A Bayesian Framework for Storm Tracking Using a Hidden-State Representation

LUCAS SCHARENBROICH

Department of Computer Science, University of California, Irvine, Irvine, California

GUDRUN MAGNUSDOTTIR

Department of Earth System Science, University of California, Irvine, Irvine, California

PADHRAIC SMYTH

Department of Computer Science, University of California, Irvine, Irvine, California

HAL STERN

Department of Statistics, University of California, Irvine, Irvine, California

CHIA-CHI WANG*

Department of Earth System Science, University of California, Irvine, Irvine, California

(Manuscript received 13 January 2009, in final form 22 July 2009)

ABSTRACT

A probabilistic tracking model is introduced that identifies storm tracks from feature vectors that are extracted from meteorological analysis data. The model assumes that the genesis and lysis times of each track are unknown and estimates their values along with the track's position and storm intensity over time. A hidden-state dynamics model (Kalman filter) characterizes the temporal evolution of the storms.

The model uses a Bayesian methodology for estimating the unknown lifetimes (genesis–lysis pairs) and tracks of the storms. Prior distributions are placed over the unknown parameters and their posterior distributions are estimated using a Markov Chain Monte Carlo (MCMC) sampling algorithm. The posterior distributions are used to identify and report the most likely storm tracks in the data. This approach provides a unified probabilistic framework that accounts for uncertainty in storm timing (genesis and lysis), storm location and intensity, and the feature detection process. Thus, issues such as missing observations can be accommodated in a statistical manner without human intervention.

The model is applied to the field of relative vorticity at the 975-hPa level of analysis from the National Centers for Environmental Prediction Global Forecast System during May–October 2000–02, in the tropical east Pacific. Storm tracks in the National Hurricane Center best-track data (HURDAT) for the same period are used to assess the performance of the storm identification and tracking model.

1. Introduction

The tropical east Pacific is a meteorologically intriguing area yet it is poorly studied compared to the other ocean basins. Synoptic-scale activity may be described in terms of two distinct yet related phenomena during the summer half-year: activity related to the intertropical convergence zone (ITCZ) and activity of westward-propagating disturbances (WPDs; Magnusdottir and Wang 2008). The east Pacific has the highest frequency of tropical cyclogenesis per unit area in the world (e.g., Gray 1968). Many of the WPDs develop as a result of tropical waves that have propagated from the Caribbean but may have originated as far away as Africa (e.g., Avila 1991). Other WPDs may have originated when individual vortices merged as a result of ITCZ breakdown (Kieu and Zhang

* Current affiliation: Research Center for Environmental Changes, Academia Sinica, Taipei, Taiwan.

Corresponding author address: Padhraic Smyth, Dept. of Computer Science, Bren School of Information and Computer Sciences, University of California, Irvine, Irvine, CA 92697-3435.
E-mail: smyth@ics.uci.edu

DOI: 10.1175/2009MWR2944.1

2008; Davis et al. 2008). A WPD may interact with the ITCZ and cause its partial or complete breakdown—furthermore, the WPD may intensify as a result of this interaction. While the National Hurricane Center (NHC) puts out a summary of the tracks of tropical depressions and named tropical storms every year, the track and intensity of each precursor system is not included. The current study grew out of efforts to describe the location and intensity of WPDs in the east Pacific as a function of time.

While numerous automatic tracking methods of meteorological systems exist in the literature, most have been developed with extratropical applications in mind. In many cases, these tracking methods are not well equipped to deal with the high spatial resolution that is required to describe tropical systems. For example, while the climate simulations described by Bengtsson et al. (2007) are at a spatial resolution of 0.7° by 0.7°, the vorticity field was degraded to 2.5° by 2.5° for the purposes of tracking. Additional fields were required to distinguish tropical cyclones from other less intense, but spatially more expansive, disturbances. Chauvin et al. (2006) discuss the challenges associated with detecting and tracking tropical disturbances and give a comprehensive summary of previous work on the subject. Automated tracking methods are useful for extracting tracks from large datasets in order to reduce the amount of human effort and provide an objective reference for subsequent analyses. The availability of new and higher-resolution reanalysis products, as well as climate models that are able to better resolve storms, provide an opportunity for the development of sophisticated tracking algorithms that take advantage of the improved model output to produce higher-quality and more detailed storm tracks.

In this paper a probabilistic storm tracking model is presented and applied to analysis output for the tropical east Pacific. The remainder of this introduction section provides an overview of our method and compares it to other tracking methods both in meteorology and other disciplines. Following the introduction the paper is organized as follows. The methodology and Bayesian approach to inference is described in section 2 and applied to the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) analysis in section 3. The model's performance is evaluated relative to the NHC best-track dataset. We conclude with a summary in section 4.

a. Model overview

We develop a probabilistic model that relates feature detections from meteorological fields to underlying storm tracks, and use this model to estimate the lifetimes and tracks of tropical disturbances contained in the 975-hPa[1] vorticity field from the NCEP GFS analysis. A storm track is defined in our model by a genesis (starting) time, a lysis (termination) time, and a sequence of states that describe the location and intensity of the storm over the interval on which it is defined. These unknown characteristics (the start times, termination times, and storm status or intensity) are estimated from a time sequence of an analyzed two dimensional meteorological field at a fixed vertical level. In what follows we will often refer to the two-dimensional meteorological field at time t as the *image* at time t.

We assume that we are given a sequence of gridded meteorological images and that each image is summarized by a set of feature vectors. The feature vectors are possible storm detections. In the following, each feature detection comprises the latitude, longitude, and intensity corresponding to the location of a local vorticity peak. For each storm track, and for each time t that the storm is present, at most one of the feature vectors corresponds to that storm track. The feature vectors (i.e., possible storm detections) are produced by a suitable *detector* that takes an image as input and returns a set of feature vectors. The detector will produce false positives [i.e., detections (local vorticity peaks) that do not correspond to any storm] and false negatives (i.e., instances in which a storm is present but there is no corresponding feature vector). Such errors are accommodated in the model probabilistically: there is a fixed probability of detection (and a complementary probability of a missed detection) when a storm is present.

For the 975-hPa relative vorticity field that we work with, these detections correspond to the location and value of local vorticity maxima in the Northern Hemisphere. Because of the limited spatial and temporal resolution of the images, there is inherent uncertainty in the feature vector values (e.g., spatial uncertainty due to the gridding of the data). This uncertainty is explicitly incorporated into the model via a probability model that relates observed feature vectors to the unknown state of the storm. This storm model also includes a dynamic component to model the expected motion of a storm over time, as well as additional variability to allow for storm-to-storm variation in tracks and the feature detection process. Our analysis considers all detections over the entire time period to infer storm locations and intensity. We use an offline approach in which inferences at a particular time use information from detections earlier and later in time. This is different from

[1] This level was chosen over a standard level such as 850 hPa because the relative vorticity anomaly due to the tropical disturbance was greater at 975 than at 850 hPa.

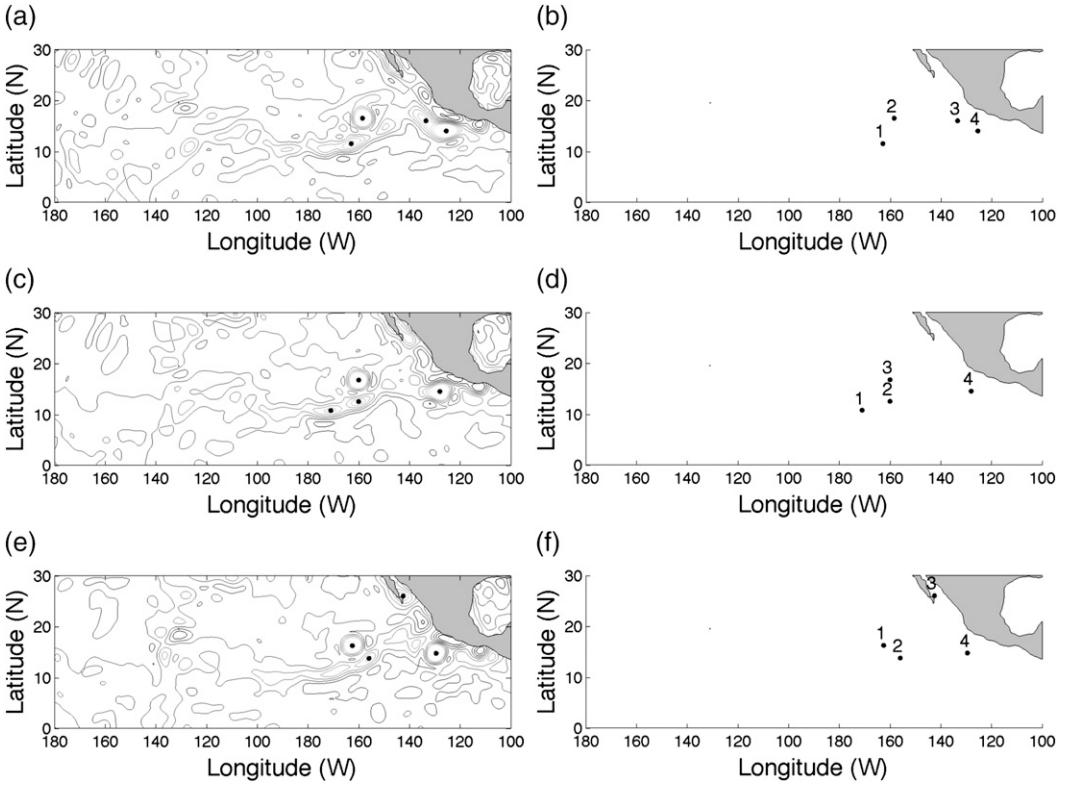

FIG. 1. Sequence of relative vorticity fields at three consecutive time steps at 0000, 0600, and 1200 UTC 5 Aug 2000. (a),(c),(e) Contour plot of the relative vorticity field along with the location of the feature detections (black dots). (b),(d),(f) Only the detections, labeled from left to right, and representing the input to the tracking model. There are two named storms Fabio (b-2, d-3, f-1) and Gilma (b-4, d-4, f-4) and two false positives in each image.

the many online algorithms that attempt to determine storm status at a given time based only on the current and earlier detections (this is discussed further next).

As an illustrative example, Fig. 1 shows a sequence of three vorticity maps (or images) at three consecutive time steps on 5 August 2000, along with the corresponding locations of the detections. The two strongest vorticity pools in each image correspond to Tropical Storm Fabio and Hurricane Gilma (Fig. 1b, detections 2 and 4, respectively). The other detections are false positives in the sense that they are vorticity maxima that do not correspond to long-lived storms. The image sequence in Fig. 1 represents a typical set of feature detections and highlights some of the difficulties involved in storm tracking. For instance, the local vorticity pools represented by detections 3 and 4 in Fig. 1b merge together into a single vorticity pool identified by detection 4 in Fig. 1d. Also, detection 1 in Fig. 1d is a spurious detection that introduces additional uncertainty in determining Fabio's optimal track. The more detections, the more possibilities there are to consider.

Some of the false detections could potentially be removed in this case by smoothing the data (e.g., with a spectral filter). However, as we show below, the current algorithm preserves the ability to track a wide range of storms. If we smooth away some false detections this will limit us to only identifying the strongest storm signatures in the sequence of images. Instead, our method relies on a probabilistic model of storm dynamics to determine which detections are relevant rather than attempt to determine this from the data a priori.

The key characteristic of our approach is a unified probabilistic framework that accounts for uncertainty in storm timing (genesis and lysis), storm location and intensity, and the feature detection process. Thus, issues such as missing observations in an otherwise strong track are easily accommodated without human intervention. The probability model is built via a series of modules that are integrated. These include the specification of prior information about storm origin, location, and length, a dynamic model for storm progression, and a probabilistic model for association of detections and storms. This makes it possible to apply the approach in other locations or other contexts by altering one or more of these components. To apply our algorithm in another location would require some prior information from area experts

on the typical storm length and storm genesis area [or the major development region (MDR)] as well as some exemplars from which a dynamic model can be developed. We view this modularity as a strong feature of our proposed approach. The probabilistic model also leads naturally to posterior inferences regarding storm location and traits that incorporate all modeled sources of variability and uncertainty. In the remainder of the introduction we describe the key computational problems in tracking and relate our work to other tracking approaches in the literature.

b. *Computational problems in tracking and related work*

There are two significant problems that must be solved in order to identify the storms from the observed feature detections. First, the *track initiation and termination* problem must be solved in order to estimate the start and termination times of the storm. Any method for solving this problem must, at a minimum, iterate over and evaluate the $T(T + 1)/2$ possible start–termination pairs in order to find the optimal one (Bar-Shalom 1987).

The second, more difficult, problem is the *data association* problem that identifies which feature vectors belong to the storm track (Rasmussen and Hager 2001; Särkkä et al. 2004). Algorithms for solving the data association problem often assume that the start and end time for a storm are known and that, for each time point in the storm's life an observed feature vector is associated with the storm track. In practice, however, it is possible that the detector fails to observe the storm at all, so that there may be no association at a particular time (a missed detection), significantly complicating the analysis.

1) TRACK INITIATION AND TERMINATION

Track initiation methods in the computer science and engineering literature are usually based on Reid's multiple hypothesis tracker (MHT; Reid 1979). The sequence of observed images, or more formally the sequence of feature detections, is scanned from beginning to end and a set of potential tracks is maintained. A track is initiated once enough evidence has accumulated to determine that a potential track is not spurious but rather represents a real object. A drawback of MHT is that the number of hypotheses can grow exponentially, so a pruning mechanism is used that limits either the total number of hypotheses under consideration or the width of the hypothesis window. Limiting the hypothesis window leads to the so-called *N*-scan algorithm and can yield significant improvements in computational speed (Cox and Hingorani 1996).

Alternative track initiation methods employ the probabilistic data association filter (PDAF), which is a "0-scan" or online method that collapses the combinatoric number of hypotheses into a single representative distribution after processing each image's detections. This distribution is used to assign probabilities to the unknown associations for the next image, which are then filtered in order to maintain a small set of track estimates (McMillan and Lim 1990).

Other track initiation techniques include rule-based methods that determine when a track begins via a set of ad hoc rules, and data transforms such as the Hough transformation (Hu et al. 1997). In some contexts, the problem can be simplified by assuming that storms always enter the detection region by transitioning across an edge. In this case, only the feature points near the boundary need to be considered as locations of potential track genesis (Chang et al. 1994).

All of the preceding track initiation techniques attempt to initiate a storm track in an *online* fashion where no information is available beyond the current time t. In contrast our model is an *offline* method that presumes access to *all* of the data (past and future) when making determinations about storm location at a particular time. Future observations are especially informative for the track initiation problem and the model can effectively incorporate this information in order to estimate likely genesis and lysis times.

2) DATA ASSOCIATION

The data association problem consists of identifying which feature vectors belong to the storm track. To illustrate its complexity consider a a storm lasting 7 days, with observations every 6 h (thus there are 28 images) and suppose there are 5 detections in each image. For this scenario there are over 6.14×10^{21} possible associations. Given the size of this space, directly computing the optimal data association sequence by iterating over the possibilities is impractical for all but the very shortest storm tracks. Most practical methods limit the number of potential associations through the use of a *gating function* that eliminates detections that violate some constraint (e.g., a weighted distance from the storm's current estimated location; Bar-Shalom 1987). Bayesian methods can also be used to provide some additional modeling flexibility in this context (Cox and Hingorani 1996).

The dynamics component of our model (described in section 2b) plays the role of a *soft-gating* function by assigning high probability to detections that are likely to be a track. Our inference method (see section 2d) leverages this information in order to focus its calculations on likely tracks. Thus, we are able to realize the computational benefits of gating without throwing away data or inadvertently committing to a suboptimal storm track.

3) JOINTLY SOLVING THE INITIATION AND ASSOCIATION PROBLEMS

The problems of track initiation and data association are intertwined. Finding the correct track initiation requires solving at least a subset of the data association problem and solving the data association problem requires that the genesis and lysis times of the storm are known. If both problems are modeled within a probabilistic framework, information is automatically and consistently shared among all the model parameters (i.e., the relative contributions of a ''good'' track initiation versus a good set of track associations is made precise). This kind of information sharing cannot be achieved if independent methods are applied to the two subproblems.

Prior nonprobabilistic approaches to solving the initiation and association problems utilize gating functions and hard constraints on certain track properties, such as storm velocity and track smoothness, in order to prune away a large amount of the search space (Hodges 1994; Gauvrit et al. 1997). These methods can produce excellent tracking results within a particular domain, but may lack flexibility because the track constraints and heuristics are often an integral part of the method. Hodges (1999) introduced an adaptive method that provides some flexibility in tracking by modifying the parameters of the objective function used to identify tracks based on a storm's location. This affords the opportunity to introduce prior knowledge of a storm's typical behavior over different geographic regions. The probabilistic framework that we describe can also in principle allow key parameters to adapt over space and time although we do not pursue this option in this paper.

There are other probabilistic methods in the literature. These methods are typically developed as ''forward only'' online target tracking algorithms (Vermaak et al. 1995; Karlsson and Gustafsson 2001; Muskulus and Jacob 2005). Even when a full probabilistic model is presented, the output of these algorithms may be a point estimate of the most likely track rather than a full posterior distribution over all possible tracks (Streit and Luginbuhl 1994; Gauvrit et al. 1997). Oh et al. (2004) propose a similar model to ours, along with a sampling strategy to explore the posterior space of data associations; however, their model does not place prior distributions over the genesis and lysis times. Also, their sampling strategy tends to be less effective when the feature vectors are spatially sparse relative to the scale of the storm dynamics, which is the case in our application.

Storlie et al. (2009) have recently (and independently) proposed a probabilistic tracking framework that is similar to our approach in that it uses a general Bayesian framework for inferring tracks, including estimation of lysis and genesis times. A major difference between our approaches is, however, that the Storlie et al. method models the genesis and lysis of storms as *local events* that occur at a rate proportional to the current number of active storms, whereas our model directly parameterizes the actual time of genesis and lysis. This allows our model to use a prior distribution to constrain the storm's lifetime in a manner that cannot be captured by a local event model. Another major difference is that the Storlie et al. inference method is based on simultaneously running multiple instances of an adapted MHT algorithm. The MHT algorithm operates in an online manner (forward pass over the data), using only past data to make inferences for each time point. In contrast to this MHT-based method, our approach can leverage both past and future data when making inferences about states and parameters for a particular time.

2. Tracking methodology

In this section, the tracking methodology is described in detail. Notation is introduced, the probabilistic storm tracking model is defined, and a sampling strategy is presented for performing Bayesian inference. The probabilistic model is developed under the assumption of tracking a single storm—we illustrate later in the experimental results section how to use the single-storm method to find tracks for multiple storms.

a. Notation and assumptions

The genesis or starting time of the storm is denoted s and the lysis or termination time is denoted τ. The state of a storm is represented by a state vector $\mathbf{x}_t \in \mathcal{R}^n$ at each discrete time $t \in [s, \tau]$, conditioned on a fixed genesis time s and lysis time τ. Let $\mathbf{X}$ represent the set of valid state vectors for the storm, $\mathbf{X} = \{\mathrm{x}_s, \ldots, \mathrm{x}_\tau\}$. The state vector $\mathbf{x}_t$ can denote for example the location and intensity of the storm at time t. In this application, $\mathbf{x}$ is a five-dimensional vector composed of the storms' position, velocity, and scalar intensity. At each time t we have a set of n_t feature vectors or detections, $\mathbf{Y}_t = \{\mathbf{y}_{t,1}, \ldots, \mathbf{y}_{t,n_t}\}$. The detections are locations and intensities of peaks in a relevant meteorological field (in our case vorticity) obtained by a separately provided detector. The collection of all the feature vector sets is denoted as $\mathbf{Y} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_T\}$. The state vector $\mathbf{x}_t$ is hidden (not observable), and it is assumed that at most one of the n_t feature vectors in each $\mathbf{Y}_t$ provides an indirect measurement of $\mathbf{x}_t$. For example, in Fig. 1b there are $n_t = 4$ feature vectors, one of which (number 2) is associated with the $\mathbf{x}_t$ state for Fabio and another (number 4) with the $\mathbf{x}_t$ state for Gilma. The other detections are assumed to be false positives, detections that do not correspond to storms of interest for our tracker. As described earlier we say ''at most'' one feature vector corresponds to a storm because we allow for the

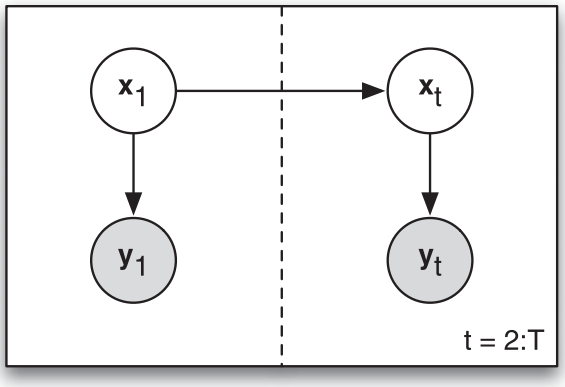

FIG. 2. A directed graphical model representing the relationship between hidden states and observed feature vectors for adjacent points in time. Each node in the graph represents a model variable. The shaded nodes represent observed variables and the unshaded nodes are unobserved (hidden) variables. There is a single feature vector $\mathbf{y}_t$ generated from a storm's hidden state $\mathbf{x}_t$. The right-hand side of the graph is replicated over time, which implies a dependency $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ for all $t > 1$.

possibility of false negatives or missed detections in which a storm is present but no detection is recorded in the appropriate location.

We assume that the dependence between state variables $\mathbf{x}_t$ is Markov over time (the state at time t depends only on the previous state) and that the observations $\mathbf{Y}_t$ are conditionally independent. The relationship between the hidden states and their corresponding observed feature vectors is presented as a directed graphical model in Fig. 2. Graphical models provide a compact representation for describing large stochastic models (Jordan 2004; Ihler et al. 2007). Each node in the graph corresponds to a variable and a directed edge establishes a direct dependency between two variables. If an edge is not present between two nodes, the variables are assumed to be conditionally independent. Shaded nodes are observed and unshaded nodes are hidden. Any portion of the graph that is enclosed within a square *plate* is replicated over the index in the lower-right corner. In Fig. 2, the right half of the plate is replicated $T - 1$ times. This implies that the Markov dependency between $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$ holds over all t from 2 to T. This conditional distribution $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is called the state transition distribution and describes how a storm's state evolves over time. The conditional distribution $p(\mathbf{y}_t|\mathbf{x}_t)$ is the *observation distribution* and defines the distribution of observed feature vectors given a known storm state.

We next introduce notation that is used to address the data association problem. Because there are multiple observations at each time t whose association with the state variable $\mathbf{x}_t$ is unknown, the model is extended to incorporate an unobserved integer association variable $q_t \in \{0, 1, \ldots, n_t\}$ that indexes into the set of n_t feature vectors and identifies the observed feature vector corresponding to the storm at time t. A nonzero value of q_t links $\mathbf{x}_t$ to feature vector $\mathbf{y}_{t,q_t}$ and a zero value indicates that none of the n_t detections correspond to the storm (i.e., a missed detection). The association variable q_t can be zero under two circumstances: either 1) a storm is present at time t but the detector did not find a corresponding feature vector, or 2) the current time t is outside of the storm's lifetime ($t \notin [s, \tau]$).

In what follows below we use the notation and assumptions above to specify a joint distribution over the unknown variables, s, τ, $\mathbf{q}$, $\mathbf{X}$, and the observed data, $\mathbf{Y}$. We then illustrate how Bayes's rule can be used to derive the posterior distribution of the unknown variables conditioned on the observed feature vectors.

b. A linear dynamic model for storm motion

We use a linear dynamical model (LDM) to characterize the dynamics $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ of the state vector $\mathbf{x}_t$. The use of a LDM is motivated by the desire to remain relatively parsimonious in terms of the statistical model, with attendant benefits in terms of both parameter estimation and computation. It should be noted that the assumed dynamics model for $\mathbf{x}_t$ is intended to provide a plausible "first order" approximation of storm motion from one image to the next, rather than a fully realistic model of storm dynamics.

The $n \times n$ matrix $\mathbf{A}$ defines the linear mapping of the storm's state at time $t - 1$ to time t. The updated state is then perturbed with zero-mean Gaussian noise with an $n \times n$ covariance matrix $\mathbf{Q}$. Because the state vector is hidden, the only information about $\mathbf{x}_t$ comes from the observed feature vector, $\mathbf{y}_t \in \mathcal{R}^m$. The feature vector $\mathbf{y}_t$ is assumed to be a linear transformation of the state vector with additional Gaussian noise:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{A}\mathbf{x}_{t-1}, \mathbf{Q}), \tag{1}$$

$$p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{C}\mathbf{x}_t, \mathbf{R}). \tag{2}$$

The $m \times n$ matrix $\mathbf{C}$ projects the hidden state onto the m-dimensional feature vector space, and $\mathbf{R}$ is an $m \times m$ covariance matrix.

Finally, a prior distribution over the initial state vector is required since the distribution $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is undefined for $t = 1$. This prior is assumed to be a Gaussian distribution,

$$p(\mathbf{x}_1) = \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \tag{3}$$

where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ specify the expected value and uncertainty of the initial state vector, $\mathbf{x}_1$ (e.g., the expected genesis location and initial intensity for storms). If the genesis time, lysis time, and associations are known, then the optimal solution for estimating each state $\mathbf{x}_t$ is given by

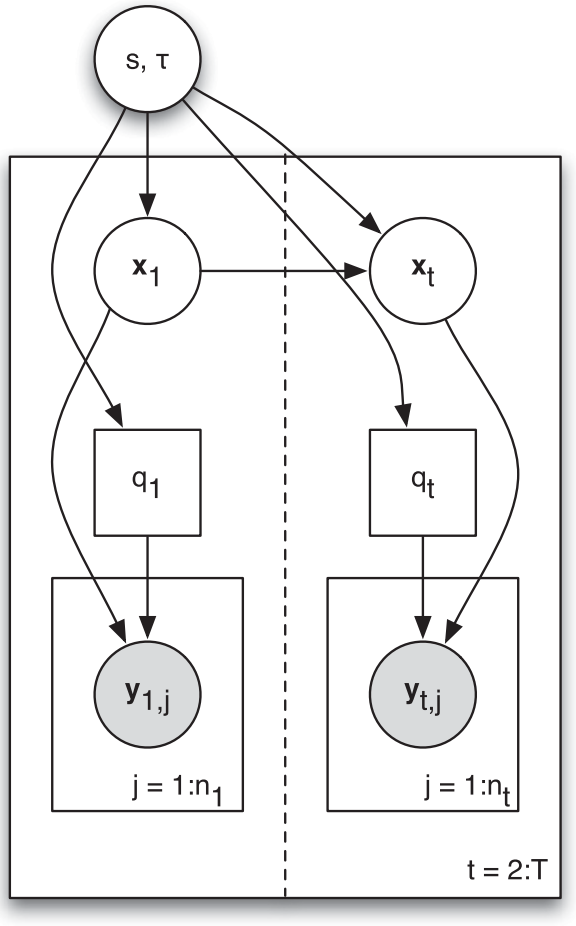

FIG. 3. A directed graphical model representation of the full tracking model. The square nodes represent discrete variables and the circular nodes are continuous. The unknown genesis and lysis times (s and τ) are shared across all times. The hidden states of the dynamics model $\mathbf{x}_t$ are coupled over time and the unknown association variable q_t links the storms with the observed feature vectors. The observed variables are shaded and the unknown variables are white.

the Rauch–Tung–Striebel smoother (Rauch et al. 1965), which has since become a standard technique in the tracking literature (Gelb 1974; West and Harrison 1997).

c. *A probabilistic model for storm association and tracking*

Figure 3 shows the full probabilistic model, which extends Fig. 2 to incorporate the storm lifetime variables, s (genesis time) and τ (lysis time), and the association variable $\mathbf{q}$. It is important to note that $\mathbf{q}$, s, and τ only identify when a storm is present and which feature vectors should be used to estimate the storm's state. They do not contain any information about the storm's state by themselves. Rather, when $\mathbf{q}$, s, and τ are known, the inference problem reduces to a standard LDM and can be solved in closed form.

TABLE 1. Table of the prior distributions over storm lifetimes. The uniform-uniform prior does not force the model to prefer any particular lifetime length, but the uniform-Poisson prior penalizes storm tracks that have a lifetime far from the mean, λ.

Storm lifetime priors		
Uniform-uniform prior	$p(s,\tau) \propto \begin{cases} \frac{1}{T}\cdot\frac{1}{u-l} & \text{for } s=1\ldots T, \text{ and } \tau \text{ s.t. } l \le \lvert\tau - s\rvert \le u \\ 0 & \text{otherwise} \end{cases}$	
Uniform-Poisson prior	$p(s,\tau) \propto \begin{cases} \frac{1}{T}\cdot\frac{\lambda^{(\tau-s)}}{(\tau-s)!}e^{-\lambda} & \text{if } s \le \tau < T \\ 0 & \text{otherwise} \end{cases}$	

Fundamentally, the *joint distribution* $p(\mathbf{Y}, \mathbf{X}, \mathbf{q}, s, \tau)$ is of interest because any conditional or marginal distribution of $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{q}$, s, and τ can be computed from it. We are typically interested in computing distributions of the unobserved variables conditioned on the observed feature vectors, $\mathbf{Y}$. Figure 3 implies that the joint distribution can be factored as

$$p(\mathbf{Y}, \mathbf{X}, \mathbf{q}, s, \tau) = p(s, \tau)p(\mathbf{X}|s, \tau)p(\mathbf{q}|s, \tau)p(\mathbf{Y}|\mathbf{X}, \mathbf{q}).$$

Thus, only the conditional distributions above need to be defined in order to fully specify the joint distribution.

The prior distribution over the storm's lifetime can be factored into two components corresponding to the genesis time of a storm and its duration, $p(s, \tau) = p(s)p(\tau|s)$. The choice of $p(\tau|s)$ can have a significant effect on the tracks found by the model. A broad, relatively flat distribution will not give any preference to tracks of a particular length and will permit the model to continue extending a track as long as there are feature detections present that follow the LDM. In contrast, a peaked distribution for $p(\tau|s)$ will force the model to try and find tracks with a duration that lies within a narrow range. Table 1 defines the two specific prior distributions that we evaluated in our experiments in section 3.

If the genesis and lysis times are known, a conditional prior distribution, $p(\mathbf{q}|s, \tau)$, can be defined over the association variables because $\mathbf{q}$ depends only on s and τ. Assuming conditional independence, $p(\mathbf{q}|s, \tau)$ can be factored as $p(\mathbf{q}|s, \tau) = \Pi_t p(q_t|s, \tau)$, where $p(q_t|s, \tau)$ is proportional to a function of the probability of detection, p_d. We make the simplifying assumption that the likelihood ratio between any nonzero and zero value of q_t is equal to the detection/nondetection ratio:

$$\frac{p(q_t = j|s,\tau)}{p(q_t = 0|s,\tau)} = \frac{p_d}{1-p_d} \quad \text{for all } j \text{ and for } t \text{ in } [s,\tau]. \qquad (4)$$

The least constrained distribution that satisfies both the above ratio and the constraint that $q_t = 0$ when a storm is not present is given by

$$p(q_t|s,\tau)=\begin{cases}\begin{cases}1 & q_t=0\\ 0 & \text{otherwise}\end{cases} & t\notin[s,\tau]\\ \begin{cases}\dfrac{1-p_d}{1+(n_t-1)p_d} & q_t=0\\ \dfrac{p_d}{1+(n_t-1)p_d} & \text{otherwise}\end{cases} & t\in[s,\tau].\end{cases} \tag{5}$$

This distribution also has the property that $p(q_t|s,\tau)$ is uniform over all nonzero values of q_t [which is implied by Eq. (4)].

The conditional distribution of the storm's state vector, $\mathbf{x}_t|\mathbf{x}_{t-1}$, also depends on s and τ. The genesis and lysis times are used to define a translation of the basic LDM to a starting time equal to s:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1},s,\tau)=\begin{cases}p(\mathbf{x}_t|\mathbf{x}_{t-1},s,\tau)=\mathcal{N}(\mathbf{x}_t;\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0), & t=s\\ p(\mathbf{x}_t|\mathbf{x}_{t-1},s,\tau)=\mathcal{N}(\mathbf{x}_t;\mathbf{A}\mathbf{x}_{t-1},\mathbf{Q}), & s<t\le\tau\end{cases}\quad t\in[s,\tau]\,.$$

and where in appendix A we show that the values of the state vector $\mathbf{x}_t$ for $t\notin[s,\tau]$ can be ignored.

Finally, we define the distribution over the feature vectors $p(\mathbf{Y}_t|\mathbf{x}_t,q_t)$. Conditional independence is assumed among the elements of the feature vector set, $\mathbf{Y}_t$, and each $\mathbf{y}_{t,j}$ is modeled as a draw from one of two probability distributions depending on whether the feature vector is associated with a storm or not:

$$p(\mathbf{Y}_t|\mathbf{x}_t,q_t)=\prod_{j=1}^{n_t}p(\mathbf{y}_{t,j}|\mathbf{x}_t,q_t), \tag{6}$$

where

$$p(\mathbf{y}_{t,j}|\mathbf{x}_t,q_t)=\begin{cases}\mathcal{N}(\mathbf{y}_t;\mathbf{C}\mathbf{x}_t,\mathbf{R}) & q_t=j\\ g(\mathbf{y}_{t,j}) & \text{otherwise}\end{cases} \tag{7}$$

and $g(\mathbf{y}_{t,j})$ is the distribution of the false positive feature vectors, which we model as a uniform distribution over the domain of $\mathbf{y}$.

With all of the conditional distributions in the graphical model defined, the joint distribution, $p(\mathbf{Y},\mathbf{X},\mathbf{q},s,\tau)$, may be written as

$$p(\mathbf{Y},\mathbf{X},\mathbf{q},s,\tau)=p(s,\tau)\prod_{t=1}^{T}p(\mathbf{Y}_t,q_t,\mathbf{x}_t|\mathbf{x}_{t-1},s,\tau), \tag{8}$$

$$=\underbrace{p(s,\tau)}_{\text{lifetime}}\prod_{t=1}^{T}\underbrace{p(\mathbf{Y}_t|\mathbf{x}_t,q_t)}_{\text{observed data}}\underbrace{p(\mathbf{x}_t|\mathbf{x}_{t-1},s,\tau)}_{\text{dynamics model}}\underbrace{p(q_t|s,\tau)}_{\text{data association}}, \tag{9}$$

$$=\underbrace{p(s,\tau)}_{\text{lifetime}}\prod_{t=1}^{T}\prod_{j=1}^{n_t}\underbrace{p(\mathbf{y}_{t,j}|\mathbf{x}_t,q_t)}_{\text{observed data}}\underbrace{p(\mathbf{x}_t|\mathbf{x}_{t-1},s,\tau)}_{\text{dynamics model}}\underbrace{p(q_t|s,\tau)}_{\text{data association}}. \tag{10}$$

d. Performing inference with the model

To extract a track in a Bayesian manner, the posterior distribution $p(\mathbf{X},\mathbf{q},s,\tau|\mathbf{Y})$ must be computed (i.e., the distribution over variables of interest given the data $\mathbf{Y}$). For notational convenience, the unknown model parameters are grouped into a single parameter vector $\theta=\{\mathbf{X},\mathbf{q},s,\tau\}$ and parameter subscripts on θ denote the subvector containing that parameter (i.e., θ_s represents the unknown parameter s). Consequently, $p(\theta_s)$ and $p(s)$ refer to the same distribution.

Invoking Bayes rule, the posterior distribution may be written as

$$p(\theta|\mathbf{Y})\propto p(\mathbf{Y}|\theta)p(\theta). \tag{11}$$

If this distribution can be computed (or approximated), then any posterior quantity of interest may be derived, such as posterior marginal distributions [e.g., $p(\mathbf{X}|\mathbf{Y})$, or modes (maxima) of the posterior distribution]. The posterior modes correspond to likely storm tracks under the dynamics model, and, once these modes are found, the corresponding tracks may be ranked according to their

likelihood. Because the posterior distribution cannot be computed in closed form, we use numerical techniques to find the posterior modes. Developing a practically efficient method to perform this estimation is addressed in the remainder of this section.

The true posterior distribution is approximated by drawing samples from the posterior distribution using the Metropolis–Hastings (M–H) sampler (Gelman et al. 2003) and then computing statistics of interest from the set of samples. The M–H algorithm (algorithm 1) defines a procedure for drawing samples from an arbitrary distribution through the use of a *proposal distribution*, $p(\theta^*|\theta)$, which suggests a new set of parameter values θ^* given a current set θ. Desirable properties of a proposal distribution are such that (a) new parameter vectors can be quickly generated, (b) the density $p(\theta^*|\theta)$ is easily computable up to a constant that does not depend on θ, and (c) the forward and reverse proposal distributions, $p(\theta^*|\theta)$ and $p(\theta|\theta^*)$, should not be too imbalanced.

ALGORITHM 1. The Metropolis–Hastings sampling procedure.

1. **for** $i = 1$ to n **do**
2. $\theta^* \sim p(\theta^*|\theta)$
3. $r \Leftarrow \dfrac{p(\theta^*|\mathbf{Y})p(\theta|\theta^*)}{p(\theta|\mathbf{Y})p(\theta^*|\theta)}$
4. $u \sim \mathcal{U}(0, 1)$
5. **if** $u < \min(r, 1)$ **then**
6. $\theta \Leftarrow \theta^*$
7. **end if**
8. **end for**

The proposal distribution we use in this paper is defined as a mixture of two proposal distributions that allow the sampler to efficiently explore a local mode of the posterior distribution. The two proposal distributions are *local proposal distributions* on s and τ that draw either a new genesis or lysis times, s^* and τ^*, respectively, using a discrete triangle distribution centered on the current genesis or lysis time. The triangle distribution is truncated to enforce the boundary conditions, $s^* \le \tau^*$, $s^* > 0$, and $\tau^* \le T$. Once s^* or τ^* is drawn, a data association proposal is drawn using the method described in Bergman and Doucet (2000) according to the following distribution:

$$p(\mathbf{q}^*|s^*, \tau^*, \mathbf{q}, s, \tau, \mathbf{Y}) = \begin{cases} p(\mathbf{q}^*|s^*, \tau^*, \mathbf{q}, s, \tau, \mathbf{Y}) & s^* = s \\ p(\mathbf{q}^*|s^*, \tau^*, \mathbf{q} = 0, s, \tau, \mathbf{Y}) & \text{otherwise.} \end{cases} \quad (12)$$

The distribution in Eq. (12) specifies separate distributions for the case in which the starting position does not change and the case in which the starting distribution does change. In the former case we use information in the current association $\mathbf{q}$ when generating a new proposal $\mathbf{q}^*$ while in the latter we do not. This is done in order to balance the likelihood ratio between the forward and reverse proposal distributions. An imbalance happens when $\mathbf{q}^*$ represents a storm track that is disjoint from $\mathbf{q}$. Evaluating $p(\theta|\theta^*)$ will result in a very low likelihood because the likelihood term at each time t (dropping the s and τ parameters for clarity) is conditioned on the elements of $\mathbf{q}$ from time 1 to $t - 1$ and the elements of $\mathbf{q}^*$ from time $t + 1$ to T:

$$p(\theta|\theta^*) = \prod_t p(\mathbf{q}_t|\mathbf{q}_{1:t-1}, \mathbf{q}^*_{t+1,T}, \mathbf{Y}).$$

Consider a worst case scenario where $\mathbf{q}$ represents a storm track that moves due west with a genesis point $\mathcal{P}$ at time t. Let the proposed association vector $\mathbf{q}^*$ represent a storm that starts at the same point $\mathcal{P}$ at time $t + 1$, but moves due east. Also, assume the likelihood that the tracks are identical [i.e., $p(\theta|\mathbf{Y}) = p(\theta^*|\mathbf{Y})$]. When we evaluate the proposal distribution at time t, the likelihood of a particular association q_t is conditioned on a set of associations from the westerly track up to time $t - 1$ and the easterly track from time $t + 1$ to T. These probabilities can become very small since the probability of a track making a large, discontinuous jump at time t is unlikely. To avoid this situation, we set $\mathbf{q} = 0$ before drawing a proposal where $s \neq s^*$. This ensures that the reverse proposal likelihood is not penalized for sampling a different track. If $s = s^*$, $\mathbf{q}$ and $\mathbf{q}^*$ are assumed to represent the same track.

Finally, one issue common to all MCMC implementations is the problem of assessing whether or not a Markov Chain has converged to the true posterior distribution after drawing M samples. If the sampler is initialized in a region of low posterior probability, then it may take many iterations for the sampler to accept enough proposals to converge in distribution to the posterior. For this reason, the first N of M samples are usually discarded as a ''burn-in'' phase in order to ensure that all the samples are drawn from the true posterior distribution (Liu 2002; Gilks 1995). Determining a precise value for N is difficult in the general case. In practice, an approximate convergence metric developed by Gelman and Rubin (Brooks and Gelman 1998; Gelman and Rubin 1992) may be used to estimate when a chain has likely converged by running multiple chains simultaneously.

3. Experiments

We demonstrate the effectiveness of our model and methodology by applying it to WPDs in the eastern

Pacific, using the 975-hPa relative vorticity field from NCEP GFS. Isolating WPDs can help identify possible interactions with ITCZ breakdown since roughly half of the ITCZ breakdown events are triggered by interactions with strong WPDs (Wang and Magnusdottir 2006). This is a challenging tracking problem due to the zonally elongated vorticity structures within the ITCZ itself, which gives rise to numerous vorticity pools. These vorticity pools are coherent structures and could reasonably be considered storm tracks in their own right. We depend on the LDM component of our model to discern between the ITCZ vorticity pools and those that represent strong WPDs. We trained our dynamics model on 10 hand-labeled WPD tracks from the year 2000 that were estimated from the Geostationary Operational Environmental Satellite (GOES) visible and GOES IR images—details are provided in appendix B. We performed the training on satellite imagery rather than the NCEP GFS data for two reasons. First, and most importantly, it is much easier for the meteorologist to track the storms using GOES imagery than vorticity. Second, by training the dynamics model on GOES data and applying it to NCEP GFS data, we demonstrate that the trained dynamics model is transferrable to different datasets.

We sequentially extract tracks from the full season of feature vectors by initializing a single-storm model at an approximate local posterior maximum and then drawing posterior samples using the Metropolis–Hastings MCMC sampling algorithm with a local proposal distribution as described in section 2. The mode of the posterior samples is used as an estimate of a storm track and its associated feature vectors are then removed from the dataset. This process is repeated until 100 storm tracks have been extracted. The mode is reported because it represents the single most likely track and is more interpretable for this model than other posterior summary statistics.

a. Feature vector detection

There is considerable freedom in selecting a method to detect a set of feature vectors from a series of raw meteorological images. We use a simple method that works well for WPD detection. More sophisticated methods, such as Hodges's multiresolution approach (Hodges 1994), the watershed segmentation method (Muskulus and Jacob 2005), or the resolution-dependent method of Walsh et al. (2007) could instead be used if desired.

First, the relative vorticity image is smoothed using a 7×7 isotropic Gaussian filter with a standard deviation of 2°. Then the smoothed image is converted into a binary image by setting any pixel with an intensity greater than the threshold of 5×10^{-5} s^{-1} to 1. The thresholding level can have a significant effect on the number of false positives and false negatives of the feature set. If the threshold is too high, then many local maxima may be missed because they fall below the threshold level. This can lead to a large number of false negatives and, consequently, a low probability of detection. If the threshold is too low, then many maxima may be found that are simply background fluctuations. In this case the detector will have a high false positive rate and, again, a low probability of detection. A useful threshold strikes a balance between these two extremes (Gelfand et al. 1996).

The next step in our detection method is to apply connected component analysis (CCA) to the binary image using an 8-point neighborhood function to find locally connected regions (Haralick and Shapiro 1992; Tarjan 1975). Finally, all the local maxima within each connected component are found by identifying the pixels, p, such that the relative vorticity $I_t(p)$ is greater than all of the neighboring pixels. Once all the local maxima are identified, a set of feature vectors $\mathbf{y}_{t,j}$ are extracted by recording the positions of the maxima and the corresponding magnitudes of the relative vorticity field.

Although the position of each maximum is constrained by the gridding of the data, the tracking model is resistant to any ''stair-stepping'' effect induced by a coarse grid. The use of an LDM, with a continuous representation of the state vector, automatically produces smoothed tracks from noisy or grid-aligned feature detections. This eliminates the need to spend additional effort improving the feature detection accuracy. Uncertainty in the true location of a storm due to gridding is absorbed into the variance terms of the linear dynamic model.

b. Parameter initialization

It is important to initialize the model parameters close to likely storms within the dataset to encourage rapid convergence to plausible tracks. It is infeasible to compute the true modes of the posterior marginal distribution $p(s, \tau|\mathbf{Y}) = \sum_{\mathbf{q}} p(s, \tau, \mathbf{q}|\mathbf{Y})$ due to the combinatoric number of association vectors. Instead, we compute an approximation of $p(s, \tau|\mathbf{Y})$ by optimizing the posterior distribution over $\mathbf{q}$ using the expectation-maximization (EM) algorithm:

$$\hat{p}(s, \tau|\mathbf{Y}) = \max_{\mathbf{q}} p(\mathbf{Y}|\mathbf{q}, s, \tau)p(\mathbf{q}, s, \tau).$$

This optimization is tractable and can be computed in time $O(kTNd^3)$, where k is the number of EM iterations,

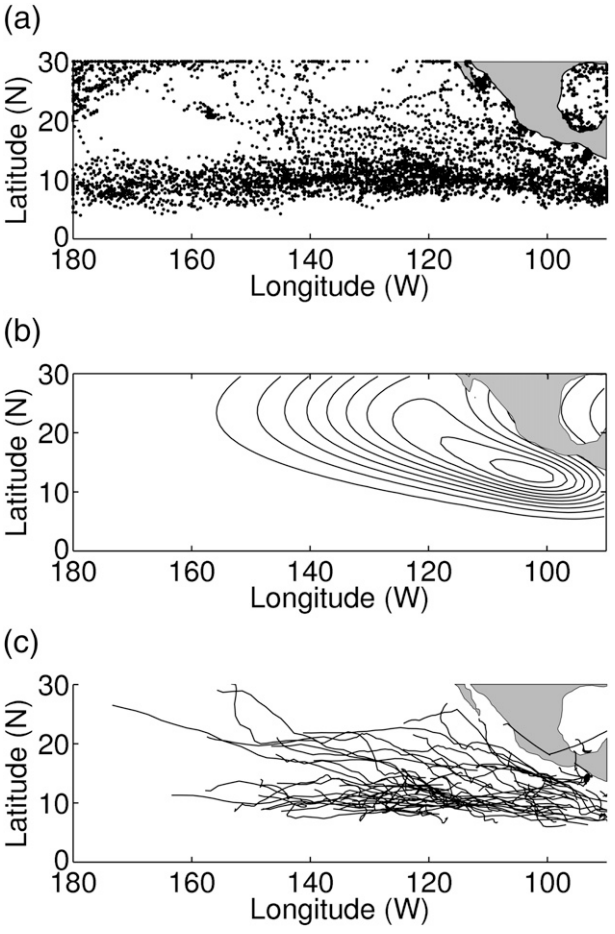

FIG. 4. The model takes the set of feature detections (a) as input. These detections were extracted from the 2001 NCEP data after segmenting the image and identifying all the local maxima. Each black dot is a feature detection. The tracking model uses a LDM in order to identify a specific class of storm tracks. (b) The expected location of the storm tracks under the chosen LDM and (c) the final set of tracks found by the model after combining the feature detections with the assumed dynamics.

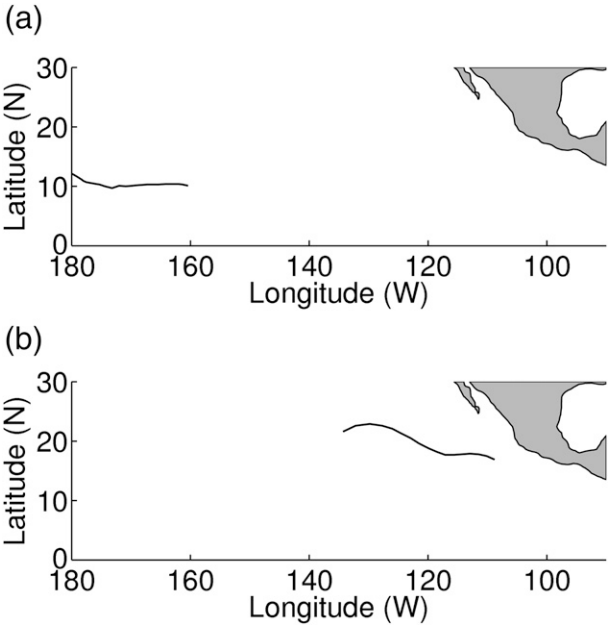

FIG. 5. Three cases were observed where the model failed to find a track corresponding to a NHC best track. The three cases could be classified into two types of tracking failures. (a) In the first type, the NHC track has a genesis too far west to be picked up by the dynamics model. (b) In the second type, the early part of the track did not follow the assumed dynamics well enough to be selected by the initialization method.

T is the number of time steps, N is the expected number of detections in each frame, and d is the dimensionality of the hidden state vector $\mathbf{x}$. Computational complexity is further reduced by only computing $\hat{p}(s, \tau|\mathbf{Y})$ for a fixed storm duration $\delta = \tau - s$. After evaluating $\hat{p}(s = t, \tau = t + \delta|\mathbf{Y})$ for $t = 1 \ldots T - \delta$, we select $t^* = \operatorname{argmax}_t \hat{p}(s, \tau|\mathbf{Y})$ and initialize the sampler at

$$\theta = (s = t^*, \tau = t^* + \delta, \mathbf{q} = 0).$$

It is important to initialize $\mathbf{q}$ to 0 in order to prevent the sampler from becoming stuck at a poor local maximum. When all the associations are unassigned, the model is free to sample a likely $\mathbf{q}$ for the given s and τ.

c. Model assumptions

The storm lifetime prior, $p(s, \tau)$, is set to be proportional to a uniform-uniform distribution over s and τ (see Table 1) with a limited range of 3 to 21 days. We also evaluated the uniform-Poisson distribution (results not shown) but found it to be less effective in terms of the overall quality of tracks found as measured by the fractional overlap of NHC tracks with the tracks found by the model. The distribution of the false positive feature vectors, $g(\mathbf{y}_{t,j})$, is also defined to be a uniform distribution over the region of interest, which spans the west coast of the Americas to the date line, from the equator to 30°N, and from 0 to 10^{-3} s^{-1} on the relative vorticity scale. The probability of detection is set to 0.95 and the width of the triangle proposal distribution is set to 5 time steps. When performing inference using Markov Chain Monte Carlo (MCMC), the sampler is run for 2000 iterations and the first 500 samples are discarded as burn-in. Each of the 1500 posterior samples is an $(s, \tau, \mathbf{q})$ tuple that defines a single track. The highest posterior mode is identified by computing a histogram of the $(s, \tau, \mathbf{q})$ tuples and keeping the tuple with the largest number of samples. In the case of a tie, the mode is chosen at random from among the tied $(s, \tau, \mathbf{q})$ tuples. After the

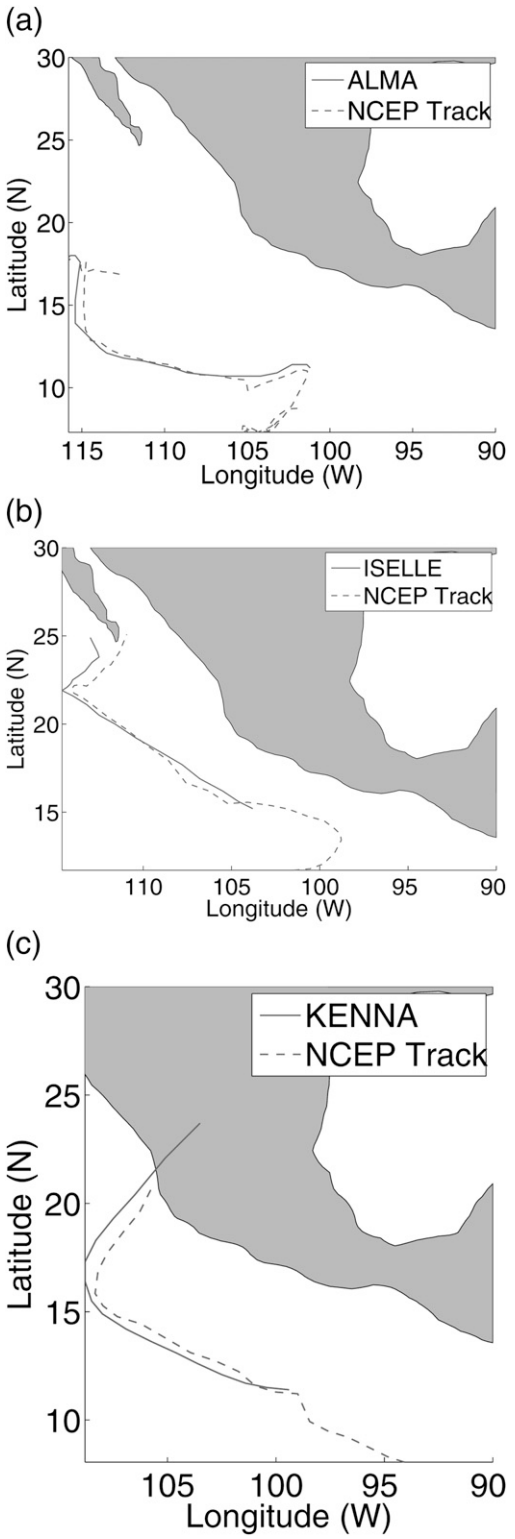

FIG. 6. The 2002 NHC best tracks (solid curves) of (a) Hurricanes Alma, (b) Tropical Storm Iselle, and (c) Hurricane Kenna compared to the tracks returned by the model (dashed curves). The track for Hurricane Kenna ends abruptly because the model is prevented from tracking over land.

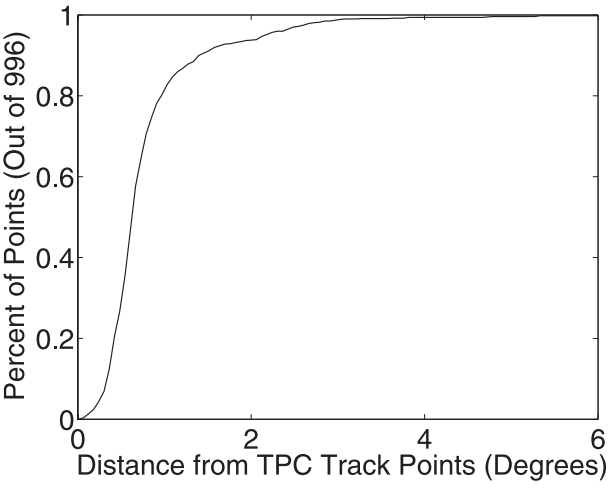

FIG. 7. A recall curve that plots the percentage of the 996 model track points that are found within a fixed distance of the NHC track points. The unmatched tracks from Tropical Storm Henriette, Hurricane Ele, and Hurricane Huko are excluded from the set. Over 90% of all the points are found within 1.45° of the NHC tracks and over 95% are found within 2.2°.

mode is identified, it is removed from the set of feature detections.

d. Application to NCEP GFS analysis

We apply our model to relative vorticity data from NCEP GFS analysis on the 975-hPa level during the summer half year (May–October) of 2000–02. The data is available at 1° by 1° spatial resolution, 4 times per day, which results in a sequence of 736 vorticity field images for each year. A total of 100 tracks were extracted and compared to a reference set of NHC best tracks.

The model can be applied to data with a different time resolution either by retraining the dynamics model at the new time scale, or by rescaling the dynamic parameters of an existing model. For example, if the new data has a temporal resolution that is an integral fraction, n, of the data used to train the original model, then one can simply update the existing model parameters: $\hat{\mathbf{A}} = \mathbf{A}^n$, $\hat{\mathbf{Q}} = \hat{\mathbf{A}}\mathbf{Q}\hat{\mathbf{A}}^{\mathrm{T}}$.

Figures 4a,b show the full set of feature detections from 2001 and a contour plot of the expected storm locations, respectively. The contour plot represents the probability of observing a storm at a given location assuming that the storm tracks follow the LDM dynamics. The dynamics model places almost no probability on observing storms in the extreme northern and western areas of the region of interest and the model does not initiate any storm tracks outside of a relatively small

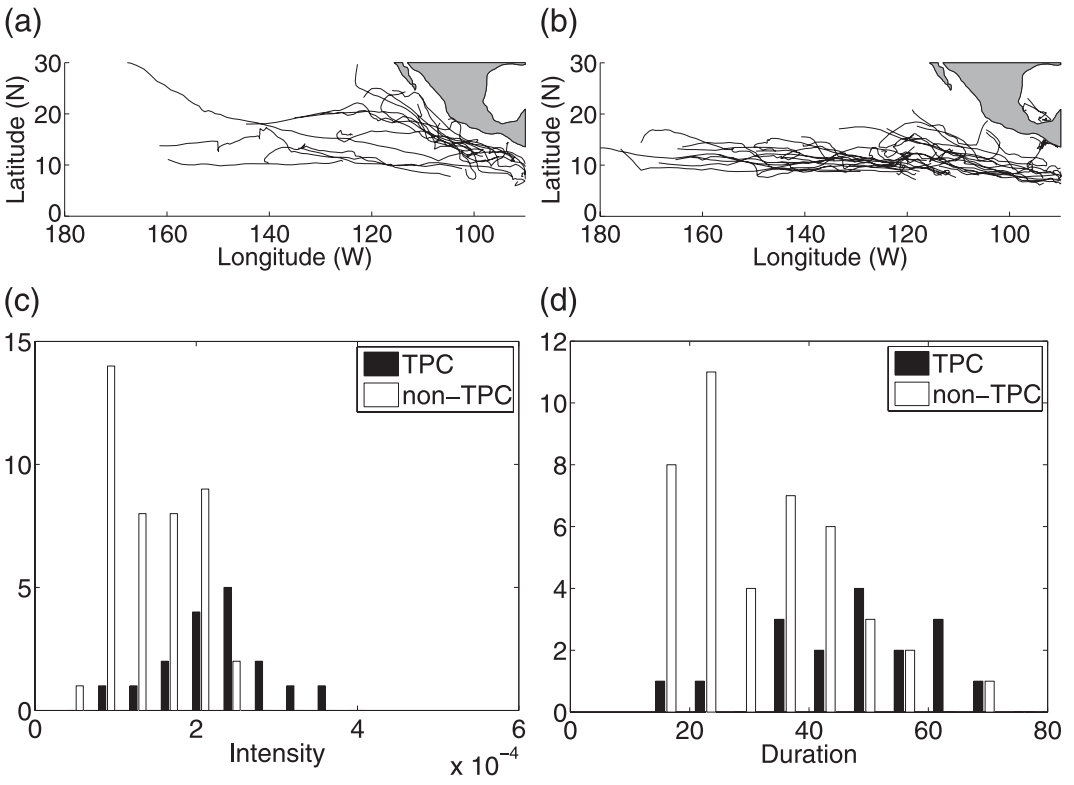

FIG. 8. (a) Comparison of the storms tracks and track statistics of the 17 identified NHC tracks in 2000 vs the other tracks returned by the model. The non-NHC tracks (b) primarily correspond to ITCZ breakdown events, have (c) a statistically weaker maximum intensity, and (d) shorter lifetime than the named storms.

genesis region. The set of all 100 extracted tracks is shown in Fig. 4c. A qualitative examination of the storm tracks shows that the extracted tracks obey the general constraints of the assumed dynamics model.

The model's tracks are quantitatively compared to the NHC best track dataset for each year. Quantitative evaluation of tracking methods is generally difficult because there is rarely an objective set of ground truth storm tracks for comparison. Because our LDM parameters prefer tracking strong storms, the NHC best-track dataset is used as a proxy for ground truth. The NHC tracks include all the named storms for a given year as well as tropical depressions, and thus, they may be considered as the set of ''important'' tracks, which could reasonably be found by any tracking method. Before beginning any analysis of the model's tracks, the NHC best-track set is filtered by removing all the storms that lie outside the region of interest or the 6-month seasonal window.[2]

To carry out the analysis, each NHC track is paired with the best corresponding track found by our model. A two-step method is used to identify a set of matching storm tracks. First, the tracks are scanned to find candidates containing an overlapping track point (latitude–longitude pair) within the same image. Once a candidate set for each NHC track is identified, the root-mean-square error (RMSE) is computed between the sets of corresponding track points. The model track with the smallest RMSE is selected as the match to the NHC track.

The set of matches are scanned again and any NHC track without a matching model track or a RMSE greater than 8° is considered a missed track and removed from further consideration. Over the three seasons, only three NHC tracks were not detected by the model—Tropical Storm Henriette in 2001, and Hurricanes Ele and Huko in 2002. These three tracks exhibited two types of detection failure (Fig. 5). Hurricanes Ele and Huko are both missed because of the extreme western location of their genesis. The model places a very low probability of storm genesis in this region and the fact that the two storms are ignored is consistent with

[2] The official hurricane season in the east Pacific lasts from 15 May through the end of November.

this model prior. In the case of Henriette, the first 2 days of its track did not match the dynamics model, so the model was never initialized near enough the true storm parameters to find its corresponding mode in the posterior distribution. If the model is manually initialized at Henriette's genesis, then the storm is found.

For the rest of the NHC tracks identified by the model, the model tracks exhibit excellent consistency with their corresponding NHC positions. A representative set of tracks from 2002—Hurricane Alma, Tropical Storm Iselle, and Hurricane Kenna—are shown in Fig. 6. Generally, each NHC track is found to be a subset of one of the model tracks where the model tracks a storm both before and after it reaches tropical depression strength. The fact that the tracks from our model are longer than the NHC tracks is a function of our feature detection threshold as it identifies vorticity maxima in storms that are weaker than tropical depressions.

The track points are used as a basis of further quantitative comparison by computing a *recall curve* (Fig. 7) for the full set of storm tracks across all 3 yr. The vertical axis is the fraction of model track points that lie within a certain distance (horizontal axis) of the associated NHC track points. The NHC and model track points both must be present at the same time for the comparison to be valid. The only way to achieve a 100% match at any distance threshold is to have a track point from our model present for *all* the NHC track points.

The algorithm exhibits very good recall, with over 75% of the model track points falling within 1° of the NHC tracks and over 90% falling within 1.5°. For comparison, the method of Kleppek et al. (2008) reported a match percentage of less than 60% for a threshold of about 1.65°. Significantly, 100% of the track points are eventually captured, which indicates that *every* NHC track point has a corresponding match in the set of model tracks.

Finally, Fig. 8 presents a statistical comparison between the set of model tracks matching the NHC tracks and the remaining set of detected model tracks. The NHC matches are shown in Fig. 8a and the non-NHC tracks in Fig. 8b. The remaining two figures are histograms comparing the the maximum track intensities (Fig. 8c) and durations (Fig. 8d) of the two sets of tracks, verifying that the NHC tracks are more intense.

The non-NHC tracks found by the model tend to be located within the ITCZ and are mostly zonally directed. The average duration of the non-NHC tracks is only slightly shorter than the NHC tracks, which indicates that the disturbances in the region tend to persist for about the same time span regardless of their strength.

4. Conclusions

We have presented a probabilistic model for automatically extracting storm tracks from a set of domain-specific feature detections. The novel aspects of our approach lie in the treatment of genesis and lysis times of storms as unknown variables in a probabilistic framework and the introduction of a flexible dynamics model in a modular fashion.

We developed an MCMC-based inference method for fitting the model parameters to observed data and a practical methodology for extracting multiple tracks from a single season of data. We then applied our method to the NCEP GFS dataset for the years 2000–02 and have shown quantitatively that our model is able to effectively find tracks that are 1) consistent with the assumed dynamics, 2) correspond to known storms from the NHC best-track set, and 3) identify a larger fraction of the track points of known storms than other published approaches. In principle the type of model proposed in this paper could also be applied to the problem of tracking storms over other geographical regions using an appropriate dynamics model.

The model could be extended to simultaneously track multiple storms at once, however, the additional complexity was not needed in this application because of the relatively large separation between storms. A multiple object tracker would be useful in applications where tracks overlap in time or the observational noise is large relative to the dynamics (e.g., $\|\mathbf{R}\| \gg \|\mathbf{C}\mathbf{Q}\mathbf{C}^{\mathrm{T}}\|$).

Acknowledgments. We thank two anonymous reviewers for comments on the manuscript. This work was supported by NSF Grants ATM-0530926 and IIS-0431085.

APPENDIX A

Marginalization of the Hidden States when $q_t = 0$ for $t \notin [s, \tau]$

It is advantageous to analytically integrate out nuisance variables from a Bayesian model before applying Monte Carlo simulation since the sampling variance will be reduced (Särkkä et al. 2007). In our model we consider the distribution of the state vector $\mathbf{x}_t$ outside of the storm's duration, $h(\mathbf{x}_t)$, as a nuisance variable.

Consider expanding the joint distribution at a single time $i \notin [s, \tau]$ and integrating out the state variable $\mathbf{x}_i$:

$$
\begin{aligned}
&\int_{\mathbf{x}_i} p(\mathbf{Y}, \mathbf{X}, \mathbf{q}, s, \tau)\, d\mathbf{x}_i \\
&= \int_{\mathbf{x}_i} p(s, \tau) \prod_{t=1}^{T} \prod_{j=1}^{n_t} p(\mathbf{y}_{t,j}|\mathbf{x}_t, q_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}, s, \tau) p(q_t|s, \tau)\, d\mathbf{x}_i \\
&= p(s, \tau) \prod_{t=1}^{T} p(q_t|s, \tau) \int_{\mathbf{x}_i} \prod_{j=1}^{n_t} p(\mathbf{y}_{t,j}|\mathbf{x}_t, q_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}, s, \tau)\, d\mathbf{x}_i \\
&= p(s, \tau) \left[\prod_{t} p(q_t|s, \tau) \right] \left[\prod_{t \neq i} \prod_{j=1}^{n_t} p(\mathbf{y}_{t,j}|\mathbf{x}_t, q_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}, s, \tau) \right] \int_{\mathbf{x}_i} \prod_{j=1}^{n_i} g(\mathbf{y}_{i,j}) h(\mathbf{x}_i)\, d\mathbf{x}_i \\
&= p(s, \tau) \left[\prod_{t} p(q_t|s, \tau) \right] \left[\prod_{t \neq i} \prod_{j=1}^{n_t} p(\mathbf{y}_{t,j}|\mathbf{x}_t, q_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}, s, \tau) \right] \prod_{j=1}^{n_i} g(\mathbf{y}_{i,j}) \int_{\mathbf{x}_i} h(\mathbf{x}_i)\, d\mathbf{x}_i \\
&= p(s, \tau) \left[\prod_{t} p(q_t|s, \tau) \right] \left[\prod_{t \neq i} \prod_{j=1}^{n_t} p(\mathbf{y}_{t,j}|\mathbf{x}_t, q_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}, s, \tau) \right] \prod_{j=1}^{n_i} g(\mathbf{y}_{i,j}) \\
&= p(s, \tau) \left[\prod_{t \neq i} \prod_{j=1}^{n_t} p(\mathbf{Y}_t|\mathbf{x}_t, q_t) p(\mathbf{x}_t|\mathbf{x}_{t-1}, s, \tau) p(q_t|s, \tau) \right] \prod_{j=1}^{n_i} g(\mathbf{y}_{i,j})
\end{aligned}
$$

This marginalization can be repeated for every $\mathbf{x}_t$ that lies outside of the storm's lifetime. This simplifies the model and matches our human intuition as well—if no storm is present, then one should not have to compute any storm-related quantities. The final form of the simplified joint distribution is

$$
\begin{aligned}
p(\mathbf{Y}, \mathbf{X}_{s,\tau}, \mathbf{q}_{s,\tau}, s, \tau) = p(s, \tau) &\left[p(\mathbf{x}_s) \prod_{t=s+1}^{\tau} p(\mathbf{x}_t|\mathbf{x}_{t-1}, s, \tau) \right] \\
&\times \left[\prod_{t \in [s,\tau]} \prod_{j=1}^{n_t} p(q_t|s, \tau) p(\mathbf{y}_{t,j}|\mathbf{x}_t, q_t) \right] \\
&\times \left[\prod_{t \notin [s,\tau]} \prod_{j=1}^{n_t} g(\mathbf{y}_{t,j}) \right]
\end{aligned}
$$

In the case where $p(\mathbf{x}_t|\mathbf{x}_{t-1}, s, \tau)$ is defined by a LDM, the state variables can be analytically integrated out within the storm's lifetime as well, which further simplifies the joint distribution (Bergman and Doucet 2000).

APPENDIX B

LDM Parameter Estimation from Labeled Examples

For our experiments, we trained the LDM using a set of 10 hand-labeled WPD tracks,

$$\mathbf{Z}_{\text{train}} = \{\mathbf{z}_{i,j}\} \quad i = 1 \ldots 10, \quad j = 1 \ldots t_i,$$

from August to October 2000 using the EM algorithm (Digalakis et al. 1993), where $\mathbf{z}_{i,j}$ is the feature vector for storm i at time j and t_i denotes the length of storm i. Because the training data contained only the WPD track positions (no intensity), we initially fit an $m = 4$, $n = 2$ LDM to the data and then manually augment the model parameters in order to add an intensity term and expand the model to $m = 5$, $n = 3$.

The model parameters are initialized with a simple first-order dynamics model that updates a storm's position by adding the current velocity. The initial state mean $\boldsymbol{\mu}_0$ and covariance $\boldsymbol{\Sigma}_0$ are initialized at the mean and covariance of the positions and velocities of the first observation across all the labeled tracks:

$$\hat{\mathbf{z}}_i = \begin{bmatrix} \mathbf{z}_{i,1} \\ \mathbf{z}_{i,2} - \mathbf{z}_{i,1} \end{bmatrix} \quad \boldsymbol{\mu}_0 = \frac{1}{10} \sum_{i=1}^{10} \hat{\mathbf{z}}_i$$

$$\boldsymbol{\Sigma}_0 = \frac{1}{9} \sum_{i=1}^{10} (\hat{\mathbf{z}}_i - \boldsymbol{\mu}_0)(\hat{\mathbf{z}}_i - \boldsymbol{\mu}_0)^{\mathrm{T}}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

After the EM algorithm converges, we augment the trained model parameters in order to add the intensity component. Under the assumption that a storm's intensity

is independent from its motion, we can perform this augmentation by simply appending a diagonal term to each LDM parameter. We model the intensity as a random walk, so the dynamics of the intensity are trivial, $A_{5,5} = 1$. Because of our independence assumption, the observed intensity is a direct measurement of the hidden state intensity, so $C_{3,5} = 1$.

The remaining intensity parameter values are estimated by identifying the closest feature detection to each of the hand-labeled track points and calculating the mean and variance of the observed intensity values. The observed variance is split among **Q** and **R**, with proportionally more variance being added to the **Q** matrix because we expect our observations to have less inherent variability than the random changes in intensity over a 6-h period.

The final set of LDM parameters that are used in our experiments are

$$\mathbf{A} = \begin{bmatrix} 1.0002 & 1.4235 & 0.0037 & -0.1383 & 0.0 \\ -0.0014 & 0.5659 & 0.0053 & -0.0771 & 0.0 \\ 0.0006 & 0.0546 & 1.0100 & 0.0546 & 0.0 \\ 0.0009 & -0.0157 & -0.0138 & 0.8950 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 0.9584 & 3.1962 & 0.6243 & 0.0235 & 0.0 \\ 0.0009 & 0.1189 & 0.9761 & 0.2602 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} 2.3143 & -0.6892 & 0.0823 & -0.0189 & 0.0 \\ -0.6892 & 0.2196 & -0.0460 & -0.0060 & 0.0 \\ 0.0823 & -0.0460 & 0.1755 & 0.0119 & 0.0 \\ -0.0189 & -0.0060 & 0.0119 & 0.3344 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 2.6 \times 10^{-9} \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 0.0380 & -0.0051 & 0.0000 \\ -0.0051 & 0.0097 & 0.0000 \\ 0.0000 & 0.0000 & 1.0 \times 10^{-9} \end{bmatrix}$$

$$\boldsymbol{\mu}_0 = \begin{bmatrix} 156.7625 \\ 34.2135 \\ 6.3930 \\ 1.4997 \\ 1.0 \times 10^{-6} \end{bmatrix} \quad \boldsymbol{\Sigma}_0 = \begin{bmatrix} 8.5446 & 1.5322 & -3.2486 & -0.5101 & 0.0 \\ 1.5322 & 0.8893 & -0.7035 & 0.0078 & 0.0 \\ -3.2486 & -0.7035 & 4.0124 & -0.3671 & 0.0 \\ -0.5101 & 0.0078 & -0.3671 & 2.7500 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \times 10^{-7} \end{bmatrix}.$$

REFERENCES

Avila, L. A., 1991: Eastern North Pacific hurricane season of 1990. *Mon. Wea. Rev.,* **119,** 2034–2046.

Bar-Shalom, Y., 1987: *Tracking and Data Association*. Academic Press Professional, Inc., 353 pp.

Bengtsson, L., K. I. Hodges, and M. Esch, 2007: Tropical cyclones in a T159 resolution global climate model: Comparison with observations and re-analyses. *Tellus,* **59A,** 396–416.

Bergman, N., and A. Doucet, 2000: Markov Chain Monte Carlo data association for target tracking. *IEEE Conf. on Acoustics, Speech and Signal Processing,* Vol. 2, Washington, D.C., IEEE Computer Society, II705–II708, doi:10.1109/ICASSP.2000.859057.

Brooks, S. P., and A. Gelman, 1998: General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.,* **7,** 434–455.

Chang, K., S. Mori, and C. Chong, 1994: Performance evaluation of track initiation in dense target environments. *IEEE Trans. Aerospace Electron. Syst.,* **30,** 213–219, doi:10.1109/7.250421.

Chauvin, F., J.-F. Royer, and M. Deque, 2006: Response of hurricane-type vortices to global warming as simulated by ARPERGE-Climate at high resolution. *Climate Dyn.,* **27,** 377–399.

Cox, I. J., and S. L. Hingorani, 1996: An efficient implementation and evaluation of Reid's multiple hypothesis tracking algorithm for visual tracking. *IEEE Trans. Pattern Anal. Mach. Intell.,* **18,** 138–150, doi:10.1109/34.481539.

Davis, C., C. Snyder, and A. C. Didlake, 2008: A vortex-based perspective of eastern Pacific tropical cyclone formation. *Mon. Wea. Rev.,* **136,** 2461–2477.

Digalakis, V., J. Rohlicek, and M. Ostendorf, 1993: ML estimation of a stochastic linear system with the EM algorithm and its application to speech recognition. *IEEE Trans. Speech Audio Process.,* **1,** 431–442.

Gauvrit, H., J. Le Cadre, and C. Jauffret, 1997: A formulation of multitarget tracking as an incomplete data problem. *IEEE Trans. Aerospace Electron. Syst.,* **33,** 1242–1257, doi:10.1109/7.625121.

Gelb, A., 1974: *Applied Optimal Estimation*. The MIT Press, 382 pp.

Gelfand, S. B., T. E. Fortmann, and Y. Bar-Shalom, 1996: Adaptive detection threshold optimization for tracking in clutter. *IEEE Trans. Aerospace Electron. Syst.,* **32,** 514–523.

Gelman, A., and D. Rubin, 1992: Inference from iterative simulation using multiple sequences. *Stat. Sci.,* **7,** 457–511.

——, J. B. Carlin, H. S. Stern, and D. B. Rubin, 2003: *Bayesian Data Analysis*. 2nd ed. Chapman & Hall/CRC, 696 pp.

Gilks, W. R., 1995: *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, 488 pp.

Gray, W. M., 1968: Global view of the origin of tropical disturbances and storms. *Mon. Wea. Rev.,* **96,** 669–700.

Haralick, R. M., and L. G. Shapiro, 1992: *Computer and Robot Vision*. Vol. I. Addison-Wesley, 28–48.

Hodges, K. I., 1994: A general method for tracking analysis and its application to meteorological data. *Mon. Wea. Rev.,* **122,** 2573–2586.

——, 1999: Adaptive constraints for feature tracking. *Mon. Wea. Rev.,* **127,** 1362–1373.

Hu, Z., H. Leung, and M. Blanchette, 1997: Statistical performance analysis of track initiation techniques. *IEEE Trans. Signal Process.,* **45,** 445–456, doi:10.1109/78.554308.

Ihler, A., S. Kirshner, M. Ghil, A. Robertson, and P. Smyth, 2007: Graphical models for statistical inference and data assimilation. *Physica D,* **230,** 72–87.

Jordan, M. I., 2004: Graphical models. *Stat. Sci.,* **19,** 140–155.

Karlsson, R., and F. Gustafsson, 2001: Monte Carlo data association for multiple target tracking. *IEE Target Tracking: Algorithms Appl.,* **1** (2001/174), 13/1–13/5.

Kieu, C. Q., and D.-L. Zhang, 2008: Genesis of tropical storm Eugene (2005) from merging vortices associated with ITCZ breakdowns. Part I: Observational and modeling analyses. *J. Atmos. Sci.,* **65,** 3419–3439.

Kleppek, S., V. Muccione, C. C. Raible, D. N. Bresch, P. Koellner-Heck, and T. F. Stocker, 2008: Tropical cyclones in ERA-40: A detection and tracking method. *Geophys. Res. Lett.,* **35,** L05707, doi:10.1029/2008GL033880.

Liu, J. S., 2002: *Monte Carlo Strategies in Scientific Computing*. Springer, 346 pp.

Magnusdottir, G., and C.-C. Wang, 2008: Intertropical convergence zones during the active season in daily data. *J. Atmos. Sci.,* **65,** 2425–2436.

McMillan, J. C., and S. S. Lim, 1990: Data association algorithms for multiple target tracking. Tech. Rep. ADA231688, Defence Research Establishment Ottawa, Ontario, Canada, 37 pp.

Muskulus, M., and D. Jacob, 2005: Tracking cyclones in regional model data: The future of Mediterranean storms. *Adv. Geosci.,* **2,** 13–19.

Oh, S., S. Russell, and S. Sastry, 2004: Markov Chain Monte Carlo data association for general multiple-target tracking problems. *Proc. 43rd IEEE Conf. on Decision and Control,* Vol. 1, Nassau, Bahamas, IEEE, 735–742, doi:10.1109/CDC.2004.1428740.

Rasmussen, C., and G. D. Hager, 2001: Probabilistic data association methods for tracking complex visual objects. *IEEE Trans. Pattern Anal. Mach. Intell.,* **23,** 560–576, doi:10.1109/34.927458.

Rauch, H. E., F. Tung, and C. T. Striebel, 1965: Maximum likelihood estimates of linear dynamic systems. *AAIA J.,* **3,** 1445–1450.

Reid, D. B., 1979: An algorithm for tracking multiple targets. *IEEE Trans. Auto. Control,* **AC-24** (6), 843–854.

Särkkä, S., T. Tamminen, A. Vehtari, and J. Lampinen, 2004: Probabilistic methods in multiple target tracking: Review and bibliography. Tech. Rep., Helsinki University of Technology, 95 pp.

——, A. Vehtari, and J. Lampinen, 2007: Rao-Blackwellized Monte Carlo data association for multiple target tracking. *Info. Fusion,* **8,** 2–15, doi:10.1016/j.inffus.2005.09.009.

Storlie, C. B., T. C. M. Lee, J. Hannig, and D. Nychka, 2009: Tracking of multiple merging and splitting targets: A statistical perspective. *Stat. Sin.,* **19,** 1–30.

Streit, R. L., and T. E. Luginbuhl, 1994: Maximum likelihood method for probabilistic multihypothesis tracking. *Signal and Data Processing of Small Targets,* O. E. Drummond, Ed., International Society for Optical Engineering (SPIE Proceedings, Vol. 2235), 394–405.

Tarjan, R. E., 1975: Efficiency of a good but not linear set union algorithm. *J. ACM,* **22,** 215–225, doi:10.1145/321879.321884.

Vermaak, J., S. J. Godsill, and P. Perez, 1995: Monte Carlo filtering for multitarget tracking and data association. *IEEE Trans. Aerospace Electron. Syst.,* **41,** 309–332.

Walsh, K. J. E., M. Fiorino, C. W. Landsea, and K. L. McInnes, 2007: Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *J. Climate,* **20,** 2307–2314.

Wang, C., and G. Magnusdottir, 2006: The ITCZ in the central and eastern Pacific on synoptic time scales. *Mon. Wea. Rev.,* **134,** 1405–1421.

West, M., and J. Harrison, 1997: *Bayesian Forecasting and Dynamic Models*. 2nd ed. Springer-Verlag, 681 pp.