

moisture) from the previous time step. In the EnKF method, the model prediction is made for each ensemble member as:

$$x_{t+1}^{i-} = f(x_t^{i+}, U_t^i, \Theta, t) + \omega_t^i, \quad i = 1,2,..,n \tag{2}$$

where $x_{t+1}^{i-}$is the ith ensemble member forecast at time t+1 and x_t^{i+} is the ith updated ensemble member at time t.

Given a priori estimated states (forecasted states x_{t+1}^{i-}) made in advance, the EnKF utilizes the observation y_{t+1} to derive the updated states (posterior estimate x_{t+1}^{i+}). To achieve this, a linear correction is used based on the KF method using the Kalman gain (K_{t+1}), which is approximated from ensembles, to update the state ensemble members as:

$$x_{t+1}^{i+} = x_{t+1}^{i-} + K_{t+1}(y_{t+1}^i - \hat{y}_{t+1}^i) \tag{3}$$

where y_{t+1}^i is the ith trajectory of the observation replicates generated by adding the noise of η_{t+1}^i with covariance Σ_{t+1}^y to the actual observation:

$$y_{t+1}^i = y_{t+1} + \eta_{t+1}^i, \eta_{t+1}^i \sim N(0, \Sigma_{t+1}^y) \tag{4}$$

The EnKF method as shown in Equation (4), treats the observation value as a random variable by generating an ensemble of data.

To perform and evaluate the EnK method, first, all of the involved variables in the modeling procedure, including forcing data, state variables, simulations, and observations are perturbed using standardized errors. Then, two loops are performed to evaluate the performance of CFE-EnKF: open loop and synthetic loop. To illustrate the ability of the developed CFE-EnKF method to estimate pre-defined states and model outputs, a synthetic scenario is designed. In this scenario, model states and outputs are generated with predefined parameters and forcing data, which are assumed to be true and can be used as observed values [1]. Then, the EnKF is performed to correct the initial state value and tackle the uncertainties of the modeling. The results of the CFE-EnKF are compared with the results of an open loop with the same error factors and parameters.

5. Results

5.1. DA-CFE

In this study, our state variable for performing DA is soil moisture storage and we aim to assimilate streamflow observation data into the model. DA-CFE is implemented to improve the performance of CFE modeling through answering the uncertainties associated with the forcing data and model parameters. The outputs of CFE-EnKF are the averaged values of the model with corrected initial state variables and the outputs of the open loop are the averaged values of the model after perturbing the variables without correcting the initial state variable. This

comparison is performed to show that the improved performance of the model is not due to averaging the output ensembles. The results (**Figure 1**) indicates that EnKF could successfully improve the performance of the model compared to open loop. According to Figure 1a, the uncertainty window for EnKF is more reliable than the open loop. Additionally, Figure 1b indicates that EnKF could correct the state variable at each time step compared to the open loop. Figure 1c shows that KGE, NSE and RMSE are improved about 4 %, 10%, and 50%, respectively, representing the efficiency of applied EnKF on CFE.

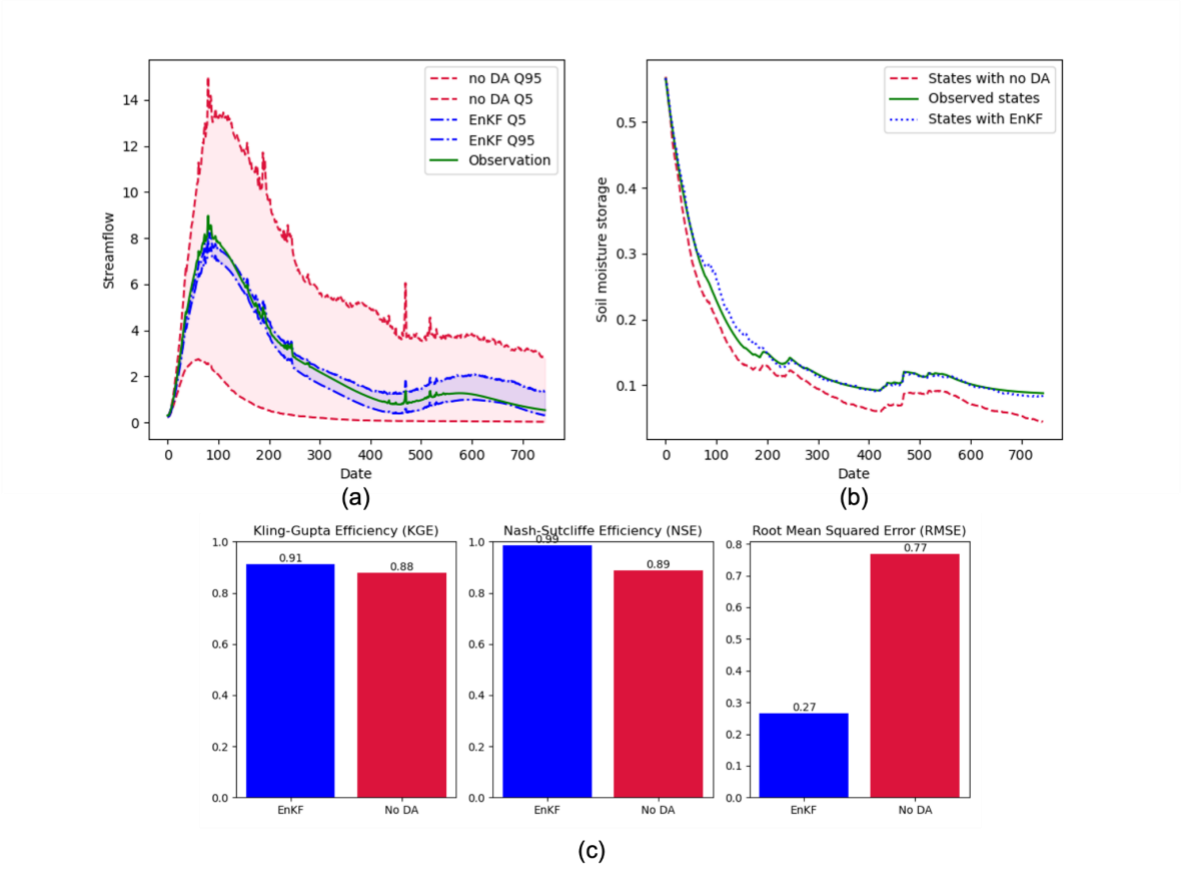


Figure 1. *Comparison of results of the synthetic loop with EnKF and no DA and observation in a) streamflow, b) soil moisture storage (state variable), and c) evaluation metrics.*

5.2. DA-Routing

The performance of data assimilation to improve streamflow predictions at ungauged locations is carried out through a holdout evaluation, utilizing the two USGS gauges located in the Carmel river watershed. In this assessment, the Kalman Filter is used to integrate data only at the upstream gauge. The resultant streamflow estimates at the downstream location are then compared with the downstream USGS sensor observations, which are the ground truth.

Figure 2 illustrates how USGS gauge observations, assimilated at upstream locations, influence the router over time. The corrections implemented by the Kalman Filter effectively propagate downstream, resulting in improved streamflow estimations at the downstream holdout site. Unlike the simulation and NWM models, which lack data assimilation and often drastically overestimate streamflow or misalign with peak timing, the router equipped with data assimilation consistently aligns well with observed data at the downstream holdout site, even when the performance of the CFE output is suboptimal (**Figure 3**).

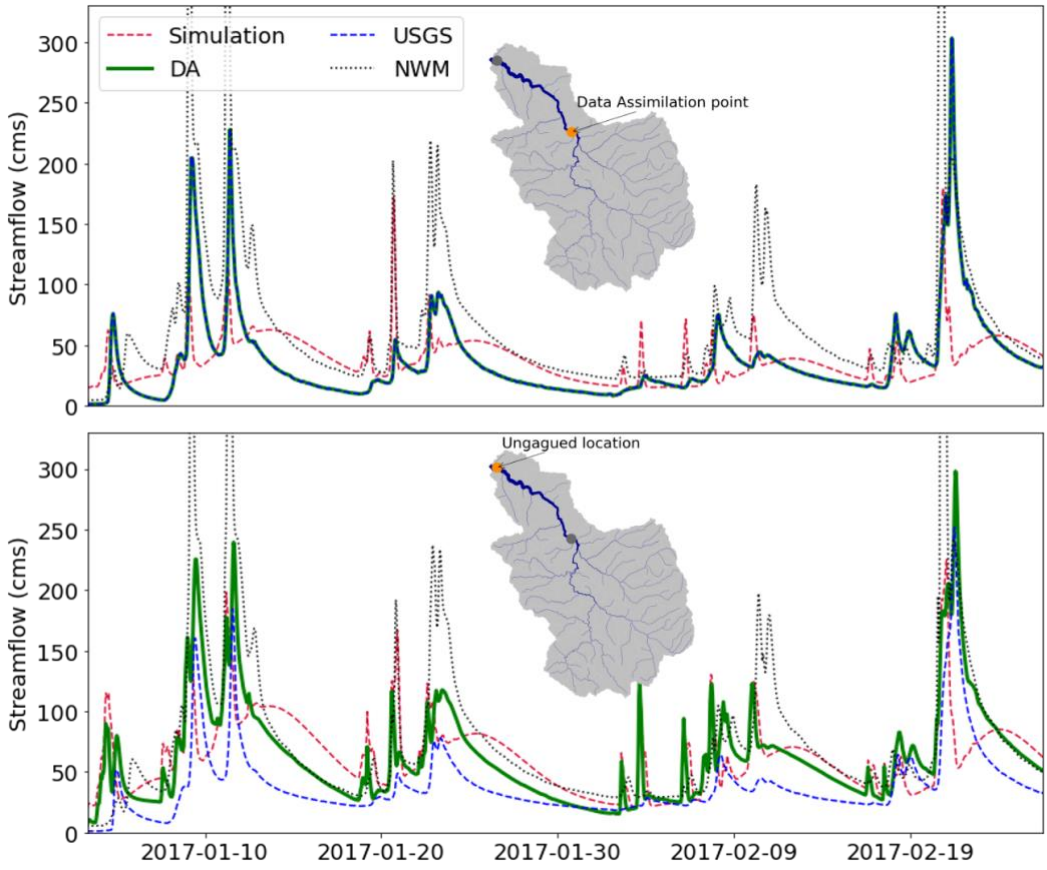


Figure 2. *Results of the Kalman Filtering holdout assessment: Simulation Output (Red), Data Assimilation Output (Green), USGS Observations (Blue), and NWM Output (Grey); Minimap show hydrograph locations in the watershed.*

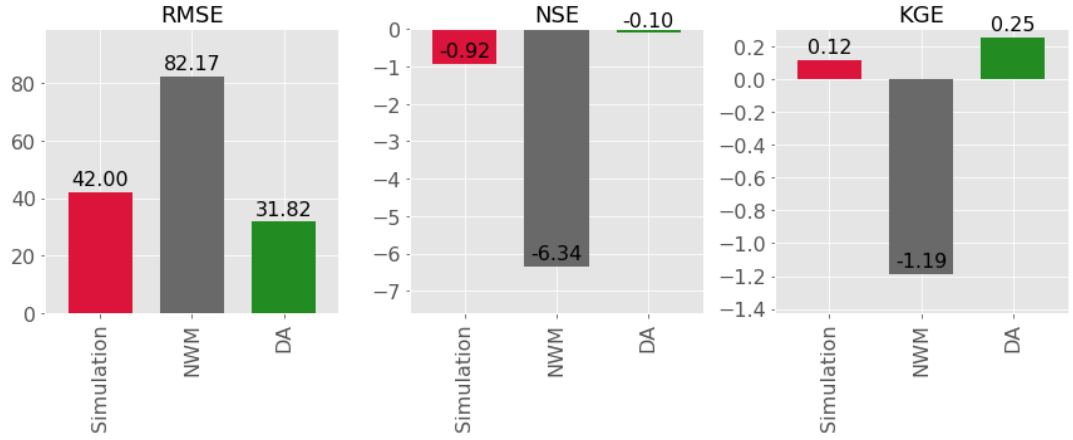


Figure 3. *The error evaluations of the models, using metrics such as RMSE, NSE, and KGE, are conducted based on the downstream gauge data; the simulation output is represented by red bars, NWM is shown in gray, and the router model with DA is depicted in green.*

6. Conclusion

We developed the DA-enabled schemes for both CFE and the Muskingum-Cunge routing. Our findings indicate that data assimilation can improve the predictive performance at both the catchment scale (DA-CFE) and the watershed scale (DA-routing). The CFE model combined with the EnKF was successful in improving prediction accuracy based on the estimation of soil moisture storage state by assimilating streamflow gauge data. Furthermore, the Muskingum-Cunge routing method, in combination with the KF, was effective in improving downstream flow forecasting by integrating the CFE output from each catchment with the assimilated upstream USGS sensor gauge. The proposed DA-CFE can be applied directly to NextGen for employing DA on CFE and other hydrological models. DA-routing also provides a DA-applied routing scheme as a potential option for T-route in NextGen. It is, however, worth noting that this option would require further development to incorporate the T-route schemes as a part of the state-space function.

Supplementary Materials: Code and data links are available at: https://github.com/NWC-CUAHSI-Summer-Institute/data_assimilation_with_bmi

References

1. Moradkhani, H.; Sorooshian, S.; Gupta, H. V.; Houser, P. R. Dual State–Parameter Estimation of Hydrological Models Using Ensemble Kalman Filter. *Advances in Water Resources* **2005**, *28* (2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002.
2. Liu, Y.; Gupta, H. V. Uncertainty in Hydrologic Modeling: Toward an Integrated Data Assimilation Framework: HYDROLOGIC DATA ASSIMILATION. *Water Resour. Res.* **2007**, *43* (7). https://doi.org/10.1029/2006WR005756.
3. Clark, M. P.; Rupp, D. E.; Woods, R. A.; Zheng, X.; Ibbitt, R. P.; Slater, A. G.; Schmidt, J.; Uddstrom, M. J. Hydrological Data Assimilation with the Ensemble Kalman Filter: Use of Streamflow Observations to Update States in a Distributed Hydrological Model. *Advances in Water Resources* **2008**, *31* (10), 1309–1324. https://doi.org/10.1016/j.advwatres.2008.06.005.
4. Seo, B.; Krajewski, W. F.; Quintero, F. Multi-Scale Hydrologic Evaluation of the National Water Model Streamflow Data Assimilation. *J American Water Resour Assoc* **2021**, *57* (6), 875–884. https://doi.org/10.1111/1752-1688.12955.
5. Ogden, F.; Avant, B.; Bartel, R.; Blodgett, D.; Clark, E.; Coon, E.; Cosgrove, B.; Cui, S.; Kindl da Cunha, L.; Farthing, M.; others. The Next Generation Water Resources Modeling Framework: Open Source, Standards Based, Community Accessible, Model Interoperability for Large Scale Water Prediction. In *AGU Fall Meeting Abstracts*; 2021; Vol. **2021**, pp H43D-01.
6. Cunha, L.; Jennings, K.; Wood, A.; Mizukami, N.; Ogden, F.; Feng, X.; Liu, Y.; Peckham, S.; Garrett, J.; Frame, J.; others. Next Generation National Water Model: Strategy and Preliminary Performance of Initial Model Formulations. In *AGU Fall Meeting Abstracts*; **2021**; Vol. 2021, pp H54G-04.

7. Alvarado-Montero, R.; Schwanenberg, D.; Krahe, P.; Helmke, P.; Klein, B. Multi-Parametric Variational Data Assimilation for Hydrological Forecasting. *Advances in Water Resources* **2017**, *110*, 182–192. https://doi.org/10.1016/j.advwatres.2017.09.026.
8. Lee, H.; Seo, D.-J.; Koren, V. Assimilation of Streamflow and in Situ Soil Moisture Data into Operational Distributed Hydrologic Models: Effects of Uncertainties in the Data and Initial Model Soil Moisture States. *Advances in Water Resources* **2011**, *34* (12), 1597–1615. https://doi.org/10.1016/j.advwatres.2011.08.012.
9. Bartos, M.; Kerkez, B. Pipedream: An Interactive Digital Twin Model for Natural and Urban Drainage Systems. *Environmental Modelling & Software* **2021**, *144*, 105120. https://doi.org/10.1016/j.envsoft.2021.105120.
10. Moradkhani, H.; Hsu, K.-L.; Gupta, H.; Sorooshian, S. Uncertainty Assessment of Hydrologic Model States and Parameters: Sequential Data Assimilation Using the Particle Filter: UNCERTAINTY ASSESSMENT OF HYDROLOGIC MODEL. *Water Resour. Res.* **2005**, *41* (5). https://doi.org/10.1029/2004WR003604.
11. Evensen, G. Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. *J. Geophys. Res.* **1994**, *99* (C5), 10143. https://doi.org/10.1029/94JC00572.
12. Huling, L. G.; others. Streamflow and Precipitation Data Assimilation into the National Water Model: An Investigative Study for Statewide Flood Forecasting. PhD Thesis, 2020.
13. Zarekarizi, M. Ensemble Data Assimilation for Flood Forecasting in Operational Settings: From Noah-MP to WRF-Hydro and the National Water Model. PhD Thesis, Portland State University, 2018.
14. El Gharamti, M.; McCreight, J. L.; Noh, S. J.; Hoar, T. J.; RafieeiNasab, A.; Johnson, B. K. Ensemble Streamflow Data Assimilation Using WRF-Hydro and DART: Novel Localization and Inflation Techniques Applied to Hurricane Florence Flooding. *Hydrology and Earth System Sciences* **2021**, *25* (9), 5315–5336.
15. Read, L. K.; Yates, D. N.; McCreight, J. M.; Rafieeinasab, A.; Sampson, K.; Gochis, D. J. Development and Evaluation of the Channel Routing Model and Parameters within the National Water Model. *J American Water Resour Assoc* **2023**, 1752-1688.13134. https://doi.org/10.1111/1752-1688.13134.
16. Georgakakos, A. P.; Georgakakos, K. P.; Baltas, E. A. A State-Space Model for Hydrologic River Routing. *Water Resour. Res.* **1990**, *26* (5), 827–838. https://doi.org/10.1029/WR026i005p00827.
17. Mazzoleni, M.; Chacon-Hurtado, J.; Noh, S. J.; Seo, D.-J.; Alfonso, L.; Solomatine, D. Data Assimilation in Hydrologic Routing: Impact of Model Error and Sensor Placement on Flood Forecasting. *J. Hydrol. Eng.* **2018**, *23* (6), 04018018. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001656.
18. Mathias Kondolf, G.; Batalla, R. J. Chapter 11 Hydrological Effects of Dams and Water Diversions on Rivers of Mediterranean-Climate Regions: Examples from California. In *Developments in Earth Surface Processes*; Elsevier, 2005; Vol. 7, pp 197–211. https://doi.org/10.1016/S0928-2025(05)80017-3.

19. Frame, J. An Integrated Surface Water-Groundwater Interaction Model for the Carmel River.
20. Modaresi Rad, A.; Sthapit, E.; Awaad, M.; Sigdel, R.; Ogden, F. L.; McMillan, H. K. Evaluation of Streamflow Generation Using Different Formulations of Conceptual Functional Equivalent Model for Semi Arid Region. **2022**, *2022*, H45I-1486.
21. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*; Allen, R. G., Food and Agriculture Organization of the United Nations, Eds.; FAO irrigation and drainage paper; Food and Agriculture Organization of the United Nations: Rome, 1998.
22. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. **1960**.

Chapter 2

On numerical methods and differentiable modeling for soil process representations in the NextGen Framework for arid regions

Ryoko Araki[1], Soelem Aafnan Bhuiyan[2], Tadd Bindas[3], and Jeremy Rapp[4]

*All authors contributed equally, and order is set alphabetically

[1]San Diego State University & University of California, Santa Barbara, California; raraki8159@sdsu.edu

[2]George Mason University, Fairfax, Virginia; sbhuiya2@gmu.edu

[3]Penn State University, University Park, Pennsylvania; tkb5476@psu.edu

[4]Michigan State University, East Lansing, Michigan; rappjer1@msu.edu

Academic Advisors: Hilary McMillan, *San Diego State University*; Viviana Maggioni, *George Mason University*; Chaopeng Shen, *Penn State University*; Anthony Kendall, *Michigan State University*

Summer Institute Theme Advisors: Jonathan M. Frame, *Floodbase*

Abstract: We hypothesize soil/physical processes in these regions are not adequately represented within Nextgen compatible streamflow prediction models. Thus, there is a need for more accurately considering soil/physical processes represented in the system from within the conceptual-type model that is relatively computationally efficient. We developed an Ordinary Differential Equation (ODE) representation of soil fluxes within the Conceptual Functional Equivalent Framework (CFE) model. We compared it to the existing framework using observed streamflow from 498 (Catchment Attributes and Meteorology for Large-sample Studies) CAMELS basins. Results were inconclusive as CFE with the ODE soil flux integration performed similarly to the CFE. To further investigate the representation of soil/physical properties in arid/semiarid regions, we created differential programming approaches for CFE and the Layered Green & Ampt with Redistribution (LGAR) model to determine physically-informed model parameters. All project deliverables, including supplemental information, are available at http://bit.ly/Summer-Institute-Ngen-Aridity.

---

1. Motivation

In recent years, a significant focus of hydrologic research has been on evaluating the performance of models for streamflow forecasting within the contiguous United States (CONUS) over a benchmark dataset (CAMELS; [1], [2]). These efforts have compared the performance of physical-conceptual-based models against emergent deep learning models across various catchment scales and hydrologic settings [3], [4]. Within these analyses, a ubiquitous performance drop in catchments under arid/semi-arid climates, where soil processes significantly control hydrologic connectivity [5], [6], has been observed for all model schema [3], [7]. Along with the core physical processes in the arid regions, key characteristics such as phase correlation, forest, and grass cover restrict model capacity to estimate regional

evapotranspiration [8] in arid regions accurately. Hence, it is evident that models need a better representation of land-surface or soil module, which Keith Beven posited in his 2023 paper on "what models and parameter sets might be considered as not fit-for-purpose" [9].

We test the null hypothesis "Models with detailed soil process representation perform no better than the ones with less detailed soil in the arid regions" in CAMELS basins. Our multimodel approach consists of a combination of process-based and differentiable models considered "fit-for-the purpose" [9] in arid regions, i.e., models with finer soil process representation empowered with a parameter-learning system for predicting hydrograph patterns.

2. Objective and Scope

We developed process-based and differentiable models with detailed representations of soil processes and evaluated their performance in arid environments. Our research questions are:

1. Can we improve CFE model predictions in arid regions by updating the soil moisture partitioning using an ordinary differential equation? (See the results in Section 5.1).
2. Can differentiable modeling show an improvement over process-based models in arid regions or assist in learning intermediate soil/physical processes within CFE and LGAR? (Section 5.2).

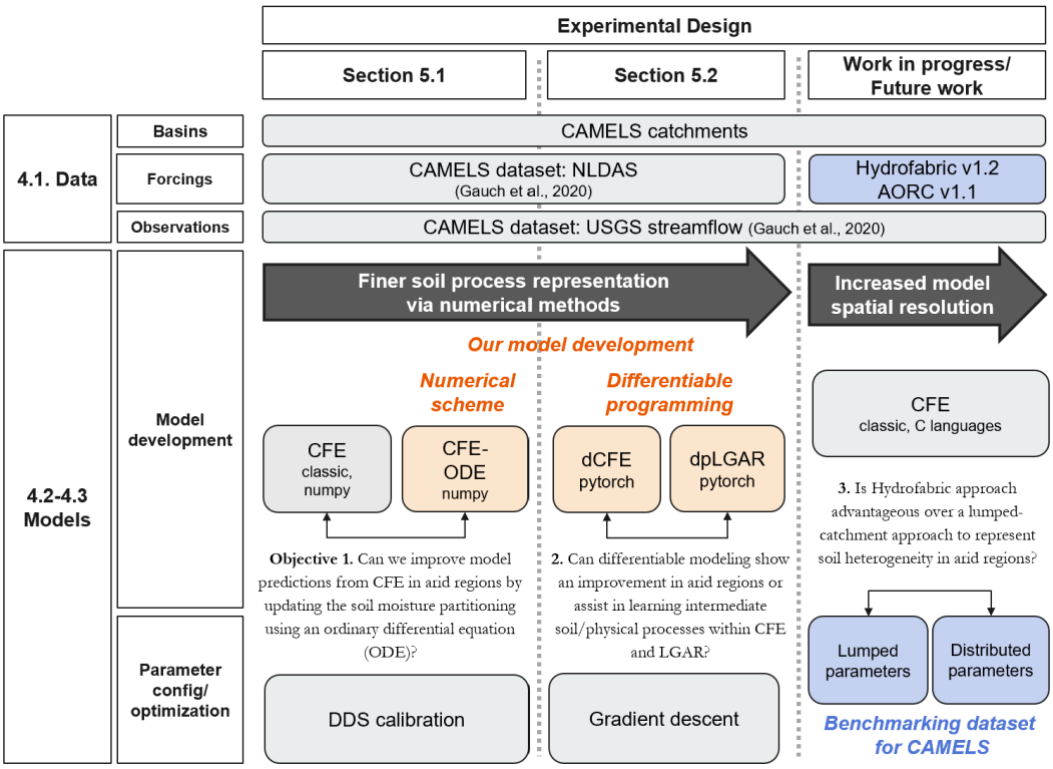


Figure 1: Overview of our experimental design, with the corresponding Method section on the left and Result section at the top. The middle shows model comparisons and research objectives/questions.

3. Previous Studies

Our work uses the NOAA-OWP NWC NextGen framework and builds upon the project by the 2022 Summer Institute team, "Automated decision support for model selection in the

Nextgen National Water Model" [10]. While their study focused on model selection over CONUS, we investigated multi-model approaches in arid regions, specifically using models with high representations of soil properties.

4. Methodology

4.1. The CAMELS dataset: Forcing and observation data:

Following the previous benchmarking studies [11]–[13], we used a dataset developed by Gauch et al. (2020, 2021), which includes 561 CAMELS basins. The dataset contains hourly streamflow data from the United States Geological Survey and forcing data (hourly total rainfall and potential evapotranspiration) from the National Land Data Assimilation System for the CAMELS basins. We further reduced the number of basins to 498 during model calibration.

4.2. Models

4.2.1. CFE model (https://github.com/NWC-CUAHSI-Summer-Institute/cfe_py)

Conceptual Functional Equivalent (CFE) is a conceptual hydrologic model that provides similar functionality to the current National Water Model with simplified soil moisture and routing expressions [15].

4.2.2. CFE-ODE model (https://github.com/NWC-CUAHSI-Summer-Institute/cfe_py)

The current version of CFE models sequentially subtracts three fluxes going out from soil moisture storage (soil evaporation, percolation, then lateral flow), reducing computational time but introducing approximation errors and inaccurately represents the internal feedback [16]. We introduced an ordinary differential equation (ODE) to the soil moisture module to solve those issues. The physics equations in the ODE scheme are equivalent to the existing CFE formulation, but the outfluxes are simultaneously calculated and subtracted from the soil moisture storage. Hereafter, we refer to the model with the previous soil moisture calculation scheme (as introduced in Section 4.2.1.) as "CFE-classic" and with the ODE scheme as "CFE-ODE."

4.2.3. CFE calibration (https://github.com/NWC-CUAHSI-Summer-Institute/calibrate-cfe)

We calibrated both CFE-classic and ODE for the nine parameters and two runoff subroutines (See Table S1) using the Dynamically Dimensioned Search optimization algorithm (DDS) [17]. Five hundred evaluation runs were used as a hyperparameter [18] with Kling-Gupta Efficiency as an objective function [19]. The calibration and testing periods were the same as the LSTM training period of WY 2000-2008 and the testing period of WY 2010, respectively, with a one-year warmup period.

4.2.4. LGAR model (https://github.com/NOAA-OWP/LGAR-C)

The Layered Green & Ampt with Redistribution (LGAR) [20] model partitions precipitation into infiltration and runoff and was designed to mimic the Richardson-Richards (RRE) equation [21], [22] for semi-arid and arid basins. LGAR can be conceptually visualized as a bucket filled with different layers of soil (Figure 2(a)). Each layer possesses different Van Genuchten [23] parameters and hydrologic conductivity values (K_{sat}), which control water storage. Through tracking water storage, soil infiltration, evapotranspiration, runoff, and percolation can be inferred. Van Genuchten parameters, and hydrologic conductivity values, are statically defined for each soil column via a soil survey to a lookup table. Figure 2(b) shows a simple example of volumetric water content and pressure head, varying with depth.

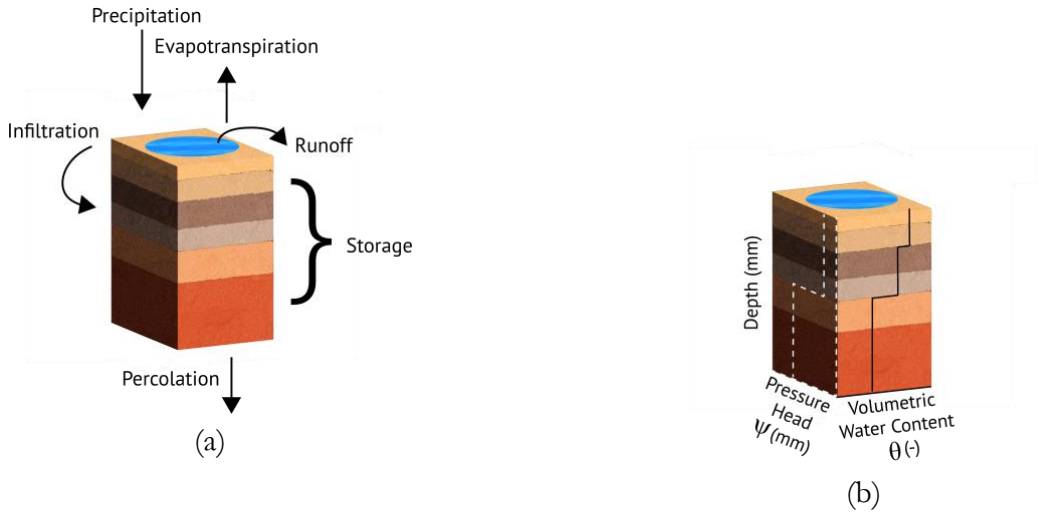


Figure 2: (a) A conceptual representation of the hydrologic properties within a soil column. (b) Two graphs show how pressure head and volumetric water content change through soil layers

4.3. Differentiable modeling

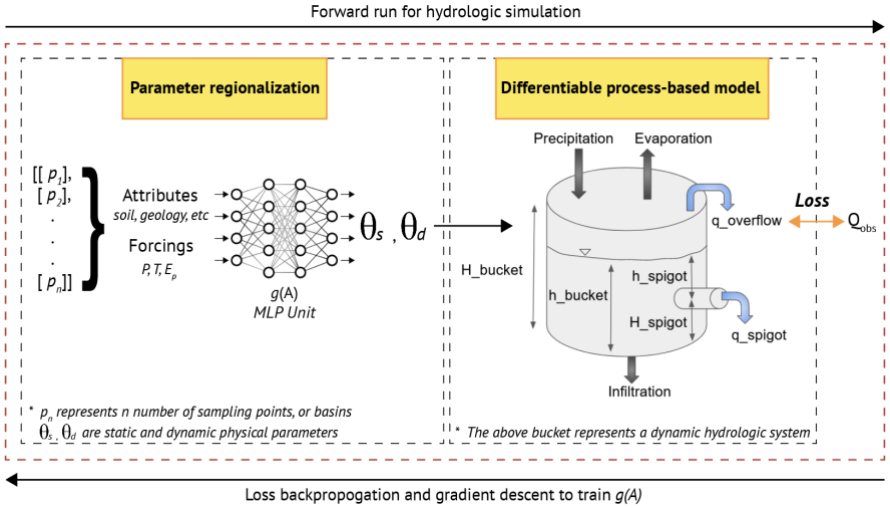


Figure 3: A conceptual differentiable parameter learning framework based off Feng et al. (2023) [24]. Parameters are calculated through the NN, then passed to the differentiable process-based model to generate discharge. Discharge is compared to observations, loss is calculated, and NN weights are updated.

4.3.1. dpLGAR (https://github.com/NWC-CUAHSI-Summer-Institute/dpLGAR)

By implementing LGAR on a differentiable platform (PyTorch), we can train a NN to produce physical soil parameters representative of each soil layer, similar to work done in previous differentiable parameter learning (dPL) studies [24]–[29]. We input a p number of SSURGO (POLARIS) Polaris [30] two-meter soil column attributes into a Multilayer Perceptron Model (MLP) [31] (Figure 3; see Figure S1 for the LGAR-specific model) to learn lumped catchment scale soil parameters and ponding depth limitations (Equation 1).

$$\alpha, n, K_{sat}, \theta_s, \theta_r, d_p = NN(c) \quad (1)$$

Attributes used for parameter prediction include soil composition percentages (clay, silt, sand), soil pH, and organic matter content (Table S2).

4.3.2. dCFE (https://github.com/NWC-CUAHSI-Summer-Institute/dCFE)

Similar to dpLGAR, and other differentiable methods [24]–[29], we implemented the CFE model in a differentiable platform (PyTorch) to learn CFE parameters (Figure 3; see Figure S2 for the CFE-specific model). The two parameters learned are refkdt and satdk, both significant determinants of the influx to soil water storage [32]. We input four static attributes from the Hydrofabric for one CAMELS basin as a test case into a Multilayer Perceptron Model (MLP) [31] (Table S3).

$$refdkdt, satdk = NN(c) \quad (2)$$

4.4. Evaluation

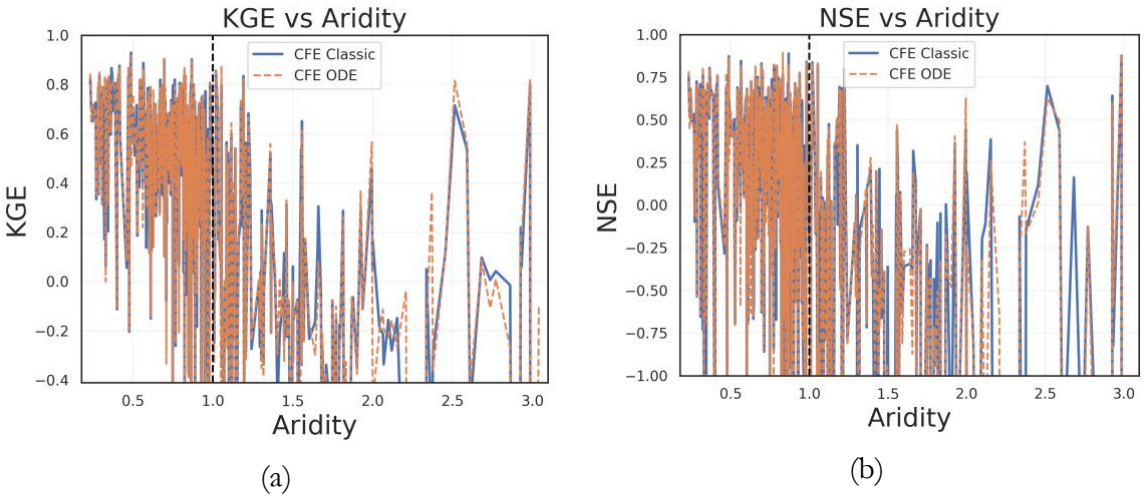
We evaluated the models' streamflow discharge using the Nash-Sutcliffe Efficiency (NSE [33]) and Kling-Gupta Efficiency (KGE [19]). We classified the aridity of the catchments using the CAMELS Aridity Index (PET/P), where it classifies regions with aridity > 1.00 as arid regions.

5. Results and Discussion

5.1. Model evaluation between CFE-classic and CFE-ODE in arid regions

Enhancing soil process representation by introducing an ordinary differential equation did not result in improved streamflow prediction in arid regions, and therefore, we pursue even finer soil representation in Section 5.2.

Similar to previous studies, we observed a CFE performance slump in arid regions (Figure 4). The performance in basins with aridity values beyond 1 rarely exceeded the KGE = -0.41 (Figure 4a) and NSE = 0.0 (Figure 4b), indicating the simulation was no better than the mean flow [34].



(a) (b)

Figure 4: The (a) KGE and (b) NSE performance with increasing aridity for the 498 CAMELS basins. Aridity values beyond 1, on the right hand side of the vertical black dashed line, are classified as arid regions. Note that extremely low KGE and NGE outliers were beyond the Y-axis limits of the figure.

The performance of the CFE-classic and ODE were similar in terms of KGE and NSE values (Figure S3-6). CFE-classic had more basins on the lower end of NSE metrics values (Figure S4

& S6). There were no apparent geographic patterns in terms of which model performed better between the CFE-classic or ODE (Figure 5).

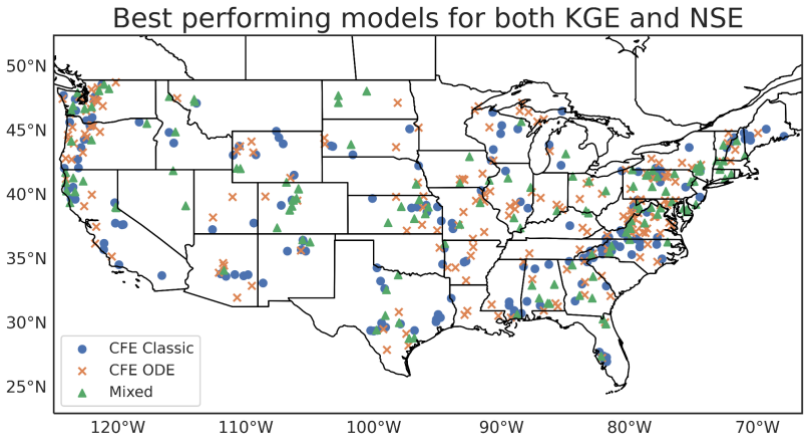


Figure 5: CAMELS basins gauge locations showing the best performing models between CFE Classic and CFE ODE for both KGE and NSE values. The blue circles show the gauges where both KGE and NSE agree that CFE Classic performed better. The orange crosses show the gauges where both KGE and NSE agree that CFE ODE performed better. The green triangles show the gauges where the best-performing models were undetermined; CFE Classic performed better for KGE, CFE ODE performed better for NSE, and vice versa.

This can be attributed to disinformative data or the conceptual design of CFE. To be able to distinguish between the possible causes, additional investigation by excluding data periods with unreliable datasets [8] and testing with additional models in Section 5.2. are needed.

5.2. Demonstrating differentiable versions of LGAR and CFE

5.2.1. dpLGAR (https://github.com/NWC-CUAHSI-Summer-Institute/dpLGAR)

The **d**ifferentiable **p**arameter learning **LGAR** (dpLGAR) module when compared to LGAR-C, with the Phillipsburg, KS test case, gives identical mass-balance and soil simulation results. The internal NN module has been implemented, but is being benchmarked against synthetic soil parameters and is considered a work in progress. See GitHub for the model code/progress.

5.2.2. dCFE (https://github.com/NWC-CUAHSI-Summer-Institute/dCFE)

The **d**ifferentiable **CFE** model (dCFE) provides identical mass balance to the original CFE code but is still being benchmarked against synthetic parameters. We plan to expand the model to incorporate multiple basins and time-varying attributes. The code for the model, gradient chain visualization (Figure S7), and the state of the project can be found in our GitHub linked above.

6. Conclusion

We found ODE solutions provided no improvement, indicating simple parameter tuning outweighs the complexity of the model. Further modeling at finer spatial scales using the Open Geospatial Consortium compliant hydrology artifact model application dataset (Hydrofabric) over the CAMELS extents could reveal greater performance differences. While this was originally intended to be conducted during the 2023 Summer Institute timeframe, data access limitations to forcing data (AORC v1.1), software versioning issues (Hydrofabric "pre-release" version versus the current Hydrofabric v1.2), and incomplete software documentation for successful deployment of NextGen pushed this task out of scope. Furthermore, our project aimed to bridge the recent divide between machine learning models and physics/conceptual-

based modeling. Differentiable modeling methods were applied to CFE and LGAR; however, more work is required in parameter recovery and model tuning.

References

[1] N. Addor, A. J. Newman, N. Mizukami, and M. P. Clark, "The CAMELS data set: catchment attributes and meteorology for large-sample studies," *Hydrol. Earth Syst. Sci.*, vol. 21, no. 10, pp. 5293–5313, Oct. 2017, doi: 10.5194/hess-21-5293-2017.

[2] Newman, Andrew, "A large-sample watershed-scale hydrometeorological dataset for the contiguous USA." UCAR/NCAR - GDEX, p. approximately 2.5 GB, 2014. doi: 10.5065/D6MW2F4D.

[3] D. Feng, K. Fang, and C. Shen, "Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales," *Water Resources Research*, vol. 56, p. 24, 2020, doi: https://doi.org/10.1029/2019WR026793.

[4] F. Kratzert, D. Klotz, G. Shalev, G. Klambauer, S. Hochreiter, and G. Nearing, "Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets," *Hydrology and Earth System Sciences*, vol. 23, no. 12, pp. 5089–5110, Dec. 2019, doi: 10.5194/hess-23-5089-2019.

[5] R. Araki, F. Branger, I. Wiekenkamp, and H. McMillan, "A signature-based approach to quantify soil moisture dynamics under contrasting land-uses," *Hydrological Processes*, vol. 36, no. 4, p. e14553, 2022, doi: 10.1002/hyp.14553.

[6] R. Araki, Y. Mu, and H. McMillan, "Evaluation of GLDAS soil moisture seasonality in arid climates," *Hydrological Sciences Journal*, vol. 0, no. ja, p. null, 2023, doi: 10.1080/02626667.2023.2206032.

[7] T. M. Lahmers *et al.*, "Evaluation of NOAA National Water Model Parameter Calibration in Semi-Arid Environments Prone to Channel Infiltration," *Journal of Hydrometeorology*, Aug. 2021, doi: 10.1175/JHM-D-20-0198.1.

[8] K. Beven and I. Westerberg, "On red herrings and real herrings: disinformation and information in hydrological inference," *Hydrol. Process.*, vol. 25, no. 10, pp. 1676–1680, May 2011, doi: 10.1002/hyp.7963.

[9] K. Beven, "Benchmarking hydrological models for an uncertain future," *Hydrological Processes*, vol. 37, no. 5, p. e14882, May 2023, doi: 10.1002/hyp.14882.

[10] E. Deardorff, A. M. Rad, J. Bales, and T. Flowers, "National Water Center Innovators Program Summer Institute Report 2022," 2022.

[11] M. Gauch, F. Kratzert, D. Klotz, G. Nearing, J. Lin, and S. Hochreiter, "Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 4, pp. 2045–2062, Apr. 2021, doi: 10.5194/hess-25-2045-2021.

[12] F. Kratzert, D. Klotz, G. Shalev, G. Klambauer, S. Hochreiter, and G. Nearing, "Benchmarking a Catchment-Aware Long Short-Term MemoryNetwork (LSTM) for Large-Scale Hydrological Modeling," Global hydrology/Modelling approaches, preprint,

Aug. 2019. doi: 10.5194/hess-2019-368.

[13] A. J. Newman, N. Mizukami, M. P. Clark, A. W. Wood, B. Nijssen, and G. Nearing, "Benchmarking of a Physically Based Hydrologic Model," *Journal of Hydrometeorology*, vol. 18, no. 8, pp. 2215–2225, Aug. 2017, doi: 10.1175/JHM-D-16-0284.1.

[14] M. Gauch, F. Kratzert, D. Klotz, G. Nearing, J. Lin, and S. Hochreiter, "Data for 'Rainfall-Runoff Prediction at Multiple Timescales with a Single Long Short-Term Memory Network.'" Zenodo, Oct. 15, 2020. doi: 10.5281/zenodo.4072701.

[15] F. L. Ogden, "Parameter Estimation for a Conceptual Functional Equivalen (CFE) Formulation of the National Water Model".

[16] M. P. Clark and D. Kavetski, "Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes," *Water Resour. Res.*, vol. 46, no. 10, p. 2009WR008894, Oct. 2010, doi: 10.1029/2009WR008894.

[17] B. A. Tolson and C. A. Shoemaker, "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration: DYNAMICALLY DIMENSIONED SEARCH ALGORITHM," *Water Resour. Res.*, vol. 43, no. 1, Jan. 2007, doi: 10.1029/2005WR004723.

[18] D. Huard, "RavenPy Documentation," 2023.

[19] H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez, "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling," *Journal of Hydrology*, vol. 377, no. 1, pp. 80–91, Oct. 2009, doi: 10.1016/j.jhydrol.2009.08.003.

[20] P. La Follette, F. L. Ogden, and A. Jan, "Layered Green & Ampt Infiltration with Redistribution," *Water Resources Research*, p. e2022WR033742, Apr. 2023, doi: 10.1029/2022WR033742.

[21] L. A. Richards, "CAPILLARY CONDUCTION OF LIQUIDS THROUGH POROUS MEDIUMS," *Physics*, vol. 1, no. 5, pp. 318–333, Nov. 1931, doi: 10.1063/1.1745010.

[22] L. F. Richardson, *Weather Prediction by Numerical Process*. University Press, 1922.

[23] M. Th. van Genuchten, "A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils," *Soil Science Society of America Journal*, vol. 44, no. 5, pp. 892–898, 1980, doi: 10.2136/sssaj1980.03615995004400050002x.

[24] D. Feng, H. Beck, K. Lawson, and C. Shen, "The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact assessment," *Hydrology and Earth System Sciences*, vol. 27, no. 12, pp. 2357–2373, Jun. 2023, doi: 10.5194/hess-27-2357-2023.

[25] W.-P. Tsai *et al.*, "From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling," *Nat Commun*, vol. 12, no. 1, p. 5988, Oct. 2021, doi: 10.1038/s41467-021-26107-z.

[26] D. Feng, J. Liu, K. Lawson, and C. Shen, "Differentiable, Learnable, Regionalized Process-Based Models With Multiphysical Outputs can Approach State-Of-The-Art Hydrologic Prediction Accuracy," *Water Resources Research*, vol. 58, no. 10, p.

e2022WR032404, 2022, doi: 10.1029/2022WR032404.

[27] T. Bindas *et al.*, "Improving large-basin river routing using a differentiable Muskingum-Cunge model and physics-informed machine learning." Preprints, May 25, 2023. doi: 10.22541/essoar.168500246.67971832/v1.

[28] C. Shen *et al.*, "Differentiable modelling to unify machine learning and physical models for geosciences," *Nat Rev Earth Environ*, Jul. 2023, doi: 10.1038/s43017-023-00450-9.

[29] D. Aboelyazeed *et al.*, "A differentiable, physics-informed ecosystem modeling and learning framework for large-scale inverse problems: demonstration with photosynthesis simulations," *Biogeosciences*, vol. 20, no. 13, pp. 2671–2692, 2023, doi: 10.5194/bg-20-2671-2023.

[30] N. W. Chaney *et al.*, "POLARIS Soil Properties: 30-m Probabilistic Maps of Soil Properties Over the Contiguous United States," *Water Resources Research*, vol. 55, no. 4, pp. 2916–2938, 2019, doi: 10.1029/2018WR022797.

[31] M. Leshno, V. Ya. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural Networks*, vol. 6, no. 6, pp. 861–867, Jan. 1993, doi: 10/bjjdg2.

[32] J. Zhang, P. Lin, S. Gao, and Z. Fang, "Understanding the re-infiltration process to simulating streamflow in North Central Texas using the WRF-hydro modeling system," *Journal of Hydrology*, vol. 587, p. 124902, Aug. 2020, doi: 10.1016/j.jhydrol.2020.124902.

[33] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models part I — A discussion of principles," *Journal of Hydrology*, vol. 10, no. 3, pp. 282–290, Apr. 1970, doi: 10.1016/0022-1694(70)90255-6.

[34] W. J. M. Knoben, J. E. Freer, and R. A. Woods, "Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores," Catchment hydrology/Modelling approaches, preprint, Jul. 2019. doi: 10.5194/hess-2019-327.

Chapter 3

Analysis of Flood Drivers Contributions to Compound Flooding Using Coupled Modeling and Machine Learning

Javed Ali[1], Sadaf Mahmoudi[2], Farnaz Yarveysi[3], and Samuel Daramola[4]

[1]University of Central Florida; *javed.ali@ucf.edu*

[2]University of Alabama, Tuscaloosa; *smahmoudikouhi@crimson.ua.edu*

[3]University of Alabama, Tuscaloosa; *fyarveysi@crimson.ua.edu*

[4]Virginia Tech, Blacksburg; *samueldaramola@vt.edu*

Academic Advisors: Thomas Wahl, *University of Central Florida*[1]; Hamed Moftakhari, *University of Alabama*[2]; Hamid Moradkhani, *University of Alabama*[3]; David Muñoz, *Virginia Tech*[4]

Summer Institute Theme Advisor: Kyle T. Mandli, Columbia University, *kyle.mandli@columbia.edu*

Abstract: This study aimed to understand the drivers of compound flooding in New York City (NYC) using advanced hydrological and hydrodynamic modeling techniques and machine learning methods. We employed the National Water Center's Conceptual Functional Equivalent (CFE) model for hydrologic modeling of eight delineated watersheds within the NYC area, and for hydrodynamic modeling, we used GeoClaw, incorporating river discharge data from CFE. By coupling these two modeling methodologies, we simulated nine historical storm events, determining the combined effects of precipitation, storm surge, and river discharge as drivers of compound flooding. To quantify their relative contributions, we trained and validated three machine learning models – Random Forest, Support Vector Machine, and Multi-Layer Perceptron – on the modeled flood depth data. Our findings demonstrated that storm surge was the main cause of compound flooding in NYC, with precipitation also playing an important role. Interestingly, river discharge doesn't have as much impact on these flooding events. These results, supported by historical data, have profound implications for urban planning, disaster management, and policy-making in NYC, providing a solid foundation for developing targeted strategies to mitigate compound flooding.

1. Motivation

Flooding, a phenomenon resulting from drivers, such as heavy rainfall, storm surge, and river discharge, poses significant risks to coastal communities in the United States [1]. The severity and frequency of these events are expected to increase due to climate change, necessitating a comprehensive understanding of the contributing drivers and their interactions [2]. Historically, the study of flooding events has often been compartmentalized, with separate analyses conducted for different types of flooding such as fluvial, pluvial, and coastal contributions. However, in many instances, these events do not occur in isolation but rather in combination, leading to compound flooding. The drivers of these events, including storm characteristics, sea-level rise, and land-use changes, can interact in complex ways to exacerbate flooding impacts. In

New York City (NYC), for instance, the observed rise in compound events has been attributed to a shift toward weather patterns characterized by storm surges that coincide with increased precipitation [3]. Traditional modeling approaches have struggled to accurately capture these interactions due to their inherent complexity and the high dimensionality of the problem. However, recent advances in computational power and the development of sophisticated coupled models that integrate atmospheric, oceanic, and hydrological processes have opened new possibilities for studying compound flooding [4]. Despite these advances, significant challenges remain in analyzing compound flooding events. Machine learning, with its ability to handle high-dimensional data and capture complex patterns [5], offers a promising tool for improving our understanding and analysis of compound flooding.

2. Objectives and Scope

This study aims to assess the relative contributions of various flood drivers that may have contributed to compound flooding in Manhattan, New York City, during various historical storms. To achieve this objective, a coupled model that integrates both hydrological and hydrodynamic processes is implemented. The hydrological component of the model utilizes the National Water Center's CFE (Conceptual Functional Equivalent) model within the NextGen Framework, while the hydrodynamic component employs the GeoClaw numerical model. By coupling these models, the study simulates the impact of hurricanes and tropical storms that have affected New York City in recent years. The outputs of the coupled model provide insights into the flood drivers associated with each simulated event. Additionally, the hydrodynamic modeling component allows for the estimation of flood depths across the study area. To further investigate the contribution of each flood driver to the flood depth at the tract-level resolution, machine learning algorithms were employed. By leveraging various machine learning algorithms, the study aims to determine the relative importance and contribution of each flood driver in different parts of Manhattan. This integrated approach combining hydrologic and hydrodynamic coupled modeling, and machine learning techniques will provide valuable insights into the relative contributions of flood drivers in Manhattan, ultimately enhancing our understanding of flood processes and facilitating effective flood management and mitigation strategies in urban coastal areas.

3. Previous Studies

GeoClaw is a two-dimensional hydrodynamic model that incorporates adaptive mesh refinement capabilities, primarily designed for simulating shallow earth-surface flows involving water-wave propagation and inundation, including scenarios such as tsunamis, storm surges, and general overland flooding [6]. High-resolution finite volume methods are employed in GeoClaw to address geophysical flow problems. Spero et al. [7] compared GeoClaw with HEC-RAS for modeling the 1976 Teton Dam failure. The evaluation of GeoClaw's suitability for dam failure modeling was based on its ability to accurately depict the extent of inundation and the arrival times of flood waves. The study found that the 2D GeoClaw dam-break model produced results that reasonably aligned with historical gauge records, field observational data, and HEC-RAS results. The model demonstrated stability and relatively low computational costs. While

GeoClaw has been predominantly utilized in dam failure, tsunami, and geo-hazard studies, there are limited examples of its application in flood modeling. The use of GeoClaw in coastal flooding and storm surge simulations is not extensively documented. Hence, in this study, we propose to employ GeoClaw and validate its performance against observational data from NOAA tide gauges to assess its efficacy in simulating coastal flooding events. The selection of the most significant features plays a crucial role in pattern recognition systems. Nowadays, when examining the combined effects of multiple variables, researchers often employ machine learning techniques to obtain relevance scores [8]. For instance, Yarveysi et al. [1] used a machine learning algorithm to objectively assign weights to variables contributing to the overall estimated vulnerability, thus reducing subjectivity in determining the impact of various social, economic, and infrastructural factors on vulnerability. Similarly, Opoku et al. [9] utilized five supervised machine learning algorithms to predict depression. Through the permutation importance method, they were able to identify influential behavioral markers in the prediction of depression.

4. Methodology

4.1 Data and Study Area

This study focuses on analyzing historic storms that occurred between 2007 and 2019 and impacted New York City (NYC) (Figure 1a). We selected eight catchments that cover the whole study area (Figure 1b). The selected time period aligns with the capabilities of the hydrofabric AORC (Atmospheric Oceanic Reanalysis and Characterization) data utilized in this study.

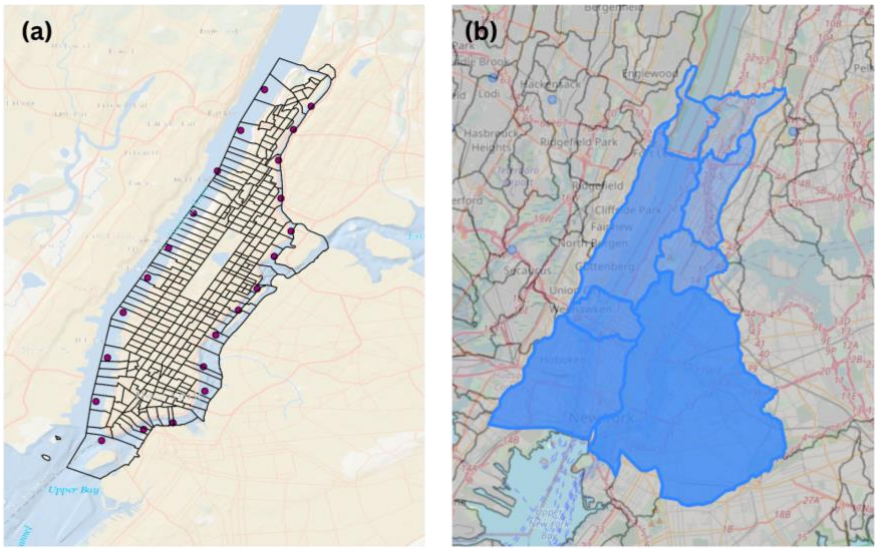


Figure 1: (a) Study area: Manhattan, New York City, and (b) eight catchments used in the study area.

Table 1 presents the information about the storms, and their ATCF (Automated Tropical Cyclone Forecasting) names. All 9 storms included in the analysis are simulated using the CFE model, and the maximum discharge associated with each storm is subsequently used as an input parameter for the GeoClaw hydrodynamic model. The storm data utilized in GeoClaw, including the storm's eye latitude, longitude, pressure, wind speed, and radius, can be obtained from the following repository: https://ftp.nhc.noaa.gov/atcf/archive/. By incorporating the storm data from these selected events, this study aims to comprehensively examine the hydrological and

hydrodynamic processes associated with historic storms in NYC, providing valuable insights into flood dynamics and contributing to enhanced flood risk management strategies.

Table 1: Information about the storm events, their respective landfall dates, the dates of impact on NYC, the hour of landfall, and their ATCF names

Storms	Date of Landfall	Date of impact in NYC	Hour of landfall	ATCF data
Tropical Storm Barry	2-Jun-2007	5-Jun-2007	02:00	AL022007
Hurricane Hannah	6-Sep-2008	6-Sep-2008	07:20	AL082008
Hurricane Irene	28-Aug-2011	28-Aug-2011	13:00	AL092011
Hurricane Sandy	29-Oct-2012	29-Oct-2012	23:30	AL182012
Hurricane Arthur	4-Jul-2014	4-Jul-2014	08:00	AL012014
Tropical Storm Jose	19-Sep-2017	20-Sep-2017	00:00	AL122017
Tropical Storm Philippe	28-Oct-2017	30-Oct-2017	22:00	AL182017
Hurricane Dorian	6-Sep-2019	7-Sep-2019	12:30	AL052019
Hurricane Ogla	27-Oct-2019	27-Oct-2019	03:00	AL172019

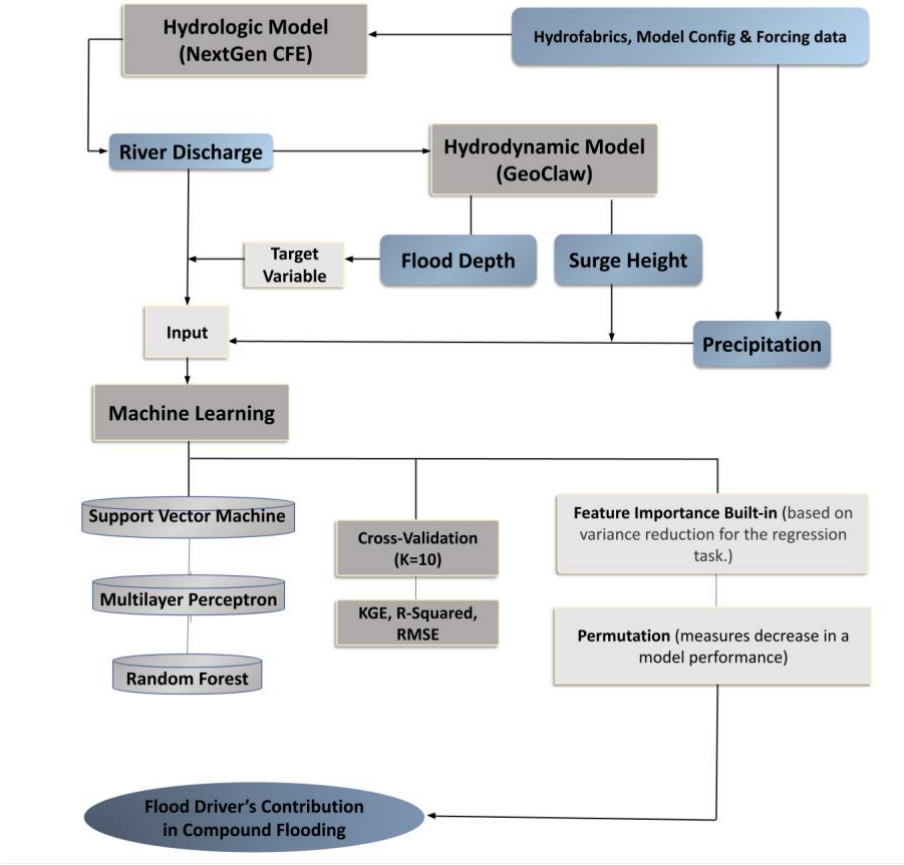


Figure 2: Flowchart showing the methodology employed in this study.

4.2 Hydrologic Modeling using the National Water Center's CFE Model

The National Water Center's Conceptual Functional Equivalent (CFE) model, a conceptual rainfall-runoff model designed to simulate rainfall conversion into runoff, was utilized in this study to understand the volume of water flowing into rivers and streams post-rain events. The initial stage of hydrologic modeling with the CFE entails obtaining and processing HydroFabric data pertinent to the eight delineated watersheds within the study area. These data are secured from the Amazon S3 Bucket and processed to construct parameter configuration files. These configuration files, pivotal for the operation of the CFE model and the Simple Logical Tautology Handler (SLoTH) within the NextGen framework, consist of model default parameters, specific formulations, detailed input and output paths, simulation time steps, and initial conditions, along with other settings relevant for the precise modeling of the hydrological system under study. Following this, basin-averaged forcing inputs are generated, each designed to correspond with distinct storm event time periods within the individual watersheds. These inputs, drawn from AORC v1.0 kerchunk header files, are specifically prepared for integration with NOAA's advanced Next Generation (NextGen) Water Resource Modeling Framework. Finally, the CFE model is executed within the NGEN framework using the 'ngen' command, which consolidates the positional arguments for running the model. More information on this process can be found in the GitHub repository.

4.3 Hydrodynamic Modeling using GeoClaw

i) Setting up the model for validation: The model was set up according to the descriptive specifications in the clawpack repository and our model repository on GitHub.

ii) Incorporating river discharge: The subroutine "src2" script, written in Fortran, was used to integrate the river discharge into the model. It sets certain geographical bounds for the river source and computes the river's discharge in cubic meters per second for cells that fall within the river source area.

iii) Bias Correction: A maximum bias correction is a statistical approach employed in hydrological and meteorological modeling to address systematic biases in modeled variables. The methodology of maximum bias correction [10] involves identifying the maximum bias between the modeled and observed data, followed by the application of a correction factor to mitigate or eliminate the bias. The primary objective of maximum bias correction is to enhance the accuracy and reliability of modeled outputs.

4.4 Machine Learning Approach

The outcome of coupled modeling (including precipitation, storm surge, and river discharge) for the nine historical storms was used as input features to the various ML algorithms tested here. These algorithms are trained and validated using the flood depth estimates from the GeoClaw hydrodynamic model, split 80/20 for training/validation. The depth estimates are obtained at 20 computation points randomly selected around Manhattan. We tested the applicability and performance of three different ML algorithms namely Random Forest (RF), Support Vector Machine Regression (SVMR), and Multi-Layer Perceptron (MLP). To select the best-performing algorithm, we consider various evaluation metrics (i.e., KGE and RMSE). RF's inherent feature

importance mechanism was used for the regression task, while the permutation importance method was utilized for SVMR, and MLP to calculate feature importance. K-fold cross-validation (K=10) was employed to verify the algorithm's performance on new data, ensuring stability in accuracy estimates.

4.5 Evaluation Metrics

Four evaluation metrics were used to assess the effectiveness of the machine learning and the hydrodynamic models.: the Nash-Sutcliffe efficiency (NSE) [11], gauging the predictive accuracy and reliability of the model; the Kling-Gupta efficiency (KGE) [12], providing an overall assessment of the model's ability to replicate observed values in terms of timing, magnitude, and variability; the coefficient of determination (R-squared), representing the proportion of variance in the dependent variable that can be attributed to the independent variables; and the root mean square error (RMSE), which assessed the accuracy and precision of the model by measuring the average discrepancy between predicted and observed values [13].

5. Results

5.1 Hydrologic Modeling with NextGen CFE Model

Figure 3 shows the river discharge for different catchments over time during Hurricane Irene (2011). The river discharge for each catchment generally increases over time, which is expected as the hurricane progresses. There is a notable peak in discharge rates for all catchments around August 28, which likely corresponds to the height of the hurricane. Catchment "cat-694852" consistently has the highest discharge rate throughout the event, reaching close to 120,000 cubic meters per second at its peak. Catchments "cat-694853" and "cat-694854" also show significant increases in discharge during the hurricane, with maximum values around 46,000 and 26,000 cubic meters per second, respectively. The rest of the catchments have lower overall discharge rates, with maximum values not exceeding 24,000 cubic meters per second. The high values of discharge could be attributed to the intense rainfall and strong winds associated with Hurricane Irene. These conditions can result in a significant increase in river discharge, as more water is transported into rivers and streams from direct rainfall and surface runoff. Results for other storm events are shown in the supplementary information.

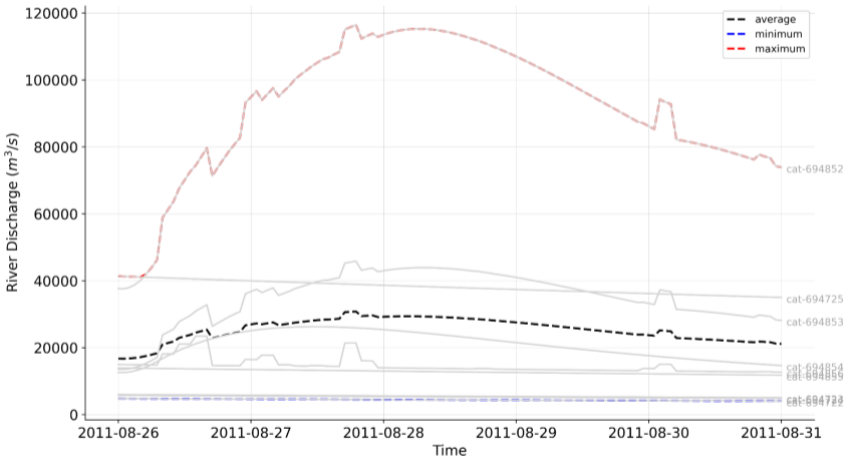


Figure 3: River discharge output from the NextGen CFE model during Hurricane Irene (2011) for different catchments in the NYC area.

5.2 Hydrodynamic Modeling using GeoClaw Model

In this study, the model's performance was validated in the context of two hurricanes, Irene (2011) and Sandy (2012). The results of Sandy are shown in the Supplementary Information. The performance of the GeoClaw model in capturing storm events exhibits notable proficiency in capturing the peak of storms. However, during the initial days leading up to the storm event, the model may exhibit variations. Consequently, to ensure a reliable validation process, the decision was made to focus solely on validating the model against the peak of the storms.

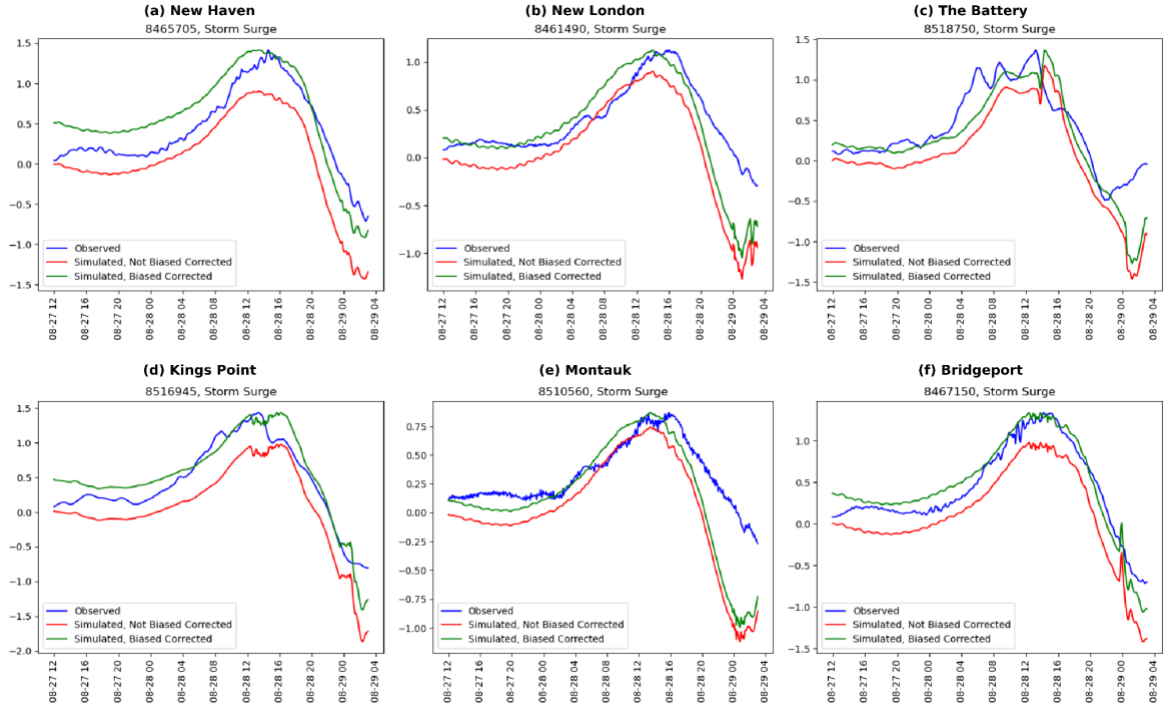


Figure 4: Comparison of observed, simulated, and maximum biased corrected simulated storm surges during Hurricane Irene (2011).

Figure 4 shows the comparison between the simulation and observation time series at the time of the storm's peak during Hurricane Irene, which provides evidence of GeoClaw's proficiency in capturing the peak of the storm. Moreover, evaluation metrics of the biased corrected and not

biased corrected simulation results of Hurricane Irene are shown in Table 2. These findings highlight the station-specific effects of bias correction on the evaluation of the models used in the study.

Table 2: The evaluation metrics of simulated storm surges during Hurricane Irene; the bold ones are after bias correction.

Hurricane Irene's Evaluation Metrics					
Station Name	Station ID	NSE	KGE	RMSE	R Squared
The Battery	8518750	0.204⇒**0.559**	0.313⇒**0.594**	0.43⇒**0.32**	0.204⇒**0.559**
Kings Point	8516945	0.515⇒**0.847**	0.344⇒**0.737**	0.391⇒**0.219**	0.515⇒**0.847**
Montauk	8510560	-0.188⇒**-0.45**	0.109⇒**0.077**	0.296⇒**0.327**	-0.188⇒**-0.45**
Bridgeport	8467150	0.647⇒**0.917**	0.428⇒**0.835**	0.292⇒**0.146**	0.647⇒**0.917**
New Haven	8465705	0.612⇒**0.685**	0.342⇒**0.568**	0.309⇒**0.278**	0.612⇒**0.685**
New London	8461490	0.353⇒**0.299**	0.324⇒**0.478**	0.291⇒**0.303**	0.353⇒**0.299**

5.3 Machine Learning

We examined a variety of performance metrics to determine which machine learning (ML) algorithms are most suited to the specific objectives of this project. We found that MLP is the best-performing algorithm (Figures 5a, 5b, and 5c). This algorithm, with a median KGE of 0.82, and a median RMSE of 0.17 m, demonstrates significantly better performance than SVM and slightly better than RF. Results from the feature importance analysis implemented using the trained MLP (Figure 5e) showed storm surge as the predominant factor contributing to the impacts of compound flooding around Manhattan, with an average relative importance score of 0.53. Local precipitation, although the secondary contributor, still holds substantial influence, exceeding storm surge in certain areas and averaging an importance score of 0.29. River discharge, however, has a relatively limited contribution to flood severity around Manhattan, with an average score of 0.17. These findings align with existing literature. For instance, comprehensive studies of compound historic floods in the lower Hudson River (such as [14]) concur that the river flow's contribution to flood variability is negligible to extremely limited. Thus, our analysis reaffirms these observations, providing valuable insights for future flood mitigation efforts in the Manhattan area. For further details, please refer to Supplementary information.

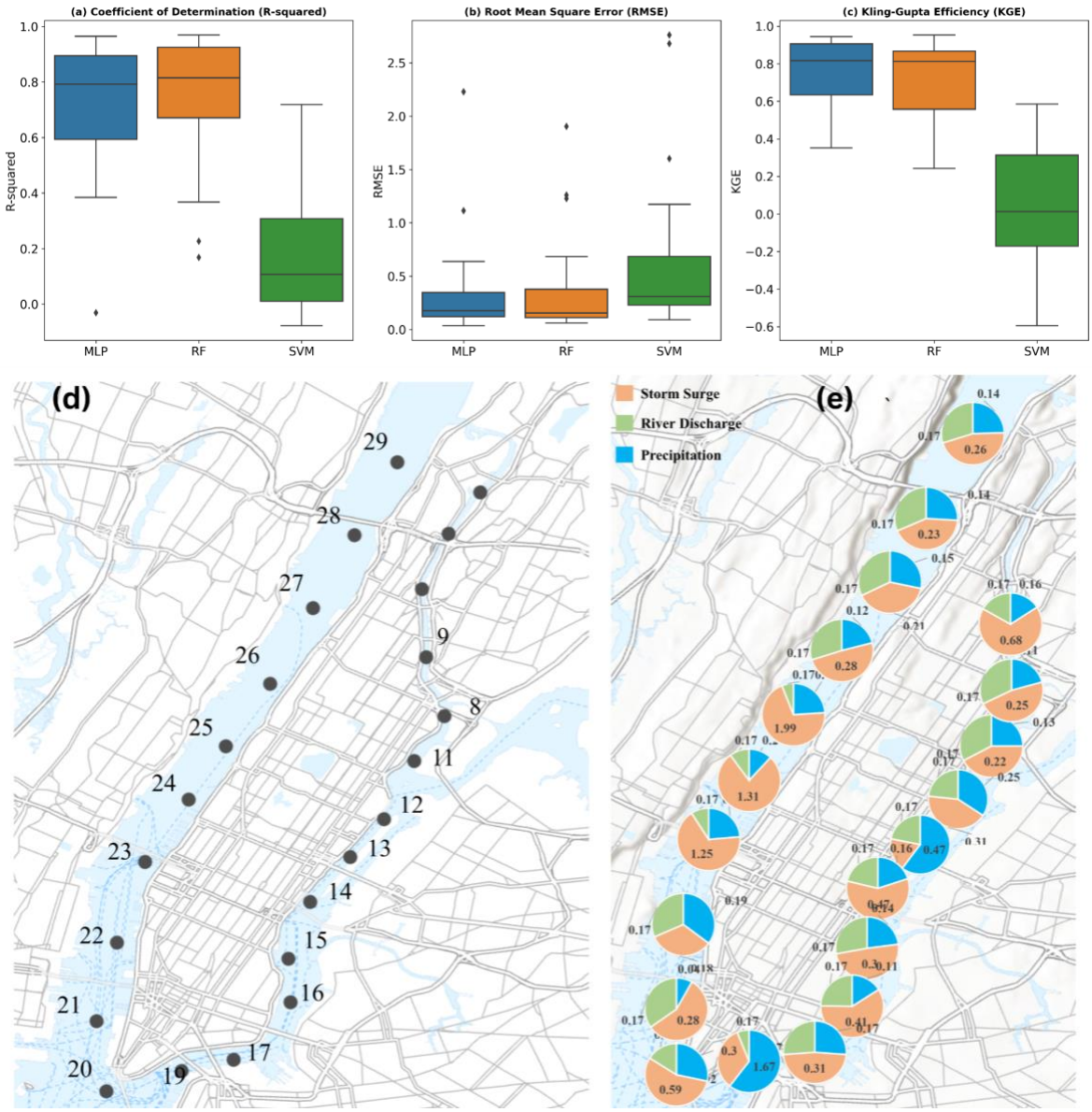


Figure 5: Performance metrics of various ML algorithms (a) R-squared, (b) RMSE, (c) KGE, (d) study locations (numbered), (e) the relative importance of different flood drivers at different locations in the Manhattan area.

6. Conclusion

In this study, we conducted a thorough analysis of nine historic storms that caused significant flooding impacts in Manhattan, New York City. We developed a coupled hydrologic-hydrodynamic modeling framework that uses NextGEN CFE for hydrologic modeling to simulate runoff generated during major storms and GeoClaw to estimate water level variation in the study area based on the fluvial flux received from the NextGEN-CFE model and other forms of reanalysis data such as ocean and atmospheric boundary forcing. The main output of the coupled modeling framework was the calculation of water levels at different locations throughout Manhattan, which serve as the target variables for the machine learning (ML) algorithms implemented within the study. By incorporating the intensity of fluvial, pluvial, and coastal drivers as input features, these sophisticated ML algorithms allowed us to analyze the contribution of each of these drivers to the overall water level variability at different locations

across the study area during compound flooding events. The study's findings underscore that storm surge emerges as the determinant factor in understanding the complex dynamics of compound flooding in Manhattan. Moreover, local precipitation emerges as a secondary, yet significant, contributor. Contrarily, river discharge does not demonstrate a substantial role in influencing the dynamic variability of the flooding regime within the study area.

Supplementary Material: The supplementary information can be found here. The computational code is accessible at our open-access GitHub repository https://github.com/javedali99/si2023-compound-flooding

References

1. Yarveisi, F., Alipour,A., Moftakhari, H., Jafarzadegan, K., Moradkhani, H.(2023). Block-Level Vulnerability Assessment Reveals Disproportionate Impacts of Natural Hazards across the Conterminous United States. Nature Communications. DOI: 10.1038/s41467-023-39853-z.
2. Ming, X., Liang, Q., Dawson, R., Xia, X., & Hou, J. (2022). A quantitative multi-hazard risk assessment framework for compound flooding considering hazard inter-dependencies and interactions. *Journal of Hydrology*, *607*, 127477. https://doi.org/10.1016/j.jhydrol.2022.127477.
3. Lacy, H. S., DeVito, A., & De Nivo, A. C. (2013). Geotechnical Aspects of Three Storm Surge Barrier Sites to Protect New York City from Flooding. In Storm Surge Barriers to Protect New York City: Against The Deluge (pp. 134-149). https://doi.org/10.1061/9780784412527.010.
4. Olbert, A. I., Moradian, S., Nash, S., Comer, J., Kazmierczak, B., Falconer, R. A., & Hartnett, M. (2023). Combined statistical and hydrodynamic modelling of compound flooding in coastal areas—Methodology and application. *Journal of Hydrology*, *620*, 129383. https://doi.org/10.1016/j.jhydrol.2023.129383
5. Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. *Archives of Computational Methods in Engineering*, *27*(4), 1071–1092. https://doi.org/10.1007/s11831-019-09344-w.
6. Berger, M. J., George, D. L., LeVeque, R. J., & Mandli, K. T. (2011). The GeoClaw software for depth-averaged flows with adaptive refinement. Advances in Water Resources, 34(9), 1195–1206. https://doi.org/10.1016/j.advwatres.2011.02.016.
7. Spero, H., Calhoun, D., & Schubert, M. (2022). Simulating the 1976 Teton Dam Failure using Geoclaw and HEC-RAS and comparing with Historical Observations (arXiv:2206.00766). arXiv. http://arxiv.org/abs/2206.00766.
8. J. M. Fontana, M. Farooq and E. Sazonov, "Estimation of feature importance for food intake detection based on Random Forests classification," 2013 35th Annual Internabltional Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 2013, pp. 6756-6759, doi: 10.1109/EMBC.2013.6611107.
9. Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting Depression From Smartphone Behavioral Markers Using Machine Learning Methods, Hyperparameter Optimization, and Feature Importance Analysis: Exploratory Study. JMIR Mhealth Uhealth 2021;9(7):e26540.

10. Teklu T. Hailegeorgis and Knut Alfredsen, "Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway", Journal of Hydrology: Regional Studies, vol. 9, pp. 104-126, 2017.
11. Nash, J.E. and Sutcliffe, J.V. (1970) River Flow Forecasting through Conceptual Model. Part 1— A Discussion of Principles. Journal of Hydrology, 10, 282-290.
12. Gupta, H. V., & Kling, H. (2009). On the evaluation of hydrological models: towards a measurable paradigm. Hydrology and Earth System Sciences, 13(12), 1869-1880.
13. Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.
14. Orton, P., Georgas, N., Blumberg, A., & Pullen, J. (2012). Detailed modeling of recent severe storm tides in estuaries of the New York City region. Journal of Geophysical Research: Oceans, 117(C9). https://doi.org/10.1029/2012JC008220

Chapter 4

Predicting Flood Inundation Susceptibility Using HAND FIM, Crowd-Sourced and Satellite Data with Machine Learning

Azadeh Hosseinzadeh[1], Elnaz Heidari[2], Monica Cardona[3], Roja Najafi[4]

[1]Southern Methodist University; *ahosseinzadeh@mail.smu.edu*
[2]Clemson University; *eheidar@clemson.edu*
[3]Florida International University; *mcard041@fiu.edu*
[4]Brigham Young University; *rnajafi@byu.edu*

Academic Advisors: Barbara Minsker, *Southern Methodist University*[1]; Vidya Samadi, *Clemson University*[2]; Michael Sukop, *Florida International University*[3]; Dan Ames, *Brigham Young University*[4].

Summer Institute Theme Advisor: Barbara Minsker, Southern Methodist University, *minsker@mail.smu.edu*

Abstract: This study introduces a hybrid flood inundation modeling approach that integrates diverse sources of information, including crowd-sourced data, hydrological information, Synthetic Aperture Radar (SAR) satellite data, Federal Emergency Management Agency (FEMA) National Flood Insurance Program (NFIP) claims, and predictions from HAND FIM (Height Above Nearest Drainage Flood Inundation Mapping). The analysis aims to enhance the accuracy of flood inundation maps, particularly in urban areas, as flood models and mapping rely significantly on physical inputs and characteristics that are often unavailable from a single data source. The method utilizes a random forest algorithm as a classification technique to predict flood susceptibility across a 30-meter grid from historical storm data. The results are inconclusive in the effectiveness of the specific approach, however considering there is a probe of concept in the probabilities to utilize classification methods that combine both physical and social features it can serve to the HAND flood inundation maps to improve accuracy compared to traditional methods. The integration of data captured by crowd-sourced software can enhance the model's adaptability to changing urban environments, accounting for local variations, and providing valuable insights for emergency response planning. These advances are needed for the evolution of emergency response operations, like advancements seen in other industries.

Keywords: Hydro informatics, Crowd-sourced Data, Machine Learning, Citizen Science, Susceptibility, Flood Inundation Mapping, Hydraulics and Hydrology**,** Random Forest Algorithm**.**

1. Motivation

Floods are recognized as among the most perilous and economically devastating natural disasters worldwide. In the continental United States alone, the Federal Emergency Management Agency

(FEMA) has reported that flooding surpasses all other severe weather-related events, causing more damages than any other severe weather-related event and many fatalities. The magnitude of the problem emphasizes the urgent need for effective flood management strategies.

To enhance flood management and improve the accuracy of hydrological models, particularly for urban flooding events, crowd-sourced data has been shown to be a valuable resource [17]. This analysis evaluates how geo-citizen, volunteered geographical information and other non-traditional data can be combined with more traditional flood inundation modeling in urban areas, where large populations are at risk.

A major challenge in flood mapping is to estimate the horizontal extent of the flooding [12]. Predictions typically rely on empirical methods and verification with numerical modeling, satellite images, aerial photographs, and surveying data [8], [10], [11],[13]. The approach taken in this report is intended to evaluate the value of several non-traditional data sources for improving existing modeling techniques.

Non-traditional data such as geo-tagged social media posts and other volunteer citizens' disaster occurrence reports have emerged as valuable sources for urban flood monitoring and prediction (e.g., [18]). During flood events, people utilize several venues to share vital information, including the location and severity of flooding, along with photos and videos that can aid emergency responders in assessing the situation and allocating resources effectively. However, there are still some challenges with this approach as citizen's observations and occurrences reporting data can be noisy and problematic to interpret, and verifying its accuracy poses difficulties, such as concerns regarding privacy and security when sharing personal information [1]. Hence.This study combines citizen data with other available data sources and model predictions in an integrated, hybrid machine learning approach.

2. Objectives and Scope

Enhance flood inundation mapping accuracy by incorporating crowd-sourced and other non-traditional data to improve the precision and reliability of flood extent estimation. Currently, most urban flood mapping relies on empirical methods, using streamflow data and rating curves produced by the USGS for model calibration and validation. This project advances the use of real observations made by citizens to improve flood mapping and modeling in urban areas, addressing challenges of complex terrain, changing drainage patterns, urban sprawl, and ongoing infrastructure modifications. The approach also incorporates satellite data and elements from traditional modeling techniques into hybrid probabilistic machine learning models to estimate flood extents with more confidence.

3. Previous Studies

By analyzing social media data, real-time flood maps can be created, enabling the identification of high-risk areas with a higher likelihood of occurrences, and tracking the movement of people and resources within and outside the flood zone. This information allows emergency response

decision-making and resource allocation [2]. Li et al. [4] propose an approach that utilizes flood-related tweets to assist in rapid flood mapping and enhance situational awareness during flooding events. By analyzing spatiotemporal patterns of flood-related tweets, patterns can be identified to improve the utilization of Twitter data for flood mapping [3].

In another study, Mazzoleni et al. [5] presents a methodology for integrating crowd-sourced data from social and physical sensors into hydrological models, aiming to enhance streamflow prediction in early warning systems. The methodology is applied to case studies in the UK, Italy, and Luxembourg, demonstrating that assimilating crowdsourced data, particularly at high frequencies and accuracies, improves streamflow prediction accuracy.

4. **Data and Research Area**

This research focuses on the analysis of ten historical storms that have caused urban flooding in Buffalo Bayou Watershed since 2015 (Table 1). This watershed is in southeastern Texas, covering most of Harris County, including the city of Houston (Figure 1). Buffalo Bayou, which is the mainstream within the catchment, serves as the primary outlet for stormwater runoff from the area. It stretches approximately eighty-five kilometers in length and is bordered on the western side by the Addicks and Barker reservoirs. The catchment is highly susceptible to urban flooding due to its urbanized nature and extensive impervious surfaces in the form of roads, buildings, and other infrastructure within the city of Houston and its suburbs [14].

During Hurricane Harvey in 2017, Buffalo Bayou Watershed (BBW) and Brays Bayou Watershed (BrBW) suffered significant impacts, with 17,090 and 23,810 houses damaged, respectively [15]. These numbers are the highest among the twenty-two watersheds in Harris County affected by Harvey. Additionally, the region has experienced other significant flood events (Table 1), such as the Memorial Day flood in May 2015, Tax Day flood in April 2016, and tropical storm Imelda in September 2019. These natural disasters resulted in significant road flooding and inundation along the Buffalo and Brays Bayous [16].

Table 2 shows the data that are processed into attributes for the machine learning model described below, including hydrologic data, Synthetic Aperture Radar (SAR) satellite data, crowd-sourced data, National Flood Insurance Program (NFIP) flood claims during historical storms, and HAND FIM (Height Above Nearest Drainage Flood Inundation Mapping) data provided by the National Water Center (Figure 2).

Satellite data are only available for the storms noted in Table 1. HAND FIM data are used to estimate the time of peak flood inundation for each storm (also shown in Table 1), which is the time used for model training, testing, and validation. For the purposes of modeling, the watershed is divided into 259,999 cells 30m by 30-meter 390,450grid. Data on building footprints are used to eliminate any grid cells that contain only buildings, resulting in 259,999 cells grid cells for analysis.

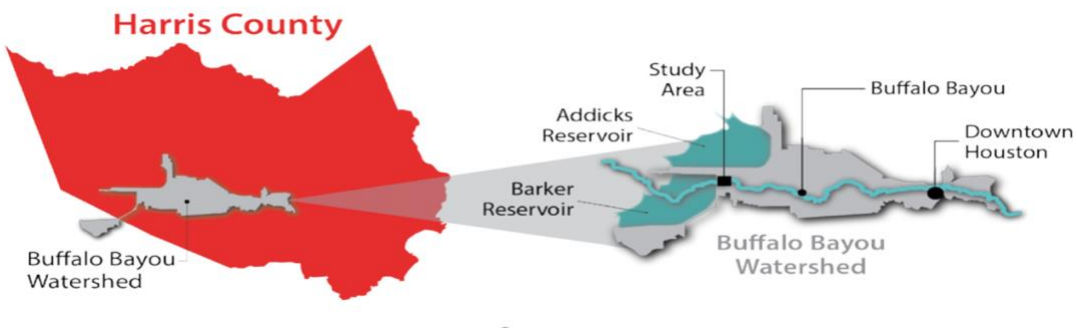


Figure 1. Research area: Buffalo Bayou watershed

5. Methodology

A probabilistic approach is taken in this research, using a random forest machine learning algorithm [6], [7] to classify each grid cell in the watershed as flooded or unflooded and provide a confidence level associated with each classification. Thus, the predicted confidence level provides an estimate of the 25 susceptibilities of each location to flood occurrence during a particular storm. The random forest algorithm's robustness, low bias, capability of handling unbalanced and high dimensional data, and its quick prediction makes it a useful tool for this research among other machine learning methods.

The probability of flood occurrence is estimated from the attribute datasets given in Table 1 (Supplementary Material), including HAND FIM predictions [9], topographic, hydrologic, satellite, and crowd-sourced features. Observations of flooding from any source (Twitter posts, 311 reports - the local authorities' citizen complaint hotline or app, FEMA flood related damages reported as claims to the National Flood Insurance Program, and SAR satellite data) are combined into a single attribute (0/1 for not flooded/flooded). This attribute is used in the training phase of the modeling and then omitted (predicted) for the testing and validation phases. Spatial correlations are considered by using crowd-sourced and satellite data from surrounding grid cells as other attributes. Additionally, any previous flood damage insurance claims at each grid cell are considered as another attribute.

An ensemble modeling approach is used, where data from nine storms are used to train and test storm-specific random forest models, whose predictions are then input into a "committee" random forest model that makes the final prediction for any storm (see Figure 2). This structure allows the storm-specific models to specialize on the hydrologic patterns unique to each storm, which are then generalized to other storm patterns in the committee model. Furthermore, satellite data can be incorporated into the storm-specific models where available without causing difficulties with missing data for the other storms.

A 10-fold cross-validation approach is used for training and testing, where the model is trained on 90 percent of the data (9 folds) and 10 percent (1-fold) are held out for testing. This process is repeated ten times, each with different training and testing folds. As noted above, during

testing, the reported flooding at each cell from all sources is the predictive variable as a surrogate for flood susceptibility.

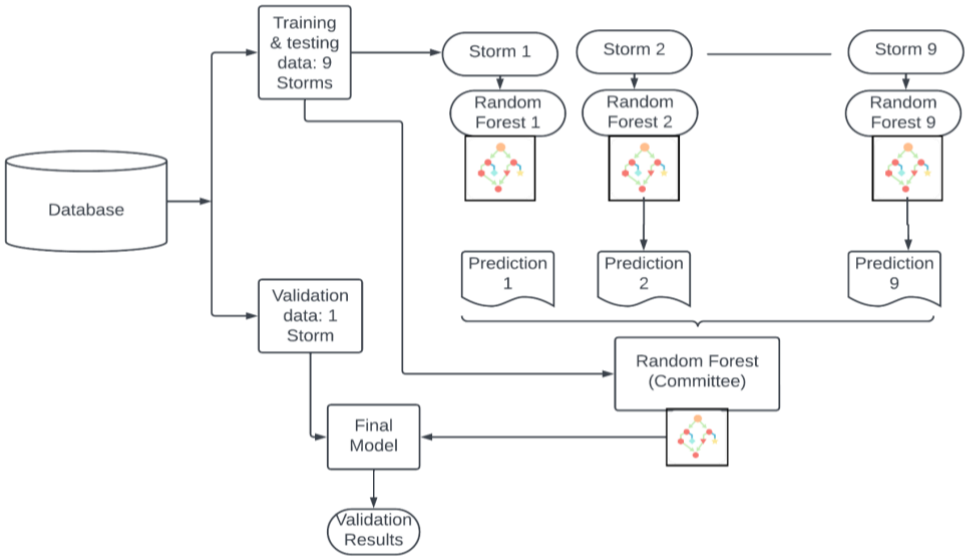


Figure 2. Diagram of Random Forest Classification Algorithm.

Finally, the performance of the model is validated using satellite, FEMA flood claims, and crowd-sourced data from one storm in the dataset (Tropical Storm Imelda) that was not used for training or testing (Figure 2).

5. Results

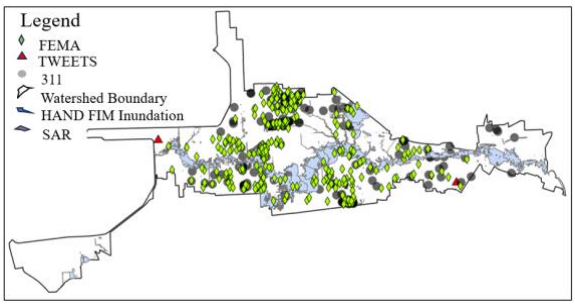
The machine learning algorithm used Cross-validation to assess the model performance by partitioning the data into multiple subsets or folds 10 by each storm - k-fold cross-validation, the model was trained on k-1 folds and tested on the remaining fold in each iteration with this process repeated k times, with each fold serving as the test set exactly once. The final evaluation metric was computed by averaging the performance results from all k iterations, providing a more robust estimate of the model's performance compared to a single train-test split.

While the results show promise, several assumptions must be considered. First, all flood observations from satellite, crowd-sourced, and HAND FIM sources are considered equally valid. All these data sources have errors that should be investigated in future research. Crowd-sourced data can be biased, as some citizens are more likely to report flooding than others, and erroneous, whether intentional or unintentional. For example, the relevance of Twitter posts to flooding was assessed by Flood Tags using their proprietary algorithms, which could have included irrelevant posts or excluded relevant posts. Second, due to limited time, only one machine learning method was investigated and default hyperparameters were used. A more thorough investigation of methods, importance of different attributes, and hyperparameter optimization is needed in future research.

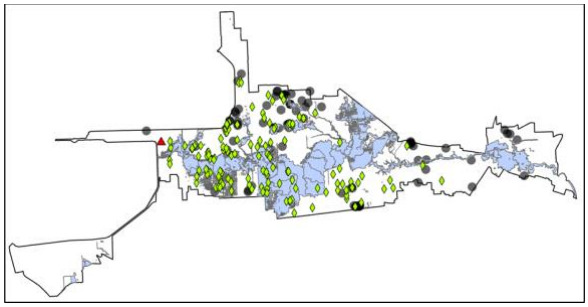
Estimating there were some limitations within the features and all the features introduced to achieve viable results and proof of concept. However, the final and conclusive results are part of final deliverables to be presented during the CUAHSI- 2023 Summer Institute capstone.

5.1. Data Exploration Results:

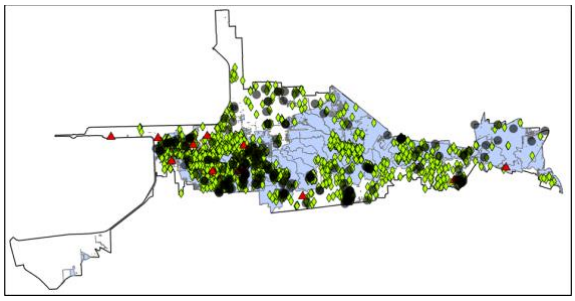
These maps (Map 1 and Map 2) of the watershed in some of the storms analyzed display the extent of the HAND inundation map and human reported occurrences, along with the SAR data when available and the FEMA flood related damages contributing to the contributions and potentials of the crowd-source data for this type of analyses. The map highlights the horizontal representation of the map inundation prediction during the peak of the storm (flood) light blue shading, the reported and documented occurrences are shown as dots, triangles and squares and urban areas marked with red symbols, and SAR data when available showing inundation areas are seen in pink, gray. Watershed and boundaries demarcate the borders between neighboring Hydrological Units.



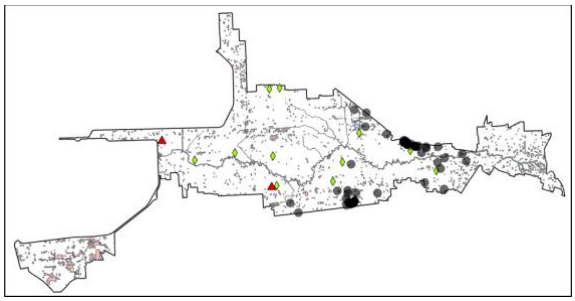
(A)TS Bill



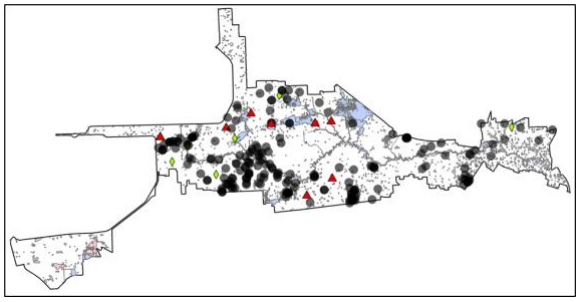
(B)Tax Day



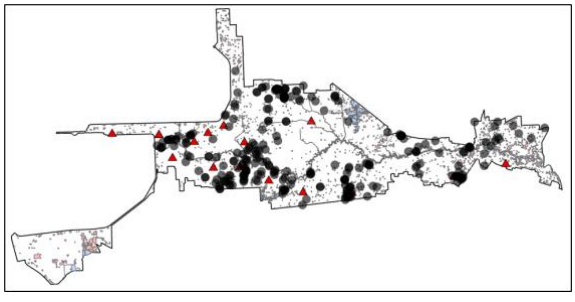
(C)EVENT2017



(D)EVENT2018

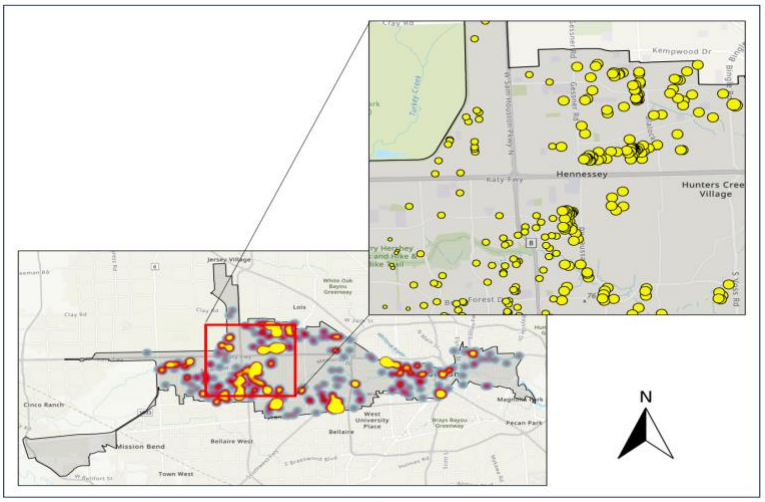


(E)TS Imelda



(F)TS Beta

Map 1. Heat map generated from all crowd-sourced data across storms, as well as highlighting the most flood-prone area within the Buffalo Bayou Watershed.



Map 2. Heat map generated from all crowd-sourced data across storms, as well as highlighting the most flood-prone area within the Buffalo Bayou Watershed.

Below Graphs (Figure 3) show rain return rates by hour for each storm which is a relationship between peak of rainfall between 6 hours before the peak and the peak of the greatest horizontal extension of the flood. This is an average time interval between the analyzed storms only considering similar or greater peak rainfall intensity causing flood conditions in the area.
The formula used for this calculation formula: Return Time = 1 / Exceedance Probability, where Exceedance Probability is the probability of exceeding the peak rainfall intensity of a given storm. The lower return time indicates a higher likelihood of storms, meaning flooding conditions with similar or greater peak rainfall intensity occurring more frequently.

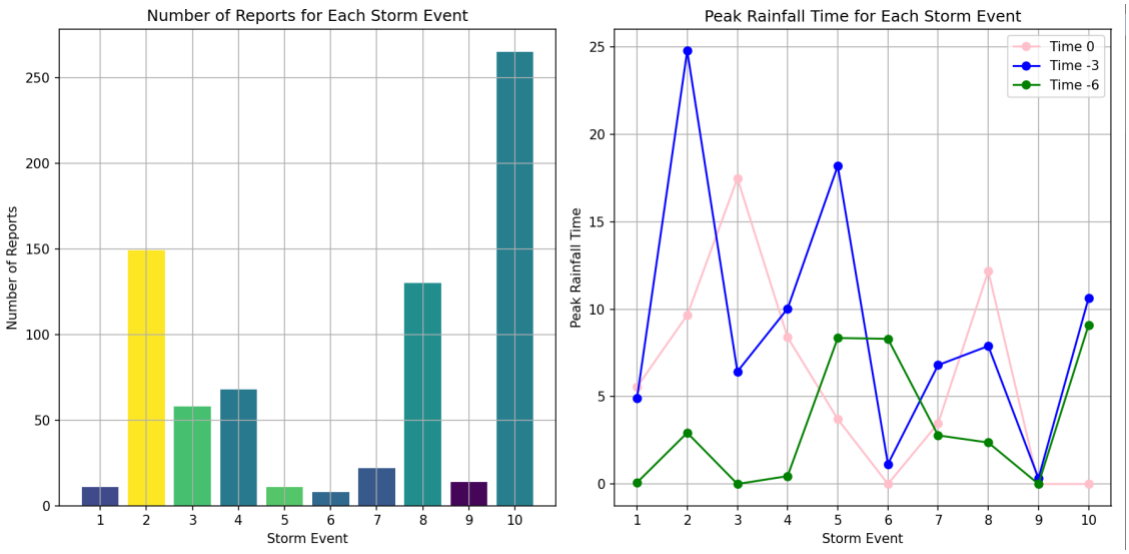


Figure 3. Rainfall returns within 0-3-6 hours window.

6. Conclusions

A novel hybrid machine learning method is developed that combines HAND FIM predictions with satellite, hydrological, and crowd-sourced data to predict flood extent in urban environments. An ensemble random forest classification model is developed with specialized models predicting flood extent for each storm in the historical record. The output from these models is fed into a committee model that predicts areas of flood impacts for all storms.

There are promising options to enable real-time and widespread data collection during flood events, providing valuable information for rapid response and decision-making, along for a substantial number of contributors, crowd-sourced data can improve the spatial resolution of flood inundation maps, capturing localized flood extents with finer detail.

Citizens data and the already existing personal technology allows for geotagged images and videos that can serve as valuable ground truth for validating flood models and enhancing their accuracy; involving citizens in data collection fosters community engagement and awareness about flood risks, empowering individuals to take preventive measures; and Real-time data from smartphones can contribute to flood monitoring systems and early warning mechanisms, alerting authorities and the public to potential hazards.

In general data fusion techniques can be applied to integrate different data sources, optimizing flood mapping processes and strengthening flood forecasting capabilities, and along with collaboration among government agencies, researchers, and the public is essential to effectively harness crowd-sourced data for flood modeling and map improvements.

There are some limitations due to facts such a Crowd-sourced data may suffer from inconsistencies and inaccuracies due to variations in data collection methods by the user, companies or agencies, data may vary across different geographic areas, leading to data gaps along with gaps and uneven spatial coverage; the frequency of data updates sources may not be consistent, which could affect the timeliness of flood inundation map updates, data might be biased towards specific areas or events, potentially leading to skewed representations of flood occurrences. Managing large volumes of crowd-sourced data requires robust data processing and validation mechanisms, adding complexity to the map generation process along with privacy and ethical considerations, necessitating transparent data usage policies and consent from contributors.

In summary, while crowd-sourced data and smart personal technology offer promising opportunities to enhance flood inundation maps, addressing data quality, privacy, and integration challenges is crucial for maximizing the benefits of these emerging technologies in flood risk management and disaster response.

7. Supplementary Materials:

Table 1. Features

Attribute Names	Number of Attributes	Source (include websites as available)
Latitude	1	Geospatial Database
Longitude	1	Geospatial Database
Storm number (1-10)	1	-
Number of hours to peak inundation	1	NOAA Affiliate(s)National Oceanic and Atmospheric Administration of the United States of America NOAA Office of Water Prediction | Geo-Intelligence Division)
Slope	1	Topographic maps
Topographic (terrain) wetness index (TWI)	1	Chowdhury, M. S. (2023). Modelling hydrological factors from DEM using GIS. MethodsX, 10, 102062
Rainfall Storm Peak =Hour.hr. Hour_0		GPM /IMERG **IMERG V06 Technical Documentation | NASA Global Precipitation Measurement Mission
Rainfall 3 hours prior peak Hour_-3	3	GPM /IMERG **IMERG V06 Technical Documentation | NASA Global Precipitation Measurement Mission
Rainfall 6 hours prior peak Hour_-6	2	GPM /IMERG **IMERG V06 Technical Documentation | NASA Global Precipitation Measurement Mission
Flow accumulation	1	Arcpy package
Crowd-sourced flood observations:		Twitter posts collected by Flood Tags B.V., City of Houston datasets (311)
Flooded (1) or not flooded (0) at this cell during this storm	1	311 datasets, Twitter, HAND, SAR
Number of reports in surrounding grid cells (90 m, 150 m) during this storm	4	Arcpy package
Number of reports in surrounding grid cells (90 m, 150 m) across all storms	2	Arcpy package
FEMA flood claims	2	FEMA flood claims/ FIP Houston
Current HAND FIM predictions at this cell (0/1 - not flooded or flooded)	1	HAND FIM. NOAA Affiliate(s)National Oceanic and Atmospheric Administration of the United States of America NOAA Office of Water Prediction | Geo-Intelligence Division)
SAR satellite data:	2	Google Earth Engine (see table 2)
Flooded (1) or not flooded (0) at this cell during this storm	1	

# of flooded cells in surrounding grid cells (90 m, 150 m) during this storm	2	Arcpy package
# of flooded cells in surrounding grid cells (90 m, 150 m) across all storms	2	Arcpy package

*** The IMERG Integrated Multi satellite retrievals provides a data field that estimates the mixed liquid and solid precipitation from June 2000 – Present. As described by NASA textual from the technical description:" The Integrated Multi-satellite Retrievals for GPM (*IMERG*) is the unified U.S. algorithm that provides the multi-satellite precipitation product for the U.S. GPM team. The precipitation estimates from the various precipitation-relevant satellite passive microwave (PMW) sensors comprising the GPM constellation are computed using the 2017 version of the Goddard Profiling Algorithm (GPROF2017), then gridded, inter calibrated to the GPM Combined Radar Radiometer Analysis product (with GPCP climatological calibration) and combined into half-hourly 0.1°x0.1° fields. These are provided to the Climate Prediction Center (CPC) Morphing-Kalman Filter (CMORPH-KF) quasi-Lagrangian time interpolation procedure and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Cloud Classification System (PERSIANN-CCS) infrared (IR) re-calibration procedure.

Table 2. Flood events and data used for the analysis.

	Storm	**Dates**	**Total Rainfall per storm (inches)**	**Rainfall on Peak Inundation Date (in)**	**Estimated Peak Time from HAND FIM**	**Storm Hrs**	**SAR Satellite Data Availability**	**Dates of Citizen Observations**	**# 311 calls**	**# Tweets**	**Total Citizen Data Points**	**HAND-FIM Dates (Sub-basin USGS 8-12040104)**
1	TS Bill	06/16/15-06/21/2015	2.63	1.04	20150617-15	15		6/17/2015-06/18/2015	11	0	11	6/17/2015-06/18/2015
2	TS Imelda	09/17/2019-09/19/2019	10.85	1.61	20190919-17	17	2019/09/19-T00:18:37 and T00:26:34; 2019/09/24-T00:26:59	09/17/2019-09/20/2019	51	98	149	09/17/2019 - 09/22/2019
3	Un-named	05/07/2019-05/09/2019	6.74	4.08	20190510-04	40	2019/05/10-T00:18:30 and T00:26:26; 2019/05/15-T00:26:51	05/07/2019-05/10/2019	10	48	58	05/07/2019 - 05/12/2019
4	Un-named	07/04/2018-07/06/2018	6.65	6	20180704-19	19	2018/07/07-T00:26:23 and T00:26:48; 2018/07/14-T00:18:28	07/04/2018-07/08/2018	60	8	68	07/04/2018 - 07/08/2018
5	Un-named	12/7/18	3.3	3.76	20181208-06	30		12/7/2018-12/8/2018	2	9	11	12/7/2018-12/9/2018

6	Un-named	6/5/19	2.68	0.85	20190605-21	21		06/05/2019-06/06/2019	2	6	8	06/05/2019 - 06/06/2019
7	TS Beta	09/17/2020-09/25/2020	12.36	7.16	20200923-14	14	2020/09/18-T00:26:40 and T00:27:05; 2020/09/25-T00:18:44	9/22/2020-09/25/2020	16	6	22	9/23/2020-09/25/2020
8	Tax Day	04/16/2016-04/17/2016	3.93	1.35	20160418-12	60		04/16/2016-04/19/2016	103	27	130	04/16/2016 - 04/19/2016
9	Un-named	06-24/2020-06/25/2020	4.44	1.28	20200625-16	40	2020/06/26-T00:26:36; T00:27:01	06/24/2020-06/27/2020	4	10	14	06/24/2020 - 06/27/2020
10	Memorial Day	05/26/2015-05/28/2015	6.96	5.89	20150526-13	6		05/26/2015-05/28/2015	182	83	265	05/26/2015 - 05/30/2015

Table 3. Acronyms

Acronyms	Description	Acronyms	Description
BMP	Best Management Practice	LID	low impact development
BWF	Base Wastewater Flow	LPS	liters per second
CFS	Cubic Feet per Second	MGD	million gallons per day
CMS	Cubic Meters per Second	MLD	million liters per day
CSO	Combined Sewer Overflow	NCDC	National Climatic Data Center
DCIA	Directly Connected Impervious Area.	NOAA	National Oceanic and Atmospheric Administration
DOI	Digital object identifier	NRCS	Natural Resources Conservation Service
EIA	effective impervious area	NWS	National Weather Service
ET	evapotranspiration	NWC	National Water Center
EPA	Environmental Protection Agency	PRMS	Precipitation-Runoff Modeling System
EVAP	daily pan evaporation	RDII	rainfall dependent inflow and infiltration
FEMA	Federal Emergency Management Agency	SCF	Snow Catch Factor
GIS	geographic information system	SCS	Soil Conservation Service
GPM	gallons per minute	SWMM	Storm Water Management Model
GWI	groundwater infiltration	TMAX	maximum daily temperature
HSP	Hydrologic Simulation Program	WE	water equivalent.
JSON	JavaScript Object Notation	GPM	Global Precipitation Measurements Mission

Table 4. Units' description

Features	Units	
Area (sub-catchment)	Acres	SF
Area of Storage	Square feet	SF
Depression Storage	Inches	ml
Depth	Feet	ft
Elevation	Feet	ft
Evaporation	Inches/day	In/d
Flow Rate	cubic feet/sec (cuffs); gallons/min (gpm); gallons/day (mgd)	cfs/ gpm/ mgd
Hydraulic Conductivity	Inches/hour	In/h
Hydraulic Head	feet	ft
Infiltration Rate	Inches/hour	In/h
Sub-Basin	Hydrology Unite Codes	
Watershed	Watershed	
Slope	Percent (%)	
Distance to or from water	meter	mt
Rainfall	Inches/hour	In/h

References

1. Albuquerque, J. P. D., Herfort, B., & Eckle, M. (2016). The tasks of the crowd: A typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. Remote Sensing, 8(10), 859.

2. Guo, H., Liu, Z., Jiang, H., Wang, C., Liu, J., & Liang, D. (2017). Big Earth Data: A new challenge and opportunity for Digital Earth's development. International Journal of Digital Earth, 10(1), 1-12.

3. Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. International Journal of Digital Earth, 3(3), 231-241.

4. Li, Z., Wang, C., Emrich, C. T., & Guo, D. (2018). A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. Cartography and Geographic Information Science, 45(2), 97-110.

5. Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M., & Solomatine, D. P. (2017). Can the assimilation of crowdsourced data in hydrological modeling improve flood prediction? Hydrology and Earth System Sciences, 21(2), 839-861.

6. Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. International Journal of Digital Earth, 3(3), 231–241.

7. Ho, T. K. 1995. Random decision forests. In Proceedings of 3rd International Conference on Document Analysis and Recognition, 278{282. Piscataway, NJ: IEEE.

8. Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., & Moradi, A. (2012). An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environmental Earth Sciences, 67, 251–264.

9.. Liu, Y. Y., Maidment, D. R., Tarboton, D. G., Zheng, X., Yildirim, A., Sazib, N. S., & Wang, S. (2016). A CyberGIS approach to generating high-resolution height above nearest drainage (HAND) raster for national flood mapping.

10. Rahmati, O., Pourghasemi, H. R., & Melesse, A. M. (2016). Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. Catena, 137, 360–372.

11. Tehrany, M. S., Jones, S., & Shabani, F. (2019). Identifying the essential flood conditioning factors for flood-prone area mapping using machine learning techniques. Catena, 175, 174–192.

12. Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., & Kim, S. (2017). Flood inundation modeling: A review of methods, recent advances and uncertainty analysis. Environmental Modelling & Software, 90, 201–216.

13. Youssef, A. M., Pourghasemi, H. R., Pourtaghi, Z. S., & Al-Katheeri, M. M. (2016). Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides, 13, 839–856.

14. HCFCD, 2020a. Buffalo bayou. https://www.hcfcd.org/Find-Your-Watershed/Buff

alo-Bayou. (Accessed 25 July 2020).
15. HCFCD, 2018. Immediate report final. https://www.hcfcd.org/Portals/62/Harvey /immediate-flood-report-final-hurricane-harvey-2017.pdf. (Accessed 25 July 2020).
16. Bukunmi-Omidiran, Titilope, and Balaji Bhaskar Maruthi Sridhar. "Evaluation of spatial and temporal water and soil quality in the Buffalo and Brays Bayou watersheds of Houston, Texas." Remote Sensing Applications: Society and Environment 21 (2021): 100455.
17. Hung K-C, Kalantari M, Rajabifard A (2016) Methods for assessing the credibility of volunteered geographic information in flood response: a case study in Brisbane, Australia. Appl Geogr 68:37–47
18. Safaei Moghadam, A., D. Tarboton, B. Minsker (2022) Estimating the likelihood of roadway pluvial flood based on crowdsourced traffic data and depression-based DEM analysis. Natural Hazards and Earth System Sciences, https://doi.org/10.5194/nhess-23-1-2023, 23: 1-19.

Chapter 5

Quantifying the Sources of Uncertainty in OWP HAND-FIM Predictions

Umanda Abeysinghe[1], Anne Holt[2], Karina Larco[3], and Meraj Sohrabi[4]

[1]University of Missouri; *umandajb@gmail.com*, *mjabnf@missouri.edu*

[2]San Diego State University; *holtannie@gmail.com*

[3]Bringham Young University; *karynina3@gmail.com*, *karinaml@byu.edu*

[4]University of Alabama; *emsohrabi1@crimson.ua.edu*, *merraj.sohrabi1369@gmail.com*

Academic Advisors: Noel Aloysius, *University of Missouri*; Hilary McMillan, *San Diego State University*; Jim Nelson, *Bringham Young University;* Hamed Moftakhari, *University of Alabama*

Summer Institute Theme Advisors: Sagy Cohen, University of Alabama, *sagy.cohen@ua.edu*, Ehab Meselhe, Tulane University, *emeselhe@tulane.edu*

Abstract: Every year, millions of people are devastated by floods, which are getting increasingly severe worldwide. Flood Inundation Mapping (FIM) frameworks have become crucial tools for mitigating flood risks and improving disaster awareness and forecasting, especially for decision-makers. In the United States (US), the NOAA Office of Water Prediction (OWP) is leading the development of a national-scale FIM forecasting framework. In its current operational state, the framework is based on the Height Above Nearest Drainage (HAND) approach, which couples the OWP National Water Model with the HAND approach through Synthetic Rating Curves (SRCs). However, recent studies have highlighted several shortcomings in the OWP HAND-FIM method, notably in regards to the technique's efficacy in coastal and low-relief settings. This study aims to evaluate OWP HAND-FIM performance and isolate uncertainties in SRCs and NWM-predicted discharge on flood extent. The Amite River Basin in the US was chosen as a location for a thorough case study due to its significant susceptibility to fluvial floods and the availability of high-fidelity HEC-RAS model for past flood episodes. We compared SRCs at USGS gauge locations with HEC-RAS rating curves and assessed the impact of different discharges on OWP HAND-FIM predictions. The study found that OWP HAND-FIM demonstrated a performance agreement of 44% when compared to observed data, while the ARB HEC-RAS model showed a significantly higher agreement of 93%. In order to introduce uncertainties into OWP HAND-FIM we found disparities between SRCs and USGS curves. We generated new FIM with different streamflows, and the flood extent differences based on the HAND-FIM method emphasized the complexity of determining the most suitable rating curve. Addressing these uncertainties is essential for improving large-scale FIM tools and enhancing flood forecasting for decision-makers. Addressing these uncertainties is essential for improving large-scale FIM tools and enhancing flood forecasting for decision-makers.

1. Motivation

Floods are one of the most frequent natural disasters in the world, affecting over 250 million people each year. According to the Intergovernmental Panel on Climate Change's 2021 report [1], floods are becoming more severe in many areas due to increasing sea levels, stronger storms, and heavier rainfall. Given the increasing frequency and severity of floods, their impacts on human lives, infrastructure, and the economy can be significant. In response to these devastating flood impacts, accurate and timely Flood Inundation Mapping (FIM) models are essential as they provide spatially explicit representation of the extent which can serve as powerful tools for decision-makers. To efficiently forecast flood inundation at the continental scale, the Office of Water Prediction (OWP) of the United States's National Oceanic and Atmospheric Administration (NOAA) utilizes Height Above Nearest Drainage (HAND) in combination with Synthetic Rating Curves (SRCs) to generate flood inundation maps based on discharge predicted by the National Water Model (NWM). The NWM integrates real-time and forecast data to simulate water balance and streamflow supporting decision-making related to water management, flood mitigation, and drought preparedness at various scales. The OWP operational FIM prediction framework, referred to as OWP HAND-FIM, have been evaluated and compared with flood inundations from observations as well as from hydraulic models that leverage detailed channel geometry such as the Hydrologic Engineering Center's River Analysis System (HEC-RAS) developed by the United States Army Corps of Engineers [1, 2, 3].

Recent studies have found that OWP HAND-FIM can have lower performance in coastal areas and low-relief regions [2, 3], and has limitations related to both underlying data and methodology. For example, factors that can contribute to OWP HAND-FIM performance and uncertainty include the accuracy of Digital Elevation Models (DEMs), errors in modeled streamflow, characterization of anthropogenic influences such as bridges or drainage systems, or the accuracy of SRCs and underlying features influencing such stage-discharge relationships such as roughness and channel geometry [4]. Isolating these sources of uncertainty and their relative influence on OWP HAND-FIM performance would contribute to the improvement of large-scale FIM tools and support more accurate flood forecasting for decision-makers.

2. Objectives and Scope

The objectives of this study are to evaluate OWP HAND-FIM performance and to isolate and characterize the impact of uncertainties in SRCs and estimated streamflow on forecasted flood extent. We conducted a case study in the Amite River Basin (ARB), located in Southeastern Louisiana and Southwestern Mississippi in the United States of America (USA; **Figure 1**). The ARB has a drainage area of 4,878 km^2 and is susceptible to flooding due to urbanization and development, low-relief topography, coastal proximity, and high annual precipitation.

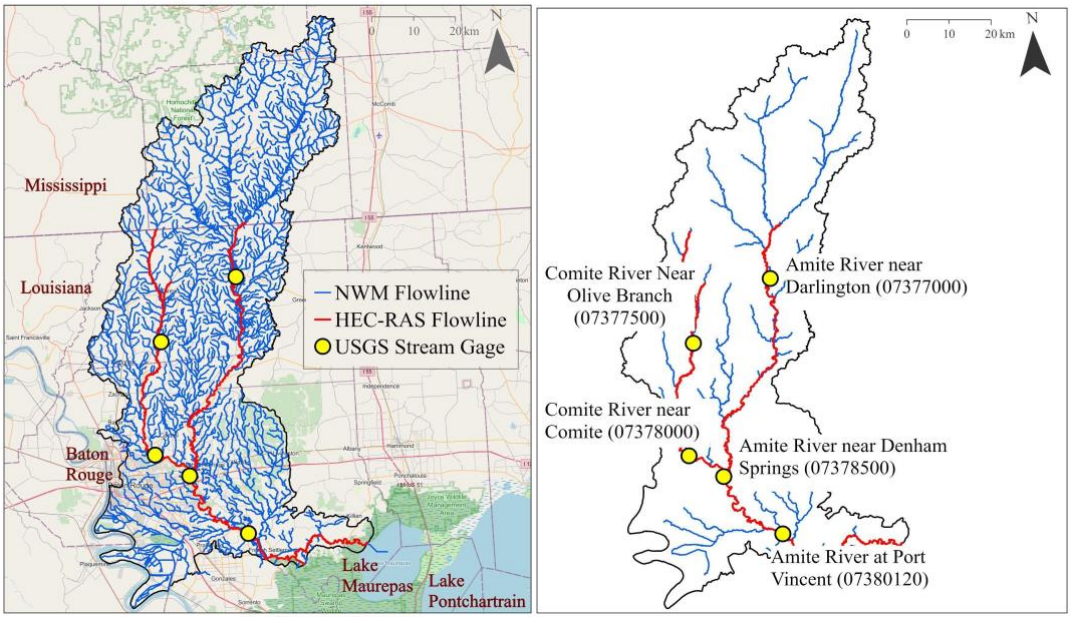


Figure 1. *Overview of Amite River Basin and USGS gauge locations selected for rating curve analysis.*

For our evaluations, we focused on flood events that occurred in March and August of 2016 (**Table 1**). The August 2016 event was a historical flood caused by a tropical storm that resulted in a record 48-hour total rainfall of 797 mm (31.39 inches) at one station. This event resulted in thirteen deaths, and an estimated $4.4 billion dollars of structural damage in the ARB [5, 6, 7].

Table 1. *Historic flood events used in this study*

Flood event	Start date	End date
Event 1	March 10, 2016 (10:00)[1]	March 12, 2016 (21:00)[1]
Event 2	August 10, 2016 (10:00)[2]	August 21, 2016 (19:00)[2]

[1] The North American Central Time Zone (CST), [2] The American Central Daylight Time (CDT)

3. Previous Studies

Previous research has examined uncertainty sources in OWP HAND-FIM predictions and have explored various approaches to enhance FIM accuracy. Aristizabal et al. (2023) demonstrated that NWM flow predictions significantly affect HAND-FIM accuracy [8]. Li et al. (2022), also emphasized accounting for uncertainty in real-time flood inundation mapping using the HAND model as a means to improve the reliability of flood extent predictions [9]. Evaluations of SRCs and HAND-FIM coastal regions have emphasized the need for improvements as HAND-FIM accuracy may be compromised due to the complex interactions between river flows, tides, and coastal processes. For example, while Zhang et al. (2019), found that SRCs exhibited good agreement with field-derived rating curves during normal flow conditions, discrepancies emerged under extreme flow in coastal areas [10]. Additionally, SRCs have a tendency to overpredict stage in coastal areas and underpredict in mountainous regions [11]. By leveraging insights from these studies, our research aims to continue isolating the sources of uncertainty in OWP HAND-FIM predictions, particularly focusing on the flow predictions and rating curve performance under extreme flow conditions.

4. Methodology

We applied the OWP HAND-FIM methodology to predict flood depths and extent in the study basin for the historical flood events based on the NWM discharge predictions. By comparing forecasted extents with observed data as well as flood extents produced by a detailed HEC-RAS model, our study evaluated OWP HAND-FIM performance and suggested sources of FIM uncertainty. The analysis used the ARB 1D/2D HEC-RAS model version 5.0.6 including detailed data for cross-sections, bridges, culverts, and lateral structures. The model assumed sub-critical flow, Manning's n values ranging from 0.035 to 0.12, and elevation data was from the 2018 LADOTD LiDAR. The analysis also utilized 449 High-Water Marks (HWM) survey data for the August 2016 event. Out of the 449 HWMs, 324 were designed as high confidence in our watershed based on their agreement with adjacent HWMs and are used for our evaluations [15]. Our study isolated impacts of discharge and SRC errors on forecasted flood extent using predicted, observed, and HEC-RAS simulated data. Differences in resulting flood extent from the OWP HAND-FIM methodology were evaluated under this study.

4.1. Comparing Rating Curves

To evaluate the performance of SRCs utilized in HAND-FIM methodology, we compared SRCs at three USGS gauge locations with rating curves from USGS and the ARB HEC-RAS model. The locations and names of the stations are shown in **Figure 1**. A rating curve relates water level (stage) to discharge (flow rate) in rivers, and is vital for flood forecasting and. The HAND-FIM approach links observed water levels to actual water flow at a specific location. Channel geometry, roughness, and obstructions are among the factors that influence this relationship.

We associated each gauge with the closest stream segment in the flowlines used in HAND-FIM as well as the closest ARB HEC-RAS cross sections in order to align the stage-discharge data sources for our comparisons. To generate ARB HEC-RAS rating curves, we extracted water level and discharge for the modeled 2016 events. We also downloaded USGS stage and discharge data utilizing the 'dataretrieval' Python package [12] to extract data from the National Water Information System (NWIS) for the 2016 events based on site IDs. As the USGS stage-discharge relationships exhibited hysteresis, we selected the lower portion of the hysteresis loops in order to finalize the USGS rating curves.

In order to achieve an estimation of the potential impact of these differences on the OWP HAND-FIM, our analysis focused on sub-catchments with available rating curves from the USGS. By isolating discharge values, we can accurately assess the extent to which variations in water level (stage) determined by rating curves influence the final outcome of the HAND FIM calculation. To accomplish this, we utilized the discharge value for the March 2016 event from USGS as the reference water level, and then determined the associated discharge value based on HEC-RAS and HAND rating curves. Subsequently, we generated OWP HAND-FIM for the three different stages.

To analyze the differences in stage heights derived from these rating curves, we utilized the Bland-Altman method. It is a graphical method used to compare two measurements or methods. The plot was introduced by J. Martin Bland and Douglas G. Altman in 1983 as an alternative to simple correlation or regression analysis when comparing two methods of measurement that produce continuous data. The Bland-Altman method plots the differences between the two measurements against their averages. If the differences are randomly scattered around zero, it indicates good agreement between the methods. If the differences are consistently above or below zero, there is a systematic bias between the methods. If the differences increase or decrease as the averages increase, it suggests that the agreement may not be constant across the measurement range. The plot provides valuable insights into the level of agreement, any potential bias, and the overall performance of the two methods, making it a useful tool for assessing measurement agreement in scientific research.

4.2. Flood Inundation Mapping

Flood inundation maps for the 2016 events were generated in this study applying the OWP HAND-FIM methodology. To ensure consistent and reproducible results, we utilized tools published on the NOAA-OWP GitHub repository and followed the recommended Docker workflow for executing the inundation functions [12]. The OWP HAND-FIM pipeline generates HAND rasters representing the height of each grid above the nearest stream grid based on a 10m DEM. To generate inundation maps based on these HAND rasters, discharge is converted to stage height using the HAND-derived SRCs and flooding occurs when the stage height is greater than the HAND grid value [11]. For our study, HAND data including the HAND rasters, SRCs, and flowlines were downloaded from the OWP AWS bucket and we executed the subsequent inundation workflows.

To first evaluate HAND-FIM performance in the ARB for the events, we generated a flood inundation map applying HAND-FIM for the HUC8 watershed using predicted streamflow data from the NWM Retrospective Dataset version 2.1. To retrieve discharge data for the 2016 event (see **Table 1**), we modified a Python code based on the work of Abdelkader, M. and J.H. Bravo Mendez (2023) [14] to download hourly predicted discharge for all reaches within the ARB based on FeatreIDs. We then retained the maximum discharge for each FeatureID for our subsequent flood inundation mapping. In order to isolate the impact of NWM discharge predictions on flood extent, we also generated flood inundation maps for August 2016 applying OWP HAND-FIM methodology at three USGS gauge locations across the ARB using observed maximum discharge from the USGS and maximum discharge from a calibrated ARB HEC-RAS model extracted at the cross section closest to the gauge location.

4.3. Evaluating Differences in Flood Inundation Maps

Results from the OWP HAND-FIM methodology were compared with maps generated using a calibrated ARB HEC-RAS model and HWM survey data for the August 2016 event. The

comparison aimed to assess and quantify differences between predicted and observed flood extents, as well as to evaluate the impact of streamflow and SRC uncertainties on the predicted extent. The March event was not used for this portion of the analysis as HWM data was not available for validation.

In order to compare HEC-RAS flood inundation results with OWP HAND-FIM simulations for the August 2016 event, flood depth from HEC-RAS were aggregated to a 10 by 10-meter resolution, matching the HAND-FIM resolution. Flood inundation maps were subsequently converted to binary flood/not flooded pixel values based on a water depth threshold of 10 cm. To validate the accuracy of the flood extent predictions, we also matched the high confidence HWM locations with flood raster pixels in order to calculate the percentage of agreement. Three metrics were subsequently used to compare flood extents derived from HAND-FIM and HEC-RAS following Aristizabal et al. (2023). The metrics were used: Critical Success Index (CSI), Probability of Detection (POD) and False Alarm Ratio (FAR) [8]. By quantifying flood extent differences, conducting visual comparisons, the study aims to understand the differences between flood inundation maps generated using different datasets and methodologies.

5. Results

5.1. OWP HAND-FIM Performance for August 2016

To generally evaluate the performance of OWP HAND-FIM across the ARB, we first compared the predicted flood inundation extent for the August 2016 flood event with the detailed flood extent produced by the ARB HEC-RAS model and the observed HWMs. We found that OWP HAND-FIM captured 44% of the high confidence HWMs, had a CSI of 0.46, and a POD of 0.57 for the August 2016 event compared to the HEC-RAS flood extent. Alternatively, HEC-RAS captured 93% of the high confidence HWMs. The results show that in the upper middle portion of the ARB, HAND-FIM has more agreement with HEC-RAS flood extent compared to the low-relief, more urbanized lower watershed that likely experienced large amounts of pluvial flooding in the August 2016 event. Lower HAND-FIM performance in low relief, urban, and coastal catchment areas agrees with previous studies, and is likely because OWP HAND-FIM captures fluvial flooding but not backwater flooding or local flooding from intense, short duration rainfall [1, 2, 3, 16, 17]. Differences in flood extent in the upper portion of the watershed and the upper tributaries occurred because the HEC-RAS model did not resolve the upper watershed channels explicitly; rather their runoff was captured through the hydrologic model and fed into the downstream tributaries. Therefore, our comparison metrics do not fully capture the differences in modeled flow neworks.

5.2. Impact of discharge on OWP HAND-FIM predictions

By generating flood extents using observed USGS and ARB HEC-RAS maximum discharge values for the August 2016 event, we removed uncertainties associated with NWM streamflow predictions and assessed the resulting impact on flood forecasts. We observed that the maximum hourly NWM streamflow closely matched the HEC-RAS and USGS maximum discharges for

the upstream gauge locations (**Figure 2a** and **2c**). However, in **Figure 2b** (downstream gauge), the NWM maximum streamflow was approximately 50% higher. This may be due to a failure at the USGS gauge, judging by the linear falling limb of the flood wave and the low values recorded compared to the upstream gauge. The event hydrographs in **Figure 2a** shows that the USGS flood peak was earlier compared to the predicted NWM discharge and HEC-RAS hydrographs, and **Figure 2c** shows a case where the NWM event hydrograph recedes earlier compared to USGS and HEC-RAS.

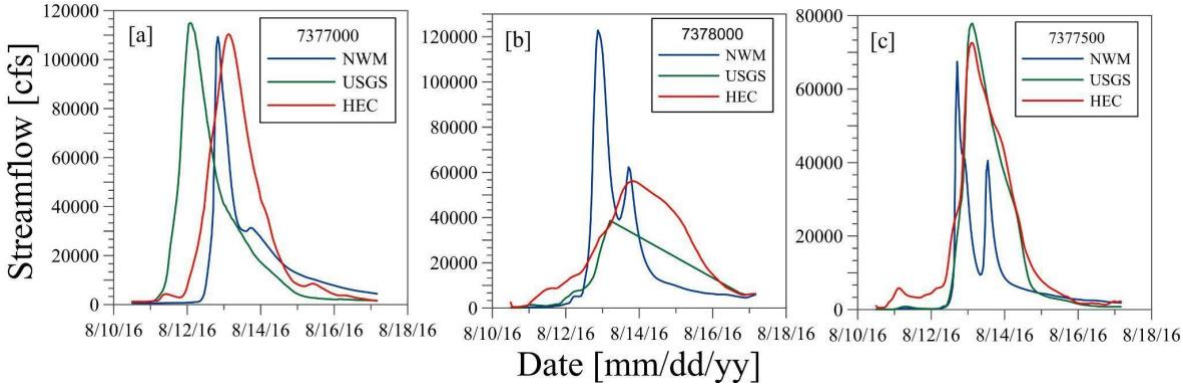


Figure 2. *Hydrographs for the August 2016 event for three gauge locations. (a) Maximum discharge in* Amite River Near Darlington *(b) Maximum discharge in Comite River Near Comite (c) Maximum discharge in Comite River Near Olive Branch*

Despite using different streamflow data sources to run OWP HAND-FIM, the maps in **Figure 3** show that the flood extent results are the same at the upstream gauges due to similar maximum discharges (**Figure 3a, 3c**). However, **Figure 3b** shows differences in flood extent in accordance with differences in maximum discharge values, demonstrating that streamflow predictions can cause uncertainties in flood extent. At the site in **Figure 2b**, the differences between NWM and HEC-RAS could be related to the presence of a tributary river above the reach as well as a bridge closeby. **Figure 3** also demonstrates that at these locations, OWP HAND-FIM methodology is generally underpredicting flood extents in comparison to 'true' flood extents simulated in the HEC-RAS model likely due to pluvial flooding not being represented or remaining uncertainties in SRC performance.

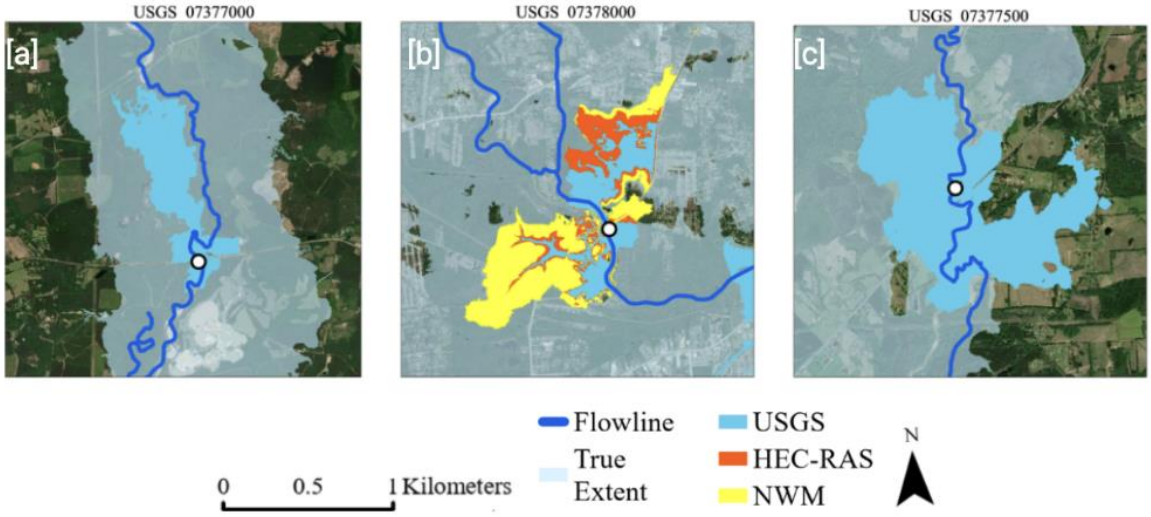


Figure 3. *Flood inundation extents generated using OWP HAND-FIM methodology with different input discharges. (a) FIM for* Amite River Near Darlington *(b) FIM for Comite River Near Comite (c) FIM forComite River Near Olive Branch. Note that (a) and (c) had matching flood extents for*

5.3. Comparing rating curves

The disparities between the SRC, the USGS and HEC-RAS rating curves for select locations are shown in Figure 4. The SRC and USGS curves do not demonstrate consistent agreement with each other. Notably, there are instances where the SRC underpredicts and other instances where it overpredicts. We posit that these discrepancies could introduce substantial uncertainty into the final HAND FIM.

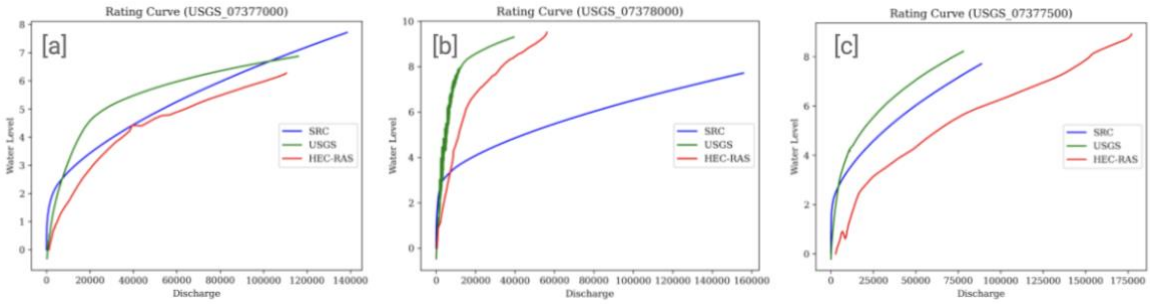


Figure 4. *Rating curves from HEC-RAS, USGS and HAND Synthetic Rating (a)* Amite River Near Darlington *(b) Comite River Near Comite (c) Comite River Near Olive Branch*

In the first two plots (a and b) presented in **Figure 5**, a trend is observed where the differences between the two curves decrease as the average of the measurements (here is water level) increases. This behavior indicates that the agreement between the curves might not be constant across the entire range of measurements. Conversely, in the plot c, the differences are observed to be randomly distributed around the mean difference line. This particular pattern signifies a favorable level of agreement between the two curves. These graphical representations, through the Bland-Altman plot, provide valuable insights into the nature of agreement, potential biases, and overall performance of the two measurement methods, contributing to a comprehensive assessment of their concordance in the context of this research.

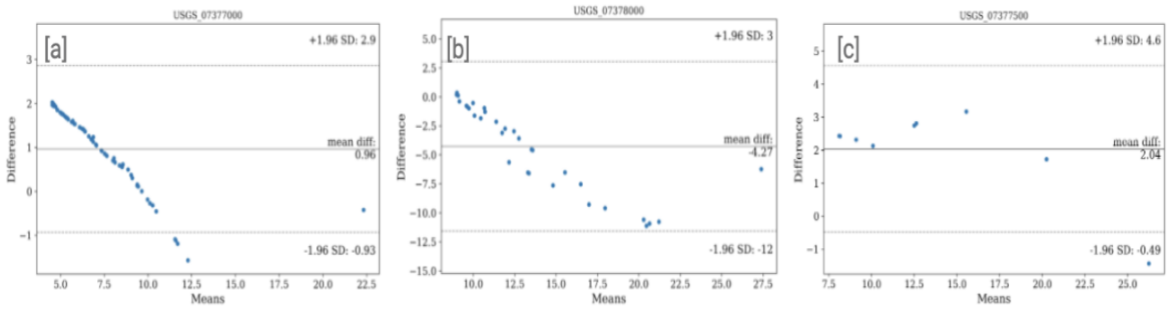


Figure 5. *Bland Altman plots for examining differences between USGS and HAND rating curve in three different locations: (a)* Amite River Near Darlington *(b) Comite River Near Comite (c) Comite River Near Olive Branch. Each plot shows the average Water Level (WL) on the x-axis and the difference between the two WL on the y-axis, with all units in feet*

Figure 6 illustrates the differences in flood extent for the same events based on the OWP HAND-FIM method. Through an examination of the flood extent, it becomes apparent that the utilization of different rating curves significantly impacts the flood mapping outcomes.

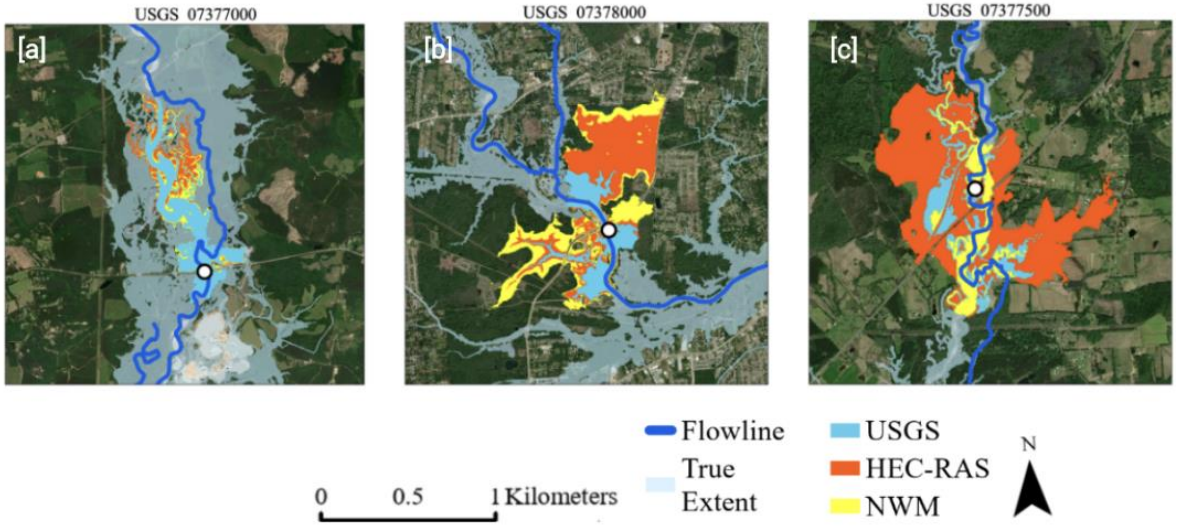


Figure 6. *Flood extent differences using HEC-RAS, HAND and USGS rating curves for March 2016.(a) FIM for* Amite River Near Darlington *(b) FIM for Comite River Near Comite (c) FIM forComite River Near Olive Branch*

6. Conclusions

This study evaluates the OWP HAND-FIM methodology's performance in predicting flood extents and examines the impact of uncertainties in SRCs and modeled discharge on flood predictions. By conducting a comprehensive analysis that involved comparing SRCs, USGS, and HEC-RAS rating curves, as well as generating flood inundation maps, our study evaluated the performance of HAND-FIM in predicting flood extents. The results revealed that OWP HAND-FIM exhibited a moderate agreement of 44% with the observed data. On the other hand, the ARB HEC-RAS model demonstrated a higher agreement of 93% when compared to the observed data, indicating its superior accuracy and reliability in simulating flood extents. To summarize, we found limited differences in FIM predictions due to input discharge since NWM and HEC-RAS streamflow predictions in our case studies exhibited good agreement with observations. However, OWO HAND-FIM is highly sensitive to the rating curves. The discrepancies observed between SRCs and USGS rating curves underscore the existence of potential uncertainties in HAND-FIM predictions. This highlights the necessity of addressing these uncertainties to ensure accurate and reliable flood forecasting.

Supplementary Materials:

https://github.com/annieholt/OWP_HAND_FIM_Eval_ARB

References

1. Zheng, X., Tarboton, D. G., Maidment, D. R., Liu, Y. Y., & Passalacqua, P. (2018). River Channel Geometry and Rating Curve Estimation Using Height above the Nearest

Drainage. JAWRA Journal of the American Water Resources Association, 54(4), 785–806. https://doi.org/10.1111/1752-1688.12661

2. Garousi-Nejad, Irene, David G. Tarboton, Mahyar Aboutalebi, Alfonso F. Torres-Rua. "Terrain analysis enhancements to the height above nearest drainage flood inundation mapping method." *Water Resources Research* 55, no. 10 (2019): 7983-8009. https://doi.org/10.1029/ 2019WR024837.
3. Johnson, J. Michael, Dinuke Munasinghe, Damilola Eyelade, and Sagy Cohen. "An integrated evaluation of the national water model (NWM)–Height above nearest drainage (HAND) flood mapping methodology." *Natural Hazards and Earth System Sciences* 19, no. 11 (2019): 2405-2420.
4. Aristizabal, Fernando, Fernando Salas, Gregory Petrochenkov, Trevor Grout, Brian Avant, Bradford Bates, Ryan Spies, Nick Chadwick, Zachary Wills, and Jasmeet Judge. "Extending Height Above Nearest Drainage to Model Multiple Fluvial Sources in Flood Inundation Mapping Applications for the US National Water Model." *Water Resources Research* (2023): e2022WR032039.
5. Jacobsen, Bob. "August 2016 Flood Preliminary Report: Amite River Basin." *Report prepared for the Amite River Basin Drainage and Water Conservation District. Available online: http://www. amitebasin. org/2016Flood/August* 202016 (2017).
6. Brown, Vincent M., Barry D. Keim, William D. Kappel, Douglas M. Hultstrand, Ashton G. Peyrefitte, Alan W. Black, Kristi M. Steinhilber, and Geoffrey A. Muhlestein. "How rare was the August 2016 South-Central Louisiana heavy rainfall event?." *Journal of Hydrometeorology* 21, no. 4 (2020): 773-790.
7. Cowles, Alexandre GH. *Effects of Historical Land-use Change on Surface Runoff and Flooding in the Amite River Basin, Louisiana, USA Using Coupled 1D/2D HEC-RAS–HEC-HMS Hydrological Modeling*. Louisiana State University and Agricultural & Mechanical College, 2021.
8. Aristizabal, Fernando, Taher Chegini, Gregory Petrochenkov, Fernando Renzo Salas, and Jasmeet Judge. "Effects of High-Quality Elevation Data and Explanatory Variables on the Accuracy of Flood Inundation Mapping via Height Above Nearest Drainage." *EGUsphere* 2023 (2023): 1-41.
9. Li, Zhouyayan, Jerry Mount, and Ibrahim Demir. "Accounting for uncertainty in real-time flood inundation mapping using HAND model: Iowa case study." *Natural Hazards* 112, no. 1 (2022): 977-1004.
10. Zhang, Guotao, Peng Cui, Yanzhou Yin, Dingzhu Liu, Wen Jin, Hao Wang, Yan Yan, Bazai Nazir Ahmed, and Jiao Wang. "Real-time monitoring and estimation of the discharge of flash floods in a steep mountain catchment." *Hydrological processes* 33, no. 25 (2019): 3195-3212.
11. Ghanghas, Ankit, Sayan Dey, and Venkatesh Merwade. "Evaluating the reliability of synthetic rating curves for continental scale flood mapping." *Journal of Hydrology* 606 (2022): 127470.
12. Hodson, T.O., Hariharan, J.A., Black, S., and Horsburgh, J.S. "dataretrieval (Python): a Python package for discovering and retrieving water data available from U.S. federal

hydrologic web services". *U.S. Geological Survey software release,* (2023): https://doi.org/10.5066/P94I5TX3.

13. NOAA Office of Water Prediction. "Inundation Mapping". Version 4.3.11.7 (2023): https://github.com/NOAA-OWP/inundation-mapping.
14. Abdelkader, M., J. H. Bravo Mendez. "NWM version 2.1 model output data retrieval." *HydroShare*, (2023): https://doi.org/10.4211/hs.c4c9f0950c7a42d298ca25e4f6ba5542.
15. Dewberry Engineers Inc. "Amite River Basin Numerical Model Project Report. Baton Rouge, LA". (2019).
16. Wing, Oliver EJ, Andrew M. Smith, Michael L. Marston, Jeremy R. Porter, Mike F. Amodeo, Christopher C. Sampson, and Paul D. Bates. "Simulating historical flood events at the continental scale: observational validation of a large-scale hydrodynamic model." *Natural Hazards and Earth System Sciences* 21, no. 2 (2021): 559-575.
17. Godbout, L., Zheng, J.Y., Dey, S., Eyelade, D., Maidment, D., Passalacqua, P., 2019. Error assessment for height above the nearest drainage inundation mapping. JAWRA J. Am. Water Resour. Assoc. 55 (4), 952–963. https://doi.org/10.1111/17521688.12783.

Chapter 6

Improving the Fidelity and Performance of OWP HAND-FIM Using a Surrogate Model (SM)

Berina Mina Kilicarslan[1], Qianqiu Longyang[2], and Victor Obi[3]
[1]Stevens Institute of Technology; bkilicar@stevens.edu
[2]Arizona State University; qlongyan@asu.edu
[3]Kent State University; vobi@kent.edu

Academic Advisors: Marouane Temimi, *Stevens Institute of Technology*; Ruijie Zeng, *Arizona State University*; Kuldeep Singh, *Kent State University*

Summer Institute Team Advisors: Ehab Meselhe (Tulane University) emeselhe@tulane.edu, Sagy Cohen (University of Alabama) sagy.cohen@ua.edu

Abstract:

This study focuses on enhancing flood inundation accuracy and mitigating limitations in the Height Above the Nearest Drainage-Flood Inundation Mapping (HAND-FIM) method through the implementation of surrogate modeling (SM) The SM is developed using machine learning techniques that replicate the relevant hydrodynamic characteristics from a high-fidelity Hydrologic Engineering Center-River Analysis System (HEC-RAS) model, integrating it with the low-fidelity HAND-FIM. The HAND-FIM is generated for a historic flood event in August 2016 using the National Water Model (NWM) streamflow and the Office of Water Predictions (OWP) HAND-FIM Synthetic Rating Curves (SRC). The flood extent predicted by HEC-RAS serves as the target for the SM. Other inputs including Digital Elevation Model (DEM), slope, aspect, and landcover are used to train the SM. For the SM, the Critical Success Index (CSI) is utilized as a loss function for grid search, optimizing parameters for the Random Forest (RF) classifier model. The model is trained on 10 sub-watersheds and tested on 4 sub-watersheds, encompassing diverse types of land use. Notably, the testing phase excludes the HEC-RAS model outputs. The results indicate that when comparing high-fidelity input data (HEC-RAS FIM) to low-fidelity input data (HAND-FIM), the latter tends to exhibit more false alarms in both the training and test sets. It is observed that the higher-resolution DEM does not directly improve HAND-FIM, but when it is included as input in the SM, it enhances prediction accuracy. Results show the SM holds potential for model transferability, leading to reduced false alarms, and achieving commendable Probability of Detection (POD) and CSI values for both training and test sets. Overall, integrating SM presents a promising approach to improve flood prediction accuracy and address HAND-FIM limitations.

Keywords: flood inundation mapping; surrogate model; low fidelity; high fidelity; machine learning

1. Motivation

Flood events pose significant risks to communities and infrastructure, with over 1,100 flood-related fatalities recorded in the US over the past decade [1]. Floods have been the major cause of property and crop damage costs in 2017 and 2019 [2,3], emphasizing the importance of effective flood management strategies. Accurate flood inundation mapping is crucial for

identifying at-risk areas, enabling proactive planning, emergency response coordination, and mitigation measures. The Office of Weather Prediction (OWP) utilizes the WRF-Hydro model as the NWM to provide streamflow forecasts for over 2.7 million river reaches across the contiguous United States (CONUS). To translate the NWM streamflow prediction to the river stage and calculate inundation extents, the OWP uses the HAND method with synthetic rating curves (SRC) [4]. Despite HAND's advantages in simplicity and computational efficiency, it has limitations, including reliance on assumptions and a lack of representation of flood dynamics, which may impact its accuracy [1-2].

2. Objectives and Scope

This study aims to enhance the fidelity of a simple flood inundation method by employing a surrogate model (SM) that bridges the gap between this simplified model and a more complex flood inundation method that utilizes a hydrodynamic solver. In this project, HEC-RAS predictions are used as a high-fidelity FIM whereas the OWP HAND-FIM is considered as the low-fidelity FIM predictions. Further, the regionalization of theSM is explored over the watershed with different characteristics and data/model scarcity.

3. Previous Studies

Flood inundation mapping can be done using various approaches, including hydrodynamic modeling, like the widely used HEC-RAS model. This method simulates flow dynamics in one-dimensional (1D) or two-dimensional (2D) settings, considering factors like topography, river geometry, and flow rates to estimate flood extents. Alternatively, simpler and less computationally demanding methods like the HAND method are used for large-scale mapping [4]. Another approach to reducing computational costs is using data-driven SMs [11-12]. These models learn from complex relationships between inputs and outputs, often using time series data to address system behavior hysteresis. Input features include topographic data [7]; and remote sensing data [8] for accurate predictions, especially in coastal urban and riverine flooding regions. While recent studies [18] suggest that higher resolution DEM may not directly enhance HAND-FIM, its influence on a SM remains unclear. Previous studies have employed data-driven techniques such as random forest algorithms [7]; artificial neural networks [9]; support vector machines [8]; and combinations of different models [10]. Event-based datasets, including historical and synthetic events, are commonly used for training, and testing SMs. Another method involves constructing a computationally efficient, low-fidelity model and then enhancing its results with a data-driven surrogate model. The SM improves the accuracy of the low-fidelity model outputs to match the high-fidelity model outputs, deciphering complex relationships between the two.

4. Study Area

The Amite River Basin (ARB), which spans an area of about 4, km^2(~1,596 mi^2), is a watershed situated in the southeastern region of the United States (Figure 1). It encompasses several important counties and parishes in Mississippi and Louisiana such as East Baton Rouge, Ascension, and others as well as several cities Darlington, Denham Springs, and Olive Branch. It originates as the East Fork Amite River and the West Fork Amite River in Mississippi, with

their headwaters merging around 2 km downstream of the Louisiana - Mississippi state line. Continuing its path, the Amite River converges with the Comite River just before the city of Denham Springs, spanning approximately 80 km until it ultimately reaches its destination, Lake Maurepas (Figure 1) In this study we concentrated on the drainage area covering 08070202-HUC8 region. ARB plays a crucial role in supporting various ecological conditions, water supply for human consumption, irrigation, navigation, land use, and recreational activities.

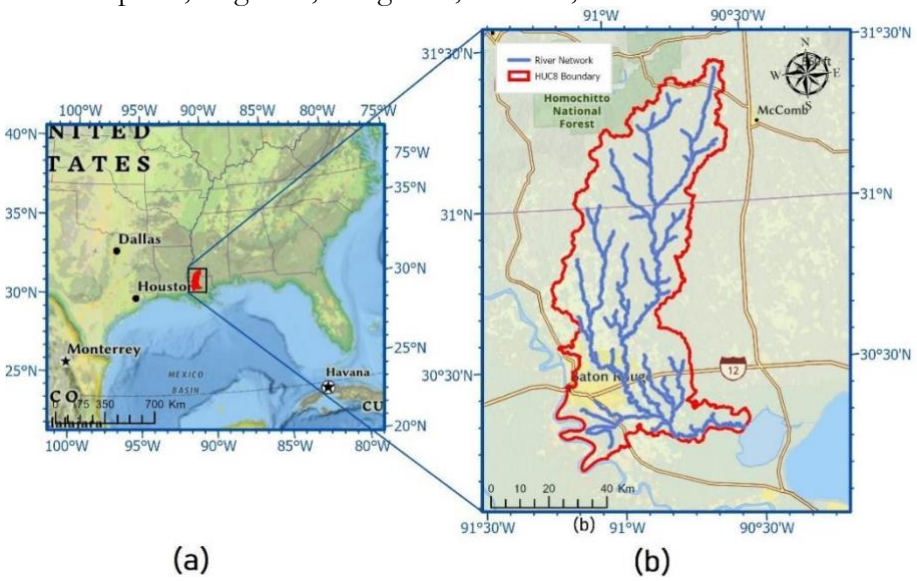


Figure 1. ***(a)*** *Study area: Louisiana, United States and* ***(b)*** *Amite River Network over HUC8 (08070202)*

5. Methodology

5.1. HAND Method

The HAND method is a topographic model most commonly used for generating flood inundation maps over large domains. It is calculated from a DEM by normalizing the difference in elevation between each grid cell and its nearest stream grid cells. This relative elevation value is called the HAND value [13]. For calculating FIM, grid cells whose HAND value is lower than the stage height for a specific flow condition are classified as inundated [14]. The difference between stage and HAND values in each cell can also be used to calculate floodwater depth. Operationally within the OWP, HAND-FIM layers are generated by utilizing the National Water Model (NWM) streamflow predictions. Stage heights for each stream reach are calculated by using a reach-specific discharge-stage synthetic rating curve [14].

5.2 HEC-RAS Model

HEC-RAS, developed by the U.S. Army Corps of Engineers [17]., is a widely used hydraulic modeling software for analyzing river systems. It uses open-channel flow principles to simulate flow dynamics within channels, and over floodplains. It applies the Saint-Venant equations to simulate flow, accounting for variables such as water depth, velocity, and so on which are solved using numerical solvers.

5.3 Data Extraction and Preprocessing

Data for this study originates from Dewberry Engineers Inc's modeling tools, commissioned by the Louisiana Department of Transportation and Development (LA DOTD). The Amite River Basin (ARB) dataset consists of 34 flood events, comprising 4 historical events and 30 designed/synthetic events in Denham Springs, Darlington, and Olive Branch cities. Precipitation depths range from 8 to 26 inches. The August 2016 historic flood event is the primary focus of this study, where dynamic characteristics are replicated from the high-fidelity HEC-RAS model. Our emphasis lies in the upstream region of HEC-RAS, excluding the Amite River diversion canal due to OWP HAND-FIM inability to represent channel diversions (*See Supplementary Materi*al). The ARB's geometry encompasses over 800 1D cross-sections.

5.4. Surrogate Model

A hybrid model approach is proposed to enhance the accuracy of the OWP HAND-FIM flood inundation mapping to achieve high-fidelity results. The HAND-FIM method is treated as a low-fidelity model and improved by mimicking the high-fidelity behavior exhibited by the HEC-RAS 2D model using a machine learning SM. 14 sub-watersheds that cover the more diverse types of land use including Developed, Cultivated, Forest, Wetlands, etc. HEC-RAS and HAND-FIMs are generated for a historical flood event that occurred in August 2016. These sub-watersheds represent a range of characteristics within the basin. The SM relies on the Random Forest (RF) algorithm, known for its robustness in the hydrology and water resources fields [16]. The workflow is outlined in Figure 2. The model is trained on data from 10 sub-watersheds and tested on 4 sub-watersheds to assess its transferability to other regions in which high-fidelity predictions are not available. It is imperative to emphasize that, during the testing phase, the HEC-RAS model outputs are excluded from the evaluation process once the model is trained. The input flood extent is generated with a HAND-FIM raster, which is re-classified as -1 for dry conditions, 1 for flooding, and 0 for grid cells outside the study area. Additional input variables include land cover classes and imperviousness from the NLCD 2016 dataset, 10-m 3DEP DEM, terrain slope and aspect (calculated from the DEM), and a river network mask. In accordance with [11], the flood extent simulated by the HEC-RAS model is resampled at 10 m resolution (to match the HAND-FIM output) and is defined by water depth exceeding 3 cm.

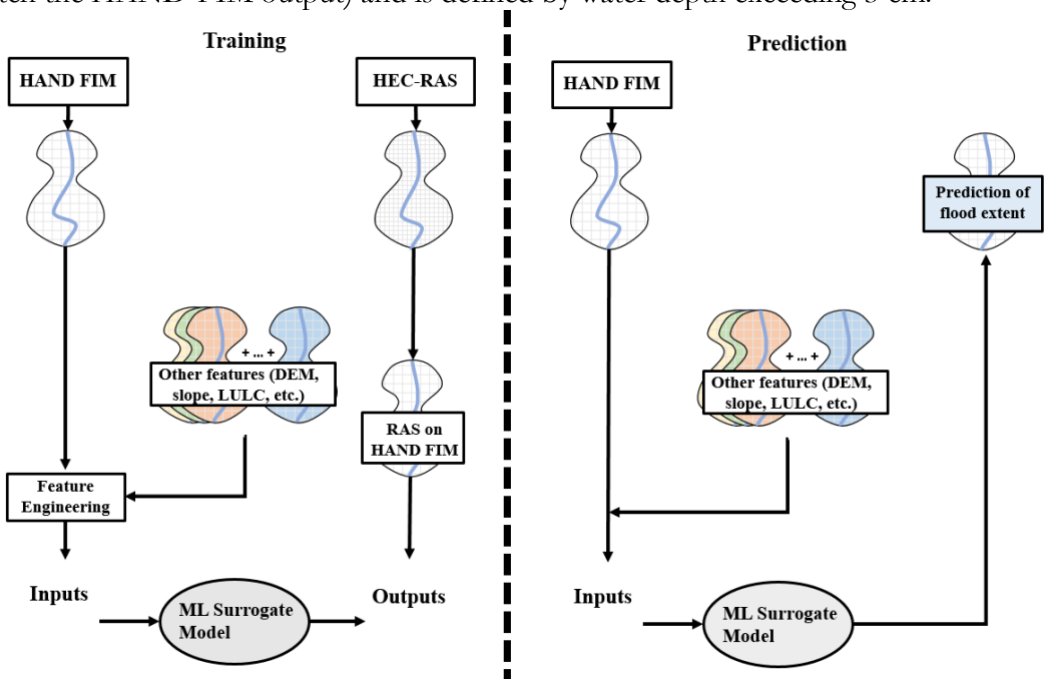


Figure 2. *Schematic of the utilized methodology, illustrating the workflow of the SM (Modified after [11]).*

5.5. Evaluation Metrics

Inundation extent evaluations involve three metrics: Probability of Detection (POD) (Eq. 1), False Alarm Ratio (FAR) (Eq. 2), and Critical Success Index (CSI) (Eq. 3. To calculate these metrics, the total number of True Positive (TP), False Positive (FP), and False Negative (FN) grid cells for categories in the contingency table are considered (*See the Supplementary Material*).

$$POD = \frac{TP}{TP+FN} \quad \text{(Eq. 1)}$$

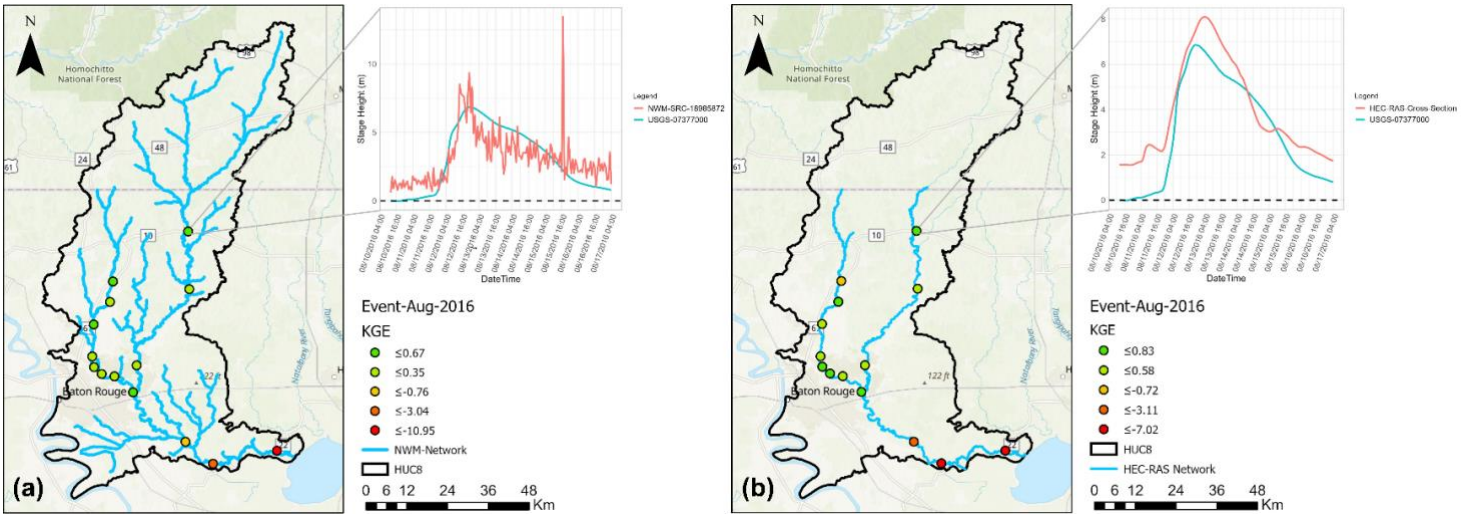
$$FAR = \frac{FP}{TP+FP} \quad \text{(Eq. 2)}$$

$$CSI = \frac{TP}{TP+FN+FP} \quad \text{(Eq. 3)}$$

6. Results and Discussion

6.1. Model Evaluation

A comparison was made between the simulated stage height data generated by the HEC-RAS and NWM-SRC frameworks and the USGS gauge height data, utilizing the Kling-Gupta Efficiency (KGE) metric for the August 2016 event. The analysis, aimed at assessing the overall model performance over the ARB, takes into account the USGS's typical selection of the gage datum based on its location beneath the streambed to accommodate stream channel changes. Consequently, it is assumed that the gauge height corresponds to the stage height. Nevertheless, the analysis considers the additional uncertainty introduced by the lack of riverbed elevation information in stage calculations. Streamflow values from NWM reaches at USGS station locations are converted to stage height using SRCs. Median stage height values from HEC-RAS cross-sections within the corresponding NWM reach are considered for comparison. The



results revealed superior performance for the upstream USGS stations (Figure 3), with HEC-RAS achieving a range of KGE of up to 0.83, compared to the NWM-SRC framework, which attained a KGE of up to 0.67.

Figure 3. *(a) Comparison of stage heights between USGS records and the NWM-SRC framework (b) comparison of USGS records with the HEC-RAS model for the August 2016 event at USGS Station-07377000.*

6.2. Surrogate Model

6.2.1 Testing the Transferability of Surrogate Model and Feature Importance

The preliminary evaluation of the SM in two watersheds shows commendable Probability of Detection (POD) values compared to HEC-RAS in both training (0.978) and test (0.995) sets (Figure 4 and Table 3a). The surrogate model also reduces the False Alarm Rate (FAR) in both training (0.326) and test (0.106) sets, indicating potential model transferability to other sub-watersheds in the ARB. Notably, the test part demonstrates a higher CSI index (0.890) compared to the training part (0.664). Permutation feature importance identifies terrain data (DEM, slope, and aspect) and Land Use and Land Cover (LULC) as crucial factors influencing predictions (Figure 4b).

Based on the results of the initial SM, the training dataset is expanded by adding more sub-watersheds (Figure 5) and excludes river network data. Using a 5-fold cross-validation approach, the SM achieves high accuracy with a POD of 0.999 and 0.999 on the training and test sets, respectively. It effectively reduces false alarms when compared to the original HAND-FIM model, with rates dropping from 0.398 to 0.283 in the training set and from 0.472 to 0.290 in the test set compared to HEC-RAS. The calculated CSI values for training and test sets are 0.717 and 0.709, respectively, indicating its efficacy in reducing false alarms associated with HAND-FIM (Table 1b). These results demonstrate the SM's potential to enhance HAND-FIM predictions.

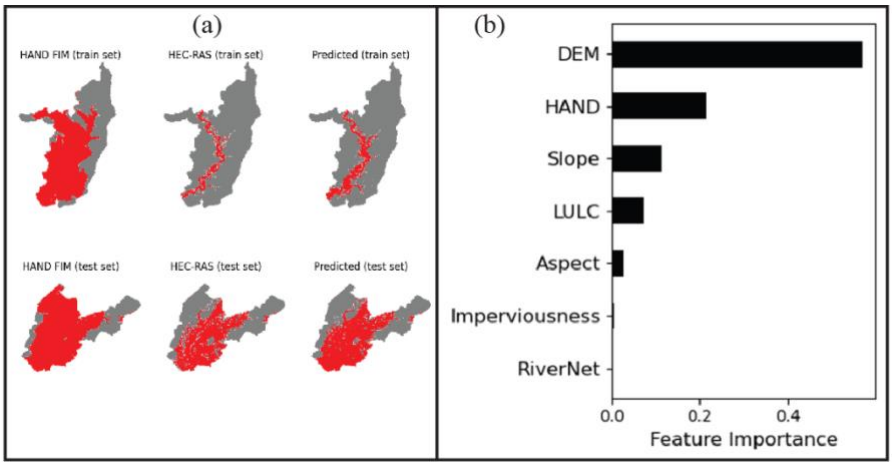


Figure 4. ***(a)****: Comparison of flood extents generated by HAND-FIM, the HEC-RAS, and predicted by the SM. Areas depicted in red represent flooded, while gray indicates dry zones and* ***(b)*** *Feature importance in the SM during the transferability test.*

Table 1. *Inundation metrics of existing (blue) and predicted (green) HAND-FIM compared to HEC-RAS-FIM for the SM trained and test for* ***(a)*** *two sub-watersheds* ***(b)*** *14 sub-watersheds in total.*

Index	Train		Test	
	HAND-FIM	SM	HAND-FIM	SM
CSI	(a) 0.156	0.664	0.626	0.890
	(b) 0.602	0.717	0.528	0.709
POD	(a) 1.000	0.978	0.999	0.995
	(b) 1.000	0.999	1.000	0.999
RFA	(a) 0.844	0.326	0.373	0.106
	(b) 0.398	0.283	0.472	0.290

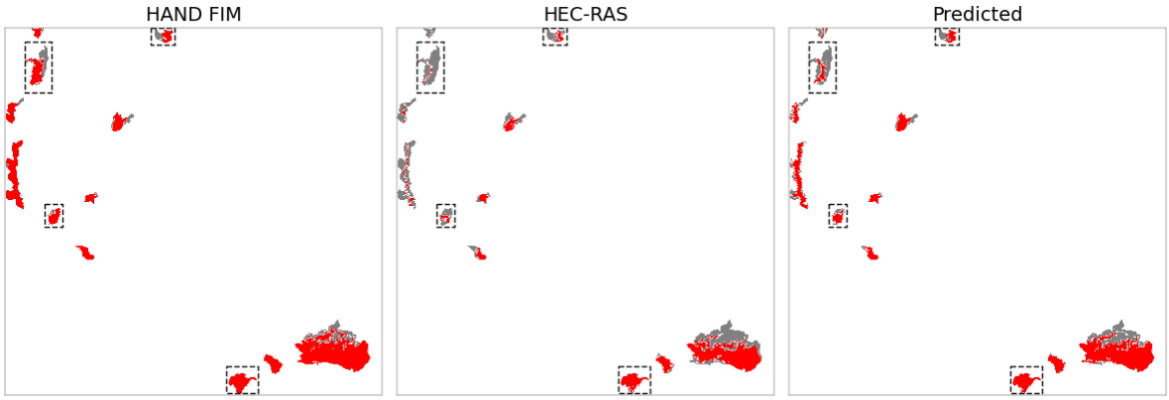


Figure 5. *Comparison of flood extents generated by HAND-FIM, the HEC-RAS, and the SM. Areas depicted in red represent flooding, while gray indicates dry zones. The sub-watershed encompassed within the dashed boundary are used for testing, while the remaining sub-watersheds are employed for model training.*

6.2.2. Computational Cost

It is important to acknowledge that the SM effectively reduced the computational burden, especially in terms of run time. The training process takes between 2 to 5 mins in an Intel Core i7-9800X CPU @ 3.8GHz, contingent on the data size, while the testing phase requires less than 1 minute. In comparison, obtaining a single time step from the simulated HEC-RAS model demands approximately 1 hour and 40 minutes, highlighting the significant time-saving advantage offered by the SM.

7.0 Conclusion

Operational flood risk management heavily relies on the accurate forecasting of the dynamic flood inundation extent. However, its data and computationally demanding structure make such a model prohibitive for continental-scale flood inundation practices. In this study, the potential utilization of the SM frameworks is explored to enhance the accuracy of a nationwide FIM method. It was observed that the HEC-RAS model outperforms the NWM and SRC combination in stage height results. Furthermore, it is found that the HAND-FIM tends to exhibit high FAR values compared to HEC-RAS. This can be attributed to the spikes seen in stage height resulting from the NWM and SRC. However, despite these differences, the SM demonstrates good performance in enhancing the FAR of the HAND-FIM method by 40%. Recent studies [18] indicate that higher resolution DEM may not directly enhance HAND-FIM, however, its integration in the SM holds huge promise for enhancing HAND-FIM's accuracy. This highlights the potential effectiveness of the SM framework in mitigating the limitations of

the HAND-FIM and improving flood inundation mapping practices on a continental scale. Considering the importance of requiring less data and computational resources in terms of operational flood management, the SM emerges as a viable solution for future flood prediction in the same region and its transferability to data/model-scarce regions. Its effectiveness can reduce risks and potential losses of life and property in regions with limited flood forecast practices.

8.0 Future Work

To fully harness the potential of the SM and to address the need for further research, our focus will be on enhancing the transferability through the implementation of sophisticated training strategies based on time series data.

The Supplementary information can be found here: https://github.com/SummerInstitute2023-FloodBusters/SummerInstitute2023_FloodBusters/blob/main/blob/Supplemental_Material.md

The GitHub link:
https://github.com/SummerInstitute2023FloodBusters/SummerInstitute2023_FloodBusters

Acknowledgments

We would like to extend our heartfelt gratitude to all individuals who played a critical role in making this research possible. Special thanks go to David Weiss, Brad Bates, and Anupal Baruah for their valuable contributions throughout this endeavor.

References

1. National Weather Service, "NWS Preliminary US Flood Fatality Statistics," 2023.
2. National Weather Service, "Summary of Natural Hazard Statistics for 2017 in the United States," 2017.
3. National Weather Service, "Summary of Natural Hazard Statistics for 2019 in the United States," 2019.
4. X. Zheng, D. G. Tarboton, D. R. Maidment, Y. Y. Liu, and P. Passalacqua, "River Channel Geometry and Rating Curve Estimation Using Height above the Nearest Drainage," 2018. [Online]. Available: https://doi.org/10.
5. P. Bates, "Fundamental limits to flood inundation modelling," *Nature Water*, Jul. 2023, doi: 10.1038/s44221-023-00106-4.
6. H. Chu, W. Wu, Q. J. Wang, R. Nathan, and J. Wei, "An ANN-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions," *Environmental Modelling and Software*, vol. 124, Feb. 2020, doi: 10.1016/j.envsoft.2019.104587.
7. F. T. Zahura and J. L. Goodall, "Predicting combined tidal and pluvial flood inundation using a machine learning surrogate model," *J Hydrol Reg Stud*, vol. 41, Jun. 2022, doi: 10.1016/j.ejrh.2022.101087.
8. F. Aristizabal, J. Judge, and A. Monsivais-Huertero, "High-resolution inundation mapping for heterogeneous land covers with synthetic aperture radar and terrain data," *Remote Sens (Basel)*, vol. 12, no. 6, Mar. 2020, doi: 10.3390/rs12060900.
9. Q. Lin, J. Leandro, W. Wu, P. Bhola, and M. Disse, "Prediction of Maximum Flood Inundation Extents With Resilient Backpropagation Neural Network: Case Study of

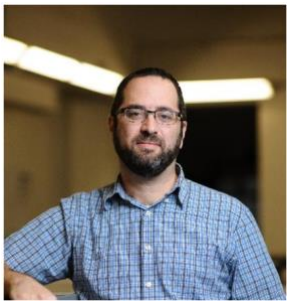
Kulmbach," *Front Earth Sci (Lausanne)*, vol. 8, Aug. 2020, doi: 10.3389/feart.2020.00332.
10. S. Kabir, S. Patidar, and G. Pender, "A machine learning approach for forecasting and visualising flood inundation information," *Proceedings of the Institution of Civil Engineers: Water Management*, vol. 174, no. 1, pp. 27–41, Feb. 2021, doi: 10.1680/jwama.20.00002.
11. N. Fraehr, Q. J. Wang, W. Wu, and R. Nathan, "Upskilling Low-Fidelity Hydrodynamic Models of Flood Inundation Through Spatial Analysis and Gaussian Process Learning," *Water Resour Res*, vol. 58, no. 8, Aug. 2022, doi: 10.1029/2022WR032248.
12. N. Fraehr, Q. J. Wang, W. Wu, and R. Nathan, "Development of a Fast and Accurate Hybrid Model for Floodplain Inundation Simulations," *Water Resour Res*, vol. 59, no. 6, Jun. 2023, doi: 10.1029/2022wr033836.
13. C. D. Rennó *et al.*, "HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia," *Remote Sens Environ*, vol. 112, no. 9, pp. 3469–3481, Sep. 2008, doi: 10.1016/j.rse.2008.03.018.
14. X. Zheng, D. G. Tarboton, D. R. Maidment, Y. Y. Liu, and P. Passalacqua, "River Channel Geometry and Rating Curve Estimation Using Height above the Nearest Drainage," *J Am Water Resour Assoc*, vol. 54, no. 4, pp. 785–806, Aug. 2018, doi: 10.1111/1752-1688.12661.
15. L. Breiman, "Random Forests," 2001.
16. H. Tyralis, G. Papacharalampous, and A. Langousis, "A brief review of random forests for water scientists and practitioners and their recent history in water resources," Water (Switzerland), vol. 11, no. 5. MDPI AG, May 01, 2019. doi: 10.3390/w11050910.
17. Brunner, G.W., 2016a, U.S. Army Corps of Engineers Hydrologic Engineering Center-River Analysis System (HEC-RAS)—2D modeling user's manual (Version 5.0): Davis, Calif., U.S. Army Corps of Engineers report CPD-68A, 171 p.
18. Aristizabal, F., Chegini, T., Petrochenkov, G., Salas, F. R., & Judge, J. (2023). Effects of High-Quality Elevation Data and Explanatory Variables on the Accuracy of Flood Inundation Mapping via Height Above Nearest Drainage. EGUsphere, 2023, 1-41.

Appendix

Team Leaders

Barbara Minsker,
Professor
Southern Methodist University

Jonathan Frame,
Senior Hydrologist
Floodbase

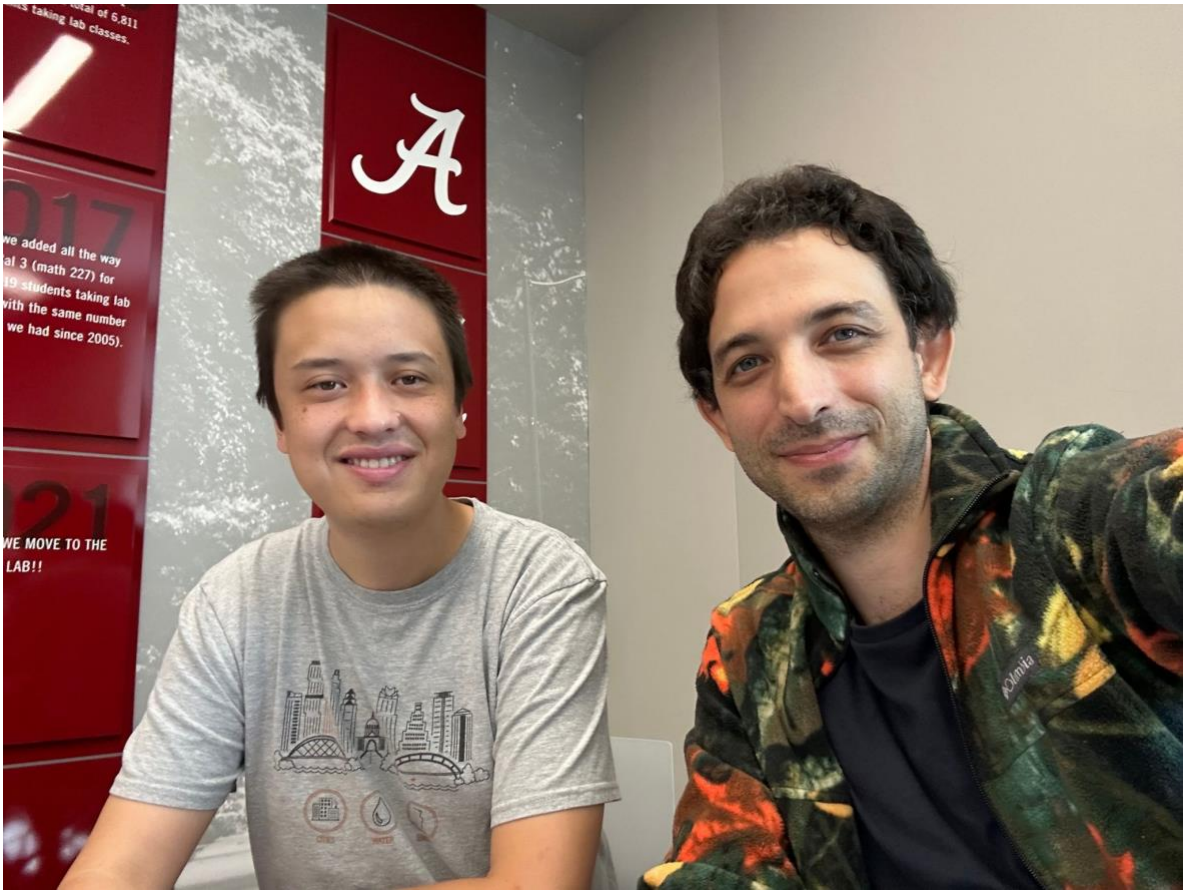


Sagy Cohen,
Professor
University of Alabama

Ehab Meselhe,
Professor
Tulane University

Kyle Mandli,
Associate Professor
Columbia University

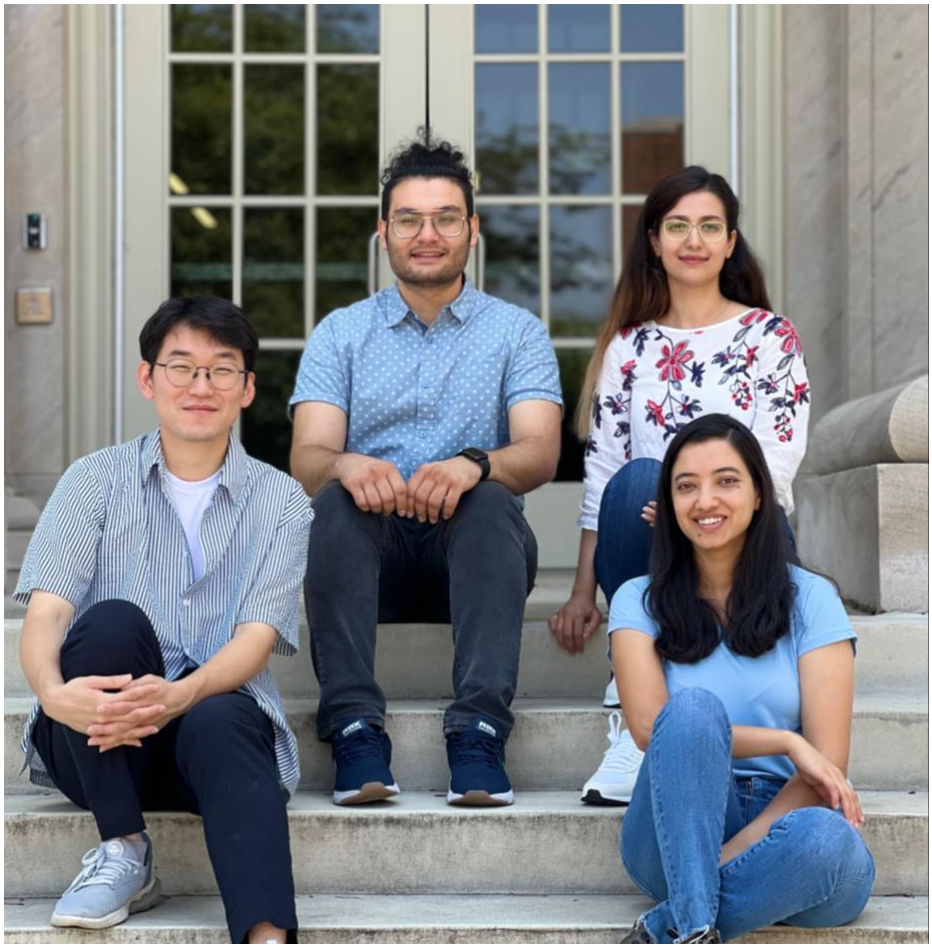
Summer Institute Course Coordinators



Mark Wang and Ebrahim Hamidi

Summer Institute Teams

Chapter 1: Data Assimilation for Improving Forecast Accuracy and Streamflow Prediction in Ungauged Basins



Ehsan Foroumandi, Parnian Ghaneei, Jeil Oh, and Sujana Timilsina

Chapter 2: On Numerical Methods and Differentiable Modeling for Soil Process Representations in the Nextgen Framework in Arid Regions

Soelem Aafnan Bhuiyan, Ryoko Araki, Jeremy Rapp, and Tadd Bindas

Chapter 3: Analysis of Flood Drivers Contributions to Compound Flooding Using Coupled Modeling and Machine Learning

Farnaz Yarveysi, Javed Ali, Sadaf MahmoudiKouhi, and Samuel Daramola

Chapter 4: Predicting Flood Inundation Susceptibility Using HAND FIM, Crowd-Sourced and Satellite Data with Machine Learning

Roja Najafi, Azadeh Hosseinzadeh, Monica Cardona, and Elnaz Heidari

Chapter 5: Quantifying the Sources of Uncertainty in OWP HAND-FIM Predictions

Meraj Sohrabi, Annie Holt, Karina Larco, and Umanda Abeysinghe

Chapter 6: Improving the Fidelity and Performance of OWP HAND-FIM Using a Surrogate Model Technique (SMT)

Berina Mina Kilicarslan, Victor Obi, and Qianqiu Longyang