University of Bristol - Explore Bristol Research
General rights
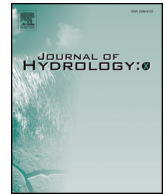
Research papers

A flood inundation forecast of Hurricane Harvey using a continental-scale 2D hydrodynamic model

Oliver E.J. Wing[a,b,*], Christopher C. Sampson[b], Paul D. Bates[a,b], Niall Quinn[b], Andrew M. Smith[b], Jeffrey C. Neal[a,b]

[a] School of Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS, UK
[b] Fathom, Engine Shed, Station Approach, Temple Meads, Bristol BS1 6QH, UK

ABSTRACT

Forecasts of tropical cyclones have seen rapid improvements in recent years as expanding computational capacity permits more runs of finer resolution meteorological models with increasing representation of physical processes. However, the utilization of a hydrodynamic component in these models is often neglected, meaning flood forecasts typically output point water levels that give little indication of a projected inundation extent on the ground. Here, we append this critical component to the forecast cascade by coupling Fathom-US, a continental-scale hydraulic model which employs the LISFLOOD-FP numerical scheme, to forecasts of streamflow, rainfall and coastal surge height from the National Oceanic and Atmospheric Administration (NOAA). Medium-term (2–15 days) flood inundation forecasts, as well as hindcasts driven by real-time observations, were executed for Hurricane Harvey by rapidly simulating pluvial and coastal flood hazard and extracting fluvial flood maps from an existing US-wide simulation library. The resultant ~30 m resolution depth grids were then validated against post-event observations collated by the US Geological Survey. Across the disaster zone, the hindcast (forecast) model captured, on average, 78% (75%) of the benchmark flood extent, obtained a Critical Success Index of 0.66 (0.57) and deviated from observed high water marks by ~1 m (~1.2 m). When compared to a simpler GIS-based approach, the hydraulic model exhibited much higher skill in replicating observations. This study shows that fully hydrodynamic approaches can be practically employed in large-scale forecast frameworks at high resolution to produce skillful projections of inundation extent without significantly affecting the forecast lead time.

1. Introduction

Flood events are among the most costly and deadly natural disasters on the planet: since 1980, they have caused economic damages of over $1 trillion and 220,000 deaths worldwide (Munich Re, 2018a). Recent devastating events, particularly flooding (both freshwater and saltwater) arising from tropical cyclones, have sharply focused the issue in the minds of the public and policy makers alike. In 2017, Hurricanes Harvey, Irma and Maria collectively caused $220 billion of damage in the Gulf of Mexico; Typhoon Haiyan claimed the lives of over 6000 people and caused $10 billion worth of damage in East Asia in 2013; and in 2008, Cyclone Nargis caused 140,000 fatalities and $4 billion of economic damage in Myanmar (Munich Re, 2018b). A growing body of evidence suggests that such tropical cyclones will become more intense (Kang and Elsner, 2016; Sobel et al., 2016; van Oldenborgh et al., 2017; Emanuel, 2017) and move more slowly once they make landfall

(Kossin, 2018) as a result of climate change. With more precipitation falling over a longer duration, tropical cyclone driven flood impacts are likely to increase in the future. On top of the freshwater component of such events, the low atmospheric pressure and strong onshore winds arising from tropical cyclones result in coastal inundation. Regardless of potential changes to these storm characteristics under climate change, increased sea levels in a warming world are likely to exacerbate coastal flood impacts.

In light of this, there is a clear need for substantial risk reduction measures to mitigate against present and future flood consequences. One facet of such measures is improved flood forecasting, which permits a short-term response to be mounted (e.g. temporary defense erection, evacuations, first-responder preparedness, reinsurance purchasing). A typical generic flood forecasting approach can be conceptualized as a source-pathway-receptor framework (e.g. Narayan et al., 2012) as outlined in Fig. 1.
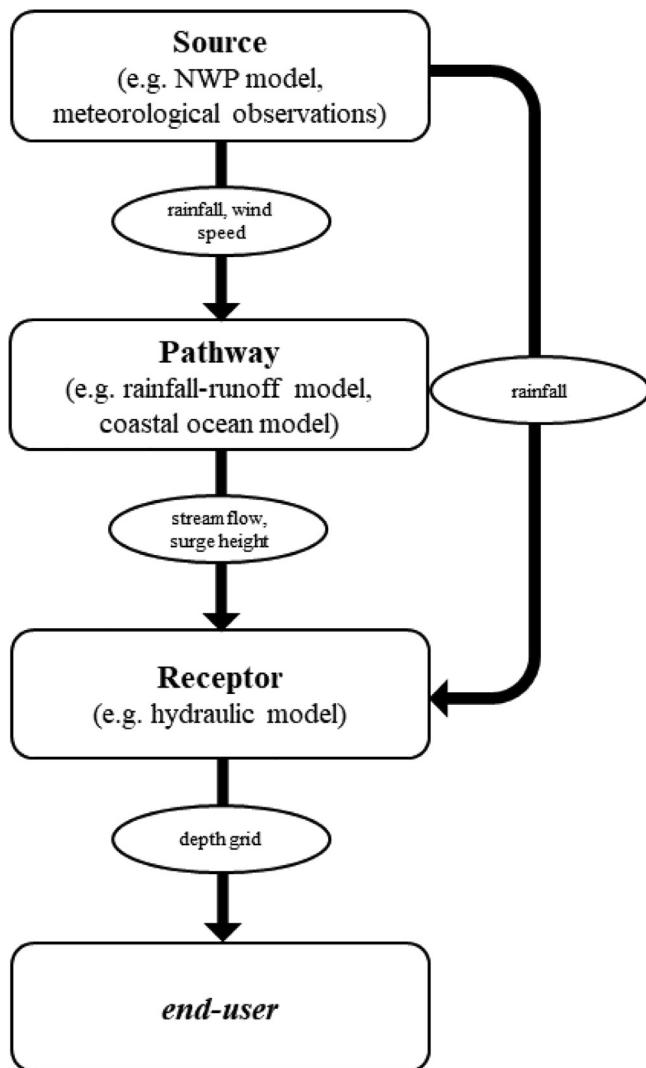
Fig. 1. Schematic of a source-pathway-receptor flood forecasting cascade. Arrows indicate the flow of data (with examples in ovals), where the outputs of one component (specified in rectangular boxes) become the inputs of another.

Direct meteorological and hydrological observations (e.g. of rainfall from gauges or radar, of wind speed from anemometers or radar, or of river flows from an upstream gauge) can form the source data of the forecast cascade if spatial coverage is sufficient, but such information is often unavailable or generates forecasts with short lead times (a few hours to a maximum of perhaps 1–2 days in very large basins), limiting their usefulness. To increase lead times, medium-term forecasts (2–15 days) use numerical weather prediction (NWP) models as the primary input data source to the model cascade, and NWP systems have benefitted from the rapid advances in computational capacity seen in recent years. High performance computing (HPC) resources have had the dual effect of permitting NWP models to run at increasingly fine spatial and temporal resolutions so that atmospheric dynamics can be more accurately represented (Buizza et al., 1999), and allowing multiple NWP simulations in an ensemble to account for underlying model uncertainties (Cloke and Pappenberger, 2009).

Pathway models translate the source-generated meteorological variables (e.g. precipitation, wind speed) to water flows (e.g. streamflow, coastal surge height), except where the rainfall output from the source forms a direct input to a hydraulic model (i.e. for pluvial flood events). For riverine flood events, a hydrological model takes meteorological inputs and computes, with varying levels of physical complexity, water fluxes at the land surface based on soils, land cover and topography to generate streamflow. As examples, the European Commission Joint Research Centre (JRC) LISFLOOD (van der Knijff et al., 2010) and the European Centre for Medium-Range Weather Forecasting (ECMWF) HTESSEL (Balsamo et al., 2011) models are used in the JRC-ECMWF forecast product GloFAS (Alfieri et al., 2013), while the National Oceanic and Atmospheric Administration (NOAA) National Water Model is driven by WRF-Hydro (Gochis et al., 2018) to forecast river discharge in the US. For coastal storm surges, forecast wind fields are the driving forces in models describing fluid motion in the ocean which simulate surge height at given coastal locations (e.g. the NOAA SLOSH model (Jelesnianski et al., 1992)).

The source and pathway components of the forecast framework have received much attention in the literature (Cloke and Pappenberger, 2009; Thielen et al., 2009; Pappenberger et al., 2010; Alfieri et al., 2012, 2013) and form the products of the world's leading forecast centers – NOAA and ECMWF – but a receptor model is an often-neglected component of the forecast cascade. A receptor model in the framework described here translates input water flows from pathway or source models (e.g. streamflow, rainfall or coastal surge height) and translates them to a 2D grid of flood depths using a hydraulic model. If a receptor model is used at all (many forecasts are point water or flows levels only), such products are not operational and only focus on a single peril (Pappenberger et al., 2005; Schumann et al., 2013); simulate over small spatial scales (Addor et al., 2011; Nguyen et al., 2016); require significant manual set-up and have demanding data requirements (Sanders et al., 2010; Bhola et al., 2018; Adams et al., 2018); or employ simplified representations of hydraulic processes to reduce computational costs (Paiva et al., 2013; Zheng et al., 2018a). This is predominantly because most of the computational time available is afforded to the NWP models, maximizing resolution and producing probabilistic ensemble simulations (Cloke and Pappenberger, 2009), alongside the prevailing view that full-physics hydraulic models are too computationally expensive to be used in operational forecasts (Leskens et al., 2014; Bhola et al., 2018). Yet, tropical cyclones demand forecasts of multiple flood drivers and end-users would greatly benefit from detailed, local predictions of flood extent and depth to enable a more complete risk calculation.

Official riverine flood forecasts in the US are issued by the NOAA National Weather Service (NWS) through Weather Forecast Offices. These forecasts are generated by River Forecast Centers (RFCs), providing accurate information to inform public alerts and warnings (for more information on operational practice, see Adams, 2016). These RFC forecasts are produced at particular points, with forecast information currently available at 3697 points across the contiguous US according to the NOAA NWS Advanced Hydrologic Prediction Service (AHPS; https://water.weather.gov/ahps/forecasts.php). Of these, only 155 have accompanying inundation maps (i.e. adopt the receptor component in Fig. 1). Adams et al. (2018) illustrate state-of-the-art practice at the Ohio RFC, where unsteady-state 1D HEC-RAS models rapidly translate forecast point discharge to inundation maps for 3200 km of continuous river reach. Reported errors in predicted stage are < 0.5 m. Furthermore, Mashriqui et al. (2014) apply a similar approach in the Middle Atlantic RFC, coupling 1D HEC-RAS to a tidal boundary at the mouth of the Potomac River. Reported accuracies were similar to Adams et al. (2018). While these 1D approaches provide rapid and accurate riverine forecasts, they require channel cross-section data which are only sparsely available and considerable manual set-up by skilled practitioners. Furthermore, their focus only on large-river flooding means the significant pluvial hazard posed by tropical cyclones remained unmodelled.

Here we present a medium-term (2–15 days) tropical cyclone flood inundation forecasting product which is capable of being used in an operational system, and demonstrate that high-resolution hydraulic models can now be practically employed at large-scale in such frameworks where accurate local forecasts are lacking. We forecasted Hurricane Harvey by coupling streamflow, rainfall and storm surge

predictions from NOAA to Fathom-US (Wing et al., 2017), a continental-scale hydraulic flood model of the US, and used this to produce daily flood depth footprints at ~30 m resolution. Updating the hydrologic inputs with post-event observations as they became available, we also produced a model hindcast in real-time. After the event, we tested the forecast and hindcast model against ground observations and derived flood extents from the US Geological Survey (USGS) (Watson et al., 2018). The results of this validation procedure are compared to those obtained when a simpler Height Above Nearest Drainage (HAND) model from the NOAA National Water Center (NWC) is used instead of the hydrodynamic model (NOAA NWC et al., 2018).

2. Methods and data

2.1. Hydrodynamic model (Fathom)

Hurricane Harvey made landfall on the east coast of Texas in August 2017, where some areas experienced 8-day rainfall totals of over 1500 mm and dozens of USGS river gauges recorded return period flows exceeding 1 in 100 years. Three important features of this tropical cyclone were forecasted by NOAA: streamflow from the NOAA National Water Model (NWM), rainfall from the NOAA Weather Prediction Center (WPC) and predicted storm surge height from the NOAA National Hurricane Center (NHC). The NOAA NWM (http://water.noaa.gov/about/nwm) has four variants: the analysis and assimilation configuration, which provides a real-time view of current streamflow conditions; short-range streamflow forecasts up to 18 hours; medium-range streamflow forecasts up to 10 days; and long-range streamflow forecasts up to 30 days. The medium-range product used here (NWM v1.1) takes meteorological variables from the NOAA Global Forecast System (https://www.emc.ncep.noaa.gov/GFS/) as inputs to the WRF-Hydro hydrological model (Gochis et al., 2018) to forecast streamflow for every river reach in the US, as defined by the USGS National Hydrography Dataset (NHD), via the Noah-MP land surface model (Niu et al., 2011) and a Muskingum-Cunge channel routing scheme. Here, we extract the maximum simulated streamflow on each river in the domain from the NWM medium-range configuration over 3 days of forecast model time, i.e. the maximum of all forecast streamflows within a 72-hour forecast horizon from the same model run (http://thredds.hydroshare.org/thredds/catalog/nwm/medium_range/catalog.html). The NOAA National Water Model is unrelated to official forecasts issued by NOAA NWS based on RFC modelling, but is employed here since streamflow is forecast for every US river. The NOAA WPC data used is the 3-day interactive forecast of 72-hour rainfall for 20 × 20 km grid cells, output by a NWP model but subject to manual adjustments by forecasters at the WPC (https://www.wpc.ncep.noaa.gov/qpf/day1-3.shtml). The NOAA NHC Probabilistic Tropical Storm Surge (P-Surge) model routes simulated meteorological variables through a SLOSH model to determine potential storm surge heights at the coast with a 3-day lead time (https://slosh.nws.noaa.gov/psurge2.0/).

For the fluvial inundation forecast, flood depths within a given river basin, as defined by HydroBASINS (Lehner and Grill, 2013), are extracted from an existing library of nationwide flood maps at ~30 m resolution (Wing et al., 2017). These fluvial flood maps are driven by discharges from a regional flood frequency analysis (RFFA) to ensure flood inundation is simulated on every US river, meaning multiple maps corresponding to a certain annual exceedance probability (e.g. 20% to 0.1% AEP) have been generated for the whole country. For each river basin in the study area, the NWM streamflow forecast is assigned an AEP, based on basin-specific data from a US variant of the global RFFA of Smith et al. (2015) using USGS river gauges, and the relevant AEP flood map is extracted from the library. Fig. 2 visualizes this process. By sampling from pre-existing flood maps of the entire continental US, simulation quality (e.g. grid resolution, physical process representation) is not hampered by the need for a forecast with reasonable latency: the extraction process takes only seconds (Leedal et al., 2010).

For fluvial flooding, the response to extreme streamflow in a river basin is relatively well-confined to that particular locality, meaning any plausible flood event will match a pre-computed, event-agnostic inundation simulation of that area provided enough runs with different boundary conditions are executed. How rainfall and surge vary in time and space, however, is less related to the hydrologic conditioning of the ground surface. In other words: simulating these perils ahead of time is inadvisable, given the *a priori* specification of suitable extraction zones is less defensible than for fluvial flooding. As such, the NOAA rainfall and surge information is input to Fathom-US to generate new event-specific depth grids, requiring ~6-hour parallel simulations of coastal and pluvial flooding for a ~400,000 km^2 domain at ~30 m resolution on a single node of 20 cores (Intel Broadwell E5 Xeon). Alongside model-building and post-processing, the final product (maximum flood depth in a pixel from the fluvial, pluvial and surge models) was produced ~24 hours after the release of the NOAA data, meaning the total lead time was ~2 days. The forecasts were updated with observed boundary conditions to produce a hindcast so that a more accurate delineation of impacted areas could be identified with immediacy. These boundary condition observations consisted of 24-hour rainfall totals from the NOAA NWS (https://water.weather.gov/precip/download.php), peak flows observed at USGS river gauging stations (https://waterdata.usgs.gov/nwis/rt) and coastal water level observations from the NOAA Tides and Currents service (https://tidesandcurrents.noaa.gov/gmap3/). During the Harvey event the forecast and hindcast flood extent data produced by the system were made freely available to first-responders and were used operationally by NASA and insurers to rapidly assess exposure.

Fathom-US is a large-scale hydraulic modelling framework, whose freshwater flooding component is set out in Wing et al. (2017). The US implementation is itself a variant of the global model first described by Sampson et al. (2015). It has been rigorously tested against thousands of bespoke flood inundation studies carried out by US government agencies, concluding that the large-scale methodology is approaching the accuracy of traditional local models (Wing et al., 2017). The framework permits rapid construction and execution of model domains as defined by the forecast footprint of US hurricane landfalls. Its computational hydraulic engine is driven by a variant of the LISFLOOD-FP hydraulic model, which solves a local inertial form of the shallow water equations in two dimensions over a regular grid (~30 m: 1 arc second resolution in this case) using a highly efficient numerical solution (Bates et al., 2010; de Almeida and Bates, 2013). This grid is populated with elevation values from the 1 arc second version of the USGS National Elevation Dataset (NED), with levee information from the US Army Corps of Engineers National Levee Database explicitly represented. This Digital Elevation Model (DEM) has complete coverage of the conterminous US and is thus the crucial component of the framework's applicability to simulate inundation for all potential hurricane landfall locations in the US. River hydrography is represented by HydroSHEDS (Lehner et al., 2008); with those rivers wider than the grid resolution being burnt directly into the DEM, and narrower streams being represented by the subgrid 1D model of Neal et al. (2012). Fluvial modelling is executed by inserting river discharge information (which is ultimately linked to an RFFA-derived AEP, an NWM streamflow forecast or a USGS river gauge observation) at the relevant inflow points in the stream network, while the pluvial model takes spatial rainfall data and drops it directly onto the land surface ("rain-on-grid"; Sampson et al., 2013). Assumptions relating to infiltration capacity are made based on soil information from the Harmonized World Soil Database in conjunction with a modified Hortonian infiltration equation (Morin and Benyamini, 1977). In urban areas, identified using satellite luminosity data (Elvidge et al., 2007), the infiltration capacity is defined using assumed urban drainage design standards. The storm surge model component was conceived by Bates et al. (2005), who adapted the 2D LISFLOOD-FP code traditionally used in fluvial settings for coastal flooding. Using LISFLOOD-FP for such an application has precedent: as
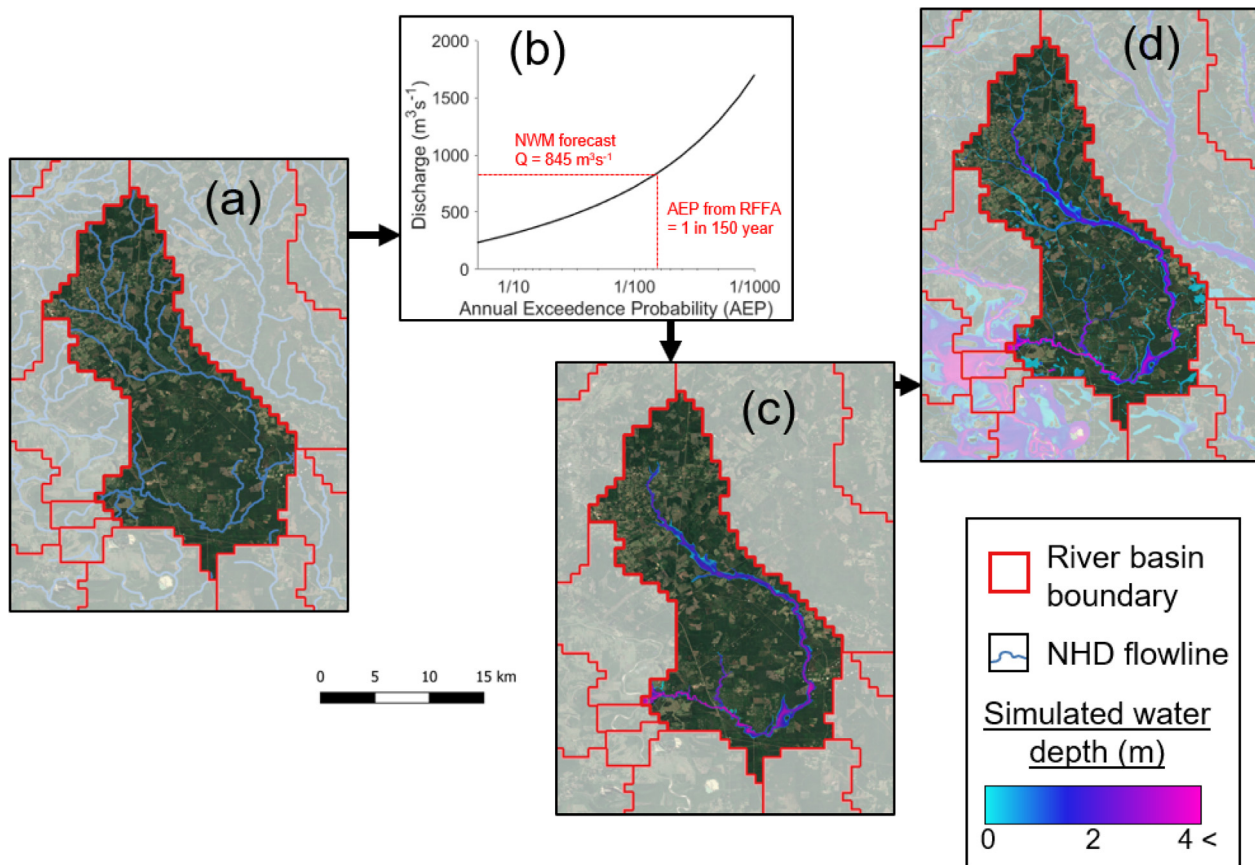
Fig. 2. An example demonstrating the process of extracting fluvial flood maps for each river basin. The unshaded river basin (ID = 17944) in (a) contains Menard Creek, a tributary of the Trinity River ~100 km NE of Houston, and has a drainage area of ~500 km². Streamflow is forecast for each flowline from the USGS NHD by the NOAA NWM. The graph in (b) shows the regional flood frequency analysis at the outlet of basin 17944. The NWM 3-day forecast peak discharge was 845 m³ s⁻¹ from a model run executed on 27 August 2017. This corresponds to the 1 in 150 year streamflow in this basin, whose inundation has already been simulated in the Fathom-US library as shown in (c). The depth grid is then extracted for this catchment. The final flood inundation map is shown in (d), where this process has been repeated for all river basins in the domain and integrated with the new pluvial simulations to represent headwater (rivers with drainage area < 50 km²) and surface water hazard.

examples, Smith et al. (2012) evaluated its suitability for coastal inundation modelling in the UK and Quinn et al. (2014) applied the code in a coastal flood risk assessment. In this component, the coastal boundary line of the model domain is set within oceanic cells just offshore of the coastal flood defenses. For each cell along the coastal boundary, the predicted peak surge height was extracted from P-surge output and used to scale the tidal time series at that location to create a fractional surge height time series. Water enters the model domain along the boundary in accordance with this time series, and the hydraulic model simulates the dynamics of the surge as it interacts with the shoreline and moves inland. 2D (burnt in the DEM) and 1D (subgrid) channels are represented here also, meaning the ability of the storm surge to propagate inland via river channels is properly represented: a crucial component of coastal flood models (e.g. Maskell et al., 2014).

2.2. NWM-HAND model (NOAA NWC)

The NOAA National Water Center has recently explored the coupling of the National Water Model to the Height Above Nearest Drainage method (Rodda, 2005; Rennó et al., 2008; Nobre et al., 2011, 2016). This effort forms part of the National Flood Interoperability Experiment (Maidment, 2016): which, like this paper, seeks to append the neglected receptor component to the flood forecasting cascade (Fig. 1). The HAND approach normalizes the DEM so that a given pixel takes the value of the vertical distance to the stream it drains to. A

rating curve is then used to translate NWM flow forecasts to stage for a given river cell, and any land cells that drain to this stream location and have a HAND value less than the stage become flooded. 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 modelling (Bates and De Roo, 2000) and again more recently (Afshari et al., 2018), owing to their omission of mass and momentum conservation laws. In isolated test cases however, particularly on confined floodplains with steep valley sides and straight river reaches, a planar approach may offer satisfactory performance (Bates and De Roo, 2000). Furthermore, the reduction in model skill may be considered acceptable where rapid solutions are required to large computation problems (Afshari et al., 2018; Liu et al., 2018). Geographies where these methods are appropriate are difficult to specify *a priori* though, since they have not undergone the same level of wide-area testing as hydrodynamic approaches (e.g. Wing et al., 2017). In evaluating the simpler HAND-based approach alongside the more complex hydraulic model presented here, the trade-off between including the physics of water flow in a forecast model and computational efficiency can be quantified.

The NWM-HAND model was executed for Hurricane Harvey (NOAA NWC et al., 2018) using the USGS NED at 1/3 arc second resolution as the DEM. The model was run each day with NWM analysis and assimilation outputs, providing a snapshot of the flood extent caused by Harvey at that time. By taking the maximum extent from all of these

HAND-based simulations, an analogous comparison can be made with the Fathom hindcast: both models intended to simulate the maximum flood extent when driven with observations.

2.3. Validation data

We evaluated the Hurricane Harvey forecast and hindcast footprints against ground observations of flood extent and depth made by the USGS (Watson et al., 2018). After the event, USGS field teams visited impacted basins and collected high water marks (HWMs) in accordance with the guidelines of Koenig et al. (2016). These are surveyed from the debris or stain lines left by the receding water on the sides of buildings, trees, fences and other structures. Horizontal co-ordinates are obtained with a GPS, while vertical heights are referenced to the NAVD88 datum. The 2123 resultant HWMs were, in combination with USGS gauging station maximum water levels, interpolated (technique described in Musser et al., 2016) across 1.4–3 m resolution DEMs built with LiDAR data for fourteen sites. This provided the current best reconstruction of flood extents from the observed water level data. Though these observation data are used as a benchmark in this study, they are not error-free. Watson et al. (2018) listed uncertainties for specific data points in their study, and these range from < 0.01 m–0.55 m (mean 0.07 m), though no information was specified regarding the method of quantification. For further details, see Watson et al. (2018).

2.4. Validation metrics

Firstly, the gridded flood depth benchmark data, derived from observed water levels, were used to test the extent to which the models captured the overall spatial pattern of the flooding. For this, the same binary pattern measures were used as in Wing et al. (2017):

$$Hit \ Rate = \frac{M_1 B_1}{M_1 B_1 + M_0 B_1} \tag{1}$$

$$False \ Alarm \ Ratio = \frac{M_1 B_0}{M_1 B_0 + M_1 B_1} \tag{2}$$

$$Critical \ Success \ Index = \frac{M_1 B_1}{M_1 B_1 + M_0 B_1 + M_1 B_0} \tag{3}$$

$$Error \ Bias = \frac{M_1 B_0}{M_0 B_1} \tag{4}$$

where M and B represent model and benchmark cells respectively, and the subscript 1 and 0 indicate if the cell considered is wet and dry respectively.

The 'hit rate' metric (HR; Eq. (1)) penalizes type II errors and is thus a measure of the model's tendency to underpredict the benchmark flood extent. It can be interpreted as the proportion of benchmark flooded areas that were replicated by the model. The 'false alarm ratio' (FAR; Eq. (2)) penalizes type I errors and so represents the tendency of the model to overpredict the benchmark flood extent. This metric can be interpreted as the proportion of modelled flooded areas that are dry in the benchmark. The 'critical success index' (CSI; Eq. (3)) penalizes both type I and type II errors, thus being a metric that accounts for both under and overprediction. It ranges between 0 (no match between model and benchmark) and 1 (perfect match between model and benchmark) and can be thought of as representing the model performance over floodplain areas only as it excludes areas that do not inundate in both the model and benchmark data. Finally, 'error bias' (EB; Eq. (4)) is the ratio of type I to type II errors. Values greater than 1 indicate the model tends to overpredict, while values less than 1 indicate a tendency to underpredict with respect to the benchmark.

Secondly, we calculated the difference in water surface elevation (WSE) specified by the USGS HWMs and the models. These differences are analyzed in three ways:

$$Root \ Mean \ Squared \ Error = \sqrt{\frac{\sum_{n=1}^{N}(O_n - M_n)^2}{N}} \tag{5}$$

$$Mean \ Absolute \ Error = \frac{\sum_{n=1}^{N}|O_n - M_n|}{N} \tag{6}$$

$$Mean \ Error = \frac{\sum_{n=1}^{N} O_n - M_n}{N} \tag{7}$$

where O_n and M_n is the WSE at a given observation point and corresponding model cell respectively, N is the number of HWMs analyzed and everything else is as above. The original dataset was trimmed down to 1134 points so that the analysis was confined to high quality HWMs that were not taken at the same location and which were referenced against the same geodetic datum.

Both Root Mean Squared Error (RMSE; Eq. (5)) and Mean Absolute Error (MAE; Eq. (6)) measure the average magnitude of the errors (where an error is $O_n - M_n$). RMSE is a quadratic scoring rule (meaning greater weight is given to larger errors), while MAE is linear (all errors have equal weight). Mean Error (ME; Eq. (7)) calculates the average error whilst still accounting for their sign: a negative ME indicates the model has a tendency to overpredict observed WSEs, while a positive one suggests underprediction.
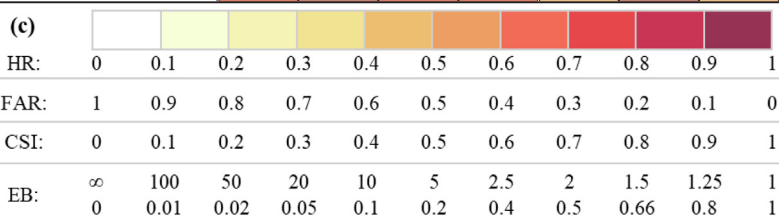
3. Results and discussion

3.1. Flood extent comparison

The results of the flood extent comparison between the Fathom hindcast model (driven with observed streamflow, rainfall and surge heights) and USGS benchmarks for each of their fourteen study sites are shown in Table 1a and Fig. 3. Most of the USGS flood extent is captured by the model, with 78% of observed wet pixels being correctly identified as such in the model on average across all sites. Many of the sites with high HRs also have relatively high FARs (e.g. Tres Palacios River (Fig. 3a), Lower San Bernard River (Fig. 3l) and Upper Brazos River (Fig. 3b)), driving a lower overall correspondence and overpredictive bias as indicated by their CSIs and EBs respectively. Conversely, many lower-HR sites have very low FARs (e.g. Upper and Lower Neches River (Fig. 3k and 3m) and Cow Bayou (Fig. 3f)), generating comparable CSIs and EBs < 1 indicating underprediction. Some high-HR sites have correspondingly high CSIs, owing to low FARs (e.g. Lower Brazos River (Fig. 3e) and Middle San Bernard River (Fig. 3j)), suggesting there is a strong match between model and benchmark flood extents. CSIs, which are the most discriminatory metric, range from a poor 0.5 (the model is correct as often as it is incorrect) on the Upper San Bernard River (Fig. 3i) to an excellent 0.9 (a 90% match between model and benchmark) on the Lower Brazos River (Fig. 3b). Across the domain, an average CSI of 0.66 indicates that roughly 2 in every 3 model pixels in the functional floodplain match the benchmark. Further to the details in Section 2.3, the USGS benchmark flood inundation extent is subject to uncertainty. It is generated via interpolation between point HWMs (themselves containing vertical error), rather than a genuine 2D observation, meaning the accuracy of these maps is heavily dependent on the spatial resolution of the HWMs. Without representing the physics of flow that generated the HWMs in a hydraulic model (e.g. timing and interaction of adjacent streamflows), the interpolation procedure may produce unrealistic inundation extent boundaries where unconstrained by point observations. To put the pattern scores into context: the Wing et al. (2017) model attained average CSIs of ~0.75 against detailed local models, with a maximum of ~0.90; global flood models average CSIs of ~0.50, obtaining up to ~0.70 in isolated test cases, against a variety of local models and satellite-derived flood extents (Sampson et al., 2015; Winsemius et al., 2016; Dottori et al., 2016; Ward et al., 2017); local flood models that have been manually built and extensively calibrated generally achieve CSI scores of 0.70–0.80 when

Table 1

A comparison of (a) the Fathom model hindcast and (b) maximum flood extent from the NWM-HAND model hindcast of the NOAA NWC to the benchmark USGS flood extents. The metrics Hit Rate (HR), False Alarm Ratio (FAR), Critical Success Index (CSI) and Error Bias (EB) are explained by Eqs. (1)–(4) in Section 2.4. To aid interpretation, (c) indicates the color scale used to classify each measure: the darker the color, the higher the performance.

Site Name	(a) Fathom hindcast model				(b) NWC hindcast model			
	HR	FAR	CSI	EB	HR	FAR	CSI	EB
Tres Palacios River	0.89	0.41	0.55	5.59	0.67	0.43	0.45	1.54
Upper Brazos River	0.89	0.18	0.74	1.68	0.52	0.19	0.46	0.26
Lower Brazos River	0.93	0.04	0.90	0.61	0.50	0.05	0.48	0.06
Cow Bayou	0.77	0.11	0.70	0.42	0.51	0.12	0.47	0.14
Big Cow Creek	0.65	0.16	0.57	0.35	0.54	0.19	0.48	0.27
East Matagorda Bay	0.76	0.15	0.67	0.55	0.33	0.25	0.30	0.17
Matagorda Bay	0.79	0.37	0.54	2.19	0.58	0.21	0.50	0.37
Upper San Bernard River	0.55	0.15	0.50	0.21	0.13	0.05	0.13	0.01
Middle San Bernard River	0.79	0.04	0.76	0.16	0.34	0.04	0.34	0.02
Lower San Bernard River	0.95	0.25	0.72	6.01	0.75	0.27	0.58	1.10
San Jacinto River	0.76	0.35	0.54	1.74	0.45	0.25	0.39	0.28
Lower Neches River	0.79	0.07	0.74	0.29	0.37	0.07	0.36	0.05
Upper Neches River	0.70	0.02	0.69	0.06	0.36	0.03	0.36	0.02
Pine Island Bayou	0.67	0.04	0.65	0.08	0.44	0.07	0.43	0.06
MEAN	**0.78**	**0.17**	**0.66**	**1.42**	**0.46**	**0.16**	**0.41**	**0.31**

(c)												
HR:	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	
FAR:	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0	
CSI:	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	
EB:	∞	100	50	20	10	5	2.5	2	1.5	1.25	1	
	0	0.01	0.02	0.05	0.1	0.2	0.4	0.5	0.66	0.8	1	

compared to satellite observations of flood inundation (Aronica et al., 2002; Pappenberger et al., 2007; Di Baldassarre et al., 2009; Wood et al., 2016) and up to 0.9 when benchmarked against very high quality data (e.g. Bates et al., 2006; Altenau et al., 2017). Fleischmann et al. (2019) propose that a hydrodynamic model provides locally relevant estimates of flood extent when CSI > 0.65. It should be noted that the CSI metric is sensitive to the selected study area, however: favoring overpredictive models of larger floods on flat terrain compared to the reverse case (Stephens et al., 2014). The results shown here are towards the higher end of those in the literature (though many of these are for smaller floods where high CSIs are difficult to obtain). The hindcast model performance is, however, shy of the very high CSIs exhibited by calibrated local model – observation comparisons. The results thus indicate that, across the model domain, the hindcast hydraulic model presented here has some skill in replicating benchmark patterns.

The results of the comparison between the NWM-HAND model and the USGS benchmark are shown in Table 1b. It is evident here that the

NWM-HAND model has lower predictive skill when benchmarked against the USGS flood extents. Mean HRs indicate 46% of flooded areas are correctly captured, with mean CSIs suggesting just over 4 in every 10 pixels in the functional floodplain are identified correctly. It should be noted that the NWM-HAND model structure is only capable of representing fluvial flood hazard, meaning both hydrologically-isolated flooding from intense local rainfall and coastal surge will not be captured. Hurricane Harvey was predominantly a pluvial event: flooding arose in many areas due to intense local rainfall on the land surface, rather than from rivers flowing out of bank. Fig. 4 shows how different the Fathom and NWM-HAND models look when this runoff component is represented in an area to the NE of Houston. HAND being a fluvial-only model perhaps explains why coastal basins (e.g. two sites in Matagorda Bay) have such low scores, since storm surge may have played an important role here. The lack of a rainfall component may explain poor performance on small streams: the Upper San Bernard River has a drainage area of ~500 km² and the San Jacinto River site
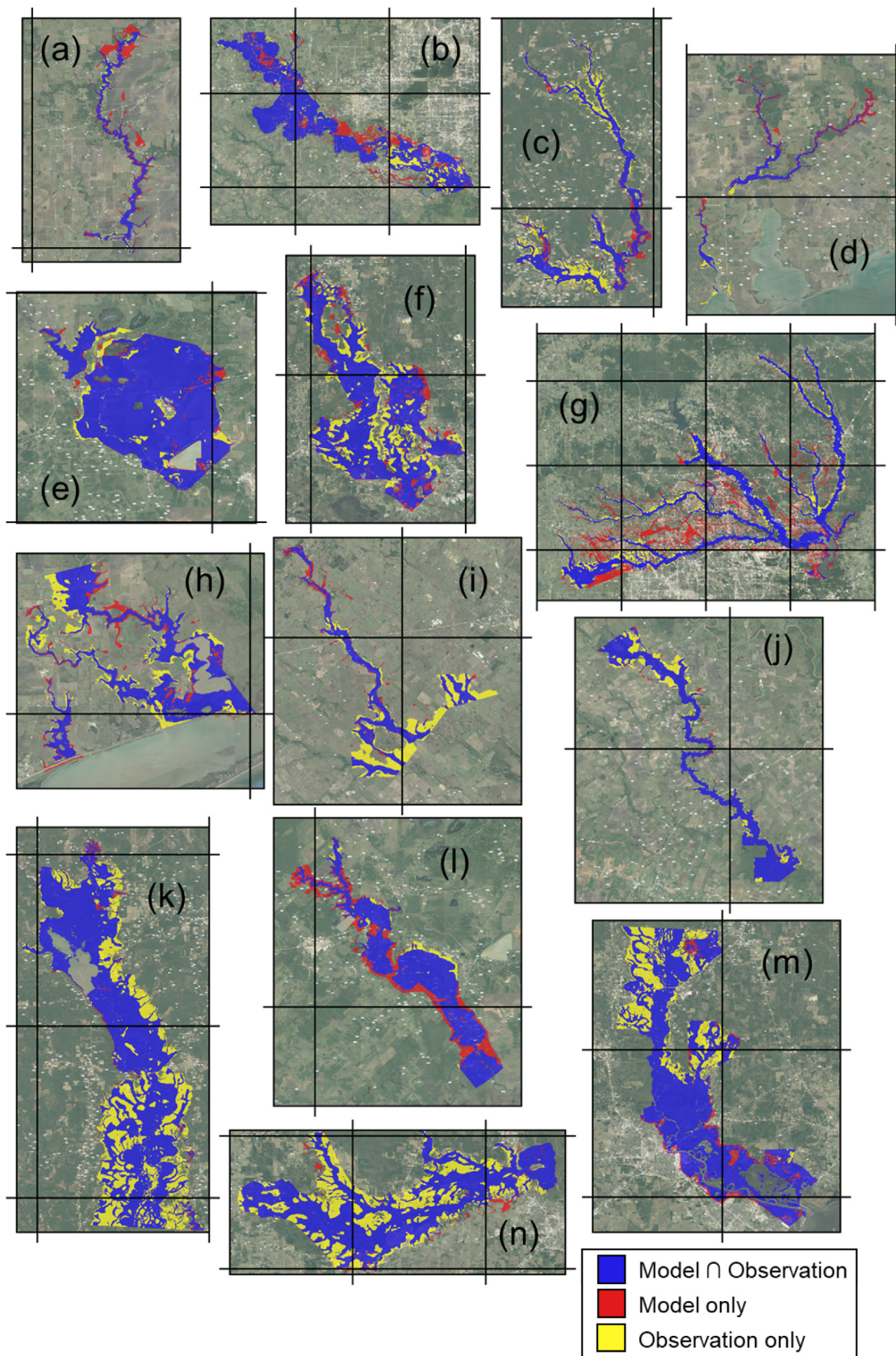
Fig. 3. Maps displaying the intersection of the Fathom hindcast modelled Hurricane Harvey flood extent with those generated by the USGS based on observed HWMs at 14 sites: (a) Tres Palacios River, (b) Upper Brazos River, (c) Big Cow Creek, (d) Matagorda Bay, (e) Lower Brazos River, (f) Cow Bayou, (g) San Jacinto River, (h) East Matagorda Bay, (i) Upper San Bernard River, (j) Middle San Bernard River, (k) Upper Neches River, (l) Lower San Bernard River, (m) Lower Neches River and (n) Pine Island Bayou. Grid scales in all panels are 0.25° (~27 km).
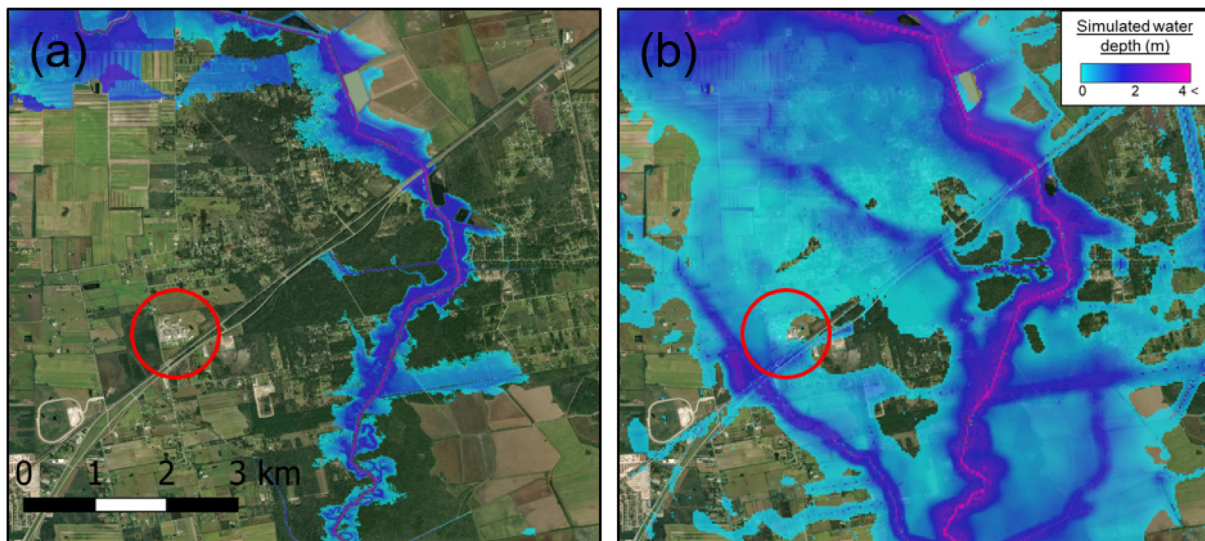
Fig. 4. An area near Crosby, NE of Houston, exhibiting the output of (a) the NWM-HAND and (b) the Fathom hindcast models. The Arkema chemical plant is within the red circle, which was inundated during Harvey and led to an industrial disaster. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

contains headwater streams with drainage areas as small as ~20 km², and both exhibit extremely poor performance with NWM-HAND. The flood hazard arising on these small streams is not driven by traditional fluvial flooding processes, where aggregation effects lead to a large, low-amplitude flood wave which propagates downstream, but by the rapid lateral surface flow of intense local rainfall generating a flash flood. This further contextualizes the scores obtained by the Fathom model at these sites too, where, despite possessing a pluvial model component, the difficulties in simulating this phenomenon are evident. With that being said, for even the relatively simple problem of modelling the Brazos River, which has a flat floodplain confined by steep valley sides, NWM-HAND correctly identifies less than half of the pixels. The hydrodynamic Fathom model, in contrast, correctly identifies between 74% and 90% of flooding from the Brazos. For reference, CSIs obtained by HAND models have been reported between 0.5 and 0.9 for a selection of watersheds in the US when comparing HAND approaches to observational data or hydrodynamic methods (Zheng et al., 2018a, 2018b; Zhang et al., 2018; Afshari et al., 2018). The results from the NWM-HAND model presented here do not appear to be as skillful as those presented in these smaller scale studies. Further, despite this approach making use of higher resolution (~10 m) terrain data, it is outperformed by a coarser (~30 m) hydrodynamic model. This suggests that the importance of higher grid resolution is only realized when physical processes are represented.

Further to testing the Fathom and NWM-HAND models that were driven with real-time observations, the Fathom forecast variant was also benchmarked against the USGS flood extents. It was driven with NWM peak flows from a 3-day forecast commencing at 1800 GMT-6 on 27th August 2017, as well as forecast rainfall and surge data from this time. Hurricane Harvey generated peak streamflows across the domain for over 5 days: beyond the horizon of the 3-day forecast model presented here. The forecast model output chosen for validation here captures (temporally) the main hurricane impact in Texas, which was felt prior to 31st August 2017. The results of this comparison are shown in Table 2, where the Fathom forecast model sees a performance drop with respect to its hindcast variant: mean CSI drops by 0.09 and mean HR drops by 0.03. While HRs actually increase for most sites when compared to hindcast performance (as high as 99.9% on the Lower San Bernard River), this is amidst a backdrop of increased EBs that indicate a heavy bias towards overprediction (mean EB of 84.02, compared to a more modest 1.42 in the hindcast). It's worth noting here, though, that EBs are measured on a logarithmic scale. Since underprediction is

Table 2
A comparison of the forecast variant of the Fathom model to the benchmark USGS flood extents. The color scale used is explained in Table 1c.

Site Name	Fathom forecast model			
	HR	FAR	CSI	EB
Tres Palacios River	0.96	0.50	0.49	25.21
Upper Brazos River	1.00	0.25	0.75	73.31
Lower Brazos River	0.99	0.12	0.87	24.51
Cow Bayou	0.19	0.15	0.18	0.04
Big Cow Creek	0.50	0.13	0.46	0.15
East Matagorda Bay	0.88	0.37	0.58	4.23
Matagorda Bay	0.98	0.57	0.42	65.05
Upper San Bernard River	0.75	0.20	0.63	0.77
Middle San Bernard River	0.96	0.14	0.83	3.62
Lower San Bernard River	1.00	0.38	0.62	977.4
San Jacinto River	0.81	0.31	0.60	1.94
Lower Neches River	0.61	0.03	0.60	0.05
Upper Neches River	0.44	0.01	0.44	0.01
Pine Island Bayou	0.46	0.03	0.45	0.03
MEAN	**0.75**	**0.23**	**0.57**	**84.02**

indicated by values less than 1 and overprediction by values greater than 1, a value of 0.5 has the same magnitude bias as a value of 2, yet the mean of these two values is 1.25 (not 1). Underrepresented by this

mean EB, therefore, are the extremely high underpredictive biases evident on the Lower and Upper Neches River, and Cow and Pine Island Bayous. The use of forecasted, rather than observed, model inputs is therefore widening the tails of the bias distribution (EBs moving further from 1) and reducing overall performance (CSIs decreasing). With that being said, capturing, on average, three-quarters of benchmark inundated pixels (HR) while keeping false alarms at 23% (FAR) is indicative of fair performance in the forecast model.

Aside from structural errors in the hydraulic model and its components (e.g. the DEM), the hindcast model contends only with errors in the measured data it is driven with. Though this is not insignificant, 25–40% error in measured flows is common and can be even higher when flows are extreme (Di Baldassarre and Montanari, 2009; Coxon et al., 2015; Westerberg et al., 2016), it is likely to be much lower than uncertainties in the boundary conditions used in the forecast hydraulic model. These uncertainties, though, are typically explored by the use of ensemble prediction systems (EPS) in meteorological (e.g. Buizza, 2005) and hydrological (e.g. Thielen et al., 2009) forecast models, where many realizations of projected weather or water levels are simulated for a single site (Cloke and Pappenberger, 2009). These probabilistic frameworks thus permit some measure of forecast predictability to be quantified. In the models presented here, there is only a single deterministic flood extent for each day's forecast. The reason for this is two-fold. Firstly, the NOAA NWM itself currently produces deterministic streamflow forecasts for its medium-range variant. NWM v1.1 was the version in existence during Hurricane Harvey, but current and future versions will improve on this functionality (e.g. NWM v2 will have a 7-member ensemble in the medium-range forecast). Were a probability distribution of projected river discharge available, the rapid extraction algorithm that samples from a pre-existing library of fluvial flood maps could feasibly produce 2D grids where each cell represents the probability density function of water depth. This process takes only seconds, so scaling to a probabilistic framework remains trivial. Secondly, the computational expense of running new hydrodynamic pluvial and coastal flood models in an EPS would add significant computational burden. With that being said, running an ensemble of the hydraulic model presented here (a single deterministic run requires ~ 120 processor hours) would be a manageable task for the HPC facilities of leading forecast centers and so probabilistic depth grids of these flood drivers could be constructed too. Incorporating the hydrodynamic model in an EPS would be a relatively straightforward addition to the method proposed here, but is one which is beyond the scope of the current paper.

Digging deeper into the forecast uncertainties, observations of peak streamflow from 63 USGS river gauges across the footprint domain are compared to the corresponding NWM streamflow forecast from 27th August 2017. All river gauges used experienced peak flow during this forecast's 3-day time horizon. By volume, the mean absolute error (as in Eq. (6), but for discharge rather than water surface elevation) comes to 2970 m^{3}s^{-1}. For reference, this error is equivalent to roughly 80% of the peak discharge experienced on the Lower Brazos River during Hurricane Harvey. Errors as a proportion of the observed discharge are shown in Fig. 5. The MAE is 290%, with a very high bias towards overprediction (mean error of –281%). Forecast discharge on the Buffalo Bayou near Addicks was an eighteenfold overestimate: an implausible 7129 m^{3}s^{-1} when 390 m^{3}s^{-1} was observed. In some cases, though, the NWM was very accurate. For instance, on the East Fork of the San Jacinto River near New Caney: 3492 m^{3}s^{-1} was forecasted while 3398 m^{3}s^{-1} was observed. It should be noted that this a single forecast from single point in time, while the medium-range NWM variant is run four times a day for 80 time horizons. The testing presented here, therefore, will not be representative of NWM performance during Harvey and should not be viewed as an authoritative assessment of model skill. It may be ill fortune that the particular Fathom flood extent forecast selected for this study was driven by anomalously poor discharge forecasts, but their benchmarking against observed flows

provides useful context for analyzing the skill of the 2D flood forecast. Furthermore, the NWM was in its infancy around the time of Harvey (v1.1). Improvements, both to date and in the future (v2), of this hydrological model after rigorous validation exercises will result in much closer replication of observed streamflows than those presented here. Official NOAA RFC forecasts are of much greater accuracy (Adams, 2016), but do not have total coverage of US rivers. In Fig. 6, the discharge errors at USGS observation sites are plotted and polygons representing the fourteen sites are colored by their error bias (from Table 2). Many of the USGS gauging stations do not relate to sites where USGS benchmark flood extents were generated, perhaps explaining why the CSIs in Table 2 are not as low as the discharge errors in Fig. 5 might suggest. Stations are heavily concentrated in Houston, of which only some are relevant to just one of the fourteen study sites (San Jacinto River (Fig. 6g)). Equally, many of the sites do not have a representative USGS gauging station: but for those that do, biases in the forecast discharge are often replicated in the inundation extent (where the site polygon and gauge point are of a similar color in Fig. 6). Forecast flood inundation in the east of the domain was driven by NWM discharges that are generally biased towards underprediction, which is reflected in the bias of the flood extent (Big Cow Creek (Fig. 6c), Upper (Fig. 6k) and Lower Neches River (Fig. 6m), Cow Bayou (Fig. 6f), Pine Island Bayou (Fig. 6n)). Most of the remaining sites have an overpredictive flood extent bias, where nearby gauges indicate that the NWM forecast discharge was overpredictive also (Upper (Fig. 6b) and Lower Brazos River (Fig. 6e) and San Jacinto River (Fig. 6g)).

The conclusions drawn in this section generally verify those found in the wider literature that models based on some formulation of the shallow water equations produce a more accurate simulation of flood extent than simple planar approximations, though a performance differential of this magnitude has not been documented previously for fluvial flooding (in Bates et al. (2005) the planar method obtained a CSI of 0.11 when simulating coastal floods). As fully hydrodynamic approaches become increasingly computationally tractable, the advantages of using GIS-based methods will further diminish. Intuitively, the performance of the Fathom forecast model is lower than that of its hindcast counterpart, partly due to some significant errors in the NWM forecast streamflows and perhaps forecast rainfall relative to observations also. Yet, these results evidence that despite these limitations the forecast model has fair skill in replicating benchmark flood extents for Harvey.

3.2. High water mark comparison

The Fathom models were further tested against the raw HWMs that the USGS used to construct the benchmark flood extents in the preceding section. These observational data represent the maximum water surface elevation surveyed by USGS field teams as a result of Hurricane Harvey at a given point. For the hindcast model, the mean absolute WSE error between the selected observation points and the modelled value at the same location is 1.03 m. The influence of a small number of large outlying error values means that the RMSE is significantly higher at 1.71 m. Fig. 7a outlines the distribution of these WSE differences, exhibiting a Gaussian distribution about a central tendency close to 0 (mean error: 0.15 m). 50% of the errors lie between −0.43 m (Q_{25}) and 0.90 m (Q_{75}), while 90% of the errors are between −2.32 m (Q_5) and 2.12 m (Q_{95}). Fig. 7b exhibits the expected shallowing and widening of the error distribution when analyzing the Fathom forecast model. MAE rises to 1.22 m (RMSE = 1.88 m). However, this does not represent a drastic reduction in skill from hindcast to forecast.

Point comparisons of surveyed vs. simulated WSEs are a particularly stringent test for a model of this scale. Indeed, such examinations are rarely carried out for analogous models, particularly uncalibrated, large-scale, high-resolution, ones which include urban areas and smaller streams, meaning context-setting for these results is quite difficult. The Wilson et al. (2007) model of the Amazon obtained an RMSE
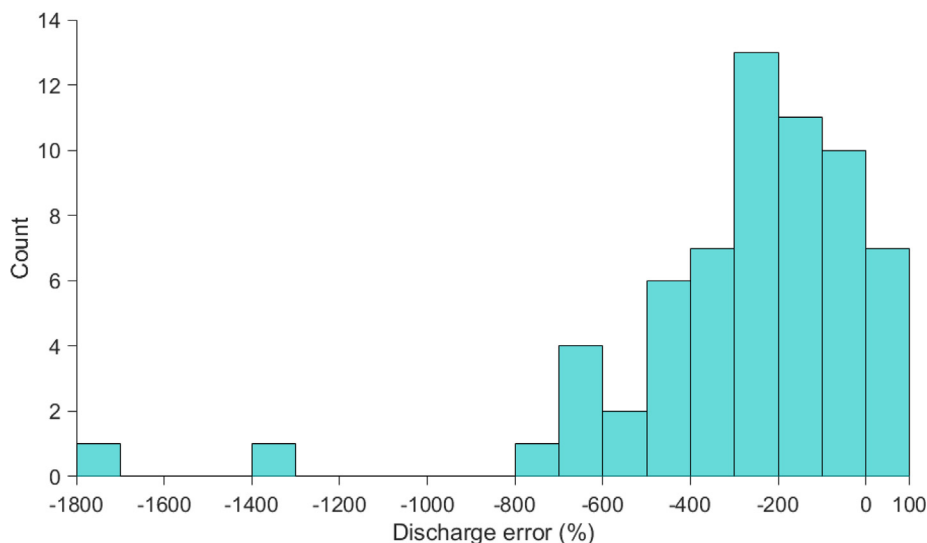
Fig. 5. Errors in National Water Model forecast discharge (Q) as a percentage of observed peak flows from USGS stream gauges. A positive error indicates NWM underprediction (observed Q > modelled Q), while a negative error indicates overprediction (observed Q < modelled Q).
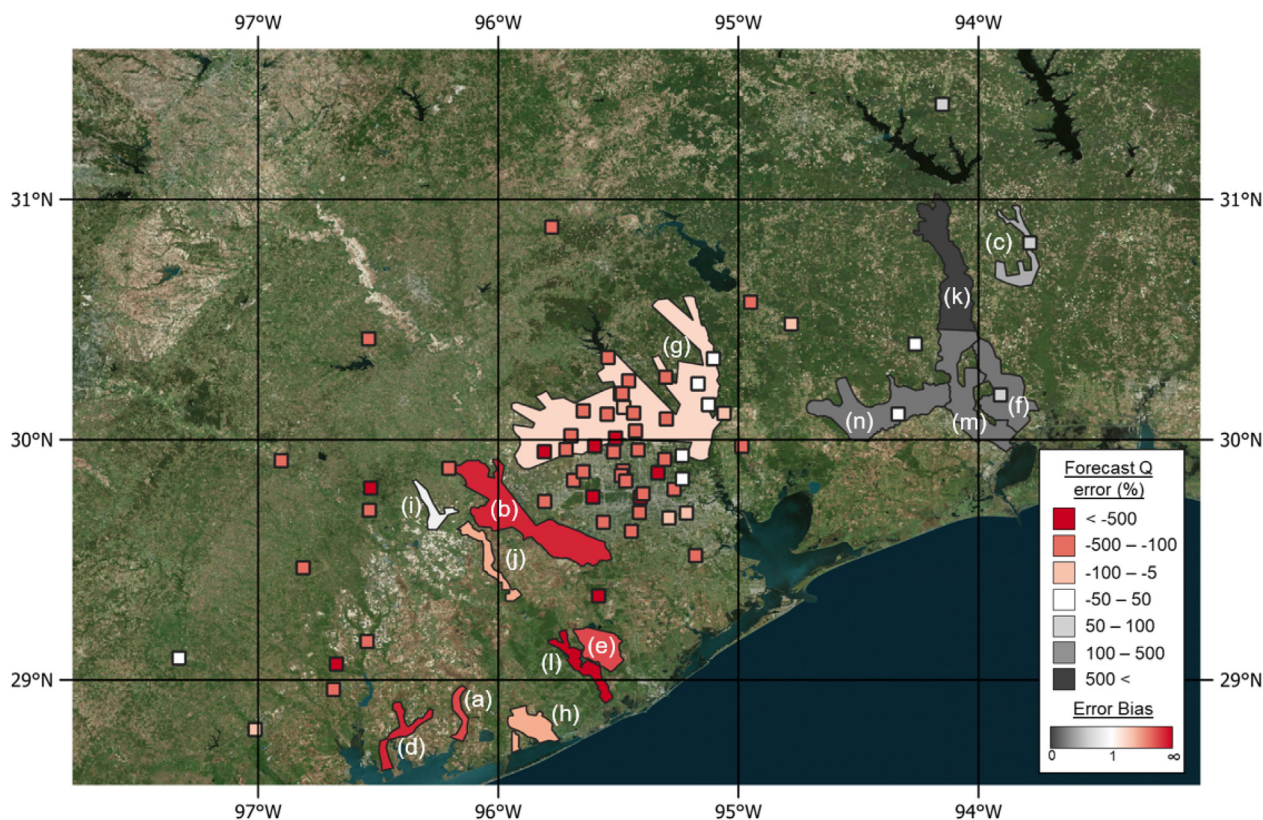

Fig. 6. Errors in National Water Model forecast discharge (Q) as a percentage of observed peak flows from USGS stream gauges (shown as points). Error Bias (Eq. (4)) is displayed for each of the 14 sites (shown as polygons): (a) Tres Palacios River, (b) Upper Brazos River, (c) Big Cow Creek, (d) Matagorda Bay, (e) Lower Brazos River, (f) Cow Bayou, (g) San Jacinto River, (h) East Matagorda Bay, (i) Upper San Bernard River, (j) Middle San Bernard River, (k) Upper Neches River, (l) Lower San Bernard River, (m) Lower Neches River and (n) Pine Island Bayou. Grey colors indicate underprediction, red colors indicate overprediction, white represents an unbiased estimate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of 2.37 m (0.99 m at high water) when comparing simulated water levels to those derived from satellite altimetry data. Schumann et al. (2013) calibrated their 1 km resolution forecasting model of the Lower Zambezi River to within 0.27 m of ICESat-derived water levels. This, though, is the product of using a smooth, coarse-resolution DEM of a very flat and wide floodplain, meaning WSEs change very little as the flood extent grows. Obtaining low errors under such circumstances is

much less challenging than in the test case presented here, especially when the Zambezi model has been calibrated to observations. Neal et al. (2009), in a high-resolution, reach-scale hydraulic model of an urban area built with local data, obtained a maximum RMSE of 0.28 m when calibrated to HWMs. As another example, the standard deviation of errors in the calibrated Mignot et al. (2006) model of an urban area in France was 0.53 m. It should be noted that hydraulic models that
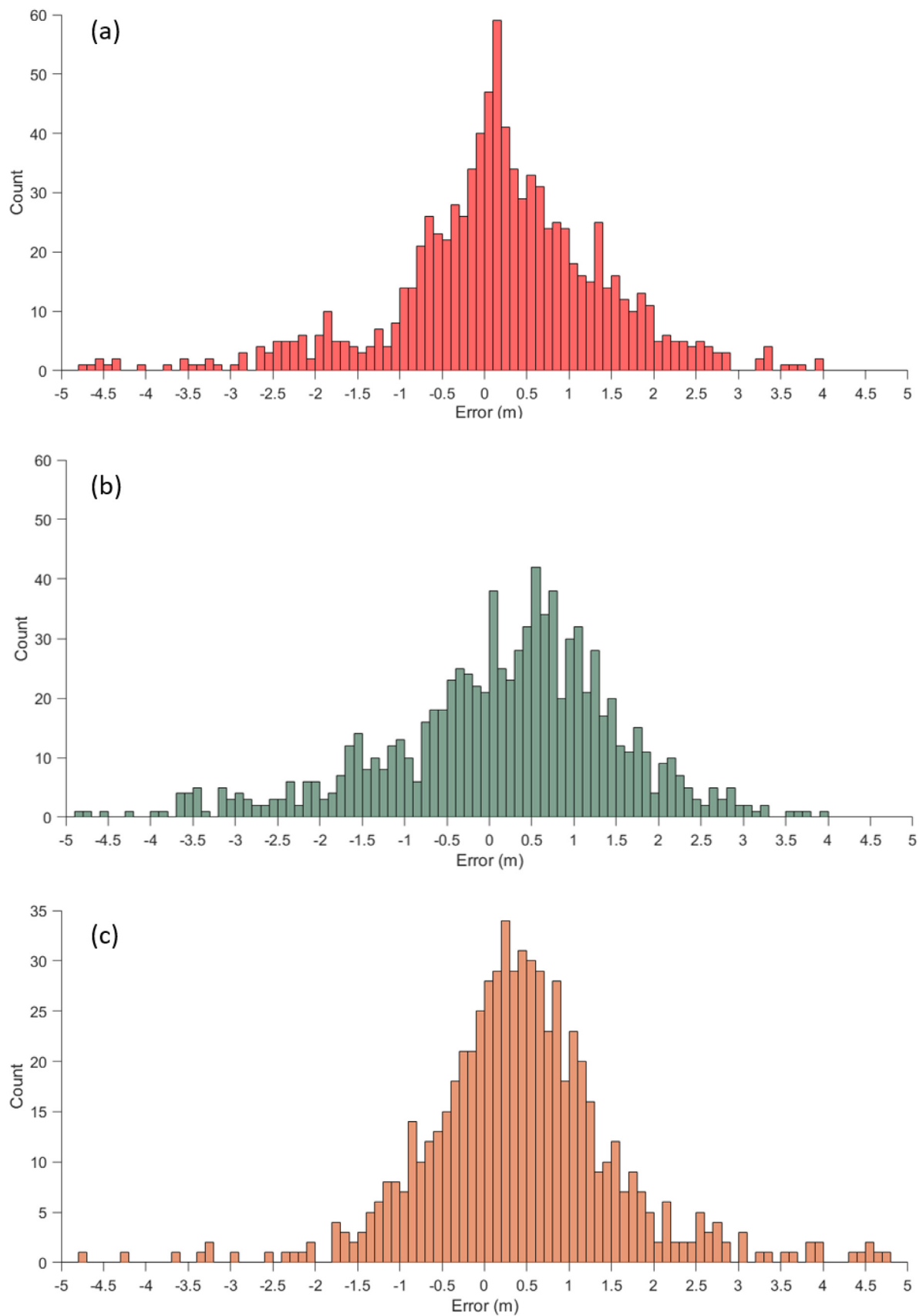
Fig. 7. Histograms of the differences between 1134 observed and simulated water surface elevations in the Fathom (a) hindcast and (b) forecast models. (c) shows the differences between 659 observed ground elevations and those in the DEM employed by the models. Note that 12, 13 and 9 data points fall outside of the displayed error range in (a), (b) and (c) respectively.

have been calibrated may not necessarily be more inherently skillful, meaning these local-scale model errors may not be analogous to the uncalibrated large-scale model presented here. As an example of an operational NOAA flood forecast, Adams et al. (2018) compare modelled hindcasts of their large-scale framework in the Midwest to USGS gauged stages and found errors of < 0.5 m. While this is more accurate than the model presented here, it required thousands of manual bathymetric surveys to build, meaning the Adams et al. (2018) model only simulates flooding on large rivers of known geometry. In the context of Hurricane Harvey, where flooding from small streams and surface runoff accounted for a considerable portion of the hazard, this NOAA exemplar model may not have attained similar errors when benchmarked against high water marks on the floodplain and particularly where flooding was predominantly pluvial. Zheng (2018) tested the NWM-HAND model of maximum inundation outlined in the preceding section (NOAA NWC et al., 2018) against this same set of USGS HWMs and found the standard deviation of model error to be 4 m. Previous studies have also quantified the errors in surveyed HWMs that such models are calibrated to or validated against. Errors are generally 0.3–0.5 m (Schumann et al., 2007; Neal et al., 2009; Horritt et al., 2010; Fewtrell et al., 2011), owing to deposition during a hiatus in flood recession (i.e. not high water), wall seepage or debris line width. These are broadly consistent with the maximum errors reported for the HWMs used in this study, but average errors are generally < 0.1 m (Watson et al., 2018).

It is clear that the automatically generated forecast and hindcast models in this paper do not obtain comparable WSE errors to bespoke hydraulic models in the wider literature or NOAA NWS RFC forecasts, though without hydraulic model calibration, local data and manual operation they were never likely to. It is also important to point out the difference in purpose between this and other models. Rather than seeking to perfectly replicate a given flood event with few limitations regarding processing and computation time, this is a forecast product designed to quickly indicate areas susceptible to flooding from an incoming storm to enable an immediate response. Rapid exposure assessments and resource allocation by first responders do not demand a highly accurate map of water depths, but rather a broad indication of the spatial patterns in flooding that may occur (Pitt, 2008; Price et al. 2012). Despite some sacrifice to point precision (e.g. with respect to the accuracy of the 155 discrete forecast inundation points in NOAA AHPS), the framework's total coverage, both in geography and flood driver, serves to fill the information gaps left by sporadic yet accurate local forecasts. Besides, efforts to reduce WSE errors in the hydraulic model are of little value when considering the commensurate or larger errors in the meteorological and hydrological models that precede it in the forecast cascade. Cangialosi (2018) noted that intensity and track errors in the NOAA NHC 3-day forecasts during the 2017 hurricane season were ~7 ms^{-1} and ~150 km respectively. The maximum total rainfall in the NOAA WPC 3-day forecast was ~1000 mm, yet a maximum of over 1500 mm was observed (Blake and Zelinsky, 2018). Hydrological models, used to translate rainfall to streamflow, may also be significantly uncertain, even if the meteorological inputs were error-free (Blöschl et al., 2013). Generally accepted benchmarks for satisfactory performance in hydrological models are: (i) within 25% error in simulated discharge and (ii) a Nash-Sutcliffe efficiency of greater than 0.5, meaning model mean square discharge error represents less than half of the observed variance (Moriasi et al., 2007; Refsgaard and Knudsen, 1996; Ritter and Muñoz-Carpena, 2013). With source and pathway uncertainties such as these propagating into the receptor component of the forecast cascade (Fig. 1), the hydraulic model deviating from observations by an average of ~1 m is unsurprising yet may still be useful for event early warning and first responder preparedness in the absence of local modelling strategies.

659 of the HWMs collected by the USGS also reported the height of the water above the ground surface, meaning a surveyed ground elevation can be calculated by subtracting this measurement of water depth from the WSE. Comparing these elevation values to those contained at the corresponding location in the DEM provides further context for the errors shown in Fig. 7a and b. Again, it should be noted that the observed data contends with error in both the surveyed water surface elevation and its height above the ground: such errors could plausibly be half a meter or more (Watson et al., 2018). Fig. 7c shows the spread of these elevation errors: the DEM RMSE is 3.77 m, though this is heavily biased by a handful of implausibly high errors (maximum negative error: −58.9 m; maximum positive error: 20.9 m) which are just as likely to be due to human error in the manual ground survey as erroneous values in the DEM. Reducing the influence of these errors by considering MAE, the quantity stands at 1.19 m. With forecast and hindcast WSE MAEs of 1.21 and 1.03 m respectively, the water levels simulated by the model are similar to or outperform the accuracy of the DEM, implying two key findings. Firstly, much of the WSE error is attributable to the quality of the USGS NED (the source of the model DEM) in this domain. It stands to reason, then, that in domains with a greater proportion of LiDAR in the NED, ground, and thus water surface, elevation error would drop. Once again, the assertion made by Horritt and Bates (2002) still holds: topography is the major control on flood inundation patterns. Secondly, although it seems counterintuitive that errors in WSE can be smaller than the DEM from which they are derived, this underlines the idea that relative, rather than absolute, DEM accuracy is much more important in hydraulic modelling. How the elevation varies between pixels in a locality controls the movement of water over a floodplain, not a pixel's elevation relative to a vertical datum (though this does not apply for simulations of coastal flooding). Typically, relative DEM errors are much lower than absolute ones: Gesch et al. (2014) quantified a ~20% reduction in the standard deviation of relative compared to absolute errors in the NED at 1/3 arc second resolution.

Fig. 8 displays the spatial distribution of the WSE errors for the hindcast. Areas of high performance (in white; errors between −0.5 and 0.5 m) are evident across the domain, particularly in Matagorda and Brazoria Counties south of Houston (approx. 29.2°N 95.5°W; to the south of the top-left panel in Fig. 8), Jefferson and Hardin Counties to Houston's east (approx. 30.1°N 94.3°W), and in parts of Houston's Harris County itself (approx. 29.9°N 95.4°W). There is clear clustering to areas of poor performance. Coastal areas in the southwest of the domain around Corpus Christi have a positive bias, indicating the model underpredicted observed HWMs in these areas. This is also the case around Village Creek, north of Beaumont (approx. 30.4°N 94.2°W). Clusters of very high negative errors are present along Houston's Buffalo Bayou (approx. 29.8°N 95.5°W; to the north of the top-left panel in Fig. 8) and the Calcasieu River in Lake Charles (30.2°N 93.2°W; to the east of the top-right panel in Fig. 8), signaling model overprediction of WSEs. Incorporating Fig. 9 into this interpretation, which exhibits the spatial distribution of the DEM errors in Fig. 7c, it may be expected that areas of high WSE error are also areas of high ground elevation error. Observing Figs. 8 and 9 in tandem suggests that this is not the case. Areas of low WSE error to the south of Houston are in fact dominated by relatively high ground elevation error (as colors move away from white towards red and black), while ground elevation errors in areas identified as having high WSE error (Houston and Lake Charles) are low or, if anything, of the opposite sign to the WSE errors. This reinforces the suggestion that relative, and not absolute, DEM accuracy is pre-eminent and draws on the well-documented challenges of hydraulic modelling in urban areas (Yu and Lane, 2006a; Mason et al., 2007; Hunter et al., 2008), especially at large scales (Wing et al., 2017). High WSE errors are generally confined to urban areas in this study, where horizontal, rather than vertical, elevation accuracy (i.e. grid resolution) controls flood model performance to a much greater extent than elsewhere. This is due to the increased prevalence of hydraulically important anthropogenic features (small-scale flow paths, building walls, levees, roads and much else) which are unresolved by the elevation data since they are smaller than the width of a grid cell. In the absence of the
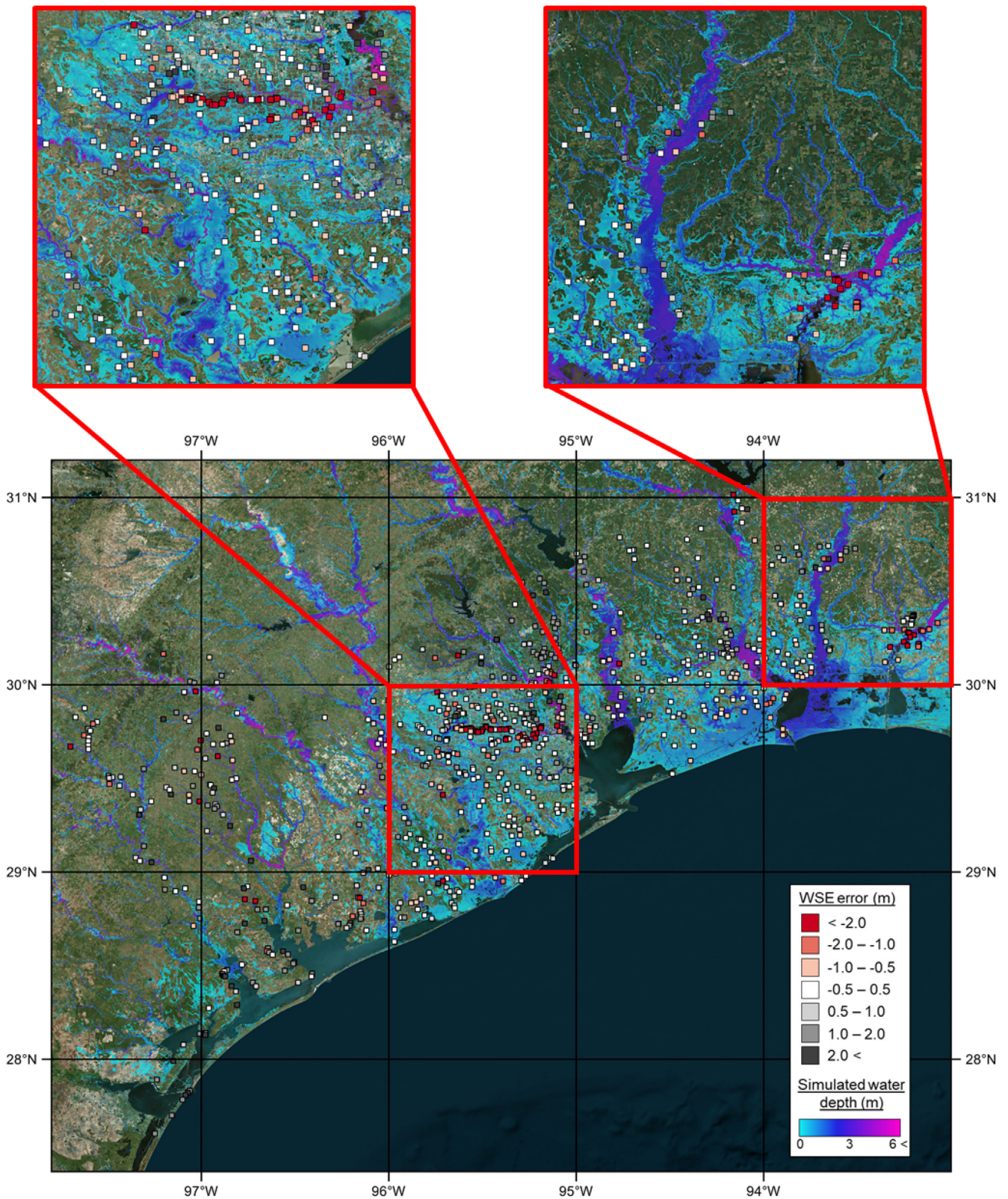
Fig. 8. The model domain containing a grid of hindcast water depths and 1134 points of WSE errors with respect to the surveyed HWMs. The top-left panel contains Houston to the north; the top-right panel contains Lake Charles to the southeast. White and lighter colored points indicate low errors; redder colors indicate model overprediction; greyer colors indicate model underprediction.

computational capacity available to run large-scale models at very high-resolution (e.g. Sampson et al., 2012), which is some distance from feasibility given doubling the granularity of grid resolution increases computation time by an order of magnitude (Savage et al.,

2016), the solution to this problem may lie with nested or variable-resolution grids where high-risk cities can be modelled at finer resolution (e.g. Sanders et al., 2010; Kim et al., 2014; Sanders and Schubert, 2019) or innovative sub-grid scale solutions (e.g. Yu and
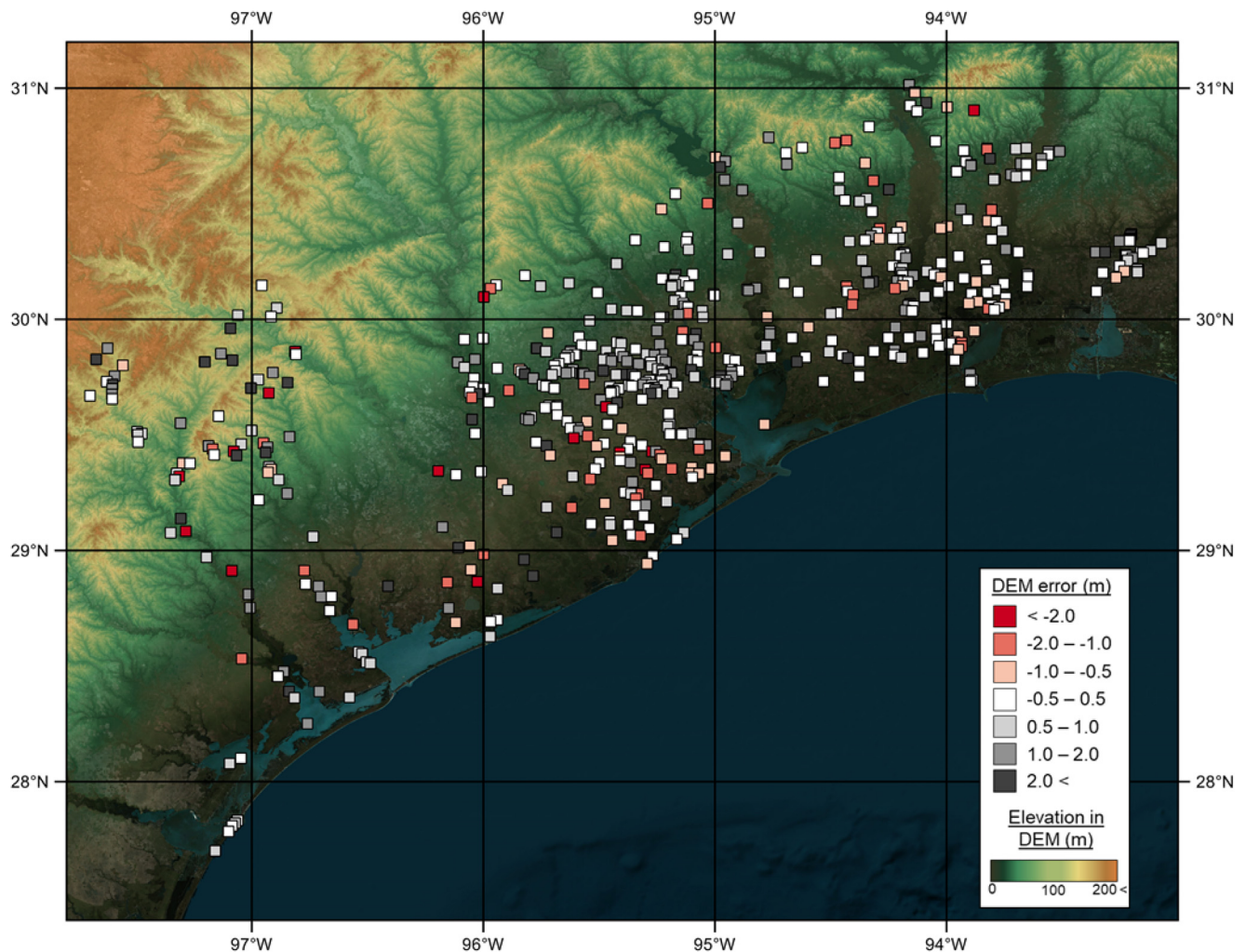
Fig. 9. Elevation differences in the DEM (derived from the USGS National Elevation Dataset) compared to 659 USGS post-event ground surveys. White and lighter colored points indicate similar elevations in the DEM and the survey; redder colors indicate higher elevations in the DEM than those surveyed; greyer colors indicate lower elevations in the DEM than those surveyed.

Lane, 2006b; Sanders et al., 2008; Schubert and Sanders, 2012; Guinot et al., 2017): the bottleneck of the latter approach being obtaining parameterization data at large scale.

4. Conclusions

This paper presents a flood forecasting product for hurricanes in the US that has the potential to be used operationally in the absence of accurate local forecasts, comprehensively testing it against ground truth data collected by the USGS. The framework takes available hydrologic NOAA forecasts as inputs to an existing continental-scale model structure (Wing et al., 2017), accounting for all primary flood drivers, in order to rapidly simulate event water depth grids for a given domain anywhere in the US. Comparing model hindcasts of Hurricane Harvey to benchmark data of its maximum flood extent indicates that the model has skill in picking up the spatial patterns of inundation (mean CSI value of 0.66). When benchmarked against surveyed water surface elevations, the model misestimates this quantity by roughly 1 m on average. This is amidst a backdrop of similar errors in the DEM and up to ~0.5 m of error in the measured HWMs. The model in forecast mode experiences only a moderate drop in performance relative to the hindcast. We conclude this to be commensurate to or beneath likely uncertainties in the preceding components of the forecast cascade, as well as similar to errors in the underlying DEM. With expanded lidar coverage within the USGS NED, the model presented here may

approach the accuracy of operational NOAA inundation forecasts (Adams et al., 2018; Mashriqui et al., 2014) but with total coverage and for a diverse range of flood drivers.

Leading forecast centers generally only produce point water levels and flows (e.g. river discharge, storm surge height), neglecting the crucial receptor component that translates this information to a 2D grid of flood depths in favor of focusing almost exclusively on improved meteorological modelling. Contemporary 1D approaches by NOAA RFCs are likely more accurate and computationally efficient in fluvial settings, but these are only available in limited areas with accurate local data. This paper shows that large-scale hydraulic modelling of fluvial, pluvial and coastal flooding can and should play a role in medium-term forecasts, outperforming simpler GIS-based approaches. With the hydrodynamics of pluvial and coastal flooding from Hurricane Harvey taking only ~6 hours to run on a single node of 20 cores and extractions from a pre-computed library of fluvial flood maps taking only seconds, forecast centers can couple such a module to their existing medium-term forecast frameworks and provide benefits to a plethora of end-users, while sacrificing only a marginal portion of available computation time. The principle of sampling from a pre-computed inventory of flood maps could permit this framework to be applied in a probabilistic ensemble forecast; an important method in accounting for model and exogenous errors. This will be addressed in forthcoming research.

Employing a true hydrodynamic model which properly represents the physics of floodplain flow is shown to outstrip the performance of a

simplified GIS-based approach. While these planar approximations have gained in popularity due to their "quick-and-dirty" solutions to computationally-intensive problems, we have demonstrated that this is not a substitute for 2D hydraulic modelling in this instance. Mass- and momentum- conserving hydraulic codes are shown to be suitably fast and much more accurate than zero-physics approaches, even over large scales where such simulations have historically been intractable.

Looking to the future, improved representation of terrestrial features – through sub-grid parameterization (e.g. Guinot et al., 2017; Sanders and Schubert, 2019), more comprehensive inventories of defense structures (Scussolini et al., 2016) and large-scale acquisition of river channel information from imminent satellite launches (e.g. NASA's Surface Water and Ocean Topography satellite in 2021; Biancamaria et al., 2016) to name but a few projects on the horizon – will herald yet another revolutionary leap in large-scale hydraulic modelling.

5. Data availability

Streamflow forecasts from the NOAA National Water Model are available from https://nomads.ncep.noaa.gov/pub/data/nccf/com/nwm/. NOAA Weather Prediction Center rainfall forecasts can be accessed at https://www.wpc.ncep.noaa.gov/qpf/day1-3.shtml. NOAA National Hurricane Center coastal surge height forecasts are accessible from https://slosh.nws.noaa.gov/psurge2.0/. USGS river gauge observations can be downloaded from https://waterdata.usgs.gov/nwis/rt. Rainfall observations from the NOAA National Weather Service are available at https://water.weather.gov/precip/download.php. Observed coastal surge heights from NOAA Tides and Currents can be accessed at https://tidesandcurrents.noaa.gov/gmap3/. The hydrodynamic model LISFLOOD-FP can be downloaded from http://www.bristol.ac.uk/geography/research/hydrology/models/lisflood/downloads/. Output from the NOAA National Water Center Hurricane Harvey HAND models is available at https://www.hydroshare.org/resource/fe85a680d0144e79b39e8c483dc1e5aa/. USGS observations and benchmark data can be downloaded from https://www.sciencebase.gov/catalog/item/5a85f30fe4b00f54eb36d3a9. The Fathom Harvey depth grids presented in this study are proprietary in nature, but can be made available for academic research purposes (i.e. not commercial, policy or regulatory applications) by contacting Christopher Sampson at Fathom (c.sampson@fathom.global).

Declaration of Competing Interest

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

Acknowledgments

References

Adams, T.E., 2016. Flood forecasting in the United States NOAA/National Weather Service. In: Adams, T.E., Pagano, T.C. (Eds.), Flood Forecasting: A Global Perspective. Academic Press, pp. 249–310.

Adams, T.E., Chen, S., Dymond, R., 2018. Results from operational hydrologic forecasts using the NOAA/~NWS OHRFC Ohio river community HEC-RAS model. J. Hydrologic Eng. 23 (7). https://doi.org/10.1061/(ASCE)HE.1943-5584.0001663.

Addor, N., Jaun, S., Fundel, F., Zappa, M., 2011. An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios. Hydrol. Earth Syst. Sci. 15, 2327–2347. https://doi.org/10.5194/hess-15-2327-2011.

Afshari, S., Tavakoly, A.A., Rajib, M.A., Zheng, X., Follum, M.L., Omranian, E., Fekete,

B.M., 2018. Comparison of new generation low-complexity flood inundation mapping tools with a hydrodynamic model. J. Hydrol. 556, 539–556. https://doi.org/10.1016/j.jhydrol.2017.11.036.

Alfieri, L., Salamon, P., Pappenberger, F., Wetterhall, F., Thielen, J., 2012. Operational early warning systems for water-related hazards in Europe. Environ. Sci. Policy 21, 35–49. https://doi.org/10.1016/j.envsci.2012.01.008.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., Pappenberger, F., 2013. GloFAS – global ensemble streamflow forecasting and flood early warning. Hydrol. Earth Syst. Sci. 17, 1161–1175. https://doi.org/10.5194/hess-17-1161-2013.

Altenau, E.H., Pavelsky, T.M., Bates, P.D., Neal, J.C., 2017. The effects of spatial resolution and dimensionality on modeling regional-scale hydraulics in a multichannel river. Water Resour. Res. 53 (2), 1683–1701. https://doi.org/10.1002/2016WR019396.

Aronica, G., Bates, P.D., Horritt, M.S., 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. Hydrol. Process. 16, 2001–2016. https://doi.org/10.1002/hyp.398.

Balsamo, G., Pappenberger, F., Dutra, E., Viterbo, P., van der Hurk, B., 2011. A revised land hydrology in the ECMWF model: a step towards daily water flux prediction in a fully-closed water cycle. Hydrol. Process. 25 (7), 1046–1054. https://doi.org/10.1002/hyp.7808.

Bates, P.D., De Roo, A.P.J., 2000. A simple raster-based model for flood inundation simulation. J. Hydrol. 236 (1–2), 54–77. https://doi.org/10.1016/S0022-1694(00)00278-X.

Bates, P.D., Dawson, R.J., Hall, J.W., Horritt, M.S., Nicholls, R.J., Wicks, J., Hassan, M.A.A.M., 2005. Simplified two-dimensional numerical modelling of coastal flooding and example applications. Coast. Eng. 52 (9), 793–810. https://doi.org/10.1016/j.coastaleng.2005.06.001.

Bates, P.D., Wilson, M.D., Horritt, M.S., Mason, D.C., Holden, N., Currie, A., 2006. Reach scale floodplain inundation dynamics observed using synthetic aperture radar imagery: data analysis and modelling. J. Hydrol. 328 (1–2), 306–318. https://doi.org/10.1016/j.jhydrol.2005.12.028.

Bates, P.D., Horritt, M.S., Fewtrell, T.J., 2010. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. J. Hydrol. 387 (1–2), 33–45. https://doi.org/10.1016/j.jhydrol.2010.03.027.

Bhola, P.K., Leandro, J., Disse, M., 2018. Framework for offline flood inundation forecasts for two-dimensional hydrodynamic models. Geosciences 8 (9), 346. https://doi.org/10.3390/geosciences8090346.

Biancamaria, S., Lettenmaier, D.P., Pavelsky, T.M., 2016. The SWOT mission and its capabilities for land hydrology. Space Science Series of ISSI In: In: Cazenave, A., Champollion, N., Benveniste, J., Chen, J. (Eds.), Remote Sensing and Water Resources, vol. 55. Springer, Cham, pp. 117–141.

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

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., Savenije, H., 2013. Runoff Prediction in Ungauged Basins. Cambridge University Press, Cambridge, UK doi:10.1017/CBO9781139235761.

Buizza, R., Miller, M., Palmer, T.N., 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Q. J. R. Meteorolog. Soc. 125 (560), 2887–2908. https://doi.org/10.1002/qj.49712556006.

Buizza, R., 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. Mon. Weather Rev. 133 (5), 1076–7097. https://doi.org/10.1175/MWR2905.1.

Cangialosi, J.P., 2018, 2017 Hurricane Season (National Hurricane Center Forecast Verification Report). Retrieved from https://www.nhc.noaa.gov/verification/pdfs/Verification_2017.pdf.

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: a review. J. Hydrol. 375 (3–4), 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005.

Coxon, G., Freer, J., Westerberg, I.K., Wagener, T., Woods, R., Smith, P.J., 2015. A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. Water Resour. Res. 51 (7), 5531–5546. https://doi.org/10.1002/2014WR016532.

de Almeida, G.A.M., Bates, P.D., 2013. Applicability of the local inertial approximation of the shallow water equations to flood modeling. Water Resour. Res. 49 (8), 4833–4844. https://doi.org/10.1002/wrcr.20366.

Di Baldassarre, G., Montanari, A., 2009. Uncertainty in river discharge observations: a quantitative analysis. Hydrol. Earth Syst. Sci. 13, 913–921. https://doi.org/10.5194/hess-13-913-2009.

Di Baldassarre, G., Schumann, G., Bates, P.D., 2009. A technique for the calibration of hydraulic models using uncertain satellite observations of flood extent. J. Hydrol. 367 (3–4), 276–282. https://doi.org/10.1016/j.jhydrol.2009.01.020.

Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F.A., Feyen, L., 2016. Development and evaluation of a framework for global flood hazard mapping. Adv. Water Resour. 94, 87–102. https://doi.org/10.1016/j.advwatres.2016.05.002.

Gesch, D.B., Oimoen, M.J., Evans, G.A., 2014, Accuracy Assessment of the US Geological Survey National Elevation Dataset, and Comparison with Other Large-Area Elevation Datasets—SRTM and ASTER (US Geological Survey Open-File Report 2014-1008, 10 p.). Retrieved from https://pubs.usgs.gov/of/2014/1008/pdf/ofr2014-1008.pdf.

Elvidge, C.D., Tuttle, B.T., Sutton, P.S., Baugh, K.E., Howard, A.T., Milesi, C., Bhaduri, B.L., Nemani, R., 2007. Global distribution and density of constructed impervious surfaces. Sensors 7 (9), 1962–1979. https://doi.org/10.3390/s7091962.

Emanuel, K., 2017. Assessing the present and future probability of Hurricane Harvey's rainfall. PNAS 114 (48), 12681–12684. https://doi.org/10.1073/pnas.1716222114.

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

Fleischmann, A., Paiva, R., Collischonn, W., 2019. Can regional to continental river hy-drodynamic models be locally relevant? A cross-scale comparison. J. Hydrol. X 3, 100027. https://doi.org/10.1016/j.hydroa.2019.100027.

Gochis, D.J., Barlage, M., Dugger, A., Fitzgerald, K., Karsten, L., McAllister, M., McCreight, J., Mills, J., RafieiNasab, A., Read, L., Sampson, K., Yates, D., Yu, W., 2018, The WRF-Hydro modeling system technical description, Version 5.0 (NCAR Technical Note, 107 p.). Retrieved from https://ral.ucar.edu/sites/default/files/public/WRF-HydroV5TechnicalDescription.pdf.

Guinot, V., Sanders, B.F., Schubert, J.E., 2017. Dual integral porosity shallow water model for urban flood modelling. Adv. Water Resour. 103, 16–31. https://doi.org/10.1016/j.advwatres.2017.02.009.

Horritt, M.S., Bates, P.D., 2002. Evaluation of 1D and 2D numerical models for predicting river flood inundation. J. Hydrol. 268 (1–4), 87–99. https://doi.org/10.1016/S0022-1694(02)00121-X.

Horritt, M.S., Bates, P.D., Fewtrell, T.J., Mason, D.C., Wilson, M.D., 2010. Modelling the hydraulics of the Carlisle 2005 flood event. Proc. Inst. Civ. Eng. 163 (WM6), 273–281. https://doi.org/10.1680/wama.2010.163.6.273.

Hunter, N.M., Bates, P.D., Neelz, S., Pender, G., Villanueva, I., Wright, N.G., Liang, D., Falconer, R.A., Lin, B., Waller, S., Crossley, A.J., Mason, D.C., 2008. Benchmarking 2D hydraulic models for urban flooding. Proc. Inst. Civil Eng. – Water Manage. 161 (1), 13–30. https://doi.org/10.1680/wama.2008.161.1.13.

Jelesnianski, C.P., Chen, J., Schaffer, W.A., 1992. SLOSH: Sea, Lake, and Overland Surges from Hurricanes (NOAA Technical Report NWS 48). National Weather Service, Silver Spring, MD.

Kang, N.-Y., Elsner, J.B., 2016. Climate mechanism for stronger typhoons in a warmer world. J. Clim. 29, 1051–1057. https://doi.org/10.1175/JCLI-D-15-0585.1.

Kim, B., Sanders, B.F., Schubert, J.E., Famiglietti, J.S., 2014. Mesh type tradeoffs in 2D hydrodynamic modeling of flooding with a Godunov-based flow solver. Adv. Water Resour. 68, 42–61. https://doi.org/10.1016/j.advwatres.2014.02.013.

Koenig, T.A., Bruce, J.L., O'Connor, J.E., McGee, B.D., Holmes Jr., R.R., Hollins, R., Forbes, B.T., Kohn, M.S., Schellekens, M., Martin, Z.W., Peppler, M.C., 2016, Identifying and Preserving High-Water Mark Data (US Geological Survey Techniques and Methods, book 3, chapter A24, 47 p.). doi:10.3133/tm3A24.

Kossin, J.P., 2018. A global slowdown of tropical-cyclone translation speed. Nature 558, 104–107. https://doi.org/10.1038/s41586-018-0158-3.

Leedal, D., Neal, J., Beven, K., Young, P., Bates, P., 2010. Visualization approaches for communicating real-time flood forecasting level and inundation information. J. Flood Risk Manage. 3 (2), 140–150. https://doi.org/10.1111/j.1753-318X.2010.01063.x.

Lehner, B., Verdin, K., Jarvis, A., 2008. New global hydrography derived from spaceborne elevation data. Eos, Trans., Am. Geophys. Union 89 (10), 93–94. https://doi.org/10.1029/2008EO100001.

Lehner, B., Grill, G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrol. Process. 27, 2171–2186. https://doi.org/10.1002/hyp.9740.

Leskens, J.G., Brugnach, M., Hoekstra, A.Y., Schuurmans, W., 2014. Why are decisions in flood disaster management so poorly supported by information from flood models? Environ. Modell. Software 53, 53–61. https://doi.org/10.1016/j.envsoft.2013.11.003.

Liu, Y.Y., Maidment, D.R., Tarboton, D.G., Zheng, X., Wang, S., 2018. A CyberGIS in-tegration and computation framework for high-resolution continental-scale flood inundation mapping. J. Am. Water Resour. Assoc. 54 (4), 770–784. https://doi.org/10.1111/1752-1688.12660.

Maidment, D.R., 2016. Conceptual framework for the national flood interoperability experiment. J. Am. Water Resour. Assoc. 53 (2), 245–257. https://doi.org/10.1111/1752-1688.12474.

Mashriqui, H.S., Halgren, J.S., Reed, S.M., 2014. 1D river hydraulic model for operational flood forecasting in the tidal Potomac: evaluation for freshwater, tidal, and wind-driven events. J. Hydraulic Eng. 140 (5). https://doi.org/10.1061/(ASCE)HY.1943-7900.0000862.

Maskell, J., Horsburgh, K., Lewis, M., Bates, P., 2014. Investigating river – surge inter-action in idealised estuaries. J. Coastal Res. 30 (2), 248–259. https://doi.org/10.2112/JCOASTRES-D-12-00221.1.

Mason, D.C., Horritt, M.S., Hunter, N.M., Bates, P.D., 2007. Use of fused airborne scan-ning laser altimetry and digital map data for urban flood modelling. Hydrol. Process. 21 (11), 1436–1447. https://doi.org/10.1002/hyp.6343.

Mignot, E., Paquier, A., Haider, S., 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.

Moriasi, D.N., Arnold, J.G., van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in wa-tershed simulations. Trans. Am. Soc. Agric. Biolog. Eng. 50 (3), 885–900.

Morin, J., Benyamini, Y., 1977. Rainfall infiltration into bare soils. Water Resour. Res. 13 (5), 813–817. https://doi.org/10.1029/WR013i005p00813.

Munich Re, 2018a, NatCatSERVICE. Retrieved from https://natcatservice.munichre.com/.

Munich Re, 2018b, Nature catastrophes 2017 (TOPICS Geo). Retrieved from https://www.munichre.com/content/dam/assets/munichre/content-pieces/documents/pdf/302-09092_en.pdf.

Musser, J.W., Watson, K.M., Painter, J.A., Gotvald, A.J. 2016, Flood-inundation maps of selected areas affected by the flood of October 2015 in central and coastal South Carolina (US Geological Survey Open-File Report 2016-1019, 81 p.). doi:10.3133/ofr20161019.

Narayan, S., Hanson, S., Nicholls, R.J., Clarke, D., Willems, P., Ntegeka, V., Mobaliu, J., 2012. A holistic model for coastal flooding using system diagrams and the Source-Pathway-Receptor (SPR) concept. Nat. Hazards Earth Syst. Sci. 12, 1431–1439. https://doi.org/10.5194/nhess-12-1431-2012.

Neal, J.C., Bates, P.D., Fewtrell, T.J., Hunter, N.M., Wilson, M.D., Horritt, M.S., 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.

Neal, J., Schumann, G., Bates, P., 2012. A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. Water Resour. Res. 48 (11), W11506. https://doi.org/10.1029/2012WR012514.

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

Niu, G.-Y., Yang, Z.-L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., Xia, Y., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model de-scription and evaluation with local-scale measurements. J. Geophys. Res. 116, D12109. https://doi.org/10.1029/2010JD015139.

NOAA National Water Center, Boghici, E., Arctur, D., 2018, NOAA NWC – Harvey NWM-HAND Flood Extents (HydroShare). doi:10.4211/hs.fe85a680d0144e79-b39e8c483dc1e5aa.

Nobre, A.D., Cuartas, L.A., Hodnett, M., Rennó, C.D., Rodrigues, G., Silveira, A., Waterloo, M., Saleska, S., 2011. Height above the nearest drainage – a hydrologically relevant new terrain model. J. Hydrol. 404 (1–2), 13–29. https://doi.org/10.1016/j.jhydrol.2011.03.051.

Nobre, A.D., Cuartas, L.A., Momo, M.R., Severo, D.L., Pinheiro, A., Nobre, C.A., 2016. HAND contour: a new proxy predictor of inundation extent. Hydrol. Process. 30 (2), 320–333. https://doi.org/10.1002/hyp.10581.

Paiva, R.C.D., Buarque, D.C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., Mendes, C.A.B., 2013. Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. Water Resour. Res. 49 (3), 1226–1243. https://doi.org/10.1002/wrcr.20067.

Pappenberger, F., Beven, K.J., Hunter, N.M., Bates, P.D., Gouweleeuw, B.T., Thielen, J., de Roo, A.P.J., 2005. Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS). Hydrol. Earth Syst. Sci. 9 (4), 381–393. https://doi.org/10.5194/hess-9-381-2005.

Pappenberger, F., Frodsham, K., Beven, K., Romanowicz, R., Matgen, P., 2007. Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations. Hydrol. Earth Syst. Sci. 11, 739–752. https://doi.org/10.5194/hess-11-739-2007.

Pappenberger, F., Cloke, H.L., Balsamo, G., Ngo-Duc, T., Oki, T., 2010. Global runoff routing with the hydrological component of the ECMWF NWP system. Int. J. Climatol. 30 (14), 2155–2174. https://doi.org/10.1002/joc.2028.

Pitt, M., 2008. Learning Lessons from the 2007 Floods: An Independent Review by Sir Michael Pitt. Cabinet Office, London, UK.

Price, D., Hudson, K., Boyce, G., Schellekens, J., Moore, R.J., Clark, P., Harrison, T., Connolly, E., Pilling, C., 2012. Operational use of a grid-based model for flood forecasting. Proc. Inst. Civil Eng. 165 (WM2), 65–77. https://doi.org/10.1680/wama.2012.165.2.65.

Quinn, N., Lewis, M., Wadley, M.P., Haigh, I.D., 2014. Assessing the temporal variability in extreme storm-tide time series for coastal flood risk assessment. J. Geophys. Res. 119 (8), 4983–4998. https://doi.org/10.1002/2014JC010197.

Rennó, C.D., Nobre, A.D., Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J., Waterloo, M.J., 2008. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. Remote Sensing Environ. 112 (9), 3469–3481. https://doi.org/10.1016/j.rse.2008.03.018.

Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of dif-ferent types of hydrological models. Water Resour. Res. 32 (7), 2189–2202. https://doi.org/10.1029/96WR00896.

Ritter, A., Muñoz-Carpena, R., 2013. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. J. Hydrol. 480, 33–45. https://doi.org/10.1016/j.jhydrol.2012.12.004.

Rodda, H.J.E., 2005. The development and application of a flood risk model for the Czech Republic. Nat. Hazards 36 (1–2), 207–220. https://doi.org/10.1007/s11069-004-4549-4.

Sampson, C.C., Fewtrell, T.J., Duncan, A., Shaad, K., Horritt, M.S., Bates, P.D., 2012. Use of terrestrial laser scanning data to drive decimetric resolution urban inundation models. Adv. Water Resour. 41, 1–17. https://doi.org/10.1016/j.advwatres.2012.02.010.

Sampson, C.C., Bates, P.D., Neal, J.C., Horritt, M.S., 2013. An automated routing meth-odology to enable direct rainfall in high resolution shallow water models. Hydrol. Process. 27 (3), 467–476. https://doi.org/10.1002/hyp.9515.

Sampson, C.C., Smith, A.M., Bates, P.D., Neal, J.C., Alfieri, L., Freer, J.E., 2015. A high-resolution global flood hazard model. Water Resour. Res. 51 (9), 7358–7381. https://doi.org/10.1002/2015WR016954.

Sanders, B.F., Schubert, J.E., Gallegos, H.A., 2008. Integral formulation of the shallow-water equations with anisotropic porosity for urban flood modelling. J. Hydrol. 362 (1–2), 19–38. https://doi.org/10.1016/j.jhydrol.2008.08.009.

Sanders, B.F., Schubert, J.E., Detwiler, R.L., 2010. ParBreZo: A parallel, unstructured grid, Godunov-type, shallow-water code for high-resolution flood inundation modeling at the regional scale. Adv. Water Resour. 33 (12), 1456–1467. https://doi.org/10.1016/j.advwatres.2010.07.007.

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

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

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

Schumann, G., Matgen, P., Hoffmann, L., Hostache, R., Pappenberger, F., Pfister, L., 2007. Deriving distributed roughness values from satellite radar data for flood inundation modelling. J. Hydrol. 344 (1–2), 96–111. https://doi.org/10.1016/j.jhydrol.2007.06.024.

Schumann, G.J.-P., Neal, J.C., Voisin, N., Andreadis, K.M., Pappenberger, F., Phanthuwongpakdee, N., Hall, A.C., Bates, P.D., 2013. A first large-scale flood inundation forecasting model. Water Resour. Res. 49, 6248–6257. https://doi.org/10.1002/wrcr.20521.

Scussolini, P., Aerts, J.C.J.H., Jongman, B., Bouwer, L.M., Winsemius, H.C., de Moel, H., Ward, P.J., 2016. FLOPROS: an evolving database of flood protection standards. Nat. Hazards Earth Syst. Sci. 16, 1049–1061. https://doi.org/10.5194/nhess-16-1049-2016.

Smith, R.A.E., Bates, P.D., Hayes, C., 2012. Evaluation of a coastal flood inundation model using hard and soft data. Environ. Modell. Software 30, 35–46. https://doi.org/10.1016/j.envsoft.2011.11.008.

Smith, A., Sampson, C., Bates, P., 2015. Regional flood frequency analysis at the global scale. Water Resour. Res. 51 (1), 539–553. https://doi.org/10.1002/2014WR015814.

Sobel, A.H., Camargo, S.J., Hall, T.M., Lee, C.-Y., Tippett, M.K., Wing, A.A., 2016. Human influence on tropical cyclone intensity. Science 353 (6296), 242–246. https://doi.org/10.1126/science.aaf6574.

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

Thielen, J., Bartholmes, J., Ramos, M.-H., de Roo, A., 2009. The European Flood Alert System – Part 1: concept and development. Hydrol. Earth Syst. Sci. 13, 125–140. https://doi.org/10.5194/hess-13-125-2009.

van der Knijff, J.M., Younis, J., de Roo, A.P.J., 2010. LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. Int. J. Geograph. Inform. Sci. 24 (2), 189–212. https://doi.org/10.1080/13658810802549154.

van Oldenborgh, G.J., van der Wiel, K., Sebastian, A., Singh, R., Arrighi, J., Otto, F., Haustein, K., Li, S., Vecchi, G., Cullen, H., 2017. Attribution of extreme rainfall from Hurricane Harvey, August 2017. Environ. Res. Lett. 12 (12), 124009. https://doi.org/10.1088/1748-9326/aa9ef2.

Ward, P.J., Jongman, B., Aerts, J.C.J.H., Bates, P.D., Botzen, W.J.W., Loaiza, A.D., Hallegatte, S., Kind, J.M., Kwadijk, J., Scussolini, P., Winsemius, H., 2017. A global framework for future costs and benefits of river-flood protection in urban areas. Nat. Clim. Change 7, 642–646. https://doi.org/10.1038/nclimate3350.

Watson, K.M., Harwell, G.R., Wallace, D.S., Welborn, T.L., Stengel, V.G., McDowell, J.S., 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 p.). doi:10.3133/sir20185070.

Westerberg, I.K., Wagener, T., Coxon, G., McMillan, H.K., Castellarin, A., Montanari, A., Freer, J., 2016. Uncertainty in hydrological signatures for gauged and ungauged catchments. Water Resour. Res. 52 (3), 1847–1865. https://doi.org/10.1002/2015WR017635.

Wilson, M., Bates, P., Alsdorf, D., Forsberg, B., Horritt, M., Melack, J., Frappart, F., Famiglietti, J., 2007. Modeling large-scale inundation of Amazonian seasonally flooded wetlands. Geophys. Res. Lett. 34 (15). https://doi.org/10.1029/2007GL030156.

Wing, O.E.J., Bates, P.D., Sampson, C.C., Smith, A.M., Johnson, K.A., Erickson, T.A., 2017. Validation of a 30 m resolution flood hazard model of the conterminous United States. Water Resour. Res. 53 (9), 7968–7986. https://doi.org/10.1002/2017WR020917.

Winsemius, H., Aerts, J.C.J.H., van Beek, L.P.H., Bierkens, M.F.P., Bouwman, A., Jongman, B., Kwadijk, J.C.J., Ligtvoet, W., Lucas, P.L., van Vuuren, D.P., Ward, P.J., 2016. Global drivers of future river flood risk. Nat. Clim. Change 6, 381–385. https://doi.org/10.1038/nclimate2893.

Wood, M., Hostache, R., Neal, J., Wagener, T., Giustarini, L., Chini, M., Corato, G., Matgen, P., Bates, P., 2016. Calibration of channel depth and friction parameters in the LISFLOOD-FP hydraulic model using medium-resolution SAR data and identifiability techniques. Hydrol. Earth Syst. Sci. 20, 4983–4997. https://doi.org/10.5194/hess-20-4983-2016.

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

Yu, D., Lane, S.N., 2006b. Urban fluvial flood modelling using a two-dimensional diffusion-wave treatment, part 2: development of a sub-grid-scale treatment. Hydrol. Process. 20 (7), 1567–1583. https://doi.org/10.1002/hyp.5936.

Zhang, J., Huang, Y.-F., Munasinghe, D., Fang, Z., Tsang, Y.-P., Cohen, S., 2018. Comparative analysis of inundation mapping approaches for the 2016 flood in the Brazos River, Texas. J. Am. Water Resour. Assoc. 54 (4), 820–833. https://doi.org/10.1111/1752-1688.12623.

Zheng, X., Maidment, D.R., Tarboton, D.G., Liu, Y.Y., Passalacqua, P., 2018a. GeoFlood: large-scale flood inundation mapping based on high-resolution terrain analysis. Water Resour. Res. 54. https://doi.org/10.1029/2018WR023457.

Zheng, X., Tarboton, D.G., Maidment, D.R., Liu, Y.Y., Passalacqua, P., 2018b. River channel geometry and rating curve estimation using height above the nearest drainage. J. Am. Water Resour. Assoc. 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). Retrieved from Texas ScholarWorks (https://repositories.lib.utexas.edu/bitstream/handle/2152/68178/ZHENG-THESIS-2018.pdf). Austin, TX: University of Texas at Austin.