

# Predicting Flood Property Insurance Claims over CONUS, Fusing Big Earth Observation Data

Qing Yang, Xinyi Shen, Feifei Yang, Emmanouil N. Anagnostou, Kang He, Chongxun Mo, Hojjat Seyyedi, Albert J. Kettner, and Qingyuan Zhang

**ABSTRACT:** Each year throughout the contiguous United States (CONUS), flood hazards cause damage amounting to billions of dollars in homeowner insurance claims. As climate change threatens to raise the frequency and severity of flooding in vulnerable areas, the ability to predict the number of property insurance claims resulting from flood events becomes increasingly important to flood resilience. Based on random forest, we develop a flood property Insurance Claims model (iClaim) by fusing records from the National Flood Insurance Program (NFIP), including building locations, topography, basin morphometry, and land cover, with data from multiple sources of hydrometeorological variables, including flood extent, precipitation, and operational river-stage and oceanic water-level measurements. The model utilizes two steps—damage level classification and claim number regression—and subsampling strategies designed accordingly to reduce overfitting and underfitting caused by the flood claim samples, which are unevenly distributed and widely ranged. We evaluate the model using 446,446 grid samples identified from 589 flood events occurring from 2016 to 2019 over CONUS, overlapping 258,159 claims out of a total of 287,439 NFIP records of the same period. Our rigorous validation yields acceptable performance at the grid/event, county/event, and event accumulative level, with  $R^2$  over 0.5, 0.9, and 0.95, respectively. We conclude that the iClaim model can be used in many application scenarios, including assessing flood impact and improving flood resilience.

**KEYWORDS:** Flood events; Hydrology; Satellite observations; Surface observations; Damage assessment; Machine learning

<https://doi.org/10.1175/BAMS-D-21-0082.1>

Corresponding author: Xinyi Shen, [xinyi.shen@uconn.edu](mailto:xinyi.shen@uconn.edu)

Supplemental material: <https://doi.org/10.1175/BAMS-D-21-0082.2>

In final form 13 October 2021

©2022 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).



This article is licensed under a [Creative Commons Attribution 4.0 license](#).

**AFFILIATIONS:** **Yang**—College of Civil Engineering and Architecture, Guangxi University, Nanning, Guangxi, China, and Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut; **Shen, Yang, Anagnostou, and He**—Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut; **Mo**—College of Civil Engineering and Architecture, Guangxi University, Nanning, Guangxi, China; **Seyyedie**—Swiss Re America Holding Corporation, Schaumburg, Illinois; **Kettner**—DFO Flood Observatory, Community Surface Dynamics Modeling System, Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, Colorado; **Zhang**—Center for Satellite Applications and Research, National Oceanic and Atmospheric Administration, College Park, Maryland

Floods have caused tremendous losses in the United States, with direct annual damage, on average, increasing from approximately \$4 billion in the 1980s (in 2019 dollars) to nearly \$17 billion after 2010 (Smith 2020). With storms and sea level rise possibly intensifying and coastal development and population projected to increase (Mendelsohn et al. 2012), the exposure to floods in the contiguous United States (CONUS) is expected to increase by more than 50% by 2050 as more and more people and economic assets move to areas that are flood-prone but economically attractive for development (Wing et al. 2018).

Property inundation has a direct socioeconomic impact, often resulting in displacement or financial loss. The assessment of flood property damage in near-real time (NRT) remains challenging, however, because the interaction between hazard and loss varies in both space and time. As adopted by mechanistic modeling chains such as HAZUS-MH (Hazards U.S. Multi-Hazard; Scawthorn et al. 2006), FLEMOps (Flood Loss Estimation Model for the private sector; Thielen et al. 2008), and HEC-FIA (Hydrologic Engineering Center Flood Impact Analysis; HEC 2012), property damage is traditionally estimated by substituting flood depth into a stage-damage curve (Wagenaar et al. 2016; Merz et al. 2010). Flood depth is unavailable at large scales, however, due to the high computational expenses of hydrological/hydraulic simulation at high spatial resolution (Hardesty et al. 2018; Shen and Anagnostou 2017; Shen et al. 2017a; Khanam et al. 2021) and to large uncertainties propagated from the model parameters (Domeneghetti et al. 2012; Oubennaceur et al. 2018) and river bathymetry (Tate et al. 2015). Moreover, the empirical stage-damage curve, derived from statistical regression and restricted by strong spatial heterogeneity (Wing et al. 2020), can also lead to considerable uncertainty when applying to a large area (Garrote et al. 2016; Yildirim and Demir 2019). Hence, it remains difficult for mechanistic model chains to meet the requirement for emergency response (Neal et al. 2018).

Remote sensing imageries are ready-to-use data sources for locating floods and characterizing flood severity (Serpico et al. 2012). State-of-the-art retrieval techniques (Grimaldi et al. 2020; Shen et al. 2019a) leverage the high resolution (1–10 m) and weather penetration capability of synthetic aperture radar (SAR). Limited by the current revisiting frequency, however (e.g., around 6 days for Sentinel-1, the only freely available SAR satellite), SAR-based measures of flood extents are likely to miss peaks for small or flash flood events (Shen et al. 2019b).

In addition, in urban areas, where properties are denser, flood detection using SAR data becomes significantly more challenging, for two reasons. First, the building flood signatures (i.e., backscatter enhancement and the loss of coherence) is less reliable than the submerged flood signature because backscatter in the former may be either enhanced or damped, depending on the relative orientation of the building and radar sight direction,

while a loss of coherence may cause too many false positives if used alone (Chini et al. 2019; Li et al. 2019). Second, the comprehensive detection of urban flooding requires very high-resolution SAR data (i.e., less than 1 m) from multiple angles because tall objects like buildings might block the sight of the radar.

SAR aside, the widely used passive sensors, such as optical spectrometers and microwave radiometers, are rendered unreliable by cloud blockage (Jones 2019) and coarse resolution (Du et al. 2018), respectively.

Although remotely sensed datasets might not be used directly to detect flood damage on properties, they can be informative on flood severity. Machine-learning (ML) algorithms can synergize different datasets to characterize complex relationships between input data and results (GFDRR 2018). Previous ML studies using survey data have concluded that training data from various flood events and regions may be more effective than using information from a single event (Merz et al. 2013; Wagenaar et al. 2018). Unfortunately, postevent property damage surveys are not widely available (Gerl et al. 2016). An alternative is using flood property claims gathered by insurance sectors for financial compensation as a proxy for flood property damage; these are available over the long term in the United States (Barredo et al. 2012).

To date, only a handful of studies attempt to characterize the relationship among meteorological, geographical, and property-related factors and insurance claims (Moncoulon et al. 2014; Sørensen and Mobini 2017). These have been limited to small geographical regions, comprising one state or a few towns (Gradeci et al. 2019). In addition, existing studies rely on postdisaster surveys to locate the flood events and associated insurance claims. None has, thus far, attempted to predict event-wise flood insurance claims blindly at high spatial resolution at the national scale. Event-wise prediction poses three main challenges: first, a lack of definition of the spatiotemporal range of a flood event; second, the complex interaction between flood intensity and flood claims; and third, a lack of effective predictors.

Recently, Yang et al. (2020) describe a hybrid ML structure by combining the classification and regression that fit for characterizing the complex interaction between natural hazards and damage. A two-step method, it first predicts the damage level (i.e., the value range of the damage) of a sample through classification and then predicts its continuous damage value using the model built for the corresponding level. The two-step framework can improve prediction by having each regressor focus on a limited data range to reduce the underfit of a possibly complex and shifting relationship.

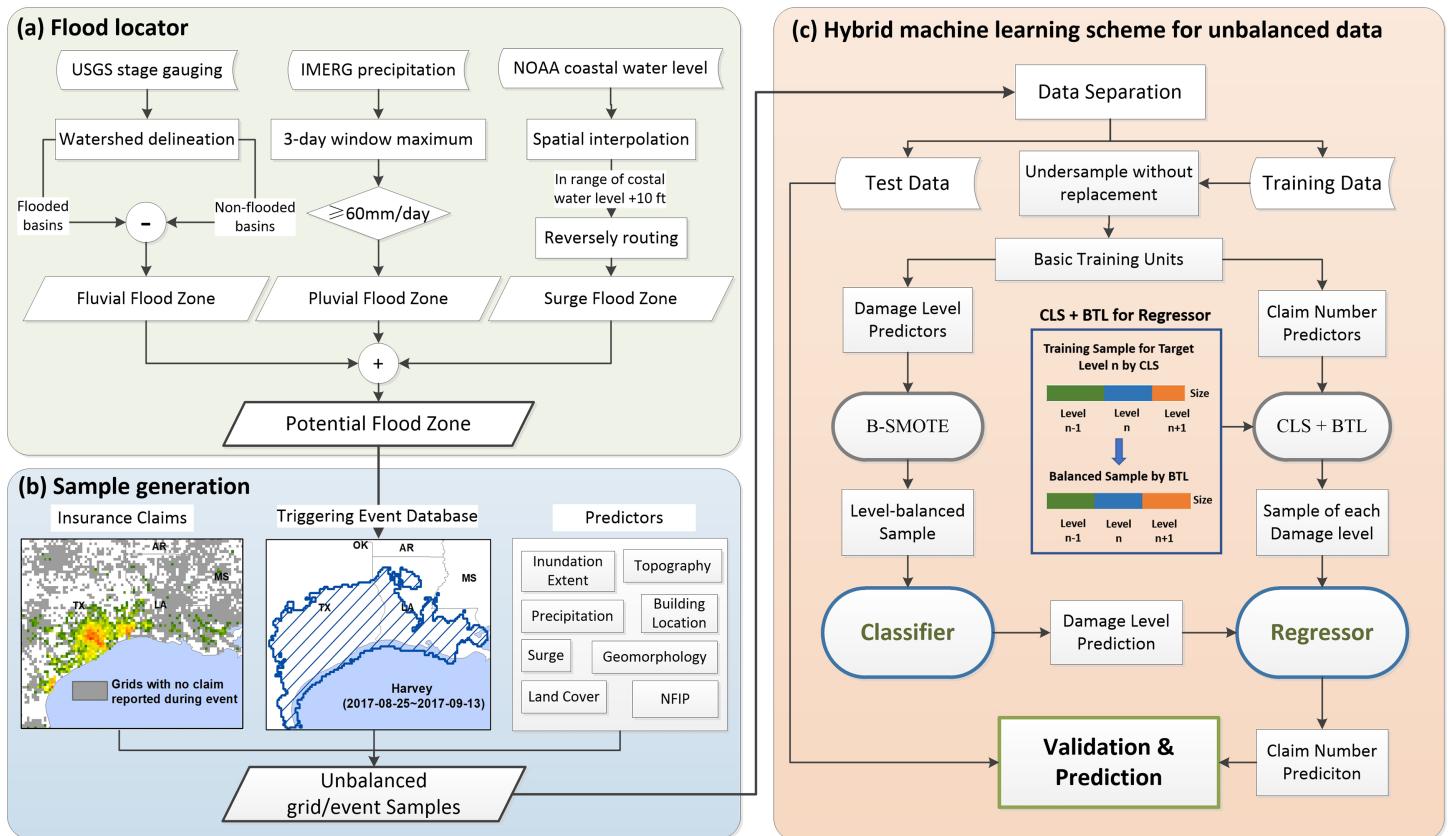
The distribution of the response variable, the flood claims, is highly unbalanced (i.e., events with high losses are significantly rarer than events with low loss); this is, unfortunately, common in natural disasters (Yang et al. 2020; Anantrasirichai et al. 2019), especially floods. This unbalanced sample can bias the model toward predicting lower impact. Previous flood risk studies address this issue mainly by undersampling the lower-impact samples (Tang et al. 2019; Woznicki et al. 2019). Training the model with a subset of data, however, can lead to underfitting (Santos et al. 2018).

For this study, we propose a hybrid ML scheme, iClaim, to predict in NRT the grid-total count of property insurance claims caused by every flood event over CONUS. The datasets and data processing are described in the second section. For a given flood event, we first locate its potential flood zone (PFZ) by using a low-cost flood locator developed in a CONUS inundation archive (Yang et al. 2021), which eliminate the need for a postflood surveyed location. Then, within the PFZ, we extract predictors and the response variable from hydrometeorological, property, land use, and topographical data and data on effective insurance policies. We develop a subsampling strategy to cope with the data imbalance issue in training, as explained in the third section. The fourth section demonstrates the validation results through rigorous leave-one-out strategies. Finally, we examine possible applications and discuss the limitations of iClaim.

## Data processing

**Event-based variable formation.** The occurrence of flooding is low across both time and space. It is, therefore, inadvisable to train a flood damage model using full time series data covering all of CONUS because the model will be overwhelmed by zero-claim samples. To focus on the potentially flooded areas of each event in our study, we only take samples and run the model within the spatial–temporal extent of the PFZ, which could be rapidly generated by a flood locator proposed by Yang et al. (2021). The original locator combines the observed networks of river discharge and standing water from precipitation to detect the fluvial and pluvial PFZs. Specifically, it utilizes a watershed algorithm to delineate the fluvial flood zone by subtracting the upstream drainage areas pouring to nonflooded stream stations from the drainage area flowing to a flooded station. The flood status of United States Geological Survey (USGS) stream stations was obtained by applying the National Weather Service (NWS) flood stage threshold, available for about 4,400 stations, to the daily mean flow. For a pluvial flooded pixel, the maximal daily accumulated precipitation needs to exceed  $60 \text{ mm day}^{-1}$  in a 3-day sliding window.

For this study, we upgrade the flood locator by including flood zones caused by storm surges, using hourly tidal station water-level measurements from the National Oceanic and Atmospheric Administration (NOAA). We delineate the surge- and tidal water-affected areas from a reverse routing algorithm (Lehner et al. 2013) by 1) interpolating the daily maximal water level at the coast at 1-km resolution and 2) routing the coastal water reversely along the flow direction to overland pixels where the elevation did not exceed 10 feet ( $\sim 3 \text{ m}$ ) above the coastal water level. The resultant daily dynamic PFZs included fluvial, pluvial, and coastal flood zones (Fig. 1a). Finally, we calculate the daily PFZ by merging spatially proximate PFZs for consecutive days, using the algorithm detailed by Yang et al. (2021). We then use the time series of PFZs for each flood event to sample the claim records as well as the predictors (Fig. 1b). Note that a PFZ could provide the maximum potential spatial extent of a flood event but not the depth.



**Fig. 1.** The framework of the flood claim prediction model for unbalanced samples. (a) The flood events locator, (b) the sample generation methods, and (c) the machine learning model structure.

**The response variable: Flood insurance claim count at the grid by event level.** We collect samples from 287,439 flood insurance claim records held by the National Flood Insurance Program (NFIP), managed by the Federal Emergency Management Agency (FEMA 2020). These claims account for \$16 billion issued in payouts from 1 January 2016 to 31 August 2019; their locations are shown in Fig. 2a. In general, most flood claims are concentrated in the coastal areas and result from events taking place between August and October (Fig. 2b), the hurricane season (Klotzbach et al. 2018). To protect privacy, the property location of each NFIP flood transaction is truncated at the source to  $0.1^\circ$  ( $\sim 10$  km) or census tract. Since it is more convenient to unify multisourced data into regular grids, we choose  $0.1^\circ \times 0.1^\circ$  grids as our spatial unit.

**Predictor selection.** We choose eight categories of predictors, including satellite-based flood extent, precipitation, coastal water level, building location, land use, topography, geomorphology, and effective policy count, as described below and listed in Table 1. It is worth mentioning that, based on the unique spatial and dynamic temporal coverage of each sample (i.e.,  $0.1^\circ$  grid per event), we use grid statistics, such as count, mean, or fraction, for most features.

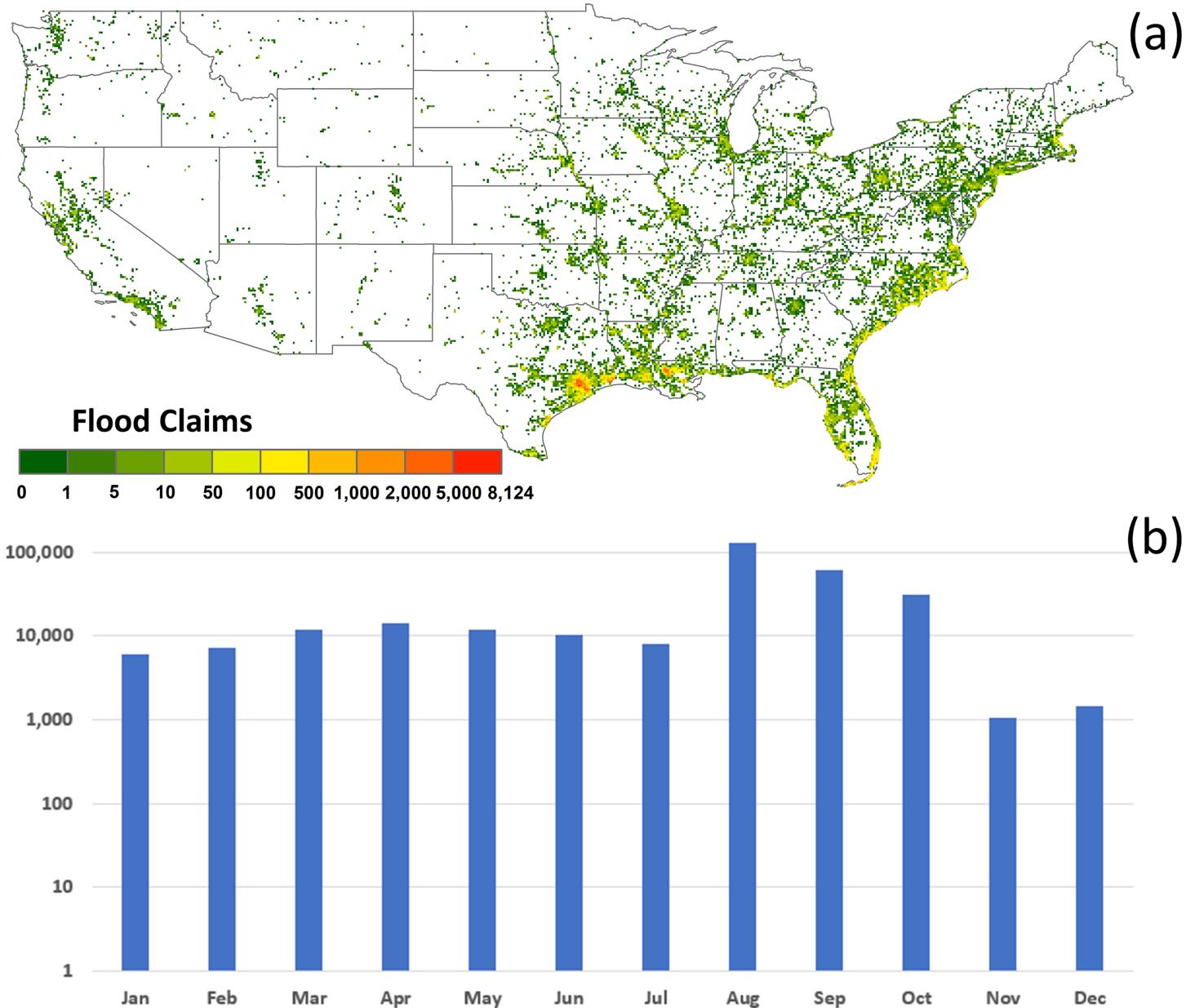


Fig. 2. The (a) spatial and (b) temporal distribution of flood insurance claims from January 2016 to August 2019.

**Table 1. Description of the 34 flood claim predictors.**

No.	Acronym	Description (unit)	Scale	Category
1, 2, 3, 4, 5	$FH_h$ , $h = 0, 1, 2, 4, 4+$	Flooded area with HAND values not exceeding $h$ meters ( $\text{km}^2$ )	$[0, +\infty)$	Flood extent
6	$P_{\text{summax5}}$	Event-maximal 5-day precipitation ( $\text{mm day}^{-1}$ )	$[0, +\infty)$	Precipitation
7, 8, 9	$PL_p$ $p = 5, 40, 80 \text{ mm day}^{-1}$	Count of daily precipitation larger than $p \text{ mm day}^{-1}$ during a period of event-maximal 7-day precipitation	$[0, +\infty)$	Precipitation
10	$P_{\text{max}}$	Event-maximal precipitation ( $\text{mm day}^{-1}$ )	$[0, +\infty)$	Precipitation
11	$P_{\text{pre}}$	7-day cumulative precipitation prior to the event ( $\text{mm day}^{-1}$ )	$[0, +\infty)$	Precipitation
12	$CWL_{\text{max}}$	Maximal coastal water level (m)	$[0, +\infty)$	Coastal water level
13, 14, 15, 16, 17	$NH_h$ , $h = 0, 1, 2, 4, 4+$	Number of properties with HAND values not exceeding $h$ meters	$[0, +\infty)$	Building location
18	$N_{\text{pro}}$	Total number of properties	$[0, +\infty)$	Building location
19	NS	Number of properties located in the SFHA zone	$[0, +\infty)$	Building location
20	LC	Length of the coastal shoreline (m)	$[0, +\infty)$	Building location
21	HC	Mean horizontal flow distance from properties to the coastal shoreline (m)	$[0, +\infty)$	Building location
22	VC	Mean vertical flow distance from properties to the coastal shoreline (m)	$[0, +\infty)$	Building location
23	$F_{\text{water}}$	Permanent water body fraction	$[0, 1]$	Land use
24	$F_{\text{wet}}$	Wetland fraction	$[0, 1]$	Land use
25	$F_{\text{imp}}$	Impervious (built-up) fraction	$[0, 1]$	Land use
26	$C_{\text{mean}}$	Mean curve number	$[30, 100]$	Land use
27	$S_{\text{mean}}$	Mean slope	$[0, \pi/2]$	Topography
28	$TWI_{\text{mean}}$	Mean topographic wetness index	$(-\infty, +\infty)$	Topography
29	$SPI_{\text{mean}}$	Mean stream power index	$(0, +\infty)$	Topography
30	$E_{\text{mean}}$	Mean properties elevation (m)	$[0, +\infty)$	Topography
31	$RE_{\text{mean}}$	Mean elongation ratio	$[0, 1]$	Geomorphology
32	$RR_{\text{mean}}$	Mean relief ratio	$(0, +\infty)$	Geomorphology
33	$DD_{\text{mean}}$	Mean drainage density	$[0, +\infty)$	Geomorphology
34	NP	Number of effective policies during the event	$[0, +\infty)$	Effective policies

**FLOOD SEVERITY.** Flood extent derived from remotely sensed imageries may not support direct counting of flooded properties, nor is it guaranteed to capture all event peaks or dynamics. The elevation statistics of inundated areas can, however, still characterize flood intensity. We overlay the event maximum inundation extent generated by Radar-Produced Inundation Diary (RAPID; Shen et al. 2019a,b; Yang et al. 2021) on the Height Above Nearest Drainage (HAND; Nobre et al. 2016) to produce grid-inundated elevation statistics for each event.

Since precipitation is the main driver for pluvial floods, we choose six statistical parameters to characterize it during and before a flood event, extracted from Integrated Multisatellite Retrievals for Global Precipitation Measurement (IMERG; Huffman et al. 2019), as listed in Table 1 (predictors 6 to 11).

Coastal inundation caused by storms and tide can be characterized by coastal water level, consisting of surge and tide data available from NOAA tidal/surge stations. We thus include the event maximum of coastal water level of the grid as the predictor of coastal flood severity.

**EXPOSURE TO FLOOD.** To represent the exposure of buildings to fluvial and pluvial floods, we compute the statistics of HAND values for the buildings as a counterpart to the statistics of inundated HAND values. We first extract the building locations from Microsoft's U.S. Building Footprints dataset (Microsoft 2018), then overlap them with HAND to compute the statistics. With the elevation statistics of both the properties and inundated areas (i.e., derived from SAR-based flood extent), the model is informed of how many buildings might be affected by the flood.

For the coastal flood exposure—since ocean flood waves propagate more rapidly through channels toward inland—we compute the average hydrological distance from properties to the nearest shoreline (predictors 21 and 22). We also use the length of the shoreline to describe the potential exposure to surge flooding.

To derive the variables related to property elevation and hydrological distance, we use the hydrography datasets from the National Hydrography Dataset Plus Version 2 (NHDPlusV2; McKay et al. 2012). Since the FEMA Special Flood Hazard Area (SFHA), including the 100-year floodplain maps, is another static measure of flood-prone areas, we also extract the total SFHA overlapped buildings as a predictor. Although SFHA only covers 61% of the area of CONUS (Woznicki et al. 2019), this predictor could improve the representation of property exposure to floods.

**LAND USE AND TOPOGRAPHY-RELATED PREDICTORS.** To characterize the capability of a grid cell to buffer or trigger a flood, we select the fractions of water, wetland, and imperviousness. We also use the average curve number derived from GCN250 to represent the grid cell's rainfall–runoff characteristics (Jaafar et al. 2019). Since topography can characterize basin response to precipitation by shaping flood-prone areas (Cook and Merwade 2009), we compute topographical predictors, such as the grid-averaged slope, property altitude, topographic wetness index, and stream power index (Tehrany et al. 2019). And since geomorphology has been proved to influence the occurrence, severity, and flashiness of flood events, we extract geomorphological features, including elongation ratio, relief ratio, and drainage density, from the global distributed basin characteristics (GDBC; Shen et al. 2016).

**THE MODULATOR.** Finally, we extract the number of effective insurance policies and building count as the modulator.

All the data sources listed in Table 1 are freely accessible to the public. Fig. ES1 in the online supplemental material demonstrates an example of a  $0.1^\circ \times 0.1^\circ$  grid with some visualizable predictors. Note that, as high-resolution inundation extents are seldom available for short-lasting flood events due to the acquisition gaps of Sentinel-1 satellites, the NaN (empty) value prevails in the samples associated with those events. Furthermore, we set the predictors related to coastal flood severity and exposure (predictors 12, 20, 21, and 22 in Table 1) to a fixed value for inland grids—i.e., the average hydrological distance to the shoreline is longer than 50 km. Specifically, we set predictors 12 and 20 to zero and predictors 21 and 22 to 999,999.

## Model setup

**The hybrid machine-learning scheme.** To reduce the underfit of high-claim samples caused by the skewed sample distribution, we propose a hybrid machine-learning scheme consisting of two steps—damage level classification and claim number regression—as depicted in Fig. 1c. We first classify a given sample to a damage level, then predict its final claim number by using the regressor associated with the predicted damage level. In other words, we train a regressor for each damage level, which help the regressor focus on a certain range of damage. We define five damage levels: level 0 refers to no claim and level  $n$  ( $0 < n \leq 5$ ) to a claim number between  $10^{n-1}$  and  $10^n$ . Figure 1c shows the overview of the model scheme.

**Subsampling strategy.** To work with unbalanced data, we develop a subsampling method, as shown in Fig. ES2. After splitting the input data into training and testing datasets, we randomly shuffle the negative samples (i.e., the zero-claim samples) of the training data into  $K$  ( $K > 1$ ) splits and make each split roughly match the size of the samples with positive claims. We form a basic training unit (BTU) by combining one negative split with all the positive samples. Since  $K > 1$ , we would finally have more than one model output to form an ensembled result, similar to that produced by the EasyEnsemble method (Liu et al. 2009). In

this way, we ensure that every zero-claim sample had been included once to eliminate the possibility of muting important samples.

To train the classifier, we balance the sample sizes between the zero-claim and all other damage levels by using the Borderline Synthetic Minority Oversampling Technique (B-SMOTE) to oversample the samples that contained actual damage inside of a BTU. Note that, since we could train multiple classifiers with  $K$  BTUs, the final damage level prediction of a sample was determined by the majority classification results by all classifiers.

Even within one damage level, we still need to reduce overfitting in training the regression model. For this hybrid scheme, we must pay special attention to avoiding discontinuous error distribution—essentially, an overfitting that comes from classification error. If each regressor is only trained with samples of its own damage level, the claim prediction of a sample classified to a wrong level will be limited to the wrong value range, which is at least one order of magnitude from the correct range. To avoid this overfitting, we develop a Cross-Level Sampling (CLS) strategy, coupled with a Balance to the Target Level (BTL) subsampling technique.

CLS utilizes samples not only within the damage level but also within adjacent levels to train a regressor. Each model, is therefore, trained using samples from three consecutive levels (two, if the model is for the lowest or highest damage level), centric at the regression model's target level. The robustness of using samples of adjacent levels can be compared to drawing a straight line; the line becomes more accurate as the two end points move farther apart from each other.

Since the sample size of each lower level was much greater than that of a higher one, we develop the BTL technique to, again, balance the sample size in the training of one regression model. Specifically, the BTL would randomly undersample the lower level and oversample (using bootstrap) the higher level to match the sample size of the target level. Theoretically, undersampling causes underfitting, while bootstrapping causes overfitting. But since we apply them only to the lower- and higher-level samples, the model performance on the target level is not affected.

**Model selection.** The random forest algorithm (RF; Breiman 2001) is a highly flexible and widely used machine-learning approach. We select RF as our baseline model for several reasons:

- 1) RF can be used for both classification and regression, as needed by this study.
- 2) RF can handle datasets with missing values (Tang and Ishwaran 2017). In our case, we need to consider that Sentinel-1 data are not available for every flood event, and the flood extent-related predictors are often missing in small events. Specifically, we adapt the surrogate split algorithm (Breiman et al. 1984) to handle missing values.
- 3) RF can capture high-order nonlinear and complex relationships (Belgiu and Drăgu 2016), as can occur between flood severity and damage (Wagenaar et al. 2020).
- 4) RF is not sensitive to the outlier, noisy samples, or correlated predictors (Louppe 2014). In this study, predictors in the same category, such as  $P_{\text{summax5}}$  and  $P_{\text{max}}$  or  $F_{\text{water}}$ ,  $F_{\text{wet}}$ , and  $F_{\text{imp}}$ , could be moderately to strongly correlated (refer to Table 1), with each predictor still providing useful information, as indicated by the variable importance test (see the feature importance assessment below for details).
- 5) RF's accuracy and efficiency is competitive with that of many other ML approaches (Gislason et al. 2006; Rodriguez-Galiano et al. 2012)—for example, it can handle thousands of predictors without the need for dimensionality reduction (Wei et al. 2019), which makes it suitable for fusing multisource big data, as in our study.

**Model training and validation.** We optimize the minimum sample size in a terminal node (minNSTO) and the number of predictors per split (NPN)—the two hyperparameters that control the randomness of the RF (Probst et al. 2019)—using a fivefold cross-validation (CV)

procedure (Santos et al. 2018). The number of trees for each classifier or regressor is fixed to 200 to trade off the convergence and computation efficiency (Probst and Boulesteix 2018). In the CV procedure, we use every fold as the validation fold to compute the loss in one iteration. The misclassification rate and mean squared error (MSE) are used as the loss function of the classifier and regressor. It is worth noting that, with the utilization of the hybrid structure and the setting of BTU, the  $K$  classifiers and  $5K$  regressors would be produced by a complete training procedure, where  $K$  is the number of BTUs. We choose the Bayesian optimization, a widely used search algorithm (Ghahramani 2015), to find the best hyperparameters. Since the optimal parameters of RF depend on the training dataset, and multiple optimal hypergeometric parameter sets might result in the same error level, we do not recommend keeping fixed values as the optimal hyperparameters for all BTUs.

We test the performance of iClaim through a fivefold ensemble leave-one-out (EnLOO) procedure. In EnLOO, every sample is left out of the training set once. The final test set is formed, therefore, by concatenating all left-out samples. To evaluate the model capability in different application scenarios, we further conduct EnLOO by randomly partitioning samples using different principles, namely, leaving samples out (LSO), leaving grids out (LGO), leaving counties out (LCO), and leaving events out (LEO). Meanwhile, we estimate the feature importance using a permutation procedure (Strobl et al. 2007), which computes the increase of the prediction error by training the model with permuting the target predictor variable. It should be noted that, since we train a different claim regression model for each damage level, the importance of a predictor might be various among levels.

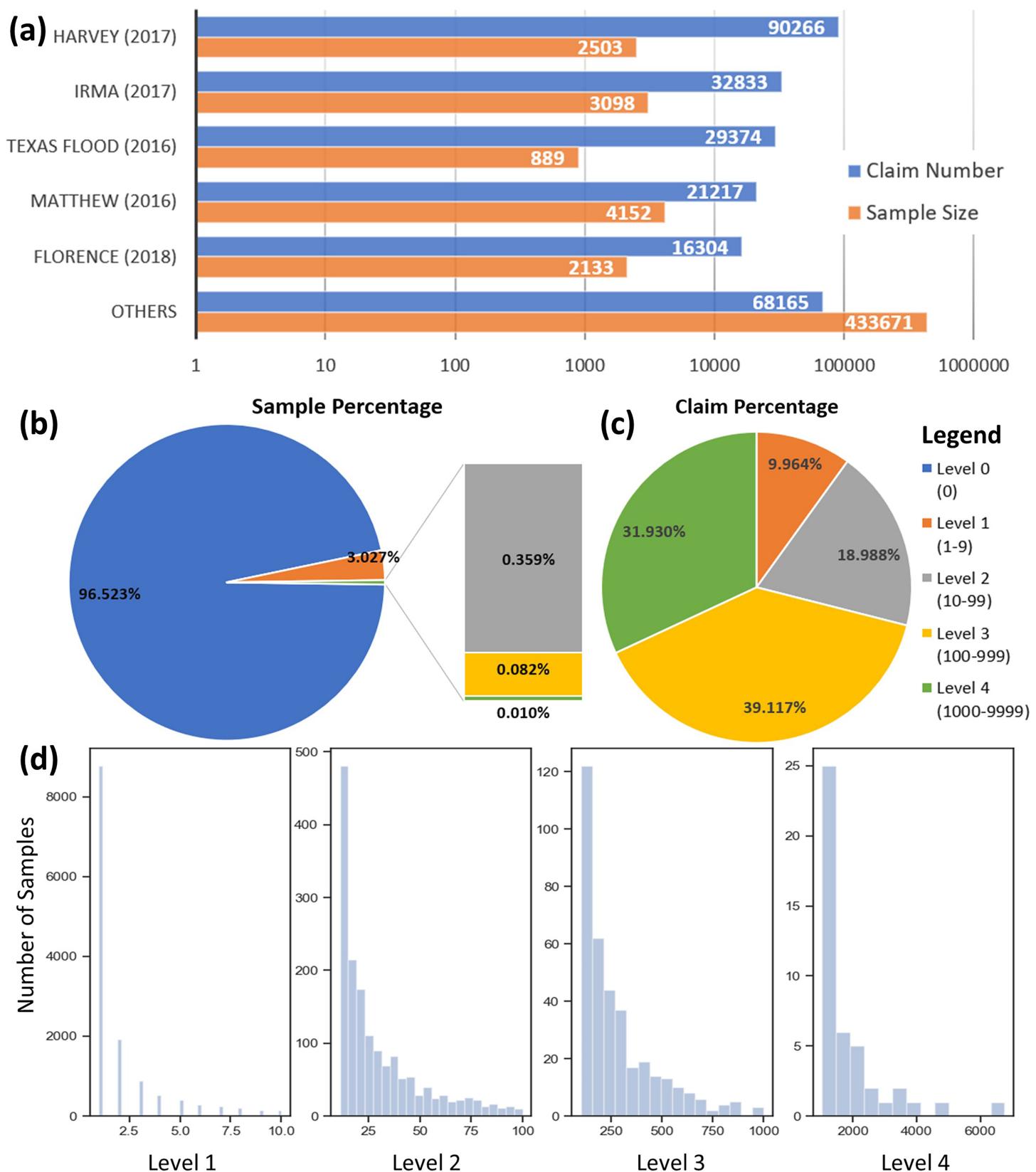
To assess the accuracy of the model, we select the Accuracy (Acc) and Cohen's kappa ( $K$  score) as the error metrics for classification and the coefficient of determination ( $R^2$ ), root-mean-square error (RMSE), and percentage of bias (Pbias) as the performance metrics for regression.

## Results and discussion

To demonstrate the effectiveness of iClaim, including the hybrid framework and the subsampling method, we compare the iClaim validation result with that derived from the simple RF.

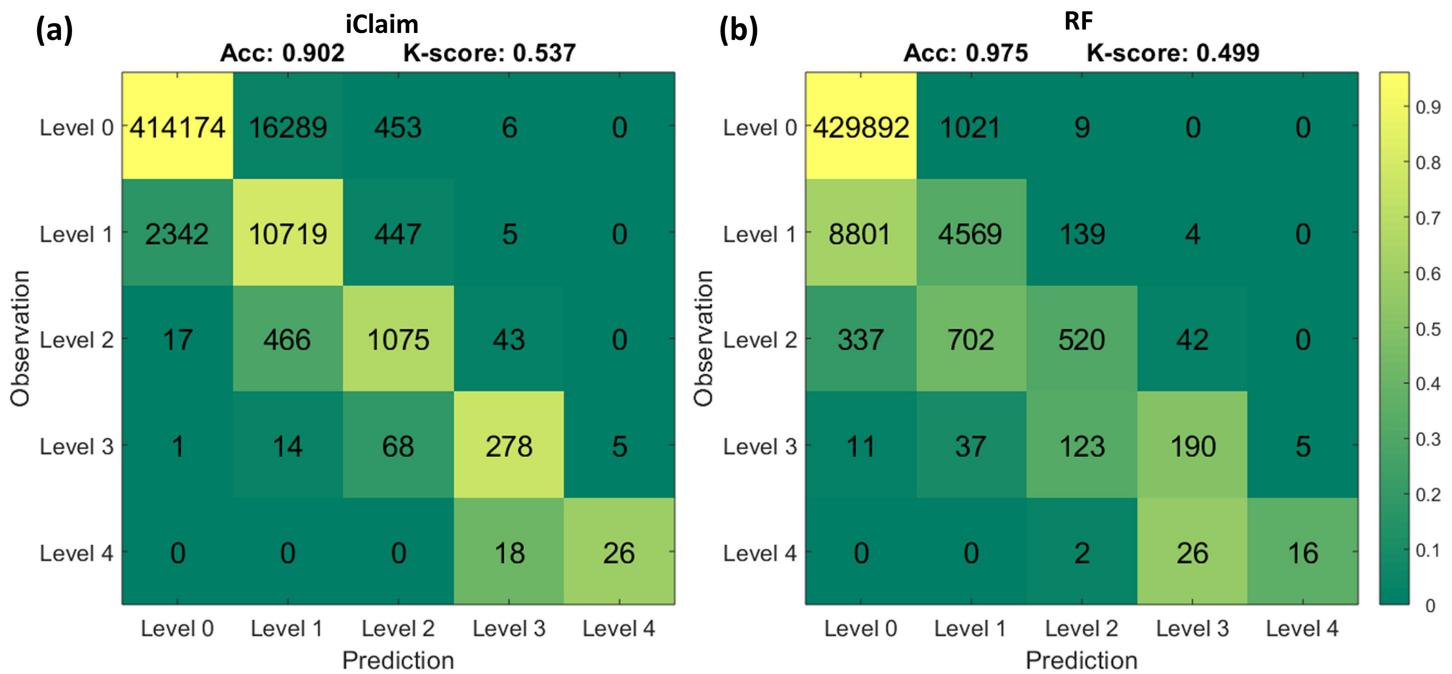
**Distribution of claims across events.** The claims data show skewed distribution, which would bias the model training because simply predicting zero-claims could minimize the loss function. The flood locator detected 24,674 potential flood events across CONUS from January 2016 to August 2019, with over 2 million samples. Due to computation limitations, we applied 10,000 km<sup>2</sup> as the PFZ threshold and selected 589 flood events to generate 446,446 grid/event samples, along with 258,159 claims (covering nearly 90% of the total claims during the same period), as the input dataset. The distribution of the dataset is shown in Fig. 3. Nearly three-quarters of the flood claims were contributed by the top five events, which take up only 3% of the total samples (Fig. 3a). Postevent survey reports from FEMA (FEMA 2018a,b) conclude that Hurricane Harvey (2017) and Hurricane Irma (2017) led to the filing of 91,000 and 33,111 flood insurance claims, respectively, a difference of only about 1% from our estimation (90,266 and 32,833; see Fig. 3a), derived using the PFZs. The grid/event samples were highly unbalanced across all damage levels, with over 93% having no claims reported and less than 1.5% (i.e., levels 3–4) contributing over 70% of the claims (Figs. 3b,c). Moreover, the sample size decreases exponentially with the claim count (Fig. 3d), which also makes the distribution of samples within each damage level uneven.

**Prediction of damage level.** Both iClaim's classification model and the simple RF achieve seemingly high accuracy on the damage level prediction, as indicated by their  $K$ -score values. But the simple RF only performs well in predicting level 0 (Fig. 4b), whereas the



**Fig. 3.** Unbalanced distribution of grid/event sample and claim: (a) the number of claim and sample among flood events; (b) the sample and (c) claim percentage for different damage levels; (d) distribution of sample number within each level.

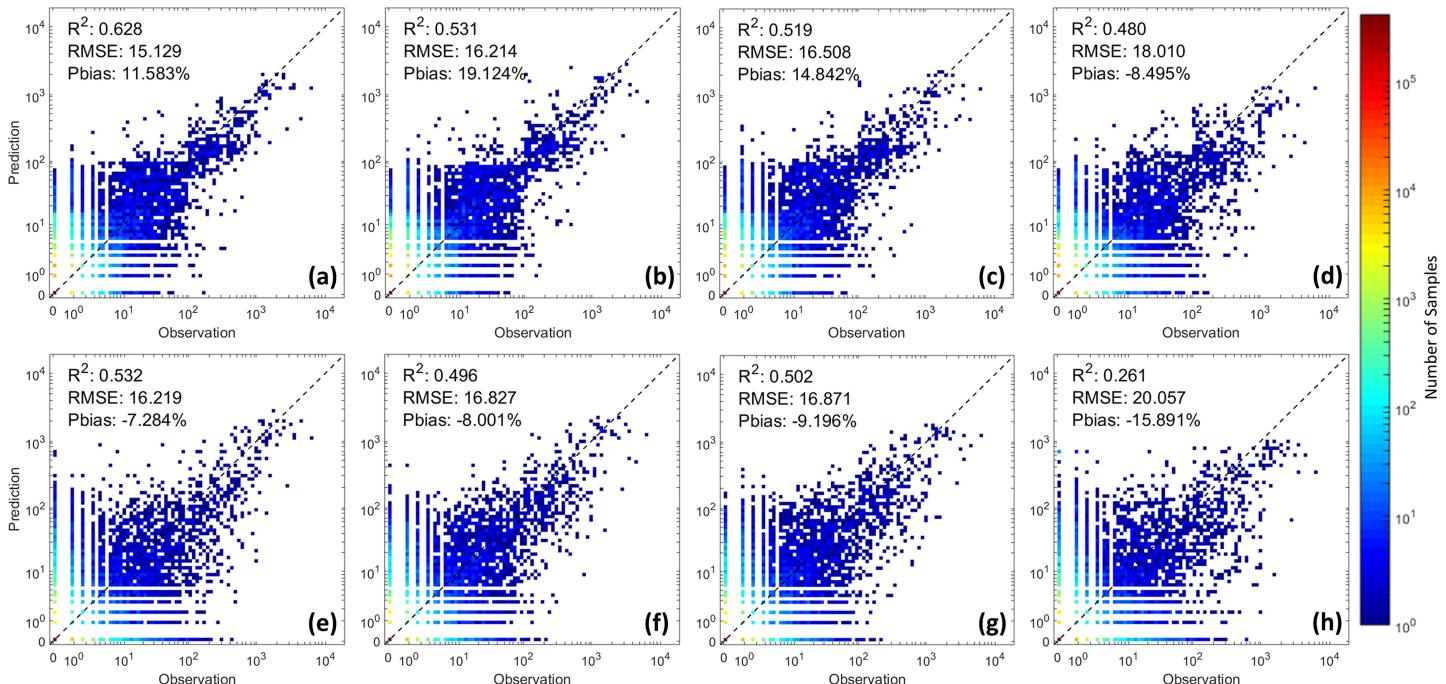
prediction by iClaim yields high accuracy across all damage levels (Fig. 4a). Note that stakeholders and policymakers have more urgent needs of the capability in predicting the damage of major flood events whose damage level is usually high (i.e., levels 3–4). We also observed that, in iClaim, most of the confusion occur between adjacent levels,



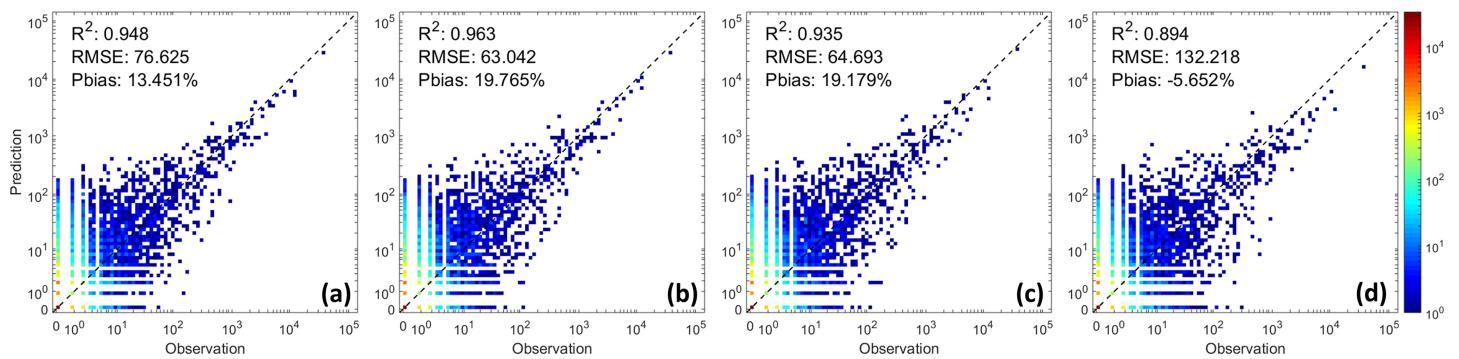
**Fig. 4.** Confusion matrix and performance indicators of damage level prediction from (a) the iClaim's classification result, which leverages the balance subsampling strategies, and (b) the simple RF. The color of each cell represents the ratio of correct predict number vs total observation for that level.

which allows the CLS in the regression step to reduce significantly the error caused by misclassification.

**Prediction of claim count.** Figure 5 shows the density scatterplots at the grid by event level using the four EnLOO scenarios, based on both the iClaim model and simple RF. For iClaim, the  $R^2$  of all validation scenarios was close to or above 0.5. The percentage of bias (i.e., Pbias) was within  $\pm 20\%$ , and the RMSE of most samples were below 20 houses, indicating an



**Fig. 5.** Density scatterplots of validation results from LSO, LGO, LCO, and LEO scenarios for the (a)–(d) iClaim and (e)–(h) simple RF at the grid/event level (sample size: 446,446) over CONUS. Black dashes indicate the 1:1 line.



**Fig. 6.** (a)–(d) Density scatterplots of validation results from LSO, LGO, LCO, and LEO scenarios for the iClaim at the county/event level (sample size: 47,235) over CONUS. Black dashes indicate the 1:1 line.

acceptable accuracy over the CONUS area. As shown in Figs. 5f–h, the simple RF significantly underperforms iClaim. In the LEO scenario, its performance is unacceptable.

Among all the EnLOO scenarios, the LSO (Figs. 5a,e) gains the highest accuracy for both iClaim and RF, with  $R^2$  being 0.628 and 0.532, respectively. In this scenario, the EnLOO take advantage of both spatial and intra-event correlations hidden in the data. The performance of both models decrease slightly in the LGO and LCO scenarios (Figs. 5b,c,f,g), indicating the claim prediction did not have strong spatial dependence (Ploton et al. 2020). As demonstrated by the significantly reduced  $R^2$  and increased RMSE in the LEO scenario (Figs. 5d,h), the model utilizes intra-event correlation easily. The reduced performance in the LEO scenario could also be attributed to the claim distribution, which contains only a handful of major flood events (see Fig. 5a). Leaving those high-claim events out means testing the model performance for unobserved extremes. Even in this most challenging scenario, iClaim prediction still shows an acceptable agreement with observations. In contrast, the simple RF model severely underestimates the high-claim samples and overestimates the low-claim samples. The primary reason may be the integration of the balanced subsampling strategies (i.e., CLS and BTL) in iClaim, which gives a less biased estimation, even with some extreme samples permuted.

In practice, the prediction result at the county by event level is more important to stakeholders and policymakers. We obtain the county-level predictions and observations by aggregating from the grid level. The overall agreement between model prediction and observation is expected to be significantly higher at the county level than at the grid level ( $R^2$  are increase from 0.5–0.6 to 0.9), indicating that much of the random error at the grid level is neutralized in the aggregation process (Fig. 6). As in the grid-level analysis, the LEO scenario yields the lowest performance (Fig. 6d), with  $R^2$  and RMSE at 0.894 and 132.218 houses, respectively, which was acceptable, proving that iClaim could characterize unobserved flood events.

To demonstrate further the predictability of unobserved major flood events, we extract four hurricanes—Harvey (2017), Irma (2017), Matthew (2016), and Florence (2018)—as well as a fluvial flood event—the 2019 Midwestern Great Flood—from the test sets (see Fig. 7). All prediction results in these five events are consistent with the observations, resulting in high  $R^2$  values. The hurricane examples suffer from similar amounts of underestimation, especially in counties with large claim numbers, exhibiting a prediction pattern common to any data-driven method when excluding the extreme samples from the training set. In the case of 2019 Midwestern Great Flood, the underestimation is concentrated in the upper Mississippi River region (Fig. 7e), which could be attributed to the lack of sufficient Sentinel-1 coverage over these areas during the event [refer to Fig. 6a from Yang et al. (2021)].

To obtain a better understanding of the performance of iClaim across different events, we compute the  $R^2$  at both the grid and county levels for each flood event from the LEO

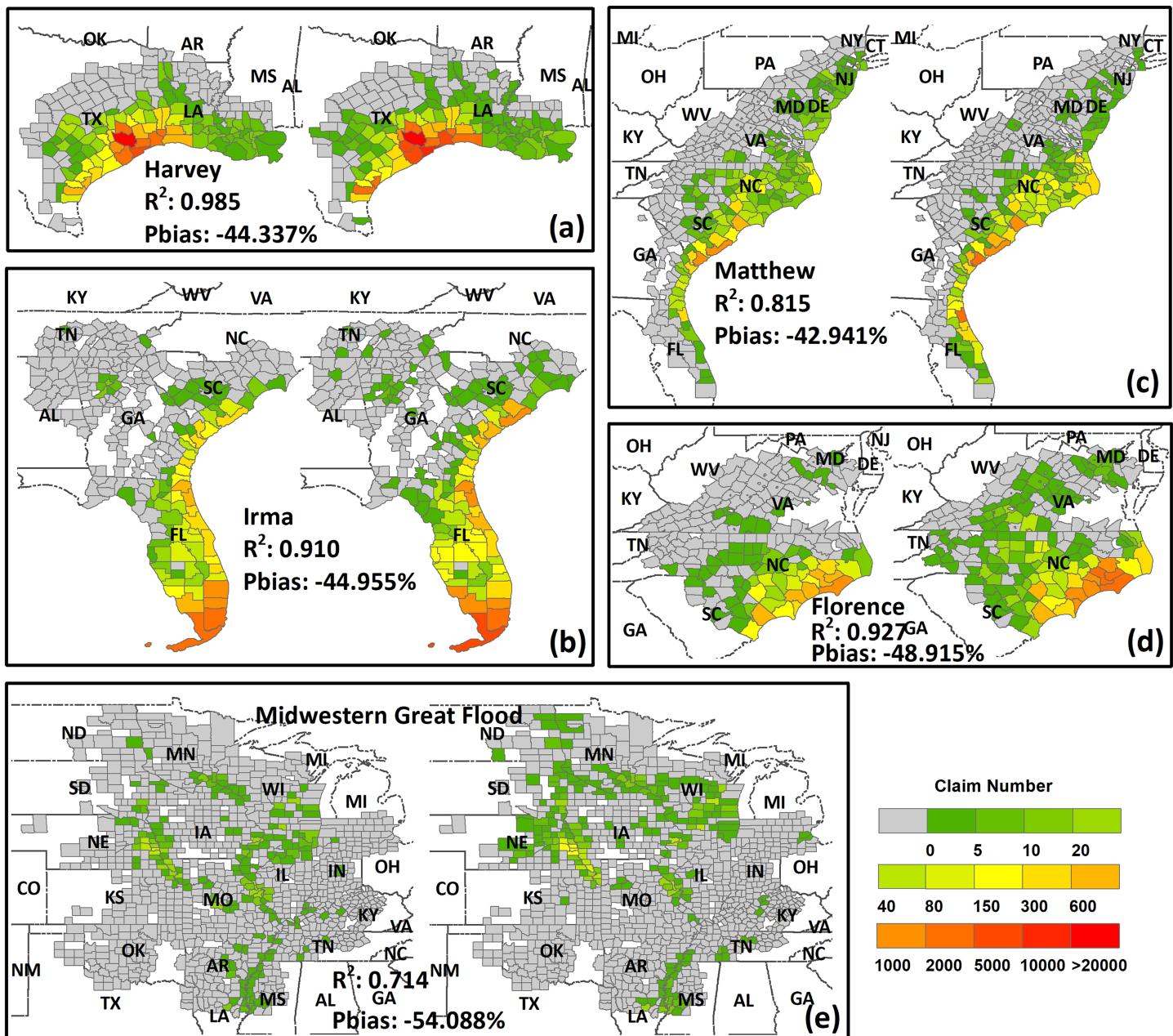
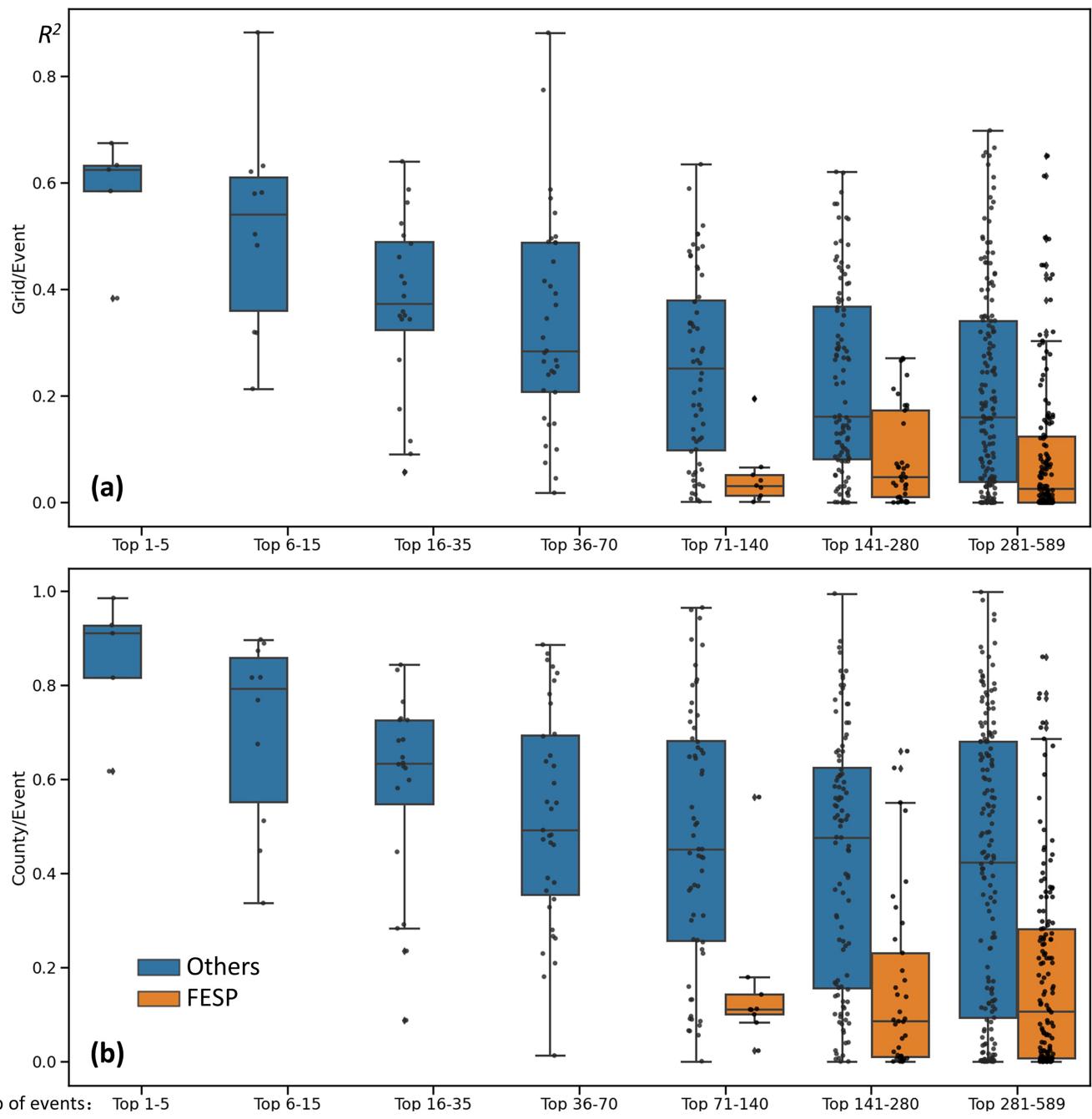


Fig. 7. Predicted (left map in each panel) vs observed (right map in each panel) claims at the county level for five selective events.

scenario, as shown in Fig. 8. The result indicates that the accuracy increases with the total claim number for each event, and better performance could be expected at the county level than the grid level. Lower performance was also reported for the events with fewer claims (e.g., rank 71–589 events). Among those events, 190 were Fluvial Events without Sufficient Predictors (FESP represents the events induced by fluvial floods while have no Sentinel-1 SAR images coverage). Most of these are found to have  $R^2$  less than 0.1 and 0.2 at the grid and county levels, respectively. This revealed that iClaim suffers from the lack of sufficient predictors in the low-damage fluvial events. Fig. ES3 shows the prediction result in the event accumulative level from the LEO scenario. The  $R^2$  of 0.952 reveals that the overall high consistency with the observed claim number. Similar to the grid/event and county/event levels, however, the accuracy drops significantly when the total claims per event were fewer than 300.

To verify iClaim's predictability for unseen locations, we compute the spatial distribution of  $R^2$  at the county level using results of the LCO scenario, as depicted by Fig. 9. Overall, the



**Fig. 8.** Boxplot of prediction performance ( $R^2$ ) among different event groups at the (a) grid and (b) county levels, derived from LEO scenario. Numbers in blue indicate the average claims of each grid/event or county/event sample.

mean and the median of the local  $R^2$  are 0.614 and 0.803, respectively. In particular, the local  $R^2$  tend to be higher for areas vulnerable to flooding, where the average of county-total claims per event was greater than five (Figs. 9a,b).

**Feature importance assessment.** Since the LEO is the benchmark validation scenario for iClaim, we choose it to compute the relative importance of features, as shown in Fig. 10. For clarity, we normalize the relative importance, so the sum of all features is 100%. All finally selected features of iClaim are, overall, important. At the low damage levels (0 and 1),  $P_{\text{summax5}}$ ,  $E_{\text{mean}}$ ,  $P_{\text{max}}$ , and NP are the most important features (Fig. 10a). At the high levels (2–4), NP,  $P_{\text{summax5}}$ ,  $CWL_{\text{max}}$ , and  $E_{\text{mean}}$  are the most important. NP is important for all levels, which confirms its role as the modulator. Topographical predictors, such as mean slope

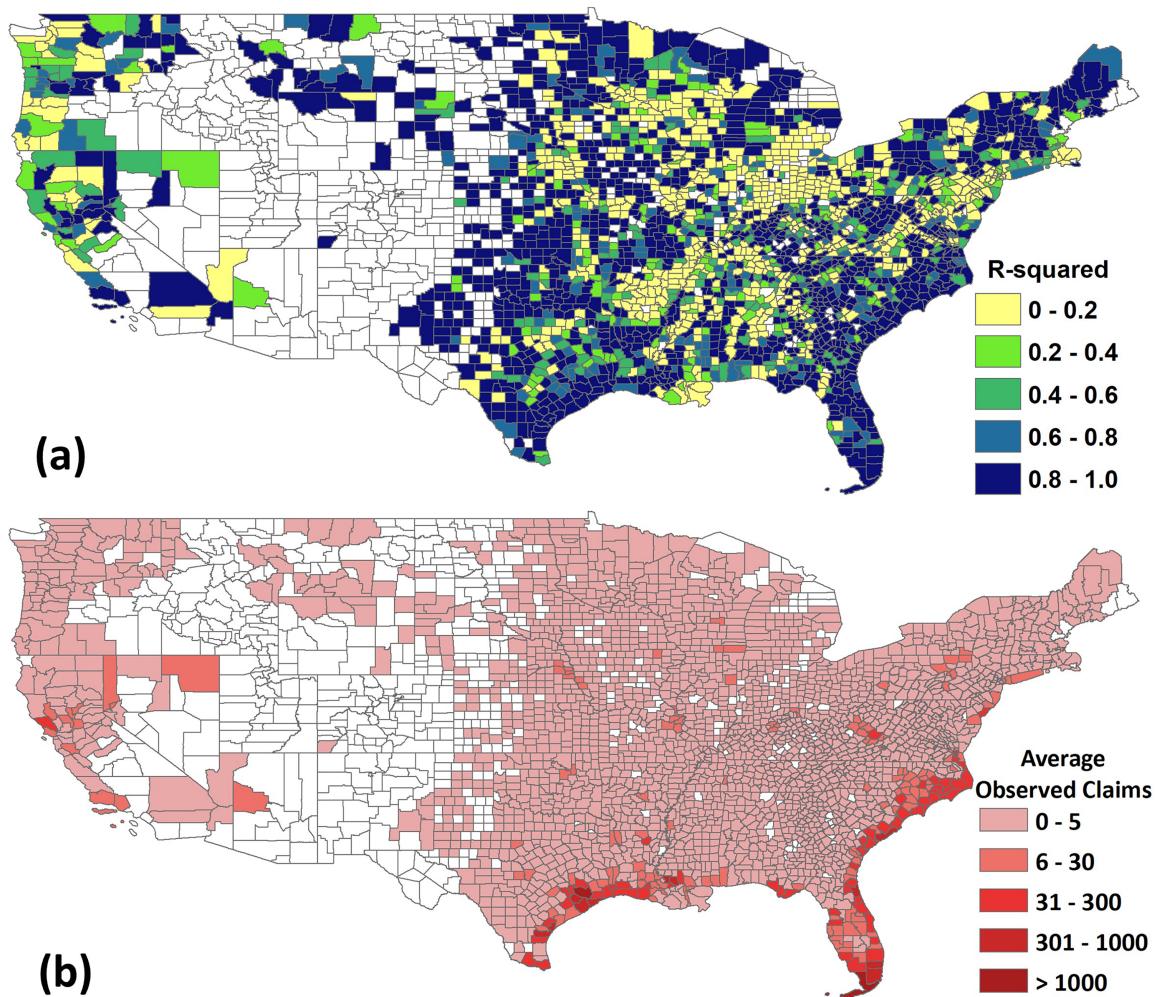


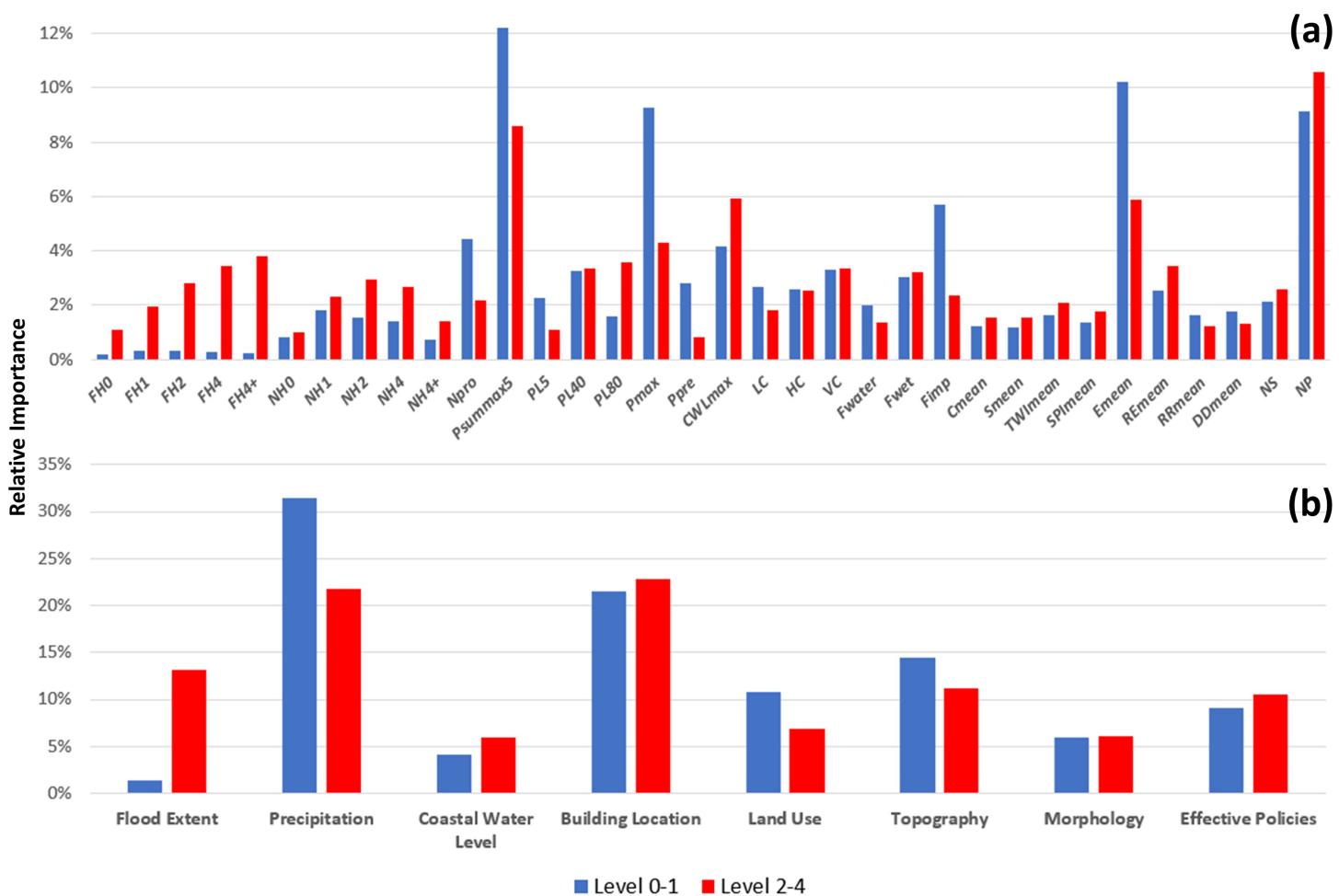
Fig. 9. Spatial distribution of (a) local  $R^2$  between claims prediction and measurement and (b) average observed claims per event from 2016 to 2019 over CONUS at the county level. Results are from the LCO scenario.

and elevation of the grids, are also found to be important components for claim prediction, which is consistent with the conclusion in Woznicki et al. (2019) and Alipour et al. (2020).

The contributions of some features are contrasting for low and high damage levels (Fig. 10b). The satellite-based flood inundation extent, for instance, is significantly more important at the high levels than the low. This result is intuitive, because a short-lasting flood event is significantly less likely to be captured by the Sentinel-1 than a long-lasting event. In fact, Sentinel-1-based flood extents are available for over 50% of the samples at levels 2–4 and less than 18% for those at levels 0–1. Similarly, coastal water level (i.e.,  $CWL_{max}$ ) is more important at the high damage levels, since samples with large claim number are likely to be induced by the coastal flood events (see Fig. 2a). In contrast, the contribution of impervious area fraction decreases significantly from low to high damage levels, indicating impervious areas may only influence the damage from small flood events. The result also reveals that some of the features could be removed to cope with possible data scarcity conditions without reducing much of the model performance. Specifically,  $TWI_{mean}$ ,  $SPI_{mean}$ ,  $RR_{mean}$ , and  $DD_{mean}$ , which contribute relatively little to the model at all damage levels, could be considered for exclusion from training.

## Summary

In this study, we develop a hybrid ML model, iClaim, to predict flood insurance claims caused by every event over CONUS. With the data imbalance issue in model construction now overcome,



**Fig. 10.** Average relative importance of (a) feature variables and (b) feature categories for different damage level ranges. The detailed description of features and the associated categories could be found in Table 1.

iClaim becomes the first “blind-prediction” model—that is, it does not rely on any postdisaster surveys. Therefore, it can aid stakeholders, policymakers, and first responders in postdisaster assessment and alleviation, emergency response, predisaster forecasting, and climate risk mitigation. The iClaim model can provide predictions at multiple spatiotemporal scales, including per grid per event, per county per event, per grid, per county, and per event, depending on the application scenario and accuracy requirement. The capability of blind prediction at fine spatiotemporal scale, which is per grid per event, had not been reported by previous studies.

Technically, iClaim works for both single-source flood events (e.g., the 2019 Midwestern Great Flood) and compound ones (e.g., hurricane-induced events) to which heavy precipitation and storm surge contribute simultaneously. By constructing a two-step working flow—damage level classification followed by claim number estimation—with a specially designed subsampling strategy, iClaim can overcome the overfitting and underfitting issues that occur when dealing with flood claim samples that show a strongly skewed distribution in both space and time. By alchemizing information from multisource satellite products, river and coastal water level monitoring systems, and flood insurance inventories, the model demonstrates acceptable predictability at the grid/event, county/event, and event accumulative levels, respectively, also including the predictability for unseen locations and events. Its integration of a low-cost flood locator makes iClaim the only blind-prediction model that eliminates the dependence on post-surveyed flood location information. It utilizes, for the first time, the NRT remotely sensed flood extent as predictors in a novel way, comparing the elevation distribution of inundated pixels with that of properties.

The following lessons are learned from the feature importance assessment:

- 1) As the primary flood driver, precipitation shows notable importance for all levels of flood events, which is consistent with Wang et al. (2015).
- 2) It is important to increase flood insurance penetration (Horn and Webel 2019) in vulnerable areas because the contribution of effective numbers of policies is stable at all damage levels.
- 3) The inclusion of satellite-retrieved flood extent information can significantly increase the predictive skill because the contribution of the flood extent predictors in reducing the prediction error of large flood events is significant. It is also suggested that we need to increase the revisiting frequency of SAR-based flood observation.

**Limitations and future developments.** iClaim is limited in five ways:

- 1) It cannot predict the claim status for each individual house, as it is limited by the truncated location information of the publicly accessible NFIP records.
- 2) We only tested it using four years of data, as we are limited by the property locations unknown to us from earlier times and by the seldom available flood extent products based on SAR (e.g., Sentinel-1) from early years (e.g., before 2016).
- 3) We have not tested its capability for predicting claimed payout because of the lack of accessible house value data and the involvement of non-physical factors in the approval of payout—for example, some properties without insurance might still acquire compensation through the Individual and Households Program (IHP; FEMA 2021).
- 4) Restricted by the revisiting frequency of Sentinel-1, iClaim is more likely to miss sufficient predictors for flash fluvial floods, leading to low prediction performance for these events.
- 5) We have not tested the model in regions other than CONUS, primarily because of the lack of comprehensive data on flood property damage. Effective flood predictors, such as the NRT flood extent information (Yang et al. 2021) and monitored flood stages, may also be less available in other parts of the world than in CONUS.

In the future, we will revisit these limitations by reaching out to authorities and vendors to access additional datasets under research-oriented and confidential agreements. These will include FEMA for the original NFIP data records, Zillow for the housing records, and ICEYE for the subdaily SAR data.

In addition, we will extend the model to include longer historical records to enable analysis of the impact of climate variability on flood vulnerability. We are also working on linking flood vulnerability to socioeconomic and demographic indicators, such as racial composition, education level, house income, population, asset value, precautions for and experience of floods, and flood mitigation structures, as suggested by other studies (Darabi et al. 2019; Duha Metin et al. 2018; Paprotny et al. 2018). Moreover, we will add the snow water equivalent (SWE) data from SNODAS to characterize snowmelt-triggered flooding, which contributes to spring floods in northern CONUS and arid mountainous areas (Shen et al. 2017b; Shen and Anagnostou 2017). Finally, we will attempt to utilize inundation map products from MODIS (Hawker et al. 2019), Landsat (Jones 2019), *Suomi-NPP* (Li et al. 2018), and passive microwave sensors (Du et al. 2018), which might be less detailed or effective than the SAR-based flood retrievals but are available going back multiple decades and/or have higher revisiting frequency.

**Acknowledgments.** The authors of this publication had research facility support from the Eversource Energy Center at the University of Connecticut. Additionally, this research was partially funded by the National Natural Science Foundation of China (NSFC), Regional Science Program, 51969004.

## References

Alipour, A., A. Ahmadalipour, P. Abbaszadeh, and H. Moradkhani, 2020: Leveraging machine learning for predicting flash flood damage in the Southeast US. *Environ. Res. Lett.*, **15**, 024011, <https://doi.org/10.1088/1748-9326/ab6edd>.

Anantrasirichai, N., J. Biggs, F. Albino, and D. Bull, 2019: A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets. *Remote Sens. Environ.*, **230**, 111179, <https://doi.org/10.1016/j.rse.2019.04.032>.

Barredo, J. I., D. Saurí, and M. C. Llasat, 2012: Assessing trends in insured losses from floods in Spain 1971–2008. *Nat. Hazards Earth Syst. Sci.*, **12**, 1723–1729, <https://doi.org/10.5194/nhess-12-1723-2012>.

Belgiu, M., and L. Drăgu, 2016: Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.*, **114**, 24–31, <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.

Breiman, L., 2001: Random forest. *Mach. Learn.*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.

—, J. H. Friedman, R. A. Olshen, and C. J. Stone, 1984: *Classification and Regression Trees*. Wadsworth and Brooks, 358 pp.

Chini, M., R. Pelich, L. Pulvirenti, N. Pierdicca, R. Hostache, and P. Matgen, 2019: Sentinel-1 InSAR coherence to detect floodwater in urban areas: Houston and Hurricane Harvey as a test case. *Remote Sens.*, **11**, 107, <https://doi.org/10.3390/rs11020107>.

Cook, A., and V. Merwade, 2009: Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping. *J. Hydrol.*, **377**, 131–142, <https://doi.org/10.1016/j.jhydrol.2009.08.015>.

Darabi, H., B. Choubin, O. Rahmati, A. Torabi Haghghi, B. Pradhan, and B. Kløve, 2019: Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. *J. Hydrol.*, **569**, 142–154, <https://doi.org/10.1016/j.jhydrol.2018.12.002>.

Domeneghetti, A., A. Castellarin, and A. Brath, 2012: Assessing rating-curve uncertainty and its effects on hydraulic model calibration. *Hydrol. Earth Syst. Sci.*, **16**, 1191–1202, <https://doi.org/10.5194/hess-16-1191-2012>.

Du, J., J. S. Kimball, J. Galantowicz, S. B. Kim, S. K. Chan, R. Reichle, L. A. Jones, and J. D. Watts, 2018: Assessing global surface water inundation dynamics using combined satellite information from SMAP, AMSR2 and Landsat. *Remote Sens. Environ.*, **213**, 1–17, <https://doi.org/10.1016/j.rse.2018.04.054>.

Duha Metin, A., N. Viet Dung, K. Schröter, B. Guse, H. Apel, H. Kreibich, S. Vorogushyn, and B. Merz, 2018: How do changes along the risk chain affect flood risk? *Nat. Hazards Earth Syst. Sci.*, **18**, 3089–3108, <https://doi.org/10.5194/nhess-18-3089-2018>.

FEMA, 2018a: Halfway through hurricane season critical need for flood insurance remains. FEMA, 3 pp., [www.fema.gov/press-release/20210318/halfway-through-hurricane-season-critical-need-flood-insurance-remains](http://www.fema.gov/press-release/20210318/halfway-through-hurricane-season-critical-need-flood-insurance-remains).

—, 2018b: Progress. Partnerships. Preparedness: Six months after Hurricane Harvey. FEMA, 3 pp., [https://www.fema.gov/press-release/20210318/progress-partnerships-preparedness-six-months-after-hurricane-harvey](http://www.fema.gov/press-release/20210318/progress-partnerships-preparedness-six-months-after-hurricane-harvey).

—, 2020: The National Flood Insurance Program (NFIP). Accessed 31 August 2020, [www.fema.gov/national-flood-insurance-program](http://www.fema.gov/national-flood-insurance-program).

—, 2021: Individuals and Households Program. Accessed 20 February 2021, [www.fema.gov/assistance/individual/program](http://www.fema.gov/assistance/individual/program).

Garrote, J., F. M. Alvarenga, and A. Díez-Herrero, 2016: Quantification of flash flood economic risk using ultra-detailed stage–damage functions and 2-D hydraulic models. *J. Hydrol.*, **541**, 611–625, <https://doi.org/10.1016/j.jhydrol.2016.02.006>.

Gerl, T., H. Kreibich, G. Franco, D. Marechal, and K. Schröter, 2016: A review of flood loss models as basis for harmonization and benchmarking. *PLOS ONE*, **11**, e0159791, <https://doi.org/10.1371/journal.pone.0159791>.

GFDRR, 2018: Machine learning for disaster risk management. GFDRR, 49 pp., [www.gfdrr.org/en/publication/machine-learning-disaster-risk-management](http://www.gfdrr.org/en/publication/machine-learning-disaster-risk-management).

Ghahramani, Z., 2015: Probabilistic machine learning and artificial intelligence. *Nature*, **521**, 452–459, <https://doi.org/10.1038/nature14541>.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson, 2006: Random forests for land cover classification. *Pattern Recognit. Lett.*, **27**, 294–300, <https://doi.org/10.1016/j.patrec.2005.08.011>.

Gradeci, K., N. Labonne, E. Sivertsen, and B. Time, 2019: The use of insurance data in the analysis of surface water flood events – A systematic review. *J. Hydrol.*, **568**, 194–206, <https://doi.org/10.1016/j.jhydrol.2018.10.060>.

Grimaldi, S., J. Xu, Y. Li, V. R. N. Pauwels, and J. P. Walker, 2020: Flood mapping under vegetation using single SAR acquisitions. *Remote Sens. Environ.*, **237**, 111582, <https://doi.org/10.1016/j.rse.2019.111582>.

Hardesty, S., X. Shen, E. Nikolopoulos, and E. Anagnostou, 2018: A numerical framework for evaluating flood inundation hazard under different dam operation scenarios—A case study in Naugatuck River. *Water*, **10**, 1798, <https://doi.org/10.3390/w10121798>.

Hawker, L., and Coauthors, 2019: Comparing earth observation and inundation models to map flood hazards. *Environ. Res. Lett.*, **15**, 124032, <https://doi.org/10.1088/1748-9326/abc216>.

HEC, 2012: HEC-FIA flood impact analysis user's manual. U.S. Army Corps of Engineers, 354 pp.

Horn, D. P., and B. W. Webel, 2019: Introduction to the national flood insurance program (NFIP). *The National Flood Insurance Program Background, Reauthorization and Reform*, SNOVA, 53–98.

Huffman, G. J., E. F. Stocker, D. T. Bolvin, E. J. Nelkin, and J. Tan, 2019: GPM IMERG Late Precipitation L3 1 day 0.1 degree × 0.1 degree V06. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 31 August 2020, <https://doi.org/10.5067/GPM/IMERGDL/DAY/06>.

Jaafar, H. H., F. A. Ahmad, and N. El Beyrouthy, 2019: GCN250, new global gridded curve numbers for hydrologic modeling and design. *Sci. Data*, **6**, 145, <https://doi.org/10.1038/s41597-019-0155-x>.

Jones, J. W., 2019: Improved automated detection of subpixel-scale inundation-revised Dynamic Surface Water Extent (DSWE) partial surface water tests. *Remote Sens.*, **11**, 374, <https://doi.org/10.3390/rs11040374>.

Khanam, M., G. Sofia, M. Koukoula, R. Lazin, E. I. Nikolopoulos, X. Shen, and E. N. Anagnostou, 2021: Impact of compound flood event on coastal critical infrastructures considering current and future climate. *Nat. Hazards Earth Syst. Sci.*, **21**, 587–605, <https://doi.org/10.5194/nhess-21-587-2021>.

Klotzbach, P. J., S. G. Bowen, R. Pielke, and M. Bell, 2018: Continental U.S. hurricane landfall frequency and associated damage: Observations and future risks. *Bull. Amer. Meteor. Soc.*, **99**, 1359–1376, <https://doi.org/10.1175/BAMS-D-17-0184.1>.

Lehner, B., K. Verdin, and A. Jarvis, 2013: HydroSHEDS Technical Documentation Version 1.2. HydroSHEDS, 26 pp., [www.hydrosheds.org](http://www.hydrosheds.org).

Li, S., and Coauthors, 2018: Automatic near real-time flood detection using Suomi-NPP/VIIRS data. *Remote Sens. Environ.*, **204**, 672–689, <https://doi.org/10.1016/j.rse.2017.09.032>.

Li, Y., S. Martinis, and M. Wieland, 2019: Urban flood mapping with an active self-learning convolutional neural network based on TerraSAR-X intensity and interferometric coherence. *ISPRS J. Photogramm. Remote Sens.*, **152**, 178–191, <https://doi.org/10.1016/j.isprsjprs.2019.04.014>.

Liu, X. Y., J. Wu, and Z. H. Zhou, 2009: Exploratory undersampling for class-imbalance learning. *IEEE Trans. Syst. Man Cybernetics*, **39**, 539–550, <https://doi.org/10.1109/TSMCB.2008.2007853>.

Louppe, G., 2014: Understanding random forests: From theory to practice. arXiv, 2111 pp., <http://arxiv.org/abs/1407.7502>.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea, 2012: NHD Plus Version 2 : User Guide. U.S. Environmental Protection Agency, 172 pp., accessed 31 August 2020, [ftp://ftp.horizon-systems.com/NHDplus/NHDPlusV21/Documentation/NHDPlusV2\\_User\\_Guide.pdf](http://ftp.horizon-systems.com/NHDplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf).

Mendelsohn, R., K. Emanuel, S. Chonabayashi, and L. Bakkensen, 2012: The impact of climate change on global tropical cyclone damage. *Nat. Climate Change*, **2**, 205–209, <https://doi.org/10.1038/nclimate1357>.

Merz, B., H. Kreibich, R. Schwarze, and A. Thieken, 2010: Review article "assessment of economic flood damage." *Nat. Hazards Earth Syst. Sci.*, **10**, 1697–1724, <https://doi.org/10.5194/nhess-10-1697-2010>.

—, —, and U. Lall, 2013: Multi-variate flood damage assessment: A tree-based data-mining approach. *Nat. Hazards Earth Syst. Sci.*, **13**, 53–64, <https://doi.org/10.5194/nhess-13-53-2013>, 2013.

Microsoft, 2018: US Building Footprints. GitHub, accessed 31 August 2020, <https://github.com/microsoft/USBuildingFootprints>.

Moncoulon, D., and Coauthors, 2014: Analysis of the French insurance market exposure to floods: A stochastic model combining river overflow and surface runoff. *Nat. Hazards Earth Syst. Sci.*, **14**, 2469–2485, <https://doi.org/10.5194/nhess-14-2469-2014>.

Neal, J., T. Dunne, C. Sampson, A. Smith, and P. Bates, 2018: Optimisation of the two-dimensional hydraulic model LISFOOD-FP for CPU architecture. *Environ. Modell. Software*, **107**, 148–157, <https://doi.org/10.1016/j.envsoft.2018.05.011>.

Nobre, A. D., L. A. Cuartas, M. R. Momo, D. L. Severo, A. Pinheiro, and C. A. Nobre, 2016: HAND contour: A new proxy predictor of inundation extent. *Hydrol. Processes*, **30**, 320–333, <https://doi.org/10.1002/hyp.10581>.

Oubennaceur, K., K. Chokmani, M. Nastev, M. Tanguy, and S. Raymond, 2018: Uncertainty analysis of a two-dimensional hydraulic model. *Water*, **10**, 272, <https://doi.org/10.3390/w10030272>.

Paprotny, D., O. Morales-Nápoles, and S. N. Jonkman, 2018: HANZE: A pan-European database of exposure to natural hazards and damaging historical floods since 1870. *Earth Syst. Sci. Data*, **10**, 565–581, <https://doi.org/10.5194/essd-10-565-2018>.

Ploton, P., and Coauthors, 2020: Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.*, **11**, 4540, <https://doi.org/10.1038/s41467-020-18321-y>.

Probst, P., and A. L. Boulesteix, 2018: To tune or not to tune the number of trees in random forest. *J. Mach. Learn. Res.*, **18**, 6673–6690.

—, M. N. Wright, and A. L. Boulesteix, 2019: Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev.: Data Min. Knowl. Discovery*, **9**, e1301, <https://doi.org/10.1002/widm.1301>.

Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sánchez, 2012: An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.*, **67**, 93–104, <https://doi.org/10.1016/j.isprsjprs.2011.11.002>.

Santos, M. S., J. P. Soares, P. H. Abreu, H. Araujo, and J. Santos, 2018: Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches. *IEEE Comput. Intell. Mag.*, **13**, 59–76, <https://doi.org/10.1109/MCI.2018.2866730>.

Scawthorn, C., and Coauthors, 2006: HAZUS-MH flood loss estimation methodology. II. Damage and loss assessment. *Nat. Hazards Rev.*, **7**, 72–81, [https://doi.org/10.1061/\(ASCE\)1527-6988\(2006\)7:2\(72\)](https://doi.org/10.1061/(ASCE)1527-6988(2006)7:2(72)).

Serpico, S. B., S. Dellepiane, G. Boni, G. Moser, E. Angiati, and R. Rudari, 2012: Information extraction from remote sensing images for flood monitoring and damage evaluation. *Proc. IEEE*, **100**, 2946–2970, <https://doi.org/10.1109/JPROC.2012.2198030>.

Shen, X., and E. N. Anagnostou, 2017: A framework to improve hyper-resolution hydrological simulation in snow-affected regions. *J. Hydrol.*, **552**, 1–12, <https://doi.org/10.1016/j.jhydrol.2017.05.048>.

—, H. J. Vergara, E. I. Nikolopoulos, E. N. Anagnostou, Y. Hong, Z. Hao, K. Zhang, and K. Mao, 2016: GDBC: A tool for generating global-scale distributed basin morphometry. *Environ. Modell. Software*, **83**, 212–223, <https://doi.org/10.1016/j.envsoft.2016.05.012>.

—, Y. Hong, K. Zhang, and Z. Hao, 2017a: Refining a distributed linear reservoir routing method to improve performance of the CREST model. *J. Hydrol. Eng.*, **22**, 04016061, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001442](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001442).

—, Y. Mei, and E. N. Anagnostou, 2017b: A comprehensive database of flood events in the contiguous United States from 2002 to 2013. *Bull. Amer. Meteor. Soc.*, **98**, 1493–1502, <https://doi.org/10.1175/BAMS-D-16-0125.1>.

—, E. N. Anagnostou, G. H. Allen, G. R. Brakenridge, and A. J. Kettner, 2019a: Near-real-time non-obstructed flood inundation mapping using synthetic aperture radar. *Remote Sens. Environ.*, **221**, 302–315, <https://doi.org/10.1016/j.rse.2018.11.008>.

—, D. Wang, K. Mao, E. Anagnostou, and Y. Hong, 2019b: Inundation extent mapping by synthetic aperture radar: A review. *Remote Sens.*, **11**, 879, <https://doi.org/10.3390/rs11070879>.

Smith, A. B., 2020: U.S. billion-dollar weather and climate disasters, 1980–present (NCEI Accession 0209268). NOAA National Centers for Environmental Information, accessed 29 December 2021, <https://doi.org/10.25921/stkw-7w73>.

Sørensen, J., and S. Mobini, 2017: Pluvial, urban flood mechanisms and characteristics – Assessment based on insurance claims. *J. Hydrol.*, **555**, 51–67, <https://doi.org/10.1016/j.jhydrol.2017.09.039>.

Strobl, C., A. L. Boulesteix, A. Zeileis, and T. Hothorn, 2007: Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinf.*, **8**, 25, <https://doi.org/10.1186/1471-2105-8-25>.

Tang, F., and H. Ishwaran, 2017: Random forest missing data algorithms. *Stat. Anal. Data Min.*, **10**, 363–377, <https://doi.org/10.1002/sam.11348>.

Tang, X., H. Hong, Y. Shu, H. Tang, J. Li, and W. Liu, 2019: Urban waterlogging susceptibility assessment based on a PSO-SVM method using a novel repeatedly random sampling idea to select negative samples. *J. Hydrol.*, **576**, 583–595, <https://doi.org/10.1016/j.jhydrol.2019.06.058>.

Tate, E., C. Muñoz, and J. Suchan, 2015: Uncertainty and sensitivity analysis of the HAZUS-MH flood model. *Nat. Hazards Rev.*, **16**, 04014030, [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000167](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000167).

Tehrany, M. S., S. Jones, and F. Shabani, 2019: Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *Catena*, **175**, 174–192, <https://doi.org/10.1016/j.catena.2018.12.011>.

Thieken, A. H., A. Olschewski, H. Kreibich, S. Kobsch, and B. Merz, 2008: Development and evaluation of FLEMOPs - A new flood loss estimation model for the private sector. *WIT Trans. Ecol. Environ.*, **315**–324, <https://doi.org/10.2495/FRIAR080301>.

Wagenaar, D., S. Lüdtke, K. Schröter, L. M. Bouwer, and H. Kreibich, 2018: Regional and temporal transferability of multivariable flood damage models. *Water Resour. Res.*, **54**, 3688–3703, <https://doi.org/10.1029/2017WR022233>.

—, K. M. De Bruijn, L. M. Bouwer, and H. De Moel, 2016: Uncertainty in flood damage estimates and its potential effect on investment decisions. *Nat. Hazards Earth Syst. Sci.*, **16**, 1–14, <https://doi.org/10.5194/nhess-16-1-2016>.

—, and Coauthors, 2020: Invited perspectives: How machine learning will change flood risk and impact assessment. *Nat. Hazards Earth Syst. Sci.*, **20**, 1149–1161, <https://doi.org/10.5194/nhess-20-1149-2020>.

Wang, Z., C. Lai, X. Chen, B. Yang, S. Zhao, and X. Bai, 2015: Flood hazard risk assessment model based on random forest. *J. Hydrol.*, **527**, 1130–1141, <https://doi.org/10.1016/j.jhydrol.2015.06.008>.

Wei, J., W. Huang, Z. Li, W. Xue, Y. Peng, L. Sun, and M. Cribb, 2019: Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. *Remote Sens. Environ.*, **231**, 111221, <https://doi.org/10.1016/j.rse.2019.111221>.

Wing, O. E. J., P. D. Bates, A. M. Smith, C. C. Sampson, K. A. Johnson, J. Fargione, and P. Morefield, 2018: Estimates of present and future flood risk in the conterminous United States. *Environ. Res. Lett.*, **13**, 034023, <https://doi.org/10.1088/1748-9326/aaac65>.

—, N. Pinter, P. D. Bates, and C. Kousky, 2020: New insights into US flood vulnerability revealed from flood insurance big data. *Nat. Commun.*, **11**, 1444, <https://doi.org/10.1038/s41467-020-15264-2>.

Woznicki, S. A., J. Baynes, S. Panlasigui, M. Mehaffey, and A. Neale, 2019: Development of a spatially complete floodplain map of the conterminous United States using random forest. *Sci. Total Environ.*, **647**, 942–953, <https://doi.org/10.1016/j.scitotenv.2018.07.353>.

Yang, F., P. Watson, M. Koukoula, and E. N. Anagnostou, 2020: Enhancing weather-related power outage prediction by event severity classification. *IEEE Access*, **8**, 60029–60042, <https://doi.org/10.1109/ACCESS.2020.2983159>.

Yang, Q., X. Shen, E. N. Anagnostou, C. Mo, J. R. Eggleston, and A. J. Kettner, 2021: A high-resolution flood inundation archive (2016–the present) from sentinel-1 SAR imagery over CONUS. *Bull. Amer. Meteor. Soc.*, **102**, E1064–E1079, <https://doi.org/10.1175/BAMS-D-19-0319.1>.

Yildirim, E., and I. Demir, 2019: An integrated web framework for HAZUS-MH flood loss estimation analysis. *Nat. Hazards*, **99**, 275–286, <https://doi.org/10.1007/s11069-019-03738-6>.