JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 108, NO. D24, 8776, doi:10.1029/2002JD002865, 2003

Physical-statistical modeling in geophysics

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

Received 20 August 2002; revised 21 January 2003; accepted 3 February 2003; published 17 September 2003.

[1] Two powerful formulas have been available to scientists for more than two centuries: Newton's second law, providing a foundation for classical physics, and Bayes's theorem, prescribing probabilistic learning about unknown quantities based on observations. For the most part the use of these formulas has been separated, with Newton being the more dominant in geophysics. This separation is arguably surprising since numerous sources of uncertainty arise in the application of classical physics in complex situations. One explanation for the separation is the difficulty in implementing Bayesian analysis in complex settings. However, recent advances in both modeling strategies and computational tools have contributed to a significant change in the scope and feasibility of Bayesian analysis. This paradigm provides opportunities for the combination of physical reasoning and observational data in a coherent analysis framework but in a fashion which manages the uncertainties in both information sources. A key to the modeling is the hierarchical viewpoint, in which separate statistical models are developed for the process variables studied and for the observations conditional on those variables. Modeling process variables in this way enables the incorporation of physics across a spectrum of levels of intensity, ranging from a qualitative use of physical reasoning to a strong reliance on numerical models. Selected examples from this spectrum are reviewed. *INDEX TERMS:* 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; 3210 Mathematical Geophysics: Modeling; *KEYWORDS:* Bayesian analysis, climate change, data assimilation, dynamics, hierarchical modeling, uncertainty

Citation: Berliner, L. M., Physical-statistical modeling in geophysics, *J. Geophys. Res.*, *108*(D24), 8776, doi:10.1029/2002JD002865, 2003.

> So far as the laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality.
>
> Albert Einstein (1921)

1. Introduction

[2] There is a prevalent view that there are two sorts of geophysical models: physical and statistical models. Physical models arise as applications of the laws of classical physics. The spirit of implementation is deterministic, and the results are mechanistic: Explanation and prediction of natural phenomena are based on the development and application of mathematical representations of the laws of physics. The view of statistical models is that they are descriptive. In particular, they do not rely on the use of physics. The perspective in this article is that these two views are endpoints of a spectrum of physical-statistical models. These models are intended to combine physical and observational views in a coherent, operational fashion.

[3] Neither the physical nor statistical view ignores the other. Physicists rely on observational data in developing and assessing physical models and in estimating key quantities in models, including initial and boundary conditions. On the other hand, the construction of statistical models is often guided by science. At the least, science is usually used to select the variables of interest and the data sets to be analyzed. However, the two views are sometimes combined in a haphazard fashion in complex settings. In some cases, there are deliberate attempts not to combine the approaches but rather only to compare and contrast model output and observations.

[4] I do not mean to suggest that either traditional view of modeling is wrong or useless. Indeed, science has made great progress based on these approaches. Rather, I mean to suggest that physical-statistical modeling, especially implemented via Bayesian analysis, can add to the tools available. One motivation for this claim is the critical issue of uncertainty management. Applications of physics in complex settings hinge on various approximations (e.g., "neglecting higher-order terms") and unsure assumptions (e.g., specifications of relevant forces). In addition, when the laws lead to certain forms of nonlinear relationships, chaotic behavior may eliminate the possibility of precise prediction, opening the door to probabilistic techniques [e.g., *Berliner*, 1992]. In modeling the atmosphere and oceans, practical, numerical implementations of the laws of physics typically treat only selected, space-time-averaged variables. This leads to concerns regarding sub-grid-scale parameterizations, up-scaling and down-scaling, etc. Furthermore, observational data are subject to a variety of errors. Failure to account for such difficulties is problematic. For example, it is customary to

0148-0227/03/2002JD002865

estimate or "compute" the values of unknown model parameters based on observations and to use the results as if they are the truth. Sampling variability of such estimates is important, especially in nonlinear contexts, where (1) standard methods of propagation of errors are too difficult to implement and (2) the model displays sensitive dependence on parameters. It is fair to claim that, in practice, only approximate physics are applied approximately and unsurely.

[5] On the other hand, purely statistical models, either traditional or Bayesian, may be inefficient in complex settings. First, though massive observational data sets are available and even more data collection is anticipated, these data do not capture the complete behavior of the Earth system. While it should be undeniable that the more we observe, the more we know, the extraction of the useful information from massive data sets representing spatial-temporal processes is challenging. Also, some aspects of the system are well represented in the observations, while others are not, for practical reasons. A second issue regarding the use of purely statistical models arises from their value in prediction, especially against a background of potential changes in the underlying system's behavior. In particular, a purely statistical approach to climate change analysis and prediction of future climate changes and their impacts seems improbable.

[6] The Bayesian statistical view is often acknowledged to be the most natural approach for combining information sources while managing their associated uncertainties. Bayesian predictive analyses focus on the production of predictive distributions for the quantities of interest, formally conditioned on the observed data. That is, the Bayesian "answer" is a distribution rather than a single forecast. Of course, such distributions are often summarized via single forecasts; to be a real value, though, these should be accompanied by measures of confidence or accuracy. Such measures of uncertainty are, by nature, sensitive to the natural variation of the process to be predicted and to errors in the data underlying the analysis as well as to other sources of uncertainty. Even in the presence of hypothesized underlying determinism, uncertainty leads to operational solutions that are fundamentally statistical/probabilistic in form [e.g., *Epstein*, 1969, 1988]. Whether or not one adopts a deterministic philosophy is irrelevant operationally if the determinism cannot be implemented with certainty.

[7] Nevertheless, Bayesian statistics has made only a few inroads in geophysical applications. (An exception is modern data assimilation techniques, which I discuss below.) One reason for this is that there have been severe practical limitations in the simultaneous application of Newton and Bayes. This situation is changing, owing in large part to advances in modeling strategies, data collection, and computational assets and algorithms.

[8] Four clarifications are needed. The first involves the scope of Bayes's theorem. The result is simply a theorem in probability theory. Its use in the context of (traditionally defined) random variables with known distributions is uncontroversial and common. I distinguish such uses from Bayesian statistics, which involves modeling unknown distributions and updating those models based on data. (This viewpoint has also been suggested in climate science [e.g., *Epstein*, 1985].) This distinction is important; some scientists may well recognize the value of Bayes's theorem but contend that it requires perfect knowledge of the "true" underlying probability distributions.

[9] Second, modern data assimilation methods go beyond the traditional issue of initial condition uncertainty and recognize the difficulties associated with forms of model uncertainty and data errors, seeking to respond to them by developing "optimal" data assimilation schemes, estimates of analysis errors, ensemble forecasting, etc. [e.g., *Ehrendorfer*, 1997]. Furthermore, these procedures combine physics and data and, in fact, are known to be related to certain Bayesian procedures [e.g., *Anderson*, 2001; *Evensen and van Leeuwen*, 2000; *Lorenc*, 1986; *Pham*, 2001; *Tarantola*, 1987]. For example, use of the Kalman filter is prevalent; it is also known to be derivable via Bayes's theorem [e.g., *Meinhold and Singpurwalla*, 1983]. Many optimal interpolation schemes also have Bayesian interpretations. However, standard applications assume model parameters to be known; a richer Bayesian statistical view treats these parameters as unknown. This is not an article about data assimilation. It is about widening the scope of the meaning of "combine physics and data" to account for alternative ways of bringing physics to bear in statistical models, while managing uncertainty in the effort.

[10] Third, physical reasoning can lead to stochastic models. There is a long tradition of stochastic modeling in turbulence theory. Stochastic models arise from various types of averaging. For example, Reynold's averaging involves behavior across an ensemble of realizations of the physical process of interest. Spatial and/or temporal filtering of processes and their associated underlying mathematical equations leads to closure problems having stochastic interpretations. Nevertheless, the philosophical basis of these methods is mechanistic: Stochastic models are derived from the laws of physics, at least in principle.

[11] Finally, as I proselyte about Bayesian hierarchical modeling, I do not mean to suggest that it easily solves all problems (regardless of my more private view). The approach can be difficult from modeling, computational, and interpretative standpoints; indeed, there are many devils in the details. One should not expect to write down Bayes's theorem and have the definitive analysis result (anymore than writing down the Navier-Stokes equations leads directly to an ideal, implemented global numerical weather forecasting system). I do believe that there is strong potential for new and effective analyses, though more research on geophysical applications is needed. Further, even if formal Bayesian methods are too difficult or can be only roughly approximated, there is value to thinking like a Bayesian in model formulation and use.

[12] Section 2 is a brief review of Bayesian hierarchical analysis and computation. Section 3 opens with a review of the Lagrangian and Eulerian views in fluids models. An exercise from *Holton* [1992] on computations using the total derivative is used to develop a Bayesian illustration. The simple analysis motivates potential applications. In section 4, climate modeling and climate change analyses are cast in the Bayesian hierarchical framework, and an example analysis is reviewed. Two examples of analyses of near-surface ocean winds are discussed in section 5. The reviews of these three examples are brief and are directed toward illustrating the development of physical-statistical

models. The source publications provide more complete discussions of the models, their analyses, and their conclusions. A concluding discussion is given in section 6.

2. Bayesian Hierarchical Modeling

[13] Let $[X]$ represent a probability distribution for a random variable X; $[X|Y]$ represents the conditional probability distribution of X given Y. (I will not use the more intricate notation of differentiating between a random variable and a possible value of that variable.) Bayes's theorem provides a mechanism for updating our prior distribution $[X]$ based on data, say, Y. That is, the posterior distribution (i.e., after observing the data) is

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

where, for a continuous variable X, $[Y] = \int [Y|X][X]\, dX$ (for a discrete variable the integral is replaced by a sum). Extensive, general discussions of Bayesian analysis as well as comparisons to other statistical approaches can be found in the work of *Berger* [1985] and of *Bernardo and Smith* [1994].

[14] The cornerstone of hierarchical modeling is the notion of conditional thinking. The basis is a fact from probability theory. The joint probability distribution of a collection of random variables $X_1, \ldots, X_k$ can be represented as a product of conditional distributions:

$$[X_1, \ldots, X_k] = [X_1|X_2, \ldots, X_k]\,[X_2|X_3, \ldots, X_k]\,\ldots\,[X_k]. \tag{2}$$

The suggestion is to develop the individual distributions on the right-hand side of equation (2), rather than taking on the more formidable task of specifying the left-hand side directly.

[15] The following view has proven useful in geophysical applications. First, we separate the unknowns into two groups: process variables and the model parameters. Process variables are to include the actual physical quantities of interest (e.g., temperature and pressure in fluids problems), while parameters are quantities introduced in model development (e.g., propagator matrices in linear dynamical models, measurement error and sub-grid-scale variances, unknown physical constants).

[16] A Bayesian hierarchical model (BHM) consists of three primary components [e.g., *Berliner*, 1996]: [data| process, parameters]; [process|parameters]; and [parameters]. Probability theory ensures us that a bona fide joint prior distribution for the process and parameters is given by the product

$$[\text{process, parameters}] = [\text{process}|\text{parameters}][\text{parameters}]. \tag{3}$$

Bayes's theorem provides [process, parameters|data], which, in turn, serves as the basis for inference and prediction.

[17] Our goal is the merger of statistical and physical reasoning in the development of all three primary components of hierarchical models. Two key points are demonstrated. First, modeling observational data conditionally on the underlying processes of interest can lead to major simplifications in the analysis. One such simplification is the use of Lagrangian data for inference on Eulerian field properties (section 3.3.2). Also, this approach enables the incorporation of a variety of diverse observational data sets. Second, physical modeling of fluids is directly related to conditional thinking, which, in turn, implies that we can embed physical reasoning into a hierarchical strategy.

[18] Bayesian analysis and its hierarchical implementation are not new. However, until recently, computational issues limited applications in large and/or complex settings. The key issue is that the computation of the denominator (i.e., normalizing constant) on the right-hand side of equation (1) is generally very difficult. The problem of calculating partition functions in statistical mechanics offers a nice parallel.

[19] Owing to this problem, much of the Bayesian literature has been dominated by the treatment of relatively simple and typically low-dimensional problems. Special models were often selected because the computations were possible (e.g., reliance on so-called conjugate priors; see *Berger* [1985] and *Epstein* [1985]). In low-dimension settings, numerical integration, analytic approximations, and Monte Carlo integration were also used. Such methods are typically inefficient in high dimensions (e.g., the example discussed in section 5.2 involves a posterior distribution in roughly 160,000 dimensions). The scope of application has been widened with the recent development of simulation techniques, chiefly the Markov Chain Monte Carlo technique (MCMC). The Bayesian use of MCMC has its roots in Monte Carlo approaches in statistical mechanics. The essence is to rig a Markov chain whose stationary, ergodic distribution coincides with the posterior distribution of interest. After a transient burn-in period, numerically simulated realizations compose an ensemble of identically distributed, though dependent, realizations from the posterior. Both applications reviewed in section 5 rely on MCMC. Additional geophysical examples include those of *Berliner et al.* [2000b, 2003]. General introductions to the subject can be found in the work of *Gilks et al.* [1996] and *Robert and Casella* [1999]. For brevity I focus the rest of this article on modeling.

3. An Elementary Example

3.1. Kinematics of Fluids

[20] I briefly review mathematical models to establish notation. A fluid, such as the atmosphere and ocean, is viewed as a continuum. Physics is applied to infinitesimal fluid parcels of the continuum. In a Lagrangian description the behavior of a fluid is based on the combined behaviors of the trajectories and properties of its parcels. In an Eulerian description we [*Salby*, 1996, p. 322] "describe atmospheric behavior in terms of field variables that represent the distributions of properties at particular instants." A Eulerian field variable (e.g., temperature, density, pressure) is a time-varying, spatial field.

[21] For smooth variables, these views are related via the total derivative (also called the material derivative), the substantial derivative, and Stokes's derivative. Suppose the fluid flows in a three-dimensional space, indexed in Cartesian coordinates $\mathbf{x} = (x, y, z)'$, where x represents an east-west coordinate, y is north-south, and z is vertical. For a given

parcel, its location in space at time t is the vector $\mathbf{x}(t) = (x(t), y(t), z(t))'$. Let ψ denote a selected variable associated with a parcel. Its value at time t can be viewed as a function of t. Assume we can differentiate $\psi(t)$; that is, the derivative $d\psi(t)/dt$ exists. To calculate this derivative, note that ψ at time t may be written as $\psi(\mathbf{x}(t),t)$. Consider its total derivative:

$$\frac{D\psi(\mathbf{x}(t),t)}{Dt} = \frac{d\psi(t)}{dt}.$$

Applying total differentiation (i.e., total differentiation of a function of several variables plus the chain rule), we obtain

$$\frac{D\psi(\mathbf{x}(t),t)}{Dt} = \frac{\partial\psi(\mathbf{x},t)}{\partial t} + \frac{\partial\psi(\mathbf{x},t)}{\partial x}\frac{dx(t)}{dt} + \frac{\partial\psi(\mathbf{x},t)}{\partial y}\frac{dy(t)}{dt} + \frac{\partial\psi(\mathbf{x},t)}{\partial z}\frac{dz(t)}{dt}. \tag{4}$$

[22] A key notion underlying much of the analysis of fluids is kinematic consistency, as described by *Salby* [1996, pp. 322–323]:

> . . .the Eulerian description is related to the Lagrangian description by a fundamental kinematic constraint: The field property at a given location and time must equal the property possessed by the material element occupying that position at that instant.

Note that equation (4) is fundamentally a Lagrangian relationship. However, the partial derivatives in equation (4) are interpreted as being Eulerian in the following sense. If we were to actually compute the total derivative for a specified ψ at a fixed time s, the partial derivative with respect to time would enter as

$$\frac{\partial\psi(\mathbf{x},t)}{\partial t}\Big|_{\mathbf{x}=\mathbf{x}(s),t=s};$$

that is, the partial derivative of the Eulerian ψ function evaluated at time s and spatial location $\mathbf{x}(s)$. This is what the mathematics dictates, though it has an important intuitive content and practical implications.

[23] As an example, consider the following exercise from the work of *Holton* [1992, pp. 29–30]:

> The surface pressure decreases by 0.3 kPa/180 km in the eastward direction. A ship steaming eastward at 10 km/h measures a pressure fall of 0.1 kPa/3 h. What is the pressure change on an island that the ship is passing?

On the basis of the first sentence, assume that surface pressure $p(x, t)$ is given by

$$p(x,t) = c(t) - \frac{1}{600}\,\text{kPa/km}\,x \tag{5}$$

for some function $c(t)$, where t is in hours and x is in km. Lagrangian measurement of pressure fall on the ship provides the total derivative:

$$\frac{Dp(x(t),t)}{Dt}\Big|_{t=t_0} = -\frac{1}{30}\,\text{kPa/h},$$

where t_0 is the time the ship passes the island (i.e., $x(t_0) = x_I$, the location of the island). From equation (4) we calculate

$$\frac{\partial p(x_I,t)}{\partial t}\Big|_{t=t_0} = -\frac{1}{30}\,\text{kPa/h} - \left(-\frac{1}{600}\,\text{kPa/km}\right)(10\,\text{km/h}) = -\frac{1}{60}\,\text{kPa/h}. \tag{6}$$

3.2. A Bayesian Analysis

[24] To transition to statistical modeling, consider actual pressure measurement. For a small time interval of length δ, suppose an islander measures pressure. At times t_0 and $t_1 = t_0 + \delta$ we expect those measurements to be $c(t_0) - 1/600$ kPa/km x_I and $c(t_1) - 1/600$ kPa/km x_I, respectively. Hence the pressure change should be $c(t_1) - c(t_0)$. Dividing by δ and then letting $\delta \to 0$, we obtain the Eulerian derivative we seek. Even if these measurements are not collected and we do not know $c(t)$, we can estimate its derivative at t_0 using the data from the ship and the formula for the total derivative. On the ship we expect corresponding pressure measurements of $c(t_0) - 1/600$ kPa/km x_I and $c(t_1) - 1/600$ kPa/km $(x_I + 10\delta)$. The pressure difference is then

$$(c(t_1) - c(t_0)) - \frac{1}{600}\,\text{kPa/km}\,10\delta,$$

again, after division by δ, as $\delta \to 0$. We obtain the Lagrangian derivative $Dp(x(t), t)/Dt|_{t=t_0}$ observed on the ship. We see that equation (6) is just the right algebra.

3.2.1. Data Model

[25] The formulation of the expectations (see equations (7) and (8)) of the four pressure observations will be done conditionally using knowledge of the space-time Eulerian pressure field and on the knowledge of the motion of the ship. This construction does not require that we truly know those properties; if we did, we would not need any data. Rather, they are variables in a conditional probability distribution.

[26] It is convenient to parameterize $c(t)$ as a linear function $c(t) = a + bt$, so a(kPa) and b(kPa/h) are unknown. (This is hardly a limiting assumption here since we consider only two pairs of observations.) Also, assume that the slope of the pressure field, say, M(kPa/km), is known and is equal to $-(1/600)$ kPa/km. (M could also be modeled as unknown.) Hence we seek inference on the parameter b based on the observations. The key step is to model the four potential observations: E_0 and E_1 on the island and L_0 and L_1 on the ship. Consider the measurement error models

$$E_0 = a + bt_0 + Mx_I + e_0, \quad E_1 = a + bt_1 + Mx_I + e_1; \tag{7}$$

$$L_0 = a + bt_0 + Mx_I + \lambda_0, \quad L_1 = a + bt_1 + M(x_I + 10\delta) + \lambda_1, \tag{8}$$

where e_0, e_1, λ_0, and λ_1 are all mutually independent zero-mean random measurement errors, for example, Gaussian (or "normal") random variables with a common, known variance σ^2. (More generally, we could allow σ^2 to be unknown; we could also model different variances for the islander's and ship's errors.) The statistical argument views these models as specifying the joint distribution of the data conditional on the parameters, which takes the following form due to the independence assumption:

$$[E_0, E_1, L_0, L_1|a, b] = [E_0|a, b][E_1|a, b][L_0|a, b][L_1|a, b]. \quad (9)$$

[27] Continuing to assume that the slope $M = -(1/600)$ kPa/km of the pressure field is known, we seek inference on the parameter b. As in *Holton*'s [1992] original problem, suppose that inference is to be based on the observations L_0 and L_1 only. Let $d = L_1 - L_0$. (This eliminates a from the analysis.) Assuming Gaussianity and conditional independence, we have

$$[d|b] \text{ is } N\left(\delta(b + 10M), 2\sigma^2\right) \quad (10)$$

or, equivalently,

$$\left[\frac{d}{\delta} - 10M|b\right] \text{ is } N\left(b, 2\frac{\sigma^2}{\delta^2}\right). \quad (11)$$

(The notation $N(\mu, s^2)$ denotes a normal distribution, with mean μ and variance s^2.) To match *Holton*'s [1992] data, assume that the observed value is $d/\delta = -(1/30)$ kPa/h. I set $\delta = 1/6$ hours (10 min) and $\sigma^2 = 0.0001$. Note that the time step δ does not disappear in these calculations because I am not letting $\delta \to 0$. Indeed, as δ decreases, the variance of the least squares estimator of b increases. This arises because I am adjusting for measurement errors and have assumed a linear model. In the presence of measurement errors and a linear model, optimal least squares estimators (in terms of smallest variance) are obtained by taking observations as far apart as possible.

3.2.2. Process Model

[28] The prior for b is assumed to also be Gaussian:

$$[b] \text{ is } N\left(\mu, \tau^2\right), \quad (12)$$

where μ and τ^2 are to be chosen based on prior information. For example, assume that the linear pressure field arises from an approximation during the passing of a weather system. Our prior mean μ could be based upon information on the rate of passage of this system or, if that information is not available, we could use climatology. Further, b and M both arise in an approximation, suggesting that they would be modeled as dependent variables. I set $\mu = -0.05$ kPa/h, and since I am not very knowledgeable about these issues, I assigned a rather large prior variance: $\tau^2 = 0.01$.

3.2.3. Posterior Analysis

[29] The posterior distribution $[b|d]$ is also Gaussian [e.g., *Berger*, 1985, pp. 127–128], with mean

$$\frac{\tau^2}{\tau^2 + 2\frac{\sigma^2}{\delta^2}}\left(\frac{d}{\delta} - 10M\right) + \frac{2\frac{\sigma^2}{\delta^2}}{\tau^2 + 2\frac{\sigma^2}{\delta^2}}\mu = -0.030 \quad (13)$$

and variance

$$\frac{2\frac{\sigma^2}{\delta^2}\tau^2}{\tau^2 + 2\frac{\sigma^2}{\delta^2}} = 0.0042. \quad (14)$$

Hence a natural posterior Bayesian point estimate of the rate of change in pressure on the island is -0.03 kPa/h; a 95% Bayesian credible interval for this rate is $(-0.157, 0.097)$.

3.3. Discussion

3.3.1. Combining Data Sets

[30] We can readily incorporate observations made on the island. Manipulating the data model, we have that

$$\left[\frac{E_1 - E_0}{\delta}|b\right] \text{ is } N\left(b, 2\frac{\sigma^2}{\delta^2}\right). \quad (15)$$

Since both $(E_1 - E_0)/\delta$ and $d/\delta - 10M$ have mean b and the same variances, their numerical average, say, $\bar{d}$, has the conditional distribution

$$\left[\bar{d}|b\right] \text{ is } N\left(b, \frac{\sigma^2}{\delta^2}\right). \quad (16)$$

(If these two statistics had different variances, we would still "average" them but not with equal weights; rather, we would average them with individual weights related inversely to the individual variances.) Once $\bar{d}$ is found, we can recompute equations (13) and (14) with d replaced by $\bar{d}$ and with the data-model variance modified appropriately.

[31] There is an alternate derivation of the above posterior distribution. Owing to the conditional independence of all four observations, we could directly apply Bayes's theorem using the data model of equation (15) and the posterior $[b|d]$, based on the ship observations as obtained earlier, as if the posterior was our prior. This is a very general fact and does not rely on the specific distributional assumptions. Such results are very important in sequential problems in which data are collected over time (e.g., weather forecasting). Intuitively, in the presence of conditionally independent data collected over time we have that "today's posterior is tomorrow's prior."

3.3.2. Lagrangian Data

[32] I claim that substantial simplifications in such settings arise if one constructs data models for Lagrangian data conditional on appropriate Eulerian fields. The basis of my claim is kinematic consistency. This is exemplified in the construction of the data model for *Holton*'s [1992] exercise. The point is that if we can get this far in modeling the observations, further references to the Lagrangian versus Eulerian origins of the data are irrelevant. In this example, instead of L_0 being taken on the ship, the observation could have been a replicated observation on the island; L_1 could have been a Eulerian observation taken on the island at location $x_I + 10\delta$ at time t_1.

[33] These arguments apply quite generally to the problem of inference for Eulerian properties based on Lagrangian data. Kinematic consistency implies that conditioning on complete space-time Eulerian fields removes some of the difficulties in processing Lagrangian data, at least in principle. Many statistical analyses of Lagrangian data are based on marginal, descriptive models of Lagrangian properties. To clarify, recall that the Bayesian calculations for *Holton*'s [1992] example used the conditional independence of the

observations. However, the observations are not marginally (unconditional on b) independent. Rather, they covary because they all respond to the variability in b. For example, unconditional on b, the implied covariance cov(E_0, E_1) between E_0 and E_1 is $\tau^2 t_0(t_0 + \delta)$; cov(L_0, L_1) is also equal to that value but cov(E_0, L_0) = $\tau^2 t_0^2$ and cov(E_1, L_1) = $\tau^2 (t_0 + \delta)^2$. Operating in a conditional mode in which independence among measurement errors was plausible meant that we did not need to concern ourselves with finding these quantities.

[34] I do not mean to imply that the issues of Lagrangian data are made trivial. That is hardly the case. Since we typically consider space-time-gridded Eulerian fields, formulating the data model can be very challenging and can involve sub-grid-scale modeling. I do suggest that the effort is well spent and may be no worse than similar sub-grid-scale modelings of Eulerian observations.

4. Climate Modeling

4.1. Bayesian Climate Modeling

[35] A general formulation of Bayesian climate modeling finds its origin in the definition that climate is average weather or, more completely, a statistical description of relevant variables. A simple sketch of the formulation begins with the consideration of three sets of variables: Let **O** represent a collection of observations to be used, and let **W** represent the values of some selection of relevant weather variables (e.g., temperature, rainfall). The third collection of variables, denoted by **C**, contains variables that serve as quantifications of climate; formally, they serve as parameters in probability distributions of weather. These three collections of variables are defined in space and time, but the extra notation is suppressed for now.

[36] To formalize the ideas, consider a Bayesian hierarchical strategy with four basic stages: $[\mathbf{O}|\mathbf{W}, \boldsymbol{\theta}]$; $[\mathbf{W}|\mathbf{C}, \boldsymbol{\theta}]$; $[\mathbf{C}|\boldsymbol{\theta}]$; and $[\boldsymbol{\theta}]$, where $\boldsymbol{\theta}$ is a set of model parameters. This framework is not revolutionary. Indeed, some of the notions are suggested by *Hasselmann* [1976]. One of the ideas there is to develop a statistical dynamical model for climate evolution using the notion of probabilistic modeling of stochastic forcings of the climate without an explicit requirement that the statistics of climate be based on averages of all weather variables. In the language of modeling here, *Hasselmann* [1976] builds a version of $[\mathbf{C}|\boldsymbol{\theta}]$. This is an important distinction in that strict adherence to the common definition seems to inevitably link climate and weather variables in the following sense: The application of physics and probabilistic thinking may appear only to lead to pairs of distributions $[\mathbf{W}|\mathbf{C}, \boldsymbol{\theta}]$ and $[\mathbf{C}|\mathbf{W}, \boldsymbol{\theta}]$ [cf. *Hasselmann*, 1976, equations (2.2) and (2.3)] but leaving $[\mathbf{C}|\boldsymbol{\theta}]$ unspecified without resorting to arbitrary closure schemes.

[37] Substantial flexibility can be achieved by viewing climate variables as parameters of a distribution for physical variables. As exemplified in the example that follows (section 4.2), "climate variables" can be based upon a useful dimension reduction.

4.2. Bayesian Fingerprinting

[38] *Berliner et al.* [2000a] use BHMs to study the climatic response to anthropogenic CO_2 forcing. Although the context is simplified, their modeling is a Bayesian approach to fingerprinting, namely, statistical inference based on **O**, the observed temperatures on a global spatial grid over a given period of time, regarding the quantity a in the relation

$$\mathbf{O} = \mathbf{g}a + \text{noise}, \tag{17}$$

where **g** is a purely spatial pattern associated with CO_2 impacts. The noise term is modeled to account for both errors in the data and natural variability.

[39] In the notation of section 4.1, **C** consists of the pattern **g** and the amplitude a; in *Berliner et al.* [2000a], **g** is assumed to be known. The "weather" variables are gridded true annual global surface temperatures, denoted by **T**. The observational data set **O** used is a portion of the Jones annual temperature data [*Jones et al.*, 1999]. Both **T** and **O** are vectors of these variables stacked over the years used in the study.

4.2.1. Data Model

[40] The data model relates the observations to the gridded true temperatures. The data are assumed to be unbiased; that is, the measurement errors all have means equal to zero. These errors are assumed to follow a normal distribution, leading to the model

$$[\mathbf{O} \mid \mathbf{T}, \Sigma_o] \text{ is } N(\mathbf{XT},\ \Sigma_o). \tag{18}$$

The notation N indicates a multivariate normal distribution, with mean vector **XT** and covariance matrix Σ_o. The matrix **X** simply takes care of the bookkeeping of matching up the observations to the global grid and accounting for missing data. The covariance Σ_o was assumed to be known and was developed from estimates given by *Jones et al.* [1997].

4.2.2. Process Model

[41] *Berliner et al.* [2000a] considered the statistical model

$$[\mathbf{T} \mid a, \mathbf{g}, \Sigma_T] \text{ is } N(\mathbf{g}a, \Sigma_T), \tag{19}$$

where the covariance Σ_T is intended to capture spatial-temporal dependence in global annual temperatures.

[42] The Bayesian treatment of detection and attribution of climate change suggests a prior distribution as a mixture of two main components, as follows: With probability p, a has a normal distribution with mean 0 and variance τ^2, and with probability $1 - p$, a has a normal distribution with mean μ_A and variance τ_A^2; that is,

$$[a] \text{ is } pN(0, \tau^2) + (1-p)N(\mu_A, \tau_A^2). \tag{20}$$

The prior probability p of the "no impact" hypothesis was allowed to vary in the analyses.

[43] The goal is then to draw posterior inferences on a, i.e., produce and inspect $[a|\mathbf{O}]$ (this distribution is again a mixture, but all parameters, including p, are updated in light of the data). I refer the reader to the source article for a description of the results. In line with the goals of this

article I indicate how physical reasoning was used to develop the statistical models. *Berliner et al.* [2000a] made substantial use of model output from the National Center for Atmospheric Research climate system model (CSM) [*Meehl et al.*, 2000]. First, to estimate Σ_T, natural climate variability was estimated from a 300-year control run of the CSM. Second, a 120-year CO_2-forced run was used in combination with the control run to construct a CO_2 pattern **g**. Specifically, *Berliner et al.* [2000a] computed differences between temporal averages of the final 100 years of the forced run and of the 100 years of the control run at each model grid box. Finally, the parameters τ^2, μ_A, and τ_A^2 in equation (20) were developed by (1) subsampling the CSM control and CO_2-forced runs into smaller batches and by (2) computing means and variances of estimates of a based on the batches.

5. Near-Surface Ocean Winds

5.1. A Stochastic Geostrophic Model

[44] *Royle et al.* [1998] considered the estimation of gridded near-surface winds over the Labrador Sea. The data set used is composed of scatterometer-based wind estimates. Let **S** denote the vector of scatterometer wind estimates. Each "element" of **S** is a two-dimensional vector of estimates of the orthogonal westerly and southerly components. These estimates are supported on grid boxes of the order of 50 × 50 km. The locations of these boxes vary with the path of the satellite bearing the scatterometer. The goal was a spatial prediction analysis of horizontal wind averages over a fixed grid. The authors considered the European Centre for Medium-Range Weather Forecasts (ECMWF) grid. Let **W** denote the vector of these gridded true winds (as above, each element of **W** is a two-dimensional vector of the u wind and v wind).

[45] Their modeling focused on the geostrophic approximation: Local geostrophic winds are proportional to the gradient of the pressure field. At a location the geostrophic wind components u_g and v_g satisfy

$$v_g \propto \frac{\partial p(x,y)}{\partial x}; \tag{21a}$$

$$u_g \propto -\frac{\partial p(x,y)}{\partial y}. \tag{21b}$$

Assuming the density of the air to be constant in space, the constant of proportionality is essentially the Coriolis parameter, which depends on the latitude of the location [*Holton*, 1992, p. 40].

[46] The key idea in constructing the hierarchical model is to view equations (21a) and (21b) as suggesting a model for winds conditional on pressure. Let **P** denote the vector of gridded (on the same grid as **W** is defined) pressures.

5.1.1. Data Model

[47] The first step is to formulate a data model:

$$[\mathbf{s} \mid \mathbf{W}, \mathbf{P}] \text{ is } N(\mathbf{KW}, \Sigma_s). \tag{22}$$

K represents a mapping matrix, mapping scatterometer observations to the **W** grid, and Σ_s is the assumed form of the covariance matrix of the measurement errors.

5.1.2. Process Model

[48] The process model is the product of the distributions defined by

$$[\mathbf{W} \mid \mathbf{P}] \text{ is } N\left(\mathbf{BP}, \Sigma_{w|p}\right) \tag{23}$$

and

$$[\mathbf{P}] \text{ is } N\left(\boldsymbol{\mu}, \Sigma_p\right). \tag{24}$$

Using a discrete approximation to differentiation, the matrix **B** in equation (23) is a sparse matrix, whose rows were parameterized to yield local, discretized spatial derivatives of the pressure field. For example, consider the gridded u wind component at an arbitrary interior grid box with coordinates i, j. With discretized derivative and geostrophic approximations,

$$\begin{aligned} u(i,j) \approx\ & \beta_1(i,j)\,[p(i,j+1) - p(i,j-1)] \\ & + \beta_2(i,j)\,[p(i-1,j) - p(i+1,j)]. \end{aligned} \tag{25}$$

Similar expressions hold for v winds. Collecting such terms and accounting for boundary sites, we are led to the linear model in equation (23). Note that although the prior process model for winds was developed based on geostrophy, the model is not geostrophic. We certainly would not expect geostrophy to hold for near-surface winds due to turbulent processes, frictional effects, etc. While the prior conditional mean for winds as modeled in equation (23) is roughly geostrophic, deviations from that mean have variability quantified by $\Sigma_{W|p}$.

[49] The modeling relies on the construction of a plausible model (equation (24)) since the analysis was constructed for use with no pressure observations. *Royle et al.* [1998] used a stochastic model based on that derived by *Thiebaux* [1985]. Finally, prior distributions for parameters in **B** and in the covariance matrices Σ_s, $\Sigma_{W|p}$, and Σ_p were formulated. It was assumed that β_1 and β_2 were constant over space, although that assumption could be relaxed easily.

[50] Since model parameters were themselves modeled as random, Bayes's theorem allows the data to control their role in the posterior distribution. In particular, a prior for the elements of **B** can be derived from consideration of the values of the Coriolis parameter and geostrophy. For example, under geostrophy we expect $\beta_1(i, j) = 0$, but this parameter was maintained in the analysis to allow the model to respond to the data and potentially capture ageostrophic behavior.

[51] This model is capable of producing estimates of gridded pressures, even though no pressure data is used; that is, Bayesian analysis yields a posterior distribution $[\mathbf{P}|\mathbf{s}]$. *Royle et al.* [1998] computed estimated pressure fields and compared them to ECMWF-analyzed pressure fields. These results were quite good. Furthermore, to demonstrate the combining of diverse data sets, the approach was

adapted to actually use ECMWF-analyzed fields as "data" to enhance the results.

5.2. Space-Time Model for Tropical Ocean Winds

[52] *Wikle et al.* [2001] develop a space-time BHM for high-resolution, near-surface tropical ocean winds. Two data sets were used: (1) scatterometer data and (2) weather center analysis fields (specifically, National Centers for Environmental Prediction (NCEP) analyses). The former data are of very high resolution but are spatially incomplete over time, being constrained by the orbits of satellites bearing the scatterometer instruments. The latter data sets are complete in space and time but are of low resolution. *Wikle et al.* [2001] consider inference at a spatial resolution between those of the two data sets. Let $\{W\}_1^T$ denote the collection of horizontal wind vectors $\mathbf{W}^t$, defined on a high-resolution grid of the spatial domain of interest for times $t = 1, \ldots, T$. These wind fields are the primary process variables of interest.

5.2.1. Data Model

[53] Let $\mathbf{s}^t$ and $\mathbf{a}^t$ denote scatterometer and weather center analysis fields available at time t, respectively. Following the notation in section 5.2, let $\{s\}_1^T$ and $\{a\}_1^T$ be the collections of data available over the times of interest. *Wikle et al.* [2001] make the following assumptions: Conditional on the wind process and measurement error covariances, (1) the scatterometer and analysis fields are independent, (2) the data are independent across time, and (3) the distribution of data at time t depends only on the winds at time t. More formally, the assumptions imply that

$$\left[\{s\}_1^T, \{a\}_1^T | \{W\}_1^T, \Sigma_s, \Sigma_a\right] = \prod_{t=1}^{T} [\mathbf{s}^t | \mathbf{W}^t, \Sigma_s]\, [\mathbf{a}^t | \mathbf{W}^t, \Sigma_a]. \quad (26)$$

[54] The simplifications indicated in equation (26) suggest the value of developing data models conditionally. The key is that assumption 1 is only plausible conditionally on the true winds. We naturally expect the two data sets to have complex dependence structures since they are measurements of the same process. Similar considerations arise in the assumption that the data are conditionally independent across time. We would not expect the data to be unconditionally independent across time since dynamic relationships among the true winds over time would lead to dependence in the data. Since such assumptions are very useful in the combination of diverse data sets, we look more closely at the essential argument. Ignoring the measurement error covariances, first note that any joint distribution $[\{s\}_1^T, \{a\}_1^T | \{W\}_1^T]$ can be written as

$$\left[\{s\}_1^T, \{a\}_1^T | \{W\}_1^T\right] = \left[\{s\}_1^T | \{a\}_1^T, \{W\}_1^T\right] \left[\{a\}_1^T | \{W\}_1^T\right].$$

This is just a fact, not an assumption, from probability theory. The justification of the conditional independence assumption focuses on the role of $\{a\}_1^T$ in $[\{s\}_1^T | \{a\}_1^T, \{W\}_1^T]$, namely, Given $\{W\}_1^T$, what additional information is available regarding $\{s\}_1^T$ by also conditioning on $\{a\}_1^T$? The assumption is that the answer is none or at least little enough to allow the assumption as an approximation. At the time *Wikle et al.* [2001] developed their model, NCEP did not use scatterometer data in developing their analysis fields, suggesting that the conditional independence assumption is quite reasonable. Currently, NCEP does use scatterometer data, so it is interesting to consider whether or not the assumption could still be reasonable; that is, is it plausible that

$$\left[\{s\}_1^T | \{a\}_1^T, \{W\}_1^T\right] = \left[\{s\}_1^T | \{W\}_1^T\right],$$

even though $\{a\}_1^T$ includes $\{s\}_1^T$? Since the NCEP fields are highly complex functions of the data they are based upon, the assumption is arguably a reasonable approximation. As a final comment, note that Bayesian analysis is not disabled if such assumptions are not justifiable, although the computations may be more difficult in complex settings.

[55] The actual forms of the component models used in the right-hand side of equation (26) are

$$[\mathbf{s}^t \mid \mathbf{W}^t] \text{ is } N\left(\mathbf{K}_s^t \mathbf{W}^t, \Sigma_s^t\right) \quad (27)$$

and

$$[\mathbf{a}^t \mid \mathbf{W}^t] \text{ is } N(\mathbf{K}_a \mathbf{W}^t, \Sigma_a), \quad (28)$$

where the $\mathbf{K}_s^t$ map the scatterometer observations to the grid of interest (these matrices vary with time, depending upon the satellite orbits) and $\mathbf{K}_a$ maps the NCEP grid to the model grid.

5.2.2. Process Model

[56] *Wikle et al.* [2001] decompose the wind process as follows:

$$\mathbf{W}^t = \boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{\theta}_1^t + \boldsymbol{\Psi}\boldsymbol{\theta}_2^t, \quad (29)$$

where $\boldsymbol{\mu}$ is a purely spatial mean, $\boldsymbol{\Phi}$ is a collection of equatorial normal modes (discretized in space), derived from the linear shallow-fluid equations on the equatorial beta plane, and $\boldsymbol{\Psi}$ are a collection of wavelet basis functions. The vectors of coefficients $\boldsymbol{\theta}_1^t$ and $\boldsymbol{\theta}_2^t$ are endowed with first-order Markov vector autoregression priors; that is, they assume that

$$\boldsymbol{\theta}_1^t = \mathbf{H}_1 \boldsymbol{\theta}_1^{t-1} + \boldsymbol{\eta}_1^t; \quad (30a)$$

$$\boldsymbol{\theta}_2^t = \mathbf{H}_2 \boldsymbol{\theta}_2^{t-1} + \boldsymbol{\eta}_2^t, \quad (30b)$$

where the noise vectors $\boldsymbol{\eta}_1^t$ and $\boldsymbol{\eta}_2^t$ are assumed to all have means equal to zero vectors, are independent across time, and are mutually independent.

[57] Substantial physical reasoning is used to parameterize the propagation matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ and model the covariance matrices of the noise vectors. I make two key points here and refer the reader to *Wikle et al.* [2001] for further details. First, the linear wave theory, if applied deterministically, specifies the form of $\mathbf{H}_1$. However, this theory is not expected to apply exactly; indeed, only a few

selected normal modes are used. Rather, the theory is used to suggest a prior model for the elements of $\mathbf{H}_1$, but these values are trained by the observational data in computation of the posterior distribution. That is, much like the claim that the *Royle et al.* [1998] model is not geostrophic, this large-scale portion of the *Wikle et al.* [2001] model is not a linear shallow-fluid model but rather a stochastic model motivated by the physics. The second point is that physical reasoning is also used to model the small-scale contributions represented via the wavelets. Specifically, notions of turbulence theory were related to the multiresolution properties of wavelets in formulating the prior distribution.

6. Discussion and Summary

[58] I have argued that the Bayesian approach to modeling in the presence of uncertainty offers opportunities for analyses founded on the combination of various information sources, while tracking the uncertainties in each. The hierarchical Bayesian approach to modeling compiles a complete model for data, physical processes of interest, and model parameters from a collection of conditional stochastic submodels. This strategy can be particularly effective when dealing with complex problems. In one basic stage of a hierarchy the formulation of statistical data models conditional on unknown processes of interest enables analyses combining diverse data sets displaying complicated dependence structures. Separate formulations of stochastic models for the physical processes of interest can be tailored to the incorporation of physics. In these modeling steps a variety of parameters are typically introduced. They appear to increase our modeling overhead. However, Bayesian learning about these parameters is of scientific value.

[59] Reviews of three published examples of BHM analyses were presented. The theme in each was how physical and statistical reasoning are combined. In the climate change analysis of *Berliner et al.* [2000a], substantial use of physical modeling was used by enticing that information from the output of a global climate model. In the spatial wind model of *Royle et al.* [1998] the familiar geostrophic approximation was to construct a stochastic model that allowed for the impacts of ageostrophic effects present in observations. In the space-time tropical-wind model of *Wikle et al.* [2001], physics operative at large spatial scales, as reflected through linear shallow-fluid equations, and at small scales, as modeled via some basic turbulence theory, were used to develop a statistical model. That model also incorporates additional dynamics reflected in the NCEP model by treating its output as observational data.

[60] I focused the discussions on modeling, leaving out two key potential warts associated with Bayesian analysis. The first of these is the difficulty in computation. Though computations can be difficult, Bayesian computation is an area of vigorous research. Further, from a comparative viewpoint, one would not claim that non-Bayesian computations in geophysics are easy. The second issue is the specification of prior distributions for parameters. Such specifications are not easy. Also, they often make scientists uneasy in that they appear to introduce subjectivity into the modeling. In response, most Bayesians claim that all scientific models dealing with uncertainty have elements that are subjective, though they are often simply called "assumptions." The Bayesian viewpoint is argued to be the preferred framework for presenting scientific models and results; one's assumptions are quantified and made clear for all to see. Bayesians are also concerned with the sensitivity of results to the specification of priors. See *Berliner et al.* [2000a] for a discussion and example treatments of this issue.

[61] While I sought to give an indication of the large scope for Bayesian modeling, I have not exhausted that scope. First, BHMs that seem primarily statistical (i.e., no direct reliance on physical models) can be developed using qualitative physical reasoning [e.g., *Wikle et al.*, 1998; *Berliner et al.*, 2000b]. Second, BMH offers novel ways of dealing with interacting physical processes. For example, *Berliner et al.* [2003] present a BHM relying upon quasi-geostrophic modeling to treat air-sea interaction. As a related example, *Wikle et al.* [2003] consider BHMs for regional models with stochastic boundary conditions. Applications of BHMs to sub-grid-scale parameterization can be pursued. Indeed, the tradition of using model output for large-scale variables at some time point to infer small-scale variables, which in turn then support the prediction of large-scale variables in the future, is conditional thinking. The formalisms of Bayesian learning-from-data and uncertainty management would enhance approaches, leading to classes of stochastic parameterizations. Similar considerations apply to problems of up-scaling and down-scaling.

[62] I close with some speculation. As background, consider three points. First, we anticipate that observations will continue to improve and increase. Second, geophysical analysis and prediction of the Earth system involve allocations of resources ranging from data collection to modeling effort to numerical computation. Finally, real (as opposed to our uncertain guesses) physics are present in observational data. These points suggest that our task is to form models which are highly efficient in their use of that data. My suggestion is that BHMs offer a framework for that task. First, it is crucial to note that BHMs lead to posterior distributions that are more general and more faithful to the observations than the prior distributions used. This suggests that the anticipated more and better observations can be combined with simple physical models to produce efficient and effective analyses. I do not mean to suggest that the massive climate system models or their atmospheric components, as used in global numerical weather forecasting, can be replaced by simple BHMs, as were reviewed here. I do suggest that a rethinking of the prediction enterprise based on a spectrum of BHMs may prove useful.

[63] **Acknowledgments.** This work was supported by the NASA/Goddard Space Flight Center under award NAG5-8455 and by the U.S. EPA's Science to Achieve Results (STAR) program, assistance agreement R827257-01-0. I am grateful to R. Milliff and C. Wikle for useful comments on earlier versions of this material and to three referees for their very helpful reviews.

References

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

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

Berliner, L. M., Statistics, probability, and chaos, *Stat. Sci.*, *7*, 69–122, 1992.

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., R. A. Levine, and D. J. Shea, Bayesian climate change assessment, *J. Clim.*, *13*, 3805–3820, 2000a.

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, 2000b.

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.

Bernardo, J. M., and A. F. M. Smith, *Bayesian Theory*, John Wiley, Hoboken, N. J., 1994.

Ehrendorfer, M., Predicting the uncertainty of numerical weather forecasts: A review, *Meteorol. Z.* [*Berlin*], *6*, 147–183, 1997.

Epstein, E. S., Stochastic dynamic prediction, *Tellus*, *21*, 739–759, 1969.

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

Epstein, E. S., Long-range predictions: Limits of predictability and beyond, *Weather Forecast*, *3*, 69–75, 1988.

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

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, 486 pp., Chapman and Hall, New York, 1996.

Hasselmann, K., Stochastic climate models, part I. Theory, *Tellus*, *28*, 305–473, 1976.

Holton, J. R., *An Introduction to Dynamic Meteorology*, 3rd ed., 507 pp., Academic, San Diego, Calif., 1992.

Jones, P. D., T. J. Osborn, and K. R. Briffa, Estimating sampling errors in large-scale temperature averages, *J. Clim.*, *10*, 2548–2568, 1997.

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor, Surface air temperature and its changes over the past 150 years, *Rev. Geophys.*, *37*, 173–199, 1999.

Lorenc, A. C., Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *112*, 1177–1194, 1986.

Meehl, G. A., W. D. Collins, B. A. Boville, J. T. Kiehl, T. M. L. Wigley, and J. M. Arblaster, Response of the NCAR Climate System Model to increased CO_2 and the role of physical processes, *J. Clim.*, *13*, 1879–1898, 2000.

Meinhold, R. J., and N. Singpurwalla, Understanding the Kalman filter, *Am. Stat.*, *37*, 123–127, 1983.

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

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

Royle, J. A., L. M. Berliner, C. K. Wikle, and R. Milliff, A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador Sea, in *Case Studies in Bayesian Statistics*, edited by C. Gatsonis et al., pp. 367–382, Springer-Verlag, New York, 1998.

Salby, M. L., *Fundamentals of Atmospheric Physics*, Academic, San Diego, Calif., 1996.

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.

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.*, *131*, 1051–1062, 2003.

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