Statistical Distribution Fits for Hurricanes Parameters in the Atlantic Basin

Seyed Reza Hosseini ([a]), Marco Marani, ([b]), Marco Scaioni ([c])

([a]) Politecnico di Milano, Dip. di Ingegneria Civile e Ambientale, *seyedreza.hosseini@polimi.it*
([b]) Università degli Studi di Padova, Dip. di Ingegneria Civile, Edile e Ambientale, *marco.marani@unipd.it*
([c]) Politecnico di Milano, Dip. di Architettura, dell'Ambiente costruito e dell'Ingegneria delle costruzioni, *marco.scaioni@polimi.it*

Abstract
Extreme weather events have become more frequent and intense in many regions due to global climate change. Changes in hurricane frequency and intensity are, in particular, at the center of a debate of significant scientific and socio-economic relevance. In this view, understanding what probability distributions may be used to describe hurricane sustained wind velocity is a necessary step towards understanding their possible changes. The purpose of this study is to identify which statistical distributions best represent observed hurricane intensities and their extremes, particularly in the Atlantic Basin. Initially, as the main parameter of hurricanes activity, the Annual Hurricane Occurrence (AHO) is fitted with the Poisson and Negative Binomial distributions. This work shows that AHO is more consistent with an inhomogeneous Poisson model. Hurricane intensity is fitted with the Weibull, Inverse Gaussian, and Generalized Pareto distributions (GPD). It is also shown that the GPD has a better fit for the hurricane intensity with respect to other statistical distributions. To this end data from historical records in the period 1886-2015 have been used.

Riassunto
A causa del cambiamento climatico globale, gli eventi meteorologici estremi sono diventati sempre più frequenti in molte regioni del nostro pianeta. In particolare, i cambiamenti della frequenza e dell'intensità degli uragani sono oggi al centro di un vivace dibattito di notevole rilevanza dal punto di vista scientifico e socio-economico. In questa direzione, lo studio della distribuzione di probabilità che meglio si adatta per descrivere la velocità del vento che causa gli uragani costituisce un passo essenziale per capirne i possibili sviluppi. Lo scopo di questa ricerca è stato dunque quello di identificare quale distribuzione di densità di probabilità possa meglio rappresentare l'intensità degli uragani e dei relativi eventi estremi, in modo specifico nel bacino dell'Oceano Atlantico. In una prima fase, il numero degli uragani per anno (Annual Hurricane Occurrence - AHO), che può essere utilizzato come parametro principale per misurarne l'attività, è stato rappresentato tramite le distribuzioni di Poisson e Binomiale Negativa. Questo studio ha dimostrato che l'AHO può essere meglio interpretato quando si utilizza un modello di Poisson non omogeneo. L'intensità degli uragani è stata rappresentata tramite le distribuzioni di Weibull, la Gaussiana Inversa e la Generalizzata di Pareto (GPD). Si è valutato inoltre che la GPD mostra un migliore adattamento all'intensità degli uragani rispetto ad altre distribuzioni statistiche di probabilità. Per questo studio sono stati utilizzati i dati presenti negli archivi storici relativi al periodo 1886-2015.

Introduction
Fluctuations in Atlantic hurricane frequency have a significant impact on human life and property. The climatology of Atlantic hurricanes has long been a subject of research, but interest in

characterizing hurricane activity has intensified in recent years in view of possible climate change impacts. Different methods have been developed to forecast hurricane activities (Gray et al., 1994; Elsner, Schmertmann 1993; Elsner et al., 1999). Modeling annual occurrence is the first step for studying hurricanes. According to domain knowledge in meteorology, the best statistical distribution of the number of hurricanes occurring per year is either the Poisson distribution or the Negative Binomial distribution (Chen et al., 2003). Accordingly, in the first part of this research these two distributions have been scrutinized to see which one may best model Annual Hurricane Occurrence (AHO).
Since 80% of all hurricane damage is caused by less than 20% of the most intense events (Jagger, Elsner, 2006), the analysis of extreme events has been the next objective of this research. Maximum sustained wind speed is a common indicator of the intensity of storms. Here explore a wide range of possible probability distributions with the aim to identify the one that accurately reproduces the observed frequency of hurricane events.

Historical Records

For the analyses carried out in this study we used a data set obtained from the National Hurricane Center (NHC) for the Atlantic basin, known as "Atlantic HURDAT2" ("The revised Atlantic hurricane database", Landsea et al., 2015). This historical database contains occurrences and six-hourly records of Maximum Sustained Wind (MSW) for all tropical cyclones from 1851 to 2015. MSW is defined as the maximum 1-min average wind associated with the tropical cyclone at an elevation of 10 m a.s.l. with an unobstructed exposure. Recorded values of MSW in HURDAT2 are given to the nearest 10 knots for the years 1851 through 1885, and to the nearest 5 knots from 1886 to date. With the intention of having a homogeneous data set, and to maximize the reliability of the data, only data recorded after 1886 have been used.

Annual hurricane occurrence

The first step when studying hurricanes is to determine the frequency of annual occurrences. The Poisson and the Negative Binomial distributions have been applied to rare events in Meteorology for a long time. The Poisson distribution has mostly been used in recent works for modeling the annual occurrence of storms (Elsner, Schmertmann, 1993; Elsner et al., 1999; Bove et al., 1998; Jagger, Elsner, 2006). The purpose of this section is to compare the performance of the Poisson and of the Negative Binomial models in describing the distribution of hurricanes in the Atlantic Ocean. All records of the annual counts of tropical cyclones in the Atlantic Ocean are acquired from HURDAT2. Among 1563 events recorded between 1886 and 2015, 731 tropical cyclones had hurricane intensity (> 64 knot) (Table 1).

Tropical cyclone	Wind speed (knot)	Counts
All recorded	0-165	1563
Tropical storm	35-60	590
Hurricane	>65	731

Table 1 – Tropical Cyclone Counts

Time series and distribution of the annual occurrence of hurricanes for 130 years from 1886 to 2015 in the Atlantic Ocean are shown in Figure 1.

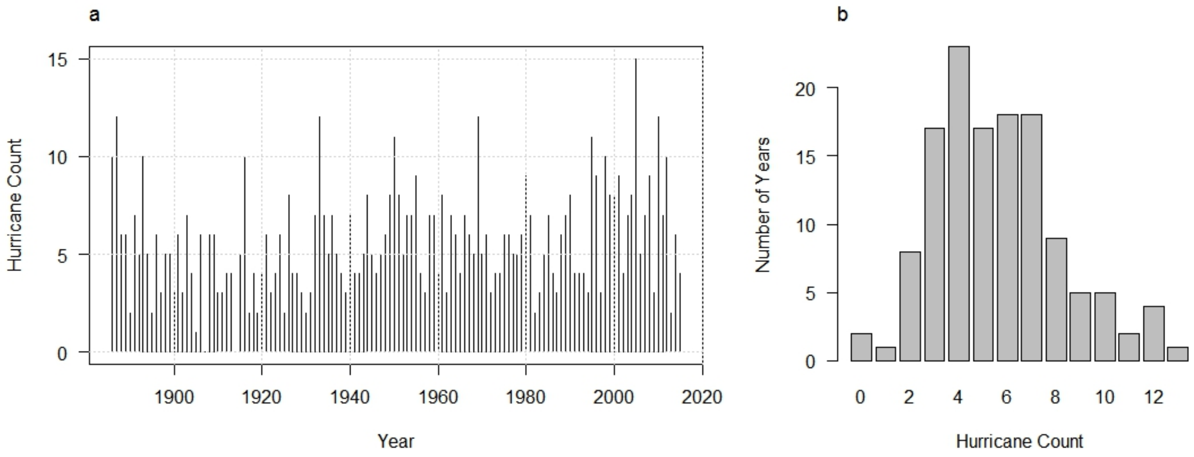

Figure 1 – Annual Hurricane Occurrence (AHO). (a) Time series and (b) frequency distribution.

The Poisson distribution describes the number of events randomly occurring within a time interval of fixed duration (1 year in the present case). The Poisson distribution is defined by a single parameter, $\lambda \geq 0$, which is also the mean of the distribution.

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}; \qquad x = 0,1,2,\ldots,\infty \qquad [1]$$

The probability mass function of Negative Binomial distribution, when R is an integer and where $q = 1 - p$ is:

$$f(x) = \binom{R + x + 1}{x} p^R q^x; \qquad x = 1,2,\ldots,\infty \qquad [2]$$

The Negative Binomial distribution is defined in terms of two parameters, R (>0, number of successes), and p ($0 < p \leq 1$, probability of success). Parameter values for both distributions are estimated using the Maximum Likelihood Estimator (MLE) (Table 2).

Distribution	Estimated distribution parameters
Negative Binomial	R = 22.29, P = 0.79
Poisson	λ = 5.62

Table 2 - Estimated distribution parameters

The goodness of fit for the two distributions is evaluated using different metrics. Here the Kolmogorov-Smirnov goodness-of-fit test is used (Table 3).

Distribution	Kolmogorov-Smirnov	
	Statistic	Rank
Poisson	0.143	1
Negative Binomial	0.150	2

Table 3 - Goodness-of-fit tests

As the statistics of goodness-of-fit test indicate the distance between data and fitted distributions, the distribution with the lowest statistic value is the best fitting. It is interesting to note that, even though the Poisson distribution is defined in terms of a lower number of parameters it does better fit the sample at hand. The basic assumption of the Poisson distribution is that the mean number of hurricanes in any two non-overlapping time intervals of the same length should be equal. From the form of the Poisson distribution, it follows that the variance should be numerically equal to the mean, i.e. the variance-to-mean ratio should be equal to one. Table 4 shows the statistics of the observed variance to mean ratios. Interestingly, the observed variance of the yearly hurricane counts is on average 27% greater than the mean.

Sample size (N)	130	Min	0
Mean	5.62	Median	5
Variance	7.15	Max	15
Variance/Mean	1.27	Std. deviation	2.67

Table 4 - Descriptive statistics of AHO

A Monte Carlo (MC) simulation is performed to check how likely it is for a Poisson distribution to generate a variance-to-mean ratio equal to 1.27 in a 130-year sample. For this simulation m_1=1000 samples are generated from a Poisson distribution with sample size n=130 and lambda λ= 5.62. For each sample, the ratio of variance to the mean is computed. The results show that only 2.4% of the ratios are larger than 1.27. In other words, if the distribution of hurricane occurrence were perfectly Poisson, a variance-to-mean ratio equal or greater to 1.27 only occurs with a probability approximately p=0.024. This suggests that hurricane occurrence is unlikely to be a Poisson process and that it exhibits significant “clustering”, i.e. years with large (low) hurricane activity tend to be followed by years with large (low) hurricane activity. Hence, our results suggest the presence of a significant inter-annual correlation in hurricane occurrence.
The higher-than-random observed variation could be due to changes in climatic conditions, which can be systematic (a trend, e.g., in Sea Surface Temperature, SST) or due to long-term oscillations (e.g. associated with the North Atlantic Oscillation, NAO) In order to test this assumption, another MC simulation has been performed by using a Gamma distribution to generate a varying rate of Poisson occurrence. The scale parameter was assumed to be 0.7 and the shape parameter was taken equal to 5.6. These values were chosen such that their product is equal to the observed mean, as required by the properties of a Gamma distribution. For this new simulation, m_2=1,000 random Gamma values have been generated and for each Gamma, 130 years of hurricane counts have been produced. The results show that variance-to-mean ratios larger than 1.27 occur with a probability of about 0.98. We thus conclude that the observed hurricane occurrence process is compatible with a variable rate of Poisson occurrences.

Lifetime Maximum Intensity

MSW is a standard measure to indicate the intensity of storms. The data set for the analysis in this section is the Lifetime Maximum Intensity (LMI) of each hurricane from 1886 to 2015. Lifetime refers to the time from genesis to dissipation and lifetime maximum refers to the highest MSW during this lifetime. Our interest is the value of MSW where the lifetime maximum occurs. For each hurricane only one value (the highest MSW during its lifetime) has been collected as the LMI of that hurricane. The histogram of LMI for this dataset is shown in Figure 2.

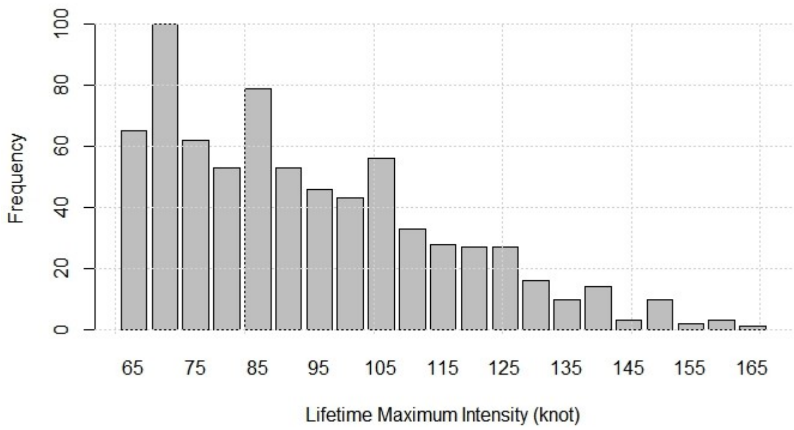

Figure 2 - Histogram of Lifetime Maximum Intensity (LMI)

Distribution of extreme wind speeds in tropical storms and hurricanes are conventionally modeled with Weibull or Generalized Pareto distribution (GPD), see (Heckret, 1998; Palutikof, 1999; Jagger, Elsner, 2006). The choice of these distributions is also consistent with LMI data. The histogram of the LMI in Figure 2 shows that data are right-skewed. In addition, the LMI cannot contain negative values so only non-negative distributions should be considered. Inverse Gaussian distribution is also used to model non-negative positively skewed data. The purpose of this section is to determine which of these three models is the most appropriate to describe the distribution of LMI in the Atlantic Ocean. PDF's and parameters of selected distributions are presented in the following (Forbes, 2011).

Weibull (Shape α and Scale β):

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}; \qquad 0 \le x \le +\infty, \quad \alpha > 0, \quad \beta > 0 \qquad [3]$$

Generalized Pareto (Shape k and Scale σ):

$$f(x) = \left(\frac{1}{\sigma}\right)\left(1 + k\,\frac{(x-\theta)}{\sigma}\right)^{-1-\frac{1}{k}}; \qquad k > 0, \quad \theta < x; \qquad k < 0, \quad \theta < x < \theta \qquad [4]$$

Inverse Gaussian (Shape λ and Scale μ):

$$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}}\, e^{\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)}; \qquad 0 \le x \le +\infty, \quad \lambda > 0, \quad \mu > 0 \qquad [5]$$

The GPD model requires a threshold intensity θ. The chosen intensity threshold should be, on one hand, high enough that the positive residual values follow a GPD, on the other hand low enough that there are enough values to accurately estimate the GPD parameters. Different techniques exist to support threshold selection. One example is Conditional Mean Exceedance (CME) graphs, also known as Mean Residual Life (MRL) graphs (Scarrott et al., 2012). For this analysis only the LMI of tropical storms of hurricane intensity (> 64 knot) have been collected, therefore, to keep all the intensities, the threshold for GPD has been selected equal to 64 knots, which is the minimum value of the data set. Similar to AHO, parameters of selected distributions have been obtained using MLE method. Estimated distribution parameters are shown in the Table 5.

Distribution	Estimated distribution parameters	
Weibull	α = 4.325	β = 101.717
Generalized Pareto	k = - 0.412	σ =42.734
Inverse Gaussian	λ = 1740.98	μ = 92.879

Table 5. Estimated distribution parameters

To determine how well the selected distributions are fitted, a goodness-of-fit test has been performed. The Kolmogorov-Smirnov test has been used to arrange the distributions in the order of performance and results are presented in Table 6.

Distribution	Kolmogorov-Smirnov	
	Statistics	Rank
Generalized Pareto	0.073	1
Inverse Gaussian	0.136	2
Weibull	0.152	3

Table 6. Goodness-of-fit tests

Statistics of Kolmogorov-Smirnov goodness-of-fit test indicate a lower value and therefore a better fit for GPD. A Q-Q plot is a plotting of quantile values of a real sample against the corresponding theoretical quantile values from the fitted distribution and is often used to visually represent the goodness of fit. The comparison of Q-Q plots for the three distributions are shown in Figure 3.

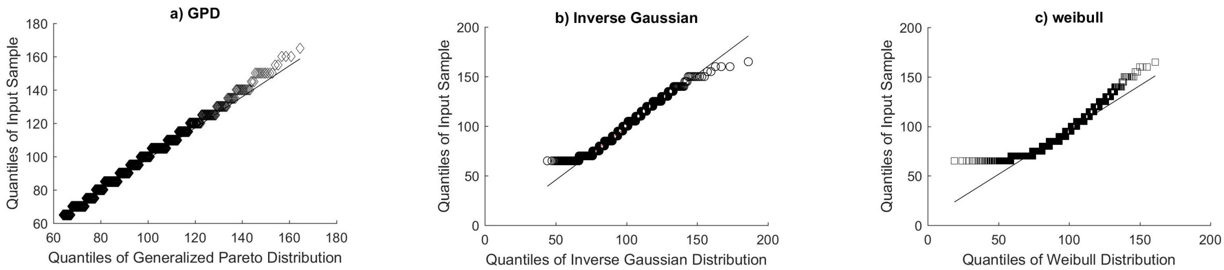

Figure 3 – Q-Q plots of (a) GPD, (b) Inverse Gaussian and (c) Weibull distribution

The Q-Q plots indicates the same ranking for the distributions. On the basis of the results of goodness-of-fit test and the Q-Q plot, the GPD with parameters k = -0.412 and σ =42.734 is the best fit for the LMI of hurricanes in the Atlantic Ocean.

Summary and Conclusions

The focus of this study is to fit different distributions to Annual Hurricane Occurrence (AHO) and Lifetime Maximum Intensity (LMI) in the Atlantic Ocean for 130 years from 1886 until 2015, and to determine the most valid statistical models. The comparison between homogeneous Poisson distribution and Negative Binomial distribution shows a better fit for the former. However, the information of the data set indicates that the equality of the variance-to-mean ratio, which is required for the homogeneous Poisson distribution, is not complied with. Therefore, two Monte Carlo simulations have been performed to understand whether this should be expected when applying the Poisson distribution. The results have shown that AHO is more consistent with an inhomogeneous Poisson model. It is mentioned that this could come from climate change conditions such as North Atlantic Oscillation or Sea Surface Temperature. The LMI for all the tropical cyclones of hurricane intensity for the period 1886-2015 have been calculated from HURDAT2 data set. Three different distributions of GPD, Weibull and Inverse Gaussian have been fitted to LMI with MLE method. The estimated distributions have been evaluated with the Kolmogorov-Smirnov goodness-of-fit test and Q-Q plot. The results of both tests have shown a better fit for GPD with respect to other statistical distributions.

Acknowledgements

This research has made use of the revised Atlantic hurricane database HURDAT2 from the National Hurricane Center (NHC).

References

Bove M, O'Brien J, Eisner J, Landsea C, Niu X. (1998), "Effect of El Niño on U.S. Landfalling Hurricanes", Revisited. Bull. Amer. Meteor.

Chen S, Gulati S, Hamid S, Huang X, Luo L, Morisseau-Leroy N, Powell M, Zhan C, Zhang C. (2003) "A Three-tier System Architecture Design and Development for Hurricane Occurrence Simulation", Proc. of the IEEE Int. Conference on Information Technology: Research and Education

Elsner J, Jagger T. (2013), *Hurricane Climatology: A Modern Statistical Guide Using R,* Oxford University Press, New York, ISBN 978-0-19-982763-3, 390 pp

Elsner J, Kara A, Owens M, (1999), "Fluctuations in North Atlantic Hurricane Frequency", J. Climate, 12, 427–437

Elsner J, Schmertmann C, (1993), "Improving Extended-Range Seasonal Predictions of Intense Atlantic Hurricane Activity", Wea. Forecasting, 8, 345–351

Forbes C, Evans M, Hastings N, Peacock B. (2011), *Statistical Distributions*, 4th Edition, Wiley, ISBN 978-0-470-39063-4, 220 pp

Gray W, Landsea C, Mielke Jr P, Berry k. (1994), "Predicting Atlantic Basin Seasonal Tropical Cyclone Activity by 1 June", *Wea. Forecasting*, 9, 103–115

Heckert N, Simiu E, Whalen T. (1998), "Estimates of Hurricane Wind Speeds by Peaks Over Threshold Method", J. Struct.

Jagger T, Elsner J, (2006), "Climatology Models for Extreme Hurricane Winds near the United States", J. Climate, 19, 3220–3236

Landsea C, Franklin J, Beven J. (2015), The revised Atlantic hurricane database (HURDAT2), nhc.noaa.gov, National Ocean and Atmosphere Administration

Palutikof J, Brabson B, Lister D, Adcock S. (1999), "A Review of Methods to Calculate Extreme Wind Speeds", Met. Apps, 6: 119–132

Scarrott C, MacDonald, A. (2012), "A review of extreme value threshold estimation and uncertainty quantification", REVSTAT - *Statistical Journal* 10, 33–60