Impact of Beliefs About Atlantic Tropical Cyclone Detection on Conclusions About Trends in Tropical Cyclone Numbers

Article · December 2011

DOI: 10.1214/11-BA621

5 authors, including:

Surya Tokdar
Duke University

Iris Grossmann
Chatham University

Mitchell J. Small
Carnegie Mellon University

Bayesian Analysis (2011) **6**, Number 4, pp. 547–572

Impact of Beliefs About Atlantic Tropical Cyclone Detection on Conclusions About Trends in Tropical Cyclone Numbers

Surya T. Tokdar*, Iris Grossmann†, Joseph B. Kadane‡,
Anne-Sophie Charest§, and Mitchell J. Small¶

Abstract. Whether the number of tropical cyclones (TCs) has increased in the last 150 years has become a matter of intense debate. We investigate the effects of beliefs about TC detection capacities in the North Atlantic on trends in TC numbers since the 1870s. While raw data show an increasing trend of TC counts, the capability to detect TCs and to determine intensities and changes in intensity has also increased dramatically over the same period. We present a model of TC activity that allows investigating the relationship between what one believes about the increase in detection and what one believes about TC trends. Previous work has used assumptions on TC tracks, detection capacities or the relationship between TC activity and various climate parameters to provide estimates of year-by-year missed TCs. These estimates and the associated conclusions about trends cover a wide range of possibilities. We build on previous work to investigate the sensitivity of these conclusions to the assumed priors about detection. Our analysis shows that any inference on TC count trends is strongly sensitive to one's specification of prior beliefs about TC detection. Overall, we regard the evidence on the trend in North Atlantic TC numbers to be ambiguous.

Keywords: Atlantic tropical cyclones, HURDAT, tropical cyclone data, tropical cyclone detection.

*Department of Statistical Science, Duke University, Durham, NC, st118@stat.duke.edu
†Climate Decision Making Center, Carnegie Mellon University, Pittsburgh, PA, irisg@andrew.cmu.edu
‡Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, kadane@stat.cmu.edu
§Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, acharest@stat.cmu.edu
¶Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, ms35@andrew.cmu.edu

 DOI:10.1214/11-BA621

1 Introduction

Whether anthropogenic global warming may be impacting tropical cyclones (TCs) has become a matter of intense debate (Knutson et al. 2010; Grossmann and Morgan 2011). Raw data from the North Atlantic show an increasing trend of annual TC counts in the region (Vecchi and Knutson 2008; Holland and Webster 2007). A similar observation holds for the number of high-intensity TCs (Webster et al. 2005). The difficulty of interpretation is that the capability to detect TCs and to determine intensities and change in intensity has also increased dramatically over the same period (e.g. Landsea et al. 2004, 2006; Landsea 2007). Consequently there is a relationship between what one believes about the increase in detection capability and what one believes about trends in TC activity. This paper investigates the effects of beliefs about TC detection capacities in the North Atlantic on trends in Atlantic TC numbers.

The inference on the trend of tropical cyclone counts shares common elements with other statistical problems in trend analysis with missing observations, particularly where there is systematic bias in observations over a portion of the temporal or spatial domain (Coles and Sparks 2006; Cornulier et al. 2011; Kéry and Royle 2010). An effective approach for dealing with nonrandom missing data is to include a representation for the observation process (or the missing data mechanism) as part of the overall statistical model (Little 1995; Zeger and Liang 1996; Ibrahim et al. 2001; Tingley and Huybers 2010). Missing data are then imputed as part of the parameter estimation process, resulting in a joint distribution for the model parameters and the missing observations (Rubin 1996; Allison 2001; Honaker and King 2010).

The dataset used for studies on North Atlantic TC trends is the "best track" dataset of the National Hurricane Center (HURDAT) (Jarvinen et al. 1984). HURDAT includes observed positions, maximum wind speeds, and some central pressure measurements for TCs dating back to 1851. The North Atlantic is widely regarded as having the most reliable TC data and the longest TC time series. Trend detection in all basins, in particular the North Atlantic, is complicated by natural variability on several time scales, including the multidecadal (Klotzbach and Gray 2008; Vecchi and Knutson 2008). Consequently, long historical coverage is essential for trend analysis. However, Atlantic TC records from the earlier period prior to the availability of reconnaissance aircraft and satellites rely on sparsely populated coasts and limited ship tracks (Landsea et al. 2004, 2006), with the consequence that some TCs likely did not get recorded. Thus, an increasing trend of the observed counts might be attributed entirely to improvements in detection technology (Landsea 2007).

The extent of data quality issues in HURDAT has been the subject of intense discussions (Landsea 2007; Holland and Webster 2007; Vecchi and Knutson 2008; Chang and Guo 2007; Mann et al. 2007; Bengtsson and Hodges 2008; Landsea et al. 2010), with some studies suggesting that the Atlantic TC record can be regarded as reasonably reliable back into the late 19th century because ships could not be warned off from approaching TCs (Mann and Emanuel 2006; Holland and Webster 2007). A number of recent studies aim at interpreting the HURDAT records by augmenting it with estimates of year-by-year missed TC counts. These studies can be categorized into three groups based on the principles they employ to estimate missed TCs. One group matches satellite-era TC tracks with earlier ship tracks and land points (Chang and Guo 2007; Vecchi and Knutson 2008). A second group analyzes time trends of the proportion of TCs possessing certain characteristics, such as TCs making landfall, which can be argued to have had enjoyed good detection even in the earlier times (Solow and Moore 2000, 2002; Landsea 2007; Nyberg et al. 2007; Elsner and Bossak 2006). A third group predicts TC counts by modeling their relationship to other climate variables with more accurate historical records (Mann et al. 2007; Solow and Beet 2008).

Building on these studies, our analysis investigates the sensitivity of conclusions about trends in TC numbers to the assumed priors about detection. Our goal is to encourage climate researchers to use the platform we develop, possibly in conjunction with other trend models, and to draw their attention to the extremely important issue of carefully quantifying one's beliefs about detection probabilities. To illustrate this latter point, we first develop a belief quantification that produces estimates of missed TC counts that match the numbers reported in Vecchi and Knutson (2008), thus recapturing their conclusion of an increasing trend of yearly TC counts since the 1870s. We then show that seemingly minor changes to this belief quantification result in either roughly constant or negative trends. This sensitivity of the inference to the prior input is not a negative feature of our approach, rather a simple reminder of the inherent ambiguity of the HURDAT records caused by missing observations. Our analysis shows that any inference on TC count trend is strongly sensitive to one's specification of prior beliefs about TC detection.

In Section 2, we begin by reviewing changes in TC detection methods and technologies over time. We then briefly discuss the assumptions that had to be made by different previously published approaches for the estimation of missed Atlantic TCs. Section 3 explains the data and methods. Our results in Sections 4 and 5 show the strong sensitivity to the assumptions made on TC detection capacities and highlight

the resulting ambiguity in the trend of TC numbers. In Section 6 we present possible ways to resolve this ambiguity.

2 TC detection and recording over the years

Tropical cyclones are warm-core low pressure systems with a closed circulation over tropical or subtropical oceans. The weakest form of a TC is called tropical depression. TCs with wind speeds of 35 knots or more are called tropical storms, TCs with speeds of 65 knots or more are called hurricanes; the latter range from the less intense category 1 to the very intense category 5 (Simpson 1974). Trend analysis of TC numbers considers "named storms", that is, TCs of tropical storm strength or greater. Two criteria determine whether wind speed observations of tropical storm strength or greater are recorded as tropical cyclones within HURDAT (Landsea et al. 2008). First, evidence of a closed circulation and the non-frontal character of the system are required to distinguish the cyclone from an extratropical or subtropical cyclone. Second, at least two wind speed measurements or estimates by independent observers are required (Landsea et al. 2008). Thus, the reasons leading to a TC not being recorded in HURDAT are twofold: first, an actual lack of observations of the TC in question, and second, a lack of information to classify the observed wind anomaly as a tropical cyclone. Figure 1 gives a timeline of how TC observation technology has changed over the years, along with the occurrences of the major events that may have impacted our ability to record TCs.

The US Signal Service has been observing Atlantic TCs since approximately 1873 (Sheets 1990; Fernández-Partagás and Diaz 1996). Several forecast offices were established in the 1930s, followed by the designation of the Miami forecast office as National Hurricane Center in 1955 (Sheets 1990). Satellite observations have been available since 1967, aircraft reconnaissance since 1945. Prior to 1945, TC records relied entirely on ship and coastal observations. In parts of the US, insufficient coastal density and limited reporting in the early part of the century may have led to a failure to detect landfalling TCs that had not been reported by ships (Landsea 2007; Landsea et al. 2008; Sheets 1990, Figure 1). This is illustrated by the recent addition of four new landfalling storms to the 10-year period from 1911-20 in the course of the ongoing reanalysis of the HURDAT records (Landsea et al. 2008).

Ships may not have sighted and reported all TCs that occurred due to insufficient coverage with regular ship routes, insufficient observation equipment, and possibly conscious avoidance of an approaching storm. Until radio became available in 1905, TC

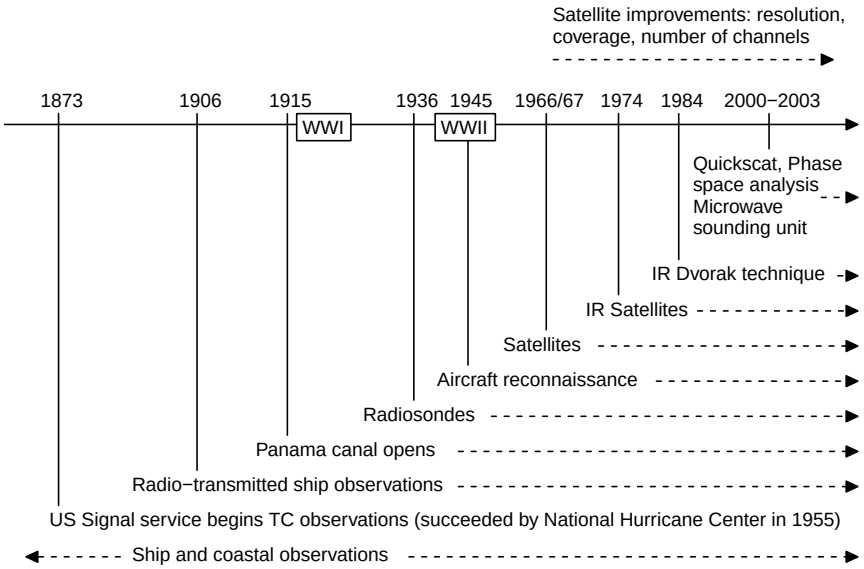

Figure 1: A timeline of Atlantic TC observation cataloging main changes in observation technology along with other major events that may have impacted TC recording.

observations relied on ship reports after the ship returned to port. TC observations were evaluated with the Beaufort scale, which specifies differences in waves and the ocean state up to category 1 hurricanes, and with marine barometers if available (Landsea et al. 2004). In 1900, approximately one quarter of ships were equipped with marine barometers; by 1930, most had barometers (C. W. Landsea, personal communication, 2007). Prior to the opening of the Panama Canal in 1915, ship routes were concentrated in the northern and eastern parts of the basin and near the US East coast. The opening of the Panama Canal significantly increased the likelihood of observation of TCs south of 32°N (Vecchi and Knutson 2008). Methods to detect approaching TCs prior to recording gale force winds were available although it is unclear to what extent mariners made use of these methods to avoid contact with approaching TCs (Bowditch 1841; Piddington 1860; Bowditch 1995).

Both limited coverage by ship tracks and the possible conscious avoidance of TCs may have resulted in missed TCs. Evidence of a closed circulation usually relied on weather maps or multiple observations of the same TC. With the limited coverage and limited quality of observations, such evidence was not always available, likely resulting in the omission of several TCs from HURDAT during this era. Aircraft reconnaissance has been used sporadically since 1944, and more regularly since 1956 after the devastation caused by several New England hurricanes (Dorst 2007). Until the 1960s, reconnaissance flights were typically dispatched to investigate TCs that had already been detected; in addition, regular patrols covered the route from Bermuda to east of St. Croix, St. Croix to Miami and back to Bermuda (Dunn and Miller 1960). Flights generally did not travel beyond 55°W. Dunn and Miller (1960, page 155), report that about half of all TCs were initially detected by ships until the 1960s. While some TCs were probably first detected by aircraft, this leaves a large number of TCs to be initially detected by islands or coastal areas, implying that at least some of those TCs that remained at a reasonable distance from islands and coasts were likely missed altogether.

During the early satellite era beginning in 1967, TC observations relied on a combination of visible satellite imagery and aircraft reconnaissance (Neumann et al. 1999). Nighttime observations became possible only with the launch of infrared (IR) satellites in 1974 and the adoption of a Dvorak-scheme for the interpretation of IR imagery in 1984 (Dvorak 1984). Further significant changes were gradual improvements in coverage and resolution (Sheets 1990; Landsea et al. 2006), and the addition of the Advanced Microwave Sounding unit (Brueske and Velden 2003), the Quick Scatterometer, or "QuikSCAT" (Atlas et al. 2001) and the Cyclone Phase Space analysis tool (Harper

and Callaghan 2006), during the years 2000 to 2003 (Landsea 2007).

Two types of systems might be underrepresented to some extent prior to these more recent improvements. First, TCs with very short lifetimes could have been missed (Landsea et al. 2010), in particular due to a lack of nighttime IR imagery and due to viewing gaps. Knapp and Kossin (2007) find that in the 1980s, satellite observations were not available during 5.5% of all 6-hour periods during which a TC was present in the North Atlantic; during the 1990s and 2000s this was reduced to 1.6% and 0.5%, respectively. Second, storms may have been detected but not classified as TCs as they exhibited tropical characteristics or tropical storm strength for only a short period of time. The capacity to classify storms that were tropical only for short time periods has also improved dramatically over time; in fact Landsea (2007) suggests that capacities may have become adequate only in recent years with the help of new tools. The original Dvorak developmental sample (Velden et al. 2006) did not include subtropical systems – storms that exhibit both tropical and extratropical characteristics and that might or might not become fully tropical at some stage of their lives.

Approaches that estimate missed TCs necessarily rely on assumptions about detection capacities, and in some cases, aspects of TC activity and the presence or absence of changes in TC activity over time. Matching ship tracks with satellite era TC tracks requires the assumption that the spatial distribution of satellite era TC tracks is similar to the distribution during earlier time periods (Vecchi and Knutson 2008). However, tracks may not be static in time if they respond to shifts in atmosphere-ocean conditions such as the recent warming of the eastern Atlantic (Holland 2007).

A second type of assumption made by approaches using ship tracks is that the number of recorded TCs during the satellite era matches the number of TCs that actually occurred. This assumption is problematic given the insufficiency of observational capacities for the correct classification of short-lived or weak tropical storms until very recently (Landsea 2007; Landsea et al. 2010). Models using climate variables are affected by a similar problem, as the assumed relationship between TCs and climatic variables has to rely on data over a certain time period (or time periods) to estimate the parameters in the model. However, available observation technologies and interpretation schemes continued to undergo significant changes and improvements during the satellite era, with likely effects on the completeness of TC records (Landsea 2007; Landsea et al. 2010). Hence even the satellite era records are likely not to be complete.

A third type of assumption made by approaches that consider ship tracks concerns observation and detection capacities and practices. Available studies assume land points

to be perfect storm detectors. This means, first, that coasts were populated at sufficient density to detect TCs, which may not have been the case in the first two or three decades of the century (Sheets 1990; Landsea et al. 2006). Second, it is assumed that land points everywhere were sufficiently equipped to observe and correctly classify TCs (Vecchi and Knutson 2008; Chang and Guo 2007). It is further assumed that ships did not alter their course to avoid encounters with TCs before tropical storm force winds and in particular evidence for a closed circulation could be reported (Vecchi and Knutson 2008). If TCs were avoided, this should have caused the record of observed TCs to be even more incomplete. It is also assumed that observation by one (Chang and Guo 2007) or two (Vecchi and Knutson 2008) observers always led to a storm being recorded. The additional requirement that evidence of the storms' tropical characteristics be available has not been incorporated into available studies (Vecchi and Knutson 2008).

3 Combining detection probability with raw counts

Let n_i denote the raw TC counts in the HURDAT records corresponding to calendar year y_i, $i = 1, 2, \cdots, N$. We cover the period from $y_1 = 1871$ through $y_N = 2008$ with $N = 138$. Let m_i denote the missed TC counts for these years, and $t_i = n_i + m_i$ the actual total TC counts. Our aim is to infer about the trend of t_i's. We achieve this within a Poisson trend model (Solow 1989; Elsner and Kara 1999):

$$t_i \sim \mathsf{Po}(\lambda_i), \quad t_i\text{'s are independent across } i, \tag{1}$$

$$\log\lambda_i = \eta_1 + \eta_2 \frac{y_i - \bar{y}}{N-1} + \phi_1 \cos\left(\pi\left\{\phi_2 + \phi_3 \frac{y_i - y_1}{y_N - y_1}\right\}\right) \tag{2}$$

where $\bar{y} = \frac{1}{N}\sum_i y_i = 1939.5$ denotes the "central year" for the period under study. In (2), $\eta_1 + \eta_2 \frac{y_i - \bar{y}}{N-1}$ reflects a linear trend (in log scale) of the total TC counts: e^{η_1} gives the mean TC count in the central year and $e^{\eta_2} - 1$ gives the relative change in mean counts over the N years with respect to year 1. The important parameter here is the slope η_2. A positive, zero or negative value of η_2 indicates an increasing, flat or decreasing trend of the total TC counts. The cosine term in (2) is introduced to allow for multidecadal oscillation in TC frequency (Klotzbach and Gray 2008; Elsner and Jagger 2006; Goldenberg et al. 2001; Vecchi and Knutson 2008; Knutson et al. 2007). Here, ϕ_1 encodes the amplitude of this oscillation; ϕ_2 encodes a phase shift, $\phi_2 = 0$ makes year 1 coincide with a peak (positive or negative depending on the sign of ϕ_1) of this oscillation; ϕ_3 encodes the number of oscillation cycles within the period of study.

For the j-th TC in the i-th year, let the detection probability be π_{ij}, $j = 1, 2, \cdots, t_i$,

$i = 1, 2, \cdots, N$. We let π_{ij} depend both on the detection technology available in year i and a scalar z_{ij} denoting some measure of strength of the TC being detected. The latter is included to reflect the notion that for any given detection method, a stronger TC was more likely to be detected than a weaker one. This measure of strength can be defined in various ways and a vector of multiple measures could also be considered. For our illustration, we consider a simple measure where z_{ij} denotes the time (in hours) the corresponding TC had wind speed in category 1 or higher. We specify

$$\pi_{ij} = \Phi(\gamma_i + \beta z_{ij}) \tag{3}$$

where the scalar parameter γ_i gives the contribution of the i-th year's detection technology, and the scalar parameter β gives the influence of TC strength on its detectability; $\Phi(x) = \int_{-\infty}^{x} (\sqrt{2\pi})^{-1} \exp(-z^2/2) dz$ denotes the cumulative distribution function of the standard normal distribution. In principle, one can replace β with a year dependent β_i, but we shall stick to a constant β to maintain simplicity.

Note that z_{ij} is unobserved for every TC that went undetected. This requires a probability distribution to describe what these missing measurements could have been, with parameters underlying the distribution that can be learned from the z_{ij} that were actually recorded. Toward this, we define $z_{ij} = \max(0, v_{ij})$ with $v_{ij} \sim \mathsf{N}(\mu_i, \sigma_i^2)$ independently of each other. This definition reflects that each z_{ij} is non-negative and can equal zero with probability $1 - \Phi(\mu_i/\sigma_i)$. The year specific means μ_i and variances σ_i^2 allow year to year variation in the overall strength of the North Atlantic TC season. Although other models for z_{ij} could be considered, we prefer the truncated normal model because it leads to simple computations for learning its parameters within a Bayesian setting with a conjugate prior specification.

The unknown parameters in our model are η_1, η_2, ϕ_1, ϕ_2, ϕ_3, $\{\mu_i\}_{i=1}^N$, $\{\sigma_i^2\}_{i=1}^N$, $\{\gamma_i\}_{i=1}^N$ and β. For our illustrations, we fix the detection parameters β and $\{\gamma_i\}_{i=1}^N$, partly to emphasize that the inference on the other parameters, particularly η_2, is quite sensitive to the choice of these parameters. Specific choices are given in the next section. Prior beliefs about each of the remaining parameters are specified by choosing an appropriate probability distribution as described in Table 1. These parameters are taken to be *a priori* independent of each other, i.e., the joint prior distribution on all these parameters is simply the product of the marginal prior distributions that appear in Table 1. The entries in Table 1 are chosen as follows. For each parameter, we look at model features (usually one) that are directly influenced by the parameter. Then a prior distribution is chosen to provide reasonable values for the *a priori* mid-point and range for each of these model features. All prior distributions are chosen from

Parameter	Prior	Related TC characteristic	TC characteristic median (95% CI)
η_1	$\mathsf{N}(2,1)$	Mean TC count in central year: e^{η_1}	7 (1, 52)
η_2	$\mathsf{N}(0,0.7^2)$	Percentage change over N years in mean TC count relative to year 1: $e^{\eta_2}-1$	0% (-75%, 294%)
ϕ_1	Discrete uniform on $\{-.5,-.475,\cdots,.5\}$	Percentage change from minimum to maximum mean TC count due to oscillation: $e^{2\lvert\phi_1\rvert}-1$	65% (3%, 165%)
ϕ_2	Discrete uniform on $\{-.5,-.4,\cdots,.5\}$	Oscillation phase shift: $\pi\phi_2$	0 ($-\pi/2$, $\pi/2$)
ϕ_3	Discrete uniform on $\{2,2.2,\cdots,8\}$	Number of oscillation cycles: $\phi_3/2$	2.5 (1, 4)
$(\mu_i,\sigma_i^2)_{i=1}^N$	$\mu_i\lvert\sigma_i^2\sim\mathsf{N}(25,\frac{\sigma_i^2}{100})$, $\sigma_i^{-2}\sim\mathsf{Ga}(50,50\cdot10^4)$, independent over i	A. Fraction of years with 40% or more non-hurricane TCs: $\frac{\#\{i:\ \Phi(-\frac{\mu_i}{\sigma_i})\geq.4\}}{N}$	51% (43%, 60%)
		B. Fraction of years with 1/6-th or more TCs staying category 1 or higher for 5 or more days: $\frac{\#\{i:\ \Phi(\frac{120-\mu_i}{\sigma_i})\geq1/6\}}{N}$	57% (49%, 64%)

Table 1: Prior distributions on model parameters (except for detection probability parameters).

simple exponential family distributions or discrete distributions. A normal-inverse-gamma prior for the (μ_i, σ_i^2)'s is chosen because of its conjugacy properties.

4 Illustration

To specify the detection probability parameters β and $\{\gamma_i\}_{i=1}^N$, we first split the study period into 10 sub-periods. The separation points are chosen to reflect significant changes in detection technology over the years as well as some other global events that are likely to have affected TC recording. These ten sub-periods are shown on the first column of Table 2, with the events determining the onset of these sub-periods described in the second column. The γ_i values for all years within a sub-period are taken to be identical.

		E1	E2	E3
y_i	Comments	γ_i	γ_i	γ_i
1871-1872	Beginning of study	-3.5	-3.5	-3.5
1873-1905	US Signal Service	-2.75	-3.25	-3.25
1906-1916	Ships with radio	-0.75	-1.5	-1.5
1917-1920	US in WWI & after-effects of WWI	-1.5	-2.25	-2.25
1921-1940	Post WWI	0.5	-1.25	-1.25
1941-1945	US in WWII	-2.5	-2.5	-2.5
1946-1966	Post WWII & aircraft	0.5	0.5	0.5
1967-1973	Early satellite era	1.5	1.5	1.5
1974-2001	Infrared satellites; better resolution & coverage	2	2	2
2003-2008	New tools (QuikSCAT, Microwave Sounding Unit, Cyclone Phase Space Analysis)	2.5	2.5	2.5
		β	β	β
		0.02	0.02	0.01

Table 2: Choice of detection probability parameters γ_i and β for experiments E1, E2 and E3.

In our first experiment (E1 hereafter), we consider values of γ_i and β as given in the E1 column of Table 2. The detection probabilities in E1 were chosen to obtain a posterior summary of missed TC counts similar to that of Vecchi and Knutson (2008). For this choice, detection probabilities of 7 hypothetical TCs are shown in the top-left

panel of Figure 2. These 7 TCs correspond to 7 different levels of strength, starting from a TC that never reached category 1 windspeed, to one that was category 1 or more for 10 days. For each sub-period, the detection probabilities of the 7 storms are shown by 7 dots in the middle. Detection probabilities are flat within a sub-period; the lines joining dots from one sub-period to the next are purely for visual assistance.

Once the detection probabilities are specified, we learn about other model parameters from data through their joint posterior distribution as determined by the likelihood function and the prior. We use a reversible jump Markov chain Monte Carlo (Green 1995) to sample from the joint posterior of the model parameters plus the missing observations m_i and z_{ij} (for unobserved TCs). A Metropolis update with Gaussian increment is used for each of η_1 and η_2, with increment size chosen to achieve an acceptance rate close to 45%. Gibbs updates are used for ϕ_1, ϕ_2, ϕ_3, and the block $\{(\mu_i, \sigma_i^2)\}_i$. The conditional posterior distribution of each of these parameters assumes a simple form thanks to either discreteness or well-known conjugacy properties of the normal-inverse-gamma prior.

The missing observations $\{(m_i, \{z_{ij}\}_{j=n_i+1}^{n_i+m_i})\}_{i=1}^N$ are updated via reversible jump Metropolis. For a randomly chosen year, an "addition" or a "deletion" is proposed with equal probability. In the case of addition, a missing TC, with z_{ij} generated from the prior, is proposed to be added to that year's TC count. For deletion, a TC from that year's list of missing TCs is chosen randomly and is proposed to be removed. These proposals are complementary to each other and lead to simple calculations of acceptance probabilities that preserve detailed balance. For a year with no missing TC currently imputed, only the addition proposal is made. We use 50 addition-deletion moves per iteration of the MCMC. Additionally, we update the z_{ij} values of all imputed TCs by a Gibbs update which is available due to our formulation of $z_{ij} = \max(0, v_{ij})$ with a normal prior on the v_{ij}'s.

Posterior sampling is done through two parallel runs of the Markov chain, with starting points overdispersed with respect to the imputed TC counts. One run starts with zero imputation ($m_i = 0$) for all years while the other run starts with m_i's generated from $\mathsf{Po}(\lambda_i(1 - \Phi(\gamma_i)))$ distributions. Each run is 20,000 iterations long. Convergence takes place within a few hundred iterations (see supplementary materials). The first quarter of each chain is discarded, the rest is thinned and the two chains are then pooled together to form a sample of 6000 draws from the posterior distribution.

The bottom-left panel of Figure 2 shows the posterior distribution over missed TC counts m_i under choice E1. The posterior means of m_i are shown by the gray line.

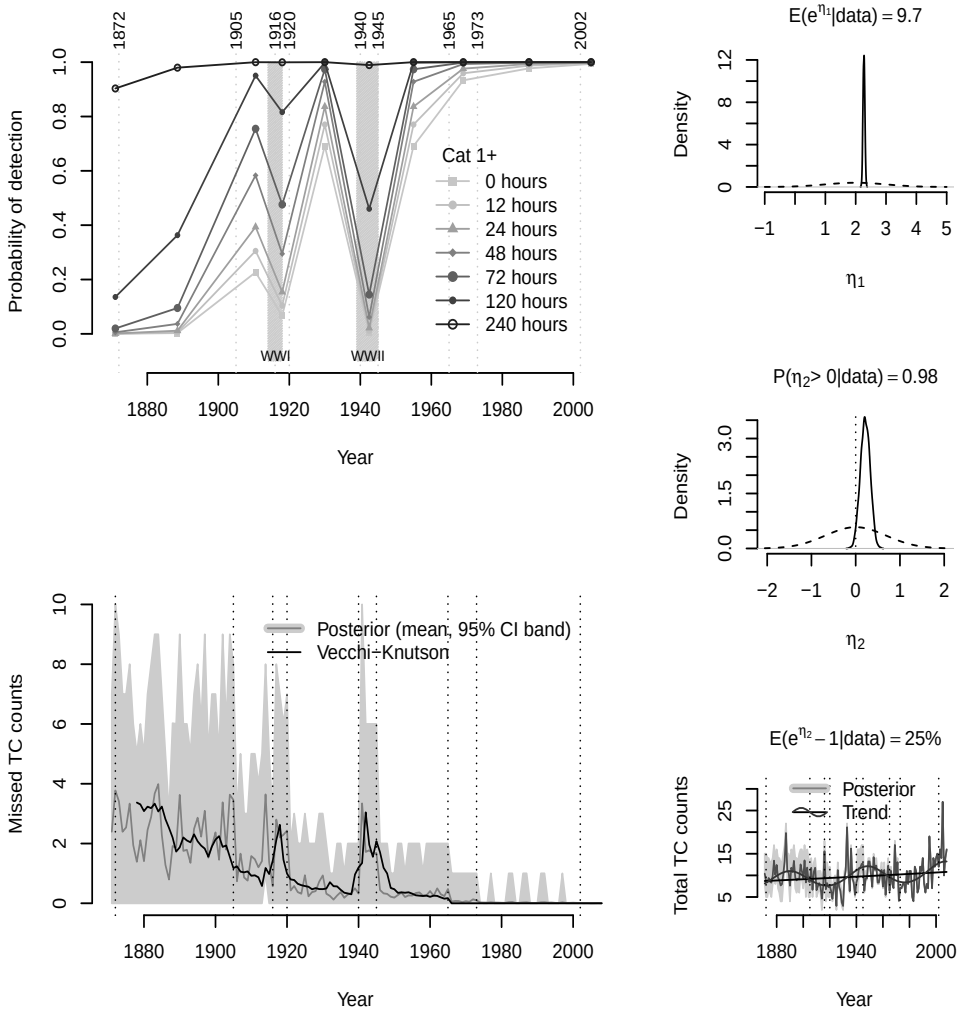

Figure 2: A visual summary of experiment E1. Top-left panel shows chosen detection probabilities across years for a set of 7 hypothetical TCs of various strength, measured by the hours each TC was a category 1 or higher. The vertical dotted lines mark the end of detection sub-periods as described in Table 2. Bottom-left panel shows posterior mean and range (95% equal-tail credible interval) for missed TC counts across years, overlaid with missed TC counts reported in Vecchi and Knutson (2008). Posterior (solid) and prior (broken) densities of η_1 and η_2 are shown on the top and middle panels on the right. Bottom-right panel shows posterior mean and 95% credible band of annual TC counts (observed + imputed).

	E1	E2	E3
$\mathsf{P}(\eta_2 > 0 \mid \text{data})$	0.98	0.81	0.31
$\mathsf{E}\,(e^{\eta_2} - 1 \mid \text{data})$	25%	11%	-4%

Table 3: Posterior inference on the trend of mean TC count under experiments E1, E2 and E3.

Overlaid on this is the black line showing the missed TC counts reported by Vecchi and Knutson (2008). The detection probabilities in E1 were chosen to obtain a posterior summary of missed TC counts similar to that of Vecchi and Knutson (2008). The top and middle panels on the right show posterior densities of η_1 and η_2. The posterior mean total TC count in the central year is approximately 10. The posterior places an overwhelmingly large probability (98%) on the slope η_2 being positive. Thus the assumptions of E1 strongly support the conclusion of increasing TC activity. The posterior mean of $e^{\eta_2} - 1$, the change in TC counts over the N years (relative to beginning of the study) is 25%. On top of this linear growth, the posterior also supports additional fluctuation in annual TC counts through multidecadal oscillation (bottom-right panel) with a mean of 2.3 (95% interval = (2.1, 2.5)) oscillation cycles ($\phi_3/2$) over the entire period. The oscillation alone accounts for an average 49% (95% interval = (28%, 82%)) relative change from minimum to maximum mean TC counts ($e^{2|\phi_1|} - 1$). Posterior summaries of several model parameters are provided in the supplementary materials.

5 Sensitivity to prior quantification

We consider two other choices of detection probability parameters, as given in the E2 and E3 columns of Table 2, to illustrate how the posterior trend critically depends on these choices. E2 is similar to E1, except for the years 1878 through 1940. E1 assigns very high detection probabilities ($\approx$ 80% or more) even to tropical storm strength TCs in the period 1921-1940, when measurements were based on ship and land records only. E2 (Figure 3) presents a somewhat less optimistic view, where the detection probabilities are more than 50% only for TCs that were category 1 or stronger for at least 3 days. E2 also lowers detection probabilities for 1878-1920, to maintain the same relative patterns in the 1878-1940 period as given by E1. E3 (Figure 4) is exactly the same as E2 except for the value of β which is halved.

Experiments E2 and E3 lead to substantially different inferences on the TC count trend relative to E1 (Table 3). Under E2, the posterior probability of η_2 being positive

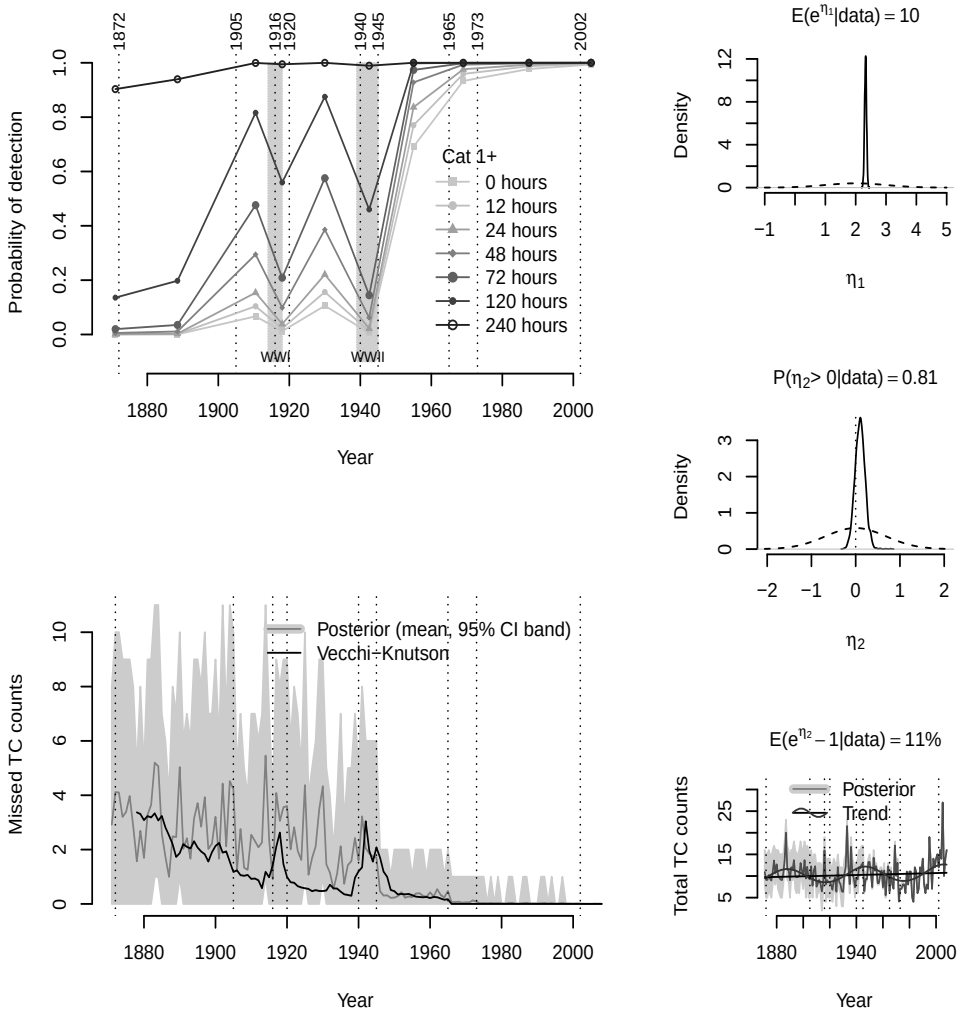

Figure 3: Visual summary of experiment E2. Conclusion of a positive trend is much weaker than that in E1.

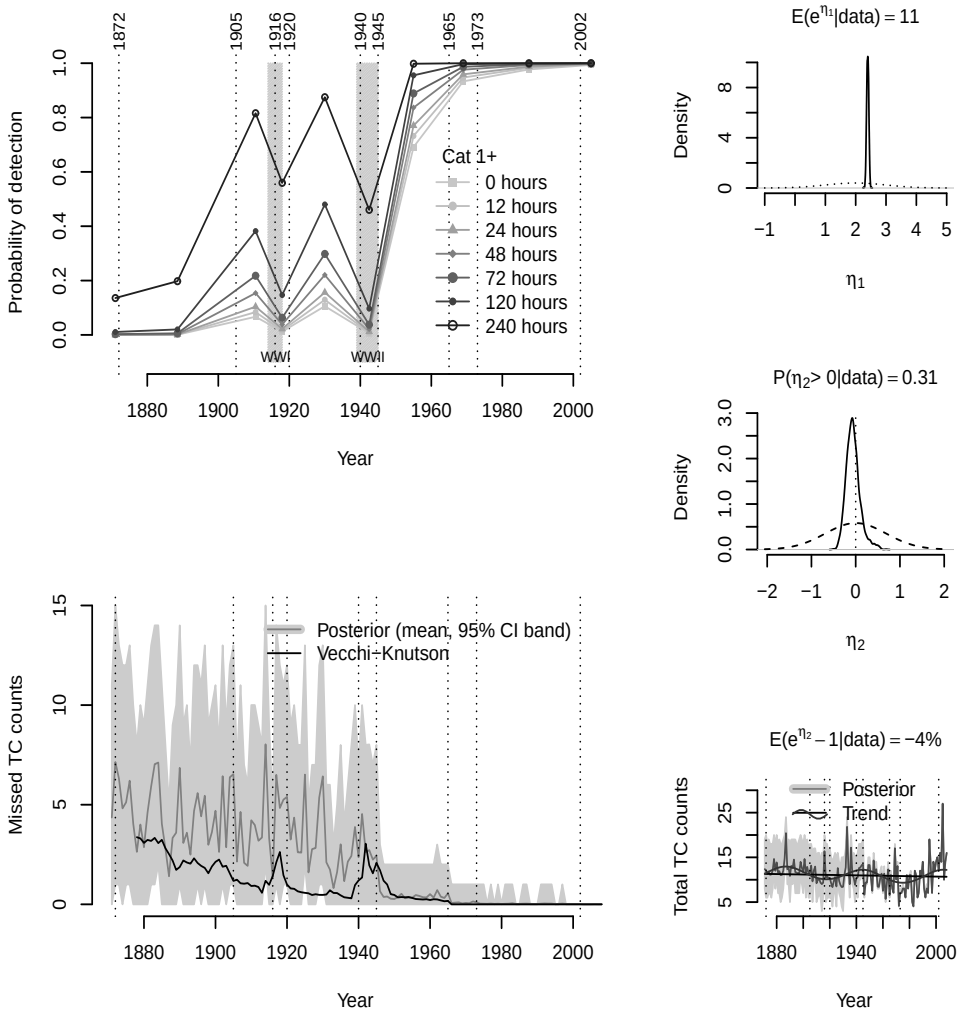

Figure 4: Visual summary of experiment E3. The posterior assigns 70% chance of a negative trend, with a mean drop of 4% in average TC counts across the period of study.

is less overwhelming (81%) than that under E1. The posterior mean of total percentage increase reduces to 10%. E3 presents a different picture, where the posterior probability of $\eta_2 > 0$ drops down to 31% and the mean trend is negative, with about a 4% drop in TC counts over the study period. Both E2 and E3 support multidecadal oscillation similar to E1, with an average 2.3 oscillation cycles, but with reduced amplitudes.

6 Concluding remarks

6.1 Resolving ambiguity with further studies

Our three experiments point to significant ambiguity present in the historical tropical cyclone occurrence records, with the assessment of TC counts trend clearly linked to assumptions regarding historic detection probabilities. Can we draw any reasonable conclusions about the trend despite this ambiguity? We feel the answer is yes, but it would involve using beliefs about detection probabilities that are deemed reasonable by the scientific community.

First, formal elicitation could be conducted with experts on past tropical cyclone detection to derive expert-specific detection probability formulation. This can be achieved by an extension of the statistical framework presented here. The main vehicle for this extension is the probability detection plot that appears on the top-left panel of Figures 2, 3, and 4. This plot provides an interface between our statistical model and quantities that can be elicited from a climate expert.

The expert would be required to generate her version of this plot by quantifying her belief about detection probabilities of a collection of TCs of various intensities, durations, tracks, etc. across different eras starting from the mid nineteenth century. These elicited quantities would be used to identify a suitable detection probability function (3) with z_{ij} possibly reflecting a multitude of TC characteristics and Φ possibly replaced with a different distribution function. It might also be necessary to use a non-trivial prior distribution on the parameters in (3) to accommodate the expert's uncertainty about them. Once a suitable detection probability function is found to match the expert's belief, the rest of the modeling, computing and summarizing can proceed in exactly the same manner as in the examples presented here.

Ideally, such a study would be carried out with many different experts, generating a catalog of conclusions about TC trend based on current expert opinions. Whether any kind of scientific consensus might result from this remains to be seen.

6.2 Sensitivity to formulation

In incorporating an expert's quantified belief, it is important to ascertain how sensitive the results will be to the particular choice of the formulation of (3). In particular, if an expert's quantified beliefs are well represented by two different formulations of (3), will the conclusions about trend depend on which formulation is used? While such a question cannot be addressed in full generality, we report below additional experiments that suggest the conclusions are indeed robust to moderate variations in the choice of (3).

We consider three additional experiments, E1*, E2* and E3* in which (3) is specified as

$$\pi_{ij} = F_3(\gamma_i^* + \beta_i^* z_{ij}) \tag{4}$$

where F_3 denotes the cumulative distribution function of the Student-t distribution with 3 degrees of freedom. We continue with our 10 sub-periods split of the study period (Table 2) and assign a common value to γ_i^* for all years i within a sub-period. The same is done for β_i^*.

In experiment E1*, we choose γ_i^*'s and β_i^*'s to match the "quantified beliefs" of E1 as displayed on the top-left panel of Figure 2. For the 7 (hypothetical) TCs displayed on that panel, with strengths $z_1^h = 0, z_2^h = 12, \cdots, z_7^h = 240$, we record their detection probabilities $\pi_{1k}^h, \pi_{2k}^h, \cdots, \pi_{7k}^h$ for all sub-periods $k = 1, \cdots, 10$, as specified under E1. Next we find $\hat{\gamma}_k$ and $\hat{\beta}_k$ for $k = 1, \cdots, 10$, by minimizing

$$f(\gamma_1, \cdots, \gamma_{10}, \beta_1, \cdots, \beta_{10}) = \left[\frac{1}{70}\sum_j \sum_k \{\pi_{jk}^h - F_3(\gamma_k + \beta_k z_j^h)\}^2\right]^{1/2} \tag{5}$$

and set $\gamma_i^* = \hat{\gamma}_k$ and $\beta_i^* = \hat{\beta}_k$ for all years i within sub-period k. This minimization is done numerically by using the `optim()` routine of `R` and the minimum value equals 0.0091, which suggests a reasonably good fit. The rest of the model is kept as before. Posterior summaries are generated by the reversible jump Markov chain sampler as discussed earlier with a suitable modification to reflect the change from (3) to (4). Experiments E2* and E3* are similar in design, but with "quantified beliefs" coming from E2 and E3, respectively. The minimum value of $f(\gamma_1, \cdots, \gamma_{10}, \beta_1, \cdots, \beta_{10})$ equals 0.0101 for E2* and 0.0102 for E3*.

Table 4 shows conclusions about trend under E1*, E2* and E3*. Full graphical summaries, as in Figure 2 etc., are included in the supplementary materials. Despite the differences in (3) and (4), the conclusions about trend in these new experiments are

	E1*	E2*	E3*
$\mathsf{P}(\eta_2 > 0 \mid \text{data})$	0.98	0.83	0.34
$\mathsf{E}\,(e^{\eta_2} - 1 \mid \text{data})$	24%	12%	-3%

Table 4: Posterior inference on the trend of mean TC count under experiments E1*, E2* and E3*.

virtually indistinguishable from the conclusions drawn in, respectively, E1, E2 and E3.

It is possible to obtain different conclusions than E1 etc. by replacing F_3 in (4) with a function that is more dissimilar to Φ. In fact, with F_1 instead of F_3, the minimum value of $f(\gamma_1, \cdots, \gamma_{10}, \beta_1, \cdots, \beta_{10})$ undergoes about a 3-fold increase to values of 0.0247, 0.0259 and 0.0275 for E1*, E2* and E3* respectively, suggesting a less satisfactory fit. The conclusion about η_2 is different, but not by a big margin (supplementary materials). Equation (4) with F_1 instead of F_3 does not provide as good a fit to the "quantified beliefs" of E1 and so the difference in conclusion is not worrying. However, determining lack of fit to an expert's quantified beliefs is a nontrivial task and should be based upon both graphical plots of detection probabilities and numerical values of f.

6.3 Toward a flexible formulation

Both (3) and (4) are limited in their ability to encode an expert's quantified beliefs. A slightly richer formulation, of which both (3) and (4) are special cases, can be obtained by taking $\pi_{ij} = F_\nu(\gamma_i^* + \beta_i^* \cdot z_{ij})$ where the degrees of freedom parameter ν is also to be included in the minimization of $f(\gamma_1, \cdots, \gamma_{10}, \beta_1, \cdots, \beta_{10}, \nu) = [\frac{1}{70} \sum_j \sum_k \{\pi_{jk}^h - F_\nu(\gamma_k + \beta_k z_j^h)\}^2]^{1/2}$. Here z_{ij} is taken to be a vector of strength measures which would possibly include summaries of duration, size and trajectory. This vector is to be decided upon in consultation with the expert to characterize the variation in TCs she chooses as the examples for her belief quantification.

6.4 Incorporating climate models

The proposed analysis could be extended further by combining expert knowledge with climate model projections of North Atlantic TC activity. Climate model studies have investigated how TC intensity, frequency and tracks may change in a warmer climate both globally and in individual basins such as the North Atlantic. Global climate models that are used to project temperature increases resulting from increased CO_2 levels have

low resolution (Knutson et al. 2010) and consequently are limited in their ability to simulate TCs. Recent approaches have combined the output from global models with regional higher-resolution models that are able to simulate more realistic TCs (Emanuel et al. 2008; Knutson et al. 2010, 2008).

Ideally, the results of such models would be used to specify priors for the vector z_{ij}, where this vector includes a multitude of TC characteristics such as intensity, duration, seasonal variation, and geographic distribution of tracks. This would allow the experts to focus on the detection probabilities during the elicitation process. However, despite the improvements in model resolution, currently available models remain somewhat limited in their capacity to simulate realistic distributions of these TC properties. In particular, projections of TC frequencies in individual basins and changes in tracks and duration are regarded as rather unreliable at present as is evident, for instance, in the disagreement on the sign of changes in these parameters (Knutson et al. 2010; Grossmann and Morgan 2011). A recent model driven by annual observed North Atlantic ocean temperatures and atmospheric parameters over the 27-year period 1980-2006 was able to simulate North Atlantic TC activity that agreed remarkably well with observed activity. However, this model first had to be calibrated to the observed basinwide TC counts, the very dataset that is sought to be corrected. This introduces a circular problem. We also note that this kind of study cannot be extended to the period prior to 1980 because the required detailed atmospheric parameters are only available from 1980 onwards (Kalnay et al. 1996).

References

Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage.

Atlas, R., Hoffman, R. N., Leidner, S. M., Sienkiewicz, J., and Yu, T.-W. (2001). "The effects of marine winds from scatterometer data on weather analysis and forecasting." *Bulletin of the American Meteorological Society*, 82: 1965–1990.

Bengtsson, L. and Hodges, K. I. (2008). "Comment on "Is the number of North Atlantic tropical cyclones significantly underestimated prior to the availability of satellite observations?" by EKM Chang and Y Guo." *Geophysical Research Letters*, 35: doi:10.1029/2007GL032251.

Bowditch, H. I. (1841). *Memoir of Nathaniel Bowditch*. James Munroe and Company.

Bowditch, N. (1995). *The American Practical Navigator*. United States Government.

Brueske, K. F. and Velden, C. S. (2003). "Satellite-based tropical cyclone intensity estimation using the NOAA-KLM series Advanced Microwave Sounding Unit (AMSU)." *Monthly Weather Review*, 131: 687–697.

Chang, E. K. M. and Guo, Y. (2007). "Is the number of North Atlantic tropical cyclones significantly underestimated prior to the availability of satellite observations?" *Geophysical Research Letter*, 34: doi:10.1029/2007GL030169.

Coles, S. G. and Sparks, R. S. J. (2006). "Extreme value methods for modeling historical series of large volcanic magnitudes." In *Statistics in Volcanology*, volume Special Publication of IAVCEI, 1, 47–56. London: Geological Society.

Cornulier, T., Robinson, R. A., Elston, D., Lambin, X., Sutherland, W. J., and Benton, T. G. (2011). "Bayesian reconstitution of environmental change from disparate historical records: hedgerow loss and farmland bird declines." *Methods in Ecology and Evolution*, 2: 86–94.

Dorst, N. M. (2007). "The National Hurricane Center research project: 50 years of research, rough rides, and name changes." *Bulletin of the American Meteorological Society*, 88(10): 1566–1588.

Dunn, G. E. and Miller, B. I. (1960). *Atlantic hurricanes*. Louisiana State University Press.

Dvorak, V. F. (1984). "Tropical Cyclone Intensity Analysis Using Satellite Data." NOAA Technical Report NESDIS 11, 47pp.

Elsner, J. B. and Bossak, B. H. (2006). "Hurricane landfall probability and climate." In Murnane, R. J. and Liu, K.-B. (eds.), *Hurricane and typhoons - past, present and future*, 333–353. Columbia University Press.

Elsner, J. B. and Jagger, T. H. (2006). "Prediction Models for Annual U.S. Hurricane Counts." *Journal of Climate*, 19: 2935–2952.

Elsner, J. B. and Kara, A. B. (1999). *Hurricanes of the North Atlantic: climate and society*. New York: Oxford University Press.

Emanuel, K. A., Sundararajan, R., and Williams, J. (2008). "Hurricanes and global warming: Results from downscaling IPCC AR4 simulations." *Bulletin of American Meteorological Society*, 89: 347–367.

Fernández-Partagás, J. and Diaz, H. F. (1996). "Atlantic hurricanes in the second half of the nineteenth century." *Bulletin of the American Meteorological Society*, 77: 2899–2906.

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.

Green, P. J. (1995). "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination." *Biometrika*, 82: 711–732.

Grossmann, I. and Morgan, M. G. (2011). "Tropical cyclones, climate change, and scientific uncertainty: what do we know, what does it mean, and what should be done?" *Climatic Change*, doi: 10.1007/s10584–011–0020–1.

Harper, B. A. and Callaghan, J. (2006). "On the importance of reviewing historical tropical cyclone intensities." 27th Conference on Hurricanes and Tropical Meteorology. American Meteorological Society: April 23-28, Monterey, CA.

Holland, G. J. (2007). "Misuse of landfall as a proxy for Atlantic tropical cyclone activity." *Eos*, 88(36): 349–356.

Holland, G. J. and Webster, P. J. (2007). "Heightened tropical cyclone activity in the North Atlantic: natural variability or climate trend?" *Philosophical Transactions of the Royal Society A*, 365(1860): 2695–2716, doi:10.1098/rsta.2007.2083.

Honaker, J. and King, G. (2010). "What to do about missing values in time-series cross-section data." *American Journal of Political Science*, 54: 561–581.

Ibrahim, J. G., Chen, M. H., and Lipsitz, S. R. (2001). "Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable." *Biometrika*, 88: 551–564.

Jarvinen, B. R., Neumann, C. J., and Davis, M. A. S. (1984). "A tropical cyclone data tape for the North Atlantic basin, 1886-1983: contents, limitations, and uses." NOAA Technical Memorandum NWS HHC 22, Coral Gables, FL, 21 pp. Available online at http://www.nhc.noaa.gov/pdf/NWS-NHC-1988-22.pdf.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J.,

Jenne, R., and Joseph, D. (1996). "The NCEP/NCAR 40-Year Reanalysis Project." *Bulletin of the American Meteorological Society*, 77: 437–471.

Kéry, M. and Royle, J. A. (2010). "Hierarchical modelling and estimation of abundance and population trends in metapopulation designs." *Journal of Animal Ecology*, 79: 453–461.

Klotzbach, P. J. and Gray, V. M. (2008). "Multidecadal variability in North Atlantic tropical cyclone activity." *Journal of Climate*, 21: 3929–3935.

Knapp, K. R. and Kossin, J. P. (2007). "New global tropical cyclone data from ISCCP B1 geostationary satellite observations." *Journal of Applied Remote Sensing*, 1: 13505–13510.

Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J. P., Srivastava, A. K., and Sugi, M. (2010). "Tropical cyclones and climate change." *Nature Geoscience*, 3: 157–163.

Knutson, T. R., Sirutis, J. J., Garner, S. T., Held, I. M., and Tuleya, R. E. (2007). "Simulation of the recent multidecadal increase of Atlantic hurricane activity using an 18-km-grid regional model." *Bulletin of American Meteorological Society*, 88: 1549–65.

Knutson, T. R., Sirutis, J. J., Garner, S. T., Vecchi, G. A., and Held, I. M. (2008). "Simulated reduction in Atlantic hurricane frequency under twenty-first-century warming conditions." *Nature Geoscience*, 1: 359–364.

Landsea, C. W. (2007). "Counting Atlantic tropical cyclones back to 1900." *Eos*, 88(18): 197–202.

Landsea, C. W., Anderson, C., Charles, N., Clark, G., Dunion, J., Fernández-Partágas, J., Hungerford, P., Neumann, C., and Zimmer, M. (2004). "The Atlantic hurricane database reanalysis project: Documentation for the 1851-1910 alterations and additions to the HURDAT database." In Murname, R. J. and Liu, K. B. (eds.), *Hurricanes and Typhoons: Past, present and future*. 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 renalysis of the 1911-1920 Atlantic hurricane database." *Journal of Climate*, 21: 2138–2168.

Landsea, C. W., Harper, B. A., Harau, K., and Knaff, J. A. (2006). "Can we detect trends in extreme tropical cyclones?" *Science*, 313: 452–454.

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.

Little, R. J. A. (1995). "Modeling the dropout mechanism in repeated-measures studies." *Journal of the American Statistical Association*, 90: 1112–1121.

Mann, M. E. and Emanuel, K. A. (2006). "Atlantic hurricane trends linked to climate change." *Eos*, 87(24): 233–244.

Mann, M. E., Sabbatelli, T. A., and Neu, U. (2007). "Evidence for a modest undercount bias in early historical Atlantic tropical cyclone counts." *Geophysical Research Letters*, 34: doi:10.1029/2007GL031781.

Neumann, C. J., Jarvinen, B. R., and McAdie, C. J. (1999). "Tropical cyclones of the North Atlantic Ocean, 1871-1998." NOAA Historical Climatology Series 6-2.

Nyberg, J., Malmgren, B. A., Jury, A. W. M. R., Kilbourne, K. H., and Quinn, T. M. (2007). "Low Atlantic hurricane activity in the 1970s and 1980s compared to the past 270 years." *Nature*, 447: 698–702.

Piddington, H. (1860). *The sailors horn-book for the law of storms*. Williams & Norgate.

Rubin, D. B. (1996). "Multiple imputation after 18+ years." *Journal of the American Statistical Association*, 91: 473–489.

Sheets, R. C. (1990). "The National Hurricane Center – Past, present and future." *Weather and Forecasting*, 5(2): 185–232.

Simpson, R. H. (1974). "The hurricane disaster potential scale." *Weatherwise*, 27: 169–186.

Solow, A. R. (1989). "Statistical modeling of storm counts." *Journal of Climate*, 2: 131–136.

Solow, A. R. and Beet, A. R. (2008). "On the incompleteness of the historical record of North Atlantic tropical cyclones." *Geophysical Research Letters*, 35: doi:10.1029/2008GL033546.

Solow, A. R. and Moore, L. (2000). "Testing for trend in a partially incomplete hurricane record." *Journal of Climate*, 13: 3696–3699.

— (2002). "Testing for trend in North Atlantic hurricane activity, 1900-1998." *Journal of Climate*, 15: 3111–3114.

Tingley, M. and Huybers, P. (2010). "A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part 1: Development and applications to paleoclimate reconstruction problems." *Journal of Climate*, 23: 2759–2781.

Vecchi, G. A. and Knutson, T. R. (2008). "On estimates of historical North Atlantic tropical cyclone activity." *Journal of Climate*, 21: 3580–3600.

Velden, C., Harper, B., Wells, F., Beven, J. L. I., Zehr, R., Olander, T., Mayfield, M., Guard, C., Lander, M., Edson, R., Avila, L., Burton, A., Turk, M., Kikuchi, A., Christian, A., Caroff, P., and McCrone, P. (2006). "The Dvorak tropical cyclone intensity estimation technique." *Bulletin of the American Meteorological Society*, 10: 1195–1210.

Webster, P. J., Holland, G. J., Curry, J. A., and Chang, H. R. (2005). "Changes in tropical cyclone number, duration and intensity in a warming environment." *Science*, 309: 1844–1846.

Zeger, S. L. and Liang, K. Y. (1996). "Longitudinal data analysis for discrete and continuous outcomes." *Biometrics*, 42: 121–130.

Acknowledgements

We thank the referees and an Associate Editor who reviewed this paper for helpful comments and suggestions. The second author was supported by the Climate Decision Making Center created through a cooperative agreement between the National Science Foundation (SES-0345798) and Carnegie Mellon University.