JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

Vol. 58, No. 2 AMERICAN WATER RESOURCES ASSOCIATION April 2022

Application of a Large-Scale Terrain-Analysis-Based Flood Mapping System to Hurricane Harvey

Xing Zheng, Claudia D'Angelo, David R. Maidment, and Paola Passalacqua

Research Impact Statement: We compare inundation estimates with high-water marks collected during Hurricane Harvey. Our system estimates depth with a 0.5-m mean error and extent covering 90% of that obtained from observations.

ABSTRACT: Flood modeling provides inundation estimates and improves disaster preparedness and response. Recent development in hydrologic modeling and inundation mapping enables the creation of such estimates in near real time. To quantify their performance, these estimates need to be compared to measurements collected during historical events. We present an application of a flood mapping system based on the National Water Model and the Height Above Nearest Drainage method to Hurricane Harvey. The outputs are validated with high-water marks collected to record the highest water levels during the flood. We use these points to compute elevation-related variables and flood extents and measure the quality of the estimates. To improve the performance of the method, we calibrate the roughness coefficient based on stream order. We also use lidar data with a workflow named GeoFlood and we compare the modeled inundation to that recorded by the high-water marks and to the maximum inundation extent provided by the Dartmouth Flood Observatory based on remotely sensed data from multiple sources. The results show that our mapping system estimates local water depth with a mean error of about 0.5 m and that the inundation extent covers over 90% of that derived from high-water marks. Using a calibrated roughness coefficient and lidar data reduces the mean error in flood depth but does not affect as much the inundation extent estimation.

(KEYWORDS: large-scale flood modeling; high-water marks; flood inundation mapping; lidar; HAND (Height Above Nearest Drainage).)

INTRODUCTION

The United States (U.S.) National Weather Service has developed the National Water Model (NWM) which continuously forecasts discharge throughout the nation's stream and river network. The headquarters of this effort is the National Water Center (NWC), located in Tuscaloosa, Alabama which recently issued a Handbook of NWC Visualization Services (National Water Center 2020). Each service is a map depicting some aspect of the forecast over the continental United States (U.S.) and in some cases, Hawaii. Three main categories of forecast products are included: those for current conditions, for a short-range forecast 18 h ahead, and for a medium range forecast 10 days ahead. The river flow forecasts on the main stem rivers are derived from regional models operated by the 12 regional river forecast centers operated by the National Weather Service. At

Paper No. JAWR-20-0095-P of the *Journal of the American Water Resources Association* (JAWR). Received July 26, 2020; accepted January 10, 2022. **Discussions are open until six months from issue publication**.

Department of Civil, Architectural and Environmental Engineering and Center for Water and the Environment (Zheng, Maidment, Passalacqua), University of Texas at Austin Austin, Texas, USA; and Department of Engineering (D'Angelo), Roma Tre University Rome, ITA (Correspondence to Passalacqua: paola@austin.utexas.edu).

Citation: Zheng, X., C. D'Angelo, D.R. Maidment, and P. Passalacqua. 2022. "Application of a Large-Scale Terrain-Analysis-Based Flood Mapping System to Hurricane Harvey." *Journal of the American Water Resources Association* 58 (2): 149–163. https://doi.org/10.1111/1752-1688.12987.

the NWC, these forecasts are overlaid on those arising from the NWM whose data define flows in the rest of the river and stream network. It is remarkable that the National Weather Service is now producing river forecast services as maps across the river and stream network, in addition to the traditional forecast hydrographs at particular points on the main stem rivers.

Included in the NWC Visualization Services are inundation mapping services developed using the Height Above Nearest Drainage (HAND) method, for which the NWC Handbook cites the work of Zheng et al. (2018) and Liu et al. (2018). Zheng et al. (2018) defined the methodology of producing inundation maps and synthetic rating curves to relate the discharge forecast to water depth above the river channel thalweg, and thus to inundation extent. Liu et al. (2018) showed how the mapping and rating curves could be developed using supercomputing throughout the continental U.S. The inundation mapping services are at present being calculated at the NWC only for the West Gulf and Northeast River Forecast Centers, with the intention to include other regions later.

The HAND-based inundation mapping approach has been implemented and improved in different studies. Shastry et al. (2019) added a hydraulic component to increase accuracy in headwater tributaries and at channel junctions where the backwater effect plays a role. Godbout et al. (2019) improved the hydraulic geometry estimation by proposing a segmentation of the National Hydrography Dataset (NHDPlus) river network, used as the default network in this approach. Viterbo et al. (2020) incorporated the HAND approach as a module of the NWM forecast framework and evaluated its performance for the May 2018 Ellicott City, Maryland, flood event.

Some authors, however, have also identified shortcomings in the HAND approach to flood inundation mapping. Johnson et al. (2019) compared inundation maps for various storms developed using the NWM and HAND with those measured by remote sensing, and concluded that inundation is systematically underpredicted in lower-order reaches and overpredicted in higher-order reaches. They suggested that the Manning's coefficient used in the HAND mapping is too small in lower-order reaches and too large in higher-order reaches. In the NWM version 2.0, a constant value of Manning's n is used, regardless of the order of the reaches. Wing et al. (2019) make the case that "Planar approximations such as these, which do not consider flow physics, have been shown to be less skillful than models which represent the dynamics of flood inundation since the inception of raster-based hydraulic modeling." They compare their physics-based continental-scale two-dimensional hydrodynamic approach with inundation mapping based on the NWM and HAND for Hurricane Harvey, and conclude that their approach produces superior flood inundation maps. In practice, the National Weather Service updates the NWM calculations for current conditions and short range forecasts hourly, and the inundation mapping part of the computation takes a small proportion of the total computation time. This means that the inundation mapping for large regions such as Texas needs to be completed in a time measured in minutes, not hours, and the HAND approach satisfies this operational criterion.

The HAND-based inundation map services were first developed and tested for Texas, and a very large-scale test of their application was provided by Hurricane Harvey, which occurred in late August and early September 2017. For rainfall of three to five days duration, Hurricane Harvey significantly exceeded all the previous worst storms in the continental U.S. In this paper, we present a study to validate the water levels and inundation extents generated from the NWM-HAND system during Hurricane Harvey, using the high-water marks collected by the U.S. Geological Survey (USGS) and the inundation extent estimated by the Dartmouth Flood Observatory (DFO) with satellite data (Brackenridge and Kettner 2017). For the comparison with high-water marks, we compute differences in ground elevation and local water depth and quantify the errors in terrain inputs and in water depths. Channel roughness coefficients adopted in the model are then further calibrated based on stream order to minimize the water depth errors, consistent with the conclusions reached by Johnson et al. (2019). The comparison with high-water marks is first performed over the entire Texas Harvey-impacted domain using 10-m terrain data, and then over part of central Texas where high-resolution topography is available from recent lidar surveys. Within the lidar coverage, an inundation extent is also computed from the high-water marks. This extent is compared with the extent generated from the model and the accuracy is quantified with performance metrics. The comparison with the DFO inundation extent is performed in the catchments with high-water marks and over the entire impacted area in Texas.

The paper is organized as follows: after introducing the flood event (Hurricane Harvey), study area (southeast Texas and part of central Texas), and datasets, we briefly review the NWM-HAND flood mapping system and the improvements brought by GeoFlood, our workflow for flood inundation mapping on high-resolution topography. We describe the method used to compare the modeled water levels and the high-water marks, the calibration of the channel roughness coefficients, and the computation of the inundation mapping performance metrics. The

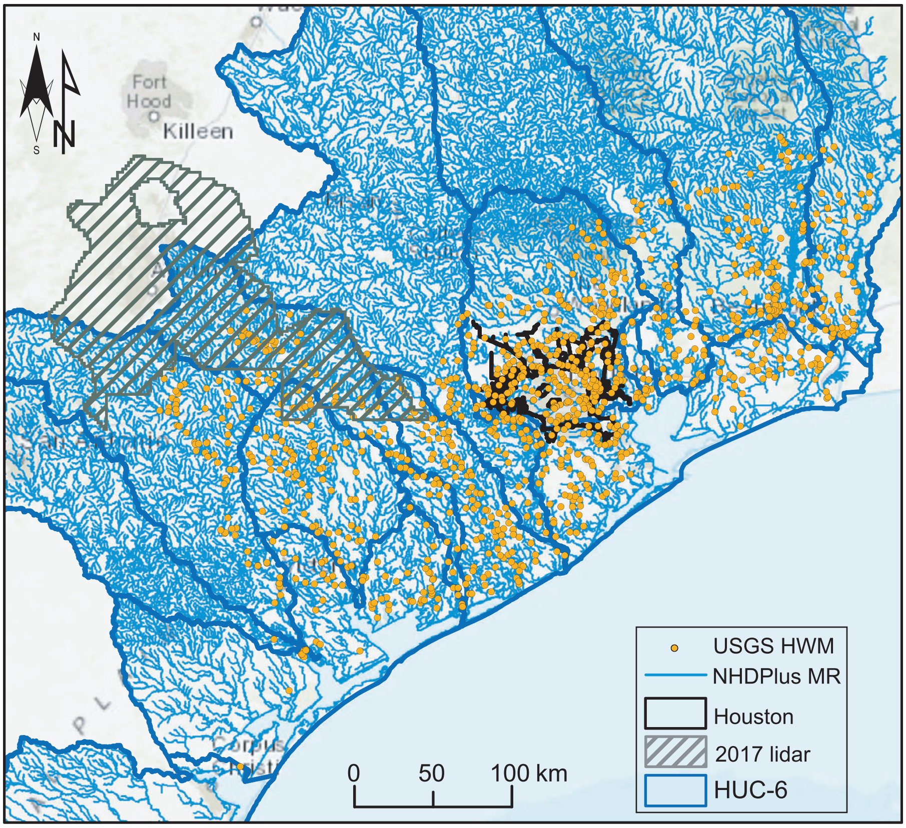

FIGURE 1. Overview of the study area: Harvey impacted areas in Texas. Flood peak summary points collected by the USGS during Harvey are used to quantify the performance of the National Water Model Height Above Nearest Drainage (NWM-HAND) flood mapping system. The gray dashed area indicates the extent of the lidar survey, while the black solid boundary indicates the extent of the City of Houston. Blue solid boundaries indicate the six-digit Hydrologic Units (HUC6) watersheds analyzed. HWM, high-water mark; NHDPlus MR, National Hydrography Dataset Plus medium resolution; USGS, United States Geological Survey.

modeled water depths and inundation extents are compared to the high-water marks and the DFO map and differences are discussed. Finally, we draw conclusions from this work on the capabilities of large-scale flood inundation mapping.

FLOOD EVENT, STUDY AREA, AND DATASETS

On August 25, 2017, Hurricane Harvey (referred to as Harvey hereinafter) made landfall near Rockport, Texas, as a Category 4 hurricane. Its inland movement resulted in the most significant tropical cyclone rainfall event in U.S. history in both scope and amounts (Blake and Zelinsky 2018). Its total eight-day rainfall depth exceeded 1,500 mm in some locations, which was about 300 mm greater than the previous historic continental U.S. record (Blake and Zelinsky 2018). As a result of the overwhelming precipitation, historic flooding occurred in Texas, causing at least 68 direct fatalities as the deadliest hurricane to hit this area since 1919 and $125 billion of damage as the second costliest U.S. tropical cyclone (Blake and Zelinsky 2018).

This study focuses on southeast Texas and part of central Texas (Figure 1). We analyze all the basins where high-water marks were collected (Watson et al. 2018), resulting in 13 six-digit Hydrologic Units (HUC6). High-water marks are the evidence of the highest water levels (peak height of high water) during a flood (Koenig et al. 2016); during and after a storm, hydrologists visit the field and flag the marks left behind in natural and man-made environments by tranquil and rapid flowing water with highly visible signs. After the flood, follow-up surveys are conducted to measure the location and height at these locations. These marks provide valuable information about recent and historical flood events and have been widely used in various flood-related research topics such as flood frequency analysis (Sweet et al. 2013), inundation mapping (Schumann et al. 2008; Cariolet 2010), indirect discharge measurement, and damage assessment. In the U.S., the USGS is the main federal agency in charge of the collection, processing, and publication of high-water marks.

The high-water mark collection during Harvey is the most extensive effort that the USGS has made. Over 2,000 high-water marks were surveyed in southeast Texas and three parishes across southwest Louisiana by over 100 USGS employees in about five weeks. At some sites, water surface elevations were measured at multiple marks and the different measurements were averaged. Therefore, these raw marks were synthesized in 1,263 high-quality peak summary points, which have the most accurate estimation of the stage level at each measured reach. Right after Harvey, the USGS and Federal Emergency Management Agency (FEMA) initiated a study to estimate the magnitude of flooding and map its extent in Texas. In that study, high-water mark data (peak summary), together with discharge information measured at USGS stream gages, were used to create 19 inundation maps for six severely flooded basins. Both the inundation extent and the water depth grid were generated at each site. Although these maps provided valuable information to emergency managers after the event, some limitations can be found when they are examined in detail. First, their spatial extent is limited as these maps only cover parts of several main stem rivers in the hurricane-affected region. According to the NHDPlus MR (medium resolution), the total river length in the impacted region is 163,228 km, while USGS maps only cover 4,762 km, which is 3% of the entire network. No maps are available for most tributaries and a large portion of the main rivers. Second, even though the Harvey high-water mark collection was extensive, its density was not high enough to accurately map local inundation across the impacted zone. Objectively describing the inundation extent of a small, rural reach requires five to 10 marks, and more are needed in urban environments with man-made structures (Koenig et al. 2016). Applying spatial interpolation techniques with a limited number of sparsely distributed high-water marks to generate large-scale flood extents can result in significant overestimation that neglects most local inundation details (Figure 2).

Within the Harvey dataset, 2,309 of 2,359 high-water marks and 1,211 of 1,263 peak summary points are located in Texas. We use the peak data to evaluate the performance of our stage level estimation and inundation mapping system. At each peak summary point, two kinds of elevation-related quantities are measured: the peak stage and the height above ground. The former is the peak height of flood water above the geodetic datum, namely the North American Vertical Datum of 1988 (NAVD 88) in this study. The latter records the local depth at the point

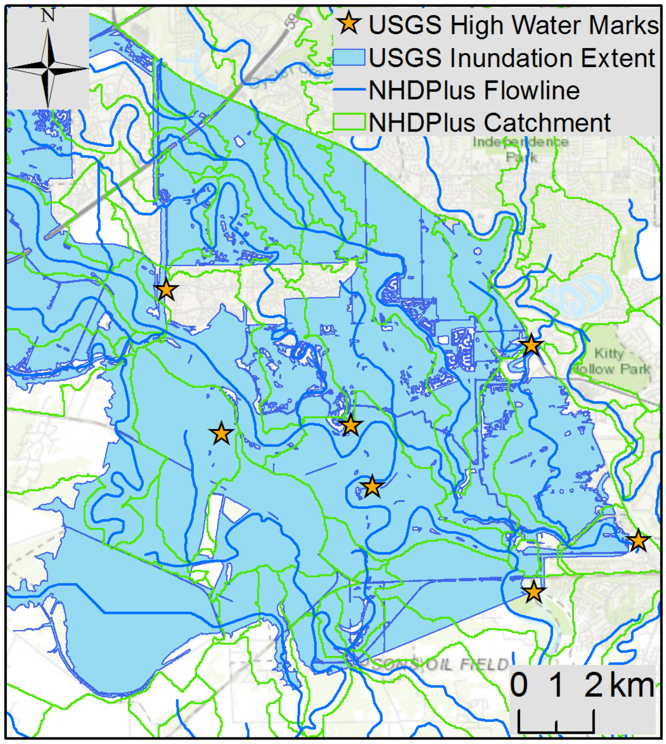

FIGURE 2. Harvey inundation map of the lower reach of the Brazos River created by the USGS using high-water marks and interpolation.

marked when the stage reached the peak level. Both quantities are used in this study to evaluate the different components of our approach.

Another independent source of Harvey inundation extent is the DFO map (Figure 3), which was produced by overlapping data from different sources (NASA MODIS, ESA Sentinel 1, ASI Cosmo SkyMed, and Radarsat 2). The flooded area captured in this map represents the maximum inundation extent during the entire event. The DFO map does not show inundation North-East and South-West of Houston (Figure 3). For this reason, despite the presence of high-water marks, these areas are excluded in the comparison with our method, resulting in 633 catchments analyzed, which include both urban and rural areas.

Recent development of hyper-resolution large-scale hydrological modeling (Archfield et al. 2015; Bierkens et al. 2015; Salas et al. 2018) has significantly increased the spatial and temporal density of streamflow estimates. Since the launch of the NWM in August 2016, simulated discharge information is available for each of the 2.7 million river segments defined in the NHDPlus MR. We use the NWM outputs in our flood mapping system as real-time input streamflow information. The NWM has four types of operational configurations, including: analysis and

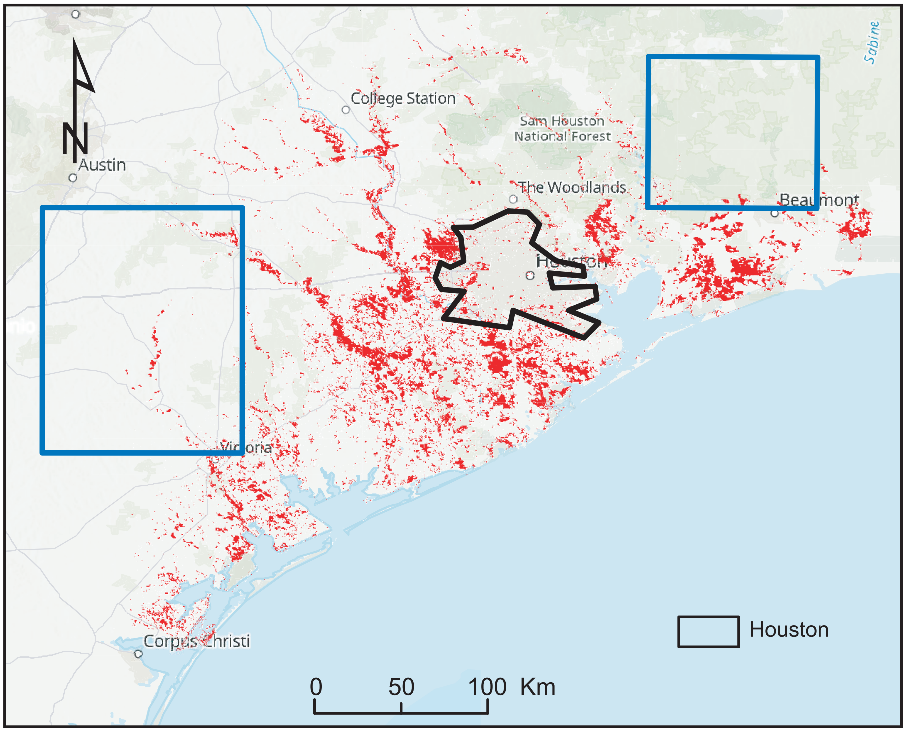

FIGURE 3. Inundation extent produced by the Dartmouth Flood Observatory (DFO) for Hurricane Harvey and published on September 8, 2017. Red indicates flooding and the two boxes mark the areas excluded from the analysis due to lack of data. The black solid boundary indicates the extent of the City of Houston.

assimilation, short-range forecast, medium-range forecast, and long-range forecast. Our study aims at estimating the best performance that the NWM and the HAND flood mapping system can achieve for this event. Therefore, we selected the most accurate streamflow output of the NWM, which is the one obtained from the analysis and assimilation model. This model uses a nudging-based data assimilation technique to ingest observations from about 7,000 USGS stream gages as real-time flowrates, and propagate a correction throughout the entire river network. The model also incorporates information from 1,506 reservoirs. The analysis and assimilation model is executed hourly and provides a snapshot of the hydrologic conditions during the previous three hours. We archived NWM analysis and assimilation products from August 23, 2017 to September 3, 2017 for each NHDPlus stream reach in the Harvey-impacted region.

Our flood mapping system relies on hydrologic terrain analyses and simplified hydraulic assumptions to allow the conversion from streamflow to water depth and inundation extent. The analyses that cover the entire Harvey-impacted area take the 1/3rd arc-second (about 10 m) national elevation dataset (NED) as terrain input and generate hydrologic terrain attributes at the same resolution. These results have been computed in a previous study (Liu et al. 2018) and published online. We also use a Digital Elevation Model (DEM) (coverage shown in Figure 1) at 1-m resolution provided by the Texas Natural Resources Information System, which was generated from their 2017 lidar survey.

METHODS

Water Depth and Inundation Estimation Procedure of the NWM-HAND System

HAND measures the relative height of a given cell above the nearest flowline cell that location drains to. This elevation difference is used to define the flood depth at that cell: when the real-time stream water depth (h) is greater than the HAND value (hand_i) of a cell i, that cell is classified as flooded and the local water depth (d_i) at that location can be computed as:

$$d_i = h - \mathrm{hand}_i. \tag{1}$$

This computation, performed for all the cells within the local drainage catchment of a stream reach, results in a water depth grid for that catchment associated with a flowline depth h. In a recent study (Zheng, Tarboton, et al. 2018), we showed that channel geometric properties, such as flood volume, inundated surface area, and inundated bed area corresponding to the depth h, can be derived from that depth grid. Furthermore, dividing these variables by the length of the river produces additional channel hydraulic information including cross sectional area, channel top width, wetted perimeter, and hydraulic radius. Repeating the calculations presented above at different water depths, the relationship between water depths and different channel hydraulic properties can be obtained (Zheng, Tarboton, et al. 2018). Under the assumption of one-dimensional steady flow, the Manning's equation is then applied to generate a synthetic rating curve, knowing the channel bed slope and the surface roughness coefficient. This rating curve is used to convert the streamflow estimate provided by the model to its corresponding water depth and compute flood inundation extent and depth (Zheng, Tarboton, et al. 2018).

Since the HAND grid relates the relative height above the nearest stream to flooding at a given location, having an accurate river network as the local datum for the HAND calculation is essential to the accuracy of the inundation map produced. Therefore, we recently coupled HAND with GeoNet (Passalacqua et al. 2010; Sangireddy et al. 2016), an advanced river network extraction approach specifically designed for leveraging the information provided by high-resolution topography data, while addressing the challenges associated with their analysis. In the combined workflow, called GeoFlood (Zheng, Maidment, et al. 2018), lidar-derived high-resolution DEMs and the flowlines retraced based on nonlinear filtering, statistical analysis of terrain properties, and a cost minimization approach are used as the input for the HAND flood mapping calculations.

Validation of Water Depth Estimates vs. High-Water Mark Field Measurements

We compute two types of errors related to two elevation-related measurements. The first type of error (e_{depth}) measures the difference between the modeled local water depth (d) and the measured local water depth ($\hat{d}$), reported as the height above ground in the USGS high-water mark dataset:

$$e_{\text{depth}} = d - \hat{d}. \tag{2}$$

The NWM provides streamflow estimates indexed by NHDPlus reaches, which are converted into reach-average flowline depths with our synthetic rating curves. A spatial intersection operation is performed between the NHDPlus catchment and the water marks to assign a flowline depth to each mark. Equation 1 is then applied to compute the modeled local water depth d. If a negative local depth d is obtained at a high-water mark position, it means that the location is not flooded according to our approach. By counting the number of high-water marks with a positive simulated depth ($N_{ed \geq 0}$) and dividing it by the total number of high-water marks (N), a hit rate index (h) can be computed to estimate the overall performance of the model:

$$h = \frac{N_{ed \geq 0}}{N}. \tag{3}$$

The mean and standard deviation of the local depth error e_{depth} are computed for all the high-water marks to quantify the accuracy of the local water depth estimation. Since the modeled water depth is obtained by converting the NWM peak discharge through HAND-derived synthetic rating curves generated with the Manning's equation, the Manning's roughness coefficient, n, adopted to derive the rating curves has a significant impact on the magnitude of the depth error (Johnson et al. 2019). Therefore, we further calibrate the Manning's n value within the generic channel roughness range to minimize the mean local depth error.

This calibration, which identifies the best case among multiple scenarios with different roughness values, is conducted by stream order (o), which indicates the hierarchical position of different reaches within the river system from headwaters to main rivers.

The second error source are the terrain inputs. Since our method estimates inundation based on topography, the terrain input error propagates through the workflow, up to the final stage level estimation. At each high-water mark, the ground elevation ($\hat{g}$) can be calculated from the measured peak stage elevation ($\hat{e}$) and the measured local water depth as:

$$\hat{g} = \hat{e} - \hat{d}, \tag{4}$$

while the ground elevation used in the estimation (g) is directly extracted from the DEM at the same location. Then, the error in the ground elevation e_{ground} can be computed as:

$$e_{\text{ground}} = g - \hat{g}. \tag{5}$$

Unlike the depth error e_{depth}, the ground elevation error e_{ground} is fixed during the channel roughness calibration.

To eliminate the effect of tides on water depth measurements in coastal areas, we created a 2-km buffer zone around the coastline in the NHDPlus dataset, determined by clustering all the high-water marks classified as "coastal" in the metadata. All the locations within this buffer zone are excluded from the analysis (139 points of 1,211 peak summary points, resulting in 1,072 used in the analysis).

In order to explore whether the performance of our mapping system can be improved by adopting high-resolution terrain data, we use the lidar DEM as input and the recently developed GeoFlood workflow to reconstruct the river network, the synthetic rating curves, and the inundation extent. We then reevaluate the performance of our approach using the same metrics already introduced.

RESULTS AND DISCUSSION

Texas Harvey-Impacted Zone Comparison with NED

Water Depth Comparison. The error in ground elevation over the entire domain has a mean of 0.06 m and a standard deviation of 3.46 m (as previously found by Zheng 2018). However, there are four points in the dataset with an error >20 m, indicating potential errors during the measurements. When these outliers are removed from the sample, the mean error shifts from 0.06 to −0.05 m, and the standard deviation drops from 3.46 to 1.53 m. Among the 1,072 peak summary points, 313 of them (29.2% of the total) have a ground elevation error <0.305 m and 707 of them (66.0%) have a ground elevation error of <1 m. These numbers demonstrate that the 1/3 arc second NED provides an acceptable estimation of the actual elevation over the region but with large uncertainties at individual locations.

When estimating the error in local water depth estimation, a reach-based channel roughness calibration is first conducted over the entire dataset. The 1,072 peak summary points are located in 892 NHDPlus catchments. The optimal Manning's n value is identified within the range 0.01–0.2 (Chow 1959), with an interval of 0.005 for each individual reach where NWM estimates are available. When multiple peak summary points are located in the same catchment, the optimal Manning's n for that reach is obtained by minimizing the reach-average depth error. This calibration effort results in a mean water depth error of −0.68 m and a standard deviation of 2.20 m. As a reference, the uncalibrated simulation with a single Manning's n value of 0.05 gives a mean error of −0.45 m and a standard deviation of 3.60 m. Among the 1,072 peak summary points, 449 of them have positive water depths and thus are flooded according to the NWM-HAND model; this value corresponds to a hit rate of 41.9%. The underperformance shown by the results is partially due to the significant randomness associated with the point sampling strategy of the high-water mark collection, which is different from the areal comparison widely used in traditional inundation extent comparisons. If only positive water depths are taken into account during the Manning's n calibration process, our system detects 653 flooded points, corresponding to a hit rate of 60.9%. At the points with positive water depths, the depth error has a mean of 0.57 m and a standard deviation of 0.94 m.

Although the reach-based channel roughness calibration can achieve optimal performance, it is not practical during modeling since no ground truth information is available. Therefore, we explore bulk calibration alternatives, dividing stream reaches into different groups and then assigning a generic channel roughness coefficient to streams in each group. We identify groups based on stream order and stream level. The results (Table 1) show that, compared to the previous reach-based calibration, a stream-order-based calibration gives better total estimation (smaller total mean error of −0.39 m) due to the error compensation among different sites. This difference in error does not necessarily mean that the bulk calibration performs better than the reach-based individual one in all cases. However, this result demonstrates the value of a bulk calibration, which is computationally advantageous, especially when the model is running in operational mode. The estimation error in first-order streams is significantly greater than in higher-order streams. When a similar analysis is applied to the ground elevation errors, such a difference cannot be detected, indicating that it is mainly due to the flow underestimation of the NWM in headwater catchments. The residual error in the mean water depth represents the uncertainty that cannot be addressed by calibrating the Manning's n coefficient. A decreasing trend with increasing stream order can be observed in the optimal Manning's n value (Table 1; Figure 4), which is consistent with the fact that headstreams usually have higher roughness than downstream rivers.

The optimal relationship between Manning's, n, and stream order, o, is found to be:

$$n = 0.2313o^{-1.325}. \tag{6}$$

An equation of this form may be useful to correct for the underprediction in streams of low order and overprediction in streams of high order identified by Johnson et al. (2019). We also calibrated the

TABLE 1. Comparison of local water depth between NWM-HAND estimates with stream-order-based calibration and USGS high-water mark measurements.

Stream order	Number of peak summary points	Optimal Manning's *n*	Depth error mean (m)	Depth error standard deviation (m)
1	354	0.2	−1.22	2.69
2	307	0.1	−0.01	3.02
3	192	0.065	−0.04	2.78
4	93	0.045	0.10	4.05
5	45	0.03	−0.06	3.77
6	67	0.01	0.34	2.75
7	14	0.025	0.13	2.91
Total	1,072		−0.39	3.05

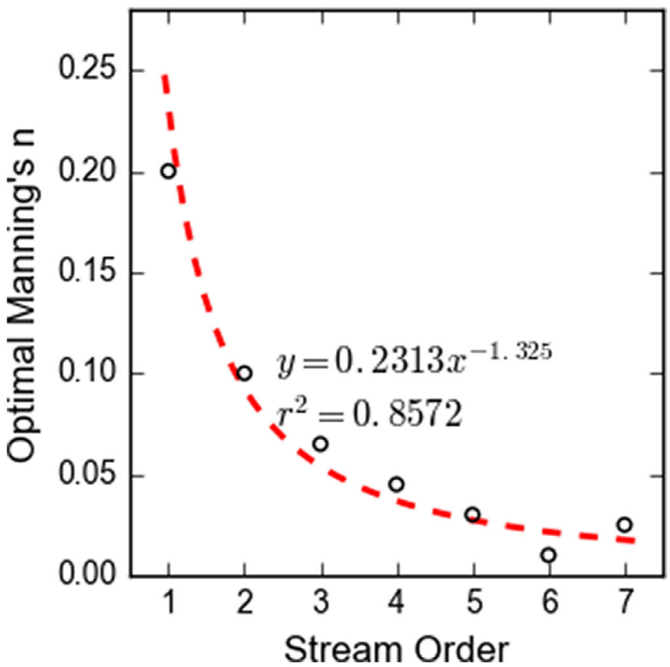

FIGURE 4. Optimal channel roughness coefficients calibrated for different stream orders.

roughness coefficient based on stream level (the opposite of stream order), another stream numbering system, which assigns the hierarchy of streams from the mouth. The results (Table A1) confirm the findings of the calibration based on stream order.

Inundation Extent Comparison. To assess the accuracy of our flood maps, for each NHDPlus catchment containing high-water marks, we reconstruct the inundation based on the measured high-water marks, using a flowline depth calculated as the sum of the high-water mark HAND value and the measured local water depth. Then, we compare it to the inundation extent generated with our optimal modeled depth (Figure 5). When the same mapping procedure is implemented in the 892 NHDPlus catchments where the 1,072 peak summary points are located, the total inundated area computed with the USGS high-water mark measured depths is 6,049 km^2 vs. 5,526 km^2 with NWM-HAND modeled depths; thus, the total modeled extent covers 91.3% of the one reconstructed with high-water marks. However, when the performance is examined at each individual site, a significant amount of variation is detected: 48.8% of the sites have an estimation/observation area ratio between 0.5% and 1.5%, and 80.1% of the sites have an area ratio between 0 and 2, suggesting that our system estimates more than double inundation coverage for nearly 20% of the sites.

Central Texas Harvey-Impacted Zone Comparison with Lidar Data

Since uniform quality high-resolution terrain data are not available for the entire study area, we focus on Central Texas where a 1 m DEM has been derived from the 2017 lidar survey part of the Strategic Mapping Program; 49 peak summary points located in 45 NHDPlus catchments are within the lidar domain. The difference between the NED and the lidar-derived DEM, and the difference between the NHDPlus MR flowline and the GeoFlood-extracted one are shown in Figure 6.

Water Depth Comparison. We compute the error in ground elevation for the 49 high-water marks and find that the mean error is now decreased from −0.47 m with NED to −0.25 m with the 1 m lidar DEM. The standard deviation of the error is also reduced from 2.19 m with NED to 1.85 m. The significant variation still present could be due to the removal of artificial structures during the hydro-enforcing process when the DEM was produced. Many of the high-water marks were collected on roads and bridges across rivers in flood, whose elevations are not retained in the DEM. Among these 49 peak summary points sampled from the lidar DEM, 22 of them (44.9% of the total) have a ground elevation error <0.305 m, and 36 of them (73.5% of the total) have a ground elevation error <1 m. The corresponding values calculated with NED are 10 (20.4%) and 24 (49.0%). The numbers listed here suggest that the lidar DEM provides a more accurate and robust estimation of land surface elevation, compared to the NED.

When estimating the error in local water depth obtained with the reach-based channel roughness calibration, using either lidar or NED results in 18 of 49 sites with positive modeled depths, corresponding to a hit rate of 36.7%. However, the lidar case has a

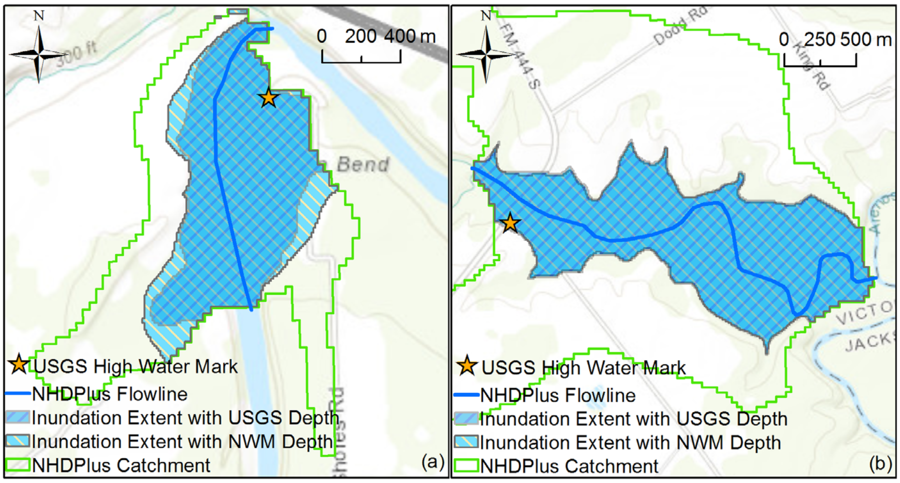

FIGURE 5. USGS high-water mark and NWM-HAND inundation extent comparison: (a) Overestimation case, Reach 5790068 on Colorado River. The USGS inundation extent is created with a flowline depth of 6.79 m and the NWM inundation extent is created with a flowline depth of 7.50 m. These two extents have an area ratio of 1.21. (b) Underestimation case, Reach 9349353 on Garcitas Creek. The USGS inundation extent is created with a flowline depth of 6.34 m and the NWM inundation extent is created with a flowline depth of 5.89 m. These two extents have an area ratio of 0.97.

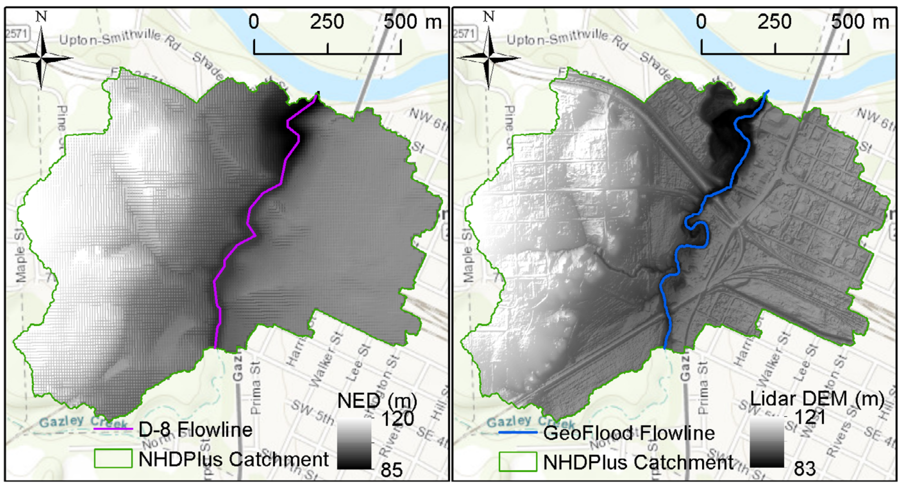

FIGURE 6. Comparison of terrain datasets and flowlines: (a) 1/3 arc-second NED and NHDPlus MR flowline, (b) 1 m lidar-derived Digital Elevation Model (DEM) and flowline extracted with GeoFlood.

smaller mean (−0.66 m) and standard deviation (1.36 m), compared to the NED case (mean of −0.84 m and standard deviation of 1.96 m). If only positive water depths are kept during the calibration process, the lidar results capture 30 of 49 sites with a mean of 0.27 m and a standard deviation of 0.35 m, while the NED results capture 22 with a mean of 0.77 m and a standard deviation of 1.06 m. When the stream-order-based calibration is applied, the results obtained with lidar data (Table 2a) and the NED (Table 2b) show that, compared to the NWM-HAND results generated with NED, the mean error in local water depth decreases from −0.74 to −0.33 m. The results of our global analysis (Section 4.1.1) are confirmed, including the difference in the depth error magnitude between first-order streams and higher-order ones.

Inundation Extent Comparison. In the 45 NHDPlus catchments in Central Texas where the 49 peak summary points are located, the total inundated area computed with the USGS high-water mark measured depths is 38.34 km^2, while that computed with NWM discharge, GeoFlood workflow and calibrated Manning's coefficients is 27.80 km^2, covering 72.5% of the one reconstructed from high-water marks. As a reference, the estimation made by the NWM-HAND approach with NED covers 70.4%. At

TABLE 2. Comparison of local water depth between NWM-HAND estimates with stream-order-based calibration and USGS high-water mark measurements using (a) lidar terrain inputs and (b) NED terrain inputs.

Stream order	Number of peak summary points	Optimal Manning's n	Depth error mean (m)	Depth error standard deviation (m)
(a)				
1	11	0.2	−1.16	1.82
2	12	0.2	−0.18	2.36
3	13	0.17	0.01	1.50
6	13	0.035	−0.10	2.37
Total	49		−0.33	2.04
(b)				
1	11	0.2	−2.10	2.17
2	12	0.2	−1.03	1.89
3	13	0.2	−0.59	1.80
6	13	0.015	0.67	3.32
Total	49		−0.74	2.48

each individual site, both NED and lidar results show 27 sites (55% of the 49 sites) with an estimation/observation area ratio between 0.5 and 1.5, and 41 sites (84% of the 49 sites) with a ratio between 0 and 2 (note that, although the numbers are the same, the specific catchments in the NED case and the lidar case are not always the same). These numbers show that, first, tuning the channel roughness coefficient cannot fully compensate the errors of the NWM, and adopting higher resolution terrain inputs with an approximate inundation mapping approach does not necessarily improve the accuracy of the estimated flood extent. The equivalence between inundation extent estimates generated with different terrain inputs can be explained by noting that the variation in local inundation extent computed with high-resolution terrain data diminishes as water level rises and more areas become flooded (Figure 7). It is important to note that using high-resolution terrain results in more accurate and detailed flood depth estimates (Figure 8), especially near artificial structures in urban environments. Local inundation details cannot be revealed with low-resolution terrain data because these landscape details are not captured.

Comparison with Maximum Extent from the DFO

We compare our estimated inundation to the DFO extent in two domains as above: in catchments containing high-water marks and over the entire impacted area in Texas. The percentages of correct estimation, underestimation, and overestimation are defined in terms of flooded area of our approach. To perform the comparison within the catchments containing high-water marks, we focus on those catchments that are flooded according to the DFO map. In order to make our results and the DFO map comparable, we resampled the DFO map at a spatial resolution of 10 m.

The total inundated area computed with the NWM-HAND depth is 2,695 km^2, while the DFO flooded area is 2,138 km^2. Results show that the overlap corresponds to 26%, the DFO-only coverage to 30%, and the NWM-HAND-only coverage to 44% of the total area. The performance does not appear to depend on land use, as we have obtained similar values after dividing the domain in rural (185 catchments) and urban areas (448 catchments). At several individual sites, the flood extent estimated by our method is similar to that of the DFO map in terms of overall inundation pattern, rather than in terms of area ratio (Figure 9). Detection of flooding from satellite data is known to be altered by the backscatter caused by buildings and by the presence of vegetation, likely playing a role in the differences here observed (Figure 9). Citizen-contributed data could be used to improve the prediction in these areas (Yang et al. 2019).

We performed the same analysis over the entire area; for each HUC6 unit, we computed synthetic rating curves with the calibrated Manning's n and the flooded area corresponding to the peak discharge for each reach from the NWM (Figure A1). We clipped the domain in order to isolate the areas that were flooded according to the DFO map, removing the permanent water bodies and the catchments close to the coast. We obtained a total inundated area of 11,621 km^2 with the NWM depth and 10,105 km^2 from the DFO map. The result shows that the overlap corresponds to 21%, 35% is DFO-only coverage, and 44% is NWM-HAND-only coverage.

The inconsistency of the results might depend on the nature of the phenomenon that we are analyzing and the limitations of both approaches. To examine the quality of the DFO inundation extent itself, we also computed its corresponding high-water mark hit rate. Only 281 of the 1,072 marks fall in the DFO coverage, corresponding to a hit rate of 26.2%. This relatively low hit rate indicates that the remote-sensing-based DFO inundation map may not be capable to capture flood conditions at higher-order streams. The inundation extent produced by the DFO has been obtained by overlaying data at different resolutions and time stamps from multiple sources. Hence, re-projection and rescaling might have reduced accuracy and generated loss of data, possibly explaining the discontinuous pattern of flooding observed in portions of the DFO map.

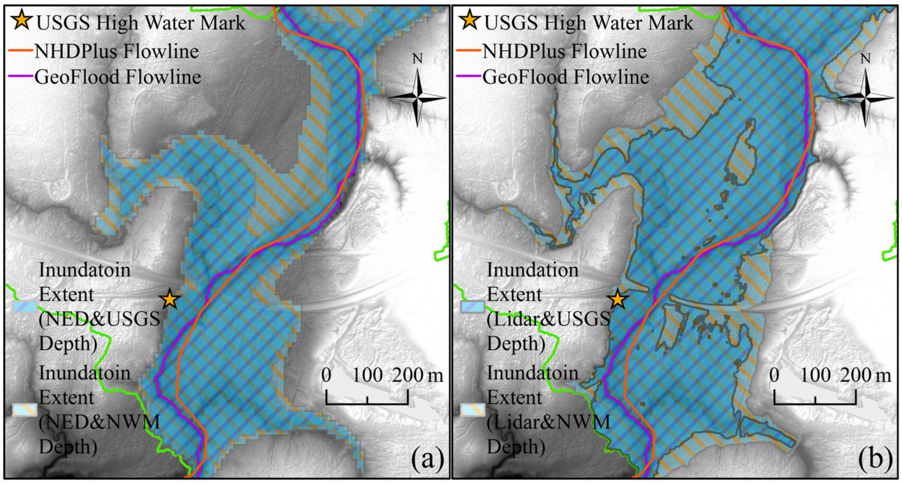

FIGURE 7. Comparison of inundation extents generated with observed and modeled water depths from different resolution terrains: (a) HAND inundation extents derived from 1/3 arc-second NED, (b) HAND inundation extents derived from 1 m lidar DEM.

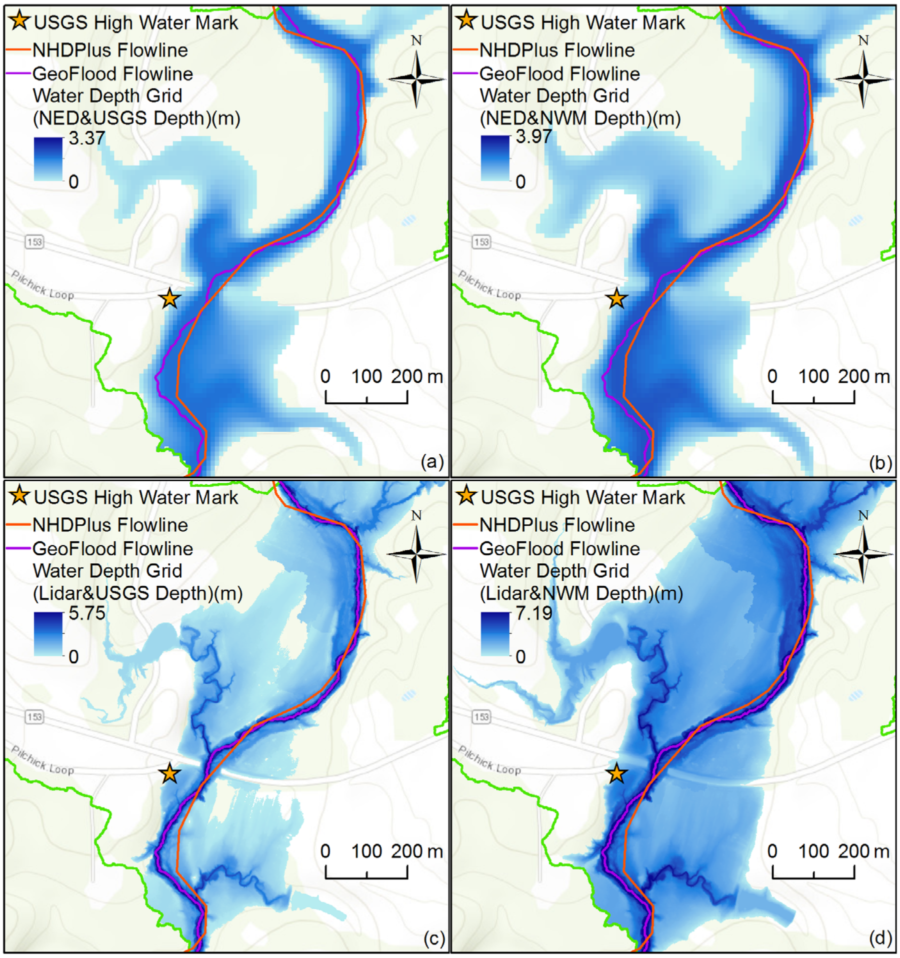

FIGURE 8. Comparison of inundation extents generated with observed and modeled water depths from different resolution terrains: (a, b) HAND inundation extents derived from 1/3 arc-second NED, (c, d) HAND inundation extents derived from 1 m lidar DEM. The left column shows results generated with measured water depths at high-water marks, while the right column shows results generated with simulated water depths.

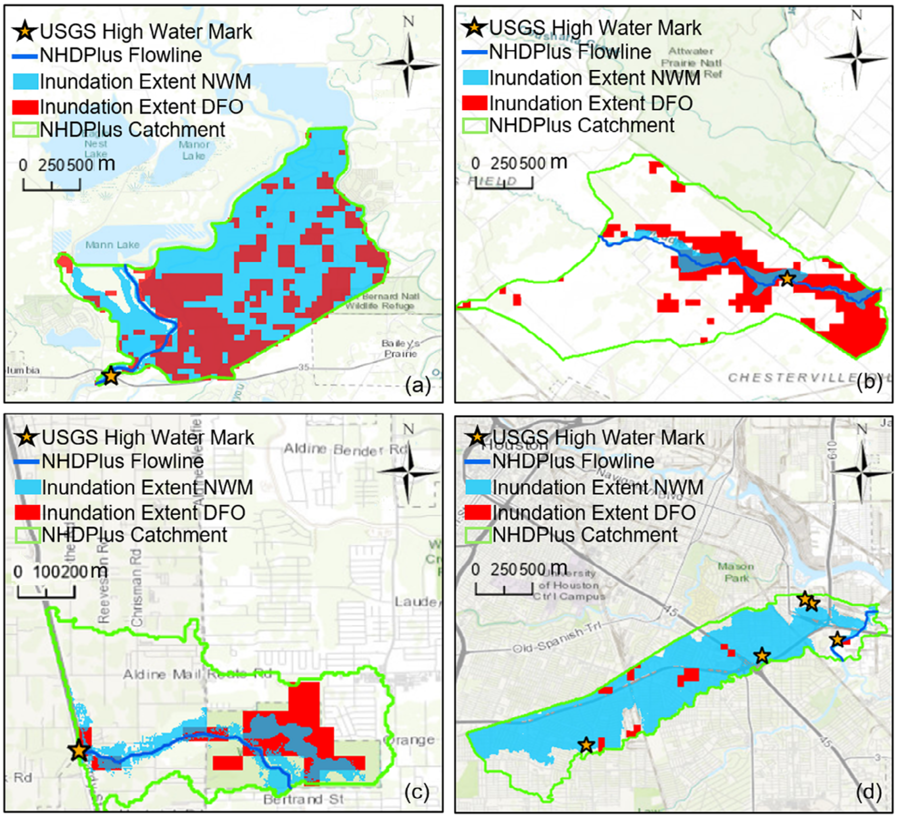

FIGURE 9. NWM-HAND and Dartmouth Flood Observatory (DFO) inundation extent comparison for rural and urban catchments: (a) rural catchment, reach 3124654 on Brazos River; area ratio of 1.54; (b) rural catchment, reach 1605414 on Middle Bernard Creek; area ratio of 0.25; (c) urban catchment, reach 1638595 on Halls Bayou; area ratio of 1.2; (d) urban catchment, reach 1439537 on Sims Bayou; area ratio of 22.5.

CONCLUSIONS

Recent development in operational continental-scale hydrologic modeling and inundation mapping has enabled the creation of regional- and continental-scale flood maps in near real time. To quantify their performance, flood estimates from these systems need to be compared with field measurements collected during historical flood events. High-water marks, which record the highest water levels during a flood, are a well-known type of reference that has been widely adopted in previous flood modeling studies. During Hurricane Harvey, the USGS carried out the most extensive high-water mark collection in recent flood events, resulting in over 2,000 points in southeast Texas and southwest Louisiana. This dataset provides an unprecedented opportunity to examine the performance of large-scale flood models.

In this paper, we presented a study to validate the water depths and inundation maps generated from the operational NWM-HAND flood mapping system vs. corresponding quantities measured at high-water marks; 1,072 peak summary points in Texas were used in this effort. Different elevation-related variables were computed to identify the quality of ground elevation in the DEM and the local water depth generated from the flood model. The results show that the 1/3 arc second NED adopted in the current NWM-HAND model provides overall acceptable estimations but suffers from great uncertainties at individual locations. When modeling local water depths, the depths converted through the HAND-derived synthetic rating curves with either reach-based or stream-order-based calibrated channel roughness coefficients, have a mean error of about 0.5 m and a standard deviation around 2–3 m. The stream-order-based channel roughness coefficient calibration shows a decreasing trend in the optimal Manning's n value associated with increasing stream order. The total simulated extent covers over 90% of the one reconstructed from high-water marks with larger uncertainties at individual sites.

We also investigated the benefits that high-resolution topography can bring to flood inundation mapping by applying the workflow GeoFlood. Water depths estimated from lidar data have smaller mean errors and standard deviations. While the estimation of flood extent is not particularly affected, adopting high-resolution terrain data in flood inundation mapping results in improved depth estimates, especially in urban environments.

A stream-order-based calibration of the channel roughness coefficient improves the results, although future work is needed to compare the findings from this Harvey study to other flood events. The residual model errors that cannot be fully eliminated with the roughness calibration are due to the errors in streamflow estimates from the large-scale hydrologic model, likely due to hydrodynamic effects, such as backwater, which are not captured in the model due to its underlying hydraulic simplifications. Therefore, the information provided by our simplified approach needs to be combined with the outputs from detailed local studies to obtain a comprehensive view of the flood impact caused by extreme flooding events.

The DFO map is a valuable instrument that leads to rapid analysis and comparison with our model; we found similarities in the overall patterns but also discrepancies at local scales. Harvey caused significant pluvial flooding that is not captured by the NWM-HAND approach, possibly explaining the observed underestimation. Also, in several areas the inundation extent in the DFO map is discontinuous, leading to overestimation of our method at those locations. The inputs to these two approaches are different and more analysis is required to understand these differences as the NWM forecasted depths and extent are comparable to those measured at high-water marks.

Overall, our study shows that the current NWM-HAND approach provides a reasonable gross estimation of the inundation caused by large coverage extreme flood events such as Hurricane Harvey. Where available, the use of lidar data is recommended as local terrain and inundation patterns are better captured, resulting in improved flood depth estimation.

APPENDIX

(1) NWM-HAND and DFO inundation extent comparison for the entire Texas impacted area (Figure A1); (2) Comparison of local water depth between NWM-HAND estimates with stream-level-based calibration and USGS high-water mark measurements (Table A1).

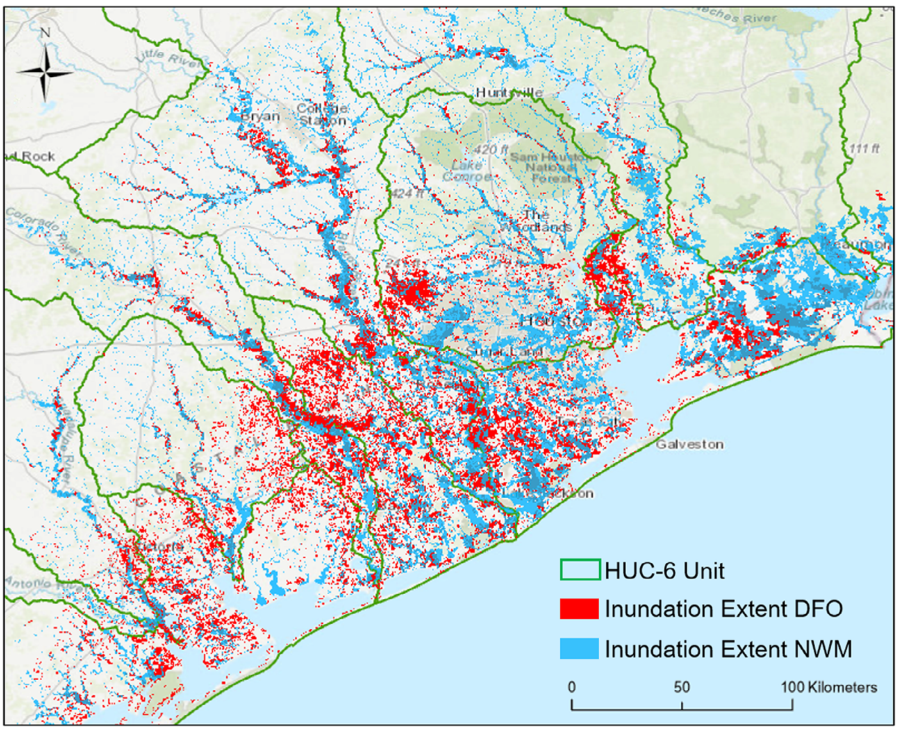

FIGURE A1. NWM-HAND and DFO inundation extent comparison for the entire Texas impacted area.

TABLE A1. Comparison of local water depth between NWM-HAND estimates with stream-level-based calibration and USGS high-water mark measurements.

Stream level	Number of peak summary points	Optimal Manning's *n*	Depth error mean (m)	Depth error standard deviation (m)
1	254	0.025	0.06	3.27
2	387	0.08	0.00	4.23
3	262	0.19	−0.01	3.54
4	124	0.185	0.01	2.46
5	41	0.2	−0.91	3.24
6	4	0.2	−2.26	3.16
Total	1,072		−0.03	3.63

DATA AVAILABILITY STATEMENT

GeoFlood is available for download on GitHub (https://github.com/passaH2O/GeoFlood/tree/master/GeoFlood). The USGS Harvey high-water mark dataset is available on HydroShare (https://www.hydroshare.org/resource/2836494ee75e43a9bfb647b37260e461/). The lidar data are obtained from the Texas Natural Resources Information System (https://tnris.org/stratmap/elevation-lidar/). The 1/3rd arc-second hydrologic terrain analysis datasets for CONUS is hosted at the Texas Advanced Computing Center at the University of Texas at Austin (https://web.corral.tacc.utexas.edu/nfiedata/).

ACKNOWLEDGMENTS

The collection of Hurricane Harvey flood datasets used in this study was supported with the NSF RAPID Grant "Archiving and Enabling Community Access to Data from Recent US Hurricanes" (1761673). We thank the Texas Natural Resources Information System for providing the lidar data and the Texas Advanced Computing Center at the University of Texas at Austin for providing the computing resources and hosting our hydrologic terrain analysis datasets. This work was supported in part by NOAA and by Planet Texas 2050, a research grand challenge initiative of The University of Texas at Austin.

AUTHOR CONTRIBUTIONS

Xing Zheng: Formal analysis; Visualization; Writing – original draft. **Claudia D'Angelo:** Formal analysis; Validation; Visualization; Writing – original draft. **David R. Maidment:** Supervision; Writing – review & editing. **Paola Passalacqua:** Supervision; Writing – review & editing.

LITERATURE CITED

Archfield, S.A., M. Clark, B. Arheimer, L.E. Hay, H. McMillan, J.E. Kiang, J. Seibert *et al*. 2015. "Accelerating Advances in Continental Domain Hydrologic Modeling." *Water Resources Research* 51 (12): 10078–91. https://doi.org/10.1002/2015WR017498.

Bierkens, M.F.P., V.A. Bell, P. Burek, N. Chaney, L. Condon, C.H. David, A. de Roo *et al*. 2015. "Hyper-Resolution Global Hydrological Modelling: What is Next?" *Hydrological Processes* 29: 310–20. https://doi.org/10.1002/hyp.10391.

Blake, E.S., and D.A. Zelinsky. 2018. "Hurricane Harvey (AL092017): 17 August–1 September 2017." National Hurricane Center Tropical Cyclone Report.

Brackenridge, G.R., and A.J. Kettner. 2017. *DFO Flood Event 4510*. Boulder, CO: Dartmouth Flood Observatory, University of Colorado. https://floodobservatory.colorado.edu/Events/2017USA4510/2017USA4510.html.

Cariolet, J.M. 2010. "Use of High Water Marks and Eyewitness Accounts to Delineate Flooded Coastal Areas: The Case of Storm Johanna (10 March 2008) in Brittany, France." *Ocean & Coastal Management* 53: 679–90. https://doi.org/10.1016/j.ocecoaman.2010.09.002.

Chow, V.T. 1959. *Open-Channel Hydraulics*. New York: McGraw-Hill Book Co.

Godbout, L., J.Y. Zheng, S. Dey, D. Eyelade, D. Maidment, and P. Passalacqua. 2019. "Error Assessment for Height above the Nearest Drainage Inundation Mapping." *Journal of the American Water Resources Association* 55: 952–63. https://doi.org/10.1111/1752-1688.12783.

Johnson, J.M., D. Munasinghe, D. Eyelade, and S. Cohen. 2019. "An Integrated Evaluation of the National Water Model (NWM)-Height above Nearest Drainage (HAND) Flood Mapping Methodology." *Natural Hazards and Earth System Sciences* 19: 2405–20. https://doi.org/10.5194/nhess-19-2405-2019.

Koenig, T.A., J.L. Bruce, J. O'Connor, B.D. McGee, R.R. Jr Holmes, R. Hollins, B.T. Forbes *et al*. 2016. *Identifying and preserving high-water mark data (No. 3-A24)*. U.S. Geological Survey. https://doi.org/10.3133/tm3A24.

Liu, Y.Y., D.R. Maidment, D.G. Tarboton, X. Zheng, and S. Wang. 2018. "A CyberGIS Integration and Computation Framework for High-Resolution Continental-Scale Flood Inundation Mapping." *Journal of the American Water Resources Association* 54 (4): 770–84. https://doi.org/10.1111/1752-1688.12660.

National Water Center 2020. "Handbook: NWC Visualization Services." 60 pp. https://www.weather.gov/media/water/Handbook_NWC-Visualization-Services_latest.pdf.

Passalacqua, P., T. Do Trung, E. Foufoula-Georgiou, G. Sapiro, and W.E. Dietrich. 2010. "A Geometric Framework for Channel Network Extraction from Lidar: Nonlinear Diffusion and Geodesic Paths." *Journal of Geophysical Research Earth Surface* 115: F01002. https://doi.org/10.1029/2009JF001254.

Salas, F.R., M.A. Somos-Valenzuela, A. Dugger, D.R. Maidment, D.J. Gochis, C.H. David, W. Yu, D. Ding, E.P. Clark, and N. Noman. 2018. "Towards Real-Time Continental Scale Streamflow Simulation in Continuous and Discrete Space." *Journal of the American Water Resources Association* 54 (1): 7–27. https://doi.org/10.1111/1752-1688.12586.

Sangireddy, H., C.P. Stark, A. Kladzyk, and P. Passalacqua. 2016. "GeoNet: An Open Source Software for the Automatic and Objective Extraction of Channel Heads, Channel Network, and Channel Morphology from High Resolution Topography Data." *Environmental Modelling & Software* 83: 58–73. https://doi.org/10.1016/j.envsoft.2016.04.026.

Schumann, G., P. Matgen, M.E.J. Cutler, A. Black, L. Hoffmann, and L. Pfister. 2008. "Comparison of Remotely Sensed Water Stages from LiDAR, Topographic Contours and SRTM." *ISPRS*

Journal of Photogrammetry and Remote Sensing 63 (3): 283–96. https://doi.org/10.1016/j.isprsjprs.2007.09.004.

Shastry, A., R. Egbert, F. Aristizabal, C. Luo, C.-W. Yu, and S. Praskievicz. 2019. "Using Steady-State Backwater Analysis to Predict Inundated Area from National Water Model Streamflow Simulations." *Journal of the American Water Resources Association* 55 (4): 940–51. https://doi.org/10.1111/1752-1688.12785.

Sweet, W., C. Zervas, S. Gill, and J. Park. 2013. "Hurricane Sandy Inundation Probabilities Today and Tomorrow." *Bulletin of the American Meteorological Society* 94 (9): S17–20.

Viterbo, F., K. Mahoney, L. Read, F. Salas, B. Bates, J. Elliott, B. Cosgrove, A. Dugger, D. Gochis, and R. Cifelli. 2020. "A Multiscale, Hydrometeorological Forecast Evaluation of National Water Model Forecasts of the May 2018 Ellicott City, Maryland." *Journal of Hydrometeorology* 21 (3): 475–99. https://doi.org/10.1175/JHM-D-19-0125.1.

Watson, K.M., G.R. Harwell, D.S. Wallace, T.L. Welborn, V.G. Stengel, and J.S. McDowell. 2018. "Characterization of Peak Streamflows and Flood Inundation of Selected Areas in Southeastern Texas and Southwestern Louisiana from the August and September 2017 Flood Resulting from Hurricane Harvey." *US Geological Survey Scientific Investigations Report 2018-5070*, 44 pp. https://doi.org/10.3133/sir20185070.

Wing, O.E.J., C.C. Sampson, P.D. Bates, N. Quinn, A.M. Smith, and J.C. Neal. 2019. "A Flood Inundation Forecast of Hurricane Harvey Using a Continental-Scale 2D Hydrodynamic Model." *Journal of Hydrology X* 4. https://doi.org/10.1016/j.hydroa.2019.10039.

Yang, D., A. Yang, H. Qiu, Y. Zhou, H. Herrero, C.S. Fu, Q. Yu, and J. Tang. 2019. "A Citizen-Contributed GIS Approach for Evaluating the Impacts of Land Use on Hurricane-Harvey Induced Flooding in Houston Area." *Land* 8 (2): 25. https://doi.org/10.3390/land8020025.

Zheng, X., D.R. Maidment, D.G. Tarboton, Y.Y. Liu, and P. Passalacqua. 2018. "GeoFlood: Large-Scale Flood Inundation Mapping Based on High-Resolution Terrain Analysis." *Water Resources Research* 54: 10013–33. https://doi.org/10.1029/2018WR023457.

Zheng, X., D.G. Tarboton, D.R. Maidment, Y.Y. Liu, and P. Passalacqua. 2018. "River Channel Geometry and Rating Curve Estimation Using Height Above the Nearest Drainage." *Journal of the American Water Resources Association* 54 (4): 785–806. https://doi.org/10.1111/1752-1688.12661.

Zheng, Y. 2018. "Hurricane Harvey: A Quantitative Approach to Assessing the Accuracy of National Water Model Forecasted Inundation." Master's thesis, Texas ScholarWorks, University of Texas at Austin, Austin, TX. https://repositories.lib.utexas.edu/bitstream/handle/2152/68178/ZHENG-THESIS-2018.pdf.