Combining Information Across Spatial Scales

Christopher K. WIKLE

Department of Statistics
University of Missouri
Columbia, MO 65203
(*wikle@stat.missouri.edu*)

L. Mark BERLINER

Department of Statistics
Ohio State University
Columbus, OH 43210-1247
(*mb@stat.ohio-state.edu*)

Spatial and spatiotemporal processes in the physical, environmental, and biological sciences often exhibit complicated and diverse patterns across different space–time scales. Both scientific understanding and observational data vary in form and content across scales. We develop and examine a Bayesian hierarchical framework by which the combination of such information sources can be accomplished. Our approach is targeted to settings in which various special spatial scales arise. These scales may be dictated by the data collection methods, availability of prior information, and/or goals of the analysis. The approach restricts to a few essential scales. Hence we avoid the challenging problem of constructing a model that can be used at all scales. This means that we can provide inferences only at the preselected special scales. However, problems involving special scales are sufficiently common to justify the trade-off between our comparatively simple modeling and analysis strategy with the formidable task of forming models valid at all scales. Specifically, our approach is based on a simple idea of conditioning the spatially continuous process on an areal average of the process at some resolution of interest. In addition, the data at prescribed resolutions are then conditioned on this areal-averaged true process. These conditioning arguments fit nicely into the hierarchical Bayesian framework. The methodology is demonstrated for the spatial prediction of an important quantity known as streamfunction based on wind information from satellite observations and weather center, computer model output.

KEY WORDS: Bayes; Change of support; Hierarchical; Poisson equation; Spatiotemporal; Streamfunction; Wind.

1. INTRODUCTION

In many spatial and spatiotemporal modeling problems, data concerning a process of interest are often of different spatial resolutions and alignments; that is, they provide information at varying spatial scales. In addition, one may be interested in the true process at some other spatial resolution, including resolutions not represented in the observations. These issues fall under the topic of "change of support" (COS) in spatial statistics and are longstanding, especially in geography (e.g., Cressie 1993, 1996). Various approaches have been considered for this problem, as reviewed by Gotway and Young (2002). Continuing interest in the problem has been fueled by technological advances in both data collection (e.g., remote sensing) and data management (e.g., geographic information systems).

Scientific understanding of spatiotemporal processes often varies with scale. For example, meteorologists have greater understanding of large-scale processes, such as fronts, than for microscale processes, such as raindrop formation. In addition, the choice of spatial scales for processes may arise for practical reasons. For example, it is increasingly common for scientists and engineers to rely on numerical models based on relevant mathematical models of the process. Such models are developed as approximations to underlying physical/biological theory. An ideal example is modern approaches to global numerical weather forecasting. Physical theories are applied to produce systems of partial differential equations for the evolution of the meteorological state process. These complicated nonlinear equations cannot be solved exactly, and thus modelers produce discretized approximations to the equations. For example, derivatives are often replaced by *finite differences*, leading to a discrete model describing the evolution of variables representing spatial averages of the weather variables.

The problem of interest here concerns the case where both the data and the process are at disparate spatial scales. In such cases, particularly when there is additional information such as the process models described earlier, the Bayesian view is natural. It has been demonstrated that hierarchical Bayesian models are well suited to the analysis of such complicated spatial and spatiotemporal problems (e.g., Wikle, Berliner, and Cressie 1998). There have been other recent examples of Bayesian approaches to the COS problem (e.g., Mugglin, Carlin, and Gelfand 2000; Gelfand, Zhu, and Carlin 2001; Wikle, Milliff, Nychka, and Berliner 2001). Our approach differs from these approaches in that we make use of the conditional models inherent in the Bayesian hierarchical approach. Specifically, our approach is based on the idea of conditioning a continuous spatial process on values (often, areal averages) of the process defined at some resolution for which inference is desired and/or prior scientific information is available. We then consider data observed at resolutions that are larger, the same, or smaller than the resolution desired for inference. These data are also conditioned on the process at the resolution of interest. Implementation of these conditioning arguments is relatively simple within the hierarchical Bayesian framework.

The hierarchical change of support methodology is described in Section 2. In Section 3 the methodology is illustrated with the prediction of atmospheric streamfunction given wind observations at different supports. Enhancements and extensions to the approach are presented in Section 4, followed by a conclusion in Section 5.

TECHNOMETRICS, FEBRUARY 2005, VOL. 47, NO. 1
DOI 10.1198/004017004000000572

2. HIERARCHICAL BAYESIAN APPROACH TO CHANGE OF SUPPORT

At the risk of some oversimplification, most Bayesian and Bayesian-like treatments of COS (e.g., Mugglin et al. 2000; Gelfand et al. 2001) are grounded in specification of a statistical model or prior distribution for the process of interest at the finest spatial resolution possible. Typically, this resolution corresponds to point support. Consider a spatial process $\{Y\} = \{Y(\mathbf{s}) : \mathbf{s} \in D\}$, where D is a bounded subset of $\mathcal{R}^d$. For a set S, let $|S| \equiv \int_S d\mathbf{s}$ denote the volume of S. (Typically $d = 2$, and volume is simply geometric area.) For much of this article, "Y with spatial support on S" is taken to mean $Y(\mathbf{s})$ averaged over $S \subset D$,

$$Y(S) \equiv \begin{cases} \dfrac{1}{|S|}\displaystyle\int_S Y(\mathbf{s})\, d\mathbf{s}, & |S| > 0 \\ \text{ave}\{Y(\mathbf{s}) : \mathbf{s} \in S\}, & |S| = 0, \end{cases} \tag{1}$$

where the discrete case accounts for point (e.g., weather station) data. Inference on $\{Y\}$ or Y supported on selected sets, based on observations, relies on specification of a spatial stochastic process model (or at least on formulation of first and second moments of the process). A common assumption is that $\{Y\}$ is a Gaussian random field with mean function $\mu(\mathbf{s})$ and covariance function $\tilde{\sigma}(\mathbf{r}, \mathbf{s})$. In principle, general approaches to the COS problem are available through application of probability theory. Specifically, a Gaussian-process assumption for Y implies a readily obtained multivariate normal joint distribution for $\{Y(S)\}$ for any selection of a finite collection of support sets.

We consider situations in which full specification of a complete model, or even of first and second moments, valid at all scales is deemed too difficult. We focus on predicting the spatial process supported on a finite collection of nonoverlapping subsets of D, all with similar areas. That is, we consider disjoint sets B_j, $j = 1, \ldots, n_b$, whose union is D such that $|B_j| > 0$ are similar for all j. Although we may observe data with supports on some of the B_j, we focus on the more difficult case where data are observed with support sets (all subsets of D) $A_i, i = 1, \ldots, n_a$, and C_k, $k = 1, \ldots, n_c$, such that $0 \le |A_i| < |B_j| < |C_k|$ for all i, j, k. That is, the C_k are at a larger scale than the B_j, which are in turn at a larger scale than the A_i. It is useful to think of $Y(B_j)$, $j = 1, \ldots, n_b$, as the process on a desired *prediction grid* and $Y(A_i)$, $i = 1, \ldots, n_a$, and $Y(C_k)$, $k = 1, \ldots, n_c$, as the *subgrid* and *supergrid* processes. We define the vector $\mathbf{Y}_A \equiv (Y(A_1), \ldots, Y(A_{n_a}))'$, and define $\mathbf{Y}_B$ and $\mathbf{Y}_C$ analogously.

In geostatistical contexts, our primary suggestion is to model in two steps: (1) conditionally model subgrid-scale *residuals* from $\mathbf{Y}_B$ and (2) separately model $\mathbf{Y}_B$. Such a decomposition is often used in the turbulence literature (e.g., Holton 1992, sec. 5.1) and numerical modeling of partial differential equations via finite differences (e.g., Holton 1992, chap. 13). More precisely, we construct the process $Y(\mathbf{s})$ conditional on $\mathbf{Y}_B$ based on the following assumption.

Assumption 1. Conditional on $\mathbf{Y}_B$, define a continuously indexed process $\{\gamma(\mathbf{s}), \mathbf{s} \in D\}$ such that for each B_j and all $\mathbf{s} \in B_j$,

$$Y(\mathbf{s}) = Y(B_j) + \gamma(\mathbf{s}), \tag{2}$$

where the γ process has mean 0 everywhere, the covariance function $\sigma(\mathbf{s}, \mathbf{r}) = \text{cov}[\gamma(\mathbf{s}), \gamma(\mathbf{r})]$ for all $\mathbf{s}, \mathbf{r} \in D$, and the distribution of the γ process does not depend on $\mathbf{Y}_B$.

This strategy does not obviate the need for a statistical model of a process varying continuously in space. However, because our modeling of the residual γ-process takes place conditionally on $\mathbf{Y}_B$, covariance structures are likely to be simpler than those used in complete joint modeling.

Note that when constructing a process $Y(\mathbf{s})$ following the recipe of Assumption 1, we have that for all $\mathbf{s} \in B_j$, $E(Y(\mathbf{s}) | \mathbf{Y}_B) = Y(B_j)$, implying that $E(Y(\mathbf{s})) = E(Y(B_j))$. That is, the overall mean is constant within each B_j. Next, note that $\text{cov}[Y(\mathbf{s}), Y(\mathbf{r}) | \mathbf{Y}_B] = \sigma(\mathbf{s}, \mathbf{r})$, and hence $\text{cov}[Y(\mathbf{s}), Y(\mathbf{r})] = \sigma(\mathbf{s}, \mathbf{r}) + \text{cov}[Y(B_j), Y(B_k)]$ for $\mathbf{s} \in B_j$ and $\mathbf{r} \in B_k$. An interesting question is whether given a complete model for $\{Y\}$ with mean function $\mu(\mathbf{s})$ and covariance function $\tilde{\sigma}(\mathbf{r}, \mathbf{s})$, does it admit a representation as described in Assumption 1? As indicated earlier, a positive answer requires that μ can be constant on the B_j. We also require that $\tilde{\sigma}(\mathbf{r}, \mathbf{s}) = \sigma(\mathbf{s}, \mathbf{r}) + \text{cov}[Y(B_j), Y(B_k)]$ for $\mathbf{s} \in B_j$ and $\mathbf{r} \in B_k$. This requirement is delicate; in the case of areal averaging, it may introduce singularities in σ.

Assumption 1 implies that we have lost some level of generality compared with traditional approaches beginning with a complete model for $\{Y\}$. The requirement that overall mean μ must be constant within each B_j relates to a fundamental issue in practical spatial statistics: the modeling trade-offs between *trend* and *spatial dependence*. Unlike the full-joint modeling approach, our approach essentially limits explanation of subgrid-scale variation to locally trend-free spatial dependence modeling. On the other hand, without strong beliefs about local trends (exactly our circumstance), spatial covariance modeling is generally believed to be capable of capturing spatial structure.

While enduring the loss of subgrid-scale trend modeling, we gain substantial generality, because we can relax the full joint normality assumption. That is, $\mathbf{Y}_B$ need not have a multivariate normal distribution. Such models would be very difficult to formulate based on a complete joint specification. Also, as we discuss later, the data models that incorporate the change of support can be developed without reference to the specific prior on $\mathbf{Y}_B$ and can be used for various spatial models as well as certain space–time models. Finally, our methodology is tailored to allow ready incorporation of numerical model output restricted to particular scales. Within the hierarchical Bayesian framework, such model outputs are viewed simply as data. This is particularly natural in the context of weather center analyses, which are statistics in the conventional definition as functions of observations.

2.1 Hierarchical Models Based on Areal Averaging

In complex problems, hierarchical Bayesian models can also appear to be complex. However, it is useful to view the basic approach as model building in three essential stages (e.g., Berliner 1996). The first step is to construct a statistical *data model*. This involves statistical modeling of the measurement process, whereby one considers the distribution of the observations conditioned on the true process of interest and model parameters. In the second stage, that process is then modeled via a *process model*, which itself may be composed of a several substages. Finally, parameters introduced in these two stages are endowed with a joint prior distribution in the *parameter model*.

The step of modeling the observations conditional on the process of interest is extremely powerful. It allows explicit treatment of measurement error separately from process variation, thereby leading to comparatively easy combination of different data sources. Furthermore, construction of the process model offers the opportunity for imputation of scientific understanding regarding the variables of interest and, as we demonstrate in this article, the treatment of various spatial scales.

2.1.1 Data Models. We let Z denote observational data. To indicate appropriate support, $Z(S)$ represents an observation of $Y(S)$. Although this notation appears to suggest that some averaging of a process $Z(\mathbf{s})$ is involved, this need not be the case. Let $\mathbf{Z}_A \equiv (Z(A_1), \ldots, Z(A_{n_a}))'$ be the vector of observations of the $\mathbf{Y}_A$ vector. Similarly let $\mathbf{Z}_C \equiv (Z(C_1), \ldots, Z(C_{n_c}))'$ be the vector of observations of the $\mathbf{Y}_C$ vector. Our task is to construct the data model,

$$[\mathbf{Z}_A, \mathbf{Z}_C | \mathbf{Y}_A, \mathbf{Y}_C, \mathbf{Y}_B, \boldsymbol{\theta}_m], \tag{3}$$

where $\boldsymbol{\theta}_m$ denotes a collection of model parameters and the brackets [ ] refer to a probability distribution.

Assumption 2. Conditional on $\mathbf{Y}_A$, $\mathbf{Y}_C$, $\mathbf{Y}_B$, and $\boldsymbol{\theta}_m$, (a) $\mathbf{Z}_A$ and $\mathbf{Z}_C$ are independent and have distributions that do not depend on $\mathbf{Y}_B$, and (b) the conditional distribution of $\mathbf{Z}_A$ does not depend on $\mathbf{Y}_C$ and the conditional distribution of $\mathbf{Z}_C$ does not depend on $\mathbf{Y}_A$. Formally, we assume that

$$[\mathbf{Z}_A, \mathbf{Z}_C | \mathbf{Y}_A, \mathbf{Y}_C, \mathbf{Y}_B, \boldsymbol{\theta}_m] = [\mathbf{Z}_A, \mathbf{Z}_C | \mathbf{Y}_A, \mathbf{Y}_C, \boldsymbol{\theta}_m] \tag{4}$$

$$= [\mathbf{Z}_A | \mathbf{Y}_A, \boldsymbol{\theta}_m][\mathbf{Z}_C | \mathbf{Y}_C, \boldsymbol{\theta}_m]. \tag{5}$$

Throughout this section, we also make the following assumption.

Assumption 3. All observations follow models of the form

$$Z(S) = Y(S) + \epsilon_S, \tag{6}$$

where ϵ_S is an associated measurement error.

In vector notation, the additive measurement error model (6) implies that

$$\mathbf{Z}_A = \mathbf{Y}_A + \boldsymbol{\epsilon}_A, \tag{7}$$

where $\boldsymbol{\epsilon}_A$ is an n_a-vector of measurement errors. Similarly, we have

$$\mathbf{Z}_C = \mathbf{Y}_C + \boldsymbol{\epsilon}_C, \tag{8}$$

where $\boldsymbol{\epsilon}_C$ is an n_c-vector of measurement errors. As mentioned earlier, we can also include observations at the B-support, but this is actually quite easy and is suppressed here for simplicity. Also, this notation does not include conditionally independent replicates (i.e., multiple observations corresponding to some A and/or C grid boxes), although adjustments in that case are simple. In some cases, replacing (7) with $\mathbf{Z}_A = \mathbf{H}_A\mathbf{Y}_A + \boldsymbol{\epsilon}_A$, where $\mathbf{H}_A$ a known design matrix, is mandated. For example, in obtaining satellite data, observations are often nonlinearly related to the underlying Y-process. A common procedure is to approximate that relationship with a linear one, quantified by $\mathbf{H}_A$. [Analogous modeling for (8) would also be performed.] To ease the notational burden, we do not incorporate these features here; the rest of the formulas in this article can be easily adapted to handle such issues.

Much of the COS literature focuses on manipulations of the Z variables to account for their different supports. In our hierarchical formulation, we view the models (7) and (8) as conditional (on Y) models, and focus on manipulation of the Y variables to account for changes of support. This differentiates the spatial modeling challenge from one involving the complexities of the dependence structures present in the observations treated marginally (i.e., unconditionally on Y) to one involving prior formulations for Y. In particular, it is often quite plausible that the $\boldsymbol{\epsilon}_A$ and $\boldsymbol{\epsilon}_C$ can be modeled as mutually independent, with each having covariance matrices that are diagonal.

To write models in a compact form, we use the following notational device: $\mathbf{X} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is read as "the random vector $\mathbf{X}$ has mean, $\boldsymbol{\mu}$, and covariance matrix, $\boldsymbol{\Sigma}$." To add a joint normality assumption, we write $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. With this notation, we combine (7) and (8) as

$$\begin{pmatrix} \mathbf{Z}_A \\ \mathbf{Z}_C \end{pmatrix} | \mathbf{Y}_A, \mathbf{Y}_C, \boldsymbol{\Sigma}_m \sim \left(\begin{pmatrix} \mathbf{Y}_A \\ \mathbf{Y}_C \end{pmatrix}, \boldsymbol{\Sigma}_m \right), \tag{9}$$

where $\boldsymbol{\Sigma}_m$ is the measurement error covariance, represented in partitioned form as

$$\boldsymbol{\Sigma}_m = \begin{pmatrix} \boldsymbol{\Sigma}_{m_a} & \boldsymbol{\Sigma}_{m_{ac}} \\ \boldsymbol{\Sigma}_{m_{ca}} & \boldsymbol{\Sigma}_{m_c} \end{pmatrix}. \tag{10}$$

As mentioned earlier, we may often argue that $\boldsymbol{\Sigma}_{m_{ac}} = \mathbf{0}$ and, further, that

$$\boldsymbol{\Sigma}_m = \begin{pmatrix} \sigma_a^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_c^2\mathbf{I} \end{pmatrix}. \tag{11}$$

Adjustments to allow these variances to depend on location are comparatively direct.

2.1.2 Completing the Hierarchy. A critical point in constructing the process model is interrelating the vectors $\mathbf{Y}_A$, $\mathbf{Y}_B$, and $\mathbf{Y}_C$, by providing

$$[\mathbf{Y}_A, \mathbf{Y}_C | \mathbf{Y}_B, \boldsymbol{\theta}_{ac}], \tag{12}$$

where $\boldsymbol{\theta}_{ac}$ is a collection of model parameters. Then modeling proceeds with development of the primary process model,

$$[\mathbf{Y}_B | \boldsymbol{\theta}_b], \tag{13}$$

and, finally, a prior on all parameters,

$$[\boldsymbol{\theta}_m, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b]. \tag{14}$$

Models (4), (12), (13), and (14) form a hierarchical model. Bayesian analysis produces the posterior distribution

$$[\mathbf{Y}_A, \mathbf{Y}_B, \mathbf{Y}_C, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b | \mathbf{Z}_A, \mathbf{Z}_C]. \tag{15}$$

For relatively simple problems, this posterior may be determined analytically. In large, complex problems, posterior inferences are typically approximated via Markov chain Monte Carlo (MCMC). Even then, analyses of the foregoing model may be extremely difficult to perform. If primary interest is on inference for $\mathbf{Y}_B$ (as is usually the case), then we may be able to reduce the size of the problem. We simply multiply (4) and (12) and integrate out $\mathbf{Y}_A$ and $\mathbf{Y}_C$, yielding

$$[\mathbf{Z}_A, \mathbf{Z}_C | \mathbf{Y}_B, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{ac}]. \tag{16}$$

Bayesian analysis based on (16), (13), and (14) then proceeds, producing the posterior

$$[\mathbf{Y}_B, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b | \mathbf{Z}_A, \mathbf{Z}_C]. \tag{17}$$

The second option, focusing on (16), (13), and (14), appears easier to manage than the first. In precisely the sort of problems that we consider here, the dimensionality of (17) is considerably lower than that of (15). Hence attacking (17) via MCMC appears to be the better choice. However, there is a delicate balance; integrations used to produce (16) may lead to complicated models. For example, such integrations are done conditionally on model parameters but often link those parameters in intricate ways, making the Bayesian calculations difficult. This issue arises quite generally in hierarchical analysis.

2.1.3 Moment Conditions Based on Areal Averaging. The primary goal of this section is the development of (12), or at least first and second moments, based on areal averaging. We do this by relying on Assumption 1. Consider integrals of both sides of expressions like (2) with respect to supports corresponding to the subgrid and supergrid processes. For example, in the subgrid case, if $A_i \subset B_j$, then for some B_j, we have that

$$\begin{aligned} Y(A_i) &= \frac{1}{|A_i|}\int_{A_i} Y(B_j)\, d\mathbf{s} + \frac{1}{|A_i|}\int_{A_i} \gamma(\mathbf{s})\, d\mathbf{s} \\ &= Y(B_j) + \frac{1}{|A_i|}\int_{A_i} \gamma(\mathbf{s})\, d\mathbf{s}. \end{aligned} \tag{18}$$

Alternatively, consider the supergrid process. For each k,

$$\begin{aligned} Y(C_k) &= \frac{1}{|C_k|}\sum_{j=1}^{n_b}\int_{C_k \cap B_j} Y(B_j)\, d\mathbf{s} + \frac{1}{|C_k|}\int_{C_k} \gamma(\mathbf{s})\, d\mathbf{s} \\ &= \mathbf{g}_C^{(k)\prime}\mathbf{Y}_B + \frac{1}{|C_k|}\int_{C_k} \gamma(\mathbf{s})\, d\mathbf{s}. \end{aligned} \tag{19}$$

The first term on the right side of (19) is a weighted combination of the relevant $Y(B_j)$ (i.e., those for which $C_k \cap B_j$ is not empty), with weights given by the areas $|B_j \cap C_k|$ divided by the area of C_k. Thus $\mathbf{g}_C^{(k)}$ is an n_b-vector of weights, many of which will be 0 (i.e., those for which $C_k \cap B_j$ is empty). Finally, a similarly derived expression holds for arbitrary A_i, in which case (18) can be written as

$$Y(A_i) = \mathbf{g}_A^{(i)\prime}\mathbf{Y}_B + \frac{1}{|A_i|}\int_{A_i} \gamma(\mathbf{s})\, d\mathbf{s}. \tag{20}$$

Nonzero entries of $\mathbf{g}_A^{(i)}$ are of the form $|B_j \cap A_i|/|A_i|$. Thus this approach is valid regardless of whether the data are "aligned" or "misaligned" relative to the prediction process support. Furthermore, note that the first terms on the right side of (19) and (20) essentially constitute the simple "area of overlap" approaches to COS as used in standard GIS packages. Clearly, if $\gamma(\mathbf{s})$ is not 0, then this simple approach is not sufficient as a stand-alone COS methodology.

From calculations like those in (19) and (20), we can obtain the following first- and second-moment information,

$$\begin{pmatrix} \mathbf{Y}_A \\ \mathbf{Y}_C \end{pmatrix} | \mathbf{Y}_B, \boldsymbol{\Sigma} \sim \left(\begin{pmatrix} \mathbf{G}_A \\ \mathbf{G}_C \end{pmatrix} \mathbf{Y}_B, \boldsymbol{\Sigma} \right), \tag{21}$$

where $\mathbf{G}_A$ is $n_a \times n_b$ and $\mathbf{G}_C$ is $n_c \times n_b$ and contain the averaging weights corresponding to the A_i and C_k. The elements of the $(n_a + n_c) \times (n_a + n_c)$ covariance matrix $\boldsymbol{\Sigma}$ are given by

$$\int_S \int_{S'} \sigma(\mathbf{r}, \mathbf{s})\, d\mathbf{r}\, d\mathbf{s} / (|S||S'|), \tag{22}$$

as S and S' vary over those sets represented in $\mathbf{Y}_A$ and $\mathbf{Y}_C$. Finally, (16) is then

$$\begin{pmatrix} \mathbf{Z}_A \\ \mathbf{Z}_C \end{pmatrix} | \mathbf{Y}_B, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_m \sim \left(\begin{pmatrix} \mathbf{G}_A \\ \mathbf{G}_C \end{pmatrix} \mathbf{Y}_B, \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_m \right). \tag{23}$$

We close this section with a technical remark. Modeling subgrid scale covariances σ of the residual process induces a *consistency* issue in modeling the process $\gamma(\mathbf{s})$. Formally, consistency requires that for each j and any sequence of (Lebesgue measurable) $A^{(m)}$ such that as $m \to \infty$, $A^{(m)} \uparrow B_j$ implies that $Y(A^{(m)}) \to Y(B_j)$, almost surely. Stated more simply, this involves a singularity condition on σ, because it requires that $\int_{B_j} \gamma(\mathbf{s})\, d\mathbf{s} = 0$ for all j. In computationally intense problems, we may not always demand consistency but rather may rely on approximation, namely ignore the consistency issue as long as all A_i are small relative to the B_j and no B_j is filled by the collection of A_i sampled. Besides technical issues of implementation, a second reason for not enforcing consistency is that definitions of all the A-, B-, and C-locations are typically themselves subject to error. Although a better approach is to model such mapping and truncation errors, an expedient adjustment is to relax consistency. Note that even without consistency, a process $Y(\mathbf{s})$ generated from (2) still exists, but such a process averaged over any B_j would no longer equal $Y(B_j)$.

3. EXAMPLE: HIERARCHICAL MODELING OF ATMOSPHERIC STREAMFUNCTION

To demonstrate the methodology, we consider the problem of determining a scalar measure of the near-surface atmospheric circulation over a limited domain, given remotely sensed winds and weather center assimilated ("analysis") wind estimates. Specifically, we are interested in the circulation over the Labrador Sea region. This domain is of interest in climatology because of its role in triggering ocean deep convection (Renfrew and Moore 1999). One contribution to the initiation of such convection is thought to be the rapid progression of relatively small-scale yet intense atmospheric cyclones across the region. That is, the triggering of ocean deep convection is likely related nonlinearly to the exchange of momentum from the atmosphere to the ocean that occurs when these storms move across the ocean. Thus small differences in the cyclone intensity may be important in understanding the convection "threshold" and the sensitivity to small changes in momentum transfer. Clearly, models for such sensitivities must account for the uncertainty in the wind observations (both COS issues and measurement error). To evaluate these issues formally, one must link an atmospheric model such as that presented here with a dynamic model of the ocean. This is beyond the scope of this article but was recently considered (without consideration of COS issues) by Berliner, Milliff, and Wikle (2003). We demonstrate how one can accommodate uncertainties in data resolution for such problems and also account for the uncertainties in the discretization of the relevant mathematical model and boundary process.

3.1 Data

Satellite-based wind estimates from the NASA scatterometer (NSCAT) were available from September 15, 1996 to June 29, 1997. These observations occurred in swaths on either side of the polar-orbiting satellite ground track. Different portions of the Labrador Sea were covered by successive orbits two times in a 24-hour period. Within the subregion of the NSCAT swath, observations were reported at 50-km spatial resolution. Figure 1(a) shows a subset of NSCAT data for December 26, 1996 over the area of interest. In addition to the satellite data, low-resolution gridded observations were available from the National Centers for Environmental Prediction (NCEP) global analysis. These "data" arose from the assimilation of relevant atmospheric observations (excluding the NSCAT observations) into a high-dimensional numerical model.

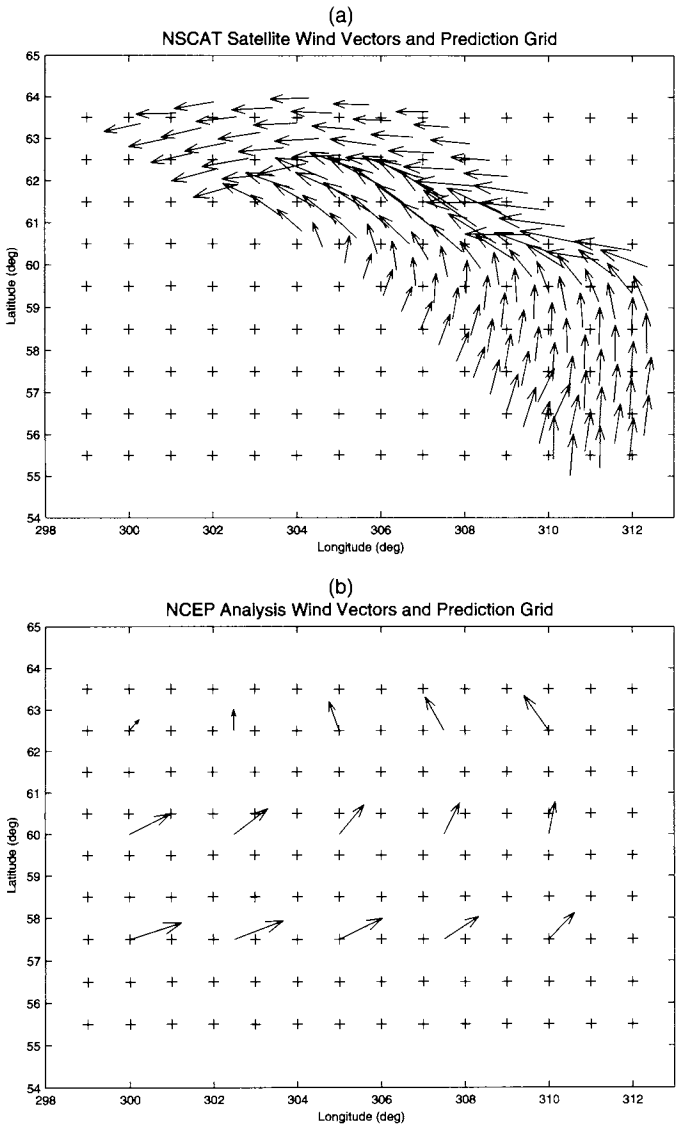

Figure 1. Wind Data. (a) The NSCAT wind vectors for December 26, 1996 over the Labrador Sea area. Note that the wind blows in the direction of the arrow and the wind speed is proportional to the length of the arrow. In addition, the "+" shows the center of the prediction grid box. (b) The NCEP wind analysis for the same period, as well as the prediction grid (+).

In our example, the assimilated observations are available at a 2.5-degree resolution (in latitude and longitude). Figure 1(b) shows the NCEP observations for December 26, 1996 over the Labrador Sea area of interest. In summary, the NSCAT observations have a resolution of about .5 degree, whereas the NCEP observations have a resolution of 2.5 degrees. The NCEP observations are on a regular grid, whereas the NSCAT observations are not. Our interest is the prediction of a scalar quantity (surface streamfunction) at an intermediate resolution of 1.0 degree (see Fig. 1), given these wind observations at differing supports and orientations relative to the prediction grid support. Streamfunction is a scalar quantity that describes the flow of the vector wind field. Streamfunction can be thought of as a family of streamlines that are everywhere tangent to the velocity (flow) field. Furthermore, a change in streamfunction corresponds to a change in the flow rate and in turn indicates the flow direction. Thus streamfunction is a useful summary measure of the circulation state of the atmosphere (or ocean).

3.2 Hierarchical Specification of a Boundary Process

Before describing the full hierarchical model, we describe the physical-based process model in some detail. This approach is interesting in its use of physical information to motivate the process model and in the specification of a hierarchical boundary process.

To get streamfunction from winds, one can use the deterministic Poisson equation (see, e.g., Holton 1992),

$$\nabla^2\psi = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \tag{24}$$

where ψ is the streamfunction [at some location (x, y) in two-dimensional Euclidean space], ∇^2 is the Laplacian operator (i.e., $\nabla^2 F = \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2}$, where F is some scalar spatial process continuous over some domain in $\mathcal{R}^2$), and the right side represents the vorticity of the flow defined by the horizontal derivatives of the north–south (v) and east–west (u) wind components. This is a boundary value problem, and thus its solution requires information about the process at the boundary. Of course, we typically do not know the boundary values exactly. Besides this uncertainty, there is uncertainty regarding the wind process, which we do not know everywhere in the domain of interest. Furthermore, with data, one must use numerical approaches to solving this equation, and such numerical discretizations introduce error. As described by Wikle, Berliner, and Milliff (2003), one can account for these uncertainties through a hierarchical formulation and stochastic boundary value specification. In summary, one can discretize (24) for a finite set of prediction locations (e.g., a regular grid) and get the following matrix representation:

$$\mathbf{L}\boldsymbol{\psi}_I = \mathbf{D}_x\mathbf{v} - \mathbf{D}_y\mathbf{u} + \mathbf{L}_{bc}\boldsymbol{\psi}_{bc} + \text{error}, \tag{25}$$

where $\boldsymbol{\psi}_I$ is the n_I-vector representation of the streamfunction process at the interior grid locations and $\boldsymbol{\psi}_{bc}$ is the n_{bc}-vector of the streamfunction process at the boundary grid locations. The $n_I \times n_I$ matrix $\mathbf{L}$ is a representation of the finite difference of the Laplacian operator applied to the interior locations, whereas $\mathbf{L}_{bc}$ is the corresponding $n_I \times n_{bc}$ matrix derived from the finite

difference of the Laplacian operator applied to the boundary locations. The $(n_I + n_{bc})$ vectors $\mathbf{u}$ and $\mathbf{v}$ contain the wind component processes at all grid locations, and $\mathbf{D}_x$ and $\mathbf{D}_y$ correspond to operators for centered first-difference calculations. Details of the formulation of $\mathbf{L}$, $\mathbf{L}_{bc}$, $\mathbf{D}_x$, and $\mathbf{D}_y$ have been provided by Wikle et al. (2003).

One can specify a hierarchical boundary value formulation of the streamfunction process according to

$$[\boldsymbol{\psi}_I, \boldsymbol{\psi}_{bc} | \mathbf{u}, \mathbf{v}] = [\boldsymbol{\psi}_I | \boldsymbol{\psi}_{bc}, \mathbf{u}, \mathbf{v}][\boldsymbol{\psi}_{bc} | \mathbf{u}, \mathbf{v}]. \tag{26}$$

In the case of (25), the first distribution on the right side of (26) would be

$$\boldsymbol{\psi}_I | \boldsymbol{\psi}_{bc}, \mathbf{u}, \mathbf{v} \sim \left(\mathbf{L}^{-1}[\mathbf{D}_x \mathbf{v} - \mathbf{D}_y \mathbf{u} + \mathbf{L}_{bc} \boldsymbol{\psi}_{bc}], \boldsymbol{\Sigma}_I\right), \tag{27}$$

where $\boldsymbol{\Sigma}_I$ is the error process covariance matrix. Depending on one's prior knowledge of the system, the distribution for the boundary streamfunction distribution may or may not depend on the wind process. For example, one might simply allow $\boldsymbol{\psi}_{bc} \sim (\boldsymbol{\mu}_{bc}, \boldsymbol{\Sigma}_{bc})$. In this case, the usual solution in numerical analysis, in which one specifies a fixed boundary, would correspond to the specification of $\boldsymbol{\mu}_{bc}$ and $\boldsymbol{\Sigma}_{bc}$ as a matrix of 0's (no variance). Very seldom do we actually know the boundary value exactly, so the nonzero specification of $\boldsymbol{\Sigma}_{bc}$ corresponds to our uncertainty in this knowledge.

3.3 Hierarchical Model

Here we present our hierarchical model for the streamfunction at intermediate support given winds at larger and smaller supports.

3.3.1 Data Model. Let the NSCAT (satellite) data be represented by the vectors $\mathbf{U}_A$ and $\mathbf{V}_A$. Similarly, let the NCEP (numerical model output) data vectors be denoted by $\mathbf{U}_C$ and $\mathbf{V}_C$. The true wind process on the prediction grid is specified by $\mathbf{u}_B$ and $\mathbf{v}_B$. The data model is then based on the hierarchical change of support formulation in (23) and is given by

$$\begin{pmatrix} \mathbf{U}_A \\ \mathbf{U}_C \end{pmatrix} | \mathbf{u}_B, \boldsymbol{\Sigma}_u, \boldsymbol{\Sigma}_m \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{G}_A \\ \mathbf{G}_C \end{pmatrix} \mathbf{u}_B, \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_m\right) \tag{28}$$

and

$$\begin{pmatrix} \mathbf{V}_A \\ \mathbf{V}_C \end{pmatrix} | \mathbf{v}_B, \boldsymbol{\Sigma}_v, \boldsymbol{\Sigma}_m \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{G}_A \\ \mathbf{G}_C \end{pmatrix} \mathbf{v}_B, \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_m\right). \tag{29}$$

The conditional normality assumption is justified based on other studies using wind components (e.g., Freilich 1997; Royle, Berliner, Wikle, and Milliff 1998; Wikle et al. 2001).

3.3.2 Process Model. As described previously, the process model is based on the hierarchical boundary specification. In addition, we specify a distribution for the wind components. As described by Wikle et al. (2003), it is reasonable to assume that these distributions are Gaussian,

$$\begin{aligned} \boldsymbol{\psi}_I | \boldsymbol{\psi}_{bc}, \mathbf{u}, \mathbf{v} &\sim \mathrm{N}\left(\mathbf{L}^{-1}[\mathbf{D}_x \mathbf{v} - \mathbf{D}_y \mathbf{u} + \mathbf{L}_{bc} \boldsymbol{\psi}_{bc}], \boldsymbol{\Sigma}_I\right), \\ \boldsymbol{\psi}_{bc} &\sim \mathrm{N}(\boldsymbol{\mu}_{bc}, \boldsymbol{\Sigma}_{bc}), \end{aligned} \tag{30}$$

and

$$\begin{pmatrix} \mathbf{u}_B \\ \mathbf{v}_B \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_u \mathbf{1} \\ \mu_v \mathbf{1} \end{pmatrix}, \boldsymbol{\Sigma}_{uv}\right). \tag{31}$$

Note that we are assuming that the streamfunction at both the interior and the boundary are at the same ("B") scale as the wind process $\mathbf{u}$ and $\mathbf{v}$.

3.3.3 Parameter Models. For a completely hierarchical analysis, we must specify distributions for $\boldsymbol{\Sigma}_u$, $\boldsymbol{\Sigma}_v$, $\boldsymbol{\Sigma}_m$, $\boldsymbol{\Sigma}_I$, $\boldsymbol{\Sigma}_{bc}$, $\boldsymbol{\mu}_{bc}$, μ_u, and μ_v, or their relevant parameterizations. To demonstrate of the COS methodology in the present application, it is sufficient to assume that these parameters are fixed and known (as described later). We discuss implementation strategies for fully hierarchical models in Section 4.

Thus the foregoing data and process models represent a multivariate spatial model (with $\mathbf{u}_B$, $\mathbf{v}_B$, $\boldsymbol{\psi}_I$, and $\boldsymbol{\psi}_{bc}$ as the various spatial processes) with a hierarchical boundary process and data at different spatial supports than the prediction process.

3.4 Posterior Distribution of Streamfunction

Although the data and process models are relatively complicated, in this case where the parameters are fixed and known, the distributions are Gaussian, and the boundary streamfunction process does not depend on the data, we can determine the marginal posterior of the interior streamfunction analytically.

For ease of notation, let $\mathbf{U} \equiv [\mathbf{U}_A' \ \mathbf{U}_C']'$, $\mathbf{V} \equiv [\mathbf{V}_A' \ \mathbf{V}_C']'$, $\mathbf{W} \equiv [\mathbf{U}' \ \mathbf{V}']'$, $\mathbf{G} \equiv [\mathbf{G}_A' \ \mathbf{G}_C']'$, $\mathbf{D} \equiv [-\mathbf{D}_y \ \mathbf{D}_x]$, $\boldsymbol{\mu}_{uv} \equiv [\mu_u \mathbf{1}' \ \mu_v \mathbf{1}']'$, and

$$\boldsymbol{\Sigma}_\epsilon \equiv \begin{pmatrix} \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_m & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_m \end{pmatrix}.$$

At this point we apply the hierarchical COS ideas to obtain the posterior distribution of the interior streamfunction given the data. It is easily shown (see the App.) that the marginal posterior of the interior streamfunction is given by

$$\boldsymbol{\psi}_I | \mathbf{U}, \mathbf{V} \sim \mathrm{N}(\boldsymbol{\mu}_{\psi|U,V}, \boldsymbol{\Sigma}_{\psi|U,V}), \tag{32}$$

with

$$\boldsymbol{\mu}_{\psi|U,V} = \mathbf{L}^{-1}\mathbf{D}\boldsymbol{\mu}_{u,v|U,V} + \mathbf{L}^{-1}\mathbf{L}_{bc}\boldsymbol{\mu}_{bc} \tag{33}$$

and

$$\boldsymbol{\Sigma}_{\psi|U,V} = \boldsymbol{\Sigma}_{\psi|U,V,\psi_{bc}} + \mathbf{L}^{-1}\mathbf{L}_{bc}\boldsymbol{\Sigma}_{bc}\mathbf{L}_{bc}'\mathbf{L}^{-1}, \tag{34}$$

where

$$\boldsymbol{\Sigma}_{\psi|U,V,\psi_{bc}} = \boldsymbol{\Sigma}_I + \mathbf{L}^{-1}\mathbf{D}\boldsymbol{\Sigma}_{u,v|U,V}\mathbf{D}'\mathbf{L}^{-1}, \tag{35}$$

$$\boldsymbol{\mu}_{u,v|U,V} = \boldsymbol{\Sigma}_{u,v|U,V}(\mathbf{G}'\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{W} + \boldsymbol{\Sigma}_{uv}^{-1}\boldsymbol{\mu}_{uv}), \tag{36}$$

and

$$\boldsymbol{\Sigma}_{u,v|U,V} = (\mathbf{G}'\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{G} + \boldsymbol{\Sigma}_{uv}^{-1})^{-1}. \tag{37}$$

3.4.1 Parameter Specification. The parameter specifications used for this example were based on preliminary data analysis and previous studies (e.g., Milliff, Niiler, Sybrandy, Nychka, and Large 2003; Wikle et al. 2003). First, we note that $n_A = 369$, $n_C = 15$, $n_I = 84$, and $n_{bc} = 42$. We let

$$\boldsymbol{\Sigma}_m = \begin{pmatrix} \sigma_A^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_C^2 \mathbf{I} \end{pmatrix},$$

with $\sigma_A^2 = 1.7$ m^{2}s^{-2} and $\sigma_C^2 = 2.7$ m^{2}s^{-2}. Furthermore, we assume that the residual covariance matrices for $\mathbf{u}$ and $\mathbf{v}$ are equal ($\boldsymbol{\Sigma}_u = \boldsymbol{\Sigma}_v$) and that the spatial covariance $\sigma(\mathbf{r}, \mathbf{s})$ is exponential and isotropic $[\sigma(\mathbf{r}, \mathbf{s}) = \sigma_\gamma^2 \exp(-\theta\|\mathbf{r} - \mathbf{s}\|)]$ with $\sigma_\gamma^2 = 1.5$ m^{2}s^{-2} and $\theta = 1/150$ km^{-1}. We let $\boldsymbol{\Sigma}_{uv} =$

$\mathbf{S}_{uv} \otimes \mathbf{R}_{uv}$, where $\mathbf{S}_{uv}$ is a 2×2 covariance matrix between the **u** and **v** wind components, with the covariance between **u** and **v** set at 12 m^2s^{-2} and the variance of **u** and **v** set at 213 m^2s^{-2} and 55 m^2s^{-2}. Furthermore, $\mathbf{R}_{uv}$ is an $(n_I + n_{bc}) \times (n_I + n_{bc})$ spatial correlation matrix assumed to be isotropic and exponential [e.g., $r(d) = \exp(-\theta_{uv}d)$], where d is the distance between two prediction grid locations and we select $\theta_{uv} = 1/300\,\text{km}^{-1}$. We also let $\mu_u = 0\,\text{ms}^{-1}$, $\mu_v = 3\,\text{ms}^{-1}$, and $\boldsymbol{\Sigma}_I = \sigma^2_\psi \mathbf{I}$, with $\sigma^2_\psi = 10^{10}\,\text{m}^4\text{s}^{-2}$. We specified the prior mean for the streamfunction boundary process to match the implied domain inflow and outflow based on visual inspection of the wind data plot (see Fig. 1) (see, e.g., Wikle et al. 2003). Finally, we set $\boldsymbol{\Sigma}_{bc} = \sigma^2_{bc}\mathbf{R}_{bc}$, where $\mathbf{R}_{bc}$ is based on an exponential covariance function with spatial dependence parameter $\theta_{bc} = 1/100\,\text{km}^{-1}$, reflecting a fairly near-range spatial dependence.

3.4.2 Computation Issues. All distances used for calculating covariances are great circle distances, to account for the distortion in longitudinal distance with latitude. Similarly, the derivative and Laplacian operators $\mathbf{D}_x$, $\mathbf{D}_y$, $\mathbf{L}$, and $\mathbf{L}_{bc}$ are adjusted to account for this effect.

For the purposes of calculating $\mathbf{G}_A$, $\mathbf{G}_C$, and $\boldsymbol{\Sigma}_u$ (and hence $\boldsymbol{\Sigma}_v$), we must specify the appropriate grid boxes at the "*A*," "*B*," and "*C*" scales. As mentioned previously, we assume that the "*B*" scale is 1 degree × 1 degree. Similarly, we assume that the satellite "footprint" for the "*A*" scale is .5 degree × .5 degree. We evaluate the sensitivity to this assumption by considering rectangular and circular "footprints" and find that the results are not sensitive to this choice. Finally, we assume that the "*C*"-scale data (NCEP) are at a 2.5-degree × 2.5-degree resolution. Again, the results are not sensitive to reasonable variations in the orientation assumption as long as the area is the same.

To obtain $\boldsymbol{\Sigma}_u$ (and hence $\boldsymbol{\Sigma}_v$), we evaluated the integrals in (22) given the covariance function specified earlier. Following Gelfand et al. (2001), we evaluated these integrals by Monte Carlo integration. Specifically, for each "*A*"-scale observation, we randomly selected 20 locations within the box, and for the "*C*"-scale data, we randomly selected 50 locations to perform the Monte Carlo evaluation. Such specifications provided sufficient Monte Carlo accuracy. For example, when we increased each by 10 locations, the difference between the estimated covariances was in the third and fourth decimal places.

To evaluate the $\mathbf{G}_A$ and $\mathbf{G}_C$ matrices, one must be able to calculate the area of overlap between the various "*A*"-, "*B*"-, and "*C*"-scale boxes as described in Section 2. Although such calculations can be done simply in most GIS software packages, we wrote our own general code to perform these calculations in cases where the grid boxes are four-sided polygons.

3.4.3 Results. Figures 2(a) and 2(b) shows the posterior mean and standard deviation of the streamfunction process. Note that the negative streamfunction indicates a counterclockwise circulation, and that the closer the contour lines, the stronger the flow. Thus the posterior mean streamfunction field clearly suggests the presence of a fairly intense low-pressure system (i.e., a "cyclone") centered around 305 degrees east longitude and 60 degrees north latitude. The posterior standard deviation indicates that the least certainty in this prediction is

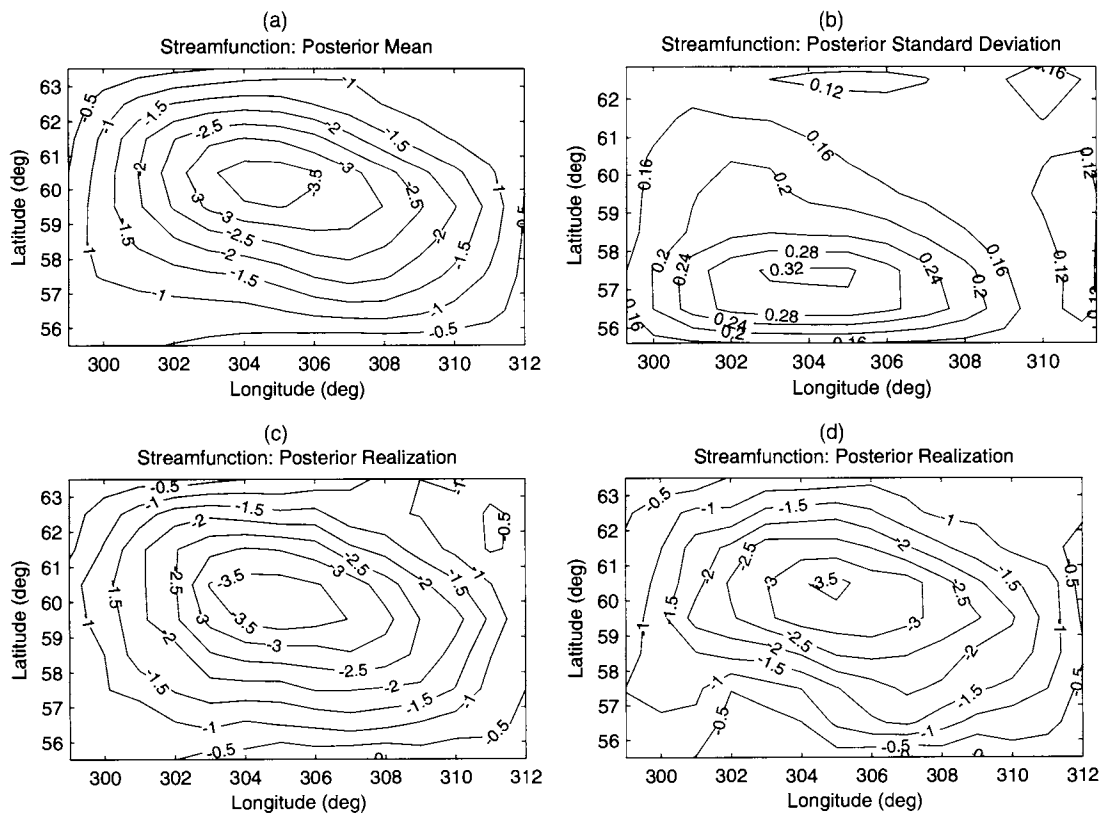

Figure 2. Streamfunction Posterior Mean (a), Standard Deviation (b), and Realizations [(c) and (d)] for December 26, 1996, Corresponding to the Wind Data Given in Figure 1. Contour values should be multiplied by 10^6 m^2s^{-1}.

in the southwest corner of the domain. Not surprisingly, this is the area with the least amount of data. For comparison, the corresponding posterior mean wind vectors are shown in Figure 3(a).

The realizations from the posterior shown in Figures 2(c) and 2(d) [and wind vector realizations in Figs. 3(b) and 3(c)] show substantial variability around the posterior mean.

To see the importance of the information at both large and small scales, we considered two additional analyses. Figure 4 shows the posterior mean and standard deviation streamfunction fields for the case where only the large-scale (NCEP) data are considered. Although the streamfunction field still shows the cyclone, the orientation of the cyclone is east–west rather than northwest–southeast, and the cyclone intensity is much weaker. Furthermore, the variability in this case is extreme (note the contour intervals are two orders of magnitude larger than in Fig. 2!). Alternatively, Figure 5 shows the posterior mean and standard deviation streamfunction fields for the case where only the small-scale (NSCAT) data are considered. In this case the orientation is more similar to the posterior mean of Figure 2, and although the intensity is weaker than that in Figure 2, it is much stronger than that in Figure 4. Furthermore, the standard deviations are about twice as large in this case as in the full-data case shown in Figure 2. Thus these results show that the scatterometer data are probably most important in determining the streamfunction, but information is definitely added by the large-scale NCEP data.

4. ENHANCEMENTS AND EXTENSIONS

The conditional COS approach described here can easily be implemented in the fully Bayesian context. Furthermore, the methodology can be extended to the spatiotemporal and multivariate settings. We describe these scenarios in the following sections.

4.1 Enhancements for Markov Chain Monte Carlo Implementation

As the example in the previous section demonstrates, with relatively small datasets, Gaussian distributions on the data models and priors, and known parameters, the hierarchical COS methodology is relatively easy to implement. However, when these assumptions are not appropriate (as is often the case), the methodology is still appropriate, although one must use MCMC methods to evaluate the posterior.

Bayesian computations, including MCMC approaches, are difficult when $n_a \times n_c$ is very large. In particular, these computations may require inversion of the $(n_a + n_c) \times (n_a + n_c)$ covariance matrix described in (23). This matrix naturally partitions into components for the $\mathbf{Z}_A$ and $\mathbf{Z}_C$. We might consider computing the overall inverse of $\boldsymbol{\Sigma}$ directly, based on familiar formulas for the inverse of a partitioned matrix. Instead, we expand on this notion further, essentially performing Bayesian updating sequentially.

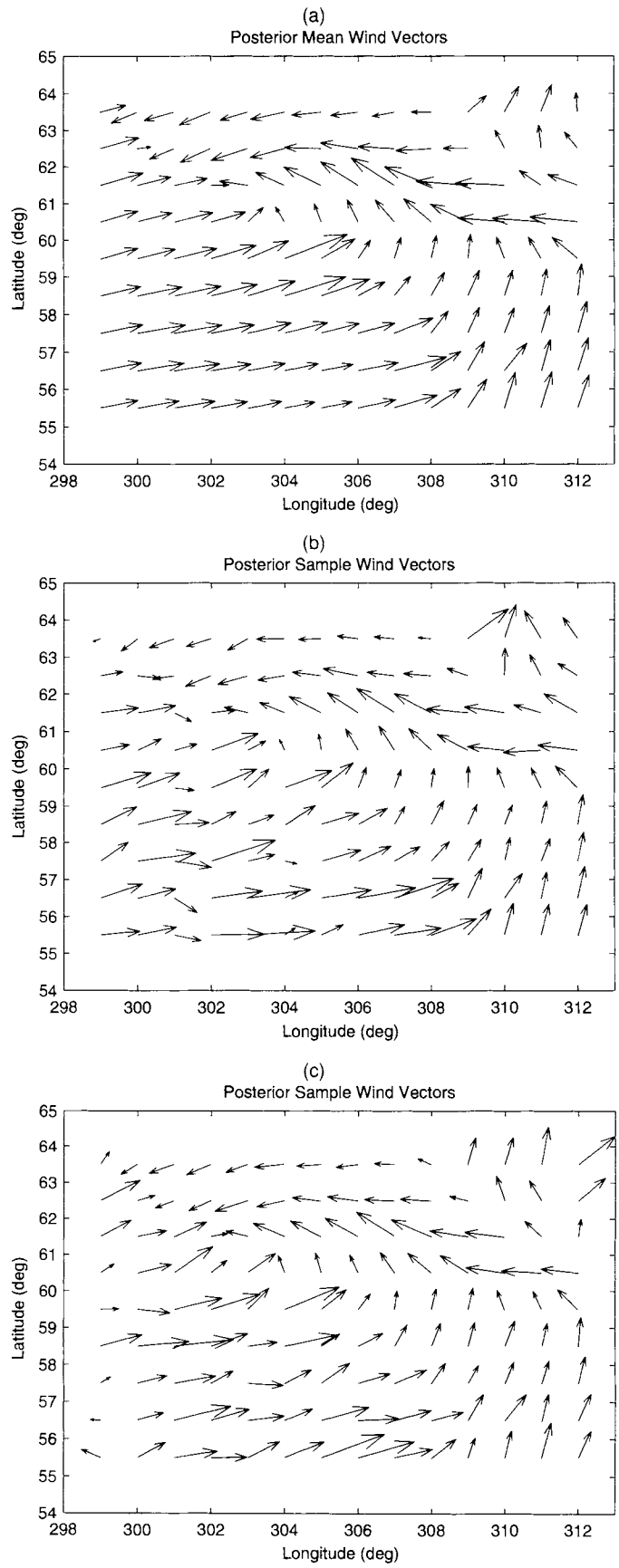

Figure 3. Wind Vector Posterior Mean (a) and Realizations [(b) and (c)] for December 26, 1996, Corresponding to the Wind Observations Given in Figure 1.

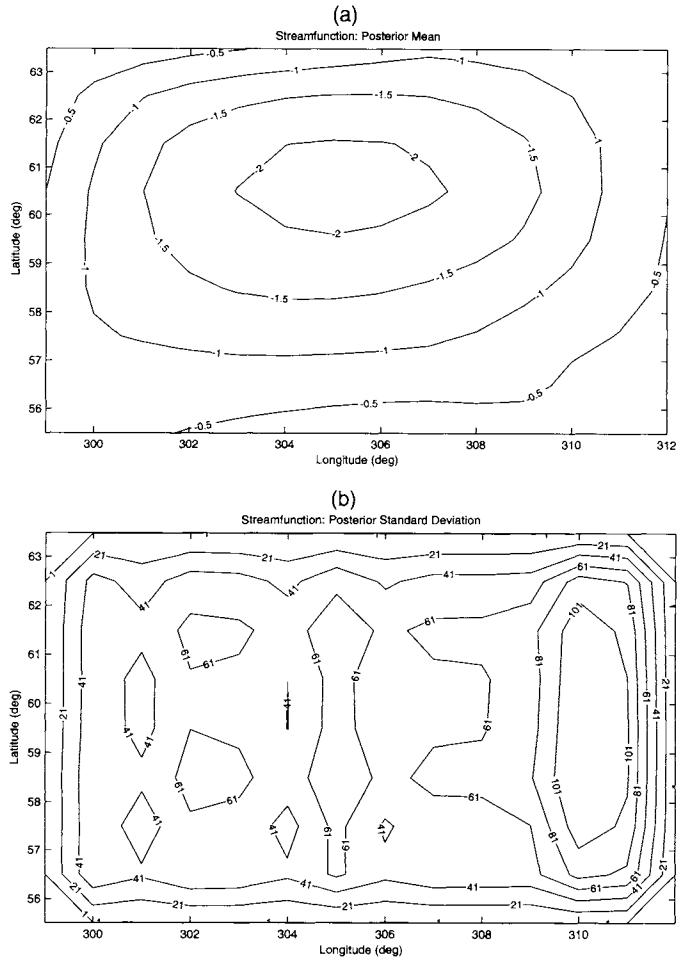

Figure 4. Streamfunction Posterior Mean (a) and Standard Deviation (b) for December 26, 1996, Using Only the Large-Scale (NCEP) Winds Shown in Figure 1(b). Contour values should be multiplied by 10^6 $m^2 s^{-1}$.

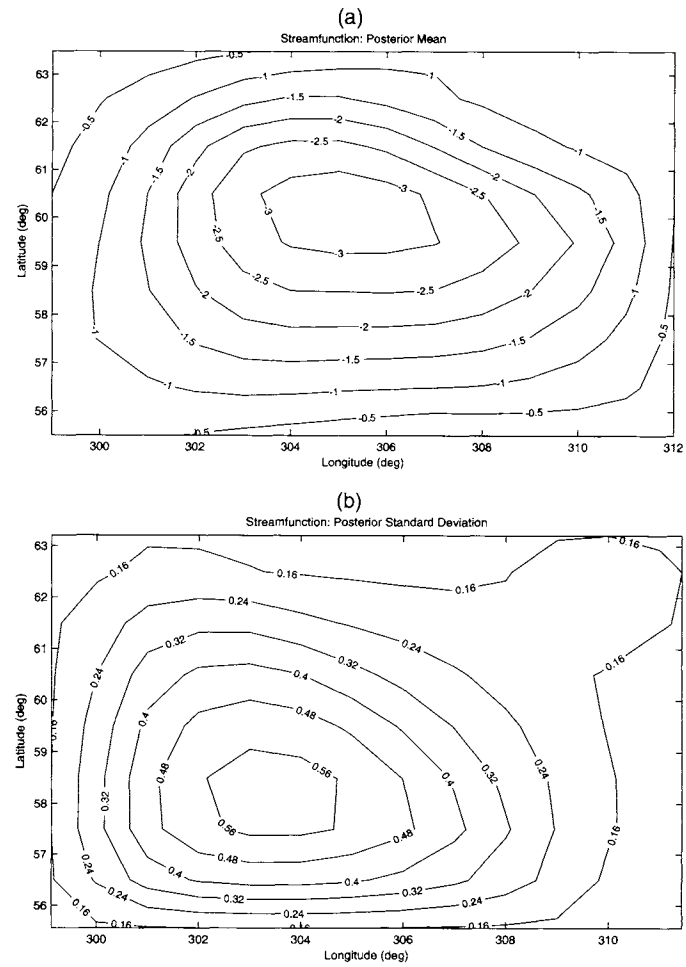

Figure 5. Streamfunction Posterior Mean (a) and Standard Deviation (b) for December 26, 1996, Using Only the Small-Scale (NSCAT) Winds Shown in Figure 1(a). Contour values should be multiplied by 10^6 $m^2 s^{-1}$.

4.1.1 Combining Datasets Sequentially. Temporarily suppressing conditioning on parameters, we may always write

$$[\mathbf{Z}_A, \mathbf{Z}_C|\mathbf{Y}_B] = [\mathbf{Z}_C|\mathbf{Z}_A, \mathbf{Y}_B][\mathbf{Z}_A|\mathbf{Y}_B]. \tag{38}$$

Adding a multivariate normality assumption and rewriting (23) in partitioned form, we have

$$\begin{pmatrix}\mathbf{Z}_A\\ \mathbf{Z}_C\end{pmatrix}|\mathbf{Y}_B, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_m \sim \mathrm{N}\left(\begin{pmatrix}\mathbf{G}_A\\ \mathbf{G}_C\end{pmatrix}\mathbf{Y}_B, \begin{pmatrix}\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_{m_a} & \boldsymbol{\Sigma}_{AC}\\ \boldsymbol{\Sigma}_{CA} & \boldsymbol{\Sigma}_C + \boldsymbol{\Sigma}_{m_c}\end{pmatrix}\right). \tag{39}$$

Of course, we have

$$\mathbf{Z}_A|\mathbf{Y}_B, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_m \sim \mathrm{N}\big(\mathbf{G}_A\mathbf{Y}_B, \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_{m_a}\big). \tag{40}$$

Next, applying the normal conditioning result, it follows that

$$\begin{aligned}\mathbf{Z}_C|\mathbf{Z}_A, \mathbf{Y}_B &\sim \mathrm{N}\big(\mathbf{G}_C\mathbf{Y}_B + \boldsymbol{\Sigma}_{CA}\big(\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_{m_a}\big)^{-1}(\mathbf{Z}_A - \mathbf{G}_A\mathbf{Y}_B),\\ &\qquad \boldsymbol{\Sigma}_C + \boldsymbol{\Sigma}_{m_c} - \boldsymbol{\Sigma}_{CA}\big(\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_{m_a}\big)^{-1}\boldsymbol{\Sigma}_{AC}\big).\end{aligned} \tag{41}$$

The main suggestion is to begin by performing MCMC using the A-data only. After a burn-in period, one may then augment the algorithm by then incorporating the C-data. Of course, the order of these data applications may be reversed. This approach offers reduced complexity of the MCMC for the burn-in phase; that is, it is a nice initialization for the full sampler. Perhaps more importantly, it also offers an opportunity to assess the value of the information added by the C-data.

This discussion is strongly related to the *Kalman filter*, although we have adjusted for COS. That is, rather than sequentially updating over time, we update across spatial scales sequentially. A similar, although not identical, view was used by Huang and Cressie (2001). Their work involves a complete specification of a joint model, whereas our "filter" is for observations conditional on $\mathbf{Y}_B$. As such, these formulas can be used to simulate synthetic observations readily, which in turn may be used in model validation studies as well as in the design of spatial sampling networks.

4.1.2 Combining Datasets via Importance Sampling. The Gaussian assumption in (39) may not always be appropriate. A more general approach is needed. In this context, the following suggestion for combining datasets may be of broad practical value, and hence we develop the basic formulas in general.

Assumption 4. First, assume that the parameters $\boldsymbol{\theta}_m$ appearing in the data model can be partitioned into two subvectors,

$\boldsymbol{\theta}_{m_a}$ and $\boldsymbol{\theta}_{m_c}$, where elements of $\boldsymbol{\theta}_{m_a}$ appear in the conditional model for $\mathbf{Z}_A$, but not in the model for $\mathbf{Z}_C$ (and vice versa).

We write (38) as

$$[\mathbf{Z}_A, \mathbf{Z}_C | \mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{m_c}, \boldsymbol{\theta}_{ac}] = [\mathbf{Z}_C | \mathbf{Z}_A, \mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{m_c}, \boldsymbol{\theta}_{ac}][\mathbf{Z}_A | \mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{ac}]. \quad (42)$$

We also assume that the prior on parameters is of the form

$$[\boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{m_c}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b] = [\boldsymbol{\theta}_{m_c}][\boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b]. \quad (43)$$

It follows that the complete joint posterior distribution is given by

$$[\mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{m_c}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b | \mathbf{Z}_A, \mathbf{Z}_C] = \frac{1}{h(\mathbf{Z}_A, \mathbf{Z}_C)} \quad (44)$$

$$\times [\mathbf{Z}_C | \mathbf{Z}_A, \mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{m_c}, \boldsymbol{\theta}_{ac}][\boldsymbol{\theta}_{m_c}] \quad (45)$$

$$\times [\mathbf{Z}_A | \mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{ac}][\mathbf{Y}_B | \boldsymbol{\theta}_b][\boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b], \quad (46)$$

where $h(\mathbf{Z}_A, \mathbf{Z}_C)$ is the normalizing constant (i.e., the integral of the numerator over $\mathbf{Y}_B$ and all parameters). Manipulation of (46) and some algebra implies that

$$[\mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{m_c}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b | \mathbf{Z}_A, \mathbf{Z}_C] = w(\mathbf{Z}_A, \mathbf{Z}_C) \quad (47)$$

$$\times [\mathbf{Z}_C | \mathbf{Z}_A, \mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{m_c}, \boldsymbol{\theta}_{ac}][\boldsymbol{\theta}_{m_c}] \quad (48)$$

$$\times [\mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b | \mathbf{Z}_A] \quad (49)$$

where (49) is the posterior on the indicated quantities conditional on $\mathbf{Z}_A$ only and $w(\mathbf{Z}_A, \mathbf{Z}_C)$ is the ratio of the corresponding marginals.

Standard results in importance sampling provide appropriate adjustment for these quantities, leading to inferences conditional on all data, based on MCMC samples from the sampler based on $\mathbf{Z}_A$. Specifically, let $(\mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b)^M$ denote the Mth realization from the MCMC conditioned only on $\mathbf{Z}_A$. Also, let $\boldsymbol{\theta}_{m_c}^M$ denote an independently generated realization from the prior $[\boldsymbol{\theta}_{m_c}]$. Assuming ergodicity of the sampler, we may asymptotically estimate expectations of integrable functions $f(\mathbf{Y}_B, \boldsymbol{\theta}_A, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b, \boldsymbol{\theta}_{m_c})$ conditional on $\mathbf{Z}_A$ *and* $\mathbf{Z}_C$ as

$$\sum_M f\left((\mathbf{Y}_B, \boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b)^M, \boldsymbol{\theta}_{m_c}^M\right) \times [\mathbf{Z}_C | \mathbf{Z}_A, \mathbf{Y}_B, (\boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b)^M, \boldsymbol{\theta}_{m_c}^M] \times \left(\sum_M [\mathbf{Z}_C | \mathbf{Z}_A, \mathbf{Y}_B, (\boldsymbol{\theta}_{m_a}, \boldsymbol{\theta}_{ac}, \boldsymbol{\theta}_b)^M, \boldsymbol{\theta}_{m_c}^M]\right)^{-1}. \quad (50)$$

Note that this result does not rely on computation of the weight w.

Just as the Gaussian case presented in Section 4.1.1 is analogous to well-known results in Kalman filtering, the importance sampling approach suggested here for the non-Gaussian case is analogous to the sequential Monte Carlo approach for time series (e.g., Doucet, de Freitas, and Gordon 2001). Again, the main differences are that updating is done not over time, but rather across spatial scales, and conditional on the process at the "B" scale.

4.2 Extensions of the Methodology

The hierarchical model summarized as the data model (3), process model composed of (12) and (13), and parameter model (14) is quite general. Thus the associated COS methodology that focuses on the data model, as presented here, also is quite general. Here we describe some generalizations.

4.2.1 Spatiotemporal Process. As described previously, one of the primary strengths of the hierarchical approach to COS outlined earlier, is that the support transformations occur in the data model, leaving the process to be modeled at another stage. Thus the methodology is directly applicable in the case of spatiotemporal data and a spatiotemporal process. That is, the data model might assume that, conditioned on the true process at time t, the observations are independent in time; for example,

$$[\mathbf{Z}_A(1), \ldots, \mathbf{Z}_A(T), \mathbf{Z}_C(1), \ldots, \mathbf{Z}_C(T) | \mathbf{Y}_B(1), \ldots, \mathbf{Y}_B(T)] = \prod_{t=1}^{T} [\mathbf{Z}_A(t), \mathbf{Z}_C(t) | \mathbf{Y}_B(t)],$$

with a subsequent spatiotemporal process model for $[\mathbf{Y}_B(1), \ldots, \mathbf{Y}_B(T)]$. An example of this was given by Wikle et al. (2001), although the COS implementation in the data model was substantially simpler than the moment condition approach outlined in Section 2.3.

4.2.2 Spatiotemporal COS. The hierarchical COS methodology can accommodate data of differing temporal as well as spatial supports. This requires straightforward modification to the notation presented previously. First, let τ represent some period of time over which the process is to be defined. Then we let

$$Y(B, \tau) = \frac{1}{|B|} \frac{1}{|\tau|} \int_\tau \int_B Y(\mathbf{s}, t)\, d\mathbf{s}\, d\mathbf{t}, \qquad |B| > 0, |\tau| > 0,$$

where $t \in \mathcal{T}$, the time domain of interest. Then a slight modification of Assumption 1 would be as follows.

Assumption 1b. Conditional on $Y(B, \tau)$, define a continuously indexed spatiotemporal process $\{\gamma(\mathbf{s}, t), \mathbf{s} \in D, t \in \mathcal{T}\}$ such that for each B_j and τ_i and all $\mathbf{s} \in B_j$, and $t \in \tau_i$,

$$Y(\mathbf{s}, t) = Y(B_j, \tau_i) + \gamma(\mathbf{s}, t), \quad (51)$$

where the γ process has mean 0 everywhere; covariance function $\sigma(\mathbf{s}, \mathbf{r}, t, t') = \text{cov}[\gamma(\mathbf{s}, t), \gamma(\mathbf{r}, t')]$ for all $\mathbf{s}, \mathbf{r} \in D$ and $t, t' \in \mathcal{T}$; and the distribution of the γ process does not depend on the Y-process at the support B and τ.

In principle, the development proceeds as for the purely spatial case. In practice, the spatiotemporal case can be more difficult, due to the problem of specifying realistic spatiotemporal covariance structures for the γ process and the dimensionality of the associated G and covariance matrices.

4.2.3 Multivariate Data/Processes. There is no requirement that all of $\mathbf{Y}_A$, $\mathbf{Y}_B$, and $\mathbf{Y}_C$ represent the same variables, defined from a basic Y. For example, $\mathbf{Y}_A$ and $\mathbf{Y}_B$ may represent one process, and $\mathbf{Y}_C$ may represent a completely different variable, albeit one for which we could still form (12). An illustration related to our wind example would be if we obtained sea surface height data (e.g., from an orbiting altimeter), which can be related physically to surface winds. (For a similar example with relatively simple COS methodology, see Royle et al. 1998.)

4.2.4 $\mathbf{Y}_B$ *as "Parameter."* In our development, $\mathbf{Y}_B$, although a vector of random parameters, is treated as a "process." Obviously, these parameters could be interpreted in the more traditional sense of parameters. For example, it might represent rate of disease or precipitation intensity distributed spatially.

4.2.5 Nonadditive Errors. We can relax Assumption 3 of additive errors. In that case, the Z distributions can still be written conditionally on the respective Y processes.

5. CONCLUSION

Due to technological advances, physical, biological, and environmental scientists and engineers are collecting many new datasets of various resolutions in space and time. These data do not always match up with the scales at which inference is desired. Furthermore, prior information regarding the processes of interest may vary in availability and quality across scales. To address such issues, we have proposed a conditional change of support approach that is well suited to the hierarchical Bayesian framework. The key elements of our approach are that we assume that the true process of interest in continuous space can be reasonably modeled conditional on values (typically, areal averages) of the process defined at some support for which inference is desired and/or prior information is available. We then assume that we have data at lesser, equal, or greater supports relative to this process scale (which we call "subgrid," "grid," and "supergrid" scales). Such data are modeled conditionally on the process defined at the supports corresponding to those of the data. These conditioning assumptions fit naturally into the hierarchical Bayesian framework. We have discussed alternative approaches for implementing this conditional COS methodology in the fully Bayesian context, suggesting sequential approaches for updating across spatial scales. Furthermore, we described how this methodology can be extended to the spatiotemporal setting.

The notion of considering data from different spatial scales in a hierarchical context is very powerful. As the modeling of complicated spatial processes is increasingly considered from a hierarchical perspective to account realistically for spatial variability, the use of hierarchical COS approaches will become increasingly important.

ACKNOWLEDGMENTS

Wikle's research was supported by National Science Foundation (NSF) grants DMS-01-39903 and ATM-02-22057. Berliner's research was supported by NSF grants DMS-01-39903. The authors thank two referees, whose comments led to improvements in this article.

APPENDIX: DERIVATION OF STREAMFUNCTION POSTERIOR DISTRIBUTION

Analogous to the example in Section 3, assume that we are interested primarily in a spatial process $\boldsymbol{\psi}_I$ with an associated boundary process $\boldsymbol{\psi}_{bc}$. In addition, assume that there is another related spatial process $\mathbf{w}_B$ with support B and associated observations $\mathbf{W}$ at arbitrary support. Consider the following hierarchical model:

$$\mathbf{W}|\mathbf{w}_B \sim \mathrm{N}(\mathbf{G}\mathbf{w}_B, \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_m),$$
$$\boldsymbol{\psi}_I|\mathbf{w}_B, \boldsymbol{\psi}_{bc} \sim \mathrm{N}(\mathbf{H}\mathbf{w}_B + \mathbf{K}\boldsymbol{\psi}_{bc}, \boldsymbol{\Sigma}_I),$$
$$\boldsymbol{\psi}_{bc} \sim \mathrm{N}(\boldsymbol{\mu}_{bc}, \boldsymbol{\Sigma}_{bc}),$$
$$\mathbf{w}_B \sim \mathrm{N}(\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w).$$

Thus we are assuming that $\mathbf{G}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_m$, $\mathbf{H}$, $\mathbf{K}$, $\boldsymbol{\Sigma}_I$, $\boldsymbol{\Sigma}_{bc}$, $\boldsymbol{\mu}_{bc}$, $\boldsymbol{\mu}_w$, and $\boldsymbol{\Sigma}_w$ are known or estimated in some other fashion (e.g., empirical Bayes).

Direct application of Bayes's rule gives the posterior distribution

$$[\boldsymbol{\psi}_I, \boldsymbol{\psi}_{bc}, \mathbf{w}_B|\mathbf{W}] \propto [\mathbf{W}|\mathbf{w}_B][\boldsymbol{\psi}_I|\boldsymbol{\psi}_{bc}, \mathbf{w}_B][\boldsymbol{\psi}_{bc}][\mathbf{w}_B].$$

To evaluate this posterior, we first note that the constant of proportionality is given by $\int [\mathbf{W}|\mathbf{w}_B][\mathbf{w}_B]\, d\mathbf{w}_B$, which is special for this model and follows from the fact that $\boldsymbol{\psi}_{bc}$ is independent of the $\mathbf{w}_B$ process. Thus we can rewrite the posterior as the equality

$$[\boldsymbol{\psi}_I, \boldsymbol{\psi}_{bc}, \mathbf{w}_B|\mathbf{W}] = [\mathbf{w}_B|\mathbf{W}][\boldsymbol{\psi}_I|\boldsymbol{\psi}_{bc}, \mathbf{w}_B][\boldsymbol{\psi}_{bc}].$$

Using the multivariate Gaussian assumptions in the hierarchical model, we obtain

$$\mathbf{w}_B|\mathbf{W} \sim \mathrm{N}\left(\boldsymbol{\mu}_{w_B|W}, \boldsymbol{\Sigma}_{w_B|W}\right),$$

where

$$\boldsymbol{\mu}_{w_B|W} = \boldsymbol{\Sigma}_{w_B|W}\left(\mathbf{G}'(\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_m)^{-1}\mathbf{W} + \boldsymbol{\Sigma}_w^{-1}\boldsymbol{\mu}_w\right)$$

and

$$\boldsymbol{\Sigma}_{w_B|W} = \left(\mathbf{G}'(\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_m)^{-1}\mathbf{G} + \boldsymbol{\Sigma}_w^{-1}\right)^{-1}.$$

Furthermore, exploiting conditioning and Gaussian assumptions gives

$$\boldsymbol{\psi}_I|\boldsymbol{\psi}_{bc}, \mathbf{W} \sim \mathrm{N}\left(\boldsymbol{\mu}_{\psi|\psi_{bc},W}, \boldsymbol{\Sigma}_{\psi|\psi_{bc},W}\right),$$

where

$$\boldsymbol{\mu}_{\psi|\psi_{bc},W} = \mathbf{H}\boldsymbol{\mu}_{w_B|W} + \mathbf{K}\boldsymbol{\psi}_{bc}$$

and

$$\boldsymbol{\Sigma}_{\psi|\psi_{bc},W} = \boldsymbol{\Sigma}_I + \mathbf{H}\boldsymbol{\Sigma}_{w_B|W}\mathbf{H}'.$$

Finally, integrating out the boundary process gives the desired posterior distribution for the interior spatial process,

$$\boldsymbol{\psi}_I|\mathbf{W} \sim \mathrm{N}\left(\boldsymbol{\mu}_{\psi|W}, \boldsymbol{\Sigma}_{\psi|W}\right),$$

where

$$\boldsymbol{\mu}_{\psi|W} = \mathbf{H}\boldsymbol{\mu}_{w_B|W} + \mathbf{K}\boldsymbol{\mu}_{bc}$$

and

$$\boldsymbol{\Sigma}_{\psi|W} = \boldsymbol{\Sigma}_{\psi|\psi_{bc},W} + \mathbf{K}\boldsymbol{\Sigma}_{bc}\mathbf{K}'.$$

[Received April 2002. Revised April 2004.]

REFERENCES

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

Berliner, L. M., Milliff, R. F., and Wikle, C. K. (2003), "Bayesian Hierarchical Modeling of Air–Sea Interaction," *Journal of Geophysical Research*, 108, 3104.

Cressie, N. A. C. (1993), *Statistics for Spatial Data* (rev. ed.), New York: Wiley.

——— (1996), "Change of Support and the Modifiable Areal Unit Problem," *Geographical Systems*, 3, 159–180.

Doucet, A., de Freitas, N., and Gordon, N. (eds.) (2001), *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag.

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

Gelfand, A. E., Zhu, L., and Carlin, B. P. (2001), "On the Change of Support Problem for Spatio-Temporal Data," *Biostatistics*, 2, 31–45.

Gotway, C. A., and Young, L. J. (2002), "Combining Incompatible Spatial Data," *Journal of the American Statistical Association*, 97, 632–648.

Huang, H.-C., and Cressie, N. (2001), "Multiscale Graphical Modeling in Space: Applications to Command and Control," in *Spatial Statistics: Methodological Aspects and Applications*, ed. M. Moore, New York: Springer-Verlag, pp. 83–113.

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

Milliff, R. F., Niiler, P. P., Sybrandy, A. E., Nychka, D., and Large, W. G. (2003), "Mesoscale Correlation Length Scales From NSCAT and Minimet Surface Wind Retrievals in the Labrador Sea," *Journal of Atmospheric and Oceanic Technology*, 20, 513–533.

Mugglin, A. S., Carlin, B. P., and Gelfand, A. E. (2000), "Fully Model-Based Approaches for Spatially Misaligned Data," *Journal of the American Statistical Association*, 95, 877–887.

Renfrew, I. A., and Moore, G. W. K. (1999), "An Extreme Cold-Air Outbreak Over the Labrador Sea: Roll Vortices and Air–Sea Interaction," *Monthly Weather Review*, 127, 2379–2394.

Royle, J. A., Berliner, L. M., Wikle, C. K., and Milliff, R. (1998), "A Hierarchical Spatial Model for Constructing Wind Fields From Scatterometer Data in the Labrador Sea," in *Case Studies in Bayesian Statistics*, eds. C. Gatsonis, J. S. Hodges, R. E. Kass, and N. D. Singpurwalla, New York: Springer-Verlag, pp. 367–382.

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

Wikle, C. K., Berliner, L. M., and Milliff, R. F. (2003), "Hierarchical Bayesian Approach to Boundary Value Problems With Stochastic Boundary Conditions," *Monthly Weather Review*, 131, 1051–1062.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001), "Spatiotemporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds," *Journal of the American Statistical Association*, 96, 382–397.