*Author:*
**Wing, Oliver E J**

*Title:*
**Understanding flood risk at large spatial scales**

*Testing, development and application of a continental flood hazard model*

# Understanding Flood Risk at Large Spatial Scales: Testing, Development and Application of a Continental Flood Hazard Model

Oliver E. J. Wing

*A dissertation submitted to the University of Bristol in accordance with the requirements for the award of the degree of Doctor of Philosophy in the Faculty of Sciences.*

*School of Geographical Sciences, September 2019*

Word count: 60,507

# Abstract

Floods pose a major global economic threat, with recorded damages exceeding $1 trillion since 1980. Our ability to mitigate these disasters is dependent on understanding the location, severity and probability of flood hazard: wrought from computer models of flood inundation. These models, though, are traditionally of individual river reaches; built locally by hydraulic engineers with accurate flow, topography and bathymetry data. By leaving the vast majority of the world's rivers unmodelled, flood risk has historically been largely unquantified globally. To fill this information gap, recent advances in data availability and computational capacity have heralded the advent of global flood models, leading to a number of previously unexplored research questions: How accurate are these emerging models? What are they useful for? How can they be improved?

In this thesis, a large-scale flood hazard model of the US is comprehensively evaluated against thousands of local-scale models as well as observations of real flood events. This procedure uncovered that the skill of large-scale structures is approaching convergence with that of traditional local analyses. Model validation revealed new applications of total-coverage flood inundation models. Flood risk estimates have been updated for the US, finding the low-coverage assemblage of government flood maps underestimates risk considerably. The model also found applicability in indicating potential inundation from incoming storms, a component often neglected by official forecasts. Validation further exposed areas where model development is needed. A solution to the identified issue of poor representation of structural flood defences is presented, ensuring protected areas are modelled as such.

This thesis illustrates the considerable utility of emerging large-scale model structures and offers an improvement to a weakly constrained component within these. A lack of data will inhibit further improvements; thus, an expansion in the scale of publicly-available accurate elevation and bathymetry data is required to replicate these methods globally.

# *Author's Declaration*

*I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.*

SIGNED.....

DATE.....

# Acknowledgements

*"Not everything that counts can be counted, and not everything that can be counted counts."*

William Bruce Cameron

*With eternal gratitude to…*

Paul Bates,

Chris Sampson,

Andy Smith,

Jeff Neal,

Mum, Dad, and Rory

*…without whom this thesis would never have been written.*

# Table of Contents

# List of Figures

# List of Tables

# 1 INTRODUCTION

Recent extreme flood events have brought the issue of flood risk to the forefront of political, corporate, scientific and public consciousness. Flooding across the Mississippi river basin and landfalling hurricanes Harvey, Irma, Maria, Florence and Michael in the US; heavy summer monsoon rains in Kerala, India; successive intense storms in Japan; and rapid snowmelt in Quebec, Canada are just some of the high profile flood events that occurred, with differing geographies and hydrological drivers, since 2017. In the past 40 years, across almost 6000 recorded events, flooding has caused economic losses in excess of $1 trillion and over 220,000 deaths (Munich Re, 2019). These totals mask a distinct upward trend in flood losses over time, primarily driven by increased socio-economic activity in hazardous areas (Jongman et al., 2012a). According to the Intergovernmental Panel on Climate Change (IPCC), a climatological signal is yet to be detected in the observational loss record, though this may be due to a scarcity of hydrological data and the confounding effect of land-use change (IPCC, 2012).

The limited evidence in support of a climate change effect in historical data, coupled with the considerable uncertainty inherent in future flood simulations by atmospheric-hydrologic model cascades (Hirabayashi et al., 2013; Winsemius et al., 2016; Alfieri et al., 2017), means the IPCC has low confidence in projections of future inland flooding (IPCC, 2012). With that being said, an enhanced hydrological cycle in a warming world is expected to lead to increased rainfall and thus flood hazard (Kundzewicz et al., 2014), but this would be highly variable within and between regions. There is, for instance, a growing body of evidence that freshwater flooding from tropical cyclones will worsen because of their increased intensity (Kang and Elsner, 2015; Sobel et al., 2016; van Oldenborgh et al., 2017) and slower movement post-landfall (Kossin, 2018) under projected climatic changes. Coastal flooding, meanwhile, is better understood as a function of sea level rise which, leaving potential changes to storminess aside, has and will continue to exacerbate extreme coastal high water levels (IPCC, 2012). More transient pluvial events, with higher sensitivity to increased rainfall under climate change than fluvial floods which can be more dependent on antecedent conditions, are also expected to intensify (Westra et al., 2014; Do et al., 2017).

Amidst this backdrop of catastrophic loss of life and livelihoods in recent history, continued development in flood-prone areas and an uncertain threat posed by climate change, it is clear

that mitigation of flood risk should be a major objective across scales and sectors. This is reflected in the Sendai Framework for Disaster Risk Reduction, a United Nations (UN) initiative which outlines actionable targets and priorities for international governments to meet this goal (UNISDR, 2015a). The first priority of the Sendai Framework is to understand risk, for which it is necessary to break it down into its component parts: hazard, exposure and vulnerability. The following definitions provided are consistent with terminology set out by the UN and the IPCC (UNISDR, 2015b; IPCC, 2012): 'hazard' refers to the nature, probability and severity of the physical phenomenon itself; 'exposure' outlines the assets, tangible or intangible, that could be affected by the hazard; and 'vulnerability' indicates the susceptibility of these entities to experience adversity. Thus, any prevailing predilection for viewing flood risk predominantly as a function of its physical characteristics does disservice to the crucial study of how and where this interacts with human interests (e.g. Jongman et al., 2012a, 2015). While this thesis generally focuses on improved understanding of inland flood hazard, the relative futility of doing so for the purposes of understanding risk without commensurate recognition of exposure and vulnerability is duly noted.

Understanding of the nature of flood hazard is principally wrought through computational models. Most, if not all, river reaches have not been observed to a level sufficient for the comprehension of their flood behaviour in isolation. To fill this gap, what limited spatial, temporal and dimensional observations are available are combined with a physical understanding of flooding to simulate synthetic representations of flood hazard. Flood inundation models require flow data, ultimately derived from discharge or precipitation measurements, and a topographic representation of the river channel, its surrounding floodplain and their friction. The digitization and expansion of these hydrological data, alongside the advent of remotely-sensed terrain information (e.g. Marks and Bates, 2000; Bates et al., 2003; Mason et al., 2003), have created enough data-richness to enable high physical process representation in hydraulic models. Validation exercises, where event-replicating model structures are compared to some real-world observations of flood inundation from the said event, demonstrate such tools have significant skill (e.g. Bates and De Roo, 2000; Bates et al., 2006; Neal et al., 2009a; Altenau et al., 2017). In this context, significant skill means observations of maximum inundation extent during a large flood event are reproduced by the model with a degree of similarity consistent with the underlying uncertainties in the boundary conditions, terrain data, model structure and benchmark observations. In cases where high quality data are available to build and calibrate the model,

they may even be more accurate and useful than the observations themselves through their spatially and temporally continuous representation of the physics of flood inundation, e.g. wetting and dry of the floodplain, lagging and attenuation of the flood wave (e.g. Bates et al., 2006; Neal et al., 2009a), effectively forming a flood event reanalysis.

The caveat that models perform well when built with quality data is significant. Areas where there is both enough information to suitably characterise the flow regime of a river and accurate measurements of channel bathymetry and floodplain topography constitute a minute proportion of the Earth's surface. If an understanding of flood hazard, and thus risk, is only accessible where bespoke, local-scale models can be built manually with *in situ* data, then Priority 1 of the Sendai Framework will not be achieved in any meaningful way at the global scale. It is unsurprising, therefore, that only 12% ($126 billion) of the ~$1 trillion in flood losses since 1980 were insured (Munich Re, 2019), since there is not enough knowledge about the hazard for insurers to confidently expand their exposure. Insurance is a crucial instrument, alongside preventative management, by which the effects of flooding can be mitigated: ensuring those impacted are compensated to rebuild livelihoods post-event.

In the past decade, hydraulic modellers have expanded the spatial scale of the domains they study. Driven by ever-increasing computational capacity and the availability of global spaceborne radar elevation datasets (e.g. Rabus et al., 2003), regional, continental and even global flood modelling exercises are now possible (e.g. Wilson et al., 2007; Alfieri et al., 2014; Winsemius et al., 2013; Sampson et al., 2015). A number of different academic and commercial institutions have produced global flood hazard models, of which six were compared in a study by Trigg et al. (2016). They concluded that these models show little agreement in defining hazardous areas. When tested over the continent of Africa, there was only 30–40% agreement in flood extent. The study was not a formal validation exercise, since no benchmark data (which the models should aspire to resemble) were used. Large-scale models generally suffer from under-validation. Papers in which their methods are presented contain comparisons to given benchmarks of a scale far from commensurate to that of the model, mostly because such data are unavailable (e.g. Alfieri et al., 2014; Sampson et al., 2015; Winsemius et al., 2016; Dottori et al., 2016; Ward et al., 2017). Furthermore, in the few cases where large-scale models have been benchmarked against the same data with consistent metrics (e.g. Po, Severn and Thames catchments), the number of model pixels in agreement with those of the benchmark dataset is broadly similar to that of when they disagree:

inundation is generally modelled incorrectly as often as it is correctly (Dottori et al., 2016; Winsemius et al., 2016; Sampson et al., 2015). There is thus a clear need for more comprehensive model validation: the current use of such structures to support planning decisions, insurance premium pricing, event forecasting, risk management strategies, land acquisition and much else is largely unjustified at present. In doing so, model validation unlocks myriad opportunities to explore applications for which they may be proven fit-for-purpose and also highlights facets of the model which require improvement.

This thesis is centred around these themes. Working with a continental-scale flood hazard model of the United States based on the global-scale variant of Sampson et al. (2015) as an exemplar, it aims to do the following:

(i) present a first comprehensive validation procedure for these emerging large-scale flood hazard models;
(ii) discover new applications for these model structures, thus revealing the purposes for which existing models are appropriate;
(iii) identify circumstances under which contemporary model structures break down and to, therefore, develop methods for their improvement.

Naturally, these aims are interdependent in that model evaluation leads to discovery of appropriate applications and limitations, and addressing limitations requires further validation which cascades into ever more advanced model applications. In the literature review that follows, historic and expected changes to flooding will be outlined; the methods by which flood hazard has historically been understood will be examined; the expansion in spatial scale of the model domains considered will be explained and critically discussed; and the contemporary translation of knowledge about flood hazard to flood risk will be explored. This evaluation of scientific literature, alongside the aims presented here, will generate thorough research objectives to be addressed in the subsequent chapters.

# 2 LITERATURE REVIEW

## 2.1 Observations and projections of flooding

### 2.1.1 Introduction

This section examines trends (past) and projections (future) of flood hazard and risk globally. Flooding can occur in a number of distinct contexts, which include: (i) *storm surges* at the coasts, where low atmospheric pressure (e.g. during an intense depression or cyclone) or strong onshore winds increase the surface sea level; (ii) *tidal flooding* also at coasts, where isostatic or eustatic change drives inundation of low-lying coastal areas even in the absence of a meteorological event; (iii) *groundwater flooding*, where permeable alluvium becomes saturated and the water table rises above the land surface; (iv) *anthropogenic floods* are driven principally by humans and most commonly relate to infrastructure failures (e.g. dams and levees); (v) *snow-* or *ice-melt floods*, initiated by an increase in air temperature which triggers a melting of the snowpack or glacier ablation; (vi) *pluvial flooding*, occurring where precipitation from intense storms cannot infiltrate the ground surface and instead runs over it (due to, e.g., antecedent soil saturation, precipitation intensity exceeding infiltration capacity, or widespread impervious surface cover); (vii) *fluvial flooding*, where river channel capacity is exceeded and water spills out onto the surrounding floodplain. Furthermore, the simultaneous occurrence of multiple flood types, *compound flooding*, can occur where statistically dependent flood drivers lead to amplified impacts compared to where they are traditionally considered in isolation (e.g. fluvial or pluvial floods coinciding with storm surge or tidal events at the coast: Wahl et al., 2015; Moftakhari et al., 2017). This thesis focuses mainly on pluvial (vi) and fluvial (vii) flooding. Even in this apparent microcosm of flood types, the drivers of these are highly heterogenous (Nied et al., 2014). Floods are driven not only by intense rainfall, but are also strongly controlled by antecedent hydrological conditions, temperature, anthropogenic interaction and catchment characteristics (Berghuijs et al., 2016, 2019; Sikorska et al., 2015; Paschalis et al., 2014). While the frequency and magnitude of extreme precipitation may have increased over time (Alexander, 2016), it has a decoupled, even dichotomous, relationship with flood magnitude (Sharma et al., 2018). For instance, Ivancic and Shaw (2015) found that only 36% of river discharge exceeding the 99$^{th}$ percentile corresponded to precipitation intensity above the 99$^{th}$ percentile in a study across the US. This makes explaining trends or modelling projections of flooding highly complex, particularly with respect to climate change, since (thermo)dynamical phenomena are only a single part of a multi-faceted problem. Confounding this further is an analysis of flood losses

amidst continual demographic and economic growth. While flood losses have increased over time, so too have the societal contexts in which the physical flooding took place (UNISDR, 2015b). Trends in flood risk are thus an entanglement of changes to hazard, exposure and vulnerability: isolating the relative contribution of each is difficult.

### 2.1.2 Trends

By its very nature, extreme flooding is difficult to observe. Calculating mean trends in streamflow draws on all available data in a given record, while extreme trends require only the tail of the flow data. There is a large gulf in statistical power between these two analyses, meaning the latter requires substantially more data to characterise patterns faithfully. Streamflow observations are also impaired by their limited spatial distribution and non-uniform operation, processing and quality control techniques (Hannah et al., 2011). Furthermore, methods are inconsistent in trend detection, probability distribution fitting, time window consideration and the choice of statistical phenomenon to measure (e.g. annual maximum, peak-over-threshold, 1 in 100 year), all of which may lead to different findings. Owing to the diversity of driving processes, analysis tools and data availability, there is no valid universal statement on observations of flooding.

Do et al. (2017) employed ~2000 gauging stations across the world from the Global Runoff Data Centre in an analysis of peak flow trends since 1966. They found that more stations exhibited a decreasing trend (western North America, Australia, Scandinavia, eastern Brazil, parts of Africa) than an increasing one (western Europe, southern Africa, parts of South America, eastern North America), yet the majority had no statistically significant trend. The analysis is limited, however, by the sparse coverage of station data, which is particularly lacking in parts of South America, Africa and Asia. The record length, too, set at a minimum of 30 years, inhibits the exploration of trends in more extreme, high-magnitude floods by necessitating an analysis of annual streamflow maxima rather than, say, 1 in 100 year floods. Interestingly, Do et al. (2017) found smaller catchments were more likely to exhibit positive trends than larger catchments. This is perhaps consistent with trends in precipitation, where sub-daily rainfall is observed to have intensified more rapidly than accumulations measured at daily time scales (e.g. Westra et al., 2014). Small catchments are much more sensitive to intense, localised storms as a flood driver: their reduced drainage area and steeper slopes relative to large catchments leads to more infiltration excess overland flow (Horton, 1933; Dunne and Black, 1970) and a shorter time to concentration, producing ‘flash floods’. Global

inconsistencies were also highlighted in an analysis by Gudmundsson et al. (2019), where a regional clustering of similar trends is evident. Increasing trends in the 90$^{th}$ percentile or maximum streamflow were charted for eastern Asia, southern South America, northern and central Europe and central North America, while decreasing trends were found in the Mediterranean, north-eastern Brazil, western North America and southern Australia.

Fragmenting into specific regional analyses, the picture is much the same. In North America, Mallakpour and Villarini (2015) found the vast majority of USGS gauging stations in the central US had no significant trend in annual flow maxima (13% increasing; 7% decreasing) since 1962. This is consistent with findings across a number of similar studies of this region (Villarini et al., 2009a, 2011; Lins and Slack, 2005; Peterson et al., 2013). Hirch and Ryberg (2012) note statistically significant decreases in trends of peak flows in the southwest US and increases in the northeast US, consistent with the global analyses of Do et al. (2017) and Gudmundsson et al. (2019) and the study of northeast US by Armstrong et al. (2014). Using a peaks-over-threshold approach, though, Mallakpour and Villarini (2015) found an increase in flood event *frequency* for 34% of central US stations. The emergence of increased frequency, rather than magnitude, of flooding in the US has been evidenced elsewhere also (Vogel et al., 2011; Hirsch and Archfield, 2015; Archfield et al., 2016). In Canada, Burn and Whitfield (2016) note decreasing trends in flood magnitude for nival catchments, increasing flood trends for pluvial catchments and earlier, more frequent flood events in mixed catchments. These observations have the signature of diminished winter snowpack and increased heavy rainfall, suggesting flood drivers are changing across Canada. This is supported by findings in a study of Canadian streamflow data by Cunderlik and Ouarda (2009).

In Europe, trends are similarly mixed (Hall et al., 2014; Medeiro et al., 2015; Kundzewicz et al., 2018; Mangini et al., 2018). The western Mediterranean region sees consistent decreasing trends in annual flow peaks across a rich collection of studies (Mediero et al., 2014; Llasat et al., 2010; López-Moreno et al., 2006; Renard et al., 2008), which was also reflected in the global study of Gudmundsson et al. (2019). Further east, though, in Greece, Diakakis (2014) found increased frequency of flood events in Athens since 1880, while Montanari (2012) found no trend in flood peaks for the Po river in Italy. In northern Spain, Medeiro et al. (2014) highlighted an increase in the magnitude of floods since 1959, as did Renard et al. (2008) and Giuntoli et al. (2012) in northern France. Further north in Atlantic western Europe, maritime areas of the British Isles have also seen an increase in high flows since the

mid-20th Century (Hannaford and Marsh, 2008; Hannaford and Buys, 2012; Murphy et al., 2013). The overall UK picture is mixed however, with strong seasonal and regional variation in high flow trends (Hannaford and Buys, 2012). Robson et al. (1998) found no trend in flood peaks across the UK, but identified similar findings to Hannaford and Marsh (2008) in their relationship to climatic oscillations. Wilson et al. (2010) did not find evidence for decreased snowmelt- and increased rainfall-driven floods in Scandinavia (unlike studies in Canada), as others found highly heterogenous patterns in time and space for the Nordic region (Lindström and Alexandersson, 2004; Korhonen and Kuusisto, 2010; Stahl et al., 2012). In Germany, Petrow and Merz (2009) found that annual peak flows have increased in southern, western and central regions, especially in winter, as Mudelsee et al. (2003) identified decreased flood frequency for eastern German winter. Meanwhile, Bormann et al. (2011) found no statistically significant trends in annual maxima or flood frequencies in Germany. In Switzerland, studies have identified distinct upward trends in annual flood peaks, though note oscillations in flood-rich and flood-poor periods align with atmospheric circulation patterns (Birsan et al., 2005; Allamano et al., 2009; Schmocker-Fackel and Naef, 2010). In the Alps, Bard et al. (2012) found earlier and higher-magnitude snowmelt-driven floods since 1960, consistent with the French analysis by Renard et al. (2008). In Poland, the Baltics and eastern Europe, a decreasing trend in high flows has been identified (Strupczewski et al., 2001; Pekárová et al., 2009; Reihan et al., 2012; Birsan et al., 2014; Stahl et al., 2012). In Austria and Hungary, there are few gauging stations exhibiting statistically significant trends (Somlyódy et al., 2010; Blöschl et al., 2011).

Analyses in regions outside of North America and Europe are sparse, predominantly due to data availability. In Australia, Ishak et al. (2013) found decreasing annual maximum flows in the south, corroborating the findings of Do et al. (2017) and Gudmundsson et al. (2019). East Asia sees large regional disparities. Delgado et al. (2010) posit that the likelihood of flooding on the Mekong river has increased, as do Jiang et al. (2008) of the Yangtze river. Yet, Bai et al. (2016) demonstrated a negative trend in annual peaks in central China and Zhang et al. (2015) found no trend in peak flows for southern China. In western Africa, Nka et al. (2015) showed Sahelian catchments exhibited positive trends in flood magnitude, while Sudanian catchments were negatively trended. In South America, Marengo (1995, 2009) found no statistically significant trend in streamflow across the continent, linking its variability to the El Niño Southern Oscillation.

Overall, then, it is impossible to generalise trends in flood hazard at the global scale. Trends differ between regions and seasons, and studies of the same region often disagree. The IPCC (Seneviratne et al., 2012) has low confidence in even asserting the direction of change in flooding globally, let alone its magnitude or cause. In spite of an unclear trend in changes to the physical phenomenon of flooding, there is an overwhelming signal in flood *risk* trends over time (Miller et al., 2008; Barthel and Neumayer, 2012; Mohleji and Pielke, 2014; Jongman et al., 2012a; Munich Re, 2019). Flood-related losses have unambiguously increased over time, yet this is against a backdrop of increasing population and wealth. Correcting loss data to account for non-stationary socio-economics is known as 'normalisation': with such changes accounted for, any remaining signal could arguably be attributed to climatology (e.g. Pielke and Landsea, 1998). However, most studies that have normalised flood losses for different flood drivers and regions find no temporal trend (Bouwer, 2011). Barredo (2009) found no trend in normalised European flood losses (1970 – 2008); Barredo et al. (2012) showed there was no trend in normalised insured flood losses in Spain (1971 – 2008); Crompton and McAneney (2008) did not find a trend in normalised weather-related insured losses in Australia (1967 – 2006); Weinkle et al. (2018) found no trend in normalised hurricane damages in the US (1900 – 2017); Downton and Pielke (2005) showed there was no trend in normalised US flood losses also (1926 – 2000); Pielke et al. (2003) found no evidence of a trend in normalised hurricane damages for Latin America (1944 – 1999); Raghavan and Rajesh (2003) showed no trend in normalised cyclone losses for India (1977 – 1998). In eastern Asia, though, studies of normalised flood losses have shown an increasing trend in China (Fengqing et al., 2005) and Korea (Chang et al., 2008), with the authors positing that changes could be due to land-use change or increased annual precipitation. At the global scale, analyses have not found any trend in normalised flood losses (Barthel and Neumayer, 2012; Mohleji and Pielke, 2014). This is seemingly consistent with the mixed and inconsistent trends in flood hazard outlined previously, and suggests increased risk is exclusively a function of socio-economic change.

The process of normalisation is non-trivial however, with sensitivity to the nature, quality and scale of the data that informs it (e.g. wealth, population, GDP, asset values, properties). Neumayer and Barthel (2011) point out that, while changes across time are accounted for, changes across space at a point in time are often not. It may be the case, for instance, that flood-prone areas have developed at different rates to non-hazardous zones, biasing the normalisation. Furthermore, with few exceptions (e.g. Crompton and McAneney, 2008),

normalisation studies do not account for changes in vulnerability. Stronger building codes and effective early warning systems can reduce flood losses with hazard and exposure held constant. If vulnerability to flooding has decreased over time (Jongman et al., 2015), then no trend in exposure-normalised losses would actually suggest increases to flood hazard. Even so, the IPCC (2012) has not attributed increasing trends in flood losses to climate change, stating that increased exposure has been the major driver of this, supported by the changes in flood exposure charted by Jongman et al. (2012a) and Ceola et al. (2014).

### 2.1.3 Projections

If assessing historical changes using observations is difficult, simulating future changes is even more so. Understanding changes to flood hazard typically involves cascading General Circulation Model (GCM) output, driven with multiple scenarios of future atmospheric greenhouse gas concentrations, through a hydrological model to simulate streamflow. If extreme flows of the future occur at a high frequency or magnitude in comparison to a baseline simulation driven with present-day atmospheric conditions, flood hazard can be projected to increase. Unfortunately, both of these model components are highly uncertain in the context of future flood projections. At ~25 km resolution, state-of-the-science GCMs start to simulate tropical cyclones and intense storms more realistically, yet they still require parameterisation of cumulus convection (Wehner et al., 2014). Kendon et al. (2012) found even a 12 km resolution climate model significantly underpredicts heavy rainfall, and that a <2 km resolution, convection-permitting model is required to faithfully represent rainfall. While very high resolution climate models (intractably fine for GCMs) are necessary for reliable projections of extremes (Kendon et al., 2017), *not* parameterising convection in such fine-grid models means the phenomenon still remains under-resolved, biasing rainfall estimates (Chan et al., 2013). Regardless, most widely cited flood projection studies use Coupled Model Intercomparison Project 5 (CMIP5) GCMs which are 1–2 orders of magnitude coarser in their horizontal resolution (Taylor et al., 2012). The resultant inability to replicate rainfall extremes necessitates bias correction, itself an uncertain and challenging task. This involves post-processing GCM output to match observations, themselves subject to error, which can violate conservation principles and unrealistically alter the physical relationship between variables (Ehret et al., 2012; Muerth et al., 2013; Hagemann et al., 2011). Similarly, ameliorating the difference in scale between GCMs and rainfall-generating processes introduces uncertainty related to downscaling (e.g. Prudhomme and Davis, 2008). As a spatial example, the smoothing effect of having rainfall totals summarised for large

areas of diverse topography does not capture the windward/leeward precipitation differential as a result of orographic uplift. In a temporal sense, GCMs, with large time-steps, do not distinguish between transient and intense convective storms and more moderate frontal events. Downscaling is a process by which the effects of spatial and temporal coarseness in GCMs are corrected, often by using empirical relationships between modelled large-scale and observed small-scale atmospheric phenomena (*statistical downscaling*) or physical relationships between a GCM and a more finely resolved and localised climate model (*dynamical downscaling*). The method by which this post-processing of deficient GCM output is achieved introduces further uncertainty to the model cascade (Chen et al., 2011; van Pelt et al., 2009; Milly and Dunne, 2011). Flood projections are also highly sensitive to the choice of GCM employed and the emission scenarios considered (Hundecha et al., 2016; Krysanova et al., 2017; Hattermann et al., 2018). Arnell and Gosling (2016) show that flood risk projections vary between a 9% decrease and a 376% increase relative to the present-day depending on the GCM used. The IPCC (Seneviratne et al., 2012) identify the choice of GCM as being the principal component of uncertainty in hydrological projections, yet the hydrological models, too, are also uncertain (e.g. Schewe et al., 2014; Orth et al., 2015). Dankers et al. (2014) show different hydrological models forced with the same GCM can disagree on even the sign of the change at the local scale, while Roudier et al. (2016) find poor correspondence between European-scale hydrological models and discharge observations. Hydrological models are generally incapable of replicating extreme (or even average: Haddeland et al., 2011; Gruell et al., 2015; Lohmann et al., 2004) stream flows over large scales without a wealth of data to calibrate to or to inform sub-grid parameterisation (e.g. Fischer et al., 2011; Prudhomme et al., 2011; Sperna Weiland et al. 2010). While there is an implication of higher performance through the drive to finer grid resolution permitting enhanced process representation (Wood et al., 2011; Bierkens et al., 2015), this is no panacea. Such a focus should not ignore the effect of epistemic uncertainty – our lack of understanding of said processes – and thus the irrevocable and perennial need for local data to calibrate to and parameterise with (Beven and Cloke, 2012; Beven et al., 2014).

With this in mind, it is unsurprising that continental and global scale studies are highly inconsistent in their projections of future flooding. Flood projections for Europe show low spatial agreement (Hall et al., 2014; Kundzewicz et al., 2017). Regarding changes in the magnitude of flooding in southern Europe, Alfieri et al. (2015) and Roudier et al. (2016) model an increase, Dankers et al. (2014) a decrease, and Giuntoli et al. (2015) no change. In

the Alps, Köplin et al. (2014) project a decrease in flood severity, while Thober et al. (2018) project an increase. Arheimer and Lindström (2015) project a decrease in snowmelt-driven spring floods and an increase in rainfall-driven autumn/winter floods for Sweden, resulting in an annual average decrease in flood magnitude. Alfieri et al. (2015) project an increase for Sweden, though agree with Veijalainen et al. (2010) in a stable or decreasing trend for Finland. Roudier et al. (2016) project a decrease in flood magnitude for most of Scandinavia. Eastern Europe sees projections of decreased flood magnitude in the Baltics and increases in river basins east of the Danube (Dankers and Feyen, 2009; Rojas et al., 2012), though Stagl and Hattermann (2015) present a temporally mixed picture for the Danube region: decreases in summer/autumn and increases in winter/spring. In northern and northeastern Europe, Dankers and Feyen (2009) and Hirabayashi et al. (2008) project a decrease in flood magnitude, Arheimer et al. (2012) no change, and Lehner et al. (2006) and Rojas et al. (2011) an increase. In the UK and Ireland, increases in flood magnitude are generally projected (Steele-Dunne et al., 2008; Kay and Jones, 2012). In Germany, both decreases (Huang et al., 2012) and increases (Hattermann et al., 2014) are projected. Decreased flood frequency is projected for western France by Lehner et al. (2006) and Dumas et al. (2013), but an increase is projected by Rojas et al. (2011). Thobler et al. (2018) project decreasing flood magnitude in the Mediterranean region, while others project increases (Rohas et al. 2012; Dumas et al., 2013; Dankers and Feyen, 2009).

In North America, Tohver et al. (2014), Salathé et al. (2014) and Hamlet et al. (2013) project increased flooding in the Pacific Northwest; Wobus et al. (2017) project increases in western and eastern-central US, while projecting decreases for southeast and southwest US and the Great Lakes; and Naz et al. (2016) project increases in central, northern and southwest US and decreases in New England and southern Florida. In South America, Sorribas et al. (2016) project increased flood magnitude in western Amazonia and decreases in the east; broadly verified by Langerwisch et al. (2013) and Zulkafli et al. (2016). Bozkurt et al. (2018) project increases in flooding over Andean basins. In China, Liu et al. (2017) project increased flood magnitudes. Shkolnik et al. (2018) simulate increased flood intensity in northern Eurasia

Global-scale modelling studies generally project increasing flood hazard in more areas than a decreasing one (Alfieri et al., 2017; Hirabayashi et al., 2013; Dankers et al., 2014; Winsemius et al., 2016; Döll et al., 2018; Arnell and Gosling, 2016). There is rough consensus on projections of increased flows over south and southeast Asia, tropical Africa,

northern Eurasia, and northern North America and decreased flows over central and eastern Europe, the Mediterranean, and the Andes. Elsewhere, inter- and intra-study signals are mixed. In many locations these sharply contrast with regional assessments. It should be noted that agreement between different models, which all share the same fundamental constraints and limitations in their ability to represent the relevant processes, does not necessarily translate to increased confidence in the conclusions they lead to. Furthermore, localised studies are also not necessarily more accurate than global ones. Their often finer grid resolution, permitted by a reduction in model domain size, and more detailed topography and land cover data do not resolve their general inability to reproduce flood-driving processes at these spatial scales: in some cases, higher spatial resolution and complexity may even amplify these uncertainties (Mateo et al., 2017; Beven et al., 2014; Orth et al., 2015; Bout and Jetten, 2018).

Physical scientists will continue to garner better understanding of flood-driving processes and computational capacity will continue to enhance modelling capabilities, leading to more refined estimates of the magnitude and frequency of future global flood hazard. Meanwhile, as with historical trends, there is a confident expectation that future flood *risk* will increase: as a function of exposure. The IPCC (Handmer et al., 2012), based on the work of Bouwer (2010), reviewed 21 regional risk projection studies, mostly in Europe (e.g. Feyen et al., 2009; Bouwer et al., 2010) and the US (e.g. Pielke, 2007; Schmidt et al., 2009). The mean increase in estimated flood losses in 2040 as a result of climate change alone was 71%, while the same quantity attributed to socio-economic change came to 189%. Global risk analyses that separate hazard- and exposure-change contributions to future risk come to similar conclusions (e.g. Hirabayashi et al., 2013; Jongman et al., 2012a; Winsemius et al., 2016; Dottori et al., 2018). For instance, Winsemius et al. (2016) found increased future flood damages were 66 – 87% attributable to GDP increase depending on the scenario analysed.

### 2.1.4 Discussion

Changes to flood hazard are dependent on catchment characteristics, flood drivers, seasons, climate and human influences, presenting much spatial heterogeneity in the magnitude of past and future changes. If confident quantification of even of the *sign* of such changes is impaired by a lack of data, poor process understanding and its modelled representation, then determining their *magnitude* and *cause(s)* is even more so. Amidst such uncertainties relating to hazard, a clearer signal elsewhere has emerged. It can be asserted with confidence that

exposure has driven up flood losses over time and will continue to do so as long as wealth and population grows. Despite this unambiguous primary driver of flood risk change, modelled exposure representation in large-scale risk models is generally poor. Smith et al. (2019) highlight this in a comparison of ubiquitous global exposure datasets (e.g. WorldPop: Tatem, 2017; LandScan: Bright et al., 2016) to emerging higher precision layers built by applying machine learning techniques to high-fidelity, building-resolving satellite imagery (e.g. High Resolution Settlement Layer: Tiecke et al., 2017). They show flood exposure is mis-estimated significantly when population counts are distributed over wide areas, as is common via their coarse representation in current risk models (populations spread over 1–5 km resolution grid cells: Winsemius et al., 2016; Hirabayashi et al., 2013; Alfieri et al., 2017; Dottori et al., 2018). With historic and projected changes to risk being more sensitive to exposure and hazard, alongside the conclusions of Smith et al. (2019), it is clear that improvements to the exposure components of risk models are needed to more faithfully support their conclusions. Equally, flood vulnerability is understudied relative to flood hazard and is subject to much uncertainty. To estimate flood risk (e.g. losses), depth-damage functions are commonly employed, yet a quantification of the reliability of these relationships has been inhibited by data availability (e.g. Jongman et al., 2012b; Apel et al., 2009; Merz et al., 2010). Few localised assessments of vulnerability have found substantial variability in the depth-damage relationship (e.g. Merz et al., 2004; Thieken et al., 2008). The vulnerability component of flood risk models has thus been termed the ‘main bottleneck’ in estimating flood damages (Freni et al., 2010).

While the preceding sections outline fundamental limitations – relating to data scarcity, epistemic uncertainty and computational capacity – to the meteorological and hydrological aspects of a risk model cascade, inhibiting their immediate improvement, the hydraulic component of contemporary risk models have not hit such a ceiling. The solutions in this instance are not to *collect more data*, *better understand physical processes* or *get bigger computers*, but can make use of recent scientific advances in the development of fast and accurate hydrodynamic codes that are applicable at-scale. As an example, the hazard model underpinning the risk framework of Winsemius et al. (2016), set out in Winsemius et al. (2013) and Ward et al. (2013), is GLOFRIS. Using the global hydrology model PCR-GLOBWB (van Beek and Bierkens, 2009) and its river routing scheme, annual exceedance probability (AEP) flood volumes are computed for 0.5° (~50 km) resolution grid cells, which are then downscaled to 30” (~1 km). The mass-conservative downscaler iteratively increases

water levels in river cells and maintains it across floodplain cells until the 0.5° flood volume is depleted. Such volume spreading approaches (e.g. Gouldby et al., 2008), which lack a representation of the physics of flood propagation, are highly uncertain amidst the low horizontal and vertical accuracy elevation data employed in global models (Sampson et al. 2012; Kesserwani and Wang, 2014). The next section charts the accumulation of scientific knowledge, data availability and computer power that have enabled the execution of more accurate algorithms for representing flood flows at large spatial scales. The need for the integration of these new tools into both risk calculation frameworks and meteorological–hydrological model cascades akin to those reviewed in this section is then outlined.

## 2.2 Hydraulic modelling

### 2.2.1 Introduction

Flood inundation modelling has enjoyed extraordinary advances in the first decades of this century. As with developments to climate and hydrology models, hydraulic models, too, are constrained by the trinity of data availability, process understanding and computational capacity. Fortunately, the Navier-Stokes equations and subsequent derivation of Saint-Venant equations describing the movement of water have been well defined since the early 1800s. While fluid dynamics in reality is obviously a 3D process, computational fluid dynamics is inhibited by data and computer availability and thus requires acceptable approximation in light of these constraints. Models based on full solutions to the 3D Navier-Stokes equations are possible (e.g. Younis, 1996; Stoesser et al., 2003; Horna-Munoz and Constantinescu, 2018), yet they may be (i) unnecessarily complex for the simulation of flood inundation, (ii) too computationally intractable to provide useful information over any meaningful spatial scale, and (iii) disproportionate to the quality of topography, bathymetry and boundary condition data available. ‘Acceptable approximation’ may be wrought from the assertion of Knight and Shiono (1996) that, for the case of floodplain inundation, in-channel flow is predominantly a 1D process and floodplain flow is clearly 2D, while 3D vortices in the shear layer between high-velocity in-channel flow and low-velocity floodplain flow facilitate substantial momentum transfer.

In the latter half of the 20$^{th}$ century, the only data available to build and test hydraulic models had limited spatial dimensionality: stream gauges measuring river flow and lateral surveyed cross-sections of floodplain terrain and channel bathymetry spaced many tens of kilometres apart. The ease with which such data could be integrated into 1D model structures facilitated

their ubiquity (Samuels, 1990). The hydraulics behind such models (e.g. ISIS, MIKE11 and HEC-RAS) are based on a simplification of the Navier-Stokes equations, the 1D Saint-Venant equations, which ensure conservation of mass (Eq. 2.1) and momentum (Eq. 2.2) between two cross-sections with a friction slope defined by (in this example) Manning's equation:

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = 0 \tag{2.1}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial \left[\frac{Q^2}{A}\right]}{\partial x} + gA\left(\frac{\partial h + \partial z}{\partial x} + \frac{Q^2 n^2}{R^{4/3}A^2}\right) = 0 \tag{2.2}$$

where $Q$ is discharge ($m^3 s^{-1}$), $t$ is time (s), $A$ is the flow cross-sectional area ($m^2$), $g$ is acceleration due to gravity ($m^2 s^{-1}$), $x$ is the distance between cross-sections (m), $h$ is water depth (m), $z$ is bed elevation (m), $n$ is the coefficient of friction ($m^{-1/3} s$), and $R$ is hydraulic radius (m).

This led, though, to highly uncertain models, where: (i) necessary interpolation between cross-sections to ameliorate the difference in dimensionality between model and reality incurred significant errors and, while conserving mass and momentum *at* each cross-section, these physical principles were violated *between* them during the interpolation of the water surface; and (ii) validation usually consisted only of determining a model's ability to replicate wave routing down a channel. This testing of bulk flow behaviour is not particularly discriminatory, meaning multiple model structures could fit available validation data equally well, yet look very different under alternative conditions. Such equifinality stunted model development, where any added complexity (or dimensionality) appeared futile (e.g. Bates et al., 1998).

The proliferation of remote sensing techniques enabled a step-change in hydraulic modelling capability, driving 1D/2D or 2D models with more granular data and providing validation data to more rigorously discriminate between different models. Principally, airborne laser altimeter data (lidar) permitted the construction of fine spatial resolution (2–5 m) elevation grids with up to 0.15 m vertical error (Marks and Bates, 2000; Bates et al., 2003). With an added dimension relative to laborious and sparse ground surveys used as input to previous

models, modellers sought codes capable of making use of these new spatially distributed data. These 2D models (some of which retained 1D representation of river channels) extended the 1D Saint-Venant equations to represent flow in both *x* and *y* Cartesian directions. These depth-averaged (the amplitude of a flood wave thought to be negligible compared to its length) Navier-Stokes equations, with Manning's equation representing energy loss, can be described as follows:

$$\frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} = 0 \tag{2.3}$$

$$\frac{\partial (hu)}{\partial t} + \frac{\partial (hu^2)}{\partial x} + \frac{\partial (huv)}{\partial y} + gh\left(\frac{h/2 + \partial z}{\partial x} + \frac{n^2 u\sqrt{u^2 + v^2}}{h^{4/3}}\right) = 0 \tag{2.4}$$

$$\frac{\partial (hv)}{\partial t} + \frac{\partial (huv)}{\partial x} + \frac{\partial (hv^2)}{\partial y} + gh\left(\frac{h/2 + \partial z}{\partial y} + \frac{n^2 v\sqrt{u^2 + v^2}}{h^{4/3}}\right) = 0 \tag{2.5}$$

where *u* and *v* are depth-averaged velocities (m $s^{-1}$) in the *x* and *y* directions respectively and all other terms are previously defined. With distributed models came distributed validation data: event-replicating models could be compared to images of their extent from the air (Bates and De Roo, 2000) or space (Horitt and Bates, 2002), providing a more robust testing procedure. The barrier to inter-model discrimination imposed by equifinality was lifted, as observations were of commensurate scale to the modelled processes. At this point, the field of flood inundation modelling started to develop rapidly. Over the forthcoming two decades, the expansion in area coverage, element granularity and vertical accuracy of these remote sensing techniques, coupled with enhanced computational capacity, permitted models of increasingly large domains and fine grid scales. The next section details this journey in scientific understanding through the lens of a diverse codebase that has been at the forefront of these developments – LISFLOOD-FP – from its inception as a simple modelling response to new remotely-sensed datasets of small stretches of river (Bates and De Roo, 2000), to its application in a model whose domain is virtually the entire planet (Sampson et al., 2015).

Other models are available (e.g. TUFLOW, JFLOW, HEC-RAS, TELEMAC, TRENT, SOBEK), yet LISFLOOD-FP is chosen for its pan-scale applicability and thus propensity to chart this century's developments to flood inundation modelling: from a 0.1 m resolution model of a 0.1 $km^2$ urban area (Sampson et al., 2012) to a 90 m resolution model of the globe (Sampson et al., 2015). Indeed, controlled intercomparisons of a wide variety of different 2D

models (Hunter et al., 2008) and different model complexities (Neal et al., 2012a) show them to yield very similar results for the case of simulating flood inundation. Thus, model choice is driven by other considerations. The computational efficiency of LISFLOOD-FP in its present-day local inertial form permits its diversity of application to sub-critical flows at very fine grid scales or over very large domains, unlike most other commercial and research 2D hydraulic codes which are confined to use at the local scale. The ease of integration of LISFLOOD-FP with available terrain and bathymetry information enables relatively simple model set-up, which is particularly crucial in its automated execution at large spatial scales. Lastly, the availability of the source code and proximity to technical support of the developers at the University of Bristol further justifies this model choice. Indeed, the results of previous 2D benchmarking studies (Hunter et al., 2008; Neal et al., 2012a) suggest that consistent conclusions should be obtainable with other hydraulic model codes.

### 2.2.2 LISFLOOD-FP

Bates and De Roo (2000) first presented LISFLOOD-FP, a storage cell flood inundation model, to take advantage of emerging distributed elevation data. Storage cell approaches, conceived by Zanobetti et al. (1970) in a model of the Mekong delta, were ubiquitous in the contemporary flood modelling community (Cunge et al., 1980; Estrela, 1994; Romanowicz et al., 1996). Terrain data would be discretized into large, irregular, polygonal units, with the node of each polygon commonly defined by a geodetic survey. The advent of remotely sensed topographic information permitted LISFLOOD-FP to employ regular elevation grids with element sizes much smaller than those of previous discretisations. In line with the assertions of Knight and Shiono (1996), channel flow was represented in 1D since both vertical and lateral variations in velocity are considered negligible, while floodplain flow had a 2D representation as lateral velocity variations became more important. Eqs. 2.1–2.5 were deemed overly complex and too computationally intensive in the Bates and De Roo (2000) iteration, so the governing equations neglected terms considered unimportant. 1D channel flow adopted the kinematic wave approximation, assuming the friction slope is equal to the bed slope: thus neglecting the advective inertia, local inertia and water slope terms. 2D floodplain flow used the diffusive approximation, where only the inertia terms are assumed insignificant. The Bates and De Roo (2000) paper, and subsequent tests to the formulation by Horritt and Bates (2001), came to the somewhat surprising conclusion that highly simplified models perform similarly to more complex ones yet outperform planar approximations. This

suggests that topographic representation constrains model accuracy to a greater extent than hydraulics, which are relatively simple in the context of flood inundation.

Inherent instabilities in numerical hydraulic models necessitated the selection of a time step to obtain stable solutions in the Bates and De Roo (2000) model. Its dependence on water depth, water slope, friction slope and grid resolution made the selection of an optimum time step – achieving maximum computational efficiency whilst maintaining stability – a non-trivial exercise. Too large a time step resulted in 'chequerboard' oscillations, where wet cells would entirely empty into a neighbour in one time step and revert to their original state in the next (Cunge et al., 1980). This instability would spread rapidly, making the simulation useless. Too small a time step would result in a computationally intractable simulation. Horritt and Bates (2001) thus imposed a flow limiter on the model to prevent too much water leaving a cell in a given time step (Eq. 2.6):

$$Q_x^{i,j} = \min\left(\frac{h_{\text{flow}}^{5/3}}{n}\left(\frac{h^{i-1,j} - h^{i,j}}{\Delta x}\right)^{\frac{1}{2}}\Delta y, \frac{\Delta x \Delta y\left(h^{i,j} - h^{i-1,j}\right)}{4\Delta t}\right) \quad (2.6)$$

where $Q_x^{i,j}$ is the flux between cells ($m^3\ s^{-1}$) at node ($i$,$j$), $\Delta x$ and $\Delta y$ are the grid size in each dimension respectively (m), $h$ is the water height (m), $h_{\text{flow}}$ is the depth between two cells that water can flow through (m), and $\Delta t$ is the time step (s). Making maximum volumetric flow rate a function of something other than Manning's equation (the first term in the flow limiter) generated model sensitivity to element size and time step rather than $n$ under these conditions. In other words, model output would look different with different grid scales and time steps but similar with different friction parameterizations.

Hunter et al. (2005) devised an adaptive time step solution to the lack of physical reality induced by the flow limiter in the next version of LISFLOOD-FP. Analogous to the Courant-Friedrichs-Lewy (CFL) condition (Courant et al., 1928) for advective schemes (Eq. 2.9, where $\alpha = 1$), Hunter et al. (2005) formulated a method to calculate the maximum stable time step at each increment in time (but constant for the whole domain) during the simulation of a diffusive wave:

$$\Delta t = \frac{\Delta x^2}{4} \min\left(\frac{2n}{h_{\text{flow}}^{5/3}}\left|\frac{\partial h}{\partial x}\right|^{\frac{1}{2}}, \frac{2n}{h_{\text{flow}}^{5/3}}\left|\frac{\partial h}{\partial y}\right|^{\frac{1}{2}}\right) \quad (2.7)$$

This new optimal time step calculator ensured results were invariant to different grid sizes and (initial) time steps, while maintaining intuitively correct floodplain friction sensitivity.

The Hunter et al. (2005) model, though, became problematic as grid resolutions became ever finer as a response to increased computational capacity and data quality. Despite the more realistic simulations generated by the adaptive time step model, Hunter et al. (2006) found this increased run times relative to the flow limited model by a factor of six. With a power law dependency on model grid size in Eq. 2.7, increasingly granular grids would pose a computationally intractable problem. The Neal et al. (2009a) model of a small city using this formulation had to be run at 25 m resolution to keep run times below even two days. Meanwhile, Fewtrell et al. (2008) identified grid scales of < 10 m to be appropriate for urban inundation modelling: a task that the Hunter et al. (2005) formulation performed more slowly than even full solutions to the Saint-Venant equations (also evidenced by Hunter et al., 2008). Fine-scale modelling was, therefore, computationally prohibitive with the Hunter et al. (2005) adaptive time step version yet physically unrealistic with the Horritt and Bates (2001) flow limited model.

Bates et al. (2010) found a surprising solution: to make the governing equations more complex, rather than simplify them yet further as would be the intuitive answer to reducing computational cost. This new "inertial" formulation preserved the local inertia term in the Saint-Venant equations unlike the diffusive formulation, though still neglected convective inertia; an acceptable assumption for sub-critical flows. Thus, Bates et al. (2010) derived Eq. 2.8, describing flow per unit width $q$ ($m^2\ s^{-1}$) in an assumed rectangular channel at the time step subsequent to time $t$ (s):

$$q_{t+\Delta t} = \frac{q_t - g h_t \Delta t \frac{\partial (h_t + z)}{\partial x}}{1 + \frac{g \Delta t n^2 |q_t|}{h_t^{7/3}}} \quad (2.8)$$

In including local inertia, modelled water now had mass (unlike Eq. 2.6) and so was less likely to produce chequerboard oscillations. As Eq. 2.8 is now a true shallow water model, the time step was estimated by a derivation of the CFL condition as a function of wave velocity and grid size:

$$\Delta t_{t+\Delta t} = \alpha \frac{\Delta x}{(g h_t)^{1/2}} \tag{2.9}$$

where α is a dimensionless coefficient which adjusts the CFL condition in circumstances where the assumption of small wave amplitude is invalid. For most floodplain flows, $0.2 \leq \alpha \leq 0.7$. The Bates et al. (2010) formulation generates time steps that are typically 1–3 orders of magnitude larger than the adaptive time step formulation of Hunter et al. (2005), as the largest stable time step scales linearly with grid size (Eq. 2.9) rather than quadratically (Eq. 2.7).

de Almeida et al. (2012) adapted the Bates et al. (2010) formulation to improve stability. They noted that errors arising from discretisation contain no diffusive terms, propagating instabilities particularly when *n* is low. This is because stability is wrought from increasing *n* in the denominator of Eq. 2.8 forcing flow to zero. At low friction values, this stabilising effect is reduced. de Almeida et al. (2012) introduced a 'q-centred' scheme which adds diffusive terms to the calculation of $\delta q/\delta t$. The Bates et al. (2010) approximation of derivatives of the discretised momentum equation (Eqs. 2.2, 2.4, 2.5) at cell interfaces was extended so that flow at cell centre *i* at the time step subsequent to *t* becomes a weighted average of flow across the interfaces of the preceding and proceeding neighbours at time *t*:

$$q_{t+\Delta t}^{i-0.5} = \frac{\left[\theta q_t^{i-0.5} + \frac{1-\theta}{2}\left(q_t^{i-1.5} + q_t^{i+0.5}\right)\right] - g h_t \Delta t \frac{\partial (h_t + z)}{\partial x}}{1 + \frac{g \Delta t n^2 |q_t^{i-0.5}|}{h_t^{7/3}}} \tag{2.10}$$

where $\theta$ is the weighting factor. Note that $q_{t+\Delta t}^{i-0.5}$ in Eq. 2.10 is the same as $q_{t+\Delta t}$ in Eq. 2.8, but further detailing that flow is calculated at the interface between cell *i* and *i* – 1. Negligible increases in computational cost relative to the Bates et al. (2010) solution were more than offset by the increase to largest stable time step. With enhanced stability at lower friction

values offered by de Almeida et al. (2012) and simulation at fine grid resolutions made computationally tractable by Bates et al. (2010), the potential for LISFLOOD-FP to simulate urban flooding was unlocked: providing rapid, stable and accurate solutions at high resolution (e.g. Neal et al., 2011; Fewtrell et al., 2011a; Sampson et al., 2012).

Sampson et al. (2013) developed the framework further to enable rainfall, rather than river flow, input to the model domain via a routing scheme. Not all flood hazard will arise from floodwaters overtopping river banks; in many cases the direct effect of rain falling onto the land surface is the flood driver (e.g. intense convective storms, tropical cyclones). Simulating this phenomenon exclusively with some derivation of the shallow water equations (e.g. Eq. 2.8) poses a number of problems: the large-wavelength, small-amplitude flood wave assumption underpinning such formulations breaks down in urban pluvial flood scenarios amidst topographic discontinuities and the number of wet cells generated by rainfall simulation will increase the compute load substantially. The large number of very shallow, non-hazardous water depths arising at $t$ following a rainfall input are therefore handled by a routing scheme. This involves the translation of the elevation data into a flow-direction grid, with water below a specified depth threshold routed to the downslope neighbour at a specified velocity. Once the depth threshold is exceeded (i.e. when the very shallow waters eventually accumulate), the normal shallow water equations initiate.

Concurrent to scientific advances in the hydraulic algorithms, ensuring LISFLOOD-FP can be implemented in parallel has commanded considerable effort. Neal et al. (2009b) translated LISFLOOD-FP to a multi-threaded process, ensuring simulation execution on the full complement of cores available to a CPU. Their 8-core simulations were >5x faster than when run as a single-threaded process: ~90% of the theoretical speed-up available according to Amdahl's Law. Neal et al. (2010) enabled parallelisation of LISFLOOD-FP using message passing between multiple CPUs with distributed memory and Neal et al. (2018) documented efficiencies relating to vectorisation of code. Not only have these algorithmic and parallel improvements enabled very fine-scale models to proliferate (e.g. the 10 cm resolution model of Sampson et al., 2012), they have also permitted the simulation of extremely large model domains.

The latter case begot further modelling challenges, with one such being that the width of small channels was often narrower than the grid scale of the model. This led Neal et al.

(2012b) to formulate the sub-grid channel scheme, where channels can be parameterised within the 2D model so that both channel and floodplain dynamics can be simulated at large spatial scales. Reformulating the Bates et al. (2010) equation (Eq. 2.8) to describe sub-grid channel flow, Neal et al. (2012b) made channel flow a function of hydraulic radius and flow area at time *t* rather than flow depth and cell width:

$$Q_{t+\Delta t}^{\mathrm{ch}} = \frac{Q_t^{\mathrm{ch}} - gA_t^{\mathrm{ch,flow}}\Delta t S_t^{\mathrm{ch}}}{1 + \frac{g\Delta t (n^{\mathrm{ch}})^2 |Q_t^{\mathrm{ch}}|}{\left(R_t^{\mathrm{ch,flow}}\right)^{4/3} A_t^{\mathrm{ch,flow}}}} \tag{2.11}$$

where *R* is the hydraulic radius, *A* is area of flow, *S* is water surface slope, and the superscript ch refers to the sub-grid channel. When bankfull stage is exceeded, an adaptation of Eq 2.8 initiates to describe floodplain flow at a 2D cell occupied by a sub-grid channel:

$$Q_{t+\Delta t}^{\mathrm{fp}} = \frac{q_t^{\mathrm{fp}} - gh_t^{\mathrm{fp}}\Delta t S_t^{\mathrm{fp}}}{1 + \frac{g\Delta t (n^{\mathrm{fp}})^2 |q_t^{\mathrm{fp}}|}{\left(h_t^{\mathrm{fp,flow}}\right)^{7/3}}} \left(\Delta x - w_t^{\mathrm{ch,flow}}\right) \tag{2.12}$$

where *w* is width and the superscript fp refers to the 2D floodplain. The difference between Eq. 2.8 (Bates et al., 2010) and Eq. 2.12 (Neal et al., 2012b) is flow width is the difference between cell width and channel width and water depth *h* is represented as depth above the sub-grid channel bank. Neal et al. (2012b) tested this formulation (1D all rivers; 2D floodplain) at 1 km grid resolution against a traditional 1D/2D model (1D larger, supra-grid rivers; 2D floodplain), a floodplain-only model (2D floodplain) and a channel-only model (1D all rivers). The test bed was an 800 km stretch of the River Niger, with observational benchmarks derived from ICESat observations. Both channel-only and floodplain-only simulations had no predictive skill. The connectivity provided by the small rivers in the sub-grid scheme resulted in increased accuracy (in the terms of inundation extent, water surface elevation and wave celerity) relative to the 1D/2D supra-grid analogue. Neal et al. (2012b) thus demonstrated the importance of full channel representation in hydraulic models, offering a method for their inclusion amidst data scarcity.

Thus, both many-element/fine-scale/local-area (e.g. Sampson et al. (2012): $10^{-1}$ m resolution of a $10^{-1}$ km$^2$ area of Alcester, UK) and many-element/coarse-scale/wide-area (e.g. Neal et al. (2012b): $10^3$ m resolution of a $10^5$ km$^2$ area of the Niger basin) models have found application within the LISFLOOD-FP framework. Simplifications have been made to fundamental hydraulic formulae, whilst retaining accurate solutions for the purpose of flood inundation modelling. Improvements to stability, accuracy and efficiency at fine grid scales and low frictions, as well as a runoff routing scheme, have ensured applicability of LISFLOOD-FP to urban flood simulation. Neal et al. (2012b) showed how LISFLOOD-FP can be applied to large areas, introducing algorithmic adaptations to ensure necessary grid coarsening does not impair predictive capability. This effort has culminated in the execution of LISFLOOD-FP at the global scale, simulating design flood events on virtually every river on Earth (Sampson et al., 2015). It is here that challenges cease to be related to specific algorithmic improvements to LISFLOOD-FP and instead are a function of data availability. Urban and large-scale models have distinct and separate data needs. Urban models are generally built in the context of data-richness, with surveyed river bathymetry, highly accurate and finely resolved airborne (or even terrestrial) lidar elevation data, local knowledge of drainage systems, well-constrained boundary conditions and much else. This thesis, though, focuses on expansion of model domain rather than refinements to element size. At large spatial scales, the data afforded to local analyses are not available (at least seamlessly). This data scarcity has not, however, been a barrier to enhancing understanding of flood hazard at global scales: it has instead placed a new set of demands on hydraulic modellers whose considered domains are global, with greater weight put upon data approximations than algorithmic developments. These demands, relative to the local-scale analogue, are set out in the proceeding section.

### 2.2.3 Hydraulic modelling over large spatial domains

#### 2.2.3.1 Introduction

Flood inundation modelling has historically been a reach-scale endeavour, but models of almost the entire world have now been built and executed (Yamazaki et al., 2011; Winsemius et al., 2013; Ward et al., 2013; Pappenberger et al., 2012; Dottori et al., 2016; Sampson et al., 2015). Increased computational capacity and the development of fast hydraulic codes set out in the previous section only explain a portion of the scale-up in spatial domain seen in recent years. The other challenges large-scale modellers work to address in facilitating this are often different to their local-scale counterparts. These are set out in this section.

#### 2.2.3.2 Elevation data

An accurate representation of topography in a hydraulic model has been shown to be the dominant control on the dynamics of flood inundation (Horritt and Bates, 2002; Bates, 2012; Neal et al., 2011). High-quality local flood models are usually based on a Digital Elevation Model (DEM) derived from lidar data. Local modellers further correct these data to produce a Digital Terrain Model (DTM) by removing vegetation and buildings. While their frequencies (5 KHz – 166 KHz), vertical accuracy (0.15 – 0.05 m) and spatial resolution (5 – 0.25 m) have improved in the first decade of this century (Marks and Bates, 2000; Bates, 2012), the impact on large-scale modelling has been low. Seamless, wide-area, corrected, lidar-based DTMs are only sparsely available for use in large-scale hydraulic models: in parts of North America, Europe and Australia. Large-scale modellers must instead turn to spaceborne radar data, with greater availability over large areas, while continuing calls for industry, governments and NGOs to support the development of a global lidar-based DEM (Schumann, 2014; Winsemius et al., 2019). The most popular source of satellite-derived elevation data is from the Shuttle Radar Topography Mission (SRTM), primarily owing to its higher accuracy and accessibility compared to other datasets (Yan et al., 2015). An entirely new field of enquiry was conceived to make SRTM suitable for flood modelling; correcting errors due to voids, speckle, noise, striping, vegetation and buildings (e.g. Lehner et al., 2008; Baugh et al., 2013; Elvidge et al., 2007; Gallant, 2011; O'Loughlin et al., 2016; Yamazaki et al., 2012). Despite this, significant elevation errors still remain, and research into hydraulic modelling amidst such uncertainty continues. Yamazaki et al. (2017) produced an SRTM-based corrected product called Multi-Error-Removed Improved-Terrain (MERIT) DEM at 3 arc sec (~90 m) resolution. Despite their novel iterative method in combining multiple incomplete-coverage, higher-accuracy DEMs (e.g. AW3D, VFP and ICESat), correcting systematic and absolute biases and removing tree heights, 42% of land areas had vertical errors exceeding 2 m. Hawker et al. (2018) propose a probabilistic approach, fitting semivariograms to SRTM data to produce multiple plausible DEMs, necessitating multiple model runs to explore the uncertainty induced by vertical errors. At the global scale, however, this poses a computationally intractable solution (given a single return period run of the Sampson et al. (2015) global model requires ~200,000 CPU hours of processing time). Ultimately, even corrected SRTM-based models will hit an accuracy ceiling lower than that imposed by other facets of global models (i.e. SRTM is the limiting factor on model performance) because of their insoluble vertical errors, coarse (~90 m) resolution and decay due to age (approaching

two decades). New data sources may provide a solution in the short–medium term. Archer et al. (2018) demonstrate promise with the use of a TanDEM-X-based product (Wessel et al., 2018) for a test case in Fiji, outperforming SRTM and MERIT DEMs when compared to hydraulic models run with lidar data. Similarly, should the photogrammetric techniques underpinning ArcticDEM be replicated globally, lidar-like accuracy models may be possible based on such data: as demonstrated by Taylor et al. (2019) for a test case in Finland.

#### 2.2.3.3 Channel geometry and hydrography

Neal et al. (2012b) and Fewtrell et al. (2011b) demonstrate the necessity of channel representation in models for producing behavioural simulations of flood events in terms of depth, extent and dynamics. Their raw representation in even lidar-based DEMs is inadequate, since conventional red-laser lidar does not penetrate water bodies. Local-scale models, then, use surveyed river bathymetry to inform bed elevation, manual digitisation of channel locations and a DEM of suitable resolution to identify bank heights. Such detailed, *in situ* information is not available at large scales, requiring large-scale modellers to approximate channel hydrography and geometry. Lehner et al. (2008) processed SRTM elevation data to derive river networks based on flow accumulation. This was updated by Yamazaki et al. (2019) who produced a version of the same data based on the new MERIT DEM. These elevation-based hydrography derivations are subject to considerable uncertainty, since they are only as accurate as the underlying DEM. Deviations from true channel location are generally more pronounced in very flat topographies, for heavily engineered streams and, due to coarse grid scales, for smaller river channels. A number of studies process global satellite imagery to produce river width data (e.g. Andreadis et al., 2013; Allen and Pavelsky, 2018; Yamazaki et al., 2014). Allen and Pavelsky (2018) apply the RivWidth tool (Pavelsky and Smith, 2008) to worldwide Landsat scenes to generate global river width data. These, too, are inhibited by underlying data quality: Allen and Pavelsky (2018) consider detected rivers with width < 90 m to be unreliably detected and measured. River bed elevations are not presently remotely detectable, so this feature of large-scale models is commonly approximated based on hydraulic geometry theory (Leopold and Maddock, 1953). This involves relating bed elevation to different hydro- and geomorphological parameters such as width, upstream area, bankfull discharge and velocity; many of which can be informed by global databases of these quantities (e.g. Frasson et al., 2019; Allen and Pavelsky, 2018). However, some of these variables require approximations themselves: bankfull discharge is unknown for unmonitored rivers, meaning many models approximate this to be a certain

return period flow calculated by their flow characterisation (see next section) tool (e.g. 1 in 1.5–2 year: Sampson et al., 2015; Winsemius et al., 2013). From this, channel bed elevation can be derived from, for example, Manning's equation. In many circumstances though, such approximations break down (Williams, 1978): the ignorance of geography in the bankfull return period assumption means every river flows out of bank at the same frequency. Overall, channel representation in large-scale hydraulic models is relatively poor: from something as seemingly simple as identifying river locations to the more complex task of representing their bathymetry, these remain fundamentally limiting features of contemporary modelling approaches.

#### 2.2.3.4 Flow characterisation

All flood inundation models require a flow input, the nature of which depends on the purpose of the model. They can be historical models, where past events are simulated, or design flood models, simulating potential events of a certain probability. Boundary condition uncertainty (of real-world conditions) is constrained by (i) errors in the raw observations (*measurement error*), (ii) approximations in non-measured areas (*errors in space*), and (iii) uncertainty in the statistical tools ascribing probabilities to flows (*errors in time*). Historical models are generally less uncertain than design ones, since they are not subject to error imposed by (iii). They often are not subject to (ii) either, with real event observation of flows being incorporated into the model when built amidst data-richness, as is common. Large-scale models, though, (and certainly global ones) tend to be design scenario based: delineating areas of a particular annual chance of flooding.

Observations of flow (i) are the most accurate data source to drive a hydraulic model, yet these are not error-free. They typically consist of directly measured water levels at a river gauging station, with a so-called 'rating curve' applied to translate this stage to a discharge. While water levels can be measured very accurately (<0.02 m; McMillan et al. 2012), stage–discharge relationships introduce errors to the resultant discharge quantification. These can be substantial (~40%), particularly for extreme flows when the observation station itself may be destroyed (Di Baldassarre and Montanari, 2009; Coxon et al., 2015; McMillan et al., 2012).

Such data are only available sparsely, though. With no measurements of flow on most of the world's rivers and a marked decline in monitoring stations since the 1980s (Vörösmarty et al., 2001; Fekete and Vörösmarty, 2007), approximations (ii) are made via two methods:

hydrological modelling and regional flood frequency analysis (RFFA). The uncertainties relating to the former were mentioned in section 2.1.2, illustrating their problematic use in large-scale hydraulic models without a large volume of distributed data to calibrate to or parameterise with (Blöschl et al., 2013). Furthermore, the precipitation input to hydrological models themselves is highly uncertain. Observational rainfall products of varying source are poorly correlated, with spatially variable biases and a tendency to underestimate extremes (McMillan et al., 2012; Dee et al., 2011; Chen et al., 2014; Kidd et al., 2012). Sampson et al. (2014) calculated flood risk in Dublin (Ireland) using satellite, reanalysis, radar and gauged rainfall data separately as input to a catastrophe model and found the resultant losses differed by an order of magnitude. Despite this, most global flood models use climate reanalysis data cascaded through a hydrological model to generate river flow boundary conditions (Yamazaki et al., 2011; Winsemius et al., 2013; Ward et al., 2013; Pappenberger et al., 2012; Dottori et al., 2016).

The alternative, favoured by Sampson et al. (2015), is a global RFFA based on an index flood method (Smith et al., 2015). This approach is, however, not applicable for 'real' event simulation (e.g. historical floods, forecasts, future scenarios), but for design inundation models which simulate floods of certain probabilities. This approach inherits uncertainties relating to (iii), outlined afterwards. The principle behind RFFA is that measured flows can be related to widely observed catchment characteristics (e.g. mean annual rainfall, drainage area, slope, climatology), and so such descriptors can be used to estimate flow on ungauged streams. These approaches have seen fairly common usage at the regional scale (e.g. Salinas et al., 2013; Ouarda et al., 2006; Merz and Blöschl, 2009) and have been extended to the global scale by Smith et al. (2015). They relate catchment descriptors of mean annual rainfall, drainage area and climate classification to over 3000 gauging stations from the Global Runoff Data Centre (GRDC), permitting any AEP discharge to be estimated given these predictors. Median errors in estimating the 1% AEP discharge were reportedly 56%. In its application to the Sampson et al. (2015) global model, river geometry is coupled to biases in the global RFFA, meaning these errors are dampened somewhat by adjusting channel conveyance accordingly. The RFFA approach is considerably more computationally efficient and less data-hungry than use of a hydrological model, though lacks the benefit of total-coverage data input in the form of satellite-measured precipitation. Some promising alternative methods of estimating of discharge on ungauged streams are on the horizon through measuring flow from

space, particularly when the NASA Surface Water and Ocean Topography satellite mission launches in 2021 (e.g. Yoon et al., 2012; Pavelsky et al, 2014; Gleason and Smith, 2014).

For quantifying design discharges, their AEP is determined through fitting statistical distributions to flow data (iii). Whether flows are generated from cascading rainfall observations through a runoff model or directly from river gauges, each subject to separate uncertainties, this statistical analysis is fundamentally impaired by non-stationarity and record length. AEP flows of the present-day are a function of historical data and are thus underpinned by the assumption that flow behaviour across time has remained stationary. Changes to land-use, climate and river morphology (anthropogenic or otherwise) mean that historical flow data may no longer be relevant in gaining understanding of present-day flows. Techniques for correcting this do exist, but their application at large spatial scales is inhibited by their manual operation and scarcity of data on the drivers of non-stationarity (Pinter et al., 2001; Villarini et al., 2009b; Xiong et al., 2015). The most fundamentally limiting factor on the characterisation of AEP flows is the length of records available. The GRDC has only 721 gauges with >80 years record length; most are ~30 years (Do et al., 2017). Estimating the, say, 1 in 100 year flow when discharge has not been monitored for ≥100 years therefore requires extrapolation, inducing significant uncertainty. Without many more centuries of observations, this source of error will not diminish.

The consideration of rain and river boundary conditions in isolation may, in many circumstances, lead to a misestimation of their magnitude and the resultant flood extent. While current inland large-scale flood models set their coastal boundaries to 0 m elevation (local mean sea level), Ward et al. (2018) find that there is significant dependence between extreme river and coastal water levels for 36% of estuarine river gauges globally. These compound extremes exacerbate flooding: extreme near-coastal fluvial floods will not always freely drain into the ocean, but back up against concurrently high sea levels (as demonstrated by the flood response to Hurricane Harvey and Cyclone Idai; Zscheischler et al., 2018). Current large-scale approaches do not account for this, and so near-coastal fluvial floods may be underestimated by these models. Couasnon et al. (2019) propose the coupling of coastal level and fluvial discharge time series to produce a bivariate dependence model from which plausible compound events can be simulated. Other large-scale models will need to follow this blueprint to prevent misestimation of inland flood hazard in deltas and estuaries.

#### 2.2.3.5 Calibration and validation

Calibration involves repeatedly running a model with unknown or uncertain parameters varied until some degree of agreement with benchmark data is found. Commonly, the friction parameter is varied (e.g. Aronica et al., 2002; Neal et al., 2009a; Bates et al., 2006; Horritt and Bates, 2001; Werner et al., 2005a; Pappenberger et al., 2005a) as this is only crudely estimated *a priori* based on empirically derived look-up tables (e.g. Chow, 1959, Arcement and Schneider, 1989) or vegetation heights derived from the difference between the raw lidar Digital Surface Model and the bare-earth DTM (e.g. Mason et al., 2003; Straatsma and Baptist, 2008). Theoretically, any facet of a hydraulic model can be calibrated, with examples including: poorly constrained bed elevation where manual surveys have not been carried out (Neal et al., 2012b; Schumann et al., 2013), uncertain stage–discharge relationships to generate flow input (Domeneghetti et al., 2012), and the DEM when based on highly uncertain SRTM derivatives (Hawker et al., 2018). The data a modeller wish their simulations to resemble – the benchmark data – can come in many forms, commonly: airborne or satellite synthetic aperture radar or photogrammetric images of flood extent (Aronica et al., 2002; Horritt and Bates, 2002; Werner et al., 2005a; Di Baldassarre et al., 2009) to facilitate a binary pattern comparison, or an extension which intersects the binary images with a DEM to yield water levels (Bates et al., 2006; Stephens and Bates, 2015); surveyed high water and wrack marks providing distributed water level measurements (Neal et al., 2009a); point water level data measured by a stream gauge (Neal et al., 2009a; Bates et al., 1998; Neal et al., 2012b); or directly observed water levels from laser altimeters (Schumann et al., 2013; Neal et al., 2012b). The calibration process therefore requires capacity for multiple model runs and collation of suitable benchmark data. It is therefore commonplace in local-scale models, where confined domains permit a more straightforward search for and processing of calibration data and their small spatial scales allow multiple model runs to explore the uncertain parameter space. At large, and certainly global, scales, a sufficient number of simulations to explore parameter uncertainty is often computationally intractable and the identification of appropriate calibration data is difficult. In many cases, the latter is impossible because large-scale model structures usually simulate something unobservable in reality: design discharges – simulated everywhere simultaneously – are not tangible phenomena that can be benchmarked. Indeed, one of the reasons these large-scale models exist is because widespread observations of flooding are lacking; so even if calibration was a routine procedure at these spatial scales, it would involve overfitting to sparse observations.

Validation is a similar process to calibration, but rather than maximising a test statistic with multiple model runs, a single model run is compared to the benchmark data to indicate model performance (or multiple runs to validate model ensembles). For the local modelling approaches outlined previously, calibration and validation are the same thing: the model run with the most favourable metrics is used to illustrate model performance unless the two procedures are specifically distinct (i.e. there are separate calibration and validation data). This split-sample validation, where independent benchmark data (if available) are compared to the calibrated optimum model run, tests whether the identified optimum parameter set is stationary between events. For the same reasons large-scale models are not calibrated, their validation is deficient also. Since their purpose is not to recreate an observable event, what little validation is done usually involves comparison to local-scale models also used to simulate design events. Assuming the local model has received extensive attention from skilled hydraulicians, with all facets of the model built with highly accurate local data, tested against available observations to ensure behavioural simulations, this is a broadly defensible approach to ensuring a larger-scale model generates output that seems sensible. By their very nature, local-scale studies are not available at a spatial scale commensurate to global assessments: hence, previous local–global model intercomparisons do not suitably scrutinise model output. Dottori et al. (2016), Winsemius et al. (2016), Sampson et al. (2015) and Alfieri et al. (2014) tested their respective large-scale flood models (all are global scale except the European model of Alfieri et al., 2014) against local-scale models of four European river basins. Winsemius et al. (2016) found their model had more incorrect than correct pixels; Alfieri et al. (2014) and Dottori et al. (2016) found broadly similar numbers of correct and incorrect pixels in the comparison; and Sampson et al. (2015) found correct pixels constituted roughly half to two-thirds of considered pixels. Dottori et al. (2016) extended their validation to satellite observations of flooding in several river basins in Asia, Africa and South America. They created cumulative maps of inundation extent in the period 2000–2013 from MODIS imagery, and re-ran their global model with maximum flows from that period (rather than derived AEP flows) for the comparison. They found their model had more incorrect than correct pixels. It should be noted that compilation of global flood extents by MODIS is subject to considerable uncertainty, where cloud cover (common during flood events) and coarse resolution imagery inhibit reliable identification of flooded areas (Schumann and Moller, 2015). Ward et al. (2017) also built on the validation of GLOFRIS performed by Winsemius et al. (2013, 2016) to six local-scale models and two events with an

assigned AEP. For these isolated test cases, the global model agreed with the benchmark data roughly as often as it disagreed.

There are a number of different metrics available to compare models to benchmarks, which depend upon the nature of the benchmark data and the aspect of model performance considered. Model errors with respect to observed water levels are commonly summarised by some central tendency (e.g. mean) to indicate bias and a central tendency of absolute errors (e.g. root mean squared error) or standard deviation to more faithfully constrain model performance where positive and negative errors do not cancel out. Typical reach-scale hydrodynamic models are calibrated to obtain average absolute errors of <0.4 m versus surveyed high water marks, imagery of flood extents intersected with a DEM or point hydrometric data (e.g. Mignot et al., 2006; Matgen et al., 2007; Apel et al., 2009; Fewtrell et al., 2011b; Rudorff et al., 2014; Bermúdez et al., 2017; Altenau et al., 2017; Adams et al., 2018; Fleischmann et al., 2019). Indeed, typical errors in surveyed high water marks are 0.3–0.5 m, suggesting model errors smaller than these would be overfit to observations (Schumann et al., 2007; Neal et al., 2009a; Horritt et al., 2010; Fewtrell et al., 2011b). For the case of binary pattern (wet/dry) comparisons, different metrics designed to emphasise different aspects of model performance can be maximised through calibration. The balance of overprediction and underprediction indicates bias; underprediction can be exclusively penalised to generate a hit rate; overprediction can be exclusively penalised to generate a false alarm rate; or an overall fit metric can illustrate performance where both types of model error are similarly penalised. Commonly applied in binary pattern comparisons, the Critical Success Index (CSI) divides correctly identified wet pixels by all relevant pixels (correct wet, incorrect wet, incorrect dry), providing an optimum score of 1. While ubiquitous, Stephens et al. (2014) point out that CSIs slightly inflate the scores of overpredictive models of large floods on flat terrain compared to the inverse case. Thus, flooding in headwater areas may not achieve a comparable CSI to flooding on large rivers despite having similar water level errors. In the wider literature, local hydrodynamic models are generally calibrated to 0.7–08 when compared to air- or spaceborne observations of flood extent (e.g. Aronica et al., 2002; Horritt and Bates, 2002; Pappenberger et al., 2007a; Di Baldassarre et al., 2009; Wood et al., 2016). Indeed, Horritt et al. (2001) compare spaceborne radar observations of flood extent to aerial imagery of the same flood, finding roughly 75% agreement between the two data sets. With validation data of this quality, obtaining CSIs beyond 0.7–0.8 would constitute overfitting to the observations. When validating against very high quality observational data,

CSIs of up to 0.9 are possible (Bates et al., 2006; Altenau et al., 2017). Fleischmann et al. (2019) posit that a hydrodynamic model achieves locally relevant results with CSIs greater than 0.65, where roughly two-thirds of relevant pixels are identified correctly.

Large-scale models, then, are apparently unable to achieve local relevance, though the scale and nature of the validation procedures performed to date which underpin this remark are far from comprehensive. A key unknown question is whether, when local model prediction errors are aggregated to large scales, these errors compound or cancel to produce a model which exacerbates or dampens local-scale errors. Furthermore, Trigg et al. (2016) compared six global models – not to benchmark data, but to each other – and found they disagree in more areas than they agree for given AEP flood extents at the continental scale (Africa). Thus, not only do these large-scale models fail to replicate benchmarks, they do not seem to do so in a consistent way. The culmination of limited validation studies to date indicate a clear need for large-scale models to be benchmarked against data of a commensurate scale.

The need for these structures to become event replicating, so that they can be validated against real-world observations, is also critical. The admirable efforts of Ward et al. (2017) and Bernhofen et al. (2018) on this front are impaired by their comparison of real event data to non-event replicating models, imposing considerable uncertainty in the AEP assignment of the given events and the unlikely assumption that this would be static across space. While the limited studies to date indicate large-scale models driven with gauge-based flows (thus, no hydrology component) and 2D hydrodynamics are more accurate than those using climate reanalysis and simplified hydraulics, wide-area event-replication will ultimately require coupled hydrology–hydraulics at large-scale. Not only will this provide a way of fairly validating against observations rather than models, it will enable large-scale flood forecasts detailing potential inundation: an often-neglected component of forecast models. The exception to this is the forecast model of the Zambezi River presented by Schumann et al. (2013), though this model achieved high accuracy (0.27 m average error when compared to water levels) through calibration to observed water levels and flood extents; not possible in a true forecast where the event has not yet occurred.

One final observation on this topic is that large-scale model validation has not yet been applied in conjunction with simpler models. The presentation of new models is often accompanied by a demonstration that simpler alternatives can be outperformed. Bates and De

Roo (2000), when LISFLOOD-FP was first introduced, found higher flood extent observation correspondence with their new model than with a planar lid fitted to water levels. Carried out separately, there is sporadic evidence that the global LISFLOOD-FP based model of Sampson et al. (2015) is more accurate than the non-momentum conserving GLOFRIS (see results of Ward et al., 2017), but even simpler approaches are being posited with inadequate scrutiny. Height Above Nearest Drainage (HAND), akin to a planar approximation and thus employing no representation of flow physics at all (e.g. Afshari et al., 2018; Liu et al., 2018), and approaches based on geomorphic descriptors (e.g. Jafarzadegan et al., 2018; Samela et al., 2016; Nardi et al., 2019) have gained in popularity due to their "quick-and-dirty" solutions to computationally intensive problems. Though shown extensively at local scales, it remains to be seen if large-scale hydraulic models outperform these low-complexity approaches.

#### 2.2.3.6 Flood defences

An accurate representation of flood defences in large-scale models is extremely challenging. Complete databases on the locations and standards of flood protection structures are available only for a handful of countries and such features are undetectable in the spaceborne radar DEMs employed in global models. Indeed, papers where global hydraulic modelling methodologies are presented highlight the need for flood defence information to be a research priority (e.g. Dottori et al., 2016; Sampson et al., 2015; Ward et al., 2015). Recently, a potential solution was proposed by Scussolini et al. (2016): FLOPROS. They developed a global database of protection levels (as an AEP) based on engineering design information, theoretical standards based on policies and a model based on wealth. These data were then applied in further global flood risk analyses, claiming flood defences were now accounted for to some degree in their models (Winsemius et al., 2016; Alfieri et al., 2017). However, while FLOPROS calculates protection standards, it makes no statement on their location other than an implicit assumption that defences have been built on every river. The stated standards are only likely to be on major rivers in urban centres; even these will likely vary within the sub-country units considered by Scussolini et al. (2016). Furthermore, the modelled information which extends flood protection standards to areas where design and policy information is lacking is validated only in the sense that it is internally consistent with the data that drive it: the relationship between defence standard and wealth is presently inadequately evidenced. FLOPROS is therefore used by recent global risk studies to stipulate that virtually all frequent (or even infrequent) flooding generates no damage (e.g. depending on the FLOPROS layer

used, Alfieri et al. (2017) assume no flood that occurs more frequently than once every 500 years generates damage for nearly all of the US). Thus, these estimates of flood risk are heavily sensitive to such data that are highly uncertain. The clear requirement for further research into defence representation still remains.

#### 2.2.3.7 Discussion

The proliferation of large-scale hydraulic models has been a function of increased data availability, computational capacity and efficiency of hydraulic codes. Particularly, the new local inertial formulation of the shallow water equations (Bates et al., 2012; de Almeida et al., 2012) and the sub-grid channel implementation (Neal et al., 2012b) enabled Sampson et al. (2015) to build a global hydraulic model based on LISFLOOD-FP. By virtue of its more physically based 2D hydraulic engine, simulation of all river channels with drainage area >50 $km^2$, pluvial model component, extensively corrected, SRTM-based, high-resolution DEM, and flow inputs based directly on river gauges, the Sampson et al. (2015) model seemingly represents the most accurate of academic global models. These other models can be characterised through their possession of some or all of the following features: employing rainfall reanalysis data cascaded through a hydrological model, having low representation of physics in river routing and no representation of floodplain flow, solving governing equations and outputting flood maps at coarse resolution, and simulating flooding only on large rivers (Yamazaki et al., 2011; Winsemius et al., 2013; Ward et al., 2013; Pappenberger et al., 2012; Dottori et al., 2016; Trigg et al., 2016). This assertion is only implicit, though; based on sporadic and incomprehensive validation studies. These model structures require more thorough testing for their output to be trusted in their many and varied potential use cases: for use in insurance pricing, zoning regulations, development planning and emergency response. This review has also highlighted the difficult and laborious challenges that lie ahead for large-scale modellers, predominantly: resolving DEM inaccuracies, improving river channel representation and compiling accurate flood defence information.

The inaccuracies of global models evidenced (albeit for a handful of test cases) in this review have not inhibited their application in risk quantification. Winsemius et al. (2016) calculate current flood risk to be equivalent to $1 trillion in flood damages each year or $160 billion when using the FLOPROS defence database with their volume spreading algorithm. Alfieri et al. (2017) place annual flood losses at $66 billion when considering FLOPROS alongside their 2D semi-inertial hydraulic model (Dottori and Todini, 2011). Winsemius et al. (2016)

and Alfieri et al. (2017) also project changes to flood risk under climate change scenarios, conflating the uncertainty inherent in such an endeavour (section 2.1.2) with those considerable hydraulic modelling uncertainties (section 2.2.3). These annual flood damages, which differ by an order of magnitude for even the present-day, are subject not only to uncertainties in the physical science but also to their representation of socio-economics. A key error source here is grid resolution, where the distribution of people and assets (exposure) is confined to ~1 $km^2$ cells. As highlighted in section 2.2.4, Smith et al. (2019) demonstrate that representing exposure at such a coarse scale results in significant misestimates of flood risk, even if the sum of the data (e.g. population count or building value) at a higher unit of aggregation is identical. The application of the modelling techniques and data processing underpinning the Sampson et al. (2015) framework to a large-scale risk model, avoiding the pitfalls of coarse exposure representation, is yet to be evidenced.

### 2.3 Thesis objectives

This review assimilates findings from a diverse array of literature on past and future changes to flooding, how we have historically understood the phenomenon through reach-scale models, and the emergence of new tools to expand the spatial scale of our knowledge. The observational record of floods (mostly from river gauges) is too deficient in both time and space to make any assertion on trends to the physical phenomenon of flooding at the global scale. The available data present a complex and spatially mixed picture, with different geographies and drivers inhibiting a universal statement. This heterogeneity translates to projections of flooding also, and is further confounded by the high uncertainties induced by the fundamental inability of climate and hydrology models to represent relevant phenomena. That said, physical reasoning based on the Clausius–Clapeyron relation between air's temperature and its water-holding capacity (7% increase per °C rise) means there is high confidence of increased precipitation in a warming world. Whether this translates to increased flooding, though, depends on how much rainfall occurs at once, antecedent conditions and catchment characteristics. Projected increases in flooding from tropical cyclones as a result of warmer seas, pluvial events arising from intense local rainfall and changes to flood seasonality are afforded a higher level of confidence than fluvial floods. Understanding of this physical science will continue, yet even if stationarity of flood hazard were to be assumed, flood losses will still increase (as they have done over the last century) as a result of exposure changes. This facet of risk modelling, though, has received

comparatively little attention by flood scientists, despite it being the overwhelming driver of increased risk.

To understand flood risk, an idea of how streamflow translates to extents, depths and velocities on the floodplain is required. This field of enquiry – hydraulic modelling – has undergone extensive developments in recent years, though the rapidity of this expansion in modelling capabilities has left several research gaps. Any deterioration in skill between traditional local engineering models and recent large-scale approaches is not well documented and the representation of flood defences in these models is very poor. Furthermore, large-scale risk models are yet to make use of recent high-resolution, high-physics, total-coverage hydrodynamic models: current understanding of risk at large spatial scales is underpinned by coarse-resolution, low-physics models which focus only on large rivers.

As such, this thesis aims to address these challenges using a hydraulic model of the US based on the Sampson et al. (2015) framework. This model is the culmination of developments to LISFLOOD-FP, involving accurate and efficient hydrodynamics via a local inertial formulation of the Saint-Venant equations and a sophisticated approach to channel representation as a sub-grid parameter, but is applied at continental-scale amidst the data-richness (in terms of elevation and flow data) of the US. This model structure is chosen for its ease of integration and model set-up with new river flow and elevation data endemic to the US, permitting rapid re-execution of a modified Sampson et al. (2015) model. Its use of extensively validated LISFLOOD-FP hydrodynamics, shown to compare favourably to analytical and field test cases as well as other 2D codes (e.g. Bates et al., 2006; Hunter et al., 2008; Neal et al., 2009a; Bates et al., 2010; de Almeida et al., 2012), means accurate representation of sub-critical flows inherent to flood inundation can be applied at high resolution without prohibitive computational burden. Of the sporadic validation studies carried out to date, the Sampson et al. (2015) model has generally performed better than other academic models. It achieved higher CSIs on the Severn and Thames against official government hazard maps than models of Dottori et al. (2016), Winsemius et al. (2016), and Alfieri et al. (2014), even achieving local relevance (according to the 0.65 CSI benchmark set by Fleischmann et al., 2019) on the Severn. The proximity of the model developers at Fathom enables access to the source code of the model builder and technical support to ensure a

comprehensive interrogation of the framework. The Sampson et al. (2015) method is thus the leading candidate for application in this thesis, which has the following detailed objectives:

1. Perform a comprehensive validation of flood inundation modelled over large spatial domains:
    a) by comparison to ubiquitous, engineering-grade hydraulic models of a number which constitute a commensurate spatial scale to the large-scale approach;
    b) by comparison to real-world flood event observations, thus necessitating coupling the hydraulics to meteorological and hydrological models;
    c) through evidencing increased accuracy relative to low-complexity, zero-physics approaches.
2. Explore new applications for a hydraulic modelling approach which avoids the pitfalls of low-coverage local models and current low-accuracy large-scale models:
    a) through incorporation in a flood risk quantification framework with a more sophisticated treatment of exposure;
    b) through the generation of operational flood inundation forecasts (also requiring the advances sought in objective 1b).
3. Tackle the issue of poor flood defence representation epidemic in current large-scale approaches.

These objectives intend to chart the development of the next generation of large-scale flood models, addressing significant problems with current approaches to make them suitable for myriad applications. More advanced large-scale hydraulic models will aid: (i) continued understanding of changing flood risk under climate, demographic and economic change; (ii) in providing rigorous, spatially consistent flood maps for regulatory zoning; (iii) insurers in ensuring their underwriting procedures are reflective of accurately quantified flood risk; and much else. These objectives are addressed in the following four research chapters.

# 3 VALIDATION OF A 30 M RESOLUTION FLOOD HAZARD MODEL OF THE CONTERMINOUS UNITED STATES

*This paper-chapter was published in Water Resources Research in 2017:*

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

*Author contributions:*
O. Wing led the study, processed the benchmark and auxiliary data, performed all analyses, and wrote the manuscript. C. Sampson and A. Smith provided the output of the continental hydraulic model simulations. All authors aided in the conceptualisation of the research and commented on manuscript drafts.

## 3.1 Background

This chapter addresses thesis objective 1a:

> *Perform a comprehensive validation of flood inundation modelled over large spatial domains by comparison to ubiquitous, engineering-grade hydraulic models of a number which constitute a commensurate spatial scale to the large-scale approach.*

With a global modelling framework set out by Sampson et al. (2015) applied to elevation data with higher vertical and horizontal accuracy over the US, this body of work sought to benchmark this new approach against the full inventory of FEMA flood maps. The comparison between an automated, inexpensive and rapid delineation of AEP floodplains and that of hundreds of thousands of manual, bespoke studies requiring billions of dollars in funding exhibited some degree of convergence between the two approaches. In a validation exercise more comprehensive than any other, model performance is benchmarked for different climatologies, land-uses, river basins, and AEPs in a continental-scale test area. This chapter also demonstrates the drawback of understanding large-scale flood hazard using a patchwork of local models, where expanses of US land remain unmodelled by FEMA.

## 3.2 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 datasets (Sampson et al., 2015;

Dottori et al., 2016; Winsemius et al., 2013; Pappenberger et al., 2012). 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 modelling from exclusively local reach-scale studies to continental- and global-scale analyses. The vertical accuracy of large-scale terrain datasets 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 modelling. 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 dataset 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. Trigg et al. (2016) conducted a continent-wide intercomparison of six global model outputs over Africa and Pappenberger et al. (2012) compared their model output to the global hazard map produced by UNISDR for the 2011 Global Assessment Report on Disaster Risk Reduction. The benchmark dataset of the latter was generated using flows from gauging data and regionalization approaches where catchments were ungauged. These were then used to simulate 1 in 100-year flood extents, and then merged with real flood events from the Dartmouth Flood Observatory to ascertain the 1 in 100-year flood hazard globally (Herold and Moulton, 2011). While these two studies adopted large-scale validation procedures, the validation data itself is 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.

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 dataset 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 dataset (Gesch et al., 2014). The USA 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 equalled or exceeded in any given year (FEMA, 2016a). Validation data are therefore available in the form of a mosaic of community-level flood hazard assessments spanning the USA. These are carried out by FEMA to determine a SFHA in a particular locality at a standard specified by NFIP.

These data in the US 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 utilised by a variety of clients: 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.

## 3.3 Data and methodology

### 3.3.1 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 regionalised 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. LISFLOOD-FP, a highly efficient inertial formulation of the shallow water equations in two dimensions (Bates et al., 2010; Neal et al., 2012b), propagates these extreme flows through the channel network and over the floodplain. These channels are delineated by the HydroSHEDS global hydrography dataset (Lehner et al., 2008), while the floodplain is represented by a digital elevation model (DEM) derived from the 1 arc second (~30m) 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 sub-grid method of channel representation (Neal et al., 2012b) is restricted to smaller rivers, while larger rivers are 'burned' directly into the DEM. The US Army Corps of Engineers (USACE) National Levee Dataset (NLD) is incorporated into the model to explicitly represent known flood defences. 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 behaviour, means they cannot be 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 for each 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. This assumption involves the 1 in 10-year pluvial flood in an intense urban area being captured by a drainage network, thus not generating hazard. In medium intensity developments, the drainage capacity is 1 in 5-year.

NED is a continuously updated dataset utilising the most accurate elevation data, 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, 2006a). Though NED is available at ⅓" (~10m) 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 provides a more practicable DEM for continent-wide hydraulic modelling. 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., 2012b). 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 towards negative or positive errors (Gesch et al., 2014), meaning vertical error at 1" is one-third of the error at ⅓" on flat terrain.

The NLD provides a map of regions protected by flood defence structures. The regions are accompanied by defence design standards, and the approach adopted by this model is to restrict flow into these regions at return period simulations below the defence standard while permitting flow for return period simulations that exceed the defence standard. This approach has an advantage over a simple post-simulation masking approach (whereby wet pixels within the defended areas are reset to zero depth after simulation) as it enables the hydraulic effects of defence structures, such as backwatering, to be captured by the model. Here, we run a preliminary "defence height estimation" simulation each return period flood, tracking the height of water at the levee locations that protect to each return period and elevating the crests to this for the primary model run.

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 sub-grid model (Neal et al., 2012b) is retained for smaller channels whose width is below the grid scale.

### 3.3.2 FEMA benchmark

The benchmark data to which the model output will be compared is primarily sourced from FEMA, whose local modelling 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 studies across the US, given the range of methodologies employed. The vector-based data consist of over 2,000,000 GIS shapefiles, 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 (FEMA, 2015). These boundary conditions are usually then routed through a 1D or 2D 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 (USACE, 2016a; Du et al., 2012) and the hydraulic model HEC-RAS (USACE, 2016b; Icaga et al., 2016). The most accurate elevation data available to FEMA must always be used and has to meet certain vertical accuracy requirements (FEMA, 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 (FEMA, 2016b). Many of these conditions, however, are policy standards specified in the last few years and so will only apply to recent and future 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 analysing historical flood data (e.g. high water marks or aerial photographs of previous flood events) (FEMA, 1995; National Research Council of the National Academies, 2015).

Although much of the US is mapped, the FEMA data contains both declared and undeclared no-data areas. By their own admission, FEMA has not studied the areas shown in Fig. 3.1. 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 of 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 the Fig. 3.2 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 'un-modelled in FEMA' areas, such as those smaller tributaries that branch northwards. 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 ~1 km buffer constructed around the SFHA. Though this will still likely capture areas FEMA hasn't studied (but which are still classified as outside the SFHA), a reasonable idea of the continental model's performance should be provided.

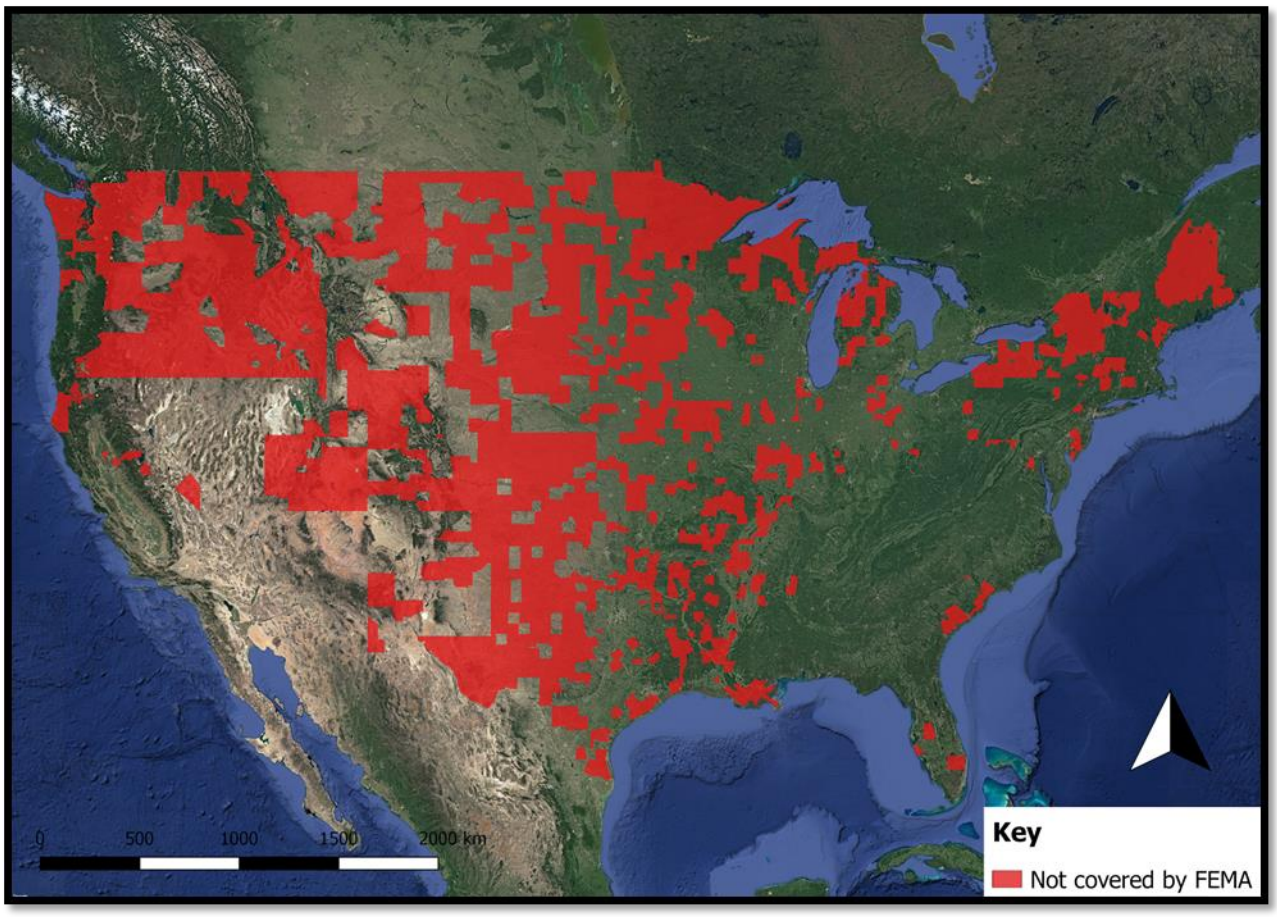


**Figure 3.1** Map of the US exhibiting areas FEMA explicitly claims not to have studied.

To undertake the analysis the 2,000,000 FEMA GIS shapefiles were converted to a 1” 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 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 of the 1 in 100-year floodplain as a result of levee construction. This information was also rasterised 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 rasterised 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.

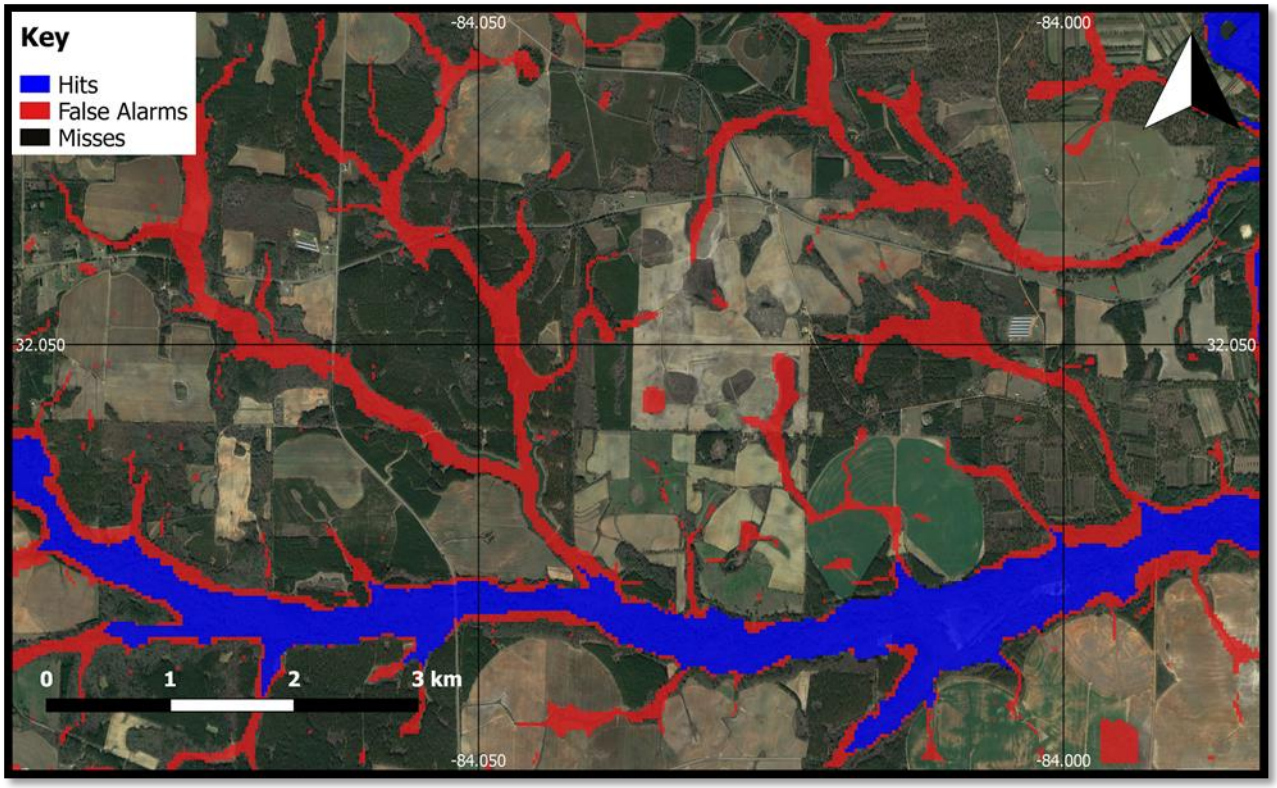


**Figure 3.2** Intersection of the defended 1 in 100-year model at 1" resolution with the FEMA benchmark in an area of Georgia to exhibit the varying nature of false positives. As per Table 3.1, 'Hits' correspond to $M_1B_1$, 'False Alarms' correspond to $M_1B_0$ and 'Misses' correspond to $M_0B_1$.

### 3.3.3 USGS benchmarks

As well as FEMA data, which represents the bulk of the validation information used here, isolated modelling 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 kilometres 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 1D hydraulic model HEC-RAS, one used its inferior counterpart HEC-2 and the other used a 2D model produced by USGS called FESWMS-2DH which uses a finite-element grid (USGS, 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 utilised 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 modelled and their associated USGS reports are detailed here:

- Albany, GA (Musser and Dyar, 2007): 1 in 100
- Battle Creek, MI (Hoard et al., 2010): 1 in 10, 1 in 50, 1 in 100, 1 in 500
- Columbus, IN (Coon, 2013): 1 in 100
- Greenville, SC (Benedict et al., 2013): 1 in 100
- Harrisburg, PA (Roland et al., 2014): 1 in 10, 1 in 50, 1 in 100, 1 in 500
- Hattiesburg, MS (Storm, 2014): 1 in 100
- Killbuck, OH (Ostheimer, 2013): 1 in 5, 1 in 10, 1 in 50, 1 in 100, 1 in 500
- Lincolnshire, IL (Murphy et al., 2012): 1 in 500
- Minneapolis, MN (Czuba et al., 2014): 1 in 100
- 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 US 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 behaviour in certain areas.

### 3.3.4 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 centimetres 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 analyse the relative number of pixels which conform to one of the states in the contingency table (Table 3.1).

| | Wet in benchmark data | Dry in benchmark data |
|---|---|---|
| Wet in modeled data | $M_1B_1$ | $M_1B_0$ |
| Dry in modeled data | $M_0B_1$ | $M_0B_0$ |

**Table 3.1** Contingency table of possible cell descriptors in a binary classification scheme.

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 towards underprediction of the flood hazard. *H* can range from 0 (none of the wet benchmark data is wet model data) to 1 (all of the wet benchmark data is wet model data).

$$H = \frac{M_1B_1}{M_1B_1 + M_0B_1} \tag{3.1}$$

The False Alarm Ratio (*F*) indicates the proportion of wet modelled 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).

$$F = \frac{M_1 B_0}{M_1 B_0 + M_1 B_1} \tag{3.2}$$

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

$$C = \frac{M_1 B_1}{M_1 B_1 + M_0 B_1 + M_1 B_0} \tag{3.3}$$

Finally, Error Bias (*E*) indicates whether the model has a tendency towards over- or underprediction. $E = 1$ would indicate no bias, $0 \leq E < 1$ indicates a tendency towards underprediction, and $1 < E \leq \infty$ indicates a tendency towards overprediction.

$$E = \frac{M_1 B_0}{M_0 B_1} \tag{3.4}$$

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

- Nationwide: all performance metrics within the buffer surrounding the SFHA
- Climate: performance analysed in the three main climate zones in CONUS
- Quality: performance where FEMA data is high quality (Zone AE) versus that where it is lower (Zone A)
- Defence: testing whether the continental model correctly identifies defended areas (as specified by FEMA) as dry
- Size of catchment upstream: analysis of whether the model performs better for rivers with larger or smaller upstream catchment areas
- Land-use: performance disaggregated between developed areas, forested areas and areas that are neither of these
- 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 defence data. Some of the metrics and scenarios are also applied to the 3" global model of Sampson et al. (2015) which utilises 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 utilise 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 km² 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.

$$E_A = \frac{\sum_1^N |M-B|}{N} \tag{3.5}$$

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

$$B_A = \frac{\sum_1^N M-B}{N} \tag{3.6}$$

If $B_A > 0$, it is an indication that the model has a tendency towards 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).

## 3.4 Results

### 3.4.1 Nationwide

The ~30 m flood model accounting for USACE levee data mapped the 1 in 100-year flood extent across CONUS. Analysing nearly 800,000,000 pixels, the nationwide results are shown in Table 3.2. 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. Fig. 3.3a 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 over- or underprediction. Fig. 3.3b illustrates where continental model performance is much poorer; there's a great deal of overprediction on the Rillito and Santa Cruz Rivers in Tucson, AZ.

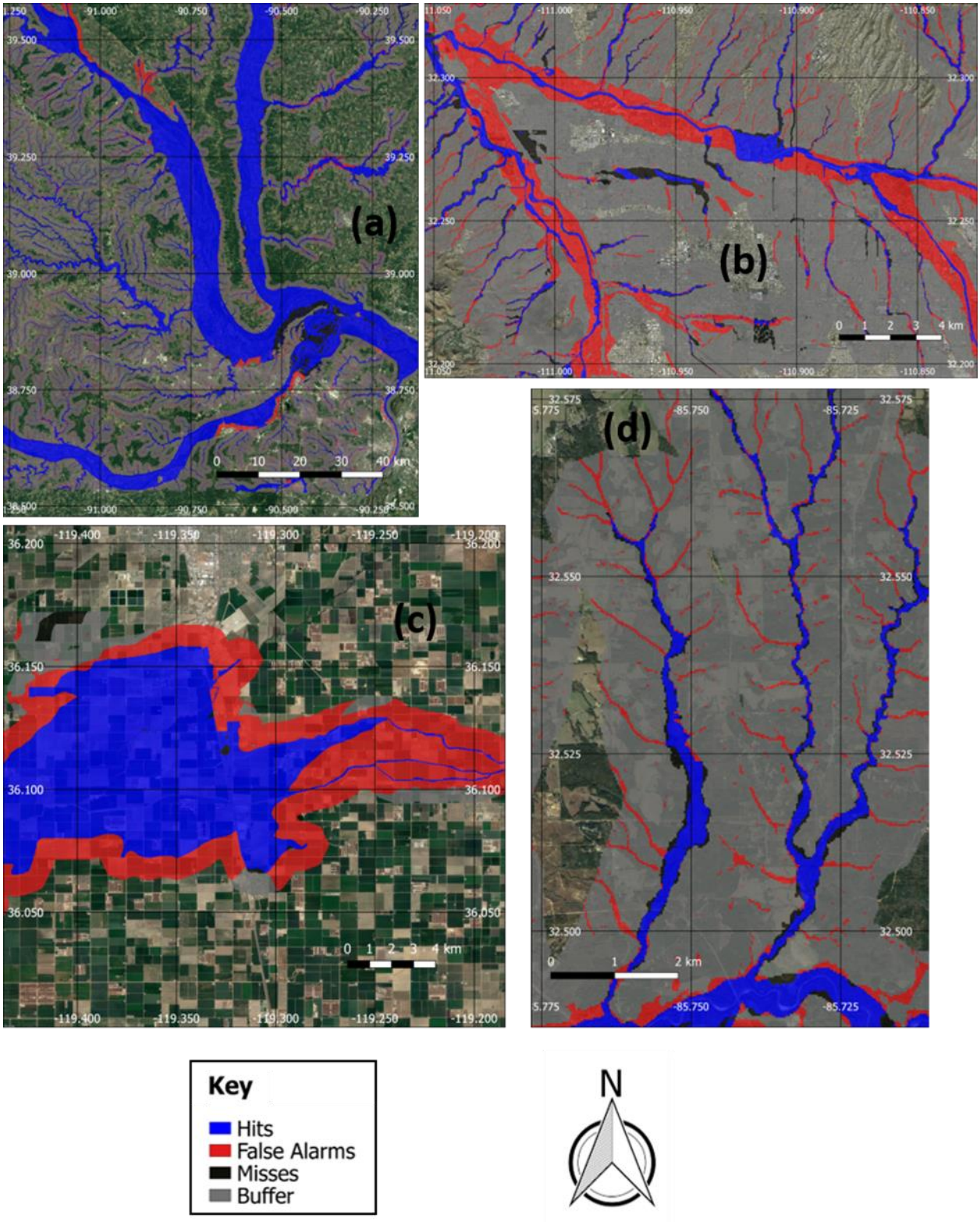


**Figure 3.3** Intersection of the defended 1 in 100-year model at 1" resolution with the FEMA benchmark in **(a)** an area of Missouri and Illinois near St. Louis; **(b)** Tucson, Arizona; **(c)** an area of California between Fresno and Bakersfield; and **(d)** an area of Alabama between Montgomery and Columbus.

| 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 |

**Table 3.2** Validation metrics for the defended, 1 in 100-year model across CONUS at 1” and 3” resolution.

Possible explanations for the differences between the continental model and FEMA data are numerous, with failure of the buffer to filter out areas un-modelled by FEMA likely bearing the most responsibility. From Fig. 3.3d, 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 hasn’t simply overshot the flood extent specified by FEMA, but has rightly captured flood hazard in the small river valleys. However, Figure 3.3c 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 3.3b, could perhaps be explained by its coarser resolution. Many flow restricting structures may not be resolved by the continental model, where a localised 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 ~30% of the nation’s levees. According to a report by the American Society for Civil Engineers (2017), the NLD contains roughly 30,000 out of an estimated 100,000 miles of US levees. This has severe consequences for the delineation of the continental-scale modelled 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 towards 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.

### 3.4.2 Climate

The contrast between Fig. 3.3a and 3.3b 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 analysing 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 3.3. Performance in temperate regions, which cover roughly two-thirds of the pixels analysed, 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 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 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 analysed, 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.

| 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 |

**Table 3.3** Validation metrics for the defended, 1 in 100-year model in three climate zones within CONUS at 1” and 3” resolution.

### 3.4.3 Quality

Some of the benchmark data has 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 3.4. 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 modelling 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 ten percentage points higher than in the 3” model.

| 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 3.4** Validation metrics for the defended, 1 in 100-year model at 1" and 3" resolution when compared against high and low quality FEMA data.

### 3.4.4 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 3.5. The key theme is that performance is notably higher in larger catchment categories, with areas around rivers with upstream catchment areas greater than 8,000 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 8,000 km$^2$ have slightly lower *H* scores of around 0.85 with false alarm ratios creeping upwards with reducing upstream catchment size. The marked increase in *F* scores from rivers with an upstream catchment area between 400 and 800 km$^2$ to those with an upstream 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 haven't 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 modelled 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).

| Approximate area of catchment upstream (km$^2$) | Size of buffer (km) | % of total pixels analysed | 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 |

**Table 3.5** River size categories, their associated descriptors and validation metrics for each using the defended, 1 in 100-year model at 1” resolution.

### 3.4.5 Land-use

Since people and assets are not distributed uniformly across the study area, it is necessary to analyse continental model performance in areas where the presence of a hazard translates into high risk separately to that 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 3.6, 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 3.6 are estimated in light of this. It is evident that the level of fit between the two datasets 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 modelling in urban areas, consequently requiring further scrutiny of the validity of the FEMA benchmark.

Urban hydraulic modelling has historically been challenging, owing to the 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, 2006a). 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 objects in the 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 these sub-grid scale, irregularly distributed, flow-obstructing objects (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 from the varied and often opaque methodologies they employ. In most instances the distinction is irrelevant, since 1D HEC-RAS models do not account for the hydraulic significance of structures on the floodplain anyway. 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's more useful to elaborate on why both of them will be subject to large errors. Extremely high *F* scores in more developed areas are perhaps explained by the incompleteness of the NLD by USACE, which will result in areas that are defended in reality being flooded in the continental model. FEMA will have accounted for these defences in ground-based field surveys, whereas the lower resolution continental model is unlikely to have captured the full effect of a levee unless specified explicitly from the NLD. 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 modelled 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 some level of agreement between the two datasets.

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 many instances this may be the case, since FEMA models will have often incorporated critical local-scale information, such as flood defences. However, much of the

divergence is likely derived from both sets of data 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.

| 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 |

**Table 3.6** Validation metrics for the defended, 1 in 100-year model at 1" resolution within land-use descriptors.

### 3.4.6 Defence

Table 3.7 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 US levee data in the model results in higher hit rates in defended areas. The continental model including defences 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.

| Model Version | Hit Rate |
|---|---|
| Defended (1”) | 0.673 |
| Undefended (1”) | 0.296 |
| Defended (3”) | 0.369 |
| Undefended (3”) | 0.367 |

**Table 3.7** Rate at which different model versions correctly identify defended areas as dry.

The defended 3” model, amongst its other differences to the 1” version, does not explicitly represent levees. Instead, defences are parameterised through the adjustment of channel conveyance based on socio-economic factors and degree of urbanization, which are assumed to be reasonable predictors of level of defence standard. The results in Table 3.7 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 mislabelling just under two-thirds of such areas as wet.

### 3.4.7 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 in Table 3.8. 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" model strongly outperforming it as usual.

| 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 |

**Table 3.8** Hit rate of the defended, 1 in 500-year model in different study areas at 1" and 3" resolution.

### 3.4.8 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 of 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 fraction at 1 km resolution between FEMA and the continental model. This value is towards 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 model to overpredict.

### 3.4.9 USGS

The isolated, local, high-quality flood hazard studies of the USGS provide excellent validation data for the model, albeit not on the grand spatial scale of the FEMA benchmark. The results of validating against the nine sites which modelled the 1 in 100-year event are shown in Table 3.9, and are graphically represented in Fig. 3.4. *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 Fig. 3.4, however, that often false alarms are generated from failure to isolate the hazard derived from the specific river modelled 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 unafflicted from high false alarm ratios are comparable to optima when high quality flood models are calibrated to real event data (Bates et al., 2006).

| 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 |

**Table 3.9** Validation metrics for the defended, 1 in 100-year model at 1” resolution when compared against USGS benchmark data.

Analysis of data on further return periods is listed in Table 3.10. 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 to significantly underpredict 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.

| 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 |

**Table 3.10** Validation metrics for the defended, 1" resolution model at different return periods when compared against USGS benchmark data.

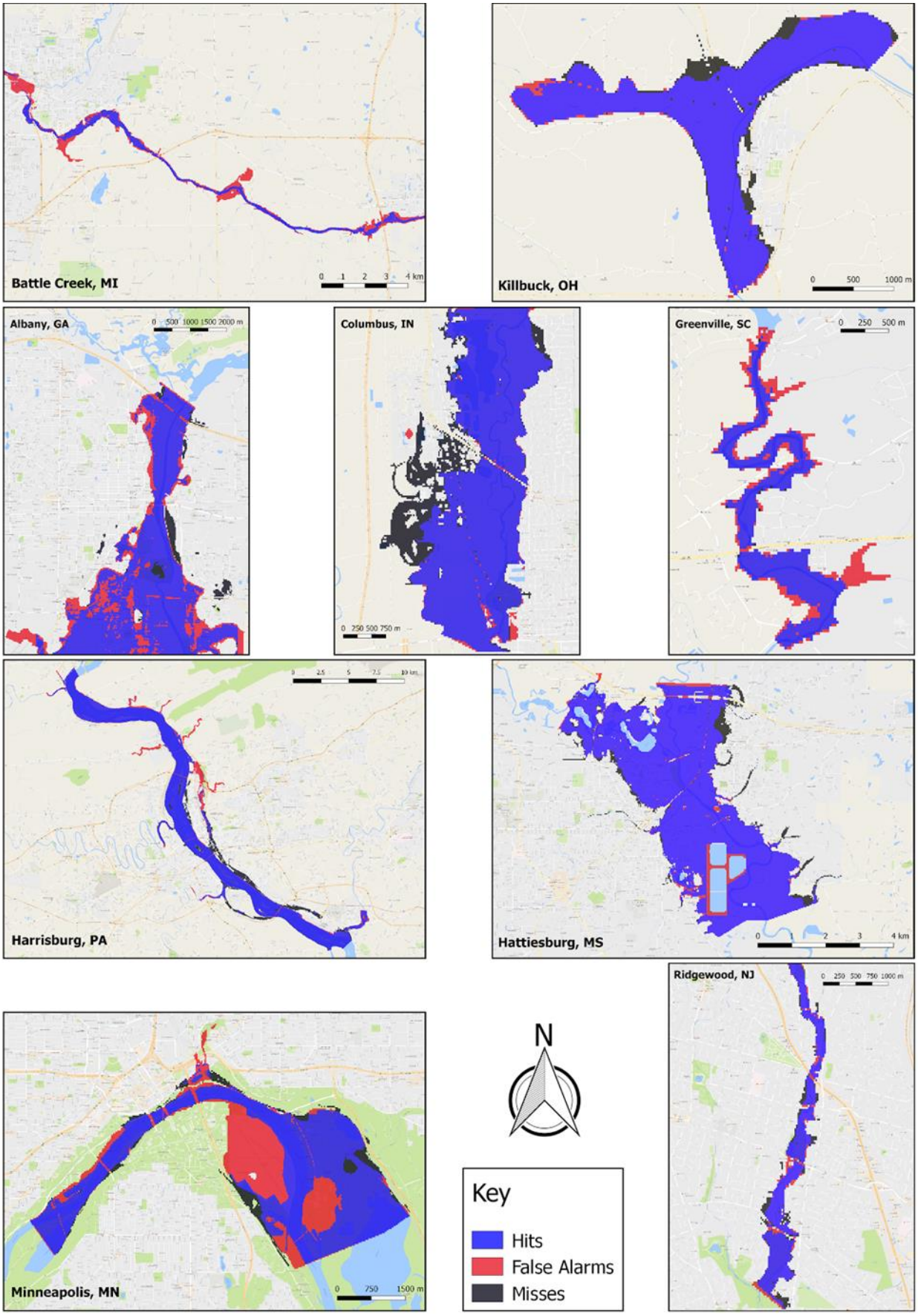
Battle Creek, MI
Killbuck, OH
Albany, GA
Columbus, IN
Greenville, SC
Harrisburg, PA
Hattiesburg, MS
Ridgewood, NJ
Minneapolis, MN
N
Key
Hits
False Alarms
Misses

**Figure 3.4** 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 modelled.

### 3.5 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 towards 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.

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 itself will 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 0.9s 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 Fig. 3.1) as well as the smaller watersheds not included in FEMA models (see Fig. 3.2). Additionally, the FEMA models have taken thousands of individual studies and many decades to develop, whilst 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 has spent between $4.5 and $7.5 billion on flood mapping up to 2013, and 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 would take approximately 5000 hours 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 re-run the continental model: either to update it with existing 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 recognised 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 clearly show the necessity for the explicit representation of defence data also. 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 here. 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 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 over- 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 simulated in catchments down to 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 accounting for by FEMA's failure to specify which headwaters are not modelled, 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 requires 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 calibrate future improvements to such data. On top of the directions for future work advocated on 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 to the field of large-scale flood hazard modelling is a demonstration that the 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 they are working with, but more cheaply, easily and quickly than alternatives of commensurate calibre. 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 is in producing quality hazard data. The authors therefore reiterate the plea of Schumann (2014) for a global terrain dataset 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 defence catalogue has been clearly demonstrated. Levee delineation is a crucial determinant of flood hazard, so an incomplete NLD has meant the modelled floodplain is an overprediction in some areas. This dataset needs to be improved and a global inventory of flood defences is required for further advancement in this field.

### 3.6 Acknowledgements

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 Centre (http://www.msc.fema.gov). USGS flood maps are available from https://water.usgs.gov/osw/flood_inundation/libraries.html. 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 dataset. In combination with the methodology, results and supporting information of Smith et al. (2015), our underlying discharge data is available from the Global Runoff Data Centre (http://www.bafg.de/GRDC/EN/01_GRDC/12_plcy/data_policy_node.html) for research purposes (requires registration). NOAA IDF data is 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 dataset 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). Our 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 non-commercial research 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.

# 4 ESTIMATES OF PRESENT AND FUTURE FLOOD RISK IN THE CONTERMINOUS UNITED STATES

*This paper-chapter was published in Environmental Research Letters in 2018:*

**Wing, O. E. J.**, Bates, P. D., Smith, A. M., Sampson, C. C., Johnson, K. A., Fargione, J., and Morefield, P. (2018a), Estimates of present and future flood risk in the conterminous United States, *Environmental Research Letters*, 13, 034023, doi:10.1088/1748-9326/aaac65.

*Author contributions:*
O. Wing led the study, processed the socio-economic data, generated vulnerability functions, performed all analyses, and wrote the manuscript. C. Sampson and A. Smith provided the output of the continental hydraulic model simulations. P. Morefield supplied the ICLUS data. All authors aided in the conceptualisation of the research and commented on manuscript drafts.

## 4.1 Background

This chapter addresses thesis objective 2a:

> *Explore new applications for a hydraulic modelling approach which avoids the pitfalls of low-coverage local models and current low-accuracy large-scale models through incorporation in a flood risk quantification framework with a more sophisticated treatment of exposure.*

With the skill of the US flood hazard model evidenced in chapter 3, surpassing that of the Sampson et al. (2015) global model and, implicitly therefore, that of other cited global models (e.g. Yamazaki et al., 2011; Winsemius et al., 2013; Ward et al., 2013; Pappenberger et al., 2012; Dottori et al., 2016), the model could find applicability in quantification of flood risk in the US. With information gaps identified in regulatory FEMA maps and other global models subject to known limitations, this chapter uncovers previously unknown or miscalculated levels of flood risk inherent to the US. Not only does this chapter provide novelty through its application of a newly validated hazard component, high-resolution data on the location of people and assets are employed to present the most accurate quantification of US flood risk to date. Beyond this, high-resolution projections of exposure changes throughout the 21$^{st}$ century are employed to assess the intensification of floodplain development. Focusing only on the expected primary driver of past and future flood risk

increases (developmental changes), the analysis finds that the number of people and value of buildings in the 1 in 100 year floodplain may almost double by 2100. The study evidences that – climate change aside – flooding is a severe problem in the US right now and will only get worse if floodplain development continues broadly unabated. This paper-chapter was originally published as a letter with additional supplementary information; this has been re-written for suitable inclusion in this thesis.

### 4.2 Introduction

In 2016, global economic losses as a result of flooding totalled $56bn (all values are 2017 USD), with $10bn of this accounted for by the August floods in Mississippi and Louisiana alone (Munich Re, 2017). In the US over the past 30 years, freshwater flooding has caused an average of $8.2bn in damages each year, though this average masks an upward trend in flood losses over time (National Weather Service, 2015; Changon, 2008; Pielke and Downton, 2000). This is also the case globally, with the major driver thought to be the increased exposure of people and assets (IPCC, 2012; Changon et al., 2000).

The Federal Emergency Management Agency (FEMA) has produced maps delineating the Special Flood Hazard Area (SFHA) for nearly all current coastal flood hazard areas in the US, and rigorous estimates have been published indicating how many people are exposed and how exposure is distributed nationally (Crowell et al., 2010; Neumann et al., 2015). Maps delineating fluvial (riverine) and pluvial (rainfall-driven) flooding, however, are only partially complete nationwide, and no comprehensive estimate of US population exposure currently exists. Where they are available, FEMA flood maps are of varying age and levels of quality. They also have notably poor coverage of smaller catchments, which is a trait shared by many of the hazard maps that are used to inform risk calculations at global or continental scales. For example, the framework for flood risk assessments set out by Winsemius et al. (2013), which is the current state-of-the-art in large-scale flood risk analytics (Winsemius et al., 2016; Muis et al., 2015; Jongman et al., 2015; Ward et al., 2013), excludes rivers below Strahler (1964) order 6 (catchments smaller than roughly 10,000 $km^2$). This means that risk generated by these smaller streams, which may be situated in or around residential or commercial areas, is not captured. Further, coarse-resolution terrain data and the simplistic representation of the physics of flood spreading are characteristics shared by a majority of existing large-scale models (Trigg et al., 2016). It is evident, therefore, that previous large-scale efforts to quantify flood exposure (in terms of population and economic assets) and risk

(in terms of deaths and economic losses) have known limitations that will lead to misestimation of these quantities.

This study presents new estimates of current and potential future flood exposure and risk using high-resolution hazard, population, asset and projected development maps of the entire conterminous United States (CONUS). These layers are of significantly higher quality and spatial coverage than those that have previously informed exposure and risk estimations. Validation of the new hazard layers (Wing et al., 2017) suggests they are of commensurate quality to local studies carried out by US government agencies. These new high-resolution analyses with a realistic representation of flood physics indicate that the population exposed to serious flooding in CONUS is 2.6 to 3.1 times higher than previous estimates. This has major consequences for flood risk management and policy in the US.

### 4.3 Terminology

Risk is defined in this paper as being the product of hazard, exposure and vulnerability; consistent with the definition of the United Nations International Strategy for Disaster Risk Reduction (2015b). Hazard refers to the nature, magnitude and probability of the flood event itself. Socio-economic data, such as where people live or where buildings are situated, are intersected with the delineation of the hazardous area to generate flood exposure. This process is described schematically in the 'exposure' box (lighter grey) of Fig. 4.1. Exposure does not account for factors such as the damage a flood may cause to a structure, the number of fatalities arising from an event, or the propensity to mount a short-term response to a flood in order to mitigate its effects (e.g. an early warning system leading to the erection of temporary defences). These factors broadly relate to the vulnerability of an area, defined as the susceptibility of the exposed people or assets to experience loss (in various forms), which creates the distinction between flood exposure and flood risk. For example, exposure may be the total value of buildings within a flood zone, while risk may be the annual expected damage to these buildings. To constrain risk in these estimates, relative depth-damage curves (representing vulnerability) are applied to the exposed assets (see the 'risk' box (darker grey) of Fig. 4.1). This relationship determines the economic damage a certain depth of water causes to an asset as a percentage of its total value. While this generates some idea of flood risk, it still does not account for the wealth of local-level factors which are ultimately required for a more accurate estimation. On top of this, the depth-damage relationships themselves are subject to much uncertainty (Jongman et al., 2012b). This and other

uncertainty is explored in section 4.7, but the key point here is that uncertainty accumulates when moving from a hazard to a risk calculation. Observing the hazard layers in isolation is already subject to much uncertainty; but a risk estimation is not only subject to uncertainty in the hazard layer, but also to uncertainty in asset values, the location of assets and the economic effect of flooding on a certain asset. Although a risk estimation may seem to provide something more tangible and interesting (e.g. expected damage from a flood event), an exposure estimation (e.g. total value of assets within a certain floodplain) will be subject to less uncertainty, and a hazard estimation (e.g. total area of the floodplain) to even less than that. The data and methodology employed to generate hazard, exposure and, ultimately, risk estimations are detailed in the following section.

## 4.4 Methods

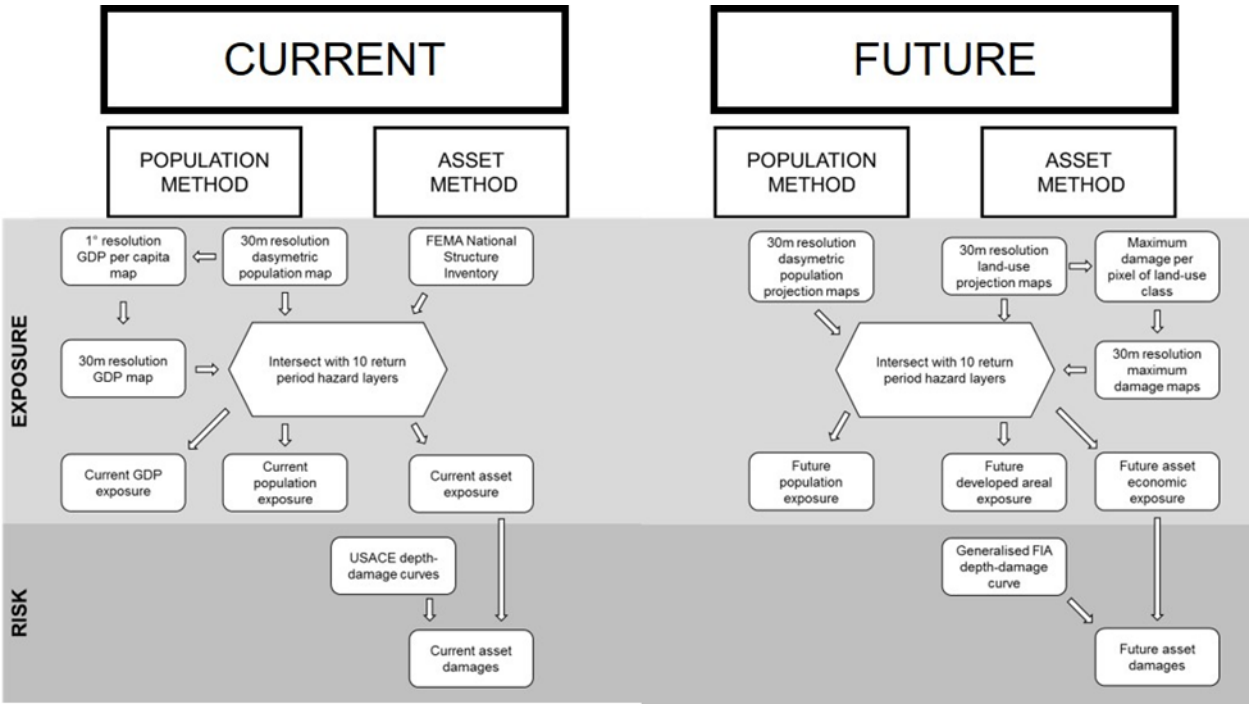


**Figure 4.1** Schematic diagram of the methodology for generating exposure and risk estimates. FIA stands for Federal Insurance Agency.

### 4.4.1 Hazard

The hazard layers employed to inform these estimates represent both fluvial (flooding from rivers) and pluvial (flash-flooding arising from direct rainfall onto the land surface) perils and are detailed more fully in Wing et al. (2017).

The fluvial model component is driven by design discharges of ten different recurrence intervals: 1 in 5-, 10-, 20-, 50-, 75-, 100-, 200-, 250-, 500- and 1000-year (20% to 0.1% annual probability). These are generated using a regionalised flood frequency analysis (RFFA; Smith et al., 2015), using roughly 6000 US Geological Survey (USGS) river gauges across the CONUS. The RFFA overcomes the issue of spatial sparsity in river gauges by transferring flow data from a gauged basin to a similar ungauged basin. This approach assumes that catchments with similar characteristics, such as climatology, upstream annual rainfall and drainage area, will have a similar flood frequency. Suitably homogenous groups are assigned a flood estimation index which details the mean annual flood and flood frequency curve of each river reach, permitting the 10 return period discharges to be generated. A synthetic triangular hydrograph is generated for each river reach based on a determined time to concentration and the RFFA-derived discharge at its peak (Sampson et al., 2015). The USGS National Elevation Dataset (NED) is used to generate the digital elevation model (DEM) at 1 arc second (~30m) resolution, and simulations are executed at this high resolution. River channels are delineated by the HydroSHEDS global hydrography dataset (Lehner et al., 2008) and if a reach is narrower than the resolution of the grid, it is represented at the subgrid level (Neal et al., 2012b). The RFFA-generated discharges are routed by means of an efficient inertial formulation of the shallow water equations in two dimensions, based on LISFLOOD-FP (Bates et al., 2010; Neal et al., 2012b), through the channels and over the floodplain. Flood defences are incorporated explicitly into the model, sourced from the US Army Corps of Engineers (USACE) National Levee Database (NLD).

The fluvial component is only executed for catchments larger than 50 $km^2$. Smaller catchments are highly heterogenous in terms of their flood frequency response, which alongside a lack of stream records for such areas, renders them unsuitable to be incorporated into the RFFA. Instead, these areas are simulated by the pluvial model, owing to their ‘flashy’ flood response to intense local rainfall events. National Oceanic and Atmospheric Administration (NOAA) Intensity-Duration-Frequency relationships form the basis of design rainfall events of 1-, 6- and 24-hour duration. The maximum water depth in each pixel from these three simulations forms the pluvial hazard layer. In recognition that not all of this rainfall will generate flood hazard, an infiltration equation is applied (Morin and Benyamini, 1977) depending on the underlying soil type (sourced from the Harmonized World Soil Database; Fischer et al., 2012) and, in urban areas (Elvidge et al., 2007), a certain drainage design standard is assumed. This assumption involves the 1 in 10-year pluvial flood in an

intense urban area being captured by a drainage network, thus not generating hazard. In medium intensity developments, the drainage capacity is 1 in 5-year.

The friction parameter is spatially uniform, but different for channels and floodplains. It is based on land cover data and sensible values from the literature (Sampson et al., 2015; Chow, 1959). No calibration of this parameter was carried out, simply because calibration data are not available at this scale. Validation against US government agency flood maps across the CONUS has been performed, with the hazard model in this study replicating up to 100% of USGS-defined flood plains and up to 90% of high quality FEMA flood maps (Wing et al., 2017). Simulations were run in unsteady-state conditions.

To simulate all components of the flood hazard model for a single return period, a server node with 20 Intel Broadwell E5 Xeon cores takes ~5000 hours. The runtime is much shorter in practice, as computations are performed over multiple nodes.

The Intergovernmental Panel on Climate Change has low confidence in even the direction of change in future flooding (Seniveratne et al., 2012). This significant uncertainty is emergent in current climate models, which have a low level of agreement with regards to changes in flood-inducing rainfall. The propagation of modelled precipitation through rainfall-runoff models further amplifies these uncertainties. Smith et al. (2014) demonstrated that the estimation of present-day extreme flow discharge in a well-gauged catchment already presents a significant challenge. Indeed, the study suggested that the incorporation of future flood projections only led to a small increase in the uncertainty that was already present in the extreme flow estimation procedures for current conditions. Given that these conclusions were reached using a well-calibrated rainfall-runoff model in a gauged catchment with a long historical record, it is likely that the scale of current uncertainties in extreme flow estimation compared to potential climate change effects will be even greater for poorly understood or ungauged catchments. Therefore, the hazard layers used to estimate future flood risk do not consider future climate change and are the same as those used for current estimates.

### 4.4.2 Current flood exposure and risk

#### 4.4.2.1 Population exposure

To estimate the total number of people exposed to floods of various magnitudes, the hazard layers are intersected with the US Environmental Protection Agency (USEPA) EnviroAtlas

map of population density, which distributes 2010 census block population counts to 30 m pixels based on land-use and slope. Using dasymetric techniques, each state is assigned a population density per land-cover classification (e.g. higher density for intense urban development) based on the NLCD. Block-level counts are then redistributed to respective 30 m land-cover pixels that have slopes gentler than 25% as calculated using the NED. Flood exposure is then calculated by summing pixel values from the population map for all 'wet' cells in the hazard map (i.e. a positive water depth). It is also combined with a GDP per capita map derived from G-Econ (Nordhaus et al., 2006), which represents the heterogeneity of GDP distribution across the USA. This provides a monetary (GDP per 30 m pixel), as well as popular, value of current flood exposure.

#### 4.4.2.2 Asset exposure and risk

The FEMA National Structure Inventory (NSI) is used to estimate the value of assets currently exposed to fluvial and pluvial flooding in the CONUS. The NSI details the nature, value and location of buildings and their contents in the US. Upon intersection with the hazard layers (to generate exposure), a specific USACE depth-damage function (Davis and Skaggs, 1992) based on building type is applied to calculate expected damages from a certain return period flood event (risk). For this risk calculation, rather than using binary hazard maps where a pixel is either 'wet' or 'dry'), simulated water depths are employed. These functions essentially relate the water depth, as per the hazard layers, at a given asset location to the likely damage that it would cause. This relationship is different for different types of buildings, so a different function is applied depending on information captured in the NSI (e.g. occupancy type, number of storeys, presence of a basement). The National Land-Use Dataset (NLUD; Theobold, 2014) is used to calculate the area of exposed developments.

### 4.4.3 Future flood exposure and risk

Future flood exposure and risk is estimated using population and land-use projections from the USEPA Integrated Climate and Land-Use Scenarios (ICLUS) project. These projections were generated using inputs and assumptions corresponding to Shared Socioeconomic Pathways (SSPs), which describe broad societal changes such as human migration patterns, fertility rates and technological innovation. Projections under SSP2 (medium population growth, historical migration patterns continue) and SSP5 (high population growth, medium-sized cities see increased in-migration) are used for the years 2050 and 2100. These scenarios attempt to capture a plausible range of socio-economic and demographic growth throughout

the coming century, with projected population at 2100 for SSP2 and SSP5 exceeding 450 and 700 million respectively. The projections are at 90 m resolution, but are resampled to 30 m in order to align with the hazard layers. A fuller explanation of ICLUS can be found in USEPA (2016).

#### 4.4.3.1 Population exposure

As with the current population-based exposure estimates, the hazard layers are intersected with the population projections to indicate the future population exposed to rainfall and river-flow driven flooding in 2050 and 2100. These projections by the ICLUS demographic model are derived from its input variables, such as fertility, mortality and migration, as quantitatively described in the literature (Samir and Lutz, 2014; van Vuuren and Carter, 2014). County-level populations projected by a cohort-component and migration model, calibrated to data from the US Census Bureau (2000, 2003), are downscaled onto the land-use projection grid based on the number of dwelling units within a particular class (Theobold, 2001; Ewing and Cervero, 2010).

#### 4.4.3.2 Asset exposure and risk

The National Structure Inventory was used to calculate asset values per pixel of each 'developed' land-use class in the present-day 30 m resolution NLUD (e.g. greater value/pixel of dense urban than rural areas). The pixel values of each class were then assigned to the corresponding pixels of the future ICLUS land-use projections. The ICLUS land-use projections were generated alongside the population maps (section 4.4.3.1) and are generated by a spatial allocation model, which creates new residential, commercial, and industrial land-use patches until the demand from the county-level population projections is satisfied. This model incorporates a vast amount of data at varying spatial scales, including city growth functions, transition probabilities, regional land-use patch shape and size parameters, travel times and transport capacity. The area and value of exposed developments are calculated by intersecting these projections with the hazard data, and the expected damage is computed using a single Federal Insurance Agency (FIA)-derived generalised depth-damage relationship since projections do not capture building type. This function is a composite of FIA vulnerability functions relating to damage to different types of residential property, and is calculated by giving greater weight to property types that make insurance claims more frequently under the National Flood Insurance Programme.

## 4.5 Results

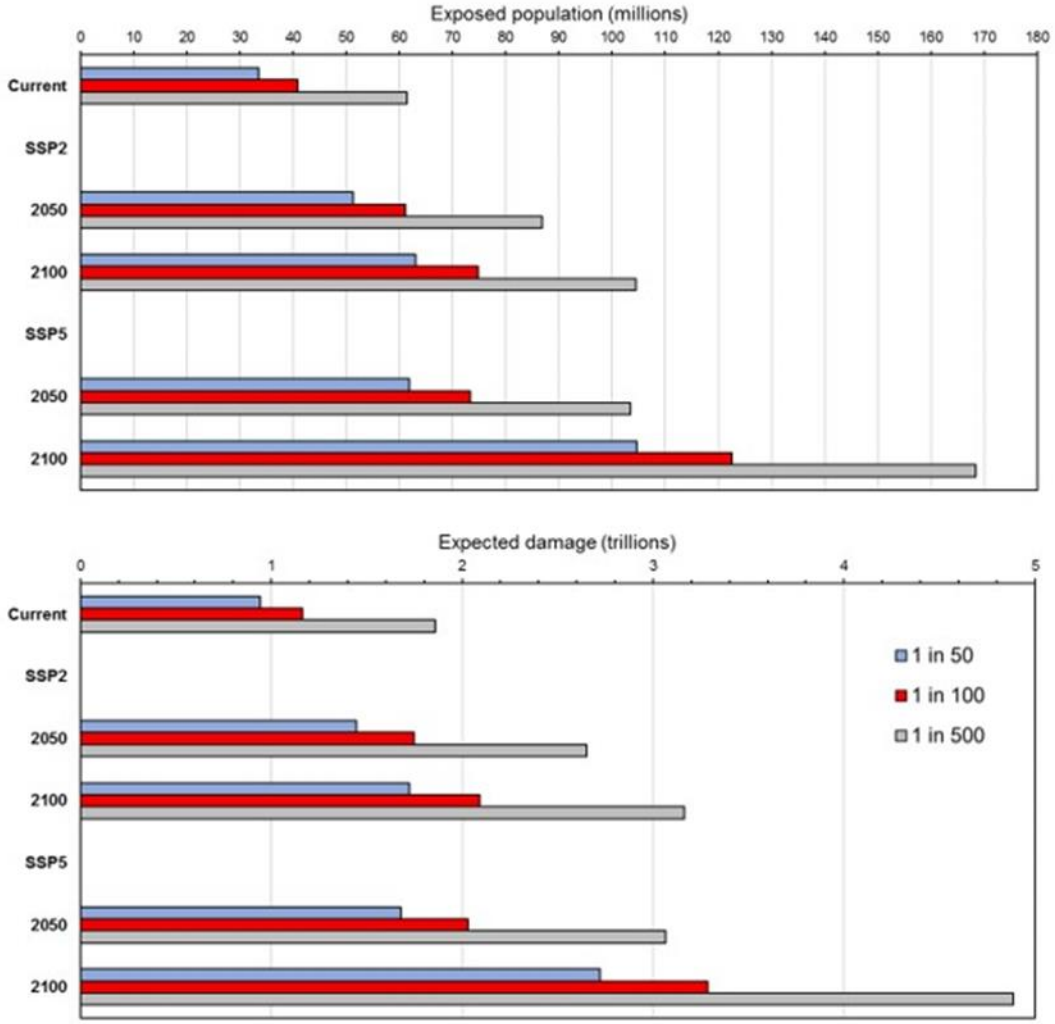


**Figure 4.2** Selected exposure (population) and risk (damage) estimates for present and future 1 in 50-, 100- and 500-year floods. SSP2 represents a medium growth scenario, in terms of population and development, while SSP5 represents a higher one.

### 4.5.1 Current flood exposure and risk

#### 4.5.1.1 Population exposure

The analysis shows that 40.8 million people (13.3% of the population) are currently exposed to a 1 in 100-year (1% annual exceedance probability) fluvial or pluvial flood in the CONUS, which translates to a GDP exposure of $2.9 trillion (15.3% of total GDP). This represents substantially higher exposure than previous estimates suggest. The World Resources Institute (WRI) Aqueduct Global Flood Analyser (hereafter, Aqueduct; Winsemius et al., 2013; Ward et al., 2013; WRI, 2015) suggests that 15.7 million people and $0.7 trillion of GDP are exposed to a 1 in 100-year flood in the US. FEMA flood maps, intersected with the exposure

data used in this study, indicate 13.0 million people are exposed (Fig. 4.3). FEMA flood maps only cover around 61% of the CONUS land area, and even within this many small headwater catchments are un-modelled (Wing et al., 2017). A growing body of research also highlights the inadequacies of FEMA data where it does exist. Blessing et al. (2017) compared FEMA flood maps to loss claims from five storms near Houston, Texas. Only one-quarter of the claims were located within the FEMA-delineated 1 in 100-year flood zone, despite none of the storms having a return period greater than a 1 in 50-year. Fig. 4.4 elucidates where the differences between our analysis and FEMA exposures arise. While our study identifies exposure missed by FEMA across the country, higher concentrations of newly identified exposure are particularly evident along the Pacific coast, in urban centres around the Great Lakes and across inland western US. With no coverage of small rivers in the Aqueduct data and incomplete coverage in the FEMA flood maps, it is not surprising that our study has identified additional areas of flood exposure. This analysis indicates that previous estimates capture roughly one-third of the exposure identified in our 1 in 100-year floodplain. Table 4.1 and Fig. 4.2 detail these current population-based exposure estimates further. It is also worth noting the similarity of results across all three return periods for the Aqueduct data. Trigg et al. (2016) observe that the underlying hazard layer of the Aqueduct data, GLOFRIS (GLObal Flood Risk with IMAGE Scenarios), displays relatively little sensitivity to the probability of the flood event in terms of total flooded area. This also appears to be the case in terms of population exposure: only 2.6 million more people are exposed to the 1 in 500-year flood compared to the 1 in 50-year. In reality, this is unlikely to be the case. Our analysis shows population exposure almost doubles from the 1 in 50- to the 1 in 500-year event.

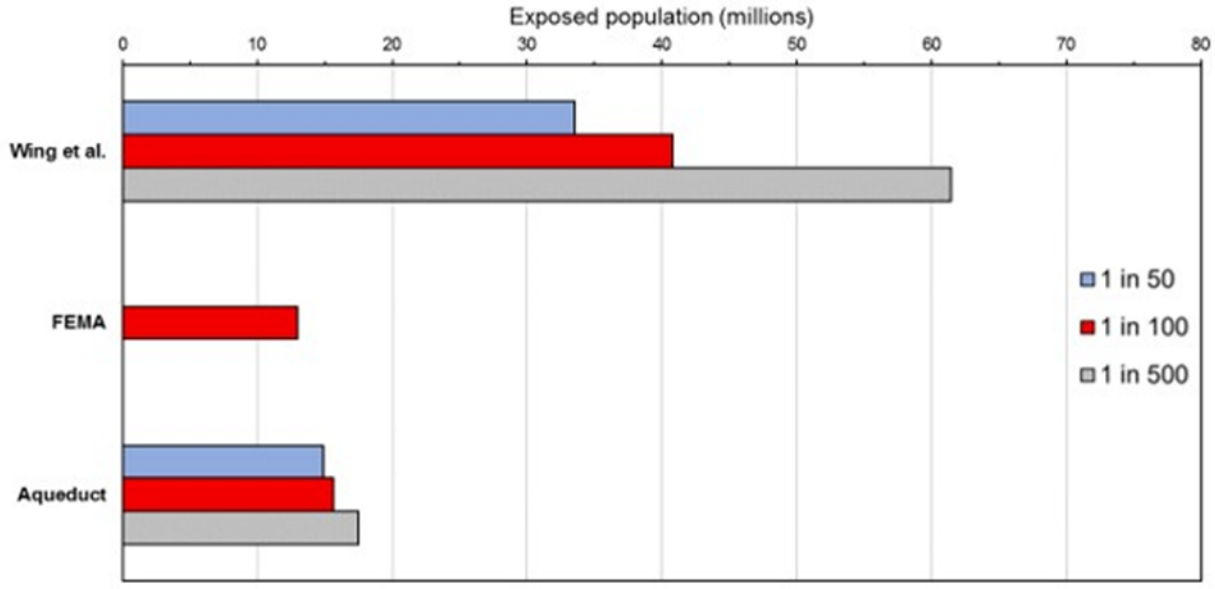


**Figure 4.3** Total population exposed to flooding in the CONUS for the present-day. As well as containing results from this analysis, the graph displays FEMA-derived and Aqueduct exposure estimates.

| Return Period (year) | Exposed population | Percentage population exposure | GDP exposure (trillions of USD) | GDP exposure (% of US GDP) |
|---|---|---|---|---|
| 50 | 33,567,281 | 11.0 | 2.4 | 12.6 |
| 100 | 40,817,905 | 13.3 | 2.9 | 15.3 |
| 500 | 61,443,761 | 20.0 | 4.4 | 23.1 |

**Table 4.1** Present-day population-based flood exposure in the CONUS based on the analysis in this paper.

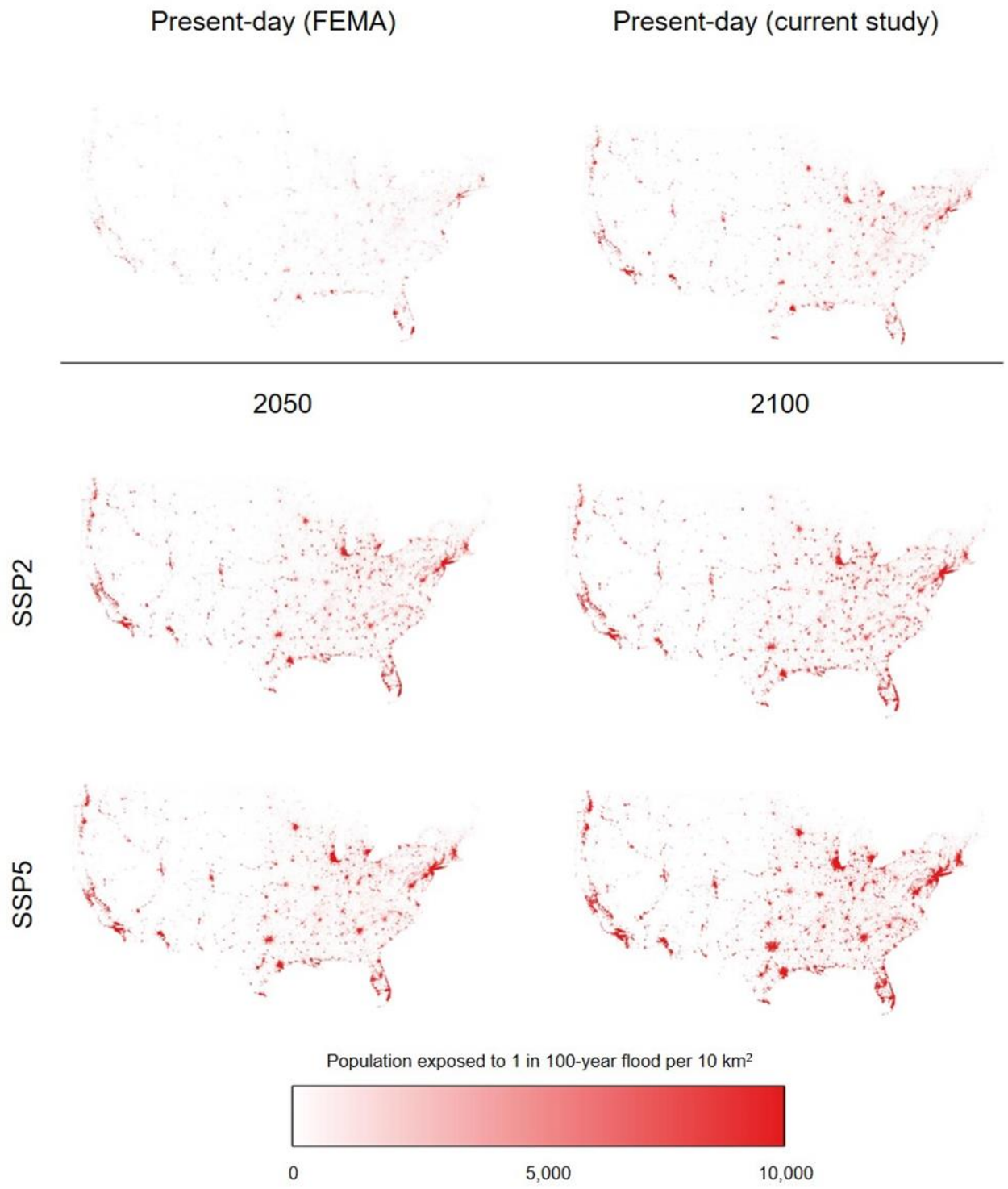


**Figure 4.4** Distribution of population exposed to a 1 in 100-year flood across the CONUS for the present and future.

#### 4.5.1.2 Asset exposure and risk

The total value of assets within the present-day CONUS 1 in 100-year floodplain is $5.5 trillion, with $1.2 trillion of this at potential risk from flood damage (Table 4.2). Fig. 4.5

indicates that Louisiana, Florida, Arizona and West Virginia are particularly over-exposed, with 32%, 28%, 26% and 25% of their total asset values situated within the 1 in 100-year floodplain respectively. From Figure 4.6, it is evident that the absolute value of assets on the Floridian floodplain is also particularly high at $714 billion: Florida is thus a hotspot of flood exposure. While their percentage of exposed assets are not particularly high, California and Texas have high absolute values of flood exposure at $763 billion and $400 billion respectively.

| Return Period (year) | Area of exposed developed land ($km^2$) | Total value of exposed assets (trillions of USD) | Expected damages (trillions of USD) | Number of exposed assets (millions) |
|---|---|---|---|---|
| 50 | 140,657 | 4.6 | 0.9 | 12.9 |
| 100 | 157,430 | 5.5 | 1.2 | 15.4 |
| 500 | 203,775 | 8.2 | 1.9 | 22.6 |

**Table 4.2** Present-day asset exposure and risk estimates in the CONUS.

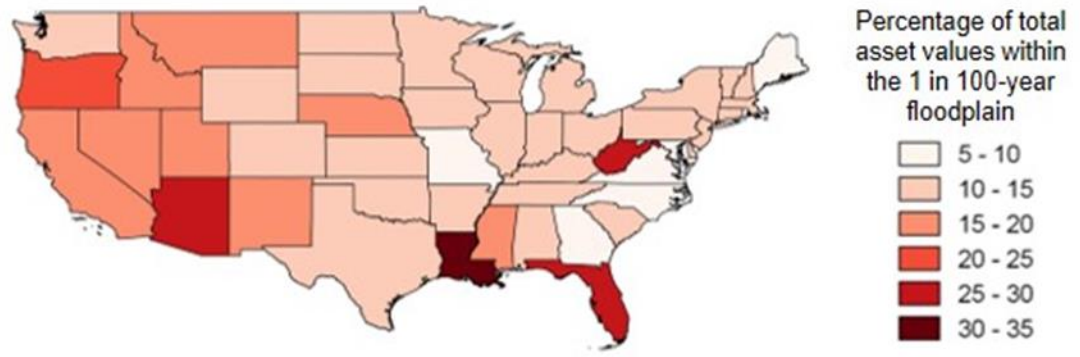


**Figure 4.5** Proportion, in terms of their economic value, of assets within the present-day 1 in 100-year defended floodplain.

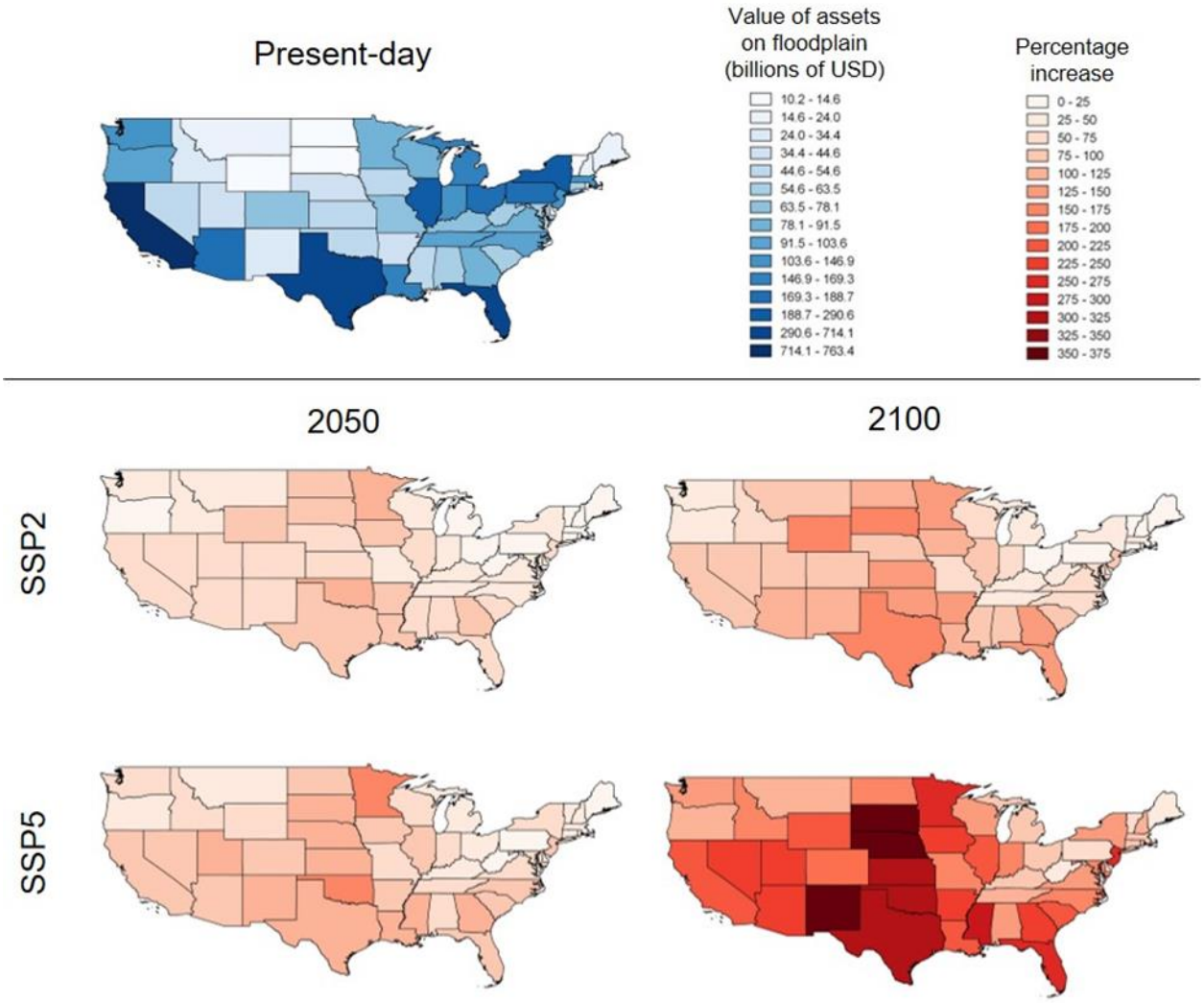


**Figure 4.6** Maps depicting the value of assets within the 1 in 100-year floodplain, split by state. The 'current' map (blue) indicates the absolute value of assets within each state's floodplain. The 'future' maps (red) indicate the proportional increase in exposed assets from the present-day to the respective year under a particular scenario.

In contrast to the comparison between our analysis and Aqueduct data for population exposure, our asset damage estimates are smaller than those of the Aqueduct data. Aqueduct estimates that $3.3 trillion of assets are at risk of damage from the 1 in 100-year flood, over double the corresponding value in our analysis. Perhaps surprisingly, with a smaller spatial flood extent, the Aqueduct hazard map translates to a larger estimate of flood risk. We hypothesize that the coarse land-use maps used by Aqueduct do not capture the nuances of where assets are situated in reality. The 5' (~10 km) grid cells, which indicate percentage urban area, used in Aqueduct to drive exposure estimates, are unlikely to be sufficiently

resolved to capture the true distribution of assets. In reality, buildings will generally be more concentrated outside the floodplain than within it, but this is not captured by 5' grid cells. In addition, differences may arise from the depth-damage curves used and asset values assigned.

The New York University (NYU) Furman Centre has merged FEMA flood maps with census block-level housing information, determining the number of housing units within the floodplain (NYU Furman Centre, 2017). The NYU data indicate that 6.8 million housing units fall within the 1 in 100-year floodplain, while our analysis finds that 15.4 million houses are situated there. This illustrates the inadequacies of using incomplete hazard maps (from FEMA) in combination with aggregated socio-economic data (from the US Census Bureau) for flood exposure estimates: the NYU methodology failed to capture even half of the properties on the floodplain identified by our method.

### 4.5.2 Future flood exposure and risk

#### 4.5.2.1 Population exposure

Future population exposure is also detailed in Fig. 4.2. The general trends are perhaps as expected, with greater exposure by 2100 and greater exposure increases in high growth scenarios (e.g. SSP5 vs. SSP2). Absolute future increases in population exposure are naturally higher for larger return periods: an increase of 17.8 million for the 1 in 50-year flood vs. 25.4 million for the 1 in 500-year flood in 2050 under the SSP2 scenario. Percentage future increases confer something more interesting: for the same scenario and year, the 1 in 50-year population exposure increases by 53% while the 1 in 500-year is 41%. This means that more frequently inundated areas are experiencing faster population growth than less frequently inundated ones. In other words, the more hazardous areas in the floodplain (1 in 50-year zone) are projected to experience a higher population growth rate than floodplain as a whole (1 in 500-year zone). Table 4.3 indicates the proportion of the population within each floodplain, where it is apparent that a larger percentage of the total CONUS population will reside in a floodplain in the future. Present-day 1 in 100-year flood exposure stands at 13.3%, but increases to 15.6–15.8% in 2050 and 16.4–16.8% in 2100. It is evident, therefore, that an increased share of future development is projected to take place within the floodplain. The SSP5 scenario places slightly more people on the floodplain as a proportion of total population than SSP2, but the low-growth SSP2 scenario still predicts a proportional increase in population exposure. Fig. 4.4 highlights that this increased exposure is generated across the CONUS as population increases. In particular, the expansion of urban

areas around the Great Lakes, Florida coast, northeast US and Texas see substantial increases in total population exposure.

| Scenario | Return Period (year) | Exposed population | Percentage exposure |
|---|---|---|---|
| 2050 (SSP2) | 50 | 51,344,662 | 13.1 |
| | 100 | 61,214,846 | 15.6 |
| | 500 | 86,842,731 | 22.2 |
| 2050 (SSP5) | 50 | 61,893,832 | 13.3 |
| | 100 | 73,409,641 | 15.8 |
| | 500 | 103,410,662 | 22.2 |
| 2100 (SSP2) | 50 | 63,094,198 | 13.9 |
| | 100 | 74,834,300 | 16.4 |
| | 500 | 104,535,565 | 23.0 |
| 2100 (SSP5) | 50 | 104,633,823 | 14.3 |
| | 100 | 122,615,504 | 16.8 |
| | 500 | 168,454,913 | 23.1 |

**Table 4.3** Future population exposure in the CONUS according to this analysis.

Aqueduct modelled future flood risk scenarios in 2030, while our projections represent changes in 2050 and 2100. In order to compare exposure and risk estimates, 2030 calculations are made by interpolating between our present-day and 2050 values, where yearly increases are assumed to be equal. The selected Aqueduct figures are from the SSP2 socio-economic scenario for 2030 with no hydrological change to make the two datasets as comparable as possible. Differences between Aqueduct data and our analysis for future population exposure have a broadly similar theme to differences in present-day estimates. As shown in Fig. 4.7, our estimate for the number of people exposed to a 1 in 100-year flood is almost triple the equivalent Aqueduct figure. Relative insensitivity of Aqueduct estimates to return period is also displayed once again.

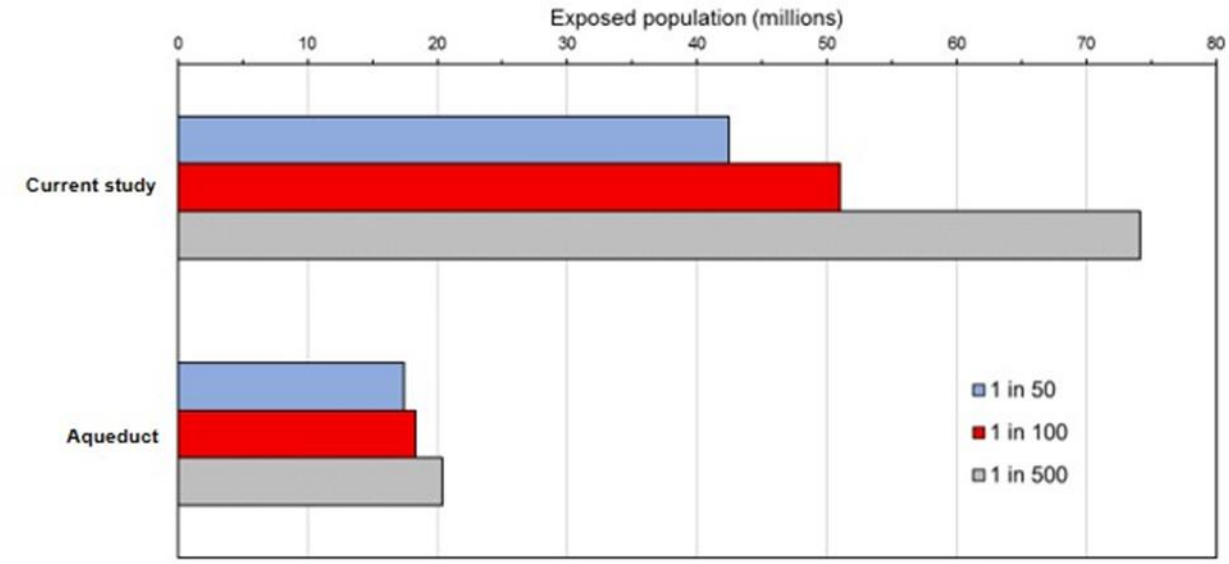


**Figure 4.7** Population exposure estimates for 2030 in CONUS under the SSP2 scenario.

#### 4.5.2.2 Asset exposure and risk

Future trends in asset flood risk are similar to the population-based estimates (Fig. 4.2 and Table 4.4). Notably, the total area of developed land within the 1 in 100-year floodplain in 2100 under the SSP5 scenario is roughly equivalent to the land area of Colorado. This equates to a projected value of exposed assets roughly equivalent to the current GDP of the US. Fig. 4.6 indicates where asset exposure increases are concentrated. In 2050, Minnesota and the Great Plains see substantial proportional increases in assets within the floodplain. In particular, exposed assets at least double by 2050 in Oklahoma. Currently, many of these states are relatively undeveloped, so a large increase in the percentage flood exposure does not necessarily amount to a high absolute increase. California, with $763 billion worth of assets on the floodplain already, sees a 50–100% increase in exposure by 2050. These patterns are even more pronounced by 2100. The currently less-developed Great Plains still experience the greatest proportional increase in exposure, with increases in South Dakota, Nebraska and New Mexico of almost five-fold in relation to the present-day. California, Florida and Texas, already with high absolute exposure, will see within-floodplain asset values triple or quadruple by 2100 under the SSP5 scenario. Interestingly, West Virginia, which is currently proportionally highly exposed with respect to other states (Fig. 4.5), sees very little change in its exposure going into the future. Even under the SSP5 scenario in 2100, the value of assets on West Virginia's 100-year floodplain only increases by one-quarter. New England experiences relatively little increase in asset exposure under both scenarios

analysed. In comparison to our data, Aqueduct simulates significantly higher flood risk in the CONUS (see Fig. 4.8).

| Scenario | Return Period (year) | Area of exposed developed land ($km^2$) | Total value of exposed assets (trillions of USD) | Expected damages (trillions of USD) |
|---|---|---|---|---|
| 2050 (SSP2) | 50 | 174,989 | 6.9 | 1.5 |
| | 100 | 195,981 | 8.1 | 1.8 |
| | 500 | 251,702 | 11.3 | 2.7 |
| 2050 (SSP5) | 50 | 193,023 | 8.1 | 1.7 |
| | 100 | 216,348 | 9.5 | 2.0 |
| | 500 | 277,233 | 13.2 | 3.1 |
| 2100 (SSP2) | 50 | 192,417 | 8.3 | 1.7 |
| | 100 | 215,900 | 9.8 | 2.1 |
| | 500 | 276,956 | 13.6 | 3.2 |
| 2100 (SSP5) | 50 | 240,941 | 10.3 | 2.7 |
| | 100 | 271,106 | 15.5 | 3.3 |
| | 500 | 346,968 | 21.2 | 4.9 |

**Table 4.4** Future asset exposure and risk estimates in the CONUS.

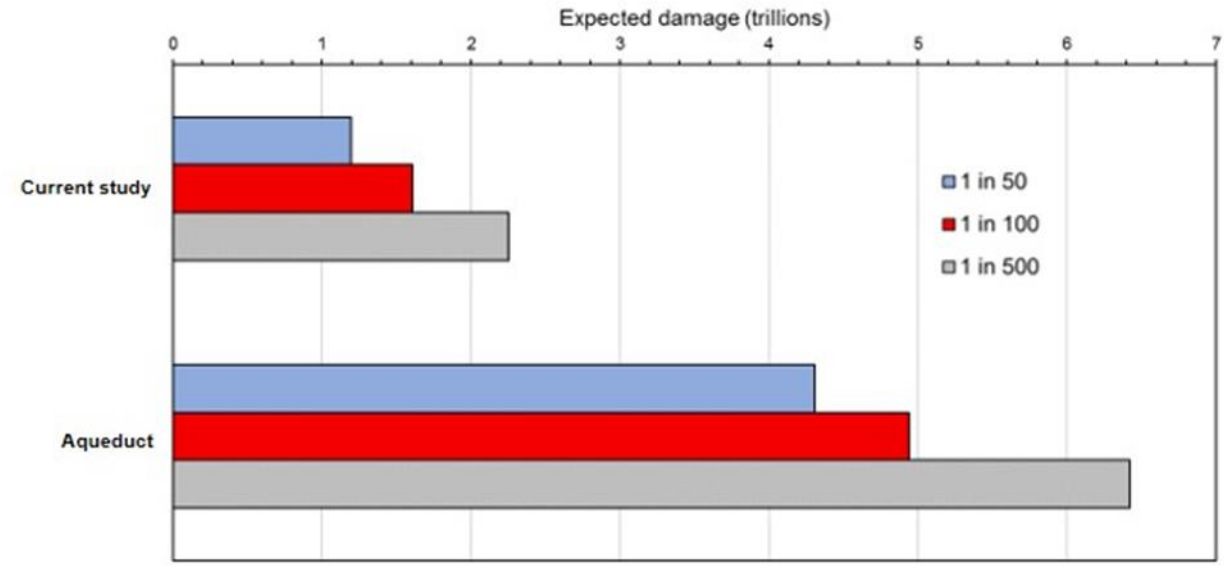


**Figure 4.8** Expected flood damages for 2030 in CONUS under the SSP2 scenario.

### 4.6 Validation

The hazard layers, in isolation, have undergone validation against flood maps produced by FEMA and USGS in a previous study (Wing et al., 2017). This work demonstrated that the continental-scale model used here captured around 90% of the floodplain delineated by these high quality flood maps and, in some instances, was a near perfect match. Validating elements of flood risk analyses that go beyond a specification of flood hazard is, however, a notoriously difficult task. The only tangible estimate produced in this study that can also be measured in reality for validation purposes is flood losses, which are themselves highly uncertain (Downton et al., 2005). The NOAA National Weather Service has compiled annual losses as a result of freshwater (fluvial and pluvial) flooding for over a century. Their annual average loss (AAL) of the 30 year period between 1985 and 2014 amounts to $8.2 billion (inflation adjusted; NWS, 2015).

In order to calculate an AAL from our analysis, an exceedance probability-impact curve must be constructed. AAL is simply the area under this curve, though it can be ‘truncated’ at a certain return period to indicate a flood defence standard (Meyer et al., 2009). In the US, river flood defences are typically built to withstand events with a return period of 1 in 100-years or greater, and storm sewer networks typically protect urban areas against rainfall flooding events with a return period of 1 in 10- or 1 in 20-years. Fig. 4.9 illustrates the exceedance probability-impact curve for our simulations. Assuming floods with a return period lower

than 1 in 100-year are not damaging, this analysis produces an AAL of $12.5 billion; somewhat aligning with NOAA observations. Under this strict assumption, it appears that our framework can replicate observed flood losses to some degree. Relaxing this assumption to a more realistic 1 in 10-year nationwide defence standard (i.e. floods smaller than those with a 1 in 10-year return period cause no damage) generates an AAL of $77.5 billion. Reasons for the differences to the NOAA observed AAL of $8.2 billion could be numerous and are explored further below and in the limitations section.

It is apparent that this model framework has a positive bias towards low return period flood events, likely owing to deficient defence representation. The USACE National Levee Database is known to be incomplete for even large structural defences (ASCE, 2017). It also fails to catalogue more 'informal' berms and other small-scale structures that may unintentionally defend against high-frequency, low-magnitude flood events. Floods in these locations will cause relatively little damage in reality, but may inundate high risk areas in the model. All flood defences, large and small, will naturally be captured by the observed AAL. Similarly, the pluvial model component may not adequately represent the drainage capacity of urban areas, meaning too much of the rainfall during a storm is modelled as hazardous. It is encouraging that there is reasonable agreement between modelled and observed values with the exceedance probability-impact curve truncated at 1 in 100-years (most high-risk areas will be defended to this standard), and it suggests that more accurate estimates of lower return period events can be produced with improved defence representation. With that being said, it is likely that the NOAA observations are negatively biased towards both high and low return period flood events. By their own admission, damages are often underreported and, for small localised flood events, the information might not ever reach the National Weather Service at all. Additionally, it is implicit that the 30-year average will not pick up very low-probability, high-impact flood events, thus underestimating nationwide risk to these. Furthermore, our analysis has calculated loss at one moment in time (the present-day), whereas the NOAA 30-year average loss contains damages from a time when the US was less developed. In other words, a flood simulated by our model may have generated substantial losses in the present-day but not in the 1990s because of urban development over this period. It is therefore prudent not to overstate the verity of the NOAA observations: both sets of data, modelled and observed, are subject to error and provide different, and useful, information. For these reasons, we restrict the focus of our analysis to higher return period floods (1 in 50-

years and above) as here we can be more confident that the model has reasonable predictive skill and that the conclusions we draw are robust.

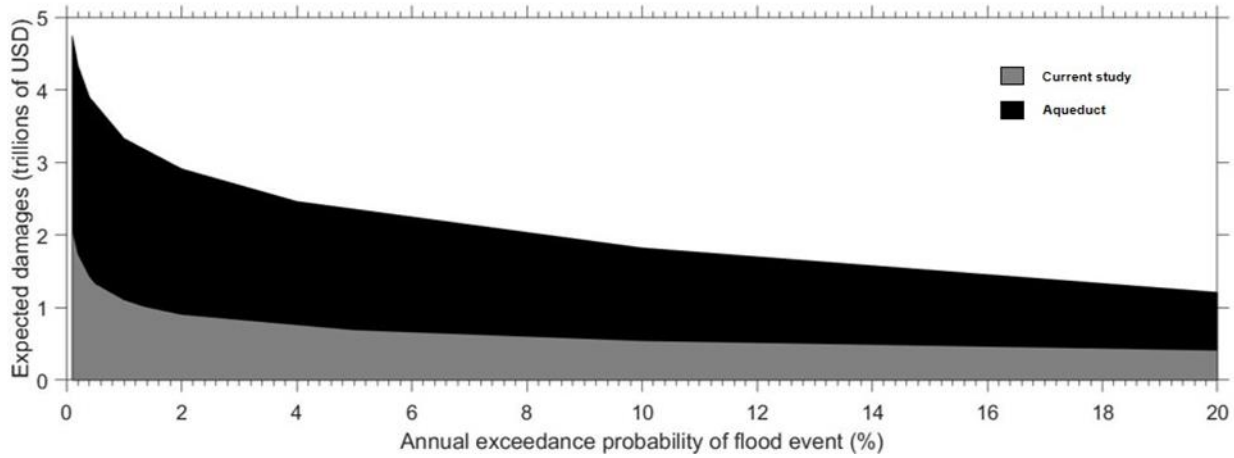


**Figure 4.9** Exceedance probability-impact curves from this analysis and with Aqueduct data.

Comparing the AAL derived from the Aqueduct model, a current state-of-the-art global flood risk framework, to NOAA observations produces a much starker deviation (see Fig. 4.9). Assuming a 1 in 10-year defence standard, the AAL based on Aqueduct data comes to $249 billion. Only when truncating the curve between the 1 in 500- and 1 in 1000-year exceedance probabilities do the NOAA observations converge with an Aqueduct AAL.

### 4.7 Limitations

The flood exposure and risk estimations presented in this analysis are subject to several limitations; the major ones are outlined in this section. As mentioned in section 4.3, some estimates are more uncertain than others (e.g. flooded area is less uncertain than expected damage). This section is structured in such a way that limitations are examined separately where they pertain to hazard, exposure or risk. Once again, we reiterate that these limitations are cumulative, in that the final risk estimation contends with uncertainty in hazard and exposure estimations also. It is also worth noting that all future estimates are significantly more uncertain than those for the present-day. The uncertainties discussed in this section do not invalidate the analyses presented here, instead they should be seen as the foci of future research to continue advancements in this field.

### 4.7.1 Uncertainty in hazard estimation

The underlying terrain data, the USGS NED, has a relative mean point-to-point accuracy of roughly 27 cm at 1" resolution (Gesch et al., 2014), though this will be lower for flatter topography where flooding predominantly occurs. Even though such errors are very low in comparison to other large-scale terrain datasets (e.g. SRTM and ASTER), they will influence the final hazard delineation. Topography is the primary control on patterns of flood inundation, more so than the governing equations of fluid dynamics (Horritt and Bates, 2002), meaning that even these relatively small errors in elevation may influence whether a cell is flooded or not.

Furthermore, the risk estimates will be most sensitive to hazard model performance in more urban areas, which is precisely where there is lower confidence in model output as per chapter 3. Correspondence between FEMA maps and the hazard model dropped substantially in the most built up areas, for a number of possible reasons: (i) the grid resolution of 30 m inadequately captures important fine-scale urban features; (ii) the vertical accuracy of the DEM is lower, particularly because of the need to make it "bare earth"; (iii) flood defences are missing (see below); (iv) the FEMA benchmark data themselves, predominantly consisting of 1D HEC-RAS models which begin to break down in urban areas, are poorer quality; and (v) surface water hazard, which is not captured by the FEMA maps, may be wrongly penalised as false alarms. As such, while confidence may be lower, the benchmark FEMA data in chapter 3 may not be of an appropriate fidelity to faithfully constrain hazard model performance. Meanwhile, the hazard model shows high correspondence with local-scale urban models of the USGS (chapter 3) and the Iowa Flood Centre (chapter 6) and real observations of flooding in Texan cities (chapter 5). As such, the inundation model employed in this analysis is suitable for its reported purpose in the computation of aggregate measures of flood risk across the CONUS.

Although our hazard model has gone further than other continental or global models in the representation of flood defences, it is still lacking in this area. The USACE National Levee Database is incorporated explicitly into the model, and validation studies show that doing this increases model performance markedly in defended areas (Wing et al., 2017). The NLD, however, is estimated to be only around 30% complete (ASCE, 2017), meaning that a number of genuinely defended areas may be inundated in the model. Since defences are built where people and assets are more concentrated, the exposure and risk estimates in this study

will be sensitive to the delineation of flood defences. Unfortunately, this limitation is not easily overcome; the incompleteness of the leading inventory of defence information (the NLD), rather than our methodology, is at fault. This study focuses on large flood events (with return period greater than 1 in 50-years) in order to minimise the effect of this, but improvements to this database and, indeed, the generation of a global one (FLOPROS (FLOod PROtection Standards) shows promise at this early stage (Scussolini et al., 2016)) are a key requirement for future developments in large-scale flood risk analysis.

Extreme flow generation is subject to substantial uncertainty in all flood hazard models irrespective of their scale and coverage. Due to the challenges associated with modelling extreme flows in ungauged catchments (a necessary step in a continental-scale model with total coverage), we opt to use a flood frequency analysis methodology based on river gauge data rather than a rainfall-driven hydrological model. We address the issue of gauge paucity in space by adopting a regionalised flood frequency analysis (RFFA) methodology, which assumes that a gauged and ungauged catchment of similar physical characteristics (e.g. size, climatology, drainage area) will also have a similar flood frequency response (Smith et al., 2015). This approach, too, is subject to high errors: global mean errors of roughly 80%, and in some cases over 300%, have been reported when comparing RFFA-derived discharges to observed ones (Smith et al., 2015). We consider these errors to be unavoidable in models of this scale, and certainly would not be addressed by methodological change (e.g. the use of rainfall-runoff models). Only with increased observation of river flows in time and space can significant advances be made in extreme flow generation. However, our approach does mitigate much of this error by calibrating channel geometry to channel bankfull discharge as estimated by the RFFA (Sampson et al., 2015). This step is crucial as it ensures that any biases in the RFFA are represented in the hydraulic model channel calibration (i.e. if the RFFA has a positive bias for a particular river, the estimated bankfull discharge will be larger than expected and so the channel will be deeper than expected to allow a higher conveyance). We therefore have confidence that the model developed in this study is adequate for the purpose to which it is put and that the conclusions from the analysis are robust.

Further limitations in the generation of extreme flows are derived from the assumption of hydrologic stationarity. The RFFA produces static return period discharges based on river gauge records stretching back at least 30 years. Land-use change and river engineering occurring within the length of the record will likely affect flood frequency and magnitude, yet

the stationary approach assumes historical readings are applicable to the river reach in its current state (Villarini et al., 2009a). If a particular reach exhibits an increase in flooding over time, the RFFA will take into account the smaller flood peaks from when the stream possessed different characteristics. In doing this, present-day flooding would be systematically underestimated (Pinter et al., 2008). Villarini et al. (2009b) examined this effect empirically in areas of the US which became increasingly urbanised during the period of flow measurement. They demonstrated in a river reach in Charlotte, NC, that a 1 in 100-year flood calculated under stationary conditions could feasibly be a 1 in 8- or 1 in 5000-year flood when non-stationarity is accounted for.

Applying non-stationary statistics to boundary conditions in a model of this spatial scale has not been attempted to date. Aside from the computational expense, which is likely to be immense, there is little indication as to how non-stationarity in ungauged catchments could be addressed. It should, however, be seen as a priority in this field to find a solution, with the likely candidate involving a 'scaling up' of present empirical approaches (Villarini et al., 2009a, 2009b; Pinter et al., 2001).

Uncertainty can also be derived from the values assigned to certain model parameters. Often the roughness parameter is the major unknown in a flood model, so it gets calibrated to some event data. However, local calibration is not performed in a model of this spatial scale, owing to the lack of data with which to calibrate. As mentioned in section 4.4.1, the values representing channel and floodplain friction are based on land cover maps and suitable numbers from the literature (Sampson et al., 2015; Chow, 1959). These choices will impact how the model propagates water through channels and over floodplains, but an extensive validation study suggests that appropriate parameter values have been chosen (Wing et al., 2017).

### 4.7.2 Uncertainty in exposure estimation

Exposure estimation, at least for the present-day, is probably the least error-prone stage of this analysis. Despite being sourced from aggregate data (e.g. census-block level), the dasymetric distribution of people in the USEPA EnviroAtlas population map and assets in the FEMA NSI is likely approaching the ceiling for accuracy of such data. There will be considerable uncertainty in the values assigned to assets, though.

Future socio-economic projections will naturally be highly uncertain. We will not ever know for sure what the CONUS will look like in a demographic and developmental sense up to 2100. USEPA ICLUS projections, also distributed dasymetrically, provide plausible, detailed views of future change, though it should be noted that these projections offer only two feasible scenarios amongst infinite possibilities. The method by which particular land-use classes are assigned economic values, however, is another major source of uncertainty. By intersecting the NSI with the NLUD, assigning per-pixel values to individual classes and iterating this across future land-use maps, we have generated reasonable approximations to future asset values, but a number of important factors are not accounted for. In particular, homogeneity of asset value within a certain class is assumed. This means a densely built-up pixel in Los Angeles has the same economic value as one in Detroit: a somewhat unrealistic assumption. Notwithstanding these limitations, our exposure methodology is a substantial improvement on that of previous large-scale estimates and produces worthy estimates for the purposes they are intended for.

### 4.7.3 Uncertainty in risk estimation

Our estimates of flood risk, which is confined to potential flood damage, accumulates the wealth of aforementioned uncertainty as well as that pertaining to depth-damage functions. Estimates of present-day flood risk employ these relationships in a more realistic way than those of the future by applying a specific USACE depth-damage curve for each type of building. In this way, information such as the number of storeys a building has or whether it has a basement are accounted for. Measures of vulnerability, however, depend also on information about the age of a building or what it is built of. Localised information is also critical: a certain depth of water would affect properties in urban New York differently to those in rural Texas. It would be prohibitively difficult to account for such characteristics in a model of this scale. Besides, it is likely that, at the aggregate scale and return periods at which it is applied, our analysis largely cancels out these local errors.

The characteristics of flood hazard that inform the resultant damage estimation is not confined to water depth. Numerous parameters, including velocity, duration, sediment load, and contamination, contribute to flood damage, yet are not represented in this framework (Merz et al., 2010). Studies have suggested that accounting for some of these factors in a flood loss estimation improves the accuracy of results (Apel et al., 2009; Ahmadisharaf et al., 2015), while others suggest that, on balance, their inclusion is not worthwhile (Kreibich et al.,

2009; McBean et al., 1998). Representing these in a flood model of any spatial scale is very difficult. The parameters are not independent and their interaction will have important effects; they are spatially and temporally heterogenous; and their quantitative effect on assets is poorly understood (Merz et al., 2010). It is evident, therefore, that without much more data to define such vulnerability relationships and greater understanding to aid in their implementation, using hazard characteristics other than depth in a framework of this scale would not be appropriate.

The FIA-derived generalised depth-damage curve is applied across the CONUS for future projections, meaning the type of structure is not accounted for. Such information simply is not available for the future, so this approach, with an aggregated relationship weighted towards more frequently flooded building types, has little scope for improvement. Nevertheless, there is still substantial value in the output of this analysis.

Future projections of risk assume temporally static vulnerability: that is, future buildings are not expected to be any more resilient and, crucially, that new developments will not be protected by flood defences. In reality, these are unrealistic assumptions and necessitates future work to build on this first-generation framework. The autonomous adaptation model of Kinoshita et al. (2018) couples socio-economic projections to likely vulnerability scenarios based on historical vulnerability data (Tanoue et al., 2016). Future model development could seek to implement this approach alongside the SSP scenarios employed in this study.

### 4.8 Conclusions

We present the most spatially detailed flood risk estimates, both present and future, of the CONUS to date. Our analysis shows that both FEMA flood maps and previous large-scale risk estimates likely significantly underestimate population exposure, while the latter simultaneously overestimates flood risk. This study is a first of its kind; utilising highly resolved, spatially comprehensive flood hazard information derived from a model that properly represents the physics of flood spreading in combination with high-resolution estimates of present and future distribution of people and assets. With this detailed spatial information on present-day flood risk, federal and state agencies can take appropriate action to mitigate losses (USEPA, 2015). Use of USEPA population and land-use projections means that particular attention can be paid to floodplains where development is projected. Steps to conserve such areas or ensure adequate defences are in place could avoid the exposure of

trillions of dollars of assets, not to mention the human suffering caused by loss of property and life.

### 4.9 Acknowledgements

Our flood hazard layers are available for non-commercial research purposes by contacting Christopher Sampson at Fathom (c.sampson@fathom.global). Components of the framework which generates these layers are available as follows: (i) the LISFLOOD-FP computational hydraulic engine is available from Paul Bates at the University of Bristol (paul.bates@bristol.ac.uk); (ii) Global Runoff Data Centre discharge data from http://www.bafg.de/GRDC/EN/01_GRDC/12_plcy/data_policy_node.html, in combination with the methodology and results of Smith et al. (2015), constitute the boundary conditions; (iii) the US Geological Survey National Elevation Dataset, at http://www.ned.usgs.gov/, is used to represent topography; (iv) to represent the river network, the HydroSHEDS global hydrography dataset (http://www.hydrosheds.org/) is used; (v) National Oceanic and Atmospheric Administration Intensity-Duration-Frequency relationships which drive the pluvial model component are sourced from http://hdsc.nws.noaa.gov/hdsc/pfds/; (vi) satellite luminosity data from http://www.ngdc.noaa.gov/eog/ and (vii) soil data from the Harmonized World Soil Database of the Food and Agriculture Organization of the United Nations (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/) are used to modulate infiltration in the pluvial model; and (viii) the USACE National Levee Database is available from http://www.nld.usace.army.mil/. Exposure data from the USEPA, which includes present-day population distributions and ICLUS projections of population and land-use, are available from the USEPA's Environmental Dataset Gateway: http://edg.epa.gov/. The National Land-Use Dataset of the present day can be accessed at http://csp inc.org/public/NLUD2010_20140326.zip. The FEMA National Structure Inventory is available from http://data.femadata.com/FIMA/NSI_2010/. The G-Econ GDP per capita map of the US can be accessed at http://gecon.yale.edu/. Depth-damage relationships are available upon request from Oliver Wing at the University of Bristol (oliver.wing@bristol.ac.uk). Paul Bates was supported by a Royal Society Wolfson Research Merit Award and a Leverhulme Research Fellowship. The work in this paper was in part supported by both UK Natural Environment Research Council grant NE/M007766/1 and the Kresge Foundation's generous support of The Nature Conservancy.

# 5 A FLOOD INUNDATION FORECAST OF HURRICANE HARVEY USING A CONTINENTAL-SCALE 2D HYDRODYNAMIC MODEL

*This paper-chapter was published in Journal of Hydrology X in 2019:*

**Wing, O. E. J.**, Sampson, C. C., Bates, P. D., Quinn, N., Smith, A. M., and Neal, J. C. (2019), A flood inundation forecast of Hurricane Harvey using a continental-scale 2D hydrodynamic model, *Journal of Hydrology X*, 4, 100039, doi:10.1016/j.hydroa.2019.100039.

*Author contributions:*
O. Wing led the study, conceived and generated the fluvial forecasts using existing hazard maps generated by C. Sampson and A. Smith, processed the benchmark data, performed all analyses, and wrote the manuscript. C. Sampson and N. Quinn developed the pluvial and coastal forecasts. All authors aided in the conceptualisation of the research and commented on manuscript drafts.

## 5.1 Background

This chapter addresses thesis objectives 1b, 1c and 2b:

> *Perform a comprehensive validation of flood inundation modelled over large spatial domains by comparison to real-world flood event observations, thus necessitating coupling the hydraulics to meteorological and hydrological models, and through evidencing increased accuracy relative to low-complexity, zero-physics approaches.*
>
> *Explore new applications for a hydraulic modelling approach which avoids the pitfalls of low-coverage local models and current low-accuracy large-scale models through the generation of operational flood inundation forecasts.*

The model framework presented in chapter 3 and Sampson et al. (2015) has more utility than just the provision of event-agnostic floodplain delineations. This chapter charts further developments to the framework where a simple coastal model is presented, AEP pluvial rainfall inputs are supplanted by forecast ones, and fluvial modelling involving sampling from the pre-computed AEP catalogue informed by hydrological forecast output. In doing so, the large-scale model framework is shown to be fit-for-purpose in highlighting impacted areas from incoming US storms. This coupling of accurate hydraulics to meteorological–

hydrological model cascades led to wide-area event replication, thus enabling a first fair comparison between a large-scale model and real event observations. This chapter finds good correspondence between observed and modelled extents, but a relatively large deviation in observed and modelled water levels broadly explained by DEM quality. A simple planar approximation is shown to have very low skill in replicating flood event extents; outperformed considerably by the hydrodynamic model.

## 5.2 Introduction

Flood events are among the most costly and deadly natural disasters on the planet: since 1980, they have caused economic damages of over $1 trillion and 220,000 deaths worldwide (Munich Re, 2018a). Recent devastating events, particularly flooding (both freshwater and saltwater) arising from tropical cyclones, have sharply focused the issue in the minds of the public and policy makers alike. In 2017, Hurricanes Harvey, Irma and Maria collectively caused $220 billion of damage in the Gulf of Mexico; Typhoon Haiyan claimed the lives of over 6000 people and caused $10 billion worth of damage in East Asia in 2013; and in 2008, Cyclone Nargis caused 140,000 fatalities and $4 billion of economic damage in Myanmar (Munich Re, 2018b). A growing body of evidence suggests that such tropical cyclones will become more intense (Kang and Elsner, 2016; Sobel et al., 2016; van Oldenborgh et al., 2017; Emanuel, 2017) and move more slowly once they make landfall (Kossin, 2018) as a result of climate change. With more precipitation falling over a longer duration, tropical cyclone driven flood impacts are likely to increase in the future. On top of the freshwater component of such events, the low atmospheric pressure and strong onshore winds arising from tropical cyclones result in coastal inundation. Regardless of potential changes to these storm characteristics under climate change, increased sea levels in a warming world are likely to exacerbate coastal flood impacts.

In light of this, there is a clear need for substantial risk reduction measures to mitigate against present and future flood consequences. One facet of such measures is improved flood forecasting, which permits a short-term response to be mounted (e.g. temporary defence erection, evacuations, first-responder preparedness, reinsurance purchasing). A typical generic flood forecasting approach can be conceptualized as a source-pathway-receptor framework (e.g. Narayan et al., 2012) as outlined in Fig. 5.1.

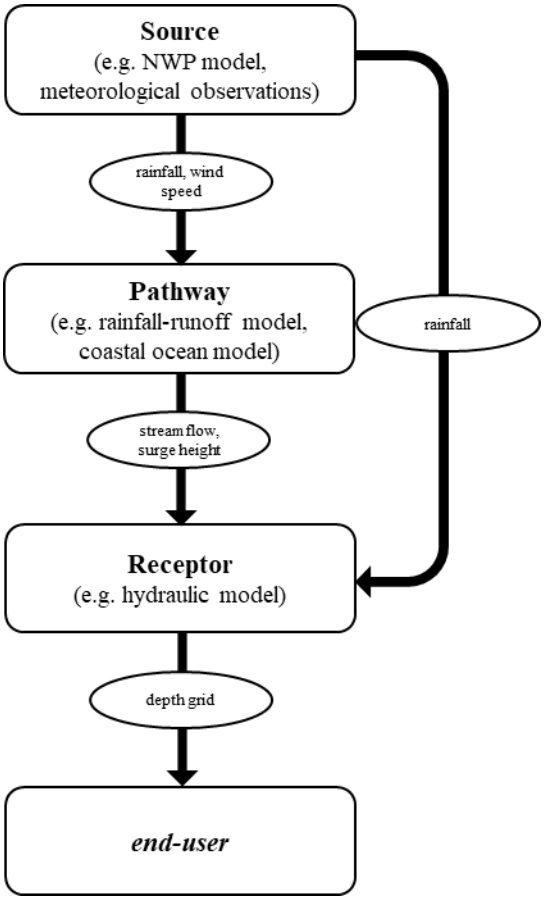


**Figure 5.1** Schematic of a source-pathway-receptor flood forecasting cascade. Arrows indicate the flow of data (with examples in ovals), where the outputs of one component (specified in rectangular boxes) become the inputs of another.

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

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

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

would greatly benefit from detailed, local predictions of flood extent and depth to enable a more complete risk calculation.

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

Here we present a medium-term (2 – 15 days) tropical cyclone flood inundation forecasting product which is capable of being used in an operational system, and demonstrate that high-resolution hydraulic models can now be practicably employed at large-scale in such frameworks where accurate local forecasts are lacking. We forecasted Hurricane Harvey by coupling streamflow, rainfall and storm surge predictions from NOAA to Fathom-US (Wing et al., 2017), a continental-scale hydraulic flood model of the US, and used this to produce daily flood depth footprints at ~30 m resolution. Updating the hydrologic inputs with post-event observations as they became available, we also produced a model hindcast in real-time. After the event, we tested the forecast and hindcast model against ground observations and derived flood extents from the US Geological Survey (USGS) (Watson et al., 2018). The results of this validation procedure are compared to those obtained when a simpler Height

Above Nearest Drainage (HAND) model from the NOAA National Water Centre (NWC) is used instead of the hydrodynamic model (NOAA NWC et al., 2018).

## 5.3 Methods and data

### 5.3.1 Hydrodynamic model (Fathom)

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

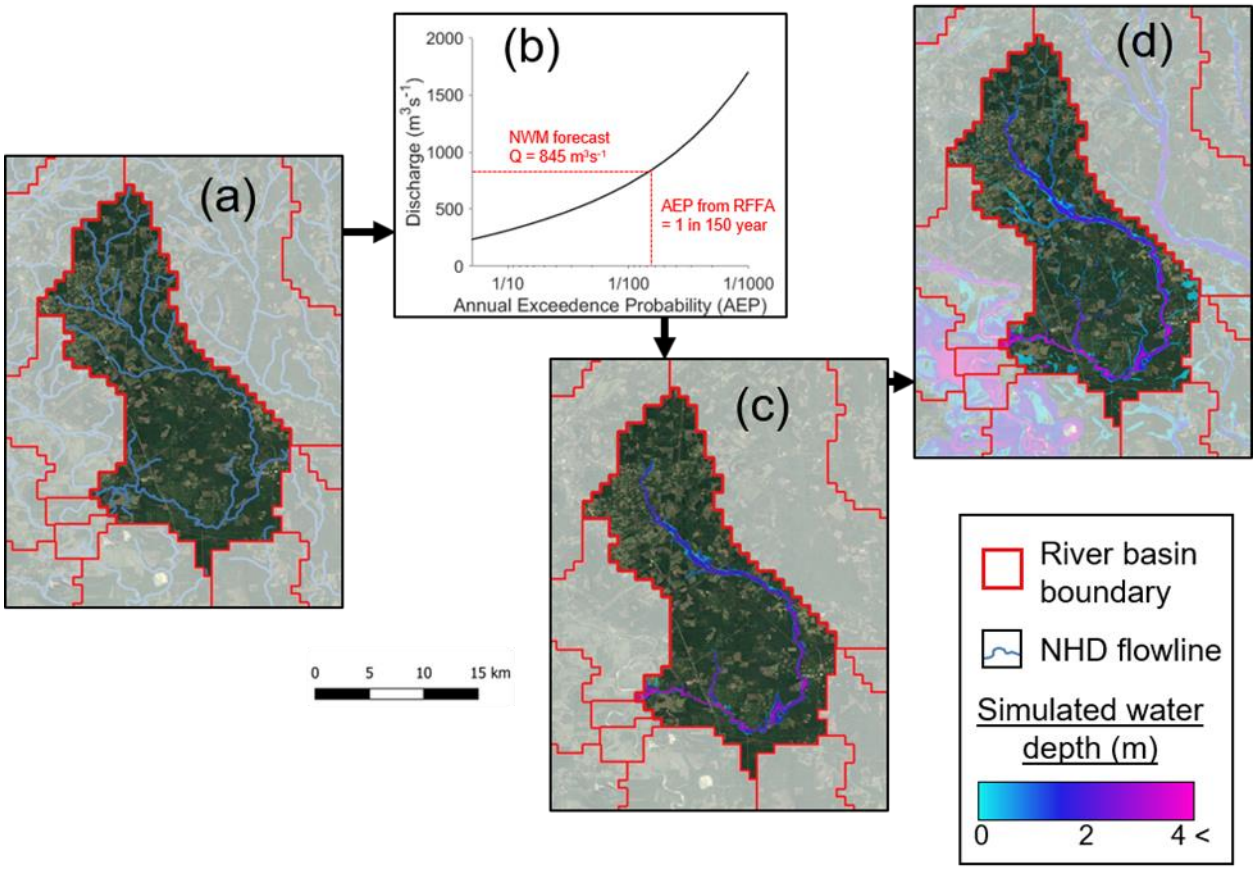


**Figure 5.2** An example demonstrating the process of extracting fluvial flood maps for each river basin. The unshaded river basin (ID = 17944) in **(a)** contains Menard Creek, a tributary of the Trinity River ~100 km NE of Houston, and has a drainage area of ~500 $km^2$. Streamflow is forecast for each flowline from the USGS NHD by the NOAA NWM. The graph in **(b)** shows the regional flood frequency analysis at the outlet of basin 17944. The NWM 3-day forecast peak discharge was 845 $m^3$ $s^{-1}$ from a model run executed on 27 August 2017. This corresponds to the 1 in 150 year streamflow in this basin, whose inundation has already been simulated in the Fathom-US library as shown in **(c)**. The depth grid is then extracted for this catchment. The final flood inundation map is shown in **(d)**, where this process has been repeated for all river basins in the domain and integrated with the new pluvial simulations to represent headwater (rivers with drainage area <50 $km^2$) and surface water hazard.

For the fluvial inundation forecast, flood depths within a given river basin, as defined by HydroBASINS (Lehner and Grill, 2013), are extracted from an existing library of nationwide flood maps at ~30 m resolution (Wing et al., 2017). These fluvial flood maps are driven by discharges from a regional flood frequency analysis (RFFA) to ensure flood inundation is

simulated on every US river, meaning multiple maps corresponding to a certain annual exceedance probability (e.g. 20% to 0.1% AEP) have been generated for the whole country. For each river basin in the study area, the NWM streamflow forecast is assigned an AEP, based on basin-specific data from a US variant of the global RFFA of Smith et al. (2015) using USGS river gauges, and the relevant AEP flood map is extracted from the library. Fig. 5.2 visualizes this process. By sampling from pre-existing flood maps of the entire continental US, simulation quality (e.g. grid resolution, physical process representation) is not hampered by the need for a forecast with reasonable latency: the extraction process takes only seconds (Leedal et al., 2010). For fluvial flooding, the response to extreme streamflow in a river basin is relatively well-confined to that particular locality, meaning any plausible flood event will match a pre-computed, event-agnostic inundation simulation of that area provided enough runs with different boundary conditions are executed. How rainfall and surge vary in time and space, however, is less related to the hydrologic conditioning of the ground surface. In other words: simulating these perils ahead of time is inadvisable, given the *a priori* specification of suitable extraction zones is less defensible than for fluvial flooding. As such, the NOAA rainfall and surge information is input to Fathom-US to generate new event-specific depth grids, requiring ~6-hour parallel simulations of coastal and pluvial flooding for a ~400,000 $km^2$ domain at ~30 m resolution on a single node of 20 cores (Intel Broadwell E5 Xeon). Alongside model-building and post-processing, the final product (maximum flood depth in a pixel from the fluvial, pluvial and surge models) was produced ~24 hours after the release of the NOAA data, meaning the total lead time was ~2 days. The forecasts were updated with observed boundary conditions to produce a hindcast so that a more accurate delineation of impacted areas could be identified with immediacy. These boundary condition observations consisted of 24-hour rainfall totals from the NOAA NWS (https://water.weather.gov/precip/download.php), peak flows observed at USGS river gauging stations (https://waterdata.usgs.gov/nwis/rt) and coastal water level observations from the NOAA Tides and Currents service (https://tidesandcurrents.noaa.gov/gmap3/). During the Harvey event, the forecast and hindcast flood extent data produced by the system were made freely available to first-responders and were used operationally by NASA and insurers to rapidly assess exposure.

Fathom-US is a large-scale hydraulic modelling framework, whose freshwater flooding component is set out in Wing et al. (2017). The US implementation is itself a variant of the global model first described by Sampson et al. (2015). It has been rigorously tested against

thousands of bespoke flood inundation studies carried out by US government agencies, concluding that the large-scale methodology is approaching the accuracy of traditional local models (Wing et al., 2017). The framework permits rapid construction and execution of model domains as defined by the forecast footprint of US hurricane landfalls. Its computational hydraulic engine is driven by a variant of the LISFLOOD-FP hydraulic model, which solves a local inertial form of the shallow water equations in two dimensions over a regular grid (~30 m: 1 arc second resolution in this case) using a highly efficient numerical solution (Bates et al., 2010; de Almeida and Bates, 2013). This grid is populated with elevation values from the 1 arc second version of the USGS National Elevation Dataset (NED), with levee information from the US Army Corps of Engineers National Levee Database explicitly represented. This Digital Elevation Model (DEM) has complete coverage of the conterminous US and is thus the crucial component of the framework's applicability to simulate inundation for all potential hurricane landfall locations in the US. River hydrography is represented by HydroSHEDS (Lehner et al., 2008); with those rivers wider than the grid resolution being burnt directly into the DEM, and narrower streams being represented by the sub-grid 1D model of Neal et al. (2012b). Fluvial modelling is executed by inserting river discharge information (which is ultimately linked to an RFFA-derived AEP, an NWM streamflow forecast or a USGS river gauge observation) at the relevant inflow points in the stream network, while the pluvial model takes spatial rainfall data and drops it directly onto the land surface ("rain-on-grid"; Sampson et al., 2012). Assumptions relating to infiltration capacity are made based on soil information from the Harmonized World Soil Database in conjunction with a modified Hortonian infiltration equation (Morin and Benyamini, 1977). In urban areas, identified using satellite luminosity data (Elvidge et al., 2007), the infiltration capacity is defined using assumed urban drainage design standards. The storm surge model component was conceived by Bates et al. (2005), who adapted the 2D LISFLOOD-FP code traditionally used in fluvial settings for coastal flooding. Using LISFLOOD-FP for such an application has precedent: as examples, Smith et al. (2012) evaluated its suitability for coastal inundation modelling in the UK and Quinn et al. (2014) applied the code in a coastal flood risk assessment. In this component, the coastal boundary line of the model domain is set within oceanic cells just offshore of the coastal flood defences. For each cell along the coastal boundary, the predicted peak surge height was extracted from P-surge output and used to scale the tidal time series at that location to create a fractional surge height time series. Water enters the model domain along the boundary in accordance with this time series, and the hydraulic model simulates the dynamics of the surge

as it interacts with the shoreline and moves inland. 2D (burnt in the DEM) and 1D (sub-grid) channels are represented here also, meaning the ability of the storm surge to propagate inland via river channels is properly represented: a crucial component of coastal flood models (e.g. Maskell et al., 2014).

### 5.3.2 NWM-HAND model (NOAA NWC)

The NOAA National Water Centre has recently explored the coupling of the National Water Model to the Height Above Nearest Drainage method (Rodda, 2005; Rennó et al., 2008; Nobre et al., 2011, 2016). This effort forms part of the National Flood Interoperability Experiment (Maidment, 2016): which, like this paper, seeks to append the neglected receptor component to the flood forecasting cascade (Fig. 5.1). The HAND approach normalizes the DEM so that a given pixel takes the value of the vertical distance to the stream it drains to. A rating curve is then used to translate NWM flow forecasts to stage for a given river cell, and any land cells that drain to this stream location and have a HAND value less than the stage become flooded. Planar approximations such as these, which do not consider flow physics, have been shown to be less skillful than models which represent the dynamics of flood inundation since the inception of raster-based hydraulic modelling (Bates and De Roo, 2000) and again more recently (Afshari et al., 2018), owing to their omission of mass and momentum conservation laws. In isolated test cases however, particularly on confined floodplains with steep valley sides and straight river reaches, a planar approach may offer satisfactory performance (Bates and De Roo, 2000). Furthermore, the reduction in model skill may be considered acceptable where rapid solutions are required to large computation problems (Afshari et al., 2018; Liu et al., 2018). Geographies where these methods are appropriate are difficult to specify *a priori* though, since they have not undergone the same level of wide-area testing as hydrodynamic approaches (e.g. Wing et al., 2017). In evaluating the simpler HAND-based approach alongside the more complex hydraulic model presented here, the trade-off between including the physics of water flow in a forecast model and computational efficiency can be quantified.

The NWM-HAND model was executed for Hurricane Harvey (NOAA NWC et al., 2018) using the USGS NED at ⅓ arc second resolution as the DEM. The model was run each day with NWM analysis and assimilation outputs, providing a snapshot of the flood extent caused by Harvey at that time. By taking the maximum extent from all of these HAND-based

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

### 5.3.3 Validation data

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

### 5.3.4 Validation metrics

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

$$\text{Hit Rate} = \frac{M_1 B_1}{M_1 B_1 + M_0 B_1} \tag{5.1}$$

$$\text{False Alarm Ratio} = \frac{M_1 B_0}{M_1 B_0 + M_1 B_1} \tag{5.2}$$

$$\text{Critical Success Index} = \frac{M_1 B_1}{M_1 B_1 + M_0 B_1 + M_1 B_0} \tag{5.3}$$

$$\text{Error Bias} = \frac{M_1 B_0}{M_0 B_1} \tag{5.4}$$

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

The 'hit rate' metric (HR; Eq. 5.1) penalizes type II errors and is thus a measure of the model's tendency to underpredict the benchmark flood extent. It can be interpreted as the proportion of benchmark flooded areas that were replicated by the model. The 'false alarm ratio' (FAR; Eq. 5.2) penalizes type I errors and so represents the tendency of the model to overpredict the benchmark flood extent. This metric can be interpreted as the proportion of modelled flooded areas that are dry in the benchmark. The 'critical success index' (CSI; Eq. 5.3) penalizes both type I and type II errors, thus being a metric that accounts for both under and overprediction. It ranges between 0 (no match between model and benchmark) and 1 (perfect match between model and benchmark) and can be thought of as representing the model performance over floodplain areas only as it excludes areas that do not inundate in both the model and benchmark data. Finally, 'error bias' (EB; Eq. 5.4) is the ratio of type I to type II errors. Values greater than 1 indicate the model tends to overpredict, while values less than 1 indicate a tendency to underpredict with respect to the benchmark.

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

$$\text{Root Mean Squared Error} = \sqrt{\frac{\sum_{n=1}^{N}(O_n - M_n)^2}{N}} \tag{5.5}$$

$$\text{Mean Absolute Error} = \frac{\sum_{n=1}^{N}|O_n - M_n|}{N} \tag{5.6}$$

$$\text{Mean Error} = \frac{\sum_{n=1}^{N} O_n - M_n}{N} \tag{5.7}$$

where $O_n$ and $M_n$ is the WSE at a given observation point and corresponding model cell respectively, *N* is the number of HWMs analysed and everything else is as above. The original dataset was trimmed down to 1134 points so that the analysis was confined to high

quality HWMs that were not taken at the same location and which were referenced against the same geodetic datum.

Both Root Mean Squared Error (RMSE; Eq. 5.5) and Mean Absolute Error (MAE; Eq. 5.6) measure the average magnitude of the errors (where an error is $O_n - M_n$). RMSE is a quadratic scoring rule (meaning greater weight is given to larger errors), while MAE is linear (all errors have equal weight). Mean Error (ME; Eq. 5.7) calculates the average error whilst still accounting for their sign: a negative ME indicates the model has a tendency to overpredict observed WSEs, while a positive one suggests underprediction.

## 5.4 Results and discussion

### 5.4.1 Flood extent comparison

The results of the flood extent comparison between the Fathom hindcast model (driven with observed streamflow, rainfall and surge heights) and USGS benchmarks for each of their fourteen study sites are shown in Table 5.1a and Fig. 5.3. Most of the USGS flood extent is captured by the model, with 78% of observed wet pixels being correctly identified as such in the model on average across all sites. Many of the sites with high HRs also have relatively high FARs (e.g. Tres Palacios River (Fig. 5.3a), Lower San Bernard River (Fig. 5.3l) and Upper Brazos River (Fig. 5.3b), driving a lower overall correspondence and overpredictive bias as indicated by their CSIs and EBs respectively. Conversely, many lower-HR sites have very low FARs (e.g. Upper and Lower Neches River (Fig. 5.3k and 5.3m) and Cow Bayou (Fig. 5.3f)), generating comparable CSIs and EBs < 1 indicating underprediction. Some high-HR sites have correspondingly high CSIs, owing to low FARs (e.g. Lower Brazos River (Fig. 5.3e) and Middle San Bernard River (Fig. 5.3j), suggesting there is a strong match between model and benchmark flood extents. CSIs, which are the most discriminatory metric, range from a poor 0.5 (the model is correct as often as it is incorrect) on the Upper San Bernard River (Fig. 5.3i) to an excellent 0.9 (a 90% match between model and benchmark) on the Lower Brazos River (Fig. 5.3b). Across the domain, an average CSI of 0.66 indicates that roughly 2 in every 3 model pixels in the functional floodplain match the benchmark. Further to the details in section 5.3.3, the USGS benchmark flood inundation extent is subject to uncertainty. It is generated via interpolation between point HWMs (themselves containing vertical error), rather than a genuine 2D observation, meaning the accuracy of these maps is heavily dependent on the spatial resolution of the HWMs. Without representing the physics of flow that generated the HWMs in a hydraulic model (e.g. timing and interaction of

adjacent streamflows), the interpolation procedure may produce unrealistic inundation extent boundaries where unconstrained by point observations. To put the pattern scores into context: the Wing et al. (2017) model attained average CSIs of ~0.75 against detailed local models, with a maximum of ~0.90; global flood models average CSIs of ~0.50, obtaining up to ~0.70 in isolated test cases, against a variety of local models and satellite-derived flood extents (Sampson et al., 2015; Winsemius et al., 2016; Dottori et al., 2016; Ward et al., 2017); local flood models that have been manually built and extensively calibrated generally achieve CSI scores of 0.70 – 0.80 when compared to satellite observations of flood inundation (Aronica et al., 2002; Pappenberger et al., 2007a; Di Baldassarre et al., 2009; Wood et al., 2016) and up to 0.9 when benchmarked against very high quality data (e.g. Bates et al., 2006; Altenau et al., 2017). Fleischmann et al. (2019) propose that a hydrodynamic model provides locally relevant estimates of flood extent when CSI > 0.65. It should be noted that the CSI metric is sensitive to the selected study area, however: favouring overpredictive models of larger floods on flat terrain compared to the reverse case (Stephens et al., 2014). The results shown here are towards the higher end of those in the literature (though many of these are for smaller floods where high CSIs are difficult to obtain). The hindcast model performance is, however, shy of the very high CSIs exhibited by calibrated local model – observation comparisons. The results thus indicate that, across the model domain, the hindcast hydraulic model presented here has some skill in replicating benchmark patterns.

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

**(c)**

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

**Table 5.1** A comparison of **(a)** the Fathom model hindcast and **(b)** maximum flood extent from the NWM-HAND model hindcast of the NOAA NWC to the benchmark USGS flood extents. The metrics Hit Rate (HR), False Alarm Ratio (FAR), Critical Success Index (CSI)

and Error Bias (EB) are explained by Eqs. 5.1–5.4 in section 5.3.4. To aid interpretation, **(c)** indicates the colour scale used to classify each measure: the darker the colour, the higher the performance.

The results of the comparison between the NWM-HAND model and the USGS benchmark are shown in Table 5.1b. It is evident here that the NWM-HAND model has lower predictive skill when benchmarked against the USGS flood extents. Mean HRs indicate 46% of flooded areas are correctly captured, with mean CSIs suggesting just over 4 in every 10 pixels in the functional floodplain are identified correctly. It should be noted that the NWM-HAND model structure is only capable of representing fluvial flood hazard, meaning both hydrologically-isolated flooding from intense local rainfall and coastal surge will not be captured. Hurricane Harvey was predominantly a pluvial event: flooding arose in many areas due to intense local rainfall on the land surface, rather than from rivers flowing out of bank. Fig. 5.4 shows how different the Fathom and NWM-HAND models look when this runoff component is represented in an area to the NE of Houston. HAND being a fluvial-only model perhaps explains why coastal basins (e.g. two sites in Matagorda Bay) have such low scores, since storm surge may have played an important role here. The lack of a rainfall component may explain poor performance on small streams: the Upper San Bernard River has a drainage area of ~500 $km^2$ and the San Jacinto River site contains headwater streams with drainage areas as small as ~20 $km^2$, and both exhibit extremely poor performance with NWM-HAND. The flood hazard arising on these small streams is not driven by traditional fluvial flooding processes, where aggregation effects lead to a large, low-amplitude flood wave which propagates downstream, but by the rapid lateral surface flow of intense local rainfall generating a flash flood. This further contextualizes the scores obtained by the Fathom model at these sites too, where, despite possessing a pluvial model component, the difficulties in simulating this phenomenon are evident. With that being said, for even the relatively simple problem of modelling the Brazos River, which has a flat floodplain confined by steep valley sides, NWM-HAND correctly identifies less than half of the pixels. The hydrodynamic Fathom model, in contrast, correctly identifies between 74% and 90% of flooding from the Brazos. For reference, CSIs obtained by HAND models have been reported between 0.5 and 0.9 for a selection of watersheds in the US when comparing HAND approaches to observational data or hydrodynamic methods (Zheng et al., 2018a, 2018b; Zhang et al., 2018; Afshari et al., 2018). The results from the NWM-HAND model presented here do not appear

to be as skillful as those presented in these smaller scale studies. Further, despite this approach making use of higher resolution (~10 m) terrain data, it is outperformed by a coarser (~30 m) hydrodynamic model. This suggests that the importance of higher grid resolution is only realised when physical processes are represented.

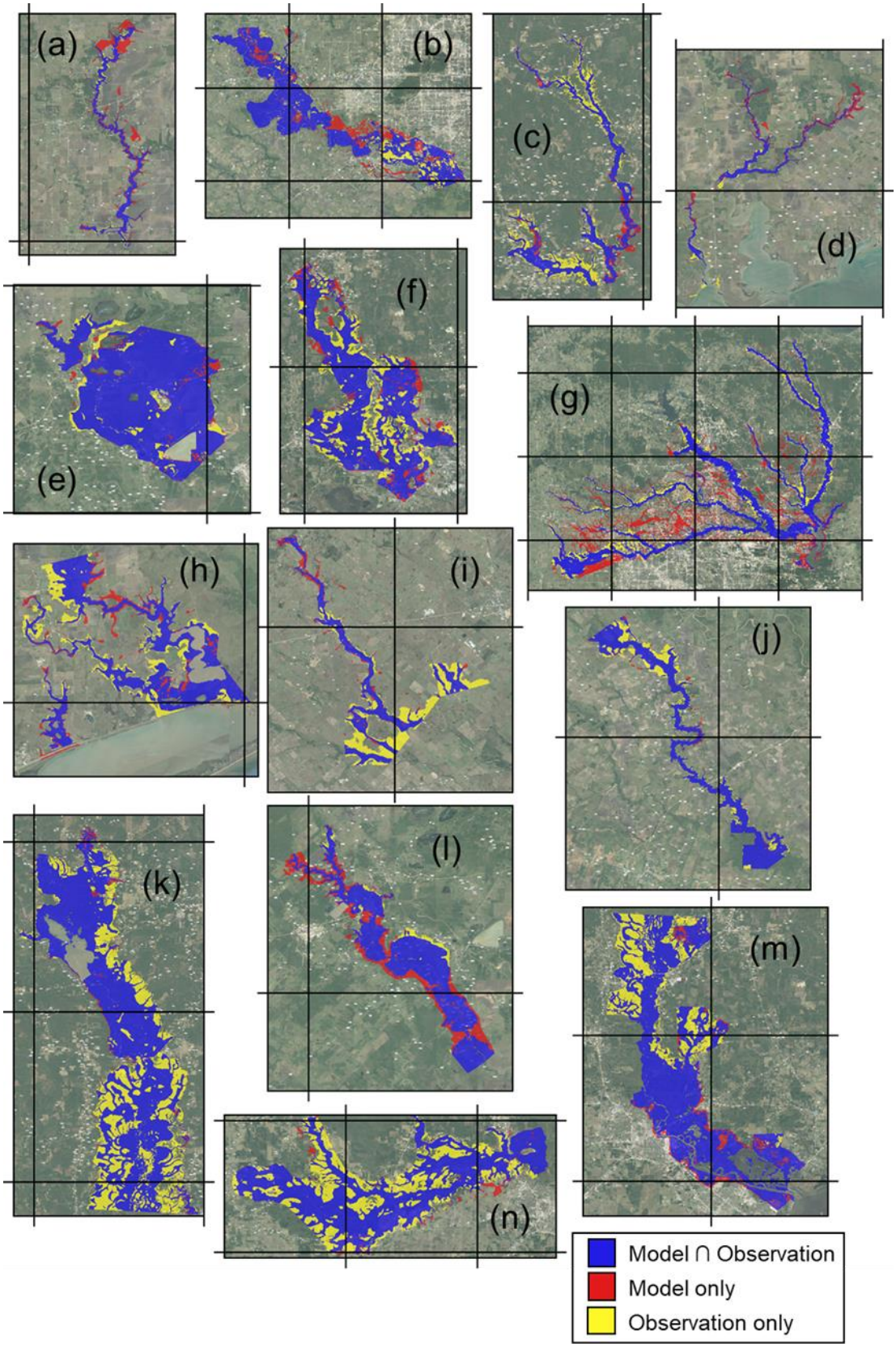


**Figure 5.3** Maps displaying the intersection of the Fathom hindcast modelled Hurricane Harvey flood extent with those generated by the USGS based on observed HWMs at 14 sites: **(a)** Tres Palacios River, **(b)** Upper Brazos River, **(c)** Big Cow Creek, **(d)** Matagorda Bay, **(e)**

Lower Brazos River, **(f)** Cow Bayou, **(g)** San Jacinto River, **(h)** East Matagorda Bay, **(i)** Upper San Bernard River, **(j)** Middle San Bernard River, **(k)** Upper Neches River, **(l)** Lower San Bernard River, **(m)** Lower Neches River and **(n)** Pine Island Bayou. Grid scales in all panels are 0.25° (~27 km).

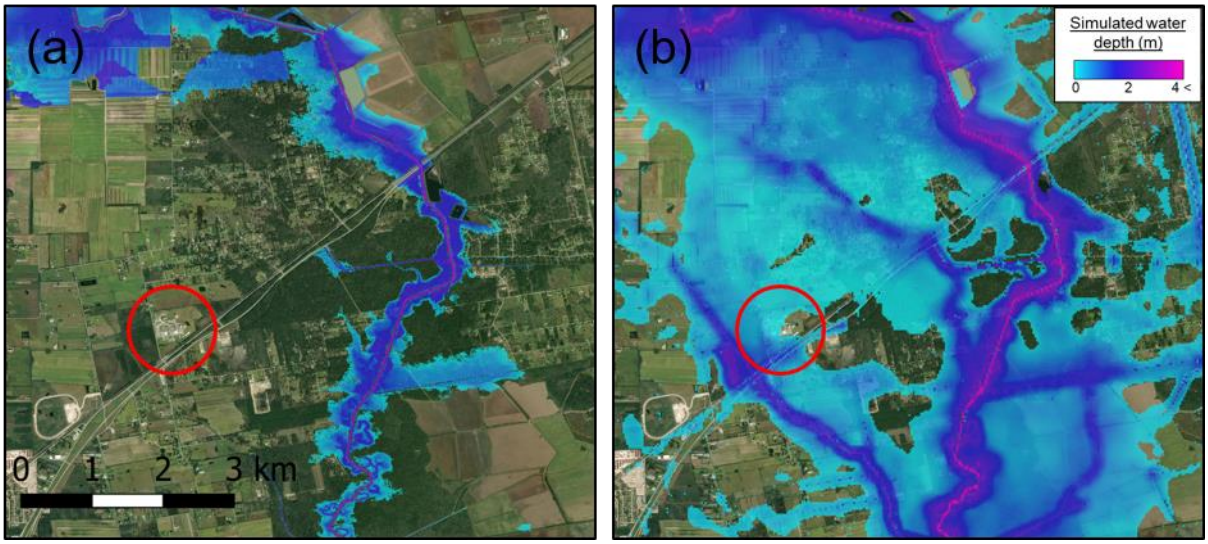


**Figure 5.4** An area near Crosby, NE of Houston, exhibiting the output of **(a)** the NWM-HAND and **(b)** the Fathom hindcast models. The Arkema chemical plant is within the red circle, which was inundated during Harvey and led to an industrial disaster.

Further to testing the Fathom and NWM-HAND models that were driven with real-time observations, the Fathom forecast variant was also benchmarked against the USGS flood extents. It was driven with NWM peak flows from a 3-day forecast commencing at 1800 GMT-6 on 27th August 2017, as well as forecast rainfall and surge data from this time. Hurricane Harvey generated peak streamflows across the domain for over 5 days: beyond the horizon of the 3-day forecast model presented here. The forecast model output chosen for validation here captures (temporally) the main hurricane impact in Texas, which was felt prior to 31st August 2017. The results of this comparison are shown in Table 5.2, where the Fathom forecast model sees a performance drop with respect to its hindcast variant: mean CSI drops by 0.09 and mean HR drops by 0.03. While HRs actually increase for most sites when compared to hindcast performance (as high as 99.9% on the Lower San Bernard River), this is amidst a backdrop of increased EBs that indicate a heavy bias towards overprediction (mean EB of 84.02, compared to a more modest 1.42 in the hindcast). It's worth noting here,

though, that EBs are measured on a logarithmic scale. Since underprediction is indicated by values less than 1 and overprediction by values greater than 1, a value of 0.5 has the same magnitude bias as a value of 2, yet the mean of these two values is 1.25 (not 1). Underrepresented by this mean EB, therefore, are the extremely high underpredictive biases evident on the Lower and Upper Neches River, and Cow and Pine Island Bayous. The use of forecasted, rather than observed, model inputs is therefore widening the tails of the bias distribution (EBs moving further from 1) and reducing overall performance (CSIs decreasing). With that being said, capturing, on average, three-quarters of benchmark inundated pixels (HR) while keeping false alarms at 23% (FAR) is indicative of fair performance in the forecast model.

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

manageable task for the HPC facilities of leading forecast centres and so probabilistic depth grids of these flood drivers could be constructed too. Incorporating the hydrodynamic model in an EPS would be a relatively straightforward addition to the method proposed here, but is one which is beyond the scope of the current paper.

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

**Table 5.2** A comparison of the forecast variant of the Fathom model to the benchmark USGS flood extents. The color scale used is explained in Table 5.1c.

Digging deeper into the forecast uncertainties, observations of peak streamflow from 63 USGS river gauges across the footprint domain are compared to the corresponding NWM streamflow forecast from 27th August 2017. All river gauges used experienced peak flow during this forecast's 3-day time horizon. By volume, the mean absolute error (as in Eq. 5.6, but for discharge rather than water surface elevation) comes to 2970 $m^3 s^{-1}$. For reference, this error is equivalent to roughly 80% of the peak discharge experienced on the Lower Brazos River during Hurricane Harvey. Errors as a proportion of the observed discharge are shown in Fig. 5.5. The MAE is 290%, with a very high bias towards overprediction (mean error of –281%). Forecast discharge on the Buffalo Bayou near Addicks was an eighteenfold overestimate: an implausible 7129 $m^3 s^{-1}$ when 390 $m^3 s^{-1}$ was observed. In some cases, though, the NWM was very accurate. For instance, on the East Fork of the San Jacinto River near New Caney: 3492 $m^3 s^{-1}$ was forecasted while 3398 $m^3 s^{-1}$ was observed. It should be noted that this a single forecast from a single point in time, while the medium-range NWM variant is run four times a day for 80 time horizons. The testing presented here, therefore, will not be representative of NWM performance during Harvey and should not be viewed as an authoritative assessment of model skill. It may be ill fortune that the particular Fathom flood extent forecast selected for this study was driven by anomalously poor discharge forecasts, but their benchmarking against observed flows provides useful context for analysing the skill of the 2D flood forecast. Furthermore, the NWM was in its infancy around the time of Harvey (v1.1). Improvements, both to date and in the future (v2), of this hydrological model after rigorous validation exercises will result in much closer replication of observed streamflows than those presented here. Official NOAA RFC forecasts are of much greater accuracy (Adams, 2016), but do not have total coverage of US rivers. In Fig. 5.6, the discharge errors at USGS observation sites are plotted and polygons representing the fourteen sites are coloured by their error bias (from Table 5.2). Many of the USGS gauging stations do not relate to sites where USGS benchmark flood extents were generated, perhaps explaining why the CSIs in Table 5.2 are not as low as the discharge errors in Figure 5.5 might suggest. Stations are heavily concentrated in Houston, of which only some are relevant to just one of the fourteen study sites (San Jacinto River (Fig. 5.6g). Equally, many of the sites do not have a representative USGS gauging station: but for those that do, biases in the forecast discharge are often replicated in the inundation extent (where the site polygon and gauge point are of a similar colour in Fig. 5.6). Forecast flood inundation in the east of the domain was driven by NWM discharges that are generally biased towards underprediction, which is reflected in the bias of the flood extent (Big Cow Creek (Fig. 5.6c), Upper (Fig. 5.6k) and Lower Neches

River (Fig. 5.6m), Cow Bayou (Fig. 5.6f), Pine Island Bayou (Fig. 5.6n)). Most of the remaining sites have an overpredictive flood extent bias, where nearby gauges indicate that the NWM forecast discharge was overpredictive also (Upper (Fig. 5.6b) and Lower Brazos River (Fig. 5.6e) and San Jacinto River (Fig. 5.6g)).

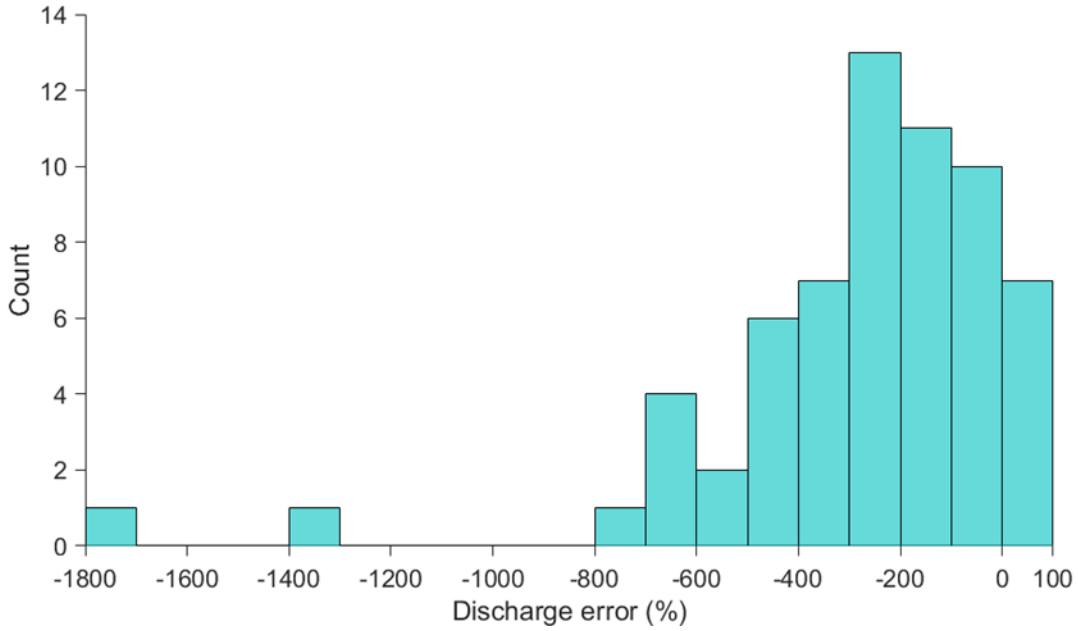


**Figure 5.5** Errors in National Water Model forecast discharge (Q) as a percentage of observed peak flows from USGS stream gauges. A positive error indicates NWM underprediction (observed Q > modelled Q), while a negative error indicates overprediction (observed Q < modelled Q).

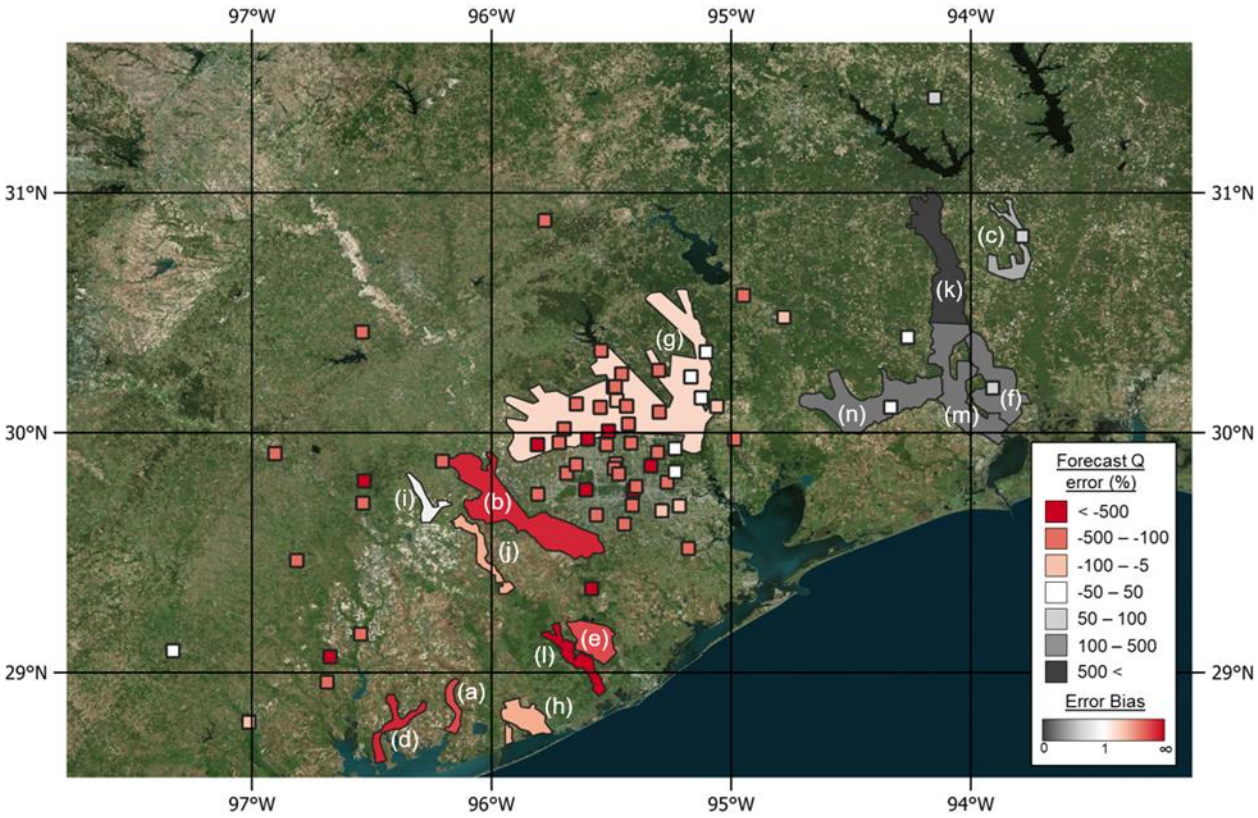


**Figure 5.6** Errors in National Water Model forecast discharge (Q) as a percentage of observed peak flows from USGS stream gauges (shown as points). Error Bias (Eq. 5.4) is displayed for each of the 14 sites (shown as polygons): **(a)** Tres Palacios River, **(b)** Upper Brazos River, **(c)** Big Cow Creek, **(d)** Matagorda Bay, **(e)** Lower Brazos River, **(f)** Cow Bayou, **(g)** San Jacinto River, **(h)** East Matagorda Bay, **(i)** Upper San Bernard River, **(j)** Middle San Bernard River, **(k)** Upper Neches River, **(l)** Lower San Bernard River, **(m)** Lower Neches River and **(n)** Pine Island Bayou. Grey colours indicate underprediction, red colours indicate overprediction, white represents an unbiased estimate.

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

observations also. Yet, these results evidence that despite these limitations the forecast model has fair skill in replicating benchmark flood extents for Harvey.

### 5.4.2 High water mark comparison

The Fathom models were further tested against the raw HWMs that the USGS used to construct the benchmark flood extents in the preceding section. These observational data represent the maximum water surface elevation surveyed by USGS field teams as a result of Hurricane Harvey at a given point. For the hindcast model, the mean absolute WSE error between the selected observation points and the modelled value at the same location is 1.03 m. The influence of a small number of large outlying error values means that the RMSE is significantly higher at 1.71 m. Fig. 5.7a outlines the distribution of these WSE differences, exhibiting a Gaussian distribution about a central tendency close to 0 (mean error: 0.15 m). 50% of the errors lie between –0.43 m (Q25) and 0.90 m (Q75), while 90% of the errors are between –2.32 m (Q5) and 2.12 m (Q95). Fig. 5.7b exhibits the expected shallowing and widening of the error distribution when analysing the Fathom forecast model. MAE rises to 1.22 m (RMSE = 1.88 m). However, this does not represent a drastic reduction in skill from hindcast to forecast.

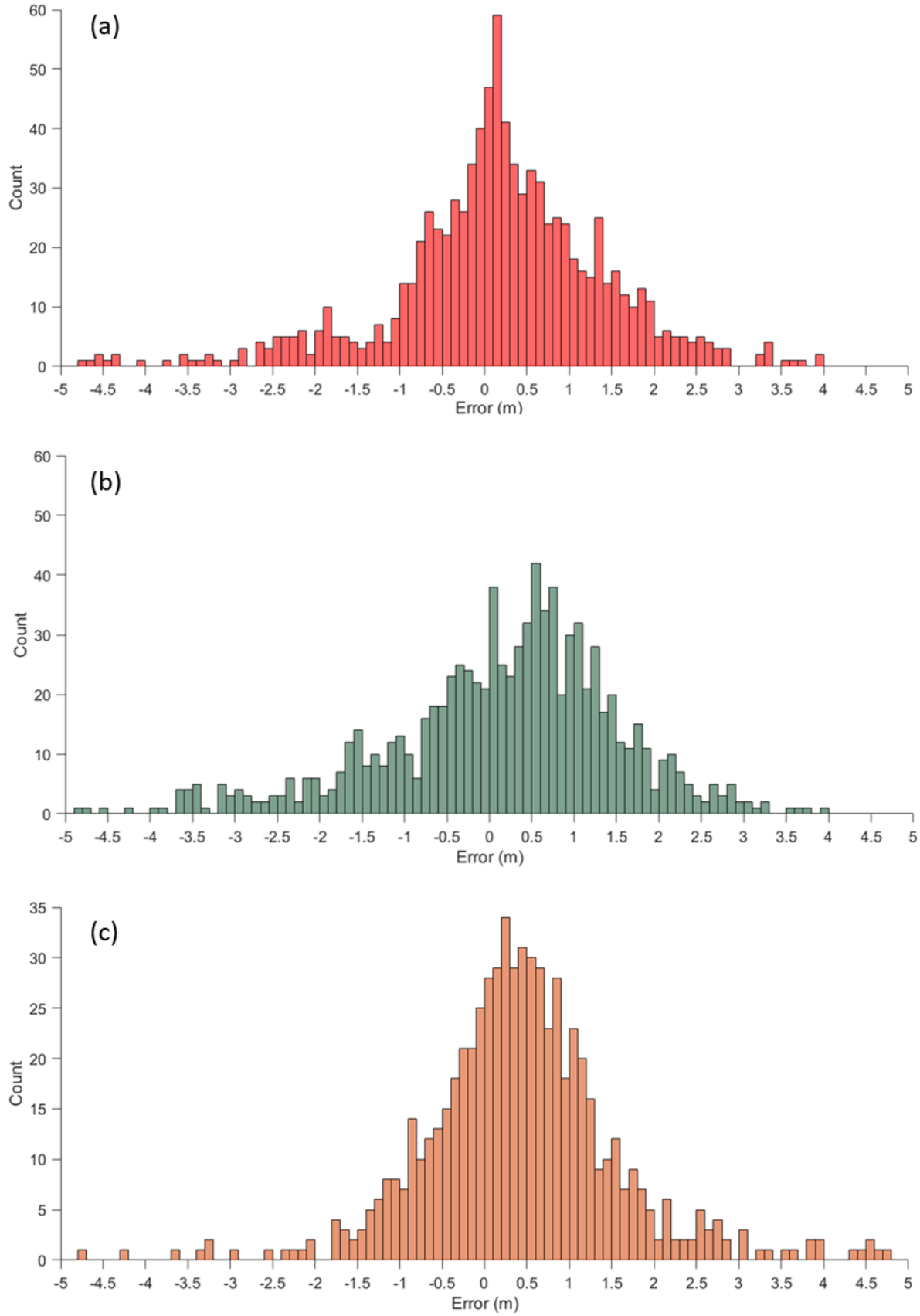


**Figure 5.7** Histograms of the differences between 1134 observed and simulated water surface elevations in the Fathom **(a)** hindcast and **(b)** forecast models. **(c)** shows the differences

between 659 observed ground elevations and those in the DEM employed by the models. Note that 12, 13 and 9 data points fall outside of the displayed error range in (a), (b) and (c) respectively.

Point comparisons of surveyed vs. simulated WSEs are a particularly stringent test for a model of this scale. Indeed, such examinations are rarely carried out for analogous models, particularly uncalibrated, large-scale, high-resolution, ones which include urban areas and smaller streams, meaning context-setting for these results is quite difficult. The Wilson et al. (2007) model of the Amazon obtained an RMSE of 2.37 m (0.99 m at high water) when comparing simulated water levels to those derived from satellite altimetry data. Schumann et al. (2013) calibrated their 1 km resolution forecasting model of the Lower Zambezi River to within 0.27 m of ICESat-derived water levels. This, though, is the product of using a smooth, coarse-resolution DEM of a very flat and wide floodplain, meaning WSEs change very little as the flood extent grows. Obtaining low errors under such circumstances is much less challenging than in the test case presented here, especially when the Zambezi model has been calibrated to observations. Neal et al. (2009a), in a high-resolution, reach-scale hydraulic model of an urban area built with local data, obtained a maximum RMSE of 0.28 m when calibrated to HWMs. As another example, the standard deviation of errors in the calibrated Mignot et al. (2006) model of an urban area in France was 0.53 m. It should be noted that hydraulic models that have been calibrated may not necessarily be more inherently skillful, meaning these local-scale model errors may not be analogous to the uncalibrated large-scale model presented here. As an example of an operational NOAA flood forecast, Adams et al. (2018) compare modelled hindcasts of their large-scale framework in the Midwest to USGS gauged stages and found errors of <0.5 m. While this is more accurate than the model presented here, it required thousands of manual bathymetric surveys to build, meaning the Adams et al. (2018) model only simulates flooding on large rivers of known geometry. In the context of Hurricane Harvey, where flooding from small streams and surface runoff accounted for a considerable portion of the hazard, this NOAA exemplar model may not have attained similar errors when benchmarked against high water marks on the floodplain and particularly where flooding was predominantly pluvial. Zheng (2018) tested the NWM-HAND model of maximum inundation outlined in the preceding section (NOAA NWC et al., 2018) against this same set of USGS HWMs and found the standard deviation of model error to be 4 m. Previous studies have also quantified the errors in surveyed HWMs that such

models are calibrated to or validated against. Errors are generally 0.3 – 0.5 m (Schumann et al., 2007; Neal et al., 2009a; Horritt et al., 2010; Fewtrell et al., 2011b), owing to deposition during a hiatus in flood recession (i.e. not high water), wall seepage or debris line width. These are broadly consistent with the maximum errors reported for the HWMs used in this study, but average errors are generally < 0.1 m (Watson et al., 2018).

It is clear that the automatically generated forecast and hindcast models in this paper do not obtain comparable WSE errors to bespoke hydraulic models in the wider literature or NOAA NWS RFC forecasts, though without hydraulic model calibration, local data and manual operation they were never likely to. It is also important to point out the difference in purpose between this and other models. Rather than seeking to perfectly replicate a given flood event with few limitations regarding processing and computation time, this is a forecast product designed to quickly indicate areas susceptible to flooding from an incoming storm to enable an immediate response. Rapid exposure assessments and resource allocation by first responders do not demand a highly accurate map of water depths, but rather a broad indication of the spatial patterns in flooding that may occur (Pitt, 2008; Price et al. 2012). Despite some sacrifice to point precision (e.g. with respect to the accuracy of the 155 discrete forecast inundation points in NOAA AHPS), the framework's total coverage, both in geography and flood driver, serves to fill the information gaps left by sporadic yet accurate local forecasts. Besides, efforts to reduce WSE errors in the hydraulic model are of little value when considering the commensurate or larger errors in the meteorological and hydrological models that precede it in the forecast cascade. Cangialosi (2018) noted that intensity and track errors in the NOAA NHC 3-day forecasts during the 2017 hurricane season were ~7 $ms^{-1}$ and ~150 km respectively. The maximum total rainfall in the NOAA WPC 3-day forecast was ~1000 mm, yet a maximum of over 1500 mm was observed (Blake and Zelinsky, 2018). Hydrological models, used to translate rainfall to streamflow, may also be significantly uncertain, even if the meteorological inputs were error-free (Blöschl et al., 2013). Generally accepted benchmarks for satisfactory performance in hydrological models are: (i) within 25% error in simulated discharge and (ii) a Nash-Sutcliffe efficiency of greater than 0.5, meaning model mean square discharge error represents less than half of the observed variance (Moriasi et al., 2007; Refsgaard and Knudsen, 1996; Ritter and Muñoz-Carpena, 2013). With source and pathway uncertainties such as these propagating into the receptor component of the forecast cascade (Fig. 5.1), the hydraulic model deviating from

observations by an average of ~1 m is unsurprising yet may still be useful for event early warning and first responder preparedness in the absence of local modelling strategies.

659 of the HWMs collected by the USGS also reported the height of the water above the ground surface, meaning a surveyed ground elevation can be calculated by subtracting this measurement of water depth from the WSE. Comparing these elevation values to those contained at the corresponding location in the DEM provides further context for the errors shown in Fig. 5.7a and 5.7b. Again, it should be noted that the observed data contends with error in both the surveyed water surface elevation and its height above the ground: such errors could plausibly be half a metre or more (Watson et al., 2018). Fig. 5.7c shows the spread of these elevation errors: the DEM RMSE is 3.77 m, though this is heavily biased by a handful of implausibly high errors (maximum negative error: –58.9 m; maximum positive error: 20.9 m) which are just as likely to be due to human error in the manual ground survey as erroneous values in the DEM. Reducing the influence of these errors by considering MAE, the quantity stands at 1.19 m. With forecast and hindcast WSE MAEs of 1.21 and 1.03 m respectively, the water levels simulated by the model are similar to or outperform the accuracy of the DEM, implying two key findings. Firstly, much of the WSE error is attributable to the quality of the USGS NED (the source of the model DEM) in this domain. It stands to reason, then, that in domains with a greater proportion of LiDAR in the NED, ground, and thus water surface, elevation error would drop. Once again, the assertion made by Horritt and Bates (2002) still holds: topography is the major control on flood inundation patterns. Secondly, although it seems counterintuitive that errors in WSE can be smaller than the DEM from which they are derived, this underlines the idea that relative, rather than absolute, DEM accuracy is much more important in hydraulic modelling. How the elevation varies between pixels in a locality controls the movement of water over a floodplain, not a pixel's elevation relative to a vertical datum (though this does not apply for simulations of coastal flooding). Typically, relative DEM errors are much lower than absolute ones: Gesch et al. (2014) quantified a ~20% reduction in the standard deviation of relative compared to absolute errors in the NED at ⅓ arc second resolution.

Fig. 5.8 displays the spatial distribution of the WSE errors for the hindcast. Areas of high performance (in white; errors between –0.5 and 0.5 m) are evident across the domain, particularly in Matagorda and Brazoria Counties south of Houston (approx. 29.2°N 95.5°W; to the south of the top-left panel in Fig. 5.8), Jefferson and Hardin Counties to Houston's east

(approx. 30.1°N 94.3°W), and in parts of Houston's Harris County itself (approx. 29.9°N 95.4°W). There is clear clustering to areas of poor performance. Coastal areas in the southwest of the domain around Corpus Christi have a positive bias, indicating the model underpredicted observed HWMs in these areas. This is also the case around Village Creek, north of Beaumont (approx. 30.4°N 94.2°W). Clusters of very high negative errors are present along Houston's Buffalo Bayou (approx. 29.8°N 95.5°W; to the north of the top-left panel in Fig. 5.8) and the Calcasieu River in Lake Charles (30.2°N 93.2°W; to the east of the top-right panel in Fig. 8), signalling model overprediction of WSEs. Incorporating Fig. 5.9 into this interpretation, which exhibits the spatial distribution of the DEM errors in Fig. 5.7c, it may be expected that areas of high WSE error are also areas of high ground elevation error. Observing Figs. 5.8 and 5.9 in tandem suggests that this is not the case. Areas of low WSE error to the south of Houston are in fact dominated by relatively high ground elevation error (as colours move away from white towards red and black), while ground elevation errors in areas identified as having high WSE error (Houston and Lake Charles) are low or, if anything, of the opposite sign to the WSE errors. This reinforces the suggestion that relative, and not absolute, DEM accuracy is pre-eminent and draws on the well-documented challenges of hydraulic modelling in urban areas (Yu and Lane, 2006a; Mason et al., 2007; Hunter et al., 2008), especially at large scales (Wing et al., 2017). High WSE errors are generally confined to urban areas in this study, where horizontal, rather than vertical, elevation accuracy (i.e. grid resolution) controls flood model performance to a much greater extent than elsewhere. This is due to the increased prevalence of hydraulically important anthropogenic features (small-scale flow paths, building walls, levees, roads and much else) which are unresolved by the elevation data since they are smaller than the width of a grid cell. In the absence of the computational capacity available to run large-scale models at very high-resolution (e.g. Sampson et al., 2012), which is some distance from feasibility given doubling the granularity of grid resolution increases computation time by an order of magnitude (Savage et al., 2016), the solution to this problem may lie with nested or variable-resolution grids where high-risk cities can be modelled at finer resolution (e.g. Sanders et al., 2010; Kim et al., 2014; Sanders and Schubert, 2019) or innovative sub-grid scale solutions (e.g. Yu and Lane, 2006b; Sanders et al., 2008; Schubert and Sanders, 2012; Guinot et al., 2017): the bottleneck of the latter approach being obtaining parameterization data at large scale.

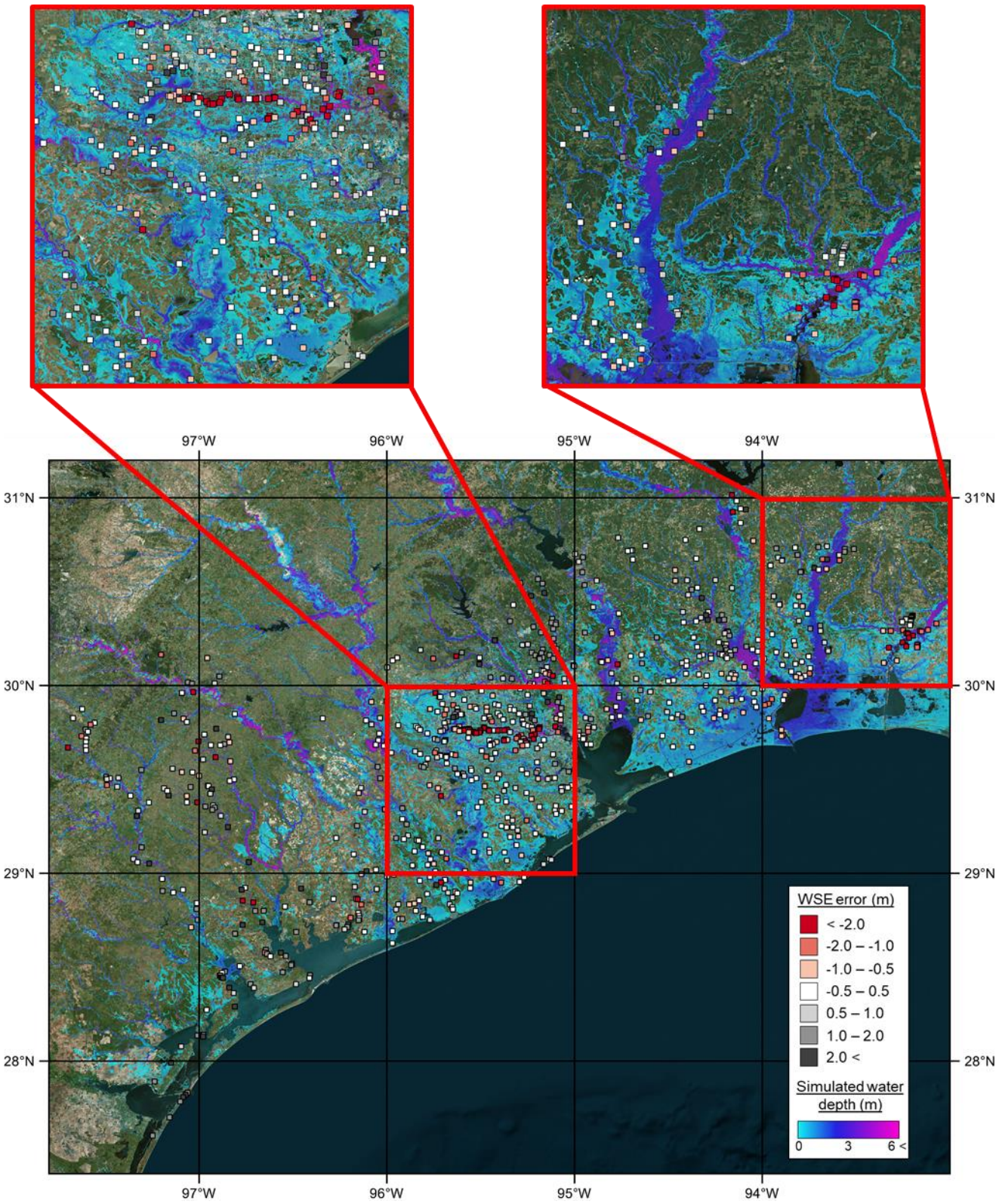


**Figure 5.8** The model domain containing a grid of hindcast water depths and 1134 points of WSE errors with respect to the surveyed HWMs. The top-left panel contains Houston to the north; the top-right panel contains Lake Charles to the southeast. White and lighter colored points indicate low errors; redder colours indicate model overprediction; greyer colours indicate model underprediction.

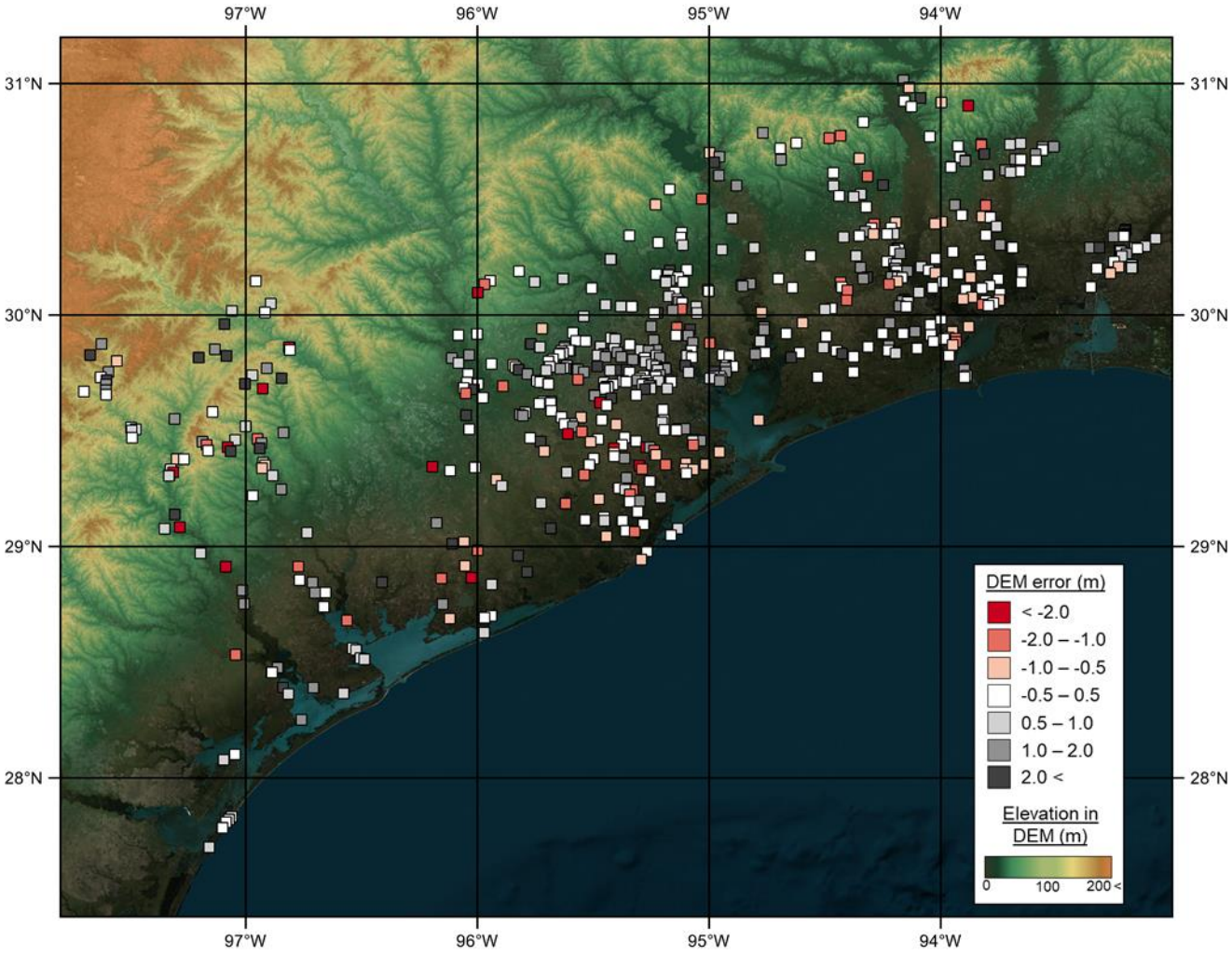


**Figure 5.9** Elevation differences in the DEM (derived from the USGS National Elevation Dataset) compared to 659 USGS post-event ground surveys. White and lighter coloured points indicate similar elevations in the DEM and the survey; redder colours indicate higher elevations in the DEM than those surveyed; greyer colours indicate lower elevations in the DEM than those surveyed.

## 5.5 Conclusions

This paper presents a flood forecasting product for hurricanes in the US that has the potential to be used operationally in the absence of accurate local forecasts, comprehensively testing it against ground truth data collected by the USGS. The framework takes available hydrologic NOAA forecasts as inputs to an existing continental-scale model structure (Wing et al., 2017), accounting for all primary flood drivers, in order to rapidly simulate event water depth grids for a given domain anywhere in the US. Comparing model hindcasts of Hurricane Harvey to benchmark data of its maximum flood extent indicates that the model has skill in picking up the spatial patterns of inundation (mean CSI value of 0.66). When benchmarked

against surveyed water surface elevations, the model misestimates this quantity by roughly 1 m on average. This is amidst a backdrop of similar errors in the DEM and up to ~0.5 m of error in the measured HWMs. The model in forecast mode experiences only a moderate drop in performance relative to the hindcast. We conclude this to be commensurate to or beneath likely uncertainties in the preceding components of the forecast cascade, as well as similar to errors in the underlying DEM. With expanded lidar coverage within the USGS NED, the model presented here may approach the accuracy of operational NOAA inundation forecasts (Adams et al., 2018; Mashriqui et al., 2014) but with total coverage and for a diverse range of flood drivers.

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

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

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

### 5.6 Acknowledgements

Streamflow forecasts from the NOAA National Water Model are available from https://nomads.ncep.noaa.gov/pub/data/nccf/com/nwm/. NOAA Weather Prediction Centre rainfall forecasts can be accessed at https://www.wpc.ncep.noaa.gov/qpf/day1-3.shtml. NOAA National Hurricane Centre coastal surge height forecasts are accessible from https://slosh.nws.noaa.gov/psurge2.0/. USGS river gauge observations can be downloaded from https://waterdata.usgs.gov/nwis/rt. Rainfall observations from the NOAA National Weather Service are available at https://water.weather.gov/precip/download.php. Observed coastal surge heights from NOAA Tides and Currents can be accessed at https://tidesandcurrents.noaa.gov/gmap3/. The hydrodynamic model LISFLOOD-FP can be downloaded from http://www.bristol.ac.uk/geography/research/hydrology/models/lisflood/downloads/. Output from the NOAA National Water Centre Hurricane Harvey HAND models is available at https://www.hydroshare.org/resource/fe85a680d0144e79b39e8c483dc1e5aa/. USGS observations and benchmark data can be downloaded from https://www.sciencebase.gov/catalog/item/5a85f30fe4b00f54eb36d3a9. The Fathom Harvey depth grids presented in this study are proprietary in nature, but can be made available for academic research purposes (i.e. not commercial, policy or regulatory applications) by contacting Christopher Sampson at Fathom (c.sampson@fathom.global). Paul Bates was supported by a Research Fellowship from the Leverhulme Trust and a Royal Society Wolfson Research Merit Award. Jeff Neal was supported by UK Natural Environment Research Council grant NE/S006079/1.

# 6 A NEW AUTOMATED METHOD FOR IMPROVED FLOOD DEFENCE REPRESENTATION IN LARGE-SCALE HYDRAULIC MODELS

*This paper-chapter was published in Water Resources Research in 2019:*

**Wing, O. E. J.**, Bates, P. D., Neal, J. C., Sampson, C. C., Smith, A. M., Quinn, N., Shustikova, I., Domeneghetti, A., Gilles, D. W., Goska, R., and Krajewski, W. F. (2019), A new automated method for improved flood defense representation in large-scale hydraulic models, *Water Resources Research*, 55(12), 11007–110034, doi:10.1029/2019WR025957.

*Author contributions:*
O. Wing led the study, conceived the methodology, generated the levee extraction algorithm, produced new continental elevation data, processed the benchmark data, performed all analyses, and wrote the manuscript. O. Wing and C. Sampson developed the new hydraulic simulations in Iowa. O. Wing and I. Shustikova developed the hydraulic simulations of the Po River, along with J. Neal and A. Domeneghetti. D. Gilles, R. Goska, and W. Krajewski provided the Iowa Flood Centre benchmark data. All authors aided in the conceptualisation of the research and commented on manuscript drafts.

## 6.1 Background

This chapter addresses thesis objective 3:

> *Tackle the issue of poor flood defence representation epidemic in current large-scale approaches.*

While the results of chapters 1 and 3 indicate the large-scale US flood inundation model obtains similar flood extents to local-scale models and observations, it is evident that a lack of structural flood defence information can hamper its utility for heavily engineered floodplains. This chapter demonstrates that previous approaches to flood defence representation are largely unsubstantiated through their assumptions of defences being afforded to every river and socio-economic variables being predictive of defence standard. The US model as presented in chapters 1–3 uses publicly-available levee data from the US Army Corps of Engineers, yet this only contains an estimated 30% of the nation's levees. Thus, a novel solution is presented here for application in the US by sampling the geomorphometric characteristics of known levees and searching for features exhibiting these

characteristics in high resolution elevation data. The elevation data of the US model, the USGS National Elevation Dataset, is available at ~10 m resolution (and ~3 m resolution for most populated areas) but simulations take place at ~30 m resolution for computational efficiency. The algorithm identifies levee-like features in the ~3 and ~10 m resolution data and ensures their elevations are preserved during coarsening to ~30 m resolution. When compared to surveyed levee crests, the algorithmically-built DEM had a median of error of 0.05 m compared to the 1.16 m error of the original DEM. When employed in a hydraulic modelling framework, the method is shown to have a positive effect on model skill when benchmarked against high-accuracy urban flood models in Iowa (US), though improvements can be obscured by uncertainties in other model features (namely, extreme flow quantification and channel hydrography and geometry). For a test on the Po River (Italy), the model replicated observations of a large historic flood event as well as an engineering-grade model, while the analogue to current large-scale approaches performed poorly.

## 6.2 Introduction

The last decade has seen a revolution in the field of flood inundation modelling. 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 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 modelling 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–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 NGOs 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 purpose in flood modelling at large scales (e.g. Archer et al., 2018; Yamazaki et al., 2017). The development of simplified hydraulic codes, which ensure physical process representation is commensurate with data availability and computational burden (Hunter et al., 2007), have 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 and Schubert, 2019). The delineation of channel networks from global DEMs (e.g. Allen and 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., 2012b) 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 defence structures within them.

Information on the location of flood defence structures on the floodplain (specifically levees, embankments, dykes, flood walls: these terms are used interchangeably hereafter) at large spatial scales is very scarce; metadata regarding their defence 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 is higher resolution (e.g. airborne lidar surveys which are usually available on

1–5 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 defences is virtually ignored by large-scale flood hazard modelers. If accounted for at all, many modelers treat flood defences as a post-processing step 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 dataset which provides design, policy and modelled flood defence standards globally at the sub-country level to inform the defence threshold above which floods cause damage. It is, however, not justified to assume defence 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 and Pinter, 2012) and downstream (Castellarin et al., 2011a; 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 US 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 defence 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 interacts with the terrain data. Datum conflicts, geolocation errors and the effect of residual levee artefacts in the baseline topographic data can result in an improper representation of defence information, even in the relatively few locations where such data are available. If defence 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 defence standard to the bank height of the river channel (Neal et al., 2012b) 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 "defence 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 defence representation leading to inundation in areas that are protected in reality.

Amidst incomplete defence inventories, some modellers make assumptions about where levees are likely to be situated based on socio-economic 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 US as an example, we demonstrate that such assumptions are generally invalid. The USACE NLD contains a wealth of levee data for the contiguous US, including ~200,000 $km^2$ 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 US Census Bureau) and government spending (USACE spending under the American Recovery and Reinvestment Act of 2009) offers insights into whether a crude predictive defence model based on nationwide socio-economic characteristics is valid. The left-hand side of Fig. 6.1 illustrates the correlation between the defence standard of the levee protecting an area of land and its characteristics in terms of urbanity, wealth and spending respectively. It is clear from Fig 6.1a 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 neighbourhoods are not offered greater protection than defended poorer ones (Fig. 6.1b). Indeed, very high return period levees (i.e. those that defend against the 1 in 1000 year flood) appear to defend poorer areas. Further, higher USACE spending since 2009 does not appear to be correlated with defence standard: if anything, the opposite case is true (Fig. 6.1c). The histograms on the right-hand side of Fig. 6.1 show the distribution of these characteristics within defended areas and in the remaining US land surface outside of them. The distribution of urbanity (Fig.

6.1d), wealth (Fig. 6.1e) and spending (Fig. 6.1f) are 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 defences.

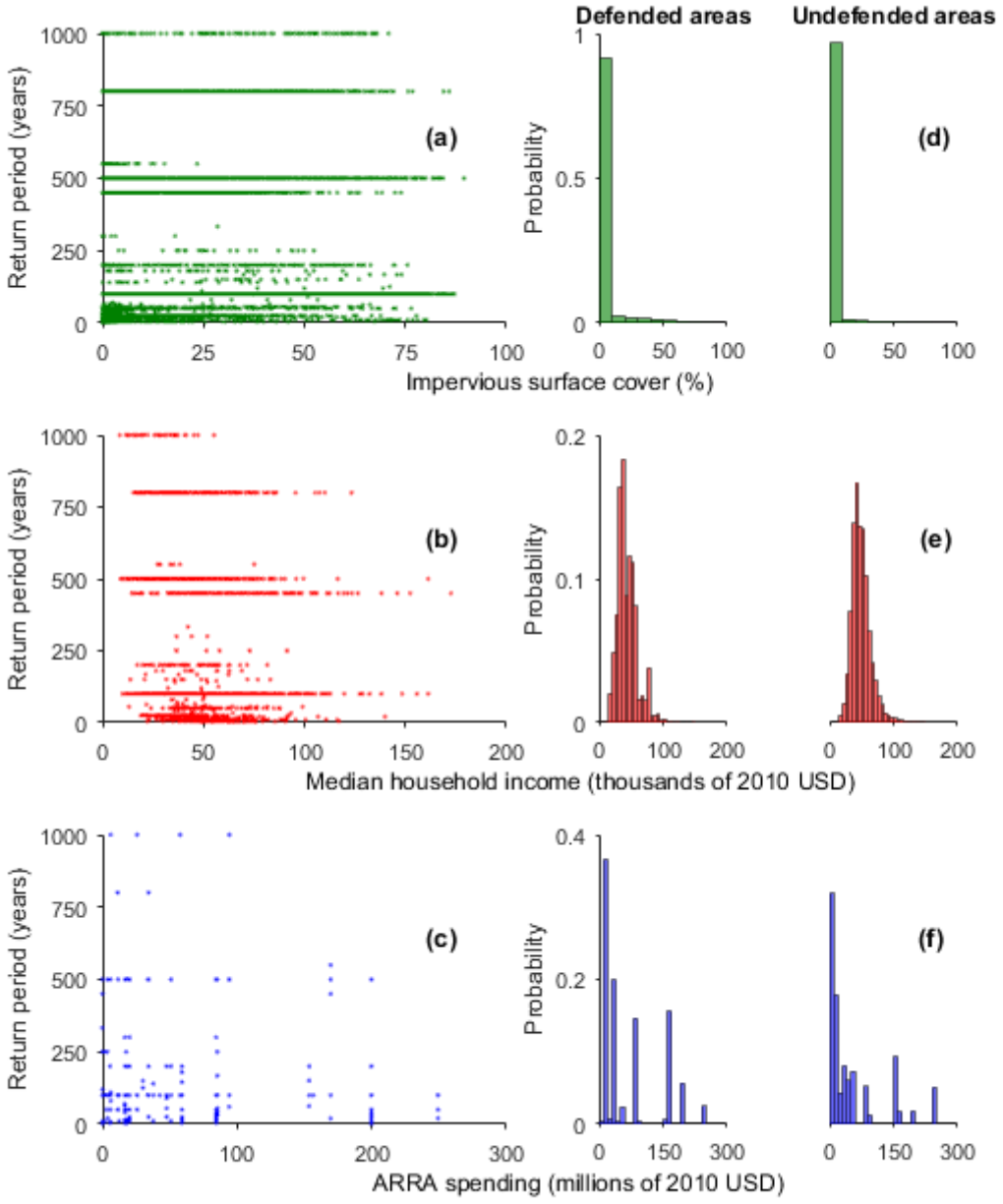


**Figure 6.1** The relationship between USACE NLD defence standard and **(a)** urbanity, **(b)** wealth, and **(c)** spending, as well as the distribution of **(d)** urbanity, **(e)** wealth and **(f)** spending within and outside USACE NLD defended areas.

In this paper, we propose a novel advance in defence 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 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 flood defence 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 defences have been extracted from DEMs in a number of studies (Bailly et al., 2008; Casas et al., 2012; Cazorzi et al., 2013; Choung, 2014; Krüger, 2010; Krüger and 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, 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 and Meinel, 2008; Sofia et al., 2014). Given these studies: (i) employed lidar-derived DEMs with resolutions of the order 100 m, which are currently unavailable at large spatial scales; (ii) were of small, isolated test cases; (iii) offered little indication regarding the computational feasibility of applying such methods at larger scales; and (iv) were not extended to analyse fitness-for-purpose in a hydraulic modelling context, such methods leave crucial research questions unanswered. Other novel approaches, including those with a post-processing 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 modelling, a view of

crest elevation or defence standard is required as well. As such, the method presented here adheres to a DEM-based approach since: (i) elevation data is available at large spatial scales, albeit at a coarser resolution than in previous levee detection studies; and (ii) 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 modelling 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 defence 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.

### 6.3 Data and methods

For this study, the test area is the entire contiguous United States (CONUS), owing to the availability of total-coverage elevation data from the US Geological Survey (USGS) National Elevation Dataset (NED) at ⅓ arc sec (~10 m) and partial coverage at ⅑ arc sec (~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 sec (~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 sec model structure. Fathom-US hydraulics are based on a version of the LISFLOOD-FP numerical scheme: a 2D simplification of the shallow water equations which approximates local inertia (Bates et al., 2010; de Almeida and Bates, 2013). AEP flows, based on the regional flood frequency analysis of Smith et al. (2015) using USGS river gauges, are routed through 1D subgrid channels (Neal et al., 2012b), which are derived from HydroSHEDS hydrography data (Lehner and 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 NOAA 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 ⅓ and ⅑ arc sec 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 National Levee Database. This is another reason why the CONUS is an ideal testbed, since it contains wide-area levee information in the form of the NLD. While not containing every levee in the US, the database contains 48,000 km of flood defence 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 = 1000, in this case) of the geomorphometric characteristics evident in the NLD drives the extraction algorithm 1000 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 1000. 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 1000), its elevation value will become the corresponding coarser 1 arc sec DEM pixel value. Where multiple higher-resolution pixels with ε above the threshold fall within a 1 arc sec 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 sec 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 1000 NLD-derived thresholds are detailed as follows (see Figs. 6.2 and 6.3):

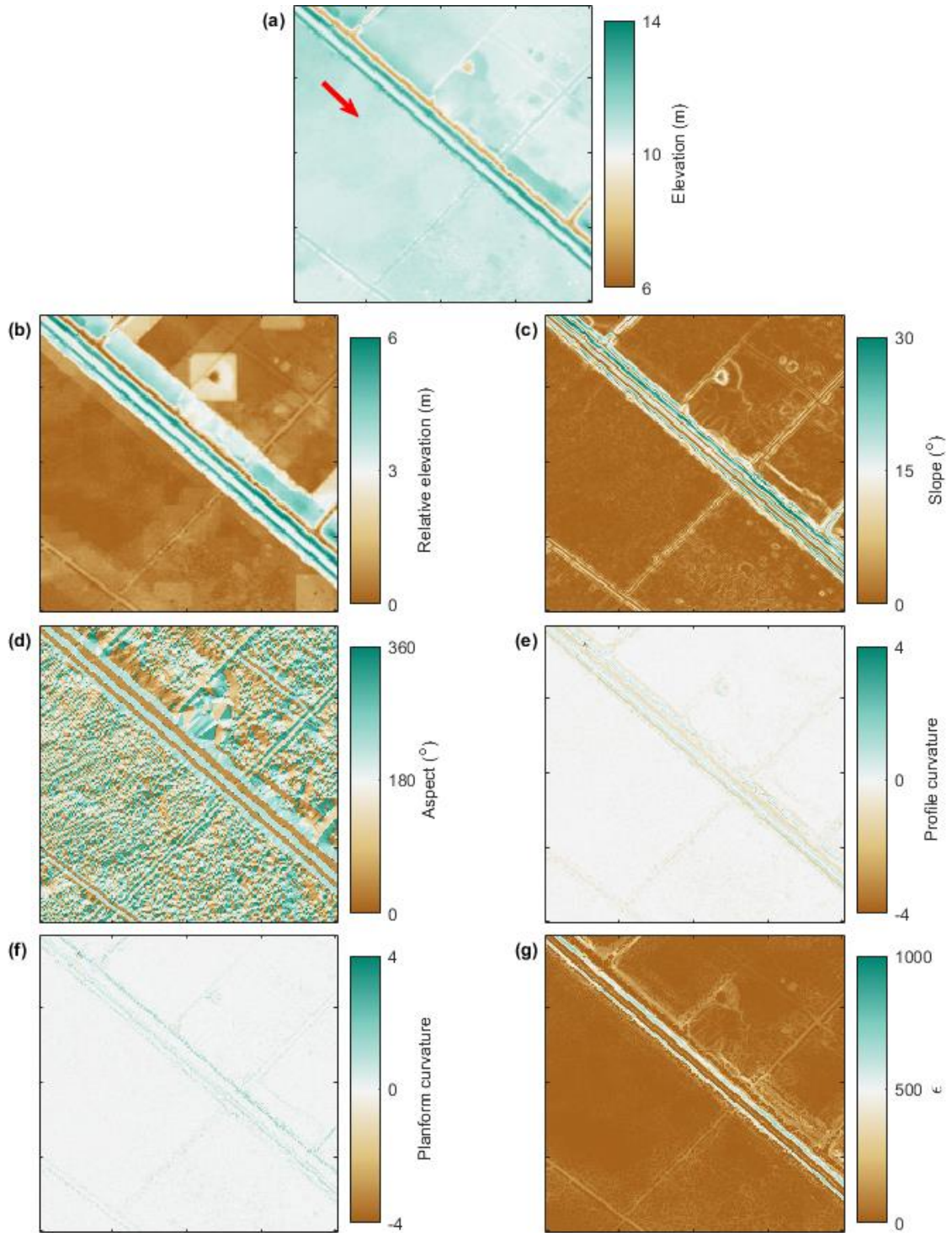


**Figure 6.2** Examples of a levee in a DEM and its processing in the extraction algorithm for a 0.6 x 0.6 km area of Brazoria County, Texas (ticks are spaced 150 m apart). **(a)** an area of raw ⅓ arc sec USGS NED DEM, with two levees bounding a river channel (in dark green) with flow direction indicated by the red arrow; **(b)** relative elevation, showing the levees' clear height differential above their neighbourhood 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; **(g)** output of the extraction algorithm when driven with geomorphometric parameter thresholds on the right-hand side of Fig. 6.3, where the levees have much higher ε values than other terrain features.

(i) 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 neighbourhood minimum. Neighbourhood in this context is defined as a 100 x 100 m square with point ($x$,$y$) at its centre. 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 Eq. 6.1).

$$z_r = z_{xy} - z_{min} \tag{6.1}$$

Our expectation is that levee crests are elevated with respect to their neighbourhood minimum, resulting in a positive $z_r$. A visual example is shown in Fig. 6.2b. Figs. 6.3a–b show the 1000 thresholds at ⅓ and ⅑ arc sec respectively, derived from known levee locations. Although both provide quantities in line with the above expectation, their distributions are different. At ⅓ arc sec, roughly 3x as many $z_r$ thresholds are < 1 m than at ⅑ arc sec. This seems to suggest that, even at ~10 m resolution, the representation of levee crests in the DEM has been diminished.

(ii) Slope

Slope ($S$) is a first-order differential of elevation: defined here as the maximum rate of change of elevation between neighbouring pixels in $x$ and $y$ directions in degrees (Eq. 6.2). 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 Fig. 6.2c).

$$S = tan^{-1}\left(\sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2}\right) \tag{6.2}$$

From Figs. 6.3c–d, 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.

(iii) Aspect difference

Aspect (*A*) is another first-order differential of elevation, indicating the direction of maximum slope (*S*). This is defined by Eq. 6.3 in degrees clockwise from north. In the case of levees, we would expect the opposing slopes to be facing polar opposite directions (see Fig. 6.2d).

$$A = tan^{-1}\left(\frac{\left(\frac{dz}{dy}\right)}{\left(\frac{dz}{dx}\right)}\right) \tag{6.3}$$

The thresholds are defined in terms of the differences between the aspects of opposing slopes. As in Eq. 6.4, 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:

$$A_{diff} = \left|A_{(x+m,y+n)} - A_{(x-m,y-n)}\right| \tag{6.4}$$

The geomorphometric parameter thresholds in Fig. 6.3e–f 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 to 0°. The slight rise in cases where the tolerance is close to 180° (indicating $A_{diff} \approx 0$) are probably due to NLD geolocation errors where the crest is placed on the slope. In such circumstances, the steep slopes either side of this misidentified crest would be facing the same direction.

(iv) Profile curvature

Curvature is a second-order differential of elevation. Profile curvature ($prof_c$) is the rate of change of slope in the maximum downslope (*S*) direction, defined in Eq. 6.5:

$$prof_c = \frac{dS}{dxy} \tag{6.5}$$

Levees are characterized by profile convexity (Fig. 6.2e), in this case being represented as negative profile curvature. In almost all cases, the algorithm considers levee crests with a convex profile (Fig. 6.3g–h). With flatter slopes in the coarser data, profile convexity is also closer to 0.

(v) Planform curvature

Finally, planform curvature ($plan_c$) is the second derivative of elevation orthogonal to the direction of maximum slope ($S'$), as defined in Eq. 6.6. This direction follows the levee crest, where very little change in slope is expected (Fig. 6.2f).

$$plan_c = \frac{dS'}{dxy} \tag{6.6}$$

From Fig. 6.3i–j, it is evident that levees in the NLD have no or very little planform curvature within a narrow tolerance.

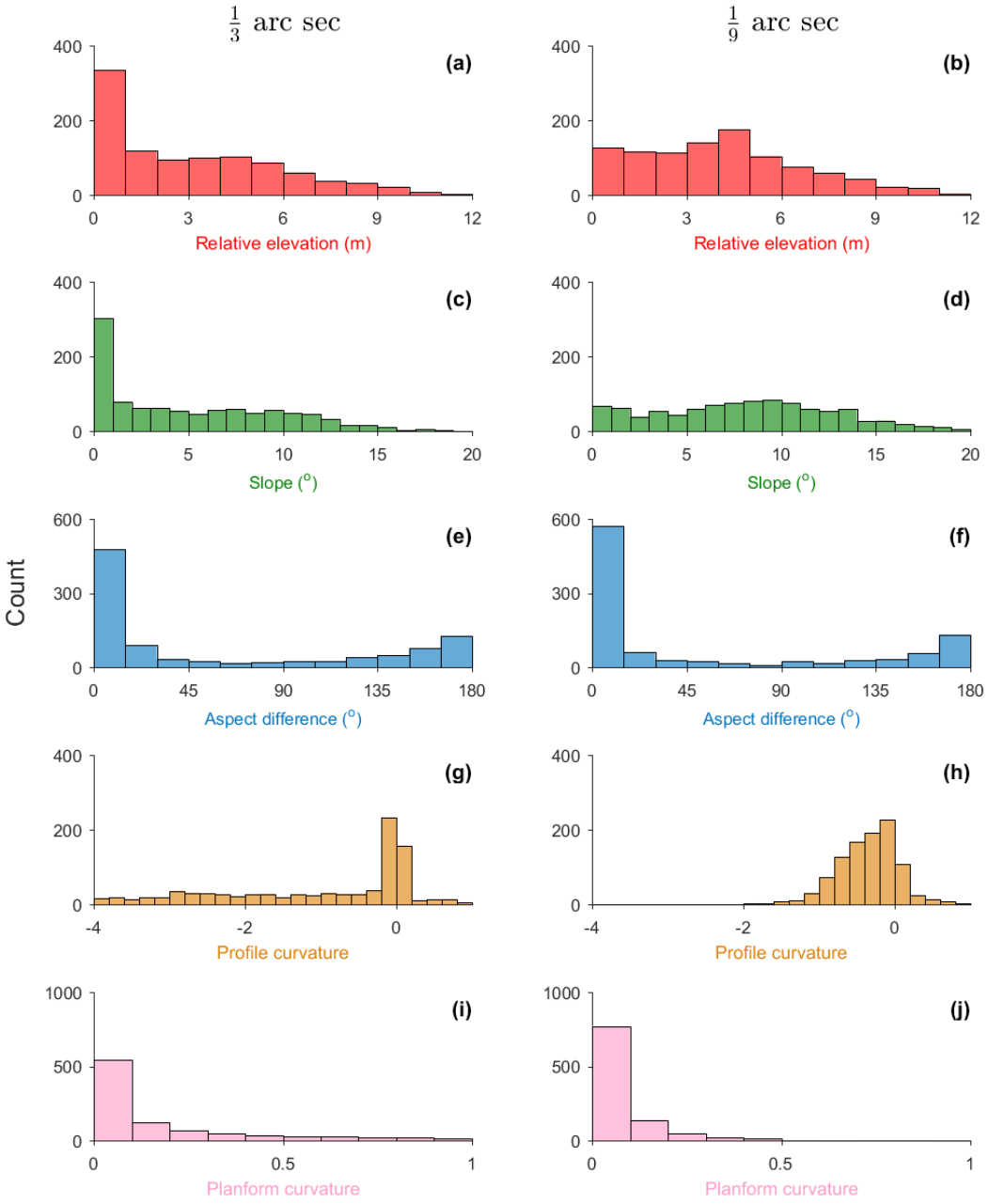


**Figure 6.3** Histograms showing the geomorphometric parameter thresholds sampled from the USACE levee data at ⅓ (left-hand side) and ⅑ arc sec (right-hand side): **(a)(b)** relative elevation; **(c)(d)** slope; **(e)(f)** aspect difference; **(g)(h)** profile curvature; and **(i)(j)** planform curvature.

Fig. 6.2g demonstrates an example of how the algorithm functions for a small section of the ⅑ arc sec 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 contiguous US at ⅓ arc sec resolution and at ⅑ arc sec resolution where available to re-generate a 1 arc sec NED-based DEM with these hydraulically-important features preserved during the coarsening process. The 1 arc sec "defended" DEM built with the feature preservation algorithm run with ⅑ arc sec 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 NAVD 88 datum for roughly 7000 km of Californian levees. Concurrently, the "undefended" 1 arc sec 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 re-run with a new 1 arc sec DEM built with algorithmic output of the ⅓ arc sec 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 Centre (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 (Krajewski et al., 2017; Chen et al., 2017; Horna-Muñoz and Constantinescu, 2018). Their maps consist of two modelling strategies: (i) complete-coverage statewide maps built with 1D HEC-RAS models to underpin National Flood Insurance Programme (NFIP) rate setting and (ii) more detailed urban flood maps at select locations, mostly simulated using coupled 1D/2D models. Both IFC modelling 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 1D channel representation or the 2D DEM. The 1D/2D HEC-RAS or MIKE FLOOD models, run at 10–20 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 modelling 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 defence 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):

$$\mathrm{HR} = \frac{M_1 B_1}{M_1 B_1 + M_0 B_1} \tag{6.7}$$

$$\mathrm{FAR} = \frac{M_1 B_0}{M_1 B_0 + M_1 B_1} \tag{6.8}$$

$$\mathrm{CSI} = \frac{M_1 B_1}{M_1 B_1 + M_0 B_1 + M_1 B_0} \tag{6.9}$$

$$\mathrm{EB} = \frac{M_1 B_0}{M_0 B_1 + M_1 B_0} \tag{6.10}$$

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; Eq. 6.7; optimum = 1) is a measure of model underprediction, penalizing only "misses" ($M_0B_1$). The False Alarm Ratio (FAR; Eq. 6.8; optimum = 0) measures model overprediction, where "false alarms" are exclusively penalised ($M_1B_0$). Critical Success Index (CSI; Eq. 6.9; optimum = 1) is the most discriminatory measure, penalising both under- and over-prediction. Finally, the Error Bias (EB; Eq. 6.10; optimum = 0.5) metric is slightly different to that employed in previous studies. It still indicates the overall balance of under- and over-prediction, but ensures the magnitude of the deviation from the optimum in either direction is consistent: EB > 0.5 is overprediction; EB = 0.5 is unbiased; 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. (2007b). Eqs. 6.7–6.10 are applied in the same way, but $M_xB_x$ are 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 US Environmental Protection Agency EnviroAtlas program, where census block counts are dasymetrically downscaled to 30 m pixels based on land-use and slope. While the pixel-count metrics indicate wide-area physical modelling ability where correctly modelling 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 ⅑ arc sec resolution lidar data (provided at 2 m resolution, but resampled to ⅑ arc sec) 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–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 2D (no sub-grid channels). Instead, surveyed channel bathymetry is burnt directly into the DEM, thus representing rivers as supra-grid features. The data associated with this model of the Po has been used in a number of hydraulic studies (e.g. Castellarin et al., 2009, 2011a, 2011b; Di Baldassarre et al., 2010; Domeneghetti et al., 2015; Schumann et al., 2010). For this test case, three different DEMs are employed whilst all other elements of the model structure are held constant:

(i) manual-defended (MD): the raw DEM is resampled to 1 arc sec resolution and manually digitised levee crest centrelines are used to elevate corresponding pixels to the ⅑ arc sec elevation value;

(ii) automated-defended (AD): to emulate large-scale model structures, where local levee data is mostly unavailable, the automated levee extraction algorithm is run over the raw ⅑ arc sec DEM, preserving the elevations of relevant pixels during the resampling to 1 arc sec resolution;

(iii) undefended: no representation of flood defences except those which are inherent in the resampled 1 arc sec DEM; while some residual representation of flood defences 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 (ii) DEM where they also hold in the MD (i) DEM while overtopping in the undefended (iii) DEM. While the manual handling of flood defences considers exclusively formal levee structures, the automatic method extracts all levee-like features regardless of their anthropogenic classification. This means the elevation of more informal features, such as small berms, river banks, roads, undocumented (often minor) dykes 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.

### 6.4 Results and discussion

The algorithm was executed across the entire contiguous US at ⅓ arc sec (~10 m) resolution, at ⅑ arc sec (~3 m) resolution where available in the US, and over the Po floodplain at ⅑ arc sec (~3 m) resolution to ultimately produce new "defended" 1 arc sec (~30 m) DEMs for the US and the Po. At ⅓ arc sec resolution, a single parameter threshold set takes, on average, 0.99 s to run for a 1 x 1° (approx. 100 x 100 km) tile on a single 2.30 GHz Intel Xeon E5-2650 core. For the same size tile at ⅑ arc sec 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 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.

### 6.4.1 Levee crest elevation validation: California, US

The 1 arc sec DEM built with consideration of levee-like features at ⅓ arc sec and an ε threshold of 300 (see section 6.4.2.1 for rationale) is compared to geodetic surveys of levee crest heights in the California Levee Database. The results of this benchmarking are shown (in blue) in Fig. 6.4, where positive (negative) errors indicate DEM overprediction (underprediction) of the CLD crest height. 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 sec DEM built only by resampling the ⅓ arc sec data are shown in orange. This DEM has a clear underpredictive bias, with a median error of –1.16 m.

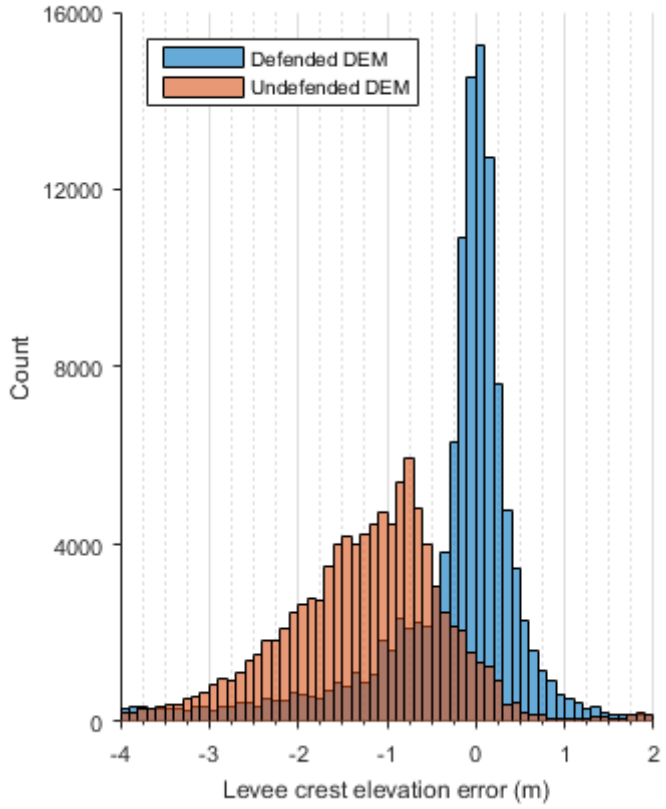


**Figure 6.4** Histograms showing the distribution of elevation errors when benchmarking the defended (blue) and undefended (orange) 1 arc sec DEMs against levee crest elevations in the California Levee Database.

It is evident, then, that the smoothing effect of standard resampling approaches result 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 zero. 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 heterogenous grid cells. Some of the more extreme errors in both DEMs are likely where the NED is not built with lidar data.

### 6.4.2 Inter-scale model comparison: Iowa, US

#### 6.4.2.1 Elevation model building and parameterization

The 1 arc sec DEMs employed in the Iowa analysis, both derived from the ⅓ arc sec NED in this instance, are exemplified in Fig. 6.5. 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 northern bank of Cedar River (compare Fig. 6.5a and 6.5b). In the DEM generated via the levee detection algorithm, the levees are represented to their full effect (compare Fig. 6.5a and 6.5c). Fig. 6.5e 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 ⅓ arc sec data. Fig. 6.5d plots the ⅓ arc sec 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. artefacts in the DEM representation of the channel bed, 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’. Fig. 6.5f 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 centrelines and ignore more modest, yet potentially crucial, levee-like features. Based on numerous visual inspections of the data across the US, an ε threshold of 300 is identified to

be suitable. Fig. 6.5f 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 ε100 and ε300 Iowa DEM is 0.12 m. In other words, pixels where $100 \leq \varepsilon < 300$ 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).

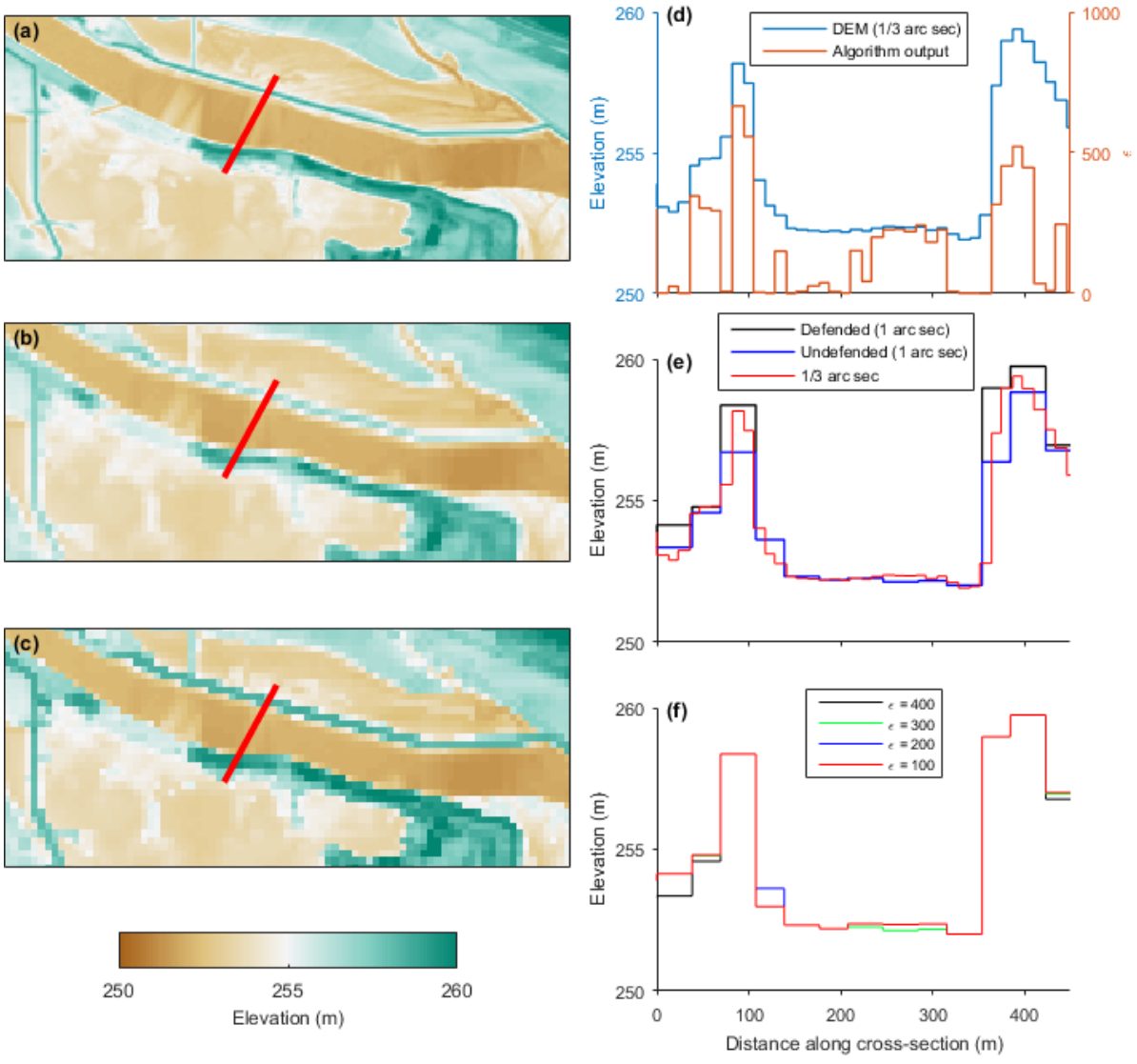


**Figure 6.5** An example of the algorithm's functionality of a 1 x 3 km area along the Cedar river in Waterloo, IA. **(a)** shows the baseline high-resolution (⅓ arc sec) DEM; **(b)** shows an "undefended" DEM, resampled to 1 arc sec with no consideration of levees; **(c)** shows a "defended" DEM generated by the algorithm presented here, where the ε threshold is 300. The red line in **(a)** – **(c)** is the cross-section for which data is 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 ⅓ arc sec 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)**; **(f)** shows the cross-section of elevation values for when different ε threshold values are used.

#### 6.4.2.2 Statewide model comparison

Firstly, the results of the comparison to lower-quality, wide-area flood maps of the entire state are shown in Table 6.1. 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 inundated 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 modelled), 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 $km^2$) are modelled and the pluvial model simulates flooding in rivers smaller than this. The IFC data is an assemblage of local studies, which generally map rivers down to ~3 $km^2$ flow accumulation, though do not model every one: often for justified reasons, because such rivers are in uninhabited areas and there simply is no need to expend resources on modelling 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 Fig. 6.6a for an example of this, where there is broad flood extent agreement in areas the IFC has modelled but non-genuine false alarms in unmodelled-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 dataset.

| 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 |

**Table 6.1** Pixel-count test scores, comparing the statewide IFC maps and Fathom-US run with the new defended DEM.

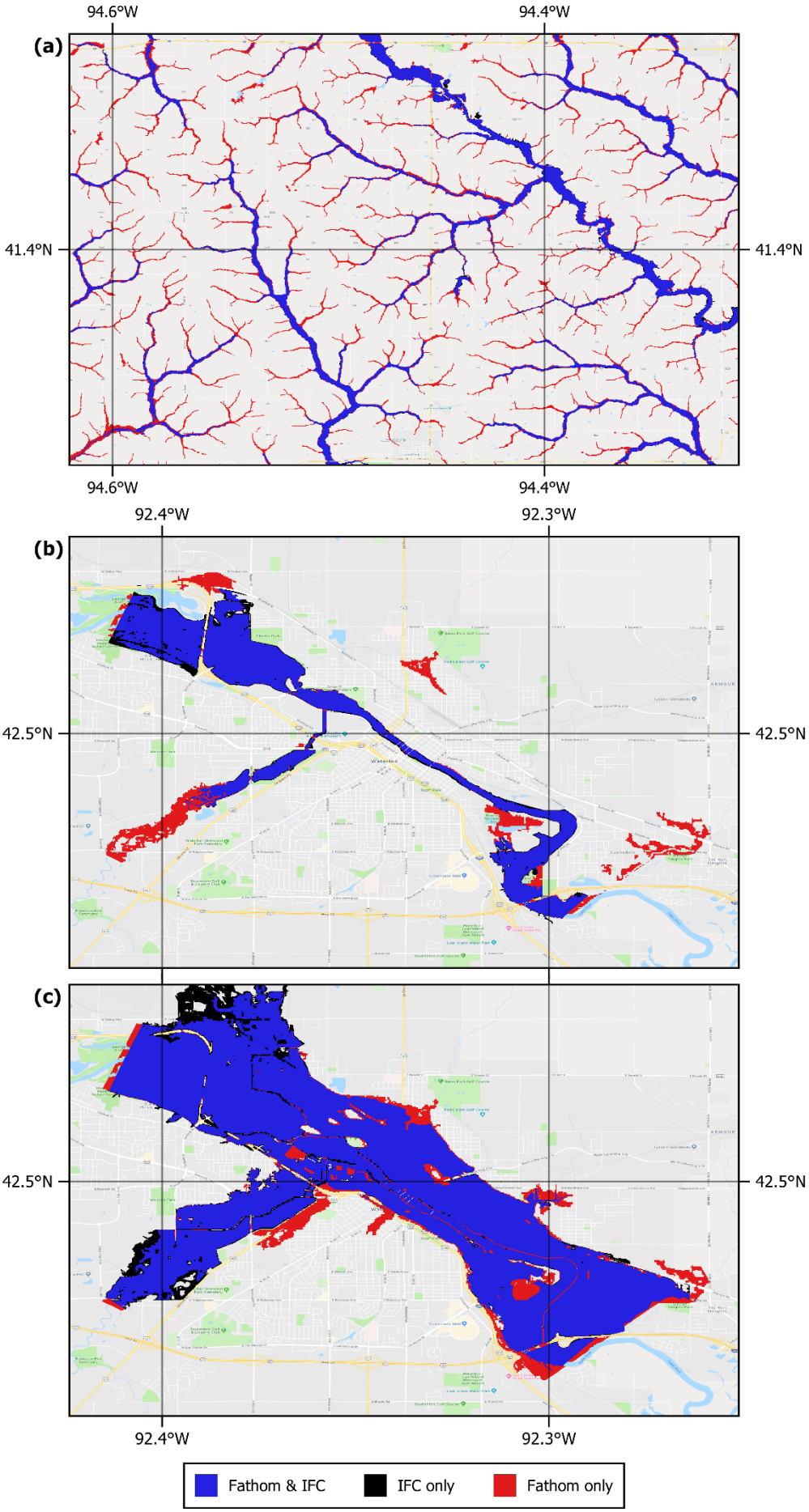
(a)
94.6°W
94.4°W
41.4°N
41.4°N
94.6°W
94.4°W
(b)
92.4°W
92.3°W
42.5°N
42.5°N
(c)
42.5°N
42.5°N
92.4°W
92.3°W
Fathom & IFC
IFC only
Fathom only

**Figure 6.6** Example of the model intercomparison between Fathom-US and the IFC's **(a)** statewide models of the 1% 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.

#### 6.4.2.3 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 have modelled a single river flowing through the urban centre, 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-modelled channel, yet inundation arises from overtopping in the (unmodelled by IFC, modelled by Fathom-US) tributary. See Fig. 6.6 for an example at the Waterloo study site. Some areas flagged as false alarms for the 10% AEP flood (red in Fig. 6.6b) are not connected to the IFC-modelled Cedar River, but around IFC-unmodelled 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 (Fig. 6.6c) 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.

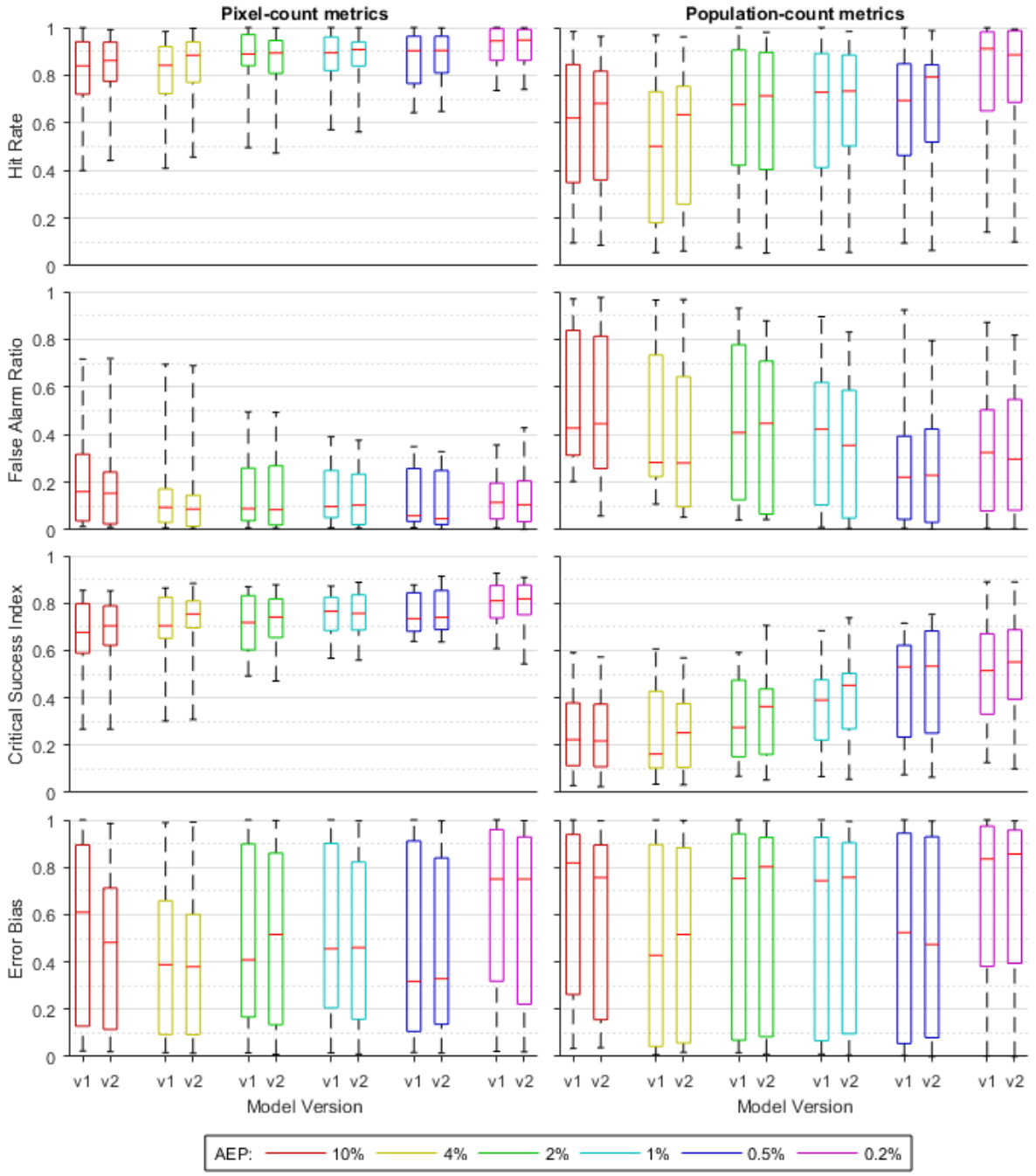


**Figure 6.7** Boxplots of the binary pattern metrics when comparing version of Fathom-US to the IFC models of 27 urban areas in Iowa. The left-hand side shows test scores using raw pixel counts in Eqs. 6.7–6.10. The right-hand side shows test scores using population counts in Eqs. 6.7–6.10. The left 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.

Fig. 6.7 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 Fig. 6.8, where performance has increased in Ottumwa (Fig. 6.8i–j), remained broadly constant at Monticello (Fig. 6.8c–d), and slightly decreased at Clarksville (Fig. 6.8e–f). 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 even 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); 0.50 to 0.55 (0.2% AEP). For instance, the 1% AEP flood in Ames (Fig. 6.8g–h) 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 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 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 defences may incorrectly overtop regardless. In Fig. 6.9, it is evident that, in certain circumstances, this method of DEM construction actually exacerbates errors in other facets of the model. In Fig. 6.9a, 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 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 centres of 3 arc sec HydroSHEDS flow accumulation data). The flood walls (Fig. 6.9b) on the W and SE banks of the Missouri have held back flood waters, while the hydrography of the tributary has broken through the flood walls in the N and E 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 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 defences 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 ⅓ arc sec 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 and Orlandini, 2018; Sangireddy et al., 2016), may ameliorate this issue if implemented in the Fathom-US model framework. Future research will explore this.

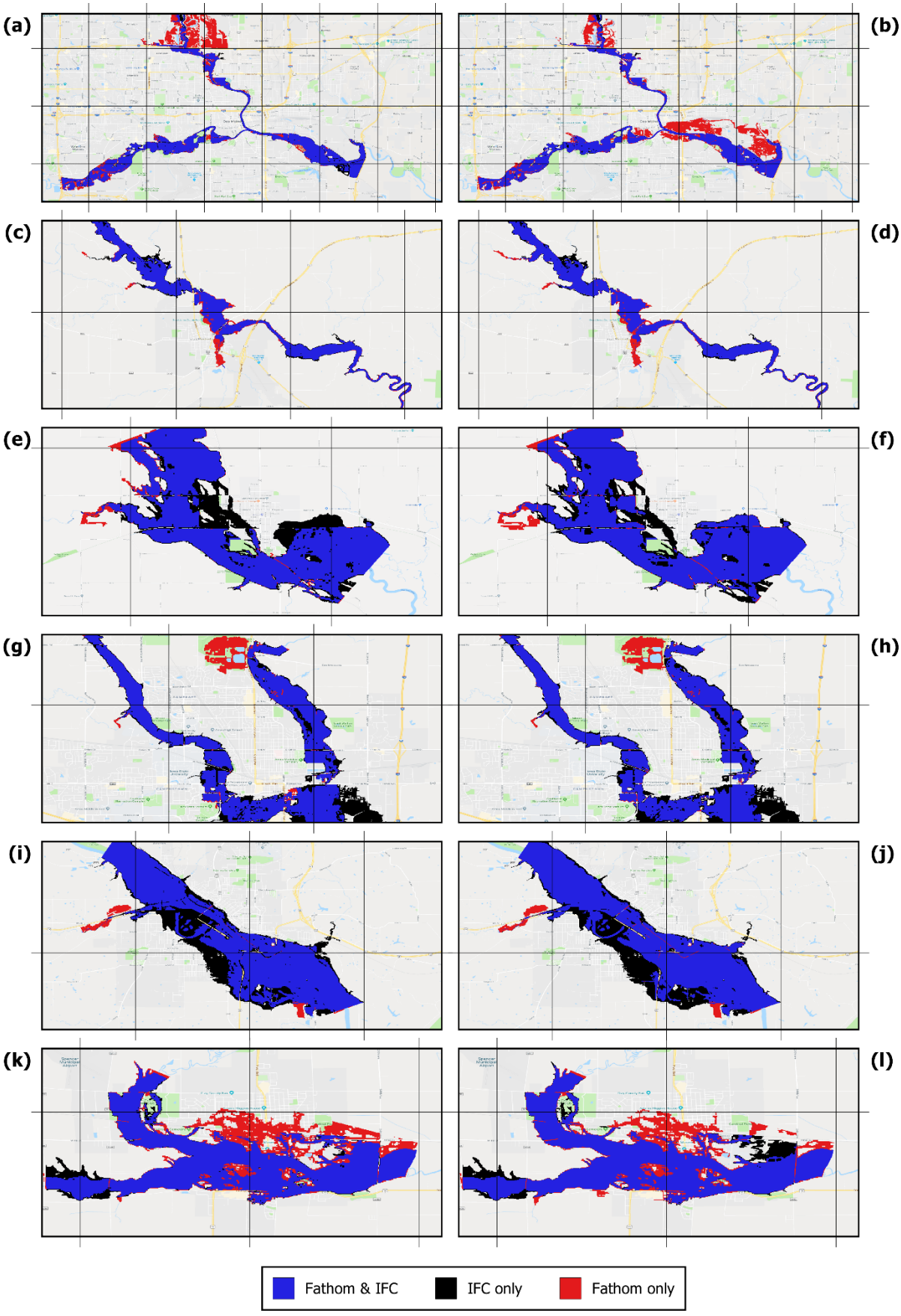
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
Fathom & IFC
IFC only
Fathom only

**Figure 6.8** Examples of model intercomparison between the IFC's urban models and Fathom-US v1 (right-hand side) and v2 (left-hand side). Each grid box is 180 arc sec (~6 km). The table below gives further information:

| Fig. 6.8 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 |

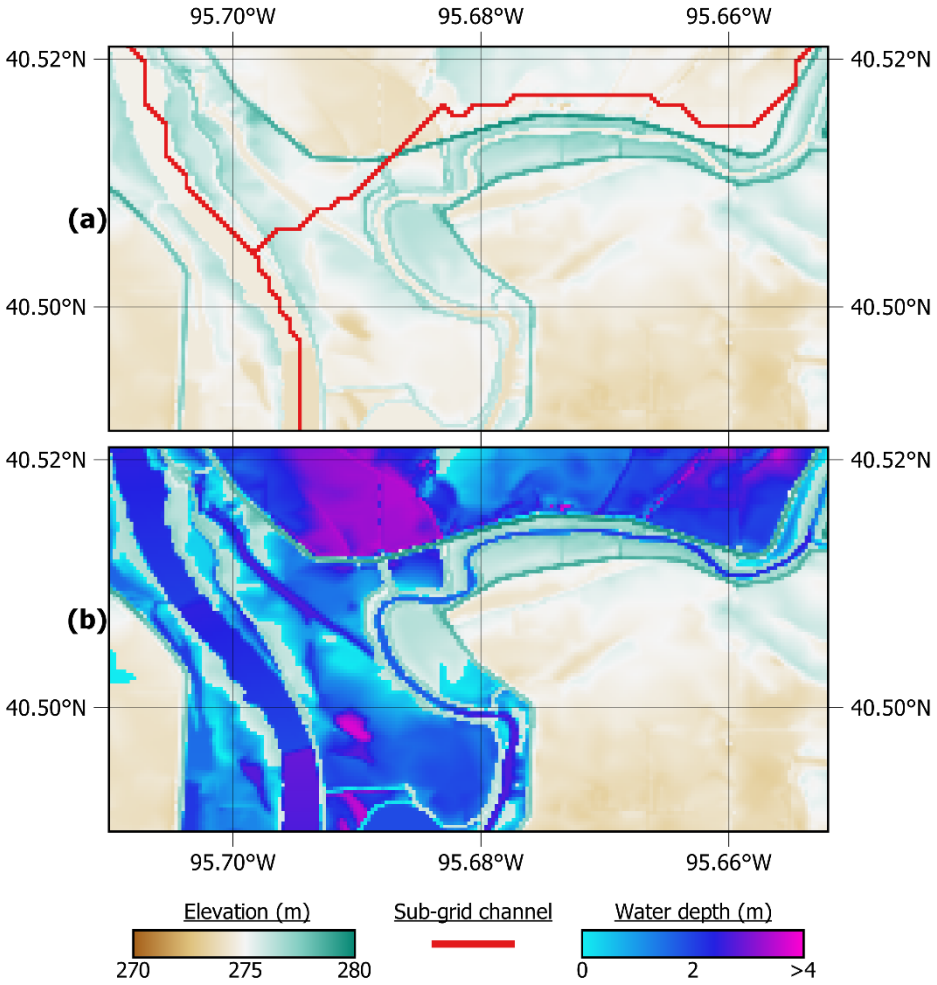


**Figure 6.9** For a stretch of the Missouri River: **(a)** the defended DEM and sub-grid channel network derived from HydroSHEDS employed by Fathom-US, with **(b)** the resultant 2% AEP flood depth grid.

More broadly, the population-weighted metrics indicate there is a lot more work to be done to resolve the scale performance gap in inundation modelling than would be suggested by the pixel-based metrics. The pixel-based scores here and in Wing et al. (2017) suggest skill between local and large-scale models is converging, with CSIs approaching a ceiling given the fundamental hydraulic modelling constraints imposed by extreme flow characterization (Wing et al., 2018b). 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 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–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 skilful model structure. Furthermore, in spite of accurate elevation data, channel hydrography and bathymetry, computational hydraulics, and defence information, they are impaired by the fundamental inability of all models to accurately characterize extreme flows. Even for models fortunate enough to have boundary conditions informed by a river gauge, their short records and subjection to non-stationarity (particularly of land-use) result in multiple defensible interpretations of given AEP flows (particularly less frequent ones). The justification, therefore, for one model being ‘validater’ 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. (2009a), 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 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 behavioural than previously. In Fig. 6.8a, levee representation in Des Moines has resulted in constrained 10% AEP flooding through the urban centre, which v1 struggled to do (Fig. 6.8b). Similarly in Waterloo, the 10% AEP flood was constrained (Fig. 6.5b). Yet, it is evident that floodplain

feature preservation is no panacea: uncertainties relating to other model components inhibit substantial gains in performance. In Fig. 6.8k, 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 1265 $m^3s^{-1}$, while the input to Fathom-US was 1580 $m^3s^{-1}$. It is unsurprising, therefore, that simply differing the method of DEM construction did not reduce the false positives evident in v1 (Fig. 6.8l).

### 6.4.3 Observation-based validation: Po River, Italy

#### 6.4.3.1 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 dyke system, but not so large that the major embankments were overtopped (estimated AEP of 2%). The most obvious test of the three model runs (manual-defended, automated-defended 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 (Fig. 6.10a–di). The AD model (ε threshold of 300 in light of its functionality in section 6.4.2 and crest elevation accuracy in section 6.4.1) does this too: flow is contained entirely within the floodplain bounded by the outer levees (Fig. 6.10a–dii). Crucially, the undefended model run simulates an overtopping of these outer levees across the domain (Fig 6.10a–diii). Though a somewhat aggregated and narrow view of model performance, levees holding in both defended model runs while overtopping in the undefended run demonstrates the algorithm presented here represents flood defences 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 defences 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 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 modelling 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 modelling challenge (Castellarin et al., 2011a). Even in

the presence of complete levee inventories, implementation of these data within the model build is sensitive and non-trivial. Meanwhile, the automated method presented reproduces experiential expectations of levee behaviuor akin to local modelling strategies.

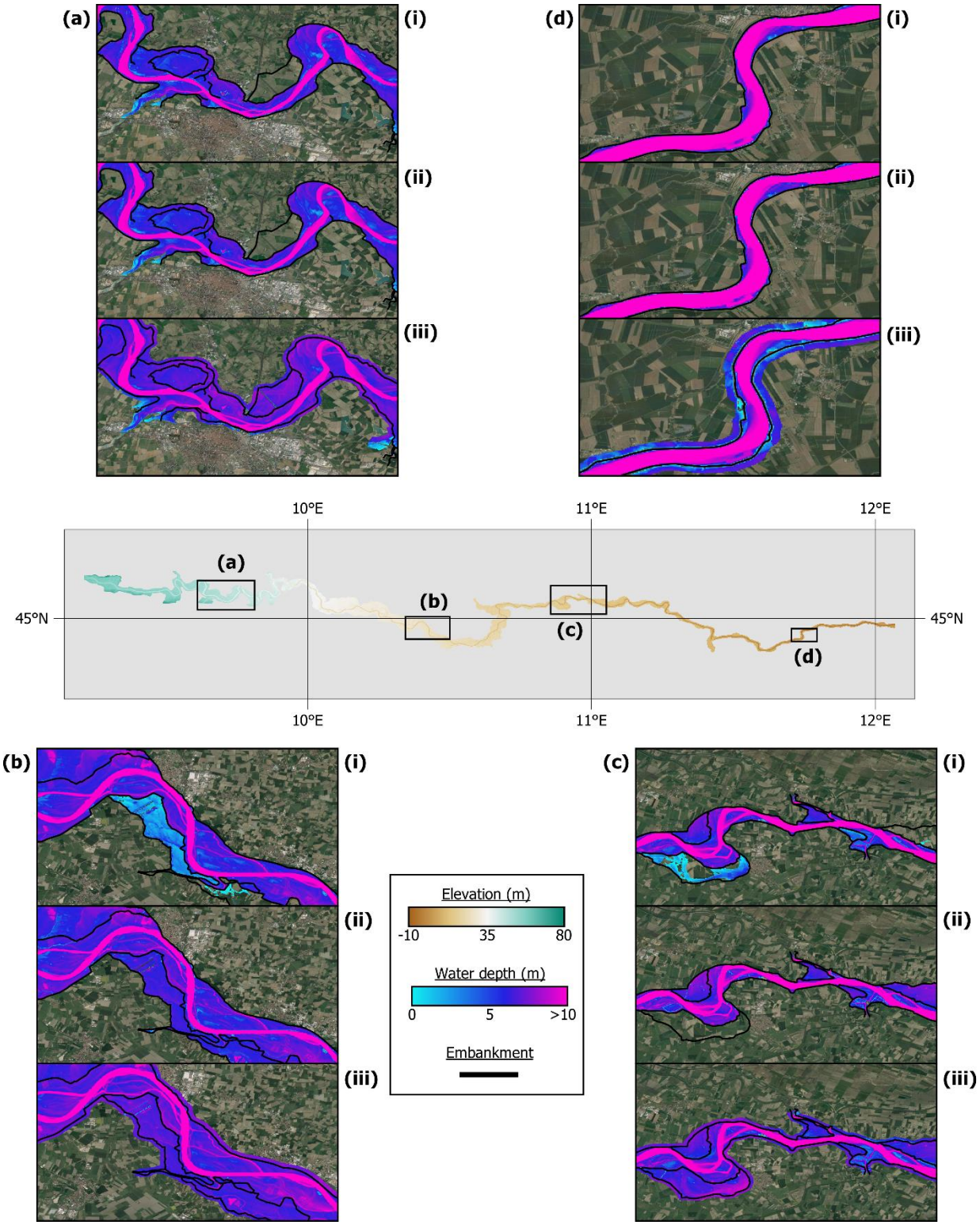


**Figure 6.10** Four areas – **(a)** Piacenza, **(b)** Casalmaggiore, **(c)** just upstream of Ostiglia, **(d)** just downstream of Pontelagoscuro – of the simulated 2000 flood event along the Po River

for the three different DEMs: **(i)** manual-defended; **(ii)** automated-defended; **(iii)** undefended. The DEM 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)**.

#### 6.4.3.2 Validation against high water marks

With the two simulations involving some treatment of defence 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 (Fig. 6.11). For the MD simulation (Fig. 6.11a), 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 (Fig. 6.11b), 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 defences than in the ubiquitous manual approach, in spite of friction being calibrated to the latter.

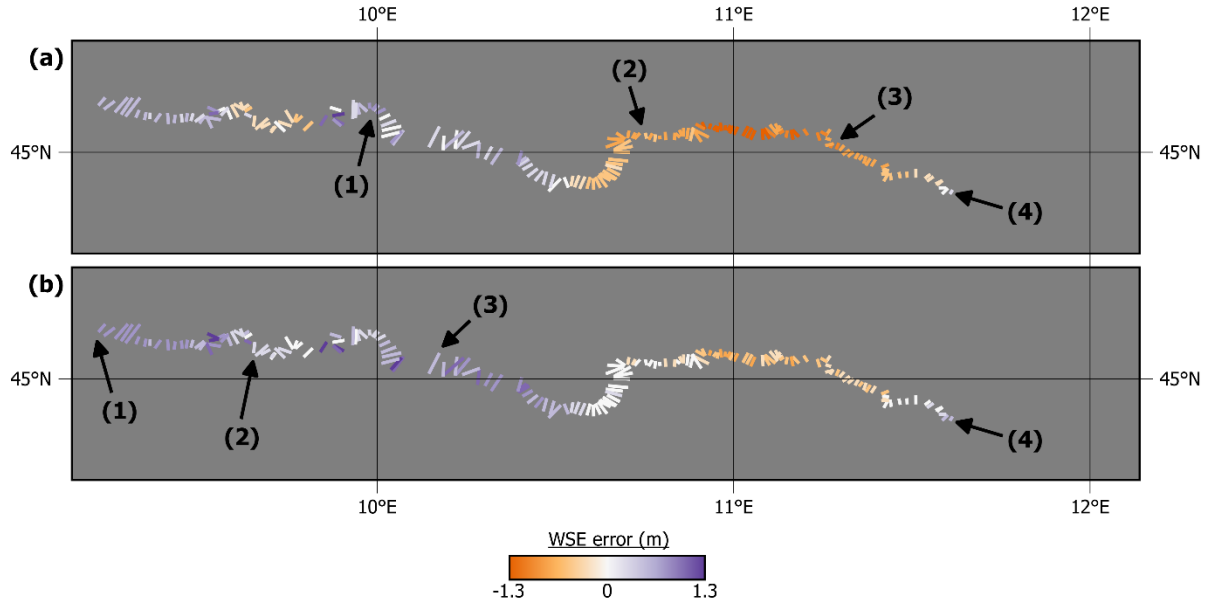


**Figure 6.11** Maps showing the spatial distribution of maximum water surface elevation errors at each cross-section when comparing observations to **(a)** manual-defended and **(b)** automated-defended model output. Positive errors (greys) indicate model overprediction; negative errors (reds) indicate model underprediction. The direction of flow is west to east. The numbers in **(a)** correspond to river gauges in Fig. 6.14. The numbers in **(b)** refer to locations in Fig. 6.12 or are referred to in the text.

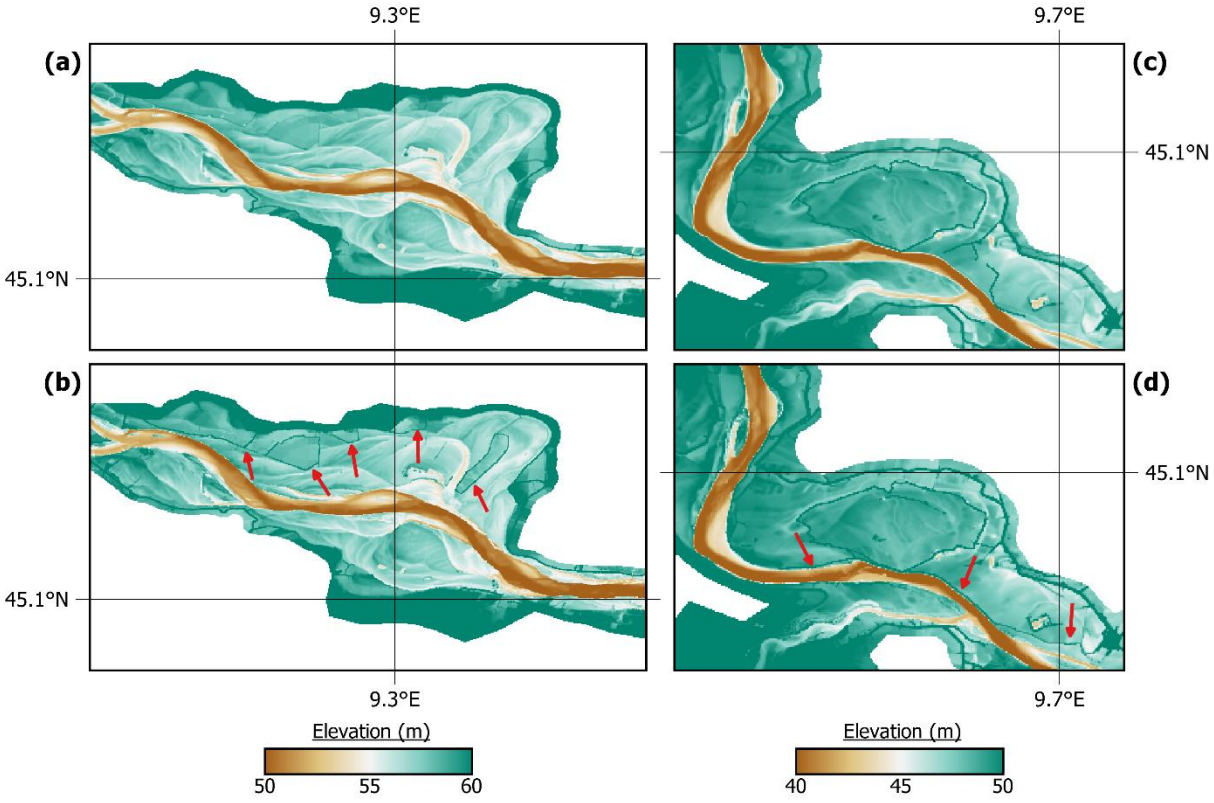

**Figure 6.12** Maps highlighting the minor, yet crucial, MD and AD DEM differences (indicated by the red arrows). **(a)** and **(b)** are MD and AD DEMs respectively of the upstream domain boundary, at the confluence of the Po and Ticino rivers downstream of Pavia (location 1 in Fig. 6.11b). **(c)** and **(d)** are MD and AD DEMs respectively of an area just upstream of Piacenza (location 2 in Fig. 11b).

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 Fig. 6.11b) and Pontelagoscuro (location 4 in Fig. 6.11b). 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 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. Fig. 6.12 exhibits this for areas in the upstream model domain. Levee compartments near the domain entrance (Fig. 6.12a–b) were seemingly absent from the levee inventory used by the MD DEM, but were automatically detected for the AD model. Near Piacenza (Fig. 6.11c–d), 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 micro-topographic features that remain sub-grid to the MD model resulted in faster conveyance of floodwaters downstream in the AD model, culminating in a more accurate representation of maximum WSE in the easterly (downstream) portion of the domain. Fig. 6.13 illustrates this assertion. The widespread areas of purple shading in Fig. 6.13a indicates such areas were inundated much earlier (6–12+ hours) 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 Fig. 6.13b, which show much deeper waters in the MD than AD simulation after 45 hours of model-time: a volumetric difference of ~2,450,000 $m^3$, 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 in-flow 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).

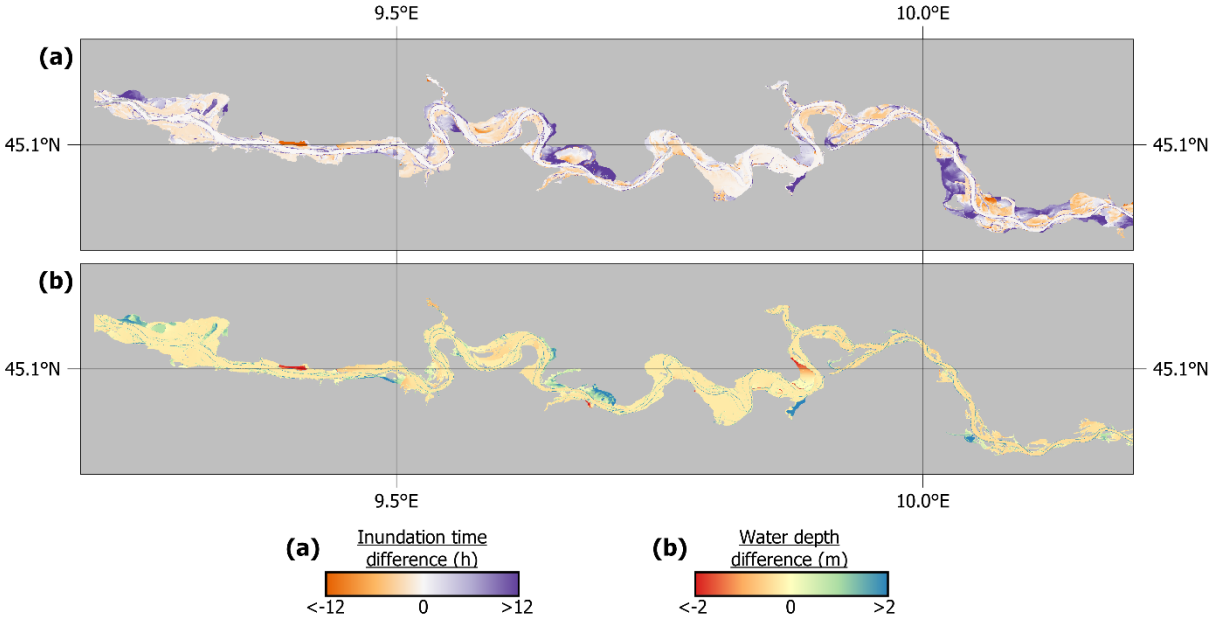


**Figure 6.13** Maps of the upstream portion of the Po River floodplain showing: **(a)** time of inundation difference between the models, where a positive (negative) difference indicates earlier (later) inundation in the MD than AD model; and **(b)** water depth difference 45 hours into the simulations, where a positive (negative) difference indicates deeper (shallower) water in the MD than AD model.

### 6.4.3.3 Validation against river gauge data

This phenomenon is yet further evidenced by comparison to river gauge observations of water surface elevation. When comparing modelled WSE to those measured by the gauge at Cremona (upstream section of the domain; location 1 in Fig. 6.11a), both the MD and AD models perform similarly (Fig. 6.14a). 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 – 2 cm), providing suitably discriminatory validation data (Hamilton and Moore, 2012; McMillan et al., 2012; Schmidt, 2002; van der Made, 1982). At Borgoforte (location 2 in Fig. 6.11a), the MD model underpredicts the peak to a much greater magnitude (Fig. 6.14b). The AD model replicates observations well here. Disparities on the rising limb are predominantly a function of model spin-up, e.g., no water reaches Borgoforte until ~50

hours of model time; the Po would not be empty before this time in reality. At Sermide (location 3 in Fig. 6.11a), the MD model is severely underpredictive: misestimating peak WSE by ~1 m (Fig. 6.14c). The AD model underpredicts the peak also, but only by ~0.5 m. The geometry reproduced by the AD approach seems to ensure a more faithful reproduction of the channel–floodplain interaction, providing an attenuation more similar to observations than with that obtained by the MD model. However, at the furthest downstream gauge (Pontelagoscuro; location 4 in Fig. 6.11a), both models become slightly overpredictive (Fig. 6.14d). 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. Pontelagoscuro 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 water surface elevation 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. Fig. 6.14e is a cross-section of the channel and floodplain at Pontelagoscuro. 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 (⅓ arc sec) source data, simply as a function of grid resolution. Preserving peaks from the ⅓ arc sec data in the 1 arc sec DEM results 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 ⅓ arc sec DEM was used. This is an important limitation to representing defences 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 Pontelagoscuro is the probable product of more accurate modelled 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 defences, but can outperform them in ensuring representation of other hydraulically important sub-grid 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 such an approach may offer benefits to local modellers, too, in improving model-build efficiency.

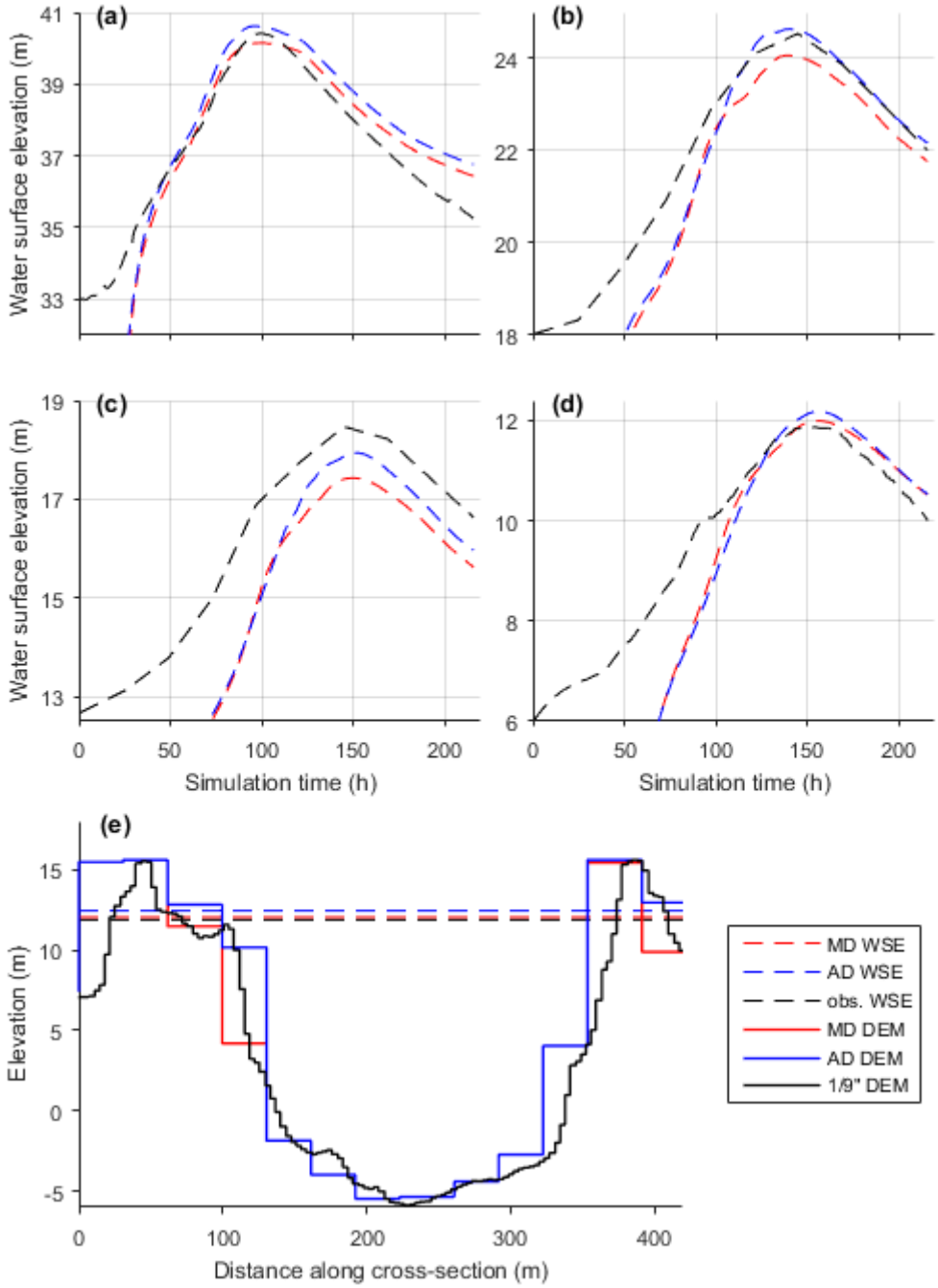


**Figure 6.14** Comparison of MD and AD modelled water surface elevations to four river gauge observations of the 2000 event at: **(a)** Cremona (hourly data); **(b)** Borgoforte (daily data); **(c)** Sermide (daily data); **(d)** Pontelagoscuro (hourly data). Locations **(1)** – **(4)** in Fig. 6.11a correspond to gauges **(a)** – **(d)**. **(e)** is a cross-section of MD, AD and baseline ⅑ arc sec DEMs as well as MD, AD and observed water surface elevations at Pontelagoscuro.

## 6.5 Conclusions

Overall, then, the new method of defence representation in DEMs presented here has significant skill in accurately defining levee crest elevations. When validated against the California Levee Database (section 6.4.1), 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 modelling 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 6.4.2), 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 inter-scale model deviations in a risk context were substantial. The Po test case (section 6.4.3), though, offers crucial evidence in support of the validity of this method when applied in conjunction with more accurate data: (i) being a real, gauged flood event, boundary condition uncertainty is much reduced relative to the theoretical design floods in the Iowa tests; (ii) complete channel bathymetry and accurate hydrography offer considerable improvement in river representation relative to the HydroSHEDS sub-grid channels employed in the US; (iii) total high-resolution lidar coverage of the Po results in a more accurate DEM than that of Iowa when using the USGS NED; and (iv) validatory benchmarks are better defined for the Po (observed high water marks, gauged stage, 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. Firstly, it requires a high-quality, high-resolution baseline DEM that a modeller would typically coarsen. This effectively rules out the majority of the globe, where the best source of free elevation data remains corrected SRTM-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 neighbourhood-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 defence locations is problematic, given this would unintentionally penalize the detection of hydraulically-important, non-levee features as false alarms, even if the reference defence 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 judgement 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 over-coarsened 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 and Bates, 2018); updates to the USGS National Hydrography Dataset may supplant the cruder global HydroSHEDS representation of river channels for the US and MERIT Hydro shows promise at the global scale (Yamazaki et al., 2019); and emerging datasets 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. Yoon et al., 2012; Pavelsky et al., 2014). Meanwhile, this method may find applicability in smaller-scale studies given the improvements to common local modelling 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 defences and other important hydraulic features in the absence of local information. It provides a crucial first step in ensuring physically realistic flood defences in large-scale models, moving away from the largely unsubstantiated assumptions underpinning previous approaches. When it is a lack of defence information that is the dominant source of error in a hydraulic model, the method presented here offers considerable improvement.

### 6.6 Acknowledgements

The US Army Corps of Engineers National Levee Database is available from https://levees.sec.usace.army.mil/. The US Geological Survey National Elevation Dataset can be accessed from https://ned.usgs.gov/. The socio-economic data in Fig. 6.1 is available from the US 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 California Levee Database. The Iowa Flood Centre benchmark inundation maps are available from the Iowa Flood Information System (https://ifis.iowafloodcenter.org/). The dasymetric population map of the US 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. OEJW, PDB, JCN, CSS, AMS and NQ are affiliated with Fathom, a flood analytics firm based in the UK. PDB was supported by a Research Fellowship from the Leverhulme Trust and a Royal Society Wolfson Research Merit Award. JCN was supported by National Environmental Research Council grant NE/S006079/1. DWG, RG and WFK were supported by the Iowa Flood Centre.

# 7 Conclusions

## 7.1 Context

High profile flood events in recent years have brought the issue to the forefront of public and political consciousness, yet our ability to mitigate their impacts is hampered by a lack of understanding of flood risk at large spatial scales. Physical models of flood inundation, required amidst a paucity of observations of flood extent, are used for planning and regulation of development, pricing flood insurance, and emergency response. Such models, though, are traditionally only built in specific locations where stringent data requirements are met, leaving the vast majority of the world's rivers un-modelled and their flood risk poorly understood. Recent advances in computational capacity and data availability have enabled a scale up in the domains considered by a hydraulic modeller. Global flood models have been presented by a number of academic and commercial entities, predominantly as crucial input components with near-global coverage became available thanks to spaceborne sensors. All global flood models in the Trigg et al. (2016) intercomparison use global SRTM data as the DEM, while the solution to quantifying flow inputs in ungauged basins is split between cascading global precipitation data through rainfall-runoff models or a statistical model relating basin characteristics to gauged river flows. These large-scale models, though, operate amidst significantly reduced data fidelity compared to local-scale analogues, which are often built with lidar-derived DEMs and directly measured river flows. The effect of this on model performance is not well understood. Sporadic global flood model evaluations suggest they have relatively low skill (as many model cells disagreeing with a given benchmark as agree), but this remains to be robustly evidenced at any commensurate spatial scale in inter-scale model comparisons or in validation against observations. These under-validated models have been employed in multiple analyses that quantify present and future flood risk (e.g. Hirabayashi et al., 2013; Winsemius et al., 2016; Alfieri et al., 2017), vulnerability (e.g. Jongman et al., 2015), the cost-effectiveness of flood defences (e.g. Ward et al., 2017), risk to transport infrastructure (e.g. Koks et al., 2019), and other applications. Applying a flood model whose skill is more robustly benchmarked in such applications will lead to increased confidence in its output. At present, conclusions drawn by these studies are subject to considerable – and, more importantly, unquantified – uncertainty.

Existing studies that focus on large-scale flood risk quantification not only suffer from inadequately justified hydraulics, but also from poor representation of socio-economics. Coarse grid cells (~1 $km^2$) imprecisely distribute quantities such as people and assets,

resulting in significant misestimates of flood risk (as evidenced by Smith et al., 2019). Thus, while the climate and hydrology models used in these large-scale analyses are fundamentally limited by data availability, process understanding and computational capacity, their hydraulic and socio-economic components are not limited in this way. Incorporating more physically-based and comprehensively validated hydrodynamics with higher resolution exposure representation can lead to increased accuracy in risk assessments at-scale. Furthermore, the effect of climate change on flood hazard is considerably uncertain. Sound physical reasoning suggests an enhanced hydrological cycle in a warming world will lead to increased precipitation and, by extension, flooding in more places than not, but examinations beyond this qualitative statement are inaccurate in the context of inland floods. Meanwhile, past and projected increased human activity in floodplains as a driver of flood risk dwarfs the uncertain climate change signal (Bouwer, 2010; Handmer et al., 2012; Winsemius et al., 2016). Thus, a more faithful representation of exposure data will lead to greater understanding of future, as well as present-day, flood risk.

New methods were required to herald the scale-up from local to global flood models: processing of elevation data, developing fast and accurate hydraulic codes, approximating channel geometry and estimating extreme flows. The handling of flood defences, though, is cited as a major deficiency in the current generation of large-scale models. Publicly available information on the location and standard of structural flood defences are sparse and incomplete, meaning large-scale models originally considered “undefended” simulations only. The advent of FLOPROS (Scussolini et al., 2016) offered an estimate of protection standards, but only at an aggregated scale (i.e. sub-country units). Some large-scale models used FLOPROS to assume all floods more frequent than the stated standard do not cause damage (e.g. Alfieri et al., 2017; Winsemius et al., 2016), but the assumption that these general protection levels apply homogenously to every flood source in a region is not well founded. Furthermore, it is unclear whether either protection standard or defence location can be reliably inferred or predicted based on socio-economic variables. Thus, while most facets of large-scale flood models have received considerable attention, adequate flood defence representation is a frontier yet to be addressed.

This brief overview of chapter 2 (literature review) motivated the research objectives specified in section 2.3:

1. Perform a comprehensive validation of flood inundation modelled over large spatial domains:
    a) by comparison to ubiquitous, engineering-grade hydraulic models of a number which constitute a commensurate spatial scale to the large-scale approach;
    b) by comparison to real-world flood event observations, thus necessitating coupling the hydraulics to meteorological and hydrological models;
    c) through evidencing increased accuracy relative to low-complexity, zero-physics approaches.
2. Explore new applications for a hydraulic modelling approach which avoids the pitfalls of low-coverage local models and current low-accuracy large-scale models:
    a) through incorporation in a flood risk quantification framework with a more sophisticated treatment of exposure;
    b) through the generation of operational flood inundation forecasts (also requiring the advances sought in objective 1b).
3. Tackle the issue of poor flood defence representation epidemic in current large-scale approaches.

### 7.2 Specific findings

This section highlights conclusions drawn in the research chapters (3–6) with respect to the thesis objectives. Limitations and constraints of current or newly presented methods are also presented, offering prospects for future research.

***Hydraulic models built and executed (semi-)automatically at large spatial scales can produce similar simulations of extreme flood inundation to traditional local-scale approaches.***

The first published example of comprehensive large-scale flood model validation was presented in chapter 3. Global flood models have been sparsely validated in a relatively small number of confined test cases, far from the spatial scale they simulate over (e.g. Dottori et al., 2016; Sampson et al., 2015; Ward et al., 2017). These give little indication of the different scenarios under which such approaches are most (or least) appropriate. This chapter employs a continental-scale subset of the Sampson et al. (2015) global model, with the only difference between the continental and global model being the 1 arc sec USGS elevation data used in place of the global model's 3 arc sec SRTM-based DEM. These model structures produce flood maps of a constant return period in space, rather than a real observable event;

necessitating model–model comparisons to evaluate performance. The chapter assumes that FEMA flood maps, built by hydraulic engineers across the US to delineate the 1 in 100 year (1% AEP) flood zone at an estimated cost of up to $7.5 billion to 2013 (ASFPM, 2013), are a benchmark which large-scale flood maps should aspire to resemble. With FEMA models commonly built with manually surveyed river bathymetry, high-resolution lidar data, manual determination of extreme discharge from local flow data and calibration to any available historic flood observations, they are expected to produce a more faithful realisation of the theoretical 1 in 100 year flood than a large-scale model built with no manual intervention, extreme flows quantified by catchment descriptors, publicly available elevation data and approximations to channel geometry.

The large-scale model obtained a hit rate of 0.82 against the FEMA flood maps, increasing to 0.86 when benchmarked against local data specifically categorised as high quality. The Sampson et al. (2015) global model obtained 0.75. Overprediction against FEMA maps was spuriously quantified, since their non-total coverage of streams meant the large-scale model was penalised for simulating flooding in areas not considered by FEMA. Against select USGS studies with specified model domains (thus enabling fair quantification of false alarms), the large-scale model obtained an average critical success index (a fit metric accounting for both type I and type II errors) of 0.76: a score akin to those obtained by validation of calibrated local models to flood extent. Further scenario-based analyses uncovered greater (lower) correspondence with the large-scale and FEMA models for large (small) rivers, in temperate (arid) climatologies, and in rural (urban) areas. Chapter 6 offered further evidence of convergence between the continental and local models in benchmarking against high quality models from the Iowa Flood Centre. Correspondence of the continental-scale model to more approximate statewide maps averaged 0.75 (CSI), with hit rates averaging 0.91. When compared to highly accurate urban flood models, test scores were broadly similar. Larger flood events generally achieved higher correspondence.

The inter-scale model comparison performed in this thesis highlights the approaching convergence of skill between large- and local-scale models. Fit statistics generally indicate that the two modelling strategies can produce virtually indistinguishable flood extents when simulating extreme events given the inherent error in quantifying extreme flows. Global models reflecting the characteristics of the continental-scale one tested here – with 2D hydrodynamics, a DEM with USGS NED-like accuracy, gauge-based river flows and

reasonable channel approximations based on hydraulic geometry theory – can obtain similar results. The clearest priority here, given the Sampson et al. (2015) model contains all but one of these features, is improved mapping of global topography. The continental-scale model was built, executed and could be re-run at a small fraction of the time and money expended on building FEMA models, while maintaining total coverage of US streams.

***Large-scale hydraulic models when executed in the context of event replication can provide similar delineations of maximum flood extent to observations, but accurate replication of observed water levels is dependent on DEM accuracy.***

The hindcast simulation of Hurricane Harvey flooding presented in chapter 5 obtained an average hit rate of 0.78 and a critical success index of 0.66 when compared to flood extents derived from observed high water marks. In an event dominated by surface runoff rather than high river flows, this presents a difficult modelling challenge: capturing roughly 4 in every 5 wet cells and keeping false alarms to <20% is indicative of reasonable model performance and provides a useful indication of impacted areas. When compared to the elevations of the high water marks themselves, the hindcast model mis-estimated water surface elevation by 1.0 m on average (with a mean bias of 0.15 m), though the observation data themselves may contain errors of up to 0.5 m. The distribution of errors in the DEM itself was found to be virtually identical to simulated water surface elevation errors. Indeed, when a fluvial event on the Po River was simulated using lidar data and continuously surveyed channel bathymetry, water level errors were roughly half those in the Harvey simulation. The effect of calibration (applied to local model benchmarks and the Po River model) should be noted, however. The measures of fit and model error are related to an uncalibrated model; calibrated models outperforming these does not necessarily indicate they are inherently more skilful.

***Models based on some formulation of the shallow water equations outperform simplified approaches: confirmed again at-scale***

When Bates and De Roo (2000) presented the first iteration of LISFLOOD-FP (1D kinematic channels; 2D diffusive floodplains), their physically-based model achieved greater correspondence with observations of flood inundation than a simple planar lid. Chapter 5 reiterates this for a flood inundation simulation of Hurricane Harvey. When driven with flow observations, the Height Above Nearest Drainage (HAND) planar approximation correctly identified 41% of cells while the mass- and momentum-conserving continental-scale model correctly identified 66%.

***Large-scale flood hazard models can provide updated estimates of flood risk, highlighting the mis-estimated quantities predicted by an amalgamation of local assessments and cruder global approaches.***

Chapter 4 employed the continental-scale model in a flood risk quantification framework, providing estimates of exposed population, assets and potential damages. The relatively sparse coverage of FEMA maps highlighted in chapter 3 (roughly 60% of CONUS land area) suggested that some areas may be at risk but lack a local flood model indicating this. In chapter 4, the total number of people living in the 1 in 100 year flood zone was calculated to be 41 million. According to FEMA maps, this quantity is 13 million. Despite the similarity of the continental model to FEMA maps where they exist, the latter's failure to model every river translates to a threefold underestimate in exposed populations. Prior to the study of Smith et al. (2019) detailing this phenomenon more fully, chapter 4 suggested coarse representations of exposure can lead to mis-estimates of flood risk also. The risk framework presented by Ward et al. (2013), with populations summarised by 1 $km^2$ cells, achieved similar underestimates of 100-year flood exposure to FEMA. Chapter 4 employed population maps generated by the US Environmental Protection Agency at 30 m resolution, providing a more precise distribution of exposure. With a more thorough delineation of potentially hazardous areas, flood maps derived from large-scale hydraulic models can ensure floodplain regulations and insurance pricing is consistent and more reflective of true risk.

***Forecasts of flood inundation with reasonable latency are now possible with emerging large-scale hydrodynamic modelling frameworks, extending traditional forecasts which typically neglect potential inundation.***

Chapter 5 presented an appendage to typical flood forecasting cascades by coupling the continental-scale model structure to output from meteorological and hydrological models. With a new domain defined by the potential landfall of an incoming tropical cyclone, NOAA Quantitative Precipitation Forecasts were input to the pluvial model structure, NOAA National Hurricane Centre storm surge forecasts were propagated inland from the coast and NOAA National Water Model forecast streamflows guided sampling of fluvial flood maps from the existing inventory. New dynamic simulations (pluvial and coastal) required roughly 6 hours of computation on a single 20-core node of a HPC cluster, while the sampling of pre-existing fluvial data was near-instant. Comparison to observations suggests model skill in identifying impacted areas, though DEM errors inhibited accurate floodplain water level prediction. In providing more relevant information for an end-user in the form of potential

flood inundation rather than at-a-point flow forecasts, emergency responses may be more effectively mounted to such events. The concept of sampling from a pre-simulated library of inundation simulations, at no additional computational cost to traditional forecast cascades, could be seamlessly integrated into a probabilistic ensemble forecast.

***Existing methods of flood defence representation in large-scale hydraulic models are unjustified.***

While most papers in which large-scale flood models are presented cite flood defence representation as a major deficiency in their framework, recent solutions remain unsubstantiated: the assumption of predictability of defence location and standard based on socio-economic data (e.g. Feyen et al., 2012; Quinn et al., 2019), the every-river-defended assumption underpinning the application of FLOPROS (e.g. Alfieri et al., 2017; Winsemius et al., 2016) and the use of publicly available defence inventories in the continental-scale model presented in this thesis. The defence database from the US Army Corps of Engineers is considered to contain roughly 30% of the nation's levees (ASCE, 2017), with these missing data evident in model performance in some urban areas in chapter 3. Chapter 6 disproves an aggregate relationship between flood defence location and standard and potential predictors such as land-use, household wealth and government spending. With resulting estimates of risk highly sensitive to assumptions relating to the frequency of flooding required to cause damage, existing risk analyses will significantly misrepresent actual risk.

***Structural flood defences can be dynamically represented in large-scale hydraulic models through the identification and preservation of levee-like features in the source elevation data.***

Chapter 6 presented a new method of flood defence representation as a response to the highlighted deficiencies in current practice. The DEM underlying the continental-scale model is available at resolutions 3–9 times finer than the 1 arc sec variant employed in the dynamic simulations, with the coarser DEM employed for computational efficiency. Information on crest elevations inherent in the higher resolution versions is therefore smoothed out of the 1 arc sec DEM. Chapter 6 presented an algorithm to capture flood defence information in these data and ensure its representation at the computationally efficient 1 arc sec resolution. The higher resolution DEMs were translated to geomorphometric parameters – derivatives of elevation such as slope, aspect and curvature – and the characteristics of known levee locations (based on the partial coverage USACE data) was sampled. The geomorphometric

traits of known levees then informed the algorithmic search for undocumented levees or features which exhibit these hydraulically relevant characteristics. Higher resolution elevations of pixels satisfying the algorithm are then preserved during the necessary grid coarsening.

When benchmarked against surveyed crest elevations in the California Levee Database, the new "defended DEM" exhibited a median underprediction of 0.05 m, compared to the 1.16 m of the original DEM which had no further conditioning beyond standard bilinear resampling. The continental-scale model re-run with the defended DEM was tested against high quality local-scale models of the Iowa Flood Centre. The updated model generally saw higher correspondence with the Iowa flood maps than the original version, particularly in populated areas. However, in areas where extreme flow quantification and river channel representation uncertainties dominated, improvements to defences were scarcely captured in the validation metrics. Further testing on the Po River, a heavily leveed river basin in northern Italy, found similar performance in the replication of an historic flood event to ubiquitous engineering-grade models, while a model with neither the new nor a manual treatment of flood defences was non-behavioural. The new automated method has key advantages over traditional approaches in not requiring *a priori* understanding of relevant flow-controlling features, meaning observed storage and conveyance of water on the floodplain is more accurately replicated when only formal levees are considered.

### 7.3 Final comments

This thesis evidences approaching convergence of skill between large- and local-scale hydraulic models as well as their ability to reasonably replicate observed events, while simplified models which do not consider flow physics are less accurate. A key deficiency with regards to flood defence representation is addressed. New insights in the quantification of flood risk at large spatial scales are revealed, with shortcomings in local-scale and previous global-scale approaches evidenced. New applicability of wide-area hydrodynamic models can be found in flood forecasting, providing crucial distributed data on potential impact to first responders. These research advances, though, also highlight a number of future research needs.

The quality of DEMs is repeatedly cited as the control on hydraulic model quality; Horritt and Bates (2002) found topography to be the primary control on flood inundation simulation

quality. This is found again here: (i) chapter 3 found increased performance of the NED-based continental-scale model against the FEMA benchmark relative to the SRTM-based global-scale model; (ii) chapter 5 found the distribution of water surface elevation errors in an event-replicating model to be near-identical to those in the underlying DEM; and (iii) in chapter 6, the baseline 2 m resolution lidar data used in the Po test case resulted in greater correspondence to a given benchmark than either the Iowa or Harvey (chapter 5) test cases. Indeed, the solution to defence representation offered in chapter 6 is (entirely) dependent on the quality of the DEM. While not necessarily a new finding or research aspiration, a lack of high accuracy global terrain data of lidar-like quality continues to be the primary barrier to greater understanding of flood hazard globally, so this thesis reiterates the multiple calls for compilation of such data (Schumann, 2014; Schumann and Bates, 2018; Winsemius et al., 2019).

The need for commensurate treatment of model components is evident throughout the research presented here. There is little necessity for the added complexity in flood defence representation (chapter 6) where elevation data are poor (indeed, the new method would not be applicable in such an instance since it requires a high resolution, high-accuracy DEM) and inaccurate channel representation drives low confidence in model output. In particular, the disparities in model performance between the Iowa and Po test cases in chapter 6 evidence the importance of accurate channel representation. The Iowa models contain river hydrographies ultimately derived from SRTM, resulting in sub-grid channels being located far from their "supra-grid" representation in the DEM in some instances. Ensuring levee crest heights are represented in the DEM will not be the dominant source of error where this is the case. Channel bathymetry is approximated based on geomorphological relationships, but these will not be universally applicable to every river. The 1 in 2 year bankfull capacity assumption may break down for rivers which flood seasonally and for engineered waterways (Williams, 1978). The Po test case illustrates the high performance of a wide-area model which contains continuously surveyed bathymetry. The upcoming SWOT mission may offer new solutions to measuring discharge, and thus channel capacity, from space to resolve this (Biancamaria et al., 2016). Emerging hydrography datasets, such as the USGS National Hydrography Dataset, MERIT-Hydro (Yamazaki et al., 2019) and those incorporating Open Street Map river data may aid in ensuring accurate channel locations.

While a framework for the validation of flood hazard models has been presented here, validation of risk estimates is more difficult. A lack of detailed data on historic flood losses inhibits a thorough interrogation of modelled risk. While flood hazard data may be valid and the importance of accurate exposure data has been evidenced by Smith et al. (2019), how hazard and exposure interact to generate risk has not been well tested. Commonly, depth-damage functions are applied to estimate the loss a building might experience for a given water depth, but their universal applicability is unknown. Continued compilation of flood loss data and, more crucially, the release of flood insurance claims data will enable the vulnerability component of large-scale risk to be better tested and constrained.

The future direction of the study of large-scale flood risk thus requires collaboration between physical modelling, remote sensing, social science and economics communities to continue the advances in understanding seen in recent years. This thesis, focussed mostly on physical modelling, indicates that hydraulic models executed at-scale may now inform some decision-making needs where local analyses have not taken place.

# 8 References


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

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

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

Afshari, S., Tavakoly, A. A., Rajib, M. A., Zheng, X., Follum, M. L., Omranian, E., and Fekete, B. M. (2018), Comparison of new generation low-complexity flood inundation mapping tools with a hydrodynamic model, *Journal of Hydrology*, 556, 539–556. doi:10.1016/j.jhydrol.2017.11.036.

Ahmadisharaf, E., Kalyanapu, A. J., and Chung, E.-S. (2015), Evaluating the effects of inundation duration and velocity on selection of flood management alternatives using multi-criteria decision making, *Water Resources Management*, 29(8), 2543–2561, doi:10.1007/s11269-015-0956-4.

Alexander, L. V. (2016), Global observed long-term changes in temperature and precipitation extremes: a review of progress and limitations in IPCC assessments and beyond, *Weather and Climate Extremes*, 11, 4–16, doi:10.1016/j.wace.2015.10.007.

Alfieri, L., Salamon, P., Pappenberger, F., Wetterhall, F., and Thielen, J. (2012), Operational early warning systems for water-related hazards in Europe, *Environmental Science and Policy*, 21, 35–49, doi:10.1016/j.envsci.2012.01.008.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F. (2013), GloFAS – global ensemble streamflow forecasting and flood early warning, *Hydrology and Earth System Sciences*, 17, 1161–1175. doi:10.5194/hess-17-1161-2013.

Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P., and Feyen, L. (2014), Advances in pan-European flood hazard mapping, *Hydrological Processes*, 28(13), 4067–4077, doi:10.1002/hyp.9947.

Alfieri, L., Feyen, L., Dottori, F., and Bianchi, A. (2015), Ensemble flood risk assessment in Europe under high end climate scenarios, *Global Environmental Change*, 35, 199–212, doi:10.1016/j.gloenvcha.2015.09.004.

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

Allamano, P., Claps, P., and Laio, F. (2009), Global warming increases flood risk in mountainous areas, *Geophysical Research Letters*, 36(24), doi:10.1029/2009GL041395.

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

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

American Society of Civil Engineers (2017), *2017 Infrastructure Report Card: Levees*. Reston, VA. Retrieved from https://www.infrastructurereportcard.org/cat-item/levees/

Andreadis, K. M., Schumann, G. J.-P., and Pavelsky, T. (2013), A simple global river bankfull width and depth database, *Water Resources Research*, 49(10), 7164–7168, doi:10.1002/wrcr.20440.

Apel, H., Aronica, G. T., Kreibich, H., and Thieken, A. H. (2009), Flood risk analyses – how detailed do we need to be?, *Natural Hazards*, 49(1), 79–98, doi:10.1007/s11069-008-9277-8.

Arcement, G. J. and Schneider, V. R. (1989), *Guide for Selecting Manning's Roughness Coefficients for Natural Channels and Flood Plains*. U.S. Geological Survey Water-Supply Paper 2339. Reston, VA: USGS.

Archer, L., Neal, J. C., Bates, P. D., and House, J. I. (2018), Comparing TanDEM-X data with frequently used DEMs for flood inundation modeling, *Water Resources Research*, 54(12), 10205–10222, doi:10.1029/2018WR023688.

Archfield, S. A., Hirsch, R. M., Viglione, A., and Blöschl, G. (2016), Fragmented patterns of flood change across the United States, *Geophysical Research Letters*, 43(19), 10232–10239, doi:10.1002/2016GL070590.

Arheimer, B., Dahné, J., and Donnelly, C. (2012), Climate change impact on riverine nutrient load and land-based remedial measures of the Baltic Sea Action Plan, *AMBIO*, 41(6), 600–612, doi:10.1007/s13280-012-0323-0.

Arheimer, B. and Lindström, G. (2015), Climate impact on floods: changes in high flows in Sweden in the past and future (1911–2100), *Hydrology and Earth System Sciences*, 19, 771–784, doi:10.5194/hess-19-771-2015.

Armstrong, W. H., Collins, M. J., and Snyder, N. P. (2014), Hydroclimatic flood trends in the northeastern United States and linkages with large-scale atmospheric circulation patterns, *Hydrological Sciences Journal*, 59(9), 1636–1655, doi:10.1080/02626667.2013.862339.

Arnell, N. W. and Gosling, S. N. (2016), The impacts of climate change on river flood risk at the global scale, *Climatic Change*, 134(3), 387–401, doi:10.1007/s10584-014-1084-5.

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

Apel, H., Aronica, G. T., Kreibich, H., and Thieken, A. H. (2009), Flood risk analyses – how detailed do we need to be?, *Natural Hazards*, 49(1), 77–98, doi:10.1007/s11069-008-9277-8.

Association of State Floodplain Managers (2013), *Flood Mapping for the Nation: A Cost Analysis for the Nation's Flood Map Inventory*. Retrieved from https://www.floods.org/ace-files/documentlibrary/2012_NFIP_Reform/Flood_Mapping_for_the_Nation_ASFPM_Report_3-1-2013.pdf

Bai, P., Liu, X., Liang, K., and Liu, C. (2016), Investigation of changes in the annual maximum flood in the Yellow River basin, China, *Quaternary International*, 392, 168–177, doi:10.1016/j.quaint.2015.04.053.

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

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

Bard, A., Renard, B., and Lang, M. (2012), Floods in the Alpine areas of Europe. *In:* Kundzewicz, Z. W. (Ed.) *Changes in Flood Risk in Europe*, 362–371, Wallingford, UK: IAHS Press.

Barredo, J. I. (2009), Normalised flood losses in Europe: 1970–2006, *Natural Hazards and Earth System Sciences*, 9, 97–104, doi:10.5194/nhess-9-97-2009.

Barredo, J. I., Sauri, D., and Llasat, M. C. (2012), Assessing trends in insured losses from floods in Spain 1971–2008, *Natural Hazards and Earth System Sciences*, 12, 1723–1729, doi:10.5194/nhess-12-1723-2012.

Barthel, F. and Neumayer, E. (2012), A trend analysis of normalized insured damage from natural disasters, *Climatic Change*, 113(2), 215–237, doi:10.1007/s10584-011-0331-2.

Bates, P. D., Stewart, M. D., Siggers, G. B., Smith, C. N., Hervouet, J. M., and Sellin, R. H. J. (1998), Internal and external validation of a two-dimensional finite element code for river flood simulations, *Proceedings of the Institution of Civil Engineers – Water, Maritime and Energy*, 130(3), 127–141, doi:10.1680/iwtme.1998.30972.

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

Bates, P. D., Marks, K. J., and Horritt, M. S. (2003), Optimal use of high-resolution topographic data in flood inundation models, *Hydrological Processes*, 17(3), 537–557, doi:10.1002/hyp.1113.

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

Bates, P. D., Wilson, M. D., Horritt, M. S., Mason, D. C., Holden, N., and 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. doi:10.1016/j.jhydrol.2005.12.028.

Bates, P. D., Horritt, M. S., and 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, doi:10.1016/j.jhydrol.2010.03.027.

Bates, P. D. (2012), Integrated remote sensing data with flood inundation models: how far have we got?, *Hydrological Processes*, 26(16), 2515–2521, doi:10.1002/hyp.9374.

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

Benedict, S. T. A., Caldwell, A. W., and Clark, J. M. (2013*), Flood-inundation maps for the Saluda River from Old Easley Bridge Road to Saluda Lake Dam near Greenville, South Carolina*. U.S. Geological Survey Scientific Investigations Map 3244. Retrieved from http://pubs.usgs.gov/sim/3244

Berghuijs, W. R., Woods, R. A., Hutton, C. J., and Sivapalan, M. (2016), Dominant flood generating mechanisms across the United States, *Geophysical Research Letters*, 43(9), 4382–4390, doi:10.1002/2016GL068070.

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., and Kirchner, J. W. (2019), The relative importance of different flood-generating mechanisms across Europe, *Water Resources Research*, 55(6), 4582–4593, doi:10.1029/2019WR024841.

Bermúdez, M., Neal, J. C., Bates, P. D., Coxon, G., Freer, J. E., Cea, L., and Puertas, J. (2017), Quantifying local rainfall dynamics and uncertain boundary conditions into a

nested regional-local flood modeling system, *Water Resources Research*, 53(4), 2770–2785, doi:10.1002/2016WR019903.

Bernhofen, M. V., Whyman, C., Trigg, M. A., Sleigh, P. A., Smith, A. M., Sampson, C. C., Yamazaki, D., Ward, P. J., Rudari, R., Pappenberger, F., Dottori, F., Salamon, P., and Winsemius, H. C. (2018), A first collective validation of global fluvial flood models for major floods in Nigeria and Mozambique, *Environmental Research Letters*, 13(10), 104007, doi:10.1088/1748-9326/aae014.

Beven, K. J. and Cloke, H. L. (2012), Comment on "Hyperresolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et al., *Water Resources Research*, 48(1), doi:10.1029/2011WR010982.

Beven, K. J., Cloke, H., Pappenberger, F., Lamb, R., and Hunter, N. (2015), Hyperresolution information and hyperresolution ignorance in modelling the hydrology of the land surface, *Science China Earth Sciences*, 58(1), 25–35, doi:10.1007/s11430-014-5003-4.

Bhola, P. K., Leandro, J., and Disse, M. (2018), Framework for Offline Flood Inundation Forecasts for Two-Dimensional Hydrodynamic Models, *Geosciences*, 8(9), 346, doi:10.3390/geosciences8090346.

Biancamaria, S., Lettenmaier, D. P., and 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. Springer, Cham.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F. (2015), Hyper-resolution global hydrological modelling: what is next?, *Hydrological Processes*, 29(2), 310–320, doi:10.1002/hyp.10391.

Birsan, M.-V., Molnar, P., Burlando, P., and Pfaundler, M. (2005), Streamflow trends in Switzerland, *Journal of Hydrology*, 314(1–4), 312–329, doi:10.1016/j.jhydrol.2005.06.008.

Birsan, M.-V., Zaharia, L., Chendes, V., and Branescu, E. (2014), Seasonal trends in Romanian streamflow, *Hydrological Processes*, 28(15), 4496–4505, doi:10.1002/hyp.9961.

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

Blessing, R., Sebastian, A., and Brody, S. D. (2017), Flood risk delineation in the United States: how much loss are we capturing?, *Natural Hazards Review*, 18(3), 04017002, doi:10.1061/(ASCE)NH.1527-6996.0000242.

Blöschl, G., Viglione, A., Merz, R., Parajka, J., Salinas, J., and Schöner, W. (2011), Climate impacts on floods and low flows, *Österreichische Wasser-und Abfallwirtschaft*, 63(1–2), 21–30, doi:10.1007/s00506-010-0269-z.

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

Bormann, H., Pinter, N., and Elfert, S. (2011), Hydrological signatures of flood trends on German rivers: flood frequencies, flood heights and specific stages, *Journal of Hydrology*, 404(1–2), 50–66, doi:10.1016/j.jhydrol.2011.04.019.

Bout, B. and Jetten, V. G. (2018), The validity of flow approximations when simulating catchment-integrated flash floods, *Journal of Hydrology*, 556, 674–688, doi:10.1016/j.jhydrol.2017.11.033.

Bouwer, L. M. (2010), *Disasters and climate change: Analysis and methods for projecting future losses from extreme weather*. PhD Thesis: Institute for Environmental Studies, Vrije Universiteit Amsterdam.

Bouwer, L. M., Bubeck, P., and Aerts, J. C. J. H. (2010), Changes in a future flood risk due to climate and development in a Dutch polder area, *Global Environmental Change*, 20(3), 463–471, doi:10.1016/j.gloenvcha.2010.04.002.

Bouwer, L. M. (2011), Have disaster losses increased due to anthropogenic climate change?, *Bulletin of the American Meteorological Society*, 92(1), 39–46, doi:10.1175/2010BAMS3092.1.

Bozkurt, D., Rojas, M., Boisier, J. P., and Valdivieso, J. (2018), Projected hydroclimate changes over Andean basins in central Chile from downscaled CMIP5 models under the low and high emission scenarios, *Climatic Change*, 150(3–4), 131–147, doi:10.1007/s10584-018-2246-7.

Bright, E. A., Rose, A. N., and Urban, M. L. (2016), *LandScan 2015 High-Resolution Global Population Data Set*. Oak Ridge, TN, USA: Oak Ridge National Laboratory.

Buizza, R., Miller, M., and Palmer, T. N. (1999), Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, *Quarterly Journal of the Royal Meteorological Society*, 125(560), 2887–2908, doi:10.1002/qj.49712556006.

Buizza, R. (2005), A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems, *Monthly Weather Review*, 133(5), 1076–1097. doi:10.1175/MWR2905.1.

Burn, D. H. and Whitfield, P. H. (2016), Changes in floods and flood regimes in Canada, *Canadian Water Resources Journal*, 41(1–2), 139–150, doi:10.1080/07011784.2015.1026844.

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

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

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

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

Castellarin, A., Domeneghetti, A., and Brath, A. (2011b), 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, doi:10.1016/j.pce.2011.02.008.

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

Ceola, S., Laio, F., and Montanari, A. (2014), Satellite nighttime lights reveal increasing human exposure to floods worldwide, *Geophysical Research Letters*, 41(20), 7184–7190, doi:10.1002/2014GL061859.

Chan, S. C., Kendon, E. J., Fowler, H. J., Blenkinsop, S., Ferro, C. A. T., and Stephenson, D. B. (2013), Does increasing the spatial resolution of a regional climate model improve the simulated daily precipitation?, *Climate Dynamics*, 41(5–6), 1475–1495, doi:10.1007/s00382-012-1568-9.

Chang, H., Franczyk, J., and Kim, C. (2009), What is responsible for increasing flood risks? The case of Gangwon Province, Korea, *Natural Hazards*, 48, 339, doi:10.1007/s11069-008-9266-y.

Changnon, S. A. (2008), Assessment of flood losses in the United States, *Journal of Contemporary Water Research and Education*, 138(1), 38–44, doi: 10.1111/j.1936-704X.2008.00007.x.

Changnon, S. A., Pielke, R. J., Changnon, D., Sylves, R. T., and Pulwarty, R. (2000), Human factors explain the increased losses from weather and climate extremes, *Bulletin of the American Meteorological Society*, 81, 437–442, doi: 10.1175/1520-0477(2000)081<0437:HFETIL>2.3.CO;2.

Chen, J., Brissette, F. P., and Leconte, R. (2011), Uncertainty of downscaling method in quantifying the impact of climate change on hydrology, *Journal of Hydrology*, 401(3–4), 190–202, doi:10.1016/j.jhydrol.2011.02.020.

Chen, S., Liu, H., You, Y., Mullens, E., Hu, J., Yuan, Y., Huang, M., He, L., Luo, Y., Zeng, X., Tang, G., and Hong, Y. (2014), Evaluation of high-resolution precipitation estimates from satellites during July 2012 Beijing flood event using dense rain gauge observations, *PLoS ONE*, 9(4), e89681, doi:10.1371/journal.pone.0089681.

Chen, B., Krajewski, W. F., Goska, R., and Young, N. (2017), Using LiDAR surveys to document floods: a case study of the 2008 Iowa flood, *Journal of Hydrology*, 553, 338–349, doi: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, doi:10.3390/rs6098696.

Chow, V. T. (1959), *Open Channel Hydraulics*. New York, US: McGraw Hill.

Ciullo, A., de Bruijn, K. M., Kwakkel, J. H., and 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*, doi:10.1111/jfr3.12532.

Cloke, H. L. and Pappenberger, F. (2009), Ensemble flood forecasting: A review, *Journal of Hydrology*, 375(3-4), 613–626, doi:10.1016/j.jhydrol.2009.06.005.

Cohen, W. B., Healy, S. P., Yang, Z., Stehman, S. V., Brewer, C. K., Brooks, E. B., Gorelick, N., Huang, C., Hughes, M. J., Kennedy, R. E., Loveland, T. R., Moisen, G. G., Schroeder, T. A., Vogelmann, J. E., Woodcock, C. E., Yang, L., and Zhu, Z. (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. Geological Survey Scientific Investigations Map 3241. Retrieved from http://pubs.usgs.gov/sim/3241

Couasnon, A., Eilander, D., Muis, S., Veldkamp, T. I. E., Haigh, I. D., Wahl, T., Winsemius, H., and Ward, P. J. (2019), Measuring compound flood potential from river discharge and storm surge extremes at the global scale and its implications for flood hazard, *Natural Hazards and Earth System Sciences Discussions*, doi.org/10.5194/nhess-2019-205.

Courant, R., Friedrichs, K., and Lewy, H. (1928), Über die partiellen Differenzengleichungen der mathematischen Physik, *Mathematische Annalen*, 100(1), 32–74, doi:10.1007/BF01448839.

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

Crompton, R. P. and McAneney, K. J. (2008), Normalized Australian insured losses from meteorological hazards: 1967–2006, *Environmental Science and Policy*, 11(5), 371–378, doi:10.1016/j.envsci.2008.01.005.

Crowell, M., Coulton, K., Johnson, C., Westcott, J., Bellomo, D., Edelman, S., and Hirsch, E. (2010), An estimate of the US population living in 100-year coastal flood hazard areas, *Journal of Coastal Research*, 36(2), 201–211, doi:10.2112/JCOASTRES-D-09-00076.1.

Cunderlik, J. M. and Ouarda, T. M. M. J. (2009), Trends in the timing and magnitude of floods in Canada, *Journal of Hydrology*, 375(3–4), 471–480, doi:10.1016/j.jhydrol.2009.06.050.

Cunge, J. A., Holly, F. M., and Verwey, A. (1980), *Practical aspects of computational river hydraulics.* London, UK: Pitman Publishing Ltd.

Czuba, C. R., Fallon, J. D., Lewis, C. R., and Cooper, D. F. (2014), *Development of flood-inundation maps for the Mississippi River in Saint Paul, Minnesota.* U.S. Geological Survey Scientific Investigations Report 2014–5079, doi:10.3133/sir20145079.

Dankers, R. and Feyen, L. (2009), Flood hazard in Europe in an ensemble of regional climate scenarios, *Journal of Geophysical Research Atmospheres*, 114(D16), doi:10.1029/2008JD011523.

de Almeida, G. A. M., Bates, P., Freer, J. E., and Souvignet, M. (2012), Improving the stability of a simple formulation of the shallow water equations for 2-D flood modeling, *Water Resources Research*, 48(5), doi:10.1029/2011WR011570.

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

de Moel, H., van Alphen, J., and Aerts, J. C. J. H. (2009), Flood maps in Europe – methods, availability and use, *Natural Hazards and Earth System Sciences*, 9, 289–301, doi:10.5194/nhess-9-289-2009.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137(656A), 553–597, doi:10.1002/qj.828.

Delgado, J. M., Apel, H., and Merz, B. (2010), Flood trends and variability in the Mekong river, *Hydrology and Earth System Sciences*, 14, 407–418, doi:10.5194/hess-14-407-2010.

Di Baldassarre, G. and Montanari, A. (2009), Uncertainty in river discharge observations: a quantitative analysis, *Hydrology and Earth System Sciences*, 13, 913–921, doi:10.5194/hess-13-913-2009.

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

Di Baldassarre, G., Castellarin, A., and 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, doi:10.1623/hysj.54.6.1007.

Diakakis, M. (2014), An inventory of flood events in Athens, Greence, during the last 130 years. Seasonality and spatial distribution, *Journal of Flood Risk Management*, 7(3), 332–343, doi:10.1111/jfr3.12053.

Do, H. X., Westra, S., and Leonard, M. (2017), A global-scale investigation of trends in annual maximum streamflow, *Journal of Hydrology*, 552, 28–43, doi:10.1016/j.jhydrol.2017.06.015.

Döll, P., Trautmann, T., Gerten, D., Schmied, H. M., Ostberg, S., Saaed, F., and Schleussner, C.-F. (2018), Risks for the global freshwater system at 1.5°C and 2°C global warming, *Environmental Research Letters*, 13(4), 044038, doi:10.1088/1748-9326/aab792.

Domeneghetti, A., Castellarin, A., and Brath, A. (2012), Assessing rating-curve uncertainty and its effects on hydraulic model calibration, *Hydrology and Earth System Sciences*, 16, 1191–1202, doi:10.5194/hess-16-1191-2012.

Domeneghetti, A., Carisi, F., Castellarin A., and Brath, A. (2015), Evolution of flood risk over large areas: quantitative assessment for the Po river, *Journal of Hydrology*, 527, 809–823, doi:10.1016/j.jhydrol.2015.05.043.

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

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

Dottori, F. and Todini, E. (2011), Developments of a flood inundation model based on the cellular automata approach: testing different methods to improve model performance, *Physics and Chemistry of the Earth, Parts A/B/C*, 36(7–8), 266–280, doi:10.1016/j.pce.2011.02.004.

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

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, doi:10.1016/j.advwatres.2016.05.002.

Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A., Mongelli, I., Frieler, K., Betts, R. A., and Feyen, L. (2018), Increased human and economic losses from river flooding with anthropogenic warming, *Nature Climate Change*, 8, 781–786, doi:10.1038/s41558-018-0257-z.

Downton, M. W. and Pielke, R. A. (2005), How accurate are disaster loss data? The case of U.S. flood damage, *Natural Hazards*, 35(2), 211–228, doi:10.1007/s11069-004-4808-4.

Downton, M. W., Barnard Miller, J., and Pielke, R. A. (2005), Reanalysis of US National Weather Service flood loss database, *Natural Hazards Review*, 6(1), 13–22, doi:10.1061/(ASCE)1527-6988(2005)6:1(13).

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

Dumas, P., Hallegatte, S., Quintana-Seguì, P., and Martin, E. (2013), The influence of climate change on flood risks in France – first estimates and uncertainty analysis, *Natural Hazards and Earth System Sciences*, 13, 809–821, doi:10.5194/nhess-13-809-2013.

Dunne, T. and Black, R. D. (1970), An experimental investigation of runoff production in permeable soils, *Water Resources Research*, 6(2), 478–490, doi:10.1029/WR006i002p00478.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J. (2012), HESS Opinions "Should we apply bias correction to global and regional climate model data?", *Hydrology and Earth System Sciences*, 16, 3391–3404, doi:10.5194/hess-16-3391-2012.

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

Emanuel, K. (2017), Assessing the present and future probability of Hurricane Harvey's rainfall, *Proceedings of the National Academy of Sciences of the United States of America*, 114(48), 12681–12684, doi:10.1073/pnas.1716222114.

Environment Agency (2013), *What is the updated Flood Map for Surface Water?* Retrieved from http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/297432/LIT_8988_0bf634.pdf

Estrela, T. and Quintas, L. (1994), Use of GIS in the modelling of flows on floodplains. *In:* White, W. R. and Watts, J. (Eds.) *Proceedings of the 2nd International Conference on River Flood Hydraulics*, 177–189, Chichester, UK: Wiley.

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

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

Ewing, R. and Cervero, R. (2010), Travel and the built environment: a meta-analysis, *Journal of the American Planning Association*, 76(3), 265–294, doi:10.1080/01944361003766766.

Federal Emergency Management Agency (1995), *Managing Floodplain Development in Approximate Zone A Areas*. Retrieved from 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*. Retrieved from 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.* Retrieved from https://www.fema.gov/media-librarydata/1458925981276b3f8ab92d6abf14365ad154954e4dc69/changepackage_508_apr1 6b.pdf

Federal Emergency Management Agency (2016b), *FEMA Policy Standards for Flood risk Analysis and Mapping*. Retrieved from https://www.fema.gov/media-library-data/1480449548025-4736d89b89d30fbf102228680c1f8acd/Standards_for_Flood_Risk_Projects_(Nov2016).pdf

Fekete, B. M. and Vörösmarty, C. J. (2007), The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *In: Predictions in Ungauged Basins: PUB Kick-off*, 129–136. IAHS Publication 309..

Fengqing, J., Cheng, Z., Guijin, M., Ruji, H., and Qingxia, M. (2005), Magnification of flood disasters and its relation to regional precipitation and local human activities since the

1980s in Xinjiang, northwestern China, *Natural Hazards*, 36(3), 307–330, doi:10.1007/s11069-005-0977-z.

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

Fewtrell, T. J., Duncan, A., Sampson, C. C., Neal, J. C., and Bates, P. D. (2011), Benchmarking urban flood models of varying complexity and scale using high resolution terrestrial LiDAR data, *Physics and Chemistry of the Earth, Parts A/B/C*, 36(7–8), 281–291, doi:10.1016/j.pce.2010.12.011.

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

Feyen, L., Barredo, J. I., and Dankers, R. (2009), Implications of global warming and urban landuse change on flooding in Europe. *In:* Feyen, L., Shannon, K., and Neville, M. (Eds.) *Water and Urban Development Paradigms*, 217–225, Taylor and Francis.

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

Fischer, E. M., Lawrence, D. M., and Sanderson, B. M. (2011), Quantifying uncertainties in projections of extremes – a perturbed land surface parameter experiment, *Climate Dynamics*, 37(7–8), 1381–1398, doi:10.1007/s00382-010-0915-y.

Fischer, G., Nachtergaele, F. O., Prieler, S., Teixeira, E., Toth, G., van Velthuizen, H., Verelst, L., and Wiberg, D. (2012), *Global Agro-ecological Zones (GAEZ v3.0) – Model Documentation.* Laxenburg, Austria: IIASA.

Fleischmann, A., Paiva, R., and Collischonn, W. (2019), Can regional to continental river hydrodynamic models be locally relevant? A cross-scale comparison, *Journal of Hydrology X*, 3, 100027, doi: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., Lion, C., Beighley, R. E., Yang, X. (2019), Global relationships between river width, slope, catchment area, meander wavelength, sinuosity, and

discharge, *Geophysical Research Letters*, 46(6), 3252–3262, doi:10.1029/2019GL082027.

Freni, G., La Loggia, G., and Notaro, V. (2010), Uncertainty in urban flood damage assessment due to urban drainage modelling and depth-damage curve estimation, *Water Science and Technology*, 61(12), 2979–2993, doi:10.2166/wst.2010.177.

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

Gesch, D. B., Oimoen, M. J., and Evans, G. A. (2014), Accuracy Assessment of the US Geological Survey National Elevation Dataset, and Comparison with Other Large-Area Elevation Datasets—SRTM and ASTER, *US Geological Survey Open-File Report 2014-1008*, doi:10.3133/ofr20141008.

Gilles, D., Young, N., Schroeder, H., Piotrowski, J., and Chang, Y.-J. (2012), Inundation mapping initiatives of the Iowa Flood Center: statewide coverage and detailed urban flooding analysis, *Water*, 4(1), 85–106, doi:10.3390/w4010085.

Giuntoli, I., Renard, B., and Lang, M. (2012), Floods in France. *In:* Kundzewicz, Z. W. (Ed.) *Changes in Flood Risk in Europe*, 199-211, Wallingford, UK: IAHS Press.

Gleason, C. J. and Smith, L. C. (2014), Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry, *Proceedings of the National Academy of Sciences of the United States of America*, 111(13), 4788–4791, doi:10.1073/pnas.1317606111.

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

Google (2016), *Google Earth Engine API Developer's guide*. Retrieved from https://developers.google.com/earth-engine/

Gouldby, B., Sayers, P., Mulet-Marti, J., Hassan, M. A. A. M., and Benwell, D. (2008), A methodology for regional-scale flood risk assessment, *Proceedings of the Institute of Civil Engineers – Water Management*, 161(3), 169–182, doi:10.1680/wama.2008.161.3.169.

Gruell, W., Andersson, J. C. M., Donnelly, C., Feyen, L., Gerten, D., Ludwig, F., Pisacane, G., Roudier, P., and Schaphoff, S. (2015), Evaluation of five hydrological models across Europe and their suitability for making projections under climate change, *Hydrology and Earth System Sciences Discussions*, 12, 10289–10330, doi:10.5194/hessd-12-10289-2015, 2015.

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., and Seneviratne, S. I. (2019), Observed trends in global indicators of mean and extreme streamflow, *Geophysical Research Letters*, 46(2), 756–766, doi:10.1029/2018GL079725.

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

Guinot, V., Sanders, B. F., and Schubert, J. E. (2017), Dual integral porosity shallow water model for urban flood modelling, *Advances in Water Resources*, 103, 16–31, doi:10.1016/j.advwatres.2017.02.009.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voss, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabar, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P. (2011), Multimodel estimate of the global terrestrial water balance: setup and first results, *Journal of Hydrometeorology*, 12(5), 869–884, doi:10.1175/2011JHM1324.1.

Hagemann, S., Chen, C., and Haerter, J. O. (2011), Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models, *Journal of Hydrometeorology*, 12(4), 556–578, doi:10.1175/2011JHM1336.1.

Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., Kjeldsen, T. R., Kriaučiūnienė, J., Kundzewicz, Z. W., Lang, M., Llasat, M. C., Macdonald, N., McIntyre, N., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Neuhold, C., Parajka, J., Perdigão, R. A. P., Plavcová L., Rogger, M., Salinas, J. L., Sauquet, E., Schär, C., Szolgay, J., Viglione, A., and Blöschl, G. (2014), Understanding flood regime changes in Europe: a state-of-the-art assessment, *Hydrology and Earth System Sciences*, 18, 2735–2772, doi:10.5194/hess-18-2735-2014.

Hamilton, A. S., and Moore, R. D. (2012), Quantifying uncertainty in streamflow records, *Canadian Water Resources Journal*, 37(1), 3–21, doi:10.4296/cwrj3701865.

Hamlet, A. F., Elsner, M. M., Mauger, G. S., Lee, S.-Y., Tohver, I., and Norheim, R. A. (2013), An overview of the Columbia basin climate change scenarios project: approach, methods, and summary of key results, *Atmosphere-Ocean*, 51, 392–415, doi:10.1080/07055900.2013.819555.

Handmer, J., Honda, Y., Kundzewicz, Z. W., Arnell, N., Benito, G., Hatfield, J., Mohamed, I. F., Peduzzi, P., Wu, S., Sherstyukov, B., Takahashi, K., and Yan, Z. (2012), *In:* Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M., and Midgley, P. M. (Eds.) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press.

Hannaford, J. and Marsh, T. J. (2008), High-flow and flood trends in a network of undisturbed catchments in the UK, *International Journal of Climatology*, 28, 1325–1338, doi:10.1002/joc.1643.

Hannaford, J. and Buys, G. (2012), Trends in seasonal river flow regimes in the UK, *Journal of Hydrology*, 475, 158–174, doi:10.1016/j.jhydrol.2012.09.044.

Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M. (2011), Large-scale river flow archives: importance, current status and future needs, *Hydrological Processes*, 25(7), 1191–1200, doi:10.1002/hyp.7794.

Hattermann, F. F., Huang, S., Burghoff, O., Willems, W., Österle, H., Büchner, M., and Kundzewicz, Z. (2014), Modelling flood damages under climate change conditions – a case study for Germany, *Natural Hazards and Earth System Sciences*, 14, 3151–3168, doi: 10.5194/nhess-14-3151-2014.

Hattermann, F. F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., Fekete, B., Flörke, F., Gosling, S. N., and Hoffmann, P. (2018), Sources of uncertainty in hydrological climate impact assessment: a cross-scale study, *Environmental Research Letters*, 13(1), 015006, doi:10.1088/1748-9326/aa9938.

Hawker, L., Rougier, J., Neal, J., Bates, P., Archer, L., and Yamazaki, D. (2018), Implications of simulating global digital elevation models for flood inundation studies, *Water Resources Research*, 54(10), 7910–7928, doi:10.1029/2018WR023279.

Heine, R. A., and Pinter, N. (2012), Levee effects upon flood levels: an empirical assessment, *Hydrological Processes*, 26, 3225–3240, doi:10.1002/hyp.8261.

Hirabayashi, Y., Kanae, S., Emori, S., Oki, T., and Kimoto, M. (2008), Global projections of changing risks of floods and droughts in a changing climate, *Hydrological Sciences Journal*, 53(4), 754–772, doi:10.1623/hysj.53.4.754.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S. (2013), Global flood risk under climate change, *Nature Climate Change*, 3, 816–821, doi:10.1038/nclimate1911.

Hirsch, R. M. and Ryberg, K. R. (2012), Has the magnitude of floods across the USA changed with global $CO_2$ levels?, *Hydrological Sciences Journal*, 57(1), 1–9, doi:10.1080/02626667.2011.621895.

Hirsch, R. M. and Archfield, S. A. (2015), Flood trends: not higher but more often, *Nature Climate Change*, 5, 198–199, doi:10.1038/nclimate2551.

Hoard, C. J., Fowler, K. K., Kim, M. H., Menke, C. D., Morlock, S. E., Peppler, M. C., Rachol, C. M., and Whitehead, M. T. (2010), *Flood-inundation maps for a 15-mile reach of the Kalamazoo River from Marshall to Battle Creek, Michigan, 2010*. U.S. Geological Survey Scientific Investigations Map 3135, doi:10.3133/sim3135.

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

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, doi:10.1002/rra.781.

Horna-Muñoz, D. and 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, doi:10.1016/j.advwatres.2018.10.014.

Horritt, M. S. and Bates, P. D. (2001), Predicting floodplain inundation: raster-based modelling versus the finite-element approach, *Hydrological Processes*, 15(5), 825–842, doi:10.1002/hyp.188.

Horritt, M. S., Mason, D. C., and Luckman, A. J. (2001), Flood boundary delineation from synthetic aperture radar imagery using a statistical active contour model, *International Journal of Remote Sensing*, 22(13), 2489–2507, doi:10.1080/01431160116902.

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

Horritt, M. S., Bates, P. D., Fewtrell, T. J., Mason, D. C., and Wilson, M. D. (2010), Modelling the hydraulics of the Carlisle 2005 flood event, *Proceedings of the Institution of Civil Engineers*, 163(WM6), 273–281, doi:10.1680/wama.2010.163.6.273.

Horton, R. E. (1933), The role of infiltration in the hydrologic cycle, *Eos*, 14(1), 446–460, doi:10.1029/TR014i001p00446.

Huang, S., Hattermann, F. F., Krysanova, V., and Bronstert, A. (2013), Projections of climate change impacts on river flood conditions in Germany by combining three different RCMs with a regional eco-hydrological model, *Climatic Change*, 116(3–4), 631–663, doi:10.1007/s10584-012-0586-2.

Hundecha, Y., Sunyer, M. A., Lawrence, D., Madsen, H., Willems, P., Bürger, G., Kriaučiūnienė, J., Loukas, A., Martinkova, M., Osuch, M., Vasiliades, L., von Christierson, B., Vormoor, K., Yücel, I. (2016), Inter-comparison of statistical downscaling methods for projection of extreme flow indices across Europe, *Journal of Hydrology*, 541(B), 1273–1286, doi:10.1016/j.jhydrol.2016.08.033.

Hunter, N. M., Horritt, M. S., Bates, P. D., Wilson, M. D., and Werner, M. G. F. (2005), An adaptive time step solution for raster-based storage cell modelling of floodplain inundation, *Advances in Water Resources*, 28(9), 975–991, doi:10.1016/j.advwatres.2005.03.007.

Hunter, N. M., Bates, P. D., Horritt, M. S., and Wilson, M. D. (2006), Improved simulation of flood flows using storage cell models, *Proceedings of the Institution of Civil Engineers – Water Management*, 159(1), 9–18, doi:10.1680/wama.2006.159.1.9.

Hunter, N. M., Bates, P. D., Horritt, M. S., and Wilson, M. D. (2007), Simple spatially-distributed models for predicting flood inundation: a review, *Geomorphology*, 90(3–4), 208–225, doi:10.1016/j.geomorph.2006.10.021.

Hunter, N. M., Bates, P. D., Neelz, S., Pender, G., Villanueva, I., Wright, N. G., Liang, D., Falconer, R. A., Lin, B., Waller, S., Crossley, A. J., and Mason, D. C. (2008), Benchmarking 2D hydraulic models for urban flooding, *Proceedings of the Institute of Civil Engineers – Water Management*, 161(1), 13–30, doi:10.1680/wama.2008.161.1.13.

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

IPCC (2012), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M., and Midgley, P. M. (Eds.). Cambridge, UK: Cambridge University Press .

Ishak, E. H., Rahman, A., Westra, S., Sharma, A., and Kuczera, G. (2013), Evaluating the non-stationarity of Australian annual maximum flood, *Journal of Hydrology*, 494, 134–145, doi:10.1016/j.jhydrol.2013.04.021.

Ivancic, T. J. and Shaw, S. B. (2015), Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge, *Climatic Change*, 133(4), 681–693, doi:10.1007/s10584-015-1476-1.

Jafarzadegan, K., Merwade, V., and Saksena, S. (2018), A geomorphic approach to 100-year floodplain mapping for the conterminous United States, *Journal of Hydrology*, 561, 43–58, doi:10.1016/j.jhydrol.2018.03.061.

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

Jiang, T., Kundzewicz, Z. W., and Su, B. (2008), Changes in monthly precipitation and flood hazard in the Yangtze river basin, China, *International Journal of Climatology*, 28, 1471–1481, doi:10.1002/joc.1635.

Jongman, B., Ward, P. J., and Aerts, J. C. J. H. (2012a), Global exposure to river and coastal flooding: long term trends and changes, *Global Environmental Change*, 22(4), 823–835, doi:10.1016/j.gloenvcha.2012.07.004.

Jongman, B., Kreibich, H., Apel, H., Barredo, J. I., Bates, P. D., Feyen, L., Gericke, A., Neal, J., Aerts, J. C. J. H., and Ward, P. J. (2012b), Comparative flood damage model assessment: towards a European approach, *Natural Hazards and Earth System Sciences*, 12, 3733–3752, doi:10.5194/nhess-12-3733-2012.

Jongman, B., Winsemius, H. C., Aerts, J. C. J. H., Coughlan de Perez, E., van Aalst, M. K., Kron, W., and Ward, P. J. (2015), Declining vulnerability to river floods and the global benefits of adaptation, *Proceedings of the National Academy of Sciences of the United States of America*, 112(18), E2271–E2280, doi:10.1073/pnas.1414439112.

Kang, N.-Y. and Elsner, J. B. (2016), Climate Mechanism for Stronger Typhoons in a Warmer World, *Journal of Climate*, 29, 1051–1057, doi:10.1175/JCLI-D-15-0585.1.

Kay, A. L. and Jones, D. A. (2012), Transient changes in flood frequency and timing in Britain under potential projections of climate change, *International Journal of Climatology*, 32(4), 489–502, doi:10.1002/joc.2288.

Kendon, E. J., Roberts, N. M., Senior, C. A., and Roberts, M. J. (2012), Realism of rainfall in a very high-resolution regional climate model, *Journal of Climate*, 25, 5791–5806, doi:10.1175/JCLI-D-11-00562.1.

Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., Evans, J. P., Fosser, G., and Wilkinson, J. M. (2017), Do convection-permitting regional climate models improve projections of future precipitation change?, *Bulletin of the American Meteorological Society*, 98, 79–93, doi:10.1175/BAMS-D-15-0004.1.

Kesserwani, G. and Wang, Y. (2014), Discontinuous Galerkin flood model formulation: luxury or necessity?, *Water Resources Research*, 50(8), 6522–6541, doi:10.1002/2013WR014906.

Kidd, C., Bauer, P., Turk, J., Huffman, G. J., Joyce, R., Hsu, K.-L., and Braithwaite, D. (2012), Intercomparison of high-resolution precipitation products over northwest Europe, *Journal of Hydrometeorology*, 13, 67–83, doi:10.1175/JHM-D-11-042.1.

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

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

Kinoshita, Y., Tanoue, M., Watanabe, S., and Hirabayashi, Y. (2018), Quantifying the effect of autonomous adaptation to global river flood projections: application to future flood risk assessments, *Environmental Research Letters*, 13(1), 014006, doi:10.1088/1748-9326/aa9401.

Knight, D. W. and Shiono, K. (1996), River channel and floodplain hydraulics. *In:* Anderson, M. G., Walling, D. E., and Bates, P. D. (Eds.) *Floodplain Processes*, 139–182. Chichester, UK: Wiley.

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

Koks, E. E., Rozenberg, J., Zorn, C., Tariverdi, M., Vousdoukas, M., Fraser, S. A., Hall, J. W., and Hallegatte, S. (2019), A global multi-hazard risk analysis of road and railway infrastructure assets, *Nature Communications*, 10, doi:10.1038/s41467-019-10442-3.

Köplin, N., Schädler, B., Viviroli, D., and Weingartner, R. (2014), Seasonality and magnitude of floods in Switzerland under future climate change, *Hydrological Processes*, 28(4), 2567–2578, doi:10.1002/hyp.9757.

Korhonen, J. and Kuusisto, E. (2010), Long-term changes in the discharge regime in Finland, *Hydrology Research*, 41(3–4), 253–268, doi:10.2166/nh.2010.112.

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

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

Krajewski, W. F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., Mantilla, R., Niemeier, J., Quintero, F., Seo, B.-C., Small, S., Weber, L., and Young, N. (2017), Real-time flood forecasting and information system for the State of Iowa, *Bulletin of the American Meteorological Society*, 98(3), 539–554, doi:10.1175/BAMS-D-15-00243.1.

Kreibich, H., Piroth, K., Seifert, I., Maiwald, H., Kunert, U., Schwarz, J., Merz, B., and Thieken, A. H. (2009), Is flow velocity a significant parameter in flood damage modelling?, *Natural Hazards and Earth System Sciences*, 9, 1679–1692, doi:10.5194/nhess-9-1679-2009.

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*. Berlin, Offenbach: Wichmann, S.130-139.

Krüger, T., and 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*. Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg, doi:10.1007/978-3-540-72135-2_6.

Krysanova, V., Vetter, T., Eisner, S., Huang, S., Pechlivanidis, I., Strauch, M., Gelfan, A., Kumar, R., Aich, V., Arheimer, B., Chamorro, A., van Griensven, A., Kundu, D., Lobanova, A., Mishra, V., Plötner, S., Reinhardt, J., Seidou, O., Wang, X., Wortmann, M., Zeng, X., and Hattermann, F. F. (2017), Intercomparison of regional-scale hydrological models and climate change impacts projected for 12 large river basins worldwide – a synthesis, *Environmental Research Letters*, 12(10), 105002, doi:10.1088/1748-9326/aa8359.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L. M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, G. R., Kron, W., Benito, G., Honda, Y., Takahashi, K., and Shersyukov, B. (2014), Flood risk and climate change: global and regional perspectives, *Hydrological Sciences Journal*, 59(1), 1–28, doi:10.1080/02626667.2013.857411.

Kundzewicz, Z. W., Krysanova, V., Dankers, R., Hirabayashi, Y., Kanae, S., Hattermann, F. F., Huang, S., Milly, P. C. D., Stoffel, M., Driessen, P. P. J., Matczak, P., Quevauviller, P., and Schellnhuber, H.-J. (2017), Differences in flood hazard projections in Europe – their causes and consequences for decision making, *Hydrological Sciences Journal*, 62(1), 1–14, doi:10.1080/02626667.2016.1241398.

Kundzewicz, Z. W., Pińskwar, I, and Brakenridge, G. R. (2018), Changes in river flood hazard in Europe: a review, *Hydrology Research*, 49(2), 294–302, doi:doi.org/10.2166/nh.2017.016.

Lamb, R., Crossley, M., and Waller, S. (2009), A fast two-dimensional floodplain inundation model, *Proceedings of the Institution of Civil Engineers – Water Management*, 162(6), 363–370, doi:10.1680/wama.2009.162.6.363.

Langerwisch, F., Rost, S., Gerten, D., Poulter, B., Rammig, A., and Cramer, W. (2013), Potential effects of climate change on inundation patterns in the Amazon basin, *Hydrology and Earth System Sciences*, 17, 2247–2262, doi:10.5194/hess-17-2247-2013.

Lashermes, B., Foufoula-Georgiou, E., and Dietrich, W. E. (2007), Channel network extraction from high resolution topography using wavelets, *Geophysical Research Letters*, 34(23), L23S04, doi:10.1029/2007GL031140.

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

Lehner, B., Döll, P., Alcamo, J., Henrichs, T., and Kaspar F. (2006), Estimating the impact of global change on flood and drought risks in Europe: a continental, integrated analysis, *Climatic Change*, 75(3), 273–299, doi:10.1007/s10584-006-6338-4.

Lehner, B., Verdin, K., and Jarvis, A. (2008), New global hydrography derived from spaceborne elevation data, *Eos*, 89(10), 93–104, doi:10.1029/2008EO100001.

Lehner, B., and 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, doi:10.1002/hyp.9740.

Leopold, L. B. and Maddock, T. (1953), *The hydraulic geometry of stream channels and some physiographic implications*. US Government Printing Office.

Leskens, J. G., Brugnach, M., Hoekstra, A. Y., and Schuurmans, W. (2014), Why are decisions in flood disaster management so poorly supported by information from flood models?, *Environmental Modelling and Software*, 53, 53–61, doi:10.1016/j.envsoft.2013.11.003.

Lindström, G. and Alexandersson, H. (2004), Recent mild and wet years in relation to long observation records and future climate change in Sweden, *AMBIO*, 33(4), 183–186, doi:10.1579/0044-7447-33.4.183.

Lins, H. F. and Slack, J. R. (2005), Seasonal and regional characteristics of U.S. streamflow trends in the United States from 1940 to 1999, *Physical Geography*, 26(6), 489–501, doi:10.2747/0272-3646.26.6.489.

Liu, L., Xu, H., Wang, Y., and Jiang, T. (2017), Impacts of 1.5 and 2°C global warming on water availability and extreme hydrological events in Yiluo and Bejiang river catchments in China, *Climatic Change*, 145(1–2), 145–158, doi:10.1007/s10584-017-2072-3.

Liu, Y. Y., Maidment, D. R., Tarboton, D. G., Zheng, X., and Wang, S. (2018), A CyberGIS Integration and Computation Framework for High-Resolution Continental-Scale Flood Inundation Mapping, *Journal of the American Water Resources Association*, 54(4), 770–784, doi:10.1111/1752-1688.12660.

Llasat, M. C., Llasat-Botija, M., Rodriguez, A., and Lindbergh, S. (2010), Flash floods in Catalonia: a recurrent situation, *Advances in Geosciences*, 26, 105–111, doi:10.5194/adgeo-26-105-2010, 2010.

Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., and Tarpley, J. D. (2004), Streamflow and water balance intercomparisons for four land surface models in the North American Land Data Assimilation System project, *Journal of Geophysical Research Atmospheres*, 109(D7), doi:10.1029/2003JD003517.

López-Moreno, J. I., Beguería, S., and García-Ruiz, J. M. (2006), Trends in high flows in the central Spanish Pyrenees: response to climatic factors or to land-use change?, *Hydrological Sciences Journal*, 51(6), 1039–1050, doi:10.1623/hysj.51.6.1039.

Maidment, D. R. (2016), Conceptual Framework for the National Flood Interoperability Experiment, *Journal of the American Water Resources Association*, 53(2), 245–257, doi:10.1111/1752-1688.12474.

Mallakpour, I. and Villarini, G. (2015), The changing nature of flooding across the central United States, *Nature Climate Change*, 5, 250–254, doi:10.1038/nclimate2516.

Mangini, W., Viglione, A., Hall, J., Hundecha, Y., Ceola, S., Montanari, A., Rogger, M., Salinas, J. S., Borzì, I., and Parajka, J. (2018), Detection of trends in magnitude and frequency of flood peaks across Europe, *Hydrological Sciences Journal*, 63(4), 493–512, doi:10.1080/02626667.2018.1444766.

Marengo, J. A. (1995), Variations and change in South American streamflow, *Climatic Change*, 31(1), 99–117, doi:10.1007/BF01092983.

Marengo, J. A. (2009), Long-term trends and cycles in the hydrometeorology of the Amazon basin since the late 1920s, *Hydrological Processes*, 23(22), 3236–3244, doi:10.1002/hyp.7396.

Marks, K. and Bates, P. (2000), Integration of high-resolution topographic data with floodplain flow models, *Hydrological Processes*, 14, 2109–2122, doi:10.1002/1099-1085(20000815/30)14:11/12<2109::AID-HYP58>3.0.CO;2-1.

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

Maskell, J., Horsburgh, K., Lewis, M., and Bates, P. (2014), Investigating river – surge interaction in idealised estuaries, *Journal of Coastal Research*, 30(2), 248–259, doi:10.2112/JCOASTRES-D-12-00221.1.

Masoero, A., Claps, P., Asselman, N. E. M., Mosselman, E., and Di Baldassarre, G. (2013), Reconstruction and analysis of the Po River inundation of 1951, *Hydrological Processes*, 27, 1341–1348, doi:10.1002/hyp.9558.

Mason, D. C., Cobby, D. M., Horritt, M. S., and Bates, P. D. (2003), Floodplain friction parameterization in two-dimensional river flood models using vegetation heights derived from airborne scanning laser altimetry, *Hydrological Processes*, 17(9), 1711–1732, doi:10.1002/hyp.1270.

Mason, D. C., Horritt, M. S., Hunter, N. M., and Bates, P. D. (2007), Use of fused airborne scanning laser altimetry and digital map data for urban flood modelling, *Hydrological Processes*, 21(11), 1436–1447, doi:10.1002/hyp.6343.

Mateo, C. M. R., Yamazaki, D., Kim, H., Champathong, A., Vaze, J., and Oki, T. (2017), Impacts of spatial resolution and representation of flow connectivity on large-scale simulation of floods, *Hydrology and Earth System Sciences*, 21, 5143–5163, doi:10.5194/hess-21-5143-2017.

Matgen, P., Schumann, G., Hentry, J.-B., Hoffmann, L., and Pfister, L. (2007), Integration of SAR-derived river inundation areas, high-precision topographic data and a river flow model toward near real-time flood management, *International Journal of Applied Earth Observation and Geoinformation*, 9(3), 247–263, doi:10.1016/j.jag.2006.03.003.

McBean, E. A., Gorrie, J., Fortin, M., Ding, J., and Moulton, R. (1988), Adjustment factors for flood damage curves, *Journal of Water Resources Planning and Management*, 114(6), doi:10.1061/(ASCE)0733-9496(1988)114:6(635).

McMillan, H., Krueger, T., and Freer, J. (2012), Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, *Hydrological Processes*, 26(26), 4078–4111, doi:10.1002/hyp.9384.

Mediero, L., Santillán, D., Garrote, L., and Granados, A. (2014), Detection and attribution of trends in magnitude, frequency and timing of floods in Spain, *Journal of Hydrology*, 517, 1072–1088, doi:10.1016/j.jhydrol.2014.06.040.

Mediero, L., Kjeldsen, T. R., Macdonald, N., Kohnova, S., Merz, B., Vorogushyn, S., Wilson, D., Alburquerque, T., Blöschl, g., Bogdanowicz, E., Castellarin, A., Hall, J., Kobold, M., Kriaučiūnienė, J., Lang, M., Madsen, H., Onuşluel Güll, G., Perdigão, R. A. P., Roald, L. A., Salinas, J. L., Toumazis, A. D., Veijalainen, N., Þórarinssonr, Ó (2015), Identification of coherent flood regions across Europe by using the longest streamflow records, *Journal of Hydrology*, 528, 341–360, doi:10.1016/j.jhydrol.2015.06.016.

Merz, B., Kreibich, H., Thieken, A., and Schmidtke, R. (2004), Estimation uncertainty of direct monetary flood damage to buildings, *Natural Hazards and Earth System Sciences*, 4, 153–163, doi:10.5194/nhess-4-153-2004.

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

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A. (2010), Review article "Assessment of economic flood damage", *Natural Hazards and Earth System Sciences*, 10, 1697–1724, doi:10.5194/nhess-10-1697-2010.

Meyer, V., Haase, D., and Scheuer, S. (2009), Flood risk assessment in European river basins – concept, methods, and challenges exemplified at the Mulde river, *Integrated Environmental Assessment and Management*, 5(1), 17–26, doi:10.1897/IEAM_2008-031.1.

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

Miller, S., Muir-Wood, R., and Boissonade, A. (2008), An exploration of trends in normalized weather related catastrophe losses. *In:* Diaz, H. F. and Murnane, R. J. (Eds.) *Climate extremes and society*, 225–247. New York, US: Cambridge University Press.

Milly, P. C. D. and Dunne, K. A. (2011), On the hydrologic adjustment of climate-model projections: the potential pitfall of potential evapotranspiration, *Earth Interactions*, 15, 1–14, doi:10.1175/2010EI363.1.

Moftakhari, H. R., Salvadori, G., AghaKouchak, A., Sanders, B. F., and Matthew, R. A. (2017), Compound effects of sea level rise and fluvial flooding, *Proceedings of the National Academy of Sciences of the United States of America*, 114(37), 9785–9790, doi:10.1073/pnas.1620325114.

Mohleji, S. and Pielke, R. (2014), Reconciliation of trends on global and regional economic losses from weather events: 1980–2008, *Natural Hazards Review*, 15(4), doi:0.1061/(ASCE)NH.1527-6996.0000141.

Montanari, A. (2012), Hydrology of the Po river: looking for changing patterns in river discharge, *Hydrology and Earth System Sciences*, 16, 3739–3747, doi:10.5194/hess-16-3739-2012.

Moriasi, D. N., Arnold, J. G., van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007), Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, *Transactions of the American Society of Agricultural and Biological Engineers*, 50(3), 885–900.

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

Moretti, G. and Orlandini, S. (2018), Hydrography-driven coarsening of grid digital elevation models, *Water Resources Research*, 54(5), 3654–3672, doi:10.1029/2017WR021206.

Mudelsee, M., Börngen, M., Tetzlaff, G., and Grünewald, U. (2003), No upward trends in the occurrence of extreme floods in central Europe, *Nature*, 425, 166–169, doi:10.1038/nature01928.

Muerth, M. J., Gauvin St-Denis, B., Ricard, S., Velásquez, J. A., Schmid, J., Minville, M., Caya, D., Chaumont, D., Ludwig, R., and Turcotte, R. (2013), On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, *Hydrology and Earth System Sciences*, 17, 1189–1204, doi:10.5194/hess-17-1189-2013.

Muis, S., Güneralp, B., Jongman, B., Aerts, J. C. J. H., and Ward, P. J. (2016), Flood risk and adaptation strategies under climate change and urban expansion: a probabilistic analysis using global data, *Science of The Total Environment*, 538, 445–457, doi:10.1016/j.scitotenv.2015.08.068.

Munich Re (2017), *Natural catastrophes 2016: analyses, assessments, positions*. Retrieved from: https://www.munichre.com/site/touch-publications/get/documents_E-271800065/mr/assetpool.shared/Documents/5_Touch/_Publications/TOPICS_GEO_2016-en.pdf

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

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

Munich Re (2019), *NatCatSERVICE*. Retrieved from https://natcatservice.munichre.com/

Murphy, E. A., Soong, D. T., and Sharpe, J. B. (2012), *Flood-Inundation Maps for a Nine-Mile Reach of the Des Plaines River from Riverwoods to Mettawa, Illinois*. U.S.

Geological Survey Scientific Investigations Report 2012–5227. Retrieved from https://pubs.usgs.gov/sir/2012/5227/

Murphy, C., Harrigan, S., Hall, J., and Wilby, R. L. (2013), Climate-driven trends in mean and high flows fro ma network of reference stations in Ireland, *Hydrological Sciences Journal*, 58(4), 755–772, doi:10.1080/02626667.2013.782407.

Musser, J. W. and Dyar, T. R. (2007), *Two-dimensional flood-inundation model of the Flint River at Albany, Georgia: Atlanta*. U.S. Geological Survey Scientific Investigations Report 2007-5107. Retrieved from https://pubs.usgs.gov/sir/2007/5107/

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

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

Nardi, F., Annis, A., Di Baldassarre, G., Vivoni, E. R., and Grimaldi, S. (2019), GFPLAIN250m, a global high-resolution dataset of Earth's floodplains, *Scientific Data*, 6, 180309, doi:10.1038/sdata.2018.309.

Naz, B. S., Kao, S.-C., Ashfaq, M., Rastogi, D., Mei, R., and Bowling, L. C. (2016), Regional hydrologic response to climate change in the conterminous United States using high-resolution hydroclimate simulations, *Global and Planetary Change*, 143, 100–117, doi:10.1016/j.gloplacha.2016.06.003.

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

National Weather Service (2015), *Hydrologic Information Center – Flood Loss Data*. Retrieved from: http://www.nws.noaa.gov/hic/

Neal, J. C., Bates, P. D., Fewtrell, T. J., Hunter, N. M., Wilson, M. D., and Horritt, M. S. (2009a), 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, doi:10.1016/j.jhydrol.2009.01.026.

Neal, J., Fewtrell, T., and Trigg, M. (2009b), Parallelisation of storage cell flood models using Open MP, *Environmental Modelling and Software*, 24(7), 872–877, doi:10.1016/j.envsoft.2008.12.004.

Neal, J. C., Fewtrell, T. J., Bates, P. D., and Wright, N. G., (2010), A comparison of three parallelization methods for 2D flood inundation models, *Environmental Modelling and Software*, 25(4), 398–411, doi:10.1016/j.envsoft.2009.11.007.

Neal, J., Schumann, G., Fewtrell, T., Budimir, M., Bates, P., and Mason, D. (2011), Evaluating a new LISFLOOD-FP formulation with data from the summer 2007 floods in Tewkesbury, UK, *Journal of Flood Risk Management*, 4(2), 88–95, doi:10.1111/j.1753-318X.2011.01093.x.

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

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

Neal, J., Dunne, T., Sampson, C., Smith, A., and Bates, P. (2018), Optimisation of the two-dimensional hydraulic model LISFLOOD-FP for CPU architecture, *Environmental Modelling and Software*, 107, 148–157, doi:10.1016/j.envsoft.2018.05.011.

Neumann, B., Vafeidis, A. T., Zimmermann, J., and Nicholls, R. J. (2015), Future coastal population growth and exposure to sea-level rise and coastal flooding – a global assessment, *PLoS ONE*, 10, e0118571, doi:10.1371/journal.pone.0118571.

Neumayer, E. and Barthel, F. (2011), Normalizing economic loss from natural disasters: a global analysis, *Global Environmental Change*, 21(1), 13–24, doi:10.1016/j.gloenvcha.2010.10.004.

New York University Furman Centre (2017), *FloodZoneData.us*. Retrieved from http://floodzonedata.us/

Nguyen, P., Thorstensen, A., Sorooshian, S., Hsu, K., AghaKouchak, A., Sanders, B., Koren, F., Cui, Z., and Smith, M. (2016), A high resolution coupled hydrologic-hydraulic

model (HiResFlood-UCI) for flash flood modeling, *Journal of Hydrology*, 541(A), 401–420, doi:10.1016/j.jhydrol.2015.10.047.

Nied, M., Pardowitz, T., Nissen, K., Ulbrich, U., Hundecha, Y., and Merz, B. (2014), On the relationship between hydro-meteorological patterns and flood types, *Journal of Hydrology*, 519(D), 3249–3262, doi:10.1016/j.jhydrol.2014.09.089.

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

Nka, B. N., Oudin, L., Karambiri, H., Paturel, J. E., and Ribstein, P. (2015), Trends in floods in West Africa: analysis based on 11 catchments in the region, *Hydrology and Earth System Sciences*, 19(11), 4707–4719, doi:10.5194/hess-19-4707-2015.

NOAA National Water Center, Boghici, E., and Arctur, D. (2018), *NOAA NWC – Harvey NWM-HAND Flood Extents* (HydroShare), doi:10.4211/hs.fe85a680d0144e79b39e8c483dc1e5aa.

Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., Waterloo, M., and Saleska, S. (2011), Height Above the Nearest Drainage – a hydrologically relevant new terrain model, *Journal of Hydrology*, 404(1-2), 13–29, doi:10.1016/j.jhydrol.2011.03.051.

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

Nordhaus, W., Azam, Q., Corderi, D., Hood, K., Victor, N. M., Mohammed, M., Miltner, A., and Weiss, J. (2006), *The G-Econ Database on Gridded Output: Methods and Data*. Retrieved from https://gecon.yale.edu/sites/default/files/files/gecon_data_20051206_1.pdf

O'Loughlin, F. E., Paiva, R. C. D., Durand, M., Alsdorf, D. E., and Bates, P. D. (2016), A multi-sensor approach towards a global vegetation corrected SRTM DEM product, *Remote Sensing of Environment*, 182, 49–59, doi:10.1016/j.rse.2016.04.018.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., and Zappa, M. (2015), Does model performance improve with complexity? A case study with three hydrological models, *Journal of Hydrology*, 523, 147–159, doi:10.1016/j.jhydrol.2015.01.044.

Ostheimer, C. J. (2013), *Development of flood profiles and flood-inundation maps for the village of Killbuck, Ohio*. U.S. Geological Survey Scientific Investigations Report 2013-1032, doi:10.3133/ofr20131032.

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

Pappenberger, F., Beven, K., Horritt, M., and Blazkova, S. (2005a), Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations, *Journal of Hydrology*, 302(1–4), 46–69, doi:10.1016/j.jhydrol.2004.06.036.

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

Pappenberger, F., Frodsham, K., Beven, K., Romanowicz, R., and Matgen, P. (2007a), Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations, *Hydrology and Earth System Sciences*, 11, 739–752, doi:10.5194/hess-11-739-2007.

Pappenberger, F., Beven, K., Frodsham, K., Romanowicz, R., and Matgen, P. (2007b), Grasping the unavoidable subjectivity in calibration of flood inundation models: a vulnerability weighted approach, *Journal of Hydrology*, 333(2–4), 275–287, doi:10.1016/j.jhydrol.2006.08.017.

Pappenberger, F., Cloke, H. L., Balsamo, G., Ngo-Duc, T., and Oki, T. (2010), Global runoff routing with the hydrological component of the ECMWF NWP system, *International Journal of Climatology*, 30(14), 2155–2174, doi:10.1002/joc.2028.

Pappenberger, F., Dutra, E., Wetterhall, F., and 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, doi:10.5194/hess-16-4143-2012.

Paschalis, A., Fatichi, S., Molnar, P., Rimkus, S., and Burlando, P. (2014), On the effects of small scale space-time variability of rainfall on basin flood response, *Journal of Hydrology*, 514, 313–327, doi:10.1016/j.jhydrol.2014.04.014.

Passalacqua, P., Belmont, P., Foufoula-Georgiou, E., (2012), Automatic geomorphic feature extraction from lidar in flat and engineered landscapes, *Water Resources Research*, 48(3), W03528, doi:10.1029/2011WR010958.

Passalacqua, P., Tarolli, P., and Foufoula-Georgiou, E., (2010), Testing space-scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscale, *Water Resources Research*, 46(11), W11535, doi:10.1029/2009WR008812.

Pavelsky, T. M. and Smith, L. C. (2008), RivWidth: a software tool for the calculation of river widths from remotely sensed imagery, *IEEE Geoscience and Remote Sensing Letters*, 5(1), 70–73, doi:10.1109/LGRS.2007.908305.

Pavelsky, T. M., Durand, M. T., Andreadis, K. M., Beighley, R. E., Paiva, R. C. D., Allen, G. H., and Miller, Z. F. (2014), Assessing the potential global extent of SWOT river discharge observations, *Journal of Hydrology*, 519(B), 1516–1525, doi:10.1016/j.jhydrol.2014.08.044.

Pekárová, P., Pacl, J., Liová, S., Miklánek, P., Škoda, P., and Pekár, J. (2009), Analysis of the maximum annual discharge regime in the high mountain basin of Belá river in Podbanské, *Acta Hydrologica Slovaca*, 10, 300–311.

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

Peterson, T. C., Heim, R. R., Hirsch, R., Kaiser, D. P., Brooks, H., Diffenbaugh, N. S., Dole, R. M., Giovannettone, J. P., Guirguis, K., Karl, T. R., Katz, R. W., Kunkel, K., Lettenmaier, D., McCabe, G. J., Ryberg, K. R., Schubert, S., Silva, V. B. S., Stewart, B. C., Vecchia, A. V., Villarini, G., Vose, R. S., Walsh, J., Wehner, M., Wolock, D., Wolter, K., Woodhouse, C. A., and Wuebbles, D. (2013), Monitoring and

understanding changes in heat waves, cold waves, floods, and droughts in the United States: state of knowledge, *Bulletin of the American Meteorological Society*, 94, 821–834, doi:10.1175/BAMS-D-12-00066.1.

Petrow, T. and Merz, B. (2009), Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002, *Journal of Hydrology*, 371(1–4), doi:10.1016/j.jhydrol.2009.03.024.

Pielke, R. A. and Landsea, C. W. (1998), Normalized hurricane damages in the United States: 1925–95, *Weather and Forecasting*, 13, 621–631, doi:10.1175/1520-0434(1998)013<0621:NHDITU>2.0.CO;2.

Pielke, R. A. and Downton, M. W. (2000), Precipitation and damaging floods: trends in the United States, 1932–97, *Journal of Climate*, 13, 3625–3637, doi:10.1175/1520-0442(2000)013<3625:PADFTI>2.0.CO;2.

Pielke, R. A., Rubiera, J., Landsea, C., Fernández, M. L., and Klein, R. (2003), Hurricane vulnerability in Latin America and the Caribbean: normalized damage and loss potentials, *Natural Hazards Review*, 4(3), 101–114, doi:10.1061/~ASCE!1527-6988~2003!4:3~101!.

Pielke, R. A. (2007), Future economic damage from tropical cyclones: sensitivities to societal and climate changes, *Philosophical Transactions of the Royal Society A*, 365(1860), doi:10.1098/rsta.2007.2086.

Pinter, N., Thomas, R., and Wlosinski, J. H. (2001), Assessing flood hazard on dynamic rivers, *Eos*, 82(31), 333–344, doi:10.1029/01EO00199.

Pinter, N., Jemberie, A. A., Remo, J. W. F., Heine, R. A., and Ickes, B. S. (2008), Flood trends and river engineering on the Mississippi river system, *Geophysical Research Letters*, 35(23), doi:10.1029/2008GL035987.

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

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

Prudhomme, C. and Davis, H. (2009), Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: baseline climate, *Climatic Change*, 93, 177–195, doi:10.1007/s10584-008-9464-3.

Prudhomme, C., Parry, S., Hannaford, J., and Clark, D. B. (2011), How well do large-scale models reproduce regional hydrological extremes in Europe?, *Journal of Hydrometeorology*, 12, 1181–1204, doi:10.1175/2011JHM1387.1.

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., and Heffernan, J. (2019), The spatial dependence of flood hazard and risk in the United States, *Water Resources Research*, 55(3), 1890–1911, doi:10.1029/2018WR024205.

Quinn, N., Lewis, M., Wadley, M. P., and Haigh, I. D. (2014), Assessing the temporal variability in extreme storm-tide time series for coastal flood risk assessment, *Journal of Geophysical Research*, 119(8), 4983–4998, doi:10.1002/2014JC010197.

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

Raghavan, S. and Rajesh, S. (2003), Trends in tropical cyclone impact: a study in Andhra Pradesh, India, *Bulletin of the American Meteorological Society*, 84, 635–644, doi:10.1175/BAMS-84-5-635.

Refsgaard, J. C. and Knudsen, J. (1996), Operational validation and intercomparison of different types of hydrological models, *Water Resources Research*, 32(7), 2189–2202, doi:10.1029/96WR00896.

Reihan, A., Kriaučiūnienė, J., Meilutyte-Barauskiene, D., and Kolcova, T. (2012), Temporal variation of spring flood in rivers of the Baltic States, *Hydrology Research*, 43(4), 301–314, doi:10.2166/nh.2012.141.

Renard, B., Lang, M., Bois, P., Dupeyray, A., Mestre, O., Niel, H., Sauquet, E., Prudhomme, C., Parey, S., Paquet, E., Neppel, L., and Gailhard, J. (2008), Regional methods for trend detection: assessing field significance and regional consistency, *Water Resources Research*, 44(8), doi:10.1029/2007WR006268.

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., and Waterloo, M. J. (2008), HAND, a new terrain descriptor using SRTM-DEM: Mapping

terra-firme rainforest environments in Amazonia, *Remote Sensing of the Environment*, 112(9), 3469–3481, doi:10.1016/j.rse.2008.03.018.

Ritter, A. and Muñoz-Carpena, R. (2013), Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments, *Journal of Hydrology*, 480, 33–45, doi:10.1016/j.jhydrol.2012.12.004.

Robson, A. J., Jones, T. K., Reed, D. W., and Bayliss, A. C. (1998), A study of national trend and variation in UK floods, *International Journal of Climatology*, 18(2), 165–182, doi:10.1002/(SICI)1097-0088(199802)18:2<165::AID-JOC230>3.0.CO;2-%23.

Rodda, H. J. E. (2005), The Development and Application of a Flood Risk Model for the Czech Republic, *Natural Hazards*, 36(1-2), 207–220, doi:10.1007/s11069-004-4549-4.

Rojas, R., Feyen, L., Dosio, A., and Bavera, D. (2011), Improving pan-European hydrological simulation of extreme events through statistical bias correction of RCM-driven climate simulations, *Hydrology and Earth System Sciences*, 15, 2599–2620, doi:10.5194/hess-15-2599-2011.

Rojas, R., Feyen, L., Bianchi, A., and Dosio, A. (2012), Assessment of future flood hazard in Europe using a large ensemble of bias-corrected regional climate simulations, *Journal of Geophysical Research Atmospheres*, 117(D17), doi:10.1029/2012JD017461.

Roland, M. A., Underwood, S. M., Thomas, C. M., Miller, J. F., Pratt, B. A., Hogan, L. G., and Wnek, P. A. (2014), *Flood-inundation maps for the Susquehanna River near Harrisburg, Pennsylvania, 2013*. U.S. Geological Survey Scientific Investigations Report 2014-5046, doi:10.3133/sir20145046.

Romanowicz, R., Beven, K. J., and Tawn, J. (1996), Bayesian calibration of flood inundation models. *In:* Anderson, M. G., Walling, D. E., and Bates, P. D. (Eds.) *Floodplain Processes*, 333–360. Chichester, UK: Wiley.

Roudier, P., Andersson, J. C. M., Donnelly, C., Feyen, L., Freuell, W. and Ludwig, F. (2016), Projections of future floods and hydrological droughts in Europe under a +2°C global warming, *Climatic Change*, 135(2), 341–355, doi:10.1007/s10584-015-1570-4.

Rudorff, C. M., Melack, J. M., and Bates, P. D. (2014), Flooding dynamics on the lower Amazon floodplain: 1. Hydraulic controls on water elevation, inundation extent, and

river-floodplain discharge, *Water Resources Research*, 50(1), 619–634, doi:10.1002/2013WR014091.

Salathé, E. P., Hamlet, A. F., Mass, C. F., Lee, S.-Y., Stumbaugh, M., and Steed, R. (2014), Estimates of twenty-first-century flood risk in the Pacific Northwest based on regional climate model simulations, *Journal of Hydrometeorology*, 15, 1881–1899, doi:10.1175/JHM-D-13-0137.1.

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivalapan, M., and Blöschl, G. (2013), Comparative assessment of predictions in ungauged basins – part 2: flood and low flow studies, *Hydrology and Earth System Sciences*, 17, 2637–2652, doi:10.5194/hess-17-2637-2013.

Samela, C., Manfreda, S., De Paola, F., and Giugni, M. (2016), DEM-based approaches for the delineation of flood-prone areas in an ungauged basin in Africa, *Journal of Hydrologic Engineering*, 21(2), doi:10.1061/(ASCE)HE.1943-5584.0001272.

Samir, K. C. and Lutz, W. (2014), Demographic scenarios by age, sex and education corresponding to the SSP narratives, *Population and Environment*, 35, 243–260, doi:10.1007/s11111-014-0205-4.

Sampson, C. C., Fewtrell, T. J., Duncan, A., Shaad, K., Horritt, M. S., and Bates, P. D. (2012), Use of terrestrial laser scanning data to drive decimetric resolution urban inundation models, *Advances in Water Resources*, 41, 1–17, doi:10.1016/j.advwatres.2012.02.010.

Sampson, C. C., Bates, P.D., Neal., J.C., Horritt, M.S., (2013), An automated routing methodology to enable direct rainfall in high resolution shallow water models, *Hydrological Processes*, 27(3), 467–476, doi: 10.1002/hyp.9515.

Sampson, C. C., Fewtrell, T. J., O'Loughlin, F., Pappenberger, F., Bates, P. D., Freer, J. E., and Cloke, H. L. (2014), The impact of uncertain precipitation data on insurance loss estimates using a flood catastrophe model, *Hydrology and Earth System Sciences*, 18, 2305–2324, doi:10.5194/hess-18-2305-2014.

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

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., and 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, doi:10.3389/feart.2015.00085.

Samuels, P. G. (1990), Cross section location in one-dimensional models. *In:* White, W. R. (Ed.) *Proceedings of the International Conference on River Flood Hydraulics*, 339–350. Chichester, UK: Wiley.

Sanders, B. F., and Schubert, J. E. (2019), PRIMo: parallel raster inundation model, *Advances in Water Resources*, 126, 79–95, doi:10.1016/j.advwatres.2019.02.007.

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

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

Sangireddy, H., Stark, C. P., Kladzyk, A., and 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, doi:10.1016/j.envsoft.2016.04.026.

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

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P. (2014), Multimodel assessment of water scarcity under climate change, *Proceedings of the National Academy of Sciences of the United States of America*, 111(9), 3245–3250, doi:10.1073/pnas.1222460110.

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.

Schmidt, S., Kemfert, C., and Faust, E. (2009), *Simulation of economic losses from tropical cyclones in the years 2015 and 2050: the effects of anthropogenic climate change and growing wealth*. DIW Berlin Discussion Paper 914, doi:10.2139/ssrn.1494425.

Schmocker-Fackel, P. and Naef, F. (2010), More frequent flooding? Changes in flood frequency in Switzerland since 1850, *Journal of Hydrology*, 381(1–2), 1–8, doi:10.1016/j.jhydrol.2009.09.022.

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

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

Schumann, G., Di Baldassarre, G., Alsdorf, D., and 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(5), W05601, doi:10.1029/2008WR007672.

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

Schumann, G. J.-P. (2014), Fight floods on a global scale, *Nature*, 507, 169, doi:10.1038/507169e.

Schumann, G. J.-P. and Moller, D. K. (2015), Microwave remote sensing of flood inundation, *Physics and Chemistry of the Earth, Parts A/B/C*, 83–84, 84–95, doi:10.1016/j.pce.2015.05.002.

Schumann, G. J.-P. and Bates, P. D. (2018), The need for a high-accuracy, open-access global DEM, *Frontiers in Earth Science*, 6, 225, doi:10.3389/feart.2018.00225.

Scussolini, P., Aerts, J. C. J. H., Jongman, B., Bouwer, L. M., Winsemius, H. C., de Moel, H., and Ward, P. J. (2016), FLOPROS: an evolving global database of flood protection standards, *Natural Hazards and Earth System Sciences*, 16, 1049–1061, doi:10.5194/nhess-16-1049-2016.

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X. (2012), Changes in climate extremes and their impacts on the natural physical environment. *In:* Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M., and Midgley, P. M. (Eds.) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press.

Sharma, A., Wasko, C., and Lettenmaier, D. P. (2018), If precipitation extremes are increasing, why aren't floods?, *Water Resources Research*, 54(11), 8545–8551, doi:10.1029/2018WR023749.

Shkolnik, I., Pavlova, T., Efimov, S., and Zhuravlev, S. (2018), Future changes in peak river flows across northern Eurasia as inferred from an ensemble of regional climate projections under the IPCC RCP8.5 scenario, *Climate Dynamics*, 50(1–2), 215–230, doi:10.1007/s00382-017-3600-6.

Sikorksa, A. E., Viviroli, D., and Seibert, J. (2015), Flood-type classification in mountainous catchments using crisp and fuzzy decision trees, *Water Resources Research*, 51(10), 7959–7976, doi:10.1002/2015WR017326.

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

Smith, A., Freer, J., Bates, P., and Sampson, C. (2014), Comparing ensemble projections of flooding against flood estimation by continuous simulation, *Journal of Hydrology*, 511, 205–219, doi:10.1016/j.jhydrol.2014.01.045.

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

Smith, A., Bates, P. D., Wing, O., Sampson, C., Quinn, N., and Neal, J. (2019), New estimates of flood exposure in developing countries using high-resolution population data, *Nature Communications*, 10, 1814, doi:10.1038/s41467-019-09282-y.

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

Sofia, G., Dalla Fontana, G., and Tarolli, P. (2014), High-resolution topography and anthropogenic feature extraction: testing geomorphometric parameters in floodplains, *Hydrological Processes*, 28(4), 2046–2061, doi:10.1002/hyp.9727.

Sofia, G., Tarolli, P., Cazorzi, F., and 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, doi:10.5194/hess-15-1387-2011.

Somlyódy, L., Nováky, B., and Simonffy, Z. (2010), Climate change, extremes and water management, *CLIMA-21 Brochures Climate change–impact–responses*, 61, 15–32.

Sorribas, M. V., Paiva, R. C. D., Melack, J. M., Bravo, J. M., Jones, C., Carvalho, L., Beighley, E., Forsberg, B., and Costa, M. H. (2016), Projections of climate change effects on discharge inundation in the Amazon basin, *Climatic Change*, 136(3–4), 555–570, doi:10.1007/s10584-016-1640-2.

Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J., and Bierkens, M. F. P. (2010), The ability of a GCM-forced hydrological model to reproduce global discharge variability, *Hydrology and Earth System Sciences*, 14, 1595–1621, doi:10.5194/hess-14-1595-2010.

Stagl, J. C. and Hattermann, F. F. (2015), Impacts of climate change on the hydrological regime of the Danube river and its tributaries using an ensemble of climate scenarios, *Water*, 7(11), 6139–6172, doi:10.3390/w7116139.

Stahl, K., Tallaksen, L. M., Hannaford, J., and van Lanen, H. A. J. (2012), Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble, *Hydrology and Earth System Sciences*, 16, 2035–2047, doi:10.5194/hess-16-2035-2012.

Steele-Dunne, S., Lynch, P. McGrath, R., Semmler, T., Wang, S., Hanafin, J., Nolan, P. (2008), The impacts of climate change on hydrology in Ireland, *Journal of Hydrology*, 356(1–2), 28–45, doi:10.1016/j.jhydrol.2008.03.025.

Steinfeld, C. M. M., Kingsford, R. T., and Laffan, S. W. (2013), Semi-automated GIS techniques for detecting floodplain earthworks, *Hydrological Processes*, 27(4), 579–591, doi:10.1002/hyp.9244.

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

Stephens, E. and Bates, P. (2015), Assessing the reliability of probabilistic flood inundation model predictions, *Hydrological Processsses*, 29(19), 4264–4283, doi:10.1002/hyp.10451.

Stoesser, T., Wilson, C. A. M. E., Bates, P. D., and Dittrich, A. (2003), Application of a 3D numerical model to a river with vegetated floodplains, *Journal of Hydroinformatics*, 5(2), 99–112, doi:10.2166/hydro.2003.0008.

Storm, J. B. (2014), *An expanded model: flood-inundation maps for the Leaf River at Hattiesburg, Mississippi, 2013*. U.S. Geological Survey Scientific Investigations Map 3300, doi:10.3133/sim3300.

Straatsma, M. W. and Baptist, M. J. (2008), Floodplain roughness parameterization using airborne laser scanning and spectral remote sensing, *Remote Sensing of Environment*, 112(3), 1062–1080, doi:10.1016/j.rse.2007.07.012.

Strahler, A. N. (1964), Quantitative geomorphology of drainage basins and channel networks. *In:* Chow, V. T. (Ed.), *Handbook of Applied Hydrology*. New York, US: McGraw Hill.

Strupczewski, W. G., Singh, V. P., and Mitosek, H. T. (2001), Non-stationary approach to at-site flood frequency modelling. III. Flood analysis of Polish rivers, *Journal of Hydrology*, 248(1–4), 152–167, doi:10.1016/S0022-1694(01)00399-7.

Tanoue, M., Hirabayashi, Y., and Ikeuchi, H. (2016), Global-scale river flood vulnerability in the last 50 years, *Scientific Reports*, 6, 36021, doi:10.1038/srep36021.

Tarolli, P., Sofia, G., and Dalla Fontana, G. (2012), Geomorphic features extraction from high-resolution topography: landslide crowns and bank erosion, *Natural Hazards*, 61(1), 65–83, doi:10.1007/s11069-010-9695-2.

Tatem, A. J. (2017), WorldPop, open data for spatial demography, *Scientific Data*, 4, 170004, doi:10.1038/sdata.2017.4.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012), An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, 93, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Taylor, H., Bates, P., and Neal, J. (2019), *ArcticDEM; DSM to DTM conversion and associated hydrodynamic flood modelling*. EGU General Assembly Conference Abstracts, vol. 21, EGU2019-1239.

Theobold, D. M. (2001), Land-use dynamics beyond the American urban fringe, *Geographical Review*, 91(3), 544–564, doi:10.2307/3594740.

Theobold, D. M. (2014), Development and applications of a comprehensive land use classification and map for the US, *PLoS ONE*, 9, e94628, doi:10.1371/journal.pone.0094628.

Thieken, A. H., Olschewski, A., Kreibich, H., Kobsch, S., and Merz, B. (2008), Development and evaluation of FLEMOps – a new flood loss estimation model for the private sector, *WIT Transactions on Ecology and the Environment*, 118, 315–324.

Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A. (2009), The European Flood Alert System – Part 1: Concept and development, *Hydrology and Earth System Sciences*, 13, 125–140, doi:10.5194/hess-13-125-2009.

Thober, S., Kumar, R., Wanders, N., Marx, A., Pan, M., Rakovec, O., Samaniego, L., Sheffield, J., Wood, E. F., and Zink, M. (2018), Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming, *Environmental Research Letters*, 13(1), 014003, doi:10.1088/1748-9326/aa9e35.

Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E. B., and Dang, H.-A. H. (2017), *Mapping the world population one building at a time*. arXiv:1712.05839, Cornell University: arXiv preprint.

Tohver, I. M., Hamlet, A. F., and Lee, S.-Y. (2014), Impacts of 21st-century climate change on hydrologic extremes in the Pacific Northwest region of North America, *Journal of the American Water Resources Association*, 50(6), 1461–1476, doi:10.1111/jawr.12199.

Trigg, M. A., Birch, C. E., Neal, J. C., Bates, P. D., Smith, A., Sampson, C. C., Yamazaki, D., Hirabayashi, Y., Pappenberger, F., Dutra, E., Ward, P. J., Winsemius, H. C., Salamon, P., Dottori, F., Rudari, R., Kappes, M. S., Simpson, A. L., Hadzilacos, G., and Fewtrell, T. J. (2016), The credibility challenge for global fluvial flood risk analysis, *Environmental Research Letters*, 11, 094014, doi:10.1088/1748-9326/11/9/094014.

UN International Strategy for Disaster Risk Reduction (2015a), *Sendai Framework for Disaster Risk Reduction 2015–2030*. Geneva, Switzerland: United Nations Office for Disaster Risk Reduction.

UN International Strategy for Disaster Risk Reduction (2015b), *Making Development Sustainable: the Future of Disaster Risk Management*. Global Assessment Report on Disaster Risk Reduction. Geneva, Switzerland: United Nations Office for Disaster Risk Reduction.

US Army Corps of Engineers (2016a), *Hydrologic Modeling System HEC-HMS User's Manual*. Retrieved from https://www.hec.usace.army.mil/software/hec-hms/documentation/HEC-HMS_Users_Manual_4.2.pdf

US Army Corps of Engineers (2016b), *HEC-RAS River Analysis System User's Manual*. Retrieved from https://www.hec.usace.army.mil/software/hec-ras/documentation/HEC-RAS%205.0%20Users%20Manual.pdf

US Census Bureau (2000), Assumptions for the components of change. *In: Methodology and Assumptions for the Population Projections of the United States: 1999-2100.* Retrieved from http://www.census.gov/population/projections/data/national/natproj2000.html/

US Census Bureau (2003), *Census 2000, Public Use Microdata Sample (PUMS)*. Retrieved from https://www.census.gov/census2000/PUMS5.html/

US Environmental Protection Agency (2015), *Flood Loss Avoidance Benefits of Green Infrastructure for Stormwater Management*. Retrieved from

http://www.epa.gov/green-infrastructure/flood-loss-avoidance-benefits-green-infrastructure-stormwater-management/

US Environmental Protection Agency (2016), *Updates to the Demographic and Spatial Allocation Models to Produce Integrated Climate and Land Use Scenarios (ICLUS) Version 2*, EPA/600/R-16/366F. Washington, DC, US: National Center for Environmental Assessment.

US Geological Survey (2016), *FESWMS-2DH: Finite-Element Surface-Water Modeling System for Two-Dimensional Flow in the Horizontal Plane*. Retrieved from https://water.usgs.gov/software/FESWMS-2DH/

van Beek, L. P. H. and Bierkens, M. F. P. (2009), *The global hydrological model PCR-GLOBWB: conceptualization, parameterization and verification*. Department of Physical Geography, Utrecht University.

van der Knijff, J. M., Younis, J., and de Roo, A. P. J. (2010), LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, *International Journal of Geographical Information Science*, 24(2), 189–212, doi:10.1080/13658810802549154.

van der Made, J. E. (1982), Determination of the accuracy of water level observations, *Proceedings of the Exeter Symposium, IAHS Publications 134*, 172–184.

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

van Pelt, S. C., Kabat, P., ter Maat, H. W., van den Hurk, B. J. J. M., and Weerts, A. H. (2009), Discharge simulations performed with a hydrological model using bias corrected regional climate model input, *Hydrology and Earth System Sciences*, 13, 2387–2397, doi:10.5194/hess-13-2387-2009.

van Vuuren, D. P. and Carter, T. R. (2014), Climate and socio-economic scenarios for climate change research and assessment: reconciling the new with the old, *Climatic Change*, 122(3), 415–429, doi:10.1007/s10584-013-0974-2.

Veijalainen, N., Lotsari, E., Alho, P., Vehviläinen, B., and Käyhkö, J. (2010), National scale assessment of climate change impacts on flooding in Finland, *Journal of Hydrology*, 391(3–4), 333–350, doi:10.1016/j.jhydrol.2010.07.035.

Villarini, G., Serinaldi, F., Smith, J. A., and Krajewski, W. F. (2009a), On the stationarity of annual flood peaks in the continental United States during the 20th century, *Water Resources Research*, 45(8), doi:10.1029/2008WR007645.

Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., Krajewski, W. F. (2009b), Flood frequency analysis for nonstationary annual peak records in an urban drainage basin, *Advances in Water Resources*, 32(8), 1255–1266, doi:10.1016/j.advwatres.2009.05.003.

Villarini, G., Smith, J. A., Baeck, M. L., and Krajewski, W. F. (2011), Examining flood frequency distributions in the midwest U.S., *Journal of the American Water Resources Association*, 47(3), 447–463, doi:10.1111/j.1752-1688.2011.00540.x.

Vogel, R. M., Yaindl, C., and Walter, M. (2011), Nonstationary: flood magnification and recurrence reduction factors in the United States, *Journal of the American Water Resources Association*, 47(3), 464–474, doi:10.1111/j.1752-1688.2011.00541.x.

Vörösmarty, C., Askew, A., Grabs, W., Barry, R. G., Birkett, C., Döll, P., Goodison, B., Hall, A., Jenne, R., Kitaev, L., Landwehr, J., Keeler, M., Leavesley, G., Schaake, J., Strzepek, K., Sundarvel, S. S., Takeuchi, K., and Webster, F. (2001), Global water data: a newly endangered species, *Eos*, 82(5), 54–58, doi:10.1029/01EO00031.

Wahl, T., Jain, S., Bender, J., Meyers, S. D., and Luther, M. E. (2015), Increasing risk of compound flooding from storm surge and rainfall for major US cities, *Nature Climate Change*, 5, 1093–1097, doi:10.1038/nclimate2736.

Ward, P. J., Jongman, B., Sperna Weiland, F., Bouwman, A., van Beek, R., Bierkens, M. F. P., Ligtvoet, W., and Winsemius, H. C. (2013), Assessing flood risk and the global scale: model setup, results, and sensitivity, *Environmental Research Letters*, 8(4), 044019, doi:10.1088/1748-9326/8/4/044019.

Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., Muis, S., Coughlan de Perez, E., Rudari, R., Trigg, M. A., and Winsemius, H. C. (2015), Usefulness and limitations of global flood risk models, *Nature Climate Change*, 5, 712–715, doi:10.1038/nclimate2742.

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

Ward, P. J., Couasnon, A., Eilander, D., Haigh, I. D., Hendry, A., Muis, S., Veldkamp, T. I. E., Winsemius, H. C., and Wahl, T. (2018), Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries, *Environmental Research Letters*, 13(8), 084012, doi:10.1088/1748-9326/aad400.

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

Watson, K. M. and Niemoczynski, M. J. (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. Geological Survey Scientific Investigations Map 3299, doi:10.3133/sim3299.

Wehner, M. F., Reed, K. A., Li, F., Prabhat, Bacmeister, J., Chen, C.-T., Paciorek, C., Gleckler, P. J., Sperber, K. R., Collins, W. D., Gettelman, A., and Jablonowski, C. (2014), The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1, *Journal of Advances in Modeling Earth Systems*, 6, 980–997, doi:10.1002/2013MS000276.

Weinkle, J., Landsea, C., Collins, D., Musulin, R., Crompton, R. P., Klotzbach, P. J., and Pielke, R. (2018), Normalized hurricane damage in the continental United States 1990–2017, *Nature Sustainability*, 1, 808–813, doi:10.1038/s41893-018-0165-2.

Werner, M., Blazkova, S., and Petr, J. (2005a), Spatially distributed observations in constraining inundation modelling uncertainties, *Hydrological Processes*, 19, 3081–3096, doi:10.1002/hyp.5833.

Werner, M. G. F., Hunter, N. M., and Bates, P. D. (2005b), Identifiability of distributed floodplain roughness values in flood extent estimation, *Journal of Hydrology*, 314(1–4), 139–157, doi:10.1016/j.jhydrol.2005.03.012.

Wessel, B., Huber, M., Wohlfart, C., Marschalk, U., Kosmann, D., and Roth, A. (2018), Accuracy assessment of the global TanDEM-X digital elevation model with GPS data, *ISPRS Journal of Photogrammetry and Remote Sensing*, 139, 171–182, doi:10.1016/j.isprsjprs.2018.02.017.

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

Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G., and Roberts, N. M. (2014), Future changes to the intensity and frequency of short-duration extreme rainfall, *Reviews of Geophysics*, 52(3), 522–555, doi:10.1002/2014RG000464.

Williams, G. P. (1978), Bank-full discharge of rivers, *Water Resources Research*, 14(6), 1141–1154, doi:10.1029/WR014i006p01141.

Wilson, M., Bates, P., Alsdorf, D., Forsberg, B., Horritt, M., Melack, J., Frappart, F., and Famiglietti, J. (2007), Modeling large-scale inundation of Amazonian seasonally flooded wetlands, *Geophysical Research Letters*, 34(15), L15404, doi:10.1029/2007GL030156.

Wilson, D., Hisdal, H., and Lawrence, D. (2010), Has streamflow changed in the Nordic countries? – Recent trends and comparisons to hydrological projections, *Journal of Hydrology*, 394(3–4), 334–346, doi:10.1016/j.jhydrol.2010.09.010.

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

Wing, O. E. J., Bates, P. D., Smith, A. M., Sampson, C. C., Johnson, K. A., Fargione, J., and Morefield, P. (2018a), Estimates of present and future flood risk in the conterminous United States, *Environmental Research Letters*, 13, 034023, doi:10.1088/1748-9326/aaac65.

Wing, O., Bates, P., Sampson, C., Smith, A., Fargione, J., and Johnson, K. (2018b), Millions more Americans face flood risks than previously thought, *Eos*, 99, doi:10.1029/2018EO106321.

Wing, O. E. J., Sampson, C. C., Bates, P. D., Quinn, N., Smith, A. M., and Neal, J. C. (2019), A flood inundation forecast of Hurricane Harvey using a continental-scale 2D hydrodynamic model, *Journal of Hydrology X*, 4(C), 100039, doi:10.1016/j.hydroa.2019.100039.

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

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

Winsemius, H. C., Ward, P. J., Gayton, I., Veldhuis, M.-C., Meiher, D. H., and Iliffe, M. (2019), Commentary: the need for high-accuracy, open-access global DEM, *Frontiers in Earth Science*, 7, 33, doi:10.3389/feart.2019.00033.

Wobus, C., Gutmann, E., Jones, R., Rissing, M., Mizukami, N., Lorie, M., Mahoney, H., Wood, A. W., Mills, D., and Martinich, J. (2017), Climate change impacts on flood risk and asset damages within mapped 100-year floodplains of the contiguous United States, *Natural Hazards and Earth System Sciences*, 17, 2199–2211, doi:10.5194/nhess-17-2199-2017.

Wood, J. D. (1996), *The geomorphological characteristics of digital elevation models.* PhD Thesis: Department of Geography, University of Leicester.

Wood, E. F., Roundy, J. K., Troy, T. J., van Been, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield J., Wade, A., and Whitehead, P. (2011), Hyperresolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resources Research*, 47(5), doi:10.1029/2010WR010090.

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

Wood, M., de Jong, S. M., and 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, doi:10.1016/j.jag.2018.04.003.

World Resources Institute (2015), *Aqueduct Global Flood Analyzer*. Retrieved from http://floods.wri.org/

Xiong, L., Du, T., Xu, C.-Y., Guo, S., Jiang, C., and Gippel, C. J. (2015), Non-stationary annual maximum flood frequency analysis using the norming constants method to consider non-stationarity in the annual daily flow series, *Water Resources Management*, 29(10), 3615–3633, doi:10.1007/s11269-015-1019-6.

Yamazaki, D., Kanae, S., Kim, H., and Oki, T. (2011), A physically based description of floodplain inundation dynamics in a global river routing model, *Water Resources Research*, 47(4), W04501, doi:10.1029/2010WR009726.

Yamazaki, D., Baugh, C. A., Bates, P. D., Kanae, S., Alsdorf, D. E., and Oki, T. (2012), Adjustment of a spaceborne DEM for use in floodplain hydrodynamic modeling, *Journal of Hydrology*, 436–437, 81–91, doi:10.1016/j.jhydrol.2012.02.045.

Yamazaki, D., O'Loughlin, F., Trigg, M. A., Miller, Z. F., Pavelsky, T. M., and Bates, P. D. (2014), Development of the Global Width Database for Large Rivers, *Water Resources Research*, 50(4), 3467–3480, doi:10.1002/2013WR014664.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D. (2017), A high-accuracy map of global terrain elevations, *Geophysical Research Letters*, 44(11), 5844–5853, doi:10.1002/2017GL072874.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M. (2019), MERIT Hydro: a high-resolution global hydrography map based on latest topography dataset, *Water Resources Research*, 55, doi:10.1029/2019WR024873.

Yan, K., Di Baldassarre, G., Solomatine, D., and Schumann, G. J.-P. (2015), A review of low-cost space-borne data for flood modelling: topography, flood extent and water level, *Hydrological Processes*, 29(15), 3368–3387, doi:10.1002/hyp.10449.

Yoon, Y., Durand, M., Merry, C. J., Clark, E. A., Andreadis, K. M., and Alsdorf, D. E. (2012), Estimating river bathymetry from data assimilation of synthetic SWOT measurements, *Journal of Hydrology*, 464–465, 363–375, doi:10.1016/j.jhydrol.2012.07.028.

Younis, B. A. (1996), Progress in turbulence modelling for open channel flows. *In:* Anderson, M. G., Walling, D. E., and Bates, P. D. (Eds.) *Floodplain Processes*, 299–332. Chichester, UK: Wiley.

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

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

Zanobetti, D., Lorgeré, H., Preissmann, A., and Cunge, J. A. (1970), Mekong delta mathematical model program construction, *Journal of the Waterways, Harbors and Coastal Engineering Division*, 96(2), 181–199.

Zhang, Q., Gu, X., Singh, V. P., Xiao, M., and Xu, C.-Y. (2015), Flood frequency under the influence of trends in the Pearl river basin, China: changing patterns, causes and implications, *Hydrological Processes*, 29(6), 1406–1417, doi:10.1002/hyp.10278.

Zhang, J., Huang, Y.-F., Munasinghe, D., Fang, Z., Tsang, Y.-P, and Cohen, S. (2018), Comparative Analysis of Inundation Mapping Approaches for the 2016 Flood in the Brazos River, Texas, *Journal of the American Water Resources Association*, 54(4), 820–833, doi:10.1111/1752-1688.12623.

Zheng, X., Maidment, D. R., Tarboton, D. G., Liu, Y. Y., and Passalacqua, P. (2018a), GeoFlood: Large-Scale Flood Inundation Mapping Based on High-Resolution Terrain Analysis, *Water Resources Research*, 54, 10013–10033 doi:10.1029/2018WR023457.

Zheng, X., Tarboton, D. G., Maidment, D. R., Liu, Y. Y., and Passalacqua, P. (2018b), River Channel Geometry and Rating Curve Estimation UsingHeight above the Nearest

Drainage, *Journal of the American Water Resources Association*, 54(4), 785–806, doi:10.1111/1752-1688.12661.

Zheng, Y. (2018), *Hurricane Harvey: A quantitative approach to assessing the accuracy of National Water Model forecasted inundation* (Master's thesis). Retrieved from Texas ScholarWorks (https://repositories.lib.utexas.edu/bitstream/handle/2152/68178/ZHENG-THESIS-2018.pdf). Austin, TX: University of Texas at Austin.

Zcheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., and Zhang, X. (2018), Future climate risk from compound events, *Nature Climate Change*, 8, 469–477, doi:10.1038/s41558-018-0156-3.

Zulkalfli, Z., Buytaert, W., Manz, B., Rosas, C. V., Willems, P., Lavado-Casimiro, W., Guyot, J.-L., and Santini, W. (2016), Projected increases in the annual flood pulse of the Western Amazon, *Environmental Research Letters*, 11, 014013, doi:10.1088/1748-9326/11/1/014013.