EXTREME HURRICANE WINDS IN THE UNITED STATES

Thomas H. Jagger and James B. Elsner *
Florida State University, Tallahassee, FL

1. INTRODUCTION

The rarity of severe coastal hurricanes implies that data are sparse and likely unreliable for empirical modelling. Extreme value theory provides a framework for statistical modelling rare wind events and a Bayesian approach allows the inclusion of measurement error and other sources of uncertainty. The utility of the Bayesian approach for modelling the mean number of coastal hurricanes is shown in Elsner and Jagger (2004). But, because parameter estimates in extreme value analysis are sensitive to rare events, the ability to include long historical records is particularly important. Here we are interested in estimating the upper limit of hurricane wind speeds after landfall. The goal is to demonstrate the usefulness of the Bayesian approach for this purpose.

2. EMPIRICAL DISTRIBUTION OF INLAND HURRICANE INTENSITY

Because the reanalysis of the HURDAT dataset does not yet contain a complete listing of landfalling events by location, time, and intensity, we develop an objective technique for estimating the wind immediately after landfall. First, a natural spline interpolation is used to obtain positions and wind speeds at 1-hr intervals from the 6-hr values for all tropical cyclones in HURDAT. Second, the maximum wind speed at the first inland location based on the 1-hr intensity interpolations is considered the maximum inland tropical cyclone intensity. U.S. landfall occurs when the hurricane's eye wall passes directly over the coast or crosses the border from Mexico. Cyclones passing over the Florida Keys, the Outer Banks of North Carolina, etc are excluded as they are not inland penetrating storms. Third, second landfalls occurring within 48 hr of the first landfall are ignored.

Figure 1 shows the empirical distribution of maximum hurricane wind speeds within 1 hr after landfall from the 211 inland penetrating events over the period 1851–2002 using the objective criteria described above. The majority (143 or 67.8%) of hurricanes have peak winds at or below 90 kt with only a few exceeding 124 kt (6 or 2.8%).

3. EXTREME VALUE THEORY

In the absence of empirical or physical evidence for assigning an extreme level to a process, an asymptotic argument is used to generate extreme value models. But, extreme values are scarce making it necessary to estimate rare levels that are much higher than what already have been observed. Extreme value analysis requires an estimation of the probability of events that are more extreme than any that have already been observed. This implies an extrapolation from observed levels to unobserved levels. Extreme value theory provides a family of models to make such extrapolation.

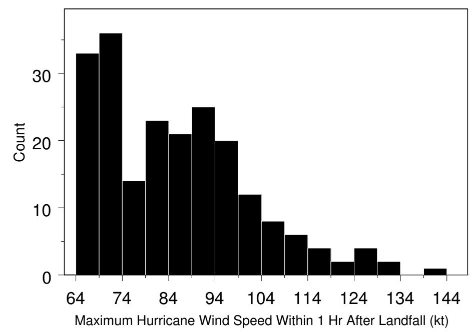

Fig. 1: Empirical distribution of maximum wind speed immediately after landfall. Counts are based on hurricanes making direct landfall in the continental United States over the period 1851–2002, inclusive. The maximum wind speed is estimated from the first on-land position using a natural spline 1 hr interpolation of the 6 hr HURDAT positions and intensities. Hurricanes affecting only the Florida Keys, Outer Banks of North Carolina, Long Island in New York, and Massachusetts' Nantucket Island are excluded.

Consider observations from a collection of independent and identically distributed random variables in which we keep only those observations that exceed a fixed threshold value. The generalized Pareto distribution (GPD) family represents the limiting behavior of this new collection of random variables. This makes the family of GPD a suitable choice for modelling extreme events. Observations below the threshold value are removed from the analysis. The choice of threshold is a compromise between retaining enough observations to properly estimate the distributional parameters, but few enough that the observations follow a GPD. Here we choose 64 kt as the threshold based on examination of linearity in a plot of mean excess versus threshold. The mean excess is the expected value of the amount that the observations exceed the threshold. Moreover, we assume that the conditional distribution of the observed wind speed given both the true wind speed and the parameters depends only on

*Corresponding author address: Thomas H. Jagger, Florida State University, Department of Geography, Tallahassee, FL; e-mail: tjagger@blarg.net

the true wind speed. Thus, given the true wind speed, the observed wind speed is uniformly distributed between the true value plus or minus 2.5 kt to reflect the 5 kt precision used on the majority of storms in HURDAT. Incorporation of the measurement error into our extreme value models is accomplished using a Bayesian approach.

4. BAYESIAN APPROACH

The Bayesian approach provides a context in which to incorporate measurement error and prior beliefs into the model. It allows a complete picture of the distribution of the extremal quantities. Extremal quantities of interest include the distribution of return levels and the joint distribution of the parameters. More technically, the Bayesian approach allows us to supply parameter posteriors when classical methods fail. For example when the shape parameter ξ is less than or equal to -0.5, the MLE cannot be used to determine the distribution of the parameter values (Coles 2001).

We employ an hierarchical specification in which the conditional distribution of the observed maximum wind speed within 1 hr after landfall depends only on the true wind speed w_i and the conditional distribution of w_i depends only on the distribution of the parameters. Mathematically, let unif(a, b) represent the uniform distribution with limits a and b, let gpd(ξ, σ) represent the GPD with parameters ξ and σ, let mvnorm$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and let u represent the threshold value, then our specification is:

$$\begin{aligned} y_i &\sim \text{unif}(w_i - 2.5, w_i + 2.5) \\ w_i &\sim u + \text{gpd}(\sigma, \xi) \\ \log(\sigma) &= \beta_1 \\ \xi &= \beta_2 \\ [\beta_1, \beta_2] &\sim \text{mvnorm}(\mu, \boldsymbol{\Sigma}) \\ \boldsymbol{\mu} &= [3.5, -0.5] \\ \boldsymbol{\Sigma} &= \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix} \end{aligned}$$

Since we use a fixed threshold, the likelihood for the threshold exceedance rate and the GPD can be separated. If the rate prior is independent of the GPD parameter priors, then the posterior parameter samples of the rate will be independent of those for the GPD parameters. This allows us to use two separate models, one for the GPD parameters, and another for the exceedance rate. From these models we generate sample rates and sample GPD parameters, which allows us to generate a series of sample seasons.

The hierarchical Bayesian specification combines the prior with the GPD as the likelihood to generate posterior estimates for β_σ and β_ζ. This is accomplished using the Bayesian inference using Gibbs sampler (BUGS). BUGS is a software for the Bayesian analysis of complex statistical models using Markov chain Monte Carlo (MCMC) methods. Currently, BUGS does not have support for the

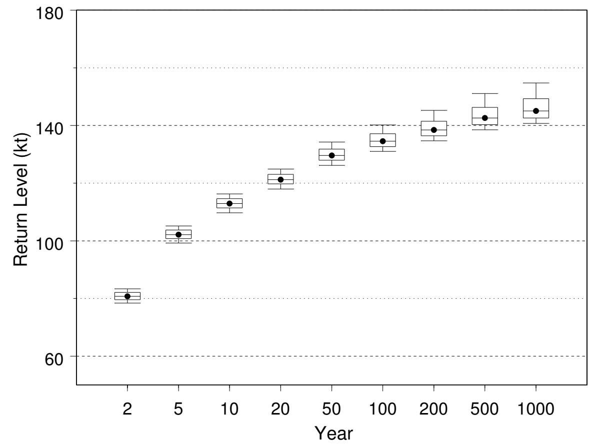

Fig. 2: Return levels for 2-, 5-, 10-, 20-, 50-, 100-, 200-, 500-, and 1000-year hurricane winds. Return levels are in knots representing the maximum U.S. hurricane wind speed within 1 hr after landfall. The box plot indicates the median value (dot), the interquartile range (box limits) and the 10th and 90th percentiles (whiskers) of the return levels.

GPD, which requires us to make several changes. To ensure stability of the results we run the Gibbs sampler for 30K updates and discard the first 10K as burn-in. The return-level plot (Fig. 2) shows the distributions of return levels for 2-year, 5-year, 10-year, 20-year, 100-year, 200-year, 500-year, and 1000-year inland hurricane winds. Up to a point, the longer we wait the higher the return level. On average we can expect 80 kt hurricane winds in the U.S. every 2 years and 122 kt winds every 20 years. Since $P(\xi \geq 0)$ is less than 2.2% in the model, coastal hypercanes are not possible under the assumption that the extreme near-coastal winds follow a distribution from the GPD family.

The model can be improved by incorporating factors such as the ENSO and NAO. Statistical significance can be examined using the notion of model selection which amounts to determining the most reasonable model given the data. In this regard, when using a Bayesian approach the Deviance Information Criterion (DIC) plays the same role as Akaike's Information Criterion (AIC) aiming to identify models that best explain the observed data.

Acknowledgement. Partial support for this study was provided by the National Science Foundation (ATM-0086958).

5. REFERENCES

- Coles, S., 2001: An Introduction to Statistical Modeling of Extreme Values. Springer, New York, 208 pp.
- Elsner, J. B., and T. H. Jagger, 2004: A hierarchical Bayesian approach to seasonal hurricane modelling. In review.