Validation of Urban Flood Inundation Models Applied Using Nationally Available Data Sets: Novel Analyses of Observed High Water Information

Michael Smith[1]; Nathan Patrick[2]; Nels Frazier[3]; and Jongkwan Kim[4]

Abstract: This paper presents novel spatial analysis techniques to evaluate simulations of urban flood inundation from two hydrologic models applied using nationally available datasets. These techniques account for differences in model computational element size and discretization when comparing simulated flood depths to surveyed high-water marks. To complement direct evaluations of predicted depths at high-water marks, our techniques provide five additional metrics to assess modeled depths in areal sectors between the surveyed high-water mark and the stream channel. Our study also demonstrates a novel technique to evaluate flood predictions at damaged structures and crowd-sourced observations of flooded locations. The work in this study is part of a more complete evaluation of two hyper-resolution hydrologic models to generate street-level flood inundation predictions. We used the 679-km^2 Sugar Creek watershed above USGS Gage 02146800, near Fort Mill, SC, which contains the Charlotte, NC, municipal area. We assessed model performance at 172 surveyed high-water marks, 373 locations of flooded structures, and nearly 2,000 crowd-sourced observations of flooded locations. Results indicate that the analyses techniques help distinguish model performance and identify model deficiencies. **DOI: 10.1061/(ASCE)HE.1943-5584.0002129.**

Introduction

Validation of urban flood prediction models requires accurate observations of flood extents and depths. Different methods are used to validate model predictions depending on the type of observed flood data. Satellite imagery and aerial photos can be used to determine flood extent at certain times during a flood event. Some satellite imagery exists for urban areas, but infrequent revisit times and locations limit their utility (Neal et al. 2009; Werner et al. 2005). Social-media data and news reports/photos are growing sources of data for validation and have the potential to provide vast volumes of flood-related information. However, the information in the photos and reports must be converted into useable forms, such as relating the pictured flood level to a local depth at a specific location and time of occurrence. This often requires visits to the pictured site and painstaking photo interpretation and data entry procedures (e.g., Macchione et al. 2019). Nonetheless, photographic high-water mark (HWM) data provide value as shown by Noh et al. (2019), Yu et al. (2016), Xing et al. (2019), Blumberg et al. (2015), Fohringer et al. (2015), Kutija et al. (2014), and McDougall and Temple-Watts (2012).

Debris lines left on the ground at the flooded edge [i.e., wrack lines; Neal et al. (2009)] provide an estimate of maximum flood extent and also give an indication of maximum surface water elevation. High-water marks (HWMs) such as mud lines on trees or buildings can be surveyed to provide point estimates of maximum surface water elevation. For evaluating predicted flood extents, binary pixel-wise metrics, such as the critical success index derived from contingency tables, have been used to quantify the error between predicted and observed wet/dry computational cells (e.g., Wing et al. 2019; Yu and Lane 2006). However, Stephens et al. (2014) noted biases in these metrics and recommended further exploration. Moreover, consensus does not exist in the literature on the best approach to evaluate simulated and observed HWMs. When both the simulated and observed high water indicates an above-ground depth at a specific location, researchers apply traditional measures, such as mean error, root mean square error (RMSE), correlation, and bias, to quantify the vertical error (e.g., Wing et al. 2019; Xing et al. 2019; Yu et al. 2016; Hartnet and Nash 2017; Nguyen et al. 2016; Blumberg et al. 2015; Horritt et al. 2010; Neal et al. 2009; Mignot et al. 2006). However, in modeling studies, it is possible that the predicted flood extent does not reach the location of an observed high-water mark. In such instances, it is not clear how to compute a goodness-of-fit metric. To the best of our knowledge, only a few studies addressed this issue. Savage et al. (2016), Smith et al. (2015), and Neal et al. (2009) computed the vertical difference between the high-water mark elevation and the water surface elevation in the nearest wet cell. In another approach, Hunter et al. (2005) computed the vertical difference between the high-water mark elevation and the digital elevation model (DEM) elevation.

[1]Research Hydrologist, National Oceanic and Atmospheric Administration/National Weather Service (NOAA/NWS) Office of Water Prediction, 1325 East West Highway, Silver Spring, MD 20910 (corresponding author). ORCID: https://orcid.org/0000-0002-7018-1009. Email: michael.smith@noaa.gov

[2]Hydrologist, National Oceanic and Atmospheric Administration/National Weather Service (NOAA/NWS) California Nevada River Forecast Center, 3310 El Camino Ave., Sacramento, CA 95821. Email: nathan.patrick@noaa.gov

[3]Computational Scientist, Cyberdata, National Oceanic and Atmospheric Administration/National Weather Service (NOAA/NWS)/Office of Water Prediction, National Water Center, 205 Hackberry Lane, Tuscaloosa, AL 35401. Email: nels.frazier@noaa.gov

[4]Postdoctoral Fellow, School of Engineering and Science, Dept. of Civil, Environmental and Ocean Engineering (CEOE), Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030. Email: jkim10@stevens.edu

Note. This manuscript was submitted on October 14, 2020; approved on June 30, 2021; published online on September 29, 2021. Discussion period open until February 28, 2022; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699.

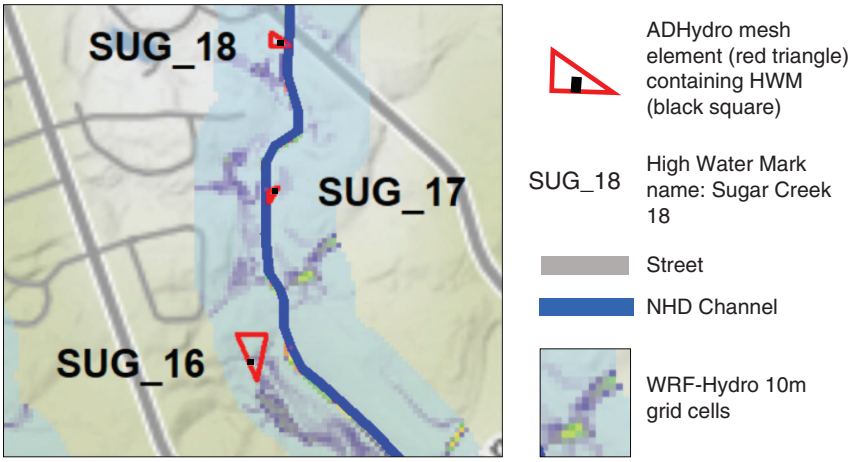

Fig. 1. (Color) Example of difference in the size of computational elements of ADHydro and WRF-Hydro. The small shaded cells are the WRF-Hydro ~10-m grid cells and correspond to the 10-m NED DEM. The larger red triangles are the ADHydro computational mesh elements containing surveyed HWMs 16, 17, and 18 along Sugar Creek. The wide blue line is the National Hydrography Dataset Plus Version 2 stream channel vector. The small black square (barely visible) in each red element is the surveyed high watermark.

Additional complexities emerge when using point-surveyed HWMs to evaluate models having different discretizations and computational element sizes. Fig. 1 illustrates these issues using the two models in the present study: the Weather Research Forecasting Model Hydrologic Extension (WRF-Hydro) (Gochis et al. 2020) and ADHydro (Ogden et al. 2015). ADHydro is an unstructured mesh model that uses smaller elements where more topographic detail is needed, such as channel-overbank boundaries, while larger elements are used to represent areas requiring less detail. In contrast, WRF-Hydro was applied at a ~10-m grid resolution throughout the study domain. In Fig. 1, the underlying DEM is the 10-m USGS National Elevation Dataset (NED). The notations SUG_16, 17, and 18 indicate the location of three surveyed HWMs along Sugar Creek. The thick blue line represents the stream channel vector. The red triangles are the ADHydro mesh elements encompassing the HWMs. Fig. 1 clearly shows the size variation among ADHydro mesh elements and the typical size difference between the ADHydro mesh elements and the WRF-Hydro 10-m grid cells. Because both models utilize the same underlying DEM, differences in elevations between the two models should be small and a function of the underlying grid structure and element size differences. That being said, elevation differences between the models are a source of uncertainty and could introduce error, especially when larger ADHydro mesh elements span areas of high elevation variability in the DEM.

These uncertainties give rise to such questions as the following:

- How does one evaluate predicted water depths originating from models having different underlying grid mesh structures and element sizes?
- How does one assess model performance when a predicted *neighborhood* water depth magnitude is approximately equal to the surveyed HWM depth but is spatially shifted?
- Likewise, how does one evaluate a modeled water depth that matches the extent of the HWM but not the magnitude?

To address these questions, this paper presents novel methods for evaluating model predictions of flood depths at surveyed high-water marks. These techniques account for differences in model element discretization and size when comparing simulated flood depths to surveyed HWMs. We also developed a novel approach to qualitatively analyze inundation predictions at the locations of flood-damaged structures and crowd-sourced observations of flooded locations. The work in this study is part of a more complete evaluation of two hyper-resolution models (HRMs) for predicting urban flooding (Smith et al. 2020). To the best of our knowledge, our evaluation is among the most comprehensive whole-city studies to date, considering the number of storm events and corresponding observations of surveyed HWMs, flood damage locations, and crowd-sourced locations of flooding.

Models

We used two models in our study. ADHydro (Ogden et al. 2015) was developed at the University of Wyoming to simulate large watershed response to climate change. ADHydro has been parallelized to run in a high-performance computing (HPC) environment and uses an unstructured mesh discretization to describe land surface and subsurface characteristics. The model partitions precipitation into runoff using the Green & Ampt redistribution method coupled to shallow groundwater using a one-dimensional (1D) finite-moisture content discretization of the advection-like term of the soil moisture velocity equation (Lai et al. 2015; Ogden et al. 2017). Two-dimensional overland flow is calculated using either the full dynamic wave or diffusive wave (zero-inertia) approximation of the de Saint-Venant equations. The full dynamic or diffusion wave approximations are also used to solve the one-dimensional de Saint-Venant equations for channel flow. Two-way coupling of the overland and channel flow is based on a source-term lateral flow connection using a broad-crested weir equation.

The second model was WRF-Hydro (Gochis et al. 2020), developed at the National Center for Atmospheric Research (NCAR). A version of WRF-Hydro forms the core of the National Weather Service (NWS) National Water Model (NWM). WRF-Hydro has also been parallelized to run in an HPC environment. The Noah-MP Multi-Parameterization (Noah-MP) model (Niu et al. 2011; Yang et al. 2011) is used to compute water balance and runoff generation. Two combinations of overland and channel routing were available. One version has diffusive wave overland flow with two-way coupling to an approximation of diffusive wave channel routing. In the second and selected version, diffusive wave routing is used for both overland and channel flow. However, this version is limited to a one-way coupling between overland and channel flow. All channel flow is retained within trapezoidal elements and cannot overflow onto the floodplain. In this project, we set up WRF-Hydro to run on a ~10-m structured grid. Kim et al. (2021)

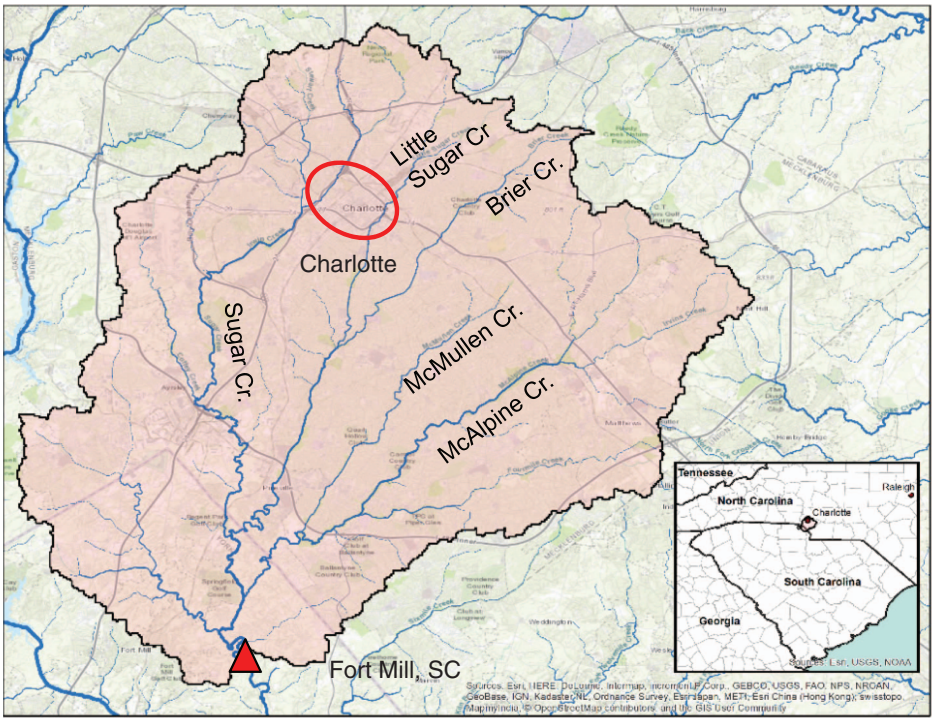

Fig. 2. (Color) Sugar Creek test basin. The outline is the drainage area above USGS Gage 02146800 in Fort Mill, South Carolina. The red ellipse indicates downtown Charlotte, NC. [Map data sources: Esri, HERE, Delorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors and the GIS User Community.]

applied a similar version of WRF-Hydro to study the impacts of spatial and temporal resolution, calibration, initial conditions, and streamflow data assimilation on outlet hydrographs for three small basins.

Study Basin

The study area was the 679-km^2 Sugar Creek watershed above the USGS Gage 02146800. This basin completely encompasses the city of Charlotte, NC (Fig. 2). The Sugar Creek basin lies almost entirely within Mecklenburg County, NC, with a small portion containing the outlet gage in Fort Mill, South Carolina. The Charlotte metropolitan area has undergone rapid urban and suburban growth since the 1960s, with urban area increasing from 31.5% in 1992 to 68.3% in 2011 (Zhou et al. 2017). During the same period, forested area decreased from 55% in 1992 to 27.7% in 2011 [see the study by Zhou et al. (2017) and references therein]. The basin is highly flood-prone, with warm-season thunderstorm systems and tropical cyclones causing the main flood-producing events. This region is served by the Flood Information and Notification System (FINS 2017), a collaborative effort between the USGS and Charlotte-Mecklenburg Storm Water Services (CMSWS) to provide data collection, monitoring, and alert services to the Charlotte metropolitan area.

The Sugar Creek study basin is covered by the NWS NEXRAD radar at Greer, SC (NWS code KGSP) and dense networks of stream and rain gages (Wright et al. 2014; Smith et al. 2002). Charlotte is approximately 130 km from the KGSP radar, but analyses showed minimal range effects (Wright et al. 2014). The data-rich Sugar Creek basin has been used to study the urban response to storm position, movement, and scale (Veldhuis et al. 2018); the hydrology of extreme events (Smith et al. 2002); urban area impacts on flood frequency (Zhou et al. 2017); and the space-time variability of extreme rainfall (Villarini et al. 2010).

Data

Forcing Data

We used the NWS Analysis of Record for Calibration (AORC; Kitzmiller et al. 2014) as the primary meteorological forcing for ADHydro and WRF-Hydro. The AORC is a multidecade, internally-consistent data set of precipitation, temperature, solar radiation (shortwave and longwave downward at the surface), dew point, wind vectors, and terrain-level pressure. AORC data cover the contiguous US (CONUS) at a 1-km grid resolution and have an hourly time interval. Noting the flashy hydrograph response of the Sugar Creek stream gages, we utilized the 2 and 5 min precipitation estimates from the NWS Multiradar Multisensor System (MRMS; Zhang et al. 2016) to time disaggregate the hourly AORC precipitation estimates into 15 min intervals.

We selected two extreme storms to evaluate the models' ability to simulate flood inundation. The flood of record in August 2008 resulted from over 28 cm of rain falling over the entire basin in a 36–48 h period. We also selected a large convective event in August 2011 in which 7–15 cm fell in 3–4 h. Hereafter, we refer to these events by year, e.g., the 2008 and 2011 events.

Geographic Data

Anticipating that there would be future expanded applications of HRMs, we selected only static geographic data sets that had national coverage. The 10 m National Elevation Data set (NED; USGS 2017) was selected to represent topography and define flow directions. Street vectors were derived using Open Street Map and

used to define major urban flow paths for both models. We used the 2011 USGS National Land Cover Dataset (NLCD; Wickham et al. 2017) for land use and land cover. Soil texture information was taken from the Soil Survey Geographic (SSURGO) data set (Soil Survey Staff 2021). We chose to forego strict definitions of building footprints to define surface flow paths. Finally, we used the National Hydrography Dataset Plus (NHD+) version 2 data to define the location of the channel network (Moore et al. 2019). Storm sewer networks were not modeled. Cross-section data were available but not used due to our self-imposed limitation to use only data sets with national coverage. Instead, both ADHydro and WRF-Hydro used empirical stream-order relationships to define channel shape parameters.

Observed High-Water Marks

The CMSWS provided three types of observed HWMs. The first type was surveyed HWMs collected by a local engineering firm on behalf of CMSWS. The vertical datum used was NAVD 88, and the horizontal datum was NAD83/2007. Map projection data utilized the North Carolina 3200 projection. Categorically, CMSWS rated the surveyed HWMs as good, fair, or poor. However, no additional information was available regarding how this rating was determined. The types of surveyed HWMs were mud lines, wrack lines, debris lines, seed lines, stain lines, and witness marks. Second, flood damage depths were geo-encoded parcel locations of damage in which interior surveys were conducted to determine the depth of water within the structure. The structures impacted include private, business, and utility locations. The residential structures impacted include single-family homes, apartments, condos, and townhomes. Business structures include retail, offices, and warehouses. Recorded water depths were relative depths measured within the structure. Considering 2008 and 2011 together, average damage depths ranged from approximately 43–78 cm depending on the type of measurement (i.e., living area, crawl space, etc). Third, flooded streets were geo-encoded locations where flooding was observed. Of the three types of inundation data, flooded streets are the least quantitative and should be considered subjective. No indication of water depth was provided. Sources of the observations were witnesses, news reports, emergency management (such as police, fire, and local government), and photographs of flooding. Appendix I presents examples of the three types of HWMs from the Charlotte basin. Table 1 presents the number of observed HWMs obtained for this study. Examination of the 2011 flood damage (100 values) and flooded streets data (1,951 values) revealed that the latitude/longitude location information between the two types was redundant. As such, the 100 flood damage depths in 2011 were treated as a subset of the 1,951 flooded street observations. Hereafter, we use the term flood damage/flooded streets to refer to this group of observations.

Model Application

We set up ADHydro and WRF-Hydro on the Sugar Creek basin to generate predictions of maximum flow depths in each computational element to compare to the observed high water information for each storm. Noting the major role that streets play in routing urban floods (e.g., Schubert and Sanders 2012), we created the mesh for ADHydro so that major streets were defined as impervious flow paths (e.g., Gallegos et al. 2009). For WRF-Hydro, the DEM corresponding to major streets was artificially lowered to ensure that flow followed street directions. After these steps, the median area of the ADHydro irregular mesh elements for the entire Charlotte basin was 1,922 m^2 or ~44 m on a regular grid side. The basin-wide ratio of ADHydro median element sizes to WRF-Hydro grid cells was ~16:1. Along the channel segments, the median area of the ADHydro mesh elements was 1,579 m^2 or approximately 40 m on a regular grid side. Thus, near the channels, the ratio of ADHydro median mesh element size to WRF-Hydro grid cell size was approximately 13:1. Trapezoidal channel dimensions for both models were defined using empirical stream order relationships that could be applied nationally.

Table 1. Number and type of high-water mark observations for the two storm events

Type of HWM	2008	2011
Surveyed HWM	131	41
Flood damage locations	373	100
Flooded streets	none	1,951

Surveyed HWM data are not available everywhere in the US; thus, we did not use these measurements for model calibration. Our goal was to calibrate model parameters using only nationally-available USGS observed hydrographs to get the hydrograph volume correct and, subsequently, to determine how well the models performed for simulating observed HWMs. Interested readers are referred to the study by Smith et al. (2020) for details regarding model calibration, simulation run periods, initial conditions, and analysis of simulated hydrographs. In this study, we focus on the analysis of inundation results from versions of ADHydro and WRF-Hydro that were calibrated to fit observed hydrographs.

The constraint to use only nationally-available data sets in our underlying feasibility study (Smith et al. 2020) precluded the explicit modeling of buildings, microtopography, storm sewer networks, and cross sections, which likely impacted the simulation accuracy. Nonetheless, the choice of which urban features to model and how to model them must be considered in light of trade-offs in computational time, expected accuracy, and model complexity. Moreover, we still do not know how much physical complexity a flood inundation model needs to address a given problem (Neal et al. 2012). Modelers are cautioned regarding the expectation that increased modeling resolution and complexity will necessarily result in greater accuracy (Dottori et al. 2013). Modeling choices must also consider project goals, end-user requirements, data availability, preprocessing demands, and implementation effort (Schubert and Sanders 2012). These considerations are important to the NWS for the operational implementation of models at a national scale. For example, end users of NWS flood forecasts, such as emergency managers, often want actionable depth information presented in general ranges as they consider what level of response is necessary, such as signage, road closures, and rescue operations.

We present several examples of the trade-off between model complexity (e.g., buildings, storm sewers, and cross sections) and project goals. Horritt et al. (2010) and Gallegos et al. (2009) determined that excessive computational demands with two-dimensional (2D) hydraulic models precluded the use of mesh sizes needed to resolve buildings. Yu et al. (2016) neglected buildings given the project scope and goals. Wing et al. (2019) did not model buildings, streets, or storm sewers in their city-scale evaluation of a 2D hydraulic model and a simple GIS-based approach for Hurricane Harvey in Houston. Even when buildings are modeled, simulation results can be contradictory and confounding. For example, Neal et al. (2009) found that RMSE errors in HWM simulations were slightly worse when buildings were modeled compared to the no-building scenario. Similarly, Grimley et al. (2017) found that representing buildings in the terrain model resulted in slightly worse results in basin outlet hydrograph simulation compared to the no-building case. On the other hand, Schubert and

Sanders (2012) found that the inclusion of buildings is important for modeling local scale velocities and depths but less important for the simulation of hydrographs and flood extents.

Regarding the importance of defining urban microtopography, Fewtrell et al. (2011) conducted a benchmarking study using two variants of a hydraulic model. Spatial resolutions of 25 cm, 50 cm, 1 m, 2 m, and 5 m were used to define the microtopography (e.g., curbs, road camber, etc.) on a very small 0.11 km^2 basin. Such modeling resolutions required the use of vehicle-mounted light detection and ranging (LiDAR) units as airborne LiDAR has been incapable of providing the resolution needed to define urban microtopography (Ozdemir et al. 2013). Furthermore, proprietary software was needed to process the LiDAR data. Clearly, such efforts are nearly impossible at present and in the near future for city-scale operationally-viable forecasting in urban areas across the US.

Studies have shown (e.g., Rafieeinasab et al. 2015; Schumann et al. 2011; Ogden et al. 2011) that in severe rainfall events, such as the two used in our study, the capacity of the subsurface drainage network pales in comparison to the flow conveyed by surface features. Moreover, it is nearly impossible to model all storm sewers in a city-wide domain in the time appropriate for operational forecasting. As a result, decisions must be made as to what level of simplification of the storm sewer network needs to be made to meet project goals (e.g., Habibi and Seo 2018; Leitao et al. 2010). Indeed, the immense complexity of the storm sewer network argues for simplicity as a first modeling step, as in our case (Gallegos et al. 2009).

It is well known that cross-section shape and spacing can have large influences on the extent and depth of flood inundation. Among others, Ali et al. (2015), Cook and Merwade (2009), and Fewtrell et al. (2011) noted differences in flood inundation extents and depths when using cross sections derived from topographic data of various resolutions.

Methodology: Analysis of Surveyed HWMs and Flood Damage/Flooded Street Locations

Analysis of Predicted Depths at Surveyed HWMs

In an idealized setting, predicted maximum water depths in the computational element containing the HWM would be extracted and directly compared to the surveyed depth at each high-water mark location for a direct or reference comparison. Uncertainties in the surveyed HWM value and quality and differences in the underlying mesh structures and size (e.g., Fig. 1) of computational elements of models such as WRF-Hydro and ADHydro, make direct comparisons complicated. Therefore, in addition to the reference comparison, we used the areal sector approach of Patrick et al. (2018) to minimize the impacts of computational mesh differences and potential differences in the representation of terrain in the models. This approach provided five additional metrics to objectively assess simulation performance when the model predicts water in the general vicinity of the HWM.

We use Fig. 3 to illustrate the derivation of areal sectors. The background image in Fig. 3 shows the WRF-Hydro maximum water depth grid cells. For ease of visualization, the background grid cells have been aggregated 4:1 in size. (Recall that all WRF-Hydro calculations were performed at the ~10 m grid scale). The blue line is the NHD+ stream channel vector. The triangular ADHydro mesh element is outlined in red. The yellow square is the WRF-Hydro element, and the black dot is the surveyed HWM location. The yellow triangle outlined in red is the point on the NHD+ stream network vector that is closest to the surveyed HWM. The steps of Patrick et al. (2018) are listed below:

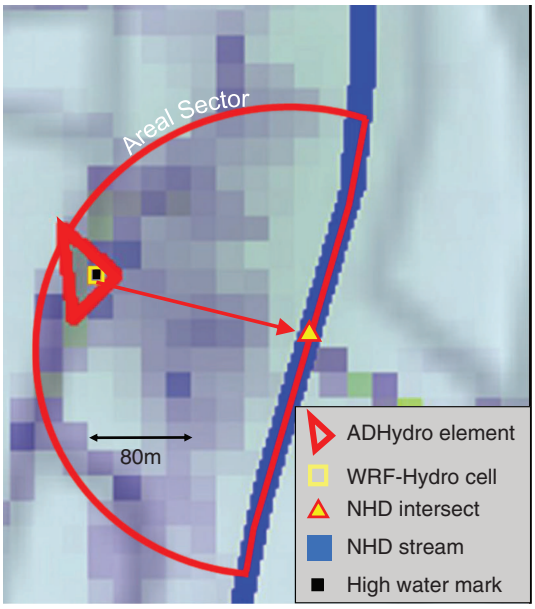

Fig. 3. (Color) Example of the areal sector. The background grid is the 10-m WRF-Hydro maximum depth grid scaled to 20 × 20 m cells for ease of visualization.

1. We computed an areal radius based on 110% of the distance from each surveyed HWM to the nearest point on the NHD+ stream segment vector. The red arrow in Fig. 3 represents the distance from the surveyed HWM to the closest point on the NHD+ network. The rationale for selecting a radius extending slightly beyond the surveyed HWM was to incorporate potential model depths that spatially extended beyond the surveyed HWM. This could help offset spatial errors with respect to on-ground surveyed HWMs (i.e., when the surveyed HWM depth = 0).
2. Using the computed areal radius, we defined a circular buffer centered at the NHD+ stream intersection for each surveyed HWM.
3. The areal sectors were computed by bisecting the circular buffers with NHD+ stream segments, as shown by the red line in the center of the blue NHD+ stream vector. The halves containing the surveyed HWMs were retained. We retained only the sectors containing the HWM because land cover characteristics and topographic profiles were more likely to differ on the opposite side of the stream channel.
4. For both WRF-Hydro and ADHydro, computational grid and mesh elements were then intersected with the computed areal sectors. Depending on the underlying computational element structure, this allowed the selected computational elements to extend beyond the NHD+ stream segment boundary. This was accounted for in later processing.

Using the areal sectors, we computed a variety of predicted water depths for each model as shown in the list below.

- Reference—the maximum water depth of the model computational element that intersects the surveyed HWM location.
- Max—the maximum water depth of any model computational element intersected by the areal sector. The computational element may extend beyond the areal radius boundary. The intent of the maximum depth is to allow for the models to capture the magnitude of the depth within some reasonable distance of the HWM while not assigning a penalty for a lateral spatial offset.
- Mean—the mean water depth of all modeled computation elements intersected by the areal sector. The computational elements may extend beyond the areal radius boundary.

- Median—the median water depth of all modeled computation elements intersected by the areal sector. The computational elements may extend beyond the areal radius boundary.
- IDW–Inverse distance-weighted water depth. Distance is from model element centroids in the areal sector to the surveyed HWM.
- Areal mean—the areal weighted mean water depth computed from model elements within the areal sector. Model elements that are intersected by the areal radius will have their area adjusted to only include the area within the areal sector on the surveyed HWM side of the channel.

One concern about using areal spatial analysis is the impact of uncertainty in ground elevations. Patrick et al. (2018; Appendix II) performed several analyses to explore this concern. Elevation analysis confirmed that there was minimal NED DEM elevation variation within the areal sectors. In addition, distributions of different NED DEM elevations within the areal sectors were explored and compared to the surveyed HWM ground elevation. The variance between surveyed HWM ground elevations and NED 10 m HWM elevations was consistent with the variance between surveyed HWM ground elevations and the areal sector mean elevations. Typically, these relationships were valid except when areal sector areas were large. Overall, median distances from the surveyed HWMs to the NHD+ stream segment intersections were less than 45 m, and the median areal sector areas are less than 4,400 m^2. This information suggests that using an areal analysis does not significantly enhance uncertainty or increase errors for most HWMs. Thus, differences in the models' computed maximum depths at surveyed HWM locations should largely be a function of model physics and discretization.

Analysis of Predicted Depths at Flood Damage/ Flooded Streets Locations

We also developed a novel method to analyze predictions at flood damage/flooded street locations. Recall that these flood reports consist of measurements of flood damage depths in structures and locations where photos, news, and witness reports indicated flooding. Referring to Fig. 4, the analysis steps are as follows:

1. The locations of flood reports were intersected with Charlotte property parcel polygons. This step minimizes the impact of multiple flood reports sharing the same latitude/longitude location (i.e., multi-family residences) and incorporates uncertainty regarding the latitude/longitude of a flood observation. For example, in 2008, there were 373 flood damage reports that intersected 280 unique Charlotte property parcel polygons. In 2011, there were 1,951 observations of flooded streets. These intersected 1,144 property parcels. Based on the geo-encoded observations, all of the parcels experienced some level of flooding, but the absolute depth relative to the DEM is unknown. We surmise that the observed flood depths were greater than nuisance flooding; otherwise, the damaged structures would not have been surveyed, and any news and social media photos would not have been taken and submitted to CMSWS.
2. For each property parcel, we computed the simulated depth as the mean areal average of the maximum depths in the models' computational elements contained in the property parcel.
3. We used discrete depth thresholds to calculate flood *hit* percentages for the parcel polygons. Five thresholds were used: <2.54, ≥2.54, ≥5.08, ≥15.24, and ≥30.48 cm. These were determined as the analyses proceeded and are very similar to those by Yu et al. (2016), who used 2, 15, and 30 cm depth thresholds.
4. We computed the parcel-threshold *hits* or the percentage of parcels at which the models predicted a maximum depth greater than a threshold depth. We also determined the percentage of parcels at which the models computed a maximum depth of less than 2.54 cm.

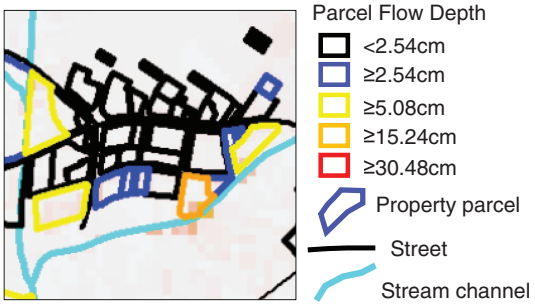

Fig. 4. (Color) Example of maximum simulated inundation depth in property parcels. Each polygon is a property parcel with one or more observations of flood damage/flooded streets. The color of the polygon indicates the depth threshold exceeded.

Results

Results at Surveyed HWMs

We computed maximum depths in each computational element to compare to the total of 172 surveyed HWMs and over 2,000 flood-damage/flooded-street observations. For WRF-Hydro, the maximum depth data files consisted of overland and channel flow depths. The ADHydro team submitted maximum depths in all non-channel computational mesh elements. These depths should be taken into account when interpreting the results. For brevity, we focus on the calibrated results. The uncalibrated results are presented by Smith et al. (2020).

Table 2 presents the analysis of predicted inundation depths at surveyed HWMs. Column 2 lists the six types of predicted maximum depths. The reference value refers to the maximum depth for the model element which contains the surveyed HWM. The other types (maximum, mean, median, IDW, and areal mean) are depths computed using the predicted maximum depths in all model elements residing within the areal sectors which contain each surveyed HWM. Note that the *maximum* category most likely included flow depths in channels for WRF-Hydro. There are 131 surveyed HWMs for the 2008 event and 41 for the 2011 event. Of these, 73 and 15 were above-ground HWMs for 2008 and 2011, respectively.

Table 2 presents root mean square errors and mean absolute errors (MAE) between the predicted and observed HWM depths. We computed these errors for the 2008 and 2011 storm events for the reference depth and the five areal sector computed depths. An observed depth of zero was used in the case of surveyed on-ground HWMs. Columns 3–6 present the errors at all surveyed HWMs, while columns 7–10 highlight the errors for only above-ground surveyed HWMs. Values in bold font indicate the smallest errors in a column.

The reference results (Row 1) in Table 2 indicate that WRF-Hydro generated lower RMSE and MAE values than ADHydro (Row 7) at all locations of surveyed HWMs. All the various areal sector RMSE and MAE values support this result by assessing the predicted depths at and in the vicinity of the surveyed HWM. In six of eight columns, the best overall results were generated by the WRF-Hydro mean depth, as seen by the values in bold in Row 3.

Table 2. RMSE and MAE errors for simulated depths at surveyed HWMs

		All HWMs (m)				Ground HWMs Removed (m)			
		2008 (131)		2011 (41)		2008 (73)		2011 (15)	
Model	Calculated Depth	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	2	3	4	5	6	7	8	9	10
WRF-Hydro									
1	Reference	0.91	0.57	0.53	0.32	0.98	0.81	0.88	0.83
2	Maximum	3.16	2.53	1.50	1.18	2.96	2.20	0.99	0.85
3	Mean	**0.76**	0.58	**0.46**	0.37	**0.79**	**0.62**	**0.67**	**0.61**
4	Median	0.84	0.57	0.52	**0.32**	1.00	0.85	0.85	0.81
5	IDW	0.83	**0.57**	0.51	0.33	0.92	0.75	0.83	0.79
6	Areal Weighted Mean	0.78	0.58	0.47	0.36	0.81	0.63	0.69	0.63
ADHydro									
7	Reference	1.50	0.96	1.50	1.01	1.58	1.08	1.32	1.04
8	Maximum	3.55	2.94	3.22	2.93	3.08	2.59	2.92	2.71
9	Mean	1.80	1.40	1.69	1.42	1.61	1.22	1.25	1.04
10	Median	2.00	1.45	1.89	1.47	1.900	1.42	1.50	1.13
11	IDW	1.45	1.05	1.64	1.27	1.36	0.93	1.35	0.98
12	Areal Weighted Mean	1.65	1.25	1.65	1.31	1.47	1.05	1.37	0.95

Note: Units are in meters. Values in parentheses next to years are the number of surveyed high-water marks. Values in bold font are the smallest errors in each column.

Table 3. Percentage of hits at surveyed HWMs by depth threshold.

		%Hits at all HWMs						%Hits: Ground HWMs Removed					
		2008 (131)			2011 (41)			2008 (73)			2011 (15)		
Model	Calculated Depth	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%
1	2	3	4	5	6	7	8	9	10	11	12	13	14
WRF-Hydro	Reference	49	43	39	46	44	44	37	26	19	7	0	0
	Areal weighted mean	89	79	69	80	66	63	81	63	47	60	20	13
ADHydro	Reference	97	89	83	80	73	71	95	81	70	53	33	27
	Areal weighted mean	100	98	96	98	98	98	100	96	93	93	93	93

Note: As an example, looking at the fifth column, WRF-Hydro predicted a max reference depth that exceeded the 50% threshold of the observed depth at 39% of the 131 surveyed high-water marks.

Considering the within-model results, it would be reasonable to expect that a model's best results would be achieved by the reference simulation at the HWM. For ADHydro, the reference depth (Row 7) did indeed generate the lowest RMSE and MAE errors in three cases, as seen in columns four through six. The IDW sector depth (Row 11) resulted in the lowest RMSE and MAE values in three other cases (columns 3, 7, and 8). For WRF-Hydro, none of the reference depths achieved the lowest RMSE and MAE values. Rather, the lowest WRF-Hydro values in six out of eight columns resulted from the mean sector depth (Row 3). The areal weighted mean depth achieved the second-lowest results (columns 3, 5, and 7–10). Our results in Table 2 suggest that the predictive strength of WRF-Hydro was achieved by mean or areal weighted mean depths rather than reference depths at the specific HWM locations. In contrast, ADHydro's best results were achieved at or near the HWM locations using reference and IDW depths, respectively.

The overall lower values of RMSE and MAE from WRF-Hydro compared to ADHydro in Table 2 are likely the result of two factors. First, because it is limited to a one-way coupling between overland and channel flow, WRF-Hydro frequently underestimated inundation depths. This resulted in lower RMSE and MAE values when observed depths were also shallow. Second, channel capacities in ADHydro were likely underestimated, causing that model to more frequently overpredict inundations depths by an amount greater than what WRF-Hydro underestimated them. This is supported by the general improvement in ADHydro's RMSE and MAE values when surveyed on-ground watermarks are removed (21 out of 24 cases in columns 7–10). Contrary to ADHydro, when the surveyed on-ground watermarks were removed, WRF-Hydro's RMSE and MAE values typically worsened (20 out of 24 cases in columns 7–10). Quantitatively, WRF-Hydro may be producing slightly lower RMSE and MAE values overall, but qualitatively, we believe ADHydro more realistically predicts inundation near the surveyed HWM and provides additional actionable information.

We present further analyses of predicted depths at surveyed HWMs in Table 3. This table shows the percentage of hits each model recorded at the surveyed HWMs. A hit occurs when the model predicts a maximum depth that is greater than the observed depth multiplied by a threshold. Hits are computed using the reference depth and the areal weighted mean depth of the sector containing the surveyed HWM. We prescribed hit categories with depth thresholds of 10%, 30%, and 50% of the observed high water depth. For example, suppose the surveyed observed depth is 1.0 m. At the 10% threshold, a model must compute a depth greater than 1.0 m × 10% or 0.1 m to record a hit. Note that for on-ground HWMs, a hit is counted anytime the flow depth is greater than zero. We interpret Table 3 as follows. The WRF-Hydro value of 39 in column 5 means that at 39% of the 131 surveyed HWMs, WRF-Hydro predicted a maximum depth which was equal to or greater than 50% of the observed high water depth.

Considering first the reference case, the results in Table 3 indicate ADHydro computed the highest number of hits at all depth thresholds. Looking at the case for all HWMs (columns 3–8) for each depth threshold, ADHydro predicted exceedance depths at nearly twice the number of HWMs compared to WRF-Hydro. The differences are greater when only above-ground HMWs are considered (columns 9–14). Similarly, ADHydro areal weighted mean flood depths exceeded the thresholds at a greater number of HWM locations than WRF-Hydro. These results indicate ADHydro properly placed floodwaters at HWM locations more often than WRF-Hydro. ADHydro also computed more hits than WRF-Hydro for uncalibrated simulations of surveyed HWMs (Smith et al. 2020).

Results at Flood-Damage/Flooded-Street Locations

Table 4 presents the analysis of simulated depths at property parcels where damage surveys were conducted, and photos and news outlets reported flooding. We used five threshold depths (<2.54, ≥2.54, ≥5.08, ≥15.24, and ≥30.48 cm) to calculate hit *flooded* percentages for the parcel polygons. We computed the parcel-threshold hits or the percentage of parcels at which the models predicted a maximum depth greater than a threshold depth.

In the case of WRF-Hydro simulations of the 2008 event, 31% of the parcel polygons had a predicted mean areal maximum water depth less than 2.54 cm, and 21% of the parcels had a predicted water depth greater than 30.48 cm. For the 2008 event, ADHydro achieved higher percentages of hits for deeper thresholds compared to WRF-Hydro. For example, 76% of the 280 parcels had a predicted maximum depth greater than 30.48 cm compared to 21% for WRF-Hydro.

For the 2011 event, ADHydro also achieved more hits at the deeper thresholds than WRF-Hydro. Interestingly, both models generated more hits for the <2.54 cm threshold in 2011 than 2008. For WRF-Hydro, 84% of the parcels in 2011 had a computed depth of less than 2.54 cm compared to 31% in 2008. ADHydro generated max depths less than 2.54 cm in 43% of the cases in 2011, compared to only 1% in 2008.

The results in Table 4 indicate WRF-Hydro tended to generate many instances of minimal flood depths (<2.54 cm) in 2008 and 2011 at locations where flooding was observed. Such large numbers of minimal depths are suspect, given the nature of the observations and storm severity. On the other hand, ADHydro computes a greater number of hits at all thresholds deeper than 2.54 cm. Hits at deeper thresholds seem to be realistic given the nature of the flooded streets/flood damage observations.

Fig. 5 shows the spatial distribution of a subset of the 1,144 flooded property parcels for the 2011 storm described in Table 4. The preponderance of black parcels in Fig. 5 shows that WRF-Hydro had many more cases of flood depths less than 2.54 cm compared to ADHydro. The large number of parcel inundation depths less than 2.54 cm for WRF-Hydro seems unrealistic because 7–15 cm of rain fell within a period of 3–4 h.

Table 4. Percentage of hits for different depth thresholds at flood-damaged/flood-damage/flooded-street locations

Model	Threshold depth	%Hits 2008 373 flood reports 280 parcels	%Hits 2011 1,951 flood reports 1,144 parcels
WRF-Hydro	<2.54 cm	31%	84%
	≥2.54 cm	69%	16%
	≥5.08 cm	61%	12%
	≥15.24 cm	44%	4%
	≥30.48 cm	21%	1%
ADHydro	<2.54 cm	1%	43%
	≥2.54 cm	99%	57%
	≥5.08 cm	92%	36%
	≥15.24 cm	84%	21%
	≥30.48 cm	76%	15%

Note: For example, 373 crowd-sourced observations of flooding correspond to 280 distinct property parcels for the 2008 event. At 61% of these 280 property parcel locations in 2008, WRF-Hydro predicted a depth equal to or greater than 5.08 cm.

Discussion

Analysis of predicted maximum depths at surveyed HWMs showed that WRF-Hydro achieved smaller RMSE and MAE values than ADHydro. Analysis of areal sector depths supported this result. As stated previously, this result is likely due to the fact that WRF-Hydro uses one-way coupling between overland and channel flow in contrast to the two-way coupling used in ADHydro. Our results suggest that WRF-Hydro predicts shallower depths than ADHydro at or in the vicinity of the surveyed HWMs, which may lead to smaller RMSE and MAE errors when observed flood depths are shallow. We investigated whether differences in the centroid distance of the element to the stream channel, differences in model element area size, or other topology-related characteristics could help explain the RMSE and MAE results for surveyed HWMs (not shown). We were unable to identify a clear signal that would highlight one model's topology-related advantage over the other, and it is important to note that both models derived their topographic representations from the same 10-m NED DEM. As stated previously, analysis of the NED grid elevations for the areal sectors typically revealed a uniform elevation profile within most areal sectors (Patrick et al. 2018; Appendix II). This does not imply that the 10-m NED DEM shared the same elevation as the ground elevation at the observed surveyed HWM but that both models used the same NED elevation data, and therefore, the relative predicted water depths of the two models should be comparable.

We place our results in light of other whole-city investigations that validated models against numerous surveyed HWMs, recognizing that differences in study contexts preclude a strict comparison of results. Wing et al. (2019) reported RMSE and MAE errors of 1.71 m and 1.03 m, respectively, using 1,123 surveyed HWMs for flooding in Houston caused by Hurricane Harvey. Xing et al. (2019) simulated inundation depths at 368 surveyed HWMs and achieved an RMSE of 0.36 m. Neal et al. (2009) reported RMSE errors of 0.32 and 0.28 m for 263 HWM simulations with and without building representations, respectively. Our reference HWM RMSE values fit within but near the high end of the range of these reported errors. It is highly likely that our results were affected by the use of empirically-derived channel properties rather than surveyed cross-section data. Neal et al. (2009) used numerous channel cross sections, which may have contributed to their low RMSE values. Neither Xing et al. (2019) nor Wing et al. (2019) mention the use of channel cross-section information.

Conclusions and Recommendations

This paper presents the application of novel techniques for the analysis of simulations of high water observations. We compared predicted maximum depths at 172 surveyed high-water marks, 373 locations of flooded structures, and nearly 2,000 observed flooded

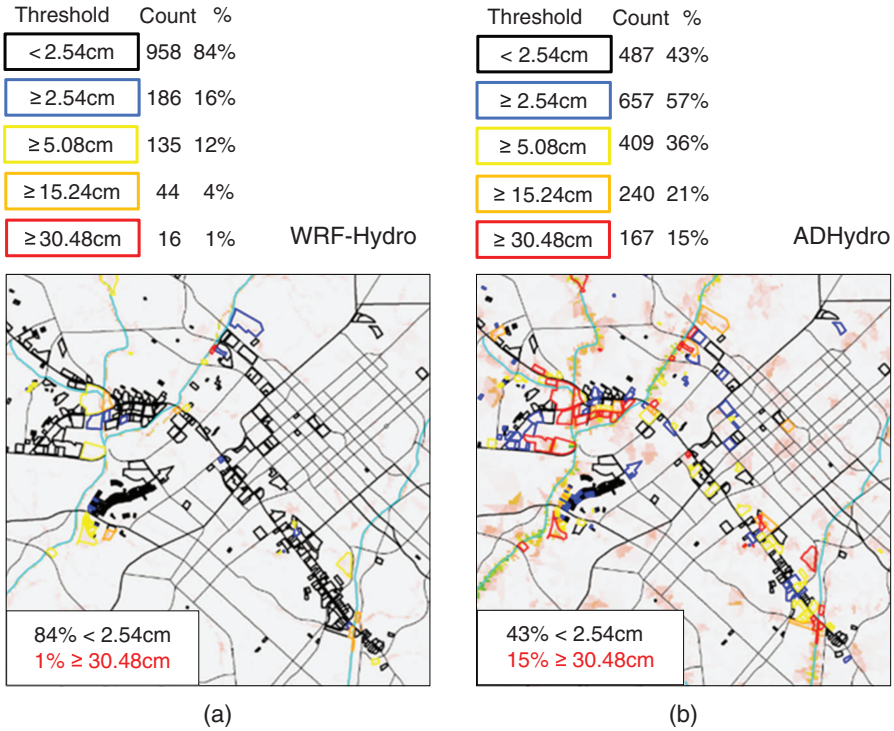

Fig. 5. (Color) Comparison of parcel flood depths for (a) WRF-Hydro; and (b) ADHydro for the 2011 storm event. The downtown area of Charlotte is shown, corresponding to the red ellipse in Fig. 2. The color of the parcel indicates the depth threshold category. For example, in 84% of the parcels, WRF-Hydro computed maximum depths that were less than 2.54 cm (black parcels).

locations to evaluate the models' ability to simulate inundation. In terms of data abundance, our study is among the most comprehensive reported in the literature to date.

Simulation results were somewhat mixed between models, highlighting the need to examine multiple metrics when evaluating models. WRF-Hydro achieved lower values of RMSE and MAE when comparing simulated and surveyed HWM depths, but we surmise that this is attributed to shallower computed water depths when observed depths are also shallow. The marked improvement for ADHydro values of RMSE and MAE when removing on-ground HWMs suggests that WRF-Hydro skews this result in cases of shallow water depths. On the other hand, ADHydro more frequently generated significant inundation when compared to WRF-Hydro for all depth thresholds at surveyed high-water marks. In addition, ADHydro more often predicted flood inundation at locations with observed flood damage and/or street inundation. Thus, we conclude ADHydro properly predicted inundation more often than WRF-Hydro.

Evaluation of simulated inundation depths and extents is complex. Our spatial analyses attempted to account for differences in model discretizations and computational element sizes and to distinguish model performance, assuming that the model predicted flooding in the vicinity of the HWMs. The techniques were predicated on an analysis of NED 10-m grid elevations, which showed minimal topographic variation in most of the areal sectors. Given the data constraints, modeling assumptions, and purpose of the study, we believe the analysis techniques helped distinguish model performance differences and identify model deficiencies. The analysis methods in our study are broadly applicable for validating and intercomparing urban flood inundation models.

Further work is recommended to diagnose the surveyed HWM results. We used highly accurate surveyed HWMs in conjunction with the 10-m NED DEM. Future work could use the surveyed HWM data in conjunction with the 1-m LiDAR DEM available for Charlotte, NC, in the hope of achieving more accurate results (e.g., Neal et al. 2009). Two LiDAR DEM versions are available: 8–9 points/m^2 and 30 points/m^2 (Josh McSwain, CMSWS, personal communication, August 27, 2020).

Both models defined channel geometry using stream-order scaling relationships. Using available surveyed cross-section information would likely have benefited both models. Surveyed cross sections were available for the Sugar Creek basin but not used as we desired to explore model performance using only data sets having national coverage.

Future related studies should be limited to those models having a two-way coupling of overbank and channel flow. ADHydro contained a two-way coupling between the overland and channel routing components. Continued development of WRF-Hydro should include a similar two-way linkage between overland routing and explicit channel routing to allow excess channel flow to move onto overbank areas.

Appendix I. Examples of High-Water Marks

The CMSWS provided three types of observed HWM information for the 2008 and 2011 storm events. First, surveyed HWMs were collected by a local engineering firm. These were classified as riverine. Both above-ground and on-ground HWMs were collected and recorded as absolute elevations above mean sea level (Fig. 6). The accuracy of the surveyed HWMs is stated as 0.0762 m (0.25 ft) vertically and 3.048 m (10 ft) horizontally with a 95% accuracy level. The surveying company described HWM quality as good, fair, or poor. We found no documentation that defines these descriptors of quality. Second, the city of Charlotte, NC, conducted flood damage surveys in which flood depths were measured relative to

Fig. 6. (Color) Examples of (a) surveyed above-ground; and (b) surveyed on-ground high watermarks. (Images courtesy of Charlotte Mecklenburg Storm Water Services.)

Fig. 7. (Color) Example of flood damage survey for the 2008 flood of record. Damage depths were recorded relative within a structure and are not referenced to a datum. (Image courtesy of Charlotte-Mecklenburg Storm Water Services.)

structures. Fig. 7 shows such a measurement. Finally, locations of flooding from news reports, citizen reports, and crowd-sourced photos were collected and geo-encoded. Fig. 8 shows a typical example.

Fig. 8. (Color) Example of flooded street high water location. (Image courtesy of Charlotte-Mecklenburg Storm Water Services.)

Appendix II. Analysis of Areal Sector NED Elevations

We analyzed the NED 10-m elevations within the areal sectors for each of the surveyed HWMs in 2011 (131) and 2008 (41). Fig. 9 shows an example analysis using Brier Creek HWM number six (BRI-06). The surveyed ground elevation at the HWM location is 190.91 m, while the elevation of the NED grid cell containing

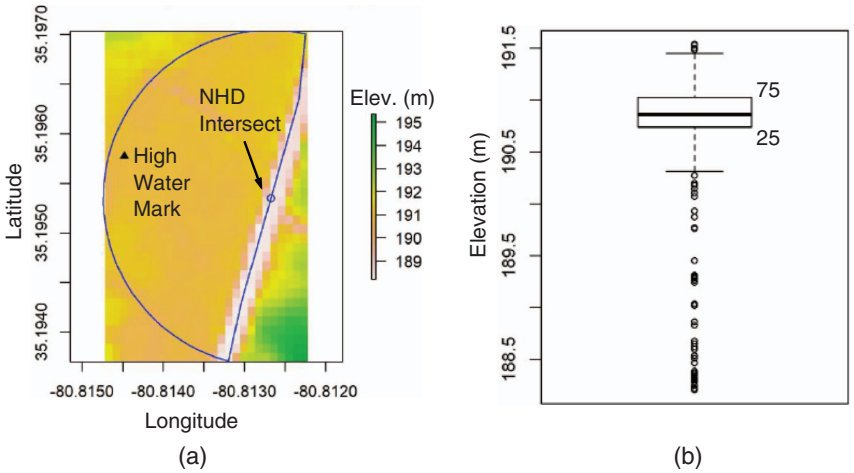

Fig. 9. (Color) Example analysis of NED 10-m elevations in the areal sector for the BRI-06 HWM. (a) The spatial variation of NED 10-m grid elevations and location of the HWM and NHD intersect. (b) Box plot of NED 10-m elevations. The heavy line in the box is the median. The lower and upper borders of the box represent the 25th and 75th percentiles, respectively.

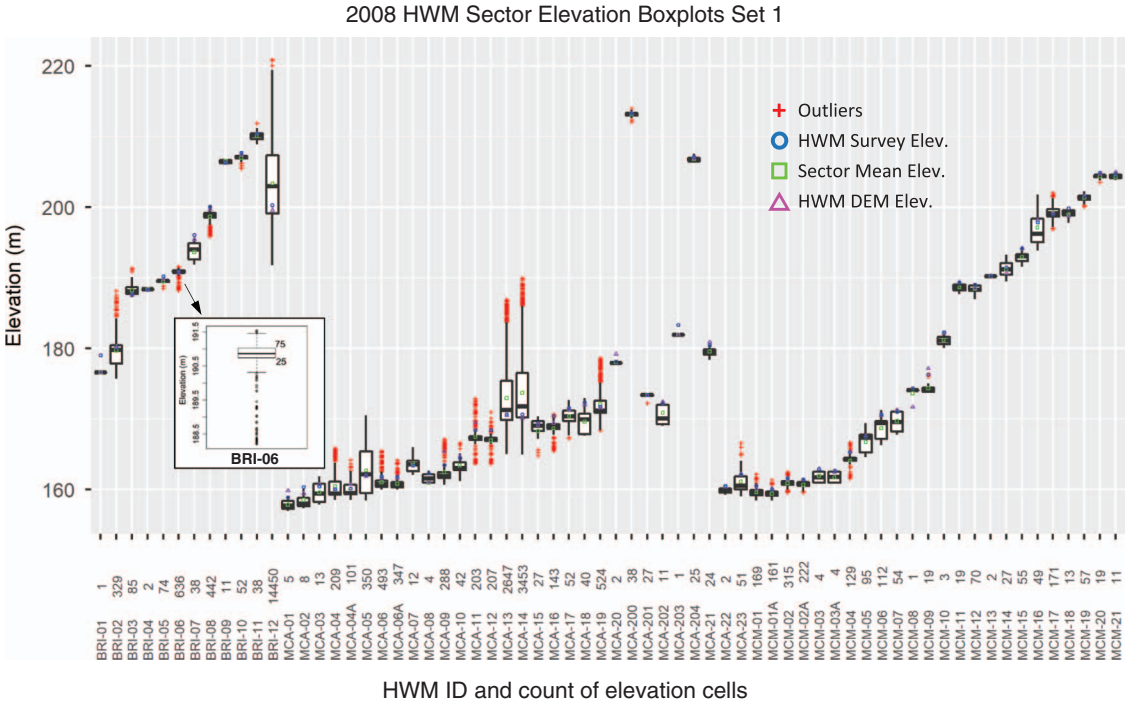

Fig. 10. (Color) Box plots of areal sector elevations for HWM Set 1 in 2008. The *X*-axis labels are the HWM identifier and the number of NED 10-m grid cells in the sector. The inset shows the box plot for BRI-06 from Fig. 9(b).

the HWM is 190.71 m. The areal mean elevation of the 636 NED grid cells in this sector is 190.78. The elevation of the NED grid cell at the NHD+ stream intersect is 188.4 m and most likely reflects a grid cell in the channel. The uniform color in the sector away from the NHD+ stream vector in Fig. 9(a) suggests minimal elevation variation among the 636 NED grid cells in this sector.

The box plot in Fig. 9(b) depicts the distribution of the 636 grid cell elevations in the areal sector for BRI-06. The lower border of the box is the 25th percentile, while the top border of the box represents the 75th percentile. The median elevation is approximately 190.80 m. The whiskers represent 1.5 times the interquartile range (IQR). The IQR (50% of the values) is quite narrow and ranges from ~190.6 m to ~191.1 m. Values below the lower whisker represent elevation outliers. Based on Fig. 9(a), these outliers are probably cells within the stream channel derived from the DEM. The box plot suggests an approximate homogenous elevation outside of the immediate stream channel in the areal sector.

Figs. 10–12 present box plots of the NED 10-m cell elevations in the areal sectors for all the surveyed HMWs. The box plots for the 131 HWMs in 2008 are displayed in two sets. Similar to BRI-06, the box plots indicate that for most of the locations, the topography in the areal sectors (assuming one is up and out of the channel) is generally homogenous, and the elevation is consistent.

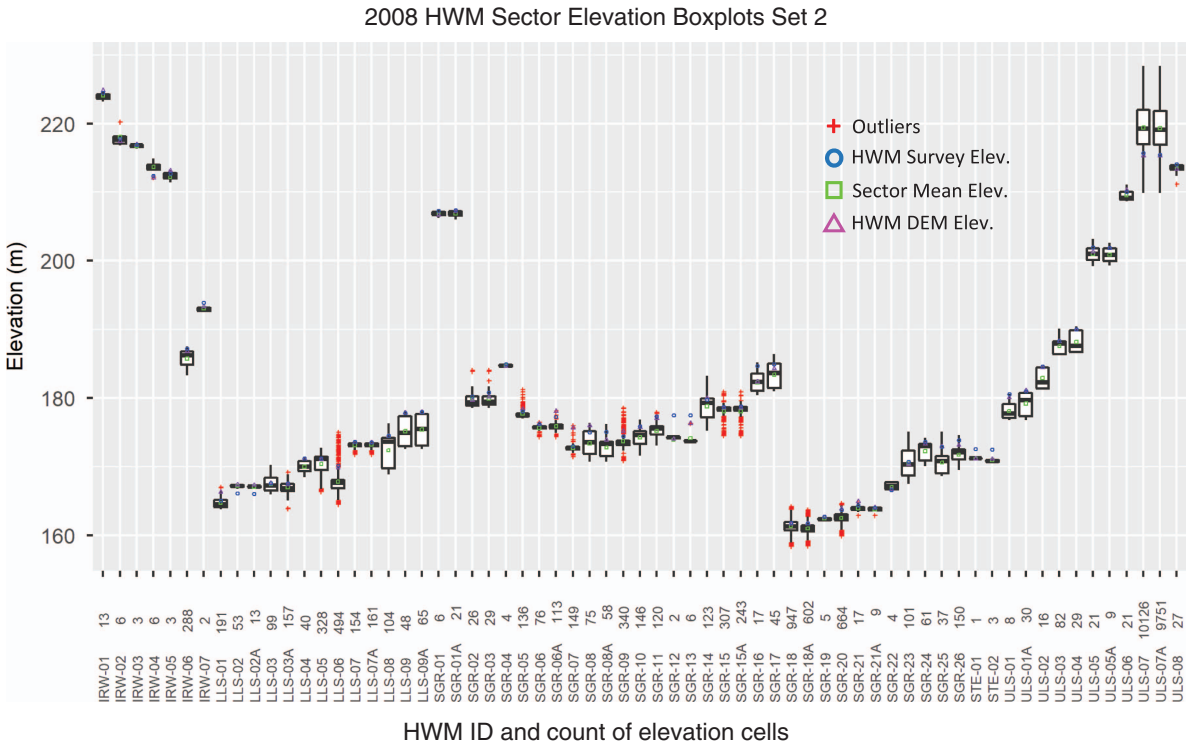

Fig. 11. (Color) Box plots of areal sector elevations for HWM Set 2 in 2008. The *X*-axis labels are the HWM identifier and the number of NED 10-m grid cells in the areal sector.

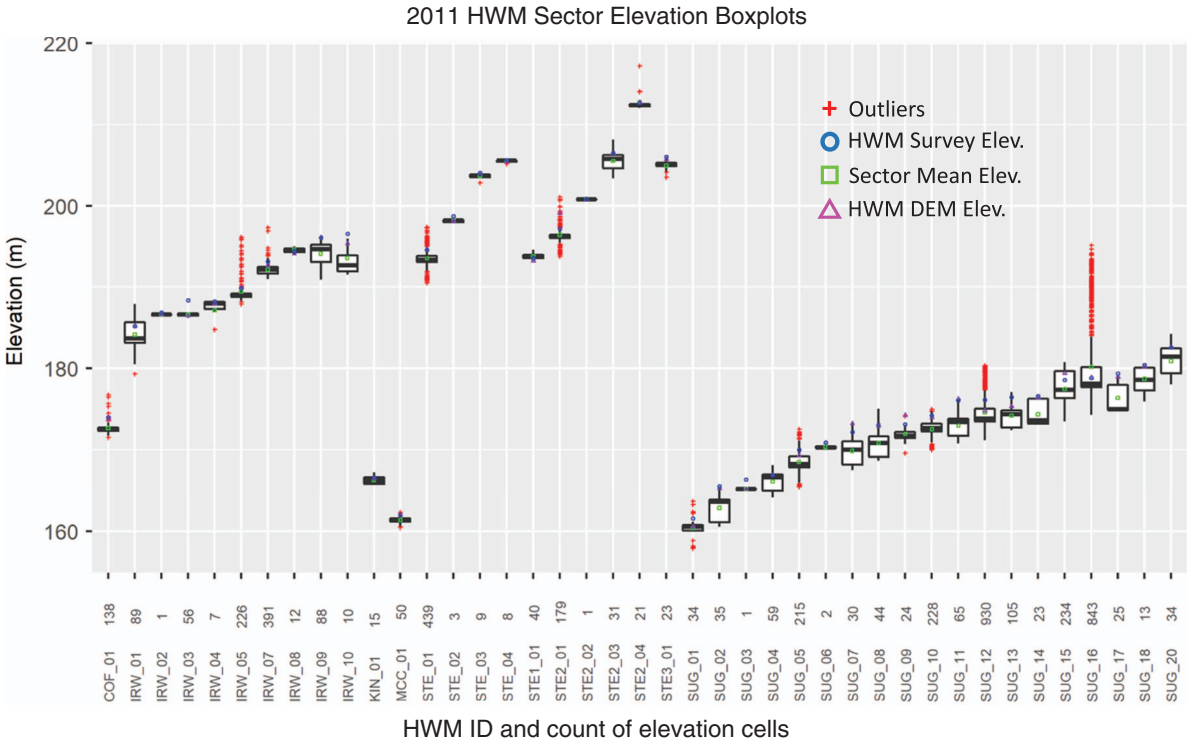

Fig. 12. (Color) Box plots of areal sector elevations for the 2011 HWMs. The *X*-axis labels are the HWM identifier and the number of NED 10-m grid cells in the sector.

Table 5 presents mean values of the various elevations in the areal sectors for 2008 and 2011. The results in Table 5 suggest an overall good agreement between the ground elevation at the surveyed HWM and the elevation of the corresponding NED 10-m grid cell. Taken together, the box plots and results in Table 5 indicate that using an areal surface water depth to compare model results may be appropriate as there should not be too many variations in the ground surface elevation once out of the channel.

Table 5. Mean elevations of areal sectors and high-water marks for 2008 and 2011

Type of elevation value	Mean elevation (m) 2008	2011
Surveyed ground elevation at HWM	180.45	183.90
NED 10-m grid elevation at HWM	180.30	183.71
Areal sector mean elevation	179.72	182.72
NED 10-m grid elevation at NHD+ intersect	179.10	181.81

Data Availability Statement

Some or all data, models, or code used during the study were provided by a third party. Direct request for these materials may be made to the provider, as indicated in the Acknowledgments.

Acknowledgments

We thank Trey Flowers of the National Water Center for his leadership throughout this project. Fred Ogden, also of the National Water Center, is acknowledged for helping to design the study and providing the ADHydro model code. Bob Steinke is recognized for his work on ADHydro code development. Dave Gochis, Laura Read, Logan Karsten, and Kevin Sampson of NCAR are gratefully acknowledged for their WRF-Hydro contributions. We thank J. J. Gourley and the staff of the NOAA National Severe Storms Laboratory for their help in acquiring MRMS data. Iowa State University is recognized for providing additional MRMS data. Jerald Robinson, chief of the Charlotte Field Office of the USGS South Atlantic Water Science Center, provided helpful guidance in interpreting the USGS observations for several events. We especially thank Joshua McSwain and his staff of the Charlotte-Mecklenburg County Storm Water Services for providing their extensive records of high-water marks and for granting permission to use photographs from their HWM archives. This work was funded by appropriations from the US Congress. This research of the NWS Office of Water Prediction (OWP) is supported by NWS Science and Technology Integration (STI)–appropriated funds applied to NOAA's Science Collaboration Program and administered by UCAR's Cooperative Programs for the Advancement of Earth System Science (CPAESS) under Awards NA16NWS4620043 and NA18NWS4620043B and Cyberdata Technologies, Inc., Contract # EG133W16NC1399.

References

Ali, A. M., G. D. Baldassarre, and D. P. Solomatine. 2015. "Testing different cross-section spacing in 1D hydraulic modelling: A case study on Johor River, Malaysia." *Hydrol. Sci. J.* 60 (2): 351–360. https://doi.org/10.1080/02626667.2014.889297.

Blumberg, A. F., N. Georgas, L. Yin, T. O. Herrington, and P. M. Orton. 2015. "Street-scale modeling of storm surge inundation along the New Jersey Hudson River Waterfront." *J. Atmos. Oceanic Technol.* 32 (8): 1486–1497. https://doi.org/10.1175/JTECH-D-14-00213.1.

Cook, A., and V. Merwade. 2009. "Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping." *J. Hydrol.* 377 (1–2): 131–142. https://doi.org/10.1016/j.jhydrol.2009.08.015.

Dottori, F., G. Di Baldassare, and E. Todini. 2013. "Detailed data is welcome, but with a pinch of salt: Accuracy, precision, and uncertainty in flood inundation modeling." *Water Resour. Res.* 49 (9): 6079–6085. https://doi.org/10.1002/wrcr.20406.

Fewtrell, T. J., J. C. Neal, P. D. Bates, and P. J. Harrison. 2011. "Geometric and structural river channel complexity and the prediction of urban inundation." *Hydrol. Processes* 25 (20): 3173–3186. https://doi.org/10.1002/hyp.8035.

FINS (Flood Information and Notification System). 2017. "Monitors real-time rain and stream gage levels." Accessed December 9, 2017. http://finslive.mecklenburgcountync.gov/finslive/?gauge=rain.=PT24H.

Fohringer, J., D. Dransch, H. Kreibich, and K. Schroter. 2015. "Social media as an information source for rapid flood inundation mapping." *Nat. Hazards Earth Syst. Sci. Discuss.* 3 (3): 4231–4264. https://doi.org/10.5194/nhessd-3-4231-2015.

Gallegos, H. A., J. E. Schubert, and B. F. Sanders. 2009. "Two-dimensional, high-resolution modeling of urban dam-break flooding: A case study of Baldwin Hills, California." *Adv. Water Resour.* 32 (8): 1323–1335. https://doi.org/10.1016/j.advwatres.2009.05.008.

Gochis, D. J., et al. 2020. "The WRF-Hydro® modeling system technical description (version 5.1.1)." Accessed July 1, 2020. https://ral.ucar.edu/sites/default/files/public/WRFHydroV511TechnicalDescription.pdf.

Grimley, L., M. Khanam, E. Tiernan, and D. Tijerina. 2017. Hyper-resolution modeling in urban landscapes." Chap. 3 in *National Water Center Innovators Program Summer Institute Report. Consortium of Universities for the Advancement of Hydrologic Science, Inc. Technical Report No. 14*, 23–31. https://doi.org/10.4211/technical.20171009.

Habibi, H., and D.-J. Seo. 2018. "Simple and modular integrated modeling of storm drain network with gridded distributed hydrologic model via grid-rendering of storm drains for large urban areas." *J. Hydrol.* 567 (Dec): 637–653. https://doi.org/10.1016/j.jhydrol.2018.10.037.

Hartnet, M., and S. Nash. 2017. "High-resolution flood modeling of urban areas using MSN_Flood." *Water Sci. Eng.* 10 (3): 175–183. https://doi.org/10.1016/J.wse.2017.10.003.

Horritt, M. S., P. D. Bates, T. J. Fewtrell, D. C. Mason, and M. D. Wilson. 2010. "Modelling the hydraulics of the Carlisle 2005 flood event." *Proc. Inst. Civ. Eng. Water Manage.* 163 (6): 273–281. https://doi.org/10.1680/wama.2010.163.6.273.

Hunter, N. M., P. D. Bates, M. S. Horritt, P. J. De Roo, and M. G. F. Werner. 2005. "Utility of different data types for calibrating flood inundation models with a GLUE framework." *Hydrol. Earth Syst. Sci.* 9 (4): 412–430. https://doi.org/10.5194/hess-9-412-2005.

Kim, S., H. Shen, S. Noh, D.-H. Seo, E. Welles, E. Pelgrim, A. Weerts, E. Lyons, and B. Philips. 2021. "High-resolution modeling and prediction of urban floods using WRF-Hydro and data assimilation." *J. Hydrol.* 598 (Jul): 126236. https://doi.org/10.1016/j.jhydrol.2021.126236.

Kitzmiller, D. H., W. Wu, Y. Zhang, D. A. Miller, and Z. Zhang. 2014. "Multi-decade analysis of record for hydrologic model calibration." In *Proc., AMS Annual Meeting: Paper 5.2*. Boston: American Meteorological Society.

Kutija, V., R. Bertsch, V. Glenis, D. Alderson, G. Parkin, C. Walsh, J. Robinson, and C. Kilsby. 2014. "Model validation using crowd-sourced data from a large pluvial flood." In *Proc., 11th Int. Conf. on Hydroinformatics*, New York: City University of New York.

Lai, W., F. I. Ogden, R. C. Steinke, and C. A. Talbot. 2015. "An efficient and guaranteed stable numerical method for continuous modeling of infiltration and redistribution with a shallow dynamic water table." *Water Resour. Res.* 51 (3): 1514–1528. https://doi.org/10.1002/2014WR016487.

Leitao, J. P., N. E. Simoes, C. Maksimovic, F. Ferreira, D. Prodanovic, J. S. Matos, and A. Sa Marques. 2010. "Real-time forecasting urban drainage models: Full or simplified networks?" *Water Sci. Technol.* 62 (9): 2106–2114. https://doi.org/10.2166/wst.2010.382.

Macchione, F., P. Costabile, C. Costanzo, and G. De Lorenzo. 2019. "Extracting quantitative data from non-conventional information for the hydraulic reconstruction of past urban flood events. A case study." *J. Hydrol.* 576 (Sep): 443–465. https://doi.org/10.1016/j.jhydrol.2019.06.031.

McDougall, K., and P. Temple-Watts. 2012. "The use of LiDAR and volunteered geographic information to map flood extents and inundation." In Vol. I-4 of *Proc., ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences: XXII ISPRS Congress*, 251–256. Melbourne, Australia: International Society of Photogrammetry and Remote Sensing.

Mignot, E., A. Paquier, and S. Haider. 2006. "Modeling floods in a dense urban area using 2D shallow water equations." *J. Hydrol.* 327 (1–2): 186–199. https://doi.org/10.1016/j.jhydrol.2005.11.026.

Moore, R. B., L. D. McKay, A. H. Rea, T. R. Bondelid, C. V. Price, and T. G. Dewald. 2019. *User's guide for the national hydrography dataset plus (NHDPlus) high resolution (USGS Numbered Series No. 2019–1096), (NHDPlus) high resolution*. Open-File Rep. Reston, VA: USGS.

Neal, J., I. Villanueva, N. Wright, T. Willis, T. Fewtrell, and P. Bates. 2012. "How much physical complexity is needed to model flood inundation?" *Hydrol. Processes* 26 (15): 2264–2282. https://doi.org/10.1002/hyp.8339.

Neal, J. C., P. D. Bates, T. J. Fewtrell, N. M. Hunter, M. D. Wilson, and M. S. Horritt. 2009. "Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations." *J. Hydrol.* 368 (1–4): 42–55. https://doi.org/10.1016/j.jhydrol.2009.01.026.

Nguyen, P., A. Thorstensen, S. Sorooshian, K. Hsu, A. AghaKouchak, B. Sanders, V. Koren, Z. Cui, and M. Smith. 2016. "A high resolution coupled hydrologic-hydraulic model (Hi ResFlood-UCI) for flash flood modeling." *J. Hydrol.* 541 (Part A): 401–420. https://doi.org/10.1016/j.jhydrol.2015.10.047.

Niu, G. Y., et al. 2011. "The community Noah land surface model with multi-parameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements." *J. Geophys. Res. Atmos.* 116 (12): D12109. https://doi.org/10.1029/2010JD015139.

Noh, S. J., J.-H. Lee, S. Lee, and D.-J. Seo. 2019. "Retrospective dynamic inundation mapping of Hurricane Harvey flooding in the Houston metropolitan area using high-resolution modeling and high-performance computing." *Water* 11 (3): 597. https://doi.org/10.3390/w11030597.

Ogden, F. L., M. B. Allen, J. Zhu, W. Lai, C. C. Douglas, M. Seo, and C. A. Talbot. 2017. "The soil moisture velocity equation." *J. Adv. Model. Earth Syst.* 9 (2): 1473–1487. https://doi.org/10.1002/2017MS000931.

Ogden, F. L., W. Lai, and R. C. Steinke. 2015. "ADHydro: Quasi-3D high performance hydrological model." In *Proc., 3rd Joint Federal Interagency Conf. (10th Federal Interagency Sedimentation Conf. and 5th Federal Interagency Hydrologic Modeling Conf.)*, 342–350. Washington, DC: Advisory Committee on Water Information.

Ogden, F. L., N. R. Pradhan, C. W. Downer, and J. A. Zahner. 2011. "Relative importance of impervious area, drainage density, width function, and subsurface storm drainage on flood runoff from an urbanized catchment." *Water Resour. Res.* 47 (12): W12503. https://doi.org/10.1029/2011WR010550.

Ozdemir, H., C. C. Sampson, G. A. M. de Almeida, and P. D. Bates. 2013. "Evaluating scale and roughness effects in urban flood modelling using terrestrial LiDAR data." *Hydrol. Earth Syst. Sci.* 17 (10): 4015–4030. https://doi.org/10.5194/hess-17-4015-2013.

Patrick, N., M. B. Smith, F. L. Ogden, J. Kim, and N. Frazier. 2018. "Hyper-resolution flood inundation modeling: Use of surveyed high water marks in the evaluation of hyper-resolution hydrologic models." In *Proc., Poster H41P-2351 Presented at the Fall AGU Meeting*. Washington, DC: American Geophysical Union.

Rafieeinasab, A., A. Norouzi, S. Kim, H. Habibi, B. Nazari, D.-J. Seo, H. Lee, B. Cosgrove, and Z. Cui. 2015. "Toward high-resolution flash flood prediction in large urban areas—Analysis of sensitivity to spatio-temporal resolution of rainfall input and hydrologic modeling." *J. Hydrol.* 531 (Part 2): 370–388. https://doi.org/10.1016/j.jhydrol.2015.08.045.

Savage, J. T. S., P. Bates, J. Freer, J. Neal, and G. Aronica. 2016. "When does spatial resolution become spurious in probabilistic flood inundation predictions?" *Hydrol. Processes* 30 (13): 2014–2032. https://doi.org/10.1002/hyp.10749.

Schubert, J. E., and B. F. Sanders. 2012. "Building treatments for urban flood inundation models and implications for predictive skill and modeling efficiency." *Adv. Water Resour.* 41 (Jun): 49–64. https://doi.org/10.1016/j.advwatres.2012.02.012.

Schumann, G. J.-P., J. C. Neal, D. C. Mason, and P. D. Bates. 2011. "The accuracy of sequential aerial photography and SAR data for observing urban flood dynamics, a case study of the UK summer 2007 floods." *Remote Sens. Environ.* 115 (10): 2536–2546. https://doi.org/10.1016/j.rse.2011.04.039.

Smith, J. A., M. L. Baeck, J. E. Morrison, and P. Sturdevant-Rees. 2002. "The regional hydrology of extreme floods in an urbanizing drainage basin." *J. Hydrometeorol.* 3 (3): 267–282. https://doi.org/10.1175/1525-7541(2002)003<0267:TRHOEF>2.0.CO;2.

Smith, L. S., Q. Liang, and P. F. Quinn. 2015. "Towards a hydrodynamic modelling framework appropriate for applications in urban flood assessment and mitigation using heterogeneous computing." *Urban Water J.* 12 (1): 67–78. https://doi.org/10.1080/1573062X.2014.938763.

Smith, M., N. Patrick, N. Frazier, J. Kim, T. Flowers, and F. Ogden. 2020. *Hyper-resolution modeling of urban flood inundation*. NOAA Technical Report NWS 56. Silver Spring, MD: National Oceanic and Atmospheric Administration. https://doi.org/10.25923/9t55-tn77.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture 2021. "Web soil survey." Accessed December 1, 2017. https://websoilsurvey.nrcs.usda.gov/.

Stephens, E., G. Schumann, and P. Bates. 2014. "Problems with binary pattern measures for flood model evaluation." *Hydrol. Processes* 28 (18): 4928–4937. https://doi.org/10.1002/hyp.9979.

USGS. 2017. "1/3 arc second DEM." Accessed December 4, 2020. https://www.usgs.gov/core-science-systems/national-geospatial-program/national-map.

Veldhuis, M. T., Z. Zhou, L. Yang, S. Liu, and J. Smith. 2018. "The role of storm scale, position, and movement in controlling urban flood response." *Hydrol. Earth Syst. Sci.* 22 (1): 417–436. https://doi.org/10.5194/hess-22-417-2018.

Villarini, G., J. A. Smith, M. L. Baeck, P. Sturdevant-Rees, and W. F. Krajewski. 2010. "Radar analyses of extreme rainfall and flooding in urban drainage basins." *J. Hydrol.* 381 (3–4): 266–286. https://doi.org/10.1016/j.jhydrol.2009.11.048.

Werner, M., S. Blazkova, and J. Petr. 2005. "Spatially distributed observations in constraining inundation modeling uncertainties." *Hydrol. Processes* 19 (16): 3081–3096. https://doi.org/10.1002/hyp.5833.

Wickham, J., S. V. Stehman, L. Gass, J. A. Dewitz, D. G. Sorenson, B. J. Granneman, R. V. Poss, and L. A. Baer. 2017. "The accuracy assessment of the 2011 National Land Cover Database (NLCD)." *Remote Sens. Environ.* 191: 328–341. https://doi.org/10.1016/j.rse.2016.12.026.

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." *J. Hydrol. X* 4 (Jul): 100039. https://doi.org/10.1016/j.hydroa.2019.100039.

Wright, D. B., J. A. Smith, G. Villarini, and M. L. Baeck. 2014. "Long-term high-resolution radar rainfall fields for urban hydrology." *J. Am. Water Resour. Assoc.* 50 (3): 713–734. https://doi.org/10.1111/jawr.12139.

Xing, Y., Q. Liang, G. Wang, X. Ming, and X. Xia. 2019. "City-scale hydrodynamic modelling of urban flash floods: The issues of scale and resolution." *Nat. Hazards* 96 (1): 473–496. https://doi.org/10.1007/s11069-018-3553-z.

Yang, Z.-L., et al. 2011. "The community Noah land surface model with multi-parameterization options (Noah-MP): 2. Evaluation over global river basins." *J. Geophys. Res. Atmos.* 116 (12): 12110. https://doi.org/10.1029/2010JD015140.

Yu, D., and S. N. Lane. 2006. "Urban fluvial flood modelling using a two-dimensional diffusion-wave treatment, part 1: Mesh resolution effects." *Hydrol. Processes* 20 (7): 1541–1565. https://doi.org/10.1002/hyp.5935.

Yu, D., J. Yin, and M. Liu. 2016. "Validating city-scale surface water flood modelling using crowdsourced data." *Environ. Res. Lett.* 11 (12): 124011. https://doi.org/10.1088/1748-9326/11/12/124011.

Zhang, J., et al. 2016. "Multi-radar multi-sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities." *Bull. Am. Meteorol. Soc.* 97 (4): 621–638. https://doi.org/10.1175/BAMS-D-14-00174.1.

Zhou, Z., J. A. Smith, L. Yang, M. L. Baeck, M. Chaney, M.-C. Ten Veldhuis, H. Deng, and S. Liu. 2017. "The complexities of urban flood response: Flood frequency analyses for the Charlotte metropolitan region." *Water Resour. Res.* 53 (8): 7401–7425. https://doi.org/10.1002/2016WR019997.