Environmental signals in property damage losses from hurricanes

Thomas H. Jagger and James B. Elsner

Abstract We demonstrate the existence of environmental signals in property damage losses from hurricanes affecting the United States. The methodology is based on a random sums model, where the number of damaging hurricane events is modeled separately from the amount of damage per event. It is shown that when the springtime north-south surface pressure gradient over the North Atlantic is weaker than normal, the Atlantic ocean is warmer than normal, there is no El Niño event, and sunspots are few, the probability of at least one loss event increases. However, given at least some losses, the magnitude of the damage per annum is correlated only to ocean temperatures in the Atlantic. The magnitude of damage losses at a return period of 50 years is largest under a scenario featuring a warm Atlantic Ocean, a weak North Atlantic surface pressure gradient, El Niño, and few sunspots.

1 Introduction

Hurricanes at landfall generate large financial losses. Hurricane climatologists have developed statistical models to anticipate the level of coastal hurricane activity from independent climate signals. In addition, these models can be used to account for changes in hurricane intensity. Thus we posit that it should be possible to detect environmental signals in historical damages. Our purpose here is to show to what

Thomas H. Jagger
Department of Geography, Florida State University, Tallahassee, FL, USA 32306
Tel.: 303-759-9776
Fax: 850-644-5913
e-mail: tjagger@fsu.edu

Thomas H. Jagger · James B Elsner
Department of Geography, Florida State University, Tallahassee, Florida, USA 32306
e-mail: jelsner@fsu.edu

extent environmental patterns that are known to influence the frequency and intensity of hurricanes over the North Atlantic can be detected in long records of damage losses from hurricanes along the U.S. coastline (Gulf and Atlantic).

Elucidating this connection between ambient environmental conditions and future economic threats from natural hazards is an important new and interesting line of inquiry (Leckebusch, et al., 2007). Lane (2008a) writes that insurance markets and capital markets are converging that they are borrowing techniques from each other to access capital and to assess and deal with risk. As the participation of the financial markets becomes more important so does the need to provide investors with sufficient and timely information. Early information is particularly valuable if it aids investors in predicting the number or severity of loss events. For example, Lane (2008b) analyzes investment returns for Insurance Linked Securities (i.e., Catastrophe Bonds) and notes, unsurprisingly, that different loss magnitudes and loss timing patterns are controlled by nature and that this can significantly affect investment results.

Although others have shown environmental signals in damage losses using bivariate relationships including El Niño and wind shear (Katz, 2002; Saunders and Lea, 2005), this paper is the first to look at the problem from a multivariate perspective. It is based on a recent study that uses pre-season environmental conditions to anticipate insured losses before the start of the hurricane season (Jagger et al., 2008). Here we focus on the set of predictors shown to be directly to U.S. hurricane activity and intensity. (Jagger and Elsner, 2006). These predictors are the most likely candidates to elucidate the multivariate relationships between the environment and losses from hurricanes. Based on recent research into Atlantic hurricane activity (Elsner and Jagger, 2008) we also introduce the solar cycle as a potentially important covariate in estimating losses.

The strategy is to model annual total economic losses associated with hurricanes as a compound stochastic point process. The process is compound since the number of hurricanes causing damage in a given year is fit using a Poisson distribution, while the amount of damage is fit using a log-normal distribution. Loss totals are thus represented as a random sum, with variations in total losses decomposed into two sources, one attributable to variations in the frequency of events and another to variations in the amount of damage from individual events. We also consider a model for losses over a longer time horizon using a generalized Pareto distribution for the amount of losses coupled with a Poisson distribution for the number of loss events exceeding a threshold amount.

We begin with an examination of the normalized damage loss data and the data associated with climate patterns. We then describe the overall modeling approach and conclude with forecasts of annual and maximum losses for a variety of climate scenarios.

2 Normalized damage losses: 1900–2007

We obtain normalized hurricane damage data from the work of Pielke et al. (2008). The normalization attempts to adjust damage amounts to what they would be if the hurricane struck in the year 2005 by accounting for inflation and changes in wealth and population, plus an additional factor that represents a change in the number of housing units that exceeds population growth between the year of the loss and 2005. The methodology produces a longitudinally consistent estimate of economic damage from past tropical cyclones affecting the U.S. Gulf and Atlantic coasts.

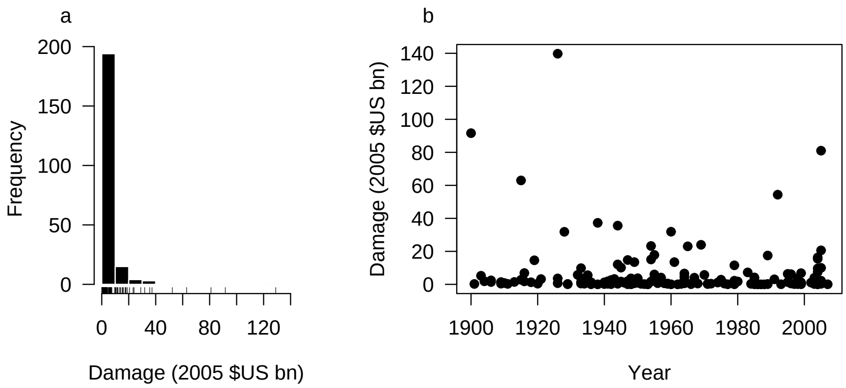
Economic damage is the direct losses associated with the hurricane's impact. It does not include losses due to business interruption or other macroeconomic effects including demand surge and mitigation. Details and caveats for two slightly different normalization procedures are provided in Pielke et al. (2008). Here we focus on the data set from the Collins/Lowe methodology, but note that both data sets are quite similar. Results presented in this study are not sensitive to the choice of data set.

We extend the data by adding the estimated economic damage losses from the three tropical cyclones during 2006 and 2007. The damage estimates are those reported in the *National Hurricane Center* (NHC) storm summaries and derived by the NHC by the *American Insurance Services* and the *Property Claim Services*. This is the same primary data source used in both normalization methods described in Pielke et al. (2008). We did not adjust these losses at this time. There were six tropical cyclones that caused at least some damage in the United Stated during this two-year period, but loss levels were quite small, especially when compared with the losses experienced in 2004 and 2005. In fact loss levels for three of the six tropical storms were below the $25 million reporting threshold (Alberto in 2006, and Barry and Gabrielle in 2007).

Tropical storm Ernesto in 2006 struck southern Florida and North Carolina. Total direct damage losses are estimated at $500 mn (million). We estimate that 4/5ths of those losses occurred in North Carolina where the storm was stronger at landfall. An estimate of the total property damage from tropical storm Erin and hurricane Humberto, both of which hit Texas in 2007, is $35 mn and $50 mn, respectively. The NHC suggests that the low damage total from Humberto was probably due to its small size and the relatively unpopulated area subjected to Humberto. In addition, the large losses in the same area from Hurricane Rita in 2005 and may have moderated the amount of damage that could have been done by Humberto. We make no attempt to normalize the losses from 2006 and 2007.

Here we assume that multiple landfalls from a single tropical cyclone produce multiple loss events. For example, in 1992 Hurricane Andrew produced a $52 bn (billion) loss in southeast Florida and a separate $2 bn loss in Louisiana. When multiple landfall events are included, the updated data set contains 221 loss events from 210 separate tropical cyclones over the period 1900–2007. Figure 1 shows the distribution and time series of the damages from all loss events. The histogram bars indicate the percentage of events with losses in groups of $10 bn. The distribution is highly skewed with 88% of the events having losses less than or equal to $10 bn and 95% of the events having losses less than $20 bn. The worst loss occurred with

Fig. 1 (a) Distribution of per storm damage losses from hurricanes in the United States (excluding Hawaii). The distribution is highly skewed with a few events generating very large damage losses. (b) Time series of the damage losses. Individual years may have more than one loss event.



the 1926 hurricane that struck southeast Florida creating an estimated damage loss adjusted to 2005 dollars of $129 bn. The Galveston hurricane of 1900 ranks second with an estimated loss of $99 bn and hurricane Katrina of 2005 ranks third with an estimated total loss of $81 bn. Years with more than one loss have more than one dot in the time series plot. There is large year-to-year variability but no obvious long-term trend, although here the data are not disaggregated into loss amount and the number of loss events.

The damage loss exceedances are shown in Table 1. Of the 221 loss events from 1900–2007, 169 exceeded $100 mn in losses and 28 of these exceeded $10 bn. The distribution of losses is similar using the Collins/Lowe (CL) method and the Pielke/Landsea (PL) methods, although the Collins/Lowe method tends to result in somewhat larger losses. The two events producing losses less than $1 mn include Gustav in 2002 and Dean in 1995. Another way to examine the data is to look at the total amount of loss for storms exceeding the Saffir-Simpson lower intensity threshold. Table 2 shows losses in billions of U.S. dollars from 1900–2007, inclusive. For example, category 0 is for the minimum tropical storm threshold (17 ms^{-1}), and category 1 is for the minumum hurricane threshold (33 ms^{-1}). So from this table, all tropical storms accounted for of $1103.9 bn 2005 adjusted $US with hurricanes accounting for 1063.1 bn. Here again we see the similarity in the two data sets and that category 4's and 5's, although rare, have historically accounted for nearly 50% of all losses.

Figure 2 shows the annual number of loss events and their distribution. There are five years with 6 loss events, the most recent being 2005. The annual rate of loss events is 1.94 events/yr with a variance of 2.48 $(\text{events/yr})^2$. There is a distinct upward trend in the number of loss events attributable to some extent to an increase

Table 1 Damage exceedances ($US adjusted to 2005$). The values are the number of events exceeding various damage loss thresholds from 1900–2007 inclusive.

Exceedance $US (2005)	Number Events (CL)[a]	Number Events (PL)[b]
1 million	219	219
10 million	207	207
100 million	169	169
1 billion	98	94
10 billion	28	27
100 billion	1	1

[a] Collins and Lowe data set
[b] Pielke and Landsea data set.

Table 2 Cumulative losses by Saffir-Simpson scale. ($US bn adjusted to 2005)

Category (Saffir/Simp)	Cumulative Losses (CL)	Cumulative Losses (PL)
0	1103.9	1125.1
1	1063.1	1084.5
2	1022.7	1045.6
3	941.4	964.4
4	533.1	557.3
5	79.4	79.3

in coastal population. As population increases so do the number of loss events from the weaker tropical cyclones. Indeed, prior to 1950 the number of loss events from tropical storms was 6% of the total number of events. This increases to nearly 38% from 1950 onward. In the present work we focus on the set of large losses from the stronger tropical cyclones. There is significant positive skewness in per storm damage amounts. If we transform the data using base-10 logarithms then the loss frequency distribution becomes more symmetric.

3 Climate and solar factors

Statistical relationships between U.S. hurricane activity and climate are well established (Elsner et al., 2004; Murnane et al., 2000). More importantly for the present work, Jagger et al. (2001) and Jagger and Elsner (2006) model the wind speeds of hurricanes at or near landfall and show that the exceedance probabilities (e.g., wind speeds in excess of 50 ms^{-1}) vary appreciably with the phase of the ENSO, the NAO, and Atlantic sea-surface temperature (SST). Recent work has also shown a

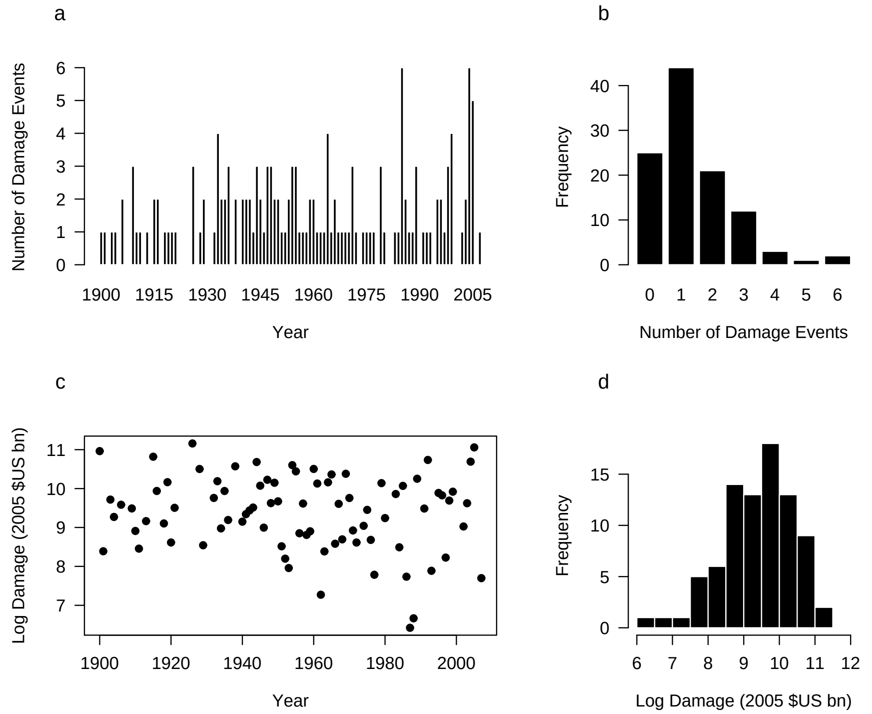


Fig. 2 Annual total damage losses. (a) Time series and (b) distribution of loss events and the (c) time series and (d) distribution of the number of the logarithm (base 10) of annual total damage losses

linkage between U.S. hurricanes and sunspot numbers (SSN) (Elsner and Jagger, 2008).

The ENSO can be characterized by basin-scale fluctuations in sea-level pressure between Tahiti and Darwin. Although noisier than equatorial Pacific ocean temperatures, pressure values are available back to 1900. The Southern Oscillation Index (SOI) is defined as the normalized sea-level pressure difference between Tahiti and Darwin (in units of standard deviation). Negative values of the SOI indicate an El Niño event. The relationship between ENSO and hurricane activity is strongest during the hurricane season so we use a August-October average of the SOI as a climate factor. The monthly SOI values (Ropelewski and Jones, 1997) are obtained from the *Climatic Research Unit* (CRU).

The NAO is characterized by fluctuations in sea level pressure (SLP) differences. Index values for the NAO (NAOI) are calculated as the difference in SLP between Gibraltar and a station over southwest Iceland (in units of standard deviation), and are obtained from the CRU (Jones et al., 1997). The values are averaged over the pre- and early-hurricane season months of May and June (Elsner et al., 2001) as this is when the relationship with hurricane activity is strongest. (Elsner and Jagger, 2006).

The SST values are a blend of modeled and observed data and, to a first order, correlate with the amount of fuel for hurricane development. Unsmoothed and undetrended Monthly North Atlantic arealy averaged SST anomalies from 0 to 70∘N (in units of °C) were computed using the base period 1951–2000. Data are obtained from the NOAA-CIRES *Climate Diagnostics Center* back to 1871. For this study we average the North Atlantic SST anomalies over the peak hurricane season months of August through October.

For SSN we use the monthly total sunspot number for September (the peak month of the hurricane season). Sunspots are magnetic disturbances of the sun surface having both dark and brighter regions. The brighter regions (plages and faculae) increase the intensity of the UV emissions. Increased sunspot numbers correspond to more magnetic disturbances. Sunspot numbers produced by the *Solar Influences Data Analysis Center* (SIDC), *World Data Center for the Sunspot Index*, at the *Royal Observatory* of Belgium are obtained from the U.S. *National Oceanic and Atmospheric Administration*.

In summary, normalized historical economic damage losses from hurricane events from the period 1900–2007 will be modeled using climate and solar data that represent optimal relationships found in previous studies on U.S. hurricane activity. By "optimal" we mean relative to what is currently understood about how environmental variables influence hurricanes. It does not mean relative to an exhaustive search for correlations across many different variables.

Figure 3 shows the time series of the climate factors that are used in the model. Upper and lower quartile values of the SOI are 0.40 and −0.90 s.d., respectively with a median (mean) value of −0.18 (−0.16) s.d. Years of above (below) normal SOI correspond to La Niña (El Niño) events and thus a higher probability of at least one U.S. hurricane. The upper and lower quartile values of the NAO are 0.40 and −1.09 s.d., respectively with a median (mean) value of −0.39 (−0.33) s.d. Years of below (above) normal values of the NAO correspond to a weak (strong) NAO phase and thus to higher (lower) probability of U.S. hurricanes. The upper and lower quartile values of the Atlantic SST anomalies are 0.22 and −0.16°C, respectively. Years of above (below) normal values of SST correspond to higher (lower) probability of hurricane activity. The upper and lower quartile values of the September SSN are 91.7 and 17.1, respectively with a median (mean) value of 50.2 (62.0). Years of below (above) normal SSN correspond to a lower (higher) probability of U.S. hurricanes. The largest correlation among the covariates occurs between the SSN and SST at a marginally statistically significant value of 0.18 (p-value = 0.064).

As an initial analysis of the damage data relative to the climate signals, here we compare locations on the distribution of per storm damage losses conditional on the various climate factors. Table 3 lists the damage amounts at the median and upper 99th percentile for both data sets and the damage ratio as the amount of damage during above normal years to the amount during below normal years. During seasons characterized by La Niña conditions (above normal values of SOI) the median losses are greater by a factor of more than two compared with years with El Niño

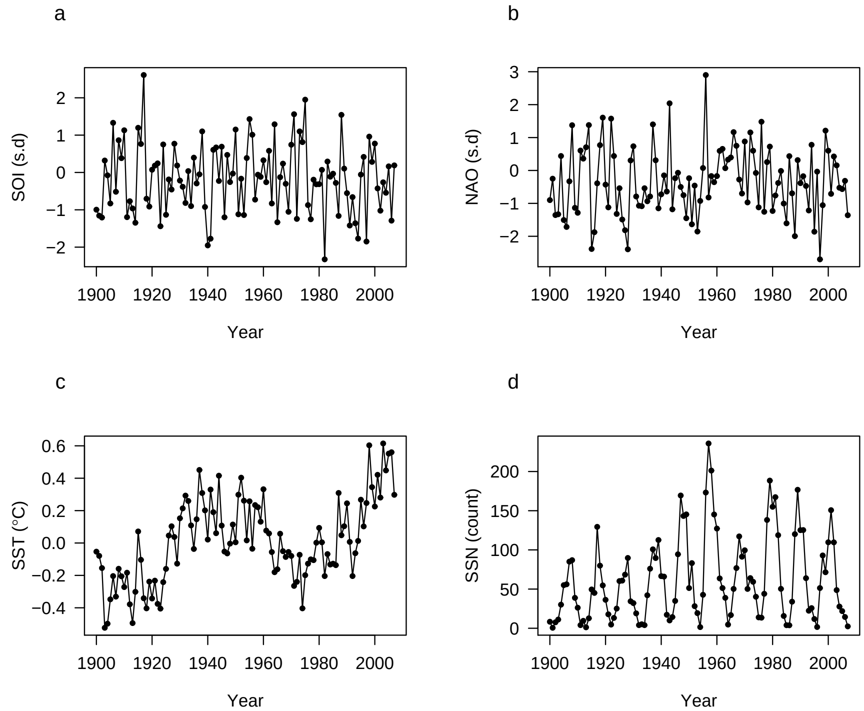


Fig. 3 Time series of the four covariates used to estimate wind damage losses from hurricanes.

Table 3 Damage amounts for Hurricanes in billions $US adjusted to 2005 along with conditional damage ratios. The damage ratio is the respective quantile amount of damage per storm during above normal years to the amount during below normal years.

	Collins/Lowe		Pielke/Landsea	
	50%	99%	50%	99%
Damage	1.326	86.308	1.216	90.181
SOI	2.460	0.578	2.860	0.560
NAO	0.829	0.354	0.711	0.319
SST	1.027	1.342	1.233	1.289
SSN	0.671	0.493	0.557	0.480

conditions. However, the extreme losses are greater during El Niño conditions. During seasons with below normal springtime NAO conditions, the damages tend to be greater at the median level and even more so at the extremes.

Interestingly, seasons characterized by higher than average SSTs show lower amounts of total damage at the median levels compared with seasons characterized by lower SSTs. There is, however, a modest increase in damage loss amounts during warm years over loss amounts during the cold years at the upper tails of the distribution. During seasons with below normal sunspots, damage losses tend to be greater at the median level and similarly so at the extremes. These results are expected from what we know about how these climate factors influence U.S. hurricane activity (Elsner and Jagger, 2006; Jagger and Elsner, 2006). Again, note that the CL and PL damage loss data sets give practically the same results.

4 Large and small losses

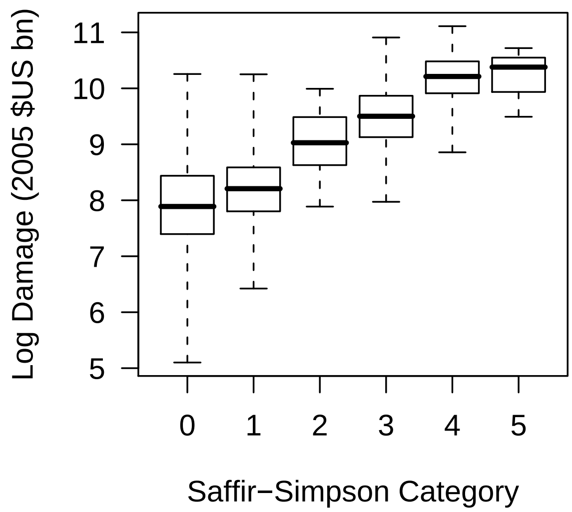


Fig. 4 Common logarithm ($\log_{10}$) of damage losses by Saffir-Simpson hurricane category.

Figure 4 shows the amount of losses by the Saffir-Simpson category of storm intensity. It is clear that both the category 0 and 1 tropical cyclones are different from the stronger tropical cyclones, in that the range of damage is larger. For instance, at the 80% interval of losses the range is from 6.8 to 9.0 for category 0 storms, and from 7.2 to 9.1 for category 1 storms, this is about 2.2 and 1.9, respectively, or approximately a factor of 100. In comparison, the category 2, 3, 4, and 5 ranges are 1.3, 1.5, 1.4, and 1.1, respectively. Thus it makes sense to model tropical storms (category 0) and category 1 hurricanes separately from category 2 and higher storms. However, there is a practical limitation in that we lose 109 of the 212 storms. Thus for this paper we restrict our analysis to category 1 and higher tropical cyclones (hurricanes only) as a compromise between removing too much data and keeping too many weaker events.

Moreover, the total damage from the 221 events (1900–2007) calibrated to 2005 is estimated at US $1.1 trillion. The large skewness in damage losses per event and per annum suggests that it might be a good strategy to separate large losses from small losses for the purpose of modeling. It is often noted that 80% of the total damage from tropical cyclones is caused by 20% of the biggest loss events. Figure 5 shows that the distribution of damage data is more skewed than that. In fact, the top 35 loss events (less than 16% of the total number of loss events) account for more than 81% of the total loss amount. The relative infrequency of the largest loss events

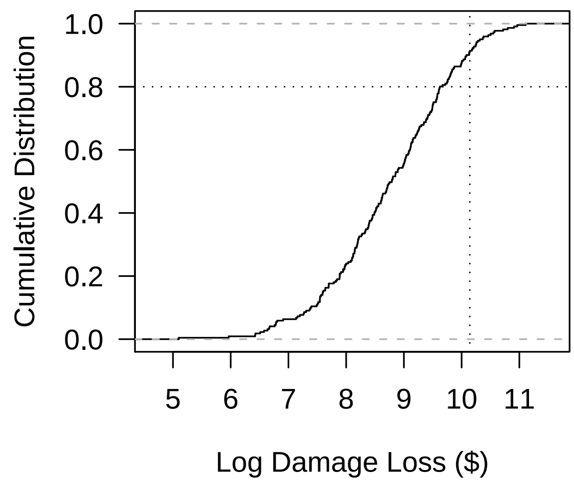


Fig. 5 Empirical cumulative distribution function of per storm damage losses expressed on the common logarithm scale. The values on the ordinate are the fraction of damage events less than or equal to the damage values on the abscissa. The horizontal dotted line is the 80th percentile and the vertical dotted line is the damage amount of the 20th worst event.

argues for a split that favors more data for modeling the largest losses. Here we use a cutoff of one billion \$US and find that 90 of the 160 hurricane events (56.3%) exceed this threshold. The remaining 70 events (43.7%) account for only 15.4% of the total damages. Thus it might be reasonable to assume that the small loss events are at the "noise" level. In summary, our focus here is on the set of large losses from the stronger tropical cyclones.

5 A model for annual expected loss

Given a loss event, the logarithm of the amount of loss on an annual basis is modeled using a linear regression with the logarithm of the loss amount modeled as a truncated Normal distribution. The only statistically significant climate signal in the loss amount is the SST. Thus given a loss event, the magnitude of the loss increases with increasing ocean warmth. This is consistent with SST acting as a proxy for upper-ocean heat; a source of energy for hurricanes (Emanuel, 1991).

To arrive at an estimate of the annual loss we need to combine this loss amount estimate given an event with the frequency of a loss event. Since we divide loss events into large and small events, we use separate models. Thus, given a mean annual rate of large (small) loss events, the annual number of large (small) loss events follows a Poisson distribution with the natural logarithm of the loss event rate given as a linear function of the climate variables. We find that SST, NAO, SOI, and SSN are all statistically significant indicators of the frequency of large losses, but none of the climate variables are important for the frequency of small losses.

Mathematically we write the model for large losses as:

$$\begin{aligned} \log L &\sim \text{Normal}(\mu, \sigma^2) \\ \mu &= \alpha_0 + \alpha_1 \text{SST} \\ N &\sim \text{Poisson}(\lambda) \\ \lambda &= \exp(\beta_0 + \beta_1 \text{SST} + \beta_2 \text{NAO} + \beta_3 \text{SOI} + \beta_4 \text{SSN}) \end{aligned} \tag{1}$$

where L is the amount of total loss for an event and λ is the yearly hurricane frequency. The symbol $\sim$ refers to a stochastic relationship and indicates that the variable on the left hand side is a random draw (sample) from a distribution specified on the right hand side. The equal sign indicates a logical relationship with the variable on the left hand side algebraically related to variable(s) on the right hand side. As mentioned, the size of the loss is modeled as a truncated Normal distribution with parameters μ and σ^2 indicating the location and scale for the distribution. Unlike the normal distribution the location and scale parameters of the truncated normal distribution are are not the same as its mean and variance. In short, the model describes a compound Poisson process with rate λ and logarithm of the jump size distributed as a truncated normal distribution with parameters μ and σ.

Chi-square goodness-of-fit statistics do not show any evidence against adequacy of the rate model. Furthermore, there is no trend in the deviance residuals implying the rate model for large losses conditioned on the climate variables chosen is statistically stationary and the addition of a trend term does not improve the model. This suggests to us that there is no significant historical under reporting of the number of loss events from hurricanes in the United States over the period considered here.

The final model that combines the frequency of loss events with the magnitude of the loss given an event uses a hierarchical Bayesian specification. Bayesian models provide posterior distributions of model parameters, as opposed to a frequentist model using maximum likelihood estimatation(MLE) which only provides the parameter estimate and prediction error. For non normal distributions these MLE estimates are biased leading to biased predictions. We chose flat (uniform) model priors for the location and model precision $(1/\sigma^2)$ parameters to minimize the influence of prior. The final model is selected from a set of possible models by comparing the Deviance Information Criterion (DIC) for each model and then choosing the model with the smallest DIC. The DIC is the formulated explicitly for model selection in Bayesian models, in the same fashion that the AIC which is used to compare models using maximum likelihood estimation (Spiegelhalter et al., 2002). It is useful in Bayesian model selection where the posterior distributions of the models are obtained by Markov Chain Monte Carlo (MCMC) simulation. The model with the smallest DIC is estimated to be the model that would best predict a replicate data set that has the same structure as the observed one.

Given the hierarchical form of the model, samples of the annual losses are generated using WinBUGS (Windows version of Bayesian inference Using Gibbs Sampling) developed at the *Medical Research Council* in the UK (Gilks et al., 1994; Spiegelhalter et al., 1996). WinBUGS chooses an appropriate MCMC sampling algorithm based on the model structure. In this way annual losses are sampled conditional on the model coefficients and the observed values of the covariates. The cost associated with a Bayesian approach is the requirement to formally specify prior beliefs. Here we take the standard route and assume noninformative priors that provide little information about the parameters of interest.

MCMC, in particular Gibbs sampling, is used to sample the parameters given the data since no closed form solution exists for the posterior distribution of the model parameters in the truncated Normal (or for the generalized Pareto distribution (GPD) used in the next section). Indeed, WinBUGS is useful in that it allows us to sample the parameters from the posteriors created from arbitrary likelihood functions. As far as we are aware, there is no software for finding the maximum likelihood estimates of the regression parameters for a truncated normal distribution.

We check for mixing and convergence by examining successive samples of the parameters. Samples from the posterior distributions of the parameters indicate relatively good mixing and quick settling as two different sets of initial conditions produce sample values that fluctuate around a fixed mean. Based on these diagnostics, we discard the first 10000 samples and analyze the output from the next 10000 samples. The utility of the Bayesian approach for modeling the mean number of coastal hurricanes is described in Elsner and Jagger (2004) and for predicting damage losses is described in Jagger et al. (2008).

Figure 6 shows the predictive posterior distributions of annually aggregated losses for 6 different climate scenarios. The set of scenarios is ordered by generally increasing favorable climate conditions for hurricanes to affect the United States. Each panel shows the probability of no losses and the probability of losses on a logarithm (base 10) scale given at least one loss event. The upper left panel shows the posterior probabilities for a year during which the SST is much below normal, the NAO is much above normal, there is a strong El Niño, and the sun is very active (many sunspots). The specific covariate values are listed in the figure caption. The posterier samples indicate a relatively large probabiliy of no damage events (37%) under this scenario. The estimated annual loss taking into account the non-zero probability of no loss events is centered in the range between \$0.1 and \$1 bn.

As the climate factors change to indicate more favorable conditions for hurricanes, the posterior samples provide a distribution of annual losses that are more ominous with the probability of no losses decreasing to less than 1% and the expected annual total loss amount exceeding \$100 billion. All estimated loss amounts are converted to 2005 dollars. The results of the model integration are rather remarkable in showing a distinct climate signal in aggregate property losses in the United States from hurricanes. The annual expected loss increases with warmer Atlantic SSTs, cooler equatorial eastern Pacific SSTs, a negative phase of the NAO, and fewer sunspots.

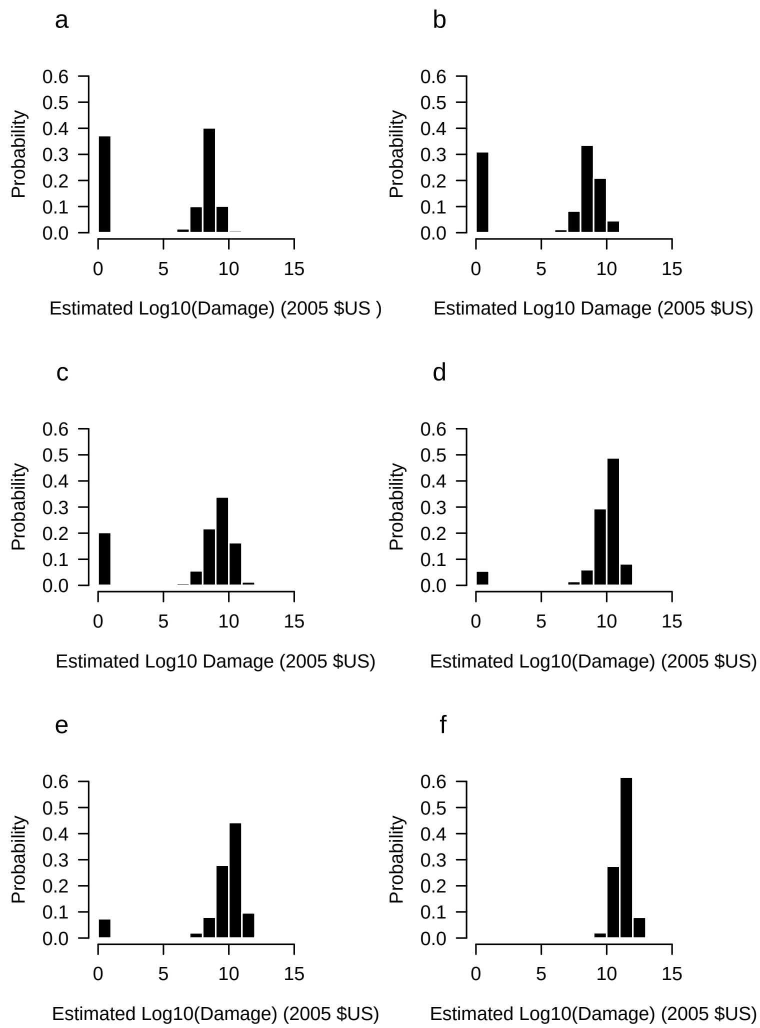


Fig. 6 Simulated annual losses for six different climate scenarios. The histograms show the probability of no losses on the left and the distribution of annual amount of loss on a logarithmic scale on the right. The panels are ordered toward conditions increasingly favorable for hurricane activity along the U.S. coast. The histograms are created from posterior samples generated from the hierarchical Bayesian model using (a) SST = -0.52°C, NAO = $+2.9$ s.d., SOI = -2.3 s.d., and SSN = 236, (b) SST = -0.24°C, NAO = $+0.7$ s.d., SOI = -1.1 s.d., and SSN = 115, (c) SST = $+0.01^\circ$C, NAO = -0.3 s.d., SOI = -0.2 s.d., and SSN = 62, (d) SST = $+0.27^\circ$C, NAO = -1.4 s.d., SOI = $+0.8$ s.d., and SSN = 9, (e) SST = $+0.43^\circ$C, NAO = -1.3 s.d., SOI = -0.1 s.d., and SSN = 5, and (f) SST = $+0.61^\circ$C, NAO = -2.7 s.d., SOI = $+2.6$ s.d., and SSN = 1.

6 A model for the probable maximum loss

While the above modeling strategy makes sense for estimating the distribution of likely annual damage associated with variations in environmental conditions from one year to the next, for financial planning it might be of greater importance to know the maximum possible storm loss. In this case, the normal distribution is replaced by an extreme value distribution for the logarithm of losses. For example, the family of Generalized Pareto Distributions (GPD) describes the behavior of individual extreme events.

Consider observations from a collection of random variables in which only those observations that exceed a fixed value are kept. As the magnitude of this value increases, the GPD family represents the limiting behavior of each new collection of random variables. This property makes the family of GPD a good choice for modeling extreme events including large losses from wind storms. The choice of threshold, above which we treat the values as extreme, is a compromise between retaining enough observations to properly estimate the distributional parameters (scale and shape), but few enough that the observations follow a GPD family.

The GPD describes the distribution of losses that exceed a threshold l but not the frequency of losses at that threshold. As we did with the annual loss model, we specify that, given a rate of loss events above the threshold, the number of loss events follows a Poisson distribution. Here there is no need to consider small loss events as we are only interested in the large ones. Combining the GPD for the distribution of large loss amounts with the Poisson distribution for the frequency of loss events above the threshold allows us to obtain return periods for given levels of losses.

Mathematically we are modeling the exceedances, $L-l$, as samples from a family of GPD distributions so that for any threshold l and any event with losses L, the probability that L exceeds some arbitrary level x above l is

$$\Pr(L > x + l | L > l) = \begin{cases} \exp(-\frac{x}{\sigma}) & \xi = 0 \\ \left(1 + \frac{\xi}{\sigma_l} x\right)^{-\frac{1}{\xi}} & \xi \neq 0 \end{cases} \tag{2}$$

$$= \mathrm{GPD}(\mathrm{x}|\sigma_{\mathrm{l}}, \xi) \tag{3}$$

where $\sigma_l > 0$, $x \geq 0$, and $\sigma_l + \xi x \geq 0$. If the exceedances above l_0 follow a GPD then the exceedances above $l > l_0$ follow a GPD with the same shape, ξ and scale that shifts linearly with the threshold:

$$\sigma_l = \sigma_0 + \xi l$$

The parameters σ_l and ξ are the scale and shape parameters respectively. For negative shape parameters the GPD family of distributions has an upper limit of $L_{\max} = l + \sigma_l / |\xi|$. The equation for σ_l specifies that if the values follow a GPD, then for any threshold the distribution of exceedances is GPD with the same value of the shape parameter (ξ) from the original distribution and a scale parameter that changes linearly with the threshold at a rate equal to the shape parameter.

We determine the threshold value at \$1 bn U.S. for the set of losses by examining the mean residual life plot. This is a plot of the mean value of the exceedances as a function of the threshold. If the data follow a GPD distribution, this plot is linear. The threshold is chosen as the smallest value where the function is linear for all larger thresholds Coles (2001).

The GPD describes the loss distribution for each wind event whose losses exceed l but not the frequency of events at that magnitude. We assume that the number of loss events in year y that exceed l has a Poisson distribution with mean (or exceedance) rate is λ_l. Thus by combining the exceedance probability and the exceedance rate with our assumption that they are independent we get a Poisson distribution for the number of loss events per year with losses exceeding m (N_m) with a rate given by

$$\lambda_m = \lambda_l \Pr(L > m | L > l). \tag{4}$$

This specification is physically realistic since it allows us to model loss occurrence separately from loss amount. Moreover from a practical perspective, rather than a return rate per loss occurrence, the above specification allows us to obtain an annual return rate on the extreme losses, which is more meaningful for the business of risk and insurance.

Now, the probability that the yearly maximum will be less than m is the probability that $N_m = 0$. Since N_m has a Poisson distribution

$$\begin{aligned}\Pr(L_{\max} \le m) &= \Pr(N_m = 0) & (5)\\ &= \exp(-\lambda_m) & (6)\\ &= \exp\{-\lambda_l \mathrm{GPD}(\mathrm{m}-\mathrm{l}|\sigma_\mathrm{l}, \xi)\} & (7)\end{aligned}$$

If we make the substitution for $\xi \neq 0$:

$$\begin{aligned}\sigma_\mu &= \lambda_l^{\xi} \sigma_l & (8)\\ \mu &= l + \frac{\sigma_\mu - \sigma_l}{\xi} & (9)\end{aligned}$$

then

$$\Pr(L_{\max} \le m) = \exp\left\{-\left[1 + \xi\left(\frac{m-\mu}{\sigma_\mu}\right)\right]^{-\frac{1}{\xi}}\right\} \tag{10}$$

has a Generalized Extreme Value (GEV) distribution, which is in canonical form. If $\xi = 0$ then we make the substitutions

$$\begin{aligned}\sigma_\mu &= \sigma_l\\ \mu &= l + \sigma_l \log(\lambda_l)\end{aligned}$$

then

$$\Pr(L_{\max} \le m) = \exp\left\{-\exp\left[-\left(\frac{m-\mu}{\sigma_\mu}\right)\right]\right\}. \tag{11}$$

We convert the peaks-over-threshold parameters λ_l, σ_l, ξ to the GEV canonical parameters μ, σ_μ, ξ, and so compare results obtained with different thresholds. Using the canonical parameters, for example we calculate the yearly (seasonal) return level, $\mathrm{rl}(r)$, corresponding to a given return period, r and GEV parameters μ, σ, ξ by solving for m in $\Pr(L_{\max} \geq m) = \frac{1}{r}$ giving

$$\mathrm{rl}(r) = \begin{array}{ll} \mu + \frac{\sigma}{\xi}\left\{\left[\log\left(\frac{r}{r-1}\right)^{-\xi} - 1\right]\right\} & \xi \neq 0 \\ \mu - \sigma \cdot \log\left\{\log\left(\frac{r}{r-1}\right)\right\} & \xi = 0. \end{array} \quad (12)$$

Additional details are given in (Coles 2001).

As with the annual loss model we use a Bayesian hierarchical specification for the model of extreme losses. MCMC samples are used to generate posterior predictive distributions. Here we are interested in the return level as a function of return period. The annual return level is determined by the return level of individual extreme events and the annual frequency of such events above a threshold rate. The annual number of extreme events follows a Poisson distribution with the natural logarithm of the rate specified as a linear function of the four covariates.

Given values for the scale (σ) and shape (ξ) parameters, the return level of individual extreme events follows a GPD. The logarithm of the scale parameter and the shape parameter are both linear functions of the the four covariates.

As before, samples of the return levels are generated using WinBUGS and we use noninformative prior distributions. Samples from the posterior distribution of the model parameters indicate good mixing and good convergence properties. We discard the first 10000 samples and analyze the output from the next 10000 samples. Applications of Bayesian extremal analysis are found in Coles and Tawn (1996), Walshaw (2000), Katz et al. (2002), Coles et al. (2003), Hsieh (2004), and Jagger and Elsner (2006). Figure 7 shows the predictive posterior distributions of extreme losses for four different climate scenarios using quantile values. For each return period the the 0.025, 0.05, 0.25, 0.5, 0.75, 0.95, and 0.975 quantile values of the maximum storm damage losses are plotted. The first scenario is characterized by covariates in favor of fewer hurricanes, the second scenario represents long-term climatological conditions, the third scenario is characterized by covariates favoring more hurricanes, and the fourth scenario is characterized by covariates favoring stronger hurricanes. The loss distribution changes substantially between the different climate and solar scenarios and in a direction that is consistent with our understanding about the relationship between climate and hurricane activity.

Under the first scenario we find the median return level of a 50-year extreme event at approximately \$18 bn, this compares with a median return level of the same 50-year extreme event loss of approximately \$793 bn under the fourth scenario. Thus the model can be useful for projecting extreme losses over time horizons longer than a year given values of the covariates. Note that the results are interpreted as the posterior distributions of the return level for a return period of 50 years of the covariate values as extreme or more extreme than 1 standard deviation. With 4 independent covariates and an annual probability of about 16% that a particular covariate

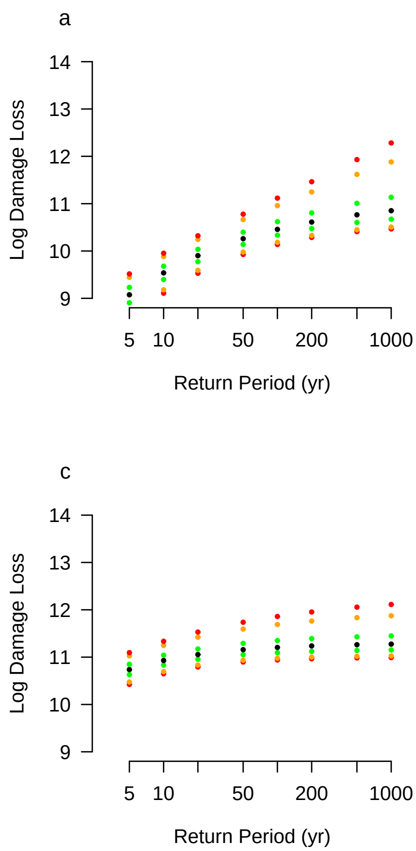


Fig. 7 Simulated extreme losses for four different climate scenarios. The points are the 0.025, 0.05, 0.25, 0.5, 0.75, 0.95, and 0.975 quantiles from the posterior distribution of the loss model. The panels are ordered toward conditions increasingly favorable for large losses. (a) SST = -0.243°C, NAO = $+0.698$ s.d., SOI = -1.087 s.d., and SSN = 115, (b) SST = $+0.012$°C, NAO = -0.331 s.d., SOI = -0.160 s.d., and SSN = 62, (c) SST = $+0.268$°C, NAO = -1.359 s.d., SOI = -0.766 s.d., and SSN = 9, (d) SST = $+0.268$°C, NAO = -1.359 s.d., SOI = -1.087 s.d., and SSN = 9. The upper quantile values in panel (d) are outside the range of the plot.

is more than 1 s.d. from its mean, the chance that all covariates will be this extreme or more in a given year is less than 0.1%.

7 Summary

Coastal hurricanes are capable of generating large financial losses to the insurance industry. Annual loss totals are directly related to the intensity and frequency of hurricanes affecting the coast. Since a measureable amount of skill exists in forecasts of coastal hurricane activity, it makes sense to investigate the potential of modeling losses directly. This paper demonstrates clear climate and solar signals in the historical estimates of property damage losses.

Two separate statistical models are specified using hierarchical Bayesian technology and predictive posterior distributions are generated using MCMC sampling. The first model can be used to estimate the expected annual loss under various environmental scenarios. The annual expected loss increases with warmer Atlantic SSTs, cooler equatorial eastern Pacific SSTs, a negative phase of the NAO, and fewer sunspots. The second model can be used to estimate the distribution of losses over a longer time horizon conditional on the values of the four covariates.

Results are consistent with current understanding of hurricane climate variability. While the models here are developed from aggregate loss data for the entire United States susceptible to Atlantic hurricanes, it would be possible to apply the techniques to model data representing a subset of losses capturing, for example, a particular reinsurance portfolio. Moreover, since the models make use of MCMC sampling they can be easily extended to include measurement error and missing data.

Hazard risk affects the profit and loss of the insurance industry. Some of this risk is transferred to the performance of securities traded in financial markets. This implies that early and reliable information concerning potential hazards will be useful to investors. This paper advances those goals.

Traditional hurricane risk models used by the insurance industry rely on a catalog of storms that represent the historical data in some way or another. While useful for estimating AAL and loss exceedance curves for aggregate and occurrence portfolio losses, these catalogs are not easily suited for anticipating losses based on an ever-changing climate. Specifically, at the core of the catalog is a set of synthetic storms and a way to assign a probability to each. However, it is not obvious how to condition the set of storm characteristics on climate. The approach demonstrated here provides an alternative way to anticipate losses on the seasonal to multi-year time scale.

Concerning the future, increases in ocean temperature will raise a hurricane's potential intensity, all else being equal. However, corresponding increases in atmospheric wind shear—in which winds at different altitudes blow in different directions—could tear apart developing hurricanes and could counter this tendency by dispersing the hurricane's heat. However, a recent study based on a set of homogenized satellite-derived wind speeds indicates the strongest hurricanes are getting stronger worldwide (Elsner, Kossin, and Jagger, 2008). This new information can be incorporated in models of the type demonstrated here by placing a discount factor on the older information relative to the more recent events.

Acknowledgements This research is supported by Florida State University's *Catastrophic Storm Risk Management Center*, the *Risk Prediction Initiative* of the Bermuda Institute for Ocean Studies (RPI-08-02-002), and by the U.S. *National Science Foundation* (ATM-0738172). The views expressed within are those of the authors and do not reflect those of the funding agency.

References

1. Coles S (2001) An Introduction to Statistical Modeling of Extreme Values. Springer, New York
2. Coles S, Tawn JA (1996) A Bayesian analysis of extreme rainfall data. Applied Statistics 45, 463–478
3. Coles S, Pericchi LR, Sisson S (2003) A fully probabilistic approach to extreme rainfall modeling. Journal of Hydrology 273, 35–50
4. Elsner JB, Jagger TH (2004) A hierarchical Bayesian approach to seasonal hurricane modeling, Journal of Climate 17, 2813–2827
5. Elsner JB, Jagger TH (2006) Prediction models for annual U.S. hurricane counts. Journal of Climate 19, 2935–2952
6. 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
7. Elsner JB, Bossak BH, Niu, X-F (2001) Secular changes to the ENSO-U.S. hurricane relationship. Geophysical Research Letters, 28, 4123–4126
8. Elsner JB, Niu, X-F, Jagger TH (2004) Detecting shifts in hurricane rates using a Markov Chain Monte Carlo approach. Journal of Climate 17, 2652–2666
9. Elsner JB, Kossin JP, Jagger TH (2008) The increasing intensity of the strongest tropical cyclones. Nature 455, 92–95
10. Emanuel KA (1991) The theory of hurricanes. Annual Review of Fluid Mechanics 23, 179–196
11. Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov Chain Monte Carlo in Practice, London: Chapman & Hall/CRC
12. Hsieh PH (2004) A data-analytic method for forecasting next record catastrophe loss. The Journal of Risk and Insurance 71, 309–322
13. Jagger TH, Elsner JB (2006) Climatology models for extreme hurricane winds near the United States. Journal of Climate 19, 3220–3236
14. Jagger TH, Elsner JB, Niu, X-F (2001) A dynamic probability model of hurricane winds in coastal counties of the United States. Journal of Applied Meteorology 40, 853–863
15. Jagger TH, Elsner JB, Saunders MA (2008) Forecasting U.S. insured hurricane losses. In: Murnane R, Diaz H (Eds.). Climate Extremes and Society, Cambridge: Cambridge University Press
16. 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
17. Katz RW (2002) Stochastic modeling of hurricane damage. Journal of Applied Meteorology 41, 754–762
18. Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. Advances in Water Resources 25, 1287–1304
19. Lane MN, (2008a) Beyond convergence, toward integration. Trade Notes, Lane Financial LLC, August 29, 2008
20. Lane MN, (2008b) The measurement of ILS returns. Trade Notes, Lane Financial LLC, August 29, 2008

21. Leckebusch GC, Ulbrich U, Fröhlich L, Pinto JG (2007) Property loss potentials for European midlatitude storms in a changing climate. Geophysical Research Letters 34, L05703, doi:10.1029/2006GL027663
22. Murnane RJ, Barton C, Collins E, Donnelly J, Elsner J, Emanuel K, Ginis I, Howard S, Landsea C, Liu, K-b, Malmquist D, McKay M, Michaels A, Nelson N, O'Brien J, Scott D, Webb T (2000) Model estimates hurricane wind speed probabilities. Eos, Transactions, American Geophysical Union 81, 433, 438
23. Pielke RA, Gratz J, Landsea CW, Collins D, Saunders MA, Musulin R (2008) Normalized hurricane damage in the United States: 1900–2005. Natural Hazards Review, 9(1), 29–42
24. Ropelewski CF, Jones PD (1997) An extension of the Tahiti-Darwin southern oscillation index. Monthly Weather Review 115, 2161–2165
25. Saunders MA, Lea AS (2005) Seasonal prediction of hurricane activity reaching the coast of the United States. Nature 434, 1005–1008
26. Spiegelhalter DJ, Best NG, Gilks WR, Inskip H (1996) Hepatitis B: A case study in MCMC methods. In: Gilks WR, Richardson S, Spiegelhalter DJ (Eds.). Markov Chain Monte Carlo in Practice, London: Chapman & Hall/CRC, 21–43
27. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society, Series B 64, Part q4, 583–639
28. Walshaw D (2000) Modeling extreme wind speeds in regions prone to hurricanes. Applied Statistics 49, 51–62