SUPPLEMENT

A STATISTICAL DETERMINISTIC APPROACH TO HURRICANE RISK ASSESSMENT

BY KERRY EMANUEL, SAI RAVELA, EMMANUEL VIVANT, AND CAMILLE RISI

This document is a supplement to "A Statistical Deterministic Approach to Hurricane Risk Assessment," by Kerry Emanuel, Sai Ravela, Emmanuel Vivant, and Camille Risi (*Bull. Amer. Meteor. Soc.*, **87,** 299–314) • ©2006 American Meteorological Society • *Corresponding author:* Kerry Emanuel, Room 54-1620, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139 • E-mail: emanuel@texmex.mit.edu • DOI: 10.1175/BAMS-87-3-Emanuel

Synthetic track generation using Markov chains consists of two steps: genesis and track propagation, described as follows.

Genesis. The probability of genesis is constructed as a three-dimensional field of latitude, longitude, and time. This construction begins by using a fine-resolution grid at 0.5° × 0.5° × 5 days and counting the number of events in the post-1970 hurricane database (HURDAT) within each cell. These estimates are then smoothed using a three-dimensional Gaussian kernel (Ravela and Manmatha 2001) with isotropic but varying scale in latitude and longitude and a fixed time of 5 days, with an extent of 15 days. The spatial scale of the Gaussian kernel is estimated by expanding the neighborhood around any point until N events are included (Wand and Jones 1994) or a 15° limit in latitude–longitude is reached. The genesis probability density function (PDF) is sampled to generate an event $e_0 = e_0(x_0, y_0, t_g)$, where x_0 and y_0 are the longitude and latitude of genesis, and t_g is a time window within which genesis occurs. Land points are not accorded any genesis probability. Formally, we write

$$p_g(x_i, y_i, t_i) = H_g(x_i, y_i, t_i) \otimes G(., \Sigma_i),$$

where H_g is the normalized histogram over space–time, G is the normalized three-dimensional space–time Gaussian kernal, and the operator is convolution. Variable-resolution smoothing is better than smoothing at a fixed resolution because the latter can underestimate and overestimate genesis frequency. This is in turn better than no smoothing at all, which leads to sampling discontinuities as an artifact of discretization. Figure S1 depicts examples of genesis PDFs, comparing variable resolution smoothing to fixed smoothing and to no smoothing.

Propagating tracks. Once initiated, a track is stepped forward in 6-h intervals as a Markov chain. There are several ways to integrate a track. The most obvious choice is to construct a PDF $p(x_{i+1}, y_{i+1} | x_i, y_i)$ at a chosen time window t_g, but this method suffers from two problems. First, tracks become very sensitive to resolution. Second, there is no easy way to interpolate between grid nodes because this method does not really capture the continuity in the intrinsic track manifold. The most striking feature of tracks is that they appear to be smooth. There is a strong relationship between a track's current and prior speed and direction, and there is continuity across space between these variables. We therefore begin model-

ing by using an intrinsic parameterization of track manifolds. Using differential properties to describe the intrinsic geometry of manifolds is natural, and the manner in which we do this is described in Fig. S2. In this figure, at a given location on the grid (x_i, y_i), our parameterization computes a conditional distribution of instantaneous rates of change of track speed and angle given prior speeds and angles. A sample drawn from this distribution is integrated forward to obtain a new location for the track. We do this because writing the transition probabilities in terms of their rates of change (see Fig. S2) produces conditional PDFs in a nearly resolution-independent manner, explained as follows. In Fig. S3, the conditional density of current speed and angle given their one-step priors is depicted. To sample the present speed and direction from a given prior requires sampling from the corresponding column of the left matrix shown in Fig. S3. To generate speeds that have smoothness similar to those of the observed tracks, a very high discretization of the conditional PDF becomes necessary, which is not advisable for the amount of space it requires and because there is simply insufficient data to populate the transition matrix. By modeling the conditional PDF as rates of change of speed and direction given prior speeds and directions, respectively, we are able to produce fairly coarse-resolution representations.

Also, observe that modeling in terms of prior rates of change of speed and direction (as opposed to speed and direction themselves) is not advisable because it does not capture the *regulation* inherent in tracks; a track can quickly "runaway" (diverge) when

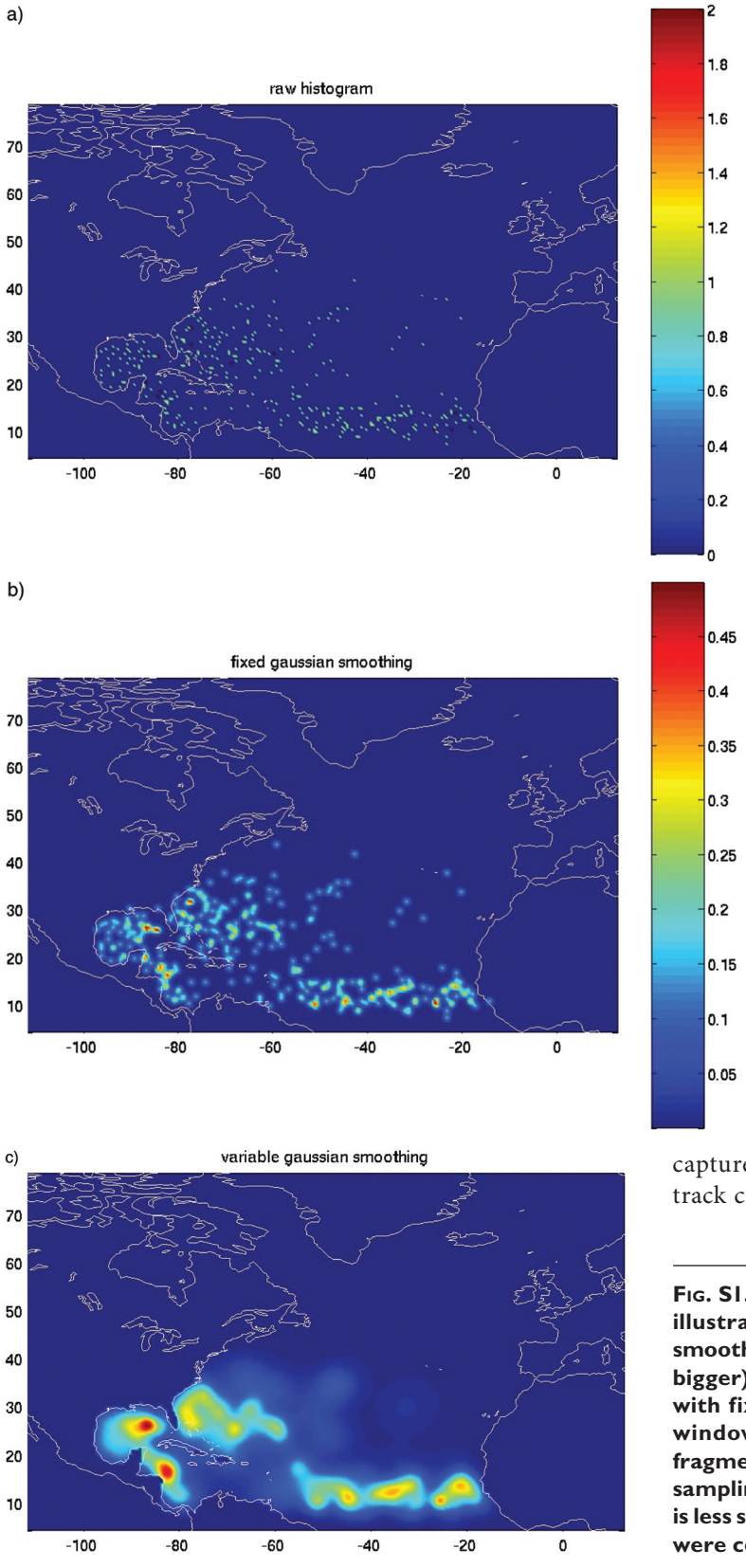

FIG. S1. The genesis PDFs for the Atlantic basin illustrate the utility of space-varying isotropic smoothing: (a) color-coded raw (warm colors are bigger) annual HURDAT genesis PDF, (b) PDF with fixed window smoothing, and (c) variable window smoothing. The unsmoothed PDF is fragmented, the fixed-window PDF still contains sampling artifacts, but the variable-window PDF is less sensitive to sampling artifacts. These PDFs were constructed using post-1970 tracks.

using prior rates of change of speed and direction. Thus, the simulation process consists of the following overall steps:

1) Compute

$$p_t(\dot{s}_i, \dot{\theta}_i | s_{i-1}, \theta_{i-1}, e_i).$$

Here, $e_i = e_i(x_i, y_i, t_g)$, s and θ are the 6-h displacement magnitude and direction, respectively, and the overdots represent time rates of change. Note that all values drawn from the conditional distribution p_t carry the same genesis time.

2) Sample from this distribution. We use a hit-or-miss method for a fixed interval of time and if no samples are available, the cumulative density function (CDF) method is used (Gentle 2004). We do this because in cases where the distributions are "broad," hit-or-miss is substantially faster than CDF.
3) Propagate the track. The new position is computed by straightforward integration from currently estimated angle and speed rates:

$$\begin{pmatrix} x_{i+1} \\ y_{i+1} \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} + \left(s_{i-1} + \dot{s}_i \delta t\right) \frac{180}{\pi a} \begin{bmatrix} \sin\left(\theta_{i-1} + \dot{\theta}_i \delta t\right) \\ \dfrac{\cos\left(\theta_{i-1} + \dot{\theta}_i \delta t\right)}{\cos\left(\dfrac{\pi y_i}{180}\right)} \end{bmatrix}.$$

Here, a is the radius of Earth, and δt is 6 h. The joint distribution in step 1 can be high-dimensional and sparse when populated using HURDAT. We enforce conditional independence to simplify the representation. This splits the joint probability to

$$p_t(\dot{s}_t, \dot{\theta}_t) = p_s(\dot{s}_i | s_{i-1}, e_i)\, p_\theta(\dot{\theta}_i | \theta_{i-1}, e_i).$$

The conditional PDFs are generated using multiresolution kernel-smoothed nonparametric density estimates produced from raw histograms (Wand and Jones 1994). Rates of change of speed and direction in the raw histograms are discretized to 8 km/6 h/6 h and 3°/6 h, respectively, and prior speed is discretized at 40 km/6 h (22 bins), and prior direction to 20° (18 bins).

A multiresolution representation of the PDFs of rates of change of speed and direction in space and time is used to sample the state variables according to a "schedule," described as follows. We do this heuristically, because one of the primary problems with a fixed space–time resolution of transition probabilities of state variables is that scarcity of HURDAT data can easily lead to dead tracks where no data are available. To avoid this, we pursue an approach that tries to gather the best possible evidence for propagating a track before letting it die. The manner

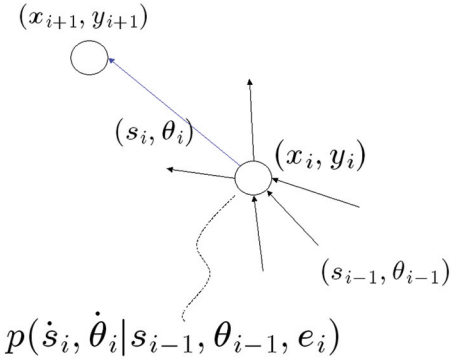

FIG. S2. The genesis event $e_0 = e_0(x_0, y_0, t_g)$ initiates a track that is propagated forward using a one-step Markov chain, conditioned by locations traversed in the (x,y) field within time window t_g.

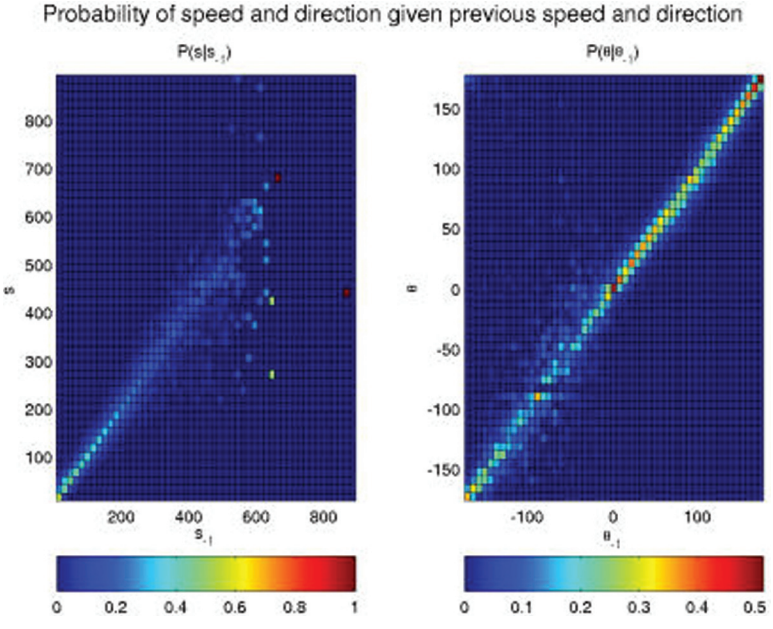

FIG. S3. There is a nearly linear relationship between current and previous states, speed (left), and direction (right). This makes a rate-based representation advantageous because it makes the transition PDFs insensitive to resolution.

Table 1. Sampling schedule.				
Space→time	0.5° × 0.5° discretization	5° × 5° discretization	Three manually constructed latitude belts	Global
Nine time periods	Priority: 1	2	3	7
One time period (annual)	4	5	6	8

in which this is done is by constructing a sampling schedule, described in the Table S1.

Transition probability density functions at any given space–time resolution are smoothed outputs of the corresponding histograms in space and state variables. As with the genesis PDFs, three-dimensional normalized Gaussians are used in *x, y,* and *z,* where *z* is the rate of change of either speed or direction. The extent in *x* and *y* is variable so as to encompass a constant number of total points. The scale (σ_z) in *z* is fixed to the state resolution that is employed and truncated to a $3\sigma_z$ extent. In more elaborate applications of this technique, the Gaussian width σ_z is determined by application of some optimization principles, but here we choose it subjectively to produce reasonable looking probability density functions. Tracks are continued until they die because of a lack of data after following the sampling schedule (which is very rare), or if the track reaches (see Fig. S4) an area with weakly observed hurricane activity in HURDAT. The intensity model provides a more principled regulation of track length, discussed in appendix C of Emanuel et al. (2006).

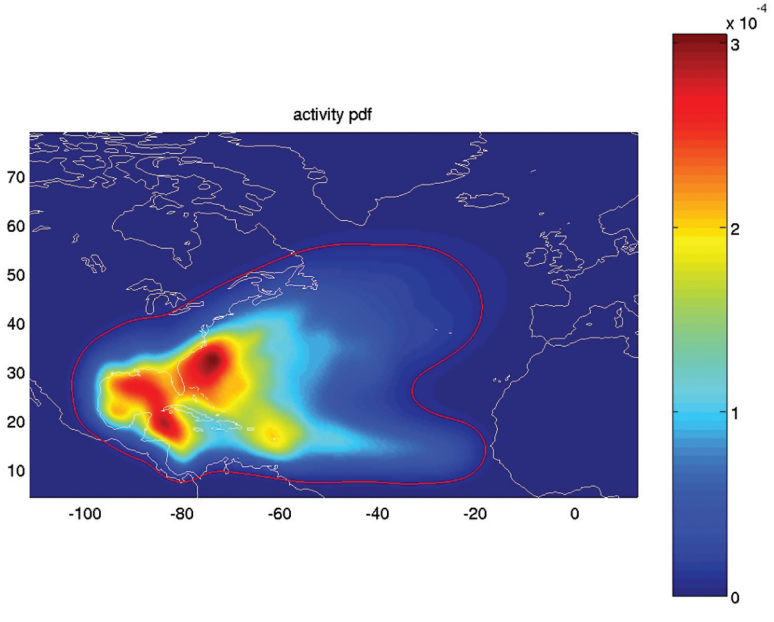

Fig. S4. The termination boundary is constructed from the marginal distribution of hurricane activity.

GENERATION OF SYNTHETIC TIME SERIES OF 250- AND 850-HPA FLOW. We begin by representing the zonal wind component at 250 hPa by its monthly mean plus a Fourier series with random phase, whose amplitude is the square root of the observed variance

$$u_{250}(x,y,\tau,t) = \overline{u}_{250}(x,y,\tau) + \sqrt{\overline{u'^2_{250}}(x,y,\tau)}F_1(t), \quad \text{(S1)}$$

where $\overline{u}_{250}(x,y,\tau)$ is the monthly mean zonal flow at 250 hPa interpolated to the date and position of the storm, $\overline{u'^2_{250}}(x,y,\tau)$ is its variance from the monthly mean, and F_1 is defined as

$$F_1 \equiv \sqrt{\frac{2}{\sum_{n=1}^{N} n^{-3}}} \sum_{n=1}^{N} n^{-3/2} \sin\left[2\pi\left(nt/T + X_{1n}\right)\right], \quad \text{(S2)}$$

where T is a time scale corresponding to the period of the lowest frequency wave in the series, N is the total number of waves retained, and X_{1n} is, for each n, a random number between 0 and 1. In (S1), τ is a slow time variable corresponding to the linearly interpolated variation of the monthly mean flow with time, while t is a fast time scale. The time series thus has the observed monthly mean and variance, while the coefficients in (S2) are chosen so that the power spectrum of the kinetic energy of the zonal flow falls off as the inverse cube of the frequency, mimicking the observed spectrum of geostrophic turbulence. We do not attempt to model the effect on the storm of higher-frequency environmental fluctuations as might, for example, be encountered in the mesoscale frequency domain, characterized by an $\omega^{-5/3}$ power spectrum.

In practice, we take $T = 15$ days and use $N = 15$. Figure S5 shows an example of such a time series, with $\bar{u}_{250} = 30$ m s^{-1} and

$$\sqrt{\overline{u'^2_{250}(x,y,\tau)}} = 10 \text{ m s}^{-1}.$$

The time series of the other flow components, $v_{250}(x,y,\tau,t)$, $u_{850}(x,y,\tau,t)$, and $v_{850}(x,y,\tau,t)$ are modeled according to

$$\begin{aligned} v_{250}(x,y,\tau,t) &= \bar{v}_{250}(x,y,\tau) + A_{21}F_1(t) + A_{22}F_2(t), \\ u_{850}(x,y,\tau,t) &= \bar{u}_{850}(x,y,\tau) + A_{31}F_1(t) + A_{32}F_2(t) + A_{33}F_3(t), \\ v_{850}(x,y,\tau,t) &= \bar{v}_{850}(x,y,\tau) + A_{41}F_1(t) + A_{42}F_2(t) + A_{43}F_3(t) + A_{44}F_4(t), \end{aligned} \tag{S3}$$

where A_{ij} are coefficients whose determination is discussed presently, and the Fs have the same form as those in (S2), but with different random phases. Thus, the different Fs are uncorrelated. We can write (S1) and (S3) in matrix form

$$\mathbf{V} = \overline{\mathbf{V}} + \boldsymbol{A}\mathbf{F}, \tag{S4}$$

where $\mathbf{V}$ is a vector containing the velocity components, $\overline{\mathbf{V}}$ is the climatological mean flow, $\mathbf{F}$ is the vector of uncorrelated time series of random phase [and amplitude of unity, as in (S2)], and $\boldsymbol{A}$ is a lower triangular matrix of coefficients that satisfies

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \mathbf{COV}, \tag{S5}$$

where **COV** is the symmetric matrix containing the variances and covariances of the flow components. In constructing the covariance matrix, we ignore any correlation between the zonal flow at 250 hPa and the meridional flow at 850 hPa, and between the meridional flow at 250 hPa and the zonal flow at 850 hPa. Because **COV** is symmetric and positive definite, the matrix $\boldsymbol{A}$ can easily be found from **COV** by Cholesky decomposition.

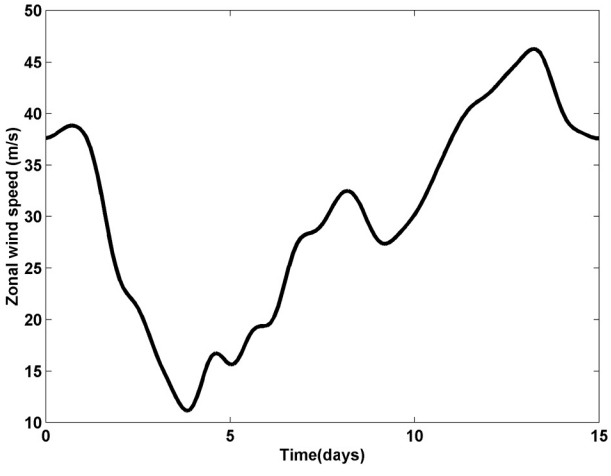

FIG. S5. Example of random time series generated using (S1) and (S2), here for the zonal wind at 250 hPa.

REFERENCES

Emanuel, K., S. Ravela, E. Vivant, and C. Risi, 2006: A statistical deterministic approach to hurricane risk assessment. *Bull. Amer. Meteor. Soc.,* **87,** 299–314.

Gentle, J. E., 2004: *Random Number Generation and Monte Carlo Methods.* 2d ed. Springer, 264 pp.

Ravela, S., and R. Manmatha, 2001: Gaussian filtered representation of images. *Wiley Encyclopedia of Electrical and Electronics Engineering,* J. G. Webster, Ed., Wiley-Interscience, 18 496 pp.

Wand, M. P., and M. C. Jones, 1994: *Kernel Smoothing.* Chapman and Hall, 224 pp.