Flood Defense Standard Estimation Using Machine Learning and Its Representation in Large-Scale Flood Hazard Modeling

Gang Zhao[1†], Paul D. Bates[1,2], and Jeff Neal [1,2]

[1]School of Geographical Sciences, University of Bristol, Bristol, UK.

[2]Fathom, Engine Shed, Station Approach, Bristol, UK

†Corresponding author: Gang Zhao (gang.zhao@bristol.ac.uk)

Key Points

- A machine learning-based approach was developed for flood defense standard estimation using publicly available datasets.
- This approach was demonstrated in the conterminous United States and England, with a Nash–Sutcliffe efficiency of 0.82 and 0.73, respectively.
- Three case studies were used to test the reliable representation of the proposed approach in large-scale flood hazard modeling.

Abstract

We propose a machine learning-based approach to estimate the flood defense standard (FDS) for ungauged sites. We adopted random forest regression (RFR) to characterize the relationship between the observed FDS and ten explanatory factors contained in publicly available datasets. We compared RFR with multiple linear regression (MLR) and demonstrated the proposed approach in the conterminous United States (CONUS) and England, respectively. The results showed the following: (1) RFR performed better than MLR, with a Nash–Sutcliffe efficiency (NSE) of 0.82 in the CONUS and 0.73 in England. A negative NSE when using MLR indicated that the relationship between the FDS and each explanatory factor did not obey an explicit linear function. (2) River flood factors had higher importance than physical and socio-economic factors in the FDS estimation. The proposed approach achieved the highest performance using all factors for prediction and could not provide satisfactory predictions (NSE < 0.6) using physical or socio-economic factors individually. (3) We estimated the FDS for all ungauged sites in the CONUS and England. Approximately 80% and 29% of sites were identified as high or highest standard (> 100-year return period) in the CONUS and England, respectively. (4) We incorporated the estimated FDS in large-scale flood modeling and compared the model results with official flood hazard maps in three case studies. We identified obvious overestimations in protected areas when flood defenses were not taken into account; and flood defenses were successfully represented using the proposed approach.

1 Introduction

Floods are the most frequent type of natural hazards worldwide, and have caused significant loss of life and severe economic impacts for populations and property during the past two decades (CRED and UNDRR, 2020). To reduce the negative impact of floods, numerous types of flood defenses, such as levee systems, have been built to protect cities, towns, and farms in almost every country (Ubilla et al., 2008; Z. Wang & Liu, 2019). According to a report by the United States (U.S.) Army Corps of Engineers (USACE), approximately 11 million people and $1.3 trillion of property value existed in flood-defended areas in the U.S. as of 2018 (USACE, 2018). The number and standard of flood defenses continuously improves over time to meet societal needs and keep pace with rapid urbanization in floodplains (Leonard, 2008; T. Zhu et al., 2007). Flood defenses have significantly changed the regional flooding distribution and also residents' exposure (Di Baldassarre et al., 2009; Ludy & Kondolf, 2012), and this needs to be considered in flood hazard assessment.

With the increase of computing power and advances in remote sensing techniques, it is now possible to map flood hazards on a large scale at high resolution (< 100 m) (Ward et al., 2015). Fine resolution global hydrography datasets, such as MERIT Hydro (Yamazaki et al., 2019) and HydroBASINS (Bernhard Lehner & Grill, 2013) have been released; however, information on detailed flood defenses for most rivers in the world is severely limited (Aerts et al., 2020; Sampson et al., 2015). Existing state-of-the-art global flood hazard models either assume a simplified high flood defense standard (FDS) or assume no protection when applied (Aerts et al., 2020; Scussolini et al., 2016; Ward et al., 2015). This assumption causes the overestimation of flood hazards when the flooded areas are actually protected by existing flood defenses, and therefore induces a distorted flood hazard and risk assessment. The first global flood defense database (called FLOPROS) was built by collecting FDS data worldwide at the sub-country scale (Scussolini et al., 2016). FLOPROS assumes that the FDSs in a vast area are the same (i.e., most states in the US or all of Australia have the same defense standard) and ignores the heterogeneity of FDSs between rivers. The coarse resolution of the FDS data in FLOPROS cannot meet the requirement of large-scale flood hazard modelling at high resolution.

To accurately represent the effect of flood defenses in large-scale flood modelling, the modeler requires the location and standard of the flood defenses. In recent decades, national flood defense inventories, such as the U.S. national levee database (USNLD) (USACE, 2021), AIMS Spatial Flood Defenses database (UK Environment Agency, 2021), and openDELvE (O'Dell et al., 2021), have been published and adopted for flood hazard modelling. As these inventories are collected by a variety of agencies for different purposes, the coverage and consistency of the FDS data remains inadequate worldwide. Remote sensing techniques provide a low-cost method for the identification of flood defenses (Choung, 2014; Özer et al., 2019). In several studies, satellite images and advanced algorithms have been adopted successfully to extract flood defense locations at the regional to national scale (Maguya et al., 2014; Steinfeld et al., 2013; Wood et al., 2021). Compared with location data, FDS data are more difficult to obtain. An example can be found in the USNLD, which records the locations of more than 9,000 levee systems in the U.S.; however, only 20% of them have detailed levee parameters relating to the FDS. Recently, LiDAR-based digital terrain data have been introduced to extract the parameters of hydraulic structures (Sofia et al., 2011), which is also promising in FDS estimation. For example, Wing et al. (2019) developed an automated method to identify levee locations and extract the levee crest height from LiDAR-based digital terrain data and successfully incorporated the extracted crest height into large-scale flood hazard modelling. This method requires high-quality, high-resolution terrain data, and one sensitive parameter for this method (called the extraction rate threshold) still needs to be defined based on visual inspection. These deficiencies make this method difficult to apply in large-scale studies, particularly for data-sparse areas without high-resolution terrain data.

Because of the spatial resolution and geodetic datum conflicts between flood defense metadata and terrain data, the FDS in large-scale studies is typically described by the overtopping annual exceedance probability (overtopping AEP) rather than real structural parameters, such as the levee crest height. The overtopping AEP defines the return period of a flood exceeding the designed FDS (i.e., overtopping the levee and causing flooding in protected areas). This definition can easily be incorporated into large-scale flood hazard modelling that evaluates the flood hazard based on return period floods (AEP is the inverse of the return period). For example, the CIMA-UNEP global flood model does not incorporate flood defenses explicitly but simply identifies protected areas around large cities (Herold and Rudari, 2013; Rudari and Silvestro, 2015). The simulation is therefore of the undefended state, but any flooding predicted to occur in the identified protected areas is removed during model post-processing until the model-driven flood exceeds the overtopping AEP (Aerts et al., 2020). Another useful strategy is adopted in the Fathom global flood model, which considers the FDS during model pre-processing. The Fathom strategy links the FDS with the channel conveyance by determining the bankfull height of channels for different overtopping AEPs using flood frequency analysis (Sampson et al., 2015; Smith et al., 2015; Wing et al., 2017). As overtopping AEP data are insufficient, even in some data-rich countries, the FDS is regressed in both of the above models with respect to social-economic factors in protected areas. Specifically, the overtopping AEP in the CIMA-UNEP model is assumed to obey a linear function of the gross domestic product (GDP) value in urban areas (Herold & Rudari, 2013) and the Fathom model also assumes that the FDS increases as protected areas become more urbanized (Quinn et al., 2019). Both assumptions are derived from empirical data from particular case studies without comprehensive validation, and therefore induce a distorted flood hazard and risk assessment. As a result, estimating the FDS for ungauged sites has been highlighted as a key issue in flood hazard modelling (Bates et al., 2018; Ward et al., 2015) which we seek to address in this study.

We attempt to estimate the FDS considering three improvements:

a) In previous studies, FDS was mainly estimated considering the social-economic conditions in protected areas (Herold & Rudari, 2013; Quinn et al., 2019). However, Wing et al. (2019) proved that social-economic factors are inappropriate for use in the FDS estimation by comparing urbanity, wealth, and spending between protected and unprotected areas in the CONUS. This result is expected because the FDS should be designed to consider the overall flood hazard, and physical and social-economic conditions (Bašić et al., 2018) which, in theory, cannot be predicted using any individual factor. In this study, we consider ten factors that cover the river flood hazard, and physical and social-economic conditions of surrounding areas for regression. We further evaluate the factor importance that contributes to FDS estimation.

b) In a real application it is very difficult to derive the extent of protected areas, particularly for flood defenses without detailed records. This difficulty can typically be simplified using the average value of an explanatory factor of the entire catchment/administration unit for model development, and thereby ignoring the heterogeneity of the FDS at the reach or local level (Scussolini et al., 2016; D. Wang et al., 2021). In this study, we regress the at-site FDS with respect to the explanatory factors of surrounding areas by defining an impact width and develop the estimation model at the grid level. The results of this study reflect the heterogeneity of the FDS between rivers.

c) FDS estimation in most previous studies was assumed to obey one explicit linear relationship with specific explanatory factors. This assumption may not always be appropriate for a large-scale study because the design criteria for flood defenses may change significantly over a study area as a result of different hydrological and social-economic conditions. Recently, machine learning models have demonstrated advantages over ordinary regression models in processing complicated nonlinear problems in flood hydrology (Lange & Sippel, 2020; Mosavi et al., 2018; Shen et al., 2021). In this study, we test a widely used machine learning model, called random forest regression (RFR) (Breiman, 2001; Tyralis et al., 2019), for the first time in FDS estimation using publicly available datasets.

The objective of this study is therefore to develop a robust approach for FDS estimation and incorporate the estimated FDS into large-scale flood hazard modelling. Specifically, we use an RFR to develop the relationship between the observed FDS (overtopping AEP) and ten global coverage explanatory factors at grid level in both the CONUS and England. We compare the proposed RFR with multiple linear regression (MLR) and validate this approach using a 5-fold cross-validation strategy. We then couple the estimated FDS with the Fathom global flood model and test flood hazard mapping in three case studies. We validate the simulated flood hazard maps from the Fathom model either using or not using the proposed approach against official flood hazard maps from regional agencies.

2 Study area and data preparation

2.1 Study area and flood defense data

The study area included the CONUS and England, which both have well documented flood defense data. The flood defense data in the CONUS were obtained from the US National Levee Database or USNLD. The USNLD is an official repository that is maintained and updated by the US Army Corps of Engineers. To date, the USNLD has recorded the location of 9,068 levee systems, approximately 20% of them with detailed levee attributes (e.g., levee height and overtopping AEP). The flood defense data in England were collected from the AIMS Spatial Flood Defenses database. This database includes both natural and man-made flood defenses managed by the UK Environment Agency or a private manager. In this study, only man-made flood defenses in the AIMS Spatial Flood Defenses database were selected for analysis. Natural flood defense structures (e.g., beaches, cliffs, or high land) and tidal defenses were not considered. The distribution of flood defenses in the CONUS and England is presented in Figure 1 (a) and (b), respectively. Details of how to obtain the flood defense data are given at the end of the paper.

Figure 1 shows that there are large differences in the distribution of FDS in the CONUS and England. If only the flood defenses with observed overtopping AEP data are considered, the FDS for most sites in the CONUS was larger than the 100-year return period (overtopping AEP < 0.01), accounting for 78.6% of the total number of levee sites. However, high or highest-standard defenses (overtopping AEP < 0.01) in England only accounted for 32.7% of the total sites, and these defenses were mainly located in London and the East Midlands (downstream of the rivers Nene and Witham). The pie chart in Figure 1 (c) shows that 44.7% of the sites in USNLD and 12.9% of the sites in the AIMS Spatial Flood Defenses database had no observed overtopping AEP data.

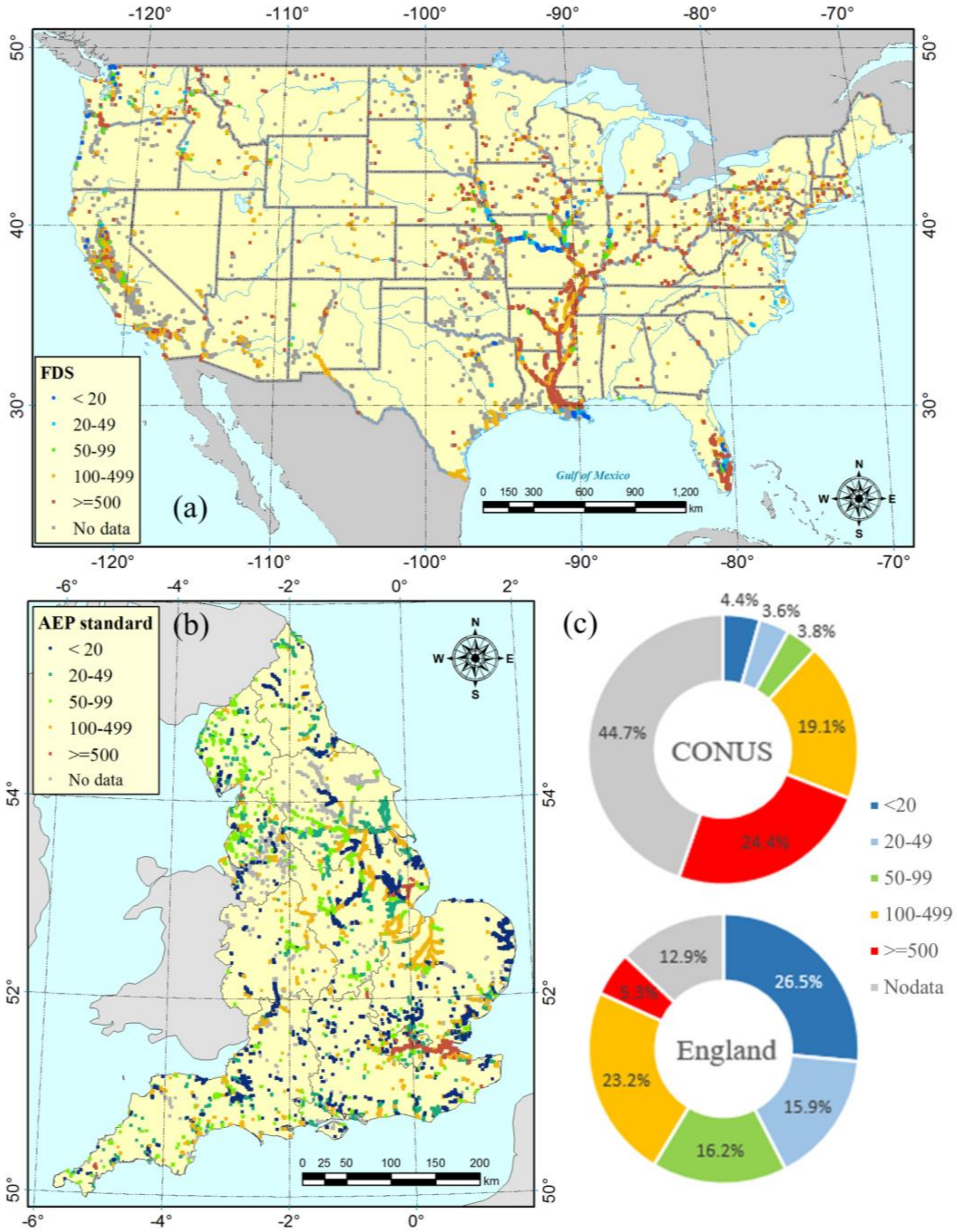

Figure 1 Location and standard of flood defenses (a) in the CONUS and (b) England. Panel (c) shows the proportion of sites in different FDS classes (return period) for the two territories.

2.2 Explanatory factors

From a review of existing publications, ten explanatory factors that reflect the hydrological conditions of river networks and physical-social conditions of the surrounding land areas were selected for this study. These factors were all derived from publicly available datasets with global coverage. The data sources of these factors are presented in Table 1.

Table 1 Explanatory factors used for model development

Aspects	No	Factor name (Abbreviation)	Data source	Data intrinsic resolution
River flood factors	**1**	Catchment area (CA)	MERIT Hydro (Yamazaki et al., 2019)	90 meters
	2	Annual precipitation (AP)	WorldClim (Fick & Hijmans, 2017)	30 seconds
	3	Curve number (CN)	NRCS CN dataset (Zeng et al., 2017)	0.1 degree
	4	Dam capacity (DC)	GRanD V1.3 (B. Lehner et al., 2011)	Points
	5	Bankfull discharge (BD)	Global RFFA (Zhao, Bates, et al., 2021)	30 seconds
Land area factors	**6**	Elevation (EL)	MERIT DEM (Yamazaki et al., 2017)	90 meters
	7	Slope (SL)	MERIT DEM (Yamazaki et al., 2017)	90 meters
	8	Population density (PD)	GPW (Doxsey-Whitfield et al., 2015)	30 seconds
	9	Crop density (CD)	Global Cropland Extent (Pittman et al., 2010)	250 meters
	10	Gross domestic product (GDP)	Gridded global GDP (Kummu et al., 2018)	30 seconds

These explanatory factors can be classified into two categories as follows:

The first category are river flood factors selected to determine the likelihood and magnitude of flood hazards along river networks. The catchment area (CA) reflects the size of upstream catchments and was derived from the flow accumulation map in the MERIT Hydro dataset (Yamazaki et al., 2019). Annual precipitation (AP) describes the average AP of the upstream catchment, and was obtained from the WorldClim V2 dataset (Fick & Hijmans, 2017). The curve number (CN) is an empirical metric that describes the runoff potential for different land uses/land cover, and hydrologic soil group classifications. The CN map was obtained from the study of Zeng et al. (2017)

used global coverage Moderate Resolution Imaging Spectroradiometer (MODIS) land cover and HWSD soil data (FAO et al., 2012) to estimate CN values. In theory, increasing CA, AP, and CN could produce a large increase in runoff, and consequently enlarge the risk of flood hazards. The dam capacity (DC) was calculated by accumulating the maximum reservoir storage capacity of dams in the upstream catchment. The maximum reservoir storage capacity was collected from the GRanD V1.3 dataset (B. Lehner et al., 2011) and this factor has been widely used to evaluate dam attenuation effects on downstream discharge (Volpi et al., 2018; Xiong et al., 2019). Bankfull discharge (BD) describes the channel conveyance by which floodwater just fills the channel without overtopping the banks (Wu et al., 2008). As the real BD is very difficult to observe on a large scale, BD is typically set using a particular return period flood in real applications (Ahilan et al., 2013; Clark et al., 2014). In this study, the 2-year return period flood, which was obtained from regional flood frequency analysis at the global scale (Zhao et al., 2021) was used to represent the BD along river networks.

The second category are land area factors that include the physical and social-economic conditions of surrounding land areas. Physical factors include elevation (EL) and slope (SL), which describe basic terrain characteristics. EL was obtained directly from the MERIT DEM dataset (Yamazaki et al., 2017) and SL was calculated as the maximum rate of change in EL from the grid to its surrounding eight neighbors. Social-economic factors comprised population density (PD), crop density (CD), and Gross Domestic Product (GDP). PD and CD were adopted to describe the density of two main flood exposures, urban areas and farmland, respectively. PD was obtained from the Gridded Population of the World dataset (GWP) (Doxsey-Whitfield et al., 2015) , and represents the average percentage of PD over the past two decades. CD contained a 0–100% cropland probability for each pixel, which was estimated by Pittman et al. (2010) using multi-year MODIS image data. GDP is a widely used metric in FDS estimation that measures the total monetary value of final goods and services in a specific time period (Scussolini et al., 2016). The adopted gridded global GDP product was developed by Kummu et al. (2018), who collected lumped GDP data from regional and national reports and distributed them to each grid cell according to the PD.

3 Methods

The research framework is presented in Figure 2 and can be divided into four parts. Part (1) describes the sample preparation procedure (section 3.1). Part (2) describes the regression model development and comparison procedures (section 3.2). Within this part, the results of RFR and MLR are compared using a 5-fold cross-validation strategy. Part (3) describes the FDS estimation for unlabeled samples in the CONUS and England using the optimal RFR in Part (2). The estimated FDS is incorporated into a large-scale flood hazard model using an enhanced flood defense module (section 3.3). As shown in Part (4), the proposed approach was demonstrated for flood hazard mapping in three case studies by comparing the obtained flood hazard maps with equivalent results from official agencies in section 5. Evaluation metrics for FDS estimation and flood hazard mapping were presented in section 3.4.

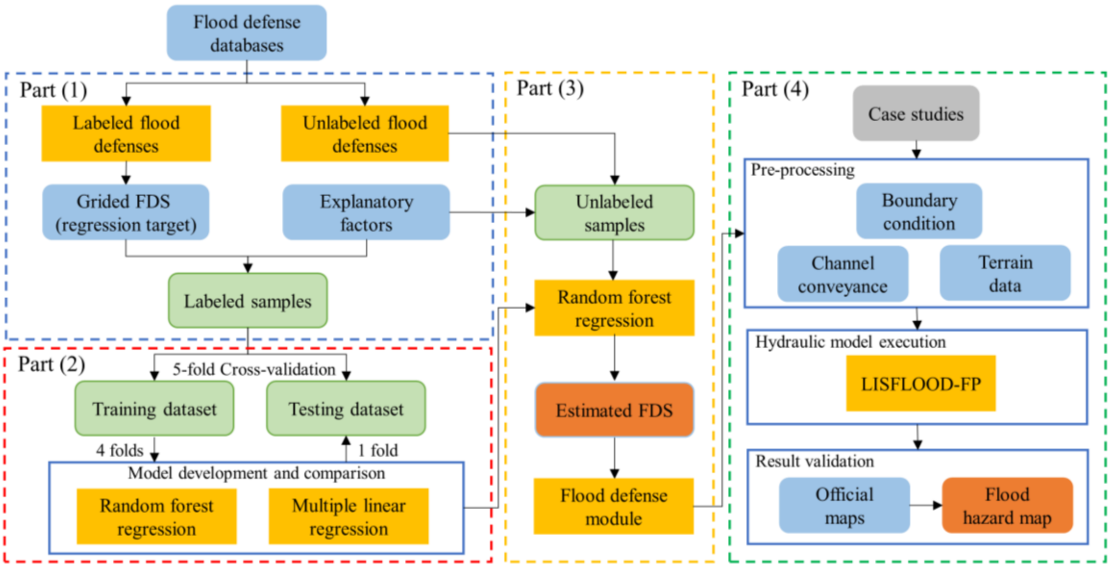

Figure 2 Model framework of this research.

3.1 Sample preparation

The polylines of flood defenses in the CONUS and England were converted into a 1 × 1 km grid layer using ArcGIS software. If there was more than one flood defense in a grid, the longest flood defense in that grid determined the grid value. After conversion, 53,955 and 11,395 grids in the CONUS and England, respectively, were identified as designated sites (sites with flood defenses). Among them, 29,835 grids in the CONUS and 9,921 grids in England had observed FDS data (overtopping AEP), and these grids were selected for model development.

All the explanatory factors were resampled to the same resolution as the FDS layer. Each grid represented the value of that factor within the area of 1 km^2. As the protected areas for some flood defenses were larger than 1 km^2, the factor value within one cell could not accurately describe the relevant physical-social conditions. Therefore, an impact width (*iw*) was defined to address this problem. For each labeled sample, the regression target was the gridded FDS, and the predictors were the explanatory factor conditions of the surrounding $iw \times iw$ km^2. The mean value of explanatory factors of the surrounding $iw \times iw$ km^2 were adopted as predictors. As *iw* increased, the surrounding information considered for model development increased.

3.2 Regression models

Two regression models, RFR and MLR, were adopted for comparison.

(a) Random forest regression

The RFR algorithm can be described using the following three steps:

Step 1: Draw *ntree* subsets from all the training samples using the bootstrapping method (Zhu, 1997), where *ntree* is the number of subsets. The samples that are not selected by the bootstrapping method are called out-of-bag (OOB) samples.

Step 2: Grow the *ntree* regression tree model (Lewis, 2000) using the bootstrapped subsets. For each regression tree, use *mtry* factors randomly for model development to reduce the correlation between the trees. Measure the best split of the tree node using the optimal residual sum of squares (RSS) in eq. (4-1) according the research of Breiman, (2001):

$$RSS = \sum_{i=1}^{n} (\bar{y}_i - y_i)^2, \qquad (4\text{-}1)$$

where $\bar{y}_i$ is the actual value and y_i is the predicted value from the model.

Step 3: Select these two parameters (*ntree* and *mtry*) using a trial method that considers the OOB error changes within the training dataset. The result of the RFR model is the average of the results from the *ntree* regression trees.

RFR was chosen because it can handle categorical and continuous samples, avoids overfitting, and has demonstrated advantages in solving complicated nonlinear problems in hydrology. The factor importance can be evaluated during regression tree development by computing the sum of the reduction of the RSS when a factor is chosen to split a tree node. The larger the average decrease in the RSS of a factor, the more important the factor is to FDS estimation. More detailed information about RFR can be found in the study by Breiman, (2001).

(b) Multiple linear regression

MLR attempts to develop a simple linear relationship between the explanatory factors and observed FDS as follows:

$$y = k_0 + k_1x_1 + k_2x_2 + \cdots + k_Nx_N \pm \epsilon, \quad (4\text{-}2)$$

where y is the regression target, N is the number of explanatory factors, k is the weight of each factor, and ϵ is the error term.

3.3 Fathom global flood hazard model

The Fathom model was used to predict floodplain inundation as can estimate flood hazards anywhere along global river networks. This model consists of four parts: terrain data pre-processing, boundary condition pre-processing, channel bathymetry pre-processing, and hydraulic model execution (Sampson et al., 2015). The terrain data in this model were obtained from a global coverage Merit DEM at 90 m resolution that improved flood hazard mapping by reducing the stripe noise, speckle noise, absolute bias, and tree height bias (Yamazaki et al., 2017). The boundary conditions adopted different return period floods and were derived based on a newly developed regional flood frequency analysis approach (Zhao et al., 2021). Because of the shortage of channel bathymetry data on a large scale, the Fathom model simplified the channel shape as a rectangle whose width and bankfull height controlled the channel conveyance (Neal et al., 2012). In the present study, the river width was obtained from the Merit Hydro datasets and the bankfull height was calculated using a gradually varied flow (GVF) method (Neal et al., 2021) according to the bankfull discharge. The LISFLOOD-FP model (Bates et al., 2010; Neal et al., 2012) designed for large-scale flood modelling was selected for hydraulic model execution. To date, the Fathom model has been successfully applied to high-resolution flood hazard mapping worldwide, including in the CONUS and England (Sampson et al., 2015; Wing et al., 2017).

The Fathom model considers the flood defense in one-dimensional flood routing by linking the channel conveyance with the FDS using flood frequency analysis. For a channel without levees, the bankfull height (H_{bf}) can be estimated using the GVF method by assuming a bankfull discharge return period of approximately 2 years (see

Figure 3 (a)) as suggested by classical geomorphologic theory. As a proof of concept, it is possible to incorporate a flood defense by increasing the channel height according to the FDS. As shown in Figure 3 (b), the new channel height is estimated according to a high FDS (i.e., 100-year return period flood). The levee height (H_{fd}) is then represented by a deeper channel height than that estimated by 2-year return period flood. This strategy avoids the problem of gross mismatches between the discharge and channel conveyance and can represent flood defenses using only the FDS data. However, it is difficult for this strategy to represent the different FDSs between the left and right sides of a channel, and it also ignores the lateral floodwater storage between the channel and levees (Wing et al., 2019). In this study, the flood defense was considered in terrain data by adding the estimated the levee height H_{fd} to the terrain data at levee sites (see Figure 3 (c)). This module can be used for a site knowing location and standard of flood defenses.

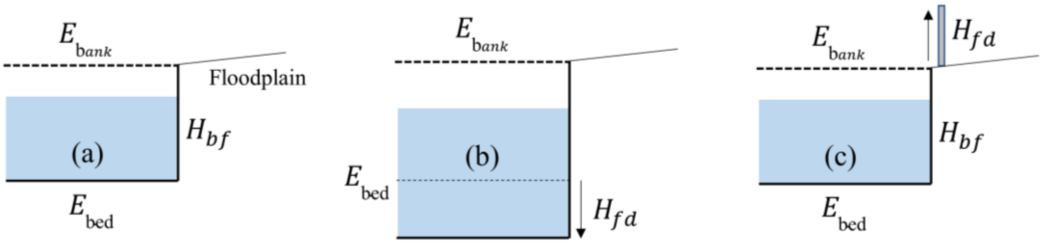

Figure 3 Representing levees in the Fathom model (a) without levees, (b) considering in channel bathymetry, and (c) considering in terrain data (where H_{fd}: levee height; H_{bf}: bank full height; E_{bank}: river bank elevation; E_{bed}: river bed elevation)

3.4 Evaluation metrics

The evaluation procedures focused on two aspects: FDS estimation and flood hazard mapping. For FDS estimation, a 5-fold cross-validation strategy was adopted (Wong & Yeh, 2020). Five models were developed in each study area, and 80% and 20% of the samples were used for training and testing, respectively, in each model. Using this strategy, each sample in the dataset was used for testing once, and model performance was described using the mean value of the evaluation metrics from the five testing datasets. Three evaluation metrics were used for the evaluation of FDS estimation: percent bias (PBIAS), Nash–Sutcliffe efficiency (NSE), and Pearson correlation coefficient (PCC), (see Table 2).

For flood hazard mapping, the simulated results from the Fathom model were compared with the official flood hazard maps from two national agencies for the scenario of a 100-year return period event. In the CONUS, the 100-year floodplain zone from the Federal Emergency Management Agency (FEMA) was adopted as the benchmark map (Bellomo & Ryon, 2010). This FEMA map was collected from simulated flood layers from numerous local models and incorporated the influence of levees in flood hazard modelling. For England, the defended flooded layer from the UK Environment Agency was used. This layer was derived by removing the areas that benefited from defenses from the undefended 100-year flood hazard map created from a similar patchwork of local models as in the US. Two metrics, the Critical Success Index (CSI) and Frequency Bias Index (FBI), were adopted for comparison (see Table 2).

Table 2 Evaluation metrics used for FDS estimation (Nos. 1–3) and flood hazard mapping (Nos. 4–5)

No.	Name	Function	Optimal value
1	Percent bias	$PBIAS = \frac{\sum_{i=1}^{N}(y_i^o - y_i^s)}{\sum_{i=1}^{N}(y_i^o)} \times 100\%$	0%
2	Nash-Sutcliffe efficiency	$NSE = 1 - \frac{\sum_{i=1}^{N}(y_i^o - y_i^s)^2}{\sum_{i=1}^{N}(\overline{y^o} - y_i^s)^2}$	1
3	Pearson correlation coefficient	$PCC = \frac{\sum_{n=1}^{N}(y_i^s - \overline{y^s})(y_i^o - \overline{y^o})}{\sqrt{\sum_{i=1}^{N}(y_i^s - \overline{y^s})^2}\sqrt{\sum_{i=1}^{N}(y_i^o - \overline{y^o})^2}}$	1
4	Critical success index	$CSI = \frac{S_1O_1}{S_1O_1 + S_0O_1 + S_1O_0}$	1
5	Frequency bias index	$FBI = \frac{S_1O_1 + S_1O_0}{S_1O_1 + S_0O_1}$	1

Table note: N is the total number of samples, y_i^o is the observed overtopping AEP for sample i, y_i^s is the simulated overtopping AEP for sample i, $\overline{y^o}$ and $\overline{y^s}$ are the mean values of all observed and simulated overtopping AEPs, respectively, S_1O_1 is the area that is flooded in both the modeled and benchmark maps, S_1O_0 is the area that is flooded in the modeled map but non-flooded in the benchmark map, S_0O_1 is the area that is non-flooded in modeled map but flooded in benchmark map, and S_0O_0 is the area that is non-flooded in both the modeled and benchmark maps.

4 Flood defense standard estimation results

4.1 Model evaluation

First, the influences of *iw* on the cross-validation results of the RFR model were tested. As shown in Figure 4, the cross-validation results of the RFR model showed a similar trend both in the CONUS and England. RFR achieved low performance at the start (*iw* = 1 km), with a mean NSE of 0.1 and 0.0 in the CONUS and England, respectively. The low performance of these 1-*iw* models in the two study areas demonstrated that factors that only described the surrounding 1 km^2 area were insufficient for model development. This is mainly because 1 km^2 could not cover the protected areas for large levees. Model performance became stable when iw was larger than 20 km. The optimal iw was therefore selected as 20 according to the highest mean NSE among all testing results. For this value, the predictors reflected the average condition of the explanatory factors around a distance of 10 km. The range of NSE in Figure 4 reflects the generalization ability in terms of different model inputs. The NSE of the 20-iw models in 5-folds cross-validation ranged from 0.76 to 0.84, and 0.70 to 0.74 respectively in the CONUS and England, which demonstrated that the RFR model provided stable results for unseen data.

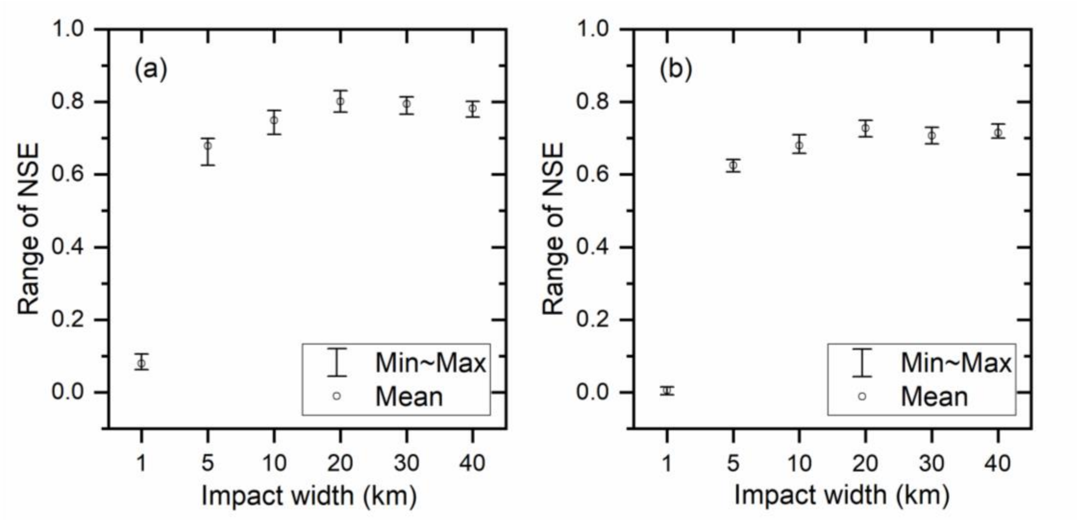

Figure 4 Five-fold cross-validation performance under different impact widths (a: CONUS; b: England).

The cross-validation results of the two regression models (RFR and MLR) were further compared using the optimal *iw*; the results are presented in Figure 5. From the evaluation metrics, RFR achieved better performance than MLR, with an average NSE in the 5-fold validation of 0.82 in the CONUS and 0.73 in England. The MLR resulted in a negative NSE for both test sites, which demonstrated that the relationship between FDS and the explanatory factors did not obey a linear function. The PCC and PBIAS metrics gave similar results to NSE, with both suggesting that RFR is a reliable approach for FDS estimation. From a comparison of the two study areas, the FDS was better estimated in the CONUS than England. This was mainly because the FDS data in the CONUS were more consistent than those in England. The FDS in the CONUS only recorded artificial levee systems along rivers. However, the FDS data in England included multiple levee structures designed for both fluvial and coastal floods which typically have different defense standards over this territory. The positive PBIAS for RFR indicated that the average tendency of the simulated FDS was smaller than the observed FDS. Overall underestimations of 14% and 21% were found in the CONUS and England, respectively.

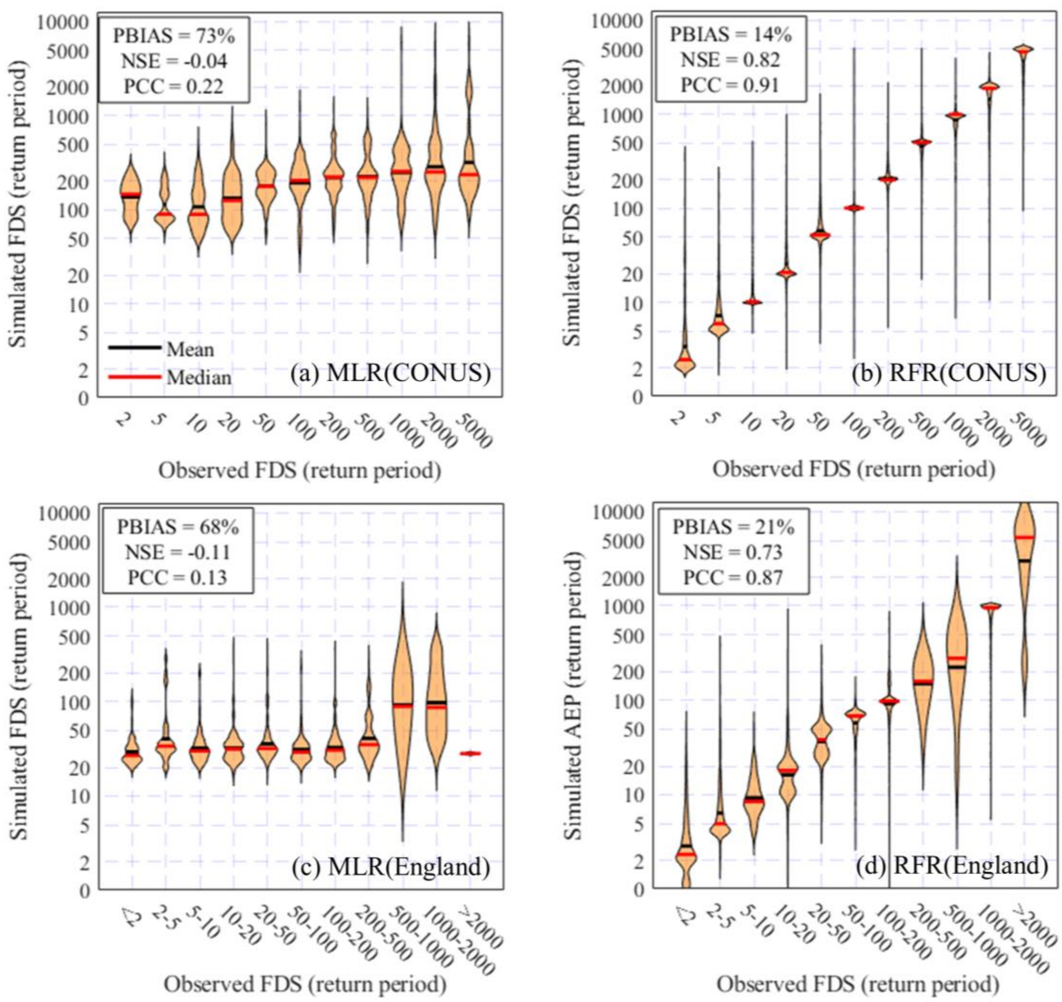

Figure 5 Violin plot of the simulated FDS for each observed FDS magnitude ((a) MLR in the CONUS, (b) RFR in the CONUS, (c) MLR in England, and (d) RFR in England).

The violin plot in Figure 5 describes the probability density of the simulated FDS in each observed FDS bin. As shown in Figure 5 (a) and (c), the mean value of the simulated FDS on the Y-axis remained stable as the observed FDS increased on the X-axis. This demonstrated that MLR did not have prediction ability for every FDS magnitude. Figure 5 (b) and (c) show that the mean value of the simulated FDS increased as that of the observed FDS magnitude increased. The variation of the probability density reflected the range of model errors for each FDS magnitude. Although some large error sites were found for each FDS magnitude, most FDSs both in the CONUS and England were correctly simulated and the simulated FDSs concentrated along the corresponding bin value on the X-axis. RFR slightly overestimated the FDS at low return periods both in the CONUS and England, and largely underestimated the FDS at the 200-1000 year return period in England.

4.2 Factor importance

Figure 6 (a) and (b) show the factor importance for FDS estimation in the CONUS and England, respectively. Although regression targets were collected from different databases, AP, CA, and BD were the top three important factors for FDS estimation at both sites. These factors determined the basic hydrological conditions of the upstream catchment directly affecting the risk of flooding. This high ranking is easy to understand because the aim of levees is to reduce the regional flood hazard, and some levee parameters are also designed based on these hydrological factors. Due to the limited resources for flood risk management, a cost-benefit analysis (CBA) is typically required in deciding the FDS considering the trade-off between the costs over the appraisal period and socio-economic benefits in the protected area (Hallegatte, 2006; Hudson & Botzen, 2019; Ward et al., 2017). Socio-economic factors are commonly applied in CBA and are regarded as key factors in deciding FDS investments in several studies (Fadel et al., 2018; Hudson & Botzen, 2019; Scussolini et al., 2016; Ward et al., 2017). However, three socio-economic factors (i.e., GDP, PD, and CD) had the lowest importance among all factors in the CONUS (Fig. 6(a)). This result is similar to that of Wing et al. (2019), who also did not identify connections between the FDS and socio-economic variables in the CONUS. Figure 1 (c) shows that the medium, low, or lowest samples only accounted for less than 5% of the total number of samples in the CONUS. As the factor importance was evaluated based on learning targets, these limited low-standard samples in the CONUS may make it difficult to provide a fair evaluation of factor importance. By contrast, GDP and PD played a more important role in England when the FDS data were more representative for each FDS magnitude (Figure 6 (b)). Both results in the CONUS and England demonstrated that CD had the lowest importance among all socio-economic factors. This mainly because croplands are typically assigned a lower weight than urban areas in any CBA.

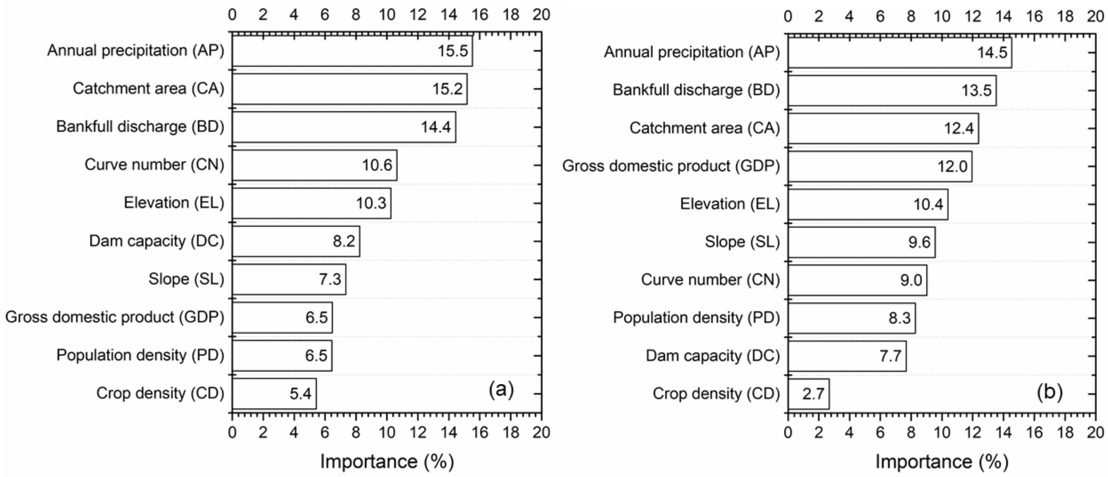

Figure 6 Factor importance evaluated using RFR in (a) the CONUS and (b) England.

Table 3 Validation accuracy of RFR using different aspects of factors

Aspects	**CONUS**			**England**		
	PBIAS	NSE	PCC	PBIAS	NSE	PCC
All factors	14%	0.82	0.91	21%	0.73	0.87
River flood factors	24%	0.72	0.88	38%	0.52	0.77
Physical factors	27%	0.58	0.77	37%	0.48	0.73
Social-economic factors	51%	0.44	0.73	63%	-0.01	0.41

Table 3 presents the cross-validation results of the RFR model using different aspects of factors. RFR provided the highest performance when all factors were used for model development both in the CONUS and England. RFR achieved satisfactory performance (NSE = 0.72) in the CONUS, but low performance (NSE < 0.60) in England when only river flood factors were used for regression. RFR achieved low performance when either physical or social-economic factors only were used for FDS estimation. This demonstrates that social-economic factors alone are inadequate for FDS estimation in the CONUS and England, which conflicts with the results of some studies (Aerts et al., 2020; Herold & Rudari, 2013; Nicholls, 2002; Quinn et al., 2019). Combining the factor importance shown in Figure 6 and the accuracy shown in Table 3, river flood factors were more important than physical factors and social-economic factors in FDS estimation.

4.3 Levee standard estimation

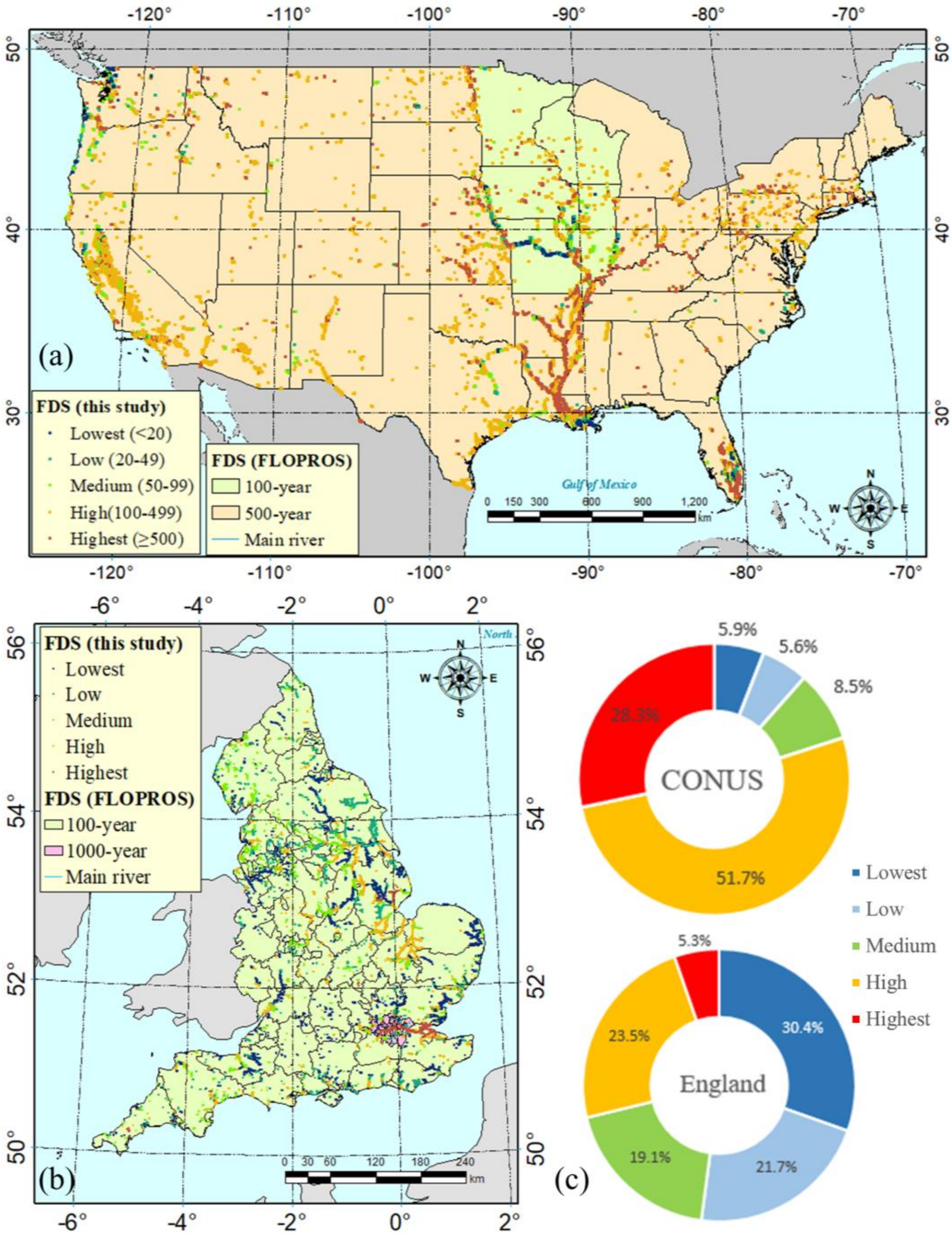

Figure 7 FDS from this study and FLOPROS in (a) the CONUS and (b) England, and (c) proportion of FDS in different classes.

The FDS was estimated for all designated sites using the optimal RFR in the two study areas (Figure 7 (a) and (b)) and the results were compared with the FDS data from the merged layer of FLOPROS. The FDS was divided into five classes (lowest: < 20 year return period, low: 20–49 year, medium: 50–99 year, high: 100–499 year, and highest: >= 500 year). The percentage of samples in each class is shown in Figure 7 (c). As the FDS in FLOPROS was processed based on administrative units (state in the CONUS and county in England), the heterogeneity of FDS cannot be reflected accurately. All administrative units were identified as high or highest classes (FDS >= 100 year) in the FLOPROS dataset, and this underestimated the flood hazard for reaches with low FDS. By contrast, in this study, FDS was estimated at 1 km resolution and the heterogeneity of FDS between rivers was considered. Only a small proportion of samples in the CONUS were estimated as the low or lowest levels, and these levees were mainly concentrated in the State of Missouri. In England, the highest-standard levees were mainly located around the city of London, and were also identified as the highest standard in FLOPROS and the research of Hall et al., (2019).

The average FDS was calculated for each hydrological unit. The results are shown in Figure S1 in Supplementary information 1. Figure S-1 (a) shows that the FDS for most hydrological units in the CONUS was larger than that of 100-year return period, accounting for approximately 64% of the total study area. Only 5.9% samples were identified as the lowest standard in the CONUS, but the areas of hydrological units identified as the lowest standard accounted for approximately 30% of the total areas. This is mainly because some hydrological units without observed flood defenses were also classified as the lowest level. As shown in Figure S1(b), the FDS for most hydrological units was lower than 100-year return period in England, accounting for 85% of the total area.

The river size and economic condition were further analyzed for hydrological units in different FDS classes. The river size and economic condition were reflected by the maximum value of flow accumulation (Max FA) and sum value of GDP (Sum GDP) in hydrological units, respectively. Figure 8 (a) and (b) describe the range of max FA in the CONUS and England, respectively. The median and mean values of max FA between the low and lowest return period levels were difficult to distinguish visually, and these two metrics gradually increased as the FDS increased from the medium to highest standard. Similar trends are also shown in Figure 8 (c) and (d), where the mean value of Sum GDP in the high and highest classes are clearly larger than that in the low and lower classes. This agrees with our experience that high flood hazard and exposure areas typically have correspondingly large FDS. However, a wide range of Max FA and Sum GDP were also found in each class. This means that this identified trend was not valid for a certain number of hydrological units, and reliable FDS estimation could not be provided for all hydrological units only considering river size or economic condition.

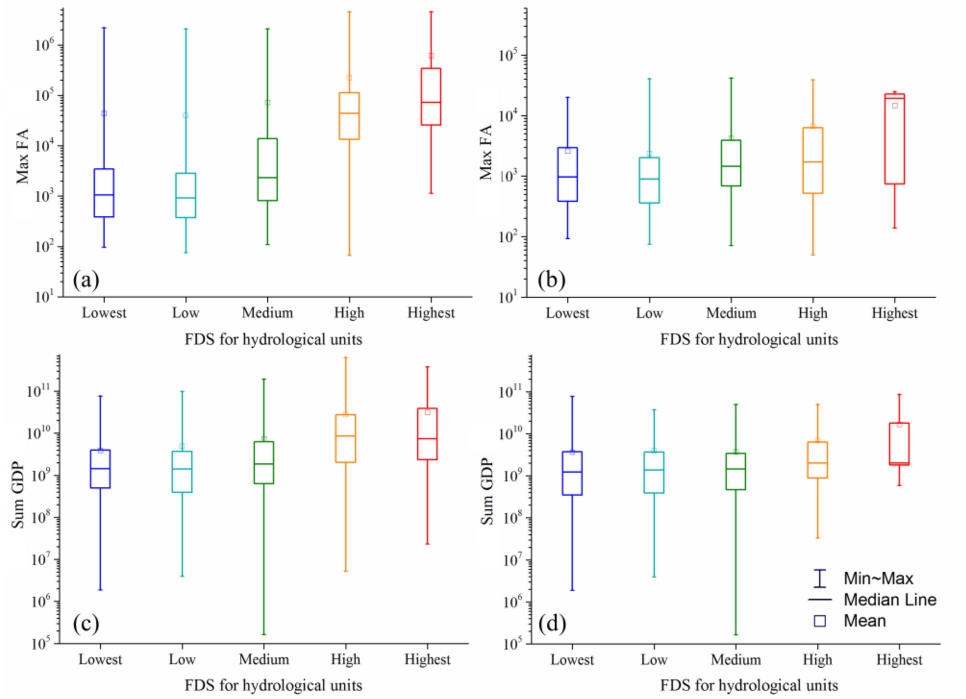

Figure 8 Max FA for hydrological units for different FDS classes in (a) the CONUS and (b) England. Sum GDP for hydrological units for different FDS classes in (c) the CONUS and (d) England.

5 Flood hazard modelling

5.1 Case study descriptions

Three case studies were selected to test the representation of flood defenses in flood hazard mapping using the proposed approach. As the flood hazard was simulated based on the scenario of a 100-year return period, the proposed approach was tested to determine whether it could correctly identify the FDS of levees exceeding 100-year return period.

Figure 9 (a) shows that case study 1 (C1) was located in the upper Mississippi river, and this area involved eight levee systems. The Sny Island levees are mainly distributed along the east side of the river, and were correctly simulated as being of high or highest standards (>100-year) when the proposed approach was used. Although the protected area of the Sny Island levees is rural, it prevents at least $51.1 million in flood damages each year. Other medium or low-standard levees were mainly distributed downstream of reach A, which were also correctly simulated.

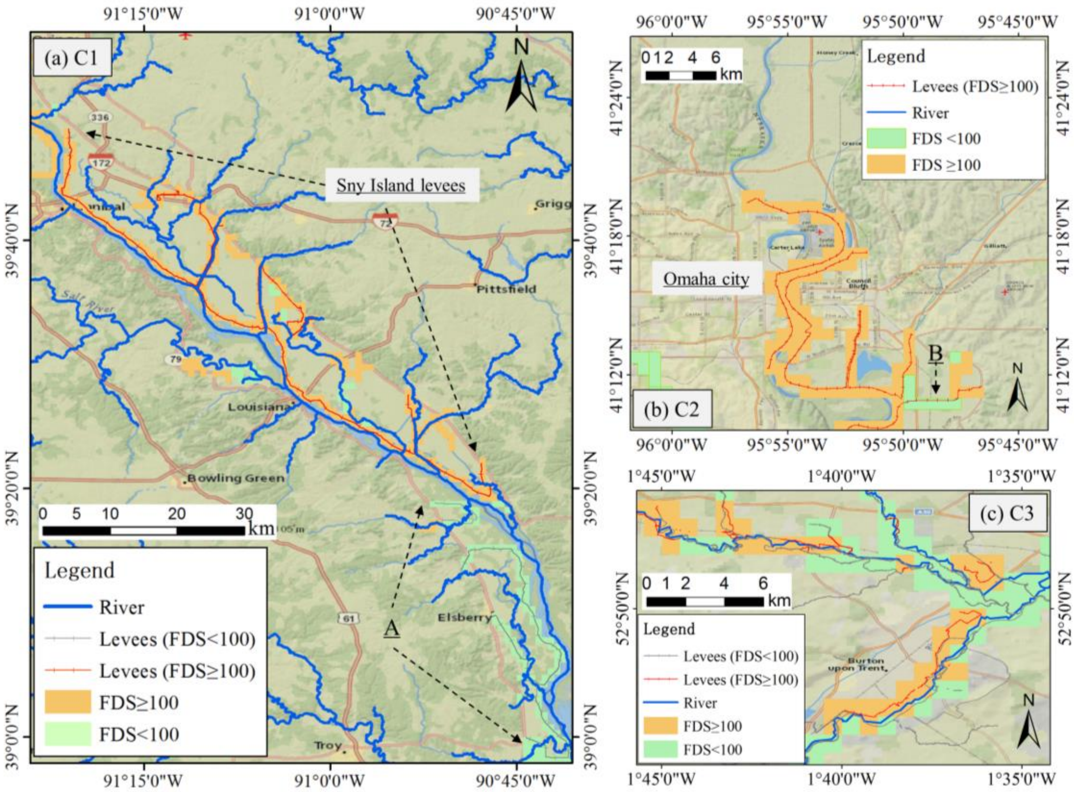

Figure 9 Observed and simulated FDSs in the three case studies: (a) case study 1 (in the CONUS), (b) case study 2 (in the CONUS), and (c) case study 3 (in England).

Unlike C1, levee systems in case study 2 (C2) were located in the middle Missouri River and were built to protect Omaha (the largest city in the U.S. state of Nebraska). Because of the high flood exposure in Omaha city, all levees in C2 were designed to a high standard. Figure 9 (b) shows that most levees in C2 were correctly identified, except for the levees along reach B. Case study 3 (C3) was located at the confluence zone of the River Trent and River Dove in England (Figure 9 (c)).

Compared with the rivers in C1 and C2, River Trent and River Dove are smaller river systems and some levee systems in C3 were located in the same grid. The heterogeneity of the FDS of some levees in C3 was therefore difficult to represent at 1 km resolution, and thereby induced both over and underestimations. The high-standard levees in C3 were mainly located to protect urban areas, such as the town of Burton and the village of Hatton, and were still correctly identified using the proposed method.

5.2 Validation results

Figure 10 shows the validation results in the three case studies (C1, C2, and C3), considering or not considering flood defenses. Overall, the CSI ranged from 0.47 to 0.51 for the three case studies where flood defenses were not taken into account. The global method thus achieved CSI values similar to those of previous large-scale modelling results in the study areas (Sampson et al., 2015; Wing et al., 2017). Mapping performance clearly improved when the proposed approach was used, and the CSI ranged from 0.62 to 0.75 in the three case studies.

Figure 10 (C1a) and (C1b) show a comparison of the flood hazard map in C1. As shown in Figure 10 (C1b), the protected areas of the Sny Island levees were almost completely flooded when the levee effects were not considered. The overestimation was corrected (Figure 10 (C1a)) and the mapping performance as determined by the CSI metric improved from 0.47 to 0.75 when the proposed method was used. Some underestimation was found in small reaches in Figure 10 (C1a) and (C1b) because only flooding of the main reaches was simulated.

Figure 10 (C2a) and (C2b) show a comparison of the flood hazard maps in C2. The CSI increased from 0.51 to 0.69 when levee effects were considered, and this improvement was mainly around the urban area of Omaha. The FBI reduced from 1.23 to 1.06, which means that the overestimation was corrected when the proposed approach was used. An obvious overestimation was still found downstream of the reach in C2. This is mainly because the FDS in the area of B was incorrectly estimated by the proposed method, and therefore caused the overestimation of the flood hazard.

Figure 10 (C3a) and (C3b) show a comparison of the flood hazard maps for case study C3. As shown in Figure 10 (C3a), two protected areas in C3 were both flooded when the levee effects were not considered. Although some levees could not be reflected at 1 km resolution, the two defended areas in C3 were still represented (see Figure 10 (C1a)) and a satisfactory CSI of 0.62 was achieved. After the improvement, the flood hazard model still showed some errors compared with the benchmark flood maps. This is mainly because, in this study, the flood hazard was mapped based on open access flood and terrain data with global coverage whereas the local models that underpin the official maps use high resolution airborne LiDAR terrain information and river gauge information. Although several efforts were made to reduce data errors, it remained difficult to achieve similar results to local and regional models that adopt observed flood and high-resolution terrain data.

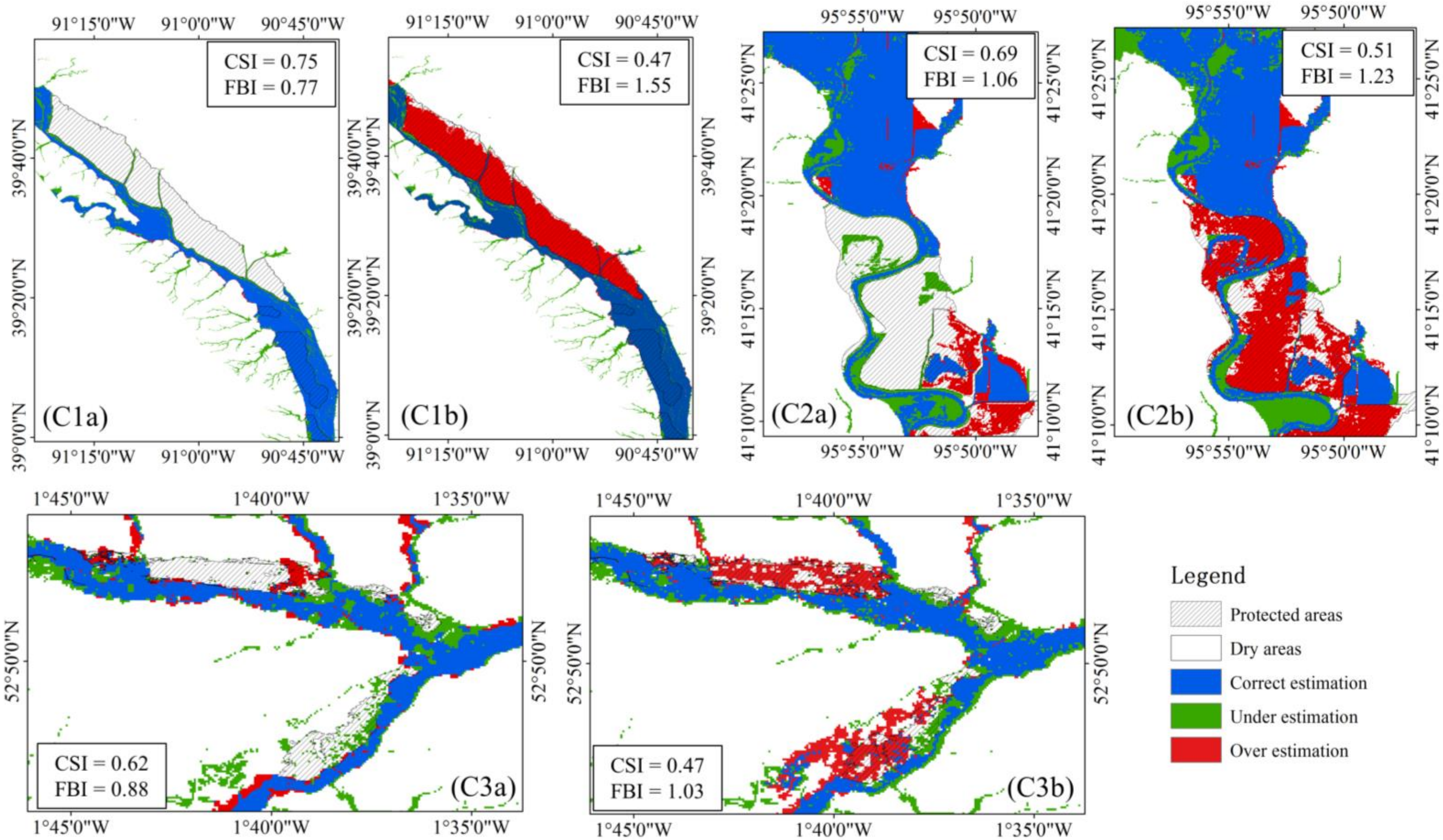

Figure 10 Validation of simulated results from Fathom model against with official flood hazard maps (a: considering and b: not considering the flood defenses in C1, C2, and C3).

6 Discussion and future work

The impact of sample selection on model performance was analysed by considering some scenarios. Scenario 1 (S1) manually reduced the size of training samples from 100%, to 75%, 25% and 10% respectively, and left the validation samples unchanged. As shown in Figure 11 (a), model performance reduced as the training sample size decreased both in the CONUS and England. The model provided satisfactory results (NSE > 0.7) when only a small number of samples were removed. However, the RFR model achieved low performance in the two study areas when a large number of samples were removed. This demonstrated that model performance was highly dependent on the representativeness of training samples, and the proposed method should not be applied to a study area that does not have a large amount of flood defense data.

As shown in the Figure 11 (a), the learning samples in the CONUS were unbalanced in terms of spatial distribution, and the highest-standard levees were located mainly along the reaches of the Mississippi River. In scenario 2 (S2), the levee samples in the Mississippi River were eliminated and a new RFR was developed using the retained samples. The model in S2 obtained good results (NSE = 0.81) similar to those of the model trained using all the samples. Figure 11 (b) shows the factor importance evaluated using the new RFR. Although some rankings changed slightly compared with the factor importance in Figure 6(a), the river flood factors still had higher importance than the physical and socio-economic factors. This demonstrated that the unbalanced FDS samples in the CONUS did not significantly impact the RFR results.

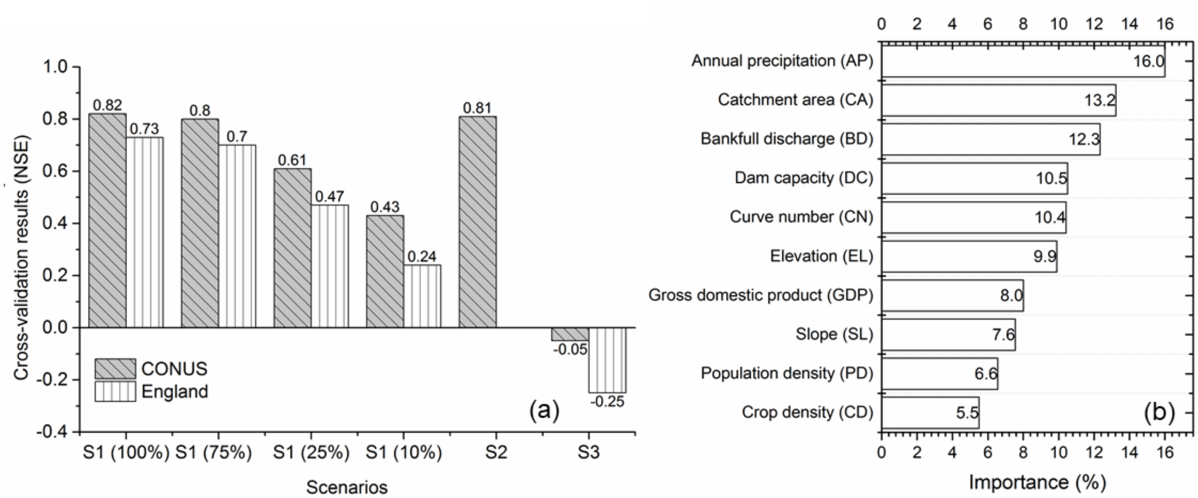

Figure 11 (a) Cross-validation results of RFR in different scenarios and (b) factor importance evaluated using RFR in scenario 2 (S2).

Some deficiencies were found during the study and could be improved in future work. Firstly, in scenario 3 (S3), RFR developed in the CONUS was used in England, and vice versa. As shown in Figure 11 (a), low performance in S3 (NSE < 0) demonstrated that a model developed in one study area cannot easily be applied to another. This result was expected because there were large differences in the FDS characteristics between the CONUS and England, and machine learning typically achieves low prediction performance outside the training domain. Recently, the transfer learning technique has been introduced into hydrology to improve the extrapolation ability of machine learning models (Ma et al., 2021; Zhao, Pang, et al., 2021). This provides a potential solution to the application of the proposed approach in data-sparse areas. Secondly, the predictors were derived using the mean value of the image-based explanatory factors, and this calculation lost the topological information of images. Meanwhile, the width of images (impact width) in theory should vary in space in terms of different river sizes. During the experiments, RFR was compared with some ordinary machine learning models, such as the support vector machine method, and RFR achieved the best results. However, it is difficult to handle image-based inputs for traditional machine learning models. In some studies, advances in machine learning techniques, such as convolutional neural networks, have shown advantages when considering unstructured image inputs (Shen et al., 2021; Zhang et al., 2020). In future work, deep learning techniques will be used to further improve the estimation results considering the variability of impact width in space. Finally, the flood defenses in this study focused on artificial levee systems. Other flood protection structures, such as dams or diversion canals, were not considered, and this led to an underestimation of the FDS for some hydrological units which are also protected by other flood defenses. In previous studies, the impact of dams on extreme flow and inundation simulation in national-scale flood modelling were considered (Zhao et al., 2020). As a next step, FDS estimation will be improved by considering multiple types of flood protection structures and attempting to incorporate compound defenses in flood hazard modelling.

7 Conclusions

In this study, we proposed an RFR-based approach to estimate FDS using publicly available datasets. We compared RFR with MLR and demonstrated this approach in the CONUS and England, respectively. We incorporated the results of the proposed approach into hydraulic modelling and improved the representation of flood defenses in large-scale flood hazard mapping.

The main conclusions are summarized as follows:

1. The RFR-based approach successfully estimated the FDS (i.e., overtopping AEP) using the explanatory factors contained in publicly available datasets. The RFR results are sensitive to the impact width (iw) parameter sensitive. RFR achieved low performance (NSE < 0.1) when iw was set to 1 km and achieved good performance (NSE > 0.7) when iw was larger than 20 km.

2. RFR performed better than MLR, with an NSE of 0.82 in the CONUS and 0.73 in England. The MLR achieved a negative NSE, which demonstrates that the FDS and explanatory factors do not obey a simple linear relationship. RFR overestimated the FDS at the low return periods both in the CONUS and England, and largely underestimated the FDS in the 200–1000 return period in England.

3. The factor importance for FDS estimation in the CONUS and England was evaluated using RFR. River flood factors (annual precipitation, catchment area, and bankfull discharge) had higher importance than physical and social-economic factors both in the CONUS and England. In the CONUS, three socio-economic factors (gross domestic product, population density, and crop density) had the lowest importance among all factors. By contrast, gross domestic product and population density played more important roles in England.

4. The FDS for all ungauged sites and hydrological units in the CONUS and England were estimated based on the proposed approach. Compared with the results of FLOPROS, in this study, the heterogeneity of FDS between rivers was considered. The FDSs of most hydrological units in the CONUS were larger than that of the 100-year return period. However, most hydrological units were lower than the 100-year return period in England.

5. The results of this study were incorporated into large-scale flood hazard mapping and the mapping results were validated in three case studies. The CSI ranged between 0.47 and 0.51 in the three case studies, without considering flood defense effects. The function of flood defenses was successfully simulated using the proposed approach, with an improved CSI ranging from 0.62 to 0.75 in the three case studies.

Data and Software

The data source of all explanatory factors is described in Table 1. The US National Levee Database is available from https://levees.sec.usace.army.mil/. Benchmark flood hazard maps in the CONUS are available from the FEMA Flood Map Service Center (https://msc.fema.gov/portal/home). The AIMS Spatial Flood Defenses Database and benchmark flood hazard maps in the England can be downloaded from https://data.gov.uk/. The Fathom global flood hazard model is available for academic research use by contacting info@fathom.global. The Random Forest regression is implemented based on the package 'randomForest' under R software environment.

Acknowledgments

Paul Bates' work on this paper was supported by a Royal Society Wolfson Research Merit award and UK Natural Environment Research Council grant NE/V017756/1. Jeff Neal is supported by UK Natural Environment Research Council grants (NE/S003061/1 and NE/S006079/1).

References

Aerts, J. P. M., Uhlemann-Elmer, S., Eilander, D., & Ward, P. J. (2020). Comparison of estimates of global flood models for flood hazard and exposed gross domestic product: a China case study. *Natural Hazards and Earth System Sciences*, *20*(12), 3245–3260. https://doi.org/10/gjsknr

Ahilan, S., O'Sullivan, J. J., Bruen, M., Brauders, N., & Healy, D. (2013). Bankfull discharge and recurrence intervals in Irish rivers. *Proceedings of the Institution of Civil Engineers - Water Management*, *166*(7), 381–393. https://doi.org/10.1680/wama.11.00078

Bašić, K., Gilja, G., Bujak, D., & Kuspilić, N. (2018). Design Criteria for Levees. Presented at the Conference: 10th Eastern European Young Water Professionals Conference, Zagreb, Croati.

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

Bates, P. D., Neal, J., Sampson, C., Smith, A., & Trigg, M. (2018). Progress Toward Hyperresolution Models of Global Flood Hazard. In *Risk Modeling for Hazards and Disasters* (pp. 211–232). Elsevier. https://doi.org/10.1016/B978-0-12-804071-3.00009-4

Bellomo, D., & Ryon, A. (2010). FEMA Flood Map Modernization Program. In *World Environmental and Water Resources Congress 2010* (pp. 2223–2230). Providence, Rhode Island, United States: American Society of Civil Engineers. https://doi.org/10.1061/41114(371)229

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10/d8zjwq

Choung, Y. (2014). Mapping Levees Using LiDAR Data and Multispectral Orthoimages in the Nakdong River Basins, South Korea. *Remote Sensing*, *6*(9), 8696–8717.

https://doi.org/10.3390/rs6098696

Clark, R. A., Gourley, J. J., Flamig, Z. L., Hong, Y., & Clark, E. (2014). CONUS-Wide Evaluation of National Weather Service Flash Flood Guidance Products. *Weather and Forecasting*, *29*(2), 377–392. https://doi.org/10.1175/WAF-D-12-00124.1

CRED, & UNDRR. (n.d.). The Human Cost of Disasters - An overview of the last 20 years 2000-2019 - World. Retrieved February 21, 2022, from https://reliefweb.int/report/world/human-cost-disasters-overview-last-20-years-2000-2019

Di Baldassarre, G., Schumann, G., & 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. https://doi.org/10/c645f7

Doxsey-Whitfield, E., MacManus, K., Adamo, S. B., Pistolesi, L., Squires, J., Borkovska, O., & Baptista, S. R. (2015). Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Papers in Applied Geography*, *1*(3), 226–234. https://doi.org/10.1080/23754931.2015.1014272

Environment Agency. (2021, December 10). AIMS Spatial Flood Defences (inc. standardised attributes). Retrieved February 21, 2022, from https://data.gov.uk/dataset/cc76738e-fc17-49f9-a216-977c61858dda/aims-spatial-flood-defences-inc-standardised-attributes

Fadel, A. W., Marques, G. F., Goldenfum, J. A., Medellín-Azuara, J., & Tilmant, A. (2018). Full Flood Cost: Insights from a Risk Analysis Perspective. *Journal of Environmental Engineering*, *144*(9), 04018071. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001414

FAO, IIASA, ISRIC, ISSCAS, & JRC. (2012). Harmonized World Soil Database - HWSD (version 1.2) [Data set]. International Institute for Applied Systems Analysis (IIASA). Retrieved from http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/index.html?sb=1

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Hall, J. W., Harvey, H., & Manning, L. J. (2019). Adaptation thresholds and pathways for tidal flood risk management in London. *Climate Risk Management*, *24*, 42–58. https://doi.org/10.1016/j.crm.2019.04.001

Hallegatte, S. (2006). A Cost-Benefit Analysis of the New Orleans Flood Protection System. *AEI-Brookings Joint Center. Regulatory Analysis*, 06–02.

Herold, C., & Rudari, R. (2013). Improvement of the Global Flood Model for the GAR 2013 and 2015. *United Nations Office for Disaster Risk Reduction (UNISDR): Geneva, Switzerland*.

Hudson, P., & Botzen, W. J. W. (2019). Cost–benefit analysis of flood-zoning policies: A review of current practice. *WIREs Water*, *6*(6), e1387. https://doi.org/10.1002/wat2.1387

Kummu, M., Taka, M., & Guillaume, J. H. A. (2018). Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Scientific Data*, *5*(1), 180004. https://doi.org/10.1038/sdata.2018.4

Lange, H., & Sippel, S. (2020). Machine Learning Applications in Hydrology. In D. F. Levia, D.

E. Carlyle-Moses, S. Iida, B. Michalzik, K. Nanko, & A. Tischer (Eds.), *Forest-Water Interactions* (pp. 233–257). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-26086-6_10

Lehner, B., ReidyLiermann, C., Revenga, C., Vorosmarty, C., Fekete, B., Crouzet, P., et al. (2011). Global Reservoir and Dam Database, Version 1 (GRanDv1): Dams, Revision 01 [Data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4N877QK

Lehner, Bernhard, & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10/f5czg3

Leonard, D. (2008). Raising the Levee: Dutch Land Use Law as a Model for U.S. Adaptation to Climate Change. *Georgetown International Environmental Law Review*, *21*, 543.

Lewis, R. J. (2000). An introduction to classification and regression tree (CART) analysis. In *Annual meeting of the society for academic emergency medicine in San Francisco, California* (Vol. 14). Citeseer.

Ludy, J., & Kondolf, G. M. (2012). Flood risk perception in lands "protected" by 100-year levees. *Natural Hazards*, *61*(2), 829–842. https://doi.org/10.1007/s11069-011-0072-6

Ma, K., Feng, D., Lawson, K., Tsai, W., Liang, C., Huang, X., et al. (2021). Transferring Hydrologic Data Across Continents – Leveraging Data-Rich Regions to Improve Hydrologic Prediction in Data-Sparse Regions. *Water Resources Research*, *57*(5). https://doi.org/10/gpczf6

Maguya, A., Junttila, V., & Kauranne, T. (2014). Algorithm for Extracting Digital Terrain Models under Forest Canopy from Airborne LiDAR Data. *Remote Sensing*, *6*(7), 6524–6548. https://doi.org/10.3390/rs6076524

Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, *10*(11), 1536. https://doi.org/10/gfr2nb

Neal, J., Schumann, G., & Bates, P. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resources Research*, *48*(11). https://doi.org/10/ggj7sm

Neal, J., Hawker, L., Savage, J., Durand, M., Bates, P., & Sampson, C. (2021). Estimating River Channel Bathymetry in Large Scale Flood Inundation Models. *Water Resources Research*, *57*(5). https://doi.org/10/gpbg23

Nicholls, R. J. (2002). Analysis of global impacts of sea-level rise: a case study of flooding. *Physics and Chemistry of the Earth, Parts A/B/C*, *27*(32), 1455–1466. https://doi.org/10.1016/S1474-7065(02)00090-6

O'Dell, J., Nienhuis, J. H., Cox, J. R., Edmonds, D. A., & Scussolini, P. (2021). *A global database of flood-protection levees on river deltas (openDELvE)* (other). pico. https://doi.org/10.5194/egusphere-egu21-13180

Özer, I. E., Leijen, F. J., Jonkman, S. N., & Hanssen, R. F. (2019). Applicability of satellite radar imaging to monitor the conditions of levees. *Journal of Flood Risk Management*, *12*(S2). https://doi.org/10.1111/jfr3.12509

Pittman, K., Hansen, M. C., Becker-Reshef, I., Potapov, P. V., & Justice, C. O. (2010). Estimating Global Cropland Extent with Multi-year MODIS Data. *Remote Sensing*, *2*(7), 1844–1863. https://doi.org/10.3390/rs2071844

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., et al. (2019). The Spatial Dependence of Flood Hazard and Risk in the United States. *Water Resources Research*, *55*(3), 1890–1911. https://doi.org/10.1029/2018WR024205

Rudari, R., & Silvestro, F. (n.d.). Input Paper prepared for the Global Assessment Report on Disaster Risk Reduction 2015, 69.

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., & Freer, J. E. (2015). A high-resolution global flood hazard model. *Water Resources Research*, *51*(9), 7358–7381. https://doi.org/10/ggjfth

Scussolini, P., Aerts, J. C. J. H., Jongman, B., Bouwer, L. M., Winsemius, H. C., de Moel, H., & Ward, P. J. (2016). FLOPROS: an evolving global database of flood protection standards. *Natural Hazards and Earth System Sciences*, *16*(5), 1049–1061. https://doi.org/10.5194/nhess-16-1049-2016

Shen, C., Chen, X., & Laloy, E. (2021). Editorial: Broadening the Use of Machine Learning in Hydrology. *Frontiers in Water*, *3*, 38. https://doi.org/10/gjzfnz

Smith, A., Sampson, C., & Bates, P. (2015). Regional flood frequency analysis at the global scale. *Water Resources Research*, *51*(1), 539–553. https://doi.org/10/f63rs4

Sofia, G., Tarolli, P., Cazorzi, F., & Dalla Fontana, G. (2011). An objective approach for feature extraction: distribution analysis and statistical descriptors for scale choice and channel network identification. *Hydrology and Earth System Sciences*, *15*(5), 1387–1402. https://doi.org/10.5194/hess-15-1387-2011

Steinfeld, C. M. M., Kingsford, R. T., & Laffan, S. W. (2013). Semi-automated GIS techniques for detecting floodplain earthworks: DETECTING FLOODPLAIN EARTHWORKS. *Hydrological Processes*, *27*(4), 579–591. https://doi.org/10.1002/hyp.9244

Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water*, *11*(5), 910. https://doi.org/10/ghwnkc

Ubilla, J., Abdoun, T., Sasanakul, I., Sharp, M., Steedman, S., Vanadit-Ellis, W., & Zimmie, T. (2008). New Orleans Levee System Performance during Hurricane Katrina: London Avenue and Orleans Canal South. *Journal of Geotechnical and Geoenvironmental Engineering*, *134*(5), 668–680. https://doi.org/10.1061/(ASCE)1090-0241(2008)134:5(668)

USACE. (n.d.). Civil Works Levee Safety Program. Retrieved February 21, 2022, from https://www.usace.army.mil/Missions/Civil-Works/Levee-Safety-Program/

Volpi, E., Di Lazzaro, M., Bertola, M., Viglione, A., & Fiori, A. (2018). Reservoir Effects on Flood Peak Discharge at the Catchment Scale. *Water Resources Research*, *54*(11), 9623–9636. https://doi.org/10.1029/2018WR023866

Wang, D., Scussolini, P., & Du, S. (2021). Assessing Chinese flood protection and its social divergence. *Natural Hazards and Earth System Sciences*, *21*(2), 743–755.

https://doi.org/10.5194/nhess-21-743-2021

Wang, Z., & Liu, C. (2019). Two-thousand years of debates and practices of Yellow River training strategies. *International Journal of Sediment Research*, *34*(1), 73–83. https://doi.org/10.1016/j.ijsrc.2018.08.006

Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., et al. (2015). Usefulness and limitations of global flood risk models. *Nature Climate Change*, *5*(8), 712–715. https://doi.org/10.1038/nclimate2742

Ward, P. J., Jongman, B., Aerts, J. C. J. H., Bates, P. D., Botzen, W. J. W., Diaz Loaiza, A., et al. (2017). A global framework for future costs and benefits of river-flood protection in urban areas. *Nature Climate Change*, *7*(9), 642–646. https://doi.org/10/gc4njk

Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., & Erickson, T. A. (2017). Validation of a 30 m resolution flood hazard model of the conterminous United States: 30 m RESOLUTION FLOOD MODEL OF CONUS. *Water Resources Research*, *53*(9), 7968–7986. https://doi.org/10/gcg3xn

Wing, O. E. J., Bates, P. D., Neal, J. C., Sampson, C. C., Smith, A. M., Quinn, N., et al. (2019). A New Automated Method for Improved Flood Defense Representation in Large-Scale Hydraulic Models. *Water Resources Research*, *55*(12), 11007–11034. https://doi.org/10.1029/2019WR025957

Wong, T.-T., & Yeh, P.-Y. (2020). Reliable Accuracy Estimates from *k* -Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering*, *32*(8), 1586–1594. https://doi.org/10.1109/TKDE.2019.2912815

Wood, D. J., Brown, C. R. M., Doyle, L., Smith, H. L., Waller, S., & Weller, E. F. (2021). Identification of River Defences from Digital Terrain Models using Deep Learning. In *Science and practice for an uncertain future* (p. null-null). Online: Budapest University of Technology and Economics. https://doi.org/10.3311/FloodRisk2020.14.4

Wu, B., Wang, G., Xia, J., Fu, X., & Zhang, Y. (2008). Response of bankfull discharge to discharge and sediment load in the Lower Yellow River. *Geomorphology*, *100*(3–4), 366–376. https://doi.org/10.1016/j.geomorph.2008.01.007

Xiong, F., Guo, S., Liu, P., Xu, C.-Y., Zhong, Y., Yin, J., & He, S. (2019). A general framework of design flood estimation for cascade reservoirs in operation period. *Journal of Hydrology*, *577*, 124003. https://doi.org/10.1016/j.jhydrol.2019.124003

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., et al. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, *44*(11), 5844–5853. https://doi.org/10/gbnxf9

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resources Research*, *55*(6), 5053–5073. https://doi.org/10/gf3x3k

Zeng, Z., Tang, G., Hong, Y., Zeng, C., & Yang, Y. (2017). Development of an NRCS curve number global dataset using the latest geospatial remote sensing data for worldwide hydrologic applications. *Remote Sensing Letters*, *8*(6), 528–536. https://doi.org/10.1080/2150704X.2017.1297544

Zhang, D., Yin, C., Zeng, J., Yuan, X., & Zhang, P. (2020). Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Medical Informatics and Decision Making*, *20*(1), 280. https://doi.org/10.1186/s12911-020-01297-6

Zhao, G., Bates, P., & Neal, J. (2020). The Impact of Dams on Design Floods in the Conterminous US. *Water Resources Research*, *56*(3). https://doi.org/10.1029/2019WR025380

Zhao, G., Bates, P., Neal, J., & Pang, B. (2021). Design flood estimation for global river networks based on machine learning models. *Hydrology and Earth System Sciences*, *25*(11), 5981–5999. https://doi.org/10.5194/hess-25-5981-2021

Zhao, G., Pang, B., Xu, Z., Cui, L., Wang, J., Zuo, D., & Peng, D. (2021). Improving urban flood susceptibility mapping using transfer learning. *Journal of Hydrology*, *602*, 126777. https://doi.org/10/gpc8wm

Zhu, T., Lund, J. R., Jenkins, M. W., Marques, G. F., & Ritzema, R. S. (2007). Climate change, urbanization, and optimal long-term floodplain protection: OPTIMAL LONG-TERM FLOODPLAIN PROTECTION. *Water Resources Research*, *43*(6). https://doi.org/10.1029/2004WR003516

Zhu, W. (1997). Making Bootstrap Statistical Inferences: A Tutorial. *Research Quarterly for Exercise and Sport*, *68*(1), 44–55. https://doi.org/10.1080/02701367.1997.10608865