International Statistical Review (2003), 71, 2, 181–199, Printed in The Netherlands

Hierarchical Models in Environmental Science

Christopher K. Wikle

Department of Statistics, University of Missouri, 222 Math Science Building, Columbia, MO 65203, USA. E-mail: wikle@stat.missouri.edu

Summary

Environmental systems are complicated. They include very intricate spatio-temporal processes, interacting on a wide variety of scales. There is increasingly vast amounts of data for such processes from geographical information systems, remote sensing platforms, monitoring networks, and computer models. In addition, often there is a great variety of scientific knowledge available for such systems, from partial differential equations based on first principles to panel surveys. It is argued that it is not generally adequate to consider such processes from a joint perspective. Instead, the processes often must be considered as a coherently linked system of conditional models. This paper provides a brief overview of hierarchical approaches applied to environmental processes. The key elements of such models can be considered in three general stages, the data stage, process stage, and parameter stage. In each stage, complicated dependence structure is mitigated by conditioning. For example, the data stage can incorporate measurement errors as well as multiple datasets with varying supports. The process and parameter stages can allow spatial and spatio-temporal processes as well as the direct inclusion of scientific knowledge. The paper concludes with a discussion of some outstanding problems in hierarchical modelling of environmental systems, including the need for new collaboration approaches.

Key words: Bayesian; CAR; Change of support; Climate; Co-kriging; Conditional; Convolution; Dynamic; Markov random field; MCMC; Ozone; PDE; Spatial; Spatio-temporal; Spectral.

1 Introduction

Global warming, El Niño, the "ozone hole", genetically modified crops, inadvertent introduction of harmful species, the decline of amphibians, emerging diseases—these are but a few of the "hot button" issues in the environmental/ecological sciences. How do we begin to make sense of these issues? Can we afford to consider them individually, or must we consider them collectively? How do we combine the "known" scientific first principles with uncertain empirically-based parameterizations? Finally, how can we do all of these things given the massive amounts of data available from remote sensing platforms? Although no single methodology can hope to provide the answer to all of these questions at once, it can be argued that a hierarchical framework holds the most promise.

Environmental processes are complicated. What does the word "complicated" mean to a statistician? It is hard to imagine an environmental process that is not ultimately some manifestation of an underlying spatio-temporal process. For example, spatial processes such as the distribution of a radioactive contaminant in a given area, can be considered as manifestations of underlying spatio-temporal processes integrated over time. Similarly, temporal trends in global temperature can result from a very complicated multivariate spatio-temporal process (i.e., the earth/ocean/atmosphere system) integrated over space and the other relevant variables. These spatio-temporal processes often encompass a very large range of spatial and temporal scales of variability, and include nonlinear interactions across domains, variables, and systems. Furthermore, a key aspect of environmental

science (by definition) is that there are multiple "factors", biological and physical, that can affect organisms or communities. Thus, environmental processes arise from interactions of various processes. These interactions are often occurring at various scales in space and time. Although it is often convenient to simplify such systems either by ignoring the multivariate interaction, or by assuming spatial/temporal stationarity, linearity, and Gaussianity, it is increasingly the case that the scientific questions of interest are becoming sufficiently complex that one can no longer justify such assumptions.

The later years of the 20th century saw the exponential increase in the amount of data related to environmental processes. Most of this increase was/is due to rapid technological development, specifically in terms of remote sensing systems, and computational technology required to pre-process and store the vast quantities of available information. In addition to remote sensing, there have been numerous projects to provide vast *in situ* monitoring networks (e.g., the ocean buoy network in the tropical Pacific; ecological monitoring systems such as the Long Term Ecological Research (LTER) network, e.g., Kaiser, 2001). Correspondingly, along with computational technology and the increased monitoring efforts, scientists in some disciplines have been able to develop ever more sophisticated deterministic computer models of environmental systems which have a tremendous amount of potentially useful output (e.g., General Circulation Models (GCMs) of the atmosphere and ocean, coupled to models of biota to study potential climate change, Blackmon *et al.*, 2001). Thus, environmental statisticians have a sometimes bewildering array of data to consider when modeling environmental systems. In addition to the sheer volume of data, these data are often of differing spatial and temporal support, orientation, and alignment, relative to the process of interest.

Folded into the issues surrounding process and data complexity is the consideration of the relevant scientific information that is available. That is, there is often a great deal of knowledge that has been developed concerning environmental processes, ranging from physical laws (such as the dynamic and thermodynamic partial differential equations governing the atmosphere and ocean), "established" empirical relationships, as well as purely subjective considerations such as "expert opinion" developed from panel surveys. Of course, depending on the problem of interest, one may have various levels of confidence in such understanding. The question is then how can this information be utilized, or should it be? Very seldom do we see "designed" experiments in environmental studies as they are typically "observational" in nature. In that case, to evaluate scientific hypotheses, one must build the relevant information into the model, and evaluate the associated parameters or processes after consideration of the data. In the case of prediction, one may very well use all the information available, whether it be subjective or not, in an attempt to obtain the best prediction possible.

Faced with very complex processes, multiple sources of data from various platforms, and various degrees of scientific knowledge, the scientist is often overwhelmed. Modeling, whether it be deterministic or stochastic, can approach such complex problems from a "total process" or "connected subprocess" perspective. Or, in the language of statistics, from either a joint or conditional viewpoint. Although it may be intuitive to consider processes from a joint perspective, it can present serious challenges to statistical modeling. It is very difficult to specify joint multivariate spatio-temporal covariance structures for complicated environmental processes. It may be much easier to factor such joint distributions into a series of conditional models. For example, in forest landscape modeling, it is common to view the relevant spatio-temporal processes from a conditional frame of reference (e.g., tree structure is conditional on climate, land use, management, fire, etc.). In this case, it is then natural to link conditional models together in a hierarchical framework, one that is probabilistically valid. On the other hand, it is common in the atmospheric sciences to think of the atmosphere as an enormous multivariate spatial state process that evolves over time (nonlinearly). In this case, scientists may not be used to thinking conditionally, yet many of their formulations can still be viewed that way (e.g., geostrophic wind conditioned on the gradient of pressure). In either case, it is often

possible to simplify modeling specifications, account for uncertainties, and use first principles in a series of conditional models, coherently linked together by simple probability rules, i.e., hierarchical models.

Thus, it is not surprising that practitioners are increasingly finding that the way to proceed given these circumstances is through the use of hierarchical models. Such models can be considered from either a classical or Bayesian perspective. However, as the level of complexity increases (or "subjective" prior information is included) the Bayesian paradigm is usually necessary. Indeed, the increased use of such approaches has coincided with the revolution in Bayesian computation exemplified by the adoption and further development of Markov chain Monte Carlo (MCMC) simulation approaches. However, there is still very much of a role for data-based or "empirical Bayesian" analysis, particularly in the early stages of an investigation when there may be little knowledge about the parameter distributions (e.g., Carlin & Louis, 2000; Smith, 1999).

The remainder of this paper will consider hierarchical models in the environmental sciences. Clearly, the exposition will be somewhat biased by my own experiences and beliefs as to what is important. The discussion is not intended to be a comprehensive review of the entire literature on hierarchical models in environmental science. Rather, it is intended to provide an introduction and glimpse of how hierarchical models have been and can be used to facilitate modeling of environmental processes. The basic principles of hierarchical modeling are well-established and the cited references can provide additional perspective as well as increased technical detail. Section 2 will describe the hierarchical approach from a simple schematic perspective. This will be followed by a more detailed description of data models in the hierarchical framework, including a discussion of the change of support problem. Section 3 will emphasize the process modeling stages in hierarchical models. As is common in the literature, the discussion will be somewhat artificially separated into spatial and spatio-temporal processes. Finally, Section 4 will give a discussion of some of the outstanding challenges for the future of hierarchical modeling in the environmental sciences.

2 Basic Hierarchical Approach

The formal ideas of hierarchical modeling arise from simple probability rules. Consequently, some early modeling efforts in the mid 20th century that are clearly hierarchical in nature, are not referred to as such (see the Bibliographic note in Gelman *et al.* (1995, pp. 154–155) for a nice summary). Although the concept is not inherently Bayesian, over time most of the literature has been developed in that context, and the best pedagogical descriptions are most often found in the Bayesian literature (e.g., Gelman *et al.*, 1995). Our purpose here is not to reproduce such general descriptions, but rather to describe hierarchical models in the context of a framework for environmental modeling. These ideas follow those in Berliner (1996) and Wikle, Berliner & Cressie (1998) and Wikle *et al.* (2001).

Hierarchical modeling is, in essence, based on the simple fact from probability that the joint distribution of a collection of random variables can be decomposed into a series of conditional models. That is, if X, Y, Z are random variables, then we can write the joint distribution in terms of a factorizations such as $[X, Y, Z] = [Z|Y, X][Y|X][X]$. [Note, throughout the paper we make use of the bracket notation for probability distribution in which $[Y]$ refers to the distribution of Y and $[Y|X]$ refers to the conditional distribution of Y given X (Gelfand & Smith, 1990)]. This simple formula is the crux of hierarchical thinking. For example, for a spatio-temporal process, the joint distribution describes the stochastic behavior of the process at all spatial locations and all times, including all possible interactions. This can be *extremely* difficult (if not impossible) to specify for many environmental processes. However, it is often much easier to specify the distribution of the relevant conditional models (e.g., conditioning the process at the present time given the past). In this case, the product of a series of relatively simple conditional models leads to a joint distribution that can be quite involved.

For complicated processes in the presence of data, the idea is to approach the problem by breaking it into three primary stages (e.g., Berliner 1996):

Stage 1. Data Model: [*data* | *process, parameters*]

Stage 2. Process Model: [*process* | *parameters*]

Stage 3. Parameter Model: [*parameters*].

The first stage is concerned with the observational process or "data model", which specifies the distribution of the data *given* the process of interest and parameters that describe the data model. The second stage then describes the process, conditional on other parameters. Finally, the last stage accounts for the uncertainty in the parameters. In applications, each of these stages may have multiple sub-stages. For example, if the process is spatio-temporal, it might be modeled as a product of several physically-motivated conditional distributions suggested by a state-space formulation. Similar decompositions are possible in the parameter stage.

Ultimately, we are interested in the distribution of the process and parameters updated by the data. We obtain this so-called "posterior" distribution via Bayes' rule:

$$[\textit{process, parameters} \mid \textit{data}] \propto [\textit{data} \mid \textit{process, parameters}][\textit{process} \mid \textit{parameters}][\textit{parameters}].$$

This formula serves as the basis for Bayesian hierarchical analysis. However, there are several critical points to consider. These will be addressed in more detail in the following subsections.

2.1 *Data Models*

Datasets commonly considered for environmental processes are complicated and usually exhibit substantial spatial, temporal, or spatio-temporal dependence. The major advantage of modeling the conditional distribution of the data given the true process is that substantial simplifications in model form are possible. For example, let D_a be data observed for some process Y, and let θ_a be parameters. The data model is written, $[D_a|Y, \theta_a]$. Usually, this conditional distribution is much simpler than the unconditional distribution of $[D_a]$ as most of the complicated structure comes from the process Y. Often, this model simply represents measurement error. Note that in this general framework the measurement error need not be additive. Furthermore, as will be shown, this framework can also accommodate data that is at a different support than the process.

This framework also provides a natural way to combine data sets. For example, assume that D_a and D_c represent data from two different sources. Again, let Y be the process of interest and θ_a, θ_c be parameters. In this case, the data model is often written

$$[D_a, D_c|Y, \theta_a, \theta_c] = [D_a|Y, \theta_a][D_c|Y, \theta_c]. \tag{1}$$

In words, conditioned on the true process, the data are assumed to be independent. Of course, this does not suggest that the two datasets are unconditionally independent, they surely are not. Rather, the majority of the dependence among the datasets is due to the process, Y. This assumption of independence is exactly that, an assumption. Although often very reasonable, it must be assessed critically for each problem.

The conditional partitioning of the datasets in (1) is often similarly applied to multivariate models. That is, say our processes of interest are denoted Y_a and Y_c, with associated observations D_a and D_c. One might write

$$[D_a, D_c|Y_a, Y_c, \theta_a, \theta_c] = [D_a|Y_a, \theta_a][D_c|Y_c, \theta_c]. \tag{2}$$

Again, this represents the assumption that given the true processes of interest, the datasets are independent. Such an assumption must be evaluated and is not required in hierarchical analysis, but it is often very reasonable and can lead to dramatic simplifications in the computations.

2.2 *Process Models*

It is usually the case that developing the process distribution is the most critical step in constructing the hierarchical model. This distribution is often further factored hierarchically into a series of submodels. For example, assume the process of interest is composed of two subprocesses, Y_a and Y_c. Perhaps Y_a represents hospital admission rates for a geographical region and Y_c might represent atmospheric particulate concentrations over the same region. Furthermore, define parameters $\theta_Y = \{\theta_{Y_a}, \theta_{Y_c}\}$ that describe these two processes. One might consider the decomposition,

$$[Y_a, Y_c|\theta_Y] = [Y_a|Y_c, \theta_Y][Y_c|\theta_Y]. \tag{3}$$

This is just a fact of probability theory and can always be written. However, it may be the case that one can assume the parameters are conditionally independent in which case the right hand side of (3) can be written as $[Y_a|Y_c, \theta_{Y_a}][Y_c|\theta_{Y_c}]$. The challenge is the specification of these component distributions. Indeed, most of the effort in the development of hierarchical models is related to constructing these distributions. It is often the case, however, that there is very good scientific insight that can suggest appropriate conditioning order and possible models for the component distributions. For example, in a model with ozone and temperature, it is natural to condition ozone on temperature due to the chemistry of ozone production (Royle & Berliner, 1999). Similarly, Y_a might represent the process of interest at time t and Y_c the same process at the previous time, $t - 1$. Natural deterministic models for process evolution could suggest the form of such models (e.g., Wikle *et al.*, 2001).

2.3 *Parameter Models*

The parameter distributions may require significant modeling effort. As is the case with the data and process models, the distribution of parameters is often partitioned into a series of distributions. For example, consider the data model (2) and process model (3). One must specify the parameter distribution $[\theta_a, \theta_c, \theta_{Y_a}, \theta_{Y_c}]$. Often, one can make reasonable independence assumptions regarding this distribution, e.g., $[\theta_a, \theta_c, \theta_{Y_a}, \theta_{Y_c}] = [\theta_a][\theta_c][\theta_{Y_a}][\theta_{Y_c}]$. Of course, this assumption must be justified. There are usually appropriate submodels for parameters as well, leading to other levels of the model hierarchy. In many cases, for complicated processes, there is substantial scientific insight that can go into developing the parameter models (e.g., Wikle *et al.*, 2001). In other cases, one does not know much about the parameter distribution, suggesting "vague priors" or data-based estimates be used. That is, it is often useful to think empirically at first and perform exploratory data analysis in order to develop understanding about the process. The emphasis in this case is on model building.

The development of parameter distributions has often been the focus of objections due to its implied subjectiveness. Of course, the formulation of the data and process models are quite subjective as well, but those choices have not generated as much concern, probably because such subjectiveness is just as much a part of classical model building as it is the Bayesian approach. One must simply recognize that a strength of the hierarchical (Bayesian) approach is the quantification of such subjective judgment. Hierarchical models provide a coherent probabilistic framework in which to incorporate explicitly in the model the uncertainty related to judgment, scientific reasoning, subjective decisions, and experience.

3 Data Models and Change of Support

Section 2.1 outlined the general framework for data distributions, in which one specifies the distribution of the data given the process. This approach also provides a coherent framework in which to include data at different support than the process.

The consideration of problems where the data have a different level of support (usually spatial) than the desired inference is long-standing in many different disciplines and known by many different

names (e.g., Cressie, 1996; see Gotway & Young (2002) for a comprehensive review). Increasing attention has been focused on this issue as technological advances have made available many new datasets at varying spatial scales and orientations, primarily from Geographical Information Systems (GIS), remotely sensed data, and massive numerical models. Several recent studies have shown that the Bayesian paradigm provides a natural framework in which to incorporate such datasets (e.g., Gelfand *et al.*, 2001; Mugglin *et al.*, 2000). We refer to this as the change of support (COS) problem and discuss a basic hierarchical approach below.

3.1 Hierarchical Change of Support

Wikle & Berliner (2002) formulate the COS problem specifically within the hierarchical framework. Although this approach can be applied in general to spatio-temporal processes and COS in space and time, the following discussion illustrates the idea in terms of a spatial process.

Wikle & Berliner (2002) consider a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ where $\mathbf{s}$ is a spatial location in domain $D \subset \mathcal{R}^2$, and $\text{var}(Y(\mathbf{s})) < \infty$. For any region $S \subset D$, they adopt the geostatistical COS notion of "block average":

$$Y(S) = \frac{1}{|S|} \int_S Y(\mathbf{s}) d\mathbf{s}, \quad |S| > 0 \tag{4}$$

where $|S| = \int_S d\mathbf{s}$ is just the area of S in two-dimensional Euclidean space. Now, consider three types of support: *"Subgrid"*, $A_i, \quad i = 1, \ldots, n_a$; *"Grid"*, $B_j, \quad j = 1, \ldots, n_b$; and *"Supergrid"*, $C_k, \quad k = 1, \ldots, n_c$, where $0 \leq |A_i| \leq |B_j| \leq |C_k| < \infty$. Note that this framework includes situations with "nested grids" for which each A is contained in only one of the B's and in turn, each B is contained in only one of the C's. Alternatively, this formulation allows situations in which the grids overlap (e.g., some of the A's may overlap with various B's, etc.). In addition, the framework accommodates cases where the grids are regularly spaced or irregularly spaced.

Assume one has observations $Z(A_i), Z(C_k)$ for $i = 1, \ldots, n_a, k = 1, \ldots, n_c$, and the process of interest is to be considered at the intermediate B level of support. Although not considered here, it is easy to include data at the B level and data with point support as well. One approach to hierarchical COS is based on the assumption that for any $\mathbf{s} \in D$ there is some $B_j \in D$ such that:

$$Y(\mathbf{s}) = Y(B_j) + \gamma(\mathbf{s}), \tag{5}$$

where $\{\gamma(\mathbf{s}) : \mathbf{s} \in D\}$ is a correlated spatial process with covariance $c_\gamma(\mathbf{r}, \mathbf{s})$ and the distribution of $\gamma(\mathbf{s})$ does not depend on $Y(B_j), \quad j = 1, \ldots, n_b$. Then, one defines a subgrid process utilizing (4) and (5):

$$\begin{aligned} Y(A_i) &= \frac{1}{|A_i|} \int_{A_i} Y(B_j) d\mathbf{s} + \frac{1}{|A_i|} \int_{A_i} \gamma(\mathbf{s}) d\mathbf{s} \\ &= \mathbf{g}'_{a,i} \mathbf{Y}_B + \frac{1}{|A_i|} \int_{A_i} \gamma(\mathbf{s}) d\mathbf{s}, \end{aligned} \tag{6}$$

where $\mathbf{Y}_B \equiv (Y(B_1), \ldots, Y(B_{n_b}))'$ and elements of $\mathbf{g}_{a,i}$ are given by $g_{a,i}(j) = \frac{|B_j \cap A_i|}{|A_i|}$, the area of overlap between B_j and A_i, normalized by the area of A_i. Note, this is similar to the *ad hoc* approach used in many GIS packages to accommodate data of differing support and alignment (see the review by Gotway & Young, 2002).

For illustration, consider a conditional Gaussian assumption for the process $\mathbf{Y}_A = (Y(A_1), \ldots, Y(A_{n_a}))'$. In that case,

$$\mathbf{Y}_A | \mathbf{Y}_B, \ \boldsymbol{\Sigma}_\gamma^A \sim N(\mathbf{G}_a \mathbf{Y}_B, \ \boldsymbol{\Sigma}_\gamma^A) \tag{7}$$

where $\mathbf{G}_a$ contains the elements $g_{a,i}(j)$ and the (p, q)-th element of Σ_γ^A is given by

$$\Sigma_\gamma^A(p, q) = \int_{A_p} \int_{A_q} c_\gamma(\mathbf{r}, \mathbf{s}) d\mathbf{r} d\mathbf{s} / (|A_q||A_p|). \tag{8}$$

Given the measurement error model:

$$\mathbf{Z}_A|\mathbf{Y}_A, \ \Sigma_a \sim N(\mathbf{Y}_A, \ \Sigma_a) \tag{9}$$

one can combine (7) and (9) by integrating out the Y_A process to get,

$$\mathbf{Z}_A|\mathbf{Y}_B, \ \Sigma_\gamma^A, \ \Sigma_a \sim N(\mathbf{G}_a\mathbf{Y}_B, \ \Sigma_\gamma^A + \ \Sigma_a). \tag{10}$$

Note that we can derive an analogous distribution for $[\mathbf{Z}_C|\mathbf{Y}_B]$, the case where the data have larger support than the process.

Typically, the covariance function c_γ and the measurement error process must be parameterized in this setting. However, given that most of the complicated structure is in the process, it is often the case that these can be represented by relatively simple processes (e.g., stationary spatial models for γ and simple white noise measurement error). Indeed, the separation of the COS and measurement processes from the environmental process is the primary advantage of the hierarchical data model. Finally, note that the approach is not limited to the Gaussian assumptions. For example, Mugglin & Carlin (1998), Mugglin *et al.* (1999), Mugglin *et al.* (2000), and Best *et al.* (2000) discuss hierarchical approaches for misaligned count data.

In many cases, the volume of spatial data and/or the density and size of the prediction domain are so large that the hierarchical approaches mentioned above are not feasible. An alternative in that case is based on multi-scale models (e.g., Chou *et al.*, 1994; Huang & Cressie, 2000). That is, one considers nested levels of support where the "parent" at one level of support is associated with several "children" with smaller levels of support. For example, assume the data Z is observed at some support S, so one obtains a data model such as $[\mathbf{Z}_S|\mathbf{Y}_S, \ \Sigma]$. Then, one considers the distribution of this process $\mathbf{Y}_S$ conditioned on its "parent" so $[\mathbf{Y}_S|\mathbf{H}, \mathbf{Y}_p, \ \Sigma_s]$, where $\mathbf{Y}_p$ represent the parent nodes, and $\mathbf{H}$ describes the relationship between "parent" and "child". Clearly, in the Gaussian case one can also write a distribution for the parent nodes conditional on the children. Chou *et al.* (1994) illustrate how the use of these conditional distributions leads to very efficient Kalman filter implementations across scale. Huang & Cressie (2000) extend this idea into the fully hierarchical setting. Examples have considered ocean surface height (Fieguth *et al.*, 1995) and global ozone measurements from satellites (Huang *et al.*, 2002). It is important to keep in mind the vast amount of data considered in such analyses.

4 Process Models

Although the process models can relate to any process of interest in environmental science, this discussion will focus on spatial and spatio-temporal processes. In particular, we will focus on a general framework for spatial processes, approaches for modeling spatial structure, and hierarchical approaches to multivariate spatial models. We will also consider hierarchical spatio-temporal models and dynamic processes.

4.1 Spatial Models

The hierarchical framework can generalize most of the standard modeling techniques used for spatial statistical modeling (e.g., Cressie, 2000a). The following describes a general framework for spatial models, but is not intended to be exhaustive.

4.1.1 *General hierarchical spatial model framework*

Assume there are observations of a spatial process, denoted by $\mathbf{Z} = (Z(\mathbf{r}_1), \ldots, Z(\mathbf{r}_m))'$. Furthermore, assume there is some underlying spatial process of interest, denoted by $\mathbf{y} = (y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n))'$ where the spatial locations of the observations may or may not correspond to the locations (or the support) of the process of interest. Thus, the general framework for hierarchical spatial modeling is simply given by the component models:

$$[\mathbf{Z}|\mathbf{y}, \boldsymbol{\theta}][\mathbf{y}|\boldsymbol{\theta}_y][\boldsymbol{\theta}, \boldsymbol{\theta}_y], \tag{11}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_y$ are parameter vectors. As simple as this seems, it is quite flexible. For example, as shown by Le & Zidek (1992) and Handcock & Stein (1993), a hierarchical version of the geostatistical "universal kriging" model can be written,

$$\mathbf{Z}|\boldsymbol{\beta}, \mathbf{y}, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{K}\mathbf{y}, \sigma^2\mathbf{I}) \tag{12}$$

$$\mathbf{y} \sim N(\mathbf{0},\ \boldsymbol{\Sigma}(\theta_y)) \tag{13}$$

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0,\ \boldsymbol{\Sigma}_\beta) \tag{14}$$

$$[\sigma^2, \boldsymbol{\theta}_y, \boldsymbol{\beta}_0,\ \boldsymbol{\Sigma}_\beta] \tag{15}$$

where the $\mathbf{X}\boldsymbol{\beta}$ term typically corresponds to large-scale trend attributable to covariates in $\mathbf{X}$, σ^2 accounts for small-scale spatial variability, and $\mathbf{y}$ is the underlying correlated spatial process. Usually, if the observations and underlying spatial processes coincide, then $\mathbf{K}$ is just the identity matrix. Otherwise, $\mathbf{K}$ might be an incidence matrix or something more complicated (e.g., Wikle *et al.*, 1998). One may specify distributions for the parameters (15); often, point mass priors are specified.

The "universal kriging" model presented above is analogous to the linear mixed model (LMM) framework in classical modeling (i.e., specifically if $\boldsymbol{\beta}$ and the parameters are assumed to be fixed but unknown). This suggests that the general framework shown in (11) might also be used to consider a generalized linear mixed model (GLMM) formulation. Indeed, Diggle, Tawn & Moyeed (1998) have shown that one can consider non-Gaussian data conditioned on a Gaussian random field from this perspective. That is, one assumes the conditional mean relationship $E(Z(\mathbf{r}_i)|\mathbf{y}, \boldsymbol{\beta}) = \mu(\mathbf{r}_i)$. Then, for some link function $g(\ )$, it is assumed that the transformed mean can be written $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{K}\mathbf{y} + \boldsymbol{\epsilon}$, where the error term $\boldsymbol{\epsilon}$ is sometimes not necessary. Critical to this approach is the assumption $\mathbf{y} \sim N(\mathbf{0},\ \boldsymbol{\Sigma}(\theta_y))$. That is, $\mathbf{y}$ follows a Gaussian random field model, depending on some parameters θ_y that describe the spatial structure. As with the Gaussian case, there may be distributions on $\boldsymbol{\beta}$ and the variance parameters as well, similar to (14) and (15). This modeling approach is quite powerful.

4.1.2 *Spatial structure*

In the hierarchical spatial model above, it is assumed that $\mathbf{y}$ has covariance matrix $\boldsymbol{\Sigma}(\theta_y)$. In simple cases where $\mathbf{y}$ is defined in continuous space and is assumed to be stationary, one of the well-known classes of spatial dependence functions (e.g., the Matern class) can be considered, conditional on a relatively few parameters (e.g., Handcock & Stein, 1993). Although the stationary assumption seems restrictive, it is important to remember that after conditioning on the $\mathbf{X}\boldsymbol{\beta}$ component, the spatial structure is often reasonably close to stationary.

Despite the usefulness of stationary spatial models, sometimes more complicated spatial structures must be considered, particularly over large spatial domains. Guttorp & Sampson (1994) and Sampson, Damian & Guttorp (2001) have provided comprehensive reviews of approaches to modeling nonstationary spatial covariance. Many of these can be considered from a hierarchical perspective, including the spatial deformation approach (Damian, Sampson & Guttorp, 2001), the convolution

kernel approach (Higdon, 1998), and spectral approaches (Nychka *et al.*, 1999; Nychka *et al.*, 2002; Fuentes, 2002; Wikle, 2002b).

Another issue of practical concern is when the spatial process is very high-dimensional. That is, if the spatial process of interest is required at a large number of prediction locations, the implementation of traditional hierarchical spatial methodology outlined by Diggle *et al.* (1998) can be problematic. Recently, Wikle (2002b) showed that spectral transform approaches can be used in the non-Gaussian hierarchical spatial setting to efficiently model spatial processes of high dimension.

4.1.3 Markov random field models

As described above, the process modeling stage in a hierarchical model may be decomposed into substages. This is one motivation for Markov random field (MRF) spatial models (Besag, 1974). These models are specified conditionally, which can afford certain computational advantages. For a review of these processes see Cressie (1993) and Besag *et al.* (1995). A very simple outline of such processes relative to spatial process models is given below.

MRF models are most useful in situations where the set of all possible spatial locations is discrete (countable). A collection of such sites is often referred to as a *lattice*. Examples include disease mapping in counties and modeling air-pollution on a grid. The first is an example of an *irregular lattice* and the latter is often a *regular lattice*. Regular lattices have neighborhoods that are primarily defined by adjoining sites and irregular lattices often have neighborhoods defined by Euclidean proximity.

Consider a spatial process defined at n spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, $\mathbf{y} = (y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n))'$. This process has joint distribution $[y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)]$. From this joint distribution, the conditional distribution of the process at each location i can be expressed in terms of all other sites ($j \neq i$) as

$$[y(\mathbf{s}_i)|\{y(\mathbf{s}_j) : j \neq i\}], \quad i = 1, \ldots, n. \tag{16}$$

The neighborhood N_i of the i-th site is then defined as the collection of locations such that

$$[y(\mathbf{s}_i)|\{y(\mathbf{s}_j) : j \neq i\}] = [y(\mathbf{s}_i)|\{y(\mathbf{s}_j) : j \in N_i\}], \quad i = 1, \ldots, n. \tag{17}$$

In other words, the conditional probability at site i depends only on nearby values of the process $\{y(\mathbf{s}_j) : j \in N_i\}$. The specification of these conditional, neighborhood-specific distributions must be made consistently so that the joint distribution is well-defined (e.g., Besag, 1974; Kaiser & Cressie, 2000). Examples of such models include auto-Gamma models for non-negative continuous processes, auto-Poisson models for spatial counts, auto-logistic models for binary spatial random variables, and auto-Gaussian models for spatial Gaussian processes on a lattice. The auto-Gaussian model is considered in more detail below.

Consider a spatial process y in which the conditional models (17) are Gaussian with,

$$\text{Prob}\{y(\mathbf{s}_i)|y(N_i)\} = \frac{1}{\sqrt{2\pi}\sigma_i} \exp[-\frac{1}{2\sigma_i^2}\{y(\mathbf{s}_i) - \mu(\mathbf{s}_i) - \sum_{j \in N_i} c_{ij}[y(\mathbf{s}_j) - \mu(\mathbf{s}_j)]\}^2], \tag{18}$$

where c_{ij} are coefficients that must obey certain restrictions (given below) and $\mu(\mathbf{s}_i) \equiv E[y(\mathbf{s}_i)]$. It can be shown that the joint distribution for $\mathbf{y}$ is then given by

$$\mathbf{y} \sim N(\boldsymbol{\mu}, (\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}), \tag{19}$$

provided $(\mathbf{I} - \mathbf{C})^{-1}$ is positive-definite, and where $\boldsymbol{\mu} \equiv [\mu(\mathbf{s}_i), \ldots, \mu(\mathbf{s}_n)]'$, $\mathbf{C}$ is an $n \times n$ matrix with (i, j)-th element c_{ij}, where $c_{ij}\sigma_j^2 = c_{ji}\sigma_i^2$ (symmetry), $c_{ii} = 0$, $c_{ik} = 0$ for k not in N_i, and $\mathbf{M}$ is a diagonal matrix with $\sigma_1^2, \ldots, \sigma_n^2$ on the main diagonal (e.g., Cressie, 1993). Note that the inverse of the joint covariance matrix, $\mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$, often called the "precision matrix", is typically sparse when the neighborhood is relatively small.

The Gaussian MRF can easily be incorporated in the hierarchical framework described previously. For example, let $\mathbf{Z} = \mathbf{X}\beta + \mathbf{y} + \epsilon$, where now $\mathbf{y} \sim N(\mathbf{0}, (\mathbf{I} - \mathbf{C})^{-1}\mathbf{M})$. The parameters are then specified at the next level of the hierarchy. In practice, the neighborhood structure of the MRF is often simplified greatly so that the number of unknown parameters is reasonably small. For example, a typical assumption is that only those neighbors that are immediately adjacent to a given location are necessary to specify the conditional distribution; this is often referred to as a first-order dependence model or conditional autoregressive (CAR) model. There are many examples of such models in the disease mapping and abundance mapping literature (e.g., Clayton & Bernardinelli, 1992). Although the CAR model is used frequently, its inherent local spatial dependence can be quite limiting. Rue & Tjelmeland (2002) show that one needs high-order dependence to obtain covariance structures similar to the classical geostatistical spatial dependence models (e.g., the Matern class).

Although data often arise in what may be viewed naturally as continuous space, one may parameterize the underlying spatial process as a discrete-space MRF. This is a natural formulation when a convenient stratification scheme can be specified, where "within-strata" measurements are assumed to be independent conditional on the stratum effect. In this case, one would express the model as $\mathbf{Z} = \mathbf{X}\beta + \mathbf{K}\mathbf{y} + \epsilon$ where $\mathbf{K}$ associates each datum with one of the discrete strata. Note the similarity of this interpretation with the change of support notions given in Section 3.1. Indeed, the hierarchical change of support can be used to specify the data model such that the point support data are given in terms of a lattice process at larger support, which can then in turn be modeled by a MRF. Thus, MRFs may be useful regardless of whether or not the data actually have discrete spatial support. However, it is important to recognize the inherent disadvantage of MRFs regarding COS. In particular, MRFs defined on point support do not inherently aggregate to larger scales. The construction of the process on the larger support is non-trivial. For example, consider an $n \times 1$ process $\mathbf{y}$ that has a MRF structure with covariance matrix $\Sigma = (\mathbf{I} - \mathbf{C})^{-1}\mathbf{M} = \{\sigma_{kl}\}_{k,l=1,\ldots,n}$. Now, consider a change of support problem in which we simply combine two elements of $\mathbf{y}$, say $z_1 = y(\mathbf{s}_i) + y(\mathbf{s}_j)$ for some $i \neq j$. Then, consider the joint distribution of the $(n-1)$-dimensional aggregated process consisting of z_1 and all of the $y(\mathbf{s}_k), k \neq i, j$. Even such a simple COS (adding two elements of original process) does not yield obvious MRF structure on the precision matrix of the new aggregated process. One can see this by noting that $\text{var}(z_1) = \sigma_{ii} + \sigma_{jj} + 2\sigma_{ij}$ and $\text{cov}(z_1, y(\mathbf{s}_k)) = \sigma_{ik} + \sigma_{jk}$, $k \neq i, j$. Thus, if we define the covariance of the new COS process as Γ, then it is not necessarily the case that the precision matrix of the aggregated process, Γ^{-1}, is suitably sparse.

Finally, we note that for the Gaussian MRF model, the spatial dependence is specified through the conditional distributions (17) and ultimately the *inverse* covariance matrix (i.e., precision matrix). This is in contrast to the continuous Gaussian random field case where one specifies the covariance matrix directly. There may be some computational advantages to the MRF approach in many practical situations due to the sparcity of the inverse covariance matrix. Furthermore, in the Bayesian setting, one can usually make use of the conditional distributions in efficient MCMC algorithms.

4.1.4 Multivariate spatial models

It is seldom the case in environmental studies that measurements are made on a single variable. For example, air pollution monitoring networks usually collect observations on several pollutants as well as meteorological variables. Clearly, there are relationships between these variables and such relationships will make up the joint spatial structure of the variables. Even if interest is in one specific variable, it is often the case that one can "borrow strength" from other related variables. In either context, one must understand the relationships between the variables.

The traditional approach to modeling multivariate spatial processes is known as "co-kriging" and involves the direct specification of the joint spatial structure among the relevant set of variables (Meyers, 1982; Ver Hoef & Cressie, 1993). Of course, it is often extremely difficult to specify

realistically the joint spatial structure for complicated environmental processes, especially the second-order spatial structure. Alternatively, one can recast the model in a hierarchical perspective as outlined in Royle & Berliner (1999). For example, in the case of a bivariate spatial process $[\mathbf{y}_1, \mathbf{y}_2]$ and parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$, one can simply consider

$$[\mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\theta}] = [\mathbf{y}_1|\mathbf{y}_2, \boldsymbol{\theta}_1][\mathbf{y}_2|\boldsymbol{\theta}_2][\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]. \tag{20}$$

Rather than consider the interaction between the processes via the second-moments (e.g., covariance), one establishes a first-order relationship between the two processes. That is, $\mathbf{y}_1$ is related to $\mathbf{y}_2$ in the conditional mean, about which one usually knows more than the joint covariance. This is particularly true if one knows that there is a mechanistic (or causal) relationship between $\mathbf{y}_1$ and $\mathbf{y}_2$. Of course, the model is valid regardless of such a relationship. Royle & Berliner (1999) illustrated the approach with the problem of predicting ozone given temperature. Since it is well known that the chemistry of ozone production is related to temperature, they conditioned ozone ($\mathbf{y}_1$) on temperature ($\mathbf{y}_2$) with a model like:

$$\mathbf{y}_1|\mathbf{y}_2, \mathbf{B}, \boldsymbol{\beta}_1, \ \boldsymbol{\Sigma}_{1|2} \sim N(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{B}\mathbf{y}_2, \ \boldsymbol{\Sigma}_{1|2}), \tag{21}$$

$$\mathbf{y}_2|\boldsymbol{\beta}_2, \ \boldsymbol{\Sigma}_2 \sim N(\mathbf{X}_2\boldsymbol{\beta}_2, \ \boldsymbol{\Sigma}_2), \tag{22}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ both contained a column of ones and a column representing the relevant latitude, $\mathbf{B} = \theta\mathbf{I}$, and the parameters were given distributions at the next stage.

A very important aspect of such spatial modeling is that the conditioning on observed variables need not be explicit. Rather, one can condition the observed processes on some unobserved (latent) process, and model the latent process at a lower stage of the hierarchy. This is particularly useful when there are a large number of variables that can be at least partially explained by a common (hopefully, lower dimensional) spatial process. For example Royle *et al.* (1999) considered such an approach for modeling wind components given an unobserved pressure process. Specifically, for the east-west ($\mathbf{U}$) and north-south ($\mathbf{V}$) wind component processes with associated observations $\mathbf{D}_u$ and $\mathbf{D}_v$, respectively, and the unobserved pressure process $\mathbf{P}$, the basic model was:

$$\mathbf{D}_u, \mathbf{D}_v|\mathbf{U}, \mathbf{V}, \sigma^2 \sim N\left(\begin{pmatrix} \mathbf{KU} \\ \mathbf{KV} \end{pmatrix}, \sigma^2\mathbf{I}\right), \tag{23}$$

$$\mathbf{U}, \mathbf{V}|\mathbf{P}, \mathbf{B}_u, \mathbf{B}_v, \ \boldsymbol{\Sigma} \sim N\left(\begin{pmatrix} \mathbf{B}_u\mathbf{P} \\ \mathbf{B}_v\mathbf{P} \end{pmatrix}, \ \boldsymbol{\Sigma}\right), \tag{24}$$

$$\mathbf{P}|\mu_p, \sigma_p^2, \mathbf{R}_\theta \sim N(\mu_p\mathbf{I}, \sigma_p^2\mathbf{R}_\theta), \tag{25}$$

where $\mathbf{K}$ was an incidence matrix relating observed locations to prediction locations, and $\mathbf{B}_u$, $\mathbf{B}_v$ specified the conditional relationship between wind and pressure (based on a physical relationship known as geostrophy). These parameters and the others were given distributions in the lower levels of the hierarchy. The critical idea in this model is that by specifying a relationship between the wind process and the pressure process in the mean, the structure of $\boldsymbol{\Sigma}$ is greatly simplified.

4.2 Spatio-Temporal Models

As mentioned previously, most processes of interest to environmental scientists are either spatio-temporal, or space or time averages of such processes. In principle, models for spatio-temporal processes are relatively easy to formulate from a joint perspective. However, lack of complete understanding and the "curse of dimensionality" make the implementation of these models challenging, to say the least. A hierarchical formulation is often preferable as it allows one to build very complicated

joint spatio-temporal structure in terms of relatively simple conditional models.

Consider spatio-temporal observations $Z(\mathbf{s}; t)$ defined for $\mathbf{s} \in D_s$, $t \in D_t$, where D_s and D_t are spatial and temporal domains, respectively. The domains may be continuous or discrete, but we typically consider D_t to be a discrete collection of times (e.g., $D_t = \{1, 2, \ldots, T\}$). Analogous to the spatial process discussion above, we may then formulate the spatio-temporal process where $[Z(\mathbf{s}; t)|\mathbf{y}]$ is a conditionally independent distribution for all $\mathbf{s}$ and t, and where $y(\mathbf{s}; t)$ is a spatio-temporal process such that $\mathbf{y} \equiv (\mathbf{y}_1' \ \ldots \ \mathbf{y}_T')'$, where $\mathbf{y}_t \equiv (y(\mathbf{s}_1; t), \ldots, y(\mathbf{s}_n; t))'$ and $\mathbf{y}$ has covariance matrix $\boldsymbol{\Sigma}_y$. In this case, $\boldsymbol{\Sigma}_y$ is an $(n + T) \times (n + T)$ covariance matrix with elements $c_y(\mathbf{s}, \mathbf{s}', t, t') = \text{cov}(y(\mathbf{s}; t), y(\mathbf{s}'; t'))$. There are practical limitations with this formulation. First, although the known classes of spatio-temporal covariance functions continues to expand (e.g., Cressie & Huang, 1999; Gneiting, 2002), most of these are not adequate for explaining complicated spatio-temporal interactions for environmental processes. Secondly, the dimensionality of the joint spatio-temporal process $\mathbf{y}$ is often prohibitively large.

4.2.1 Hierarchical spatio-temporal models

Complicated joint spatio-temporal process can often be decomposed into relatively simple conditional models. Early examples of this include the work by Handcock & Wallis (1994), Brown *et al.* (1994), Waller *et al.* (1997), Knorr-Held & Besag (1998), and Wikle *et al.* (1998). A general model for a spatio-temporal process $y(\mathbf{s}; t)$ takes the form:

$$y(\mathbf{s}; t) = \mu(\mathbf{s}; \boldsymbol{\theta}_{\mu,t}) + \gamma(t; \boldsymbol{\theta}_{\gamma,\mathbf{s}}) + \alpha(\mathbf{s}, t; \boldsymbol{\theta}_{\alpha,s,t}) + \epsilon(\mathbf{s}; t). \tag{26}$$

In general, the first term on the right-hand side, $\mu(\mathbf{s}; \boldsymbol{\theta}_{\mu,t})$ represents a spatial trend surface, but with parameters that may vary with time. Similarly, the second term $\gamma(t; \boldsymbol{\theta}_{\gamma,s})$ represents a temporal trend that may have spatially varying parameters. The third term on the right-hand side, $\alpha(\mathbf{s}; t; \boldsymbol{\theta}_{\alpha,s,t})$ typically represents a complicated (usually non-separable) spatio-temporal process of primary interest; often this is a dynamic process, as discussed below. The last term, $\epsilon(\mathbf{s}; t)$ is usually an uncorrelated spatio-temporal error process.

Obviously, models such as (26) may not be identifiable unless substantial prior knowledge is used in the construction of the component distributions for μ, γ, α, and ϵ. In essence, this is the fundamental challenge of space-time modeling. For example, one must decide if a process is better represented as a spatial process with time-varying parameters or a temporal process with spatially-varying parameters. In some cases, the dimensionality of the relevant processes suggests the proper form. However, in most cases, the decision is based on prior understanding.

In Wikle *et al.* (1998) the process of interest was monthly average maximum temperature in Iowa. In that case it was clear from climatological understanding of the process that μ should simply depend on space and it was assigned a conditionally Gaussian MRF with a mean depending on latitude and longitude according to a simple linear relationship. Furthermore, γ was used to account for the very strong sinusoidal seasonal cycle of atmospheric temperature. Since the phase and amplitude of the seasonal cycle changes as a function of space, the parameters describing the sinusoidal cycle were allowed to vary as a linear function of latitude and longitude. Finally, the process of primary interest was α which was assumed to be a dynamic process, accounting for month-to-month circulation changes in the atmosphere. Such processes are described in more detail in the next section. In summary, the spatio-temporal process for monthly temperature included spatio-temporal variability from climatological, seasonal, and shorter-term dynamical processes. A realistic joint specification of this structure was not possible, but rather was "constructed" by the aforementioned component models.

4.2.2 *Dynamic models*

It is natural for many environmental processes to assume that the immediate past state of the process affects the current state. Indeed, many spatio-temporal models have included such a dynamic component (e.g., Guttorp *et al.*, 1994; Huang & Cressie, 1996; Waller *et al.*, 1997; Knorr-Held & Besag, 1998; Mardia *et al.*, 1998; Meiring *et al.*, 1998; Wikle *et al.*, 1998; Wikle & Cressie, 1999; Wikle *et al.*, 2001; Stroud *et al.*, 2001). In general, the joint spatio-temporal process α, corresponding to $\alpha(\mathbf{s}; t)$ in (26) can often be further factored into a series of conditional models. For many processes, a dynamical factorization based on a Markov assumption in time is appropriate. That is,

$$[\alpha] = [\alpha_0] \prod_{t=1}^{T} [\alpha_t | \alpha_{t-1}, \theta_t], \tag{27}$$

where $\alpha_t = (\alpha(\mathbf{s}_1; t), \ldots, \alpha(\mathbf{s}_n; t))'$ and the conditional distributions $[\alpha_t | \alpha_{t-1}, \theta_t]$ depend on a collection of parameters θ_t that describe the dynamical evolution and the structure of the associated driving noise process. These parameters may or may not be time-varying depending on the specific process of interest. For example, this process might follow a first-order vector autoregressive model such as $\alpha_t = \mathbf{H}_{\theta_t} \alpha_{t-1} + \eta_t$, where $\eta_t \sim N(\mathbf{0}, \Sigma_\eta)$ is a spatial noise process, and $\mathbf{H}_{\theta_t}$ is a collection of parameters that describe the evolution of the α process. In settings where the parameters are known or can be estimated, one may implement such a model in a Kalman filter framework (e.g., West & Harrison, 1989; Huang & Cressie, 1996; Wikle & Cressie, 1999).

When the number of spatial locations is large, the number of parameters in the "propagator matrix" $\mathbf{H}_{\theta_t}$ may prohibit likelihood-based estimation procedures. One may have scientific knowledge that suggests relatively simple parameterizations of $\mathbf{H}_{\theta_t}$ (e.g., Ghil *et al.*, 1981). Alternatively, one can model $\mathbf{H}_{\theta_t}$ in a hierarchical fashion, by specifying its distribution, or the distribution of the parameters θ_t. In general, even a simple model in which $\mathbf{H}_{\theta_t}$ is diagonal with parameters varying by spatial location (i.e., a spatially indexed first-order autoregressive process), provides a simple way to model processes with non-separable spatio-temporal interaction. Such formulations don't allow inherently for the propagating phenomena that are often present in geophysical processes. However, relatively simple modifications can allow for such things. For example, in Wikle *et al.* (1998) it was assumed that the dynamic structure was based on a "lagged nearest neighbor" structure. In that case, the value of $\alpha(\mathbf{s}; t)$ was related to its immediate past value at location $\mathbf{s}$ and its four nearest neighbors at the previous time. The parameter corresponding to the location $\mathbf{s}$ was then spatially varying, but the lagged-neighbor coefficients were not.

4.2.3 *Motivation for spatio-temporal dynamic models*

Although it has been apparent for sometime that dynamic models are essential in describing spatio-temporal processes in the environmental sciences, it is only relatively recently that statisticians have been attempting to motivate the dynamical dependence in such models. For many biological and physical processes, one can make use of partial differential equations (PDEs) that govern (approximately) the underlying process in order to develop prior distributions for $\mathbf{H}_{\theta_t}$. For example, Wikle *et al.* (2001) used the shallow-water PDEs that approximate atmospheric processes in the tropics to motivate priors for predicting high-resolution wind fields over the tropical ocean.

As an example of using PDE information to motivate a dynamic process model, consider the study of Wikle (2003) who used a classic diffusion PDE to develop priors for modeling the spread of an invasive species. In particular, he considered the general diffusion PDE for the spatio-temporal process $u(x, y; t)$,

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\delta(x, y)\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta(x, y)\frac{\partial u}{\partial y}\right) + \beta u, \tag{28}$$

where the spatial location is defined by the Euclidean coordinates $\mathbf{s} = (x, y)$, $\delta(x, y)$ is a spatially varying diffusion coefficient, and β is a growth coefficient. This is a generalization of Skellam's (1951) classic diffusion model plus Malthusian growth. Although analytical studies of such equations have shown them to be reasonable for explaining the spread of invasive organisms, it is by no means a perfect description of such processes. Thus, it is important to recognize that this model is only the basis for a stochastic dynamic model. That is, it is simply the motivation for our prior formulation, and will ultimately be modified by the data and the allowance of random parameters at the next stage. Specifically, the model is made much more flexible since the parameters $\delta(x, y)$ and β are given distributions at the next level of the hierarchy. These random parameters account for effects such as environmental and demographic stochasticity, which are difficult to consider in the traditional deterministic PDE-based framework.

Discretization of (28) using first-order forward differences in time and centered differences in space yields a space-time difference equation in which the value of the u-process at a given time is related to its past value at that location and its four nearest neighbors (e.g., Wikle, 2003). Adding an error term to account for uncertainties due to the discretization and other model misspecifications then gives

$$\mathbf{u}_t = \mathbf{H}(\boldsymbol{\delta}, \beta)\mathbf{u}_{t-1} + \mathbf{H}_B(\boldsymbol{\delta})\mathbf{u}_{t-1}^B + \boldsymbol{\gamma}_t, \tag{29}$$

where $\mathbf{u}_t$ corresponds to an arbitrary vectorization of the gridded u-process at time t, and $\mathbf{H}(\boldsymbol{\delta}, \beta)$ is a sparse matrix with five non-zero diagonals that depend on $\boldsymbol{\delta}$ and β. Furthermore, $\mathbf{u}_{t-1}^B$ is a vector of boundary values for the u-process, and $\mathbf{H}_B(\boldsymbol{\delta})$ is a sparse matrix with elements corresponding to the boundary elements of the discretization. It is possible, and indeed desirable in many cases, to model $\mathbf{u}_t^B$ as a random process. One can show that this suggests a hierarchical approach to accommodating boundary conditions in stochastic solutions to PDE models (see Wikle *et al.*, 2003).

The formulation (29) is similar to the space-time autoregressive (STAR) models outlined in Cliff *et al.* (1975). However, STAR models can be difficult to fit in high dimensions and don't account explicitly for boundary conditions. Wikle *et al.* (1998) showed that these models are relatively easy to consider from the hierarchical perspective. In the present case, the hierarchical structure of the propagator matrices $\mathbf{H}$ and $\mathbf{H}_B$ is suggested by the PDE process parameters. This motivation greatly strengthens the hierarchical model and the justification of specific priors on parameters.

Note that in addition to PDEs, recent work has shown that for continuous space/discrete time processes, one can use integro-difference equations in a hierarchical context to model dynamical systems (Wikle & Cressie, 1999; Brown *et al.*, 2000; Brown *et al.*, 2001; Wikle, 2001; Wikle, 2002a). These models have great potential for motivating dynamical variability in cases where PDEs are not appropriate.

4.2.4 Dimension reduction

Although the hierarchical representation of spatio-temporal dynamic processes simplifies the modeling, it is often the case that the process vectors $\boldsymbol{\alpha}_t$ are of such high dimension that implementation is still problematic. Recasting the problem in a spectral context can enable implementation either through the use of fast computational algorithms, or by facilitating dimension reduction. That is, let

$$\boldsymbol{\alpha}_t = \boldsymbol{\Phi}\mathbf{a}_t \tag{30}$$

where $\boldsymbol{\Phi}$ is a matrix of spectral basis functions (e.g., Fourier, wavelet, empirical) and $\mathbf{a}_t$ are the associated spectral coefficients. In many cases, after such a transformation to spectral-space, one can then model the $\mathbf{a}_t$ process dynamically, e.g., $\mathbf{a}_t = \tilde{\mathbf{H}}\mathbf{a}_{t-1} + \tilde{\boldsymbol{\eta}}_t$. Because of the spectral transformation, $\tilde{\mathbf{H}}$ and $\tilde{\boldsymbol{\eta}}_t$ may be substantially simpler in structure than for the corresponding terms of the $\boldsymbol{\alpha}_t$ process. This, in conjunction with the possible use of fast spectral transform routines, can greatly facilitate

computation and model specification (e.g., Wikle *et al.*, 2001). In addition, it is often the case that only a few spectral coefficients are necessary to adequately describe the dynamical behavior of the process of interest. In that case, rather than (30) one might assume $\boldsymbol{\alpha}_t = \tilde{\boldsymbol{\Phi}}\mathbf{a}_t + \boldsymbol{\nu}_t$, where $\mathbf{a}_t$ has dimension much smaller than $\boldsymbol{\alpha}_t$, $\tilde{\boldsymbol{\Phi}}$ represents some subset of the basis functions in $\boldsymbol{\Phi}$, and and $\boldsymbol{\nu}_t$ can often be represented by relatively simple structure (e.g., Wikle & Cressie, 1999; Berliner *et al.*, 2000).

4.2.5 Hidden processes

As described above in the case of multivariate spatial models, the complicated joint structure in spatio-temporal models can often be simplified by conditioning on some other process. Typically, this process is not observed and is considered to be a latent or hidden process. An early example in environmental science is the work of Hughes & Guttorp (1994) who used hidden Markov models of unobserved weather states to model spatio-temporal variability in atmospheric precipitation. More recently, Berliner *et al.* (2000) used a hidden climate state process to facilitate long-lead forecasting of Pacific sea surface temperatures and ultimately El Niño and La Niña conditions. In particular, their model included a component like, $\mathbf{a}_{t+\tau} = \mathbf{H}_{\theta_{t,\tau}}\mathbf{a}_t + \boldsymbol{\eta}_{t+\tau}$, where the propagator matrix $\mathbf{H}_{\theta_{t,\tau}}$ was time-varying and depended on the current state of the ocean and the likely future state. The likely future state was determined by a hidden climate process.

5 Challenges

The use of complicated hierarchical models for environmental processes is relatively new, corresponding for the most part to the popularization of MCMC techniques in statistics, beginning in the late 1980s. Since that time, a vast array of problems have been considered from this perspective ranging from disease mapping to weather prediction. Although the use of these techniques has led to analyses that would not have been thought possible a decade before, there are still many issues to be considered. A recent workshop at The Ohio State University ("Workshop on Hierarchical Modeling in Environmental Statistics, May 2000) produced a "position paper" that describes many of the outstanding issues related to the topic (Cressie, 2000b). Problem areas discussed included methodological issues, practical issues, and even issues concerning the inherent collaborative nature of hierarchical modeling and science. Some of these issues are discussed below.

Although statistical computation has exhibited remarkable progress over recent years, there are still many issues related to computation and implementation of hierarchical models. For example, it is very difficult, if not impossible, to establish convergence for MCMC applications in very high-dimensional settings. Although much has been done to develop diagnostics for such models (e.g., see review in Robert & Casella, 1999, Chap. 8) much remains to be done. In addition, the use of parallel computing strategies in hierarchical models has not been adequately explored. As models become ever more complicated and applied in higher dimensional settings, the use of parallel computing may become practically essential.

Hierarchical models are subjective; there is more than one way to model a complicated system. Consequently, there is a need for some "objective" measures of comparing models. Although some very nice approaches to model selection are now being considered (e.g., Spiegelhalter *et al.*, 2002), much more remains to be done. Furthermore, model averaging is often seen as a useful approach for dealing with model uncertainty (e.g., Clyde, 1999). This is also an active area of research. Finally, issues related to model sensitivity must be considered. For example, at some point in a model hierarchy, model choices are subjective, and may simply be given point-mass priors. In most cases, there are too many parameters to evaluate sensitivity to such priors by complete factorial experiments. There is a need for more objective approaches to evaluating such sensitivities.

As discussed above, dimensionality of spatial and spatio-temporal processes often prohibits efficient implementation of hierarchical formulations. In those cases, it is sometimes the case that the essential process variability can be efficiently modeled in a lower dimensional manifold. Although empirical orthogonal functions (EOFs) have proven useful in this regard (e.g., Wikle & Cressie, 1999; Berliner *et al.*, 2000), there is a need for additional dimension reduction strategies.

Going back to the Introduction, the environment encompasses many different interacting processes. It follows that one of the biggest challenges and growth areas in environmental statistics will be concerned with *linking* the many processes contributing to specific environmental questions of concern. Although the approaches outlined above will be useful, many additional methodological and practical innovations must be developed before such linkages can easily be accomplished in practice. A study by Berliner, Milliff & Wikle (2003) involves linking a simple atmospheric model with a simple ocean model in the Labrador Sea. This coupled model is complicated by the non-linear nature of the ocean flow and especially, the non-linear interaction between the atmosphere and the ocean. Ultimately, this coupling is made possible by using physically-motivated process and prior models, and the use of a hybrid Gibbs Sampling/Importance Sampling algorithm. Much more work is required to implement such models over varying spatial and temporal scales.

Many of the studies cited in this article involve both subject matter scientists and statisticians. Of course, such collaborations are not at all unusual in statistics. However, if one looks closely they will find that many collaborations involving hierarchical models are in fact different than traditional collaborative efforts. The chief difference is that the statistician and the subject matter scientist must both be involved in the detailed construction of the model. Since the essence of a good hierarchical model is strong prior information at each stage, it follows that the scientists must be able to convey their knowledge in this conditional framework. The statistician must be able to translate this scientific knowledge into reasonable statistical models, in some cases facilitating the scientist to "think conditionally", and ultimately decide whether the model can in fact be implemented. Thus, there is constant feedback and the endeavor is dynamic.

Of course, as anyone who has engaged in serious multi-disciplinary research can attest, it is easier said than done. Not the least problem is the issue of communication. Clearly, jargon issues can be significant and scientists need to pay ever increasing attention to using well-defined scientific language. Furthermore, only in rare instances are all of the collaborating scientists located at the same physical facility. Although the Internet is invaluable to such enterprises, it is not a solution in itself. For example, e-mail, although nearly instantaneous, can easily lead to misunderstandings and unintentional annoyances. New communication infrastructures are being developed to facilitate more seamless remote collaboration in science. Statisticians and scientists must overcome their initial reluctance to utilize such technology.

In conclusion, we reiterate the primary theme that hierarchical modeling provides a framework by which to simplify complicated environmental systems, so that uncertainty can be linked in a coherent fashion. Quite simply, by the power of conditioning, one can consider less complicated dependence structures, whether they be in the data model, process model, or parameters. Often, such conditional thinking follows naturally from the subject matter. In other cases, subject matter scientists and statisticians must work together to establish appropriate conditional models. If the immediate past is any indication, the future of hierarchical modeling in environment science is very bright indeed. As the human race places more and more demands on the planet and its component systems, we will never lack important environmental problems on which to work. We can take some consolation in the fact that we are well on our way to having the tools to examine them.

Acknowledgements

The development of this paper was supported by the U.S. EPA's Science to Achieve Results (STAR) program, Assistance Agreement R827257-01-0 and the NASA/Goddard Space Flight Center under Award NAG5-8344. I would like to thank Mark Berliner, Noel Cressie, Ralph Milliff and Andy Royle whose various collaborative efforts have contributed immensely to my understanding of the power of hierarchical models in environmental statistics. I would also like to thank Noel Cressie for his comments on an early draft and an anonymous reviewer who provided several valuable comments.

References

Berliner, L.M. (1996). Hierarchical Bayesian time series models. *Maximum Entropy and Bayesian Methods*, Eds K. Hanson and R. Silver, pp. 15–22. Kluwer Academic Publishers.

Berliner, L.M., Wikle, C.K. & Cressie, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian Dynamic Modeling. *Journal of Climate*, **13**, 3953–3968.

Berliner, L.M., Milliff, R.F. & Wikle, C.K. (2003). Bayesian hierarchical modeling of air-sea interaction. *Journal of Geophysical Research - Oceans*, **108**, to appear.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, B*, **36**, 192–236.

Besag, J., Green, P., Higdon, D. & Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, **10**, 3–66.

Best, N.G., Ickstadt, K. & Wolpert, R.L. (2000). Spatial Poisson regression for health and exposure data measured at disparate resolutions. *Journal of the American Statistical Association*, **95**, 1076–1088.

Blackmon, M. & Colleagues (2001). The community climate system model. *Bulletin of the American Meteorological Society*, **82**, 2357–2376.

Brown, P.E., Karesen, K.F., Roberts, G.O. & Tonellato, S. (2000). Blur-generated non-separable space-time models. *J.R. Statistical Society, B*, **62**, 847–860.

Brown, P.E., Diggle, P.J., Lord, M.E. & Young, P.C. (2001). Space-time calibration of radar rainfall data. *Applied Statistics*, **50**, 221–241.

Brown, P.J., Le, N.D. & Zidek, J.V. (1994) Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics*, **22**, 489–509.

Carlin, B.P. & Louis, T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis, Second Edition*. Boca Raton: Chapman and Hall/CRC.

Chou, K.C., Willsky, A.S. & Nikoukhah, R. (1994). Multiscale systems, Kalman filters, and Riccati equations. *IEEE Transactions on Automatic Control*, **39**, 479–492.

Clayton, D.G. & Bernardinelli, L. (1992). Bayesian methods for mapping disease risk. In *Geographical and Environmental Epidemiology: Methods for Small Area Studies*, Eds. P. Elliott, J. Cuzick, D. English and R. Stern. Oxford: Oxford University Press.

Cliff, A.D., Hagget, P., Ord, J.K., Bassett, K.A. & Davies, R.B. (1975). *Elements of Spatial Structure: A Quantitative Approach*. Cambrigde: Cambridge University Press.

Clyde, M. (1999). Bayesian Model Averaging and Model Search Strategies (with discussion). In *Bayesian Statistics 6*. Eds. J.M. Bernardo, A.P. Dawid, J.O. Berger and A.F.M. Smith, pp. 157–185. Oxford University Press.

Cressie, N.A.C. (1993). *Statistics for Spatial Data, Revised Edition*. Wiley & Sons, 900 pp.

Cressie, N. (1996). Change of support and the modifiable areal unit problem. *Geograph. Systems*, **3**, 159–180.

Cressie, N. (2000a) Spatial statistics and environmental sciences. *Proceedings of the Section on Statistics and the Environment*, pp. 1–10. Alexandria: American Statistical Association.

Cressie, N. (2000b). Position Paper from the Workshop on Hierarchical Modeling in Environmental Statistics. Columbus, Ohio, May 14–16. (available from `http://www.stat.ohio-state.edu/sses/WS2K/`)

Cressie, N, & Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94**, 1330–1340.

Damian, D., Sampson, P.D. & Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics*, **12**, 161–178.

Diggle P.J., Tawn, J.A. & Moyeed, R.A. (1998). Model-based geostatistics (with discussion). *Applied Statistics*, **47**, 299–350.

Fieguth, P.W., Karl, W.C., Willsky, A.S. & Wunsch, C. (1995). Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON satellite altimetry. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 280–292.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, **89**, 197–210.

Gelfand, A.E. & Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

Gelfand, A.E., Zhu, L. & Carlin, B.P. (2001). On the change of support problem for spatio-temporal data. *Biostatistics*, **2**, 31–45.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.

Ghil, M., Cohn, S.E., Tavantzis, J., Bube, K. & Isaacson, E. (1981). Applications of estimation theory to numerical weather prediction. *Dynamic meteorology: Data assimilation methods*, Eds. L. Bengtsson, M. Ghil and E. Källén, pp. 139–224.

New York: Springer-Verlag, .
Gneiting, T. (2002) Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, **97**, 590–600.
Gotway, C.A. & Young, L.J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, **97**, 632–648.
Guttorp, P., Meiring, W. & Sampson, P.D. (1994). A space-time analysis of ground-level ozone data. *Environmetrics*, **5**, 241–254.
Guttorp, P. & Sampson, P.D. (1994). Methods for estimating heterogeneous spatial covariance functions with environmental applications. In *Handbook of Statistics*, Eds. G.P. Patil and C.R. Rao, Vol. 12, pp. 661–689. New York: Elsevier Science.
Handcock, M.G. & Stein, M.L. (1993). A Bayesian analysis of kriging. *Technometrics*, **35**, 403–410.
Handcock, M.S. & Wallis, J.R. (1994). An approach to statistical spatial-temporal modeling of meteorological fields (with discussion). *Journal of the American Statistical Association*, **89**, 368–390.
Higdon, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic ocean. *Environmental and Ecological Statistics*, **5**, 173–190.
Huang, H.-C. & Cressie, N. (1996). Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Comp. Statist. Data Anal.*, **22**, 159–175.
Huang, H.-C. & Cressie, N. (2000). Multiscale graphical modeling in space: Applications to command and control. In *Proceedings of the Spatial Statistics Workshop*, Ed. M. Moore. New York: Springer Lecture Notes in Statistics, Springer Verlag Publishers.
Huang, H.-C., Cressie, N. & Gabrosek, J. (2002). Fast spatial prediction of global processes from satellite data. *Journal of Graphical and Computational Statistics*, **11**, 63–88.
Hughes, H.P. & Guttorp, P. (1994). Incorporating spatial dependence and atmospheric data in a model of precipitation. *Journal of Applied Meteorology*, **33**, 1503–1515.
Kaiser, J. (2001). An experiment for all seasons. *Science*, **293**, 624–627.
Kaiser, M. & Cressie, N. (2000). The Construction of Multivariate Distributions from Markov Random Fields. *Journal of Multivariate Analysis*, **73**, 199–220.
Knorr-Held, L. & Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine*, **17**, 2045–2060.
Le, N.D. & Zidek, J.V. (1992). Interpolation with uncertain spatial covariances: A Bayesian alternative to kriging. *Journal of Multivariate Analysis*, **43**, 351–374.
Mardia, K.V., Goodall, C.R., Redfern, E. & Alonso, F.J. (1998). The kriged Kalman filter (with Discussion). *Test*, **7**, 217–285.
Meiring, W., Guttorp, P. & Sampson, P.D. (1998). Space-time estimation of grid-cell hourly ozone levels for assessment of a deterministic model. *Envir. Ecol. Statist.*, **5**, 197–222.
Meyers, D.E. (1982). Matrix formulation of Co-kriging. *Mathematical Geology*, **14**, 249–257.
Mugglin, A.S. & Carlin, B.P. (1998). Hierarchical modeling in geographic information systems: Population interpolation over incompatible zones. *Journal of Agricultural, Biological, and Environmental Statistics*, **3**, 117–130.
Mugglin, A.S., Carlin, B.P., Zhu, L. & Conlon, E. (1999). Bayesian areal interpolation, estimation, and smoothing: an inferential approach for geographic information systems. *Environment and Planning, A*, **31**, 1337–1352.
Mugglin, A.S., Carlin, B.P. & Gelfand, A.E. (2000). Fully model-based approaches for spatially misaligned data. *Journal of the American Statistical Association*, **95**, 877–887.
Nychka, D., Wikle, C.K. & Royle, J.A. (1999). Large spatial prediction problems and nonstationary random fields. Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, CO, research report (`http://www.cgd.ucar.edu/stats/manuscripts/krig5.ps`).
Nychka, D., Wikle, C.K. & Royle, J.A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling: An International Journal*, **2**, 315–331.
Robert, C.P. & Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.
Royle, J.A. & Berliner, L.M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, **4**, 29–56.
Royle, J.A., Berliner, L.M., Wikle, C.K. & Milliff, R. (1999). A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador Sea. *Case Studies in Bayesian Statistics*, Eds. C. Gatsonis *et al.*, pp. 367–382. Springer-Verlag.
Rue, H. & Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, **29**, 31–49.
Sampson, P., Damian, D. & Guttorp, P. (2001). Advances in modeling and inference for environmental processes with nonstationary spatial covariance. NRCSE Technical Report Series, NRCSE-TRS No. 061.
Skellam, J.G. (1951). Random dispersal in theoretical populations. *Biometrika*, **38**, 196–218.
Smith, R.L. (1999). Bayesian and frequentist approaches to parametric predictive inference (with discussion). In *Bayesian Statistics 6*, Eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, pp. 589–612. Oxford University Press.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, B*, **64**, 583–639.
Stroud, J.R., Muller, P. & Sanso, B. (2001). Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society, B*, **63**, 673–689.
Ver Hoef, J.M. & Cressie, N. (1993). Multivariable spatial prediction. *Mathematical Geology*, **25**, 219–239.
Waller, L.A., Carlin, B.P., Xia, H. & Gelfand, A.E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, **92**, 607–617.
West, M. & Harrison, J. (1989). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.
Wikle, C.K. (2001). A kernel-based spectral approach for spatio-temporal dynamic models. *Proceedings of the 1st Spanish*

Workshop on Spatio-Temporal Modelling of Environmental Processes (METMA), Benicassim, Castellon (Spain), 28–31 October 2001, pp. 167–180.
Wikle, C.K. (2002a). A kernel-based spectral model for non-Gaussian spatio-temporal processes. *Statistical Modelling: An International Journal*, **2**, 299–314.
Wikle, C.K. (2002b). Spatial modeling of count data: A case study in modelling breeding bird survey data on large spatial domains. In *Spatial Cluster Modelling*, Eds. A. Lawson and D. Denison, pp. 199–209. Chapman and Hall.
Wikle, C.K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, **84**, to appear.
Wikle, C.K. & Berliner, L.M. (2002). Combining information across spatial scales. In review.
Wikle, C.K., Berliner, L.M. & Cressie, N. (1998). Hierarchical Bayesian space-time models. *J. Envir. Ecol. Stat.*, **5**, 117–154.
Wikle, C.K., Berliner, L.M. & Milliff, R.F. (2003). Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions. *Monthly Weather Review*, **131**, in press.
Wikle, C.K. & Cressie, N. (1999). A dimension reduced approach to space-time Kalman filtering. *Biometrika*, **86**, 815–829.
Wikle, C.K., Milliff, R.F., Nychka, D. & Berliner, L.M. (2001). Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association*, **96**, 382–397.

Résumé

Les systèmes environnementaux sont complexes. Ils incluent des processus spatio-temporels très complexes, interagissant sur une large variété d'échelles. Il existe des quantités de plus en plus grandes de données sur de tels processus, provenant des systèmes d'information géographiques, des plateformes de télédétection, des réseaux de surveillance et des modèles informatiques. De plus, il y a souvent une grande variété de connaissance scientifique disponible sur de tels systèmes, depuis les équations différentielles partielles jusqu'aux enquêtes de panels. Il est reconnu qu'il n'est généralement pas correct de considérer de tels processus d'une perspective commune. Au contraire, les processus doivent souvent être examinés comme des systèmes de modèles conditionnels liés de manière cohérente. Cet article fournit un bref aperçu des approches hiérachiques appliquées aux processus environnementaux. Les éléments clés de tels modèles peuvent être examinés à trois étapes principales: l'étape des données, celle du traitement et celle des paramètres. A chaque étape, la structure complexe de dépendance est atténuée par le conditionnement. Par exemple, le stade des données peut incorporer des erreurs de mesure ainsi que de multiples ensembles de données sous divers supports. Les stades du traitement et des paramètres peuvent admettre des processus spatiaux et spatio-temporels ainsi que l'inclusion directe du savoir scientifique. L'article conclut par une discussion de quelques problèmes en suspens dans la modélisation hiérarchique des systèmes environnementaux, incluant le besoin de nouvelles approches de collaboration.

[*Received January 2002, accepted November 2002*]