Spatiotemporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds

Christopher K. WIKLE, Ralph F. MILLIFF, Doug NYCHKA, and L. Mark BERLINER

Spatiotemporal processes are ubiquitous in the environmental and physical sciences. This is certainly true of atmospheric and oceanic processes, which typically exhibit many different scales of spatial and temporal variability. The complexity of these processes and the large number of observation/prediction locations preclude the use of traditional covariance-based spatiotemporal statistical methods. Alternatively, we focus on conditionally specified (i.e., hierarchical) spatiotemporal models. These methods offer several advantages over traditional approaches. Primarily, physical and dynamical constraints can be easily incorporated into the conditional formulation, so that the series of relatively simple yet physically realistic conditional models leads to a much more complicated spatiotemporal covariance structure than can be specified directly. Furthermore, by making use of the sparse structure inherent in the hierarchical approach, as well as multiresolution (wavelet) bases, the models can be computed with very large datasets. This modeling approach was necessitated by a scientifically meaningful problem in the geosciences. Satellite-derived wind estimates have high spatial resolution but limited global coverage. In contrast, wind fields provided by the major weather centers provide complete coverage but have low spatial resolution. The goal is to combine these data in a manner that incorporates the space-time dynamics inherent in the surface wind field. This is an essential task to enable meteorological research, because no complete high-resolution surface wind datasets exist over the world oceans. High-resolution datasets of this type are crucial for improving our understanding of global air–sea interactions affecting climate and tropical disturbances, and for driving large-scale ocean circulation models.

KEY WORDS: Climate; Combining information; Conjugate gradient algorithm; Dynamical model; Fractal process; Gibbs sampling; Numerical model; Ocean model; Satellite data; Turbulence; Wavelets.

1. INTRODUCTION

Fierce storms in California, floods in Peru, drought in Australia and Indonesia—just a few of the extreme weather events attributed to the 1997–1998 El Niño event (e.g., Kerr 1998). This El Niño brought unprecedented public attention to the interaction between the tropics and extratropics, and perhaps more important, the interaction between the ocean and the atmosphere. These interactions have been a focus of climate research over the past decade. Changes in weather around the world, such as those associated with the recent El Niño, have been linked to variations in the atmospheric circulation that at a fundamental level are affected by exchanges in heat, moisture, and momentum between the atmosphere and ocean. This exchange across the air/sea boundary is critically related to small-scale spatiotemporal features of sea-surface winds.

Climatologists and oceanographers use wind information in two main ways: (1) to improve fundamental knowledge about atmospheric phenomena such as El Niño (e.g., Liu, Tang, and Wu 1998), tropical cyclones (e.g., Gray 1976), and large-scale tropical oscillations (e.g., Madden and Julian 1994), and (2) to provide input (forcing) for deterministic models of the coupled ocean/atmosphere system (e.g., Milliff, Large, Morzell, Danabasoglu, and Chin 1999 and references therein). In both cases, one must know something about the behavior of the surface wind field and its horizontal derivatives at small scales. For example, it has been shown through the use of simulated datasets that deterministic models of the ocean are sensitive to both the temporal (Large, Holland, and Evans 1991) and spatial (Milliff, Large, Holland, and McWilliams 1996) resolution of the surface wind forcing (see also Chen, Liu, and Witter 1999). Indeed, although the deterministic coupled ocean/atmosphere models used for prediction of the 1997–1998 El Niño were more accurate than for previous El Niño events, indications are that many of these models would have performed better had uniformly high-resolution tropical wind fields been available (Kerr 1998).

Unfortunately, there are no spatially and temporally complete high-resolution observations of surface winds over the tropical oceans. Thus the major scientific challenge here is the development of physically realistic high-resolution tropical wind fields. Our fundamental scientific contribution is the development and implementation of a statistical approach to generate high-resolution wind distributions over large expanses of the tropical ocean. To that end, we develop a hierarchical Bayesian spatiotemporal dynamic model that combines wind data from different sources, and background physics, to produce realizations of high-resolution surface wind fields. The Bayesian approach is ideal for this application because (1) it provides a mechanism for combining data from very different sources, (2) it provides a natural framework in which to include scientific knowledge in the model, and (3) it provides posterior distributions on quantities of interest that can be used for scientific inference.

Our statistical analyses use two strikingly different datasets. The first dataset involves satellite-derived wind estimates that

Christopher K. Wikle is Assistant Professor, Department of Statistics, University of Missouri, Columbia, MO 65211. Ralph F. Milliff is Scientist, Colorado Research Associates, Boulder, CO 80301. Doug Nychka is Senior Scientist, Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, CO 80307. L. Mark Berliner is Professor, Department of Statistics, Ohio State University, Columbus, OH 43210. The authors thank T. Nakazawa for discussions regarding the validation of the model and for providing the GMS cloud imagery used in this presentation, T. Hoar for assistance with the acquisition of datasets, J. Tribbia for helpful discussions pertaining to linear wave theory, and N. Cressie for comments on an early draft. The authors also thank the reviewers, the editor, and the associate editor for valuable comments. Support for this research was provided by the National Center for Atmospheric Research (NCAR) Geophysical Statistics Project, sponsored by the National Science Foundation (NSF) under grant DMS93-12686, and by the NCAR NSCAT Science Working Team cooperative agreement with the National Aeronautics and Space Administration's Jet Propulsion Laboratory.

Journal of the American Statistical Association
June 2001, Vol. 96, No. 454, Applications and Case Studies

have high resolution in space but are limited in areal coverage at any given time. Milliff and Morzel (2001) demonstrated that the information from a single instrument of this type is not sufficient to resolve all of the meteorological events in the surface wind field. The second dataset comprises wind estimates, known as analyses, provided by the major weather centers. These provide complete wind fields but have low spatial resolution. Although the large-scale features of the tropical atmosphere are generally well represented by these analysis fields, these fields are unable to resolve many of the small- to medium-scale features in the wind fields needed to understand the dynamics of the tropical ocean and atmosphere (e.g., Milliff et al. 1996). Hence, in isolation, neither of these two datasets provides the breath of scientific information sought by climatologists. Our Bayesian model combines these data to yield information about winds at a useful spatial scale, and in a manner that incorporates physical theory about the spatiotemporal dynamics inherent in tropical surface winds.

We demonstrate (see Fig. 5) that our posterior wind fields contain much more finely resolved features than do the current state-of-the-art weather center wind fields over the tropics. Furthermore, based on external verification with remotely sensed cloud imagery, these higher-resolution features in the wind fields correspond to physically meaningful features of the atmosphere. We emphasize that until satellite wind data are assimilated adequately into numerical weather prediction models of similar resolution, a Bayesian procedure of the kind that we derive here provides the only source of high-resolution tropical wind field information sufficient for many aspects of research regarding air–sea interactions and their effects on climate. Furthermore, the probability distributions for wind fields that we provide will for the first time allow scientists to consider the distributional nature of phenomena that depend on air–sea interaction.

The datasets used here are described in Section 2. The physically based spatiotemporal model that we have developed is described in Section 3. By "physically based," we mean that substantial physical modeling and background science were used in both development of the model and specifications of priors on model parameters. The Bayesian implementation and specific computational issues related to our analysis are discussed in Section 4. The huge datasets used and the large number of unknowns modeled necessitated the development of special algorithms. These developments are of general interest in large-scale Bayesian analyses. Model verification and inference based on our wind model are described in Section 5. A brief discussion is presented in Section 6.

2. WIND DATA

Because winds are vector quantities, they can be split into orthogonal components. We use the standard decomposition in which u represents the east–west ("x-direction") component and v represents the north–south ("y-direction") component. Although other decompositions are possible, we selected this Cartesian decomposition for physical reasons; the equator is a fundamental line of symmetry in the equatorial dynamics that govern weather in the tropics and is a source of anisotropy that discourages the use of any coordinate system other than Cartesian (e.g., Gill 1982, pp. 436–463). We consider surface wind components over a spatial domain in the western Pacific Ocean from 107° to 170° E longitude and 23°S to 24°N latitude, as shown in Figure 1. This portion of the equatorial Pacific contains the "warm pool region" and is critical to the forcing and maintenance of many weather and climate-scale phenomena (e.g., Philander 1990). We focus on 6-hour increments during the 2-week period from October 28, 1996, through November 10, 1996. Tropical variability consistent with these scales include, for example, westerly wind bursts, equatorial Rossby wave propagation, and tropical storms. The 2-week time period is sufficient to capture up to five such events.

Although some in situ observations of ocean surface winds are made from buoys and ships, they are rather sparsely distributed in space and time relative to land-based observation networks. The world's major meteorological centers take these few observations and insert them into global-scale numerical weather prediction models to produce tropical wind field analyses (e.g., Daley 1991). Hence the resulting data are not measurements or observations in the traditional sense, but rather are *statistics* computed as highly complex functions of observations.

We consider weather center wind fields from the National Centers for Environmental Prediction (NCEP). These data represent surface winds (actually, 10 m above the surface) and have a reporting period of 6 hours and spatial resolution of nearly 2°, or about 200 km in equatorial regions. NCEP u-winds for three consecutive 6-hour periods in early November 1996 are shown in Figure 2(a).

Wind data from the NASA scatterometer (NSCAT) instrument are also used here. A scatterometer is a satellite-borne instrument that emits radar pulses at specific frequencies and polarizations toward the sea surface, where they are backscattered by surface capillary waves (e.g., Naderi, Freilich, and Long 1991). The backscattering is detected and related, through a "geophysical model function," to wind speed and direction near the surface (usually 10 m; see, e.g., Stoffelen and Anderson 1997; Wentz and Freilich 1997; Wentz and Smith 1999). That is, as in the case of analysis fields, these data are not direct measurements of winds, but rather are functions of backscatter detections.

Due to the polar orbit of these satellite platforms, the temporal resolution of these data are relatively sparse and, over the span of several hours, the spatial coverage area is relatively small; see Figure 2(b). Each "snapshot" in time includes all observations within a 6-hour window centered on the corresponding analysis time. The NSCAT surface (i.e., 10-m) wind data used here were produced by the NSCAT-1 model function (Wentz and Freilich 1997). These data have a 50-km nominal spatial resolution, although the reported winds are actually derived by applying the model function to an average of several backscatter observations within a 50-km by 50-km observational "cell."

Notation. Let $V_a(\mathbf{r}_i; t)$ and $U_a(\mathbf{r}_i; t)$ denote the NCEP analysis north–south and east–west, wind components at spatial location $\{\mathbf{r}_i : i = 1, \ldots, m\}$ and time $\{t : t = 1, \ldots, T\}$. The scatterometer (NSCAT) north–south (east–west) wind component is denoted by $V_s(\tilde{\mathbf{r}}_j; t)$ $(U_s(\tilde{\mathbf{r}}_j; t))$ at location $\{\tilde{\mathbf{r}}_j : j = 1, \ldots, p_t\}$ and time $\{t : t = 1, \ldots, T\}$. (The number

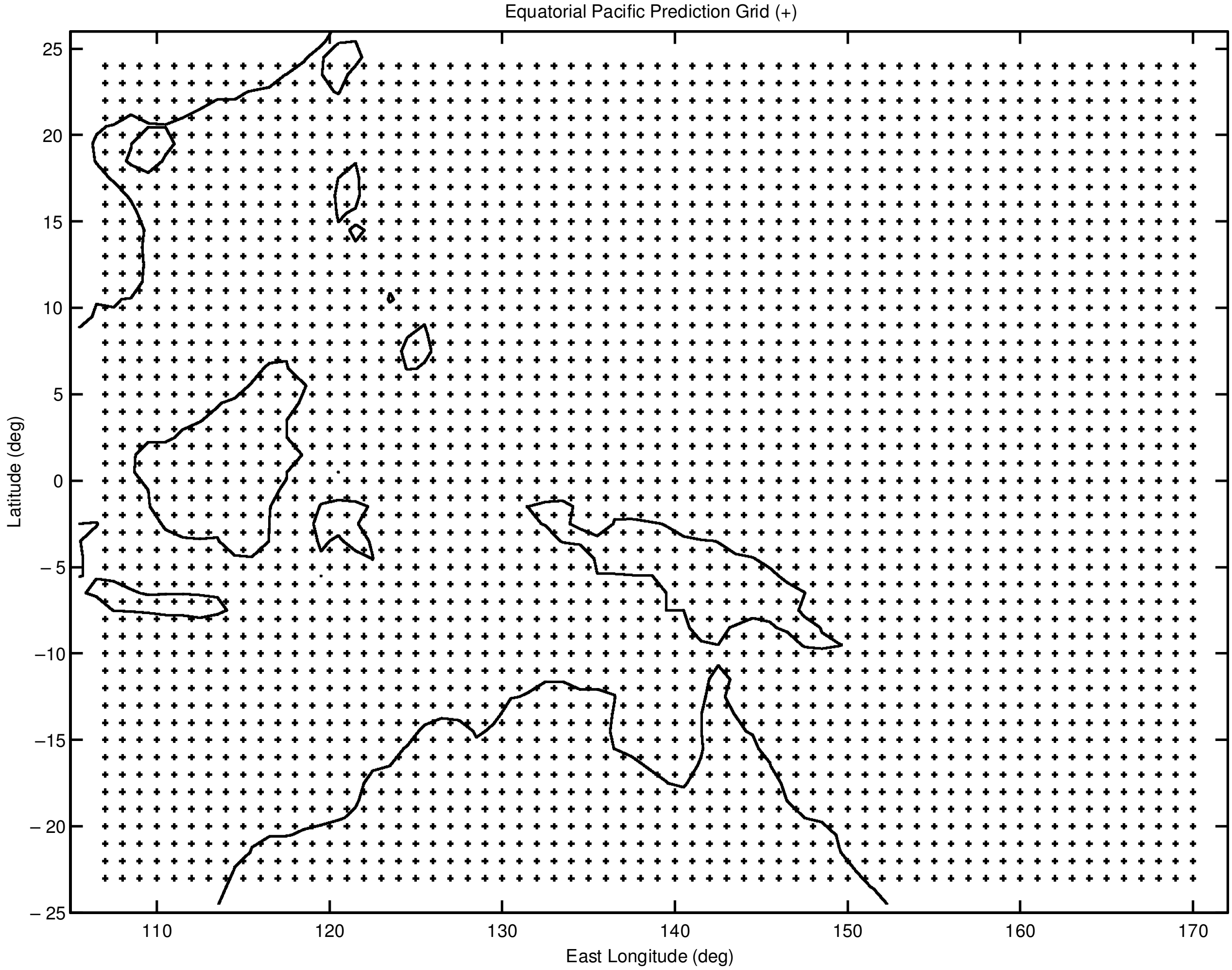

Figure 1. Prediction Grid Locations (+) Over the Equatorial Pacific Study Region.

of NSCAT observations, p_t can be highly variable, see Fig. 2.) We define the "true" (i.e., noiseless) wind components as $v(\mathbf{s}_i; t)$ and $u(\mathbf{s}_i; t)$ at spatial locations $\{\mathbf{s}_i : i = 1, \ldots, n\}$ and times $\{t : t = 1, \ldots, T\}$. The cumbersome notation of indexing spatial locations is needed because we are faced with a "change of support" problem: The NCEP and NSCAT data represent different spatial scales, both of which differ from the desired prediction sites $\mathbf{s}_i$.

In the present example, we choose a 1-degree regular prediction grid (Fig. 1) and consider 54 6-hour time increments over the period from 0600 UTC (Coordinated Universal Time) on October 28, 1996, to 1200 UTC on November 10, 1996. We neglect small displacements in the prediction lattice that are due to the curvature of the earth.

Next, let $\mathbf{V}_t$ denote an $m + p_t$ vectorization of the north–south weather center and scatterometer observations at time t. Similarly, $\mathbf{U}_t$ is the combined list of the data corresponding to the east–west component. Also, let $\mathbf{v}_t$ and $\mathbf{u}_t$ be n vectors of the "true" north–south and east–west wind components, at prediction locations at time t.

Finally, we use the following notation to denote matrices composed of columns of vectors representing intervals of time: let $\{\mathbf{V}\}_A^B$ be the collection of vectors $\{\mathbf{V}_t : t = A, \ldots, B\}$.

3. HIERARCHICAL SPACE-TIME MODELS

A major difficulty in the application of statistical spatiotemporal models in geophysical problems has been adequately describing the complicated spatiotemporal covariance structures inherent in these contexts. (For an overview of traditional spatiotemporal modeling approaches, see Wikle and Cressie 1999.) These methods are not suitable to the present problem in that they cannot easily account for propagation of synoptic-scale weather disturbances, fill "gaps" in the observations with realistic variance at all spatial scales, include multiple measurement errors and change of support for different data sources, or incorporate huge amounts of data.

3.1 The Hierarchical Approach

Hierarchical models are ideal for extremely complex and/or high-dimensional problems. In essence, the strategy is based on the formulation of three primary statistical models or stages:

- Stage 1: Data model: $[data|process, \theta_1]$
- Stage 2: Process model: $[process|\theta_2]$
- Stage 3: Prior on parameters: $[\theta_1, \theta_2]$

Here the bracket notation denotes probability distribution (e.g., Gelfand and Smith 1990), and θ_1 and θ_2 generically

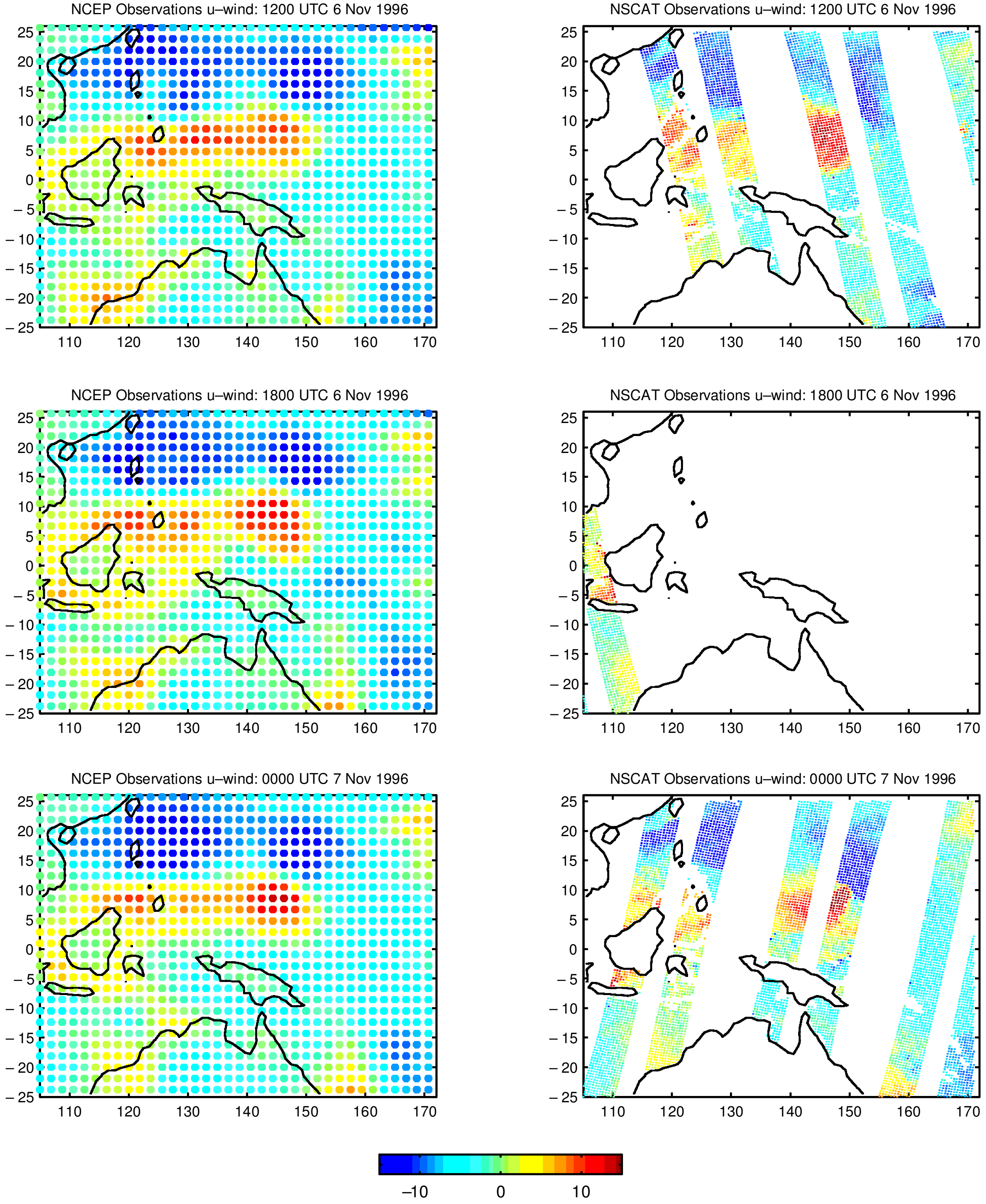

Figure 2. NCEP and NSCAT Sampling Locations and u-Wind Component Value (ms^{-1}) Within 6-Hour Time Windows Centered on 1200 UTC on November 6, 1996, 1800 UTC on November 6, 1996, and 0000 UTC on November 7, 1996.

represent parameters introduced in the modeling. The idea is to approach complex problems by breaking them into pieces—in this case, a series of conditional models (e.g., Berliner 1996). The stage 2 model for the *process* (in our case, true winds) can itself be specified as a product of physically motivated conditional distributions. By treating the spatiotemporal variability as a series of relatively simple, yet physically based conditional models, we can obtain spatiotemporal dependence structures that are much more complicated (and more realistic physically) than could be specified directly. Bayesian analysis relies on the posterior distribution of the process of interest and parameters given data: [*process*, θ_1, θ_2 | *data*]. Recent examples of hierarchical Bayesian spatiotemporal models include that used by Waller, Carlin, Xia, and

Gelfand (1997) for mapping disease rates. An overview of hierarchical spatiotemporal dynamic models along with a geophysical application has been provided by Wikle, Berliner, and Cressie (1998).

3.2 Stage 1: Data Model

We expect the wind data to be replete with complicated spatiotemporal dependencies. However, conditional on the true winds, we expect the complexity of this dependence to be dramatically reduced. That is, stage 1 models only *measurement errors*, not that portion of the complex structure present in the data due to the structure of the true winds. The fundamental assumptions are that, conditional on the true process $\{\mathbf{u}\}_1^T, \{\mathbf{v}\}_1^T$, the data are independent with respect to time, and the set of U observations is independent of the set of V observations. Specifically, we have that

$$[\{\mathbf{V}\}_1^T, \{\mathbf{U}\}_1^T | \{\mathbf{v}\}_1^T, \{\mathbf{u}\}_1^T; \theta_1] = \prod_{t=1}^{T} [\mathbf{V}_t | \mathbf{v}_t; \theta_1][\mathbf{U}_t | \mathbf{u}_t; \theta_1]. \quad (1)$$

In particular, we assume normally distributed errors,

$$\mathbf{V}_t | \mathbf{v}_t, \Sigma_t \sim \text{Gau}(\mathbf{K}_t \mathbf{v_t}, \Sigma_\mathbf{t}) \quad \text{and } \mathbf{U}_t | \mathbf{u}_t, \Sigma_t \sim \text{Gau}(\mathbf{K}_t \mathbf{u}_t, \Sigma_t), \quad (2)$$

where $\text{Gau}(\nu, A)$ refers to a multivariate Gaussian distribution with mean ν and covariance matrix A. We assume that the covariance matrices Σ_t are diagonal with unknown variances σ_B^2, for NCEP data at sites on the boundary of the NCEP grid, σ_I^2 for NCEP data at interior sites on the NCEP grid, and σ^2 for NSCAT observations; that is, the first m diagonal elements of Σ_t are equal to either σ_I^2 or σ_B^2, and the remaining p_t are equal to σ^2. Further, for each t, $\mathbf{K}_t$ is a specified $(m+p_t) \times n$ matrix that maps the prediction grid locations to the observation locations.

Several issues surround our assumptions about the data-acquisition process. First, we assume that conditional on true winds, the scatterometer errors and the NCEP analysis errors are independent. This is quite plausible, because the NCEP did not use scatterometer data in producing wind fields. Second, evidence in the literature points to the plausibility of our assumptions that the scatterometer errors are mutually conditionally independent and have homogeneous variance, and that the east–west and north–south component errors are independent. For example, Freilich (1997) demonstrated that an independent and normally distributed random error model for scatterometer velocity components is consistent with observed distributions for wind speed. Freilich and Dunbar (1999) considered comparisons between collocated satellite wind estimates and direct measurements from ocean buoys in a validation study. They concluded that the independent-component error model, with standard deviations equal to 1.3 m/s (for both components), is appropriate for NSCAT data. Furthermore, over the relatively small geographical region considered here, these references suggest that the homogeneous variance assumption is reasonable. Haslett and Raftery (1989) showed that application of a square-root transformation may enhance both homogeneity and normality in wind measurements. As suggested by Freilich (1997), such a homogeneity-of-variance transformation of wind speed is consistent with the independent, homogeneous normal random measurement error model for the Cartesian wind components. Finally, the assumption that NCEP analysis errors are mutually independent seems to be the least tenable assumption in view of the complex nature of the numerical and statistical methods used in the production of such information. The formulation of genuine covariances for analyzed fields is a major research area in its own right, and well beyond the scope of this article. We believe that the independence assumption is not critical for our results.

Mapping Matrices. We partition the mapping matrices as $\mathbf{K}_t = [\mathbf{K}_a', \mathbf{K}_s'(t)]'$, where $\mathbf{K}_a$ and $\mathbf{K}_s(t)$ are $m \times n$ and $p_t \times n$ matrices. Because the prediction grid is at a finer resolution than the NCEP data, $\mathbf{K}_a$ acts by assuming that the conditional means of the data are smoothed versions of the "true" winds on the lattice. This "change-of-support" approach is further justified because NCEP data have been shown to be too smooth at large scales (e.g., Milliff et al. 1999; Wikle, Milliff, and Large 1999). Specifically, the $\mathbf{K}_a$ matrix considers the nearest nine prediction grid locations within some distance D ($D = 165$ km) and weights those locations linearly by $w_i = (D - d_i)/w^*$, where d_i is the distance between the ith prediction grid location and the NCEP datum location and w^* normalizes the weights to sum to 1.

Each $\mathbf{K}_s(t)$ is an incidence matrix of 0's and 1's that simply maps the conditional mean of an NSCAT observation to the nearest grid process location. The error induced by this mapping is related to the chosen prediction grid resolution. Effectively, by using the mapping matrix, $\mathbf{K}_t$, we allow the wind process to "live" on a fine-resolution regular grid. The resolution of this grid could be so high as to allow the NSCAT data points to each correspond to a unique lattice location. Practically, a balance must be sought between computational expense, grid resolution, and resolution of the physics that one is seeking to describe or model. More-complicated approaches to parameterizing both $\mathbf{K}_a$ and $\mathbf{K}_s(t)$ are possible (see Wikle and Berliner 2001). However, for computationally purposes, these mapping matrices must be very sparse (see Sec. 4.1).

3.3 Stage 2: Priors on the Process

Our task is to formulate a joint probability model for the gridded wind process, $\{\mathbf{u}\}_1^T, \{\mathbf{v}\}_1^T$. We begin by decomposing each of the wind processes into three physically meaningful components. The decomposition and models for the resulting components were developed based on our physical and statistical understanding of the problem. After a review of that reasoning in the following section, we present the specific statistical models used for each of the three components.

3.3.1 Decomposition of the Wind Process. In the equatorial region, much of the large-scale variability in wind fields can be represented by treating the atmosphere as a thin fluid; that is, the depth of the atmosphere is much smaller than characteristic horizontal length scales (e.g., Gill 1982; Holton 1992). However, the thin-fluid approximation is incomplete in that it excludes small-scale motions that are fundamentally three-dimensional, and it is based on a zero-mean background flow. The following decompositions for our statistical model address these deficiencies while retaining the convenience of the thin-fluid approximation:

$$\mathbf{u}_t = \boldsymbol{\mu}_u + \mathbf{u}_t^E + \tilde{\mathbf{u}}_t \tag{3}$$

and

$$\mathbf{v}_t = \boldsymbol{\mu}_v + \mathbf{v}_t^E + \tilde{\mathbf{v}}_t. \tag{4}$$

Here $\boldsymbol{\mu}_u$ and $\boldsymbol{\mu}_v$ are spatial means for the respective wind components, $\mathbf{u}_t^E$ and $\mathbf{v}_t^E$ are the component contributions from the thin-fluid approximation, and $\tilde{\mathbf{u}}_t$ and $\tilde{\mathbf{v}}_t$ represent small-scale motions.

We assume that the components $\{\boldsymbol{\mu}_u, \boldsymbol{\mu}_v, \mathbf{u}_t^E, \mathbf{v}_t^E, \tilde{\mathbf{u}}_t, \tilde{\mathbf{v}}_t\}$ are mutually independent. The assumption of independence between the elements of $\mathbf{u}_t$ and $\mathbf{v}_t$ requires physical justification, which is discussed in Section 3.3.2.

Large-Scale Wind Components. The thin-fluid approximation for large-scale tropical dynamics also involves companion approximations. Important among these are the neglect of nonlinear terms in the momentum equations and the simplification of spherical effects to a linear dependence on latitude. These approximations lead to a system referred to in the geophysical literature as the "linear shallow-water equations on the equatorial beta plane" (e.g., Gill 1982; Holton 1992). Looking for solutions in the form of two-dimensional waves in the Cartesian (x, y) plane leads to an ordinary differential equation for $v^E(x, y; t)$, from which corresponding solutions for $u^E(x, y; t)$ can be derived. The solutions for $v^E(x, y; t)$ can be written as

$$v^E(x, y; t) = \sum_p \sum_l v_{l,p}^E(x, y; t), \tag{5}$$

where the $v_{l,p}^E(x, y; t)$ represent the equatorial normal mode (ENM) orthogonal basis set (Matsuno 1966). The waves associated with individual ENMs are identifiable in observations (e.g., Wheeler and Kiladis 1999), and they form the foundation for much of our understanding of tropical dynamics in the atmosphere and ocean.

In practical applications, the infinite series (5) is often truncated to a few leading modes, such that

$$v^E(x, y; t) \approx \sum_{p=1}^{P} \sum_{l=0}^{L} v_{l,p}^E(x, y; t) \tag{6}$$

for some choice of P and L; here we use set of $P = 2$ and $L = 3$, yielding eight modes for v^E. The ENM theory applies to motions with length scales as long as the circumference of the planet. The prediction domain size limits the maximum length scale in our problem to a small fraction of the circumference. In theory, energy can be distributed across an infinity of modes in the series (5). But Wheeler and Kiladis (1999) demonstrated that most of the energy is distributed in clusters of a relatively few modes, suggesting that the truncation used in (6) is not too severe.

It can be shown that each mode can be written as

$$v_{l,p}^E(x, y; t) = V_l(y) \cos(k_p x - \omega_{l,p} t), \tag{7}$$

where $V_l(y)$ describes the north–south structure of the lth mode; $k_p = 2\pi p/D_x$, where p is the east–west wave number and D_x is the east–west domain length; and $\omega_{l,p}$ is the dispersion frequency of the (l, p)th wave mode solution (i.e., it describes the propagation speed and direction of the ENM). Further, the north–south structure can be shown to be proportional to Hermite polynomials that are exponentially damped away from the equator (e.g., Gill 1982),

$$V_l(y) = H_l(y^*) \exp(-.5y^{*2}), \tag{8}$$

where $H_l()$ is the lth Hermite polynomial (with l corresponding to the number of nodes in the north–south direction) and y^* is the "normalized" latitudinal distance from the equator. Specifically, $y^* = \beta_0 y/(\sqrt{gh_e}/\beta_0)^{.5}$, where β_0 is a constant related to the ratio of the earth's angular velocity to its radius, g is the gravitational acceleration, and h_e is the "equivalent depth" of the thin fluid. Of the parameters considered here, D_x (and thus k_p), β_0, and g are fixed and known. The dispersion frequency $(\omega_{l,p})$ and the equivalent depth parameter (h_e) cannot be precisely determined from the thin-fluid approximation theory; however, plausible values can be estimated via data analysis (e.g., Wheeler and Kiladis 1999). In the case of the dispersion frequency, we consider a reparameterization using random components (as discussed later) with priors determined from historical data analysis (see Sec. 3.4.1). For the equivalent depth parameter, it is natural (as Bayesians) to model h_e as random and use this historical information to construct a prior distribution. But in view of the complex way in which h_e enters the model through the Hermite polynomials and the already complicated scope of our model, a fully Bayesian analysis seems prohibitive. We simply set $h_e = 25$ m which is what our prior mean would be based on the discussion of Wheeler and Kiladis (1999). Fortunately, the analysis does not seem particularly sensitive to the value of h_e (see Sec. 4.4).

An elementary trigonometric identity allows us to rewrite (7) as

$$v_{l,p}^E(x, y; t) = \cos(\omega_{l,p} t)[V_l(y)\cos(k_p x)] + \sin(\omega_{l,p} t)[V_l(y)\sin(k_p x)]. \tag{9}$$

In view of the approximations associated with this development, that real winds are very unlikely to propagate like perfect sinusoids as suggested in (9). Furthermore, these expressions were obtained in continuous space and time; our statistical model is for gridded winds defined on a limited domain. To account for such sources of uncertainty, we embed the physical modeling into a stochastic model. Specifically, we replace the leading cosine and sine terms in (9) with random coefficients. That is, for each of our grid points $\mathbf{s}_i \equiv (x_i, y_i)$, $i = 1, \ldots, n$, we let

$$v_t^E(\mathbf{s}_i) = \sum_{p=1}^{P} \sum_{l=0}^{L} \{a_{l,p;1}(t)[V_l(y_i)\cos(k_p x_i)] + a_{l,p;2}(t)[V_l(y_i)\sin(k_p x_i)]\}, \tag{10}$$

where $a_{l,p;1}(t)$ and $a_{l,p;2}(t)$ are assumed to be random coefficients. Allowing these parameters to be random greatly increases the flexibility of our model. In addition, we see that the cosine and sine terms that they replace suggest a natural, physically based prior. The model for the a's is described in the following section.

Our stochastic version of (6) takes the form

$$\mathbf{v}_t^E = \mathbf{\Phi}\mathbf{a}_t^v, \tag{11}$$

where $\mathbf{v}_t^E$ is the vector of v^E winds for all prediction grid locations at time t and $\mathbf{a}_t^v$ is a vector of pairs of a's for each of the $J = P \times (L+1)$ combinations of p and l (recall that $P = 2$ and $L = 3$, so $\mathbf{a}_t^v$ is of length 16). The matrix $\mathbf{\Phi}$ is obtained by evaluating the ENM basis functions at grid points. Specifically, for a total of J combinations, $\mathbf{\Phi}$ is an $n \times 2J$ matrix with columns $\phi_{2(j-1)+1}(x, y) = V_j(y)\cos(k_j x)$ and $\phi_{2(j-1)+2}(x, y) = V_j(y)\sin(k_j x)$ for $j = 1, \ldots, J$, evaluated at the coordinates of the n prediction grid locations. Figure 3 shows the structure of two of these basis functions, $(l, p) = (0, 1)$ and $(l, p) = (2, 1)$. A similar model, $\mathbf{u}_t^E = \mathbf{\Phi}\mathbf{a}_t^u$, is also used.

Small-Scale Wind Components. The small-scale wind components $\tilde{\mathbf{v}}_t$ and $\tilde{\mathbf{u}}_t$ represent scales and types of dynamical processes not explained by the thin-fluid approximation near the equator. We would like these processes to represent the scales that are resolved in the NSCAT sampling and are commonly thought to display multiresolution spatial behavior associated with fractal processes. We chose to represent them in terms of wavelet basis functions with compact support,

$$\tilde{\mathbf{v}}_t = \mathbf{\Psi}\mathbf{b}_t^v, \tag{12}$$

where $\mathbf{b}_t^v$ is an n-vector of temporally evolving random coefficients and $\mathbf{\Psi}$ is an $n \times n$ matrix containing Daubechies wavelet basis functions of order two (evaluated on the prediction grid), modified for closed domains (e.g., Cohen, Daubechies, and Vial 1993); the "order" is the number of vanishing moments of the wavelets. A similar decomposition is specified for $\tilde{\mathbf{u}}_t$.

Our use of wavelets is motivated by the observation that these small-scale processes are typically localized in space and time. The specific choice of the foregoing multiresolution wavelet basis is based on its ability to represent fractal processes (e.g., Wornell 1993). This is critical in attempting to explain the multiscale turbulence structure of wind fields (see Sec. 3.4). Moreover, this wavelet basis has advantages in terms of computational efficiency (see Sec. 4).

Spatial Mean. The spatial mean processes $\boldsymbol{\mu}_v$ and $\boldsymbol{\mu}_u$ account for the climatological mean wind structure. In the tropical western Pacific Ocean, the climatological winds are easterly (i.e., out of the east, toward the west). Note that our domain contains land areas (see Fig. 1). Given that near-surface wind behaves differently over land and sea (e.g., surface heating and/or frictional differences), the spatial mean field should include a dichotomous variable to delineate whether a prediction grid location is over land or sea. Finally, although the climatological wind structure can change with season and horizontal extent, our spatial domain is small enough and our temporal domain short enough (approximately 2 weeks) that we need not consider more complicated spatial or time-varying mean fields in this analysis.

3.3.2 Process Model Specification. The decompositions (3) and (4) and subsequent modeling lead to the statistical models

$$\mathbf{u}_t = \boldsymbol{\mu}_u + \mathbf{\Phi}\mathbf{a}_t^u + \mathbf{\Psi}\mathbf{b}_t^u \tag{13}$$

and

$$\mathbf{v}_t = \boldsymbol{\mu}_v + \mathbf{\Phi}\mathbf{a}_t^v + \Psi\mathbf{b}_t^v. \tag{14}$$

Our hierarchical Bayesian model at this stage requires specification of a parameterized joint distribution

$$[\boldsymbol{\mu}_u, \boldsymbol{\mu}_v, \{\mathbf{a}_t^u\}_0^T, \{\mathbf{a}_t^v\}_0^T, \{\mathbf{b}_t^u\}_0^T, \{\mathbf{b}_t^v\}_0^T | \theta], \tag{15}$$

where $\{\mathbf{a}_t^u\}_0^T$ represents the collection $\{\mathbf{a}_t^u : t = 0, \ldots, T\}$, and so on, and θ generically denotes a collection of parameters to be specified. The crucial point is that the dynamic aspect of our modeling is through time series models for the a and b vectors. We use autoregressive models for these evolutions. Hence we have appended their initial states to the collection of unknowns. Priors for these initial states are discussed at the end of this section.

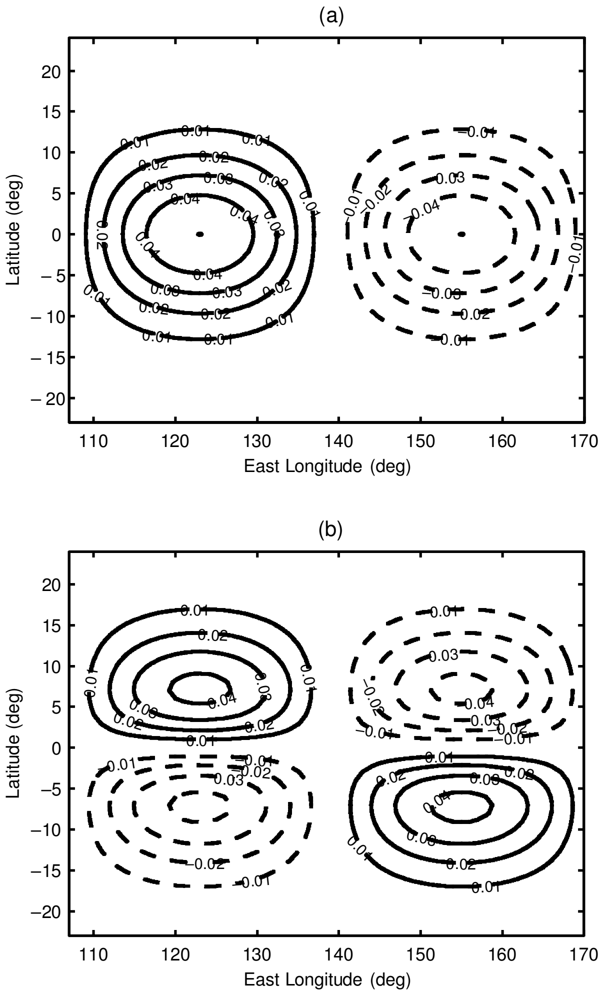

Figure 3. Examples of Shallow-Water Equatorial Normal Mode Basis Functions Used in the Analysis. (a) North–south Hermite mode $l = 0$; East–west Fourier mode domain wavenumber $p = 1$. (b) $l = 2$, $p = 1$.

As noted earlier, a critical modeling assumption is that all six components of the gridded winds in (15) are mutually conditionally independent; that is, (15) is factored as

$$[\boldsymbol{\mu}_u, \boldsymbol{\mu}_v|\theta][\{\mathbf{a}_t^u\}_0^T|\theta][\{\mathbf{a}_t^v\}_0^T|\theta][\{\mathbf{b}_t^u\}_0^T|\theta][\{\mathbf{b}_t^v\}_0^T|\theta]. \quad (16)$$

Our justification of the priori independence assumption is based primarily on physical grounds. The theory of nondivergent two-dimensional turbulence implies that the velocity components are uncorrelated across all spatial scales (e.g., Freilich and Chelton 1986). As discussed in Section 3.4.2, we rely strongly on theoretical and empirical results suggesting that tropical surface wind fields behave like turbulent fields. In addition, Freilich and Chelton (1986) showed that the empirical cross-spectral densities of tropical surface wind components are very small, justifying the general prior modeling assumption of independence. Of course, dependence can arise a posteriori, especially in the presence of physically meaningful structures (e.g., storms). For example, for our 2-week study period, posterior analysis yields a correlation between wind components, averaged over both time and space, of .3.

We next describe the prior distributions in (16). For economy in presentation, we describe in detail only the models for the v-components and hence suppress dependence on v. The models for the u-components were developed similarly and are summarized at the end of this section.

Spatial Mean. We chose a simple spatial regression model for $\boldsymbol{\mu}$,

$$\boldsymbol{\mu} = \mathbf{P}\boldsymbol{\gamma}, \quad (17)$$

where $\mathbf{P}$ is a specified design matrix. In our analysis, this includes an overall intercept term and a land/sea indicator variable (1, land; 0, sea). The regression coefficient vector $\boldsymbol{\gamma}$ is then length 2 and is assigned a bivariate normal prior distribution, $\boldsymbol{\gamma} \sim \text{Gau}(\boldsymbol{\gamma}_o, \boldsymbol{\Sigma}_\gamma)$. The hyperparameters of this distribution were specified based on an ordinary least squares regression of NCEP data from a 4-month period roughly centered around, but excluding, our study period. Specifically, for the v and u components the prior means were $(-.4, .02)$ and $(-2.7, 1.9)$. We assumed that the prior variance-covariance matrices were diagonal with relatively small variances. We used preliminary data analysis in developing these specifications. Because our study period is only 2 weeks long, genuine climatological means (even seasonal means) would not serve well in centering the model. Further, these mean parameters were not of interest in and of themselves; they merely offered a simple method for adjusting for a land-versus-sea effect.

Dynamic Models. One of the key features of our approach is that we seek to model empirically the atmospheric dynamics, so that wind information observed at time t can in principle propagate to nearby locations at time $t+1$, where there may be fewer observations (e.g., see Fig. 2). Thus we assume that the coefficient vectors ($\mathbf{a}$'s and $\mathbf{b}$'s) are conditionally independent and follow first-order Markov vector autoregression (VAR) models: for $t = 1, \ldots, T$,

$$\mathbf{a}_t|\mathbf{H}_a, \mathbf{a}_{t-1}, \boldsymbol{\Sigma}_{\eta_a} \sim \text{Gau}(\mathbf{H}_a\mathbf{a}_{t-1}, \Sigma_{\eta_a}) \quad (18)$$

and

$$\mathbf{b}_t|\mathbf{H}_b, \mathbf{b}_{t-1}, \boldsymbol{\Sigma}_{\eta_b} \sim \text{Gau}(\mathbf{H}_b\mathbf{b}_{t-1}, \Sigma_{\eta_b}), \quad (19)$$

where $\mathbf{H}_a$ and $\mathbf{H}_b$ are VAR parameter matrices for the ENM and wavelet coefficients and $\boldsymbol{\Sigma}_{\eta_a}$ and $\boldsymbol{\Sigma}_{\eta_b}$ are the associated VAR innovation covariance matrices.

To initialize these VAR models, we assumed that $\mathbf{a}_0 \sim \text{Gau}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$ and $\mathbf{b}_0 \sim \text{Gau}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$. The hyperparameters $\boldsymbol{\mu}_a$, $\boldsymbol{\Sigma}_a$, $\boldsymbol{\mu}_b$, and $\boldsymbol{\Sigma}_b$ were specified based on an assumption of mean 0 and diagonal covariance matrices with large variances. Specifically, $\boldsymbol{\Sigma}_a$ was assumed to have variance 100 and $\boldsymbol{\Sigma}_b$ was given prior variance corresponding to the multiresolution scaling discussed in Section 3.4.2.

3.4 Stage 3: Priors on Parameters

We assume that the parameters $\sigma_I^2, \sigma_B^2, \sigma^2, \mathbf{H}_a, \mathbf{H}_b, \boldsymbol{\Sigma}_{\eta_a}$, and $\boldsymbol{\Sigma}_{\eta_b}$ are mutually independent. Similar formulations are used for the parameters relevant to the u-component model.

3.4.1 Autoregressive Parameter Matrices. As suggested by the derivations in Section 3.3.1, to describe wave structures that propagate in time, each pair of coefficients $a_{l,p;1}$ and $a_{l,p;2}$ must be dependent. A simple model for such evolution is a first-order vector autoregression,

$$\begin{bmatrix} a_{l,p;1}(t) \\ a_{l,p;2}(t) \end{bmatrix} = \mathbf{H}_{l,p}^a \begin{bmatrix} a_{l,p;1}(t-\delta_t) \\ a_{l,p;2}(t-\delta_t) \end{bmatrix} + \boldsymbol{\eta}_{l,p}^a(t), \quad (20)$$

where $\mathbf{H}_{l,p}^a$ is a 2×2 *propagator matrix*, the $\boldsymbol{\eta}_{l,p}^a(t)$ are vectors of random innovations, and δ_t is some time interval (.25 days in our case). Application of simple trigonometric identities for $\cos(\omega_{l,p}(t+\delta_t))$ and $\sin(\omega_{l,p}(t+\delta_t))$ suggests physically based prior information for the structure of $\mathbf{H}_{l,p}^a$:

$$\mathbf{H}_{l,p}^a = \begin{bmatrix} \cos(\omega_{l,p}\delta_t) & -\sin(\omega_{l,p}\delta_t) \\ \sin(\omega_{l,p}\delta_t) & \cos(\omega_{l,p}\delta_t) \end{bmatrix}. \quad (21)$$

Given an equivalent depth h_e, $\omega_{l,p}$ can be determined from data analysis. For the v-wind, we used the values suggested by Wheeler and Kiladis (1999) as prior means, namely $\omega_{l,p} = 2\pi[-.133, -.18, -.08, -.05, .67, .59, .75, .75]$ for (l, p) = [(0, 1), (0, 2), (1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3,2)]. Our prior knowledge regarding the last two modes is comparatively uninformed. Note that in some (l, p) combinations, wave modes with the identical horizontal structure (i.e., basis function) can have different propagation characteristics. For simplicity, we chose for our prior the "dominant" wave mode suggested by the data analysis of Wheeler and Kiladis (1999). Thus $\text{vec}(\mathbf{H}_{l,p}^a)$ is specified to be Gaussian with means given by (21) and diagonal covariance structure with relatively large prior variances all set to 100. Sensitivity analysis showed that the posterior wind fields were not sensitive to these specifications. Similar priors were developed for the u-components.

Our specification for the prior on the VAR matrix $\mathbf{H}_b$ is based more on a subjective sense of the dynamics. We expect small-scale features to have some persistence over the 6-hour time intervals considered in this model. However, it is not clear from theory what the prior means and variances should be or whether we should allow spectral interaction. Interaction of the spectral modes would be implied if we allowed nonzero off-diagonal elements in $\mathbf{H}_b$. The added level of complexity required to implement such a formulation was not justified in the current application. Instead, an effective interaction of scales is parameterized by a fractal innovation variance structure, as described in Section 3.4.2. We assume that the elements of $\mathbf{H}_b \equiv \text{diag}([h_b(1), \ldots, h_b(n)]')$ are distributed as independent Gaussians,

$$h_b(j)| \sim N(\mu_{h_b}(j), \sigma^2_{h_b}(j)) : j = 1, \ldots, k_a. \tag{22}$$

For the hyperparameters, we chose $\mu_{h_b}(j) = .4$ and $\sigma^2_{h_b}(j) = .01$ for all j. These values reflect our subjective physical prior of persistence in small-scale modes. Sensitivity analyses on these hyperparameters showed that the posterior wind fields were not extremely sensitive to the specification.

3.4.2 Autoregressive Innovation Covariance Matrices. The VAR conditional covariance matrix $\mathbf{\Sigma}_{\eta_a}$ is assumed to be block diagonal, with J 2×2 covariance matrices, $\mathbf{\Sigma}_{\eta_a}(l, p)$ on the diagonal. For each $l = 0, \ldots, L$ and $p = 1, \ldots, P$, these covariance matrices are assumed to be mutually independent and distributed as

$$\mathbf{\Sigma}_{\eta_a}(l, p)^{-1} \sim W((\kappa_a \mathbf{S}_{\eta_a}(l, p))^{-1}, \kappa_a), \tag{23}$$

where $W()$ is a Wishart distribution with degrees of freedom κ_a and expectation $\mathbf{S}_{\eta_a}(l, p)^{-1}$. For the v-component, these hyperparameters were specified to be $\kappa_a = 2$ and $\mathbf{S}_{\eta_a}(l, p) = \sigma^2_{\eta_a}(l, p)\mathbf{I}$, where $\sigma^2_{\eta_a}(l, p) = (s^2(l, p)/2)[1 - (\cos(\omega_{l,p}\delta_t))^2]$. In this case, $s^2(l, p)$ are climatological variances for each wave mode as observed by Wheeler and Kiladis (1999); that is, $s^2(l, p) = (2{,}133, 2{,}681, 3{,}047, 7{,}922, 305, 335, 200, 200)$, for the eight ENMs used here. The posterior wind fields are not overly sensitive to the choice of these hyperparameters. A similar specification was developed for the u-component portion of the model.

For the wavelet coefficient innovation covariances, we assumed that

$$\mathbf{\Sigma}_{\eta_b} \equiv \text{diag}(\sigma^2_{\eta_b}(1), \ldots, \sigma^2_{\eta_b}(n)). \tag{24}$$

The choice of the hyperparameters were based on physical ideas. The spatial energy spectrum of tropical surface winds has been shown to behave like a self-similar random fractal process (Freilich and Chelton 1986; Wikle et al. 1999), in which the energy spectrum is proportional to the inverse of the spatial frequency taken to some power,

$$S_v(k) \propto \frac{\sigma_v^2}{|k|^d}, \tag{25}$$

where $S_v(k)$ is the spatial energy spectrum of v at spatial frequency k, σ_v^2 is the wind component variance, and d is the decay rate (e.g., Wornell 1993). In the tropical surface wind case, d has been shown to be approximately equal to 5/3 over a broad range (1–1000 km) of spatial scales (Wikle et al. 1999). This spectral decay rate is consistent with famous results from turbulence theory (Kolmogorov 1941a,b; see also Rose and Sulem 1978). It is also a robust empirical result in that recent observational studies of surface winds (Freilich and Chelton 1986; Milliff et al. 1999; Wikle et al. 1999) and winds aloft (Lindborg 1999; Nastrom and Gage 1985) demonstrate a similar power law relation without the conditions for two-dimensional isotropic turbulence and an inertial subrange required by the theory of Kolmogorov. Wornell (1993) derived the relationship for variances of such a fractal process in terms of scales of a wavelet multiresolution analysis. Chin, Milliff, and Large (1998) extended this result to the two-dimensional case by assuming identical distribution of the "diagonal," "horizontal," and "vertical" multiresolution wavelet coefficients. They showed that the variance of two-dimensional wavelet coefficients is proportional to $2^{-l(1+d)-1}$, where l is the level of the multiresolution decomposition ($l = 1, \ldots, N_l$). We use these results, along with the result that the innovation variance for a first-order autoregressive process can be written in terms of the autoregressive coefficient and marginal variance (e.g., $\sigma^2_{\eta_b} = [1 - h_b^2]\sigma_b^2$), to derive the prior variances for each multiresolution level in the η_b process,

$$\sigma^2_{\eta_b}(l) \propto [1 - h_b^2(l)][2^{-l(1+d)-1}], \tag{26}$$

where we substitute the prior mean $\mu_{h_b} = .4$ for $h_b(l)$ and let $d = 5/3$. We use this relationship to determine the inverse gamma (IG) priors,

$$\sigma^2_{\eta_b}(j)|q_{\eta_b}(j), r_{\eta_b}(j) \sim \text{IG}(q_{\eta_b}(j), r_{\eta_b}(j)) \quad : j = 1, \ldots, k_b. \tag{27}$$

That is, we define all spectral indices within a given multiresolution scale (l) to have independent inverse gamma distributions with parameters $q_{\eta_b}(l)$ and $r_{\eta_b}(l)$ determined by assuming a mean given in (26) and a suitable variance. For instance, we give a large variance to the largest wavelet scales, which overlap with the large-scale equatorial modes and can adequately be determined by the data. Alternatively, we assign small (inverse gamma) prior variances for small and medium wavelet scales where observational data are less abundant. This is the most critical prior assumption in the Bayesian analysis! Sensitivity analysis has shown that if we do not give narrow priors on the small- and medium-scale wavelet modes, the posterior spectrum will not follow the 5/3 slope over all spatial scales, as is necessary for realistic wind fields. This is simply because some large spatial regions are not sampled by the scatterometer. Thus, by using the narrow priors, we are in effect constraining the posterior to physical reality, but in such a way that it can be informed by the data, if available.

3.4.3 Measurement Error Variances. The measurement error variances for the data model were assigned inverse gamma distributions $\sigma^2 \sim \text{IG}(q, r)$, $\sigma_I^2 \sim \text{IG}(q_I, r_I)$, and $\sigma_B^2 \sim \text{IG}(q_B, r_B)$. As noted in Section 3.2, Freilich and Dunbar (1999) showed that the NSCAT measurement error variance is approximately 1.7 (m/s)^2. Because we ignored "gridding error" in both space and time, we inflated this value to a prior mean of 2.0 (m/s)^2 and assumed a prior variance of .1. Hence we set $q = 42$ and $r = .0122$. There is little information in the literature concerning NCEP measurement error variances. We have partially accounted for the overly smooth nature of NCEP winds via the $\mathbf{K}_a$ matrix, and so suggest that the measurement error variance should be about the same as found for NSCAT [1.7 $(\mathrm{m/s})^2$] at interior NCEP locations and twice that [3.4 $(\mathrm{m/s})^2$] at boundary grid locations. This latter assumption follows because fewer prediction grid locations are available for the change of support averaging (see Sec. 3.2). However, to reflect our lack of certainty, we assigned larger prior variances (.3) than for the NSCAT variance. Thus we set $q_I = 11.63$, $r_I = .0553$, $q_B = 40.53$, and $q_B = .0074$. Our posterior wind fields were not extremely sensitive to these choices.

4. BAYESIAN ANALYSIS

The fundamental product of a Bayesian analysis is the posterior distribution of all unknowns. Explicit formulas for the posterior distribution for large complicated hierarchical models such as those presented here are intractable. Hence we use a Markov chain Monte Carlo (MCMC) method—specifically, a Gibbs sampler.

4.1 Computation

Our example analysis involve $64 \times 48 \times 54 \approx 166,000$ prediction locations in space (i.e., 64×48) and time (i.e., 54), and we have a large amount of data to ingest into the model ($\sim 200,000$ observations over 14 days). Derivation of the full conditional distributions used in a basic Gibbs sampler implementation is straightforward; the relevant full conditionals are available on the Web at http://www.stat.missouri.edu/~wikle/trop_wind_pap.html. But the high dimensionality of some of these distributions precludes the use of traditional sampling algorithms. For instance, consider the full conditional distribution for the wavelet coefficients,

$$\mathbf{b}_t | \cdot \sim \text{Gau}[\mathbf{A}_t^{-1}\mathbf{g}_t, \mathbf{A}_t^{-1}], \tag{28}$$

for $t = 1, \ldots, T$, where

$$\mathbf{A}_t \equiv (\boldsymbol{\Psi}'\mathbf{K}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{K}_t\boldsymbol{\Psi} + \boldsymbol{\Sigma}_{\eta_b}^{-1} + \mathbf{H}_b'\boldsymbol{\Sigma}_{\eta_b}^{-1}\mathbf{H}_b)^{-1} \tag{29}$$

and

$$\mathbf{g}_t \equiv ((\mathbf{V}_t - \mathbf{K}_t\boldsymbol{\mu}_v - \mathbf{K}_t\boldsymbol{\Phi}\mathbf{a}_t^v)'\boldsymbol{\Sigma}_t^{-1}\mathbf{K}_t\boldsymbol{\Psi} + \mathbf{b}_{t-1}^{v'}\mathbf{H}_b^{v'}\boldsymbol{\Sigma}_{\eta_b}^{-1} + \mathbf{b}_{t+1}^{v'}\boldsymbol{\Sigma}_{\eta_b}^{-1}\mathbf{H}_b)'. \tag{30}$$

Each $\mathbf{A}_t$ is a $3{,}072 \times 3{,}072$ matrix, and many of the matrices from which it is computed are huge (e.g., $\mathbf{K}_t$ can be as large as $3{,}072 \times 6{,}481$). Standard methods for the generation of very high-dimensional multivariate normal random variates (see, e.g., Ripley 1987) are impractical, because we must sample from such high-dimensional distributions for each time t and over many Gibbs iterations. Fortunately, the sparse specification of $\mathbf{K}_t$ can be exploited computationally (e.g., Press, Flannery, Teukolsky, and Vetterling 1986, sec. 2.10). Similarly, the models for temporal evolution parameters (e.g., $\mathbf{H}_b$ and $\boldsymbol{\Sigma}_{\eta_b}$) involve sparse (e.g., diagonal) matrices. Further, computations for the multiresolution wavelet transform are fast (order-n operations). The net result is that matrix multiplications of the form $\mathbf{A}_t\mathbf{w}$, for any n-vector $\mathbf{w}$, can be performed in order-n operations.

To make sampling from such a distribution practical on a high-end workstation, we use iterative linear methods. Specifically, we use a conjugate gradient solver (e.g., Golub and van Loan 1996, sec. 10.2). Details of this sampling approach are given in the Appendix. A key strength of the conjugate gradient approach is that the sparse operations described in the previous paragraph can be exploited. The iterative solver terminates after a preselected convergence criterion is met. The sample obtained is an *approximate* sample from the true full-conditional distribution. We can control the degree of approximation by selecting a more or less rigorous convergence criterion. For the results presented here, we have prescribed a rather rigorous convergence criterion (see the Appendix). If larger spatiotemporal domains are of interest, then trade-off between computation time and the degree of convergence becomes important.

4.2 Gibbs Sampler Convergence

The Gibbs sampler was implemented separately on both the east–west (u) and north–south (v) wind components. (This is valid under all of the conditional independence assumptions described earlier.) Strategies to assess the convergence of a Gibbs sampler in high-dimensional models (e.g., $\sim 10^5$ parameters) such as presented here are not well developed. We base our convergence diagnosis on visual assessment of randomly and subjectively chosen model parameters obtained from pilot simulations with varying starting values. Along with performing a visual assessment, we examined the Gelman and Rubin (1992) convergence monitor. These assessments suggested no reason to reject convergence after about 700 iterations. We then ran a single chain (2,400 iterations) and discarded the first 800 iterations. We based inference on the remaining 1,600 samples.

4.3 Posterior Wind Process

A particularly interesting time period in our data centers on the mature phase of tropical cyclone Dale. In particular, consider the u-component posterior mean wind field for 0000 UTC on November 7, 1996 Figure 4(a) shows the NCEP weather center u-wind component field for this period, Figure 4(b) shows our (estimated) posterior mean for the u-wind component, Figure 4(c) shows the field of posterior means for the u-wind spatial mean plus the equatorial wave modes (i.e., $\boldsymbol{\mu}_u + \boldsymbol{\Phi}\mathbf{a}_t^u$), and Figure 4(d) shows the associated wavelet mode posterior mean component (i.e., $\boldsymbol{\Psi}\mathbf{b}_t^u$). The posterior wind field has significantly more small-scale spatial structure than the NCEP field. Recalling the NSCAT sampling for this period (see Fig. 2), it is clear that there is small-scale structure in regions for which small-scale observations were not available. This is a crucial and desirable feature of our modeling strategy.

4.4 Sensitivity

Assessing sensitivity to our prior/model specifications is extremely difficult because of the model's size and complexity. We performed some sensitivity analyses one parameter at a time, by rerunning the Gibbs sampler with different values for each parameter, albeit with fewer iterations. We would expect interactions among sensitivities of various models and priors on parameters at various levels, but performing "complete factorial" sensitivity experiments is not feasible. We investigated sensitivities mainly by visually inspecting the wind

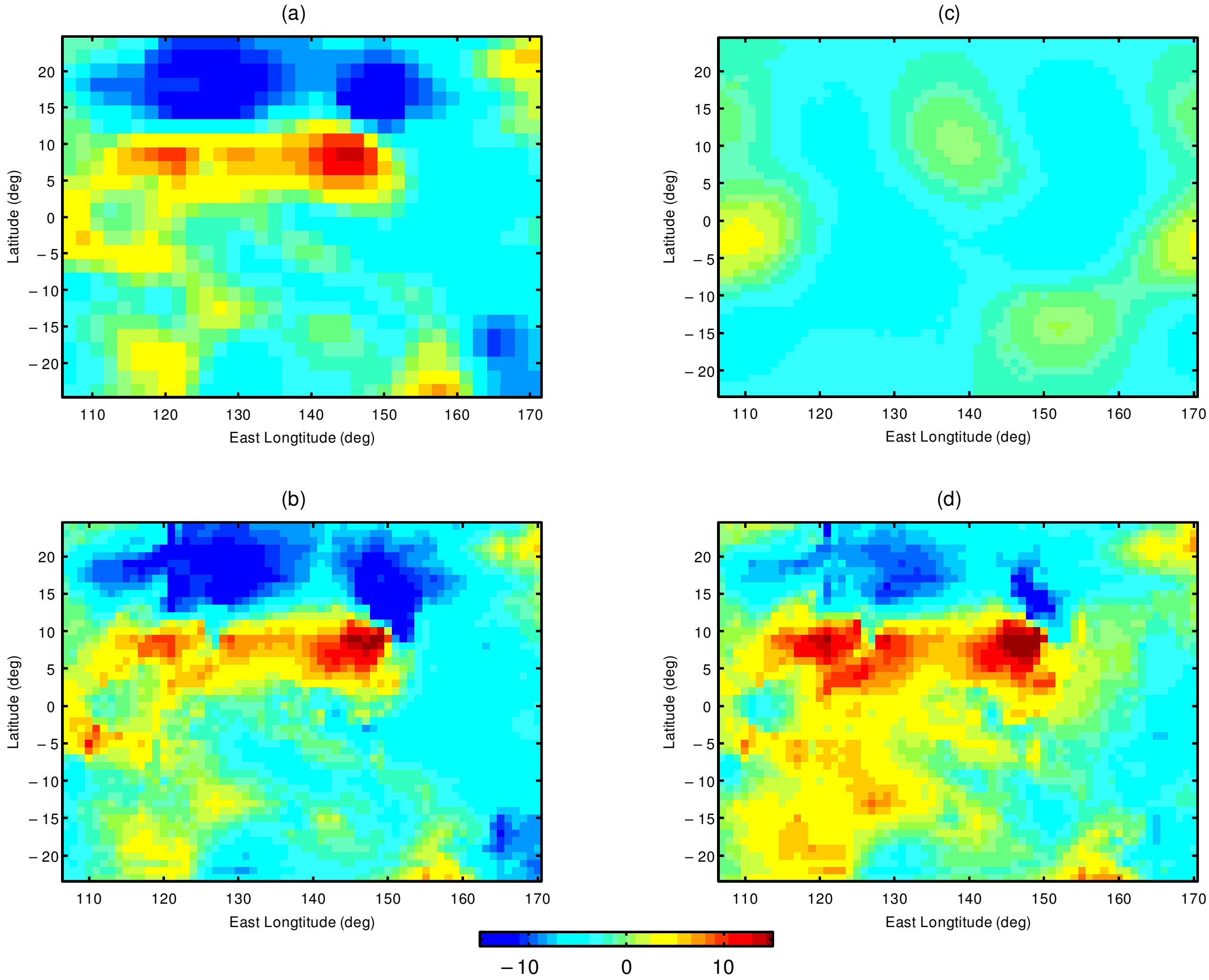

Figure 4. East-West (u) Component of Wind at 0000 UTC on November 7, 1996 (in ms^{-1}). (a) NCEP u-wind component; (b) posterior mean of total "blended" u-wind [i.e., sum of components shown in (c) and (d)]; (c) posterior mean of u-wind spatial mean component ($\boldsymbol{\mu}$) plus equatorial mode components ($\boldsymbol{\Phi a}_t$); (d) posterior mean of wavelet mode u-wind components ($\boldsymbol{\Psi b}_t$).

fields and examining the empirical spatial spectrum of the posterior winds to see how it compared to the desired 5/3 slope discussed in Section 3.4.2. The posterior wind fields are not sensitive to reasonable choices of the equivalent depth h_e, NCEP weighting scheme ($\mathbf{K}_a$), and hyperparameters on measurement error variances. Similarly, the posterior wind fields are not overly sensitive to the hyperparameters for $\boldsymbol{\gamma}$, $\mathbf{H}^a_{l,p}$, $\boldsymbol{\Sigma}_{\eta_a}(l, p)$, and $\mathbf{H}_b$. But as mentioned in Section 3.4.2, the posterior wind fields are very sensitive to the priors on $\boldsymbol{\Sigma}_{\eta_b}$, which must be narrowly centered around the required fractal variances that give the desired 5/3 spatial spectra. This is necessary to ensure proper variability in the posterior winds over areas and time periods where NSCAT sampling is absent.

5. INFERENCE AND MODEL ASSESSMENT

Though again limited by model size and complexity, we considered three "validations": external/physical, internal/physical, and NSCAT data hold out/resample.

5.1 External Physical Verification and Inference

As stated in Section 1, to understand convective processes in the tropical atmosphere, one needs a detailed view of the surface wind field and its horizontal derivatives. Specifically, we consider the *divergence* of the surface wind field. The divergence, defined at a point as $\partial u/\partial x + \partial v/\partial y$, measures the overall rate at which air is being transported away from that point. Conversely, if the sign of the divergence at a location is negative, then air is converging on the point. Convergence at the surface can be related, through a continuity equation, to upward vertical motion. If sufficient moisture is available in the atmosphere, then this rising motion leads to the formation of clouds and, through nonlinear dynamic and thermodynamic processes, the possibility of a tropical storm associated with deep convection. This suggests that external verification of our model would involve comparing cloud imagery with divergence fields calculated from our posterior wind fields.

Figure 5(a) shows wind vectors and gridded estimates of divergence for a subset of the spatial domain at 0000 UTC on November 7, 1996 based on the low-resolution NCEP data only. This period corresponds to the mature phase of tropical cyclone Dale. The NCEP field represents the state-of-the-art wind and divergence fields currently available. Figure 5(b) shows a cloud top (or "brightness") temperature image for the same period as observed from the Japanese GMS

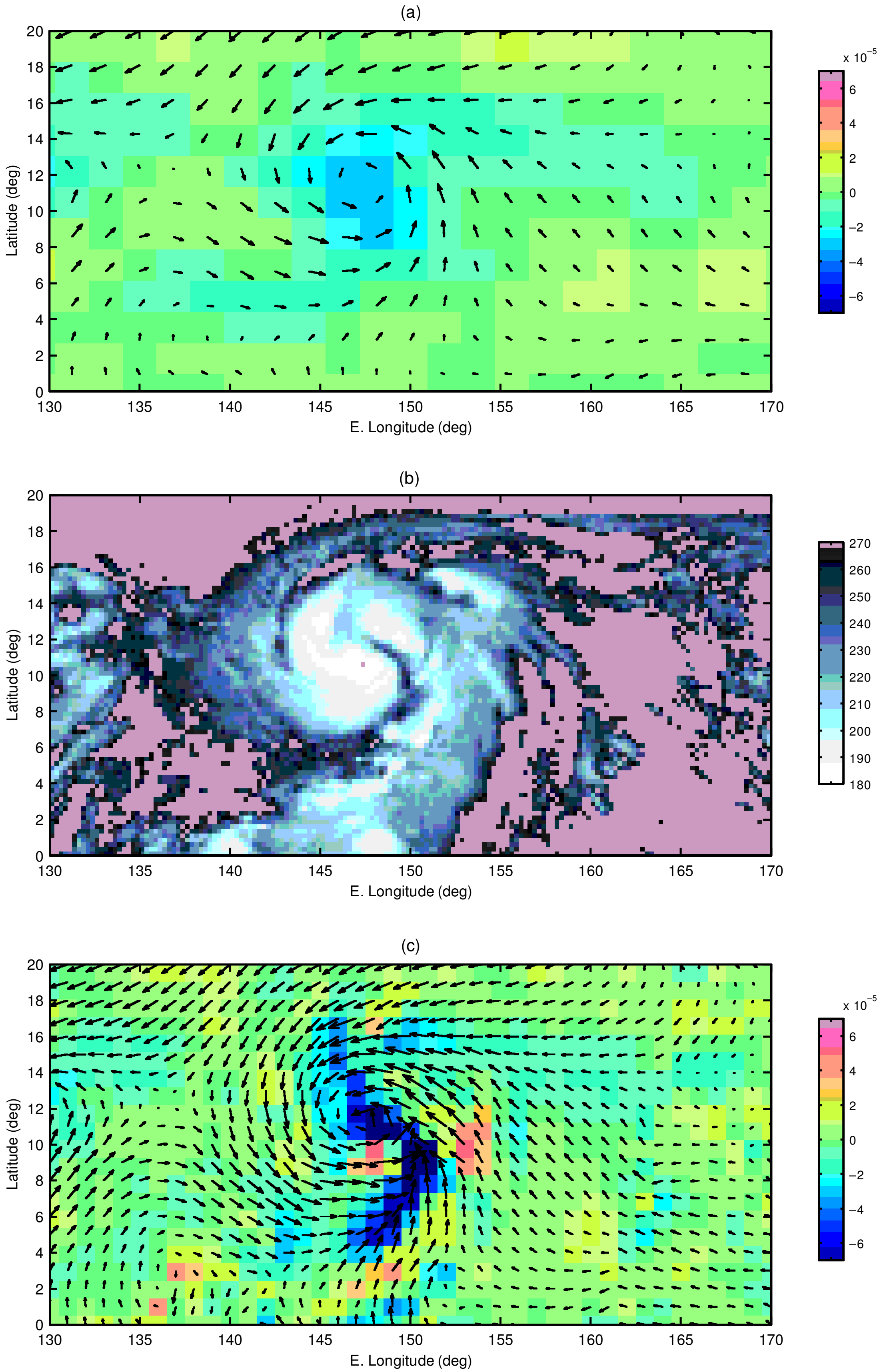

Figure 5. (a) NCEP Divergence (s^{-1}) and Wind Fields (direction of arrows correspond to wind direction, and length corresponds to magnitude) for a Subregion of the Prediction Grid at 0000 UTC on November 7, 1996; (b) Cloud Top Temperature (deg K) Satellite Imagery for the Same Period; (c) Corresponding Blended Posterior Mean Divergence and Wind Fields.

satellite. Colder cloud top temperatures on this plot generally correspond to higher clouds, which in turn are indicative of deep convection and tropical storm activity. Thus areas of clouds in Figure 5(b) should be associated with darker blue areas (convergence) in Figure 5(a). The comparison between the NCEP divergence field and this cloud imagery clearly shows that the NCEP field does not capture the convergence associated with the cloud structures and bands of deep convection associated with the tropical storm. Alternatively, Figure 5(c) shows the posterior mean wind vectors and surface divergence for the same period from our analysis. The use of NSCAT winds and a model capable of space-time propagation have added detail not present in the NCEP analysis. In particular, note the substantial agreement between areas of convergence in the wind field and cloud bands in the tropical cyclone. The physical agreement between convergence and cloud imagery shown here provides very strong physical evidence that the model is performing well.

5.2 Internal Physical Verification

An important check on our model is obtained by examining realizations from the posterior distribution. Figure 6 shows divergence/wind plots for two Gibbs-sampled realizations (widely separated in "Gibbs time") for the cyclone Dale period shown in Figure 5. These realizations are physically realistic, suggesting no reasons for questioning the plausibility of the posterior distribution. Furthermore, Figure 6(c) shows the posterior standard deviation for divergence at this same time. Note that, as expected, the "tracks" of low standard deviation correspond to the satellite sampling paths (see Fig. 2).

5.3 Hold Out/Resample Verification

Although it would be useful to inspect residuals from our model, we do not have residuals in the traditional sense. Our data sources reflect winds at either coarser (NCEP) or much finer (NSCAT) spatial scales. The modeled wind process is *never* observed! Nonetheless, we investigated the model's ability to generate plausible observational data.

Consider the time period represented in Figure 5. We ran a separate Gibbs sampler but left out the NSCAT data for this period. We then compared NSCAT observations to posterior means (and realizations) at the NSCAT locations by mapping the posterior output to those locations via the appropriate $\mathbf{K}_s(t')$. Figure 7(a) shows the relationship when all NSCAT data are included in the analysis. Figure 7(b) shows the result when the NSCAT data for this time period are excluded. Similarly, Figures 7(c) and 7(d) show the same plots, but for a realization from the posterior distribution. Given the amount of data removed (over 5×10^3 observations), the linear associations shown in these figures suggest that the model is reasonable.

6. DISCUSSION

The wind fields from these analysis are currently being used in studies of tropical cyclone development and its relationship to intraseasonal and interseasonal phenomena such as the Madden–Julian oscillation and El Niño, and the seasonal prediction of El Niño. Additional studies of this kind will be possible when the methodology is extended to cover the entire tropical region. We are currently "porting" this model to a supercomputing environment, which will allow such calculations for larger domains. Because the posterior wind fields generated by the current model show realistic small- and medium-scale variability, the results from these analyses can then be used to provide distributional forcing to tropical ocean general circulation models.

APPENDIX: HIGH-DIMENSIONAL MULTIVARIATE NORMAL SAMPLING

Consider the full conditional distribution for some $n \times 1$ vector $\mathbf{x}$,

$$\mathbf{x}|\cdot \sim N(\mathbf{Q}^{-1}\mathbf{g}, \mathbf{Q}^{-1}), \tag{A.1}$$

where $\mathbf{Q} \equiv \boldsymbol{\Psi}'\mathbf{K}'\mathbf{K}\boldsymbol{\Psi} + \mathbf{D}$ is known and has dimensions $n \times n$ and $\mathbf{g}$ is a known $n \times 1$ vector. Define $n \times 1$ random vectors $\mathbf{e}_1, \mathbf{e}_2 \sim$ iid $N(\mathbf{0}, \mathbf{I})$ and let

$$\mathbf{f} \equiv \boldsymbol{\Psi}'\mathbf{K}'\mathbf{e}_1 + \mathbf{D}^{1/2}\mathbf{e}_2. \tag{A.2}$$

Consider the linear system

$$\mathbf{Q}\mathbf{x} = \mathbf{g} + \mathbf{f}. \tag{A.3}$$

Because $\mathbf{Q}$ is invertible by hypothesis and, with probability 1, $\mathbf{g} \neq -\mathbf{f}$, the unique (with probability 1) solution to (A.3) is $\tilde{\mathbf{x}} = \mathbf{Q}^{-1}(\mathbf{g}+\mathbf{f})$. It can be easily shown that $\mathrm{E}(\mathbf{f}) = \mathbf{0}$ and $\mathrm{var}(\mathbf{f}) = \mathbf{Q}$, and hence $\mathrm{E}(\tilde{\mathbf{x}}) = \mathbf{Q}^{-1}\mathbf{g}$ and $\mathrm{var}(\tilde{\mathbf{x}}) = \mathbf{Q}^{-1}$. Thus, with simulated $\mathbf{e}_1$ and $\mathbf{e}_2$, the corresponding solution to (A.3) is a sample from (A.1).

For n very large, we use iterative approaches to solve (A.3), rather than attempt the indicated matrix inversion directly. Specifically, we use the *conjugate gradient* algorithm (e.g., Golub and Van Loan 1996, sec. 10.2). Especially in the case of sparse systems as arising in our model, this approach has computational advantages related to storage and efficiency. The basis of the algorithm is that the solution to (A.3) coincides with the minimizer of the expression

$$\min_{\mathbf{x}} \left\{ \frac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x} + \mathbf{x}'(\mathbf{g}+\mathbf{f}) \right\}. \tag{A.4}$$

Improvements over direct iteration, such as Newton's method or steepest descent, come to mind. The conjugate gradient method is similar, but has the property that all successive differences, $\mathbf{x}^{i+1} - \mathbf{x}^i$, between iterates are mutually $\mathbf{Q}$-orthogonal (or "conjugate"); that is, $(\mathbf{x}^{i+1} - \mathbf{x}^i)'\mathbf{Q}(\mathbf{x}^{j+1} - \mathbf{x}^j) = 0$.

As with most iterative procedures, a key computational issue is the rapid computation of powers of $\mathbf{Q}$. Indeed, we can write (A.3) as

$$(\boldsymbol{\Psi}'\mathbf{K}'\mathbf{K}\boldsymbol{\Psi} + \mathbf{D})\mathbf{x} = \mathbf{g} + \boldsymbol{\Psi}'\mathbf{K}'\mathbf{e}_1 + \mathbf{D}^{1/2}\mathbf{e}_2, \tag{A.5}$$

where $\mathbf{D}^{1/2}$ is sparse for our models. Thus we do not have to store $\mathbf{Q}$, and must only perform a series of vector multiplications. By making use of sparseness from our hierarchical implementation and spectral and multiresolution representations, we can do these multiplications very efficiently (e.g., in our case $\boldsymbol{\Psi}\mathbf{x}$ corresponds to the inverse discrete wavelet transform).

With an iterative approach, a choice must be made as to starting values. We typically use the value for the previous Gibbs iteration or the one-step-ahead "prediction" from the appropriate Markov model. Furthermore, although the conjugate gradient algorithm is known to converge to the solution in at most n steps, n is far too large for the algorithm to be run convergence for each MCMC iteration. Hence one must choose an approximate-convergence criterion. In our implementation, this criterion is specified to be $\varepsilon\|\mathbf{g}+\mathbf{f}\|$, where $\varepsilon = .0005$; this criterion is usually met after 15–30 iterations.

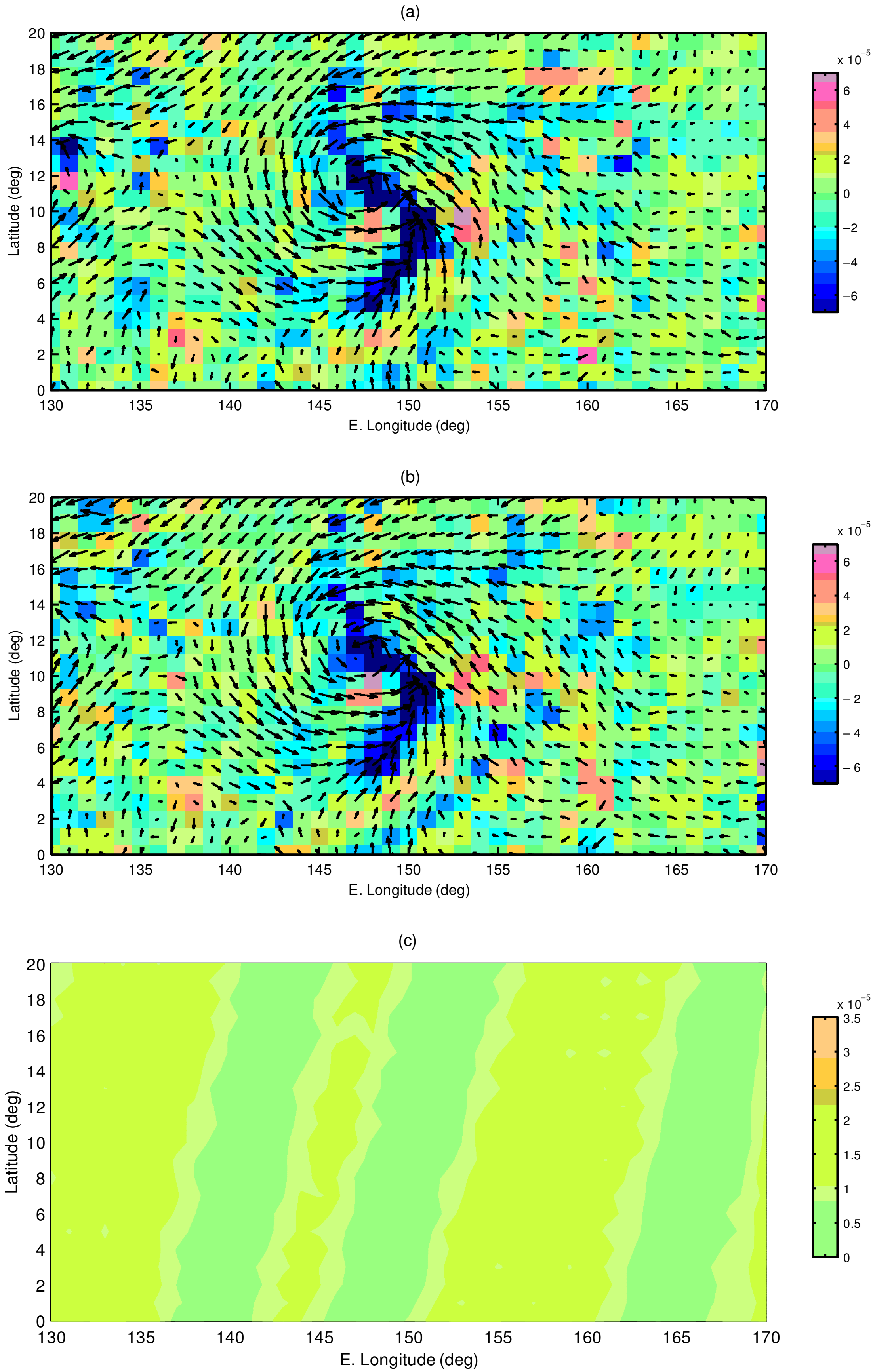

Figure 6. *(a), (b) Wind and Divergence Field Realizations From the Posterior Distribution at 0000 UTC on November 7, 1996; (c) Posterior Standard Deviation for Divergence (s^{-1}) at the Same Time.*

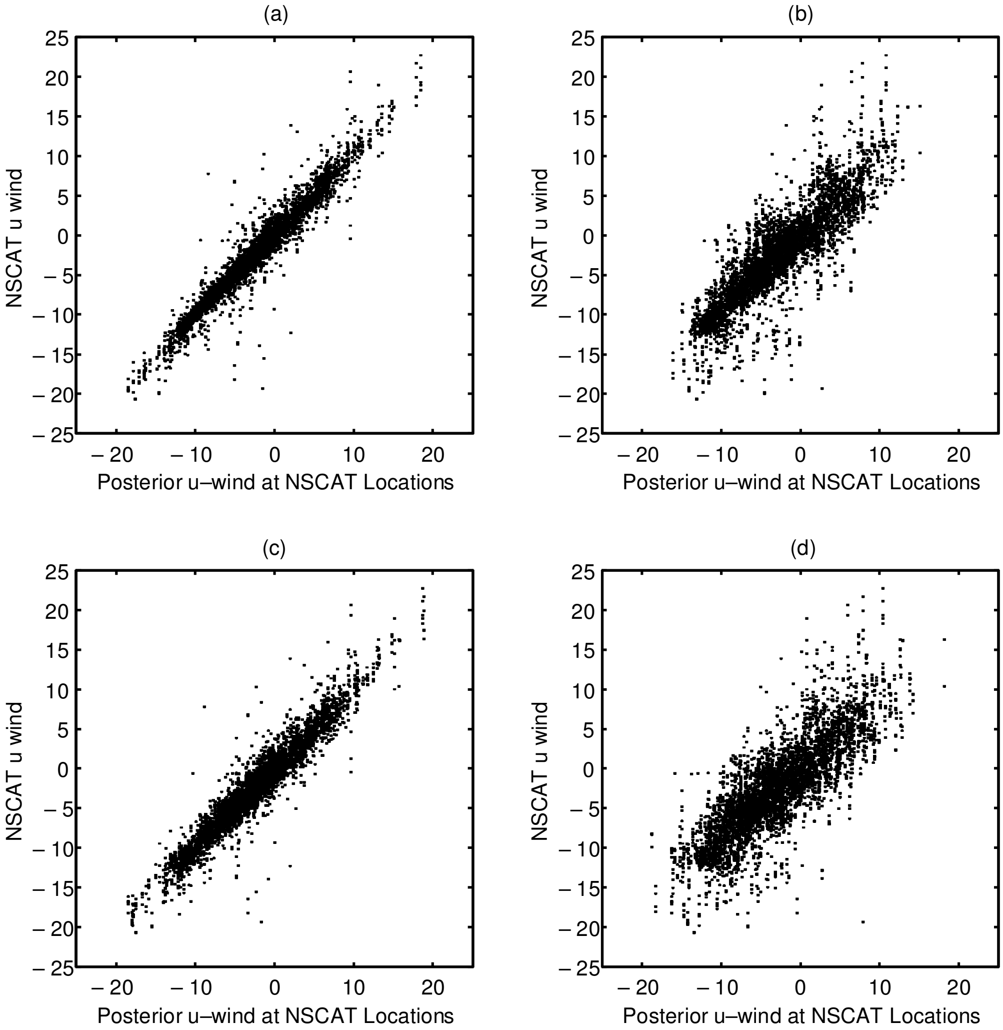

Figure 7. NSCAT u-wind Component at 0000 UTC November 7, 1996 Versus the Posterior u-Wind Downscaled to NSCAT Locations; (a) Data Versus Posterior Mean With NSCAT Data Included for This Period; (b) Data Versus Posterior Mean with NSCAT Data Deleted for This Period; (c) Same as (a), Except That a Realization From the Posterior Is Used; (d) same as (b), Except That a Realization From the Posterior Is Used. All data are in m/s.

[Received July 1998. Revised September 2000.]

REFERENCES

Berliner, L. M. (1996), "Hierarchical Bayesian Time Series Models," in *Maximum Entropy and Bayesian Methods*, eds. K. M. Hanson and R. N. Silver, Kluwer, Dordrecht: pp. 15–22.

Chen, D., Liu, W. T., and Witter, D. (1999), "Sensitivity of the Tropical Pacific Ocean Simulation to the Temporal and Spatial Resolution for the Wind Forcing," *Journal of Geophysical Research, Oceans*, 104, 11261–11272.

Chin, T. M., Milliff, R. F., and Large, W. G., (1998), "Basin-Scale, High-Wavenumber Sea Surface Wind Fields From a Multiresolution Analysis of Scatterometer Data," *Journal of Atmospheric and Oceanic Technology*, 15, 741–763.

Cohen, A., Daubechies, I., and Vial, P. (1993), "Wavelets on the Interval and Fast Wavelet Transforms," *Applied and Computational Harmonic Analysis*, 1, 54–81.

Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge, U.K.: Cambridge University Press.

Freilich, M. H. (1997), "Validation of Vector Magnitude Datasets: Effects of Random Component Errors," *Journal of Atmospheric and Oceanic Technology*, 14, 695–703.

Freilich, M. H., and Chelton, D. B. (1986), "Wavenumber Spectral of Pacific Winds Measured by the Seasat Scatterometer," *Journal of Physical Oceanography*, 16, 741–757.

Freilich, M. H., and Dunbar, R. S. (1999), "The Accuracy of NSCAT-1 Vector Winds: Comparisons With National Data Buoy Center Buoys," *Journal of Geophysical Research, Oceans*, 104, 11231–11246.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science* 7, 457–472.

Gill, A. E. (1982), *Atmosphere-Ocean Dynamics*, New York: Academic Press.

Golub, G. H., and van Loan, C. F. (1996), *Matrix Computations* (3rd ed.), Baltimore: John Hopkins University Press.

Gray, W. M. (1976), "Global View of the Origin of Tropical Disturbances and Storms," *Monthly Weather Review*, 96, 669–700.

Haslett, J., and Raftery, A. E. (1989), "Space-Time Modelling With Long-Memory Dependence: Assessing Ireland's Wind Power Resource," *Applied Statistics*, 38, 1–50.

Holton, J. R. (1992), *An Introduction to Dynamic Meteorology*, (3rd ed.), San Diego: Academic Press.

Kerr, R. A. (1998), "Models Win Big in Forecasting El Niño," *Science*, 280, 522–523.

Kolmogorov, A. N. (1941a), "The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds Numbers," ISSN 0002-3264, *Doklady Akademii Nauk SSSR*, 30, 301–305.

Kolmogorov, A. N. (1941b), "On Degeneration of Isotropic Turbulence in an Incompressible Viscous Liquid," ISSN 0002-3264 *Doklady Akademii Nauk SSSR*, 31, 538–541.

Large, W. G., Holland, W. R., and Evans, J. C. (1991), "Quasigeostrophic Ocean Response to Real Wind Forcing: The Effects of Temporal Smoothing," *Journal of Physical Oceanography*, 21, 998–1017.

Lindborg, E. (1999), "Can the Atmospheric Kinetic Energy Spectrum be

Explained by Two-Dimensional Turbulence?," *Journal of Fluid Mechanics*, 388, 259–288.

Liu, W. T., Tang, W., and Hu, H. (1998), "Spaceborne Sensors Observe El Niño's Effects on Ocean and Atmosphere in the North Pacific," *EOS, Transactions of the American Geophysical Union*, 79, 249–252.

Madden, R. A., and Julian, P. R. (1994), "Observations of the 40 to 50-day Tropical Oscillation—A Review," *Monthly Weather Review*, 122, 814–837.

Matsuno, T. (1966), "Quasi-Geostrophic Motions in the Equatorial Area," *Journal of the Meteorological Society of Japan*, 41, 25–42.

Milliff, R. F., Large, W. G., Holland, W. R., and McWilliams, J. C. (1996), "The General Circulation Responses of High-Resolution North Atlantic Ocean Models to Synthetic Scatterometer Winds," *Journal of Physical Oceanography*, 26, 1747–1768.

Milliff, R. F., Large, W. G., Morzel, J., Danabasoglu, G., and Chin, T. M. (1999), "Ocean General Circulation Model Sensitivity to Forcing From Scatterometer Winds," *Journal of Geophysical Research, Oceans*, 104, 11337–11358.

Milliff, R. F., and Morzel, J. (2001), "The Global Distribution of the Time-Average Wind-Stress Curl From NSCAT," *Journal of Atmospheric Science*, 58, 109–131.

Naderi, F. M., Freilich, M. F., and Long, D. G. (1991), "Spaceborne Radar Measurement of Wind Velocity Over the Ocean—An Overview of the NSCAT Scatterometer System," *Proceedings of the IEEE*, 79, 850–866.

Nastrom, G. D., and Gage, K. S. (1985), "A Climatology of Atmospheric Wavenumber Spectra of Wind and Temperature Observed by Commercial Aircraft," *Journal of the Atmospheric Sciences*, 42, 950–960.

Philander, G. S. (1990), *El Niño, La Niña, and the Southern Oscillation*. San Diego: Academic Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986), *Numerical Recipes: The Art of Scientific Computing*, Cambridge, U.K.: Cambridge University Press.

Ripley, B. D. (1987), *Stochastic Simulation*, New York: Wiley.

Rose, H. A., and Sulem, P. L. (1978), "Fully Developed Turbulence and Statistical Mechanics," *Journal de Physique*, 39, 441–482.

Stoffelen, A., and Anderson, D. (1997), "Scatterometer Data Interpretation; Measurement Space and Inversion," *Journal of Atmospheric and Oceanic Technology*, 14, 1298–1313.

Waller, L. A., Carlin, B. P. Xia, H., and Gelfand, A. E. (1997), "Hierarchical Spatiotemporal Mapping of Disease Rates," *Journal of the American Statistical Association*, 92, 607–617.

Wentz, F. J., and Freilich, M. H. (1997), "Finalizing the NSCAT-1 Model and Process Towards NSCAT-2," in *Proceedings of the NASA Scatterometer Science Symposium*, NSCAT Project, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, pp. 61–63.

Wentz, F. J., and Smith, D. K. (1999), "A Model Function for the Ocean-Normalized Radar Cross-Section at 14 GHz Derived From NSCAT Observations," *Journal of Geophysical Research, Oceans*, 104, 11499–11514.

Wheeler, M., and Kiladis, G. N. (1999), "Convectively-Coupled Equatorial Waves: Analysis of Clouds and Temperature in the Wavenumber-Frequency Domain," *Journal of the Atmospheric Sciences*, 56, 374–399.

Wikle, C. K., and Berliner, L. M. (2001), "Change of Support in Hierarchical Spatial Models," technical report, University of Missouri, Dept. of Statistics.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998), "Hierarchical Bayesian Space-Time Models," *Environmental and Ecological Statistics*, 5, 117–154.

Wikle, C. K., and Cressie, N. (1999), "A Dimension-Reduction Approach to Space-Time Kalman Filtering," *Biometrika*, 86, 815–829.

Wikle, C. K., Milliff, R. F., and Large, W. G. (1999), "Observed Wavenumber Spectra From 1 to 1,000 km of Near-Surface Winds During the TOGA COARE IOP," *Journal of Atmospheric Science*, 56, 2222–2231.

Wornell, G. W. (1993), "Wavelet-Based Representations for the 1/f Family of Fractal Processes," *Proceedings of the IEEE*, 81, 1428–1450.