Modeling Extreme Hurricane Damage in the United States

Asim Kumer Dey* and Kumer Pial Das, Ph.D.†

Abstract: Hurricanes are one of the costliest natural disasters in the United States. After more than two decades of relatively little hurricane activity, the past decade saw heightened hurricane activity and more than $200 billion in damage in 2004 and 2005.This study analyzes U.S. hurricane damages from 1900-2012. Based on this analysis we propose an extreme value model for predicting extreme hurricane damage. Finally, a simulated hurricane series are generated by Bootstrap sampling to quantify the uncertainty in the inference of extreme return levels of hurricane losses.

Keywords: Extreme value model,Generalized Pareto distribution, Uncertainty,Monte Carlo Simulation, Bootstrap sampling.

1 Introduction

Hurricanes are responsible for the enormous loss of life and the massive monetary costs around the U.S. Gulf Coast. The purpose of this study is to quantify the economic risk of hurricanes. More specifically we want to predict magnitude of hurricane damage (return level) for a specific return period.Monte Carlo simulation is used to generate hurricane series based on proposed empirical model. A total 10000 hurricane event is generated by bootstrap resampling approach to measure the variability of estimates obtained from the empirical model.

The paper is organized into five sections: the first describes the damage data that are used in the analysis; the second introduces a brief description of extreme value model and generalized Pareto distribution(GPD); the third proposes a GPD model for hurricane damage; the fourth describes the uncertainty measurement in the inference of extreme return levels of the proposed model through simulated hurricane series; and the fifth discusses the conclusion of the paper.

*Graduate Student, Department of Mathematics, Lamar University, Texas, USA.Email: adey@lamar.edu

†Associate Professor, Department of Mathematics, Lamar University, Texas, USA.Email: kumer.das@lamar.edu

2 Data

We analyze the hurricane losses in the United States. Pertinent data span the years 1900-2012. Data have been taken from the National Oceanic and Atmospheric Association's National Hurricane Center (NOAA). In order to accurately compare historic loss amounts, a proper normalization method has been used to account for inflation, changes in population and wealth along the affected areas. Pielke-Landsea(2008)[1] normalization procedure has been used to normalize the data to 2014 values.

3 Extreme Value Model and Generalized Pareto Distribution(GPD)

Extreme value analysis is a branch of statistics dealing with the extreme deviations from the median of probability distributions. It seeks to assess, from a given ordered sample of a given random variable, the probability of events that are more extreme than any previously observed.

3.1 Classical Extreme Value Model

Let $X_1, X_2, \cdots, X_n$ be a sequence of independent random variables with common distribution function F. Extreme value analysis focuses on the statistical behavior of

$$M_n = max\{X_1, X_2, \cdots, X_n\}$$

In applications, the X_i usually represent values of a process measured on a regular time-scale, for example hourly measurements of sea level, or daily mean temperature so that M_n represents the maximum of the process over n times units of observation[2]. If n is the number of observations in a year, then M_n corresponds to the annual maximum.

In theory the distribution of M_n can be derived exactly for all values of n:

$$\begin{aligned} Pr\{M_n\} &= Pr\{X_1 \leq z, \cdots, X_n \leq z\} \\ &= Pr\{X_1 \leq z\}, \cdots, Pr\{X_n \leq z\} \\ &= \{F(z)\}^n \end{aligned} \tag{3.1}$$

In practice, we might not have the distribution function F but according to the extremal eypes theorem (Fisher and Tippett, 1928 [5]) if there exist sequences of constants $\{a_n\} > 0$ and $\{b_n\}$ such that

$$Pr\{(M_n - b_n)/a_n \leq z\} \longrightarrow G(z) \quad \text{as} \quad n \longrightarrow \infty$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:
I: Gumbel family:

$$G(z) \;=\; \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty \tag{3.2}$$

II: Frechet family:

$$G(z) \;=\; \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, \quad z > b; \quad 0, z \leq b \tag{3.3}$$

III: Weibull family:

$$G(z) \;=\; \exp\left\{-\left[-\left(\frac{z-b}{a}\right)^{-\alpha}\right]\right\}, \quad z < b; \quad 0, z \geq b \tag{3.4}$$

for parameters $a > 0$, b and, in case of families II and III, $\alpha > 0$.

Families I, II and III are termed as extreme value distributions. These three classes can be combined into a single family of models having distribution function of the form.

$$G(z) \;=\; \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)^{-1/\alpha}\right]\right\} \tag{3.5}$$

defined on the set $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, where the parameters satisfy $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \sigma < \infty$. This is the generalized extreme value (GEV) family of distributions.

For modeling extremes of a series of independent observations data are blocked into series of block maxima to which the GEV distribution can be fitted.Often the blocks are chosen to correspond to a time period of length one year, in which case n is the number of observations in a year and the block maxima are the annual maxima. Parameters of the model can be estimated by maximum likelihood estimation method. From the estimated parameters the return level for different return periods can be easily obtained.

3.2 Threshold Model and Generalized Pareto distribution(GPD)

Modeling only block maxima is a wasteful approach to extreme value analysis if other data on extremes are available. In this case the r largest order statistic model(Threshold Model) is a better alternative.

Consider again $X_1, X_2, \cdots, X_n$ be a sequence of independent random variables with common distribution function F. Then for suitably high threshold u, the execedences $y(= x - u)$, conditional of $X > u$, approximately follow GPD having the distribution function:

$$H(y) = 1 - (1 + \frac{\xi y}{\sigma})^{-1/\xi} \tag{3.6}$$

defined on $\{y : y > 0 \text{ and } (1 + \xi y/\sigma) > 0\}$ and for some $\sigma > 0$ and ξ.

To fit the GPD model we need to select an appropriate threshold. A number of methods are available to select the threshold of the data. Among them mean excess plot and threshold choice plot are widely used. The return levels for different return period can be easily calculated from inferred generalized Pareto model. The m-year return level is

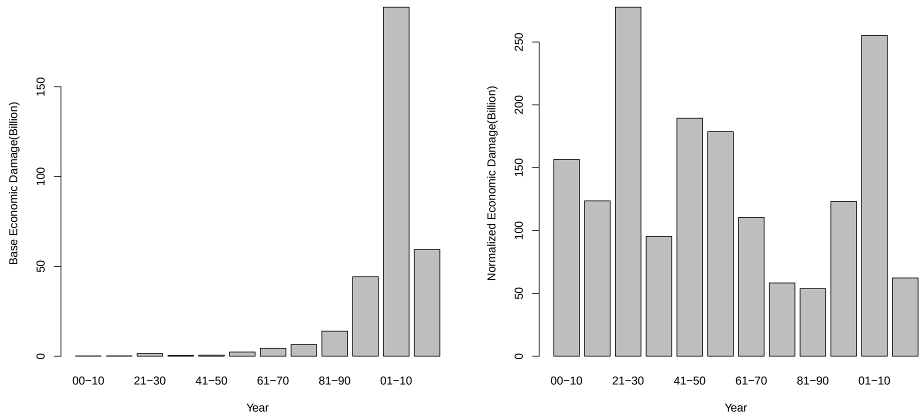
$$x_m = \mu + \frac{\sigma}{\xi}[(m\zeta_u)^\xi - 1] \tag{3.7}$$

where $\zeta_u = \lambda \times Pr(X > u)$ and $\lambda =$ rate for extreme events.

To assess the accuracy of the GPD model various diagnostic plots are used. Among them probability plot, quantile plot and density plot are commonly used to check the goodness-of-fit of the model [3].

4 Modeling Hurricane data with Generalized Pareto distribution (GPD)

This study focuses on the economic damage related to hurricane from 1900-2012.A histogram of the aggregate nominal losses is shown in Fig 4.1(a). The obvious upward trend is deceptive as it does not take into account change in inflation, population, or wealth. A bar plot of the normalized losses is shown in Fig 4.1(b). The bar plot shows that historically second highest hurricane lose occurred in the last decade, 2001-2010, whereas the highest hurricane damage happen in 1921-1930, the decade of Great Miami.



(a) Total base damage over period (b) Total normalized damage over period

Figure 4.1: Total hurricane damage, 1900-2012

The probability histogram (Fig 4.2) shows that the data follow the long tailed distribution. A standard technique for modeling such long tail data is the GPD.

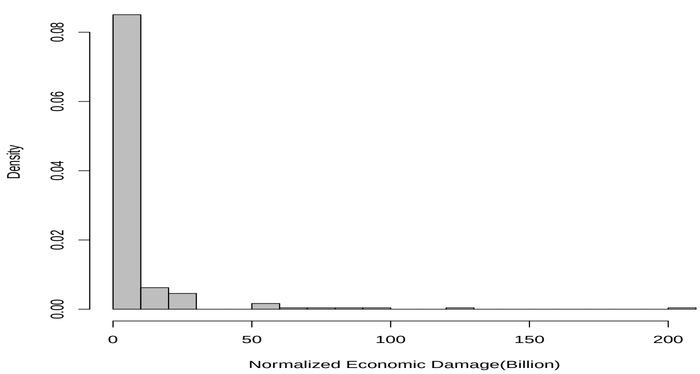


Figure 4.2: Probability Histogram

To fit the GPD model we need to select appropriate threshold first. Several graphical approaches are available to select the threshold. However, the most commonly used method is the mean excess (ME) plot. From the ME plot (Fig.4.3) we can see that after 25 the graph shows approximate linearity.Therefore a a better threshold could be 25.

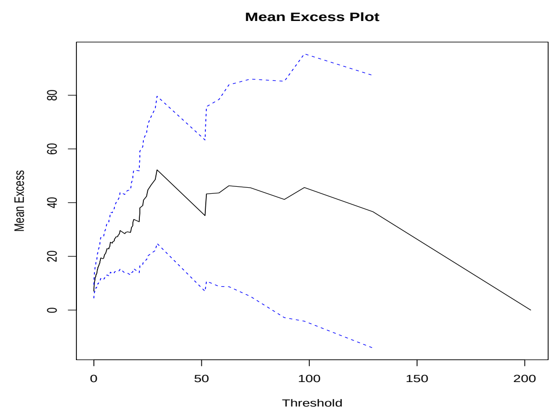


Figure 4.3: Mean Excess plot

Having determined a threshold, the parameter of the generalized Pareto distribution can be estimated by maximum likelihood method or by other estimation techniques.From our data we get the maximum likelihood estimates as

$$\hat{\sigma} = 41.90623 \text{ and } \hat{\xi} = 0.06729$$

Now using the return level formula (Eq.3.7) we can obtain approximate hurricane damage for a specific return period. Table 4.1 shows hurricane losses $X(T = t)$ for different return periods T. For example, at return period 50 the estimated hurricane damage is approximately 106.740 billion.

Table 4.1: Magnitude of hurricane damage for a return period

Return period (Year)	Return level(Billion)
20	64.614
30	82.935
50	106.740
100	140.378
200	175.622

Diagnostic plots for the fitted GPD are shown in Fig. 4.4. The plots suggest that the proposed Pareto model is reasonable for modeling this data set.

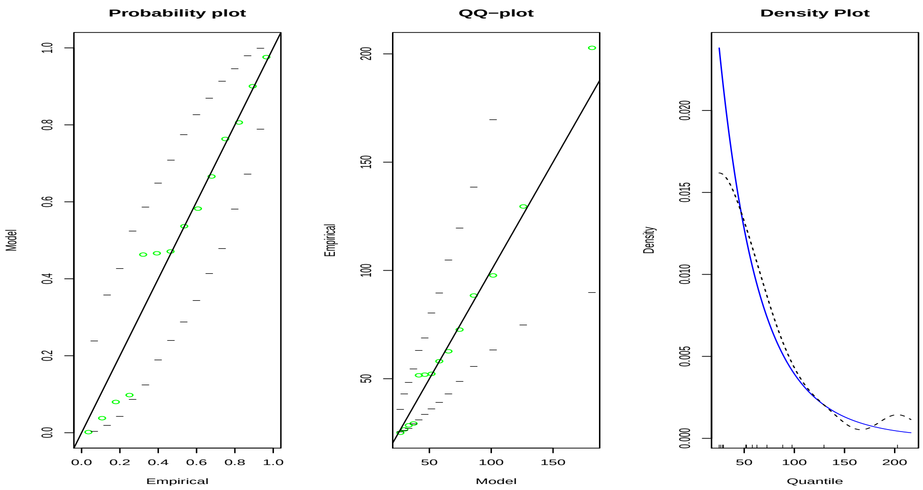


Figure 4.4: Diagnostic plots

5 Quantifying uncertainty of the estimates of empirical model

Using generalized Pareto model to infer extremal behavior of hurricane damage data is a common practice. A difficulty arises, however, in measuring statistical properties such as mean, variance, etc. of such return level estimates.We can find these properties of the estimates by drawing repeated samples of same size from the population of interest, a large number of times [4]. This resampling method is known as “naive” sampling. But the problem is that sampling again and again with Monte Carlo simulation from the same population causes to arise only a small degree of variability of the estimates.

However, potentially much greater uncertainties arise from sampling error in the the inference of the hurricane model itself.Therefore we generate a sample from the population and this sample is used as a “surrogate population”, for next steps of measuring properties of a statistic. A number of samples are then taken with replacement from the sample at hand. For each case we estimate the parameter of interest. And so we have a set of values of the estimator of the parameter. From this set of values we can calculate different statistical properties of the estimate of the parameter of interest.This resampling method is known as bootstrapping.

For uncertainty measure of extreme return levels of the empirical model we generate 10000 extreme hurricane series based on the proposed model. Here we make the following assumptions. First, that hurricane model is based on 112 years of historical records. Second, that the number of extreme hurricanes affecting is 241, that is the hurricane model is based on 241 data.

Then we generate 100 samples of size 241 from the original 10,000 simulated values allowing with replacement. From each sample we calculate the maximum likelihood estimates $\hat{\sigma}$ and $\hat{\xi}$ of the parameters, and then return levels for return periods $m = 20$, 30, 50, 100, 200, and 500. Therefore for each return period we have 100 values of return level. From this set of values we calculate the mean return level, sample standard deviation of these return level (standard error).For each of the specified return levels, the information is summarized in Table 5.1. We also report the coefficient of variation, and 95% confidence intervals for return level estimates.

Table 5.1: Results obtained from Bootstrap samples

Return Period	Mean Return Level	Standard error	Coefficient of variation	95% Lower Limit	95% Upper Limit
20	58.25	0.54	0.93	57.19	59.31
30	77.99	0.79	1.01	76.45	79.54
50	103.89	1.14	1.09	101.66	106.14
100	141.03	1.81	1.28	137.48	144.58
200	180.70	2.82	1.56	175.17	186.23
500	237.57	4.84	2.04	228.07	247.06

From table 5.1 we can see the location of the distribution increases with the return period. The variability also increases since there is greater uncertainty with increased model extrapolation. Finally, we observe that the coefficients of variation for return level estimates increase with the return period, reflecting the magnification of uncertainty under extrapolation.We can see clearly the expected return level with 95% upper and lower limit for a specific return period from the return level plot (Fig. 5.1).

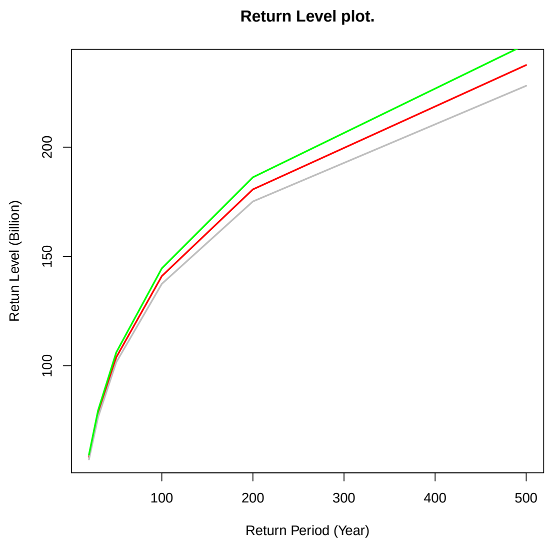


Figure 5.1: Return Level plot

6 Conclusion

The analysis highlights the forming of a model which predicts the possible economic damage associated with a future hurricane event in United States Gulf Coast and Atlantic shore of Florida.The propose model shows that the potential damages from hurricanes will be growing at an alarming rate that may place severe burdens on society.The analysis should provide a cautionary warning for hurricane policy makers.

References

[1] Roger A. Pielke Jr., Joel Gratz, Christopher W. Landsea, Douglas Collins, Mark A. Saunders and Rade Musulin (2008), "Normalized Hurricane Damage in the United States:1900-2005", *Natural Hazards Review*, **29**, 1527-6988.

[2] Coles, Stuart (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag.

[3] Enrique Castillo, Ali S. Hadi, N. Balakrishnan, and Jose M. Sarabia (2004) *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley-Interscience.

[4] Stuart Coles and Emil Simiu (2003), "Estimating Uncertainty in the Extreme Value Analysis of Data Generated by a Hurricane Simulation Model", *Journal of Engineering Mechanics*, **1288**, 0733-9399.

[5] Fisher, R and Tippet, L (1928), "Limiting Forms of the Largest or Smallest Member of a sample", Proceedinds of the Cambridge Philosophical Society, **24**, 180-190.