Tellus (2007), 59A, 486–498
Printed in Singapore. All rights reserved

© 2007 The Authors
Journal compilation © 2007 Blackwell Munksgaard

TELLUS

Statistical modelling of North Atlantic tropical cyclone tracks

By TIMOTHY M. HALL[1]* and STEPHEN JEWSON[2], [1]*NASA GISS, New York, NY, USA;* [2]*Risk Management Solutions, London, UK*

(Manuscript received 15 August 2006; in final form 19 February 2007)

ABSTRACT

We present a statistical model of North Atlantic tropical cyclone tracks from genesis site through lysis. To propagate tracks we use the means and variances of latitudinal and longitudinal displacements and model the remaining anomalies as autoregressive. Coefficients are determined by averaging near-neighbour historical track data, with 'near' determined optimally by using jackknife out-of-sample validation to maximize the likelihood of the observations. The number of cyclones in a simulated year is sampled randomly from the historical record, and the cyclone genesis sites are simulated with a spatial probability density function using kernels with optimized bandwidths. Simulated cyclones suffer lysis with a probability again determined from optimal averaging of historical lysis rates. We evaluate the track model by comparing an ensemble of 1950–2003 simulations to the historical record using several diagnostics, including landfall rates. In most regions, but not all, the observations fall within the variability across the ensemble members, indicating that the simulations and observations are statistically indistinguishable. An intensity component to the TC model, necessary for risk assessment applications, is currently under development.

1. Introduction

Powerful tropical cyclones (TCs) are among the most devastating natural phenomena, and there has been great interest in estimating and forecasting the risk of wind, rainfall and flooding damage. Numerical weather forecast models, which integrate the fundamental equations of motion from observed initial conditions, form the basis for predicting the evolution of particular TCs days in advance. At seasonal and longer time scales, however, weather is unpredictable, and TCs must be treated stochastically, although ensemble simulations with numerical weather models may still be useful. Various approaches to statistical TC risk assessment have been developed over the past 20 yr by private sector and academic researchers. Generally, detailed descriptions of private-sector models are not publicly available, with only broad outlines published in conference abstracts. The models published with more detailed descriptions include Darling (1991), Chu and Wang (1998), Vickery et al. (2000), Casson and Coles (2000), James and Mason (2005), Emanuel et al. (2006) and Rumpf et al. (2007).

The characteristics of TCs at landfall are of primary interest in risk assessment, and a natural approach to modelling landfall statistically is to use historical data on TCs exclusively at landfall (e.g. Jagger et al., 2001). However, in many coastal regions there are few or no historical events, making assessment of the risk difficult. One way to overcome this limitation is to make use of entire historical TCs, from genesis to lysis, thereby enhancing by roughly two orders of magnitude the amount of data on which to construct a statistical model. Of course, the majority of the data is less relevant to landfall. Historical TC behaviour in the central Atlantic contains less information useful for assessing landfall rates than does TC behaviour at or near a coastal region. Nonetheless, in regions of rare landfall, historical TCs that merely pass within 100 s of kilometres of landfall provide constraints for risk assessment.

A crucial component of basin-wide TC modelling is the TC track, the geographic trajectory from genesis to lysis, and it is the modelling of this component that we focus on here. Basin-wide approaches have been taken by several researchers, for tracks as well as intensity. Vickery et al. (2000) use an autoregressive model for increments in track speed and direction, with a random error term acting as the innovation. Separate regression coefficients are fit and gridded for eastward and westward heading TCs. James and Mason (2005) also fit an autoregressive model for TCs in the Coral Sea near northeastern Australia, but they model the latitudinal and longitudinal increments, rather than the velocity increment. Their coefficients do not vary spatially, and an ad hoc term is added to the latitude model to inhibit tracks from propagating too near the equator. Emanuel et al. (2006)

*Corresponding author.
e-mail: thall@giss.nasa.gov
DOI: 10.1111/j.1600-0870.2007.00240.x

propagate tracks (in one of their two approaches) by sampling a transition matrix, populated from historical analysis, that relates prior track speed and direction to the new speed and direction. Casson and Coles (2000) simply draw from the set of complete historical tracks, translating the tracks by small random displacements. Rumpf et al. (2007) separate TCs into independent classes based on geographic characteristics, then sample kernel probability density functions (pdfs) build from historical speed and direction increments to propagate the simulated TCs.

Simulated TC genesis in these studies is performed in several ways. Emanuel et al. (2006) sample from a time- and space-dependent pdf constructed by binning historical genesis events and smoothing. James and Mason (2005) apply a scheme that interpolates historical genesis. Vickery et al. (2000) simply sample directly the historical genesis sites. Rumpf et al. (2007) use a near-neighbour approach similar to ours to develop and sample a genesis kernel pdf.

We also take a basin-wide approach to North-Atlantic TC track modelling. Our model is non-parametric, in the sense that simulations are derived by spatially averaging historical data, rather than fitting parametric forms to the data. We have strived to make maximal use of historical data, without over-fitting the model, by using out-of-sample validation to optimize data averaging. In contrast to many other studies, we document explicitly this objective procedure to average historical data. At a given location **r**, we base the genesis, propagation, and lysis (death) of TCs on data near **r**. For the tracks, mean six-hourly displacement increments are computed, as are variances about the mean and autocorrelations of the anomalies of track displacements. The magnitude of the random noise forcing (the 'innovation') depends on the variance and the autocorrelation, and is ultimately drawn from model residuals. Genesis and lysis rates are modelled by sampling pdfs built from historical events using kernel techniques. For all elements, the length scales over which historical data are averaged (the definition of 'near') are chosen to maximize the jackknife out-of-sample likelihood of the observations.

This paper summarizes are present state of model development. In contrast to many other published work on stochastic TC modelling no intensity is simulated here, nor is any intensity information used in the track modelling. Intensity modelling is clearly indispensable for TC risk assessment, and we are presently developing an intensity model to complement the track model. However, we believe it worthwhile first to describe and evaluate the TC track component of the model in detail.

After reviewing the historical data we outline our modelling procedure. We then describe in detail each of the modelling elements: mean displacements, variance, and autocorrelation for the propagation; and genesis and lysis. Subsequently, we compare simulated TCs to historical TCs using the large-scale diagnostics of track-point density, latitude and longitude crossing rates, and landfall rates.

2. Data

Following other North Atlantic TC modelling efforts (Vickery et al., 2000; Emanuel et al., 2006) we use the HURDAT 'best track' historical tropical cyclone data set compiled by NOAA's National Hurricane Center (Javinen et al., 1984; www.aoml.noaa.gov/hrd/hurdat). HURDAT provides date, time, longitude, latitude, central pressure and maximum wind speed every 6 h for TCs rated tropical disturbance and higher back to the 1800s. However, only from 1950 was aircraft reconnaissance used routinely to monitor TCs. Information on earlier cyclones is less reliable. In this study we use 524 HURDAT TCs from 1950 to 2003, inclusive. This represent all TCs in this period, except for a small number (order 10) that had spurious six-hourly displacements (e.g. 10s of degrees latitude or longitude).

Figure 1 shows these 524 historical TCs. Most North Atlantic TCs are born in the subtropical middle Atlantic and the Caribbean. Their trajectories follow a general sweep northwestward in the subtropics, then veer northeastward at mid-latitudes. Superposed on this average behaviour is considerable pseudo-randomness; many TCs move in directions opposing the average trajectory. The randomness makes a stochastic approach to simulation necessary.

3. Outline of the tropical cyclone track model

Our simulation of a TC track can be summarized by the following steps.

(i) Generate the first point on the track from the genesis model. This is the first 'current point' of the simulation.

(ii) Compute mean latitude and longitude displacements from the current point by averaging historical displacements with a weight that declines with distance from the current point.

(iii) Compute variances about the mean in directions parallel and perpendicular to the mean, again weighting historical data inversely with distance from the current point.

(iv) Simulate a displacement as $u = \overline{u} + \epsilon_u u_{\mathrm{rms}}$ for the direction parallel to the mean track and $v = \epsilon_v v_{\mathrm{rms}}$ for the direction perpendicular, where overline indicates mean quantities and the subscript 'rms' indicates root-mean-squared variances. The ϵ are innovations, that is, random forcing drawn from a standard normal distribution of zero mean and unit variance (or, subsequently, from model residuals).

(v) Use u and v to propagate the TC to the second point. For the second point onward the simulation is summarized by the following steps.

(vi) Find the local mean and variances as above and store.

(vii) Find autocorrelation coefficients in the parallel direction (ϕ_u) and perpendicular direction (ϕ_v) by regressing all historical displacements against their previous displacements. The historical elements in the regression are weighed inversely with distance from the current point.

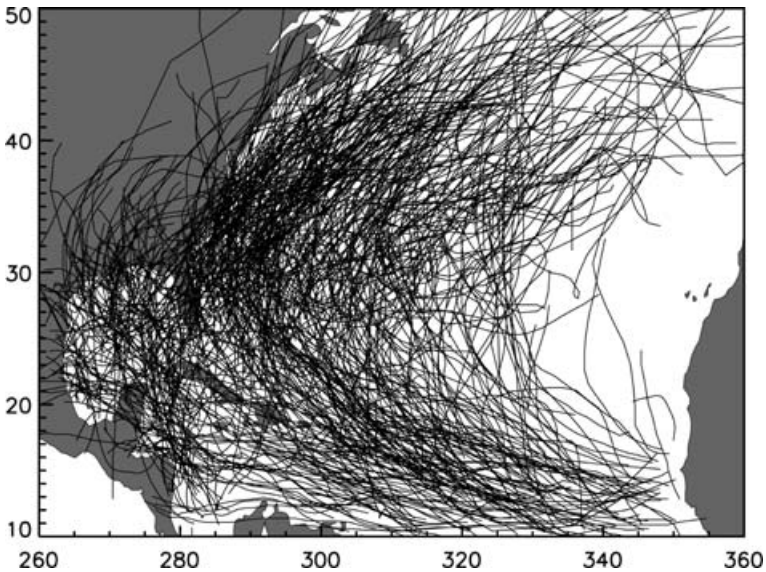

Fig. 1. The 524 HURDAT historical tropical cyclone tracks from 1950 to 2003.

(viii) Compute the magnitudes s_u and s_v of noise forcing using $s^2 = 1 - \phi^2$. (This is the statement that the total variance is equal to the variance correlated with the previous step plus the variance of the noise.)

(xi) Compute the parallel anomalies $\tilde{u}_{n+1} = \phi_u \tilde{u}_n + s_u \epsilon_u$ and perpendicular anomalies $\tilde{v}_{n+1} = \phi_v \tilde{v}_n + s_v \epsilon_v$, where tilde indicates a regularized anomaly quantity (zero mean and unit variance).

(x) Multiply anomalies by the rms variance and add the mean: $u_{\mathrm{rms}}\tilde{u} + \overline{u}$ and $v_{\mathrm{rms}}\tilde{v}$.

(xi) Rotate the displacements to the zonal-meridional orientation.

(xii) Update the current point with these displacements.

(xiii) Apply the lysis model, and terminate the TC if lysis occurs.

The mean, variance and autocorrelation coefficients, as well as elements of the genesis and lysis models, are computed from spatial averages of historical data. The form of the averaging is built on the premise that historical TC displacements closer to the current point should carry more weight. Climatological conditions vary spatially, and nearby historical displacements carry information most relevant to the local climate. On the other hand, the more restrictive is the weight about the current point the less historical information is used, and the averages suffer from sampling error. Thus, there is an optimal averaging length scale that balances the requirements of having local information but avoiding sampling error. This optimal length scale is computed using jackkife out-of-sample validation, as described in more detail in the sections below.

4. Model components

We now describe in detail the elements of the TC modelling. In the process analysis of the historical data is presented.

4.1. Means

The x (zonal) and y (meridional) components of the mean 6-h displacement vector are computed by averaging historical displacements:

$$\overline{x}(\mathbf{r}) = \frac{\Sigma_i x_i e^{-d_i^2/2L^2}}{\Sigma_i e^{-d_i^2/2L^2}}, \tag{1}$$

where d_i is the great-circle distance between the location $\mathbf{r}$ of the current point and the location $\mathbf{r}_i$ of the ith historical TC point, and similarly for $\overline{y}$. The length scale L is optimized using an out-of-sample jackknife procedure. Consider a six-hourly point $\mathbf{r}$ of a historical TC in calendar year j. Select a scale L and sweep over all six-hourly displacements over all historical TCs for all years $k \neq j$, forming $\overline{x}$ and $\overline{y}$. The average displacement vector so computed differs from the historical displacement vector emanating from $\mathbf{r}$, and the magnitude of the difference is the forecast error. We compute the average forecast error over all points for all TCs in all years j in the range 1950–2003. The averaged forecast error is minimized in L, within some tolerance.

Figure 2 shows the mean forecast error as a function of L. There is a minimum at 300 ± 10 km. Figure 3 illustrates the impact of the averaging scale on mean tracks. Ten mean tracks of equal duration are shown originating from evenly spaced longitudes along a line of constant latitude. For the case of $L = 100$ km the tracks display irregularities caused by excessive sensitivity to individual TCs. For $L = 1000$ km the tracks have very little structure, as regions of different climatological properties are averaged together. The optimal case, $L = 300$ km, is intermediate.

Typical scatter of the displacement components about their means can be seen in Fig. 4, which shows the historical displacements vectors within 300 km of two sample locations. Also shown are the pdfs of zonal displacements. The pdfs look rea-

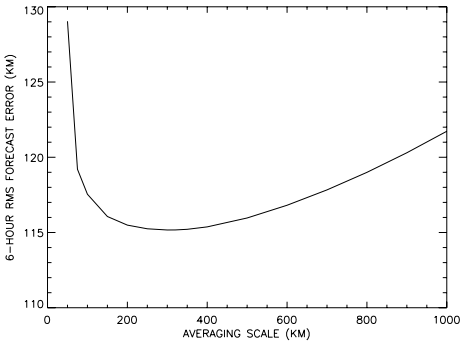

Fig. 2. Mean-track average forecast error versus averaging length-scale.

sonably normal, encouraging the use of a normal distribution to compute variances and their likelihoods, as described next.

4.2. Variance

The next step of the TC model construction is the computation of the variance of 6-h displacements about the mean. The pdf of the general multivariate normal distribution can be written

$$f(z) = \frac{1}{(2\pi)^{p/2}\sqrt{D}} \exp\left[-\frac{1}{2}(z-\mu)^T \Sigma^{-1}(z-\mu)\right], \tag{2}$$

where p is the dimensionality, Σ is the $p \times p$ covariance matrix, D is the determinant of Σ, superscript T indicates transpose and z and μ are p-length vectors, with μ being the mean. Here, $p = 2$ and $\mu = (0, 0)$ because the means are removed before modelling the variances of the two-component vector, z.

We have found that an anisotropic model, in which the variances in the u and v directions are distinct, has higher likelihood than an isotropic model. We consider $z = (u, v)$, where u and v are deviations from the mean in directions parallel and perpendicular to the mean. The covariance matrix is

$$\Sigma = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}, \tag{3}$$

so that (2) becomes

$$f = \frac{1}{2\pi\sigma_u\sigma_v} \exp\left(-\frac{u^2\sigma_v^2 + v^2\sigma_u}{2\sigma_u^2\sigma_v^2}\right). \tag{4}$$

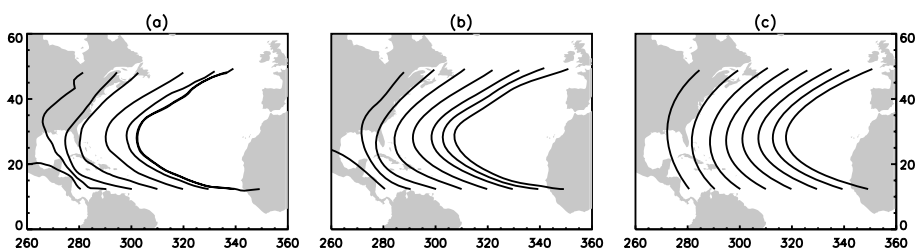

Fig. 3. Eight mean tracks originating from evenly spaced longitudes and having equal durations. The averaging scales are 100 km (left), 300 km (centre), and 1000 km (right). The optimal scale is 300 km (centre).

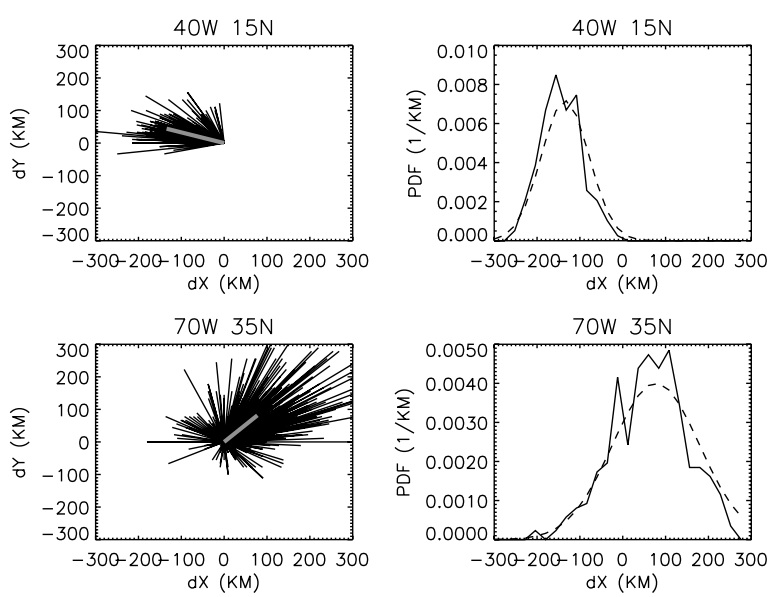

Fig. 4. Distribution of historical longitude (X) and latitude (Y) displacements at two locations, as labeled (left column, black line segments). Also shown are the mean displacements (thick grey). Shown in the right column are the corresponding distributions of longitude displacements (solid) and a fitted normal distribution (dash).

We have also examined an anisotropic correlated variance model, which, turns out to have a higher likelihood than the uncorrelated model. However, for simplicity of analysis of autocorrelation (next section) we retain the uncorrelated model for all subsequent analysis.

The coefficients σ_u and σ_v must be computed. This is done using weighted averages of deviations of historical TC displacements about the optimal mean displacement. That is, at location $\mathbf{r}$

$$\sigma_u(\mathbf{r}) = \left(\frac{\Sigma_i u_i^2 e^{-d_i^2/2L^2}}{\Sigma_i e^{-d_i^2/2L^2}} \right)^{1/2} \tag{5}$$

and similarly for $\sigma_v(\mathbf{r})$, where L is the length scale to be optimized, d_i is the great-circle distance between $\mathbf{r}$ and $\mathbf{r}_i$, and the summations are over historical storm points.

The scale L is optimized in a similar fashion similar to the mean-displacement averaging scale, but now by maximizing the likelihood rather than minimizing the forecast error. The likelihood of a historical deviation (u_i, v_i) in year j at position $\mathbf{r}$ is the distribution f evaluated at u_i and v_i. The summation in the coefficients (5) is taken over all storm points in years $k \neq j$. The log-likelihoods are summed over all historical points $\mathbf{r}$ in all years j, forming the total log-likelihood for a particular scale L for the variance model. Figure 5 shows the total log-likelihood as a function of L, as well as the analogous log-likelihood function for the isotropic variance model. The optimal scales for the anisotropic model is 300 km, identical to the scale for the mean track. Figure 6 shows the spatial distributions of the anisotropic variance components in kilometres. The variances are largest in the northern Atlantic, with values of 100–150 km. This is also the region where the mean-track propagation speed is greatest.

4.3. Autocorrelation

We now turn to the analysis and modelling of the autocorrelation, or 'memory', of the anomalies from the mean track. We first analyse the historical data to determine the nature of the autocorrelations. We then select an autocorrelation model and evaluate it by analysing its residuals.

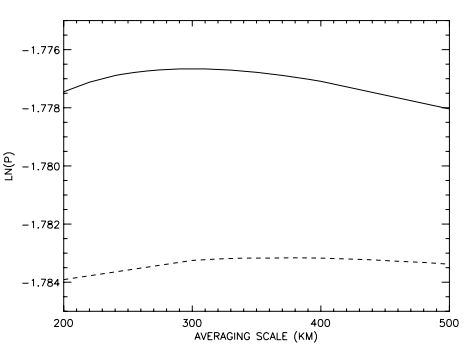

Fig. 5. Log likelihood of the variance versus spatial averaging scale for the isotropic model (dash) and the anisotropic model (solid). Values listed on the vertical axis are divided by 10^5.

4.3.1. Historical autocorrelations. We examine the autocorrelation of historical TC displacement anomalies. First, optimal means are subtracted from the historical track data, and the remainders are divided by optimal variances of the anisotropic variance model, leaving standardized anomalies. Figure 7 shows scatter plots among u and v anomalies (denoted $\tilde{u}$ and $\tilde{v}$) and their values at the previous 6-h step. The anomalies $\tilde{u}$ and $\tilde{v}$ are each strongly correlated with their values at the previous time step. On the other hand, no structure is apparent between contemporaneous $\tilde{u}$ and $\tilde{v}$, or between $\tilde{u}$ and $\tilde{v}$ with either lagged, suggesting that independent modelling of the $\tilde{u}$ and $\tilde{v}$ anomalies is appropriate. A similar conclusion was reached about Coral Sea TCs by James and Mason (2005), although these authors considered correlations among raw track increments, rather than regularized anomalies.

Our variance model assumes that the deviations about the mean are normally distributed. The 'quantile–quantile' (QQ) plot of Fig. 8 reveals the degree to which normality is maintained by the anomalies. Both $\tilde{u}$ and $\tilde{v}$ are normal within ± 2 standard deviations, while outside this range they exhibit 'fat tails;' that is, anomalies greater than ± 2 standard deviations are more common than would be expected normally.

In order to decide how many 6-h lags to consider in the autocorrelation modelling we regress all $\tilde{u}$ and $\tilde{v}$ against a range of lagged values. Figure 9 shows the resulting regression coefficients for 10 lags. The lag-one coefficients are 0.75–0.8 for $\tilde{u}$ and $\tilde{v}$, while the coefficients for all greater lags have magnitudes less than 0.05. This indicates that consideration of just one lag is sufficient to model the anomalies.

A lag-one autocorrelation function has the form $f(t) = e^{-t/\tau}$, where t is the lag and τ is the decorrelation time scale. The $\tilde{u}$ and $\tilde{v}$ time scales τ are plotted spatially in Fig. 10, using an averaging length scale of 900 km to compute the autocorrelation coefficients (justified below). There is spatial structure, with τ for $\tilde{u}$ peaking at about 2.25 d in mid-latitudes and τ for $\tilde{v}$ peaking at about 1.75 d in the subtropics.

4.3.2. Autocorrelation model. The analysis of the autocorrelations suggests the modelling of the u and v anomalies as independent lag-one autoregressive processes, denoted AR(1). Given anomalies at step $n-1$, the anomalies at step n are

$$\begin{aligned} \tilde{u}_n &= \phi_n^u \tilde{u}_n + s_n^u \epsilon^u \\ \tilde{v}_n &= \phi_n^v \tilde{v}_n + s_n^v \epsilon^v, \end{aligned} \tag{6}$$

where ϕ are the autocorrelation coefficients, s the magnitude of the innovations, and ϵ the innovation, drawn from a normal distribution of zero mean and unit variance (and, subsequently, from model residuals). Autoregressive models are also used by Vickery et al. (2000) and James and Mason (2005), though applied to the raw track increments, rather than the anomalies. Note that in (6) s and ϕ are not independent. Squaring and averaging both

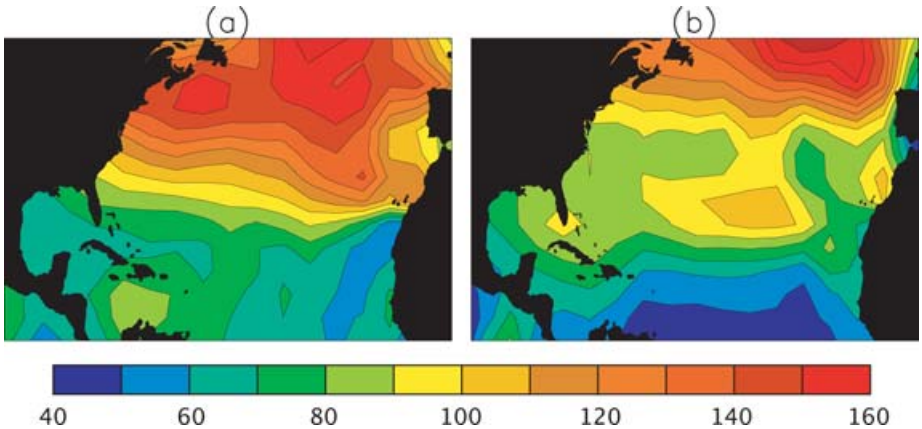

Fig. 6. The spatial distributions of the rms displacement variances parallel (a) and perpendicular (b) to the mean displacement vector. Units are kilometres. The averaging scale is the optimal value of 300 km.

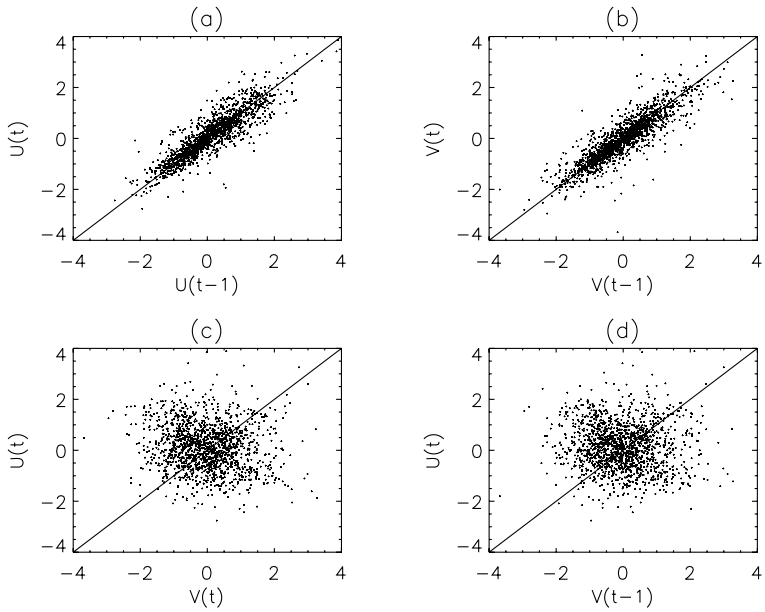

Fig. 7. Scatter plots of anomalies. u refers to direction parallel to mean displacements and v to direction perpendicular to mean displacement. (a) $u(t)$ versus $u(t - \Delta t)$, where Δt is one 6-h time step; (b) $v(t)$ versus $v(t - \Delta t)$; (c) $u(t)$ versus $v(t)$; and (d) $u(t)$ versus $v(t - \Delta t)$. Points are a random subset of the historical data set. One-to-one line is plotted for reference.

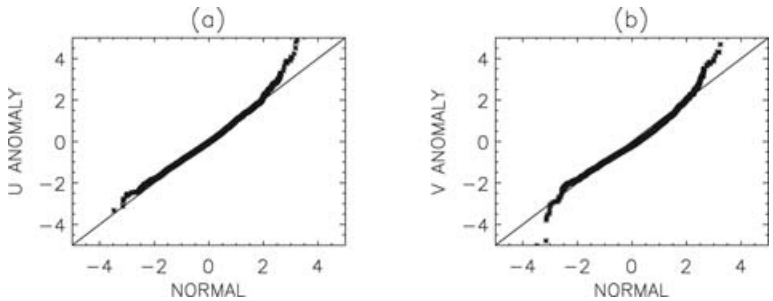

Fig. 8. Quantile–quantile plots of (a) displacement anomalies parallel and (b) perpendicular to mean displacement vectors. The anomalies would be distributed normally if the scatter fell on the straight lines. This is the case within ±2 standard deviations. Outside this, the anomalies exhibit 'fat tails.' That is, large anomalies are more probable than normal.

sides of eq. (6) and using the fact that the regularized anomalies have unit variance, one finds that $s^2 = 1 - \phi^2$.

The coefficients ϕ_u and ϕ_v at a current point $\mathbf{r}$ are computed by weighted spatial averaging of historical autocorrelations. We regress all $\tilde{u}_n$ against all $\tilde{u}_{n-1}$ anomalies ($n > 1$), using a weight $e^{-d_i^2}/2L^2$ for each historical $(\tilde{u}_n, \tilde{u}_{n-1})$ pair, where d_i is the great-circle distance between $\mathbf{r}$ and the position $\mathbf{r}_i$ of the historical pair. (An identical calculation is performed for v.) Once again the optimal averaging scale, $L = 900$ km, is obtained by maximizing the jackknife out-of-sample likelihood. To compute the likelihood, the multivariate normal distribution (2) is used to represent the probability of the entire series of $\tilde{u}$ or $\tilde{v}$ along a TC

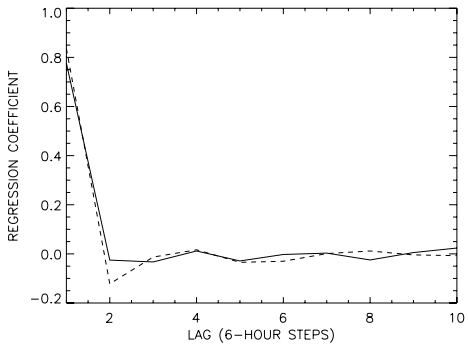

Fig. 9. Coefficients obtained regressing anomalies against themselves at 10 successive six-hourly lags in the direction parallel (solid) and perpendicular (dash) to mean displacements. In this case all historical data are used with equal weight. Only lag one is significantly different than zero, supporting the use of the AR(1) model.

track. The symmetric covariance matrix has dimensions $m \times m$, where m is the number of steps in the track. It can be written

$$\Sigma = \begin{pmatrix} 1 & \phi_1 & \phi_1\phi_2 & \phi_1\phi_2\phi_3 & \dots \\ . & 1 & \phi_1 & \phi_1\phi_2 & \dots \\ . & . & 1 & \phi_1 & \dots \\ . & . & . & 1 & \dots \\ . & . & . & . & 1 \end{pmatrix}, \tag{7}$$

where the subscripts on ϕ refer to the step number along the track. (The diagonal elements are all 1, because the anomalies have been standardized to unit variance.)

4.3.3. Residuals. To evaluate how well AR(1) models the anomalies we examine the residuals; that is, the difference of an AR(1) predicted anomaly from a historical TC point and the

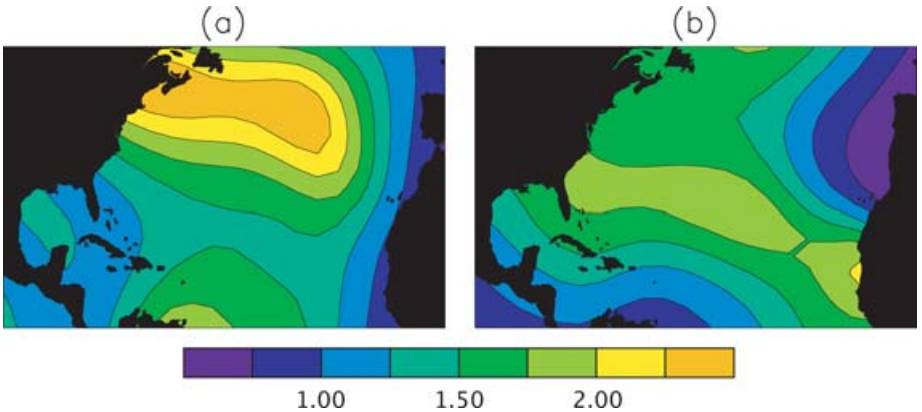

Fig. 10. Autocorrelation timescale τ for anomalies parallel (a) and perpendicular (b) to mean displacements. Units are days. For a perfect AR(1) model the autocorrelation function (ACF) is $f(t) = e^{-t/\tau}$. The timescales plotted here are $\tau = 0.25$ day/ln [f(0.25 day)], where $f(t)$ is the empirical ACF computed locally from the historical data using the optimal averaging scale of 900 km.

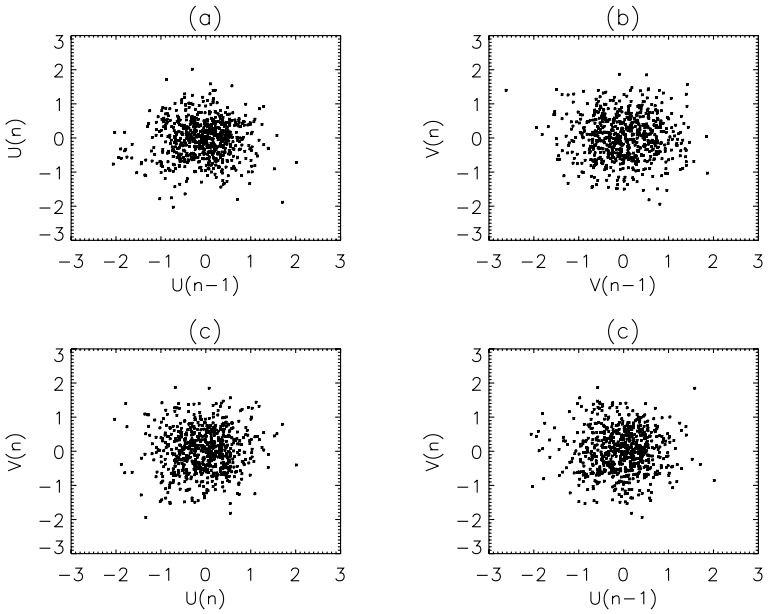

Fig. 11. As in Fig. 7, but now for the residuals of the lag-one autocorrelation model. No structure is apparent, indicating that the lag-one model captures well the u and v correlations and autocorrelations.

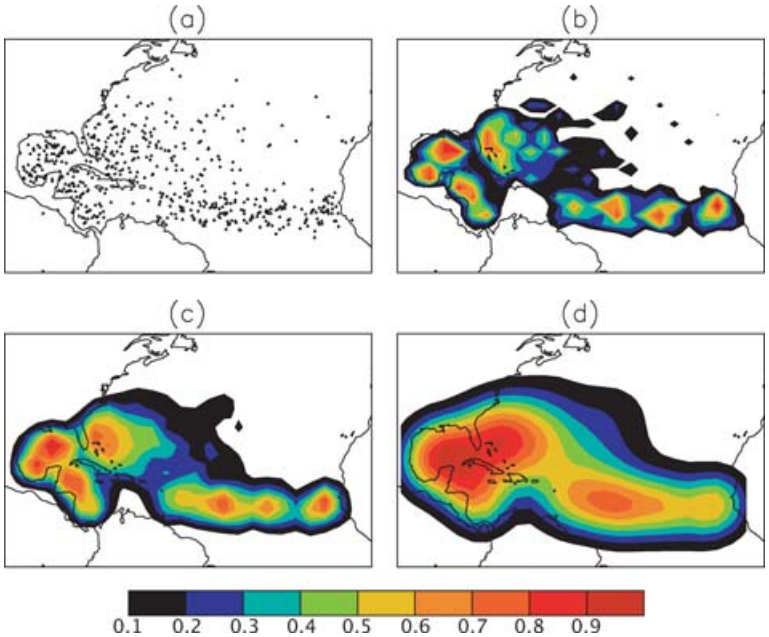

Fig. 12. (a) Historical genesis sites from 1950 to 2003. (b) Kernel genesis pdf with length-scale 100 km, (c) 210 km and (d) 500 km. The case of 210 km is optimal. The pdfs are each normalized to unit maximum.

actual historical anomaly. Figure 11 contains scatter plots analogous to Fig. 7, but now for the residuals. The residual scatter displays no structure, indicating that the relationships have been captured well by the AR(1) model. A QQ analysis of the residuals (not shown) shows behaviour similar to the QQ analysis of the anomalies (Fig. 8): normal behaviour within ±2 standard deviations, and 'fat tails' outside.

To accommodate the effects of the fat tails in track simulations we proceed as follows: At each historical TC point we simulate the $\tilde{u}$ and $\tilde{v}$ using the optimized AR(1) model with standard normal innovations. The residuals of these simulations are stored in a table. In all subsequent TC simulations the innovations consist of sampling randomly from the residual table. In this way anomalies in simulations will have larger than normal magnitude outside ±2 standard deviations. Using residuals for innovations turns out to provide a small increase in the realism of the simulated tracks, as diagnosed by the model-observation comparisons shown in Section 5, below.

4.4. Genesis.

The TC genesis model consists of two separate components: (1) simulating the number of TCs in a year, and (2) simulating the geographical sites of these TCs. Substantial effort has been spent developing forecasts for TC number at seasonal and longer leads (e.g. Gray et al., 1992), with much interest generated by increased Atlantic hurricane frequency and intensity in recent years (Webster et al., 2005). These issues are not addressed here for TC number, and we have chosen a simple expedient: random sampling from the historical number of TCs per year in the 1950–2003 period. For comparison, Vickery et al. (2000) sample a fitted negative binomial distribution. Whether or not annual TC number simulation is necessary depends on the nature of model evaluation against historical TCs. Some studies do not simulate TC number, simply generating an large arbitrary number of TCs for comparisons to historical TCs (e.g. James and Mason, 2005; Emanuel et al., 2006).

Given the number of TCs in a year, we model the genesis sites using a two-dimensional (latitude, longitude) pdf comprised of sums of kernels about out-of-sample historical genesis sites. The kernels are Gaussian, with isotropic variance length scale L, which is referred to as the bandwidth of the genesis pdf. (By contrast, Rumpf et al., 2007 use Epanichnikov kernels.) Thus, the pdf for genesis at position $\mathbf{r}$ is

$$f(\mathbf{r}) = \frac{1}{2\pi N\sigma^2} \sum_{i=1}^{N} e^{-d_i^2/2L^2}, \tag{8}$$

where d_i is great-circle distance between $\mathbf{r}$ and the location of the ith genesis site. The optimal bandwidth $L = 210$ km is obtained from the jackknife out-of-sample likelihood maximization. Figure 12 shows the 524 historical genesis sites, along with the pdf $f(\mathbf{r})$ computed using values $L = 100$ km, 210 km (optimal) and 500 km. For $L = 100$ km the pdf has detailed structure and many local maxima. For $L = 500$ km the pdf is much smoother and has just three maxima. The optimal case, $L = 210$ km, is intermediate.

To simulate genesis the pdf is normalized to unit maximum. A random value of $\mathbf{r}$ is chosen from a uniform distribution over a region that encompasses the entire domain. Genesis occurs at $\mathbf{r}$ randomly with a probability given by the normalized f. The procedure is continued until the desired number of genesis events is realized. Fig.13 shows the historical genesis sites from 1950 to 2003 (524 events) and three simulations of genesis sites over the

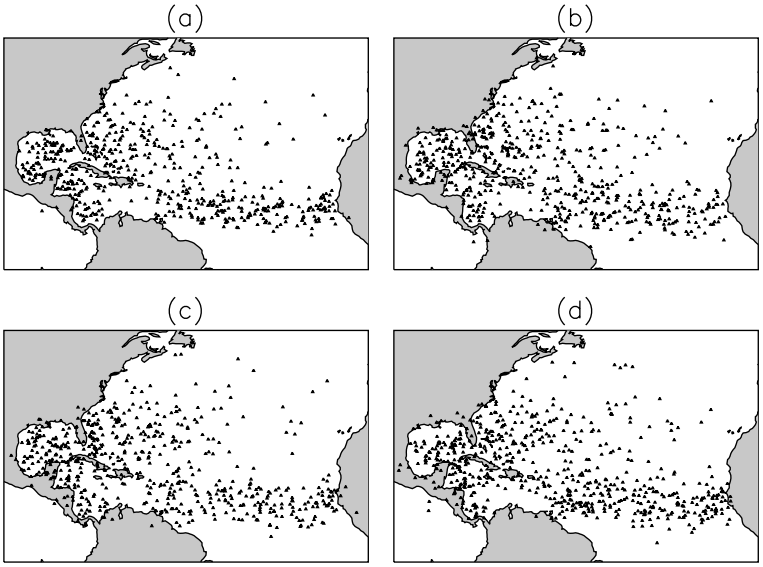

Fig. 13. (a) The 524 historical genesis sites from 1950 to 2003 used in this analysis, as in Fig. 12a. (b), (c) and (d) Three simulations of genesis for the same period. The number of simulated TCs are 516, 510 and 554, respectively.

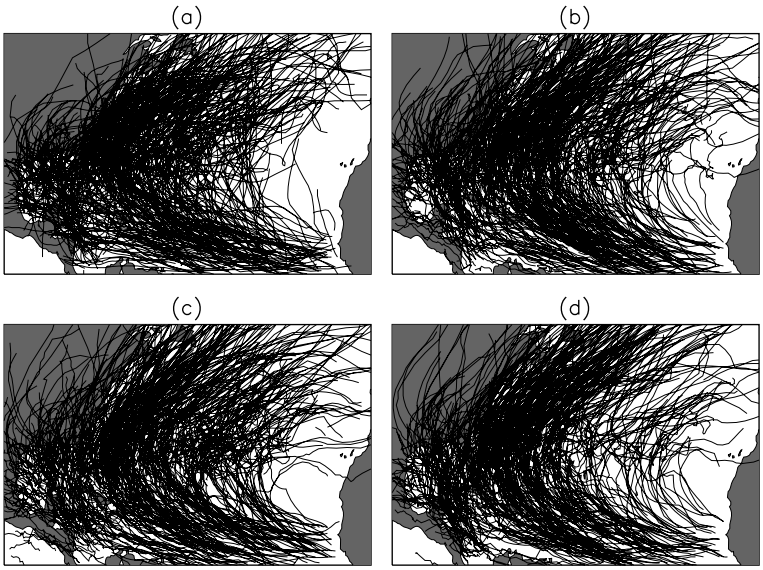

Fig. 14. (a) Historical TC tracks from 1950 to 2003 (as in Fig. 1). (b), (c) and (d) Three simulations of TC tracks over the same period.

same period (516, 510 and 554 events). The simulations capture the historical genesis pattern well. The simulated genesis events scatter over a latitude band in the eastern subtropical Atlantic that is somewhat broader than the historical distribution.

4.5. Lysis.

Ultimately the lysis of a TC should be linked to the simulated evolution of its intensity. A stochastic intensity model is currently being developed. In the meantime, we construct a lysis model from historical lysis rates. The probability $p_L(\mathbf{r})$ of suffering lysis at a current point $\mathbf{r}$ is modelled as

$$p_L(\mathbf{r}) = \frac{\Sigma_i \Theta_i e^{-d_i^2/2L^2}}{\Sigma_i e^{-d_i^2/2L^2}}, \tag{9}$$

and the probability of not suffering lysis is $p_{NL} = 1 - p_L(\mathbf{r})$, where $\Theta_i = 1$ if the storm point i is the last point of a TC, and $\Theta_i = 0$ otherwise. The optimal averaging scale is $L_L =$ 360 km, obtained again by jackknife out-of-sample likelihood maximization.

5. Simulations and diagnostics

We now present simulated TCs and compare diagnostics of the simulations to analysis of historical TCs. Figure 14 shows the historical track from 1950 to 2003 and three sets of simulations of this period. The general features of the historical storms are well captured by the model, including the northwestward motion at low latitudes, the northeastward motion at mid-latitudes, the degree of penetration into the continent, and the degree of ran-

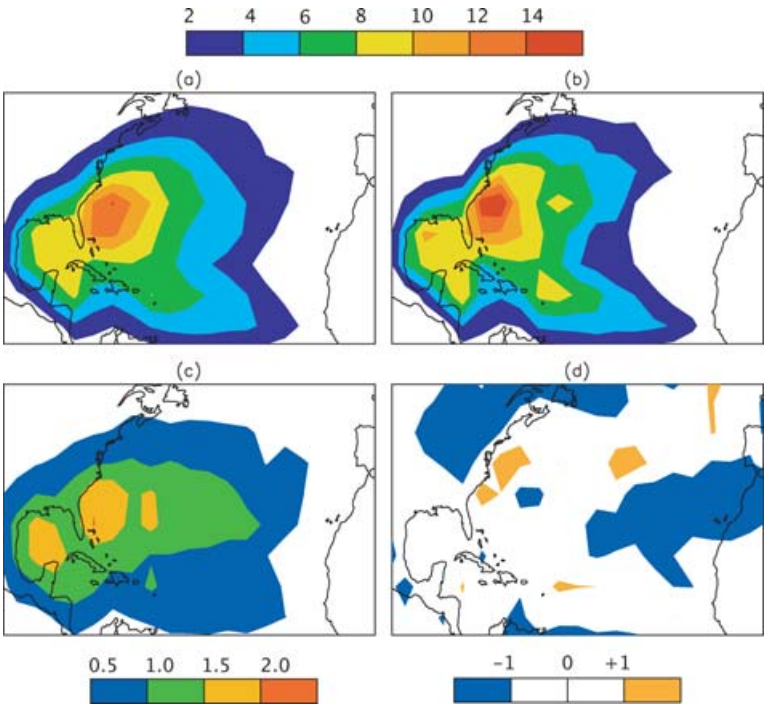

Fig. 15. TC track-point density, defined as the number of six-hourly TC positions ('points') per area accumulated over the period 1950–2003. (a) Mean density over 50-member ensemble of simulations of the 1950–2003 period. (b) Historical density over same period. (c) Root-mean-squared variance of density over 50-member ensemble. (d) 'Z score;' that is, historical minus simulated ensemble mean divided by historical. Units in (a), (b) and (c) are track points per 100 × 100 km box, while the Z-score in (d) is dimensionless.

domness in the tracks. However, there is a higher historical track concentration off the mid Atlantic coast than is simulated, and the simulated TCs tend to sweep too far east at mid-latitudes. Emanuel et al. (2006) also show examples of simulated tracks, which have features qualitatively similar to ours. However, they report no further large-scale diagnostics, making quantitative comparison difficult. Vickery et al. (2000) do not show their simulated tracks, although they evaluate the landfall characteristics of the simulated TCs.

5.1. Density.

In order to evaluate the model performance quantitatively we have devised several diagnostics. The first is the spatial density of 'storm points' (six-hourly longitude–latitude positions) computed over the 54-yr 1950–2003 period. We have simulated 50 such periods, forming an ensemble, whose average density is shown in Fig. 15a. Figure 15b shows the density for the historical TCs. Figure 15c shows the distribution of rms variance of density, computed across the ensemble. Finally, Fig. 15d shows the model's 'Z score,' the historical distribution minus the ensemble mean divided by the rms variance. The Z score is a simple test of the statistical significance of the historical-model differences. If the model were unbiased then the historical TC tracks would simply be one sample of the distribution of the simulated ensemble, and the Z score magnitude would surpass unity only infrequently. More sophisticated significance tests would be worthwhile, such as asymptotic distribution-free goodness of fit, which do not depend on normality assumptions in certain limits (e.g. Bishop and Chakraborti, 1989), but we do not pursue them here.

The basic features of the historical density are replicated by the simulations. The largest difference occurs off the US eastern seaboard just south of Cape Hatteras, where the historical density reaches a sharp maximum. The simulated TCs also have a maximum here, but it is not quite as sharp and is located further off shore. Figure 15d shows this difference to be above one standard deviation (about 1.6). Apparently, the simulations do not sufficiently 'focus' TC trajectories into the region adjacent to the mid-Atlantic eastern US coast, a discrepancy that can be seen in the tracks themselves (Fig. 14). By contrast, the simulated density is too high by more than one standard deviation in the eastern subtropical Atlantic and the American interior northeast, although both these regions have few historical or simulated TCs.

5.2. Latitude and longitude crossing.

Our second diagnostic is the number of storms crossing various latitude and longitude lines. Figure 16 shows the count of TCs crossing five latitude lines as a function of longitude. Northward and southward crossings are tallied separately. The general northward sweep of TCs can be seen, with the maximum northward penetration occurring at 30°N and about 75°E. Very few storms cross north to south. The simulated crossings are shown as the ensemble-mean rate ± one standard deviation. The spread in simulated crossing rates bounds the historical rates in most places.

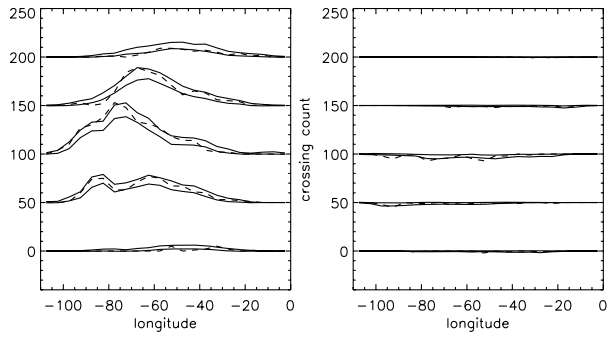

Fig. 16. Number of historical (dash) and simulated (solid) TCs over 1950–2003 period crossing latitude lines as a function of longitude. Counts are made in 5° longitude bins. The latitudes are 10°, 20°, 30°, 40° and 50°N. Northward crossings (left column) and southward crossings (right column) are tallied separately. Simulated crossings are shown as the 50-member ensemble mean ±1 standard deviation. In most regions the historical crossings fall within the variability of the simulations. Units on the vertical axis are 10-count increments, and should be taken to start at zero for each latitude.

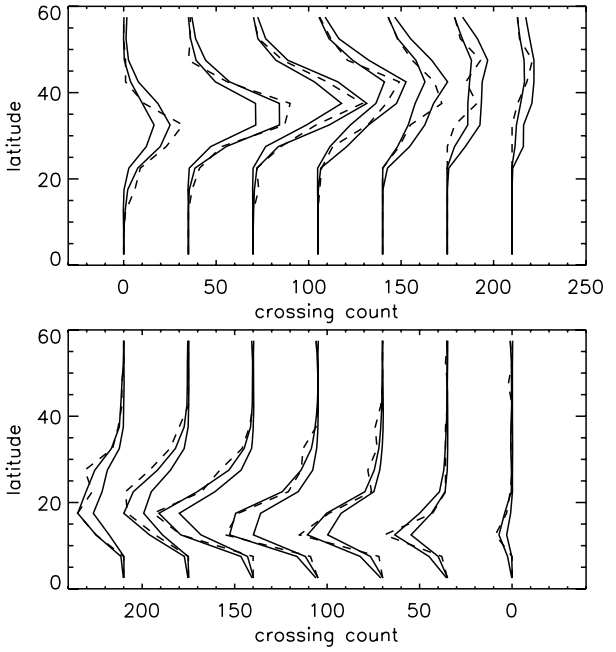

Fig. 17. Number of historical (dash) and simulated (solid) TCs over 1950–2003 period crossing longitude lines as a function of latitude. Counts are made in 5° latitude bins. The longitudes are 280°, 290°, 300°, 310°, 320°, 330° and 340°. Westward crossings (bottom) and eastward crossings (top) are tallied separately. Simulated crossings are shown as the 50-member ensemble mean ±1 standard deviation. The simulations slightly underestimate the westward crossing rate in the subtropics. Units on the horizontal axis are 10-count increments, and should be taken to start at zero for each longitude.

Figure 17 shows the westward and eastward longitude crossings. The westward penetration of TCs occurs primarily 5°N–25°N, while the eastward penetration occurs 25°–50°. In most places the simulation spread encompasses the historical crossings. In the western Atlantic, however, the simulations underestimate the westward penetration at 15°N–20°N and 25°N–30°N. This is consistent with the underestimate in TC-point density off the southern part of the US eastern seaboard (Fig. 15).

5.3. Landfall.

Landfall rates are of major interest for risk assessment. To be useful in this regard a TC model needs an intensity component, which we have not yet developed, to separate potentially catastrophic storms from minor storms. Nonetheless, for evaluating our track model, landfall rates for all named TCs taken together, is still a valuable diagnostic. Figure 18 shows historical and simulated landfall rates along the North American east coast and Gulf coast. A coarse model of the coastline is constructed using 39 line segments ('gates') of various lengths, and the number of TC displacement vectors crossing the segments from ocean to land is counted. For this tallying we treat displacement vectors independently; a single TC can make multiple landfalls. The landfall rates are plotted as a function of distance along the coarse-grained coastline from northeast to southwest. The rates are stated in units of TC crossings per year per 100 km of coastline. Shown in Fig. 18 are the coastline map and its gates, the historical landfall rate, the simulated landfall rates for the 50 ensemble members, the mean simulated rate across the ensemble and its standard deviation, and the Z score.

The historical coastal landfall-rate profile is highly structured. It is highest just south of Cape Hatteras, where it reaches 0.2 per 100 km per year, but there are several other local maxima. The landfall rate on a particular coast segment is highly sensitive to the segment's orientation, with a maximum rate occurring when the segment's normal vector is parallel to the local mean track and a minimum occurring when it is perpendicular to the mean track. Thus, the landfall-rate profile depends sensitively on the definition of the coastline gates, which are somewhat arbitrary. Nonetheless, historical regional landfall rates along the US coast shown Vickery et al. (2006) for intense hurricanes display similarly located minimums and maximums.

The simulated TC landfall rates capture much of the structure of the historical rates. The simulated peak occurs near Cape Hatteras, and the locations of the other maxima and minima match the historical profile. Because simulations are stochastic the landfall-rate profile of one 54-yr simulation differs from another. In this light, the historical profile can be viewed as only one possibility among many. The 54-yr periods prior to and following the 1950–2003 period would exhibit different landfall profiles, even in the absence of climate variability or secular change. Taken together the simulated profiles provide a sense of the range of possible landfall rates. Were the simulations unbiased, the range of simulated profiles would bound the historical

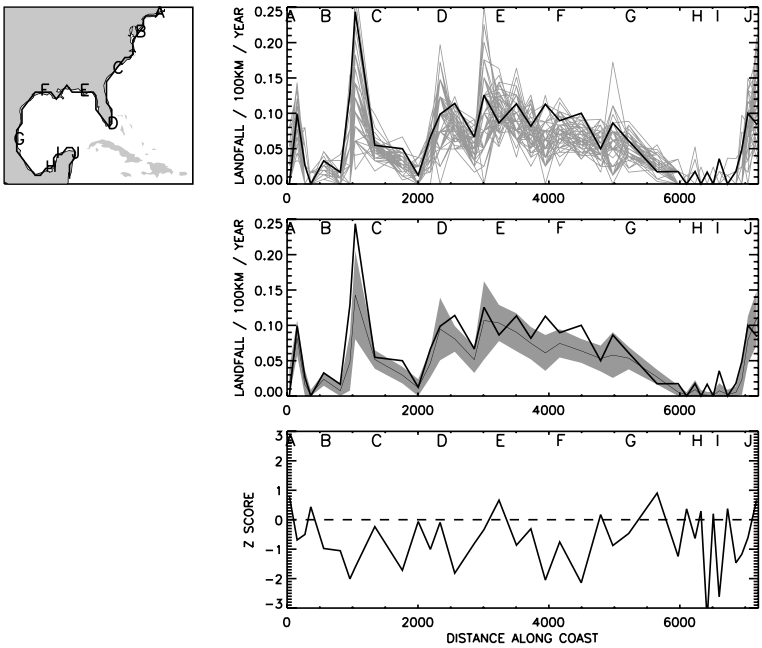

Fig. 18. Top: historical landfall rate over 1950–2003 period (thick black) and 50 simulations over same period (thin grey), both plotted as a function of distance along the coast from northeast to southwest. Rates are expressed as counts per year per 100 km of coastline, as computed over the 39 gates shown at left. Letters are shown for reference. Middle: Historical rates (thick black) and mean over 50-member simulation ensemble (thin black). The grey region represents ±1 standard deviation about the ensemble mean. Bottom: 'Z score,' that is, the simulation mean minus historical rate divided by historical rate.

profile. However, while the simulation profiles of Fig. 18 have shapes very similar to the historical, there is a negative bias in many places, which is most pronounced near Cape Hatteras, the west coast of Florida and the north Gulf Coast ($Z \approx -2$).

These landfall rates are not categorized by TC intensity, as intensity has not been simulated, and no intensity information has gone into the tracks beyond inclusion in HURDAT. Consequently, the successes and limitations of our track model to simulate landfall realistically cannot, at this point, be taken as a complete evaluation of a TC risk assessment model. When we have finished development of an intensity model (including TC wind speed) to complement the track model we will re-evaluate landfall rates at different intensity thresholds.

6. Summary

We have developed a basin-wide statistical model of North Atlantic TC tracks. The model is non-parametric, using near-neighbour historical information to propagate TCs from one 6-h time step to the next and to simulate their genesis and lysis. 'Near' is defined optimally, by maximizing the likelihood in an out-of-sample jack-knife procedure. The propagation consists of computing mean 6-h displacement increments, variances about the mean, and lag-one autocorrelations. Innovations are drawn from model residuals. The genesis model consists of sampling a kernel pdf build from historical genesis whose bandwidth (210 km) is selected to optimize the likelihood of the historical genesis. Ultimately the lysis of the simulated TCs should be based on intensity, the modelling of which is under development. At present the probability of lysis is determined from local historical lysis events.

We have used the statistical model to perform multiple simulations of the 54-yr historical period on which the model is based (1950–2003). Because the processes are stochastic, one simulation differs in detail from another. The ensemble-mean simulation exhibits large-scale features, such as track-point density, crossing rates of lines of constant latitude and longitude, and landfall rates, that match their historical counterparts reasonably well. For most diagnostics in most regions the historical quantity falls within one standard deviation of the ensemble-mean simulation, indicating that the simulations are statistically indistinguishable from the historical record. There are, however, regions of bias in the model, such as an underestimate in the track density and landfall rate in the central mid-Atlantic US coast.

There are several approaches that could be taken to further improve the realism of the tracks. We have neglected a weak correlation between anomalies perpendicular and parallel to the mean TC displacements. More generally, one should use a vector-autoregressive model to relate $(\tilde{u}, \tilde{v})$ anomaly vectors at steps n and $n-1$ by a 2×2 matrix. We have neglected any time-of-year dependence. The genesis and the track formation could be conditioned on date, for example, forming averages with a weight that declines from the date of the current point. Genesis kernels could be made one sided near coastlines to reduce negative bias from inland regions where no genesis occurs.

Finally, we emphasize that at present we have neither modelled TC intensity, nor included any track dependence on intensity. The model evaluation described here should not be considered as an evaluation of a TC risk assessment model, since risk assessments clearly require TC wind speed. Development of a statistical TC model is presently underway using non-parametric techniques similar to those described here. Our strategy is to develop

components individually, thoroughly documenting and testing each separately. We have reported on the track component in this paper.

7. Acknowledgments

This work was supported in part by NASA. The HURDAT data was obtained online courtesy of the Hurricane Research Division of NOAA's Atlantic Oceanographic and Meteorological Laboratory. We thank an anonymous reviewer for comments.

References

Bishop, J. A., Chakraborti, S. and Thistle, P. D. 1989. Asymptotically distribution-free statistical inference for generalized Lorenz curves. *Rev. Econ. Stat.* **71**, 725–727.

Casson, E. and Coles, S. 2000. Simulation and extremal analysis of hurricane events. *Appl. Stat.* **49**, 227–245.

Chu, P. and Wang, J. 1998. Modeling return periods of tropical cyclone intensities in the vicinity of Hawaii. *J. Appl. Meteorol.* **37**, 951–960.

Darling, R. 1991. Estimating probabilities of hurricane wind speeds using a large-scale empirical model. *J. Clim.* **4**, 1035–1046.

Emanuel, K. A., Ravela, S., Vivant, E. and Risi, C. 2006. A statistical-deterministic approach to hurricane risk assessment. *Bull. Amer. Meteor. Soc.* **87**, 299–312.

Gray, W. M., Landsea, C. W., Mielke, P. W. and Berry, K. J. 1992. Predicting Atlantic seasonal hurricane activity 6–11 months in advance. *Wea. Forecasting* **7**, 440–455.

Jagger, T., Elsner, J. B. and Niu, X. 2001. A dynamic probability model of hurricane winds in coastal counties of the United States. *Appl. Stat.* **40**, 853–863.

James, M. K. and Mason, L. B. 2005. Synthetic tropical cyclone database. *J. Wtrwy., Port, Coastal, and Oc. Engrg.* **131**, 181–192.

Jarvinen, B. R., Neumann, JC. J. and Davis, M. A. S. 1984. A tropical cyclone data tape for the North Atlantic Basin, 1886–1983, contents, limitations, and uses. *NOAA Tech. Memo., NWS NHC 22*.

Rumpf, J., Weindl, H., Höppe, P., Rauch, E. and Schmidt, V. 2007. Stochastic modelling of tropical cyclone tracks. *Math. Meth. Oper. Res.* **65**, in press.

Vickery, P. J., Lin, J., Skerlj, P. F., Twisdale, L. A. and Huang, K. 2006. HAZUS-MH hurricane model methodology. I: hurricane hazard, terrain, and wind load modeling. *Natural Hazards Rev.* **7**, 82–93.

Vickery, P. J., Skerlj, P. F. and Twisdale, L. A. 2000. Simulation of hurricane risk in the US using an empirical track model. *J. Struct. Engin.* **126**, 1222–1237.

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.