Bayesian hierarchical modeling of air-sea interaction

L. Mark Berliner
Department of Statistics, Ohio State University, Columbus, Ohio, USA

Ralph F. Milliff
Colorado Research Associates, A Division of NorthWest Research Associates, Bellevue, Washington, USA

Christopher K. Wikle
Department of Statistics, University of Missouri, Columbia, Missouri, USA

Received 27 March 2002; revised 27 August 2002; accepted 30 August 2002; published 1 April 2003.

[1] By means of Bayesian hierarchical modeling (BHM), we develop a model for aspects of vigorous air-sea interactions on a basin scale. The approach relies on both physical reasoning and statistical techniques for data processing and uncertainty management. The theory and current practice of BHM of air-sea interaction physics will be introduced and demonstrated in this paper. The demonstration is in the context of an observing system simulation experiment. An ocean "truth" simulation is driven by idealized surface winds in a testbed domain abstracted from the Labrador Sea. Artificial observations analogous to scatterometer and altimeter data are supplied to a BHM and comparisons made with the evolution of the "truth" simulation over a ten day experiment. Substantial attention is devoted to description of computational techniques. *INDEX TERMS:* 4504 Oceanography: Physical: Air/sea interactions (0312); 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; 3210 Mathematical Geophysics: Modeling; *KEYWORDS:* data assimilation, Monte Carlo, quasigeostrophy, uncertainty

Citation: Berliner, L. M., R. F. Milliff, and C. K. Wikle, Bayesian hierarchical modeling of air-sea interaction, *J. Geophys. Res.*, *108*(C4), 3104, doi:10.1029/2002JC001413, 2003.

1. Introduction

[2] The combination of model and data is a familiar problem in the atmosphere and ocean sciences. For example, rich theories and sophisticated algorithms have been developed under the rubric of data assimilation [e.g., *Bennett*, 1993; *Daley*, 1991; *Robinson et al.*, 1998], and this continues to be an area of intense research. Three of the most outstanding issues are (1) the quantification and reflection of uncertainty in the results; (2) incorporation of diverse observational data sets relevant to different processes and coupling models for interacting processes (e.g., atmosphere and ocean); and (3) inclusion of stochastic elements to adjust for model uncertainty, unmodeled aspects of the problem, etc. As is recognized in much of the data assimilation literature, Bayesian modeling is useful in the development and analysis of models that incorporate physical reasoning and respond to observations while accounting for the uncertainties in both. The Bayesian framework, particularly when formulated hierarchically, provides an inference engine that can address the three challenges above.

[3] First, in the Bayesian view, all unknowns are modeled as random variables; that is, a fundamental property of Bayesian modeling is that model inputs and outputs are probability distributions. In the context of space-time gridded models analogous to finite difference approaches to the solution of differential equations, this means that at each point in a domain, and for every output time step, we can obtain estimates of a mean value and measures of uncertainty for every predicted variable, as well as measures of dependence (e.g., covariances) among all variables in the model system. Further, we can produce estimates and uncertainty measures for arbitrary (e.g., nonlinear) functions of the modeled variables. The mechanism for producing these output distributions, known as posterior distributions, is Bayes' Theorem. This theorem is a recipe for updating prior distributions for quantities of interest based on observational data. The construction of prior distributions relies on our knowledge about the variables, but reflects uncertainties in such knowledge.

[4] Regarding the second challenge, Bayesian hierarchical modeling is an application of principles of conditional probability theory. As we will indicate, this leads to models that are adaptable to multiplatform observational data sets including temporally sparse, but spatially abundant, remote sensing observations. The reliance on conditional probability theory also enables the coupling of different process models as well as probabilistic treatment of uncertain boundary and initial conditions. In our example here, the essence of the prior modeling links a parameterized, stochastic model for near-surface atmospheric winds and a parameterized, stochastic ocean-streamfunction model that is developed conditional on atmospheric winds.

0148-0227/03/2002JC001413

[5] As perspective on the third challenge, in our application of Bayes' Theorem, the definition of probabilistic models and methods for computation differ from those described recently in the data assimilation literature [e.g., *van Leeuwen and Evensen*, 1996; *Miller et al.*, 1999; *Evensen and van Leeuwen*, 2000; *Anderson*, 2001; *Pham*, 2001; *van Leeuwen*, 2001]. The sources of uncertainty in many of these applications include initial and boundary conditions as well as model uncertainty. Model uncertainty is often reflected by the addition of a stochastic error term to equations that are otherwise identical to deterministic approximations of continuous, governing partial differential equations. The BHM view includes such modeling but also enables extensions. First, the distributions of model errors involve parameters (e.g., model error covariances) which are themselves unknown. Further, analogues of diffusion coefficients and other quantities arising in forward models can be treated as unknown. Such unknown parameters are viewed as random variables endowed with prior distributions and hence subject to updating in light of the observations. In this fully probabilistic framework, the various sources of uncertainty can be treated formally in computing methods quite different from deterministic model integration (see *Scipione and Berliner* [1993] for an example based on the Lorenz system). In the development and implementation of dynamical priors in the air-sea model described here, we use spatial operators with discrete forms similar to those of conventional forward models. However, these priors are not forward models as in the sense of practical implementations of, say, the extended Kalman filter in data assimilation. Time steps and spatial grid dimensions in the BHM are larger than stability criteria would permit in a forward model sense. That is because we do not solve approximations to stiff, quasi-elliptic, partial differential equations. Nevertheless, some notions of stability are relevant. We return to this issue in section 5.

[6] We implement a Bayesian Hierarchical Model (BHM) in a testbed configuration in order to demonstrate the approach, and to indicate characteristics of a BHM in the context of an observing system simulation experiment (OSSE). The OSSE consists of an ocean "truth" simulation, driven by idealized surface wind fields. The ocean simulation is developed in the context of primitive equation, shallow water equations (PE-SWE), using the model described by *Milliff and McWilliams* [1994]. The ocean "truth" and idealized atmospheric forcing are subsampled in space and time and corrupted with artificial noise to supply the BHM with simulated observations analogous to altimeter and scatterometer data. The Bayesian model itself is based on quasi-geostrophic modeling as opposed to PE-SWE. Our intent is not to assess nor to suggest quasi-geostrophic modeling as an approximation. Rather, we simply seek to mimic the practical issue that no feasible model can be claimed to capture perfectly the truth in a real application.

[7] For Bayesian models as complex as those developed here, exact calculations of the posterior distributions are infeasible, leading to the use of simulation-based approaches. The computational method used here is a combination of two powerful techniques: Importance Sampling Monte Carlo (ISMC) and Markov chain Monte Carlo (MCMC). We present these methods in two appendices. Appendix A reviews these techniques in a comparative fashion. Specifically, we develop popular ensemble smoothing methods from the perspective of ISMC. Appendix B is a technical description of the actual algorithms used in this article.

[8] To clarify both the strategy and the specifics, we present the modeling in two steps. In section 2 we provide a comparatively nontechnical overview of Bayesian analysis and our hierarchical modeling for the air-sea problem. Section 3 is devoted to the actual specification and analyses for the model applied to the air-sea problem. Section 4 describes results from the OSSE. Discussion is provided in section 5.

2. Bayesian Analysis and Model Overview

2.1. Basics of Bayesian Analysis

[9] The following notation is used to indicate probability distributions: for random variables X and D, $[X, D]$ denotes their joint distribution; $[D|X]$ is the conditional distribution of D given X. The goal is production of probability distributions that combine our scientific understanding of the phenomena of interest, X (including both processes and parameters), with the information contained in observational data, D. The steps leading to the analysis are (1) scientific understanding and relevant past data concerning X are summarized in a prior distribution $[X]$ for X; (2) quantification of the random nature of the data D and their relationships to X are developed in a data model consisting of the probability distribution of D given X: $[D|X]$; and (3) Bayes' Theorem yields the updated distribution for X having observed D. The result, known as the posterior distribution, is the conditional distribution for X given D,

$$[X|D] = \frac{[D|X]\,[X]}{[D]}. \tag{1}$$

The denominator in this expression is the integral of the numerator with respect to X. It is a normalizing constant insuring that $[X|D]$ has total probability equal to 1. General reviews of Bayesian statistics are given by *Berger* [1985] and *Bernardo and Smith* [1994]. Also, see *Epstein* [1985] and *Tarantola* [1987] for introductions targeted to geophysical applications, and see *Wikle et al.* [1998], *Berliner et al.* [1999, 2000], and *Wikle et al.* [2001] for examples of Bayesian analyses in geophysical settings.

2.2. Bayesian Hierarchical Modeling

[10] The simplicity of Bayes' formula belies the rich potential for application. Extremely complex probability models can be formulated, particularly via the strategy of hierarchical modeling. Our use of the word "hierarchical" involves the formulation of probability models for large collections of random variables as products of conditional probability distributions. We note that one can adopt the hierarchical view without being explicitly Bayesian and vice versa. Also, the common use of the appellation hierarchical in the geosciences refers to a sequence of increasingly complex models for the phenomenon of interest. While this view can be included in the statistical definition, the latter definition also includes a variety of other modeling strategies. Rather than delving into the

theory, we turn to illustration in our air-sea modeling context. We use the following variables:

w atmospheric winds, later prescribed as gridded, horizontal wind vectors (w stands for "wind," rather than referring to vertical velocity);

ψ oceanic variables, later developed as gridded, ocean streamfunction values;

D_w atmospheric wind data (e.g., scatterometer winds estimates);

D_ψ ocean data (e.g., altimetry data).

All these variables are distributed in space and time, though the extra notation is suppressed for now.

[11] Production of a stochastic model is the construction of a joint probability distribution for all unknown quantities, in our case D_w, D_ψ, w, ψ, and model parameters. For each of the four collections of variables, we introduce four sets (not necessarily disjoint) of model parameters (to be specified later): Let θ_w, θ_ψ denote parameters associated with the statistics of the data collection methods (e.g., measurement error variances), and let η_w, η_ψ represent two sets of model parameters associated with the probability models for the atmospheric and oceanic processes.

[12] Bayes' Theorem (equation (1)) provides the posterior distribution of the unknowns conditional on the observed data, $[w, \psi, \eta_w, \eta_\psi, \theta_w, \theta_\psi | D_w, D_\psi]$. Application of the result requires specification of the joint distribution of all random quantities: D_w, D_ψ, w, ψ, η_w, η_ψ, θ_w, and θ_ψ. This is a formidable task if attempted directly, so we suggest a hierarchical strategy. The following view of the primary components of a hierarchical model has proven useful [*Berliner*, 1996]:

Data model

$$\left[D_w, D_\psi | w, \psi, \theta_w, \theta_\psi\right],$$

Process model

$$\left[w, \psi | \eta_w, \eta_\psi\right],$$

Parameter model

$$\left[\theta_w, \theta_\psi, \eta_w, \eta_\psi\right].$$

The hierarchy specifies the desired joint model as the product

$$\left[D_w, D_\psi, w, \psi, \eta_w, \eta_\psi, \theta_w, \theta_\psi\right] = \left[D_w, D_\psi | w, \psi, \theta_w, \theta_\psi\right]\left[w, \psi | \eta_w, \eta_\psi\right] \cdot \left[\theta_w, \theta_\psi, \eta_w, \eta_\psi\right]. \quad (2)$$

That such a representation can be written is a simple fact from probability theory; specifications of the components of the hierarchy are our challenges.

2.2.1. Data Model: Combining Data Sets

[13] We anticipate the use of complex data sets replete with spatio-temporal dependence structures (e.g., strong interrelationships between atmospheric and oceanic data). An advantage of modeling the conditional distribution of the data given the true values of the processes they represent is that major simplifications in model form are typically quite plausible. In our case, we assume that the two chief data sets are conditionally independent. Our data model takes the form

$$\left[D_w, D_\psi | w, \psi, \theta_w, \theta_\psi\right] = \left[D_w | w, \theta_w\right]\left[D_\psi | \psi, \theta_\psi\right]. \quad (3)$$

Note that conditional on w and ψ, we view the data model as primarily representing random, but unobserved, measurement errors. Further, the measurement errors made in developing the scatterometer data and those made in developing the altimeter data are independent. This is not the same as suggesting that the two data sets are unconditionally independent. Rather, the lion's share of the dependence among these data sets is thought to arise through the dependence of the true processes (e.g., winds and ocean streamfunction). This dependence is then modeled for the true processes in later modeling steps.

[14] Also, note that we assumed that conditional on w, θ_w, the distribution of D_w does not depend upon ψ, and we made an analogous assumption for D_ψ. In general such conditional independence assumptions are not unequivocal, and should be defended and assessed on a case-by-case basis. Finally, such assumptions are not required in Bayesian analyses, though they are often quite reasonable and lead to major simplifications in the computations.

2.2.2. Process and Parameter Prior Models

[15] Thinking hierarchically, we develop a process-model prior $[w, \psi | \eta_w, \eta_\psi]$ as the product

$$\left[w, \psi | \eta_w, \eta_\psi\right] = \left[\psi | w, \eta_\psi\right]\left[w | \eta_w\right]. \quad (4)$$

In words, the joint air-sea prior is the product of an ocean-given-atmosphere model and an atmospheric model. The construction and subsequent analysis of the two distributions on the right-hand side of equation (4) are primary themes of this article.

[16] Define $\mathbf{\Psi}$ to be gridded (in space and time), vectorized ocean streamfunction defined on a bounded domain. We partition $\mathbf{\Psi} = (\mathbf{\Psi_I}, \mathbf{\Psi_B})$ where $\mathbf{\Psi_I}$ represents values in the interior of the domain and $\mathbf{\Psi_B}$ are boundary values. Also, let $\mathbf{U}$ ($\mathbf{V}$) be similarly gridded zonal (meridional) wind components. Our tasks are (1) to build stochastic prior models for the evolution of the processes, $(\mathbf{U}^t, \mathbf{V}^t, \mathbf{\Psi}^t)$: $t \in \mathcal{T}\}$, where $\mathcal{T} = \{1, \ldots, T\}$ represents discrete time steps, and (2) to update the prior based on observations. For convenience, let

$$\mathbf{U} = \{\mathbf{U}^t : t \in \mathcal{T}\}; \mathbf{V} = \{\mathbf{V}^t : t \in \mathcal{T}\}; \mathbf{\Psi} = \{\mathbf{\Psi}^t : t \in \mathcal{T}\}. \quad (5)$$

[17] We formulate a (first-order) Markovian evolution model for streamfunction conditional on winds. That is, we assume that for each $t + 1$, conditional on past values of the winds and the model parameters, the distribution of $\mathbf{\Psi}^{t+1}$ depends on the past streamfunction only through $\mathbf{\Psi}^t$ and on past winds only through $\mathbf{U}^t$ and $\mathbf{V}^t$. (Note, for example, that this does not mean that $\mathbf{\Psi}^t$ and $\mathbf{\Psi}^{t+2}$ are assumed to be independent.) Symbolically, we develop a probability distribution, $[\mathbf{\Psi}^{t+1} | \mathbf{\Psi}^t, \mathbf{U}^t, \mathbf{V}^t, \eta_\psi]$. It follows that the more formal form of $[\psi | w, \eta_\psi]$ is given by

$$\left[\mathbf{\Psi} | \mathbf{U}, \mathbf{V}, \eta_\psi\right] = \left[\mathbf{\Psi}^1 | \eta_\psi\right] \prod_{t=1}^{T-1} \left[\mathbf{\Psi}^{t+1} | \mathbf{\Psi}^t, \mathbf{U}^t, \mathbf{V}^t, \eta_\psi\right], \quad (6)$$

where $[\mathbf{\Psi}^1|\eta_\psi]$ provides temporal initialization of the stochastic model.

[18] A further factorization of $[\mathbf{\Psi}^{t+1}|\mathbf{\Psi}^t, \mathbf{U}^t, \mathbf{V}^t, \eta_\psi]$ is used to deal with dependence of the interior on the boundary as well as the evolution of the boundary [also see *Wikle et al.*, 2003]. Specifically, we formulate the component models

$$\left[\mathbf{\Psi}^{t+1}|\mathbf{\Psi}^t, \mathbf{U}^t, \mathbf{V}^t, \eta_\psi\right] = \left[\mathbf{\Psi}_I^{t+1}|\mathbf{\Psi}_B^{t+1}, \mathbf{\Psi}^t, \mathbf{U}^t, \mathbf{V}^t, \eta_\psi\right] \times \left[\mathbf{\Psi}_B^{t+1}|\mathbf{\Psi}^t, \mathbf{U}^t, \mathbf{V}^t, \eta_\psi\right]. \quad (7)$$

Here and elsewhere, we make use of a basic fact from probability theory: Any conditional joint distribution for x and y can be factored as $[x, y|z] = [x|y, z][y|z]$. Note how the condition z is held constant in all terms.

3. Stochastic Air-Sea Interaction Models

[19] The primary topics of this section are (1) the production of the models on the right-hand side of equation (7) based on physical reasoning, and (2) a review of the requisite numerical computations.

3.1. Quasi-Geostrophic Ocean Models: Finite Differences

[20] Let ψ denote upper ocean streamfunction. A basic quasi-geostrophic (QG) model, with a conventional β-plane approximation, for the space-time evolution of ψ is

$$\left(\nabla^2 - \frac{1}{r^2}\right)\frac{\partial \psi}{\partial t} = -J\left(\psi, \nabla^2\psi\right) - \beta\frac{\partial \psi}{\partial x} + \frac{1}{\rho H}\mathrm{curl}_z\tau - \gamma\nabla^2\psi + a_h\nabla^4\psi, \quad (8)$$

where $r = \sqrt{gH}/f$ is the radius of deformation parameter; τ represents wind stress; ρ is density; H is the fluid depth; $-\gamma\nabla^2\psi$ represents the effect of bottom friction; and $a_h\nabla^4\psi$ is the lateral dissipation of relative vorticity by eddy processes (see below). The operator J is the Jacobian; $J(a, b) = \frac{\partial a}{\partial x}\frac{\partial b}{\partial y} - \frac{\partial b}{\partial x}\frac{\partial a}{\partial y}$.

[21] Conventional conversion to a discrete-time system approximates the time derivative on the left-hand side of equation (8) as $(\nabla^2\psi(t+1) - r^{-2}\psi(t+1)) - (\nabla^2\psi(t) - r^{-2}\psi(t))\}/\Delta$, where Δ is the selected time step.

[22] We consider a two-dimensional domain $\mathcal{D}$ consisting of $(n+2) \times (n+2)$ rectangular grid boxes, each measuring h_x units in the zonal (x) direction and h_y units in the meridional (y) direction. Let $N = n \times n$ denote the number of interior grid boxes; and $N_B = 4(n+1)$ denote the number of boundary grid boxes. Define $\mathbf{\Psi}^t$ to be the $(N + N_B)$-vector of space-time gridded ocean-streamfunction values composed of the N-vector $\mathbf{\Psi}_I^t$ and the N_B-vector $\mathbf{\Psi}_B^t$. It is convenient to view each element of a $\mathbf{\Psi}$ vector as representing the spatial average of ψ over the associated grid box.

[23] The next step involves approximation of the spatial derivatives in equation (8). For example, the conventional five-point approximation for the two-dimensional Laplacian ∇^2 at a point (x, y) is given by

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} \approx \frac{\psi(x+h_x, y) - 2\psi(x, y) + \psi(x-h_x, y)}{h_x^2} + \frac{\psi\left(x, y+h_y\right) - 2\psi(x, y) + \psi\left(x, y-h_y\right)}{h_y^2}. \quad (9)$$

[24] Combining these space- and time-derivative approximations, we obtain an approximation to equation (8) of the form

$$\mathbf{\Psi}_I^{t+1} = \left\{\mathbf{I} + \Delta\tilde{\mathbf{G}}\left(-\beta\mathbf{D}_x - \gamma\mathbf{G} + a_h\mathbf{G}^2\right)\right\}\mathbf{\Psi}_I^t + \Delta\tilde{\mathbf{G}}\left(-\mathcal{J}\left(\mathbf{\Psi}_I^t\right) + \frac{1}{\rho H}\mathcal{C}(\mathbf{U}^t, \mathbf{V}^t)\right) + \mathbf{B}\mathbf{\Psi}_B^{t+1}, \quad (10)$$

where $\mathbf{I}$ is the $N \times N$ identity matrix; $\mathcal{J}(\mathbf{\Psi}_I^t)$ is a discretized Jacobian evaluated at $\mathbf{\Psi}_I^t$; $\mathcal{C}(\mathbf{U}^t, \mathbf{V}^t)$ is a discretized contribution of the wind stress; $\mathbf{B}\mathbf{\Psi}_\mathbf{B}^{t+1}$ is an N_B-vector arising from boundary conditions; $\mathbf{D}_x$ is the matrix operator corresponding to discretized, zonal spatial derivatives; and $\tilde{\mathbf{G}} = (\mathbf{G} - r^{-2}\mathbf{I})^{-1}$, where $\mathbf{G}$ is the matrix operator for a discretized Laplacian ∇^2.

3.2. Stochastic Wind-Driven Ocean Models

[25] We seek prior and posterior distributions for the gridded "true" streamfunction. We now view $\mathbf{\Psi}$ as an idealized quantification of an aspect of the "real" ocean, rather than as idealized numerical approximations to solutions of a system of partial differential equations. Further, $\mathbf{\Psi}$ is modeled as random.

[26] With equation (10) serving as motivation, our suggestion for the first distribution on the right-hand side of equation (7) is

$$\mathbf{\Psi}^{t+1} = \mathbf{A}(\mathbf{l})\mathbf{\Psi}^t - j\tilde{\mathbf{G}}\mathcal{J}\left(\mathbf{\Psi}_I^t\right) + c\tilde{\mathbf{G}}\mathcal{C}(\mathbf{U}^t, \mathbf{V}^t) + b\mathbf{B}\mathbf{\Psi}_B^{t+1} + \mathbf{e}_{t+1}, \quad (11)$$

where $\mathbf{e}_{t+1}$ is a vector of random errors. Note that we account for variability and some uncertainty by inclusion of additive random errors $\mathbf{e}$. Typical assumptions for the $\mathbf{e}_t$'s are that they each have a multivariate Gaussian distribution with mean equal to the zero vector and covariance matrix Σ_e; as shorthand, write $\mathbf{e}_t \sim N(\mathbf{0}, \Sigma_e)$. It is common to assume that these vectors are independent across time. The use of additive errors and Gaussian assumptions are convenient, rather than limitations of the Bayesian viewpoint.

[27] We next describe modeling the various terms and parameters (what we have been labeling generically as η_ψ is comprised of the quantities $\mathbf{l}$, j, c, b, and Σ_e) indicated in equation (11).

3.2.1. Linear Term

[28] Our suggestion for $\mathbf{A}$ in equation (11) is based on the linear contributions in equation (10),

$$\mathbf{A}(\mathbf{l}) = l_1\mathbf{I} + l_2\tilde{\mathbf{G}}\mathbf{D}_x + l_3\tilde{\mathbf{G}}\mathbf{G}^2, \quad (12)$$

where $\mathbf{l} = (l_1, l_2, l_3)'$ are parameters to be modeled. Before describing the modeling, we point out that we have first applied a simple approximation. Namely, comparing equation (12) to equation (10), it appears that we omitted $\gamma\tilde{\mathbf{G}}\mathbf{G}$ corresponding to the discretized bottom friction contribution. In our example, we find that $\tilde{\mathbf{G}}\mathbf{G} \approx \mathbf{I}$, with the approximation good to at least four decimal places in each element. Hence, this term is simply combined in the development of l_1.

[29] The essence of the thought process is the comparison of the terms in equation (10) with those of equation (12). Our introduction of $\mathbf{l}$ can be viewed as the use of "free parameters" that can partially control the eigenvalues of $\mathbf{A}$, but in a fashion that is responsive to the observations.

Rather than specifying parameters exactly, we view them as random, but trainable by the observations through Bayesian updating. For example, direct comparison suggests that l_1 ought to be set to $1 - \Delta\gamma$. We treat this as prior information regarding l_1. We do impose the condition that $0 < l_1 < 1$. Similarly, comparing terms suggests that $l_2 \approx -\Delta\beta$ and $l_3 \approx \Delta a_h$, but we treat l_2 and l_3 as a random quantities. This approach essentially absorbs the gridding parameters h_x, h_y and Δ into statistical parameters. This is highly nonstandard from the perspectives of conventional deterministic modeling and most modern data assimilation schemes.

[30] Two important enhancements to this strategy are available in principle. First, the scalars l_1, l_2, and l_3 used here could themselves be replaced by matrices, leading to models capable accounting for varying spatial behaviors. This seems particularly important when the domain $\mathcal{D}$ is large enough to suggest that different "physics" are relevant within the domain. Second, these control parameters can be modeled as time-varying, perhaps with clocks much slower that the clock of the streamfunction process. The idea here is that the model can be responsive to local space-time fluctuations of the physical processes of interest. For example, the operative physics are highly nonstationary during ocean deep convection events as opposed to those during nonconvective periods. We should mention that such enhancements can introduce intense modeling and computational challenges. As in all modeling efforts, we must trade off fidelity to the physics and the data with parsimony of the model and feasibility of computations. Note that the radius of deformation parameter r^2 can also be modeled as random.

3.2.2. Jacobian Term

[31] The quantity $-j\tilde{\mathbf{G}}\,\mathcal{J}$ in equation (11) is to account for the nonlinear evolution of QG streamfunction. The inclusion of the parameter j, apparently replacing Δ in equation (10), follows the spirit of our discussion of $\mathbf{L}$ and will not be repeated.

3.2.3. Wind Stresses

[32] The term $c\tilde{\mathbf{G}}\mathcal{C}(\mathbf{U}^t, \mathbf{V}^t)$ is critical to our goal of linking atmosphere and ocean processes. Note that again we have used a random control parameter c to moderate the contribution of wind-stress curl to the model.

3.2.4. Boundary

[33] We next formulate the boundary model (in equation (7)). As indicated, this step is amenable to hierarchical Bayesian modeling [cf. *Wikle et al.*, 2003].

[34] Our prior for Ψ_B depends upon changes in the implied Sverdrup streamfunction in the vicinity of the western boundary, based on updated wind forcing information. First, we specify an initial condition Ψ_B^0 based on the coarsened PE-SWE "truth" simulation. We allow the boundary to evolve according to

$$\Psi_B^{t+1} = \Psi_B^0 + \alpha_{sv}\left(\bar{S}_v^t - \bar{S}_v^0\right), \tag{13}$$

where $\bar{S}_v^t$ is the average Sverdrup streamfunction on the western boundary at time t, and α_{sv} is a random parameter that moderates the contribution of the Sverdrup streamfunction relative to its initial value. Specifically, we numerically approximate

$$S_v^t(x_W) = \frac{-1}{\rho H\beta}\int_{x_E}^{x_W} \mathrm{curl}_z\tau\, dx,$$

where $S_v^t(x_W)$ is the Sverdrup streamfunction at a western boundary location x_W. We set $\bar{S}_v^t = N_x^{-1}\Sigma_{x_W} S_v^t(x_W)$, where the sum is over all of the western boundary locations. The prior distribution for α_{sv} is set to be uniform over the interval -0.1 to 0.1.

3.2.5. Errors

[35] As mentioned above, a typical assumption is that the errors have a Gaussian distribution $\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_e)$. Our prior understanding suggests that the errors are likely to be correlated and we have a rough sense of the variability corresponding to these errors. Such background information can be used in formulating reasonable models for $\mathbf{\Sigma}_e$. Preliminary computational experiments to analyze the differences between PE-SWE solutions and coarsely gridded QG solutions could provide additional information. Fortunately, modeling spatial covariance matrices is a well-studied problem in spatial statistics [e.g., *Cressie*, 1993]. Also, departures from the Gaussian assumptions can be accommodated. On the other hand, some sacrifices may be necessitated due to computational limitations.

4. Observing System Simulation Experiment

[36] An observing system simulation experiment (OSSE) was designed to demonstrate the effectiveness and sensitivities of the BHM air-sea interaction model. The OSSE consists of an ocean "truth" simulation, driven by idealized surface wind fields. The ocean "truth" and idealized atmospheric forcing are realistically subsampled in space and time, and corrupted with artificial noise, to supply the BHM with simulated observational data.

4.1. Ocean Spin-Up

[37] The ocean "truth" simulation is developed in a primitive equation, shallow water approximation context, using the model described by *Milliff and McWilliams* [1994]. A 1-year spin-up calculation is driven by relatively low-amplitude constant winds to build a basin-scale background circulation upon which a 10-day OSSE is conducted. The surface wind forcing for the spin-up calculation is designed to drive a cyclonic basin-scale circulation characterized by boundary currents along three sides of a closed domain, with an eddy field on the basin interior.

[38] Ocean model parameters for the spin-up calculation are listed in Table 1. The testbed domain is centered at 55°W, 60°N, and is discretized at 10 km resolution into a 101×101 square grid. Given that we have selected barotropic scaling, the shallow-water, primitive equation model time step is 15 s. Dissipation in the model is achieved by bottom friction with a 10-day spin-down timescale, and Laplacian lateral viscosity which is chosen to be as low as numerical stability will permit. These parameters are selected to be appropriate for the 10-day wind-driven experiment to be described below.

[39] The constant surface wind field used for spin-up is depicted in Figure 1a, and it consists of two parts. Part one is the classical sinusoidal distribution of zonal wind with latitude known to force a basin-scale subpolar gyre circulation with an equatorward western boundary current. Part two is superposed on part one in the eastern boundary current region of the basin. There, the northward winds

Table 1. Shallow Water Primitive Equation Model Parameters

Parameter	Value
H, basin depth	5000 m
g, gravitational acceleration	9.81 m s^{-2}
f_o, Coriolis parameter at mid-basin	1.263×10^{-4} s^{-1}
β, variation in f with latitude	1.144×10^{-11} m^{-1} s^{-1}
a_h, eddy coefficient (Laplacian lateral viscosity)	135 m^2 s^{-1}
γ, bottom friction coefficient	1.1574×10^{-6} s^{-1}

increase exponentially toward the coast across a narrow nearshore band. This induces an eastern boundary intensification in wind-stress curl that in turn drives a poleward boundary current along the domain eastern boundary. Eastern boundary wind-stress curl extrema have been observed in eastern boundary current regions, all over the globe, in annual average surface wind-stress curl derived from satellite scatterometer data [*Milliff and Morzel*, 2001].

[40] The ocean circulation after a 1-year spin-up calculation is represented in Figure 1b as the perturbation pressure field, $p' = \frac{p}{\rho} - gH$. The desired boundary currents and interior eddy fields are evident. The maximum and minimum velocity component amplitudes corresponding to the pressure gradients in Figures 1 and 2 are -0.19 to 0.29 m s^{-1} in the eastward, u component; and -0.24 to 0.33 m s^{-1} in the northward, v component. Most of these extrema occur in a closed cyclonic circulation that is slowly building, over the course of the spin-up calculation, in the southwest corner of the domain. While this quasi-permanent cyclonic flow departs from an idealized circulation based for example on the Labrador Sea, it provides a ''marker'' circulation with which differences in the BHM posterior distributions can be quantified. The oceanic cyclone in the southwestern corner of the domain is consistent with β-plume response to the wind-stress curl forcing on the eastern boundary [see it *Spall*, 2000].

4.2. Simulated Data and Data Models

[41] The p' field in Figure 1b represents the initial condition for the ocean ''truth'' circulation to be used in the 10-day OSSE. Atmospheric forcing for the OSSE period is designed to be vigorous, emulating the momentum effects of surface forcing from polar low pressure systems in the Labrador Sea region in winter [e.g., *Renfrew et al.*, 1999]. An idealized atmospheric cyclone superposes the spin-up forcing, and is propagated across the testbed domain in 6*d*. Figure 2 consists of two columns depicting the simulated scatterometer samples of the atmospheric forcing data (right column) and the simulated ocean response p' fields with overlying simulated altimeter tracks (left column). Fields are depicted in Figure 2 for days 1.0, 3.0, 5.0, and 7.0 (four rows). The idealized atmospheric cyclone propagates across and exits the OSSE domain within this time period.

4.2.1. Scatterometer Data

[42] Simulated scatterometer data occur in alternating ascending and descending swathes that span large portions of the domain at 0 and 12 hours every day. The simulated orbits are declined 8° with respect to due north. The swath width, within-swath resolution, and the over-flight frequency are all comparable to the present-day QuikSCAT system.

[43] The formulation of the scatterometer data model is given by

$$\begin{pmatrix} \mathbf{D}_u^t \\ \\ \mathbf{D}_v^t \end{pmatrix} = \mathbf{K}_w^t \begin{pmatrix} \mathbf{U}^t \\ \\ \mathbf{V}^t \end{pmatrix} + \epsilon_w^t, \epsilon_w^t \sim N\left(\mathbf{0}, \sigma_{w,\epsilon}^2 \mathbf{I}\right), \qquad (14)$$

where $\mathbf{D}_u^t$ and $\mathbf{D}_v^t$ are the vectors of u- and v-component wind observations at time t; the vectors of measurement errors ϵ_w^t are assumed to be independent across time; and the error variance $\sigma_{w,\epsilon}^2$ is known for both wind components (e.g., from calibration/validation studies for NSCAT by *Freilich and Dunbar* [1999]). The measurement operator $\mathbf{K}_w^t$ is an incidence matrix (a sparse matrix of zeros and ones) that maps the prediction locations to the observation locations such that all observations within a prediction location grid box are assumed to be noisy observations of the true process at that prediction location [e.g., *Wikle et al.*, 1998]. More complicated formulations of this incidence matrix are possible [e.g., *Wikle et al.*, 2001].

4.2.2. Altimeter Data

[44] Simulated altimeter tracks occupy a repeating pattern of 18 tracks that cross the OSSE domain in alternating ascending (i.e., from southeast to northwest) and descending (i.e., from northeast to southwest) directions every 3 days. Simulated altimeter orbits cross portions of the OSSE domain at 0 and 6 *hr* every day. Either two or three orbits are reported at each of those times. The simulated altimeter reports p' directly from the ocean ''truth'' simulation as described below. The along-track resolution is 12 km. While this along-track resolution is about half the resolution of the standard SSH products from the Topex/Poseidon data set, the number of tracks that cover the domain at any given over-flight time is about twice the realistic count.

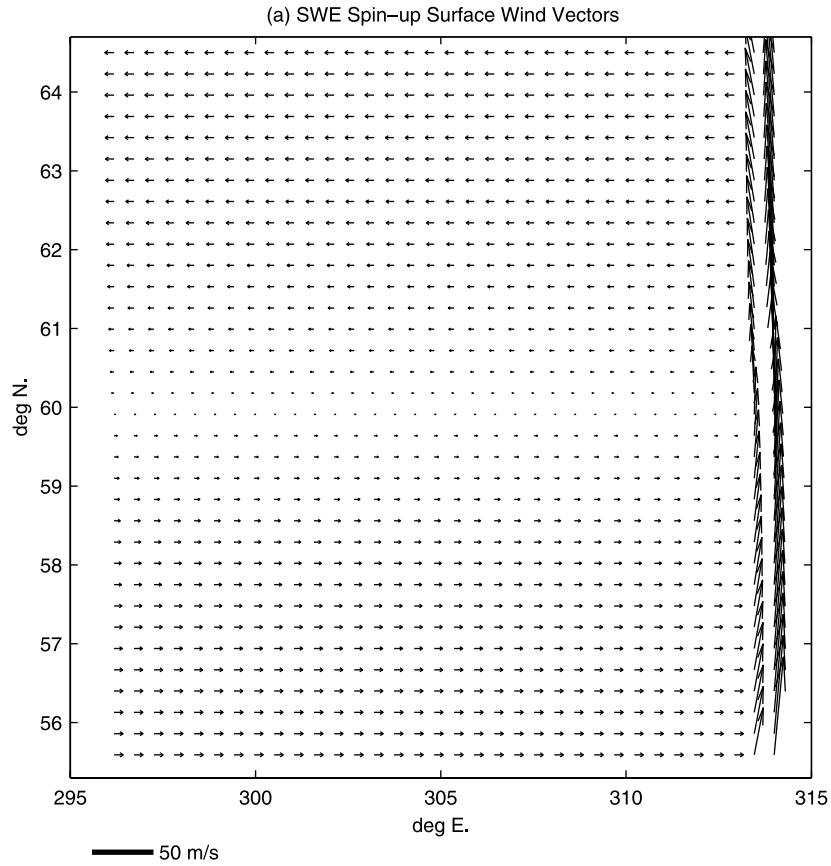
[45] The data model for the simulated altimeter observations is similar to that used for the wind data,

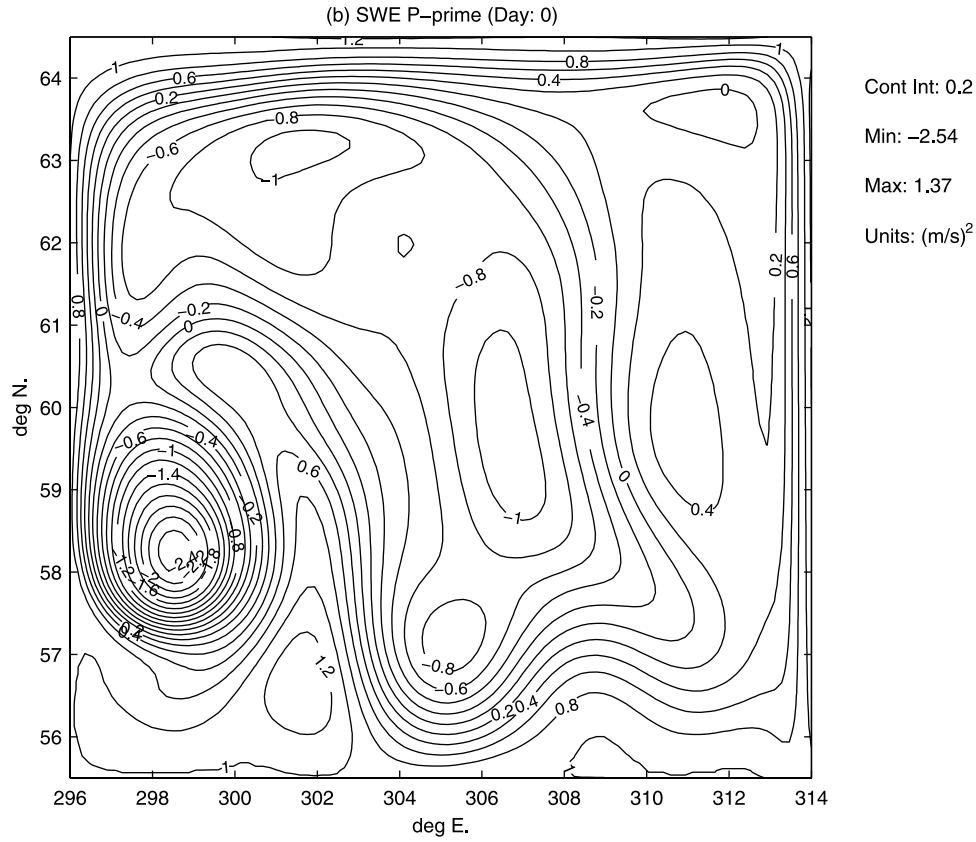
$$\mathbf{D}_{p'}^t = \mathbf{K}_o^t \mathbf{\Psi}^t f_o + \epsilon_o^t, \epsilon_o^t \sim N\left(\mathbf{0}, \sigma_{o,\epsilon}^2 \mathbf{I}\right), \qquad (15)$$

where $\mathbf{D}_{p'}^t$ is a vector of ocean dynamic pressure observations at time t; $\mathbf{\Psi}^t$ is a vector of the ocean streamfunction anomaly process at prediction grid locations for time t, f_o is the Coriolis parameter; the vectors of measurement errors ϵ_o^t are assumed to be independent across time; the error variance $\sigma_{o,\epsilon}^2$ is again assumed known; and $\mathbf{K}_o^t$ is an incidence matrix that maps the prediction grid locations to the observation locations, analogous to the atmospheric data model formulation.

[46] Note that this model implies that $p'_{equiv} = f_o \Psi$, which is valid for a quasi-geostrophic system. We distinguish

Figure 1. (opposite) Shallow-water model ''truth'' simulation spin-up forcing and response to establish initial conditions for the 10-day Observing System Simulation Experiment. (a) Surface wind vectors held constant for a 1-year spin-up period. An eastern boundary positive wind-stress curl extremum is superposed upon a basin scale circulation consistent with circulation in the sense of a subpolar gyre. (b) Perturbation dynamic pressure contours at 1 year in the spin-up calculation.


(a) SWE Spin–up Surface Wind Vectors
deg N.
deg E.
50 m/s


(b) SWE P–prime (Day: 0)
deg N.
deg E.
Cont Int: 0.2
Min: –2.54
Max: 1.37
Units: (m/s)2

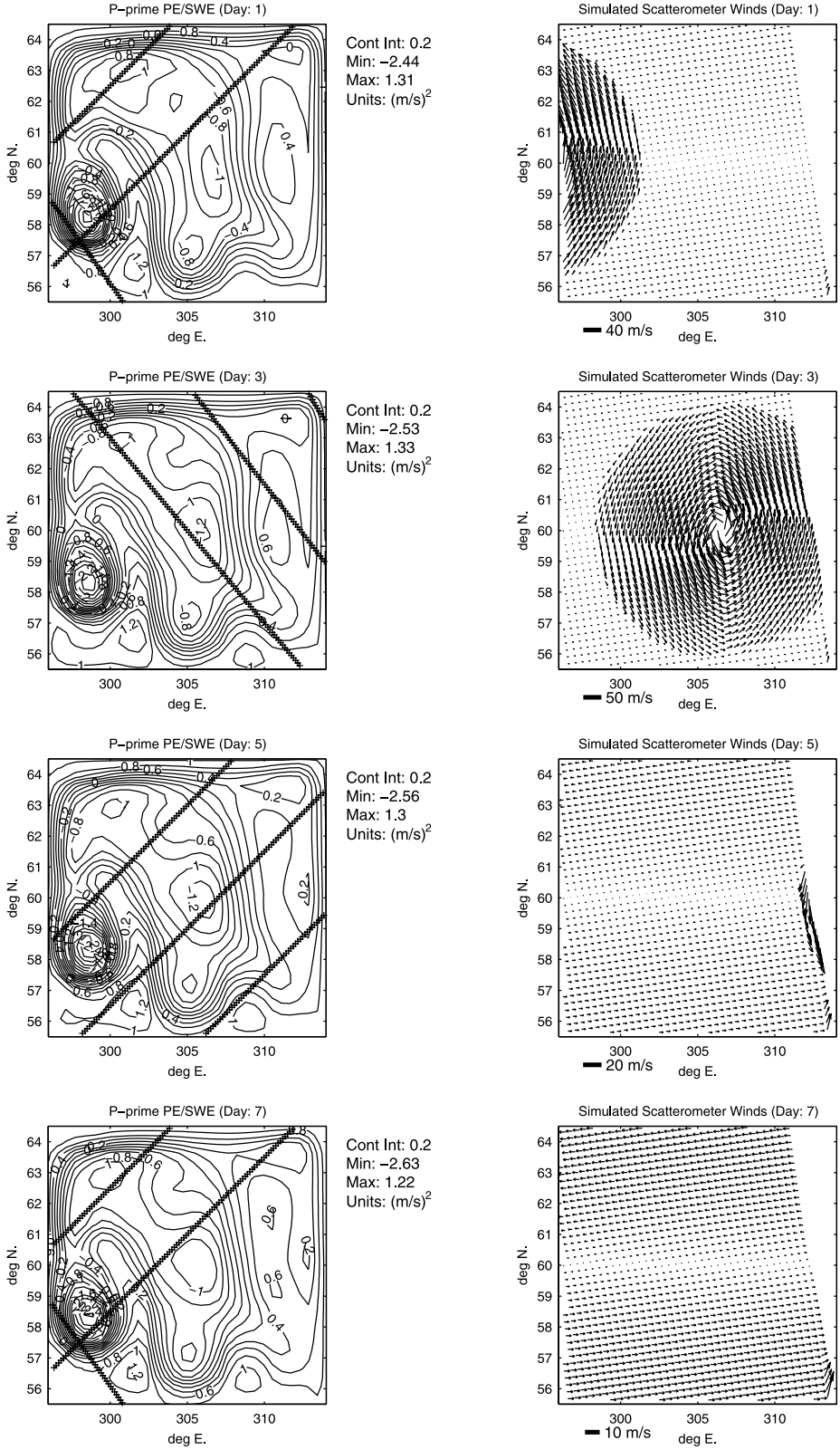


Figure 2. (left-hand column) Ground tracks for simulated altimeter and (right-hand column) simulated scatterometer inputs to the BHM. Rows, top to bottom, correspond to days 1, 3, 5, and 7 of the 10-day OSSE. The simulated altimeter tracks are superposed on the perturbation dynamic pressure (p') fields from the ''truth'' simulation. These p' data are artificially degraded to simulate observational errors before being supplied to the BHM. The simulated scatterometer fields depict the rapid transit across the domain of an idealized atmospheric cyclone. The simulated scatterometer data are also artificially contaminated with errors before being supplied to the BHM.

between p'_{equiv} and p' from the PE-SWE system where a divergent circulation changes the definition of dynamic pressure. However, in the OSSE the divergent circulation component is small, and the data model presented above is reasonable. More complicated data models for altimetry observations will be explored elsewhere (e.g., including geoid and general circulation effects).

4.3. BHM Specifications

[47] We next describe the specific assumptions used in the OSSE. First, we set the BHM time step to be 12 hours.

4.3.1. Atmospheric Process Model

[48] As shown by *Royle et al.* [1999] and *Wikle et al.* [2001], the BHM framework can be used to obtain realistic wind fields given sparsely sampled scatterometer data. Since our focus here is on the atmosphere-ocean coupling and the ocean response, and since we have assumed that relatively large amounts of wind data are available, we employ a relatively simple BHM for the atmospheric component. We make use of a "stochastic geostrophy" argument, similar to *Royle et al.* [1999]. For each time t, we specify the distribution of the horizontal wind components given the atmospheric pressure field as follows:

$$\mathbf{U}^t|\mathbf{P}^t, \sigma^2_{u|P} \sim N\left(-(1/\rho_a f_0)\mathbf{D}_y\mathbf{P}^t, \sigma^2_{u|P}\mathbf{I}\right) \tag{16}$$

$$\mathbf{V}^t|\mathbf{P}^t, \sigma^2_{v|P} \sim N\left((1/\rho_a f_0)\mathbf{D}_x\mathbf{P}^t, \sigma^2_{v|P}\mathbf{I}\right), \tag{17}$$

where $\mathbf{P}^t$ is the atmospheric pressure field at time t, ρ_a was set to 1, $\mathbf{D}_x$ and $\mathbf{D}_y$ are the difference operators in the x- and y-directions, respectively, and $\sigma^2_{u|P}$, $\sigma^2_{v|P}$ are variances of random errors representing differences between the true field and the geostrophic approximation. Note that more sophisticated (i.e., correlated) error structures can be used for this error process and the geostrophic relationship can be made less explicit within this framework [*Royle et al.*, 1999; *Berliner et al.*, 1999].

[49] We next specify the distribution for the atmospheric pressure process. (Note that expressions such as equation (6) should formally include the pressure process; we will not adjust the notation.) In the current setting we have no observations of atmospheric pressure. Thus, pressure is in the model as a "hidden process." We specify a distribution for the pressure process that can simulate realistic pressure fields. Specifically, we let $\mathbf{P}^t \sim N(\boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P)$ where $\boldsymbol{\mu}_P$ is the mean pressure field and $\boldsymbol{\Sigma}_P$ is the covariance matrix of a correlated noise process. One could obtain $\boldsymbol{\Sigma}_P$ by choosing a covariance function appropriate to the atmosphere [e.g., *Thiebaux*, 1985] with correlation lengths appropriate for the Labrador Sea region. In the limited domain of the Labrador Sea, analyses of surface winds from NSCAT and Minimet drifting buoys [*Milliff et al.*, 2003] suggest that a reasonable correlation function for the pressure field is the exponential covariance function, $c(d) = \sigma^2_p \exp(-\theta_p d)$ where d is distance and θ_p is a parameter describing the correlation decay over space. To facilitate computation, we reparameterize the pressure process by letting $\mathbf{P}^t = \mu_p\mathbf{1} + \mathbf{E}(\theta_p)\boldsymbol{\alpha}^t$, where μ_p is the baseline pressure (assumed known), $\mathbf{E}(\theta_p)$ is a matrix of the first n_p EOFs from the covariance matrix $\boldsymbol{\Sigma}_P$ (which depends on the parameter θ_p), and $\boldsymbol{\alpha}^t$ are the random variances associated with the EOFs $\mathbf{E}$. We assume that the $\boldsymbol{\alpha}^t$ all have the same distribution and are independent over time: $\boldsymbol{\alpha}^t \sim N(\mathbf{0}, \text{diag}(\boldsymbol{\lambda}))$ where $\text{diag}(\boldsymbol{\lambda})$ is a diagonal matrix of variances. We assume the λ_j, $j = 1,\ldots, n_p$ are independent, each distributed according to an inverse gamma distribution, $IG(q_j, r_j)$, where the parameters q_j, r_j are known and can be obtained from the EOF decomposition of $\boldsymbol{\Sigma}_P$.

[50] We must specify prior distributions and/or parameter values for this model. For example, $\sigma^2_{u|P}$ and $\sigma^2_{v|P}$ can be given inverse gamma priors with known parameters or we can simply specify the variances directly, in which case they are assumed to be fixed. We take the latter approach and specify $\sigma^2_{u|P} = \sigma^2_{v|P} = 3\ \text{m}^2\ \text{s}^{-2}$ in the OSSE. Similarly, σ^2_P can be given a distribution but is fixed in the OSSE to be $\sigma^2_P = 10^5 Pa^2$. Furthermore, we fix $\theta_P = .003\ \text{km}^{-1}$ (corresponding to distance in kilometers) and let $n_p = 10$. Our results are not extremely sensitive to these choices.

[51] Note that there is no dynamical time-dependence in this atmospheric model. Dynamical models for wind fields can be considered in the BHM framework [e.g., *Wikle et al.*, 2001] but are not needed for the OSSE. A time-varying "stochastic quasi-geostrophic" model in the BHM framework can be used in situations with realistic scatterometer sampling and will be discussed elsewhere.

4.3.2. Ocean Model Error Distribution

[52] We consider the distribution $\mathbf{e}_t \sim N(\mathbf{0}, \sigma^2_e\mathbf{R}(\theta))$, where σ^2_e is the variance (assumed to be homogeneous) and $\mathbf{R}(\theta)$ is a correlation matrix which depends on a spatial dependence parameter θ. Given the relatively small spatial domain, we assume an isotropic exponential correlation model in defining $\mathbf{R}$ [*Cressie*, 1993, section 2.3]. Implementation is problematic however, given the relatively high-dimensional prediction grid for which samples from this multivariate normal distribution must be obtained. In testbed mode, we make an additional simplification for the sake of computational efficiency. We use an approximation,

$$\tilde{\mathbf{R}}(\theta) = \mathbf{F}(\theta)(\theta)\mathbf{F}'(\theta), \tag{18}$$

where $\mathbf{F}(\theta)$ is an $N \times n_e$ dimensional matrix of eigenvectors (EOFs) of $\mathbf{R}(\theta)$), and $\boldsymbol{\Lambda}(\theta)$ is the corresponding diagonal matrix of spectral variances. The implied correlation function is approximately exponential, if n_e is relatively large. Further, $\tilde{\mathbf{R}}(\theta)$ is singular, but suffices for the illustrative intent of the OSSE.

[53] We must specify distributions for σ^2_e and θ. We let $\sigma^2_e \sim Unif(1.1 \times 10^6, 1.6 \times 10^6)$, where *Unif* indicates a uniform distribution on the indicated range, and $\theta \sim DUnif([.001,.002,\ldots,.01])$, where *DUnif* refers to a discrete uniform distribution. Our selection of parameters corresponds to the exponential correlation, $\exp(-\theta d)$ where d is distance in kilometers.

4.3.3. Initial State Prior

[54] The initial condition $\boldsymbol{\Psi}^1$ was assumed to be Gaussian with prior mean that is the coarsened spin-up value from the truth simulation and covariance matrix that is the same as that of the model errors, with the exception that the variance is reduced by a factor of 10 (i.e., $0.10\ \sigma^2_e\mathbf{R}(\theta)$).

4.3.4. Altimetry Data Model Revisited

[55] The computational algorithm used here (see Appendices A and B) is valid, but may be inefficient. One byproduct of this inefficiency was that we needed to inflate

$\sigma^2_{o,\epsilon} = .001$ (see equation (15)) by a factor of 10^4 in the numerical calculations. In terms of standard deviation (i.e., in the same units as the altimetry data), this is an increase by a factor of 100. In part, this inflation is a result of subgrid-scale effects. Namely, the simulated data were produced by adding noise to the original PE-SWE results before applying space-time gridding up to the BHM specifications. Our data analysis of the PE-SWE results suggest spatial subgrid-scale variations leading to a standard deviation of about 0.1, explaining 1 order of magnitude of our inflation. Temporal effects may explain more of that inflation. We will report on new algorithms now being developed and tested elsewhere.

4.4. Results

[56] The basin-scale ocean circulation response to forcing from the idealized atmospheric cyclone is compared, for days 1, 3, 5, and 7 during the OSSE period, in Figure 3. The right-hand column depicts the $p'_{equiv} = f_o\psi$ field for the posterior mean of the BHM simulations (based on 1000 realizations), and the left-hand column depicts p' from the PE-SWE ''truth'' simulation. The field evolutions over the 10-day OSSE are very similar in feature morphology and response to the intermittent forcing.

[57] The amplitude fluctuations when the atmospheric cyclone inhabits most of the basin (see Figure 2) are larger in the posterior mean p'_{equiv} from the BHM than for p' from PE-SWE ''truth.'' The atmospheric cyclone imprint on p' contours at midbasin is apparent in the fields from days 3 and 5. The BHM posterior mean response amplitude is 30% larger than for the PE-SWE ''truth'' field on day 3. Boundary currents implied by gradients in p' in the ''truth'' simulation are at most 10% stronger than for the BHM posterior mean. The oceanic cyclone in the southwestern corner of the domain is 40% deeper than for the BHM case on days 5 and 7. On days 1 and 3 the difference is only 20%. Note that the oceanic cyclone is the strongest oceanic signal in both the ''truth'' and the BHM posterior mean.

[58] Difference fields with respect to p' from the PE-SWE ''truth'' simulation are depicted for two BHM results in Figure 4. Again, the rows correspond to differences on days 1, 3, 5, and 7. The left-hand column shows differences from ''truth'' on those days for p'_{equiv} from the posterior mean field described in Figure 3. The right-hand column shows differences for the BHM calculation from which all simulated altimeter data have been excluded.

[59] As expected, the largest differences, for both comparisons, occur in the oceanic cyclone in the southwestern corner of the ocean domain. However, the difference by day 5 in the BHM case with altimeter data is about 60% smaller than for the case when altimeter data are excluded. The dynamical priors for both BHM simulations are the same, so the difference field comparisons in Figure 4 quantify and add spatial information as to the value of simulated altimeter data in the Bayesian framework. This hints at an array design utility for BHM that will be taken up elsewhere.

[60] We also note that the differences in the boundary region in both BHM simulations (i.e., with and without simulated altimeter data) are uniformly small throughout the OSSE period. This implies that, in a posterior mean sense, the boundary vector approach to stochastic boundary condition specification is working [*Wikle et al.*, 2003]. The implementation of time-dependent quasi-geostrophic boundary conditions in deterministic modeling is somewhat nontrivial [e.g., *Pinardi and Milliff*, 1989], and the present success in this regard within the BHM framework indicates another potential strength of the approach.

[61] Figure 5 shows estimates of the standard deviation in p'_{equiv} from the BHM simulation in Figure 4, for days 1, 3, 5, and 7 (note change in contour interval). These estimates of standard deviation are based on 1000 realizations of the posterior distribution for ocean streamfunction (converted to equivalent dynamic pressure). They reflect regions of greater and lesser uncertainty in the posterior mean fields of Figure 3. Again, the region of the oceanic cyclone in the southwestern corner of the domain appears as a local extremum. Higher uncertainties are also associated with the domain boundaries. Similarly robust field measures of uncertainty are expensive to produce in conventional deterministic modeling.

[62] Next, we consider posterior distributions for model parameters. Prior distributions for many of these parameters have been described in the text, and they are summarized in Table 2. Values used for specified quantities are given in Table 3. Figure 6 compares the posterior and prior distributions for four parameters discussed in sections 3.2 and 3.3. (These plots are of estimates of the posterior density functions based on the Monte Carlo analysis. Throughout this article, the estimated densities are kernel density estimates [cf. *Silverman*, 1986].) These include; γ the bottom friction parameter, a_h the lateral dissipation parameter, j a parameter on the nonlinear evolution component of the BHM, and c a parameter on the wind-stress curl component. Posterior distributions for these parameters have evolved from initially uniform priors (e.g., the flat lines in each panel of Figure 6). The ranges for these priors were deemed reasonable, and were in some cases influenced by experience in deterministic model settings.

[63] The posterior distribution for the BHM term identified with the bottom friction coefficient γ (Figure 6a) implies a spin-down time longer than the bottom friction parameter used in the PE-SWE ''truth'' simulation (Table 1). This suggests a tendency in the BHM for lower overall dissipation. A tendency of this kind can arise because of an inconsistency that might be indicated between large dissipation in the model and the more energetic circulation implied by the observations. Figure 6b depicts the posterior distribution for the BHM term identified with eddy viscosity (a_h). The posterior mode for this distribution is less than zero, with a magnitude much smaller than the positive value required in the PE-SWE ''truth'' simulation for numerical stability ($a_h = 135$ m^2 s^{-1} in PE-SWE). Only by trial and error did we arrive at a uniform prior for a_h that included both positive and negative values. In BHM considered here, the a_h term is a random variable that scales a nearest-neighbor process in the evolution of the stochastic streamfunction process. Apparently, given the BHM grid spacing and time step size, this nearest neighbor process does not take up the role of a Fickian diffusion, at least in the presence of the observations.

[64] The posterior distributions for c and j (Figures 6c and 6d) are also difficult to relate to experience in deterministic modeling. The posterior distribution for c is bimodal, with extrema occupying the low and high ends of the range prescribed in the prior. The extremum in the largest ampli-

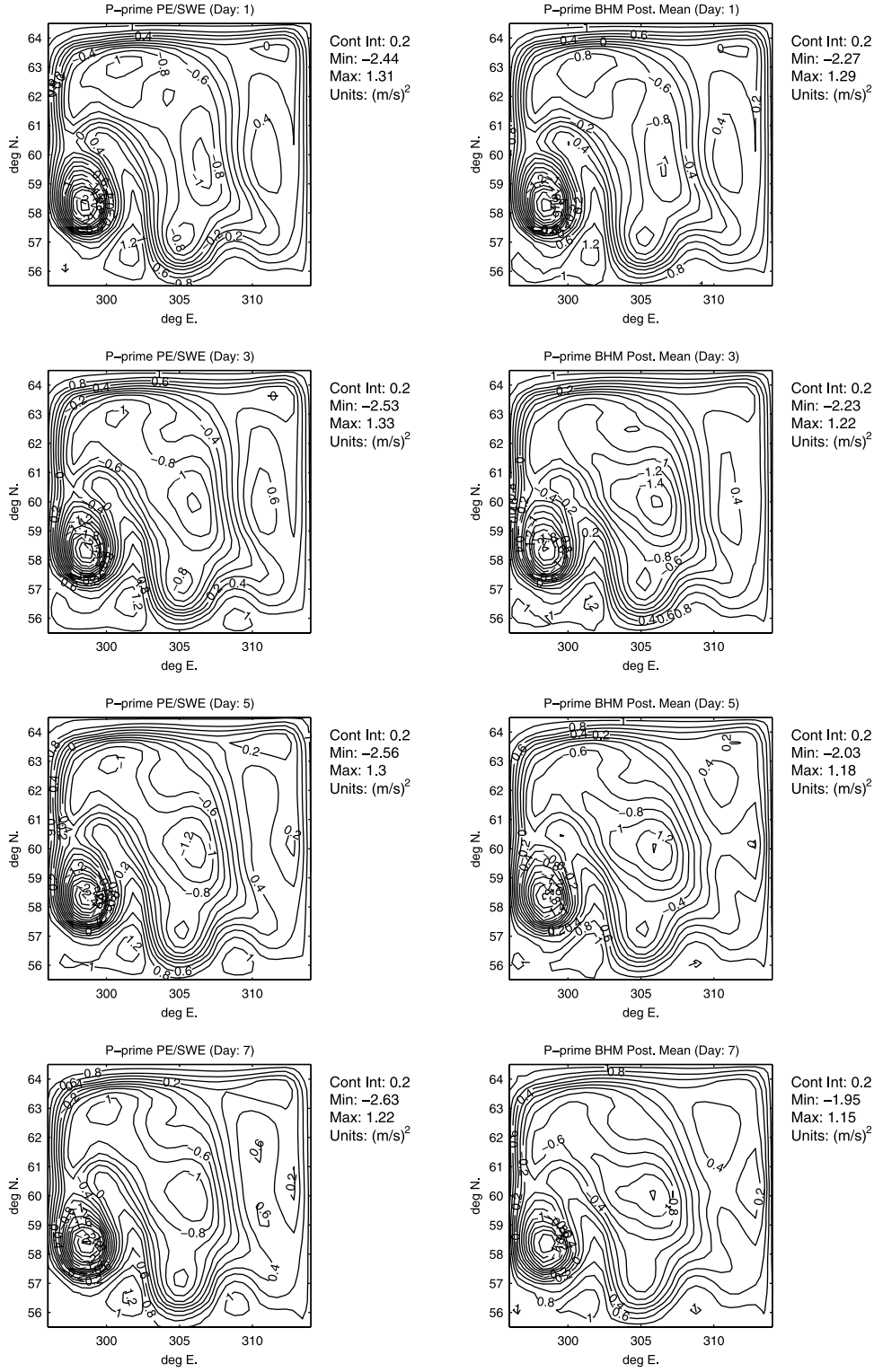


Figure 3. (right-hand column) Posterior mean fields versus (left-hand column) "truth" simulation for days (rows from top to bottom) 1, 3, 5, and 7 in the 10-day OSSE. For each day, the perturbation dynamic pressure (p') is compared with the posterior mean for an equivalent field ($p'_{equiv} = f_o\psi$).

tude range might also infer a BHM tendency toward a higher energy state. The posterior distribution for j does not identify a clear mode. The distribution is shifted toward the higher nonlinearity end of the prior range.

[65] As a final indication of potential calculations based on the BHM output, consider basin-integrated kinetic energy $KE(t)$ as a function of time t. The KE(t) trace from the PE-SWE "truth" simulation is marked by a bold solid line in

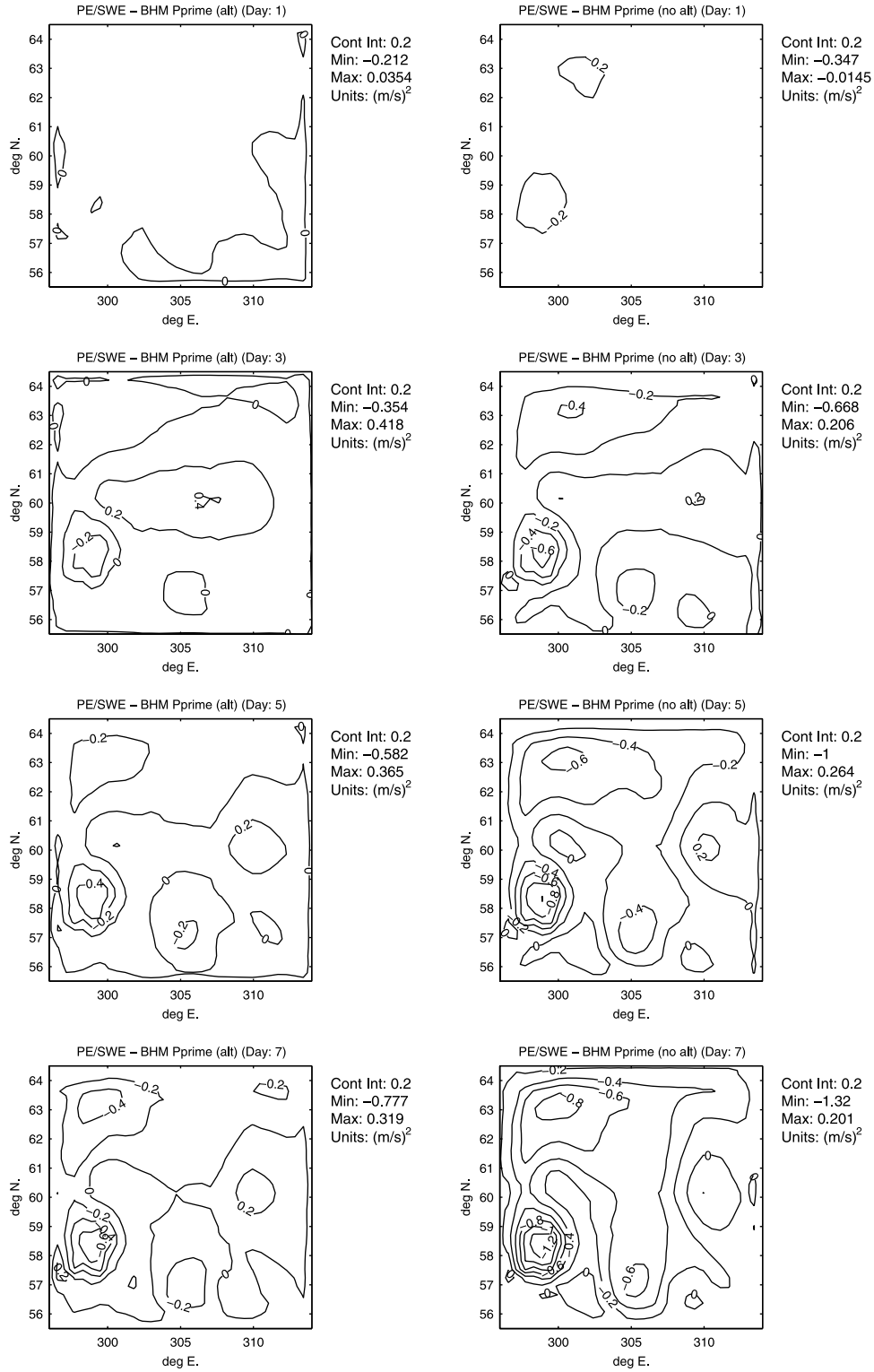


Figure 4. Difference fields with respect to the p' "truth" simulation for the posterior mean fields from two BHM OSSE calculations. (left-hand column) Differences (in $m^2\ s^{-2}$) for the BHM OSSE that includes both simulated scatterometer and simulated altimeter data sets. (right-hand column) BHM OSSE difference fields from posterior mean field comparisons in an experiment excluding simulated altimeter data. Difference fields are shown for days (rows from top to bottom) 1, 3, 5, and 7.

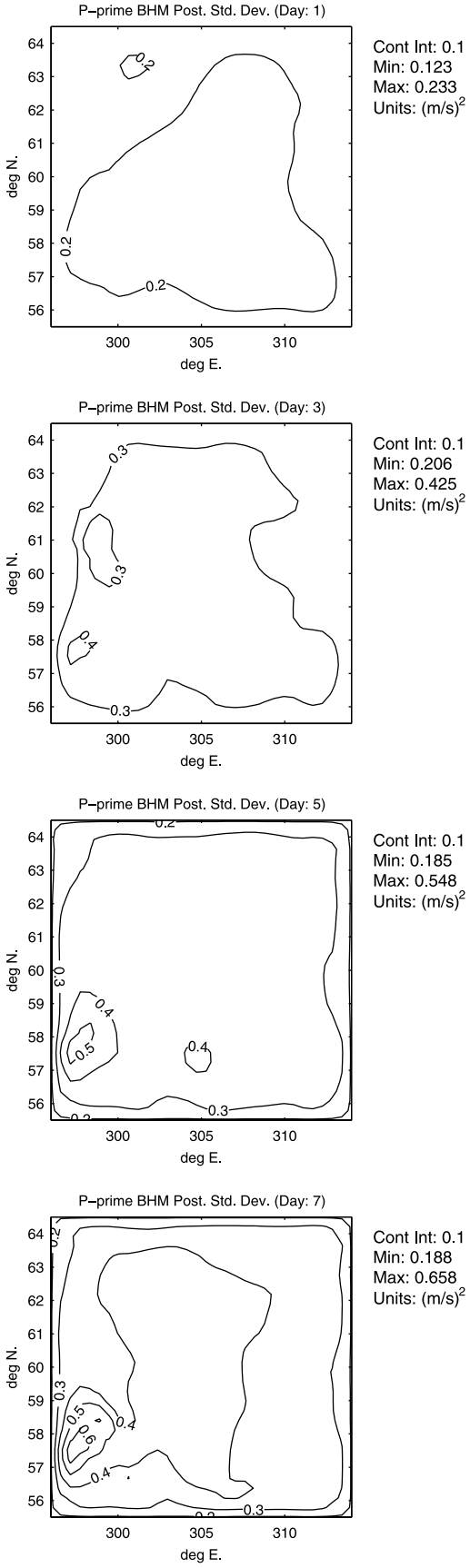


Figure 5. Maps of posterior standard deviations of p'_{equiv} for days 1, 3, 5, and 7 from the 10-day BHM OSSE including both simulated scatterometer and simulated altimeter data.

Table 2. BHM Parameter Prior Distributions

Parameter	Prior Distribution
l_1	$1 - \Delta\gamma$ where $\gamma \sim Unif(0, 1.6534 \times 10^{-6})$
l_2	$Unif(-\Delta\beta - (0.04)\Delta\beta, -\Delta\beta + (0.04)\Delta\beta)$
l_3	Δa_h, where $a_h \sim Unif(-215, 215)$
j	$Unif(\Delta - (0.04)\Delta, \Delta + (0.04)\Delta)$
c	$Unif(\Delta - (0.04)\Delta, \Delta + (0.04)\Delta)$
b	$Unif(.97, 1)$
σ_e^2	$Unif(1.1 \times 10^6, 1.6 \times 10^6)$
α_{sv}	$Unif(-0.1, 0.1)$
θ	$DUnif([0.001, 0.002, \ldots, 0.01])$

Figure 7. The BHM yields posterior distributions of kinetic energy at each BHM output time. These distributions are indicated by gray-scale shading in Figures 7a and 7b. The means of the posterior distributions for KE(t) are marked by bold dashed lines. Figure 7a compares "truth" KE(t) with the BHM posterior distributions for an example simulation relying on both simulated-scatterometer and simulated-altimeter data. Figure 7b depicts the comparison for a BHM simulation from which all simulated altimeter data have been withheld. Figures 7c–7f depict KE(t) distribution comparisons for selected times (days 1, 3, 5, and 7).

[66] First, note that for the simulation based on both data sets, the true values of $KE(t)$ are contained in regions of reasonably high posterior probability for each t. The posterior means of $KE(t)$ from the BHM simulations depart from the PE-SWE "truth" $KE(t)$ with time, though this effect is less prominent when simulated-altimeter data are included (Figure 7a). Simulated-altimeter data are incorporated in the BHM every 12 hours, beginning 12 hours after the model initialization. The posterior mean KE trace can be seen to respond by jumping back toward the SWE "truth" $KE(t)$ in the BHM output times following simulated-altimeter updates. The posterior distributions for $KE(t)$ dissipate in amplitude with time, demonstrating in part an increase in uncertainty in the BHM simulation with time due to dynamical effects in the presence of uncertain initial conditions. However, as expected the spread in the posteriors relying on both data sets is less than those based on the simulated-scatterometer data only.

[67] We note that there is a bias toward a lower KE state in the BHM that is overcome in part by simulated-altimeter data. Possible causes for this bias in the posterior means include (1) the BHM underestimation of the oceanic cyclone

Table 3. BHM Fixed Parameters

Parameter	Fixed Value
H	5000 m
ρ	1000 kg m^{-3}
r	1.753×10^6 m
f_o	1.26301×10^{-4} s^{-1}
Δ	43,200 s
h_x	32.3×10^3 m
h_y	32.3×10^3 m
$\sigma^2_{o,\epsilon}$	0.001 m^4 s^{-4}
n_e	30
n_p	10
$\sigma^2_{y,\epsilon}$	0.1 m^2 s^{-2}
$\sigma^2_{u\|P}$	3 m^2 s^{-2}
$\sigma^2_{v\|P}$	3 m^2 s^{-2}
θ_p	0.003 km^{-1}
σ^2_p	10^5 Pa^2
μ_p	100,000 Pa

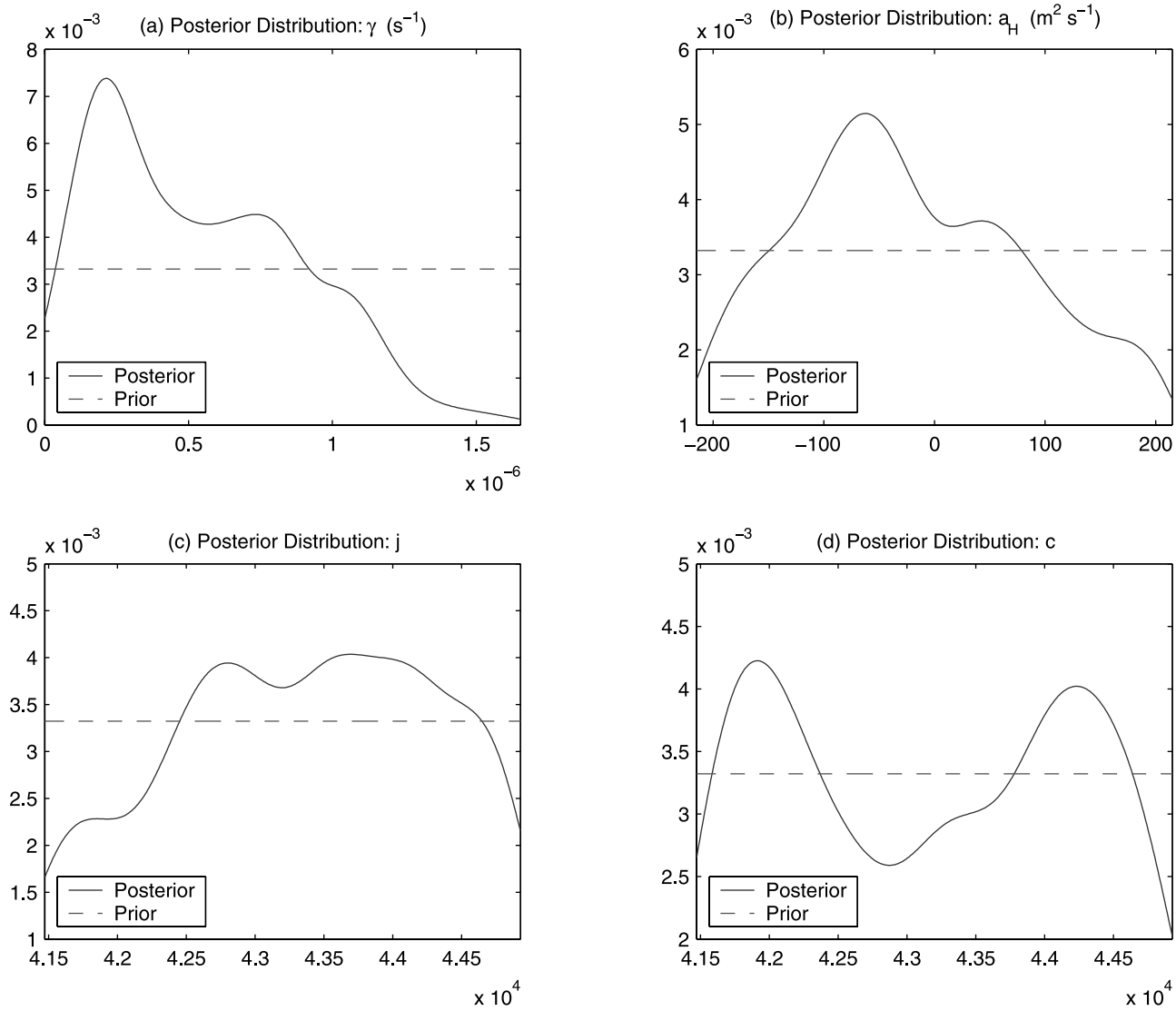


Figure 6. Plots of prior and posterior distributions for four BHM parameters described in the text. The prior distributions for all parameters are set to be uniform (i.e., flat lines) over the intervals depicted in each panel. (a) Bottom friction parameter, γ. (b) Lateral viscosity parameter, a_h. (c) Coefficient of the nonlinear terms in the BHM, j. (d) Coefficient of the wind-stress curl term, c.

in the southwestern corner of the domain, (2) the inconsistency between large dissipation implied by the BHM prior model and the more energetic circulation suggested by the observations, and (3) the fact that we deprived the model of some of its ability to respond to the truth by basing the ocean-data model (equation (15)) on a QG-like approximation. On the other hand, note that an alternative summary, the posterior mode (most likely value), outperforms the mean at some time points (e.g., Figure 7d). Furthermore, preliminary computations based on more sophisticated Monte Carlo methods (also, see the conclusion of Appendix B) than those used here also suggest that the true posterior results are less biased than those presented here.

5. Conclusion

[68] We have demonstrated how Bayesian hierarchical modeling can be applied to form stochastic models for complex processes, including the coupling of atmospheric and oceanic processes in the presence of unknown boundary conditions. In contrast to most views of "statistical models," these stochastic models can be based strongly on physical reasoning. Such prior process models are combined with observations by application of Bayes' Theorem. Throughout all stages of the model development and analysis, the primary objects of interest are probability distributions describing predata uncertainties and postdata uncertainties.

[69] To illustrate the strategy, we presented an OSSE based on a truth simulation of a basin-scale, wind-driven PE-SWE ocean model. The "true" processes were sampled and corrupted with noise to provide artificial data sets. This experiment was formulated to parallel the collection of scatterometer and altimeter data sets over the Labrador Sea. We then considered a QG-based, stochastic model and updated the model based upon data. We then provided extensive discussion of the results of the Bayesian analysis. These included summaries of posterior distributions of

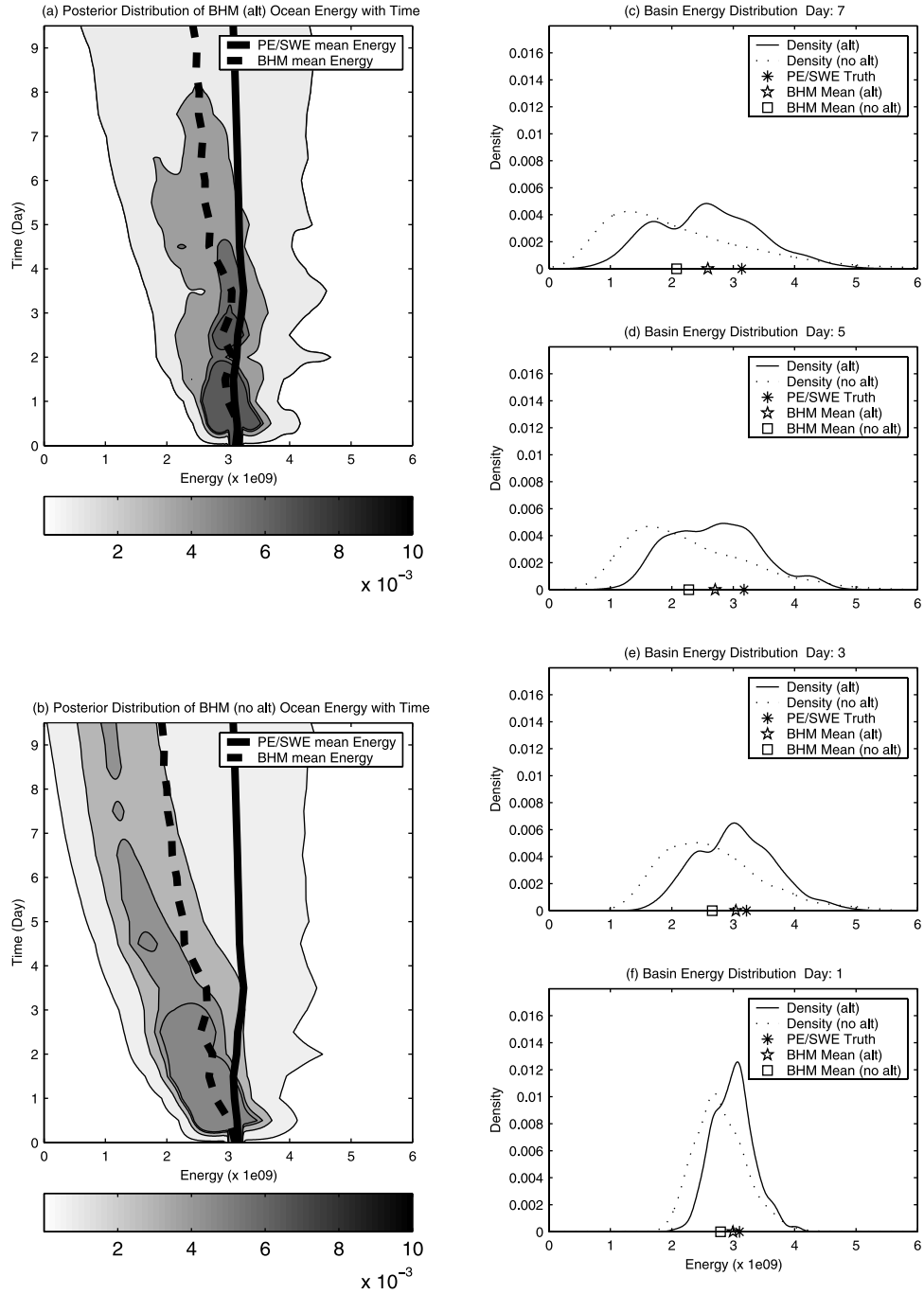


Figure 7. (a, b) Basin-integrated kinetic energy posterior distributions (gray scale), posterior mean traces (bold line), and "truth" simulation traces (bold dashed line). Figure 7a depicts the 10-day evolution of these posterior results for a BHM simulation incorporating simulated scatterometer and altimeter data sets. Figure 7b depicts the results when simulated altimeter data have been excluded. (c–f) Posterior densities for days 7, 5, 3, and 1.

primary variables of interest (ocean streamfunction) as well as functions of these variables (e.g., basin-scale total kinetic energy *KE*) and model parameters.

[70] While our overall impression of the results is quite favorable, we also indicated aspects of the results that depart from perfection (e.g., an apparent bias toward smaller-than-true *KE*). The following issues are important in gauging the results. First, we have tried to truly "forget" the true model in developing the statistical models. In particular, we did not direct the process models or the data models to account for behaviors beyond those associated with a QG-ocean, except for the inclusion of the stochastic error terms. Though error terms were included, they were not modeled using our knowledge that the truth was PE-SWE as opposed to QG. Similarly, we did little "tuning" of parameters or their priors based on the data nor our knowledge of the truth.

[71] A second issue is that our intent in BHM is to form effective combinations of information present in observa-

tions and physics, rather than more typical notions of seeking approximate solutions of uncertain physical models. This suggests that a typical BHM cannot be expected to maintain all the properties associated with traditional deterministic modeling and include uncertainty management. For example, while we investigated the implied posterior distributions on QG-model parameters such as γ and a_h, we never expected to obtain excellent statistical estimates of these parameters based on the model used. Indeed, we made clear that the statistical model involved time steps far too large to permit the interpretation that our model is a discretized QG model. On the other hand, this is not a limitation of the BHM approach in principle. Rather, we sought to indicate approximate modeling in the presence of serious computational limitations. Without such limitations, BHM can be used in a mode similar to that of solving a deterministic system in the presence of uncertainty.

[72] Though we did not seek to solve a deterministic system, some notion of stability regarding the model (equation (11)) is relevant. For example, recall that the OSSE was based on a 10-day model. What happens if we extend the study period beyond 10 days? While the formulas of Bayesian methods apply for arbitrary study periods, their numerical implementation may be difficult. In particular, a component of the numerical strategy described in Appendix B involves simulation of realizations from a stochastic dynamical system based on equation (11). Hence, we are concerned with stochastic stability. Unfortunately, this notion cannot be defined uniquely [e.g., *Meyn and Tweedie*, 1993; *Borovkov*, 1998]. *Meyn and Tweedie* [1993] consider forms of stochastic stability for Markov chains, though these results are difficult to transfer to our context since our model includes random parameters and equation (11) is merely conditionally Markov given the winds; that is, stochastic stability, however defined, applies to the entire model, not just to the streamfunction variables. This general topic, particularly as it relates to the efficiency of numerical procedures, is an important subject for future research.

[73] Finally, it is important to separately assess the value of the BHM paradigm and any particular implementation of that paradigm; that is, we presented a BHM with particular choices of the model components. In general, there are no unequivocal choices for such components, just as there are no unequivocal choices of parameterizations and data assimilation schemes in deterministic numerical models. Limitations of knowledge and practical implementation apply.

Appendix A

A1. Bayesian Computations

[74] Recalling the notation of section 2.1, the calculation of the normalizing constant $[D]$ in equation (1) is intractable in many applications. A variety of computational techniques have been considered [see *Berger*, 1985; *Bernardo and Smith*, 1994]. We focus on simulation-based, Monte Carlo methods.

A2. Importance Sampling Monte Carlo

[75] One of the main uses of conventional Monte Carlo is the estimation of integrals or expectations in the case of probability models. Let f be some function of the variable X of interest and assume a Bayesian context in which we are to estimate the expected value of $f(X)$ conditional on the data D,

$$E(f(X)|D) = \int f(x)[x|D]dx = \frac{\int f(x)[D|x][x]dx}{\int [D|x][x]dx}, \qquad \text{(A1)}$$

assuming this integral exists. A Monte Carlo estimate can be obtained by (1) generating M pseudo-random realizations or ensemble members from $[x|D]$, and (2) evaluating f for each realization and computing the arithmetic average of the results. Under independent sampling, this average tends to $E(f(X)|D)$ in a very strong sense as $M \to \infty$. Further, such convergence also holds if the realizations are stationary (or ergodic), though not necessarily independent. (This fact is critical in MCMC.)

[76] Typically, ISMC is suggested as a method for improving the efficiency of Monte Carlo (i.e., achieving the same or better expected precision with smaller sample sizes). The technique is also useful in settings in which simulation from $[x|D]$ is difficult. Consider another distribution, say q, which is comparatively easy to simulate from. One generates M ensemble members from q and evaluates f for each. To use these values to estimate equation (A1), each must be weighted to adjust for the fact that the ensemble members are not from the target posterior. A useful formula for an ISMC estimate of equation (A1) is [e.g., *Berger*, 1985; *Berliner*, 2001]

$$\hat{E}(f(X)|D) = \frac{\sum_i f(x_i)[x_i|D]/q(x_i)}{\sum_i [x_i|D]/q(x_i)}. \qquad \text{(A2)}$$

[77] Note that equation (A2) can be implemented even when normalizers of $[x|D]$ and/or q cannot be computed. A common selection of an importance sampler is the prior $[X]$. For a sample of size M from $[X]$, equation (A2) reduces to

$$\hat{E}(f(X)|D) = \frac{\sum_i f(x_i)[D|x_i]}{\sum_i [D|x_i]}. \qquad \text{(A3)}$$

[78] In our setting as well as much of the data assimilation literature, the primary variables of interest are a time series, say $\mathbf{X} = (X^1, \ldots, X^T)$. Suppose further that (1) our prior for the series is a Markov process model of order 1,

$$[\mathbf{X}] = [X^0] \prod_{t=0}^{T-1} [X^{t+1}|X^t]; \qquad \text{(A4)}$$

and (2) observations $D^1, \ldots, D^T$ are available. If we assume that these observations are conditionally independent and that the conditional distribution of any D_t depends only on X_t, we have that

$$[\mathbf{X}|D] \propto [X^0] \prod_{t=0}^{T-1} [D^{t+1}|X^{t+1}][X^{t+1}|X^t].$$

For Monte Carlo samples $\mathbf{x}_i = (x_i^0, \ldots, x_i^T)$ generated from the prior equation (A4), the estimate (equation (A3)) is

$$\hat{E}(f(\mathbf{X})|D) = \frac{\sum_i f(\mathbf{x}_i) \prod_{t=1}^{T} [D^t|x_i^t]}{\sum_i \prod_{t=1}^{T} [D^t|x_i^t]}. \qquad \text{(A5)}$$

Equation (A3) (with notation changes) appears at the top of page 265 of the work by *Berger* [1985]. It is the same (again, with notation changes) as the basis of ensemble

methods as described by *van Leeuwen and Evensen* [1996, p. 2903, equation (34)]. Indeed, equation (A5) is an ensemble smoother.

A3. Markov Chain Monte Carlo

[79] For many distributions, including settings in which the normalization of densities is not possible, one cannot readily generate realizations. In such cases, MCMC offers an approximate approach to generating realizations. The idea is to construct an ergodic Markov chain whose stationary, ergodic distribution coincides with the distribution of interest. Then, after a transience or burn-in period, numerically generated realizations of the chain are (in a limiting sense) realizations from the target distribution. However, these ensemble members are not independent, and hence, we require larger samples to achieve the same accuracy as can be expected with independent sampling. Theories have been developed that enable us to write out prescriptions for Markov chains with ergodic distributions that are posteriors from complex Bayesian models; that is, there are recipes for doing so, even though we cannot display the posterior. A general discussion of the approach is given by *Robert and Casella* [1999].

[80] Formal comparisons between ISMC and MCMC are difficult and beyond the scope of this article. There is one simple intuition, however. The intent in any case is to produce as efficiently as possible ensemble members which are representative of the true posterior. MCMC accomplishes this, but at a computational cost of enduring a transience period and potentially very large ensemble sizes compared to independent sampling. It can also be very difficult to implement for nonlinear models, especially in the case of coupled models as considered here. ISMC or ensemble filter/smoother methods may be preferred to MCMC in settings in which the posterior and prior distributions are not too dissimilar. The idea is that if the prior $[\mathbf{X}]$ and data model $[D|\mathbf{X}]$ (but viewed as a function of $\mathbf{X}$ for fixed D) favor very different regions of the $\mathbf{X}$-space, most of the values of $[D|\mathbf{x}_i]$ in equation (A5) would be very small. That is, most of the samples are wasted, and we again would expect the need for large ensemble sizes to achieve good accuracy. One hopes that this is partially alleviated by use of a sequential procedure. The general issue can be particularly problematic in large hierarchical models.

Appendix B

B1. Posterior Computations for the OSSE

[81] The posterior distribution of the primary physical variables of interest $\mathbf{U}$, $\mathbf{V}$ and $\boldsymbol{\Psi}$ (recall equation (5)) and all model parameters, conditional on the observations is given by

$$\begin{aligned}&\left[\mathbf{U},\mathbf{V},\psi,\theta_w,\theta_\psi,\eta_w,\eta_\psi|D_\psi,D_w\right]\\&\quad\propto\left[D_\psi|\psi,\theta_\psi\right]\left[\boldsymbol{\Psi}^1|\mathbf{U},\mathbf{V},\eta_\psi\right]\\&\qquad\prod\left[\boldsymbol{\Psi}^{t+1}|\boldsymbol{\Psi}^t,\mathbf{U},\mathbf{V},\eta_\psi\right]\left[\eta_\psi,\theta_\psi\right]\\&\qquad[D_w|\mathbf{U},\mathbf{V},\theta_w][\mathbf{U},\mathbf{V}|\eta_w][\eta_w,\theta_w].\end{aligned}\tag{B1}$$

The proportionality constant for this distribution is the integral of the right-hand side with respect to all variables except the data D_ψ, D_w.

[82] Though recipes for developing MCMC analyses exist, the highly nonlinear structures associated with our version of the wind-driven QG model render them intractable. However, a combined ISMC-MCMC approach is available. The key step is to apply Bayes' Theorem to the third line of equation (B1) only, isolating the wind process for the moment. It follows that

$$\begin{aligned}&\left[\mathbf{U},\mathbf{V},\boldsymbol{\Psi},\theta_w,\theta_\psi,\eta_w,\eta_\psi|D_\psi,D_w\right]\\&\quad\propto\left[D_\psi|\boldsymbol{\Psi},\theta_\psi\right]\left[\boldsymbol{\Psi}^1|\mathbf{U},\mathbf{V},\eta_\psi\right]\\&\qquad\prod\left[\boldsymbol{\Psi}^{t+1}|\boldsymbol{\Psi}^t,\mathbf{U},\mathbf{V},\eta_\psi\right]\left[\eta_\psi,\theta_\psi\right]\\&\qquad[\mathbf{U},\mathbf{V},\theta_w,\eta_w|D_w].\end{aligned}\tag{B2}$$

Though we cannot produce the normalizing constant for the third line of equation (B2), we can perform an MCMC analysis of this distribution [*Royle et al.*, 1999]. The output of this "atmosphere-only" MCMC are not realizations from the target posterior. (Intuitively, this must be so since such output could not be affected by D_ψ.) We (1) substitute a realization of $\mathbf{U}$, $\mathbf{V}$ from this MCMC into the first and second lines of equation (B2), along with randomly generated values of the model parameters, and (2) generate a realization of $\boldsymbol{\Psi}$ from the resulting model. Again, the resulting $\boldsymbol{\Psi}$ is not a realization from the final posterior. However, we can generalize equation (A5) to achieve the desired inferences.

B2. ISMC-MCMC Algorithm

[83] Our algorithm consists of two primary steps.

B2.1. Step A: MCMC the Atmospheric Model

[84] We construct and run an MCMC algorithm for the atmospheric model, which after transience, yields M samples from equation (B2), $\mathbf{U}_i$, $\mathbf{V}_i$, $\theta_{w,i}$, $\eta_{w,i}$: $i = 1,\ldots, M$.

B2.2. Step B: MCMC-ISMC Linkage

[85] Let $f = f(\mathbf{U}, \mathbf{V}, \boldsymbol{\Psi}, \theta_w, \theta_\psi, \eta_w, \eta_\psi)$ be any (integrable) function of the unknowns. To estimate posterior expectations of f, $E(f|D_\psi, D_w)$, perform the following.

B2.2.1. Step 1

[86] Choose an MCMC generated $\mathbf{U}_i$, $\mathbf{V}_i$ (step a). Independently generate $\theta_{\psi,i}$, $\eta_{\psi,i} \sim [\theta_\psi, \eta_\psi]$ (step b). Generate the full time series of ocean streamfunction from the process model, with parameters and winds from steps a and b (step c):

$$\boldsymbol{\Psi}_i \sim \left[\boldsymbol{\Psi}^1|\mathbf{U}_i,\mathbf{V}_i,\eta_{\psi,i}\right]\prod\left[\boldsymbol{\Psi}^{t+1}|\boldsymbol{\Psi}^t,\mathbf{U}_i,\mathbf{V}_i,\eta_{\psi,i}\right].$$

B2.2.2. Step 2

[87] Compute

$$\hat{E}(f)=\frac{\sum_i f_i\left[D_\psi|\boldsymbol{\Psi}_i,\theta_{\psi,i}\right]}{\sum_i\left[D_\psi|\boldsymbol{\Psi}_i,\theta_{\psi,i}\right]},\tag{B3}$$

where $f_i = f(\mathbf{U}_i, \mathbf{V}_i, \boldsymbol{\Psi}_i, \theta_{w,i}, \theta_{\psi,i}, \eta_{w,i}, \eta_{\psi,i})$.

[88] To clarify the form in equation (B3), note that we generated a realization from

$$\left[\mathbf{U},\mathbf{V},\boldsymbol{\Psi},\theta_w,\theta_\psi,\eta_w,\eta_\psi\right] \propto \left[\boldsymbol{\Psi}^1|\mathbf{U},\mathbf{V},\eta_\psi\right]\prod\left[\boldsymbol{\Psi}^{t+1}|\boldsymbol{\Psi}^t,\mathbf{U},\mathbf{V},\eta_\psi\right]\left[\eta_\psi,\theta_\psi\right][\mathbf{U},\mathbf{V},\theta_w,\eta_w|D_w].$$

This plays the role of q in our explanation of ISMC above. Dividing the correct posterior, p in the simpler version, as given in equation (B2) by this distribution, we see that only the data model $[D_\psi|\boldsymbol{\Psi},\theta_\psi]$ survives in the analogue of equation (A5).

[89] In the case of no ocean data, the correct full posterior reduces to

$$\left[\mathbf{U},\mathbf{V},\boldsymbol{\Psi},\theta_w,\eta_w,\eta_\psi|D_w\right] = \left[\boldsymbol{\Psi}^1|\mathbf{U},\mathbf{V},\eta_\psi\right]\prod\left[\boldsymbol{\Psi}^{t+1}|\boldsymbol{\Psi}^t,\mathbf{U},\mathbf{V},\eta_\psi\right]\cdot\left[\eta_\psi\right][\mathbf{U},\mathbf{V},\eta_w,\theta_w|D_w]. \quad \text{(B4)}$$

Note that this is an equality, not a proportionality relationship. Hence, inference for ocean streamfunction is accomplished directly with no need for reweighting. That is, the algorithm produces realizations from the correct posterior. Posterior expectations are estimated by

$$\hat{E}(f) = \frac{\sum_i f_i}{M}. \quad \text{(B5)}$$

One can compute both equations (B3) and (B5) based on a single simulation. This enables assessments of the impact of the addition of the ocean data in the presence of the atmospheric data.

[90] **Acknowledgments.** We are grateful for research support from the Oceanography Program of the NASA Earth Science Enterprise administered by Eric Lindstrom. Specifically, this material is based upon work supported by the NASA/Goddard Space Flight Center under Awards NAG5-8344 (MU), NAG5-8455 (OSU), and a contract from the Jet Propulsion Laboratory JPL-1860 to CoRA/NWRA. We are grateful to two referees for their useful comments.

References

Anderson, J. L., An ensemble adjustment Kalman filter for data assimilation, *Mon. Weather Rev.*, *129*, 2884–2903, 2001.

Bennett, A. F., *Inverse Methods in Physical Oceanography*, 134 pp., Cambridge Univ. Press, New York, 1993.

Berger, J. O., *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1985.

Berliner, L. M., Hierarchical Bayesian time series models, in *Maximum Entropy and Bayesian Methods*, edited by K. Hanson and R. Silver, pp. 15–22, Kluwer Acad., Norwell, Mass., 1996.

Berliner, L. M., Monte Carlo based ensemble forecasting, *Stat. Comput.*, *11*, 269–275, 2001.

Berliner, L. M., Z.-Q. Lu, and C. Snyder, Statistical design for adaptive weather observations, *J. Atmos. Sci.*, *56*, 2536–2552, 1999.

Berliner, L. M., C. K. Wikle, and N. Cressie, Long-lead prediction of Pacific SST's via Bayesian dynamic modeling, *J. Clim.*, *13*, 3953–3968, 2000.

Bernardo, J. M., and A. F. M. Smith, *Bayesian Theory*, John Wiley, New York, 1994.

Borovkov, A. A., *Ergodicity and Stability of Stochastic Processes*, 585 pp., John Wiley, New York, 1998.

Cressie, N. A. C., *Statistics for Spatial Data*, rev. ed., 900 pp., John Wiley, New York, 1993.

Daley, R., *Atmospheric Data Analysis*, Cambridge Univ. Press, New York, 1991.

Epstein, E. S., *Statistical Inference and Prediction in Climatology: A Bayesian Approach*, Am. Meteorol. Soc., Boston, 1985.

Evensen, G., and P. J. van Leeuwen, An ensemble Kalman smoother for nonlinear dynamics, *Mon. Weather Rev.*, *128*, 1852–1867, 2000.

Freilich, M. H., and R. S. Dunbar, The accuracy of NSCAT-1 vector winds: Comparisons with national buoy center buoys, *J. Geophys. Res.*, *104*, 11,231–11,246, 1999.

Meyn, S. P., and R. L. Tweedie, *Markov Chains and Stochastic Stability*, 550 pp., Springer-Verlag, New York, 1993.

Miller, R. N., E. F. Carter, and S. T. Blue, Data assimilation into non-linear stochastic models, *Tellus, Ser. A*, *51*, 167–194, 1999.

Milliff, R. F., and J. C. McWilliams, The evolution of boundary pressure in ocean basins, *J. Phys. Oceanogr.*, *24*, 1317–1338, 1994.

Milliff, R. F., and J. Morzel, The global distribution of the time-average wind stress curl from NSCAT, *J. Atmos. Sci.*, *58*, 109–131, 2001.

Millif, R. F., P. P. Niiler, A. E. Sybrandy, D. Nyehka, and W. G. Large, Mesoscale correlation length scales from NSCAT and Minimet surface wind retrievals in the Labrador Sea, *J. Atmos. Oceanic Technol.*, *20*, 513–533, 2003.

Pham, D. T., Stochastic methods for sequential data assimilation in strongly nonlinear systems, *Mon. Weather Rev.*, *129*, 1194–1207, 2001.

Pinardi, N., and R. F. Milliff, A note on consistent quasi-geostrophic boundary conditions in partially open, simply and multiply connected domains, *Dyn. Atmos. Oceans*, *14*, 65–76, 1989.

Renfrew, I. A., G. W. K. Moore, T. R. Holt, S. W. Chang, and P. Guest, Mesoscale forecasting during a field program: Meteorological support of the Labrador Sea deep convection experiment, *Bull. Am. Meteorol. Soc.*, *80*, 605–620, 1999.

Robert, C. P., and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, New York, 1999.

Robinson, A. R., P. F. J. Lermusiaux, and N. Q. Sloan III, *Data Assimilation, The Sea: The Global Coastal Ocean, I, Processes and Methods*, vol. 10, edited by K. H. Brink and A. R. Robinson, pp. 541–594, John Wiley, New York, 1998.

Royle, J. A., L. M. Berliner, C. K. Wikle, and R. F. Milliff, A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador sea, in *Case Studies in Bayesian Statistics IV,* edited by C. Gatsonis et al., pp. 367–382, Springer-Verlag, New York, 1999.

Scipione, C. M., and L. M. Berliner, Bayesian inference in nonlinear dynamical systems, paper presented at Section on Bayesian Statistical Science, Am. Stat. Assoc., Washington, D. C., 1993.

Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, 1986.

Spall, M. A., Buoyancy forced circulations around islands and ridges, *J. Mar. Res.*, *58*, 97–116, 2000.

Tarantola, A., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier Sci., New York, 1987.

Thiebaux, H. J., On approximations to geopotential and wind-field correlation structures, *Tellus, Ser. A*, *37*, 126–131, 1985.

van Leeuwen, P. J., An ensemble smoother with error estimates, *Mon. Weather Rev.*, *129*, 709–728, 2001.

van Leeuwen, P. J., and G. Evensen, Data assimilation and inverse methods in terms of a probabilistic formulation, *Mon. Weather Rev.*, *124*, 2898–2913, 1996.

Wikle, C. K., L. M. Berliner, and N. Cressie, Hierarchical Bayesian space-time models, *J. Environ. Ecol. Stat.*, *5*, 117–154, 1998.

Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner, Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds, *J. Am. Stat. Assoc.*, *96*, 382–397, 2001.

Wikle, C. K., L. M. Berliner, and R. F. Milliff, Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions, *Mon. Weather Rev.*, in press, 2003.

L. M. Berliner, Department of Statistics, Ohio State University, 404 Cockins Hall, 1958 Neil Avenue, CH 205B, Columbus, OH 43210-1247, USA. (mb@stat.ohio-state.edu)

R. F. Milliff, Colorado Research Associates, A Division of NorthWest Research Associates, 3380 Mitchell Lane, Boulder, CO 80301, USA. (milliff@colorado-research.com)

C. K. Wikle, Department of Statistics, University of Missouri, 222 Math Science Building, Columbia, MO 65211, USA. (wikle@stat.missouri.edu)