QUANTIFYING PRECIPITATION, STREAMFLOW, AND FLOODPLAIN FORECASTING SKILLS DURING EXTREME WEATHER EVENTS IN BRAYS BAYOU, HOUSTON, TEXAS

A Thesis

by

CHERYL HOLMES

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Huilin Gao
Committee Members,	Francisco Olivera
	Inci Güneralp
Head of Department,	Robin Autenrieth

August 2019

Major Subject: Civil Engineering

ABSTRACT

Extreme precipitation and increased urban land cover have increased the frequency and severity of urban flooding events in recent years. Accurate precipitation, streamflow, and floodplain inundation forecasts are necessary to decrease the damage from these events via reservoir operations planning, evacuation of residents, and mobilization of relief efforts. In this study, Quantitative Precipitation Forecasts (QPFs) developed by the National Weather Service (NWS) were analyzed for their skill in predicting precipitation in Brays Bayou in Houston, Texas. This forecasted data were used to force the Distributed Hydrological Soil and Vegetation Model (DHSVM), a physically-based, distributed hydrological model, and the resulted streamflows were assessed for accuracy. Then, a 2-dimensional hydraulic model, Flood2D-GPU, was employed to produce forecasted floodplains, also with skill assessment. This study focuses on three major flood events in the last decade with an emphasis on Hurricane Harvey. Results were focused on three aspects: 1) identifying changes in forecast accuracy with increased lead time; 2) quantifying skill scores of the forecasts through the flood forecasting system; and 3) comparing DHSVM forecasts with those used by the West Gulf River Forecasting Center (WGRFC) to identify optimal forecasting lead time during extreme events.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my committee chair and advisor of the last two years, Dr. Huilin Gao. I am so grateful for the opportunity she has afforded me in studying with her at Texas A&M University. I am in awe of her dedication to her field and her students. I appreciate all of the help and advice she has given me, both in regards to my academic work and my future career. I would also like to thank my committee members, Dr. Francisco Olivera, Dr. Inci Güneralp, and Dr. Jacob Torres. I appreciate your insight, encouragement, and support throughout this process. I also want to thank Dr. Anthony Cahill for his contributions.

A huge thank you also goes out to the Gao research group: Gang Zhao, Manqing Shao, Xudong Li, and Yao Li. I have received so much help from all of you and am so grateful for your willingness to be another set of eyes to help figure out solutions to my problems.

I would like to thank the departmental faculty and staff for their assistance with all aspects of my degree, especially Laura Byrd and Chris Grunkemeyer. My research has also been supported by the TAMU High Performance Research and Computing. I am thankful to the Texas A&M system as a whole, for cultivating my interest in civil engineering and allowing me to pursue graduate studies.

Thank you also to Dr. Shih-Sheh Kao, Sudershan Gangrade, Kris Lander, and Nathalie Voisin, for your many contributions to this work. I am very grateful for your help and advice.

Thank you to my dear parents and brother for their loving encouragement and advice throughout my life, including my graduate studies. And, finally, thank you to my

fiancé, Will Rankin. Your support and encouragement have meant the world to me and helped me persevere to the end.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a thesis committee including Dr. Huilin Gao and Dr. Olivera of the Department of Civil Engineering, as well as Dr. Güneralp of the department of Geosciences and Dr. Jacob Torres of Lockwood, Andrews, and Newnam, Inc. Additional contributors external to Texas A&M University include Kris Lander of NOAA, Dr. Shih-Sheh Kao and Sudershan Gangrade of Oak Ridge National Laboratory, and Dr. Nathalie Voisin of PNNL. The inundation modeling in Section 4.3 was conducted by Sudershan Gangrade. Within the Gao Research Group, Dr. Gang Zhao and Manqing Shao contributed to this work. All other work conducted for the thesis was completed by the student independently.

This work was funded by the U.S. National Science Foundation Grant 1805584 and a fellowship from Texas A&M University. The contents of this thesis are solely the responsibility of the author and do not necessarily represent the official views of Texas A&M University, the Civil Engineering Department, nor the National Science Foundation.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

1. INTRODUCTION

Extreme precipitation events are occurring with greater frequency and intensity in recent years due to climate change (Lehmann et al. 2015). Severe flooding is one of the most devastating consequences of these events and can result in massive environmental and economic losses and even loss of life. In the last 30 years, the average number of weather fatalities caused by floods was second only to heat fatalities in the United States (NWS 2017a) and the floods have caused an estimated $275 billion in damages in 2016 dollars (OECD 2017). Furthermore, flood damages in the United States continue to increase over time despite engineered efforts to combat them (Pielke et al. 2002). Floodwaters can move sediment and contaminate water supplies, posing an additional threat to the environment and human life.

Because floods occur in a relatively short period of time and have the potential to cause massive amounts of destruction, it is important to be able to predict when and where floods will occur in advance. Hydrological forecasting can be used as a tool to effectively plan for and mitigate emergency situations. Hydrological forecasting before and during extreme weather events can be used to aid in strategic planning for first responders and to determine resident evacuation needs (Selvanathan et al. 2018). Additionally, reservoir operations can be assessed, and pre-releases can be made to make room for additional water from ongoing precipitation to avoid dam failure (Chen et al. 2017). Thus, it is important to study flood forecasting systems to determine (and therefore utilize in practice) the longest skillful lead time to provide useful information about the impending and ongoing event.

However, the usefulness of flood forecasts to mitigate the effects of flooding is dependent upon the quality and accuracy of the three types of forecasts that are being made throughout the flood forecasting system: precipitation, streamflow, and floodplain inundation forecasts. It has been shown that the uncertainty in precipitation forecasts is the main source of uncertainty in flood forecasting systems (Zappa et al. 2011). As Steenbergen and Willems (2014) indicated in their study of the Rivierbeek catchment in Belgium, about 30 percent of streamflow forecast uncertainty could be explained by rainfall input data uncertainty. Further, the errors from the precipitation forecasts will be propagated and exacerbated through the streamflow and flood inundation forecasts (Cluckie and Xuan 2008, Cloke and Pappenberger 2009, Rossa et al. 2011, Steenbergen and Willems 2014) due to model, input, and initial condition uncertainty (Cloke and Pappenberger 2009, Zappa et al. 2011). Thus, it is important to study the quality and accuracy of forecasts throughout the flood forecasting system to determine their usefulness.

Streamflow forecasts are produced using computerized hydrological models that simulate the hydrological cycle, including the movement of water over the land and infiltration into the soil. While lumped models spatially average the key characteristics of the watershed, distributed models discretize the watershed into distinct cells with varying characteristics. This is an advantage for the distributed model, as it can capture the spatial heterogeneity of the watershed conditions, including incoming precipitation, land surface cover, and soil type. Thus, distributed models are considered to generally produce more realistic and accurate streamflow results that reduce model error when compared with lumped model results (Koren et al. 2003, Zhang et al. 2004, Carpenter and Georgakakos

2006). Furthermore, the streamflow outputs from the hydrological models can be used to drive hydraulic models that produce simulated floodplains. Like the difference between lumped and distributed hydrological models, 1D hydraulic models represent the cross sections of a stream only, while 2D hydraulic models include a computational mesh or digital elevation model (DEM) on which the model is calculated. Studies acknowledge that it is difficult to determine whether or not the results from 2D models are improved over the results of 1D models (Horritt and Bates 2002, ShahiriParsa 2016). However, the additional details captured by 2D models are generally thought to be useful in areas of complex topology and can reduce errors in modeling flood extent (Cook and Merwade 2009).

To quantify the errors in flood forecasting systems, it is common practice to calculate statistical skill scores of the various forecasts: precipitation, streamflow, and flood inundation. Skill scores indicate the quality and appropriateness of these forecasted results. The skill scores of the data and simulations are typically reported based on the lead time of the forecasts. Generally, with increased lead time, the skill scores, and thus ability to provide useful information, of each type of forecast decline (Voisin et al. 2011, Shukla et al. 2012, Nguyen et al. 2015, Seo et al. 2016). These studies indicate the uncertainties associated with precipitation forecasting and follow this uncertainty through the other components of flood forecasts in an effort to analyze the effectiveness of flood forecasting systems.

Hurricane Harvey was a catastrophic event that caused major flooding in and around Houston, Texas, and affected many of the coastal states on the Gulf of Mexico, including Louisiana, Mississippi, as well as Tennessee, and Kentucky. With a return

period estimated to be greater than 9,000 years (van Oldenborgh et al. 2017), Hurricane Harvey was considered the most significant tropical cyclone rainfall event in United States history, both in scope and peak rainfall amounts (Blake and Zelinsky 2018). This category 4 hurricane made landfall in Texas on August 25, 2017 and had direct effects on Houston from August 25 – 30, 2017. According to the National Hurricane Service report on Hurricane Harvey, the hurricane caused $125 billion in damage, mostly due to extreme flooding. This is more damage than any single storm in US history except for Hurricane Katrina in 2005. It is estimated that 203,000 homes were damaged or destroyed, and 68 people died from the direct effects of the storm (Blake and Zelinsky 2018). Infrastructure suffered severe damage, including washed out roads and collapsed bridges. The safe operation of dams upstream of the city was also affected. The ramifications of this storm will be experienced for years, as the city and surrounding areas will need time and investment to rebuild and recover.

For advance warning of flooding in Houston and its surrounding areas, the National Weather Service (NWS) West Gulf River Forecast Center (WGRFC) provides river forecasts up to 12 hours in advance under normal circumstances. Quantitative Precipitation Forecasts (QPFs) produced by the National Centers for Environmental Prediction (NCEP) Weather Prediction Center (WPC) are used to produce these forecasts. The WPC QPFs are based on NWS ensemble weather forecasts and are manually adjusted by modelers based on their experience and judgement (https://www.wpc.ncep.noaa.gov/html/fam2.shtml#qpf). The QPFs, which are sometimes manually adjusted a second time by WGRFC modelers, are used to drive the Community Hydrologic Prediction System's Flood Early Warning System (CHPS-FEWS model). The

CHPS-FEWS model is a lumped hydrological model that is used to produce deterministic streamflow forecasts. Then, rating curves from the United States Geologic Survey (USGS) are used to convert the forecast streamflow into flood stage. During extreme events, including Hurricane Harvey, the WGRFC modelers' judgement resulted in extending their hydrological forecast lead time from 12 to 48, and then to 72 hours. However, it is not quantified if this lead time extension resulted in skillful streamflow predictions and flood warning during extreme events like Hurricane Harvey.

Therefore, the key objectives of this study are to quantify the skills of forecasts through a flood forecasting system and determine the effect of lead time on forecasts during storm events. The study will analyze the three most extreme flood events in Houston in the last decade (i.e., the Memorial Day flood (2015), the Tax Day flood (2016), and Hurricane Harvey (2017)), with a focus on Hurricane Harvey. The following questions will be answered:

1) What is the magnitude of error throughout the flood forecasting system? What types of events have more or less errors in their forecasts?
2) Do forecasts with lead times beyond 12 hours provide useful information to forecasters? Are there certain characteristics for storm events that provide more useful forecasts?
3) Does a distributed hydrological model provide improved forecasts when compared with the forecasts from a lumped hydrological model?

2. STUDY AREA

The state of Texas has experienced multiple extreme storms that have resulted in massive flooding damages and deaths in the past few decades. In the year 2017, Texas had more flood-related deaths than any other state (NWS 2017b). In particular, the city of Houston has experienced multiple severe flooding events in the last few decades. Since 2000, there have been five major flood events: Tropical Storm Allison (2001, 35 inches), the west side flood (2009), the Memorial Day Flood (2015, 11 inches), the Tax Day Flood (2016, 17 inches), and Hurricane Harvey (2017, 50-60+ inches). Damage to property and infrastructure has been devastating, as well as the toll on human life.

The Brays Bayou watershed is located in southwest Houston and is mostly in Harris County, with the southwestern-most part belonging to Fort Bend County, as shown in Figure 1 (29.37-29.45°N, 95.16-95.41°W). Brays Bayou watershed includes three primary streams: Brays Bayou, Keegans Bayou, and Willow Waterhole Bayou. According to the Harris County Flood Control District, the drainage area of the watershed is 329 km^2, with 195 km open stream. The elevation in the watershed ranges from -0.12 to 49.72 meters, with the land sloping downward from west to east as the watershed approaches the coast (Figure 2a). On the northwestern side, outside of the watershed, is Barker Reservoir, which is used for flood control purposes, and makes releases to Buffalo Bayou to the north of Brays Bayou.

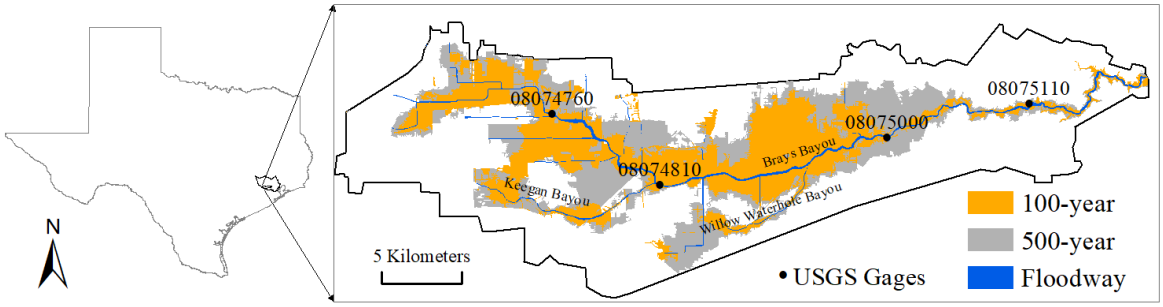

Figure 1 Map of the study area, including locations of USGS streamflow gages and FEMA map of the floodway, 100-year, and 500-year floodplain. Adapted from Bass et al. (2017).

The climatology of the watershed is generally wet and subtropical, characterized by humid, warm summers and mild winters. The average rainfall is 1,415 mm per year with July and August being the wettest months, on average (Figure 3). The average daily temperature is 22.3°C and the average relative humidity is 76.1%. The primary land cover type is urban land (Figure 2b) and the primary soil type is clay. Clay has a very low hydraulic conductivity. Thus, the water does not easily infiltrate into the soil, making the area prone to flooding.

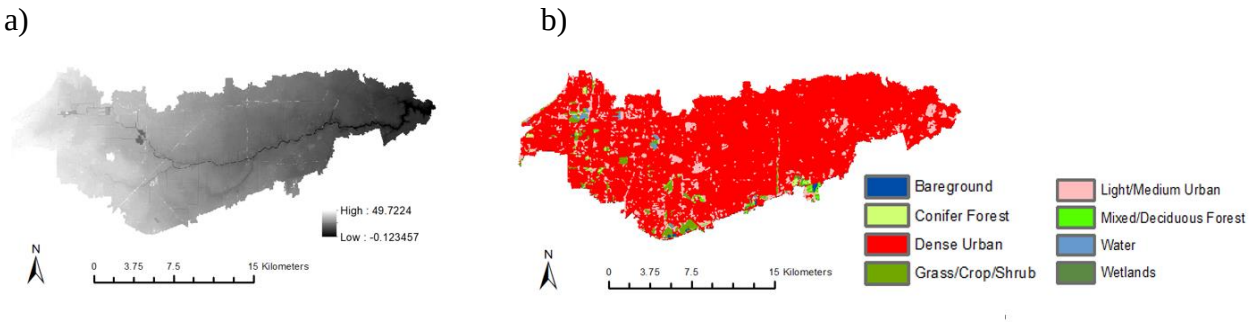

Figure 2 Spatial distribution of data in the watershed, including (a) digital elevation model (DEM) in meters and (b) land use/land cover data. Sources: (a) USGS 3D Elevation Program (3DEP) and (b) National Land Cover Database (NLCD) 2011.

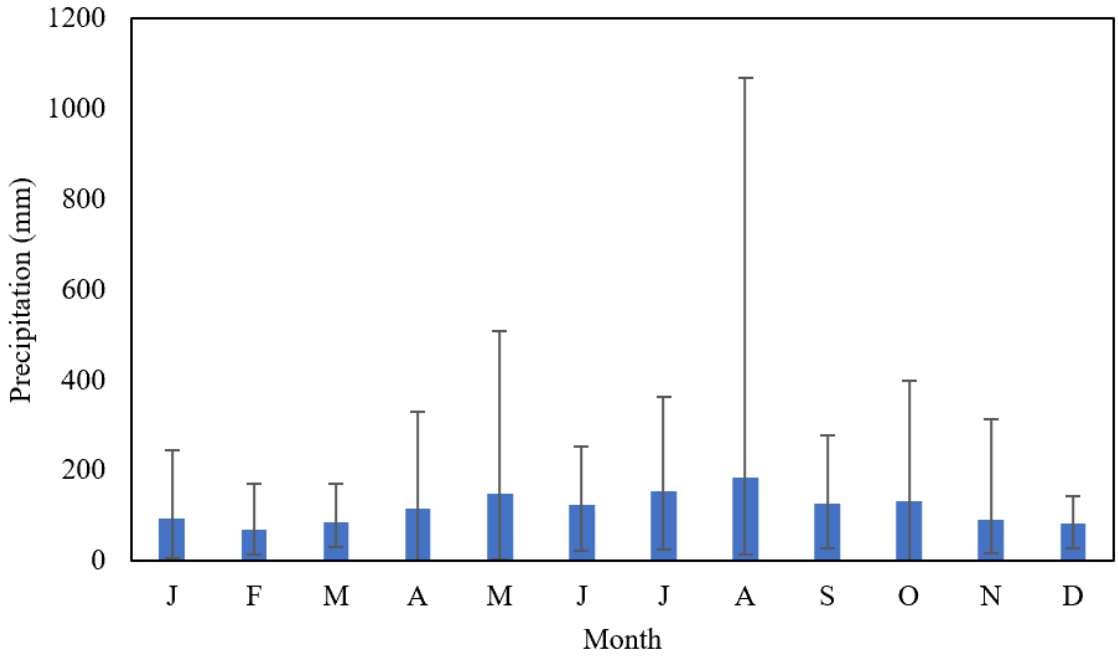

Figure 3 Monthly average rainfall in Brays Bayou. Data is spatially averaged over the watershed from 2003-2017. The error bars represent the maximum and minimum monthly rainfall. Source: National Centers for Environmental Protection (NCEP) Stage IV Quantitative Precipitation Estimates (ST4 QPE).

In recent years, Brays Bayou (and the city of Houston in general) has been experiencing increased urbanization and population growth. According to the 2017 Metro Houston Population Forecast, the population in the city is projected to almost double between 2010 to 2050. Brays Bayou is a heavily urbanized area with a population of over 0.7 million in the Harris County portion of the watershed (U.S. Census 2010 cited in HCFCD website). While Brays Bayou is primarily composed of urbanized land (95.9% urbanized, Figure 2b), additional population growth will result in further urbanization in the watershed. The effect of urbanized land on the hydrological cycle is primarily tied to increased impervious surface area, which results in less water infiltration into the ground, and thus more runoff. Increased runoff has been shown to result in increased flooding

when compared to natural landcover (Olivera and DeFee 2007, Zhao et al. 2016, Muñoz et al. 2018).

Brays Bayou has experienced flooding from each of the major storm events in Houston listed above. Particularly in the Meyerland area, which sits adjacent to the Brays Bayou stream, many houses have flooded repeatedly during these storm events. During Hurricane Harvey, 26,752 structures experienced damage (HCFCD 2018). Figure 1 shows that a large portion of the watershed is included in the 100-year Federal Emergency Management Agency (FEMA) floodplain. Thus, it is important to be able to spatially and temporally forecast flooding events in this watershed.

3. DATA AND METHODS

This section introduces the precipitation data, hydrological and hydraulic models, and methods for assessing precipitation, streamflow, and floodplain forecasting skills.

3.1 Precipitation Data

In this study, two types of precipitation data were used: Stage IV Quantitative Precipitation Estimates (ST4 QPEs) and NWS QPFs. The ST4 QPE product served two purposes in this study: 1) to evaluate QPF performance, and 2) to drive the DHSVM model with observed precipitation. ST4 QPEs were taken as the "ground truth" in this study. ST4 QPEs are produced from the National Weather Center (NWC) River Forecast Centers (RFCs) Stage 3 data (Lin 2011) and were originally developed to be assimilated into atmospheric forecast models to produce improved QPFs (Lin and Mitchell 2005). The ST4 QPEs are available in an hourly time step over 4 km grids from 2002 to present. The data is produced in near-real time, as it becomes available within 1 hour of receiving data from the RFCs. These are gage-corrected radar data with manual quality control from the RFCs and have been used in many studies to represent estimated precipitation (Sapiano and Arkin 2009, Ashouri et al. 2014, Nelson et al. 2015, Kao et al. 2019).

QPFs from the National Weather Service were analyzed for skill and then used to drive the DHSVM model to produce streamflow forecasts. The QPFs were acquired in GRIB format from ftp://ftp.hpc.ncep.noaa.gov/qpf_archive/. QPFs provide precipitation forecasts over the continental U.S. in 6-hour and 48-hour increments with spatial resolutions of 2.5 and 5 km, depending on the forecast. The 6-hour product provides forecasts until 84 hours in the future; the 48-hour accumulation product is available up to 168 hours into the future. The 6-hour product was used in this study in order to assess

model performance on a relatively short-range, fine temporal scale. Furthermore, the WGRFC uses the 6-hour product to monitor for flood alerts. Weather Prediction Center (WPC) modelers use their experience and guidance from operational and ensemble models to determine which models produce a reasonably likely amount of precipitation and then make manual adjustments. NWS QPFs consistently earn higher threat scores than other forecasted models, including the North American Mesoscale Forecast System (NAM), the Global Forecast System (GFS), and the European Center for Medium-Range Weather Forecasts (ECMRWF) (https://www.wpc.ncep.noaa.gov/html/hpcverif.shtml#medmin). Thus, NWS QPFs were used in this study to determine their forecasting skill and appropriateness for use in a flood forecasting system.

3.2 DHSVM

The Distributed Hydrology Soil and Vegetation Model (DHSVM) was employed in this study to simulate streamflow in Brays Bayou. As explained in Wigmosta et al. (1994), the model is fully-distributed and physically-based with high spatial (10-200 meters) and temporal resolution (hourly to subdaily/daily time steps). Hydrological processes such as evapotranspiration, infiltration, snowmelt, and urban area detention are represented in the model. The energy and water balance are solved at each grid cell for every time step. The evapotranspiration algorithm is based on the Penman–Monteith method (Shuttleworth 1992). Water movement is determined by topography via a high-resolution DEM, with cells draining to and receiving water from other cells. Unsaturated and saturated water movement are governed by Darcy's Law. The kinematic runoff routing method is used to route overland flow to the channel. This method employs a

simplification of the Saint Venant equations and finite difference solution scheme. The linear reservoir routing method is used to represent streamflow through the river reach.

An advantage of the DHSVM is that it has an urban module which can represent the effects of impervious cover and detention on hydrological processes (Cuo et al. 2008). This is especially important in this highly urban watershed. In the module, two aspects of urban storm water management are represented: runoff generated from impervious surfaces and storage due to detention basins in manmade channels. The precipitation that falls on impervious surfaces in the watershed becomes surface runoff in the module. However, depending on the land cover type, a portion of the runoff is routed directly to the channel, while the rest of the runoff is released slowly to the channel following linear storage theory. This slow release mimics storm water detention that exists in most manmade channels with detention basins. There are two land cover types in the model that allow this type of relationship: dense urban and light urban. In this study, 80% and 40% impervious cover for dense urban and light urban land, respectively, were used during the calibration and validation processes. Additionally, the user can specify the percentage of water to be stored in flood detention and released slowly according to the linear storage theory.

Input data to the model includes the DEM, soil type, land use/land cover, and meteorological forcing data. The input data were resampled to 20 m resolution and summarized in Table 1 below.

Table 1 Data and sources used as input to DHSVM.

Type of Data	Source	Spatial Resolution
1/3 Arc-Second Digital Elevation Model	National Elevation Dataset	Resampled to 20 meters
Soil Type	SSURGO Database from NRCS and USDA	Resampled to 20 meters
Land Use Land Cover Data for 2011	National Land Cover Database	Resampled to 20 meters
Meteorological Forcing Data (other than precipitation)	North America Land Data Assimilation System (NLDAS)	1 kilometer, averaged over the watershed

A DEM with 10-meter resolution was obtained from the USGS's 3D Elevation Program (3DEP) (https://viewer.nationalmap.gov/basic/). The 3DEP used LIDAR to map elevation data for the coterminous United States. The data was projected into the NAD 1983 UTM Zone 15N coordinate system and resampled to 20 meters. The DEM was used to create the basin mask and flow direction in ArcGIS 10.4. Then, Python code utilized the DEM and basin mask to generate the stream network and soil depth files (Duan 2018).

Soil type spatial data was acquired from the United States Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS) Soil Survey Geographic Database (SSURGO). This dataset was created through both field observations and laboratory sample analysis. Due to the data's spatially detailed nature, it is used in this study to identify soil type in each grid cell. Based on the soil type identified from this dataset, the other soil parameters are specified in the DHSVM configuration file. The spatial soil data was also processed to the correct projection and spatial resolution.

Land use-land cover data for the year 2011 was obtained from the Multi-Resolution Land Characteristics Consortium (MRLC) National Land Cover Database (NLCD) (https://www.mrlc.gov/). Data for 2011 was the most current data available from this source at the time of this work. Since land use/land cover does not change dramatically from year to year, this data is sufficient to represent the current land use in the watershed. NLCD 2011 utilizes Landsat satellite data and conducts a decision-tree classification to differentiate land cover types into 16 classifications at a 30-meter spatial resolution. The data was re-projected into the appropriate coordinate system and resampled to 20 meters using the majority resampling method. These land cover types were then reclassified into DHSVM classifications (Figure 2b).

In addition to precipitation data (Section 3.1), other meteorological forcing data are required to drive the DHSVM model, including air temperature, wind speed, relative humidity, and incoming shortwave and longwave radiation. These data were obtained from the National Land Data Assimilation Systems-2 (NLDAS-2) dataset. They are available hourly in 1/8-degree spatial resolution and were aggregated over the watershed in Google Earth Engine. These variables have less spatial heterogeneity than other model inputs (i.e., precipitation, land use/land cover, soil) and affect the water budget to a smaller extent. The wind speed was calculated from the horizontal and vertical wind components, and relative humidity was calculated from specific humidity and air temperature data. The same meteorological forcing data was used for each precipitation station across the watershed.

For each timestep, DHSVM outputs the current timestep's baseflow and the previous timestep's streamflow. This mismatch in calculation timing could affect the total

streamflow produced by the model, particularly during large events where baseflow has an impact on the total streamflow. Thus, the DHSVM output was manually adjusted to shift the total streamflow by 1 timestep.

3.2.1 Calibration and Validation of DHSVM

USGS observed 3-hourly streamflow data collected at the three most downstream gages (Figure 1, Stations 08074810, 08075000, and 08075110) were used to calibrate and validate the model. Since the most upstream gage only captures a small portion of the watershed area, its data were not used for evaluating the model. Specifically, streamflow data from 2017 were used to calibrate the model while those from 2015-2016 were used for validation. Simulations during 2014 were used for spin up. In this study, a spatial resolution of 20 meters was selected in order to capture the spatial variability of the watershed and a temporal resolution of 3 hours was used to represent temporal variability of the floods.

To calibrate DHSVM, soil and vegetation parameters were adjusted to produce optimal error statistics and reliable validation results. The model calibration and validation were assessed with traditional statistical methods against USGS streamflow data, including coefficient of determination (R^2), Nash-Sutcliffe Efficiency (NSE), and Relative Bias (RB) calculations after equations 1, 2, and 3 below:

$$R^2 = 1 - \frac{\sum_t (S_t - O_t)^2}{\sum_t (S_t - \bar{S})^2}, \tag{1}$$

$$NSE = 1 - \frac{\sum_t (S_t - O_t)^2}{\sum_t (O_t - \bar{O})^2}, \tag{2}$$

$$RB = \frac{(\bar{S} - \bar{O})}{\bar{O}}, \tag{3}$$

where S_t represents the simulated value and O_t represents the observed value at time t, and $\bar{S}$ and $\bar{O}$ represent the average of the simulated and observed values over the whole time series.

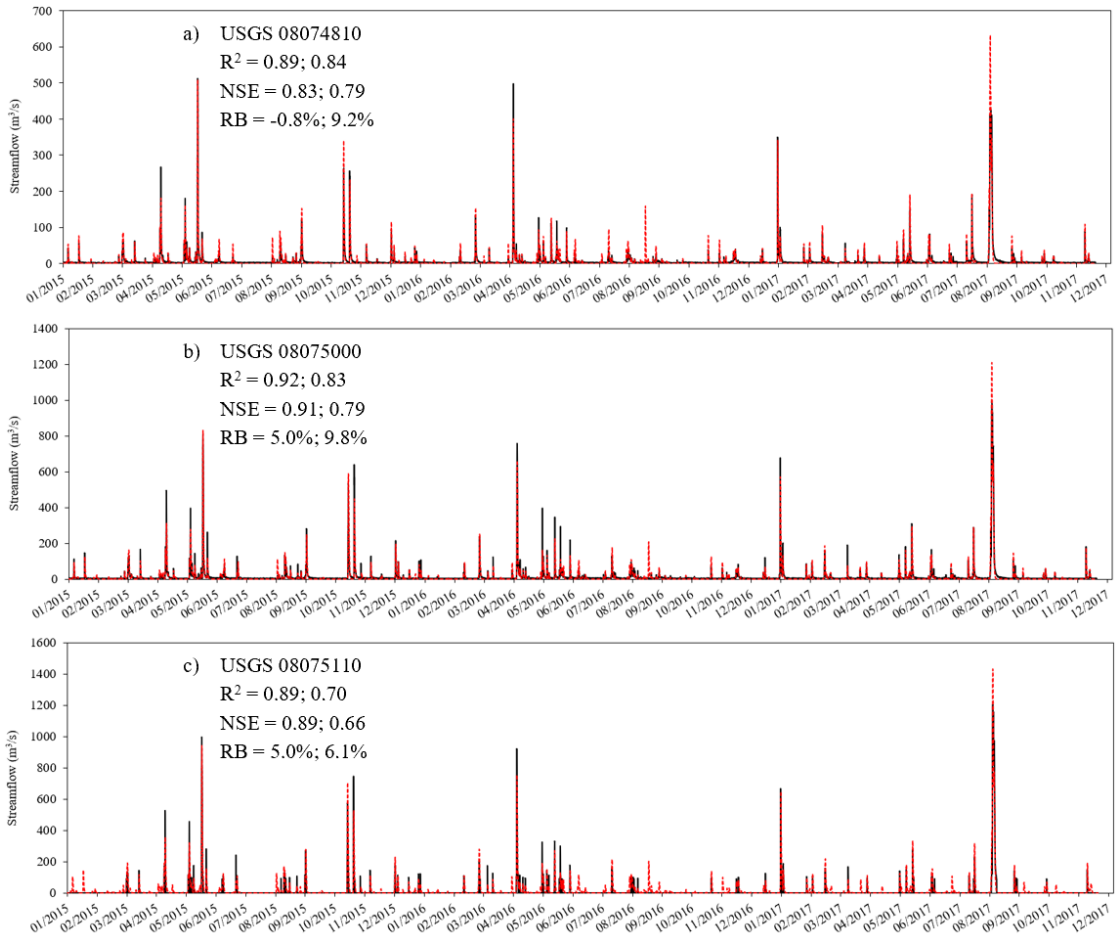

Figure 4 Results of 3-hourly calibration (2017) and validation (2015-2016) in Brays Bayou by USGS gage: a) USGS 08074810, b) USGS 08075000, and c) USGS 08075110. The first listed statistic describes the calibration; the second listed statistic describes the validation. Note that gage USGS 08075110 has an incomplete dataset.

Figure 4 compares the simulated streamflow against the gage data during both the calibration and validation periods. Note that USGS gage 08075110 has an incomplete dataset and only has recorded data during storm events. USGS gage 08075000 appears to

perform the best across all metrics except for relative bias, and all the gages show good agreement with the gage data. The calibration results are better than the validation results, which is to be expected in modeling. Therefore, the 3-hourly DHSVM model can be used for streamflow simulations.

Furthermore, the 3-hourly DHSVM simulations for each of the three storm events of interest are evaluated at USGS gage 08075000, which is located near the WGRFC's output location for streamflow simulations in Brays Bayou. Figure 5 shows that the model underestimates the peak for the Memorial Day and Tax Day floods but overestimates the peak for Hurricane Harvey. Overall, the error statistics (Table 2) are very high, indicating high correlation between the gage data and the simulated streamflow. With the exception of the high relative bias for the Memorial Day flood, the error statistics have more favorable values for these events than for the 3-year calibration and validation period. Thus, this analysis suggests that the model can robustly represent the hydrological processes during flood events in Brays Bayou.

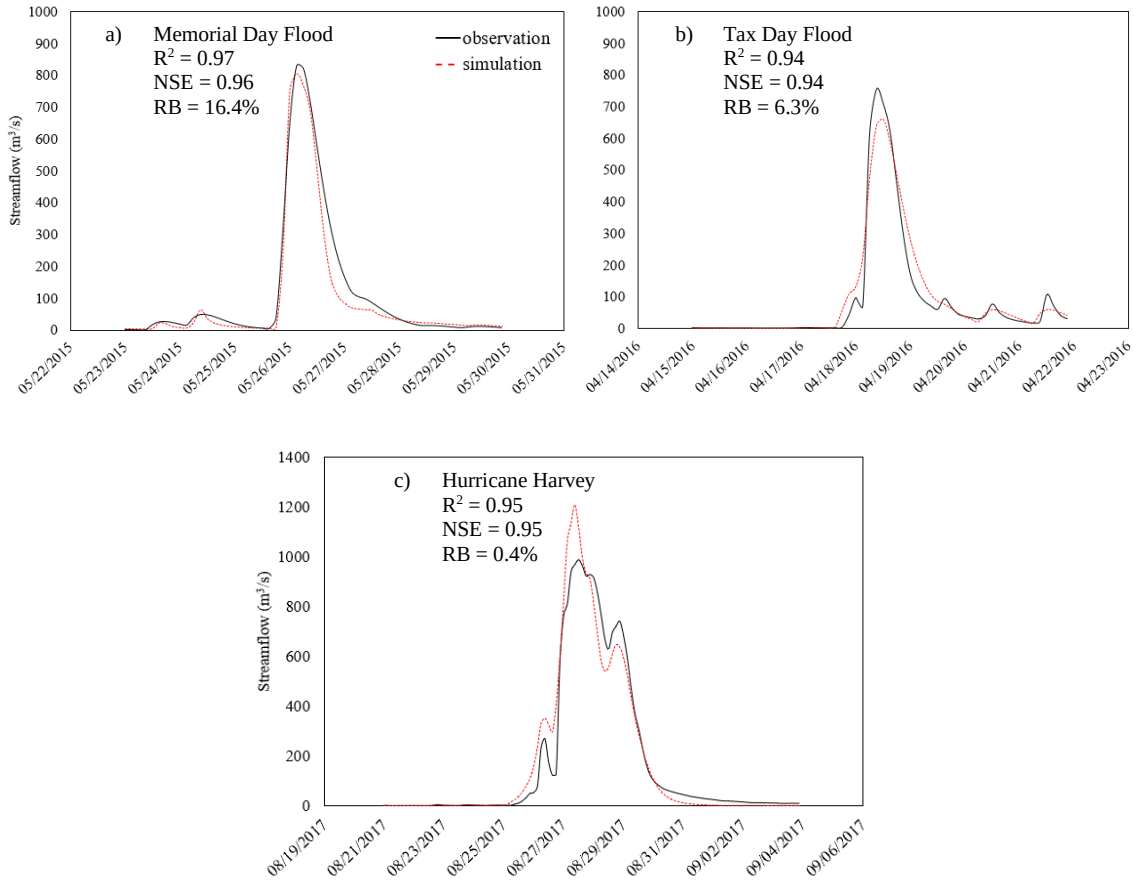

Figure 5 Evaluation of DHSVM at gauge 08075000 across three storm events of interest: a) Memorial Day Flood results, b) Tax Day Flood results, and c) Hurricane Harvey results.

3.3 Flood2D-GPU

The computationally enhanced version of the 2D hydraulic model Flood2D-GPU (Marshall et al. 2018) was employed in this study to generate the maximum floodplain inundation under forecasted streamflow conditions. The hydraulic model was originally developed at the University of Utah and is described in Kalyanapu et al. (2011). A first-order upwind difference scheme that solves the non-linear hyperbolic shallow water Saint Venant equations, which were derived from the Navier-Stokes equations, is used.

A 20-meter DEM that extended downstream of the watershed, along with model default Manning's surface roughness (n = 0.035) served as the base setup for the flood model. The stream network from DHSVM was utilized to identify channel segments across the entire watershed. Inflow locations for the flow hydrographs are shown in Figure 7. The input flow hydrographs extended for 10 days and captured the peak of the event at the end of the 5th day. The dates for each model run are given in Section 4.3. Outputs were saved at a 30-minute step across the entire computational domain (Figure 6).

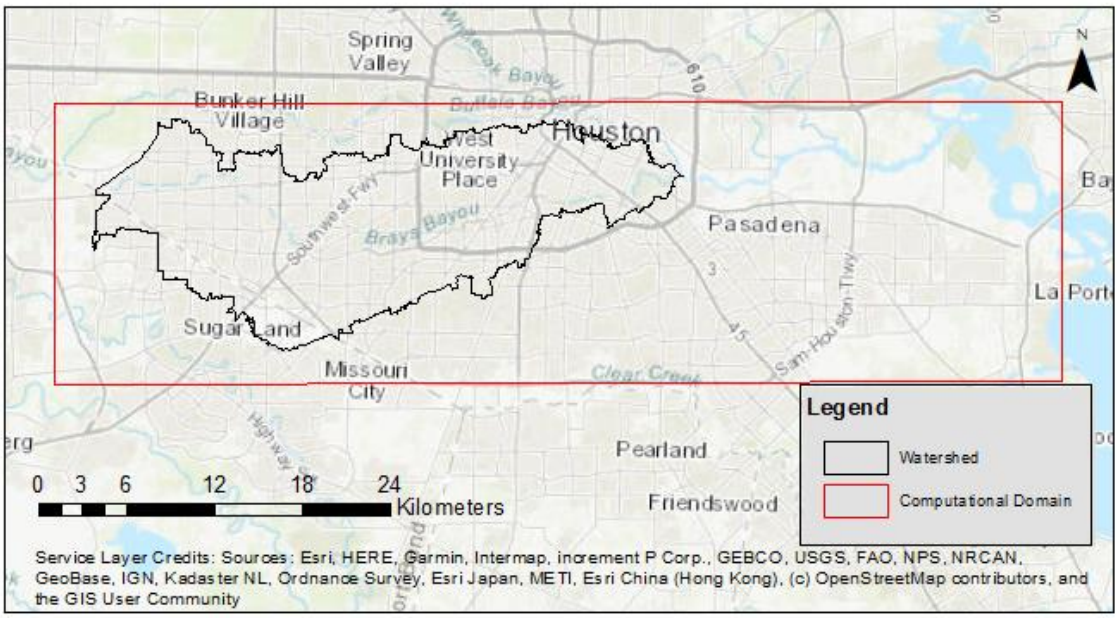

Figure 6 Watershed and computational domain used in Flood2D-GPU simulations.

3.3.1 Performance of Flood2D-GPU Model

To analyze the performance of the Flood2D-GPU model in the study domain, comparisons were first made between the baseline simulation and the 100-year FEMA

flood map (Zone A/AE; Wing et al. 2017, Gangrade et al. in review). The hydrographs from the DHSVM baseline run were rescaled to match the 100-year peak discharge value for the watershed obtained from the FEMA Flood Insurance Study (FIS) Report, 2017 at USGS gage 08075000. The Flood2D-GPU model outputs driven by the rescaled baseline hydrographs were the compared to the FEMA 100-year floodplain (Figure 7). Visual inspection reveals that the Flood2D-GPU simulation underestimates inundation extent in the upstream part of the watershed. This could be due to the low upstream streamflow predicted by DHSVM. The simulation has the most agreement in the central part of the watershed. However, the Flood2D-GPU simulation overestimates inundation extent in the downstream part of the watershed. This could be due the flat topography in the watershed, the highly urban land cover, which leads to overestimation of flood inundation, backwater effects from the extent of the DEM used in the study, and overestimation of the DHSVM hydrographs in the downstream portion of the watershed.

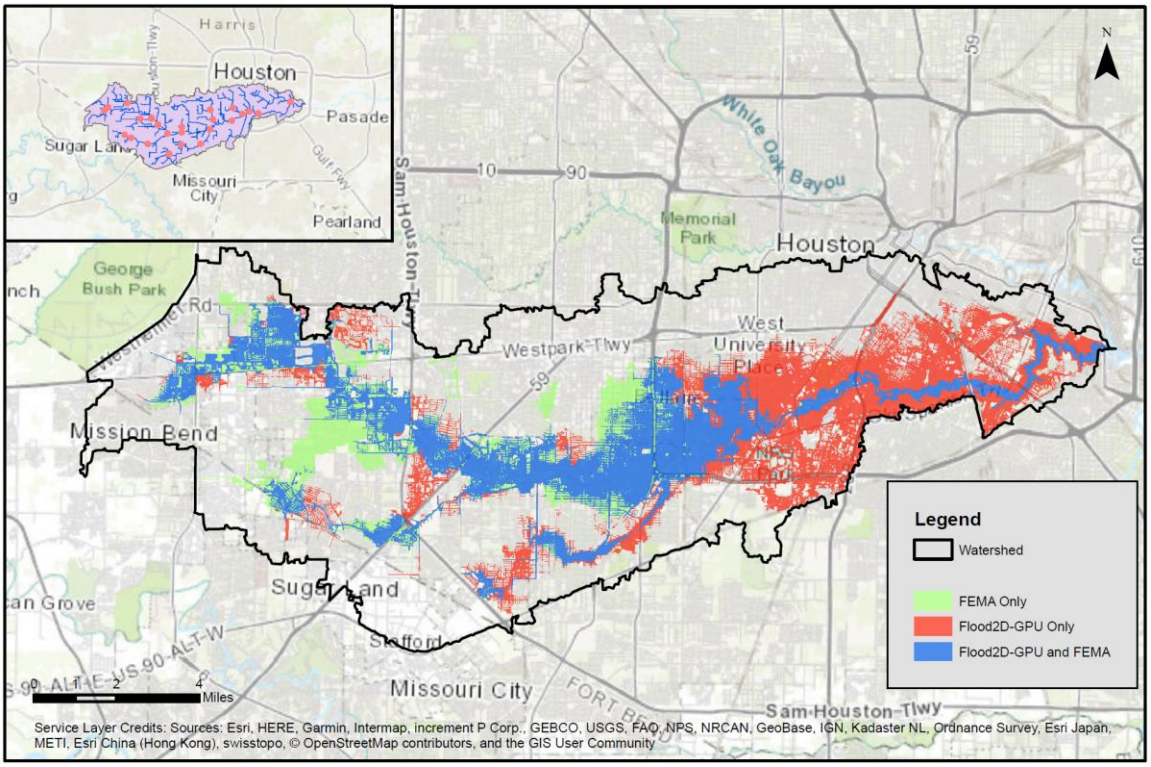

Figure 7 Validation of Flood2D-GPU model run with 100-year FEMA floodplain. Image in upper left shows inflow points used in the simulations.

A binary classification system –as defined in Table 2– was used to identify cells as either flooded or not flooded in the Flood2D-GPU simulation and the FEMA 100-year flood map, respectively. The statistics to analyze the performance of the Flood2D-GPU 100-year flood simulation against the 100-year FEMA flood map are summarized in Table 3. These include the Hit Rate (HR), False Alarm Rate (FAR), Critical Success Index (CSI), and Error (E). The HR indicates the 68% of cells that are correctly identified as flooded. The FAR indicates 47% of the cells are falsely reported as flooded. The CSI measures the overall fit of the spatial data with penalties for under/overprediction. In this simulation, the CSI of 0.42 is lower than desired (i.e., 0.5-

1.0), but is acceptable. The E value of 1.89 indicates that the model has a tendency to overpredict the flooding extent.

Table 2 Contingency table for cell identification.

Cells	Wet in Model (M1)	Dry in Model (M0)
Wet in Baseline (B1)	M1B1	M0B1
Dry in Baseline (B0)	M1B0	M0B0

Table 3 Flood2D-GPU performance statistics compared with 100-year FEMA floodplain and HCFCD maximum inundation map during Hurricane Harvey. Adapted from Wing et al. (2017).

Criterion	**Formula**	**Range**	**Description**	**Comparison with 100-year FEMA Floodplain**	**Comparison with HCFCD maximum inundation map**
Hit rate (HR)	M1B1 / (M1B1+M0B1)	0 – 1	Measure of tendency of model to accurately predict the benchmark flood extents	0.68	0.83
False alarm ratio (FAR)	M1B0 / (M1B0+M1B1)	0 – 1	Measure of tendency to overpredict flood extent	0.47	0.67
Critical success index (CSI)	M1B1 / (M1B1+M0B1 +M1B0)	0 – 1	Measure of fit with penalty for overprediction and underprediction	0.42	0.31
Error (E)	M1B0 / M0B1	0 – infinity	Measure of tendency toward overprediction or underprediction	1.89	9.97

Additionally, statistics were calculated to validate the performance of the Flood2D-GPU simulation during Hurricane Harvey with a maximum inundation map from the Harris County Flood Control District (HCFCD) in Brays Bayou. Results are also summarized in Table 3. While the HR shows good agreement, the FAR is high, CSI is

low, and the error is high. Thus, the Flood2D-GPU model overestimates the flooding extent during Hurricane Harvey as well.

3.4 Methods for Assessing Precipitation, Streamflow, and Floodplain Forecasting Skills

The methods for assessing precipitation, streamflow, and floodplain forecasting skills are summarized in Figure 8. Throughout each assessment, uncertainty analysis (skill score calculations) were conducted in order to demonstrate the uncertainty and skill of each forecast throughout the flood forecasting process.

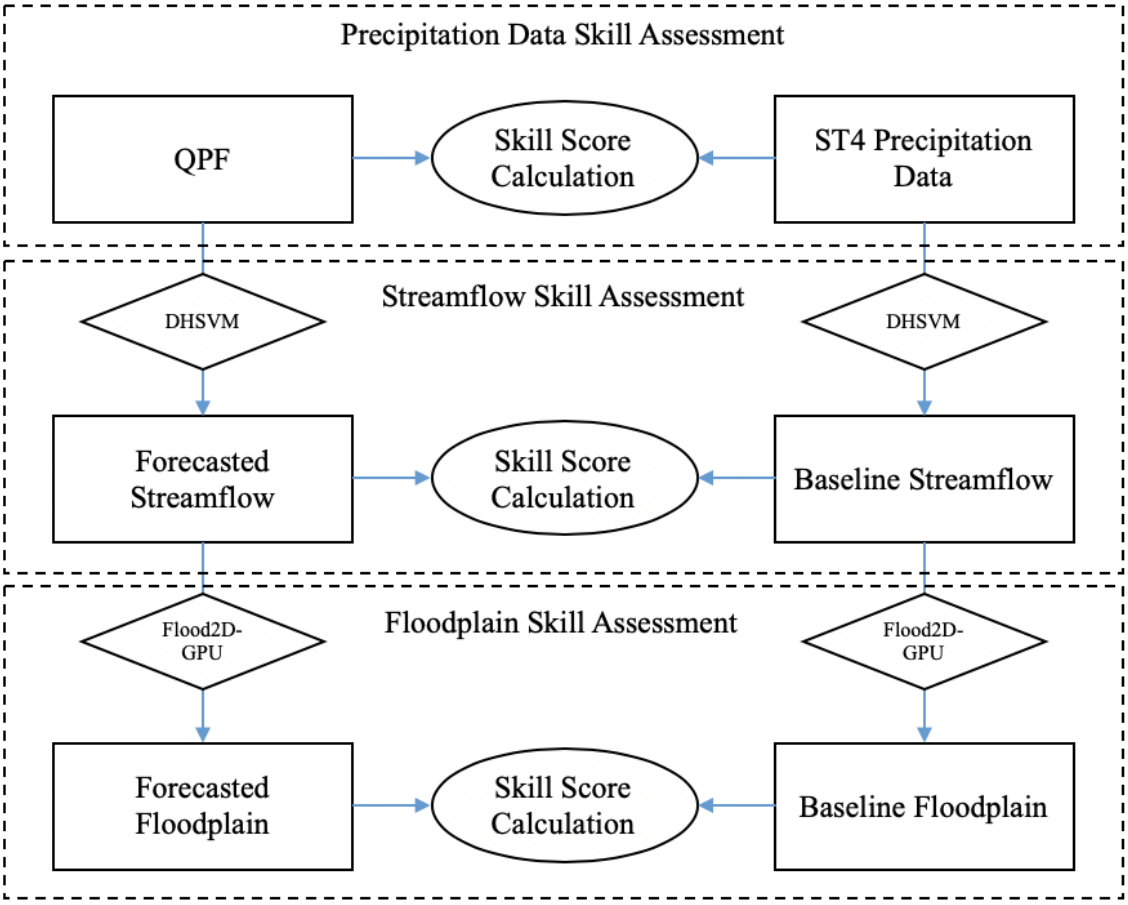

Figure 8 Schematic overview of QPF, hydrological modeling, and hydraulic modeling evaluation procedures.

3.4.1 QPF Skill Quantification to Test for the Accuracy of NWS QPF Data by Lead Time

In this study, the QPFs were first assessed to determine their spatial and temporal accuracy with respect to ST4 QPE data via skill score calculation under scenarios defined by forecast lead time. The QPF data for the years 2015 and 2016 have a spatial resolution of 5 km, while the QPF data for 2017 has a spatial resolution of 2.5 km. In order to be compared with the ST4 QPE data (whose spatial resolution is 4 km), QPF data was compared with the closest ST4 QPE grid cell, based on the distance between the centroids of the grids of both sets of data. Thresholds were set to determine the distance threshold between two grids that were too far apart to be compared with accuracy. The threshold was 2 km (included 89% of points) for 2015-2016 and 1.3 km (73% of points) for 2017. Additionally, the ST4 QPE data was processed to a 6-hourly time step in order to be comparable with the QPF data. Data for the storm events extended from May 23 to May 29, 2015 for the Memorial Day flood, from April 15 to April 21, 2016 for the Tax Day flood, and August 21 to September 3, 2017 for Hurricane Harvey.

The skill scores for precipitation forecasts were adopted from Seo et al. (2018), which include: hit rate (HR), false alarm rate (FAR), and frequency bias (FB; Equation 4). Additionally, the Critical Success Index (CSI) was used as an indicator of the overall performance of the forecasts. These skill scores were determined by counting the number of QPF grid cells identified as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) when compared to the ST4 QPE data, according to the definitions in Table 4. These scores were calculated for each 6-hourly forecast from 6 hours to 84

hours after the forecast was published. The results were used to analyze the effect that lead time has on precipitation forecasting skills.

Table 4 Contingency table after Seo et al. (2018) to determine QPF grid cell designations.

Cells	Wet in Forecast	Dry in Forecast
Wet in Observation	TP	FN
Dry in Observation	FP	TN

$$FB = \frac{TP + FP}{TP + FN} \tag{4}$$

The HR is equivalent to the Probability of Detection (POD) and represents the probability of correctly identifying cells with precipitation, with a value of 1 being optimal. The FAR measures forecast failure and indicates the proportion of incorrectly forecasted grid cells to the total number of grid cells. The FB measures the over/underestimation of precipitation on a spatial scale. The CSI measures the overall performance of the data with a penalty for overprediction and underprediction. Thus, these skill scores determine the accuracy of the spatial distribution of forecasted precipitation but do not assess the accuracy of the forecasted precipitation depth.

In terms of analyzing the skills of the forecasted rainfall depth, the ST4 QPE and QPF data were compared for each time step by lead time in a graphical manner. This allows for visually representing the overestimation/underestimation of the QPF data by lead time.

3.4.2 Streamflow Forecast Design and Methods to Test the Effect of Lead Time on Streamflow Forecasts

In this section, the design of the streamflow forecasts is described as well as the methods to assess forecasting skill by lead time.

3.4.2.1 Streamflow Forecast Design

The QPF precipitation were used to construct various lead time forcing scenarios for driving DHSVM, with the resulting streamflow mimicking the five-day forecasts produced by the WGRFC. This is because the operational WGRFC forecasts for Brays Bayou only included 72-hour forecasts during Hurricane Harvey, and 12- and 24-hour forecasts during the Memorial Day and Tax Day flood events. By generating a suite of results at a full spectrum of lead times, the optimal lead time(s) for an event can be identified. Specifically, DHSVM simulations were set up using 10 days of previous observed precipitation, forecasted precipitation for the duration of forecast interest, and the remaining time steps have no precipitation for a total of 5 days. Table 5 uses Hurricane Harvey as an example to show the streamflow forecast set-up for each lead time from 6 to 72 hours (at a 6-hour interval) for forecasts from 8/25/2017-01 to 9/1/2017-00. Similarly, forecasts under each of these 12 lead time scenarios were also generated from 5/23/2015 to 5/29/2015 for the Memorial Day flood, from 4/15/2016 to 4/21/2016 for the Tax Day flood, and from 8/21/2017 to 9/3/2017 for Hurricane Harvey. These forecasts were compared directly with the WGRFC forecast outputs and are described in Section 3.4.4.

During the time periods listed above for each storm event, the DHSVM forecasts were synthesized to produce hydrographs composed fully of forecasts with consistent

lead times. This was achieved by constructing new hydrographs out of only the forecasted time periods of the hydrographs, not including the periods driven by ST4 QPE data or the 0 precipitation data. Additionally, the hydrographs were pieced together with consistent lead times such that at the end of 1 forecast, another begins. For example, in the 6-hour lead time hydrograph, the first 6 hours are from forecast 1 followed by the next 6 hours from forecast 2 … until the end of the time period. These synthesized hydrographs allowed for analysis that isolated the forecasted lead time.

Table 5 Modeling of forecasted streamflow design using Hurricane Harvey as an example. All dates are for the year 2017 and hours are included on pertinent dates for disambiguation.

Forecast Lead Time (hours)	Dates of Observed Precipitation	Dates of Forecasted Precipitation	Dates of 0 Precipitation
6	08/15-01 to 08/25-00	08/25-01 to 08/25-06	08/25-07 to 08/30-00
12	08/15-01 to 08/25-00	08/25-01 to 08/25-12	08/25-13 to 08/30-00
18	08/15-01 to 08/25-00	08/25-01 to 08/25-18	08/25-19 to 08/30-00
24	08/15-01 to 08/25-00	08/25-01 to 08/26-00	08/26-01 to 08/30-00
30	08/15-01 to 08/25-00	08/25-01 to 08/26-06	08/26-07 to 08/30-00
36	08/15-01 to 08/25-00	08/25-01 to 08/26-12	08/26-13 to 08/30-00
42	08/15-01 to 08/25-00	08/25-01 to 08/26-18	08/26-19 to 08/30-00
48	08/15-01 to 08/25-00	08/25-01 to 08/27-00	08/27-01 to 08/30-00
54	08/15-01 to 08/25-00	08/25-01 to 08/27-06	08/27-07 to 08/30-00
60	08/15-01 to 08/25-00	08/25-01 to 08/27-12	08/27-13 to 08/30-00
66	08/15-01 to 08/25-00	08/25-01 to 08/27-18	08/27-19 to 08/30-00
72	08/15-01 to 08/25-00	08/25-01 to 08/28-00	08/28-01 to 08/30-00

3.4.2.2 Streamflow Forecast Skill Quantification

A DHSVM baseline run (driven by ST4 QPE) for each storm event was compared with forecasted streamflow outputs to assess the streamflow forecasting skill under different lead times. The skill statistics that were employed include R^2, NSE, RB of the

mean streamflow, Relative Root Mean Square Error (RRMSE; Equation 5), RB of peak streamflow, and mean streamflow.

$$RRMSE = \frac{1}{\bar{O}}\sqrt{\frac{\sum_t (S_t - O_t)^2}{n}}, \quad (5)$$

where S_t represents the simulated value and O_t represents the observed value at time t, $\bar{O}$ represents the average of the observed values over the whole time series, and n represents the number of observations.

3.4.3 Forecasted Inundation Mapping Skill Quantification

Forecasted inundation maps were generated only for Hurricane Harvey, not for the other two events (Memorial Day and Tax Day floods). This is due to the lack of improvement in skill of the streamflow forecasts as the lead time was extended. Thus, it can be assumed that the floodplain forecasts would not have improved skill with increased lead time either. The skill of the forecasted floodplains during Hurricane Harvey were assessed against a baseline floodplain simulate by the Flood2D-GPU (with input hydrographs adopted from DHSVM results driven by ST4 QPE). The metrics used to quantify the skill of the forecasts include hit rate (HR), false alarm rate (FAR), critical skill index (CSI) and error (E), as described in Section 3.3.1.

3.4.4 Streamflow Forecast Comparison with WGRFC Forecasts

The forecasted streamflow results from the DHSVM model were compared against those from the lumped CHPS-FEWS model utilized by the NWS WGRFC to generate flooding forecasts during the three flood events. Output from the CHPS-FEWS model can be accessed through the NWS WGRFC website. Forecasts are only published when the streamflow approaches the Action Stage (for Brays Bayou, this is 38 feet; Kris

Lander, personal communication, March 7, 2019). Thus, there were 2, 6, and 18 forecasts for Brays Bayou (USGS station 08075000) during the Memorial Day flood, the Tax Day flood, and Hurricane Harvey, respectively. It is important to note that the forecasts reflect manual adjustments made by the NWS forecasters. One common adjustment is to utilize the observed precipitation (ST4 QPE) for the first time step, rather than the forecasted precipitation (QPF). Therefore, the modeler has the advantage of using the observed precipitation rather than the forecast. In this study, QPF's were used for each timestep in to drive the DHSVM model.

The comparisons of the DHSVM streamflow forecasts with the available WGRFC forecasts during the three storm events help evaluate if WGRFC had selected the appropriate forecast lead time in each case. Because WGRFC operational forecasts only provided results at one given lead time per forecast, the suite of DHSVM results (with set up described in Table 5) can fill in the missing scenarios to facilitate such analysis. Skill statistics (including R^2, NSE, and RB) are calculated for 12-hour, 24-hour, 48-hour, and 72-hour forecasts simulated via DHSVM versus the gage data, the 72-hour DHSVM forecasts versus the 72-hour RFC forecasts, and the gage data versus the 72-hour RFC forecasts. Thus, the lead time of the best streamflow forecasts can be identified.

4. RESULTS

In this section, the results of the precipitation, streamflow, and inundation analyses as well as the comparison with the WGRFC forecasts are presented.

4.1 QPF Skill Quantification Results by Lead Time

The QPF skill statistics against ST4 QPE for the Memorial Day flood, the Tax Day flood, and Hurricane Harvey in Brays Bayou are shown in Figure 9. QPFs from the Memorial Day flood have the worst skill, while the Tax Day flood and Hurricane Harvey have better and generally comparable skill statistics. This is likely due to the majority of the precipitation from the Memorial Day flood occurring over a time period of 10 hours, while the other two events were more prolonged (23+ hours for the Tax Day flood and 108 hours for Hurricane Harvey). Inspection of the QPF and ST4 QPE data revealed that the forecasts for the Memorial Day flood underestimated the rainfall intensity and instead forecasted a more prolonged event with less rainfall at each timestep. In general, Figure 9 shows that, as lead time increases, the QPF skill statistics tend to worsen. One notable exception is the HR: as lead time increases, the hit rate approaches 1. This is likely due to the uncertainty in the forecasts at longer lead times. These forecasts tend to overestimate the number of timesteps with precipitation, as evidenced in the general increase in FB with lead time. However, the FAR, FB, and CSI generally worsen with increased lead time. The values of CSI above 0.5 indicate that more than 50% of the precipitation was correctly forecasted in a spatial sense. The Tax Day flood and Hurricane Harvey have CSI values above 0.5 for almost the entire forecast horizon, while the Memorial Day flood only has a CSI value above 0.5 at the 6-hour lead time. Therefore, for the events of longer duration, the QPFs are more skillful than for those of shorter duration.

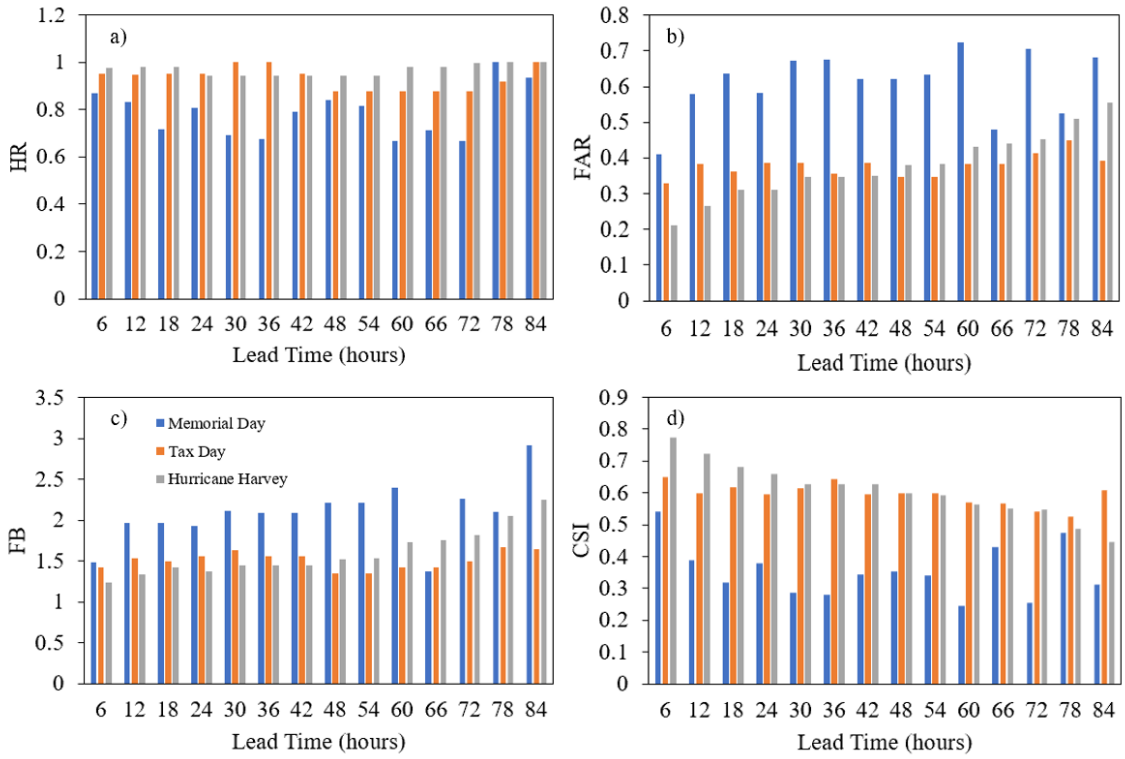

Figure 9 a) Hit rate (HR), b) false alarm rate (FAR), c) frequency bias (FB), and d) critical success index (CSI) of QPF data for the dates of the Memorial Day flood, Tax Day flood, and Hurricane Harvey.

Because the precipitation during Hurricane Harvey lasted for such a prolonged period of time, it was of interest to determine if the beginning, middle, and end of the event showed varying forecasting skills. The event was thus divided into three time periods (beginning: August 21 to 24, middle: August 25 to 29, and end: August 30 to September 3), and the skill statistics were calculated for each time period. Figure 10 shows that the middle of the event had the best skill statistics, while the beginning and end have the worst skill statistics. This indicates that, as the event progressed and it became evident that the storm was going to last multiple days, the precipitation forecasts were more accurate in terms of timing. This could be indicative of accuracy in the

climatic models coupled with the forecaster's manual adjustments used to produce the QPFs. It is interesting to note that in the last period, the QPF generally overpredicted the time periods with precipitation. Thus, the FB for the latter lead times is generally high and increasing, with the exception of the 84-hour forecast. Additionally, the CSI shows very desirable values (and thus skillful forecasts) during the middle period only. The CSI values are low at the beginning of the event and are 0 during the end of the event due to the lack of precipitation (the numerator of the calculation is 0).

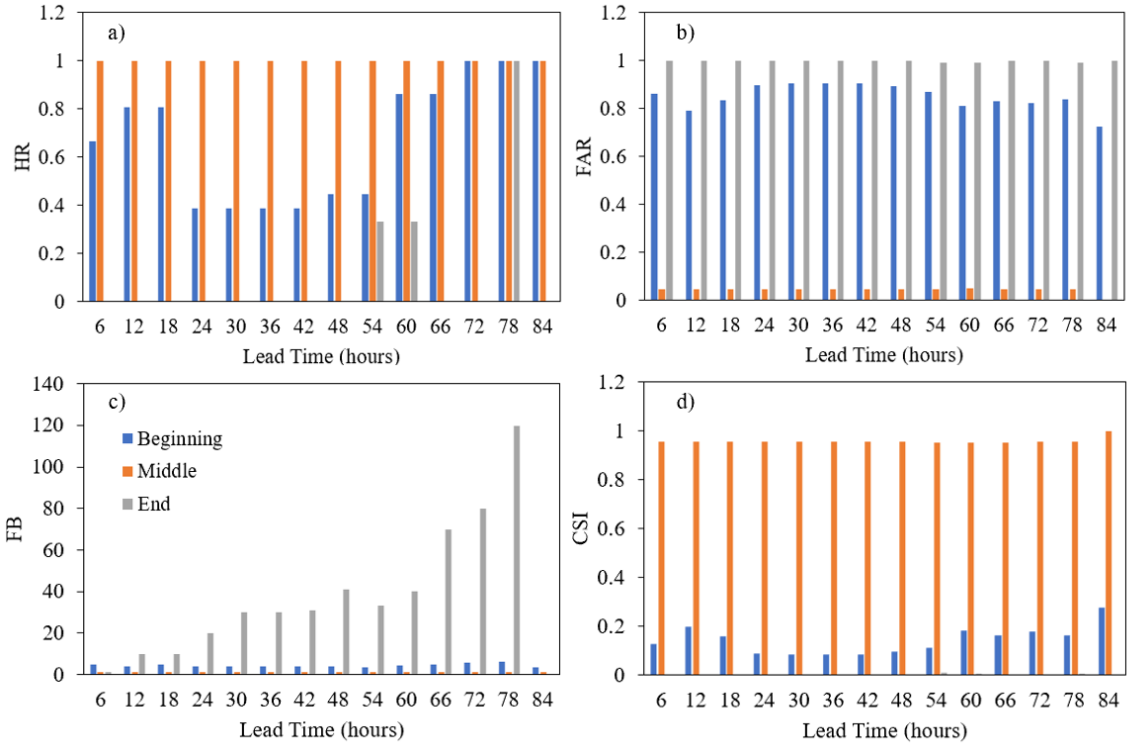

Figure 10 a) Hit rate (HR), b) false alarm rate (FAR), c) frequency bias (FB), and d) critical success index (CSI) for three periods of Hurricane Harvey event (Beginning, Middle, and End).

Figures 11-13 compare the forecasted precipitation amount with the ST4 QPE under different lead times (i.e., 12-, 24-, 48-, and 72-hour) for each storm event. The

regression line of best fit in each case shows that the QPF, on average, underestimates the amount of rainfall for each 6-hourly timestep and does not capture the magnitude of the peak rainfall. This is in agreement with Sukovich et al. (2014), which found that the NWS QPFs for the contiguous United States underestimated rainfall in events of extreme precipitation. In fact, the negative bias is more pronounced for extreme precipitation events than for events of lesser magnitude. Furthermore, the regression line of best fit generally declines as the lead time increases, which indicates that the magnitude of the precipitation forecast is less accurate as lead time increases.

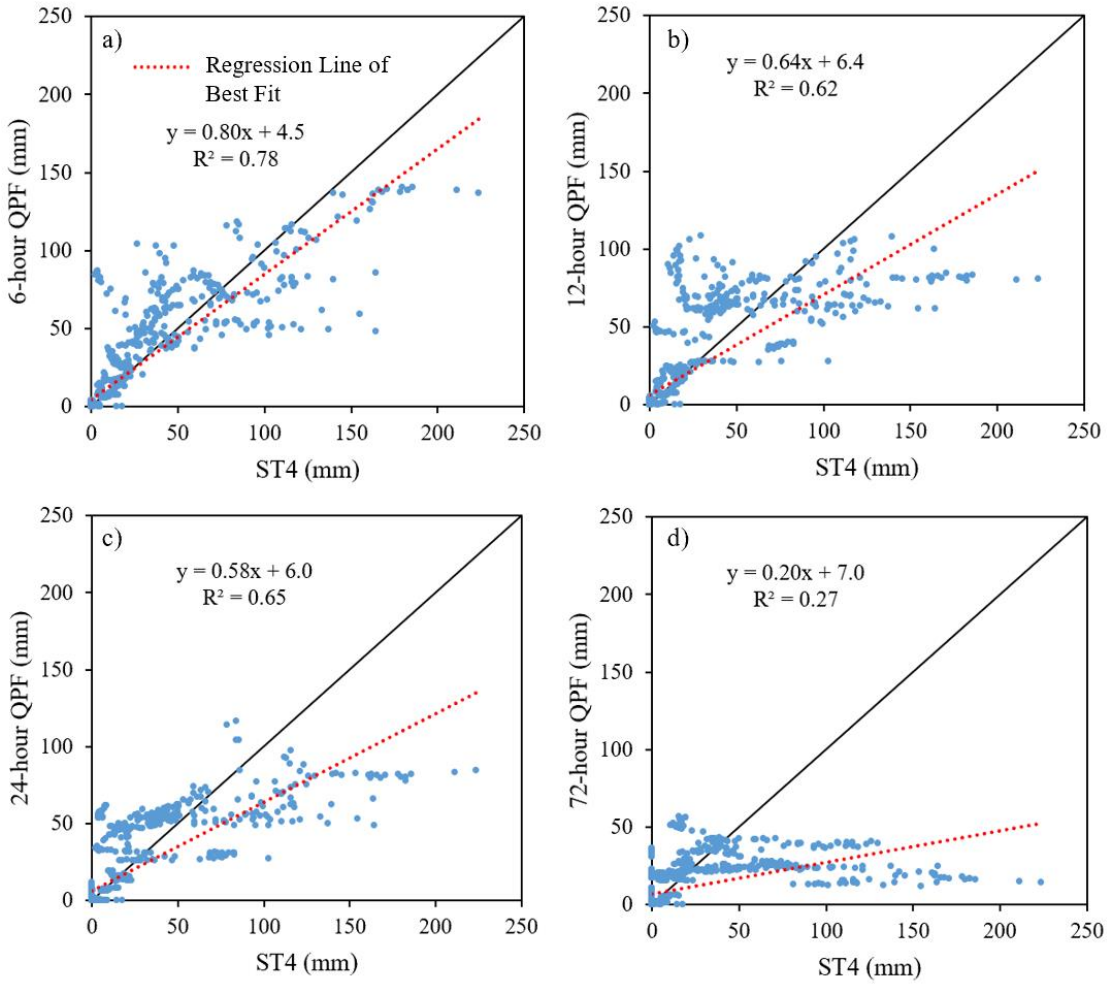

Figure 11 Discrepancy in precipitation amounts for Hurricane Harvey by lead time: a) 6-hour, b) 12-hour, c) 24-hour, and d) 72-hour lead times. Each data point represents 6-hour accumulated precipitation.

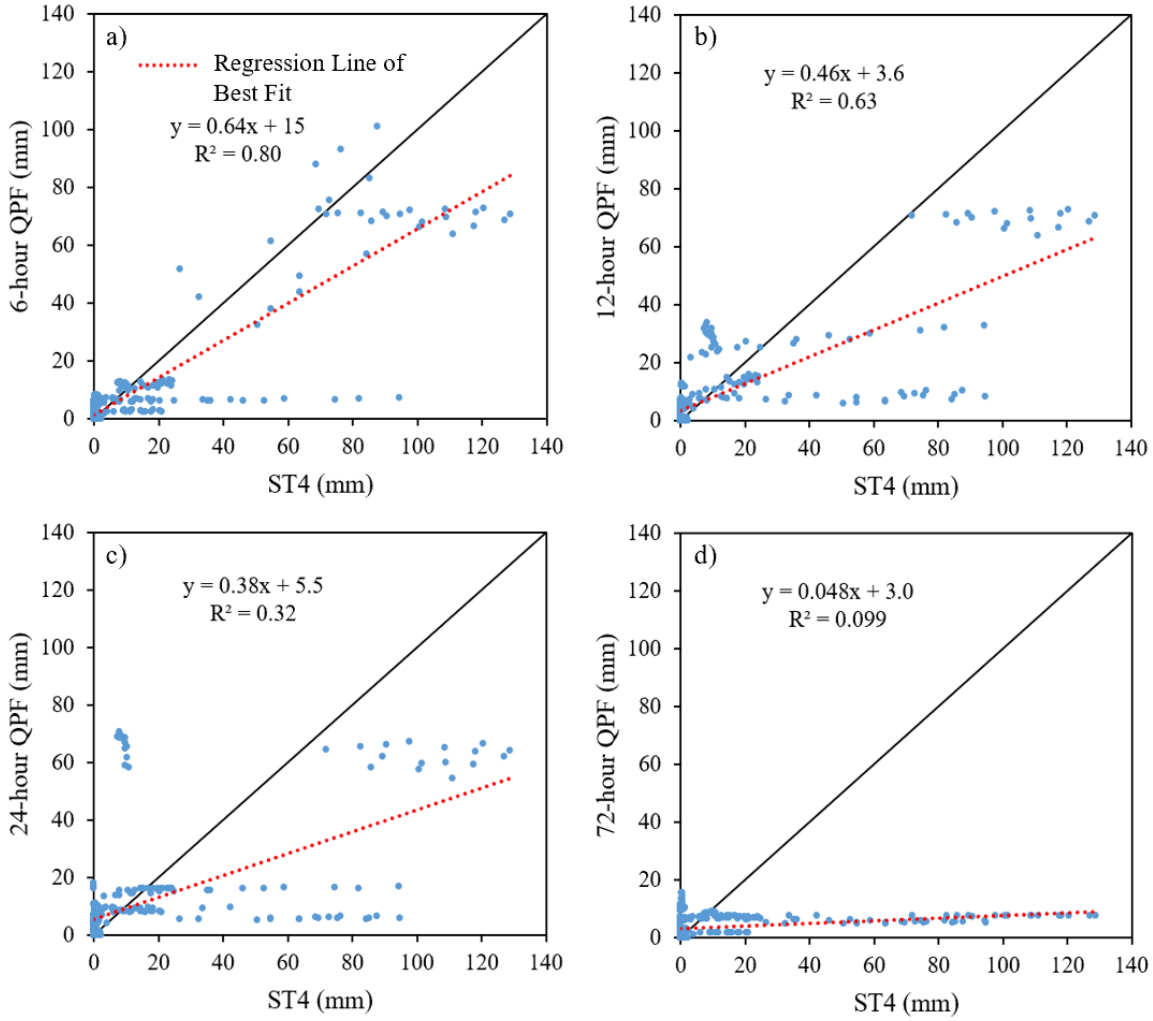

Figure 12 Discrepancy in precipitation amounts for the Tax Day flood by lead time: a) 6-hour, b) 12-hour, c) 24-hour, and d) 72-hour lead times. Each data point represents 6-hour accumulated precipitation.

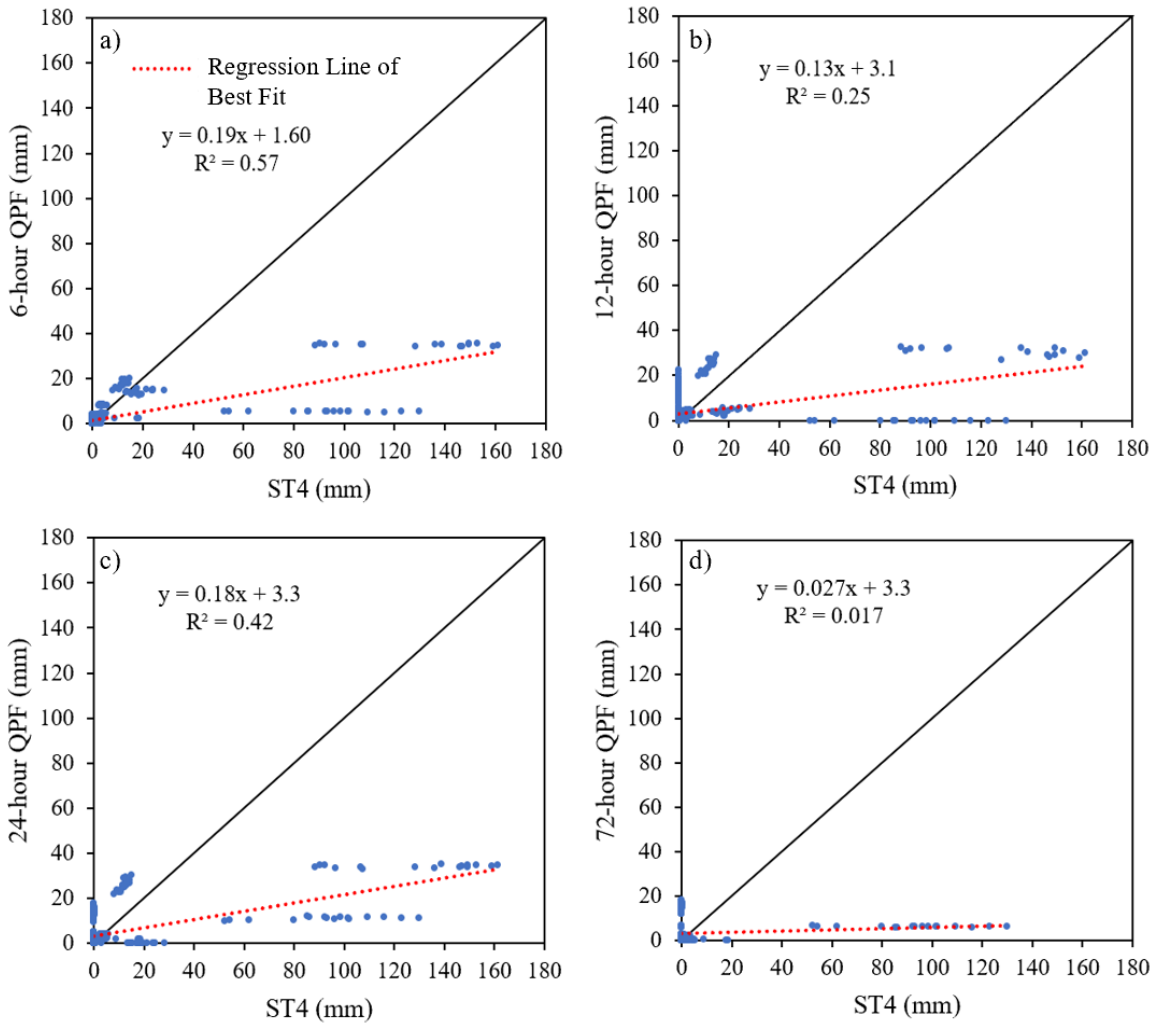

Figure 13 Discrepancy in precipitation amounts for the Memorial Day flood by lead time: a) 6-hour, b) 12-hour, c) 24-hour, and d) 72-hour lead times. Each data point represents 6-hour accumulated precipitation.

The R^2 values for the 6-hour forecast for both Hurricane Harvey and the Tax Day flood are very high (0.78 and 0.80 respectively), indicating that the fit between the QPF and ST4 precipitation data is good. However, the best R^2 value for the Memorial Day flood is 0.57, also for the 6-hourly forecast. This indicates that even the best Memorial Day flood forecast is not as accurate as for the other events. Figure 13 shows that the rainfall is most underpredicted in the case of the Memorial Day flood, even in the

shortest-range forecasts. Again, this can be attributed to the short duration and intensity of the event that was not captured by the QPFs, which predicted a longer event of smaller intensity. Thus, it is the most difficult event to capture by the precipitation forecasts.

Overall, the results of the precipitation analysis indicate that the forecasts with shorter lead times generally produce more accurate QPF in terms of timing and amount of rainfall. Additionally, the events with longer duration tend to have higher forecasting skills. However, the NWS QPFs as a whole tend to underestimate the amount of rainfall for extreme storm events.

4.2 Streamflow Forecast Skill Quantification Results by Lead Time

Figure 14 shows the skill statistics for the streamflow forecasts by lead time when compared with the baseline model run for each storm event. The skill statistics at the two USGS gages (08074810 and 08075000; 08075110 has an incomplete dataset and thus will not be analyzed in this section) follow largely the same trends. The skill statistics generally worsen as the lead time increases. This is in agreement with the worsening of precipitation QPF skill statistics with increased lead time. The underestimation of the QPFs is also reflected in the streamflow forecasts, particularly in the relative bias statistics. However, since the streamflow forecasts are generated with ST4 QPE prior to the forecasted precipitation, the RB is higher than it would be if QPFs were used for the entire simulation. Additionally, the skill statistics for Hurricane Harvey are the best, while the Memorial Day flood skill statistics are the worst. This is also a reflection of the quality of the QPFs for those two events. Further, the finer spatial resolution of the QPF data for Hurricane Harvey likely contributes to the higher skill statistics of the streamflow forecasts.

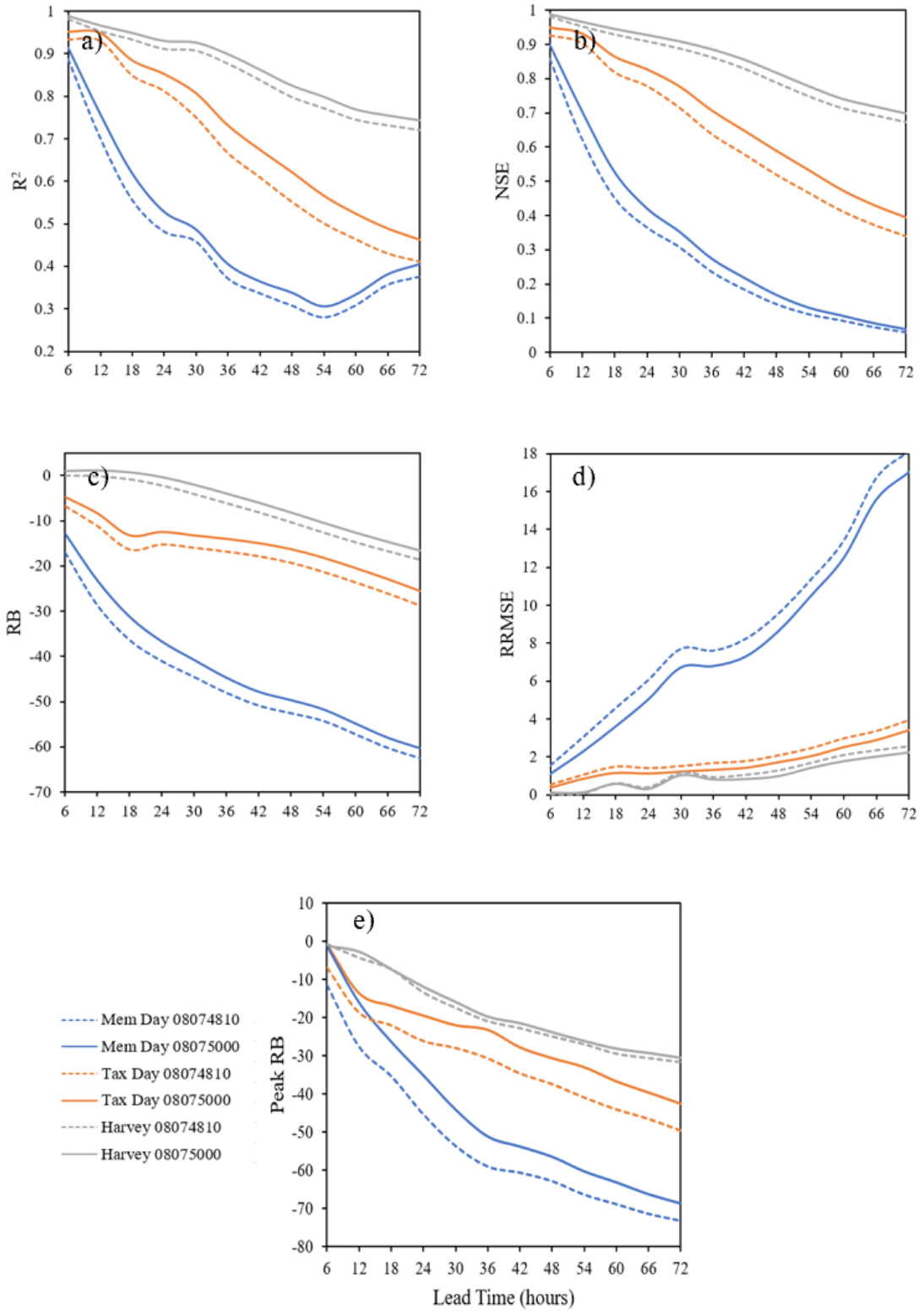

Figure 14 Skill statistics of the forecasted streamflow by lead time for gages 08074810 and 08075000: a) R-Squared (R^2), b) Nash-Sutcliffe Efficiency (NSE), c) Relative Bias (RB), d) Relative Root Mean Square Error (RRMSE), and e) Peak Relative Bias (Peak RB).

The correlation statistics (R^2 and NSE) indicate that the very short-term forecasts perform fairly well and then drop off significantly thereafter (Figure 14a and b). Moriasi et al. (2007) note that R^2 values greater than 0.5 are acceptable for watershed simulations and NSE values greater than 0 are acceptable. By those standards for the R^2, the forecasts for the Memorial Day flood are acceptable only for the 6 to 24 hour forecasts and unacceptable for longer forecasts, while the forecasts for the Tax Day flood are acceptable from 6 to 60 hours and are unacceptable thereafter (gage 08075000; Figure 14a). The forecasts for Hurricane Harvey are acceptable during the entire forecast horizon (Figure 14a). In addition, the NSE values for all of the events are greater than 0 and are thus considered acceptable (Figure 14b). The RB is negative for all lead times except for 6-18 hour lead times at gage 08075000. This positive RB is due to forecast overestimation of streamflow at the beginning and end of the event. The peak RB is negative for all of the flooding events, indicating that the peaks are underestimated by the streamflow forecasts at all lead times (Figure 14e). Figure 14d shows that the RRMSE becomes quite large for the Memorial Day flood, reflecting the loss in forecast accuracy with increased lead time. The RRMSE also increases for the Tax Day flood and Hurricane Harvey, but to a lesser extent.

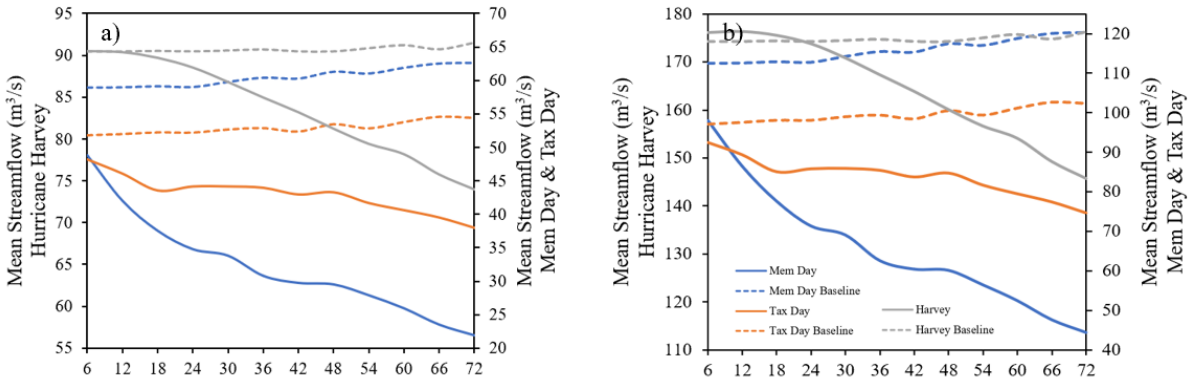

Figure 15 Forecasted mean streamflow compared with baseline mean streamflow for USGS stations a) 08074810 and b) 08075000.

Figure 15 shows the forecasted mean streamflow as compared to the baseline model mean streamflow for each time period for the two USGS gage locations. The mean streamflow for the Tax Day and Memorial Day floods are underestimated for all lead times at both gage locations. For Hurricane Harvey, the mean streamflow is underestimated at all lead times for gage 08074810. However, the mean streamflow is overestimated at short lead times (6-18 hours) and is underestimated for the longer lead times at gage 08075000. This suggests that the negative bias of the QPF data is driving the hydrological simulations as the lead time gets longer, as forecasts at longer lead times are driven by longer time periods of low QPF data when compared to the shorter lead times that are driven by fewer time periods of the QPF data.

Therefore, it is concluded that the negative bias of the QPF data results in underestimated streamflow forecasts. This negative bias is exacerbated in forecasts as the lead time increases. This is because those streamflow forecasts are more affected by the accumulation of the QPF errors as time goes on, while the benefit of initializing the model using the ST4 QPE data becomes less significant. The forecasted streamflow for

extended precipitation events have the best skill statistics (Hurricane Harvey), whereas the shorter-term events have the worst skill statistics (Memorial Day flood). This can also be most likely attributed to the higher skill of the QPF data for extended events as well as the higher spatial resolution of the QPF data for Hurricane Harvey.

4.3 Results from Inundation Forecasts in Terms of Lead Time

Selected streamflow forecasts with 12- and 72-hour lead times were used to drive the Flood2D-GPU model to generate floodplains at three points during Hurricane Harvey: before the event began, the rising limb, and the flood peak. Thus, the effect of extending the lead time during extreme events could be analyzed at each of the three points during the event. These 8 forecasted inundation cases are summarized in Table 6, among which the control run (Case 2) – whose streamflow is based on ST4 QPE data – provides a baseline against which the other forecasts were compared.

Table 6 Inundation forecast cases used in Flood2D-GPU.

Case	Description	Forecast Duration (hours)	Hydrograph Dates
1	100-year flood	-	-
2	Hurricane Harvey control run	-	08/22-03 to 09/01-03
3	Forecast beginning from 08/25/2017-01	12	08/20-12 to 08/30-12
4	Forecast beginning from 08/25/2017-01	72	08/22-12 to 09/01-12
5	Forecast beginning from 08/26/2017-01	12	08/21-12 to 08/31-12
6	Forecast beginning from 08/26/2017-01	72	08/22-06 to 09/01-06
7	Forecast beginning from 08/27/2017-01	12	08/22-06 to 09/01-06
8	Forecast beginning from 08/27/2017-01	72	08/22-06 to 09/01-06

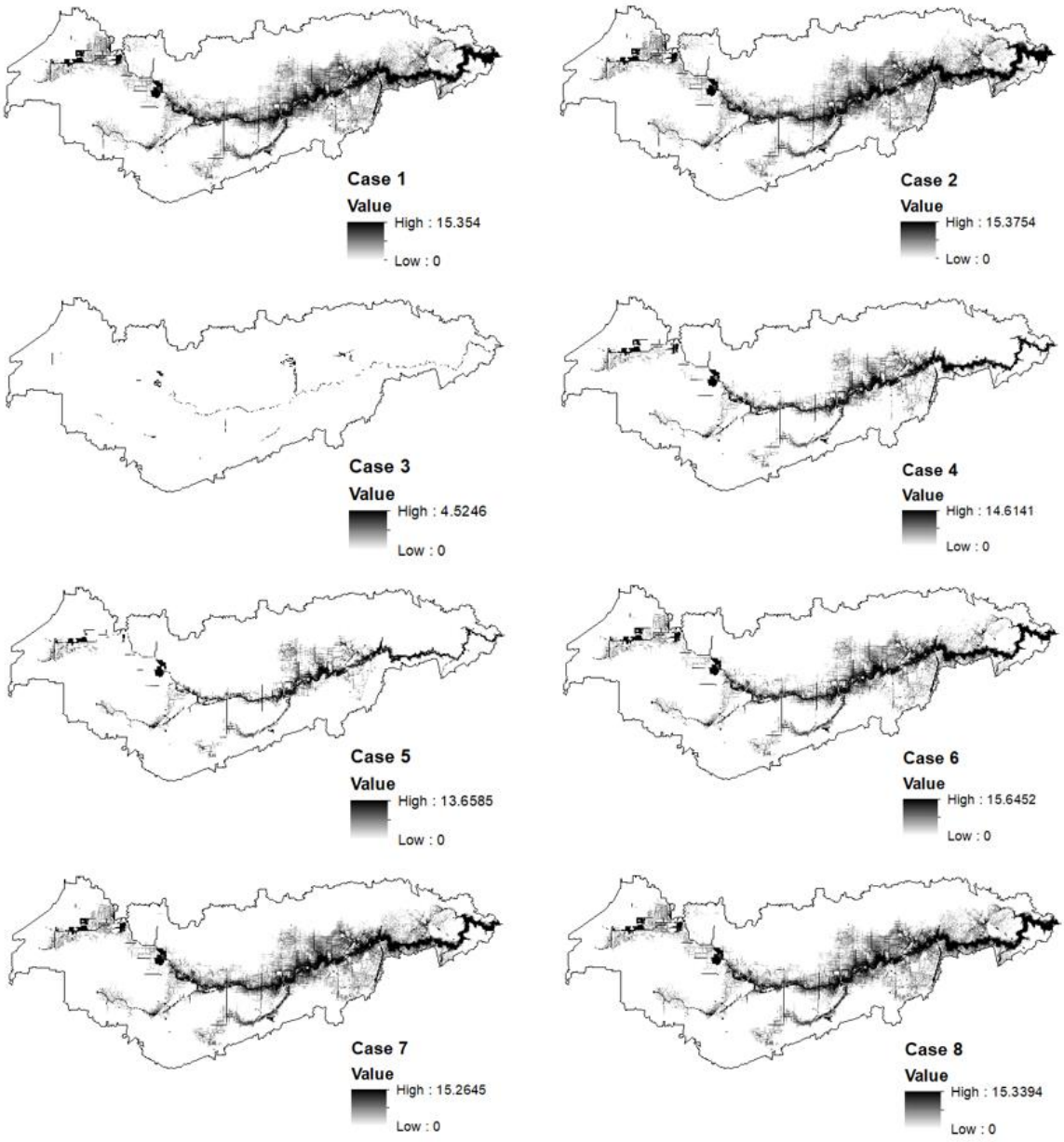

Figure 16 Results of inundation model. Cases correspond to Table 6.

Table 7 Evaluation metrics for forecasted floodplains using Case 2 as benchmark.

Criterion	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
Hit rate (HR)	0.02	0.47	0.33	0.70	0.87	0.90
False alarm ratio (FAR)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
Critical success index (CSI)	0.023	0.47	0.33	0.70	0.87	0.90
Error (E)	0.0000	0.0000	0.0000	0.0001	0.0001	0.0005

According to Case 3 (12-hour forecast) in Figure 16, it can be seen that the inundation model did not predict flooding, as the streamflow forecast was so low. However, extending the forecast to 72 hours (Case 4) produced a more useful depiction of the flooding to come (Figure 16). For each forecast, the hit rate (HR) and critical success index (CSI), increased significantly with the extension of the forecast from 12 to 72 hours. However, the FAR and E are very close to 0. This is due to the underestimated streamflow forecasts when compared to the baseline streamflow. Thus, cells did not fill with water in the forecasts that were not filled with water in the baseline. Thus, it is clear from the above results that increasing the lead time of the forecasts during Hurricane Harvey produced skillful results in the inundation forecasts throughout the event. However, it would be useful to produce forecasted floodplains for the end of the event as well to determine if the forecasts are still skillful with increased lead time as the event ends.

4.4 Comparison between WGRFC Streamflow Forecasts and DHSVM Forecasts to Determine Optimal Lead Time

To compare the DHSVM streamflow forecasts with the WGRFC streamflow forecasts produced by the CHPS-FEWS model (hereafter referred to as RFC forecasts),

17 72-hour streamflow forecasts for Brays Bayou during Hurricane Harvey were obtained from the WGRFC website (Table 8). The 18th available RFC forecast is a 24-hour forecast and, for consistency, were not be evaluated in this section. Streamflow under alternate lead times (12-, 24-, and 48-hours) were simulated using the DHSVM model to analyze the effect of extended forecasts on the forecasts' accuracy and usefulness during extreme events. The available forecasts for the Memorial Day flood and the Tax Day flood were also obtained and analyzed; however, due to their short duration, the statistical skill of the forecasts was not improved as lead time increased beyond 12 hours. Thus, this section focuses on analyzing the forecasts for Hurricane Harvey.

Table 8 WGRFC 72-hour forecasts for Brays Bayou during Hurricane Harvey.

Forecast Number	Issuance Date (CDT)	Starting Time of Forecasted Precipitation (CDT)
1	08/25 at 02:15 pm	08/25 at 12:00 pm
2	08/25 at 08:36 pm	08/25 at 06:00 pm
3	08/26 at 02:01 am	08/26 at 12:00 am
4	08/26 at 08:30 am	08/26 at 06:00 am
5	08/26 at 02:23 pm	08/26 at 12:00 pm
6	08/26 at 03:09 pm	08/26 at 12:00 pm
7	08/26 at 08:30 pm	08/26 at 06:00 pm
8	08/26 at 11:38 pm	08/26 at 06:00 pm
9	08/27 at 07:55 am	08/27 at 06:00 am
10	08/27 at 02:17 pm	08/27 at 12:00 pm
11	08/27 at 08:51 pm	08/27 at 06:00 pm
12	08/28 at 02:21 am	08/28 at 12:00 am
13	08/28 at 08:27 am	08/28 at 06:00 am
14	08/28 at 01:55 pm	08/28 at 12:00 pm
15	08/28 at 08:01 pm	08/28 at 06:00 pm
16	08/29 at 02:45 am	08/29 at 12:00 am
17	08/29 at 08:22 am	08/29 at 06:00 am

Figure 17 depicts the results of the 17 forecasts (Table 8) from a total of six pairs of comparisons (four comparing gage data with DHSVM forecasts at different lead times,

one comparing gage data with RFC forecasts at a 72-hour lead time, and one comparing the 72-hour DHSVM forecasts with the 72-hour RFC forecasts) in terms of the R^2, NSE, and RB, respectively. The last row of each figure represents the statistics for the 72-hour RFC forecast versus the 72-hour DHSVM forecast. Since the statistics are very good (high R^2 and NSE, low RB) for each forecast in this scenario, it suggests that the DHSVM forecasts can be used as a substitute for the RFC forecasts at lead times that were not provided by the RFC (12-, 24-, and 48-hour forecasts).

The results in Figure 17 indicate that generally, at the beginning and middle of the event, the R^2 and NSE increase with increased lead time. Towards the end of the event (starting around forecast 12), the R^2 and NSE generally decrease as lead time increases. From the beginning to the middle of the event, the R^2 and NSE first increase, then decrease, and later improve again at the end of the event. This is likely due to the duration of the event: the QPF performed well in the middle of the event and had very high frequency bias towards the end of the event. At the very end of the event, when it was apparent that the precipitation had ended, the streamflow forecasts perform very well.

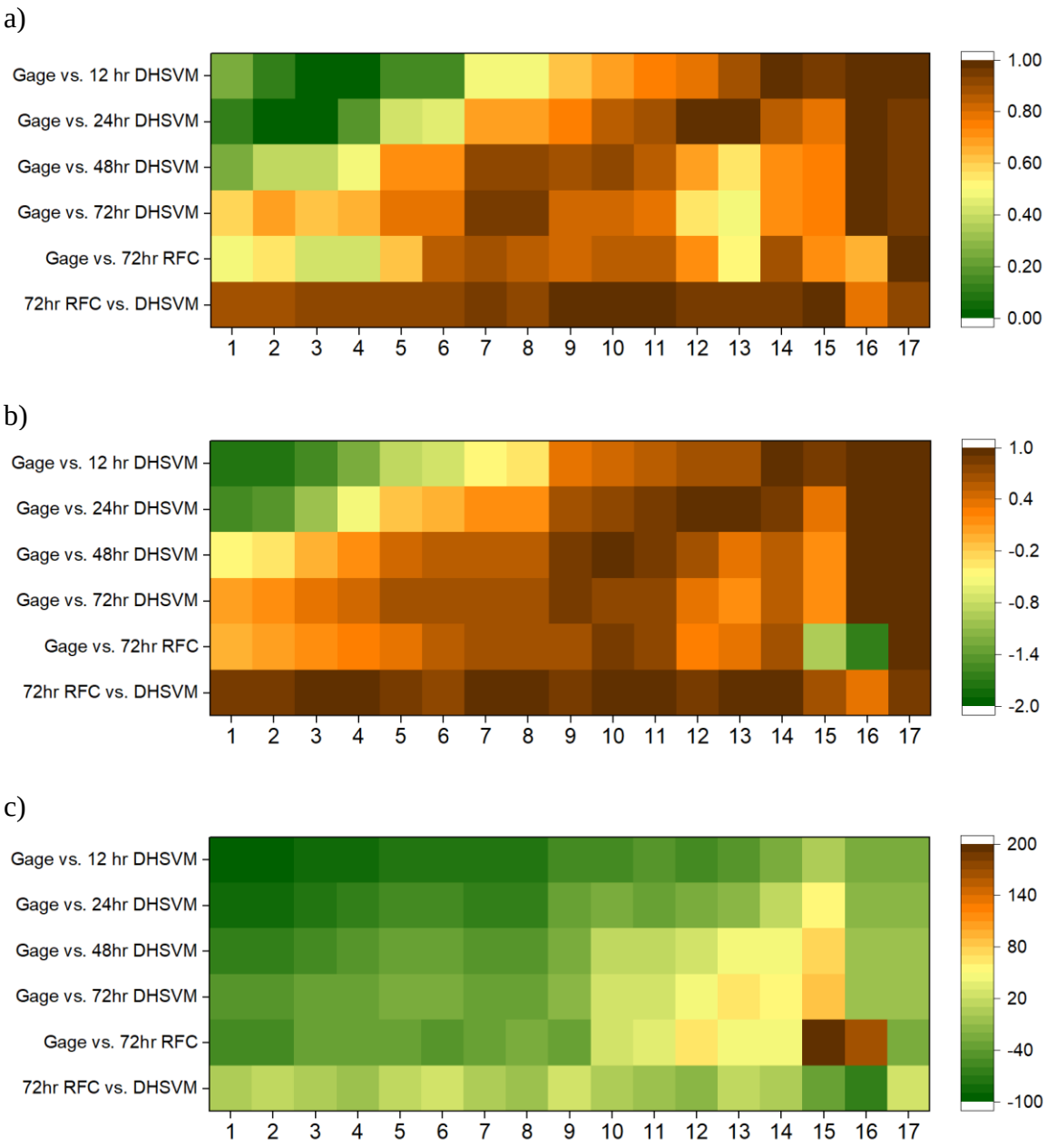

Figure 17 a) R^2, b) NSE, and c) Relative Bias (RB) for comparison of forecasts with different lead times during Hurricane Harvey by issuance time. The x-axis refers to the RFC forecasts (Table 8).

In Figure 17c, the trend of the RB is such that, in general, the DHSVM and RFC forecasts are negatively biased in comparison to the gage data in the beginning and middle of the event. At the end, the models are positively biased. This is likely due to

QPF underestimation of the event in the beginning and middle of the event, and overestimation of precipitation frequency at the end of the event, as evidenced by the high FB of the QPF data at the end of the event. Bias is generally lowest for shortest lead times. This is likely due to the forecast design: after the short period of forecasted precipitation data, the rest of the 5-day period has 0 precipitation. From Figure 18, it is apparent that the QPF underestimated the volume of rainfall, as the forecasted streamflow is mostly underestimated, including the peaks.

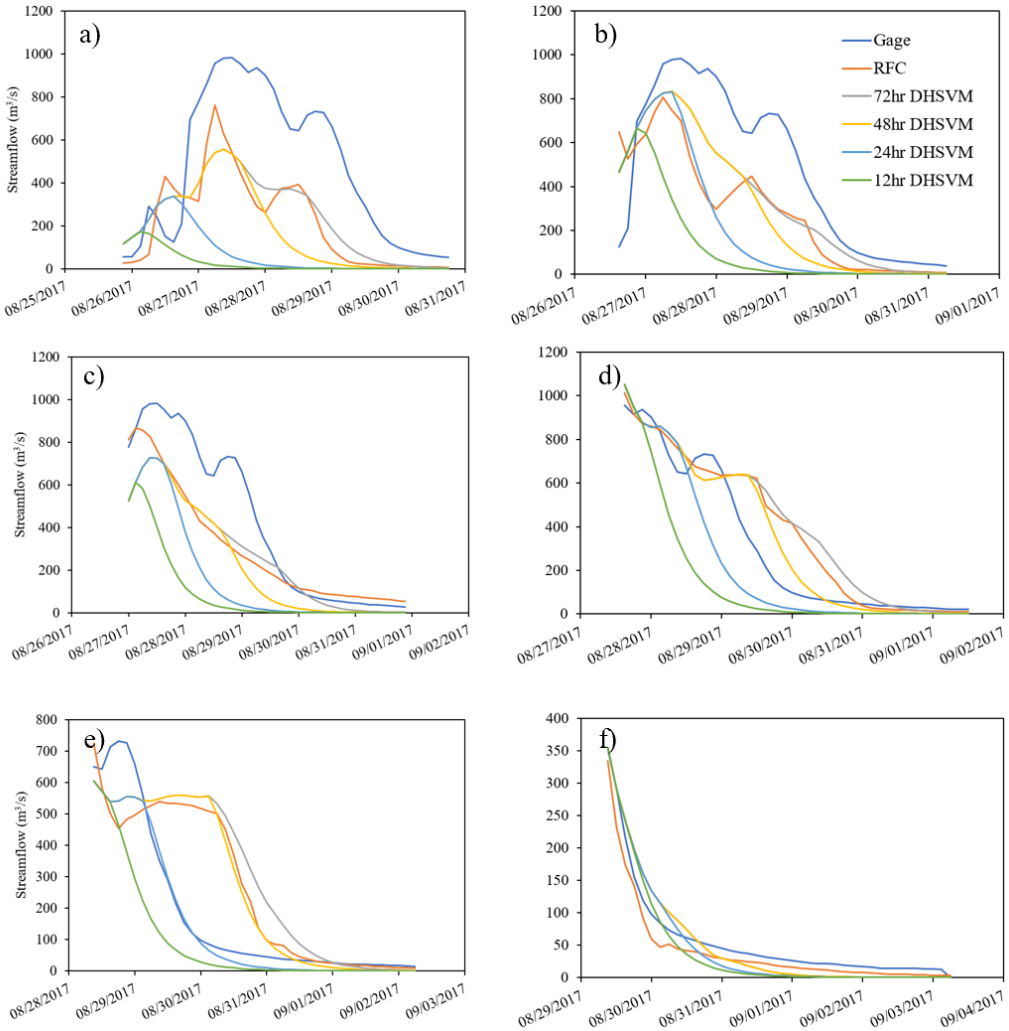

Figure 18 Selected forecasts for (a-b) beginning, (c-d) middle, and (e-f) end of Hurricane Harvey: a) Forecast 2, b) Forecast 5, c) Forecast 8, d) Forecast 10, e) Forecast 13, f) Forecast 17. Forecast numbers correspond to those in Table 8.

Forecasts 15 and 16 show anomalous results when comparing the gage data versus the 72-hour RFC forecast (Figure 17). Both forecasts predict streamflow that is positively biased when compared with the gage data and DHSVM forecasts. One possibility for this is the modeler judgement that was used to adjust the QPF data.

In summary, the decision to extend streamflow forecasts from 12 hours to 48 hours and then 72 hours during extreme events, like Hurricane Harvey, results in skillful gains at the beginning and middle of the event. Towards the end of the event, the QPFs overestimated the frequency and amount of precipitation, thus losing skill in the streamflow forecasts. This was not found to be the case for the Memorial Day flood and the Tax Day flood, which were shorter, less intense events. This finding would need to be tested on additional hurricane-level events to determine if this is true for all extreme, hurricane-level events under the current WGRFC flood forecasting system.

5. DISCUSSION

This study demonstrated the added value in extending the lead time for streamflow forecasts used by the WGRFC flood forecasting system for hurricane-level events. It also revealed the issues of underestimations in QPF data and thus error in the flood forecasting system. To the best of our knowledge, this is the first study to evaluate the benefit of extending the lead time in the WGRFC system. The methods and framework developed in this study are generally applicable to any geographical region where flood forecasting is critical.

An important limitation in this study is the difference in spatial resolution between the QPF data and the ST4 QPE data. The QPF data has 5 km resolution for the Memorial Day and Tax Day floods, and 2.5 km resolution for Hurricane Harvey, while the ST4 QPE data has 4 km resolution. To analyze the accuracy of the precipitation forecasts, a threshold was used to compare the QPF grids with the ST4 QPE grids. Thus, uncertainty is inherent in this method of comparison, as the grids do not overlap perfectly. The skill scores that are reported for the QPF data therefore have a level of uncertainty due to this method of comparison. Additionally, the skill scores for the QPF data during Hurricane Harvey are likely positively influenced by the higher spatial resolution of the QPF data.

Furthermore, input data and the initial model conditions in the hydrologic and hydraulic models are sources of uncertainty in the streamflow and floodplain forecasts (Cloke and Pappenberger 2009, Zappa et al. 2011). While precautions were taken to use high-quality input data (e.g. 10-meter DEM resampled to 20 meters, high resolution soil and precipitation data), all datasets introduce sources of error into the modeling process.

Specifically, the errors associated with the DEM could be exacerbated by the flat topography in the watershed, which could lead to errors in the streamflow and floodplain modeling. Furthermore, for the streamflow modeling, the initial model states were saved from a long model run and used to drive the forecasts. However, these model states are a product of the model, which itself makes approximations and has errors. In the case of the hydraulic model, the model initializes from a dry state. Since this was not the case for Hurricane Harvey (the watershed had received rain before the event and thus would have had higher soil moisture content), there are errors associated with this modeling setting.

In addition, uncertainty and errors from the model input data and initial conditions are exacerbated throughout the flood forecasting system, as the forecasted precipitation is used to drive the hydrological model to produce forecasted streamflow, which is in turn used to drive the hydraulic model (Cloke and Pappenberger 2009). Because models introduce errors associated with model physics and numerical solution methods, our evaluation could be affected. Thus, the contributions of model errors in the streamflow forecasts produced by DHSVM and the floodplain forecasts produced by Flood2D-GPU need to be acknowledged. To recognize the effects of such model errors on results analysis, the output for baseline simulations were compared against USGS gage data (for the streamflow modeling) and FEMA floodplain maps (for the floodplain modeling).

Furthermore, DHSVM models urban detention and runoff but does not explicitly account for urban stormwater sewage systems. Therefore, it can be concluded that there are errors associated with this omission, as this watershed is highly urbanized with a complex stormwater system. However, because the magnitude of the selected storm events is very large, the effect of the stormwater system is likely relatively minor. The

simulated streamflow performs well when compared against the USGS gage data in the calibration and validation processes; thus, the model performs robustly enough to simulate the major hydrological processes in the watershed for these selected storm events.

There are a few extensions of this work that would provide valuable insights from other perspectives. One such extension would be to study the effect that increased urbanization has on the forecasted streamflow and floodplains in Brays Bayou during these storm events. Since population is expected to continue growing in the watershed, quantification of the forecasts with increased impervious cover would allow for studying the quality of the forecasts under the new conditions. It is possible that the low bias in QPF data would be exacerbated by increasing urbanization in the streamflow and inundation forecasts.

Another extension of this work would be to use ensemble precipitation and streamflow forecasts to produce probable maximum flood inundation forecasts. As Cloke and Pappenberger (2009) assert, ensemble forecasting is generally agreed upon to improve hydrological forecasts over deterministic forecasting, particularly in regard to early flood warning systems, as model outputs from the ensemble members can be analyzed as a whole to assess the likelihood of flooding events. Thus, it would be useful to see if the effect of extending lead time that was found in this study also pertains to an adjusted flood forecasting system that utilizes ensemble forecasts. Furthermore, it would be important to determine if the WGRFC flood forecasting system could be improved as a whole if the system were to use ensemble precipitation data to produce ensemble streamflow forecasts and probable maximum flood inundation forecasts.

Lastly, one of the capabilities of the Flood2D-GPU model is to produce the flood depth in addition to the flood extent. While not analyzed in this study, the flood depth would provide important information about the magnitude of the flooding forecast. Thus, we believe that analyzing the flood depth would be an important extension of this work.

6. CONCLUSIONS

In this study, the flood forecasting system used by the WGRFC was analyzed in Brays Bayou for three of the largest storm events in Houston, TX in the last decade: the Memorial Day flood, the Tax Day flood, and Hurricane Harvey. Skill statistics were calculated throughout the flood forecasting system: QPF data for forecasted precipitation, streamflow forecasts produced by driving the DHSVM model with the QPF data, and forecasted floodplains produced by driving the 2D hydraulic model, Flood2D-GPU, with forecasted streamflow. Then, forecasted streamflow comparisons were made between the DHSVM forecasts and the RFC forecasts (from the lumped CHPS-FEWS model) to determine the performance of the models as well as the effect of lead time to produce skillful forecasts. The key conclusions are:

1. The QPFs for shorter, less intense events (e.g., Memorial Day flood) have worse skill statistics than the QPFs for longer, more intense events (e.g., Hurricane Harvey). The skill of the QPF data generally decreases with increased lead time and the QPF data is negatively biased.
2. The negative bias of the QPF data is reflected in the generally underestimated forecasted streamflow results, as the forecasted streamflow relies on the accuracy of the QPF data. Thus, the forecasted streamflow also shows decreased skill with increased lead time. For the extended event (Hurricane Harvey), the streamflow skill scores are higher than for the shorter events. This could be due to the long duration of the event, the uniform precipitation coverage of the watershed, and the better ability of the QPF data to forecast the depth of the rainfall in this event.

3. The forecasted floodplain results show increased skill when increasing the lead time from 12 hours to 72 hours during Hurricane Harvey.
4. The QPF data is the key factor in determining the forecasting skills, as the precipitation data is the key input for the streamflow model, which in turn is the key input for the floodplain model. Therefore, the quality of the streamflow and floodplain forecasts is dependent upon the quality of the QPF data.
5. Increasing the lead time for streamflow forecasts improved the skill scores at the beginning and middle of the event for Hurricane Harvey. However, at the end of the storm event, the forecasts overestimated the amount of streamflow that occurred. Therefore, we conclude that the decision by the RFC to extend the streamflow forecasts from 12 hours to 48 hours and then to 72 hours is a skillful and useful forecasting decision for hurricane-level events. However, it does not add value to shorter, smaller-scale event forecasts, as the spatial and temporal variability is difficult to accurately predict in the forecasting system.
6. The benefits of extending forecast lead time in hurricane-level events should be tested on similar events in the Brays Bayou watershed to confirm our results. Furthermore, they should be tested on similar events in other watersheds to determine if this finding can be extended to other watersheds.

REFERENCES

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R., and Prat, O. P. (2014). "PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies." *Bulletin of the American Meteorological Society*, 96(1), 69-83.

Bass, B., Juan, A., Gori, A., Fang, Z., and Bedient, P. (2017). "2015 Memorial Day Flood Impacts for Changing Watershed Conditions in Houston." *Natural Hazards Review*, 18(3), 05016007.

Blake, E. S., and Zelinsky, D. A. (2018). "Hurricane Harvey." National Hurricane Center Tropical Cyclone Report. https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf

Carpenter, T. M., and Georgakakos, K. P. (2006). "Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales." *Journal of Hydrology*, 329(1), 174-185.

Chen, J., Zhong, P.-a., Wang, M.-l., Zhu, F.-l., Wan, X.-y., and Zhang, Y. (2018). "A Risk-Based Model for Real-Time Flood Control Operation of a Cascade Reservoir System under Emergency Conditions." *Water*, 10(2).

Cloke, H. L., and Pappenberger, F. (2009). "Ensemble flood forecasting: A review." *Journal of Hydrology*, 375(3), 613-626.

Cluckie, I. D., and Xuan, Y. (2008). "Uncertainty Propagation in Ensemble Rainfall Prediction Systems used for Operational Real-Time Flood Forecasting." *Practical Hydroinformatics: Computational Intelligence and Technological Developments*

in Water Applications, R. J. Abrahart, L. M. See, and D. P. Solomatine, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, 437-447.

Cook, A., and Merwade, V. (2009). "Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping." *Journal of Hydrology*, 377(1), 131-142.

Cuo, L., Lettenmaier, D. P., Mattheussen, B. V., Storck, P., and Wiley, M. (2008). "Hydrologic prediction for urban watersheds with the Distributed Hydrology-Soil-Vegetation Model." *Hydrological Processes*, 22(21), 4205-4213.

Duan, Zhuoran. (2018). Github Repository. Retrieved from https://github.com/pnnl/DHSVMPNNL/tree/master/CreateStreamNetwork_PythonV

Gangrade, S., Kao, S.-C., Dullo, T., Kalyanapu, A., and Preston, B. (2019). "Ensemble-based flood vulnerability assessment for probable maximum flood in a changing environment." In review.

Harris County Flood Control District (HCFCD 2018). "Brays Bayou - Bond Program." < https://www.hcfcd.org/2018-bond-program/watersheds/brays-bayou-bond-program/> (accessed May 29, 2019).

Harris County Flood Control District (HCFCD 2019). "Brays Bayou." <https://www.hcfcd.org/projects-studies/brays-bayou/> (accessed May 23, 2019).

Horritt, M. S., and Bates, P. D. (2002). "Evaluation of 1D and 2D numerical models for predicting river flood inundation." *Journal of Hydrology*, 268(1), 87-99.

Kalyanapu, A. J., Shankar, S., Pardyjak, E. R., Judi, D. R., and Burian, S. J. (2011). "Assessment of GPU computational enhancement to a 2D flood model." *Environmental Modelling & Software*, 26(8), 1009-1016.

Kao, S.-C., DeNeale Scott, T., and Watson David, B. (2019). "Hurricane Harvey Highlights: Need to Assess the Adequacy of Probable Maximum Precipitation Estimation Methods." *Journal of Hydrologic Engineering*, 24(4), 05019005.

Koren, V., Reed, S., Smith, M., Zhang, Z., and Seo, D.-J. (2004). "Hydrology Laboratory Research Modeling System (HL-RMS) of the US National Weather Service." *Journal of Hydrology*, 291, 297-318.

Lehmann, J., Coumou, D., and Frieler, K. (2015). "Increased record-breaking precipitation events under global warming." *Clim. Change*, 132(4), 501-515.

Lin, Y., and Mitchell, K. E. (2005). "The NCEP stage II/IV hourly precipitation analyses: Development and applications." 19th Conf. on Hydrology, San Diego, CA. Meteor. Soc., 1.2.

Lin, Y. (2011). GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Version 1.0. UCAR/NCAR – Earth Observing Laboratory. https://doi.org/10.5065/D6PG1QDD. Accessed 14 Feb 2019.

Marshall, R., Ghafoor, S., Rogers, M., Kalyanapu, A., and T. Dullo, T. (2018). "Performance Evaluation and Enhancements of a Flood Simulator Application for Heterogeneous HPC Environments." *International Journal of Networking and Computing*, 8(2), 387-407.

"Metro Houston Population Forecast: Projections to 2050." Greater Houston Partnership Research Department. April 2014.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007). "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations." *Trans. ASABE*, 50(3), 885-900.

Muñoz, L. A., Olivera, F., Giglio, M., & Berke, P. (2018). The impact of urbanization on the streamflows and the 100-year floodplain extent of the Sims Bayou in Houston, Texas. *International Journal of River Basin Management, 16*(1), 61-69.

NWS (2017a). National Weather Service Hazard Statistics Weather Fatalities 2017. http://www.nws.noaa.gov/os/hazstats/images/weather_fatalities.pdf

NWS (2017b). 2017 Flash Flood Report. http://www.nws.noaa.gov/om/hazstats/flood17.pdf.

Nelson, B. R., Prat, O. P., Seo, D. J., and Habib, E. (2015). "Assessment and Implications of NCEP Stage IV Quantitative Precipitation Estimates for Product Intercomparisons." *Weather Forecast.*, 31(2), 371-394.

Nguyen, P., Thorstensen, A., Sorooshian, S., Hsu, K., and AghaKouchak, A. (2015). "Flood Forecasting and Inundation Mapping Using HiResFlood-UCI and Near-Real-Time Satellite Precipitation Data: The 2008 Iowa Flood." *Journal of Hydrometeorology*, 16(3), 1171-1183.

OECD (2017). *Coping with the financial consequences of devastating floods*. Available at: http://www.oecd.org/finance/coping-with-the-financial-consequences-of-devastating-floods.htm [Accessed 17 Jan. 2019].

Olivera, F., & DeFee Buren, B. (2007). Urbanization and Its Effect On Runoff in the Whiteoak Bayou Watershed, Texas. *JAWRA Journal of the American Water Resources Association, 43*(1), 170-182. doi:10.1111/j.1752-1688.2007.00014.x

Pielke, Jr., R.A., M.W. Downton, and J.Z. Barnard Miller, 2002: Flood Damage in the United States, 1926–2000: A Reanalysis of National Weather Service Estimates. Boulder, CO: UCAR.

Rossa, A., Liechti, K., Zappa, M., Bruen, M., Germann, U., Haase, G., Keil, C., and Krahe, P. (2011). *The COST 731 Action: A review on uncertainty propagation in advanced hydro-meteorological forecast systems*.

Sapiano, M. R. P., and Arkin, P. A. (2009). "An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data." *Journal of Hydrometeorology*, 10(1), 149-166.

Seo, B. C., Quintero, F., and Krajewski, W. F. (2018). "High-Resolution QPF Uncertainty and Its Implications for Flood Prediction: A Case Study for the Eastern Iowa Flood of 2016." *Journal of Hydrometeorology*, 19(8), 1289-1304.

Selvanathan, S., Sreetharan, M., Lawler, S., Rand, K., Choi, J., and Mampara, M. (2018). "A Framework to Develop Nationwide Flooding Extents Using Climate Models and Assess Forecast Potential for Flood Resilience." *J. Am. Water Resour. Assoc.*, 54(1), 90-103.

ShahiriParsa, A., Noori, M., Heydari, M., and Rashidi, M. (2016). "Floodplain Zoning Simulation by Using HEC-RAS and CCHE2D Models in the Sungai Maka River." *Air, Soil and Water Research*, 9, ASWR.S36089.

Shukla, S., Voisin, N., and Lettenmaier, D. P. (2012). "Value of medium range weather forecasts in the improvement of seasonal hydrologic prediction skill." *Hydrology and Earth System Sciences*, 16(8), 2825-2838.

Shuttleworth, W. (1992). 'Evaporation'. *Handbook of Hydrology*, D. Maidment, McGraw-Hill, New York, 4.2-4.18.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at https://websoilsurvey.nrcs.usda.gov/. Accessed [February 14, 2019].

Steenbergen, N., and Willems, P. (2014). *Rainfall Uncertainty in Flood Forecasting: Belgian Case Study of Rivierbeek.*

Sukovich, E. M., Ralph, F. M., Barthold, F. E., Reynolds, D. W., and Novak, D. R. (2014). "Extreme Quantitative Precipitation Forecast Performance at the Weather Prediction Center from 2001 to 2011." *Weather Forecast.*, 29(4), 894-911.

U.S. Geological Survey, The National Map, 2017, 3DEP products and services: The National Map, 3D Elevation Program Web page, accessed October 6, 2018 at https://nationalmap.gov/3DEP/3dep_prodserv.html.

van Oldenborgh, G. J., van der Wiel, K., Sebastian, A., Singh, R., Arrighi, J., Otto, F., Haustein, K., Li, S., Vecchi, G., and Cullen, H. (2017). "Attribution of extreme rainfall from Hurricane Harvey, August 2017." *Environmental Research Letters*, 12(12), 124009.

Voisin, N., Pappenberger, F., Lettenmaier, D. P., Buizza, R., and Schaake, J. C. (2011). "Application of a Medium-Range Global Hydrologic Probabilistic Forecast Scheme to the Ohio River Basin." *Weather Forecast.*, 26(4), 425-446.

Wigmosta, M.S., Vail, L.W. and Lettenmaier, D.P. (1994). “A Distributed Hydrology-Vegetation Model for Complex Terrain.” *Water Resources Research* 30(6): 1665 – 1679.

Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., and Erickson, T. A. (2017). "Validation of a 30 m resolution flood hazard model of the conterminous United States." *Water Resources Research*, 53(9), 7968-7986.

Zappa, M., Jaun, S., Germann, U., Walser, A., and Fundel, F. (2011). *Superposition of three sources of uncertainties in operational flood forecasting chains*.

Zhang, Z., Koren, V., Smith, M., Reed, S., and Wang, D. (2004). "Use of Next Generation Weather Radar Data and Basin Disaggregation to Improve Continuous Hydrograph Simulations." *Journal of Hydrologic Engineering*, 9(2), 103-115.

Zhao, G., Gao, H., and Cuo, L. (2016). "Effects of Urbanization and Climate Change on Peak Flows over the San Antonio River Basin, Texas." *Journal of Hydrometeorology*, 17(9), 2371-2389.