Journal of Hydrology 612 (2022) 128168

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

A flood predictability study for Hurricane Harvey with the CREST-iMAP model using high-resolution quantitative precipitation forecasts and U-Net deep learning precipitation nowcasts

Mengye Chen [a,b], Zhi Li [a,b], Shang Gao [b], Ming Xue [c], Jonathan J. Gourley [d], Randall L. Kolar [a], Yang Hong [a,b,*]

[a] School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, United States
[b] Hydrometeorology and Remote Sensing Laboratory, University of Oklahoma, Norman, OK, United States
[c] School of Meteorology, University of Oklahoma, Norman, OK, United States
[d] NOAA National Severe Storm Laboratory, Norman, OK, United States

ARTICLE INFO

This manuscript was handled by Jiri Simunek, Editor-in-Chief

Keywords:
Quantitative precipitation forecast
Flood
Deep learning
Hydrologic-hydraulic model
Forecasts
Hurricane Harvey

ABSTRACT

A flood is one of the most hazardous natural disasters, and it commonly causes fatalities and socioeconomic damages. The advances of modeling techniques and observation data in flood prediction have found success in field operations. This paper presents a comprehensive flood prediction of Hurricane Harvey in 1-hour lead-time that is not limited to 1D streamflow forecast but also 2D flood extent and 3D inundation depth. It uses high-resolution quantitative precipitation forecasts (QPFs, from operational Rapid Refresh-RAP, and High Resolution Rapid Refresh-HRRR models) and deep learning nowcasts (AI nowcasts). The results show that the QPFs have a well-known displacement issue and the AI nowcast cannot predict the precipitation intensity, and an attempt to combine the two methods (AI hybrid) failed to improve the overall accuracy. However, the 2D flood extent predictions with the HRRR and AI hybrid forcings can provide information indicating the future flooded area with about 50% accuracy (hit rate) and stream flow prediction showed that the HRRR QPF can provide relatively accurate flood peak prediction (-8%). In contrast, the AI nowcast reveals minimal displacement errors but underpredicts precipitation intensity. The deep learning method also indicates that the binary tests with low threshold, which are commonly employed in the deep learning field, neglect the importance of precipitation intensity errors for extreme event studies.

1. Introduction

The increase of frequency and magnitude of extreme weather and precipitation caused by climate change has received much attention (Meehl et al., 2000; USGCRP, 2017; van Oldenborgh et al., 2018). Further, 'tropicalization', a global trend towards a heavier precipitation climate, has been modeled in Europe (Gobiet et al., 2014). Consequentially, this change would lead to more flood events in the future, which is the second-deadliest and most common natural hazard in the United States and world (Ashley and Ashley, 2008; Barredo, 2007; Benito et al., 2004; Smith and Ward, 1998). According to Brauer et al. (2020), the rate and duration of precipitation, land use and soil moisture content, as well as topography all have impacts on the severity of flooding. Moreover, tropicalization combining with rising sea levels and ocean temperatures will only intensify flooding especially along coastal areas (Wing et al., 2019). To quantify and mitigate flood risks, many tools have been developed to simulate flooding at post-event (Bates et al., 2010; Bates and De Roo, 2000; de Almeida and Bates, 2013; Srinivasan and Arnold, 1994; Wang et al., 2011; Wood et al., 1992; Xue et al., 2016), and in real-time (Cohen et al., 2018; Gochis et al., 2017; Sampson et al., 2015; Wing et al., 2019, 2017). Instead of post-event hydrological reanalysis or real-time streamflow simulation, this study explores the options for forecasting flood events, which could potentially provide a longer lead-time for emergency response officials or even the general public to proactively respond to possible future flood risks.

Almost 40 years ago, Georgakakos and Hudlow (1984) discuss the

* Corresponding author at: School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, United States.
E-mail addresses: mchen15@ou.edu (M. Chen), li1995@ou.edu (Z. Li), shang.gao@ou.edu (S. Gao), mxue@ou.edu (M. Xue), jj.gourley@noaa.gov (J.J. Gourley), kolar@ou.edu (R.L. Kolar), yanghong@ou.edu (Y. Hong).

https://doi.org/10.1016/j.jhydrol.2022.128168
Received 7 March 2022; Received in revised form 30 June 2022; Accepted 2 July 2022
Available online 11 July 2022

use of Quantitative Precipitation Forecasts (QPFs) from the Numerical Weather Prediction (NWP) models for hydrological forecasting; the operational Limited-Area Fine Mesh (LFM) model (Newell and Deaver, 1981) from the National Weather Service (NWS) at the time had a horizontal grid spacing of 127 km, and the on-demand Movable Fine Mesh (MFM) model had an improved grid spacing up to 60 km. While promising, the lack of spatial specificity needed for hydrological forecasting and the relatively low accuracy of QPF limited the application of model-based QPF to drive hydrological models. With the advancement of computing and climate observation technologies, the NWP models have improved dramatically in recent decades in terms of their spatial and temporal resolutions, accuracy, and coverage (Trenberth, 2010). The current operational High-Resolution Rapid Refresh (HRRR, Benjamin et al., 2016; Lee et al., 2019) has a horizontal grid spacing of 3 km with hourly updated forecasts over the contiguous United States (CONUS). Convection-allowing, cloud-resolving NWP models have been run at 1-km grid spacing (Loken et al., 2017; Xue et al., 2013). However, accurate QPF remains a challenge because of many complex factors involved in precipitation processes (Ebert and McBride, 2000; Golding, 2000). Major errors of NWP QPFs include spatial displacement errors (Ebert and McBride, 2000), and errors in intensity especially for extreme events (Cuo et al., 2011). For short-range precipitation forecasting, the proper initialization of existing precipitation fields for the forecasting model is also an important issue (Kain et al., 2010; Sun et al., 2014).

Precipitation nowcasting typically refers to very short-range (1–6 hr) forecasting of precipitation, and the traditional nowcasting is mostly done by extrapolating observed radar reflectivity fields. Ligda (1953) demonstrated the possibility of providing reasonable forecasts based on the persistence and movement of radar echoes. This technology has a recent technological improvement with the combination of High Performance Computing (HPC) and Artificial Intelligence (AI), resulting in successful applications in real-world problems including precipitation nowcasting (Agrawal et al., 2019; LeCun et al., 2015). In 2019, Google Research published a study showing the advantage of using deep learning nowcasts to predict 1-hour lead-time precipitation over the CONUS, outperforming traditional optical flow extrapolation method, persistence, and 1-hour HRRR forecasts (Agrawal et al., 2019). Moreover, many recent studies also show successes in deep learning precipitation nowcasting. Franch et al. (2020) introduce a new deep learning model, TAASRAD19, to forecast 1-hour lead-time radar reflectivity which is then converted to precipitation rate, yielding a Critical Success Index (CSI) of 0.5 over a 9-year sample. While most AI nowcast studies evaluate binary coverage as the performance test, Kumar et al. (2020) demonstrate their deep learning model reduces the Root-Mean-Square Error (RMSE) between forecast and satellite precipitation estimates to 0.8 mm/hr on 30 min lead-time and up to 1.4 mm/hr on 150 min lead-time. Using Global Precipitation Mission (GPM) Integrated Multi-satellitE Retrievals (IMERG) datasets, they identify the most significant error source being underestimation with extreme events. However, the hydrological performance of AI nowcasts is generally underexplored.

Flood prediction has a history of different methods from upstream level exceedance thresholding, and level-to-level correlations, to hydrological and hydrodynamic models (Adams and Pagano, 2016). For modern operational flood forecasting systems, it is common to use Quantitative Precipitation Estimates (QPEs), precipitation nowcasts, and NWP QPFs to drive hydrological models (Cuo et al., 2011; Golding, 2000; Hapuarachchi et al., 2011). Since there is a lag time between the precipitation peak and the streamflow peak (Bedient et al., 1988), some operational systems are capable of predicting the flood by using the real-time observed precipitation products as the forcing dataset. For example, the Scottish Environment Protection Agency uses radar and in-situ precipitation data to provide streamflow forecasts (Werner and Cranston, 2009); the National Center of Environmental Prediction (NCEP) adapts the National Water Model (NWM) (Cohen et al., 2018; Gochis et al., 2017), while the National Severe Storm Laboratory (NSSL) utilizes the Ensemble Framework For Flash Flood Forecasting (EF5) as part of the FLASH project (Flamig et al., 2020; Gourley et al., 2017). Both NWM and EF5 use the Multi-Radar Multi-Sensor (MRMS) QPEs to drive streamflow forecasts for the entire CONUS and outer territories. To provide longer range (more than a few hours) and more comprehensive flood predictions, it is necessary to utilize nowcasts or QPFs, and the flood predictions can be extended beyond streamflow prediction to also include two-dimensional (2D) flood extent and flood depth forecasting. However, due to the accuracy issues of QPF, the skill of longer flood range prediction is still limited, especially at the scales of urban watersheds (Hapuarachchi et al., 2011). More recent hydrological evaluation of operational HRRR QPF indicates that even though this advanced product is capable of forecasting mesoscale convective systems well, QPFs associated with smaller scale precipitation systems often contain significant errors making hydrological prediction errors to become too large (Lee et al., 2019; Seo et al., 2018).

This study is an attempt to utilize advanced NWP QPFs and AI precipitation nowcasting for the prediction of flooding triggered by the landfalling Hurricane Harvey in 2017. The comprehensive predictions include 1D stream discharge, 2D flood extent, and 3D flood depth, using the Coupled Routing and Excessive STorage – inundation MApping and Prediction (CREST-iMAP) model. This work evaluates the ability of state-of-art operational QPFs to predict not only streamflow but also flood extents and flood depths, and it presents one of the first studies to evaluate the hydrological performance of AI precipitation nowcast techniques. The precipitation forecasts and flood prediction results are compared with MRMS QPE, US Geological Survey (USGS) stream gauge data, MRMS QPE simulated flood extent, Federal Emergency Management Agency (FEMA) flood insurance claims, and USGS High-Water Mark post-event survey data. The rest of this paper is organized as follows. Section 2 describes the QPF datasets, study area, CREST-iMAP model, and the deep learning architecture for AI nowcasting. Section 3 demonstrates the results of the predicted streamflow hydrographs, flood extent, and inundation depth compared to the benchmark and other ground observations. Section 4 discusses the major findings from the results. Section 5 concludes and proposes future studies.

2. Methodology

2.1. Forecast rainfall

This study relies on the applications of operational NWP models that provide the QPF products, which are archived in NOAA's National Operational Model Archive and Distribution System. The Rapid Refresh (RAP) was first developed in May 2012, built upon the first hourly updated operational NWP system in the world, Rapid Update Cycle (RUC), back in 1998 (Benjamin et al., 2016, 2004). After two major improvements in 2014 and 2016, RAP now consists of multiple meteorological data and models to increase accuracy, such as the NOAA Gridpoint Statistical Interpolation (GSI), a version of Weather Research and Forecasting (WRF) regional model, and the National Center of Environmental Prediction (NCEP) Unified Post Processor (UPP). The current RAP QPF product can provide hourly updated precipitation forecasts out to 18 h at 13-km spatial resolution over the entire North America. In this study, the 1-hour lead-time data was obtained from the NOAA's National Center for Environmental Information (NCEI) website (https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/rapid-refresh-rap). The HRRR was first released in September 2014 and finished its first upgrade in 2016, then was deployed as a new operational model in July 2018 (Lee et al., 2019). HRRR is a nested model that heavily relies on RAP data assimilation, which covers the Continental United States (CONUS) and provides hourly forecasts up to 18 h in the future at 3-km spatial resolution (Benjamin et al., 2016). In this study, the 1-hour lead-time HRRR v2 forecast data obtained from the University of Utah HRRR data archive (https://home.chpc.utah.edu/~u0553130/Brian_Blaylock/cgi-bin/hrrr_download.cgi). All

forecast products are 1-hour lead-time since the deep learning precipitation nowcasting method in this study can only produce a 1-hour forecast due to the limitation of computational resources. The basic statistics of precipitation forecast datasets are listed in Table 1.

2.2. Deep learning precipitation nowcasting

The advancement in machine learning (deep learning) recently has been claimed to be successful when applied to the short-range nowcasting of precipitation (Shi et al., 2015). AI precipitation nowcasting, as a data-driven and localized method, is adapted to the local environment and requires no prior knowledge about weather systems from the model developer, which predicts next-step precipitation distribution by a series of previous frames from a well-trained model structure. In this study, we select the U-Net model structure for the following reasons. First, it is a widely adopted structure for precipitation estimation and prediction (Li et al., 2020b; Sadeghi et al., 2020). Second, the lightweight framework necessitates a fewer number of parameters compared to other models, and requires less system memory to train the model. Before the model training, some preprocessing steps are conducted for unifying precipitation estimates from multiple events. Due to the nature of the non-Gaussian distribution of precipitation per frame, one way to normalize precipitation data is via logarithmic transformation. Similar to Sønderby et al. (2020), precipitation data is transformed and normalized based on the following equation.

$$R = \log(x + 0.01)/4$$

where $\times$ is the input rainfall rate, and R is the normalized rainfall rates. Subsequently, rainfall data is grouped every 10 frames at 6-min increments to target next-hour precipitation forecasts. The Mean Square Error (MSE) and Adam, a deep-learning optimizer (Kingma and Ba, 2017), are chosen as the objective function and optimizer in this setting. The initial learning rate (α)[1] is set at 0.001 but is scheduled to decrease exponentially with training process to avoid blowing up. Due to the limitation of the Graphics Processing Unit (GPU) memory capacity (2 GB), this study can only predict a 1-hour lead time given the size of the study area, which is 4 time-steps ahead (15 min resolution). In this study, the U-Net structured deep learning model is trained by the MRMS QPEs for 16 precipitation events listed in NOAA Storm Report (Table 2) that caused flooding in the Houston area from 2015 to 2019, except Hurricane Harvey, with a total of over 30,000 images. The intensive-precipitation-focused deep learning model is then used to forecast 1-hour lead-time from 24/08/2017 to 02/09/2017 (10 days). The experiment is conducted using a single Nvidia GTX 960 M GPU card. The training process took 48 h and the forecasting process took around 300 s for 1-hour lead time (4 time steps).

AI-based nowcasting can effectively capture the spatiotemporal correlation with observations, yet it may misrepresent the event magnitude, which is a common weakness of AI nowcast (Kumar et al., 2020). On the other hand, the physical simulations are apt at identifying storm cores and thus the event rainfall magnitude. It is presumably advantageous to take the respective advantages of both to produce a hybrid product. In this study, a conventional non-parametric probability matching scheme is taken at each nowcast time step to increase the rainfall magnitude of the AI-produced product close to the HRRR forecasts. The basic idea is to conserve the ranks of the initial rainfall rate for a given frame, and its empirical cumulative density function (CDF) is modified to match the target CDF.

[1] Learning rate is a hyperparameter that controls how much change the deep-learning model in response to estimated error each time the model weights are updated. Some considers the learning rate is the most important hyper-parameter for neural network configuring.

2.3. Observed rainfall, flood benchmark, and data

The observed precipitation product is chosen to be the MRMS radar-only, 2-min QPE, which showed high correlations and small error compared to the Harris County Flood Control District rain gauge network in previous studies (Chen et al., 2020; Li et al., 2020a). This data is obtained from the Iowa Environmental Mesonet NWS data archive (https://mesonet.agron.iastate.edu/nws/). The benchmark flood map is simulated by CREST-iMAP using MRMS 2-min QPE as the forcing data. To create the benchmark flood map, the CREST-iMAP goes through a warmup period from 01/04/2017 to 24/08/2017 using the MRMS 1-hour gauge corrected QPE and then simulates the Hurricane Harvey induced flood from 24/08/2017 to 02/09/2017 (10 days), driven by the MRMS 2-min QPE. The 10-meter resolution Manning's roughness coefficient field is derived from the landcover data from Multi-Resolution Land Characteristics Consortium (MRLC, https://www.mrlc.gov/), using a look-up table from the literature (Liu et al., 2019; McCuen, 2005).

Other benchmark data include the two USGS stream gauge (Fig. 1) data downstream of Spring (08068500) and Cypress Creek (0809000), which are obtained from the USGS Water Information System (USGS WIS, https://waterdata.usgs.gov/). The USGS High-Water-Marks (HWM) data are manually surveyed post-event water depths by measuring the residual water stain or mudlines left on buildings and other infrastructure, which is considered as the true ground observations of flood depth (Feaster and Koenig, 2017).

2.4. The CREST-iMAP model

CREST-iMAP is a hydrological and hydraulic coupled model developed by the Hydrometeorology and Remote Sensing Laboratory at the University of Oklahoma, and is designed to operate in near-real-time. It is the newly added member of a well-documented CREST modeling family, which has been widely applied to multiple operations and applications around the world (Clark et al., 2017; Flamig et al., 2020; Gourley et al., 2017; Wang et al., 2011). The CREST-iMAP coupled the water balance component of CREST and the 2D hydraulic routing using a fully solved Shallow Water Equation with a finite volume method. The model can be nested with the CREST/EF5 framework for near-real-time operation or as a standalone operation. The model receives precipitation data as the driving input over topographical datasets and the CREST water balance component generates excessive rainfall, which is then routed through the unstructured triangular mesh to generate channel flow rate, flood extent, and flood depth. The model calibration is conducted for the period from 01/04/2017 to 24/08/2017, using the DREAM algorithm (Vrugt, 2016; Vrugt et al., 2009) to optimize all water-balance parameters by targeting five different USGS stream gauges at midstream (08068275) and downstream (08068500) of Spring Creek, as well as upperstream (08068720), midstream (08068800), and downstream (08069000) of Cypress Creek. After the model is set and warmed-up, the different QPFs and AI nowcasting products were applied to CREST-iMAP as the forcing data to produce 1D streamflow, 2D flood extent, and the 3D flood depth. The simulation results were saved in the model native NetCDF format and are available through HydroShare (https://www.hydroshare.org/resource/fae24734d6fc47be8bf0b54d6a175d86/).

2.5. Study area

Harris County, TX was the most impacted area during Hurricane Harvey with 103 casualties, over 40,000 people evacuated, and over 30,000 water rescues conducted (Murphy, 2018). Fig. 1 shows the impacted area of Hurricane Harvey, the hurricane track, and the study area of this research. The statistical analysis of different QPFs is conducted throughout the entire Harris County area, and the flood analysis is conducted at the Spring basin, which is located at the northern part of

Table 1
List of precipitation estimation and forecasting productions and the basic statistics during Hurricane Harvey.

Name	Temporal resolution	Spatial resolution	Precipitation rate, mm/hr		Total precipitation (25/8/2017 to 31/8/2017), mm		
			Max	Mean	Max	Mean	Min
MRMS QPE	2 mins	1 km	202.60	6.16	1793	830	287
RAP QPF	1 h	13 km	49.83	7.45	1469	1081	730
HRRR QPF	1 h	3 km	236.87	6.65	1640	928	502
AI Nowcast	6 mins	1 km	71.27	3.18	851	190	80
AI Hybrid	6 mins	1 km	236.87	7.10	2633	930	419

Table 2
List of heavy precipitation events used for deep-learning training.

Index	Start location	Flood date	County	Fatality	Damage (million dollars)
1	Orr St.	31/10/2015	Harris, TX	2	1.7
2	SW Houston St.	18/03/2016	Harris, TX	0	0
3	NW Hockley St.	18/04/2016	Harris, TX	8	51
4	Hooks homepark	01/06/2016	Harris, TX	0	0.005
5	SW Almeda Rd.	18/01/2017	Harris, TX	0	0.5
6	S Houston St.	29/03/2017	Harris, TX	2	0.62
7	NE McNair St.	04/06/2017	Harris, TX	0	0.01
8	NE Little York Rd.	24/06/2017	Harris, TX	0	0.001
9	Satsuma Dr.	09/07/2017	Harris, TX	0	0
10	S Deer Park	21/09/2017	Harris, TX	0	0.005
11	NE Spring St.	04/07/2018	Harris, TX	0	0
12	SE Englewood St.	08/12/2018	Harris, TX	0	0
13	NW Huffman	07/05/2019	Harris, TX	0	0.25
14	NW Katy	09/05/2019	Harris, TX	0	0.05
15	E Wallis Rd.	23/08/2019	Harris, TX	0	0
16	NE Humble	19/09/2019	Harris, TX	2	565

Harris County. The hydrologic and flood inundation analysis is focused on the Spring Basin, because the two rivers are connected the Lake Houston, which is not directly entering the Trinity Bay and Gulf of Mexico. Therefore, this focusing basin is less impacted by wind-induced storm surges and this study can focus on the precipitation differences and its corresponding pluvial and fluvial flood.

2.6. Statistical metrics

To compare each forecasting product and evaluate the flood predictability, four levels of statistical tests are conducted in this study. First, the benchmark precipitation in 2.3 is used to compare QPFs and AI-based nowcasts using statistical measures. Second, the CREST-iMAP simulated hydrographs are analyzed using not only standard statistic metrics but also hydrological modeling efficiency measures. Third, the CREST-iMAP simulated flood extents are analyzed using standard binary pattern measures. Fourth, the CREST-iMAP simulated flood depth is analyzed against the USGS HWM. All four levels of statistical metrics are listed in Table 3.

The correlation coefficient (CC) over a time series measures the strength of an estimate to capture the temporal pattern of the observation. Traditionally, the relative bias is commonly used to measure the errors of the estimate as a fraction of the observation value. However, for precipitation and streamflow data, the values are either positive or zero, so the relative bias value can range from -1 to $+\infty$ especially when the predicted value is much higher than the benchmark value, which could causes confusion during aggregation when few large positive relative bias value offset most negative values. To avoid the confusion, the Normalized bias (NB, fraction) is introduced to normalize the bias from -1 to $+1$. The Root-mean-square error (RMSE, mm/hr for precipitation, and m^3/s for streamflow) measures the distance between the estimates and the observation. The Probability of detection (POD) measures the ability to predict the benchmark flood extent. The False Alarm Rate (FAR) reflects the tendency to overpredict the benchmark flood extent. The Critical Success Index (CSI) measures the overall performance of the flood predictions compared to the benchmark. The Nash-Sutcliffe coefficient of efficiency (NSCE, unitless) measures the effectiveness of the model prediction compared to the stream gauge observation. The peak flow error (PE, m^3/s) calculates the difference between predicted and observed peak flow. The peak time error (PTE, hour) calculates the arriving time difference between predicted and observed peak flow.

3. Results

3.1. Precipitation analysis

To observe the first level differences of all the precipitation forecast products, the accumulated precipitation from 25/08/2017 to 31/08/2017 is plotted in Fig. 2.

The first clear observation appears to be the spatial resolution differences, as RAP and HRRR show coarser resolution grid cells across the map compared to MRMS and the AI nowcasts. Due to the low spatial resolution, RAP can only provide a basic depiction of the precipitation core and has a smooth transition from the high precipitation (southeast) to the lower precipitating amounts (northwest). As listed in Table 1, RAP's highest mean precipitation rate (7.45 mm/hr), the highest mean accumulated precipitation (1081 mm), but low maximum precipitation rate (49.83 mm/hr) means that it does not simulate extremes like MRMS. The second clear observation is the underprediction of the AI nowcasts. The average precipitation rate of AI nowcast during Hurricane Harvey is 3.18 mm/hr and its mean accumulated precipitation is only 190 mm, which is much less than the values of MRMS QPE, which are 6.16 mm/hr and 830 mm, respectively. Google shows the U-Net framework providing well-performed nowcasting precipitation prediction from July 2017 to July 2019 (Agrawal et al., 2019). However, for an extreme event like Hurricane Harvey, the result indicates there are difficulties for machine learning methods to predict the precipitation intensity. Magnitude-wise, HRRR QPF is comparable with MRMS QPE with the mean precipitation rate of 6.65 and 6.16 mm/hr respectively (Table 1). However, the spatial displacement of large precipitation amounts from the HRRR is visually clear, where the red circles (Fig. 2) provide simple examples where MRMS has low accumulated precipitation but the HRRR has a large amount of total precipitation forecast and vice versa. The AI hybrid method shows improved precipitation magnitude from 3.18 mm/hr to 7.10 mm/hr and the mean accumulated precipitation increases from 190 to 930 mm. The CDF-matching method to hybrid machine learning and numerical modeling causes multiple

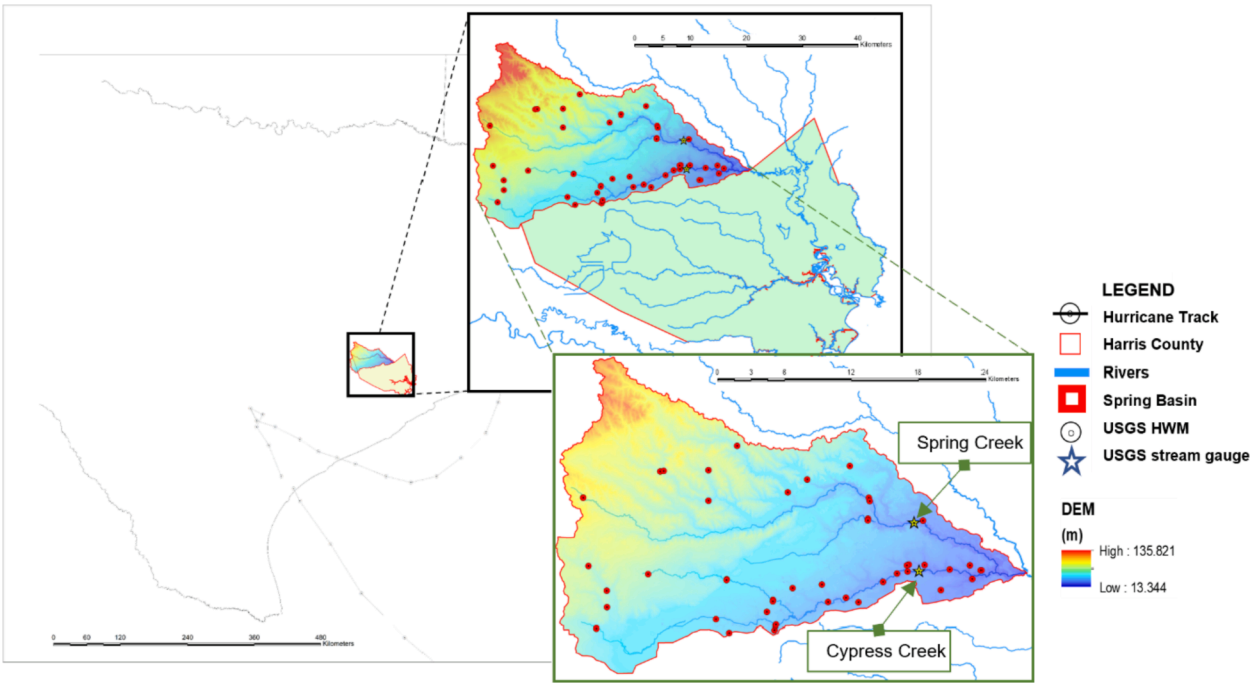

Fig. 1. The Harris County, TX and Spring Basin, showing the track of Hurricane Harvey, WHM locations, USGS gauge locations, and the DEM of Spring Basin.

Table 3
List of statistical metrics used in this study.

Statistic metrics	Equation[a]	Value range	Perfect value
Correlation coefficient (CC)	$CC = \frac{\sum_{n=1}^{N}(f_n-\bar{f})(r_n-\bar{r})}{\sqrt{\sum_{n=1}^{N}(f_n-\bar{f})^2}\sqrt{\sum_{n=1}^{N}(r_n-\bar{r})^2}}$	−1, 1	1
Normalized bias (NB)	$NB = \frac{1}{N}\sum_{n=1}^{N}\frac{f_n-r_n}{f_n+r_n}$	−1, 1	0
Root-mean-square error (RMSE)	$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(f_n-r_n)^2}$	0, +∞	0
Probability of detection (POD)	$POD = \frac{F_1 \wedge R_1}{F_1 \wedge R_1 + F_0 \wedge R_1}$	0,1	1
False alarm ratio (FAR)	$FAR = \frac{F_1 \wedge R_0}{F_1 \wedge R_0 + F_1 \wedge R_0}$	0,1	0
Critical success index (CSI)	$CSI = \frac{F_1 \wedge R_1}{F_1 \wedge R_1 + F_0 \wedge R_1 + F_1 \wedge R_0}$	0,1	1
Nash-Sutcliffe coefficient efficiency (NSCE)	$NSCE = 1 - \frac{\sum_{n=1}^{N}(f_n-r_n)^2}{\sum_{n=1}^{N}(r_n-\bar{r})^2}$	-∞, 1	1
Peak flow error (PE)	$PE = f_{max} - r_{max}$	-∞, +∞	0
Peak time error (PTE)	$PTE = t(r_{max}) - t(f_{max})$	-∞, +∞	0

[a] Variables: n and N, sample index and a total number of samples, f represents the precipitation forecast products from the numerical modeling and AI nowcasts, r represents the reference including the MRMS QPE and USGS stream gauge observations, F and f represent the model simulation results of binary classification and values respectively; R and r represent the reference data of binary classification and values respectively; 1 and 0 means positive (wet) and negative (dry) classifications.

problems: first, the high values of accumulated precipitation appeared to be concentrated at the east and northeast pixels at the edge; second, the horizontal, belt-shaped high-precipitation artifacts appear at the northern part of Harris County.

To quantify the difference between the benchmark precipitation and the precipitation forecasts, the first level statistics are applied to the county-wide averages as well as to each pixel from 25/08/2017 to 31/08/2017. The statistical results are listed in Table 4.

Both county-averaged and pixel-level statistics agree that the AI nowcast precipitation forecast has the highest correlation coefficient (0.98 and 0.8), the large negative normalized bias (-52.84% and –22.96%), and the least RMSE (5.18 mm/hr and 11.87 mm/hr). Among all precipitation forecasts, AI nowcast can capture the temporal pattern very well, despite that there is a large deficit in precipitation magnitude. HRRR QPFs show a very low correlation with benchmark precipitation both at county average (0.55) and pixel-level (0.13). The HRRR QPFs also have the largest RMSE among all other precipitation forecasts despite having the clostest to optimal bias (NB is 8.45% and 1.03% for county-averaged and pixel-level analysis). RAP QPFs have a good correlation with the benchmark when averaged over Harris County (0.76) but poor correlation at each pixel (0.31), which might be caused by its coarse spatial resolution. The AI hybrid method is proved to be less correlated with the benchmark compared to AI nowcasts and has higher RMSE. The attempt to combine AI nowcast and HRRR provides worse statistical results, which can be caused by the artifacts found in Fig. 2. To visualize the differences, the county-averaged precipitation estimates and forecasts are plotted in Fig. 3.

The AI nowcast (green) appears to follow all the peaks as the benchmark (black) and match well when the precipitation is light. However the magnitude of the peaks are only equivalent to a fraction of the benchmark data, and the accumulated rainfall amount indicates a severe underprediction. All other precipitation forecasts (RAP, HRRR, and AI hybrid) overpredict most of the precipitation peaks before 27/08/2017 and after 29/08/2017, while underpredicting during 28/08/2019, which explains the poor statistical results displayed in Table 4. As shown in Fig. 3, the HRRR QPF (blue) and AI hybrid (plum) forecasts are overlapped in most days, which indicates that the CDF-matching method not only increases the magnitude of AI nowcasts but also picks up a lot of information from the HRRR. Since the HRRR has the spatial displacement problem, this property also passes along to AI nowcast causing the reduced predictability.

The pixel-level statistical results are plotted in Fig. 4. As shown in the plot and Table 4, RAP and HRRR have a low temporal correlation with benchmark data throughout the area, and AI nowcasts and AI hybrid predictions have a higher correlation. Deep learning predictions appear to create minor artifacts in multiple radial circle patterns on the temporal correlation with the benchmark data. For bias, the numerical

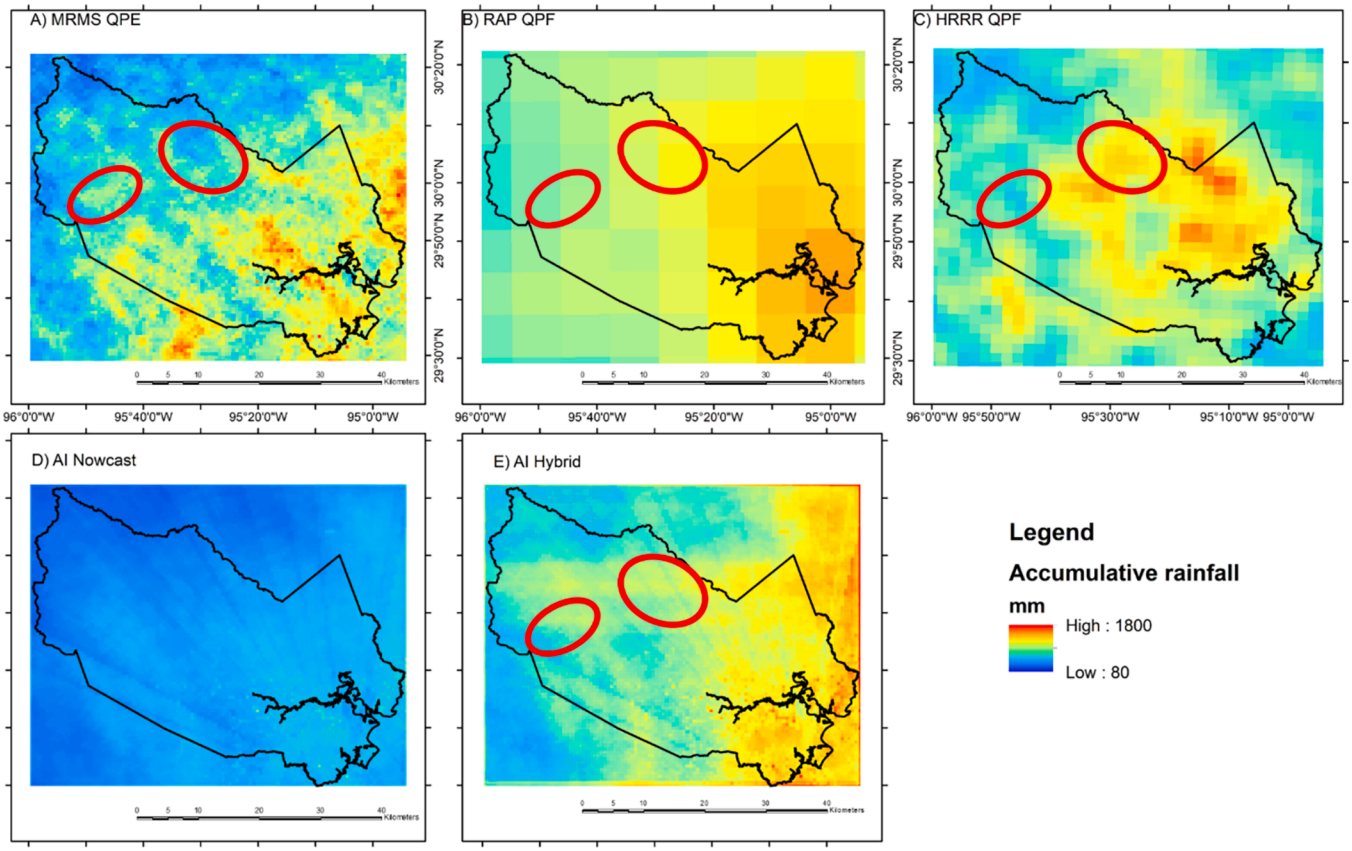

Fig. 2. The accumulative precipitation during Hurricane Harvey of A) MRMS QPE, B) RAP QPF, C) HRRR QPF, D) AI nowcast, and E) AI hybrid, the red ovals circle the two examples of displacements.

Table 4
The first level statistical results between precipitation forecasts and the benchmark precipitation estimation.

Name	County averaged statistic			Pixel statistic		
	CC	NB(%)	RMSE, mm/hr	CC	NB(%)	RMSE, mm/hr
RAP QPF	0.76	24.38	5.41	0.31	34.29	15.72
HRRR QPF	0.55	8.45	7.31	0.13	1.03	20.06
AI nowcast	0.98	−52.84	5.36	0.80	−22.69	11.87
AI hybrid	0.62	17.54	6.51	0.66	7.55	13.16

modeling forecasts (RAP and HRRR) have most of the high bias concentrated at the southeastern corner, which is towards the storm core. This indicates overprediction of precipitation at the highest intensity cores for the numerical modeling forecasts. The artifact of the AI hybrid found in Fig. 2 is magnified in the bias analysis, where the horizontal stripes of high bias values concentrate at the northern half of the map. In this study, RAP has lower RMSE than HRRR, despite the fact that the HRRR is a newer and higher-resolution version of RAP, and the large RMSE values are concentrated towards the storm core.

Based on the statistical analysis, the results indicate that there are differences between the precipitation forecasts and the benchmark, where the RAP QPF is too coarse to represent the details of the storm, the

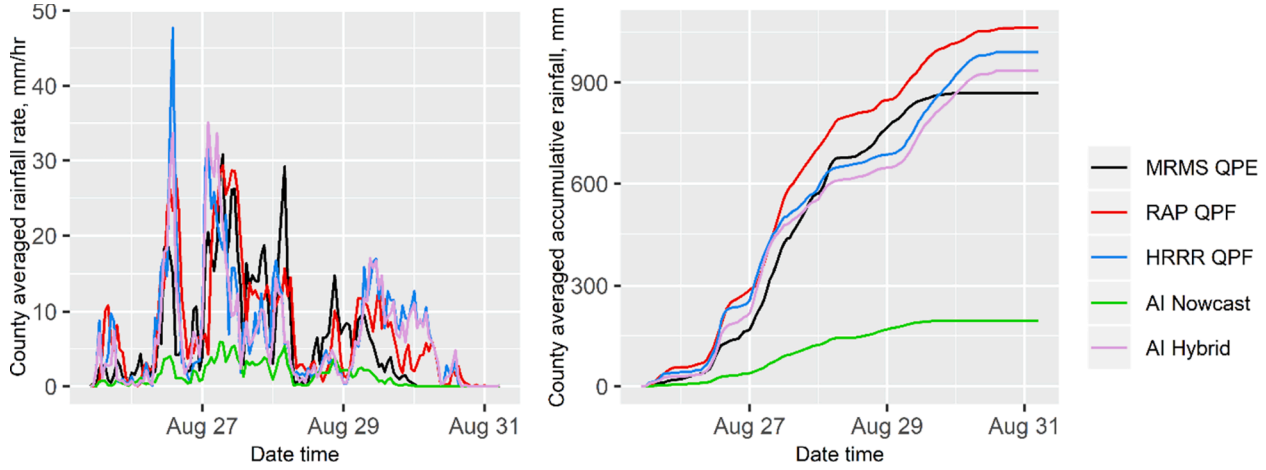

Fig. 3. The county averaged precipitation rate (left) and the county averaged accumulative precipitation (right).

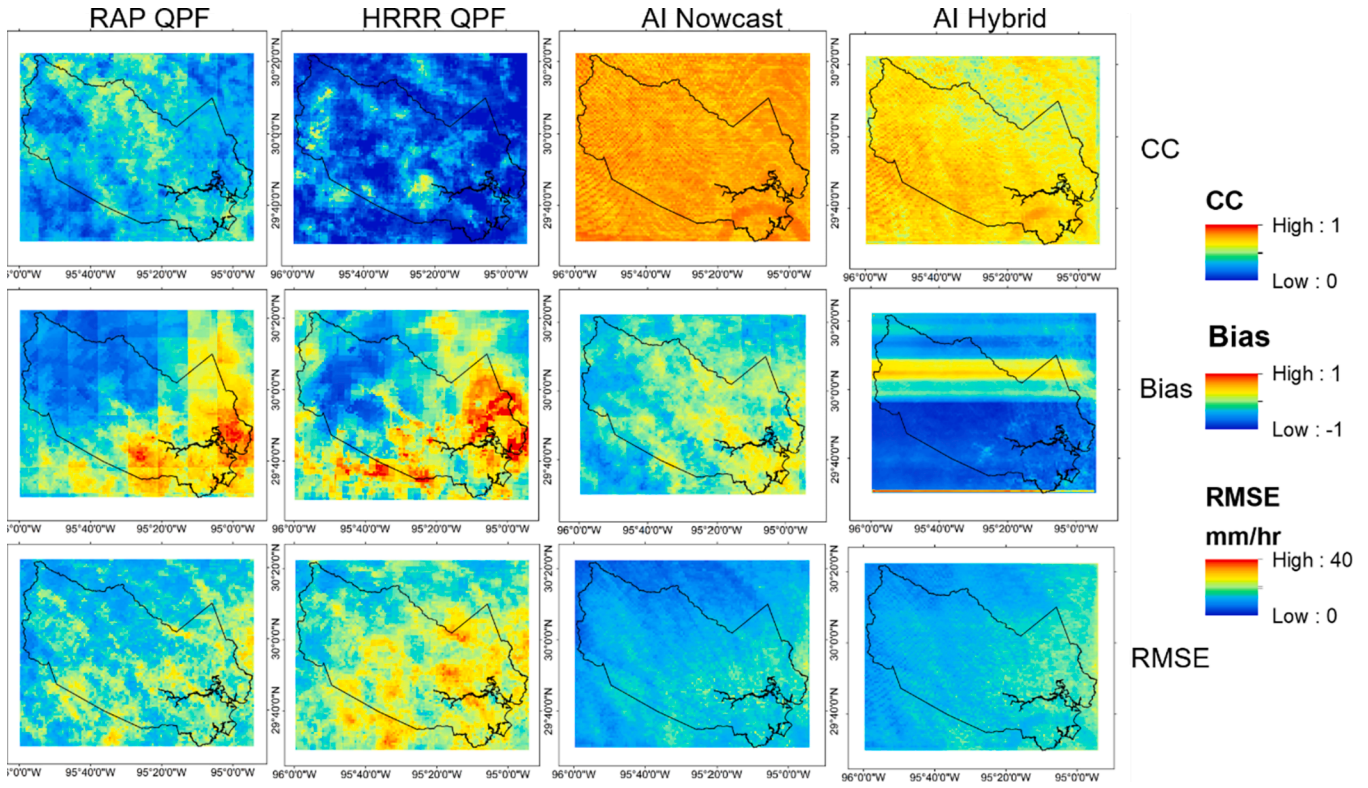

Fig. 4. The statistic results (CC, Bias, and RMSE) at each pixel for RAP QPF, HRRR QPF, AI nowcast, and AI hybrid, compared to the benchmark precipitation (MRMS QPE).

HRRR QPF has displacement error, the AI nowcast fails to predict the precipitation intensity, and the AI hybrid forecast has boundary and artificial noise issue. Yet, a complete agreement between the weather forecast and the weather radar observation is not expected, it would be worth to explore the potentially useful information from the QPF driven flood prediction. As the forecast technology improves in the future, the flood prediction improvement will hence to be examined based on the presented results in section 3.2.

3.2. Hydrological analysis

The CREST-iMAP yields simulated hydrographs using forcing from the MRMS benchmark and all the precipitation forecasts from 15/08/2017 to 03/09/2017 at two USGS gauge locations at Cypress Creek and Spring Creek. Spring Creek is located at the northern part of the Spring basin, which includes an underdeveloped area and did not show too much overbank flow during Hurricane Harvey; while Cypress Creek is located at the southern part of the basin, across multiple developed urban areas, and had obvious overbank flow during Hurricane Harvey (Chen et al., 2020). The simulation results are plotted and listed in Fig. 5 and Table 5.

In Fig. 5, with the exception of the AI nowcasts, all precipitation forecasts generate a second flood peak at both Cypress and Spring creek, which was not observed by USGS stream gauges. As AI nowcast underpredicts the total precipitation amounts and rates, it generates a smaller flood peak consequentially, but it has a high score on NSCE (0.71) and CC (>0.9) compared with the USGS gauge data as shown in Table 4. The AI hybrid provides an undesirable flood forecast with NSCE less than 0.7

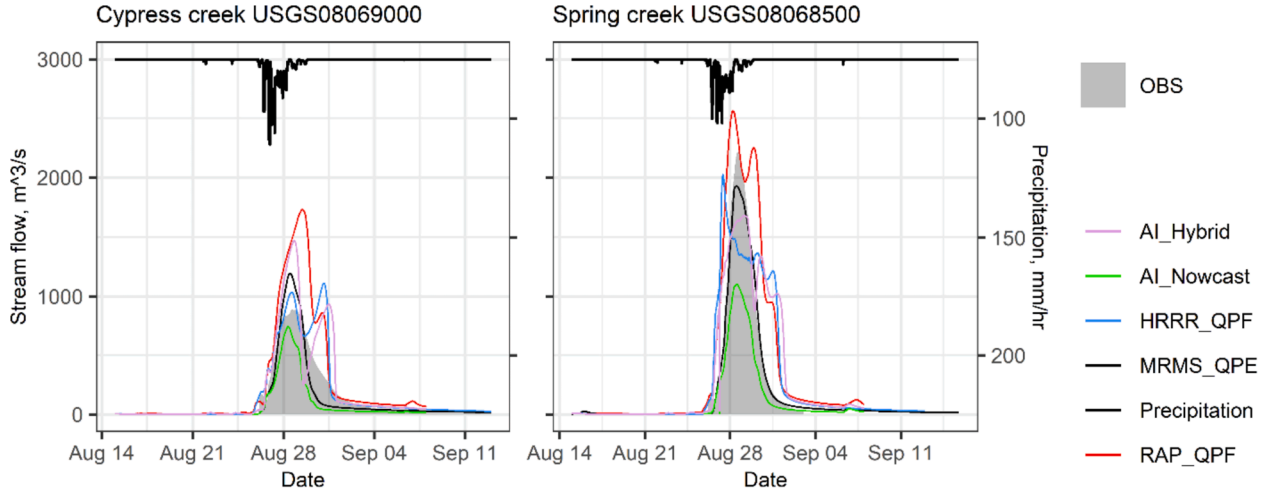

Fig. 5. The flowrate simulation during Hurricane Harvey at the Cypress Creek (southern stream) and Spring Creek (northern stream).

Table 5
The hydrological analysis of precipitation forecasts and benchmark compared to the USGS stream gauge data.

	Cypress Creek 08069000				
	MRMS	RAP	HRRR	AI Nowcast	AI Hybrid
NSCE	0.85	−0.05	0.66	0.71	0.44
NB(%)	4.23	36.96	13.54	−25.54	5.05
CC	0.94	0.97	0.91	0.92	0.90
RMSE (m^3/s)	90	255	135	135	187
Peak Error (m^3/s)	301	834	219	−149	579
Peak Time Error (hour)	−6	16.75	56.5	−9.25	2
	Spring Creek 08068500				
NSCE	0.98	0.50	0.48	0.71	0.67
NB(%)	18.48	52.66	13.08	−8.82	16.65
CC	0.99	0.92	0.77	0.99	0.85
RMSE (m^3/s)	59	588	627	441	367
Peak Error (m^3/s)	−298	340	−198	−1118	−544
Peak Time Error (hour)	−3.25	−11.25	−30.5	−2.75	11.5

and positive PTE, which means the forecasted flood peak is later than the observed flood peak. The AI hybrid (plum line in Fig. 5) carries some features from the HRRR, which causes a non-existent second flood peak. The NWP modeled QPFs produce lower NSCE scores on flood prediction comparing to the deep-learning methods, in general. However, HRRR provides relatively more accurate predictions on flood peak and less flood peak time error, which are significant for the emergency responses and NWS field operations as the flood peak and flood time is the most important factors to mitigate the damage and protect properties. Overall, the precipitation forecasts cannot match the performance of the benchmark as expected, and HRRR QPF provides a relatively better flood magnitude prediction among four precipitation forecasts in this study. The benchmark precipitation data (MRMS) is utilized in the FLASH project as the forcing data for the CREST-EF5 model to produce the near-real-time streamflow (Flamig et al., 2020; Gourley et al., 2017). The FLASH project has shown its benefit as a supporting tool during Hurricane Harvey according to the NWS service assessment report (Murphy, 2018), therefore, with the consideration of the error propagated from the precipitation forecasts, the 1-hour lead-time flood prediction can be a very useful tool. The AI nowcast predicts the flood timing better than others, so if the first responders know the semi-accurate flood time with hours of lead-time, it is a successful step for flood prediction only if the method could improve in the precipitation intensity forecast.

As shown in the hydrological analysis, the current structure of using precipitation forecasts to drive CREST-iMAP has considerable value for the automated real-time flood prediction operation. Even though, no forecast data can provide a comparable flood prediction as MRMS benchmark observation, but the errors from the precipitation forecasts are expected and the prediction results do provide flood information with much longer lead-time. As MRMS is a real-time weather radar precipitation product, it can provide very short to no lead-time for flood prediction, depending on the lag time of the individual flood event. Therefore, the weather forecast based flood prediction has the advantage of lead-time over the benchmark but the disadvantage on accuracy, which can be potentially balanced in flood warning operations.

3.3. *Flood extent analysis*

The CREST-iMAP can not only simulate the streamflow but also the 2D flood extent and flood depth. We take the maximum flood depth at every pixel through all the time steps to construct the maximum flood extent map for all precipitation forecasts and benchmark-forced simulations, which is shown in Fig. 6. The binary statistical results are listed in Table 6.

Although previous statistical and hydrological analysis shows drastic differences between the different precipitation forecasts, the maximum flood extent maps are very similar, and the binary statistics in Table 6

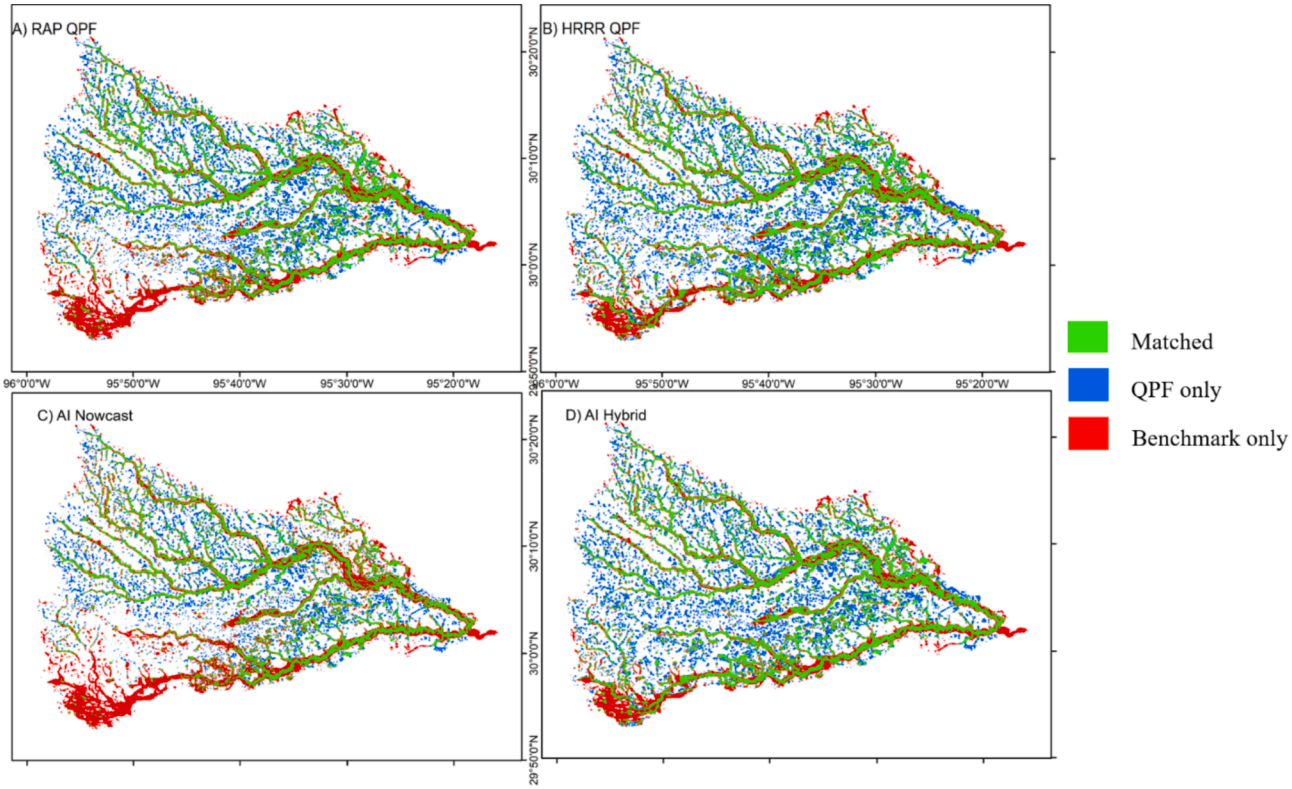

Fig. 6. The 2D flood extents maps display the intersection of A) RAP QPF, B) HRRR QPF, C) AI nowcast, D) AI hybrid predictions with those from the benchmark flood map.

Table 6
The flood extent binary statistics and FEMA flood claim coverage results.

	POD	FAR	CSI
RAP QPF	0.57	0.46	0.38
HRRR QPF	0.62	0.47	0.40
AI Nowcast	0.48	0.43	0.35
AI Hybrid	0.63	0.47	0.40
MRMS	1.00	0.00	1.00

are numerically close, too. The AI nowcast is the only one that has a lower POD (0.48) and a lower CSI (0.35), which is likely due to its underestimation of precipitation amount as shown in Table 1 and Fig. 3. Based on the results in Fig. 6, the CREST-iMAP appeared to neutralize most of the differences between different precipitation forecasts to produce similar maximum flood maps since the maximum flood extent has positive correlation with the total precipitation amount during the flood events (O'Hara et al., 2019). There are no successful predictions for the upper Cypress Creek (southwestern part of the map). All flood predictions other than the AI nowcasts can achieve a CSI of about 0.4 and POD about 0.6, which can still provide limited information and guidelines for future flooding. However, the over 45% false alarming could mislead preparatory action to the flood.

The analysis of the initial inundation (water depth $>$ 5 cm) time is listed in Fig. 7, where the color ramp represents the unified time steps, which is equivalent to 30 mins for easier display due to high temporal resolution of 2D flood simulation (10 s). The dark blue color represents the first day of the study period (25/08/2017 00:00) and the dark red color represents the end of the last day of the study period (29/08/2017 24:00). For example, in the figure, all maps show dark blue color as the center of the main streamlines, which represent the river channels that have water running at the beginning of the period. The bar plot on the right column of Fig. 7 represents the distribution of the flood time difference between the flood prediction and the benchmark flood extent, where the red dashed line is the 0-difference line and the black dashed line represents the mean differences. Negative values indicate the predicted flood comes earlier than the benchmark and the positive value means the predicted flood comes later than the benchmark.

For the benchmark flood map, most of the inundation occurs around 88 $\sim$ 150 time-steps, which are late in the 26th to the morning of the 28th of August 2017. This result corresponds with the county-average precipitation results in Fig. 3, where the benchmark (black line) shows the first peak of precipitation about 20 mm/hr around 1200 UTC on the 26th and two large rainfall peaks (30 mm/hr) on the 27th and early on the 28th of August. The flow dynamic analysis also picks up the overbank flow of the rivers where the color ramp changes from the center outward, from dark blue (had water in channel on 25th) to light green (overbank flooded on 27th) then to dark red (over bank flooded on 29th), which represents the river experiencing multiple overbank flows throughout the study period. The flood prediction from the RAP and HRRR QPF show much more blue (earlier inundation) compared to the benchmark, which corresponds to the time when the RAP (red) and HRRR (blue) have their largest peak precipitation rates (46 mm/hr and 34 mm/hr respectively) on the 26th of August in Fig. 3. These simulated (artifitial) extreme precipitation rates can cause flood inundation

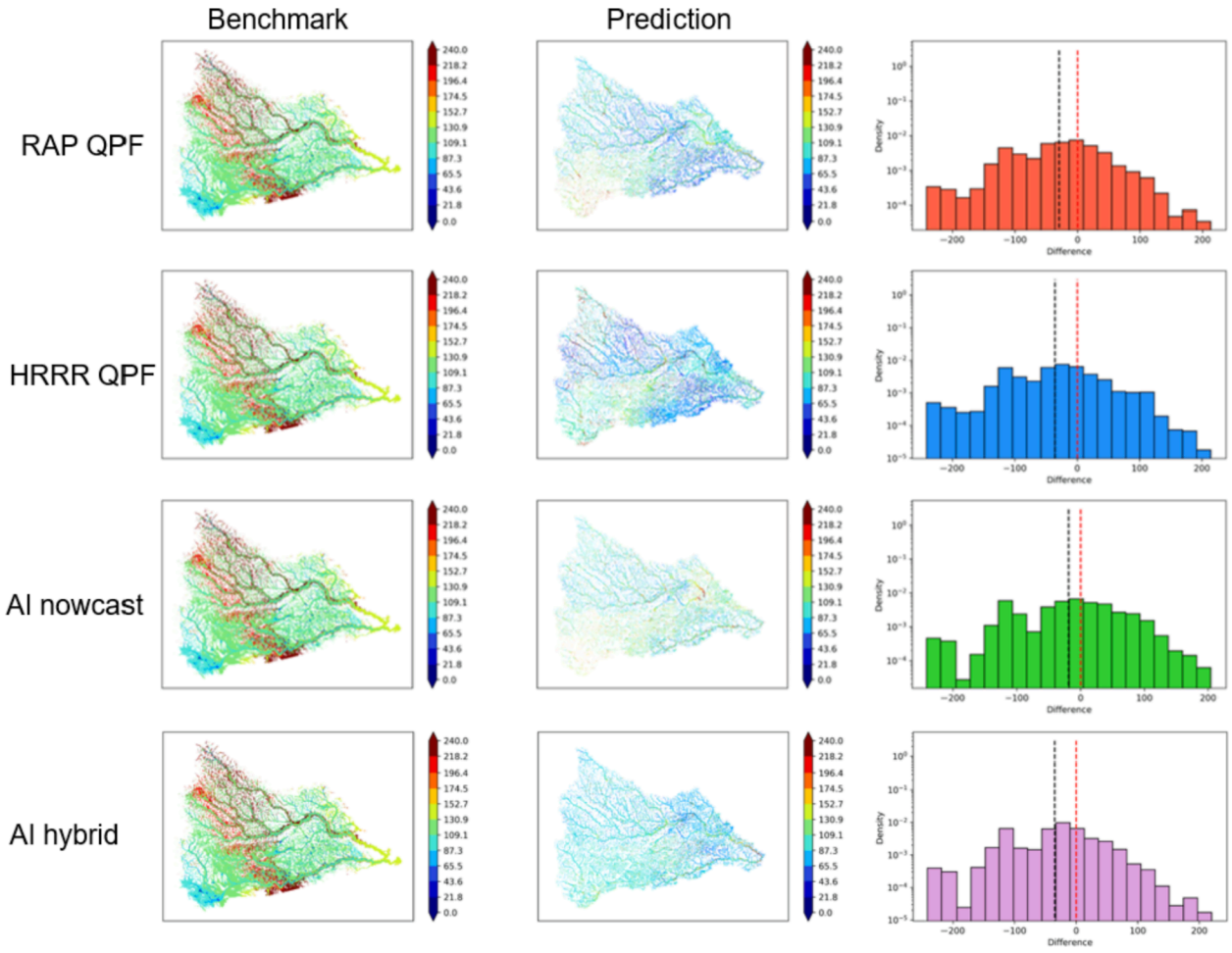

Fig. 7. The half-hourly flood inundation time maps between the benchmark flood map (left) and the predictions (middle), as well as the time differences distribution (right).

immediately, which leads to an earlier inundation than the benchmark. Since the HRRR has higher spatial resolution and greater peak precipitation rates than the RAP presented in Table 1, the initial inundation time map of the HRRR has more late (red) inundation along the river channels than the RAP predicted inundation time, which indicates that there is more overbank flood with reasonable time lags represented in the HRRR than the RAP. Due to the spatial resolution differences, low resolution RAP provides uniform precipitation on river channel, upper and downer portion of a reach segment, as well as the basin catchment. On the other hand, the higher resolution HRRR could provide higher precipitation rate at catchment but lower rate at channel at the same time step as the RAP, but water routing from land to channel could take longer time, so the higher resolution forecast (HRRR) showed more delayed flood than the lower resolution forecast (RAP). A more comprehensive study will be needed to provide explicit understanding of this relationship, and the aforementioned hypothetical scenario is provided to help understanding the possible causes of the flood time differences. The AI nowcast-predicted inundation time map shows more light blue (32 hr) and green (48 hr) compared to other predictions, which is caused by the low precipitation intensity that led the delayed water accumulation on the surface. The AI hybrid prediction shows earlier inundation than the benchmark, which is very similar to the HRRR prediction. Overall, AI hybrid and HRRR predictions have over 70% of wet pixels (74% and 72% respectively) that are flooded earlier than the benchmark, and RAP prediction has 66% of wet pixels flooded earlier than the benchmark. Even with the significant underprediction of precipitation rate from AI nowcast, its flood prediction has about 50% of wet pixels flooded earlier than the benchmark.

The predicted flood extent provides more useful information than the hydrograph predictions. The flood map predicted by the HRRR and RAP QPFs can provide information about the location of possible fluvial and pluvial floods, however, the flood prediction times are not reliable. Machine learning prediction methods show little to no value in the flood extent analysis due the underprediction of the precipitation intensity.

3.4. Flood depth analysis

50 USGS High-Water Mark (HWM) sites are located in the Spring basin (Fig. 1), where the water depth values are extracted from all flood maps including the benchmark and predicted flood maps. In Fig. 8, the scatter plot of the benchmark flood depth and predicted flood depth against the HWM data is presented. As the previous study indicates the error of simulated Harvey flood depth is up to 1 m (Wing et al., 2019), it is no surprise that the benchmark (black dots) cannot align perfectly along the diagonal, and the predicted flood depths only perform worse. As shown in Fig. 8, almost all predicted flood depths are below the diagonal except one location that is predicted by HRRR QPF with CREST-iMAP (blue dots). Since AI nowcast underpredicts the precipitation and associated flood, there are 38 out of 50 HWM sites that are predicted to be not inundated under this scenario. Other above-the-diagonal dots are the ones that HWM data shows 0 but the flood predictions indicate positive values. Overall, all the flood predictions underpredict the inundation depth and provide little value for flooding emergency responses.

The error distributions of inundation depths are plotted in Fig. 9. The plots show that no flood prediction error is as centered to 0 as the benchmark estimation (MRMS). Second to the benchmark, the RAP QPF can lead to a predicted flood depth that is slightly better than HRRR QPFs and the AI hybrid method-predicted flood depths. This result, again, demonstrates the similarity of HRRR QPF and AI hybrid forecasting, where their flood depth prediction errors are almost identical. It raises the speculation that using CDF-matching technique to hybrid machine learning and numerical modeling might pick up too much unnecessary information from the target dataset (HRRR in this study).

In general, using RAP QPF, HRRR QPF, AI nowcast, and AI hybrid forecast can only provide limited information about the upcoming flood inundation depths for an extreme event. However, the flood depth estimation is a common insufficiency of many popular flood models, which has been shown in previous studies (Chen et al., 2021; Wing et al., 2019). Therefore, the error from the precipitation forecasts combines the hydraulic model error that deviate the predicted flood depth further away from the HWM measurements. At the current state of improvement for hydrologic modeling and weather prediction, the overarching goal of higher accuracy on flood depth predictions is yet to be reached.

4. Discussion

The NWP-based QPFs show poor predictability in precipitation

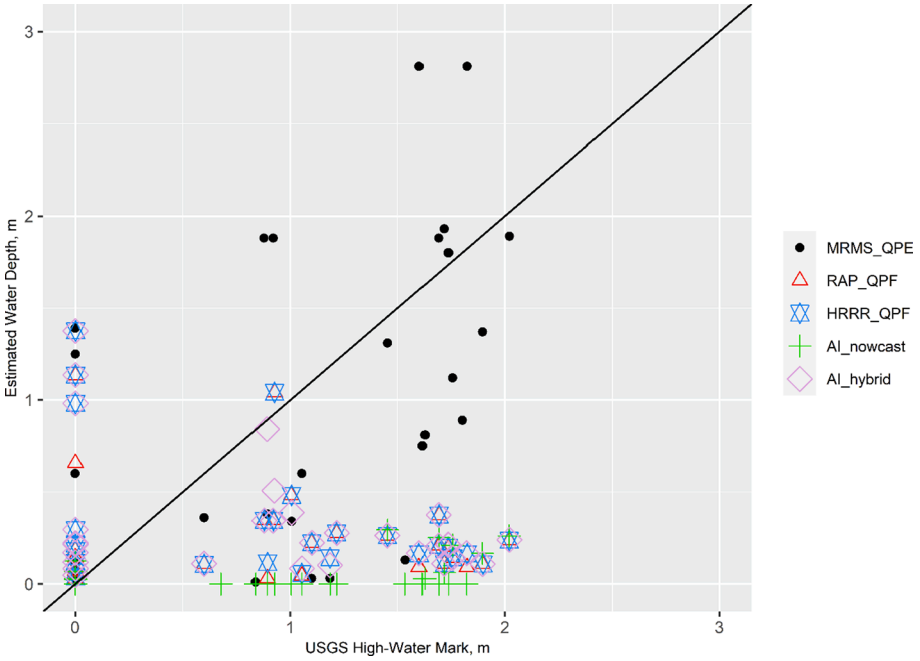

Fig. 8. The scatter plot of flood inundation depths of benchmark flood map and predicted flood map against USGS HWM data.

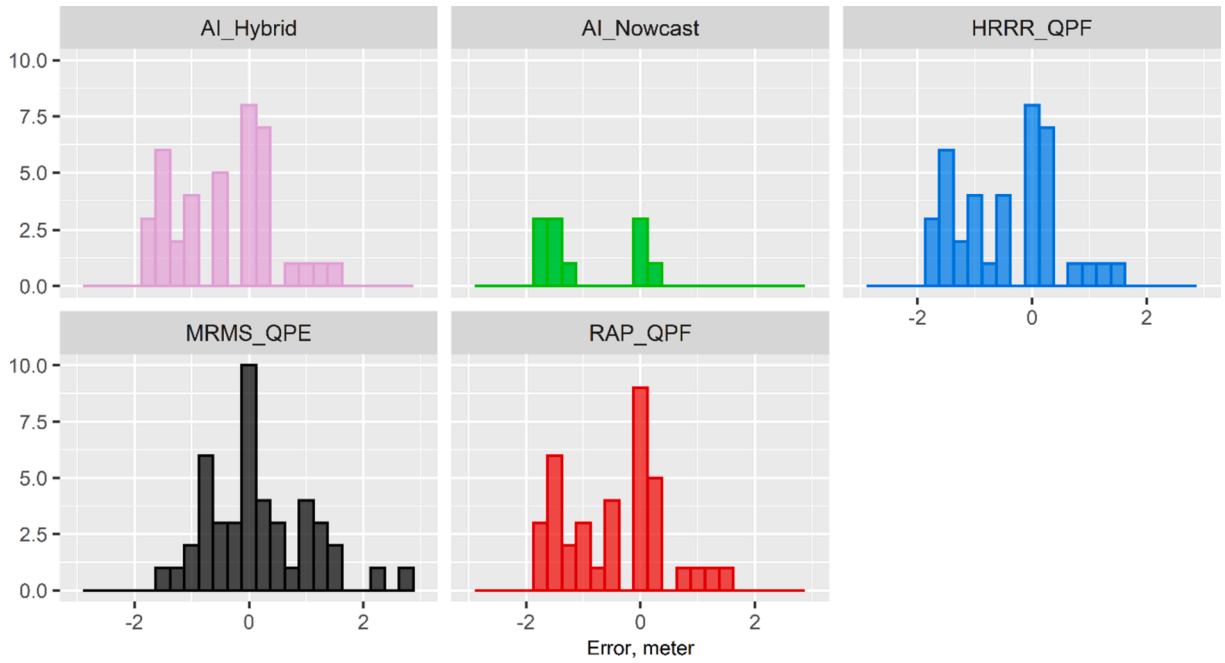

Fig. 9. The flood depth error distribution between USGS HWMs and benchmark flood map (MRMS QPE) and predicted flood maps of RAP QPF, HRRR QPF, AI nowcast, and AI hybrid.

spatial and temporal distribution, even though their predicted total quantity is relatively accurate. One possible cause could be the selection of the forecasted rain fields, which we select the + 1-hour lead time forecasts while RAP and HRRR can forecast up to + 18 h. The study of Seo et al. (2018) suggests that the RAP and HRRR QPFs perform better with + 4 ~ +6 h lead time, indicating that the + 1-hour forecasts are not necessarily the most reliable QPF product. Another recent study (Yue and Gebremichael, 2020) indicated that all the HRRR forecast including + 1 hr lead-time has strong agreement with the observation, while slightly over-prediction the magnitude for a half of the hurricane event. However, due to the limited computational resources, the requirement of hyperresolution precipitation for hydrological forecast, and the heavy GPU memory requirements, the AI methods will not likely forecast>3 h of lead time (only limited to 1-hour lead time in this study) for any machine at current time. Therefore, certain amount of error from numerical modeling forecasts are expected, and the predictability study over different lead-times will be conducted in the future. However, the well-known QPF issues still exist, such as the underforecasting the convective system rainfall (RAP) and spatial displacement errors (HRRR), which are likely to present in the other forecast periods. The HRRR Ensemble (HRRRE) project was proven to mitigate those errors (Carlberg et al., 2020), which will lead to a better flood forecast once the HRRRE being an operational product.

The AI nowcast method lacks the ability to forecast the extreme precipitation event, such as Hurricane Harvey. The AI model is trained by all precipitation events in the Houston area from 2015 to 2019 that are listed in NOAA storm report and have caused urban flooding, except Hurricane Harvey, which means it is trained specifically for high-intensity precipitation scenarios. But the results indicate the method cannot reproduce the precipitation intensity of Hurricane Harvey and provides no useful information for local flood prediction. Based on the study done by Google Research, the U-Net machine learning method outperforms HRRR QPF and the Optical Flow method using precision-recall curve analysis (Agrawal et al., 2019). This statement still holds in this case study, however, as a binary analysis conducted in the aforementioned study, the lack of predicted precipitation intensity is not revealed in the precision-recall curve analysis. Although the precision-recall curve is a common and powerful scoring tool for machine learning, this test is not sufficient for hydrological and hydrometeorological studies, since the precipitation intensity is as important, if not more, as its spatial–temporal placement. Therefore, deep learning methods for hydrological prediction still require much effort from the scientific community to improve the technology.

As a preliminary attempt to combine the results of AI nowcasts and numerical model outputs, the CDF-matching technique is used in this study to produce the AI hybrid forecasts, but failed to perform well in this case study. It has been shown in the previous section that the technique creates artifacts and border effects, but it also shows that the technique alters the spatial–temporal distribution of the AI nowcast towards the HRRR, which leads to the similarities between the HRRR and the AI hybrid in section 3.3 and 3.4. Since the AI nowcast is able to capture the spatial–temporal pattern of the MRMS estimation, improving the precipitation intensity forecast but not changing its pattern could be a future direction for deep-learning nowcasting approaches. Given that the AI nowcast model could produce high-resolution (3 km, 15 mins) forecast within minutes for the 10 days period for 276 by 216 cells using a single GPU unit, this method still has the advantage on computational efficiency, compared to HRRRE experiment that has used 144 parallel nodes to power WRF-ARW to simulate 1580 by 1000 meshes over the full CONUS (Dowell et al., 2018). With improvement on its forecast performance, AI nowcasting method could potentially be a computationally efficient forecast method in the future.

Lastly, the maximum flood map analysis using CREST-iMAP neutralizes or smooths much of the differences between precipitation forecasts as the total accumulated precipitation dictates the maximum flood extent. The results indicate the QPF with a higher spatial-resolution would potentially improve the flood timing forecast, yet, a more comprehensive study is needed to examine the importance and impacts of spatial resolution to flood extent forecast. Since the RAP and HRRR QPFs have a reasonable forecast of total precipitation amount over the region, the flood extent analysis results for numerical modeling are acceptable and provide indications for flood forecasts. Therefore, even with certain missing inundation scenarios, the QPF + CREST-iMAP could still potentially provide preliminary information about the final outcome of a future flood event.

5. Conclusion

The ability of precipitation forecasts to predict flood discharges, inundation extents, and flood depth has been tested using numerically modeled QPFs (RAP and HRRR) and deep learning nowcasts (AI nowcast and AI hybrid) with + 1-hour lead time. None of the precipitation forecasts can provide comparable flood information as the radar-based benchmark (MRMS), which neither the NWP models nor any nowcasting method is expected to be as accurate as weather radar observation. The 'benchmark' in this study is an ultimate and overarching goal for the weather forecasting community. When the real forecasting is needed for the future and the observation is not available, the numerical weather prediction model QPFs can provide general information about flood return period and inundation outcomes, where the HRRR slightly outperforms the RAP. The AI nowcast is incapable to capture the precipitation intensity of Hurricane Harvey, which indicates the potential inability of such method, as well as the inability of common machine learning performance tests to reveal such information.

The AI nowcast method can forecast the spatial–temporal pattern in high-resolution and extent of the precipitation in a very short amount of time; it will be the future objective to combine the AI nowcasts and QPFs using different methods, as the CDF-matching technique fails to improve the performance of the hybrid approach evaluated in this study. The NWP QPFs have spatial displacement errors underlying most of the QPF total error, but the ensembled weather forecasting products (e.g. HRRR-Ensemble) should reduce such errors by theory, which should be further studied to provide improved flood severity prediction. Further, flood predictability analysis with longer lead-time is needed to study the possible best combination of QPF + CREST-iMAP to provide the best possible flood forecast for local emergency response agencies.

CRediT authorship contribution statement

Mengye Chen: Conceptualization, Methodology, Software, Visualization. **Zhi Li:** Software. **Shang Gao:** Writing – review & editing. **Ming Xue:** Supervision. **Jonathan J. Gourley:** Writing – review & editing. **Randall L. Kolar:** Supervision. **Yang Hong:** Conceptualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The authors would like to thank the support from Oklahoma University Hydrology and Water Security program.

References

Adams, T.E., Pagano, T.C. (Eds.), 2016. Flood forecasting: a global perspective. Elsevier/AP, Academic Press is an imprint of Elsevier, Amsterdam Boston.

Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C., Hickey, J., 2019. Machine Learning for Precipitation Nowcasting from Radar Images. arXiv:1912.12132 [cs, stat].

Ashley, S.T., Ashley, W.S., 2008. Flood Fatalities in the United States. J. Appl. Meteorol. Climatol. 47, 805–818. https://doi.org/10.1175/2007JAMC1611.1.

Barredo, J.I., 2007. Major flood disasters in Europe: 1950–2005. Nat. Hazards 42, 125–148. https://doi.org/10.1007/s11069-006-9065-2.

Bates, P.D., De Roo, A.P.J., 2000. A simple raster-based model for flood inundation simulation. J. Hydrol. 236, 54–77. https://doi.org/10.1016/S0022-1694(00)00278-X.

Bates, P.D., Horritt, M.S., Fewtrell, T.J., 2010. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. J. Hydrol. 387, 33–45. https://doi.org/10.1016/j.jhydrol.2010.03.027.

Bedient, P.B., Huber, W.C., Vieux, B.E., 1988. Hydrology and floodplain analysis, 5th ed. Prentice Hall, Upper Saddle River, NJ.

Benjamin, S.G., Devenyi, D., Weygandt, S.S., Brundage, K.J., Brown, J.M., Grell, G.A., Kim, D., Schwartz, B.E., Smirnova, T.G., Smith, T.L., 2004. An hourly assimilation–forecast cycle: The RUC. Mon. Weather Rev. 132, 495–518. https://doi.org/10.1175/1520-0493(2004)132,0495:AHACTR.2.0.CO;2.

Benjamin, S.G., Weygandt, S.S., Brown, J.M., Hu, M., Alexander, C.R., Smirnova, T.G., Olson, J.B., James, E.P., Dowell, D.C., Grell, G.A., Lin, H., Peckham, S.E., Smith, T.L., Moninger, W.R., Kenyon, J.S., Manikin, G.S., 2016. A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. Mon. Weather Rev. 144, 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1.

Brauer, N.S., Basara, J.B., Homeyer, C.R., McFarquhar, G.M., Kirstetter, P.E., 2020. Quantifying Precipitation Efficiency and Drivers of Excessive Precipitation in Post-Landfall Hurricane Harvey. J. Hydrometeorol. 21, 433–452. https://doi.org/10.1175/JHM-D-19-0192.1.

Carlberg, B., Franz, K., Gallus, W., 2020. A Method to Account for QPF Spatial Displacement Errors in Short-Term Ensemble Streamflow Forecasting. Water 12, 3505. https://doi.org/10.3390/w12123505.

Chen, M., Nabih, S., Brauer, N.S., Gao, S., Gourley, J.J., Hong, Z., Kolar, R.L., Hong, Y., 2020. Can Remote Sensing Technologies Capture the Extreme Precipitation Event and Its Cascading Hydrological Response? A Case Study of Hurricane Harvey Using EF5 Modeling Framework. Remote Sensing 12, 445. https://doi.org/10.3390/rs12030445.

Chen, M., Li, Z., Gao, S., Luo, X., Wing, O.E.J., Shen, X., Gourley, J.J., Kolar, R.L., Hong, Y., 2021. A comprehensive flood inundation mapping for Hurricane Harvey using an integrated hydrological and hydraulic model. J. Hydrometeorol. https://doi.org/10.1175/JHM-D-20-0218.1.

Clark, R.A., Flamig, Z.L., Vergara, H., Hong, Y., Gourley, J.J., Mandl, D.J., Frye, S., Handy, M., Patterson, M., 2017. Hydrological Modeling and Capacity Building in the Republic of Namibia. Bull. Am. Meteorol. Soc. 98, 1697–1715. https://doi.org/10.1175/BAMS-D-15-00130.1.

Cohen, S., Praskievicz, S., Maidment, D.R., 2018. Featured Collection Introduction: National Water Model. J. Am. Water Resour. Assoc. 54, 767–769. https://doi.org/10.1111/1752-1688.12664.

Cuo, L., Pagano, T.C., Wang, Q.J., 2011. A Review of Quantitative Precipitation Forecasts and Their Use in Short- to Medium-Range Streamflow Forecasting. J. Hydrometeorol. 12, 713–728. https://doi.org/10.1175/2011JHM1347.1.

de Almeida, G.A.M., Bates, P., 2013. Applicability of the local inertial approximation of the shallow water equations to flood modeling: applicability local inertial. Water Resour. Res. 49, 4833–4844. https://doi.org/10.1002/wrcr.20366.

Dowell, D.C., Alexander, C.R., Alcott, T., Ladwig, T., 2018. HRRR Ensemble (HRRRE) Guidance 2018 HWT Spring Experiment. (Guide). NOAA.

Ebert, E.E., McBride, J.L., 2000. Verification of precipitation in weather systems: determination of systematic errors. J. Hydrol. 239, 179–202. https://doi.org/10.1016/S0022-1694(00)00343-7.

Feaster, T.D., Koenig, T.A., 2017. Field manual for identifying and preserving High-Water Mark data (Open-File Report No. 67). U.S. Geological Survey, Reston, Virginia.

Flamig, Z.L., Vergara, H., Gourley, J.J., 2020. The Ensemble Framework For Flash Flood Forecasting (EF5) v1.2: Description and Case Study (preprint). Hydrology. https://doi.org/10.5194/gmd-2020-46.

Franch, G., Maggio, V., Coviello, L., Pendesini, M., Jurman, G., Furlanello, C., 2020. TAASRAD19, a high-resolution weather radar reflectivity dataset for precipitation nowcasting. Sci. Data 7, 234. https://doi.org/10.1038/s41597-020-0574-8.

Georgakakos, K.P., Hudlow, M.D., 1984. Quantitative Precipitation Forecast Techniques for Use in Hydrologic Forecasting. Bull. Am. Meteorol. Soc. 65, 1186–1200. https://doi.org/10.1175/1520-0477(1984)065<1186:QPFTFU>2.0.CO;2.

Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., Stoffel, M., 2014. 21st century climate change in the European Alps—A review. Sci. Total Environ. 493, 1138–1151. https://doi.org/10.1016/j.scitotenv.2013.07.050.

Gochis, D.J., Dugger, A.L., Yu, W., Yates, D.N., Sampson, K., Barlage, M., Pan, L., Zhang, Y., McCreight, J.L., RafieeiNasab, A., Karsten, L., Read, L., Gaydos, A., McAllister, M., Mills, J., Towler, E., Grim, J., FitzGerald, K., 2017. The NOAA National Water Model: Research to Operations to Research (Presentation). CUAHSI, Boulder CO.

Golding, B.W., 2000. Quantitative precipitation forecasting in the UK. J. Hydrol. 239, 286–305. https://doi.org/10.1016/S0022-1694(00)00354-1.

Gourley, J.J., Flamig, Z.L., Vergara, H., Kirstetter, P.-E., Clark, R.A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J.M., Hong, Y., Howard, K.W., 2017. The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States. Bull. Am. Meteorol. Soc. 98, 361–372. https://doi.org/10.1175/BAMS-D-15-00247.1.

Hapuarachchi, H.A.P., Wang, Q.J., Pagano, T.C., 2011. A review of advances in flash flood forecasting. Hydrol. Process. 25, 2771–2784. https://doi.org/10.1002/hyp.8040.

Kain, J.S., Xue, M., Coniglio, M.C., Weiss, S.J., Kong, F., Jensen, T.L., Brown, B.G., Gao, J., Brewster, K., Thomas, K.W., Wang, Y., Schwartz, C.S., Levit, J.J., 2010. Assessing Advances in the Assimilation of Radar Data and Other Mesoscale Observations within a Collaborative Forecasting-Research Environment. Weather Forecasting 25, 1510–1521. https://doi.org/10.1175/2010WAF2222405.1.

Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. arXiv: 1412.6980 [cs].

Kumar, A., Islam, T., Sekimoto, Y., Mattmann, C., Wilson, B., Stoean, R., 2020. Convcast: An embedded convolutional LSTM based architecture for precipitation nowcasting using satellite data. PLoS One 15 (3), e0230114.
Benito, G., Lang, M., Mariano Barriendos, M.C.L., Frances, F., Ouarda, T., Thorndycraft, V.R., 2004. Use of systematic, palaeoflood and historical data for the improvement of flood risk estimation. Review of scientific methods. Nat. Hazards 31 (3), 623–643.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.
Lee, T.R., Buban, M., Turner, D.D., Meyers, T.P., Baker, C.B., 2019. Evaluation of the High-Resolution Rapid Refresh (HRRR) Model Using Near-Surface Meteorological and Flux Observations from Northern Alabama. Weather Forecasting 34, 635–663. https://doi.org/10.1175/WAF-D-18-0184.1.
Li, Z., Chen, M., Gao, S., Hong, Z., Tang, G., Wen, Y., Gourley, J.J., Hong, Y., 2020. Cross-Examination of Similarity, Difference and Deficiency of Gauge, Radar and Satellite Precipitation Measuring Uncertainties for Extreme Events Using Conventional Metrics and Multiplicative Triple Collocation. Remote Sensing 12, 1258. https://doi.org/10.3390/rs12081258.
Li, Z., Wen, Y., Schreier, M., Behrangi, A., Hong, Y., Lambrigtsen, B., 2021. Advancing Satellite Precipitation Retrievals with Data Driven Approaches: Is black box model explainable? Earth Space Sci. 8 (2) https://doi.org/10.1029/2020EA001423.
Ligda, M.G.H., 1953. Horizontal Motion of Small Precipitation Areas as Observed by Radar. Massachusetts Institute of Technology. PhD Thesis.
Liu, Z., Merwade, V., Jafarzadegan, K., 2019. Investigating the role of model structure and surface roughness in generating flood inundation extents using one- and two-dimensional hydraulic models. J. Flood Risk Manage. 12 (1), e12347.
Loken, E.D., Clark, A.J., Xue, M., Kong, F., 2017. Comparison of Next-Day Probabilistic Severe Weather Forecasts from Coarse- and Fine-Resolution CAMs and a Convection-Allowing Ensemble. Weather Forecasting 32, 1403–1421. https://doi.org/10.1175/WAF-D-16-0200.1.
McCuen, R.H., 2005. Hydrologic analysis and design, 3rd ed. Pearson Prentice Hall, Upper Saddle River, N.J.
Meehl, G.A., Zwiers, F., Evans, J., Knutson, T., Mearns, L., Whetton, P., 2000. Trends in Extreme Weather and Climate Events: Issues Related to Modeling Extremes in Projections of Future Climate Change. Bull. Am. Meteorol. Soc. 81, 427–436. https://doi.org/10.1175/1520-0477(2000)081<0427:TIEWAC>2.3.CO;2.
Murphy, J.D., 2018. Service assessment August–September 2017 Hurricane Harvey (US DOC). NOAA National Weather Service, Silver Spring, Maryland.
Newell, J.E., Deaver, D.G., 1981. The LFM-II model–1980 (Technical memo No. 66). NOAA, Springfield, VA.
O'Hara, R., Green, S., McCarthy, T., 2019. The agricultural impact of the 2015–2016 floods in Ireland as mapped through Sentinel 1 satellite imagery. Irish J. Agric. Food Res. 58, 44–65. https://doi.org/10.2478/ijafr-2019-0006.
Sadeghi, M., Nguyen, P., Hsu, K., Sorooshian, S., 2020. Improving near real-time precipitation estimation using a U-Net convolutional neural network and geographical information. Environ. Modell. Software 134, 104856. https://doi.org/10.1016/j.envsoft.2020.104856.
Sampson, C.C., Smith, A.M., Bates, P.D., Neal, J.C., Alfieri, L., Freer, J.E., 2015. A high-resolution global flood hazard model: a high-resolution global flood hazard modeL. Water Resour. Res. 51, 7358–7381. https://doi.org/10.1002/2015WR016954.
Seo, B.-C., Quintero, F., Krajewski, W.F., 2018. 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.
Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., Woo, W., 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. arXiv: 1506.04214 [cs].
Smith, K., Ward, R., 1998. Floods: Physical processes and human impact. John Wiley, Chichester.
Sønderby, C.K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Agrawal, S., Hickey, J., Kalchbrenner, N., 2020. MetNet: A Neural Weather Model for Precipitation Forecasting. arXiv:2003.12140 [physics, stat].
Srinivasan, R., Arnold, J.G., 1994. Integration of a basin-scale water quality model with GIS. J. Am. Water Resourc. Assoc. 30, 453–462. https://doi.org/10.1111/j.1752-1688.1994.tb03304.x.
Sun, J., Xue, M., Wilson, J.W., Zawadzki, I., Ballard, S.P., Onvlee-Hooimeyer, J., Joe, P., Barker, D.M., Li, P.-W., Golding, B., Xu, M., Pinto, J., 2014. Use of NWP for Nowcasting Convective Precipitation: Recent Progress and Challenges. Bull. Am. Meteorol. Soc. 95, 409–426. https://doi.org/10.1175/BAMS-D-11-00263.1.
Trenberth, K.E. (Ed.), 2010. Climate system modeling. Cambridge University Press, Cambridge.
USGCRP, 2017. Climate Science Special Report: Fourth National Climate Assessment, Volume I. U.S. Global Change Research Program, Washington, DC, USA. https://doi.org/10.7930/J0J964J6.
van Oldenborgh, G.J., van der Wiel, K., Sebastian, A., Singh, R., Arrighi, J., Otto, F., Haustein, K., Li, S., Vecchi, G., Cullen, H., 2018. Corrigendum: Attribution of extreme rainfall from Hurricane Harvey, August 2017 (2017 *Environ. Res. Lett.* **12** 124009). Environ. Res. Lett. 13, 019501. https://doi.org/10.1088/1748-9326/aaa343.
Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. Environ. Modell. Software 75, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013.
Vrugt, J.A., Braak, C.J.F., Gupta, H.V., Robinson, B.A., 2009. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stoch. Env. Res. Risk Assess. 23, 1101–11026. https://doi.org/10.1007/s00477-008-0274-y.
Wang, J., Hong, Y., Li, L., Gourley, J.J., Khan, S.I., Yilmaz, K.K., Adler, R.F., Policelli, F. S., Habib, S., Irwn, D., Limaye, A.S., Korme, T., Okello, L., 2011. The coupled routing and excess storage (CREST) distributed hydrological model. Hydrol. Sci. J. 56, 84–98. https://doi.org/10.1080/02626667.2010.543087.
Werner, M., Cranston, M., 2009. Understanding the Value of Radar Rainfall Nowcasts in Flood Forecasting and Warning in Flashy Catchments. Met. Apps 16, 41–55. https://doi.org/10.1002/met.125.
Wing, O.E.J., Bates, P.D., Sampson, C.C., Smith, A.M., Johnson, K.A., Erickson, T.A., 2017. Validation of a 30 m resolution flood hazard model of the conterminous United States: 30 m resolution flood model of CONUS. Water Resour. Res. 53, 7968–7986. https://doi.org/10.1002/2017WR020917.
Wing, O.E.J., Sampson, C.C., Bates, P.D., Quinn, N., Smith, A.M., Neal, J.C., 2019. A flood inundation forecast of Hurricane Harvey using a continental-scale 2D hydrodynamic model. J. Hydrol. X 4, 100039. https://doi.org/10.1016/j.hydroa.2019.100039.
Wood, E.F., Lettenmaier, D.P., Zartarian, V.G., 1992. A land-surface hydrology parameterization with subgrid variability for general circulation models. J. Geophys. Res. 97, 2717. https://doi.org/10.1029/91JD01786.
Xue, M., Kong, F., Thomas, K.W., Gao, J., Wang, Y., Brewster, K., Droegemeier, K.K., 2013. Prediction of Convective Storms at Convection-Resolving 1 km Resolution over Continental United States with Radar Data Assimilation: An Example Case of 26 May 2008 and Precipitation Forecasts from Spring 2009. Adv. Meteorol. 2013, 1–9. https://doi.org/10.1155/2013/259052.
Xue, X., Zhang, K., Hong, Y., Gourley, J.J., Kellogg, W., McPherson, R.A., Wan, Z., Austin, B.N., 2016. New multisite cascading calibration approach for hydrological models: case study in the Red River Basin using the VIC model. J. Hydrol. Eng. 21 https://doi.org/10.1061/(ASCE)HE.1943-5584.0001282.
Yue, H., Gebremichael, M., 2020. Evaluation of high-resolution rapid refresh (HRRR) forecasts for extreme precipitation. Environ. Res. Commun. 2 (6), 065004.