A Hydro-EVT Approach to Flood Insurance Pricing
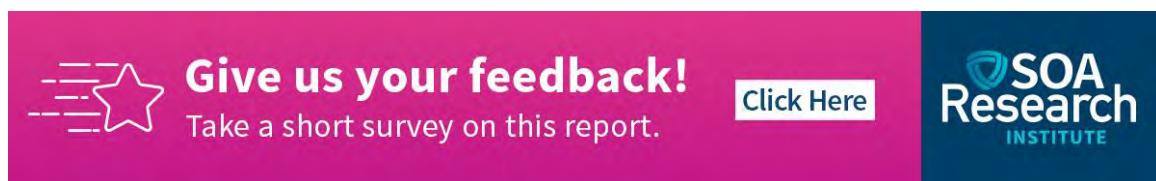
SEPTEMBER | 2022

A Hydro-EVT Approach to Flood Insurance Pricing

AUTHORS   Chaopeng Shen, PhD

Zhongyi Yuan, ASA, PhD

SPONSOR   Catastrophe and Climate Strategic Research Program Steering Committee

CONTENTS

A Hydro-EVT Approach to Flood Insurance Pricing

Executive Summary

Over the past decades, the world has witnessed more and more flood events that cause enormous amounts of economic losses. Such catastrophic floodings have been increasing in both frequency and intensity at alarming rates. It is now imperative to investigate the risk more closely and work towards better risk mitigation measures and risk sharing mechanisms. The goal of this research is to contribute to the effort by providing a modeling framework for assessing both the inundation risk of properties and their expected flood damages, and thereby laying a foundation for flood insurance pricing.

To this end, we propose to use a combination of extreme value statistics and hydrology models to assess properties' inundation probabilities and a mixed (generalized) beta model to model insurance claim losses. We use a regional analysis of flood risk to introduce a hierarchical model for assessing the inundation probability, consisting of an extreme value model that produces a heavy-tailed distribution for the annual peak streamflow at a nearby gage and a hydrologic model that produces inundation levels for various streamflow rates and across various locations. The hydrologic analysis produces a probabilistic classification model, which can be considered as a conditional relation between the inundation probability and geological and hydrological variables. Together with the streamflow model, this enables us to evaluate the inundation probability.

Furthermore, to investigate insurance losses to flood risk, we study over 50 years of claims data from the National Flood Insurance Program with 2.5 million records. We identify the property characteristics that predict flood damages for inundated properties and derive a mixed beta model that links the conditional property damage, conditional on the property's inundation, with relevant property characteristics.

Section 1: Introduction

1.1 FLOOD RISK AND FLOOD INSURANCE

The warming globe has undergone a series of changes in climate and environment that have rendered floods a growing risk worldwide. The climate and environmental changes have given rise to, among others, extreme precipitations, rising sea levels, and an intensifying global water cycle, each of which can contribute to extreme flooding (Brunner et al. (2021) and Masson-Delmotte et al. (2021)).

In recent years, devastating floods have been occurring with an increasing frequency and have taken heavy economic tolls. In fact, during the past decades, for every few years, there would be some major flood events that would cause catastrophic losses. According to Federal Emergency Management Agency (FEMA)'s National Flood Insurance Program (NFIP), insured losses to significant flood events were $23.6 billion in 2005, $11.0 billion in 2012, and $11.5 billion in 2017[1].

Moreover, as is shown in Figure 1 below, insured losses due to floods in the U.S have now become much more substantial and more volatile than they were just two decades ago. As the climate trends toward more frequent extreme hydrologic events for many parts of the U.S. (Stocker et al. (2013) and Hirsch and Archfield (2015)), floods will continue to be a major risk for property owners.

Figure 1
FLOOD INSURANCE PAYMENTS (IN BILLIONS) BY THE NFIP DURING THE YEARS 1978–2020. THE AMOUNTS ARE INFLATION ADJUSTED TO 2020 DOLLARS USING THE CPI DATA FROM THE U.S. BUREAU OF LABOR STATISTICS.
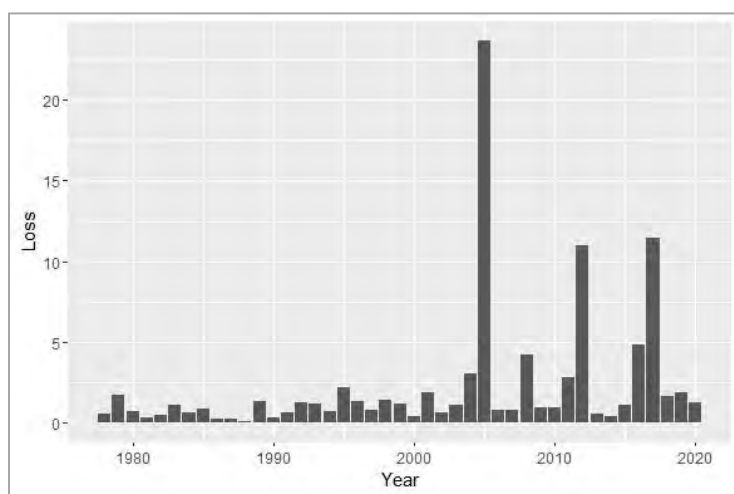
Data Source: Federal Emergency Management Agency (FEMA) and U.S. Bureau of Labor Statistics (BLS)

Due to the potentially high flood risk many property owners must face, it is extremely important that affordable flood insurance is available in the market. In an effort to improve flood insurance availability and affordability to the general public, FEMA created the NFIP in 1968, and for a long time, the NFIP has been the only marketplace for property owners to purchase flood insurance. It has so far sold millions of

---

[1] All numbers are in 2020 dollars, adjusted to inflation using CPI data from U.S. Bureau of Labor Statistics.

flood insurance policies to Americans and has contributed greatly to the growth of the flood insurance market.

On the other hand, due to the more frequent and more devastating floods over the last two decades, the NFIP has paid billions more in reimbursement than the premium collected and is currently deep in debt. Apparently, for many NFIP policies, the rates are not actuarially fair and are likely heavily subsidized. Realizing that the NFIP has failed to self-sustain as it hopes to, Congress has been planning to reform the program in recent years. Now the program relies on Congress's periodic authorization to remain functioning, and, at the time of writing, the authorization is set to expire on September 30, 2022.

This unclear future of the NFIP makes developing and establishing a sustainable private flood insurance market extremely important. The private market may not replace the NFIP anytime soon but will likely coexist with NFIP by providing extra coverage capacity and, for individual policies, different levels of coverage from the limited coverages by NFIP. The private flood insurance market is already growing, but, despite its importance and the various promotional measures from the government[2], it is still underdeveloped, insuring roughly 5%[3] of primary residential policies.

Obviously, the development of a private market hinges on insurers' ability to better understand and price the flood risk. In this research, we will take a closer look at flood risk modeling and flood insurance pricing.

1.2 STANDARD FLOOD INSURANCE POLICY

The flood insurance we shall study has coverages similar to those of a typical NFIP policy. In this section, we briefly describe the coverages of an NFIP Standard Flood Insurance Policy (SFIP).

First, under an SFIP, a flood must reach a certain threshold for it to be covered. Specifically, a covered flood is defined by the NFIP as a general and temporary condition of partial or complete inundation of two or more acres of normally dry land area or of two or more properties.

SFIP coverages, limitations, and exclusions are specified using three forms: the Dwelling Form, the General Property Form, and the Residential Condominium Building Association Policy (RCBAP) Form. While the General Property Form and RCBAP Form are used for covering residential buildings with five or more units, residential condominium association buildings, and non-residential buildings, the Dwelling Form is used for covering single-family homes, two-to-four families, residential condominium buildings, and residential townhouses/rowhouses. We shall focus on the coverages by the Dwelling Form, which is used by the majority of the NFIP policies.

The coverages under the Dwelling Form include the building property (Coverage A), the personal property (Coverage B), other coverages (Coverage C), and increased cost of compliance (ICC, Coverage D). Each coverage is subject to a max coverage limit, and so is the combined coverage. For example, for single family homes, the coverage limits are $250,000, $100,000, and $30,000 for building, content, and ICC, respectively. Claims on ICC are handled separately from those on the building and content but the combined payment for the three types of claims cannot exceed $250,000.

Despite being a mandatory part of most policies since 1997 and automatically included in new and renewed NFIP policies, the ICC coverage is used less than the other two. The vast majority of ICC claims

---

[2] The efforts include, among others, reforming the NFIP (e.g., by introducing Risk Rating 2.0 in October 2021) to deliver rates that are actuarially fair—private insurers are not willing to compete with a government program that offers subsidized rates—and publishing NFIP data to facilitate research on flood risk.
[3] See https://www.forbes.com/advisor/homeowners-insurance/flood-insurance/.

qualify because they satisfy the substantial damage criterion—although repetitive loss properties[4] are also eligible for ICC funds—meaning the flood damage is more than 50% of the property value. Overall, only a very small portion of claims satisfy one or both of the two criteria. Among the over 2 million single-family-home claim records in our dataset, only 1.8% of them include ICC claims.[5] For our claims analysis in Chapter 5, we shall study the claims on building coverage, which are by far the largest component.

Last, we point out that almost all NFIP policies have a one-year policy term, with some exceptions being Group Insurance Policies, which have a three-year term. Therefore, we shall also consider one-year policies and, for this reason, the inundation probability we calculate in Chapter 4 will be one-year inundation probability.

1.3 RESEARCH GOAL AND METHOD

In this study, we propose an integrated framework for assessing properties' flood risk and pricing their flood insurance coverage, using extreme value theory (EVT) and hydraulic modeling.

We first point out that existing research on flood risk usually employs distinct methodologies, such as extreme value statistics, machine learning, and large model ensembles, without investigating the possibility of combining them. Our study is the first to combine hydrologic engineering with extreme value statistics to create an ensemble model for flood risk modeling. A similar idea of integrating extreme value statistics with large model ensembles has been explored in the literature to investigate weather events; see Sippel et al. (2015). Using temperature data and rainfall data, Sippel et al. (2015) show that a combination of extreme value statistics and ensemble simulations can inform each other and improve overall model performance, especially in tail assessments. As the authors point out, the reasons for the improvement are that the ensemble simulations may help estimate the EVT parameters for samples of small size, and that the utilization of EVT allows one to significantly reduce the number of ensemble simulations needed, which could otherwise be prohibitively large for tail assessments. This is precisely our motivation of combining EVT with hydrological models for flood risk modeling.

In the meantime, we point out that although there is extensive research in flood risk modeling, research on how to translate the understanding of flood risk to flood insurance pricing has been rather limited in the academic literature. Some recent research in the insurance and actuarial research literature are Czajkowski et al. (2012), Kousky et al. (2017), Furman et al. (2019), and Boudreault et al. (2020). Our research is a contribution to this strand of literature.

In this research, we will treat flood frequency risk and property flood damage risk separately, using different models. It is natural to calculate the pure premium of flood insurance (with one year coverage) as the product of annual inundation probability and the conditional damage to the building[6]. The calculation of flood insurance premium will be straightforward once both of the two components are figured out.

To this end, we use the proposed ensemble model that combines extreme value statistics and hydrology methods for regional analysis of flooding probability. The extreme value model produces a distribution for the annual peak streamflow at a nearby gage, which captures the heavy tail of the streamflow. The hydrologic analysis, together with the probabilistic classification results in a conditional relation between

---

[4] Defined as properties that were damaged by floods two times in the past 10 years, and the cost of repairing the flood damage averages at least 25% percent of the property value at the time of each flood.

[5] See, e.g., FEMA (2017) and Kousky and Lingle (2017) for more details about ICC coverage.

[6] Although an NFIP policy coverages not only building damage, but also personal belongings, increased cost of compliance, etc., we focus on building damage only. The methodology should be easily extendable for other coverages.

the inundation probability and hydrologic variables, i.e., the streamflow and the property height above nearest drainage. Together with the streamflow model, this gives us the annual inundation probability.

Moreover, we will study over 50 years of claims data from the National Flood Insurance Program with 2.5 million records to investigate insurance losses to flood risk. This leads to a model for the insurance loss given that the property's inundation. Again, the two components above used together will give us a framework for flood risk pricing.

1.4 PREVIEW OF RESULTS

In this section, we preview pricing results that we will be able to obtain with our pricing models using two examples. Specifically, we consider the prices of one-year flood insurance coverage of two single-family homes located in Bellefonte, Pennsylvania.

As the first example, we consider a property located in Bellefonte, PA, at latitude 40.91 and longitude -77.78. Using the Z-estimate on Zillow[7], we estimate that its property value is $145,000. Since the property value is estimated to be below the NFIP single-family home coverage limit of $250,000, we assume that the flood insurance coverage limit is also below the limit and chosen as the estimated property value of $145,000. According to FEMA map[8] and Zillow, this property is located in a high-risk zone, has two floors and no basement, and was built 112 years ago. Its height above nearest drainage is 13.53 feet. The estimated annual inundation probability is 2.86%. Moreover, based on Model III in Section 5.3 and a median property value of $191,600 for this census tract, we estimate the building damage given inundation—assuming no community rating credits—is $38,556. Thus, the estimated annual premium for coverage of building damage is $1,102.

As the second example, we consider another property in Bellefonte, PA, at latitude 40.92 and longitude -77.78. We obtain the property characteristics in a similar way to the above. This property has an estimated value of $98,400—which also assume is the coverage limit—not located in a high-risk zone, has two floors and a finished basement, and was built 162 years ago. Its height above nearest drainage is 7.92 feet. The estimated annual inundation probability is 11.2%. Again, based on Model III in Section 5.3 and the median property value of $191,600 and assuming no community rating credits, we estimate the building damage given inundation is $21,488. Hence, the estimated annual premium for coverage of building damage is $2,404[9].

As a side note, the rates calculated here do not account for certain provisions set in statute, such as the cross-subsidies between policyholders,[10] how the subsidies will be phased out, and the rate increase limit.[11]

Below we shall elaborate on our models that we employ to produce these pricing results. The rest of the report consists of six sections. Sections 2, 3 and 4 focus on assessing the inundation probability, where

---

[7] See https://www.zillow.com/.

[8] See https://msc.fema.gov/portal/home.

[9] The rates seem quite high, but for comparison, some property owners in the same area pay even higher rates. As examples, according to FEMA policy data, one policy covering a single-family home in Bellefonte, PA, with building coverage limit $17,7600, effective date July 08, 2020, has an annual policy premium of $5,296 and policy cost of $6,549—The policy cost includes policy premium, reserve fund assessment, and federal policy fee, and Homeowner Flood Insurance Affordability Act (HFIAA) surcharge—and another policy, also covering a single-family home in Bellefonte, PA, with building coverage limit $20,2000, effective date June 22, 2020, has an annual policy premium of $6,760 and policy cost of $8,052.

[10] Subsidies are provided by NFIP to, e.g., pre-FIRM buildings, newly mapped buildings, and grandfathered buildings. These subsidies are set to phase out gradually under Risk Rating 2.0. See, e.g., Horn (2021) for related discussions.

[11] For example, under the current statute, the premium rate for individual primary residence policies cannot increase by more than 18% per year, and that for other properties such as non-primary residences and business properties cannot increase by more than 25% per year.

Section 2 introduces the EVT method for modeling peak streamflow, Section 3 describes the hydrological variables used for assessing conditional inundation probability, Section 4 derives both the unconditional and unconditional inundation probabilities. Section 5 discusses claims modeling, i.e., models for expected flood damage given that the property is inundated. Section 6 consists of acknowledgements. The Appendices contain a review of EVT, a description of machine learning methods for predicting streamflow for ungauged basins, and some additional discussions on our claims models.

Section 2: Extreme Value Modeling of Peak Streamflow

Our approach to assessing a property's inundation probability is built upon two components: a model of peak streamflow for the water source nearby, which we shall establish using extreme value theory (EVT), and regional analysis of the conditional relation, conditional on the peak flow, between the inundation probability and hydrologic variables. We now describe each of the two components and the underlying models, starting with extreme value modeling of peak streamflow.

2.1 EVT

EVT has been applied extensively to both actuarial science and hydrology. For example, in actuarial science, it has found applications in developing life tables, modeling large losses, extreme mortality events, natural catastrophes, extreme weather events, as well as their tail dependences; see, e.g., McNeil (1997), Watts et al. (2006), Zimbidis et al. (2007), Gbari et al. (2017), Tang and Yuan (2019), and Huang et al. (2020). Monographic treatments on extreme value statistics can be found in, e.g., Coles et al. (2001), Beirlant et al. (2004), and Embrechts et al. (2013). In hydrology, it has been used to model, e.g., extreme precipitation, river peak streamflow, inundation risk, and flood damage. See, e.g., Smith (1987), Wang (1991), Katz et al. (2002), El Adlouni et al. (2007), Williams et al. (2007), Villarini and Smith (2010), Agilan and Umamahesh (2015), Lombardo et al. (2019), Curceac et al. (2020), Lee et al. (2020), and Tabari (2021).

Extreme value analysis of flood risk is usually performed using the Annual Maximum Series (AMS) approach or the Partial Duration Series (PDS) approach, which correspond, respectively, to the Block Maximum (BM) method and Peaks over Threshold (POT) method in the statistics literature. The former has its root in the Fisher–Tippett–Gnedenko theorem and the latter in the Pickands-Balkema-de Haan theorem. We provide a brief review of EVT in Appendix A.

Our purpose is to utilize EVT described in Appendix A to model peak flow and then estimate inundation probabilities. Specifically, in our pricing model, we aim at a relation between properties' inundation probabilities and hydrologic variables. The relation will be a conditional relation derived from regional analysis and takes the peak flow as an input. In the end, this conditional relation coupled with EVT model for peak flow enables us to estimate the inundation probabilities.

In the following section, we demonstrate how EVT can be applied to model peak discharge at a gage in Spring Creek near Axemann, PA. This will later serve as a component of our pricing model in our case study.

2.2 MODELING PEAK DISCHARGE FOR SPRING CREEK

Based on over 80 years (from 1940 to 2022) of data on daily mean discharge collected for USGS gage 01546500, Spring Creek near Axemann, PA, we model the distribution of the annual peak of daily mean flows at the gage.[12] We point out that we choose to model annual, rather than monthly or daily, peak flow out of the following reasons:

- The BM method requires the block size to be relatively large so that the limiting distribution can serve as a good approximation. Annual blocks can be understood as blocks of size 365 and thus are reasonably large.

---

[12] The code and data for implementing the extreme value modeling are available for download from the SOA web page that contains this report.

- The flood insurance policies we shall consider are of one-year term and, as a result, will need annual inundation probability. Although deriving annual inundation probabilities from, e.g., monthly peak flows is certainly feasible, deriving that from peak flows with matched frequency is more convenient.

- The thinning resulting from the use of annual maxima as opposed to monthly/daily maxima reduces serial dependence among the data.

- The usual concern of data scarcity with the annual maxima method discarding many observations is less worrisome since we have more than 80 years of data.

Figure 2 presents the discharge rate data, where the left graph shows the daily mean flow rates, in cubic feet per second (cfs), from October 1, 1940, to April 18, 2022, and the right graph shows the annual maximal flow rates from water-year (WY) 1941 to 2021. Here, we follow USGS's definition of water-year, which is the period from October of the previous calendar year to September of the current year. We exclude the 2022 WY data as we only have partial-year observations.

Figure 2
STREAMFLOW (CUBIC FEET PER SECOND) AT GAGE 01546500, SPRING CREEK NEAR AXEMANN, PA, FROM 1940 TO 2022. THE LEFT GRAPH SHOWS THE DAILY MEAN FLOW AND THE RIGHT GRAPH SHOWS THE ANNUAL MAXIMA.
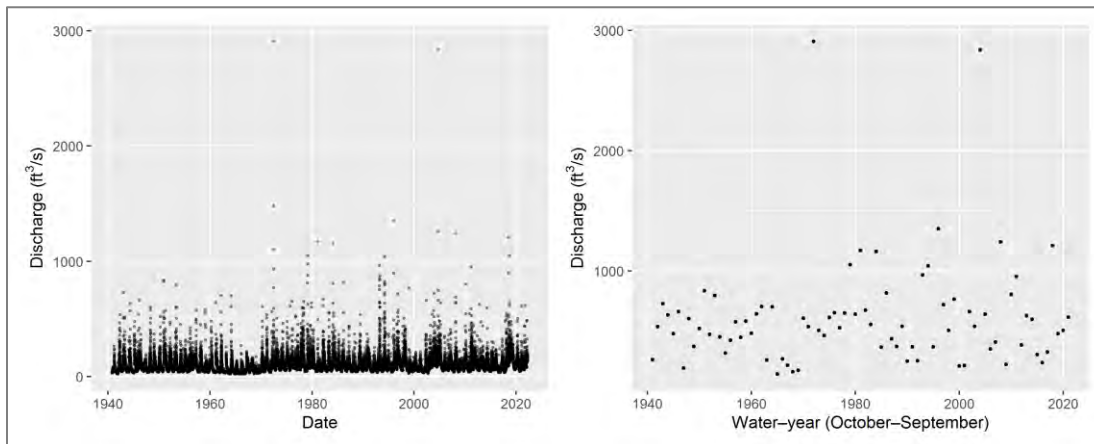
Data Source: United States Geological Survey (USGS)

Visual inspection of the data reveals that some observations of flow rates are extremely large as compared to others, motivating a model with heavy tail for the distribution of annual maximum flow rate. The nature of the observations, being maxima of blocks of observations, motivates the use of GEV. As such, we estimate a GEV distribution for the annual peak flow. Using maximum likelihood estimation (MLE), we find the estimates of the parameters μ, σ, and ξ are, respectively,

$$\hat{\mu} = 421, \quad \hat{\sigma} = 221, \quad and \quad \hat{\xi} = 0.22,$$

with corresponding standard errors of, respectively, 27.9, 22.2, and 0.09.

An estimated shape parameter of 0.22 indicates that the annual peak flow is indeed heavy tailed. Obviously, annual peak flows at different gages may have distinct tails and therefore their shape parameters may be different. Although estimates with very distinct values have been obtained by some researchers—for example, Smith (1987) estimated, using records up to 1986, that the shape parameter of

Potomac River annual peak flow is 0.42—our estimate of 0.22 is consistent with estimates for some rivers. For example, for Potomac River (with more data than used by Smith (1987)), the shape parameter is estimated as 0.19, and for Salt River, estimated as 0.28 (Katz et al. 2002).

Moreover, we also perform formal hypothesis testing regarding the heavy-tailedness of the peak flow by considering

- H_0: the annual maximum peak flow follows a GEV distribution with shape parameter $\xi = 0$; i.e., a Gumbel distribution, versus

- H_1: the annual maximum peak flow follows a GEV distribution with shape parameter $\xi \in \mathbb{R}$.

This can be easily done with a likelihood ratio test. The maximized log-likelihood is obtained as $l_0 = -580.4$ under H_0 and as $l_1 = -575.2$ under H_1. The likelihood ratio test statistics is given by

$$-2(l_0 - l_1) = 10.4$$

and is large compared to the values of a chi-squared random variable with degree of freedom 1, leading to a p-value of 0.1%. This confirms that the annual peak flow at the gage is heavy tailed.

Figure 3
DIAGNOSTICS GRAPHS FOR THE GEV FITTING. THE TOP LEFT IS A P-P PLOT, TOP RIGHT IS A Q-Q PLOT, BOTTOM LEFT IS A COMPARISON BETWEEN THE EMPIRICAL DENSITY, OBTAINED WITH A BANDWIDTH OF 84.78, AND THE MODELED DENSITY, AND BOTTOM RIGHT SHOWS THE PREDICTED LEVELS FOR VARIOUS RETURN PERIODS.
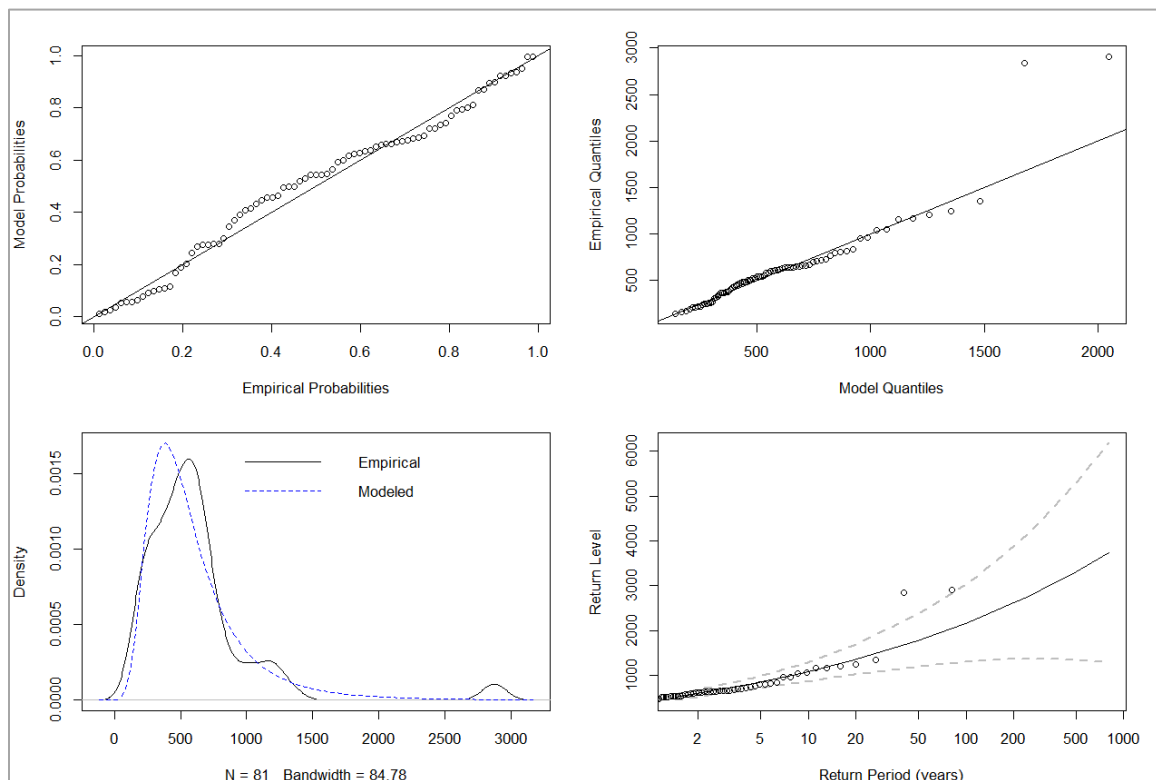
Data Source: Author's calculation

We present graphical diagnostics in Figure 3 to examine the goodness of fit using the GEV distribution. The points largely falling along the diagonals of the P-P plot (Figure 3, top left) and the Q-Q plot (Figure 3, top right) and the alignment between the empirical and modeled densities (Figure 3, bottom left) suggest that the GEV distribution is indeed a good fit to the data. Also presented in Figure 3 (bottom right) is a plot of the return levels for return periods ranging from 1 year to 1000 years. The point estimates of 20-year, 100-year, and 500-year return levels are 1,342 cfs, 2,164 cfs, and 3,318 cfs, respectively. Obviously, estimates of longer return periods mean more uncertainty with prediction and lead to wider interval estimates.

We note that in our estimation procedure, we have omitted the temporal dependence among the observations. This is less of a problem when the use of WY maxima thins the data and reduces serial dependence. In general, statistical inference for extreme values presents some challenges when there is serial dependence. For example, for likelihood-based inference, the impact of dependence on the likelihood function is unclear and so there may be substantial model risk in specifying the likelihood function. In practice, methods such as decluttering—that is, identifying the clusters of extremes and using observations across different clusters that can be roughly considered independent for statistical inference—have been widely used, although that could mean subjective determination of clusters and discard of useful information.

In addition, our model does not stress the possible nonstationary induced by climate change. We have used a simple stationary model for the annual peak flow since the data exhibit reasonable stationarity. Visual inspection reveals no trend over the years. Despite the evidence supporting an increasing trend in some closely related hydrometeorological variables, such as extreme precipitation, and an increased hydrological cycle, peak flows do not seem to exhibit similar increasing trends, at least not across the board. Analysis of peak flows for different rivers has led to different conclusions. See, e.g., Katz et al. (2002) for a summary of some research findings on peak flow trends.

Furthermore, as some authors have pointed out, despite the profound changes that have occurred to drainage basins throughout the U.S., proving the demise of stationarity of flood peaks has been a challenge. See, e.g., Villarini et al. (2009), who conclude that, although a statistically significant change point was detected in the series they analyzed, overall, no monotonic temporal patterns of the annual maximum instantaneous peak discharge were detected from the series.

To formally test the temporal trend of the peak flows at gage 01546500, we first estimate a GEV distribution with linear location function; that is, the location parameter of the GEV distribution is assumed to be $\mu(t) = \mu_0 + \mu_1(t - 1941)$, where t is in years. The MLEs of the parameters μ_0, μ_1, σ, and ξ are found to be, respectively,

$$\widehat{\mu_0} = 413, \quad \widehat{\mu_1} = 0.32, \quad \hat{\sigma} = 223, \quad and \quad \hat{\xi} = 0.21,$$

with corresponding standard errors of, respectively, 47.5, 0.95, 22.6, and 0.09. The large standard error of $\widehat{\mu_1}$ as compared to the estimated value indicates that the trend is not significant. This can be confirmed with the following hypothesis test:

- H_0: the location parameter of the GEV distribution of the annual peak flow is a constant; versus

- H_1: the location parameter of the GEV distribution of the annual peak flow is a linear function of time, with $\mu(t) = \mu_0 + \mu_1(t - 1941)$, $\mu_1 \neq 0$.

Again, this can be performed with a likelihood ratio test. The maximized log-likelihood is obtained as $l_0 = -575.23$ under H_0 and as $l_1 = -575.15$ under H_1. The likelihood ratio test statistics is given by

$$-2(l_0 - l_1) = 0.16.$$

With a test statistic following a chi-squared distribution with degree of freedom 1, this leads to a p-value of 0.70, meaning that there is no evidence that the annual peak flows at gage 01546500 exhibits a temporal trend. Of course, conclusions for different regional analyses may be different.

Note that although models with trends in other parameters and tests about other trends of the GEV parameters are possible, it makes intuitive sense and is customary in the literature to focus on the trend in the location parameter.

Lastly, we point out that although this is only one example of regional analysis of peak flow, for Spring Creek near Axemann, PA, regional analysis for other areas can be performed easily given the ample data available through USGS. Additionally, for ungagged locations, where there are no gages nearby, streamflow data may be generated using machine learning algorithms; see Appendix B. In some cases, the properties to be underwritten may be exposed to flood risks from multiple drainages along one or more river lines. Regional analysis can be tweaked and muti-dimensional or infinite-dimensional EVT (such as multivariate EVT and max-stable processes) can be applied to account for the interactions between the risks from the multiple drainages. The caveat is that, although deterministic numerical methods forecast large-scale pattern better, there are many statistical modeling/learning tools at our disposal for regional analysis and much finer characterization can be achieved. One implicit assumption for regional analysis to be useful or feasible at all, however, is that some level of homogeneity exists within a region or across regions. For example, when analyzing the tails of peak flows for multiple regions, it is customary to assume that the location and shape parameters vary spatially but the shape parameter is fixed across the regions. Further exploration along this direction is outside the scope of this report.

Section 3: Spring Creek Inundation Modeling

3.1 OVERVIEW

This section describes how we model inundation extent using observed discharge data and a well-established, off-the-shelf hydraulic model (HEC-RAS), allowing statistical analysis of flooding depth. This is a reliable procedure that depends on the availability of nearby streamflow gauges with long-term daily streamflow records. However, in places without such data (ungauged locations), we have run hydrologic models (and perhaps, most recently, machine-learning-based models) that can be used to provide alternative estimates. For readers without much background in hydrologic and hydraulic modeling, Appendix B provides the background information, discussion about recent progress including machine learning models, and an example for the Susquehanna River basin.

The main goal here is to demonstrate how discharge can be converted into inundation extent and relate it to a concept called height above nearest drainage (HAND). The case study is Bellefonte in the U.S. state of Pennsylvania, an area that has the potential to be flooded with varying frequency. Spring Creek is one of the tributaries of the Susquehanna River which passes through Bellefonte, and the residences in the vicinity of this river are under flood risk. The study area covers approximately a 3-mile reach of Spring Creek River that goes through Bellefonte as shown in Figure 4.

Figure 4
STUDY AREA (CIRCLED IN BLUE) IN BELLEFONTE, PA, USA.

This study consists of the following steps:

1. Finding the nearest streamflow gage as upstream inflow to the Bellefonte municipality area
2. Downloading the digital elevation model (DEM) file for the study area
3. Creating the mesh-grids for modeling and performing the HEC-RAS simulations
4. Calculating the flooded area for different flow rates
5. Locating the residence areas under the flood
6. Extracting HAND raster map for the area for statistical analysis

3.2 STREAMFLOW GAGE

We obtained daily historical discharge data from streamflow gages in the vicinity of the study area. Gages 01547100, 01546500, and 01546400 are the stream gages with daily observed data that are supported by the U.S. Geological Survey (USGS 2016) and Pennsylvania Department of Environmental Protection (Figure 5). The nearest streamflow gage, 01546500, was selected to provide the upstream inflow conditions to the model. The USGS website provides the following information for this station:

- STATION: 01546500 SPRING CREEK NEAR AXEMANN, PA

- LOCATION: Lat 40`53'23", long 77`47'40", Centre County, Hydrologic Unit 02050204, on right bank at upstream side of bridge on SR 3001, 1.6 mi west of Axemann, 1.8 mi southwest of Bellefonte, and 2.5 mi upstream from Logan Branch.

- DRAINAGE AREA: 87.2 square miles.

- PERIOD OF RECORD: October 1940 to current year.

- GAGE: Water-stage recorder and crest-stage gage. Datum of gage is 788.81 ft above National Geodetic Vertical Datum of 1929. Prior to Nov. 19, 1940, non-recording gage at same site and datum. Non-recording gage Mar. 6 to Sept. 30, 1995.

Figure 5
STREAMFLOW GAGES IN THE VICINITY OF BELLEFONTE.
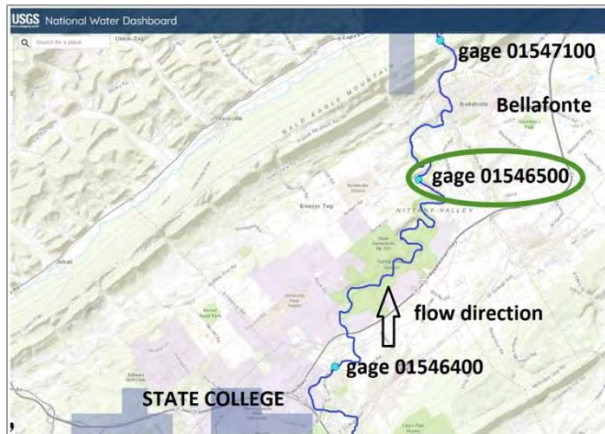
3.3 DIGITAL ELEVATION MODELS (DEM) FILE

A DEM file is a representation of topography excluding trees, buildings, and other objects on the surface, which can be used for flood modeling. DEM files are generally rasters that provide the elevation as discrete values on a grid. In this study, we used a 10-meter (1/9 arc-sec) resolution DEM to be used in HEC-RAS for flood modeling, which was downloaded from the USGS National Map (USGS, 2022).

3.4 FLOOD MODELING

This model demonstrates the flooded areas under different flood frequencies in the study domain using version 6.1 of Hydrologic Engineering Center-River Analysis System (HEC-RAS), a classical and widely employed flood inundation modeling software. We created a 2D mesh-grids (5 meter by 5 meter) along the

river starting from gage 01546500 and ending at a position downstream of Bellefonte (Figure 6). This grid is refined and helps to improve the model's accuracy close to the river.

We defined the upstream condition as a constant flow hydrograph, which changes for different scenarios to represent the different flood scenarios. For the downstream boundary, the flow condition was set to the normal depth.

Figure 6
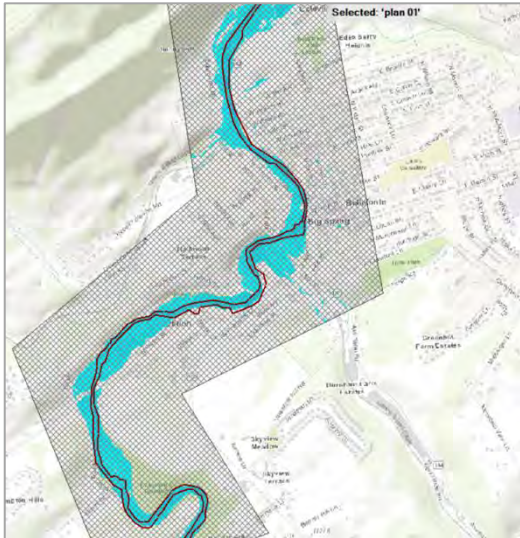
RIVER BOUNDARIES AND 2D MESH-GRID IN THE STUDY DOMAIN.

3.5 FLOOD AREAS FOR DIFFERENT SCENARIOS

In the last section, the base of the hydraulic model was created. We connected HEC-RAS model to Python so as to run it with multiple flood scenarios, and used the pyHMT2D package (Liu, 2022) and the examples in it to automatically run HEC-RAS for different scenarios. We saved the results, flooded area for each scenario, in raster TIFF files for water surface elevation (WSE) and flood depth. The resolution of these raster files is 10 meters by 10 meters.

According to a USGS report (Roland & Stuckey, 2008), the streamflow values for different recurrence intervals are provided for USGS stream gages as well as gage 01546500. We ran the model for flow rates ranging from 20 cubic meter per second (cms) to 400 cms (Table 1):

Table 1
FLOOD SCENARIOS MODELED FOR GAGE 01546500.

Scenario	Flow rates (cms)	Recurrence interval[13]
1	20	2-year
2	37	5-year
3	40	—
4	55	10-year
5	60	—

[13] These are recurrence levels as reported by Roland and Stuckey (2008) and may be different from those derived from the EVT model we used in Section 3.

6	80	—
7	100	—
8	118	50-year
9	120	—
10	140	—
11	161	100-year
12	200	—
13	250	—
14	318	500-year
15	400	—

3.6 RESIDENTIAL AREAS

We identified residential areas based on satellite maps of Bellefonte, which excluded, for example, parks and streets (Figure 7).

Figure 7

RESIDENTIAL AREA IN BELLEFONTE CLOSE TO SPRING CREEK.

The overlaps between these polygons and water depth raster files will demonstrate the risk to the residential area. The resolution of the result is 10 meters by 10 meters.

3.7 CALCULATING HAND

Height Above Nearest Drainage (HAND) describes the vertical distance between a location (pixel in raster DEM file) on a map and its nearest stream. First, flow directions and accumulations were calculated based on the DEM to get flow paths in the study area. This analysis was done in QGIS (an open-source Geographic Information System software). The next step was to calculate the sub-catchments of each river pixel, which was done using the sub-catchment toolbox in QGIS (Figure 8-a and b). Then, the elevation of the drainage was calculated using the minimum elevation in each zone. The last step was to subtract the DEM value from the elevation of the drainage (Figure 8-c).

Figure 8
HAND CALCULATION GENERAL WORKFLOW.
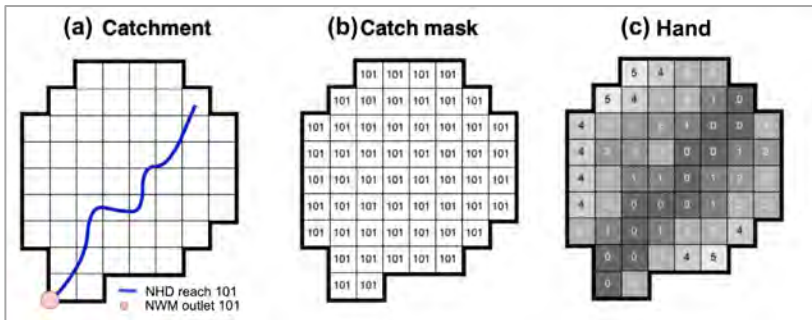
The HAND calculation result for Bellefonte area was a raster TIFF file with a 10-meter by 10-meter resolution (Figure 9).

Figure 9
HAND MAP FOR BELLEFONTE.
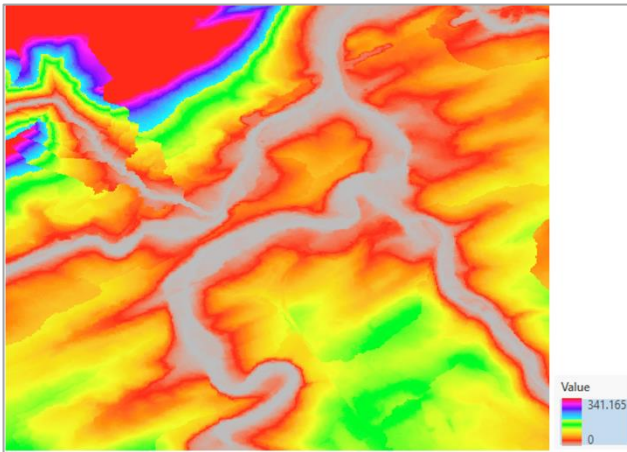
3.8 INUNDATION MODELING RESULTS

We ran the model for 15 flood scenarios, corresponding to different recurrence intervals (Table 1). Table 2 describes the overall results for the number of pixels (each 10 meter by 10 meter) under the flood and the average water depth in the residential area:

Table 2
RESULTS FOR DIFFERENT FLOOD SCENARIOS.

Scenario	Flow estimates (cms)	Number of pixels in residence area under flood	Average water depth among the flooded residential pixels (meter)
1	20	484	0.2018
2	37	688	0.2991
3	40	725	0.3162
4	55	844	0.3772
5	60	883	0.4005
6	80	1002	0.4903

7	100	1101	0.5666
8	118	1146	0.6408
9	120	1162	0.6443
10	140	1233	0.7021
11	161	1279	0.7698
12	200	1388	0.8645
13	250	1467	0.9919
14	318	1567	1.2067
15	400	1681	1.2924

The depth maps below show the inundation areas and depths for flood magnitudes of 55, 118, and 161 cms. Again, the recurrence level of 10-year, 50-year, and 100-year are as calculated by Roland and Stuckey (2008). Clearly, the flooded area increases with higher-magnitude floods (Figures 10 and 11):

Figure 10

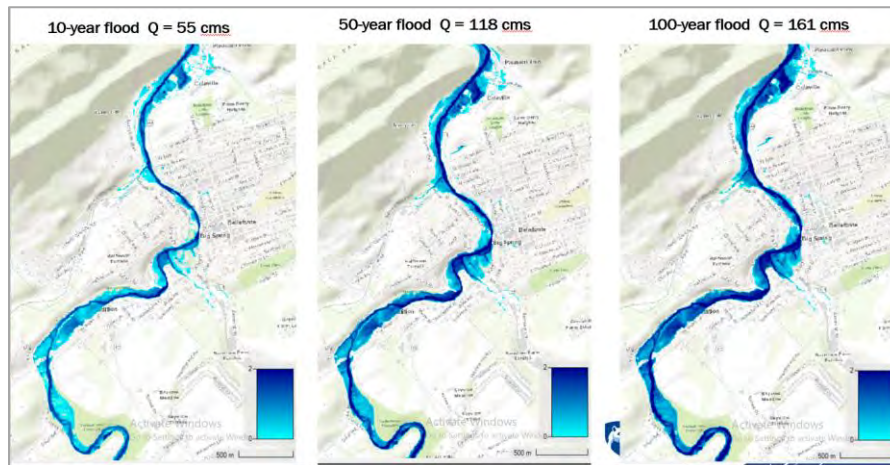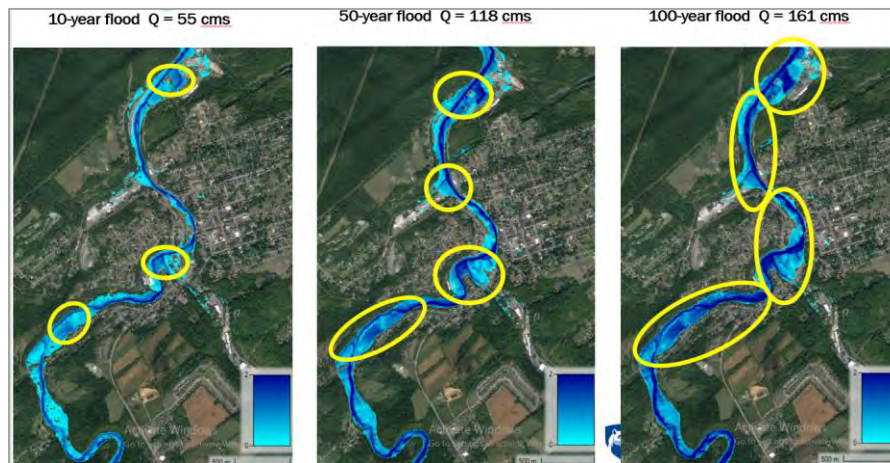FLOODED AREA FOR 10, 50, AND 100-YEAR RECURRENCE INTERVALS.

Figure 11

HOW FLOOD AREA INCREASES WHEN THE FLOW INCREASES.

As we mentioned before, Figures 10 and 11 only illustrate the total flooded area. In order to show the residential area under the flood, we cropped the flood map for the residential area. The simulated data will be used for further flood probability assessment in the next section.

# Section 4: Inundation Probability Assessment

4.1 PROBABILISTIC CLASSIFCIATION OF INUNDATION

The first step of our regional analysis led to a distribution of annual peak streamflow at a nearby gauge (Section 2). Our second step is to derive a relation that, given the peak streamflow rate, will probabilistically classify whether each property in the region will be inundated in a year, based on the inundation extent data generated by HEC-RAS (Section 3).

The data contain records for 1,681 locations, including their longitudes, latitudes, and HANDs, near Spring Creek around Axemann/Bellefonte of Pennsylvania, as shown in Figure 12.

Figure 12

LOCATIONS INCLUDED IN THE REGIONAL ANALYSIS FOR SPRING CREEK NEAR AXEMANN AND BELLEFONTE, PENNSYLVANIA.
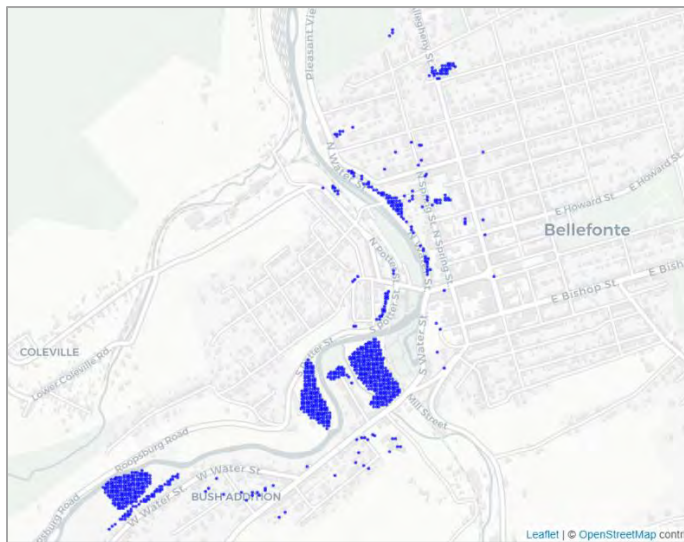
Data Source: Authors' selections

As a recap, we ran HEC-RAS for different rates of streamflow at USGS gage 01546500 to obtain the inundation levels at each of the locations. The inundation level changes with respect to the streamflow rate. We run the model for 15 different levels of streamflow rates ranging from 20 cms to 400 cms.

A visualization of the data through Figure 13 and Figure 14 helps us better understand the relation between the inundation probability/level and variables such as peak streamflow and the HAND at the location.

Figure 13 shows the boxplots of inundation levels at the locations, grouped by the streamflow rates. The groups in the left graph are equally spaced on the x-axis, while the spacing for those in the right graph is scaled based on the flow rate values and is likely more revealing about the functional relation between the flow rate and inundation level. In Figure 14, we use the red dots to represent the locations that are not inundated, and the light-blue ones to represent those that are inundated. Larger dots mean higher water levels.

Figure 13

BOXPLOTS OF THE INUNDATION LEVELS AT THE LOCATIONS INCLUDED IN THE REGIONAL ANALYSIS, GROUPED BY PEAK FLOW RATES. THE GROUPS IN THE LEFT GRAPH ARE EQUALLY SPACED ON THE X-AXIS, AND THE SPACING FOR THOSE IN THE RIGHT GRAPH IS SCALED BASED ON THE FLOW RATES.
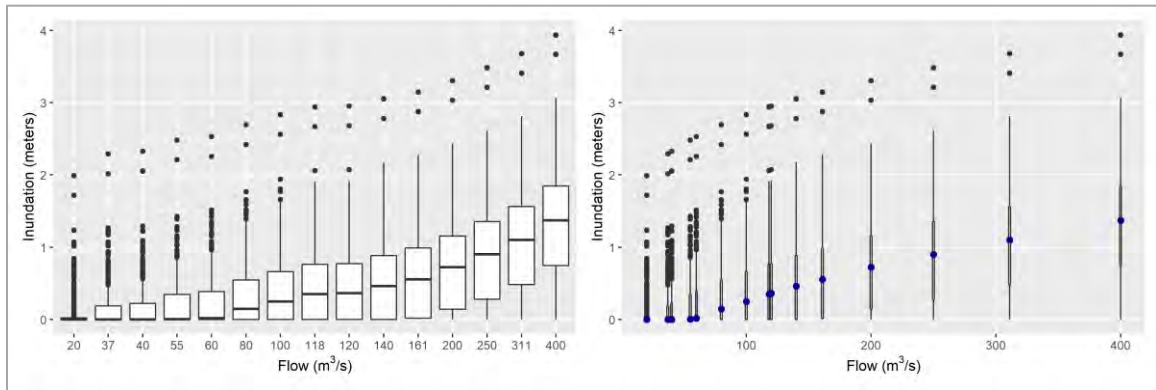
Data Source: HEC-RAS models

Figure 14

INUNDATION LEVELS AT EACH OF THE LOCATIONS IN THE REGIONAL ANALYSIS. THE BROWN DOTS ARE THE LOCATIONS THAT ARE INUNDATED AND THE LIGHT-BLUE DOTS ARE THOSE THAT ARE NOT. THE LARGER LIGHT-BLUE DOTS INDICATE HIGHER WATER LEVELS.
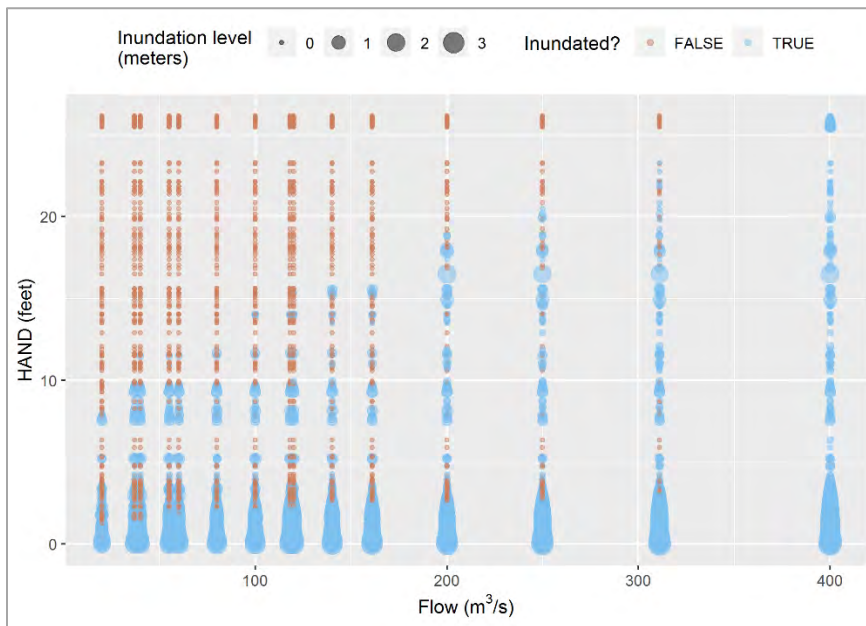
Data Source: HEC-RAS models

Apparently, both figures show that the greater the flow rate, the more likely the location is inundated and the higher the inundation level. The same can be concluded if the HAND at the location is lower.

In our pricing model, we focus more on the inundation probability than the actual inundation level, mainly because we treat inundation probability and not inundation level as a component of the pricing model. It is worth pointing out that despite some recent research that shows the importance of inundation level in determining building damage, their exact relation is far from clear.

As a proof of concept, we propose to use the annual peak flow rate and the HAND as the predictors of annual inundation probability; obviously, other variables such as distance to water can also be included in the model. In fact, some of the other variables have been found important in determining flood risk; for example, Giovannettone et al. (2018) find that elevation and distance to water have strong influence on flood risk in the urban and coastal subregions, whereas distance to water and surface geology have dominating influence in the rural subregion. As a comparison, the HAND, as a measurement of the height above the nearest drainage, is likely to have a more important role to play than elevation (above sea level) when evaluating in-land flood risks.

We randomly choose 70% of the data to train the following logistic regression model as a probabilistic classifier about whether a property is inundated:

$$logit\big(P(Inundated|HAND, Flow)\big) = \beta_0 + \beta_1 HAND + \beta_2 Flow,$$

where HAND and Flow are the HAND (in feet) at the location and the annual peak flow rate (in cms) at the nearest gage. The estimation results are summarized in Table 3.

Table 3
LOGISTIC REGRESSION ESTIMATION RESULTS.

Coefficients					
		Estimate	Std. Error	z value	p value
(Intercept)		-0.287	0.034	-8.41	0.000***
HAND		-0.260	0.008	-33.42	0.000***
Flow		0.015	0.000	47.53	0.000***
Significance code:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	
Deviance Residuals					
Min		1Q	Median	3Q	Max
-2.980		-0.990	0.322	0.814	2.339
AIC	16,998.3				
BIC	17,021.6				

The negative coefficient of HAND and positive coefficient of Flow are consistent with our intuition that larger flows and lower HAND lead to higher probabilities of inundation. Moreover, a test with the remaining 30% of the data shows that the predication accuracy of the model is high at 77%.

4.2 UNCONDITIONAL PROBABILITY OF INUNDATION

With a model for the annual peak flow rate and a model for conditional inundation probability that takes the flow rate as an input—and noting that the HAND at a location is deterministic—we are ready to derive the unconditional inundation probability. Specifically, the annual inundation probability for a property in the region with HAND h is given by

$$p(h) = \int_{\mu - \frac{\sigma}{\xi}}^{\infty} \frac{\exp\{\beta_0 + \beta_1 h + \beta_2 x\}}{1 + \exp\{\beta_0 + \beta_1 h + \beta_2 x\}} \frac{d}{dx} G_{\xi;\mu,\sigma}(x)$$

$$= \frac{1}{\xi} \int_{\mu - \frac{\sigma}{\xi}}^{\infty} \frac{\exp\{\beta_0 + \beta_1 h + \beta_2 x\}}{1 + \exp\{\beta_0 + \beta_1 h + \beta_2 x\}} \exp\left\{-\left[1 + \frac{\xi(x - \mu)}{\sigma}\right]^{-\frac{1}{\xi}}\right\}\left[1 + \frac{\xi(x - \mu)}{\sigma}\right]^{-\frac{1}{\xi}-1} dx$$

$$= \frac{\exp\{\beta_0 + \beta_1 h\}}{\xi} \int_{\mu-\frac{\sigma}{\xi}}^{\infty} \frac{\exp\{\beta_2 x - [1 + \xi(x-\mu)/\sigma]^{-1/\xi}\}}{1 + \exp\{\beta_0 + \beta_1 h + \beta_2 x\}} \left[1 + \frac{\xi(x-\mu)}{\sigma}\right]^{-\frac{1}{\xi}-1} dx$$

where $\beta_0, \beta_1, \beta_2, \mu, \sigma$, and ξ are the estimated parameter values from the logistic regression and fitting of the generalized extreme value distribution. We numerically evaluate the integration to obtain the inundation probabilities for different locations with given HANDs.

In the examples given in Section 1.4, the flood insurance premia are calculated using the inundation probability given above. It is important to remember that the inundation probability function is derived from this particular regional analysis and will need to be adjusted accordingly for a different target region.

Section 5: Claims Modeling

This section aims at modeling the expected claim amounts for buildings that are inundated.

5.1 DATA

We mainly rely on two datasets for our claims modeling: the American Community Survey (ACS) data on property values and the NFIP claims data.

The property value dataset is collected from the U.S. Census Bureau[14], which estimates the median property values for housing units through the ACS at census tract level. For census tracts where the estimated median property value is missing, we use mean imputation that imputes the value with the average of the estimates across the census tracts in the same county[15]. Moreover, the estimated median property values are censored from below at $10,000 or from above at $2,000,000. For those censored records, the censored values of $10,000 and $2,000,000 are used for our normalization.

We remove those census tracts for which median property values are not available. In total, our data set contains median property values for 74,001 of the 84,414 census tracts in the U.S.

The NFIP claims dataset used for our analysis is publicly available at FEMA. The records for our analysis were accessed on June 25, 2021 and were last refreshed by FEMA in May 2021.[16]

The claims data range from 1970 to 2021, consisting of around 2.5 million records. The policies cover single-family homes, multi-unit residential buildings, as well as non-residential buildings. Among the 2.5 million records, over 2 million records are claims from single-family homes. Given that the majority of policies are for single-family homes, and multi-unit buildings and non-residential buildings could have coverage limits that are very different from single-family homes—for example, for residential properties, the current coverage limit is $250,000 for the building and $100,000 for the contents, while for commercial properties, the current coverage limit is $500,000 for both the building and the contents—we focus our analysis on single-family homes.

The dataset contains forty features[17]. We ignore the features that are not indicative of the flood risk, such as the indicator of whether a building is reported as being a non-profit in the policy application, the indicator of whether a building is reported as being a house of worship in the policy application, the original date of the flood policy, and the unique ID assigned to the record.

Among the variables selected for our study, in addition to the continuous variables and date variables, there are nine categorical variables: basement type, whether the property is elevated, whether it has an elevation certificate, the number of floors, the location of contents, obstruction type, whether the property is the primary residence, whether its construction started before or after publication of the Flood Insurance Rate Map (FIRM), and the flood zone the property is located in.

---

[14] See https://data.census.gov/cedsci/table?hidePreview=true&tid=ACSDT1Y2019.B25077. The estimates provided by the U.S. Census Bureau are not available for every year; the estimates are calculated every few years. We use the 2019 property values.
[15] There are three counties where median property value data are not available for any of the census tracts. For census tracts in those counties, we use the state-level average to impute the missing records.
[16] The dataset was retrieved through https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims-v1. This report uses the Federal Emergency Management Agency's OpenFEMA API but is not endorsed by FEMA. The Federal Government or FEMA cannot vouch for the data or analyses derived from these data after the data have been retrieved from the Agency's website(s).
[17] See https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims-v1 for more details.

A property's flood zone has been a good indicator of its flood risk and has been a rating factor before FEMA implemented its new rating system, Risk Rating 2.0, in October 2021. FEMA designated flood zones include those in the Special Flood Hazard Areas (SFHAs) (i.e., Zones A, AE, A1-A30, AH, AO, AR, A99, V, VE, V1-V30), the areas of moderate flood hazard (i.e., Zones B and X (shaded)), the areas of minimal flood hazard (i.e., Zones C and X (unshaded)), and the areas of undetermined flood hazard where flooding is possible.[18] Among those in the A Zones, about 89% are in Zones AE, A1-A30, or AH, and among those in the V Zones, about 98% are in Zones VE or V1-V30. This means that most of those properties have Base Flood Elevation (BFE) derived.[19]

Following the literature (see, e.g., Wing et al. (2020)), we group the flood zones into high-risk areas, which include the SFHAs, and low-risk areas, which include all other areas.[20] Among all the records, about 30% are from properties in low-risk areas and about 70% in high-risk areas.

We also create a variable out of the features in the data: the age of the property at the date of loss, calculated as the date of loss minus the date when the original construction date of the property. Apparently, the age of the property reflects the risk of damage better than the dates *per se*.

We remove entries that are obviously erroneous. For example,

- we remove claim amounts and coverage limits that are greater $250,000, which is the upper limit of NFIP coverage on building for single family homes,

- we remove the records with claim amount greater than coverage,

- we remove entries with zero amount on building[21], and

- to produce the figures for visualization, we removed records with latitude and longitude coordinates outside the U.S. territory.

In addition, we notice that there are many missing or unreasonable values of elevation difference, which we believe has a strong impact on our damage estimation. Therefore, we manually calculate the elevation difference by subtracting the BFE from the lowest floor elevation of the property—this is how elevation difference is defined. We remove the records with unreasonable elevation differences: those above 500 feet or below -500 feet. These values are most likely entered with error, some are possibly in inches and others hard to explain; see Wing et al. (2020) for similar discussions about possible errors in FEMA datasets.

In our models, if there are missing values in the response variable, we omit those records. In the end, the clean dataset has about 1.4 million records for our analysis.

We note that what we are analyzing is the insurance payouts, not the losses. To derive the actual losses from the floods, one would need to the deductible levels of the corresponding policies. According to FEMA policy data[22] that contain about 50 million records of policies on single family homes, for coverage on building for single family homes, about 82% of the policies have deductible of 1000, 1250, or 1500, with

---

[18] See, e.g., https://www.grar.org/wp-content/uploads/2017/07/Realist-Flood-Zone-Codes.pdf.
[19] BFE is provided, possibly at selected intervals, for Zones AE, A1-A30, AH, VE, and V1-V30.
[20] Note that we treat zones AR and A99 as low-risk zones, even though they appear under FEMA's list of high-risk zones. This is consistent with the treatment by the Community Rating System. Note that AR zones are areas with a temporarily increased flood risk (due to the building or restoration of a flood control system), and A99 zones are defined as areas "with a 1% annual chance of flooding that will be protected by a Federal flood control system where construction has reached specified legal requirements."
[21] Some claims only have payments for content loss or increased cost of compliance.
[22] Accessed on July 29, 2021, through https://www.fema.gov/openfema-data-page/fima-nfip-redacted-policies-v1.

the former two accounting for about 72%. As of November 2021, the minimum deductible levels for building coverage of $100,000 or less is $1,000 and that for building coverage over $100,000 is $1,250, except that for Pre-FIRM buildings (i.e., buildings that were constructed or substantially improved on or before December 31, 1974, or before the effective date of an initial FIRM), the minimum deductible for building coverage is $1,500[23]. For coverage on content for single-family homes, about 80% of the policies have deductible of $500, $1,000, or $1,250. The options have changed in the history of NFIP, and other options were available. Currently, there are six levels of choice for deductibles for single family homes, up to $10,000 for both building and contents, but there is not great heterogeneity in policyholders' choices. Therefore, studying the claims does shed light on the ground-up losses.

5.2 VISUALIZATION

We first show a few figures to visualize the geospatial characteristics of the flood insurance claims. Figure 15 shows the inflation-adjusted claims paid on building (in 2020 dollars), aggregated over the years of 1970–2021, at county level across the nation.

Figure 15

COUNTY-LEVEL PLOT OF AMOUNT PAID ON BUILDING CLAIMS, SMOOTHED USING POWER TRANSORMATION, ACROSS THE CONTIGUOUS UNITED STATES. THE AMOUNTS ARE INFLATION ADJUSTED TO 2020 DOLLARS USING CPI DATA FROM THE U.S. BUREAU OF LABOR STATISTICS.[24]
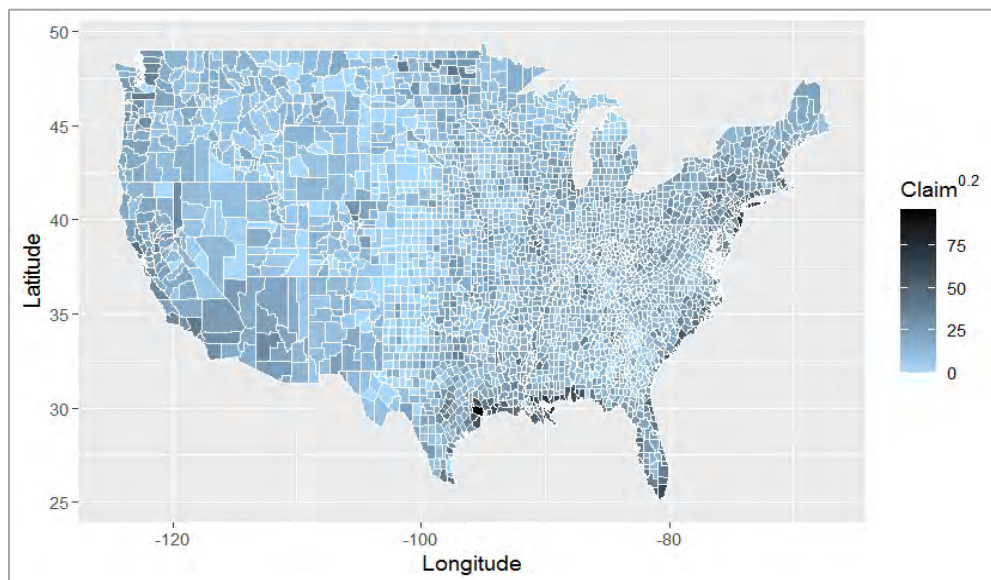
Data Source: Federal Emergency Management Agency (FEMA) and U.S. Bureau of Labor Statistics (BLS)

It is clear that the largest amounts of claims come from coastal areas as well as the Mississippi river basin. The county with the highest aggregated claims across the nation is Harris County in Texas. Figure 16 shows in more detail the claims from Texas, one of the states that are most susceptible to flood risk.

---

[23] See page 3-42 of https://www.fema.gov/sites/default/files/documents/fema_nfip-flood-insurance-manual-sections-1-6_oct2021.pdf for more details.
[24] For a better visual effect, we smooth the claim amounts by taking a one fifth power.

In Figure 16, the left graph is a county-level plot of the paid-on building in Texas. The right graph is a scatter plot of the inflation-adjusted amounts paid on building in Texas[25]. Scatter points of larger size indicate larger claim amounts and darker color indicates higher claim frequencies. We see that the Houston area (mainly Harris County, but also Fort Bend and Montgomery, etc.) has by far the largest amount of claims. Substantial amounts of claims also come from Beaumont, Dallas, and San Antonio areas.

Figure 16

AGGREGATED CLAIM AMOUNTS PAID ON BUILDING (TEXAS). THE AMOUNTS ARE INFLATION ADJUSTED TO 2020 DOLLARS USING CPI DATA FROM THE U.S. BUREAU OF LABOR STATISTICS.

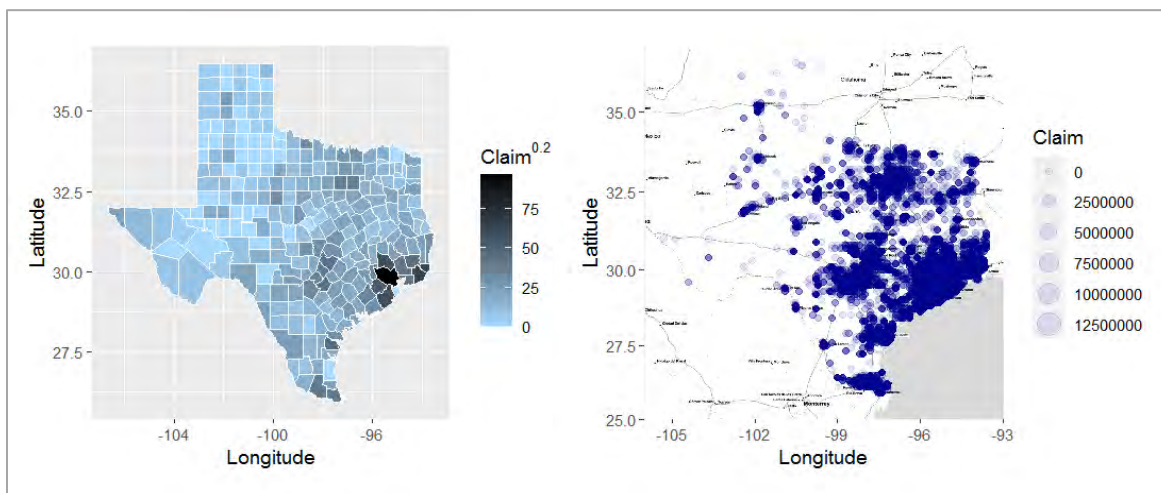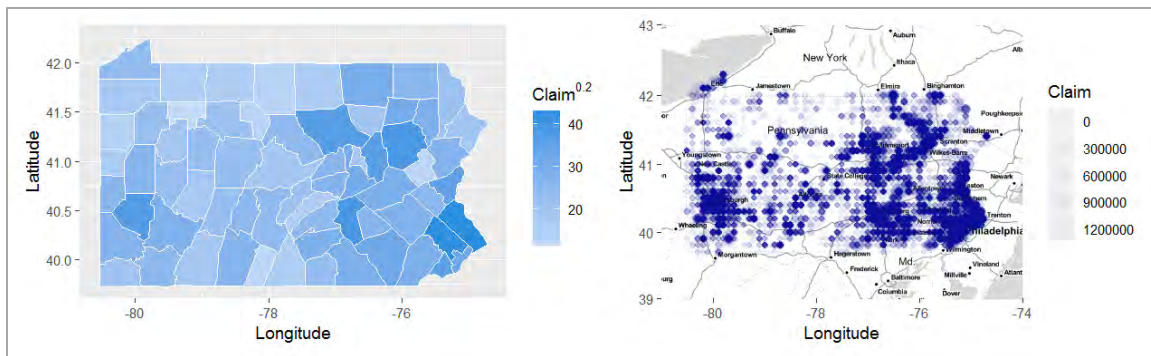Data Source: Federal Emergency Management Agency (FEMA) and U.S. Bureau of Labor Statistics (BLS)

Figure 17

AGGREGATED CLAIM AMOUNTS PAID ON BUILDING (PENNSYLVANIA). THE AMOUNTS ARE INFLATION ADJUSTED TO 2020 DOLLARS USING CPI DATA FROM THE U.S. BUREAU OF LABOR STATISTICS.

Data Source: Federal Emergency Management Agency (FEMA) and U.S. Bureau of Labor Statistics (BLS)

Lastly, we show the claims from Pennsylvania in Figure 17 as another example. Similarly, the left graph is a county-level plot of the inflation adjusted and smoothed claim amounts paid on building in Pennsylvania. The right graph is a scatter plot of the inflation-adjusted claim amounts paid on building. It is clear that the

---

[25] Again, the amounts are smoothed by raising to the one fifth power.

Philadelphia area (Bucks and Montgomery Counties) has produced the largest amounts of claims while some counties in central Pennsylvania also face significant flood risk.

5.3 A MIXED (GENERALIZED) BETA MODEL FOR CLAIMS MODELING

We focus on the claims paid on building and therefore exclude the variables that are only related to the content coverage of the policy. Ultimately, we hope the model the claim amounts as a percentage of the property value. However, because of the Privacy Act passed by Congress in 1974, FEMA does not release property value data. We instead use the median property value in the county as a benchmark and define a response variable by the ratio of the claim amount paid on building to the median property value.

We now explore a few candidate models for claim amounts and investigate the link between the expected damage and property characteristics. Recent research by Wing et al. (2020) studies the relation between flood damage and inundation depth concludes that flood damages follow a beta distribution and the distribution shifts when the water depth varies. Motivated by this research, we fit a conditional generalized beta model to the flood damages. In our model, the distribution of flood damage shifts according to, not only possible water depth, but also other variables and their interactions. This can be viewed as a generalization of the model in Wing et al. (2020). Specifically, we consider a (generalized) beta regression model (see Ferrari and Cribari-Neto (2004)) that regresses the damage ratio—that is, the damage as a percentage of property value—against the relevant variables.

Note that a standard beta distribution with parameters $a > 0$ and $b > 0$ has density function

$$f(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1}, \qquad 0 < y < 1,$$

where $\Gamma(\cdot)$ is the gamma function. A different parametrization of $\mu = a/(a+b)$—which represents the mean of the distribution—and $\phi = a + b$ enables us to write it as

$$f(y) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \qquad 0 < y < 1.$$

We assume that conditional on the feature variable $\mathbf{X}$ with m features, the response variables $Y_1, \dots, Y_n$ are independent, following beta distributions with respective means $\mu_1, \dots, \mu_n$[26]. Moreover, suppose that, for a link function $g(\cdot)$, the conditional mean of the response variable satisfies

$$g(E[Y_i|\mathbf{X}]) = \sum_{j=1}^{m} \mathbf{X}_{ij}\beta_j, \qquad i = 1, \dots, n,$$

or, simply,

$$g(\mu_i) = \sum_{j=1}^{m} x_{ij}\beta_j, \qquad i = 1, \dots, n,$$

[26] Independence may be a strong assumption for claims from the same geographical region. Nonetheless, spatial dependence would be more of a problem for the risk of inundation than for flood damages, because, for the latter, property characteristics of the building in the region could be diverse enough to justify the independence assumption.

where $\beta = (\beta_1, \dots, \beta_m)$ is the vector of regression parameters. The explanatory variables we use for our beta regression model are listed in Table 4.

Table 4
DEFINITION OF VARIABLES.

Variable	Value	Description
damageRatio	Continuous	The ratio of claim amount to the median property value for the census tract, standardized to values in (0,1).
coverage	Continuous	Total Insurance Amount in dollars on the building
elevationDifference	Continuous, in number of feet	Difference between the elevation of the lowest floor used for rating or the floodproofed elevation and the base flood elevation (BFE). BFE is the elevation, in feet, at which there is a 1% chance per year of flooding from the elevation certificate.
age	Continuous, in years	Age of the building at the time of loss. Calculated as the year of loss minus the year of building construction.
postFIRM	1 for post FIRM building, 0 otherwise	If the construction of the building or a substantial improvement of the building was started after publication of the Flood Insurance Rate Map (FIRM) on December 31, 1974, and after the effective date of the initial FIRM for the community, it is considered as a post FIRM building.
crs	Continuous	Rating of the policy by the Community Rating System (CRS), ranging from 0 to 10
SFHA	1 if in a SFHA, 0 otherwise	Indicator of whether the building is in a SFHA zone
crsSFHA		Interaction term between CRS and SFHA
basementType1	1 if type 1 basement, 0 otherwise	Indicator of whether the building has a finished basement/enclosure
basementType2	1 if type 2 basement, 0 otherwise	Indicator of whether the building has an unfinished basement/enclosure or crawlspace (above-grade or below-grade)
floors2	1 if two-story property, 0 otherwise	Indicator of whether the building is a two-story building
floors3	1 if three or more floors, 0 otherwise	Indicator of whether the building has three or more stories

Note that the standard beta regression requires the response variable to be in $(0,1)$, while for our response variable, the estimated median property value could be smaller than the claim amount, leading to responses that exceed 1. We do a uniform transformation to bring the response variable into $(0,1)$. Using a logit link function, we estimate the following three models and summarize the results in Table 5–Table 7.

First, we estimate Model I, given as follows:

$$g(damageRatio) = intercept + \beta_1 coverage + \beta_2 SFHA + \beta_3 basementType1 + \beta_4 basementType2 + \beta_5 crs + \beta_6 floors2 + \beta_7 floors3 + \beta_8 postFIRM + \beta_9 age + \beta_{10} elevationDifference.$$

The results are presented in Table 5.

Most of the estimation results for Model I make intuitive sense. For example, the damage ratio is expected to be larger for buildings with higher coverage limits and for older buildings. On average, damage ratios for buildings with basement are lower, because, presumably, the building materials and other covered items are generally less costly than those above grade.

Table 5
BETA REGRESSION RESULTS FOR MODEL I.

Model I				
Coefficients				
	Estimate	Std. Error	z value	p value
(Intercept)	-2.921	0.068	-43.024	0.000***
coverage	5.962e-07	5.297e-08	11.256	0.000***
SFHA	0.005	0.065	0.080	0.936
basementType1	-0.408	0.014	-29.384	0.000***
basementType2	-0.297	0.009	-31.758	0.000***
crs	0.092	0.002	39.115	0.000***
floors2	-0.364	0.009	-39.424	0.000***
floors3	-0.495	0.012	-41.352	0.000***
postFIRM	-0.119	0.008	-15.183	0.000***
age	6.596e-06	5.264e-07	12.529	0.000***
elevationDifference	-0.005	0.001	-4.652	0.000***
Significance code:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1			
Pseudo R-squared:	0.1707			
Standardized weighted residuals				
Min	1Q	Median	3Q	Max
-5.4462	-0.5168	0.0538	0.6131	4.3146
AIC	-304,209.3			
BIC	-304,098.1			

Buildings with a better CRS rating (i.e., a lower value) are expected to experience a lower damage ratio. This is because CRS credits are usually obtained when, e.g., flood protection techniques such as dry and/or wet floodproofing are used, where dry floodproofing refers to, e.g., making walls and floor watertight so that flood cannot enter the building, and wet floodproofing refers to, e.g., taking measures to reduce the damage once water enters the building.[27] These flood protection techniques are often effective in reducing flood damages.

Moreover,

- the more floors a building has, the smaller damage ratio we generally expect, because, for the same water level, the proportion of under water is expected to be smaller for a building with more floors;

- post FIRM buildings, when inundated, are expected to result in smaller damages, because they are subject to more stringent building codes;

- with building materials aging, it renders older buildings less likely to withstand water damage;

---

[27] See NFIP Community Rating Coordinator's Manual 2017 at https://www.fema.gov/sites/default/files/documents/fema_community-rating-system_coordinators-manual_2017.pdf.

- the higher the building is elevated, the smaller the damage ratio is expected to be. A higher elevation means a lower water level for a given flood and hence a smaller damage, as is documented in the literature (Wing et al. (2020)).

Note that the coefficient of SFHA is not significant. The expected damage for buildings in SFHAs is not clear as compared to those outside SFHAs. It is possible that buildings in SFHA are subject to more stringent building codes[28] and are thus more resilient to flood damage, but also possible that the water level is likely higher for SFHA properties when inundated. While Model I seems to suggest the latter—although not conclusively—Models II and III below are more indicative of the former.

It is worth pointing out that in FEMA's new rating system Risk Rating 2.0, flood zone is no longer used for flood insurance premium calculation; see Horn (2021). It is, however, not entirely clear that ignoring the building's flood zone when pricing its flood insurance is a wise decision. As we will see in our Models II and III, flood zone does play a significant role in determining the property's damage ratio when its interaction with CRS rating is accounted for.

Table 6
BETA REGRESSION RESULTS FOR MODEL II.

Model II				
Coefficients				
	Estimate	Std. Error	z value	p value
(Intercept)	-2.049	0.286	-7.172	0.000***
coverage	5.96E-07	5.30E-08	11.245	0.000***
SFHA	-0.869	0.286	-3.039	0.002**
basementType1	-0.407	0.014	-29.340	0.000***
basementType2	-0.297	0.009	-31.751	0.000***
crs	-0.038	0.427	-0.899	0.368
floors2	-0.364	0.009	-39.414	0.000***
floors3	-0.495	0.012	-41.363	0.000***
postFIRM	-0.119	0.008	-15.219	0.000***
age	6.57E-06	5.27E-07	12.486	0.000***
elevationDifference	-0.005	0.001	-4.667	0.000***
crsSFHA	1.308	0.043	3.059	0.002**
Significance code:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1			
Pseudo R-squared:	0.1709			
Standardized weighted residuals				
Min	1Q	Median	3Q	Max
-5.446	-0.5170	0.0539	0.613	4.316
AIC	-304,217.4			
BIC	-304,096.9			

Second, considering that premium credits earned by the same CRS rating are different for SFHA properties and non-SFHA properties, we introduce an interaction term between CRS and flood zone. We then estimate Model II, given by

---

[28] See, e.g., https://www.teamcomplete.com/wp-content/uploads/2017/02/Floodplain-Management-Regulations.pdf.

$$g(damageRatio)$$
$$= intercept + \ \beta_1 coverage + \beta_2 SFHA + \beta_3 basementType1 + \beta_4 basementType2$$
$$+ \beta_5 crs + \beta_6 floors2 + \beta_7 floors3 + \beta_8 postFIRM + \beta_9 age$$
$$+ \beta_{10} elevationDifference + \beta_{11} crsSFHA.$$

The results are presented in Table 6. As mentioned above, in Model II, flood zone plays a significant role in determining the expected flood damage. The model also indicates that, for SFHA properties, a better CRS rating means a lower expected flood damage. For non-SFHA properties, the CRS rating is not significant, meaning that in low flood risk areas, a better rating does not necessarily mean a lower damage ratio for an inundated property. This is roughly in line with NFIP's rating system: for properties not in SFHA's, FEMA will give no more than 10% premium credit even for the best CRS rating—the three premium credit levels are 0, 5%, and 10%—as compared to the 45% premium credit for those in SFHA with best CRS rating[29].

To compare the two models above, we observe that Model I is favored based on the BIC criterion and Model II is favored based on the AIC criterion. We recommend Model II, because it accounts for the different impacts the change of CRS rating could have on the expected claim amount for properties in different flood zones.

Table 7
BETA REGRESSION RESULTS FOR MODEL III.

Model III				
Coefficients				
	Estimate	Std. Error	z value	p value
(Intercept)	-2.593	1.714	-151.30	0.000***
coverage	7.25E-07	2.542-08	28.81	0.000***
SFHA	-0.218	0.019	-11.48	0.000***
basementType1	-0.531	0.007	-79.42	0.000***
basementType2	-0.397	0.006	-66.68	0.000***
crs	0.032	0.002	13.79	0.000***
floors2	-0.322	0.006	-53.92	0.000***
floors3	-0.425	0.007	-59.19	0.000***
postFIRM	-0.062	0.004	-14.15	0.000***
age	8.79E-06	2.03E-07	43.23	0.000***
crsSFHA	0.028	0.003	10.20	0.000***
Significance code:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1			
Pseudo R-squared:	0.1975			
Standardized weighted residuals				
Min	1Q	Median	3Q	Max
-7.4318	-0.4888	0.0527	0.5697	8.6971
AIC	-1,017,707			
BIC	-1,017,582			

Lastly, note that despite the significance of elevation difference as rating variable in the two models above, it may not always be available, because, e.g., the insured may not have an elevation certificate or the BFE

---

[29] The implementation of Risk Rating 2.0 did not result in changes to these credits.

of the building may be unknown. It is important to be able to price the insurance without knowing the building's elevation difference. Therefore, we fit a model, Model III, with elevation difference excluded from the explanatory variables. We keep the interaction term between the CRS rating and the property's flood zone in the model. Model III is given by:

$$g(damageRatio)$$
$$= intercept + \beta_1 coverage + \beta_2 SFHA + \beta_3 basementType1 + \beta_4 basementType2$$
$$+ \beta_5 crs + \beta_6 floors2 + \beta_7 floors3 + \beta_8 postFIRM + \beta_9 age + \beta_{10} crsSFHA.$$

We present the estimation results in Table 7. As we can see, the coefficients of Model III estimations are all significant and consistent with our intuition.

Among the three models, we favor Model III according to both the AIC and BIC criteria. Nonetheless, we argue that Model II could be an alternative when accurate data on the property's elevation difference is available. As pointed out by Wing et al. (2020), inundation depth is a crucial variable in determining building damage. Everything else being equal, a larger elevation difference would mean a lower inundation depth and hence a smaller damage ratio. We will further explore the relevance of elevation difference as a pricing factor in Appendix C. As suggested by the regression tree and random forest in Appendix C, elevation difference is one of the most important variables for claim modeling.

In summary, our choice of model is Model II when elevation difference is known and Model III when it is not.

Section 6: Acknowledgments

The researchers' deepest gratitude goes to those without whose efforts this project could not have come to fruition: the Project Oversight Group and others for their diligent work overseeing questionnaire development, analyzing and discussing respondent answers, and reviewing and editing this report for accuracy and relevance.

Project Oversight Group members:

R. Dale Hall, FSA, MAAA, CERA, CFA

Steve Kolk, ACAS, MAAA

Andrea Marcovici, FSA, PhD

Jianxi Su, FSA, PhD

Kenneth Wolstrup

At the Society of Actuaries:

W. Scott Lennox, FSA, FCAS, FCIA, Staff Fellow – General Insurance

Rob Montgomery ASA, MAAA, FLMI, Consultant – Research Project Manager

Erika Schulty, Research Administrator

The researchers would like to also acknowledge the excellent research assistance provided by:

Farshid Rahmani

Weilun Zhang, ASA

Appendix A: A Review of EVT

In this appendix, we briefly review univariate EVT, the application of which is rooted in the Fisher–Tippett–Gnedenko theorem and the Pickands-Balkema-de Haan theorem.

By the Fisher–Tippett–Gnedenko theorem, under some mild conditions, the normalized maximum asymptotically follows a generalized extreme value (GEV) distribution. Precisely, let F be the distribution function of a risk variable X, and let M_n be the maximum of an independent and identically distributed (i.i.d.) sample of size n from F. We say that F belongs to the max-domain of attraction of a non-degenerate distribution G, written as $F \in MDA(G)$, if there are normalizing constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$\frac{M_n - d_n}{c_n} \to^d G, \qquad n \to \infty,$$

where $\to^d$ denotes convergence in distribution. In fact, the distribution G must be a member of the family of GEV distributions, with standard form

$$G_\xi(x) = \exp\left\{-(1 + \xi x)^{-\frac{1}{\xi}}\right\}, \qquad 1 + \xi x > 0,$$

where in the case of $\xi = 0$ the term $(1 + \xi x)^{-1/\xi}$ is interpreted as its limit when ξ approaches 0, that is e^{-x}. The GEV distribution above is a general representation for three types of extreme value distributions: Fréchet ($\xi > 0$), Gumbel ($\xi = 0$), and Weibull ($\xi < 0$). The distribution has a heavy tail when $\xi > 0$, an exponential tail when $\xi = 0$, and an upper-bounded tail when $\xi < 0$. Given the explicit parametric structure of the GEV distribution, likelihood-based approaches are usually used to estimate the parameters, although in the hydrology literature, the probability weighted moments (PWM) method, owing largely to its computational simplicity and small-sample performance, is also a popular choice.

In both actuarial literature and hydrology literature, it is customary to base statistical inference on the limiting distributions, i.e., the GEV distribution, as an approximation of the true underlying distribution. In some cases, one would need to estimate the normalizing constants for the purpose of, e.g., transforming the marginal distribution to a standard distribution, such as Fréchet distribution with shape parameter 1, as is usually done in multivariate extreme analysis. In other cases, the normalizing constants can be absorbed into the location and scale parameters of the distribution without need for special treatment.

In practice, given a sample of $n \times m$ observations, one could divide them into m blocks and assume the common distribution of the block maxima $M_{nj}, j = 1, \ldots, m$, is

$$G_{\xi;\mu,\sigma}(x) = \exp\left\{-\left[1 + \frac{\xi(x - \mu)}{\sigma}\right]^{-\frac{1}{\xi}}\right\}, \qquad 1 + \frac{\xi(x - \mu)}{\sigma} > 0,$$

which belongs to a three-parameter family with shape parameter $\xi \in \mathbb{R}$, location parameter $\mu \in \mathbb{R}$, and scale parameter $\sigma > 0$.

Under the GEV model, one can easily derive the return level. For $0 < p < 1$, its $1/p$-year return level is given by

$$G_{\xi;\mu,\sigma}^{-1}(1 - p) = \mu + \frac{\sigma}{\xi}\left((-\ln(1 - p))^{-\xi} - 1\right).$$

For example, if the annual maximum streamflow is modeled by the GEV above, then the 100-year streamflow is given by

$$G_{\xi;\mu,\sigma}^{-1}\left(1 - \frac{1}{100}\right) = \mu + \frac{\sigma}{\xi}\left(\left(-\ln\left(1 - \frac{1}{100}\right)\right)^{-\xi} - 1\right).$$

When the interest is in modeling, e.g., high water level or inundation level, researchers have employed the PDS/POT approach, which fits a generalized Pareto distribution (GPD) to the excess over a high threshold. The underlying theory justifying such an application—the Pickands-Balkema-de Haan theorem—states that, for a random variable X following distribution $F \in MDA(G_{\xi;\mu,\sigma})$ with upper endpoint x_F (finite or infinite), there exists a positive scale function $a(\cdot)$ such that,

$$\lim_{y\uparrow x_F} P\left(\frac{X - y}{a(y)} \leq x \middle| X > y\right) = 1 - (1 + \xi x)^{-\frac{1}{\xi}}, \quad x > 0, 1 + \xi x > 0.$$

This means that the scaled excess over a high threshold y converges weakly to the GPD, justifying the use of the exceedances in the data to estimate the tail of the risk. The asymptotic result allows us to approximate the conditional distribution of $X - y$ given $X > y$ by

$$H_{\xi,\beta}(x) = 1 - \left(1 + \xi\frac{x}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}, \quad x > 0, 1 + \xi\frac{x}{\tilde{\sigma}} > 0,$$

where $\tilde{\sigma} = \sigma + \xi(y - \mu)$.

Note the correspondence above between the GEV parameters and the GPD parameters. Such duality enables one to, for example, estimate the GEV parameters using a PDS. In fact, both the AMS and PDS methods can be unified using a point process approach. The theory underlying the point process approach concludes that if the risk variables $X_1, \dots, X_n$ are i.i.d. following common distribution $F \in MDA(G_{\xi;\mu,\sigma})$, such that, for constants $c_n > 0$ and $d_n \in \mathbb{R}$, the distributions of the normalized maxima $\left(\max_{1\leq i\leq n} X_i - d_n\right)/c_n$ converge to $G_{\xi;\mu,\sigma}$, then we have the weak convergence of the normalized point processes to a Poisson point process; that is,

$$N_n := \left(\frac{i}{n+1}, \frac{X_i - d_n}{c_n}\right)_{1\leq i\leq n} \Rightarrow N, \quad n \to \infty$$

on $(0,1) \times [u, \infty)$ for any u above the lower endpoint of $G_{\xi;\mu,\sigma}$, where $\Rightarrow$ denotes weak convergence, and N is a Poisson point process with mean measure Λ defined by

$$\Lambda(A) = (t_2 - t_1)\left(1 + \xi\frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}, \quad A = [t_2 - t_1] \times [x, x_G),$$

with x_G being the upper endpoint of $G_{\xi;\mu,\sigma}$. See Chapter 7 of Coles (2001) for more details. Such a characterization enables us to estimate the GEV parameters alternatively with likelihood functions derived from the limiting point process.

Appendix B. Background Information About Flood Modeling

B.1 BACKGROUND INFORMATION

This section serves as a concise introduction into the pieces of flood modeling to help the reader understand the background of flood inundation modeling and the relevant science employed in the study. It is not meant to be thorough, but as an introduction or refresher as needed.

B.1.1. HYDROLOGIC MODELING

Physical flood modeling's tasks include predicting the frequency of floods, their magnitudes, and the potential inundation extent. Few models can handle all of these tasks well, so specialization is required. Here we focus on predicting the long-term discharge of a river, or its future distribution of flows (from which we determine the risk of flooding under future climates).

To this end, hydrologic models operate on long time scales (a few years to decades) and typically predict daily or hourly discharge along different points in a river network. These models accept input information about meteorological climate forcings (rainfall, temperature, radiation, humidity, etc.) and attributes about the basins of interest (slope, land cover, soil texture, geology, etc.). They output discharge and potentially other diagnostic variables, e.g., soil moisture, water storage, groundwater, evapotranspiration (ET), to provide a full narrative of the simulation history.

Then, flood inundation models take discharge (either from direct observations or outputs from hydrologic models) and use topographic information to predict where the water might be given a certain discharge condition. Such flooding processes occur rapidly, so dynamical flood inundation models operate on minute time steps and often use a high spatial resolution. It should be noted that the resolution of the model does not guarantee its quality or accuracy.

Different types of hydrologic models have been developed in the past. Traditionally there have been conceptual hydrologic models (with simple, box-type descriptions of hydrology), statistical hydrologic models (which try to describe streamflow distributions using past data patterns), and physics-based models (which solve lab-verified partial differential equations). These different model types have come in and out of favor throughout different time periods and the debates have never fully. The general consensus is that when one needs detailed descriptions or narratives of the progression of events, physics-based models are typically more useful, while data-driven (statistical or machine learning) models tend to have higher accuracy. Despite the many of types of models, however, accurate prediction of floods, to this day, very much remains a challenging task.

### B.1.2. DEEP LEARNING HYDROLOGIC MODELS

Recently, there has been a surging interest in deep learning (DL) hydrologic models. DL is a rebranding of neural networks featuring large capacity (lots of neurons) and depth (with two or more hidden layers). The main reason for this surge of interest is that, quite refreshingly, DL models have vastly outperformed our traditional hydrologic models. For a discharge time series that we are interested in predicting, a typical metric of success is the Nash-Sutcliffe model efficiency coefficient (NSE)[30]. Recent benchmarks using 571 basins across the United States showed that traditional models typically produced a median NSE of around 0.63 even when they were calibrated to each

---

[30] NSE can be interpreted as roughly the variance explained by the model divided by the total variance of the variable. Its value can be negative but is 0 if equivalent to using the mean of a time series as the predictor for the whole time series and is 1 for a perfect prediction.

basin individually, whereas the new DL models can reach 0.74 for the same metric (Feng et al., 2020; Kratzert et al., 2019) (Figure 18). It should be noted, however, that even though the present DL models are much stronger than previous models, they are still not perfectly predicting discharge. There are many other sources of errors with forcing data (weather), physical attributes that describe the landscape (slope, soil texture, land cover, geology, etc.), and also model imperfections (unknown mechanisms). Nonetheless, this step change in our predictive capacity has revealed that, when properly constrained, DL models can serve as practical and strong tools for flooding prediction.

Figure 18

MAP OF 531 BASINS IN THE CAMELS DATASET. COLORS INDICATE THE NSE VALUES FROM A DL MODEL FROM FENG ET AL. (2020).
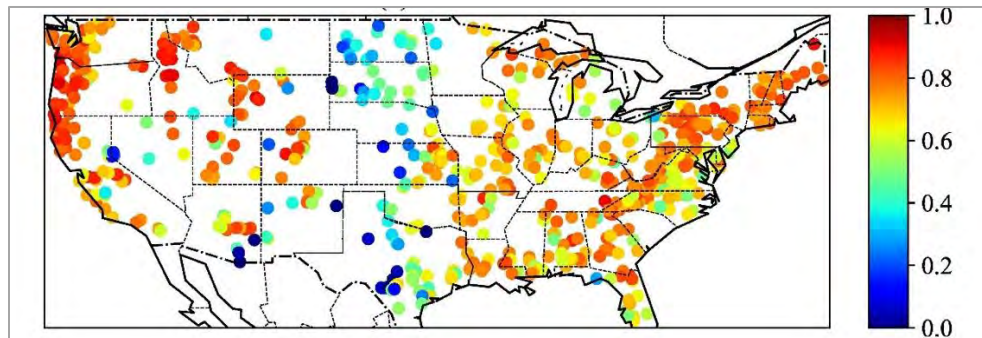
Data Source: Feng et al., 2020

B.1.3. PREDICTION IN UNGAUGED BASINS

Hydrologic models often perform better when they are "calibrated", meaning some parameters for these models have been tuned so that the model output matches the observations at a site. For DL models, we do not calibrate. Rather, we train the model on a large number of sites, as data-fitting is an inherent component of the model building process. Because not all the attributes of a basin (like geologic conditions) are known in detail, available inputs can approximately but not exactly describe a basin. To calibrate a model requires discharge observations, which implies that calibration can only occur at places where discharge data are available, which is only a very small subset of the locations where we need the predictions.

Calibrated and trained models can technically be applied to "ungauged" basins because the model simulation itself does not require discharge data, and a prediction can be made as long as we have the input data (meteorological forcings and basin attributes). Typically, this results in a deterioration of model performance, because each basin behaves differently. Even if there is a neighboring basin where the model is calibrated, there will be slight differences between the neighbors, which may not be captured in the available basin attributes. How to effectively extrapolate/interpolate the parameters one obtained from the calibrated basins to the ungauged ones is called the Prediction in Ungauged Basins (PUB) problem, which has long been a central research topic in hydrology. Because of the limited discharge data available, PUB performance is the true performance we can expect to generally have across the entire domain.

B.2. DEEP LEARNING STREAMFLOW MODELS FOR THE SUSQUEHANNA RIVER BASIN

The hydraulic modeling part of this project (Section 3) used observed streamflow. However, here we explain that streamflow can be generated from a model for ungauged locations. We focus on obtaining streamflow predictions in the Susquehanna River basin, for not only gauged basins but also ungauged basins, and for not only "reference"

basins (having minimal anthropogenic impacts) but also basins with major reservoirs[31]. This system would be generically applicable across the US and can produce daily streamflow predictions under known climate conditions, or future streamflow distributions under future climate scenarios.

Figure 19
MAP OF THE SUSQUEHANNA RIVER BASIN.
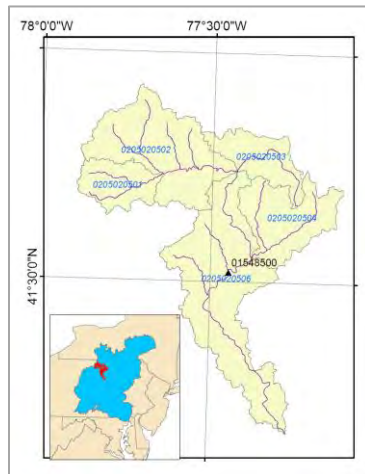
Data Source: Author's visualization using ArcMap.

B.2.1. THE LONG SHORT-TERM MEMORY (LSTM) NETWORK

LSTM is a type of recurrent neural network, which, in layman's words, can be explained as a self-trained memory system. It can learn patterns from sequential data and predict future sequences. LSTM was previously the underlying technology behind artificial intelligence tools such as Google Translate and Siri, until it was replaced by more advanced algorithms such as Transformers. These networks are composed of memory cells, the keys to which are the "cell states" and "gates" that control information flow within the LSTM algorithm. Cell states allow information to be stored over long time periods, which is important for modeling catchment processes like snow, subsurface flow, and reservoir storage.

In hydrology, LSTM has been shown to be highly effective for modeling a range of problems such as soil moisture, streamflow, and water quality (Fang et al. 2017; Feng, Fang, and Shen 2020; Rahmani et al. 2021; Zhi et al. 2021).

B.2.2. SETUP OF THE LSTM MODEL

We have collected input data for thousands of basins in the United States and trained one uniform LSTM network. This dataset contains attributes, forcings, reservoir attributes and streamflow data for 3557 basins from GAGES-II (Figure 20). Among these, reservoir attributes were an original contribution. Basins with complete streamflow records from 1 January 1990 through 31 December 2009 were selected from the Geospatial Attributes of Gages for Evaluating Streamflow II (GAGES-II) dataset (Falcone 2011).

---

[31] Reservoirs exert substantial control on streamflow and also tend to reduce model accuracy downstream. Acting as storages, they shave off the discharge peaks and can release water when flow is too low. Different types of reservoirs (for example, hydroelectric vs irrigation) have different logic behind their operations.

We selected 30 static physical attributes which fit into six categories: (1) basic identification and topographic characteristics, (2) percentages of land cover in the watershed, (3) soil characteristics, (4) geological characteristics, (5) local and cumulative dam variables, and (6) other disturbance variables. Basin mean forcing data for the period 01/01/1990–12/31/2009 was generated using the same method as for the CAMELS dataset, which was done by mapping a daily, gridded meteorological dataset, Daymet Version 3 (Thornton et al., 2016) to the chosen basin polygons.

We trained the LSTM model on the large dataset and fine-tuned it mildly for the subbasins in the Susquehanna River watershed. We carefully controlled the hyperparameters so that they are not overfitted. To reduce overfitting, we employed dropout regularization, which stochastically sets some network connections to zero. We trained the model on sequences of a fixed length (365 days) and neglected the warm-up period as warm-up did not show any advantages.

To obtain the reservoir attributes, dams listed in the National Inventory of Dams (NID) database (US Army Corps of Engineers 2018) were extracted to separate basins. For every basin, the sum of the reservoir's normal capacity associated with each dam purpose was calculated. The purpose with the largest associated capacity was considered to be the major purpose of the collective dams in the basin.

Figure 20
THE LOCATIONS OF THE BASINS IN OUR TRAINING DATASET. THE COLORS INDICATE THE TOTAL NORMAL RESERVOIR STORAGE VOLUME IN A BASIN, MEGALITERS OF TOTAL STORAGE PER SQ KM.
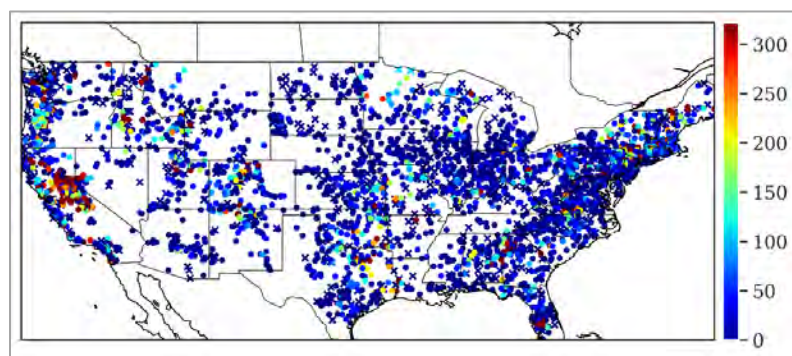
Data Source: Author's visualization based on GAGES-II dataset.

B.2.3. STREAMFLOW RESULTS

We show the NSE values from LSTM for the continental United States (Figure 21). To show the level of improvement over traditional models, we compare LSTM with an operational hydrologic model, Sacramento Soil Moisture Accounting Model (SAC-SMA), over a subset of the basins known as the CAMELS dataset with 531 mostly-reference basins (this is because SAC-SMA is only available for this subset. Here reference basins means they do not have major reservoirs). We show the time series comparisons for some of the basins in the Susquehanna River Basin. We also compared our results with some previous models with a reservoir component (Ouyang et al. 2021).

We notice that LSTM achieved much higher performance than SAC-SMA over the CAMELS dataset. In addition, it performed decently well for the larger dataset that contains mostly basins with major reservoirs (median NSE=0.74). There is a mild deterioration of performance from the CAMELS-only dataset (median NSE=0.75) to the >3000 basin dataset (Figure 20) due to the prevalence of reservoirs in the other basins not included in CAMELS. However, even if the case of major reservoirs, LSTM is stronger than SAC-SMA (median NSE=0.63) by a large margin. If SAC-SMA is

applied to the basins with major reservoir, one can expect large deterioration in performance. LSTM is highly competitive compared to the models in the literature.

This result means that LSTM well captured the pattern of reservoir management. Even though LSTM did not know anything about the management policies or priorities, it was not informed of the water demand in the basin, even though human operators sometimes do not follow their own handbooks, LSTM was able to predict their future behaviors based on what it has seen in the past. In layman terms, LSTM "understood" human operators as long as there are records of what they did in the past, and humans appear to be more predictable than we think. We further investigate the daily distribution of flows (Figure 22) comparing the observed value, LSTM and SAC-SMA.

Figure 21

CUMULATIVE DENSITY FUNCTION OF THE NSES OF THE MODELS. THE LEFT GRAPH SHOW THE LSTM PERFORMANCE OVER THE ENTIRE CONUS. RED LINE: THE LSTM TRAINED ON >3000 BASINS AND TESTED ON CAMELS; BLUE LINE: TRAINED AND TESTED ON 523 CAMELS BASINS; BLACK-DOTTED LINE: TRAINED AND TESTED ON >3000 BASINS (INCLUDING REFERENCE BASINS AND BASINS WITH MAJOR RESERVOIRS). THE RIGHT GRAPH COMPARES LSTM AND SAC-SMA OVER THE CAMELS DATASET. A CURVE TOWARD THE RIGHT-HAND SIDE IS BETTER.
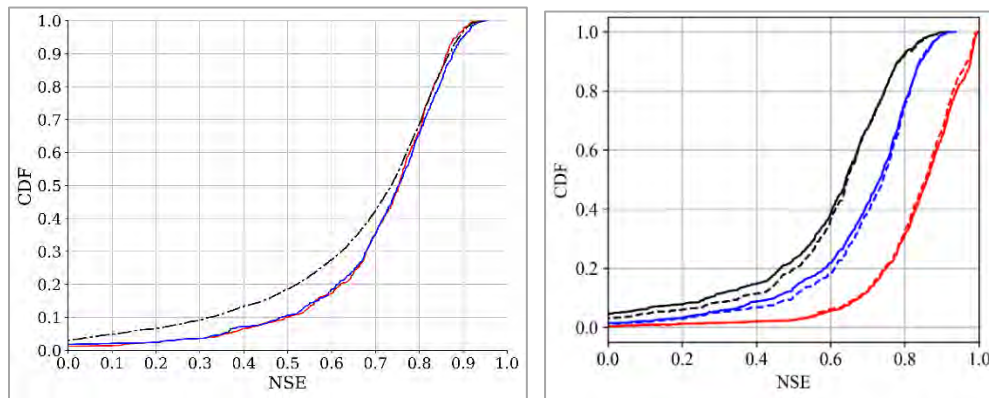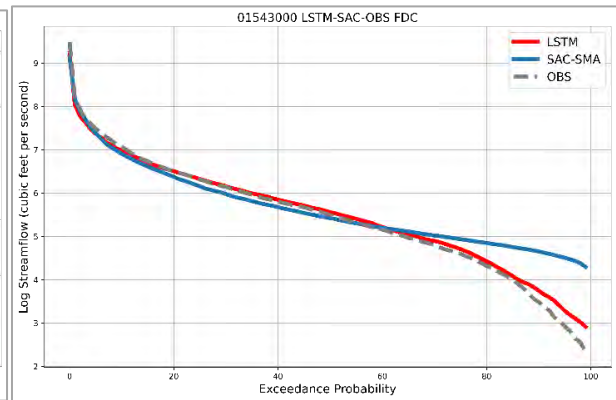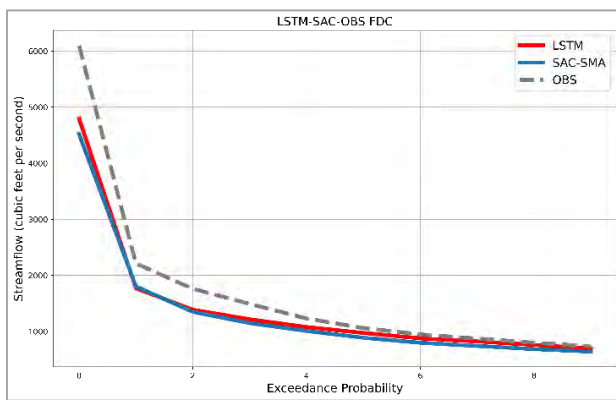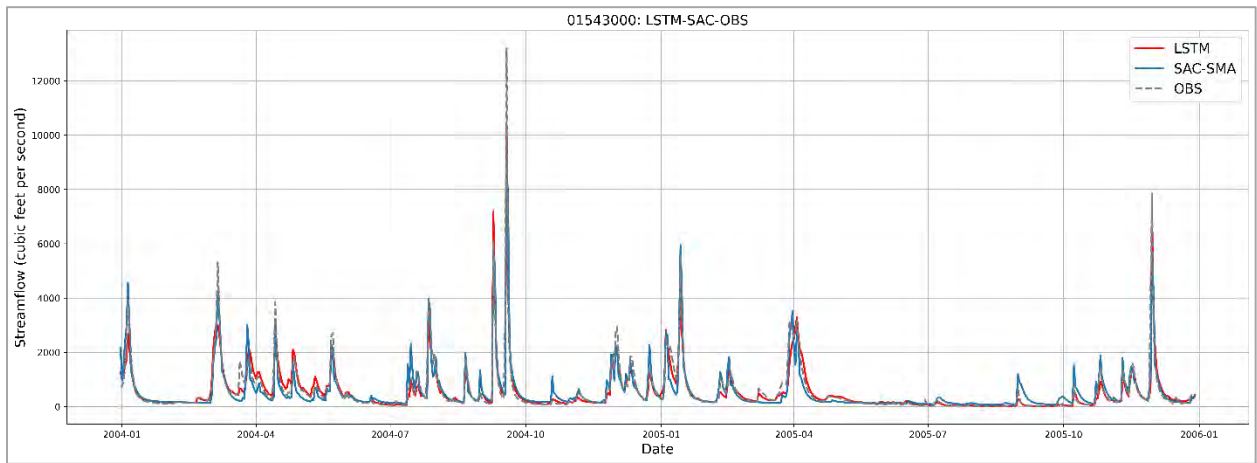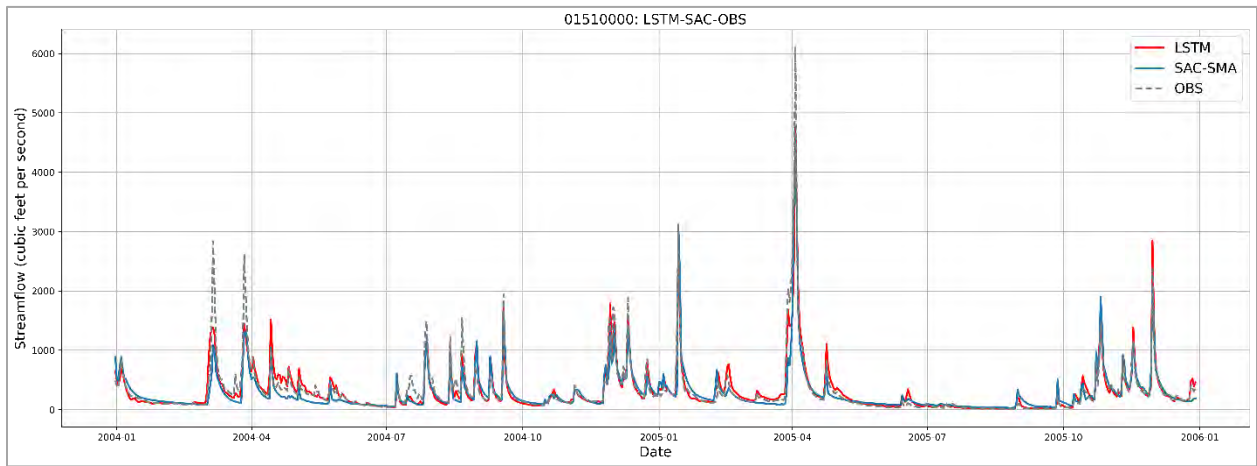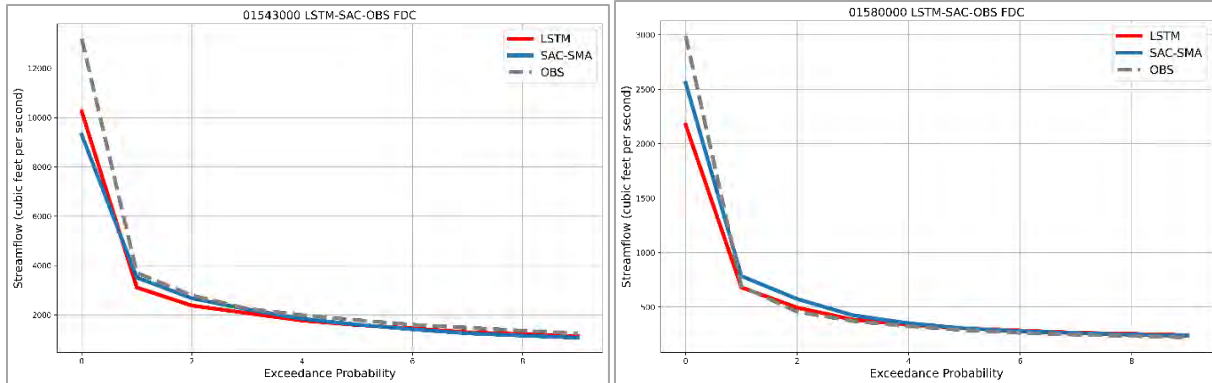

Data Source: Author's calculation

Figure 22

COMPARISONS OF SIMULATED VS. OBSERVED TIME SERIES FOR LSTM AND SAC-SMA. THE TOP PANEL SHOWS THE DAILY TIME SERIES. THE SECOND PANEL SHOWS THE DISTRIBUTION OF STREAMFLOW (FLOW DURATION CURVE). THE THIRD PANEL SHOWS THE EXCEEDANCE PROBABILITY CURVE AND DESCRIBES THE FREQUENCY (X-AXIS, IN

PERCENTAGE) OF THE TIME A GIVEN FLOW MAGNITUDE (Y-AXIS) IS EXCEEDED. THE BOTTOM PANEL SHOWS THE FLOW DURATION CURVE ZOOMED TO THE EXTREME PART.

Data Source: Author's calculation

B.2.4. IMPLICATIONS

We have trained a powerful deep learning neural network that outputs daily discharge value at unprecedentedly high (yet still not perfect) accuracy. We customized and fine-tuned it for the Susquehanna River Basin. The implications of the work for the insurance industry are trifold: (1) Improved predictions of streamflow with the help of DL will soon be available widely for large spatial scales; (2) Even given the state-of-the-art models, precisely predicting the magnitudes of the events is still going to be challenging given the available data quality (especially of precipitation). Predicting the distributions of the flow will be a little easier but still not perfect. This could serve as a starting point for thinking about nationwide strategies for flood insurance modeling.

There are also limitations. For gauged locations, the results suggest there still exists a noticeable gap between observed and simulated, even if it is now much narrower than the traditional hydrologic model (a site-by-site calibrated SAC-SMA here serves as the example). Hence, where long-term records exist, the observed data remains the best source of information for flood inundation modeling for gauged locations, and this is what we choose to demonstrate in Section 3. Deep learning models can fill the gap for locations away from gauges.

# Appendix C: More About Elevation Difference as a Rating Factor

In this appendix, we present some more evidence supporting the use of Model II derived in Section 5 when elevation difference is available.

C.1 GRAPHICAL INSPECTION

We inspect the distributions of claim amounts paid on building damage for properties with different levels of elevation difference. The graphs in Figure 23, from top to bottom, are the histograms of the claim amounts for elevation differences ranging from -2 to 1.5 feet, from 1.5 to 5 feet, 5 to 8.5 feet, and 8.5 to 12 feet.

Figure 23
HISTOGRAMS GROUPED BY ELEVATION DIFFERENCES OF THE PROPERTIES.
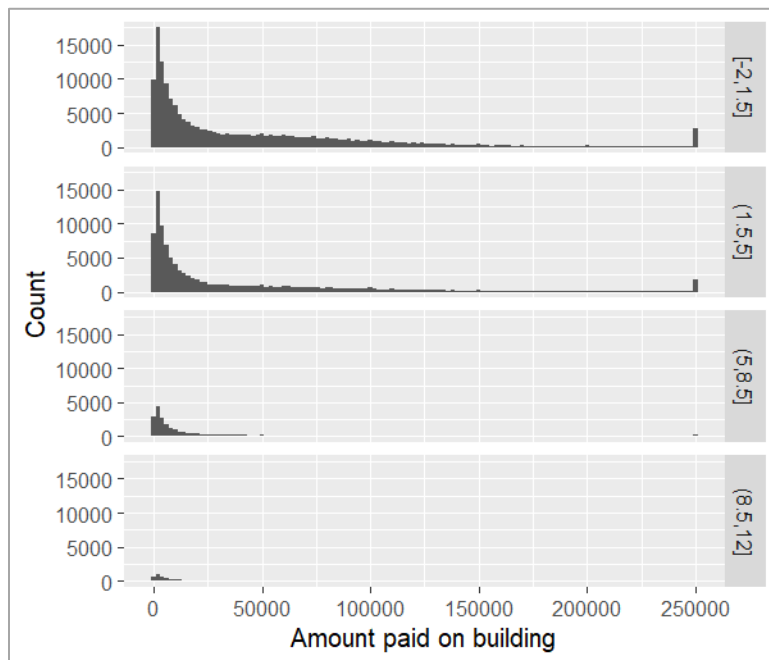
Data Source: Federal Emergency Management Agency (FEMA) and U.S. Bureau of Labor Statistics (BLS)

Figure 23 seems to suggest that, given that there is a flood damage to the building, those with larger elevation difference tend to experience smaller losses, with the claim distribution concentrated more on the left. This is consistent with the observations by Wing et al. (2020).

C.2 RESULTS FROM REGRESSION TREE AND RANDOM FOREST

As is observed in the mixed beta model, interactions of variables can play important roles in estimating the flood damage. To account for the interactions, we employ a regression tree as an alternative model to investigate the significance of variables such as elevation difference in determining the flood risk.

We first use a single tree to explore if the splits are logical. We use a tree with maximum depth of 15, complexity parameter of 0.0005 and a minimum bucket size of 5. After inspecting splits with elevation difference, we do observe that more elevated building experience less damage.

Next, we use random forest to model damage to get a more consistent result and to see how important elevation difference is compared to other features. The random forest method can only take complete data. Thus, we remove

two insignificant features from previous models and all the claims with missing data. We are remained with 30,910 claims for random forest modeling.

The forest has 200 trees, each built with 60% of the claims and each split uses 6 out of 8 of the variables. The trees stop splitting with a minimum number of 100 claims in each leaf. The variable importance plots below show that, based on importance measurements using both the increase in MSE and increase in node purity, elevation difference is among the factors with greatest impact on building damage.

Figure 24

VARIABLE IMPORTANCE PLOTS, WITH THE LEFT PLOT SHOWING THE PERCENTAGE INCREASE IN MSE AND THE RIGHT PLOT SHOWING THE INCREASE IN NODE PURITY.
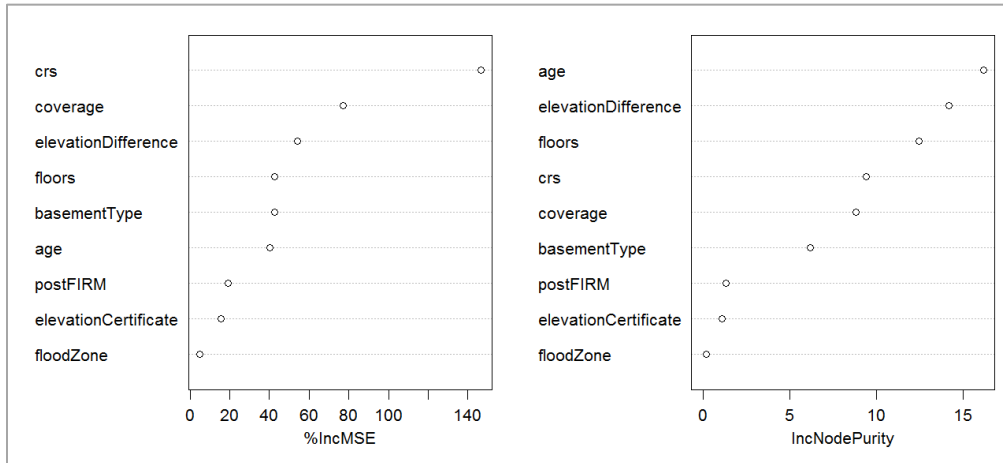
Data Source: Federal Emergency Management Agency (FEMA)

References

Agilan, V. and Umamahesh, N.V., 2015. Detection and attribution of non-stationarity in intensity and frequency of daily and 4-h extreme rainfall of Hyderabad, India. Journal of Hydrology, 530, pp.677-697.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L., 2004. Statistics of extremes: theory and applications (Vol. 558). John Wiley & Sons.

Boudreault, M., Grenier, P., Pigeon, M., Potvin, J.M. and Turcotte, R., 2020. Pricing Flood Insurance with a Hierarchical Physics-Based Model. North American Actuarial Journal, 24(2), pp.251-274.

Brunner, M.I., Swain, D.L., Wood, R.R., Willkofer, F., Done, J.M., Gilleland, E. and Ludwig, R., 2021. An extremeness threshold determines the regional response of floods to changes in rainfall extremes. Communications Earth & Environment, 2(1), pp.1-11.

Coles, S., Bawa, J., Trenner, L. and Dorazio, P., 2001. An introduction to statistical modeling of extreme values (Vol. 208, p. 208). London: Springer.

Curceac, S., Atkinson, P.M., Milne, A., Wu, L. and Harris, P., 2020. An evaluation of automated GPD threshold selection methods for hydrological extremes across different scales. Journal of Hydrology, 585, p.124845.

Czajkowski, J., Kunreuther, H. and Michel-Kerjan, E., 2012. A methodological approach for pricing flood insurance and evaluating loss reduction measures: Application to Texas. CoreLogic: Risk Management and Decision Processes Center, the Wharton School, University of Pennsylvania.

El Adlouni, S., Ouarda, T.B., Zhang, X., Roy, R. and Bobée, B., 2007. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resources Research, 43(3).

Embrechts, P., Klüppelberg, C. and Mikosch, T., 2013. Modelling extremal events: for insurance and finance (Vol. 33). Springer Science & Business Media.

Falcone, James A. 2011. "GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow." Report. Reston, VA. USGS Publications Warehouse. https://doi.org/10.3133/70046617.

Fang, Kuai, Chaopeng Shen, Daniel Kifer, and Xiao Yang. 2017. "Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental U.S. Using a Deep Learning Neural Network." *Geophysical Research Letters* 44 (21): 11,030-11,039. https://doi.org/10/gcr7mq.

FEMA 2017. "Answers to Frequently Asked Questions about Increased Cost of Compliance." Federal Emergency Management Agency. https://www.fema.gov/sites/default/files/2020-11/fema_p1080_icc_faq_20170817.pdf.

Feng, Dapeng, Kuai Fang, and Chaopeng Shen. 2020. "Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks with Data Integration at Continental Scales." *Water Resources Research* 56 (9): e2019WR026793. https://doi.org/10.1029/2019WR026793.

Ferrari, Silvia, and Francisco Cribari-Neto. 2004. "Beta Regression for Modelling Rates and Proportions." Journal of Applied Statistics 31 (7): 799–815.

Furman, E., Su, J., Chen, Y., Santoshkumar, S. and Zhang, L. 2019. Modeling, measuring, and pricing the flood risk: Survey and actuarial perspective. The Society of Actuaries, 1-29.

Gbari, S., Poulain, M., Dal, L. and Denuit, M., 2017. Extreme value analysis of mortality at the oldest ages: a case study based on individual ages at death. North American Actuarial Journal, 21(3), pp.397-416.

Giovannettone, J., Copenhaver, T., Burns, M. and Choquette, S., 2018. A statistical approach to mapping flood susceptibility in the Lower Connecticut River Valley Region. Water Resources Research, 54(10), pp.7603-7618.

Hirsch, Robert M., and Stacey A. Archfield. 2015. "Flood trends: Not higher but more often." Nature Climate Change 5 (3): 198–99. https://doi.org/10.1038/nclimate2551.

Horn, D. 2021. "National Flood Insurance Program: The Current Rating Structure and Risk Rating 2.0." Congressional Research Service. https://sgp.fas.org/crs/homesec/R45999.pdf.

Huang, F., Maller, R. and Ning, X., 2020. Modelling life tables with advanced ages: An extreme value theory approach. Insurance: Mathematics and Economics, 93, pp.95-115.

Katz, R.W., Parlange, M.B. and Naveau, P., 2002. Statistics of extremes in hydrology. Advances in water resources, 25(8-12), pp.1287-1304.

Kousky, Carolyn, and Brett Lingle. 2017. "The NFIP's Increased Cost of Compliance (ICC) Coverage." Wharton Center for Risk Management; Decision Processes.

Kousky, C., Lingle, B. and Shabman, L., 2017. The pricing of flood insurance. Journal of Extreme Events, 4(02), p.1750001.

Kratzert, Frederik, Daniel Klotz, Guy Shalev, Günter Klambauer, Sepp Hochreiter, and Grey Nearing. 2019. "Towards Learning Universal, Regional, and Local Hydrological Behaviors via Machine Learning Applied to Large-Sample Datasets." *Hydrology and Earth System Sciences* 23 (12): 5089–5110. https://doi.org/10.5194/hess-23-5089-2019.

Lee, O., Sim, I. and Kim, S., 2020. Application of the non-stationary peak-over-threshold methods for deriving rainfall extremes from temperature projections. Journal of Hydrology, 585, p.124318.

Liu, Xiaofeng. 2022. "PyHMT2D — PyHMT2D 1.0.0 Documentation." Python Hydraulic Modeling Tools - 2D. 2022. https://psu-efd.github.io/pyHMT2D_API_Web/.

Lombardo, F., Napolitano, F., Russo, F. and Koutsoyiannis, D., 2019. On the exact distribution of correlated extremes in hydrology. Water Resources Research, 55(12), pp.10405-10423.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I. and Huang, M., 2021. Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change, p.2.

McNeil, A.J., 1997. Estimating the tails of loss severity distributions using extreme value theory. ASTIN Bulletin: The Journal of the IAA, 27(1), pp.117-137.

Ouyang, Wenyu, Kathryn Lawson, Dapeng Feng, Lei Ye, Chi Zhang, and Chaopeng Shen. 2021. "Continental-Scale Streamflow Modeling of Basins with Reservoirs: Towards a Coherent Deep-Learning-Based Strategy." *Journal of Hydrology* 599 (August): 126455. https://doi.org/10.1016/j.jhydrol.2021.126455.

Rahmani, Farshid, Kathryn Lawson, Wenyu Ouyang, Alison Appling, Samantha Oliver, and Chaopeng Shen. 2021. "Exploring the Exceptional Performance of a Deep Learning Stream Temperature Model and the Value of Streamflow Data." *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/abd501.

Roland, Mark, and Marla Stuckey. 2008. "Regression Equations for Estimating Flood Flows at Selected Recurrence Intervals for Ungaged Streams in Pennsylvania." 2008–5102. U.S. Geological Survey. https://pubs.usgs.gov/sir/2008/5102/.

Sippel, S., Mitchell, D., Black, M.T., Dittus, A.J., Harrington, L., Schaller, N. and Otto, F.E., 2015. Combining large model ensembles with extreme value statistics to improve attribution statements of rare events. Weather and climate extremes, 9, pp.25-35.

Smith, J.A., 1987. Estimating the upper tail of flood frequency distributions. Water Resources Research, 23(8), pp.1657-1666.

Stocker, T. F., D. Qin, G.-K. Plattner, L. V. Alexander, S. K. Allen, N. L. Bindoff, F.-M. Bréon, et al. 2013. "Technical summary. Climate Change 2013: The Physical Science Basis." Edited by Stocker T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley. Cambridge University Press, Cambridge, United Kingdom; New York, NY, USA. http://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5{_}TS{_}FINAL.pdf.

Tabari, H., 2021. Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation. Journal of Hydrology, 593, p.125932.

Tang, Q. and Yuan, Z., 2019. CAT bond pricing under a product probability measure with POT risk characterization. ASTIN Bulletin: The Journal of the IAA, 49(2), pp.457-490.

Thornton, P.E., M.M. Thornton, B.W. Mayer, Y. Wei, R. Devarakonda, R.S. Vose, and R.B. Cook. 2016. *Daymet: Daily Surface Weather Data on a 1-Km Grid for North America, Version 3*. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1328.

US Army Corps of Engineers. 2018. "National Inventory of Dams (NID)." https://nid.sec.usace.army.mil/.

U.S. Geological Survey. 2022. "TNM Download V2." Accessed May 1, 2022. https://apps.nationalmap.gov/downloader/#/.

USGS. 2016. "USGS Surface-Water Data for the Nation." United States Geological Survey. 2016. https://doi.org/10.5066/F7P55KJN.

Villarini, G., Serinaldi, F., Smith, J.A. and Krajewski, W.F., 2009. On the stationarity of annual flood peaks in the continental United States during the 20th century. Water Resources Research, 45(8).

Villarini, G. and Smith, J.A., 2010. Flood peak distributions for the eastern United States. Water Resources Research, 46(6).

Wang, Q.J., 1991. The POT model described by the generalized Pareto distribution with Poisson arrival rate. Journal of Hydrology, 129(1-4), pp.263-280.
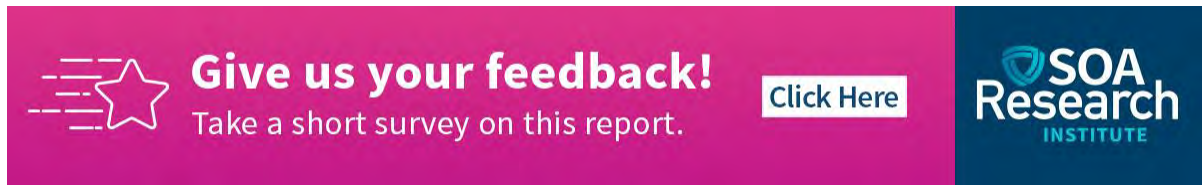
Watts, K.A., Dupuis, D.J. and Jones, B.L., 2006. An extreme value analysis of advanced age mortality data. North American Actuarial Journal, 10(4), pp.162-178.

Wing, Oliver EJ, Nicholas Pinter, Paul D Bates, and Carolyn Kousky. 2020. "New Insights into US Flood Vulnerability Revealed from Flood Insurance Big Data." Nature Communications 11 (1): 1–10.

Zhi, Wei, Dapeng Feng, Wen-Ping Tsai, Gary Sterle, Adrian Harpold, Chaopeng Shen, and Li Li. 2021. "From Hydrometeorology to River Water Quality: Can a Deep Learning Model Predict Dissolved Oxygen at the Continental Scale?" *Environmental Science & Technology* 55 (4): 2357–68. https://doi.org/10.1021/acs.est.0c06783.

Zimbidis, A.A., Frangos, N.E. and Pantelous, A.A., 2007. Modeling earthquake risk via extreme value theory and pricing the respective catastrophe bonds. ASTIN Bulletin: The Journal of the IAA, 37(1), pp.163-183.

Feedback

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org