

Categorical Changepoints and their Application to Atlantic Tropical Cyclone Changes

MICHAEL ROBBINS*, ROBERT LUND, and COLIN GALLAGHER

Department of Mathematical Sciences

Clemson University

Clemson, SC 29634-0975

QIQI LU

Department of Mathematics and Statistics

Mississippi State University

Mississippi State, MS 39762

August 27, 2009

Abstract: This paper studies changepoint detection in time-ordered sequences of categorical data. When the data are 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 data set 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 significant. The end conclusions here are opposite of those in Dr. Kelvin Droegemeier's July 28, 2009 Senate testimonial.

Key words and phrases: Atlantic Hurricanes, Brownian Bridge, Changepoints, Chi-Square Statistics, Climate Change, CUSUM, Tropical Cyclones.

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 & 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, & Williams 2008; Saunders & Lea 2008) and others arguing that no firm conclusions can yet be made (Pielke, Landsea, Mayfield, Laver, & Pasch 2005; Landsea 2007). Those acknowledging recent changes have differing opinions about the nature of the change. Saunders & 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 & Knutson (2008) and Landsea, Vecchi, Bengtsson, & Knutson (2009) believe that recent increases in storm counts are attributable to the increased number of weak cyclones included in the recent record (these storms are typically of short duration and are difficult to detect without modern sensing techniques).

Rigorous statistical studies on cyclone changes are relatively sparse. Some notable papers include the Markov Chain Monte Carlo methods of Elsner, Niu, & Jaeger (2004) and the Bayesian-based changepoint approach of Jewson & Penzer (2008). While the methods employed by these authors differ from our asymptotic results, these authors agree with our end conclusions of shifts circa 1900, 1935, and 1995. The circa 1995 changepoint is controversial amongst scientists. Indeed, Dr. Kelvin Droegemeier testified to the Senate

on July 28, 2009 essentially purporting the opposite conclusions that we make. Landsea, Pielke, Mestas-Nuñez, and Knaff (1999), Jarrell, Hebert, & Mayfield (1992), and Neumann, Jarvinen, McAdie, & Elms (1999) provide convincing explanations for the circa 1900 and 1935 changepoints in terms of data collection technique changes.

The literature on statistical and mathematical approaches to changepoint problems is vast. Page (1954; 1955) is widely credited with introducing undocumented changepoint problems. Quandt (1958; 1960) extended the methods to linear models and suggested the use of likelihood ratio (LR) tests. Yao & Davis (1984) established the asymptotic null hypothesis properties of a LR statistic for a mean shift in independent Gaussian data. MacNeill (1974) introduced a cumulative sum (CUSUM)-type 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 & Horváth (1997), a comprehensive but technical reference on large sample changepoint testing, provides asymptotic results for LR statistics under general IID settings.

After Hinkley & Hinkley (1970) studied changepoints in binomial data, changepoint detection in multinomial (categorical) data has become an active area of research. The most popular techniques include Bayesian methods (Smith 1975; Carlin, Gelfand & Smith 1992; Qian, Pan & King 2004; Girón, Ginebra & Riba 2005), CUSUM-type methods (Pettitt 1980; Wolfe & Chen 1990) and maximum likelihood methods (Fu & Curnow 1990). Most of these authors analyze the effectiveness (detection power) of changepoint estimators; reliable discussions on null hypothesis distributions of the test statistics are sparse. Two authors (Horváth & Serbinowska 1995; Hirotsu 1997) also study maximums of chi-square statistics and discuss their approximate equivalence to LR tests. Horváth & Serbinowska (1995),

the paper which is most closely related to ours, provide asymptotic distributions of maximums of chi-square statistics that are slightly different from ours. Our technical arguments also differ, our crux essentially being a multivariate version of Donsker's theorem (see (2.8) below). Other authors introduce a nonparametric method for testing for changes in the marginal distribution of a univariate sequence of IID random variables via empirical distribution functions (Csörgő & Horváth 1987; Carlstein 1988). By partitioning such data into categories, our test can also be viewed as a marginal distribution changepoint test.

The objective of this paper is to introduce a χ_{\max}^2 statistic (a maximum of correlated chi-square random variates) as a general categorical changepoint detection statistic and see how it works in identifying times of Atlantic Basin cyclone changes. The rest of this work is organized as follows. Section 2 reviews CUSUM changepoint methods. Section 3 introduces the χ_{\max}^2 statistic and establishes its asymptotic null hypothesis properties. Here, the proposed methods are also linked to classical CUSUM tests. The methods are then applied to the historical record of tropical cyclones in Section 4, where multiple changepoints are found in storm frequencies and wind speeds. Remarks in Section 5 conclude the paper.

2 CUSUM Review

Suppose we wish to determine whether or not there is a mean shift in the observed data X_1, \dots, 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*, then a way of discerning if a mean shift occurred at time c would be to 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_{\max} . The null hypothesis is rejected when T_{\max} is too large to be explained by chance variation. For discussion on the performance of the changepoint estimator (\hat{c}) under alternatives when using statistics similar to the ones presented in this paper see Csörgő and Horváth (1997) and Wolfe and Chen (1990) among others.

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 defined via

$$\text{CUSUM}_k = \frac{1}{\sqrt{n}} \left(\sum_{j=1}^k X_j - \frac{k}{n} \sum_{j=1}^n X_j \right). \quad (2.1)$$

One views CUSUM_k as a scaled difference between $k^{-1} \sum_{t=1}^k X_t$ 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, then $E[\text{CUSUM}_k] = 0$ for all k . A simple calculation gives $\text{Var}(\text{CUSUM}_k) = \sigma^2(k/n)(1-k/n)$, where σ^2 is the variance of all X_i . Hence, if there are no mean shifts, $\text{Var}(\text{CUSUM}_k)$ is maximized at the center of the data record and is relatively smaller near the boundaries of 1 and n . As Robbins, Gallagher, Lund, & Aue (2009) and many earlier authors show, this feature means that changepoints occurring near the data boundaries are seldom detected; hence, the CUSUM test has trouble detecting mean shifts occurring away from the middle of the data sequence.

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 & Horváth 1997) via

$$-2 \log \Lambda_k = \frac{\text{CUSUM}_k^2}{\sigma^2 \frac{k}{n} \left(1 - \frac{k}{n}\right)}. \quad (2.2)$$

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 where the changepoint location is than the CUSUM statistic. Motivated by (2.2), we define an adjusted CUSUM statistic via

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

To detect a single mean shift at an unknown time, one simply examines

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

where $\hat{\sigma}$ is any consistent null hypothesis estimator of σ (e.g., the sample standard deviation). One should not use $\max_{1 \leq k \leq n} T_k^2 / \hat{\sigma}^2$ as a test statistic as this maximum diverges to infinity as $n \rightarrow \infty$ if the maximum is taken over the whole of $\{1, \dots, n\}$. This said, one can scale $\max_{1 \leq k \leq n} T_k^2 / \hat{\sigma}^2$ to an extreme value Gumbel law after subtracting $\ln(\ln(n))$ (see Darling & Erdős (1956) and Csörgő & Horváth (1997)); however, power from extreme value tests is comparatively poor. An alternative approach truncates the set of times which we allow as changepoints (we call this the admissible set K) at its boundaries. Specifically, one examines

$$T_{\max}^2 = \frac{\max_{\ell \leq \frac{k}{n} \leq h} T_k^2}{\hat{\sigma}^2}, \quad (2.3)$$

for some fixed ℓ and h satisfying $0 < \ell < h < 1$.

The asymptotic distributions for the CUSUM and adjusted LR statistics are quantified in Theorem 2.1. The first part is a consequence of the functional central limit theorem (see also MacNeill 1974). For a detailed proof that (2.5) follows from (2.4), see Robbins *et al.* (2009).

Theorem 2.1. *Assume that $\{X_1, \dots, X_n\}$ is an IID sequence of random variables with finite variance σ^2 . Then*

$$\text{CUSUM}_{\max} \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} |B(t)|, \quad (2.4)$$

and

$$T_{\max}^2 \xrightarrow{\mathcal{D}} \sup_{\ell \leq t \leq h} \frac{B^2(t)}{t(1-t)}, \quad (2.5)$$

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 $\mathbf{X}_j = \{X_{1,j}, \dots, X_{d,j}\}'$ and suppose that \mathbf{X}_j has uncorrelated components (independence is not needed). Let $\text{Var}(X_{i,j}) = \sigma_i^2$ be the variance of the i^{th} component. Define a CUSUM statistic for component i and time k via

$$\text{CUSUM}_{i,k} = \frac{1}{\sqrt{n}} \left(\sum_{j=1}^k X_{i,j} - \frac{k}{n} \sum_{j=1}^n X_{i,j} \right). \quad (2.6)$$

Joint convergence of quadratic forms of the components in (2.6) can be obtained from a multi-dimensional version of the functional central limit theorem generally attributed to

Donsker. For a recent paper containing general statements and proofs of such results, see Meerschaert & Sepanski (2002).

To quantify the multivariate version of (2.5), let $\{B_1(t)\}, \{B_2(t)\}, \dots, \{B_d(t)\}$ be d independent Brownian bridge processes and set

$$B^{(d)}(t) = \sum_{j=1}^d B_j(t)^2. \quad (2.7)$$

Using the multi-dimensional functional central limit theorem, under the null hypothesis that $\{\mathbf{X}_j\}_{j=1}^n$ is IID with uncorrelated components and $\text{Var}(X_{i,j}) = \sigma_i^2$, one can show that

$$\max_{\ell \leq \frac{k}{n} \leq h} \left\{ \frac{\sum_{i=1}^d \sigma_i^{-2} \text{CUSUM}_{i,k}^2}{\frac{k}{n} \left(1 - \frac{k}{n}\right)} \right\} \xrightarrow{\mathcal{D}} \sup_{\ell \leq t \leq h} \frac{B^{(d)}(t)}{t(1-t)}. \quad (2.8)$$

for each $0 < \ell < h < 1$. This result will be our major technical tool later. To obtain null hypothesis percentiles of such statistics, Csörgő & Horváth (1997) utilize a result of Vostrikova (1981) to show that

$$P \left\{ \sup_{\ell \leq t \leq h} \left(\frac{B^{(d)}(t)}{t(1-t)} \right) \geq x \right\} = \frac{x^{d/2} e^{-x/2}}{2^{d/2} \Gamma(d/2)} \left\{ \left(1 - \frac{d}{x}\right) \log \left(\frac{(1-\ell)h}{\ell(1-h)} \right) + \frac{4}{x} + O\left(\frac{1}{x^2}\right) \right\} \quad (2.9)$$

as $x \rightarrow \infty$. Here, $O(x^{-2})$ denotes a remainder term that goes to zero no slower than x^{-2} as $x \rightarrow \infty$. When the order term is disregarded, simulations show that (2.9) provides accurate tail probability approximations. For discussion on selection of ℓ and h , see Miller & Siegmund (1982), Andrews (1993), and Csörgő & Horváth (1997).

Two points are worth making before proceeding. First, while LR-type statistics such as T_{\max}^2 require some truncation of the set $\{1, \dots, n\}$, they often have significantly higher

power than CUSUM statistics when the changepoint occurs near 1 or n (the boundaries). One should not expect to detect a mean shift with good power when there are relatively few observations before or after the changepoint time. For a power comparison of CUSUM and Gaussian LR statistics, see Robbins *et al.* (2009).

Second, this work resides in the at most one changepoint (AMOC) domain. While multiple changepoints are frequently encountered in practice, we will be able to make conclusions here by simply subsegmenting the data once changepoints are found. Of course, this procedure is not without drawbacks. On the other hand, multiple changepoint methods are by no means unflawed. For instance, Jewson & 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.

3 The χ_{\max}^2 Test

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

3.1 Detecting Changes in Multinomial Data

Suppose we wish to assess whether or not changes have occurred in a multinomial 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 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 used to make conclusions. We call the procedure a χ_{\max}^2 test. In general, the values of χ_k^2 will be correlated in the time index k .

Maximums of χ^2 random variables have been previously used in the literature, primarily in biostatistics (Halpern 1982; Koziol 1991; Betensky & 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 $\mathcal{I}_1, \dots, \mathcal{I}_m$. 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 $t = 1, \dots, n$, $\mathbf{N}_t = \{N_{1,t}, \dots, N_{m,t}\}'$ is a multinomial observation with one trial and probability vector $\mathbf{p}(t) = \{p_1(t), \dots, p_m(t)\}$. 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) $p_i(t)$'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 $E[\widehat{O}_{i,k}] = k\hat{p}_i = kO_i/n$ and $E[\widehat{O}_{i,k}^*] = (n-k)\hat{p}_i = (n-k)O_i/n$, the χ^2 statistic for a change at time k is thus

$$\chi_k^2 = \sum_{i=1}^m \frac{(O_{i,k} - E[\widehat{O}_{i,k}])^2}{E[\widehat{O}_{i,k}]} + \sum_{i=1}^m \frac{(O_{i,k}^* - E[\widehat{O}_{i,k}^*])^2}{E[\widehat{O}_{i,k}^*]}. \quad (3.1)$$

Our major result is the following.

Theorem 3.1. *Let χ_k^2 be as in (3.1) and $\chi_{\max}^2 = \max_{\ell \leq \frac{k}{n} \leq h} \chi_k^2$. Then under a null hypothesis that $\mathbf{p}(t)$ does not change in t ,*

$$\chi_{\max}^2 \xrightarrow{\mathcal{D}} \sup_{\ell \leq t \leq h} \frac{B^{(m-1)}(t)}{t(1-t)},$$

where $B^{(m-1)}(t)$ is defined in (2.7). *P-values for this test are approximated using (2.9) with $d = m - 1$.*

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

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

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

$$\sum_{t=1}^k N_{i,t} - \frac{k}{n} \sum_{t=1}^n N_{i,t} = k \left(1 - \frac{k}{n}\right) (\hat{p}_{i,k} - \hat{p}_{i,k}^*), \quad (3.2)$$

we can reexpress χ_k^2 as

$$\chi_k^2 = \frac{k(n-k)}{n} \sum_{i=1}^m \frac{(\hat{p}_{i,k} - \hat{p}_{i,k}^*)^2}{\hat{p}_i}. \quad (3.3)$$

Now define X_k^2 by replacing \hat{p}_i by p_i in (3.3):

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

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 $\mathbf{P} = (p_1, p_2, \dots, p_{m-1})'$, $\hat{\mathbf{P}}_k = (\hat{p}_{1,k}, \dots, \hat{p}_{m-1,k})'$, and $\hat{\mathbf{P}}_k^* = (\hat{p}_{1,k}^*, \dots, \hat{p}_{m-1,k}^*)'$. Also, let $\mathbf{D} = \text{diag}(\mathbf{P})$. 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 $(N_{1,t}, \dots, N_{m-1,t})'$ and \mathbf{A}^{-1} exists and does not depend on n or k . For estimating the category m probabilities, we use

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

Using these facts, we see that

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

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 (3.2), we see that each component of \mathbf{Z}_k has the form

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

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

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

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

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

where CUSUM_j refers to a cumulative sum in $\{Y_{j,t}\}$. Using Slutsky's theorem and (2.8)

finishes our work. \square

Two remarks regarding Theorem 3.1 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. In our applications, the use of Theorem 3.1 is set in the latter circumstances; here, $n = 1410$ is the total number of observed cyclones.

Second, the χ_{\max}^2 test reinforces why the admissible set should be truncated away from the boundaries. Elaborating, the standard convention needed to apply Pearson's test (because of its asymptotic nature) requires that $E[O_{i,k}]$ and $E[O_{i,k}^*]$ all exceed unity and at least 80% of $E[O_{i,k}]$ and $E[O_{i,k}^*]$ are 5 or greater for each k . These conditions are clearly violated when k is sufficiently close to 1 or n .

3.2 Tests for Poisson Data

Tropical cyclone counts are frequently modeled with Poissonian dynamics (Mooley 1981; Thompson & Guttorp 1986; Solow 1989; Lund 1994). While one can add batch arrivals, periodic features, and meteorological covariates 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. 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), $\{X_t\}$ should be IID with marginal Poisson distributions. To test a null hypothesis of a constant Poisson 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 (2.3):

$$D_{\max} = \max_{\ell \leq \frac{k}{n} \leq h} D_k = \max_{\ell \leq \frac{k}{n} \leq h} \frac{\text{CUSUM}_k^2}{\hat{\lambda} \frac{k}{n} \left(1 - \frac{k}{n}\right)}, \quad (3.4)$$

where

$$D_k = \frac{\text{CUSUM}_k^2}{\hat{\lambda} \frac{k}{n} \left(1 - \frac{k}{n}\right)}.$$

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 $E[\widehat{C}_k] = k\hat{\lambda} = kC_n/n$ and $E[\widehat{C}_k^*] = (n - k)\hat{\lambda} = (n - k)C_n/n$, one can algebraically show that

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

Note that D_k itself has the classic chi-square form in the summands: observed minus expected squared over expected. Theorem 3.1 now gives the following result.

Theorem 3.2. *If X_1, \dots, X_n is IID and Poisson, then*

$$D_{\max} \xrightarrow{\mathcal{D}} \sup_{\ell \leq t \leq h} \frac{B^2(t)}{t(1-t)}, \quad (3.5)$$

and therefore (2.9) with $d = 1$ can be used to find p -values of this test.

Many covariates are available for our cyclone data. For instance, each cyclone is accompanied with an estimate of the maximum wind speed achieved by the storm. Suppose

that the j^{th} arriving storm has covariate Y_j and that this covariate is independent of all storm arrival process dynamics. It is of interest to test whether changes are occurring in the cyclone counts, their meteorological covariates, 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 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 most severe storm. For example, a Saffir-Simpson category 1 storm is a tropical cyclone with peak wind speeds from 74 - 95 mph, inclusive. 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 estimating the categorical wind speed of the storm may be more accurate than estimating the actual wind speed. A categorical approach to handle the wind speed covariate seems reasonable and will be our approach.

Suppose that Y_j is the covariate (wind speed here) for the j^{th} storm (the storms are time ordered) and that Y_j lies in one of m disjoint categories (the categories need not be Saffir-Simpson categories). Consider the m -dimensional vector $\mathbf{X}_t = \{X_{1,t}, X_{2,t}, \dots, X_{m,t}\}'$. Here, $X_{i,t}$ is the total number of category i storms during the t^{th} year. For example, in the next section, $X_{1,t}$ will be the number of tropical storms (the storm never achieved

hurricane status) occurring during the t^{th} year. Of course, $\sum_{i=1}^m X_{i,t}$ 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 $i = 1, 2, \dots, m$ and $t = 1, 2, \dots, n$.

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, $E[\widehat{C}_{i,k}] = k\widehat{\lambda p_i} = kC_{i,n}/n$ and $E[\widehat{C}_{i,k}^*] = (n-k)\widehat{\lambda p_i} = (n-k)C_{i,n}/n$. The χ^2 statistic for testing for a changepoint at time k is

$$\chi_k^2 = \sum_{i=1}^m \frac{(C_{i,k} - E[\widehat{C}_{i,k}])^2}{E[\widehat{C}_{i,k}]} + \sum_{i=1}^m \frac{(C_{i,k}^* - E[\widehat{C}_{i,k}^*])^2}{E[\widehat{C}_{i,k}^*]}. \quad (3.6)$$

Computations as before show that

$$\chi_k^2 = \frac{n}{k(n-k)} \sum_{i=1}^m \frac{1}{\widehat{\lambda p_i}} \left(C_{i,k} - \frac{k}{n} C_{i,n} \right)^2 = \frac{1}{\frac{k}{n} (1 - \frac{k}{n})} \sum_{i=1}^m \frac{1}{\widehat{\lambda p_i}} (\text{CUSUM}_{i,k})^2, \quad (3.7)$$

where $\text{CUSUM}_{i,k}$ is as in (2.6). 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 . Furthermore, $\text{Cov}(X_{i,t}, X_{i',t}) = 0$ when

$i \neq i'$; therefore, $X_{i,t}$ and $X_{i',t}$ are uncorrelated. Applying (2.8) gives the following theorem.

Theorem 3.3. *Under the above setup, if $\lambda(t)p_i(t)$ is constant in t ,*

$$\chi_{\max}^2 = \max_{\ell \leq \frac{k}{n} \leq h} \chi_k^2 \xrightarrow{\mathcal{D}} \sup_{\ell \leq t \leq h} \frac{B^{(m)}(t)}{t(1-t)}, \quad (3.8)$$

and therefore (2.9) with $d = m$ can be used to find p -values of this test.

4 North Atlantic Basin Cyclones

This section uses the developed categorical changepoint methods to study the North Atlantic Basin tropical cyclone record. We use the HURDAT dataset available on the National Oceanic Atmospheric Administration’s website. This data contain information on 1410 Atlantic basin cyclones that attained tropical storm-level intensity between 1851 and 2008. The data have been reanalyzed several times (Landsea *et al.* 2004, Landsea *et al.* 2008) and are believed to contain inconsistencies due to ongoing advances in measurement techniques. For instance, counts of landfalling cyclones before 1900 are considered unreliable (Landsea *et al.* 1999 and Jarrell *et al.* 1992) due to sparse population along coastlines. Also, as Landsea *et al.* (1999) and Neumann *et al.* (1999) observe, aircraft reconnaissance towards the end of World War II (around 1944) improved detection of non-landfalling storms. Satellites were fully implemented for cyclone surveying in the mid 1960’s, and their introduction also likely improved abilities to accurately measure wind speeds. Landsea, Harper, Hoaran, & Knaff (2006) state that measurement techniques are continually improving and data on wind speeds as recently as the 1980’s 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 climate change.

We begin our work by jointly testing for changes in the yearly cyclone (which are assumed to have a Poisson distribution) and/or their categorical wind speeds. A natural partition of wind speeds involves the Saffir-Simpson Scale. Specifically, we partition wind speeds into five classes as follows: the first category represents storms whose peak wind speed never reached hurricane status (40-73 mph), the second corresponds to a category 1 hurricanes (74-95 mph), the third is for category 2 hurricanes (96 - 110 mph), the fourth is for category 3 hurricanes (111-130 mph) and the fifth corresponds to a category 4 or 5 hurricane (131 or greater mph). We combine category 4 and 5 storms because there are so few category 5 storms.

Applying Theorem 3.3 with $m = 5$ to the entire data sequence gives $\chi_{\max}^2 = 109.182$ with a p -value that is bounded above by 10^{-5} . The estimated changepoint time is $\hat{c} = 80$ (1930). Because the pre-1900 data is somewhat unreliable, we reran this analysis with only the 1900-2008 data. This analysis gives $\chi_{\max}^2 = 61.567$ and a p -value that is bounded by 10^{-5} . In this case, $\hat{c} = 95$ (1994). Figure 1 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 or wind speeds, we now examine these two variables separately.

First, we look at the yearly storm counts. Figure 2 plots the yearly storm counts from 1851-2008. An increase is visually evident circa 1995. Lund (1994) examined the annual

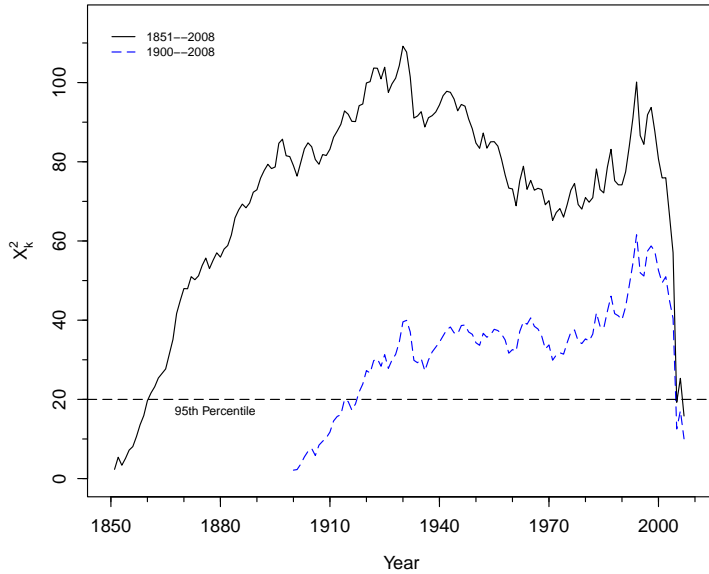


Figure 1: Chi-square Statistics for Joint Changepoints of Yearly Storm Counts and Wind Speeds

storm counts from 1871-1990 and found a change in frequencies circa 1931. To confirm this, we applied Theorem 3.2 to the annual storm counts from 1871-1990. Here, $n = 120$ and a test statistic of $D_{\max} = 20.015$ with a p -value of approximately 0.00047 were obtained; the estimated changepoint time is $\hat{c} = 60$ (1930). Repeating this test for the storm counts in the 1931-2008 segment only 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 3.2 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 3.

The CUSUM test in Theorem 2.1, which does not require any Poissonian assumptions,

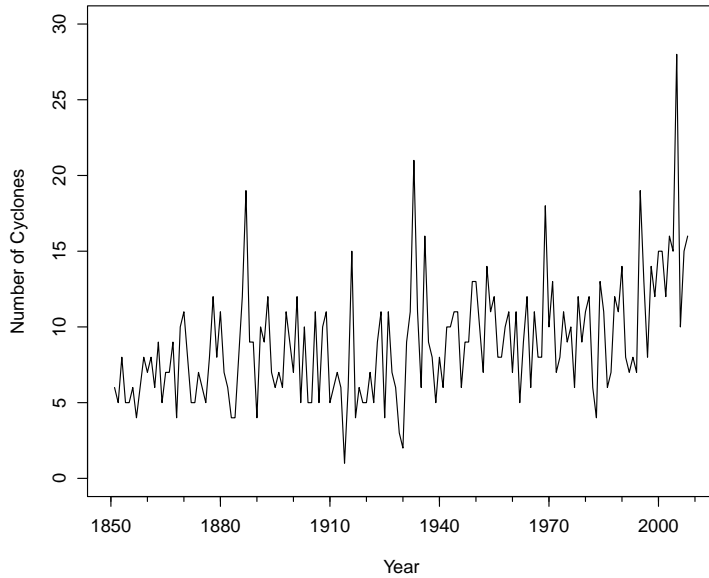


Figure 2: Annual Cyclone Counts from 1851-2008

can also be used to test for changes in the yearly storm counts. Figure 4 shows the CUSUM statistics for the three time segments in the last paragraph. For the 1871-1990 segment, $\text{CUSUM}_{\max} = 1.930$ with a p -value of 0.00116 and $\hat{c} = 60$ (1930). For the 1931-2008 segment, $\text{CUSUM}_{\max} = 1.703$ with a p -value of 0.00606 and $\hat{c} = 64$ (1994). For the 1851-2008 data, $\text{CUSUM}_{\max} = 2.719$ 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 3.2 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 2.1 identifies 1930 as the most significant changepoint whereas the Theorem 3.2 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

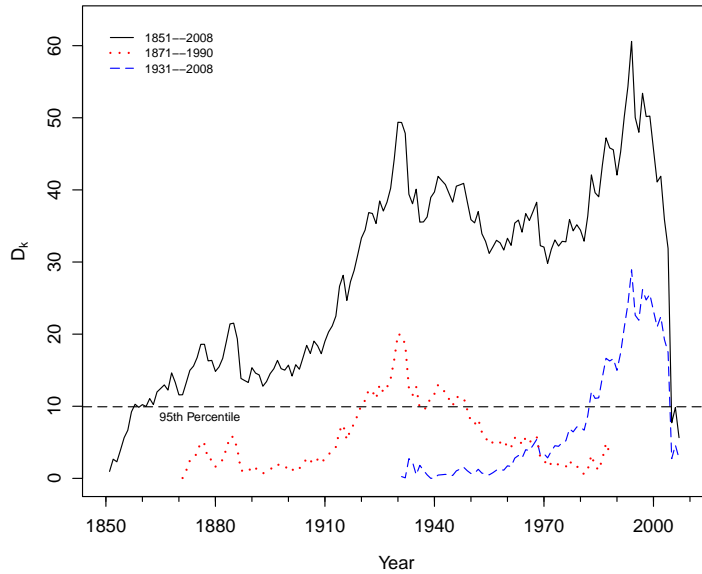


Figure 3: Chi-square Statistics for Changepoints of Yearly Storm Counts

simply illustrates the difficulty that CUSUM methods have at detecting changepoints that occur near the boundaries.

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 first ordered via their time of arrival. 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.

The two graphics in Figure 5 plot the storm number k versus $|\text{CUSUM}_k|/\hat{\sigma}$ 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 5 shows a 95th percentile confidence threshold. The vertical lines in the bottom plot depict the boundary truncations of $\ell = 1 - h = 0.05$. The CUSUM test shows that $\text{CUSUM}_{\max} = 0.960$ with a p -value of

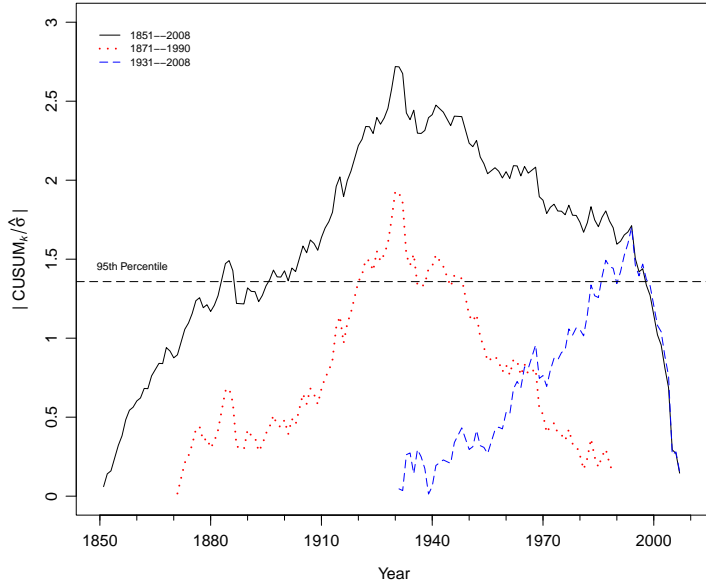


Figure 4: CUSUM Statistics for Changepoints of Yearly Storm Counts

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 χ_{\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 3.1 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 6. 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 that the pre-1900 data has not been reanalyzed or heavily quality checked. But one may ask why this changepoint was undetected by generic CUSUM methods but flagged with strong significance by the

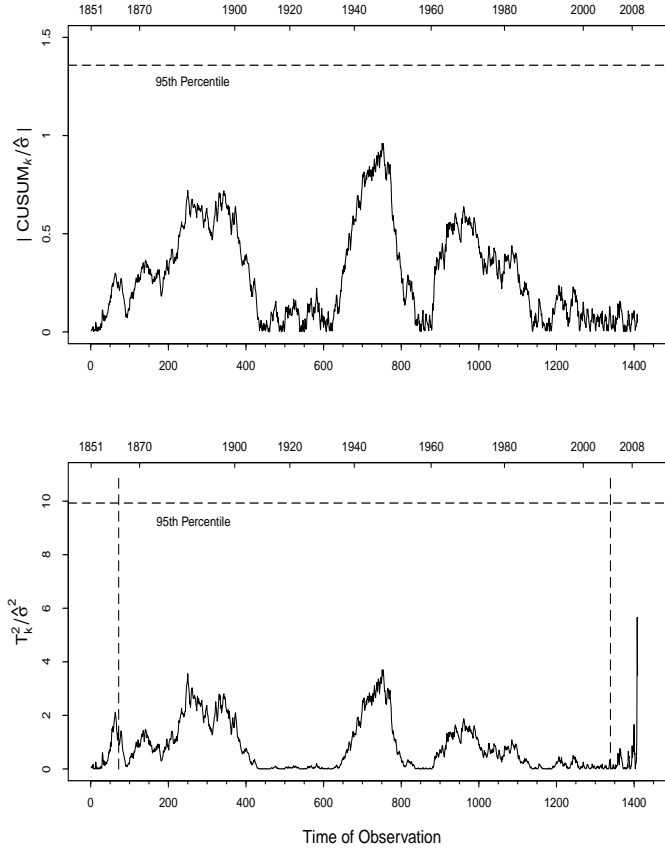


Figure 5: $|\text{CUSUM}_k|/\hat{\sigma}$ (top) and $T_k^2/\hat{\sigma}^2$ (bottom) Statistics for Peak Wind Speed in the 1851-2008 Data ($n = 1410$).

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

Segmenting the data to the 1040 storms arriving in 1900-2008, a changepoint in 1956 ($\hat{c} = 471$) is found; here, $\chi_{\max}^2 = 21.038$ and has a p -value of 0.01482. Hence, there appears

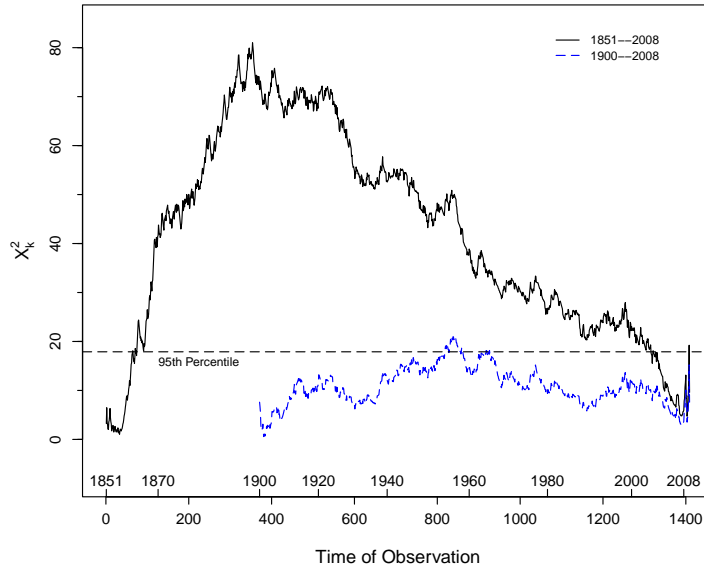


Figure 6: Chi-square statistics for Peak Wind Speeds in Atlantic Cyclones from 1851-2008 ($n = 1410$) and 1900-2008 ($n = 1040$).

Table 1: Categorical probabilities before and after the changepoint for the χ_{\max}^2 test.

Location	Tropical storm	Category 1 hurricane	Category 2 hurricane	Category 3 hurricane	Category 4-5 hurricane
Before \hat{c}	0.280	0.302	0.260	0.130	0.028
After \hat{c}	0.437	0.214	0.120	0.112	0.116

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 6 provides graphical support.

We conclude this analysis with some comments. The tests were applied to several other cyclone covariates for the years 1900-2008 including storm duration (in days), longitude

(and latitude) at which each storm's peak wind speed was first reached, and the probability of making landfall in the continental United States. Each covariate test positively for a changepoint circa 1960. As Goldenberg, Landsea, Mestas-Nuñez, & Gray (2001) and Neumann *et al.* (1999) note, tropical cyclone data after 1965, which is the year in which satellite surveillance was first in full use, is generally considered reliable. This belief is also supported in that our tests do not find any changepoints in any covariate when only data from the years 1965-2008 are examined.

The storm counts are a different matter. When the annual cyclone counts from the 44 year period from 1965-2008 are tested for a changepoint, we find $D_{\max} = 25.164$ with a p -value of less than 10^{-5} . Here, the changepoint is flagged at $\hat{c} = 30$ (1995). Figure 7 provides graphical support of this conclusion.

One can ask whether it is valid 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 3.2 and (2.9) 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 sufficiently 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. 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*

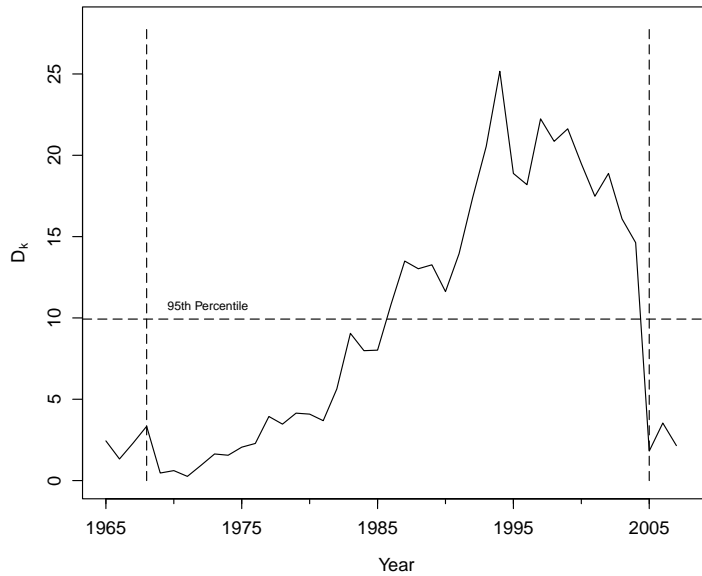


Figure 7: Chi-square Statistics for Peak Wind Speeds in Atlantic Cyclones from 1965-2008 ($n = 44$).

al. (2009) demonstrate that asymptotic changepoint tests similar to the ones used in this paper tend to be conservative.

Table 2: Simulated Type I error rates for 0.05-level test based on Theorem 3.2

Sample size	# rejections/10000
$n = 1000$	0.0433
$n = 158$	0.0345
$n = 44$	0.0234

5 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 tests miss. Contrary to some theories, we find no evidence of significant recent increases in storm strength or US 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 satellite surveillance. 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.* 2009) and/or climate change.

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

References

- Andrews, D. W. K. (1993), “Tests for Parameter Instability and Structural Change with Unknown Change Point,” *Econometrica*, 61, 821-856.
- 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.
- Betensky, R. A., and Rabinowitz, D. (1999), “Maximally Selected χ^2 Statistics for $k \times 2$ Tables,” *Biometrics*, 55, 317-320.

- Burt, C. C. (2004), *Extreme Weather*, W. W. Norton & Company, London.
- 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.
- Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992), "Hierarchical Bayesian Analysis of Changepoint Problems," *Applied Statistics*, 41, 389-405.
- Carlstein, E. (1988), "Nonparametric Change-Point Estimation," *The Annals of Statistics*, 16, 188-197.
- Csörgő, M., and Horváth, L. (1987), "Nonparametric Tests for the Change Point Problem", *Journal of Statistical Planning and Inference*, 17, 1-9.
- Csörgő, M., and Horváth, L. (1997), *Limit Theorems in Changepoint Analysis*, Chichester: John Wiley and Sons.
- 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.
- 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.
- Emanuel, K. A. (1987), "The Dependence of Hurricane Intensity on Climate," *Nature*, 326, 483-485.
- Emanuel, K. A. (2005), "Increasing Destructiveness of Tropical Cyclones Over the Past 30 Years," *Nature*, 436, 686-688.

Emanuel, 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.

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

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.

Goldenberg, S. B., Landsea, C. W., Mestas-Nuñez, A. M., and Gray, W. M. (2001), "The Recent Increase in Atlantic Hurricane Activity: Causes and Implications," *Science*, 293, 474-479.

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

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

Hinkley, D. V., and Hinkley, E. A. (1970), "Inference about the Change-Point in a Sequence of Binomial Variables," *Biometrika*, 57, 477-478.

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

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*, Cam-

bridge: Cambridge University Press.

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, Florida, 152 pp.

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

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

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

Landsea, C. W. (2007), "Counting Atlantic Tropical Cyclones Back to 1900," *EOS, Transactions, American Geophysical Union*, 88, 197-208.

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*, edited by R. J. Murname and K.-B. Liu, pp. 177-221, New York: Columbia University Press.

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.

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.

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.

Landsea, C. W., Vecchi, G. A., Bengtsson, L., and Knutson, T. R. (2009) "Impact of Duration Thresholds on Atlantic Tropical Cyclone Counts," In Press, *Journal of Climate*.

Lund, R. B. (1994), "Annual Arrival Cycle of Atlantic Tropical Cyclones," *Journal of Applied Statistics*, 21, 195-204.

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

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

Mooley, D. A. (1981), "Applicability of the Poisson Probability Model to the Severe Cyclonic Storms Striking the Coast around the Bay of Bengal," *Sankhyā Series B*, 43, 187-197.

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 Mathématique*, 71, 55-62.

Neumann, C. J., Jarvinen, B. R., McAdie, C. J., and Elms, J. D. (1999), *Tropical Cyclones of the North Atlantic Ocean, 1871-1998*. National Climatic Data Center in cooperation

with the National Hurricane Center, Asheville, NC, 206 pp.

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41, 100-105.

Page, E. S. (1955), "A Test for a Change in a Parameter Occurring at an Unknown Point," *Biometrika*, 42, 523-527.

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

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.

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.

Quandt, R. E. (1958), "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53, 873-880.

Quandt, R. E. (1960), "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324-330.

Robbins, M. W., Gallagher, C. M., Lund, R. B., and Aue, A. (2009), "Mean Shift Testing with Correlated Data," Pre-Print.

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.
- Smith, A. F. M. (1975), “A Bayesian Approach to Inference about a Change-Point in a Sequence of Random Variables”, *Biometrika*, 62, 407-416.
- Solow, A. R. (1989), “Reconstructing a Partially Observed Record of Tropical Cyclone Counts”, *Journal of Climate*, 2, 1253-1257.
- 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.
- Vecchi, G. A., and Knutson, T. R. (2008), “On Estimates of Historical North Atlantic Tropical Cyclone Activity,” *Journal of Climate*, 21, 3580-3600.
- Wolfe, D. A., and Chen, Y.-S. (1990), “The Changepoint Problem in a Multinomial Sequence”, *Communications in Statistics – Simulation and Computation*, 19, 603-618.
- Yao, Y.-C., and Davis, R. A. (1984), “The Asymptotic Behavior of the Likelihood Ratio Statistic for Testing a Shift in Mean in a Sequence of Independent Normal Variates,” *Sankhyā Series A*, 48, 339-353.
- Vostrikova, L. Y. (1981), “Detection of Disorder in a Wiener Process,” *Theory of Probability and its Applications*, 26, 356-362.