

## Hurricane Clusters in the Vicinity of Florida

THOMAS H. JAGGER AND JAMES B. ELSNER

*Department of Geography, The Florida State University, Tallahassee, Florida*

(Manuscript received 18 May 2011, in final form 14 December 2011)

### ABSTRACT

Models that predict annual U.S. hurricane activity assume a Poisson distribution for the counts. Here the authors show that this assumption applied to Florida hurricanes leads to a forecast that underpredicts both the number of years without hurricanes and the number of years with three or more hurricanes. The underdispersion in forecast counts arises from a tendency for hurricanes to arrive in groups along this part of the coastline. The authors then develop an extension to their earlier statistical model that assumes that the rate of hurricane clusters follows a Poisson distribution with cluster size capped at two hurricanes. Hindcasts from the cluster model better fit the distribution of Florida hurricanes conditional on the climate covariates including the North Atlantic Oscillation and Southern Oscillation index. Results are similar to models that parameterize the extra-Poisson variation in the observed counts, including the negative binomial and the Poisson inverse Gaussian models. The authors argue, however, that the cluster model is physically consistent with the way Florida hurricanes tend to arrive in groups.

### 1. Introduction

Seasonal forecast models that predict U.S. hurricane activity by coastal region were developed in Elsner and Jagger (2006). The models capture the historical distribution of hurricane counts and are widely used by the insurance and finance industries. They are based on the Poisson distribution, where the rate is conditioned on climate variables including the Southern Oscillation index (SOI), North Atlantic Ocean sea surface temperature (SST), the North Atlantic Oscillation (NAO), and sunspot number.

Here we show that the Poisson assumption may not be adequate for describing hurricane frequency in the vicinity of Florida. Indeed, over the period 2000–11 there were only two years with hurricanes in Florida, but the 2004 and 2005 seasons featured a total of seven strikes to the state. This suggests Florida hurricane activity is clustered.

Here we investigate this further and show evidence for clustering by comparing a prediction model for hurricane counts with a prediction model for hurricane occurrence, where both models assume the hurricane

numbers follow a conditional Poisson distribution. Occurrence is whether or not there is at least one hurricane, and count is the total number of hurricanes. We find that the predicted number of hurricanes from the occurrence model is lower than the predicted number from the count model.

To be specific, Florida hurricane counts have greater variability (dispersion) than is expected from a Poisson model. This extra dispersion results in a greater number of years without hurricanes and a greater number of years with three or more hurricanes. We find no strong evidence of extra dispersion for hurricanes affecting the Gulf of Mexico or East Coast regions, however.

We then derive a model that accounts for clustering using a natural extension of the count and occurrence models. That is, the parameters of the cluster model are derived from the parameters of the count and occurrence models. We compare the fit of the cluster model with the fit from a negative binomial model and a Poisson inverse Gaussian model.

The paper is organized as follows. In section 2 we examine the hurricane counts by region. In section 3 we examine the evidence for clustering and find it in the vicinity of Florida only. In section 4 we explore the possibility that clustering results from the influence of climate factors on the underlying hurricane rate. In section 5 we develop a cluster model for the Florida counts

---

*Corresponding author address:* Thomas H. Jagger, Dept. of Geography, Rm. 323, Bellamy Bldg., 113 Collegiate Loop, P.O. Box 3062190, The Florida State University, Tallahassee, FL 32306-2190.  
E-mail: tjagger@blarg.net

that considers the number of clusters separate from the cluster size. In section 6 we show how the model parameters can be fit and demonstrate how the fit better matches the distribution of hurricane count years in Florida. In section 7 model results for Florida and for the Gulf Coast region are compared. In section 8 results for Florida using the cluster model are compared with results using two other models that parameterize the dispersion in observed counts without specifying clusters. A summary and conclusion are provided in section 9.

## 2. Hurricane counts by region

Hurricanes form only over certain parts of the ocean. Hurricanes originating from the same area often take similar paths. This grouping, or clustering, increases the potential for multiple landfalls in a given year above what one might expect from random events.

A statistical model for landfall probability may capture this clustering through a covariate like the NAO, which relates a steering mechanism (position and strength of the subtropical high) to regional hurricane activity (Elsner and Jagger 2006; Kossin et al. 2010). There could still be additional serial correlation that is not related to the covariates, however. A model that does not account for this extra variation will underestimate the potential for multiple hits in a season.

Following the method of Jagger and Elsner (2006) we consider three coastal regions: the Gulf Coast, Florida, and the East Coast (Fig. 1). The regions are large enough to capture enough hurricanes but are not so large as to include too many noncoastal strikes. A natural spline interpolation is used to obtain positions and wind speeds at 1-h intervals from the 6-h values (Jagger and Elsner 2006) for all tropical cyclones in the Atlantic basin hurricane database (HURDAT; Jarvinen et al. 1984). For tropical cyclones in the dataset, we note the maximum wind in each region. If the maximum wind exceeds  $33 \text{ m s}^{-1}$  then we count it as a hurricane for the region. A tropical cyclone that affects more than one region at hurricane intensity is counted in each region. Because of this, the sum of the regional counts is larger than the total count, although we are only concerned here with regional counts.

Figure 2 shows the time series and histograms of the regional counts. Counts by year range from 0 to 4 for the Gulf Coast, 0 to 5 for Florida, and 0 to 3 for the East Coast. There are no significant trends in the near-coastal hurricane rates. The most common counts are 0 and 1.

## 3. Evidence for clustering

The shape of the histograms in Fig. 2 suggests a Poisson distribution (Elsner and Schmertmann 1993;

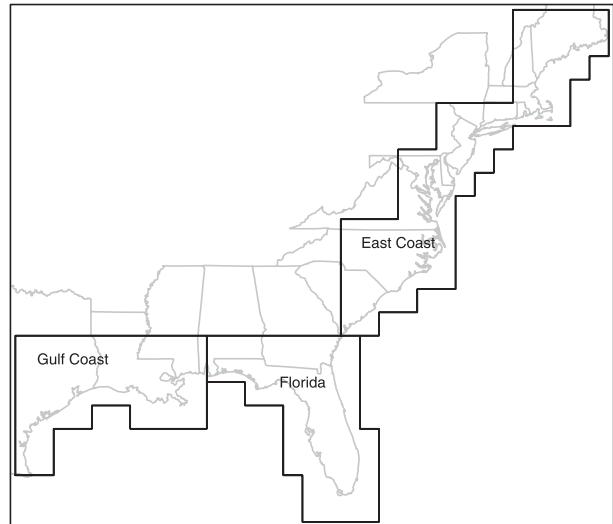


FIG. 1. Coastal hurricane regions. Regions are large enough to capture enough hurricanes but are small enough that the hurricanes correspond to actual threats to the coastal environments. Regional boundaries align with whole-number parallels and meridians. The region delineations are identical to those used in Jagger and Elsner (2006).

McTaggart-Cowan et al. 2008) as is typically the case with count data. The density function of the Poisson distribution indicates that the probability of obtaining a count  $x$  when the mean count (rate) is  $\lambda$  is given by

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}. \quad (1)$$

Thus the probability of no hurricanes is  $p(0) = e^{-\lambda}$ .

We begin by comparing the observed with expected number of years for two groups of hurricane counts. The groups include years with no hurricanes and years with three or more. The expected number is from a Poisson distribution. The idea is that for regions that show a cluster of hurricanes, the observed number of years with no hurricanes and the observed number of years with three or more hurricanes should be greater than the corresponding expected number. Said another way, a Poisson distribution with a hurricane rate estimated from counts over all years will underestimate the number of years with no hurricanes and the number with many hurricanes in regions with clustering.

Table 1 shows the results of the comparison for the Gulf Coast, Florida, and East Coast regions. For the Gulf and East Coast regions, the observed number of years is relatively close to the expected number of years in each of the three groups. For the Florida region, however, we find that the observed number of years exceeds the expected number of years in years with no hurricanes and in years with three or more hurricanes.

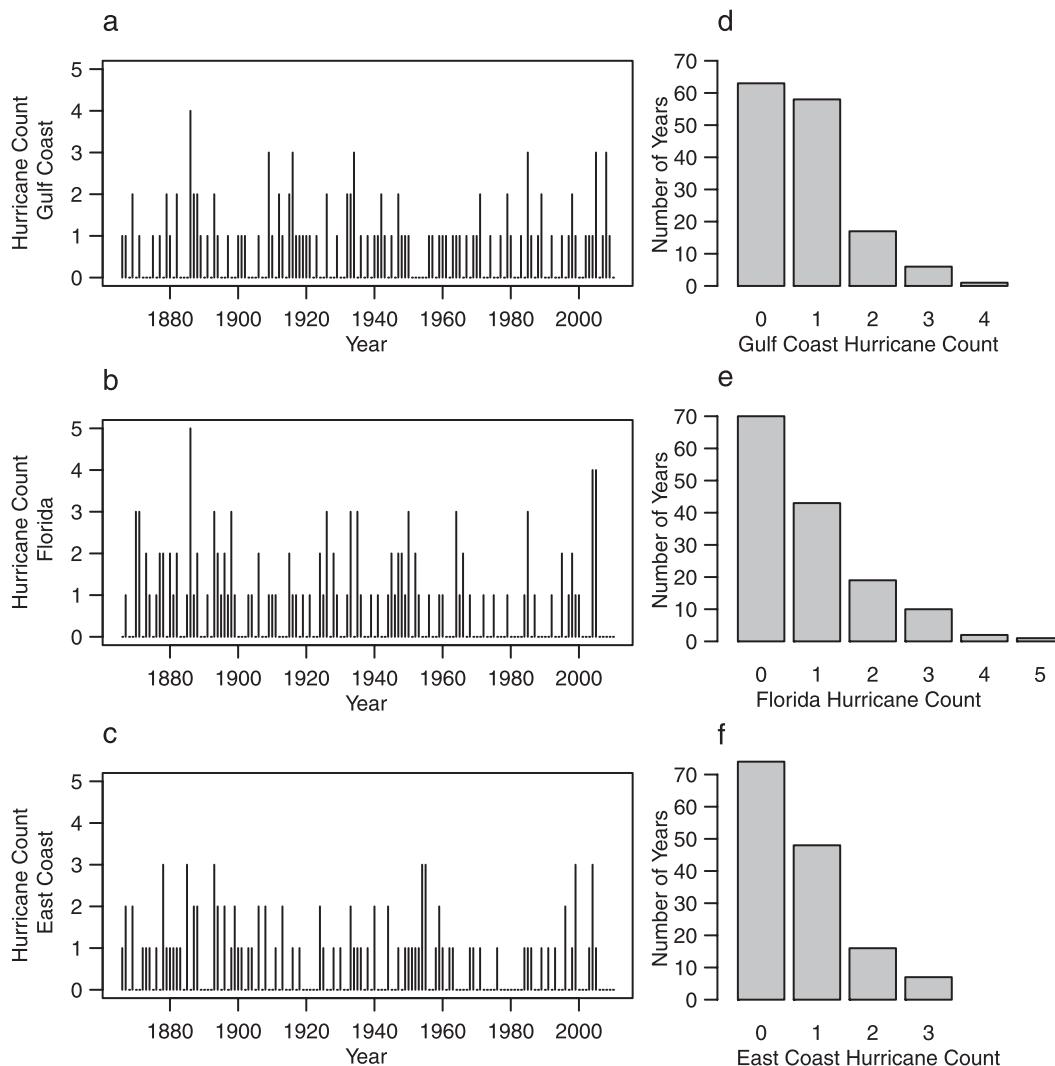


FIG. 2. Annual hurricane occurrence by region: time series of annual hurricane counts in the (a) Gulf Coast, (b) Florida, and (c) East Coast regions and (d)–(f) the distributions of hurricane counts in the corresponding regions.

The difference between the observed and expected numbers in each region is used to assess the statistical significance of the clustering. This is done using Pearson residuals and the  $\chi^2$  statistic. The Pearson residual is the difference between the observed count and expected rate divided by the square root of the variance. The  $p$  value is evidence in support of the null hypothesis of no clustering as indicated by no difference between the observed and expected numbers in each group. The  $p$  values for the Gulf and East Coast are greater than 0.05, indicating little support for clustering. In contrast the  $p$  value for the Florida region is 0.009 using the Pearson residuals and is 0.044 using the  $\chi^2$  statistic. These values provide moderate to convincing evidence that hurricane occurrences in the vicinity of Florida are not completely independent in time.

#### 4. Is clustering due to climate factors?

Having provided evidence of Florida hurricane clusters, we ask what might be causing it. Some of the extra variation in annual hurricane counts (more years with 0

TABLE 1. Observed  $O$  vs expected  $E$  number of hurricane years by count groups. The observed values are based on hurricanes over the period 1866–2010. The expected number of years is based on a Poisson distribution. The Pearson and  $\chi^2$  test statistics along with the corresponding  $p$  values are given.

Region	$O(\leq 0)$	$E(\leq 0)$	$O(\geq 3)$	$E(\geq 3)$	Pearson	$p$	$\chi^2$	$p$
Gulf	63	66.1	7	6.6	135	0.6858	0.264	0.894
Florida	70	61.7	13	8.1	187	0.0092	6.475	0.039
East	74	72.3	7	4.9	150	0.3440	1.172	0.557

counts and more years with 3 or more counts) might be due to variation in hurricane rates. We examine this possibility by assuming a conditional Poisson model in which the logarithm of the hurricane rate is estimated with a linear combination of known climate covariates and then grouping the model residuals into low and high sets as we did above.

Elsner and Jagger (2006) show that Florida hurricane activity depends on the NAO averaged during May and June and on the SOI averaged over the peak of the hurricane season (August–October) as an indicator of El Niño–Southern Oscillation (ENSO) but does not depend on Atlantic Ocean sea surface temperature or on sunspot number. The NAO is characterized by fluctuations in sea level pressure (SLP) differences. Index values for the NAO are calculated as the difference in SLP between Gibraltar and a station over southwest Iceland (Jones et al. 1997) and are obtained from the Climatic Research Unit (CRU). Monthly values can be considered to be an indicator of the strength and/or position of the subtropical Bermuda high (Elsner et al. 2001). We speculate that the relationship might result from a teleconnection between the midlatitudes and tropics whereby a below-normal NAO during the spring leads to dry conditions over the continents and to a tendency for greater summer/autumn midtropospheric ridging (enhancing the dry conditions). Ridging over the eastern and western sides of the North Atlantic basin tends to keep the midtropospheric trough, responsible for hurricane recurvature, farther to the north during the peak of the season (Elsner and Jagger 2006).

ENSO is characterized by basin-scale fluctuations in sea level pressure across the equatorial Pacific Ocean. The SOI is defined as the normalized sea level pressure difference between Tahiti and Darwin, Australia, and values are available back through the mid-nineteenth century. The SOI is strongly anticorrelated 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 tropical cyclones develop, and negative SOI values imply greater shear and subsidence. The monthly SOI values (Ropelewski and Jones 1987) are obtained from the CRU.

We follow the approach of Elsner and Jagger (2006) in using a generalized linear model (GLM) with the Poisson family. Thus the annual hurricane count model is

$$H_i \sim \text{dpois}(\lambda_i) \quad \text{with}$$

$$\log(\lambda_i) = \beta_0 + \beta_{\text{soi}}\text{SOI}_i + \beta_{\text{nao}}\text{NAO}_i,$$

where  $H_i$  is the hurricane count in year  $i$  simulated ( $\sim$ ) from a random Poisson distribution (dpois) with a rate  $\lambda$

TABLE 2. Observed vs expected number of hurricane years by count groups. The observed values are based on hurricanes over the period 1866–2010. The expected number of years are based on a GLM approach using the Poisson family with a logarithmic link function. The Pearson and  $\chi^2$  test statistics along with the corresponding  $p$  values are given.

Region	$O(\leq 0)$	$E(\leq 0)$	$O(\geq 3)$	$E(\geq 3)$	Pearson	$p$ value	$\chi^2$	$p$ value
Florida	70	63.6	13	9.0	172.4	0.042	3.85	0.134

that depends on the year  $i$ . The logarithm of the rate depends in a linear way on the SOI and NAO covariates.

The count model gives an expected number of hurricanes each year. This expectation is compared with the observed number as before. Results indicate that clustering is somewhat ameliorated by conditioning the rates on the covariates. In particular, the Pearson residual reduces to 172.4 with an increase in the corresponding  $p$  value to 0.042. The  $p$  value remains below 0.15, however, indicating that the conditional model, while an improvement, fails to capture the extra variation in Florida hurricane counts (Table 2).

To further examine this result we fit a double GLM that models the mean and the dispersion separately using the Poisson family with the logarithmic link function for both the mean and dispersion. The dispersion is the variance divided by the mean. The link function provides the relationship between the additive covariates and the Poisson rate. We fit the mean to the SOI and NAO and the dispersion only to a constant, since there was no indication of a relationship between the dispersion and either the SOI or the NAO. The estimation procedure provided in the R software package `dglm` (Dunn and Smyth 2009) alternates between one iteration for the mean and one iteration for the dispersion until convergence. The dispersion coefficient from the model has an estimated value of 1.28 with a 95% confidence interval of (1.01, 1.61). The dispersion value is 1 under the null hypothesis of no clustering.

Thus we find overdispersion in predictive hurricane counts in the vicinity of Florida even after accounting for the important predictors of activity. Note that we assume (reasonably) that both the NAO and the SOI, being derived from measured pressures, are accurate from 1866, as well as is the count of Florida hurricanes. We also fit the model iteratively using different start times for the data record and with a nonlinear response of the mean to the SOI. Results of this sensitivity exercise show a consistent estimate of the dispersion coefficient of about 1.3, giving us confidence that our evidence for clustering is not likely to be due to model misspecification or data bias.

### 5. Cluster model

Having presented evidence that Florida hurricanes arrive in clusters, we turn our attention to a model that can account for this. In the simplest case we assume that each hurricane cluster has one or two storms and the number of clusters (rather than the number of hurricanes) in each year follows a Poisson distribution with some underlying rate  $r$ . We also assign a probability of  $p$  that each cluster will have two hurricanes so that  $1 - p$  represents the probability of a single-hurricane “cluster.”

In other words, we assume that hurricane clusters arrive randomly over Florida, with each cluster contributing one or two hurricanes to the season count. A zero-count year means no clusters, and we can assume, without loss of generality, that each cluster contributes at least one hurricane. We further assume that the number of hurricanes in each cluster given  $p$  is statistically independent and that  $p$  does not change from cluster to cluster but may change from year to year. Caution: this parameter  $p$  should not be confused with a  $p$  value that we use as evidence in support of the null hypothesis in our statistical tests.

In formal terms, let  $N$  be the number of clusters in a given year and  $X_i, i = 1, \dots, N$  be the number of hurricanes in each cluster minus 1. Then the number of hurricanes in a given year is given by  $H = N + \sum_{i=1}^N X_i$ . Conditional on  $N, M = \sum_{i=1}^N X_i$  has a binomial distribution since the  $X_i$ s are independent Bernoulli variables and  $p$  is constant. That is,  $H = N + M$ , where the annual number of clusters  $N$  has a Poisson distribution with cluster rate  $r$ , and  $M$  has a binomial distribution with proportion  $p$  and size  $N$ . Here the binomial distribution describes the number of occurrences of at least one hurricane in a sequence of  $N$  independent years, with each year having a probability  $p$  of observing at least one hurricane.

In summary our cluster model has the following properties:

- 1) The expected number of hurricanes  $E(H) = r(1 + p)$ .
- 2) The variance of  $H$  is given by

$$\begin{aligned} \text{var}(H) &= E[\text{var}(H | N)] + \text{var}[E(H | N)] \\ &= E\{N[p(1 - p)]\} + \text{var}[(1 + p)N] \\ &= rp(1 - p) + r(1 + p)(1 + p) \\ &= r(1 + 3p). \end{aligned}$$

- 3) The dispersion of  $H$  is given by  $\text{var}(H)/E(H) = \phi = (1 + 3p)/(1 + p)$ , which is independent of cluster rate. Solving for  $p$  gives  $p = (\phi - 1)/(3 - \phi)$ .
- 4) The probability mass function for the number of hurricanes  $H$  is

$$\begin{aligned} P(H = k | r, p) &= \sum_{i=0}^{\lfloor k/2 \rfloor} \text{dpois}(k - i, r) \text{dbinom}(i, k - i, p); \\ k &= 0, 1, \dots, \text{ and} \end{aligned}$$

$$\begin{aligned} P(H = 0 | r, p) &= e^{-r} \text{dpois}(k - i, r) \\ &= e^{-r} \frac{r^{k-i}}{(k - i)!} \text{dbinom}(i, k - i, p) \\ &= \frac{k - i}{i} p^i (1 - p)^{k-2i}. \end{aligned}$$

- 5) The model has two parameters  $r$  and  $p$ . A better parameterization is to use  $\lambda = r(1 + p)$  with  $p$  to separate the hurricane frequency from the cluster probability. The parameters do not need to be fixed and can be functions of the covariates.
- 6) When  $p = 0, H$  is Poisson, and when  $p = 1, H/2$  is Poisson, the dispersion is 2, and the probability that  $H$  is even is 1.

Next we consider a method to estimate the model parameters  $r$  and  $p$ .

### 6. Parameter estimation

Our goal is a hurricane count distribution. For that we need an estimate of the annual cluster rate  $r$  and the probability  $p$  that the cluster size is two. Continuing with the GLM approach we separately estimate the annual hurricane frequency  $\lambda$  and the annual cluster rate  $r$ . The ratio of these two parameters minus 1 is an estimate of the probability  $p$ .

This is reasonable if  $p$  does not vary much, since the annual hurricane count variance is proportional to the expected hurricane count [i.e.,  $\text{var}(H) = r(1 + 3p) \propto r \propto E(H)$ ]. Above, we estimated the parameters of the annual count model using Poisson regression, which implies that the variance of the count is proportional to the expected count. Thus, under the assumption that  $p$  is constant, Poisson regression can be used for estimating  $\lambda$  in the cluster model.

As before we regress the cluster rate onto the predictors NAO and SOI. The model is given by

$$N_i \sim \text{dpois}(r_i) \quad \text{with} \tag{2}$$

$$\log(r_i) = \alpha_0 + \alpha_1 \text{SOI}_i + \alpha_2 \text{NAO}_i + \varepsilon_i.$$

The parameters of this annual cluster count model cannot be estimated directly, since the observed hurricane count does not furnish information about the number of clusters.

Consider the observed set of annual Florida hurricane counts. Because the annual frequency is very small, the

majority of years have either no hurricanes or a single hurricane. We create a “reduced” dataset by using an indicator of whether there was at least one hurricane. Formally let  $I_i = I(H_i > 0) = I(N_i > 0)$ ; then  $I$  is an indicator of the occurrence of a hurricane cluster for each year. We assume that  $I$  has a binomial distribution with size parameter of 1 and a proportion equal to  $\pi$ . This leads to a logistic regression model for  $I$ .

Note that since  $\exp(-r)$  is the probability of no clusters [see Eq. (1)],  $\pi = 1 - \exp(-r)$ . Thus the cluster rate is  $r = -\log(1 - \pi)$ . If we use a logarithmic link function on  $r$ , then  $\log(r) = \log[-\log(1 - \pi)] = \text{clog log}(\pi)$ , called the complementary log–log function, which is a natural choice for a model link since its domain matches the range of the distribution function’s mean. Thus we model  $I$  using the clog log function to obtain  $r$ .

In short, our cluster model is a combination of two models, one for the counts and another for the clusters. In formal terms, it is given by

$$I_i \sim \text{dbern}(\pi_i) \quad \text{with} \quad (3)$$

$$\text{clog log}(\pi_i) = \alpha_0 + \alpha_1 \text{SOI}_i + \alpha_2 \text{NAO}_i + \varepsilon_i,$$

where dbern is the Bernoulli distribution with mean  $\pi$ . The covariates are the same as those used in the cluster count model. Given these equations, one has the following relationships for  $r$  and  $p$ :

$$\log[\hat{r}(1 + \hat{p})] = \hat{\beta}_0 + \hat{\beta}_1 \text{SOI} + \hat{\beta}_2 \text{NAO} \quad \text{and} \quad (4)$$

$$\log(\hat{r}) = \hat{\alpha}_0 + \hat{\alpha}_1 \text{SOI} + \hat{\alpha}_2 \text{NAO}. \quad (5)$$

By subtracting the coefficients in Eq. (5) for the annual cluster count model from those in Eq. (4) for the annual hurricane count model, one obtains a regression model for the probabilities given by

$$\begin{aligned} \log(1 + \hat{p}) &= \hat{\beta}_0 - \hat{\alpha}_0 + (\hat{\beta}_1 - \hat{\alpha}_1) \text{SOI} \\ &+ (\hat{\beta}_2 - \hat{\alpha}_2) \text{NAO}. \end{aligned} \quad (6)$$

## 7. Model diagnostics

We diagnose the model in the following ways. First we compare the fitted values from the hurricane count and cluster occurrence models. Let  $H_i$  be the hurricane count in year  $i$  and  $\hat{r}_i$  and  $\hat{\lambda}_i$  be the fitted annual cluster frequency and annual hurricane frequency, respectively. Then let  $\tau_0$  be a test statistic given by

$$\tau_0 = \frac{1}{n} \sum_{i=1}^n (H_i - \hat{r}_i) = \frac{1}{n} \sum_{i=1}^n (\hat{\lambda}_i - \hat{r}_i). \quad (7)$$

The value of  $\tau_0$  will be greater than 1 if there is clustering. We test the significance of  $\tau_0$  by generating random samples of length  $n$  from a Poisson distribution with rate  $\lambda_i$  and computing  $\tau_j$  for  $j = 1, \dots, J$ , where  $J$  is the number of samples. A one-sided  $p$  value or test of the null hypothesis that  $\tau_0 \leq 0$  is the proportion of simulated  $\tau$ s that are at least as large as  $\tau_0$ .

Second, we can use the fitted values of  $\hat{r}$  and  $\hat{\lambda}$  to estimate  $\hat{p}$ . A linear regression through the origin of the fitted hurricane count frequency to the hurricane cluster frequency under the assumption that  $p$  is constant yields an estimate of  $1 + p$ . Third, we use Eq. (6) to get an estimate for  $\hat{p}$  when we assume  $p$  is a function of the covariates.

We demonstrate the reasonableness of the model by applying it to data from the Gulf Coast and to data from Florida. As shown previously there is evidence for clustering of hurricanes in Florida but not in the Gulf region. In Florida  $\tau_0 = 0.104$  with a  $p$  value of 0.024, whereas along the Gulf Coast  $\tau_0 = -0.062$  with a  $p$  value of 0.797, in clear agreement with the evidence presented earlier indicating hurricane clusters in the vicinity of Florida but not along the Gulf Coast.

A linear regression through the origin of the fitted count rate on the cluster rate under the assumption that  $p$  is constant yields an estimate for  $1 + p$ . We plot the annual count and cluster rates and draw the regression line for Florida and Gulf Coast hurricanes in Fig. 3. The black line is the  $y = x$  line, and we expect cluster and hurricane rates to align along this axis if there is no clustering. The red line is the regression of the fitted hurricane rate onto the fitted cluster rate with the intercept set to zero. The slope of the line is an estimate of  $1 + p$ .

The slope is 1.138 for the Florida region, giving 0.138 as an estimate for  $p$ . The regression slope is 0.942 for the Gulf Coast region, which we interpret as a lack of evidence for hurricane clusters in this region.

Our focus is now exclusively on Florida hurricanes. We continue by looking at the coefficients from the count and cluster models. The output coefficients are shown in Table 3. Results show that the NAO and SOI covariates are significant in the hurricane count model but that only the NAO is significant in the hurricane cluster model.

The difference in coefficient values from the two models is an estimate of  $\log(1 + p)$ . The difference in the NAO coefficient is 0.043, and the difference in the SOI coefficient is 0.035, indicating that the NAO contributes slightly more to clustering in the vicinity of Florida than does ENSO. Lower values of the NAO lead to a larger rate increase for the Poisson model relative to the binomial model. Using a bootstrap procedure, we find  $p$  significantly greater than zero, but neither the NAO nor the SOI are significant (at the  $\alpha = 0.1$  level) in explaining  $p$ .

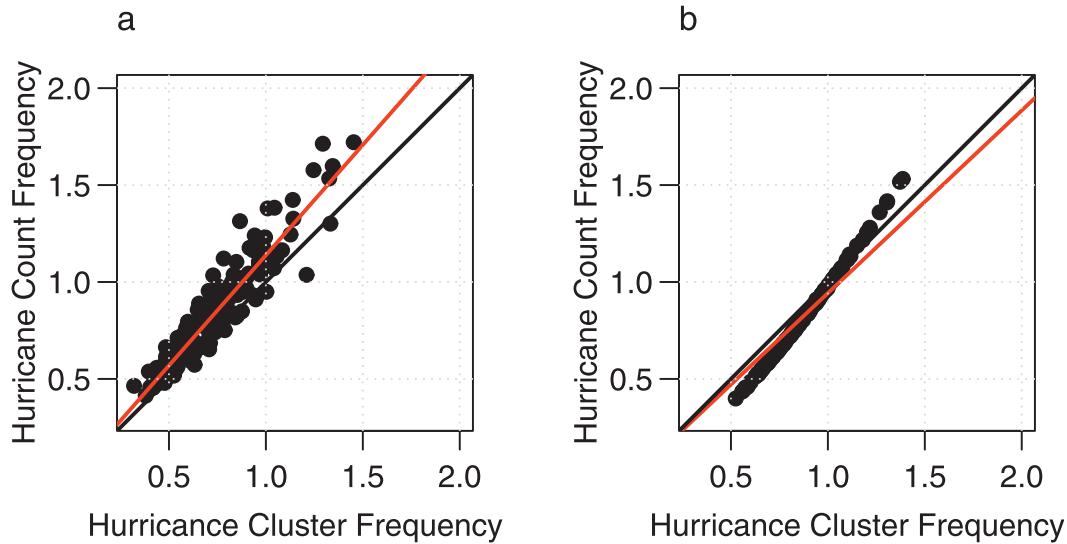


FIG. 3. Scatterplot of count vs cluster rates using hurricanes in the (a) Florida and (b) Gulf Coast regions. The black line is  $y = x$ . The red line is the regression of fitted hurricane rate onto fitted cluster rate with the intercept equal to zero.

The cluster model is used to hindcast the distribution of hurricane counts each year over the period 1866–2010. Two sets of hindcasts, one with  $p = 0.138$  estimated from the slope method and the other with  $p = 0.160$  estimated using the double GLM, are compared with a set of hindcasts from a Poisson model (a cluster model with  $p = 0$ ) in Table 4. For the double-GLM approach,  $p$  is based on a dispersion value of 1.28.

Results show that the cluster model, with  $p$  estimated either way, fits the observed counts better than does the Poisson model, but particularly for the low- and high-count years. Using the Poisson count model, the difference between the observed and expected is greater than six years for years without storms and is about four years for years with three or more hurricanes. With the cluster model, the differences are within one or two years.

### 8. Comparison with other models

A final diagnostic is a comparison of our cluster model with two standard overdispersed models, the negative binomial (NB) and the Poisson inverse Gaussian (PIG). The underlying NB and PIG distributions are formed from a continuous mixture of Poisson distributions over the Poisson rate parameter. The rate has a gamma distribution in the NB model and an inverse Gaussian distribution in the PIG model. These distributions and their maximum likelihood estimators are provided in the `gamlss` and `gamlss.dist` R packages (Rigby and Stasinopoulos 2005; Stasinopoulos et al. 2011).

The NB and PIG models have two parameters  $\mu$  and  $\sigma$ , the location and scale parameters, respectively. Each of these parameters may be specified as functions of the model covariates. For the NB distribution the mean is equal to  $\mu$  and the variance is equal to  $\mu(1 + \sigma)$ . For the PIG model, the mean is also equal to  $\mu$  but the variance is equal to  $\mu(1 + \mu\sigma)$ . The PIG model has slightly fatter tails than the NB model since the mixing distribution has a larger kurtosis. For the NB distribution we use the “NBII” formulation of the `gamlss` package since it can fit the mean separately from the dispersion.

The fitted model means and variances from the NB and PIG models are not significantly different than those from the cluster model. The model coefficients for the  $\log(\mu)$  are similar to those under the Poisson model. The coefficients on the climate covariates for the  $\log(\sigma)$  component were not significant in either model, and so  $\sigma$  was fit as a constant. For the NB model  $\sigma = 0.234$  leads to  $p = 0.133$  in the cluster model, close to the 0.138

TABLE 3. Covariate coefficients for the hurricane count model and the cluster model.

Term	Estimate	Std error	$z$ value	$p$ value
Hurricane rate [ $r(1 + p)$ ]				
Intercept	-0.268	0.105	-2.546	0.011
NAO	-0.226	0.090	-2.502	0.012
SOI	0.056	0.030	1.856	0.063
Cluster rate ( $r$ )				
Intercept	-0.419	0.135	-3.113	0.002
NAO	-0.269	0.121	-2.227	0.026
SOI	0.021	0.041	0.523	0.601

TABLE 4. Observed vs expected number of hurricane years for Florida by count groups from the Poisson model and the cluster model with two different estimates of  $p$ . The observed values are based on hurricanes over the period 1866–2010.

Model	$p$	$O(\leq 0)$	$E(\leq 0)$	$O(\geq 3)$	$E(\geq 3)$	Pearson	$p$ value	$\chi^2$	$p$ value
Poisson	0	70	63.6	13	9.0	172.4	0.042	3.85	0.134
Cluster	0.138	70	70.0	13	11.5	139.7	0.586	0.23	0.902
Cluster	0.160	70	71.0	13	11.8	135.9	0.673	0.13	0.957

estimated by the slope method, and for the PIG model  $\sigma = \exp(-1.334)$  leads to a mean estimate of  $p = 0.130$ .

Table 5 shows the observed and expected counts from the four models grouped by number of years with  $H$  hurricanes. The results show all three alternatives would be an improvement over the Poisson model as they each adequately handle the dispersion in the tails of the observed count distribution. Only the cluster model uses a formulism that mimics clustering in the occurrence of hurricanes, however.

Last, it is informative to compare hindcasts of Florida hurricanes on a graph using the Poisson and cluster models. Here we set  $p = 0.138$  for the cluster model. We can use the same two-model formulation for the Poisson model by setting  $p = 0$ . Results are shown in Fig. 4 and demonstrate that the cluster model fits the observed counts better than does the Poisson model. Again, this is most evident at the low- and high-count years.

## 9. Summary and conclusions

Over the period 2000–11 there have only been two years with hurricanes in Florida, but the 2004 and 2005 seasons featured a total of seven hurricane strikes to the state. Seasonal forecast models that predict U.S. hurricane activity assume a Poisson distribution (Elsner and Jagger 2006). Here we show that the Poisson assumption leads to a forecast that underpredicts both the number of years without hurricanes and the number of years with three or more hurricanes in the vicinity of Florida. This lack of fit

TABLE 5. Observed and expected hurricane years for Florida by count year. The expected counts are from models whose location parameters vary with the NAO and SOI. The cluster model uses a value for  $p$  that is based on the slope method. The models are based on data over the period 1866–2010.

No. hurricanes	Obs no. yrs	Hurricane models			
		Poisson	Cluster	NB	PIG
0	70	63.6	70.0	68.9	68.5
1	43	50.6	42.7	44.9	45.8
2	19	21.8	20.8	20.0	19.7
3	10	6.8	7.8	7.5	7.2
4	2	1.7	2.6	2.6	2.5
5+	1	0.5	1.1	1.1	1.3

arises because of clustering of hurricanes along this part of the coast.

Here we developed an extension to our earlier model (Elsner and Jagger 2006) that assumes that the rate of hurricane clusters follows a Poisson distribution with the size of the cluster limited to two hurricanes. The model is shown to better fit the distribution of Florida hurricanes conditional on the climate covariates including the NAO and SOI. Results are similar to mixture models that parameterize the extra variation, including the negative binomial and Poisson inverse Gaussian models. We argue, however, that as a natural extension to the Poisson distribution—which is a good fit in general—our cluster model provides a better physical basis than the overdispersed alternatives. Moreover, the model could be used on lightning and tornado data, for which a mixture distribution might fail.

The model could be extended to include cluster sizes that are greater than two. The authors have derived both likelihood estimators and Bayesian posterior estimates for the model parameters in addition to the moment estimators described in this paper. The point estimates are similar. In particular, the Bayesian approach is useful for estimating the regression coefficients and credible intervals for the probability parameter  $p$ . In our case, the posterior median for  $p$  is 0.124 with a 90% credible interval of (0.032, 0.257).

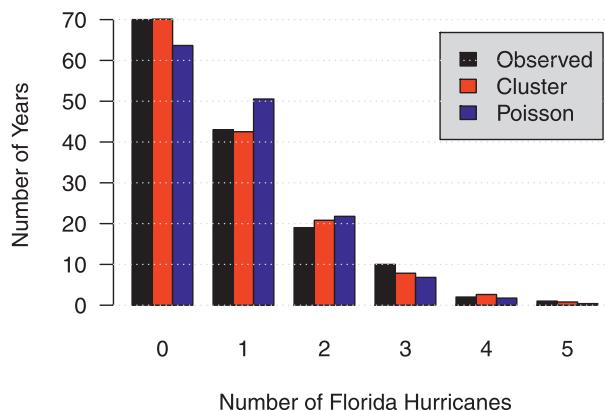


FIG. 4. Observed vs expected number of Florida hurricane years. The expected numbers are based on a cluster model ( $p = 0.138$ ) and on a Poisson model ( $p = 0$ ). The values are based on hurricanes over the period 1866–2010.

*Acknowledgments.* The research was sponsored by the Risk Prediction Initiative of the Bermuda Institute for Ocean Sciences and by a contract from the Strategic Environmental Research and Development Program (SERDP SI-1700). All statistical analyses were performed using the software environment known as R (<http://www.r-project.org>).

## REFERENCES

- Dunn, P. K., and G. K. Smyth, 2009: DGLM: Double generalized linear models. R package version 1.6.1, 13 pp. [Available online at <http://cran.r-project.org/web/packages/dglm/dglm.pdf>.]
- Elsner, J. B., and C. P. Schmertmann, 1993: Improving extended-range seasonal predictions of intense Atlantic hurricane activity. *Wea. Forecasting*, **8**, 345–351.
- , and T. H. Jagger, 2006: Prediction models for annual U.S. hurricane counts. *J. Climate*, **19**, 2935–2952.
- , B. H. Bossak, and X.-F. Niu, 2001: Secular changes to the ENSO–U.S. hurricane relationship. *Geophys. Res. Lett.*, **28**, 4123–4126.
- Jagger, T. H., and J. B. Elsner, 2006: Climatology models for extreme hurricane winds near the United States. *J. Climate*, **19**, 3220–3236.
- Jarvinen, B. R., C. J. Neumann, and M. A. S. Davis, 1984: A tropical cyclone data tape for the North Atlantic basin, 1886–1983: Contents, limitations, and uses. NOAA Tech. Memo. NWS NHC 22, 24 pp. [Available online at <http://www.nhc.noaa.gov/pdf/NWS-NHC-1988-22.pdf>.]
- Jones, P. D., T. Jonsson, and D. Wheeler, 1997: Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *Int. J. Climatol.*, **17**, 1433–1450.
- Kossin, J. P., S. J. Camargo, and M. Sitkowski, 2010: Climate modulation of North Atlantic hurricane tracks. *J. Climate*, **23**, 3057–3076.
- McTaggart-Cowan, R., G. D. Deane, L. F. Bosart, C. A. Davis, and T. J. Galarneau Jr., 2008: Climatology of tropical cyclogenesis in the North Atlantic (1948–2004). *Mon. Wea. Rev.*, **136**, 1284–1304.
- Rigby, R. A., and D. M. Stasinopoulos, 2005: Generalized additive models for location, scale and shape, (with discussion). *Appl. Stat.*, **54**, 507–554.
- Ropelewski, C. F., and P. D. Jones, 1987: An extension of the Tahiti–Darwin Southern Oscillation index. *Mon. Wea. Rev.*, **115**, 2161–2165.
- Stasinopoulos, M., B. C. A. Rigby, G. Heller, R. Ospina, and N. Motpan, cited 2011: gamlss.dist: Distributions to be used for GAMLSS modelling. R package version 4.0-5. [Available online at <http://CRAN.R-project.org/package=gamlss.dist>.]