Evaluating precipitation, streamflow, and inundation forecasting skills during extreme weather events: A case study for an urban watershed

Xudong Li[1], Cheryl Rankin[1], Sudershan Gangrade[2], Gang Zhao[1], Kris Lander[3], Nathalie Voisin[4], Manqing Shao[1], Mario Morales-Hernández[5], Shih-Chieh Kao[2], and Huilin Gao[1, *]

1. Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX, United States

2. Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States

3. West Gulf River Forecast Center, National Oceanic and Atmospheric Administration, Fort Worth, TX, United States

4. Pacific Northwest National Laboratory, Seattle, WA, United States

5. Computational Science and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States

* Author to whom any correspondence should be addressed.

E-mail: hgao@civil.tamu.edu

Abstract: Integrated forecasting systems for precipitation, streamflow, and floodplain inundation are of critical importance in mitigating the impacts of destructive floods caused by extreme weather events. However, the skills of streamflow and floodplain inundation forecasts derived from various Quantitative Precipitation Forecast (QPF) require a greater level of understanding. In this study, a set of QPF developed by the National Weather Service (NWS) were used to drive a flood modeling system obtained through offline coupling of a physics-based distributed hydrological model, the Distributed Hydrology Soil and Vegetation Model (DHSVM), and a hydrodynamic model, the Two-dimensional Runoff Inundation Toolkit for Operational Needs (TRITON). This flood modeling system was used to produce forecasts of streamflow and floodplain inundation maps during three major flood events in the Brays Bayou Watershed (Houston, Texas, USA) for a range of QPF durations (6–72 h). Then, to investigate the effects of increasing QPF durations on the forecasts, the forecasting skills of precipitation, streamflow, and floodplain inundation were quantified. The results show that: 1) QPF skills for more intense and sustained events such as hurricanes and tropical storms are higher than for shorter, less intense events; 2) while QPF and streamflow forecasting skills decrease as QPF durations increase, inundation forecasts under longer QPF durations (24 or 72 h) show higher skills; 3) extending the maximum QPF duration in operational hydrologic modeling from 24 h (under normal circumstances) to 72 h (for extreme events) may increase the skills of long lead time forecasts for large-scale events like Hurricane Harvey.

Keywords: flood forecasting skill, extreme weather events, quantitative precipitation forecast (QPF), QPF duration, inundation mapping

1. Introduction

Extreme precipitation events are occurring with greater frequency and intensity due to climate change (Lehmann et al. 2015), resulting in increasingly disastrous floods. Floods can cause massive destruction in a short period of time, but are challenging to forecast given the complex relationships between meteorological and hydrologic systems. Accurate flood predictions with sufficiently long lead times are desired to support emergency managers and first responders in determining a best strategy for fast action and evacuation (Georgakakos and Hudlow, 1984; Selvanathan et al. 2018).

The accuracy of flood forecasts depends on the cascading prediction skills across precipitation, streamflow, and floodplain inundation. With the exception of snowmelt-driven floods (Fang et al., 2014), the main source of uncertainty in flood forecasting systems is precipitation. Other secondary factors include initial soil moisture conditions and model structural uncertainty (Zappa et al. 2011). The skills of precipitation forecasts may vary with storm drivers, duration, and intensity. For example, Sikder et al. (2019) reported that hurricane-induced storms are more challenging to forecast than the less intense and more frequent winter frontal systems. Considering that the most destructive floods are usually brought on by hurricanes (such as Hurricane Harvey in 2017; Fernandez-Rivera et al. 2019; Kao et al. 2019), the accuracy of precipitation forecasts during these storms is particularly important (Droegemeier et al. 2000; Ko et al. 2020; Marchok et al. 2007; Sikder et al. 2019).

Moreover, the errors from precipitation forecasts may be exacerbated in streamflow predictions (Cloke and Pappenberger 2009; Cluckie and Xuan 2008; Rossa et al. 2011; Van Steenbergen and Willems 2014) due to the uncertainties associated with hydrologic model structures, parameterizations, inputs, and initial conditions (Cloke and Pappenberger 2009; Zappa

et al. 2011; Demargne et al. 2014; Jain et al. 2018). While the relationships between predicted precipitation and streamflow have been well recognized (Parodi and Ferraris 2004; Jung and Merwade 2015), their relative forecasting skills—and how the uncertainties propagate—warrant further investigation. Considering the declining tendency of precipitation forecast skills as lead-time increases, Adams and Dymond (2019) questioned the benefits of applying longer duration QPF in streamflow forecast. Therefore, the effects of QPF duration on the best practices of streamflow forecasts should not be overlooked (Gourley and Vieux 2005; Ibbitt et al. 2005; Li et al. 2017).

The skills of floodplain inundation forecasts associated with varying QPF durations are also not well understood. The 2D hydrodynamic models are commonly used for floodplain inundation forecasts for emergency management (Apel et al., 2009; Bhola et al., 2018), although 1D models, and other methods like Height Above Nearest Drainage (HAND), are also adopted in practice (Adams et al., 2018; Johnson et al., 2019). Accurate hydrodynamic simulations typically demand high-resolution models (Merwade et al. 2008). Since inundation forecasts are much more computationally expensive than streamflow forecasts, they are usually produced less frequently during extreme events (Bhola et al. 2018; Jain et al. 2018). Efforts have been made to resolve the trade-off between modeling accuracy and computational burden in inundation forecasting (Schumann et al., 2013; Sanders and Schubert, 2019). Recently, an open-source 2D hydrodynamic model—the Two-dimensional Runoff Inundation Toolkit for Operational Needs (TRITON; Morales-Hernández et al. 2020 and 2021)—was developed to leverage the multi-GPU architecture of modern high-performance computers (HPC) for accelerated inundation modeling. Nevertheless, given the high computing cost, it is not realistic to expect that one may simulate all possible ensemble inundation scenarios (which is often done for ensemble precipitation and streamflow

87 forecasting). Overall, the skills of inundation forecasting methods need to be better understood in
88 order to prioritize the limited computing resources during operation.

89 The objectives of this study are to explore integrated flood forecasting skills and their
90 relationships to QPF duration, as well as to identify optimal flood forecasting strategies. Here, the
91 "QPF duration" refers to the maximum length of QPF used in the subsequent hydrologic modeling,
92 which can range from 6 hours to the full length of the precipitation forecast lead time. "Lead time"
93 is defined as the time range since a forecast is made to the forecasted future timing (Georgakakos
94 and Hudlow, 1984). Specifically, we address the following questions in this research:

95 1) How do the errors of forecasted precipitation propagate throughout the flood forecasting
96 system during different types of extreme weather events?
97 2) What are the skills of forecasted precipitation, streamflow, and floodplain inundation
98 across different QPF durations?
99 3) Should longer-duration forecasted precipitation be curtailed when making streamflow
100 and floodplain inundation predictions? How does the timing of floodplain inundation
101 forecasts initiation affect the forecasting skill?
102 4) Do flood forecasting skills differ between different types of weather events as defined by
103 their rainfall intensities and durations?

104 The remainder of this paper is organized as follows: In Section 2, the study area and
105 selected flood events are introduced. Section 3 describes the methods used in this study, including
106 models to produce the forecasts and metrics used to evaluate forecasting skills. In Section 4, the
107 skills of forecasted precipitation, streamflow, and floodplain inundation under multiple QPF
108 durations are compared. In Section 5, issues related to the varying forecasting skills are discussed.
109 Lastly, the conclusions are presented in Section 6.

2. Study area and selected flood events

a. Study area

The Brays Bayou Watershed is located in southwest Houston (mostly in Harris County), Texas, USA, as shown in Fig. 1 (29.37–29.45°N, 95.16–95.41°W). According to the Harris County Flood Control District (HCFCD), the drainage area of the watershed is 329 km^2. The watershed slopes downward gradually from west to east as the watershed approaches the coast. The climatology of the watershed is generally wet and subtropical, characterized by humid, warm summers and mild winters. The average rainfall is 1415 mm per year, with July and August being the wettest months (on average). The primary land cover type is urban (95.9% urbanized; USGS, 2014) and the primary soil type is clay, which has a very low hydraulic conductivity. Thus, water does not easily infiltrate into the soil. These topographical, climatic, and geographical characteristics make the watershed prone to flooding.

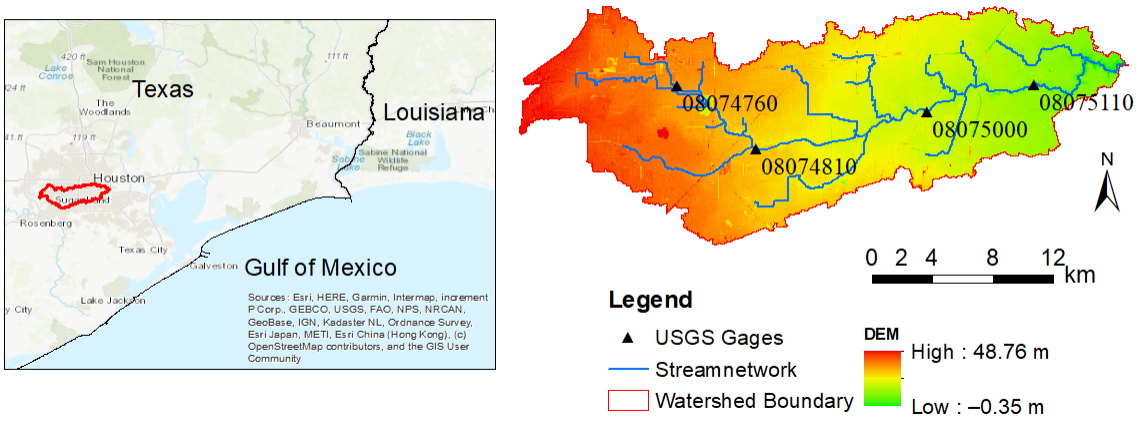


Fig. 1. Location (left) and map (right) of the study area, including Digital Elevation Model (DEM) information and locations of USGS streamflow gages.

b. Selected flood events

The Brays Bayou Watershed has experienced multiple severe flood events in the last few decades. Three recent major flood events with different driving storm types were considered in

this study: the Memorial Day Flood (May 25–26, 2015), the Tax Day Flood (April 17–18, 2016), and Hurricane Harvey (August 25–31, 2017). Both the Memorial Day Flood and the Tax Day Flood events were caused by short-term convective storms (Furl *et al*., 2018; Nielsen and Schumacher, 2020) which brought in 226 mm and 243 mm of rainfall during 9 h and 24 h periods. The Hurricane Harvey flood event was caused by a slow-moving Category 4 hurricane (Kao et al., 2019) which brought in 893 mm of rainfall during a 108 h period.

3. Methods and data

a. Precipitation observation and forecasts

Both observed and forecasted precipitation data were used in this study. The National Center for Environmental Prediction (NCEP) Stage IV (ST4) Quantitative Precipitation Estimates (QPE; Lin 2011; Lin and Mitchell 2005) were considered as the "ground truth" used to evaluate the accuracy of forecasted precipitation—and also to support hydrologic model calibration and validation. The ST4 data merges raw radar-based estimates with automatic rainfall gage observations and is further quality-controlled by National Weather Service (NWS) River Forecasting Centers (RFCs). The ST4 data are available at an hourly time step with a 4-km spatial grid resolution from 2002 to the present. The values are produced in near-real time as they become available, typically within one hour after receiving data from the RFCs. Given its high quality, ST4 has been used in many studies to represent observed precipitation (Ashouri et al. 2015; Kao et al. 2019; Nelson et al. 2016; Sapiano and Arkin 2009).

The NWS Weather Prediction Center's (WPC) Quantitative Precipitation Forecasts (QPF; https://www.wpc.ncep.noaa.gov/html/fam2.shtml#qpf) were used for forecasted precipitation data. Using a suite of operational models, WPC researchers leverage their experience to estimate the probable amount of precipitation across various future durations. The QPF products are provided

four times a day at 6 h temporal and 5 km / 2.5 km spatial resolutions for a maximum of a 7-day lead time. Specifically, QPF is manually created for 6-hour periods on Days 1, 2 and 3; while on Days 4-7, QPF is created for 24-hour periods and then equally distributed to 6- hour periods. Overall, both ST4 and QPF were processed 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.

b. DHSVM streamflow simulation

A physically based hydrological model, the Distributed Hydrology Soil and Vegetation Model (DHSVM, Wigmosta et al. 1994), was employed in this study to simulate streamflow in the Brays Bayou Watershed. DHSVM is an open-source model that couples full water and energy balances to calculate the hydrologic processes at high spatial and temporal resolutions. The model has been revised with parameters to characterize the hydrological processes in urban areas (Cuo et al. 2008), and thus is adequate for flood simulation in watersheds with urban land covers (Zhao et al. 2016; Shao et al. 2020; Li et al. 2020). In this study, the model was established with a spatial resolution of 20 m and a temporal resolution of 3 h. All input data are summarized in **Error! Reference source not found.**. Driven by ST4 and validated by observed streamflow, the DHSVM model parameters were calibrated (as described in Appendix A, Text 1 and Fig. A.1). The DHSVM performance of the three selected flood events is shown in Fig. 2, and are considered satisfactory for the purposes of this study.

Table 1

Summary of DHSVM input data

Type of Data	Source	Spatial Resolution
Digital elevation model (1/3 arc-second, ~10m)	USGS National Elevation Dataset (Sugarbaker et al. 2017)	Resampled to 20 m

Soil type	USDA NRCS SSURGO Database (Nauman et al. 2018)	Resampled to 20 m
Land use land cover (status of 2011)	USGS National Land Cover Database (USGS 2014)	Resampled to 20 m
Observed precipitation	NCEP Stage IV Quantitative Precipitation Estimates (ST4)	4 km
Forecasted precipitation	WPC Quantitative Precipitation Forecasts (QPF)	5 km (in 2015 and 2016) 2.5 km (in 2017)
Additional meteorological forcing data (other than precipitation)	North America Land Data Assimilation System (NLDAS; Xia et al. 2012)	1 km, averaged over the watershed

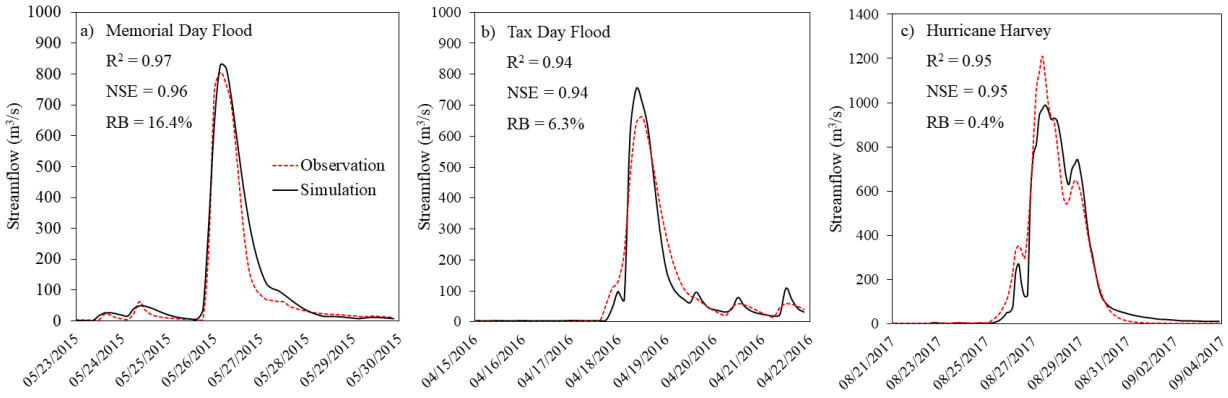


Fig. 2. Evaluation of DHSVM flow simulations at gage 08075000 across the three storm events of interest: (a) the Memorial Day Flood, (b) the Tax Day Flood, and (c) Hurricane Harvey.

Further, QPF was utilized as precipitation inputs for simulating forecasted streamflow via DHSVM. While intuitively one may expect to simply include QPF across all durations (up to 7 days) in hydrologic modeling, in practice RFCs utilized QPF with a default duration (typically 24 or 48 h) under normal circumstances in their operational streamflow forecasting (i.e., zero precipitation after the default duration). In the case of the West Gulf River Forecast Center (WGRFC), the QPF webpage (https://www.weather.gov/wgrfc/wgrfcqpfpage) states: "*In general, the first 24 hours of QPF is used as input in the creation of river forecasts; however, dependent on specific weather events, up to the entire 72 hours may be used.*" The 1–3 day QPF is then utilized by WGRFC to create the 5-day Significant Flood Outlook product.

185 Although the rationale of this practice is not specifically explained, there is a notable trade-
186 off between forecast duration and accuracy. Given the rapidly decreasing skills of QPF in longer
187 QPF durations (Cuo et al. 2011a; Seo et al. 2018a), the inclusion of medium-range QPF may lead
188 to worse streamflow forecasting skills (Adams and Dymond 2019). However, longer lead time
189 warnings (e.g., 1–2 days before the event) are useful for emergency responders and infrastructure
190 operators (Golding 2009; Li et al. 2017; Parker et al. 2009). For a long-term event like Hurricane
191 Harvey, there can be value in using longer duration QPF, even if they are less accurate. For each
192 event, we tested and compared different maximum QPF durations from 6 h to 72 h for streamflow
193 forecasts. In other words, these additional DHSVM simulations can help us understand the impact
194 of increasing maximum QPF durations on streamflow forecasting skills and the potential benefits.
195 A summary of all DHSVM simulations is provided in Table 2. For each flood event, a base
196 simulation driven by ST4 was conducted. Then ST4 was replaced with QPF produced at different
197 start times and with different QPF durations to simulate forecasted streamflow.

198 **Table 2**
199 Summary of all DHSVM simulations

Event Names & Time (YYYYMMDD_HH)	Source of Precipitation	QPF Start time	Maximum QPF Length in DHSVM Simulation (hours)	Number of DHSVM Simulations
Memorial Day Flood (20150523_00–20150529_23)	1 set of ST4		--	1
	ST4 and 28 sets of QPF*	(20150523_00, 20150523_06, …, 20150529_18)	12 max. QPF lengths (6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66, 72)	336 (=28×12)
Tax Day Flood (20160415_00–20160421_23)	1 set of ST4		--	1
	ST4 and 28 sets of QPF	(20160415_00, 20160415_06, …, 20160421_18)	12 max. QPF lengths (6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66, 72)	336 (=28×12)
	1 set of ST4		--	1

Hurricane Harvey, (20170821_00–20170903_23)	56 sets of QPF	(20170821_00, 20170821_06, …, 20170903_18)	12 max. QPF lengths (6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66, 72)	672 (=56×12)

Note: The QPF forecasts are created four times a day (00, 06, 12, 18 hours), labeled by YYYYMMDD_HH

c. TRITON inundation modeling

A multi-GPU open source 2D hydraulic model, TRITON (Morales-Hernández et al. 2020 and 2021), was employed in this study to simulate surface inundation driven by the forecasted streamflow conditions. TRITON solves the full 2D shallow water equations with source terms using an explicit numerical scheme. The model is designed for modern HPC and can leverage multiple CPUs or GPUs to largely accelerate hydrodynamic simulations.

The TRITON model was selected in this study (instead of the simplified models—diffusive and quasi-inertial) based on two considerations. First, spatially distributed and physics-based models are essential for obtaining acceptable results (in terms of accuracy and stability) for rainfall-runoff problems (Costabile et al., 2013; Costabile et al., 2015; Caviedes et al., 2020). TRITON is not only computationally efficient, but also has proven to work under any flow conditions. Second, TRITON provides reasonable results—with minimal constraints—when simulating an urbanized watershed with highly variable land surface conditions. Moreover, a main advantage of TRITON is the use of the DEM as the computational mesh, such that the user does not have to create a mesh for each case. This is important because the creation of an efficient mesh—such as an unstructured triangular mesh, a quad-tree mesh or something similar (Ferraro, et al., 2020; Costabile and Costanzo, 2021)—is both time consuming and requires modeling expertise. Furthermore, the data exchange between subdomains is crucial when having an unstructured mesh, and requires algorithms that could directly impact the performance.

The performance portability for multiple GPUs is ensured, provided that the supercomputer contains NVDIA-GPUs in its hardware. The speed of the calculations will only depend on the capabilities of the graphic card. Given that more and more supercomputers utilize GPU-like accelerators, this proposed methodology can be applied elsewhere.

The TRITON setup for the Brays Bayou Watershed included a 10-m resolution DEM (Sugarbaker et al. 2017) and a spatially varied Manning's roughness coefficient from Dullo et al. (2021). The TRITON computational domain extends downstream beyond the watershed to minimize any backwater effects. A similar process used by Gangrade et al. (2019) was followed to convert the simulated DHSVM streamflow into corresponding TRITON inputs. The input DHSVM streamflow hydrographs extended for 10 days and captured the peak streamflow around the end of the fifth day. The TRITON outputs were saved at a 30-min timestep across the entire computational domain. The simulations were conducted using the Summit supercomputer managed by the Oak Ridge Leadership Computing Facility (https://www.olcf.ornl.gov/summit/). Further details on the TRITON setup and validation are provided in Appendix A, text 2.

Even with the enhanced computing speed, it is still not feasible to conduct TRITON simulations for all ensemble DHSVM streamflow scenarios reported in Table 2. Lower frequency is also an issue that decision makers are to be faced with when carrying out inundation forecasts in practice, as compared to the frequency in streamflow forecasts. Therefore, TRITON scenarios were designed to investigate the performance of forecasts made at different stages of the flood events under different QPF durations. Forecasts were made for each of the three events at three separate stages: at the onset of flooding, around the time of the flood peak, and during the recession period. For the Memorial Day flood and the Tax Day flood, an extra scenario was tested to see whether forecasts made slightly (i.e., 6 h) before the rainfall events achieve similar forecast skill

as the forecasts made at the beginning (which could add additional time for emergency actions). Meanwhile, because Hurricane Harvey lasted much longer—and the rainfall did not reach its peak until about 2 days later—we did not see much difference in the forecasted rainfall totals (i.e., between forecasts made 6 h earlier, or 6 h later. Therefore, five scenarios were set for Hurricane Harvey at an interval of 12 h—three for the prementioned stages, and two additional scenarios distributed into the rising limb and recession limb periods. A summary of all of the TRITON simulations is provided in Table 3. For each flood event, a base TRITON simulation was conducted using the ST4-driven DHSVM streamflow. Since the timeseries of the actual inundation is unavailable, the base simulation was used as a proxy. Then, the forecasted inundations were produced, driven by different QPF-driven streamflow forecasts. To provide an outlook for the whole event and to capture the maximum inundation extent, a single streamflow forecast with a long lead time (5-day) was adopted to drive an inundation forecast. Finally, skills of these inundation forecasts were evaluated.

Table 3
Summary of all TRITON simulations

Event Names & Time (YYYYMMDDHH)	Source of Precipitation	Forecast Start Time (YYYYMMDD_HH)	Maximum QPF length in DHSVM simulation (hours)	Number of TRITON simulations (Total 41)
Memorial Day Flood (2015052118–2015053112)	1 set of ST4	--	--	1
	QPF	20150524_06	(72)	1
		20150525_12	(12, 24)	2
		20150525_18	(12, 24, 72)	3
		20150526_06	(12, 24, 72)	3
		20150527_12	(12, 24)	2
Tax Day Flood (2016041400–2016042318)	1 set of ST4	--	--	1
	QPF	20160416_12	(72)	1
		20160417_12	(12, 24, 72)	3
		20160417_18	(12, 24, 72)	3
		20160418_12	(12, 24, 72)	3

		20160419_12	(12, 24)	2
Hurricane Harvey, (2017082300–2017090118)	1 set of ST4	--	--	1
	QPF	20170825_12	(12, 24, 72)	3
		20170826_12	(12, 24, 72)	3
		20170827_12	(12, 24, 72)	3
		20170828_12	(12, 24, 72)	3
		20170829_12	(12, 24, 72)	3

Note: The QPF forecasts are created four times a day (00, 06, 12, 18 hours), labeled by YYYYMMDD_HH

d. Forecasting skill metrics

The overall workflow to evaluate the forecasting skills of the flood forecasting system is shown in Fig. 3. For each assessment, skill score calculations were conducted to demonstrate the uncertainty and skill of each forecast throughout the flood forecasting process.

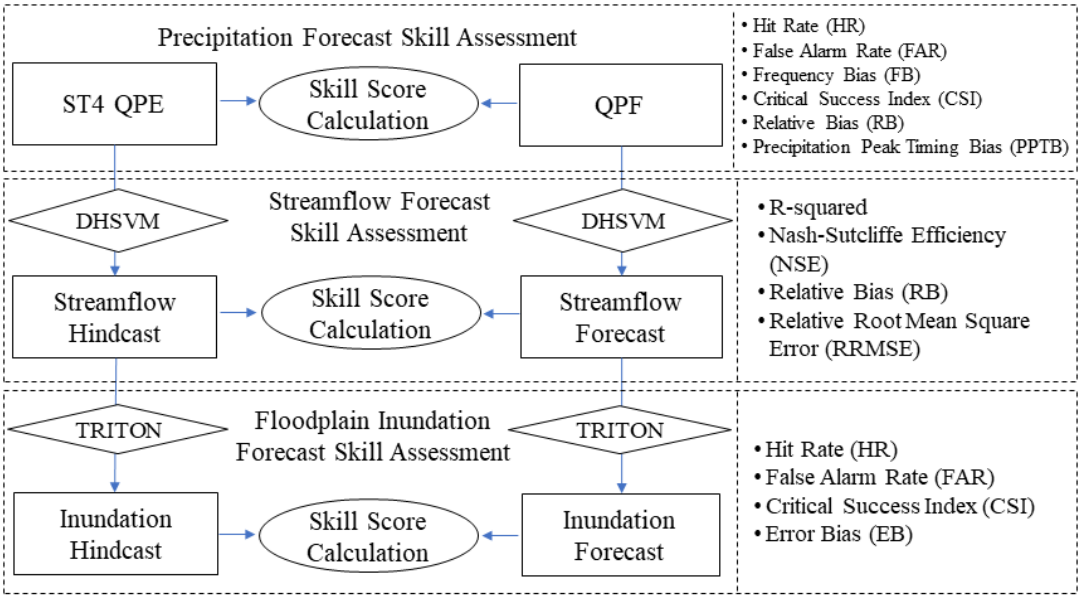


Fig. 3. Schematic overview of precipitation, streamflow, and inundation forecast evaluation

1) Precipitation

In this study, the QPF were first assessed to determine their spatial and temporal accuracy with respect to ST4. Given the different spatial resolutions between QPF and ST4, each QPF grid cell was compared with the closest ST4 grid cell. Additionally, the hourly ST4 data was aggregated to a 6-h time step in order to be comparable with the QPF. The skill scores for precipitation forecasts were adopted from Seo et al. (2018b), and include the hit rate (HR), the false alarm rate

(FAR), and frequency bias (FB). Then, the Critical Success Index (CSI) was used as an indicator of the overall performance of the forecasts in perspective of spatial and temporal accuracy of rainfall (see details in Table A.2).

In addition, the accuracy with regard to total rainfall and peak timing are also important since it directly affects the performances of streamflow and inundation forecasts. The relative bias (RB) of total precipitation over the watershed, and the precipitation peak timing bias (PPTB), are also examined for individual QPF made at different times during the events. Specifically, the RB of the total precipitation is defined as follows:

$$RB = \left(\sum_{i=1}^{T} O_i + \sum_{i=T+1}^{N} F_i\right) \Big/ \sum_{i=1}^{N} O_i$$

in which, O_i and F_i refer to observed and forecasted precipitation rates over the watershed at the i^{th} time interval (6 h); notation T is the number of time intervals before the forecast; and N is the number of time steps until the end of the event.

2) Streamflow

DHSVM simulations were used as substitutes for the WGRFC streamflow forecasts in this study, since WGRFC streamflow forecasts are only available for specific QPF durations. The simulations were carried out to mimic the five-day forecasts produced by the WGRFC (using the CHPS-FEWS model, hereafter referred to as RFC forecasts); specifically, QPF of various durations was incorporated in the hydrologic modeling scenarios. Skills of streamflow forecasts are to be evaluated in two contexts. The first context is the skill of an instantaneous (or "real-time") streamflow forecast system. The DHSVM simulations based on QPF with each particular duration (Table 2) were synthesized to produce hydrographs composed fully of forecasts with the same

duration (see details in Appendix A, text 3), which allows for the determination of the overall forecasting skills of an instantaneous streamflow forecasting system across the events. In the instantaneous streamflow system, the streamflow forecasts are updated every interval of the QPF duration.

The other one is the skill of single streamflow forecasts with a long lead time. The RFC's operational streamflow forecasts are usually made with a lead time much longer than the QPF duration they adopt. At the WGRFC, the lead time of the streamflow forecasts is 5 days, no matter how frequently the forecasts are made. For the operational purpose, only one set of streamflow forecasts is made using a fixed QPF duration for a given event (Table A.3). Performance of these single pieces of streamflow forecast of long lead times deserves our attention because they are to provide a prediction of the entire flood event and can be taken as inputs for floodplain inundation forecasts. A DHSVM baseline run (driven by ST4) for each storm event was compared with the forecasted streamflow outputs to assess the streamflow forecasting skill under different QPF durations. The skill statistics that were employed include R^2, NSE, RB of the mean streamflow. Additionally, relative root mean square error (RRMSE) and RB of peak streamflow are also used to the instantaneous streamflow system skill evaluation.

3) Inundation

Forecasted inundation timeseries maps were generated for all scenarios reported in Table 3. The skill of the forecasted floodplains during these three flood events was assessed against corresponding baseline floodplain simulated by TRITON using ST4-driven DHSVM streamflow. Additionally, forecasted floodplains driven by DHSVM with different maximum QPF durations (12, 24, and 72 h) were also simulated to evaluate the potential benefit of increasing the maximum QPF duration in operational streamflow forecasting. The metrics used in Wing et al. (2017) were

adopted to quantify the skill of the forecasts (see Table A.1), including the hit rate (HR), the false alarm rate (FAR), the critical skill index (CSI), and the error bias (EB).

4. Results

a. Skills of forecasted precipitation

As the duration of forecasted precipitation increases, the QPF skill statistics generally tend to worsen. The QPF skill statistics compared against ST4 QPE for the three extreme weather events in Brays Bayou are shown in Fig. 4. FAR, FB, and CSI generally perform worse with an increased QPF duration. However, longer QPF duration leads to higher HR skill due to the probability of rain accumulating for each grid as time passes. Fig. 5 compares the forecasted precipitation amount in mm/(6hr) with the ST4 QPE under different durations of interest (i.e., 12, 24, 48, and 72 h). The results indicate that the longer the QPF duration, the larger the error of the QPF rates.

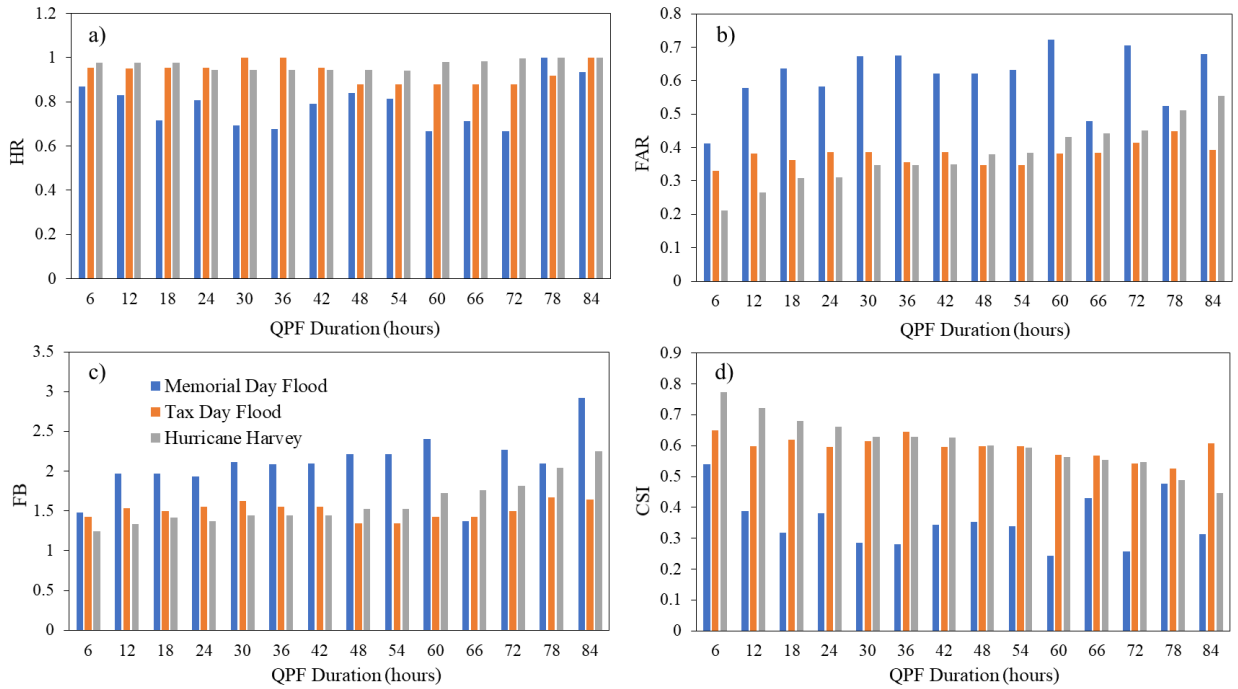


Fig. 4. (**a**) Hit rate (HR), (**b**) false alarm rate (FAR), (**c**) frequency bias (FB), and (**d**) critical success index (CSI) of QPF data for the dates of the Memorial Day flood, Tax Day flood, and Hurricane Harvey.

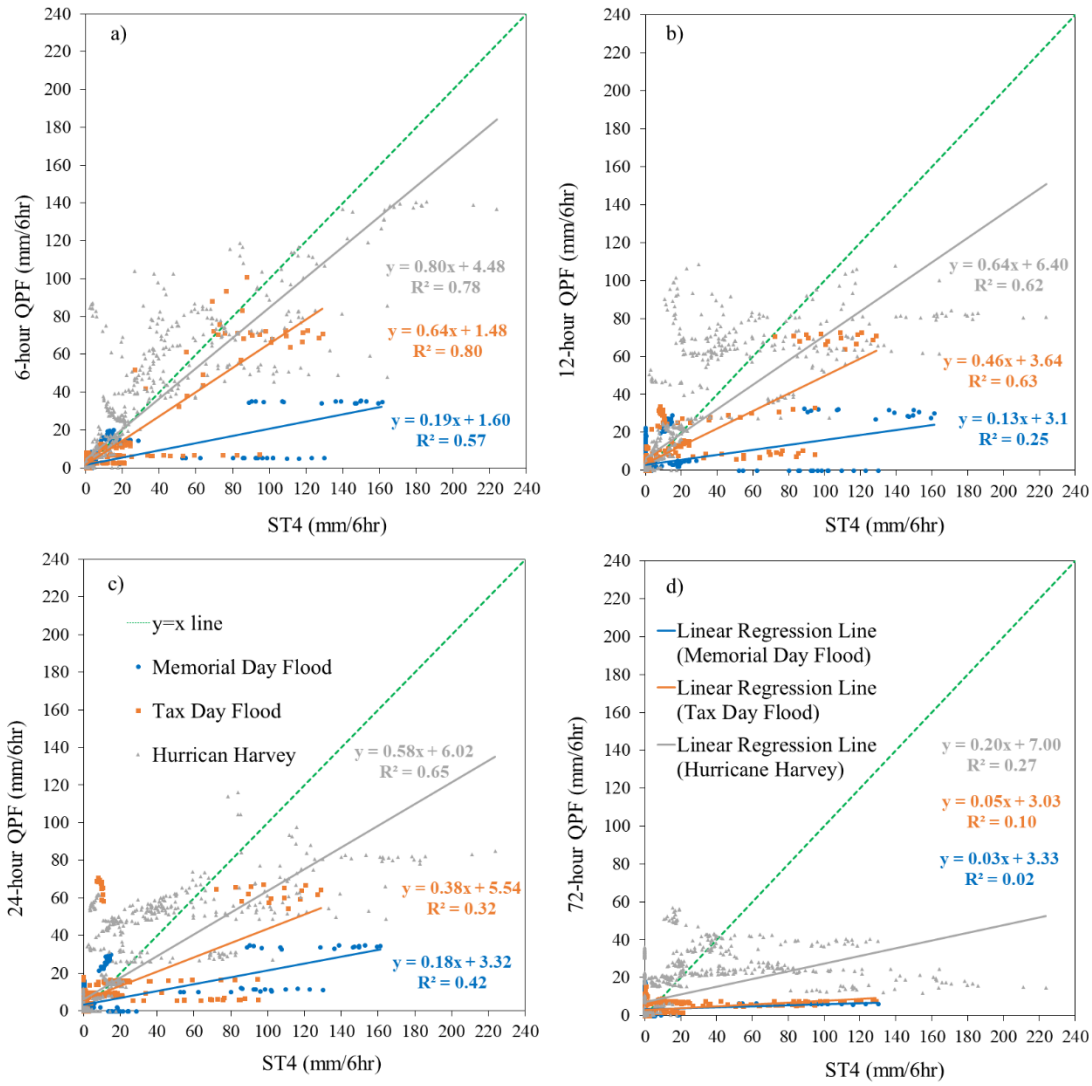


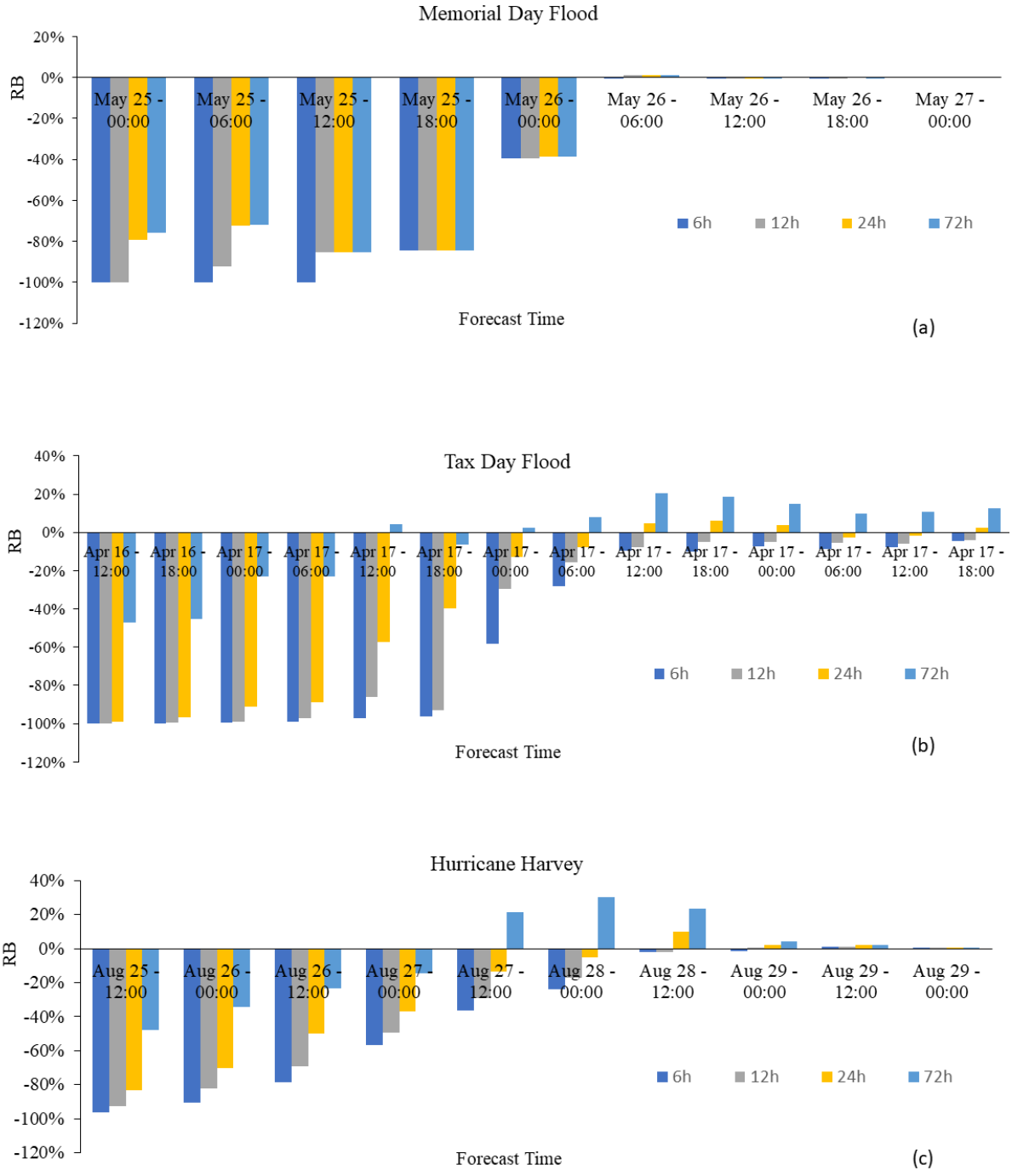
Fig. 5. Discrepancy in precipitation rates for the three case flood events by QPF duration: (**a**) 6 h, (**b**) 12 h, (**c**) 24 h, and (**d**) 72 h duration. Each data point represents 6-h accumulated precipitation.

QPF tend to overestimate the storm duration but underestimate the precipitation rates. The FB values are over 1 for the three events under all the QPF durations and tend to increase as the duration increases (Fig. 4). The results in Fig. 5 show that the QPF tend to underestimate the precipitation rate, and such underestimation increases as the rain rate increases. These findings agree with previous studies by Brown et al. (2012) and Sukovich et al. (2014), indicating that the NWS QPF underestimated rainfall during extreme precipitation events. Also, Chong et al. (2021)

indicates that the low skills in convective parametrization, and coarse resolution of modeling grids within NWP models might be causes of this phenomenon.

The QPF shows higher performance on forecasted rainfall totals as QPF duration increases—even though the QPF forecast skill metrics with regard to spatial and temporal distribution decreased. The RB of the rainfall totals of the QPF under 6, 12, 24, and 72 h durations for Hurricane Harvey are shown in Fig. 6. Using Hurricane Harvey as an example, there are notable underestimations in the forecasts made before the precipitation peak (Aug 27 3:00 am - 6:00 pm). However, this underestimation gets smaller as the QPF duration increases (Fig. 6c). Additionally, the PPTB also narrows as the QPF duration increases (Appendix Table A.4). Similar patterns were found for the Tax Day Flood, but not for the Memorial Day Flood.

QPF during the Memorial Day flood have the worst skill, while the Tax Day flood and Hurricane Harvey have comparable skill statistics (Fig. 4). This is likely due to the Memorial Day flood occurring over a short time period (9 h), while the other two events were more prolonged (24 h for the Tax Day flood and 108 h for Hurricane Harvey). In addition, the QPF rates for Hurricane Harvey are better than for the Tax Day flood (Fig. 5). High uncertainty with regard to forecasting convective events might be a reason for the lower skill of the Tax Day flood QPF (Sun et al., 2014; Chong et al., 2021). Furthermore, the higher spatial resolution of the 2017 QPF (as compared to that in 2016) might also play a role in this comparison.



357

358 **Fig. 6.** Relative Bias (RB) of the QPF at different times under 6h, 12h, 24h, and 72h QPF
359 durations during (a) the Memorial Day flood; (b) the Tax Day flood; and (c) Hurricane Harvey

b. Skills of forecasted streamflow

For the instantaneous streamflow forecasts, the streamflow forecast skills generally worsen as the duration increases (Fig. 7), which is consistent with the precipitation QPF statistics. In addition, streamflow forecasts show higher skill in Hurricane Harvey and the Tax Day flood than the Memorial Day flood, similar to QPF skill evaluation results.

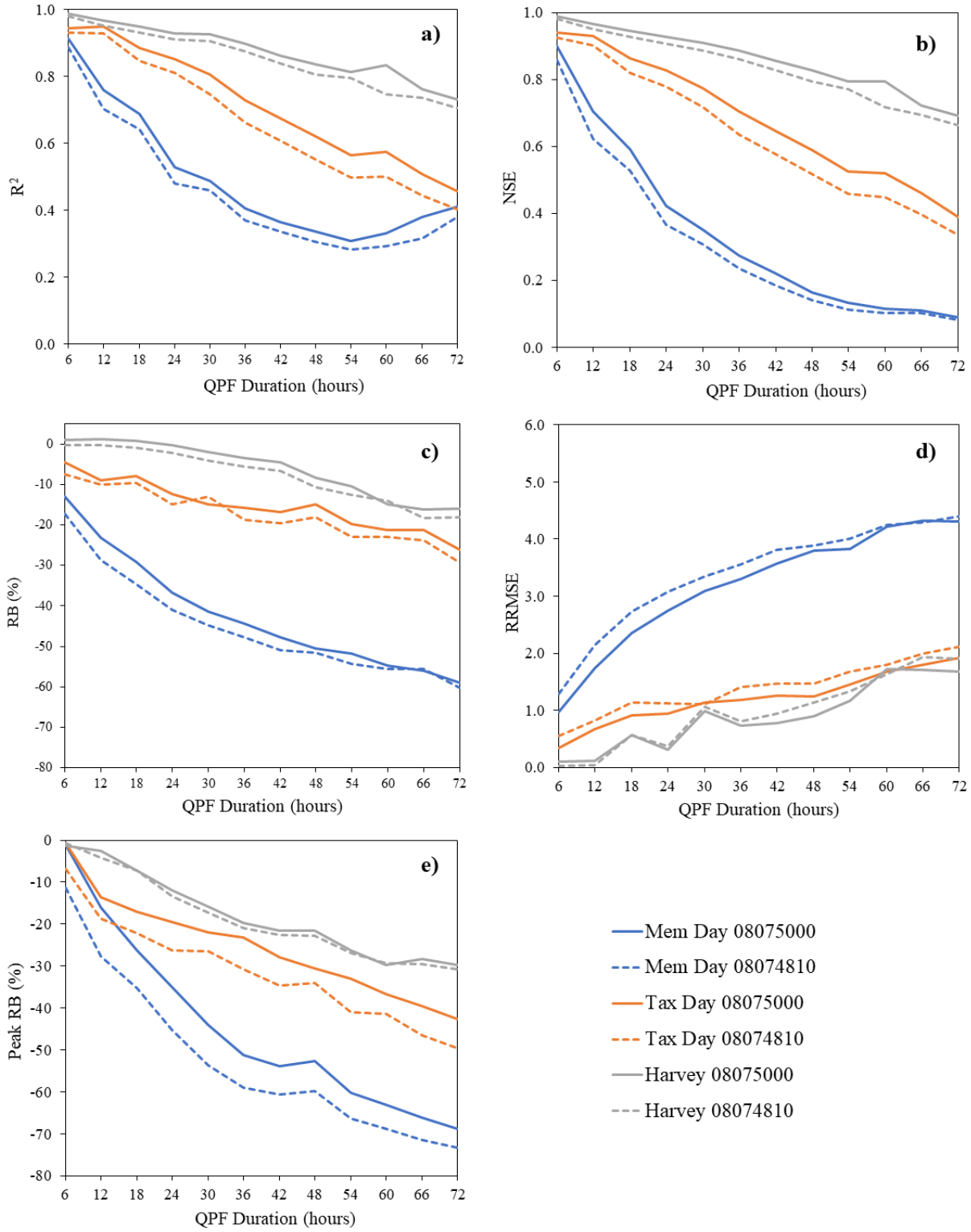


Fig. 7. Skill statistics for the instantaneous streamflow forecasts by QPF duration when compared with the baseline model run for each storm event for gages 08074810 and 08075000: (**a**) R-Squared (R^2), (**b**) Nash–Sutcliffe Efficiency (NSE), (**c**) Relative Bias (RB), (**d**) Relative Root Mean Square Error (RRMSE), and (**e**) Peak Relative Bias (Peak RB).

The mean streamflow is underestimated for almost all QPF durations during the three events, as indicated by comparisons between the forecasted mean streamflow and the baseline modeled streamflow (Fig. 8). This is attributed to the negative bias of QPF, as discussed in Section 4a. This negative bias is exacerbated as the QPF duration increases (Fig. 7(c) and Figure 8). This is because the streamflow forecasts are more affected by the accumulation of the QPF errors over time, while the benefit of initializing the model using the ST4 QPE data becomes less significant (the streamflow forecasts are generated with ST4 QPE prior to the forecasted precipitation).

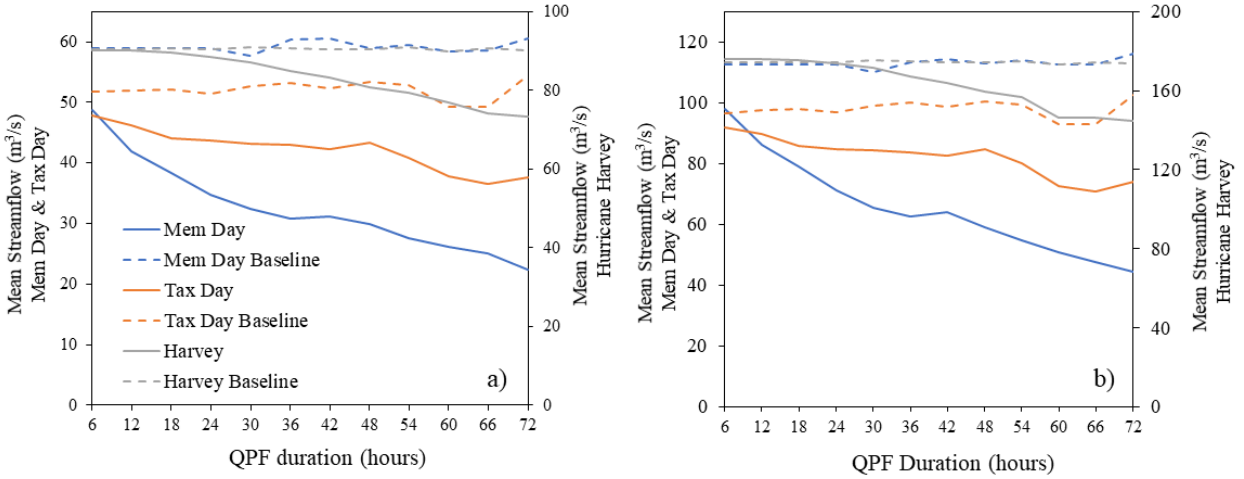


Fig. 8. Forecasted mean streamflow compared with baseline mean streamflow for USGS stations (**a**) 08074810 and (**b**) 08075000.

In instantaneous streamflow forecasts, only hydrography segments within the period of the QPF duration are used (Fig. A3). The streamflow forecast performance within the period of a QPF duration rests on the accuracy of the precipitation rate estimated for this period. As the QPF duration increases, the underestimation of precipitation rates in QPF gets exacerbated (Fig. 5). Therefore, the evaluation skills of instantaneous streamflow forecasts are in consistence of the QPF forecast skills.

386 However, for long lead time streamflow forecasts made at the onset and the middle of
387 Hurricane Harvey (Fig. A.5), as the QPF duration increases, the performance of the streamflow
388 forecasts gets better, as indicated by higher NSE, R^2, and lower RB (Fig. 9). For the forecasts made
389 at the onset of the Tax Day Flood (Forecast 18 in Table 4), no noticeable differences were found
390 in the R^2 and NSE under different QPF durations. However, the RB values gradually approached
391 zero when the QPF duration was extended from 12 to 24, then 48, and finally 72 h (Fig. 9 (c)).
392 These trends indicate that the long lead time streamflow forecasting skills increase as the
393 QPF duration increases for prolonged extreme events (like Hurricane Harvey and the Tax Day
394 Flood). This is because shorter duration QPF tends to underestimate rainfall totals since it has a
395 lower probability to cover an entire extreme weather event. For instance, due to the particularly
396 long extent of Hurricane Harvey, the 24-h duration period is only about one fifth of the total storm
397 duration. In the case of the Tax Day Flood, the 24-h duration QPF covers most of the event, but
398 still underestimates the total precipitation amount because the QPF underestimates the
399 precipitation rates (Section 4a). This underestimation of rainfall totals in QPF (Fig. 6) leads to an
400 underestimation of streamflow volumes.

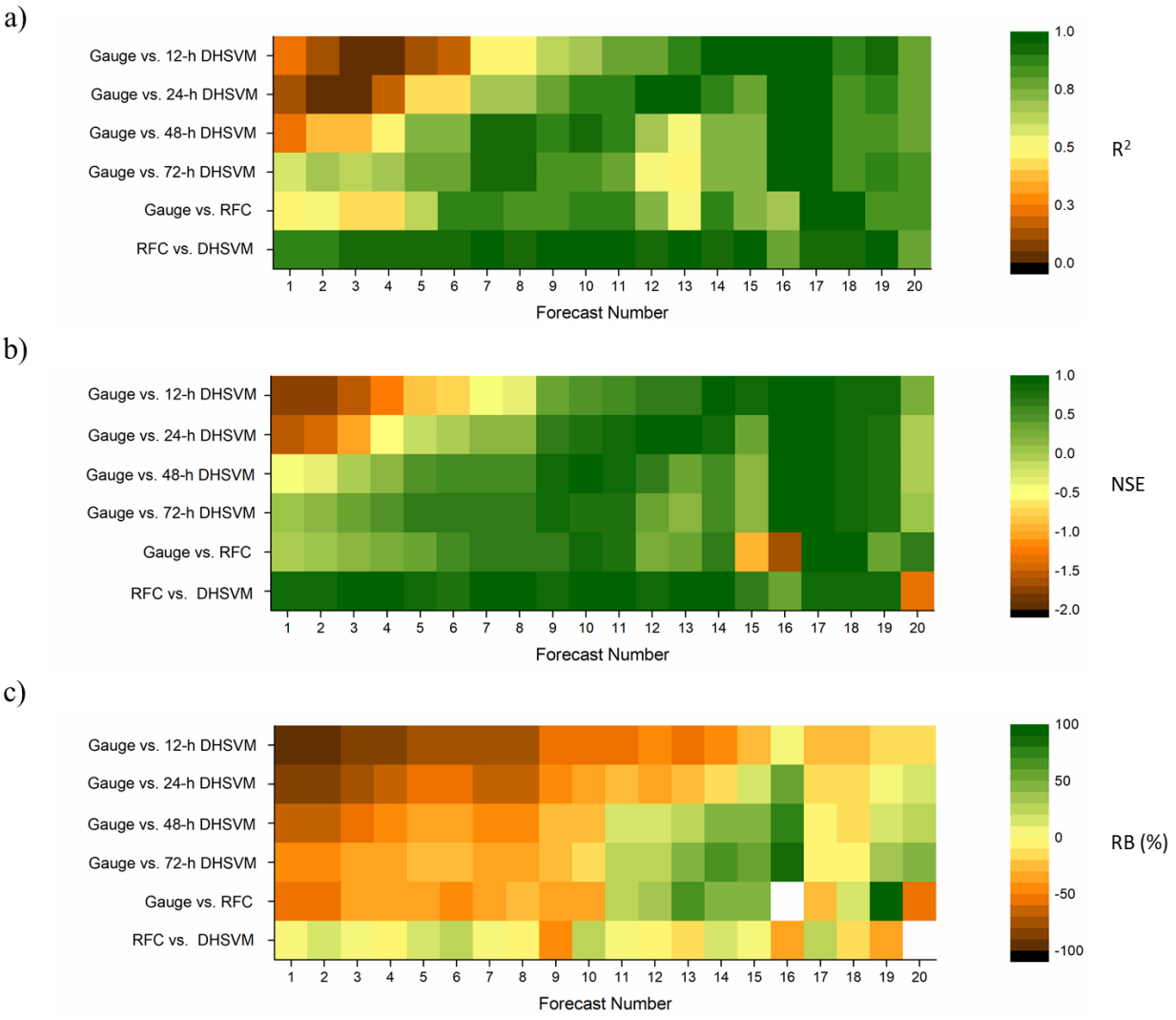


Fig. 9 To compare the DHSVM streamflow forecasts with the RFC forecasts, 17 72-h duration QPF based streamflow forecasts during Hurricane Harvey and three 12-h duration QPF based streamflow forecasts during the Tax Day Flood for Brays Bayou (USGS Gage 08075000) were obtained from the WGRFC website (Table A3). (a) R^2, (b) NSE, and (c) relative bias (RB, in %) for comparison of forecasts with different lead times during Hurricane Harvey (Forecast Number from 1 to 17) and Tax Day Flood (Forecast Number from 18 to 20) by issuance time (Table A3). The last row of each figure represents the statistics for the RFC forecasts (72-h for Hurricane Harvey and 12-h for the Tax Day Flood) versus the DHSVM forecasts. The DHSVM forecasts can be used as a substitute for the RFC forecasts because the statistics are

410 good (high R^2 and NSE, and low RB) between the DHSVM and RFC forecasts. Additionally, two blank
411 blocks in (C) indicate that the values of RB are larger than 100%.

412 *c. Skills of forecasted inundation*

413 The evaluation results of the floodplain inundation forecasts for the three flood events are
414 listed in Table 4.

415 **Table 4**

416 Inundation forecast cases and evaluation metrics

Case Events	Lead time	Case	Forecast time	HR	FAR	CSI	EB
Memorial Day Flood	12-h	1	20150525_12	0.1871	0.0000	0.1871	0.0000
		2	20150525_18	0.2002	0.0000	0.2002	0.0000
		3	20150526_06	0.9995	0.0016	0.9980	3.3000
		4	20150527_12	0.9998	0.0002	0.9996	0.9444
	24-h	5	20150525_12	0.1871	0.0000	0.1871	0.0000
		6	20150525_18	0.2002	0.0000	0.2002	0.0000
		7	20150526_06	0.9995	0.0016	0.9980	3.3000
		8	20150527_12	0.9998	0.0002	0.9996	0.9444
	72-h	9	20150524_06	0.2132	0.0000	0.2132	0.0000
		10	20150525_18	0.2013	0.0000	0.2013	0.0000
		11	20150526_06	0.9996	0.0016	0.9980	3.5164
Tax Day Flood	12-h	12	20160417_12	0.1827	0.0000	0.1827	0.0000
		13	20160417_18	0.1282	0.0000	0.1282	0.0000
		14	20160418_12	0.9990	0.0088	0.9903	8.6427
		15	20160419_12	0.9997	0.0005	0.9991	1.5497
	24-h	16	20160417_12	0.4390	0.0000	0.4390	0.0000
		17	20160417_18	0.6313	0.0000	0.6313	0.0000
		18	20160418_12	0.9987	0.0123	0.9867	9.7827
		19	20160419_12	0.9997	0.0006	0.9991	1.9384
	72-h	20	20160416_12	0.4263	0.0000	0.4263	0.0000
		21	20160417_12	0.8622	0.0011	0.8612	0.0083
		22	20160417_18	0.8007	0.0002	0.8005	0.0012
Hurricane Harvey	12-h	23	20170825_12	0.1149	0.0000	0.1149	0.0000
		24	20170826_12	0.5406	0.0000	0.5406	0.0000
		25	20170827_12	0.9936	0.0173	0.9767	2.7081
		26	20170828_12	0.9997	0.0002	0.9995	0.7563
		27	20170829_12	0.9998	0.0002	0.9996	0.9396
	24-h	28	20170825_12	0.2359	0.0000	0.2359	0.0000

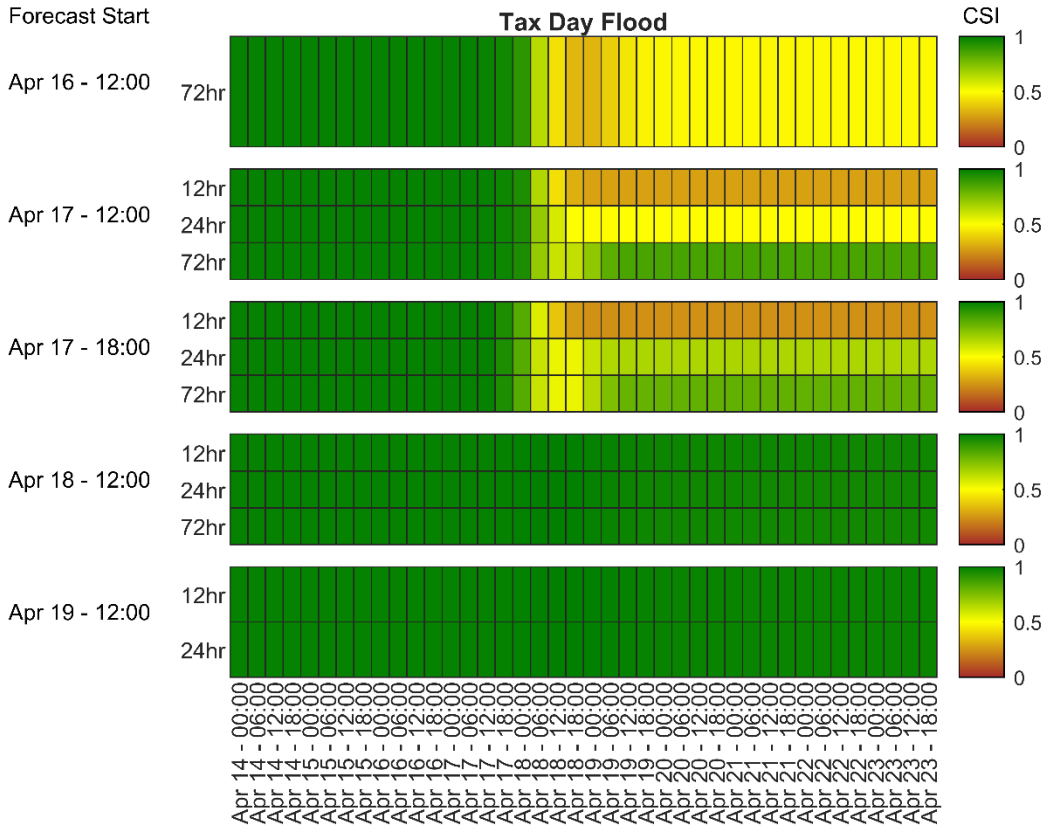
		29	20170826_12	0.7351	0.0000	0.7351	0.0000
		30	20170827_12	0.9977	0.0202	0.9779	8.5962
		31	20170828_12	0.9998	0.0006	0.9992	3.2872
		32	20170829_12	0.9998	0.0002	0.9996	0.9396
	72-h	33	20170825_12	0.5237	0.0000	0.5237	0.0001
		34	20170826_12	0.7632	0.0001	0.7631	0.0003
		35	20170827_12	0.9977	0.0204	0.9777	8.9491
		36	20170828_12	0.9998	0.0008	0.9991	3.9091
		37	20170829_12	0.9998	0.0002	0.9996	0.9396

1) Inundation forecast skill under different QPF durations

The floodplain inundation forecast skills increased as the QPF duration increased from 12 to 24 h for the Tax Day Flood and Hurricane Harvey. For example, the CSI value increased from 0.123 to 0.631 for the forecasts made at the onset of the Tax Day Flood when the QPF duration increased from 12 h to 24 h of (Case 13 vs. Case 17). Similar results are found in the forecasts that were made during the rising limb.

Moreover, for both Hurricane Harvey and the Tax Day Flood, the floodplain inundation forecast generated from the 72 h QPF at the onset of the event shows better skill than that from the 24 h duration QPF: for instance, the CSI increased from 0.236 to 0.524 (Case 28 vs. Case 33). For inundation forecasts with a start time at the middle or during the recession of the events, high skills are achieved but no significant differences exist in the skills of inundation forecasts as the QPF duration increases (Table 4).

In addition to analyzing the forecast skills of the maximum inundation extents, the skills throughout the events were also examined. Fig. 10 shows the variations in the Tax Day Flood CSI values since the start of the forecasts under different QPF durations. Inundation forecasts that are based on 72-h duration QPF achieve higher CSI values within the time of the events, than that based on 24-h QPF and 12-h QPF. This is also true for Hurricane Harvey (Fig. A.6).



435

436 Fig. 10. The CSI timeseries for each forecast of the Tax Day Flood

437 The inundation forecasts for prolonged extreme events benefit from longer duration QPF,
438 as indicated by cases of Hurricane Harvey and the Tax Day Flood. This trend is consistent with
439 the increasing trends with regard to the skill of single piece streamflow forecasts (with a long lead
440 time). Although the shorter duration QPF provide better forecasting for the precipitation rates
441 within its duration, they result in larger underestimation of the rainfall totals (as compared to the
442 long duration QPF, as suggested in Section 4.b). This leads to low flood streamflow estimates,
443 which ultimately leads to an underestimation of the inundation extent (volume).

444 The Memorial Day Flood is an exception due to its short duration. The precipitation during
445 the second 12 h was zero in the 24-h duration QPF. This leads to the same inundation forecast

skill for the results with the 12- and 24-h duration QPF. Additionally, the HR and CSI for inundation forecasts based on 72 h duration QPF do not noticeably improve.

2) Skill of forecasts beginning from different timings

Floodplain inundation forecasts made before the events and at the onset of the events are more beneficial for emergency response, although the forecasts made around peaks and during recession limbs of the floods typically reach extraordinarily high forecast skill (such as Cases 3, 4, 7, 8, 15, 19, 25, and 30 in Table 4). For the Memorial Day flood and Tax Day flood, the inundation forecasts created at the onset of the event and 6 hours before the event are very similar (Case 1 vs Case 2; Case 16 vs Case 17). Additionally, two scenarios (Cases 9 and 20 in Table 4) were tested to see whether the 72 h QPF would make reliable inundation forecasts for early warning (at least 24 h before the flood events) of such extreme events. The HR values reach only 0.213 and 0.426 for the Memorial Day Flood (Case 9) and the Tax Day Flood (Case 20), respectively, with the FAR being 0.

For flood events with particularly long durations like Hurricane Harvey, the skill of the forecasts beginning from the event onset is very low with 12- and 24-h duration QPF. CSI values are 0.115 and 0.236 in Case 23 and 24 (Table 4), respectively, and increased to 0.524 with the 72-h duration QPF. However, the forecast skill significantly increased when the forecasts were made during the rising limb of the event, which was closer to the flood peak. For inundation forecasts after the flood peak, there is not much space for improvement of HR; however, FAR decreased significantly under all three QPF durations. Again, this is not only for the maximum inundation extent forecasts, but also for the forecasted inundation throughout the flood events, as shown in Fig. 11.

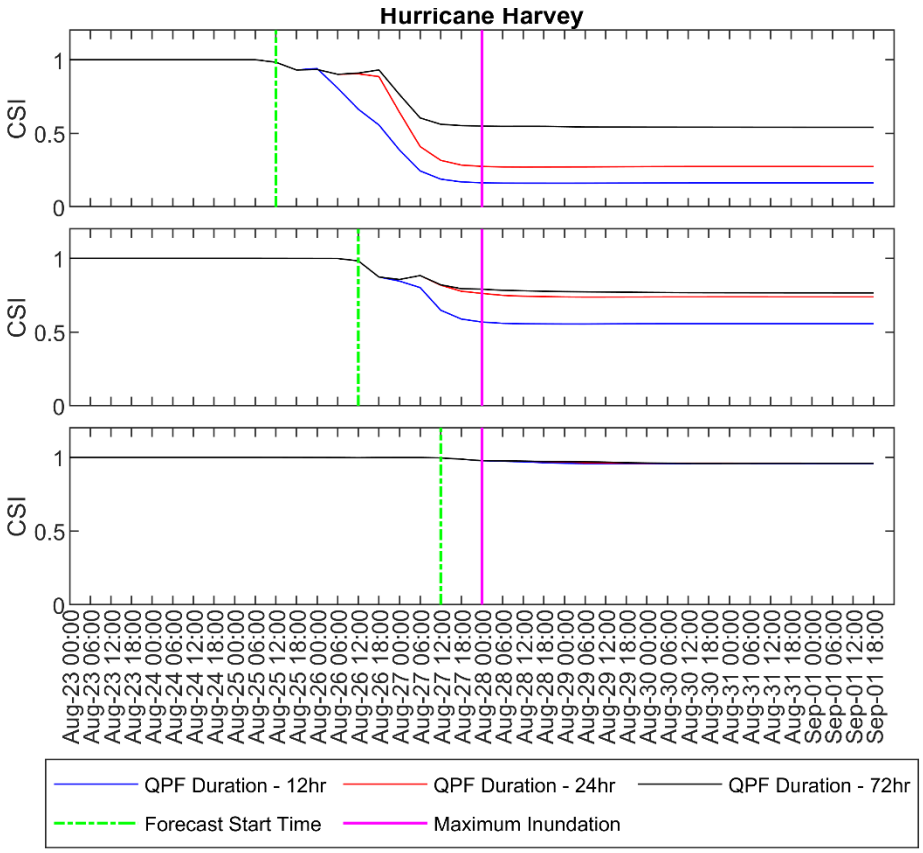


Fig. 11. CSI time series of inundation forecasts made at the onset, during the rising limb, and around the flood peak of Hurricane Harvey

Lastly, the negative bias of the streamflow forecasts, originating from the negative bias of the QPF, propagates through the flood forecasting system to the inundation forecasts. This is indicated by the tendency of inundation forecasts to underpredict the maximum floodplain extent, as evidenced by the extremely low values of EB ($0 \leq EB < 1$ indicates a underprediction tendency).

5. Discussion

a. Uncertainty associated with the forecasting skill evaluations

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; Demargne et al. 2014). While precautions were taken to use high-quality input data (e.g., 10-m DEM resampled to 20 m, high-resolution soil and precipitation data), all datasets

introduce sources of error into the modeling process. 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 error. In the case of the hydraulic model, the model is initialized 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 a higher soil moisture content), there are errors associated with this modeling setup. Despite these uncertainties, the modeling system are reliable in streamflow and floodplain inundation simulation and suffice the study goals in this research.

Moreover, Berghuijs et al. (2016) have shown that the flood-generating mechanism can vary among different regions across the USA. This case study was conducted in Houston, where the single largest precipitation event is the dominant flood-generating mechanism (Berghuijs et al. 2016). It is not clear whether comparable results will be obtained in regions with other flood-generating mechanisms such as the single largest series of precipitation events (Berghuijs et al. 2016). Furthermore, the study watershed is highly urbanized, and the flooding response to precipitation tends to be different with catchments with low urbanization levels (Munoz et al. 2018; Shao et al. 2020). In addition, various flood forecasting systems adopt various hydrological and hydrodynamic models, so the forecasting skill performance of other flood forecasting systems might be different due to structural model uncertainty.

Meanwhile, this work is conducted with deterministic QPF, though probabilistic QPF has been available from NWS. Future work should consider probabilistic QPF versus probabilistic flow forecasts as the ensemble also propagates non-linearly.

b. Relationship between event total precipitation and inundation areas/volumes

Based on the inundation forecasting scenarios and control runs of all three flood events (Table 4), a nonlinear relationship is found between the total event precipitation (in mm) and the inundation area/volume (Fig. 12). Here, inundation volume means the volume of floodwater in the inundation area. For the control runs, the total event precipitation was solely based on the ST4 estimates. For each inundation forecasting scenario, the total event precipitation equals to the sum of the ST4 estimates (before the forecast period) and the QPF estimates (during the forecast period). For extreme events with total event precipitation less than 600 mm, the inundated area increases drastically with increasing precipitation. However, when the total event precipitation is greater than 600 mm, the additional rainfall does not lead to noticeable additional increases in inundation area/volume.

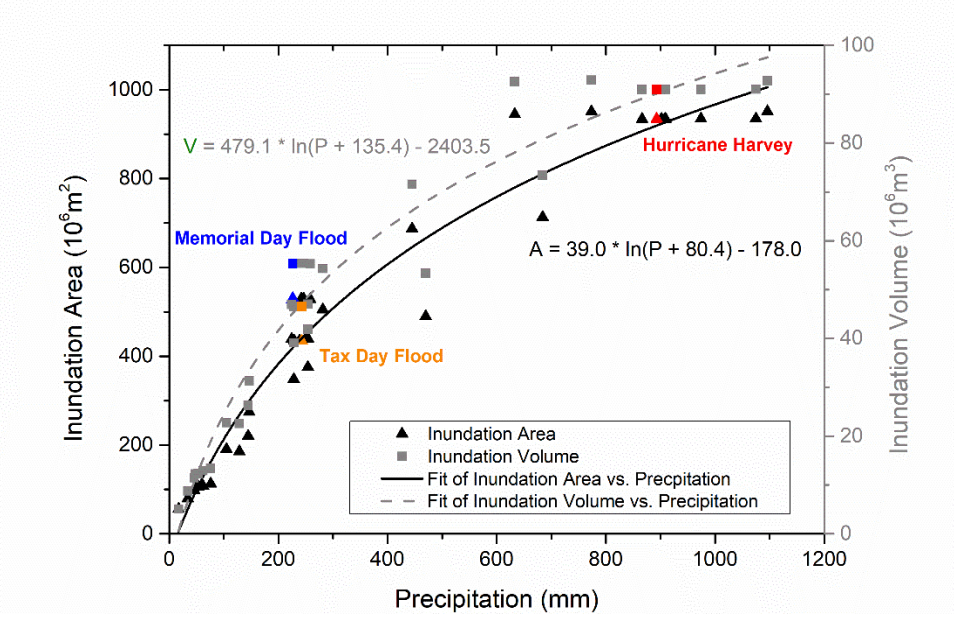


Fig. 12. The scatter plots of Precipitation-Area (P-A) relationship and Precipitation-Volume (P-V) relationship based on the inundation simulations from the control runs (labeled in different colors) and the forecast cases.

The implication of this phenomenon is that the relative bias (RB) of the QPF will not lead to significantly undermined performances of inundation forecasts for extreme events with precipitation at magnitudes comparable with Harvey. This is verified by the fact that, although the forecasted precipitation varied significantly for Hurricane Harvey under different scenarios, high inundation forecast skills were achieved for most scenarios. For instance, the HR reached 0.735 under Case 29 (Table 4), with the forecasted total event precipitation being only 49.8% of the ground truth. However, for events with precipitation of smaller magnitudes, the RB of event precipitation will lead to much more significant impacts on the inundation forecasting skills. For example, when the forecasted total event precipitation of the Tax Day Flood reached 53.0% of the ground truth, the HR reached only 0.426 (Case 20, Table 4).

c. Implications on streamflow and inundation forecasting during prolonged extreme events

Adams and Dymond (2019) examined the impacts of QPF duration on the skills of real-time hydrological forecasting over the Ohio River Basin for a 600-d period on daily basis. Overestimations of flood stage values were found in the forecasts during periods when the river above the flood level, and these overestimations were exacerbated as the QPF duration increased. It was further concluded that QPF with a longer duration is likely to lead to overall lower streamflow forecast skills. Thus, they suggested that the use of deterministic QPF should be restricted to short durations (i.e., 6-12-h). In contrast, our study focused on streamflow forecasts during selected extreme events (instead of daily streamflow across a period of nearly 20 months). We found that streamflows during the selected extreme events were underestimated. Also, it shows higher forecasting skill under shorter QPF durations when the performance of the instantaneous streamflow forecast system (i.e., a forecast system where the lead time is the same as the QPF duration) is examined. This agrees with the findings by Adam and Dymond (2019). Nevertheless,

to identify a single streamflow forecast which can best represent the overall event, a long lead time (e.g., 5 days) is preferred. The adoption of long duration QPF (e.g., 24 h or longer) can benefit long lead time streamflow forecasts during prolonged extreme events. For instance, streamflow forecasts made at the onset and the rising limb of these events showed higher R^2, NSE, and lower RB measurements when longer duration QPF was used. More importantly, such extended QPF duration contributes to better performances of the inundation forecasts (from hydrodynamic modeling) during these extreme events.

Currently, river authorities typically provide flood outlooks with streamflow forecasts and river levels that are converted from forecasted streamflow rates based on empirical methods. Thus, existing research has paid more attention to the effects of QPF on hydrologic forecasts than on hydrodynamic forecasts for flood management (Adams and Dymond, 2019). Sikder et al (2019) examined the performance of inundation forecasting performance based on QPF. However, the forecast inundation extents were derived by predeveloped rating curves. Although this approach avoids the low update frequency issue that 2D inundation modeling is faced with, this method is only applicable at locations where dense water level stations are available. To the best of our knowledge, this study is the first to investigate the impacts of increased QPF duration on forecasting skill throughout the entire process—from precipitation, to streamflow, to floodplain inundation forecast (based on hydrodynamic modeling).

6. Summary and Conclusions

In this study, a flood forecasting system was analyzed in the Brays Bayou Watershed (Houston, TX) for three extraordinary storm events from the last decade: the Memorial Day flood (2015), the Tax Day flood (2016), and Hurricane Harvey (2017). Skill statistics were calculated at each step of the flood forecasting system: precipitation forecasts (QPF), streamflow forecasts from

563 the DHSVM model driven by the QPF data, and floodplain inundation forecasts produced from
564 the TRITON model, forced by the forecasted streamflow. The main conclusions are:

565 1) The QPF for shorter, less intense events (e.g., the Memorial Day flood) have lower skill
566 statistics than the QPF for longer, more intense events (e.g., Hurricane Harvey). The skill
567 of the QPF data generally decreases with increased durations and the QPF data tend to be
568 negatively biased.
569 2) The negative bias of the QPF data is reflected in the generally underestimated forecasted
570 streamflow results, as the forecasted streamflow relies on the accuracy of the QPF data.
571 Thus, the instantaneous streamflow forecasting system also shows decreased skill with
572 increased QPF duration. For extended events (e.g., Hurricane Harvey), the streamflow skill
573 scores are higher than for shorter events. This could be due to the long duration of the event,
574 the uniform precipitation coverage of the watershed, and the better ability of the QPF data
575 to forecast the rainfall depth in this event. The underestimated streamflow forecasts led to
576 universal underprediction of floodplain inundation area.
577 3) The inundation forecasts before the flood peaks showed low skill under multiple QPF
578 durations for flood events with short durations (e.g., the Memorial Day flood). For extreme
579 flood events of relatively longer durations (e.g., the Tax Day flood), inundation forecasts
580 based on single streamflow forecasts with longer duration QPF achieve higher skill.
581 Forecasts under 24- and 72-h QPF durations might provide decent inundation forecasts
582 around the onset of a flood event. For extreme events with especially long durations like
583 Hurricane Harvey, the inundation forecasts made at a timing closer to the flood peak
584 achieved much higher skill.

4) Increasing the QPF duration to 72-h improved the skill scores for long lead time (e.g., 5-day) streamflow forecasts at the beginning and middle of Hurricane Harvey. The forecasted inundation maps based on a single streamflow forecast with 72-h duration QPF showed higher skill than those based on streamflow forecasts with 12-h and 24-h duration QPF. Therefore, the decision by the WGRFC to base the streamflow forecast on 72-h instead of 12-h duration QPF is a skillful and useful forecasting decision for hurricane-level events.

This study demonstrates the value of adopting long duration QPF in long lead time streamflow forecasts for prolonged extreme events. Furthermore, these long lead time streamflow forecasts are more beneficial for flood inundation mapping. This finding is important for QPF application in flood forecasts, since there is an increasing demand from emergency response programs on flood inundation predictions. In addition, we recommend deriving floodplain inundation forecasts when the forecasted rainfall totals during an extreme event exceed a certain threshold, rather than when streamflow forecasts exceed a certain threshold, to reduce the negative impacts from the forecast error propagation (from rainfall to streamflow, and then to inundation).

Acknowledgments

This study was supported by the U.S. National Science Foundation (Grants CBET-1454297 and CBET-1805584). Cheryl Rankin was also partially supported by a Minority Fellowship from Texas A&M University. The research used the resources of the Oak Ridge Leadership Computing Facility at Oak Ridge National Laboratory, which is a Department of Energy Office of Science User Facility. The work has also benefitted from the usage of the Texas A&M Supercomputing Facility (http://hprc.tamu.edu).

References

Adams, T. E., Chen, S., and Dymond R., 2018: Results from Operational Hydrologic Forecasts Using the NOAA/NWS OHRFC Ohio River Community HEC-RAS Model, *J. Hydrol. Eng.,* **23,** https://doi.org/10.1061/(ASCE)HE.1943-5584.0001663.

Adams T. E., and Dymond R., 2019: The Effect of QPF on Real-Time Deterministic Hydrologic Forecast Uncertainty, *J. Hydrometeorol.,* **20,** 1687-705, https://doi.org/10.1175/JHM-D-18-0202.1.

Apel H., Aronica G. T., Kreibich H., and Thieken A. H., 2009: Flood risk analyses-how detailed do we need to be? *Nat. Hazards.* **49** 79-98, https://doi/10.1007/s11069-008-9277-8.

Ashouri H., Hsu K. L., Sorooshian S., Braithwaite D. K., Knapp K. R., Cecil L. D., Nelson B. R., and Prat O. P., 2015: PERSIANN-CDR Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies, *B. Am. Meteorol. Soc.,* **96,** 69-+, https://doi.org/10.1175/BAMS-D-13-00068.1.

Bass B., Juan A., Gori A., Fang Z. and Bedient P., 2017 2015: Memorial Day Flood Impacts for Changing Watershed Conditions in Houston, *Nat. Hazards. Rev.,* **18,** https://doi.org/10.1061/(ASCE)NH.1527-6996.0000241.

Berghuijs W. R., Woods R. A., Hutton C. J., and Sivapalan M., 2016: Dominant flood generating mechanisms across the United States, *Geophys. Res. Lett.,* **43**, 4382-90, https://doi.org/10.1002/2016GL068070.

Bhola P. K., Leandro J. and Disse M., 2018: Framework for Offline Flood Inundation Forecasts for Two-Dimensional Hydrodynamic Models, *Geosciences,* **8**, https://doi.org/10.3390/geosciences8090346.

629 Brown, J. D., Seo, D. J., and Du, J. 2012. Verification of precipitation forecasts from NCEP's
630 Short-Range Ensemble Forecast (SREF) system with reference to ensemble streamflow
631 prediction using lumped hydrologic models. Journal of Hydrometeorology, 13(3), 808-
632 836. https://doi.org/10.1175/JHM-D-11-036.1.
633 Carpenter T. M., and Georgakakos K,. P., 2006: Intercomparison of lumped versus distributed
634 hydrologic model ensemble simulations on operational forecast scales, *J. Hydrol.,* **329,**
635 174-85, https://doi.org/10.1016/j.jhydrol.2006.02.013.
636 Caviedes-Voulliéme, D., Fernández-Pato, J., Hinz, C., 2020. Performance assessment of 2d zero-
637 inertia and shallow water models for simulating rainfall-runoff processes. J. Hydrol. 584,
638 124663.
639 Chong, M. L., Wong, Y. C., Woo, W. C., Tai, A. P., and Wong, W. K. 2021. Calibration of
640 High-Impact Short-Range Quantitative Precipitation Forecast Through Frequency-
641 Matching Techniques. Atmosphere, 12(2), 247. https://doi.org/10.3390/atmos12020247.
642 Cloke H. L., and Pappenberger F., 2009: Ensemble flood forecasting: A review, *J. Hydrol.,* **375**,
643 613-26, https://doi.org/10.1016/j.jhydrol.2009.06.005.
644 Cluckie I. D., and Xuan Y., 2008: Uncertainty Propagation in Ensemble Rainfall Prediction
645 Systems used for Operational Real-Time Flood Forecasting. *Wtr Sci Tec Libr*, **68**, 437-
646 447pp., https://link.springer.com/content/pdf/10.1007%2F978-3-540-79881-1.pdf
647 Costabile, P., Costanzo, C., Macchione, F., 2013. A storm event watershed model for surface
648 runoff based on 2d fully dynamic wave equations. Hydrol. Process. 27, 554–569.
649 Costabile, P., Macchione, F., Natale, L., Petaccia, G., 2015. Flood mapping using lidar dem.
650 Limitations of the 1-d modeling highlighted by the 2-d approach. Nat. Hazards 77, 181–
651 204.

652 Costabile, P. and Costanzo, C., 2021. A 2D-SWEs framework for efficient catchment-scale
653 simulations: Hydrodynamic scaling properties of river networks and implications for non-
654 uniform grids generation. Journal of Hydrology, 599, p.126306.
655 Cuo L., Giambelluca T. W., and Ziegler A. D., 2011: Lumped parameter sensitivity analysis of a
656 distributed hydrological model within tropical and temperate catchments, *Hydrol Process*
657 **25**, 2405-21, https://doi.org/10.1002/hyp.8017.
658 Cuo L., Lettenmaier D. P., Mattheussen B. V., Storck P., and Wiley M., 2008: Hydrologic
659 prediction for urban watersheds with the Distributed Hydrology-Soil-Vegetation Model,
660 *Hydrol. Process.,* **22,** 4205-13, https://doi.org/10.1002/hyp.7023.
661 Demargne J., Wu L. M., Regonda S. K., Brown J. D., Lee H., He M. X., Seo D. J., Hartman R.,
662 Herr H. D., Fresch M., Schaake J., and Zhu Y. J., 2014: The Science of NOAA's
663 Operational Hydrologic Ensemble Forecast Service, *B. Am. Meteorol. Soc.,* **95,** 79-98,
664 https://doi.org/10.1175/BAMS-D-12-00081.1.
665 Droegemeier K. K., Smith J. D., Businger S., Doswell C., Doyle J., Duffy C., Foufoula-Georgiou
666 E., Graziano T., James L. D., Krajewski V., LeMone M., Lettenmaier D., Mass C., Pielke
667 R., Ray P., Rutledge S., Schaake J., and Zipser E., 2000: Hydrological aspects of weather
668 prediction and flood warnings: Report of the Ninth Prospectus Development Team of the
669 US Weather Research Program, *B. Am. Meteorol. Soc.,* **81**, 2665-80,
670 https://www.jstor.org/stable/26215474?seq=1#metadata_info_tab_contents.
671 Dullo T. T., Gangrade S., Morales Hernández M., Sharif M. B., Kao S. C., Kalyanapu A. J.,
672 Ghafoor S, and Evans K J, 2021: Simulation of Hurricane Harvey Flood Event through
673 Coupled Hydrologic-Hydraulic Models: Challenges and Next Steps*, J. Flood Risk*
674 *Manag.*, **e12716**, https://doi.org/10.1111/jfr3.12716.

675 Fang S. F., Xu L. D., Pei H., Liu Y. Q., Liu Z. H., Zhu Y. Q., Yan J. W., and Zhang H. F., 2014:
676 An Integrated Approach to Snowmelt Flood Forecasting in Water Resource Management,
677 *Ieee. T. Ind. Inform.,* **10**, 548-58, https://doi.org/ 10.1109/TII.2013.2257807.
678 FEMA 2017 Flood Insurance Study, Harris County, Texas and Incorporated Areas **5** (12),
679 http://www.cityofshoreacres.us/documents/FEMA/48201CV005C.pdf.
680 Fernandez-Rivera D. C., Rodriguez-Rincon J. P., Alcocer-Yamanaka V. H., Brena-Naranjo J. A.,
681 and Pedrozo-Acuna A., 2019: Hydro-meteorological approach for the estimation of
682 hurricane-induced floods, *J. Flood Risk Manag.,* **12**, https://doi.org/10.1111/jfr3.12454.
683 Ferraro, D., Costabile, P., Costanzo, C., Petaccia, G. and Macchione, F., 2020. A spectral
684 analysis approach for the a priori generation of computational grids in the 2-D
685 hydrodynamic-based runoff simulations at a basin scale. Journal of Hydrology, 582,
686 p.124508.
687 Gangrade S., Kao S. C., Dullo T. T., Kalyanapu A. J., and Preston B. L., 2019: Ensemble-based
688 flood vulnerability assessment for probable maximum flood in a changing environment.
689 *J. Hydrol.*, **576**, 342-355, https://doi.org/10.1016/j.jhydrol.2019.06.027.
690 Georgakakos, K. P., and Hudlow, M. D., 1984. Quantitative Precipitation Forecast Techniques
691 for Use in Hydrologic Forecasting. B. Am. Meteorol. Soc. 65, 1186-200.
692 https://doi.org/10.1175/1520-0477(1984)065<1186:QPFTFU>2.0.CO;2
693 Golding B. W., 2009: Long lead time flood warnings: reality or fantasy? *Meteorol. Appl.*, **16**, 3-
694 12, https://doi.org/10.1002/met.123.
695 Gourley J. J., and Vieux B. E., 2005: A method for evaluating the accuracy of quantitative
696 precipitation estimates from a hydrologic modeling perspective, *J. Hydrometeorol.,* **6**,
697 115-33, https://doi.org/10.1175/JHM408.1.

Ibbitt R., Thompson C., and Turner R., 2005: Skill assessment of a linked precipitation-runoff flood forecasting system. *Journal of Hydrology (New Zealand)*, 91-104, https://www.jstor.org/stable/43944919?seq=1#metadata_info_tab_contents.

Jain S. K., Mani P., Jain S. K., Prakash P., Singh V. P., Tullos D., Kumar S., Agarwal S. P. and Dimri A. P., 2018: A Brief review of flood forecasting techniques and their applications, *Int. J. River Basin Ma.* **16**, 329-44, https://doi.org/10.1080/15715124.2017.1411920.

Johnson J. M., Munasinghe D., Eyelade D. and Cohen S., 2019: An integrated evaluation of the National Water Model (NWM)-Height Above Nearest Drainage (HAND) flood mapping methodology, *Nat. Hazard Earth Sys.* **19**, 2405-20, https://doi.org/10.5194/nhess-19-2405-2019.

Jung Y. and Merwade V., 2015: Estimation of uncertainty propagation in flood inundation mapping using a 1-D hydraulic model, *Hydrol. Process,* **29**, 624-40, https://doi.org/10.1002/hyp.10185.

Kao S. C., DeNeale S. T., and Watson D. B., 2019: Hurricane Harvey Highlights: Need to Assess the Adequacy of Probable Maximum Precipitation Estimation Methods, *J. Hydrol. Eng.,* **24**, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001768.

Ko M. C., Marks F. D., Alaka G. J., and Gopalakrishnan S. G., 2020: Evaluation of Hurricane Harvey (2017) Rainfall in Deterministic and Probabilistic HWRF Forecasts, *Atmosphere-Basel,* **11**, https://doi.org/10.3390/atmos11060666.

Lehmann J., Coumou D. and Frieler K. 2015 Increased record-breaking precipitation events under global warming *Climatic Change,* **132**, 517-8, https: // doi.org/10.1007/s10584-015-1434-y

720 Li J., Chen Y. B., Wang H. Y., Qin J. M., Li J., and Chiao S., 2017: Extending flood forecasting
721 lead time in a large watershed by coupling WRF QPF with a distributed hydrological
722 model *Hydrol. Earth Syst. Sc.,* **21**, 1279-94, https://doi.org/10.5194/hess-21-1279-2017.
723 Li X. D., Zhao G., Nielsen-Gammon J., Salazar J., Wigmosta M., Sun N., Judi D., and Gao H.
724 L., 2020: Impacts of urbanization, antecedent rainfall event, and cyclone tracks on
725 extreme floods at Houston reservoirs during Hurricane Harvey, *Environ. Res. Lett.* **15**,
726 https://doi.org/10.1088/1748-9326/abc4ff.
727 Lin Y., 2011: GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage
728 IV Data. Version 1.0. UCAR/NCAR – Earth Observing Laboratory.
729 https://doi.org/10.5065/D6PG1QDD. Accessed 14 Feb 2019
730 Lin Y., and Mitchell K. E., 2005: "The NCEP stage II/IV hourly precipitation analyses:
731 Development and applications." 19th Conf. on Hydrology, San Diego, CA. Meteor. Soc.,
732 1.2.
733 https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.182.2080&rep=rep1&type=pd
734 f
735 Marchok T., Rogers R., and Tuleya R., 2007: Validation schemes for tropical cyclone
736 quantitative precipitation forecasts: Evaluation of operational models for US landfalling
737 cases *Weather Forecast,* **22**, 726-46, https://doi.org/10.1175/WAF1024.1.
738 Merwade V., Olivera F., Arabi M. and Edleman S., 2008: Uncertainty in flood inundation
739 mapping: Current issues and future directions, *J. Hydrol. Eng.,* **13,** 608-20,
740 https://doi.org/10.1061/(ASCE)1084-0699(2008)13:7(608).
741 Morales-Hernandez M., Sharif M. B., Gangrade S., Dullo T. T., Kao S. C., Kalyanapu A.,
742 Ghafoor S. K., Evans K. J., Madadi-Kandjani E., and Hodges B. R., 2020: High-

performance computing in water resources hydrodynamics, *J. Hydroinform.,* **22**, 1217-35, https://doi.org/10.2166/hydro.2020.163.

Morales-Hernández M., Sharif M. B., Kalyanapu A., Ghafoor S. K., Dullo T. T., Gangrade S., Kao S. C., Norman M. R., and Evans K. J., 2021: TRITON: A Multi-GPU Open Source 2D Hydrodynamic Flood Model, *Environ. Modell. Softw.,* **105034**, https://doi.org/10.1016/j.envsoft.2021.105034.

Munoz L. A., Olivera F., Giglio M., and Berke P., 2018: The impact of urbanization on the streamflows and the 100-year floodplain extent of the Sims Bayou in Houston, Texas *Int J. River. Basin Ma.,* **16**, 61-9, https://doi.org/10.1080/15715124.2017.1372447.

Nelson B. R., Prat O. P., Seo D. J., and Habib E., 2016: Assessment and Implications of NCEP Stage IV Quantitative Precipitation Estimates for Product Intercomparisons, *Weather Forecast,* **31**, 371-94, https://doi.org/10.1175/WAF-D-14-00112.1.

Nauman R., Dangermond M., and Frye C., 2018: United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database for Texas. Retrieved from https://websoilsurvey.nrcs.usda.gov/

Parker D. J., Priest S. J., and Tapsell S. M., 2009: Understanding and enhancing the public's behavioural response to flood warning information, *Meteorol. Appl.,* **16**, 103-14, https://doi.org/10.1002/met.119.

Parodi U., and Ferraris L., 2004: Influence of stage discharge relationship on the annual maximum discharge statistics, *Nat. Hazard*, **31**(3), 603-611, https://link.springer.com/content/pdf/10.1023/B:NHAZ.0000024893.57284.0e.pdf.

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, *Atmos. Res.,* **100**, 150-67, https://doi.org/10.1016/j.atmosres.2010.11.016.

Sanders B. F., and Schubert J. E., 2019: PRIMo: Parallel raster inundation model, *Adv. Water Resour.* **126**, 79-95, https://doi.org/10.1016/j.advwatres.2019.02.007.

Sapiano M. R. P., and Arkin P. A., 2009: An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data, *J. Hydrometeorol.* **10**, 149-66, https://doi.org/10.1175/2008JHM1052.1.

Schumann G. J. P., Neal J. C., Voisin N., Andreadis K. M., Pappenberger F., Phanthuwongpakdee N., Hall A. C., and Bates P. D., 2013: A first large-scale flood inundation forecasting model *Water Resour. Res.,* **49**, 6248-57, https://doi.org/10.1002/wrcr.20521.

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. As.,* **54**, 90-103, https://doi.org/10.1111/1752-1688.12613.

Seo B. C., Krajewski W. F., Quintero F., ElSaadani M., Goska R., Cunha L. K., Dolan B., Wolff D. B., Smith J. A., Rutledge S. A., and Petersen W. A., 2018a: Comprehensive Evaluation of the IFloodS Radar Rainfall Products for Hydrologic Applications, *J. Hydrometeorol.,* **19**, 1793-813, https://doi.org/10.1175/JHM-D-18-0080.1.

Seo, B. C., Quintero, F., Krajewski, W. F. 2018b: High-resolution QPF uncertainty and its implications for flood prediction: A case study for the eastern Iowa flood of 2016. *J. Hydrometeorol.*, **19**, 1289-1304. https://doi.org/10.1175/JHM-D-18-0046.1.

788 Shao M. Q., Zhao G., Kao S. C., Cuo L., Rankin C., and Gao H. L., 2020: Quantifying the
789 effects of urbanization on floods in a changing environment to promote water security - A
790 case study of two adjacent basins in Texas, *J. Hydrol.,* **589**,
791 https://doi.org/10.1016/j.jhydrol.2020.125154.
792 Sikder M. S., Ahmad S., Hossain F., Gebregiorgis A. S., and Lee H., 2019: Case Study: Rapid
793 Urban Inundation Forecasting Technique Based on Quantitative Precipitation Forecast
794 for Houston and Harris County Flood Control District, *J. Hydrol. Eng.,* **24**,
795 https://doi.org/10.1061/(ASCE)HE.1943-5584.0001807.
796 Sugarbaker L. J., Eldridge D. F., Jason A. L., Lukas V., Saghy D. L., Stoker J. M., and Thunen
797 D. R., 2017: Status of the 3D Elevation Program, 2015 (No. 2016-1196) US Geological
798 Survey
799 Sukovich E. M., Ralph F. M., Barthold F. E., Reynolds D. W., and Novak D. R., 2014: Extreme
800 Quantitative Precipitation Forecast Performance at the Weather Prediction Center from
801 2001 to 2011, *Weather Forecast,* **29**, 894-911, https://doi.org/10.1175/WAF-D-13-
802 00061.1.
803 Sun, J., Xue, M., Wilson, J.W., Zawadzki, I., Ballard, S.P., Onvlee-Hooimeyer, J., Joe, P.,
804 Barker, D.M., Li, P.W., Golding, B. and Xu, M., 2014. Use of NWP for nowcasting
805 convective precipitation: Recent progress and challenges. Bulletin of the American
806 Meteorological Society, 95(3), pp.409-426, https://doi.org/10.1175/BAMS-D-11-00263.1
807 Vieux B. E., Cui Z. T., and Gaur A., 2004: Evaluation of a physics-based distributed hydrologic
808 model for flood forecasting, *J. Hydrol.,* **298**, 155-77,
809 https://doi.org/10.1016/j.jhydrol.2004.03.035.
810 USGS (US Geological Survey) 2014 NLCD 2011 Land Cover (2011 Edition, amended 2014)

Wigmosta M. S., Vail L. W., and Lettenmaier D. P., 1994: A Distributed Hydrology-Vegetation Model for Complex Terrain, *Water Resour. Res.,* **30**, 1665-79, https://doi.org/10.1029/94WR00436.

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 Resour. Res.,* **53**, 7968-86, https://doi.org/10.1002/2017WR020917.

Xia Y. L., Mitchell K., Ek M., Sheffield J., Cosgrove B., Wood E., Luo L. F., Alonge C., Wei H. L., Meng J., Livneh B., Lettenmaier D., Koren V., Duan Q. Y., Mo K., Fan Y., and Mocko D., 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res-Atmos.,* **117**, https://doi.org/10.1029/2011JD016048.

Zappa M., Jaun S., Germann U., Walser A., and Fundel F., 2011: Superposition of three sources of uncertainties in operational flood forecasting chains, *Atmos Res.,* **100**, 246-62, https://doi.org/10.1016/j.atmosres.2010.12.005.

Zhang Z. Y., 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, *J. Hydrol. Eng.,* **9**, 103-15, https://doi.org/10.1061/(ASCE)1084-0699(2004)9:2(103).

Zhao G., Gao H. L., and Cuo L., 2016: Effects of Urbanization and Climate Change on Peak Flows over the San Antonio River Basin, Texas, *J. Hydrometeorol*., **17**, 2371-89, https://doi.org/10.1175/JHM-D-15-0216.1.