

Water Resources Research



RESEARCH ARTICLE

10.1029/2019WR025957

Key Points:

- Flood defense representation is presently poor in large-scale flood models, impairing their ability to map flood hazard accurately
- A new method is presented, which automatically identifies hydraulic structures in terrain data and accurately preserves their elevations
- Hydraulic simulations where a lack of defense data is the dominant error show significant improvements in skill when incorporating this method

Correspondence to:

O. E. J. Wing,
oliver.wing@bristol.ac.uk

Citation:

Wing, O. E. J., Bates, P. D., Neal, J. C., Sampson, C. C., Smith, A. M., Quinn, N., et al. (2019). A New Automated Method for Improved Flood Defense Representation in Large-Scale Hydraulic Models. *Water Resources Research*, 55. <https://doi.org/10.1029/2019WR025957>

Received 9 JUL 2019

Accepted 16 OCT 2019

Accepted article online 11 NOV 2019

A New Automated Method for Improved Flood Defense Representation in Large-Scale Hydraulic Models

Oliver E. J. Wing^{1,2} , Paul D. Bates^{1,2} , Jeffrey C. Neal^{1,2} , Christopher C. Sampson², Andrew M. Smith², Niall Quinn² , Iuliia Shustikova³, Alessio Domeneghetti³, Daniel W. Gilles⁴, Radoslaw Goska⁴, and Witold F. Krajewski⁴

¹School of Geographical Sciences, University of Bristol, Bristol, UK, ²Fathom, Bristol, UK, ³DICAM, University of Bologna, Bologna, Italy, ⁴IHR-Hydrosience & Engineering, The University of Iowa, Iowa City, IA, USA

Abstract The execution of hydraulic models at large spatial scales has yielded a step change in our understanding of flood risk. Yet their necessary simplification through the use of coarsened terrain data results in an artificially smooth digital elevation model with diminished representation of flood defense structures. Current approaches in dealing with this, if anything is done at all, involve either employing incomplete inventories of flood defense information or making largely unsubstantiated assumptions about defense locations and standards based on socioeconomic data. Here, we introduce a novel solution for application at scale. The geomorphometric characteristics of defense structures are sampled, and these are fed into a probabilistic algorithm to identify hydraulically relevant features in the source digital elevation model. The elevation of these features is then preserved during the grid coarsening process. The method was shown to compare favorably to surveyed U.S. levee crest heights. When incorporated into a continental-scale hydrodynamic model based on LISFLOOD-FP and compared to local flood models in Iowa (USA), median correspondence was 69% for high-frequency floods and 80% for low-frequency floods, approaching the error inherent in quantifying extreme flows. However, improvements versus a model with no defenses were muted, and risk-based deviations between the local and continental models were large. When simulating an event on the Po River (Italy), built and tested with higher quality data, the method outperformed both undefended and even engineering-grade models. As such, particularly when employed alongside model components of commensurate quality, the method here generates improved-accuracy simulations of flood inundation.

Plain Language Summary Traditional flood risk assessments are carried out using computer models built with local data, but their spatial coverage is impaired by how expensive and time-consuming they are to produce. Recent advances in data availability, understanding of necessary physical process representation, and computational capacity have enabled hydraulic models of the entire globe to be built in an automated fashion at a fraction of the financial and human cost. However, their accuracy can be significantly impaired by a lack of information on flood defenses. As the model is built, elevation data are coarsened to reduce the number of calculations required to simulate flooding over such wide areas. This results in flood defense structures being smoothed out of the terrain information used in the model. Publicly available defense inventories are of insufficient coverage to ameliorate this issue. In this paper, a method is presented, which automatically detects levee-like features in high-resolution elevation data and accurately represents their heights during this necessary coarsening process. Simulating flood inundation over this “defended” topography results in high correspondence between local models and observations for test cases in the United States and Italy, with improvements particularly felt where a lack of defense information is the dominant source of error.

1. Introduction

The last decade has seen a revolution in the field of flood inundation modeling. Historically, hydraulicians have focused on custom-building local models of individual river reaches, but the dual effect of enhanced computational capacity and the advent of “big data” have expanded the size of model domains considered to entire regions, continents, and even the globe (Dottori et al., 2016; Sampson et al., 2015; Wilson et al., 2007; Wing et al., 2017; Winsemius et al., 2013; Yamazaki et al., 2011). However, a significant performance gap exists between small- and large-scale models. Event-replicating local models can closely match the

©2019. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

real-world observations they attempt to emulate (e.g., Bates et al., 2006; Mignot et al., 2006; Neal et al., 2009), while global-scale models deviate significantly from flood extent observations, local models, and even each other (e.g., Dottori et al., 2016; Sampson et al., 2015; Trigg et al., 2016; Ward et al., 2017). Regional- and continental-scale models—with substantial, but not worldwide, model domains usually affording them more localized, higher quality data—outperform global models (e.g., Fleischmann et al., 2019; Wilson et al., 2007; Wing et al., 2017), but the “spatial scale performance gap” is still evident.

Attempts to achieve performance parity between models of all spatial scale are well underway in the field. These generally involve emulating the features of local-scale modeling strategies that enable their accurate simulation of flood inundation. This typically requires manual interventions in the model-building process by a skilled practitioner and the collection of accurate local information: neither of which are possible when operating at large spatial scales. Viable alternatives to local-scale approaches that can be employed in a larger scale model therefore must be available at regional to global scales and incorporated into an automated model-building process. Much of the research into these alternatives is either in or approaching a state of maturity: related to the processing of elevation data, improvements to computational hydraulics, representation of rivers, characterization of the extreme flow inputs, and model validation. While calls for industry, governments, and nongovernmental organizations to support the development of a publicly available, global Digital Elevation Model (DEM) built from laser altimetry (LiDAR) data are ongoing (Sampson et al., 2016; Schumann, 2014; Winsemius et al., 2019), corrections to errors in spaceborne radar-based DEMs have made such data fit for the purpose in flood modeling at large scales (e.g., Archer et al., 2018; Yamazaki et al., 2017). The development of simplified hydraulic codes, which ensure that physical process representation is commensurate with data availability and computational burden (Hunter et al., 2007), has enabled the rapid spatial “scale-up” seen in the past decade (e.g., Bates et al., 2010; Lamb et al., 2009; Neal et al., 2018; Sanders & Schubert, 2019). The delineation of channel networks from global DEMs (e.g., Allen & Pavelsky, 2018; Lehner et al., 2008) and approximations regarding their geometry (e.g., Frasson et al., 2019) ensure the necessary channel representation (Neal et al., 2012) in hydraulic models amidst data scarcity at large spatial scales. Characterizing extreme flows, given the paucity of river gauges in time and space, is commonly performed by cascading climate reanalysis data through a hydrological model (e.g., Dottori et al., 2016; Pappenberger et al., 2012) or by forging statistical relationships between catchment characteristics and gauge data and transferring these to ungauged regions (e.g., Smith et al., 2015). Finally, the validation of these model structures, inhibited by the scarcity of benchmark data at a commensurate spatial scale, is becoming more common place (e.g., Dottori et al., 2016; Ward et al., 2017; Wing et al., 2017). In contrast, a critical component of large-scale hydraulic models that has received relatively little attention is the representation of flood defense structures within them.

Information on the location of flood defense structures on the floodplain (specifically levees, embankments, dikes, and flood walls: these terms are used interchangeably hereafter) at large spatial scales is very scarce; metadata regarding their defense standards or crest heights are scarcer still. The horizontal resolution of the elevation data employed in large-scale flood models is typically too low to fully capture the effect of such structures. Even if the source elevation data are higher resolution (e.g., airborne LiDAR surveys, which are usually available on 1- to 5-m grids), the resampling process required to reduce the resolution to a scale tractable for large-scale simulations imparts a smoothing effect that reduces the crest height of levees, often to the point of their disappearance. In light of this, the explicit effect of flood defenses is virtually ignored by large-scale flood hazard modelers. If accounted for at all, many modelers treat flood defenses as an afterthought by simulating the undefended case only and assuming floods smaller than a specified magnitude are unimportant when calculating risk (Alfieri et al., 2017; Dottori et al., 2018; Feyen et al., 2012; Quinn et al., 2019; Winsemius et al., 2013). FLOPROS (Scussolini et al., 2016) is a commonly used data set, which provides design, policy, and modeled flood defense standards globally at the subcountry level to inform the defense threshold above which floods cause damage. It is, however, not justified to assume that defense standards remain constant across such vast areas nor that every river within a region is afforded some standard of protection. Furthermore, representing the hydraulic effect of levees during flood events is shown to have a significant effect on peak flows, both upstream (Heine & Pinter, 2012) and downstream (Castellari et al., 2011; Di Baldassarre et al., 2010) of the levee, and the assumption that an undefended simulation behaves in the same way as in a simulation where a levee has been overtopped is not valid (e.g., Ciullo et al., 2019; Masoero et al., 2013).

The continental-scale model presented in Wing et al. (2017) incorporated the U.S. Army Corps of Engineers (USACE) National Levee Database (NLD) into its structure, though a report by the American Society of Civil Engineers (2017) indicated that only 48,000 of an estimated 160,000 km (30%) of the nation's levees are contained in the USACE database. Crest heights and defense standards are only sporadically specified also, meaning assumptions must be made to fill these gaps: only knowing the location of a levee is not enough information for its explicit inclusion in a dynamic simulation. "Burning" crest heights into the DEM is not straightforward though, owing to the specification of how the levee metadata interact with the terrain data. Datum conflicts, geolocation errors, and the effect of residual levee artifacts in the baseline topographic data can result in an improper representation of defense information, even in the relatively few locations where such data are available. If defense standards are specified (e.g., as an annual exceedance probability [AEP] flood that the structure defends against), to incorporate these standards in a hydrodynamic simulation, a crest elevation will still be required. For the model structure in Wing et al. (2017), this can be derived by linking the AEP defense standard to the bank height of the river channel (Neal et al., 2012) via the flood frequency analysis, which determines particular AEP discharges (Smith et al., 2015). Although computationally efficient, this process ignores the defensive potential of lateral floodwater storage between the channel and the levee, which in many cases can be substantial (Hooijer et al., 2004). The solution to this employed in Wing et al. (2017) is to run a preliminary "defense height estimation" simulation for certain AEP design floods, tracking the height of water at the levee location and elevating the crests to this for the primary model run. The upshot of these issues—most overwhelmingly, the 30% levee capture rate—is impaired performance in the Wing et al. (2017) model. For this model, false alarms were reportedly higher in urban, rather than rural, areas, resulting in lower model skill in the very areas where accurate risk calculations are most urgently required, with the likely culprit being inadequate defense representation leading to inundation in areas that are protected in reality.

Amidst incomplete defense inventories, some modelers make assumptions about where levees are likely to be situated based on socioeconomic or land-use data; the general idea being that wealthier, built-up areas are afforded a higher standard of protection compared with the inverse case (Feyen et al., 2012; Quinn et al., 2019; Sampson et al., 2015). Using the United States as an example, we demonstrate that such assumptions are generally invalid. The USACE NLD contains a wealth of levee data for the contiguous United States (CONUS), including ~200,000 km² of land specified as being defended from flooding. Sampling the characteristics of these lands in terms of degree of urbanity (from the National Land Cover Database), wealth (median household income from the U.S. Census Bureau), and government spending (USACE spending under the American Recovery and Reinvestment Act of 2009) offers insights into whether a crude predictive defense model based on nationwide socioeconomic characteristics is valid. The left-hand side of Figure 1 illustrates the correlation between the defense standard of the levee protecting an area of land and its characteristics in terms of urbanity, wealth, and spending, respectively. It is clear from Figure 1a that, of areas that are offered some degree of structural flood protection, urban areas are not protected to a higher standard than less developed areas. Similarly, defended wealthier neighborhoods are not offered greater protection than defended poorer ones (Figure 1b). Indeed, very high return period levees (i.e., those that defend against the one in 1,000-year flood) appear to defend poorer areas. Further, higher USACE spending since 2009 does not appear to be correlated with defense standard: if anything, the opposite case is true (Figure 1c). The histograms on the right-hand side of Figure 1 show the distribution of these characteristics within defended areas and in the remaining U.S. land surface outside of them. The distribution of urbanity (Figure 1d), wealth (Figure 1e), and spending (Figure 1f) is virtually indistinguishable between defended and undefended lands, indicating such variables would be inappropriate for use in a statistical model to predict the location and standard of defenses.

In this paper, we propose a novel advance in defense representation capable of integration in automated large-scale model-building frameworks. Using the Fathom-US model (Wing et al., 2017) as a springboard from which to make these advances, we take the source data of the DEM employed and apply an algorithm to preserve levee crest heights during the DEM resampling process. This DEM coarsening is a necessary step in ensuring that the model grid scale is computationally tractable given the vast spatial coverage of the model domain. To prevent loss of information during this process, the method automatically ensures that flood defense structures remain represented in the terrain data employed. The supposition that levees can be detected based on their topographic signature is not a new one: flood defenses have been extracted

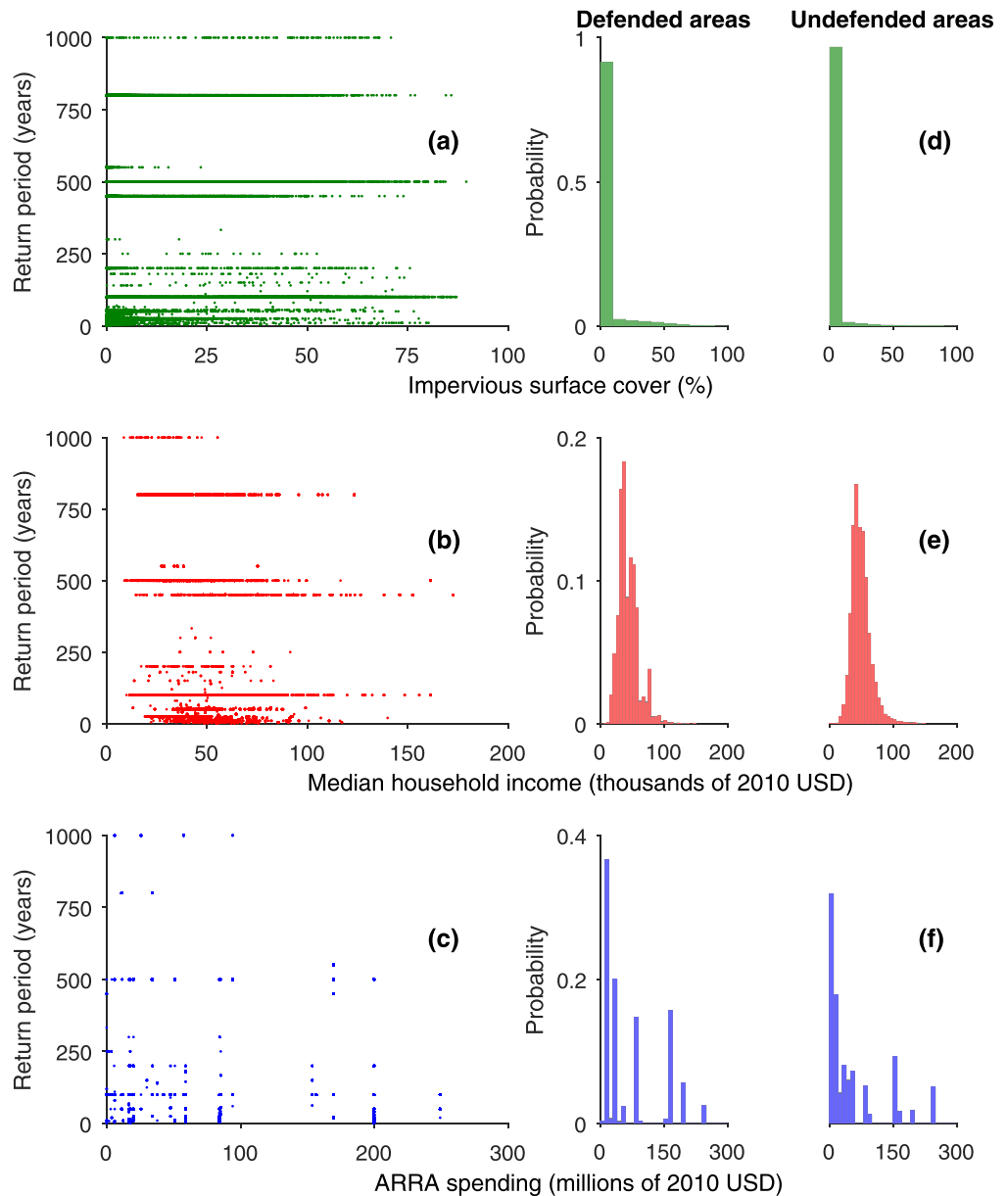


Figure 1. The relationship between U.S. Army Corps of Engineers National Levee Database defense standard and (a) urbanity, (b) wealth, and (c) spending, as well as the distribution of (d) urbanity, (e) wealth, and (f) spending within and outside U.S. Army Corps of Engineers National Levee Database defended areas.

from DEMs in a number of studies (Bailly et al., 2008; Casas et al., 2012; Cazorzi et al., 2013; Choung, 2014; Krüger, 2010; Krüger & Meinel, 2008; Sofia et al., 2014; Steinfeld et al., 2013), while feature extraction from DEMs more generally is a relatively well-established field (e.g., Lashermes et al., 2007; Passalacqua et al., 2010; Passalacqua et al., 2012; Sofia et al., 2011; Tarolli et al., 2012). The characteristics of levees with relation to elevation and its derivatives (Evans, 1979, 1980; Wood, 1996) are distinct, meaning they can be described by geomorphometric parameters (after Sofia et al., 2014). All previous studies are broadly similar in their characterization of levees with such parameters. To actually isolate levee features though, qualitative descriptions of these elevation-derived parameters (e.g., “linear features of high elevation bounded by two opposing steep slopes”) must be translated to quantitative definitions of the parameter thresholds at which certain DEM pixels are considered relevant. It is here that previously cited studies fall short: their choice of deterministic geomorphometric parameter thresholds relevant only to the geographical domain and grid resolution of each study-specific DEM limits the applicability

of their findings more widely (e.g., Choung, 2014; Krüger, 2010; Krüger & Meinel, 2008; Sofia et al., 2014). Given these studies (a) employed LiDAR-derived DEMs with resolutions of the order 10^0 m, which are currently unavailable at large spatial scales; (b) were of small, isolated test cases; (c) offered little indication regarding the computational feasibility of applying such methods at larger scales; and (d) were not extended to analyze fitness for purpose in a hydraulic modeling context; such methods leave crucial research questions unanswered. Other novel approaches, including those with a postprocessing step to fill-in line breaks in DEM-based approaches (e.g., Choung, 2014), those related to processing image spectra, texture, and shape (e.g., Steinfeld et al., 2013), and tracking observed flood edges alongside river gauge data to infer levee presence (e.g., Wood et al., 2018), are heavily impaired by data availability and only delineate the location of levees. For use in flood modeling, a view of crest elevation or defense standard is required as well. As such, the method presented here adheres to a DEM-based approach since (a) elevation data are available at large spatial scales, albeit at a coarser resolution than in previous levee detection studies, and (b) detection this way implicitly extracts a crest elevation in the form of peaks in the DEM, avoiding the need for further processing or assumptions to implement the results in a hydraulic model.

We initially test the accuracy of this approach by comparing a new “defended” DEM to surveyed levee crest elevations in the State of California, calculating whether the method captures crest heights more accurately than standard DEM resampling approaches. We then incorporate this approach into the Fathom-US hydrodynamic modeling framework and compare hydraulic simulations to a broad-scale amalgamation of high-quality local models in the State of Iowa, charting the degree to which local- and large-scale models converge in their simulation of design AEP flood events with improved flood defense representation. We finally then validate the application of the method in a high-quality real event simulation of the well-defended Po River floodplain in northern Italy.

2. Data and Methods

For this study, the test area is the entire CONUS, owing to the availability of total coverage elevation data from the U.S. Geological Survey (USGS) National Elevation Dataset (NED) at 1/3 arc second (~ 10 m) and partial coverage at 1/9 arc second (~ 3 m). The NED is based on LiDAR data for 39% of the CONUS, containing 67% of its population. The hydraulic model structure that this levee extraction method will be integrated within is Fathom-US (Wing et al., 2017), which uses the NED at 1 arc second (~ 30 m) as its DEM. Halving the grid resolution increases computation time by an order of magnitude (Savage et al., 2016), making running Fathom-US on the higher resolution NED variants computationally intractable at national scales. The proposed method intends to preserve hydraulically important information at these higher resolutions for use within the 1 arc-second model structure. Fathom-US hydraulics are based on a version of the LISFLOOD-FP numerical scheme, a 2-D simplification of the shallow water equations which approximates local inertia (Bates et al., 2010; de Almeida & Bates, 2013). AEP flows, based on the regional flood frequency analysis of Smith et al. (2015) using USGS river gauges, are routed through 1-D subgrid channels (Neal et al., 2012) which are derived from HydroSHEDS hydrography data (Lehner & Grill, 2013), and across NED-derived floodplains. Fathom-US also has a pluvial model component, where the direct effect of intense rainfall onto the land surface is simulated. Boundary conditions here are informed by National Oceanic and Atmospheric Administration intensity-duration-frequency relationships. Further details are available in Wing et al. (2017) and Sampson et al. (2015).

Five geomorphometric parameters relevant for the identification of levees were sampled from the 1/3 and 1/9 arc-second NED variants: relative elevation, slope, aspect, profile curvature, and planform curvature. The crucial component of this method is the automated sampling of the parameter thresholds from known levee locations, essentially “training” the extraction algorithm against ground truths derived from the USACE NLD. This is another reason why the CONUS is an ideal test bed, since it contains wide-area levee information in the form of the NLD. While not containing every levee in the United States, the database contains 48,000 km of flood defense locations. Geomorphometric parameters inherent in the NED at NLD locations form the parameter thresholds required to identify levees elsewhere in the NED which are not in the NLD. The algorithm inevitably captures any features within the NED that exhibit the geomorphometric characteristics of levees in the NLD, meaning other “informal” features, which may still be of hydraulic relevance, are also captured. In recognition that no single set of parameter thresholds will adequately

capture levee or levee-like features in the elevation data, a random sample ($n = 1,000$, in this case) of the geomorphometric characteristics evident in the NLD drives the extraction algorithm 1,000 times. This number is likely in excess of what is required to ensure beyond-adequate sampling of the parameter space. This generates a pseudo-probabilistic surface, assigning each DEM pixel an “extraction rate” (ϵ) out of 1,000. Specifically, ϵ is the number of times a pixel exceeds all five thresholds of a given parameter set. Where ϵ is greater than a given threshold (between 0 and 1,000), its elevation value will become the corresponding coarser 1 arc-second DEM pixel value. Where multiple higher resolution pixels with ϵ above the threshold fall within a 1 arc-second cell, the maximum elevation value is taken. In cases where ϵ is below the threshold, some central tendency of elevation values of the higher resolution pixels, which constitute a single 1 arc-second pixel, will be used. To this end, we use bilinear resampling, an aggregation process shown to produce the most accurate and physically realistic results in an analysis by Fewtrell et al. (2008). The five geomorphometric parameters, their application in the detection algorithm, and their 1,000 NLD-derived thresholds are detailed as follows (see Figures 2 and 3):

a Relative elevation

Relative elevation (z_r) is first computed to refine the number of pixels upon which later parameters are computed (e.g., adjacent slopes are only calculated for elevated pixels). It is defined as the difference between elevation (z) at point (x,y) and the neighborhood minimum. Neighborhood in this context is defined as a $100 \times 100 \text{ m}^2$ with point (x,y) at its center. This kernel size is selected so that, should point (x,y) be a levee crest, a representation of the true ground surface is captured for an accurate z_r calculation (see equation (1)).

$$z_r = z_{xy} - z_{min} \quad (1)$$

Our expectation is that levee crests are elevated with respect to their neighborhood minimum, resulting in a positive z_r . A visual example is shown in Figure 2b. Figures 3a and 3b show the 1,000 thresholds at 1/3 and 1/9 arc second, respectively, derived from known levee locations. Although both provide quantities in line with the above expectation, their distributions are different. At 1/3 arc second, roughly three times as many z_r thresholds are $<1 \text{ m}$ than at 1/9 arc second. This seems to suggest that, even at $\sim 10 \text{ m}$ resolution, the representation of levee crests in the DEM has been diminished.

b Slope

Slope (S) is a first-order differential of elevation, defined here as the maximum rate of change of elevation between neighboring pixels in x and y directions in degrees (equation (2)). Opposing slopes are permitted to be a maximum of 20 m apart (i.e., levee crests can be 20 m wide). At opposite sides of the levee crest, we would expect to see DEM pixels with a positive S (see Figure 2c).

$$S = \tan^{-1} \left(\sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2} \right) \quad (2)$$

From Figures 3c and 3d, this is shown to be the case in the S thresholds selected. There is a similar and expected trend as with z_r in that the coarser DEM produces flatter slope thresholds than the finer one.

c Aspect difference

Aspect (A) is another first-order differential of elevation, indicating the direction of maximum slope (S). This is defined by equation (3) in degrees clockwise from north. In the case of levees, we would expect the opposing slopes to be facing polar opposite directions (see Figure 2d).

$$A = \tan^{-1} \left(\frac{\left(\frac{dz}{dy}\right)}{\left(\frac{dz}{dx}\right)} \right) \quad (3)$$

The thresholds are defined in terms of the differences between the aspects of opposing slopes. As in equation (4), where (x,y) is a potential levee crest and m and n are the distance of the opposing slopes from this crest in x and y directions, respectively:

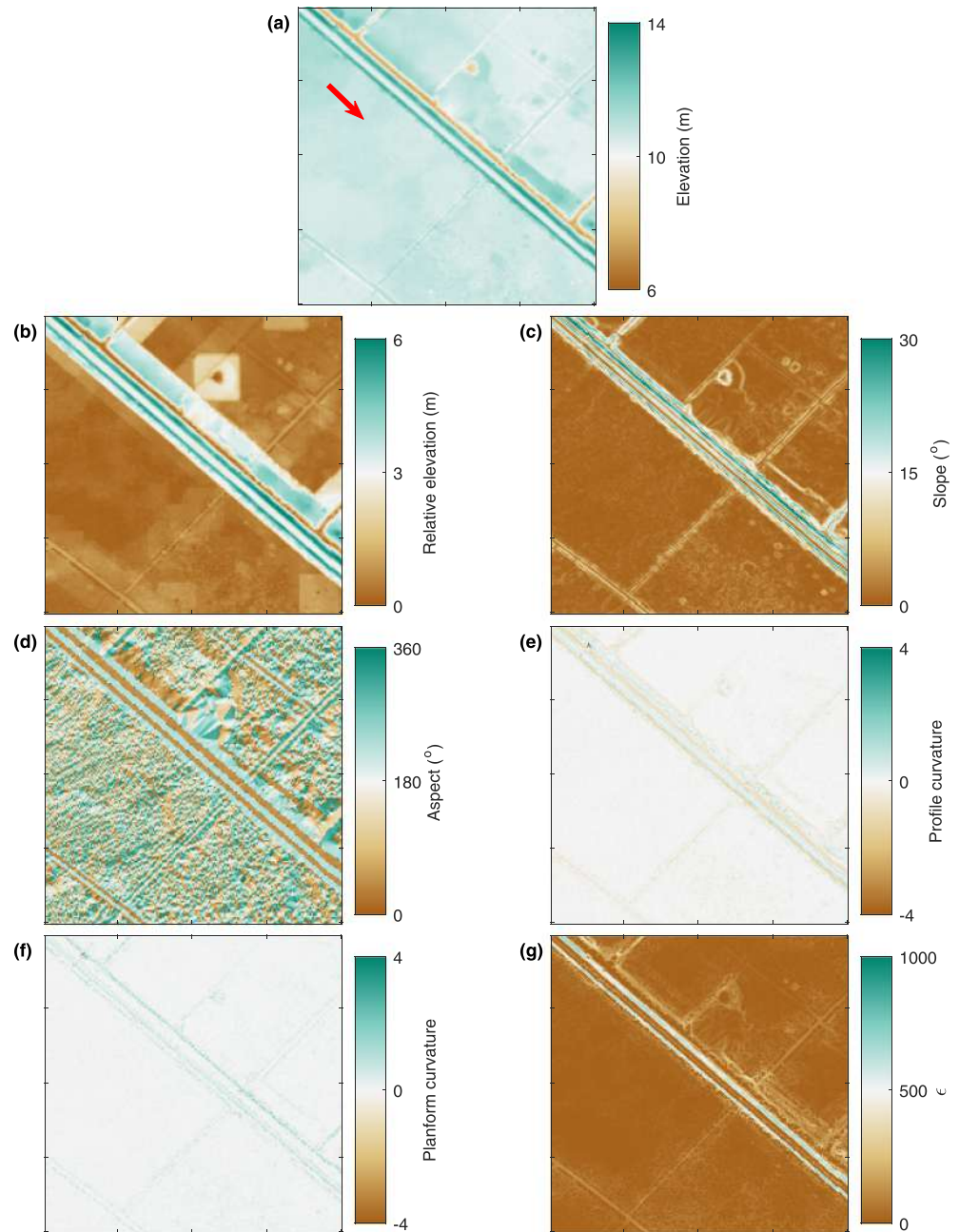


Figure 2. Examples of a levee in a digital elevation model and its processing in the extraction algorithm for a 0.6×0.6 km area of Brazoria County, Texas (ticks are spaced 150 m apart). (a) An area of raw 1/9 arc-second U.S. Geological Survey National Elevation Dataset digital elevation model, with two levees bounding a river channel (in dark green) with flow direction indicated by the red arrow; (b) relative elevation, showing the levees' clear height differential above their neighborhood minima; (c) slope, where the regions of high relative elevation are bounded by steep slopes on either side; (d) aspect, illustrating that these slopes broadly face opposite directions; (e) profile curvature, with the levee crests exhibiting surface convexity across the direction of maximum slope; (f) planform curvature, showing negligible change in slope along the levee crests; and (g) output of the extraction algorithm when driven with geomorphometric parameter thresholds on the right-hand side of Figure 3, where the levees have much higher ϵ values than other terrain features.

$$A_{\text{diff}} = |A_{(x+m,y+n)} - A_{(x-m,y-n)}| \quad (4)$$

The geomorphometric parameter thresholds in Figures 3e and 3f are the deviation of A_{diff} from 180° , permitting some tolerance in the definition of “facing the opposite direction.” In most cases, this tolerance is close

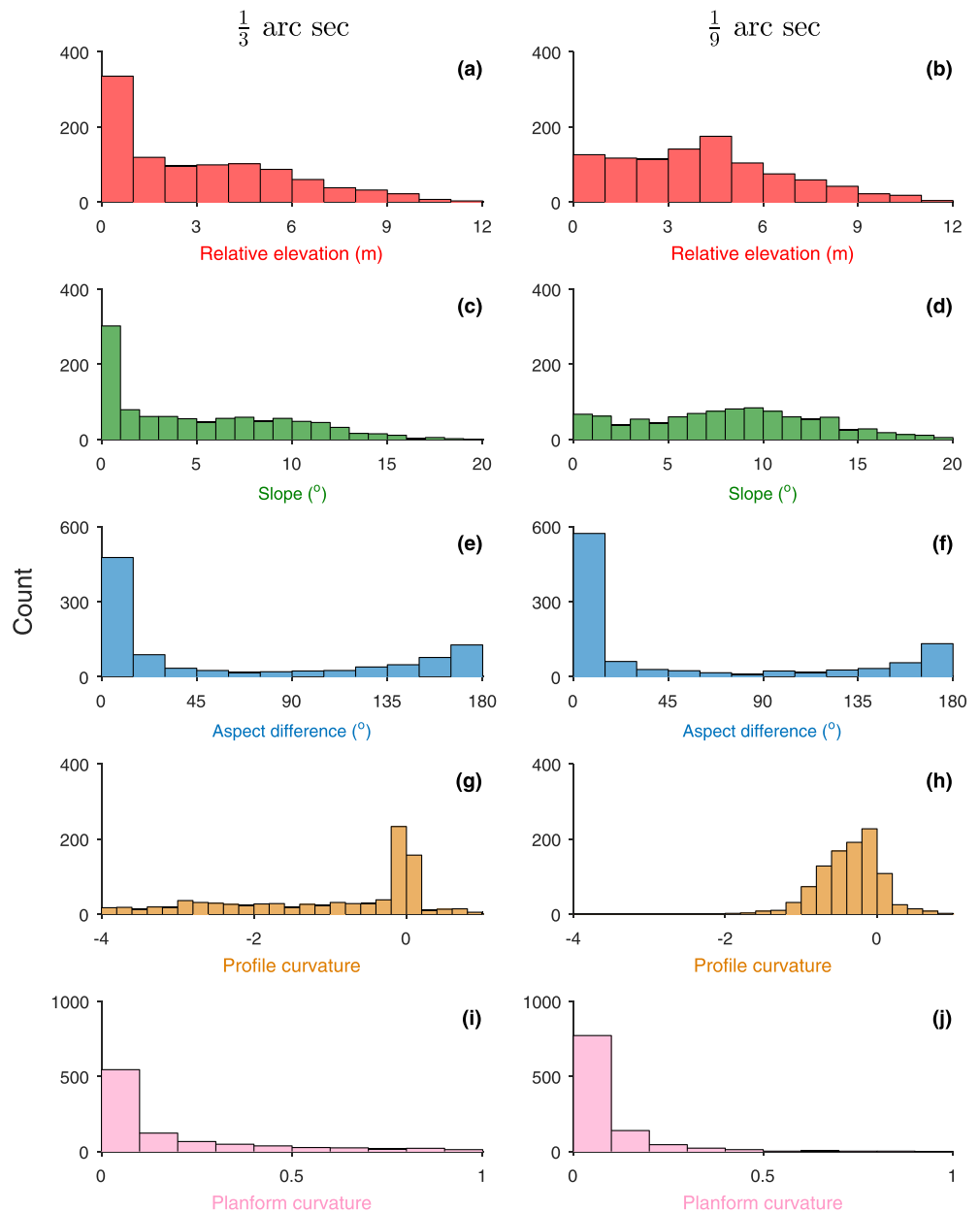


Figure 3. Histograms showing the geomorphometric parameter thresholds sampled from the U.S. Army Corps of Engineers levee data at 1/3 (left-hand side) and 1/9 arc second (right-hand side): (a and b) relative elevation; (c and d) slope; (e and f) aspect difference; (g and h) profile curvature; and (i and j) planform curvature.

to 0°. The slight rise in cases where the tolerance is close to 180° (indicating $A_{diff} \approx 0$) is probably due to NLD geolocation errors where the crest is placed on the slope. In such circumstances, the steep slopes on either side of this misidentified crest would be facing the same direction.

d Profile curvature

Curvature is a second-order differential of elevation. Profile ($prof_c$) curvature is the rate of change of slope in the maximum downslope (S) direction, defined in equation (5):

$$prof_c = \frac{dS}{dxy}. \quad (5)$$

Levees are characterized by profile convexity (Figure 2e), in this case being represented as negative profile curvature. In almost all cases, the algorithm considers levee crests with a convex profile (Figures 3g and 3h). With flatter slopes in the coarser data, profile convexity is also closer to 0.

e Planform curvature

Finally, planform curvature ($plan_c$) is the second derivative of elevation orthogonal to the direction of maximum slope (S'), as defined in equation (6). This direction follows the levee crest, where very little change in slope is expected (Figure 2f).

$$plan_c = \frac{dS'}{dxy} \quad (6)$$

From Figures 3i and 3j, it is evident that levees in the NLD have no or very little planform curvature within a narrow tolerance.

Figure 2g demonstrates an example of how the algorithm functions for a small section of the 1/9 arc-second DEM shown, with higher values of extraction rate (ϵ) along the known levee crests than for other floodplain features. The algorithm was run for the CONUS at 1/3 arc-second resolution and at 1/9 arc-second resolution where available to regenerate a 1 arc-second NED-based DEM with these hydraulically important features preserved during the coarsening process. The 1 arc-second “defended” DEM built with the feature preservation algorithm run with 1/9 arc-second data is validated against surveyed levee crest heights in the California Levee Database (CLD) provided by the State of California Department of Water Resources. The CLD contains geodetically surveyed crest elevations referenced to the NAVD88 datum for roughly 7,000 km of Californian levees. Concurrently, the “undefended” 1 arc-second DEM using standard bilinear resampling and no additional algorithmic consideration of levee-like features is benchmarked against the CLD to chart the improvement in DEM crest elevation representation when the new method is employed.

The continental-scale hydrodynamic model Fathom-US was rerun with a new 1 arc-second DEM built with algorithmic output of the 1/3 arc-second data, owing to the seamless coverage of this higher resolution variant. Performance when run with the original (undefended) DEM is compared to that when simulated on the new DEM, using local flood maps built by the Iowa Flood Center (IFC) as a benchmark. The IFC, established by the State of Iowa, is charged with producing and sharing inundation maps for the purposes of flood research, mitigation, prediction, and insurance (Chen et al., 2017; Horna-Muñoz & Constantinescu, 2018; Krajewski et al., 2017). Their maps consist of two modeling strategies: (a) complete-coverage statewide maps built with 1-D HEC-RAS models to underpin National Flood Insurance Program rate setting and (b) more detailed urban flood maps at selected locations, mostly simulated using coupled 1-D/2-D models. Both IFC modeling approaches are driven with multiple AEP flows calculated from USGS gauging stations or using standard USGS regression equations for ungauged streams. For the statewide maps, water surface profiles generated via step backwater calculations using the AEP discharge and surveyed or LiDAR-derived channel cross sections are intersected with 1-m resolution LiDAR-derived DEMs. The detailed urban flood models utilize surveyed channel bathymetry merged with LiDAR elevation data, representing both rivers and floodplains. Hydraulic structures such as levees, weirs, and bridges are surveyed and stitched into either the 1-D channel representation or the 2-D DEM. The 1-D/2-D HEC-RAS or MIKE FLOOD models, run at 10- to 20-m resolution, are calibrated to flow observations, water surface profiles, or high-water marks where available. Further details on IFC flood mapping procedures can be found in Gilles et al. (2012). Comparing Fathom-US output to both IFC modeling strategies permits performance benchmarking to wide-area studies, which employ cruder representations of flow physics yet are built with accurate local data, and also to extremely high-quality, engineering-grade inundation models whose high data requirements result in prohibitively high financial expense to produce them at larger scales. Should the large-scale model with improved defense representation achieve similar realizations of flood extent across different AEPs, it will be a vindication of the methodology presented.

The testing consists of the employment of ubiquitous binary pattern fit statistics (e.g., Aronica et al., 2002; Sampson et al., 2015; Werner et al., 2005; Wing et al., 2017):

$$\text{HR} = \frac{M_1B_1}{M_1B_1 + M_0B_1}, \quad (7)$$

$$\text{FAR} = \frac{M_1B_0}{M_1B_0 + M_1B_1}, \quad (8)$$

$$\text{CSI} = \frac{M_1B_1}{M_1B_1 + M_0B_1 + M_1B_0}, \quad (9)$$

$$\text{EB} = \frac{M_1B_0}{M_0B_1 + M_1B_0}, \quad (10)$$

where M and B describe pixels of the model being tested (Fathom-US) and those in the benchmark data (IFC) respectively, and subscripts 1 and 0 indicate whether a pixel is wet or dry in each model respectively. Hit Rate (HR; equation (7); optimum = 1) is a measure of model underprediction, penalizing only “misses” (M_0B_1). The False Alarm Ratio (FAR; equation (8); optimum = 0) measures model overprediction, where “false alarms” are exclusively penalized (M_1B_0). Critical Success Index (CSI; equation (9); optimum = 1) is the most discriminatory measure, penalizing both underprediction and overprediction. Finally, the Error Bias (EB; equation (10); optimum = 0.5) metric is slightly different to that employed in previous studies. It still indicates the overall balance of underprediction and overprediction but ensures the magnitude of the deviation from the optimum in either direction is consistent: $\text{EB} > 0.5$ is overprediction; $\text{EB} = 0.5$ is unbiased; and $\text{EB} < 0.5$ is underprediction.

To further discriminate between the IFC and Fathom models, we employ exposure-weighted metrics in a similar vein to those proposed by Pappenberger et al. (2007). Equations (7)–(10) are applied in the same way, but M_xB_x is population counts within such areas rather than pixel counts (e.g., M_1B_1 is the total number of people within areas the model and benchmark data agree is inundated). Population counts are derived from the demographic map of the U.S. Environmental Protection Agency EnviroAtlas program, where census block counts are dasymmetrically downscaled to 30 m pixels based on land use and slope. While the pixel-count metrics indicate that wide-area physical modeling ability where correctly modeling relatively simple phenomena (e.g., capturing river channels and flooding in the large portions of undeveloped floodplain) is rewarded, the population-count metrics indicate model performance in the most important (and difficult to model) areas in a risk context.

To demonstrate transferability of this method to other geographic regions, indicate the effect of it on large-scale models where similar data are available, and perhaps illustrate a use case in smaller scale studies also, the extraction algorithm was run over 1/9 arc-second resolution LiDAR data (provided at 2 m resolution but resampled to 1/9 arc second) of the Po River floodplain in northern Italy. Such a test case enables greater isolation of the effect of this new levee representation method in a hydraulic model, since crucial flood model components—a seamless LiDAR DEM, surveyed channel bathymetry and hydrography, and well-constrained inflow boundary conditions—are substantially more accurate than in the larger scale, AEP-simulating US model. The ~350 km stretch of the middle to lower Po considered is a large alluvial floodplain bounded by embankments containing a system of minor levees within. The model is driven with boundary conditions from an historical flood event in October 2000 using the same computational hydraulic engine as in Fathom-US but executed fully in 2-D (no subgrid channels). Instead, surveyed channel bathymetry is burnt directly into the DEM, thus representing rivers as supragrid features. The data associated with this model of the Po have been used in a number of hydraulic studies (e.g., Castellarin et al., 2009, Castellarin, Di Baldassarre, & Brath, 2011, Castellarin et al., 2011; Di Baldassarre et al., 2010; Domenghetti et al., 2015; Schumann et al., 2010). For this test case, three different DEMs are employed while all other elements of the model structure are held constant:

- a Manual Defended (MD): the raw DEM is resampled to 1 arc-second resolution, and manually digitized levee crest centerlines are used to elevate corresponding pixels to the 1/9 arc-second elevation value.
- b Automated Defended (AD): to emulate large-scale model structures, where local levee data are mostly unavailable, the automated levee extraction algorithm is run over the raw 1/9 arc-second DEM, preserving the elevations of relevant pixels during the resampling to 1 arc-second resolution.

- c Undefended: no representation of flood defenses except those which are inherent in the resampled 1 arc-second DEM; while some residual representation of flood defenses may be present, albeit dampened, no additional enhancement beyond their coarsened representation is performed.

The methodology presented here would be validated if levees hold (i.e., are not overtopped by floodwaters) in the AD (b) DEM where they also hold in the MD (a) DEM while overtopping in the undefended (c) DEM. While the manual handling of flood defenses considers exclusively formal levee structures, the automatic method extracts all levee-like features regardless of their anthropogenic classification. This means that the elevation of more informal features, such as small berms, river banks, roads, undocumented (often minor) dikes, and other natural features, is captured, while remaining “smoothed out” of the MD DEM. As such, we can further examine the hydraulic effect (e.g., on flood peak attenuation) of accounting for these features in the AD vs MD DEM using observations of flow and high water during the event considered.

3. Results and Discussion

The algorithm was executed across the entire CONUS at 1/3 arc-second (~10 m) resolution, at 1/9 arc-second (~3 m) resolution where available in the United States, and over the Po floodplain at 1/9 arc-second (~3 m) resolution to ultimately produce new “defended” 1 arc-second (~30 m) DEMs for the United States and the Po. At 1/3 arc-second resolution, a single parameter threshold set takes, on average, 0.99 s to run for a $1 \times 1^\circ$ (approximately 100×100 km) tile on a single 2.30 GHz Intel Xeon E5-2650 core. For the same size tile at 1/9 arc-second resolution, a single parameter threshold set takes 2.54 s to run on average. Given each parameter threshold set simulation is independent, the algorithm lends itself ideally to parallelization. Executing a hydrodynamic model at this scale and either resolution would require many orders of magnitude greater than simulation time. Thus, preserving any relevant high-resolution information during the necessary coarsening of the DEM is computationally plausible and should enable improvements in model performance, particularly for lower magnitude, higher frequency flooding, which is very sensitive to small-scale topographic features. This section will examine whether this is the case.

3.1. Levee Crest Elevation Validation: California, USA

The 1 arc-second DEM built with consideration of levee-like features at 1/9 arc-second and an ε threshold of 300 (see section 3.2.1 for rationale) is compared to geodetic surveys of levee crest heights in the CLD. The results of this benchmarking are shown (in blue) in Figure 3, where positive (negative) errors indicate DEM overprediction (underprediction) of the CLD crest height. Crest elevation errors appear broadly normally distributed around a central tendency close to 0. The defended DEM exhibiting a median error of -0.05 m means that, on average, crest elevations are underpredicted by 5 cm. Median absolute error (meaning positive and negative errors do not cancel out) comes to 0.26 m. Errors in the undefended 1 arc-second DEM built only by resampling the 1/9 arc-second data are shown in orange. This DEM has a clear underpredictive bias, with a median error of -1.16 m (Figure 4).

It is evident, then, that the smoothing effect of standard resampling approaches results in substantial truncation of levee crest heights. The upshot of this is that flood inundation models which have no *post hoc* techniques in the DEM processing for ameliorating this issue will overtop levees and inundate areas they otherwise protect at shallower water depths than in reality. Meanwhile, the new approach presented here appears to be highly effective in the preservation of crest heights, reducing average crest elevation errors to almost 0. That said, there is a hint of negative skewness in the defended DEM (mean error of -0.37 m), with some levee crests still being substantially underestimated. This illustrates the inherent challenges and inevitable smoothing effect of constructing DEMs (even at high resolution) where single elevation values often represent heterogeneous grid cells. Some of the more extreme errors in both DEMs are likely where the NED is not built with LiDAR data.

3.2. Interscale Model Comparison: Iowa, USA

3.2.1. Elevation Model Building and Parameterization

The 1 arc-second DEMs employed in the Iowa analysis, both derived from the 1/3 arc-second NED in this instance, are exemplified in Figure 5. It is evident that in this particular area of Iowa, a lack of consideration of levees results in a diminution of their elevation upon resampling to a coarser resolution, particularly on the

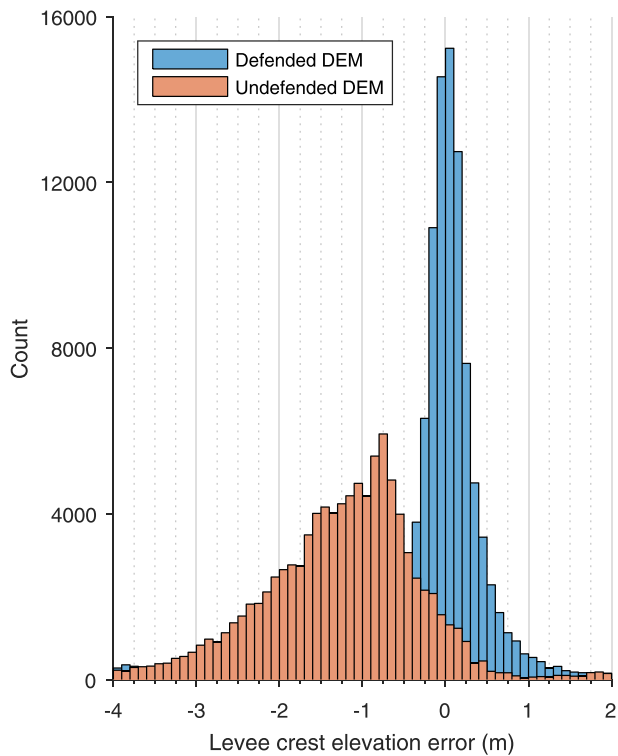


Figure 4. Histograms showing the distribution of elevation errors when benchmarking the defended (blue) and undefended (orange) 1 arc-second digital elevation models against levee crest elevations in the California Levee Database.

northern bank of Cedar River (cf. Figures 5a and 5b). In the DEM generated via the levee detection algorithm, the levees are represented to their full effect (cf. Figures 5a and 5c). Figure 5e provides a lateral view of this: both the undefended and defended DEMs are identical, except for their representation of the two levees either side of the river channel. The defended DEM has elevated the relevant pixels to more closely match the peaks captured in the 1/3 arc-second data. Figure 5d plots the 1/3 arc-second elevation data again but with the corresponding algorithm output. It is clear that levee crests are afforded higher values of ϵ than other features, but in some instances extraneous objects (e.g., artifacts in the DEM representation of the channel bed and points at deceleration of the levee slope) have notable ϵ values. Even if a threshold of ϵ is employed which transpires to capture these objects however, the effect of this on the DEM is minimal, demonstrating the tendency of this method to “fail to safety.” Figure 5f illustrates this for different thresholds of ϵ : the resultant DEM scarcely changes in any meaningful way when examining a range of suitable values. ϵ thresholds greater than 400 tend to produce DEMs with broken levee centerlines and ignore more modest, yet potentially crucial, levee-like features. Based on numerous visual inspections of the data across the United States, an ϵ threshold of 300 is identified to be suitable. Figure 5f indicates that so long as the algorithm’s geomorphometric parameters are considered to some extent, DEM generation appears to have little sensitivity to the choice of threshold. Evidencing this across a much larger scale, the median non-zero difference between an ϵ_{100} and ϵ_{300} Iowa DEM is 0.12 m. In other words, pixels where $100 \leq \epsilon < 300$ are considered cause an average elevation increase of 12 cm compared to an instance where they are not considered at all. This is well within the errors in the raw DEM itself (Gesch et al., 2014).

3.2.2. Statewide Model Comparison

First, the results of the comparison to lower quality, wide-area flood maps of the entire state are shown in Table 1. Generally, the two models show good correspondence, with between 82% and 96% of the IFC floodplain captured by Fathom-US. Accounting for false alarms, which range between 15% and 27% of correctly identified pixels, the Fathom model attains a 63–82% fit to the IFC data (based on CSIs). Error biases indicate that 64–80% of incorrect pixels are false alarms. There is a clear trend in performance with AEP. As the floods get larger (AEP reduces; a lower frequency, higher-magnitude flood is modeled), the fit to IFC data increases. This is perhaps unsurprising, as larger floods are often simply constrained by large-scale topographic features, while small floods are much more sensitive to the control exerted by small-scale topographic features. Differences due to boundary condition derivation on ungauged streams will also be more evident for smaller floods. Even slight differences in the definition of a 10% AEP streamflow can result in dramatically different flood extents since low-gradient floodplain land is still available for inundation. Differences in the 0.2% streamflow, so long as it is large enough to fill the valley, will make negligible difference to resultant flood extents. Stephens et al. (2014) also note that larger floods generally enjoy inflated CSI scores simply because there are more pixels to count as “hits”: misestimating the flood edge by a given distance is penalized more heavily for smaller floods. Performance disparities between large and small floods also arise through approximations of channel capacity. The IFC maps benefit from local bathymetric surveys, while the Fathom-US model must approximate this since remote sensing of channel beds is not yet possible. Approximations are made based on drainage area and an assumed AEP bankfull discharge of 50%. In reality, this assumption will not hold, particularly for engineered waterways. Smaller floods will be much more sensitive to channel conveyance, as a greater proportion of the total discharge will be held in-channel compared to larger floods. Uncertainties relating to what the channel conveyance actually is will thus hold greater sway over the resultant extent of a smaller flood. One further phenomenon of note is the difficulty in isolating false alarms, noted also by Wing et al. (2017), due to differences in model domain. Through its nature as an automated, large-scale flood inundation model, all rivers of a certain drainage area ($>50 \text{ km}^2$) are modeled, and the pluvial

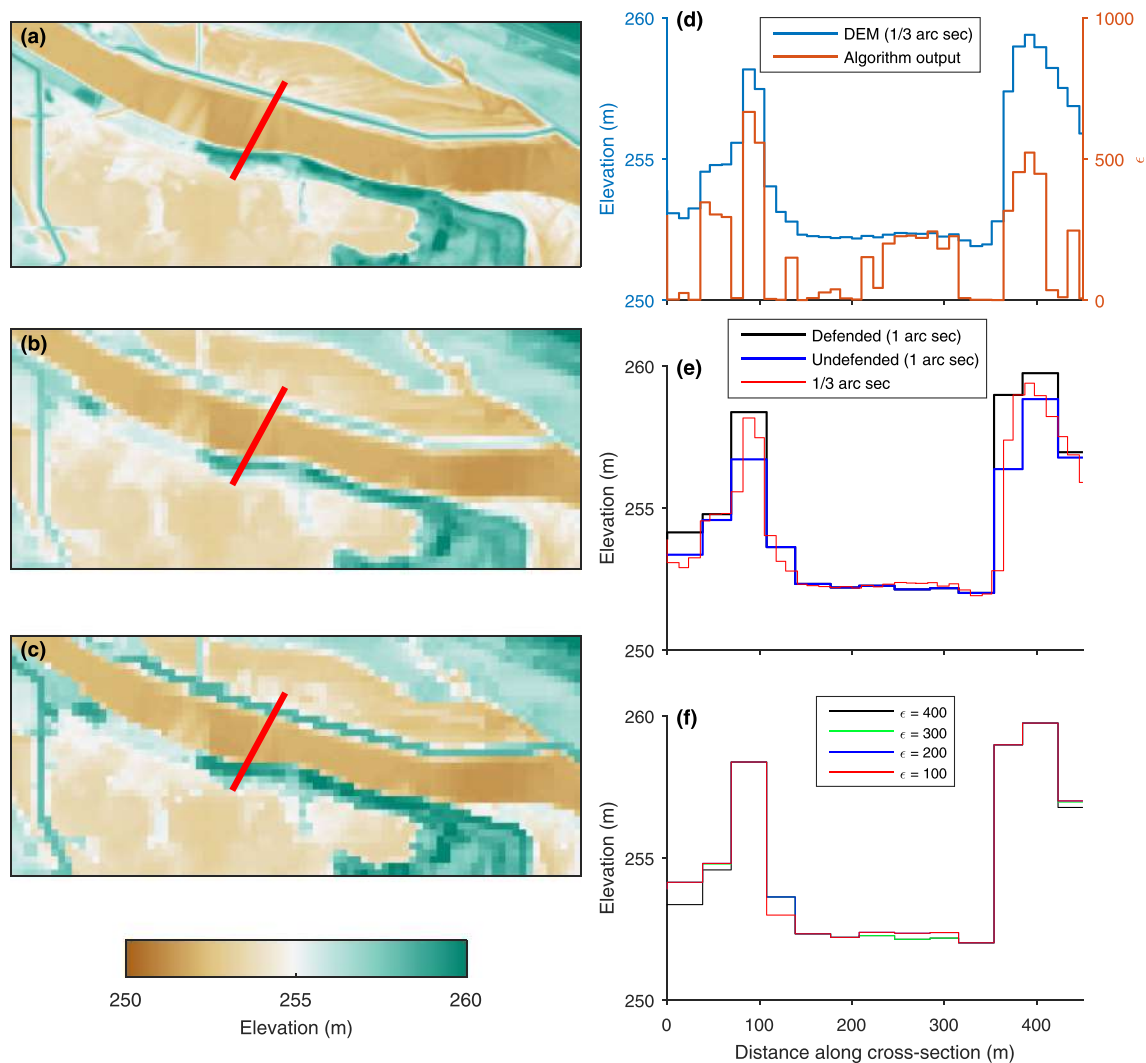


Figure 5. An example of the algorithm’s functionality of a 1×3 km area along Cedar river in Waterloo, IA. (a) shows the baseline high-resolution (1/3 arc second) digital elevation model (DEM); (b) shows an “undefended” DEM, resampled to 1 arc-second with no consideration of levees; and (c) shows a “defended” DEM generated by the algorithm presented here, where the ϵ threshold is 300. The red line in (a)–(c) is the cross section for which data are shown in (d)–(f) (increasing distance along the x axis relates to NE-SW movement along the cross section). (d) shows the elevation in the 1/3 arc-second DEM alongside raw algorithm output of values of ϵ for each corresponding pixel; (e) shows the cross sections of elevation values for each of the DEMs in (a)–(c); and (f) shows the cross section of elevation values for when different ϵ threshold values are used.

model simulates flooding in rivers smaller than this. The IFC data are an assemblage of local studies, which generally map rivers down to ~ 3 km² flow accumulation, though do not model every one, often for justified

Table 1
Pixel-Count Test Scores, Comparing the Statewide Iowa Flood Center Maps and Fathom-US Run With the New Defended Digital Elevation Model

Annual exceedance probability, %	Hit rate	False alarm ratio	Critical Success Index	Error bias
10	0.82	0.27	0.63	0.64
4	0.88	0.22	0.71	0.68
2	0.92	0.20	0.75	0.73
1	0.93	0.18	0.78	0.75
0.5	0.95	0.18	0.79	0.79
0.2	0.96	0.15	0.82	0.80

reasons, because such rivers are in uninhabited areas and there simply is no need to expend resources on modeling them. Large-scale models are agnostic of such priorities and therefore have total coverage. With no easy way of excluding these areas from the analysis, many pixels flagged as false alarms are truly in areas the IFC have not studied. See Figure 6a for an example of this, where there is broad flood extent agreement in areas the IFC has modeled but non-genuine false alarms in unmodeled-by-IFC headwater areas. Thus, while the error bias scores indicate a relatively high tendency towards overprediction, metrics accounting for false alarms (FAR, CSI, and EB) should be viewed in light of the limitations of the benchmark data set.

3.2.3. Urban Model Comparison

With a more manageable number of high-quality IFC urban flood maps, model domains and the rivers they include are easier to isolate manually, enabling a more faithful intercomparison. However, there are still instances where the IFC has modeled a single river flowing through the urban center while Fathom-US also simulated flooding on its tributaries that fall within the specified domain. For larger floods, areas around these tributaries are inundated by the main river and require consideration. For smaller floods, these areas are unaffected by the IFC-modeled channel, yet inundation arises from overtopping in the (unmodeled by IFC and modeled by Fathom-US) tributary. See Figure 6 for an example at the Waterloo study site. Some areas flagged as false alarms for the 10% AEP flood (red in Figure 6b) are not connected to the IFC-modeled Cedar River but around IFC-unmodeled Black Hawk Creek (SW), Virden Creek (NE), and Elk Run (SE). The reason these inundated areas are considered (e.g., only a small portion of Fathom-US flooding in Virden Creek) is because the 0.2% AEP flood (Figure 6c) arising from the Cedar River overtopping extends to these areas and so falls within the bounds of the study area. In reality, the 10% AEP flood appears a close match along the Cedar River at Waterloo, but its CSI (and especially the population-weighted CSI) will not be a true reflection of this. Thus, precise quantification of false alarms still remains elusive in some cases. The comparison involved 27 individual studies encompassing the main urban areas within Iowa: Ames, Cedar Falls, Cedar Rapids, Charles City, Clarksville, Columbus Junction, Des Moines, Elkader, Fort Dodge, Greene, Hills, Humboldt, Independence, Iowa City, Kalona, Manchester, Maquoketa, Mason City, Monticello, Ottumwa, Plainfield, Red Oak, Rock Rapids, Rock Valley, Spencer, Waterloo, and Waverly.

Figure 7 gives an overview of the intercomparison results, where both v2 (defended DEM presented in this paper) and v1 (standard DEM resampling) iterations of Fathom-US are benchmarked against the IFC models. Selected examples are shown as maps in Figure 8, where performance has increased in Ottumwa (Figures 8i and 8j), remained broadly constant at Monticello (Figures 8c and 8d), and slightly decreased at Clarksville (Figures 8e and 8f). Based on common pixel-count metrics, performance of the new v2 model is generally high. Median hit rates range from 86% at the 10% AEP flood to 95% at the 0.2% AEP flood (all AEPs experience a maximum HR of >99%). Median CSI at the 10% AEP is 0.69, up to 0.80 for the 0.2% AEP. Maximum CSIs for each AEP flood range from 0.85 (10% AEP) to 0.91 (1% AEP). False alarms are generally at ~10% of correct pixels in magnitude and tend towards accounting for 40–60% of incorrect pixels across all AEPs. However, it is evident that increases in performance based on simulating over a standard DEM are either minimal or non-existent. This is because the bulk of the floodplain has essentially remained unchanged between the two models, clouding small-scale and important changes to inundation extent by the defended DEM. Viewing population-weighted metrics goes some way in addressing this issue by being much more sensitive to small changes in the flood edge in inhabited areas, but these scores are generally dominated by inundation in areas where the two DEMs are virtually identical or where baseline DEM representation of important features is already adequate. Even so, increases in median population-count CSIs from v1 to v2 are quite marked: 0.20 to 0.22 (10% AEP); 0.14 to 0.25 (4% AEP); 0.23 to 0.36 (2% AEP); 0.33 to 0.45 (1% AEP); 0.50 to 0.53 (0.5% AEP); and 0.50 to 0.55 (0.2% AEP). For instance, the 1% AEP flood in Ames (Figures 8g and 8h) has a muted pixel-based CSI improvement from v1 (0.70) to v2 (0.72) but a relatively large increase in population-weighted CSI (0.39 to 0.47). This suggests that model performance is increasing in the very areas we are most interested in (from a risk-based perspective). Interestingly, this does not appear to be driven by a reduction in false alarms (population-count FARs remain fairly constant across model versions) as might be expected but by increases in the rate of correct identification of inundated areas (population-count HRs increase). This suggests that a more accurate representation of flow conveyance across the floodplain, driven by heightened representation of hydraulic structures, correctly inundates areas previously left dry by the model at a greater rate than preventing incorrectly inundated areas from flooding. On the other hand, the failure of this new method to reduce false alarms suggests that there are other causes of overprediction inherent to large-scale models. Approximations to channel capacity may induce out-of-bank flow to occur more frequently than in reality. With the proportions of floodwater in-channel and on-floodplain skewed for a given AEP event, correctly represented flood defenses may incorrectly overtop regardless. In Figure 9, it is evident that, in certain circumstances, this method of DEM construction actually exacerbates errors in other facets of the model. In Figure 9a, important hydraulic features have been elevated by the levee detection algorithm around this stretch of the Missouri River, enabling their accurate representation in the inundation

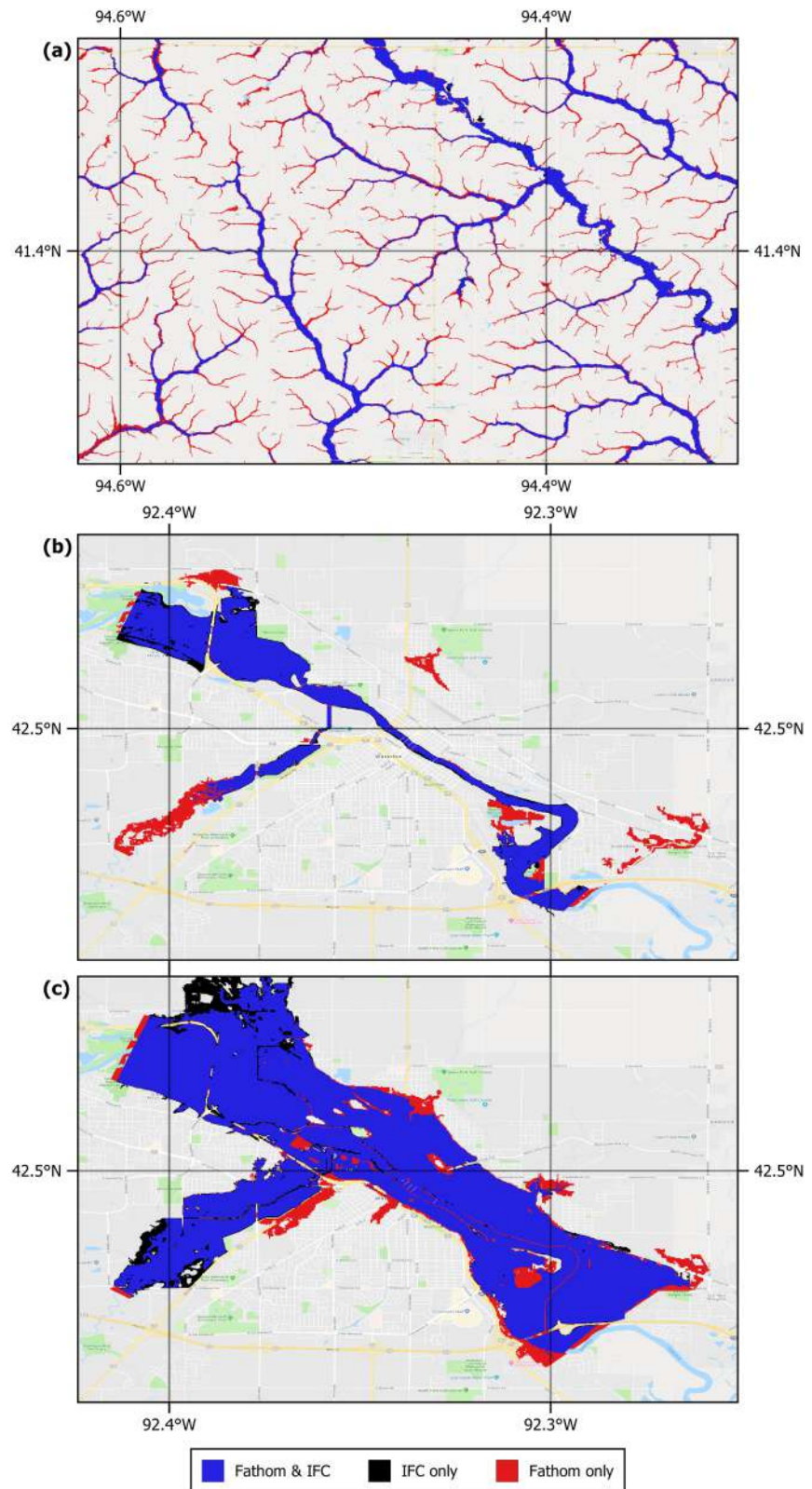


Figure 6. Example of the model intercomparison between Fathom-US and the Iowa Flood Center's (a) statewide models of the 1% annual exceedance probability (AEP) flood for an area of Iowa in between Des Moines and Omaha; urban model at Waterloo of the (b) 10% AEP flood and (c) 0.2% AEP flood.

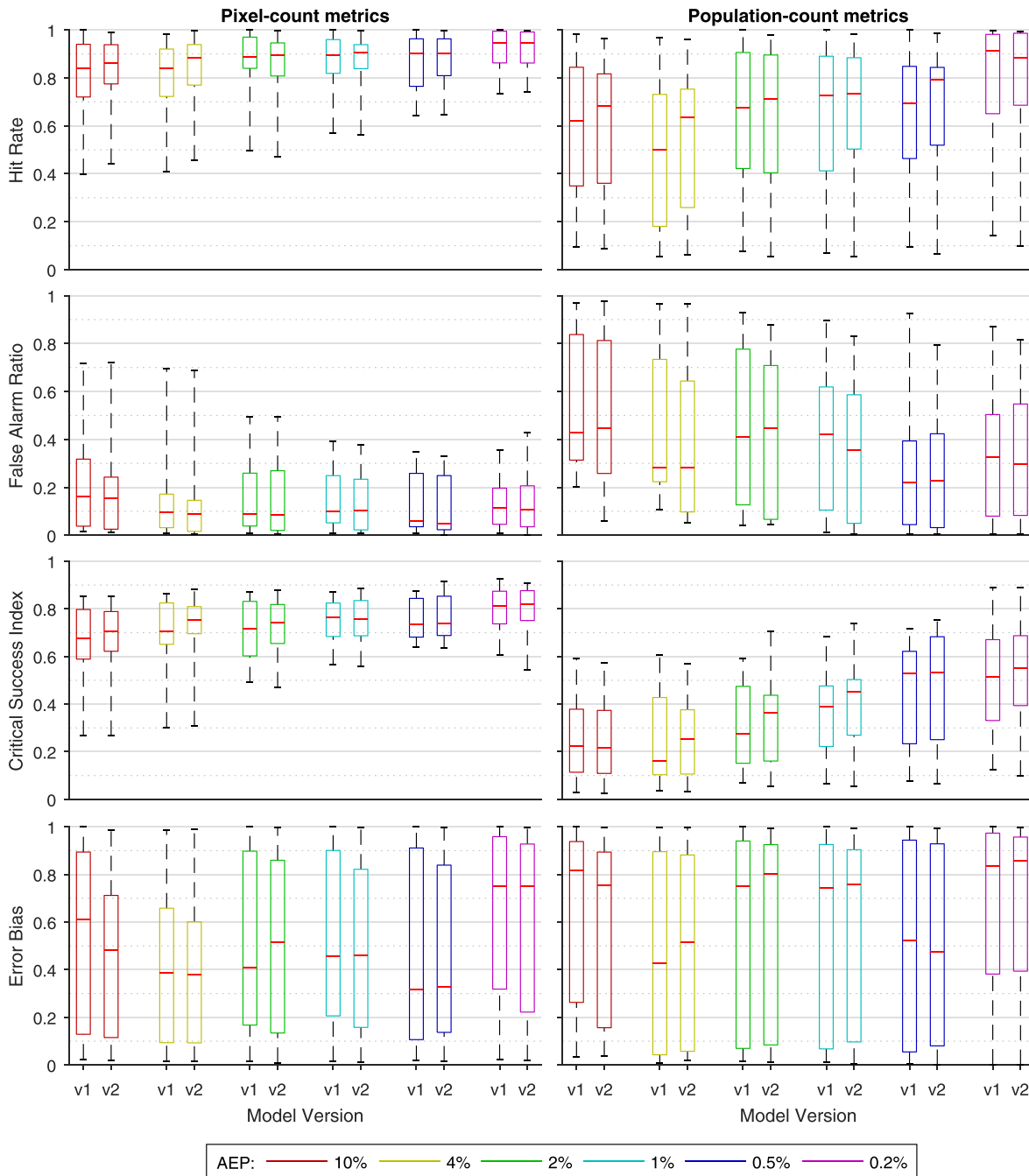


Figure 7. Boxplots of the binary pattern metrics when comparing versions of Fathom-US to the Iowa Flood Center models of 27 urban areas in Iowa. The left-hand side shows test scores using raw pixel counts in equations (7)–(10). The right-hand side shows test scores using population counts in equations (7)–(10). The left boxplot (v1) within each of the six AEP groupings is the original version of Fathom-US presented in Wing et al. (2017), while the right boxplot (v2) in each group is the Fathom-US model when executed using the new defended DEM presented here.

model. Yet, as indicated by the red line, the HydroSHEDS-derived channel network deviates substantially from the evident tributary branching eastwards from the Missouri (channel networks are derived from interpolating across the cell centers of 3 arc-second HydroSHEDS flow accumulation data). The flood walls (Figure 9b) on the west and southeast banks of the Missouri have held back floodwaters, while the hydrography of the tributary has broken through the flood walls in the north and east of the figure. As bankfull flow was exceeded, the floodwaters backed up against the levee, meaning it could not even enter the true channel or inundate the floodplain engineered to store and convey the water, and proceeded to

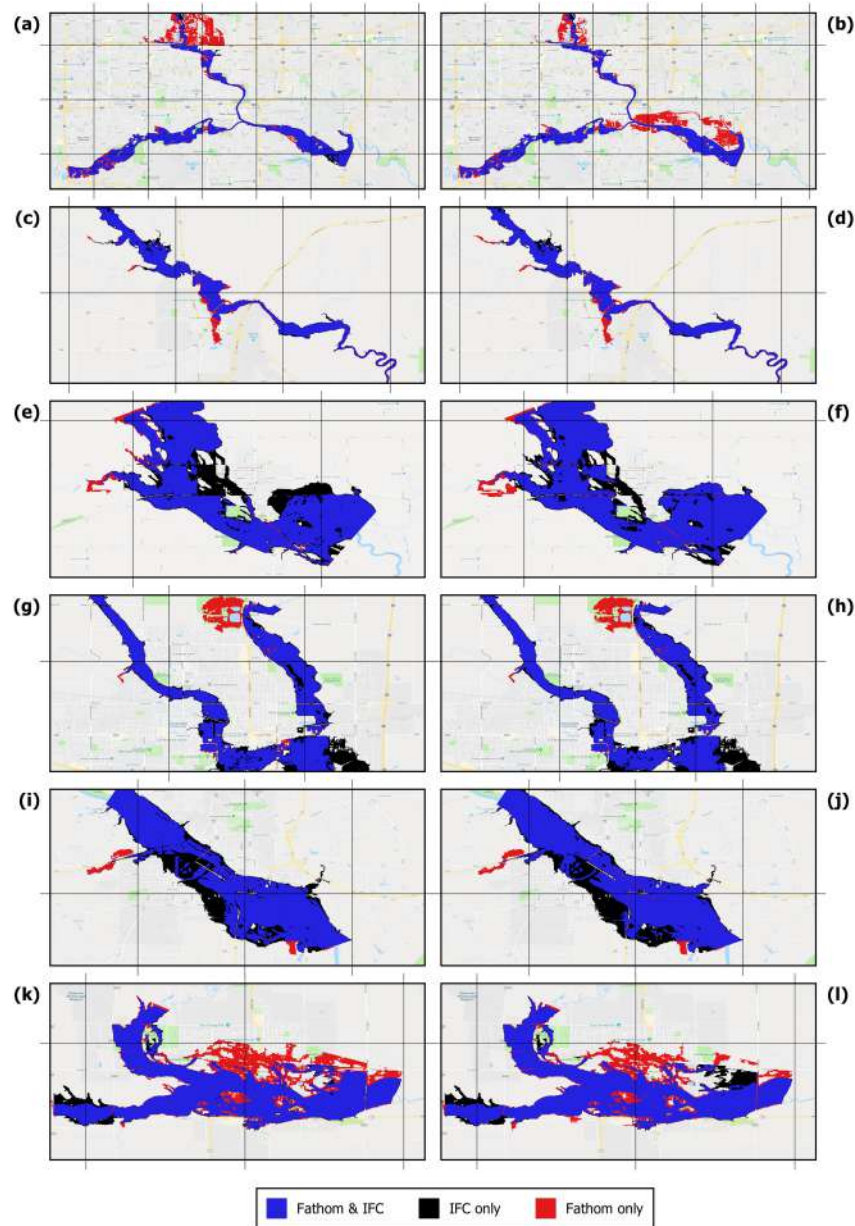


Figure 8. Examples of model intercomparison between the Iowa Flood Center's urban models and Fathom-US v1 (right-hand side) and v2 (left-hand side). Each grid box is 180 arc seconds (~6 km). AEP = annual exceedance probability; CSI = Critical Success Index. The table below gives further information:

Figure 8 panel	Location	Fathom-US version	AEP, %	CSI (pixel based)	CSI (population based)
(a)	Des Moines	v2	10	0.71	0.43
(b)	Des Moines	v1	10	0.65	0.31
(c)	Monticello	v2	4	0.86	0.44
(d)	Monticello	v1	4	0.84	0.43
(e)	Clarksville	v2	2	0.78	0.10
(f)	Clarksville	v1	2	0.81	0.09
(g)	Ames	v2	1	0.72	0.47
(h)	Ames	v1	1	0.70	0.39
(i)	Ottumwa	v2	0.5	0.81	0.52
(j)	Ottumwa	v1	0.5	0.74	0.26
(k)	Spencer	v2	0.2	0.74	0.45
(l)	Spencer	v1	0.2	0.76	0.51

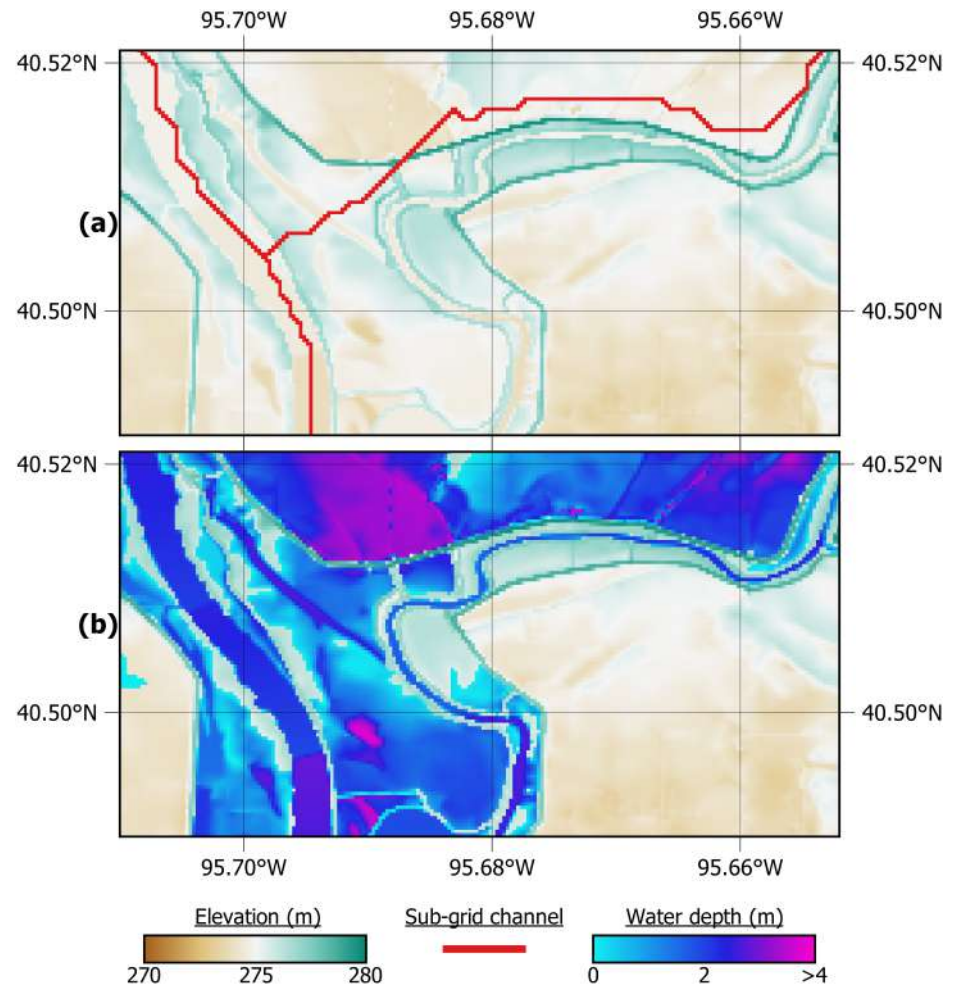


Figure 9. For a stretch of the Missouri River: (a) the defended digital elevation model and subgrid channel network derived from HydroSHEDS employed by Fathom-US, with (b) the resultant 2% AEP flood depth grid.

flood an erstwhile defended area. A smoother DEM with a discontinuous representation of these features is more forgiving of hydrographic errors: at least, by permitting water to inundate areas it would do in reality. There is an implication, then, that representing defenses in this way is overzealous in some situations given the errors posed by poor channel representation. Put another way, the use of channel hydrography defined by HydroSHEDS is inadequate in this sophisticated model setup. The integration of new data sources, such as the USGS National Hydrography Dataset based on the 1/3 arc-second NED (NHDPlus HR) or MERIT Hydro (Yamazaki et al., 2019) or DEM-based methods similar to the levee preservation method presented here which instead preserve hydraulically relevant flow paths and channel networks (e.g., Moretti & Orlandini, 2018; Sangireddy et al., 2016), may ameliorate this issue if implemented in the Fathom-US model framework. Future research will explore this.

More broadly, the population-weighted metrics indicate that there is a lot more work to be done to resolve the scale performance gap in inundation modeling than would be suggested by the pixel-based metrics. The pixel-based scores here and in Wing et al. (2017) suggest that skill between local and large-scale models is converging, with CSIs approaching a ceiling given the fundamental hydraulic modeling constraints imposed by extreme flow characterization (Wing et al., 2018). This makes charting model improvements to seemingly already within-error models difficult. However, while such models may appear to be within or approaching error on a pixel-count basis, the exposure-weighted metrics illustrate stark deviations in relevant inundation extent. An idea of what a high performance model should score in this context has little precedent though, so expectations (e.g., pixel-based CSI of >0.75 amounting to excellent

performance; Fleischmann et al., 2019) should be tempered, as population-based metrics will not be inflated by a large number of “easy hits.” Regardless, the updated Fathom-US model finding median correspondence with local IFC studies of 20–55% of exposed populace suggests that there is still plenty of room for improvement.

It is also important to note that validating hydraulic models of this scale is notoriously difficult. Such models are not event replicating but instead simulate something unobservable in reality (static-AEP-in-space floods), and, even if they were, observations of real events are unavailable at a spatial scale to adequately interrogate them. This necessitates model to model intercomparisons as a pseudo-validation procedure. The engineering-grade IFC models, though treated as a benchmark, are themselves uncertain. They generally benefit from calibration, where model parameters are adjusted until the model resembles specified benchmark data (e.g., high-water marks). The higher model skill attributed to local-scale compared to large-scale models, then, may be a product of this non-physical tuning rather than possessing a more fundamentally skillful model structure. Furthermore, in spite of accurate elevation data, channel hydrography and bathymetry, computational hydraulics, and defense information, they are impaired by the fundamental inability of all models to characterize extreme flows. Even for models fortunate enough to have boundary conditions informed by a river gauge, their short records and subjection to nonstationarity (particularly of land use) result in multiple defensible interpretations of given AEP flows (particularly less frequent ones). The justification, therefore, for one model being “validator” and the other being “validatee” is weakened. This perhaps explains traces of equifinality in the pixel-based comparisons, with improvements to model performance between defended and undefended models virtually undetectable amidst insoluble boundary condition uncertainty. See Smith et al. (2015), Villarini et al. (2009), and Blöschl et al. (2013) for further discussion of discharge estimation errors in time and space. That said, the lower quartile of pixel-based CSI scores for high-frequency design floods noticeably increasing from v1 to v2 does suggest that poorer models are being brought closer to the performance ceiling. Furthermore, conventional pixel-based metrics would consider Fathom-US a high performance model, and it is clear that, on the whole, the representation of hydraulically important features by the algorithm presented here has improved the model further. For sporadic, qualitative evaluations, the updated model appears more behavioral than previously. In Figure 8a, levee representation in Des Moines has resulted in constrained 10% AEP flooding through the urban center, which v1 struggled to do (Figure 8b). Similarly, in Waterloo, the 10% AEP flood was constrained (Figure 5b). Yet it is evident that floodplain feature preservation is no panacea: uncertainties relating to other model components inhibit substantial gains in performance. In Figure 8k, overprediction is still rife at Spencer regardless of DEM representation owing to disagreement over boundary conditions. The IFC 0.2% AEP model here used a peak discharge of $1,265 \text{ m}^3 \text{ s}^{-1}$, while the input to Fathom-US was $1,580 \text{ m}^3 \text{ s}^{-1}$. It is unsurprising, therefore, that simply differing the method of DEM construction did not reduce the false positives evident in v1 (Figure 8l).

3.3. Observation-Based Validation: Po River, Italy

3.3.1. Qualitative Examination of Levee Overtopping

The 2000 flood event in the Po floodplain was large enough in magnitude to submerge areas of the floodplain protected by the minor dike system but not so large that the major embankments were overtopped (estimated AEP of 2%). The most obvious test of the three model runs (MD, AD, and undefended), therefore, is a simple evaluation of any overtopping of the outer levees. The MD model, as expected, passes this test in overtopping only minor internal levees (Figures 10a to 10d(i)). The AD model (ϵ threshold of 300 in light of its functionality in section 3.2 and crest elevation accuracy in section 3.1) does this too: flow is contained entirely within the floodplain bounded by the outer levees (Figures 10a–10d(ii)). Crucially, the undefended model run simulates an overtopping of these outer levees across the domain (Figures 10a–10d(iii)). Though a somewhat aggregated and narrow view of model performance, levees holding in both defended model runs while overtopping in the undefended run demonstrate that the algorithm presented here represents flood defenses in a similar way to a laborious, manual, unscalable method. The undefended model here is essentially a proxy for how existing large-scale models operate: with no manual treatment of flood defenses possible, levee representation remains diminished in the DEM, and so floods that would be constrained in reality are not in the model. It likely stands to reason that all typical large-scale models fail to adequately model the Po River or any other defended river reaches globally in light of this analysis. Moreover, even when built with LiDAR topography and continuous surveyed bathymetry, the undefended model fails to simulate this event on the

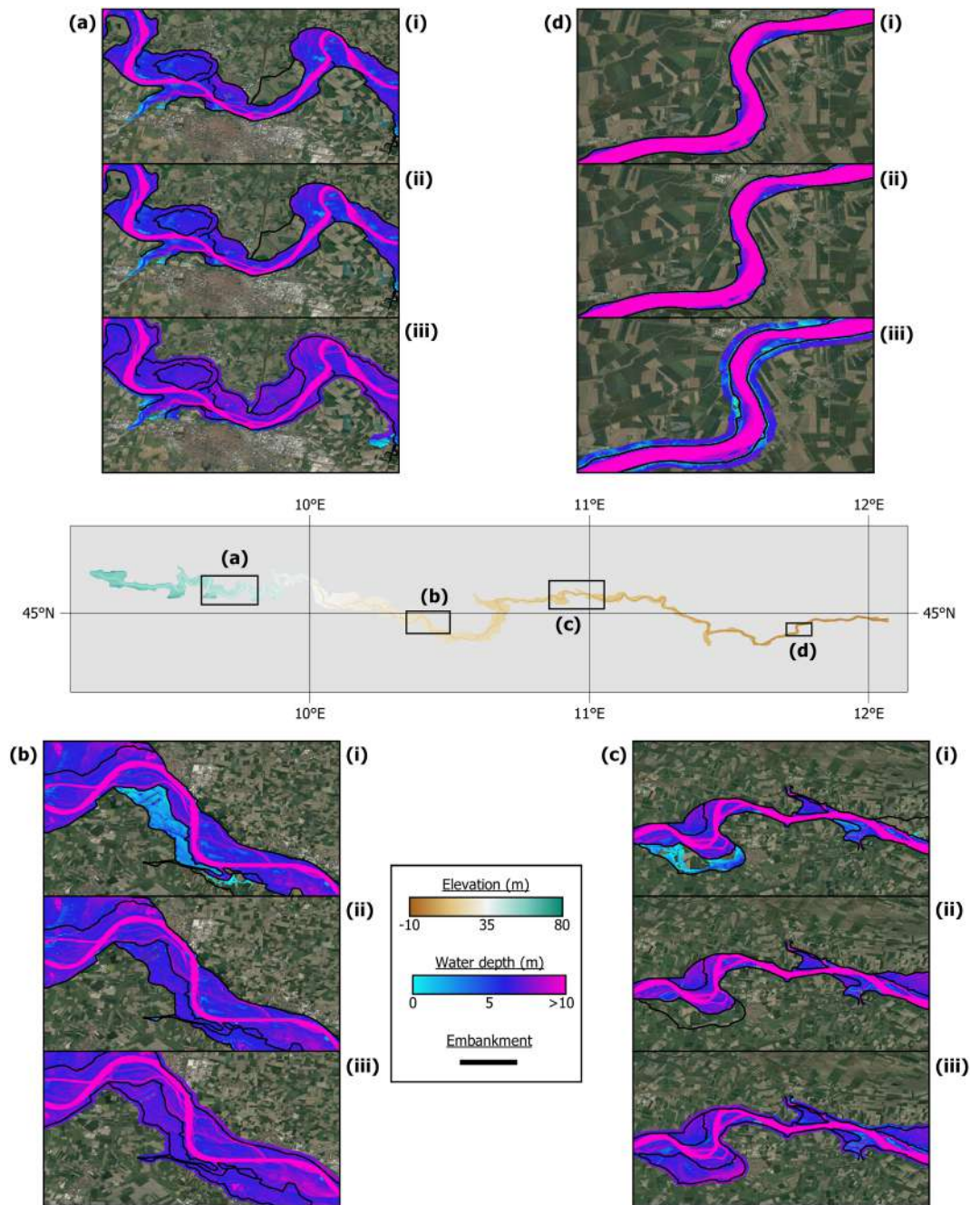


Figure 10. Four areas—(a) Piacenza, (b) Casalmaggiore, (c) just upstream of Ostiglia, and (d) just downstream of Pontelagoscuro—of the simulated 2000 flood event along the Po River for the three different digital elevation models: (i) manual defended; (ii) automated defended; and (iii) undefended. The digital elevation model only extends ~200 m from the outer embankments, making overtop visualization difficult. In all four example panels (and, indeed, across the entire domain) there is no outer levee overtopping in (i) and (ii), while overtopping in (iii).

Po accurately since major outer embankments overtop. Since they (correctly) did not overtop in the AD and MD models, this inaccuracy in inundation simulation is attributable to the DEM resampling process. Large-scale models, which would typically not benefit from even these data, are likely of little utility in modeling the Po. The specific topography of the basin—a levee-protected floodplain, which is flooded only when the stable channel experiences significant stage—poses a difficult modeling challenge (Castellarin et al., 2011a). Even in the presence of complete levee inventories, implementation of these data within the model build is

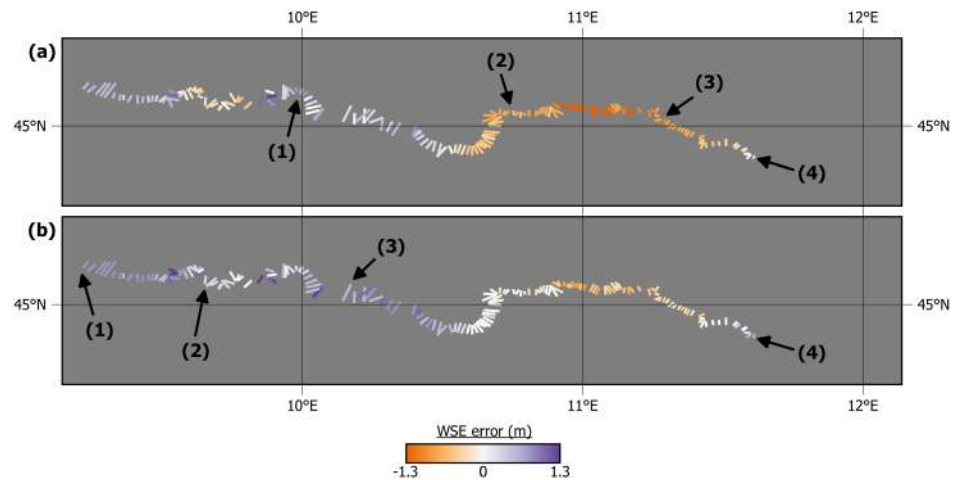


Figure 11. Maps showing the spatial distribution of maximum water surface elevation errors at each cross section when comparing observations to (a) manual-defended and (b) automated-defended model output. Positive errors (purple) indicate model overprediction; negative errors (orange) indicate model underprediction. The direction of flow is west to east. The numbers in (a) correspond to river gauges in Figure 14. The numbers in (b) refer to locations in Figure 12 or are referred to in the text.

sensitive and non-trivial. Meanwhile, the automated method presented reproduces experiential expectations of levee behavior akin to local modeling strategies.

3.3.2. Validation Against High-Water Marks

With the two simulations involving some treatment of defense structures performing similarly in terms of levee overtopping, we can further discriminate between the two models with some observational data. These take the form of post-event field surveys of maximum water surface elevation (WSE) for the 2000 flood event. In this analysis, we use 171 cross sections of maximum WSE distributed across the length of the Po domain studied (Figure 11). For the MD simulation (Figure 11a), the root-mean-squared error (RMSE) is 0.61 m with a slight bias towards underprediction (mean error = -0.25 m). With a DEM generated using the algorithm presented here (Figure 11b), the RMSE comes to 0.46 m with a minor tendency towards overprediction (mean error = 0.15 m). These errors are commensurate with those reported when validating high-resolution, calibrated, local models built with quality topography, bathymetry, and flow data (e.g., Mignot et al., 2006; Neal et al., 2009). Furthermore, errors in observed high-water marks during post-event surveys are typically 0.3–0.5 m (Fewtrell et al., 2011; Horritt et al., 2010), meaning both models are within or approaching observational error. Both models employ distributed roughness values, which are calibrated to the MD model, resulting in physically realistic drag coefficients similar to previous studies (e.g., Castellarin et al., 2011a): Manning’s n in channels is set at 0.0320–0.0395 and floodplains at 0.0430–0.0950. Importantly, errors are lower in the automated method of representing defenses than in the ubiquitous manual approach, in spite of friction being calibrated to the latter.

In both models, errors are generally low at the upstream (westerly) end of the domain, but performance begins to diverge at roughly 10.5°E, between Casalmaggiore (Location 3 in Figure 11b) and Pontelagoscuro (Location 4 in Figure 11b). The MD simulation becomes very underpredictive of maximum WSE here (up to 1.3 m at around 11°E), while performance remains high in the AD model. This suggests that the propagation of the flood wave is more accurately represented in the latter simulation than the former. The likely reason for this is that the method presented here not only preserves levee crest elevations in an efficient manner but also represents other hydraulically important features: to the extent that flood peaks across the domain are better captured than in models with a manual treatment of solely formal levee structures. Figure 12 exhibits this for areas in the upstream model domain. Levee compartments near the domain entrance (Figures 12a and 12b) were seemingly absent from the levee inventory used by the MD DEM but were automatically detected for the AD model. Near Piacenza (Figures 11c and 11d), levee-like features on the river bank are considered in the AD DEM, particularly on the northern banks, while remaining unresolved in the MD version. Maintaining river bank heights and elevating other microtopographic features that remain subgrid to the MD model resulted in faster conveyance of floodwaters downstream in the AD model,

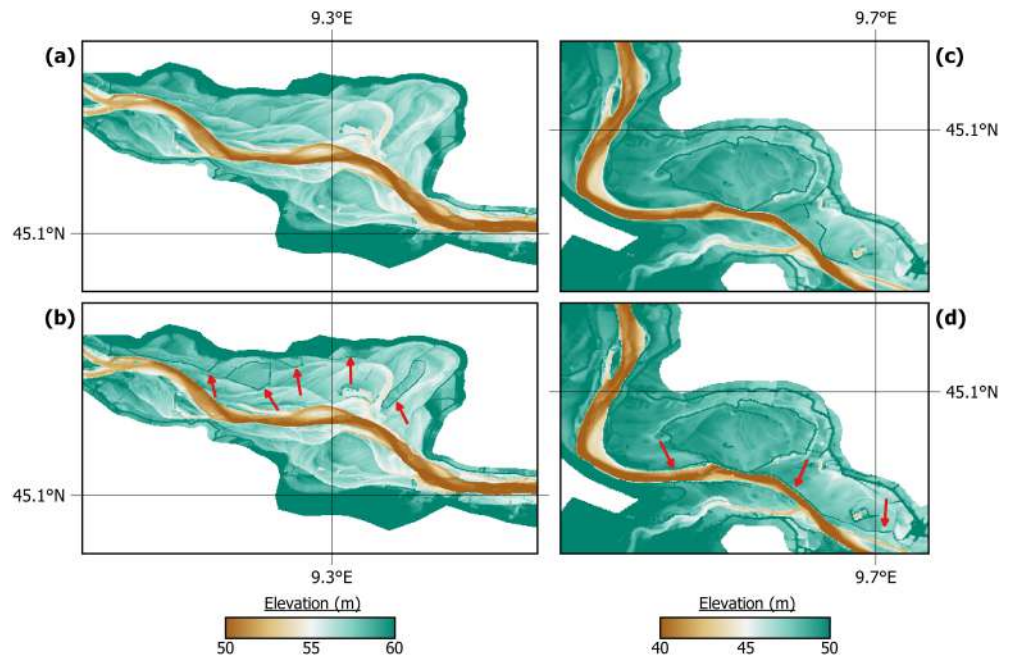


Figure 12. Maps highlighting the minor, yet crucial, manual-defended (MD) and automated-defended (AD) digital elevation model (DEM) differences (indicated by the red arrows). (a) and (b) are MD and AD DEMs, respectively, of the upstream domain boundary, at the confluence of the Po and Ticino rivers downstream of Pavia (Location 1 in Figure 11b). (c) and (d) are MD and AD DEMs, respectively, of an area just upstream of Piacenza (Location 2 in Figure 11b).

culminating in a more accurate representation of maximum WSE in the easterly (downstream) portion of the domain. Figure 13 illustrates this assertion. The widespread areas of purple shading in Figure 13a indicate that such areas were inundated much earlier (6–12+ hr) in the MD model than the AD model. The flood wave in the AD model thus propagated further downstream, rather than being stored and conveyed more slowly out of bank in these areas. This is corroborated by the blue compartments of Figure 13b, which show much deeper waters in the MD than AD simulation after 45 hr of model time, a volumetric difference of $\sim 2,450,000 \text{ m}^3$, despite identical boundary conditions. Out-of-bank flow or extensive floodplain water storage upstream that did not occur as early in the 2000 event (or the AD simulation) seemingly resulted in underprediction of maximum WSE downstream in the MD model. Representation of flood dynamics is less important in the upstream domain, since performance here is constrained by proximity to the inflow points. This is perhaps why both models perform similarly west of 10.5°E . Further downstream, however, a realistic conveyance of floodwaters begins to hold greater sway over performance (when comparing to WSE maxima).

3.3.3. Validation Against River Gauge Data

This phenomenon is yet further evidenced by comparison to river gauge observations of WSE. When comparing modeled WSE to those measured by the gauge at Cremona (upstream section of the domain; Location 1 in Figure 11a), both the MD and AD models perform similarly (Figure 14a). The MD model is slightly underpredictive at Cremona, consistent with the reasoning set out previously, while the AD model is marginally overpredictive here. It should be noted that gauged observations of water level are highly accurate ($<1\text{--}2 \text{ cm}$), providing suitably discriminatory validation data (Hamilton & Moore, 2012; McMillan et al., 2012; Schmidt, 2002; van der Made, 1982). At Borgoforte (Location 2 in Figure 11a), the MD model underpredicts the peak to a much greater magnitude (Figure 14b). The AD model replicates observations well here. Disparities on the rising limb are predominantly a function of model spin-up: for example, no water reaches Borgoforte until $\sim 50 \text{ hr}$ of model time; the Po would not be empty before this time in reality. At Sermide (Location 3 in Figure 11a), the MD model is severely underpredictive, misestimating peak WSE by $\sim 1 \text{ m}$ (Figure 14c). The AD model underpredicts the peak also but only by $\sim 0.5 \text{ m}$. The geometry reproduced by the AD approach seems to ensure a more faithful reproduction of the channel-floodplain

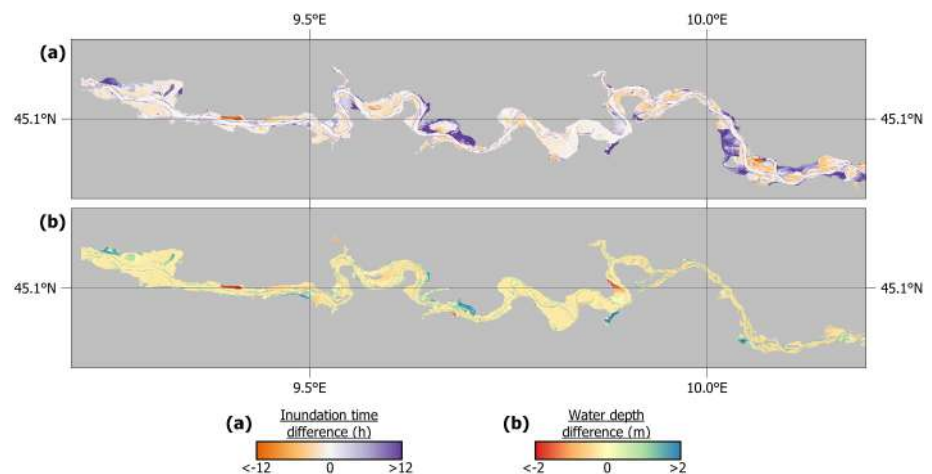


Figure 13. Maps of the upstream portion of the Po River floodplain showing (a) time of inundation difference between the models, where a positive (negative) difference indicates earlier (later) inundation in the manual-defended model than automated-defended model and (b) water depth difference 45 hr into the simulations, where a positive (negative) difference indicates deeper (shallower) water in the manual-defended model than automated-defended model.

interaction, providing an attenuation more similar to observations than with that obtained by the MD model. However, at the furthest downstream gauge (Pontelagosuro; Location 4 in Figure 11a), both models become slightly overpredictive (Figure 14d). Indeed, the MD model replicates observed WSE well at this gauge. This is inconsistent with the previously supported interpretation of results whereby the MD model fails to propagate the flood wave as faithfully as the AD model. The likely culprit here is the horizontal resolution of the DEM. Pontelagosuro is a location where the channel narrows to 150–200 m, with the large outer embankments only ~50 m from the river bank on either side. The Po channel in this domain can be up to 500 m wide, bounded by up to 3 km of floodplain. This is effectively a “bottleneck,” and so WSE will be very sensitive to the total storage volume afforded by the DEM. In such a narrow corridor, the horizontal resolution of the DEM becomes more important in defining this. Figure 14e is a cross section of the channel and floodplain at Pontelagosuro. It is evident that both the MD and AD DEMs reduce the capacity of the channel and floodplain here when compared to the high-resolution (1/9 arc second) source data, simply as a function of grid resolution. Preserving peaks from the 1/9 arc second data in the 1 arc-second DEM result in overrepresentation of these features in the horizontal plane. It is evident that there is less room for water storage in the coarser DEMs than would be present if the baseline 1/9 arc-second DEM was used. This is an important limitation in representing defenses in this way but crucially is endemic to both the common-practice, engineering-grade, manual techniques and the automated approach presented here. The accurate replication of observed WSE by the MD model is likely incidental therefore: the result of the previously asserted underprediction of discharge combined with lower storage capacity. An overpredictive bias in the AD model at Pontelagosuro is the probable product of more accurate modeled flow (higher than in the MD model) but with reduced storage volume available. Overall, therefore, there is consistency in the idea that the automated model not only performs similarly to manual methods by appropriately elevating flood defenses but can outperform them in ensuring representation of other hydraulically important subgrid scale features. At least, this is apparent on the Po, where representation of the minor embankment system is critical in the accurate replication of flood dynamics in the downstream portion of the domain. It should be noted that an MD approach can still obtain AD-like performance if these minor embankments are considered manually. In this test case, MD performance is an incidental product of the levee inventory containing only major levees. Should a modeler obtain and process extra information on these apparently critical floodplain features, the MD model would be indistinguishable from the AD model. The AD model, though, has the advantage of requiring no prior consideration of river dynamics in assessing which features to take note of. This suggests that such an approach may offer benefits to local modelers, too, in improving model-build efficiency.

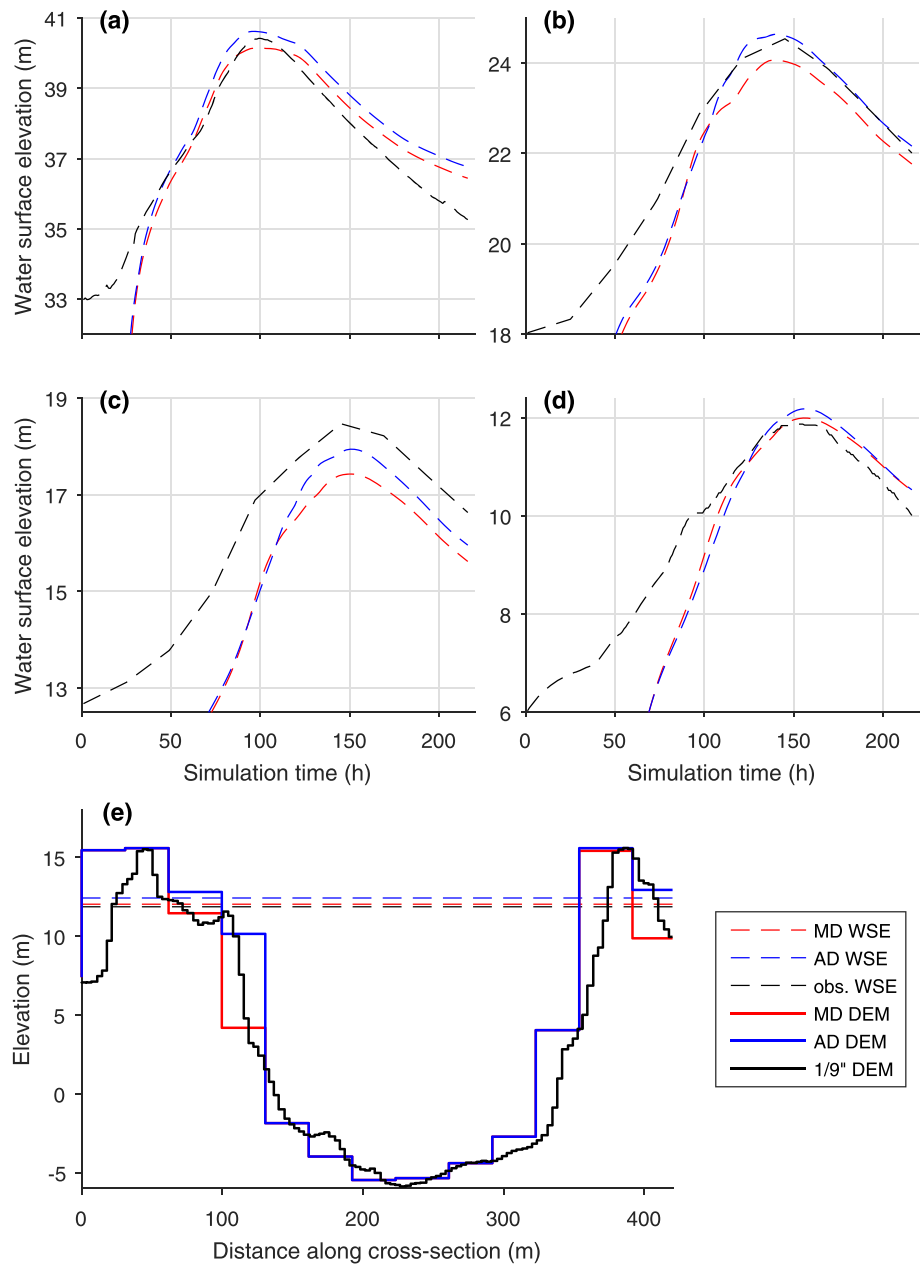


Figure 14. Comparison of manual-defended (MD) and automated-defended (AD) modeled water surface elevations to four river gauge observations of the 2000 event at (a) Cremona (hourly data); (b) Borgoforte (daily data); (c) Sermide (daily data); and (d) Pontelagoscuero (hourly data). Locations 1–4 in Figure 11a correspond to gauges (a)–(d). (e) is a cross section of MD, AD, and baseline 1/9 arc-second digital elevation models as well as MD, AD, and observed water surface elevations at Pontelagoscuero.

4. Conclusions

Overall, then, the new method of defense representation in DEMs presented here has significant skill in accurately defining levee crest elevations. When validated against the CLD (section 3.1), crest heights were underestimated by the defended DEM by 0.05 m on average, compared to 1.16 m by the undefended DEM. When incorporated into a hydraulic modeling framework, the presented method of DEM building generally improved flood extent and water level fits to model and observational benchmarks. Conventional performance metrics remained high for the Iowa tests (section 3.2), with a trend towards convergence with

local-scale models. However, increases in skill versus an undefended model were relatively muted, partly due to deficiencies in the comparison procedure itself, and interscale model deviations in a risk context were substantial. The Po test case (section 3.3), though, offers crucial evidence in support of the validity of this method when applied in conjunction with more accurate data: (a) being a real, gauged flood event, boundary condition uncertainty is much reduced relative to the theoretical design floods in the Iowa tests; (b) complete channel bathymetry and accurate hydrography offer considerable improvement in river representation relative to the HydroSHEDS subgrid channels employed in the United States; (c) total high-resolution LiDAR coverage of the Po results in a more accurate DEM than that of Iowa when using the USGS NED; and (d) validity benchmarks are better defined for the Po (observed high-water marks, gauged stage, and no outer levee overtopping) than for Iowa (local-scale models).

This leads to a number of limitations in the application of this method to large-scale models. First, it requires a high-quality, high-resolution baseline DEM that a modeler would typically coarsen. This effectively rules out the majority of the globe, where the best source of free elevation data remains corrected Shuttle Radar Topography Mission–based products, which have low vertical accuracy and are employed at native resolution. Applicable areas—those with wide-area LiDAR (or similar accuracy) coverage—are confined to North America, western Europe, and parts of Australia. The method itself is grid- and neighborhood-scale dependent, requiring due consideration for different resolution source and simulation DEMs. Furthermore, the extraction rate threshold is difficult to define *a priori*. Based on visual inspections, we identify 300 to be appropriate, but there is no formal justification for this. Derivation of a threshold from known defense locations is problematic, given that this would unintentionally penalize the detection of hydraulically important, non-levee features as false alarms, even if the reference defense inventory was complete. Despite this, we have shown the resultant DEM to be relatively insensitive to the threshold chosen. So long as a suitably low threshold is chosen to capture the features of interest, the effect of heightening extraneous features is negligible. Future research could seek to calibrate the extraction rate (ϵ) threshold to geodetic surveys of all hydraulically relevant features or even treat ϵ as an uncertain parameter in hydraulic model calibration studies, maximizing some measure of fit to a benchmark by varying the ϵ threshold. Further, the method is agnostic towards the features it identifies and elevates; there is no judgment on their characteristics except height. The fragility of these features is thus unknown: they are not all engineered structures designed to hold back floodwaters, but could fail under such conditions. In which case, undefended model runs may still be useful to explore impacts should the elevated features not remain so during a flood. One final limitation, evident in the Po versus Iowa tests, is that other features of the hydraulic model need to be of commensurate quality for any benefit to be felt. The method will have no meaningful benefit for models utilizing hydraulic codes with poor physical representation of flow, an overcoarsened DEM (e.g., $>10^2$ m resolution), inaccurately defined boundary conditions, and deficient representation of river channels.

Fathom-US generally performs well in Iowa, yet there is clear room for further improvement. The Po model provides an indication of simulation quality should other model components be improved. While improvements to extreme flow quantification are likely impossible without denser river gauge networks and centuries of observations, large-scale modelers can anticipate developments elsewhere. Increased coverage of high-quality elevation data is eminently possible with relatively modest investments from global governments (Schumann & Bates, 2018); updates to the USGS National Hydrography Dataset may supplant the cruder global HydroSHEDS representation of river channels for the United States, and MERIT-Hydro shows promise at the global scale (Yamazaki et al., 2019); and emerging data sets detailing fluvial geomorphological relationships (e.g., Frasson et al., 2019) from satellite data can be used to constrain approximations to channel capacity, while NASA's imminent Surface Water and Ocean Topography satellite mission (Biancamaria et al., 2016) may provide new opportunities for estimating river bathymetry from space (e.g., Pavelsky et al., 2014; Yoon et al., 2012). Meanwhile, this method may find applicability in smaller scale studies given the improvements to common local modeling strategies of the Po. Aside from being a more faithful DEM representation of the floodplain, raw algorithm output can be used to identify areas requiring further survey, and automation provides important advantages over the laborious and non-trivial process of manual burning.

Thus, the method is shown to provide skill in the detection of flood defenses and other important hydraulic features in the absence of local information. It provides a crucial first step in ensuring

physically realistic flood defenses in large-scale models, moving away from the largely unsubstantiated assumptions underpinning previous approaches. When it is a lack of defense information that is the dominant source of error in a hydraulic model, the method presented here offers considerable improvement.

Acknowledgments

The U.S. Army Corps of Engineers National Levee Database is available from <https://levees.sec.usace.army.mil/>. The USGS NED can be accessed from <https://ned.usgs.gov/>. The socioeconomic data in Figure 1 are available from the U.S. Census Bureau (<https://www.census.gov/data.html>) and the Multi-Resolution Land Characteristics Consortium (<https://www.mrlc.gov/>). The authors thank Agenzia Interregionale per il Fiume Po and Autorità di Bacino del Fiume Po for granting access to their data of the Po River basin. They are also grateful to the State of California Department of Water Resources for granting us access to the CLD. The IFC benchmark inundation maps are available from the Iowa Flood Information System (<https://ifis.iowafloodcenter.org/>). The dasymetric population map of the United States can be obtained from the EPA Environmental Dataset Gateway (<https://edg.epa.gov/>). Fathom-US is a proprietary model but can be made available for academic research use (i.e., not for commercial, policy, or regulatory purposes) by contacting info@fathom.global. O. E. J. W., P. D. B., J. C. N., C. S. S., A. M. S., and N. Q. are affiliated with Fathom, a flood analytics firm based in the United Kingdom. P. D. B. was supported by a Research Fellowship from the Leverhulme Trust and a Royal Society Wolfson Research Merit Award. J. C. N. was supported by National Environmental Research Council grant NE/S006079/1. D. W. G., R. G., and W. F. K. were supported by the Iowa Flood Center.

References

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., et al. (2017). Global projections of river flood risk in a warmer world. *Earth's Future*, 5, 171–182. <https://doi.org/10.1002/2016EF000485>

Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science*, 361(6402), 585–588. <https://doi.org/10.1126/science.aat0636>

American Society of Civil Engineers (2017). *2017 Infrastructure report card: Levees*. VA. [Available at: Reston. <https://www.infrastructurereportcard.org/cat-item/levees/>]

Archer, L., Neal, J. C., Bates, P. D., & House, J. I. (2018). Comparing TanDEM-X data with frequently used DEMs for flood inundation modeling. *Water Resources Research*, 54(12), 10,205–10,222. <https://doi.org/10.1029/2018WR023688>

Aronica, G., Bates, P. D., & Horritt, M. S. (2002). Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processes*, 16(10), 2001–2006. <https://doi.org/10.1002/hyp.398>

Bailly, J. S., Lagacherie, P., Millier, C., Puech, C., & Kosuth, P. (2008). Agrarian landscapes linear features detection from LiDAR: Application to artificial drainage networks. *International Journal of Remote Sensing*, 29(12), 3489–3508. <https://doi.org/10.1080/01431160701469057>

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. *Journal of Hydrology*, 387(1–2), 33–45. <https://doi.org/10.1016/j.jhydrol.2010.03.027>

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. *Journal of Hydrology*, 328(1–2), 306–318. <https://doi.org/10.1016/j.jhydrol.2005.12.028>

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

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (2013). *Runoff prediction in ungauged basins*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9781139235761>

Casas, A., Riaño, D., Greenberg, J., & Ustin, S. (2012). Assessing levee stability with geometric parameters derived from airborne LiDAR. *Remote Sensing of Environment*, 117, 281–288. <https://doi.org/10.1016/j.rse.2011.10.003>

Castellarin, A., Di Baldassarre, G., Bates, P. D., & Brath, A. (2009). Optimal cross-sectional spacing in Preissmann scheme 1D hydrodynamic models. *Journal of Hydraulic Engineering*, 135(2), 96–105. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2009\)135:2\(96\)](https://doi.org/10.1061/(ASCE)0733-9429(2009)135:2(96))

Castellarin, A., Di Baldassarre, G., & Brath, A. (2011). Floodplain management strategies for flood attenuation in the River Po. *River Research and Applications*, 27(8), 1037–1047. <https://doi.org/10.1002/rra.1405>

Castellarin, A., Domeneghetti, A., & Brath, A. (2011). Identifying robust large-scale flood risk mitigation strategies: A quasi-2D hydraulic model as a tool for the Po River. *Physics and Chemistry of the Earth, Parts A/B/C*, 36(7–8), 299–308. <https://doi.org/10.1016/j.pce.2011.02.008>

Cazorzi, F., Dalla Fontana, G., De Luca, A., Sofia, G., & Tarolli, P. (2013). Drainage network detection and assessment of network storage capacity in agrarian landscape. *Hydrological Processes*, 27(4), 541–553. <https://doi.org/10.1002/hyp.9224>

Chen, B., Krajewski, W. F., Goska, R., & Young, N. (2017). Using LiDAR surveys to document floods: A case study of the 2008 Iowa flood. *Journal of Hydrology*, 553, 338–349. <https://doi.org/10.1016/j.jhydrol.2017.08.009>

Choung, Y. (2014). Mapping levees using LiDAR data and multispectral orthoimages in the Nakdong river basins, South Korea. *Remote Sensing*, 6(9), 8696–8717. <https://doi.org/10.3390/rs6098696>

Ciullo, A., de Bruijn, K. M., Kwakkel, J. H., & Klijn, F. (2019). Accounting for the uncertain effects of hydraulic interactions in optimizing embankment heights: Proof of principle for the IJssel River. *Journal of Flood Risk Management*. <https://doi.org/10.1111/jfr3.12532>

de Almeida, G. A. M., & Bates, P. (2013). Applicability of the local inertial approximation of the shallow water equations to flood modeling. *Water Resources Research*, 49, 4833–4844. <https://doi.org/10.1002/wrcr.20366>

Di Baldassarre, G., Castellarin, A., & Brath, A. (2010). Analysis of the effects of levee heightening on flood propagation: Example of the River Po, Italy. *Hydrological Sciences Journal*, 54, 1007–1017. <https://doi.org/10.1623/hysj.54.6.1007>

Domeneghetti, A., Carisi, F., Castellarin, A., & Brath, A. (2015). Evolution of flood risk over large areas: Quantitative assessment for the Po River. *Journal of Hydrology*, 527, 809–823. <https://doi.org/10.1016/j.jhydrol.2015.05.043>

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. *Advances in Water Resources*, 94, 87–102. <https://doi.org/10.1016/j.advwatres.2016.05.002>

Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., et al. (2018). Increased human and economic losses from river flooding with anthropogenic warming. *Nature Climate Change*, 8(9), 781–786. <https://doi.org/10.1038/s41558-018-0257-z>

Evans, I. S. (1979). An integrated system of terrain analysis and slope mapping. In *Final report on grant DA-ERO-591-73-G0040*. Durham, UK: University of.

Evans, I. S. (1980). An integrated system of terrain analysis and slope mapping. *Zeitschrift für Geomorphologie Suppl.-Bd*, 36, 274–295.

Fewtrell, T. J., Bates, P. D., Horritt, M., & Hunter, N. M. (2008). Evaluating the effect of scale in flood inundation modelling in urban environments. *Hydrological Processes*, 22(26), 5107–5118. <https://doi.org/10.1002/hyp.7148>

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. *Hydrological Processes*, 25(20), 3173–3186. <https://doi.org/10.1002/hyp.8035>

Feyen, L., Dankers, R., Bódis, K., Salamon, P., & Barredo, J. I. (2012). Fluvial flood risk in Europe in present and future climates. *Climatic Change*, 112(1), 47–62. <https://doi.org/10.1007/s10584-011-0339-7>

- Fleischmann, A., Paiva, R., & Collischonn, W. (2019). Can regional to continental river hydrodynamic models be locally relevant? A cross-scale comparison. *Journal of Hydrology X*, 3, 100027. <https://doi.org/10.1016/j.hydroa.2019.100027>
- Frasson, R. P. d. M., Pavelsky, T. M., Fonstad, M. A., Durand, M. T., Allen, G. H., Schumann, G., et al. (2019). Global relationships between river width, slope, catchment area, meander wavelength, sinuosity, and discharge. *Geophysical Research Letters*, 46(6), 3252–3262. <https://doi.org/10.1029/2019GL082027>
- 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. *U.S. Geological Survey Open-File Report*, 2014–1008. <https://doi.org/10.3133/ofr20141008>
- Gilles, D., Young, N., Schroeder, H., Piotrowski, J., & Chang, Y.-J. (2012). Inundation mapping initiatives of the Iowa Flood Center: Statewide coverage and detailed urban flooding analysis. *Watermark*, 4(1), 85–106. <https://doi.org/10.3390/w4010085>
- Hamilton, A. S., & Moore, R. D. (2012). Quantifying uncertainty in streamflow records. *Canadian Water Resources Journal*, 37(1), 3–21. <https://doi.org/10.4296/cwrj3701865>
- Heine, R. A., & Pinter, N. (2012). Levee effects upon flood levels: An empirical assessment. *Hydrological Processes*, 26, 3225–3240. <https://doi.org/10.1002/hyp.8261>
- Hooijer, A., Klijn, F., Pedroli, G. B. M., & Van Os, A. G. (2004). Towards sustainable flood risk management in the Rhine and Meuse river basins: Synopsis of the findings of IRMA-SPONGE. *River Research and Applications*, 20(3), 343–357. <https://doi.org/10.1002/rra.781>
- Horna-Muñoz, D., & Constantinescu, G. (2018). A fully 3-D numerical model to predict flood wave propagation and assess efficiency of flood protection measures. *Advances in Water Resources*, 122, 148–165. <https://doi.org/10.1016/j.advwatres.2018.10.014>
- 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. *Proceedings of the Institution of Civil Engineers*, 163(WM6), 273–281. <https://doi.org/10.1680/wama.2010.163.6.273>
- Hunter, N. M., Bates, P. D., Horritt, M. S., & Wilson, M. D. (2007). Simple spatially-distributed models for predicting flood inundation: A review. *Geomorphology*, 90(3–4), 208–225. <https://doi.org/10.1016/j.geomorph.2006.10.021>
- Krajewski, W. F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., et al. (2017). Real-time flood forecasting and information system for the State of Iowa. *Bulletin of the American Meteorological Society*, 98(3), 539–554. <https://doi.org/10.1175/BAMS-D-15-00243.1>
- Krüger, T. (2010). Algorithms for detecting and extracting dikes from digital terrain models. In A. Car, G. Griesebner, & J. Strobl (Eds.), *Geospatial Crossroads @ GI Forum-10 Proceedings of the Geoinformatics Forum Salzburg* (pp. 130–139). Berlin, Offenbach: Wichmann, S.
- Krüger, T., & Meinel, G. (2008). Using raster DTM for dike modelling. In P. van Oosterom, S. Zlatanova, F. Penninga, & E. M. Fendel (Eds.), *Advances in 3D Geoinformation Systems*. Berlin, Heidelberg: Lecture Notes in Geoinformation and Cartography. Springer. https://doi.org/10.1007/978-3-540-72135-2_6
- Lamb, R., Crossley, M., & Waller, S. (2009). A fast two-dimensional floodplain inundation model. *Proceedings of the Institution of Civil Engineers: Water Management*, 162(6), 363–370. <https://doi.org/10.1680/wama.2009.162.6.363>
- Lashermes, B., Foufloula-Georgiou, E., & Dietrich, W. E. (2007). Channel network extraction from high resolution topography using wavelets. *Geophysical Research Letters*, 34, L23S04. <https://doi.org/10.1029/2007GL031140>
- Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15), 2171–2186. <https://doi.org/10.1002/hyp.9740>
- Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions of the American Geophysical Union*, 89(10), 93–104. <https://doi.org/10.1029/2008EO100001>
- Masoero, A., Claps, P., Asselman, N. E. M., Mosselman, E., & Di Baldassarre, G. (2013). Reconstruction and analysis of the Po River inundation of 1951. *Hydrological Processes*, 27, 1341–1348. <https://doi.org/10.1002/hyp.9558>
- McMillan, H., Krueger, T., & Freer, J. (2012). Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrological Processes*, 26(26), 4078–4111. <https://doi.org/10.1002/hyp.9384>
- Mignot, E., Paquier, A., & Haider, S. (2006). Modeling floods in a dense urban area using 2D shallow water equations. *Journal of Hydrology*, 327(1–2), 186–199. <https://doi.org/10.1016/j.jhydrol.2005.11.026>
- Moretti, G., & Orlandini, S. (2018). Hydrography-driven coarsening of grid digital elevation models. *Water Resources Research*, 54(5), 3654–3672. <https://doi.org/10.1029/2017WR021206>
- Neal, J., Dunne, T., Sampson, C., Smith, A., & Bates, P. (2018). Optimisation of the two-dimensional hydraulic model LISFLOOD-FP for CPU architecture. *Environmental Modelling and Software*, 107, 148–157. <https://doi.org/10.1016/j.envsoft.2018.05.011>
- 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 Resources Research*, 48, W11506. <https://doi.org/10.1029/2012WR012514>
- 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. *Journal of Hydrology*, 368(1–4), 42–55. <https://doi.org/10.1016/j.jhydrol.2009.01.026>
- Pappenberger, F., Beven, K., Frodsham, K., Romanowicz, R., & Matgen, P. (2007). Grasping the unavoidable subjectivity in calibration of flood inundation models: A vulnerability weighted approach. *Journal of Hydrology*, 333(2–4), 275–287. <https://doi.org/10.1016/j.jhydrol.2006.08.017>
- Pappenberger, F., Dutra, E., Wetterhall, F., & Cloke, H. L. (2012). Deriving global flood hazard maps of fluvial floods through a physical model cascade. *Hydrology and Earth System Sciences*, 16, 4143–4156. <https://doi.org/10.5194/hess-16-4143-2012>
- Passalacqua, P., Belmont, P., & Foufloula-Georgiou, E. (2012). Automatic geomorphic feature extraction from LiDAR in flat and engineered landscapes. *Water Resources Research*, 48, W03528. <https://doi.org/10.1029/2011WR010958>
- Passalacqua, P., Tarolli, P., & Foufloula-Georgiou, E. (2010). Testing space-scale methodologies for automatic geomorphic feature extraction from LiDAR in a complex mountainous landscape. *Water Resources Research*, 46, W11535. <https://doi.org/10.1029/2009WR008812>
- Pavelsky, T. M., Durand, M. T., Andreadis, K. M., Beighley, R. E., Paiva, R. C. D., Allen, G. H., & Miller, Z. F. (2014). Assessing the potential global extent of SWOT river discharge observations. *Journal of Hydrology*, 519(B), 1516–1525. <https://doi.org/10.1016/j.jhydrol.2014.08.044>
- Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., et al. (2019). The spatial dependence of flood hazard and risk in the United States. *Water Resources Research*, 55(3), 1890–1911. <https://doi.org/10.1029/2018WR024205>
- 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 Resources Research*, 51, 7358–7381. <https://doi.org/10.1002/2015WR016954>
- Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., & Trigg, M. A. (2016). Perspectives on open access high resolution digital elevation models to produce global flood hazard layers. *Frontiers in Earth Science*, 3, 85. <https://doi.org/10.3389/feart.2015.00085>

- Sanders, B. F., & Schubert, J. E. (2019). PRiMo: Parallel raster inundation model. *Advances in Water Resources*, *126*, 79–95. <https://doi.org/10.1016/j.advwatres.2019.02.007>
- Sangireddy, H., Stark, C. P., Kladzyk, A., & Passalacqua, P. (2016). GeoNet: An open source software for the automatic and objective extraction of channel heads, channel network, and channel morphology from high resolution topography data. *Environmental Modelling & Software*, *83*, 58–73. <https://doi.org/10.1016/j.envsoft.2016.04.026>
- Savage, J. T. S., Bates, P., Freer, J., Neal, J., & Aronica, G. (2016). When does spatial resolution become spurious in probabilistic flood inundation predictions? *Hydrological Processes*, *30*(13), 2014–2032. <https://doi.org/10.1002/hyp.10749>
- Schmidt, A. R. (2002). Analysis of stage-discharge relations for open-channel flow and their associated uncertainties. PhD Thesis: Department of Civil and Environmental Engineering, University of Illinois at Urban-Champaign, Urbana-Champaign, IL.
- Schumann, G., Di Baldassarre, G., Alsdorf, D., & Bates, P. D. (2010). Near real-time flood wave approximation on large rivers from space: Application to the River Po, Italy. *Water Resources Research*, *46*, W05601. <https://doi.org/10.1029/2008WR007672>
- Schumann, G. J.-P. (2014). Fight floods on a global scale. *Nature*, *507*(7491), 169. <https://doi.org/10.1038/507169e>
- Schumann, G. J.-P., & Bates, P. D. (2018). The need for a high-accuracy, open-access global DEM. *Frontiers in Earth Science*, *6*, 225. <https://doi.org/10.3389/feart.2018.00225>
- 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 global database of flood protection standards. *Natural Hazards and Earth System Sciences*, *16*, 1049–1061. <https://doi.org/10.5194/nhess-16-1049-2016>
- Smith, A., Sampson, C., & Bates, P. (2015). Regional flood frequency analysis at the global scale. *Water Resources Research*, *51*, 539–553. <https://doi.org/10.1002/2014WR015814>
- Sofia, G., Dalla Fontana, G., & Tarolli, P. (2014). High-resolution topography and anthropogenic feature extraction: Testing geomorphometric parameters in floodplains. *Hydrological Processes*, *28*(4), 2046–2061. <https://doi.org/10.1002/hyp.9727>
- Sofia, G., Tarolli, P., Cazorzi, F., & Dalla Fontana, G. (2011). An objective approach for feature extraction: Distribution analysis and statistical descriptors for scale choice and channel network identification. *Hydrology and Earth System Sciences*, *15*, 1387–1402. <https://doi.org/10.5194/hess-15-1387-2011>
- Steinfeld, C. M. M., Kingsford, R. T., & Laffan, S. W. (2013). Semi-automated GIS techniques for detecting floodplain earthworks. *Hydrological Processes*, *27*(4), 579–591. <https://doi.org/10.1002/hyp.9244>
- Stephens, E., Schumann, G., & Bates, P. (2014). Problems with binary pattern measures for flood model evaluation. *Hydrological Processes*, *28*(18), 4928–4937. <https://doi.org/10.1002/hyp.9979>
- Tarolli, P., Sofia, G., & Dalla Fontana, G. (2012). Geomorphic features extraction from high-resolution topography: Landslide crowns and bank erosion. *Natural Hazards*, *61*(1), 65–83. <https://doi.org/10.1007/s11069-010-9695-2>
- Trigg, M. A., Birch, C. E., Neal, J. C., Bates, P. D., Smith, A., Sampson, C. C., et al. (2016). The credibility challenge for global fluvial flood risk analysis. *Environmental Research Letters*, *11*, 094014. <https://doi.org/10.1088/1748-9326/11/9/094014>
- van der Made, J. E. (1982). Determination of the accuracy of water level observations. *Proceedings of the Exeter Symposium, IAHS Publications*, *134*, 172–184.
- Villarini, G., Serinaldi, F., Smith, J. A., & Krajewski, W. F. (2009). On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resources Research*, *45*, W08417. <https://doi.org/10.1029/2008WR007645>
- Ward, P. J., Jongman, B., Aerts, J. C. J. H., Bates, P. D., Botzen, W. J. W., Loaiza, A. D., et al. (2017). A global framework for future costs and benefits of river-flood protection in urban areas. *Nature Climate Change*, *7*(9), 642–646. <https://doi.org/10.1038/nclimate3350>
- Werner, M. G. F., Hunter, N. M., & Bates, P. D. (2005). Identifiability of distributed floodplain roughness values in flood extent estimation. *Journal of Hydrology*, *314*(1–4), 139–157. <https://doi.org/10.1016/j.jhydrol.2005.03.012>
- Wilson, M., Bates, P., Alsdorf, D., Forsberg, B., Horritt, M., Melack, J., et al. (2007). Modeling large-scale inundation of Amazonian seasonally flooded wetlands. *Geophysical Research Letters*, *34*, L15404. <https://doi.org/10.1029/2007GL030156>
- Wing, O., Bates, P., Sampson, C., Smith, A., Fargione, J., & Johnson, K. (2018). Millions more Americans face flood risks than previously thought. *Eos*, *99*. <https://doi.org/10.1029/2018EO106321>
- Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. J., & Erickson, T. A. (2017). Validation of a 30 m resolution flood hazard model of the conterminous United States. *Water Resources Research*, *53*, 7968–7986. <https://doi.org/10.1002/2017WR020917>
- Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J., & Bouwman, A. (2013). A framework for global river flood risk assessments. *Hydrology and Earth System Sciences*, *17*, 1871–1892. <https://doi.org/10.5194/hess-17-1871-2013>
- Winsemius, H. C., Ward, P. J., Gayton, I., Veldhuis, M.-C., Meijer, D. H., & Iliffe, M. (2019). Commentary: The need for high-accuracy, open-access global DEM. *Frontiers in Earth Science*, *7*, 33. <https://doi.org/10.3389/feart.2019.00033>
- Wood, J. D. (1996). The geomorphological characteristics of digital elevation models. PhD Thesis: Department of Geography, University of Leicester.
- Wood, M., de Jong, S. M., & Straatsma, M. W. (2018). Locating flood embankments using SAR time series: A proof of concept. *International Journal of Applied Earth Observation and Geoinformation*, *70*, 72–83. <https://doi.org/10.1016/j.jag.2018.04.003>
- Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT Hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resources Research*, *55*. <https://doi.org/10.1029/2019WR024873>
- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., et al. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, *44*, 5844–5853. <https://doi.org/10.1002/2017GL072874>
- Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, *47*, W04501. <https://doi.org/10.1029/2010WR009726>
- Yoon, Y., Durand, M., Merry, C. J., Clark, E. A., Andreadis, K. M., & Alsdorf, D. E. (2012). Estimating river bathymetry from data assimilation of synthetic SWOT measurements. *Journal of Hydrology*, *464–465*, 363–375. <https://doi.org/10.1016/j.jhydrol.2012.07.028>