OPEN ACCESS

University of BRISTOL

Wing, O., Bates, P., Sampson, C., Smith, A., Johnson, K., & Erickson, T. (2017). Validation of a 30m resolution flood hazard model of the conterminous United States. *Water Resources Research*. https://doi.org/10.1002/2017WR020917

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
10.1002/2017WR020917

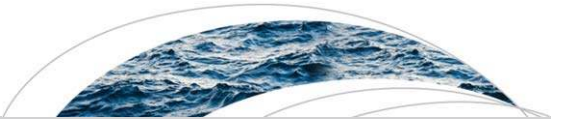
Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

AGU PUBLICATIONS

Water Resources Research



RESEARCH ARTICLE

10.1002/2017WR020917

Validation of a 30 m resolution flood hazard model of the conterminous United States

Oliver E. J. Wing[1,2], Paul D. Bates[1,2], Christopher C. Sampson[1,2], Andrew M. Smith[1,2], Kris A. Johnson[3], and Tyler A. Erickson[4]

[1]School of Geographical Sciences, University of Bristol, Bristol, UK, [2]Fathom, Engine Shed, Station Approach, Bristol, UK, [3]The Nature Conservancy, Minneapolis, Minnesota, USA, [4]Google Incorporated, Mountain View, California, USA

Key Points:
- A 30 m resolution flood hazard model of the entire conterminous United States is built using publicly available data
- Delineations of flood hazard are comprehensively validated against United States government agency benchmarks
- Model performance is largely comparable to quality local models, offering cheaper hazard information with complete spatial coverage

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

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

Received 10 APR 2017
Accepted 24 JUL 2017
Accepted article online 1 AUG 2017

Abstract This paper reports the development of a ~30 m resolution two-dimensional hydrodynamic model of the conterminous U.S. using only publicly available data. The model employs a highly efficient numerical solution of the local inertial form of the shallow water equations which simulates fluvial flooding in catchments down to 50 km^2 and pluvial flooding in all catchments. Importantly, we use the U.S. Geological Survey (USGS) National Elevation Dataset to determine topography; the U.S. Army Corps of Engineers National Levee Dataset to explicitly represent known flood defenses; and global regionalized flood frequency analysis to characterize return period flows and rainfalls. We validate these simulations against the complete catalogue of Federal Emergency Management Agency (FEMA) Special Flood Hazard Area (SFHA) maps and detailed local hydraulic models developed by the USGS. Where the FEMA SFHAs are based on high-quality local models, the continental-scale model attains a hit rate of 86%. This correspondence improves in temperate areas and for basins above 400 km^2. Against the higher quality USGS data, the average hit rate reaches 92% for the 1 in 100 year flood, and 90% for all flood return periods. Given typical hydraulic modeling uncertainties in the FEMA maps and USGS model outputs (e.g., errors in estimating return period flows), it is probable that the continental-scale model can replicate both to within error. The results show that continental-scale models may now offer sufficient rigor to inform some decision-making needs with dramatically lower cost and greater coverage than approaches based on a patchwork of local studies.

1. Introduction

Large-scale hydraulic analyses have come to the fore in recent years as a result of advances in computational capacity and availability of global terrain data sets [*Sampson et al.*, 2015; *Dottori et al.*, 2016; *Winsemius et al.*, 2013]. In particular, the release of NASA's Shuttle Radar Topography Mission (SRTM), providing elevation data across the world [*Rabus et al.*, 2003], has permitted the expansion of hydraulic modeling from exclusively local reach-scale studies to continental-scale and global-scale analyses. The vertical accuracy of large-scale terrain data sets remains the greatest barrier to obtaining accurate flood inundation projections [*Schumann et al.*, 2014], with root-mean-square errors in SRTM well exceeding depths at which water can damage property [*Gesch et al.*, 2014]. Alongside accuracy issues, voids, speckle, and significant biases in urban and forested areas hamper the utility of SRTM in its application to hydraulic modeling. Even with major conditioning, such as void removal [*Lehner et al.*, 2008], systematic vegetation and urbanization correction [*Baugh et al.*, 2013; *Elvidge et al.*, 2007], and noise reduction [*Gallant*, 2011], the data set still deviates significantly from highly accurate geodetic measurements [*Sampson et al.*, 2015].

A further issue with hydraulic analyses at continental to global scales is that they have rarely undergone testing against high-quality data of commensurate coverage. For instance, *Trigg et al.* [2016] conducted a continent-wide intercomparison of six global model outputs over Africa. While they adopted a large-scale validation procedure, the validation data itself are not derived from high-quality flood hazard assessments. *Sampson et al.* [2015] compared their global model to three Canadian urban river reaches and two UK catchments, with high-quality flood hazard data provided for these areas by their respective government agencies. In this instance, the benchmark data were of high quality but not of adequate spatial scale to comprehensively evaluate their global model.

© 2017. 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.

In light of this, there is a clear need for large-scale flood hazard models constructed using accurate topographic data and for a high-quality benchmark data set of similar spatial coverage with which to validate them. An area that can satisfy these requirements is the United States of America. The United States Geological Survey (USGS) produces the National Elevation Dataset (NED), which has a vertical accuracy far superior to any global data set [*Gesch et al*., 2014]. The U.S. also possesses flood hazard information across ~61% of its contiguous land area. Its National Flood Insurance Programme (NFIP) exists to mitigate the impacts of flooding on public and private property. The specification of areas within a hazard zone is therefore necessitated and is fulfilled by the Federal Emergency Management Agency (FEMA) who determine the Special Flood Hazard Area (SFHA). In its legal sense, the SFHA is where NFIP stipulates that the purchase of flood insurance is compulsory. In its hydrological sense, the SFHA delineates the area that would be inundated by a so-called 1 in 100 year flood, which is an event that has a 1% chance of being equaled or exceeded in any given year [*Federal Emergency Management Agency*, 2016a]. Validation data are therefore available in the form of a mosaic of community-level flood hazard assessments spanning the U.S. These are carried out by FEMA to determine a SFHA in a particular locality at a standard specified by NFIP.

These data in the U.S. present an excellent opportunity to comprehensively validate a continental-scale flood model built with accurate topographic data. As flood models of this scale continue to be developed, it is crucial that their output is properly scrutinized to ensure their delineations of flood hazard are trustworthy. These models can then be utilized by a variety of end-users: from insurers adjusting their premiums, to planners selecting appropriate sites for development; all of whom will require assurances that the hazard data are accurate. A number of binary pattern measures will be used to ascertain the level of fit between the continental model under assessment and the nationwide amalgamation of high-quality local flood hazard studies carried out by FEMA.

2. Data and Methodology

2.1. Continental Model Description

The model used to produce the full-coverage flood hazard layers of the conterminous United States (CONUS) is an evolution of the global flood hazard model detailed by *Sampson et al*. [2015]. Extreme discharge estimates are generated using the regionalized flood frequency analysis of *Smith et al*. [2015], which clusters homogenous catchments based on climate zone, catchment area, and upstream annual rainfall. A flood estimation index is applied to these clusters, providing mean annual flood and growth curves to estimate return period discharges of any magnitude. This regionalization approach is critical for hydraulic models of this scale, since a great number of catchments are ungauged. This methodology essentially relates the characteristics of gauged catchments to ungauged ones and, if they are suitably similar, assumes the flood frequency response will be similar too. Channel and floodplain flow are propagated by means of a highly efficient inertial formulation of the shallow water equations in two dimensions using the algorithms developed for the LISFLOOD-FP code as a blueprint [*Bates et al*., 2010; *Neal et al*., 2012]. River channels are delineated by the HydroSHEDS global hydrography data set [*Lehner et al*., 2008], while the floodplain is represented by a digital elevation model (DEM) derived from the 1 arc sec (~30 m) USGS NED. These simulations are executed at the native DEM resolution to remove any requirement for downscaling simulated water surfaces onto a finer grid. The use of the subgrid method of channel representation [*Neal et al*., 2012] is restricted to smaller rivers, while larger rivers are "burned" directly into the DEM. The U.S. Army Corps of Engineers (USACE) National Levee Database (NLD) is incorporated into the model to explicitly represent known flood defenses. Both "defended" (with the NLD) and "undefended" (without the NLD) versions of the model are run.

Further to these fluvial model components, pluvial simulations also contribute to the final delineation of the floodplain. Flooding from rainfall directly onto the land surface can be a significant contributor to flood hazard in its own right, but the pluvial model is also required for simulating flood hazard in small headwater channels. The limited availability of observed stream gauge records for very small catchments (<50 km^2), coupled with their highly heterogenous behavior, means they cannot be adequately represented within the RFFA and are therefore not simulated by the fluvial model. Flood hazard for these catchments is instead captured by the pluvial model, as such flooding is typically flashy and driven by intense local rainfall events. The pluvial model uses rainfall scenarios derived from Intensity-Duration-Frequency (IDF) relationships

described by the National Oceanic and Atmospheric Administration (NOAA). These IDF data were pooled by Köppen-Geiger climate zone and regressed against annual average rainfall to generate extreme rainfall estimations for every cell in the DEM. Not all rainfall will flow over the surface, so allowances are made for infiltration and urban drainage. For the former, a modified Hortonian infiltration equation of *Morin and Benyamini* [1977] is applied in conjunction with the Harmonized World Soil Database (HWSD) of the Food and Agriculture Organization of the United Nations (FAO). Urban drainage is accounted for by assuming a design standard depending on the degree of urbanization, based on the luminosity data of *Elvidge et al*. [2007], and the duration and intensity of the event.

NED is a continuously updated data set utilizing the most accurate elevation data available, meaning it is an amalgamation of many data sources; predominantly LiDAR and IfSAR. Its availability at high resolution offers significant advantages over the 3″ SRTM DEM employed in the *Sampson et al*. [2015] global model which, aside from its poor accuracy in urban areas, is too coarse a resolution to accurately simulate inundation in cities [*Yu and Lane*, 2006]. Though NED is available at $1/3''$ ($\sim$10 m) resolution, 1″ resolution offers advantages in both vertical accuracy and computational expense. Halving grid resolution increases simulation time by an order of magnitude [*Savage et al*., 2016], so the 1″ data provide a more practicable DEM for continent-wide hydraulic modeling. Elevation errors are also essentially reduced by averaging when resolution is coarsened, if flat terrain and a normal distribution of errors are assumed [*Neal et al*., 2012]. Sampling error will reduce proportional to $1/\sqrt{N}$, where N is the number of cells with a combined area equivalent to that of one cell of the coarser resolution. A USGS accuracy study claims NED is not biased toward negative or positive errors [*Gesch et al*., 2014], meaning vertical error at 1″ is one third of the error at $1/3''$ on flat terrain.

The NLD provides a map of regions protected by flood defense structures. The regions are accompanied by defense design standards, and the approach adopted by the continental model is to restrict flow into these regions at return period simulations below the defense standard, while permitting flow for return period simulations that exceed the defense standard. This approach has an advantage over a simple postsimulation masking approach (whereby wet pixels within the defended areas are reset to zero depth after simulation) as it enables the hydraulic effects of defense structures, such as backwatering, to be captured by the model.

Simulation at the native DEM resolution has been enabled by further improvements to the parallel efficiency of the code by better implementation of optimizations for the Intel Broadwell architecture. This yielded significant runtime reductions over the implementation used by *Sampson et al*. [2015] and permits simulation at 1″ resolution. This increased grid resolution means large rivers are better represented by directly burning them into the DEM, while the subgrid model [*Neal et al*., 2012] is retained for smaller channels whose width is below the grid scale.

2.2. FEMA Benchmark

The benchmark data to which the model output will be compared are primarily sourced from FEMA, whose local modeling studies delineate the 1 in 100 year flood extent in a particular community. It is difficult to provide specific details on the vast assemblage of FEMA studies across the U.S., given the range of methodologies employed. The vector-based data consist of over 2,000,000 individual GIS shapefiles with limited meta-data, and so instead some common practices and minimum standards will be outlined.

Extreme flows, which drive the models that produce FEMA flood maps, are typically generated in one of three ways: flood frequency analyses, where gauges exist; regionalized regression equations, where they do not; or rainfall-runoff models [*Federal Emergency Management Agency*, 2015]. These boundary conditions are usually then routed through a 1-D or 2-D hydraulic model. FEMA stipulates which hydrologic and hydraulic models meet NFIP specifications for flood hazard mapping. The most widely used are those developed by USACE, particularly the rainfall-runoff model HEC-HMS [*United States Army Corps of Engineers*, 2016a; *Du et al*., 2012] and the hydraulic model HEC-RAS [*United States Army Corps of Engineers*, 2016b; *Icaga et al*., 2016]. The most accurate elevation data available to FEMA must always be used and have to meet certain vertical accuracy requirements [*Federal Emergency Management Agency*, 2016b]. In most cases, the topographic data will be LiDAR. Calibration of both hydrologic and hydraulic models is also mandatory if good quality data are available [*Federal Emergency Management Agency*, 2016b]. Many of these conditions, however, are policy standards specified in the last few years and so will only apply to recent and future

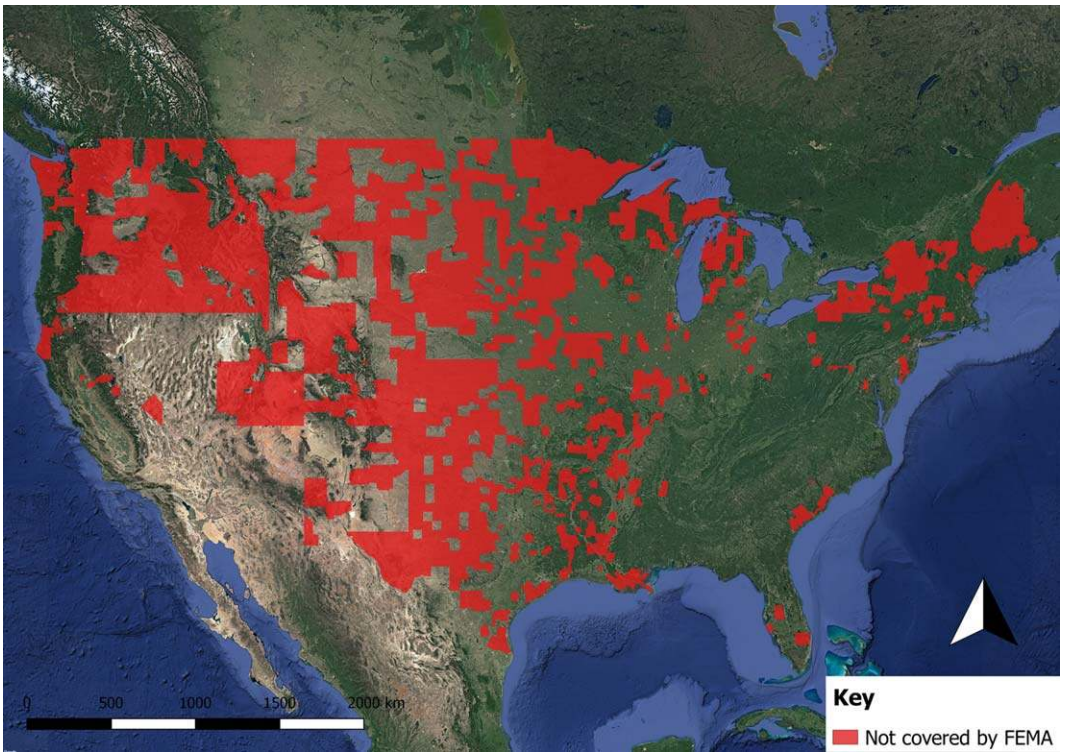


Figure 1. Map of the U.S. exhibiting areas FEMA explicitly claims not to have studied.

studies. Much of the national SFHA is classified as Zone A: approximate areas. These are areas where time and money constraints prevent detailed analyses from taking place, or more often because they are sparsely populated areas which are unlikely to be developed further in the future. In order to approximate a SFHA, FEMA employs a wide range of methods: from using Quick-2, a simplified version of HEC-RAS, to simply analyzing historical flood data (e.g., high water marks or aerial photographs of previous flood events) [*Federal Emergency Management Agency*, 1995; *National Research Council of the National Academies*, 2015].

Although much of the U.S. is mapped, the FEMA data contain both declared and undeclared no-data areas. By their own admission, FEMA has not studied the areas shown in Figure 1. These can quite easily be excluded from the validation analysis. However, even a simple examination of the FEMA data shows that some areas explicitly specified as not being within the SFHA (i.e., outside the 1 in 100 year flood extent) are clearly river valleys and floodplains. These areas are generally in smaller catchments, and while their exclusion from the SFHA may be legitimate due to the lack of development, and hence risk, occurring there, it means assessing false alarms in the continental model becomes problematic. To illustrate this point, in areas around the larger river in the south of Figure 2 the continental model exceeds the SFHA boundary and overpredicts flooding with respect to FEMA. However, these legitimate false alarms (assuming FEMA as truth for this analysis) become muddled with clearly "unmodeled in FEMA" areas, such as those smaller tributaries that branch northward. Some flooding is likely in these rivers and this is picked up by the continental model but is missed by FEMA. To combat this issue, and thus generate a better idea of performance compared to the FEMA data, the continental model output was clipped within the bounds of a $\sim$1 km buffer constructed around the SFHA. Though this will still likely capture areas FEMA has not studied (but which are still classified as outside the SFHA), a reasonable idea of the continental model's performance should be provided.

To undertake the analysis the 2,000,000 FEMA GIS shapefiles were converted to a 1 in. raster; each cell with a value representing wet, dry or no-data. Every wet cell in this raster was classified as such if a FEMA shapefile representing the SFHA covered over 50% of its area. Any shapefiles not used in this analysis were classed as no-data. Examples include areas at risk of coastal flooding, since the continental model has no coastal component, and areas of open water, since we are only interested in model performance on the

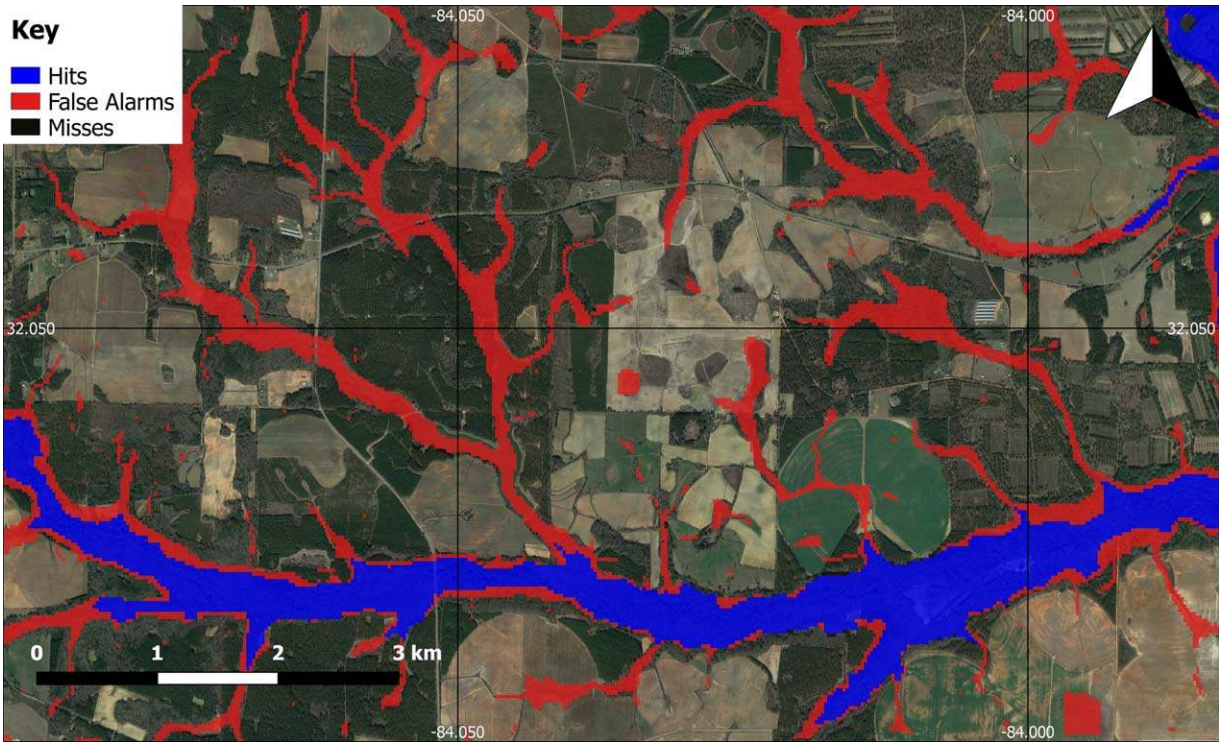


Figure 2. Intersection of the defended 1 in 100 year model at 1″ resolution with the FEMA benchmark in an area of Georgia to exhibit the varying nature of false positives. As per Table 1, "Hits" correspond to M_1B_1, "False Alarms" correspond to M_1B_0, and "Misses" correspond to M_0B_1.

floodplain. Some areas outside the SFHA were specified by FEMA as being within the 1 in 500 year flood zone, though this is not the case everywhere. A 1″ raster representing this was created, though no dry cells were specified due to the sporadic specification of a 1 in 500 year floodplain. This means that tendency of the model to overpredict a 1 in 500 year event could not be measured as only "hits" could be determined. Extra information on other areas outside the SFHA was also provided by FEMA: for example, those which were outside the 1 in 100 year floodplain as a result of levee construction. This information was also rasterized at 1″ resolution to test whether the continental model correctly identifies these areas as dry. Lastly, parts of the SFHA that are Zone A (areas where the 1 in 100 year flood was determined by approximate methodologies) were rasterized separately from Zone AE (parts of the SFHA determined by detailed methods). In doing this, model performance against high-quality data can be compared to the performance where only lower quality data are available.

2.3. USGS Benchmarks

As well as FEMA data, which represent the bulk of the validation information used here, isolated modeling studies carried out by the USGS were selected to assess model performance against high-quality benchmarks of known specification. Ten sites, with study areas usually representing just tens of kilometers of a single stream, were chosen; none of them further west than Minneapolis, MN. Nine of the sites had vector data detailing the inundation extent of a 1 in 100 year design event, three of the sites had further data on design events of varying magnitude, and one site detailed only the 1 in 500 year floodplain.

The river reaches examined by the USGS range from 6 to 40 km with upstream catchment size varying between 60 and 13,700 km^2. Eight of the studies employed the 1-D hydraulic model HEC-RAS, one used its inferior counterpart HEC-2 and the other used a 2-D model produced by USGS called FESWMS-2DH which uses a finite-element grid [*United States Geological Survey*, 2016]. All boundary conditions were derived from USGS stream gauges. All DEMs were sourced from high-resolution LiDAR data, with hydraulically important structures included from bridge plans and aerial photography. Models were run with a grid resolution between 1 and 10 m, with most run at 3 m. Half of the studies utilized bathymetry data derived from channel cross sections surveyed by USGS field teams. Most studies calibrated the energy loss coefficient

(Manning's *n*) to stage-discharge relationships derived from gauging data, high water marks from actual flood events, or FEMA flood insurance studies. The data were maintained in their original vector formats to preserve their high resolutions, and the area over which these local models were compared to the continental model was determined manually for each site. The study locations, the return periods modeled and their associated USGS reports are detailed here:

1. Albany, GA [*Musser and Dyar*, 2007]: 1 in 100.
2. Battle Creek, MI [*Hoard et al*., 2010]: 1 in 10, 1 in 50, 1 in 100, and 1 in 500.
3. Columbus, IN [*Coon*, 2013]: 1 in 100.
4. Greenville, SC [*Benedict et al*., 2013]: 1 in 100.
5. Harrisburg, PA [*Roland et al*., 2014]: 1 in 10, 1 in 50, 1 in 100, and 1 in 500.
6. Hattiesburg, MS [*Storm*, 2014]: 1 in 100.
7. Killbuck, OH [*Ostheimer*, 2013]: 1 in 5, 1 in 10, 1 in 50, 1 in 100, and 1 in 500.
8. Lincolnshire, IL [*Murphy et al*., 2012]: 1 in 500.
9. Minneapolis, MN [*Czuba et al*., 2014]: 1 in 100.
10. Ridgewood, NJ [*Watson and Niemoczynski*, 2014]: 1 in 100.

For the purposes of these analyses, the benchmark FEMA and USGS data are being treated as truth. Given the quality of the input data (especially that of the USGS), as well as the significantly greater amount of time and money expended on producing these benchmarks by U.S. government agencies in relation to that devoted to developing the continental model, it is assumed that these should more closely approximate the locally observed 1 in 100 year flood extent. It is important to note, however, that all model structures have limitations and, particularly in the case of older FEMA data, it is possible that the continental model may better approximate real behavior in certain areas.

2.4. Validation Procedure

Given the vector-based nature of both the FEMA and USGS source data, binary pattern measures are employed to enable comparison to the continental model across the CONUS. The continental model output gives the water depth for each ~30 m cell, which is then converted to one of two states: wet or dry. For the fluvial model component, cells are classified as wet where the water depth is greater than zero. This is because even a few centimeters of fluvial flooding can cause damage to basements. The pluvial model, however, has a threshold of 15 cm; in line with the way surface water masks are commonly generated [*Environment Agency*, 2013]. The primary reason is because the pluvial model produces a positive water depth for every cell, albeit mostly small ones, and so a threshold is needed. The other is that surface water flooding does not behave in the same way fluvial flooding does; in that there is not a clearly defined flood boundary as when water leaves the channel and flows over the floodplain. When the pluvial model starts to exceed water depths of 15 cm, roughly the height of a doorstep or a curb, then there can be more confidence that a significant hazard is posed.

Four basic measures of fit to the benchmark data were used, which analyze the relative number of pixels which conform to one of the states in the contingency table (Table 1).

The first of these is hit rate (*H*) which tests the proportion of wet benchmark data that was replicated by the model, ignoring whether the benchmark flood boundaries were exceeded. In its simplest sense, this measure examines the model's tendency toward underprediction of the flood hazard. *H* can range from 0 (none of the wet benchmark data are wet model data) to 1 (all of the wet benchmark data are wet model data).

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

Table 1. Contingency Table of Possible Cell Descriptors in a Binary Classification Scheme

	Wet in Benchmark Data	Dry in Benchmark Data
Wet in modeled data	M_1B_1	M_1B_0
Dry in modeled data	M_0B_1	M_0B_0

The false alarm ratio (*F*) indicates the proportion of wet modeled pixels that are not wet in the benchmark data. This metric gives an idea of whether the model has the tendency to overpredict flood extent and can range from 0 (no false alarms) to 1 (all false alarms).

$$F = \frac{M_1 B_0}{M_1 B_0 + M_1 B_1} \tag{2}$$

Third, the Critical Success Index (C) accounts for both overprediction and underprediction and can range from 0 (no match between modeled and benchmark data) to 1 (perfect match between modeled and benchmark data). C ignores the extensive areas that are dry in both the modeled and benchmark data, as these can be easily predicted by the continental model and so would bias the analysis results.

$$C = \frac{M_1 B_1}{M_1 B_1 + M_0 B_1 + M_1 B_0} \tag{3}$$

Finally, error bias (E) indicates whether the model has a tendency toward overprediction or underprediction. $E = 1$ would indicate no bias, $0 \leq E < 1$ indicates a tendency toward underprediction, and $1 < E \leq \infty$ indicates a tendency toward overprediction.

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

These metrics were applied in a number of different scenarios, which are broadly described as follows:

1. *Nationwide*: all performance metrics within the buffer surrounding the SFHA.
2. *Climate*: performance analyzed in the three main climate zones in CONUS.
3. *Quality*: performance where FEMA data are high quality (Zone AE) versus that where it is lower (Zone A).
4. *Defense*: testing whether the continental model correctly identifies defended areas (as specified by FEMA) as dry.
5. *Size of catchment upstream*: analysis of whether the model performs better for rivers with larger or smaller upstream catchment areas.
6. *Land use*: performance disaggregated between developed areas, forested areas, and areas that are neither of these.
7. *USGS*: all performance metrics applied to the ten USGS study sites.

The default continental model output used in the analysis was the 1 in 100 year 1″ hazard layer which incorporates flood defense data. Some of the metrics and scenarios are also applied to the 3″ global model of *Sampson et al*. [2015] which utilizes a SRTM-derived DEM, as well as 1 in 500 or undefended versions of the 1″ hazard layers.

Additionally, an aggregate measure of similarity to the FEMA data was computed. A pixel-to-pixel comparison is a reasonably tough test for a hydraulic model of this scale. It is perhaps more useful to know that the model is getting broadly the correct answer at a scale at which most end-users would utilize the data. In data-poor regions, for instance, where the large-scale model will be most serviceable, uncertainty over the location of a site of interest may be considerable. The performance of the model at ~30 m resolution is therefore not so relevant in this instance, since the site of interest may not be known to this level of accuracy. Instead, an aggregate performance metric may be more pertinent. Data from both the default model hazard layer and FEMA were resampled to 1 km resolution and each 1 km^2 pixel took a value between 0 and 1 to represent the proportion of its area that is covered by the 1 in 100 year event. The modulus of the differences between the model (M) and the FEMA benchmark (B) was then averaged to produce the mean absolute error (E_A). This was calculated within the bounds of the ~1 km buffer constructed around the SFHA.

$$E_A = \frac{\sum_1^N |M - B|}{N} \tag{5}$$

The aggregate error bias (B_A) was calculated, where the differences between the two data sets were of their original sign.

$$B_A = \frac{\sum_1^N M - B}{N} \tag{6}$$

If $B_A > 0$, it is an indication that the model has a tendency toward overprediction, while $B_A < 0$ indicates underprediction.

The analyses detailed in this study were performed in Google Earth Engine [*Google*, 2016], a cloud-based geoprocessing application that permits rapid spatial analysis on a global scale. This platform enabled

validation of the continental model with unprecedented efficiency. It has been employed in a number of recent studies involving large-scale analysis of the Earth's surface, particularly relating to surface water [*Donchyts et al*., 2016; *Pekel et al*., 2016] and land cover [*Cohen et al*., 2017; *Dong et al*., 2016].

3. Results

3.1. Nationwide

The ~30 m flood model accounting for USACE levee data mapped the 1 in 100 year flood extent across CONUS. Analyzing nearly 800,000,000 pixels, the nationwide results are shown in Table 2. The *H* score of 0.815 indicates that over 80% of the SFHA specified by FEMA is captured by the model. The *C* score drops relative to *H* as a result of model overprediction with respect to the FEMA data; the extent of which is highlighted by *F* and *E* scores. The *F* score essentially means for approximately every three pixels identified correctly as wet, one pixel will be incorrectly identified as such. Figure 3a exemplifies an area of good continental model performance. This area where the Illinois and Mississippi Rivers meet sees much agreement between FEMA and the continental model, with very few areas of overprediction or underprediction. Figure 3b illustrates where continental model performance is much poorer; there is a great deal of overprediction on the Rillito and Santa Cruz Rivers in Tucson, AZ.

Possible explanations for the differences between the continental model and FEMA data are numerous, with failure of the buffer to filter out areas unmodeled by FEMA likely bearing the most responsibility. From Figure 3d, it is evident that the arbitrary ~1 km buffer still picks up some of the areas it was designed to exclude. The overprediction exhibited further north than 32.50°N is not genuine: the model has not simply overshot the flood extent specified by FEMA but has rightly captured flood hazard in the small river valleys. However, Figure 3c shows where this ~1 km buffer is prohibitively small: continental model overprediction has actually been constrained. Thus, the buffer appears to be an imperfect solution to a complex issue. One must therefore interpret the metrics accounting for overprediction with a degree of caution. Areas where the model has genuinely exceeded the 1 in 100 year flood extent specified by FEMA, such as those in Figure 3b, could perhaps be explained by its coarser resolution. Many flow restricting structures may not be resolved by the continental model, where a localized FEMA study may have accounted for these. Examples of such a phenomenon could be unincorporated 1 in 100 year levees arising from their absence from the NLD, as well as lower profile berms, bridges and roads. A comprehensive evaluation of the completeness of the NLD has not taken place, but estimates suggest it contains only ~30% of the nation's levees. According to a report by the *American Society of Civil Engineers* [2017], the NLD contains roughly 30,000 out of an estimated 100,000 miles of U.S. levees. This has severe consequences for the delineation of the continental-scale modeled floodplain; the most obvious of which is the accumulation of false alarms.

Comparison to the test scores for the 3″ SRTM-based model [*Sampson et al*., 2015] demonstrates that the higher resolution NED-based model captures much more of the SFHA (*H* score differential of 0.130), though has a slightly increased tendency toward overprediction (*F* score differential of 0.024). In contrast to the 1″ model, the 3″ model does not exhibit much bias with an *E* score very close to 1. This is simply because its tendency to underpredict is much greater relative to the finer resolution model.

3.2. Climate

The contrast between Figures 3a and 3b shows that the national outlook does not tell the whole story. Thus, these nationwide results have been spatially disaggregated with four themes in mind: regional climate, quality of FEMA data, size of catchment upstream, and land use classification. The first of these, climate, involved analyzing model performance in each of the three main Köppen-Geiger climate zones within CONUS: temperate, continental, and arid [*Kottek et al*., 2006]. The results in each of these zones are listed in Table 3. Performance in temperate regions, which cover roughly two-thirds of the pixels analyzed, is better than the overall average. *C* scores, even given much uncertainty over the number of genuine false alarms, far exceed those achieved by the 3″ model of

Table 2. Validation Metrics for the Defended, 1 in 100 Year Model Across CONUS at 1 and 3″ Resolution Against FEMA Special Flood Hazard Area Data

Area of Study	Hit Rate (*H*)	False Alarm Ratio (*F*)	Critical Success Index (*C*)	Error Bias (*E*)
Nationwide (1″)	0.815	0.368	0.552	2.565
Nationwide (3″)	0.685	0.344	0.504	1.142

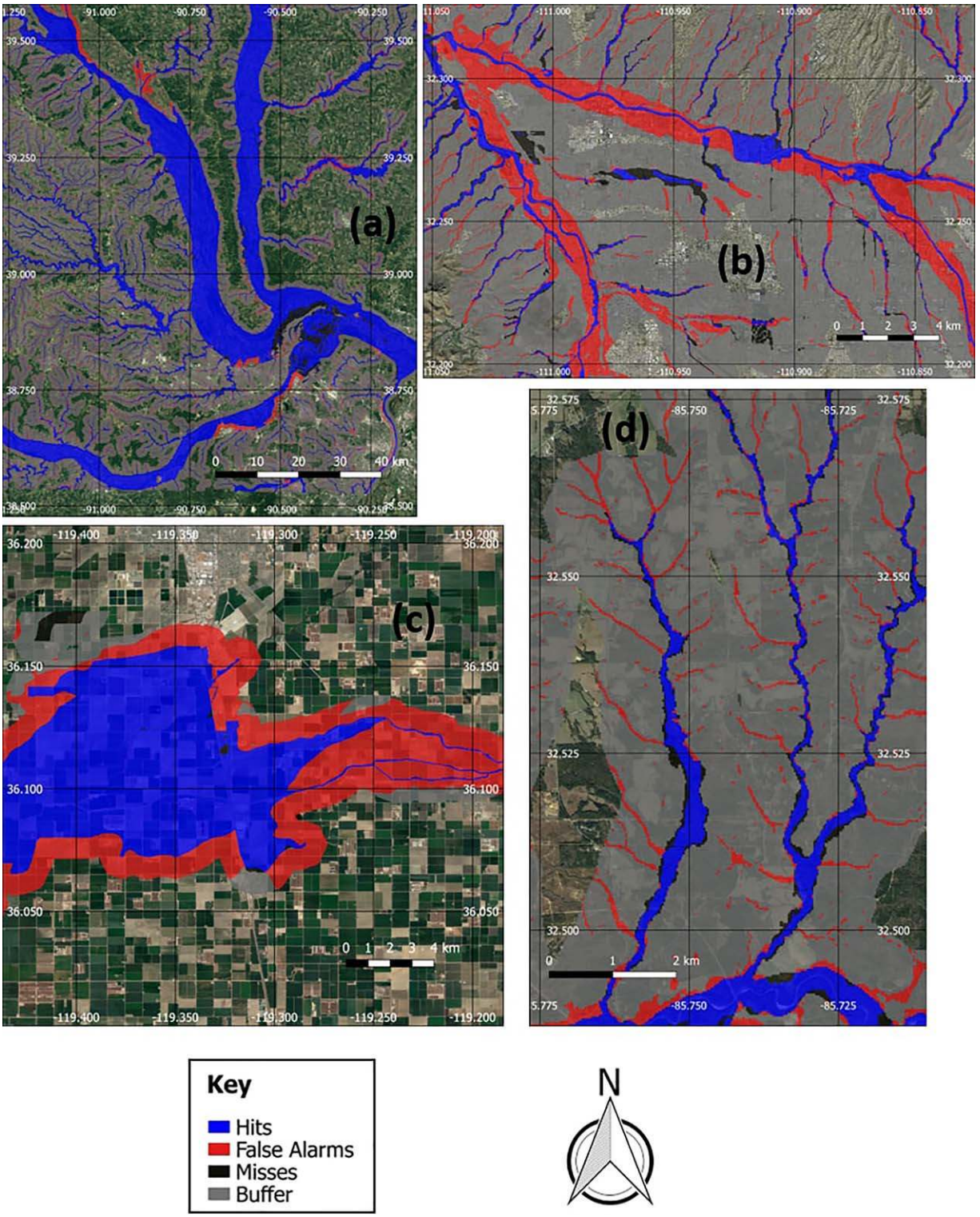


Figure 3. Intersection of the defended 1 in 100 year model at 1″ resolution with the FEMA benchmark in (a) an area of Missouri and Illinois near St. Louis; (b) Tucson, Arizona; (c) an area of California between Fresno and Bakersfield; and (d) an area of Alabama between Montgomery and Columbus.

Sampson et al. [2015] when tested on the Thames and Severn catchments in the United Kingdom (*C* score differentials of ∼0.22). Continental climate zones, covering just over one fifth of the area studied, experience a dip in performance compared to the nationwide average; *F* scores are particularly high. Poorer

Table 3. Validation Metrics Against FEMA Data for the Defended, 1 in 100 Year Model in Three Climate Zones Within CONUS at 1 and 3″ Resolution

Climate Zone	Hit Rate (*H*)	False Alarm Ratio (*F*)	Critical Success Index (*C*)	Error Bias (*E*)
Temperate (1″)	0.841	0.332	0.593	2.626
Continental (1″)	0.779	0.451	0.475	2.901
Arid (1″)	0.727	0.434	0.466	2.039
Temperate (3″)	0.705	0.315	0.533	1.099
Continental (3″)	0.686	0.398	0.472	1.446
Arid (3″)	0.571	0.430	0.399	1.007

performance in continental compared to temperate regions is consistent with the conclusions of a meta-analysis enquiry by *Salinas et al*. [2013], who summarized the findings of numerous studies into flow and flood prediction in ungauged basins using RFFA. A likely culprit for poorer performance is the varying nature of water storage types creating a more complex hydrology in these colder climates. Precipitation falling as snow or water being stored as ice means factors additional to total precipitation are likely to control extreme flows (e.g., temperature dictating snow and ice thaw). Arid climate zones, which only make up just over 10% of the total area analyzed, stand out as areas where the model performs much worse than the national average (*H* score differential of 0.09). This is, again, consistent with RFFA studies from the wider literature. The RFFA methodology of *Smith et al*. [2015], which is employed by this model, produced larger errors in replicating the 1 in 100 year discharge of arid catchments. Both *Salinas et al*. [2013] and *Smith et al*. [2015] believe this is due to the heterogeneity of dryer regions. With that being said, capturing almost three quarters of the SFHA in arid regions still represents good model performance. For the 3″ model, performance in continental climates is higher relative to its national average than the 1″ one. Continental false alarms in the 3″ model are so much lower than those in the 1″ version that *C* scores are virtually the same despite the *H* score differential of 0.09. It is notable that 1″ model arid zone *H* scores are higher than 3″ model temperate ones.

3.3. Quality

Some of the benchmark data have been specified by FEMA as being generated through detailed methods, while the bulk of it has been determined through approximate methodologies. The disaggregation of performance across CONUS between these two data categories is shown in Table 4. When validated against high-quality data, performance markedly improves compared to the national average; hit rates are up and false alarms are down. *E* scores are increased only because misses reduced at a greater rate than false alarms. This means the studies in which FEMA has devoted most of its flood modeling efforts, in both a temporal and monetary sense, more closely resemble the continental model than the approximate studies. Put simply: where FEMA is more confident in its work, the continental model agrees with them more. In lower quality areas, which cover triple the area of higher quality ones, the model deviates more from the delineation of the SFHA. Hit rates against high-quality data in the 1″ model are over 10 percentage points higher than in the 3″ model.

3.4. Size of Catchment Upstream

Continental model performance has also been split depending on the size of the river responsible for the hazard. Streams were partitioned, using their respective upstream area, into eight groups. Buffers of varying size were constructed around the rivers, depending on their grouping, to delineate the floodplains they are likely responsible for flooding. In areas of overlap, the buffer of the river with the larger upstream catchment area took precedence. The categories are detailed alongside their results in Table 5. The key theme is that performance is notably higher in larger catchment categories, with areas around rivers with upstream catchment areas greater than 8000 km^2 enjoying hit rates of almost 90% and *F* scores around half those of the national average. *C* scores for these areas are approaching those found in validation studies of good local models with real event data [*Wood et al*., 2016]. Moderately sized river reaches with upstream catchments between 80 and 8000 km^2 have slightly lower *H* scores of around 0.85 with false alarm ratios creeping upward with reducing upstream catchment size. The marked increase in *F* scores from rivers with an upstream catchment area between 400 and 800 km^2 compared to those with an upstream

Table 4. Validation Metrics for the Defended, 1 in 100 Year Model at 1 and 3″ Resolution When Compared Against High-Quality and Low-Quality FEMA Data

Quality of FEMA Benchmark Data	Hit Rate (*H*)	False Alarm Ratio (*F*)	Critical Success Index (*C*)	Error Bias (*E*)
High Quality (1″)	0.862	0.343	0.594	3.252
Low Quality (1″)	0.778	0.396	0.515	2.294
High Quality (3″)	0.752	0.302	0.567	1.137
Low Quality (3″)	0.657	0.390	0.463	1.224

Table 5. River Size Categories, Their Associated Descriptors, and Validation Metrics Against FEMA Data for Each Using the Defended, 1 in 100 Year Model at 1″ Resolution

Approximate Area of Catchment Upstream (km^2)	Size of Buffer (km)	% of Total Pixels Analyzed	Hit Rate (*H*)	False Alarm Ratio (*F*)	Critical Success Index (*C*)	Error Bias (*E*)
400,000+	10	4.6	0.899	0.233	0.707	2.698
80,000–400,000	5	4.5	0.878	0.260	0.671	2.525
40,000–80,000	2	2.5	0.892	0.188	0.739	1.904
8,000–40,000	2	8.1	0.894	0.204	0.727	2.152
800–8,000	1	16.0	0.856	0.238	0.675	1.853
400–800	0.5	5.6	0.845	0.255	0.655	1.864
80–400	0.5	17.3	0.841	0.336	0.590	2.671
0.8–80	0.25	41.4	0.759	0.480	0.446	2.906

catchment area between 80 and 400 km^2 lacks a coincident reduction in *H* score. This is likely explained by the latter category containing some of the illegitimate false positives derived from FEMA's failure to specify certain headwater areas that they have not studied. Even with some of such areas missing, headwater areas (rivers with an upstream catchment area of between 0.8 and 80 km^2) still make up the bulk of this analysis and performance is markedly poorer here. *F* scores greatly increase almost to the extent that half of the modeled wet pixels are falsely identified as such. It should be borne in mind, however, that many of these false alarms are not genuine due to the lack of complete headwater coverage by FEMA. The substantial drop in *H* score is not excused by this, however. The summary of RFFA studies by *Salinas et al*. [2013] noted that errors in the generation of 1 in 100 year discharges increased with decreasing catchment size. *Sampson et al*. [2015] found that *F* scores were greatly reduced and *C* scores dramatically increased when areas of the Severn and Thames catchments with upstream areas less than 500 km^2 were excluded from their analysis of the 1 in 100 year flood extent. A likely reason for such trends is that more data are available for larger catchments, as there is a greater chance that a flow gauge exists as stream order increases. This means the frequency curve used to generate the 1 in 100 year flow in large catchments will be derived from a greater number of gauges than flows in smaller catchments. Also, the processes that generate floods on larger catchments experience aggregation effects, which results in a tendency for the floods to not be so flashy and therefore more predictable [*Salinas et al*., 2013].

3.5. Land Use

Since people and assets are not distributed uniformly across the study area, it is necessary to analyze continental model performance in areas where the presence of a hazard translates into high risk separately to areas where it does not. To achieve this, the National Land Cover Database (NLCD) is used to disaggregate performance based on land use classification [*Homer et al*., 2015]. The results are displayed in Table 6, where it is evident that the level of fit between the continental model and FEMA data is lower in more developed areas. *Gesch et al*. [2014] carried out an accuracy assessment of NED and also provided an absolute vertical root-mean-square error (RMSE) for each NLCD class. Since the DEM of the continental model uses NED aggregated to 1″, the errors listed in Table 6 are scaled accordingly. It is evident that the level of fit between the two data sets improves with increasing NED accuracy. Vertical accuracy does not tell the whole story however, as forested areas and medium intensity developments have similar RMSEs but very different *H* and *F* scores. This highlights the difficulty of hydraulic modeling in urban areas, consequently requiring further scrutiny of the validity of the FEMA benchmark.

Urban hydraulic modeling has historically been challenging, owing to complex flow paths that require the representation of micro-scale features such as curbs and walls [*Hunter et al*., 2008]. The horizontal, rather than vertical, accuracy of the continental model inhibits the resolution of such features: ∼30 m pixels are not fine enough to capture the elevation difference between a building and a road, for instance [*Yu and Lane*, 2006]. Rather than smoothing over features of developed areas, which would result in all urban areas appearing as hills in the DEM, the lowest values in the LiDAR point cloud are used to construct a "bare earth" DEM with buildings stripped away. As such, the continental model ignores potentially critical surface objects in its determination of flow paths, yet presently there is no alternative to explicitly represent these at this resolution and spatial coverage. A growing body of research into porosity-based models, however, may provide a future solution through the parameterization of sub-grid-scale features, such as buildings, as

Table 6. Validation Metrics Against FEMA Data for the Defended, 1 in 100 Year Model at 1″ Resolution Within Land Use Descriptors

NLCD Class	% of Total Pixels Analyzed	NED Absolute Vertical RMSE Estimate (m)	Hit Rate (*H*)	False Alarm Ratio (*F*)	Critical Success Index (*C*)	Error Bias (*E*)
Developed: high intensity	0.35	0.817	0.642	0.743	0.225	5.179
Developed: medium intensity	0.95	0.627	0.660	0.722	0.243	5.035
Developed: low intensity	1.93	0.500	0.707	0.621	0.327	3.954
Developed: open space	4.32	0.440	0.752	0.522	0.413	3.307
Forested	18.39	0.624	0.810	0.465	0.475	3.701
Not developed or forested	74.06	0.486	0.822	0.322	0.591	2.191

irregularly distributed, flow-obstructing objects if appropriate national scale parameterization data can be obtained [*Sanders et al*., 2008; *Dottori and Todini*, 2012; *Kim et al*., 2015; *Guinot and Delenne*, 2014]. The FEMA data are likely derived from a "bare earth" DEM also, though it is difficult to confirm this from the varied methodologies employed and the lack of meta-data. In most instances, the distinction is irrelevant, since 1-D HEC-RAS models do not account for the hydraulic significance of structures on the floodplain either. In these circumstances, FEMA will extrapolate a channel water surface elevation from a discharge and assume, in areas of relatively simple topography, that the water surface elevation on the floodplain is largely the same as that in the channel [*National Research Council of the National Academies*, 2015]. All areas at elevations at or below this water surface will be classified as within the SFHA, based on high-resolution terrain data.

As such, the determination of an urban flood hazard is open to much more interpretation than elsewhere, meaning the range of possible flood extents that different methodologies provide will be much broader. It is therefore not surprising that the continental model and FEMA deviate more significantly from one another in areas of development, despite such areas having comparable vertical DEM accuracies to forests. Instead of treating FEMA as a benchmark for continental model performance in these instances, it is more useful to elaborate on why both of them will be subject to errors. High *F* scores in more developed areas are perhaps explained by the incompleteness of the USACE National Levee Database, which will result in areas that are defended in reality being flooded in the continental model. FEMA will have accounted for these defenses in ground-based field surveys, whereas the lower resolution continental model is unlikely to have captured the full effect of a levee unless it is specified explicitly in the NLD or its form is represented in the NED. The inclusion of a surface water hazard in the continental model will also incur many false positives with respect to the FEMA data, which ignores pluvial events. In many instances, FEMA will have accounted for hydraulically significant structures in the channel from aerial or ground surveys and incorporated these explicitly into their models. These include bridges, floodways and dams, which will alter the modeled surface water elevation and, by consequence, the extrapolation of it onto the floodplain. Since the continental model will not have accounted for such structures, its floodplain delineation may be different. With that being said, the continental model captures roughly two thirds of the FEMA delineated 1 in 100 year flood extent in urban areas and around three quarters of it in more rural developments; there is therefore a reasonable level of agreement between the two data sets.

Forested areas will have been subject to the same stripping of trees as urban areas are with buildings to produce the "bare earth" DEM. FEMA and the continental model agree much more broadly on the 1 in 100 year flood extent here than they do in developed areas, perhaps indicating that these areas are less hydrologically complex. Unsurprisingly, in undeveloped areas where the range of likely solutions provided by the data is narrower, the models are very similar. *F* scores are mostly explained by the incomplete coverage of headwater areas by FEMA. It is therefore evident that the national outlook of continental model performance is skewed by high levels of agreement in the low-risk areas that occupy over 90% of the study area. Where model performance matters most, there is an implication that performance is poorer. In some instances, this may be because FEMA models will have often incorporated critical local-scale information, such as flood defenses. However, much of the divergence is likely derived from both sets of data

Table 7. Rate at Which Different Model Versions Correctly Identify Defended Areas as Dry

Model Version	Hit Rate
Defended (1″)	0.673
Undefended (1″)	0.296
Defended (3″)	0.369
Undefended (3″)	0.367

Table 8. Hit Rate Against FEMA Data of the Defended, 1 in 500 Year Model in Different Study Areas at 1 and 3″ Resolution

Area of Study	Hit Rate (*H*)
Nationwide (1″)	0.862
Temperate (1″)	0.900
Continental (1″)	0.846
Arid (1″)	0.697
Nationwide (3″)	0.741
Temperate (3″)	0.770
Continental (3″)	0.762
Arid (3″)	0.571

providing different answers to a very complex question. It would be unfair to heavily criticize the continental model in light of this, since there is no evidence that FEMA is any closer to the "truth" in these areas than the model being tested. Real event data are required to comprehensively scrutinize the continental model in developed areas.

3.6. Defenses

Table 7 outlines performance of the different versions of the model in areas defended by a levee. Hit rates here represent the proportion of total cells correctly identified as dry. It is unsurprising that the explicit inclusion of U.S. levee data in the model results in higher hit rates in defended areas. The continental model including defenses classifies just under one third of the total defended area across CONUS as being within the 1 in 100 year flood extent, compared to the undefended model identifying over two thirds of such areas. The incorrect identification of one third of defended areas (by the defended model) is likely due to the incompleteness of the NLD provided by USACE, but also because FEMA does not account for pluvial hazard.

The defended 3″ model, among its other differences to the 1″ version, does not explicitly represent levees. Instead, defenses are parameterized through the adjustment of channel conveyance based on socioeconomic factors and degree of urbanization, which are assumed to be reasonable predictors of level of defense standard. The results in Table 7 show that this methodology has a negligible effect on hit rates in defended areas. 0.2% more of the defended area is correctly identified as dry in the defended versus the undefended 3″ model, both of which perform fairly poorly in mislabeling just under two thirds of such areas as wet.

3.7. 1 in 500 Year Floodplain

The next set of validation tests against FEMA data concern the 1 in 500 year flood event, the results of which are shown in in Table 8. Only hit rates are calculated here, since the 1 in 500 year floodplain is only specified sporadically across CONUS by FEMA. A nationwide hit rate of 86% is very high, though perhaps unsurprising for an event of this magnitude since in many cases the flood will be constrained by valley sides, making it easier to predict. The relationship of performance in temperate and continental regions to the national average takes much the same form as in the 1 in 100 year analysis, but the *H* score in arid climate zones deviates from the national one even more dramatically. A *H* score differential of 0.165 between arid zones and the national average is almost double that of the 1 in 100 year equivalent. Poorer performance in arid areas at higher return periods is perhaps explained by the high extreme flood variability in such regions, which is well documented in the literature. The RFFA by *Smith et al*. [2015] saw streams in arid regions having more variable discharge than wetter regions at higher return periods. Crucially, *Merz and Blöschl* [2009] point out that runoff responses in arid catchments are more temporally variable than in wetter ones. To produce the discharge of a certain return period therefore, the RFFA has to contend with spatial (between-catchment) and temporal (within-catchment) variability in arid catchments. This is reflected in the poorer-than-average model performance for the 1 in 100 year event and the even worse performance for the 1 in 500 year event in arid zones. The picture is much the same for the 3″ model, with the 1 inch model strongly outperforming it as usual.

Table 9. Validation Metrics for the Defended, 1 in 100 Year Model at 1″ Resolution When Compared Against USGS Benchmark Data

1 in 100 Year Flood Location	Hit Rate (*H*)	False Alarm Ratio (*F*)	Critical Success Index (*C*)	Error Bias (*E*)
Albany, GA	0.938	0.195	0.764	3.656
Battle Creek, MI	0.989	0.486	0.511	88.073
Columbus, IN	0.833	0.018	0.821	0.092
Greenville, SC	0.997	0.295	0.704	128.250
Harrisburg, PA	0.881	0.093	0.809	0.762
Hattiesburg, MS	0.937	0.039	0.903	0.605
Killbuck, OH	0.896	0.007	0.890	0.017
Minneapolis, MN	0.910	0.310	0.646	4.547
Ridgewood, NJ	0.886	0.069	0.831	0.578

3.8. Aggregate

The final comparison of the model output to FEMA data involves aggregating the analysis to 1 km^2 pixels. This took place within the ~1 km buffer around the SFHA, meaning any aggregate cell that included an area outside this was ignored. E_A of the defended, 1 in 100 year model originally at 1″ resolution was 0.098. This can be interpreted as a ~10% difference, on average, in flooded

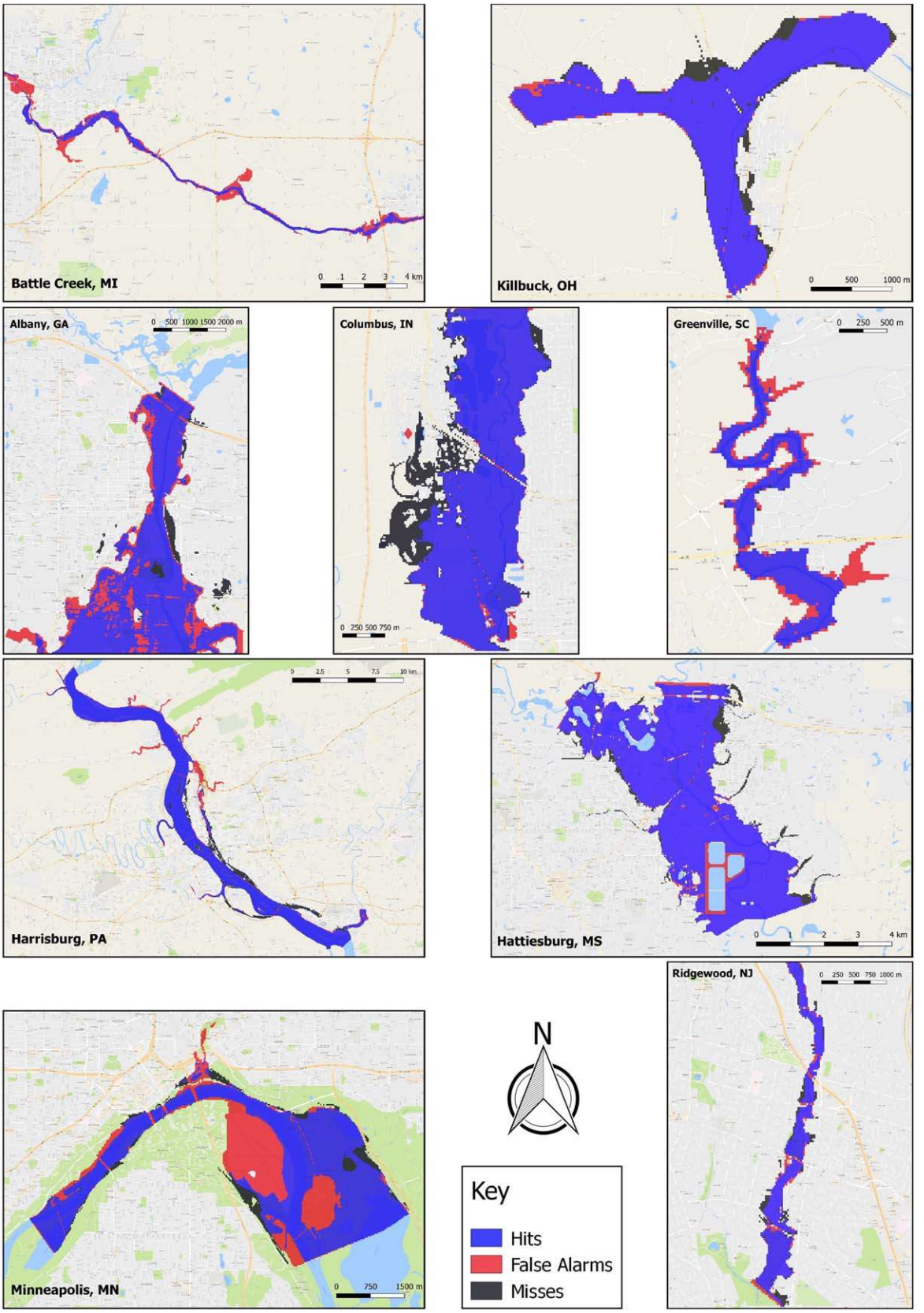


Figure 4. Intersection of the defended, 1 in 100 year model at 1″ resolution with the USGS benchmark data at the nine sites where such an event was modeled.

Table 10. Validation Metrics for the Defended, 1″ Resolution Model at Different Return Periods When Compared Against USGS Benchmark Data

Return Period	Study Site	Hit Rate (*H*)	False Alarm Ratio (*F*)	Critical Success Index (*C*)	Error Bias (*E*)
1 in 5	Killbuck, OH	0.867	0.013	0.857	0.089
1 in 10	Battle Creek, MI	0.915	0.458	0.516	9.026
	Harrisburg, PA	0.958	0.123	0.844	3.230
	Killbuck, OH	0.903	0.009	0.896	0.088
1 in 50	Battle Creek, MI	0.986	0.453	0.543	59.725
	Harrisburg, PA	0.931	0.114	0.832	1.729
	Killbuck, OH	0.902	0.009	0.894	0.081
1 in 500	Battle Creek, MI	0.993	0.500	0.498	138.525
	Harrisburg, PA	0.839	0.077	0.785	0.435
	Killbuck, OH	0.880	0.002	0.878	0.017
	Lincolnshire, IL	0.536	0.014	0.531	0.017

fraction at 1 km resolution between FEMA and the continental model. This value is toward the higher end of aggregate errors found by *Sampson et al.* [2015] on the Thames and Severn catchments, despite their coarser resolution model built with poorer topography data. It is likely that the mismatch in coverage of headwater areas holds much of the responsibility for this. Aggregate error bias (B_A) is 0.060, which supports this assertion, though it still reflects the tendency of the continental model to overpredict relative to FEMA.

3.9. USGS

The isolated, local, high-quality flood hazard studies of the USGS provide excellent validation data for the continental model, albeit not on the grand spatial scale of the FEMA benchmark. The results of validating against the nine sites which modeled the 1 in 100 year event are shown in Table 9, and are graphically represented in Figure 4. *H* scores indicate very good model performance at all sites, with the model at Greenville, SC capturing almost all of the 1 in 100 year flood extent defined by USGS. Underprediction is not prevalent at any of the sites, though overprediction is an issue for a few: notably, Battle Creek, MI, Greenville, SC, and Minneapolis, MN. It is evident from Figure 4, however, that false alarms are often generated from failure to isolate the hazard derived from the specific river modeled by USGS. For instance, the overprediction at Greenville, SC, is mainly at confluences between the Saluda River and its tributaries. This is because flood hazard derived from these tributaries has not been excluded from that caused by the Saluda River in the continental model, but has been in the USGS model. The case is the same for certain instances of overprediction in Battle Creek, MI, and Harrisburg, PA. *C* scores for sites unaffected by high false alarm ratios are comparable to optima when high-quality flood models are calibrated to real event data [*Bates et al.*, 2006].

Analysis of data on further return periods is listed in Table 10. The trends are largely the same as for the 1 in 100 year event validations. Overprediction is clearly an issue for all return periods at Battle Creek, MI, while the only site where the continental model underpredicts significantly relative to the USGS data is Lincolnshire, IL. Interestingly, when the FEMA-derived 1 in 100 year layer was compared to that of the USGS at Battle Creek, an *F* score of 0.429 was calculated; similar to that of the continental model. The USGS incorporated dams present on the Kalamazoo river into their model [*Hoard et al.*, 2010], while the continental and FEMA models were unlikely to represent these correctly. Generally, performance of the continental model at all return periods when validated against USGS data is very good. In all cases, the 1″ model outperformed the 3″ version: both *H* and *C* score average differentials between the two versions were roughly 0.17.

4. Conclusions

The results of this study can be viewed as a guide for future foci of large-scale, high-resolution flood model development, since many of the broader themes are unlikely to be specific to the particular model used here. Other features of this analysis may point toward areas where the continental model may be improved. More generally, these results are a vindication of the flood model tested. Large-scale flood models to date have not been of high enough quality to supersede detailed local studies where data are available, but the model employed here is getting close to such a position.

Height Above Nearest Drainage (HAND) methodologies have emerged as a possible alternative in providing flood hazard information over large scales [*Rennó et al.*, 2008; *Nobre et al.*, 2011, 2016]. HAND analyses use a raster grid, based on a DEM, with values containing the relative height of a particular cell from the nearest river channel. Simulated water depths therefore inundate cells within a catchment that have a value smaller than this depth. The low complexity of HAND provides a computationally inexpensive way of providing

real-time flood forecasts when coupled to a hydrological model but comes with the cost of having no physical representation of flood flow. The primary sources of uncertainty in this study pertain to boundary condition generation and errors in the topography data. These would still be present in a HAND analysis and would further contend with inaccuracies arising from the lack of a hydrodynamic element. As such techniques become more popular and refined, it is crucial that they, too, undergo a similar level of scrutiny to the flood hazard model in this study.

The benchmark data provided by FEMA were a mosaic of local studies at continental-scale. Although there were widespread issues relating to the identification of false alarms, the benchmark provided excellent validation data for assessing how capable the model is of identifying the 1 in 100 year flood hazard over the entire CONUS. The model captured 82% of FEMA's delineation of the 1 in 100 year flood, rising to 86% both where the SFHA was derived from high-quality data and where the 1 in 500 year flood was specified. This is indicative of very good model performance, particularly given the FEMA data will itself contain errors and is not "truth." A handful of USGS studies of single river reaches provided very high-quality hazard data for model validation, but at nothing close to the spatial scale of the FEMA benchmark. *H* and *C* scores here are generally in the 0.9 and 0.8s, respectively across multiple return periods; results which are unprecedented for models of this scale to the best of the authors' knowledge. Unlike FEMA, the continental model covers the entire CONUS (see Figure 1) as well as smaller watersheds not included in FEMA models (see Figure 2). Additionally, the FEMA models have taken thousands of individual studies and many decades to develop, while the continental model was built over a period of several months only from freely available data. A report by the *Association of State Floodplain Managers* [2013] claims FEMA spent between $4.5 and $7.5 billion on flood mapping up to 2013, and it will cost between $116 and $275 million per year to maintain the existing spatial coverage (i.e., prevent "decay" of the current flood maps). The continental model takes approximately 5000 h to simulate a single return period for all event types (fluvial defended, fluvial undefended, and pluvial) on a single server node with 20 Intel Broadwell E5 Xeon cores; in practice, the runtime is shorter as the compute load is distributed over multiple nodes on a HPC cluster where runtime scales linearly with the number of nodes. It would therefore be relatively straightforward to rerun the continental model: either to update it with new data or to implement different scenarios, such as climate change analysis. The former, as mentioned, is proving to be very costly for FEMA, while the latter would be prohibitively difficult for them to achieve.

The 3″ model can be replicated across the globe, but is inferior to its US-exclusive 1″ counterpart that incorporates levee and NED terrain data. Performance in all scenarios is significantly higher for the latter. It is likely that the solution to the shallow water equations at ∼30 m resolution produces a better answer than at ∼1 km (the resolution of the 3″ model before downscaling), though the greater vertical and horizontal accuracy of NED compared to SRTM is probably the primary reason for the performance discrepancy between the model versions. Even from studies carried out over a decade ago, it is recognized that the quality of the topographic data is the dominant control on flood model performance [*Horritt and Bates*, 2002]. To replicate the high performance of the 1″ model across the world therefore, high-quality topographic data must be obtained. The model tests in defended areas also clearly show the necessity for the explicit representation of defenses. Again, such data are not available across the globe but are required for a global flood model to produce hazard data of the accuracy displayed by the continental model. Nevertheless, where no better terrain data are available it is clear from the benchmarking of the SRTM-based 3″ global model against the 1″ NED-based US-only model described here that the global model does have useful skill.

Some of the other test scenarios permit identification of areas where the continental model is particularly good or particularly bad. This means areas of poor performance can be the focus of future work in improving the model. With performance disaggregated based on upstream catchment size, rivers with an upstream catchment area between 0.8 and 80 km^2 were overpredicted and underpredicted at a much greater rate than the other categories. It is therefore evident that source areas should be targeted for improvement in the model. Headwater flood hazard is primarily simulated in the pluvial model (fluvial flooding is only simulated in catchments above 50 km^2), since the RFFA is particularly poor for such small rivers owing both to the lack of data and to their heterogeneity. Some of the overprediction around such rivers is likely accounted for by FEMA's failure to specify which headwaters are not modeled, but also because FEMA is unlikely to have included surface water hazard in their studies. The pluvial model

principally simulates overland flow directly from heavy rainfall, and if FEMA has not represented this then false alarms will be incurred in the validation procedure. The underprediction, however, is not excused by these, and so the pluvial model needs refining to better represent flooding in these headwater zones. Performance in catchments above 80 km^2 is significantly better, with *H* scores almost touching 0.9 and Critical Success Indices approaching those found when local studies are validated against high-quality real event data.

Performance in arid climate zones is largely as expected based on previous RFFA studies [*Salinas et al*., 2013; *Smith et al*., 2015]. Though it is clear that such areas require improvement, this is unlikely to be achieved because of fundamental limitations in the core methodology. More gauging data to fuel RFFA in these regions would help, but the spatial and temporal heterogeneity of these regions perhaps renders them unsuitable for such methods. A 73% hit rate, however, is still indicative of good performance. The gulf in arid region performance between the 1 and 3″ models shows that huge improvements can be achieved using higher quality topography data at a finer resolution, so a change of methodology is not wholly justified. Performance is better in temperate regions and is more than satisfactory in continental climate zones.

Notwithstanding the recommendations given here, it is important to stress again that these benchmarks are subject to error and that care must be taken to not wholly base future improvements on such data. On top of the directions for future work advocated for the flood models themselves, further validation studies should also be carried out at similar scales where appropriate data are available (e.g., continental Europe), and also against real event data at varying return periods across the globe. Such studies will be able to verify the conclusions drawn here.

The wider implication of this study for large-scale flood hazard modeling is a demonstration that this field of enquiry is worthwhile. Performance of the model is approaching that of good quality local analyses; providing end-users with faith in the output, but more cheaply, easily and quickly than alternatives of commensurate caliber. Examples of future studies that this work makes possible include intersecting the hazard layer with a land use map to get an impression of the assets that are exposed to a certain flood, or applying a depth-damage function to generate a flood risk map [*de Moel et al*., 2009]. Comparison of 1 and 3″ model performance demonstrates how crucial accurate terrain data are in producing quality hazard data. The authors therefore reiterate the plea of *Schumann et al*. [2014] for a global terrain data set of comparable horizontal and vertical accuracy to NED, so that hazard layers exhibiting the quality of those developed here for CONUS can be replicated across the world. Further to this, the necessity of a comprehensive flood defense catalogue has been clearly demonstrated. Levee delineation is a crucial determinant of flood hazard, so an incomplete NLD has meant the modeled floodplain is an overprediction in some areas. This data set needs to be improved and a global inventory of flood defenses is required for further advancement in this field.

Acknowledgments
The authors acknowledge the importance of Google Earth Engine in the generation of these findings, without which they would have taken significantly longer to produce. FEMA flood maps are available from the FEMA Flood Map Service Center (http://www.msc.fema.gov). USGS flood maps are available from https://water.usgs.gov/osw/flood_inundation/libraries.html. The National Land Cover Database is available from the Multi-Resolution Land Characteristics Consortium at http://www.mrlc.gov/nlcd2011.php. To access topography data used by the continental and global models, visit http://www.ned.usgs.gov for NED and the Consortium for Spatial Information for the Version 4 release of SRTM (http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1). The USACE NLD is available at http://www.nld.usace.army.mil. Visit http://www.hydrosheds.org for access to the HydroSHEDS global hydrography data set. In combination with the methodology, results and supporting information of *Smith et al*. [2015], our underlying discharge data are available from the Global Runoff Data Center (http://www.bafg.de/GRDC/EN/01_GRDC/12_plcy/data_policy_node.html) for research purposes (requires registration). NOAA IDF data are available from the NOAA Precipitation Frequency Data Server (http://www.hdsc.nws.noaa.gov/hdsc/pfds). The FAO HWSD can be accessed from the FAO Soils Portal (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en). The impervious surface area data set of *Elvidge et al*. [2007] is derived from satellite luminosity data available from the NOAA Earth Observation Group website (http://www.ngdc.noaa.gov/eog). The hydraulic model LISFLOOD-FP is available for non-commercial research purposes by contacting Paul Bates at the University of Bristol (paul.bates@bristol.ac.uk). The continental and global model data are proprietary in nature but can be made available for research use (i.e., not commercial, policy, or regulatory applications) by contacting Christopher Sampson at Fathom (c.sampson@fathom.global). Paul Bates was supported by a Royal Society Wolfson Research Merit Award. The work in this paper was in part supported by UK Natural Environment Research Council grant NE/M007766/1.

References

American Society of Civil Engineers (2017), *2017 Infrastructure Report Card: Levees*, Reston, Va. [Available at https://www.infrastructurereportcard.org/cat-item/levees/.]

Association of State Floodplain Managers (2013), *Flood Mapping for the Nation: A Cost Analysis for the Nation's Flood Map Inventory*, Madison, Wis. [Available at http://www.floods.org/acefiles/documentlibrary/2012_NFIP_Reform/Flood_Mapping_for_the_Nation_ASFPM_Report_3-1-2013.pdf.]

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

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

Baugh, C. A., P. D. Bates, G. Schumann, and M. A. Trigg (2013), SRTM vegetation removal and hydrodynamic modelling accuracy, *Water Resour. Res.*, *49*, 5276–5289, doi:10.1002/wrcr.20412.

Benedict, S. T., A. W. Caldwell, and J. M. Clark (2013), Flood-inundation maps for the Saluda River from Old Easley Bridge Road to Saluda Lake Dam near Greenville, South Carolina, *U.S. Geol. Surv. Sci. Invest. Map*, *3244*, 7 p.

Cohen, W. B., et al. (2017), How similar are forest disturbance maps derived from different Landsat time series algorithms?, *Forests*, *8*(4), 98, doi:10.3390/f8040098.

Coon, W. F. (2013), Flood-inundation maps for the Flatrock River at Columbus, Indiana, 2012, *U.S. Geol. Surv. Sci. Invest. Map*, *3241*, 12 p.

Czuba, C. R., J. D. Fallon, C. R. Lewis, and D. F. Cooper (2014), Development of flood-inundation maps for the Mississippi River in Saint Paul, Minnesota, *U.S. Geol. Surv. Sci. Invest. Rep.*, *2014–5079*, 24 p.

de Moel, H., J. van Alphen, and J. C. J. H. Aerts (2009), Flood maps in Europe—Methods, availability and use, *Nat. Hazards Earth Syst. Sci.*, *9*, 289–301.

Donchyts, G., F. Baart, H. Winsemius, N. Gorelick, J. Kwadijk, and N. van de Giesen (2016), Earth's surface water change over the past 30 years, *Nat. Clim. Change*, *6*, 810–813, doi:10.1038/nclimate3111.

Dong, J., X. Xiao, M. A. Menarguez, G. Zhang, Y. Qin, D. Thau, C. Biradar, and B. Moore III (2016), Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine, *Remote Sens. Environ.*, *185*, 142–154, doi:10.1016/j.rse.2016.02.016.

Dottori, F., and E. Todini, (2012), Testing a simple 2D hydraulic model in an urban flood experiment, *Hydrol. Processes*, *27*(9), 1301–1320, doi:10.1002/hyp.9370.

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

Du, J., L. Qian, H. Rui, T. Zuo, D. Zheng, Y. Xu, and C.-Y. Xu (2012), Assessing the effects of urbanization on annual runoff and flood events using an integrated hydrological modelling system for Qinhuai River Basin, China, *J. Hydrol.*, *464–465*, 127–139, doi:10.1016/j.jhydrol.2012.06.057.

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

Environment Agency (2013), *What is the updated Flood Map for Surface Water?*, Bristol, U. K. [Available at http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/297432/LIT_8988_0bf634.pdf.]

Federal Emergency Management Agency (1995), *Managing Floodplain Development in Approximate Zone A Areas*, Washington, D. C. [Available at https://www.fema.gov/media-library-data/20130726-1545-20490-4110/frm_zna.pdf.]

Federal Emergency Management Agency (2015), *Guidance for Flood Risk Analysis and Mapping: Data Capture—Workflow Details*, Washington, D. C. [Available at https://www.fema.gov/media-library-data/1449866207645-91e63a2d581b55c63166ca5cad03596f/Data_Capture_Guidance_Workflow_Details_Nov_2015.pdf.]

Federal Emergency Management Agency (2016a), *NFIP Flood Insurance Manual*, Washington, D. C. [Available at http://www.fema.gov/media-library-data/1458925981276-b3f8ab92d6abf14365ad154954e4dc69/changepackage_508_apr16b.pdf.]

Federal Emergency Management Agency (2016b), *FEMA Policy Standards for Flood Risk Analysis and Mapping*, Washington, D. C. [Available at https://www.fema.gov/media-library-data/1480449548025-4736d89b89d30fbf102228680c1f8acd/Standards_for_Flood_Risk_Projects_(Nov2016).pdf.]

Gallant, J. (2011), *Adaptive Smoothing for Noisy DEMs*, Geomorphometry, Canberra.

Gesch, D. B., M. J. Oimoen, and G. A. Evans (2014), *Accuracy Assessment of the US Geological Survey National Elevation Dataset, and Comparison with Other Large-Area Elevation Datasets—SRTM and ASTER*. [Available at https://pubs.usgs.gov/of/2014/1008/pdf/ofr2014-1008.pdf].

Google (2016), *Google Earth Engine API Developer's Guide*, Mountain View, Calif. [Available at https://developers.google.com/earth-engine/.]

Guinot, V., and C. Delenne (2014), Macroscopic modelling of urban floods, *La Houille Blanche Rev. Int. l'eau, EDP Sci.*, *6*, 19–25, doi:10.1051/lhb/2014058.

Hoard, C. J., K. K. Fowler, M. H. Kim, C. D. Menke, S. E. Morlock, M. C. Peppler, C. M. Rachol, and M. T. Whitehead (2010), Flood-inundation maps for a 15-mile reach of the Kalamazoo River from Marshall to Battle Creek, Michigan, *U.S. Geol. Surv. Sci. Invest. Map*, *3135*, 6 p.

Homer, C. G., J. A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. D. Herold, J. D. Wickham, and K. Megown (2015), Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information, *Photogramm. Eng. Remote Sens.*, *81*(5), 345–354.

Horritt, M. S., and P. D. Bates (2002), Evaluation of 1D and 2D numerical models for predicting river flood inundation, *J. Hydrol.*, *268*(1–4), 97–99, doi:10.1016/S0022-1694(02)00121-X.

Hunter, M. N., et al. (2008), Benchmarking 2D hydraulic models for urban flood simulations, *Proc. Inst. Civ. Eng. Water Manage.*, *161*(1), 13–30, doi:10.1680/wama.2008.161.1.3.

Icaga, Y., E. Tas, and M. Kilit (2016), Flood inundation mapping by GIS and a hydraulic model (HEC RAS): A case study of Akarcay Bolvadin subbasin, in Tukey, *Acta Geobalcanica*, *2*(2), 111–118, doi:10.18509/AGB.2016.12.

Kim, B., B. F. Sanders, J. S. Famiglietti, and V. Guinot (2015), Urban flood modeling with porous shallow-water equations: A case study of model errors in the presence of anisotropic porosity, *J. Hydrol.*, *523*, 680–692, doi:10.1016/j.jhydrol.2015.01.059.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel (2006), World Map of the Köppen-Geiger climate classification updated, *Meteorol. Z.*, *15*, 259–263, doi:10.1127/0941-2948/2006/0130.

Lehner, B., K. Verdin, and A. Jarvis (2008), New global hydrography derived from spaceborne elevation data, *Eos Trans. AGU*, *89*(10), 93–94, doi:10.1029/2008EO100001.

Merz, R., and G. Blöschl (2009), A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria, *Water Resour. Res.*, *45*, W01405, doi:10.1029/2008WR007163.

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

Murphy, E. A., D. T. Soong, and J. B. Sharpe (2012), Flood-inundation maps for a nine-mile reach of the Des Plaines River from Riverwoods to Mettawa, Illinois, *U.S. Geol. Surv. Sci. Invest. Rep.*, *2012–5227*, 17 p.

Musser, J. W., and T. R. Dyar (2007), Two-dimensional flood-inundation model of the Flint River at Albany, Georgia: Atlanta, Georgia, *U.S. Geol. Surv. Sci. Invest. Rep.*, *2007–5107*, 49 p.

National Research Council of the National Academies (2015), *Tying Flood Insurance to Flood Risk for Low-Lying Structures in the Floodplain*, The Natl. Acad. Press, Washington, D. C., doi:10.17226/21720.

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

Nobre, A. D., L. A. Cuartas, M. Hodnett, C. D. Rennó, G. Rodrigues, A. Silveira, M. Waterloo, and S. Saleka (2011), Height Above Nearest Drainage—A hydrologically relevant new terrain model, *J. Hydrol.*, *404*, 13–29, doi:10.1016/j.jhydrol.2011.03.051.

Nobre, A. D., L. A. Cuartas, M. R. Momo, D. L. Severo, A. Pinheiro, and C. A. Nobre (2016), HAND contour: A new proxy indicator for inundation extent, *Hydrol. Processes*, *30*, 320–333, doi:10.1002/hyp.10581.

Ostheimer, C. J. (2013), Development of flood profiles and flood-inundation maps for the Village of Killbuck, Ohio, *U.S. Geol. Surv. Open File Rep.*, *2013–1032*, 49 p.

Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward (2016), High-resolution mapping of global surface water and its long-term changes, *Nature*, *540*, 418–422, doi:10.1038/nature20584.

Rabus, B., M. Eineder, A. Roth, and R. Bamler (2003), The Shuttle Radar Topography Mission—A new class of digital elevation models acquired by spaceborne radar, *ISPRS J. Photogramm. Remote Sens.*, *57*(4), 241–262, doi:10.1016/S0924-2716(02)00124-7.

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

Roland, M. A., S. M. Underwood, C. M. Thomas, J. F. Miller, B. A. Pratt, L. G. Hogan, and P. A. Wnek (2014), Flood-inundation maps for the Susquehanna River near Harrisburg, Pennsylvania, 2013, *U.S. Geol. Surv. Sci. Invest. Rep.*, *2014–5046*, 17 p.

Salinas, J. L., G. Laaha, M. Rogger, J. Parajka, A. Viglione, M. Sivapalan, and G. Blöschl (2013), Comparative assessment of predictions in ungauged basins—Part 2: Floods and low flow studies, *Hydrol. Earth Syst. Sci.*, *17*, 2637–2652, doi:10.5194/hess-17-2637-2013.

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

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

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

Schumann, G. J.-P., P. D. Bates, J. C. Neal, and K. M. Andreadis (2014), Technology: Fight floods on a global scale, *Nature*, *507*, 169–169, doi:10.1038/507169e.

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

Storm, J. B. (2014), An expanded model—Flood-inundation maps for the Leaf River at Hattiesburg, Mississippi, 2013, *U.S. Geol. Surv. Sci. Invest. Map*, *3300*, 8 p.

Trigg, M. A., et al. (2016), The credibility challenge for global fluvial flood risk analysis, *Environ. Res. Lett.*, *11*, 1–10, doi:10.1088/1748-9326/11/9/094014.

United States Army Corps of Engineers (2016a), *Hydrologic Modeling System HEC-HMS User's Manual* , Washington, D. C. [Available at https://www.hec.usace.army.mil/software/hec-hms/documentation/HEC-HMS_Users_Manual_4.2.pdf.]

United States Army Corps of Engineers (2016b), *HEC-RAS River Analysis System User's Manual*, Washington, D. C. [Available at https://www.hec.usace.army.mil/software/hec-ras/documentation/HEC-RAS%205.0%20Users%20Manual.pdf.]

United States Geological Survey (2016), *FESWMS-2DH: Finite-Element Surface-Water Modeling System for Two-Dimensional Flow in the Horizontal Plane*, Reston, Va. [Available at https://water.usgs.gov/software/FESWMS-2DH/.]

Watson, K. M., and M. J. Niemoczynski (2014), Flood-inundation maps for the Saddle River in Ho-Ho-Kus Borough, the Village of Ridgewood, and Paramus Borough, New Jersey, 2013, *U.S. Geol. Surv. Sci. Invest. Map*, *3299*, 10 p.

Winsemius, H. C., L. P. H. Van Beek, B. Jongman, P. J. Ward, and A. Bouwman (2013), A framework for global river flood risk assessments, *Hydrol. Earth Syst. Sci.*, *17*, 1871–1892, doi:10.5194/hess-17-1871-2013.

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

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