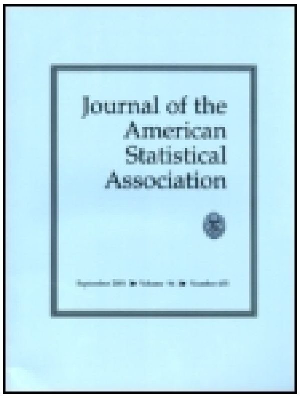


Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/uasa20

Changepoints in the North Atlantic Tropical Cyclone Record

Michael W. Robbins[a], Robert B. Lund[a], Colin M. Gallagher[a] & QiQi Lu[a]

[a] Michael W. Robbins is Postdoctoral Fellow, National Institute of Statistical Sciences, Research Triangle Park, NC 27709-4006 . Robert B. Lund is Professor and Colin M. Gallagher is Associate Professor, Department of Mathematical Sciences, Clemson University, Clemson, SC 29634-0975. QiQi Lu is Assistant Professor, Department of Mathematics and Statistics, Mississippi State University, MS 39762. This work was done while the first author was a Ph.D. Student at Clemson University. Robert Lund's research was supported by National Science Foundation grant DMS 0905770. Comments made by three referees significantly strengthed this manuscript.
Published online: 01 Jan 2012.

Changepoints in the North Atlantic Tropical Cyclone Record

Michael W. Robbins, Robert B. Lund, Colin M. Gallagher, and QiQi Lu

This article examines the North Atlantic tropical cyclone record for statistical discontinuities (changepoints). This is a controversial area and indeed, our end conclusions are opposite of those made in Dr. Kelvin Droegemeier's July 28, 2009 Senate testimonial. The methods developed here should help rigorize the debate. Elaborating, we develop a level-α test for a changepoint in a categorical data sequence sampled from a multinomial distribution. The proposed test statistic is the maximum of correlated Pearson chi-square statistics. This test statistic is linked to cumulative sum statistics and its null hypothesis asymptotic distribution is derived in terms of the supremum of squared Brownian bridges. The methods are used to identify changes in the tropical cyclone record in the North Atlantic Basin over the period 1851–2008. We find changepoints in both the storm frequencies and their strengths (wind speeds). The changepoint in wind speed is not found with standard cumulative sum mean shift changepoint methods, hence providing a dataset where categorical probabilities shift but means do not. While some of the identified shifts can be attributed to changes in data collection techniques, the hotly debated changepoint in cyclone frequency circa 1995 also appears to be significant.

KEY WORDS: Atlantic hurricanes; Brownian bridge; Chi-square statistics; Climate change; CUSUM.

1. INTRODUCTION

Climate change is a contentious and active area of research. Anthropogenic increases in global air temperature are now widely recognized (Houghton et al. 2001; Karl and Trenberth 2003). Higher sea surface temperatures (SSTs) have also been reported (Cane et al. 1997). Although tropical cyclones are powered by warm waters, climatologists do not uniformly agree that rising SSTs are increasing tropical cyclone counts and strengths, with some arguing that there have been recent increases (Anthes et al. 2006; Emanuell, Sundararajan, and Williams 2008; Saunders and Lea 2008) and others arguing that no firm conclusions can yet be made (Pielke et al. 2005; Landsea 2007). Those acknowledging recent changes have differing opinions about the nature of the change. Saunders and Lea (2008) purport an increase in cyclone frequency, while Emanuell (1987, 2005) claims that any change would manifest itself as an increase in the strength of storms. Vecchi and Knutson (2008) and Landsea et al. (2010) believe that recent increases in storm counts are attributable to an increased number of weak cyclones included in the record (these storms are typically of short duration and are difficult to detect without modern sensing techniques). Regarding the effects of a warming climate on hurricanes, Dr. Kelvin Droegemeier stated in a July 28, 2009 testimony to the United States Senate, "…we're seeing a shift not in the total number of storms, but a larger number of more intense hurricanes and a smaller number of less intense hurricanes." The full video can be found from the link *http://commerce.senate.gov/public/index.cfm?p=Hearings*.

Rigorous statistical studies on cyclone changes are relatively sparse. Some notable articles include the Markov Chain Monte Carlo methods of Elsner, Niu, and Jaeger (2004) and the Bayesian-based changepoint approach of Jewson and Penzer (2008). While the methods employed by these authors differ from our asymptotic results, these authors agree with our end conclusions of storm frequency changepoints circa 1900, 1935, and 1995. The circa 1995 changepoint is controversial. Landsea et al. (1999), Jarrell, Hebert, and Mayfield (1992), and Neumann et al. (1999) provide convincing explanations for the circa 1900 and 1935 changepoints in terms of data collection technique changes.

The literature on changepoint problems is vast. Page (1954, 1955) is widely credited with introducing undocumented changepoint problems. MacNeill (1974) studied a cumulative sum (CUSUM)-type changepoint statistic and established convergence of this statistic to a Brownian bridge in the null hypothesis cases of independent and identically distributed (IID) model errors. The monograph by Csörgő and Horváth (1997) provides asymptotic results for likelihood ratio (LR) statistics under general IID settings.

After Hinkley and Hinkley (1970) examined changepoints in binomial data, changepoint detection in categorical data has been an active research area. The most popular techniques include Bayesian methods (Smith 1975; Chib 1988; Carlin, Gelfand, and Smith 1992; Qian, Pan, and King 2004; Girón, Ginebra, and Riba 2005), CUSUM-type methods (Pettitt 1980; Wolfe and Chen 1990) and maximum likelihood methods (Fu and Curnow 1990). Most of these authors analyze detection power; reliable discussions on null hypothesis distributions of the test statistics studied are comparatively sparse. Two authors (Horváth and Serbinowska 1995; Hirotsu 1997) also study maximums of chi-square statistics and discuss their approximate equivalence to LR tests. Horváth and Serbinowska (1995), the article most methodologically related to ours, provide asymptotic distributions of maximums of chi-square statistics that differ slightly from ours. Other authors introduce a nonparametric method for testing for changes in the marginal distribution via empirical distribution functions (Csörgő and Horváth 1987;

Michael W. Robbins is Postdoctoral Fellow, National Institute of Statistical Sciences, Research Triangle Park, NC 27709-4006 (E-mail: *robbins@niss.org*). Robert B. Lund is Professor and Colin M. Gallagher is Associate Professor, Department of Mathematical Sciences, Clemson University, Clemson, SC 29634-0975. QiQi Lu is Assistant Professor, Department of Mathematics and Statistics, Mississippi State University, MS 39762. This work was done while the first author was a Ph.D. Student at Clemson University. Robert Lund's research was supported by National Science Foundation grant DMS 0905770. Comments made by three referees significantly strengthed this manuscript.

Journal of the American Statistical Association
March 2011, Vol. 106, No. 493, Applications and Case Studies
DOI: 10.1198/jasa.2011.ap10023

Carlstein 1988). By partitioning such data into categories, our categorical changepoint test can also be viewed as a marginal distribution changepoint test.

The rest of this work is organized as follows. Section 2 overviews the data that we study. Section 3 then summarizes CUSUM changepoint methods. Section 4 introduces our test statistic and establishes its asymptotic null hypothesis properties. Here, the methods are linked to classical CUSUM tests. Section 5 applies the methods to the historical North Atlantic Basin cyclone record and justifies the technical assumptions made in Sections 3 and 4. Several changepoints are found in storm frequencies and wind speeds. In fact, our findings are essentially opposite of those purported by Dr. Droegemeier in his Senate testimony. Remarks in Section 6 and an Appendix conclude the article.

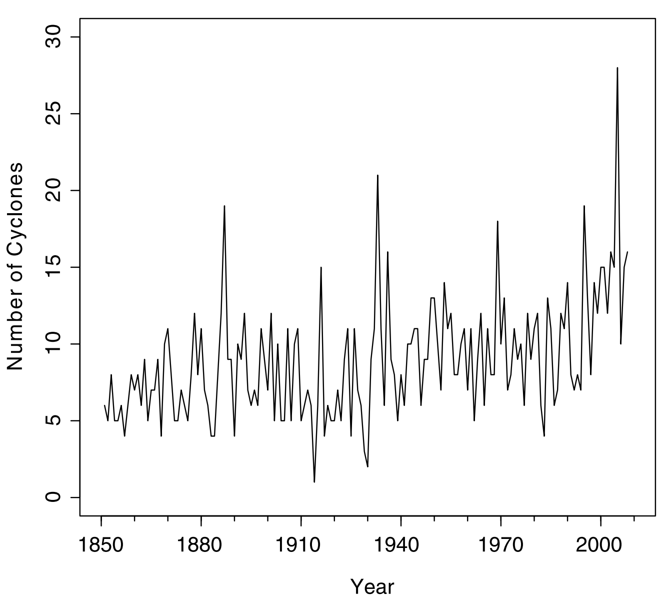
2. THE HURDAT DATA

The data that we use is the HURDAT dataset, which is available on the National Oceanic Atmospheric Administration's website. These data contain information on 1410 Atlantic basin cyclones that achieved tropical storm-level intensity or higher between 1851 and 2008. The data have been reanalyzed several times (Landsea et al. 2004, 2008) and are believed to contain inconsistencies due to advances in measurement techniques. For instance, counts of landfalling cyclones before 1900 are considered unreliable (Landsea et al. 1999, 2007; and Jarrell, Hebert, and Mayfield 1992) due to sparse coastline populations. Also, as Landsea et al. (1999) and Neumann et al. (1999) observe, aircraft reconnaissance improved detection of nonlandfalling storms towards the end of World War II. Satellites were fully implemented for cyclone surveying in the middle 1960s, and their introduction also likely improved the accuracy of measured wind speeds. Landsea et al. (2006) state that measurement techniques are continually improving and data on wind speeds as recently as the 1980s may be misleading. Our aim is to confirm or deny the existence of such inconsistencies using the developed changepoint methods. In this pursuit, we also hope to identify regime shifts caused by other causes such as climate change.

Many covariates are available in the HURDAT record. For instance, each cyclone has an estimate of the maximum wind speed achieved by that storm. Figure 1 provides time series plots of annual storm counts and maximum storm wind speeds. Both plots in Figure 1 show discontinuities—this aspect will be verified later.

The HURDAT record also provides information relating to a storm's geographic course, its duration, if (and where) it struck land and so forth. While we will examine some of these covariates later, we are primarily interested in testing for changes in the cyclone counts, storm strengths, or both. For example, it is feasible that cyclone counts and their wind speeds are increasing, that cyclone counts are increasing but their wind speeds are not, or some other combination. Hence, we would like to develop a "joint test" that can signal changes in either the Poisson rate or the distribution of any particular covariate.

Measuring individual cyclone strengths is also controversial and many of the documented wind speeds of strong storms may differ from their true strengths by as much as 20 mph (Neumann et al. 1999 and Landsea 2007 discuss data quality). Meteorologists frequently classify tropical cyclone severity via the Saffir–Simpson scale, which is a categorical (ordinal) scale ranging from 1 to 5, with 5 representing the strongest storm. Wind tunnel tests have established what type of damage each category storm typically does to buildings (see page 201 of Burt 2004 for this listing); hence, in a general sense, the categorical wind speed may be easier to correctly estimate than the true wind speed to say 10 mph. A categorical approach to handle the wind speed covariate seems reasonable and will be adopted here. While our categories are based on the Saffir–Simpson scale, one can always rerun the analysis with a finer categorical partition if desired. Categorization is also helpful in the development of the joint test. We will consider a multivariate series



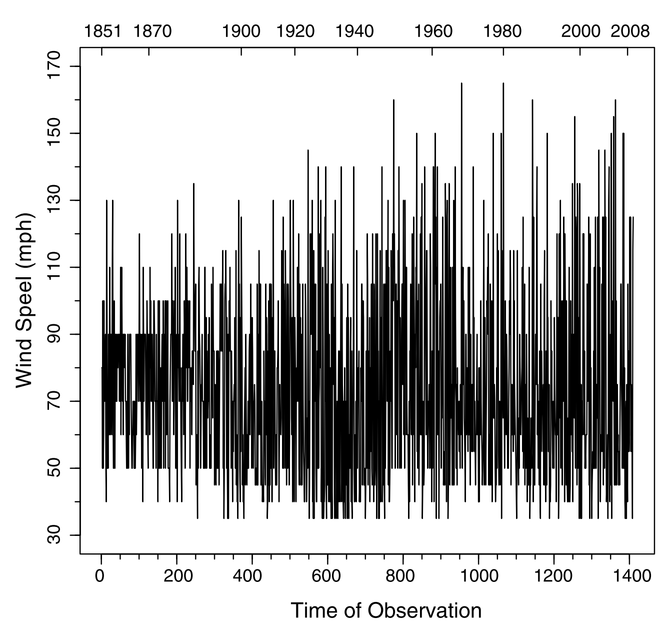


Figure 1. Time series plots of storm counts by year (left) and max windspeed by storm (right).

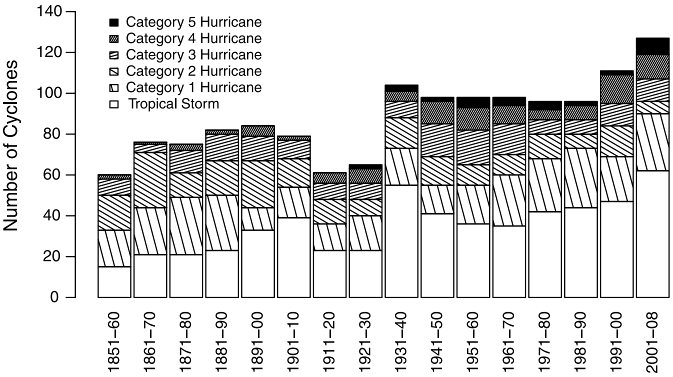


Figure 2. The number of tropical cyclones in each decade broken down by category for the years 1851–2008.

which gives the annual number of storms occurring in each category. To help visualize this series, Figure 2 shows the number storms of each category by decade.

We next introduce and develop level-α changepoint tests that are appropriate for the HURDAT data. Each series analyzed is assumed to be IID under the null hypothesis that no changepoints occur. Poissonian arrival dynamics are also assumed for the annual storm counts. Although we briefly mention tests (and corresponding results) which generalize these assumptions, we empirically justify the IID Poisson assumption for the storm counts series at the end of Section 5.

3. CUSUM REVIEW

Suppose we wish to determine whether or not there is a mean shift in the data $X_1, \ldots, X_n$ at some unknown time c. Our null hypothesis is that $\{X_t\}_{t=1}^n$ is IID and our alternative is that $E[X_t]$ shifts at time c. If c were known *a priori*, one would simply compare $c^{-1}\sum_{t=1}^c X_t$ and $(n-c)^{-1}\sum_{t=c+1}^n X_t$ with a standard t- or z-test. If T_c denotes the statistic from such a procedure, then the null is rejected when $|T_c|$ is large. When the time of the change is unknown, a natural test statistic is $T_{\max} = \max_{k\in K} T_k$, where K is the set of all (admissible) changepoint times that will be considered. The estimated changepoint time $\hat{c}$ is an argument of $k \in K$ that maximizes T_k. The null hypothesis is rejected when $T_{\max}$ is too large to be explained by chance variation.

To make inferences, the null hypothesis distribution of $T_{\max}$ is needed. While this distribution is generally intractable for finite n, MacNeill (1974) quantified the asymptotics of a scaled version of $T_{\max}$ via CUSUM statistics. The CUSUM statistic at index k is

$$\mathrm{CUSUM}_k = \frac{1}{\sqrt{n}}\left(\sum_{j=1}^k X_j - \frac{k}{n}\sum_{j=1}^n X_j\right). \tag{3.1}$$

One views CUSUM_k as a scaled difference between $k^{-1} \times \sum_{t=1}^k X_t$ and $(n-k)^{-1}\sum_{t=k+1}^n X_t$, weighting for differences in the sample sizes of these two segments. Should there be no mean shift, $E[\mathrm{CUSUM}_k] = 0$ for all k. A simple calculation gives $\mathrm{var}(\mathrm{CUSUM}_k) = \sigma^2(k/n)(1-k/n)$, where σ^2 is the variance of all X_t. As many authors show, this nonuniform variance means that changepoints occurring near the data boundaries are more difficult to detect; hence, the CUSUM test has trouble (comparatively) in detecting mean shifts occurring away from the middle of the data.

If the data are Gaussian with known variance σ^2, then the LR test statistic for a change in mean at index k, denoted by Λ_k, is related to CUSUM_k (Csörgő and Horváth 1997) via

$$-2\log\Lambda_k = \mathrm{CUSUM}_k^2 \Big/ \left(\sigma^2 \frac{k}{n}\left(1-\frac{k}{n}\right)\right). \tag{3.2}$$

The essential difference between the LR and CUSUM statistics at index k is the denominator factor of $(k/n)(1-k/n)$ in the LR statistic. This should be viewed as follows: the LR test incorporates more information about the changepoint location than the CUSUM statistic. Motivated by (3.2), an adjusted CUSUM statistic is defined via

$$T_k^2 = \mathrm{CUSUM}_k^2 \Big/ \left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right).$$

To detect a single mean shift at an unknown time, one can examine

$$\mathrm{CUSUM}_{\max} = \max_{1\le k\le n} \frac{|\,\mathrm{CUSUM}_k\,|}{\hat{\sigma}},$$

where $\hat{\sigma}$ is any consistent null hypothesis estimator of σ (e.g., the sample standard deviation). For asymptotics, the statistic $\max_{1\le k\le n} T_k^2/\hat{\sigma}^2$ converges to infinity as $n \to \infty$, the divergence being attributable to the fact that the maximum is taken over the whole of $\{1, \ldots, n\}$. Whereas one can scale $\max_{1\le k\le n} T_k^2/\hat{\sigma}^2$ to an extreme value Gumbel law after appropriate standardization (see Darling and Erdős 1956 and Csörgő and Horváth 1997), power from such extreme value tests is comparatively poor. An alternative approach truncates the set of times allowed as changepoints (we call this the admissible set K) at its boundaries. Specifically, one examines

$$T_{\max}^2 = \frac{\max_{\ell\le k/n\le h} T_k^2}{\hat{\sigma}^2} \tag{3.3}$$

for some fixed ℓ and h satisfying $0 < \ell < h < 1$. While it may not be prudent to truncate data in settings where a changepoint must be detected quickly after it occurs (e.g., a cancer patient seeking to diagnose the disease as soon as possible after its onset), there is little harm in truncating boundary data in our applications. Hence, we proceed in this manner.

The asymptotic distributions for the CUSUM and adjusted LR statistics are quantified next; the result is essentially a consequence of the functional central limit theorem.

Theorem 1. Assume that $\{X_1, \ldots, X_n\}$ is an IID sequence of with finite nonzero variance σ^2. Then

$$\mathrm{CUSUM}_{\max} \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1} |B(t)|;$$
$$T_{\max}^2 \xrightarrow{\mathcal{D}} \sup_{\ell\le t\le h} \frac{B^2(t)}{t(1-t)}, \tag{3.4}$$

where $\{B(t)\}$ denotes a Brownian bridge process on [0, 1].

The multivariate case will become important later. Let $\{\mathbf{X}_j\}_{j=1}^n$ be an IID sequence of d-dimensional vectors with $\mathbf{X}_j = \{X_{1,j}, \ldots, X_{d,j}\}'$ and suppose that $\mathbf{X}_j$ has uncorrelated components (independence is not needed). Let $\mathrm{var}(X_{i,j}) = \sigma_i^2$

be the variance of the ith component. Define a CUSUM statistic for component i and time k via

$$\mathrm{CUSUM}_{i,k} = \frac{1}{\sqrt{n}}\left(\sum_{j=1}^{k} X_{i,j} - \frac{k}{n}\sum_{j=1}^{n} X_{i,j}\right). \tag{3.5}$$

Joint convergence of quadratic forms of the components in (3.5) can be obtained from a multi-dimensional version of the functional central limit theorem generally attributed to Donsker. See Meerschaert and Sepanski (2002) for general statements and proofs of such results.

To quantify the multivariate version of (3.4), let $\{B_1(t)\}$, $\{B_2(t)\}, \ldots, \{B_d(t)\}$ be d independent Brownian bridge processes and set $B^{(d)}(t) = \sum_{j=1}^{d} B_j(t)^2$. Using the multidimensional functional central limit theorem, under the null hypothesis that $\{\mathbf{X}_j\}_{j=1}^{n}$ is IID with uncorrelated components and $\mathrm{var}(X_{i,j}) = \sigma_i^2$, one can show that

$$\max_{\ell \le k/n \le h}\left\{\sum_{i=1}^{d} \sigma_i^{-2}\,\mathrm{CUSUM}_{i,k}^2 \Big/ \left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)\right\} \xrightarrow{\mathcal{D}} \sup_{\ell \le t \le h} \frac{B^{(d)}(t)}{t(1-t)} \tag{3.6}$$

for each $0 < \ell < h < 1$. This result is our major technical tool. For the null hypothesis percentiles of such statistics, Csörgő and Horváth (1997) use a result of Vostrikova (1981) to show that

$$\begin{aligned} P&\left\{\sup_{\ell \le t \le h}\left(\frac{B^{(d)}(t)}{t(1-t)}\right) \ge x\right\} \\ &= \frac{x^{d/2}e^{-x/2}}{2^{d/2}\Gamma(d/2)} \\ &\quad \times \left\{\left(1-\frac{d}{x}\right)\log\left(\frac{(1-\ell)h}{\ell(1-h)}\right) + \frac{4}{x} + O\left(\frac{1}{x^2}\right)\right\} \end{aligned} \tag{3.7}$$

as $x \to \infty$. Here, $O(x^{-2})$ denotes a remainder term that goes to zero no slower than x^{-2} as $x \to \infty$. When the order term is disregarded, simulations show that (3.7) provides accurate (and even conservative) tail probability approximations. For discussion on selection of ℓ and h, see Miller and Siegmund (1982), Andrews (1993), and Csörgő and Horváth (1997).

4. THE $\chi^2_{\max}$ TEST

This section introduces a changepoint detection statistic for multinomial and Poisson data and derives its asymptotic null hypothesis distribution. Since any variable can be partioned into categories, the methods are perhaps best viewed as a nonparametric test for changes in distribution.

4.1 Detecting Changes in Multinomial Data

Suppose we wish to assess whether or not changes have occurred in a discrete sequence. One should not expect a CUSUM mean shift procedure to work well in all cases. For example, consider a random sequence where each observation can be 1, 2, or 3, with respective probabilities of 1/3, 1/3, and 1/3. The mean of such data is 2, and this remains so should the respective categorical probabilities shift to 1/4, 1/2, and 1/4. A more powerful procedure would partition the outcomes into categories and then consider categorical frequencies, say with Pearson's χ^2 test. Specifically, our methods will construct a χ^2 variate for each admissible changepoint time. The maximum of these statistics over all admissible changepoint times is then used to make conclusions. In general, χ_k^2 will be correlated in k.

Maximums of χ^2 random variables have been previously used in the literature, primarily in biostatistics (Halpern 1982; Koziol 1991; Betensky and Rabinowitz 1999) where they are called maximally selected chi-square statistics and are used to compare the sampling distributions of two or more independent samples.

Our analysis is similar to a goodness-of-fit test. Partition the real number line into m classes (or categories) labeled $\mathcal{I}_1, \ldots, \mathcal{I}_m$. Let $N_{i,t} = 1_{\mathcal{I}_i}(X_t)$ be an indicator variable that is unity when X_t falls into category $\mathcal{I}_i$. Then for each $t = 1, \ldots, n$, $\mathbf{N}_t = \{N_{1,t}, \ldots, N_{m,t}\}'$ is a multinomial observation with one trial and probability vector $\mathbf{p}(t) = \{p_1(t), \ldots, p_m(t)\}'$, where $p_i(t) = \mathrm{E}(N_{i,t})$. We will test the null hypothesis that $\mathbf{p}(t)$ is constant in t against the alternative that one or more (and hence at least two) of the $p_i(t)$'s change at an unknown time c.

If k is an admissible changepoint time, let $O_{i,k} = \sum_{t=1}^{k} N_{i,t}$ be the frequency of category i over the first k observations and $O_{i,k}^* = \sum_{t=k+1}^{n} N_{i,t}$ be the frequency of category i over the last $n-k$ observations. Let $O_i = O_{i,n} = \sum_{t=1}^{n} N_{i,t}$ be the category i frequency over the entire data record. Under the null hypothesis that the categorical probabilities are constant in time, an estimator of $p_i \equiv p_i(t)$ is $\hat{p}_i = O_i/n$. When the alternative is true with a changepoint at time k, an estimator of the category i probability before the changepoint time is $\hat{p}_{i,k} = O_{i,k}/k$; an estimator of the category i probability after the changepoint time is $\hat{p}_{i,k}^* = O_{i,k}^*/(n-k)$. Letting $\widehat{E[O_{i,k}]} = k\hat{p}_i = kO_i/n$ and $\widehat{E[O_{i,k}^*]} = (n-k)\hat{p}_i = (n-k)O_i/n$, the χ^2 statistic for a change at time k is thus

$$\chi_k^2 = \sum_{i=1}^{m} \frac{(O_{i,k} - \widehat{E[O_{i,k}]})^2}{\widehat{E[O_{i,k}]}} + \sum_{i=1}^{m} \frac{(O_{i,k}^* - \widehat{E[O_{i,k}^*]})^2}{\widehat{E[O_{i,k}^*]}}. \tag{4.1}$$

Our major result is the following. The result is proven in the Appendix.

Theorem 2. Let χ_k^2 be as in (4.1) and $\chi^2_{\max} = \max_{\ell \le k/n \le h} \chi_k^2$. Then under a null hypothesis that $\mathbf{p}(t)$ does not change in t,

$$\chi^2_{\max} \xrightarrow{\mathcal{D}} \sup_{\ell \le t \le h} \frac{B^{(m-1)}(t)}{t(1-t)}.$$

p-values for this test are approximated using (3.7) with $d = m-1$.

Two remarks regarding Theorem 2 should be made. First, the asymptotic approximation is applicable whether the sample size is deterministic or Poisson. For example, if one multinomial observation is sampled at each epoch, then the theorem applies. Also, if the total number of observations, say N, has a Poisson distribution, then conditional on $N = n$ where n is large, the theorem still applies. Later, our use of Theorem 2 is set in the latter circumstances; here, $n = 1410$ is the total number of observed cyclones.

Second, the $\chi^2_{\max}$ test reinforces why the admissible set should be truncated away from the boundaries. Elaborating, the standard conventions needed to apply Pearson's test (i.e., to have χ_k^2 approximately chi-squared distributed) require that $E[O_{i,k}]$ and $E[O_{i,k}^*]$ exceed unity for all i and k and at least 80% of the $E[O_{i,k}]$'s and $E[O_{i,k}^*]$'s are 5 or greater for each k. These conditions are clearly violated when k is sufficiently close to unity or n.

4.2 Tests for Poisson Data

Tropical cyclone counts are frequently modeled with Poissonian dynamics (Mooley 1981; Thompson and Guttorp 1986; Solow 1989; Lund 1994). While one can add batch arrivals, periodic features, and meteorological covariates such as El-Nino and The North Atlantic Oscillation activity indices to stationary Poisson process models to help explain the slight overdispersion seen in the actual year-to-year counts, Poisson models are fundamental in a rudimentary sense and will later be seen to describe the annual counts well. Let X_t denote the number of cyclones that occur in year t. Under general Poisson arrival assumptions (the arrival rate can vary in time but is periodic with a period of one year), $\{X_t\}$ is IID with a marginal Poisson distribution. To test a null hypothesis of a constant mean $\lambda \equiv E[X_t]$ against an alternative consisting of a shift in the mean at an unknown time, we simply examine a version of (3.3):

$$D_{\max} = \max_{\ell \le k/n \le h} D_k, \quad \text{where } D_k = \text{CUSUM}_k^2 \Big/ \left(\hat{\lambda} \frac{k}{n} \left(1 - \frac{k}{n} \right) \right).$$

Some similarities to the chi-square test statistic of the last section are evident. Specifically, let $C_k = \sum_{t=1}^k X_t$ be the number of cyclones in the first k years and let $C_k^* = \sum_{t=k+1}^n X_t$ be the number of cyclones in the last $n-k$ years. Then $\hat{\lambda} = C_n/n$ estimates λ when there are no changes and $\hat{\lambda}_k = C_k/k$ and $\hat{\lambda}_k^* = C_k^*/(n-k)$ estimate the Poisson mean before and after a changepoint at time k, respectively. With $\widehat{E[C_k]} = k\hat{\lambda} = kC_n/n$ and $\widehat{E[C_k^*]} = (n-k)\hat{\lambda} = (n-k)C_n/n$, one can algebraically show that

$$D_k = \left(C_k - \frac{k}{n} C_n \right)^2 \Big/ \left(n\hat{\lambda} \frac{k}{n} \left(1 - \frac{k}{n} \right) \right) = \frac{(C_k - \widehat{E[C_k]})^2}{\widehat{E[C_k]}} + \frac{(C_k^* - \widehat{E[C_k^*]})^2}{\widehat{E[C_k^*]}}.$$

Note that D_k itself has the classic chi-square form in the summands: observed minus expected squared over expected. Theorem 1 now gives the following result.

Theorem 3. If $X_1, \ldots, X_n$ is IID and Poisson, then

$$D_{\max} \xrightarrow{\mathcal{D}} \sup_{\ell \le t \le h} \frac{B^2(t)}{t(1-t)}, \tag{4.2}$$

and therefore (3.7) with $d = 1$ can be used to find p-values of this test.

Since we are assuming Poisson marginals, a likelihood ratio test also merits exploration. It can be shown that such a likelihood ratio statistic, when cropped to the same admissibility set as the $\chi_{\max}^2$ statistic, has the same limiting distribution as that in (4.2).

Suppose that Y_j is a covariate (we work with wind speed, but the methods could apply to other covariates) for the jth storm (the storms are time ordered) and that Y_j lies in one of m disjoint categories. We assume that Y_j does not depend on the Poisson arrival times of the storms. Consider the m-dimensional vector $\mathbf{X}_t = \{X_{1,t}, X_{2,t}, \ldots, X_{m,t}\}'$. Here, $X_{i,t}$ is the total number of category i storms during the tth year and $\sum_{i=1}^m X_{i,t}$ is the total number of cyclones reported in year t, which is assumed to follow a Poisson distribution with parameter $\lambda(t)$. Also, $p_i(t)$ denotes the probability that any year t storm has a covariate that falls into category i. From the assumed independence of arrivals and wind speeds, $E(X_{i,t}) = \lambda(t)p_i(t)$. Under a null hypothesis of no changes in arrival rate or covariates, $\lambda(t)p_i(t) = \lambda p_i$ for all $i = 1, 2, \ldots, m$ and $t = 1, 2, \ldots, n$.

Set $C_{i,k} = \sum_{t=1}^k X_{i,t}$ and $C_{i,k}^* = \sum_{t=k+1}^n X_{i,t}$ for any admissible changepoint time k. Observe that $\widehat{\lambda p_i} = C_{i,n}/n$ estimates λp_i under the null hypothesis and that $\widehat{\lambda p}_{i,k} = C_{i,k}/k$ and $\widehat{\lambda p}_{i,k}^* = C_{i,k}^*/(n-k)$ estimate λp_i before and after a changepoint at time k, respectively. The estimator of p_i under the null hypothesis is $\hat{p}_i = C_{i,n}/\sum_{i=1}^m C_{i,n}$. Also, $\widehat{E[C_{i,k}]} = k\widehat{\lambda p_i} = kC_{i,n}/n$ and $\widehat{E[C_{i,k}^*]} = (n-k)\widehat{\lambda p_i} = (n-k)C_{i,n}/n$. The χ^2 statistic for testing for a changepoint at time k becomes

$$\chi_k^2 = \sum_{i=1}^m \frac{(C_{i,k} - \widehat{E[C_{i,k}]})^2}{\widehat{E[C_{i,k}]}} + \sum_{i=1}^m \frac{(C_{i,k}^* - \widehat{E[C_{i,k}^*]})^2}{\widehat{E[C_{i,k}^*]}}. \tag{4.3}$$

Computations as before show that

$$\chi_k^2 = \frac{n}{k(n-k)} \sum_{i=1}^m \frac{1}{\widehat{\lambda p_i}} \left(C_{i,k} - \frac{k}{n} C_{i,n} \right)^2 = 1 \Big/ \left(\frac{k}{n} \left(1 - \frac{k}{n} \right) \right) \sum_{i=1}^m \frac{1}{\widehat{\lambda p_i}} (\text{CUSUM}_{i,k})^2, \tag{4.4}$$

where $\text{CUSUM}_{i,k}$ is as in (3.5). Under a null hypothesis of no changes in the Poisson arrival rate or the covariate categorical probabilities, the following hold. By thinning properties of Poisson processes, $X_{i,t}$ has a Poisson distribution with mean λp_i. Hence, $\widehat{\lambda p_i} = n^{-1} \sum_{t=1}^n X_{i,t}$ consistently estimates λp_i. Since $\text{cov}(X_{i,t}, X_{i',t}) = 0$ when $i \ne i'$, $X_{i,t}$ and $X_{i',t}$ are uncorrelated. Applying (3.6) gives the following theorem.

Theorem 4. Under the above setup, if $\lambda(t)p_i(t)$ is constant in t,

$$\chi_{\max}^2 = \max_{\ell \le \frac{k}{n} \le h} \chi_k^2 \xrightarrow{\mathcal{D}} \sup_{\ell \le t \le h} \frac{B^{(m)}(t)}{t(1-t)}, \tag{4.5}$$

and therefore (3.7) with $d = m$ can be used to find p-values of this test.

In comparing Theorems 2 and 4, note the additional degree of freedom, $d = m$, in the asymptotic distribution in Theorem 4. The Theorem 4 statistic signals changes in the Poisson rate and/or the categorical probabilities. In our applications, the asymptotics are in terms of the number of years of data.

Before proceeding, we make several comments. First, our work resides in the at most one changepoint (AMOC) domain. While multiple changepoints are frequently encountered in practice and expected here, we will be able to make conclusions by subsegmenting the data once changepoints are found. By subsegmenting, we mean that once a first changepoint is found, the series is subdivided into two segments about the flagged changepoint time. These two segments are then examined for additional changepoints. The procedure is repeated until no subsegment is judged to contain additional changepoints. While subsegmenting typically works well for series when the mean shifts are all in the same direction (either increasing or decreasing), the procedure can also be fooled. On the other hand, multiple changepoint methods are by no means unflawed. For instance, Jewson and Penzer (2008) find the optimal changepoint times conditional on the number of changepoints, and then concede difficulty with estimating the number of changepoints. Frequently, multiple changepoint methods do not readily provide p-values for simple tests. For treatments of this issue, see Albert and Chib (1993), Giron, Moreno, and Casella (2007), and Fearnhead and Vasileiou (2009) for Bayesian approaches and Lu, Lund, and Lee (2010) for a time series perspective.

We also mention techniques that relax the IID Poisson assumption. Theorem 1 is still applicable to independent non-Poisson data. Also, a short-memory stationary dependence structure in the model errors can be accommodated in CUSUM tests as long as a correlation-adjusted variance estimate is used in place of σ^2 (see Berkes, Gombay, and Horváth 2009 and Robbins et al. 2011 for the technical spirit of the details). One can also fit an autoregressive moving-average (ARMA) model to the data and then apply CUSUM tests to the ARMA residuals (Bai 1993; Robbins et al. 2011); however, one must proceed with caution when fitting ARMA models to count data as it is not always clear that stationary ARMA models with prescribed marginal distributions exist.

5. NORTH ATLANTIC BASIN CYCLONES

Our work will partition the storm wind speeds into five classes based on the common Saffir–Simpson scale. The first category contains storms whose peak wind speed never reached hurricane status (40–73 mph), the second corresponds to Saffir–Simpson category 1 hurricanes (74–95 mph), the third is for Saffir–Simpson category 2 hurricanes (96–110 mph), the fourth is for Saffir–Simpson category 3 hurricanes (111–130 mph) and the fifth contains Saffir–Simpson category 4 or 5 hurricanes (131 or greater mph). We combine Saffir–Simpson category 4 and 5 storms because there are so few category 5 storms. We use $\ell = 1 - h = 0.05$ and $\alpha = 0.05$ throughout this study.

5.1 Changes in Storm Frequency and Strengths

We proceed by jointly testing for changes in the yearly cyclone counts (which are assumed to have a Poisson distribution) and/or their categorical wind speeds. Applying Theorem 4 with $m = 5$ to the entire data sequence gives $\chi^2_{\max} = 109.182$ with a p-value that is less than 10^{-5}. The estimated changepoint time is $\hat{c} = 80$ (1930). Because the pre-1900 data are somewhat unreliable, we reran this analysis with only the 1900–2008 data. This analysis gives $\chi^2_{\max} = 61.567$, $\hat{c} = 95$ (1994), and a p-value that is less than 10^{-5}. Figure 3 plots the year versus its chi-square statistic for the 1851–2008 and 1900–2008 data segments along with 95% confidence thresholds (they are approximately the same for both tests). Clearly, the no change null hypothesis is rejected, indicating potential changepoints circa 1930 and 1995.

As the above tests do not indicate whether the changes are due to arrival rates, wind speeds, or both we now examine

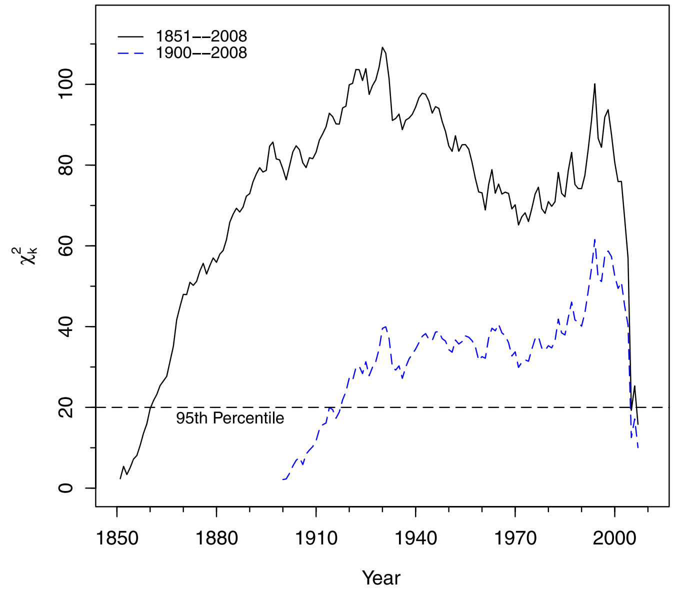


Figure 3. Chi-square statistics for joint changepoints of yearly counts and wind speeds. The online version of this figure is in color.

these two variables separately. First, we look at the yearly storm counts. Lund (1994) examined the annual storm counts from 1871–1990 via least squares methods and found a change in frequencies circa 1931. To confirm this, we applied Theorem 3 to the annual storm counts from 1871–1990. Here, $n = 120$, $D_{\max} = 20.015$, and a p-value of approximately 0.00047 was obtained; the estimated changepoint time is $\hat{c} = 60$ (1930). Repeating this test for the storm counts in the 1931–2008 segment gives $D_{\max} = 28.920$. Here, $n = 78$ and a p-value of approximately 0.00001 was obtained with an estimated changepoint at $\hat{c} = 64$ (1994). Theorem 3 applied to the entire data set (1851–2008, $n = 158$) also signals a changepoint in 1994 ($\hat{c} = 144$) with $D_{\max} = 60.593$ and a p-value that is less than 10^{-5}. A plot of year k versus the D_k statistic for 1851–2008, 1931–2008, and 1871–1990 storm counts is presented in Figure 4.

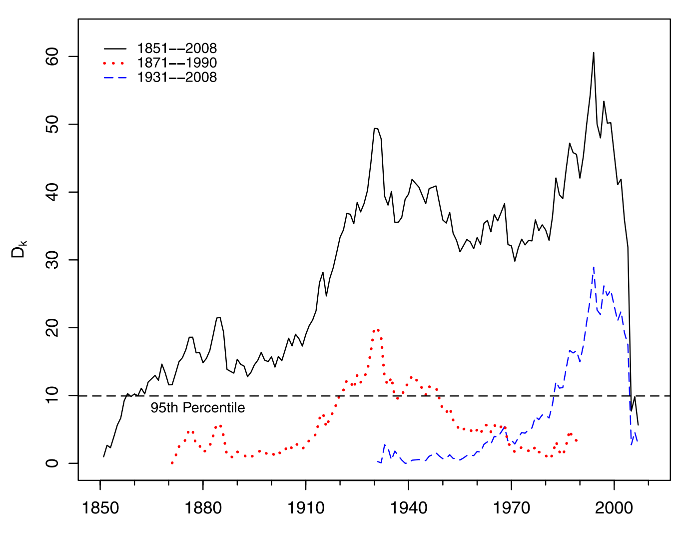


Figure 4. Chi-square statistics for changepoints of yearly counts. The online version of this figure is in color.

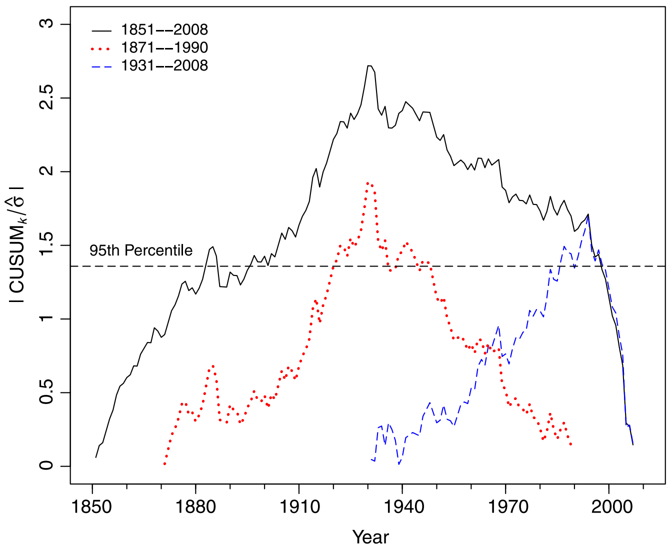


Figure 5. CUSUM statistics for changepoints of yearly counts. The online version of this figure is in color.

The CUSUM test in Theorem 1, which does not require any Poissonian assumptions, can also be used to test for changes in the yearly storm counts. Figure 5 shows the CUSUM statistics for the three segments in the above paragraph. For the 1871–1990 segment, $\mathrm{CUSUM}_{\max} = 1.930$ with a p-value of 0.00116 and $\hat{c} = 60$ (1930). For the 1931–2008 segment, $\mathrm{CUSUM}_{\max} = 1.703$ with a p-value of 0.00606 and $\hat{c} = 64$ (1994). For the 1851–2008 data, $\mathrm{CUSUM}_{\max} = 2.719$ with a p-value less than 10^{-5} and $\hat{c} = 80$ (1930). The CUSUM tests for the 1871–1990 and 1931–2008 segments essentially give the same estimated changepoint times as the Theorem 3 tests. Conclusions about the most significant changepoint differ, however, when the entire storm count sequence is considered. In particular, the generic CUSUM test in Theorem 1 identifies 1930 as the most significant changepoint whereas the Theorem 3 categorical analysis flags 1994 as the most prominent changepoint. As 1994 is somewhat close to the last year of data, we believe that the example simply illustrates the difficulty that CUSUM methods have at detecting changepoints that occur near the boundaries.

To check on the multiple changepoint aspect and the segmentation procedure, a minimum description length (MDL) procedure was coded with a Poisson likelihood as in Lu, Lund, and Lee (2010). This procedure estimates two changepoints in total in the annual storm count record at times 1930 and 1995 again. Hence, the segmentation appears reasonable. We again caution the reader that segmentation procedures can be fooled when the mean shifts orient themselves in differing directions. As another check, we also multiplied the pre-1930 data by a constant to bring the sample mean of this segment to that of the 1930–1994 segment. Then we applied the CUSUM test in Theorem 1 (the data are no longer Poisson) and found the 1995 changepoint again with a p-value of 0.000063.

We now consider changes in storm strength as measured by their peak wind speed during the life of the storm. The peak wind speeds of all storms (1851–2008, $n = 1410$) are ordered via their time of arrival and are plotted in Figure 1. To avoid any confusion, we plot both the arrival date and the index of the storm (from 1 to 1410) on the abscissa scale in future graphs.

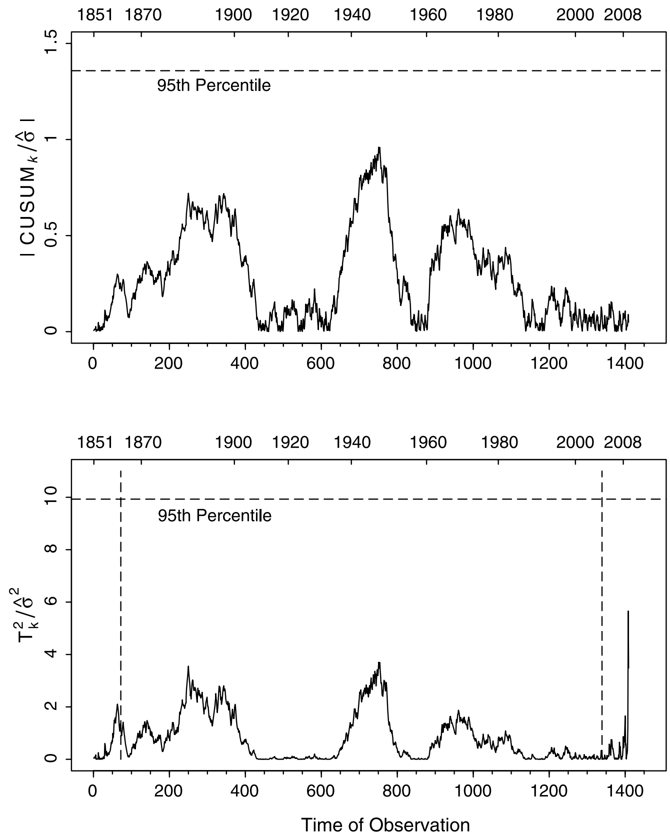


Figure 6. $|\mathrm{CUSUM}|/\hat{\sigma}$ (top) and $T_k^2/\hat{\sigma}^2$ (bottom) statistics for peak wind speed in the 1851–2008 data ($n = 1410$).

The two graphics in Figure 6 plot the storm number k versus $|\mathrm{CUSUM}_k|/\hat{\sigma}$ and $T_k^2/\hat{\sigma}^2$. Here, we are simply applying CUSUM and likelihood methods to the raw wind speeds in an effort to identify changepoints. The horizontal lines in Figure 6 depict a 95% confidence thresholds. The vertical lines in the bottom plot depict the boundary truncations of $\ell = 1 - h = 0.05$. The CUSUM test shows that $\mathrm{CUSUM}_{\max} = 0.960$ with a p-value of 0.3152 at $k = 751$ (1948). The adjusted CUSUM method has $T_{\max}^2 = 3.703$ with a p-value of 0.6483 at $k = 751$ (1948). These simple tests do not show significant evidence of a mean shift in wind speeds.

Next, we apply the $\chi_{\max}^2$ test to peak wind speeds when the 1410 wind speeds are partitioned into the same five classes used in the joint test. Using Theorem 2 with $m = 5$, we find $\chi_{\max}^2 = 81.003$ with a p-value that is less than 10^{-5}. Here, $\hat{c} = 354$ (in 1898). This highly significant changepoint is graphically displayed in Figure 7. In short, the categorical test finds a changepoint that was missed by generic CUSUM techniques.

A changepoint circa 1900 in the wind speeds seems plausible given the purported unreliability of the pre-1900 data. But one may ask why the circa 1900 changepoint was undetected by generic CUSUM methods but flagged with strong significance by the $\chi_{\max}^2$ method. Table 1 shows the estimated categorical probabilities before and after the estimated changepoint time. Observe that the categorical probabilities of storms with low (tropical storms) and high (Saffir–Simpson category 4 and 5

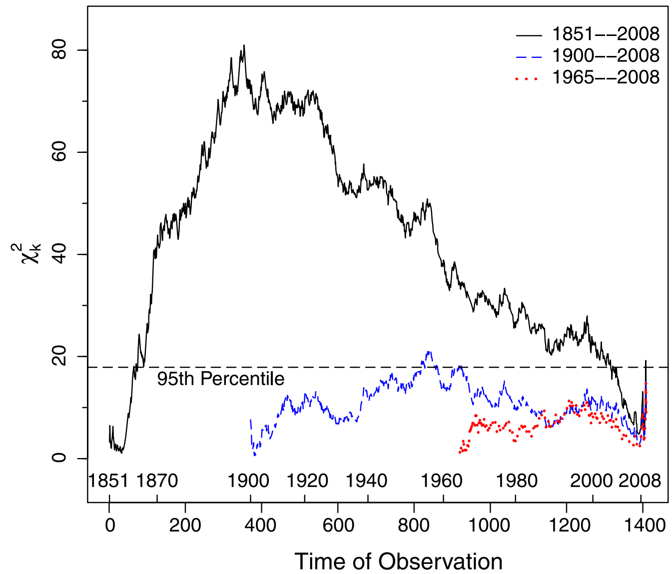


Figure 7. Chi-square statistics for peak wind speeds in Atlantic cyclones from 1851–2008 ($n = 1410$) and 1900–2008 ($n = 1040$). The online version of this figure is in color.

hurricanes) wind speeds increase markedly after the changepoint time. Phrased another way, this seems to be a distributional change that did not cause a change in mean.

Segmenting the data to the 1040 storms arriving in 1900–2008, a changepoint in 1956 ($\hat{c} = 471$) is found; here, $\chi^2_{\max} = 21.038$ and has a p-value of 0.01482. Hence, there appears to be a another mean shift in the wind speeds circa 1956, though the shift does not appear to be as significant as the circa 1900 shift. Figure 7 provides graphical support.

5.2 Changes in Storm Covariates

Changes in tropical cyclone data are commonly thought to be due to advances in storm data collection techniques, and it seems reasonable that any changes in surveying methods may also cause changes in several of the covariate series. Such covariates include each storm's latitude and longitude (at the time it first achieved its maximum windspeed), duration (in days), and occurrence of landfall within the continental United States. Landsea (2007) posits that cyclone activity occurring over the open Atlantic Ocean often went undetected prior to the advent of aircraft and satellite reconnaissance. If this theory is correct, then changepoints in the covariates seem plausible.

The adjusted CUSUM test of Theorem 1 test was applied to latitude, longitude, duration, and landfall probability for the years 1900–2008. Mean shifts are detected in each of the series with respective p-values of 0.00585, 2.01×10^{-11}, 0.00033, and 0.00431. Figure 8 displays values of $T_k^2/\hat{\sigma}^2$ for each covariate.

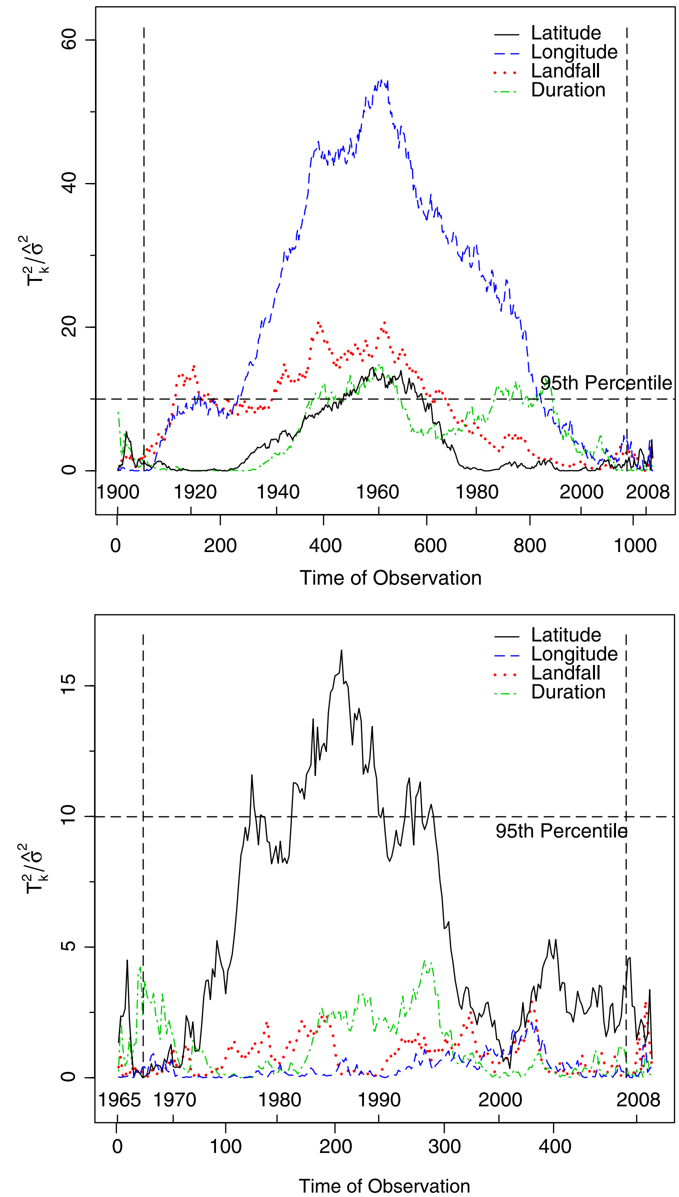


Figure 8. Plots of $T_k^2/\hat{\sigma}^2$ for each covariate for 1900–2008 (top) and 1965–2008 (bottom). The online version of this figure is in color.

The average latitude shifts up (northward), the average longitude shifts down (eastward), the average duration shifts up, and the landfall rate shifts down. Note that the most drastic change occurs in the longitude series. Each of these observations is in agreement with the theories in Landsea (2007). The estimated changepoint for each of these covariate series is circa 1960, which coincides in the introduction of satellite technology (this perhaps also explains the circa 1956 change in storm strengths).

Table 1. Categorical probabilities before and after the changepoint for the $\chi^2_{\max}$ test

Location	Tropical storm	Category 1 hurricane	Category 2 hurricane	Category 3 hurricane	Category 4 & 5 hurricane
Before 1900	0.280	0.302	0.260	0.130	0.028
After 1900	0.437	0.214	0.120	0.112	0.116

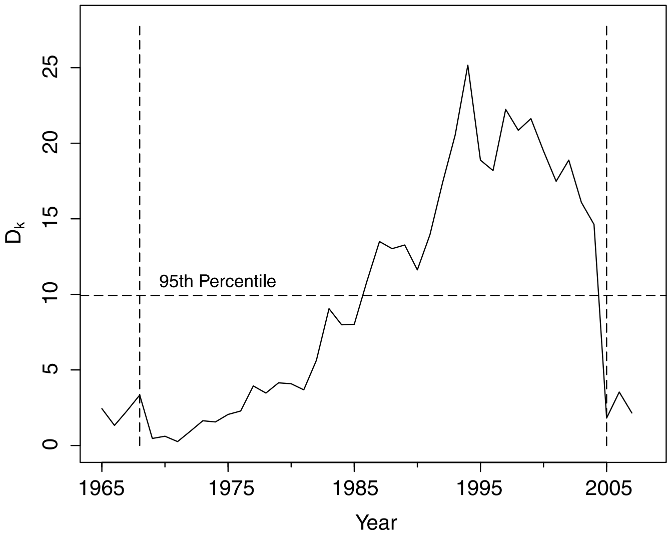


Figure 9. Chi-square statistics for Atlantic cyclone counts from 1965–2008 ($n = 44$).

The reliability of the satellite era data (1965–2008) is also of interest. Each of the previous tests were rerun using data from these years only. Figures 7 and 8 suggest no changes in storm strength, longitude, duration, and landfall probability during this time period. However, a (somewhat inexplicable) downward mean shift in the latitude series was detected circa 1986. The resulting p-value is 0.00260. Overall, these results support the notion that the satellite era data are reliable.

5.3 The Circa 1995 Changepoint

A question about the circa 1995 changepoint is whether it remains significant when only the 44 years of reliable (1965–2008) data are tested. The test finds $D_{\max} = 25.164$ with a p-value of less than 10^{-5}. Here, the changepoint is flagged at $\hat{c} = 30$ (1994). Figure 9 provides graphical support of this conclusion.

One can also ask whether it is appropriate to apply asymptotic methods when $n = 44$ (or when $n = 158$ for the 1851–2008 segment). To address this question, we simulated 100,000 series of Poisson data with a homogeneous mean of 10 for each of $n = 1000$, $n = 158$, and $n = 44$. The results are summarized in Table 2. For a target Type I error of 0.05, Theorem 3 and (3.7) suggest $P(D_{\max} > 9.929) = 0.0500$ using $\ell = 1 - h = 0.05$. The simulations for $n = 1000$ estimated $P(D_{\max} > 9.929) = 0.0433$ using $\ell = 1 - h = 0.05$. In this case, the simulated Type I error is reasonably close to its target value. The simulations for $n = 44$ resulted in an estimate of $P(D_{\max} > 9.929) = 0.0243$ using $\ell = 1 - h = 0.05$. Here, the simulated Type I error probability is significantly less than the target value when $n = 44$, implying that the test is conservative.

Table 2. Simulated Type I error rates for 0.05-level test based on Theorem 3

Sample size	# rejections/100,000
$n = 1000$	0.0433
$n = 158$	0.0345
$n = 44$	0.0234

Table 3. Goodness-of-fit tests for Poisson marginals, 1931–1994 counts

Data Series	$\bar{x}$	s^2	χ^2 p-value	IOD p-value
1931–1994	9.688	9.679	0.500	0.957

This only enhances the significance of the circa 1995 changepoint. Finally, the simulated Type I error probability for $n = 158$ is $P(D_{\max} > 9.929) = 0.0345$ using $\ell = 1 - h = 0.05$. Further simulations in Robbins et al. (2011) demonstrate that asymptotic changepoint tests similar to the ones used in this article tend to be conservative.

5.4 Test Assumptions: Robustness and Validation

We briefly consider changepoint tests for the annual storm counts which do not require the assumption of IID Poisson data. The CUSUM test of Theorem 1 (which only requires IID data) when applied to 1965–2008 segment finds the 1995 changepoint with a p-value of 0.00357. The CUSUM test of Berkes, Gombay, and Horváth (2009), which allows for autocorrelation, also detects the 1995 changepoint at the 5% significance level. For this test, we used a Bartlett-based expression to estimate the long-run series variance; the results are somewhat dependent on a bandwidth parameter used in the calculation of this estimate. In applying the methods of Bai (1993) to the residuals from a first order autoregressive fit of the 1965–2008 storm counts, the 1995 changepoint is flagged with a p-value of 0.0166. The 1930s changepoint is also evident with these methods. As these test do not assume independent Poisson data they are probably not very powerful. Hence, we now investigate the validity of the IID Poisson assumptions.

To validate the Poisson assumption, we first examine the constant mean 1931–1994 segment of annual counts. We consider two methods for testing the goodness of fit of the Poisson distribution: the classic chi-square goodness-of-fit test (χ^2, with bins 0–5, 6–7, 8–10, 11–14, and 15+) and the index of dispersion test (IOD) of Sukhatme (1938) and Okamoto (1955), which examines the ratio of the sample variance (s^2) and the sample mean ($\bar{x}$). Table 3 reports p-values for these tests and suggests that the 1931–1994 counts are reasonably Poisson. The other segments test similarly as Poisson.

We also investigated annual count independence. Figure 10 shows the sample autocorrelation structure of the annual counts after subtracting the mean of the three segments. The bounds are pointwise 95% bounds (pointwise) for white noise. Overall there does not seem to be much correlation. Of course, if there was significant correlation in these counts, we would have an easier time forecasting annual storm counts years in advance.

6. CONCLUDING REMARKS

The categorical changepoint tests have worked well in identifying changes in the Atlantic Basin hurricane record, illuminating features that standard CUSUM and LR mean shift tests miss. Contrary to some theories, we find no evidence of significant recent increases in storm strength or U.S. landfall strike probability. We do, however, find recent increases in storm frequencies circa 1995. Changepoints in many of the cyclone covariates are found circa 1960, which coincides with the onset of

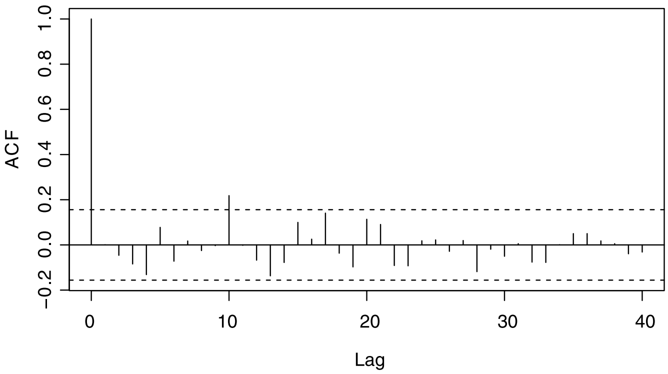


Figure 10. Sample autocorrelation of storm counts for 1851–2008 data after changepoint adjustment.

satellite surveillance; however, the post-1960 data appear reliable. We also find changepoints in the peak wind speeds of the storms circa 1900 and 1960. The circa 1995 changepoint in frequency is possibly explained by the increase of short-duration weak storms in the recent record (Vecchi 2008; Landsea et al. 2010) and/or climate change.

As a nonparametric test for changes in distribution, the $\chi^2_{\max}$ test introduced here seems preferable to standard CUSUM and LR methods. Specifically, the $\chi^2_{\max}$ tests can detect changes in distribution rather than simple changes in mean.

APPENDIX: PROOF OF THEOREM 2

From the definition of $N_{i,t}$, we have

$$\chi_k^2 = \sum_{i=1}^m \left(O_{i,k} - \frac{k}{n}O_i\right)^2 \Big/ (k\hat{p}_i) + \sum_{i=1}^m \left(O_{i,k}^* - \frac{n-k}{n}O_i\right)^2 \Big/ ((n-k)\hat{p}_i).$$

Since $O_{i,k} + O_{i,k}^* = kO_i/n + (n-k)O_i/n$, we obtain $O_{i,k}^* - \frac{n-k}{n}O_i = -(O_{i,k} - kO_i/n)$. Therefore,

$$\begin{aligned}\chi_k^2 &= \sum_{i=1}^m \frac{1}{\hat{p}_i}\left(O_{i,k} - \frac{k}{n}O_i\right)^2 \left(\frac{1}{k} + \frac{1}{n-k}\right) \\ &= \frac{n}{k(n-k)} \sum_{i=1}^m \frac{1}{\hat{p}_i}\left(O_{i,k} - \frac{k}{n}O_i\right)^2.\end{aligned}$$

Using the fact that

$$\sum_{t=1}^k N_{i,t} - \frac{k}{n}\sum_{t=1}^n N_{i,t} = k\left(1 - \frac{k}{n}\right)(\hat{p}_{i,k} - \hat{p}_{i,k}^*), \tag{A.1}$$

we can reexpress χ_k^2 as

$$\chi_k^2 = \frac{k(n-k)}{n} \sum_{i=1}^m \frac{(\hat{p}_{i,k} - \hat{p}_{i,k}^*)^2}{\hat{p}_i}. \tag{A.2}$$

Now define X_k^2 by replacing $\hat{p}_i$ by p_i in (A.2):

$$X_k^2 = \frac{k(n-k)}{n} \sum_{i=1}^m \frac{(\hat{p}_{i,k} - \hat{p}_{i,k}^*)^2}{p_i}.$$

Under the null hypothesis, $\hat{p}_i$ is a $\sqrt{n}$-consistent estimator of p_i; hence, χ_k^2 and X_k^2 have the same limiting distribution. Define the $(m-1)$-dimensional vectors $\mathbf{P} = (p_1, p_2, \ldots, p_{m-1})'$, $\hat{\mathbf{P}}_k = (\hat{p}_{1,k}, \ldots, \hat{p}_{m-1,k})'$, and $\hat{\mathbf{P}}_k^* = (\hat{p}_{1,k}^*, \ldots, \hat{p}_{m-1,k}^*)'$. Also, let $\mathbf{D} = \operatorname{diag}(\mathbf{P})$. Then under the null hypothesis, $\hat{\mathbf{P}}_k - \hat{\mathbf{P}}_k^*$ has zero mean and covariance matrix

$$\mathbf{M} = \frac{n}{k(n-k)}(\mathbf{D} - \mathbf{P}\mathbf{P}').$$

Then $\mathbf{M}^{-1} = k(n-k)\mathbf{A}^{-1}/n$, where

$$\mathbf{A}^{-1} = (\mathbf{D} - \mathbf{P}\mathbf{P}')^{-1} = \left(\mathbf{D}^{-1} + \frac{\mathbf{J}}{p_m}\right).$$

Here, $\mathbf{J}$ is a matrix consisting entirely of unit entries. Observe that under the null hypothesis, $\mathbf{A}$ is the covariance matrix of $(N_{1,t}, \ldots, N_{m-1,t})'$ and $\mathbf{A}^{-1}$ exists and does not depend on n or k. For estimating the category m probabilities, we use

$$\hat{p}_{m,k} = 1 - \hat{p}_{1,k} - \cdots - \hat{p}_{m-1,k} \qquad \text{and}$$
$$\hat{p}_{m,k}^* = 1 - \hat{p}_{1,k}^* - \cdots - \hat{p}_{m-1,k}^*.$$

Using these facts, we see that

$$\begin{aligned} X_k^2 &= \frac{k(n-k)}{n}\left(\sum_{i=1}^{m-1} \frac{(\hat{p}_{i,k} - \hat{p}_{i,k}^*)^2}{p_i} + \frac{(\sum_{i=1}^{m-1} \hat{p}_{i,k}^* - \hat{p}_{i,k})^2}{p_m}\right) \\ &= \frac{k(n-k)}{n}(\hat{\mathbf{P}} - \hat{\mathbf{P}}^*)'\left(\mathbf{D}^{-1} + \frac{1}{p_m}\mathbf{J}\right)(\hat{\mathbf{P}} - \hat{\mathbf{P}}^*) \\ &= (\hat{\mathbf{P}} - \hat{\mathbf{P}}^*)'\mathbf{M}^{-1}(\hat{\mathbf{P}} - \hat{\mathbf{P}}^*) \\ &= \left(\mathbf{M}^{-1/2}(\hat{\mathbf{P}} - \hat{\mathbf{P}}^*)\right)'\left(\mathbf{M}^{-1/2}(\hat{\mathbf{P}} - \hat{\mathbf{P}}^*)\right) \\ &= \mathbf{Z}_k'\mathbf{Z}_k, \end{aligned}$$

where the components in $\mathbf{Z}_k$ are uncorrelated under the null hypothesis. Using $\mathbf{M}^{-1/2} = \sqrt{\frac{k(n-k)}{n}}\mathbf{A}^{-1/2}$ and (A.1), we see that each component of $\mathbf{Z}_k$ has the form

$$\begin{aligned} Z_{j,k} &= \sum_{i=1}^{m-1} \sqrt{\frac{k(n-k)}{n}} a_{ij}\left(\sum_{t=1}^k N_{i,t} - \frac{k}{n}\sum_{t=1}^n N_{i,t}\right) \Big/ (k(1-k/n)) \\ &= \left(\sum_{t=1}^k Y_{j,t} - \frac{k}{n}\sum_{t=1}^n Y_{j,t}\right) \Big/ \sqrt{n(k/n)(1-k/n)}, \end{aligned}$$

where the coefficients $\{a_{ij}\}$ come from $\mathbf{A}^{-1/2}$ and do not depend on n or k. Letting

$$\mathbf{Y}_t = (Y_{1,t}, \ldots, Y_{m-1,t})' = \mathbf{A}^{-1/2}(N_{1,t}, \ldots, N_{m-1,t})',$$

we see that $\mathbf{Z}_k$ consists of scaled CUSUMs of IID vectors with uncorrelated components and a unit variance. It now follows that

$$X_k^2 = \sum_{j=1}^{m-1} \mathrm{CUSUM}_{j,k}^2 \Big/ \left(\frac{k}{n}\left(1 - \frac{k}{n}\right)\right),$$

where CUSUM_j refers to a cumulative sum in $\{Y_{j,t}\}$. Using Slutsky's theorem and (3.6) finishes our work.

[Received January 2010. Revised August 2010.]

REFERENCES

Albert, J. H., and Chib, S. (1993), "Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts," *Journal of Business & Economic Statistics*, 11, 1–15. [94]

Andrews, D. W. K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856. [92]

Anthes, R. A., Corell, R. W., Holland, G., Hurrell, J. W., MacCracken, M. C., and Trenberth, K. E. (2006), "Hurricanes and Global Warming: Potential Linkages and Consequences," *Bulletin of the American Meteorological Society*, 87, 623–628. [89]

Bai, J. (1993), "On the Partial Sums of Residuals in Autoregressive and Moving Average Models," *Journal of Time Series Analysis*, 14, 247–260. [94,97]

Berkes, I., Gombay, E., and Horváth, L. (2009), "Testing for Changes in the Covariance Structure of Linear Processes," *Journal of Statistical Planning and Inference*, 139, 2044–2063. [94,97]

Betensky, R. A., and Rabinowitz, D. (1999), "Maximally Selected χ^2 Statistics for $k \times 2$ Tables," *Biometrics*, 55, 317–320. [92]

Burt, C. C. (2004), *Extreme Weather*, London: W. W. Norton & Company. [90]

Cane, M. A., Clement, A. C., Kaplan, A., Kushnir, Y., Pozdnyakov, D., Seager, R., Zebiak, S. E., and Murtugudde, R. (1997), "Twentieth-Century Sea Surface Temperature Trends," *Science*, 275, 957–960. [89]

Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992), "Hierarchical Bayesian Analysis of Changepoint Problems," *Applied Statistics*, 41, 389–405. [89]

Carlstein, E. (1988), "Nonparametric Change-Point Estimation," *The Annals of Statistics*, 16, 188–197. [90]

Chib, S. (1998), "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 86, 221–241. [89]

Csörgő, M., and Horváth, L. (1987), "Nonparametric Tests for the Change Point Problem," *Journal of Statistical Planning and Inference*, 17, 1–9. [89]

——— (1997), *Limit Theorems in Changepoint Analysis*, Chichester: Wiley. [89,91,92]

Darling, D. A., and Erdős, P. (1956), "A Limit Theorem for the Maximum of Normalized Sums of Independent Random Variables," *Duke Mathematical Journal*, 23, 143–155. [91]

Elsner, J. B., Niu, X., and Jagger, T. H. (2004), "Detecting Shifts in Hurricane Rates Using a Markov Chain Monte Carlo Approach," *Journal of Climate*, 17, 2652–2666. [89]

Emanuell, K. A. (1987), "The Dependence of Hurricane Intensity on Climate," *Nature*, 326, 483–485. [89]

——— (2005), "Increasing Destructiveness of Tropical Cyclones Over the Past 30 Years," *Nature*, 436, 686–688. [89]

Emanuell, K. A., Sundararajan, R., and Williams, J. (2008), "Hurricanes and Global Warming: Results From Downscaling IPCC AR4 Simulations," *Bulletin of the American Meteorological Society*, 89, 347–367. [89]

Fearnhead, P., and Vasileiou, D. (2009), "Bayesian Analysis of Isochores," *Journal of the American Statistical Association*, 104, 132–141. [94]

Fu, Y.-X., and Curnow, R. N. (1990), "Maximum Likelihood Estimation of Multiple Change Points," *Biometrika*, 77, 563–573. [89]

Girón, J. B., Ginebra, J., and Riba, A. (2005), "Bayesian Analysis of a Multinomial Sequence and Homogeneity of Literary Style," *The American Statistician*, 59, 19–30. [89]

Girón J. B., Moreno, E., and Casella, G. (2007), "Objective Bayesian Analysis of Multiple Changepoints for Linear Models," *Bayesian Statistics*, 8, 1–27. [94]

Halpern, J. (1982), "Maximally Selected Chi Square Statistics for Small Samples," *Biometrics*, 38, 1017–1023. [92]

Hinkley, D. V., and Hinkley, E. A. (1970), "Inference About the Change-Point in a Sequence of Binomial Variables," *Biometrika*, 57, 4777–4788. [89]

Hirotsu, C. (1997), "Two-Way Change-Point Model and Its Applications," *Australian Journal of Statistics*, 39, 205–218. [89]

Horváth, L., and Serbinowska, M. (1995), "Testing for Changes in Multinomial Observations: The Landisfarne Scribes Problem," *Scandinavian Journal of Statistics*, 22, 371–384. [89]

Houghton, J. H., Ding, Y., Griggs, D. J., Nogue, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. A. (eds.) (2001), *Climate Change 2001: The Scientific Basis*, Cambridge: Cambridge University Press. [89]

Jarrell, J. D., Hebert, P. J., and Mayfield, M. (1992), "Hurricane Experience Levels of Coastal County Populations From Texas to Maine," NOAA Technical Memorandum NWS NHC 46, Coral Gables, FL, 152 p. [89,90]

Jewson, S., and Penzer, J. (2008), "An Objective Change-Point Analysis of Historical Atlantic Hurricane Numbers," available at *arXiv:physics/0611071v1*. [89,94]

Karl, T. R., and Trenberth, K. E. (2003), "Modern Global Climate Change," *Science*, 302, 1719–1723. [89]

Koziol, J. A. (1991), "On Maximally Selected Chi-Square Statistics," *Biometrics*, 47, 1557–1561. [92]

Landsea, C. W. (2007), "Counting Atlantic Tropical Cyclones Back to 1900," *EOS, Transactions, American Geophysical Union*, 88, 197–208. [89,90,96]

Landsea, C. W., Anderson, C., Charles, N., Clark, G., Dunion, J., Fernandez-Partagas, J., Hungerford, P., Neumann, C., and Zimmer, M. (2004), "The Atlantic Hurricane Database Re-Analysis Project: Documentation for 1851–1910 Alterations and Additions to the HURDAT Database," in *Hurricanes and Typhoons: Past, Present and Future*, eds. R. J. Murname and K.-B. Liu, New York: Columbia University Press, pp. 177–221. [90]

Landsea, C. W., Glenn, D. A., Bredemeyer, W., Chenoweth, M., Ellis, R., Gamache, J., Hufstetler, L., Mock, C., Perez, R., Prieto, R., Sanchez-Sesma, J., Thomas, D., and Woolcock, L. (2008), "A Reanalysis of the 1911–20 Atlantic Hurricane Database," *Journal of Climate*, 21, 2138–2168. [90]

Landsea, C. W., Harper, B. A., Hoaran, K., and Knaff, J. A. (2006), "Can We Detect Trends in Extreme Tropical Cyclones," *Science*, 313, 452–454. [90]

Landsea, C. W., Pielke, R. A., Mestas-Nuñez, A. M., and Knaff, J. A. (1999), "Atlantic Basin Hurricanes: Indices of Climatic Changes," *Climatic Change*, 42, 89–129. [89,90]

Landsea, C. W., Vecchi, G. A., Bengtsson, L., and Knutson, T. R. (2010), "Impact of Duration Thresholds on Atlantic Tropical Cyclone Counts," *Journal of Climate*, 23, 2508–2519. [89,98]

Lu, Q., Lund, R., and Lee, T. C. M. (2010), "An MDL Approach to the Climate Segmentation Problem," *The Annals of Applied Statistics*, 4, 299–319. [94, 95]

Lund, R. (1994), "Annual Arrival Cycle of Atlantic Tropical Cyclones," *Journal of Applied Statistics*, 21, 195–204. [93,94]

MacNeill, I. B. (1974), "Tests for Change of Parameter at Unknown Times and Distributions of Some Related Functionals of Brownian Motion," *The Annals of Statistics*, 2, 950–962. [89,91]

Meerschaert, M. M., and Sepanski, S. J. (2002), "Regular Variation and the Functional Central Limit Theorem for Heavy Tailed Random Vectors," *Publications De L'Institut Mathematique*, 71, 55–62. [92]

Miller, R., and Siegmund, D. (1982), "Maximally Selected Chi Square Statistics," *Biometrics*, 38, 1011–1016. [92]

Mooley, D. A. (1981), "Applicability of the Poisson Probability Model to the Severe Cyclonic Storms Striking the Coast Around the Bay of Bengal," *Sankhyā, Ser. B*, 43, 187–197. [93]

Neumann, C. J., Jarvinen, B. R., McAdie, C. J., and Elms, J. D. (1999), *Tropical Cyclones of the North Atlantic Ocean, 1871–1998*, Asheville, NC: National Climatic Data Center, p. 206. [89,90]

Okomoto, E. S. (1955), "Fit of a Poisson Distribution by the Index of Dispersion," *Osaka Mathematical Journal*, 7, 7–13. [97]

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41, 100–105. [89]

——— (1955), "A Test for a Change in a Parameter Occurring at an Unknown Point," *Biometrika*, 42, 523–527. [89]

Pettitt, A. N. (1980), "A Simple Cumulative Sum Type Statistic for the Change-Point Problem With Zero–One Observations," *Biometrika*, 67, 79–84. [89]

Pielke, R. A., Landsea, C. W., Mayfield, M., Laver, J., and Pasch, R. (2005), "Hurricanes and Global Warming: Potential Linkages and Consequences," *Bulletin of the American Meteorological Society*, 87, 628–631. [89]

Qian, S. Q., Pan, Y., and King, R. S. (2004), "Soil Total Phosphorus Threshold in the Everglades: A Bayesian Changepoint Analysis for Multinomial Response Data," *Ecological Indicators*, 4, 29–37. [89]

Robbins, M., Gallagher, C., Lund, R., and Aue, A. (2011), "Mean Shift Testing in Correlated Data," *Journal of Time Series Analysis*, to appear. DOI: *10.1111/j.1467-9892.201000707.x*. [94,97]

Saunders, M. A., and Lee, A. S. (2008), "Large Contributions of Sea Surface Warming to Recent Increase in Atlantic Hurricane Activity," *Nature*, 451, 557–560. [89]

Smith, A. F. M. (1975), "A Bayesian Approach to Inference About a Change-Point in a Sequence of Random Variables," *Biometrika*, 62, 407–416. [89]

Solow, A. R. (1989), "Reconstructing a Partially Observed Record of Tropical Cyclone Counts," *Journal of Climate*, 2, 1253–1257. [93]

Sukhatme, P. V. (1938), "On the Distribution of χ^2 in Samples of the Poisson Series," *Journal of the Royal Statistical Society, Supplement*, 5, 75–79. [97]

Thompson, M. L., and Guttorp, P. (1986), "A Probability Model for the Severe Cyclonic Storms Striking the Coast Around the Bay of Bengal," *Monthly Weather Review*, 114, 2267–2271. [93]

Vecchi, G. A., and Knutson, T. R. (2008), "On Estimates of Historical North Atlantic Tropical Cyclone Activity," *Journal of Climate*, 21, 3580–3600. [89,98]

Wolfe, D. A., and Chen, Y.-S. (1990), "The Changepoint Problem in a Multinomial Sequence", *Communications in Statistics—Simulation and Computation*, 19, 603–618. [89]

Vostrikova, L. Y. (1981), "Detection of Disorder in a Wiener Process," *Theory of Probability and Its Applications*, 26, 356–362. [92]