Spatial Statistical Modeling in Biology

C.K. Wikle, Department of Statistics, University of Missouri-Columbia, USA
J.A. Royle, U.S. Fish and Wildlife Service, Laurel, MD, USA

Keywords: Bayesian, Cokriging, Convolution, Generalized linear model, Generalized linear mixed model, GIS, GLMM, Hierarchical, Kriging, LMM, Maximum likelihood, Markov random fields, Markov chain Monte Carlo, MCMC, Mixed model, Poisson, REML, Spectral, Variogram

Contents

1. Introduction
2. Gaussian Random Process Models
   2.1 Linear Mixed Model Framework
   2.2 Covariance Models
   2.3 Estimation and Prediction
   2.4 Markov Random Field Spatial Models
3. Non-Gaussian Random Process Models
   3.1 Generalized Linear Mixed Model Framework
   3.2 Spatial GLMM Estimation and Prediction
   3.3 Spatial GLMM Example
4. Multivariate Spatial Models
   4.1 Cokriging
5. Spatiotemporal Models
6. Computation
   6.1 High-Level or Low-Level Language
   6.2 Bayesian Computation
   6.3 Geographic Information Systems
   6.4 Technology
7. Future Directions
   7.1 Nonstationary Spatial Processes
   7.2 Multivariate Non-Gaussian Formulations
Bibliography

Glossary

Bayesian Inference: A statistical philosophy based on Bayes' theorem, which acknowledges uncertainty about parameter values by assuming those parameters to be random variables with prior distributions. Inference is based on the conditional distribution of the parameters given the data, a quantity referred to as the posterior distribution.

BLUP: Best Linear Unbiased Predictor.

Cokriging: Kriging applied to 2 or more variables.

Gibbs Sampling: A type of MCMC algorithm.

Isotropy: A property of spatial processes that are rotationally invariant. If a process is not isotropic it is said to be anisotropic.

Kriging: Geostatistical term for best linear unbiased spatial prediction.

Lattice: A fixed network of spatial locations.

Stationary: The property of translation invariance of the moment structure of a random process.

Variogram: A description of spatial dependence, primarily for intrinsically stationary spatial processes.

MCMC: Markov Chain Monte Carlo; A Monte Carlo procedure that involves successive simulations from a Markov chain constructed in a fashion that permits the assertion that its stationary distribution coincides with the target distribution.

MLE: Maximum Likelihood Estimator (or Estimation)

REML: Restricted Maximum Likelihood.

Summary

Spatial modeling is increasingly prominent in the biological sciences as scientists attempt to characterize variability of processes that are spatially indexed. This article shows that the mixed model framework is useful for characterizing spatial statistical methodology. In particular, the classical geostatistical approach known as kriging can be cast as a linear mixed model. Furthermore, the generalized linear mixed model provides a natural framework for extending the methodology to allow modeling of non-Gaussian spatial processes. The mixed model framework is also useful for describing multivariate spatial models and many spatiotemporal models. These are discussed in the article as well as specific issues related to covariance modeling, estimation and prediction, computation, and Bayesian methodologies.

1 Introduction

Since its inception as a discipline, statistics has provided tools by which scientists can better understand complex processes. Primarily this is because statistics is concerned with the study of variability, and all natural processes exhibit variability. As scientists seek to answer ever more challenging questions concerning processes that vary over space, the traditional statistics methods that one might learn in introductory statistics courses are not sufficient to adequately account for this variability. However, at least in principle, relatively simple extensions to simple statistical concepts such as linear models and regression, provide the foundation for basic spatial statistical analysis.

Although not universally true, objects in close proximity are more alike. Consequently, one must include the effects of spatial proximity when performing statistical inference on such processes, or at least show that there is no need to do so. Including these spatial effects is important for efficient estimation of parameters, prediction, and the design of sampling networks. As a simple illustration, consider some spatial process, denoted by Y, at three locations, A, B, C such that A and B are very close together in space (i.e., adjacent plots in a field trial) and C is widely separated from both A and B. Assume the spatial process has zero mean and variance σ^2 at all spatial locations. It is then the case that $\text{var}[Y(A) - Y(B)] = 2\sigma^2 - 2\text{cov}[Y(A), Y(B)]$ and $\text{var}[Y(A) - Y(C)] = 2\sigma^2 - 2\text{cov}[Z(A), Z(C)]$. If the covariance is positive and decreases with distance (that is, things close together are more alike), then $\text{cov}[Y(A), Y(B)] > \text{cov}[Y(A), Y(C)]$ and thus $\text{var}[Y(A) - Y(B)] < \text{var}[Y(A) - Y(C)]$. Clearly, inference on the differences should include the effects of the spatial dependence. Such effects of spatial dependence in statistical inference have been known for a very long time. In fact, one of the arguments in favor of randomization for agricultural field trials is to mitigate the effects of such dependence. However, in many environmental and biological applications, one typically considers observational studies in which randomization is not a viable option. It is in these situations that one seeks to model the spatial dependence through the use of random field models.

The later twentieth century and beginning of the twenty-first century has seen a tremendous growth in spatial statistical methodological development and application. This is primarily a function of the rapid progression of computational technology, hardware, software and algorithmic, and the need to solve challenging problems. The corresponding propagation of Bayesian methodology into mainstream statistics has been responsible for a sizeable portion of this development. Although there is still an undercurrent of tension between traditional frequentist proponents and Bayesian

proponents, most practicing statisticians recognize the advantages and disadvantages of both views and approach spatial modeling from a pragmatic perspective, using the methodology appropriate for the given problem. Thus, in this article, we consider a broad range of methodologies from both traditional and Bayesian perspectives.

Kriging and its derivatives constitute the most common class of spatial models used in diverse disciplines such as crop and soil science, geology, atmospheric science, and more recently in ecology and the biological sciences. Many software packages have "kriging" routines, and kriging is the core of many contemporary graduate level courses on spatial statistics. Much of the terminology common in spatial statistics today first arose within the field of geostatistics.

Kriging can be viewed as arising under a linear mixed model (LMM), which have been intensively studied and have a well developed theory. Thus, understanding the basics of conventional mixed models is helpful for understanding spatial statistical models. In fact, it can be argued that the LMM perspective is natural since LMM's are widely used in biological, medical and epidemiological fields, particular in relation to *longitudinal* data, of which which spatial data are a special case. When viewed from a LMM perspective, estimation and prediction of spatially correlated processes poses no additional complexity beyond that required for LMMs. This is in contrast to conventional developments of kriging, where these problems were derived independently, essentially outside of the field of statistics. Because of this, many ad hoc procedures exist within the geostatistical paradigm, and the terminology is cluttered with jargon. Conversely, linear mixed models are well-known to both statisticians and practitioners of statistics alike, and so this formulation is often simpler as an introductory framework. An additional benefit of the LMM development is that extension to non-Gaussian problems is straightforward by way of the generalized linear mixed model (GLMM) extension of the normal, linear case. The discipline of disease mapping makes widespread use of GLMMs within a spatial modeling context.

This article focuses on Gaussian spatial models as considered from a LMM approach, along with classical and Bayesian estimation issues. An important part of such modeling is related to the specification of realistic covariance structures, and so a discussion of this topic is included as well. We consider extensions to non-Gaussian spatial models, as well as multivariate and spatio-temporal processes. Finally, we present some topics in which there is substantial current research interest. Computational issues will be discussed as they arise.

2 Gaussian Random Process Models

Consider a spatial process $Y(\mathbf{s})$ where $\mathbf{s} \in D$, some domain in d-dimensional Euclidean space. In this article, we will only consider two-dimensional spatial processes. Furthermore, we assume that the process $Y(\mathbf{s})$ has a Gaussian (normal) distribution with mean $\mu(\mathbf{s})$ and is correlated so that $c_y(\mathbf{s}, \mathbf{s}') \equiv \text{cov}[Y(\mathbf{s}), Y(\mathbf{s}')]$ for some $\mathbf{s}, \mathbf{s}' \in D$ where $\mathbf{s} \neq \mathbf{s}'$. We refer to such a process as a Gaussian random process or Gaussian random field.

2.1 Linear Mixed Model Framework

The classical linear mixed model generalizes the traditional linear model to include random effects. In the present context, we will equate the random effect to a correlated spatial process. A common statement of the LMM is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{H}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of responses, $\mathbf{X}$ and $\mathbf{H}$ are known matrices of independent, explanatory, or regression variables ($n \times p$ and $n \times q$, respectively), $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients or *fixed effects*, and $\boldsymbol{\alpha}$ and $\boldsymbol{\epsilon}$ are $q \times 1$ and $n \times 1$ random vectors, respectively. Typically, columns of $\mathbf{H}$ are *indicator* variables, so that each observation is associated with a particular element of $\boldsymbol{\alpha}$. The usual assumption on these random effects is multivariate normality:

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim \text{Gau}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_\alpha & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\epsilon \end{bmatrix} \right).$$

In many statistical problems, including spatial statistics, one often assumes independence of the random errors, in which case $\boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2 \mathbf{I}_{n\times n}$, where $\mathbf{I}_{n\times n}$ is the n-dimensional identity matrix. The variance component σ_ϵ^2 is *measurement error variance*, and may additionally include effects of small-scale spatial variability – that is, anything unexplained by the random effect. In the field of geostatistics, σ_ϵ^2 is called the "nugget effect".

The application of this model to spatial settings is straightforward. Suppose that the response vector is spatially indexed, so that $\mathbf{y} = [y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)]'$ for spatial locations $\mathbf{s}_i, i = 1, \ldots, n$. Let the elements of $\boldsymbol{\alpha}$ represent "spatial effects", then $\boldsymbol{\Sigma}_\alpha$ is a $q \times q$ *spatial* covariance matrix where q is the number of spatial locations. In many spatial statistical problems $q = n$; i.e., there is a single response observation at each site, in which case $\mathbf{H} = \mathbf{I}_{n\times n}$. This is the essence of the model used in conventional

kriging applications. The other consideration in the context of spatial applications is that prediction of "unobserved" data is of primary interest. This is in contrast to most mixed-model applications, where the primary interest is in estimation of the vectors $\boldsymbol{\beta}$, perhaps the variance components σ_α^2 and σ_ϵ^2, and, to a lesser extent, $\boldsymbol{\alpha}$.

For the model (1) note that $\boldsymbol{\alpha} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$ and $\boldsymbol{\epsilon} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$. One can think of this model hierarchically, as

$$\mathbf{y}|\boldsymbol{\alpha} \sim \text{Gau}(\mathbf{X}\boldsymbol{\beta} + \mathbf{H}\boldsymbol{\alpha}, \boldsymbol{\Sigma}_\epsilon) \tag{2}$$

$$\boldsymbol{\alpha} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha). \tag{3}$$

The joint distribution is given by $f(\mathbf{y}, \boldsymbol{\alpha}) = f(\mathbf{y}|\boldsymbol{\alpha})f(\boldsymbol{\alpha})$. One can obtain the marginal distribution for $\mathbf{y}$ by integrating out the random effects, $f(\mathbf{y}) = \int f(\mathbf{y}|\boldsymbol{\alpha})f(\boldsymbol{\alpha})d\boldsymbol{\alpha}$, which is easily shown in this case to be

$$\mathbf{y} \sim \text{Gau}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_\alpha + \boldsymbol{\Sigma}_\epsilon). \tag{4}$$

Thus, the marginal model (4) follows from the hierarchical formulation (2) and (3). In traditional LMM applications (e.g., longitudinal analysis) it is often convenient (and arguably more general) to proceed in terms of the marginal model, without the need for specific inference or estimation concerning the random effects. That is, one accounts for spatial dependence but is not interested in the underlying process that generates such dependence. However, for most traditional spatial applications, one is interested in performing inference on the random effects (spatial process) and the hierarchical formulation is more appropriate.

One might think of $\boldsymbol{\alpha}$ as discrete levels of a random factor, but it is more often regarded, in spatial problems, as a *spatial random process*. Instead of the discrete multivariate normal model specification given in (1), one might instead write $y(\mathbf{s}) = \sum_j x_j(\mathbf{s})\beta_j + \alpha(\mathbf{s}) + \epsilon(\mathbf{s})$ and $\alpha(\mathbf{s}) \sim \text{Gau}(0, \sigma_\alpha^2)$, with $\text{cov}[\alpha(\mathbf{s}), \alpha(\mathbf{s}')] = \sigma_\alpha^2 r_\alpha(\mathbf{s}, \mathbf{s}')$, for some correlation function $r_\alpha()$, thus relating *observed* "levels" of the random process y with *unobserved* levels (those values which we wish to predict) of the spatial random effects process, a subtle but important aspect with regard to spatial problems. The distinction between the more traditional, vector representation (common in statistics), and the "process" representation (common in geosciences) is more a matter of tradition.

2.2 Covariance models

The essence of spatial statistics is spatial correlation, and consequently it is important to model this aspect of the problem adequately. Unfortunately, there are many limitations (having to do with both data and covariance models) which make this a difficult task. To guarantee that the covariance matrix is positive definite, the spatial covariance matrix $\mathbf{\Sigma}_\alpha$ is assumed to be of some parametric form, indexed by the parameter θ (possibly a vector). To be more precise, the spatial covariance matrix is expressed as $\mathbf{\Sigma}_\alpha(\theta)$. Much of the detail concerning *implementation* of contemporary spatial statistics focuses on the choice of the covariance function, and estimation of its parameters.

The covariance function describes the spatial association between the random effect at any two locations in space, say $\mathbf{s}$ and $\mathbf{s}'$:

$$\text{cov}[\alpha(\mathbf{s}), \alpha(\mathbf{s}')] = c_\alpha(\mathbf{s}, \mathbf{s}'; \theta).$$

If the variance of α is homogeneous, we may write $c_\alpha(\mathbf{s}, \mathbf{s}') = \sigma_\alpha^2 r_\alpha(\mathbf{s}, \mathbf{s}'; \theta)$ where $r_\alpha()$ is the *correlation function*, being scaled by the variance component σ_α^2. Since elements of $\boldsymbol{\alpha}$ are indexed by space, the covariance function allows one to "fill-in" the elements of $\mathbf{\Sigma}_\alpha(\theta)$. Thus, given c_α, spatial parameters θ, and any two locations in space, $\mathbf{s}$ and $\mathbf{s}'$ (sample locations, or not), the covariance between $\alpha(\mathbf{s})$ and $\alpha(\mathbf{s}')$ may be determined.

Typically, assumptions are imposed on the process to facilitate estimation of parameters (as will be discussed further below), but also because there is a severe shortage of more general covariance models. The two usual assumptions are *second-order stationarity* and *isotropy*, the former being *translation invariance* of the second-moment structure of α, and the latter being *rotation invariance*. Normally the stationarity assumption would imply a similar constraint on the first moment structure, but we have assumed α to have mean 0, accommodating any mean nonstationarity in $\mathbf{X}\boldsymbol{\beta}$. Thus, under these assumptions, the covariance between any two points is only a function of the *distance* separating them:

$$\text{cov}[\alpha(\mathbf{s}), \alpha(\mathbf{s}')] = \sigma_\alpha^2 r_\alpha(||\mathbf{s} - \mathbf{s}'||; \theta)$$

where $||\mathbf{s} - \mathbf{s}'||$ is the distance between points $\mathbf{s}$ and $\mathbf{s}'$, say Euclidean, geographic distance, etc.. This simplified correlation structure conveniently dictates the covariance between observed and unobserved values of y for which predictions are desired, a quantity required to formulate the predictor (discussed below). A common correlation

model is the exponential model given by

$$r_\alpha(||\mathbf{s}-\mathbf{s}'||;\theta) = \exp(\frac{-||\mathbf{s}-\mathbf{s}'||}{\theta}).$$

There are many other models in widespread use.

In geostatistics, spatial modeling is often considered in terms of the *variogram* rather than the covariance. The variogram is an alternative description of spatial dependence and is defined as:

$$2\gamma(\mathbf{h}) \equiv \mathrm{var}[\alpha(\mathbf{s}+\mathbf{h}) - \alpha(\mathbf{s})], \tag{5}$$

for all $\mathbf{s}, \mathbf{s}+\mathbf{h} \in D$, where D is the spatial domain of interest and $\mathbf{h}$ is some spatial lag. Note that the variogram must satisfy the condition of conditional-negative semidefiniteness to guarantee that all model-based variances are nonnegative. As with the covariance function, when the process is isotropic, the variogram is only a function of distance, so $2\gamma(\mathbf{h}) = 2\gamma(||\mathbf{h}||)$. Under the condition (5) and if the mean is constant, the process is said to be *intrinsically stationary.* The class of processes that are second-order stationary are a subset of those that are intrinsically stationary. In fact, if the process is second-order stationary, then there is a simple relationship between the variogram and the covariance function, $2\gamma(\mathbf{h}) = 2[c(\mathbf{0}) - c(\mathbf{h})]$. In general, the condition of second-order stationarity is sufficient for most processes of concern to the biological and environmental sciences. Thus, the remainder of this article assumes second-order stationarity.

2.3 Estimation and Prediction

A common method of parameter estimation under the mixed model is maximum likelihood estimation from the marginal distribution of $\mathbf{y}$. Under the simplifications discussed above, the marginal distribution of $\mathbf{y}$ is:

$$\mathbf{y} \sim \mathrm{Gau}(\mathbf{X}\boldsymbol{\beta}, \sigma_\alpha^2\mathbf{R}_\theta + \sigma_\epsilon^2\mathbf{I}).$$

Letting $\boldsymbol{\lambda}$ be the vector of second-moment parameters, $(\sigma_\alpha^2, \sigma_\epsilon^2, \theta)$, and setting $\boldsymbol{\Sigma}(\lambda) = \sigma_\alpha^2\mathbf{R}_\theta + \sigma_\epsilon^2\mathbf{I}$, the log likelihood is:

$$l(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\lambda}) = -\frac{1}{2}\log(|\boldsymbol{\Sigma}(\lambda)|) - \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}(\lambda)^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}).$$

To obtain parameter estimates for $\boldsymbol{\lambda}$ we can maximize this marginal log-likelihood with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$. The likelihood is nonlinear in the parameters and thus

must be maximized numerically. The use of MLEs in spatial problems has been questioned by some practitioners since MLEs are generally biased for the (co)variance parameters. Consequently, *restricted maximum likelihood estimation* (REML) is often recommended. This leads to less bias in plug-in kriging variances.

If $\boldsymbol{\lambda}$ is known, the MLE of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}\boldsymbol{\Sigma}(\boldsymbol{\lambda})^{-1}\mathbf{X}]^{-1}\mathbf{X}'\boldsymbol{\Sigma}(\boldsymbol{\lambda})^{-1}\mathbf{y}, \tag{6}$$

the generalized least squares (GLS) estimate. In practice, one uses an estimate, say $\hat{\boldsymbol{\lambda}}$, in place of $\boldsymbol{\lambda}$ producing:

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}\boldsymbol{\Sigma}(\hat{\boldsymbol{\lambda}})^{-1}\mathbf{X}]^{-1}\mathbf{X}'\boldsymbol{\Sigma}(\hat{\boldsymbol{\lambda}})^{-1}\mathbf{y}. \tag{7}$$

Both estimation of the random effects (i.e., those which coincide with data locations), and prediction of "unobserved" random effects, is based on the Best Linear Unbiased Predictor (BLUP). One can derive the BLUP from several perspectives, including "distribution-free", Bayesian, and multivariate normal methods. All of these derivations either implicitly or explicitly assume that only first and second moment properties of the underlying spatial distribution are sufficient to describe the distribution.

For estimating a vector of random effects, say say $\boldsymbol{\alpha}_{pred}$, which may include random effects corresponding to both sample locations and locations for which predictions are desired, one must compute the quantities $\boldsymbol{\Omega} = \text{cov}[\boldsymbol{\alpha}_{pred}, \boldsymbol{\alpha}; \boldsymbol{\lambda}]$ and $\boldsymbol{\Sigma}_{pred} = \text{var}[\boldsymbol{\alpha}_{pred}; \boldsymbol{\lambda}]$ which requires $\boldsymbol{\lambda}$. The BLUP for known $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ is

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\Omega}\boldsymbol{\Sigma}(\lambda)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{8}$$

When the vector to be predicted are those random effects corresponding to data locations (i.e., $\boldsymbol{\alpha}_{pred} = \boldsymbol{\alpha}$), then $\boldsymbol{\Omega} = \sigma_\alpha^2\mathbf{R}_\theta$. In general, the prediction variance is

$$\text{var}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = \boldsymbol{\Sigma}_{pred} - \boldsymbol{\Omega}\boldsymbol{\Sigma}(\lambda)^{-1}\boldsymbol{\Omega}'. \tag{9}$$

Although the expressions for the predictor and its variance for known $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are convenient, these quantities are seldom known. In that case, one may plug $\hat{\boldsymbol{\beta}}$ into (8), and it remains the BLUP; the prediction variance must be adjusted accordingly to account for the uncertainty in estimation of $\boldsymbol{\beta}$. See the references for details. Accounting for unknown $\boldsymbol{\lambda}$ is much more difficult. In practice, the estimates obtained from MLE (or other approaches) are simply plugged into the known-$\boldsymbol{\lambda}$ expressions. This is the so-called "plug-in" predictor, or the estimated BLUP (EBLUP). Unfortu-

nately, the variance computed from an analogous plug-in procedure is not, in fact, the correct variance of the EBLUP. In that case, the spatial prediction variance (9) underestimates the true variability in the predictions since it does not take into account the variability introduced by the estimation of the (co)variance parameters. Since the predictor depends on $\boldsymbol{\lambda}$ in a nonlinear fashion, accounting for uncertainty due to its estimation is problematic. Various approaches have been suggested to mitigate this concern (bootstrapping, empirical Bayesian, and Bayesian methods).

2.3.1 Bayesian Estimation

The approach outlined above, in which estimation of $\boldsymbol{\lambda}$ is based on the marginal likelihood, and those estimates are plugged into the known-$\boldsymbol{\lambda}$ expressions for the BLUP and its variance, is essentially what is known as an *empirical Bayesian* approach. More formal Bayesian approaches may be considered, and we believe them to have advantages for prediction in that uncertainty in covariance parameter estimation is more easily accounted for.

Although a complete discussion of Bayesian analysis is beyond the scope of this article, the general idea is quite simple in principle. Given a likelihood $f(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\lambda})$ and prior $\pi(\boldsymbol{\alpha}, \boldsymbol{\lambda})$, Bayesian inference is based on the *posterior* distribution $f(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\mathbf{y})$. Various marginal posterior distributions, that is distributions of $\alpha_j|\mathbf{y}$ or $\lambda_i|\mathbf{y}$, are typically used to make inferences about individual parameters or unknowns. We note that further flexibility is obtained if we consider a fully hierarchical representation. In that case, one notes that the joint distribution can be decomposed $f(\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = f(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\lambda})\pi(\boldsymbol{\alpha}|\boldsymbol{\lambda})\pi(\boldsymbol{\lambda})$. Thus, one can apply Bayes rule,

$$f(\boldsymbol{\alpha}, \boldsymbol{\lambda}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\alpha}, \boldsymbol{\lambda})\pi(\boldsymbol{\alpha}|\boldsymbol{\lambda})\pi(\boldsymbol{\lambda}), \tag{10}$$

where the normalizing constant is the integral of $f(\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$. Unfortunately, the posterior distribution is often sufficiently complex so as to prevent direct analysis (i.e., the normalizing constant integral cannot be solved analytically). Instead, various methods have been developed which permit simulation from the posterior distribution. Then, simulated values of the unknowns are used to estimate the quantities of interest. One set of simulation tools falls under the category of Markov Chain Monte Carlo (MCMC) and includes such algorithms as Metropolis-Hastings and the Gibbs sampler. We note that a fully Bayesian analysis might also specify a prior distribution for $\boldsymbol{\beta}$ as well. The advantage of the fully Bayesian approach to LMM relative to the marginal approach is that it accounts for the variability of all

parameters in the model. The disadvantage is the extra computational burden that may be necessary to obtain estimates and predictions.

2.4 Markov Random Field Spatial Models

In the previous discussion, it has been assumed that the spatial process can occur at any spatial location in some two-dimensional Euclidean space, a continuous region. Special classes of spatial models known as Markov random fields (MRFs) have been developed to account for the situation where the set of all possible spatial locations is discrete (countable). We sometimes say that the collection of such sites is a *lattice*. Examples include mapping disease 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 often 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 $P[y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)]$. 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

$$P[y(\mathbf{s}_i)|\{y(\mathbf{s}_j) : j \neq i\}], \quad i = 1, \ldots, n. \tag{11}$$

We then define the neighborhood N_i of the i-th site as the collection of locations such that

$$P[y(\mathbf{s}_i)|\{y(\mathbf{s}_j) : j \neq i\}] = P[y(\mathbf{s}_i)|\{y(\mathbf{s}_j) : j \in N_i\}], \quad i = 1, \ldots, n. \tag{12}$$

That is, the the conditional probability at site i only depends 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. 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.

2.4.1 Gaussian Markov Random Field Model

A natural model when the spatial process $y(\ )$ is continuous is Gaussian MRF model. The conditional models (12) are Gaussian with,

$$P[y(\mathbf{s}_i)|y(N_i)] = \frac{1}{\sqrt{2\pi}\sigma_i} \exp[-\frac{1}{2\sigma_i^2}\{y(\mathbf{s}_i) - \mu(\mathbf{s}_i) - \sum_{j \in N_i} c_{ij}[y(\mathbf{s}_j) - \mu(\mathbf{s}_j)]\}^2], \tag{13}$$

where $c_{ij}\sigma_j^2 = c_{ji}\sigma_i^2$, $c_{ii} = 0$, $c_{ik} = 0$ for k not in N_i, 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

$$\mathbf{y} \sim \text{Gau}[\boldsymbol{\mu}, (\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}], \tag{14}$$

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}, and $\mathbf{M}$ is a diagonal matrix with $\sigma_1^2, \ldots, \sigma_n^2$ on the main diagonal.

The Gaussian MRF can easily be incorporated in the LMM framework described previously. For example, let $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha} + \boldsymbol{\epsilon}$, where now $\boldsymbol{\alpha} \sim \text{Gau}(\mathbf{0}, (\mathbf{I} - \mathbf{C})^{-1}\mathbf{M})$, a MRF with zero mean. The parameters of the LMM covariance structure $\boldsymbol{\lambda}$ are now the elements of $\boldsymbol{C}$ and $\mathbf{M}$. In practice, the neighborhood structure of the MRF is often simplified greatly so that the number of unknown parameters is reasonable 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.

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 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{H}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$ where $\mathbf{H}$ associates each datum with one of the discrete strata. Thus, MRFs may be useful regardless of whether or not the data actually have discrete spatial support.

We note that for the Gaussian MRF model, the spatial dependence is specified through the conditional distributions (13) and ultimately the *inverse* covariance matrix $(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}$. This is in contrast to the continuous Gaussian random field case where one specifies the covariance matrix directly (not its inverse). There are 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 make use of the conditional distributions in efficient MCMC algorithms.

3 Non-Gaussian Random Process Models

The linear mixed model presented above may be considered free of formal distributional assumptions (which is the usual kriging development). That is, first and second moment assumptions alone lead to the same estimators and predictors for parameters and random effects as if the processes were assumed to be Gaussian. Thus, such

approaches may be viewed as having an implicit assumption of normality.

Such informality leads to ambiguity in applying the basic linear mixed model approach when there is a strong belief as to the distribution of the random variable under study, other than Gaussian. For example, count data are common in many biological, ecological, and epidemiological problems. In such problems, there are usually discernible mean-variance relationships, and the variable is clearly discrete and positive valued. It is natural to consider Poisson and Binomial models in these situations, and treatment in the usual LMM/kriging context will lead to some inefficiency in both estimation and prediction. Similarly, many problems in the atmospheric and environmental sciences involve positive-valued variables, often right-skewed, again with strong mean-variance relationships. In this case, the log-normal distribution is a natural candidate for a model-based analysis, though other possibilities exist (e.g., Gamma).

3.1 Generalized Mixed Model Framework

Non-Gaussian spatial problems may be formally analyzed within the context of generalized linear mixed models (GLMM). Formulation of a GLMM requires specification of the likelihood of the random variable $y(\mathbf{s})$. As in classical generalized linear models (GLMs), there is a so-called *canonical parameter* corresponding to the distribution, which is nominally a function $g(\ )$ (the *link function*) of the location parameter for the distribution. It is this quantity that is assumed to be linear in the explanatory variables. In the classical formulation of GLMs containing only fixed effects, $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is the matrix of explanatory variables. To incorporate a spatial process, we assume $y(\mathbf{s}_i|\boldsymbol{\alpha})$ is conditionally independent for any location $\mathbf{s}_i$ with conditional mean $E[y(\mathbf{s}_i)|\boldsymbol{\alpha}] = \mu(\mathbf{s}_i)$. Then, the spatially correlated random effect is incorporated into the linear predictor:

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{H}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \tag{15}$$

where the additional noise term $\boldsymbol{\epsilon}$ may or may not be included, depending on the application. Nominally, such an error term accommodates *over-dispersion* relative to the mean-variance relationship implied by the distribution under consideration. As before, $\boldsymbol{\alpha} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}))$ and $\boldsymbol{\epsilon} \sim \text{Gau}(\mathbf{0}, \sigma_\epsilon^2\mathbf{I})$, with spatial correlation parameterized by $\boldsymbol{\theta}$ in $\boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta})$.

3.2 Spatial GLMM Estimation and Prediction

As with the LMM and GLM, one would like to use maximum likelihood methods to obtain parameter estimates. In the GLMM case, high-dimensional integration is generally required to evaluate the likelihood – the random effects must be integrated from the conditional likelihood $f(\mathbf{y}|\boldsymbol{\alpha})$. Although in simple random effects models, this is possible using standard numerical procedures, it is not practical for spatial problems.

Due to the difficulties of implementing traditional likelihood methods, other approaches have been considered. These include generalized estimating equations (GEE), penalized quasi-likelihood methods, and conditional likelihood methods. For the most part, these methods are essentially distribution free and do not require explicitly specification of the covariance structure of the random effects (spatial process). Although in many cases such methods work well for getting unbiased (or nearly so) and reasonably efficient parameter estimates, they typically are not as useful for prediction of the random effects. This is problematic in spatial data problems, where spatial prediction is often the primary interest. For that reason, the preference in spatial problems has traditionally been Bayesian estimation. As discussed previously, the popularization of MCMC methods has allowed the implementation of such models. The Bayesian approach for GLMMs is illustrated in the following example.

3.2.1 Spatial GLMM Example: Mapping Bird Counts

Many biological studies give rise to spatially indexed count data, typically involving counts of organisms at a collection of survey locations. Objectives are varied, but may simply involve mapping abundance, or assessing the relationship of abundance to covariates which describe habitat and landscape structure.

We consider data collected as part of the annual North American Breeding Bird Survey (BBS). This survey is conducted in May-June of each year. Volunteer observers traverse a roadside sampling route containing 50 stops. At each stop, the observer records the number of birds (by species) seen and heard. There are several thousand BBS routes in North America. We focus on mourning dove (*Zenaida macroura*) counts from 103 routes in the state of Pennsylvania. Let $y(\mathbf{s})$ be the total count of doves on the BBS route centered at $\mathbf{s}$ (aggregated over the 50 stops). Our goal is to produce a map of dove relative abundance within the state.

For large counts, one might consider applying standard Gaussian spatial process models (i.e. kriging). However, this ignores the positive support of $y(\mathbf{s})$, and the

fact that BBS data typically exhibit strong mean-variance relationships common in count data, such as might be explained by a Poisson distribution assumption on $y(\mathbf{s})$. A transformation can be used to account for the latter phenomenon. Instead of the Gaussian kriging model applied to $y(\mathbf{s})$, one might apply that model to $\log(y(\mathbf{s}))$ (so-called log-normal kriging). With this approach however, zeros are problematic, and the more appropriate Poisson mean-variance linkage is disregarded. Consequently, modeling and prediction within the Poisson framework is preferred.

Following the GLMM framework described previously, we first specify a model for $y(\mathbf{s})$ conditional on the Poisson mean $\lambda(\mathbf{s})$:

$$y(\mathbf{s})|\lambda(\mathbf{s}) \sim \text{Poisson}(\lambda(\mathbf{s})).$$

Then, we employ a Gaussian spatial process model to describe spatial variation in $\lambda(\mathbf{s})$. The log-link is the canonical link for the Poisson distribution. Thus, defining $u(\mathbf{s}) \equiv \log(\lambda(\mathbf{s}))$, we specify a normal model for $u(\mathbf{s})$ conditional on the spatial process $\alpha(\mathbf{s})$:

$$u(\mathbf{s})|\alpha(\mathbf{s}) \sim \text{Gau}(\mu + \alpha(\mathbf{s}), \sigma_\epsilon^2).$$

Here, μ is the mean, which may be generalized to accommodate interesting covariates, and σ_ϵ^2 serves to account for over-dispersion relative to the Poisson assumption. This over-dispersion arises from several sources – notably, "observation error" (i.e. not all observers are equally competent) and "small-scale variation", such as may arise from local (i.e. within-route) variation in habitat structure and so forth. Finally, we specify $\alpha(\mathbf{s})$ to be Gaussian with mean 0, variance σ_α^2, and correlation function $r_\alpha(\mathbf{s}, \mathbf{s}'; \theta)$. As usual, one may consider many correlation models. We specify $r_\alpha(\mathbf{s}, \mathbf{s}'; \theta)$ to be the single parameter exponential model: $r_\alpha(\mathbf{s}, \mathbf{s}'; \theta) = \exp(-||\mathbf{s} - \mathbf{s}'||/\theta)$ where θ controls the rate of decay in correlation as distance between sites increases.

Estimation of the parameters μ, σ^2, σ_α^2, and θ, and prediction of $\alpha(\mathbf{s})$ at sampled and/or unsampled routes is easily carried out within a Bayesian framework using MCMC techniques. We first must specify prior distributions for the parameters μ, σ^2, σ_α^2, and λ with prior distributions. We used conventional "flat" priors for this purpose with the exception of θ. For this parameter, we specified a discrete uniform distribution on 400 equally spaced values in $(0, 8)$ on $\log(\theta)$. Other details, including those on implementation of the MCMC algorithm may be found in the references.

A map of $\alpha(\mathbf{s})$ was created by estimating the posterior mean on a grid of points over the region (Figure 1, top panel). Uncertainty is quantified with the posterior standard deviations (not shown). Of more general interest are the corresponding

quantities on the natural count scale. That is, the expected number of birds, $\lambda(\mathbf{s})$, and its standard deviation. Estimation of these quantities is easily enough accomplished within the MCMC framework. Maps of both quantities are shown in Figure 1 (middle and bottom panels). The dominant feature apparent in these maps is the strong relationship between the predicted mean, and its standard deviation – a consequence of modeling within the Poisson framework.

4 Multivariate Spatial Models

In environmental and biological studies, it is seldom the case that measurements are made on a single variable. For example, most air pollution monitoring networks collect data on several pollutants in addition to relevant covariates such as temperature and precipitation. The breeding bird survey described in the previous example produces counts on over 200 species of birds, many of which have similar habitat and resource requirements. It stands to reason that relationships exist among these variables, and these relationships should manifest themselves in the joint spatial structure of the variables.

The most obvious benefit of having multiple spatial variables is the use of a covariate to aid in prediction of a primary variable. For example, temperature is informative about ozone concentration. At the same time, temperature data is more abundant and cheaper to collect. Consequently, more efficient prediction of ozone concentration may be achieved by inclusion of the relationship between ozone and temperature within a bivariate spatial model. In the context of bird monitoring, some species are difficult to observe and thus it is difficult to construct precise maps of abundance without the aid of additional information. Habitat data, or information on the abundance of similar species can serve this purpose.

The need to incorporate inter-dependence among two or more spatial variables gives rise to several strategies for modeling spatial dependence. For the most part, these are straightforward extensions of univariate methods and thus, much of that material applies in the multivariate setting.

4.1 Cokriging

The most common multivariate procedure, often referred to as *cokriging*, involves direct specification of the joint variance-covariance structure among a set of variables. In the bivariate case, with variables $y_1(\mathbf{s})$ and $y_2(\mathbf{s})$, the cokriging model can be

formulated as a joint-normality assumption on the response vectors $\mathbf{y}_1$ and $\mathbf{y}_2$:

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim \text{Gau}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_2 \end{pmatrix} \right).$$

Here, the mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ may be related to regression variables in the usual manner. As before, the marginal covariance matrices are defined as $\boldsymbol{\Sigma}_1 = \sigma_1^2 \mathbf{R}_{\theta_1}$ and $\boldsymbol{\Sigma}_2 = \sigma_2^2 \mathbf{R}_{\theta_2}$, where the correlation matrices $\mathbf{R}_{\theta_1}$ and $\mathbf{R}_{\theta_2}$ are parameterized by correlation functions $r_1(\mathbf{s}, \mathbf{s}'; \theta_1)$ and $r_2(\mathbf{s}, \mathbf{s}'; \theta_2)$. These may be chosen to be of the same parametric form, with different parameters, or of different parametric forms. Similarly, the cross-covariance matrix $\boldsymbol{\Sigma}_{12} = \sigma_{12} \mathbf{R}_{\theta_{12}}$, where σ_{12} is the covariance between $y_1(\mathbf{s})$ and $y_2(\mathbf{s})$ and $\mathbf{R}_{\theta_{12}}$ is the cross-correlation matrix. The key to the cokriging model is specification of a cross-correlation function $\text{corr}[y_1(\mathbf{s}), y_2(\mathbf{s}')] = r_{12}(\mathbf{s}, \mathbf{s}'; \theta_{12})$ used to parameterize the matrix $\mathbf{R}_{\theta_{12}}$.

In general, estimation and prediction under this model proceeds as in the univariate case. That is, parameter estimation based on MLE (or other methods) from the joint model, and then use of a plug-in procedure where these estimates are substituted into the expressions for the BLUP and its variance. Regarding $\mathbf{y} = [\mathbf{y}_1', \mathbf{y}_2']'$ as the data vector, the formulae and associated discussion given in Section 2.3 apply here. Note however, that we have not given the "conditional on $\boldsymbol{\alpha}$" development of cokriging here, as the joint distribution specification is conventional. However, the expression for the BLUP of $\boldsymbol{\alpha}_{pred}$ in Section 2.3 is equivalent to that for predicting values of $\mathbf{y}$.

There is some practical difficulty in specifying the marginal and cross-correlation functions so that the resulting joint variance-covariance matrix is valid in the sense of being positive-definite. Imposing this constraint on arbitrary functions and their estimates is not a simple matter. Indeed, in many applications the correlation functions appear to be estimated more or less independently of one another, which does not guarantee that the resulting estimates are valid. Occasionally, the ad hoc practice of *a posteriori* "checking" that resulting estimates yield positive definite covariance matrices is done. More formal approaches such as specifying the covariance functions to be scaled versions of the same basic model, forming linear combinations of multiple covariance functions, where the coefficients satisfy certain constraints, and even nonparametric approaches, exist to remedy this difficulty.

4.2 Hierarchical Models

Notwithstanding the difficulty in constructing valid joint variance-covariance models, cokriging becomes unwieldy when several variables are considered. In addition, specification of the various cross-correlation functions is awkward and typically little scientific basis exists to guide construction of the joint second-moment model. Cokriging is essentially an empirical approach. Alternatively, one may construct a joint model for two or more variables by construction of a sequence of conditional models (the hierarchical approach). Sticking to the bivariate case for simplicity, a simple hierarchical specification of the bivariate model consists of the following two components:

$$\mathbf{y}_1|\mathbf{y}_2 \sim \text{Gau}(\boldsymbol{\mu}_1 + \beta\mathbf{y}_2, \sigma^2_{1.2}\mathbf{R}_{\theta_{1.2}})$$

and

$$\mathbf{y}_2 \sim \text{Gau}(\boldsymbol{\mu}_2, \sigma^2_2\mathbf{R}_{\theta_2}).$$

As usual, one must specify correlation models for $\mathbf{R}_{\theta_{1.2}}$ and $\mathbf{R}_{\theta_2}$, but there is no cross-correlation term. Instead, the cross-correlation between the two variables is accommodated in the *conditional mean* of y_1 via the regression parameter β. General parameterizations of this conditional mean may be considered. The point is that, under this model, one avoids having to model the cross-correlation function explicitly, thereby avoiding difficulty in specifying models which produce valid joint variance-covariance structures. It is easy to show that the implied joint variance-covariance is valid, assuming that the correlation models used to construct the conditional correlation matrices are.

The hierarchical formulation is particularly useful when there is a mechanistic basis for the conditioning, such as causal relationship between y_1 and y_2. However, there need not be scientific basis — the model is valid regardless. Consequently, this strategy may be viewed as a convenient framework for building valid joint models.

Conditional models are naturally fitted, and perhaps more easily so, within a Bayesian framework. Higher order hierarchical models are also straight-forward to analyze, but require additional regression terms in the conditional stages. Often, science might suggest independence among some of the variables, thus simplifying the problem considerably.

Also, the conditioning on *observed* variables need not be explicit. Instead, one may construct models conditional on latent (that is, unobserved) variables, and model the latent structure separately. This is convenient when there are a large number of variables, but which may be explained by a smaller-dimensional spatial process.

5 Spatiotemporal Models

Spatiotemporal processes are ubiquitous in the biological sciences. In principle, models for such processes are relatively easy to formulate in the traditional LMM or GLMM framework. However, a lack of understanding of the underlying processes and the "curse of dimensionality" make the implementation of these models challenging.

Consider a spatiotemporal process $y(\mathbf{s};t)$ 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 a discrete collection of times (e.g., $D_t = \{t_1, t_2, \ldots, t_T\}$). In the remainder of this section, we will assume the discrete temporal domain. We may then formulate the spatiotemporal process as a GLMM, where $f(y(\mathbf{s};t)|\boldsymbol{\alpha})$ is a conditionally independent distribution for all $\mathbf{s}$ and t, and where $\alpha(\mathbf{s};t)$ is a spatiotemporal random process such that $\boldsymbol{\alpha} \equiv [\boldsymbol{\alpha}_1' \ \ldots \ \boldsymbol{\alpha}_T']'$, where $\boldsymbol{\alpha}_t \equiv [\alpha(\mathbf{s}_1;t), \ldots, \alpha(\mathbf{s}_n;t)]'$ and $\boldsymbol{\alpha} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$. Note that $\boldsymbol{\Sigma}_\alpha$ is an $(n+T)\times(n+T)$ covariance matrix with elements $c_\alpha(\mathbf{s},\mathbf{s}',t,t') = \text{cov}[\alpha(\mathbf{s},t), \alpha(\mathbf{s}',t')]$. Clearly, if n or T is large, this matrix will be very large. Thus, although easy to formulate, the limitation with this approach is that the known class of realistic and valid spatiotemporal covariance functions, c_α, is very small and the dimensionality of the joint spatiotemporal process $\boldsymbol{\alpha}$ is prohibitively large. One alternative is to further factorize the joint spatiotemporal distribution for $\boldsymbol{\alpha}$ as a series of conditional models. For many processes, a dynamical factorization based on a Markov assumption in time is appropriate. That is,

$$f(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}_0) \prod_{t=1}^{T} f(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \boldsymbol{\theta}), \tag{16}$$

where the conditional distributions $f(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}, \boldsymbol{\theta})$ depend on a collection of parameters $\boldsymbol{\theta}$ that describe the dynamical evolution and the variance/covariance structure of the associated spatial noise process. For example, a conditional model might follow a first-order vector autoregressive model such as $\boldsymbol{\alpha}_t = \mathbf{H}_\theta \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t$, where $\boldsymbol{\eta}_t \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\eta(\theta))$, is the spatial noise process, and $\mathbf{H}_\theta$ is a collection of parameters that describe the evolution of the $\boldsymbol{\alpha}$ process (i.e., vector-autoregression parameter matrix). One can implement such a model in a Kalman filter framework. However, when the number of spatial locations is large, the number of parameters in $\mathbf{H}_\theta$ may prohibit likelihood-based estimation procedures. In some cases, one may have scientific knowledge that suggests relatively simple parameterizations of $\mathbf{H}_\theta$. Alternatively, one can model $\mathbf{H}_\theta$ in a hierarchical fashion, by specifying its distribution, or the distribution

of the parameters θ. This is ideally suited for a Bayesian inference approach. The Bayesian hierarchical methodology is very flexible and can accommodate a wide-range of complicated spatiotemporal processes.

6 Computation

Implementation of linear mixed models is possible on most of the standard statistical software packages and some specialized programs. These packages can be very useful for relatively simple spatial problems, and have the advantage of synergism with other statistical methods contained in the particular package. However, these packages typically assume relatively simple spatial models (e.g., isotropic and stationary) and are difficult to implement in dense and/or extensive prediction domains. This is even more pronounced in the context of generalized linear mixed models. Thus, for "complicated" spatial processes or high-dimensional prediction applications one often must write custom software for a particular application. For example, we note that although there is growing interest in spectral-based non-stationary spatial models (discussed in the next section), these approaches have not been implemented in the traditional statistical software packages.

6.1 High-Level or Low-Level Language

Practitioners typically favor either so-called "low-level" programming language implementations (such as C, C++, FORTRAN) or "high-level" languages (such as S, R, MATLAB, GAUSS). Although preference for one over the other is often an issue of familiarity, there are distinct advantages to each. For example, the "high-level" languages are often matrix-oriented and as such are very efficient when it comes to matrix calculations, but inefficient for loop-intensive programs. Conversely, the low-level languages are very efficient when it comes to traditional programming structures such as loops but often less efficient for matrix-oriented calculations.

Many practitioners have found that combinations of high- and low-level languages work best for implementation of complicated spatial models in practice. In most cases, prototype algorithms and test cases are most efficiently implemented in a matrix-oriented language. When one is satisfied with the code, one may then translate the matrix-oriented program to a low-level language for "operational" implementation. In addition, many practitioners use the high-level language as a computational "shell" and then call low-level routines for portions of the code that are less efficient.

Of course, there are significant differences between the various matrix language

programs as well. Again, preference seems to follow familiarity. However, there are some differences that can influence the choice. The authors are most familiar with R and MATLAB and thus will limit comments to those environments. R, being designed as a statistical language from the start, is ideal if one is considering multiple statistical analyses for a particular application. MATLAB, a more general purpose language, does not have the statistical functionality of R. However, MATLAB incorporates sparse data structures in a nearly seemless fashion, whereas R does not have sparse functionality built in (at the time of this writing). In many spatial problems for high-dimensional processes, sparse formulations abound out of necessity. In those cases, it is often impossible to implement a model in R due to matrix storage requirements, but with the sparse structures of MATLAB, it might be relatively simple.

6.2 Bayesian Computation

Many of the advances in spatial and spatio-temporal modeling have been due to the application of MCMC techniques to Bayesian formulations. Although there are ever improving Bayesian software packages (e.g., Bayesian inference Using Gibbs Sampling, "BUGS"), and likely to be more in the near future, many problems still require model-specific implementations. The issues raised in this context are much the same as discussed above.

6.3 Geographic Information Systems

Geographic Information Systems (GIS) are a common platform for analyzing and displaying spatial data, and many biologists are familiar with GIS technology. However, there is often a perception within the community of GIS users that GIS problems are not statistical problems. Instead, GIS is used more often for describing or displaying spatial data, and not statistical modeling. There are several reasons for this. First, there is a lack of spatial statistical modeling functionality in most GIS packages, particularly that which is useful for modeling very large spatial data sets. Moreover, there is uncertainty as to what constitutes "data" within a GIS environment. GIS "data layers" often result from applying one or more procedures to data collected from some sampling procedure ("true data"). Often, these intermediate procedures are essentially statistical models but emphasis is on producing a map, and not rigorous development or application of the procedures that transform true data to GIS data layers. Indeed, some of these procedures are in fact spatial statistical in origin (kriging and other smoothing techniques), but formal model fitting and assessment

activities are deemphasized in their application. An important barrier to the synthesis of GIS and spatial statistical modeling is that GIS users often have little or no interest in quantifying uncertainty about map predictions, or they are unsure of how to apply uncertainty estimates. Having been shown a map of predictions and model-based standard errors displayed in a GIS environment, a colleague commented on the standard error map: "What do I do with that? I only need to know how much is out there." This sentiment is endemic within the GIS community, and its root cause is the lack of exposure to or understanding of spatial statistical methods within the GIS-user community, and the failure of the statistical community to articulate both the meaning and utility of error-qualified prediction.

6.4 Technology

Although it is folly to predict the future of computational technology, one can safely assume that computing and storage power will increase dramatically in the near future. Many of the issues discussed above may lose some relevance. However, typically as the technology improves, modelers attempt ever more complicated implementations that test the limits of the current technology. Thus, the general concerns raised above are likely to be relevant for some time. Of course, software development is not static and the advantages of various software packages and implementations will change dramatically. Practitioners should continue to experiment with different packages to find ever more efficient implementations.

One area that has been under-utilized in statistics in general, and spatial statistics in particular, is parallel processing. With the increasing reliance on computationally-intensive methods, there is likely to be increasing need to explore parallel algorithms for these problems. There is significant need for research in this area.

7 Future Directions

The statistical analysis of spatial data has made tremendous progress in recent years. There remain several areas of vigorous research interest. Some of these are outlined below.

7.1 Nonstationary Spatial Processes

Classical spatial analysis typically requires that the spatial process be stationary so that parameter estimation can be accomplished. However, it is unrealistic to expect

that spatial processes one might encounter in the biological sciences be stationary. That is, the spatial dependence structure may be very different in different regions of the spatial domain. In situations where one has large amounts of spatial data and/or replications, several approaches can be used to model nonstationary covariance structures. The most promising approaches appear to be those based on morphological transformation, spectral models, and process-convolution models.

In the transformation approach one finds an appropriate warping of physical space such that the nonstationary process in physical space is stationary in the transformed space. These approaches work well in relatively small problems but are limited in that they are extremely computationally demanding in large problems.

The spectral approach is also transformation-based, but in this case one simply models the process in spectral space, that is, after performing a spectral decomposition. For example, one might model the n-dimensional spatial process $\boldsymbol{\alpha} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$ as $\boldsymbol{\alpha} = \boldsymbol{\Psi}\mathbf{a}$, where $\boldsymbol{\Psi}$ is an $n \times m$ matrix of spectral basis functions, and $\mathbf{a}$ is a $m \times 1$ vector of spectral coefficients such that $\alpha(\mathbf{s}_i) = \sum_{j=1}^{m} \psi_j(\mathbf{s}_i) a_j$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Although typically $m = n$, in some situations it may be that the spectral decomposition also suggests dimension reduction and $m < n$. The spectral coefficients have the distribution $\mathbf{a} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_a)$. If the basis functions are orthogonal, one can write $\mathbf{a} = \boldsymbol{\Psi}'\boldsymbol{\alpha}$ and then $\boldsymbol{\Sigma}_a = \boldsymbol{\Psi}'\boldsymbol{\Sigma}_\alpha\boldsymbol{\Psi}$. One has flexibility in choosing the basis functions (e.g., Fourier basis functions, orthogonal polynomials, empirical orthogonal functions, or wavelets). Depending on the choice, the basis functions have the added effect of being a decorrelator, so that the structure of $\boldsymbol{\Sigma}_a$ is much simpler than $\boldsymbol{\Sigma}_\alpha$. Perhaps more importantly, such specifications can easily lead to non-stationary covariance functions. These spectral decompositions can also be included in fully Bayesian analyses and, if chosen properly, can lead to efficient Bayesian computation in large spatial domains.

The process-convolution approach is also inherently a spectral approach, but need not be written that way. In this approach one notes that the Gaussian spatial process, $y(\mathbf{s})$ can be written as a spatial moving average,

$$y(\mathbf{s}) = \int k(\mathbf{s} - \mathbf{u}; \boldsymbol{\theta}) \gamma(\mathbf{u}) d\mathbf{u}, \tag{17}$$

where $k(\mathbf{s} - \mathbf{u}; \boldsymbol{\theta})$ is a convolution kernel that depends on parameters $\boldsymbol{\theta}$ and $\gamma(\mathbf{u})$ is a Gaussian white noise process. One can show that such processes imply valid stationary spatial covariance models. It follows that if one allows the convolution kernel to vary with space, then the implied covariance function is nonstationary. That is, let $k(\mathbf{s} - \mathbf{u}; \boldsymbol{\theta}_s)$, for all $\mathbf{s}$ in the spatial domain of interest. Typically, one specifies

a slowly varying process for $\boldsymbol{\theta}_s$ and considers a hierarchical Bayesian implementation. One must be careful that the kernel parameter model is relatively simple lest one replace a univariate spatial problem with an equally complicated multivariate problem.

7.2 Multivariate Non-Gaussian Formulations

The ability to implement non-Gaussian spatial models is relatively recent, primarily because the computational tools necessary to implement them are relatively new. Perhaps the greatest frontier in spatial modeling in the near future will be modeling *spatial systems*. That is, multivariate processes, some of which may be non-Gaussian. Such models are essential for understanding complicated biological systems. For example, consider the problem of predicting the abundance of waterfowl over a continental scale six months to a year in advance. Such a model would have to consider seasonal to annual climate predictions of sea surface temperature, relate these predictions to climatological variables such as precipitation and temperature, which must then be related to habitat, and eventually waterfowl abundance. Each component of such a model is potentially a very complicated spatial or spatiotemporal process, typically non-Gaussian. Each subprocess has associated uncertainties in the data, knowledge of the process, and the statistical model. Thus, any reasonable modeling strategy must be able to account for these uncertainties. The hierarchical Bayesian linkage of GLMM formulations provides such a framework.

Acknowledgments

The authors are grateful to Professor J. Maindonald for the invitation to participate in the EOLSS project. The writing of this manuscript was supported in part by a grant from the U.S. Environmental Protection Agency's Science to Achieve Results (STAR) program, Assistance Agreement No. R827257-01-0.

Bibliography

Casella G. and George E.I. (1992). Explaining the Gibbs sampler. *The American Statistician* **46**, 167-174. [An easy-to-read explanation of one of the most popular MCMC approaches.]

Chib D.D., Cox L.H., and Ensor K.B. (1995). Understanding the Metropolis-Hastings

algorithm. *The American Statistician* **49**, 327-335. [An accessible explanation of a powerful MCMC algorithm.]

Clayton D.G. (1995) Generalized linear mixed models. In *Markov Chain Monte Carlo in Practice* (Gilks W.R., Richardson S., and Spiegelhalter D.J., editors). London: Chapman and Hall. [An excellent overview of generalized linear mixed models from both frequentist and Bayesian perspectives.]

Cressie N.A.C, (1993). *Statistics for Spatial Data, Revised Edition*, New York: John Wiley and Sons. [This book is the seminal reference on spatial statistics.]

Diggle P.J., Tawn J.A., and Moyeed, R.A. (1998). Model-based geostatistics (with discussion). *Applied Statistics* **47**, 299-350. [This paper provides the foundation for non-Gaussian spatial modeling.]

Gilks W.R., Richardson S., and Spiegelhalter D.J. (1995). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall. [One of the most practical compilations explaining MCMC.]

Higdon D. (1998) A process-convolution approach to modelling temperatures in the North Atlantic ocean. *Environmental and Ecological Statistics* **5**, 173-190. [This paper illustrates how one can model non-stationary spatial processes via a convolution approach.]

McCulloch C.E., and Searle S.R. (2001) *Generalized, Linear, and Mixed Models*, New York: John Wiley and Sons. [This text provides a nice unified framework of the modern view of linear models, linear mixed models, generalized linear models, and generalized linear mixed models.]

Robinson G.K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science* **6**, 15-51. [Classic paper describing best linear unbiased prediction in linear mixed models.]

Royle, J.A. and Berliner, L.M. (1999) A hierarchical approach to multivariate spatial modeling and prediction, *J. Agr. Biol. and Env. Statistics*, **4**(1):29-56. [Provides details on the hierarchical formulation of multivariate spatial models.]

Royle, J.A., Link, W.A., and Sauer, J.R. (2001) Statistical mapping of count survey data. In Predicting Species Occurrences: Issues of Scale and Accuracy, (Scott, J. M., P. J. Heglund, M. Morrison, M. Raphael, J. Haufler, B. Wall, editors). Island Press. Covello, CA. [Contains additional details on the BBS modeling example]

Stein, A., and Corsten, L.C.A. (1991) Universal kriging and cokriging as a regression procedure, *Biometrics*, **47**, 575-587. [Introduction to kriging ideas from a more traditional regression perspective.]

Ver Hoef, J.M. and Cressie, N. (1993) Multivariable spatial prediction, *Mathematical Geology*, **25**(2), 219-239. [Provides the traditional formulation of cokriging.]

Wikle C.K., Berliner L.M., and Cressie N. (1998) Hierarchical Bayesian space-time models, *Environmental and Ecological Statistics* **5**, 117-154. [This paper provides an extensive overview of hierarchical modeling of space-time dynamic processes.]

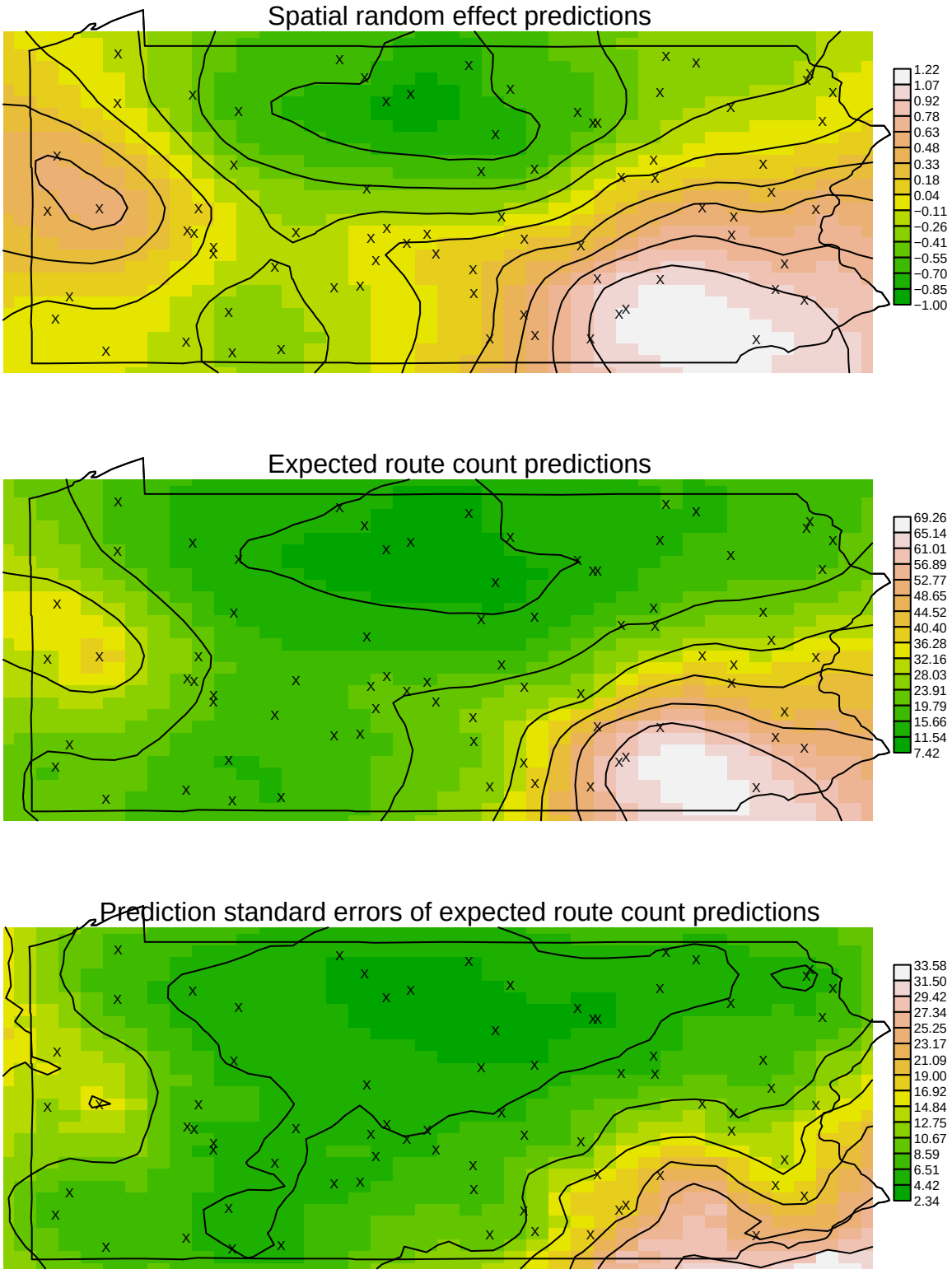

Figure 1: Results from MCMC analysis of BBS mourning dove counts. Top panel: Spatial random effect posterior mean at grid locations. Middle panel: Posterior mean map of expected route counts at the grid locations. Bottm panel: Prediction standard errors of expected route counts at grid locations. Route locations are indicated with "X".