A global flood risk modeling framework built with climate models and machine learning

David A. Carozza[1] and Mathieu Boudreault[1]

[1]Université du Québec à Montréal

November 23, 2022

Abstract

Large scale flood risk analyses are fundamental to many applications requiring national or international overviews of flood risk. While large-scale climate patterns such as teleconnections and climate change become important at this scale, it remains a challenge to represent the local hydrological cycle over various watersheds in a manner that is physically consistent with climate. As a result, global models tend to suffer from a lack of available scenarios and flexibility that are key for planners, relief organizations, regulators, and the financial services industry to analyze the socioeconomic, demographic, and climatic factors affecting exposure. Here we introduce a data-driven, global, fast, flexible, and climate-consistent flood risk modeling framework for applications that do not necessarily require high-resolution flood mapping. We first use statistical and machine learning methods to examine the relationship between historical (from the Dartmouth Flood Observatory) flood occurrence and impact, and climatic, watershed, and socioeconomic factors at over 4700 watersheds globally. Using bias-corrected output from the NCAR CESM Large Ensemble from 1980 to 2020, and the fitted statistical relationships, we simulate one million years of events worldwide along with the population displaced. We discuss potential applications of the model and present global flood hazard and risk maps. The main value of this global flood model lies in its ability to quickly simulate realistic flood events at a resolution that is useful for large-scale socioeconomic and financial planning, yet we expect it to be useful to climate and natural hazard scientists who are interested in socioeconomic impacts of climate.

A global flood risk modeling framework built with climate models and machine learning

David A. Carozza[1], **Mathieu Boudreault**[1]

[1]Département de mathématiques, Université du Québec à Montréal, Montréal, QC, Canada

Key Points:

- We present a global flood model built using machine learning methods fitted with historical flood occurrences and impacts
- Forced with a climate model, the global flood model is fast, flexible and consistent with global climate
- We provide global flood hazard (occurrence) and risk (population displaced) maps over 4734 watersheds

Corresponding author: David A. Carozza, `david.carozza@gmail.com`

Abstract

Large scale flood risk analyses are fundamental to many applications requiring national or international overviews of flood risk. While large-scale climate patterns such as teleconnections and climate change become important at this scale, it remains a challenge to represent the local hydrological cycle over various watersheds in a manner that is physically consistent with climate. As a result, global models tend to suffer from a lack of available scenarios and flexibility that are key for planners, relief organizations, regulators, and the financial services industry to analyze the socioeconomic, demographic, and climatic factors affecting exposure. Here we introduce a data-driven, global, fast, flexible, and climate-consistent flood risk modeling framework for applications that do not necessarily require high-resolution flood mapping. We first use statistical and machine learning methods to examine the relationship between historical (from the Dartmouth Flood Observatory) flood occurrence and impact, and climatic, watershed, and socioeconomic factors at over 4700 watersheds globally. Using bias-corrected output from the NCAR CESM Large Ensemble from 1980 to 2020, and the fitted statistical relationships, we simulate one million years of events worldwide along with the population displaced. We discuss potential applications of the model and present global flood hazard and risk maps. The main value of this global flood model lies in its ability to quickly simulate realistic flood events at a resolution that is useful for large-scale socioeconomic and financial planning, yet we expect it to be useful to climate and natural hazard scientists who are interested in socioeconomic impacts of climate.

Plain Language Summary

Flood is among the deadliest and most damaging natural disasters. To protect against flood risk at large scales, stakeholders need to understand how floods can occur and their potential impacts. Stakeholders rely heavily on global flood models to provide them with plausible flood scenarios around the world. For a flood model to operate at the global scale, climate effects must be represented in addition to hydrological ones to demonstrate how rivers can overflow throughout the world each year. Global flood models often lack the flexibility and variety of scenarios required by many stakeholders because they are computationally demanding. Designed for applications where detailed local flood impacts are not required, we introduce a rapid and flexible global flood model that can generate hundreds of thousands of scenarios everywhere in the world in a matter of minutes. The model is based on a historical flood database that is represented using an algorithm that learns from the data. With this model, the output from a global climate model is used to simulate a large sample of floods for risk analyses that are coherent with global climate. Maps of the annual average number of floods and number of displaced people illustrate the models results.

1 Introduction

Flood is consistently among the most damaging natural disasters in terms of economic losses (Gall et al., 2009) and mortality (Hu et al., 2018). Impacts generated by flood result from a complex set of interactions between climatic, hydrological, demographic, and economic factors. Despite improvements in flood defenses and other technologies resulting in reduced vulnerability (Paprotny et al., 2018), nominal flood-related economic losses have increased rapidly in recent decades due to developments in exposure such as total wealth and urban area (Jongman et al., 2012), and rising prices. After normalizing relative to exposure, Barredo (2009) and Neumayer and Barthel (2011) did not identify statistically significant increasing trends in economic flood losses, yet short time series, challenges with data, and the inability to control for changes in flood defenses challenged these studies. Trends in insured losses, which are further complicated by the extent to which exposure is insured, were not found for atmospheric natural disasters at the global scale (Barthel & Neumayer, 2011). However, the same study detected a positive trend in normalized flood-driven insured losses from 1980-2007 in the United States. Paprotny et al. (2018) reconstructed floods in Europe since 1870 and uncovered increasing trends in normalized flooded area and persons affected but decreases in normalized fatalities and financial losses in recent decades.

Large scale flood risk analyses are pivotal to disaster management and relief planning at regional, national, and international levels. Flood risk analyses can build resilience by informing investment needs in mitigation and financial mechanisms such as insurance (Vorogushyn et al., 2018). By the nature of their business, insurance and reinsurance companies are heavily exposed to flood risk across the globe through private and/or public insurance programs (OECD, 2016), whereas banks are subject to mortgage defaults following floods (FRBSF, 2019; Ouazad & Kahn, 2019). With mounting pressure coming from regulators and other bodies worldwide, the financial services industry (banks, insurers and reinsurers) will soon need to disclose and stress test their solvency and stability to various climate scenarios (Bank of England, 2019; Task Force on Climate-related Financial Disclosures, 2017), which includes how future flood risk will affect their profitability.

A major methodological challenge to designing global models is to guarantee that flood risk is consistent from climate, hydrological, hydraulic, and exposure standpoints, such that physically-consistent global climate patterns drive the local hydrological cycle over many watersheds (Vorogushyn et al., 2018). This is a particularly important issue for the financial services industry as their global portfolios are impacted by large-scale climate patterns affecting people over distances of thousands of kilometers. Top down approaches typically force low-resolution hydrological models with meteorological or climate model outputs that simulate runoff that is consistent with simulated climate patterns (Yamazaki et al., 2011; Winsemius et al., 2013). Such approaches are not just global because they represent the entire planet, but because connections between basins in space that are driven by climate are resolved. Top down methods have been used to delve into large-scale flood risk questions such as examining patterns of interannual climate variability (Ward et al., 2014) and to project the impacts of future climate and socioeconomic change (Jongman et al., 2014; Dottori et al., 2018; Ward et al., 2020). An important weakness of this approach is its lack of focus on flooding occurrence and impact in itself, and the inability to resolve small scale floods.

Bottom up approaches consider higher resolution processes that employ a combination of rainfall-runoff or hydrological modeling to drive a hydraulic component and calculate flood damage over watersheds (de Bruijn et al., 2014; Sampson et al., 2015; Falter et al., 2016). These models are typically forced by meteorological (historical, simulated, or projected) or discharge distributions. This more detailed approach is closer to assessing localized impacts of flood but is challenged by high computational demands and data requirements that are not necessarily available globally (Ward et al., 2015). For both approaches, the number of scenarios available is limited and they lack the flexibility required by planners, relief organizations, regulators, and the financial services industry to analyze the socioeconomic, demographic, and climatic factors affecting exposure.

In this paper, we introduce a data-driven, global, fast, flexible and climate-consistent flood risk modeling framework for applications that do not necessarily require high-resolution flood mapping. Our framework is unique in that it is driven by historical flood and environmental observations. It takes advantage of the speed of statistical models to quickly generate large global catalogues of flood events that are physically consistent with climate. Distributions of occurrence and impact can then be analyzed in terms of climatic and socioeconomic factors and over spatial scales of interest. The framework is therefore capable of examining interannual climate variability and looking into the future, accounting for global change over various greenhouse gas emission and socioeconomic scenarios, in addition to accounting for climate-driven connections between basins. Applications of the model include socioeconomic studies, climatic research of the impacts on population or wealth affected, risk analyses in poorly sampled watersheds (Hrachowitz et al., 2013), and stress testing risk portfolios for the financial services industry.

To expand upon the limited observational record (Munoz & Dee, 2017), we generate a large sample of flood occurrence probabilities and impacts using bias-corrected precipitation and temperature output from the National Center for Atmospheric Research's (NCAR) Community Earth System Model (CESM) Large Ensemble (LE) (Kay et al., 2015) for each watershed, ensemble member, and model hydrological year for the time period 1980-2020. The occurrence and impact components are fitted with large databases of past flood history that associate observed flood events to historical precipitation, temperature and watershed information such as topography, land use, soil type, and bedrock features using a machine learning method. Using the fitted occurrence and impact models, we use stochastic simulation to generate a large global catalog of synthetic flood events along with impacts, expressed in terms of the population displaced and the gross domestic product affected in a watershed.

Section 2 presents the datasets used and Section 3 the model development. We evaluate the quality and realism of the flood model in Section 4, present results that illustrate the model's capabilities in Section 5, and conclude in Section 6. A Supporting Information document is available online that presents supplementary description and validations.

2 Data

We base statistical models of flood occurrence and impact on two databases detailed here that associate flood events and their consequences to the driving environmental and demographic conditions. The global flood model is represented in terms of watersheds from HydroBASINS (Lehner & Grill, 2013). Observations of flood occurrence and impact, in terms of population displaced, are derived from the Dartmouth Flood Observatory Global Active Archive of Large Flood Events (Brakenridge, 2010). Environmental quantities that drive flood are represented in terms of a variety of sources that include climatological quantities such as precipitation (Xie et al., 2007; Funk et al., 2015) and temperature (Shi, 2007), and watershed characteristics such as: topography and location (Lehner & Grill, 2013; Marthews et al., 2015b), land cover and vegetation state (Latham et al., 2014), soil type (Shangguan et al., 2014), depth to bedrock (Shangguan et al., 2017), and hydrogeologic properties (Gleeson et al., 2014). Population (Doxsey-Whitfield et al., 2015; Klein Goldewijk et al., 2017) and wealth (Kummu et al., 2018) are used as demographic characteristics (Table 1).

To generate the global flood catalogue of events, we force the flood occurrence and impact models with output from the CESM Large Ensemble (CESM-LE) Community Project (Kay et al., 2015) that is driven by the NCAR Community Earth System Model (CESM1) (Hurrell et al., 2013). We apply precipitation quantities from the Community Land Model 2.0 (Lawrence et al., 2011) and temperature from the Community Atmosphere Model 5.2 (Neale et al., 2012) (Table 2).

3 Model

A riverine flood and its impact is driven by 1. an excess of precipitation less evapotranspiration relative to the storage capacity of the watershed, and 2. interaction with the population affected. We built a statistical framework for flood risk that relates flood occurrence and impact to environmental and demographic predictor variables (Table 1) at the watershed scale. Wolock et al. (2004) and Rumsey et al. (2015) used a similar statistical framework to define hydrologic-landscape regions and estimate baseflow, respec-

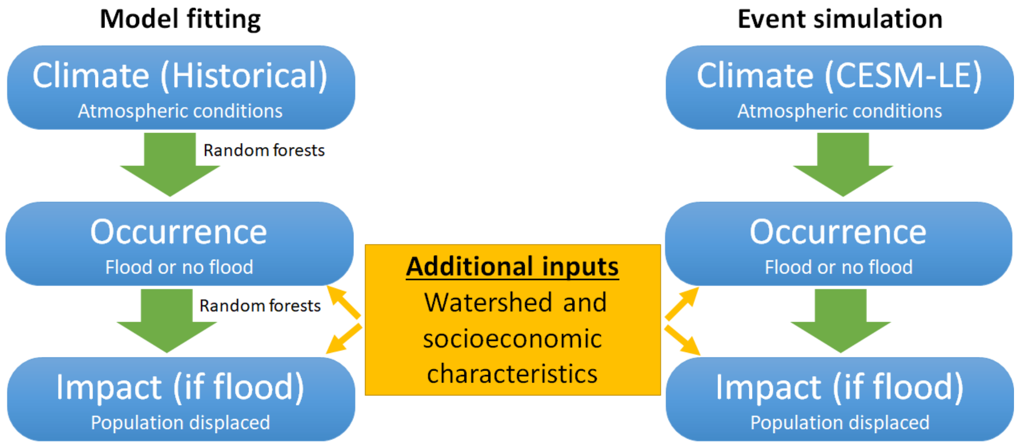


Figure 1. Schematic of the model components. The fitting process is detailed on the left, whereas the simulation process is presented on the right.

tively, but to the authors' knowledge such an approach has not been applied to flood risk modeling at the global scale. While our approach falls into the class of lumped models (Bevin, 2012; Perrin et al., 2013) since the forcing data we used is averaged over a watershed, instead of considering a system output quantity such as discharge, we directly model watershed 1. flood hazard and 2. impact. We achieve this by building databases and statistical models of 1. flood occurrence and 2. the fraction of population displaced, and express them in terms of environmental and demographic predictor variables. The model components are summarized in Figure 1.

3.1 Data Inputs

3.1.1 Observational Data

Historical flood events are provided by the Dartmouth Flood Observatory (DFO) Global Active Archive of Large Flood Events (Brakenridge, 2010). DFO flood events are derived from news, government, and instrumental sources, and validated by satellite observations. Floods are represented in space by means of a polygon that bounds the flooded areas (Figure 2). While this inherently overestimates flooded areas, it represents the synoptic and climatic scales over which riverine flood is driven and so is an appropriate quantity to quantify the association between climate variables and observed large-scale floods. All events with a non-atmospheric cause (Jökulhaup, tsunami, tides, avalanche, storm surge, barrier break or release, ice jam or ice break-up or ice melt) were ignored. We considered the years 1985-2017, during which there were 4499 flood events globally.

Flood hazard and risk are represented at the level of watersheds whose boundaries are defined by the HydroBASINS (Lehner & Grill, 2013) dataset at Pfafstetter level 5, which unless otherwise stated we refer to simply as watersheds. The databases built here and all analyses are conducted over the 4734 watersheds at this scale (Figure 3). The historical flood occurrence and impact databases described are based on the spatial intersection of the DFO events and the HydroBASINS watersheds, which we refer to as watershed-floods. A single observed DFO flood event generally occurs over several watersheds and therefore results in one or more watershed-floods. Due to the bounding polygon nature of the DFO flood events, we assumed that watershed-floods that are less than 5% of the watershed area are "no flood" in our database. Based on an investigation into a subset of DFO events with 0 reported people displaced, we treated such floods as miss-

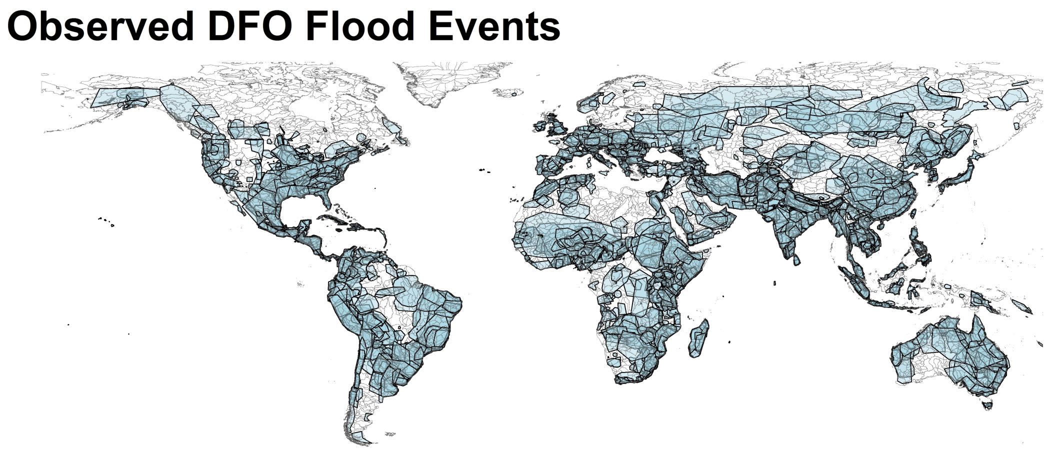


Figure 2. Dartmouth Flood Observatory flood events from 1985-2017.

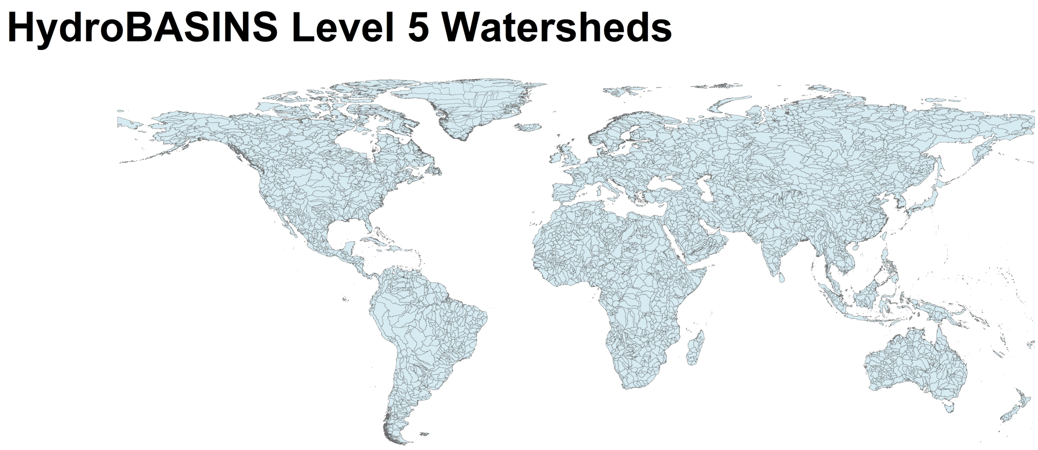


Figure 3. Graphical representation of the 4734 HydroBASINS level 5 watersheds.

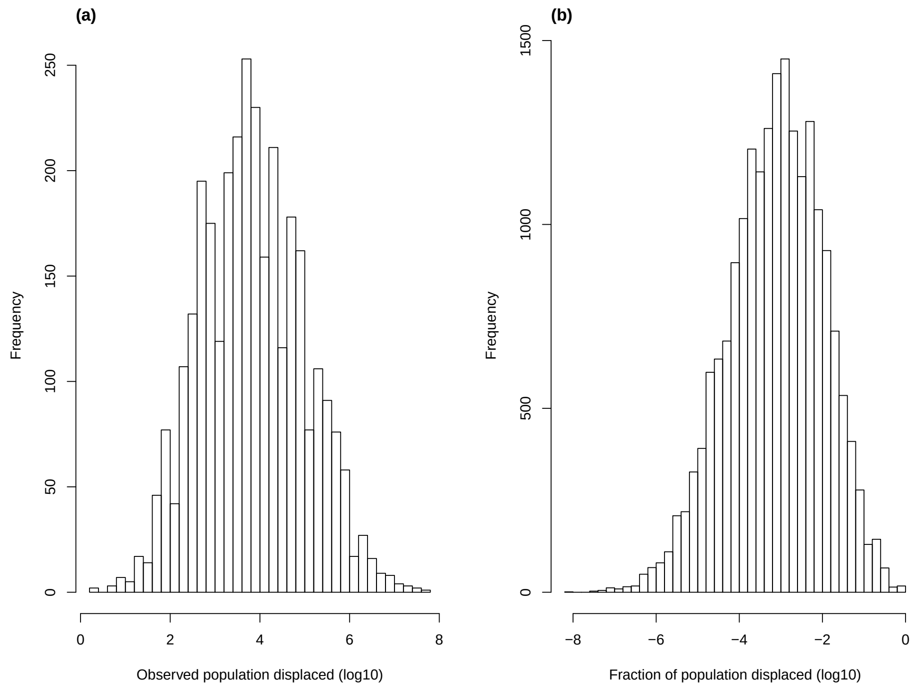


Figure 4. Distributions of population displaced impact variable. (a) Observed population displaced (log10) for each DFO event. (b) Fraction of population displaced (log10) for each watershed-event.

ing data instead of events with zero impact. These steps resulted in 3160 of the 4499 DFO floods with at least 1 person displaced (Figure 4a). For each flood, we distributed the population displaced over the watershed-floods in proportion to the population of the watershed in the year of the flood (Figure 4a).

We explored time-varying environmental predictors and chose datasets with the following features: 1. global spatial coverage, 2. temporal coverage that contained the DFO flood event dataset (1985-2017), 3. at least 1° spatial resolution, 4. at least daily temporal resolution, 5. resolved in the CESM-LE climate product that drives the catalogue simulation. While observational products such as soil moisture (Gruber et al., 2019) and terrestrial water storage (Tapley et al., 2004) were of interest, our restrictions limited our analyses to precipitation and temperature products.

Precipitation is the key driving predictor to flood occurrence and impact. We represent it using the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) dataset (Funk et al., 2015) for latitudes from 50°S to 50°N, and the CPC Global Unified Gauge-Based Analysis of Daily Precipitation (CPC Precipitation) dataset (Xie et al., 2007) for all other latitudes. Temperature plays a key role in evapotranspiration (Li et al., 2016) and we represent it with the CPC Global Daily Temperature (CPC Temperature) dataset (Shi, 2007). We considered several timescales of precipitation and temperature to capture potential regimes of climatic trajectories that can result in flood, namely averages over the 7, 8-30, 31-60, and 61-120 days prior to an event.

The remaining environmental predictors represent watershed storage capacity. Watershed topography is a crucial characteristic in hydraulic modeling that determines flows along the surface and subsurface and influences infiltration into the subsurface (Farr et al., 2007). We represent topographic features and watershed location with the HydroBASINS (Lehner & Grill, 2013) and High-resolution global topographic index values (TI) (Marthews et al., 2015b) datasets. Land cover characteristics alter infiltration into the subsurface and can additionally influence evapotranspiration (Nie et al., 2011). We represent land cover with the Global Land Cover SHARE (GLC-SHARE) (Latham et al., 2014) dataset. Soil permeability affects infiltration (Wolock et al., 2004) and subsurface drainage (Yu et al., 2000), whereas soil depth is indicative of soil water storage. We represent soil type with the Global Soil Database (GSD) (Shangguan et al., 2014) and depth with the Global Depth to Bedrock (GDB) (Shangguan et al., 2017) dataset. Bedrock porosity and permeability proxy subsurface storage and drainage, respectively (Wolock et al., 2004). We represent these quantities using the GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0) (Gleeson et al., 2014) dataset.

The exposure predictors population density and GDP (Gross Domestic Product) per capita are interpreted as time-varying proxies of urbanization and flood control that can additionally capture features of land cover and socioeconomic change that are not represented in the time-invariant GLC-SHARE dataset, and as such affect both hazard and impact. Population is represented with the Gridded Population of the World (GPW) (Doxsey-Whitfield et al., 2015) and Anthropogenic land-use estimates for the Holocene (HYDE3.2) (Klein Goldewijk et al., 2017) datasets, and wealth with the Gridded global datasets for GDP and HDI (Human Development Index) over 1990-2015 (GDPHDI) (Kummu et al., 2018). Temporal extrapolations of these variables are described in the Supporting Information.

3.1.2 Climate Model Output

The climate component used in our global model applies the NCAR-CESM1 (Hurrell et al., 2013) Large Ensemble (CESM-LE) (Kay et al., 2015). The CESM-LE was designed to examine interannual climate variability in the context of anthropogenic climate change, and consists of 40 ensemble members run from 1920 to 2100. Each member is initialized with a roundoff error perturbation to the atmosphere in model year 1850, so that by 1920 the members are independent of one another yet driven by the same forcing. The NCAR-LE employs a single historical forcing scenario up to 2005 and the RCP8.5 from 2006 to 2100. In this paper, to be consistent with the flood observational record from DFO, we consider the years 1980 to 2020. This results in 40 years from 40 members, which we refer to as 1600 member-years of global climate model output that drive the flood occurrence and impact models to generate a stochastic catalog of floods. As with the predictor variables used in the statistical fit step, for each climate forcing variable (Table 2) we aggregate by taking the average of grid points in each level 5 watershed.

CESM-LE atmospheric rain and snow are summed to model precipitation. Snow is assumed to melt immediately to water when it touches the ground at a bulk weight density of 100 kg m^{-3} corresponding to that of fresh snow (Meløysund et al., 2007). Biases in the CESM-LE precipitation are corrected relative to CHIRPS and CPC Precipitation, and biases in temperature are corrected relative to CPC Temperature, both using the methodology of Hempel et al. (2013). When considering the correction of the number of dry months, we instead use the fraction of dry months since our simulated dataset (CESM-LE, with 40 years of data) has a different sample size to the observations (CHIRPS, with 37 years of data). This bias correction approach is widely used in the hydrological and flood impact literature and separately corrects monthly means and daily variability about the means, thereby preserving long-term observed trends. We apply the correction to precipitation and temperature aggregated over each level 5 basin.

3.2 Occurrence Component

This component of the model represents the occurrence of a riverine flood ("flood" or "no flood") in a level 5 watershed given a set of observed or simulated predictor environmental variables. Watershed hydrology and river hydraulics that result in floods are complex processes with inherent nonlinear features and interactions between forcing variables. To break down this problem, we first represent how climatological, hydrological and basin-scale factors determine riverine flood occurrence globally and regionally in the model fitting step. Once these relationships are established at the watershed level, we use output from the climate component in addition to the basin-scale predictors to simulate flood occurrences. Combining these two steps results in a global flood hazard model.

3.2.1 Model Fitting

Given the observed watershed-floods, we assumed that a level 5 watershed can be in one of two states ("flood" or "no flood") in a given hydrological year dating from October 1 - September 30. The 32 hydrological years considered in the DFO database and 4734 watersheds resulted in a potential of 151488 occurrence observations. Taking missing data in the predictors into consideration left us with 128494 observations for the fitting process. For the occurrence problem climate predictors, we use the mean annual temperature and for precipitation take the annual maxima of each of the 4 timescales considered over the hydrological year.

The statistical problem at hand is therefore a classification problem with the binary response variable ("flood","no flood"). We considered classical and machine learning methods such as logistic regressions (LR), random forests (RF, Breiman (2001)) and artificial neural networks (NN, McCulloch and Pitts (1943)) to solve this problem. Hastie et al. (2009) discuss the three methods in chapters 4, 15, and 11, respectively. Since it is difficult to define explicit functional forms and interaction terms between the predictor variables a priori, in particular given the variety of flood regimes that exist globally, we adopted a machine learning approach that builds such relationship from observations. Validation of the occurrence models are presented in Section 4.1 for the chosen model, as well as in the Supporting Information for the other models.

We first fit each of the three models globally, thereby providing one set of parameter estimates per model. To guarantee that local dynamics were appropriately represented in the global fit, we also made fits of each model on aggregations of watersheds at Pfafstetter level 2 (HS2). Since there are 62 watersheds at HS2, 62 parameter sets were fitted.

For the global occurrence fit we took a random sample of 70% of the observations as the training set for all models and conducted out-of-sample validation on the remaining 30%, the test set. Although normalization of the variables is only required for the NN models, to compare the models we normalized all predictor variables to the range $[0, 1]$. For the RF model we considered 500 decision trees and 5 randomly selected predictors, from among the 38, to decide which predictor is selected for each split. Regional occurrence fits by HS2 watersheds are also based on normalized predictors and to compare with the global fit we used the same training and test set. Given the smaller samples involved in fitting the regional models, we only apply a fit if more than 3% of observations in the region are "flood". The structure of the RF models follows the global model, with 500 decision trees and 5 predictors considered for each split.

3.2.2 Event Simulation

To simulate floods we apply the fitted occurrence model to compute predicted flood occurrence probabilities (1600 member-years $\times$ 4734 watersheds) using the bias corrected precipitation and temperature outputs from the CESM-LE and the basin-scale environmental and socioeconomic variables. Flood occurrence, for each member-year and each watershed, thus follows a Bernoulli distribution whose only parameter represents the probability of a riverine flood. The flood occurrence probability is unique for every ensemble member year and watershed (40 members x 40 years x 4734 watersheds). To generate a large sample of flood events at the global scale, we sample from each of these Bernoulli distributions 625 times, resulting in a million simulated years.

3.3 Impact Component

The impact component models the number of people affected by a given flood. It aims to approximate the combined effects of the exposure (population, wealth, etc.) and its vulnerability. Combining the occurrence and impact components thus yields the global flood *risk* model.

3.3.1 Model Fitting

Combining the DFO population displaced with the level 5 watershed population (Doxsey-Whitfield et al., 2015; Klein Goldewijk et al., 2017), we model the fraction of the population displaced (displaced population / watershed population) to proxy the impact. We applied a log10 transformation to the fraction displaced since it spans 8 orders of magnitude (Figure 4b). The predictor variables used are the same as for the occurrence model, with the important difference that the four timescales for temperature and precipitation are calculated for the final day of the event as reported in DFO. Overall, the impact model aims to explain the log10 of the fraction of population displaced as a function of demographics, wealth, climatological, and watershed predictors. The climatological and watershed predictors aim to capture the effects of the intensity of a flood on the fraction of population displaced whereas population density and wealth per capita aim to capture the vulnerability of a population.

There are 19746 watershed-floods in the impact database with complete response and predictor information. As in the statistical modeling for occurrence, we considered classical and machine learning methods, namely linear regressions, random forests and neural networks (see above references for details). Validation of the impact models are presented in Section 4.2 for the chosen model, as well as in the Supporting Information for the others.

The global impact model fitting procedure follows the occurrence model and takes 70% of the observations to form the training set. All validations are done by applying the fitted model to the remaining 30% of observations. As the impact model is a regression problem, both predictors and the response variable were normalized to the range $[0, 1]$. For the random forest fits, since there are 41 predictors, we consider 6 randomly selected predictors at each tree split and repeat this for 500 regression trees. Regional impact observations were also normalized and used the same training observations as the global fit. A regional model was fitted when there were at least 30 watershed-flood impact observations in the HS2 watershed.

3.3.2 Event Simulation

For each simulated flood event, we in turn simulate impact in terms of the fraction of the watershed population displaced. The impact depends on the bias-corrected temperature and precipitation from the CESM-LE, watershed characteristics (which are assumed to not change in time), and the population density and wealth observed in the watershed for the year of the event. We assume the log10 of the fraction of population displaced is normally distributed (Figure 4b). The mean of the flood impact distribution is taken from the maximum of daily impact predictions over the hydrological year.

The standard deviation parameter is determined by calculating the root mean squared prediction error in each of 9 groups determined by watershed population density. This grouping preserves the observed structure of increasing and then flat prediction errors for watershed-floods that depend on population density (Supporting Information).

In summary, to model impact in any watershed and year, we sample from a normal distribution whose 1. mean parameter is the largest daily simulated impact of the fraction of population displaced, and 2. standard deviation parameter is the root mean squared prediction error from the fitted model for the population density group of the watershed. This process is repeated for each simulated flood over the million years of the catalogue.

3.4 Validations and Variable Importance

The flood occurrence model is a binary classification problem. A good classification model should predict an event when there is really an event (a true positive). However, when occurrences are rare, it is easy to be accurate most of the time by simply predicting the event always (or never) occurs. As such, one needs to evaluate models by balancing true positives (TP) and true negatives (TN) with false positives (FP) and false negatives (FN) (Fawcett, 2006; Powers, 2011). Such analyses are commonly summarized using the receiver operating characteristic (ROC) curve, a plot of the true positive rate (TP/(TP+FN)) versus the false positive rate (FP/(FP+TN)), which are both determined as functions of the cutoff probability used to define a predicted "flood". The area under the ROC curve (AUC) is a summary measure that ranges from 0 to 1 and indicates the likelihood that the classification model can differentiate between "flood" and "no flood." Values above 0.5 indicate that the model in question has the ability to differentiate between classes. We report the AUC aggregated over HS2 watersheds if there are at least 10 observations and at least 5 floods in the test set (Section 4.1 and Figure 5). Other model evaluation metrics are discussed in the Supporting Information.

To assess the quality of the impact models for the linear model (LM), random forest (RF), and neural network (NN), we consider two metrics using out-of-sample observations : the root mean square error (RMSE, lower is better) and the R-squared (higher is better, with 1 being the maximum). The RMSE summarizes the model error whereas the R-squared measures the proportion of variance explained by a model (Hastie et al., 2009) (Section 4.2 and Figures 6 and 7).

For the random forest models, we consider two variable importance measures for each of the flood classification and regression problems using the R `randomForest` package (Liaw & Wiener, 2002). The first measure considers how the accuracy changes in reaction to permuting the observations of each predictor in the out-of-bag (a test set) observations. For occurrence, accuracy is defined as the fraction of observations that are correctly classified $(TP+TN)/(TP+TN+FP+FN)$, whereas for impact the mean squared error is used. The second measure employs node purity, which rewards homogeneity in predictions. For flood occurrence, the Gini impurity index is used, whereas for impact the residual sum of squares is used.

3.5 Code and Computations

Our work was coded using the R software environment (R Core Team, 2018). We used the package `data.table` (Dowle & Srinivasan, 2018) for data processing and merging, and `velox` (Hunziker, 2017) to calculate aggregations of the predictor and climate variables over the HydroBASINS watersheds. The packages `sp` (E. J. Pebesma & Bivand, 2005) and `sf` (E. Pebesma, 2018) were applied for spatial analyses such as spatial intersections. The linear model and logistic regressions fits were achieved with the `stats` package core functions `lm()` and `glm()` (R Core Team, 2018), and we applied the `randomForest`

(Liaw & Wiener, 2002) and `RSNNS` (Bergmeir & Benítez, 2012) packages for the RF and NN models. `ROCR` (Sing et al., 2005) was used for the occurrence model validation calculations and `doSNOW` (Microsoft & Weston, 2017) was used to parallelize computations. We used `cartography` (Giraud & Lambert, 2016) for choropleth maps and `RColorBrewer` (Neuwirth, 2014) for the color schemes.

Table 3 details the steps involved in simulating the flood catalogue. The durations reported are for a single processor thread on an Intel Xeon E5-2650 v3 at 2.30 GHz. Step 2 takes approximately 1 day to compute daily impact predictions over the model years 1980-2020 of the CESM-LE. While this is the most computationally demanding step, it takes only 37 minutes per ensemble member and so the duration of the user's calculations depend on the number of ensemble members of interest. All other steps are rapid, with the simulation of 1 million years of flood occurrence requiring only 14 minutes (Step 1) and the simulation of corrected impacts and merging with occurrence taking only 39 minutes (Step 8). While we worked with a single thread, step 2 can be parallelized given sufficient system memory, easily reducing the calculation by a factor of 6 to 8.

Simple shocks to the occurrence or impact components through the predictor variables, such as precipitation or temperature, can be conveniently considered to examine model sensitivity. In particular, for monotonic shocks to individual predictors, Step 2 need not be repeated since the predictors that generate the annual maximum are already known. The occurrence and impact prediction functions of the model can also be used with alternative precipitation and temperature output (for example another climate model, reanalysis product, or temperature or precipitation product), or socioeconomic data products. For such an application, the user should first compare the statistical properties of the new forcing quantities over the watersheds of interest with those used in the model fitting. Based on those analyses, the user should consider applying a bias-correction before proceeding with the steps described in Table 3.

4 Results and Model Validation

This section presents validations for the occurrence and impact models that we describe in terms of goodness of fit, parsimony, ease of use, and interpretation (Sections 4.1 and 4.2). Results are presented on world maps over aggregated HydroBASINS watersheds. We analyze the realism of the entire simulated flood catalogue in Section 4.3. Additional validations and other results can be found in the Supporting Information.

4.1 Occurrence

Overall, we find that logistic regressions do poorly for the global fit but their quality significantly improves when fitted regionally (Table 4 and Supporting Information). RF and NN models are built to capture complex non-linear relationships and interactions between predictors and given the problem at hand it is not a surprise that such non-linearities and interactions appear. As such, RF and NN perform similarly globally and over regional aggregations of watersheds. That said, for fits over HS2 watersheds, all three methods perform similarly.

Neural networks do not significantly improve the quality of the fit when compared to random forests. With only 128494 observations, the dataset is likely not in the appropriate sample size regime to observe the benefits of NNs. Given the difficulty in interpreting NNs, we cannot recommend their use for this application. We generally find that the RF, fitted globally or regionally, is a solid approach in terms of in-sample and out-of-sample fit of global flood occurrences. Given that the RF is composed of individual decision trees applied to random samples of the observations, it is also easier to interpret than the other models. If a user has a preference toward a more statistical approach, we recommend the use of logistic regressions in combination with regional fits.

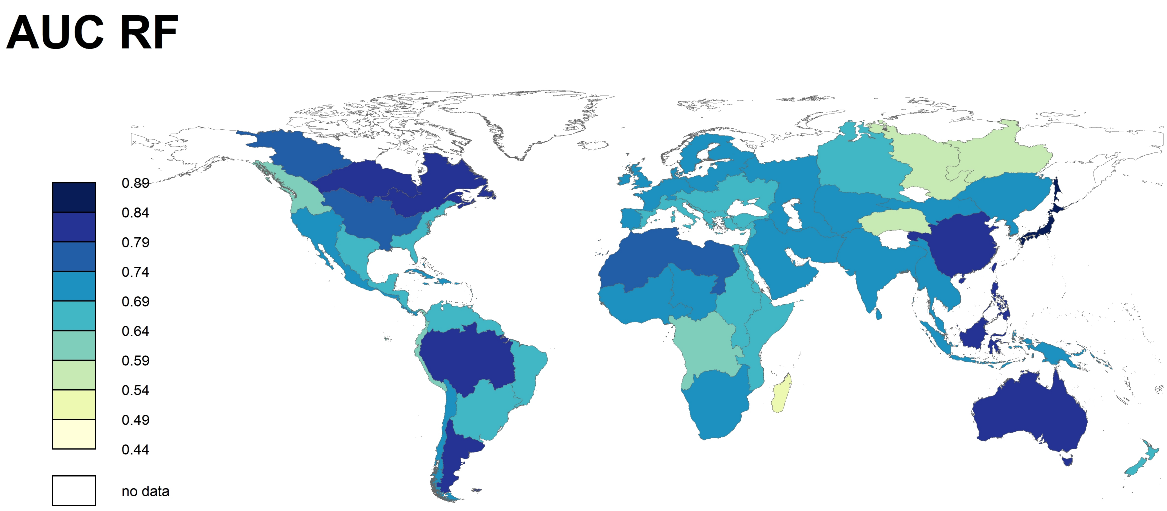


Figure 5. World map of the area under the receiver operating characteristic curve (AUC) for the globally-fitted random forest model. The AUC is evaluated using out-of-sample observations aggregated over each level 2 watershed (HS2) and indicates the probability that a model can differentiate between "flood" and "no flood".

In the Supporting Information, we present the ROC curve and other validation measures used to support these results.

Figure 5 demonstrates the performance of the globally fitted random forest in terms of the area under the ROC curve aggregated over HS2 watersheds. The occurrence model does particularly well in Eastern Canada, Eastern China, Japan, Indonesia, Australia, and the Amazon. The global random forest model shows predictive skill (AUC more than 0.5) over all regional HS2 watersheds, and generally has the ability to identify the environmental and socioeconomic features that generate flood occurrence. We are encouraged that only 6 of the 62 HS2 watersheds have an AUC less than 0.6.

Table 5 lists the 10 most significant predictors found to explain flood occurrence with the globally fitted RF model. Regardless of the measure chosen, the two exposure predictors of GDP per capita and population density are prominent. Unsurprisingly, the precipitation over various timescales are key predictors. Annual mean temperature is also an important predictor, driven by its link to interannual patterns in evapotranspiration. The remaining predictors represent the residual components of flood hydrology. Topographic effects are represented by the topographic position index and aspect, soil content by gravel, and land usage by cropland. Bedrock porosity provides a proxy of longer term storage.

4.2 Impact

As with the occurrence model, we fit each model globally and by HS2 regional subsets of watersheds. In terms of the quality of the out-of-sample fit, we again cannot claim that the NN models outperform the others, so given their complexity we do not recommend them. Linear models (LM) perform poorly globally but their fit improve once fitted over different subsets of watersheds (Table 6). However, LM are outperformed by RF in terms of global and local fit and hence we recommend the latter method for this application. The Supporting Information presents the RMSE and R^2 for all three models fit globally and by regional subsets of watersheds.

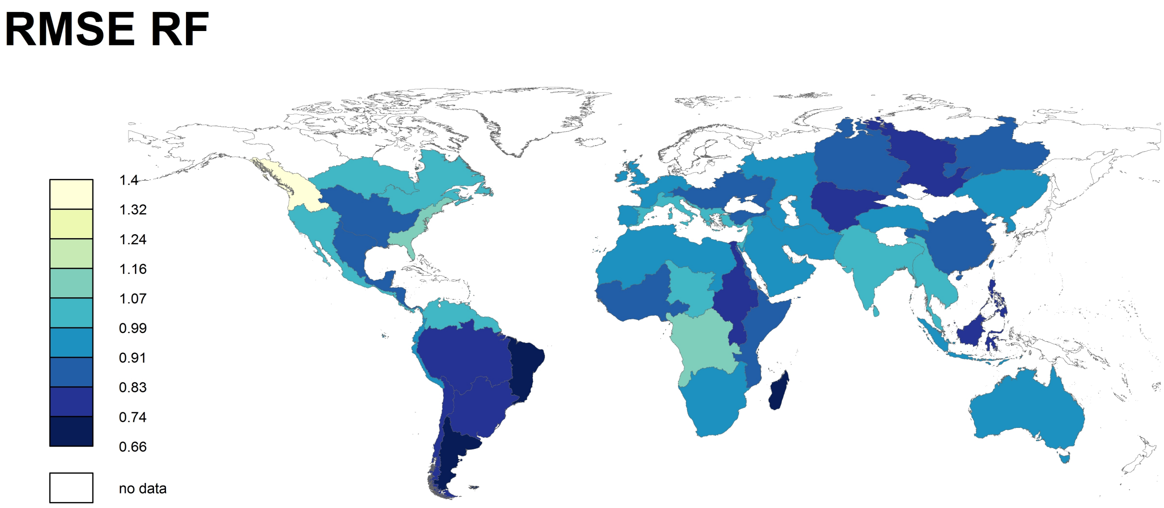


Figure 6. World map of the out-of-sample RMSE for each level 2 watershed for the globally fitted random forest model.

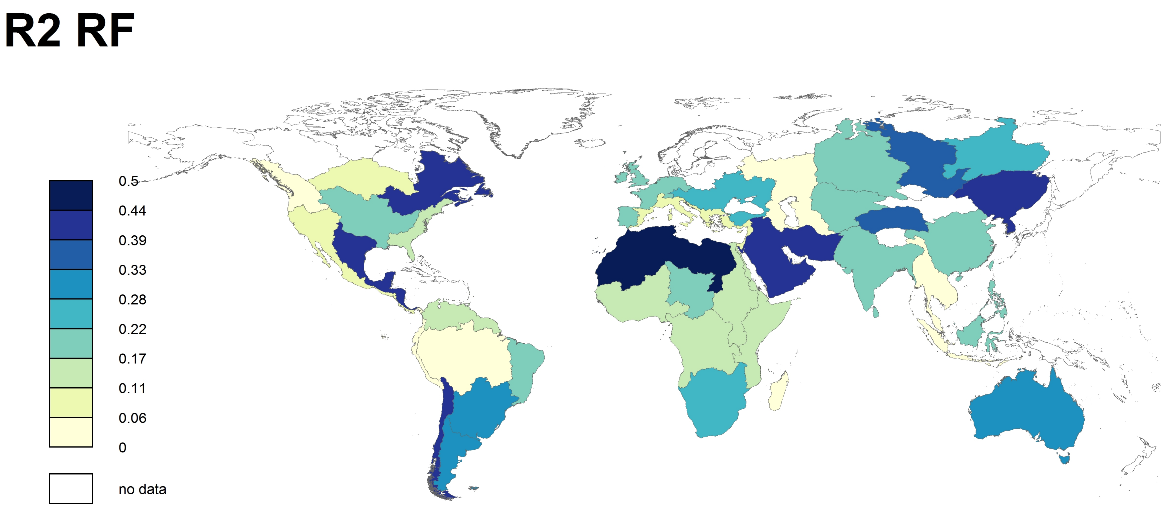


Figure 7. World map of the out-of-sample R-squared for each level 2 watershed for the globally fitted random forest model.

Figure 6 shows the root-mean-square error (RMSE) aggregated by HS2 watershed regions for the random forest fitted globally. Almost all HS2 regions have an out-of-sample RMSE below 1, which given the log10 scale of the fraction of population displaced used as the response variable, indicates that the predictions are within an order of magnitude of the observations. This result is excellent given that impact ranges over 8 orders of magnitude (Figure 4b). It is also important to note that in the observations, the fraction of displaced people is very small (mainly in the range of $[-5, -2]$ on a log10 scale) meaning that absolute errors, rather than relative errors, are quite small. Moreover, only one quarter of the HS2 watersheds analyzed have an out-of-sample R^2 below 10%, while the majority are above 20% (Figure 7) and many of them are above 40%. Given the challenge of predicting the fraction of the population affected by a flood over the globe, we find these to be promising results. A number of the poorer predictive results are influenced by limited observations, such as for Madagascar, Papua New Guinea, and New Zealand (see Supporting Information). However, for the more challenging watersheds the predictors are not representing the fraction displaced because 1. other predictors represent the relevant flood regimes in those watersheds, 2. there are biases or inaccuracies in the impact observations, 3. there are biases in the manner in which the persons displaced are associated to particular watersheds. For the impact model, we are reassured by good levels of out-of-sample variance explained over the majority of the globe, and by a lack of systemic patterns in the watersheds with poorer variance explained.

Table 7 lists the 10 most significant predictors in the globally fitted random forest impact model. Regardless of the measure used, we find that most timescales of precipitation and temperature variables need to be included in the model, as well as population density and GDP per capita. The resulting predictors are similar to those identified in the occurrence model, with lagged temperature predictors taking on an important explicative role. Overall, we find that lagged precipitation and temperature variables, when applied with the exposure predictors of GDP per capita and population density, capture the majority of the resolved signal of seasonal flooding.

4.3 Global Model

This section assesses the entire flood catalogue, which consists of 1 million simulated years of flood occurrences and impacts over 4734 watersheds globally. We first evaluate the displaced fraction of population and then the simulated occurrence and population displaced. Despite the bias correction of precipitation and temperature, the time-varying CESM-LE output and basin-scale predictors do not sufficiently capture impact extremes. This results in underestimated densities for both low and high values of the fraction displaced (Supporting Information). To remedy this issue in a simple way, we scaled the standard deviation parameters of the impact distribution to match the observed standard deviation (Supporting Information).

Figure 8 compares the mean simulated and observed occurrence and impact aggregated over level 3 watersheds (HS3). It is important to note that such a comparison is ambitious since the performance of the global model depends on the performance of the CESM-LE to generate conditions favorable for flood. As such, the clustering of points along the 45° lines indicates that the global model works well overall to simulate a realistic number of floods and displaced people compared to the observations. For low risk areas though, the model tends to slightly overestimate flood hazard. This could be because no flood was observed over the 32-year observational record even if the true flood probability is non-zero. This bias results in overestimates of the number of displaced people for small population watersheds. Nevertheless, for HS3 watersheds with an annual mean of more than 300 people displaced (the great majority of basins where risk analysis is relevant), the model predictions fit the observations well.

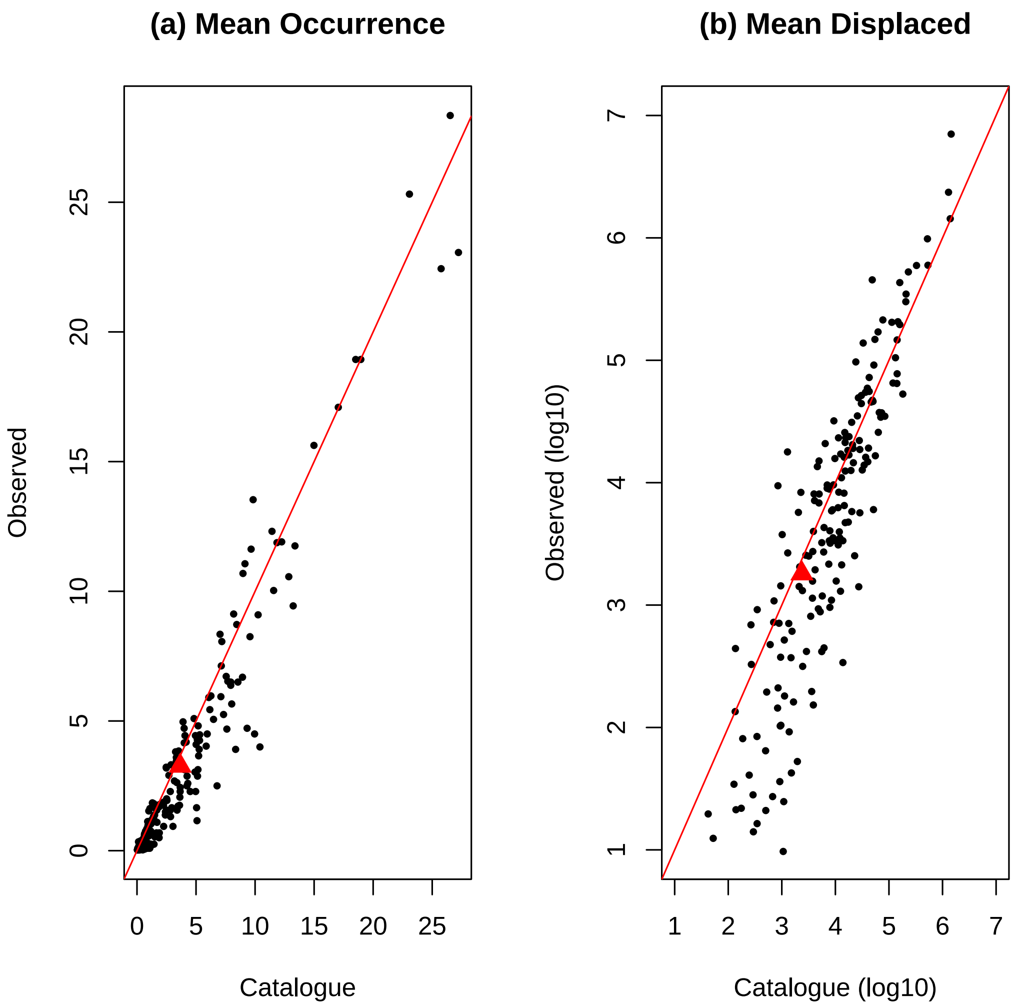


Figure 8. Observed versus simulated flood occurrence and population displaced aggregated by level 3 watersheds. (a) Mean observed flood frequency versus mean simulated frequency. (b) Mean observed versus simulated population displaced, expressed on a log10 scale. The red triangle represents the mean. Averages of observations are taken over the 32 hydrological years available from the DFO and over 1 million years for the catalogue.

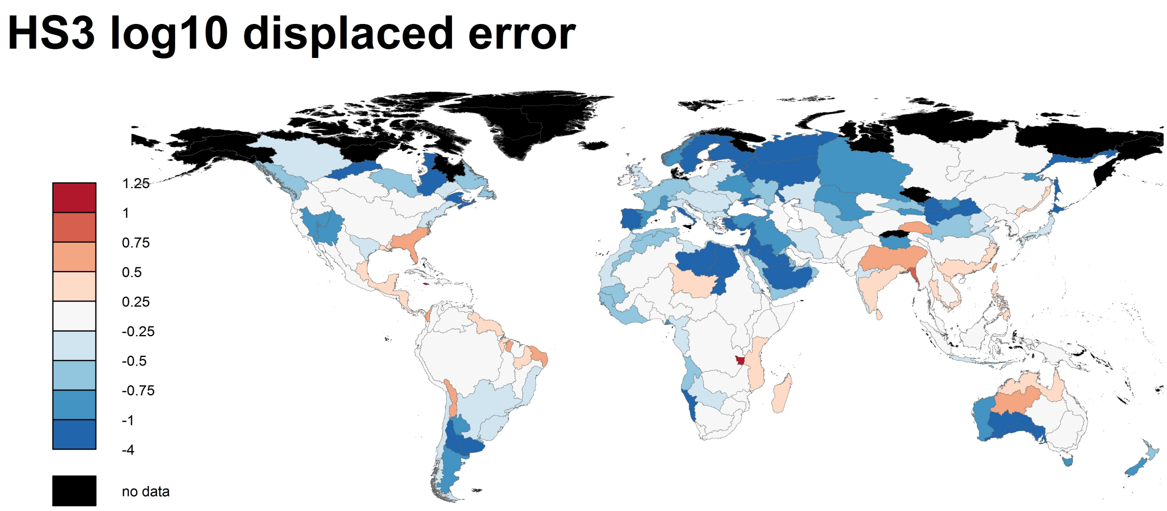


Figure 9. World map of the error on the mean number of displaced people. Represented as the mean observed displaced population less the mean displaced population in the catalog (with each member of the difference being on a log10 scale). Negative values indicate the model overestimates observations.

The world map in Figure 9 illustrates the average error in the number of displaced people. The darkest blue shade represents locations where the overestimation is larger than a factor of 10. Of these regions, we observe that they are either in far northern regions (Northern Canada, Sweden, Finland, North-Western Russia) or dry climates (Libya, Egypt, Saudi Arabia, and Southwest Australia). While flood is not common in these regions, biases could be due to underreporting in DFO in these regions as they are sparsely populated, to overestimates in the CESM-LE flood-generating conditions, or a lack of fit of the occurrence/impact statistical models due to a lack of observations. Looking back to Figures 5 and 7, the fit of both statistical models is good in dry climates but not as successful over northern regions. Overall, there are less cases of underestimated population displaced. The model underestimates population displaced in a few South Asian watersheds in Bangladesh, India, and Nepal. This region, particularly Bangladesh, is known for extreme floods resulting in millions of displaced driven by the unique combination of precipitation extremes from the annual Indian Monsoon, low-lying and complex hydrology, high population, and poor infrastructure (Dewan, 2015), which is beyond the ability of our model to capture.

5 Applications

To illustrate potential uses for the model, we show global flood hazard and risk maps based on the stochastic catalogue comprised of one million years of events. Flood hazard is expressed as the annual flood probability whereas flood risk combines flood hazard with population and wealth exposed.

Figure 10 presents the annual mean flood frequency in the catalogue for each HS5 watershed, which can be interpreted as the annual flood probability over each HS5 watershed. Darker colors point to flood hot spots, with the highest flood probabilities in northeastern India, Bangladesh, and Myanmar being driven by the annual South Asian monsoons. Southeastern China, Japan, and certain areas of Southeast Asia are also high hazard areas. Other regional peaks are found in eastern USA, southern Mexico and Cen-

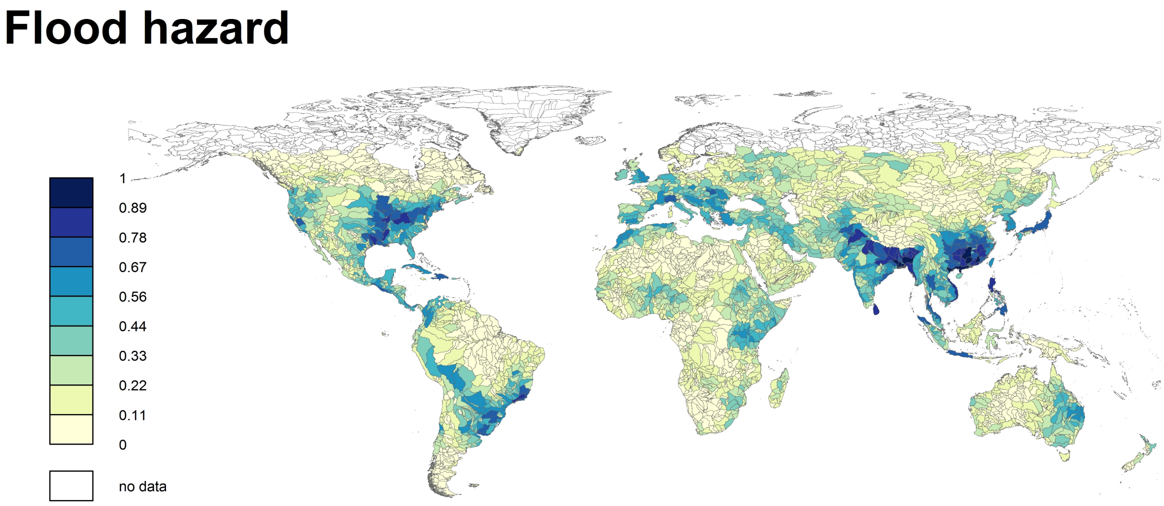


Figure 10. Global flood hazard map by level 5 watershed expressed as the annual flood probability.

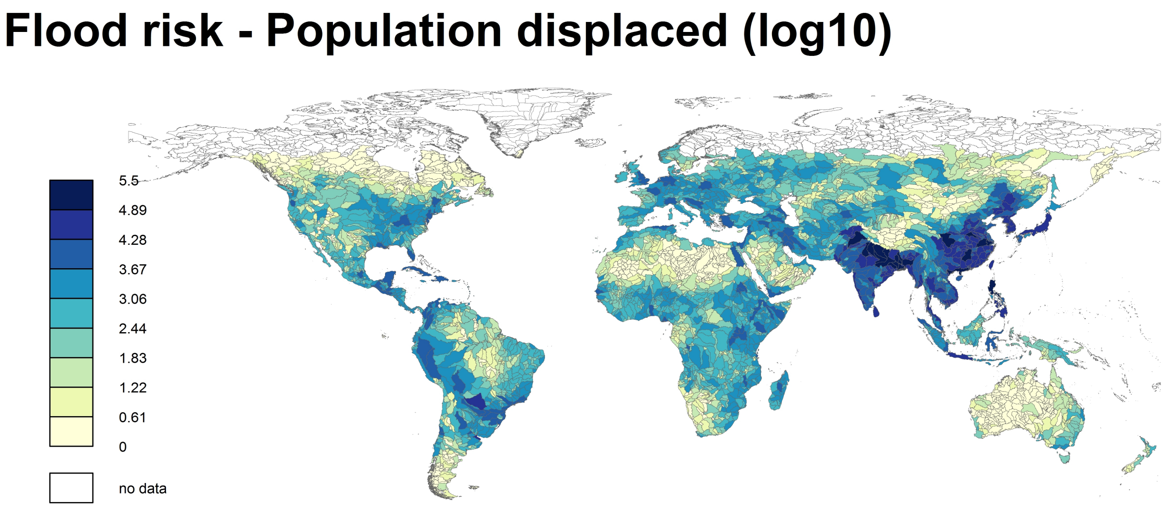


Figure 11. Global flood risk map, expressed as the average population displaced (log10).

tral America, southern Brazil, and parts of Europe. Although these probabilities seem high, they represent the likelihood that at least one river within an HS5 watershed overflows sufficiently so that a minimum number of people are displaced. It should not be compared to typical return periods.

The annual average population displaced over each HS5 watershed (Figure 11) combines flood hazard (flood probability) and exposure (population) to yield the average number of displaced people in any given year (flood risk). For example, over Northern India, Bangladesh and China, many highly densely populated watersheds yield an annual average number of displaced people of over 100,000. Over Central Europe, for example, many watersheds have an average annual number of displaced people of about 1,000. We expect this model of flood risk to be highly relevant to risk management and socioeconomic studies.

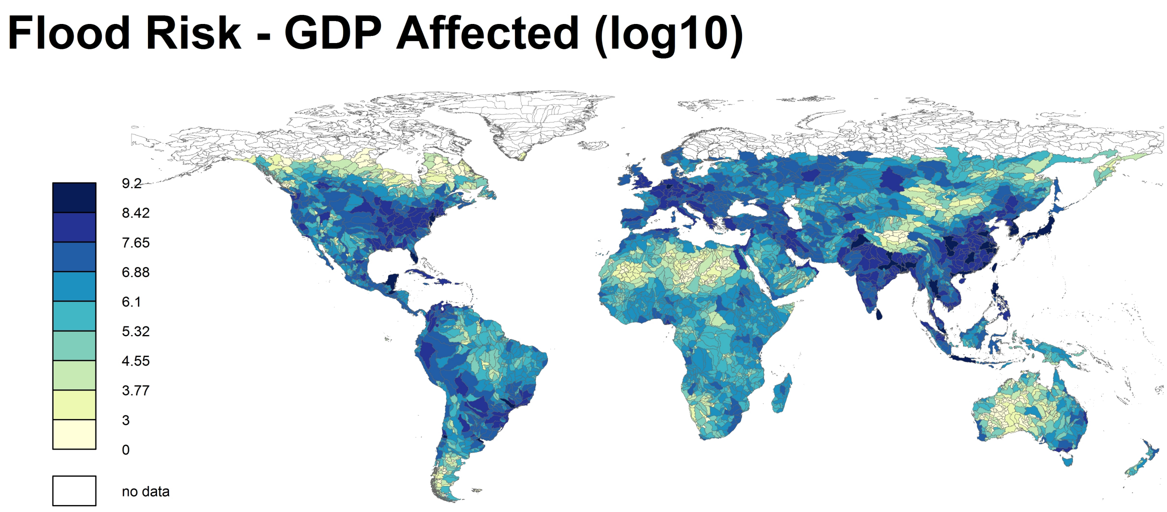


Figure 12. Global flood risk map, expressed as the average GDP produced by the displaced population (log10, US dollars)

To illustrate the global scale economic loss potential of flood, we translate population displaced into the GDP of the population displaced by simply multiplying by the annual GDP per capita (Kummu et al., 2018). Figure 12 presents this alternative perspective of flood risk, with exposure being the GDP of the population over a given watershed. This view of flood risk is more targeted to studies of economic loss and financial risk management. In terms of GDP affected, less developed high population watersheds remain important, yet hot spots appear in wealthier regions in Europe (UK, East of France, Belgium, Netherlands, Germany, and Italy), North America (Eastern USA and West Coast USA, Southern Canada), South America (Southern Brazil and Northeastern Argentina), and Australia (Brisbane and Melbourne).

6 Conclusion

By integrating the Dartmouth Flood Observatory database of historical flood events with the HydroBASINS database of watersheds, we created novel flood occurrence and impact databases that describe flood frequency and intensity over a watershed in terms of the climatic, watershed, and socioeconomic drivers. We then fitted classical regression and machine learning techniques to these data, and adopted the random forest model fitted to observations at the global scale. Finally, we generated a global catalogue of flood events by forcing the empirical model with bias-corrected precipitation and temperature output from the large ensemble of the NCAR CESM climate model.

The unique value of this global flood model lies in its ability to quickly simulate realistic flood events at a resolution that is useful for large-scale socioeconomic and financial planning. Translating outputs from a climate model into flood events facilitates the creation of scenarios and projections of impacts over various time horizons. One could apply global weather or seasonal forecasts to simulate flood impacts over time horizons from days to months, or focus on different time horizons from the NCAR CESM climate model and investigate the impacts of climate change on flood hazard and risk. Alternatively, subsets of the catalogue could be extracted to investigate the impacts of various climate oscillations on flood hazard and various measures of risk. Including population and wealth per capita in the model allows for sensitivity testing and experimentation of the dependence of flood hazard and risk to changes in the spatial population and wealth patterns. Finally, one could conveniently consider shocks to the outputs of the climate model (such as significant changes in precipitation over a given area in a year) and evaluate different climate scenarios, as will be required by regulators of the financial services industry. We expect this model to be a useful empirically-based though climatically-consistent complement to the mechanistic and other approaches available.

Acknowledgments

The data used in our model development is detailed in section 2. The HydroBASINS dataset is available at `https://www.hydrosheds.org/`. The Dartmouth Flood Observatory (DFO) Global Active Archive of Large Flood Events can be downloaded at `https://floodobservatory.colorado.edu/Archives/index.html`. The CHIRPS global precipitation dataset can be downloaded at `https://doi.org/10.15780/G2RP4Q`. CPC Global Temperature data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at `https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html`. CPC Global Unified Precipitation data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at `https://www.esrl.noaa.gov/psd/`. The High-resolution global topographic index (TI) dataset is available at `https://doi.org/10.5285/6b0c4358-2bf3-4924-aa8f-793d468b92be`. The Global Soil Dataset for Earth System Modeling (GSD) can be downloaded at `http://globalchange.bnu.edu.cn/research/soilw`. The Global Depth to Bedrock Dataset for Earth System Modeling is available at `http://globalchange.bnu.edu.cn/research/dtb.jsp`. The GLobal HYdrogeology MaPS data is available at `https://doi.org/10.5683/SP2/TTJNIU`. The FAO Global Land Cover (GLC-SHARE) Beta-Release 1.0 database available at `http://www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/en/c/1036355/`. The Gridded Population of the World (GPW) Version 4 Release 11 population count data is available at `https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev11`. The Anthropogenic land-use estimates for the Holocene; HYDE 3.2 dataset is available at `https://doi.org/10.17026/dans-25g-gez3`. Gridded global datasets for Gross Domestic Product and Human Development Index (GDPHDI) over 1990-2015 are available at `https://doi.org/10.5061/dryad.dk1j0`. The CESM Large Ensemble dataset is available at `https://www.earthsystemgrid.org/` and the authors acknowledge CESM Large Ensemble Community Project and supercomputing resources provided by NSF/CISL/Yellowstone. The data supplement to this article is available at `https://doi.org/10.5281/zenodo.3873422` and contains datasets, fitted statistical models, and an analysis script (Carozza & Boudreault, 2020).

This work was supported by Mitacs through the Mitacs Accelerate program. This work was partially funded by AXA XL, the property & casualty and specialty risk division of AXA. We acknowledge the support of the Fonds de recherche du Québec – Nature et technologies (FRQNT), [funding reference number 119908]. Cette recherche a été financée par le Fonds de recherche du Québec – Nature et technologies (FRQNT), [numéro de référence 119908]. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number PDF-502939-2017]. Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), [numéro de référence PDF-502939-2017].

Both authors contributed equally to this work.

The authors declare that there is no conflict of interest regarding the publication of this article.

The authors would like to thank Tom Philp from AXA-XL for his support and feedback on this work, Robert Erhardt for discussions on the model methodology, and Manuel Grenier for comments on the manuscript.

References

Bank of England. (2019). *The 2021 biennial exploratory scenario on the financial risks from climate change.* Bank of England.

Barredo, J. I. (2009). Normalised flood losses in europe: 1970–2006. *Natural Hazards and Earth System Sciences*, *9*(1), 97–104. Retrieved from `https://www.nat-hazards-earth-syst-sci.net/9/97/2009/` doi: 10.5194/nhess-9-97-2009

Barthel, F., & Neumayer, E. (2011, Dec). A trend analysis of normalized insured damage from natural disasters. *Climatic Change*, *113*(2), 215–237. Retrieved from `http://dx.doi.org/10.1007/s10584-011-0331-2` doi: 10.1007/s10584-011-0331-2

Bergmeir, C., & Benítez, J. M. (2012). Neural networks in R using the stuttgart neural network simulator: RSNNS. *Journal of Statistical Software*, *46*(7), 1–26. Retrieved from `http://www.jstatsoft.org/v46/i07/`

Bevin, K. (2012). Down to basics: Runoff processes and the modelling process. In *Rainfall-runoff modelling* (p. 1-23). John Wiley & Sons, Ltd. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119951001.ch1` doi: 10.1002/9781119951001.ch1

Brakenridge, G. (2010). *Global active archive of large flood events.* Dartmouth Flood Observatory, University of Colorado. Retrieved from `http://floodobservatory.colorado.edu/Archives/index.html` (Accessed: 2018-5-22)

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi: 10.1023/A:1010933404324

Carozza, D. A., & Boudreault, M. (2020, June). *A global flood risk modeling framework built with climate models and machine learning - Submission - Data Supplement.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.3873422` doi: 10.5281/zenodo.3873422

CIESIN. (2018). *Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11.* Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). (Accessed: 2019-11-22) doi: 10.7927/H4JW8BX5

de Bruijn, K. M., Diermanse, F. L. M., & Beckers, J. V. L. (2014). An advanced method for flood risk analysis in river deltas, applied to societal flood fatality risk in the netherlands. *Natural Hazards and Earth System Sciences*, *14*(10), 2767–2781. Retrieved from `https://www.nat-hazards-earth-syst-sci.net/14/2767/2014/` doi: 10.5194/nhess-14-2767-2014

Dewan, T. H. (2015). Societal impacts and vulnerability to floods in bangladesh and nepal. *Weather and Climate Extremes*, *7*, 36 - 42. Retrieved from `http://www.sciencedirect.com/science/article/pii/S2212094714000930` (SI: IGBP APN) doi: https://doi.org/10.1016/j.wace.2014.11.001

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

Dowle, M., & Srinivasan, A. (2018). data.table: Extension of 'data.frame' [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=data.table` (R package version 1.11.8)

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. doi: 10.1080/23754931.2015.1014272

Falter, D., Dung, N., Vorogushyn, S., Schröter, K., Hundecha, Y., Kreibich, H., ... Merz, B. (2016). Continuous, large-scale simulation model for flood risk assessments: proof-of-concept. *Journal of Flood Risk Management*, *9*(1), 3-

21. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/jfr3.12105` doi: 10.1111/jfr3.12105

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., … Alsdorf, D. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, *45*(2). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005RG000183` doi: 10.1029/2005RG000183

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, *27*(8), 861 - 874. (ROC Analysis in Pattern Recognition) doi: https://doi.org/10.1016/j.patrec.2005.10.010

FRBSF. (2019). Strategies to address climate change risk in low- and moderate-income communities. *Federal Reserve Bank of San Francisco Community Development Innovation Review*, *14*, 001–168. doi: 10.24148/cdir2019-01

Funk, C. (2015). *CHIRPS v2.0.* Climate Hazards Group. doi: 10.15780/G2RP4Q

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., … Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*.

Gall, M., Borden, K. A., & Cutter, S. L. (2009). When do losses count? *Bulletin of the American Meteorological Society*, *90*(6), 799-810. Retrieved from `https://doi.org/10.1175/2008BAMS2721.1` doi: 10.1175/2008BAMS2721.1

Giraud, T., & Lambert, N. (2016, aug). cartography: Create and integrate maps in your r workflow. *The Journal of Open Source Software*, *1*(4). Retrieved from `http://dx.doi.org/10.21105/joss.00054` doi: 10.21105/joss.00054

Gleeson, T., Moosdorf, N., Hartmann, J., & van Beek, L. P. H. (2014). A glimpse beneath earth's surface: Global hydrogeology maps (glhymps) of permeability and porosity. *Geophysical Research Letters*, *41*(11), 3891-3898. (Accessed data: 2018-5-7) doi: 10.1002/2014GL059856

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., & Dorigo, W. (2019). Evolution of the esa cci soil moisture climate data records and their underlying merging methodology. *Earth System Science Data*, *11*(2), 717–739. Retrieved from `https://www.earth-syst-sci-data.net/11/717/2019/` doi: 10.5194/essd-11-717-2019

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning.* Springer New York. doi: 10.1007/978-0-387-84858-7

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., & Piontek, F. (2013). A trend-preserving bias correction; the isi-mip approach. *Earth System Dynamics*, *4*(2), 219–236. Retrieved from `https://www.earth-syst-dynam.net/4/219/2013/` doi: 10.5194/esd-4-219-2013

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., … Cudennec, C. (2013). A decade of predictions in ungauged basins (pub)—a review. *Hydrological Sciences Journal*, *58*(6), 1198-1255. Retrieved from `https://doi.org/10.1080/02626667.2013.803183` doi: 10.1080/02626667.2013.803183

Hu, P., Zhang, Q., Shi, P., Chen, B., & Fang, J. (2018). Flood-induced mortality across the globe: Spatiotemporal pattern and influencing factors. *Science of The Total Environment*, *643*, 171 - 182. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0048969718322745` doi: https://doi.org/10.1016/j.scitotenv.2018.06.197

Hunziker, P. (2017). velox: Fast raster manipulation and extraction [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=velox` (R package version 0.2.0)

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., … Marshall, S. (2013). The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339-1360. Retrieved from `https://doi.org/10.1175/BAMS-D-12-00121.1` doi: 10.1175/BAMS-D-12-00121.1

Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J. C. J. H., Mechler, R., Botzen, W. J. W., … Ward, P. J. (2014, Mar). Increasing stress on disaster-risk finance due to large floods. *Nature Climate Change*, *4*(4), 264–268. Retrieved from `http://dx.doi.org/10.1038/nclimate2124` doi: 10.1038/nclimate2124

Jongman, B., Ward, P. J., & Aerts, J. C. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, *22*(4), 823 - 835. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0959378012000830` doi: https://doi.org/10.1016/j.gloenvcha.2012.07.004

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., … Vertenstein, M. (2015, August). The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society*, *96*, 1333-1349. (Data accessed: 2018-6-29) doi: 10.1175/BAMS-D-13-00255.1

Klein Goldewijk, K. (2017). *Anthropogenic land-use estimates for the Holocene; HYDE 3.2.* Data Archiving and Networked Services (DANS). (Accessed: 2020-01-17) doi: 10.17026/DANS-25G-GEZ3

Klein Goldewijk, K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic land use estimates for the holocene – hyde 3.2. *Earth System Science Data*, *9*(2), 927–953. doi: 10.5194/essd-9-927-2017

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*(180004). doi: 10.1038/sdata.2018.4

Kummu, M., Taka, M., & Guillaume, J. H. A. (2019). *Gridded global datasets for Gross Domestic Product and Human Development Index over 1990-2015, v2.* Dryad, Dataset. (Accessed: 2018-11-14) doi: 10.5061/dryad.dk1j0

Latham, J., Cumani, R., Rosati, I., & Bloise, M. (2014). FAO Global Land Cover (GLC-SHARE) Beta-Release 1.0 Database. *FAO Land and Water Division*. (Accessed data: 2018-3-30)

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., … Slater, A. G. (2011). Parameterization improvements and functional and structural advances in version 4 of the community land model. *Journal of Advances in Modeling Earth Systems*, *3*(1). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011MS00045` doi: 10.1029/2011MS00045

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171-2186. (Data accessed: 2018-4-20) doi: 10.1002/hyp.9740

Li, S., Kang, S., Zhang, L., Zhang, J., Du, T., Tong, L., & Ding, R. (2016). Evaluation of six potential evapotranspiration models for estimating crop potential and actual evapotranspiration in arid regions. *Journal of Hydrology*, *543*, 450 - 461. doi: https://doi.org/10.1016/j.jhydrol.2016.10.022

Liaw, A., & Wiener, M. (2002). Classification and regression by randomforest. *R News*, *2*(3), 18-22. Retrieved from `https://CRAN.R-project.org/doc/Rnews/`

Marthews, T., Dadson, S., Lehner, B., Abele, S., & Gedney, N. (2015a). *High-resolution global topographic index values.* NERC Environmental Information Data Centre. (Accessed: 2018-8-6) doi: 10.5285/6b0c4358-2bf3-4924-aa8f-793d468b92be

Marthews, T., Dadson, S. J., Lehner, B., Abele, S., & Gedney, N. (2015b). High-resolution global topographic index values for use in large-scale hydrological modelling. *Hydrology and Earth System Sciences*, *19*(1), 91–104. doi:

10.5194/hess-19-91-2015

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. doi: 10.1007/BF02478259

Meløysund, V., Leira, B., Høiseth, K. V., & Lisø, K. R. (2007). Predicting snow density using meteorological data. *Meteorological Applications*, *14*(4), 413-423. Retrieved from `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/met.40` doi: 10.1002/met.40

Microsoft, C., & Weston, S. (2017). dosnow: Foreach parallel adaptor for the 'snow' package [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=doSNOW` (R package version 1.0.16)

Munoz, S. E., & Dee, S. G. (2017, May). El niño increases the risk of lower mississippi river flooding. *Scientific Reports*, *7*(1). doi: 10.1038/s41598-017-01919-6

Neale, R., Chen, C.-C., & Gettelman, A. (2012). Description of the ncar community atmosphere model (cam5), technical report ncar/tn-486+st. *National Center for Atmospheric Research, Boulder, Colorado*, 268. Retrieved from `https://doi.org/10.1175/1520-0469(1999)056<0642:TOSEOT>2.0.CO;2`

Neumayer, E., & Barthel, F. (2011). Normalizing economic loss from natural disasters: A global analysis. *Global Environmental Change*, *21*(1), 13 - 24. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0959378010001019` doi: https://doi.org/10.1016/j.gloenvcha.2010.10.004

Neuwirth, E. (2014). Rcolorbrewer: Colorbrewer palettes [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=RColorBrewer` (R package version 1.1-2)

Nie, W., Yuan, Y., Kepner, W., Nash, M. S., Jackson, M., & Erickson, C. (2011). Assessing impacts of landuse and landcover changes on hydrology for the upper san pedro watershed. *Journal of Hydrology*, *407*(1), 105 - 114. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0022169411004549` doi: https://doi.org/10.1016/j.jhydrol.2011.07.012

OECD. (2016). *Financial management of flood risk.* Author. doi: 10.1787/9789264257689-en

Ouazad, A., & Kahn, M. (2019, Sep). Mortgage finance in the face of rising climate risk. *National Bureau of Economic Research Working Paper*. doi: 10.3386/w26322

Paprotny, D., Sebastian, A., Morales-Nápoles, O., & Jonkman, S. N. (2018, May). Trends in flood losses in europe over the past 150 years. *Nature Communications*, *9*(1). Retrieved from `http://dx.doi.org/10.1038/s41467-018-04253-1` doi: 10.1038/s41467-018-04253-1

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*. Retrieved from `https://journal.r-project.org/archive/2018/RJ-2018-009/index.html`

Pebesma, E. J., & Bivand, R. S. (2005, November). Classes and methods for spatial data in R. *R News*, *5*(2), 9–13. Retrieved from `https://CRAN.R-project.org/doc/Rnews/`

Perrin, C., Michel, C., & Andréassian, V. (2013). A set of hydrological models. In *Mathematical models* (p. 493-509). John Wiley & Sons, Ltd. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118557853.ch16` doi: 10.1002/9781118557853.ch16

Powers, D. (2011). Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, *2*(1), 37 - 63.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rumsey, C. A., Miller, M. P., Susong, D. D., Tillman, F. D., & Anning, D. W.

(2015). Regional scale estimates of baseflow and factors influencing baseflow in the upper colorado river basin. *Journal of Hydrology: Regional Studies*, *4*, 91 - 107. Retrieved from http://www.sciencedirect.com/science/article/pii/S2214581815000373 doi: https://doi.org/10.1016/j.ejrh.2015.04.008

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. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR016954 doi: 10.1002/2015WR016954

Shangguan, W., Dai, Y., Duan, Q., Liu, B., & Yuan, H. (2014). A global soil data set for earth system modeling. *Journal of Advances in Modeling Earth Systems*, *6*(1), 249-263. (Accessed data: 2018-8-6) doi: 10.1002/2013MS000293

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., & Dai, Y. (2017). Mapping the global depth to bedrock for land surface modeling. *Journal of Advances in Modeling Earth Systems*, *9*(1), 65-88. (Accessed data: 2018-8-6) doi: 10.1002/2016MS000686

Shi, W. (2007). Global daily surface air temperature analyses. *Presentation*. Retrieved from ftp://ftp.cpc.ncep.noaa.gov/precip/PEOPLE/wd52ws/global_temp/CPC-GLOBAL-T.pdf (Data Accessed: 2018-6-15)

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). Rocr: visualizing classifier performance in r. *Bioinformatics*, *21*(20), 7881. Retrieved from http://rocr.bioinf.mpi-sb.mpg.de

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). Grace measurements of mass variability in the earth system. *Science*, *305*(5683), 503–505. Retrieved from https://science.sciencemag.org/content/305/5683/503 doi: 10.1126/science.1099192

Task Force on Climate-related Financial Disclosures. (2017). *Final report: Recommendations of the task force on climate-related financial disclosures.* Task Force on Climate-related Financial Disclosures.

Vorogushyn, S., Bates, P. D., de Bruijn, K., Castellarin, A., Kreibich, H., Priest, S., … Merz, B. (2018). Evolutionary leap in large-scale flood risk assessment needed. *WIREs Water*, *5*(2), e1266. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1266 doi: 10.1002/wat2.1266

Ward, P. J., Jongman, B., Kummu, M., Dettinger, M. D., Sperna Weiland, F. C., & Winsemius, H. C. (2014). Strong influence of el niño southern oscillation on flood risk around the world. *Proceedings of the National Academy of Sciences*, *111*(44), 15659–15664. doi: 10.1073/pnas.1409822111

Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., … Winsemius, H. C. (2015, Jul). Usefulness and limitations of global flood risk models. *Nature Climate Change*, *5*(8), 712–715. Retrieved from http://dx.doi.org/10.1038/nclimate2742 doi: 10.1038/nclimate2742

Ward, P. J., Winsemius, H. C., Kuzma, S., Bierkens, M. F., Bouwman, A., De Moel, H., … Luo, T. (2020). Aqueduct floods methodology. *Technical Note*. Retrieved from www.wri.org/publication/aqueduct-floods-methodology

Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J., & Bouwman, A. (2013). A framework for global river flood risk assessments. *Hydrology and Earth System Sciences*, *17*(5), 1871–1892. Retrieved from https://www.hydrol-earth-syst-sci.net/17/1871/2013/ doi: 10.5194/hess-17-1871-2013

Wolock, D. M., Winter, T. C., & McMahon, G. (2004, Apr). Delineation and evaluation of hydrologic-landscape regions in the united states using geographic information system tools and multivariate statistical analyses. *Environmental Management*, *34*(S1), S71–S88. Retrieved from http://dx.doi.org/10.1007/s00267-003-5077-9 doi: 10.1007/s00267-003-5077-9

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., & Liu, C.

(2007). A gauge-based analysis of daily precipitation over east asia. *Journal of Hydrometeorology*, *8*(3), 607-626. (Data Accessed: 2018-6-18) doi: 10.1175/JHM583.1

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*, *47*(4). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010WR009726` doi: 10.1029/2010WR009726

Yu, Z., Gburek, W. J., & Schwartz, F. W. (2000). Evaluating the spatial distribution of water balance in a small watershed, pennsylvania. *Hydrological Processes*, *14*(5), 941-956. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291099-1085%2820000415%2914%3A5%3C941%3A%3AAID-HYP2%3E3.0.CO%3B2-S` doi: 10.1002/(SICI)1099-1085(20000415)14:5⟨941::AID-HYP2⟩3.0.CO;2-S

Database	Variable	Data Reference
	Response variables	
DFO - Dartmouth Flood Observatory Global Active Archive of Large Flood Events		Brakenridge (2010)
	Flood event	
	Population displaced	
	Predictor variables	
CHIRPS - Climate Hazards group Infrared Precipitation with Stations		Funk (2015)
	Precipitation (Daily, 1981-present) (0.05°) For sites between $[-50°, +50°]$ latitude	
CPC Precipitation - CPC Global Unified Gauge-Based Analysis of Daily Precipitation		Xie et al. (2007)
	Precipitation (Daily, 1979-present) (0.5°) For sites outside $[-50°, +50°]$ latitude	
CPC Temperature - CPC Global Daily Temperature		Shi (2007)
	Temperature (Daily, 1979-present) (0.5°)	
HydroBASINS		Lehner and Grill (2013)
	Aspect	
	Elevation	
	Hillslope	
	Slope	
	Topographic position index	
	Latitude of watershed centroid	
	Longitude of watershed centroid	
TI - High-resolution global topographic index values		Marthews et al. (2015a)
	Topographic index	
GLC-SHARE - Global Land Cover SHARE		Latham et al. (2014)
	Artificial surfaces	
	Cropland	
	Grassland	
	Tree covered areas	
	Shrubs covered areas	
	Herbaceous vegetation	
	Aquatic or regularly flooded	
	Mangroves	
	Sparse vegetation	
	Bare soil	
	Snow and glaciers	
	Water bodies	
GSD - Global Soil Dataset for Earth System Modeling		Shangguan et al. (2014)
	Sand	
	Silt	
	Clay	
	Gravel	
	Bulk soil density	
GDB - Global depth to bedrock		Shangguan et al. (2017)
	Global depth to bedrock	
GLHYMPS 2.0 - GLobal HYdrogeology MaPS 2.0		Gleeson et al. (2014)
	Bedrock porosity	
	Bedrock permeability	
GPW - Gridded Population of the World		CIESIN (2018)
	Population count (2.5 min)	
HYDE3.2 - Anthropogenic land-use estimates for the Holocene		Klein Goldewijk (2017)
	Population count (5 min)	
GDPHDI - Gridded global datasets for GDP and HDI over 1990-2015		Kummu et al. (2019)
	Gross domestic product (PPP) per capita (5 min)	

Table 1. Response and predictor variables used in statistical fits of occurrence and impact models.

Variable Long Name	Variable	Model Quantity
Atmospheric rain	RAIN	Precipitation
Atmospheric snow	SNOW	Precipitation
Reference height temperature	TREFHT	Temperature

Table 2. Forcing climate variables from the NCAR CESM Large Ensemble (CESM-LE) Community Project.

Step	Description	Duration	Repetitions
1	Occurrence - Simulation	1.34 s / simulation	625
2	Intensity - Distribution mean	36.9 min / member	40
3	Intensity - Distribution standard deviation	3.1 s	1
4	Intensity - Merge mean and standard deviation	6.8 s	1
5	Simulate intensity and combine with occurrence	3.75 s / simulation	50
6	Calculate standard deviation correction	1.68 min	1
7	Repeat step 4 with corrected standard deviation	6.8 s	1
8	Simulate corrected intensity and combine with occurrence	3.75 s / simulation	625

Table 3. Duration of simulation computation steps.

Model	Global Fit	HS2 Fit
LR	0.735	0.789
RF	0.788	0.792
NN-24	0.787	-
NN-12-12	0.781	-
NN-24-24	0.776	-
NN-12	-	0.777

Table 4. Out-of-sample area under the ROC curve for occurrence models and fitting approaches considered.

	Mean decrease in accuracy	Mean decrease in Gini impurity index
1	GDP per capita	Population density
2	Population density	Precipitation (previous 7 days)
3	Temperature (annual mean)	GDP per capita
4	Precipitation (previous 7 days)	Precipitation (previous 8-30 days)
5	Topographic position index	Precipitation (previous 31-60 days)
6	Precipitation (previous 8-30 days)	Precipitation (previous 61-120 days)
7	Gravel	Temperature (annual mean)
8	Precipitation (previous 61-120 days)	Cropland
9	Aspect	Latitude
10	Porosity	Longitude

Table 5. Most significant predictors of flood occurrence ranked by two methods for the globally fitted random forest model.

Model	Global Fit	HS2 Fit
LM	0.172	0.209
RF	0.325	0.353
NN-26	0.291	-
NN-13-13	0.287	-
NN-26-26	0.303	-
NN-13	-	0.308

Table 6. Out-of-sample R^2 for impact models and fitting approaches considered.

	Increase in mean squared error	Increase in residual sum of squares
1	Precipitation (previous 8-30 days)	GDP per capita
2	GDP per capita	Temperature (previous 61-120 days)
3	Precipitation (previous 7 days)	Temperature (previous 31-60 days)
4	Population density	Temperature (previous 8-30 days)
5	Precipitation (previous 31-60 days)	Temperature (previous 7 days)
6	Precipitation (previous 61-120 days)	Population density
7	Temperature (previous 61-120 days)	Precipitation (previous 7 days)
8	Temperature (previous 7 days)	Latitude
9	Longitude	Precipitation (previous 31-60 days)
10	Temperature (previous 31-60 days)	Precipitation (previous 8-30 days)

Table 7. Most significant predictors of impact model ranked by two methods.

Supporting Information for "A global flood risk modeling framework built with climate models and machine learning"

David A. Carozza[1], Mathieu Boudreault[1]

[1]Département de mathématiques, Université du Québec à Montréal, Montréal, QC, Canada

Contents of this file

1. Text S1 to S3
2. Figures S1 to S11

Introduction

The texts below present details on the neural network model topologies considered (S1), the validation techniques used (S2), and the interpolations and extrapolations applied to the population and GDP per capita data sets (S3). The supporting figures provide a variety of content such as maps of dataset test sample sizes (Figures S1 and S2), a map of population exposure (Figure S3), validations of globally and regionally fitted occurrence models (Figures S4 and S5), maps of aggregated occurrence model validations for the logistic regression and neural network models (Figures S6 and S7), aggregated validations of the globally and regionally fitted impact models (Figures S8 and S9), and analyses of the standard deviation correction applied to the impact model (Figures S10 and S11).

Text S1 Neural Network Model Topology

For the global occurrence model, we examined 3 topologies for the NN model: 1. a single hidden layer with 24 neurons (a 38-24-2 network), 2. two hidden layers with the same total of hidden neurons (a 38-12-12-2 network), and 3. two hidden layers each with 24 neurons (a 38-24-2 network). The choice of 24 neurons for the hidden layers is based on a commonly used rule of 2/3 as the ratio of hidden to input neurons. We consider a maximum of 2000 iterations to achieve convergence of the learning algorithm. Due to the smaller sample sets for the fits over HS2 watersheds, we consider a single hidden layer with 10 neurons (a 38-10-2 network).

For the global impact model, we examined 3 topologies for the NN impact model: 1. a single hidden layer with 26 neurons (a 41-26-1 network), 2. two hidden layers with the same total of hidden neurons (a 41-13-13-1 network), and 3. two hidden layers each with 26 neurons (a 41-26-26-1 network). NN models are fit with a single hidden layer with 13 neurons in the regional fits due to smaller datasets in the HS2 watersheds.

Text S2 Validation

In addition to the ROC curve and the AUC, we here plot the precision-recall curve, which is the positive predictive value (TP/(TP+FP)) versus the true positive rate (TP/(TP+FN)), and the F_1 Score, the harmonic mean of positive predictive value and true positive rate as a function of the cutoff probability that defines a prediction as a “flood”. Since 17% of the flood occurrence observations are “flood”, one could simply always predict “no flood” and be correct 83% of the time. The precision, however, focuses on the predictions of “flood” by considering the proportion of predicted floods that are truly floods.

Text S3 Population and GDP per capita data

Population from the GPW database is available from 2000-2020. Since growth rates from the early 2000's are not appropriate to back-extrapolate as far back as the 1980's, we use the HYDE3.2 database gridded population from 1980 to relate the GPW population in 2000 to that in 1980. For each grid point (i, j), we assume that

$$\mathrm{GPW}_{i,j,1980} = \frac{\mathrm{HYDE}_{i,j,1980}}{\mathrm{HYDE}_{i,j,2000}} \mathrm{GPW}_{i,j,2000}, \tag{1}$$

which simply implies that the GPW population in 1980 ($\mathrm{GPW}_{i,j,1980}$) is linearly scaled from that of year 2000 by the HYDE3.2 population ratio between 1980 and 2000.

GDP per capita data from the GDPHDI database is available from 1990-2015. For the years 2015-2020, we forward extrapolate at each grid point assuming a continuation of exponential growth based on years 2010-2015. For the years 1980-1990, we back-extrapolate assuming exponential growth at each grid point by applying the same parameters as fitted for 1990-1995.

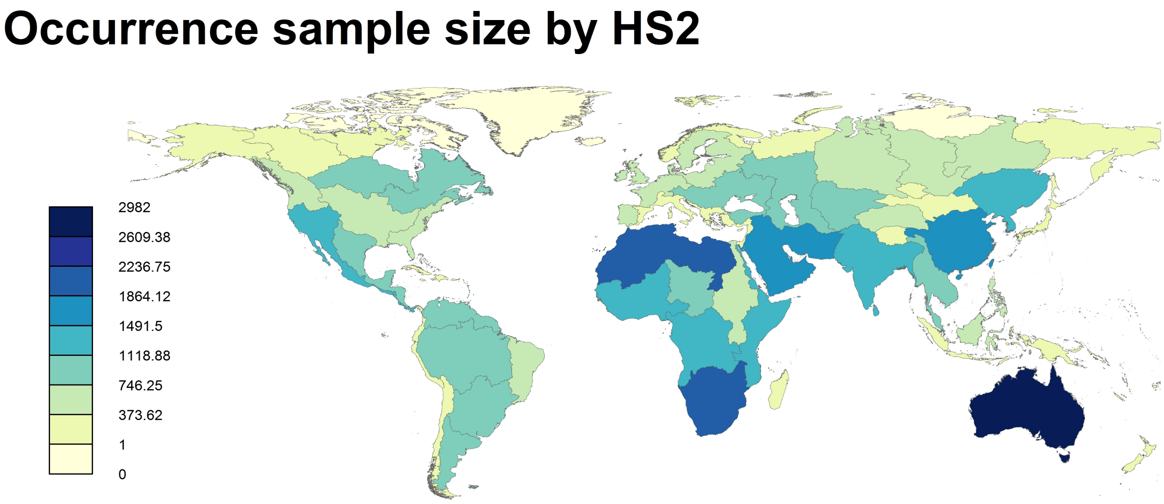


Figure S1. World map of sample size of occurrence model used in model comparison (test sets) aggregated over level 2 watersheds.

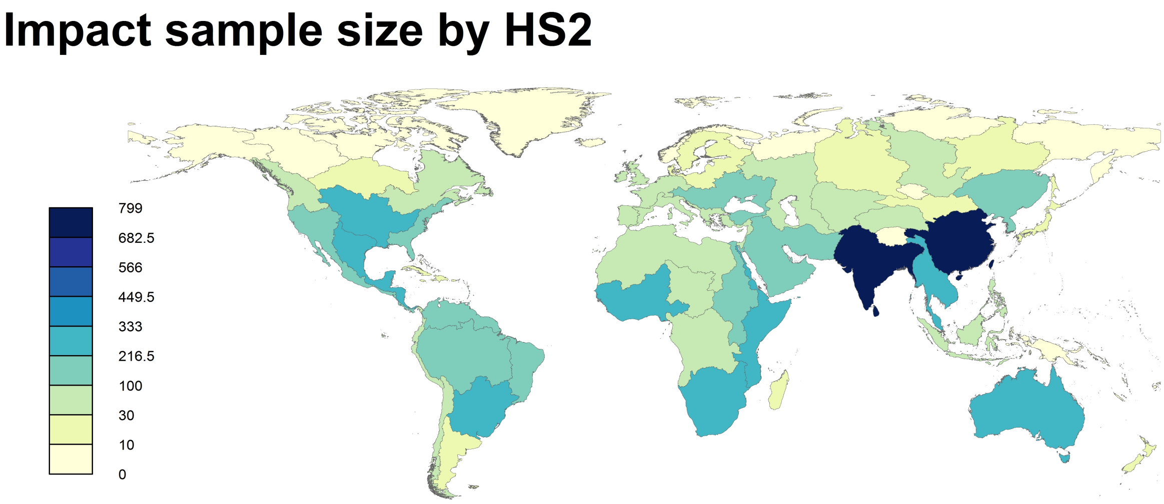


Figure S2. World map of sample size of impact model used in model comparison (test sets) aggregated over level 2 watersheds.

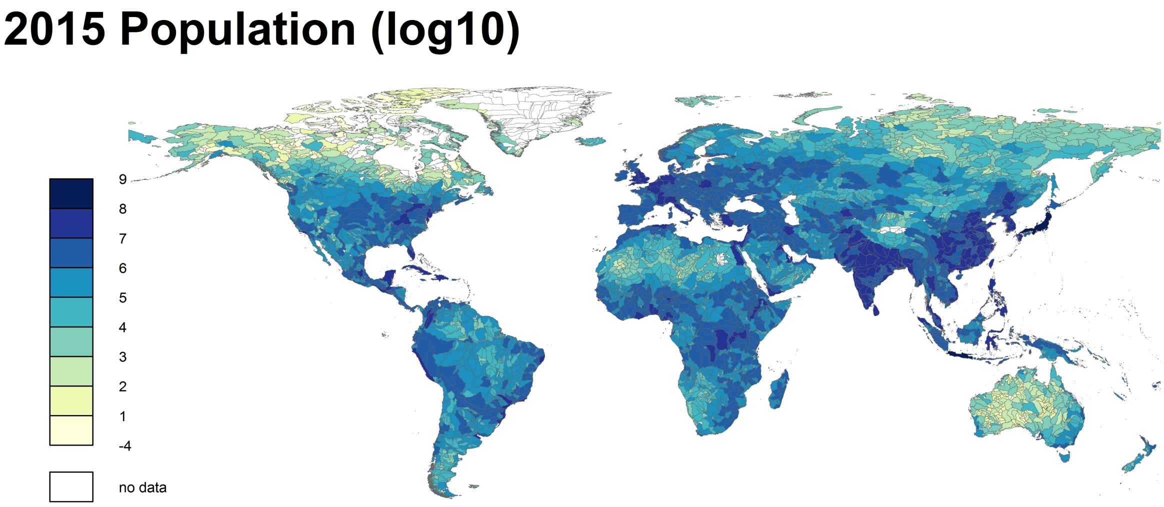


Figure S3. Population (Gridded Population of the World) in 2015 by HydroBASINS level 5 watershed.

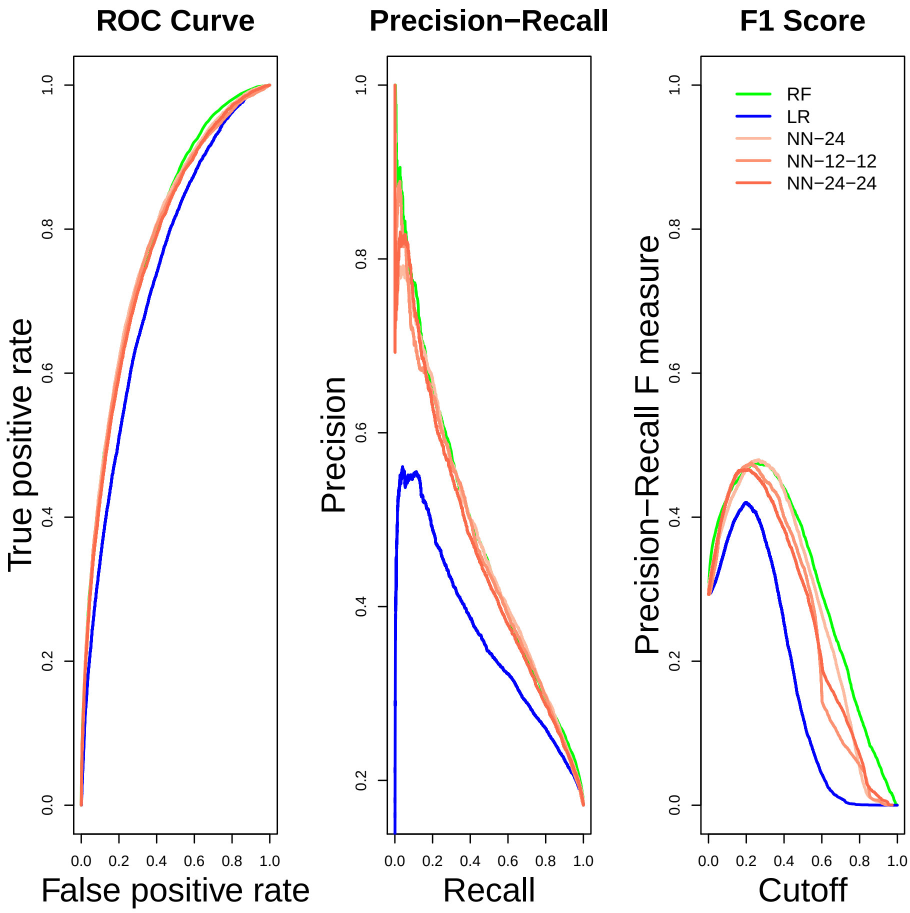


Figure S4. Occurrence validation curves fitted globally. ROC curves (left), Precision-recall curves (middle), F1 scores (right) for globally fitted models.

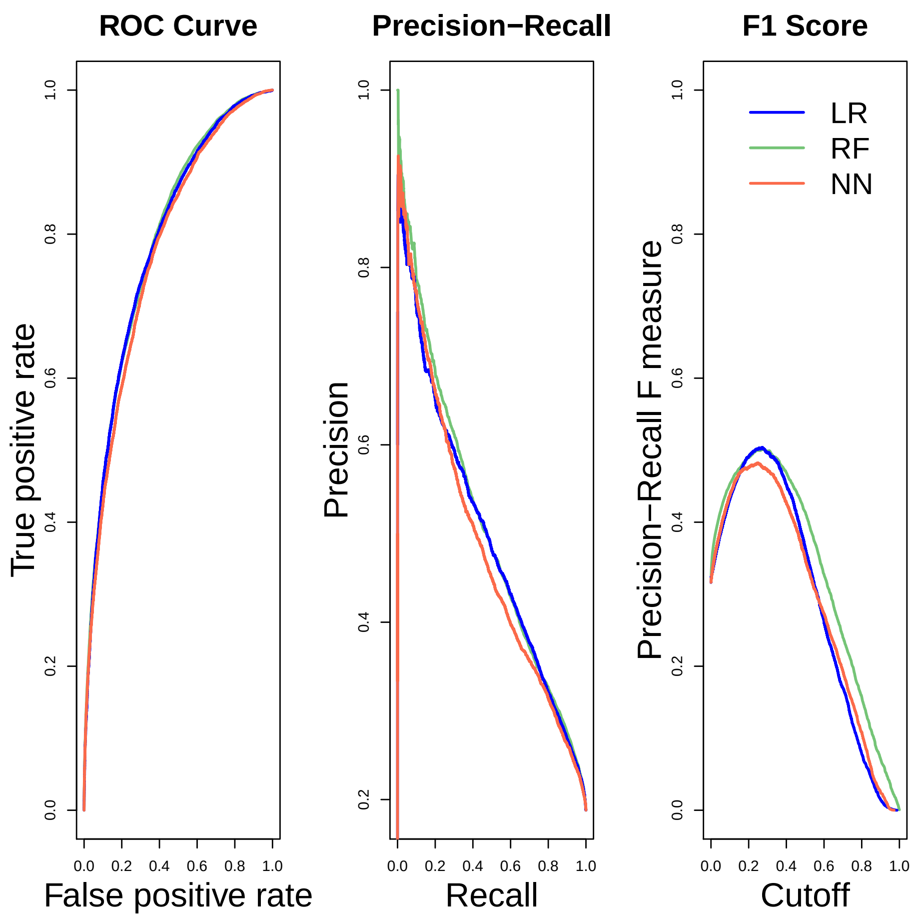


Figure S5. Occurrence validation curves fitted regionally. ROC curves (left), Precision-recall curves (middle), F1 scores (right) for models fitted over HS2 watersheds.

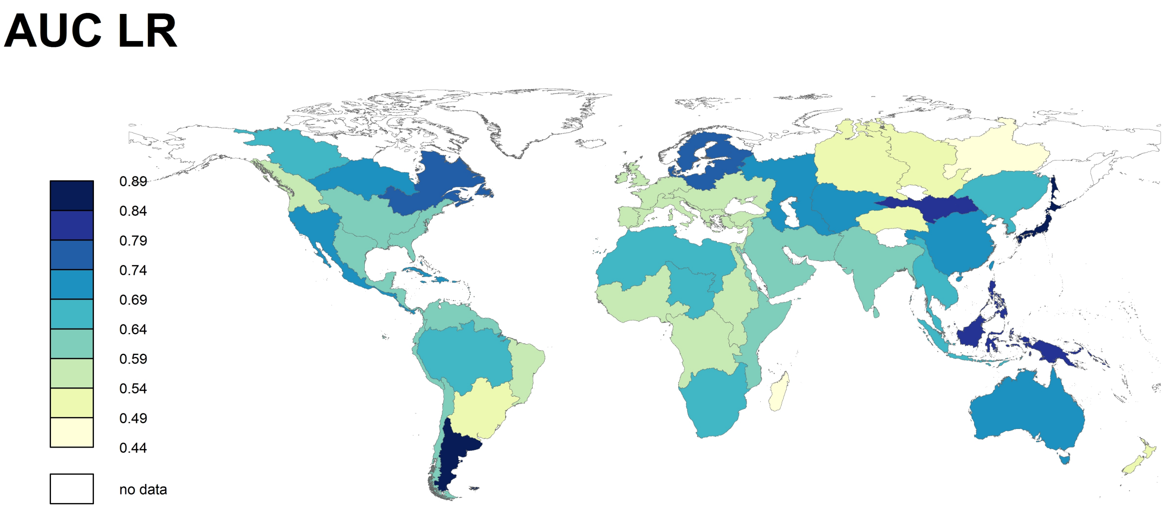


Figure S6. World map of the area under the receiver receiver operating characteristic curve, which measures probability of a model to differentiate between “flood” and “no flood”, evaluated at each level 2 watershed (HS2) for the globally fitted logistic regression model.

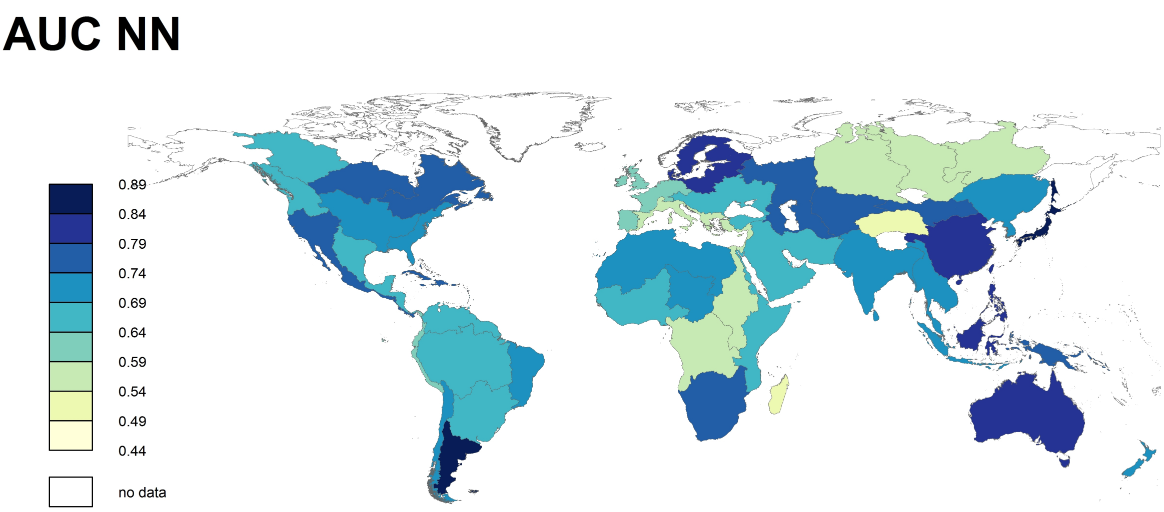
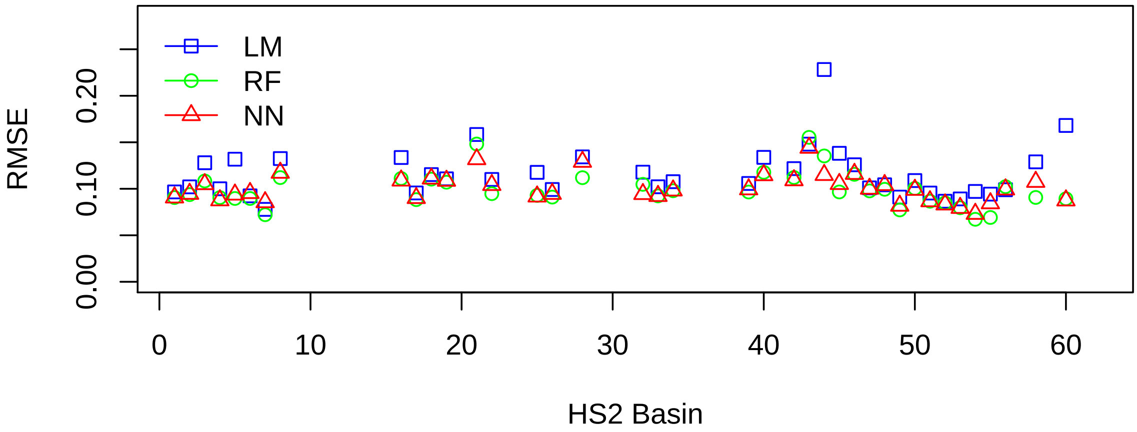


Figure S7. World map of the area under the receiver receiver operating characteristic curve, which measures probability of a model to differentiate between “flood” and “no flood”, evaluated at each level 2 watershed (HS2) for the globally fitted neural network model.



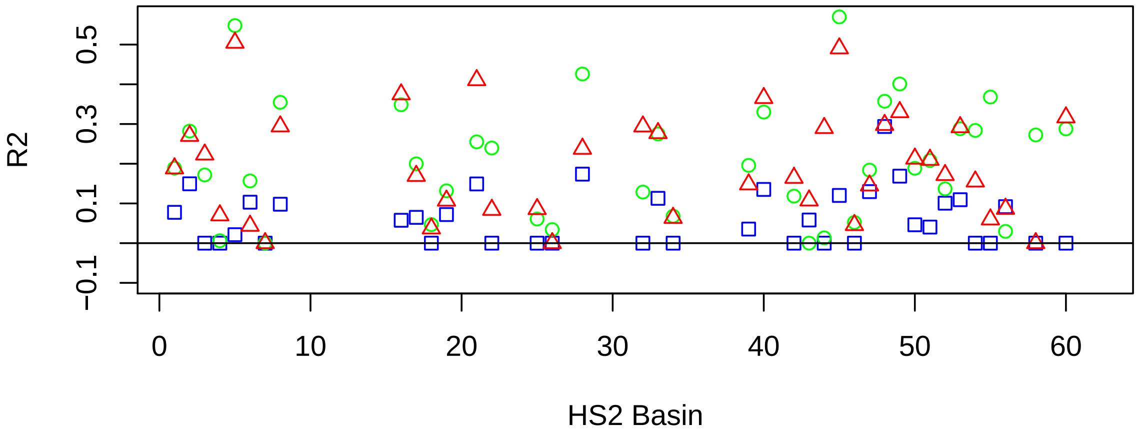


Figure S8. RMSE (top) and R-squared (bottom) by HS2 watershed for globally fitted models (LM, RF and NN).

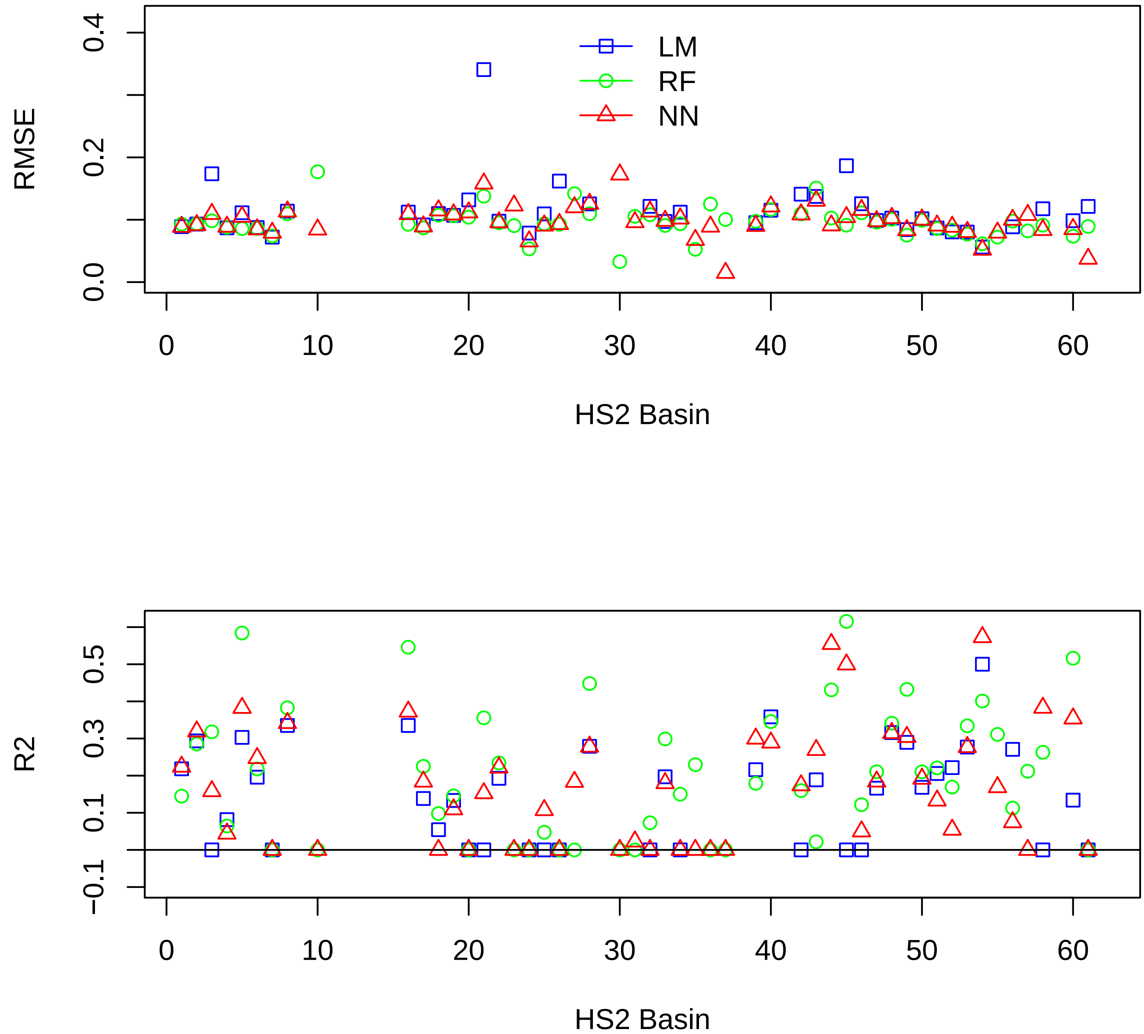


Figure S9. RMSE (top) and R-squared (bottom) by HS2 watershed for models (LM, RF and NN) fitted over subsets of watersheds.

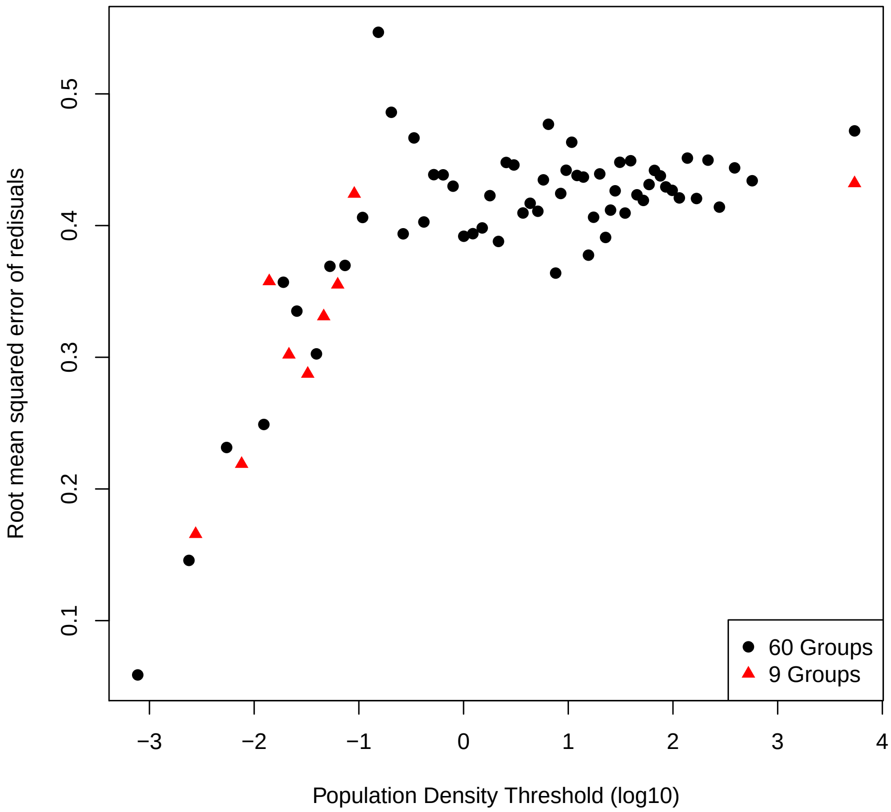


Figure S10. Comparison of the residual root mean squared errors when partitioning into 9 and 60 groups by mean watershed population density. The 60 group partitioning shows that the RMSE increases up to a threshold of -1, from where it remains flat. This allows us to use a simpler standard deviation structure with only 9 groups that preserves the increasing and then flat structure.

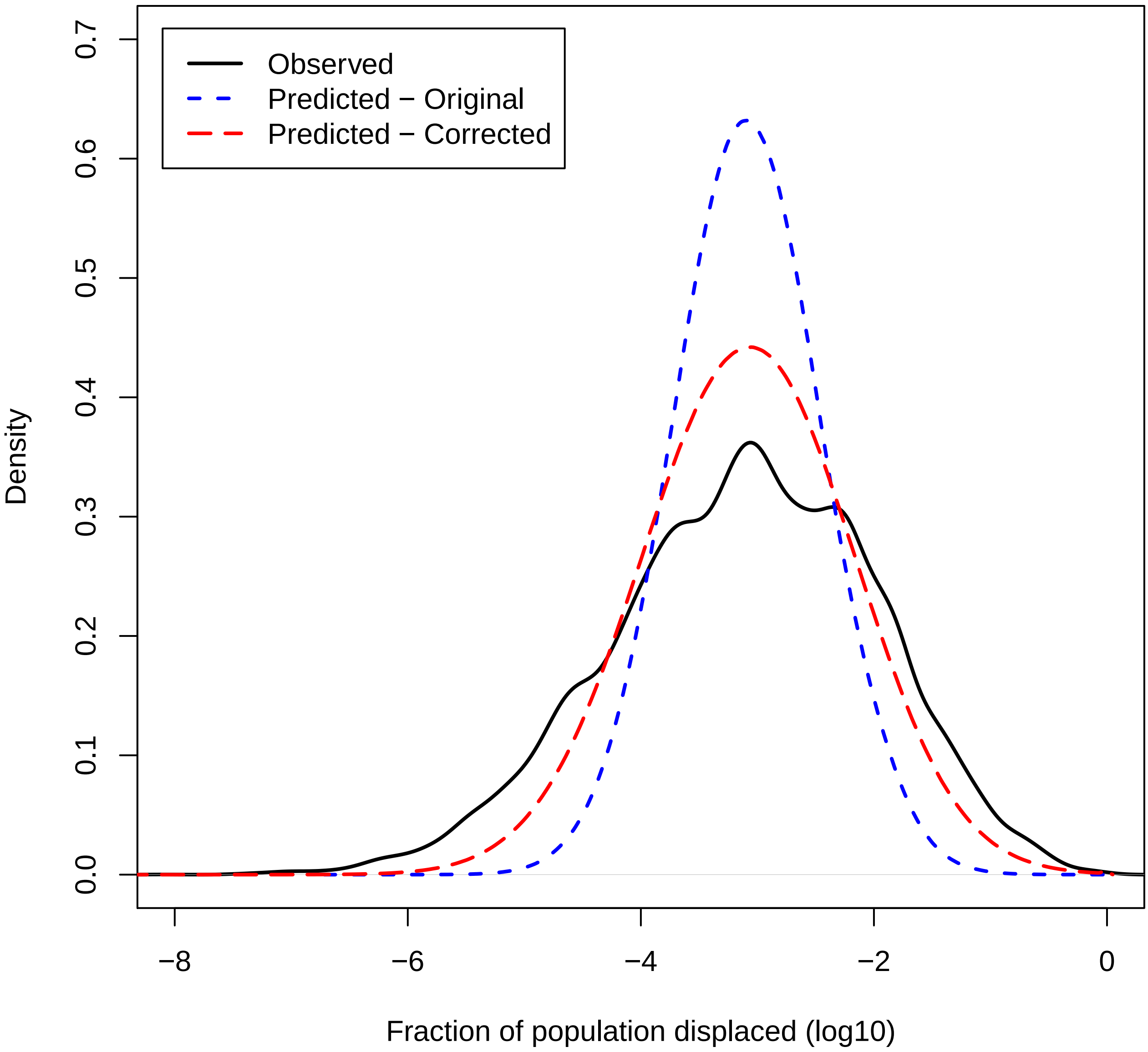


Figure S11. Comparison of observed impact distribution (log10 scale) with those simulated for the flood event catalogue. Observed impacts are solid black, whereas the original and corrected impact predictions are blue (short dash) and red (long dash), respectively.

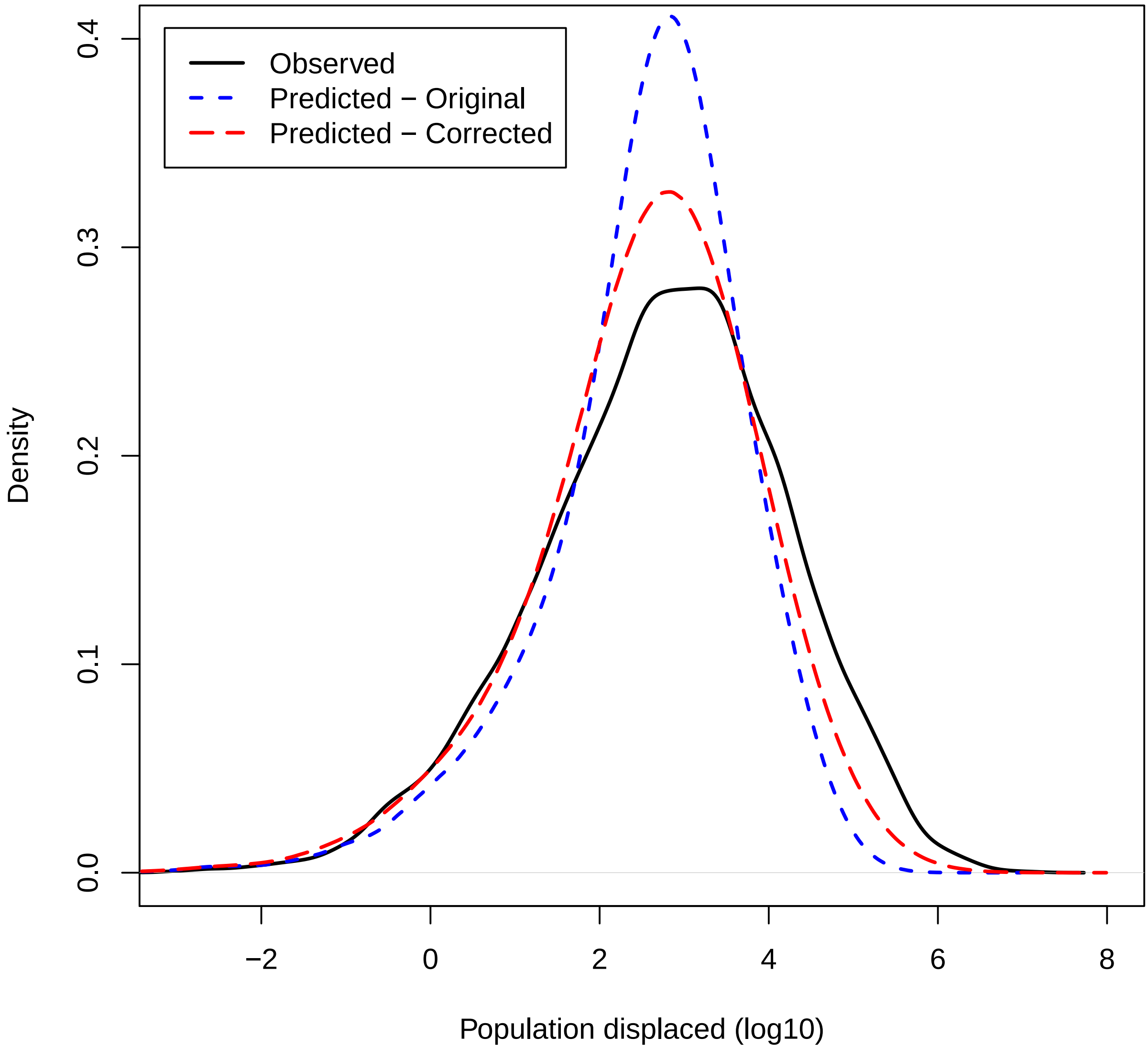


Figure S12. Comparison of observed population displaced distribution (log10 scale) with those simulated for the flood event catalogue. Observed population displaced is solid black, whereas the original and corrected impact predictions are blue (short dash) and red (long dash), respectively. Values are truncated at -3, or 0.001 people displaced. Population displaced of less than one can arise when the population is low and a low fraction of population affected is predicted. For example, a watershed with a population of 100 and and impact of 0.001 will result in 0.1 people displaced.