A Data-driven Spatial Approach to Characterize Flood Hazard

Rubayet Bin Mostafiz[1], Adilur Rahim[1], Carol Freidland[1], and Robert Rohli[1]

[1]Affiliation not available

January 16, 2023

Abstract

United States Federal Emergency Management Agency provides model-output localized flood grids that are useful in characterizing flood hazards for properties located in the Special Flood Hazard Area (SFHA - areas expected to experience a 1% or greater annual chance of flooding). But these flood grids are often unavailable or fail to include return periods for particular applications, such as understanding flood risk of properties during the 70-year useful building life cycle. Furthermore, due to the unavailability of higher-return-period flood grids, the flood risk of properties located outside the SFHA cannot be quantified. Here, we present a method to estimate the flood hazard for U.S. properties that are located both inside and outside the SFHA. The flood hazard is characterized by the Gumbel extreme value distribution to project flood elevations to extreme flood events for which an entire area is assumed to be submerged. Spatial interpolation techniques impute elevation values in the extreme flood elevation surfaces and therefore can estimate the flood hazard for areas outside the SFHA. The proposed method can improve the assessment of flood risk for properties located in both inside and outside the SFHA and therefore, the decision-making process regarding flood insurance purchases, mitigation strategies, and long-term planning for enhanced resilience to one of the world's most ubiquitous natural hazards.

A Data-driven Spatial Approach to Characterize Flood Hazard

Rubayet Bin Mostafiz[1,2*], Md Adilur Rahim[1,2], Carol J. Friedland[1,2], Robert V. Rohli[2]

[1]Louisiana State University Agricultural Center, [2]Louisiana State University

Introduction

The overarching goal of this research is to characterize flood hazards at locations both inside and outside the Special Flood Hazard Area. More specifically, the research addresses the question, "If no modeled flood data exist for some or all return periods, what are the flood characteristics?" To that end, this research introduces a method for describing flood hazards whereby the flood is characterized using the Gumbel extreme value distribution, and flood elevations are projected at higher return periods. The gaps in flood surfaces due to limited data are filled by spatial interpolation techniques. These filled elevation values are then used to estimate floods for the locations inside the shaded or unshaded X Zones.

The contribution of this research is the development of a novel method to estimate flood hazard characteristics based on existing hydrologic-modeled flood surfaces. Ultimately, this technique will help government agencies and community officials to formulate policies and homeowners to make more informed decisions regarding insurance purchase, mitigation strategy, and long-term planning.

Methods

The method consists of extrapolating flood depths using the Gumbel extreme value distribution at the locations where a Gumbel fit is possible because flood depths for at least two return periods are known. Extreme return periods are selected where most of the study area is assumed to be submerged (Figure 1). Then, spatial interpolation techniques, including moving average, inverse distance weighting, natural neighbor, and kriging, are used to estimate the flood elevation for the extreme return periods at grid cells for which no data-derived distribution can be fit confidently. It is necessary to use flood elevation rather than flood depth for spatial interpolation because flood depth cannot be smoothed across space, while flood elevation is generally insensitive to differences in surface elevation. The imputed extreme-return-period flood elevations are then fit with the Gumbel distribution and used to estimate flood depth for locations that are unflooded at shorter return periods to verify that negative values, confirming that the surface is not flooded at that return period) are returned. Through this method, the flood depth vs. annual non-exceedance probability relationships are established for all locations in the study area, which can then be used to develop flood hazard estimates that are more reasonable to expect within the useful life of the building or settlement. The overall schematic summary of the flood hazard characterization method is shown in Figure 2.

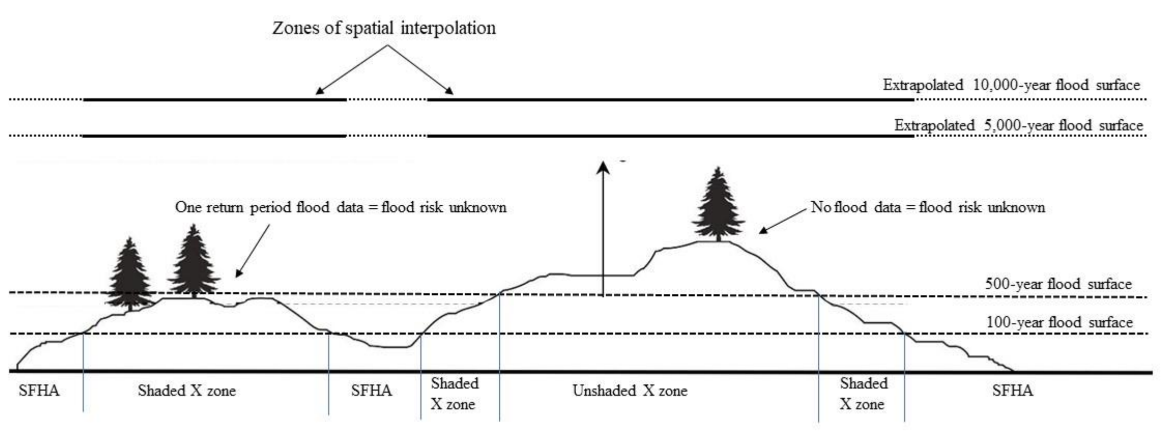

Figure 1: Schematic representation of the concept behind the flood depth surface estimating method.

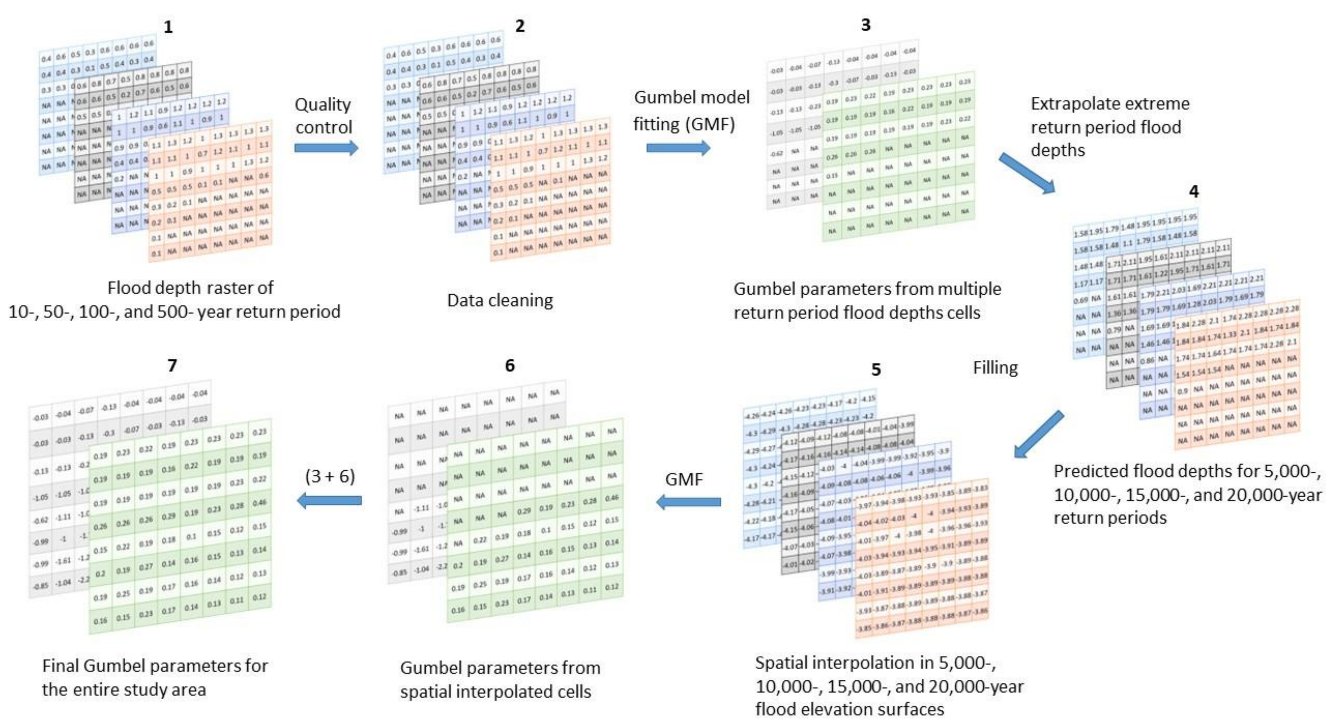

Figure 2: Schematic summary of the flood hazard characterization method.

Study Area and Data

A frequently-flooded residential neighborhood in Metairie, Louisiana (Jefferson Parish), bounded by the area shown in Figure 3, is used for this case study. This site is chosen primarily because of the availability of model-output flood depth grids for four return periods – 10, 50, 100, and 500 years – developed at a scale of 3.048 m x 3.048 m, by FEMA through its Risk Mapping, Assessment and Planning (Risk MAP) program. Although recent research has noted issues with FEMA methodologies and has enhanced flood characterization, these data are considered here due to the wide availability in the United States. The grid cells located within SFHA have at least two flood depth values (i.e., 100– and 500-year return periods) for which the Gumbel distribution can be fit initially. For the grid cells located in shaded-X zone (i.e., only 500-year flood depth is available) or unshaded-X zone (i.e., no flood information available), spatial interpolation is conducted to characterize flood in these grids. The study area consists of 44 census blocks with a total area of approximately 1.126 km2. The mean elevation in this below-sea-level, levee-protected area is –5.5 feet with a standard deviation of 0.71 and a range of –9.0 to –2.9 feet.

Figure 3: Study area in Metairie, Louisiana.

Results

Data Cleaning

The data cleaning process is run on the 121,215 cells in the study area. Data cleaning identifies 32 cells with flood depth (*D*) equal to zero (no cells have negative *D*), 3,575 cells for which a shorter return period *D* equals or exceeds a longer return period *D*, and 2,365 cells for which a positive shorter return period *D* is accompanied by a "null" longer return period *D* (Table 1). The original *D* values in these 5,972 cells (4.9% of the initial cells) are thus unused in the analysis because they fail one or more of these data cleaning tests.

Table 1: Number of cells in the study area removed by each data cleaning criterion.

Data Cleaning Rule	Number of Cells
10-year flood depth ≤ 0	13
50-year flood depth ≤ 0	16
100-year flood depth ≤ 0	1
500-year flood depth ≤ 0	2
10-year flood depth ≥ 50-year flood depth	776
10-year flood depth ≥ 100-year flood depth	0
10-year flood depth ≥ 500-year flood depth	2
50-year flood depth ≥ 100-year flood depth	530
50-year flood depth ≥ 500-year flood depth	4
100-year flood depth ≥ 500-year flood depth	2,263
10-year flood depth ≥ 0 and 50-year flood depth is NULL	7
10-year flood depth ≥ 0 and 100-year flood depth is NULL	0
10-year flood depth ≥ 0 and 500-year flood depth is NULL	0
50-year flood depth ≥ 0 and 100-year flood depth is NULL	4
50-year flood depth ≥ 0 and 500-year flood depth is NULL	1
100-year flood depth ≥ 0 and 500-year flood depth is NULL	2,353
Total	5,972

Table 2: Descriptive statistics of α and u for the location (cells) flooded by more than one return period.

Gumbel Parameter	Mean	Standard Deviation	Minimum	Maximum
α	0.24	0.08	0.08	0.82
u	–0.33	0.37	–3.16	0.00

Table 3: Descriptive statistics for α and u, after implementing a 31x31 moving average and a 3x3 smoothing.

Gumbel Parameter	Mean	Standard Deviation	Minimum	Maximum
α	0.28	0.22	0.07	2.08
u	–1.72	1.41	–12.96	–0.39

Gumbel Model Fitting

Descriptive statistics for the scale (α) and location (u) parameters are shown in Table 2. Once the α and u parameters are corrected for all cells, they are used to extrapolate *D* for the 5,000-, 10,000-, 15,000-, and 20,000-year return periods in their respective cells. The smallest possible moving-average window that interpolates all flood elevation values at extreme return periods is 31x31 cells. Descriptive statistics for the spatially interpolated and smoothed Gumbel parameters are shown in Table 3. A negative value is found for u in every cell. The Risk MAP-modeled 500-year *D* spuriously exceeds the spatially interpolated 5,000-year depth in 36 cells (0.03% of the study area), so correction procedures are implemented.

Validation

Table 4 shows the descriptive statistics and root-mean-square error (RMSE) of the difference between estimated and Risk MAP-modeled data for cells having at least two non-null *D* values. These results verify that a relatively small amount of error is introduced in the estimation procedure, if it can be assumed that the Risk MAP data are "correct." For cells having only a 500-year Risk MAP-modeled *D*, the relative correspondence between the spatially interpolated estimated 500-year *D* and that from Risk MAP is calculated by spatial interpolation technique. Because of the strong correspondence across spatial interpolation methods, values are expressed in inches (Table 5). Results suggest that the selection of spatial interpolation technique has little impact on the results.

Table 4: Descriptive statistics and root-mean-square error for Risk MAP-modeled minus predicted *D*, for cells having two or more originally-modeled *D* from among 10-, 50-, 100-, and 500-year return periods.

	Mean (ft.)	Standard Deviation (ft.)	Minimum (ft.)	Maximum (ft.)	RMSE (ft.)
10-year	0.17	0.21	–0.25	1.58	0.27
50-year	–0.01	0.09	–0.33	0.53	0.09
100-year	0.13	0.07	–0.00	0.85	0.15
500-year	–0.10	0.11	–0.95	0.57	0.14

Table 5: Descriptive statistics and root-mean-square error for Risk MAP-modeled minus predicted 500-year *D*, for cells having only 500-year return period flood depth, by moving average (31x31) and smoothing (3x3), inverse distance weighting, natural neighbor, and ordinary kriging.

Interpolation Technique	Mean (in.)	Standard Deviation (in.)	Minimum (in.)	Maximum (in.)	RMSE (in.)
Moving Average and Smoothing	–1.14	1.30	–11.43	6.90	1.73
Inverse Distance Weighting	–1.12	1.32	–11.43	6.92	1.73
Natural Neighbor	–1.11	1.33	–11.43	6.92	1.73
Ordinary Kriging	–1.12	1.32	–11.43	6.93	1.73

Sensitivity Analysis

The sensitivity analysis quantifies the rationality of using Gumbel extreme value distribution even as the number of known points decreases to two (Table 6). Results suggest that, not surprisingly, the increased magnitudes of the 500-year D leave a wider range from which the estimate can deviate from the actual *D*. Also, it is not surprising that the largest standard deviation of this modeled-vs.-estimated difference occurs for predicting the 500-year *D* when *D* is known at only two return periods. Nevertheless, even in such cases, the RMSE falls within a half-foot.

Table 6: Descriptive statistics and root-mean-square error of the difference (Δ) between the Gumbel model-based flood depth (*D*) estimation and Risk MAP-modeled *D*, when using *D* at known return periods to predict *D* at another known return period.

Scenario	Mean (ft.)	Standard Deviation (ft.)	Minimum (ft.)	Maximum (ft.)	RMSE (ft.)
Δ 500-year depth using 10-, 50-, and 100-year depth as predictors	0.32	0.22	–0.26	1.87	0.39
Δ 100-year depth using 10- and 50-year depth as predictors	–0.02	0.20	–0.46	1.09	0.20
Δ 500-year depth using 10- and 50-year depth as predictors	0.28	0.38	–0.46	2.65	0.47

Discussion

This method offers a means for circumventing the ever-present dilemma of how to ensure high-quality modeling to support planning for preventing, mitigating, and/or adapting to future flood events when little measured data are available, for locations where advanced hydrological and hydraulic modeling has been conducted to determine estimate *D* at multiple return periods. In the case study area in Metairie, Louisiana, only approximately 5 percent of the cells failed the "data cleaning" tests, which suggests that the modeled data are reasonable. Nearly all of the spurious data occurred when shorter return period *D* exceeds longer return period *D* or longer return period *D* is null.

If it can be assumed that the Risk MAP-modeled data are the "correct" values, the Gumbel distribution-generated flood parameters are shown to be remarkably stable for simulating and imputing *D* for various return periods. The fact that *u* remains negative in all cases verifies that the correction algorithm succeeded in ensuring that all terrestrial cells are not submerged under normal conditions. The much smaller standard deviation for α than for u is likely an artifact of the small, homogeneously-elevated study area. As *α* represents the slope of the Gumbel fit line, each cell in the study area will have a similar relationship between *D* and *P*. This contrasts with *u*, which can have a wider range of values, suggesting that some cells are more susceptible to flooding than others, even within the same neighborhood.

Validation and sensitivity analysis confirm that the method is relatively insensitive to the spatial interpolation technique chosen, at least for this study area. The relatively small errors, as evidenced by the small RMSE values (see Table 4), even for 500-year *D* and even when *D* values for only two return periods are known, are interpreted as evidence that the procedure is successful. The Gumbel distribution is deemed to provide an acceptable result. However, the present work does not consider the uncertainty in the Gumbel parameters. Moreover, the relatively small RMSE values, even between estimated vs. modeled 500-year *D* and even when *D* values for only two return periods are known, imply that *D* can be estimated relatively accurately and precisely. Such estimates can provide engineers and planners with useful information for enhancing infrastructure to accommodate low-frequency, large-magnitude flood events. Although the method is computationally intensive, it can be automated for improved *D* estimates for any location that is "data rich" regarding *D* grids at multiple return periods. Refinements in the modeled data for short or long return periods may allow for further improved understanding of infrastructure needs for accommodating floodwaters. .

Summary and Conclusion

The specific findings of the case study include that:

- the presented method is able to characterize flood hazards in areas of low to moderate flood risk; for example, 100-year *D* were predicted for cells with known 100-year *D* with RMSE of 0.15 feet
- spatial interpolation of extrapolated surfaces functioned well, regardless of technique; for example, 500-year *D* were imputed using spatial interpolation for cells with known 500-year *D* with RMSE of 1.73 inches
- using 10-, 50-, and 100-year *D* as predictors, the estimated 500-year *D* had an RMSE of 0.39 feet while the estimated 100- and 500-year *D* had an RMSE of 0.20 and 0.47 feet, respectively, when using 10- and 50-year *D* as predictors

Limitation and Future Work

Flood hazard estimation is, by necessity, based on such a limited number of data points, but the availability of model output at only a small number of locations and return periods necessitates use of this technique. Moreover, the rounding of original FEMA-modeled values to the tenth of a foot restricts the precision with which the results can be presented. This method was applied to a relatively limited geographical extent with homogeneous topography. Future work should evaluate the performance of the method across a larger geographical extent with more heterogeneous topography. In addition, the effect of climate change on flood hydroclimatology is not considered. Changing climate may alter the log-linear shape of the Gumbel distribution, particularly if forecasts of increasing frequency of extreme precipitation events prove to be accurate. Likewise, differences in local land cover may cause differences in the Gumbel parameters for *D* as a function of return period and in generating a continuous surface using the spatial interpolation techniques. Despite the fact that caution should be exercised in the interpretation of results for these and other reasons, the approach offers an advantageous "next step" in planning for, forecasting, and mitigating the world's most destructive natural hazard.

Funding

This research was funded by the USDA National Institute of Food and Agriculture, Hatch project LAB 94873, accession number 7008346, U.S. Department of Homeland Security (Award Number: 2015-ST-061-ND0001-01), the Louisiana Sea Grant College Program (Omnibus cycle 2020–2022; Award Number: NA18OAR4170098; Project Number: R/CH-03; Omnibus cycle 2022–2024; Award Number: NA22OAR4710105; Project Number: R/CH-05), the Gulf Research Program of the National Academies of Sciences, Engineering, and Medicine under the Grant Agreement number: 200010880, "The New First Line of Defense: Building Community Resilience through Residential Risk Disclosure," and the U.S. Department of Housing and Urban Development (HUD; 2019–2022; Award No. H21679CA, Subaward No. S01227-1). Any opinions, findings, conclusions, and recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the official policy or position of the funders.

The National Academies of SCIENCES ENGINEERING MEDICINE
GULF RESEARCH PROGRAM

United States Department of Agriculture
National Institute of Food and Agriculture