TECHNICAL REPORT 20

July 2024

OWP | OFFICE OF WATER PREDICTION

CUAHSI

allied for water science

www.cuahsi.org

National Water Center Innovators Program

Summer Institute Report 2024

National Water Center Innovators Program
Summer Institute Report 2024

Editors:
Karina Larco
Sadaf Mahmoudi

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

CUAHSI Technical Report No. 20

Version 1.0

July 2024

Funding for this project 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.

Suggested Citation:

Larco, K., Mahmoudi, S., et al. (2024). National Water Center Innovators Program - Summer Institute, CUAHSI Technical Report, HydroShare, https://www.hydroshare.org/resource/51e37a340ff04720b4c2d3d8912b0a28/

Contents

Preface

This report summarizes the research conducted during the 2024 National Water Center Innovators Program Summer Institute (NWC-SI). The NWC-SI is the result of a partnership between the National Weather Service's Office of Water Prediction (OWP) and the Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI). OWP is a part of the National Oceanic and Atmospheric Administration (NOAA). 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) facilitates the partnership between OWP and CUAHSI. 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 National Water Center (NWC) on the University of Alabama campus has hosted the NWC-SI every year. Graduate students come from different universities across the United States, and senior academic faculty, federal scientists, staff, private companies, and contractors meet in Tuscaloosa for seven weeks. This year, they worked on projects related to the NWC goals to support the Next Generation National Water Model (NextGen) framework.

Student fellows establish teams based on shared research interests, and theme leads supervise them while they work on their projects. This year, the fellows formed seven teams, each of which addressed a specific research question. The projects included impact-based flood forecasting and mapping flood inundation, the development of applications for GOES satellite imagery-based snow water equivalent (SWE) prediction, and applied topics that join flood inundation mapping with social needs.

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 in that it fosters teamwork, networking, and the development of lasting friendships.

Fellows

The 2024 cohort included 24 graduate students from 14 different universities across the United States. They began their journey with virtual meetings, which culminated in a two-week boot camp on coding, data retrieval, and understanding the National Water Model. The fellows and theme lead hail from various academic departments, including civil engineering, geography, and earth sciences.

Themes and Theme Leads

The NWC-SI 2024 themes and theme leads were:

- The "Assimilation of Geosynchronous Satellite Imagery into NextGen for Improved Modeling" theme led by Jonathan Frame (University of Alabama). Additional technical support was provided by Fred Ogden (NOAA OWP), Dan Lindsey (NOAA), Andrew Bennett (University of Arizona), James Halgren (Alabama Water Institute), Alan Rhoades (Lawrence Berkeley National Lab), Nels Frazier (Lynker), Mike Johnson (Lynker), and Austin Raney (Lynker).

- The "Flood Inundation Mapping (FIM) Uncertainty Analysis and Model Intercomparison" theme was led by Xingong Li (Kansas University) and Sagy Cohen (University of Alabama). Additional technical support was provided by Jude Kastens (University of Kansas, Kansas Biological Survey), James Halgren (University of Alabama), David Weiss (University of Kansas), Anupal Baruah (University of Alabama), Junho Song (University of Alabama), Fernando Salas (NOAA), Carson Pruitt (NOAA), and Royce Fontenot (NOAA).

- The "Geospatial Data Integration to Identify High-flow Thresholds for Improved Flood Risk Characterization" theme was led by Kelsey R. McDonough (FloodMapp) and Sanjib Sharma (Howard University). Additional technical support was provided by Sagy Cohen (University of Alabama), Anupal Baruah (University of Alabama), Dipsikha Devi (University of Alabama), Edward Clark (NWC), Carson Pruitt (NWC), Hemal Dey (University of Alabama), Sadra Seyvani (University of Alabama), and Wanyun Shao (University of Alabama).

Project Summaries

The 2024 NWC-SI projects are summarized below. Chapters 1-7 present the complete reports.

1. Projects within the Assimilation of Geosynchronous satellite imagery Into NextGen For improved modeling theme:

Chapter 1: "Leveraging Geosynchronous Satellite Imagery and Machine Learning to Predict the Magnitude and Extent of Snow water Equivalent". This chapter leverages geosynchronous satellite imagery and machine learning to predict snow water equivalent (SWE) at the NextGen hydrofabric scale. The team developed convolutional neural network (CNN) and long short-term memory (LSTM) models using GOES-16 satellite imagery and in-situ SWE measurements from California. The LSTM model showed higher accuracy (Kling-Gupta Efficiency of 0.63) than the CNN model (0.52).

Chapter 2: "Identifying Atmospheric Rivers on the West Coast of the United States with Geostationary Operational Environmental Satellite Imagery". This chapter explores the identification of Atmospheric Rivers (ARs) along the United States West Coast using Geostationary Operational

Environmental Satellite (GOES) imagery. Employing a U-Net Convolutional Neural Network (CNN) and a Random Forest (RF) model, the research aimed to leverage GOES's high temporal and spatial resolution for accurate AR detection. The U-Net CNN, trained on labeled AR data, outperformed the RF model, achieving an Intersection over Union (IoU) of 0.23 compared to the RF model's IoU of 0.02.

Chapter 3: "Probabilistic Streamflow Prediction Using the Model-Agnostic NextGen Framework". This section shows the NextGen Framework to probabilistically predict streamflow within gauged catchments, focusing on three selected CAMELS. By implementing the Dynamically Dimensioned Search (DDS) calibration method for the Conceptual Functional Equivalent (CFE) Model and utilizing various forcing data, the research aims to enhance prediction reliability. The results demonstrated improved accuracy compared to USGS streamflow data.

2. Projects within the Flood Inundation Mapping (FIM) Uncertainty Analysis and Model Intercomparison theme:

Chapter 4: "Enhancing Remote Sensing Flood Inundation Mapping by Leveraging a Terrain-Based Model". This chapter enhances flood inundation mapping using a hybrid approach combining remote sensing data with terrain-based models. Applied to the Verdigris River flood in May 2019, their method, FLDSensing, integrates satellite imagery and FLDPLN model data to accurately fill gaps caused by cloud cover. The results showed improved accuracy in flood extent and depth mapping compared to traditional methods.

Chapter 5: "Flood Inundation Mapping Using Terrain-Based Models for Flash Floods due to Dam Operations". This section evaluates terrain-based models, OWP HAND-FIM and FLDPLN, for rapid flood inundation mapping due to dam failures, comparing them with the HEC-RAS benchmark. The terrain-based models proved effective for near real-time mapping and scalability, though improvements in depth prediction are needed with a fast computation time.

3. Projects within the Geospatial Data Integration to Identify High-flow Thresholds for Improved Flood Risk Characterization theme:

Chapter 6: "Transition from Hazard to Impact-based Riverine Flood Forecasting". This chapter presents an automated system for transitioning from hazard-based to impact-based flood forecasting. By using forecasted streamflow data from the National Water Model, the system generates Flood Inundation Maps (FIMs), calculates flood depths, and integrates a Social Vulnerability Index (SVI) to assess flood impacts.

Chapter 7: "Leveraging Geospatial Data and Machine Learning to Predict Insurance-derived Flood Damage Cost". This chapter uses eXtreme Gradient Boosting Regression (XGBR) to predict building flood damage from Hurricane Harvey by integrating insurance claims and geospatial data. The model achieved an R^2 of 0.74 and highlights the importance of water depth, building value, and height.

Acknowledgments

Since the first NWC-SI in 2015, over 180 graduate student fellows have worked together to complete research projects during the Summer Institute. The innovations produced are not as significant as the relationships and partnerships that have been developed in Tuscaloosa. The NWC and NOAA OWP are to be thanked for giving us the chance to help improve water forecasting in the United States. The Summer Institute 2024 would not have been possible without the commitment and hard work of many people:

Above all, we would also like to express our gratitude to the CUAHSI staff who were instrumental in making this program a reality. Jordan Read, Julia Masterman, Emily Clark, Irene Garousi-Nejad, and Tony Castronova. Ed Clark, Fred Ogden, Fernando Salas, Betsy Morrison, and Emily Crisci of the NWC provided technical guidance and administrative support. Special thanks to CIROH, specifically Sagy Cohen, Lanna Nations, and Meredith Marsh, who managed the logistics of living on the University of Alabama campus, including travel arrangements, accommodations, and institutional relationships. Technical training was provided by Irene Garousi-Nejad and Tony Castronova (CUAHSI), Jordan Read (CUAHSI), Fred Ogden (NOAA OWP), Fernando Salas (NOAA), Jonathan Frame (University of Alabama), Sagy Cohen (University of Alabama), Xingong Li (Kansas University), Sanjib Sharma (Howard University), and Kelsey McDonough (FloodMapp). Also, thanks to James Halgren from Alabama Water Institute, who provided guidance to the groups and space in the laboratory to the cohort. Thank you to Eric Ezel and Jeremy Pate who helped to provide electronic resources, and Sepehr Karimi from Alabama Water Institute, who held office hours for compute resources. Course Coordinators Karina Larco and Sadaf Mahmoudi helped plan training sessions, research deliverables, lived and worked with the fellows for the duration of the NWC-SI.

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 contributed to the NWC-SI.

Karina Larco
Student Course Coordinator, NWC Innovators Program - Summer Institute 2024
Ph.D. Student, Brigham Young University

Sadaf Mahmoudi
Student Course Coordinator, NWC Innovators Program - Summer Institute 2024
Ph.D. Candidate, University of Alabama

Chapter 1
Leveraging Geosynchronous Satellite Imagery and Machine Learning to Predict Snow Water Equivalent at the NextGen Hydrofabric Scale

Savalan Nasar Neisary[1], Hassan Saleh[2], Raymond J. Hess[3], Helaleh Khoshkam[4]

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

[2] Western Michigan University; *h.saleh@wmich.edu*

[3] Rutgers University; *raymond.hess@rutgers.edu*

[4] University of Hawai'i at *Mánoa*; *helaleh@hawaii.edu*

Academic Advisors: Steven J. Burian, *The University of Alabama*; Mohamed Sultan, Western Michigan University; Nathaniel A. Bogie, *San José State University*; Sayed M. Bateni, *University of Hawai'i at Mánoa*

Summer Institute Theme Advisors: Jonathan Frame, University of Alabama, *jmframe@ua.edu*

Abstract: Year to year, reliable estimates of snow water equivalent (SWE) are vital for freshwater management, as they govern the accuracy of annual streamflow predictions. Direct measurements of SWE are often restricted to point source locations, which neglect changes in terrain, and bring to the fore a need for broader spatial coverage. Our study aims to address this limitation by developing a deep learning (DL) model to assess the feasibility of predicting SWE at the NextGen hydrofabric catchment scale. We conducted a correlation analysis between the point-scale instrumented measurements of SWE and the visual and short-wave infrared (SWIR) imagery from the geostationary operational environmental satellites (GOES-R) series at both the watershed and CONUS scales. We hypothesize that this relationship can be exploited through a data-driven modeling approach. To test this hypothesis, we trained two machine learning models, a convolutional neural network (CNN) and a long short-term memory (LSTM), with GOES-16 CONUS-scale satellite imagery (visual and infrared). In-situ measurements of SWE from three California Data Exchange Center (CDEC) stations were used for training, validating, and testing the model. We found a high positive and negative correlation between SWE and the radiance of visible and short-wave infrared bands, respectively. Moreover, the LSTM model predicted SWE with a Kling–Gupta Efficiency (KGE) of 0.63, compared to 0.52 from the CNN model. The findings from both correlation and machine learning provide a potential for improving SWE accuracy within the national water model and to benefit water resources management in snowpack driven streamflow systems.

---

1. Motivation

1.1 Water Resources & Mountainous Snowpack

Cold high-elevation snowpack functions as a natural water tower, and in California, Sierra Nevada snowmelt supplies 30% of the state's freshwater resources annually [1, 2]. Year to year, demand for freshwater increases due to population growth, industrial expansion, and agricultural production. Snow water equivalent (SWE) is a metric representing the volume of liquid water contained within a volume of solid snow. A significant source of freshwater is runoff from

snowmelt; SWE and snowmelt rate affect how much and when snowmelt runoff occurs [3]. Accurate estimates of snow depth and SWE are essential for hydrological modeling and streamflow prediction. SWE is typically measured at ground stations, which are limited in distribution, resulting in discontinuous monitoring [4] and significant uncertainties [5]. To address this problem, we explored training a Machine Learning (ML) model with satellite imagery in the visible and infrared spectral bands from geostationary operational environmental satellites (GOES-R) to predict SWE in poorly monitored catchments.

Currently, there is no observation-based, spatially complete, SWE product with high temporal resolution over large-scale regions. The daily SWE product from this study provides a unique opportunity to improve the accuracy of the SWE output from the National Water Model (NWM).

1.2 Geostationary Environmental Satellites

The National Oceanic and Atmospheric Administration (NOAA) currently employs four GOES-R, namely GOES-16 (East) and GOES-17 (West) (with one backup each). Although GOES-R series satellites need to be more utilized in hydrology, they are highly efficient. These orbiters provide continuous spatial coverage across the North American and South American continents, for near real-time (5-minute interval) environmental monitoring and weather forecasting. With the advanced baseline imager (ABI), instrumentation on board GOES yields pixel resolutions of 0.5 to 1.0 km for visible bands and 2 km for infrared channels. Comparing this spatial resolution with more conventional environmental satellites (MODIS and VIIRS), GOES series satellites produce coarser resolution imagery due to placement in geostationary orbit (35,800 km above Earth's surface). That said, the average lifespan of sensors aboard a GOES satellite is 15 years. This is nearly twice that of instrumentation on board conventional environmental satellites, MODIS and VIIRS, with 10-year and 7-year lifespans, respectively. These advantages suggest a promising upside and a significant cost-benefit to using GOES satellites for hydrologic research.

2. Objectives and Scope

The objectives of this study are to identify GOES bands that are correlated to SWE, and to assess the feasibility of assimilating GOES spectral bands (visual and infrared) to hydrofabric for a NextGen Water Model. To do this, we set out to develop a ML model that predicts SWE at the NextGen hydrofabric catchment-scale. Using GOES imagery as the primary training dataset, we propose the following questions to guide this study: Research Question 1, what are the benefits and limitations of using GOES for SWE modeling? Research Question 2, can we use GOES visual & infrared imagery to elicit NextGen hydrofabric catchment-scale dynamic hydrologic responses, specifically, which bands are correlated to SWE? Additionally, we propose the following hypotheses to test: Hypothesis 1, the GOES visible and infrared bands related to snow depth can be used to quantify SWE for short-term prediction and long-term monitoring of water resources in the southern Sierra Nevada. Hypothesis 2, GOES visible and infrared imagery provide a broad spatial context of SWE, enabling prediction for ungauged catchments in the Kern River watershed.

3. Previous Studies

At present, employing satellite remote sensing to monitor snow dynamics is a commonly utilized approach [6]. Over the past thirty years, remotely sensed passive microwave (PM) data have delivered valuable SWE observations worldwide [7, 8]. Yet, applications of PM SWE data products are limited by many factors, such as the mixed pixel problem at coarse spatial resolutions (e.g., 25 km), the saturation of the PM signal over deep snow, the many-to-one relationships between the PM signal, snow grain size, and SWE, and the impacts of forest cover [9]. As a result, PM-based SWE products alone are unable to account for snow depth [9], particularly in mountainous areas.

Snow exhibits a significant reflection in visible bands and a substantial absorption in the shortwave infrared spectra around 1.5 and 2.0 μm [3, 10]. Satellite observations in visible and infrared spectral channels provide useful information on fractional snow cover in various environments, which is correlated with snow depth. Many studies have tried to retrieve snow depth and SWE data using snow cover information from satellite imagery. Romanov & Tarpley (2004) evaluated the possibility of snow depth estimation over open prairie environments using an empirical relationship between snow depth and GOES-based fractional snow cover estimations. The results of this research reported a 30% error compared to ground stations for depths below 30 cm. Snow depth estimation is feasible since any alteration in snow depth impacts the fraction of the land surface covered by snow; consequently, this causes variation in the land surface's visible reflectance [11]. The authors improved their retrieval algorithm to expand snow depth estimates to forested regions [12].

By comparing the SWE output of the NWM with the ground truth SWE data at snow telemetry (SNOTEL) sites throughout the western United States, Garousi-Nejad and Tarboton (2022) found that the NWM usually underestimates SWE. For example, the authors reported that the NWM shows snow melting 6 to 19 days earlier than observed by SNOTEL [13]. However, it is worth noting that SNOTEL stations can provide SWE measurements at several locations with limited measurement frequencies. Thus, they do not allow us to evaluate the NWM's SWE product comprehensively.

We developed a new approach to estimate SWE using GOES imagery. By using radiance signals from bands one and three as inputs and SWE measurements from three CDEC stations as outputs, we tested two ML models. To our knowledge, no study has directly estimated SWE from GOES data. We selected the Kern River watershed as our study site due to its high contrast of snow-covered and surface-exposed terrain and its southerly trend from high to low elevation.

4. Data and Methodology

4.1 Hydrologic Setting

In California's southern Sierra Nevada mountains (36.578, -118.293), the Kern River is sourced by spring snowmelt from summits near Mount Whitney (the highest peak in the contiguous United States at 4,420 meters above sea level) (Figure 1). The Upper Kern River drainage basin encompasses 9,350 km^2 of sloped terrain, with the primary channel reaching 120 km from the headwaters to its terminus at Lake Isabella (a dammed storage reservoir with a water holding capacity of 0.7 km^3 or 570,000 acre-ft). Downstream, Kern County and the greater Bakersfield area depend on dam releases to source the Lower Kern River, which supplies local freshwater

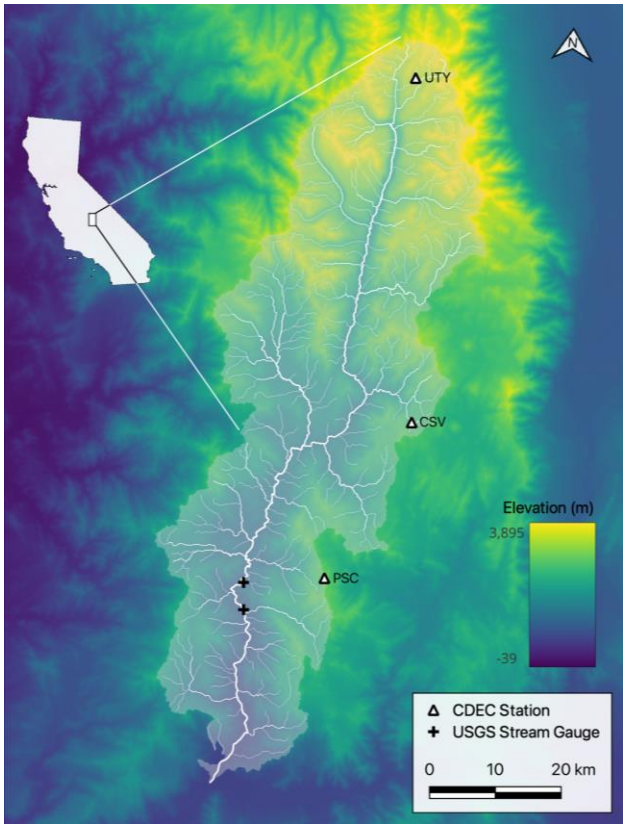


Figure 1. Inset topographic map of the Kern River watershed at the NextGen hydrofabric-scale showing two U. S. Geological Survey stream gauges (upstream and downstream) flanked by three CDEC stations: Upper Tyndall Creek (UTY), Casa Vieja Meadows (CSV), and Pascoes (PSC).

demand. These releases are critical, as they provide irrigation water to 100,000 acres of fruits and vegetables throughout the region [14], generating an average annual economic value of 320 million dollars (USD). Snowpack in the Sierra Nevada remains a critical freshwater resource for California's expanding population and prolific agricultural economy.

4.2 Ground Observations of SWE

We downloaded daily measurements of SWE from the California Data Exchange Center (CDEC) for years 2021, 2022, and 2023 at the following three sites: Upper Tyndall Creek (UTY), Casa Vieja Meadows (CSV), and Pascoes (PSC) (Figure 1; Table 1). These stations provide a pillow sensor measurement of SWE, which utilizes a pressure transducer to weigh the overlying snow, whereby sheltered instruments then convert this weight into an electrical signal of SWE. Years 2021 and 2022 were particularly dry years in California, with peak SWE in the Kern River watershed averaging 28 and 31 cm, respectively (Figure 2). Alternatively, the year 2023 was a considerably wet year, with an average peak SWE of 170 cm. These order of magnitude differences highlight the importance of SWE prediction for accurate streamflow estimates.

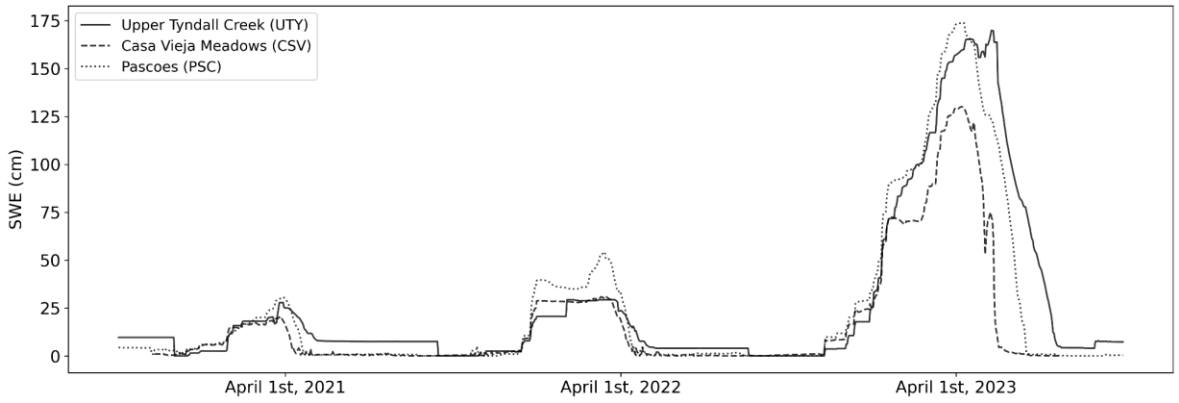


Figure 2. Peak SWE in the Kern River watershed at stations UTY, CSV, and PSC. Years 2021 and 2022 show relatively low snow conditions compared to the four-fold increase of SWE magnitude shown in the 2023 wet year.

The streamflow values measured by the U.S. Geological Survey (USGS) versus SWE measured at three CDEC sites are depicted in Figure 3. According to Figures 2 and 3, the peak value of SWE leads to a significant increase in streamflow. Moreover, during the summer, when there is

no snow on the mountain, the recorded streamflow is at its lowest range. This indicates that the downstream streamflow is highly correlated with the upstream SWE in the study basin.

Table 1. California Data Exchange Center stations in the Kern River watershed.

Station ID	Station Name	Latitude	Longitude	Elevation (m)
UTY	Upper Tyndall Creek	36.65	-118.40	3,475
CSV	Casa Vieja Meadows	36.2	-118.27	2,530
PSC	Pascoes	35.97	-118.35	2,790

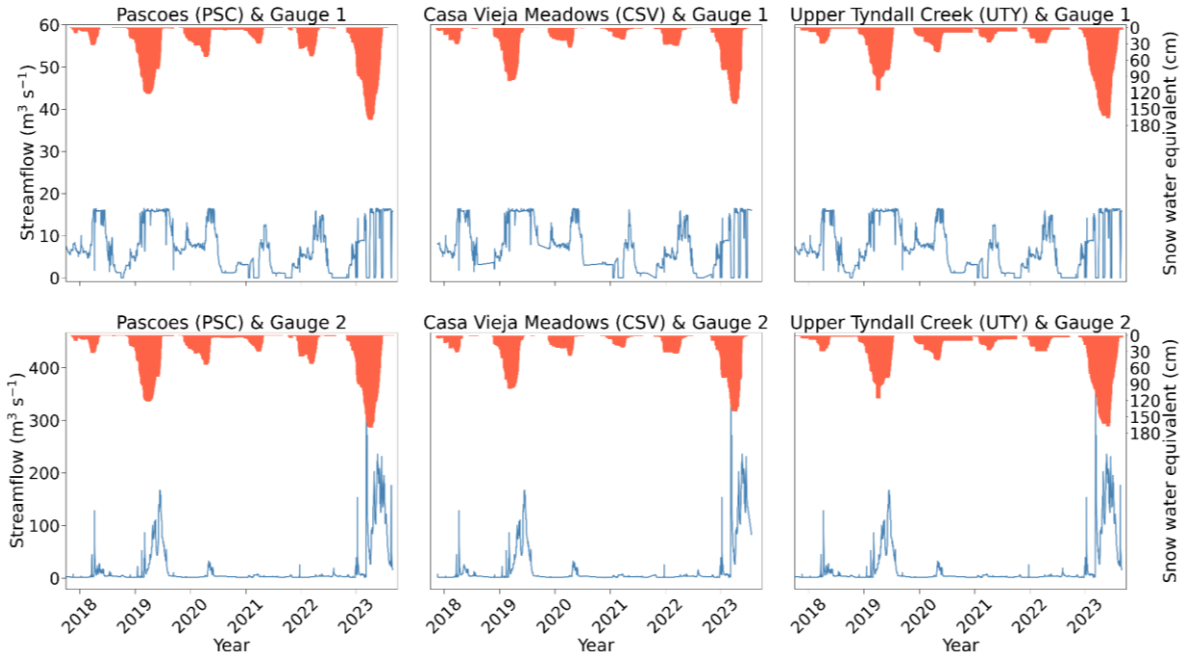


Figure 3. Streamflow at two USGS gauges versus SWE at three CDEC sites (Gauge 1 is upstream and Gauge 2 is downstream). Streamflow is shown in blue on the left axis, and SWE is depicted in red on the right axis.

4.3 GOES-16 Imagery

Visible and infrared imagery from GOES-R series satellites were downloaded using the Goes2Go Python package (version 2024.7.0). In this study, we selected bands one, three, and five from the L1b product that represent blue, veggie, and short-wave infrared (SWIR), respectively. Radiance measurements for each band were extracted at 5-minute intervals and then stacked, without applying atmospheric corrections. The stacked cube was then resampled to a daily time step to match the temporal measurements of SWE data. Then, we performed a Pearson correlation analysis between pixel radiance and in situ SWE data to test Hypothesis 1. Finally, the data were processed and prepared as an input to the ML model.

4.4 Pixel Correlation Analysis

To explore relationships between SWE data and GOES spectral channels, we performed a pixel correlation analysis with imagery from bands one, three, and five and the time series of SWE on a daily time scale. Although a one-to-one correlation is not expected between radiance and SWE, we expect radiance to have an apparent response to snow depth and can be processed (through

dynamic modeling) to a corresponding value of SWE. We evaluated the Pearson Correlation between each band and SWE stations at the watershed and CONUS scales.

4.5 Machine Learning Model

In this study, we examined the performance of two Machine Learning (ML) models, including a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM), to predict SWE values using the Radiance of GOES imagery.

An image or other grid-like data can be processed using a ML technique called a CNN algorithm. It consists of fully connected layers for prediction, pooling layers to lower dimensionality, and convolutional layers that use filters to identify features. CNNs are useful for applications like object identification and picture recognition because they automatically learn the spatial hierarchies of information [15].

Recurrent neural networks (RNNs) with LSTM are intended to simulate temporal data and sequences. LSTM utilizes gates to regulate the information flow to solve the vanishing gradient problem. The input, forget, and output gates enable the network to store or discard data over extended periods. For sequential data applications, such as time-series forecasting, LSTMs work incredibly well [16].

Radiance data of GOES bands one and three were initially selected as inputs to the ML models, while the SWE time series from each CDEC station were the model outputs. Due to limited storage capacity, we used data from the years 2022 and 2023 for training and testing purposes, respectively. The values of the hyperparameters that were set for each developed model are reported in Table S1.

4.6 Performance Evaluation Metrics

To test the model's performance, we selected the Kling–Gupta Efficiency (KGE) metric to evaluate the linear correlation, the mean ratio, and the ratio of standard deviations between the modeled SWE data and the observed SWE data [17]. Additionally, we used Percent Bias (PBias) to evaluate the over- or underestimation of the predictions, and Root Mean Square Error (RMSE) to assess the prediction error.

5. Results & Discussion

Our goal was to determine whether we can extract useful information from CONUS and watershed-scale GOES imagery, which has high temporal but moderate spatial resolution, for SWE prediction. We employed a pixel correlation analysis and tested two ML models to achieve this goal. Correlation analysis provides a preliminary understanding of how different bands may represent SWE magnitude at various locations. It also highlights which regions in CONUS and the watershed are more connected with SWE. This information is critical for developing and interpreting the results of our ML models in the subsequent steps.

5.1 Watershed Pixel Correlation Analysis

For watershed-scale analysis, we first extracted the pixels within our study basin, and then used a Pearson correlation to analyze the relationship between each band and CDEC station for 2023. The correlation maps in Figure 4 show that bands one and three have a high positive correlation at sites PSC (Figures 4a and b), CSV (Figures 4d and e), and UTY (Figures 4g and h). In contrast,

band five has a high negative correlation (Figures 4c, f, and i), particularly in high-elevation regions with significant snow cover. Snow absorbs the wavelengths of band five and intensifies the reflection of bands one and three, which results in these strong correlations. Furthermore, correlation results demonstrate that GOES imagery at the catchment-scale contains information that can help accurately predict SWE, without additional dynamic- and static-data.

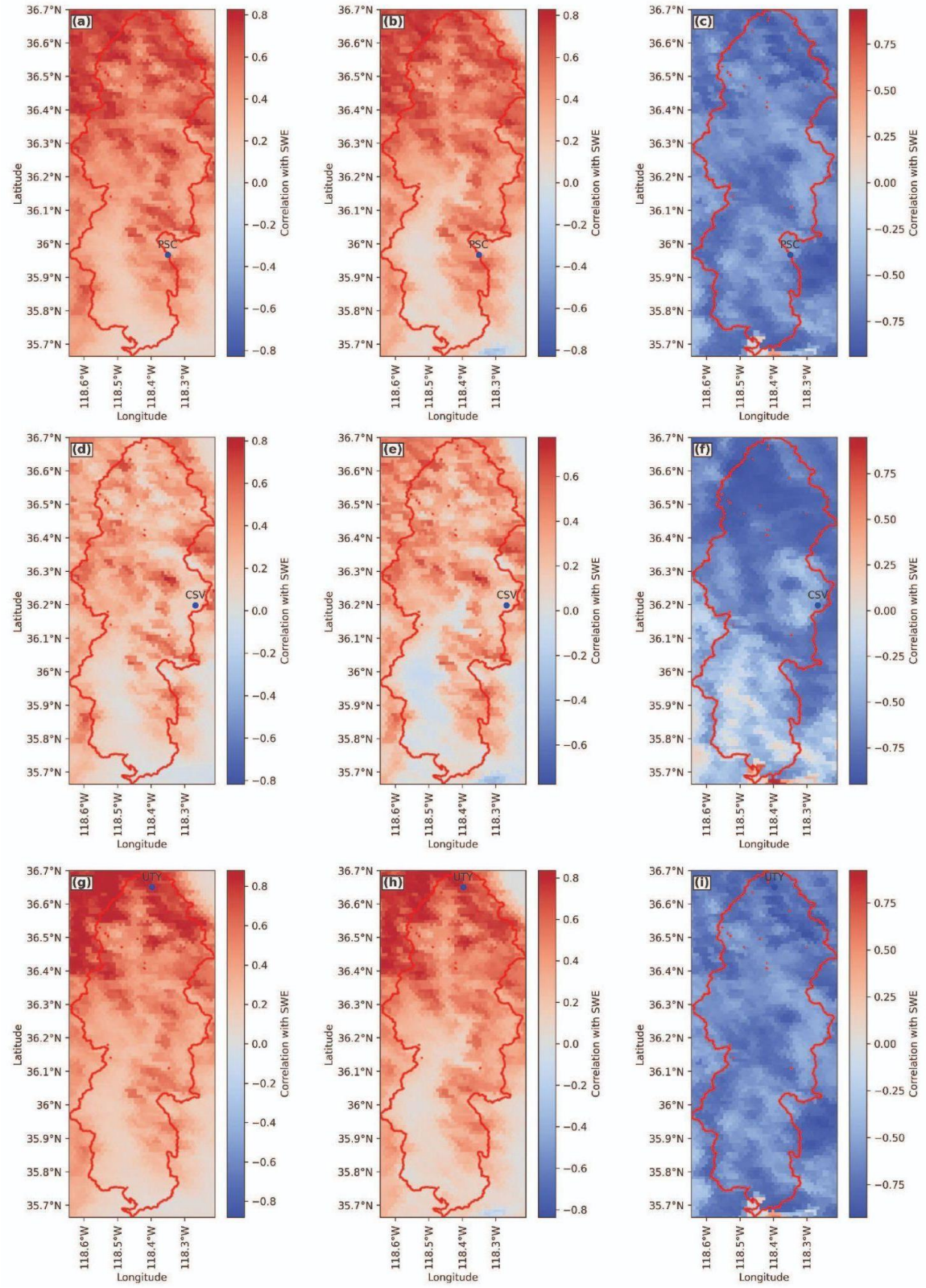


Figure 4. Watershed-scale pixel correlation analysis between CDEC stations and GOES radiance for bands 1, 3,

and 5. Subplots (a), (b), and (c) show correlations with station PSC for bands 1, 3, and 5, respectively. Subplots (d), (e), and (f) show correlations with station CSV, while (g), (h), and (i) show correlations with station UTY.

5.2 CONUS Pixel Correlation Analysis

One of our primary goals is to evaluate whether CONUS-scale GOES imagery can be used to extract information related to micro and macro atmospheric phenomena for SWE prediction at targeted locations. Thus, we performed a pixel correlation analysis between SWE values at three different high-elevation stations and the radiance of bands one, three, and five.

Figure 5 shows high positive and negative correlations for band one and stations PSC, CSV, and UTY, indicating that certain phenomena across CONUS influence SWE values at the catchment scale. To better visualize the correlation, we removed cells with correlation values between -0.4

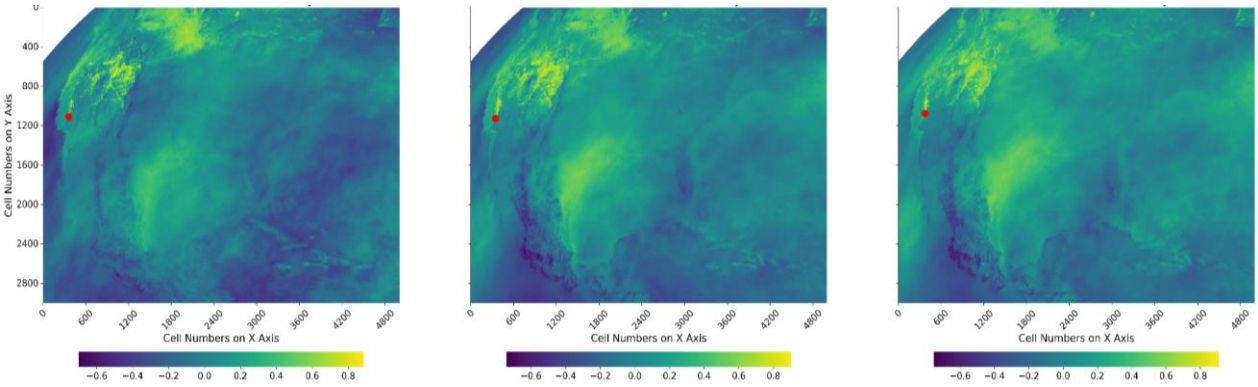


Figure 5. CONUS-scale correlation analysis from left to right for CSV, PSC, and UTY stations, respectively, and GOES band one.

and 0.4, highlighting positive correlations in red and negative correlations in blue (Figure 6). Comparing results across stations reveals spatial trends in correlated regions, suggesting that these phenomena impact various elevations differently.

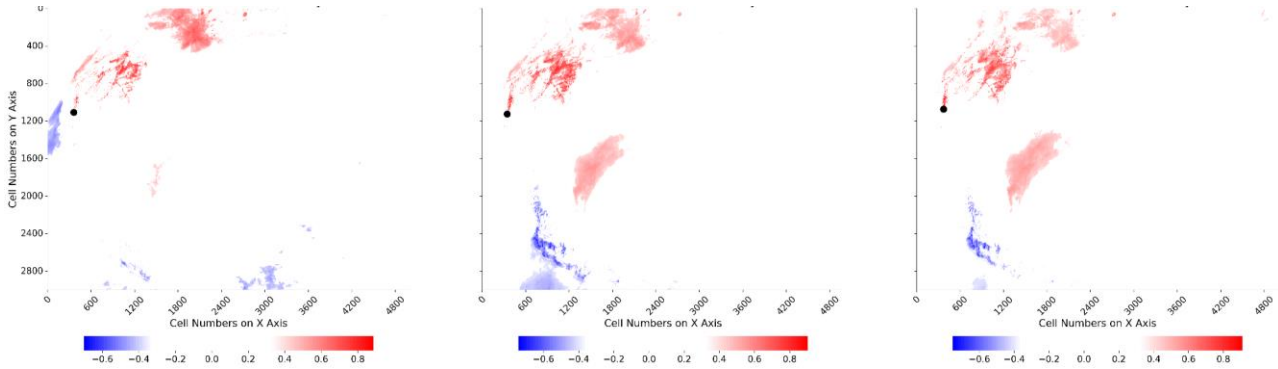


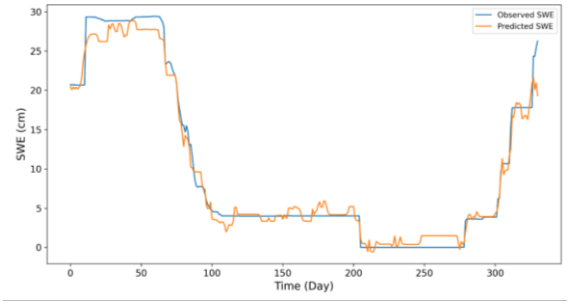
Figure 6. CONUS-scale correlation analysis from left to right for stations CSV, PSC, and UTY, respectively, and GOES band one for pixels with values less than -0.4 and greater than 0.4.

Figure 6 shows three regions of distinct high correlation for band one and for band three (Figure S1). The region with the highest correlation is the western United States, specifically the Rocky Mountains and the Cascade Range, which have significant snow cover. Correlations in this location show opposite values for different bands due to the varying response of different wavelengths to snow. The second high-correlation region is the Gulf of Mexico, attributed to

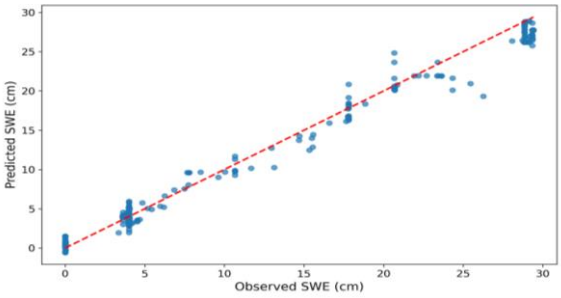
moisture movement caused by atmospheric rivers and tropical systems, affecting snowfall and moisture in the Sierra Nevada Mountains [2]. The third region is the Sierra Madre Occidental Mountains along the west coast of Mexico, which are influenced by seasonal changes in vegetation and soil moisture. These findings support our hypothesis that GOES imagery can reveal hydrologic processes and atmospheric phenomena that influence SWE magnitude across CONUS, which ML models can leverage for prediction. A detailed analysis of the physical processes connecting areas of high correlation in GOES imagery to SWE at our study sites could offer insights into atmospheric dynamics and large-scale hydrological processes.

5.3. LSTM and CNN Predictions

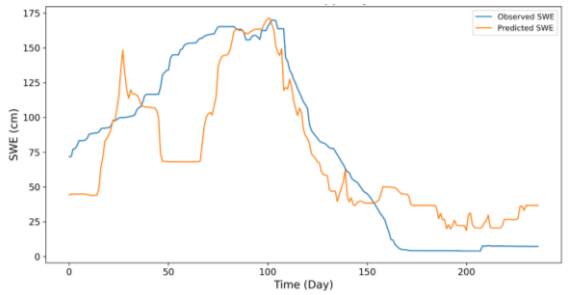
We developed two DL models, including CNN and LSTM, to predict SWE at the watershed scale using GOES bands one and three as inputs. We present the results of the UTY station here and the results for other stations in the supplementary materials section. For the LSTM model, we extracted the cells with a correlation higher than 0.5, then smoothed the radiance time series of these cells using a 15-day moving average. The LSTM model architecture has three layers, including an LSTM, a dropout, and linear layers with Mean Square Error (MSE) as the loss function. We fed the data with a 20-day look back at the model and used a dry year (2022) for training and a wet year (2023) for testing. Figure 7-A shows the time series of the observed and predicted SWE, and Figure 7-B shows the 1:1 scatter plot of the train results. In Figures 7-C and 7-D, we show the test results, and in Table 2, we summarize the evaluation values of KGE, RMSE, and PBias of the train and test results. Results show a 0.63 KGE for test duration, indicating great alignment of prediction and observation data considering the limitation of data and computational power for model training and testing.



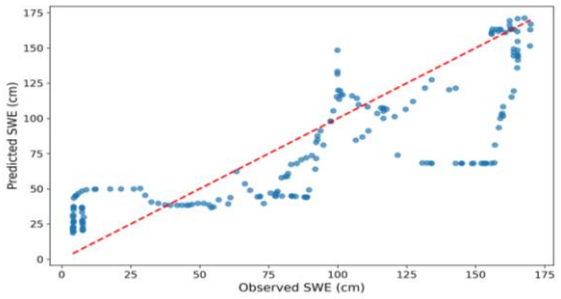
(A)



(B)



(C)



(D)

Figure 7. Shows the results of LSTM for A and B training period (2022) and C and D for the testing period (2023) for the UTY station. In B and D, the dashed line shows the 1:1 relationship between the observed and predicted values.

For the CNN model, we used all the cells as inputs and let the model analyze the dataset. Here, we used a 5-day moving average to smooth the dataset. Our tuning process showed that a longer moving average decreases the model performance in the CNN model. We used CNN, max pool, and linear layers in our architecture, with MSE as our loss function. Figure 8 presents the time series and scatter plot of the CNN model, which shows high training accuracy and reasonable testing accuracy. Table 2 provides values for the different metrics, which show more than 0.5 KGE for the model.

Table 2. Presents the metric values for both the train and test periods of both models for the UTY station. CNN shows lower accuracy compared to the LSTM.

Eval. Metric	LSTM (train)	CNN (train)	LSTM (test)	CNN (test)
KGE	0.91	0.94	0.63	0.65
RMSE	0.53	0.5	12.8	14.2
PBias	0.93	6.3	-3.35	-2.32

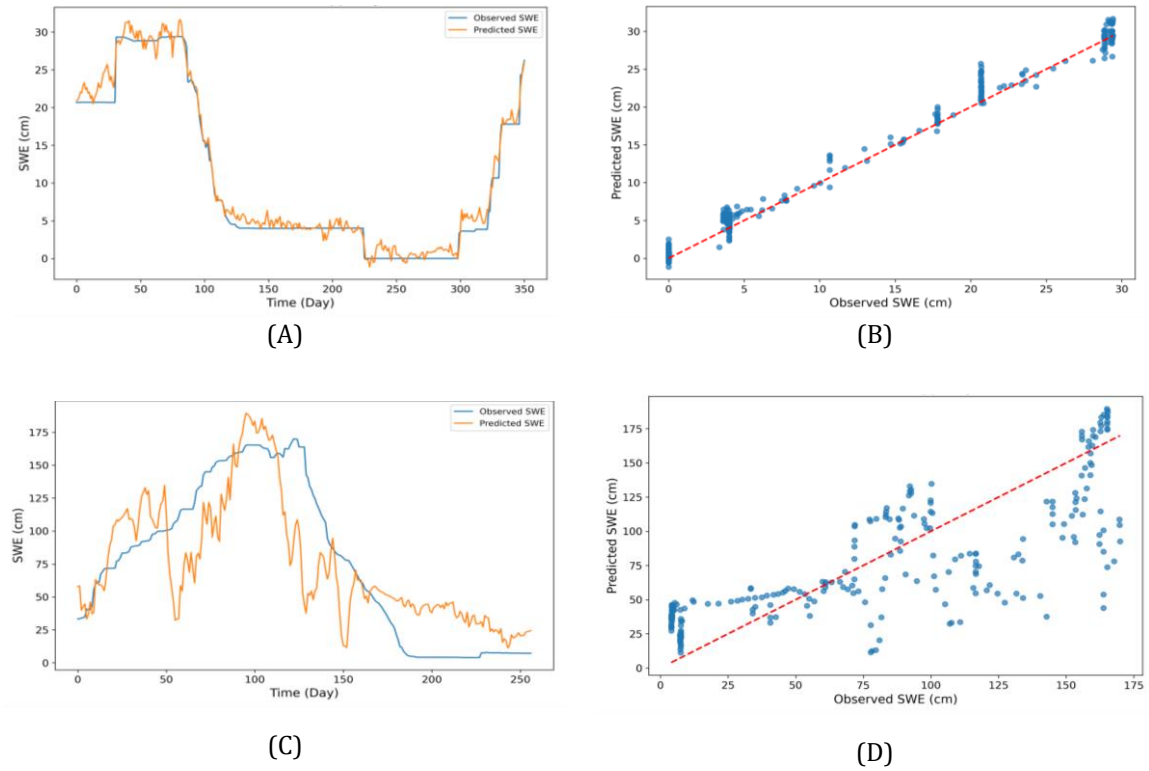


Figure 8. Results of the CNN for A and B training period (2022) and C and D for the testing period (2023) for the UTY station. In B and D, the dashed line shows the 1:1 relationship between the observed and predicted values.

6. Conclusion

Our study extracts a relationship between radiance from GOES spectral channels and SWE magnitude, at both the watershed and continental scales. We have estimated a high positive correlation between daily radiance from visible bands and observed SWE over areas covered by snow within the Kern River watershed. Similarly, a high negative correlation was found between band five (short-wave infrared) and SWE. To build on that, we trained two ML models, LSTM and CNN, to predict SWE from radiance. Model results indicate good performance (KGE: 0.63 and 0.65) for the LSTM and CNN models, respectively. The model results could be further improved by training the model with more bands (i.e. bands 2 and 5) and incorporating other hydroclimatic and static layers as inputs for the ML models. This relationship can be exploited and incorporated into the basic model interface for a NextGen National Water Model. Future efforts will derive these relationships from international geostationary satellites and similar snow telemetry networks abroad.

Acknowledgments

We thank Fred Ogden and Dan Lindsey for project guidance and helpful suggestions, Andrew Bennett for conversations regarding model architecture, and James Halgren for technical support.

Supplementary Materials

GitHub Repository: *https://github.com/NWC-CUAHSI-Summer-Institute/SI24_GOES-SWE*

Table S1. Optimal hyperparameters of LSTM and CNN models.

Model Hyperparameter	LSTM	CNN
Batch Size	10	10
Learning Rate	0.01	0.001
Number of Epochs	200	50
Weight Decay	1e-3	1e-9
Lookback	20	-
Dropout	0.7	-

1. CONUS scale correlation results.

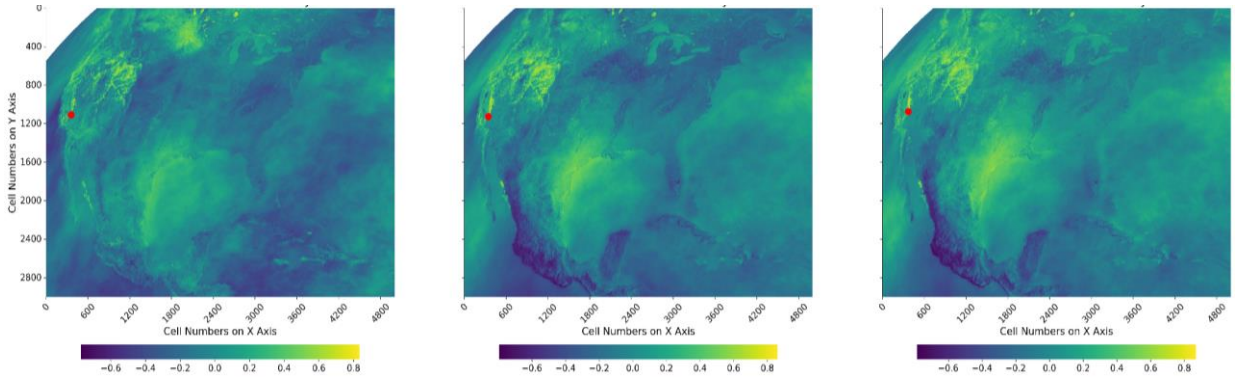


Figure S1. CONUS-scale correlation analysis from left to right for CSV, PSC, and UTY stations, respectively, and GOES band three.

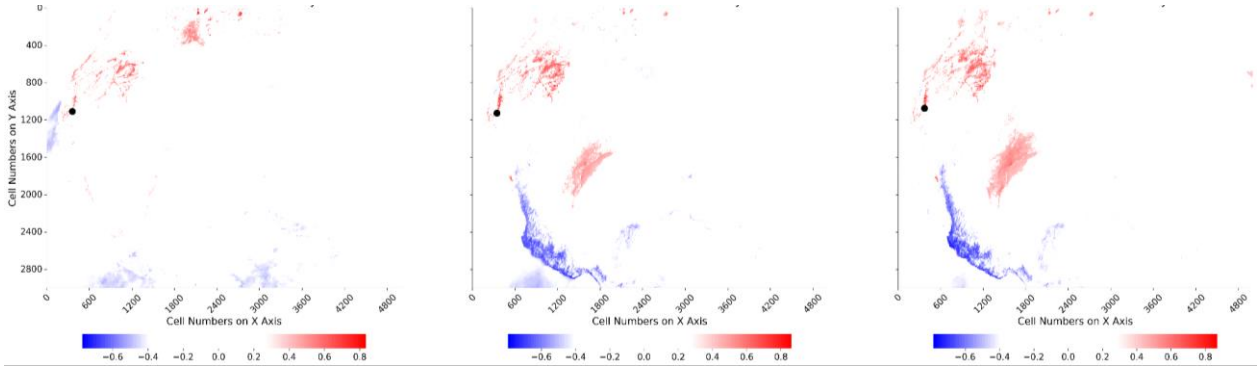


Figure S2. CONUS-scale correlation analysis from left to right for CSV, PSC, and UTY stations, respectively, and GOES band three for pixels with values less than -0.4 and greater than 0.4. We provide this plot to visualize the correlations better.

2. Model Results

2.1. PSC station results.

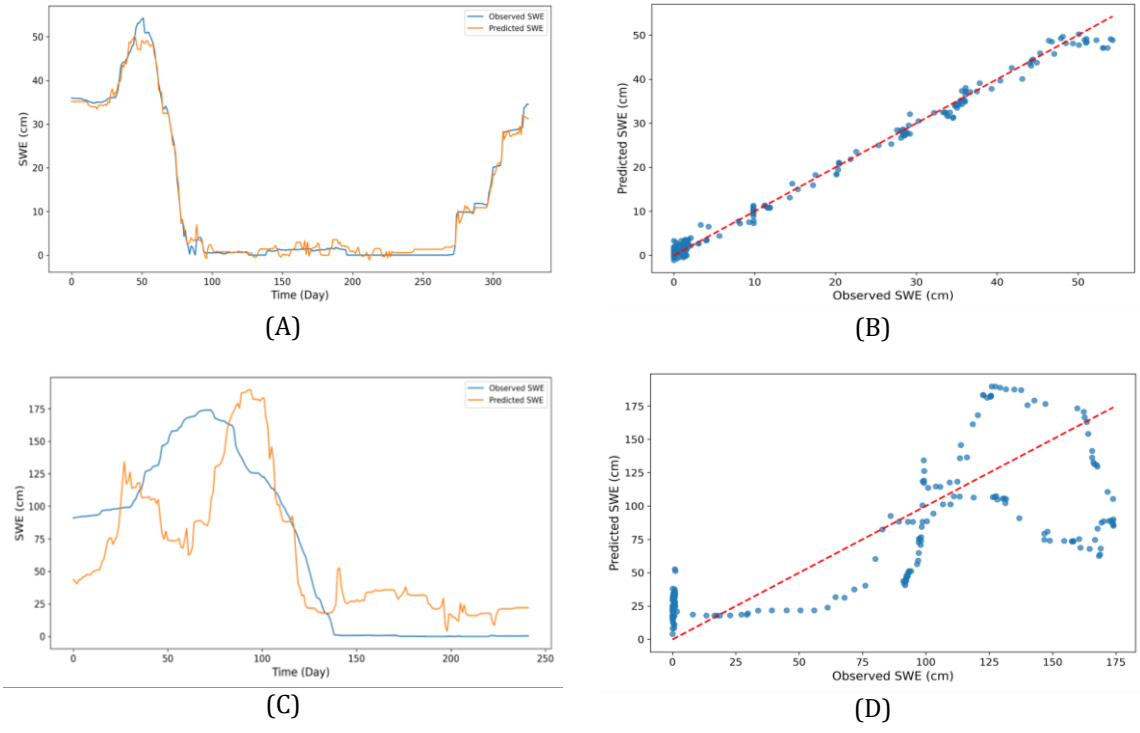


Figure S3. Shows the results of LSTM for A and B training period (2022) and C and D for the testing period (2023) for the PSC station. In B and D, the dashed line shows the 1:1 relationship between the observed and predicted values.

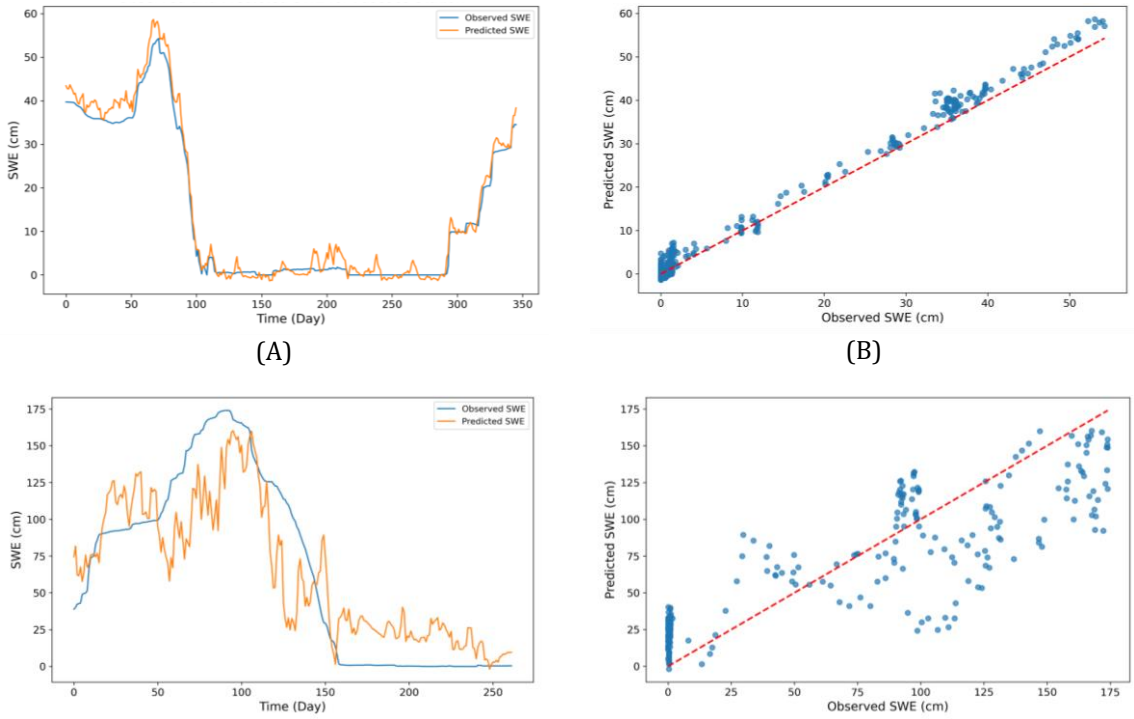


Figure S4. Shows the results of CNN for A and B training period (2022) and C and D for the testing period (2023) for the PSC station. In B and D, the dashed line shows the 1:1 relationship between the observed and predicted values.

2.2. CSV station results.

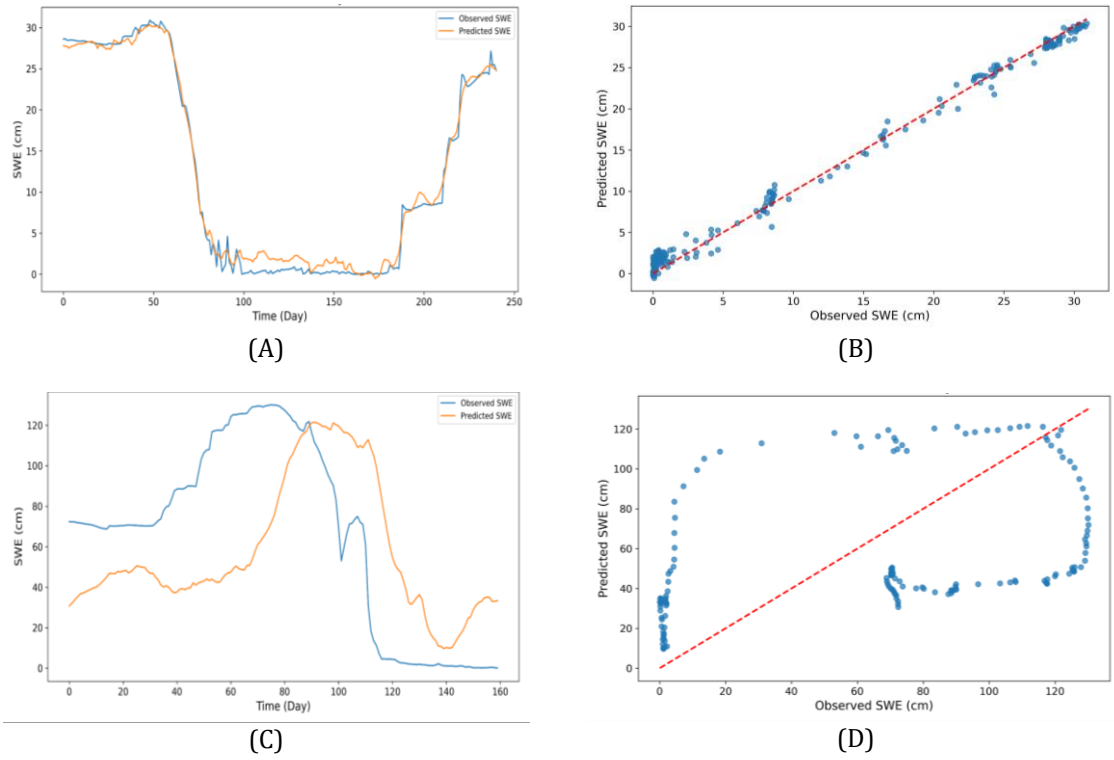


Figure S5. Shows the results of CNN for A and B training period (2022) and C and D for the testing period (2023) for the CSV station. In B and D, the dashed line shows the 1:1 relationship between the observed and predicted values.

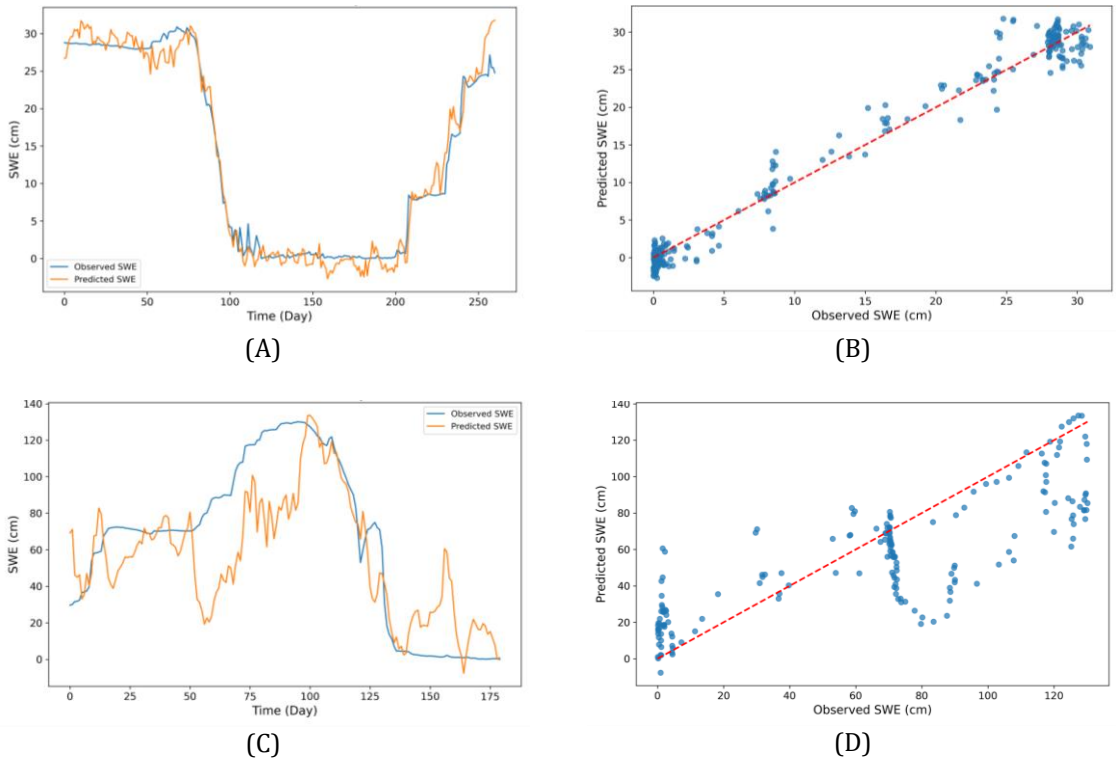


Figure S6. Shows the results of CNN for A and B training period (2022) and C and D for the testing period (2023) for the CSV station. In B and D, the dashed line shows the 1:1 relationship between the observed and predicted values.

References

[1] J. E. Halofsky, D. L. Peterson, L. Y. Buluç, and J. M. Ko, "Climate change vulnerability and adaptation for infrastructure and recreation in the Sierra Nevada," U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station, Albany, CA, PSW-GTR-272, Sep. 2021. doi: 10.2737/PSW-GTR-272.

[2] E. R. Siirila-Woodburn et al., "A low-to-no snow future and its impacts on water resources in the western United States," Nat Rev Earth Environ, vol. 2, no. 11, pp. 800–819, Oct. 2021, doi: 10.1038/s43017-021-00219-y.

[3] S. Awasthi and D. Varade, "Recent advances in the remote sensing of alpine snow: a review," GIScience & Remote Sensing, vol. 58, no. 6, pp. 852–888, Aug. 2021, doi: 10.1080/15481603.2021.1946938.

[4] J. Muñoz, "Synergistic Use of Remote Sensing for Snow Cover and Snow Water Equivalent Estimation," BJECC, vol. 3, no. 4, Nov. 2013, doi: 10.9734/BJECC/2013/7699.

[5] K. Belinska, G. Fischer, G. Parrella, and I. Hajnsek, "The Potential of Multifrequency Spaceborne DInSAR Measurements for the Retrieval of Snow Water Equivalent," IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, vol. 17, pp. 2950–2962, 2024, doi: 10.1109/JSTARS.2023.3345139.

[6] Y. Lu, T. James, C. Schillaci, and A. Lipani, "Snow detection in alpine regions with Convolutional Neural Networks: discriminating snow from cold clouds and water body," GIScience & Remote Sensing, vol. 59, no. 1, pp. 1321–1343, Dec. 2022, doi: 10.1080/15481603.2022.2112391.

[7] J. Pulliainen, "Retrieval of Regional Snow Water Equivalent from Space-Borne Passive Microwave Observations," Remote Sensing of Environment, vol. 75, no. 1, pp. 76–85, Jan. 2001, doi: 10.1016/S0034-4257(00)00157-7.

[8] N. Saberi, R. Kelly, M. Flemming, and Q. Li, "Review of snow water equivalent retrieval methods using spaceborne passive microwave radiometry," International Journal of Remote Sensing, vol. 41, no. 3, pp. 996–1018, Feb. 2020, doi: 10.1080/01431161.2019.1654144.

[9] K. Yang, K. N. Musselman, K. Rittger, S. A. Margulis, T. H. Painter, and N. P. Molotch, "Combining ground-based and remotely sensed snow data in a linear regression model for real-time estimation of snow water equivalent," Advances in Water Resources, vol. 160, p. 104075, Feb. 2022, doi: 10.1016/j.advwatres.2021.104075.

[10] J. Luo, C. Dong, K. Lin, X. Chen, L. Zhao, and L. Menzel, "Mapping snow cover in forests using optical remote sensing, machine learning and time-lapse photography," Remote Sensing of Environment, vol. 275, p. 113017, Jun. 2022, doi: 10.1016/j.rse.2022.113017.

[11] P. Romanov, and D. Tarpley, "Estimation of snow depth over open prairie environments using GOES imager observations," *Hydrol. Process,* vol. 18, pp. 1073-1087, Mar. 2004, doi: 10.1002/hyp.5508.

[12] P. Romanov, and D. Tarpley, "Enhanced algorithm for estimating snow depth from geostationary satellites," *Remote Sensing of Environment,* vol. 108, no. 1, pp. 97-110, May 2007, doi: 10.1016/j.rse.2006.11.013.

[13] I. Garousi-Nejad, and D. Tarboton, "Data for a comparison of national water model retrospective analysis snow outputs at SNOTEL sites across the Western U.S.," Hydrological Processes, vol. 36, no. 1, e. 14469, Jan. 2022, doi: 10.1002/hyp.14469.

[14] K. Bourque et al., "Balancing agricultural production, groundwater management, andbiodiversity goals: A multi-benefit optimization model of agriculture in Kern County, California," Science of The Total Environment, vol. 670, pp. 865–875, Jun. 2019, doi: 10.1016/j.scitotenv.2019.03.197.

[15] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," Dec. 02, 2015, arXiv: arXiv:1511.08458. Accessed: Jul. 17, 2024. [Online]. Available: http://arxiv.org/abs/1511.08458.

[16] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[17] H. Kling, M. Fuchs, and M. Paulin, "Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios," *Journal of Hydrology*, vol. 424–425, pp. 264–277, Mar. 2012, doi: 10.1016/j.jhydrol.2012.01.011.

Chapter 2

Identifying Atmospheric Rivers on the West Coast of the United States with Geostationary Operational Environmental Satellite Imagery

Anshul Yadav[1], Janani Kandasamy[2], Meklit Berihun Melesse[3], and Surabhi Upadhyay[4]

[1] Texas A&M University; *anshulya@tamu.edu*

[2] George Mason University; *jkandasa@gmu.edu*

[3] Washington State University; *m.melesse@wsu.edu*

[4] Colorado School of Mines, *surabhi_upadhyay@mines.edu*

Academic Advisors: Huilin Gao, *Texas A&M University*; Viviana Maggioni, *George Mason University*, Yonas Demissie. *Washington State University;* Adrienne Marshall*, Colorado School of Mines*

Summer Institute Theme Advisors: Jonathan Frame, Lynker & University of Alabama, *jmframe@ua.edu*

Abstract: Atmospheric Rivers (ARs) are extreme weather events characterized by intense moisture transport, which, upon landfall, produce precipitation that can either alleviate drought or cause significant flooding and socio-economic damage. The study focuses on ARs along the Pacific West Coast of North America, from Washington to California, due to their diverse climatic conditions and significant agricultural and economic activities. Identifying ARs using GOES imagery with high temporal and spatial resolution is crucial for timely and accurate prediction, as it provides continuous and detailed observations essential for monitoring these events. In this project, GOES-R series satellites and ERA5 reanalysis data were utilized to analyze the identification of ARs. A U-Net CNN was employed for AR detection, trained on labeled AR data, and validated against benchmark datasets. Additionally, a Random Forest (RF) model was used as a secondary approach for AR identification. The CNN outperformed the RF model across most metrics, achieving an Intersection over Union (IoU) of 0.23, while the RF model achieved IoU of 0.02. Training a robust U-Net CNN model with real-time monitoring GOES satellite imagery and hand-labeled data proved effective in identifying ARs with high accuracy.

1. Motivation

An atmospheric river (AR) is a long, narrow, and transient corridor of strong horizontal water vapor transport, typically associated with a low-level jet (LLJ) stream ahead of the cold front of an extratropical cyclone [1]. These ARs are significant conveyors between oceanic evaporation and continental precipitation, often responsible for 90% of poleward moisture transport [2]. For example, in the North Pacific, they transport an average of 700 kg m^{-1} s^{-1}, which is more than twice and 25 times the mean annual discharge of the Amazon and Mississippi rivers, respectively [1], [3]. ARs can cause heavy precipitation when forced upward by mountains or within warm conveyor belts, contributing 20-30% of annual precipitation in western Europe and USA [3]. While crucial for water supplies, ARs are also responsible for over 85% of flood events along the U.S. West Coast. Their absence can increase drought occurrence by up to 90% [3], [4].

Identifying ARs is critical for water resource management, especially as climate change predicts larger and more intense ARs [4]. Historically, AR identification has depended on algorithms and human skills. Still, recent advancements have shown the potential of deep learning, particularly convolutional neural networks (CNNs), to efficiently identify ARs in large satellite images [5]. CNNs have been applied to AR identification and segmentation using climate reanalysis datasets (i.e., ERA5, HR-CESM, MERRA-2.0, LR-CMIP6) on both global and regional scales, showing promising results in improving detection accuracy [6], [7], [8]. However, these reanalysis-based methods face challenges due to uncertainties in initial conditions, numerical approximations, and model deficiencies, leading to decreased forecast skill over time [9]. To address these limitations, identifying ARs directly from optical sensor weather satellites like Geostationary Operational Environmental Satellite (GOES) presents an unexplored opportunity for more immediate, observationally based AR analysis.

GOES provides high-frequency (5-minute resolution for some of its products), near-real-time observations with extensive spatial coverage, which is crucial for the timely detection of atmospheric phenomena. On the other hand, ERA5, a widely used reanalysis dataset, offers detailed and comprehensive atmospheric information, though with some latency [10]. By utilizing direct observation and the near-real-time capabilities of raw GOES images for AR identification and using the detailed data from ERA5 for comparison, we hope to improve the identification and analysis of ARs. It should also be noted that when a method that utilizes adaptive timestep is applied, the fine temporal resolution of GOES can be highly useful, especially when AR is about to make landfall that is unparalleled by other resources.

2. Objectives and Scope

The primary objective of this project is to explore the potential of GOES satellite data in accurately identifying ARs along the Pacific coast of North America. We developed and tested two machine learning models—a U-Net model and a random forest model—to identify ARs from GOES satellite imagery spanning the period from 2018 to 2020. The U-Net model is utilized for its capabilities in image segmentation, while the random forest model is leveraged for its robust classification abilities. The research questions guiding this investigation include:

- Can optical satellite data from GOES, with its unique spatio-temporal resolution, be effectively utilized for the detection of ARs? Can this potential be extended to real-time monitoring capabilities?
- How accurately can U-Net and random forest models identify atmospheric rivers from GOES satellite imagery? Which of the two models performs best regarding accuracy, precision, recall, and other relevant metrics for AR identification?

By addressing these questions, the project aims to enhance the understanding of using satellite data for AR identification and contribute valuable insights into developing reliable and accurate AR detection systems.

3. Previous Studies

There has been an increase in interest in ARs in recent years [11] about addressing diverse issues of climatology and hydrology. At the basis of most studies is the detection of ARs [12], [13], [14], [15], [16]. Earlier and recent studies used threshold-based algorithms as a main tool for AR

detection. For instance, Osorio et al. [17] utilized integrations of the NASA Goddard Institute for Space Studies global climate model ModelE version 2.1 (GISSE2.1) for assessing the impact of climate change on ARs on the South Pacific Ocean and found that ARs respond differently at different latitudes of the basin. Guan and Waliser [18] applied the Tracking Atmospheric Rivers Globally as Elongated Targets (tARget) algorithm on ERA5 to develop a global ARs database to handle ARs in tropical and polar areas better. Others [18], [19], [20], [21], [22] also used different threshold-based algorithms for the identification of ARs.

In contrast to threshold-based algorithms, there has been an increased use of deep learning techniques for identification and tracking of ARs in recent times. This has been mostly due to two related challenges: inconsistent results by an algorithm on varying datasets and inconsistency on a given dataset across different algorithms [23]. One such work using deep learning is by Higgins et al. [24], who utilized CNN adapted from CGNet using a climate dataset which resulted in the successful identification of ARs as compared to the threshold-based algorithms. Mahesh et al. [23], using existing Atmospheric River Tracking Method Intercomparison Project (ARTMIP) data, also utilized CNN for the identification of ARs successfully [1]. Other studies [23], [27], [28] have also explored deep learning for AR identification and tracking.

Compared to other studies utilizing deep learning on AR, our study will fill two gaps. The first is the usage of direct observations from continuous GOES satellite data, which has unparalleled temporal resolution. The second is raw satellite data, which has not been utilized for AR identification based on reviewed studies.

4. Methodology

4.1 Study Area

Our research focuses on the Pacific West Coast of North America, as illustrated in **Figures 1 (a) and 1 (b)**. The study area encompasses a broad region from the state of Washington in the north to California in the south, extending westward over the Pacific Ocean. This region is significantly impacted by atmospheric rivers (ARs), which contribute approximately 50% of the water supply [29]. Historical data indicates that from 1950 to 2010, ARs played a crucial role in alleviating drought conditions in the Pacific Northwest, ending three out of four droughts during this period [29]. More recently, in 2022, ARs were instrumental in mitigating California's prolonged drought, demonstrating their importance in sustaining water resources in this area [30].

Additionally, this research considers the broader impacts of ARs on neighboring countries, including Canada and Mexico. British Columbia in Canada and the Baja California Peninsula in Mexico also experience significant weather events influenced by ARs [31]. These regions, like the western United States, benefit from the water resources provided by ARs, which are crucial for their water supply and drought mitigation efforts. The GOES satellite imagery used in this

[1] ARTMIP is established to assess uncertainties in AR science based on detection/tracking methodology and works that analyze Ars based on the MERRA-2 reanalysis data set [25]. It utilizes five categories when assessing studies on ARs; computation type, geometry requirements, threshold requirements, temporal requirements, and regions [26]. All the categories of ARTMIP are utilized in either the identification stage of our work or in the tracking and prediction part of our work.

research provides a comprehensive view of the atmospheric conditions over this entire region, allowing for detailed analysis and monitoring of ARs.

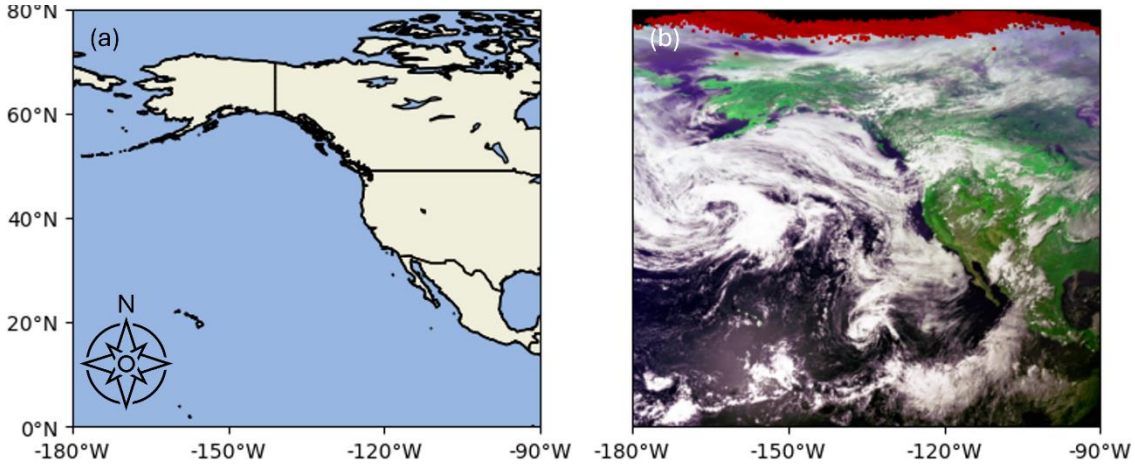


Figure 1: Study Area and GOES Satellite Imagery. (a) Pacific West Coast region including the US, Canada, and Mexico. (b) GOES satellite image capturing atmospheric conditions over the study area (false color composite using bands 2,3,1).

4.2 Dataset

The datasets used in the research, and the time period and spatial and temporal resolutions, can be found in Table 1.

Table 1: Dataset used in the research

Dataset	Period	Temporal Resolution	Spatial Resolution	Reference
GOES Imagery	2018 to 2020	15-min	4 km	[32]
Labeled AR	1979 to 2022	Event-based & 6-hour time step		[33]
ERA5	1996 to 2022	Hourly	31 km at 137 pressure levels	Dataset available at: https://climate.copernicus.eu/climate-reanalysis

4.2.1 GOES Satellite Imagery

The GOES series is operated by the National Oceanic and Atmospheric Administration (NOAA) of the United States. This study utilizes data from the GOES-R series, specifically employing the Advanced Baseline Imager (ABI) with 16 channels, offering a spatial resolution of 2 km for most channels and a temporal resolution of 15 minutes for full disk scans [34]. Data retrieval for the GOES-17 and GOES-18 satellites is facilitated by the GOES-2-go Python package (Version 2022.07.15), which accesses data through Amazon Web Services (AWS) as part of NOAA's Open Data Dissemination Program [32]. More information regarding the bands in GOES pre-2016 and post-2016 can be found in the supplementary material - section S1. The raw data from the GOES satellites required conversion into a standardized projection format. This step involved transforming the native GOES projection [32] into a rectilinear latitude-longitude projection to ensure consistency with the analysis requirements and facilitate accurate

spatial referencing. Additionally, the GOES images were resized from their original resolution of 4500x4996 pixels to 512x512 pixels to facilitate machine learning training and improve computational efficiency.

4.2.2. Labeled AR data

The dataset consists of continuous labeled data from 1979 to 2022, recorded at 6-hour intervals. This extensive temporal coverage provides a comprehensive view of the AR events. The data was sourced from Rhoades [33], who utilized the advanced TempestExtremes v2.1 algorithm to detect these features [35]. This algorithm is known for its robust capability in identifying and characterizing AR events, ensuring high-quality and reliable data for research and analysis.

4.2.3. ERA5 climate reanalysis product

The retrieved data from ERA5 include wind components (u, v) and specific humidity (q) at multiple pressure levels (300 hPa to 1000 hPa). These data were used to calculate Integrated Water Vapor (IWV) and Integrated Vapor Transport (IVT) (refer to equations in supplementary material - section S2 equations 1-2). IWV represents the total column of water vapor by integrating specific humidity over the pressure levels, while IVT quantifies the horizontal transport of water vapor using specific humidity and wind components across these levels [36].

We utilized ERA-5 reanalysis data to calculate IVT and IWV and developed a thresholding algorithm to identify AR objects. This algorithm applies specific thresholds for IVT and IWV and evaluates the minimum length and width of potential AR regions to meet typical AR criteria. The algorithm starts by applying threshold values for IVT and IWV to create an initial mask of potential AR regions. Regions meeting these thresholds are marked. Minimum length and width requirements for ARs are converted from kilometers to grid points, ensuring precise identification of contiguous AR regions. Each connected region in the initial mask is labeled and evaluated against these criteria, resulting in a robust final binary mask representing AR regions. This mask is used for further analysis and benchmarking, providing a reliable method for identifying ARs in the ERA-5 reanalysis dataset.

4.3 Models

4.3.1 CNN architecture

The U-Net CNN architecture used in this study comprises two main components: the encoder and the decoder, each tailored to handle different aspects of the segmentation task [37]. The encoder extracts meaningful features from the input images through a series of convolutional layers, capturing spatial hierarchies and reducing dimensionality. The decoder is responsible for reconstructing the segmented images from the encoded features. It performs the upsampling process, which involves a series of deconvolutional layers that gradually restore the spatial resolution of the data to its original form. The detailed U-Net CNN architecture used in this study can be seen in **Figure 2**. The input images of size 512x512 are passed through the encoder, which consists of convolutional layers with ReLU activation and max pooling layers to reduce dimensionality. The bottleneck layer captures the most abstract features before passing them to the decoder. The decoder uses bilinear upsampling and convolutional layers to gradually restore the original resolution, culminating in a final 1x1 convolution layer to produce the segmented output.

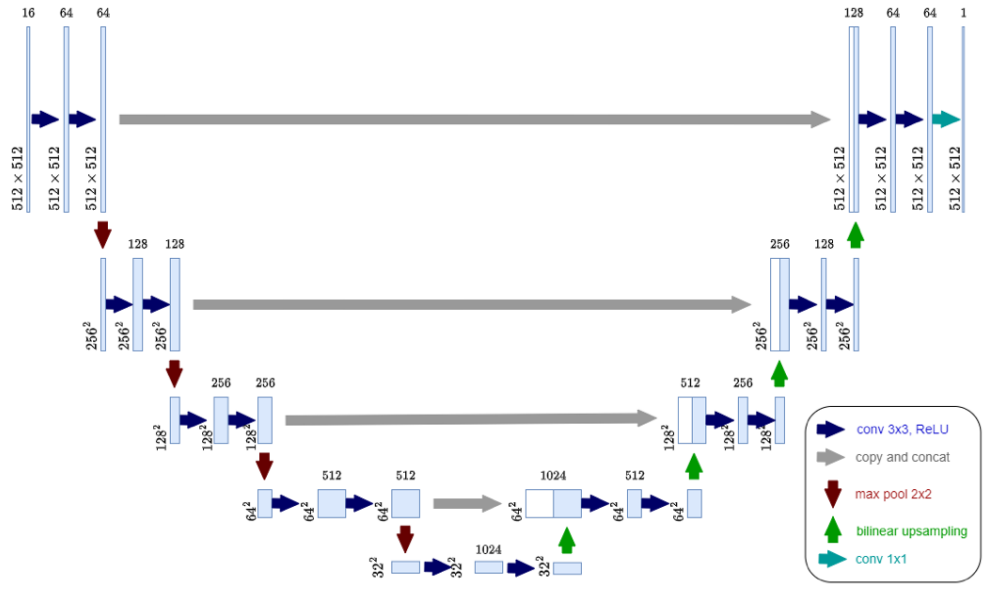


Figure 2: U-Net CNN Architecture. The model features an encoder that reduces dimensionality through convolutional and max-pooling layers, and a decoder that reconstructs the image using up sampling and convolutional layers to restore the original resolution.

The learning rate for our U-Net model was initialized to 0.0001 and dynamically reduced if accuracy metrics plateaued, ensuring a balanced speed of convergence and stability during the training process. A batch size of 16 was selected to manage memory usage efficiently while maintaining stable training dynamics. We employed the Adam optimizer for its adaptive learning rate capabilities, which assist in effectively navigating the loss landscape. The Dice coefficient was used as the primary metric for monitoring model performance. The loss function combined Binary Cross-Entropy (BCE) and Focal Cross-Entropy. The Focal Binary Cross-Entropy loss function was specifically designed to address the class imbalance, as most pixels in our dataset belonged to the non-AR class. This function down-weights the loss for well-classified examples and focuses the model's learning on harder, misclassified examples by applying a modulating factor to the cross-entropy loss. This approach ensures that the model pays more attention to the minority class, improving overall accuracy in AR detection. The combined loss function sums the standard binary cross-entropy loss with the focal loss, enhancing the model's ability to handle class imbalance effectively.

Training was conducted for 20 epochs, which provided sufficient time for the model to converge while minimizing the risk of overfitting. The training process was executed on a GPU-equipped high-performance computing (HPC) system to accelerate computations. The dataset was divided into training and validation sets to evaluate the model's performance, with 80% of the data allocated for training and 20% for testing because 80-20 split strikes a balance between having enough data to train the model and having enough data to validate its performance [37].

4.3.2 Random Forest (RF) Architecture

RF is a versatile classification algorithm that operates by constructing an ensemble of decision trees. Each decision tree in the random forest is trained on a random subset of the features and examples from the dataset, introducing variability and reducing overfitting. In RF, the classification process involves each decision tree making its prediction, and the final classification is determined by aggregating the predictions from all the trees. In our study, RF was used as a secondary approach for identifying ARs because of its straightforward training process and the

relatively short time required to train on the entire dataset. We trained the RF model with all 1420 samples with 100 decision trees (around 7 hours). We used 80% of the data for training and 20% for testing.

4.4. Evaluation metrics

The approach was evaluated by comparing it to labeled masks. Performance evaluation was conducted using six metrics: precision (to assess the proportion of correctly identified AR pixels among those predicted), recall (to measure the proportion of actual AR pixels that were correctly identified), F1 score (to balance precision and recall), intersection over union (IoU, to quantify the overlap between predicted and true AR regions), accuracy (to determine the overall correctness of predictions), and Dice coefficient (to measure the similarity between predicted and true AR masks, with detailed formulas provided in equations 3-7 of the Supplementary materials). Each metric ranges from 0 to 1, with 0 indicating the worst possible performance and 1 indicating the best possible performance.

5. Results

5.1 ERA-5 Benchmark Dataset

The results from the algorithm that we developed (details in section 4.2.3) are promising, accurately identifying AR regions by applying specific IVT and IWV thresholds. The conversion of geometrical criteria from kilometers to grid points ensures precise detection of contiguous AR regions. The final binary mask, representing AR regions, is robust and reliable for further analysis and benchmarking in the ERA-5 reanalysis dataset. The algorithm's effectiveness, showing the IVT, IWV, and resulting AR mask for December 13, 1997, at 12:00 UTC can be seen in Figure 3.

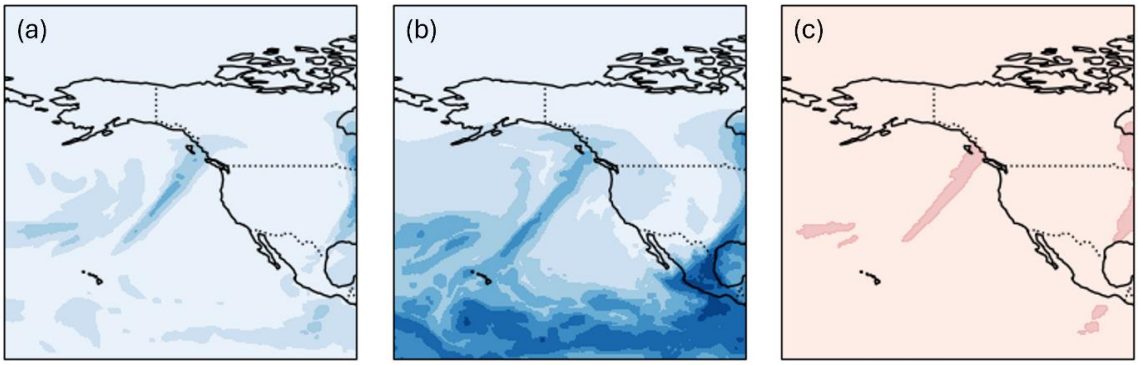


Figure 3: Atmospheric River Identification for December 13, 1997, at 12:00 UTC. (a) Integrated Vapor Transport (IVT), (b) Integrated Water Vapor (IWV), and (c) AR mask indicating identified atmospheric rivers.

The figure illustrates accurate AR region identification, validating the algorithm's ability to differentiate between AR and non-AR areas. This method enhances our understanding of AR events and improves AR detection accuracy in historical climate data.

5.2 Image Segmentation using RF

The RF model was trained using the configurations discussed in section 4.3.1 tailored to optimize performance in segmenting AR events. The AR event snapshots can be seen in **Figure 4**, where **Figure 4(a)** shows the RF model prediction using the full dataset, **Figure 4(b)** showing the labeled dataset that we used as ground truth as a target variable and **Figure 4(c)** as evaluating

both of their performance using a confusion matrix. The average precision was 0.3448, the F1-score was 0.113, and the IoU was 0.02, demonstrating significant improvements in precision and overlap between predicted and actual AR regions.

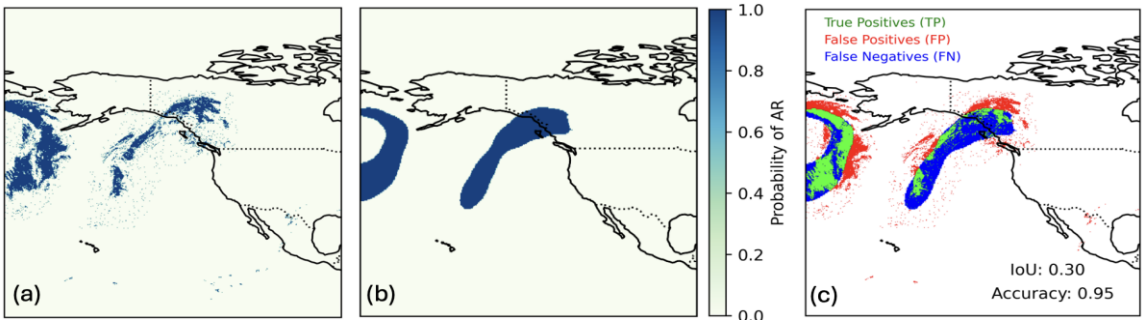


Figure 4: Random Forest (RF) model predictions compared with labeled data and evaluation results. The top row shows (a) RF model prediction, (b) labeled data, and (c) evaluation of the model which has an IoU of 0.34 and an accuracy of 0.94.

A low Dice coefficient in segmentation tasks suggests poor spatial overlap between predicted and ground truth masks. This can happen if the model fails to capture the correct boundaries or details of the segmented objects, leading to mismatches with the ground truth. If the model is too simplistic (underfitting), it may not capture the complexities of the data, resulting in poor performance across all metrics, including the F1 score and Dice coefficient. Low-quality data, noise, or artifacts (such as speckles in radar or SAR images) can adversely affect segmentation and classification tasks, leading to lower F1 scores and Dice coefficients.

5.3 Image Segmentation using CNN

The performance of the CNN model for image segmentation is illustrated in **Figure 5**. This figure compares the GOES imagery, ground truth, and model output for specific times and locations. The key metrics for evaluating the model's performance include the Intersection over Union (IoU) and accuracy. The original GOES imagery, which captures the atmospheric conditions over the Pacific West Coast using bands 8, 9, and 10 in the RGB composite can be seen in **Figure 5(a)**. These bands highlight lower brightness temperatures associated with water vapor, often appearing as black filament-like structures in the raw data, which correspond to ARs. The ground truth data, depicted in **Figure 5(b)**, highlights the actual locations of ARs based on manually annotated or benchmark data. The output of the model predicting the AR presence probability across the region can be seen in **Figure 5(c)**.

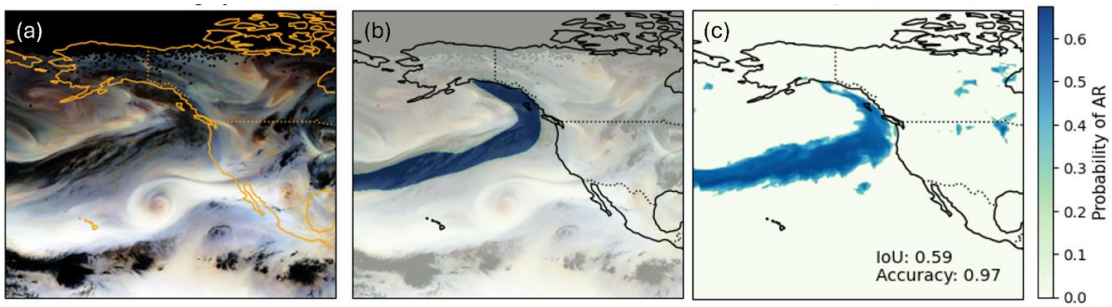


Figure 5: AR identification using our U-Net model trained on GOES imagery, showing (a) brightness temperature from the water vapor bands (Bands 8-10) as a False Color Composite showing complex cloud patterns and atmospheric features, (b) The labeled dataset provided for training, highlighting the atmospheric river region in white, and (c) the prediction made by the U-Net model, indicating the probability of an atmospheric river (AR), where lighter shades represent higher probabilities.

The model achieved an Intersection over Union (IoU) of 0.23 over the entire dataset, indicating the overlap between the predicted AR regions and the ground truth. Additionally, the model demonstrated a high accuracy of 0.92, reflecting the proportion of correctly identified pixels in the segmentation task. The CNN model shows robust performance in identifying the geometry associated with AR systems, accurately capturing the elongated and filamentous nature of these structures. However, the IoU is low, partly due to the nature of the ground truth, which comprises smooth objects. The clouds that are part of the AR system often have irregular and non-smooth boundaries, as predicted by our model, leading to a decreased IoU. Further improved performance could be achieved by training the model for more than 20 epochs, allowing it to better learn the complex patterns associated with ARs.

Despite the model's robustness, there are instances where it fails. For example, the input dataset sometimes contains corrupted images with striped artifacts, as shown in Figure S1-3 in the supplementary information. These artifacts degrade the model's performance by providing inaccurate data during training and inference. Additionally, discrepancies between the ground truth mask and the actual AR locations can also lead to a loss in accuracy. This misalignment is evident in Figure S3, where the ground truth does not perfectly match the AR structures apparent in the raw GOES imagery. While the CNN model demonstrates strong potential for real-time AR detection and monitoring, addressing these challenges and refining the training process can further enhance its precision and reliability.

5.4 Evaluation Metrics

The evaluation metrics assessed our overall model performance from different and complementary angles, as seen in **Figure 6**. While the accuracy shows a high performance for both models, that is not reflected by the rest of the metrics equivalently. Our precision and recall indicate a large number of false positives and false negatives, respectively, in both models. Our F1 score, IoU, and Dice coefficient values further reflect this. Nevertheless, the metrics indicate better performance by CNN showing its superiority over the RF mode. This shows the CNN model's potential for the identification of ARs using GOES satellite data with further improvement, as discussed in section 5.3.

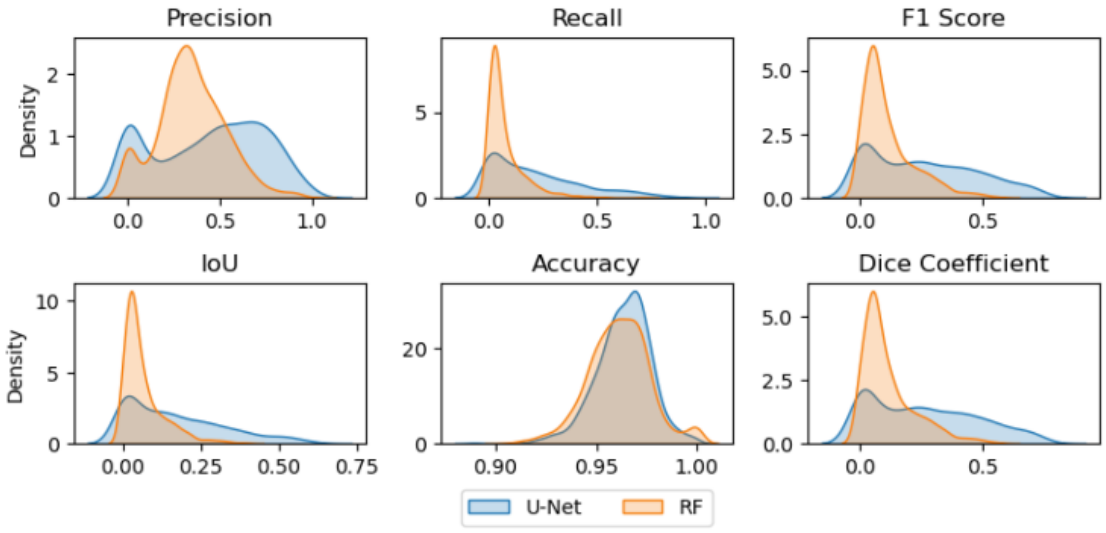


Figure 6: Density plots of evaluation metrics (Precision, Recall, F1 Score, IoU, Accuracy, and Dice Coefficient) for U-Net and RF. These plots illustrate the performance distribution of each model across the different metrics.

6. Conclusion

ARs are important climate phenomena in global moisture transport and regional precipitation, with dual roles of mitigating droughts and causing floods. This study demonstrates the potential of using GOES imagery combined with machine learning techniques to identify ARs along the Pacific coast of North America. By leveraging the high temporal resolution of GOES satellite data, our U-Net CNN and random forest models provided valuable insights into AR detection and segmentation. The U-Net CNN model exhibited robust performance, achieving high accuracy in segmenting ARs and demonstrating the feasibility of using satellite imagery for real-time AR detection. Despite some challenges in capturing precise boundaries and details of ARs, the model's ability to identify the general structure of ARs is promising for operational monitoring. Although not on a par with the CNN model, the random forest model showcased the utility of simpler classification methods as well.

While this project has made significant strides in leveraging GOES satellite data for AR identification, several areas remain for future exploration. Further research could focus on improving the preprocessing techniques for satellite imagery to enhance model accuracy, especially removing striping in the input imagery. Additionally, incorporating more advanced machine learning models, such as deep learning architectures beyond U-Net, may improve AR detection performance. Extending the temporal analysis beyond 2022 and including more comprehensive climatic data could offer deeper insights into AR patterns and long-term trends. Moreover, real-time implementation and operational testing of these models in practical settings will be essential to evaluate their robustness and reliability in dynamic environments. The potential for GOES satellite data in atmospheric river detection can further be reinforced by exploring these future research directions.

Acknowledgements

We would like to thank James Halgren (Alabama Water Institute), Fred Ogden (NOAA OWP), Alan Rhoades (Lawrence Berkeley National Lab), Andrew Bennett (University of Arizona), and GOES-SWE Group (Group 1) members for their help and insightful discussions. We also extend our sincere appreciation to Karina Larco and Sadaf Mahmoudi for their all-encompassing help throughout the study period.

Supplementary Materials

Code and data links for this research are available at: https://github.com/NWC-CUAHSI-Summer-Institute/SI24_GOES-ARs

S1: About GOES imagery

GOES 16 to 18 (GOES-R Series): https://www.noaa.gov/jetstream/satellites/goes-west-goes-17

- Visible bands: Channel 1 and 2 are Blue and Red bands - this is the highest resolution band coming from GOES 16
- Near-infrared bands: Channel 3 (Veggie band), 4 (Cirrus band), 5 (Snow/Ice band), 6 (Cloud particle size), and 7 (Shortwave window)

- Water Vapor bands: Channel 8 (Upper-level Water Vapor band), 9 (Mid-level Water Vapor band), 10 (Lower-level Water Vapor band).
- Infrared bands: Channel 11 (Cloud-top phase), 12 (Ozone band), 13 (Clean infrared band), 14 (Traditional Infrared band), 15 (The Dirty Infrared band), 16 (Carbon Dioxide band), and True Color

S2: Calculation of IWV and IVT

$$IWV = \frac{1}{g}\int_{1000\ hPa}^{300\ hPa} q \cdot dp \tag{1}$$

$$IVT = \frac{1}{g}\sqrt{\left(\int_{1000\ hPa}^{300\ hPa} q \cdot u\ dp\right)^2 + \left(\int_{1000\ hPa}^{300\ hPa} q \cdot v\ dp\right)^2} \tag{2}$$

where g is the acceleration due to gravity, wind components (u, v) and specific humidity (q) at multiple pressure levels (300 hPa to 1000 hPa).

S2: Evaluation Metrics

$TP, TN, FP\ and\ FN$ are true positive, true negative, false positive, and false negative respectively.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1\ Score = 2\ X \frac{Precision\ X\ Recall}{Precision + Recall} \tag{5}$$

$$Jaccard\ Index = \frac{TP}{TP + FP + FN} \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Dice\ coeficient = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{8}$$

S3: CNN architecture & Image segmentation using CNN

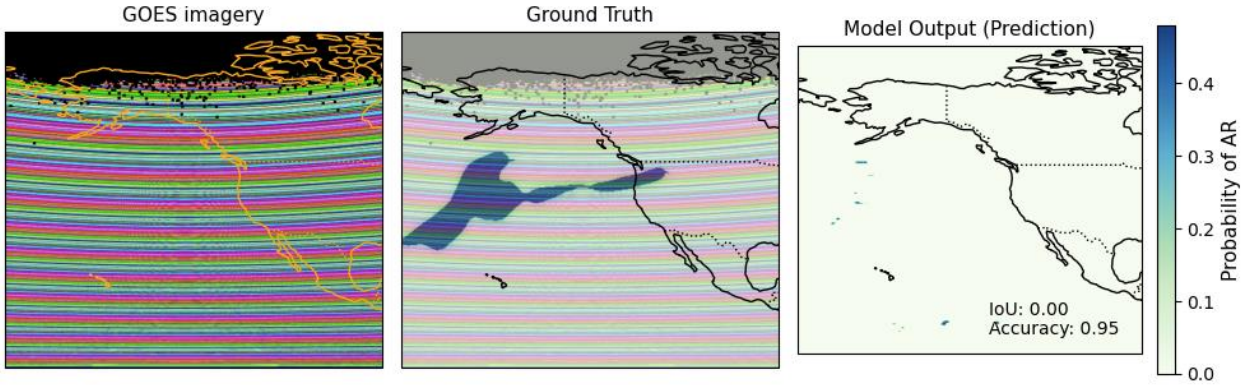


Figure S1: Example of corrupted input data with striping artifacts. (a) GOES imagery showing artifacts, (b) ground truth AR locations, and (c) model output predictions. The presence of artifacts negatively impacts the model's performance, resulting in an IoU of 0.00 despite a high accuracy of 0.95.

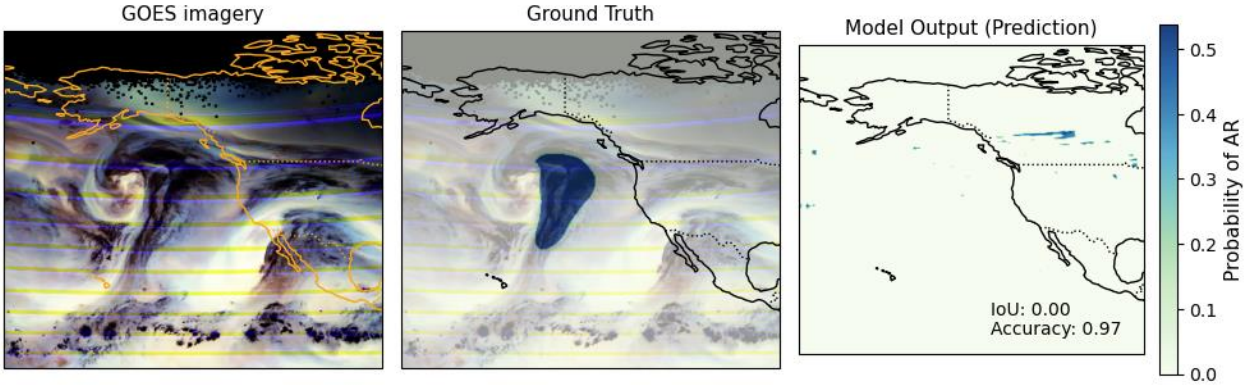


Figure S2: Even slight striping effects cause a considerable decrease in accuracy. (a) GOES imagery showing artifacts, (b) ground truth AR locations, and (c) model output predictions. The presence of artifacts negatively impacts the model's performance, resulting in an IoU of 0.00 despite a high accuracy of 0.97.

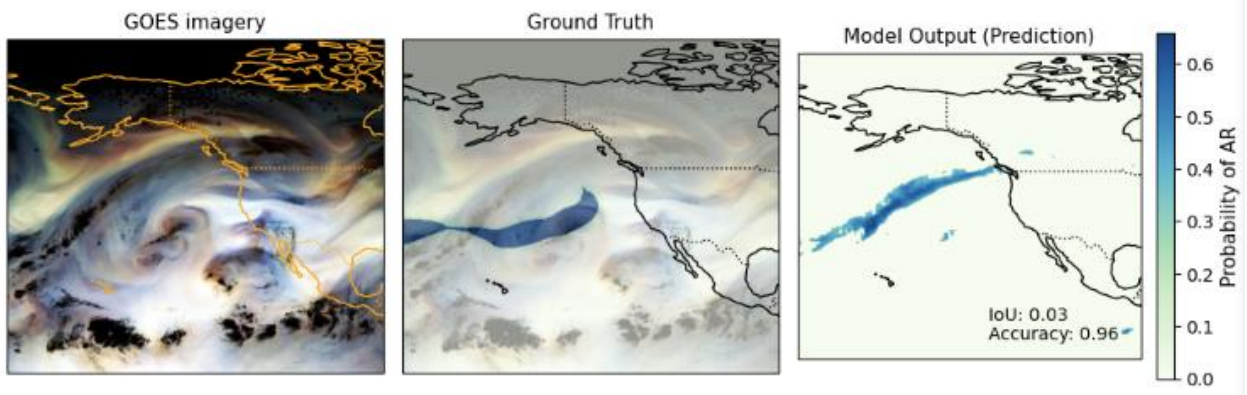


Figure S3: Comparison of (a) GOES imagery, (b) ground truth AR locations, and (c) model output predictions. This figure highlights discrepancies due to misalignment between ground truth and actual AR structures, leading to an IoU of 0.03 and an ccuracy of 0.96.

References

[1] F. M. Ralph, M. D. Dettinger, J. J. Rutz, and D. E. Waliser, Eds., *Atmospheric Rivers*. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-28906-5.

[2] L. Gimeno, R. Nieto, M. Vázquez, and D. A. Lavers, "Atmospheric rivers: a mini-review," *Front. Earth Sci.*, vol. 2, Mar. 2014, doi: 10.3389/feart.2014.00002.

[3] A. E. Payne *et al.*, "Responses and impacts of atmospheric rivers to climate change," *Nat. Rev. Earth Environ.*, vol. 1, no. 3, pp. 143–157, Mar. 2020, doi: 10.1038/s43017-020-0030-5.

[4] S. Wang *et al.*, "Extreme atmospheric rivers in a warming climate," *Nat. Commun.*, vol. 14, no. 1, p. 3219, Jun. 2023, doi: 10.1038/s41467-023-38980-x.

[5] V. Espinoza, D. E. Waliser, B. Guan, D. A. Lavers, and F. M. Ralph, "Global Analysis of Climate Change Projection Effects on Atmospheric Rivers," *Geophys. Res. Lett.*, vol. 45, no. 9, pp. 4299–4308, 2018, doi: 10.1029/2017GL076968.

[6] C. A. Shields *et al.*, "Atmospheric River Tracking Method Intercomparison Project (ARTMIP): project goals and experimental design," *Geosci. Model Dev.*, vol. 11, no. 6, pp. 2455–2474, Jun. 2018, doi: 10.5194/gmd-11-2455-2018.

[7] M. Viale, R. Valenzuela, R. D. Garreaud, and F. M. Ralph, "Impacts of Atmospheric Rivers on Precipitation in Southern South America," *J. Hydrometeorol.*, vol. 19, no. 10, pp. 1671–1687, Oct. 2018, doi: 10.1175/JHM-D-18-0006.1.

[8] D. Galea, H.-Y. Ma, W.-Y. Wu, and D. Kobayashi, "Deep Learning Image Segmentation for Atmospheric Rivers," *Artif. Intell. Earth Syst.*, vol. 3, no. 1, Jan. 2024, doi: 10.1175/AIES-D-23-0048.1.

[9] V. Thandlam, A. Rutgersson, and E. Sahlee, "Spatio-temporal variability of atmospheric rivers and associated atmospheric parameters in the Euro-Atlantic region," *Theor. Appl. Climatol.*, vol. 147, no. 1, pp. 13–33, Jan. 2022, doi: 10.1007/s00704-021-03776-w.

[10] W. E. Chapman, A. C. Subramanian, L. Delle Monache, S. P. Xie, and F. M. Ralph, "Improving Atmospheric River Forecasts With Machine Learning," *Geophys. Res. Lett.*, vol. 46, no. 17–18, pp. 10627–10635, 2019, doi: 10.1029/2019GL083662.

[11] S. Asharaf, B. Guan, and D. E. Waliser, "ROTATE: A Coordinate System for Analyzing Atmospheric Rivers," *Geophys. Res. Lett.*, vol. 51, no. 7, p. e2023GL106736, 2024, doi: 10.1029/2023GL106736.

[12] F. M. Ralph *et al.*, "ARTMIP-early start comparison of atmospheric river detection tools: how many atmospheric rivers hit northern California's Russian River watershed?," *Clim. Dyn.*, vol. 52, no. 7, pp. 4973–4994, Apr. 2019, doi: 10.1007/s00382-018-4427-5.

[13] Y. Zhou, T. A. O'Brien, P. A. Ullrich, W. D. Collins, C. M. Patricola, and A. M. Rhoades, "Uncertainties in Atmospheric River Lifecycles by Detection Algorithms: Climatology and Variability," *J. Geophys. Res. Atmospheres*, vol. 126, no. 8, p. e2020JD033711, 2021, doi: 10.1029/2020JD033711.

[14] L. S. Huning, S. A. Margulis, B. Guan, D. E. Waliser, and P. J. Neiman, "Implications of Detection Methods on Characterizing Atmospheric River Contribution to Seasonal

Snowfall Across Sierra Nevada, USA," *Geophys. Res. Lett.*, vol. 44, no. 20, p. 10,445-10,453, 2017, doi: 10.1002/2017GL075201.

[15] H.-R. Wang *et al.*, "Occurrence Frequency of Global Atmospheric River (AR) Events: A Data Fusion Analysis of 12 Identification Data Sets," *J. Geophys. Res. Atmospheres*, vol. 129, no. 5, p. e2023JD039730, 2024, doi: 10.1029/2023JD039730.

[16] K. J. Reid, A. D. King, T. P. Lane, and E. Short, "The Sensitivity of Atmospheric River Identification to Integrated Water Vapor Transport Threshold, Resolution, and Regridding Method," *J. Geophys. Res. Atmospheres*, vol. 125, no. 20, p. e2020JD032897, 2020, doi: 10.1029/2020JD032897.

[17] C. Ordaz Osorio, J. F. Booth, A. N. LeGrande, and C. M. Naud, "Assessing the Impact of Climate Change on Atmospheric Rivers: A Modeling Perspective," *J. Geophys. Res. Atmospheres*, vol. 129, no. 12, p. e2023JD040074, 2024, doi: 10.1029/2023JD040074.

[18] B. Guan and D. E. Waliser, "A regionally refined quarter-degree global atmospheric rivers database based on ERA5," *Sci. Data*, vol. 11, no. 1, p. 440, May 2024, doi: 10.1038/s41597-024-03258-4.

[19] S. Brands, J. M. Gutiérrez, and D. San-Martín, "Twentieth-century atmospheric river activity along the west coasts of Europe and North America: algorithm formulation, reanalysis uncertainty and links to atmospheric circulation patterns," *Clim. Dyn.*, vol. 48, no. 9, pp. 2771–2795, May 2017, doi: 10.1007/s00382-016-3095-6.

[20] G. Xu, X. Ma, P. Chang, and L. Wang, "A Comparison of Northern Hemisphere Atmospheric Rivers Detected by a New Image-Processing Based Method and Magnitude-Thresholding Based Methods," *Atmosphere*, vol. 11, no. 6, Art. no. 6, Jun. 2020, doi: 10.3390/atmos11060628.

[21] B. Guan, D. E. Waliser, and F. M. Ralph, "Global Application of the Atmospheric River Scale," *J. Geophys. Res. Atmospheres*, vol. 128, no. 3, p. e2022JD037180, 2023, doi: 10.1029/2022JD037180.

[22] M. Pan and M. Lu, "A Novel Atmospheric River Identification Algorithm," *Water Resour. Res.*, vol. 55, no. 7, pp. 6069–6087, 2019, doi: 10.1029/2018WR024407.

[23] A. Mahesh, T. A. O'Brien, B. Loring, A. Elbashandy, W. Boos, and W. D. Collins, "Identifying atmospheric rivers and their poleward latent heat transport with generalizable neural networks: ARCNNv1," *Geosci. Model Dev.*, vol. 17, no. 8, pp. 3533–3557, May 2024, doi: 10.5194/gmd-17-3533-2024.

[24] T. B. Higgins *et al.*, "Using Deep Learning for an Analysis of Atmospheric Rivers in a High-Resolution Large Ensemble Climate Data Set," *J. Adv. Model. Earth Syst.*, vol. 15, no. 4, p. e2022MS003495, 2023, doi: 10.1029/2022MS003495.

[25] F. Lan, B. Gamelin, L. Yan, J. Wang, B. Wang, and H. Guo, "Topological Characterization and Uncertainty Visualization of Atmospheric Rivers," *Comput. Graph. Forum*, vol. 43, no. 3, p. e15084, 2024, doi: 10.1111/cgf.15084.

[26] J. J. Rutz *et al.*, "The Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Quantifying Uncertainties in Atmospheric River Climatology," *J. Geophys. Res. Atmospheres*, vol. 124, no. 24, pp. 13777–13802, 2019, doi: 10.1029/2019JD030936.

[27] S. Meghani, S. Singh, N. Kumar, and M. K. Goyal, "Predicting the spatiotemporal characteristics of atmospheric rivers: A novel data-driven approach," *Glob. Planet. Change*, vol. 231, p. 104295, Dec. 2023, doi: 10.1016/j.gloplacha.2023.104295.

[28] S. Singh and M. K. Goyal, "An innovative approach to predict atmospheric rivers: Exploring convolutional autoencoder," *Atmospheric Res.*, vol. 289, p. 106754, Jul. 2023, doi: 10.1016/j.atmosres.2023.106754.

[29] M. D. Dettinger, "Atmospheric Rivers as Drought Busters on the U.S. West Coast," *J. Hydrometeorol.*, vol. 14, no. 6, pp. 1721–1732, Dec. 2013, doi: 10.1175/JHM-D-13-02.1.

[30] S. D. Schubert *et al.*, "Insights into the Causes and Predictability of the 2022/23 California Flooding," *J. Clim.*, vol. 37, no. 13, pp. 3613–3629, Jun. 2024, doi: 10.1175/JCLI-D-23-0696.1.

[31] N. P. Gillett *et al.*, "Human influence on the 2021 British Columbia floods," *Weather Clim. Extrem.*, vol. 36, p. 100441, Jun. 2022, doi: 10.1016/j.wace.2022.100441.

[32] Blaylock, B. K., "GOES-2-go: Download and display GOES-East and GOES-West data (Version 2022.07.15) [Computer software]." https://github.com/blaylockbk/goes2go.

[33] Rhoades, Alan, "Labeled ARs using TempestExtremes Algorithm." https://portal.nersc.gov/archive/home/a/arhoades/Shared/www/TE_ERA5_ARs.

[34] K. R. Knapp and S. L. Wilkins, "Gridded Satellite (GridSat) GOES and CONUS data," *Earth Syst. Sci. Data*, vol. 10, no. 3, pp. 1417–1425, Aug. 2018, doi: 10.5194/essd-10-1417-2018.

[35] P. A. Ullrich, C. M. Zarzycki, E. E. McClenny, M. C. Pinheiro, A. M. Stansfield, and K. A. Reed, "TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in large datasets," *Geosci. Model Dev.*, vol. 14, no. 8, pp. 5023–5048, Aug. 2021, doi: 10.5194/gmd-14-5023-2021.

[36] C. A. Ramseyer *et al.*, "Identifying Eastern US Atmospheric River Types and Evaluating Historical Trends," *J. Geophys. Res. Atmospheres*, vol. 127, no. 17, p. e2021JD036198, 2022, doi: 10.1029/2021JD036198.

[37] Z. Xu *et al.*, "An attention U-Net model for detection of fine-scale hydrologic streamlines," *Environ. Model. Softw.*, vol. 140, p. 104992, Jun. 2021, doi: 10.1016/j.envsoft.2021.104992.

Chapter 3
Probabilistic Streamflow Prediction Using the Model-Agnostic NextGen Framework

Ali Sattari[1], Reza Morovati[2], Hossein Gholizadeh[1]

[1]University of Alabama; asattaripirsoltan@crimson.ua.edu/ hgholizadeh@crimson.ua.edu

[2]Utah State University; reza.morovati@usu.edu

Academic Advisors: Hamid Moradkhani, *University of Alabama*; David Tarboton, *Utah State University*, Yong Zhang, University of Alabama

Summer Institute Theme Advisors: Jonathan M. Frame, *University of Alabama* jmframe@ua.edu

Abstract: Accurate streamflow prediction is influenced by uncertainties in initial conditions, forcing data, model structure, and parameters. It is crucial for water supply management, flood control, and various environmental applications. The accuracy of these predictions is influenced by multiple sources of uncertainty, including initial conditions, forcing data, model structure, and parameters. This study aims to quantify the uncertainty of forcing inputs to probabilistically predict streamflow within gauged catchments. Utilizing the NextGen Framework and the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) basin dataset, we implemented the Dynamically Dimensioned Search (DDS) calibration method for the Conceptual Functional Equivalent (CFE) Model. By identifying appropriate perturbation factors and ensemble sizes in the training dataset, we enhanced the reliability of streamflow predictions. The results were compared against USGS streamflow data, that is a reliable source of accurate data, demonstrating improved prediction accuracy in gauged basins.

---

1. Motivation

The primary motivation of this study is the lack of streamflow probabilistic analysis by the Next Generation Water Resources Modeling (NextGen) Framework. To do this, we propose a probabilistic approach using the NextGen Framework to quantify the uncertainty in forcing data for the streamflow predictions across three CAMEL selected basins. By leveraging this framework, we aim to enhance the accuracy and reliability of streamflow predictions, ultimately contributing to improved water resource management and decision-making processes.

Additionally, this study seeks to further our understanding of the NextGen Framework as a model-agnostic tool in the context of hydrological modeling. By employing NextGen, we aim to demonstrate its versatility and robustness in handling various hydrological models, thereby broadening its applicability and utility in the field.

To this end, the CFE Model with the Basic Model Interface (BMI) was utilized for this project analysis. The integration of CFE-BMI within the NextGen Framework allows handling model inputs and outputs to be dealt with facilitating a comprehensive evaluation of the framework's performance in streamflow prediction [1]. This approach promises to yield valuable insights into the capabilities of NextGen while advancing the state of the art in hydrological modeling and uncertainty quantification. The study will also contribute to the broader scientific community by showcasing the benefits of a model-agnostic framework, promoting collaboration and

interoperability among researchers. Additionally, it will enable evidence-driven model selection [2].

2. Objectives and Scope

2.1. Assess CFE Model Performance with NextGen

The CFE model is used in the NextGen framework to assess its performance in three CAMEL-selected basins. This implementation will leverage resources from the CFE repository [A1], which is accessible. The goal is to utilize the NextGen Framework's model-agnostic capabilities to integrate and evaluate the CFE model, ensuring comprehensive performance analysis in diverse hydrological conditions.

2.2. Model Calibration Using Ngen-cal

Calibrate and validate the NextGen model for the selected catchments using the ngen-cal tool [A2]. The calibration will involve adjusting model parameters to align simulated streamflow with observed data, followed by rigorous validation to ensure the model's accuracy and robustness. This step is critical for fine-tuning the model to accurately reflect real-world hydrological dynamics.

2.3 Uncertainty Quantification

Quantify the uncertainties associated with forcing files to enhance the reliability of streamflow predictions. This involves systematically analyzing the variability in input data, such as precipitation and temperature, and their impact on model outputs. By identifying and characterizing these uncertainties, we aim to improve the model's predictive capabilities and provide more reliable streamflow forecasts.

2.4 Probabilistic Prediction

Another objective of this project is to develop probabilistic predictions of streamflow based on the quantified uncertainties. This approach involves generating a range of possible streamflow scenarios, rather than a single deterministic forecast, to better capture the inherent variability and uncertainty in hydrological processes. Probabilistic predictions will support more informed water resource management and decision-making by providing potential outcomes and their associated probabilities.

3. Previous Studies

In previous studies, the National Water Center's CFE model, which is a simplified hydrological model, has been applied to estimate the volume of water flowing into rivers and streams following rain events [3-4]. This model has demonstrated effectiveness in capturing the essential dynamics of rainfall-runoff processes. Additionally, the NextGen Framework has been used for data assimilation, integrating real-time observations with model simulations to enhance the accuracy of hydrological forecasts [5]. Wolkeba et al. [6] also conducted data assimilation of USGS streamflow using CFE model. Our research aims to build on these studies by quantifying the uncertainty in streamflow predictions within the CAMELS basins, utilizing the strengths of both the CFE model and the NextGen Framework to improve prediction reliability.

4. Methodology

4.1. Study area

For this project, three CAMELS basins located in the northwest of Sacramento, California, were selected based on the findings by Bolotin et al. [7] and are shown in Figure 1. Their study indicated that the Normalized Nash-Sutcliffe Efficiency Error (NNSE) showed satisfying performance in these basins, with values close to one indicating better performance. According to that study, we chose three basins with promising NNSE values.

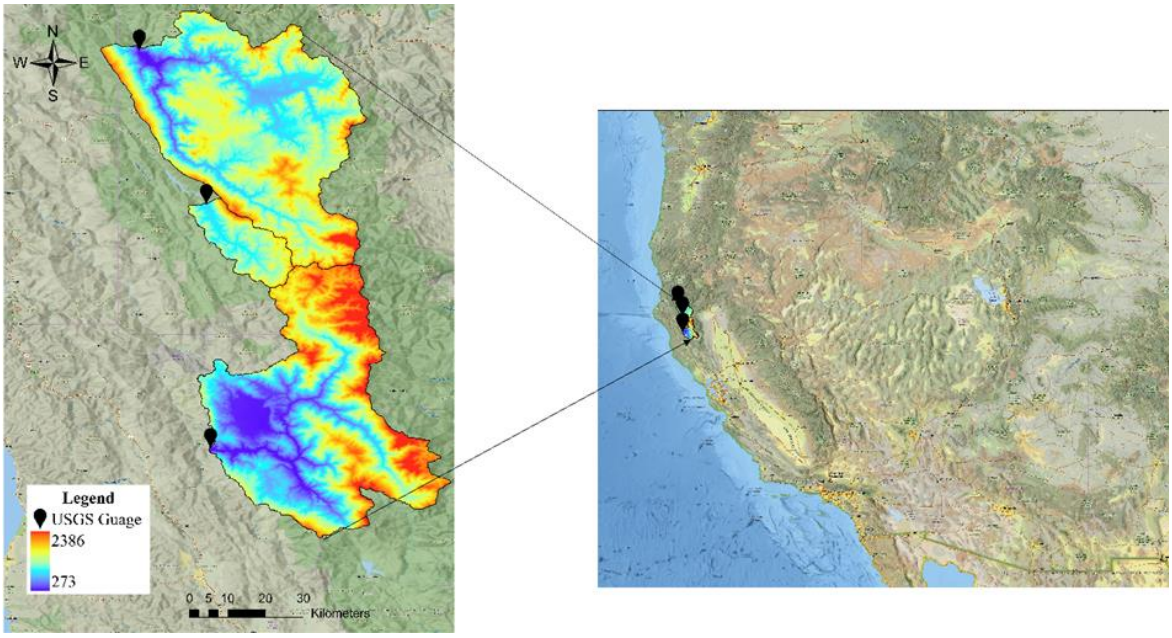


Figure 1. Location of the three selected CAMELS basins northwest of Sacramento, California.

Detailed information about these basins, including their physiographic characteristics, is provided in Table 1.

Table1. Physiographic characteristics of basins

Gauge ID (CAMEL ID)	Mean elevation (m)	Mean Slope (m/km)	Area (km²)	Average precipitation (mm)	Average Temperature (℃)
11480390	1049.64	76.61899	242.6	1327.07	10.2
11528700	1019.07	107.622	1980.08	1106.92	10.5
11473900	1037.85	126.7067	1925.01	1232.26	11.3

4.2. Hydrofabric for Next Generation Water Resource Modeling

The NextGen hydrofabric artifacts are distributed by NHDPlusV2 Vector Processing Units. These artifacts are generated from national reference datasets developed in collaboration between the USGS and NOAA for federal water modeling efforts. Designed for easy updates, manipulation, and quality control, they aim to meet diverse modeling needs while utilizing the best possible input data. Moreover, to better understand the hydrological characteristics of the

basins, hydrofabric outputs [8-9] of these basins are provided in Figure 2. In this figure, NextGen hydrofabric provides information on divides, flowlines, and nexuses.

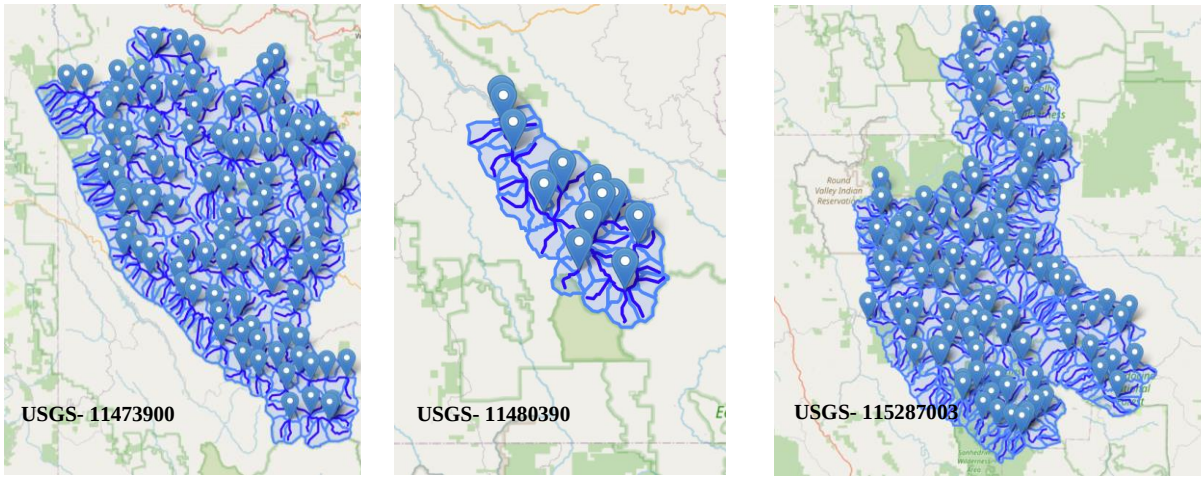


Figure 2. Hydrofabric output of the three selected basin

4.2. Data Preparation

Forcing data is used from North American Land Data Assimilation System (NLDAS-2) and also for Potential Evapotranspiration (PET) data used from Gauch et al. [10] research. NLDAS-2 dataset encompasses various meteorological elements, including shortwave and longwave radiation, specific humidity, air temperature, surface pressure, near-surface wind (u and v components), and precipitation. The NLDAS-2 datasets are notable for their fine spatial resolution of approximately 12.5 km and an hourly temporal resolution, extending from January 1979 to the present day. The project period, selected from 2008 to 2012, utilized hourly data. Additionally, USGS streamflow data is employed for the analysis. Additionally, a comparison between the National Water Model (NWM) and USGS streamflow observations for three selected basins is provided in the supplementary materials (Figures 3-5) to demonstrate its performance.

4.3. NextGen Framework

The NextGen Framework is a groundbreaking initiative undertaken by the NOAA National Weather Service (NWS), Office of Water Prediction (OWP) [11]. Launched in 2019, NextGen represents a significant evolution in hydrological modeling, aiming to enhance the predictive capabilities of the U.S. National Water Model. This refactoring effort integrates two critical standards: the Community Surface Dynamics Modeling System (CSDMS) Basic Model Interface (BMI) version 2.0 for model coupling and the Open Geospatial Consortium's WaterML version 2.0, part 3, Hydrologic Features (HY_Features) conceptual data model. These standards ensure a high level of interoperability and data consistency across diverse modeling systems.

Since early 2020, the NextGen prototype has been actively developed by NOAA/OWP staff and contractors, including ERT and Lynker. In September 2023, a contract was awarded to Raytheon to develop the NextGen Framework further, Improved Flood Inundation Mapping (FIM), and Enhanced Hydrofabric, which is provided on GitHub [A3]. Collectively, these advancements are known as the Next Generation Water Prediction System Capability (NGWPC). The NGWPC aims to operationalize Version 4 of the National Water Model and

enhance flood inundation mapping, thereby improving the nation's ability to predict and respond to water-related hazards.

The development of NextGen is characterized by its collaborative and standards-based approach [10], ensuring compatibility between the various iterations created by different stakeholders, including NOAA/OWP, Raytheon, and the Cooperative Institute for Research to Operations in Hydrology (CIROH). NextGen's development and governance emphasize a high level of technological readiness and operational value, with a structured pathway for transitioning research advancements into operational capabilities. Regular interactions and updates among CIROH, NOAA/OWP, and Raytheon ensure that the standards for model coupling, and data modeling are consistently applied, facilitating the smooth integration of new technologies and methodologies.

Overall, NextGen represents a significant advancement in water resources modeling, leveraging cutting-edge technologies and collaborative efforts to enhance the predictive accuracy and operational capabilities of water prediction systems in the United States [12-13].

4.3. Dynamic Dimensioned Search (DDS)

Dynamic Dimensioned Search (DDS) [14-16] is an advanced global optimization algorithm suited for complex systems' calibration problems. It builds upon the foundational principles of the Direct Dimensioned Search (DDS) method but introduces additional strategies to handle high-dimensional search spaces more efficiently.

4.4 Analyzing Hydrological Processes Using the CFE Model

The CFE model is a conceptual adaptation of the original NWM, comprising seventeen parameters [11]. For those interested in a more detailed exploration of the CFE model, the GitHub repository [A4] provides comprehensive information, including source code, documentation, and examples.

5. Results

For this project, our group's GitHub repository (https://github.com/NWC-CUAHSI-Summer-Institute/NextGen-Uncertainty-quantification/tree/main) stands out as a significant achievement. The main goal of this repository is to ensure the reproducibility of our processes for future CUAHSI Summer Institute fellows, thereby accelerating their progress. While it is not a tutorial, it meticulously documents all the issues and challenges we faced during the project, providing valuable insights and easing future participants' paths.

5.1. CFE-Py Calibrated Model

The CFE model was calibrated for the USGS 11480390-gauge basin, which, due to its similar regional climate and proximity to the other two basins in the study area (see Table 1), is considered representative for this project. The calibration period spans from 2008 to the end of 2009, the validation period is from 2010 to 2011, and the testing period is from 2011 to 2012. Results of the calibration and testing phase are provided in Figure 6 and Figure 7. Also, a calibrated model is implemented for 2012-01 to 2012-12 as the testing period (Figure 7).

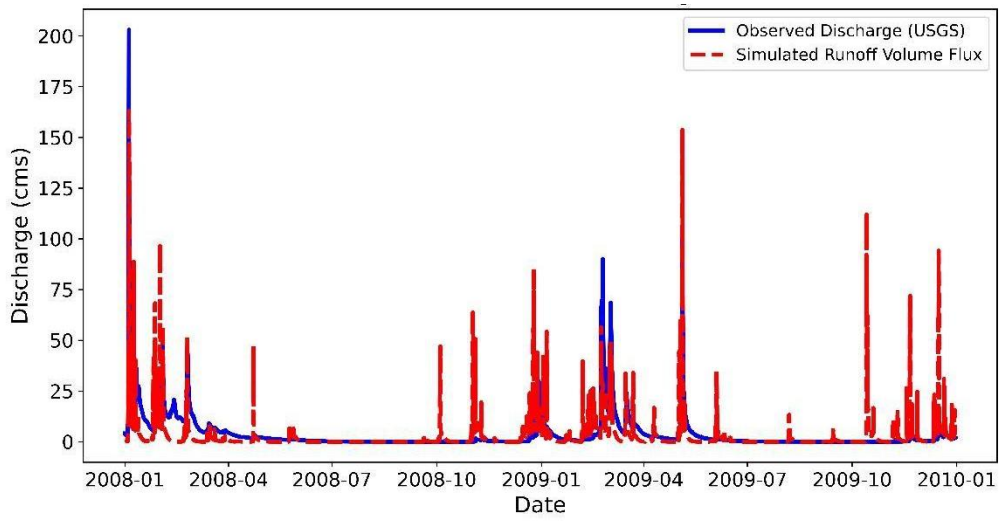


Figure 6. CFE calibrated model result and USGS observations of the selected basin

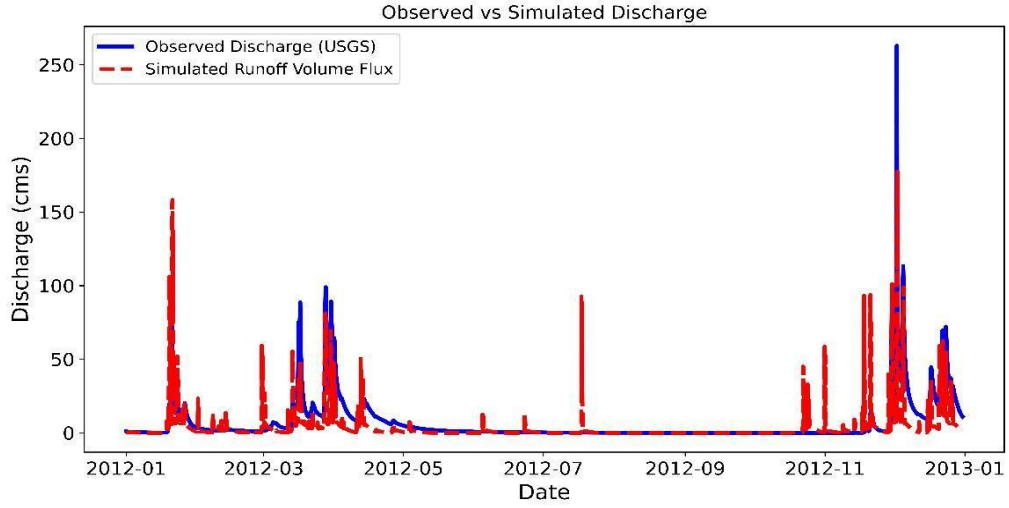


Figure 7. CFE test result and USGS observations of the selected basin (NNSE = 0.63)

5.2 Hyperparameter Tuning for Uncertainty Quantification

5.2.1 Hyperparameter Tuning

In this study, we performed hyperparameter tuning to find the optimal perturbation factor and ensemble size. This was based on the score of the probabilistic performance metric, such as the Normalized Root Mean Square Error Ratio (NRR), in the validation dataset. This tuning process helps to understand the appropriate perturbation factor and ensemble size to quantify the forcing uncertainty. This process involved testing various combinations of those hyperparameters, calibrating the model and then calculating NRR in the validation dataset. The range of perturbation factors was between 5-25% and 10-100 for ensemble size. This tuning process helps to understand the appropriate perturbation factor and ensemble size to quantify the forcing uncertainty.

5.2.1 Uncertainty Quantification

The uncertainty quantification of forcing data used a perturbation-based method. This approach involves varying the input forcing data by specific percentages and ensemble size and analyzing the predictive interval in the validation dataset based on probabilistic performance metrics. The perturbation process was implemented by creating ensembles of forcing datasets, with ensemble sizes ranging from 10 to 100 with different perturbation factors. The resulting predictive

intervals for streamflow predictions were then analyzed to assess the uncertainty. Therefore, this method provides the uncertainty quantification of forcing data. The predictive interval for the test period is shown in Figure 8, while the event-based predictive interval is depicted in Figure 10. To enhance the clarity of these predictions, precipitation rates are included in Figures 9 and 11. Additionally, the predictive intervals for the simulated streamflow using the CFE model are matched against the USGS streamflow data in general.

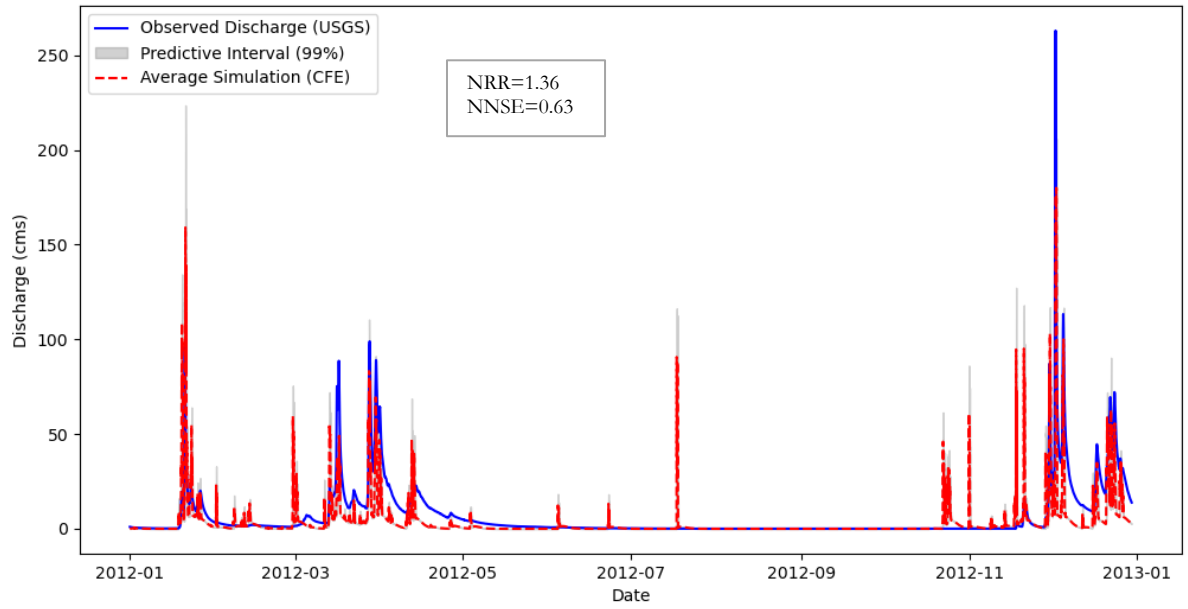


Figure 8. CFE model test and predictive interval for a one-year time period (test) for gauge-11480390

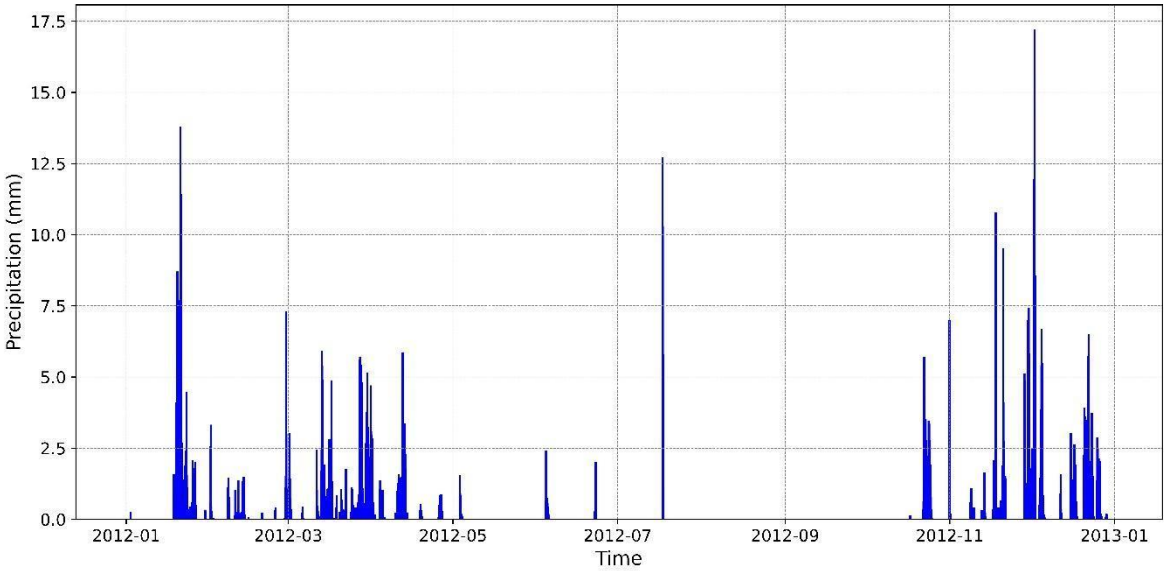


Figure 9. Precipitation over the 11480390 CAMEL basin during test period (one year)

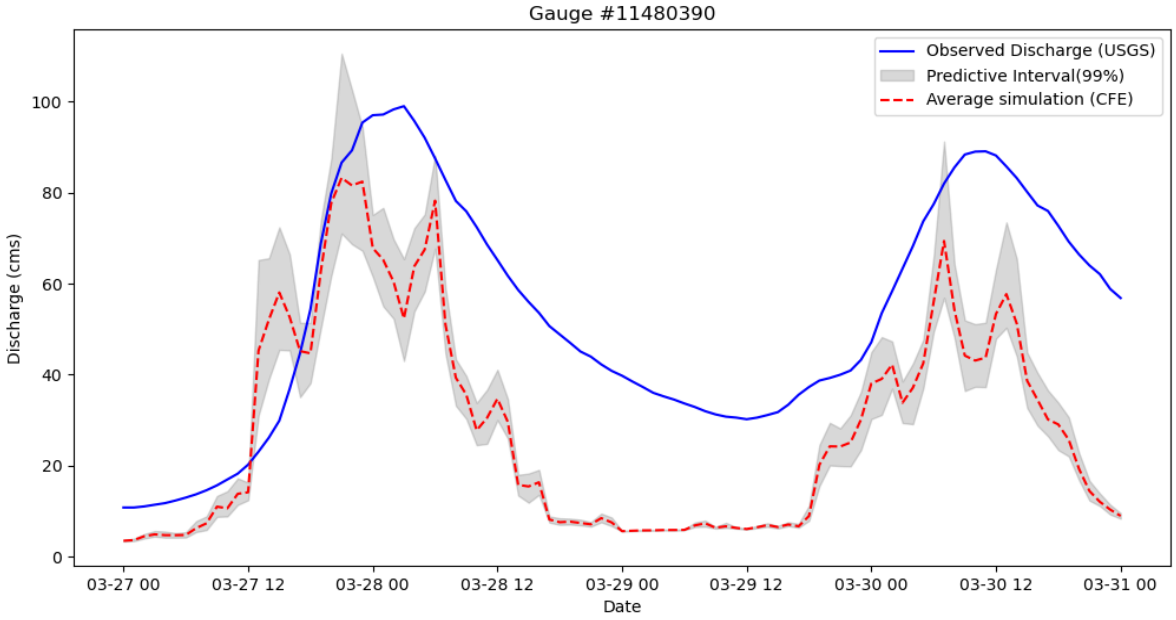


Figure 10. CFE model test and predictive interval for two events

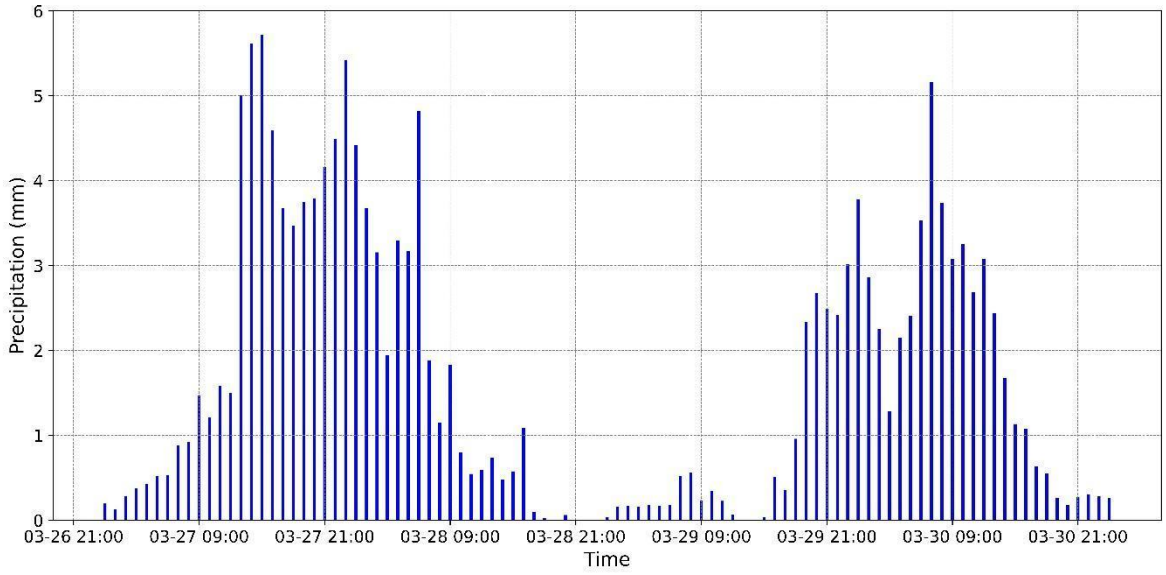


Figure 11. Precipitation over the 11480390 CAMEL basin during two events

6. Conclusion

We quantified the uncertainty of the forcing data (PET and precipitation) for the studied basin. We determined that the optimal perturbation factor is 0.15 and the ensemble size is 25. The results indicated that the predictive interval is reliable during peak events. However, the probabilistic performance metric (NRR) score of 1.36 suggests that the predictive interval is too narrow. This highlights the need to consider additional sources of uncertainty, such as parameters, model structure, initial conditions, and calibration parameters. For future studies, the use of various hydrologic models and the inclusion of different sources of uncertainty could help generate a wider predictive interval, closer to an NRR of 1.

References

1. Henrichsen, A.H., Hamidi, S.M.E., Sandquist, A., Zhang, H., Ferreira, C. and Mandli, K.T., 2022, December. Coupling coastal and hydrologic models through the first coastal Basic Model Interface in the Next Generation National Water Model Framework in low gradient coastal regions of Galveston Bay, Texas, USA. In AGU Fall Meeting Abstracts (Vol. 2022, pp. H53B-05).

2. Frazier, N., Tubbs, C., Johnson, D., Mattern, D., Ogden, F., Halgren, J.S., Williamson, M., Bartel, R., Cui, S., Wlostowsk, A. and Flowers, T., 2021, December. Joint Computational Domain Partitioning of Rainfall-Runoff Models with Hydrologic (Kinematic) Streamflow Routing Models. In AGU Fall Meeting Abstracts (Vol. 2021, pp. H54G-01).

3. Cunha, L., Jennings, K., Wood, A., Mizukami, N., Ogden, F., Feng, X., Liu, Y., Peckham, S., Garrett, J., Frame, J., Cui, S., Johnson, D., Mattern, D., Johnson, J., Frazier, N., Bartel, R., Avant, B., McDaniel, R., & Williamson, M. "Next Generation National Water Model: Strategy and Preliminary Performance of Initial Model Formulations." In AGU Fall Meeting Abstracts, vol. 2021, pp. H54G-04. 2021

4. Ogden, F.L. Justification of and Parameter Estimation for a Conceptual Functional Equivalent (CFE) Formulation of the NOAA-NWS National Water Model (version 2.1) 2021.

5. Foroumandi, E., Oh, Jeil., Ghaneei, P., Timilsina, S. Data Assimilation for Improving Forecast Accuracy and Streamflow Prediction in Ungauged Basins. National Water Center Innovators Program Summer Institute Report. 2023.

6. Wolkeba T.F., Ekpetere, O.K., Abualqumboz S.M., Butler, J.Z. Data assimilation of USGS streamflow in the Nextgen Framework. National Water Center Innovators Program Summer Institute Report. 2022.

7. Bolotin, L., Haces-Garcia, F., Liao, M., Liu, Q.,. Automated decision support for model selection in the Nextgen National Water Model. "National Water Center Innovators Program Summer Institute Report. 2023.

8. Johnson, J.M., Kindl da Cunha, L.L., Jennings, K.S., Frazier, N.J., Gibbs, R., Flowers, T. and Ogden, F.L., 2022, December. The NOAA Next Generation Water Resources Modeling Framework Hydrofabric: Version 1.0. In AGU Fall Meeting Abstracts (Vol. 2022, pp. H52B-04).

9. Johnson, J.M., Eyelade, D., Singh-Mohudpur, J., Rad, A.M., Coll, J., Spies, R. and Yeghiazarian, L., 2024. Enhancing Synthetic Rating Curve Development Through Empirical Roughness Built for Hydrofabric Datasets. Authorea Preprints.

10. Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J. and Hochreiter, S., 2021. Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network. Hydrology and Earth System Sciences, 25(4), pp.2045-2062.

11. Johnson, J.M., Fang, S., Sankarasubramanian, A., Rad, A.M., Kindl da Cunha, L., Jennings, K.S., Clarke, K.C., Mazrooei, A. and Yeghiazarian, L., 2023. Comprehensive analysis of the NOAA National Water Model: A call for heterogeneous formulations and diagnostic model selection. Journal of Geophysical Research: Atmospheres, 128(24), p.e2023JD038534.

12. Frame, J.M., Ogden, F.L., Peckham, S.D., Garrett, J., Frazier, N.J., McDaniel, R., Avant, B., Blodgett, D., Clark, E., Cosgrove, B. and Cui, S., 2022. In preparation as a section in an upcoming paper on the Next Generation US National Water Model. DEEP LEARNING FOR OPERATIONAL STREAMFLOW FORECASTS, p.131.

13. Ogden, Fred, et al. "The Next Generation Water Resources Modeling Framework: Open Source, Standards Based, Community Accessible, Model Interoperability for Large Scale Water Prediction." AGU Fall Meeting Abstracts. Vol. 2021. 2021.

14. Araki, R., Bhuiyan, S.A., Bindas T., Rapp J. On numerical methods and differentiable modeling for soil process representations in the NextGen Framework for arid regions. National Water Center Innovators Program Summer Institute Report. 2023.

15. Tolson, B.A. and Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research, 43(1).

16. Arsenault, R., Poulin, A., Côté, P. and Brissette, F., 2014. Comparison of stochastic optimization algorithms in hydrological model calibration. Journal of Hydrologic Engineering, 19(7), pp.1374-1384.

Supplementary Materials

A1. https://github.com/NWC-CUAHSI-Summer-Institute/cfe_py

A2. https://github.com/NOAA-OWP/ngen-cal

A3. https://github.com/NOAA-OWP/ngen

A4. https://github.com/NOAA-OWP/cfe

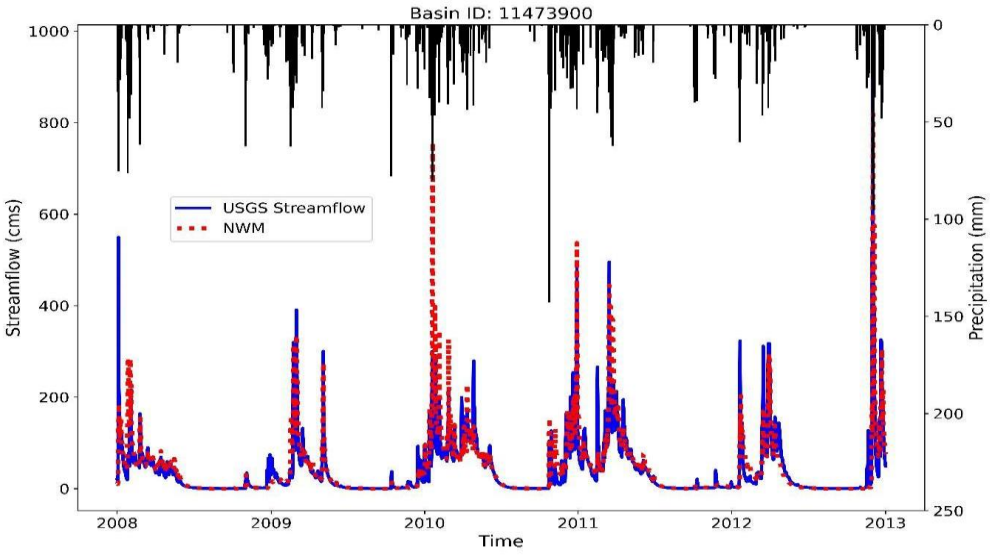


Figure 3. NWM result and USGS observations of the basin ID-11473900

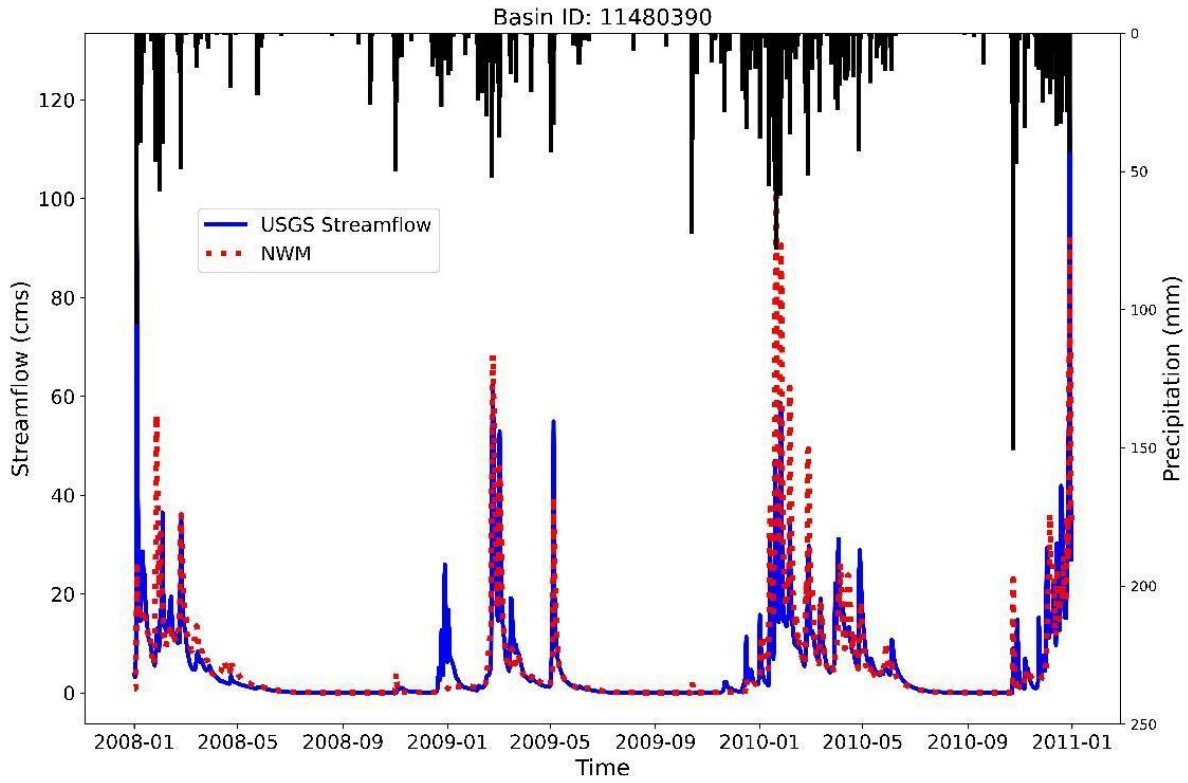


Figure 4. NWM result and USGS observations of the basin ID-11480390

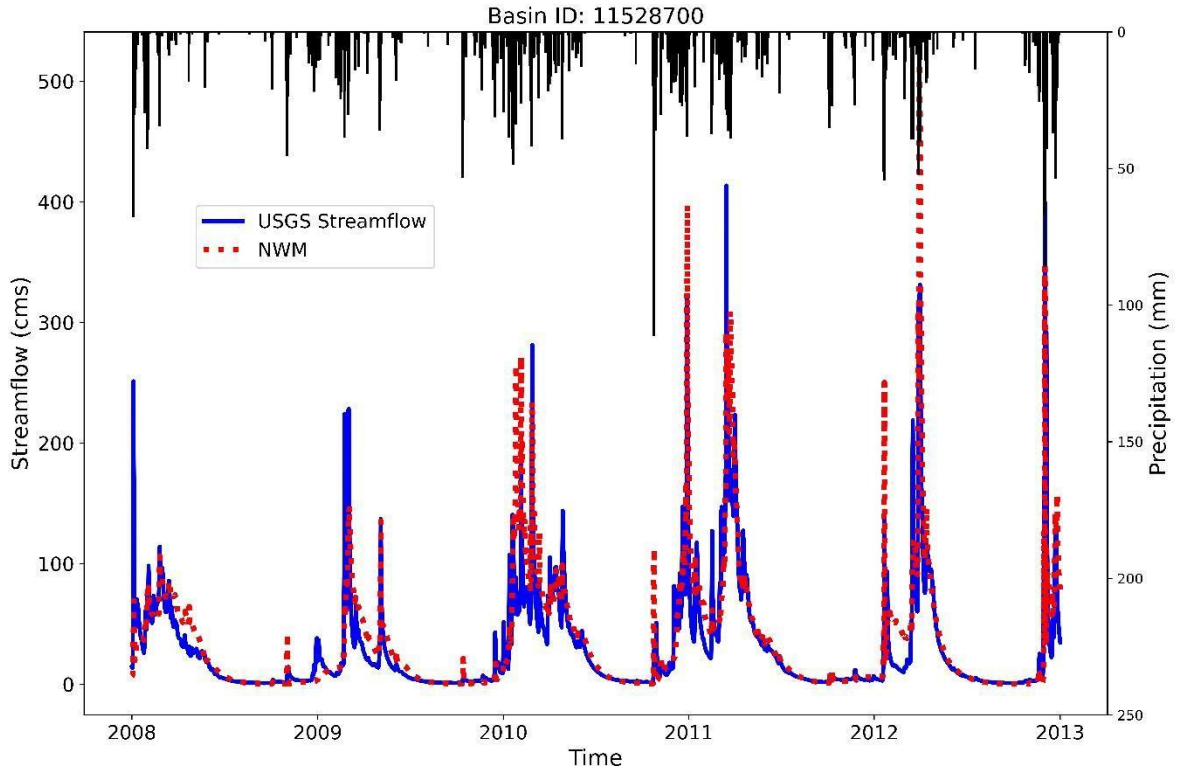


Figure 5. NWM result and USGS observations of the basin ID-111528700

Chapter 4
Enhancing Remote Sensing Flood Inundation Mapping by Leveraging a Terrain-Based Model

Jack Edwards[1], Francisco Gomez[2], and Kim Son Do[3]

[1]University of Kansas; *j375e293@ku.edu*

[2]University of Alabama; *fjgomez1@crimson.ua.edu*

[3]University of Virginia; *pgw5jd@virginia.edu*

Academic Advisors: Xingong Li, *University of Kansas*; Hamid Moradkhani, *University of Alabama*; Venkataraman Lakshmi, *University of Virginia*

Summer Institute Theme Advisors: Xingong Li, *University of Kansas*, lixi@ku.edu and Sagy Cohen, *University of Alabama*, sagy.cohen@ua.edu

Abstract: Flood mapping is essential for effective disaster management, requiring rapid and accurate delineation of flood extents. This study explores a hybrid approach combining remote sensing data and terrain-based models to generate accurate flood inundation maps. Remote sensing, especially using satellite imagery, provides data for large spatial extents but faces challenges like cloud and canopy cover. Terrain-based models, such as FLDPLN (pronounced floodplain) and Height Above Nearest Drainage (HAND), provide flood inundation mapping by leveraging information from Digital Elevation Models (DEM) and water surface elevation to predict flood inundation extent and depth. This research integrated remote sensing imagery with a terrain-based model to enhance flood mapping accuracy and efficiency. A remote sensing analysis procedure product was developed using Google Earth Engine to extract flood edge pixels from satellite imagery. The edge pixels were then used to estimate stream flood stage which serves as input to the FLDPLN model to estimate water depth and generate flood inundation maps. The proposed methodology (FLDSensing) was applied over the Verdigris River to map a flood event experienced in late May 2019. FLDSensing results show an accurate match of flood inundation map extension compared to a HEC-RAS 2D benchmark with performance metrics of (CSI, F-1 Score, PBIAS) and vertical deviations between (-1m - +2.5m) along the hydraulic profile. Integrating remote sensing data with the FLDPLN model addresses issues using remote sensing imagery, providing a comprehensive flood mapping solution.

---

1. Motivation

Rapid and accurate methods to delineate flood extent are crucial for decision-makers to provide effective measures to preserve human lives and identify areas prone to infrastructural damage. Using satellite imagery for flood mapping is widespread for this purpose, considering that it provides a valuable tool for accessing large scale and near-real-time information with high spatial and temporal resolutions [1], [2]. However, satellites are often not positioned over the area of interest at the right time, have a high cost (for commercial platforms), and do not capture the depth of the water, which often makes their use difficult in practical terms. Other challenges when using remotely sensed satellite imagery, especially in optical imagery, are the presence of clouds or canopy cover, which makes it difficult to identify the full extent of flooding. Synthetic

Aperture Radar (SAR) imagery is a useful tool in flood mapping, especially when dealing with atmospheric interference [3]. However, part of the drawbacks of SAR-based flood mapping is that it can be negatively affected by water-like and smooth surfaces, image speckling, dense vegetation, and urban development [4].

Flood inundation maps can also be obtained through models that seek to represent conceptually or numerically the extent and depth of water during a flood event. For example, 2D models, such as HEC-RAS 2D, allow the use of governing equations to simulate a flood hydrograph and identify potentially affected areas. While numerical models provide detailed representations of flow properties, they are computationally intensive. These models require equipment capable of handling large computational volumes to optimize simulation times, particularly for large domains or areas needing special detail [5]. Conceptual models like HAND and FLDPLN aim to provide a simplified representation of flood mapping for river segments. These models consider terrain and flow direction characteristics and are designed to be computationally efficient. However, they sacrifice accuracy because they cannot fully capture some hydraulic and hydrodynamic processes in the floodplain [6], [7].

The combination of remote sensing flood inundation maps and terrain-based model tools to determine the extent and depth of flood can be useful in complementing details that both lack effectiveness. For remote sensing, post-processing of imagery using terrain-based models can fill gaps inside the flooded domain in areas obstructed by clouds or canopies, increasing the extent accuracy and allows the estimation of water depths. Utilizing a near-real-time FLDPLN terrain-based model we can derive a synthetic stage value from remote sensing imagery as soon as they are available. This could improve model performance alongside improving remote sensing-based FIM, especially in areas where the stage is unknown, or the discharge-to-stage method is unreliable. This project will focus on improving remote sensing flood inundation mapping using the FLDPLN model.

2. Objectives and Scope

This study aims to develop a remote sensing image analysis procedure product in tandem with terrain-based models to generate accurate flood inundation maps from satellite imagery. The operational tools to replicate and apply the study's framework to other study areas (e.g., flood edge extraction app, remote sensing - FLDPLN integration framework) are also prepared for operational purposes. The project objectives include:

1. Develop a methodology for extracting flood edge pixels that includes detecting uncertainties due to land cover, cloud coverage, and terrain features from optical images during flood events.
2. Generation of a tool for delineating flood edge pixels using Google Earth Engine given user-defined parameters.
3. Estimation of the depth of flow at some flood source pixels along the stream from the flood edge pixels using the FSP-FPP inundation relationship of the FLDPLN model.
4. Develop a script that generates flood inundation maps by integrating remote sensing and FLDPLN and validates them against the ground truth, given the availability of data.

3. Previous Studies

At the operational level within the United States, the Office of Water Prediction (OWP) developed a Flood Inundation Mapping (FIM) forecasting framework based on HAND approach (referred to as the OWP HAND-FIM). The OWP HAND-FIM methodology is effective and quick in generating flood inundation map results and with a certain level of confidence considering its limitations and assumptions [7]. This method has been widely applied over multiple sets of domains to test the capabilities considering a variety of terrain features and drainage configurations [8], [9]. Terrain-based models like OWP HAND-FIM can generate results over large domains at a fraction of the time and computational resources that 2D hydrodynamic models consume, making these not feasible for operational CONUS-scale applications [7]. Similar to the OWP HAND-FIM, other terrain-based models have been developed, like AutoRoute [10], FLDPLN [6], GeoFlood [11], and more, with modifications of the methodology or other assumptions to capture flow patterns and estimate flood extent and depths.

The FLDPLN model was chosen for this study because of its ability to capture more complex flood inundation relationships, such as backfill and spillover processes [6]. More importantly, the FLDPLN model assigns each point within a floodplain a flood source pixel (FSP). The FSP is a stream pixel from where floodwaters originate that can inundate a floodplain pixel (FPP) at the shallowest depth to flood (DTF). FPP can be flooded from FSP through backfill and spillover processes noted above, and the model identifies the FPP-FSP flooding relationships and their associated DTF. This allows for interpolated depth of flood (DOF) values between gauged locations (i.e., stage) to create a custom flood extent map with depth during specific flood events [12].

This study builds off of similar research combining a remotely sensed image with the FLDPLN model to get an improved flood extent map [12]. That study utilized the flood edge pixels derived from an HEC-RAS 2D model output as synthetic gauge input for the FLDPLN model. The study found that combining a limited number of flood edge points with the FLDPLN model yielded promising accuracy. However, the accuracy had an upper limit, which varied based on the quantity of flood edge pixels available. The study also suggested utilizing land cover data to eliminate forested areas, which provided the basis for this project.

Compared to the proof-of-concept in [12], this study uses a remotely sensed flood inundation map as a source of the flood edge and HEC-RAS 2D model output as the benchmark for validation. Additionally, the assignment of stages at the river pixels (flood source pixel) is performed with a more rigorous algorithm instead of taking the mean of possible choices.

4. Methodology

The proposed methodology has two main steps: extracting flood edge pixels from remote sensing and generating flood inundation maps derived from FLDPLN. The following sections describe the study area, the steps for generating remote sensing plus FLDPLN flood inundation maps, and validation.

4.1 Study Area

The study area is located in southeast Kansas, over the watershed of the Verdigris River (Figure 1). In May 2019, a series of severe thunderstorms hit southeastern Kansas, which caused substantial flooding in the region, affecting cities such as Independence and Coffeyville, which sit along the Verdigris River. The study area river is a tributary of the Arkansas River and runs from central Kansas into Oklahoma. There are two USGS gauges in the study area, one in Independence Kansas (Station ID: 07070500, Latitude: 37.22, Longitude: -95.67) and one in Coffeyville Kansas (Station ID: 07170990, Latitude: 37.005, Longitude: -95.59), that provide a continuous record of gauge height. These gauges will be used later in HEC-RAS model calibration.

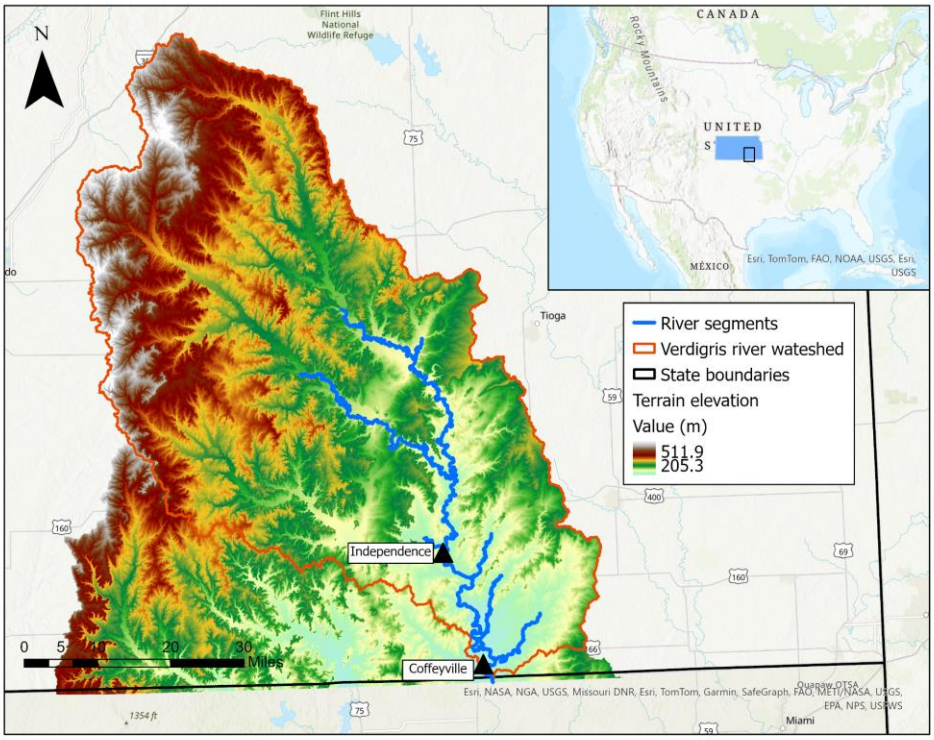


Figure 1. Study area location.

4.2 Derive Clean Flood Boundary Points

A clean flood edge pixel is a pixel that is the boundary of a flood, or a pixel that does not fall underneath clouds or border aerial-view-obscuring land cover such as trees. If a pixel falls underneath a cloud cover or borders an aerial-view-obscuring land cover, it is classified as an unclean edge or an edge pixel with low confidence on being a clean edge pixel. This "unclean" edge pixel likely has flood extent continuing underneath the land cover. If we select an unclean edge then an incorrect water surface elevation will be derived which will affect the FLDPLN-predicted FIM and depth map, so care is needed to determine accurate, clean edge pixels.

We developed a Google Earth Engine (GEE) application where the user can define the study area location and date to derive flood extent maps from Sentinel-2 imagery and select flood edge pixels with high confidence. User-defined values include study date, area of interest, and masking threshold values (Normalized Difference Water Index (NDWI), Normalized Difference Vegetation Index (NDVI), cloud masking, slope thresholding, small water body masking, and

landcover masking). The script can be accessed in the supplementary materials section of the report.

In this study, we applied a land cover mask for urban, forest, permanent water bodies, and cultivated crop pixels because these pixels contain more uncertainties for remote sensing based FIM (e.g., canopy interception may lead to underestimation of the full extent of flooding). We also applied a 1-pixel buffer around the masks and removed pixels with a slope value below 2.5 degrees and above 9.5 degrees.

4.3 Integration of Remote Sensing Flood Inundation Maps into FLDPLN

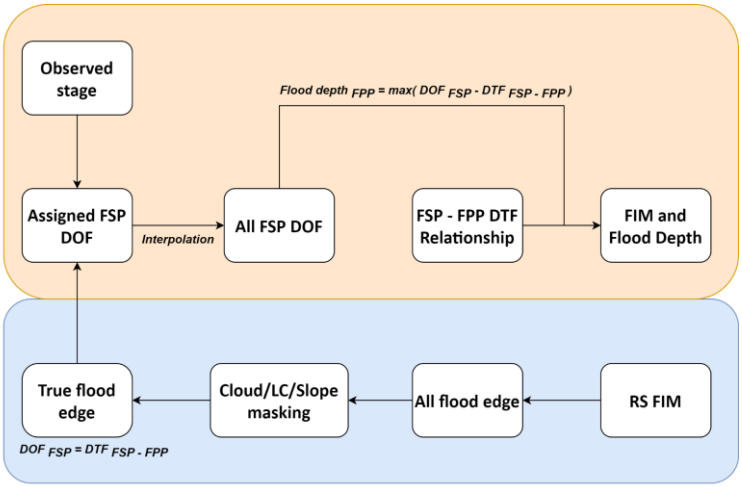


Figure 2. FLDPLN scheme in a range box, and proposed methodology in the blue box

FLDPLN uses observed stages from USGS or the National Water Model (NWM) results to assign and interpolate the stage (i.e., water level) of FSP locations along the study area. The stages are then compared to the associated depth to flood (DTF) between FSPs and clean edge pixels (FPP), and a pixel is considered to be flooded if any stage of associated FSPs is higher than the DTF relationship between them. This study assigns water levels at FSPs are assigned using the clean flood edge pixels' DTF relationship. At the clean flood edge (i.e., the boundary between water and non-water pixel), the flood depth is ideally 0. The flood depth at the flooded pixel is calculated as the difference between the associated stage and DTF relationship. Therefore, the stage of the FSPs can be assigned to equal to the DTF value between the FSPs and clean flood edge pixels (Figure 2).

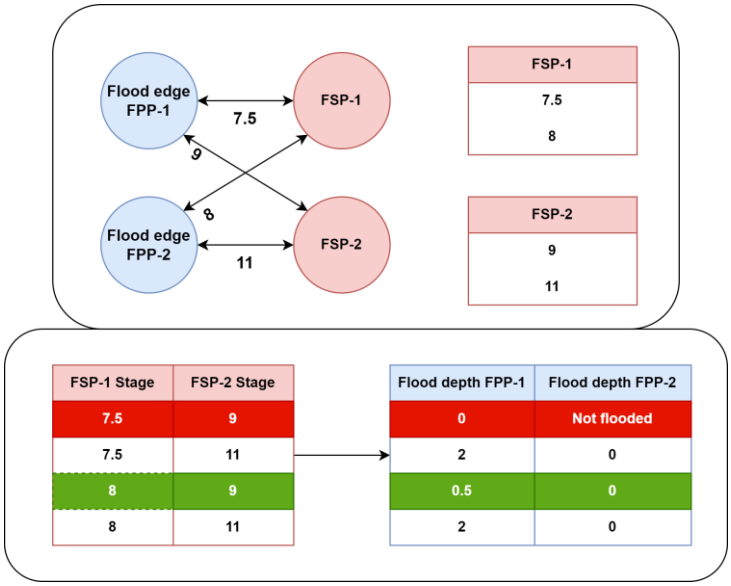


Figure 3. Example of FSPs' DOF assignment using clean edge pixel FPP methodology

We assigned the FSPs' stages based on clean edge pixels' DTF relationship by iterating through all possible combinations of stage assignment and selecting the choice that results in the flooding with the lowest sum of flood depth (i.e., flood depth simulation closest to 0) at all of the clean edge pixels. Figure 3 illustrates an example of how the FSPs' stages are assigned for the scenario where there are two FSPs and two associated clean flood edge pixels. Compared to assigning the mean value of all possible DTFs to each FSP [12], optimizing the assignment of the FSPs stage through all possible combinations requires longer computational power and time but can ensure that all clean flood edges are flooded. For example, with the same clean edge map with 122 clean edge pixels, taking the median DTFs leaves 22 edge pixels unflooded.

Because FLDPLN stores all possible flooding pathways via DTF relationships, some flood edge pixels at the outermost layer of the flooded domain can contain very high DTF values (i.e., the pixel is only flooded at extreme flood events). While clean flood edges are carefully selected by applying various masks, including land cover, slope, and cloud probability conditions, flood edge pixels with unreasonably high DTF may still remain and result in unreasonably high FSP stage assignment. To avoid overestimating FSP stages, we applied a threshold that removes flood edges with any DTF relationship higher than 14 meters. The threshold was chosen based on USGS stage information during the flood event (the maximum recorded stage was ~ 12.5 meters) and the fact that FLDPLN libraries were generated with the maximum flood depth of 15 meters. It is uncertain that this threshold can be applied in all cases. Additionally, we only used assigned FSPs on the mainstems for interpolation for the remaining FSPs as FSPs in the tributaries (FLDPLN stream order > 1) lead to overestimation of stages on the mainstems. For the FIM of tributaries catchment, the process should run separately.

4.4 Ground Truthing

To verify the results from the proposed methodology, a HEC-RAS 2D (HR2D) model (version 6.5) was generated using Shallow Water Equations ELM for governing equations. The model is delimited between Independence and Coffeyville cities, which allows it to be constrained with United States Geological Survey (USGS) data of hydrograph and water surface elevation boundary conditions, respectively. Manning roughness coefficients were assigned based on the

National Land Cover Dataset (NLCD) 2019 land cover data and reclassified into 5 categories (Table S1 in supplementary material).

4.5 Evaluation Metrics

For flood extent evaluation, we used the Critical Success Index (CSI) and F-1 score to compare against the HEC-RAS benchmark results. CSI, also known as the Threat Score, measures the accuracy of the model's positive predictions. On the other hand, the F-1 score is the harmonic mean of precision and recall, providing a balance between the two. The F-1 score is extremely useful when dealing with an imbalanced dataset (e.g., there are usually more non-water than water pixels). The two metrics are calculated using a confusion matrix of true positives (TP) (i.e., the model correctly labels water/flooded pixel), true negatives (TN) (i.e., the model correctly labels non-flooded pixel), false positive (FP) (i.e., model incorrectly label a dry pixel as flooded), and false negative (FP) (i.e., model misses a flooded pixel). For flood depth evaluation, we used PBIAS to quantify the average tendency of predicted values to differ from observed values (flood depth in this case), expressed as a percentage, with positive values indicating overestimation and negative values indicating underestimation. A table of equations used can be found in the supplementary material section as Figure S1.

5. Results

5.1 Remote Sensing Clean Flood Edges

Sentinel 2 optical imagery of the study area, captured on May 27th (Fig. 4a), shows the flood extent close to the peak discharge of the event with relatively few clouds. The shown Sentinel 2 image is used as an input into the FLDPLN model. For flood edge pixel extraction, the methodology discussed in Section 4.2 was applied. The Sentinel 2 image processing yielded a total of 227 clean edge pixels for the study area. These pixels are then reprojected to the same coordinate system used by FLDPLN (UTM 14N) and identified as FPP through a Python subroutine. Figure 4b shows the initial FIM from Sentinel 2 image after NDWI reclassification, and the third image shows flood edge pixels extracted using the GEE application.

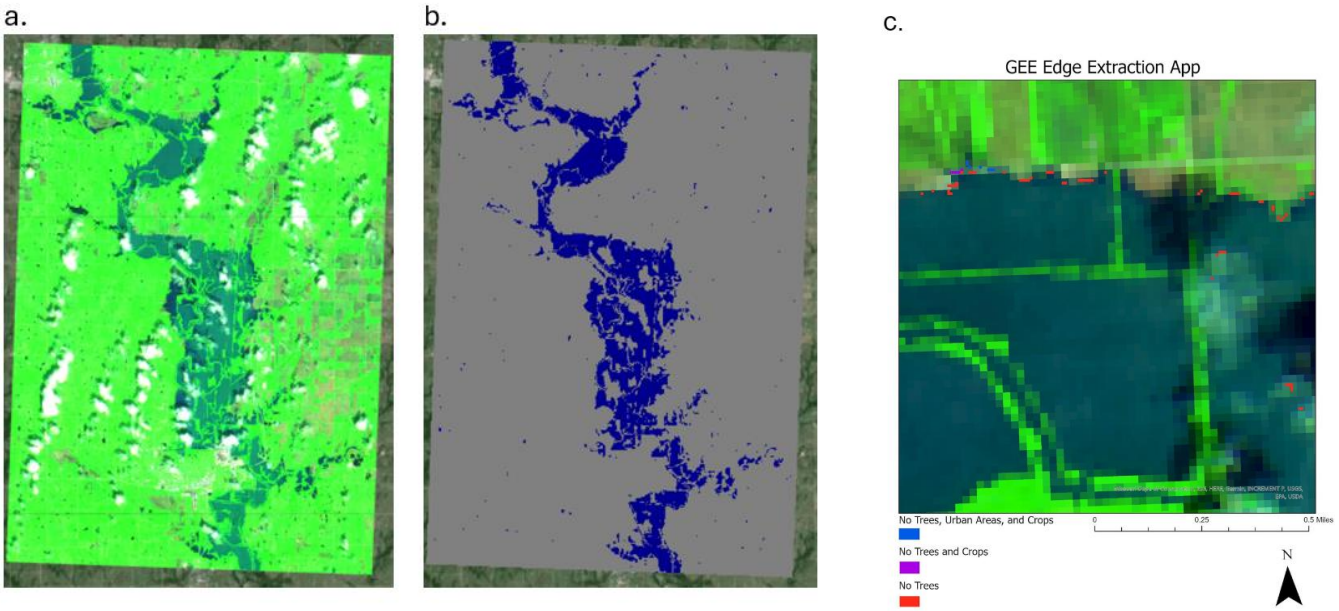


Figure 4. (a) Sentinel-2 image over the study area. Captured on 05/27/2019; (b) Flood extent map derived from Sentinel-2 image using NDWI threshold; (c) and examples of how masking out different land cover will affect the amount of clean edge pixels.

5.2 Remote sensing - FLDPLN combination FIM

Figure 5 shows Sentinel-2 and FLDSensing FIM predictions compared to the HEC-RAS 2D benchmark. Due to cloud and canopy interception, there are multiple gaps within the flooded domain of Sentinel-2 FIM (Figure 5a). Using FLDSensing methodology, the flood extent gaps/underestimations were filled in using terrain information and the DTF relationship between FSP and clean flood edge. However, simply using all clean flood edges led to a substantial overestimation of flood extent (Figure 5b). With the exclusion of flood edges with maximum DTFs > 14 meters and tributaries FSPs assignment (Section 4.3), FLDSensing result achieved much better flood extent than using solely remote sensing (Sentinel-2), improving CSI and F-1 score from 0.59 to 0.80 and from 0.74 to 0.89, respectively. The high-performance metrics underscore the effectiveness of this methodology and validate the accuracy of combining remote sensing flood edge with FLDPLN model in representing flood extents. However, it is important to note that upper and lower regions are categorized as false positives, these discrepancies are due to the HR2D domain definition using USGS gauges, which does not match exactly with the comparison area. Despite these False Positives, the overall high CSI and F1 scores indicate that the model and methodology used are highly effective for the majority of the analyzed segment.

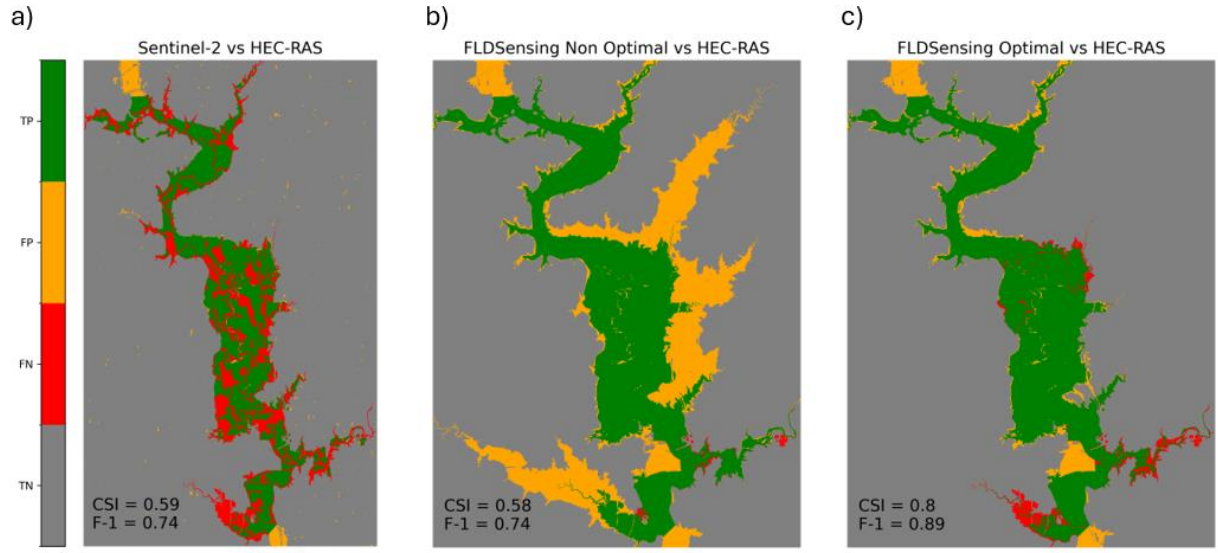


Figure 5. Performance of flood extent from Sentinel-2 (a) and FLDSensing with (c) and without (b) thresholding compared to HEC-RAS 2D benchmark.

5.3 Flood Depth Comparison

Figure 6 shows the flood depth differences of FLDPLN generic run (using USGS gauges) (Fig 6a) and FLDSensing (Fig 6b) versus HEC-RAS benchmark results. For both FLDPLN and FLDSensing, flood depths were overestimated compared to the benchmark, especially in the northern region of Coffeyville city. Additionally, discrepancies in the river channels (i.e., misalignment in the river profile) between FLDPLN and HEC-RAS could be observed at the high flood depth difference running across the floodplain. The decrease in PBIAS from base FLDPLN to FLDSensing (i.e., FLDSensing simulates lower flood depth) could be due to the fact that FLDPLN was run with the maximum stages obtained from USGS on May 29th. On the other hand, FLDSensing relied on Sentinel-2 imagery, which might not have been captured at the absolute peak of the flood on that day. As both FLDPLN and FLDSensing were overestimating the flood depth, it is highly possible that the uncertainties and assumptions in the interpolation of FSPs' stage across all segments were the main culprit for the high flood depth.

The hydraulic profile shows longitudinal dynamics in water surface elevation along the river main channel (Figure 6c). The HR2D (green) and FLDSensing (blue) results were extracted along 56 kilometers between Coffeyville and Independence. The results show that the developed methodology overestimates water surface elevation by up to 2.5 meters in some sectors. The values are consistent with the spatial results presented in Figure 5, where the areas with overestimates are identified.

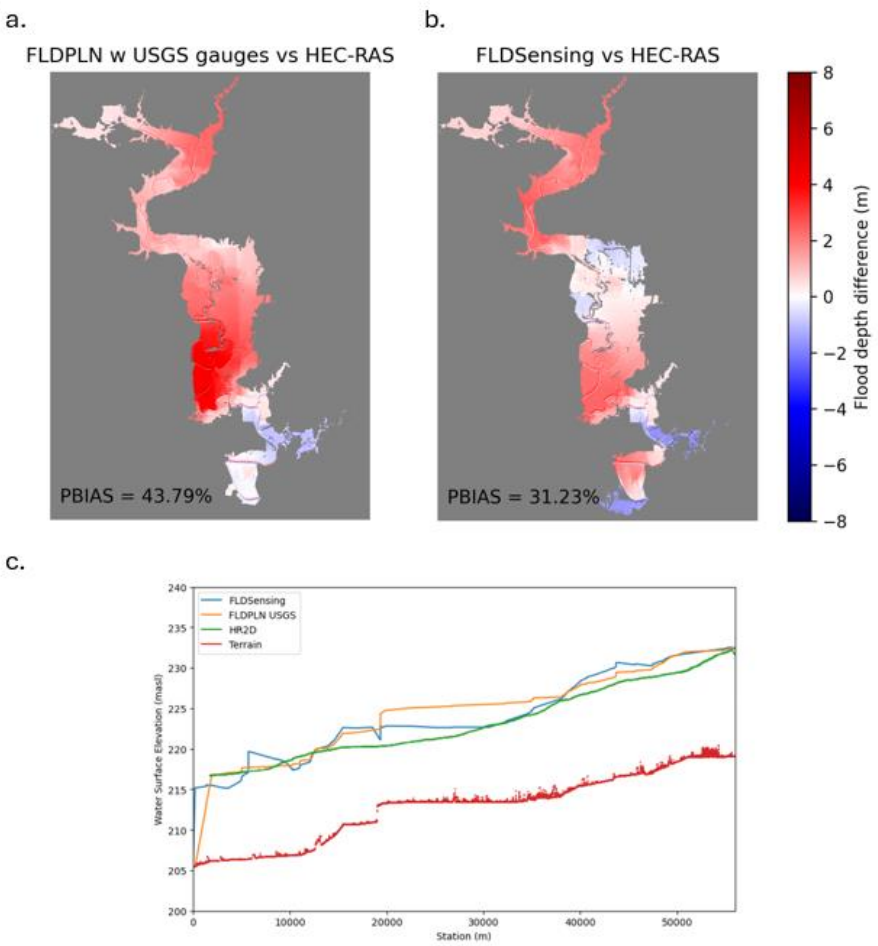


Figure 6. Flood depth difference compared to HEC-RAS benchmark of (a) FLDPLN using USGS gauges (b) FLDSensing result (c) and Hydraulic profile comparison between FLDPLN results against HR2D benchmark.

6. Conclusion

The coupling of remote sensing and terrain-based models improves the prediction/extraction of the extent and depth of floodwater. The FLDSensing tool approximates the extent and depth map from pixels at the edge of the flood captured by the satellite image. These pixels, together with the pre-calculated relationships between the FLDPLN model's FSP and FPP, determine the levels of the water layer reached at the stream and flood total extension.

The application of the methodology proposed in the Verdigris River flooding in 2019, shows that it is feasible to greatly improve a remote sensing generated flood map using the FLDSensing approach. The performance metrics indicate satisfactory results, with both the CSI and F1 score exceeding 0.8 for the analyzed segment. This demonstrates the effectiveness and reliability of the model in the specific context studied in order to capture a general flood extent and also providing water depths along the study area. The methodology presented exhibits an overestimation on flooding the extent and depth of water in some specific areas. This occurs mainly due to the incorporation of misclassified pixels as a clean edge. Part of the proposed future work includes detailed research on the categorization and number of pixels to be incorporated. Including a way to automatically mask out pixels with an unrealistic DTF value for

the study area. Future work could include the testing of different model DOF interpolation methods along FSPs and river segments to enhance model accuracy.

Acknowledgements

Dr. Jude Kastens- University of Kansas, Kansas Biological Survey

Dr. James Halgren- University of Alabama, Alabama Water Institute

David Weiss- University of Kansas

OWP, NWC, CIROH, AWI, CUAHSI, and UA staff

Supplementary Materials:

The Google Earth Engine app developed can be accessed through:

https://code.earthengine.google.com/20cef9951fa58dc223c81b76cc71dc16

Github repository with notebooks can be accessed through: https://github.com/NWC-CUAHSI-Summer-Institute/FLDSensing

Supplementary materials that could not fit in the main report can be found here:

https://docs.google.com/document/d/1H7A7AQLKm_CwxHapBhScaaYZwBpPjQn9LY57259pUdk/edit?usp=sharing

Metric	Formula	Target Score	Answers the Question*
Critical Success Index (CSI)	$\frac{True\ Positives}{True\ Positives + False\ Negatives + False\ Positives}$	1	How well did the forecast correspond to the observed?
Probability of Detection (POD)	$\frac{True\ Positives}{True\ Positives + False\ Negatives}$	1	What fraction of the observed was correctly forecast?
False Alarm Ratio (FAR)	$\frac{False\ Positives}{True\ Positives + False\ Positives}$	0	What fraction of the predicted actually did not occur (i.e., were false alarms)?
*https://www.cawcr.gov.au/projects/verification/#Methods_for_dichotomous_forecasts			

Figure S1. Equations for Evaluation Metrics, picture retrieved from OWP-Inundation Mapping Wiki https://github.com/NOAA-OWP/inundation-mapping/wiki/6.-Evaluating-HAND-Performance

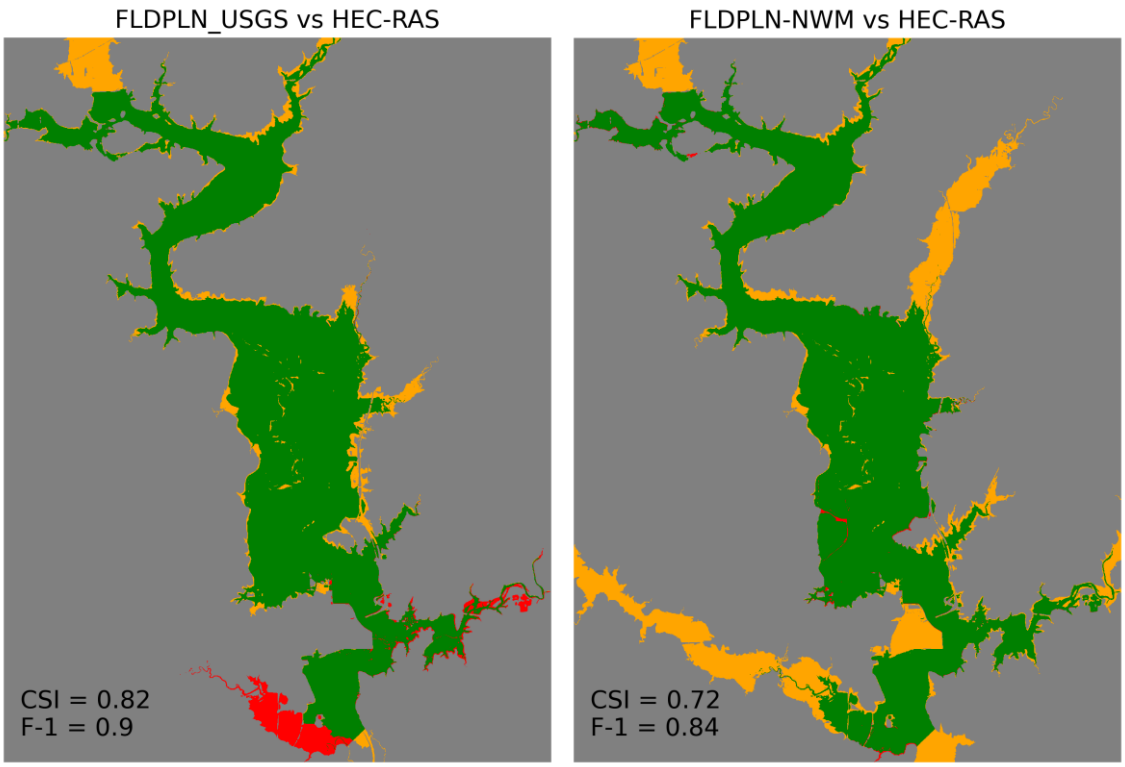


Figure S2. Comparing RS FIM and FLDPLN Gauge and NWM Results with our benchmark HEC-RAS model output.

Table S1. Manning roughness coefficients assigned by land cover.

Land cover	Manning roughness coefficient
Open water	0.035
Developed areas	0.018
Barren land	0.04
Forests/Wetlands	0.1
Cultivated crops	0.05

References

[1] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, Dec. 2017, doi: 10.1016/j.rse.2017.06.031.

[2] E. Hamidi, B. G. Peter, D. F. Munoz, H. Moftakhari, and H. Moradkhani, "Fast Flood Extent Monitoring With SAR Change Detection Using Google Earth Engine," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–19, 2023, doi: 10.1109/TGRS.2023.3240097.

[3] X. Shen, D. Wang, K. Mao, E. Anagnostou, and Y. Hong, "Inundation Extent Mapping by Synthetic Aperture Radar: A Review," *Remote Sens.*, vol. 11, no. 7, p. 879, Apr. 2019, doi: 10.3390/rs11070879.

[4] A. Betterle and P. Salamon, "Water depth estimate and flood extent enhancement for satellite-based inundation maps," Feb. 06, 2024. doi: 10.5194/nhess-2024-22.

[5] P. D. Bates, "Flood Inundation Prediction," *Annu. Rev. Fluid Mech.*, vol. 54, no. 1, pp. 287–315, Jan. 2022, doi: 10.1146/annurev-fluid-030121-113138.

[6] J. Kastens, "Some New Developments on Two Separate Topics: Statistical Cross Validation and Floodplain Mapping," University of Kansas, 2008.

[7] 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 Resour. Res.*, vol. 59, no. 5, p. e2022WR032039, May 2023, doi: 10.1029/2022WR032039.

[8] A. D. Nobre *et al.*, "Height Above the Nearest Drainage – a hydrologically relevant new terrain model," *J. Hydrol.*, vol. 404, no. 1–2, pp. 13–29, Jun. 2011, doi: 10.1016/j.jhydrol.2011.03.051.

[9] 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.

[10] M. L. Follum, R. Vera, A. A. Tavakoly, and J. L. Gutenson, "Improved accuracy and efficiency of flood inundation mapping of low-, medium-, and high-flow events using the AutoRoute model," *Nat. Hazards Earth Syst. Sci.*, vol. 20, no. 2, pp. 625–641, Feb. 2020, doi: 10.5194/nhess-20-625-2020.

[11] 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," *Water Resour. Res.*, vol. 54, no. 12, Dec. 2018, doi: 10.1029/2018WR023457.

[12] K. Dobbs, "Toward Rapid Flood Mapping Using Modeled Inundation Libraries," University of Kansas, 2017.

Chapter 5
Rapid Flood Inundation Mapping for Catastrophic Floods Due to Dam Failures

Shivakumar Balachandaran[1], Ayman Mokhtar Nemnem[2], Reza Saleh Alipour[3] and Parvaneh Nikrou[4]

[1]The University of South Carolina; sb106@email.sc.edu

[2]The University of South Carolina; amokhtar@email.sc.edu

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

[4]The University of Alabama; pnikrou@crimson.ua.edu

Academic Advisors: Erfan Goharian, University of South Carolina, goharian@cec.sc.edu ; Jasim Imran, University of South Carolina, imran@sc.edu; Steven J. Burian, The University of Alabama, sburian@ua.edu; Sagy Cohen, University of Alabama.

Summer Institute Theme Advisors: Sagy Cohen, University of Alabama, sagy.cohen@ua.edu; Xingong Li, Kansas University, lixi@ku.edu

Abstract:
Dam operations and catastrophic failures can severely impact lives and properties. While conventional hydrodynamic models can generate flood inundation maps using numerical methods to solve shallow water equations, these models are complex and computationally intensive. This study uses two terrain-based models, OWP HAND-FIM and FLDPLN, to generate the inundation maps for near real-time operational applications. The dam-break flood hydrograph is empirically calculated as a function of the normal reservoir pool level. The Fall River Dam in Kansas is used as a case study. The peak discharges are attenuated along downstream reaches using a simple analytical solution derived from the diffusive wave equation. These are used to generate flood inundation maps using the two terrain-based models. The maps are evaluated using a benchmark HEC-RAS model with quantitative metrics and flood impacts are analyzed. The results show that terrain-based models can effectively generate flood inundation maps, with the OWP HAND-FIM model achieving an accuracy of 92% and the FLDPLN model with an accuracy of 88%. The computation runtime for the approach is very short, just a few minutes, compared to 17 hours for HEC-RAS. The proposed approach of mapping dam-break inundation can provide near real-time flood impact assessment and scalability for dams across the USA.

1. Motivation

Dam operations can lead to a sudden release of water causing damage to property and life losses downstream of the dam (Devi et al., 2022, Ge et al., 2022). In unforeseen situations, these dams can fail and cause catastrophic flooding. In this study, rapid Flood Inundation Maps (FIMs) are generated using terrain-based models downstream of the dam. Rapid FIM can be useful in developing quick flood inundation maps for dam operations causing significant flash floods or during catastrophic failure of dams due to overtopping or piping. This study uses a simple analytical model derived from the diffusive wave equation with empirical closures for determining the peak discharge (Qpeak) for downstream reaches in the modeled domain. This

can help in developing inundation maps for different scenarios related to dam operations in a short time. Also, these terrain-based models give the advantage of scalability and can be used for the entire CONUS. The low-complex terrain-based models can generate FIMs in a few minutes while the hydrodynamic models take several hours.

2. Objectives and Scope

The objective of this study is to investigate the ability of terrain-based models to accurately generate FIMs for flooding caused by dam operations and failures. The Qpeak are determined using a simple analytical solver and are provided as inputs to the OWP HAND-FIM (Office of Water Prediction (OWP) Height Above Nearest Drainage (HAND) and FLDPLN models. The analytical Qpeak solver allows scalability, as it can be solved for most large dams in the US. The FIMs generated by these terrain-based models are compared with the U.S. Army Corps of Engineers, Hydrologic Engineering Center's River Analysis System (HEC-RAS) model. Qpeak simulated by the HEC-RAS simulation is also used to allow direct comparison between the models. Finally, a quantitative analysis of these maps is studied to evaluate the accuracy of terrain-based models for flash floods due to dams.

3. Previous Studies

The FIM and dam failures can be generated broadly in two ways, hydrodynamic models, and data-driven models. Hydrodynamic models are physically based models that solve mass and momentum equations. HEC-RAS has been widely used in performing hydrodynamic and dam break simulations (Xiong, 2011). In recent times, data-driven models have been used for developing dam breach FIMs. Pianforini et al., (2024) developed a surrogate machine learning model using video frame prediction techniques for near real-time inundation maps using the results of hydrodynamic models for dam break flows. The ability of predictions through such models might be limited to the cases used for training and is challenging to scale up due to the lack of dam breach FIM observations. The present work proposes a new methodology where the analytical model for determining the peak discharge is combined with the terrain-based models.

4. Methodology

In this section, we first introduce our workflow which involves a five-step process (Figure 1), followed by a description of the case study, the analytical model for diffusive wave routing, dam break discharge estimation, OWP HAND-FIM, FLDPLN, and HEC-RAS models.

A hydro-conditioned Digital Elevation Model (DEM), retrieved from the Kansas Lidar project by the Kansas Biological Survey Center, with a 10-meter resolution was employed for the FLDPLN and HEC-RAS models, while a separate 10-meter DEM from USGS (3DEP) was used by default in the operational pipeline of the OWP HAND-FIM model. Both hydro-conditioned DEMs ensured continuous water flow across the terrain, including through culverts and bridges.

In the model setup stage, the HEC-RAS model is used to simulate two dam break scenarios which will be used as a benchmark for the terrain-based model evaluations.

The peak discharge attenuation is determined using an analytical solver. The Qpeak for the two scenarios—overtopping and normal pool level (sunny day failure)—is determined through breach simulation and empirical methods, respectively. The performance of the two terrain models is assessed through both quantitative and flood impact measures.

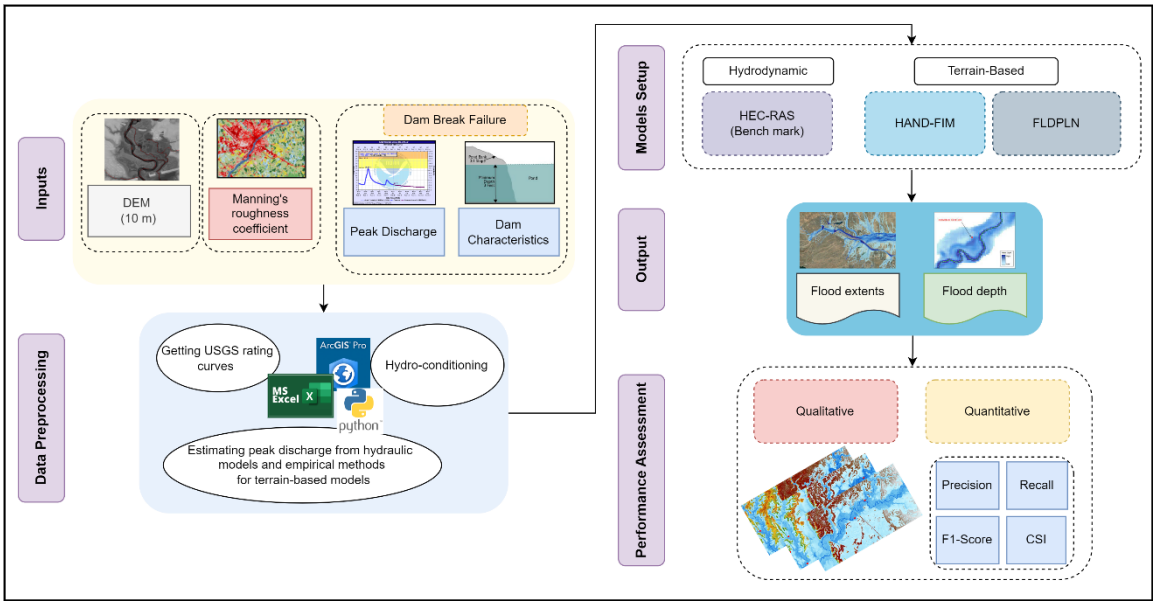


Figure 1: The schematic diagram of the workflow of the study.

4.1. Case Study

The Fall River Lake Dam (latitude 37.6098° N and longitude 96.0575° W), situated in Greenwood County, Kansas (Figure 2), is an earthen gravity dam with a height of 28.6 meters. According to the National Inventory of Dams, the dam has a maximum normal operation storage capacity of 316 million cubic meters. The watershed area drained by the dam spans approximately 1,515 square kilometers. This dam is classified as having a High Hazard Potential, indicating that failure or misoperation will likely result in the loss of human life and significant property damage (U. S. Army Corps of Engineers, 2024). The classification is due to the potential downstream impact on populated areas and critical infrastructure. Figure 2 shows the location of the Fall River dam as well as the U.S. Geological Survey (USGS) gages and building footprints located in the vicinity of the Fall River, which are represented by red points.

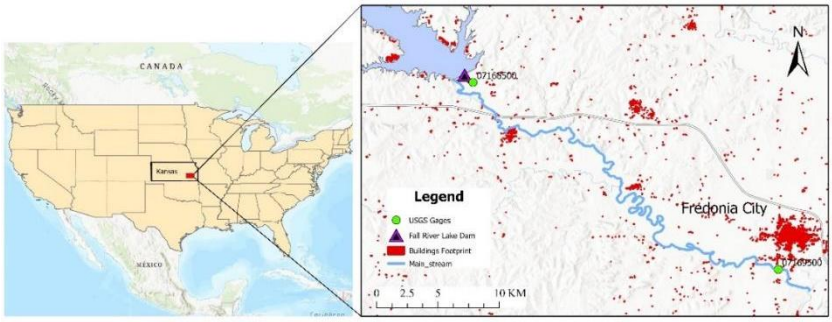
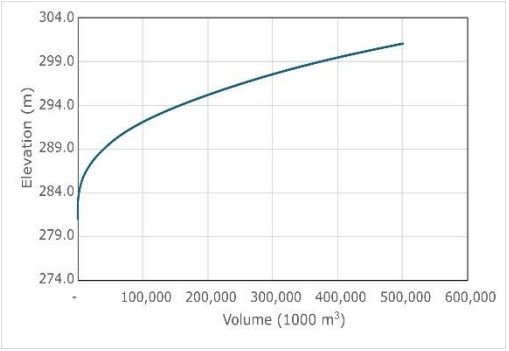


Figure 2: The Fall River Dam location along with the downstream river reach, USGS gages, and building footprints in the vicinity of the Fall River.

We calculated the Fall River Dam storage elevation curve (Figure 3) from the bathymetric survey of the Fall River Lake reservoir conducted by the Kansas Biological Survey (KBS) in 2010 (Kansas Biological Survey, 2011). The storage elevation curve is processed using the Storage Capacity geoprocessing tool under the ArcGIS Spatial Analyst package. The total reservoir capacity at the dam crest was determined to be 501 million cubic meters.



4.2. Qpeak Solver Using Diffusive Wave Equation

The attenuation of the peak discharge is determined using the following equations (Paiva & Lima, 2024).

$$\frac{Q(x)}{Q_0} = \frac{1}{[1+\varphi x]^{\frac{\beta}{3}}} \tag{1}$$

Where, φ is the peak attenuation factor and β is the flow rating exponent. The peak attenuation factor is determined using the following equation,

$$\varphi = \frac{3}{\beta}\frac{D_{h0}}{c_0^3}\left(\frac{w_t}{w}\right)^2 C'\frac{Q_0^2}{V^2} \tag{2}$$

The w_t is the width of the flood plain and w is the width of the channel.

C' is the relative curvature of the flood hydrograph at the peak considered constant as $5.78S^{-0.33}$ where S is the asymmetry parameter of the hydrograph. The hydraulic diffusivity D_{h0} is determined using the following equation:

$$D_{h0} = \frac{Q_0}{2wS_0}(1-\vartheta^2) \tag{3}$$

The Vedernikov number in the above Eq. 3 is determined using the following equation:

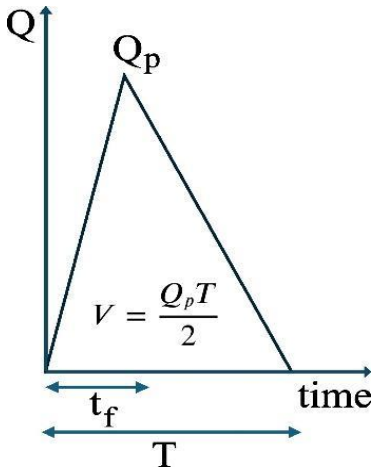
$$\vartheta = \frac{2}{5}\frac{c_0}{c_d} \tag{4}$$

Where c_0 is the flood wave celerity accounting for the floodplain storage and c_d is the dynamic wave celerity. The width of the river is determined as 75 m from the digital elevation model using HEC-RAS tools and the ratio of channel width to the floodplain width is determined as 10. The widths are generalized here to be a constant but in future work the floodplain widths can be determined from the hydrotable of OWP HAND FIM. The slope of the channel is 0.001 taken from the DEM longitudinal profile using the HEC-RAS model.

4.3. Estimating Dam Break Flow

The estimation of dam breach parameters—such as breach width, formation time, and peak flow—is important for simulating downstream flood wave routing. Several empirical methods have been developed to predict these parameters, including those by MacDonald & Langridge-Monopolis (1984), the Bureau of Reclamation (1988), Von Thun & Gillette (1990), and Froehlich (1995 & 2008).

Two scenarios were considered for estimating the dam break peak flow. In the first scenario, the breach process is simulated using HEC-RAS for the overtopping failure, and the resulting breach hydrograph is routed using the analytical equation for terrain-based models. In the second scenario, instead of simulating the breach, a synthetic breach hydrograph (Figure 4) is developed using Froehlich's (1995) empirical equation (Equation 5) as an initial condition for dam break simulation in HEC-RAS. The empirical equation is used to determine the peak dam breach flood as it is found to have the least uncertainty bandwidth among the methods for estimating the peak breach outflow Q_P (Wahl, 2004).



The dam breach formation time t_f, defined as the time taken from the breach's initial erosion to its final development, is calculated using Froehlich's (2008) empirical equation (Equation 6). Consequently, the synthetic hydrograph is developed assuming a triangular shape (Equation 7).

$$Q_P = 0.607\ (V_w{}^{0.295}\ h_w{}^{1.24}) \tag{5}$$

$$t_f = 63.2\ \sqrt{\frac{V_w}{g\ h_b{}^2}} \tag{6}$$

$$V = \frac{Q_P\ T}{2} \tag{7}$$

Where V_w and h_w are the volume and the depth of water, respectively, at the time of the dam failure above the lowest breach cross-section level. h_b is the final breach height. V and T are the total volume and the synthetic hydrograph base time respectively.

4.4. OWP HAND-FIM

The OWP HAND-FIM model is typically used to convert the streamflow from the National Water Model to the inundation map. The OWP HAND-FIM version 4 is a no-physics model that produces static flood inundation maps using Synthetic Rating Curves (SRC) which convert streamflow predictions in a river reach to water height (stage; Aristizabal et al., 2023). The model requires dependencies such as Docker and GitHub. The first step involves identifying the HUC-8 boundaries within which the study area is located. The Fall River study area is within 11070102 HUC-8. After determining the HUC-8 boundaries, input folders containing the processed HAND raster for the mainstem, and branches are downloaded from an Amazon S3 bucket

(NOAA Office of Water Prediction, 2023). Detailed FIM hydrofabric information, including hydraulic radius, wet area, volume, and SRCs for each sub-catchment and stream network, is stored in a 'hydrotable.csv' file. Each stream network within HUC-8 has a unique ID. In this project, we aim to simulate the effect of a dam break on the Fall River. To achieve this, we identify the specific stream IDs related to the reaches downstream of the dam and assign the calculated discharges from the analytical method to these stream IDs in OWP HAND-FIM.

4.5. FLDPLN Model

The FLDPLN model (Kastens, 2008) is a low-complexity flood inundation model designed for mapping flood extents and depths, particularly useful for rapid assessments using observed or forecasted stream stages. The model operates by segmenting the stream networks and utilizing mechanisms such as backfill and spillover flooding to determine flood inundation. This process includes identifying stream networks, generating segments, and iteratively determining the minimum depth to flood (DTF) at the flood source (i.e., stream) pixels (FSPs) to inundate floodplain pixels (FPPs). The flood inundation relations are organized by tiles, with each tile containing all the FSPs that can flood the FPPs within the tile. Flood inundation mapping involves estimating the depth of flow (DOF) for FSPs from gauge observations or forecasts and mapping flood depths for FPPs by calculating the maximum flood depth from multiple FSPs. The attenuated Qpeak for each river reach is converted into depths using the SRC from the OWP HAND-FIM model.

4.6. Hydrodynamic Modeling using HEC-RAS

A 2D hydrodynamic flood model has been developed using HEC-RAS, version 6.4. HEC-RAS solves the 2D full momentum Shallow Water Equations (SWE) using the Eulerian-Lagrangian method (ELM). The objective of developing this computationally intensive numerical model is to serve as a benchmark for comparison with low-complexity, terrain-based FIM models.

The HEC-RAS model incorporated the Fall River Dam break simulation and hydraulic routing of the surge wave for the 50 km downstream river to Fredonia City. The hydrodynamic model used the same 10-meter DEM as that of the FLDPLN model. The model utilized a 2D unstructured mesh with a 30-meter cell size, and a refinement region for the river floodplain with a 10-meter cell size. Break lines were defined for the main river as well as the tributaries downstream of the dam. The total number of cells in the 2D domain was approximately 1.5 million. The 2D SWE-ELM equation set was used with a computational interval set to 30 sec. A Manning's roughness coefficient of 0.06 $s/m^{-1/3}$, retrieved from the OWP HAND-FIM hydro tables, was applied uniformly over the entire domain. The Qpeak for two scenarios, as explained in section 4.3, is routed using the HEC-RAS model.

4.7. FIM Evaluation

Quantitative and impact analyses are used to evaluate the performance of the OWP HAND-FIM and FLDPLN models, using the HEC-RAS as the benchmark FIM. One common approach for quantitative analysis of flood extent involves binary contingency statistics, comparing true positives (TPs), false positives (FPs), false negatives (FNs), and true negatives (TNs).

The contingency statistics are used to calculate metrics like the critical success index (CSI), F1-score, Recall, false alarm rate (FAR), false negative rate (FNR), overall accuracy (ACC), and

precision. CSI measures the accuracy of positive predictions, considering false alarms and missed events. The Recall, also known as sensitivity, and FAR play crucial roles in flood modeling, indicating a model's ability to accurately detect flooding and the extent of over-prediction, respectively. Accuracy measures the proportion of correctly identified pixels (both positive and negative) out of the total instances. Precision measures the proportion of true positive predictions out of all positive predictions made by the model. The F1-score is the harmonic mean of precision and sensitivity, providing a single metric that balances both metrics. The FNR indicates the percentage of real flood occurrences that the model fails to predict.

In addition to flood extent, we also evaluated the flood depth predicted by the models against the benchmark. The metrics used to evaluate accuracy include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Nash-Sutcliffe Efficiency (NSE) assesses efficiency, while Bias and Percent Bias (PBIAS) indicate a model's tendency to produce systematic errors.

RMSE measures the average magnitude of the errors between predicted and observed values. It's always non-negative, with lower values indicating better accuracy. A RMSE close to 0 is desirable. NSE measures the relative magnitude of residual variance compared to observed data variance. Its values range from $-\infty$ to 1. Generally, NSE > 0.5 is considered acceptable for hydrological models, indicating the model captures more variance than the mean. A negative value means model predictions are worse than the mean of observed data. Bias measures the average difference between predicted and observed values. Positive values indicate overestimation, while negative values indicate underestimation. Ideally, bias should be close to 0 for an unbiased model.

Moreover, as the study area includes urban land use, specifically Fredonia City in the southeast of Kansas, we investigated the building footprints (Microsoft, 2018) to determine which model provides a better estimation of the number of inundated buildings compared to the benchmark map.

5. Results and Discussions

In this section, first the peak discharge and attenuations followed by the quantitative, and impact FIM evaluations are presented for the second scenario, where the dam break discharge is estimated using empirical equations. The Qpeak estimated using Eq. 5 at the dam was 10,000 m^3/s. The attenuation determined using the 2D HEC-RAS simulation along a 45 km downstream section is 25% while the analytical model estimated the attenuation to be 20% as shown in Figure 5. Figure 6 shows the inundation extent and depth for all three models. Figure 7 (a) and (b) show the comparison of the OWP HAND-FIM and FLDPLN models, respectively, with the HEC-RAS model. The quantitative metrics (precision, recall, F1-score, CSI, accuracy, FAR, and FNR) are shown in Figure 9. The OWP HAND-FIM predicted the inundation areas with an accuracy of 92.1%. The model also achieved a high recall rate of 83% and a reasonable false alarm rate of 11.5%, indicating its robustness in accurately identifying flood-prone areas. It is worth noting that the HEC-RAS simulation time was about 17 hours in a high-performance computing system while HAND-FIM was less than a minute. This shows that the methodology of routing the discharge using the analytical model and generating the inundation maps using OWP HAND-FIM is robust and can potentially be adopted for rapid and large-scale FIM caused by dam break failures and dam operations. The FLDPLN model yielded an accuracy of 87.5%

when compared to the HEC-RAS model. The reason for the overestimation of the flood extent in the FLDPLN model can likely be attributed to errors in estimating stages using the SRCs from OWP HAND-FIM. When compared to the stages from the HEC-RAS model, the SRCs overestimate the stages for the corresponding peak discharges computed from the analytical model. It is worth noting that the FLDPLN model's inundation map is estimated using both backfill and spillover techniques making it more sensitive to overprediction of stage. The OWP HAND-FIM model, on the other hand, does not account for spillover resulting in an underestimation of the inundation extent for the same stages in the downstream reaches. The stages from HEC-RAS model can be determined and used as an input for the FLDPLN model to estimate the error originating from the HAND-FIM SRCs. This will be included in the future scope of this project.

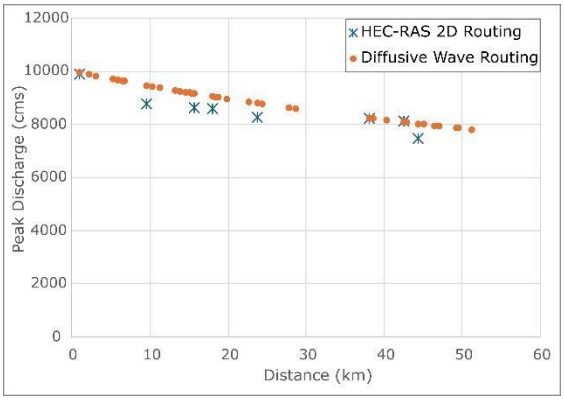


Figure 5: Peak discharge attenuation using HEC-RAS and Diffusive wave equation.

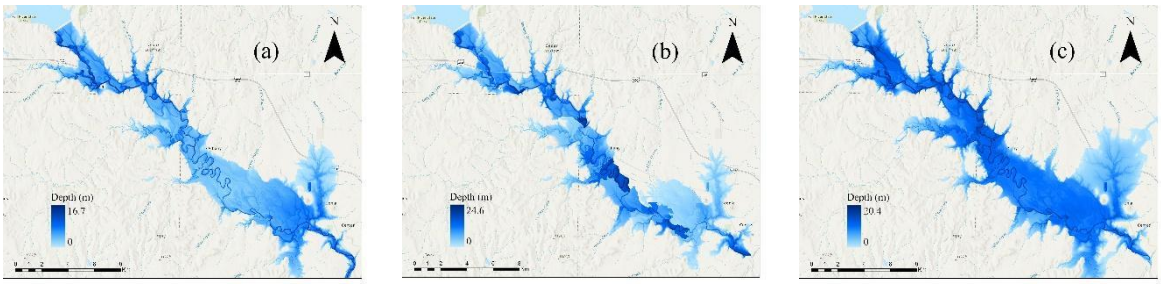


Figure 6: Inundation extent and depths for a) HEC-RAS, b) HAND-FIM, c) FLDPLN

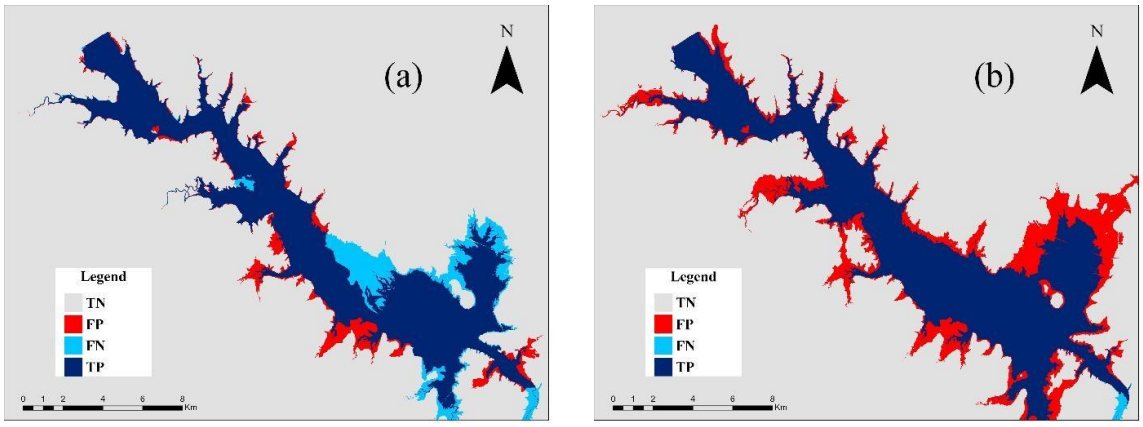


Figure 7: Confusion matrix results comparing the performance of low-complexity FIM models to the HEC-RAS model. (a) shows the results for the OWP HAND-FIM model and (b) presents the results for the FLDPLN model.

Though the inundation extents are more reliable with the OWP HAND-FIM model the depths are exaggerated due to the SRC errors and inherent limitations in the model's depth-predictions (multi 'level paths' approach). This limits the usage of OWP HAND-FIM, in its current version,

for predicting floodwater depth. Figure 8 shows the difference in depth between terrain-based models and the HEC-RAS model. Results for evaluating depth are shown in Figure 10. The RMSE of 4.34 m for the HAND model, compared to 4.63 m for the FLDPLN model, suggests that the HAND model exhibits slightly higher accuracy. Additionally, lower MAE and MAPE values for HAND compared to FLDPLN further support its superior accuracy. The Bias of 1.55 for the HAND map indicates an overestimation of flood depths. In contrast, the FLDPLN map shows a much higher Bias of 4.24, suggesting a significant overestimation of flood depths on average. The HAND model predictions are 1.56 units higher than the observed values, indicating an overestimation. This is confirmed by the PBIAS of 41.34%, indicating a significant overestimation bias. The negative NSE values (-2.15 for HAND and -2.59 for FLDPLN) indicate poor performance for both models, as their predictions are worse than the mean of the observed values.

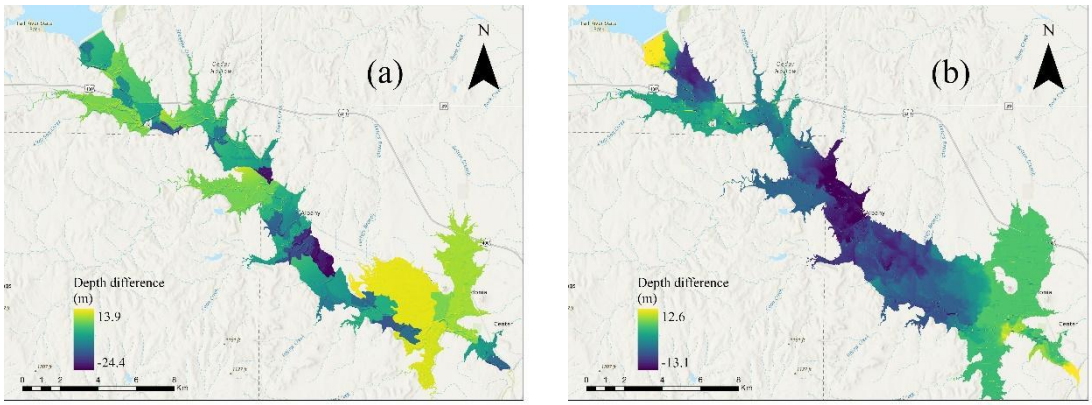


Figure 8: Flood depth differences between HEC-RAS and a) HAND-FIM, b) FLDPLN

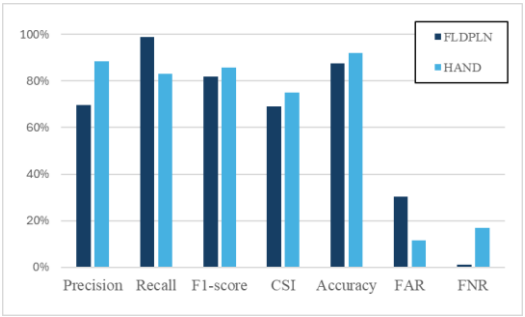


Figure 9: Evaluating model extent against HEC-RAS

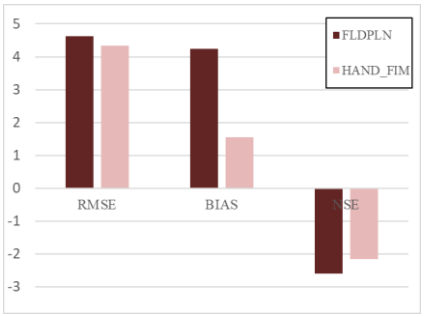


Figure 10: Evaluating model depth against HEC-RAS

A quick impact analysis (Figure 11) show that the number of buildings flooded by the models is 679, 577, and 1146 for HEC-RAS, OWP HAND-FIM, and FLDPLN models respectively. This confirms the overestimation of flood-prone buildings by the FLDPLN model. However, the underestimation of buildings by the HAND model also poses challenges in using it as a reliable source for providing accurate flood risk maps.

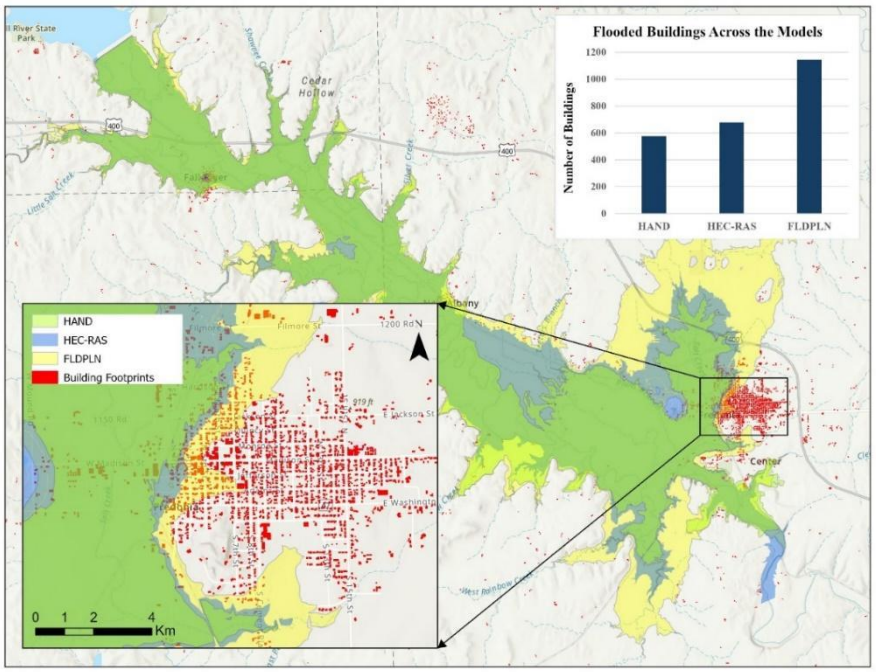


Figure 11: Flood impact assessment results for the Fall River Dam break using different FIM models (HEC-RAS, OWP HAND-FIM, FLDPLN)

6. Conclusion

This study demonstrates the effectiveness of terrain-based models OWP HAND-FIM and FLDPLN in generating rapid flood inundation maps for floods caused by dam failures and operations. The models were evaluated using the benchmark HEC-RAS model using various quantitative metrics. These findings underscore the potential of using low-complexity, terrain-based models for near real-time flood mitigation, significantly reducing computation time compared to traditional hydrodynamic models. Therefore, this study highlights the following findings: the practical applicability of terrain-based models in enhancing flood preparedness and response, particularly in regions where computational resources are limited, supports the development of terrain-based FIM for dam related floods.

Comments from authors:

Graduate students should embrace simplicity and interdisciplinary learning for innovative and comprehensive flood management solutions and build networks with peers and professionals to foster collaboration and innovation in flood management. NWS/NWC leadership could integrate terrain-based models like HAND-FIM and FLDPLN for enhanced real-time flood response and support rapid research initiatives. Future research could focus on integrating these models with other predictive tools, evaluating their performance across different geographies, leveraging machine learning for improved accuracy, and developing user-friendly platforms for community engagement and resilience.

Acknowledgement:

We thank CIROH, CUASHI, and the NOAA National Water Center (NWC) for their support during the NWC Innovators Program Summer Institute 2024. We deeply thank our team leads

and academic advisors, Dr. Xingong Li, Dr. Sagy Cohen, Dr. Jasim Imran, Dr. Erfan Goharian, and Dr. Steven J. Burian, for their unwavering support. We acknowledge Dr. Anupal Baruah and Dr. Junho Song for their assistance with the OWP HAND-FIM and FLDPLN models. We also appreciate the valuable feedback and ideas from Dr. Fernando Salas, Carson Pruitt, and Royce Fontenot. Our gratitude extends to Dr. James Halgren for providing computing resources, which greatly facilitated our work environment. Finally, we are thankful to Sadaf Mahmoudi and Karina Larco for coordinating the Summer Institute and making it fun-filled.

Supplementary Materials:

The following performance metrics are used for evaluating the inundation extent (Equations S1 to S7) and depth (Equations S8 to S13) of the terrain-based models (OWP HAND-FIM & FLDPLN) with the HEC-RAS model.

$$precision = \frac{TP}{TP + FP} \tag{S1}$$

$$Recall = \frac{TP}{TP + FN} \tag{S2}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \tag{S3}$$

$$CSI = \frac{TP}{TP + FP + TN} \tag{S4}$$

$$ACC = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{S5}$$

$$FAR = \frac{FP}{TP + FP} \tag{S6}$$

$$FNR = \frac{FN}{FN + TP} \tag{S7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left[Pi - O_i\right] \tag{S8}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{Oi - Pi}{Oi}\right| 100 \tag{S9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (Pi - Oi)^2} \tag{S10}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} (O_i - O)^2} \tag{S11}$$

$$BIAS = \frac{1}{n}\sum_{i=1}^{n} (Pi - Oi) \tag{S12}$$

$$PBIAS = \frac{\sum_{i=1}^{n} (Oi - Pi)}{\sum_{i=1}^{n} Oi} 100 \tag{S13}$$

Where O represents the observed data (HEC-RAS model) and P indicates predicted data (terrain-based FIM).

Root Mean Square Error (RMSE) measures the average magnitude of the errors between predicted and observed values. It's always non-negative, and lower values indicate better accuracy. It means an RMSE close to 0 is desirable. Nash-Sutcliffe Efficiency (NSE) measures the relative magnitude of residual variance compared to observed data variance. Its Values range from -∞ to 1. Generally, NSE > 0.5 is considered acceptable for hydrological models, indicating the model captures more variance than the mean. A negative value means Model predictions are worse than the mean of observed data.

Bias measures the average difference between predicted and observed values. Positive values indicate overestimation, while negative values indicate underestimation. Ideally, bias should be close to 0 for an unbiased model. Mean Absolute Error (MAE) measures the average magnitude of errors in a set of predictions, without considering their direction (i.e., whether the predictions are overestimations or underestimations). MAE values can range from 0 to ∞, where 0 indicates perfect accuracy.

An RMSE of 4.34 for the HAND map vs. 4.63 for the FLDPLN map means the HAND map is slightly more accurate. The MAE of 3.08 for the HAND map suggests that it has moderate accuracy. The higher MAE of 4.36 for the FLDPLN map compared to the HAND map suggests that the FLDPLN map has lower accuracy and requires more significant improvements to reduce the prediction errors. The Bias of 1.55 for the HAND map shows that it is overestimating flood depths. The FLDPLN map significantly overestimates flood depths by 4.24 units on average. The reason for the lesser amount of Bias in the HAND map is that it has both overestimations and underestimations, which counterbalance each other. In contrast, the FLDPLN map is completely overestimating, which is why the RMSE, MAE, and Bias values are almost the same. The negative NSE values in this case (-2.15 for HAND and -2.59 for FLDPLN) indicate that both models perform poorly, predicting worse than the mean of the observed values. This poor performance can be attributed to the extreme depth values in the models, which significantly impact the calculation when squared.

References

1. Aristizabal, F., Salas, F., Petrochenkov, G., Grout, T., Avant, B., Bates, B., Spies, R., Chadwick, N., Wills, Z., & Judge, J. (2023). 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, 59(5), e2022WR032039. https://doi.org/10.1029/2022WR032039

2. Bureau of Reclamation. (1988). *Downstream hazard classification guidelines*. Acer Technical memorandum nº11.

3. Devi, D., Baruah, A., & Sarma, A. K. (2022). Characterization of dam-impacted flood hydrograph and its degree of severity as a potential hazard. *Natural Hazards*, *112*(3), 1989–2011. https://doi.org/10.1007/s11069-022-05253-7

4. Froehlich, D. C. (1995). Peak outflow from breached embankment dam. *Journal of Water Resources Planning and Management*, *121*(1), 90–97.

5. Froehlich, D. C. (2008). Embankment Dam Breach Parameters and Their Uncertainties. *Journal of Hydraulic Engineering*, *134*(12), 1708–1721. https://doi.org/10.1061/(ASCE)0733-9429(2008)134:12(1708)

6. Ge, W., Jiao, Y., Wu, M., Li, Z., Wang, T., Li, W., Zhang, Y., Gao, W., & Van Gelder, P. (2022). Estimating loss of life caused by dam breaches based on the simulation of floods routing and evacuation potential of population at risk. *Journal of Hydrology*, *612*, 128059. https://doi.org/10.1016/j.jhydrol.2022.128059

7. Kansas Biological Survey. (2011). *Bathymetric and Sediment Survey of Fall River Reservoir, Greenwood County, Kansas* (2010-02 (May 2011); Applied Science and Technology for Reservoir Assessment (ASTRA) Program).

8. Kastens, J. H. (2008). *Some New Developments on Two Separate Topics: Statistical Cross Validation and Floodplain Mapping* [PhD Dissertation]. University of Kansas.

9. MacDonald, T. C., & Langridge-Monopolis, J. (1984). Breaching Charateristics of Dam Failures. *Journal of Hydraulic Engineering*, *110*(5), 567–586. https://doi.org/10.1061/(ASCE)0733-9429(1984)110:5(567)

10. NOAA Office of Water Prediction. (2023). *Inundation Mapping* (4.4.2.1) [Computer software]. https://github.com/NOAA-OWP/inundation-mapping

11. Paiva, R. C. D., & Lima, S. G. (2024). A Simple Model of Flood Peak Attenuation. *Water Resources Research*, *60*(2), e2023WR034692. https://doi.org/10.1029/2023WR034692

12. Pianforini, M., Dazzi, S., Pilzer, A., & Vacondio, R. (2024). Real-time flood maps forecasting for dam-break scenarios with a transformer-based deep learning model. *Journal of Hydrology*, *635*, 131169. https://doi.org/10.1016/j.jhydrol.2024.131169

13. U. S. Army Corps of Engineers. (2024). *National Inventory of Dams*.

14. Von Thun, J. L., & Gillette, D. R. (1990). *Guidance on breach parameters*. US Department of the Interior, Bureau of Reclamation.

15. Wahl, T. L. (2004). Uncertainty of Predictions of Embankment Dam Breach Parameters. *Journal of Hydraulic Engineering*, *130*(5), 389–397. https://doi.org/10.1061/(ASCE)0733-9429(2004)130:5(389)

16. Xiong, Y. (2011). A Dam Break Analysis Using HEC-RAS. *Journal of Water Resource and Protection*, *03*(06), 370–379. https://doi.org/10.4236/jwarp.2011.36047

17. Microsoft. (2018). *US Building Footprints* [Dataset]. https://github.com/Microsoft/USBuildingFootprints

Chapter 6
Transition From Hazard to Impact-based Riverine Flood Forecasting

Hesham Elhaddad[1], Duc Tran[2], and Lyn Watts[3]

*All authors contributed equally, and the order is set alphabetically.

[1]Geological and Environmental Sciences Department, Western Michigan State University, Kalamazoo, MI 49008, USA; *h.elhaddad@wmich.edu*

[2]Department of Civil and Environmental Engineering, University of Virginia, Charlottesville, VA 22904, USA; *syu3cs@virginia.edu*

[3]Department of Earth, Geographic, and Climate Sciences, University of Massachusetts Amherst, Amherst, MA 01003, USA; *cwatts@umass.edu*

Academic Advisors: Mohamed Sultan, *Western Michigan State University*; Venkataraman Lakshmi, *University of Virginia;* Christine Hatch, *University of Massachusetts Amherst*.

Summer Institute Theme Advisors: Sanjib Sharma, Howard University, *sanjib.sharma@howard.edu;* Kelsey McDonough, FloodMapp, *kelsey@floodmapp.com*

Abstract: When floods occur, first responders face challenges in identifying where vulnerable populations will be. However, hazard forecasts lack detailed information about exposure and vulnerability. In this study, we create an automated workflow that transitions from hazard to impact-based flood forecasting. Using the National Water Model (NWM) forecasted streamflow, we generate Flood Inundation Maps (FIMs) using the Office of Water Prediction HAND-FIM, calculate the forecasted water depth, and incorporate a Social Vulnerability Index (SVI) to predict the flood impacts. These forecasts are shared via an easy-to-use, interactive web platform accessible to the community. This framework will allow stakeholders and decision-makers to plan for flood events.

---

1. Motivation

Flood forecasting represents a significant advancement in flood risk management, offering insights that enhance preparedness and response strategies [1], [2]. However, current flood forecasts in the United States do not predict where a flood will have the greatest impact on humans. These forecasts do not provide information about flood forecast confidence, flood depth, or the potential risk to life and property. This is information that would benefit emergency responders and planners. We aim to fill this gap by using forecasted data to produce impact maps that will include information about flood hazards and population vulnerability.

2. Objectives and Scope

This work aims to provide a framework for forecasting flood impact in different temporal scales. This workflow uses the National Water Model (NWM) forecast from the National Oceanic and Atmospheric Administration (NOAA) as a base and builds on it with three main innovations: (1) calculate the flood forecast confidence from the NWM hourly forecasts, (2) incorporate

depth maps in addition to an inundation forecast, and (3) calculate social vulnerability to flooding by considering indicators such as access to transportation, population density, age, and disability. By shifting the focus from traditional flood hazard metrics to a more comprehensive analysis of community-specific vulnerabilities, our study aims to improve the effectiveness of emergency response efforts. The results will provide a valuable scientific basis to stakeholders and authorities in planning sustainable strategies to mitigate flood impacts.

3. Previous Studies

Several previous studies have combined flood hazards and social vulnerability to create comprehensive impact maps, which provide valuable insights for disaster management and mitigation efforts [3], [4], [5], [6]. For instance, a study [3] developed a method to assess social vulnerability to floods by integrating various exposure and coping capacity indicators. This approach was applied in Ponferrada, Spain, using public data and an analytic hierarchy process to create detailed social vulnerability maps. These maps help in flood risk planning and management. Similarly, another study [4] introduced a method for estimating flood hazard, vulnerability, and risk at the household level in Nan Province, Thailand. They used flood simulation models for different return periods and incorporated field data and surveys to produce flood risk maps that reflect real-world conditions and guide effective flood mitigation measures. A study [5] employed dasymetric mapping techniques and flood maps to estimate that 41 million people are exposed to the 100-year floodplain across the contiguous United States. This research highlighted the higher exposure and susceptibility of socially vulnerable populations living in flood-prone areas, such as those with lower incomes and more mobile homes, compared to national averages. Additionally, a study [6] focusing on Harris County, Texas, combined flood susceptibility and social vulnerability mapping to enhance flood risk analysis. This study used advanced geospatial data and flood modeling to assess risk disparities and inform effective flood mitigation measures, particularly in socially vulnerable communities. Previous studies provide very important insights into extreme event mapping, climate impact assessment and retrospective impact analysis. Here, we expand on previous studies by developing, evaluating, and mainstreaming a workflow to transition from hazards to impact-based flood forecasting.

4. Methodology

In this study, we constructed an automated workflow (operational) that can produce real-time, and historical observations of flood impact maps in two temporal scales (18 hours and 5 days). This operational workflow is ready to be published on an interactive platform (ArcGIS instant app) providing real-time forecast impact maps. The impact maps are created by coupling the flood forecast confidence (only on an hourly-basis scale), flood depth maps (flood hazards), and Social Vulnerability Index (SVI) maps. The workflow (Fig. 1) includes the following steps: (1) retrieve the streamflow data, (2) generate FIM and depth maps using OWP HAND-FIM (Office of Water Prediction Height Above Nearest Drainage - Flood Inundation Model), (3) create SVI maps, (4) couple the flood depth, flood forecast confidence, and SVI maps to produce impact maps, then overlay buildings, hospital, and schools layers over the impact maps, and (5) publish the maps (FIMs and depth maps, SVI, and impact maps with building layer on ArcGIS Instant App.

Additionally, we tested an approach to mitigate the inconsistency in the hourly forecast data by creating a flood forecast confidence map. We tested this approach on Tropical storm Beryl but it was not included in the main workflow.

4.1. Retrieve the streamflow data

The NWM became operational in the United States in 2016, predicting streamflow in all streams and reaches across the country [7]. This large-scale hydrological model uses meteorological inputs to provide streamflow forecasts nationwide. The model delivers forecasts at different temporal scales: short-range forecasts (hourly; up to 18 hours) forced with meteorological data from the High-Resolution Rapid Refresh and Rapid Refresh models, medium-range forecasts (5 day forecast) forced with Global Forecasting System model output [8]. The streamflow data in our workflow are divided into two categories: real-time forecast (18 hour and 5 days) and historical streamflow observations.

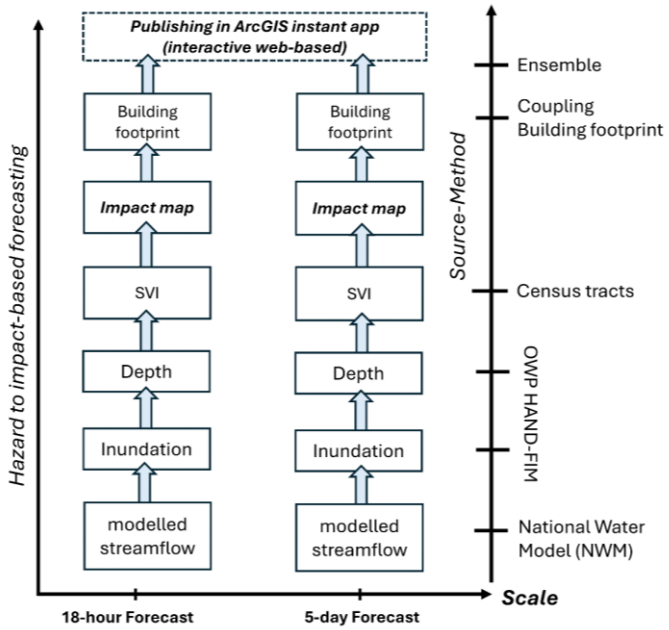


Figure 1. Proposed framework for creating the operational forecasted impact maps using depth and SVI maps.

4.2. Generate Flood Inundation Maps using OWP HAND-FIM

Beginning in 2023, the National Water Center (NWC) started deploying Flood Inundation Maps (FIMs) generated from the NWM streamflow forecasts, which will be available nationally by 2026 [7] [8]. Instead of hydrodynamic models, the NWC employs a simplified model for estimating flood inundation, referred to as OWP HAND-FIM (height above nearest drainage - flood inundation model) using the streamflow forecasts as input. There are a myriad of models that forecast flood inundation, each varying in complexity and dimensionality. Hydrodynamic models, for instance, solve the shallow water equations to simulate water movement and inundation patterns. However, these models require extensive input data and significant computational resources to process these inputs and solve the equations, making real-time implementation and large-scale application challenging [9]. Here lies the advantage of using HAND-FIM for short-term forecasting.

HAND is a terrain-based method that normalizes elevations within Digital Elevation Models (DEMs) to the nearest flow path and extends the stage elevation across that area [1], [10], [11].

It conserves neither mass nor momentum and does not attempt to solve the flow equations used in more complex models [9], [12]. This model is computationally much simpler to perform, making it an easy choice for creating real-time or near real-time flood inundation maps [9], [11], [13]. Our framework uses this model on the 18-hour streamflow forecasts retrieved from the National Water Model to produce flood inundation maps and depth maps.

Estimate the flood depths

For the 5 day forecast, a flood inundation map is provided by the National Water Model. We use the FIM to identify flooded pixels (flooded = 1, non-flooded = 0). We then extract ground elevation values for each flooded pixel from the DEM. The workflow subtracts the ground elevation from the flood surface elevation for each pixel using the formula: Flood Depth = Flood Surface Elevation - Ground Elevation. Next, we generate a depth map showing water depth across the inundated area. Finally, the depths are classified based on Federal Emergency Management Agency (FEMA) Guidance for Flood Risk Analysis and Mapping [14]. FEMA categorizes the depth severity into four classes (<0.4 m is low severity, 0.4-0.8 m is moderate severity, 0.8-1.8 m is high severity, and >1.8 m is high severity.

4.5. Create Social Vulnerability maps

The United States Center for Disease Control (CDC) calculates a social vulnerability index (SVI) that utilizes the American Community Survey (ACS) census data to determine the relative social vulnerability of every census tract in the United States [15]. The SVI considers indicators such as socioeconomic status, household composition, and housing type to evaluate the ability of a targeted community to prepare for, respond to, and recover from disasters [16], [17]. The SVI ranks each U.S. Census tract based on social indicators grouped into four themes: (1) socioeconomic status (poverty, unemployment, housing cost burden, no health insurance), (2) household (aged 65 or older, aged 17 or younger, disability), (3) language barrier, and (4) housing type and transportation (multi-unit structures, mobile homes, crowding, no vehicle, group quarters) [15], [18]. The SVI calculation involves cleaning the data, determining the percentage of a specified indicator, ranking it, setting flags for high vulnerability, and summing these flags to obtain the final SVI score.

The process begins with data cleaning, where tracts with missing data are replaced by averaging the values of neighboring tracts. The percentage of the specified indicator is calculated using the formula: Percentage of indicator = (raw estimate of indicator/relevant total population) ×100. These percentages are then ranked within the dataset to determine their percentile ranks Flags for high vulnerability are set by comparing the percentile rank to the 90th percentile threshold: if the percentile of the indicator is greater than or equal to 0.90, the flag is set to 1, indicating high vulnerability; otherwise, it is set to 0. The final SVI score for each tract is obtained by summing the flags for all indicators. We create a classified (number of classes equal to the maximum SVI) raster map that shows the geospatial distribution of the calculated SVI.

4.6. Impact map

We designed the impact map to comprehensively assess flood impacts by integrating both hydrological and social data. The impact map is obtained by multiplying the depth map, flood forecast confidence map, and SVI map. This approach ensures each pixel reflects the severity of flooding, the likelihood of occurrence, and the vulnerability of the population. The resulting

impact map highlights areas with varying levels of vulnerability, combining both physical and social dimensions of flood risk.

Flood Forecast Confidence Map

This process was tested on the hourly basis forecast, and it was not included in the workflow. A flood forecast confidence map indicates the reliability of flood predictions, showing the likelihood of an area being flooded. It is important because it helps decision-makers assess the certainty of flood forecasts, improving preparedness and response strategies. To create the flood forecast confidence map using the Flood Inundation Maps (FIMs), we followed the process outlined in figure 2. Every hour, the NWM provides a forecast for each of the next 18 hours. This means that a single hour will be forecasted 18 times, at each of the previous hours. Within the 18 iterations, the single-hour forecast suffers from inconsistency, where some iterations predict widespread flooding and other iterations predict little. For example, a forecast published at 7:00 am (real time) for 5:00 pm shows a specific area as flooded, but the same forecast for 5:00 pm at 8:00 am (real-time) shows the same area as non-flooded. This inconsistency poses a challenge for decision-makers, watching flood predictions change from hour to hour. To address this inconsistency, we calculated the hourly flood forecast confidence based on the previous forecasts, to produce a forecast confidence map (Fig. 2).

First, we collected all the forecasted streamflow data from the NWM for the same forecast hour. We calculated a FIM raster from each, where flooded pixels are assigned a value of 1 and non-flooded pixels a value of 0. Next, we calculated the percentage of times each pixel was predicted to be flooded across all forecasts. This results in a confidence value for each pixel, ranging from 0 to 1.

During this process, we encountered and resolved two specific problems. First, we needed to retain the confidence value of pixels that were non-flooded (value of zero) in the new FIM but had been flooded in previous FIMs. To address this, we ensured that the confidence values of such pixels were carried forward rather than reset to zero. Second, for pixels that were newly flooded in the latest FIM and had not been flooded before, we calculated a new confidence value by assigning a value of 1 divided by the total number of FIMs (including the new one). This approach ensured that new flood occurrences were incorporated appropriately into the flood forecast confidence map. Finally, the flood forecast confidence map was used to adjust the new FIM. We multiplied the new FIM by the flood forecast confidence map, creating an adjusted FIM that integrates previous flood patterns and confidence values. This adjusted FIM provides a more accurate representation of potential flooding by incorporating both current and previous predictions with a confidence interval for every area. This innovation will help to better identify areas at risk and improve flood risk management strategies.

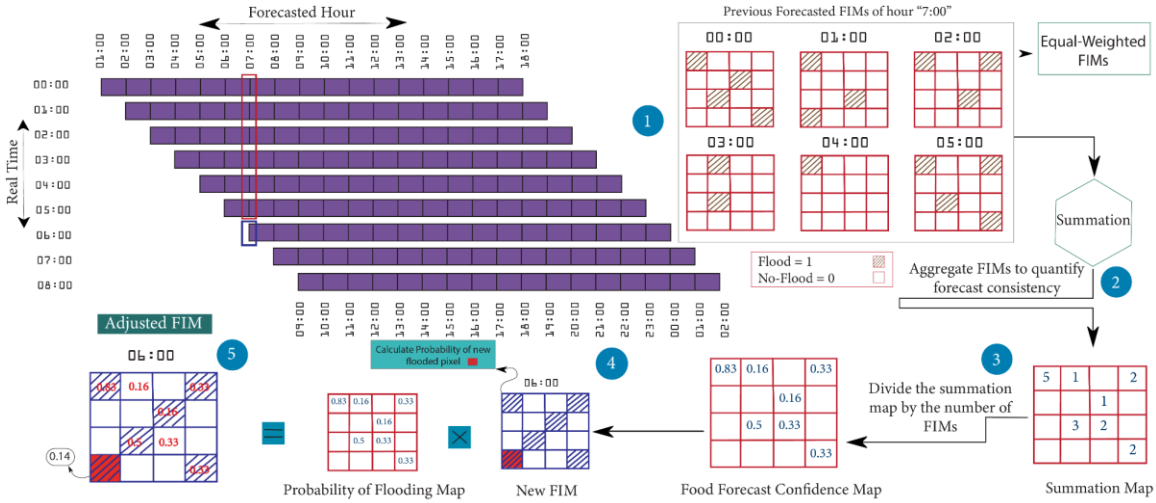


Figure 2. Proposed framework for creating a flood forecast confidence map for each forecasted timestep to adjust the resulting forecasted FIM.

Operational Impact Forecast

The previously described workflow was compiled as a Python script (see Supplementary Materials) to allow it to be run hourly. The script follows the following steps: (1) Download 5-day forecast data from the NWM, (2) Import DEM and SVI, (3) Calculate depth from the inundation extent 4) Calculate the forecasted impact for the forecast, and (5) upload the depth map and the impact map to a web-hosted map that can be shared with interested parties. This script can be run locally on any machine and can be scheduled to run hourly as the NWM forecast updates. The current iteration of our script focuses on Houston, TX but a user can adjust the workflow to any area of interest in the country.

5. Results: Case Study

Our workflow requires minimal inputs to generate operational forecasted impact maps on both an 18-hour basis and for 5-day forecasts. Designed for ease of use, this workflow can be applied to any location across the United States. Additionally, it can produce historical FIMs and impact maps for any past flood events nationwide. The primary inputs for this process include the SVI data geodatabase, Hydrologic Unit Code (HUC) ID, a shapefile of the study area, and a DEM. The workflow is structured to ensure that it can be utilized effectively by a range of users, from emergency responders to urban planners. The integration of SVI data allows for a better understanding of flood impacts, considering not only the physical extent of flooding but also the social vulnerabilities of affected communities. This dual approach enhances the relevance and applicability of the generated maps, providing crucial information for decision-making in disaster management.

To validate the workflow, we tested it on two significant events in Houston: (1) Post-Tropical Cyclone Beryl for real-time forecasts (hourly-basis forecast and 5-day forecast) (July 08-13, 2024), and (2) Hurricane Harvey (August 25-29, 2017) using historical streamflow observations. These case studies demonstrate the capability of the workflow to generate detailed and actionable flood impact maps. For Post-Tropical Cyclone Beryl, we used real-time streamflow data from the National Water Model (NWM) to produce hourly FIMs and impact maps, showcasing the workflow's real-time application. For Hurricane Harvey, we used historical streamflow

observations to reconstruct the event's flood impacts, highlighting the workflow's utility in analyzing past events.

Post-Tropical Cyclone Beryl (2024)

Beryl began as a hurricane, striking southeastern Texas in early July 2024 as a Category 1 storm. It caused extensive damage, with nearly 2 million homes in the Houston, TX area losing power due to strong winds and heavy rainfall. The federal government declared an emergency to aid recovery efforts [22]. Using the NWM hourly streamflow forecast, we calculated the flood forecast confidence every hour to account for hourly forecast inconsistencies (Fig. 4). We compared the output non-adjusted FIMs (Figs 3a, b, c; no confidence included) adjusted FIMs (Figs 3d, e, f; confidence included) for 6:00 pm at three consecutive hours (6:00, 7:00 and 8:00 am) on July 8, 2024. Each timestep (6:00, 7:00, and 8:00 am) forecast for 6:00 pm takes into consideration the confidence calculation of the previous forecast (started from 12:00 am). This means that at 6:00 am, six forecasts of 6:00 pm were incorporated in the confidence estimations. Figures (3a, b, and c; defined by red rectangle) highlight inconsistencies between flooded and non-flooded areas. Figures (3d, e, and f; defined by red rectangle) show adjusted FIMs with confidence estimates demonstrating run-to-run consistency.

The adjusted impact map (Fig. 4a) categorizes impacts into four levels (Low, Medium, High, and Very High), integrating flood depth with four classes of severity and SVI with nine levels of vulnerabilities. Additionally, a 5-day forecast impact map for Beryl on July 8th incorporates flood depth and SVI, shown in Fig. 4b.

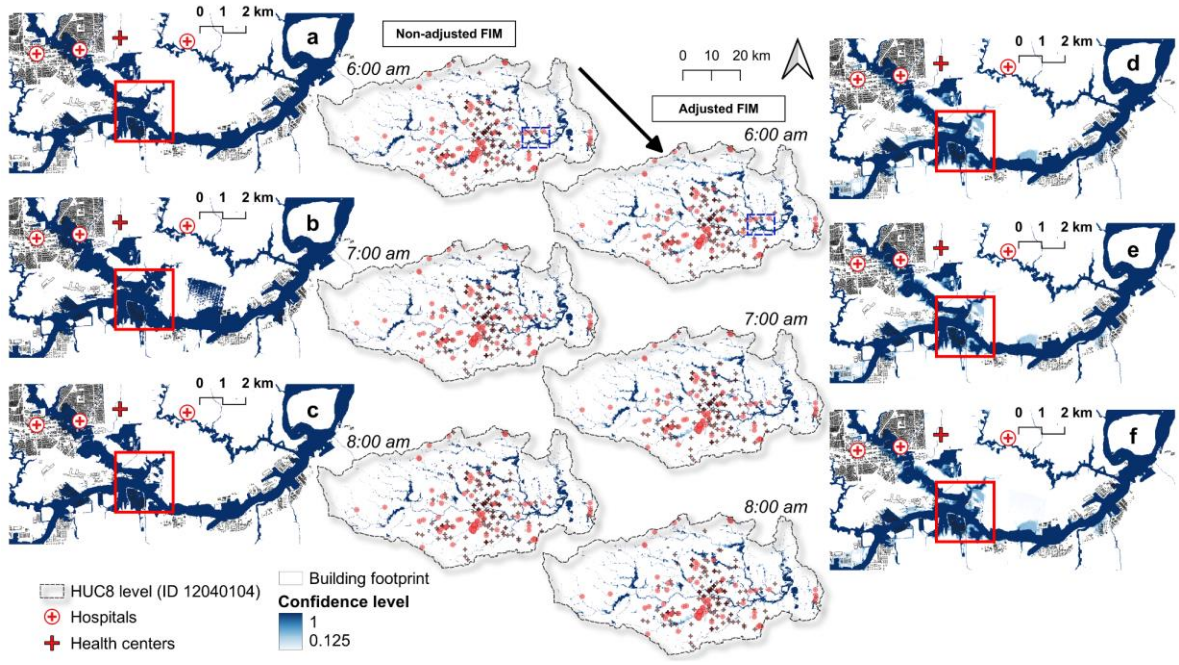


Figure 3. A comparison between non-adjusted (a, b, c) and adjusted FIMs (d, e, f) at three consecutive hours (6:00 am, 7:00 am, and 8:00 am) for the same forecasted hour (6:00 pm). The area (defined by the red rectangle) shows forecast inconsistency in the non-adjusted FIMs, while showing consistency, in addition to flood forecast confidence estimates in the adjusted FIM.

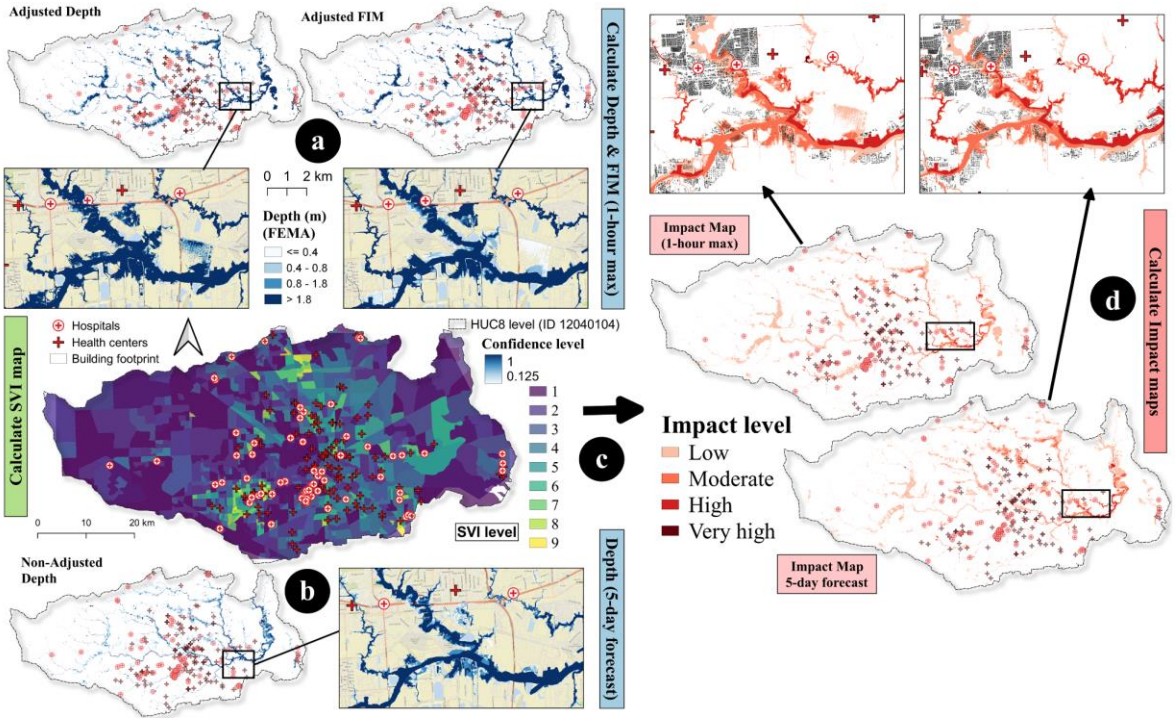


Figure 4. shows two temporal scale forecasts of Post-Tropical cyclone Beryl. (a) the adjusted FIM (confidence map) and depth map with four classes of flood severity for the hourly-basis forecast. (b) the depth map with four classes for the five days forecast. c) SVI map shows nine levels of vulnerability (lowest =1, highest =9). d) the output impact map with four levels of impact for the hourly-basis forecast and 5-day forecast.

Hurricane Harvey (2017)

Hurricane Harvey, which struck Texas in late August 2017, was a catastrophic event. It intensified into a Category 4 hurricane and made landfall near Rockport, Texas. Harvey's slow movement brought record rainfall in Houston, with over 60 inches in some areas. The resulting floods caused extensive damage and significant displacement, amounting to over $125 billion in damages [23], [24]. We used historical streamflow observations, focusing on the peak streamflow on August 27, 2017, at 3:00 pm. The flood depth and SVI maps were combined to produce the impact map for Hurricane Harvey (Fig. 6), showcasing the integration of these elements to highlight areas of varying flood impacts.

These impact maps in both case studies illustrate how integrating flood forecast data (depth and flood forecast confidence) with social vulnerability indices can provide a more comprehensive view of flood risks, aiding in more effective emergency response and planning.

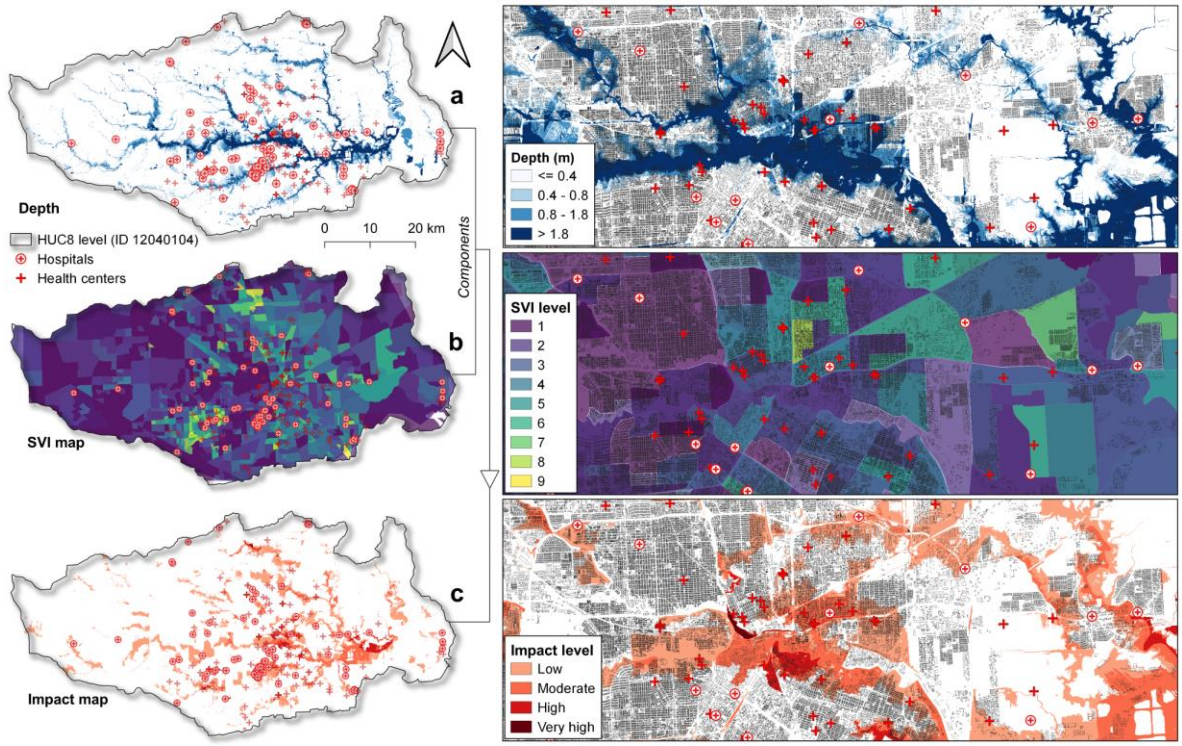


Figure 5. shows the output of Hurricane Harvey using historical streamflow observations at the peak time of the flood. a) shows the depth map with four classes of flood severity b) SVI map shows 9 levels of vulnerability (lowest =1, highest =9). c) the output impact map with four levels of impact for the peak time of the flood.

5.1. Limitations and future works

HAND-FIM is the model used by the NWM because it is low complexity and can be used for forecasting across larger spatial domains. However, it has several drawbacks which impact the accuracy of the impact forecasts. First, HAND-FIM only models fluvial flood events. It excludes coastal and compound flooding, so our impact forecasts likely underpredict areas of high impact during compound flood events. Secondly, HAND-FIM underpredicts smaller catchments (4th order and lower) while overpredicting large-order reaches [25]. The accuracy of our impact map is only as good as the accuracy of the NWM HAND-FIM. So our impacts also underpredict in small catchments and overpredict in large reaches.

It is hard to validate a flood forecast. There are efforts to do this for major historical floods, but they have their limitations. It is even harder to find a metric that can validate the human impact of a flood. The focus of our SVI is on a population's immediate risk during a flood event. We do not attempt to capture the long-term impacts of a community being flooded. There are some proxy datasets that can be used to validate the impact. For example, locations of 911 calls, insurance claims, and disaster relief funds, can all be used to approximate the social vulnerability to a particular flood. Future work could analyze these datasets for particular flood events and compare them to the forecasted impact.

Depth maps were generated using HAND-FIM [13]. The depths were calibrated against High Water Marks data from Hurricane Harvey and an alternative depth calculation model, FwDET v2.1 (Fig. 6). Despite the poor correlation between the two depth models and the HWMs, the outputs have similar trends to the HWM and both over-predict the depths. The FwDET v2.1 model performs slightly better compared to the observed high water marks but requires

significantly more computational time. When comparing the two model methods to the high-water marks based on depth classes, the two perform similarly.

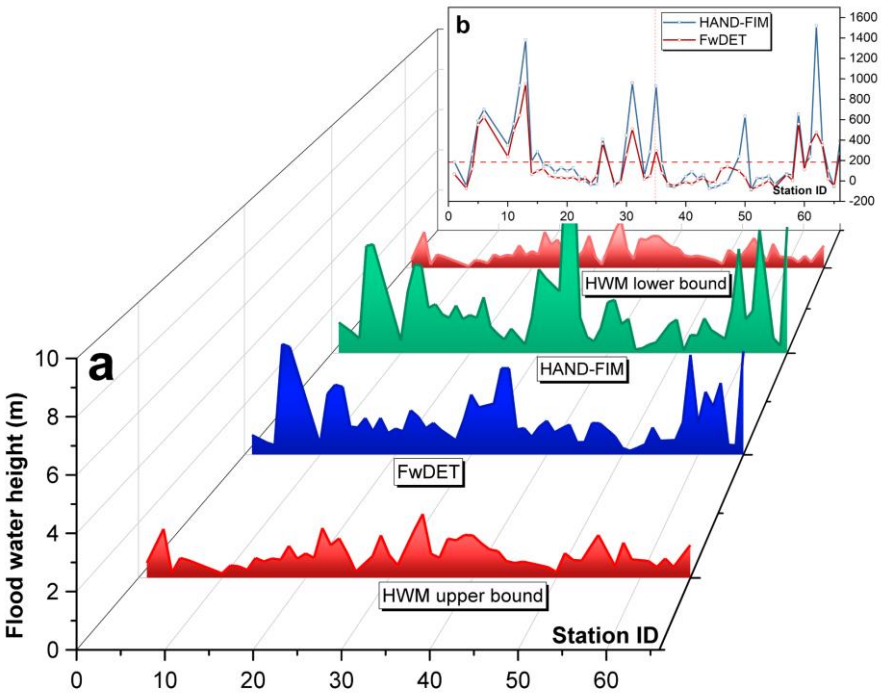


Figure 6. Comparison between depths in meters calculated using HAND-FIM and FwDET against the HWM, in which (a) the difference in water height (m) among HWM upper bound, lower bound, HAND-FIM, FwDET; (b) difference (in percentage) between HAND-FIM and FwDET.

6. Conclusion

This study developed an automated workflow to transition from hazard-based to impact-based flood forecasting. By using the National Water Model (NWM) streamflow predictions, we created Flood Inundation Maps (FIMs) with OWP HAND-FIM, estimated flood depths, and calculated social vulnerability index. Combined, these elements compose a framework for forecast impact maps. The maps can be published on an interactive web platform, providing accessible tools for communities and decision-makers to plan for forecasted floods. Our case studies on post-tropical cyclone Beryl and Hurricane Harvey demonstrated the system's effectiveness. Future work should address limitations such as the exclusion of coastal and compound flooding and explore ways to validate flood impact.

Acknowledgement:

Firstly, we want to thank our team leads, Dr. Kelsey McDonough and Dr. Sanjib Sharma for their guidance through this whole project. We acknowledge the support of Dr. Cohen Sagi, Dr. Anupal Baruah, and Dr. Dipsikha Devi from the Surface Dynamics Modeling Lab at the University of Alabama (UA) for their assistance during the HAND-FIM training. We also want to express our special thanks to CIROH, CUASHI, and the NOAA National Water Center for their great support during the National Water Center Innovators Program Summer Institute 2024.

Supplementary Materials:

The scripts for the framework proposed here, and the data used in the case studies can be found at *https://github.com/clynwatts/Impact_Map_Forecasting*

Instant App:

https://arcg.is/1jj8i12

References

1. A. Hu and I. Demir, "Real-time flood mapping on client-side web systems using hand model," Hydrology, vol. 8, no. 2, Jun. 2021, doi: 10.3390/hydrology8020065.
2. Md. S. Sikder, S. Ahmad, F. Hossain, A. S. Gebregiorgis, and H. Lee, "Case Study: Rapid Urban Inundation Forecasting Technique Based on Quantitative Precipitation Forecast for Houston and Harris County Flood Control District," J Hydrol Eng, vol. 24, no. 8, Aug. 2019, doi: 10.1061/(asce)he.1943-5584.0001807.
3. Tascón-González, Laura, Montserrat Ferrer-Julià, Maurici Ruiz, and Eduardo García-Meléndez. "Social vulnerability assessment for flood risk analysis." Water 12, no. 2 (2020): 558.
4. Tingsanchali, Tawatchai, and Thanasit Promping. "Comprehensive assessment of flood hazard, vulnerability, and flood risk at the household level in a municipality area: A case study of Nan Province, Thailand." Water 14, no. 2 (2022): 161.
5. Tate, E., Rahman, M.A., Emrich, C.T. and Sampson, C.C., 2021. Flood exposure and social vulnerability in the United States. Natural Hazards, 106(1), pp.435-457.
6. Dey, Hemal, Wanyun Shao, Md Munjurul Haque, and Matthew VanDyke. "Enhancing Flood Risk Analysis in Harris County: Integrating Flood Susceptibility and Social Vulnerability Mapping." Journal of Geovisualization and Spatial Analysis 8, no. 1 (2024): 19.
7. B. Cosgrove et al., "NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling," J Am Water Resour Assoc, vol. 60, no. 2, pp. 247–272, Apr. 2024, doi: 10.1111/1752-1688.13184.
8. NOAA, "The National Water Model." Accessed: Jul. 05, 2024. [Online]. Available: https://water.noaa.gov/about/nwm
9. P. D. Bates "Flood Inundation Prediction" in Annual Review of Fluid Mechanics, 2022. doi: https://doi.org/10.1146/annurev-fluid-030121-113138
10. J. Michael 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.
11. C. D. Rennó et al., "HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia," Remote Sens Environ, vol. 112, no. 9, pp. 3469–3481, Sep. 2008, doi: 10.1016/j.rse.2008.03.018.

12. M. Guidolin, A. S. Chen, B. Ghimire, E.C. Keedwell, S. Djordjević, D. A. Savić " A Weighted Cellular Automata 2D inundation Model for Rapid FLood Analysis" in Environmental Modeling and Software, October 2016. doi: https://doi.org/10.1016/j.envsoft.2016.07.008

13. Cohen, Sagy, Austin Raney, Dinuke Munasinghe, J. Derek Loftis, Andrew Molthan, Jordan Bell, Laura Rogers et al. "The Floodwater Depth Estimation Tool (FwDET v2.0) for improved remote sensing analysis of coastal flooding." Natural Hazards and Earth System Sciences 19, no. 9 (2019): 2053-2065.

14. Guidance for Flood Risk Analysis and Mapping Flood Depth and Analysis Grids 2018

15. ATSDR, "CDC/ATSDR Social Vulnerability Index (CDC/ATSDR SVI): Overview," Jun. 2024.

16. B. E. Flanagan, E. W. Gregory, E. J. Hallisey, J. L. Heitgerd, and B. Lewis, "A Social Vulnerability Index for Disaster Management," J Homel Secur Emerg Manag, vol. 8, no. 1, 2011, doi: 10.2202/1547-7355.1792.

17. A. Fekete, "Validation of a social vulnerability index in context to river-floods in Germany," 2009. [Online]. Available: www.nat-hazards-earth-syst-sci.net/9/393/2009/

18. Hinojos, Selena, Lauren McPhillips, Peter Stempel, and Caitlin Grady. "Social and environmental vulnerability to flooding: Investigating cross-scale hypotheses." Applied geography 157 (2023): 103017, doi: 10.1016/j.apgeog.2023.103017

19. C. Iadanza, A. Trigila, P. Starace, A. Dragoni, T. Biondo, and M. Roccisano, "IdroGEO: A collaborative web mapping application based on REST API services and open data on landslides and floods in Italy," ISPRS Int J Geoinf, vol. 10, no. 2, Feb. 2021, doi: 10.3390/ijgi10020089.

20. M. Wang, R. Lai, R. Xia, M. Wang, and M. Yang, "A Parallel Computation and Web Visualization Framework for Rapid Large-scale Flood Risk Mapping," in Journal of Physics: Conference Series, Institute of Physics Publishing, Aug. 2019. doi: 10.1088/1742-6596/1288/1/012065.

21. M. Makinano-Santillan et al., "Near-real time hazard monitoring and information dissemination through integration of remote sensing, GIS, numerical modelling, web applications and social media," in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Copernicus GmbH, Mar. 2019, pp. 25–32. doi: 10.5194/isprs-annals-IV-3-W1-25-2019.

22. National Hurricane Center. (2024). Post-Tropical Cyclone Beryl Report. Retrieved July 17, 2024 from https://www.nhc.noaa.gov/graphics_at2.shtml?start

23. National Hurricane Center. (2017). Hurricane Harvey Report. Retrieved July 17, 2024 from https://www.nhc.noaa.gov/graphics_at2.shtml?start

24. NOAA. 2017 Atlantic hurricane season. Retrieved July 10, 2024, from https://www.nhc.noaa.gov/data/tcr/index.php?season=2017&basin=atl

25. J.M. Johnson, D. Munasinghe, D. Eyelade, S. Cohen, "An integrated evaluation of the National Water Model (NWM)–Height Above Nearest Drainage (HAND) flood mapping

Chapter 7
Leveraging Geospatial Data and Machine Learning to Predict Insurance-derived Flood Damage Cost

Reihaneh Zarrabi[1], Corrine Liu[2], Aylar Samadi[1]

[1]University of Alabama; *rzarrabi@crimson.ua.edu, asamadiahari@crimson.ua.edu*

[2]University of Colorado, Boulder; *corrine.liu@colorado.edu*

Academic Advisors: Dr. Sagy Cohen, *The University of Alabama;* Dr. Shemin Ge, *University of Colorado, Boulder*; Dr. Hamid Moradkhani, *The University of Alabama*

Summer Institute Theme Advisors: Dr. Kelsey McDonough, FloodMapp, *kelsey@floodmapp.com,* Dr. Sanjib Sharma, Howard University, *sanjib.sharma@howard.edu*

Abstract:

Current flood research relies primarily on environmental parameters to communicate flood risk, often overlooking the complex reality in which socioeconomic identities and systemic social conditions inform flood risk and exposure. This study aims to leverage the recent advances in geospatial datasets and machine learning algorithms to predict flood damage. We employ insurance-derived, terrain and land cover, socioeconomic, and hydroclimatic data to predict building damage cost as a flood risk proxy during Hurricane Harvey using an eXtreme Gradient Boosting Regression (XGBR) algorithm. We selected building damage cost as a proxy for flood risk, as it can serve as a direct metric of flood impact while being an understandable parameter to a wide array of stakeholders and encompassing the social and economic effects of flooding. The XGBR model demonstrated excellent fit to the data while maintaining good predictive capability with R^2 of 0.74 and bias of 1.59 % in validation through testing. The model's predictions were closely aligned with the observed distribution, particularly the median, 25th, and 75th percentiles. Feature importance analysis demonstrated that the most important features that impacted subsequent building damage were recorded water depth, building property value, and percent of buildings with two floors. The calculated importance of the insurance-derived data for our algorithm's predictive capabilities emphasizes the interaction between flood damage, flood hazard, exposure, and vulnerability, suggesting that the implementation of such variables can be used for more robust and inclusive flood risk assessment and mitigation strategies.

Motivation

Floods pose devastating impacts on life and property. Flood mitigation strategies and emergency response rely on the ability to make sound predictions of flood impacts at decision-relevant scales. Conventional process-based flood models provide information focused primarily on hazards, such as flood depth and inundation extent. However, the complex non-linear interactions among hazard, exposure, and vulnerability shape the overall flood risk. Neglecting exposure and vulnerability in flood risk assessment can lead to suboptimal decisions about risk management.

In disaster events, the socially vulnerable are less likely to recover and more likely to die, and thus, effectively addressing social vulnerability can decrease both human suffering and economic loss due to disasters (Flanagan et al. 2011). Advances in the study of flooding that consider socioeconomic parameters can facilitate more comprehensive risk mitigation policies and clearer risk communication (Aerts et al., 2018). Traditional flood risk products, such as flood inundation maps (FIMs) or high-water marks, can be difficult to interpret for the public, thus developing alternative, robust methods for assessing and communicating risk is essential (Bales & Wagner, 2008). In this study, mean building damage cost over a zip code is the target variable identified as a proxy for flood risk. We believe that using mean building damage cost as a flood risk proxy can allow for a greater understanding of flood risk that does not require in-depth knowledge of traditional flood hazard assessment standards or hydrologic principles.

Objectives and Scope

This study demonstrates a novel approach to propose a flood risk proxy based on predicting average building damage cost at a zip code scale using Machine Learning (ML) techniques. This research leverages large datasets, including insurance claims data from the National Flood Insurance Program (NFIP), which is recognized as the foremost and most comprehensive flood insurance program in the United States today (Blessing et al., 2017). The method is applied to areas affected by Hurricane Harvey, a category 4 hurricane that devastated economically disadvantaged communities in Texas and Louisiana in 2017, resulting in $158 billion in damages and displacing over 30,000 individuals (NOAA, 2024). Through rigorous data preprocessing, advanced XGB Regression (XGBR) modeling techniques, and thorough performance assessments, this research aims to offer an alternative method of predicting flood impacts to enhance flood resilience strategies and mitigate future flood impacts across the United States.

Literature Review

The increase in severe flooding is exacerbated by rapid urban development in flood-prone areas. Urban expansion, infrastructure growth, and economic activities in vulnerable regions heighten the risk and impact of flood-related damage (Park et al., 2021). While conventional flood inundation mapping and flood risk assessments focus on flood depth and extent, there is growing recognition of the need for methodologies assessing flood-related losses or impacts (Elmer et al., 2010). Evaluating the efficacy of mitigation measures necessitates a holistic understanding of both physical and non-physical factors influencing damage to floodplain elements (Scorzini & Frank, 2017).

Preliminary damage assessments are initial assessments of damage after an emergency event done at the local and state level to assist the Federal Emergency Management Agency (FEMA) and the federal government in determining damage and allocating resources during emergency response (FEMA, 2021). Assessing building damage, such as through preliminary damage assessments, is critically important for several reasons. Buildings are central to the functionality and stability of communities, serving as homes, businesses, and infrastructure. Additionally, understanding building damage helps identify vulnerable areas and structures, guiding future urban planning, infrastructure improvements, and flood risk mitigation strategies. By focusing on building damage, communities can enhance their resilience to floods, minimize losses, and

improve recovery outcomes after such natural disasters (Urban Floods Community of Practice, 2017).

ML has been increasingly recognized as a high-power tool for handling large spatial datasets and for its improved predictive capabilities for natural hazards compared to other methods (Nachappa et al., 2020, Mosavi et al., 2018). Tree-based ML algorithms, such as eXtreme Gradient Boosting (XGB), demonstrate significant potential in improving the accuracy of flood risk predictions. XGB has been used widely for assessing pluvial flooding risk and extent in a wide variety of contexts (Wang et al., 2023, Liu et al., 2022, Ghosh et al., 2022, Ren et al., 2024). XGB has also been used in conjunction with non-traditional flood data, such as Waze alerts, to predict flash flooding susceptibility (Safaei-Moghadam et al., 2024) and to optimize urban configuration as a flood mitigation strategy (Yuan et al., 2024, Zhou et al., 2022).

To our knowledge, our methodology is the first to propose the implementation of hydrological, topographic, socioeconomic, and insurance claims data as input variables for an XGBR algorithm to predict mean building damage cost (in USD) on a zip code scale. Current literature underscores the importance of considering the social dimension of natural disaster occurrence and mitigation, and the predictive efficacy of XGBR algorithms (Blessing et al., 2017, Flanagan et al., 2017, Ren et al., 2024) . Therefore, we aim to demonstrate how publicly available insurance claims data can be used to predict building damage, which can be used as a robust proxy to predict flood risk and exposure.

Methodology

Data Collection

The Federal Insurance and Mitigation Administration (FIMA) National Flood Insurance Program (NFIP) Redacted Claims version 2 dataset is provided by the (FEMA) to help reduce future flood losses through flood hazard identification, floodplain management, and the provision of insurance protection. This dataset contains over 2.6 million records of claims made under the NFIP, removing identifying information to protect privacy and covering the contiguous United States (CONUS). This dataset includes the assessed building damage dollar amount for a property, and damage is classified as tangible and direct, resulting from the building's response to flood hazard factors such as flood depth (Merz et al., 2010, de Moel et al., 2015). As NFIP data includes corresponding zip codes for each claim, using a zip code scale represents a convenient and representative scale for our project. We calculated the mean damage amount for buildings within each zip code (Fig S1), which served as our target variable. Our model's independent (predictive) variables are categorized into the following four main classes, discussed below.

Insurance-Derived Characteristics

The NFIP Redacted Claims version 2 dataset includes various characteristics related to insured buildings and the recorded flood situations for which insurance claims were filed. This dataset contains three types of data: binary (e.g., whether the building is elevated or not), categorical (e.g., different categories for the number of building stories/floors), and continuous values (e.g., building price value). To manage dataset variation, we calculated the percentage for binary and categorical variables within each zip code, while averaging the continuous value over zip code (Fig S1).

Terrain and Land Cover Characteristics

Terrain characteristics were calculated directly from 10-meter resolution digital elevation models (DEMs) from the USGS 3D Elevation Program (3DEP). These include surface slope, distance to the nearest water body, topographic wetness index (TWI), Soil Drainage Index (SDI), and Stream Power Index (SPI). These parameters were chosen based on previous studies identifying them as key factors impacting flooding (Meraj et al., 2015; Dey et al. 2024). TWI represents soil wetness based on topography (Kirkby, 1975), SDI reflects long-term soil wetness and soil drainage capacity (Hole and Campbell, 1986; Schaetzl, 1986), and SPI measures the erosive power of flowing water (Bagnold, 1966). Land cover characteristics were derived from the National Land Cover Database 2016 (NLCD 2016). Previous studies by Amini et al. (2011), Fox et al. (2012), and Zope et al. (2016) have examined the impact of LULC changes on flooding, demonstrating how increased urbanization can promote flooding conditions. The dataset was categorized into agriculture, developed, forest, water, and other categories. The percentage area covered by each category relative to the total zip code area was calculated for each zip code (Fig S1).

Hydro-Climatic Characteristics

The hydro-climatic characteristic of interest in this study was precipitation, chosen for its primary influence on pluvial flooding and its significant spatial and temporal variability. Data was sourced from the North American Land Data Assimilation System (NLDAS) dataset, known for its high-resolution 1/8th-degree grid spacing and continuous coverage since 1979. The dataset provides hourly total precipitation data in kg/m^2, recognized for its precision in hydrological applications. Historical data for Hurricane Harvey spanning August 24 to September 19, 2017, was extracted and subsequent analysis included deriving mean and maximum values for accumulated precipitation over 1-, 3-, 5-, and 7-day periods, as well as for the entire event duration (Fig S1).

Socioeconomic Characteristics

The Social Vulnerability Index (SoVI) identifies flood-prone communities based on socioeconomic factors, aiding in prioritizing resources for flood mitigation and response efforts (Cutter et al., 2003). In this study, the Social Vulnerability Index (SoVI) is constructed following the algorithm proposed by Cutter et al. (2003). Key variables representing social vulnerability—such as population, housing, age, gender, wealth, race, language barriers, transportation, education, and disability—were extracted from the 2017 American Community Survey (ACS) five-year estimates available from the US Census Bureau at the zip code level (refer to Table S1). A Kaiser–Meyer–Olkin (KMO) test was then employed to assess the data's suitability for Principal Component Analysis (PCA), a statistical technique used to simplify and reduce the dimensionality of large datasets while preserving essential information (Abson et al., 2012). The data were normalized using Z-scores before Principal Component Analysis (PCA) was performed. PCA for each component of the SoVI was performed using IBM SPSS Statistics software version 29.0.2.0. The contribution of each component was assessed based on the proportion of variance explained by the selected components (for further details, see Dey et al., 2023). Equation 1 represents the obtained SoVI equation, and Table S2 summarizes the social vulnerability component summary. Using Equation 1, the SoVI is calculated for each zip code (Fig S1).

$$SoVI = 0.541F_1 + 0.193F_2 - 0.151F_3 + 0.115F_4 \quad (1)$$

Where F_1, F_2, F_3, and F_4 are the components calculated through PCA.

Data Preprocessing

After gathering and extracting the specified characteristics, they were merged using the zip code number as a common identifier across all provided datasets, resulting in an input dataset with 554 data points. Data points related to zip codes with no reported building damage or those with misreported information were removed. Additionally, zip codes that were sparsely distributed (see Fig 1) or considered spatial outliers were excluded. Errors in recorded water depth from the NFIP Redacted Claim version 2 were identified, and the related data points were excluded. Data points with NAN values in any features were also removed. After this preprocessing, the dataset was reduced to 368 data points (see Table S3 for dataset description, units of measurement, and statistical summary; Fig S2 shows the correlation matrix between attributes). The input dataset was then split into training and testing sets in a 75:25% ratio, resulting in 276 training points and 92 testing points. Fig 1 illustrates the spatial distribution of the training and testing datasets.

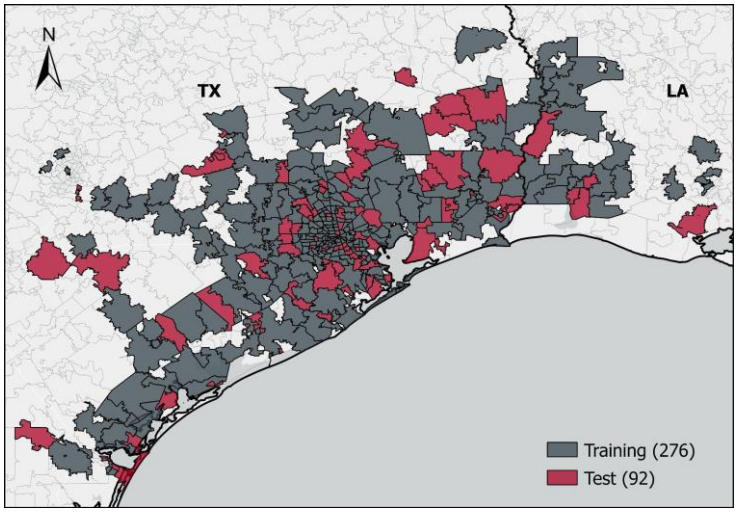


Fig 1. Spatial distribution of the training and testing datasets.

Model Development

XGB Regression (XGBR) models were developed using the input dataset. XGBR, a supervised algorithm within the gradient boosting framework, is known for its robustness, noise resistance, and high prediction accuracy (Ren et al., 2024; Liu et al., 2022). It iteratively improves predictions based on previous errors, offering increased speed and accuracy with the ability to handle correlated features while mitigating model complexity and overfitting (Zhou et al., 2022; Yuan et al., 2024).

Using this algorithm, feature importance analysis based on a decrease in impurity was conducted to identify the most significant features. This analysis evaluates each feature's contribution to reducing error and increasing accuracy. The model was then validated against the testing dataset using performance metrics such as coefficient of determination (R^2), Percent Bias (PBias), and Root Mean Square Error (RMSE). The process of fine-tuning XGBR hyperparameters continued until a satisfactory level of accuracy was achieved.

Results and Discussion

The XGBR model demonstrated strong performance on the training dataset, achieving a high R^2 value of 0.998, indicating an excellent fit to the data (Table 1). On the testing dataset, the model maintained good predictive capability with an R^2 value of 0.742. Both the training (0.092%) and testing (1.586%) datasets exhibited minimal Percent Bias (PBias), suggesting low bias in predictions. However, there was a notable difference in Root Mean Square Error (RMSE) between the training set ($2575.042) and the testing set ($21037.271), indicating variance in prediction accuracy across datasets.

Table 1. Performance metrics of the developed model using eXtreme Gradient Boosting Regression (XGBR) algorithms on the test dataset to estimate mean building damage amount ($).

Performance Metric	Training	Testing
R^2	0.998	0.742
PBias (%)	0.092	1.586
RMSE ($)	2575.042	21037.271

The violin plots of both datasets show that the model's predictions closely align with the distribution of the data in terms of median, 25th, and 75th percentiles (Fig 2). However, the model exhibits less accuracy in predicting minimum and maximum values. Scatter plots for both datasets indicate a generally good fit across the data range, although there is a tendency to overestimate at lower values. This discrepancy may be due to the limited number of data points at the extremes, potentially hindering the model's ability to fully capture trends and patterns in those regions.

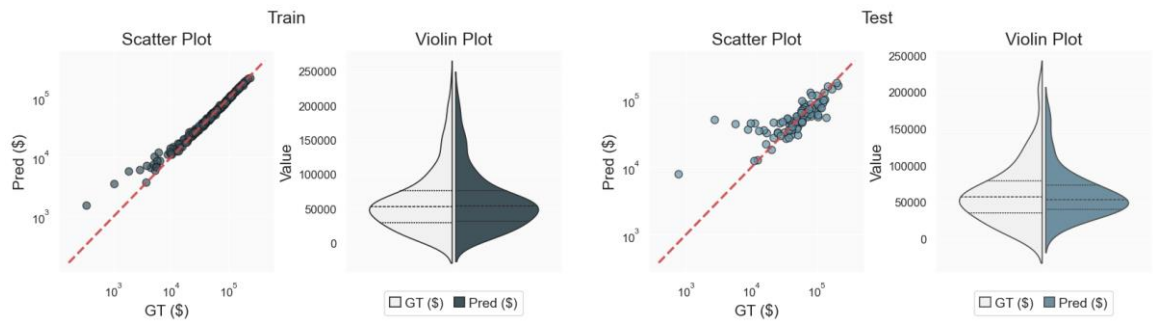


Fig 2. Scatter and violin plots of observations against predictions obtained by the developed model using the eXtreme Gradient Boosting Regression (XGBR) algorithm on the test dataset to estimate the mean building damage amount ($).

Feature importance analysis was performed through the mean decrease in impurity (MDI) (Fig 3). Recorded water depth in buildings, contributing about 22% to the model's predictive power, highlights its crucial role in determining flood damage. This aligns with depth-damage functions (Pistrika et al., 2014b), which quantitatively relate flood water depth to structural or asset damage.

The value of the building price is closely followed, accounting for nearly 16% of the feature importance. This suggests that insurance coverage for more expensive houses may be more comprehensive due to higher repair expenses and more valuable contents, potentially resulting in higher assessed damages. Also, individuals with more expensive houses may possess a better understanding of "good" insurance, ensuring they are adequately covered. They are also more likely to afford insurance compared to those of lower socioeconomic status. Additionally, the percentage of buildings with 2 floors/stories, the third most important factor, showed significant importance, contributing approximately 12%. This highlights the vulnerability of multi-story buildings to flood damage, where each additional floor increases exposure to potential flooding impacts.

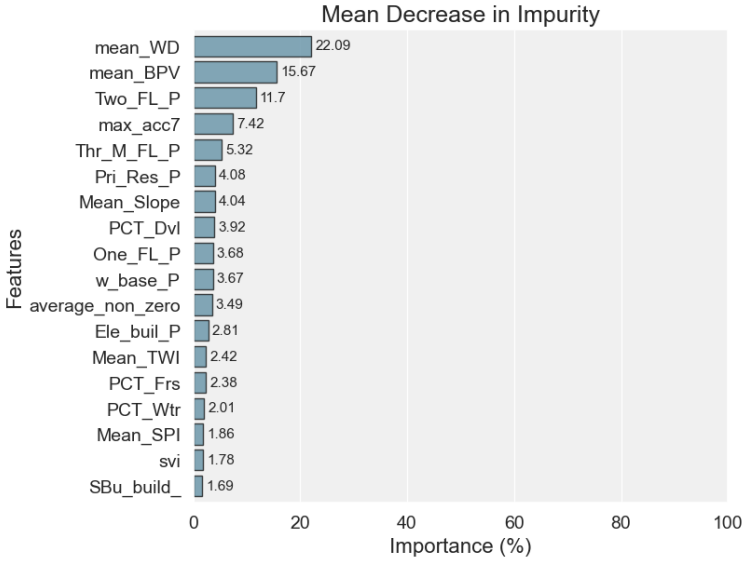


Fig 3. Results of the Mean Decrease in Impurity (MDI) analyses were obtained by applying the developed model using the eXtreme Gradient Boosting Regression (XGBR) algorithm on the test dataset to estimate the mean building damage amount ($). (Check Table S3 for attribute description)

Future Works

As mentioned, the model has limitations in capturing the distribution at both the minimum and maximum ends. This issue arises from the need for more data points in these ranges for training the model. Future work could expand the model spatially and temporally by considering more events, thereby providing a more comprehensive dataset for model development. Exploring other machine learning algorithms, such as Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP), could help develop models that better capture the extremes. These models could also be ensembled with the current XGBR model to characterize model structural uncertainty and enhance overall prediction accuracy.

Due to limitations in the NFIP Claim Redacted version 2 dataset, the model was developed on a zip code scale. However, this broader scale might result in inaccurate damage predictions by neglecting variations within incorporating codes. More precise data could enhance the model's precision assessments of flood risk and structural damage. To overcome these constraints and improve the credibility of flood risk predictions, future efforts should concentrate on integrating data with higher resolutions.

Although the model demonstrates good accuracy, its applicability is limited by its reliance on data derived from the NFIP Claim Redacted version 2, which only includes zip codes with filed insurance claims, leaving many zip codes unrepresented. Future work could focus on developing models that do not rely on the NFIP dataset or replacing it with other similar and accessible features. This approach could expand the model's applicability to a broader range of zip codes.

The developed model can be applied to other flood events and independently evaluated to ensure its robustness as a predictive tool for future occurrences. Its reliability can be quantitatively assessed by comparing its outcomes with Federal Emergency Management Agency (FEMA) 500—and 1,000-year flood maps and high-water marks.

Conclusion

This study aimed to develop predictive models for estimating building damage caused by Hurricane Harvey at a zip code scale, utilizing XGB Regression (XGBR) and integrating insurance-derived, socioeconomic, terrain-land cover, and hydro-climate characteristics. Key findings emphasized the significant influence of recorded water depth, building price value, and the percentage of buildings with 2 floors/stories on damage prediction, highlighting their direct correlation with flood hazard, exposure, and vulnerability.

XGBR emerged as a robust algorithm due to its high predictive accuracy, ability to handle non-linear relationships, and minimal dependency on extensive data preprocessing. However, the model's applicability is limited by its reliance on NFIP dataset features, restricting coverage to zip codes with available insurance claims data. Future research directions should focus on enhancing model generalizability by incorporating alternative datasets and methodologies. This approach aims to extend the model's utility across diverse geographic regions and various flood events, supporting more comprehensive flood risk assessments and enabling targeted mitigation strategies to bolster community resilience against natural disasters.

Acknowledgments

We would like to thank Hemal Dey, Sadra Seyvani, Dr. Wanyun Shao, and the Summer Institute team for their input and guidance, which was integral to our progress with this project.

Supplementary Materials

Codes and script are available at https://github.com/Reizrb/SI2024

Table S1. Description of attributes considered to assess social vulnerability

Attribute	Description
POP_65_P	Percentage of population over 65 years old (%)
POP_5_P	Percentage of population under 5 years old (%)
Fem_POP_P	Percentage of total female population (%)
Med_Hou_inc	Median households' income (%)

Attribute	Description
Hou_no_Veh	Number of households with no available vehicle
POP_no_SC	Percentage of population with no schooling completed (%)
DisAble	Number of disabled populations
Hisp_P	Percentage of Hispanic households (%)
White_P	Percentage of White households (%)
Spe_lv_P	Percentage of people with speak English less than very well
POP_Den	Population density (person/sqkm)
Housing_Den	Housing density (household/sqkm)

Table S2. Social vulnerability component summary

Factor	Cardinality	Name	Variance explained (%)	Dominant variables	Factor loading
F1	+	Race and Language Barrier	40.636	Hisp_P	0.868
				White_P	-0.804
				Spe_lv_P	0.846
F2	+	Population and Housing	14.533	POP_Den	0.896
				Housing_Den	0.936
F3	-	Wealth	11.324	Med_Hou_inc	0.798
F4	+	Gender	8.640	Fem_POP_P	0.839

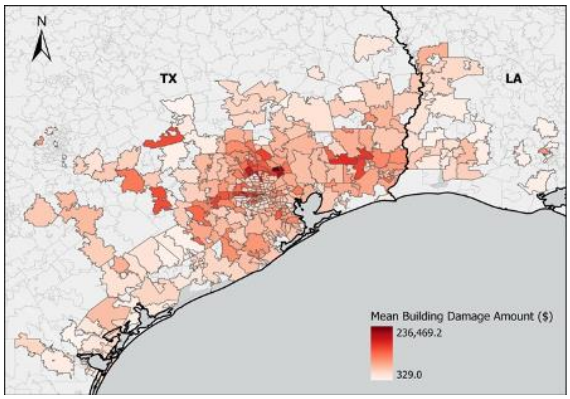
Table S3. Dataset description, units of measurement, and some descriptive statistics summary. The attributes indicated by * are those involved in the final best-developed model.

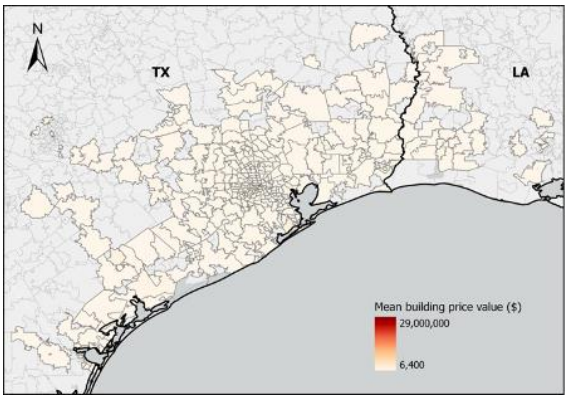
	Attribute	Description	Mean	Min	Max	StdDev
*	mean_BDA	Mean Building Damage Amount ($)	62938.96	767.00	236469.21	44019.45

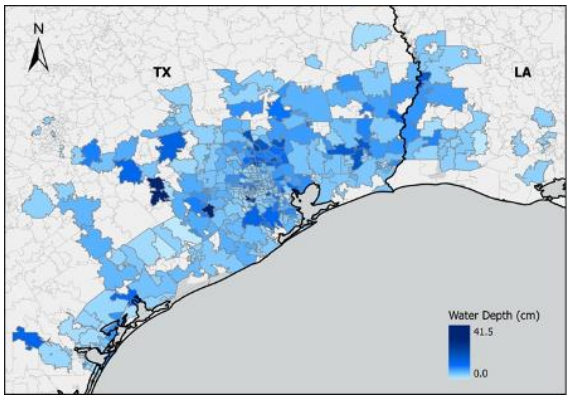
	Attribute	Description	Mean	Min	Max	StdDev
*	mean_BPV	Mean Building Price Value ($)	489455.95	6378.00	28956518.04	176299 5.20
*	mean_WD	Mean Recorded Water Depth (cm)	5.65	0.00	41.55	4.90
	Agr_stru_P	Percentage of Agricultural Buildings (%)	0.05	0.00	7.69	0.48
	Wor_Hou_P	Percentage of Worship Houses (%)	0.79	0.00	33.33	2.84
	NP_buil_P	Percentage of Non-Profit Buildings (%)	0.42	0.00	14.29	1.46
*	SBu_build_	Percentage of Small Businesses (%)	1.77	0.00	100.00	6.37
*	Pri_Res_P	Percentage of Buildings with Primary Residency (%)	78.46	0.00	100.00	17.91
*	Thr_M_FL_P	Percentage of Buildings with 3 or more Floors (%)	3.15	0.00	100.00	10.47
	St_Own_P	Percentage of State-Owned Buildings (%)	0.02	0.00	2.08	0.13
*	Ele_buil_P	Percentage of Elevated Buildings (%)	17.83	0.00	100.00	23.42
*	w_base_P	Percentage of Buildings with Basement (%)	61.31	0.00	100.00	19.91
*	One_FL_P	Percentage of Buildings with One Floor (%)	65.05	0.00	100.00	26.26
*	Two_FL_P	Percentage of Buildings with Two Floors (%)	25.75	0.00	100.00	21.98
	Res_buil_P	Percentage of Residential Buildings (%)	92.63	0.00	100.00	12.62
	PCT_Agr	Percentage of Agricultural Land Cover (%)	20.21	0.00	86.72	25.21

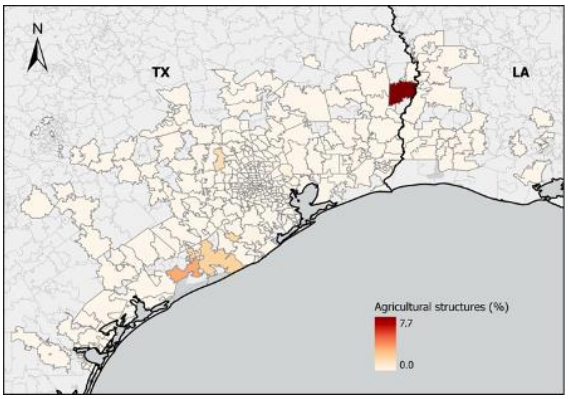
	Attribute	Description	Mean	Min	Max	StdDev
*	PCT_Dvl	Percentage of Developed Land Cover (%)	47.66	1.91	100.00	37.11
*	PCT_Frs	Percentage of Forest Land Cover (%)	12.05	0.00	60.39	15.81
*	PCT_Wtr	Percentage of Water Land Cover (%)	3.17	0.00	41.59	5.73
	PCT_Othr	Percentage of Others Land Cover (%)	16.91	0.00	76.93	15.94
	mean_acc1	1-day Mean Accumulated Precipitation (kg/sqkm)	27.96	2.12	42.99	10.43
	max_acc1	1-day Max Accumulated Precipitation (kg/sqkm)	254.93	23.05	396.22	104.14
	mean_acc3	3-day Mean Accumulated Precipitation (kg/sqkm)	90.71	5.88	140.52	34.91
	max_acc3	3-day Max Accumulated Precipitation (kg/sqkm)	507.89	42.28	729.52	192.09
	mean_acc5	5-day Mean Accumulated Precipitation (kg/sqkm)	143.09	6.71	242.08	58.65
	max_acc5	5-day Max Accumulated Precipitation (kg/sqkm)	605.43	42.97	921.50	229.47
	mean_acc7	7-day Mean Accumulated Precipitation (kg/sqkm)	171.59	7.62	320.73	73.46
*	max_acc7	7-day Max Accumulated Precipitation (kg/sqkm)	612.22	43.23	944.15	229.66
	mean_pcp	Mean Precipitation during Event (kg/sqkm)	1.22	0.09	1.87	0.45
	tot_pcp	Total Precipitation during Event (kg/sqkm)	615.04	46.71	945.80	229.48
	max_pcp	Max Precipitation during Event (kg/sqkm)	32.73	2.90	66.63	14.86

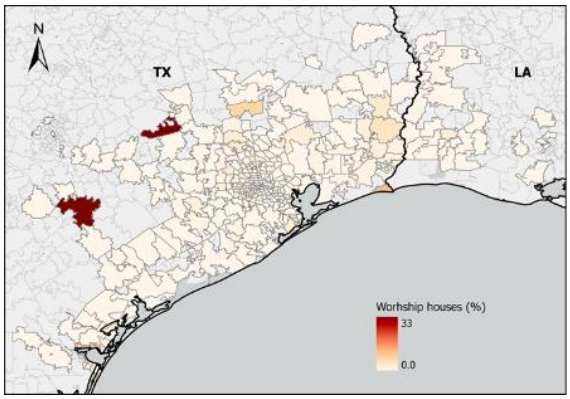
	Attribute	Description	Mean	Min	Max	StdDev
	non_zero_hours	During of the Event where Precipitation is not Zero (hrs)	148.57	87.00	204.00	26.17
*	average_non_zero	Mean Precipitation during non_zero_hours (kg/sqkm)	4.18	0.36	6.30	1.51
*	Mean_Slope	Mean Slope (Degree)	1.05	0.13	7.37	0.72
	Max_Slope	Max Slope (Degree)	24.44	10.11	64.40	7.59
*	Mean_TWI	Mean Topographic Wetness Index	3.46	2.16	6.64	0.80
	NEAR_DIST	Distance Between Zip Code Centroid and Nearest Stream River (m)	1142.99	1.70	13099.22	1245.84
*	Mean_SPI	Mean Stream Power Index	-5.86	-7.78	-2.70	0.89
	Mean_SDI	Mean Soil Drainage Index	28.95	0.00	62.05	9.71
*	svi	Social Vulnerability Index	0.05	-1.33	2.34	0.63

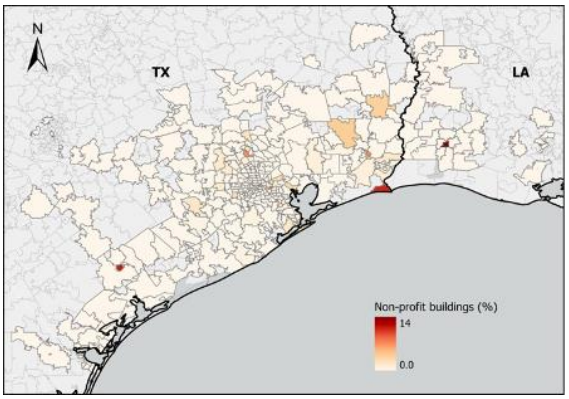


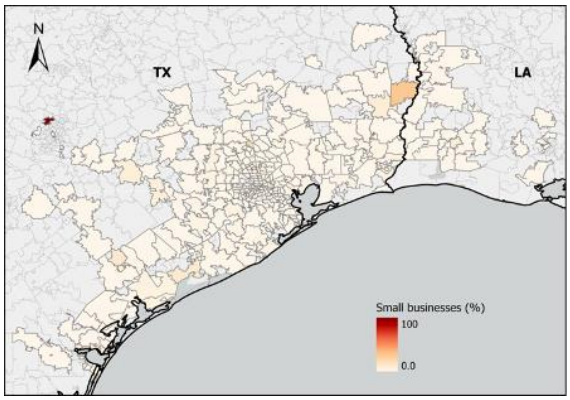


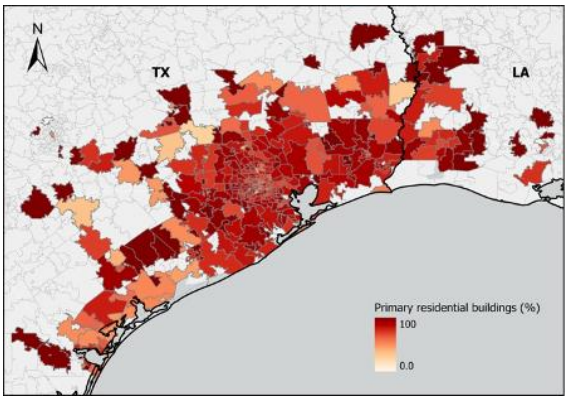

N
TX
LA
Water Depth (cm)
41.5
0.0

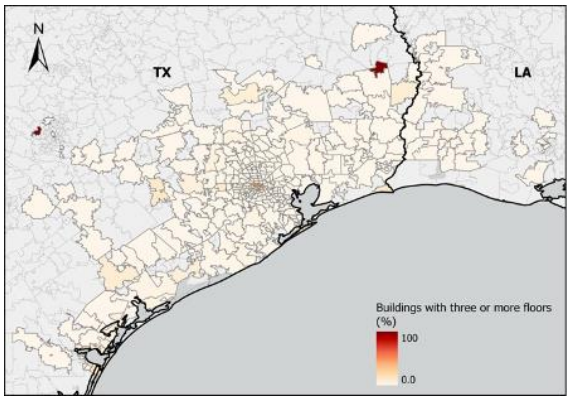

N
TX
LA
Agricultural structures (%)
7.7
0.0

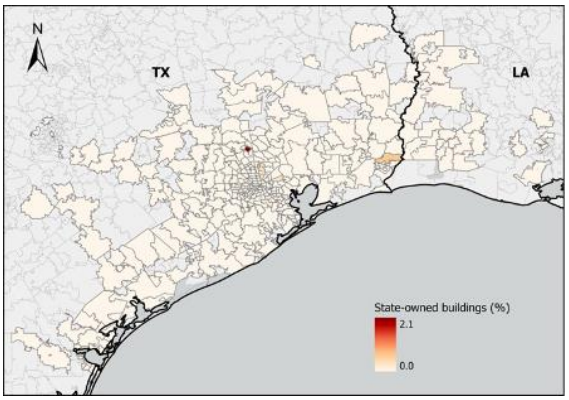

N
TX
LA
Worhship houses (%)
33
0.0

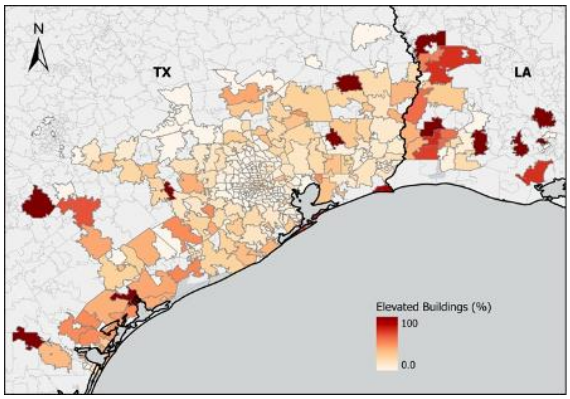

N
TX
LA
Non-profit buildings (%)
14
0.0

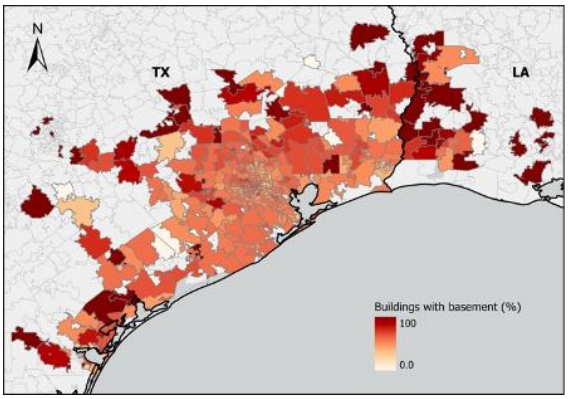

N
TX
LA
Small businesses (%)
100
0.0

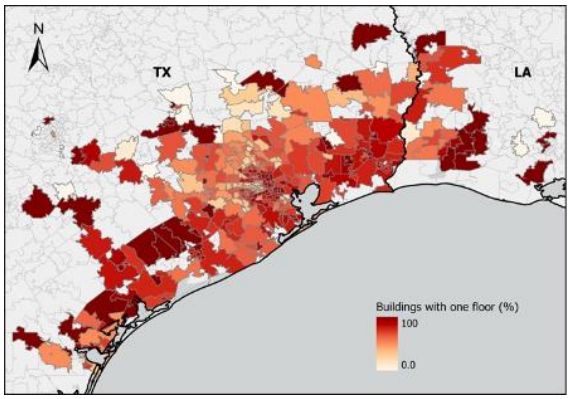

N
TX
LA
Primary residential buildings (%)
100
0.0

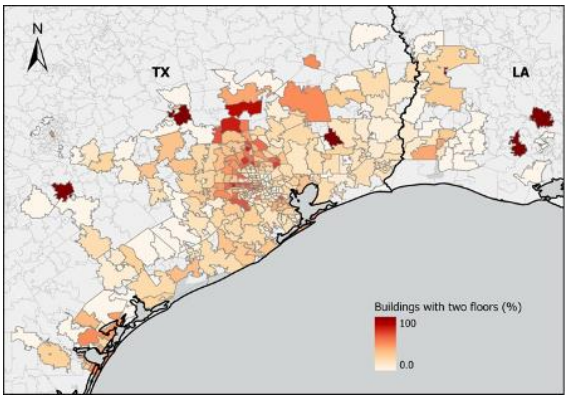

N
TX
LA
Buildings with three or more floors (%)
100
0.0

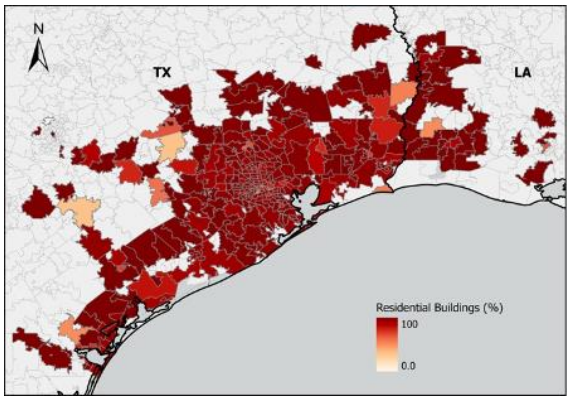

N
TX
LA
State-owned buildings (%)
2.1
0.0

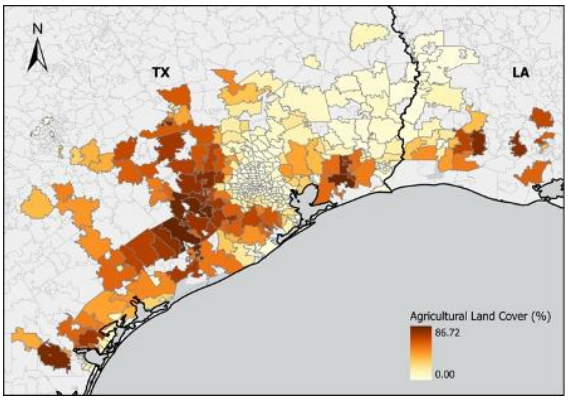

N
TX
LA
Elevated Buildings (%)
100
0.0

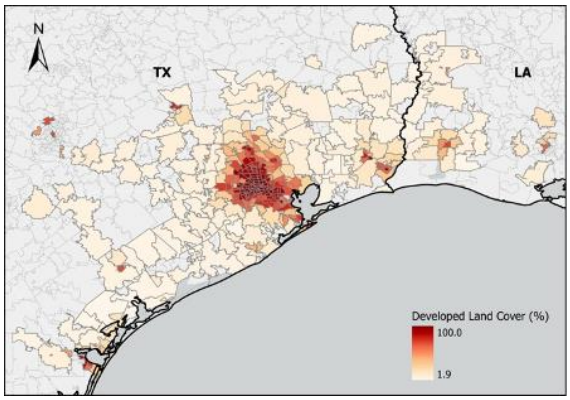

N
TX
LA
Buildings with basement (%)
100
0.0

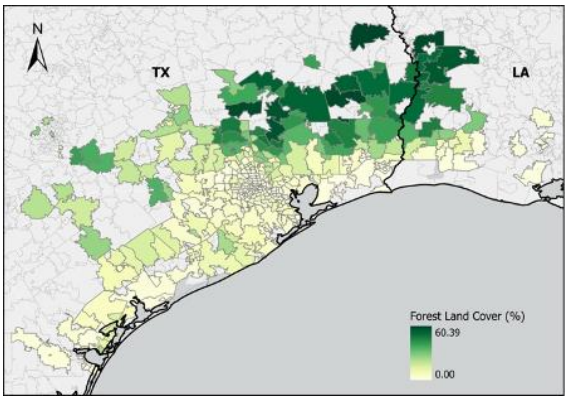

N
TX
LA
Buildings with one floor (%)
100
0.0

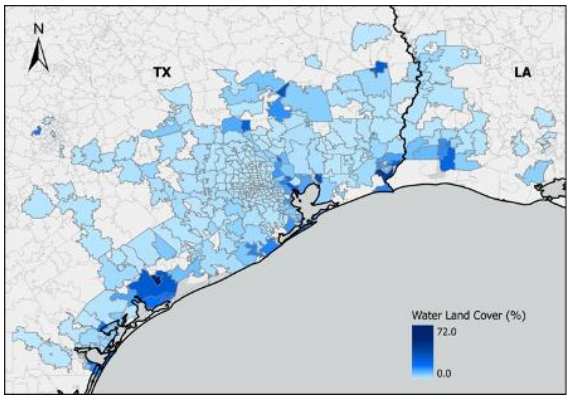

N
TX
LA
Buildings with two floors (%)
100
0.0

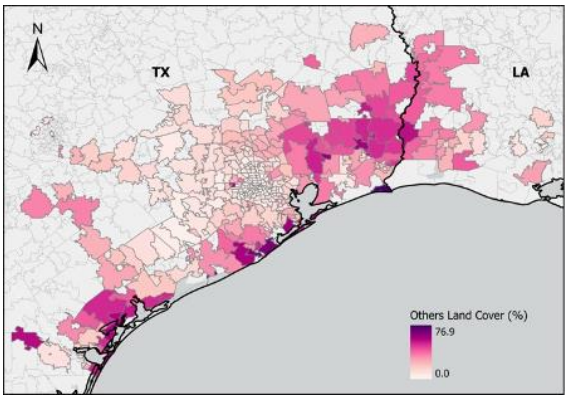

N
TX
LA
Residential Buildings (%)
100
0.0

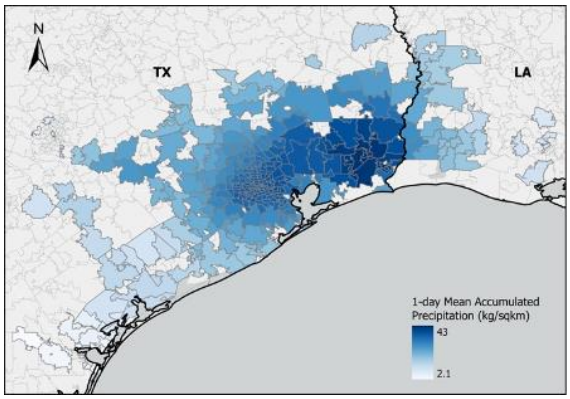

N
TX
LA
Agricultural Land Cover (%)
86.72
0.00

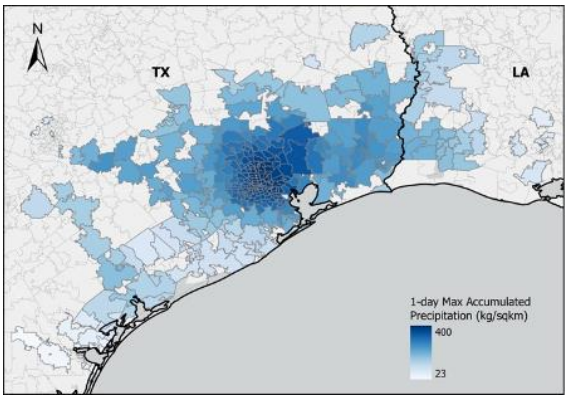

N
TX
LA
Developed Land Cover (%)
100.0
1.9

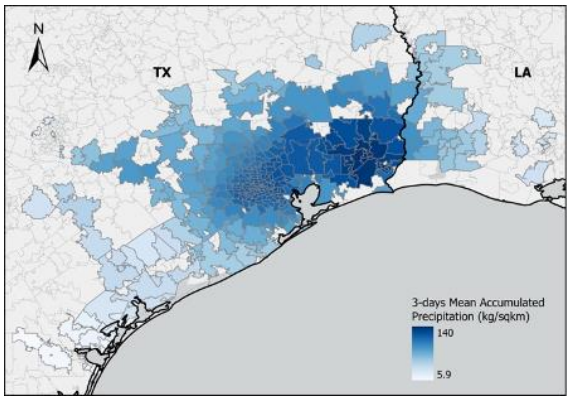

N
TX
LA
Forest Land Cover (%)
60.39
0.00

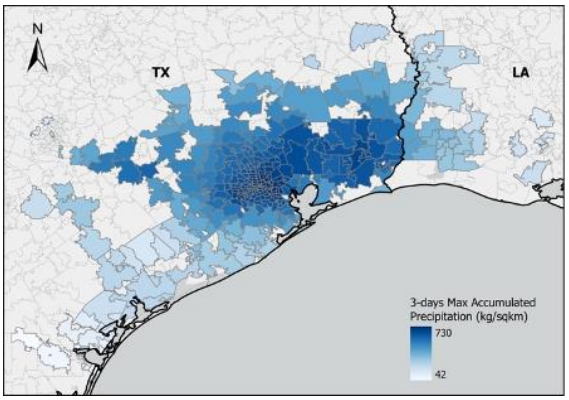

N
TX
LA
Water Land Cover (%)
72.0
0.0

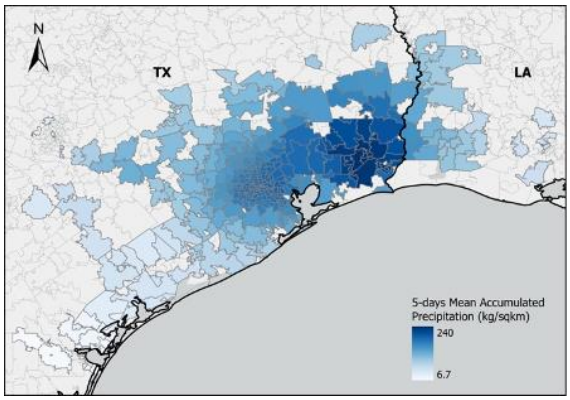

N
TX
LA
Others Land Cover (%)
76.9
0.0

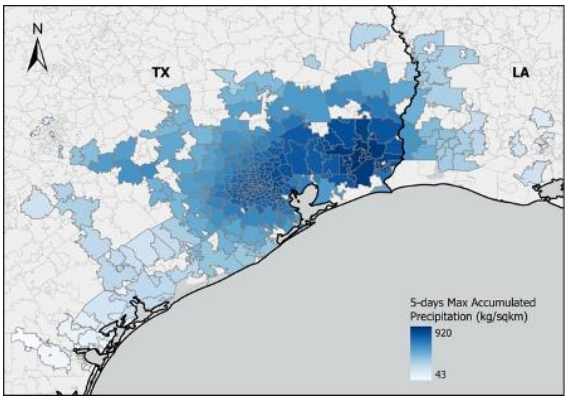

N
TX
LA
1-day Mean Accumulated
Precipitation (kg/sqkm)
43
2.1

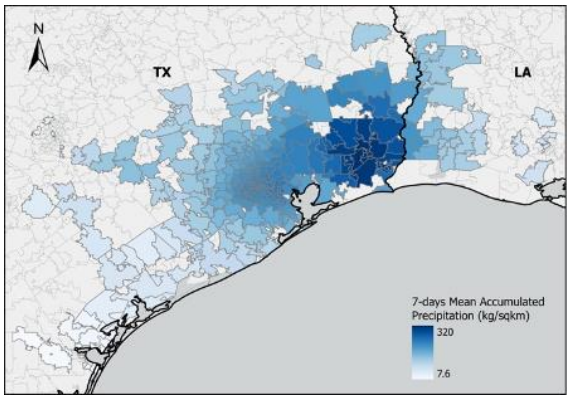

N
TX
LA
1-day Max Accumulated
Precipitation (kg/sqkm)
400
23

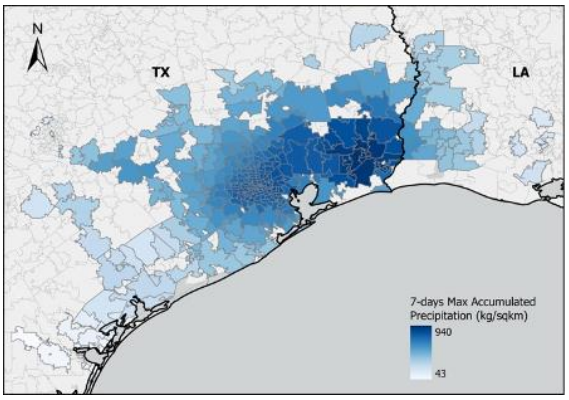

N
TX
LA
3-days Mean Accumulated
Precipitation (kg/sqkm)
140
5.9

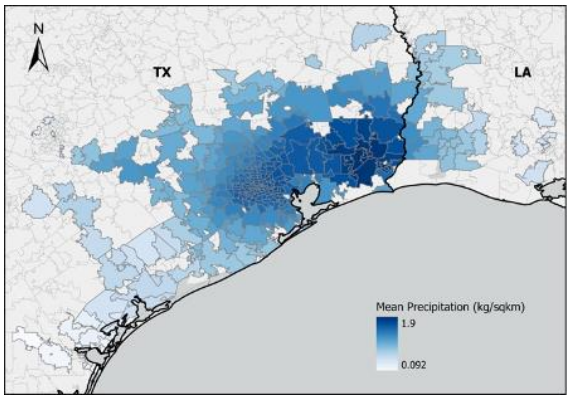

N
TX
LA
3-days Max Accumulated
Precipitation (kg/sqkm)
730
42

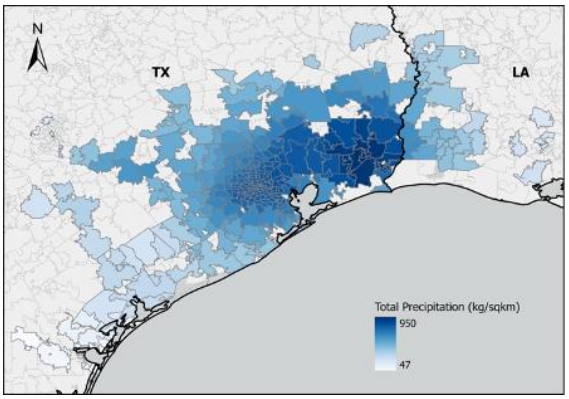

N
TX
LA
5-days Mean Accumulated
Precipitation (kg/sqkm)
240
6.7

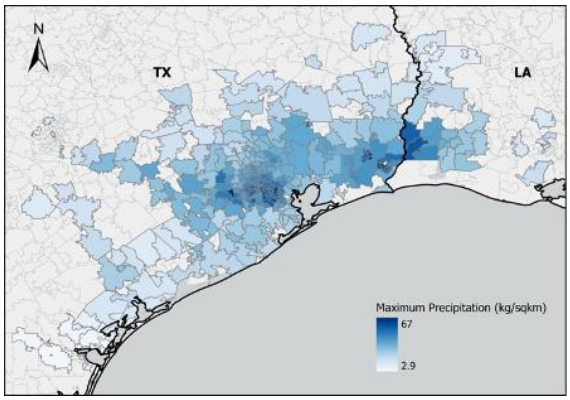

N
TX
LA
5-days Max Accumulated
Precipitation (kg/sqkm)
920
43

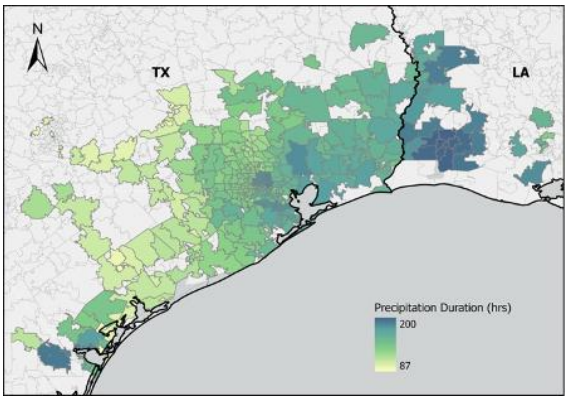

N
TX
LA
7-days Mean Accumulated Precipitation (kg/sqkm)
320
7.6

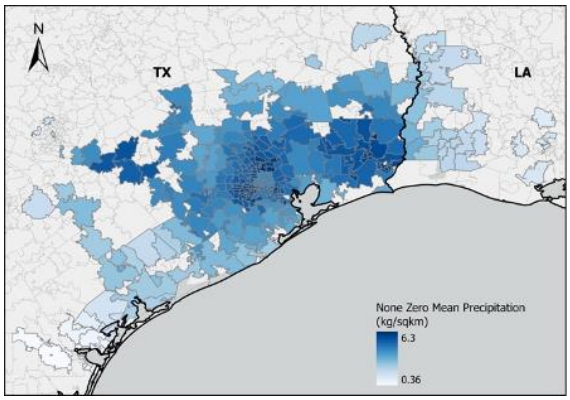

N
TX
LA
7-days Max Accumulated Precipitation (kg/sqkm)
940
43

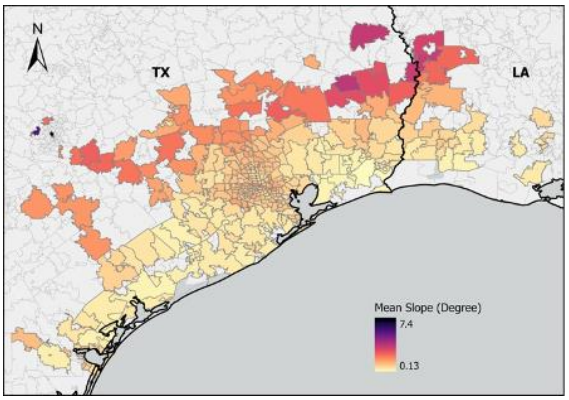

N
TX
LA
Mean Precipitation (kg/sqkm)
1.9
0.092


N
TX
LA
Total Precipitation (kg/sqkm)
950
47


N
TX
LA
Maximum Precipitation (kg/sqkm)
67
2.9


N
TX
LA
Precipitation Duration (hrs)
200
87


N
TX
LA
None Zero Mean Precipitation (kg/sqkm)
6.3
0.36


N
TX
LA
Mean Slope (Degree)
7.4
0.13

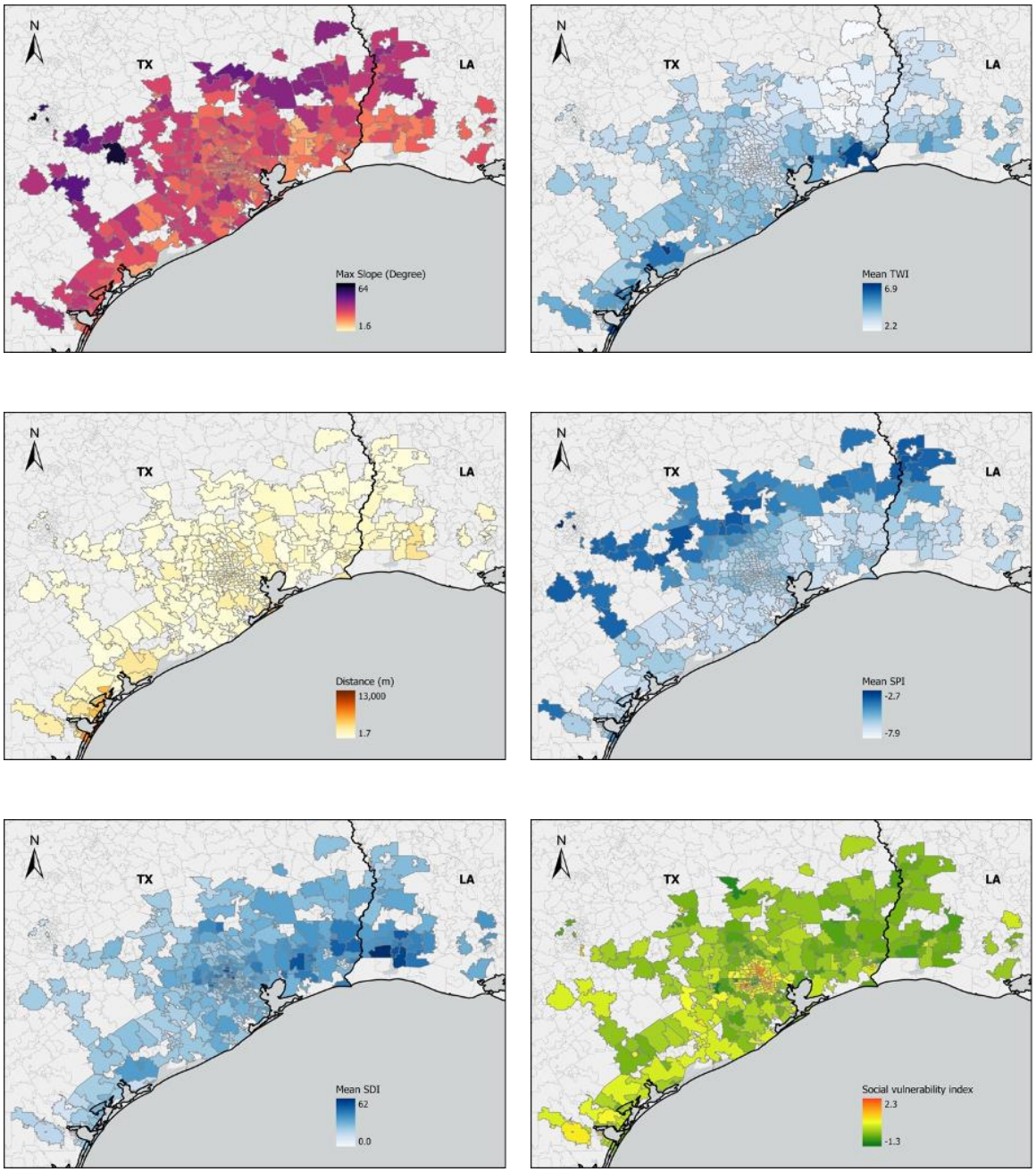


Figure S3. Spatial distribution maps of input dataset

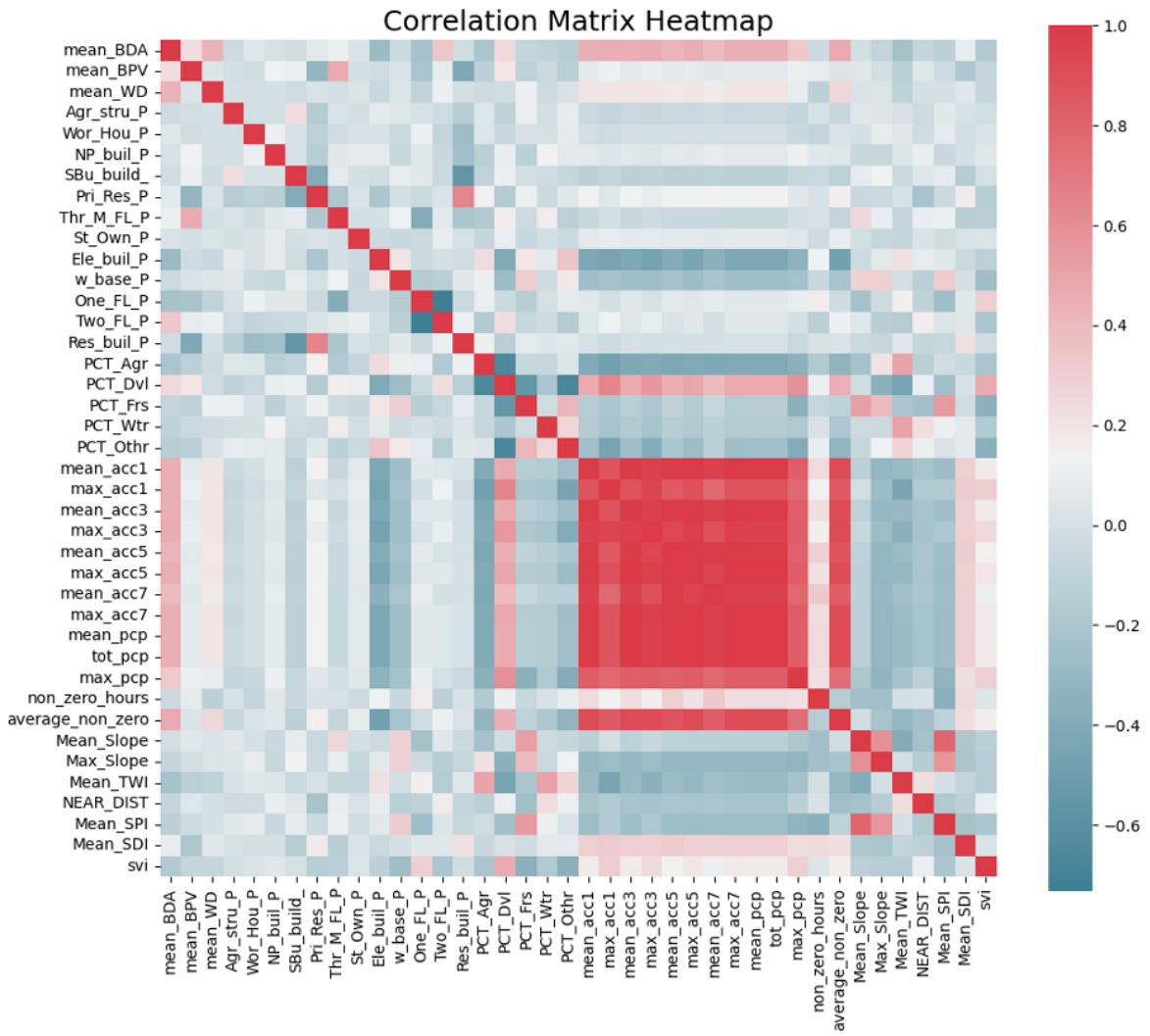


Figure S2. Correlation matrix between variables (Check Table S3 for variable description)

References

1. Amini, A., Ali, T. M., Ghazali, A. H. B., Aziz, A. A., & Akib, S. M. (2011). Impacts of Land-Use change on streamflows in the Damansara Watershed, Malaysia. *Arabian Journal for Science and Engineering. Section B, Engineering*, *36*(5), 713–720. https://doi.org/10.1007/s13369-011-0075-3

2. Bagnold, R. A. (1966b). An approach to the sediment transport problem from general physics. *U.S. Geological Survey Professional Papers/U.S. Geological Survey Professional Paper*. https://doi.org/10.3133/pp422i

3. Blessing, R., Sebastian, A., & Brody, S. D. (2017). Flood Risk Delineation in the United States: How Much Loss Are We Capturing? *Natural Hazards Review*, *18*(3). https://doi.org/10.1061/(asce)nh.1527-6996.0000242

4. Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards*. *Social Science Quarterly*, *84*(2), 242–261. https://doi.org/10.1111/1540-6237.8402002

5. De Moel, H., Jongman, B., Kreibich, H., Merz, B., Penning-Rowsell, E. C., & Ward, P. J. (2015). Flood risk assessments at different spatial scales. *Mitigation and Adaptation Strategies for Global Change*, *20*(6), 865–890. https://doi.org/10.1007/s11027-015-9654-z

6. Dey, H., Shao, W., Haque, M. M., & VanDyke, M. (2024). Enhancing flood risk analysis in Harris County: Integrating flood susceptibility and social vulnerability mapping. *Journal of Geovisualization and Spatial Analysis*, *8*(1). https://doi.org/10.1007/s41651-024-00181-5

7. Dey, H., Shao, W., Pan, S., & Tian, H. (2023). The Spatiotemporal Patterns of Community Vulnerability in the U.S. Mobile Bay from 2000–2020. *Applied Spatial Analysis and Policy*. https://doi.org/10.1007/s12061-023-09549-4

8. Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2020). A Social Vulnerability Index for Disaster Management. *Journal of Homeland Security and Emergency Management*, *8*(1). https://doi.org/10.2202/1547-7355.1792

9. Fox, D., Witz, E., Blanc, V., Soulié, C., Penalver-Navarro, M., & Dervieux, A. (2012). A case study of land cover change (1950–2003) and runoff in a Mediterranean catchment. *Applied Geography*, *32*(2), 810–821. https://doi.org/10.1016/j.apgeog.2011.07.007

10. Ghosh, S., Saha, S., & Bera, B. (2022). Flood susceptibility zonation using advanced ensemble machine learning models within Himalayan foreland basin. *Natural Hazards Research*, *2*(4). https://doi.org/10.1016/j.nhres.2022.06.003

11. Gudiyangada Nachappa, T., Tavakkoli Piralilou, S., Gholamnia, K., Ghorbanzadeh, O., Rahmati, O., & Blaschke, T. (2020). Flood susceptibility mapping with machine learning, multi-criteria decision analysis, and ensemble using Dempster Shafer Theory. In *Journal of Hydrology* (Vol. 590). https://doi.org/10.1016/j.jhydrol.2020.125275

12. Hole, F. D., & Campbell, J. B. (1985). *Soil Landscape Analysis*. Rowman and Allanheld, Totowa, NJ 196 pp.

13. Kirkby, M. J. (1975). "Hydrograph modeling strategies". Progress in physical and human geography, 689 R. F. Peel, M. D. Chisholm, and P. Haggett, eds., Heinemann, London, 69-90.

14. Liu, X., Zhou, P., Lin, Y., Sun, S., Zhang, H., Xu, W., & Yang, S. (2022). Influencing Factors and Risk Assessment of Precipitation-Induced Flooding in Zhengzhou, China, Based on Random Forest and XGBoost Algorithms. *International Journal of Environmental Research and Public Health*, *19*(24). https://doi.org/10.3390/ijerph192416544

15. Meraj, G., Romshoo SA, Yousuf AR, Altaf S, Altaf F. 2015. Assessing the influence of watershed characteristics on the flood vulnerability of the Jhelum basin in Kashmir Himalaya. Nat Hazards. 77(1):153–175.

16. Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article " Assessment of economic flood damage; *Natural Hazards and Earth System Sciences*, *10*(8), 1697–1724. https://doi.org/10.5194/nhess-10-1697-2010

17. Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: A literature review. In *Water (Switzerland)* (Vol. 10, Issue 11). https://doi.org/10.3390/w10111536

18. Pistrika, A., Tsakiris, G., & Nalbantis, I. (2014b). Flood Depth-Damage functions for built environment. *Environmental Processes*, *1*(4), 553–572. https://doi.org/10.1007/s40710-014-0038-2

19. Ren, H., Pang, B., Bai, P., Zhao, G., Liu, S., Liu, Y., & Li, M. (2024). Flood Susceptibility Assessment with Random Sampling Strategy in Ensemble Learning (RF and XGBoost). *Remote Sensing*, *16*(2). https://doi.org/10.3390/rs16020320

20. Schaetzl, R. J. (1986). Soilscape analysis of contrasting glacial terrains in Wisconsin. *Annals of the Association of American Geographers*, *76*(3), 414–425. https://doi.org/10.1111/j.1467-8306.1986.tb00128.x

21. Scorzini, A.R. and Frank, E. (2017), Flood damage curves: new insights from Italy. J. Flood Risk Manage, 10: 381-392. https://doi.org/10.1111/jfr3.12163

22. United States Economic Development Administration. *Recovering from Harvey: EDA Leads Coordination of Federal Resources,* last revised January 2021. https://www.eda.gov/impact/success-stories/disaster-recovery/recovering-harvey-eda-leads-coordination-federal-resources (accessed 2024-07-07).

23. Urban Floods Community of Practice (UFCOP). *Land Use Planning for Urban Flood Risk Management* last revised April 2017. https://www.gfdrr.org/sites/default/files/publication/UFCOPKnowledgeNoteMay.pdf (accessed 2024-07-17).

24. Wang, M., Li, Y., Yuan, H., Zhou, S., Wang, Y., Adnan Ikram, R. M., & Li, J. (2023). An XGBoost-SHAP approach to quantifying morphological impact on urban flooding susceptibility. *Ecological Indicators*, *156*. https://doi.org/10.1016/j.ecolind.2023.111137

25. Zhou, S., Liu, Z., Wang, M., Gan, W., Zhao, Z., & Wu, Z. (2022). Impacts of building configurations on urban stormwater management at a block scale using XGBoost. *Sustainable Cities and Society*, *87*. https://doi.org/10.1016/j.scs.2022.104235

26. Zope, P. E., Eldho, T. I., & Jothiprakash, V. (2014). Impacts of urbanization on flooding of a coastal urban catchment: a case study of Mumbai City, India. *Natural Hazards*, *75*(1), 887–908. https://doi.org/10.1007/s11069-014-1356-4

Appendix

Team Leaders

Assimilation of Geosynchronous satellite imagery Into NextGen For improved modeling

"I am currently affiliated with University of Alabama. I have a PhD at the University of Alabama. I hold a Master of Science from the University of California, Irvine, and a Bachelor of Science from California State University, Monterey Bay. My research interests lie in large-scale hydro/hydro-dynamic modeling and machine learning physical process representations. Originally from California, I feel equally at home in Alabama. My favorite hobby is skateboarding, though I have paused it to train for endurance events. I also enjoy chess, jazz, and philosophy."	**Jonathan Frame** 

Developing applications that effectively characterize the uncertainty in Flood Inundation Mapping (FIM)

"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."

Sagy Cohen

Xingong Li

"I am a Professor at the University of Kansas, holding a Ph.D. from the University of South Carolina. My research interests include flood inundation mapping, surface water mapping, and the Map Compute framework. Born and raised in western China, I have been a dedicated member of the University of Kansas for over 20 years. In my free time, I enjoy running and hiking. Additionally, I am deeply fascinated by the origins of life and the role that information plays in its development."

Geospatial Data Integration to Identify High-flow Thresholds for Improved Flood Risk Characterization

"I work for the Flood Solutions Lead at FloodMapp, based in Boulder, CO. I completed a Fulbright Postdoc at the University of Newcastle in Australia and another postdoc at the University of Bayreuth in Germany. I hold a Ph.D. and an M.S. from Kansas State University and a B.S. from North Carolina State University. My research interests include flood inundation mapping, flood early warning and operational forecasting, ecosystem services, and applied research. I grew up in Rhode Island and have lived in four countries. In my free time, I love hiking, gardening, and traveling. I share my life with two dogs, Winston and Oliver."	***Kelsey Mcdonough*** 
Sanjib Sharma 	"I am an Assistant Professor at Howard University, holding a Ph.D. from Pennsylvania State University. My research interests encompass hydrometeorological forecasting, machine learning, uncertainty quantification, and risk and decision analysis. With a background in engineering, I have experience in both consulting and academia. In my free time, I enjoy playing soccer, hiking, and swimming."

Summer Institute Course Coordinators

From left to right: Sadaf Mahmoudi, and Karina Larco.

Summer Institute Teams

Chapter 1: Leveraging Geosynchronous Satellite Imagery and Machine Learning to Predict the Magnitude and Extent of Snow Water Equivalent

Group SWE Home Alabama

From left to right: Raymond Hess, Helaleh Khoshkam, Hassan Saleh, and Savalan Naser Neisary.

Chapter 2: Identifying Atmospheric Rivers on the West Coast of the United States with Geostationary Operational Environmental Satellite Imagery

Group AetheR

From left to right: Janani Kandasamy, Surabhi Upadhyay, Meklit Berihun Melesse, and Anshul Yadav.

Chapter 3: Probabilistic Streamflow Prediction Using the Model-Agnostic NextGen Framework

Group ProNextGen

From left to right: Ali Sattari, Reza Morovati and Hossein Gholizadeh

Chapter 4: Enhancing Remote Sensing Flood Inundation Mapping by Leveraging a Terrain-Based Model

Group FLDSensing

From left to right: Francisco Gomez, Jack Edwards, and Kim Son Do

Chapter 5: Flood Inundation Mapping Using Terrain-Based Models for Flash Floods due to Dam Operations

Group SHARP_ FIM

From left to right: Ayman Mokhtar Nemnem, Reza Saleh Alipour, Parvaneh Nikrou and Shivakumar Balachandaran

Chapter 6: Transition From Hazard to Impact-based Riverine Flood Forecasting

Group Flood Fighters

From left to right: Duc Tran, Hesham Elhaddad, and Lyn Watts

Chapter 7: Leveraging Geospatial Data and Machine Learning to Predict Insurance-derived Flood Damage Cost

Group hydroHERoes

From left to right: Aylar Samadi, Reihaneh Zarrabi, and Corrine Liu