A Data-driven, Probabilistic, Multiple Return Period Method of Flood Depth Estimation

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

Louisiana State University

Introduction

Analysis of extreme flood probabilities is of great importance in planning for development that is expected to serve a long period of usefulness, such as residential and commercial construction, roads, and bridges. However, few if any historical records of flood depths exist. Thus, reliance on hydrologic modeling of flood depths as a function of return period is necessary. But in such models, a flood depth at smaller return periods will be "null" if floods are not expected at that frequency. Likewise, errors may compound for flood depth estimates at successively longer return periods. Therefore, accurate stochastic statistical methods are important enhancements to the hydrologic-modeled data for projecting flood depths in order to provide construction specialists, architects, developers, and urban and regional planners with adequate information to build more resilient facilities and communities. The main objective is to characterize floods for the study area (i.e., flood depth, return periods). To that end, the purpose of this research is to introduce an improved method for developing the statistical parameters necessary to enhance the accuracy of estimates of flood elevation (i.e., flood depth + site elevation), particularly at long return periods. More specifically, the research addresses the question, "If no modeled flood data exist for some or all return periods, how can the return period be estimated as a function of flood elevation?"

This research is motivated by a need for higher-resolution return period flood depths in order to make more accurate estimation of annual average loss due to flood. More specifically, for locations in which flood depths can be made confidently for at least three return periods (such as 10-, 100-, and 500-year return periods), a Gumbel log-linear distribution is fit through the available data. Then, the associated regression parameters (slope (α) and y-intercept (u)) from this fitted distribution of flood depth vs. return period are used to calculate the Gumbel log-linear distribution for extreme return periods beyond which would be expected, such 50,000-year return period, at points for which no data-derived distribution can be made confidently. At such points, this contrived flood depth can then be used to develop extreme return periods of flood depth that are more reasonable to expect within the useful life of the building or settlement, such as the 100- and 500-year return periods.

Methods

Approach

Using flood depth data on the known surfaces, we can fit the Gumbel parameters and derive the relationship (Figure 1), which is used to extrapolate higher return period flood surfaces (e.g., 1000-year, 5000-year), which can be spatially interpolated across the hill to fill the zone with no flood data.

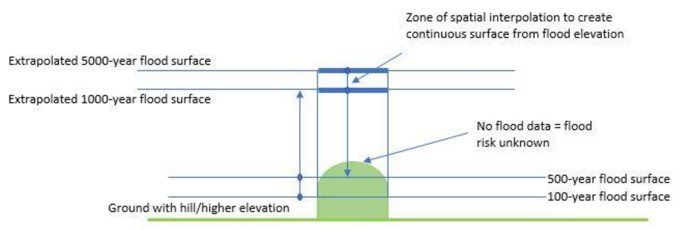

Figure 1: Flood risk estimation

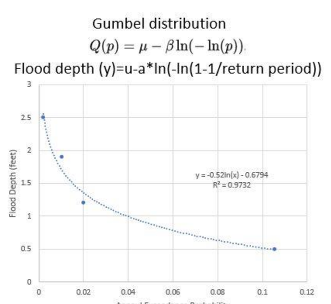

Figure 2: Flood depth-return period relationship

Data Cleaning

Each pixel will have non-null flood depths for either zero (i.e., flood return period exceeds 500 years), one (i.e., 500-year), two (i.e., 100- and 500-year), three (i.e., 50-, 100-, and 500-year), or four (i.e., 10-, 50-, 100-, and 500-year) return periods. Initial quality checks were performed to flag pixels with unrealistic data. These include three types of spurious values: 1) any pixel with a reported flood depth less than or equal to zero for any return period; 2) any pixel in which a flood depth for a shorter return period equals or exceeds that for any longer return period; and 3) any pixel in which a shorter-duration return period has a reported flood depth but a longer-duration return period has a null (i.e., flood free) value.

Gumbel Model Fitting

For each cell having non-null flood depth for at least two return periods, all available flood depth return periods were used to fit the Gumbel distribution (Figure 2). Check the 10-year flood depth with gumble parameter for those location where we fit the gumble with three cell (i.e. 10-year flood depth null). Separate the 10-year flooded cell which are originally non-flooded in the model data and put them -0.1 flood depth as barrier to fit better gumble distribution. We decrease the 10-year flood depth in the original model data with an increment of 0.1 until the gumble distribution parameter predict the 10-year flood depth as negative or non-flooded. Similarly, check the 50-year flood depth with gumble distribution parameter for those location where we fit the gumble with two cell (10- and 50-year flood depth is null) and apply the similar correction process.

Flood Depth Extrapolation

At each cell, the unique slope and intercept values of the trend line were then used to extrapolate the flood depth at that cell for floods of small probabilities (i.e., long return periods, such as 5,000-year, 10,000-year, and 50,000-year), over which the entire study area is assumed to have flooded.

Flood Elevation Surface Creation

A moving average filter was used to impute all missing cells in the study area, by experimenting with different window sizes (e.g., 35x35, and 60x60). The dimensions of the final window were determined as the smallest window that can impute all missing cells in the study area for that return period. Then, because the flood elevation surface of a completely flooded surface should be relatively smooth, a 3x3 window was run to smooth the flood elevation surface (i.e., reduce the undulations of waves over the flooded terrain).

Gumbel Model Fitting: Flood parameter

Then, for each cell, another Gumbel distribution was fit using the spatially interpolated rasters at only the longer (i.e., 5,000-, 10,000-, 50,000-, and 100,000-year) return periods. It should be noted that the Gumbel distribution is most appropriate in this research because of its established success in approximating the most extreme values. Next, flood depth was extrapolated for 10-, 50-, 100-, and 500-year return periods using the new slope and intercept of the fitted distribution. The flood depth for flood-free pixels were then estimated from these extrapolated data.

Case Study

Study Area and Data

A frequently-flooded residential neighborhood in Metairie, Louisiana (Jefferson Parish) is used for this case study. 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. Descriptive statistics of the Risk MAP-output flood depths by return period are shown in Table 1. The spurious maximum value for the 100-year return period, which is equal to that of the 500-year return period, suggests that data cleanup is necessary. 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 (FEMA 2021). This study area was selected also because its low relief necessitates only relatively short return periods (e.g., 5,000 years) for modeling of flood covering the entire study area, which introduces less error than extrapolating longer return periods

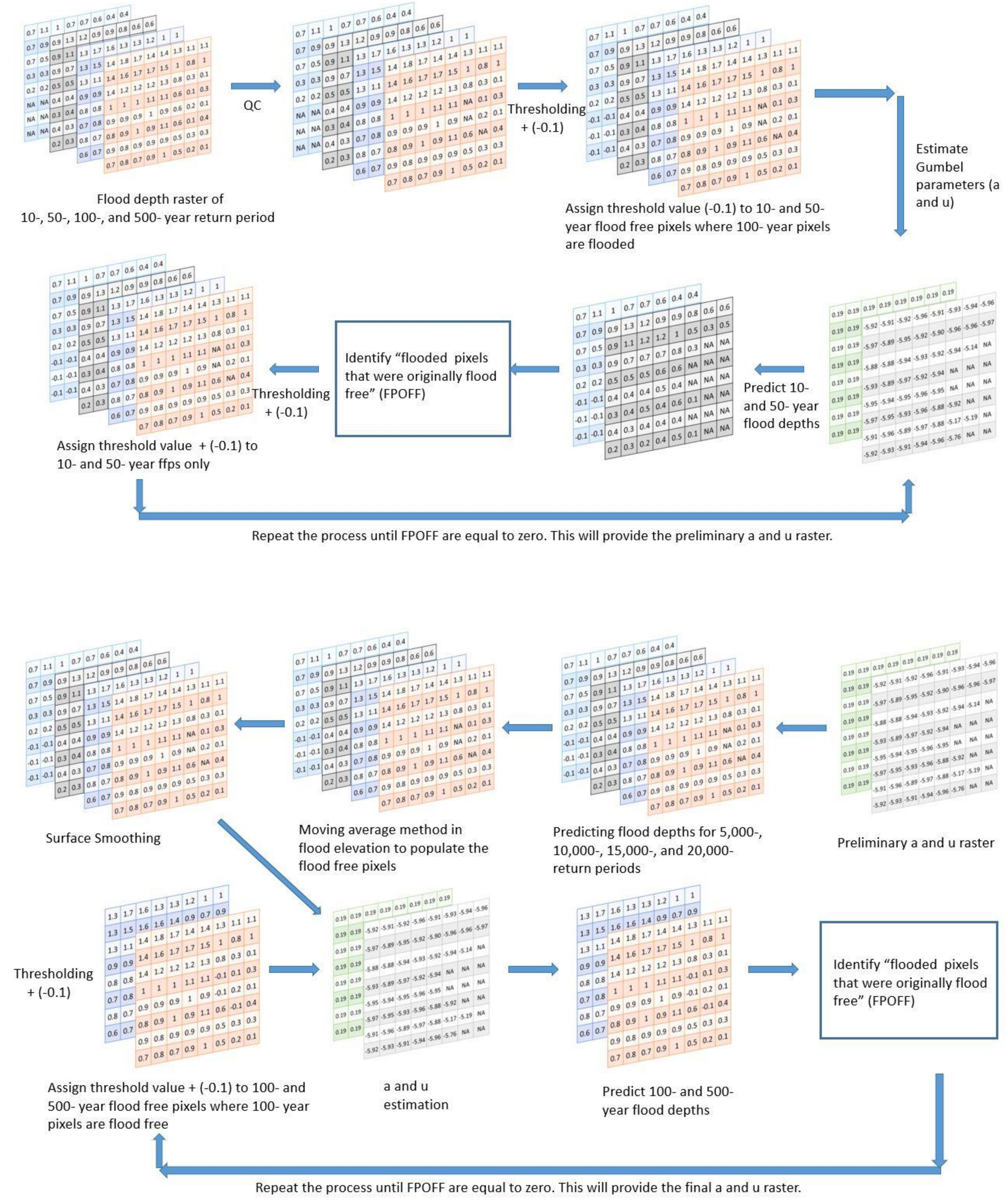

Figure 3: Summary of Case Study

Data Cleaning

The data cleaning process is run on the 121,204 cells in the study area. Data cleaning identifies 32 cells with a flood depth less than zero, 3,575 cells for which a lower-return-period flood depth equals or exceeds a higher-return-period flood depth, and 2,365 cells for which a positive shorter return period flood depth is accompanied by a "null" longer return period flood depth. The original flood depth values in a total of 5,972 cells (4.9% of the initial cells) are unused in the analysis because they fail one or more of these data cleaning tests, leaving 115,232 remaining cells with plausible flood depth and DEM data. For the 5,972 spurious cells, temporal or spatial interpolation is used to impute flood depths from values for the same cell from adjacent return periods or cells, respectively.

Gumbel Model Fitting

For each cell with cleaned, non-null flood elevation data for two or more of the four return periods, the Gumbel distribution is fit. No further processing is required for cells having derived flood depths at all four return periods.

Flood Depth Extrapolation to Longer Return Periods

Once the α and u parameters are corrected for all cells, they are used to extrapolate flood elevation surfaces for the 5,000-, 10,000-, 50,000-, and 100,000-year flood elevations in their respective cells.

Elevation Surface Creation

Several spatial interpolation techniques were applied, separately for each of the four longer return periods (i.e., 5,000 or more years). First, moving-average windows of 3x3 cells increasing incrementally from 3x3 to 5x5, 7x7, … 31x31 cells were used to impute values for missing cells, with the same-sized window implemented across the entire study area, for each iteration. The largest window (31x31 cells) is determined as the minimum size required to interpolate within the middle of the largest cluster of missing cells in the study area. The moving average window only populated the "NULL "cells. A smoothing operation (3x3) was performed to reduce the undulation of the flood elevation surface for entire study area. Nearest neighbor, kriging, co-kriging is also used (separately) to interpolate the missing cells at each of the longer return periods, separately.

Gumbel Model Fitting: Flood Parameter

For each cell, a new Gumbel distribution is fitted, using the 5,000-, 10,000-, 15,000-, and 20,000-year return period flood depths, with the new α and u parameters used to calculate the 10-, 50-, 100-, and 500-year depths. A comparison is then made with the Risk MAP depths; "null" cells in Risk MAP that have positive modeled flood depths are flagged to be corrected.

Discussion

The technique appears to provide improved estimates of flood depths in many cases, especially when the co-kriging spatial interpolation method is used, provided that the FEMA-generated originally modeled data can be assumed to be the "correct" values. In heavily populated areas, such as in the example of Metairie, Louisiana, refinements in the flood depths for short or long return periods may allow for improved understanding of infrastructure needs for accommodating floodwaters. Although the method is computationally intensive, it can be automated for improvement in flood depth estMost of the places in the U.S. have 100-year flood depth. Some places there are no flood data available. A few places have multiple return period of flood. Jefferson parish, Louisiana (study area) have multiple flood depths of different return period (10-, 50-, 100-, and 500-year). Some of the places in the study area are not affected by 500-year flood depth. That does not mean that specific location will not have any risk or vulnerability of flood. Higher return of flood (low probability) may damage that specific location. We will characterize the flood of an area using multiple return period of flood depth and elevation of that area using Quantile function gumbel distribution.imates for any location that is "data rich" regarding flood depth grids at multiple return periods.

Limitation and Future Work

As with any research, there are limitations to the analysis and interpretation of results. First, 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 (Intergovernmental Panel on Climate Change 2014, p. 8) prove to be accurate. A second concern is the uncertainty of extrapolating the Gumbel distribution based on only a small number of available return period/flood depth modeled data points, particularly for the most extreme events. Third, differences in local land cover (e.g., streets, roofs with vs. without gutters, and lawns) may cause differences in the Gumbel parameters for flood depths as a function of return period and in generating a continuous surface in 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 and mitigating extreme flood events.

Summary and Conclusion

Existing flood depth grids based on FEMA-generated model output provide communities with guidance data for preparing for and minimizing the flood hazard. However, these depth grids are only available for limited locations at a limited number of return periods. This study suggests a method for imputing flood depths for cases in which the FEMA-generated model output are not available or are spurious. The method involves fitting rasterized flood data to the Gumbel log-linear distribution of flood depth as a function of return period, by cell. The method then uses the Gumbel parameters of slope (α) and y-intercept (u) for flood elevations at extreme return periods for which it can be assumed that the study area is entirely flooded, and re-calculates the flood depths from the flood elevations based on these parameters, for 10-, 50-, 100-, and 500-year return periods through the use of spatial interpolation algorithms. Validation and sensitivity analyses are possible through comparison with FEMA-generated data. A case study of Metairie, Louisiana, is used to illustrate the technique. For the study area, the co-kriging technique offers the smallest RMSE, when compared to FEMA-generated model output flood depth grids. Validation and sensitivity analyses in the case study illustrate that the method offers improvements in estimation of flood depths for enhanced flood mitigation planning.

Future work should be conducted to provide estimates of similar flood depths in locations where FEMA-generated model output flood depth grids are unavailable. Flood risk analysis in such "data poor" locations might use the Gumbel parameters for adjacent "data rich" areas, or it might use other techniques. Additional work that considers the relationship between flood loss and flood depth is also needed, particularly for varying time intervals over which the flood depth is experienced. The present research offers the first step toward such analyses, in the interest of enhancing protection of life and property.

Funding

This project was funded by the Louisiana Sea Grant College Program (LSG) (Omnibus cycle 2020–2022; Award Number: NA18OAR4170098; Project Number: R/CH-03), and the Gulf Research Program of the National Academies of Sciences, Engineering, and Medicine under the Grant Agreement number: 2000-10880 "The New First Line of Defense: Building Community Resilience through Residential Risk Disclosure". Any opinions, findings, and conclusions or recommendations expressed in this research are those of the authors and do not necessarily reflect the views of the Gulf Research Program or the National Academies of Sciences, Engineering, and Medicine or LSG.