3-20-2014

Distribution Fits for Various Parameters in the Hurricane Model

Victoria Oxenyuk
Florida International University, vikusek@gmail.com

DOI: 10.25148/etd.FI14040863

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

DISTRIBUTION FITS FOR VARIOUS PARAMETERS

IN THE HURRICANE MODEL

A thesis submitted in partial fulfillment of

the requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS

by

Victoria Oxenyuk

2014

To: Dean Kenneth G. Furton
College of Arts and Sciences

This thesis, written by Victoria Oxenyuk, and entitled Distribution Fits for Various Parameters in the Hurricane Model, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

____
Shahid S. Hamid

____
B.M. Golam Kibria, Co-Major Professor

____
Sneh Gulati, Co-Major Professor

Date of Defense: March 20, 2014

The thesis of Victoria Oxenyuk is approved.

____
Dean Kenneth G. Furton
College of Arts and Sciences

____
Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2014

ABSTRACT OF THE THESIS

DISTRIBUTION FITS FOR VARIOUS PARAMETERS

IN THE HURRICANE MODEL

by

Victoria Oxenyuk

Florida International University, 2014

Miami, Florida

Professor Sneh Gulati, Co-Major Professor

Professor B.M. Golam Kibria, Co-Major Professor

The FPHLM is the only open public hurricane loss evaluation model available for assessment of hazard to insured residential property from hurricanes in Florida. The model consists of three independent components: the atmospheric science component, the vulnerability component and the actuarial component. The atmospheric component simulates thousands of storms, their wind speeds and their decay once on land on the basis of historical hurricane statistics defining wind risk for all residential zip codes in Florida.

The focus of the thesis was to analyze atmospheric science component of the Florida Public Hurricane Loss Model, replicate statistical procedures used to model various parameters of atmospheric science component and to validate the model. I establish the distribution for modeling annual hurricane occurrence, choose the best fitting distribution for the radius of maximum winds and compute the expression for the pressure profile parameter Holland B.

TABLE OF CONTENTS

I. INTRODUCTION

Background

Hurricanes are one of the greatest natural hazards; although relatively rare in occurrence they can cause colossal economic losses. In 1992, "when Hurricane Andrew struck Florida it caused over $30 billion in direct economic losses" (Lokupitiya et al., 2005). Hurricane modeling has become a widely used tool for assessing risks associated with windstorm catastrophes. Since the groundbreaking studies of Russell (1968, 1971) and Tryggvason et al. (1976) the modeling methods have improved significantly as a consequence of increased computing capabilities, new advanced physical and statistical models and vast growth in quantity and quality of available data. Several private models for simulating hurricane loss have been developed in the recent years for use in the State of Florida but such models are typically commercial and are not available to research community and public. The Florida Public Hurricane Loss Model (FPHLM) is a notable exception.

The FPHLM is an open public hurricane loss evaluation model which was developed as a joint effort of specialists in fields of meteorology, engineering, computer science, actuarial science, finance and statistics from Florida International University, NOAA Hurricane Research Division, University of Miami, Florida State University, Florida Institute of Technology and University of Florida. The model was created "for a purpose of probabilistic assessment of risk to insured residential property associated with wind damage from hurricanes" (Hamid et al., 2010). The FPHLM consists of three main components (Figure 1): first – the atmospheric science component which models the track and intensity of hurricanes in Florida threat area; second – the engineering component which models vulnerability of insured property; and third – the actuarial science component which models the insured loss. The atmospheric science component simulates thousands of storms, their wind speeds and their decay once on land on the basis of historical hurricane statistics defining wind risk for all residential zip codes in Florida. The wind risk information is then passed on to the engineering and actuarial science components to assess damage and annual loss. Each component is developed independently and delivered as a one-way input to the next component in line until the end result is achieved.

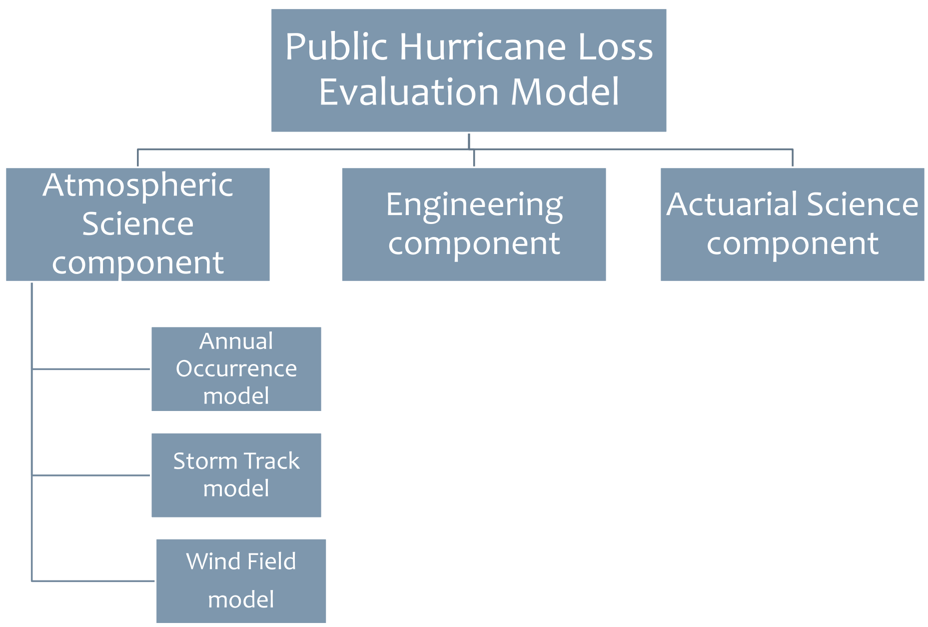


Figure 1. Structure of FPHLM

Problem Description

The aim of my thesis is to analyze atmospheric science component of FPHLM, repeat statistical procedures that were used to model parameters of meteorological component and to validate the model.

Theoretical Perspective

The atmospheric component of FPHLM includes annual occurrence model, which simulates the number of storms in a year, the storm track model, demonstrating the trajectory and intensity of hurricanes, and the wind field model.

Modeling annual hurricane occurrence (AHO) is the first step of atmospheric science component. Since the available historical data of documented hurricanes are limited, simulation that replicates fundamental characteristics of existing data has to be run to supplement the number of hurricanes. Statistical distribution of the number of hurricanes occurring per year is essential for such a simulation. "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).

One of the goals of my thesis is to determine which of the two distributions is the best for modeling AHO: Poisson distribution that assumes homogenous hurricane frequencies (the mean number of hurricanes in any two years is the same) or Negative binomial distribution that assumes non-homogenous annual occurrence rate.

Wind field model is another component within a hurricane risk model, which is dedicated to simulating hurricanes, their wind speeds and their decay once on land on the basis of historical data. The wind field model is later used for engineering simulation of the damage to insured property and actuarial calculation of the resulting loss.

Two fundamental components of the wind field model are radius of maximum winds (R_{max}) and central pressure at landfall. These two variables are most relevant for estimating loss since the greater the area of strike the greater the damage and the lower the central pressure the more intense the hurricane. The radius of maximum winds has a substantial impact on the area affected by the hurricane and modeling of R_{max} influences the likelihood of the location experiencing strong winds in cases of near misses. Modeling the distribution of R_{max} is therefore critical for estimating the possible losses for insurance pricing purposes.

The FPHLM finds the Gamma distribution to be the best fit for the R_{max}. In the present thesis I will determine how well Gamma distribution fits the R_{max} data and try to find if there are distributions that fit the data better than the Gamma distribution.

Holland B is an additional parameter defining the pressure field and maximum wind speeds in a hurricane. It was introduced by Holland (1980) and has been used in many hurricane threat studies since (Powell et al., 2005, Emanuel et al., 2006, Lee and Rosowsky, 2007, Hall and Jewson, 2008, and Vickery et al., 2009).

As a pressure profile parameter Holland B allows for the distinction in the maximum wind speeds observed in hurricanes for a given Δp (difference between central minimum sea level pressure and an outer peripheral pressure) all else being equal. The omission of B results in maximum wind speeds proportional to $\sqrt{\Delta p}$, whereas with B the maximum wind speed in the simulated hurricane is proportional to $\sqrt{B\Delta p}$.

The FPHLM shows that Holland B parameter is inversely correlated with both the size and latitude of the hurricane. Finding an exact statistical relationship between B and radius of maximum winds and latitude is another goal of the present research.

Data

The analysis of annual hurricane occurrence and radius of maximum winds will use data obtained from historical record for the Atlantic tropical cyclone basin (known as

“HURDAT”) for the period from 1901 until 2010. Earlier data are available but not used because of the lack of population centers and uncertainties about meteorological measurements before the start of 20th century.

A model for the Holland B pressure profile parameter will be developed on the basis of a subset of the data published by Willoughby and Rahn (2004) and obtained by NOAA and U.S. Air Force Reserve aircraft between 1977 and 2000.

To find the best fitting distribution the preliminary analysis of the data will be done through the use of EasyFit software which allows us to easily fit a large number of distributions to the data. Estimated parameters of the best fitting distributions will then be found using maximum likelihood estimator (MLE) method. In order to determine how well the selected distributions fit the data they will be tested for goodness-of-fit using Kolmogorov-Smirnov, Anderson-Darling and Chi-Square tests. Along with the goodness of fit tests the probability density function graphs, Q-Q and P-P plots will be used to visually assess the goodness of fit and empirically compare several fitted models. In order to determine the model for estimating Holland B, multiple regression analysis will be performed using the Proc REG procedure in SAS.

Project Organization

The purpose of my research is to examine the first part of the Florida Public Hurricane Loss Model – the atmospheric component - and check distributions of several parameters of the model. The thesis consists of six chapters and four appendices.

In the second chapter I look at the available data from the historical prospective and check for increasing trends in hurricane intensity, size or number of hurricanes striking Florida. The third chapter establishes the distribution for modeling annual hurricane occurrence. In the fourth chapter the best fitting distribution for the radius of maximum winds is identified, and finally in the fifth chapter the expression for the pressure profile parameter Holland *B* is computed.

The last chapter presents the final results and conclusions. The appendices contain databases for radius of maximum winds and Holland B as well as SAS codes and outputs.

II. TIME TRENDS

One of the important questions asked by scientists when discussing the FPHLM is "whether the distribution of hurricane loss should reflect climate change (i.e., an increasing trend in hurricane intensity)" (Katz, 2010). Indeed climate change has been a growing topic of discussion so it would only be reasonable to suspect increasing trend in hurricane intensity, size or number of hurricanes striking Florida. The parameters of the FPHLM are assessed from the historical record under the assumption of stationarity and the validation of this assumption will be an important aspect of my research.

The debate about whether warming tropical sea surface temperatures are producing more intense and long-lived cyclones has been going on for over a decade. Although Emanuel (2005) and Webster et al. (2005) have found that intensity and number of hurricanes show an increasing trend, studies by Klotzbach (2006) and Shapiro and Goldenberg (1998) conclude that most of this increase is most likely a result of improved observational technology.

First I will look at the hurricane occurrence and see if the number of hurricanes has been higher in the recent years.

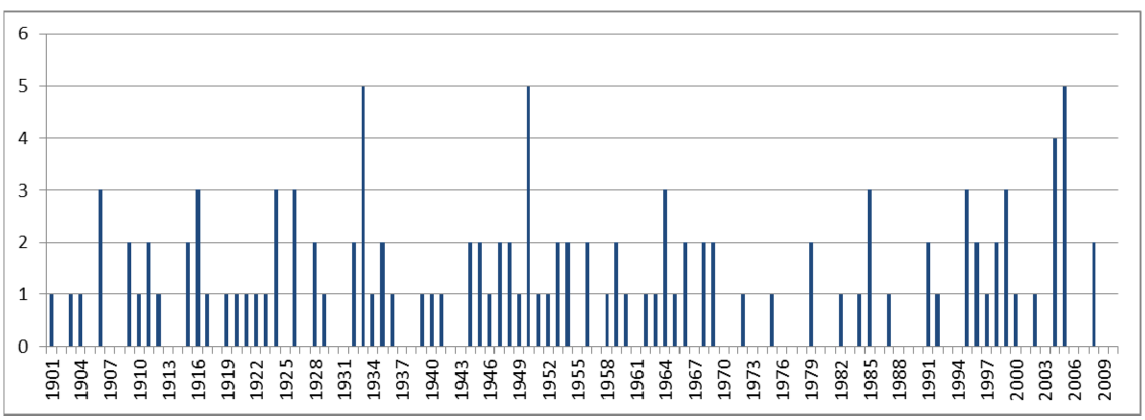


Figure 2. Number of hurricanes

The time series of annual hurricane counts (Figure 2) does not suggest a growing trend. There are also no other visible patterns. The number of hurricanes occurring in a year does not appear to be increasing.

If frequency of damaging hurricanes is rising then it should be seen on the plot of radius of maximum winds in time or the plot of central pressure in time.

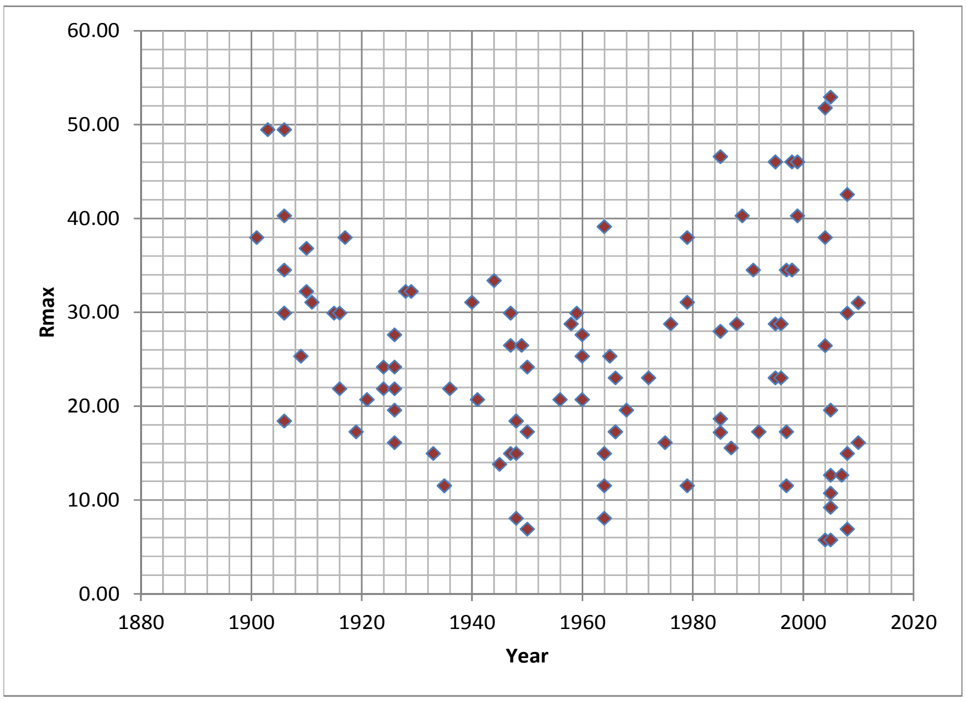


Figure 3. Plot of Radius of Maximum Winds vs. Time

The plot of radius of maximum winds (Figure 3) reveals higher values of Rmax in 1910s as well as 1990s and 2000s but no clear increasing trend can be detected.

The plot of central pressure in time (Figure 4) does not show any increase in hurricane strength. There are few higher values in 1980s but those indicate hurricanes of lower intensity. The strongest hurricanes with lowest central pressures are in 1920s, 1960s and 1990s-2000s. Thus I conclude that there is no evidence of increasing trend in hurricane intensity or size.

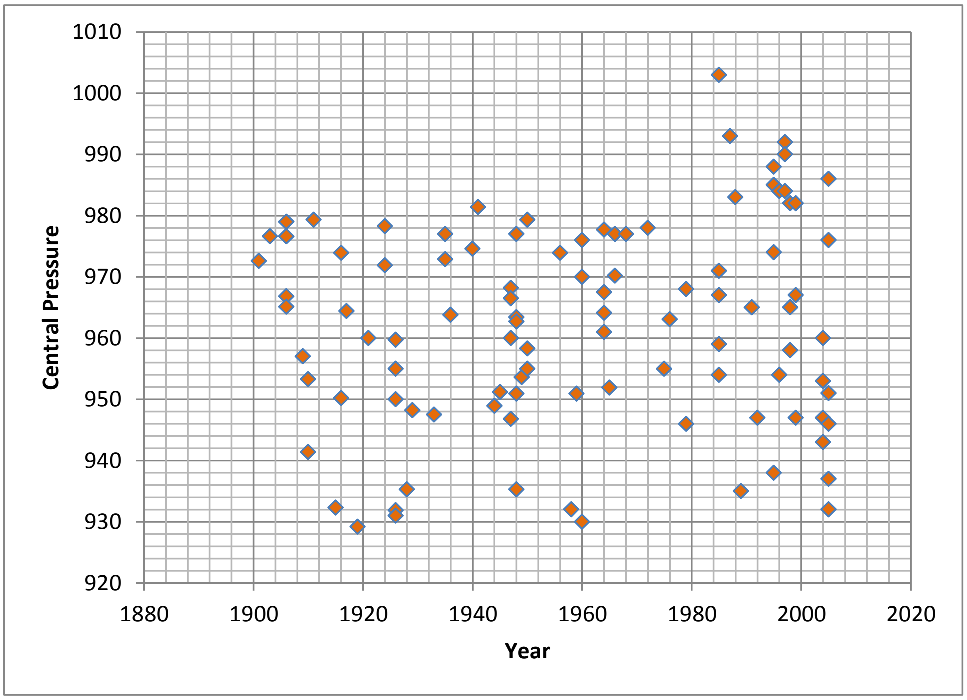


Figure 4. Plot of Central Pressure vs. Time

My findings do not support the argument that global tropical cyclone intensity, frequency and longevity have undergone increases in recent years. I conclude that no significant increasing trend is evident.

III. ANNUAL HURRICANE OCCURRENCE

The first step in study of hurricanes and their impacts is to determine the frequency with which they occur. Annual Hurricane Occurrence (AHO) rate estimates "the frequency of hurricanes occurring in a series of years based on an associated hurricane occurrence probability distribution, which is obtained through statistical analysis and calculation on the basis of historical hurricane records" (Chen et al., 2003). Substantial research in the area of modeling occurrence of hurricanes has been done in recent years by Chen et al. (2003 and 2004), Gray et al. (1992), Elsner and Schmertmann (1993), Elsner and Jagger (2004). The basic principle of these papers was to generate the statistical models from the available historical data in order to estimate AHO. Using obtained probability distributions the number of hurricanes per year in the future is produced for a desired number of years.

Rare events in meteorology are classically described by the Poisson and the Negative Binomial distributions. The rate of occurrence of a stochastic process is typically described by the use of the Poisson distribution. However, Poisson distribution assumes the mean number of storms in any two non-overlapping time intervals of the same length to be equal. To allow those means to be unequal will lead to the annual occurrence modeled by a Negative Binomial distribution. General guiding principles as to the adequacy of the two distributions have been discussed (Thom, 1966) but one cannot accurately determine which model is appropriate until tests are conducted.

In this section I will determine whether either Poisson or Negative Binomial is adequate in describing the distribution of the annual hurricane occurrence.

For the assessment of the AHO distribution to be conducted, a suitable dataset has to be obtained. Annual counts of tropical storms and hurricanes in the Atlantic Ocean are obtained from HURDAT database, which is maintained by the National Hurricane Center in Miami, Florida and the National Climatic Data Center in Asheville, North Carolina. The historical record for the Atlantic tropical cyclone basin contains six hourly record positions and intensities of tropical storm and hurricane for the period from 1851 to 2010. Only data beginning the 1901 are going to be used in my research because of unreliability of 19th century data. To focus on storms capable of affecting residential property in Florida, only storms in threat area (Figure 5) - within 1000 km of a location (26.0 N, 82.0 W) are being counted.

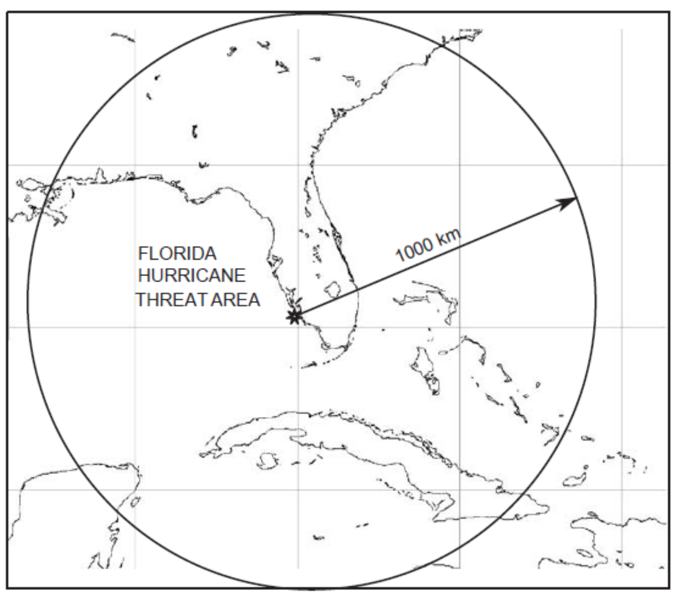


Figure 5. Florida Hurricane Threat Area

In order to obtain the number of hurricanes in each year from 1901 to 2010 I looked at each hurricane and its six hourly positions recorded by HURDAT. If hurricane entered threat area at any time during its track it had been counted so that any hurricanes could only be counted once. The results are presented in Table 1.

Table 1. Annual Number of Hurricanes

Year	Total hurricanes	Year	Total hurricanes	Year	Total hurricanes	Year	Total hurricanes	Year	Total hurricanes	Year	Total hurricanes
1901	1	1921	1	1941	1	1961	0	1981	0	2001	0
1902	0	1922	1	1942	0	1962	1	1982	1	2002	1
1903	1	1923	1	1943	0	1963	1	1983	0	2003	0
1904	1	1924	3	1944	2	1964	3	1984	1	2004	4
1905	0	1925	0	1945	2	1965	1	1985	3	2005	5
1906	3	1926	3	1946	1	1966	2	1986	0	2006	0
1907	0	1927	0	1947	2	1967	0	1987	1	2007	0
1908	0	1928	2	1948	2	1968	2	1988	0	2008	2
1909	2	1929	1	1949	1	1969	2	1989	0	2009	0
1910	1	1930	0	1950	5	1970	0	1990	0	2010	0
1911	2	1931	0	1951	1	1971	0	1991	2		
1912	1	1932	2	1952	1	1972	1	1992	1		
1913	0	1933	5	1953	2	1973	0	1993	0		
1914	0	1934	1	1954	2	1974	0	1994	0		
1915	2	1935	2	1955	0	1975	1	1995	3		
1916	3	1936	1	1956	2	1976	0	1996	2		
1917	1	1937	0	1957	0	1977	0	1997	1		
1918	0	1938	0	1958	1	1978	0	1998	2		
1919	1	1939	1	1959	2	1979	2	1999	3		
1920	1	1940	1	1960	1	1980	0	2000	1		

Historical data are retrieved and denoted by $X = \{x_i\}$ $(I = 1,2,\ldots,N)$, where N =110 is the number of years of data available and x_i is the number of storms occurred in the ith year. Values of x range in between 0 and 5 with mean 1.1091 and standard deviation 1.1704 (Table 2).

Table 2. Descriptive Statistics of Annual Occurrence Rate

Sample size (N)	110	Min	0
Mean	1.1091	Median	1
Variance	1.3699	Max	5
Std. deviation	1.1704	Range	5

Each storm is considered as a point event in time, occurring independently. If λ is a measure of the historically determined number of events per year, then the probability

P(X=x|λ) defines the probability of having x events per year, which is given by the Poisson probability density function (PDF)

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}.$$

The parameter of the Poisson distribution λ can be estimated from data by the maximum likelihood estimator

$$\hat{\lambda} = \frac{\sum_{i=1}^{N} x_i}{N}.$$

The Negative Binomial distribution PDF is given by

$$P(x) = \frac{\Gamma(x+k)}{\Gamma(x+1)\Gamma(k)} \left(\frac{k}{m+k}\right)^k \left(\frac{m}{m+k}\right)^x,$$

where Γ is the gamma function, m and k are parameters of the distribution. The maximum likelihood estimates of parameters can be obtained as

$$\hat{m} = \frac{\sum_{i=1}^{N} x_i}{N} \text{ and } \hat{k} = \frac{\hat{m}^2}{s^2 - \hat{m}},$$

where s^2 is the sample variance.

The parameters of both Poisson and Negative Binomial distributions were estimated using annual number of hurricanes dataset and results are presented in Table 3.

Table 3. Estimated distribution parameters for *AHO* data

Year	Total hurricanes
Poisson	λ = 1.1091
Negative Binomial	n = 4, p = 0.8096

After the estimation of parameters of both Poisson and Negative Binomial distributions, goodness-of-fit tests are performed to select the best fitting model.

The Kolmogorov-Smirnov test is used to decide if a sample comes from a hypothesized continuous distribution. It is derived from the empirical cumulative distribution function (CDF). Assume that we have a random sample $x_1, x_2, \ldots, x_n$ from some distribution with CDF $F(x)$. The empirical CDF is denoted by

$$F_n(x) = \frac{Number\ of\ observations\ \leq x}{n}.$$

The Kolmogorov-Smirnov statistic (D) is derived from the largest vertical difference between the theoretical ($F(x_i)$) and the empirical cumulative distribution function:

$$D = \max_{1 \leq i \leq n} \left(F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right).$$

The Kolmogorov-Smirnov statistic is thus only concerned with the maximum vertical distance between the cumulative distribution function of the fitted distribution and the cumulative distribution of the data. The Kolmogorov-Smirnov statistic's value is only determined by the one largest discrepancy and takes no account of the lack of fit across the rest of the distribution.

The null and the alternative hypotheses are: H_0: the data follow the specified distribution vs. H_A: the data do not follow the specified distribution. The P-value is calculated from the test statistic, and denotes the threshold value of the significance level in the sense that the null hypothesis (H_0) will be accepted for all values of α less than the P-value.

The Anderson-Darling test compares the fit of an observed cumulative distribution function to an expected cumulative distribution function. The A-D test gives more weight to the tails than the Kolmogorov-Smirnov test.

The Anderson-Darling statistic (A^2) is defined as

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1) \times \left[\ln F(x_i) + \ln\big(1 - F(x_{n-i+1})\big)\right].$$

The Chi-Square (χ^2) goodness-of-fit test measures how well the expected frequency of the fitted distribution compares with the observed frequency of a histogram of the observed data.

The Chi Square statistic is calculated as follows:

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i},$$

where O_i is the observed frequency of the ith histogram class or bar and E_i is the expected frequency from the fitted distribution for the ith histogram bar.

Since the χ^2 statistic sums the squares of all of the errors it can be disproportionately sensitive to any large errors. The χ^2 statistic is also very dependent on the number of bars N that are used and by changing the number of bars one can quite easily switch ranking between two distribution types.

Table 4. Goodness-of-fit tests for *AHO* data

Distribution	Chi-Squared			Kolmogorov-Smirnov		Anderson-Darling	
	Statistic	P-value	Rank	Statistic	Rank	Statistic	Rank
Poisson	1.71979	0.88640	1	0.32986	1	16.465	1
Neg. Binomial	2.83815	0.58527	2	0.42963	2	28.094	2

The results of goodness-of-fit tests for both Poisson and Negative Binomial distributions are presented in Table 4 and according to all 3 tests, the Poisson distribution is showing a better fit than Negative Binomial.

For visual assessment and an empirical comparison of the goodness of fit the distribution graphs can be used. Figure 6 shows the occurrence rates of historical and modeled hurricane data. Poisson model does appear to have a better agreement with historical occurrences than the Negative Binomial.

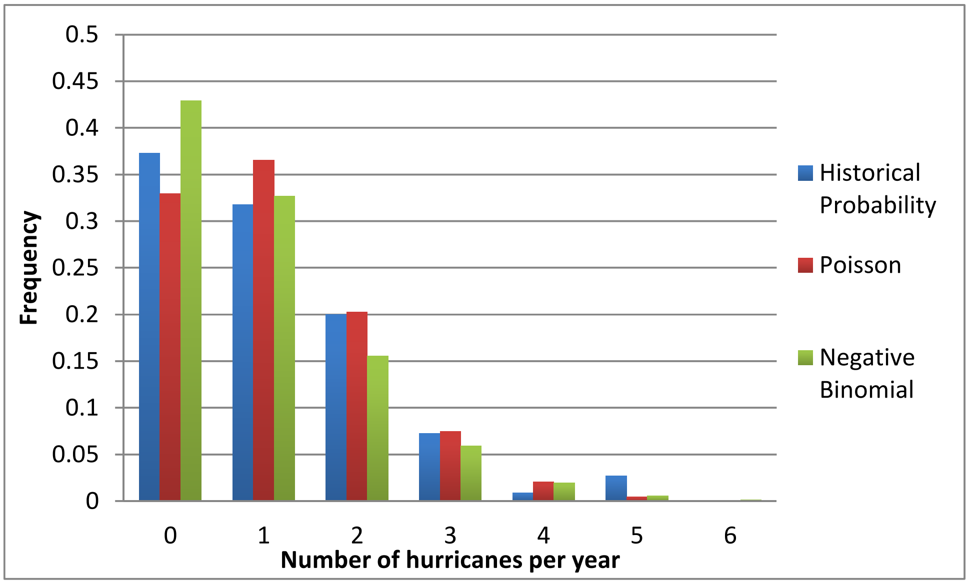


Figure 6. Comparison of simulated vs. historical occurrences

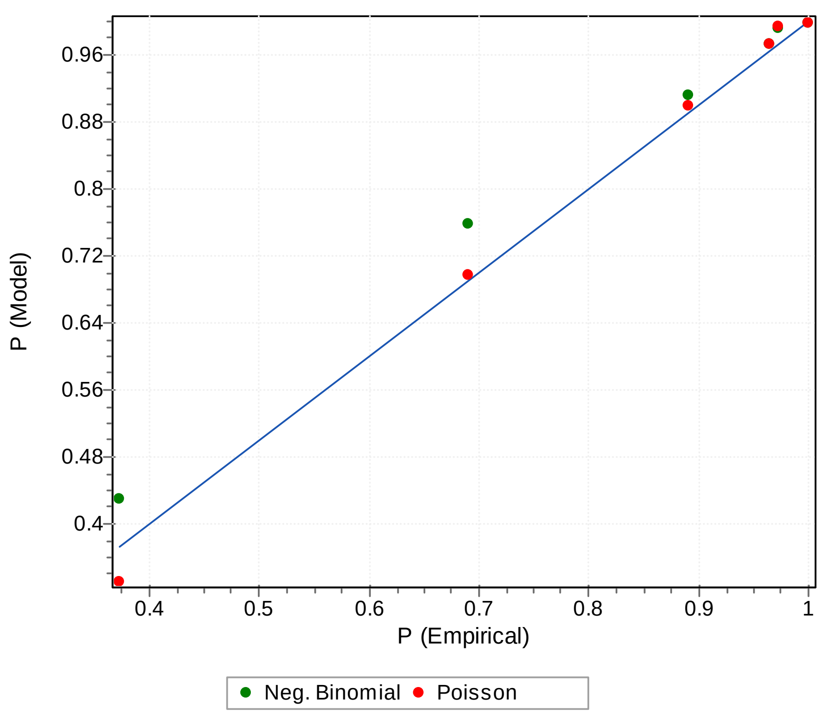


Figure 7. P-P plot

In order to see how well Poisson and Negative Binomial Distributions fit AHO data we can also look at the P-P plot (Figure 7), which is a graph of the empirical CDF values plotted against the fitted CDF values and the closer to linear it is, the better the distribution fits the data. Points of the Poisson distribution are closer to the straight line than the Negative Binomial which means that the Poisson distribution is the better choice for AHO model. This is consistent with the goodness-of-fit tests.

I conclude that the best fitting distribution for the annual hurricane occurrence on the basis of the results of goodness-of-fit tests, histogram of historical and modeled occurrences and P-P plot is Poisson distribution with parameter λ=1.1091.

IV. RADIUS OF MAXIMUM WINDS

The next part of the atmospheric science component of the hurricane model is the wind field model. Here hurricanes are simulated using historical data in order to record wind speeds and decay of the storm once on land. Recorded data from the wind field model are later used by engineers and actuarial scientists in assessment of the likely damage to insured property and losses associated with it.

Radius of maximum winds (R_{max}) is one of the random variables used to characterize the wind field. The radius of maximum winds at landfall is the distance between the center of a cyclone and its band of strongest winds. I am going to look closely at R_{max} and select a statistical distribution that is best for describing R_{max}.

The statistical information used to develop an R_{max} model (landfall R_{max} database) is created using the historical record for the Atlantic tropical cyclone basin (known as "HURDAT") and applying the annual occurrence model and the storm track model. The database includes 112 measurements of radius of maximum wind, central pressure and location at landfall for storms from 1901 till 2010 (Appendix 1).

Values of R_{max}, measured in statue miles, range in between 5.75 and 52.9 with mean 25.65 and standard deviation 11.2 (Table 5).

Table 5. Descriptive Statistics of Radius of Maximum Winds

Sample size	112	Min	5.75
Mean	25.649	Median	24.725
Variance	125.31	Max	52.9
Std. deviation	11.194	Range	47.15

There are numerous probability distributions each developed to address various data analysis needs, therefore the candidate distributions to fit should be chosen according to the nature of the data. The R_{max} dataset is continuous. Another way to classify the distributions considers the range. Radius of maximum winds cannot contain negative values so only non-negative distributions should be considered.

Another way of identifying the proper distribution is by looking at the histogram of the data and determining whether the data are symmetric, left-skewed, right-skewed and using the distributions which have the same shape. According to the histogram the R_{max} data are right-skewed (Figure 8).

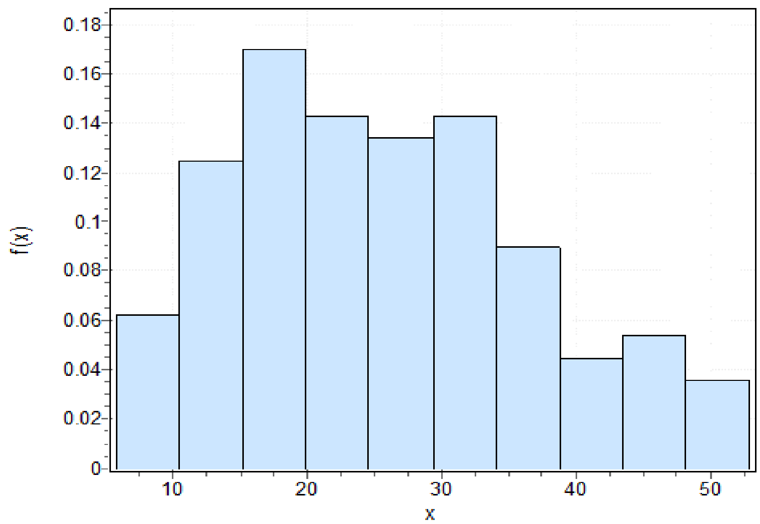


Figure 8. Probability Density Function of Radius of Maximum Winds

Using EasyFit software I have done preliminary analysis of the R_{max} landfall database on the basis of its semiboundness and skewness. Only distributions with maximum of 2 parameters were considered because extra parameters will make the use for the wind field model over complicated and not practical. Also distributions with more parameters may well fit the data better because of a lot more flexibility in shape than a 2-parameter and apparent improvement may be spurious due to over-fitting.

Five distributions that were found to be a good fit for modeling R_{max} on the basis of the provided criteria: Gamma, Lognormal, Rayleigh, Weibull and Inverse Gaussian. Gamma and Lognormal are the distributions that were considered in the Florida Public Hurricane Loss Model and Gamma was chosen as the best fit. Probability density functions of selected distributions are presented in the table (Table 6).

Table 6. Probability density functions of distributions to be fitted to R_{max} data

Gamma	$f(x) = \frac{x^{\alpha-1}}{\beta^{\alpha}\,\Gamma(\alpha)}\, e^{-\frac{x}{\beta}}$	$0 \le x \le +\infty$ $\alpha > 0, \quad \beta > 0$
Lognormal	$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\delta}\right)^2}}{x\,\delta\,\sqrt{2\pi}}$	$0 \le x \le +\infty$ $\delta > 0$
Rayleigh	$f(x) = \frac{x-\gamma}{\delta^2}\, e^{-\frac{1}{2}\left(\frac{x-\gamma}{\delta}\right)^2}$	$\gamma \le x \le +\infty$ $\delta > 0$
Weibull	$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$	$0 \le x \le +\infty$ $\alpha > 0, \quad \beta > 0$
Inverse Gaussian	$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}}\, e^{\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)}$	$0 \le x \le +\infty$ $\lambda > 0, \quad \mu > 0$

Parameters of selected distributions were obtained using maximum likelihood estimators and results are presented in the table (Table 7).

Table 7. Estimated distribution parameters for R_{max} data

Distribution	Parameters
Gamma	α=5.2501, β=4.8855
Lognormal	δ=0.49202, μ=3.1363
Weibull	α=2.4736, β=28.666
Rayleigh	δ=17.293, γ=3.8794
Inverse Gaussian	λ=134.66, μ=25.65

In order to determine how well the selected distributions fit the R_{max} data I have tested them for a goodness-of-fit and the results are presented in Table 8.

Table 8. Goodness-of-fit tests for R_{max} data

Distribution	Kolmogorov-Smirnov			Anderson-Darling	
	Statistic	P-value	Rank	Statistic	Rank
Weibull	0.04939	0.93492	1	0.32264	2
Rayleigh	0.05608	0.85301	2	0.30063	1
Gamma	0.07027	0.61237	3	0.5349	3
Lognormal	0.09036	0.30146	4	1.0419	4
Inverse Gaussian	0.0953	0.24495	5	1.8773	5

The idea behind the goodness-of-fit tests is to measure the distance between the data and the tested distribution. And although the logic of applying various goodness-of-fit tests is the same, they differ in how the test statistic is calculated. The most commonly used goodness of fit tests are Kolmogorov-Smirnov, Anderson-Darling and Chi-Square. The two goodness-of-fit tests that were used are Kolmogorov-Smirnov and Anderson-Darling. The chi-square test is not considered because the test has low power for continuous data.

The Kolmogorov-Smirnov test was used to arrange the distributions in the order of performance according to that test. Since the goodness-of-fit test statistics indicate the distance between the data and the fitted distributions, it is obvious that the distribution with the lowest statistic value is the best fitting.

Lognormal and Inverse Gaussian distributions show poor fit for R_{max} data with P-values of Kolmogorov-Smirnov test below 0.5. Other distributions show better fits according to both Kolmogorov-Smirnov and Anderson-Darling tests. I conclude that Lognormal and Inverse Gaussian distributions are not good fits and exclude them from further consideration.

The three distributions for be considered further are Weibull, Rayleigh and Gamma. Gamma distribution was used to fit the radius of maximum winds in the Florida Public Hurricane Loss Evaluation Model, however we see that other distributions perform better than the Gamma distribution.

Along with the goodness of fit tests, the distribution graphs can be very helpful to determine the best fitting model. They enable us to visually assess the goodness of fit and empirically compare several fitted models.

First I consider the Probability Density Function Graph which displays the theoretical PDFs of the fitted distributions and the histogram of the R_{max} data (Figures 9 and 10). Since the histogram depends on how the data are sorted into bins, two histograms are displayed with the R_{max} values binned in 10 and 15 intervals for comparative analysis. All 3 distributions are plotted on the same graphs. Displaying several distributions at the same time will allow us to visually compare the models and determine how they differ.

Although it can be difficult to come to a decision about better fit on the basis of these graphs as they require the arbitrary grouping of the data, the Weibull and Rayleigh distributions do appear to fit data better that the Gamma distribution.

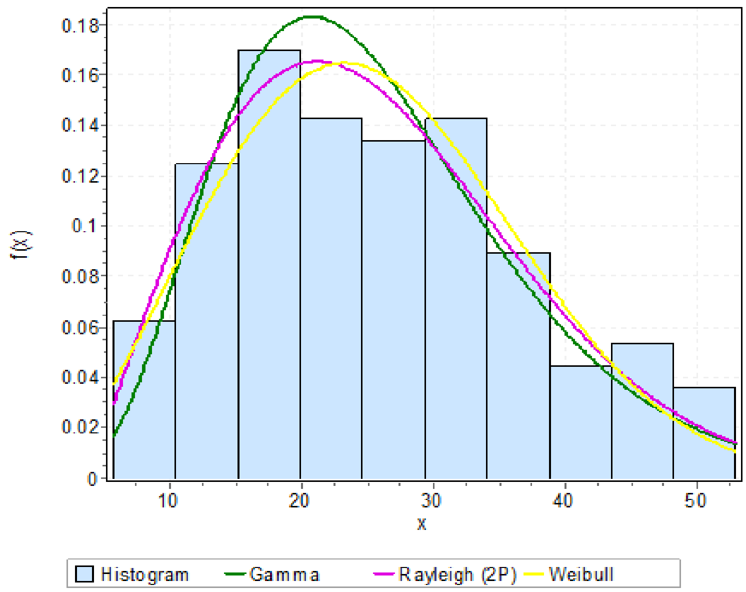


Figure 9. PDF Graph with Rmax values binned in 10 intervals

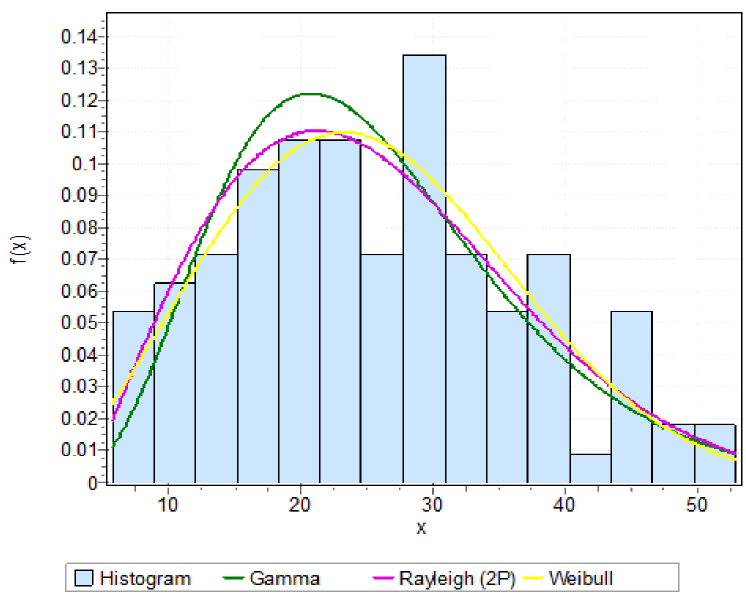
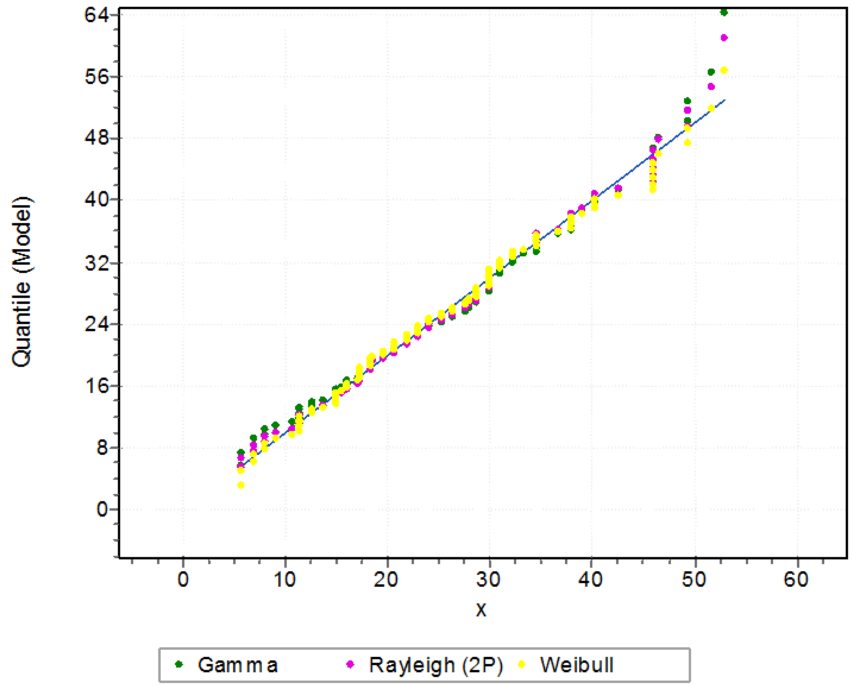


Figure 10. PDF Graph with Rmax values binned in 15 intervals

To avoid grouping of the data we can look at the Q-Q plot (Figure 11). In the quantile-quantile graph the input data values are plotted against the quantiles of the fitted distribution and both axes of this graph are in statue miles - units of the Rmax. Weibull, Rayleigh and Gamma distributions are plotted on the same plot.

Figure 11. Q-Q plot



If the distribution is the correct model, the graph points will lie on an approximately straight line. All 3 distributions have Q-Q plots that make us believe that they are good fits but Gamma and Rayleigh distributions have points further away from the straight line as values of R_{max} get larger. This is consistent with the results of the Kolmogorov-Smirnov test.

On the basis of the results of Goodness-of-fit test, the Probability Density Function Graph and the Q-Q plot, the Weibull distribution with parameters α=2.4736 and β=28.666 is the best fit for the Radius of maximum winds.

Gamma and Weibull distributions are commonly encountered in reliability analysis and it is often difficult to choose between the two. Nevertheless, as explained by Bain and Engelhardt (1980), "even though the two models may offer similar data fits even for moderate sample sizes, it is still desirable to select the correct (or more nearly correct) model, if possible, since inferences based on the model will often involve tail probabilities where the affect of the model assumption will be more critical”.

Although the Gamma distribution cannot be rejected for modeling R_{max} in the wind field model, I show that the Weibull distribution is a better fit for the radius of maximum winds.

V. HOLLAND B

Another important parameter of the wind field model is the Holland B parameter. Holland B is an additional parameter defining the pressure field and maximum wind speeds in a hurricane. It was introduced by Holland in 1980 and since been used in hurricane threat studies by many researchers including Powell et al. (2005), James and Mason (2005), Emanuel et al. (2006), Lee and Rosowsky (2007), Hall and Jewson (2008) and Vickery et al. (2009) among others.

The pressure $p(r)$ is defined as:

$$p(r) = p_c + \Delta p \, e^{-\left(\frac{R_{max}}{r}\right)^B},$$

where r is the distance from the center of the storm, p_c is the pressure at the center of the storm, Δp is the difference between central minimum sea level pressure (p_c) and an outer peripheral pressure (1013 mb), and R_{max} is the radius of maximum winds.

Introduction of B parameter results in the maximum wind speed in the simulated hurricane be proportional to $\sqrt{B\Delta p}$ compared to $\sqrt{\Delta p}$ without the Holland B.

A model for the Holland B pressure profile parameter will be developed using a subset of the data published by Willoughby and Rahn (2004). Data consist of winds and geopotential heights obtained by NOAA and U.S. Air Force Reserve aircraft between 1977 and 2000 and supplemented with Δp pressure deficit and R_{max} values. We retained 116 profiles with latitudes 20°-34°N, longitudes 70°-95°W, flight level winds V_{max} >30m/s and values of B 0.5-2.2 (Appendix 2).

Least squares fits of the Holland B model to the data will offer assessment of the parameters' distributions.

The FHPLM considers 2 models: in first model Holland B is correlated with the radius of maximum winds (R_{max}) and latitude of the hurricane (Lat)

$$B = \beta_0 + \beta_1\, Lat + \beta_2\, R_{max} + \varepsilon,$$

in the second model, the B parameter is also correlated with Δp^2 the square difference between central minimum sea level pressure and an outer peripheral pressure (1013 mb)

$$B = \beta_0 + \beta_1\, Lat + \beta_2\, R_{max} + \beta_3\, \Delta p^2 + \varepsilon.$$

Multiple regression analysis was performed on the dataset using the Proc REG procedure in SAS (Appendix 3A). On the basis of the least squares parameter estimates (Appendix 4A) the model for estimation of Holland B using the radius of maximum winds and latitude of the hurricane is:

$$\hat{B} = 1.55384 + 0.00015058\, Lat - 0.00439\, R_{max}.$$

Testing for significance of the regression equation using ANOVA (Appendix 4A) showed F=7.14 with P-value=0.0012, which means this regression is significant.

Coefficient of determination $R^2 = 0.1121$, which means that only 11.21% of the total variability in the B parameter is explained by the fitted equation.

Similar regression analysis was performed on the model for estimation of Holland B using the radius of maximum winds, latitude of the hurricane and the square difference between central minimum sea level pressure and an outer peripheral pressure using the Proc REG procedure in SAS (Appendix 3B). On the basis of the least squares parameter estimates (Appendix 4B) the new model is:

$$\hat{B} = 1.50264 + 0.00086116\, Lat - 0.00423\, R_{max} + 0.00000885\, \Delta p^2.$$

Testing for significance of the regression equation using ANOVA (Appendix 4B) showed F=4.95 with P-value=0.0029, which means this regression is significant.

Coefficient of determination $R^2 = 0.1171$, which means that only 11.71% of the total variability in the *B* parameter is explained by the fitted equation. Model including Δp^2 has a slightly higher coefficient of determination but still does not explain most of variability in Holland *B*.

In order to obtain a model which can explain a larger portion of variability in Holland *B* parameter we can include all available predictor variables in the regression model:

$$B = \beta_0 + \beta_1\, Lat + \beta_2\, Lon + \beta_3\, R_{max} + \beta_4\, \Delta p^2 + \beta_5\, V_{max} + \varepsilon,$$

where *Lat* is the latitude of the hurricane, *Lon* is the longitude of the hurricane, R_{max} is the radius of maximum winds, Δp^2 is the square difference between central minimum sea level pressure and an outer peripheral pressure (1013 mb), V_{max} is the maximum wind and ε is the error term.

Using SAS Proc REG procedure (Appendix 3C) multiple regression analysis was performed. On the basis of the least squares parameter estimates (Appendix 4C) the model for estimation Holland *B* using all available predictor variables is:

$$\hat{B} = 0.25712 - 0.001\, Lat - 0.00308\, Lon - 0.00157\, R_{max} - 0.00006305\, \Delta p^2 + 0.02652\, V_{max}.$$

Testing for significance of the regression equation using ANOVA (Appendix 4C) showed F=13.17 with P-value<0.0001, which means this regression is significant.

Coefficient of determination in this case $R^2 = 0.3745$, which is significantly higher than the previous two models and means that 37.45% of the total variability in the *B* parameter is explained by the fitted equation.

Although this model is good it is not convenient. The model includes terms with large individual t-test p-values: P-value (*Lat)* = 0.877 and P-value (*Lon)* = 0.337 (Appendix 4C). This suggests that perhaps the model is more complicated than it needs to be and includes some redundant terms. We should check if it can be reduced. In order to simplify the model a stepwise procedure on the basis of the partial F-value was chosen and performed on the dataset using the Proc Stepwise in SAS (Appendix 3C). Obtained results (Table 9) suggest the following optimal reduced model:

$$\hat{B} = \beta_0 + \beta_1 R_{max} + \beta_2 \Delta p^2 + \beta_3 V_{max} + \varepsilon.$$

Table 9. Results of the stepwise procedure

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	vmax		1	0.2277	0.2277	23.8152	33.60	<.0001
2	delp2		2	0.1249	0.3525	3.8546	21.80	<.0001
3	rmax		3	0.0163	0.3689	2.9842	2.90	0.0915

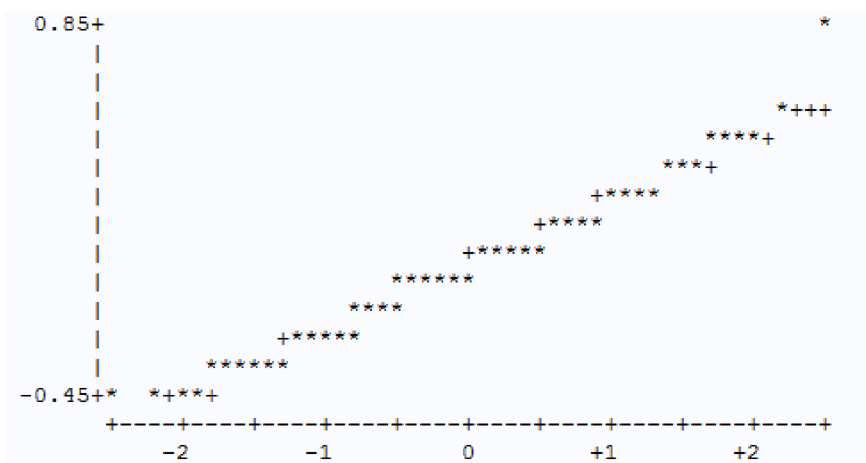
On the basis of the least squares parameter estimates obtained using SAS multiple regression analysis (Appendix 3D, 4D) the model for estimation Holland *B* using R_{max} the radius of maximum winds, Δp^2 the square difference between central minimum sea level pressure and an outer peripheral pressure (1013 mb) and V_{max} the maximum wind is:

$$\hat{B} = 0.50274 - 0.00181\, R_{max} - 0.00006228\, \Delta p^2 + 0.02612\, V_{max}$$

Removing two indicators from the model increased the value of adjusted R_A^2 to 0.3520 compared to 0.3460 for the model with five predictors. Unlike R^2, the adjusted R_A^2 increases only if the new term improves the model more than would be expected by chance. Compared to the full model coefficient of determination R^2 was reduces from 0.3745 to 0.3689, which is still significantly higher than the first two models.

Throughout my analysis I have assumed that the errors are normally and independently distributed with mean zero and constant variance σ^2 as well as that the observations are adequately described by the model. Residual analysis is the key tool in model adequacy checking. The most effective method of checking the normality assumption is constructing a normal probability plot of the residuals. If the errors are normally distributed this plot should resemble a straight line. While investigating this plot the focus should be on the central values of the plot rather than the extremes. The normal probability plot of the residuals for Holland *B* with R_{max}, Δp^2 and V_{max} as predictors (Figure 12) resembles a straight line with all values being in (-2.7,2.7) z- range.

Figure 12. Normal Probability Plot



Plot of residuals vs. predicted values (Figure 13) does not reveal any obvious patterns. Data are scattered randomly around 0. This supports the assumption that the error distribution for Holland *B* with R_{max}, Δp^2 and V_{max} as predictors is approximately normal.

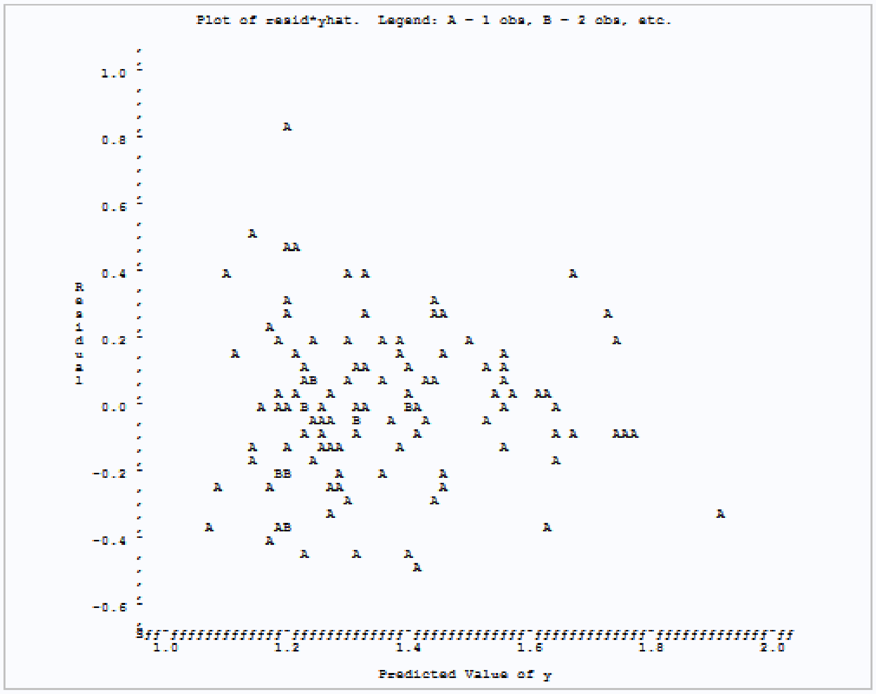


Figure 13. Plot of Residuals vs. Predicted values

Four models for Holland B parameter were considered. First two models, both used in the FHPLM, could only explain 11.21% and 11.71% of the total variability in the B parameter. The first model used the radius of maximum winds and the latitude of the hurricane and predictor variables; the second model also considered the square difference between central minimum sea level pressure and an outer peripheral pressure. The third model explained 37.45% of the total variability in Holland B but included five predictor variables. The fourth model was chosen to be the most optimal for use in predicting the Holland B parameter. It explains 36.89% of the total variability in the predicted parameter and correlated Holland B with the radius of maximum winds, the square difference between central minimum sea level pressure and an outer peripheral pressure and the maximum wind:

$$\hat{B} = 0.50274 - 0.00181\, R_{max} - 0.00006228\, \Delta p^2 + 0.02612\, V_{max}.$$

VI. FINAL RESULTS AND CONCLUSIONS

The FPHLM is the only open public hurricane loss evaluation model available for assessment of hazard to insured residential property related to damage from hurricanes in Florida. Atmospheric science component is the first part of this model; it simulates thousands of storms, their wind speeds and their decay once on land on the basis of historical hurricane statistics defining wind risk for all residential zip codes in Florida.

The focus of my thesis was to analyze the atmospheric science component of the Florida Public Hurricane Loss Model, replicate statistical procedures used to model various parameters of atmospheric component and to validate the model.

First I looked at the available data from the time point prospective and checked for increasing trends in hurricane intensity, size or number of hurricanes striking Florida. The time series of annual hurricane counts, the plot of radius of maximum winds in time or the plot of central pressure in time show no visible patterns in the data and nothing suggests an increasing trend. My findings do not support the argument that global tropical cyclone intensity, frequency and longevity have undergone increases in the recent years. I concluded that no significant increasing trend is evident.

Next I studied the frequency with which hurricanes occur and generated statistical distribution from the available historical data in order to estimate annual hurricane occurrence. Two distributions were considered: Poisson and Negative Binomial. On the basis of the results of goodness-of-fit tests, histograms of historical and modeled occurrences and P-P plots, I concluded that the best fitting distribution for the annual hurricane occurrence is the Poisson distribution with parameter $\lambda=1.1091$.

Further I modeled the distribution of radius of maximum winds which is a critical parameter for estimating the possible losses for insurance pricing purposes. The radius of maximum winds has a substantial impact on the area affected by hurricane and modeling of the R_{max} influences the likelihood of the location experiencing strong winds in cases of near misses. Five distributions were considered: Gamma, Lognormal, Rayleigh, Weibull and Inverse Gaussian. On the basis of the results of the Goodness-of-fit test, Probability Density Function Graph and the Q-Q plot, the Weibull distribution with parameters α=2.4736 and β=28.666 was chosen as the best fit for the Radius of maximum winds. The FPHLM currently uses Gamma distribution for modeling radius of maximum winds and although the Gamma distribution cannot be rejected for modeling R_{max} in the wind field model, I showed that the Weibull distribution is better fit.

Finally, the expression for finding an exact statistical relationship between the pressure profile parameter Holland and its' predictor variables was computed. Four models for the Holland B parameter were considered. The first two models, both considered in the FHPLM, could only explain 11.21% and 11.71% of the total variability in the B parameter. The third model explained 37.45% of the total variability in Holland B but included five predictor variables. The fourth model was chosen to be the most optimal to use in predicting Holland B parameter. It explains 36.89% of the total variability in the predicted parameter and correlated Holland B with the radius of maximum winds, the square difference between central minimum sea level pressure and an outer peripheral pressure and the maximum wind: $\hat{B} = 0.50274 - 0.00181\, R_{max} - 0.00006228\, \Delta p^2 + 0.02612\, V_{max}$.

REFERENCES

L.J. Bain, M. Engelhardt, "Probability of correct selection of Weibull versus gamma based on likelihood ratio", *Communications in Statistics – Theory and Methods*, A9, p. 375-381, 1980.

S.-C. Chen, S. Gulati, S. Hamid, X. Huang, L. Luo, N. Morisseau-Leroy, M. Powell, C. Zhan, C. Zhang, "A three-tier system architecture design and development for hurricane occurrence simulation", *Proc. of the IEEE Int. Conference on Information Technology: Research and Education*, 2003.

S.-C. Chen, S. Gulati, S. Hamid, X. Huang, L. Luo, N. Morisseau-Leroy, M. Powell, C. Zhan, C. Zhang, "A Web-based distributed system for hurricane occurrence projection", *Software – Practice and Experience*, vol.34, 2004.

J.B. Elsner, S.C. Schmertmann, "Improving Extended-range Seasonal Predictions of Intense Atlantic Hurricane Activity", *Weather Forecasting,* vol. 8, p. 345-351, 1993.

J.B. Elsner, T.H. Jagger, "A Hierarchical Bayesian Approach to Seasonal Hurricane Modeling", *J. Climate,* vol. 17, p. 2813-2827, 2004.

K.A. Emanuel, "Increasing destructiveness of tropical cyclones over the past 30 years", *Nature*, vol. 326, p. 686-688, 2005.

K.A. Emanuel, S. Ravela, E. Vivant, C. Risi, "A Statistical-Deterministic Approach to Hurricane Risk Assessment", *Bulletin of the American Meteorological Society,* vol. 19, p. 299-314, 2006.

W.M. Gray, C.W. Landsea, P.W. Mielke Jr., K.J. Berry, "Predicting Atlantic Seasonal Hurricane Activity 6-11 Months in Advance", *Weather Forecasting,* vol. 7, p. 440-455, 1992.

T. Hall, S. Jewson, "Comparison of Local and Basin-wide Methods for Risk Assessment of Tropical Cyclone Landfall", *Journal of Applied Meteorology and Climatology*, vol. 47, p. 361-367, 2008.

S. Hamid, G. Kibria, S. Gulati, M. Powell, B. Annane, S. Cocke, J. Pinelli, K. Gurley, S. Chen, "Predicting Losses of Residential Structures in the State of Florida by the Public Hurricane Loss Evaluation Models", *Statistical Methodology*, vol. 7, issue 5, p. 552-573, September 2005.

G.J. Holland, "An Analytic Model of the Wind and Pressure Profiles in Hurricanes", *Monthly Weather Review,* vol.108, 1980.

Z. Huang, D.V. Rosowsky, P.R. Sparks, "Hurricane simulation techniques for the evaluation of wind-speeds and expected insurance losses", *Journal of Wind Engineering and Industrial Aerodynamics,* 89, 2001.

HURDAT data, *http://www.nhc.noaa.gov/data/hurdat/hurdat2-atlantic-1851-2012-060513.txt*

HURDAT easy to read data, *http://www.aoml.noaa.gov/hrd/hurdat/easyread-2012.html*

M.K. James, L.B. Mason, "Synthetic Tropical Cyclone Database", *Journal of Waterway, Port, Coastal, and Ocean Engineering,* vol. 131, p. 181-192, 2005.

R. W. Katz, "Discussion on "Predicting losses of residential structures in the state of Florida by the public hurricane loss evaluation model" by S. Hamid et al.", *Statistical Methodology*, vol.7, p. 592-595, 2010.

P.J. Klotzbach, "Trends in global tropical cyclone activity over the past twenty years (1986-2005)", *Geophysical Research Letters,* vol. 33, 2006.

K.H. Lee, D.V. Rosowsky, "Synthetic Hurricane Wind Speed Records: Development of Database for Hazard Analyses and Risk Studies", *National Hazards Review,* vol. 8, p. 23-34, 2007.

R. Lokupitiya, L. Borgman, R. Anderson-Sprecher, "Simulation of Storm Occurrences Using Simulated Annealing", *Journal of Climate*, vol. 18, 2005.

M. Powell, G. Soukup, S. Cocke, S. Gulati, N. Morisseau-Leroy, S. Hamid, N. Dorst, L. Axe, "State of Florida hurricane projection model: Atmospheric science component", *Journal of Wind Engineering and Industrial Aerodynamics,* 93, p. 651-674, 2005.

L.R. Russell, "Probability distribution for Texas gulf coast hurricane effects of engineering interest", Ph.D. Thesis, Stanford University, 1968.

L.R. Russell, "Probability distributions for hurricane effects", *Journal of Waterways, Harbors, and Coastal Engineering Division*, 97, p.139-154, 1971.

A.H. El-Shaarawi, S.Chenouri, "Discussion on "Public Hurricane Loss Evaluation Models: Predicting losses of residential structures in the state of Florida" by S. Hamid et al..", *Statistical Methodology*, 7 (5), p. 583-586, September 2010.

L.J. Shapiro, S.B. Goldenberg, "Atlantic sea surface temperatures and tropical cyclone formation", *Journal of Climate*, vol. 11, p. 578-590, 1998.

H.C.S. Thom, "Some Methods of Climatological Analysis", *Technical Note*, vol. 81, p. 30-34, 1966.

V.J. Tryggvason, A.G. Davenport, D. Surry, "Predicting Wind-Induced Response in Hurricane Zones", *Journal of the Structural Division*, 102, p.2333-2350, 1976.

P.J. Vickery, F. Masters, M. Powell, D. Wadhera, “Hurricane hazard modeling: The past, present, and future”, *Journal of Wind Engineering and Industrial Aerodynamics*, 2009.

P.J. Vickery, D. Wadhera, “Statistical Models of Holland Pressure Profile Parameter and Radius to Maximum Winds of Hurricanes from Flight-Level Pressure and H*Wind Data”, *Journal of Applied Meteorology and Climatology*, vol. 47, p. 2497–2517, October 2008.

P.J. Webster, G.J. Holland, J.A. Curry, H.-R. Chang, “Changes in tropical cyclone number and intensity in a warming environment”, *Science,* vol. 309, p. 1844-1846, 2005.

H.E. Willoughby, M.E. Rahn, “Parametric Representation of the Primary Hurricane Vortex. Part I: Observations and Evaluation of the Holland (1980) Model”, *American Meteorological Society*, p. 3033-3048, December 2004.

APPENDICES

Appendix 1. Landfall R_{max} database

Year	Name	Rmax (sm)	Po (mb)	LFLat	LFLon
1901	No_name	37.95	972.6	30.4	88.8
1903	No_name	49.45	976.6	26.1	80.1
1906	No_name	29.90	979	27.4	80.1
1906	No_name	29.90	979	25.1	81
1906	No_name	34.50	976.6	33.3	79.2
1906	No_name	40.25	976.6	26.4	80.1
1906	No_name	18.40	966.8	24.9	81
1906	No_name	49.45	965.1	30.4	88.7
1909	No_name	25.30	957	24.7	81
1910	No_name	36.80	953.3	26	81.7
1910	No_name	32.20	941.4	24.4	82.7
1911	No_name	31.05	979.3	32.2	80.6
1915	No_name	29.90	932.3	29.2	90
1916	No_name	21.85	973.9	30.3	87.5
1916	No_name	29.90	950.2	30.4	88.3
1917	No_name	37.95	964.4	30.4	86.7
1919	No_name	17.25	929.2	24.6	82.9
1921	No_name	20.70	960	27.9	82.8
1924	No_name	24.15	978.3	25.5	81.7
1924	No_name	21.85	971.9	24.6	82.9
1926	No_name	16.10	959.7	29.9	81.3
1926	No_name	19.55	955	30.3	87.5
1926	No_name	27.60	950	26.4	81.9
1926	No_name	24.15	931.9	23.9	80.4
1926	No_name	21.85	931	25.6	80.3
1928	No_name	32.20	935.3	26.7	80
1929	No_name	32.20	948.2	25	80.5
1933	No_name	14.95	947.5	26.9	80.1
1935	No_name	11.50	977	25.2	81.1
1935	No_name	11.50	972.9	25.9	80.1
1936	No_name	21.85	963.8	30.4	86.4
1940	No_name	31.05	974.6	32.1	80.8
1941	No_name	20.70	981.4	29.8	84.7
1944	No_name	33.35	948.9	24.6	82.8
1945	No_name	13.80	951.2	25.3	80.3
1947	No_name	14.95	968.2	31.9	81.1
1947	No_name	26.45	966.5	29.6	89.5
1947	No_name	29.90	960	26.3	81.8
1947	No_name	29.90	946.8	26.3	80.1
1948	No_name	18.40	977	25.9	80.1
1948	No_name	18.40	963.4	27.2	80.2
1948	No_name	14.95	962.7	24.8	81
1948	No_name	18.40	950.9	25.9	81.7

Year	Name	Rmax (sm)	Po (mb)	LFLat	LFLon
1948	No_name	8.05	935.3	24.6	81.7
1949	No_name	26.45	953.6	26.9	80
1950	Baker	24.15	979.3	30.2	88.1
1950	Easy	17.25	958.3	28.6	82.7
1950	King	6.90	955	26.1	80.1
1956	Flossy	20.70	973.9	30.4	86.4
1958	Helene	28.75	932	32.7	78.7
1959	Gracie	29.90	950.9	32.5	80.4
1960	Ethel	25.30	976	30.3	89.3
1960	Donna	27.60	970	29.5	81.1
1960	Donna	20.70	930	24.8	80.9
1964	Isbell	14.95	977.7	26.9	80
1964	Cleo	8.05	967.5	25.7	80.2
1964	Isbell	11.50	964.1	25.8	81.3
1964	Dora	39.10	961	29.9	81.3
1965	Betsy	25.30	951.9	25	80.5
1966	Alma	23.00	977	30.1	84.2
1966	Inez	17.25	977	24.1	84.1
1966	Alma	17.25	970.2	24.6	82.9
1968	Gladys	19.55	977	28.6	82.7
1972	Agnes	23.00	978	29.9	85.4
1975	Eloise	16.10	955	30.3	86.5
1976	Belle	28.75	963.1	32.5	75.2
1979	David	11.50	968	31.6	81.2
1979	David	31.05	968	27.1	80.1
1979	Frederic	37.95	946	30.4	88.3
1985	Bob	27.95	1003	32.2	80.5
1985	Elena	46.58	971	28.8	83.8
1985	Kate	27.95	967	30	85.4
1985	Elena	17.20	959	30.4	89.2
1985	Elena	18.63	954	29.4	85.9
1987	Floyd	15.53	993	24.8	81
1988	FLORENCE	28.75	983	28.7	89.3
1989	HUGO	40.25	935	31.7	78.8
1991	BOB	34.50	965	33	76.1
1992	ANDREW	17.25	947	25.8	83.1
1995	ERIN	46.00	988	29.6	83.4
1995	ERIN	23.00	985	30.6	87.5
1995	ERIN	46.00	985	27.7	80.4
1995	ERIN	23.00	974	29.8	86.6
1995	OPAL	28.75	938	29	87.7
1996	BERTHA	28.75	984	31.2	78.6
1996	FRAN	23.00	954	31	77.2
1997	DANNY	34.50	992	29.2	89.9
1997	DANNY	11.50	990	29.5	89.4
1997	DANNY	17.25	984	30.3	88
1998	GEORGES	34.50	982	23.9	81.3

Year	Name	Rmax (sm)	Po (mb)	LFLat	LFLon
1998	GEORGES	34.50	965	30.4	88.9
1998	BONNIE	46.00	958	30.8	76.4
1999	IRENE	46.00	982	27.8	80.1
1999	DENNIS	46.00	967	30.8	78.4
1999	FLOYD	40.25	947	30.6	79.1
2004	FRANCES	37.95	960	27.2	80.2
2004	JEANNE	51.75	953	27.3	80.2
2004	CHARLEY	5.75	947	26.56	82.29
2004	IVAN	26.45	943	30.2	87.8
2005	KATRINA	10.72	986	25.88	80.13
2005	RITA	12.65	976	23.75	82
2005	WILMA	52.90	951	25.92	81.58
2005	DENNIS	5.75	946	30.38	87.05
2005	DENNIS Cuba	9.19	937	22.1	80.6
2005	KATRINA	19.55	932	30.25	89.62
2007	Humberto	12.65	985	29.5	94.4
2008	Dolly	6.90	967	26.4	97.2
2008	Paloma_Cuba	14.95	970	20.7	78
2008	Gustav	29.90	941	22.4	83.1
2008	Ike	42.55	950	29.3	94.7
2010	Paula_Cuba	16.10	1000	22.7	83.9
2010	Earl	31.00	935	30.1	74.8

Appendix 2. Holland *B* database

	Hurricane Name	Pmin	B	Rmax	DelP^2	Vmax	Lat	Lon
1	LILI96_18H1	978	0.69	78	1225	30.02	21.83	-82.13
2	ERIN95_01U1	990	1.24	42.5	529	30.47	25.42	-76.32
3	GEORGES98_25I1	981	1.5	72.5	1024	30.59	23.43	-80.15
4	ROXANNE95_16U1	982	1.18	38	961	30.86	20.35	-92.11
5	FLORENCE88_09H	991	1.36	36	484	30.98	28.67	-89.34
6	DAVID79_02H	993	1.24	46.5	400	31.29	24.41	-78.03
7	ALICIA83_17I	982	1.21	36	961	31.3	27.68	-93.83
8	BONNIE98_22U1	990	1.16	30.5	529	31.32	22.69	-70.14
9	BOB91_18U1	979	0.78	44	1156	31.38	30.7	-76.82
10	DAVID79_02I	988	1.47	48.5	625	31.56	23.67	-77.04
11	DANNY85_15I1	993	1.69	58.5	400	31.73	29.61	-92.7
12	DANNY85_15H1	993	2.06	62	400	32.1	28.59	-92.33
13	DANIELLE98_30I	988	1.32	37.5	625	32.16	27.84	-74.16
14	FRAN96_03U1	976	1.28	76.5	1369	32.17	24.68	-70.65
15	DANIELLE98_31U3	976	1.01	32.5	1369	32.35	31.14	-73.23
16	DANIELLE98_30U2	988	1.2	32.5	625	32.46	28.13	-74.27
17	ERIN95_01U2	986	1.45	37.5	729	32.89	26.47	-78.09
18	GEORGES98_28U2	966	0.93	32.5	2209	33.29	30.45	-88.9
19	OPAL95_03U2	967	1.09	27	2116	33.75	24.13	-90.34
20	DAVID79_03F	974	1.33	38	1521	33.81	25.57	-79.25
21	ELENA85_1_30I	980	0.85	70	1089	33.9	27.47	-87.44
22	BERTHA96_09U2	963	0.87	78	2500	33.95	23.63	-72.5
23	DENNIS99_28U1	969	1.18	60.5	1936	34.01	27.01	-76.93
24	OPAL95_03U1A	969	1.53	42.5	1936	34.23	22.26	-92.2
25	ERIN95_02U2	987	1.51	33.5	676	34.44	29.15	-85.52
26	GEORGES98_25H1	981	1.31	66.5	1024	35.11	24.27	-81.81
27	GEORGES98_25U1	986	1.18	46.5	729	35.15	24.83	-83.15
28	DANIELLE98_31U1	982	1.27	28	961	35.19	29.68	-73.81
29	DANIELLE98_01U1	973	0.83	77.5	1600	35.55	32.38	-71.6
30	GEORGES98_26U1	974	1.24	46	1521	35.74	25.43	-84.41
31	GEORGES98_27U2	966	0.84	61	2209	35.92	29.03	-88.32
32	GEORGES98_26U3	975	1.09	38.5	1444	36.29	27.2	-86.79
33	GEORGES98_26U2	974	1.06	38	1521	36.5	26.32	-85.75
34	BERTHA96_12U2	979	1.15	74.5	1156	36.71	32.75	-78.1
35	BERTHA96_09U1	965	1.14	34	2304	37.24	21.95	-70.09
36	GEORGES98_28U1	962	1.2	40.5	2601	37.46	29.8	-88.73
37	ELENA85_2_31H	975	1.31	68	1444	37.55	28.77	-84.27
38	EMILY93_29U4	978	1.45	47.5	1225	37.59	31.43	-70.37
39	GEORGES98_27U1	970	1.02	42.5	1849	37.71	28.21	-87.73
40	ELENA85_2_31I	975	1.04	68	1444	37.72	28.78	-83.82
41	ERIN95_03U1	982	1.57	24.5	961	37.91	30.02	-86.77
42	BOB91_18U2	976	1.17	29.5	1369	38.01	32.99	-76.07
43	ELENA85_2_30I2	976	1.39	66.5	1369	38.24	28.52	-85.36
44	GLORIA85_3_26I	945	1	45.5	4624	38.69	29.26	-75.13
45	DENNIS99_29U2	971	1.44	58	1764	39.21	32.09	-78.04

	Hurricane Name	Pmin	B	Rmax	DelP^2	Vmax	Lat	Lon
46	FLOYD99_15U3	945	1.66	54.5	4624	39.25	33.59	-77.96
47	DENNIS99_28U3	979	1.62	75	1156	39.4	28.96	-77.93
48	GILBERT88_15I1	947	1.02	68.5	4356	39.53	21.97	-92.07
49	ALICIA83_17H	982	1.43	32.5	961	39.7	27.97	-94.37
50	BONNIE98_25U2	963	1.21	86.5	2500	39.72	31.31	-76.9
51	ALICIA83_17I2	979	1.16	20.5	1156	40.13	28.49	-94.69
52	BONNIE98_26U1	961	1.32	84	2704	40.25	32.54	-77.67
53	LILI96_18H2	975	1.52	42	1444	40.73	23.24	-77.14
54	DENNIS99_29U1	970	1.73	70	1849	40.81	30.34	-78.27
55	GILBERT88_15I2	950	1.22	54.5	3969	40.86	22.7	-94.18
56	HORTENSE96_13U1	938	0.98	17	5625	41	27.96	-71.19
57	DENNIS99_28U2	974	1.27	90.5	1521	41.08	27.95	-77.5
58	BONNIE98_22U2	984	1.25	40	841	41.27	23.62	-71.26
59	BONNIE98_24U2	961	1.17	83.5	2704	41.38	27.16	-73.24
60	DIANA84_2_12I	960	1.26	16	2809	41.55	33.96	-77.17
61	FRAN96_05U1	949	0.97	83	4096	41.64	30.82	-77.13
62	DIANA84_2_11I	958	1.57	24	3025	41.74	31.69	-78.73
63	FRAN96_05U2	953	0.99	92.5	3600	41.77	33	-77.86
64	EMILY93_30U2	975	1.52	35	1444	42.04	32.35	-73.07
65	FRAN96_04U3	954	0.8	90.5	3481	42.15	29.09	-76.31
66	BONNIE98_26U2	964	1.31	78.5	2401	42.29	33.8	-77.86
67	DIANA84_2_12H2	968	1.49	29.5	2025	42.33	33.86	-77.6
68	FREDERIC79_11H	976	1.76	37.5	1369	42.67	24.69	-85.15
69	EDOUARD96_31U1	947	1.09	56	4356	42.86	32.11	-70.15
70	ELENA85_2_01H	962	1.39	36	2601	42.98	28.58	-84.07
71	FLOYD99_15U2	939	1.42	63	5476	43.16	31.14	-78.88
72	BONNIE98_23U3	957	0.97	92.5	3136	43.35	24.99	-71.9
73	EMILY93_31U1	971	1.63	39.5	1764	43.41	33.05	-74.23
74	FRAN96_03U2	974	1.75	56	1521	43.77	25.48	-72.64
75	FREDERIC79_11IA	976	1.69	37.5	1369	43.8	25.44	-85.5
76	FLOYD99_15I	939	1.68	49	5476	43.86	32.16	-78.55
77	BONNIE98_23U1	961	1.39	33.5	2704	44.14	24.2	-71.56
78	FRAN96_04U1	957	1.46	45	3136	45.03	27.7	-75.4
79	BONNIE98_26I	961	1.33	79.5	2704	45.03	33.69	-77.87
80	BONNIE98_24U1	957	1.33	91.5	3136	45.19	25.73	-72.44
81	HORTENSE96_11I1	976	1.73	27	1369	45.58	21.86	-70.84
82	BONNIE98_24I	960	1.4	76	2809	46.04	26.68	-72.97
83	FRAN96_05H1A	952	1.43	72	3721	46.48	33.44	-77.97
84	OPAL95_04U1	944	0.96	17.5	4761	46.76	26.45	-88.95
85	FREDERIC79_11IC	976	1.65	32.5	1369	48.33	26.06	-86.15
86	ELENA85_2_01I	962	1.46	28.5	2601	48.72	28.89	-84.63
87	FREDERIC79_12F	951	0.95	68	3844	48.77	28.14	-87.42
88	EDOUARD96_30U2	938	0.88	65	5625	48.83	29.62	-70.45
89	ANDREW92_25U2	947	1.22	31	4356	49.28	27.07	-87.96
90	HORTENSE96_12H1	956	1.62	23	3249	49.75	23.28	-71.58
91	FLOYD99_15U1	934	1.71	65	6241	49.91	29.25	-78.88
92	FLOYD99_14U1	926	1.25	35.5	7569	49.96	25.05	-75.66

	Hurricane Name	Pmin	B	Rmax	DelP^2	Vmax	Lat	Lon
93	FREDERIC79_11IB	976	2.06	37.5	1369	50.58	25.81	-85.7
94	ELENA85_2_02I	956	1.6	30	3249	50.66	29.66	-87.15
95	ANDREW92_22U2	988	1.69	16	625	51.04	25.67	-71.68
96	DIANA84_2_11H	966	1.59	16.5	2209	51.28	32.83	-78.23
97	FREDERIC79_12H2	951	1.64	46	3844	51.38	29.97	-88.2
98	ANDREW92_24U1	933	1.32	17	6400	51.39	25.85	-83.54
99	FREDERIC79_12H1B	951	1.58	32	3844	51.96	27.05	-86.8
100	DIANA84_2_11I2	956	1.49	17.5	3249	52.81	33.72	-77.71
101	ANDREW92_25U4	943	1.59	27	4900	53.13	29.15	-91.3
102	FLOYD99_14U2	928	1.22	69.5	7225	53.23	26.05	-76.86
103	ALLEN80_06I	955	1.64	24	3364	53.4	20.15	-81.42
104	EDOUARD96_30U1	940	1.5	29.5	5329	54.12	27.83	-70.28
105	HORTENSE96_12U1	964	1.66	16.5	2401	54.92	25.09	-71.66
106	FLOYD99_13I	923.5	1.57	35	8010.25	55.6	24.41	-73.73
107	FLOYD99_14U3	930	1.71	66	6889	56.79	27.62	-77.87
108	ANDREW92_23U2	940	1.29	12	5329	56.88	25.43	-77.05
109	ALLEN80_08I	940	1.56	15	5329	57.58	24.28	-92.47
110	FREDERIC79_12H1A	951	2	30.5	3844	58.35	26.54	-86.51
111	ANDREW92_25U3	948	1.66	27.5	4225	59.64	28.09	-90.09
112	FLOYD99_13U1	927	1.68	35	7396	60.66	23.78	-70.6
113	ANDREW92_23U3	927	1.68	20	7396	62.26	25.46	-79.54
114	ANDREW92_23U1	954	1.58	13.5	3481	63.42	25.38	-75.04
115	GILBERT88_14H1	889	1.38	12	15376	63.61	20.16	-85.75
116	ALLEN80_07H	905	1.95	16.5	11664	76.28	21.77	-86.46

Appendix 3. SAS codes

A. SAS code for estimation Holland *B* using the radius of maximum winds and latitude

```
title 'hollandB';
data holland;
infile "c:\hollandB.dat";
input y x1 x2 x3 x4 x5;
proc print;
proc reg;
model y=x1 x3 / p r xpx i covb dw;
output out=new p=yhat r=resid;
proc plot data=new;
plot resid*yhat;
proc univariate normal plot data=new;
var resid;
run;
```

B. SAS code for estimation Holland *B* using the radius of maximum winds, the square difference between central minimum sea level pressure and an outer peripheral pressure and latitude

```
title 'hollandB';
data holland;
infile "c:\hollandB.dat";
input y x1 x2 x3 x4 x5;
proc print;
proc reg;
model y=x1 x3 x4 / p r xpx i covb dw;
output out=new p=yhat r=resid;
proc plot data=new;
plot resid*yhat;
proc univariate normal plot data=new;
var resid;
run;
```

C. SAS code for estimation Holland *B* using the latitude, the longitude, the radius of maximum winds, the square difference between central minimum sea level pressure and an outer peripheral pressure and the maximum wind

```
title 'hollandB';
data holland;
infile "c:\hollandB.dat";
input y lat lon rmax delp2 vmax;
proc print;
```

```
proc reg;
model y=lat lon rmax delp2 vmax / p r xpx i covb dw;
output out=new p=yhat r=resid;
proc plot data=new;
plot resid*yhat;
proc univariate normal plot data=new;
var resid;
proc stepwise data=holland;
model y=lat lon rmax delp2 vmax / stepwise;
run;
```

D. SAS code for estimation Holland B using the radius of maximum winds, the square difference between central minimum sea level pressure and an outer peripheral pressure and the maximum wind

```
title 'hollandB';
data holland;
infile "c:\hollandB.dat";
input y lat lon rmax delp2 vmax;
proc print;
proc reg;
model y=rmax delp2 vmax / p r xpx i covb dw;
output out=new p=yhat r=resid;
proc plot data=new;
plot resid*yhat;
proc univariate normal plot data=new;
var resid;
run;
```

Appendix 4. SAS output

A. SAS output using the radius of maximum winds and latitude

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1.07527	0.53763	7.14	0.0012
Error	113	8.51376	0.07534		
Corrected Total	115	9.58903			

Root MSE	0.27449	R-Square	0.1121
Dependent Mean	1.35422	Adj R-Sq	0.0964
Coeff Var	20.26894		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.55384	0.20482	7.59	<.0001
x1	1	0.00015058	0.00750	0.02	0.9840
x2	1	-0.00439	0.00119	-3.70	0.0003

B. SAS output using the radius of maximum winds, the square difference between central minimum sea level pressure and an outer peripheral pressure and latitude

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.12264	0.37421	4.95	0.0029
Error	112	8.46639	0.07559		
Corrected Total	115	9.58903			

Root MSE	0.27494	R-Square	0.1171
Dependent Mean	1.35422	Adj R-Sq	0.0934
Coeff Var	20.30250		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.50264	0.21511	6.99	<.0001
x1	1	0.00086116	0.00757	0.11	0.9096
x2	1	-0.00426	0.00120	-3.55	0.0006
x3	1	0.00000885	0.00001118	0.79	0.4302

C. SAS output using the latitude, the longitude, the radius of maximum winds, the square difference between central minimum sea level pressure and an outer peripheral pressure and the maximum wind

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	3.59077	0.71815	13.17	<.0001
Error	110	5.99826	0.05453		
Corrected Total	115	9.58903			

Root MSE	0.23352	R-Square	0.3745
Dependent Mean	1.35422	Adj R-Sq	0.3460
Coeff Var	17.24352		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.25712	0.37454	0.69	0.4938
lat	1	-0.00100	0.00646	-0.16	0.8770
lon	1	-0.00308	0.00319	-0.97	0.3365
rmax	1	-0.00157	0.00110	-1.42	0.1588
delp2	1	-0.00006305	0.00001434	-4.40	<.0001
vmax	1	0.02652	0.00395	6.72	<.0001

D. SAS output using the radius of maximum winds, the square difference between central minimum sea level pressure and an outer peripheral pressure and the maximum wind

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3.53710	1.17903	21.82	<.0001
Error	112	6.05193	0.05404		
Corrected Total	115	9.58903			

Root MSE	0.23245	R-Square	0.3689
Dependent Mean	1.35422	Adj R-Sq	0.3520
Coeff Var	17.16514		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.50274	0.16479	3.05	0.0028
rmax	1	-0.00181	0.00106	-1.70	0.0915
delp2	1	-0.00006228	0.00001417	-4.39	<.0001
vmax	1	0.02612	0.00391	6.69	<.0001