TECHNICAL REPORT 21

July 2025

CUAHSI

allied for water science

THE UNIVERSITY OF ALABAMA® | Alabama Water Institute

Water Prediction Innovators Summer Institute Report 2025

Water Prediction Innovators Summer Institute Report 2025

Editors:

Francisco Gomez

Parvaneh Nikrou

Prepared in cooperation with the Consortium of Universities for the Advancement of Hydrologic Science, Inc.

CUAHSI Technical Report No. 21

Version 1.0

July 2025

This research was supported by the Cooperative Institute for Research to Operations in Hydrology (CIROH) with funding under award NA22NWS4320003 from the NOAA Cooperative Institute Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the opinions of NOAA.

Suggested Citation:

Gomez, F., Nikrou, P., et al. (2025). Water Prediction Innovators Program - Summer Institute, CUAHSI Technical Report, HydroShare, http://www.hydroshare.org/resource/91214cbfb1d84dbeaf69282bb530c52e.

Contents

Preface

This report summarizes the research conducted during the 2025 Water Prediction Innovators Summer Institute (WPI-SI). The WPI-SI is the result of a partnership between the Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI) and The University of Alabama. 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 Cooperative Institute for Research to Operations in Hydrology (CIROH) funds this programming and supports the WPI-SI as a part of its mission to "to innovate research that enhances our understanding of how coupled atmosphere-oceanland-biosphere components interact hydrologically and transform this new knowledge into operational research products benefitting society." Funding for the Summer Institute is provided by the National Oceanic and Atmospheric Administration (NOAA), awarded to the Cooperative Institute for Research on Hydrology (CIROH) through the NOAA Cooperative Agreement with The University of Alabama, NA22NWS4320003.

Since 2015, The University of Alabama campus has hosted the WPI-SI. Graduate students come from different universities across the United States, and work alongside academic faculty, federal scientists, researchers, private companies, and contractors in Tuscaloosa over the seven weeklong program. The research themes that students work under are ideated with input from the Office of Water Predictions (OWP) which has been supportive of this program from its inception.

Student fellows formed teams based on shared research interests and worked under the guidance of theme leaders throughout the program. This year, six teams were established, each focused on a specific research question. Projects included the application of deep learning techniques for flood inundation mapping and discharge forecasting, as well as the development of tools for visualizing flood extent in urban areas at higher spatial resolutions.

In addition to their research, participants took part in social events and outings, forming lasting connections with their peers. Course Coordinators, both of whom were fellows in the past, offered assistance and support throughout the summer. For the fellows, the Summer Institute is more than a research program. It fosters teamwork, networking, and the development of lasting friendships.

Fellows

The 2025 cohort included 22 graduate students from 17 different universities across the United States. They began their journey with virtual meetings, which culminated in a two-week bootcamp to form teams, and participated in training on coding, data retrieval, and understanding key topics to their research projects. Then research teams mentored by theme leaders worked to complete their research projects, while giving weekly research talks to share their progress and receive feedback and guidance. The program culminates with this report and a Capstone Event wherein research teams present their work to a broad audience of researchers and peers. The fellows and theme lead hail from various academic departments, including civil engineering, geography, and earth sciences.

Themes and Theme Leads

The WPI-SI 2025 themes and theme leads were:

- The "Operational Flood Inundation Mapping" theme was led by Sagy Cohen (The University of Alabama) and Yinphan Tsang (University of Hawai'i at Mānoa). Additional technical support was provided by members of the Surface Dynamics Modeling Lab (SDML - The University of Alabama).

- The "LSTM for NextGen" theme was led by Jonathan Frame (The University of Alabama) and Amobichukwu Amanambu (The University of Alabama).

- The "Pluvial Urban Flooding" theme was led by Marouane Temimi (Stevens University) and Jonathan Frame (University of Alabama). Additional technical support was provided by Mohamad Abdelkader (University of Iowa).

- The "Visualization of Urban Road Flooding" theme was led by David Maidment (University of Texas at Austin) and Kelsey R. McDonough (The Water Institute). Additional technical support was provided by Matt Bartos (University of Texas Austin) and Dean Djokic (ESRI).

Project Summaries

The 2025 WPI-SI projects are summarized below. Chapters 1-6 present the complete reports.

1. Project within the Operational Flood Inundation Mapping theme:

Chapter 1: "A Comparative Analysis of Multi-Method Flood Inundation Mapping: Quantifying Fidelity–Accuracy Tradeoffs and Bridging Gaps via Hybrid Deep Learning". This report evaluates three operational Flood Inundation Mapping (FIM) methods such as the NOAA HAND-FIM, Ripple1D, and Pre-Canned hydraulic libraries against satellite-derived benchmarks. While Pre-Canned FIMs offer higher spatial accuracy, they suffer from limited coverage. HAND-FIMs cover the full stream network but often underpredict extents. To improve temporal continuity, a novel deep learning model called FIM-BridgeNet was used. The approach produces accurate, scalable, and temporally continuous FIMs, supporting improved flood preparedness applications.

2. Projects within the LSTM for NextGen theme:

Chapter 2: "Scoring Rule-Based LSTM Models for Flow Forecasting in the NextGen Framework. This report presents a deep learning framework for predicting surface runoff in distributed catchments using a distribution-based loss function that bypasses physics-based routing during training. The novel ML-based strategy selects optimal modeling structures to capture basin hydrological signatures. The method enables efficient transfer of spatiotemporal information, supporting scalable calibration within the Next Generation Water Observing System and the National Water Model.

3. Projects within the Pluvial Urban Flooding theme:

Chapter 3: "Towards Representing Pluvial Flooding within NOAA's NextGen Modeling Framework". This report evaluates the capability of the Next Generation National Water Modeling Framework to detect pluvial flooding at local scales. By analyzing Pluvial Flood Indices across varying catchment sizes and land cover types, results reveal higher PFIs in small, urbanized basins, with variability depending on urban extent. While PFIs generally decrease with increasing catchment

area, they rise again when multiple urban zones are included. Findings provide benchmarks for improving the NextGen spatial reliability in pluvial flood forecasting.

Chapter 4: "Demonstrating the Feasibility of DL-Based Pluvial Flood Mapping in Urban Settings". This report explores the use of deep learning as a surrogate for high-resolution, street-scale pluvial flood modeling in urban environments. Applied to Hoboken, NJ, the framework combines historical and synthetic rainfall with spatial inputs using 2D and 1D Convolutional Neural Networks to generate binary flood maps. The model achieved strong performance demonstrating the potential of deep learning models for rapid, accurate pluvial flood forecasting within future platforms like the NOAA NextGen framework.

4. Projects within the Visualization of Urban Road Flooding theme:

Chapter 5: "Towards a Flood Navigation and Safety Decision Support Tool: A Pilot for Emergency Responders in Travis County, Texas". This report presents FLO-NAVSAFE, an operational decision support tool that translates National Water Model forecasts into county-level flood inundation maps for emergency response. Using the Height Above Nearest Drainage method and time-lagged ensembles, it offers near real-time and short-term forecasts with uncertainty characterization. FLO-NAVSAFE integrates stakeholder priorities and vulnerability data through a user-friendly Esri interface.

Chapter 6: "Estimating Flood Inundation Using a Densified Stream Network in Travis County, TX". This report assesses the impact of stream network density on flood modeling by downscaling National Water Model runoff to a finer-resolution network in Travis County, Texas. Compared to the original NWM network, the densified network better captures low-water crossings and yields improved streamflow estimates during a July 2025 storm. The densified model also produces more extensive and continuous inundation patterns, demonstrating its value for enhancing community-scale flood forecasting in urban areas.

Acknowledgments

Since the first Summer Institute in 2015, more than 250 graduate student fellows have collaborated on research projects focused on advancing the water prediction capabilities for the United States. Now in its 10th year, the Summer Institute has grown into a nationally recognized program that brings together emerging scientists, engineers, and professionals to develop innovative solutions for water-related challenges. The 2025 Water Prediction Innovators Summer Institute builds on this legacy and would not have been possible without the commitment and hard work of many people:

We would like to express our sincere gratitude to the CUAHSI staff whose efforts were instrumental in making this program a reality: Jordan Read, Julia Masterman, Emily Clark, Irene Garousi-Nejad, and Tony Castronova. We also thank Ed Clark, former director of the National Water Center (NWC), for sharing his valuable insights and extensive experience with the fellows. Thanks to Fred Ogden, Fernando Salas, Carson Pruitt, and Greg Petrochenkov from the NWC for their technical support throughout the program. We also want to give special thanks to CIROH, especially Sagy Cohen and Lanna Nations, for handling logistics at the University of Alabama, including travel, housing, and campus coordination.

Technical training was generously provided by Charity McCalpin (CIROH); Irene Garousi-Nejad and Tony Castronova (CUAHSI); Jordan Read (CUAHSI); Matt Bartos (UT Austin); Dean Djokic (Esri); Dirk Schwanenberg (KISTERS AG); and Dr. Joanna Krajewski and Dr. Witold Krajewski (The University of Iowa). We also appreciate all the Team Leaders who led the bootcamp sessions and continued to check in on the progress of the fellows. Members of the Surface Dynamics Modeling Lab (University of Alabama) offered multiple training sessions and were crucial in addressing numerous technical questions throughout the seven weeks of the Summer Institute. We also thank James Halgren and the Alabama Water Institute team for their guidance and for providing technical assistance and support with computational resources.

We extend our sincere gratitude to everyone whose name may not appear here but who contributed to this program's great success.

Thanks to everyone who made a contribution to the WPI-SI.

Francisco Gomez

Student Course Coordinator, Water Prediction Innovators Summer Institute 2025

Ph.D. Candidate, University of Alabama

Parvaneh Nikrou

Student Course Coordinator, Water Prediction Innovators Summer Institute 2025

Ph.D. Candidate, University of Alabama

Chapter 1:
A Comparative Analysis of Multi-Method Flood Inundation Mapping: Quantifying Fidelity–Accuracy Tradeoffs and Bridging Gaps via Hybrid Deep Learning

Faezeh Maghsoodifar[1], Mohamed Mowafy[2], Md Shadman Sakib[3], Haotian Wang[4], Sagy Cohen[5*], Yinphan Tsang[6*]

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

[2] University of Cincinnati; *mowafymd@mail.uc.edu*

[3] Virginia Tech; *shadman@vt.edu*

[4] University of Connecticut; *haotian.wang@uconn.edu*

[5*] The University of Alabama, *sagy.cohen@ua.edu*

[6*] University of Hawai'i at Mānoa, *tsangy@hawaii.edu*

* Theme Leader

Abstract: Flood Inundation Map (FIM) is essential for flood preparedness, yet operational products often face trade-offs between spatial coverage and temporal continuity. This study evaluates the performance of three common mapping approaches, the NOAA Office of Water Prediction operational Water Prediction Height Above Nearest Drainage (OWP HAND-FIM), Ripple1D (R1D), and Pre-Canned (PC) hydraulic libraries, against satellite-derived flood extents (benchmarks). The model's predictions are evaluated across varied and stream orders. Results show that while PC-FIMs provide greater spatial accuracy, they are constrained by limited coverage and discontinuity. HAND-FIMs offer full stream network coverage but often underpredict flood extent due to bias in discharge input. R1D-FIMs perform moderately but remain sensitive to flow input assumptions. To address the temporal sparsity between pre-canned RPs, we introduce FIM-BridgeNet, a hybrid Deep Learning (DL) model that couples an attention Residual UNet (ResUNet) with optical flow warping. This architecture learns spatial deformations between low and high RP PC-FIMs to generate intermediate flood maps. Outputs are translated into water depth and discharge using elevation-based post-processing. The model achieves strong performance across RP ranges. These findings support the use of hybrid models to generate scalable, temporally continuous FIMs.

---

1. Motivation

The NOAA OWP operational Flood Inundation Mapping (FIM) forecasting framework is based on three complementary products: HAND-FIM, HEC-RAS-based models such as R1D and PC-FIMs for specific Return Periods (RPs). Each of these products demonstrates strengths in either computational efficiency, spatial coverage, or mapping accuracy; however, none simultaneously achieves all three. Additionally, Pre-Canned (PC) FIMs are limited by discrete return period (RP) intervals. The growing archive of satellite-derived flood extent offers a unique opportunity to

evaluate these methods across events ranging from nuisance to extreme. This convergence of complementary models and independent benchmark FIM motivates our project: to establish a clear, data-driven understanding of current inundation-mapping skill and to create a scalable hybrid Deep Learning (DL) technique that fills the temporal gaps between existing PC-FIM RP layers.

2. Objective and Scope

The project evaluates how accurately PC-, R1D-, and HAND-FIM predictions using RS-based benchmarks, to determine whether broader low-fidelity or localized high-fidelity models offer greater operational value. Building on those findings, we develop FIM-BridgeNet, a hybrid deep-learning framework, to generate realistic intermediate inundation maps and estimate the associated discharges. The scope of work includes compiling RS benchmarks, computing accuracy metrics, analyzing trade-offs among accuracy and coverage, and training and validating FIM-BridgeNet. We believe that this research will guide agencies producing operationally accurate and temporally continuous, reliable FIM in near real time.

3. Previous Studies

Disastrous river flooding affects approximately 125 million people globally each year [1]. Global projections indicate that flood frequency is expected to continue rising over the next two decades [2]. This upward trend is evident in the United States (U.S.), where floods are among the most frequent and financially devastating natural disasters [3, 4]. According to the International Disaster Database, 691 flood-related fatalities occurred in the U.S. between 2000 and 2024 [5]. Most recently, a flash flood in Central Texas along the Guadalupe River claimed over 120 lives. Between 2010 and 2022, the U.S. experienced 21 billion-dollar flood events, exceeding the total recorded in the 1980s, 1990s, and 2000s combined, adjusted to 2023 dollars [6]. Although flood damage cannot be entirely prevented, early warnings, reliable forecasts, and accessible, actionable information can significantly reduce impacts [7, 8]. Providing emergency responders with accurate and timely flood extent maps enables more effective decision-making, conserves time, resources, and potentially saves lives [9]. These conditions collectively highlight the urgent need to improve flood modeling and forecasting capabilities across the United States. To meet this need, various approaches have been developed for real-time and forecast-based FIM. These efforts broadly fall into three categories: empirical, conceptual, and hydrodynamic. Empirical methods are observation-based, relying on high-water marks, surveys, aerial photography, and satellite imagery [10, 11]. While not predictive, they are critical for validating model performance. Conceptual methods, in contrast, are terrain-based and simulate inundation using simplified hydraulic assumptions rather than solving physical flow equations. These include planar surface methods, inclined planes, and the HAND approach [12]. Their primary appeal lies in their low computational cost, which makes them practical for large-scale or time-sensitive applications. Among them, HAND has shown robust predictive skill, particularly when derived using reliable streamflow and topography data [12-15]. The NOAA Office of Water Prediction (OWP) has developed a robust framework based on the HAND approach, which was coupled with streamflow predictions by the National Water Model (NWM) to support rapid FIM forecasting [16]. This integration, now part of the operational U.S. flood forecasting framework, has shown better performance in higher-order streams, albeit with some underestimation in lower-order channels [17]. Conceptual models like HAND remain well-suited for real-time applications and data-sparse regions, given their operational efficiency and scalability [18].

In contrast, hydrodynamic (physics-based) models simulate water movement by solving fluid dynamics equations in one (1D), two (2D) dimensions. These models, such as HEC-RAS 1D (https://www.hec.usace.army.mil/), are capable of capturing the complex behavior of floods across varied landscapes and infrastructure scenarios. They can be integrated with hydrologic models to simulate flood extent under different meteorological or infrastructural conditions. However, this physical realism comes at a high computational cost, especially when run at fine spatial resolutions (<10 m) or over large domains (>1000 km²). The computational intensity increases significantly in probabilistic or ensemble modeling applications. Despite the improved accuracy of 1D simulations, real-time operational use remains limited [19]. To address these limitations, pre-simulated or "pre-canned" flood map libraries have been developed, leveraging the accuracy of hydrodynamic models while avoiding on-demand simulation. These libraries typically contain flood extents for a discrete set of return periods (RPs), ranging from bank-full conditions to extreme flood scenarios [20]. By interpolating between pre-generated maps, practitioners can provide rapid flood guidance without incurring the full computational burden. These libraries have emerged as promising tools for operational forecasting, particularly when real-time decision-making is critical [21].

FEMA has developed an extensive archive of one-dimensional (1D) HEC-RAS models to support Base Level Engineering (BLE) initiatives and the National Flood Hazard Layer. While these models are robust and widely accepted for floodplain mapping, their computational demands make them impractical for near-real-time applications. To address this limitation, the Ripple1D (R1D) framework (https://github.com/NGWPC/flows2fim) repurposes the existing HEC-RAS models to generate NWM reach-specific FIM libraries and rating curves. The core objective of R1D is to execute these models once across a predefined range of streamflow scenarios and store the outputs in a reusable library to be used for rapid generation of FIMs in near real-time. Nevertheless, these FIM products are generally limited to a few discrete RPs, such as the 2-, 5-, 10-, 25-, 50-, and 100-year events, requiring assumptions about how flood extents evolve between them. Recent geospatial advances have begun to loosen these constraints, such as Delaunay triangulation, have been employed to guide image warping between temporally distant satellite scenes, reconstructing plausible intermediate flood states while preserving fine-scale channel geometry [23]. For example, the utility of warping satellite imagery to bridge observation gaps in river and delta morphodynamics has been previously demonstrated [24]. These innovations suggest a rethinking of flood extent interpolation, not as a linear blend, but as a constrained image-interpolation task that respects hydraulic monotonicity. Deep learning approaches have embraced this shift. Conditional Generative Adversarial Networks (cGANs), such as floodGAN and cGAN-Flood, generate high-resolution flood maps from rainfall fields orders of magnitude faster than traditional models, even generalizing to unseen catchments with meter-scale accuracy [25, 26]. However, such models are not explicitly designed to generate intermediate RP-FIMs and may fail to preserve the physical structure of flood extents. Addressing this gap, the present study introduces a hybrid framework that couples a ResUNet-based semantic segmentation model [23, 26] with deformation-informed warping. This architecture incorporates attention mechanisms [27] and geometric constraints, enabling the generation of intermediate FIMs e.g., 3- and 4-year RP events from PC 2- and 5-year RP FIMs. The resulting outputs, in standard raster format, can be seamlessly processed by tools like FwDET-GEE for depth estimation and subsequent discharge computation, offering a scalable, computationally efficient alternative to exhaustive simulations. Despite widespread use of linear interpolation between pre-canned FIMs, limited research has formally assessed their accuracy or defined the conditions under which they perform best. Prior studies have emphasized the need for comprehensive evaluations of these libraries [28]. To that end, this study benchmarks HEC-RAS-based pre-simulated FIMs against OWP HAND-FIMs using Remote Sensing (RS)-derived

flood extents. Additionally, it introduces and evaluates hybrid DL-based proof of concept as an alternative to simple interpolation, aiming to improve the physical realism and reliability of real-time flood mapping.

4. Methodology

4.1. Study Area and Data Resources

The case studies used to evaluate the FIM prediction techniques are located in southeastern Texas, encompassing several watersheds within the state boundaries (**Figure 1**). These watersheds span across regions near major cities such as Houston and Austin, and represent diverse hydrological and land use characteristics, including both urbanized and rural landscapes. Multiple Areas of Interest (AOIs) were delineated within the study domain to examine flood behavior, watershed structure, and hydrological responses. The detailed subfigures (a–c) show the distribution and spatial context of each AOI within the corresponding watershed in (d) CONUS. (a) and (c) are used in evaluation, and (b) is used in both evaluation and deep learning gap-filling. This study area was selected for its varied topography, land cover, and historical flood vulnerability. The data and models developed in this region aim to support hydrological analysis, flood mapping, and data-driven prediction efforts using remote sensing and machine learning approaches. PlanetScope imagery from Planet's smallsat constellation was acquired via NASA's Commercial Smallsat Data Acquisition (CSDA) Program for flood events cataloged in the event repository, providing 3-m spatial resolution imagery with a daily revisit time [29]. The data are orthorectified, radiometrically corrected, and terrain-adjusted multispectral products. Cloud- and shadow-contaminated pixels are excluded from the analysis. Four-band analytic imagery is used, composed of blue, green, red, and near infrared (NIR). Sentinel-1 [30] and NOAA Emergency Response Imagery [31] are also used.

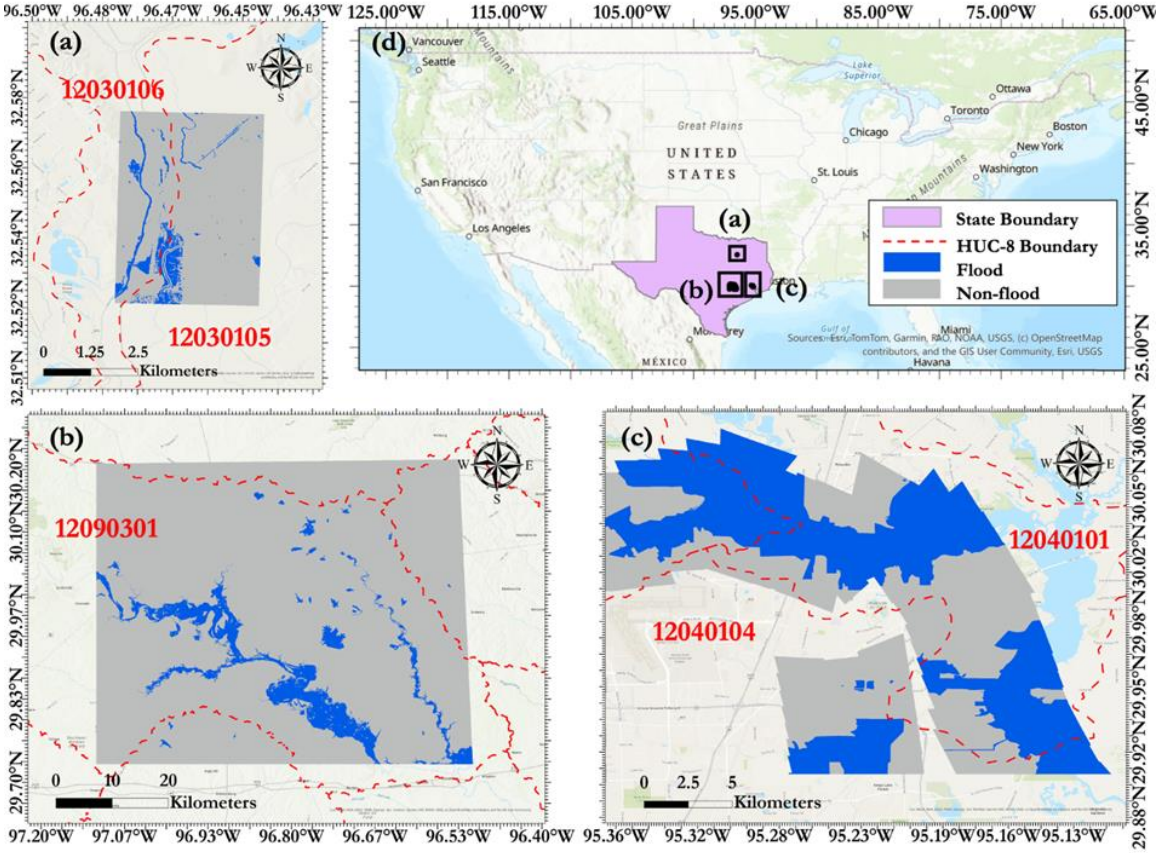


Figure 1. *The study area locations of (a) Lower Brazos (12030105) and San Bernard (12030106), (b) Lower Colorado (12090301), and (c) San Jacinto River (12040101) and West Fork San Jacinto (12040102) of (d) CONUS.*

4.2. Remote Sensing-Based Benchmarking

To generate a remote sensing (RS)-based benchmark, a comprehensive Flood Inundation Benchmark Event Repository (FIBER) was developed by aggregating data from multiple established sources. Flood event records were drawn from the Dartmouth Flood Observatory (DFO) Global Archive, which catalogs large flood events globally from 1985 onward, with a focus on post-2013 events within the CONUS region. General flood locations were obtained from the RAdar Produced Inundation Diary (RAPID), which provides high-resolution (but low accuracy) flood maps across the U.S. from 2016 to 2019 [32]. Additionally, historical storm and precipitation records from the NOAA Storm Events Database (1950–2025) were used to link flood events with corresponding atmospheric drivers. Through a series of filtering, validation, and curation steps, these sources were consolidated into FIBER, which serves as the authoritative benchmark for assessing the spatial accuracy of flood inundation products. FIMs derived from remote sensing (RS-FIMs) were produced using high-resolution optical imagery. Initial water classifications were performed by applying the Otsu thresholding method [33] to the near-infrared (NIR) band and comparing the results to those obtained using the Normalized Difference Water Index (NDWI); the delineations from the NIR band were ultimately selected due to better performance. To reduce misclassifications caused by small land features, a supervised classification model was employed to refine the initial water masks. Subsequently, the identified flood regions served as seed areas for the spatial expansion of the inundation extent. This expansion was informed by both global water level observations and terrain constraints derived from a high-resolution digital elevation model. A hydrologically guided region-growing algorithm, as described in [34], was applied to improve detection in regions where floodwaters

were likely present but visually obscured in the imagery, such as beneath forest canopies or urban infrastructure. The final RS-FIMs underwent visual inspection and were evaluated for hydro-morphological coherence to assess the reliability of the delineated flood extents.

4.3. Determining Event Return Period

USGS streamflow gauges (site a: 08062500, site b: 08161000, site c: 08068090) located near the downstream end of the inundated area were used as the primary source for determining event RP. However, in the majority of the benchmarks these downstream gauges were situated on lakes or following a dam. As a result, these gauges cannot accurately capture the true flood dynamics. In these cases, an upstream gauge along the main river channel was selected as an alternative source of observed data. Using the peak discharge data from these identified USGS gauges, the RP was computed through the Expected Moments Algorithm (EMA) following the procedures outlined in Bulletin 17C of the U.S. Army Corps of Engineers' HEC-SSP software. In the absence of USGS gauge nearby (either downstream or upstream), the analysis relied on discharge data from NWM. For events prior to January 2023, discharge time series were obtained from NWMv3 retrospective simulations. For events beyond this period, short-range NWM forecasts were used instead. In both cases, the peak discharge at the most downstream NWM reach of the inundated area was extracted and compared against its corresponding frequency distribution to determine the event RP.

4.4. Generation of HAND, PC and R1D-FIMs

To generate the HAND-FIMs we utilized the open-source *fimserve* package [35]. This package generates operational FIMs based on the HAND method developed by NOAA's OWP. HAND-FIM generation (**Figure 2**) begins with identifying the HUC code, event start and end date corresponding to each benchmark flood event. The benchmark event's image acquisition time, hereafter referred to as the Benchmark Timestamp (BMT), serves as the initial reference for HAND-FIM generation. However, NWM simulations, especially retrospective runs, are known to underpredict or overpredict streamflow when compared to observed USGS discharge values. To address this bias, we employed a simplistic temporal alignment strategy to synchronize modeled and observed conditions.

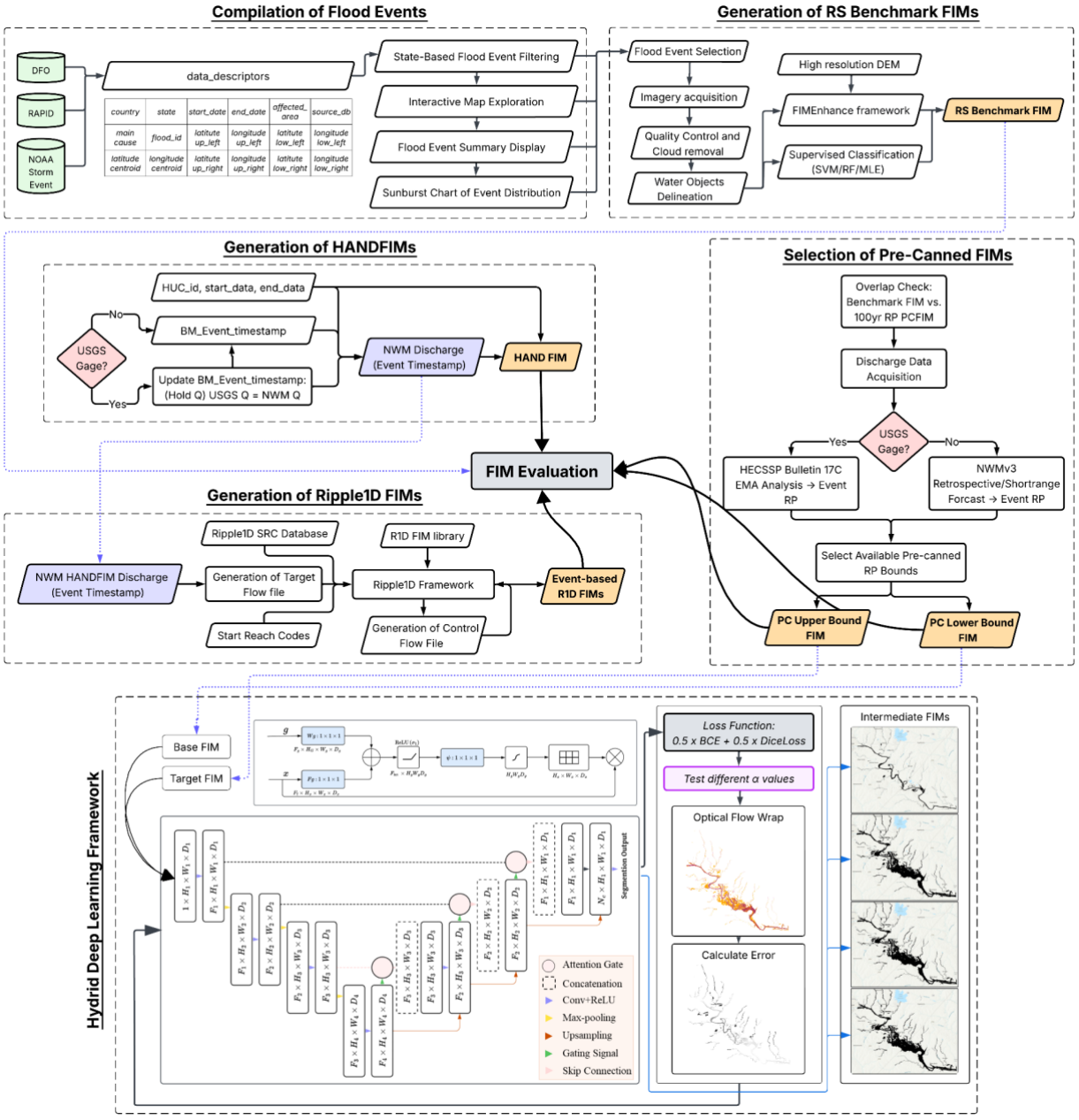


Figure 2. *Schematic diagram workflow for generating and evaluating three types of FIMs: HAND-FIMs (HAND), Ripple1D event-based FIMs (R1D-FIMs), and pre-canned FIMs (PC-FIMs)*

Specifically, the discharge from the USGS gaging station used for RP calculation was compared against the NWM discharge data at its corresponding stream *feature_id*. If the observed BMT discharge fell within the NWM discharge range, we selected the timestamp where the NWM discharge most closely matched the USGS value. This selection was done preferentially on the rising or falling limb of the hydrograph, depending on which stage the BMT aligned within the observed USGS hydrograph. However, if the USGS discharge exceeded the maximum NWM discharge within the event period, the timestamp corresponding to the NWM peak discharge was selected for HAND-FIM generation. In cases where no USGS gauge was available, the BMT was used directly without adjustment. This final (adjusted or default) timestamp was then used

to extract NWM discharge for the entire HUC, serving as input for both the HAND-FIM and the target discharge for R1D-FIM generation.

The R1D-FIM generation (**Figure 2**) is based on a set of start reach(es), and a target flow file. For each start reach, the framework identifies the closest matching rating curve from a precomputed rating curve database by comparing the input flow to available scenarios. The resulting stage is then used as a boundary condition for the next upstream reach, combined with the corresponding flow from the target file. This process is repeated iteratively, selecting closest Synthetic Rating Curve (SRC) discharge at each step based on updated boundary conditions, and progresses upstream until the full network extent is covered. In our research, the adjusted or default NWM discharge at the BMT was used to define the target flow input for generating event-specific R1D-FIMs. The discrete pre-canned FIMs corresponding to 2-, 5-, 10-, 25-, and 50-year RP were provided by Dewberry Inc. (https://www.dewberry.com/), the developers of the R1D framework. These FIMs were generated using representative discharges for each return period scenario. Based on the RP determined for each benchmark event, we selected the two closest PC-FIMs that represent the upper [PC(UB)] and lower [PC(LB)] RPs for each case study.

4.5. Generation of AOIs

Within the PC and R1D-FIMs we had several missing sub-reach models or regions with lakes or artificial reservoirs. These missing FIM sections were sparsely distributed across lower-order streams (<3). However, several higher-order streams (e.g., orders 5–8) exhibited gaps, often accounting for a substantial portion of the FIM coverage. Therefore, when delineating the AOIs, it was important to account for these discontinuities. Including all missing PC-FIM regions would inadvertently favor the HAND-FIM model, which offers near-complete coverage across stream networks. Selecting AOIs with extensive discontinuities in the PC-FIM would thus create an imbalanced comparison. On the other hand, restricting the analysis to regions covered solely by PC models would introduce bias in their favor, as it disregards areas where these pre-canned sub-models are absent. While such an approach enables a more controlled comparison of model accuracy between the high-fidelity PC-FIMs and the lower-fidelity HAND-FIMs, it simultaneously conceals a key limitation of the pre-canned framework, its limited spatial coverage.

Therefore, to ensure fair tradeoff analysis, two sets of AOIs were created for a single FIM region. The first one represents a broader area delineated based on the extent of benchmark inundation while ignoring the spatial discontinuities of pre-canned models. In defining acceptable discontinuities, only missing pre-canned FIM segments associated with order 2 or lower were retained within an AOI. These AOIs were standardized using a Convex Hull (CH) approach to define a minimum bounding polygon covering the inundated pixels. This method allows for a uniform and objective delineation of AOIs based on benchmark flood extent. They served as the primary spatial reference for identifying regions with continuous pre-canned model coverage, predominantly along streams of order 5 and above. Within each primary AOI, a second AOI is delineated by intersecting the 100yr RP PC-FIM wet cells with the benchmark raster inundation extent. Since the full lateral extent of the original HEC-RAS sub-model is unknown, a 100 m buffer was applied to approximate the dry floodplain and infer the model's operational boundary. These secondary AOIs provide a comparison that reduces biases due to gaps in the HEC-RAS simulations, allowing us to compare the differences in the FIM forecasting approaches (dynamic, pre-canned and hybrid) rather than specific modeling implementations.

4.6. Synthesizing Intermediate FIMs

This study proposes a Deep Learning (DL)-based framework to generate intermediate FIMs between existing benchmark scenarios defined by different RPs. The goal is to produce realistic flood extents for intervals not directly available in traditional flood mapping products. The approach integrates spatial transformation techniques with a customized neural network architecture capable of learning from known flood scenarios and inferring intermediate extents. Known flood maps were preprocessed to ensure consistency in spatial resolution and alignment and then converted into binary representations suitable for model training. To manage high-resolution inputs, a patch-based processing strategy was used, preserving spatial fidelity while optimizing memory use during training. The DL model combines semantic segmentation with spatial deformation priors derived from observed flood extents. A neural network with encoder-decoder architecture and enhanced feature selection modules was trained to learn flood extent patterns between pairs of low and high RPs. The training process utilized displacement fields computed between these scenarios to provide geometric guidance, helping the model predict how flood boundaries evolve. Intermediate flood extents were then generated at variable resolution depending on user-defined interpolation controls. Model training was conducted with a combination of loss functions to ensure balanced learning across diverse flood patterns and was optimized using standard techniques for segmentation tasks. Performance evaluation relies on established classification metrics that quantify both detection accuracy and false prediction rates. The predicted flood extents were further translated into hydrologic variables such as depth and discharge. Water depths were estimated using a terrain-differencing method, and discharge was calculated by matching predicted depths with synthetic or empirical rating curve relationships, using parameters from existing hydrologic databases. This end-to-end process enables enhanced temporal resolution in flood mapping and supports hydrologically meaningful decision-making, while the technical components remain abstracted to limit direct reproducibility.

4.7. FIM Evaluation

FIM evaluation was conducted using the open-source FIM Evaluation Framework (FIMeval) [36]. It evaluates the FIMs based on a pixel-to-pixel comparison between Model-predicted and Benchmark FIMs (M-FIM and B-FIM). First the M- and B-FIM are converted into binary raster representing flooded (1) and non-flooded (0) pixels. Next, permanent water bodies (PWBs) are removed from the binary raster to isolate and evaluate only the flood-affected pixels. Finally, binary B-FIM and M-FIM are compared using map algebra to produce a confusion matrix raster with each pixel classified as True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Based on the pixel classifications, five performance metrics are calculated – F1 Score, Critical Success Index (CSI), Probability of Detection (POD), Accuracy, and Precision. The F1 Score is the harmonic means of Precision and Sensitivity, which is useful for evaluating performance on imbalanced data in this case flooded and non-flooded pixel ratio. POD measures flood detection without penalizing false alarms, while CSI measures correctly predicted floods while penalizing misses and false alarms. Accuracy is the proportion of correctly classified pixels (both flood and dry). Lastly, Precision is the fraction of predicted flood pixels that are truly flooded in the B-FIM raster.

5. Results

Our approach of presenting the results include the producing of 4 contingency maps and performance metrics: (i)HAND-; (ii)PC(LB)-, (iii) PC(UB)-, and the event-based R1D-FIM, via two different approaches for the map boundaries (Convex Hull at the top panel and traced to

the Pre-canned boundaries with a 100 meter buffer on the lower panel), as shown in **Figure 3** and **Figure 4**.

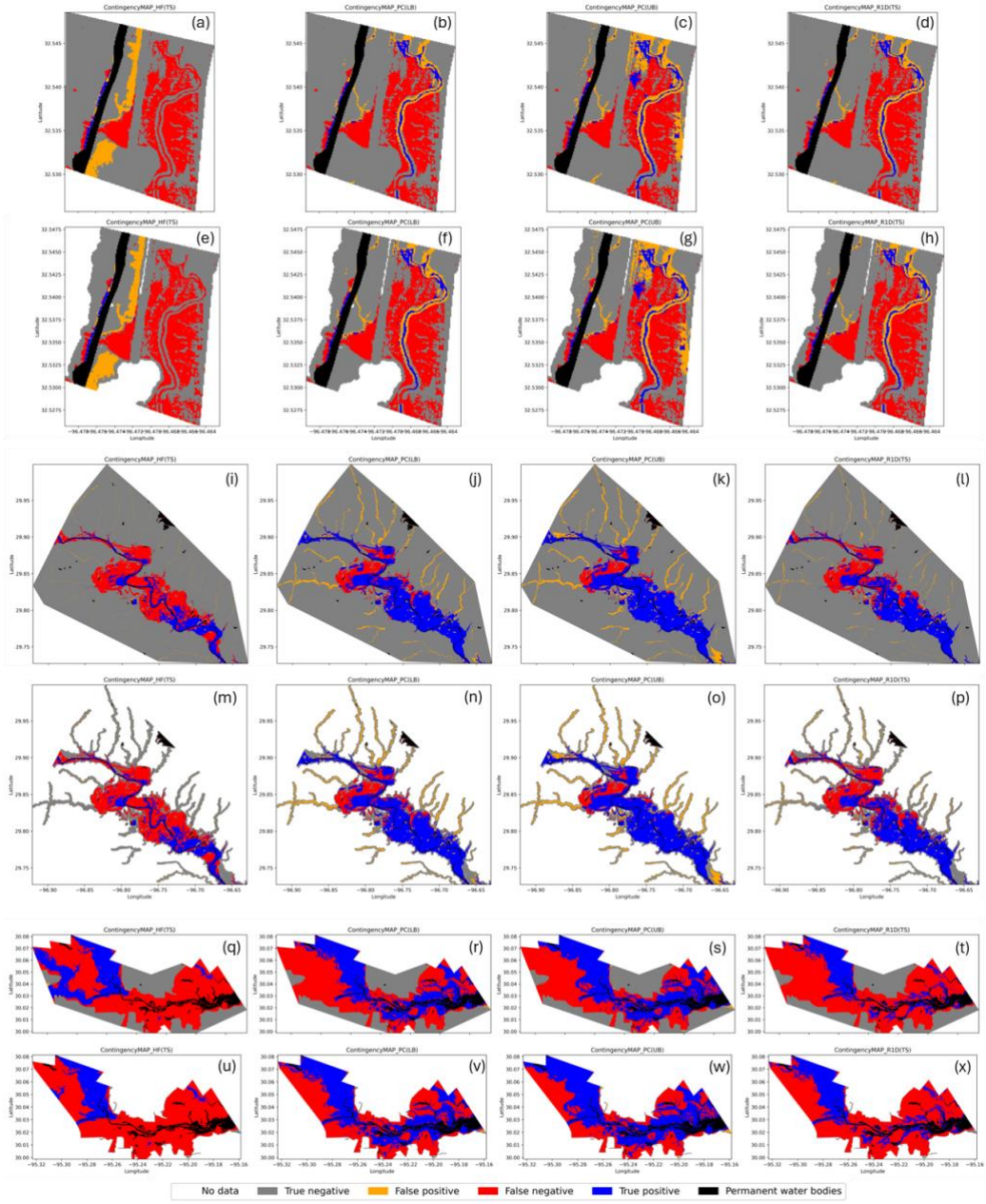


Figure 3. *Contingency maps for the four compared models for low RP event (a-h), intermediate RP event (i-p), and high RP event (q-x)*

The East Fork Trinity River has a long history of flooding, with major events impacting Dallas and Fort Worth, particularly in the early 20th century. An event of the magnitude of a 3-yr RP flood for the Trinity River was used to extract a case study within the East Fork Trinity River, representing a low RP event. The contingency maps (**Figure 3a-d**), generated using the Convex Hull method, indicate that the event-based R1D approach statistically matches the PC(LB) in terms of accuracy (0.72) as the models with the top performance (**Figure 3a & 4a**), slightly outperforming the accuracies of PC(UB) and HAND (0.69 and 0.68, respectively). It is worth mentioning that the identical performance of the event-based R1D and the PC(LB) in Figure 4a

is due to the limited discrete discharge simulation within the R1D library. When looking at the traced-to-pre-canned AOI approach (**Figure 3e-h**), a comparable performance pattern emerges across all four models, with a slight reduction within the accuracy magnitude. This traced-to-pre-canned approach isolates differences in model fidelity, allowing the performance assessment to focus solely on the quality of inundation predictions rather than on discrepancies in model coverage.

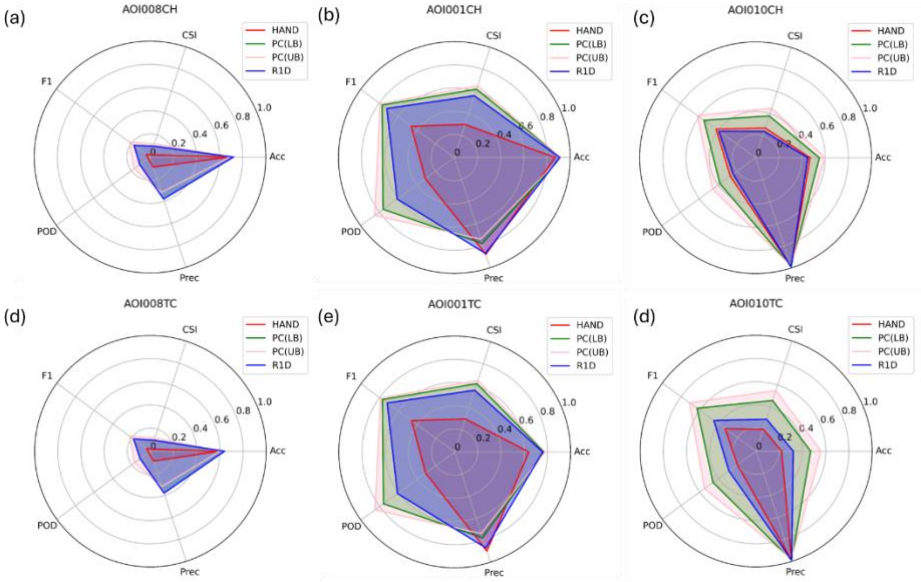


Figure 4. *Radar charts for the performance metrics of the 4 explored models for 3 ranges of return period (a) low RP (b)intermediate RP (c) high RP for the Convex Hull method (top panel) and traced to pre-canned FIM method (bottom panel)*

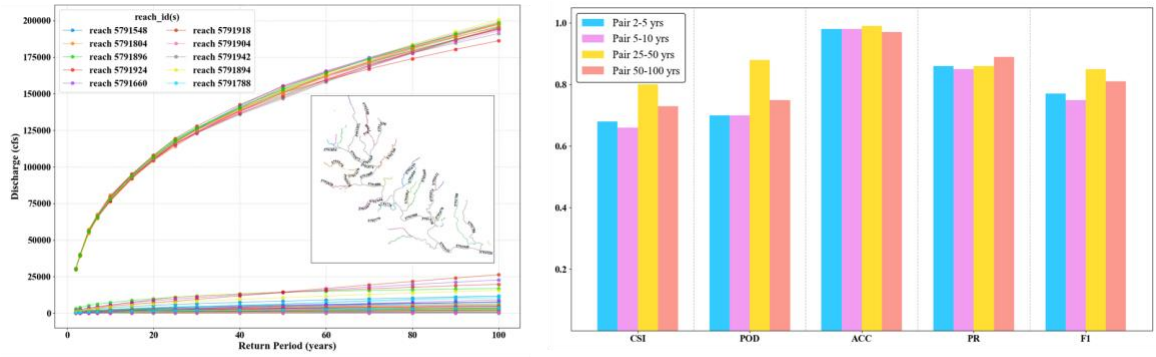
For the intermediate RP event, we have the Lower Colorado-Cummins watershed, where both severe droughts as well as extreme precipitation events have been taking place more frequently in recent years. The captured event, which occurred on August 29th, 2017, represents a 42-yr RP flood. The contingency maps produced through the Convex Hull method (**Figure 3i-l**), show that HAND underpredicts, with significant areas of flood underprediction along the main stem of the river, with an accuracy of 0.865. The PC-FIMs seem to outperform all other methods, with the accuracies of PC(UB) and PC(LB) almost identical (0.917 and 0.915, respectively). Although the event-based R1D-FIM performs relatively better than HAND (**Figure 4b**), it is still outperformed by the PC approach, despite having a similar performance at the downstream of the area of interest. Looking into the traced-to-pre-canned AOI approach (**Figure 3m-p**), a similar pattern of performance is repeated across the four models. This similarity of performance is portrayed within the radar plot in Figure 4b, where the shape of the radar and order of models are almost identical. Interestingly, when looking at the precision, HAND (0.897) outperforms all other models (0.776, 0.739, and 0.868) due to the underprediction of flood inundation extent within the lower order streams, which is due to bias in the NWM discharge prediction.

The West Fork of the San Jacinto River watershed has experienced significant flooding, particularly during Hurricane Harvey in 2017 and the October 1994 flood. The 2017 event produced the highest water levels ever recorded on the West Fork, surpassing the previous record set in 1994. The estimated 62-yr RP flood event, which took place on August 27th, 2017, is utilized here as our case study for the high RP event. The contingency maps generated (Figure 3q-x) using the Convex Hull method indicate that the PC(UB) and PC(LB) models have the highest accuracy (0.602 and 0.549, respectively) while HAND (0.468) slightly outperforms the

event-based R1D (0.452). Although both HAND and event-based R1D underpredict inundation extents along the main channel and particularly the downstream areas, HAND still outperforms the event-based R1D due to the discontinuity within the R1D sub-model library that results in leaving out an entire tributary out of the flood mapping at the upstream region. However, Figure 4f shows that the penalization of discontinuity within the R1D is eliminated when using the traced-to-pre-canned AOI approach, which restores the capabilities of R1D as a more accurate model than HAND. It is worth noting that the precision metric is almost identical for all models due to the disproportionality of wet to dry pixels within the benchmark, resulting from the severity of the high RP event.

5.1. Deep Learning Derived Intermediate FIMs

Figure 5 summarizes the evaluation of the DL-derived FIMs. Figure (a) demonstrates Discharge–return period curves for selected NWM reaches. These discharge values are computed for generated interpolated flood extents via FIM-BridgeNet. Upper curves represent mainstem reaches with larger flows, while lower ones correspond to smaller tributaries. Figure (b) shows the performance of the hybrid DL model across RP pairs, that 5–10 yr interpolation yields the most consistent metrics. These metrics were calculated for the hybrid DL model-projected upper threshold RP FIM against the true PC-FIMs for each interval brackets. The 25–50 yr pair achieves higher POD and F1, indicating better detection but slightly reduced precision. The 50–100 yr pair peaks in precision but dips in CSI. Minor variations in a second 5–10 yr run reflect model randomness.



(a) Synthetic FIMs discharge (cfs) **(b) DL Model Performance**

Figure 5. *Evaluation of Synthetic FIMs Discharges (a) and Deep Learning Model Performance (b) DL Model Performance*

Figure 6 demonstrates the capability of FIM-BridgeNet to produce intermediate FIMs for return periods (RPs) not included in the original inputs. As illustrated in Figures 6a-e, the generated maps for the 3-, 15-, 20-, 30-, and 70-year RP events, respectively, exhibit a clear and continuous spatial expansion of inundation areas with increasing RP. These outputs visually confirm that the model captures the expected progression in flood extent without introducing abrupt spatial artifacts between the known intervals.

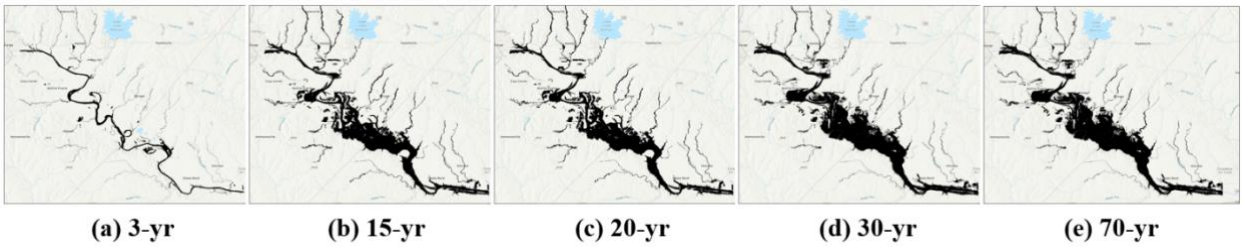


Figure 6. *Intermediate flood inundation maps generated by FIM-BridgeNet for return periods of (a) 3-year, (b) 15-year, (c) 20-year, (d) 30-year, and (e) 70-year. These maps fill the temporal gaps between standard pre-canned intervals, capturing smooth spatial progression of flood extent with increasing magnitude.*

5.2. Study Limitations

Several limitations affected the scope and generalizability of this study. Despite using a structured evaluation framework, analysis was constrained by limited benchmark datasets, sparse PC-FIM coverage, and a selected set of RP (3.5-, 37-, and 64-year), which restricted flood event diversity. Spatial discontinuities in PC-FIMs and their focus on higher-order streams limited consistent comparisons across sites and hindered evaluation in smaller tributaries, especially against HAND-FIMs. Additionally, uncertainty in NWM discharge inputs introduced bias into HAND and R1D performance metrics, with only basic temporal adjustments applied. From a computational standpoint, GPU memory constraints required splitting the study area and processing return-period pairs in parallel across multiple jobs. Input rasters were tiled into small patches for efficient training, validation, and testing. The hybrid ResUNet model, with attention and optical flow warping, increased runtime to ~23 hours due to deeper architecture and complex design. Hyperparameter tuning was necessary to balance warp strength and learning rate. Inference remained the most resource-intensive step, with sliding window predictions consuming substantial GPU memory despite pipeline-level parallelization.

6. Conclusion

This study analyzed the performance of multiple flood inundation mapping (FIM) approaches, including HAND-FIMs, pre-canned (PC) FIM libraries, and event-based Ripple1D (R1D) FIMs. Our results demonstrate considerable variability with an overall greater accuracy by PC-FIMs across return period (RP) ranges, followed by R1D-FIMs. HAND-FIMs consistently underpredicted due to limitations in NWM streamflow estimates and spatial aggregation across hydrologic units. Given the aforedescribed limitations in the analysis and relatively small number of case studies, we cannot conclude a consistent advantage of any of the FIM prediction approaches. To address the coarse RP intervals of PC-FIMs, we introduced FIM-BridgeNet, a hybrid deep learning model combining an attention ResUNet with optical flow warping, to generate intermediate flood extents. These outputs were post-processed into water depth and discharge using FwDET and NWM-based rating curves. The model demonstrated robust spatial continuity and accuracy across low, intermediate, and high RP ranges, supporting its value as a scalable solution for intermediate FIM generation. Future work will focus on extending benchmark coverage and improving RP diversity. A bidirectional DL + image warping framework with consistency-based loss will be developed to refine predictions across RP triplets. This will improve model scalability and computational efficiency, supporting large-scale implementation and a future methodological publication.

Acknowledgment

The authors gratefully acknowledge the support and guidance, and data provided by Dr. Dinuke Munasinghe, Dr. Dipsikha Devi, Dr. Anupal Baruah, Dr. Dan Tian, and Supath Dhital from the Surface Dynamics Modeling Lab (SDML) at The University of Alabama. We also extend our sincere thanks to our academic advisors, Dr. Hamed Moftakhari (The University of Alabama), Dr. Patrick Ray (University of Cincinnati), Dr. David F. Muñoz (Virginia Tech), and Dr. Fangni Lei (University of Connecticut) whose broader support and encouragement have contributed to our academic development. We sincerely thank Parvaneh Nikrou and Francisco Gomez for their excellent organization of the Summer Institute and for making it a truly rewarding experience.

This research was supported by the Cooperative Institute for Research to Operations in Hydrology (CIROH) under award NA22NWS4320003 from the NOAA Cooperative Institute Program. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA.

Supplementary Materials

1. https://github.com/NOAA-OWP/inundation-mappin
2. ghttps://github.com/NGWPC/flows2fim
3. https://stac2.dewberryanalytics.com/stacmap

References

[1] B. Merz *et al.*, "Causes, impacts and patterns of disastrous river floods," *Nature Reviews Earth & Environment,* vol. 2, no. 9, pp. 592-609, 2021.

[2] J. Buszta, K. Wójcik, C. A. Guimarães Santos, K. Kozioł, and K. Maciuk, "Historical Analysis and Prediction of the Magnitude and Scale of Natural Disasters Globally," (in en), *Resources,* vol. 12, no. 9, p. 106, 2023-09-05 2023, doi: 10.3390/resources12090106.

[3] C. Kousky, H. Kunreuther, M. LaCour-Little, and S. Wachter, "Flood Risk and the U.S. Housing Market," (in en), *Journal of Housing Research,* vol. 29, no. sup1, pp. S3-S24, 2020-12-09 2020, doi: 10.1080/10527001.2020.1836915.

[4] J. Tyler, A.-A. Sadiq, and D. S. Noonan, "A review of the community flood risk management literature in the USA: lessons for improving community resilience to floods," (in en), *Natural Hazards,* vol. 96, no. 3, pp. 1223-1248, 04/2019 2019, doi: 10.1007/s11069-019-03606-3.

[5] "International Disaster Database."

[6] A. B. Smith, "2023 U.S. Billion-dollar weather and climate disasters in historical context," in *104th Annual AMS Meeting 2024*, 2024-01-01 2024, vol. 104, p. 428624.

[7] J. Johnson, P. Ruess, and J. Coll, "OPERA–Operational Platform for Emergency Response and Awareness: Reimagining Disaster Alerts," *National water center innovators program summer institute report,* vol. 4, no. 97, p. 2016, 2016.

[8] J. M. Johnson, J. M. Coll, P. J. Ruess, and J. T. Hastings, "Challenges and Opportunities for Creating Intelligent Hazard Alerts: The "FloodHippo" Prototype," (in en), *JAWRA*

Journal of the American Water Resources Association, vol. 54, no. 4, pp. 872-881, 08/2018 2018, doi: 10.1111/1752-1688.12645.

[9] *Mapping the Zone: Improving Flood Map Accuracy*. Washington, D.C.: National Academies Press, 2009.

[10] C. Huang, Y. Chen, S. Zhang, and J. Wu, "Detecting, extracting, and monitoring surface water from space using optical sensors: A review," *Rev. Geophys.*, vol. 56, no. 2, pp. 333–360, 2018.

[11] G. Salmoral, M. Rivas Casado, M. Muthusamy, D. Butler, P. P. Menon, and P. Leinster, "Guidelines for the use of unmanned aerial systems in flood emergency response," *Water*, vol. 12, no. 2, p. 521, 2020.

[12] S. Afshari *et al.*, "Comparison of new generation low-complexity flood inundation mapping tools with a hydrodynamic model," (in en), *Journal of Hydrology,* vol. 556, pp. 539-556, 01/2018 2018, doi: 10.1016/j.jhydrol.2017.11.036.

[13] H. McGrath, J.-F. Bourgon, J.-S. Proulx-Bourque, M. Nastev, and A. Abo El Ezz, "A comparison of simplified conceptual models for rapid web-based flood inundation mapping," (in en), *Natural Hazards,* vol. 93, no. 2, pp. 905-920, 09/2018 2018, doi: 10.1007/s11069-018-3331-y.

[14] C. D. Rennó *et al.*, "HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia," (in en), *Remote Sensing of Environment,* vol. 112, no. 9, pp. 3469-3481, 09/2008 2008, doi: 10.1016/j.rse.2008.03.018.

[15] X. Zheng, D. R. Maidment, D. G. Tarboton, Y. Y. Liu, and P. Passalacqua, "GeoFlood: Large-Scale Flood Inundation Mapping Based on High-Resolution Terrain Analysis," (in en), *Water Resources Research,* vol. 54, no. 12, 12/2018 2018, doi: 10.1029/2018wr023457.

[16] D. R. Maidment, "Conceptual Framework for the National Flood Interoperability Experiment," (in en), *JAWRA Journal of the American Water Resources Association,* vol. 53, no. 2, pp. 245-257, 04/2017 2017, doi: 10.1111/1752-1688.12474.

[17] J. M. Johnson, D. Munasinghe, D. Eyelade, and S. Cohen, "An integrated evaluation of the National Water Model (NWM)–Height Above Nearest Drainage (HAND) flood mapping methodology," (in en), *Natural Hazards and Earth System Sciences,* vol. 19, no. 11, pp. 2405-2420, 2019-11-04 2019, doi: 10.5194/nhess-19-2405-2019.

[18] J. Teng, A. J. Jakeman, J. Vaze, B. F. Croke, D. Dutta, and S. Kim, "Flood inundation modelling: A review of methods, recent advances and uncertainty analysis," *Environmental modelling & software,* vol. 90, pp. 201-216, 2017.

[19] A. T. Murphy, B. Gouldby, S. J. Cole, R. J. Moore, and H. Kendall, "Real-time flood inundation forecasting and mapping for key railway infrastructure: a UK case study," *E3S Web Conf.*, vol. 7, p. 18020, 2016, doi: 10.1051/e3sconf/20160718020.

[20] T. Cox *et al.*, "Real-time flood impacts mapping technical report," Environment Agency - Flood and Coastal Erosion Risk Management Research and Development Programme, Bristol, BS1 5AH, Technical Report SC120023/R1, 2019 2019. Accessed: 2025-07-12 14:32:18.

[21] W. Wang, Q. J. Wang, R. Nathan, and C. Velasco-Forero, "Rapid prediction of flood inundation by interpolation between flood library maps for real-time applications," *J. Hydrol.*, vol. 609, p. 127735, Jun. 2022, doi: 10.1016/j.jhydrol.2022.127735.

[22] S. Cohen *et al.*, "Estimating Floodwater Depths from Flood Inundation Maps and Topography," (in en), *JAWRA Journal of the American Water Resources Association,* vol. 54, no. 4, pp. 847-858, 08/2018 2018, doi: 10.1111/1752-1688.12609.

[23] D. T. Lee and B. J. Schachter, "Two algorithms for constructing a Delaunay triangulation, *Int. J. Comput. Inf. Sci.*, vol. 9, no. 3, pp. 219–242, Jun. 1980, doi: 10.1007/bf00977785.

[24] Y. Wang, Y. Li, and A. B. Limaye, "Bridging Gaps in Satellite Observations of River and Delta Landscapes Using Image Warping," (in en), *Water Resources Research,* vol. 61, no. 5, 05/2025 2025, doi: 10.1029/2024wr039854.

[25] C. A. F. Do Lago, M. H. Giacomoni, R. Bentivoglio, R. Taormina, M. N. Gomes, and E. M. Mendiondo, "Generalizing rapid flood predictions to unseen urban catchments with conditional generative adversarial networks," (in en), *Journal of Hydrology,* vol. 618, p. 129276, 03/2023 2023, doi: 10.1016/j.jhydrol.2023.129276.

[26] J. Hofmann and H. Schüttrumpf, "floodGAN: Using Deep Adversarial Learning to Predict Pluvial Flooding in Real Time," (in en), *Water,* vol. 13, no. 16, p. 2255, 2021-08-18 2021, doi: 10.3390/w13162255.

[27] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," 2018 2018, doi: 10.48550/ARXIV.1804.03999.

[28] J. D. Bales and C. R. Wagner, "Sources of uncertainty in flood inundation maps," (in en), Journal of Flood Risk Management, vol. 2, no. 2, pp. 139-147, 06/2009 2009, doi: 10.1111/j.1753-318x.2009.01029.x.

[29] S. Marta, Planet Imagery Product Specifications, Planet Labs, San Francisco, CA, USA, 2018, pp. 91–170.

[30] E. S. A. (ESA), "Sentinel-1 SAR Level-1 Ground Range Detected (GRD) products," 2014.

[31] NOAA, Aerial Imagery Data Viewer, National Oceanic and Atmospheric Administration, 2022. [Online]. Available: https://storms.ngs.noaa.gov/. [Accessed: Jul. 16, 2025].

[32] Q. Yang, X. Shen, E. N. Anagnostou, C. Mo, J. R. Eggleston, and A. J. Kettner, "A high-resolution flood inundation archive (2016–the present) from Sentinel-1 SAR imagery over CONUS," Bull. Amer. Meteorol. Soc., vol. 102, no. 5, pp. E1064–E1079, May 2021, doi: 10.1175/BAMS-D-19-0319.1.

[33] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. SMC-9, no. No. 1, pp. 62–66, 1975

[34] D. Tian, H. Liu, L. Wang, S. Cohen, and T. Mandal, "RS-Floodxdepth: Enhancing remote sensing-derived flood extent and estimating flood depth using a hydrologically guided region-growing method and high-resolution DEMs," *SSRN*, [Online]. Available: https://ssrn.com/abstract=5336977 or http://dx.doi.org/10.2139/ssrn.5336977.

[35] A. Baruah *et al.*, "FIMserv v.1.0: A tool for streamlining Flood Inundation Mapping (FIM) using the United States operational hydrological forecasting framework," *Environ. Model. Softw.*, vol. 192, p. 106581, Aug. 2025, doi: 10.1016/j.envsoft.2025.106581.

[36] D. Devi *et al.*, "A Framework for the Evaluation of Flood Inundation Predictions Over Extensive Benchmark Databases," Jun. 27, 2025, *Wiley*. doi: 10.22541/essoar.175105672.23211828/v1.

Chapter 2:
Scoring Rule-Based LSTM Models for Flow Forecasting in the NextGen Framework

Mohamad Ali Farmani[1], Arman Oliazadeh[2], Habtamu Tamiru[3], Fatemeh Yavari[4], Amobichukwu Amanambu[5*], and Jonathan M. Frame[6*]

[1] University of Arizona, Tucson, Arizona; farmani@arizona.edu

[2] University of Georgia, Athens, Georgia; arman.Oliazadeh@uga.edu

[3] University of Alabama, Tuscaloosa, Alabama; hdagne1@crimson.ua.edu

[4] The City University of New York, New York City, New York; fyavari@ccny.cuny.edu

[5*] University of Alabama, Tuscaloosa, Alabama; acamanambu@ua.edu

[6*] University of Alabama, Tuscaloosa, Alabama; jmframe@ua.edu

[*]Theme Leader

Abstract: This technical report introduces a novel deep learning approach for predicting surface runoff from distributed catchments. We developed a distribution-based loss function that utilizes probabilistic scoring rules to avoid explicit physics-based routing during training. A single long short-term memory model processes both static and dynamic inputs, trained on diverse gauged watersheds using a defined spatiotemporal batching method. The loss quantifies agreement between observed downstream streamflow and predicted upstream runoff distributions. We introduced an innovative machine learning based method for selecting the optimal spatiotemporal modeling strategy to capture hydrological signatures of basins. This framework efficiently transfers spatiotemporal information from gauged sites to contributing catchments, providing an operationally viable pathway for calibrating hydrological predictions within the Next Generation Water Observing System in the National Water Model.

---

1. Motivation

Accurate prediction of surface runoff from spatially distributed catchments is a basis of effective water-resources management and flood risk mitigation. Surface runoff integrates complex interactions among precipitation, land cover, soil properties, and topography, resulting in highly nonlinear and site-specific responses. Deep learning models, particularly Long Short-Term Memory (LSTM) networks, have emerged as powerful alternatives for modeling hydrological processes. By ingesting time series of meteorological forcings alongside static catchment attributes, such architectures can learn implicit representations of soil moisture dynamics, infiltration, and storage directly from data. Studies (e.g., [1], [2]) demonstrated that an LSTM trained on multiple basins achieves higher median Nash–Sutcliffe efficiencies than traditional conceptual models. Also, transfer learning between catchments can reduce the need for site-specific calibration while maintaining forecast accuracy [3], [4].

To ensure that the model not only fits historical observations but also generalizes new conditions. We will evaluate its performance using metrics that assess both accuracy and

consistency across the entire spatial domain, as well as considering a Rising and Falling Limb (RFL), as a new loss function, to capture the hydrologic signatures, following the methods of previous research [17]. Such a framework can transfer information effectively from subcatchments to neighboring catchments, ultimately supporting more informed flood forecasting and watershed planning without the prohibitive calibration effort of conventional approaches to feed into the NextGen framework and make it generalizable in hydrological modeling.

2. Objectives and Scope

The main objective is to develop an LSTM framework that predicts daily and hourly surface runoff simultaneously across multiple subcatchments converging at a common downstream catchment output. By sharing spatial training model parameters across all contributing catchments, the network controls inter-catchment similarities in hydrologic signatures. This multi-catchment training pattern reduces calibration difficulty and enables robust predictions that consider the upstream contribution of catchments as a whole system.

2.1. LSTM Model Performance for Streamflow Modeling with NextGen

We evaluated the proposed LSTM model based on the new training method and various loss functions. The goal of this step is to make sure that the model can capture the streamflow and produce reliable results by considering various measures of success (e.g., R^2). Then, LSTM was integrated into the NextGen framework to assess its performance in some selected basins. The goal is to utilize the LSTM model for calibration of the Next Generation Water Observing System (NextGen) Framework to make it applicable in diverse hydrological conditions.

2.2. Hydrologic Signature Analysis

Another objective of this project is to develop predictions of streamflow based on the quantified hydrologic signatures. Hydrologic signatures, such as flashiness, baseflow index, and flow duration characteristics [5], provide insights into catchment behavior. This approach involves generating a range of possible hydrologic signatures, rather than a single deterministic streamflow value, to better capture the inherent variability and catchment behavior in hydrological modeling. Hydrologic signatures will support more informed water resource management and decision-making by providing a comprehensive understanding behavior of a catchment even for future scenarios.

3. Previous Studies

We used the NextGen framework and built upon the project by the 2024 Summer Institute team, named "Probabilistic Streamflow Prediction Using the Model-Agnostic NextGen Framework" [6]. While their study focused on the Conceptual Functional Equivalent (CFE) model approach, we specifically investigated appropriate spatiotemporal batching methods and ensemble sizes in the training dataset for each batch, and developed a new probabilistic scoring rule loss function, RFL, to capture hydrologic signatures in our simulation.

4. Methodology

We trained a single LSTM model to capture the relationship between area-normalized flow and a concatenated vector of inputs containing (m) static attributes (θ) of the unique hydrologic response unit and (n) dynamic forcing variables (x_n):

$$X_t = [\theta_1, \theta_2, \dots, \theta_m, x_{1,t}, x_{2,t}, \dots, x_{n,t}] \quad (1)$$

We used backpropagation of a loss function to train the model to predict runoff at a hydrologic response unit (e.g., y_c at a contributing catchment) from the input vector:

$$y_c = f(X_t) \quad (2)$$

Some upstream catchments contributed to predicting streamflow at the gauged location q_g.

Batch Definition: In deep learning, the model is trained with batches (N_b) which are a subset of the total number of data samples (N). A batch (B) consists of a subset of data points:

$$B = \{X_{t_1}, X_{t_2}, \dots, X_{t_b}\} \quad (3)$$

where $t_1, t_2, \dots, t_b$ are consecutive time steps, and b is the batch size (usually several timesteps but can also be split by watersheds).

The total dataset consists of all time steps for all contributing catchments:

$$D = \{X_t | \, t \in \{1, 2, \dots, T\}\} \quad (4)$$

where T is the total number of time steps for all watersheds.

Batch Iteration and Epochs: During training, the model processes the dataset in smaller subsets called batches, as using the entire dataset at once would be computationally expensive. Each batch contains data from a subset of the total time steps (T) for all gauged watersheds, and the total number of batches in one epoch is:

$$N_b = \lceil T/b \rceil \quad (5)$$

where N_b is the total number of batches, b is the batch size (number of data points in each batch), and $\lceil \cdot \rceil$ denotes the ceiling function. For a watershed dataset, the total number of time steps (T) is the sum of the time steps for all gauged watersheds:

$$T = \sum_G \; T_G \quad (6)$$

where G is the total number of gauged watersheds, and T_G is the number of time steps for a single gauged watershed. The batch size (b) is defined as:

$$b = G_b \times T_b \quad (7)$$

where G_b is the number of gauged watersheds included in the batch, and T_b is the number of time steps for each gauged watershed in the batch.

This batching approach is implemented for computationally manageable subsets of the data, while still learning from the whole dataset. It also allows the model to generalize across diverse watersheds with varying numbers of gauged locations and time steps. We need to ensure that the scoring rule-based loss function can work with batches. For modeling, the size of available data should be considered for training the model as well. We will begin to do spatiotemporal batching based on a variety of catchment categories in order to include every single catchment cluster in every batch. The clustering of all used Camel catchments is represented in **Figure 1**.

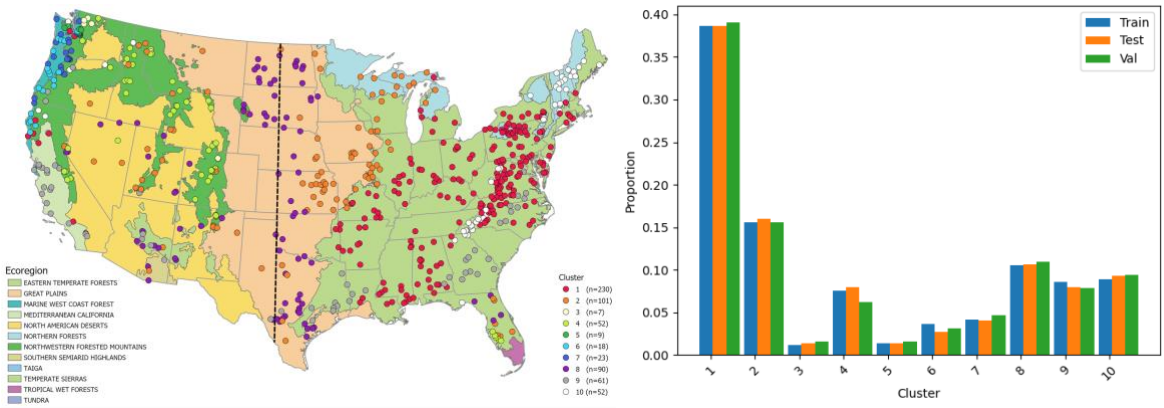


Figure 1. *(Left) Graphical clustered CAMELS catchments in the U.S. into ten clusters based on hydrological signatures with high spatial predictability. The clustering reveals that catchments with similar hydrological behavior (color coded) often exist across different regions. This highlights that while climate is a dominant driver overall, combinations of other catchment attributes can produce similar hydrological responses in geographically separate areas [16]. (Right) Clustering catchments for the new sampling method for the selected CAMELS basins*

Scoring Rule as a Loss Function: The scoring rule S(Q, P) was designed to evaluate the model's performance across the spatial and temporal distributions of watersheds and their upstream contributing catchments. Backpropagation was used to optimize the weights of the model by the scoring rule.

$$L(B) = S(Q_B, P_B) \quad (8)$$

where Q_B is observed distribution of downstream streamflow for all gauged watersheds in the batch, accounting for lag τ_g. P_B is predicted distribution of runoff contributions from all upstream catchments in the batch, accounting for lag τ_c. $S(Q_B, P_B)$ is a scoring rule quantifying the agreement between the observed (Q_B) and predicted (P_B) batch-level distributions.

- Observed Distribution (Q_B): The observed downstream distribution was constructed by aggregating the streamflow from all gauged watersheds (G_b) in the batch over time:

$$Q_B = p(y_g \mid g \in \{1, \ldots, G_b\}, t \in B, \tau_g) \quad (9)$$

where y_g is observed downstream streamflow at gauged watershed g, $g \in \{1, \ldots, G_b\}$ spans all gauged watersheds in the batch, $t \in B$ specifies the time steps included in the batch and τ_g represents the time lag for each gauged watershed.

- Predicted Distribution (P_B): The predicted upstream distribution was generated by aggregating the runoff contributions from all contributing catchments (c) for each gauged watershed (g):

$$P_B = \mathrm{p}(\mathrm{f}(x_c) \mid \mathrm{c} \in \mathrm{g}, \mathrm{g} \in \{1, \ldots, G_b\}, \mathrm{t} \in \mathrm{B}, \tau_c) \qquad (10)$$

where p(f(x_c)) represents the distribution of predictions for upstream catchments c of g gauged watersheds in the batch (G_b). t ∈ B denotes the time steps in the batch and τ_c represents the maximum lag for upstream runoff predictions.

The model was trained by cycling through many different gauged watersheds, with different numbers of contributing catchments, ranging from 10 to about 1000, an example is depicted in Figure 2.

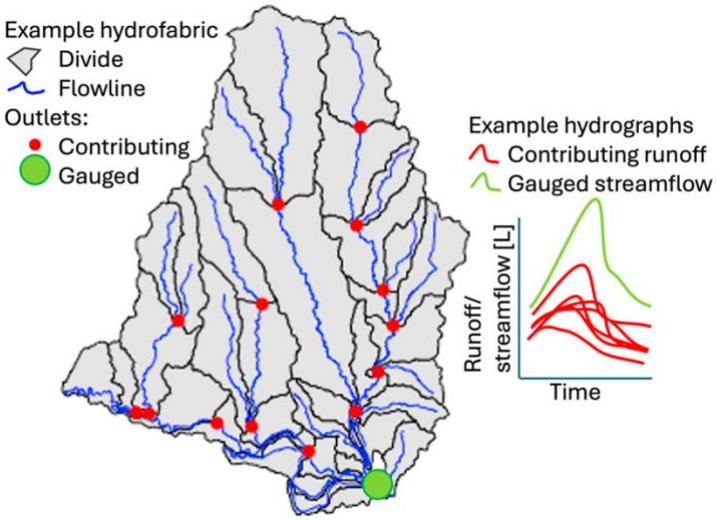


Figure 2. *Graphical depiction of a gauged watershed(green) with contributing upstream catchments (red)*

We also used two developed loss functions for our LSTM modeling including **F**low **D**uration **C**urve (FDC) and **R**ising **F**alling **L**imbs (RFL). The novel FDC divergence score, $d_{FDC}(P, Q)$ is a powerful tool for evaluating hydrological models by comparing the entire distribution of simulated and observed streamflow [7-9]. While the theoretical formulation of this score is based on an integral, for practical application with real-world time series data, the formulation is revisited as follows:

$$d_{FDC}(P, Q) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - \omega_j| - \frac{1}{2}\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \{|y_i - y_j| + |\omega_i - \omega_j|\} \qquad (11)$$

Whereas n: represents the total number of data points (e.g., days) in both the observed and simulated streamflow time series. y_i: denotes the i^{th}simulated streamflow value from the output of the model. ω_j: denotes the j^{th} observed streamflow value from the real-world data.$|y_i - \omega_j|$:

This term calculates the absolute difference between each simulated and observed value. $\sum_{i=1}^{n} \quad \sum_{j=1}^{n} \quad |y_i - \omega_j|$: This difference is computed for every possible pair of simulated and observed values. This formulation is based on a Monte Carlo estimate derived from a decomposition of the FDC divergence, which itself is equivalent to the Continuous Ranked Probability Score (CRPS) divergence [9].

To enhance the distribution-based scoring rules loss function is based on the "Rising/Falling Limbs" method, which specifically analyzes the dynamic characteristics of hydrographs within each batch. This method incorporates the concept of hydrological signatures derived from the rate of change in streamflow, which is indicative of basin response. For each hydrograph within a given batch, the process begins by calculating the slope for each discrete time interval. This slope represents the instantaneous rate of change of discharge with respect to time. So, a positive dQ/dt signifies a rising limb, indicating increasing flow, while a negative dQ/dt points to a falling limb with decreasing flow (eq. 12). Once these dQ/dt values are computed for all time intervals across all hydrographs in a batch, they form a new dataset. To create a distribution of these slopes, we follow a process conceptually like FDC but applied to the rates of change instead of discharge magnitudes. By transforming the hydrograph's temporal behavior into a distribution of its rates of change, the scoring rule was then used to quantify the agreement between the observed dQ/dt distribution and the predicted dQ/dt distribution.

$$d_{RFL}\left(\frac{\partial Q_{SIM}}{\partial t}, \frac{\partial Q_{OBS}}{\partial t}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|q_i - z_j| - \frac{1}{2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{|q_i - q_j| + |z_i - z_j|\} \quad (12)$$

Whereas n: represents the total number of data points (e.g., days divided into time intervals) in both the observed and simulated streamflow time series. q_i: denotes the i^{th} simulated streamflow rate over the specific time interval from the output of the model. z_i: denotes the i^{th} observed streamflow rate over the specific time interval from the real-world data.$|q_i - z_j|$: This term calculates the absolute difference between each simulated and observed value derived from streamflow rates. $\sum_{i=1}^{n} \quad \sum_{j=1}^{n} \quad |q_i - z_j|$: This difference is computed for every possible pair of simulated and observed rates. This formulation is based on a Monte Carlo estimate derived from a decomposition of the FDC divergence, which itself is equivalent to the Continuous Ranked Probability Score (CRPS) divergence.

4.1. The CAMELS observation data

We used a dataset developed by Frame et al. [10], which includes 515 Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) basins [11]. The dataset contains hourly and daily streamflow data from the United States Geological Survey (USGS) and forcing data (hourly total rainfall and potential evapotranspiration) from the National Land Data Assimilation System (NLDAS) for the CAMELS basins [12]. We further reduced the number of basins to 499 during model calibration based on the available data. CAMELS data include corresponding daily streamflow records from US Geological Survey (USGS) gauges and meteorological data.

4.2. The NOAA AORC forcing data

For our hydrological simulations, the NOAA Analysis of Record for Calibration (AORC) Version 1.1 dataset was utilized, providing hourly meteorological variables. These essential inputs, incorporated directly into our deep learning model, comprise: Total Precipitation (APCP_surface), Air Temperature (TMP_2maboveground), Specific Humidity (SPFH_2maboveground), Downward LongWave Radiation Flux (DLWRF_surface), Downward Short-Wave Radiation Flux (DSWRF_surface), and Pressure (PRES_surface). The selection of AORC ensures access to a high-quality, consistent, and spatially extensive set of forcing variables, providing a robust foundation for training and evaluating our distributed deep learning watershed model.

4.3. The NextGen Framework

The Next Generation National Water Model Prototype Framework is a continental-scale modeling framework that makes it simple to incorporate cutting-edge research (the best models and modeling techniques). The Office of Water Prediction (OWP) launched the Next Generation Water Resources Modeling Framework, or NextGen, an interagency effort that allows for regionally customized model formulations and addresses the present NWM performance. This framework promotes model interoperability, standardizes data and setup workflows, and eases the evaluation of diverse modeling approaches, with the NextGen NWM representing a specific configuration within it [13]. The enforcement of this conceptual model, in conjunction with the Basic Model Interface [14], offers an open-source, standards-based framework that enables modeling approaches to be regionally tailored for streamflow generation processes [15].

4.4. Code development

For this project, we developed our group's GitHub repository, which stands out as a main achievement(https://github.com/NWC-CUAHSI-Summer-Institute/NeuralNgen/tree/main). The main goal of this repository is to keep all codes and updates to ensure the reproducibility of our processes for future CUAHSI Summer Institute fellows. Though it is not a step-by-step guide, it provides a comprehensive record of the obstacles we encountered, delivering key insights to update the experience for those who follow.

4.5. Distribution-based training and testing splits

We hypothesized that using train/test splits based on the distribution of basin attributes would lead to models with more reliable generalization. To test this, we compared the resulting performance distributions from our proposed batching method against a standard cross-validation approach. The LSTM model was trained on 499 CAMELS basins; however, only a single catchment is presented here for demonstration. The training period spans 1999–2008, the validation period 1980–1989, and the testing period 2008–2014. Basins not used for training were reserved for validation and testing.

5. Results

Figure 3 the uncertainty envelopes for the cumulative distributions of KGE scores from two methods of splitting training and testing data. The blue curve representing the distributed splits shows a trend towards better performance, but more importantly, its uncertainty envelope is much tighter than that of the standard method. Based on this visual result, we failed to reject the hypothesis that using train/test splits based on the distribution of basin attributes provides a more reliable generalization.

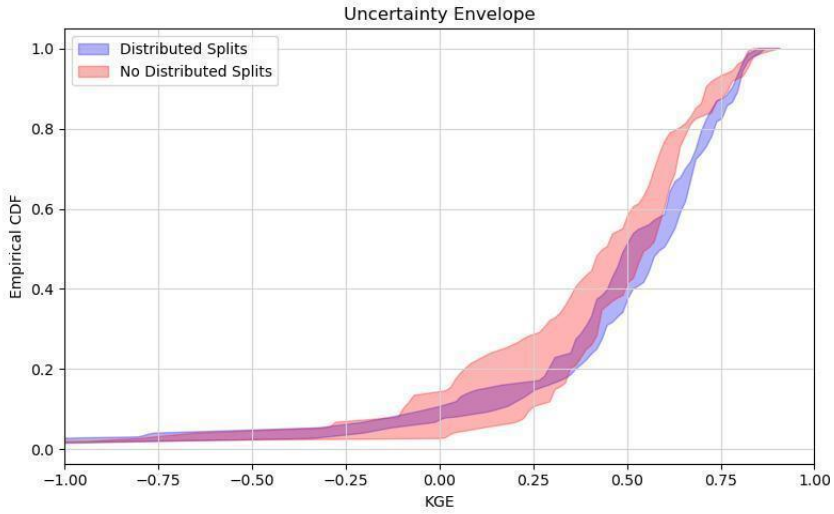


Figure 3. *Uncertainty envelopes of the empirical cumulative distribution functions (CDFs) for Kling-Gupta Efficiency (KGE) scores*

The results of a preliminary demonstration are shown in **Figure 4** for NextGen on a sample site and in **Figure 5** for comparing three methods for hydrologic signatures.

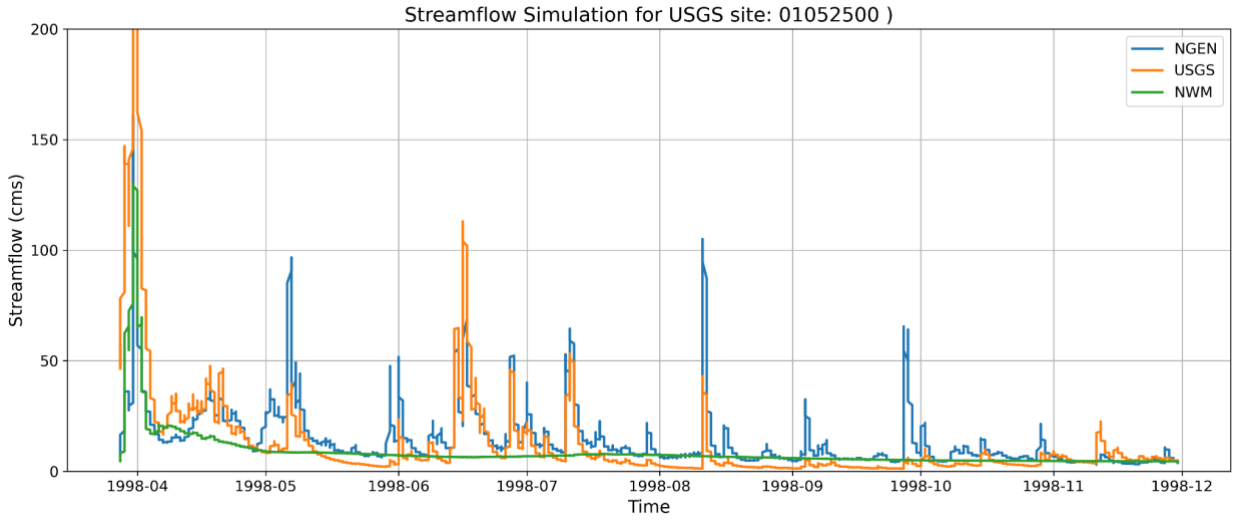


Figure 4. *NextGen and NWM simulated hydrograph at USGS gauge 01052500*

Figure 4 presents a time-series comparison of simulated and observed streamflow at USGS site 01052500 by using hourly data and the RFL as the loss function in LSTM to be used for

NextGen. Overall, the NextGen model demonstrates a strong ability to capture the timing of hydrological events. The simulated flow rises in response to storm events at nearly the same time as the observed flow, indicating that the runoff generation processes are well-timed. Similarly, the representation of baseflow and the shape of the recession limbs following most peaks align reasonably well with the observed data, suggesting the model simulates periods between events effectively. However, NextGen's primary weakness was the estimation of peak flow rates. Conversely, the model severely overestimates some events, generating a simulated peak that is more than double the corresponding observed flow. These errors in predicting the magnitude of high-flow events suggest that while the model is useful for capturing the general trend of the catchment behavior, it is not yet reliable for applications requiring accurate flood peak prediction for flood forecasting. Overall, NGEN offers better modeling of both peak and low-flow dynamics, whereas the NWM provides a smoother hydrograph. These differences highlight areas for potential recalibration to improve NWM peak responsiveness and refine NGEN's flood scaling to reduce its high-flow bias.

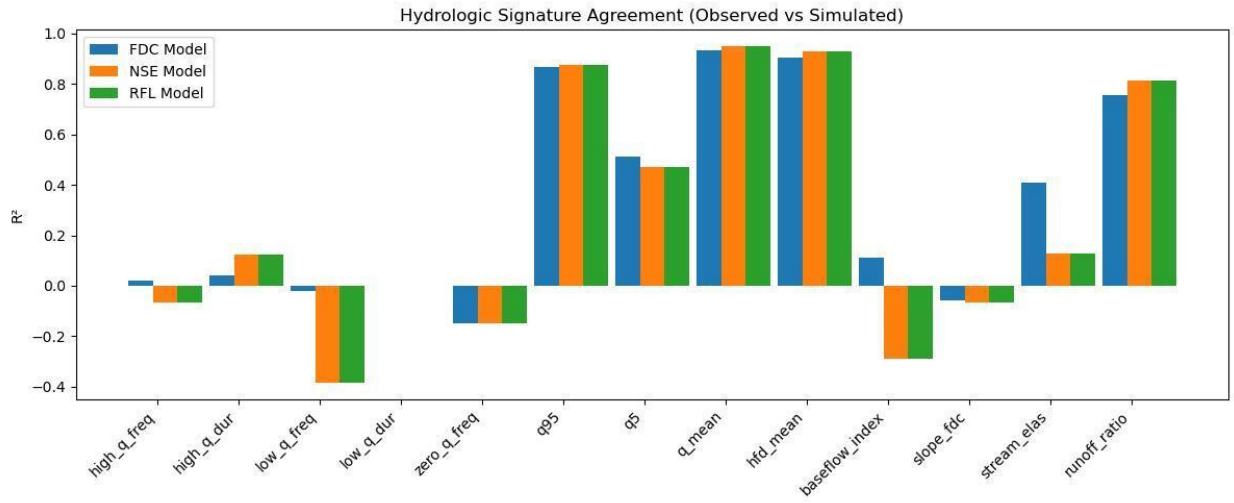


Figure 5. *Performance comparison of the FDC , NSE and RFL loss functions using the Coefficient of Determination between observed and simulated hydrologic signatures*

An analysis of the FDC, NSE and RFL models is presented in Figure 5, revealing a critical performance of various hydrologic signatures. While all models demonstrate considerable strength in simulating metrics like q_mean and hfd_mean with high R^2 values, they share a significant weakness, which is also seen in conceptual models and the NWM [17]. All of them failed profoundly to perform on the slope_fdc (slope of the flow duration curve), again, similar to conceptual models and the NWM [17]. The highly negative R^2 values for this signature indicate that their predictions of flow variability are inadequate, and there is a need to train the model with more new weights and stations.

Limitations of this work: While this study provides valuable insights, its conclusions should be considered in light of several limitations. First, the project was limited by significant time constraints, which narrowed the scope of our investigation and limited the number of simulations we could perform. This was compounded by the coding challenge inherent in translating our framework into a functional model; developing and debugging the necessary scripts was a time-intensive process. Furthermore, the theory's complexity required us to make certain simplifying assumptions to ensure the model was computationally feasible. Finally, a practical challenge was the process of running NextGen, our simulation platform. The high computational cost and long processing times for each simulation restricted our ability to conduct a more exhaustive sensitivity analysis across the entire

parameter space. Future research with dedicated computational resources and an extended timeline could build upon our work by addressing these constraints.

6. Conclusion

This work suggests a reliable approach that uses a distribution-based loss function to train a distributed deep learning watershed model. We present a computationally efficient method that moves data from gauged sites to contributing catchments without the need for explicit routing during training by utilizing probabilistic scoring rules. This approach offers a useful route for operational runoff prediction within the NextGen system and advances deep learning applications in hydrology.

Acknowledgements:

It is with great gratitude that the authors acknowledge the support of our academic advisors, Dr. Ali Behrangi and Dr. Guo-Yue Niu (The University of Arizona), Dr. Deepak Mishra (University of Georgia), Dr. Mesfin Mekonnen (The University of Alabama) and Dr. Naresh Devineni (The City University of New York), which has contributed to our professional development. Josh Cunningham of the Alabama Water Institute provided valuable guidance and support for running NGIAB. This research was supported by the Cooperative Institute for Research to Operations in Hydrology (CIROH) under award NA22NWS4320003 from the NOAA Cooperative Institute Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA.

References

[1] T. Lees, M. Buechel, B. Anderson, L. Slater, S. Reece, G. Coxon, and S. J. Dadson, "Benchmarking data-driven rainfall–runoff models in Great Britain: A comparison of long short-term memory (LSTM)-based models with four lumped conceptual models," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 10, pp. 5517–5534, 2021.

[2] J. Li, G. Wu, Y. Zhang, and W. Shi, "Optimizing flood predictions by integrating LSTM and physical-based models with mixed historical and simulated data," *Heliyon*, vol. 10, no. 13, 2024.

[3] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks," *Hydrol. Earth Syst. Sci.*, vol. 22, no. 11, pp. 6005–6022, 2018, doi: 10.5194/hess-22-6005-2018.

[4] F. Kratzert, B. Pflug, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using LSTM networks—Evaluating the significance of meteorological forcing data and exploring transfer learning capabilities," *Hydrol. Earth Syst. Sci.*, vol. 23, no. 1, pp. 65–84, 2019, doi: 10.5194/hess-23-65-2019.

[5] H. K. McMillan, "A review of hydrologic signatures and their applications," *Wiley Interdiscip. Rev. Water*, vol. 8, no. 1, e1499, 202.

[6] A. Sattari, R. Morovati, and H. Gholizadeh, "Data assimilation of USGS streamflow in the NextGen Framework probabilistic streamflow prediction using the model-agnostic NextGen framework," *National Water Center Innovators Program Summer Institute Report*, 2024.

[7] T. Lan, J. Zhang, H. Li, H. Zhang, X. Gong, J. Sun, Y. D. Chen, and C.-Y. Xu, "Flow duration curve prediction: A framework integrating regionalization and copula model," *J. Hydrol.*, vol. 647, p. 132364, 2025, doi: 10.1016/j.jhydrol.2024.132364.

[8] J. B. Swain and K. C. Patra, "Streamflow estimation in ungauged catchments using regional flow duration curve: Comparative study," *J. Hydrol. Eng.*, vol. 22, no. 7, Art. no. 04017010, 2017, doi: 10.1061/(ASCE)HE.1943-5584.0001509.

[9] J. A. Vrugt, "Distribution-based model evaluation and diagnostics: Elicitability, propriety, and scoring rules for hydrograph functionals," *Water Resour. Res.*, vol. 60, no. 6, 2024, doi: 10.1029/2023wr036710.

[10] J. M. Frame, F. Kratzert, D. Klotz, M. Gauch, G. Shalev, O. Gilon, et al., "Deep learning rainfall–runoff predictions of extreme events," *Hydrol. Earth Syst. Sci.*, vol. 26, no. 13, pp. 3377–3392, 2022.

[11] 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, 2017.

[12] K. E. Mitchell, D. Lohmann, P. R. Houser, E. F. Wood, J. C. Schaake, A. Robock, et al., "The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system," *J. Geophys. Res. Atmos.*, vol. 109, no. D7, 2004.

[13] F. L. Ogden et al., "The Next Generation Water Resources Modeling Framework: Open Source, Standards Based, Community Accessible, Model Interoperability for Large Scale Water Prediction," presented at the AGU Fall Meeting, Dec. 2021, Abstract H43D-01. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2021AGUFM.H43D..01O. Accessed: Jul. 15, 2025.

[14] S. D. Peckham, M. Stoica, E. Jafarov, A. Endalamaw, and W. R. Bolton, "Reproducible, component-based modeling with TopoFlow, a spatial hydrologic modeling toolkit," *Earth and Space Science*, vol. 4, no. 6, pp. 377–394, Jun. 2017.

[15] J. M. Johnson, S. Fang, A. Sankarasubramanian, A. M. Rad, L. Kindl da Cunha, K. S. Jennings, et al., "Comprehensive analysis of the NOAA National Water Model: A call for heterogeneous formulations and diagnostic model selection," *J. Geophys. Res. Atmos.*, vol. 128, no. 24, Art. no. e2023JD038534, 2023.

[16] F. U. Jehn, K. Bestian, L. Breuer, P. Kraft, and T. Houska, "Using hydrological and climatic catchment clusters to explore drivers of catchment behavior," *Hydrol. Earth Syst. Sci.*, vol. 24, no. 3, pp. 1081–1100, 2020.

[17] Frame, Jonathan M., et al. "Post-processing the national water model with long short-term memory networks for streamflow predictions and model diagnostics." *JAWRA Journal of the American Water Resources Association* 57.6 (2021): 885-905.

Chapter 3:
Towards Representing Pluvial Flooding within NOAA's NextGen Modeling Framework

Yogesh Bhattarai[1], Supath Dhital[2], and Samrin Sauda[3,] Mohamed Abdelkader[4*], Jonathan Frame[5*], Marouane Temimi [6*]

[1] Howard University; *yogesh.bhattarai@bison.howard.edu*

[2] The University of Alabama; *sdhital@crimson.ua.edu*

[3] Pennsylvania State University; *sss6318@psu.edu*

[4*] University of Iowa; *mohamed-abdelkader@uiowa.edu*

[5*] The University of Alabama; *jmframe@ua.edu*

[6*] Stevens Institute of Technology; *mtemimi@stevens.edu*

[*]Theme Leader

Abstract: Among natural disasters worldwide, flooding stands as one of the most catastrophic hazards. Changes in environmental conditions affect flood frequency and intensity. Relative occurrence of flood events is further intensified by anthropogenic land modifications, particularly urban expansion and the development of impermeable surfaces. Pluvial flooding poses challenges due to its site-specific characteristics and dynamic relationships between localized precipitation, topography, and subsurface elements. Currently, the Next Generation National Water Modeling Framework provides an operational framework for forecasting fluvial floods; however, its applicability at the local scale for pluvial flooding has not been thoroughly tested. This study assesses the NextGen's spatial reliability and predictive accuracy for localized flooding. The observed results show variation in Pluvial Flood Indices (PFI) across different catchment areas and land cover types. PFIs are found to be higher in small area catchments with urban land cover. The indices decrease as catchment area increases, and other land cover classes are included. However, PFIs suddenly increase when multiple urban areas are included in larger catchments. Observed findings provide benchmarks for pluvial flood detection within the NextGen framework, and the scalability of these indices across different catchment sizes and urban configurations will strengthen their operational utility.

---

1. Motivation

The motivation of this study is to address the missing pluvial component within the Next Generation National Water Resources Modeling Framework (NextGen) framework. A previous study demonstrates that pluvial flooding accounts for 87.1% of all flood-related claims examined within the National Flood Insurance Program (NFIP) database [1]. Pluvial flooding from intense rainfall overwhelms local drainage systems and poses a significant urban threat, particularly as human and environmental patterns continue to evolve. These localized events remain difficult to predict due to short lead times, complex topographic controls, and sub-catchment-scale variability [1]. NOAA's Next Generation Water Resources framework provides a scalable, flexible, and model-agnostic platform for large-scale hydrological applications using standardized

catchments of 3-15 km² [2]. In the present framework, NextGen focuses on stream-scale forecasting [3], but its ability to capture fine-scale urban flood signals is not clear. High-fidelity hydraulic models like Personal Computer Storm Water Management Model (PCSWMM) offer detailed urban simulations at block-scale resolution but are computationally intensive and typically limited to site-specific applications [4], [5].

Currently, no defined indices exist for quantifying pluvial floods within the NOAA NextGen framework. The indices from this study will provide benchmarks for detecting pluvial flood signals from NextGen. This approach will assess the spatial reliability and detection capabilities of the NextGen framework for pluvial flooding. We expect the results will guide the development of hybrid modeling strategies for more effective early warning systems for localized pluvial flooding events.

2. Objectives and Scope

The primary objective of this project is to investigate the effectiveness of NextGen in the detection of pluvial flood signals at the current hydrofabric standard (https://mikejohnson51.github.io/hyAggregate/). We define and test standardized indices across different catchments. The variation of the defined indices is evaluated across catchment areas, dominant land cover, and forcing data such as precipitation. The research questions guiding this investigation include:

a) To what extent can the NOAA NextGen-NWM framework be utilized to simulate and detect pluvial flooding events?

b) How does the spatial discretization of the NextGen hydrofabric influence the accuracy and sensitivity of pluvial flood forecasting?

By addressing these questions, the project aims to enhance the understanding of the application of NextGen's models in the localized flood event detection. More specifically, the following research objectives are addressed

RO1: We investigate the suitability of GIUH-derived runoff from CFE simulations within the NextGen framework for detecting pluvial flood signals at finer spatial resolutions.

RO2: We develop an index-based methodological framework to quantify pluvial floods and analyze how land cover characteristics and spatial scales influence the variability of pluvial flood detection using these indices.

3. Previous Studies

The National Water Model (NWM) is a comprehensive hydrologic modeling framework that emerged from the development efforts of the joint OWP (Office of Water Prediction)-NCAR (National Center for Atmospheric Research) development team to provide continental-scale high-resolution water modeling capability [6], [7]. The system uses the WRF-Hydro framework to integrate meteorological data and simulate water movement processes, including runoff, infiltration, groundwater flow, and streamflow across over 2.7 million stream reaches. Operating at high spatial resolution (1 km for land surface, 250 m for terrain routing) and providing forecasts from hourly to 30-day scales, the NWM combines land surface modeling, routing models, and real-time data assimilation to generate streamflow forecasts at the continental scale [8].

Building upon the NWM foundation, the NextGen framework is a flexible, open-source, and standards-based platform operating at the intersection of hydrologic, computational, and data sciences [2], [6]. NextGen is built upon a spatially derived network of catchments and flow paths that divide the U.S. landscape and river systems into discrete computational units. Each of these computational units can independently run hydrologic and/or hydraulic models and communicate with neighboring elements. The modular structure allows NextGen to deliver a unified, model-agnostic hydrologic forecast across the United States, aligning with the objectives of OWP. The primary strength of NextGen lies in fluvial flooding forecasts [9], [10]. NextGen predicts floods caused by river and stream overflow through its catchment-based modeling approach and streamflow routing capabilities. However, the framework currently lacks a comprehensive approach for localized pluvial flood detection. Pluvial floods caused by intense rainfall overwhelming local drainage systems in urban and low-lying areas requires selective modeling techniques focused on surface water accumulation rather than river network routing.

Indices and frameworks for assessing fluvial and coastal flood risk are well established and widely used in hazard mapping and risk assessment, such as the Federal Emergency Management Agency's (FEMA) Special Flood Hazard Area (SFHA) maps in the United States. In contrast, tools and indices specifically developed for pluvial flooding have historically been limited, with traditional approaches focusing on riverine and coastal flooding due to their clearer boundaries and longer data records. However, there is a growing recognition of the need to systematically address pluvial flood hazard [11], [12]. For example, a recent study proposed a pluvial flood index (PFI) that integrates precipitation, hydrological, and hydrodynamic processes based on pluvial flood hazard areas (PFHA). While comprehensive, this approach can be time-intensive to implement [13]. To address this, our study introduces simpler and faster indices that can be efficiently derived within the NextGen modeling framework, facilitating broader application and timely assessment of pluvial flood risk.

4. Methodology

In this study, we adopted the methodological framework represented in **Figure 1**. We develop an operational framework that integrates the local storm reports, insurance data for isolating pluvial flood events, runs the NextGen Conceptual Functional Equivalent model for the selected events, and determines the pluvial events based on the indices developed. The following subsections describe the data preparation and determination of the pluvial indices in detail:

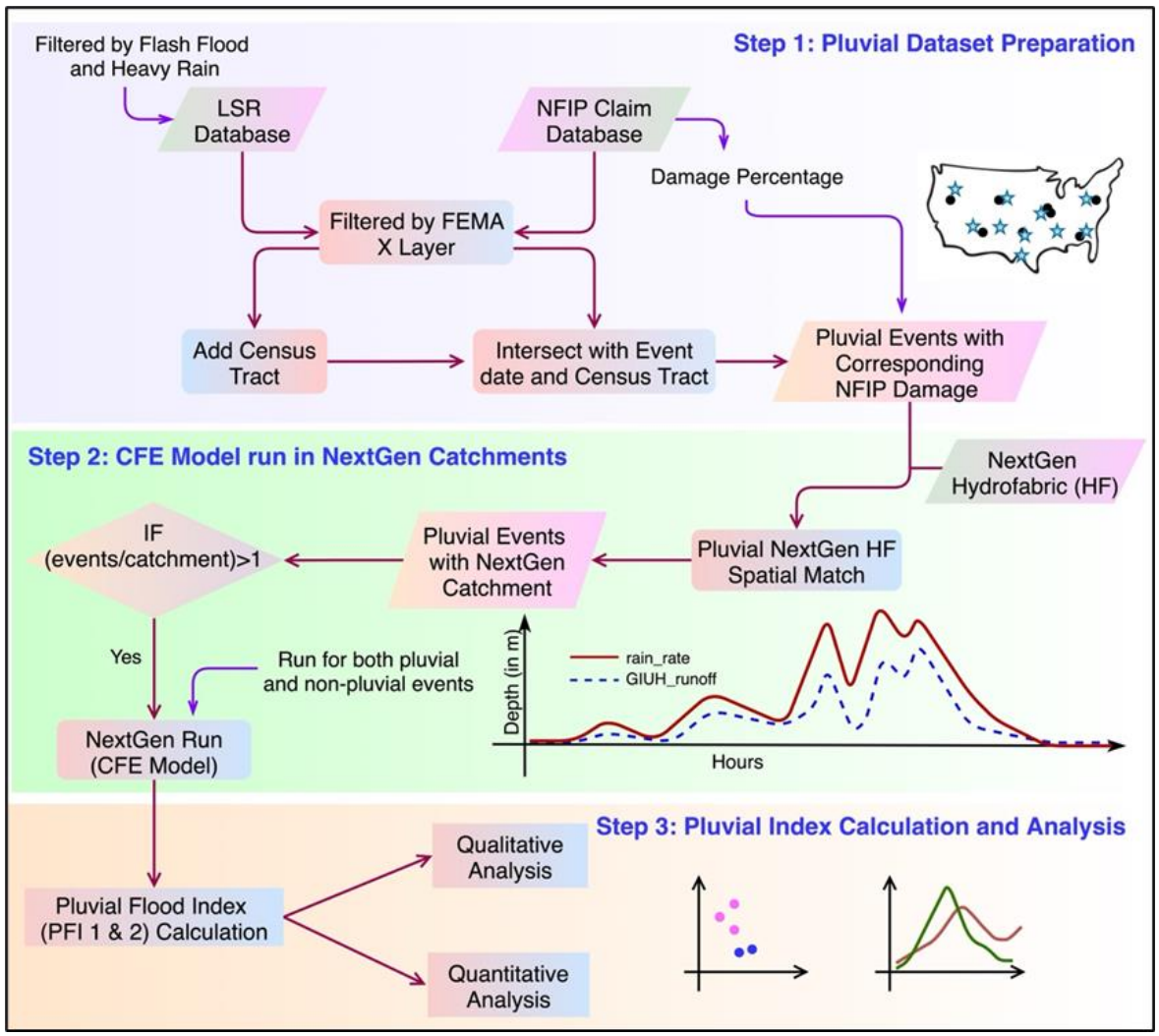


Figure 1. *Methodological framework adopted for the study. Step-1 defines the preparation of pluvial datasets from multiple sources, Step-2 includes the setup for the NextGen CFE module and simulation in selected catchments, and final Step-3 involves the visual verification and determination of volume and peak-based indices.*

4.1 Dataset

4.1.1. Local Storm Reports

The Local Storm Report (LSR) is a text-based product issued by the National Weather Service (NWS) that provides details on specific weather events, typically based on reports from external sources [14]. Covering the CONUS region from 2005 to 2024, this dataset includes more than 2.8 million reported storm events. It encompasses over 50 types of storms, such as various forms of flooding, tornadoes, lightning, and hail. The data is gathered through a range of channels, including law enforcement, storm spotters, media, and the general public. While the initial information is collected at the Local Weather Forecast Offices, it is later processed, quality-checked, and distributed in a standardized format for use by the public and decision-makers.

To identify pluvial flood events, LSRs were initially filtered to include only event types most indicative of pluvial flooding, specifically focusing on "flash flood" incidents. We filtered approximately 98,000 reports, which were then screened geographically using FEMA's X flood layer. This allowed us to focus on locations that fall outside the 100-year floodplain and are thus

more indicative of "true" pluvial flood occurrences. To enable spatial matching, census tract information was appended to each LSR event.

4.1.2. National Flood Insurance Claims Dataset

Flood insurance data were sourced from the Federal Insurance and Mitigation Administration (FIMA) National Flood Insurance Program (NFIP) Redacted Claims [15]. The NFIP Redacted Claims dataset covers over 2.5 million insurance claims dating back to 1970. For this analysis, we focused on over 189,000 claims filed between 2005 and 2021. The dataset provides important attributes such as date of loss, flood zone, census tract, building and contents coverage, and claim amount. The growing availability of NFIP claim records has made large-scale flood risk assessments increasingly feasible in the U.S. [1], [16]

NFIP insurance claims were first filtered to include only those filed from 2005 onward. To focus on claims most likely associated with pluvial flooding, we further refined the dataset using FEMA's X flood layer, which excludes properties located within the 100-year floodplain. The filtered claims were then matched to LSRs based on both the event date and census tract. Events that overlapped in both space and time, meaning they occurred within the same census tract and on the same date, were identified as pluvial flood events.

4.1.3. National Landcover Database

The National Land Cover Database (NLCD), developed by the U.S. Geological Survey (USGS) in collaboration with the Multi-Resolution Land Characteristics (MRLC) Consortium, provides a nationally standardized, 30-meter resolution raster inventory of land cover across the United States [17]. Using a modified Anderson Level II classification system, the NLCD distinguishes 20 land cover classes, including various vegetation types, development densities, agricultural uses, as well as water, ice, snow, and barren land. The database is derived from multi-date Landsat satellite imagery, enabling consistent analysis of land cover and change. In this study, we utilized the 2021 NLCD layer to determine the dominant land cover type within each catchment based on area.

4.2. Study Area

In this study, we utilize catchments from the NextGen hydro fabric across the entire United States. We select the catchments through the intersection of Local Storm Reports (LSR) events and National Flood Insurance Program (NFIP) data. This approach enables the isolation of high-confidence cases of pluvial flooding. We identify 595 distinct flooding events occurring between 2005 and 2021. From these events, we extract 251 unique NextGen catchments from the hydro fabric. These catchments were selected based on their occurrence within the compiled LSR events. Only catchments experiencing multiple flooding events during the study period were included in the final dataset (**Figure 2**).

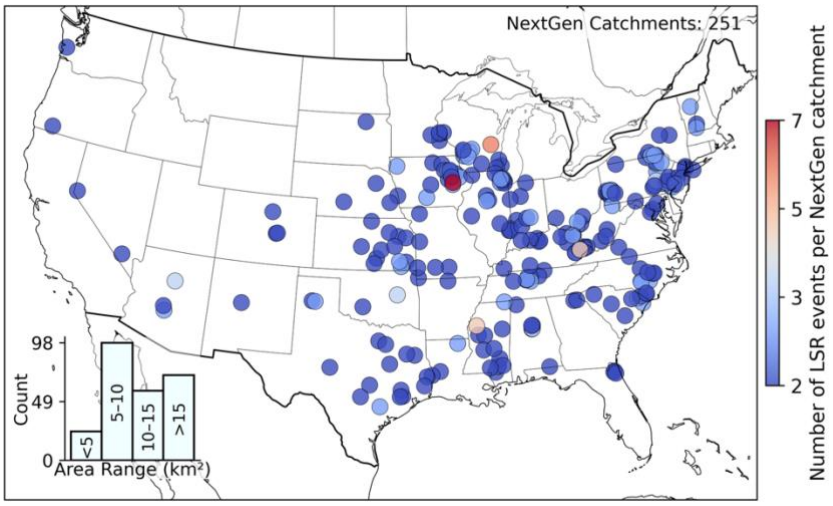


Figure 2. *Distribution of final LSR events corresponding to NFIP damage percent per NextGen catchment*

4.3. NextGen Framework

The NextGen Framework is a model-agnostic, standards-based software platform that enables explicit coupling of domain science models through sequential sharing of computed states and fluxes[18]. The NextGen framework facilitates water resources model interoperability, intercomparison, and testing of research hypotheses [19]. The framework relies on a specified hydrofabric, which delineates the hydrologic landscape into discrete catchments and flow paths. The NextGen hydrofabric, realized at a scale of 3–15 km² (although some catchments may fall outside this range due to local hydrographic constraints or network topology requirements) with a minimum channel reach length of 500 m, enables catchment-based aggregation of hydrometeorological forcings and supports model interoperability through its standards-based, modular architecture. This hydrofabric structure ensures spatial consistency for runoff generation and routing and is derived from the NHDPlus dataset [20], [21].

The National Water Center's CFE model is a conceptual hydrological model inspired by the WRF-Hydro physical processes that has been applied to estimate the volume of water flowing into rivers and streams following rain events[19]. The WRF-Hydro system has a 250m routing grid, making it computationally expensive. CFE conceptual representation retains core hydrological behaviors, including mass balance fidelity, with the same rainfall portioning scheme and nonlinear groundwater reservoir as the NWM [22]. The CFE model uses a Geomorphological Instantaneous Unit Hydrograph (GIUH) [23] for the surface runoff routing (GIUH_runoff), and lateral subsurface is handled via a Nash Cascade Model [24]. During rainfall events, potential evaporation is subtracted, and the remaining precipitation is partitioned into infiltration and surface runoff, which is routed using GIUH ordinates derived from watershed attributes and convolved with effective rainfall [22]. In this study, the NextGen CFE v1.0 model is simulated for the selected 251 catchments. This simulation includes pluvial as well as non-pluvial events to derive the indices and validate false alarms.

4.4. Pluvial flood detection approaches

In this study, we select qualitative and quantitative approaches for detecting the signal of pluvial flood from the NextGen framework. Integration of both approaches creates a more robust and interpretable analysis pipeline. For the qualitative analysis, we focus on visual verification of the

pluvial event based on the LSR database and CFE simulations. For LSR database events, we plot the time series of rainfall and corresponding GIUH runoff with LSR event markers. We validate the pluvial events by assessing the temporal alignment between GIUH runoff peaks and documented LSR events. We can confirm the accuracy of the methodology in capturing flood signals based on the close correspondence on **Figure 3**.

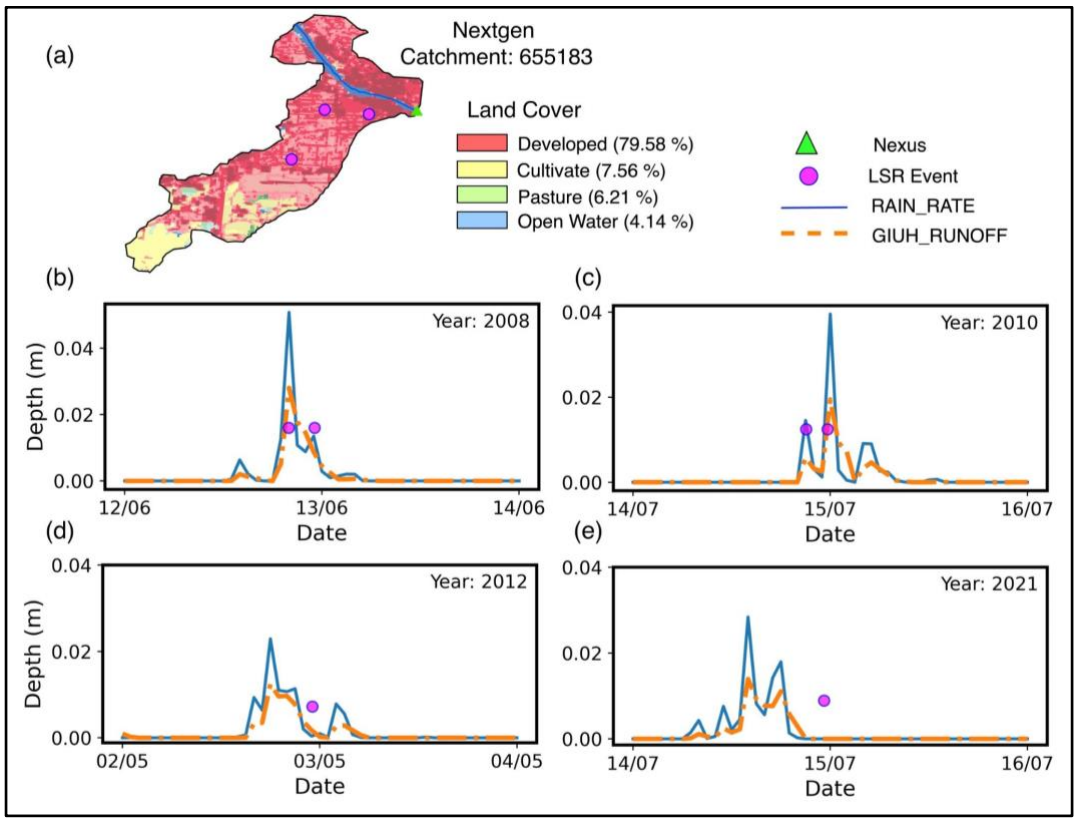


Figure 3. *Qualitative visualization of pluvial floods against the CFE simulation generated for the NextGen Catchment 655183. (a) Catchment area representing the dominant landcover to be developed (urban) followed by cultivation, pasture and open water. Time series plot obtained for the catchment (b) for year 2008 (c) for year 2010 (d) for year 2012 (e) for year 2021.*

For the quantitative analysis, we employ an index-based characterization approach using hydrological parameters from the CFE simulation, specifically the GIUH runoff and peak GIUH runoff, alongside rainfall information, including the cumulative and peak rainfall. Further, the duration between LSR events and GIUH initiation was used as a proxy to define the time span for indices calculations. We derive composite indices based on these parameters. These indices, powered by hydrological models, enable consistent, scalable, and automated flood detection and forecasting. We adopt Pluvial Flood Index-1 (PFI-1), defined as the ratio of accumulated GIUH runoff to accumulated rainfall, as in equation 1. We select Pluvial Flood Index-2 (PFI-2), defined as the ratio of peak GIUH runoff to peak rainfall within an 18-hour timeframe, as in equation 2. We chose 18 hours for volume determination to mimic NOAA's High Resolution Rapid Refresh (HRRR) forecasting time (**Figure 4**). For both indices, we go back 24 hours from the event occurrence timeline. From this 24-hour start time, we move forward in the time series. We mark the point where GIUH runoff values begin to rise from zero as "backtime" (Figure 3). We calculate the difference between the LSR event time and backtime and mark it as "simulation time (**Figure 4**). We determine the first quartile (Q1) and third quartile (Q3) for each simulation time observed across events. We define α_1 as the percentage ratio of GIUH runoff to rainfall at

the particular backtime. For additional alpha determination, we find the number of hours between the median and Q3. For the first run case, we select every event and determine values for simulation time and alphas. For the final PFI determination run, we fix α based on the median after removing outliers from the concatenation of α_1 to α_n.

$$PFI - 1 = \frac{Accumulated\ Runoff}{Accumulated\ Rainfall} \quad (1)$$

$$PFI - 2 = \frac{Peak\ Runoff}{Peak\ Rainfall} \quad (2)$$

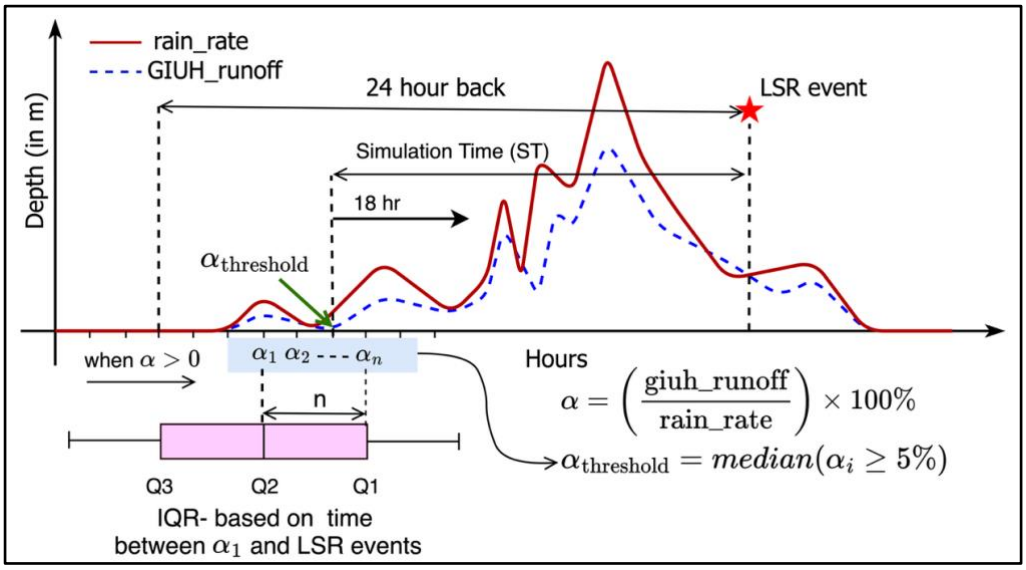


Figure 4. *Diagrammatic representation of evaluation metrics for the pluvial flood event detection*

5. Results

Figure 5 provides a statistical assessment of α and simulation time distributions, along with an evaluation of pluvial flood indices and damage percentages quantified across different area bins. **Figure 5a** shows a histogram representing the distribution of α values. We observe a right-skewed distribution with a high concentration of α values in the lower range (approximately below 20), followed by an extended distribution towards higher values (up to approximately 90). We find that most events possess relatively smaller α values, with fewer instances of large α values representing extreme conditions within the selected events. The observed distribution suggests that catchments generally have moderate to high infiltration capacities that limit runoff in most scenarios. However, critical or extreme scenarios remain with limited infiltration capacity or higher antecedent moisture conditions, resulting in significant runoff production. The distribution highlights that while severe runoff-producing events may be infrequent, their magnitude could be substantial when they occur. For further analysis, we set the α value to 11.2 based on the median of observed values. **Figure 5b** illustrates a histogram depicting the distribution of simulation times in hours. For simulation time, we observe a left-skewed distribution with concentration of simulation durations towards longer time frames, notably around 20–24 hours. We find the median simulation time to be 19 hours, approximately fitting the time window (18 hours) selected for PFI determination. The selected alignment suggests that the chosen 18-hour forecasting window effectively captures the onset of runoff and fulfills pluvial operational forecasting needs.

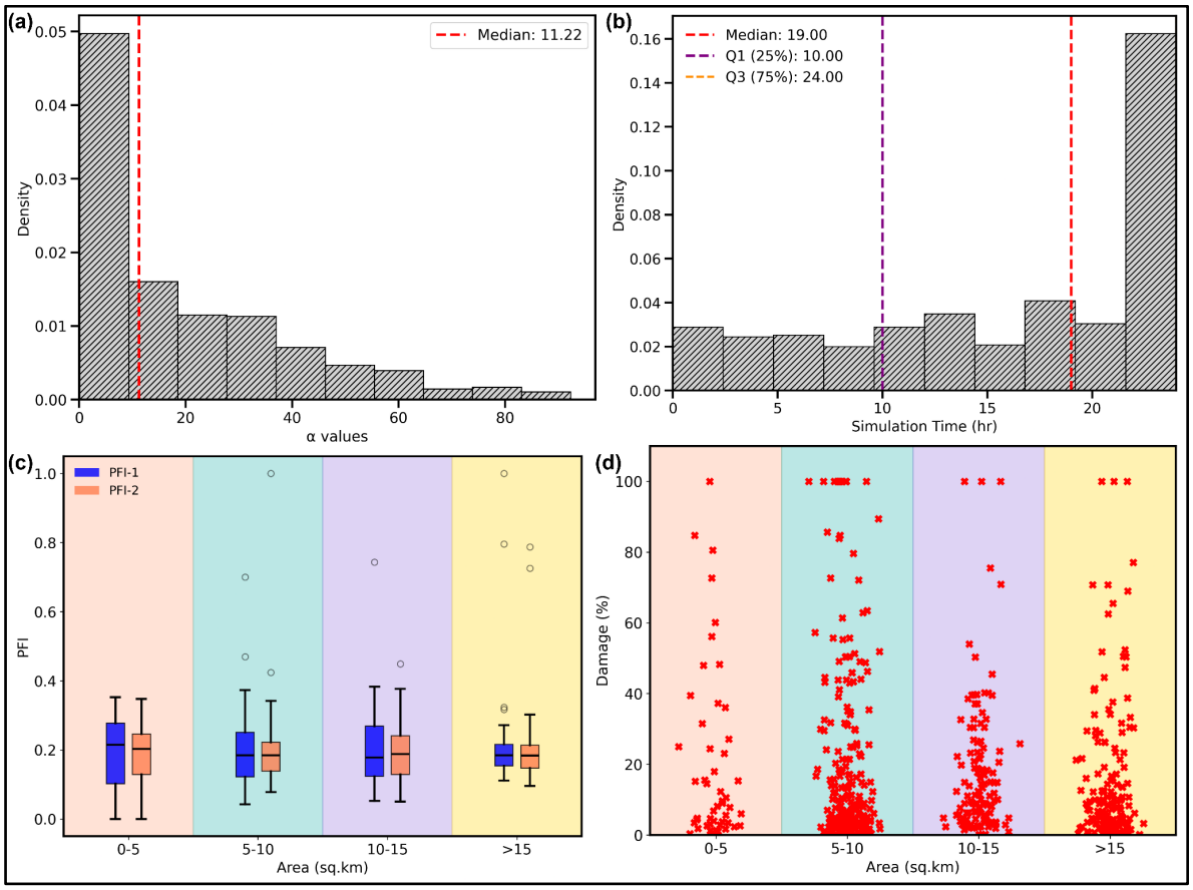


Figure 5. *Statistical analysis of pluvial flood characteristics and impacts: (a) Distribution of alpha values across all analyzed events, (b) Distribution of simulation times for all events, (c) Quantification of Pluvial Flood Indices (PFIs) for all events, and (d) Observed damage percentage for different events based on National Flood Insurance Program datasets, categorized by area division.*

Figure 5c represents the variation of PFIs across the analyzed events. We observe that both indices (PFI-1 and PFI-2) exhibit higher median values in smaller area categories. However, we find a slight decrease in median values, increased variability, and more outliers in mid-sized catchments (5–10 km² and 10–15 km²). This indicates heterogeneous hydrological behaviors. This pattern aligns with the observed clusters across different landcover types shown in **Figure** 6. We observe the dominance of developed areas in each landcover class within smaller area divisions. We also find clusters occurring in higher area divisions. This suggests increased pluvial indices in lower area catchments with developed landcover, as infiltration reduces due to higher impermeability. As the catchment area increases, multiple landcover classes are included, as seen in **Figure 6** (d), (e), and (f). This suggests an increase in permeability and decreased indices. However, when the area exceeds 15 km², multiple urban areas are included. This prompts an increment in the pluvial indices. **Figure 5d** represents the damage percentage across different catchment areas. As NFIP rounds off the exact location coordinates, damage data cannot be effectively utilized for quantifying the intensity of pluvial flooding.

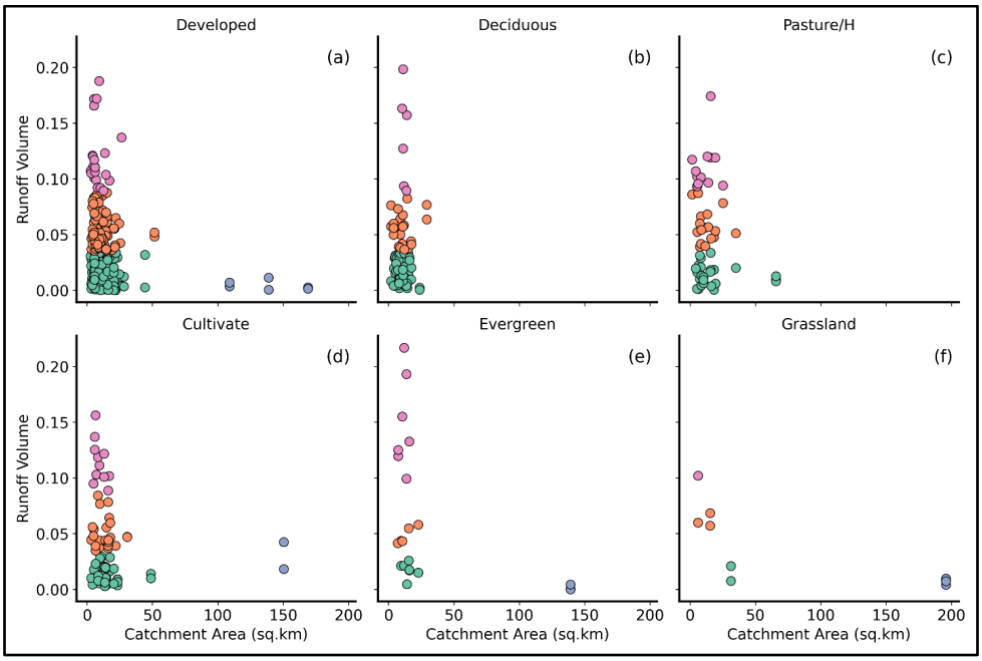


Figure 6. *Clusters observed for runoff across different landcover and area divisions*

Figure 7 demonstrates area-dependent behavior with distinct patterns across catchment sizes. We observe that small catchments (0-5 km², **Figure 7a**) are most responsive, with both indices (PFI-1 and PFI-2) showing identical 51% exceedance rates above the 0.2 threshold due to minimal storage and short travel times that allow moderate rainfall to quickly trigger alert conditions. As catchment size increases to 5-10 km² (**Figure 7b**), attenuation effects begin to emerge, with PFI-1 exceedance dropping to 41% while PFI-2 remains relatively stable at 38%, indicating that while the cumulative response becomes more damped, peak flows can still occur through sub-catchment contributions. The 10-15 km² (**Figure 7c**) range exhibits mixed behavior where cumulative response continues to attenuate (PFI-1 at 40%) but peak response increases (PFI-2 at 42%), suggesting that larger routing distances smooth flood waves, yet sufficient rainfall intensity can still produce threshold-exceeding peaks. In the largest catchments (>15 km², **Figure 7d**), storage effects dominate cumulative runoff, reducing PFI-1 exceedance to 34% despite higher mean values, while PFI-2 maintains 39% exceedance through a concentrated distribution around 0.19, demonstrating that individual storm cells or urbanized tributaries within large basins can still generate significant peak ratios even when total volumes are substantially damped. We find the choice between indices depends on the specific catchment size and whether the priority is monitoring cumulative volume response or peak flow detection.

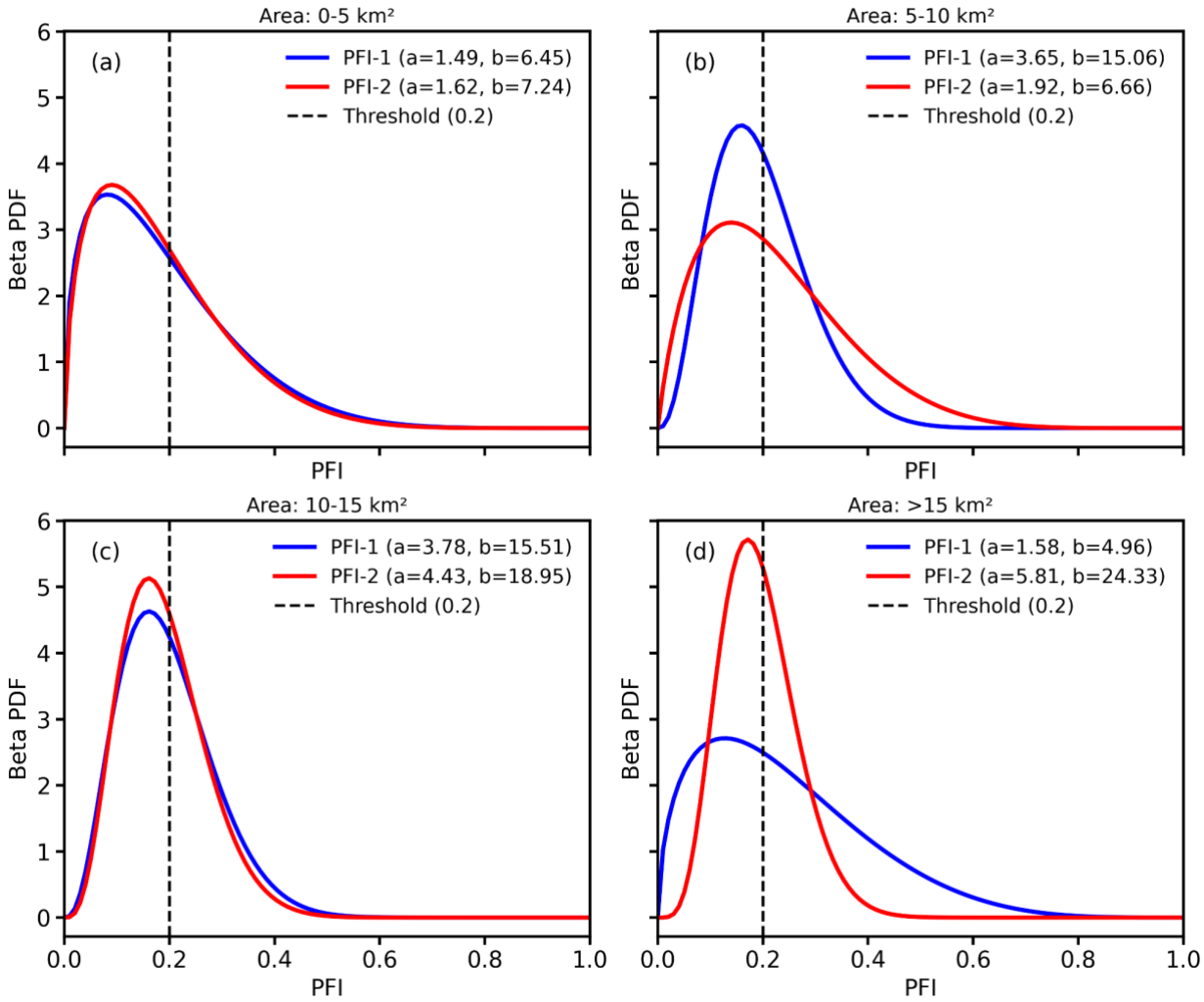


Figure 7. *Fitted Beta distribution for PFI-1 and PFI-2 across selected catchment area (a) for 0-5 km² (b) for 5-10 km² (c) for 10-15 km² (d) for >15km²*

6. Conclusion

In this study, we demonstrate the ability of the NOAA NextGen-NWM framework to simulate and detect pluvial flooding events. We prepare the pluvial flood database utilizing the LSR and the NFIP database. We establish pluvial flood indices for detecting pluvial flood signals by integrating results from CFE simulations, including runoff and rainfall data. The distribution of pluvial flood indices across different catchments with variable land cover produces distinct patterns. We find that PFIs are higher in small area catchments with urban land cover. The indices decrease as catchment area increases, and other land cover classes are included. However, PFIs suddenly increase when multiple urban areas are included in larger catchments. Observed findings provide benchmarks for pluvial flood detection within the NextGen framework. The results demonstrate the potential of NextGen for capturing fine-scale urban flood signals at operational scales.

Future research should validate these indices across diverse geographic regions and environmental conditions. Integration of real-time forcing data could improve the accuracy of α parameter estimation, simulation time, and pluvial flood indices. Likewise, integration of machine learning algorithms for automatic detection of pluvial flood signals using these indices

represents another promising direction. Additionally, coupling these indices with damage assessment models could enhance early warning system capabilities. Further investigation into the scalability of these indices across different catchment sizes and urban configurations will strengthen their operational utility. Additionally, the discretization of only urban catchment within NextGen hydrofabrics for pluvial flood indices determination can be useful.

Acknowledgements

We thank CIROH, CUAHSI, and the NOAA National Water Center (NWC) for their support during the Water Prediction Innovators Summer Institute 2025. We would like to extend our deepest gratitude to our theme leads, Dr. Marouane Temimi, Dr. Jonathan Frame, and Dr. Mohamed Abdelkader, for their unwavering support. Likewise, we would like to express our gratitude to our academic advisors, Dr. Sanjib Sharma, Dr. Sagy Cohen, and Dr. Manzhu Yu. Our gratitude extends to Dr. James Halgren for providing computing resources, which greatly facilitated our work environment. Additionally, we would like to thank Josh Cunningham and Dr. Anthony Castronova for helping us to set up the NextGen modeling framework for our study. Finally, we are grateful to Parvaneh Nikrou and Francisco Gomez for coordinating the Summer Institute and making it a fun-filled experience.

This research was supported by the Cooperative Institute for Research to Operations in Hydrology (CIROH) under award NA22NWS4320003 from the NOAA Cooperative Institute Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA.

Supplementary Materials

All the codes associated with this study can be found in this GitHub repository- https://github.com/NWC-CUAHSI-Summer-Institute/hydropulse

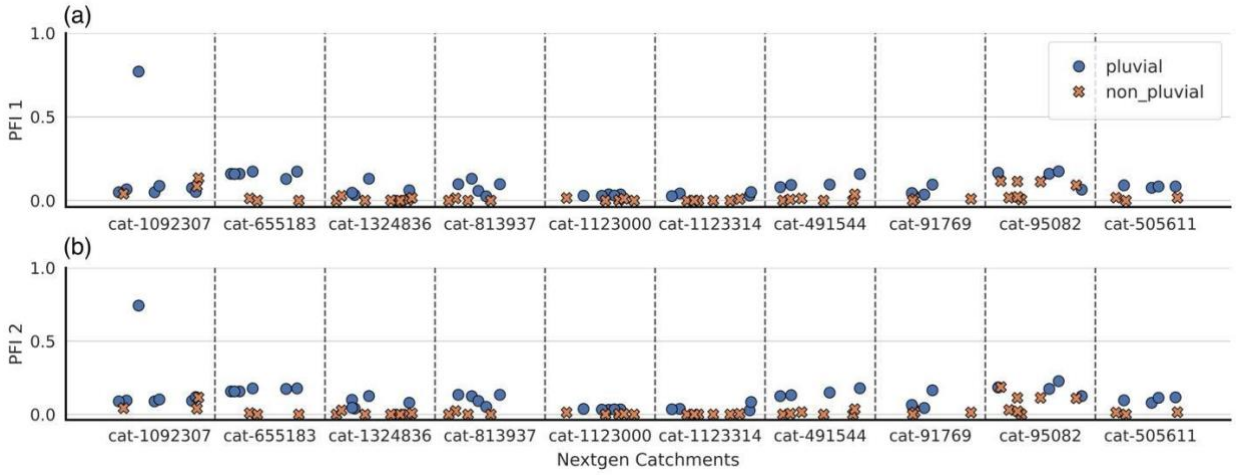


Figure S1. *Comparison between pluvial and non-pluvial events in terms of pluvial flood indices*

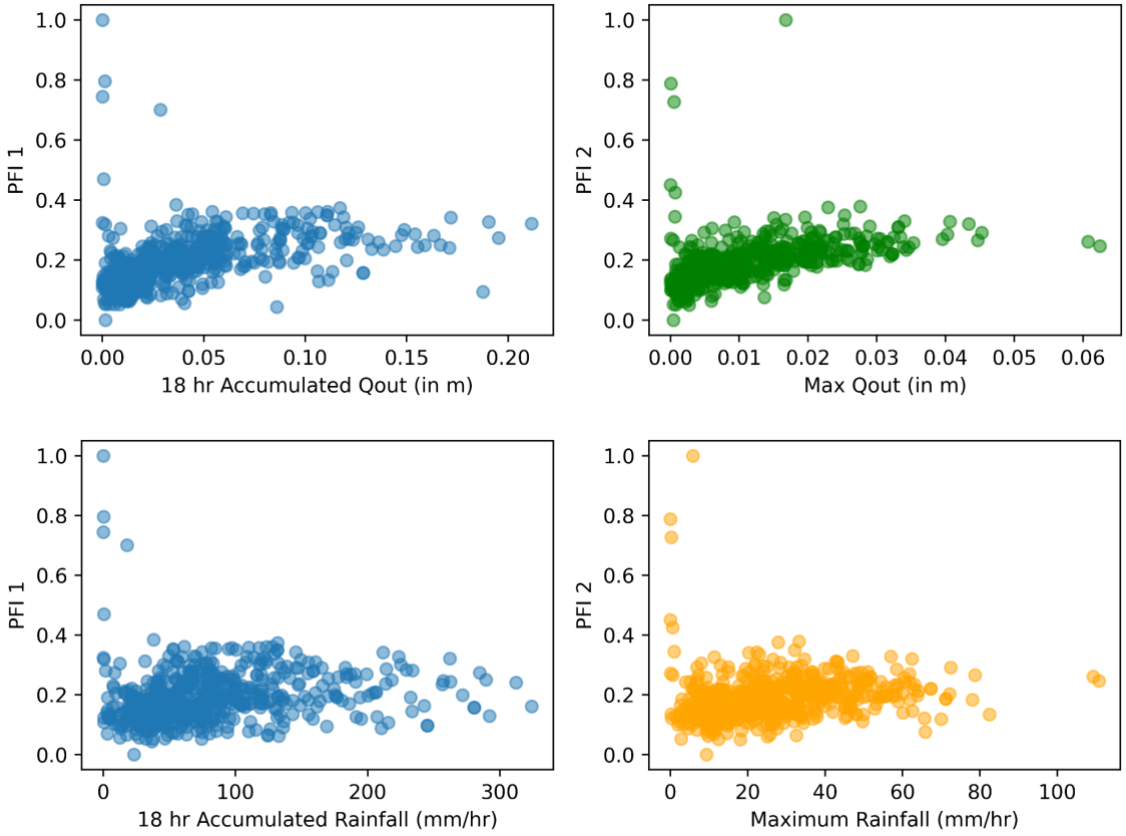


Figure S2. *Comparison between pluvial flood indices and selected forcings*

References

[1] B. Nelson-Mercer, T. Kim, V. N. Tran, and V. Ivanov, "Pluvial flood impacts and policyholder responses throughout the United States," *Npj Nat. Hazards*, vol. 2, no. 1, Jan. 2025, doi: 10.1038/s44304-025-00058-7.

[2] J. M. Frame *et al.*, "Machine Learning for a Heterogeneous Water Modeling Framework," *JAWRA J. Am. Water Resour. Assoc.*, vol. 61, no. 1, Feb. 2025, doi: 10.1111/1752-1688.70000.

[3] E. Foroumandi, H. Moradkhani, W. F. Krajewski, and F. L. Ogden, "Ensemble data assimilation for operational streamflow predictions in the next generation (NextGen) framework," *Environ. Model. Softw.*, vol. 185, p. 106306, Feb. 2025, doi: 10.1016/j.envsoft.2024.106306.

[4] Y. Bhattarai, S. Bista, R. Talchabhadel, S. Duwal, and S. Sharma, "Rapid prediction of urban flooding at street-scale using physics-informed machine learning-based surrogate modeling," *Total Environ. Adv.*, vol. 12, p. 200116, Dec. 2024, doi: 10.1016/j.teadva.2024.200116.

[5] F. T. Zahura, J. L. Goodall, J. M. Sadler, Y. Shen, M. M. Morsy, and M. Behl, "Training Machine Learning Surrogate Models From a High-Fidelity Physics-Based Model: Application for Real-Time Street-Scale Flood Prediction in an Urban Coastal Community," *Water Resour. Res.*, vol. 56, no. 10, p. e2019WR027038, 2020, doi: 10.1029/2019WR027038.

[6] B. Cosgrove *et al.*, "NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling," *JAWRA J. Am. Water Resour. Assoc.*, vol. 60, no. 2, pp. 247–272, Apr. 2024, doi: 10.1111/1752-1688.13184.

[7] B. Cosgrove *et al.*, "Operational Hydrology with NOAA's National Water Model: Current Capabilities and Future Enhancements," in *104th Annual AMS Meeting 2024*, in American Meteorological Society Meeting Abstracts, vol. 104. Jan. 2024, p. 427941.

[8] B.-C. Seo, M. Rojas, F. Quintero, W. F. Krajewski, and D. H. Kim, "Expanding and Enhancing Streamflow Prediction Capability of the National Water Model Using Real-Time Low-Cost Stage Measurements," Oct. 2022, doi: 10.1175/WAF-D-22-0050.1.

[9] T. E. Adams, "Chapter 10 - Flood Forecasting in the United States NOAA/National Weather Service," in *Flood Forecasting*, T. E. Adams and T. C. Pagano, Eds., Boston: Academic Press, 2016, pp. 249–310. doi: https://doi.org/10.1016/B978-0-12-801884-2.00010-4.

[10] A. Rostami *et al.*, "Forecasting Flood Inundation in U.S. Flood-Prone Regions Through a Data-Driven Approach (FIER): Using VIIRS Water Fractions and the National Water Model," *Remote Sens.*, vol. 16, no. 23, Art. no. 23, Jan. 2024, doi: 10.3390/rs16234357.

[11] P. D. Bates *et al.*, "Combined Modeling of US Fluvial, Pluvial, and Coastal Flood Hazard Under Current and Future Climates," *Water Resour. Res.*, vol. 57, no. 2, p. e2020WR028673, 2021, doi: 10.1029/2020WR028673.

[12] W. S. Lincoln, "2017 LMRFC Gridded Flash Flood Guidance Update," 2017.

[13] M. Weiler *et al.*, "The Pluvial Flood Index (PFI): a new instrument for evaluating flash flood hazards and facilitating real-time warning," Apr. 08, 2025, *Copernicus GmbH*. doi: 10.5194/egusphere-2025-1519.

[14] NOAA, "Storm Events Database." Accessed: Jul. 15, 2025. [Online]. Available: https://www.ncdc.noaa.gov/stormevents/

[15] Horn, D. P. & Brown, "Introduction to the National Flood Insurance Program (NFIP)." Accessed: Jul. 15, 2025. [Online]. Available: https://www.congress.gov/crs-product/R44593

[16] O. E. J. Wing, N. Pinter, P. D. Bates, and C. Kousky, "New insights into US flood vulnerability revealed from flood insurance big data," *Nat. Commun.*, vol. 11, no. 1, p. 1444, Mar. 2020, doi: 10.1038/s41467-020-15264-2.

[17] US Geological Survey, "National Land Cover Database." 2023.

[18] F. Ogden *et al.*, "The Next Generation Water Resources Modeling Framework: Open Source, Standards Based, Community Accessible, Model Interoperability for Large Scale Water Prediction," presented at the AGU Fall Meeting Abstracts, Dec. 2021, pp. H43D-01. Accessed: Jul. 15, 2025. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2021AGUFM.H43D..01O

[19] F. Ogden *et al.*, "Technical guidance for the inclusion of models/modules in the NextGen Water Resources Modeling Framework".

[20] J. Johnson, "National Hydrologic Geospatial Fabric (hydrofabric) for the Next Generation (NextGen) Hydrologic Modeling Framework | CUAHSI HydroShare." Accessed: Jul. 15, 2025. [Online]. Available: https://www.hydroshare.org/resource/129787b468aa4d55ace7b124ed27dbde/

[21] J. M. Johnson, R. Gibbs, K. S. Jennings, L. L. Kindl da Cunha, T. Flowers, and F. L. Ogden, "End-to-end Hydrofabric workflows for the NextGen Water Resources Modeling Framework," in *Frontiers in Hydrology 2022*, Jun. 2022, pp. 142–03.

[22] R. Araki, F. L. Ogden, and H. K. McMillan, "Testing Soil Moisture Performance Measures in the Conceptual-Functional Equivalent to the WRF-Hydro National Water Model," *JAWRA J. Am. Water Resour. Assoc.*, vol. 61, no. 1, Feb. 2025, doi: 10.1111/1752-1688.70002.

[23] I. Rodríguez-Iturbe and J. B. Valdés, "The geomorphologic structure of hydrologic response," *Water Resour. Res.*, vol. 15, no. 6, pp. 1409–1420, 1979, doi: 10.1029/WR015i006p01409.

[24] J. E. Nash, "The Form of Instantaneous Unit Hydrograph," 1957, [Online]. Available: https://iahs.info/uploads/dms/045011.pdf

Chapter 4: Demonstrating the Feasibility of DL-Based Pluvial Flood Mapping in Urban Settings

Mostafa Saberian[1], Sadra Seyvani[2], Adam Smith[3], Keivan Tavakoli[4], Mohamed Abdelkader[5*], Jonathan Frame[6*], and Marouane Temimi[7*]

[1] Clemson University; *mostafs@clemson.edu*

[2] The University of Alabama; *sseyvani@crimson.ua.edu*

[3] Rutgers University; *adam.m.smith@rutgers.edu*

[4] University of Texas at Austin; *keivan.tavakoli@utexas.edu*

[5*] Iowa University; *mohamed-abdelkader@uiowa.edu*

[6*] The University of Alabama; *jmframe@ua.edu*

[7*] Stevens Institute of Technology; *mtemimi@stevens.edu*

[*]Theme Leader

Abstract: Urban pluvial flooding poses a growing threat to infrastructure and public safety, exacerbated by increasing impervious surface extent and intensifying precipitation. Traditional hydrodynamic models, while accurate, are computationally intensive and ill-suited for rapid forecasting in urban environments. The goal of this study is to investigate the feasibility of using Deep Learning (DL) techniques as a surrogate model for high-fidelity, street-scale pluvial flood models. In this study, a combination of historical and synthetic rainfall events was employed to generate flood inundation maps using the Personal Computer Storm Water Management Model. The suggested methodological framework was applied to the city of Hoboken, New Jersey, a highly urbanized area susceptible to pluvial flooding. The proposed machine learning architecture integrates spatial features such as digital elevation models and building footprints through a 2D Convolutional Neural Network (CNN), and temporal rainfall inputs through a 1D CNN. The combined features are then used as inputs to generate high-resolution binary flood maps indicating flooded and non-flooded areas. The DL model evaluation showed high performance during the testing phase, achieving a Critical Success Index of 0.89, Probability of Detection of 0.92, and False-Alarm Ratio of 0.03. The obtained results highlight the potential of DL techniques to emulate high fidelity models and advance current forecasting capabilities by integrating innovative pluvial flooding modeling tools. The CNN surrogate offers a pathway toward real-time flood forecasting through an integration into upcoming water modeling platforms such as the NOAA Next Generation Water Resources Modeling framework.

1. Motivation

Floods are one of the most common and costly natural disasters in terms of both economic damage and human lives[1]. In the United States, floods kill more people per year than any other hazardous weather phenomenon [2] and threaten many areas of modern life, including electricity, water treatment, communications, and airports [3]. The urban environment plays a crucial role in exacerbating pluvial flooding [4]. The high proportion of impervious surfaces reduces infiltration, thereby increasing the volume of surface runoff [5], [6]. Pluvial floods already comprise the majority of flood claims in the United States [7], and climate change is causing

floods to increase in frequency, intensity, and flashiness [6], [8]. Therefore, calls for better strategies of flood prediction and mitigation are increasing in response to this elevated threat [4], [5], [6] as drainage systems built for past storm events may struggle to accommodate increased future demands [9].

There are three primary types of floods: fluvial, pluvial, and coastal. Coastal influence is confined to where there is tidal influence, while fluvial and pluvial events can impact anywhere inland. Fluvial flooding occurs when stream discharge overtops the banks of a channel and inundates the floodplain, in contrast to pluvial flooding, which occurs when the volume of precipitation exceeds the drainage capacity of the landscape [10]. Therefore, it is not dependent on channel routing and is sensitive to local factors[11].

The National Water Model (NWM) is a modeling framework that simulates streamflow over the entire United States [12]. Currently, the NWM does not yet integrate both pluvial and fluvial flood forecasting [13]. Therefore, pluvial flood integration into the NWM would be a step forward for flood prediction in the United States. Hydrodynamic models offer valuable insights into inundation dynamics; however, these models often require substantial parameterization and long simulation times. As a result, these types of models are not good for rapid forecasting and response. Consequently, there is a critical need to develop alternative approaches that can maintain analytical accuracy while reducing the computational demand.

2. Objectives and Scope

The primary objective of this study is to develop a data-driven surrogate model capable of rapidly predicting pluvial flood inundation maps with high spatial accuracy. By leveraging a combination of synthetically generated rainfall hyetographs, digital elevation models (DEMs), and building footprint data, the model aims to replicate the outputs of the physics-based Personal Computer Storm Water Management Model (PCSWMM) hydrodynamic simulations. Specifically, the study focuses on training a convolutional neural network (CNN) to classify each pixel of an urban domain as flooded or non-flooded based on its input characteristics. Also, the scope of this work is confined to the urban area of Hoboken, New Jersey. The model is trained and evaluated using 150 synthetic rainfall events and their corresponding maximum flood depth outputs from PCSWMM. The resulting surrogate model is designed to significantly reduce computational time while maintaining acceptable accuracy, making it suitable for near real-time flood forecasting and early warning applications in data-scarce or time-constrained environments. The proposed workflow is illustrated in **Figure 1**.

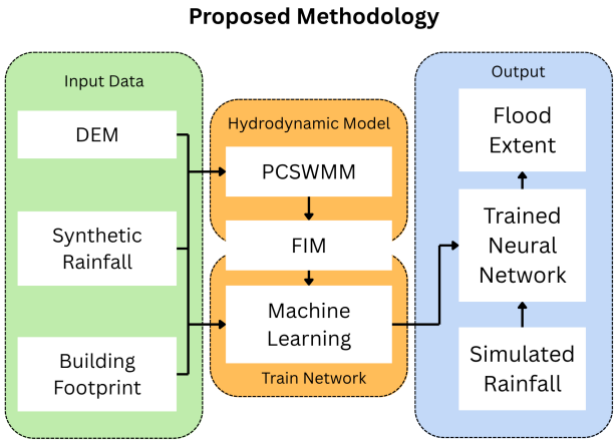


Figure 1. *Workflow of the proposed methodology integrating hydrodynamic modeling and machine learning to predict flood extent.*

3. Previous Studies

3.1 Urban Pluvial Flooding

The urban environment plays a crucial role in exacerbating pluvial flooding [4]. Storm sewer systems can become overwhelmed by intense precipitation as well as improper maintenance [14]. Other factors include the flooding of small local watercourses and the generation of increased rainfall from the urban heat island effect [15], [16]. To summarize, the key drivers of urban flooding are the built environment and anthropogenic climate change [17]. Once overland flow is generated, micro-topography plays a significant role in influencing water accumulation and flow patterns [18]. Therefore, hyper-resolution modeling is required to capture the dynamics of pluvial flood events.

3.2 Modeling Approaches for Real-Time Forecasting

The three primary approaches to flood inundation modeling are hydrodynamic models (HEC-RAS, SWMM, MIKE FLOOD), empirical models, and conceptual models (Rapid Flood Spreading Method, Height Above Nearest Drainage) [19], [20], [21]. Empirical models can only provide hindcasting based on observations and rely on assuming the underlying processes. As we cannot observe the future, this type of modeling is not applicable to forecasting pluvial floods. Conceptual models are fast but more applicable to large areas rather than providing the necessary site-tailored information needed for actionable flood prediction. Current flood forecasting relies heavily on physically based hydrodynamic models. While these models offer high accuracy and a clear understanding of underlying physical mechanisms, their computational intensity and requirement for accurate parameterization limit their utility for real-time, high-resolution urban applications [20].

Therefore, recent advances in urban pluvial flood prediction have focused on artificial intelligence (AI), machine learning (ML), and DL techniques to improve computational efficiency while preserving accuracy [22]. Traditional machine learning techniques such as random forest [23], logistic regression, support vector machine [24], and gradient boosting decision tree [25] have been explored for flood forecasting and can significantly speed up computation while preserving accuracy to a large degree. However, these training procedures can struggle with overfitting, limiting their applicability to unseen data [23], [26]. DL can serve to bridge this gap, learning physical relationships from data without explicit hydrodynamic computations [27], [28], [29].

3.3 CNNs for Urban Pluvial Flood Forecasting

DL surrogate models trained on the results of hydrodynamic models have been shown to provide comparable results in a fraction of the time to support real-time decision making [30], [31], [32], [33]. DL modes have shown the ability to successfully predict pluvial floods with high accuracy and with significant reductions in computation time [30], [34] CNN surrogate models have emerged as a valuable tool for urban pluvial flood forecasting due to their ability to identify relationships from input data to learn computational reductions to physical simulations. CNNs have been shown to outperform other types of ML models [23], [28]. Collectively, these studies illustrate the effectiveness of DL surrogate models and support the notion that this concept can be utilized for real-time flood forecasting.

4. Methodology

4.1. Study Area

The study area for this research is Hoboken, New Jersey, a highly urbanized environment (**Figure 2**) with high levels of imperviousness, low topography, and a sewer system with insufficient capacity to effectively drain the area during wet weather. Approximately 100 combined sewer overflows discharge per year into the Hudson River [35]. Hoboken benefits from extensive flood documentation available through NOAA's Local Storm Reports database, providing valuable observational data for model evaluation. Additionally, the city is equipped with an in-town rain gauge that has recorded precipitation data at 5-minute intervals since 2017, offering high-resolution rainfall input for hydrodynamic simulations. Time series flood data, including one flood event with measurements from multiple locations, supports both the calibration and validation of the model, enhancing the robustness of the simulation and subsequent machine learning training.

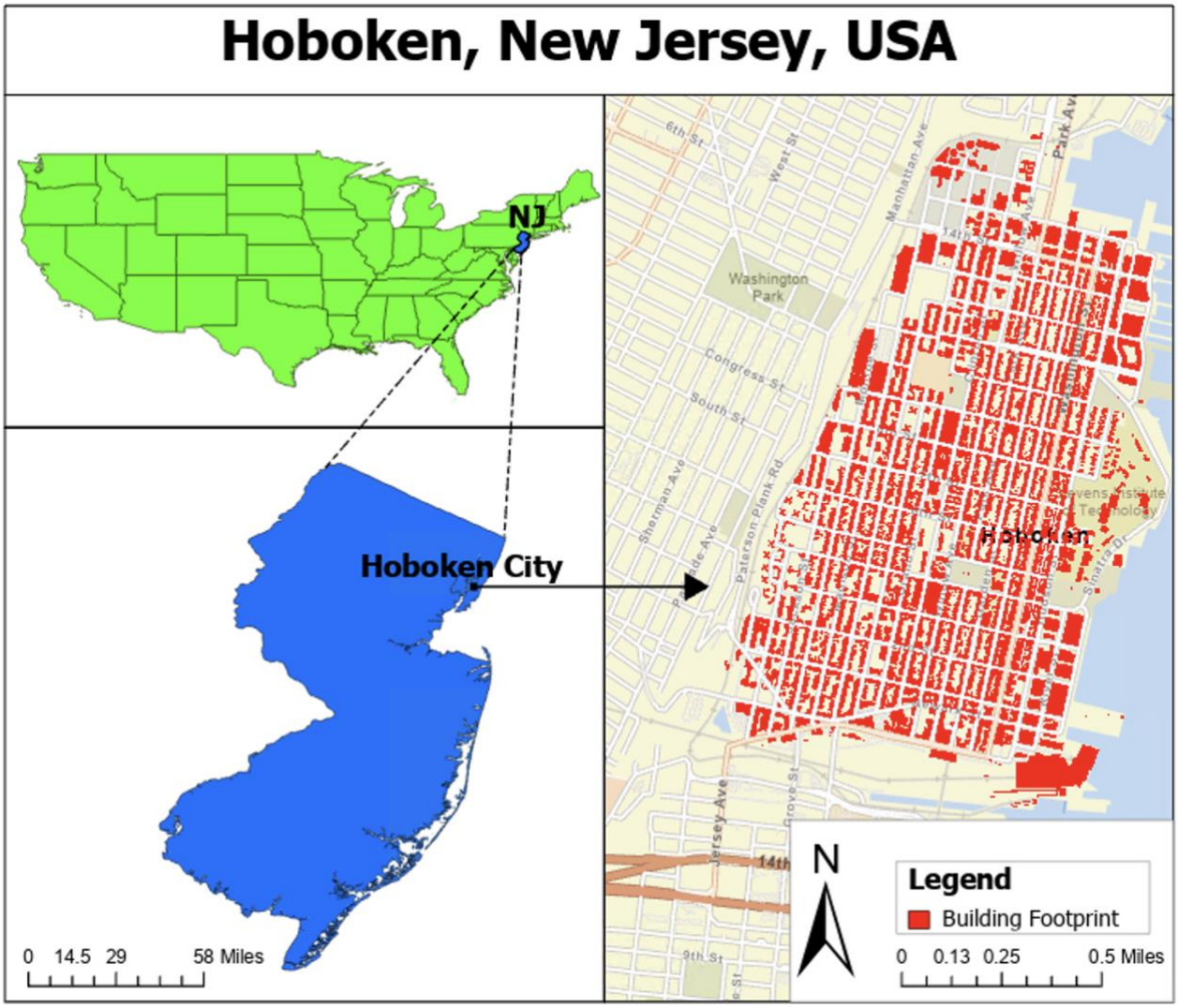


Figure 2. *Study area - City of Hoboken*

4.2. Development of Synthetic Rainfall Hyetographs Using Historical Storm Data

To develop a diverse ensemble of synthetic rainfall hyetographs, we implemented a stochastic perturbation approach based on four historical storm events recorded in Hoboken: Hurricane Barry 2019, Tropical Storm Fay 2020, Hurricane Henri 2021, and Hurricane Ida 2021. Each real rainfall series was augmented through random intensity scaling, temporal resampling using non-integer interpolation factors (via scipy.ndimage.zoom), temporal shifting, and Gaussian noise injection. This procedure generated 150 synthetic events with variable durations and intensities, mimicking the nonstationary and heterogeneous nature of urban pluvial rainfall (Figure 3). Such synthetic rainfall generation is critical for stress-testing hydrological models and training data-driven flood prediction systems, particularly in urban environments where event-based variability governs flood dynamics. The final synthetic dataset was exported in tabular format for subsequent modeling applications.

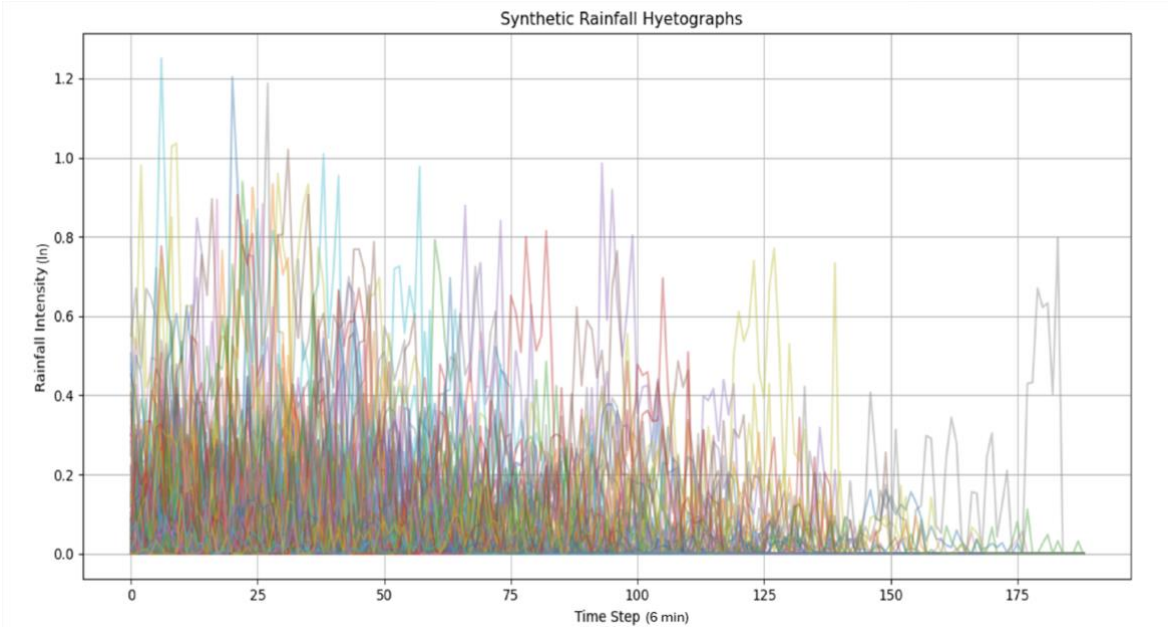


Figure 3. *Synthetic rainfall hyetographs generated from four historical events using stochastic perturbation.*

4.3. High fidelity FIM with PCSWMM

To generate training targets for the surrogate model, the PCSWMM hydrodynamic model was executed 150 times using the synthetically generated rainfall events as inputs. Each simulation produced a maximum flood depth map, representing the spatial distribution of peak inundation across the city. Rainfall events spanned a range of return periods from <1 year to >1000 years. These maps serve as the ground-truth outputs for supervised learning, enabling the development of a DL-based surrogate capable of replicating PCSWMM outputs with significantly reduced computational time. This approach supports rapid flood forecasting and real-time decision-making in urban flood management applications.

PCSWMM, developed by Computational Hydraulics International, is a GIS-based extension of EPA-SWMM designed to simulate single-event and continuous rainfall-runoff processes. It is a physically based, discrete-time model grounded in the principles of mass, energy, and momentum conservation. This study utilizes a 2D PCSWMM model with a semi-distributed routing approach, where the 2D component captures overland runoff. Sub-catchments are represented to model runoff generation and conveyance [36].

The Hoboken model covers 3,191,121 m^2 (approximately 1.2 mi^2). In the model, there are 7695 subcatchments and 3460 nodes. The size of the mesh is 10m. The roughness value is 0.014. PCSWMM utilizes various input data such as a digital elevation mode (DEM) from the USGS, building footprint from Microsoft building footprint, and storm drainage network details from the North Hudson Sewerage Authority. It processes this information using a dynamic wave routing approach. Figure 1 illustrates the overall workflow of PCSWMM.

A 1-foot threshold was applied to the depth output based on a review of flood thresholding literature [37], [38]. These maps are then used to train a DL model.

To quantitatively evaluate model performance, the pluvial flood event of September 2023 was simulated. The corresponding rainfall time series was used as the internal input, and water depth was subsequently modeled. The simulated water depths were compared with observed values at two different locations in Hoboken, as shown in Figure 4. The Root Mean Square Error (RMSE) was calculated and displayed in **Table 1**. The results demonstrate that the model is capable of accurately simulating water depth in this area in these two locations of the city.

Table 1. *Summary of evaluation metrics calculated during the Sept. 29, 2023 event*

Observation Group	Time	Observed (in)	Simulation (in)	Percent Bias (%)	RMSE (in)
A	11:45	13	10.83	-11	2.362
	11:45	7.5	10.83		
	11:30	6.5	9.95		
	11:21	9	8.9		
	11:42	10	10.42		
B	10:39	3	1.64	31	1.236
	10:36	3	1.64		
	10.:38	3	1.64		
	10:40	6.4	5.65		
Overall[1]		-		-0.16	1.944

Figure 4. *Ground truth observations collected during the Sept. 29, 2023 event. Measurements are above ground level.* [39]

After model evaluation, 150 simulations were conducted to generate corresponding flood inundation maps maximum depth), which were used to train the machine learning model.

4.4. DL Model

To predict spatial flood extents based on rainfall time series and geospatial information, we designed a dual-path DL architecture that integrates a two-dimensional (2D) CNN for spatial raster inputs and a one-dimensional (1D) CNN for temporal rainfall data, as shown in Figure 5. The model leverages feature-level fusion to combine spatial and temporal information and generate high-resolution binary flood maps.

The first path of the model processes geospatial raster data, specifically the digital elevation model (DEM) and building footprint layers, which are stacked into a two-channel input tensor of shape (256, 256, 2). This input is passed through a sequence of 2D convolutional layers designed to extract hierarchical spatial features. The raster stream begins with a convolutional layer with 16 filters of size 3×3 and ReLU activation, followed by batch normalization and a 2×2 max pooling layer. This is followed by a second convolutional layer with 32 filters (also 3×3), again followed by batch normalization and max pooling. A final convolutional layer with 64 filters is applied, followed by a dropout layer (rate = 0.3) to reduce overfitting. The output feature maps are then flattened to form a fixed-length feature vector representing the spatial context.

The second path is designed to extract features from the rainfall time series, provided as a one-dimensional vector with 189 time steps and a single channel. The input is processed using a 1D convolutional layer with 16 filters of size 3 and ReLU activation, followed by batch normalization and 1D max pooling to reduce temporal dimensionality. A second convolutional layer with 32 filters is applied, followed by a dropout layer (rate = 0.3) and flattening of the output sequence. This pathway is tailored to learn temporal rainfall patterns relevant to flood generation.

The feature vectors from the 2D CNN (raster data) and 1D CNN (rainfall) branches are concatenated to form a combined representation. This merged feature vector is passed through

a fully connected (dense) layer with 1024 units and ReLU activation, followed by a dropout layer (rate = 0.4) to enhance generalization. The output layer is a dense layer with 65,536 units (corresponding to 256×256 pixels). This flat output is reshaped into a 2D flood map of dimensions (256, 256, 1).

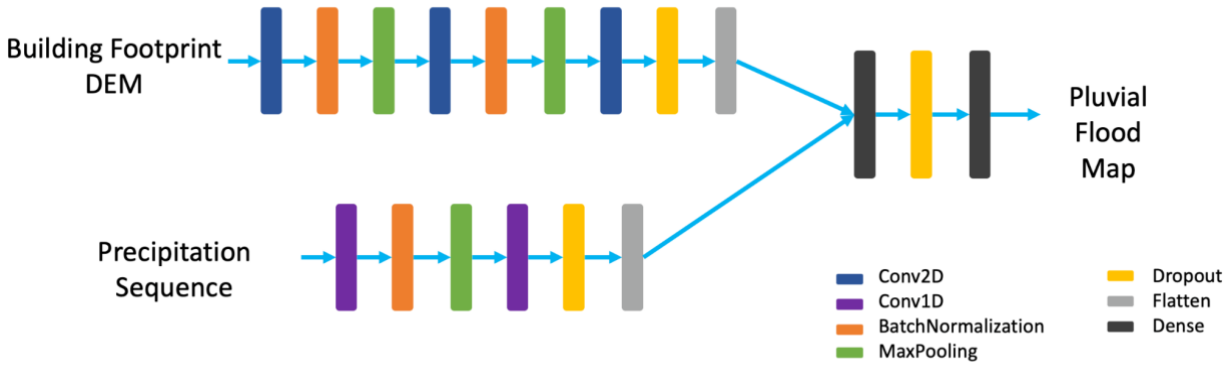


Figure 5. *DL model architecture*

5. Results

5.1. Hydrodynamic Model Results

To assess the performance of the hydrodynamic model, it was evaluated using a specific rainfall event that occurred on September 29, 2023. The simulated water depths were compared against observed data at distinct points. The results indicate that the model exhibits a Root Mean Square Error (RMSE) value of 1.944.

5.2. DL Model Results

The evaluation results for the neural network surrogate model indicate an accuracy of 0.9994. However, due to the highly imbalanced nature of the dataset—with a significantly larger number of non-flooded pixels compared to flooded ones—this high accuracy alone is insufficient to fully represent the model's predictive performance. Therefore, additional metrics must be examined. Precision (0.9669) indicates that the model rarely classifies non-flooded areas incorrectly as flooded. Recall (0.9218) demonstrates the model's effectiveness in accurately detecting actual flooded regions. The F1 Score (0.9438) combines precision and recall into a single balanced metric, further confirming the model's robust predictive capability. Additionally, the Probability of Detection is 0.92, False-Alarm Ratio is 0.03, and the Critical Success Index (CSI or Threat Score) is 0.8936. CSI measures the fraction of correctly predicted flooded pixels out of all pixels predicted or observed as flooded, thus indicating strong reliability in flood prediction. These metrics and confusion matrix values are summarized in **Table 2**.

Table 2. *The evaluation metrics and confusion matrix of the trained neural network*

Metric	Accuracy	Precision	Recall	F1 Score	CSI	True Negative	False Positive	False Negative	True Positive
Value	0.9994	0.9669	0.9218	0.9438	0.8936	75,774,690	13,577	33,649	396,452

In addition to testing our neural network surrogate model with unseen synthetic rainfall scenarios, we validated the model's performance using unseen real-world precipitation data from Hurricane Barry, which occurred on July 18, 2019. The predicted flood inundation map generated by our neural network closely aligns with the benchmark PCSWMM model results, demonstrating that the model effectively captures the spatial distribution and extent of flooded areas. **Figure 6** illustrates the PCSWMM and our neural network surrogate model results for Hurricane Barry, July 18, 2019.

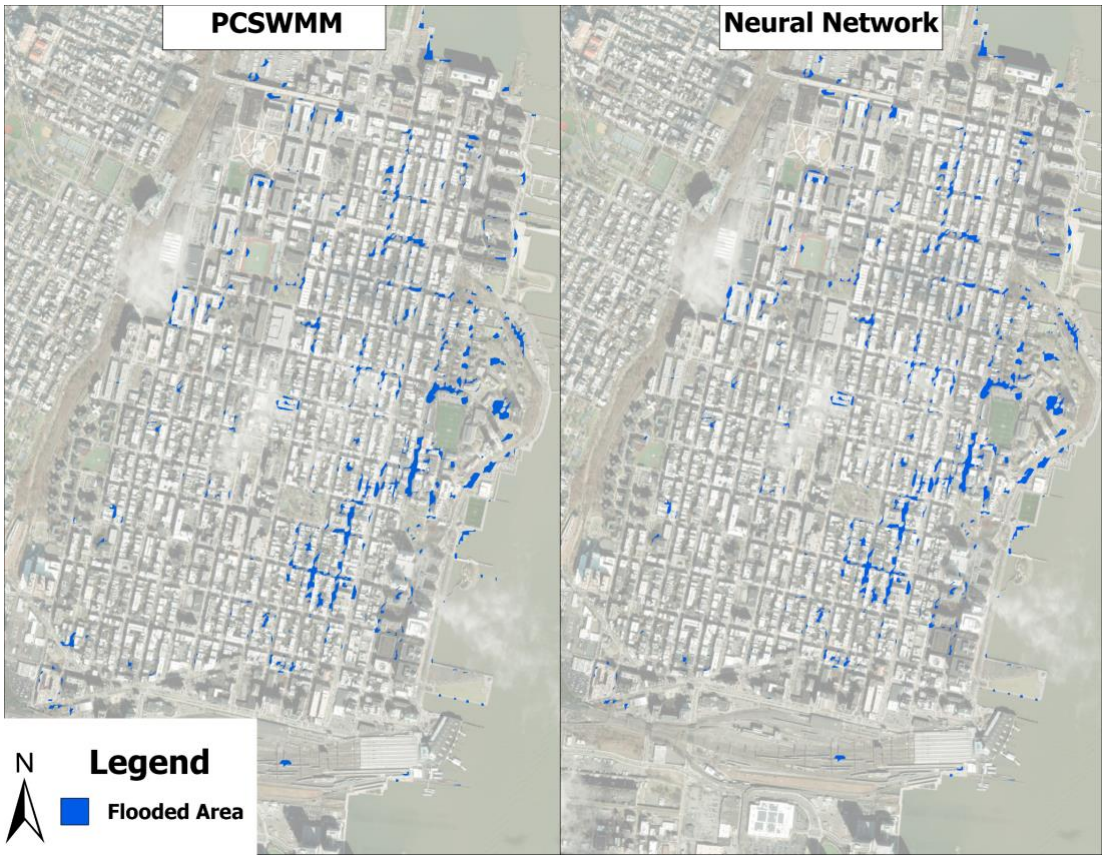


Figure 6. *Comparison between PCSWWM and Neural Network Surrogate Model for Hurricane Barry, July 18, 2019*

Evaluation metrics calculated against the PCSWMM benchmark results for Hurricane Barry indicate excellent predictive performance, with accuracy of 0.9990, precision of 0.9958, recall of 0.9314, and an F1 Score of 0.9626. These metrics demonstrate that the model not only predicts flooded regions with very high precision, thereby minimizing false-positive predictions, but also maintains a robust recall rate, effectively identifying most areas that were indeed flooded. The

high F1 Score further confirms the balanced and reliable nature of our neural network surrogate model, even when tested against actual flood events. Visual inspection of the flood maps from PCSWMM and our neural network further supports these quantitative findings, underscoring the model's capability for accurate near-real-time flood inundation mapping.

6. Conclusion

This study presents a framework for integrating pluvial flood prediction into the NWM through the use of a CNN. The CNN was trained using outputs from a physics-based hydrodynamic model as reference data and demonstrated strong predictive performance, achieving a Critical Success Index (CSI) of 0.893. The results indicate that the proposed approach is capable of accurately reproducing flood extents and can be adapted to other urban areas if a sufficiently accurate hydrodynamic model for the target region is available.

Future work should aim to implement the forecasting capabilities by feeding the trained neural network a forecast precipitation time series and then assessing the predicted to observed flood extent. Moving past deterministic flood map outputs to quantify uncertainty both in the forecast and the CNN will enhance the framework's utility for real-time flood risk assessment [40]. Furthermore, this is a data-driven CNN and must learn the underlying physical mechanisms to generate flood extent. Future work can investigate the CNN to more explicitly link the predicted flood to the specific underlying processes [41]. Additionally, training the neural network model across diverse locations can help develop a more generalized model that is transferable to other regions.

The authors acknowledge the use of ChatGPT to refine the English language during the preparation of this manuscript. All authors reviewed and edited the content following its use and take full responsibility for the final content of the publication.

Acknowledgements

We would like to acknowledge Dr. Vidya Samadi from Clemson University, Dr. Carlos E. Ramos Scharrón from the University of Texas at Austin, Dr. Ying Fan-Reinfelder from Rutgers University - New Brunswick, and Dr. Sagy Cohen from The University of Alabama for their support in all of our academic endeavors.

The PCSWMM model and the in situ observations were obtained from Stevens Institute of Technology, Theme Lead Temimi, for the purpose of this project.

This research was supported by the Cooperative Institute for Research to Operations in Hydrology (CIROH) under award NA22NWS4320003 from the NOAA Cooperative Institute Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA.

References.

[1] P. Hu, Q. Zhang, P. Shi, B. Chen, and J. Fang, "Flood-induced mortality across the globe: Spatiotemporal pattern and influencing factors," *Sci. Total Environ.*, vol. 643, pp. 171–182, Dec. 2018, doi: 10.1016/j.scitotenv.2018.06.197.

[2] US Dept of Commerce, NOAA, and National Water Service, "Turn Around Don't Drown®." Accessed: Jul. 10, 2025. [Online]. Available: https://www.weather.gov/safety/flood-turn-around-dont-drown

[3] R. Pant, S. Thacker, J. W. Hall, D. Alderson, and S. Barr, "Critical infrastructure impact assessment due to flood exposure," *J. Flood Risk Manag.*, vol. 11, no. 1, pp. 22–33, Mar. 2018, doi: 10.1111/jfr3.12288.

[4] R. H. Falconer, D. Cobby, P. Smyth, G. Astle, J. Dent, and B. Golding, "Pluvial flooding: new approaches in flood warning, mapping and risk management," *J. Flood Risk Manag.*, vol. 2, no. 3, pp. 198–208, Sep. 2009, doi: 10.1111/j.1753-318X.2009.01034.x.

[5] L. Cea, E. Sañudo, C. Montalvo, J. Farfán, J. Puertas, and P. Tamagnone, "Recent advances and future challenges in urban pluvial flood modelling," *Urban Water J.*, vol. 22, no. 2, pp. 149–173, Feb. 2025, doi: 10.1080/1573062X.2024.2446528.

[6] M. Weiler *et al.*, "The Pluvial Flood Index (PFI): a new instrument for evaluating flash flood hazards and facilitating real-time warning," *EGUsphere*, vol. 2025, pp. 1–20, Apr. 2025, doi: 10.5194/egusphere-2025-1519.

[7] G. Tonn and J. Czajkowski, "Evaluating the Risk and Complexity of Pluvial Flood Damage in the U.S.," *Water Econ. Policy*, vol. 08, no. 03, p. 2240002, Jul. 2022, doi: 10.1142/S2382624X22400021.

[8] C. Wasko, R. Nathan, L. Stein, and D. O'Shea, "Evidence of shorter more extreme rainfalls and increased flood variability under climate change," *J. Hydrol.*, vol. 603, p. 126994, Dec. 2021, doi: 10.1016/j.jhydrol.2021.126994.

[9] J. Bezdan *et al.*, "Impact of Climate Change on Extreme Rainfall Events and Pluvial Flooding Risk in the Vojvodina Region (North Serbia)," *Atmosphere*, vol. 15, no. 4, 2024, doi: 10.3390/atmos15040488.

[10] K. Beven, "Robert E. Horton's perceptual model of infiltration processes," *Hydrol. Process.*, vol. 18, no. 17, pp. 3447–3460, Dec. 2004, doi: 10.1002/hyp.5740.

[11] M. Qi, H. Huang, L. Liu, and X. Chen, "An Integrated Approach for Urban Pluvial Flood Risk Assessment at Catchment Level," *Water*, vol. 14, no. 13, 2022, doi: 10.3390/w14132000.

[12] B. Cosgrove *et al.*, "NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling," *JAWRA J. Am. Water Resour. Assoc.*, vol. 60, no. 2, pp. 247–272, Apr. 2024, doi: 10.1111/1752-1688.13184.

[13] P. D. Bates *et al.*, "Combined Modeling of US Fluvial, Pluvial, and Coastal Flood Hazard Under Current and Future Climates," *Water Resour. Res.*, vol. 57, no. 2, p. e2020WR028673, Feb. 2021, doi: 10.1029/2020WR028673.

[14] M. A. Jemberie, A. M. Melesse, and B. Abate, "Urban Drainage: The Challenges and Failure Assessment Using AHP, Addis Ababa, Ethiopia," *Water*, vol. 15, no. 5, p. 957, Mar. 2023, doi: 10.3390/w15050957.

[15] S. Reinstaller, F. Funke, A. W. König, M. Pichler, M. Kleidorfer, and D. Muschalla,

"Resilient Urban Flood Management: A Multi-Objective Assessment of Mitigation Strategies," *Sustainability*, vol. 16, no. 10, p. 4123, May 2024, doi: 10.3390/su16104123.

[16] B. M. Steensen, L. Marelle, Ø. Hodnebrog, and G. Myhre, "Future urban heat island influence on precipitation," *Clim. Dyn.*, vol. 58, no. 11–12, pp. 3393–3403, Jun. 2022, doi: 10.1007/s00382-021-06105-z.

[17] E. C. O'Donnell and C. R. Thorne, "Drivers of future urban flood risk," *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, vol. 378, no. 2168, p. 20190216, Apr. 2020, doi: 10.1098/rsta.2019.0216.

[18] P. Fiener, K. Auerswald, and K. Van Oost, "Spatio-temporal patterns in land use and management affecting surface runoff response of agricultural catchments—A review," *Earth-Sci. Rev.*, vol. 106, no. 1, pp. 92–104, May 2011, doi: 10.1016/j.earscirev.2011.01.004.

[19] J. Teng, A. J. Jakeman, J. Vaze, B. F. W. Croke, D. Dutta, and S. Kim, "Flood inundation modelling: A review of methods, recent advances and uncertainty analysis," *Environ. Model. Softw.*, vol. 90, pp. 201–216, Apr. 2017, doi: 10.1016/j.envsoft.2017.01.006.

[20] Y. Yan, N. Zhang, and H. Zhang, "Applications of Advanced Technologies in the Development of Urban Flood Models," *Water*, vol. 15, no. 4, 2023, doi: 10.3390/w15040622.

[21] A. D. Nobre *et al.*, "Height Above the Nearest Drainage – a hydrologically relevant new terrain model," *J. Hydrol.*, vol. 404, no. 1, pp. 13–29, Jun. 2011, doi: 10.1016/j.jhydrol.2011.03.051.

[22] C. Agonafir, T. Lakhankar, R. Khanbilvardi, N. Krakauer, D. Radell, and N. Devineni, "A review of recent advances in urban flood research," *Water Secur.*, vol. 19, p. 100141, Aug. 2023, doi: 10.1016/j.wasec.2023.100141.

[23] O. Seleem, G. Ayzel, A. Bronstert, and M. Heistermann, "Transferability of data-driven models to predict urban pluvial flood water depth in Berlin, Germany," *Nat Hazards Earth Syst Sci*, vol. 23, no. 2, pp. 809–822, Feb. 2023, doi: 10.5194/nhess-23-809-2023.

[24] J. T. Bersabe and B.-W. Jun, "The Machine Learning-Based Mapping of Urban Pluvial Flood Susceptibility in Seoul Integrating Flood Conditioning Factors and Drainage-Related Data," *ISPRS Int. J. Geo-Inf.*, vol. 14, no. 2, 2025, doi: 10.3390/ijgi14020057.

[25] Z. Wu, Y. Zhou, H. Wang, and Z. Jiang, "Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse," *Sci. Total Environ.*, vol. 716, p. 137077, May 2020, doi: 10.1016/j.scitotenv.2020.137077.

[26] T. Cache, M. S. Gomez, T. Beucler, J. Blagojevic, J. P. Leitao, and N. Peleg, "Enhancing generalizability of data-driven urban flood models by incorporating contextual information," *Hydrol Earth Syst Sci*, vol. 28, no. 24, pp. 5443–5458, Dec. 2024, doi: 10.5194/hess-28-5443-2024.

[27] X. Wang, M. Xiao, Y. Liu, J. Guo, Y. Qin, and Y. Zhang, "A rapid and efficient method for flash flood simulation based on deep learning," *Eng. Appl. Comput. Fluid Mech.*, vol. 18, no. 1, p. 2407016, Dec. 2024, doi: 10.1080/19942060.2024.2407016.

[28] Y. Liao, Z. Wang, X. Chen, and C. Lai, "Fast simulation and prediction of urban pluvial floods using a deep convolutional neural network model," *J. Hydrol.*, vol. 624, p. 129945, Sep. 2023, doi: 10.1016/j.jhydrol.2023.129945.

[29] H. Woo, H. Choi, M. Kim, and S. J. Noh, "Physics-Guided Deep Learning for Spatiotemporal Evolution of Urban Pluvial Flooding," *Water*, vol. 17, no. 8, 2025, doi: 10.3390/w17081239.

[30] F. T. Zahura, J. L. Goodall, J. M. Sadler, Y. Shen, M. M. Morsy, and M. Behl, "Training Machine Learning Surrogate Models From a High-Fidelity Physics-Based Model: Application for Real-Time Street-Scale Flood Prediction in an Urban Coastal Community," *Water Resour. Res.*, vol. 56, no. 10, p. e2019WR027038, Oct. 2020, doi: 10.1029/2019WR027038.

[31] J. Hou, N. Zhou, G. Chen, M. Huang, and G. Bai, "Rapid forecasting of urban flood inundation using multiple machine learning models," *Nat. Hazards*, vol. 108, no. 2, pp. 2335–2356, Sep. 2021, doi: 10.1007/s11069-021-04782-x.

[32] H. Taysi, Y. C. E. Yang, S. Gangrade, T. Chegini, S.-C. Kao, and H.-Y. Li, "Enhancing 2D hydrodynamic flood models through machine learning and urban drainage integration," *J. Hydrol.*, vol. 659, p. 133258, Oct. 2025, doi: 10.1016/j.jhydrol.2025.133258.

[33] C.-C. Lee *et al.*, "Predicting peak inundation depths with a physics informed machine learning model," *Sci. Rep.*, vol. 14, no. 1, Jun. 2024, doi: 10.1038/s41598-024-65570-8.

[34] E. Fidan, J. Gray, B. Doll, and N. G. Nelson, "Machine learning approach for modeling daily pluvial flood dynamics in agricultural landscapes," *Environ. Model. Softw.*, vol. 167, p. 105758, Sep. 2023, doi: 10.1016/j.envsoft.2023.105758.

[35] R. Šakić Trogrlić, J. Rijke, N. Dolman, and C. Zevenbergen, "Rebuild by Design in Hoboken: A Design Competition as a Means for Achieving Flood Resilience of Urban Areas through the Implementation of Green Infrastructure," *Water*, vol. 10, no. 5, p. 553, Apr. 2018, doi: 10.3390/w10050553.

[36] W. C. Huber, R. E. Dickinson, T. O. Barnwell Jr, and A. Branch, "Storm water management model; version 4," *Environ. Prot. Agency U. S.*, 1988.

[37] T.-H. Yang, Y.-C. Chen, Y.-C. Chang, S.-C. Yang, and J.-Y. Ho, "Comparison of Different Grid Cell Ordering Approaches in a Simplified Inundation Model," *Water*, vol. 7, no. 2, pp. 438–454, Jan. 2015, doi: 10.3390/w7020438.

[38] N. D. Diaz, Y. Lee, B. L. M. Kothuis, I. Pagán-Trinidad, S. N. Jonkman, and S. D. Brody, "Mapping the Flood Vulnerability of Residential Structures: Cases from The Netherlands, Puerto Rico, and the United States," *Geosciences*, vol. 14, no. 4, p. 109, Apr. 2024, doi: 10.3390/geosciences14040109.

[39] M. Abdelkader, "High Water Marks and Flood Extent During the September 29, 2023 Rainfall Event in Hoboken, NJ." HydroShare, Jun. 27, 2024. Accessed: Jul. 15, 2025. [Online]. Available: https://www.hydroshare.org/resource/656b74b4eb494550b8ffff5b596680d0

[40] S. Liu, D. Lu, S. L. Painter, N. A. Griffiths, and E. M. Pierce, "Uncertainty quantification of machine learning models to improve streamflow prediction under changing climate and environmental conditions," *Front. Water*, vol. 5, Apr. 2023, doi: 10.3389/frwa.2023.1150126.

[41] T. Zhang, R. Zhang, J. Li, and P. Feng, "Deep learning of flood forecasting by considering interpretability and physical constraints," Mar. 10, 2025, *Copernicus GmbH*. doi: 10.5194/hess-2024-393.

Chapter 5:
Towards a Flood Navigation and Safety Decision Support Tool: A Pilot for Emergency Responders in Travis County, Texas

Kayode Adebayo[1], Nana Oye Djan[2], Ali Farshid[3], Saide Zand[4], David Maidment[5*], and Kelsey McDonough[6*]

[1] South Dakota State University; *kayode.adebayo@jacks.sdstate.edu*

[2] Carnegie Mellon University; *ndjan@andrew.cmu.edu*

[3] Utah State University: *ali.farshid@usu.edu*

[4] University of Alabama: *szand@crimson.ua.edu*

[5] Professor Emeritus, University of Texas, Austin: *maidment@utexas.edu*

[6] FloodID® Product Lead, The Water Institute: *mcdonough.kelseyr@gmail.com*
*Theme Leader

Abstract: In this study we demonstrate a pipeline to distilling hydrologic forecasts to county level for emergency responders with an operational decision support tool: FLO-NAVSAFE. By combining short-range forecasts from the National Water Model (NWM) with the Height Above Nearest Drainage method, FLO-NAVSAFE generates flood inundation maps (FIMs) for near real-time, 1-, 2-, and 10-hour forecast horizons. Flood inundation uncertainty is addressed through time-lagged ensembles of NWM short range forecasts, offering both most-likely and worst-case scenarios to better reflect operational risk. In our analysis of the July 4 - 7 flood event, ~71% of flood incident reports in Travis County were captured by our FIM. Evaluation of our FIM against the best estimates of flooding yielded a Critical Success Index of 88%. FLO-NAVSAFE incorporates stakeholder-defined priorities through the Analytic Hierarchy Process and visualizes social vulnerability and critical infrastructure exposure. Built using Esri Experience Builder, FLO-NAVSAFE provides real-time and forecasted flood intelligence through an interactive, user-friendly interface. Although challenges remain with data latency and workflow automation, the platform demonstrates a scalable approach to bridging hydrologic forecasts with emergency response needs in flood-prone communities.

---

1. Motivation

While the actual costs of floods remain uncertain, it has been estimated that annual flooding costs in the United States (U.S.) range between $180 - $490 billion in 2023, including infrastructure needs for solutions and structural damage to commercial physical infrastructure [1]. The risk of flooding is expected to increase over time due to climate change and population growth [2]. This, however, does not include the loss of life, which is estimated at 127 fatalities per year [3]. The State of Texas experiences more floods and vehicle-related flood fatalities than any other state in the U.S.[4]. In 2015, Jessica Hollis, a deputy at the Travis County Sheriff's Office, was one of such fatalities, when she was swept away while checking roadways for high water on a routine patrol [5].

To aid in operational water management and to decrease the likelihood of flood fatalities, the National Water Model (NWM), a service forecasting flows in the approximately 2.7 million river reaches in the Contiguous United States (CONUS), was operationalized in 2016 [6] [7]. It has different configurations for different forecasting needs, including Short-Range (SR) deterministic forecasts spanning 18 hours, Medium-Range ensemble forecasts covering 10 days, and Long-Range ensemble forecasts extending to 30 days [6]. While many flood alert systems utilize this information to notify users of the possibility of flooding at the county level or for specific rivers and streams, they do not lend themselves much to actionable decision making. [8].

2. Objectives and Scope

Our primary aim for this project was to leverage the existing capabilities of the NWM to support emergency responders preventative disaster risk management and mitigation of flood-related damage during extreme hydrological events. Our project was scoped to Travis County, Texas (TX) and the needs of deputies (our emergency responders of focus) in the Travis County Sheriff's Office (TCSO) during fluvial flooding events. Our central objective was guided by two research questions:

- How can the NWM nowcasts and forecasts support real-time decision making by local emergency responders in the U.S.?
- How can emergency responder priorities be integrated into the development of a decision-support tool that leverages the NWM to improve its translation from research to operational application for disaster preparedness and mitigation?

To address these questions, we created a web application using NWM nowcasts and forecasts to visualize flood inundation maps showing the resultant impact of extreme flood events on critical infrastructure, roadways, and low water crossings.

3. Previous Studies

To our knowledge, operational flood inundation decision support tools specifically targeted towards emergency response within an operational context remain few and far between.

The National Oceanic and Atmospheric Administration (NOAA) Dynamic Flood-Inundation Mapping (FIM) Service, launched in September 2024, converts National Water Model (NWM) and River Forecast Center (RFC) flows into stage and extent using the Height-Above-Nearest-Drainage (HAND) framework and synthetic rating curves. Currently, covering ~30 % of the U.S. population, it delivers (i) a latest-analysis map, (ii) an RFC 5-day maximum FIM forecast, and (iii) an NWM-National Blend of Models (NBM) 5-day maximum FIM forecast [9].

To further support accessibility and operational integration of NOAA's HAND-based Dynamic FIM products, researchers at the University of Alabama, in partnership with NOAA and the Cooperative Institute for Research to Operations in Hydrology (CIROH), developed FIMserv, an open-source, notebook-based toolkit that generates custom inundation maps from NWM, the Group on Earth Observation (GEO) Global Water Sustainability (GEOGloWS), or the United States Geological Survey (USGS) discharge without the original Docker overhead [9]. FIMserv is a flexible, open-source toolkit designed to streamline and extend the operational use of the Office of Water Prediction (OWP) HAND-FIM framework [10].

While Dynamic FIM and FIMserv enable more analyst-driven workflows, Pin2Flood is a mobile- and web-based application that allows field personnel to generate instant flood maps by simply dropping a pin at the water's edge [11]. Developed by the Center for Water & the Environment

at the University of Texas at Austin, Pin2Flood, leverages the same HAND-based framework used in NOAA's Dynamic FIM service to instantly return the estimated inundation boundary, impacted address points, and affected roadway, creating a real-time common operating picture shared between field responders and emergency managers. However, instead of using streamflow forecasts, it delivers a rapid inundation estimate based on current field conditions, making it particularly valuable for localized use. A previous Summer Institute developed a workflow combining NWM forecasts, HAND mapping, and social vulnerability data to assess flood impacts in Houston during Hurricane Harvey and Hurricane Beryl [12].

Our work builds on these services or products to provide more than deterministic FIMs only for current conditions (real-time) and/or too far into the future (5 day forecasts); providing probabilistic FIMs for short range forecasts (up to 18 hours) while considering the social impact of flooding.

4. Methodology

4.1. Data and Study Area

Our study focuses on Travis County, TX, (as seen in **Figure 1**) which spans approximately 1023 square miles, of which 994 square miles is land area and 29 square miles (3 %) is covered by water [13] [14]. Nearly the entire county lies within the Colorado River Basin, which drains around 39,900 mi², originating in the Texas High Plains [15]. Numerous smaller local watersheds or catchments exist, including Onion Creek, Barton Creek, Walnut Creek, Bull Creek, and Shoal Creek, which drain into the Colorado River within Travis County. Watersheds like Barton Creek and Shoal Creek lie within "Flash Flood Alley", one of the continent's most flood-prone regions [16].

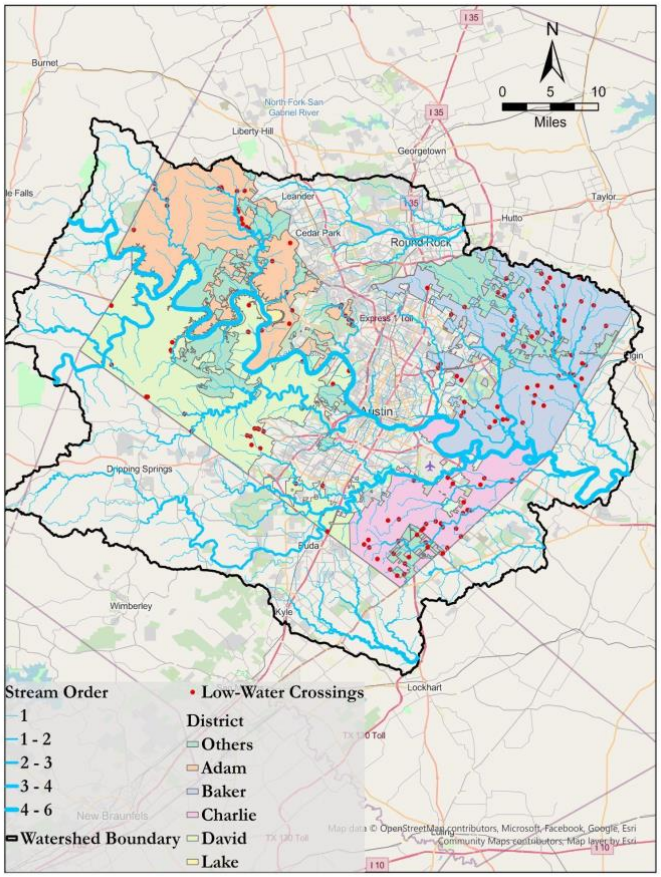


Figure 1. *Map of study area, displaying the Travis County watershed boundary (black line), stream network (blue gradients), Travis County Sheriff's Office patrol districts (colored polygons), and the locations of documented low-water crossings (red points) used in this analysis.*

A complete description of the data used in this study, including low-water crossings, road networks, flowlines, flood stacks, catchments, and jurisdictional boundaries is provided in the Supplementary Material (S1). The framework detailing our approach to create a pilot of an operational map is as seen in **Figure 2**.

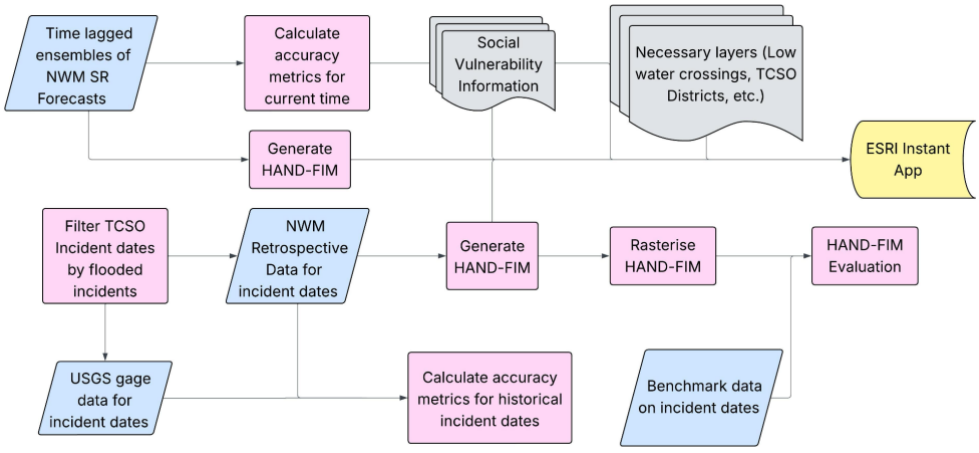


Figure 2. *Our proposed framework for creating an operational pilot tool for first responders in Travis County Texas. The pink rectangles represent tasks; the blue parallelograms represent downloaded used in completing tasks and the grey shapes represent data used to create both FIMs and maps for impact visualization*

4.2. Stakeholder Engagement and Criteria Elicitation

To align the development of our flood decision-support tool with the operational needs of local emergency responders, we conducted a structured elicitation of decision-maker preferences using a multi-criteria decision making (MCDM) method known as the Analytic Hierarchy Process (AHP). AHP allows decision makers to perform pairwise comparisons of criteria important to them, producing a consistent set of relative weights that reflect their perceived importance [17]. However, human judgements are often inconsistent and we make use of one of many adjustment methods: the skew-symmetric-bilinear (SSB) maximum-entropy approach, which finds the probability mix that no other mix can dominate and then converts it to a fully consistent weight vector. This technique preserves every original ratio, needs no manual repairs, and defaults to the standard AHP result when the matrix is already consistent [17]. A further breakdown of these methods can be viewed in the Supplementary Information (S2). To aggregate the responses of respondents, we used two methods: an element-wise geometric mean of the Saaty matrices and a geometric mean of the weights after computing the weights from each respondent.

An initial meeting with TCSO after the Summer Institute Bootcamp yielded eight distinct but related criteria/features needed in our decision support tool: FLO-NAVSAFE. The preference weights obtained through AHP were subsequently used to guide the design and feature development of FLO-NAVSAFE, ensuring that the system directly addresses at least the top five of the eight responder needs given the time constraint of the project.

4.3. FIM Uncertainty Using Time-Lagged Ensembles

To communicate flood forecast uncertainty to emergency responders in Travis County, we used time-lagged ensembles (TLEs) from the Short-Range NWM. For each forecast horizon, 1, 2, and 10 hours ahead, we retrieved 10 streamflow forecasts initialized at different times, capturing short-term variability. We generated a streamlined set of FIMs tailored to the needs of the TCSO. For both 1-hour and 2-hour forecasts, we produced: (1) a most likely scenario based on the maximum extent from the two most recent ensemble members, and (2) a worst case scenario from the maximum streamflow across all 10 members. For the 10-hour forecast, we created a single worst case FIM from the maximum streamflow across all ensemble members and forecast hours. This summarized flood uncertainty while making it easier for users to interpret and act on the results. The conceptual framework can be seen in **Figure 3** [18].

Field teams stressed that they did not want complex probability maps, and instead preferred visualization of what is likely to flood and what could flood in a worse-case scenario. This framing matches how responders make decisions, where any water on the road is treated as hazardous regardless of depth. Instead of statistical probabilities, we provided clear, consequence-based scenarios to support timely action. Prior research shows forecasts are most effective when uncertainty is translated into simple, actionable insights [19] [20]. The uncertainty shown reflects differences in streamflow forecasts, not uncertainties associated with the NWM, its forcing inputs, or the HAND-based mapping method.

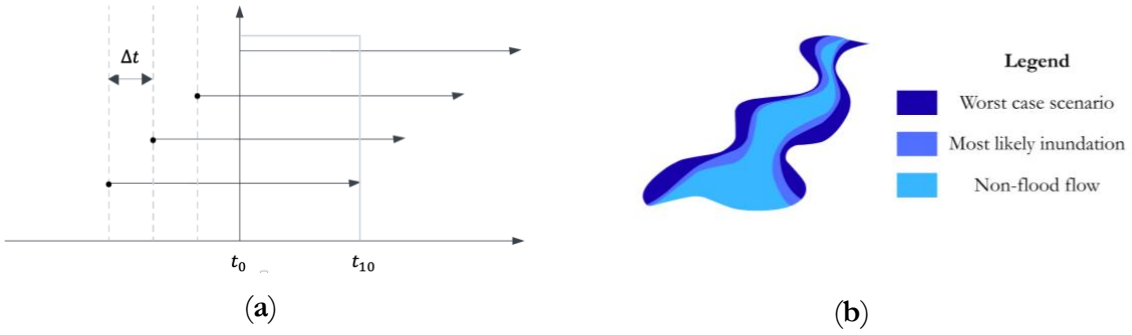


Figure 3. *(**a**) Conceptual diagram of the TLE used to generate flood-inundation scenarios for Travis County. Each horizontal arrow represents a Short-Range NWM forecast run; darker points mark the previous initialization runs. At any forecast time t_0, the 10 most recent runs (spaced Δt = 1 hour apart) form the ensemble. Adapted from Oh and Bartos's paper in review. (**b**) 1h and 2h flood inundation scenarios resulting from the time lagged ensemble*

4.4. Evaluation Metrics

The predictive reliability of the NWM was assessed using the Constant Ranked Probability Score (CRPS) for current times. CRPS was chosen due to its ability to evaluate the accuracy of ensemble members against observational data (in this study, USGS gauges) [19]. Percentage bias was used to determine whether the NWM overestimated or underestimated current conditions [21].

For historical data, the Kling Gupta Efficiency (KGE) as well as the Nash–Sutcliffe Efficiency (NSE) were used to compare USGS discharge values at three sites (Brushy Creek at Cedar Park - 08105872, Lake Creek at Lake Creek Parkway near Austin - 08105886 and Onion Creek near Driftwood - 08158700) to those obtained from the NWM retrospective dataset. Event dates to perform these analyses were obtained from flooded roadway and high water incidents reported by TCSO between 2015 and 2016. This evaluation was performed using OWP's Python package for hydrological evaluation and validation, hydrotools [22].

To evaluate our FIMs, we used fimeval, a Python-based FIM evaluation framework capable of automatically evaluating flood maps [23]. Event dates were obtained by filtering TCSO Incident data for flooded roadway or high water events. HAND-FIMs were generated for the July 4th - 5th events that occurred this year in Travis County using the forecasts made for those days in the NWM Amazon short range forecast bucket. Due to a lack of remote sensing based benchmark data, FIMs from the NWM analysis assimilation with data assimilation were generated to be used as a benchmark. The metrics used to evaluate our FIMs included the Critical Success Index (CSI), Probability of Detection (POD) and False Alarm Rate (FAR). Furthermore, another method we used to evaluate our FIM for July 4th and 5th, involved finding the distance between our FIM polygons and flooded roadway incidents reported by the TCSO on July 4th and 5th 2025.

4.5. Impact Visualization

The American Community Survey (ACS), conducted annually by the U.S. Census Bureau, collects data on over 3.5 million U.S. households to capture social, economic, housing, and demographic characteristics[24]. For this study, we used the 2019–2023 ACS 5-year estimates to

extract four indicators of social vulnerability during floods: households below the poverty line, with at least one person with a disability, without Internet access, and without vehicles. These variables are commonly used in flood vulnerability research because they highlight barriers to evacuation, preparedness, and recovery [25]. For example, people with disabilities may face mobility or communication challenges, while low-income households often lack resources to recover after disasters [26]. Lack of Internet access can limit awareness and timely response [27]. Households without vehicles are more highly exposed to flood risk, especially in areas where private transportation is critical for evacuation and access to aid [28].

In this study, Environmentally Sensitive Critical Infrastructure (ESCI) refers to facilities or services that, if flooded, could trigger widespread environmental contamination threatening drinking water supplies, harming ecosystems, and exposing nearby populations to hazardous substances during and after the event [29] [30]. We focused on five categories of ESCI in the study area: landfills, National Pollutant Discharge Elimination System (NPDES) permit holders, hazardous waste handlers, public water systems, and Toxics Release Inventory (TRI) reporters. Landfills manage the processing and disposal of solid waste. NPDES permittees such as wastewater treatment plants and industrial stormwater dischargers are authorized under the Clean Water Act to discharge pollutants from point sources into U.S. waters under regulated conditions. Hazardous waste handlers include facilities involved in the generation, treatment, storage, or disposal of hazardous materials. Public water systems encompass infrastructure such as treatment plants and pump stations that supply water to the public. TRI reporters are facilities in designated industries that use, produce, manage, or release listed toxic chemicals [31]. We obtained the locations (shapefiles) of facilities from the U.S. Environmental Protection Agency's Facility Registry Service [32].

4.6. Application Development

The development of the flood monitoring and forecast web application was carried out using Esri Experience Builder, integrating spatial and temporal flood data into a user-friendly interface. The application was structured into five main interactive pages: home, monitor, forecast, historical, and social vulnerability, each serving a specific purpose in communicating current, forecasted, and historical flood conditions.

To create the web maps and visualizations, ArcGIS Pro was used to process flood inundation layers based on forecasted discharges. HAND and Synthetic Rating Curves (SRCs) were used to delineate flood extents across Travis County. Forecasted discharge outputs from the NWM were integrated into the HAND model to produce hourly flood inundation maps and low water crossings within the FIMs were labeled flooded. The FIMs, low water crossings, and impact visualizations were subsequently uploaded as feature layers to ArcGIS Online for real-time accessibility.[1] Because ACS data are aggregated by census geography, we used the Enrich Tool in ArcGIS Pro 3.5.1 to align the data with the sheriff's office sector boundaries.

5. Results

5.1. Stakeholder Criteria Elicitation

[1] Impact visualizations were uploaded as static maps and did not possess any real-time updates due to the frequency of collections by ACS and EPA's Facility Registry Service.

Both our aggregation strategies yielded the same ranking of features, although with different weights. When the two comparison matrices were first combined element-wise, the skew-bilinear algorithm produced a consistent solution (Consistency Ratio (CR) ≈ 0.08); the first two criteria each received a weight of 0.324, while the remaining six criteria obtained individual weights ≤ 0.067. Applying the alternative strategy, deriving weights for each responder and then averaging them reduced the leading weights slightly to 0.278 and 0.228, and elevated the third-ranked criterion to 0.134, but left the rank ordering of the top five criteria unchanged.

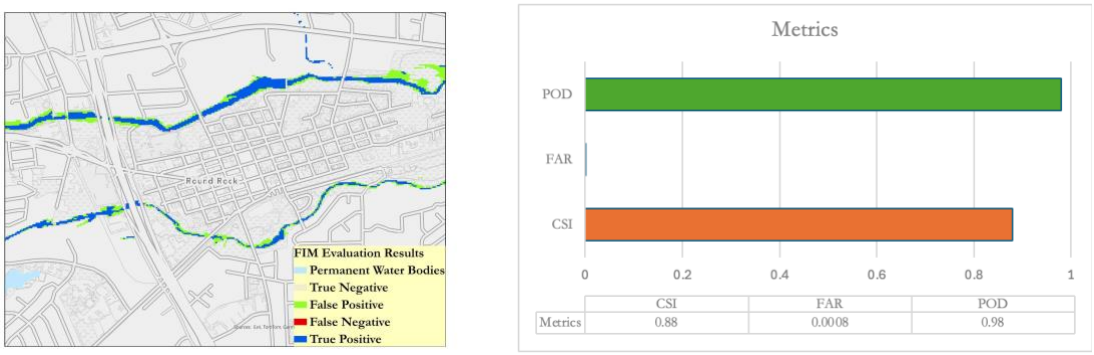
Across both methods, these five criteria accounted for more than 80 % of total importance, confirming their centrality to first-responder needs and guiding their prioritization in the initial prototype. These five criteria (in order) were: *Simple and Concise Alerts, Reliable Predictive Capability, Time and Distance Information, Export tools* and *Historical Information*. When the prototype was first scoped, however, only one survey response had been received. Given this time constraint we based the minimum-viable feature set of our app on that single matrix's top five criteria: *Simple and Concise Alerts, Reliable Predictive Capability, Time and Distance Information, Uncertainty Display*, and *Contextual Information* to ensure that development proceeded with the most critical user needs already addressed while awaiting additional feedback.

5.2. FIM Uncertainty and Evaluation Metrics

The uncertainty scenarios for each forecast horizon were integrated into the Flo-NAVSAFE web app through the Forecast page, where users can toggle between flood maps for 1-hour, 2-hour, and 10-hour lead times. Each forecast, except for the 10-hour lead times, displays both a most likely and worst case scenario, helping responders see what is expected to flood and what could flood under more severe conditions. Low-water crossings and address points are labeled and color coded as Safe or Flooded based on the selected FIM layer, providing instant visual feedback. The 10-hour lead time forecast layer highlights all areas that may be at risk using the maximum values across the ensemble to generate the FIM, thus aiding in the support of proactive planning. This option is illustrated in the application interface shown in **Figure 7**.

Figure 4[2] shows the evaluation of our HAND-FIM for July 4 - 7, 2015 in Travis County against the NWM analysis assimilation with data assimilation for the same day.

[2] Figure 4a shows only a portion of Travis County. The entire watershed was too large to view well in this report context.



(**a**) (**b**)

Figure 4. *NWM Short range forecast evaluation against NWM analysis assimilation with USGS data assimilation; (**a**) Visual representation of True Positives (Dark Blue), True Negatives (Grey), False Positives (Green) and False Negatives (Red) (**b**) Performance metrics of our FIM against the NWM analysis assimilation with DA FIM, CSI = 0.88, FAR = 0.0008, POD = 0.98.*

Table 1 shows the CRPS and Percentage Bias of the NWM short range forecast against the instantaneous USGS gauge data for 14th July 2025 at 4 PM UTC[3]. **Table 2** shows the KGE and NSE of the NWM Retrospective Data against the mean of the 15 minute USGS instantaneous values for 9th May 2015. The percentage bias indicates that the NWM tends to underestimate streamflows, the negative NSE values suggest poor model accuracy and limited predictive skill, the negative KGE values demonstrates that the NWM struggles to accurately capture the variability, correlation, and bias of the streamflows, while the CRPS also demonstrates poor accuracy and precision of the short range forecast ensemble.

Table 1. *CRPS and Percentage Bias of the NWM short range forecast against the instantaneous USGS gauge data for 14th July 2025 at 4 PM UTC*

USGS Site Number	NWM Feature ID	Percentage Bias	CRPS[1]
08105872	5671619	-35.84	1.72
08105886	5673157	-28.52	0.039
08158700	5780099	22.34	0.027

[1] CRPS ranges from 0 to +∞, the closer to 0, the better.

Table 2. *KGE and NSE of the NWM Retrospective Data against the mean of the 15 minute USGS instantaneous values for 9th May 2015*

USGS Site Number	NWM Feature ID	NSE	KGE
08105872	5671619	-0.59	0.052
08105886	5673157	-25.70	-3.32
08158700	5780099	-0.087	0.43

[3] This was the current time when this evaluation was made.

Using the incident reports provided by TCSO for July 4 - 7, 2025, we determined how close incident locations[4] are from flooded polygons as seen in **Figure 5**. ~71% of recorded incidents were within 300 m of flood polygons. This buffer zone represents the uncertainty in caller geolocation triangulation as well as the distance between caller and the actual flood inundation. Reported incidents within this buffer zone are assumed to be captured by HAND-FIM.

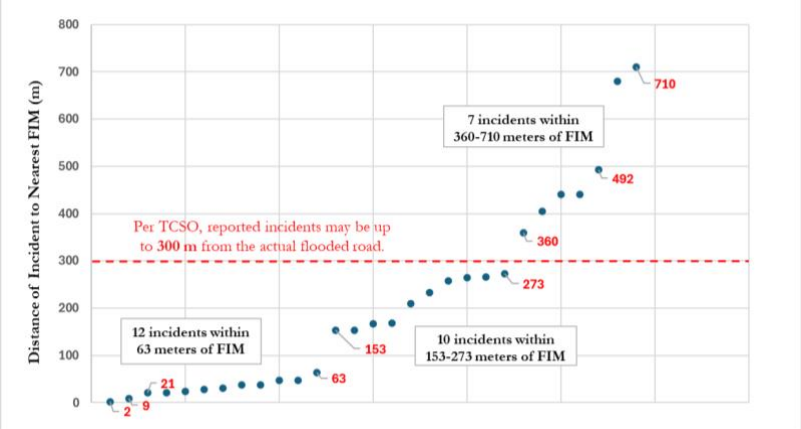


Figure 5. *Distance from reported flood incidents (July 4–7, 2025) to the nearest Flood Inundation Mapping (FIM) area. Each blue dot represents a flood report, sorted in ascending order by its proximity to the nearest FIM polygon.*

5.3. Demonstrated Capabilities of FLO-NAVSAFE

FLO-NAVSAFE (FLOod NAVigation & SAFEty for Emergency Responders) is an operational decision support tool designed to support real-time flood awareness and response. **Figure 6a** displays the Home page of the developed application. The Home page of the FLO-NAVSAFE was designed to support real-time (nowcast) flood monitoring by integrating a live feed of flood inundation data from the NWM Dynamic FIM service. This feature service is updated hourly. It is visualized on the map as red polygons representing areas of inundation.

Users can zoom into any region of interest for a more detailed view of these polygons. A Table component is provided, which populates the flood polygons across CONUS upon interaction, with built-in filtering options. The Query tool allows users to draw a polygon around any desired area to retrieve a list of flooded locations within the selection. The Legend clearly indicates the symbology used for each layer on the map. The Map Layers panel displays all available layers, with key ones (such as low water crossings, Travis County watershed boundary, and flood inundation polygons) enabled by default. Users can toggle additional layers as needed. A Basemap Selector is also included, allowing users to switch between different basemap styles.

Navigation is user-friendly: users can zoom using the mouse scroll wheel or the on-screen zoom controls and return to the default map extent using the Home button. Additionally, the Directions Widget enables users to generate routes, view distances, and estimate travel time from their current location or any specified address to another point of interest. Address inputs can include standard U.S. addresses, address points within Travis County, or the locations of low water crossings.

[4] Caller locations were assumed to be incident locations by TCSO. Callers included residents of Travis County as well as TCSO deputies.

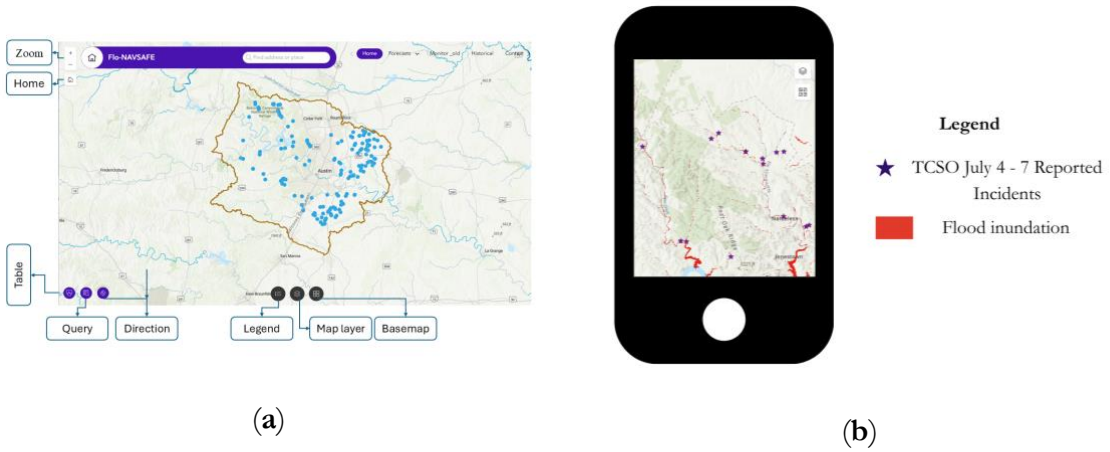


Figure 6. *FLO-NAVSAFE desktop and mobile configurations (**a**) Desktop configuration of the home page of Flo-NAVSAFE, displaying the main interface and core interactive widgets. Key functions are labeled, including zoom controls, home button, table view, query tool, direction widget, legend, map layer toggles, and basemap selector. (**b**) Mobile view of the historical page for the July 4, 2025 flood event showing the maximum NWM FIM extent over the entire flood event and the TCSO reported flood incidents.*

Figure 7 presents the Forecast page, which was designed to visualize flood uncertainty by displaying both most likely and worst-case flood scenarios. This page provides forecasts for 1-hour ahead and 2-hours ahead conditions, with the most likely scenario in orange and the worst-case scenario in red. Additionally, a separate forecast is included for the worst-case scenario over the next 10 hours, offering extended outlooks for planning and preparedness. By default, both the most likely and worst-case forecast layers are activated.

Users can toggle these layers on or off using the Map Layers control to tailor the visualization to their needs. The low water crossings layer is also enabled by default, using purple to represent flooded crossings and green for safe ones. These indicators provide immediate visual feedback on roadway conditions under different forecast scenarios.

Complementing the map, interactive charts display the number of low water crossings and address points to be inundated under each forecast (1-hour, 2-hour, and 10-hour). To support analysis and reporting, users can export the chart data as CSV files, which include the number of flooded and safe low water crossings and address points, disaggregated by districts.

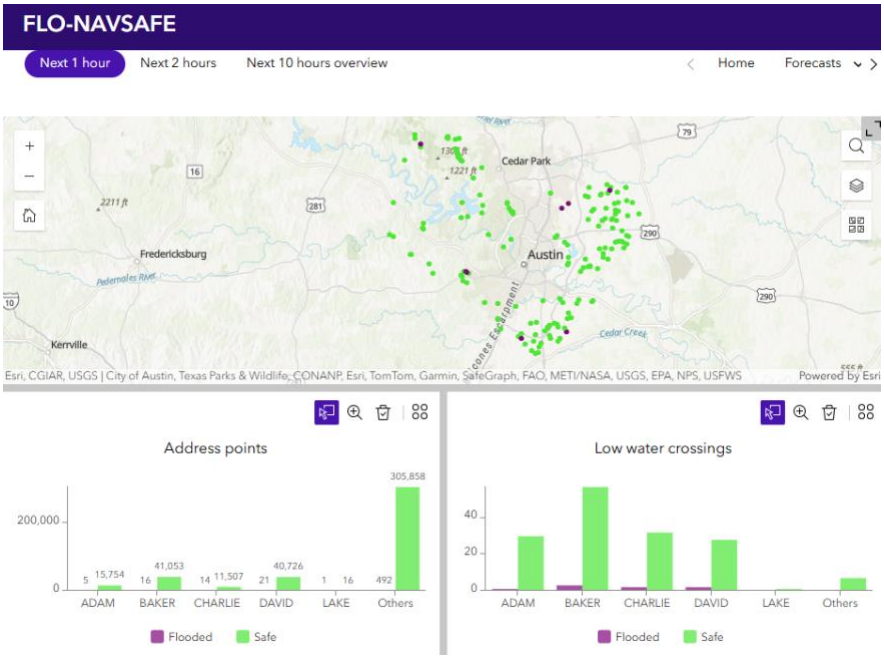


Figure 7. *Forecast page of the Flo-NAVSAFE, displaying predicted flood inundation areas and low water crossings. The map visualizes flood scenarios alongside flooded (purple) and safe (green) low water crossings. The interactive bar charts below summarize the number of inundated and safe address points and low water crossings by county, supporting short-term flood impact assessment for the next 1 to 10 hours.*

5.3.1. Impact Visualization

We analyzed five socioeconomic vulnerability indicators including age, income, disability, internet access, and vehicle ownership across five law enforcement jurisdictions in Travis County: Adam, Baker, Charlie, David, and Others. The "Others" district showed the highest vulnerability, with 40.4% of the population aged 85 or older, 27.2% of households below the poverty line, 40.1% including at least one person with a disability, and nearly one-third lacking both internet and a vehicle. Baker followed closely, with 44.6% of households below the poverty line and similarly high levels of limited internet and vehicle access. David had 29.7% of its population aged 85 or older but moderate vulnerability otherwise. Adam and Charlie showed consistently lower vulnerability levels. The analysis reveals that areas with higher poverty also tend to have higher disability rates and reduced access to internet and transportation, increasing their flood-related risks due to overlapping vulnerabilities that may significantly hinder evacuation and recovery. Detailed social vulnerability maps are provided in Supplementary Material S3.

Our analysis also identified 114 Environmentally Sensitive Critical Infrastructure (ESCI) facilities under the National Pollutant Discharge Elimination System, including 99 Industrial Stormwater Dischargers and 10 Publicly Owned Wastewater Treatment Works. The Toxics Release Inventory program includes 44 TRI Reporter facilities, while the Safe Drinking Water Information System lists 37 facilities with 33 Water Treatment Plants and 4 Community Water Systems. Hazardous waste sites account for 21 entries, including 12 Transporters, 6 Large Quantity Generators, and 3 Transfer Facilities. One landfill is also present. The most common facilities are Industrial Stormwater Dischargers (99), TRI Reporters (44), and water system sites (33). These facilities are vital to public health and are especially vulnerable during floods, which can trigger chemical releases or water contamination. A detailed ESCI map is provided in Supplementary Material S3.

5.3. Limitations

5.3.1. Issues with HAND-FIM and Rating Curves

A significant technical limitation in our approach stems from the use of HAND-based flood inundation mapping and synthetic rating curves. While the HAND method enables rapid and scalable flood estimation, it relies on simplified assumptions, such as uniform flow, no lag time and constant slope, which may not hold true in urban settings or complex river networks. Similarly, the rating curves used to convert streamflow into stage height are often synthetic and uncalibrated, which can introduce errors in the estimated inundation extent. Furthermore, HAND-FIM only models fluvial flood events; it excludes compound flooding. HAND-FIM also underpredicts smaller catchments (4th order and lower) and overpredicts for larger reaches [33].

5.3.2. Server Issues for Automation

The Home page remains fully functional through automatic updates from the NWM Dynamic FIM service, which refreshes approximately every 55 minutes. In contrast, the Forecast pages rely on automated workflows for downloading discharge data and generating FIMs. Since FIM generation is performed using ArcGIS Pro, automation currently requires a computer with ArcGIS Pro installed. At present, a local machine at The University of Alabama runs this workflow using Windows Task Scheduler to execute an arcpy script within a cloned ArcGIS Python environment. To enhance scalability and reliability, a future improvement would involve migrating this automation to the cloud, such as Amazon Web Services (AWS), where ArcGIS Pro is now supported.

5.3.3. Benchmark for FIM evaluation

Another key limitation is our lack of benchmarks for validating HAND-FIMs generated on our event dates. Our "benchmark" was also HAND-FIM generated using the NWM analysis assimilation with data assimilation, which assimilates observed meteorological forcings as well as discharge from USGS gauges, representing the best possible estimate for current conditions. This "benchmark" is also subject to model uncertainty as well as the uncertainty in HAND and synthetic rating curves.

5.3.4. Latency

A primary limitation of the system is latency, the delay between actual hydrologic conditions and when they appear in the app. This results from the NWM forecast latency, forecast posting schedule, data transfer from the AWS S3 bucket (typically ~25 minutes post-forecast), FIM generation in ArcGIS Pro, and upload to ArcGIS Online, all of which can take an additional 40 minutes. Consequently, the 1-hour forecast may reflect conditions that are now happening and the next 2-hour might be more reliable for the next hour. Also the live FIM from NWM used for the home page has a latency of 55 mins and it is updated every hour [34].

5.3.5. Incidents outside Flood Polygons

In **Figure 5**, ~71% of incidents reported during the July 4 - 7, 2025 events in Travis County were within the defined 300m buffer zone. The remaining 29% could be on other flowlines not captured by the NWM due to the coarseness of the 10 m resolution digital elevation model used to define the flowlines used in generating our HAND-FIMs, which reduces its usability for community-level applications, particularly in urban settings such as Travis County.

5.3.6. Limitations of ACS-Based Demographic Estimates

As with any survey-based dataset, the ACS provides estimates of socioeconomic and demographic characteristics rather than exact population or household counts. These estimates are subject to uncertainties stemming from sampling error, non-response bias, measurement error, and occasional changes in survey methodology.

6. Conclusion

FLO-NAVSAFE serves as a proof-of-concept for a scalable decision support tool that distills forecast science into actionable insights for emergency responders by incorporating their needs into the design process. Future work should focus on refining automation, reducing latency, and enhancing ground-truthing methods for evaluation. FIM generation could also be improved by replacing HAND with better low complexity models such as Kansas' FLDPLN (Floodplain) model for nowcasts [35]. In addition, incorporating work on the densified network using NWM runoff, Muskingum routing, and data assimilation (as seen in Chapter 5) would improve the precision of forecasts for local decision making. We would also recommend incorporating data from other existing tools used by Travis County such as ATX Floods to make the platform more robust and locally relevant.

Acknowledgements:

We acknowledge Dean Djokic (*ESRI*) who contributed particularly to the development of the FIM generation pipeline. We also acknowledge Dr. Anupal Baruah (*The University of Alabama*), James Halgreen (*Alabama Water Institute*), Fred Ogden, Carson Pruitt and Greg Petrochenkov (*NOAA*), Dan Tian (*The University of Alabama*) and Dipsikha Devi (*The University of Alabama*) for their expertise and support. In addition, we acknowledge the TCSO for their insights and support. We are also grateful to Lanna Nations (*Alabama Water Institute*), Parvaneh Nikrou and Francisco Gomez (*The University of Alabama*), Julia Masterman, Jordan Read and all the staff at CUAHSI for making this Summer Institute happen. Lastly, we acknowledge the support of our academic advisors: Sushant Mehan: *South Dakota State University*, Paulina Jaramillo: *Carnegie Mellon University*, Sarah Null: *Utah State University*, Hamed Moftakhari: *The University of Alabama*

This research was supported by the Cooperative Institute for Research to Operations in Hydrology (CIROH) under award NA22NWS4320003 from the NOAA Cooperative Institute Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA.

We acknowledge the use of ChatGPT, an AI language model developed by OpenAI, for helping to refine sections of our code and improve the overall readability of our writing. Specifically, we consulted ChatGPT for suggestions on code optimization, as well as on stylistic and structural enhancements to our drafts. All final decisions regarding the content and interpretation of the research remain our own.

Supplementary Materials:

The data used in this report, supplementary information documents containing additional methodological details as well as supporting results can be found here: http://www.hydroshare.org/resource/41b23520d92c4d6e8d7934f106e54fd3, http://www.hydroshare.org/resource/23aa7866ab614687811bb70ffb13fcfe and http://www.hydroshare.org/resource/c95e654312204ce0b4d8e31e71cd4354.

All Python notebooks and scripts created for this project can be found here: https://github.com/Kaysharp-cloud/Flo_NAVSAFE

References

[1] U. S. J. E. Committee, "Flooding Costs the U.S. Between $179.8 and $496.0 Billion Each Year - Flooding Costs the U.S. Between $179.8 and $496.0 Billion Each Year - United States Joint Economic Committee." Accessed: July 15, 2025. [Online]. Available: https://www.jec.senate.gov/public/index.cfm/democrats/2024/6/flooding-costs-the-u-s-between-179-8-and-496-0-billion-each-year

[2] L. Chu *et al.*, "Floods and cause-specific mortality in the United States applying a triply robust approach," *Nat. Commun.*, vol. 16, p. 2853, Mar. 2025, doi: 10.1038/s41467-025-58236-0.

[3] N. US Department of Commerce, "Severe Weather Awareness Week--Flash Flood Safety." Accessed: July 15, 2025. [Online]. Available: https://www.weather.gov/shv/awarenessweek_severe_flashflood

[4] Z. Han and H. O. Sharif, "Vehicle-Related Flood Fatalities in Texas, 1959–2019," *Water*, vol. 12, no. 10, Art. no. 10, Oct. 2020, doi: 10.3390/w12102884.

[5] "Senior Deputy Jessica Laura Hollis," The Officer Down Memorial Page (ODMP). Accessed: July 04, 2025. [Online]. Available: https://www.odmp.org/officer/22194-senior-deputy-jessica-laura-hollis

[6] B. Cosgrove *et al.*, "NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling," *JAWRA J. Am. Water Resour. Assoc.*, vol. 60, no. 2, pp. 247–272, 2024, doi: 10.1111/1752-1688.13184.

[7] A. Heldmyer, B. Livneh, J. McCreight, L. Read, J. Kasprzyk, and T. Minear, "Evaluation of a new observationally based channel parameterization for the National Water Model," *Hydrol. Earth Syst. Sci.*, vol. 26, no. 23, pp. 6121–6136, Dec. 2022, doi: 10.5194/hess-26-6121-2022.

[8] S. Cohen, S. Praskievicz, and D. R. Maidment, "Featured Collection Introduction: National Water Model," *JAWRA J. Am. Water Resour. Assoc.*, vol. 54, no. 4, pp. 767–769, Aug. 2018, doi: 10.1111/1752-1688.12664.

[9] "Public Handbook V2.3.1: NWC Visualization Services." Office of Water Prediction, June 24, 2025. Accessed: July 17, 2025. [Online]. Available: https://www.weather.gov/media/owp/operations/Public_Handbook_NWC_Visualization_Services.pdf

[10] A. Baruah *et al.*, "FIMserv v.1.0: A tool for streamlining Flood Inundation Mapping (FIM) using the United States operational hydrological forecasting framework," *Environ. Model.*

Softw., vol. 192, p. 106581, Aug. 2025, doi: 10.1016/j.envsoft.2025.106581.

[11] "Pin2Flood." Accessed: July 15, 2025. [Online]. Available: https://gis.tdem.texas.gov/portal/apps/storymaps/stories/72f0ec81a7654da688518f486122abed

[12] K. Larco and S. Mahmoudi, "National Water Center Innovators Program Summer Institute Report 2024," 2024.

[13] "Travis County, Texas - Census Bureau Profile." Accessed: July 15, 2025. [Online]. Available: https://data.census.gov/profile/Travis_County,_Texas?g=050XX00US48453

[14] T. S. H. Association (TSHA), "Travis County," Texas Almanac. Accessed: July 15, 2025. [Online]. Available: https://www.texasalmanac.com/places/travis-county

[15] "Colorado River Watershed." Accessed: July 15, 2025. [Online]. Available: https://www.traviscountytx.gov/tnr/environmental-quality/water-quality/colorado-river-watershed

[16] "Water FAQs." Accessed: July 15, 2025. [Online]. Available: https://www.traviscountytx.gov/tnr/environmental-quality/water-faqs#flooding

[17] S. Carpitella, V. Kratochvíl, and M. Pištěk, "Multi-criteria decision making beyond consistency: An alternative to AHP for real-world industrial problems," *Comput. Ind. Eng.*, vol. 198, p. 110661, Dec. 2024, doi: 10.1016/j.cie.2024.110661.

[18] J. Oh and M. Bartos, "Flood early warning system with data assimilation enables site-level forecasting of bridge impacts," In review.

[19] S. Joslyn and J. LeClerc, "Decisions With Uncertainty: The Glass Half Full," *Curr. Dir. Psychol. Sci.*, vol. 22, no. 4, pp. 308–315, Aug. 2013, doi: 10.1177/0963721413481473.

[20] T. Kox, L. Gerhold, and U. Ulbrich, "Perception and use of uncertainty in severe weather warnings by emergency services in Germany," *Atmospheric Res.*, vol. 158–159, pp. 292–301, May 2015, doi: 10.1016/j.atmosres.2014.02.024.

[21] S. Timilsina and P. Passalacqua, "A comparative analysis of national water model versions 2.1 and 3.0 reveals advances and challenges in streamflow predictions during storm events," *J. Hydrol. Reg. Stud.*, vol. 58, p. 102196, Apr. 2025, doi: 10.1016/j.ejrh.2025.102196.

[22] J. A. Regina and A. Raney, *OWPHydroTools*. (Apr. 2025). Python. Accessed: July 17, 2025. [Online]. Available: https://github.com/noaa-owp/hydrotools

[23] D. Devi *et al.*, "A Framework for the Evaluation of Flood Inundation Predictions Over Extensive Benchmark Databases," June 27, 2025, *Preprints*. doi: 10.22541/essoar.175105672.23211828/v1.

[24] U. C. Bureau, "Understanding and Using American Community Survey Data: What All Data Users Need to Know," Census.gov. Accessed: July 17, 2025. [Online]. Available: https://www.census.gov/programs-surveys/acs/library/handbooks/general.html

[25] S. L. Cutter, *Hazards Vulnerability and Environmental Justice*. London: Routledge, 2012. doi: 10.4324/9781849771542.

[26] J. Bailie, V. Matthews, R. Bailie, M. Villeneuve, and J. Longman, "Exposure to risk and experiences of river flooding for people with disability and carers in rural Australia: a cross-sectional survey," Aug. 2022, doi: 10.1136/bmjopen-2021-056210.

[27] H. Hao, Y. Wang, and S. Kang, "Examining 'digital' vulnerability to flooding among subsidized housing residents in Florida," *Int. J. Disaster Risk Reduct.*, vol. 82, p. 103302, Nov. 2022, doi: 10.1016/j.ijdrr.2022.103302.

[28] D. Rhubart and Y. Sun, "The social correlates of flood risk: variation along the US rural–urban continuum," *Popul. Environ.*, vol. 43, no. 2, pp. 232–256, Dec. 2021, doi: 10.1007/s11111-021-00388-4.

[29] F. Aziz, X. Wang, M. Q. Mahmood, and O. Juliette, "Assessing human health risks associated with wastewater flooding," *Environ. Impact Assess. Rev.*, vol. 115, p. 108031, Aug. 2025, doi: 10.1016/j.eiar.2025.108031.

[30] Z. Liu and A. Mostafavi, "Collision of Environmental Injustice and Sea Level Rise: Assessment of Risk Inequality in Flood-induced Pollutant Dispersion from Toxic Sites in Texas," Apr. 27, 2023, *arXiv*: arXiv:2301.00312. doi: 10.48550/arXiv.2301.00312.

[31] O. US EPA, "FRS Data Sources." Accessed: July 15, 2025. [Online]. Available: https://www.epa.gov/frs/frs-data-sources

[32] O. US EPA, "Facility Registry Service." Accessed: July 15, 2025. [Online]. Available: https://www.epa.gov/frs

[33] J. M. Johnson, D. Munasinghe, D. Eyelade, and S. Cohen, "An integrated evaluation of the National Water Model (NWM)–Height Above Nearest Drainage (HAND) flood mapping methodology," *Nat. Hazards Earth Syst. Sci.*, vol. 19, no. 11, pp. 2405–2420, Nov. 2019, doi: 10.5194/nhess-19-2405-2019.

[34] Dynamic FIM Services Factsheet [Public Domain]," 2024.

[35] D. Weiss, J. Song, J. Kastens, X. Li, J. Halgren, and J. Edwards, "Evaluation of FLDPLN Model Performance for Operational Flood Inundation Mapping in and beyond Kansas," presented at the AGU24, AGU, Dec. 2024. Accessed: July 17, 2025. [Online]. Available: https://agu.confex.com/agu/agu24/meetingapp.cgi/Paper/1741015

Chapter 6:
Estimating Flood Inundation Using a Densified Stream Network in Travis County, TX

Victor Oladoja[1], Alex Simpson[2], Megan Vardaman[3], David Maidment[4*], and Kelsey McDonough[5*]

[1] University of Connecticut; *victor.oladoja@uconn.edu*

[2] University of Oregon; *asimps10@uoregon.edu*

[3] University of New Hampshire; *megan.vardaman@unh.edu*

[5*] University of Texas, Austin: maidment@utexas.edu

[6*] The Water Institute: mcdonough.kelseyr@gmail.com

* Theme Leader

Abstract: Accurate flood forecasting is critical for hazard response and the mitigation of human, infrastructure, and economic losses. The National Water Model (NWM), developed by the NOAA Office of Water Prediction, National Water Center, provides nationwide hydrologic forecasts, including flood inundation extents. However, its stream network, derived from a 10 m resolution digital elevation model, is too coarse for community-level applications, particularly in urban settings such as Travis County, Texas. Here, we investigate how stream network density impacts simulated discharge and inundation extent and quantify the potential benefits of downscaling NWM runoff outputs onto a higher-resolution stream network. While only ~27% of low-water crossings in the county lie within 45 m of stream reaches in the NWM network, ~89% are captured within the same proximity in the densified network. Modeled streamflows were produced for reaches in the original NWM and the densified stream networks within the Shoal and Waller Creek watersheds by routing NWM runoff outputs using Muskingum routing and data assimilation schemes. For a heavy rainfall event on July 4 – 7, 2025, streamflow estimates from the densified network model (KGE = 0.31) are more similar to observed gauge data (withheld from data assimilation) than those from the original NWM network model (KGE = -0.27). During the peak of the event, the flood inundation extent estimated using the NWM network model was 0.35 km² (total reach length = 32.6 km), compared to 1.04 km² (total reach length = 69.3 km) for the densified network model. While inundations for the NWM network were smaller in some locations and larger in others, the densified network produced smoother and more continuous inundation patterns.

---

1. Motivation

Flooding is the deadliest severe weather hazard in the United States, costing between $179.8 billion and $496 billion each year [1], [2]. Texas experiences a higher number of floods and vehicle-related flood fatalities than any other U.S. state [3], [4]. Many factors contribute to the increased number of fatalities, including the physical environment (e.g., geographic features, weather patterns, topography), highly urbanized areas, fast population growth, and an extensive road network, which includes culverts, bridges, and low water crossings [3], [4], [5].

These risks are particularly apparent to first responders in Texas who often have to navigate around dangerous flood conditions during active storms and/or after dark. In 2014, Senior Deputy Jessica Hollis of the Travis County Sheriff's Department was washed away in her patrol car while checking roadways for high water [6]. Tragic accidents such as this emphasize the need for accurate and timely forecasting of floods, paired with communication of the associated risks, to mitigate associated catastrophic human, property, and economic losses.

The accuracy of flood inundation forecasts depends greatly on the reliability of the hydrologic and hydraulic modeling framework. A range of hydrologic models has emerged in recent years, spanning both process-based and data-driven approaches [7], [8]. Among these is the National Water Model (NWM), which provides streamflow forecasts for over 2.7 million river reaches in CONUS [9]. Although its spatial coverage is extensive, the current NWM stream network produces data that is too coarse for real-time decision-making at the community scale. In this report, we employ NWM runoff, Muskingum routing, and Kalman Filter data assimilation to estimate streamflow for a portion of the Travis County densified stream network (**Figure 1**) to model flooding at the street scale.

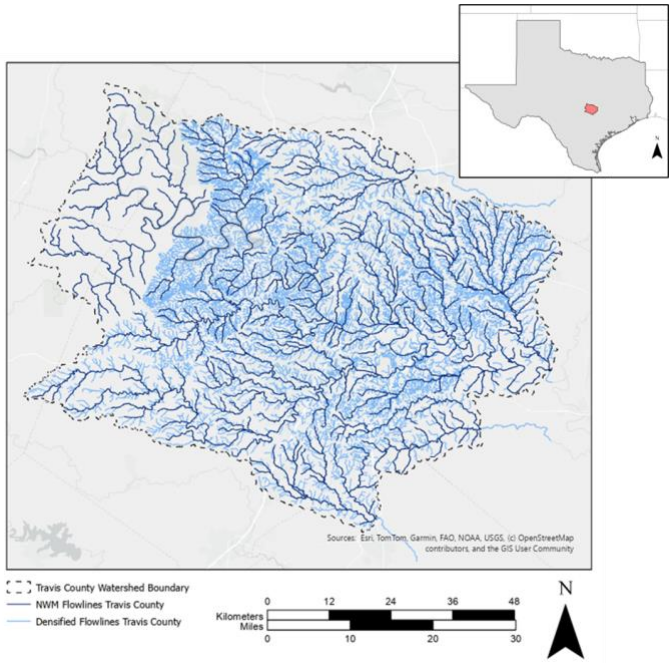


Figure 1. *Map of NWM stream network and densified stream network for the watersheds in Travis County, Texas.*

2. Objectives and Scope

This study aims to downscale NWM streamflow to a higher resolution stream network to improve flood response capabilities in urban areas where greater spatial detail is needed for emergency planning and response. Travis County, Texas, was selected as the study area due to (1) its vulnerability to flash flooding and (2) large number of low water crossings on small streams outside of the NWM network in the county. Our study addresses the research question: How does stream network density influence discharge and flood inundation extent during heavy rainfall or flood events? To investigate this, a densified stream network was prepared for the Shoal and Waller Creek watersheds in Travis County. Discharge outputs from the NWM analysis and assimilation open-loop configuration were used to drive Muskingum routing and Kalman Filter data assimilation schemes and applied to two versions of the stream network: the original NWM network and the densified network for Shoal Creek and Waller Creek. By comparing

simulated stream discharges in each network against nearby observed gauge data from the City of Austin's Flood Early Warning System (FEWS), and flood inundation maps (FIMs) generated using the Height Above Nearest Drainage (HAND) method, we assess the potential benefits of downscaling NWM runoff to the community (i.e., county) scale.

3. Previous Studies

Currently, flood prediction from the NOAA Office of Water Prediction (OWP) comes from the NWM. In the operational configuration of the NWM version 3.0, WRF-Hydro simulates land surface processes at a 1 km resolution, with overland and subsurface flow routed on a 250 m grid across the contiguous United States (CONUS). The Muskingum-Cunge approach [9] is used to route flows over a stream network derived from the National Hydrography Dataset Plus Version 2 (NHDPlus V2; [10]). NWM uses the Medium Resolution NHDPlus to provide streamflow for over 2.7 million river reaches in CONUS [11]. While this represents extensive spatial coverage, the 1:100,000 scale flowlines employed are often too coarse for community-scale applications. This limitation is especially critical in urban areas such as Travis County, Texas, where finer-resolution streamflow and flood inundation data are essential for effective emergency response and preparedness during flood events.

NWM produces both data assimilation and open-loop analyses, which incorporate the best available observed forcings to drive simulations [12]. The nudging-based data assimilation approach used in the NWM has been shown to only improve streamflow estimates downstream of gauge stations [13]. Previous work introduced an alternative data assimilation framework that combines Muskingum routing with Kalman filtering, which demonstrated superior performance compared to the current nudging approach employed by the NWM [14]. Our research does not aim to evaluate the performance of the Muskingum routing or data assimilation techniques, as these have been established in previous works [15], [16], [17], [18]. Instead, it builds on previously developed schemes [14] and establishes a methodology for applying coarse-scale runoff from NWM catchments to a densified stream network, and assesses the potential benefits of downscaling NWM runoff to the community (i.e., county) scale. To the best of the authors' knowledge, no such comparison has been made before.

4. Methodology

This study applies the routing and data assimilation schemes developed in previous works to a modified subsection of a high-resolution, densified stream network generated by the Texas Strategic Coordinating Office (TSCO), using overland flow and baseflow outputs from the NWM open-loop configuration as forcings. To evaluate differences between the NWM's native stream network and the densified network, FIMs were derived for modeled peak flows using synthetic rating curves (SRCs) via the HAND method. New SRCs and flood data stack were generated for the densified stream network, while existing SRCs produced for the University of Texas at Austin's Center for Water & the Environment and the Texas Division of Emergency Management's Pin2Flood project [19] were used for the NWM stream network. The workflow for generating FIM on the densified network is summarized in **Figure 2**. The same workflow was used to prepare the NWM network model, excluding the calculation of area fraction.

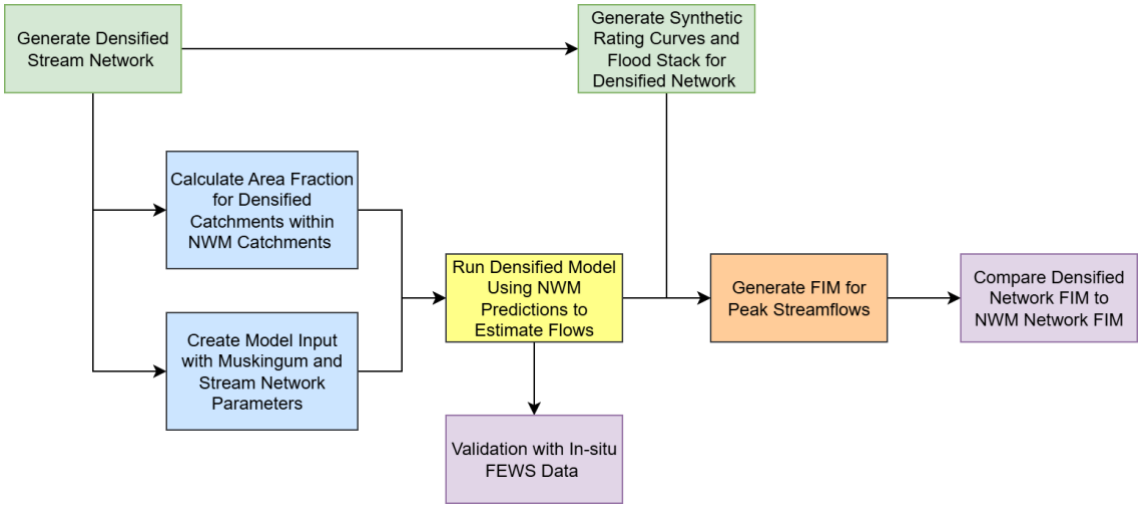


Figure 2. *Overall workflow for generating FIM on the densified stream network.*

4.1. Data and study area

Travis County is located in Central Texas (**Figure 3**) and encompasses an area of ~395 km² (~1,023 mi²) [20] with a population of ~1.31 million people [21]. Austin, TX, the state capital and a major metropolitan area, is located within the county. We created FIM for four NWM stream reaches in Shoal and Waller Creek located in downtown Austin. A stream network of 1510 reaches (henceforth referred to as the "more-densified" network) was produced based on known locations of infrastructure like sewers and culverts from the City of Austin, ground survey data from the Texas Department of Transportation (TxDOT), and prepared for analysis by Dr. Matt Bartos (*University of Texas at Austin*).

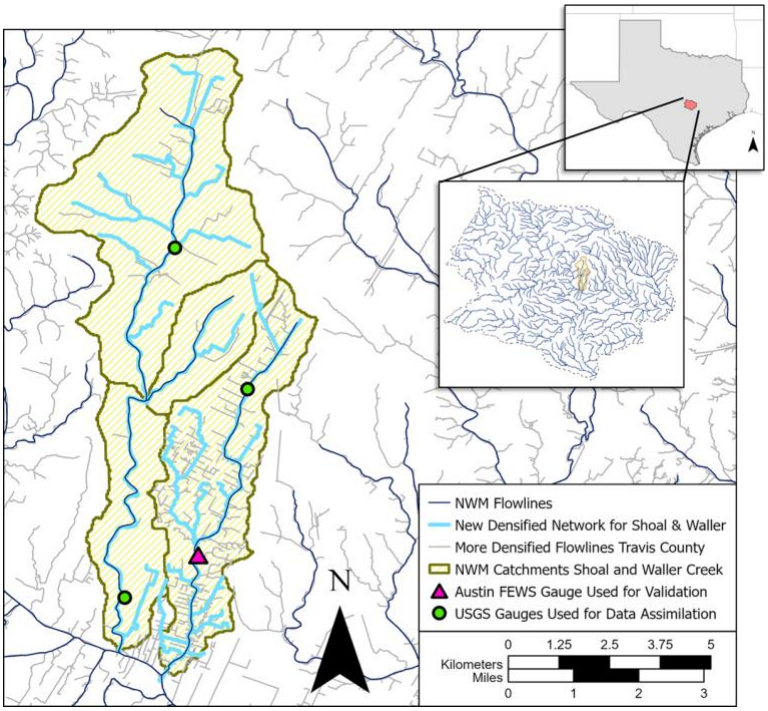


Figure 3. *Map of NWM stream network and densified stream network for the Shoal and Waller Creek watersheds Locations of USGS gauges and Austin Flood Early Warning System gauges used in the analysis are also shown.*

Shoal and Waller creeks are highly urbanized catchments. Only 16% of stream reaches in Shoal and Waller creeks are natural streams (62% across the entire network). The rest are artificial flow paths, road-side ditches, ponds, and 63% are underground culverts and storm drains (19% across the entire network). For this study, these underground flow paths were assumed to behave as open channels for FIM. This data is displayed in **Table 1.**

Table 1. *More-Dense Network Stream Reach Classifications*

Stream Classification	Shoal and Waller Creeks	Entire Network
Altered Earthen	112	5268
Altered Road-Side Ditch	36	1976
Altered Structural	54	1676
Artificial Flowpath	71	625
Concentrated Flowpath	3	593
Natural Alluvial	71	1604
Natural Rockbed	152	1308
Natural Unclassified	18	36652
Pond	47	936
Underground Culvert	158	9525
Underground Stormdrain	788	2472
Waterbody Centerline	0	679

Instantaneous discharge data from 3 U.S. Geological Survey (USGS) stream gauges, accessed using the dataretrieval package in Python [22], were used for data assimilation. Model outputs are validated with stage data from a City of Austin Flood Early Warning System (FEWS) gauge. **Table 2** summarizes the USGS gauges used for data assimilation and the FEWS gauge used for validation. For this study, we model a flood event on July 5th, 2025.

Table 2. *Summary of gauges used for data assimilation and validation.*

Data Provider	Gage ID	Use
USGS	08156675	Data Assimilation
USGS	08156800	Data Assimilation
USGS	08156910	Data Assimilation
Austin FEWS	248	Validation

4.2. Updating NWM Runoff Using Muskingum Routing and Kalman Filter

The Muskingum model and data assimilation code were prepared by the Future Water Systems Lab and are available at: https://github.com/future-water/tx-fast-hydrology. Runoff (i.e., streamflow, qBucket, and qSfcLatRunoff) from the National Water Model open-loop analysis and assimilation product was routed through the densified network using the Muskingum method. The current outflow, $Q_{j+1}^{t+\Delta t}$, at a reach j is a function of the previous timestep outflow (Q_{j+1}^{t}), current inflow ($Q_{j}^{t+\Delta t}$), previous timestep inflow (Q_{j}^{t}), and lateral inflow (q_{j}^{t}), given by the equation:

$$Q_{j+1}^{t+\Delta t} = \alpha_j Q_j^{t+\Delta t} + \beta Q_j^t + \chi Q_{j+1}^t + \gamma q_j^t \quad (1)$$

The variables $\alpha, \beta, \chi,$ and γ are described by:

$$\alpha = \frac{KX + \Delta t/2}{K(1-X) + \Delta t/2} \quad (2)$$

$$\beta = \frac{KX - \Delta t/2}{K(1-X) + \Delta t/2} \quad (3)$$

$$\chi = \frac{K(1-X) - \Delta t/2}{K(1-X) + \Delta t/2} \quad (4)$$

$$\gamma = \frac{\Delta t}{K(1-X) + \Delta t/2} \quad (5)$$

where K is the flood wave travel time in the reach [s], Δt is the simulation time delta, and X is the wave attenuation constant.

$$K = \Delta x / \underline{c} \quad (6)$$

$$X = 0.5\,[\,1 - \underline{Q}\,/\,(\underline{c}\,\underline{B}\,S\,\Delta x)\,] \quad (7)$$

where Δx is the length of the reach, $\underline{c}$ is the kinematic wave celerity (2.12 m/s), $\underline{Q}$ is discharge, $\underline{B}$ is the cross-sectional top-width associated with $\underline{Q}$, and S is the energy slope approximated by the water surface slope [17], [18], [23]. For a large system of reaches, the Muskingum equations are expressed in the state-space form:

$$x^{t+\Delta t} = Ax^t + Bu^t \quad (8)$$

A deterministic Kalman Filter approach is applied to the state-space model at the start of each forecast window to correct the runoff input. A Kalman Filter minimizes the mean squared error between the true and estimated states by assuming that measurement (i.e., gauge data) and process (i.e., model) noise follows a Gaussian distribution with known covariances, and that the initial error covariance is known. The equation for the state-space Muskingum routing model with data assimilation is:

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + L_{k+1}(y_{k+1} - \hat{y}_{k+1}) \quad (9)$$

where L is the Kalman gain, y is the sensor measurement, and $\hat{y}$ is the modeled output.

4.3. Preparation of the Densified Network Model

Flowlines in the more-densified network were separated into individual segments based on their channel and infrastructure type rather than the contributing watershed. Many flowlines were less than 1 km in length (**Figure 4**). To support rating curve generation, short channel segments are aggregated by first identifying headwater locations for all known streams and then manually selecting origination points to combine shorter streams to create an overall stream density that is valid for floodplain modeling purposes. We chose 2 km/km^2 (3 mi/mi^2) for this study, but the appropriate density for other networks should be selected by the analyst.

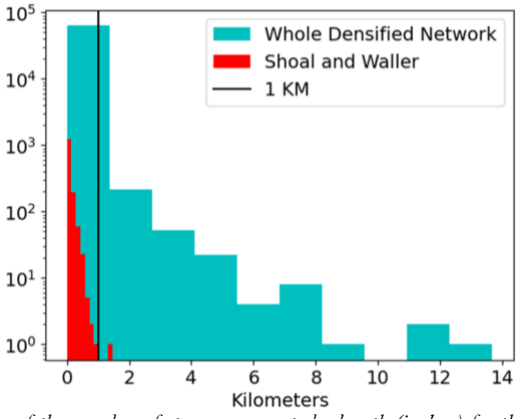


Figure 4. *Histogram of the number of stream segments by length (in km) for the more-densified network.*

Flowpaths are derived from a Texas Geographic Information Office (TxGIO) digital elevation model (DEM) of ~0.3 m (~1 ft) spatial resolution [24]. The DEM is hydroconditioned using a technique provided by Dr. Dean Djokic (ESRI). Since the more-densified network was derived from surveyed data for pipe locations rather than a DEM, extensive hydro-conditioning was required to use this network for FIM generation. As part of the hydro-conditioning, the streamlines were "burned" into a bare-earth LIDAR DEM to weight flow routing in favor of the network. Flow direction and accumulation are calculated, and flow routed through this network from the starting nodes. A length limit of ~914 m (3,000 ft) was specified for creating the vectorized streamlines. Stream reaches were decreased from 1510 to 58.

The aggregated flowlines (henceforth referred to as "densified network") were used to create new model input files to run the Muskingum routing model. The first of these inputs described the relationship between the densified network catchments and the NWM catchments based on the fraction of the NWM catchment occupied by the densified catchment. Catchment areas were

computed in ArcGIS Pro and the "tabulate intersection" tool was used to calculate the percent area for each sub-catchment. NWM runoff was partitioned to the densified catchments within each NWM watershed based on their fractional area prior to Muskingum routing.

The second required input file described the stream network topology (i.e., channel length) and spatial locations, Muskingum parameters (K and X), and connections between stream channels. The appropriate timedelta for the model was determined using the stability condition:

$$\Delta t \leq 2K(1-X) \quad (10)$$

where X is defined such that $0 < X \leq 0.5$. We adopted $X = 0.25$ for every reach. Due to differences in the minimum channel length in each network, the model timedelta was set to 120 seconds for the NWM network input and to 20 seconds for the densified network input. Input files created for other stream networks should include computation of an appropriate timedelta using this stability condition. USGS gauge data used for data assimilation were resampled to match the model simulation's time step.

4.4. Development of Synthetic Rating Curves and the Flood Data Stack for the Densified Network

The methodology for developing the SRCs and flood stack for the densified stream network was provided by Dr. Dean Djokic (ESRI). Flood extents were generated at incremental heights derived from the DEM using the HAND methodology. SRCs were generated from the HAND raster using the default values for slope (0.0015) and roughness (Manning's n = 0.045). The outputs from this process were a set of SRCs for unique flowlines within the densified network and a flood data stack containing flood inundation extent polygons for each flowline for a range of depth values (as integers) between ~0.3 m and 24 m (1 and 80 ft).

Flood inundation extent for the modeled streamflows was estimated using the SRCs and the flood data stack. Peak streamflows corresponding to each unique stream segment over the selected event period were identified from the model output and converted to stage height using the corresponding SRC. The stage height for each stream segment was rounded up or down to the nearest integer value and used to select the flood inundation polygon from the flood data stack that best matched the stage height for each stream segment. Flood inundation polygons corresponding to the stage height in each stream segment were combined into a single polygon layer of flood inundation (i.e., FIM).

5. Results

5.1. The Densified Network

Estimation of flooding at low water crossings is of particular importance in Travis County due to their contribution to road flooding and risk to drivers during flooded conditions [3]. However, many recorded flood events and low-water crossings in Travis County do not fall near NWM flowlines. Only 27% of low-water crossings (**Figure 5c**) and 1% of reported flood incidents (**Figure 5d**) are within 45 m (150 ft) of a NWM stream reach. In contrast, 89% of low-water crossings and 11% of reported flood incidents are within 45 m (150 ft) of a more-densified stream reach (**Figure 5c, 5d**). The mean distance from a low-water crossing to a streamline is approximately 846 m for the NWM network and 60 m for the more-dense network (**Figure 5a**). The mean distance for reported flood incidents is approximately 1338 m for the NWM network and 420 m for the more-dense network.

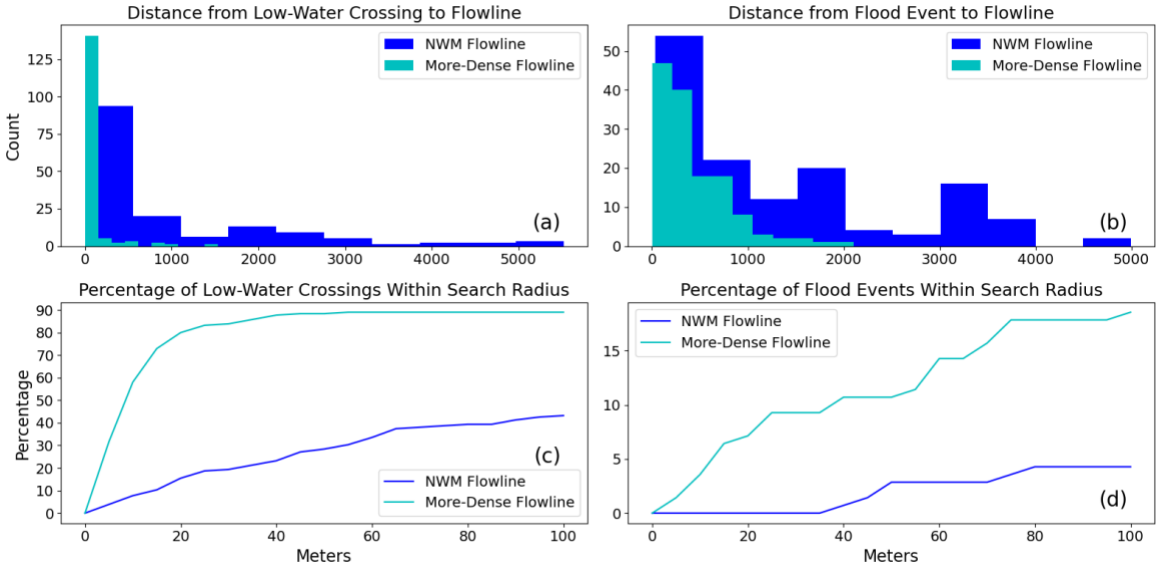


Figure 5. *Distance from low water crossings (left) and flood events (right) from nearest NWM flowlines (blue) and Densified Network flowlines (light-blue) in Travis county, Texas. Flood event records were provided by the Travis County Sheriff's Office (TCSO) for 2015, 2016, 2022, 2023, 2024, and the May 28th and July 4th weekend floods.*

5.2. Performance of the Densified Model Streamflow Estimation

Austin FEWS data was withheld from data assimilation and used to compare modeled and measured streamflows. Most FEWS gauges record stage height, but no measured rating curves are available to convert them to streamflow. One of the FEWS gauges in the Shoal and Waller Creek study region (FEWS ID 248) recorded both stage and discharge. **Figure 6** compares the modeled vs measured discharge for the nearest stream reach for each network corresponding to this gauge. The resulting Kling-Gupta efficiency (KGE) for the NWM network model and densified network model are -0.27 and 0.31, respectively. KGE values greater than -0.41 indicate that the model performance is better than using the mean of the observations, and a KGE closer to 1 indicates a better match between modeled and observed data [25]. For the heavy rainfall event on July 4 – 7, 2025 event considered for this comparison, both KGE values were greater than -0.41 and the densified network reach showed improved performance over the NWM network reach compared to the observations.

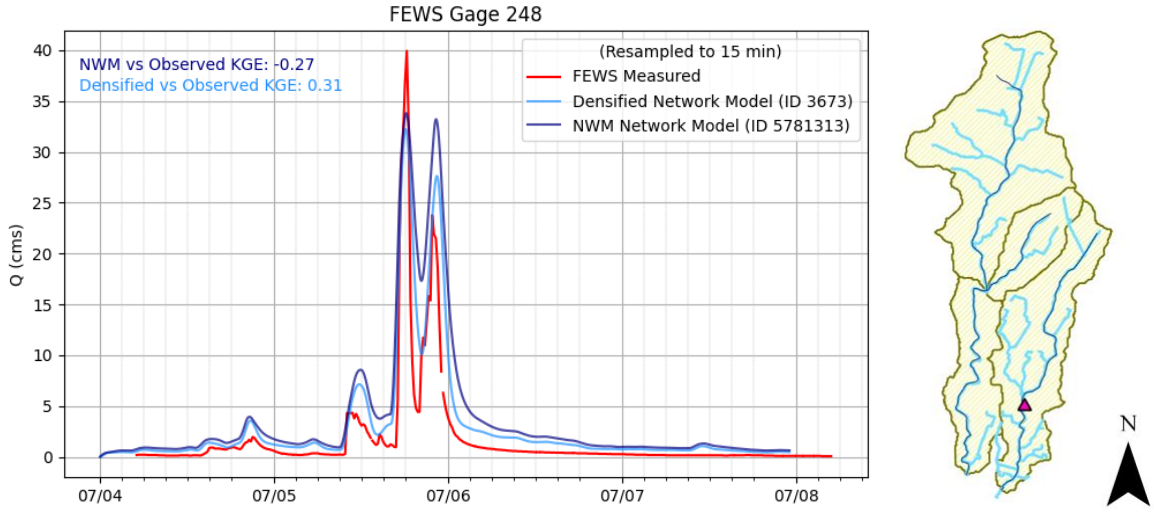


Figure 6. *Comparison between measured discharge at FEWS gauge ID 248 (red), compared to modeled discharge from the nearest stream reach in the densified network model (light blue) and NWM network model (dark blue).*

5.3. Comparison of NWM FIM and Densified Network FIM

During the heavy rainfall event on July 4 – 7, 2025, the NWM network model produced a smaller inundation extent compared to the densified network model for peak flows in Shoal and Waller Creeks (**Figure 7a-d**). Within the study area, the total stream length in the NWM network is 32.6 km, while that of the densified network is 69.3 km. For the peak flow event, the resulting flood inundation extent was 0.35 km² using the NWM network and 1.04 km² using the densified network. Thus, the NWM-based inundation extent represents approximately 67% of that produced by the densified network, while covering only about 53% of the densified network's total stream reach length. The flood inundation extent derived from the NWM network is, in some locations, larger than that from the densified network (e.g., **Figure 7b**), and in other areas, smaller (e.g., **Figure 7c** middle right panel). Overall, the inundation patterns produced using the densified network appear smoother and more continuous, likely due to the use of higher-resolution topographic data (i.e., 0.3 m DEM rather than 3 m DEM).

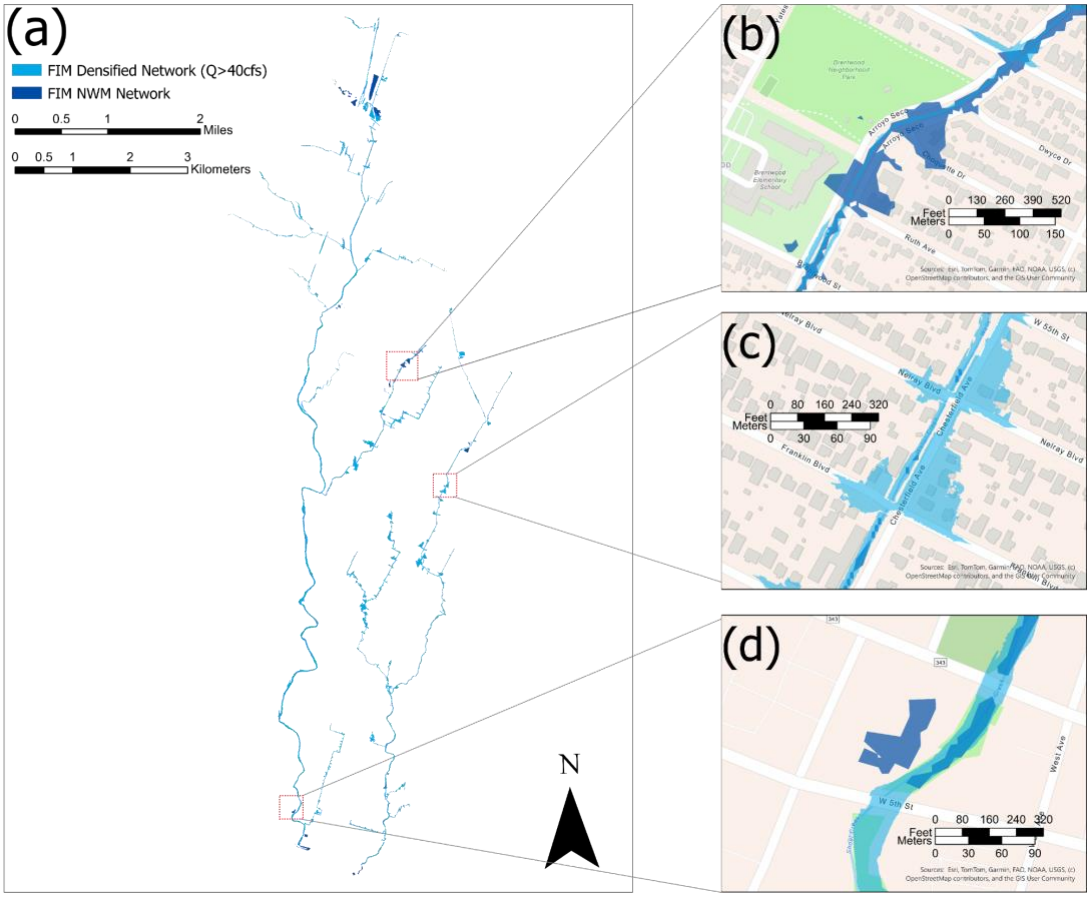


Figure 7. *FIMs generated using the densified stream network overlaid on the NWM stream network for the July 4–7, 2025 heavy rainfall event in Shoal and Waller Creek (a). The right panel (b-d) presents zoomed-in views of selected locations, illustrating differences in spatial extent and alignment between the two networks.*

5.4. Limitations

5.4.1. Stream Network Creation

This work demonstrated the potential to combine an existing high-resolution stream network using DEM-created flowpaths to create a densified stream network suitable for modeling streamflows at the community scale. However, it is still unclear what level of densification provides the best tradeoffs between computational time and model performance; future work should explore this. The fact that the more-densified network was not derived from a DEM was a serious limitation to its practicality. Flow paths derived from the DEM, which were needed for flood-stack generation, did not always follow the surveyed network, even with heavy hydro-conditioning biasing flow routing through it. The creation of future densified networks must be heavily informed by DEM-derived flow routing. Ongoing efforts to generate more detailed stream networks based on high-quality DEM data, such as the USGS 3D Hydrography Program (3DHP) [26], present a promising alternative for preparing a densified stream network for modeling streamflows and flood inundation. An additional consideration is what types of flow paths should be included in the modeling streams. The more-densified network included culverts, underground storm drainage, and artificial flow paths. The inclusion of these flowpaths requires different assumptions about flow velocity, roughness, and flood inundation compared to natural surface flowpaths. For this study, all densified flowpaths were assumed to behave as open channels, but this assumption should be evaluated by future work.

Refining the original, more-densified network to the densified network used in this study included several steps that could not be automated. Subsetting of the origination points for each stream reach was done manually to select the desired network density, meaning that the output will vary somewhat based on the analyst who performs it. The same consideration applies to the selection of stream split points at confluences and along longer reaches. Selection of channel segment length can impact the performance of SRCs and the outcome of the HAND method [27]. Channel segments that are too short can lead to catchments of impractical size, while those that are too long suggest unrealistic assumptions of uniform discharge and water depth along a channel [28]. Despite this, there is limited guidance on the preferred channel segment length for deriving SRCs. SRC performance is also decreased in urbanized environments [29], and local changes in terrain or flowpaths due to human activity (e.g., construction, flood control) may not be represented by HAND rasters used to estimate flood inundation.

5.4.2. Synthetic Rating Curves and Use of the HAND Method for FIM Generation

The HAND method is recommended for flood mapping that is aimed at guidance and not cell-by-cell decision making [30]. The quality of topographic data and accuracy of friction coefficients in the SRCs affect the performance of simulated flood inundations, particularly in flat terrain [31]. The generation of SRCs is the source of the most error. Currently, the skill of the HAND method is more limited by the terrain and SRCs than by NWM accuracy. The HAND method tends to underestimate flooded area at the floodplain level and overestimate at the catchment scale, likely due to underprediction in lower-order reaches [30].

HAND only produces inundation from the nearest flowpath and is unable to source inundation from multiple fluvial sources [32]. This is a key limitation for a densified stream network where flowlines may be closer together. Furthermore, the current HAND method projects the water surface from the nearest stream across the landscape, disregarding hydraulic conductivity [33], the travel time of water coming out of the bank, and the wave propagating to the edge of the floodplain [34]. This is particularly important for operational flood forecasting in urban areas,

where this lag may result in overestimation of flood inundation at the predicted timestep, therefore decreasing the reliability of flood warnings and communication.

5.4.3. FIM Comparison Between Networks

While the current analysis compared FIMs from the NWM network to the densified network, we did not account for uncertainties in the SRCs and the generation of the HAND resulting from the differences in DEM resolutions. The floodstacks created from HAND rasters were generated at discrete intervals so measured or modeled stage heights in each reach were rounded up or down to whole integer values to match the nearest flood inundation polygon. Due to limitations on processing time, the intervals used to generate the flood polygons at stage heights greater than 3.66 m were different in each network, so different rounding methods were used to match the modeled stage heights to the flood inundation polygons. This may result in under or overestimation of flood inundations during higher flows in the densified network model if stage heights fall between large gaps in available floodstack polygons.

While both models used the same input data to force the Muskingum routing, the NWM network model assigns catchment runoff to each corresponding reach, whereas the densified network model assigns these values to each subcatchment based on their percent contributing area to the original NWM catchment. The accuracy of sub-catchment delineation will impact the partitioning of these flows in the densified network model. Future work should explore these uncertainties to improve the comparisons between FIMs generated for each network.

5.4.3. Validation and Benchmarks

Validation of streamflows using stage height data from the Austin FEWS gauges was limited by the reliability of the SRCs generated for the modeled reaches. Measured rating curves for these gauges are not available; however, the City of Austin FloodPro website provides HEC-RAS models for each of these reaches. These models could be used to generate rating curves for each gauge location as an alternative to using SRCs. This approach should be explored to improve validation of the modeled discharge within the study region.

While we were able to produce FIM for each network, we were not able to validate these extents using benchmark data. Future work should incorporate additional data on measured flood extent, such as remote sensing observations of flooding and associated damages or measurements of high water (e.g., incident reports, crowdsourced data, high water marks). Flooded roadway records from the Travis County Sheriff's Office (TCSO) are available for Travis County, however, none of the records overlapped the watersheds used in this analysis. With sufficient resources and using the methodology outlined here, a densified network can be produced for other watersheds in Travis County that include these data.

6. Conclusion

Communicating flood hazards for operational use at the community-scale is critical for flood hazard response and preparedness. Estimation of flood inundation at a densified-network scale has the potential to provide more realistic estimates of flood impacts on roadways and infrastructure and facilitate more widespread adoption of NWM products by end users working at local and regional scales. This work demonstrates an approach for estimating flood inundation on a densified stream network in Travis County, TX, using NWM runoff, Muskingum routing, and data assimilation.

More work is needed to build upon the foundation that has been laid by this research. We recommend that future work prioritize: (1) unification of the methodology used to produce the NWM network and densified network, (2) comparison of modeled streamflows and FIM to in-situ and benchmark data, (3) expansion of the densified network model to other parts of Travis County, and (4) operationalization, communication, and visualization of densified network FIM. Resolving the differences between the two networks and performing further validation of the modeled discharge will make comparisons between flood extent in each network more meaningful. In addition, expansion of the network to the whole of Travis County will accomplish the original goals of this project to improve understanding of flood hazards at the community scale and provide more use cases for NWM that go beyond research and show its efficacy for operational use. This includes assessing the trade-offs between added computational time for densified network FIM generation and improved estimation of flood hazards for situational awareness and hazard mitigation. While the current NWM network has vastly improved our understanding of flooding across the U.S., its operational use is limited by the scale of the stream network for which it produces data. Increasing the resolution of the stream network in NWM, such as to the scale of the densified network demonstrated here, will capture smaller streams and improve operational acceptance. This additional level of detail may change the way communities respond to flood hazards in the future by providing more granular and actionable guidance. Lastly, any communication and visualization tools resulting from this work should be integrated into existing flood early warning systems and emergency operations to ensure uniform communication to first responders and other end users.

Acknowledgements:

The research team gratefully acknowledges Matt Bartos (*UT Austin*) and Dean Djokic (*ESRI*), who contributed greatly to developing the methodology and ensuring the success of this project. We also acknowledge the members of Theme 4 Goal 1 (Kayode Adebayo, Nana Oye Djan, Ali Farshid, Saide Zand) who developed the tools necessary for communicating the findings from this work to the Travis County Sheriff's Department. We are grateful for the support of the Alabama Water Institute Staff, including Lanna Nations, James Halgren, and Trupesh Patel, and the CUAHSI organizers, including Julia Masterman and Jordan Read. We are also grateful for the guidance of Jeil Oh, Ehsan Foroumandi, Fred Ogden, and Edward Clark, who greatly improved our understanding of the methodology. Lastly, the team also acknowledges the support of their academic advisors: Jennifer Jacobs (*University of New Hampshire*); Lijing Wang (*University of Connecticut*); Leif Karlstrom (*University of Oregon*).

This research was supported by the Cooperative Institute for Research to Operations in Hydrology (CIROH) under award NA22NWS4320003 from the NOAA Cooperative Institute Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA.

This work used Jestream2 at CIROH through allocation EES240087 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by U.S. National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296 [35].

We acknowledge the use of ChatGPT, an AI language model developed by OpenAI, for helping to refine sections of our code. Specifically, we consulted ChatGPT for suggestions on code optimization and stylistic enhancements to our figures. We also acknowledge the use of Grammarly (which leverages AI) for language editing and proofreading of this manuscript. Grammarly was used to improve grammar, clarity, and readability. All intellectual contributions, data interpretation, and conclusions remain the sole responsibility of the authors.

Supplementary Materials:

The code for the Muskingum model and data assimilation are available at: https://github.com/future-water/tx-fast-hydrology. The example analysis code provided by Matt Bartos for implementation and visualization of the Muskingum model and data assimilation are available at: https://github.com/future-water/cuahsi-summer-institute-2025. GIS layers for Travis County were provided by David Maidment and are available on HydroShare: http://www.hydroshare.org/resource/c95e654312204ce0b4d8e31e71cd4354. Examples of the ArcGIS Pro analysis provided by Dean Djokic are also available on HydroShare: http://www.hydroshare.org/resource/23aa7866ab614687811bb70ffb13fcfe.

The modified versions of this code used for the analysis outlined in this report are available at: https://github.com/oladojavictor/DensiFIMcation. Data used in this report are available at:

https://www.hydroshare.org/resource/a1018cc65aa341deb3b0644023d793db/.

References

[1] P. Lin, L. J. Hopper, Z.-L. Yang, M. Lenz, and J. W. Zeitler, "Insights into Hydrometeorological Factors Constraining Flood Prediction Skill during the May and October 2015 Texas Hill Country Flood Events," Aug. 2018, doi: 10.1175/JHM-D-18-0038.1.

[2] United States Joint Economic Committee, "Flooding Costs the U.S. Between $179.8 and $496.0 Billion Each Year." Accessed: July 09, 2025. [Online]. Available: https://www.jec.senate.gov/public/index.cfm/democrats/2024/6/flooding-costs-the-u-s-between-179-8-and-496-0-billion-each-year

[3] Z. Han and H. O. Sharif, "Vehicle-Related Flood Fatalities in Texas, 1959–2019," *Water*, vol. 12, no. 10, 2020, doi: 10.3390/w12102884.

[4] H. O. Sharif, T. L. Jackson, M. M. Hossain, and D. Zane, "Analysis of Flood Fatalities in Texas," *Natural Hazards Review*, vol. 16, no. 1, p. 04014016, Feb. 2015, doi: 10.1061/(ASCE)NH.1527-6996.0000145.

[5] P. Showalter and Y. Lu, "Flash Flood, Low Water Crossings and Swift Water Rescues in Texas," presented at the Applied Geography Conference 33, Dallas, Texas, 2010.

[6] "Senior Deputy Jessica Laura Hollis," The Officer Down Memorial Page (ODMP). Accessed: July 09, 2025. [Online]. Available: https://www.odmp.org/officer/22194-senior-deputy-jessica-laura-hollis

[7] J. M. Van Der Knijff, J. Younis, and A. P. J. De Roo, "LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation," *International Journal of Geographical Information Science*, vol. 24, no. 2, pp. 189–212, Feb. 2010, doi: 10.1080/13658810802549154.

[8] J. J. Gourley *et al.*, "The Flooded Locations And Simulated Hydrographs (FLASH) project: improving the tools for flash flood monitoring and prediction across the United States," *Bulletin of the American Meteorological Society*, vol. 98(2), pp. 361–372, 2017, doi: https://doi.org/10.1175/BAMS-D-15-00247.1.

[9] V. T. Chow, *Open Channel Hydraulics*. New York: McGraw-Hill, 1959.

[10] L. McKay, T. Bondelid, R. Moore, and A. Rea, "NHDPlus version 2: User Guide," US Environmental Protection Agency, 745, 2012.

[11] B. Cosgrove *et al.*, "NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling," *JAWRA Journal of the American Water Resources Association*, vol. 60, no. 2, pp. 247–272, 2024, doi: 10.1111/1752-1688.13184.

[12] B. Cosgrove and C. Klemmer, "The National Water Model," National Water Prediction Service. Accessed: July 16, 2025. [Online]. Available: https://water.noaa.gov/about/nwm

[13] B.-C. Seo, W. F. Krajewski, and F. Quintero, "Multi-Scale Hydrologic Evaluation of the National Water Model Streamflow Data Assimilation," *JAWRA Journal of the American Water Resources Association*, vol. 57, no. 6, pp. 875–884, 2021, doi: 10.1111/1752-1688.12955.

[14] J. Oh and M. Bartos, "Flood early warning system with data assimilation enables site-level forecasting of bridge impacts," (in review).

[15] E. Foroumandi, J. Oh, P. Ghaneei, and S. Timilsina, "Chapter 1: Data Assimilation for Improving Forecast Accuracy and Streamflow Prediction in Ungauged Basins," 2023.

[16] L. K. Read, D. N. Yates, J. M. McCreight, A. Rafieeinasab, K. Sampson, and D. J. Gochis, "Development and evaluation of the channel routing model and parameters within the National Water Model," *JAWRA Journal of the American Water Resources Association*, vol. 59, no. 5, pp. 1051–1066, 2023, doi: 10.1111/1752-1688.13134.

[17] C. M. Emery *et al.*, "Underlying Fundamentals of Kalman Filtering for River Network Modeling," Mar. 2020, doi: 10.1175/JHM-D-19-0084.1.

[18] D. L. Fread and K. S. Hsu, "Applicability of Two Simplified Flood Routing Methods: Level-Pool and Muskingum-Cunge," in *1993 ASCE National Hydraulic Engineering Conference*, July 1993.

[19] D. Maidment, "Pin2Flood," National Water Model. [Online]. Available: https://www.caee.utexas.edu/prof/maidment/NWM.htm

[20] Texas State Historical Association, "Travis County," Texas Almanac. Accessed: July 07, 2025. [Online]. Available: https://www.texasalmanac.com/places/travis-county

[21] "Travis County, TX | Data USA." Accessed: July 07, 2025. [Online]. Available: https://datausa.io/profile/geo/travis-county-tx?redirect=true

[22] T. O. Hodson, J. A. Hariharan, S. Black, and J. S. Horsburgh, *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). U.S. Geological Survey. [Online]. Available: https://doi.org/10.5066/P94I5TX3

[23] C. H. David, D. R. Maidment, G.-Y. Niu, Z.-L. Yang, F. Habets, and V. Eijkhout, "River Network Routing on the NHDPlus Dataset," Oct. 2011, doi: 10.1175/2011JHM1345.1.

[24] Strategic Mapping Program (StratMap), "Bexar & Travis Counties Lidar." TxGIO DataHub, Mar. 07, 2021. [Online]. Available: https://data.tnris.org/collection/?c=447db89a-58ee-4a1b-a61f-b918af2fb0bb

[25] 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," *Hydrology and Earth System Sciences*, vol. 23, no. 10, pp. 4323–4331, Oct. 2019, doi: 10.5194/hess-23-4323-2019.

[26] "3D Hydrography Program," U.S. Geological Survey. Accessed: July 17, 2025. [Online]. Available: https://www.usgs.gov/index.php/3DHP

[27] S. Dey, O. Eyelade, L. Godbout, and J. Zheng, "Chapter 5: Integrity Check of Synthetic

Rating Curves for HAND Inundation Mapping," National Water Center, Tuscaloosa, AL, Oct. 2017. doi: 10.4211/technical.20171009.

[28] C.-W. Yu, W.-J. Yang, and D. Feng, "Establishing Synthetic Rating Curves by Integrating the Height Above the Nearest Drainage with Hydrodynamic Computation," *Authorea (preprint)*, July 2024, doi: https://doi.org/10.22541/au.172206042.26373625/v1.

[29] A. Ghanghas, S. Dey, and V. Merwade, "Evaluating the reliability of synthetic rating curves for continental scale flood mapping," *Journal of Hydrology*, vol. 606, p. 127470, Mar. 2022, doi: 10.1016/j.jhydrol.2022.127470.

[30] J. M. Johnson, D. Munasinghe, D. Eyelade, and S. Cohen, "An integrated evaluation of the National Water Model (NWM)–Height Above Nearest Drainage (HAND) flood mapping methodology," *Natural Hazards and Earth System Sciences*, vol. 19, no. 11, pp. 2405–2420, Nov. 2019, doi: 10.5194/nhess-19-2405-2019.

[31] A. Rostami *et al.*, "Forecasting Flood Inundation in U.S. Flood-Prone Regions Through a Data-Driven Approach (FIER): Using VIIRS Water Fractions and the National Water Model," *Remote Sensing*, vol. 16, no. 23, Art. no. 23, Jan. 2024, doi: 10.3390/rs16234357.

[32] F. Aristizabal *et al.*, "Extending Height Above Nearest Drainage to Model Multiple Fluvial Sources in Flood Inundation Mapping Applications for the U.S. National Water Model," *Water Resources Research*, vol. 59, no. 5, p. e2022WR032039, 2023, doi: 10.1029/2022WR032039.

[33] P. D. Bates, "Flood Inundation Prediction," *Annual Review of Fluid Mechanics*, vol. 54, no. Volume 54, 2022, pp. 287–315, Jan. 2022, doi: 10.1146/annurev-fluid-030121-113138.

[34] V. Merwade, F. Olivera, M. Arabi, and S. Edleman, "Uncertainty in Flood Inundation Mapping: Current Issues and Future Directions," *Journal of Hydrologic Engineering*, vol. 13, no. 7, pp. 608–620, July 2008, doi: 10.1061/(ASCE)1084-0699(2008)13:7(608).

[35] T. J. Boerner, S. Deems, T. R. Furlani, S. L. Knuth, and J. Towns, "ACCESS: Advancing Innovation: NSF's Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support," in *Practice and Experience in Advanced Research Computing 2023: Computing for the Common Good*, in PEARC '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 173–176. doi: 10.1145/3569951.3597559.

Appendix

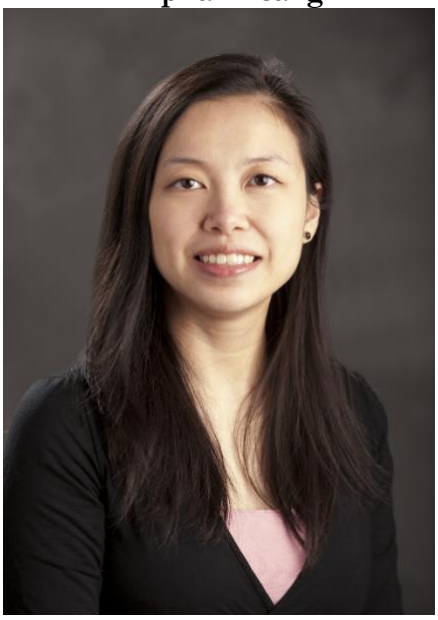
Team Leaders

Operational Flood Inundation Mapping

Sagy Cohen

"I am a Professor at the University of Alabama, with a Ph.D. from the University of Newcastle, Australia. My research interests focus on flood inundation mapping, flood remote sensing and analysis, and global and continental riverine modeling. This marks my seventh involvement in the Summer Institute, with eight of my graduate students participating to date and nine journal papers published stemming from this collaboration. In my free time, I enjoy sailing, kayaking, hiking, and traveling."

Yinphan Tsang



"I am an Associate Professor at the University of Hawai'i at Mānoa. I am an ecological modeler with a great interest in linking hydrology with ecosystems. I investigate surface and subsurface water processes and their implications in fluvial ecosystems. I use varied statistical and modeling techniques to combine disparate spatial and ecoinformatics datasets, validated with field measurements, to describe complex interaction patterns between biotic and abiotic processes. Lately, my focus has been on extreme events and their applications in Hawaii floods, as well as their ecosystem implications."

LSTM for NextGen

Jonathan Frame

"I am Assistant Professor at University of Alabama, Department of Geological Science. My research interests include large-scale hydro/hydro-dynamic modeling, machine learning for physical process representations and water resources for expanding AI infrastructure. Originally from California, I feel equally at home in Alabama. My favorite hobby is skateboarding, and I also enjoy chess, jazz, and philosophy."

Amobichukwu Amanambu

"I'm Dr. Amobichukwu Chukwudi Amanambu, Assistant Professor of Geography at the University of Alabama. I apply machine learning and AI to study hydrological drought, floodplain dynamics, and river connectivity, examining how channel networks influence flood propagation, sediment transport, and drought resilience. I integrate LiDAR and IoT data to capture river corridor changes. I enjoy kayaking."

Pluvial Urban Flooding

Marouane Temimi

"I am Gallaher Chair Associate Professor at Stevens Institute of Technology, Department of Civil, Environmental, and Ocean Engineering. My research interests include large- and small-scale hydrology, remote sensing, and water resources management. I have worked in different countries throughout my career addressing different engineering and research challenges."

Mohamed Abdelkader

"I am a Postdoctoral Researcher at IIHR—Hydroscience and Engineering, University of Iowa. My research interests include real-time flood forecasting, remote sensing of hydrologic processes, and the application of informatics in hydrology. Having experienced several natural hazards firsthand, I am committed to advancing predictive systems to improve forecasting lead time and support disaster preparedness."

***Dr. Jonathan Frame also participated in this team**

Visualization of Urban Road Flooding

David Maidment

"I am a Professor Emeritus of Civil Engineering at the University of Texas at Austin, where I served on the faculty for 38 years. I proposed the idea of the Summer Institutes in 2014 and served as the technical director of the first three Summer Institutes in 2015-2017. I am interested in the development of better flood forecasting science and technology, and in better communication of the resulting warning information to those at risk during flood emergencies."

Kelsey Mcdonough

"I am the FloodID Product Lead for The Water Institute, based in Boulder, CO, leading product strategy and growth of our operational flood forecasting technology. My research interests include flood inundation mapping, early flood warning and operational forecasting, and technology transformation for varied applications. In my free time, I love hiking, gardening, and traveling."

Summer Institute Course Coordinators

From left to right: Francisco Gomez and Parvaneh Nikrou.

Summer Institute Teams

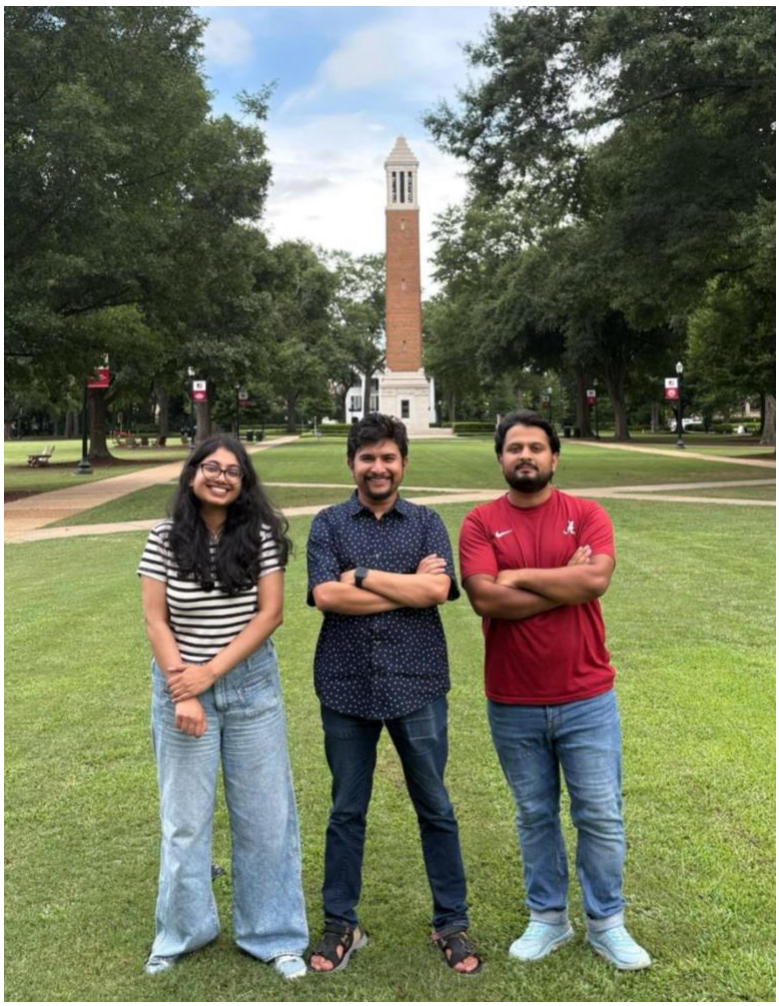
Chapter 1: A Comparative Analysis of Flood Inundation Mapping Forecasting Approaches: Quantifying Fidelity–Accuracy Tradeoffs and Bridging Gaps via Hybrid Deep Learning

From left to right:. Faezeh Maghsoodifar, Md Shadman Sakib, Mohamed Mowafy and Haotian Wang

Chapter 2: Catchment behavior analysis by developing a spatial-temporal loss function on LSTM

From left to right: Fatemeh Yavari, Habtamu Tamiru, Mohamad Ali Farmani and Arman Oliazadeh

Chapter 3: Investigating Approaches for Representing Pluvial Flooding within NOAA's NextGen Modeling Framework



From left to right: Samrin Sauda, Yogesh Bhattarai and Supath Dhital

Chapter 4: Demonstrating the Feasibility of ML-Based Pluvial Flood Mapping in Urban Settings

From left to right: Adam Smith, Sadra Seyvani, Keivan Tavakoli and Mostafa Saberian

Chapter 5: Towards a Flood Navigation and Safety Decision Support Tool: A Pilot for Emergency Responders in Travis County, Texas

From left to right: Ali Farshid, Kayode Adebayo, Nana Oye Djan and Saide Zand

Chapter 6: Estimating Flood Inundation Using a Densified Stream Network in Travis County, TX

From left to right: Megan Vardaman, Alex Simpson and Victor Oladoja