INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.1804

Modeling tropical cyclone intensity with quantile regression

Thomas H. Jagger* and James B. Elsner
Department of Geography, Florida State University, Tallahassee, Florida, USA

ABSTRACT: Wind speeds from tropical cyclones (TCs) occurring near the USA are modeled with climate variables (covariates) using quantile regression. The influences of Atlantic sea-surface temperature (SST), the Pacific El Niño, and the North Atlantic oscillation (NAO) on near-coastal TC intensity are in the direction anticipated from previous studies using Poisson regression on cyclone counts and are, in general, strongest for higher intensity quantiles. The influence of solar activity, a new covariate, peaks near the median intensity level, but the relationship switches sign for the highest quantiles. An advantage of the quantile regression approach over a traditional parametric extreme value model is that it allows easier interpretation of model coefficients (parameters) with respect to changes to the covariates since coefficients vary as a function of quantile. It is proven mathematically that parameters of the Generalized Pareto Distribution (GPD) for extreme events can be used to estimate regression coefficients for the extreme quantiles. The mathematical relationship is demonstrated empirically using the subset of TC intensities exceeding 96 kt (49 m/s). Copyright © 2008 Royal Meteorological Society

KEY WORDS tropical cyclones; quantile regression; ENSO; solar cycle; extreme value models

Received 30 January 2008; Revised 20 June 2008; Accepted 11 October 2008

1. Introduction

Coastal tropical cyclones (TCs) pose a serious threat to society and the economy. Strong winds, heavy rainfall, and storm surge kill people and destroy property. The rarity of intense TCs implies that empirical estimates of return periods will be unreliable. Fortunately, extreme value theory provides parametric models for rare events and a justification for extrapolating to intensity levels that are greater than the historically observed ones. Jagger and Elsner (2006) have developed extreme value models for USA hurricane intensity, based on the method of peaks over thresholds, using data over the period 1899–2006. They show how the models can be used to assess the probability of extremely intense hurricanes conditional on climate factors.

Quantile regression offers another way to model extreme TC events that is yet to be examined. Quantile regression, introduced by Koenker and Bassett (1978), extends the ordinary least squares regression model to conditional quantiles (e.g. 90th percentile) of the response variable. It can be considered a semiparametric technique because it relies on nonparametric quantiles, but uses parameters to assess the relationship between the quantiles and the covariates.

Ordinarily we think of parametric models as more informative, with nonparametric models useful for an initial look at the data. A parametric model involves more stringent assumptions, but it is usually a good idea to start with stronger assumptions and back off toward weaker assumptions when necessary. However, parametric models are generally more sensitive to outlying data values, which can be problematic for models of extreme values. Also, with parametric models care must be taken in specifying the distribution.

A drawback of parametric models is that the parameters can be more difficult to interpret physically. It is this difficulty in interpreting the parameters of the extreme value models with respect to issues of climate's influence on TC activity that prompts the present study. It is our contention that extreme value models are valuable for quantifying the probability of high winds from TCs conditional on climate covariates, but that quantile regression can be useful as an exploratory tool.

The present study is motivated by the importance of providing accurate statistical estimates of the probability of the next big hurricane conditioned on climate variability and change, and by the possibility that a model that is simpler to implement and interpret will likely be adopted over one that requires greater sophistication to use. The purpose here is not to argue in favor of one methodology over the other. Circumstances will dictate the choice. Instead, the goal is to introduce quantile regression as an alternative method for analyzing how quantiles of TC intensity distribution change as a function of climate variables and to show how it relates to an extreme value model.

The article is organized as follows: In the next section we briefly review the mathematics behind the quantile regression models. Following this we provide a description of the data sets used in the present work. We then perform exploratory analysis on the data and define the quantile function. We also describe, using conditional

* Correspondence to: Thomas H. Jagger, Department of Geography, Florida State University, Tallahassee, FL 32306, USA.
E-mail: tjagger@blarg.net

quantile plots, the bivariate relationships between the TC intensity and covariates (predictors). With the exception of the solar cycle, the choice of covariates is based on previous research about the USA hurricane threat. This is followed by an examination of quantile regression models for these data. We begin with the bivariate case and then look at results from the multivariate case. Finally, we examine the relationship between an extreme value model previously used on near-coastal wind speeds and quantile regression.

2. Background

Quantile regression is an extension of median regression based on estimating the value of the parameter vector $\boldsymbol{\beta}$ from the set of allowable vectors that minimizes the mean loss function

$$L_\tau(\boldsymbol{\beta}, \mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n} p_\tau\,(y_i - \mu(\mathbf{x_i}, \boldsymbol{\beta})) \quad (1)$$

where y_i; $i = 1, \ldots, n$ are the response values, μ is the estimate of the τ quantile, and $\boldsymbol{x}_i$ and $\boldsymbol{\beta}$ are the covariate vector and parameter vector, respectively. The loss function is $p_\tau(\cdot)$, where

$$p_\tau(z) = |z|\{\tau \cdot I(z > 0) + (1 - \tau) \cdot I(z < 0)\} \quad (2)$$

and $I(\cdot)$ is the indicator function, which is one when the argument is true and zero otherwise. The loss function is non-negative taking a minimum value of zero only when $z = 0$.

Given a series of samples with μ constant (intercept-only model), the resulting value of $\boldsymbol{\beta}$ (a scalar in this case) that minimizes the total loss function occurs only when μ is equal to the τ quantile of the response. If the model fits well, a plot of fitted *versus* actual values will show that τ percentage of observed values should be less than the fitted values, with $1 - \tau$ percentage of the observed values greater than that of the fitted values (Yu *et al.*, 2003). The total loss function is an unbiased sample estimate of the expected value of $p_\tau[Y - \mu(\boldsymbol{x} \cdot \boldsymbol{\beta})]$, and the minimization over $\boldsymbol{\beta}$ is a consistent estimate of the minimization of this expected value. For the fit, we employ a linear model for the regression function of the form

$$\hat{\mu} = \beta_0 + \sum_{i=1}^{p} \beta_i \cdot x_i \quad (3)$$

where x_i is climate covariate i and there are p of them.

In the meteorological literature, Bremnes (2004) shows how to produce reliable probability of precipitation forecasts using quantile regression and Friederichs and Hense (2007) show how to use quantile regression to downscale forecasts of extreme precipitation from the reanalysis data. In the present study we show the value of going beyond models for the conditional mean in climate studies. We are not aware of any climate study that makes use of quantile regression although we note that Gray *et al.* (1992) have used median regression to model Atlantic hurricane count data.

3. Data

3.1. Maximum near-coastal tropical cyclone intensity

Here TC wind speed estimates are derived from the HURricane DATa base (HURDAT or best track) maintained by the *National Oceanic and Atmospheric Administration* (NOAA) *National Hurricane Center* (NHC) of USA. Of interest is the fact that HURDAT is the official NOAA record of TC information for the Atlantic Ocean, Gulf of Mexico, and Caribbean Sea, including those that have made landfall in the United States. HURDAT consists of the 6-hr position central pressure and maximum sustained wind estimates for TCs dating back to 1851 (Jarvinen *et al.*, 1984; Neumann *et al.*, 1999). For TCs prior to 1931, the 6-hr positions and intensities are interpolated from once-daily (12 UTC) estimates. Here we use the latest version of HURDAT as of December 2006, which includes a reanalysis of all TCs prior to 1911 (Landsea *et al.*, 2004).

A natural spline interpolation is used to obtain positions and wind speeds at 1-h intervals from the 6-h values in HURDAT. Since a complete data set of all land falling TCs is not available and to be consistent, we use the near-coast region outlined in Jagger and Elsner (2006) and keep only the TC's single highest wind speed in the region. We use the term "intensity" as shorthand for "wind speed." Thus the dataset we analyze and model in this study contains $N = 422$ TCs over a 108-year period (1899–2006). The raw wind speed values in HURDAT are given in 5 kt (2.5 m/s) increments and knots (kt) are the operational unit used for reporting TC intensity to the public in the United States. Thus, all analysis and modeling done here use wind speed values in knots (1 kt = 0.51 m/s).

3.2. Climate covariates

On the seasonal timescale, and to a first order, it is known that a warm ocean fuels TC genesis, a calm atmosphere (low values of wind shear) allows a TC to intensify, and the position and strength of the subtropical high-pressure region functions as a steering mechanism paving a track for a TC that does form. Thus, regional hurricane activity responds to changes in large-scale climate conditions [e.g., El Niño-Southern Oscillation (ENSO)]. Initially we choose the same covariates used in Jagger and Elsner (2006) including the Southern Oscillation Index (SOI) as an indicator of ENSO, the North Atlantic Oscillation (NAO) index as an indicator of storm steering, and Atlantic sea-surface temperature (SST) as an indicator of cyclone energy.

ENSO is characterized by basin-scale fluctuations in sea-level pressure (SLP) across the equatorial Pacific Ocean. The SOI is defined as the normalized SLP difference between Tahiti and Darwin, and values are

DOI: 10.1002/joc

available back through the mid-19th century. The SOI is strongly anti-correlated with equatorial Pacific SSTs so that an El Niño warming event is associated with negative SOI values. Units are standard deviations. ENSO is an indicator of vertical wind shear and subsidence in the environment where TCs develop and negative SOI values imply greater shear and subsidence. As expected, the relationship is strongest during the TC season, so we use an August–October average of the SOI as our covariate. The monthly SOI values (Ropelewski and Jones, 1997) are obtained from the *Climatic Research Unit* (CRU).

The NAO is characterized by fluctuations in SLP differences. Index values for the NAO are calculated as the difference in SLP between Gibraltar and a station over southwest Iceland, and are obtained from the CRU (Jones *et al.*, 1997). The values are averaged over the preseason and early season months of May and June (Elsner *et al.*, 2001) and can be considered an indicator of the strength and/or position of the subtropical Bermuda High. We speculate that the relationship might result from a teleconnection between the mid-latitudes and tropics whereby a below-normal NAO during the spring leads to dry conditions over the continents and to a tendency for greater summer/fall middle tropospheric ridging (enhancing the dry conditions). Ridging over the eastern and western sides of the North Atlantic basin tends to keep the middle tropospheric trough, responsible for hurricane recurvature, farther to the north during the peak of the season (Elsner and Jagger, 2006).

The Atlantic SST covariate is an area-weighted average based on monthly SST values in 5-degree latitude–longitude boxes (Enfield *et al.*, 2001) from the equator to 70°N latitude. For this study we average the SST monthly anomalies (monthly means subtracted) over the peak hurricane season months of August through October. The covariate is the standard Atlantic Multidecadal Oscillation (AMO) index.

We also consider the influence variations in the sun might have on near-coastal TC intensity. This is motivated by a recent study of ours (Elsner and Jagger, 2008). We speculate that an increase in solar UV radiation during periods of strong solar activity will have a suppressing effect on TC intensity as the temperature near the tropopause will warm through absorption of radiation by ozone and modulated by dynamic effects in the stratosphere. For the solar covariate we use the August through October averaged sunspot number (SSN). The SSNs produced by the *Solar Influences Data Analysis Center* (SIDC), World Data Center for the Sunspot Index, and the *Royal Observatory of Belgium* are obtained from USA NOAA.

In summary, the distribution of near-coastal TC intensity will be modeled with quantile regression using data from the period 1899–2006. Additionally, the August through October averaged values of the SOI, SST, and SSN along with the May through June averaged values of the NAO will be used to examine changes to intensity quantiles depending on how these climate variables fluctuate over this 108-year period. All four covariates have been statistically linked to USA coastal hurricane activity (Elsner and Jagger, 2004, 2008).

4. Exploratory analysis

In this section we define some terms associated with quantiles, and perform exploratory analysis on the wind speed and covariate data that will motivate the later use of quantile regression.

4.1. Quantiles and the quantile function

Quantiles are points taken at regular intervals from the cumulative distribution function (CDF) of a random variable. The quantiles mark a set of ordered data into equal-sized data subsets. For example, of the 422 TC intensity values in our near-coastal data set, 25% of the values are less than 43 kt, while 50% are less than 61 kt. Thus there is an equal number of TCs with intensities between 10 and 43 kt (10 kt is the lowest maximum wind speed value in the data set) as there are between 43 and 61 kt. When we state that the median near-coastal maximum storm intensity is 61 kt, we mean that half of all TCs have intensities less than this value and half have intensities greater. Similarly, the quartiles (deciles) divide the sample of TC intensities into four (ten) groups with equal proportions of the sample in each group. The quantiles or percentiles refer to the general case.

The cumulative distribution and quantile functions of the 422 TC intensities are shown in Figure 1. The CDF in Figure 1(a) gives the empirical probability of observing a value in the record less than a given intensity. The quantile function [Figure 1(b)] is the inverse of the CDF allowing us to determine the TC intensity for specified quantiles. Both functions are monotonically nondecreasing. Thus, given a sample of intensities $w_1, \ldots, w_n$, the τ th sample quantile is the τ th quantile of the corresponding empirical CDF. Formally, let W be a random TC intensity, then the kth "q"-quantile is defined as the value "w" such that

$$P(W \leq w) \geq \tau \text{ and } P(W \geq w) \geq 1 - \tau \qquad (4)$$

where $\tau = \text{k/q}$.

4.2. Conditional quantile plots

Quantile regression is an extension of these ideas to the estimation of conditional quantiles. A model for the conditional quantile response is expressed as a function of the observed covariates in a linear way (typically). For instance, suppose we are interested in the quantiles of TC intensity for different values of SST. We first divide the covariate SST data into equally spaced quantiles. The choice of the set of quantiles is a compromise between having enough intensity values for a given range of SSTs and having enough quantiles to assess the fit of the model. Here we choose deciles so we divide the set of 422 SST values into 10 levels labeled Q1, Q2, ..., Q10. Note that although we only have 108 years (1899–2006), each

DOI: 10.1002/joc

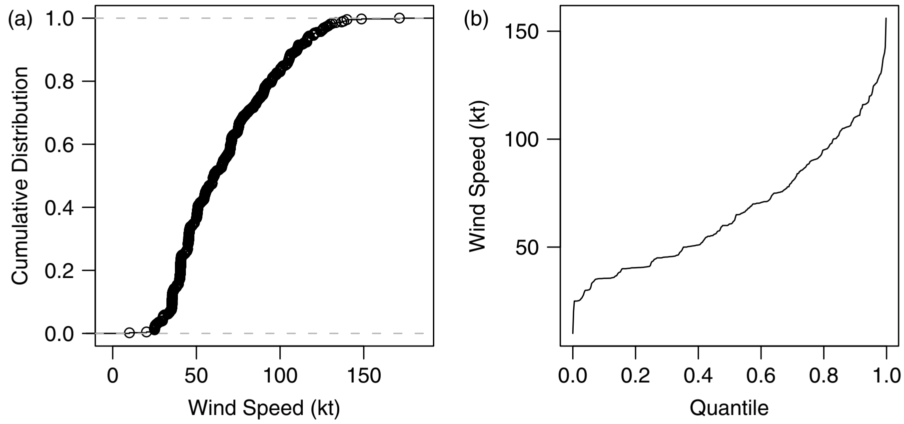

Figure 1. The cumulative distribution function (CDF) (a) and the corresponding quantile function (b) of the sample of the 422 near-coastal TC wind speeds (TC intensity) from the period 1899–2006.

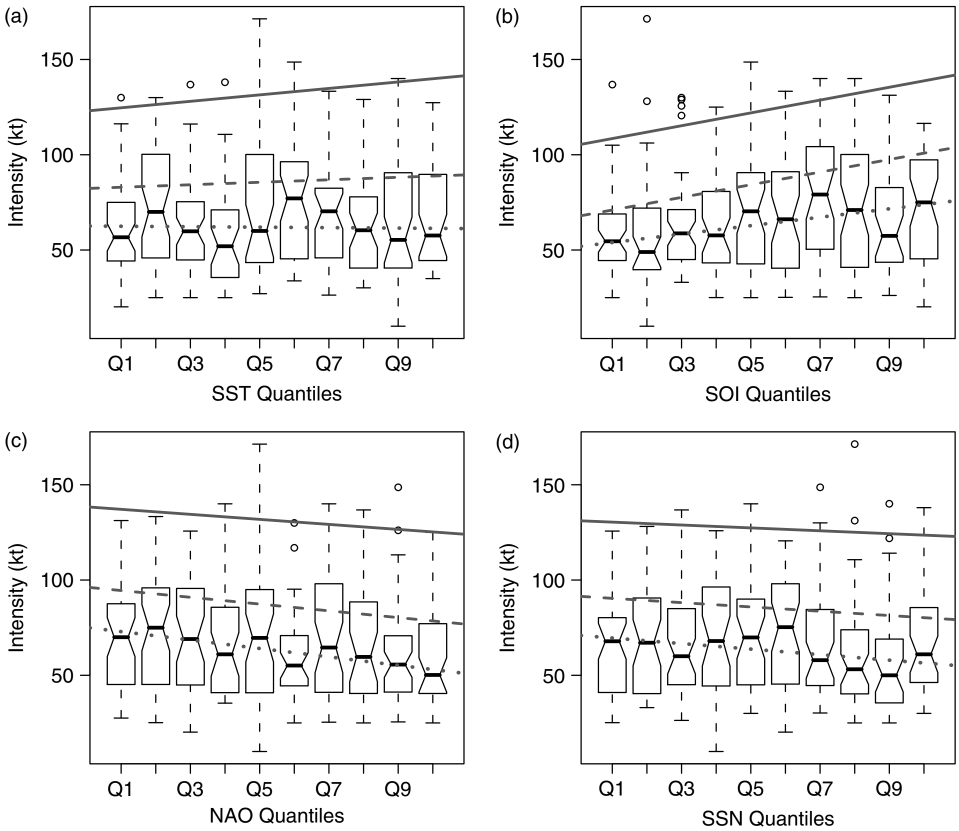

Figure 2. Tropical cyclone intensity as a function of the covariates; (a) SST, (b) SOI, (c) NAO, and (d) SSN. The box plot provides a summary of the distribution of TC intensity by deciles of the covariate. The first decile is the lowest 10% of all covariate values. The upper and lower limits of the boxes represent the first and third quartiles of TC intensity (25th and 75th quantiles). The median is a horizontal bar in the middle of the box. Notches on the box sides represent an estimated confidence interval for each median estimate. The extremes are horizontal bars at the end of the dashed "whiskers." In cases where the whiskers would extend more than one and a half times the interquartile range, they are truncated at that level and the remaining outlying points are indicated by open circles. The solid line is the best-fit line through the upper whisker values for each NAO covariate quantile, the dashed line is the best-fit line through the upper quartile values, and the dotted line is through the medians.

storm in a given season gets assigned the corresponding seasonally averaged SST value.

We then look at the distribution of TC intensity for each SST decile with box plots (Figure 2). The plots show a tendency for the upper quantiles of intensities to increases with SST. They also show that the increase in intensity with SST is more pronounced for higher TC intensity quantiles. The relationship is stronger for the SOI and NAO. There is also a tendency for the dispersion in intensity values to increase with SOI as can be seen by the widening of the interquartile range (range of values between the 25th and 75th percentiles). The plots indicate that the classical conditional mean model (e.g. linear regression) may not adequately capture the full range

DOI: 10.1002/joc

of relationships between the covariates and coastal TC intensity.

5. Quantile regression

The quantile function and the conditional box plots shown in the previous section are useful for exploratory analysis and are adequate for describing and comparing single-variable (univariate) distributions. However, since we are interested in modeling the relationship between a response variable and the covariates, it is necessary to introduce a regression-type model for the quantile function. The quantile regression model is an extension of the classical regression model. Quantile regression allows us to examine the relationship without the need to consider discrete levels of the covariate. Note that we are now interested in the quantiles of W conditional on the values of the covariates rather than those conditional on the quantiles of the covariates. The classical regression model specifies how the conditional mean changes with changes in the covariates while the quantile regression model specifies how the conditional quantile changes with changes in the covariates. We first consider the bivariate case where the quantiles of TC intensity are regressed onto each of the covariates separately.

5.1. Bivariate case

Figure 3 shows quantile regression lines for bivariate models for TC intensity using τ values of 0.5, 0.75, and 0.95. The quantile regression equation with SST as the lone covariate is given by

$$\hat{\mu}(\tau|\text{SST}) = \hat{\beta}_0(\tau) + \hat{\beta}_{\text{SST}}(\tau) \cdot \text{SST} \qquad (5)$$

where $\hat{\mu}(\tau|\text{SST})$ is the predicted conditional quantile of TC intensity (W). The intercept ($\hat{\beta}_0$) and slope ($\hat{\beta}_1$) are obtained by minimizing the piecewise linear least absolute deviance function given by

$$\frac{1-\tau}{n}\sum_{w_i<q_i}|w_i - q_i| + \frac{\tau}{n}\sum_{w_i>q_i}|w_i - q_i|. \qquad (6)$$

where q_i is the predicted τ quantile corresponding to observation i. This function gives the same value as the mean loss function given by Equations (1) and (2). Here we see how quantile regression reveals the broader nature of the relationship between TC intensity and climate covariates and underscores the limited view afforded by ordinary regression (black line). The quantile regression model reveals the nature of the TC–SST relationship on the seasonal time scale with TC intensity dependent on SST only for intensities in the upper quantiles for a given value of SST.

The quantile relationships between the SOI and near-coastal TC intensity are also interesting. Here we see the results of the well-known inhibiting effect that ENSO has on the US hurricane activity. As SOI increases (toward

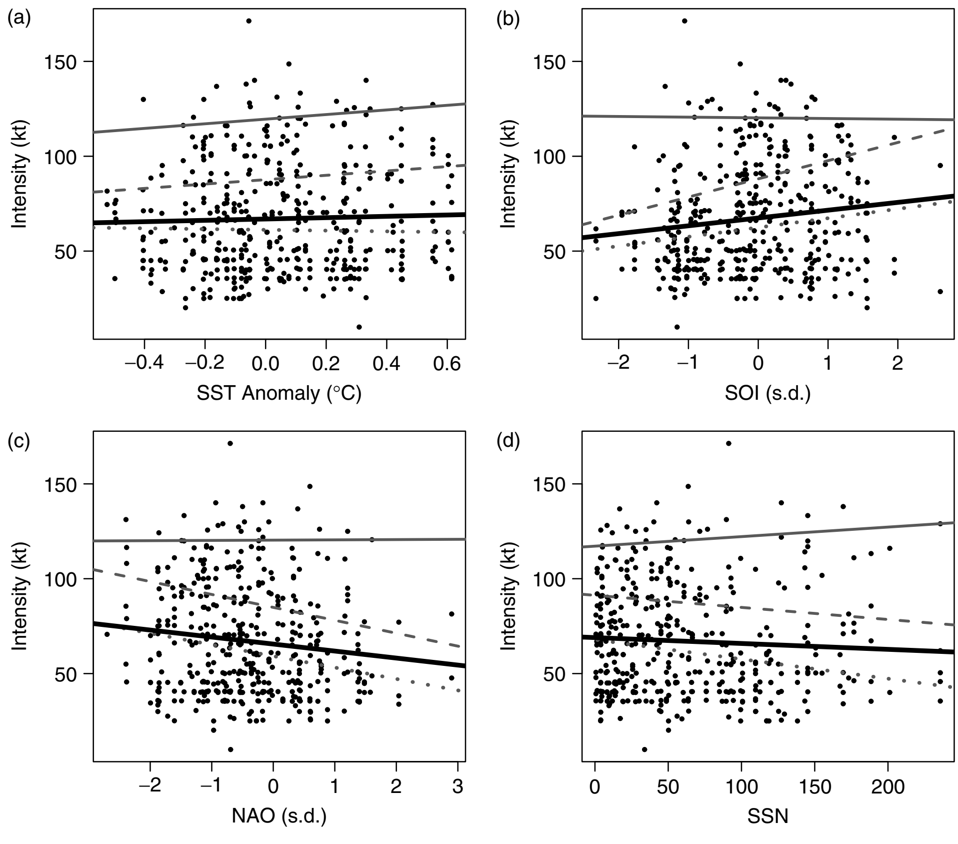

Figure 3. Tropical cyclone intensity as a function of the same covariates as Figure 1, using quantile regression. The conditional quantiles include the 50th percentile (dotted line), the 75th percentile (dashed line), the 95 percentile (solid). The least-squares regression line is shown as a thick solid line.

La Niña like conditions) the TC intensity increases in the mean and, for quantiles, in the range between the 50th and the75th percentiles. Interestingly we see that at the 95th percentile, a slight negative relationship that becomes more negative with still higher quantiles. Thus while the influence of shear and subsidence caused by a Pacific El Niño event limits the development of TCs, a cyclone capable of overcoming these debilitating influences does, on average, go on to become a strong hurricane.

There is an analogy here with tornadic thunderstorms. Although a shallow lapse rate of temperature inhibits the development of thunderstorms, if a storm is able to break through this "cap", the potential for a tornado-producing supercell increases. While the analogy is not tight, we feel it helps convey our thinking about the differential role ENSO plays in modulating TC activity.

Note that this result was reported in Jagger and Elsner (2006) using a parametric model and questions about its robustness was one of the motivations behind the present work. As with the classical regression model, we can go beyond looking at bivariate relationships with quantile regression and consider the conditional quantiles of TC intensity as a function of more than one covariate.

5.2. Multivariate case

The quantile regression model for TC intensity with the four covariates is given by

$$\hat{\mu}(\tau|\text{SOI,SST,NAO,SSN}) = \hat{\beta}_0(\tau) + \hat{\beta}_{\text{SOI}}(\tau) \cdot \text{SOI} + \hat{\beta}_{\text{SST}}(\tau) \cdot \text{SST} + \hat{\beta}_{\text{NAO}}(\tau) \cdot \text{NAO} + \hat{\beta}_{\text{SSN}}(\tau) \cdot \text{SSN} \quad (7)$$

Since we are interested in comparing changes to the conditional quantile function across different variables, we first scale each of the covariates to have a zero mean and a standard deviation of one. Interpretation of the coefficient values from the quantile regression is the same as for standard regression so that, for example, the estimated value for $\beta_{SST}(0.9)$ of 6.2 indicates that for every one s.d. increase in SST (standardized), there is a 6.2 kt increase in TC intensity at that quantile holding the other covariates constant. The unconditional 90th percentile intensity is 110 kt, so with a 1 s.d. increase in SST, TC intensity increases to 116.2 kt.

The plots in Figure 4 quantify how the quantiles of TC intensity are conditioned on each of the four covariates holding the other three constant. A coefficient value of zero indicates no relationship between intensity and the covariate after accounting for the other covariates. We see the influence of SST on intensity is most pronounced for quantiles above about the upper quartile (75th percentile) although there is some dependency for the weakest cyclones (lowest quantiles). This indicates that the SST affects the TC intensity more when the cyclone is closer to its maximum potential intensity (Emanuel, 1988). The influence of ENSO on TC intensity is more dramatic, especially for the middle and upper quantiles. However, for the most extreme hurricanes we see only a weak relationship.

The influence of the NAO on TC intensity is quite similar to that of ENSO, only that the relationship is negative. The relationship with solar activity is the weakest and generally negative indicating a tendency for stronger near-coastal TCs with a "cooler" sun. Less ultraviolet radiation with a cooler sun implies a cooler upper troposphere that enhances vertical instability assuming everything else being equal.

We increase the resolution on the regression plots and consider only the upper quantiles corresponding to the strongest TCs (Figure 5). Here we see that for the SST covariate the influence on TCs gets stronger with higher quantile intensities, but weaker with quantile intensities for the ENSO covariate. Interestingly, the sun's influence on TCs reverses for the highest quantiles [see also Figure 4(d)]. The reversal refers to a change in the sign of the coefficient relative to the coefficient sign from a standard regression about the mean (straight line in Figures 4 and 5).

The model uses all four covariates for each quantile, but the plots of the predicted intensity are displayed as a function of covariate pentiles with no regard to the value of the other covariates (Figure 6). The predicted values are summarized with box plots showing the median value, the maximum and minimum values, and the interquartile range (difference between the top and bottom of the box). The predictions do not take into account the uncertainty surrounding the estimated value of the coefficients. Here we see the effect of SST on TC intensity is stronger at the 95th percentile compared with the 75th percentile, but the effect of ENSO is stronger at the 75th percentile than at the 95th percentile. The effect of solar activity changes the sign between the 75th and 95th percentiles.

If we restrict the model to having two covariates, we can examine the relationships between the covariates and TC intensity with a contour plot. Figure 7 shows contours of TC intensity at three quantile values using separate models. At the 75th percentile, the gradient of intensity is directed from high values of NAO and low values of SOI toward low values of NAO and high values of SOI. At the 90th percentile the gradient is directed from the low values of SST and SOI toward the high values of SST and SOI. At the 95th percentile the gradient is directed toward higher values of SST and toward higher values of SSN only for SSN values less than the mean.

As shown in this section, quantile regression is useful in modeling conditional quantiles of TC intensity. It is a tool that goes beyond conditional mean models that helps in the understanding of how TC activity is related to climate variability. Since parametric extreme value models are also used in this regard, next we demonstrate how conditional extreme upper quantiles are related to parameters of a generalized Pareto distribution (GPD).

6. Extreme quantiles and the GPD

It is important to show that quantile regression gives the same predictions when the data follow a particular

DOI: 10.1002/joc

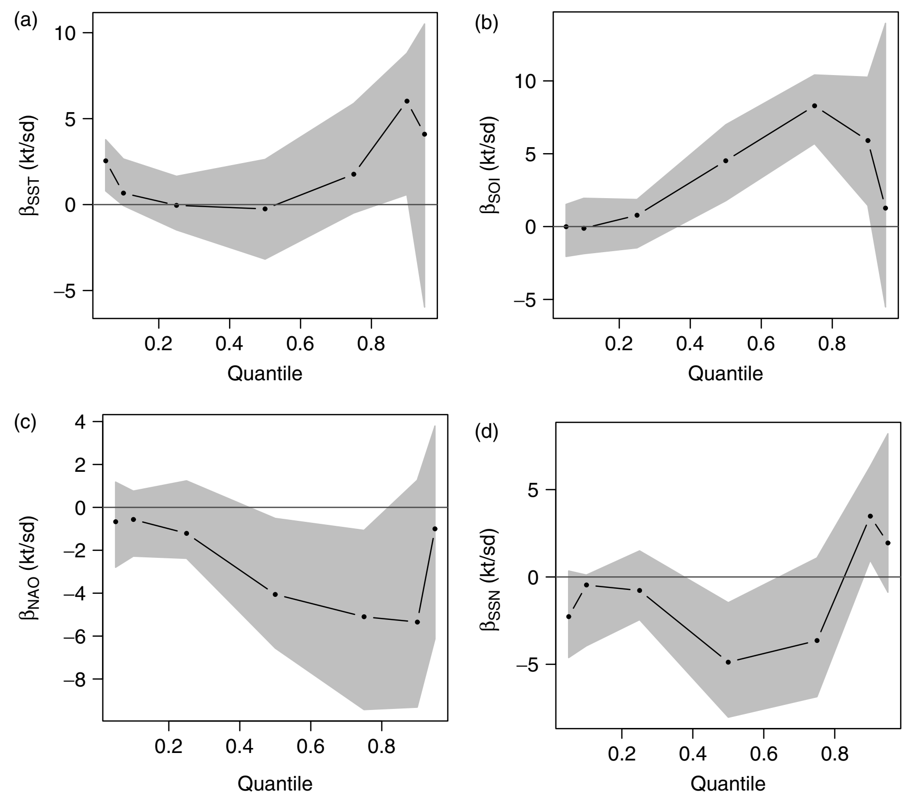

Figure 4. Coefficients of quantile regressions using quantiles of TC intensity as the response variable. The covariates include SST, SOI, NAO, and SSN. The regression coefficients express the change in the expected quantile for a 1 s.d. change in the covariate as a function of the quantiles [τ = (0.05, 0.1, 0.25, 0.5, 0.75, 0.90, and 0.95)]. The gray shading indicates the 90th pointwise confidence intervals about the coefficients.

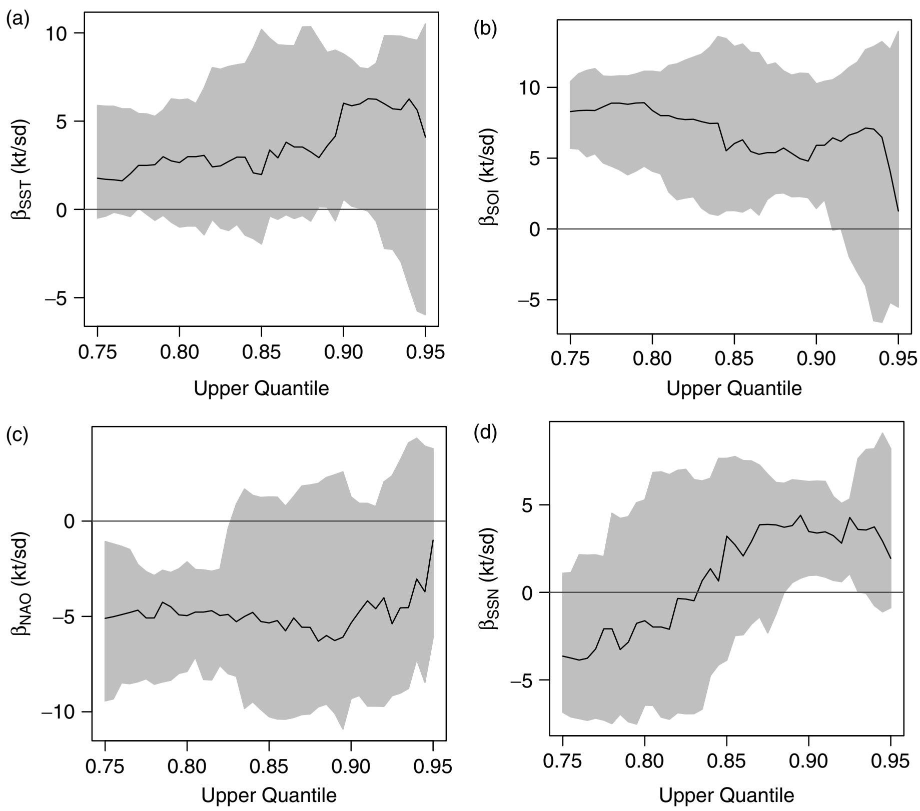

Figure 5. Same as Figure 4 except quantiles in the range of 0.75 to 0.95 in quantile increments of 0.005.

DOI: 10.1002/joc

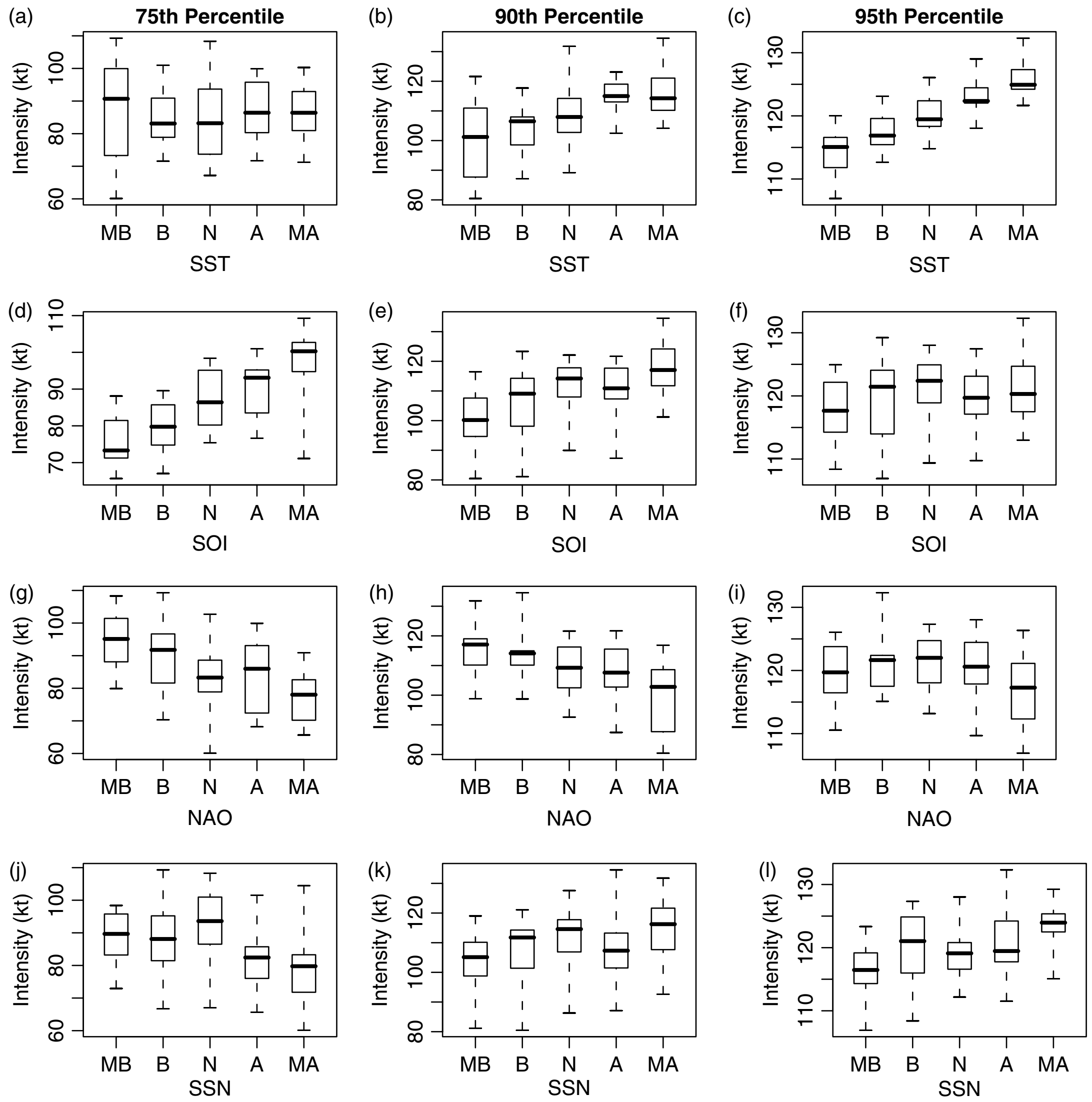

Figure 6. In-sample model predictions of the conditional quantile values of TC intensity for each of the covariates.

parametric form, in this case the GPD distribution. The mathematical derivation that follows shows us that under certain circumstances we can derive a relationship between the GPD parameters and the quantile regression parameters. Using a subset of our hurricane data set and all covariates we estimate both the GPD and quantile regression parameters. Using this relationship, we derive a new set of quantile regression parameters from the GPD parameters and show that this new set of parameters is similar to the same set estimated from the quantile regression.

As shown in Jagger and Elsner (2006), the set of extreme (hurricanes with wind speeds exceeding 96 kt) near-coastal winds follow a GPD. The GPD distribution is used to characterize the extremes or tails of a continuous distribution. In order that the GPD be appropriate a suitable threshold, $c = 96$ kt, is selected and a new data set is created by keeping only those observations that exceed this threshold. If c is large enough then resulting distribution of this data set can usually be approximated by a GPD.

This distribution has three parameters, c, σ, ξ the threshold, scale, and shape parameters respectively. The GPD distribution can be given in its exceedance form as:

$$p(\mu) = P(W > \mu | W > c) = \begin{cases} \exp([\mu - c]/\sigma) & \text{when } \xi = 0 \\ \left(1 + \frac{\xi}{\sigma}[\mu - c]\right)^{-1/\xi} & \text{otherwise} \end{cases} \quad (8)$$

where $\mu > c$, $\sigma > 0$, and when $\xi < 0$, the distribution is bounded with $\mu < c + \sigma/|\xi|$.

Since the distribution of extreme hurricane winds follows a GPD, the form of the quantile regression is constrained. To see this, let $p = 1 - \tau$ be the conditional probability that a random observed TC wind speed exceeds w given x. Let $\mu(p|x)$ be the conditional quantile given x. In other words, the probability that an observed wind speed exceeds $\mu(p)$ given x is p. Let us also assume that the quantile regression is linear in some covariate x. Now we relate the conditional quantiles to

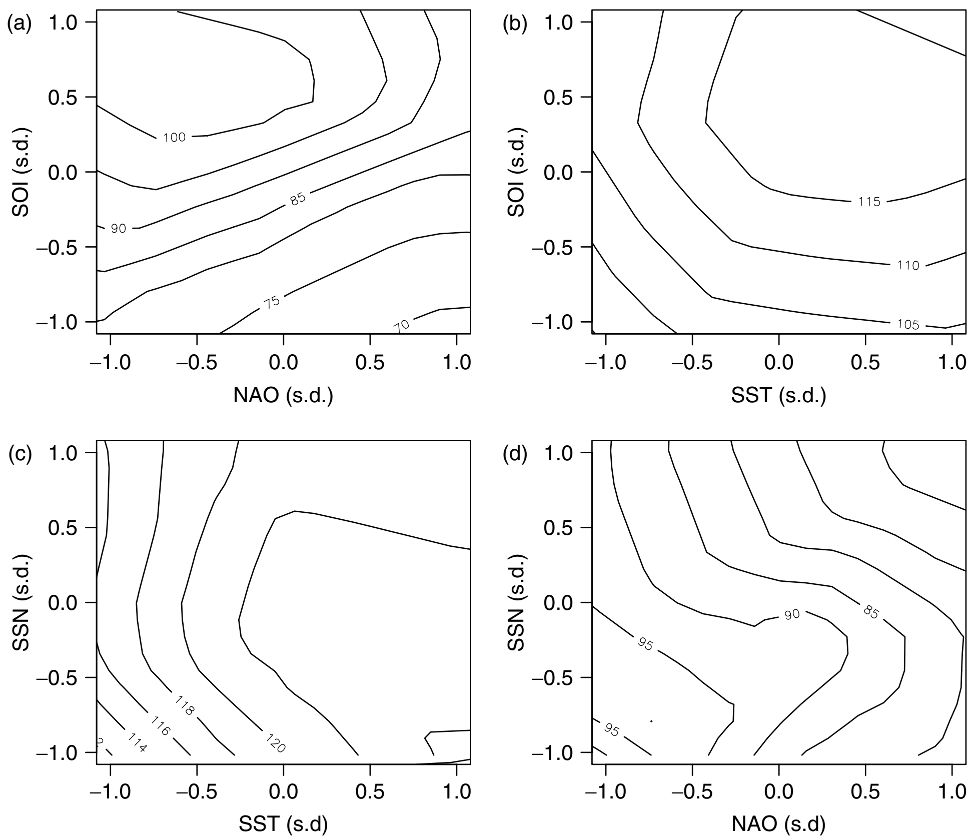

Figure 7. Quantile surfaces of TC intensity conditional on two covariates. The values are smoothed using a bandwidth of 0.5. Only predictions within ±1 s.d. of the covariates are displayed. (a) NAO *vs* SOI at $\tau = 0.75$, (b) SST *vs* SOI at $\tau = 0.9$, (c) SST *vs* SSN at $\tau = 0.95$, and (d) NAO *vs* SSN at $\tau = 0.75$.

the GPD quantiles as a function of p and x:

$$\mu(p|x) = \beta_0(p) + x \cdot \beta_1(p) \tag{9}$$

$$\mu(p|x) = c(x) + \frac{\sigma(x)}{\xi(x)} \cdot \left\{ \frac{1}{p^{\xi(x)}} - 1 \right\} \tag{10}$$

The first Equation (9) is the quantile regression model with one covariate x. The second Equation (10) is the quantile function for the GPD distribution for fixed x and exceedance probability p. This equation is derived from Equation (8) by solving for μ as a function of p, and allowing the parameters to be functions of the covariate. Under the assumption that the observations conditional on x follow a GPD distribution then these two equations can be used to constrain the functional form of $\boldsymbol{\beta_0}(\cdot)$ and $\beta_1(\cdot)$, so that one can compare extreme quantiles e.g. $p < 0.10$, where it is assumed that the conditional distribution given the observed covariate follows a GPD. Examining the function form for Equations (9) and (10) one can see that the quantile regression parameters $\boldsymbol{\beta_0}$ and β_1 are functionally independent of the covariate but dependent on the exceedance probability, whereas the GPD parameters c, σ and ξ are functionally dependent of the covariate and independent of exceedance probability.

In the simplest case, suppose the shape (ξ) and threshold (c) parameters of the GPD are constant. If so, the form of the quantile regression, which is linear in x requires that the functional form for σ also be linear in x. In other words, $\sigma(x) = s_0 + s_1 \cdot x$. In this case we can derive the quantile regression parameters as a function of p. If we let $r(p) = \frac{1}{\xi(x)} \cdot \left\{ \frac{1}{p^{\xi(x)}} - 1 \right\}$ and $\sigma(x) = s_0 + s_1 \cdot x$ in Equation (10) and equate the coefficients of x in right hand sides of Equations (9) and (10) we can solve for the conditional regression intercept and slope, respectively, as:

$$\beta_0(p) = c + s_0 \cdot r(p)$$
$$\beta_1(p) = s_1 \cdot r(p). \tag{11}$$

These equations allow us to make a direct comparison between quantile regression and extreme value models for the case where the GPD shape parameter is assumed to be constant and the GPD scale parameter is assumed to be a linear combination of the model covariates plus an intercept (Int); for example, $\sigma_i = \text{Int} + \beta \cdot x_i$.

As mentioned, the set of TC intensities near the USA coast exceeding 96 kt (49 m/s) follows a GPD. Thus, in order to make a fair comparison with the quantile regression results, we need to first select those TCs that exceed this threshold. This reduces the set of TCs from 422 down to 79. The comparison between quantile regression and a GPD model will be done using only this subset of cyclones. The GPD parameters are specified following a Bayesian approach as outlined in Jagger and Elsner (2006) using the set of scaled NAO, SST, SOI, and

DOI: 10.1002/joc

Table I. Parameters from a GPD extreme value model of near-coastal TC intensity conditional on climate covariates. The full model includes all four covariates and the final model includes only SOI and SSN. Int is the intercept term. The p-values are one-sided.

	Full model			Final model		
Parameter	Mean	S.E.	p-value	Mean	S.E.	p-value
Scale Int (kt)	33.189	5.331	<0.001	30.591	4.445	<0.001
Scale NAO (kt/sd)	0.242	2.255	0.455	–	–	–
Scale SOI (kt/sd)	−7.746	2.858	0.002	−6.431	2.063	0.001
Scale SST (kt/sd)	−1.529	2.009	0.210	–	–	–
Scale SSN (kt)	5.396	3.310	0.025	5.805	3.705	0.032
Shape Int (kt)	−0.501	0.128	0.001	−0.459	0.106	<0.001

SSN covariates. Note that the significant covariates for the most extreme intensities will, in general, be different from those based on all the data as is used in the quantile regression approach described in the previous section.

Table I shows the posterior means and p-values for the parameters of the full model (all covariates) as well as the posterior mean and standard error (S.E.) from the final model. The final model includes only the covariates that were found significant in the full model. Those covariates are the SOI and SSN. The SOI is most significant and the scale value increases with decreasing SOI (toward El Niño conditions), that is; decreasing SOI (more El Niño like conditions) leads to stronger TCs given that the cyclone strength already exceeds 96 kt. This is consistent with the results shown in the previous section.

To close the loop on the comparison, we use Equations (10) and (11) together with the posterior mean values of the GPD parameters to derive quantile regression coefficients for quantiles $\tau = 0.1$, 0.25, 0.5, 0.75 and 0.9. We then compare these GPD-derived quantile coefficients with quantile coefficients estimated directly from an application of quantile regression on the subset of the most intense TCs and the same two covariates (SOI and SSN). Note again that the quantile regression performed in the previous section used all the data. The GPD-derived and quantile-regression-estimated coefficients are shown in Table II. The corresponding values are reasonably close and the GPD-derived parameter estimates are all within the 90% confidence interval on the quantile regression estimates.

This is the first comparison of results from these two different approaches to modeling extremes that we are aware of. It is our contention that each method has its place. In the GPD model we assume the data follow a distribution that allows us to make extrapolations to the largest quantiles (and beyond). With the quantile regression approach there is no need to specify a distribution for the data, but predicted quantiles have larger variance for the same set of data. Thus, as shown with the reduced data set, the GPD model is better suited for estimating the most extreme quantiles, whereas quantile regression, which uses all the data, is better suited for estimating changes in quantiles for middle quantiles at a level where TC intensities conditioned on the covariates do not conform to a GPD.

Table II. Parameter comparisons. The GPD-derived parameters are from a GPD extreme value model of TC intensity for the subset of cyclones exceeding 96 kt scaled using Equations (10) and (11). The quantile regression parameters are based on a quantile regression model using the same subset of extreme TC intensities. The upper and lower bounds are from the quantile regression using a 90% confidence interval.

	Quantile (τ)				
Coefficient	0.10	0.25	0.50	0.75	0.90
GPD Int	99.1	104.2	114.2	127.4	139.5
QR Int	101.1	104.9	112.9	125.2	134.4
Lower bound	97.8	103.1	108.5	120.0	128.6
Upper Bound	101.6	108.0	119.0	128.2	151.2
GPD SOI	−0.66	−1.73	−3.82	−6.60	−9.14
QR SOI	−1.83	−0.71	−3.24	−5.44	−7.06
Lower bound	−2.32	−4.29	−9.02	−6.86	−19.5
Upper bound	0.36	1.67	2.16	−1.45	−1.25
GPD SSN	0.60	1.56	3.45	5.95	8.25
QR SSN	0.24	1.09	1.33	3.04	7.67
Lower bound	−5.27	−1.86	−1.05	0.06	2.82
Upper bound	3.08	3.65	4.69	7.52	23.9

7. Summary and conclusions

Coastal TCs pose a serious threat to social and economic institutions. Statistical models derived from historical data provide a useful and skillful way to evaluate the risk of the next catastrophic cyclone. Extreme value theory provides parametric models for rare events and a justification for extrapolating to levels that are greater than what has been observed. Here we show that the quantile regression provides another way to model TC intensity conditional on climate variables. Quantile regression extends the ordinary least squares regression model to conditional quantiles (e.g. 90th percentile) of the response variable. We also showcase quantile regression models of near-coastal TC intensity conditional on seasonal values of SST, SOI, NAO, and SSN.

The influence of Atlantic SST, the Pacific El Niño, and the NAO on near-coastal TC intensity is in the direction anticipated from previous studies and is generally strongest for higher intensity quantiles. The influence of solar activity peaks near the median intensity level, but

DOI: 10.1002/joc

the relationship switches the sign for the highest quantiles. The principal findings about near-coastal hurricanes in USA are as listed below:

- TC intensity increases with SST for intensities in the upper quantiles. That is, the strongest TCs get stronger with increasing SST.
- TC intensity increases with the SOI in the mean and for quantiles in the range between the 50th and the 95th percentiles. However, at extreme quantile levels, TC intensity has a significant decreasing relationship with the SOI.
- TC intensity decreases with the NAO in the mean and for quantiles in the range between the 50th and 95th percentiles.
- TC intensity decreases with SSN for quantiles in the range between the 10th and 80th percentiles, but increases with SSN for the highest quantiles.
- The quantile regression approach to modeling extreme TC wind speeds is reconcilable with the parametric GPD approach described in Jagger and Elsner (2006). In comparison, the quantile regression approach is one way to model TC activity that makes it easier to interpret the results, but it will produce estimates with larger variability at the highest quantile levels.

Acknowledgements

We acknowledge the SIDC for the sunspot data and the US National Hurricane Center for the TC data. The work was supported by the US National Science Foundation (ATM-0738172) and Risk Prediction Initiative of the Bermuda Institute for Ocean Sciences (RPI-08-02-002).

References

Bremnes JB. 2004. Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. *Monthly Weather Review* **132**: 338–347.

Elsner JB, Bossak BH, Niu XF. 2001. Secular changes to the ENSO-US hurricane relationship. *Geophysical Research Letters* **28**: 4123–4126.

Elsner JB, Jagger TH. 2004. A hierarchical Bayesian approach to seasonal hurricane modeling. *Journal of Climate* **17**: 2813–2827.

Elsner JB, Jagger TH. 2006. Prediction models for annual US hurricane counts. *Journal of Climate* **19**: 2935–2952.

Elsner JB, Jagger TH. 2008. United States and Caribbean tropical cyclone activity related to the solar cycle. *Geophysical Research Letters* **35**: L18705. (DOI:10.1029/2008/GL034431, September 2008).

Emanuel KA. 1988. The maximum intensity of hurricanes. *Journal of the Atmospheric Sciences* **45**: 1143–1155.

Enfield DB, Mestas-Nunez AM, Trimble PJ. 2001. The Atlantic multidecadal oscillation and it's relation to rainfall and river flows in the continental U.S. *Geophysical Research Letters* **28**: 2077–2080.

Friederichs P, Hense A. 2007. Downscaling of extreme precipitation events using censored quantile regression. *Monthly Weather Review* **135**: 2365–2378.

Gray WM, Landsea CW, Mielke PW Jr, Berry KJ. 1992. Predicting Atlantic seasonal hurricane activity 6–11 months in advance. *Weather and Forecasting* **7**: 440–455.

Jagger TH, Elsner JB. 2006. Climatology models for extreme hurricane winds near the United States. *Journal of Climate* **19**: 3220–3226.

Jarvinen BR, Neumann CJ, Davis MAS. 1984. A tropical cyclone data tape for the North Atlantic Basin, 1886–1983: Contents, limitations, and uses, NOAA Tech. Memo. NWS NHC-22, Coral Gables, FL, 21.

Jones PD, Jónsson T, Wheeler D. 1997. Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and South-West Iceland. *International Journal of Climatology* **17**: 1433–1450.

Koenker R, Bassett GW. 1978. Regression quantiles. *Econometrica* **46**: 33–50.

Landsea CW, Anderson C, Charles N, Clark G, Dunion J, Fernandez-Partagas J, Hungerford P, Neumann C, Zimmer M. 2004. The Atlantic hurricane database reanalysis project: Documentation for 1851–1910 alterations and additions. In *Hurricanes and Typhoons: Past Present, and Future*, Murnane R, Liu K-B (eds). Columbia University Press; New York, NY: USA; 177–221.

Neumann CJ, Jarvinen BR, McAdie CJ, Hammer GR. 1999. *Tropical Cyclones of the North Atlantic Ocean, 1871–1998*. National Oceanic and Atmospheric Administration; Miami, FL: USA; 206.

Ropelewski CF, Jones PD. 1997. An extension of the Tahiti-Darwin Southern oscillation index. *Monthly Weather Review* **115**: 2161–2165.

Yu K, Lu Z, Stander J. 2003. Quantile regression: applications and current research areas. *Statistician* **52**: 331–350.

DOI: 10.1002/joc