

Chapter 1

Gaussian random field models for spatial data

Murali Haran

1.1 Introduction

Spatial data contain information about both the attribute of interest as well as its location. Examples can be found in a large number of disciplines including ecology, geology, epidemiology, geography, image analysis, meteorology, forestry, and geosciences. The location may be a set of coordinates, such as the latitude and longitude associated with an observed pollutant level, or it may be a small region such as a county associated with an observed disease rate. Following Cressie (1993), we categorize spatial data into three distinct types: (i) *geostatistical or point-level data*, as in the pollutant levels observed at several monitors across a region, (ii) *lattice or ‘areal’ (regionally aggregated) data*, for example U.S. disease rates provided by county, and (iii) *point process data*, where the locations themselves are random variables and of interest, as in the set of locations where a rare animal species was observed. Point processes where random variables associated with the random locations are also of interest

are referred to as *marked point processes*. In this article, we only consider spatial data that fall into categories (i) and (ii). We will use the following notation throughout: Denote a real-valued spatial process in d dimensions by $\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$ where \mathbf{s} is the location of the process $Z(\mathbf{s})$ and \mathbf{s} varies over the index set D , resulting in a multivariate random process. For point-level data D is a continuous, fixed set, while for lattice or areal data D is discrete and fixed. For spatial point processes, D is stochastic and usually continuous. The distinctions among the above categories may not always be apparent in any given context, so determining a category is part of the modeling process.

The purpose of this article is to discuss the use of Gaussian random fields for modeling a variety of point-level and areal spatial data, and to point out the flexibility in model choices afforded by Markov chain Monte Carlo algorithms. Details on theory, algorithms and advanced spatial modeling can be found in Cressie (1993), Stein (1999), Banerjee et al. (2004) and other standard texts. The reader is referred to the excellent monograph by Møller and Waagepetersen (2004) for details on modeling and computation for spatial point processes.

1.1.1 Some motivation for spatial modeling

Spatial modeling can provide a statistically sound approach for performing interpolations for point-level data, which is at the heart of ‘kriging’, a body of work originating from mineral exploration (see Matheron, 1971). Even when interpolation is not the primary goal, accounting for spatial dependence can lead to better inference, superior predictions, and more accurate estimates of the variability of estimates. We describe toy examples to illustrate two general scenarios where modeling spatial dependence can be beneficial: when there is dependence in the data, and when we need to adjust for an unknown spatially varying mean. Learning about spatial dependence from observed data may also be of interest in its own right, for example in research questions where detecting spatial clusters is of interest.

Example 1. Accounting appropriately for dependence. Let $Z(s)$ be a random variable indexed by its location $s \in (0, 1)$, and $Z(s) = 6s + \epsilon(s)$ with dependent errors, $\epsilon(s)$, generated via a simple autoregressive model: $\epsilon(s_1) = 7$, $\epsilon(s_i) \sim N(0.9\epsilon(s_{i-1}), 0.1)$, $i = 2, \dots, 100$ for equally spaced locations x_1, \dots, x_{100} in $(0, 1)$. Figure 1.1(a) shows how a model that assumes the

errors are dependent, such as a linear Gaussian process model (solid curves) described later in Section 1.2.1, provides a much better fit than a regression model with independent errors (dotted lines). Note that for spatial data, s is usually in 2-D or 3-D space; we are only considering 1-D space here in order to better illustrate the ideas.

Figure 1.1 near here.

Example 2. Adjusting for an unknown spatially varying mean. Suppose $Z(s) = \sin(s) + \epsilon(s)$ where, for any set of locations $s_1, \dots, s_k \in (0, 1)$, and $\epsilon(s_1), \dots, \epsilon(s_k)$ are independent and identically distributed normal random variables with 0 mean and variance σ^2 . Suppose $Z(s)$ is observed at ten locations. From Figure 1.1(b) the dependent error model (solid curves) is superior to an independent error model (dotted lines), even though there was no dependence in the generating process. Adding dependence can thus act as a form of protection against a poorly specified model.

The above example shows how accounting for spatial dependence can adjust for a misspecified mean, thereby accounting for important missing spatially varying covariate information (for instance, the $\sin(x)$ function above). As pointed out in Cressie (1993, p.25), “What is one person’s (spatial) covariance structure may be another person’s mean structure.” In other words, an interpolation based on assuming dependence (a certain covariance structure) can be similar to an interpolation that utilizes a particular mean structure ($\sin(x)$ above). Example 2 also shows the utility of Gaussian processes for modeling the relationship between ‘inputs’ (s_1, \dots, s_n) and ‘outputs’ ($Z(s_1), \dots, Z(s_n)$) when little is known about the parametric form of the relationship. In fact, this flexibility of Gaussian processes has been exploited for modeling relationships between inputs and outputs from complex computer experiments (see Currin et al., 1991; Sacks et al., 1989). For more discussion on motivations for spatial modeling see, for instance, Cressie (1993, p.13) and Schabenberger and Gotway (2005, p.31).

1.1.2 MCMC and spatial models: a shared history

Most algorithms related to Markov chain Monte Carlo (MCMC) originated in statistical physics problems concerned with lattice systems of particles, including the original Metropo-

lis et al. (1953) paper. The Hammersley-Clifford Theorem (Besag, 1974; Clifford, 1990) provides an equivalence between the local specification via the conditional distribution of each particle given its neighboring particles, and the global specification of the joint distribution of all the particles. The specification of the joint distribution via local specification of the conditional distributions of the individual variables is the Markov random field specification, which has found extensive applications in spatial statistics and image analysis, as outlined in a series of papers by Besag and co-authors (see Besag, 1974, 1989; Besag et al., 1995; Besag and Kempton, 1986), and several papers on Bayesian image analysis (Amit et al., 1991; Geman and Geman, 1984; Grenander and Keenan, 1989). It is also the basis for variable-at-a-time Metropolis-Hastings and Gibbs samplers for simulating these systems. Thus, spatial statistics was among the earliest fields to recognize the power and generality of MCMC. A historical perspective on the connection between spatial statistics and MCMC along with related references can be found in Besag and Green (1993).

While these original connections between MCMC and spatial modeling are associated with Markov random field models, this discussion of Gaussian random field models includes both Gaussian process (GP) and Gaussian Markov random field (GMRF) models in Section 1.2. In Section 1.3, we describe the generalized versions of both linear models, followed by a discussion of non-Gaussian Markov random field models in Section 1.4 and a brief discussion of more flexible models in Section 1.5.

1.2 Linear spatial models

In this section, we discuss linear Gaussian random field models for both geostatistical and areal (lattice) data. Although a wide array of alternative approaches exist (see Cressie, 1993), we model the spatial dependence via a parametric covariance function or, as is common for lattice data, via a parameterized precision (inverse covariance) matrix, and consider Bayesian inference and prediction.

1.2.1 Linear Gaussian process models

We first consider geostatistical data. Let the spatial process at location $\mathbf{s} \in D$ be defined as

$$Z(\mathbf{s}) = X(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s}), \text{ for } \mathbf{s} \in D, \quad (1.1)$$

where $\mathbf{X}(\mathbf{s})$ is a set of p covariates associated with each site \mathbf{s} , and $\boldsymbol{\beta}$ is a p -dimensional vector of coefficients. Spatial dependence can be imposed by modeling $\{w(\mathbf{s}) : \mathbf{s} \in D\}$ as a zero mean stationary Gaussian process. Distributionally, this implies that for any $\mathbf{s}_1, \dots, \mathbf{s}_n \in D$, if we let $\mathbf{w} = (w(\mathbf{s}_1), \dots, w(\mathbf{s}_n))^T$ and Θ be the parameters of the model, then

$$\mathbf{w} \mid \Theta \sim N(0, \Sigma(\Theta)), \quad (1.2)$$

where $\Sigma(\Theta)$ is the covariance matrix of the n -dimensional normal density. We need $\Sigma(\Theta)$ to be symmetric and positive definite for this distribution to be proper. If we specify $\Sigma(\Theta)$ by a positive definite parametric covariance function, we can ensure that these conditions are satisfied. For example, consider the exponential covariance with parameters $\Theta = (\psi, \kappa, \phi)$, with $\psi, \kappa, \phi > 0$. The exponential covariance $\Sigma(\Theta)$ has the form $\Sigma(\Theta) = \psi I + \kappa H(\phi)$, where I is the identity matrix, the i, j th element of $H(\phi)$ is $\exp(-\|\mathbf{s}_i - \mathbf{s}_j\|/\phi)$, and $\|\mathbf{s}_i - \mathbf{s}_j\|$ is the Euclidean distance between locations $\mathbf{s}_i, \mathbf{s}_j \in D$. Alternatives to Euclidean distance may be useful, for instance geodesic distances are often appropriate for spatial data over large regions (Banerjee, 2005). This model is interpreted as follows: the “nugget” ψ is the variance of the non-spatial error, say from measurement error or from a micro-scale stochastic source associated with each location, and κ and ϕ dictate the scale and range of the spatial dependence respectively. Clearly, this assumes the covariance and hence dependence between two locations decreases as the distance between them increases.

The exponential covariance function is important for applications, but is a special case of the more flexible Matérn family (Handcock and Stein, 1993). The Matérn covariance between $Z(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)$ with parameters $\psi, \kappa, \phi, \nu > 0$ is based only on the distance x

between \mathbf{s}_i and \mathbf{s}_j ,

$$\text{Cov}(x; \psi, \kappa, \phi, \nu) = \begin{cases} \frac{\kappa}{2^{\nu-1}\Gamma(\nu)} (2\nu^{1/2}x/\phi)^\nu K_\nu(2\nu^{1/2}x/\phi) & \text{if } x > 0 \\ \psi + \kappa & \text{if } x = 0 \end{cases} \quad (1.3)$$

where $K_\nu(x)$ is a modified Bessel function of order ν (Abramowitz and Stegun, 1964), and ν determines the smoothness of the process. As ν increases, the process becomes increasingly smooth. As an illustration, Figure 1.1 compares prediction (interpolation) using GPs with exponential ($\nu = 0.5$) and gaussian ($\nu \rightarrow \infty$) covariance functions (we use lower case for “gaussian” as suggested in Schabenberger and Gotway, 2005, since the covariance is not related to the Gaussian distribution). Notice how the gaussian covariance function produces a much smoother interpolator (dashed curves) than the more ‘wiggly’ interpolation produced by the exponential covariance (solid curves). Stein (1999) recommends the Matérn since it is flexible enough to allow the smoothness of the process to also be estimated. He cautions against GPs with gaussian correlations since they are overly smooth (they are infinitely differentiable). In general the smoothness ν may be hard to estimate from data; hence a popular default is to use the exponential covariance for spatial data where the physical process producing the realizations is unlikely to be smooth, and a gaussian covariance for modeling output from computer experiments or other data where the associated smoothness assumption may be reasonable.

Let $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$. From (1.1) and (1.2), once a covariance function is chosen (say according to (1.3)) \mathbf{Z} has a multivariate normal distribution with unknown parameters Θ, β . Maximum likelihood inference for the parameters is then simple in principle, though strong dependence among the parameters and expensive matrix operations may sometimes make it more difficult. A Bayesian model specification is completed with prior distributions placed on Θ, β . “Objective” priors (perhaps more appropriately referred to as “default” priors) for the linear GP model have been derived by several authors (Berger et al., 2001; De Oliveira, 2007; Paulo, 2005). These default priors are very useful since it is often challenging to quantify prior information about these parameters in a subjective manner. However, they can be complicated and computationally expensive, and proving posterior propriety often necessitates analytical work. To avoid posterior impropriety when

building more complicated models, it is common to use proper priors and rely on approaches based on exploratory data analysis to determine prior settings. For example, one could use a uniform density that allows for a reasonable range of values for the range parameter ϕ , and inverse gamma densities with an infinite variance and mean set to a reasonable guess for κ and ψ (see, for e.g., Finley et al., 2007), where the guess may again depend on some rough exploratory data analysis like looking at variograms. For a careful analysis, it is critical to study sensitivity to prior settings.

MCMC for linear GPs

Inference for the linear GP model is based on the posterior distribution $\pi(\Theta, \beta \mid \mathbf{Z})$ that results from (1.1) and (1.2) and a suitable prior for Θ, β . Although π is of fairly low dimensions as long as the number of covariates is not too large, MCMC sampling for this model can be complicated by two issues: (i) the strong dependence among the covariance parameters, which leads to autocorrelations in the sampler, and (ii) by the fact that matrix operations involved at each iteration of the algorithm are of order N^3 , where N is the number of data points. Reparameterization-based MCMC approaches such as those proposed in Cowles et al. (2009); Yan et al. (2007) or block updating schemes, where multiple covariance parameters are updated at once in a single Metropolis-Hastings step (cf. Tibbits et al., 2009), may help with the dependence. Also, there are existing software implementations of MCMC algorithms for linear GP models (Finley et al., 2007; Smith et al., 2008). A number of approaches can be used to speed up the matrix operations, including changing the covariance function in order to induce sparseness or other special matrix structures that are amenable to fast matrix algorithms; we discuss this further in Section 1.5.

Predictions of the process, $\mathbf{Z}^* = (Z(\mathbf{s}_1^*), \dots, Z(\mathbf{s}_m^*))^T$ where $\mathbf{s}_1^*, \dots, \mathbf{s}_m^*$ are new locations in D , are obtained via the posterior predictive distribution,

$$\pi(\mathbf{Z}^* \mid \mathbf{Z}) = \int \pi(\mathbf{Z}^* \mid \mathbf{Z}, \Theta, \beta) \pi(\Theta, \beta \mid \mathbf{Z}) d\Theta d\beta. \quad (1.4)$$

Under the Gaussian process assumption the joint distribution of \mathbf{Z}, \mathbf{Z}^* given Θ, β is

$$\begin{bmatrix} \mathbf{Z} \\ \mathbf{Z}^* \end{bmatrix} \mid \Theta, \beta \sim N \left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right), \quad (1.5)$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the linear regression means of \mathbf{Z} and \mathbf{Z}^* (functions of covariates and β), and $\Sigma_{11}, \Sigma_{12}, \Sigma_{21}, \Sigma_{22}$ are block partitions of the covariance matrix $\Sigma(\Theta)$ (functions of covariance parameters Θ). By basic normal theory (e.g. Anderson, 2003), $\mathbf{Z}^* \mid \mathbf{Z}, \beta, \Theta$, corresponding to the first term in the integrand in (1.4), is normal with mean and covariance

$$E(\mathbf{Z}^* \mid \mathbf{Z}, \beta, \Theta) = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{Z} - \boldsymbol{\mu}_1), \text{ and } \text{Var}(\mathbf{Z}^* \mid \mathbf{Z}, \beta, \Theta) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}. \quad (1.6)$$

Note in particular that the prediction for \mathbf{Z}^* given \mathbf{Z} has expectation obtained by adding two components: (i) the mean $\boldsymbol{\mu}_2$ which, in the simple linear case, is $\beta\mathbf{X}^*$, where \mathbf{X}^* are the covariates at the new locations, and (ii) a product of the residual from the simple linear regression on the observations $(\mathbf{Z} - \boldsymbol{\mu}_1)$ weighted by $\Sigma_{21}\Sigma_{11}^{-1}$. If there is no dependence, the second term is close to 0, but if there is a strong dependence, the second term pulls the expected value at a new location closer to the values at nearby locations. Draws from the posterior predictive distribution (1.4) are obtained in two steps: (i) Simulate $\Theta', \beta' \sim \pi(\Theta, \beta \mid \mathbf{Z})$ by the Metropolis-Hastings algorithm, (ii) Simulate $\mathbf{Z}^* \mid \Theta', \beta', \mathbf{Z}$ from a multivariate normal density with conditional mean and covariance from (1.6) using the Θ', β' draws from step (i).

Example 3. Haran et al. (2009) interpolate flowering dates for wheat crops across North Dakota as part of a model to estimate crop epidemic risks. The flowering dates are only available at a few locations across the state, but using a linear GP model with a Matérn covariance, it is possible to obtain distributions for interpolated flowering dates at sites where other information (weather predictors) are available for the epidemic model, as shown in Figure 1.2. Although only point estimates are displayed here, the full distribution of the interpolated flowering dates are used when estimating crop epidemic risks.

Figure 1.2 near here.

1.2.2 Linear Gaussian Markov random field models

A direct specification of spatial dependence via $\Sigma(\Theta)$, while intuitively appealing, relies on measuring spatial proximity in terms of distances between the locations. When modeling areal data, it is possible to use measures such as intercentroid distances to serve this purpose, but this can be awkward due to irregularities in the shape of the regions. Also, since the data are aggregates, assuming a single location corresponding to multiple random variables may be inappropriate. An alternative approach is a conditional specification, by assuming that a random variable associated with a region depends primarily on its neighbors. A simple neighborhood could consist of adjacent regions, but more complicated neighborhood structures are possible depending on the specifics of the problem. Let the spatial process at location $\mathbf{s} \in D$ be defined as in (1.1) so $Z(\mathbf{s}) = X(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s})$, but now assume that the spatial random variables (‘random effects’) \mathbf{w} are modeled *conditionally*. Let \mathbf{w}_{-i} denote the vector \mathbf{w} excluding $w(\mathbf{s}_i)$. For each \mathbf{s}_i we model $w(\mathbf{s}_i)$ in terms of its *full conditional distribution*, that is, its distribution given the remaining random variables, \mathbf{w}_{-i} :

$$w(\mathbf{s}_i) \mid \mathbf{w}_{-i}, \Theta \sim N \left(\sum_{j=1}^n c_{ij} w(\mathbf{s}_j), \kappa_i^{-1} \right), \quad i = 1, \dots, n, \quad (1.7)$$

where c_{ij} describes the neighborhood structure. c_{ij} is non-zero only if i and j are neighbors, while the κ_i s are the precision (inverse variance) parameters. To make the connection to linear Gaussian process model (1.2) apparent, we let Θ denote the precision parameters. Each $w(\mathbf{s}_i)$ is therefore a normal random variate with mean based on neighboring values of $w(\mathbf{s}_i)$. Just as we need to ensure that the covariance is positive definite for a valid Gaussian process, we need to ensure that the set of conditional specifications result in a valid joint distribution. Let Q be an $n \times n$ matrix with i th diagonal element κ_i and i, j th off-diagonal element $-\kappa_i c_{ij}$. Besag (1974) proved that if Q is symmetric and positive definite (1.7) specifies a valid joint distribution,

$$\mathbf{w} \mid \Theta \sim N(0, Q^{-1}), \quad (1.8)$$

with Θ the set of precision parameters (note that c_{ij} s and κ_i s depend on Θ). Usually a common precision parameter, say τ , is assumed so $\kappa_i = \tau$ for all i , and hence $Q(\tau) = \tau(I + C)$ where C is a matrix which has 0 on its diagonals and i, j th off-diagonal element $-c_{ij}$, though a more attractive smoother may be obtained by using weights in a GMRF model motivated by a connection to thin-plate splines (Yue and Speckman, 2009). To add flexibility to the above GMRF model, some authors have included an extra parameter in the matrix C (see Ferreira and De Oliveira, 2007). Inference for the linear GMRF model specified by (1.1) and (1.8) can therefore proceed after assuming a prior distribution for τ, β , often an inverse gamma and flat prior respectively. An alternative formulation is an improper version of the GMRF prior, the so called “Intrinsic Gaussian Markov random field” (Besag and Kooperberg, 1995):

$$f(\mathbf{w} \mid \Theta) \propto \tau^{(N-1)/2} \exp\{-\mathbf{w}^T Q(\tau) \mathbf{w}\}, \quad (1.9)$$

where Q has $-\tau c_{ij}$ on its off-diagonals (as above) and i th diagonal element $\tau \sum_j c_{ij}$. The notation $j \sim i$ implies that i and j are neighbors. In the special case where $c_{ij} = 1$ if $j \sim i$ and 0 otherwise, (1.9) simplifies to the “pairwise-difference form,”

$$f(\mathbf{w} \mid \Theta) \propto \tau^{(N-1)/2} \exp\left(-\frac{1}{2} \sum_{i \sim j} \{w(\mathbf{s}_i) - w(\mathbf{s}_j)\}^2\right),$$

which is convenient for constructing MCMC algorithms with univariate updates since the full conditionals are easy to evaluate. Q is rank deficient so the above density is improper. This form is a very popular prior for the underlying spatial field of interest. For instance, denote noisy observations by $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))^T$, so $y(\mathbf{s}_i) = w(\mathbf{s}_i) + \epsilon_i$ where $\epsilon_i \sim \text{Normal}(0, \sigma^2)$ is independent error. Then an estimate of the smoothed underlying spatial process \mathbf{w} can be obtained from the posterior distribution of $\mathbf{w} \mid \mathbf{y}$ as specified by (1.9). If the parameters, say τ and σ^2 , are also to be estimated and have priors placed on them, inference is based on the posterior $\mathbf{w}, \tau, \sigma^2 \mid \mathbf{y}$. The impropriety of the intrinsic GMRF is not an issue as long as the posterior is proper. If $c_{ij} = 1$ when i and j are neighbors and 0 otherwise, this corresponds to an intuitive conditional specification:

$$f(w_j \mid \mathbf{w}_{-j}, \tau) \sim N\left(\frac{\sum_{j \in N(i)} w(\mathbf{s}_j)}{n}, \frac{1}{n_i \tau}\right),$$

where n_i is the number of neighbors for the i th region, and $N(i)$ is the set of neighbors of the i th region. Hence, the distribution of $w(\mathbf{s}_i)$ is normal with mean given by the average of its neighbors and its variance decreases as the number of neighbors increases. See Rue and Held (2005) for a discussion of related theory for GMRF models, and Sun et al. (1999) for conditions under which posterior propriety is guaranteed for various GMRF models.

Although GMRF-based models are very popular in statistics and numerous other fields, particularly computer science and image analysis, there is some concern about whether they are reasonable models even for areal or lattice data (McCullagh, 2002). The marginal dependence induced can be complicated and counter-intuitive (Besag and Kooperberg, 1995; Wall, 2004). In addition, a GMRF model on a lattice is known to be inconsistent with the corresponding GMRF model on a subset of the lattice, that is, the corresponding marginal distributions are not the same. However, quoting Besag (2002), this is not a major issue if “The main purpose of having the spatial dependence is to absorb spatial variation (dependence) rather than produce a spatial model with scientifically interpretable parameters.” GMRF models can help produce much better individual estimates by “borrowing strength” from the neighbors of each individual (region). This is of particular importance in small area estimation problems (see Ghosh and Rao, 1994), where many observations are based on small populations, for instance disease rate estimates in sparsely populated counties. Spatial dependence allows the model to borrow information from neighboring counties which may collectively have larger populations, thereby reducing the variability of the estimates. Similar considerations apply in disease mapping models (Mollié, 1996) where small regions and the rarity of diseases have led to the popularity of variants of the GMRF-based Bayesian image restoration model due to Besag et al. (1991). More sophisticated extensions of such models in the context of environmental science and public health are described in several recent books (see, for instance, Lawson, 2008; Le and Zidek, 2006; Waller and Gotway, 2004). Several of these models fall under the category of spatial generalized linear models, as discussed in Section 1.3.

MCMC for linear GMRFs

The conditional independence structure of a GMRF makes it natural to write and compute the full conditional distributions of each $w(\mathbf{s}_i)$, without any matrix computations. Hence MCMC algorithms which update a single variable at a time are easy to construct. When this algorithm is efficient, it is preferable due to its simplicity. Unfortunately, such univariate algorithms may often result in slow mixing Markov chains. In the linear GMRF model posterior distribution, it is possible to analytically integrate out all the spatial random effects (\mathbf{w}), that is, it is easy to integrate the posterior distribution $\pi(\mathbf{w}, \Theta, \boldsymbol{\beta} \mid \mathbf{Z})$ with respect to \mathbf{w} to obtain the marginal $\pi(\Theta, \boldsymbol{\beta} \mid \mathbf{Z})$ in closed form. This is a fairly low dimensional distribution, similar to the linear GP model posterior, and similar strategies as described for sampling from the linear GP model posterior may be helpful here. However, unlike the linear GP model posterior, all matrices involved in linear GMRF models are sparse. A reordering of the nodes corresponding to the graph can exploit the sparsity of the precision matrices of GMRFs, thereby reducing the matrix operations from $\mathcal{O}(n^3)$ to $\mathcal{O}(nb^2)$ where b^2 is the bandwidth of the sparse matrix; see Rue (2001) and Golub and Van Loan (1996, p.155). For instance, Example 5 (described in a later section) involves $n = 454$ data points, but the reordered precision matrix has a bandwidth of just 24. The matrix computations are therefore speeded up by a factor of 357 each, and the ensuing increase in computational speed is even larger.

1.2.3 Section summary

Linear Gaussian random fields are a simple and flexible approach to modeling dependent data. When the data are point-level, GPs are convenient since the covariance can be specified as a function of the distance between any two locations. When the data are aggregated or on a lattice, GMRFs are convenient as dependence can be specified in terms of adjacencies and neighborhoods. MCMC allows for easy simulation from the posterior distribution for both categories of models, especially since the low dimensional posterior distribution of the covariance (or precision) parameters and regression coefficients may be obtained in closed form. Relatively simple univariate Metropolis-Hastings algorithms may work well, and ex-

isting software packages can implement reasonably efficient MCMC algorithms. When the simple approaches produce slow mixing Markov chains, reparameterizations or block updating algorithms may be helpful. Many strategies are available for reducing the considerable computational burden posed by matrix operations for linear GP models, including the use of covariance functions that result in special matrix structures amenable to fast computations. GMRFs have significant computational advantages over GPs due to the conditional independence structure which naturally result in sparse matrices and greatly reduced computations for each update of the MCMC algorithm.

1.3 Spatial generalized linear models

Linear GP and GMRF models are very flexible, and work surprisingly well in a variety of situations, including many where the process is quite non-Gaussian and discrete, such as some kinds of spatial count data. When the linear Gaussian assumption provides a poor fit to data, transforming the data say via the Box-Cox family of transformations and modeling the transformed response via a linear GP or GMRF may be adequate (see ‘trans-Gaussian kriging’, for instance, in Cressie, 1993, with the use of delta method approximations to estimate the variance and perform bias-correction). However, when it is important to model the known sampling mechanism for the data, and this mechanism is non-Gaussian, spatial generalized linear models (SGLMs) may be very useful. SGLMs are generalized linear models (McCullagh and Nelder, 1983) for spatially associated data. The spatial dependence (the error structure) for SGLMs can be modeled via Gaussian processes for point-level (‘geostatistical’) data as described in the seminal paper by Diggle et al. (1998). Here, we also include the use of Gaussian Markov random field models for the errors, as commonly used for lattice or areal data. Note that the SGLMs here may also be referred to as spatial generalized linear mixed models since the specification of spatial dependence via a generalized linear model framework always involves random effects.

1.3.1 The generalized linear model framework

We begin with a brief description of SGLMs using Gaussian process models. Let $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ and $\{w(\mathbf{s}) : \mathbf{s} \in D\}$, be two spatial process on $D \subset \mathbb{R}^d$ ($d \in \mathbb{Z}^+$.) Assume the $Z(\mathbf{s}_i)$ s are conditionally independent given $w(\mathbf{s}_1), \dots, w(\mathbf{s}_n)$, where $\mathbf{s}_1, \dots, \mathbf{s}_n \in D$, and the $Z(\mathbf{s}_i)$ conditionally follow some common distributional form, for example Poisson for count data or Bernoulli for binary data, and

$$E(Z(\mathbf{s}_i) \mid \mathbf{w}) = \mu(\mathbf{s}_i), \text{ for } i = 1, \dots, n. \quad (1.10)$$

Let $\eta(\mathbf{s}) = h\{\mu(\mathbf{s})\}$ for some known link function $h(\cdot)$ (for example, the logit link, $h(x) = \log\left(\frac{x}{1-x}\right)$ or log link, $h(x) = \log(x)$). Furthermore, assume that

$$\eta(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s}), \quad (1.11)$$

where $\mathbf{X}(\mathbf{s})$ is a set of p covariates associated with each site \mathbf{s} , and $\boldsymbol{\beta}$ is a p -dimensional vector of coefficients. Spatial dependence is imposed on this process by modeling $\{w(\mathbf{s}) : \mathbf{s} \in D\}$ as a stationary Gaussian process so $\mathbf{w} = (w(\mathbf{s}_1), \dots, w(\mathbf{s}_n))^T$ is distributed as

$$\mathbf{w} \mid \Theta \sim N(0, \Sigma(\Theta)). \quad (1.12)$$

$\Sigma(\Theta)$ is a symmetric, positive definite covariance matrix usually defined via a parametric covariance such as a Matérn covariance function (Handcock and Stein, 1993), where Θ is a vector of parameters used to specify the covariance function. Note that with the identity link function and Gaussian distributions for the conditional distribution of the $Z(\mathbf{s}_i)$, we can obtain the linear Gaussian process model as a special case. The model specification is completed with prior distributions placed on $\Theta, \boldsymbol{\beta}$, where proper priors are typically chosen to avoid issues with posterior impropriety. There has been little work on prior settings for SGLMs, with researchers relying on a mix of heuristics and experience to derive suitable priors. Prior sensitivity analyses are, again, crucial, as also discussed in Section 1.6. It is important to carefully interpret the regression parameters in SGLMs *conditional* on the underlying spatial random effects, rather than as the usual marginal regression coefficients

Diggle et al. (1998, p.302).

The GMRF version of SGLMs are formulated in similar fashion so (1.10) and (1.11) stay the same but (1.12) is replaced by (1.8). Inference for the SGLM model is based on the posterior distribution $\pi(\Theta, \beta, \mathbf{w} \mid \mathbf{Z})$. Predictions can then be obtained easily via the posterior predictive distribution. In principle, the solution to virtually any scientific question related to these models is easily obtained via sample-based inference. Examples of such questions include finding maxima (see the example in Diggle et al., 1998), spatial cumulative distribution functions when finding the proportion of area where $Z(\mathbf{s})$ is above some limit (Short et al., 2005), and integrating over subregions in the case of Gaussian process SGLMs when inference is required over a subregion.

1.3.2 Examples

Binary data

Spatial binary data occur frequently in environmental and ecological research, for instance when the data correspond to presence or absence of a certain invasive plant species at a location, or when the data happen to fall into one of two categories, say two soil types. Interpolation in point-level data and smoothing in areal/lattice data may be of interest. Often, researchers may be interested in learning about relationships between the observations and predictors while adjusting appropriately for spatial dependence, and in some cases learning about spatial dependence may itself be of interest.

Example 4. The coastal marshes of the mid-Atlantic are an extremely important aquatic resource. An invasive plant species called *Phragmites australis* or “phrag” is a major threat to this aquatic ecosystem (see Saltonstall, 2002), and its rapid expansion may be the result of human activities causing habitat disturbance (Marks et al., 1994). Data from the Atlantic Slopes Consortium (Brooks et al., 2006) provide information on presence or absence of phrag in the Chesapeake Bay area, along with predictors of phrag presence such as land use characteristics. Accounting for spatial dependence when studying phrag presence is important since areas near a phrag dominated region are more likely to have phrag. Of interest is esti-

inating both the smoothed probability surface associated with phrag over the entire region as well as the most important predictors of phrag presence. Because the response (phrag presence/absence) is binary and spatial dependence is a critical component of the model, there is a need for a spatial regression model for binary data. This can be easily constructed via an SGLM as discussed below.

An SGLM for binary data may be specified following (1.10) and (1.11).

$$\begin{aligned} Z(\mathbf{s}) \mid p(\mathbf{s}) &\sim \text{Bernoulli}(p(\mathbf{s})) \\ \Phi^{-1}\{p(\mathbf{s})\} &= \boldsymbol{\beta}\mathbf{X}(\mathbf{s}) + w(\mathbf{s}), \end{aligned} \tag{1.13}$$

where $\Phi^{-1}\{p(\mathbf{s})\}$ is the inverse cumulative density function of a standard normal density so $p(\mathbf{s}) = \Phi\{\boldsymbol{\beta}\mathbf{X}(\mathbf{s}) + w(\mathbf{s})\}$. $\mathbf{X}(\mathbf{s})$, as before, is a set of p covariates associated with each site \mathbf{s} , and $\boldsymbol{\beta}$ is a p -dimensional vector of coefficients. w is modeled as a dependent process via a GP or GMRF as discussed in Subsection 1.3.1. The model described by (1.13) is the clipped Gaussian random field (De Oliveira, 2000) since it can equivalently be specified as:

$$Z(\mathbf{s}) \mid Z^*(\mathbf{s}) = \begin{cases} 1 & \text{if } Z^*(\mathbf{s}) > 0 \\ 0 & \text{if } Z^*(\mathbf{s}) \leq 0 \end{cases}$$

The $Z^*(\mathbf{s})$ is then modeled as a linear GP or GMRF as in Section 1.2. This is an intuitive approach to modeling spatial binary data since the underlying latent process may correspond to a physical process that was converted to a binary value due to the detection limits of the measuring device. It may also just be considered a modeling device to help smooth the binary field, when there is reason to assume that the binary field will be smooth. Alternatively, a logit model may be used instead of the probit in the second stage in (1.13) so $\log \left\{ \frac{p(\mathbf{s})}{1-p(\mathbf{s})} \right\} = \boldsymbol{\beta}\mathbf{X}(\mathbf{s}) + w(\mathbf{s})$.

Several of the covariance function parameters are not identifiable. Hence, for a GP model the scale and smoothness parameters are fixed at appropriate values. These identifiability issues are common in SGLMs, but are made even worse in SGLMs for binary data since they contain less information about the magnitude of dependence. A potential advantage of GMRF-based models over GP-based models for binary data is that they can aggregate

pieces of binary information from neighboring regions to better estimate spatial dependence.

Count data

SGLMs are well suited to modeling count data. For example, consider the model:

$$\begin{aligned} Z(\mathbf{s}) \mid \mu(\mathbf{s}) &\sim \text{Poisson}(E(\mathbf{s})\mu(\mathbf{s})) \\ \log(\mu(\mathbf{s})) &= \beta X + w(\mathbf{s}), \end{aligned} \tag{1.14}$$

where $E(\mathbf{s})$ is a known expected count at \mathbf{s} based on other information or by assuming uniform rates across the region, say by multiplying the overall rate by the population at \mathbf{s} .

Example 5. Yang et al. (2009) study infant mortality rates by county in the southern U.S. States of Alabama, Georgia, Mississippi, North Carolina and South Carolina (Health Resources and Services Administration, 2003) between 1998 and 2000. Of interest is finding regions with unusually elevated levels in order to study possible socio-economic contributing factors. Since no interpolation is required here, the purpose of introducing spatial dependence via a GMRF model is to improve individual county level estimates using spatial smoothing by “borrowing information” from neighboring counties. The raw and smoothed posterior means for the maps are displayed in Figure 1.3. Based on the posterior distribution, it is possible to make inferences about questions of interest, such as the probability that the rate exceeds some threshold, and the importance of different socio-economic factors.

Figure 1.3 near here.

The two main examples in Diggle et al. (1998) involve count data, utilizing a Poisson and Binomial model respectively. SGLMs for count data are also explored in Christensen and Waagepetersen (2002), which also develops a Langevin-Hastings Markov chain Monte Carlo approach for simulating from the posterior distribution. Note that count data with reasonably large counts may be modeled well by linear GP models. Given the added complexity of implementing SGLMs, it may therefore be advisable to first try a linear GP model before using an SGLM. However, when there is scientific interest in modeling a known sampling mechanism, SGLMs may be a better option.

Zero-inflated data

In many disciplines, particularly ecology and environmental sciences, observations are often in the form of spatial counts with an excess of zeros (see Welsh et al., 1996). SGLMs provide a nice framework for modeling such processes. For instance, Rathbun and Fei (2006) describe a model for oak trees which determines the species range by a spatial probit model which depends on a set of covariates thought to determine the species' range. Within that range (corresponding to suitable habitat), species counts are assumed to follow an independent Poisson distribution depending on a set of environmental covariates. The model for isopod nest burrows in Agarwal et al. (2002) generates a zero with probability p and a draw from a Poisson with probability $1 - p$. The excess zeros are modeled via a logistic regression and the Poisson mean follows a log-linear model. Spatial dependence is imposed via a GMRF model.

Example 6. Recta et al. (2009) study the spatial distribution of Colorado potato beetle (CPB) populations in potato fields where a substantial proportion of observations were zeros. From the point of view of population studies, it is important to identify the within-field factors that predispose the presence of an adult. The distribution may be seen as a manifestation of two biological processes: incidence, as shown by presence or absence; and severity, as shown by the mean of positive counts. The observation at location \mathbf{s} , $Z(\mathbf{s})$, is decomposed into two variables: incidence (binary) variable $U(\mathbf{s})=1$ if $Z(\mathbf{s}) > 0$, else $U(\mathbf{s}) = 0$, and severity (count) variable $V(\mathbf{s}) = Z(\mathbf{s})$ if $Z(\mathbf{s}) > 0$ (irrelevant otherwise). Separate linear GP models can be specified for the $U(\mathbf{s})$ and $V(\mathbf{s})$ processes, with different covariance structures and means. This formulation allows a great deal of flexibility, including the ability to study spatial dependence in severity, and spatial dependence between severity and incidence, and the potential to relate predictors specifically to severity and independence.

For convenience, we order the data so that the incidences, observations where $U(\mathbf{s}) = 1$, are the first n_1 observations. Hence, there are n_1 observations for V , corresponding to the first n_1 observations of U and $n_1 \leq n$. Our observation vectors are therefore $\mathbf{U} = (U(\mathbf{s}_1), \dots, U(\mathbf{s}_n))^T$ and $\mathbf{V} = (V(\mathbf{s}_1), \dots, V(\mathbf{s}_{n_1}))^T$. Placing this model in an SGLM frame-

work:

$$\begin{aligned} U(\mathbf{s}) &= \text{Bernoulli}(A(\mathbf{s})), \quad \text{so } Pr(U(\mathbf{s}) = 1 \mid w_U(\mathbf{s}), \boldsymbol{\alpha}) = A(\mathbf{s}) \\ V(\mathbf{s}) &= \text{TruncPoisson}(B(\mathbf{s})), \quad \text{so } E(V(\mathbf{s}) \mid w_V(\mathbf{s}), \boldsymbol{\beta}) = \frac{B(\mathbf{s})}{1 - e^{-B(\mathbf{s})}}, \end{aligned} \quad (1.15)$$

TruncPoisson is a truncated Poisson random variable (cf. David and Johnson, 1952) with $Pr(V(\mathbf{s}) = r \mid B(\mathbf{s})) = \frac{B(\mathbf{s})^r e^{-B(\mathbf{s})}}{r!(1 - e^{-B(\mathbf{s})})}$, $r = 1, 2, \dots$, and $w_U(\mathbf{s}), w_V(\mathbf{s}), \boldsymbol{\alpha}, \boldsymbol{\beta}$ are described below. Furthermore, using canonical link functions,

$$\begin{aligned} \log\left(\frac{A(\mathbf{s})}{1 - A(\mathbf{s})}\right) &= X_U(\mathbf{s})\boldsymbol{\alpha} + w_u(\mathbf{s}), \\ \log(B(\mathbf{s})) &= X_V(\mathbf{s})\boldsymbol{\beta} + w_v(\mathbf{s}), \end{aligned} \quad (1.16)$$

where $X_U(\mathbf{s}), X_V(\mathbf{s})$, are vectors of explanatory variables, $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are regression coefficients, and $\mathbf{w}_U = (w_U(\mathbf{s}_1), \dots, w_U(\mathbf{s}_n))^T$ and $\mathbf{w}_V = (w_V(\mathbf{s}_1), \dots, w_V(\mathbf{s}_{n1}))^T$ are modeled via Gaussian processes. The resulting covariance matrices for $\mathbf{w}_U, \mathbf{w}_V$ are Σ_U, Σ_V respectively, specified by exponential covariance functions as described in Section 1.2.1. The parameters of our model are therefore $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \Theta)$, with Θ representing covariance function parameters. Note that the model structure also allows for a flexible cross-covariance relating these two latent processes (Recta et al., 2009), though we do not discuss this here. Priors for $\boldsymbol{\alpha}, \boldsymbol{\beta}, \Theta$ are specified in standard fashion, with a flat prior for the regression parameters and log-uniform priors for the covariance function parameters. Inference and prediction for this model is based on the posterior $\pi(\mathbf{w}_U, \mathbf{w}_V, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Theta \mid \mathbf{U}, \mathbf{V})$. An algorithm for sampling from this distribution is discussed in the next section.

1.3.3 MCMC for SGLMs

For SGLMs, unlike spatial linear models, marginal distributions are not available in closed form for any of the parameters. In other words, for linear spatial models it is possible to study $\pi(\Theta, \boldsymbol{\beta} \mid \mathbf{Z})$ (‘marginalizing out’ \mathbf{w}), while for SGLMs, inference is based on the joint distribution $\pi(\Theta, \boldsymbol{\beta}, \mathbf{w} \mid \mathbf{Z})$. Hence, the dimension of the distribution of interest is typically of the same order as the number of data points. While one can easily construct variable-at-a-

time Metropolis-Hastings samplers for such distributions, the strong dependence among the spatial random effects (w_i s) and the covariance/precision parameters (Θ) typically results in heavily autocorrelated MCMC samplers, which makes sample-based inference a challenge in practice. In addition, expensive matrix operations involved in each iteration of the algorithm continue to be a major challenge for GP-based models, though some recent approaches have been proposed to resolve this (see Section 1.5). Hence not only are standard MCMC algorithms for SGLMs slow mixing, each update can also be computationally expensive, leading to very inefficient samplers.

A general approach to improve mixing in the Markov chain is to update parameters jointly in large blocks. This is a well known approach for improving mixing (see, for example, Liu et al., 1994) and is particularly useful in SGLMs due to the strong dependence among the components of the posterior distribution. However, constructing joint updates effectively can be a challenge, especially in high dimensions. Approaches proposed for constructing joint updates for such models often involve deriving a multivariate normal approximation to the joint conditional distribution of the random effects, $\pi(\mathbf{w} \mid \Theta, \boldsymbol{\beta}, \mathbf{Z})$. We now discuss briefly two general approaches for constructing MCMC algorithms for SGLMs.

Langevin-Hastings MCMC

Langevin-Hastings updating schemes (Roberts and Tweedie, 1996) and efficient reparameterizations for GP-based SGLMs are investigated in Christensen et al. (2006) and Christensen and Waagepetersen (2002). The Langevin-Hastings algorithm is a variant of the Metropolis-Hastings algorithm inspired by considering continuous-time Markov processes that have as their stationary distributions the target distribution. Since it is only possible to simulate from the discrete-time approximation to this process and the discrete-time approximation does not result in a Markov chain with desirable properties (it is not even recurrent), Langevin-Hastings works by utilizing the discrete-time approximation as a proposal for a standard Metropolis-Hastings algorithm (Roberts and Tweedie, 1996). Hence, Langevin-Hastings, like most algorithms for block updating of parameters, is an approach for constructing a multivariate normal approximation that can be used as a proposal for

a block update. The significant potential improvement offered by Langevin-Hastings over simple random walk type Metropolis algorithms is due to the fact that the local property of the distribution, specifically the gradient of the target distribution, is utilized. This can help move the Markov chain in the direction of modes. To illustrate the use of MCMC block sampling approaches based on Langevin-Hastings MCMC, we return to Example 6.

Example 7. Langevin-Hastings MCMC for a two-stage spatial ZIP model: This description follows closely the more detailed discussion in Recta et al. (2009). Simple univariate Metropolis-Hastings updates worked poorly and numerous Metropolis random walk block update schemes for the random effects resulted in very slow mixing Markov chains as well. Hence, Langevin-Hastings updates were applied to the random effects. Borrowing from the notation and description in Christensen et al. (2006) and Diggle and Ribeiro (2007), let

$$\nabla(\gamma) = \frac{\partial}{\partial \gamma} \log \pi(\gamma \mid \dots) = -\gamma + (\Sigma^{1/2})^T \left[\begin{array}{l} \left\{ (U(\mathbf{s}_i) - A(\mathbf{s}_i)) \frac{h'_c(A(\mathbf{s}_i))}{h'(A(\mathbf{s}_i))} \right\}_{i=1}^n \\ \left\{ (V(\mathbf{s}_j) - B(\mathbf{s}_j)) \frac{g'_c(B(\mathbf{s}_j))}{g'(B(\mathbf{s}_j))} \right\}_{j=n+1}^{n+n_1} \end{array} \right]$$

denote the gradient of the log target density evaluated at γ (denoted by $\pi(\gamma \mid \dots)$) where h'_c and g'_c are the partial derivatives of the canonical link functions for the Bernoulli and Truncated Poisson distributions, respectively, h' and g' are partial derivatives of the actual link functions used, and $\Sigma^{1/2}$ is the Choleski factor of the joint covariance matrix for $\mathbf{w}_U, \mathbf{w}_V$. Since we used canonical links in both cases, $\frac{h'_c\{A(\mathbf{s}_i)\}}{h'\{A(\mathbf{s}_i)\}} = \frac{g'_c\{B(\mathbf{s}_i)\}}{g'\{B(\mathbf{s}_i)\}} = 1$ for each i . However, since the Langevin-Hastings algorithm above is not geometrically ergodic (Christensen et al., 2001), we use a truncated version where the gradient is as follows,

$$\nabla^{trunc}(\gamma) = \frac{\partial}{\partial \gamma} \log \pi(\gamma \mid \dots) = -\gamma + (\Sigma^{1/2})^T \left[\begin{array}{l} \{U(\mathbf{s}_i) - A(\mathbf{s}_i)\}_{i=1}^n \\ \{V(\mathbf{s}_j) - (B(\mathbf{s}_j) \wedge H)\}_{j=n+1}^{n+n_1} \end{array} \right], \quad (1.17)$$

where $H \in (0, \infty)$ is a truncation constant. This results in a geometrically ergodic algorithm (Christensen et al., 2001) so a Central Limit theorem holds for the estimated expectations and a consistent estimate for standard errors can be used to provide a theoretically justified stopping rule for the algorithm (Jones et al., 2006). The binomial part of the gradient does not need to be truncated because the expectation, $A(\mathbf{s})$, is bounded. Given that the current value of the random effects vector is γ , the Langevin-Hastings update for the entire

vector of spatial random effects, involves using a multivariate normal proposal, $N(\gamma + \frac{h}{2} \nabla (\gamma)^{trunc}, hI)$, $h > 0$. The tuning parameter h may be selected based on some initial runs, say by adapting it so acceptance rates are similar to optimal rates given in Roberts and Rosenthal (1998).

Unfortunately, the above MCMC algorithm still mixes slowly in practice, which may be due to the fact that Langevin-Hastings works poorly when different components have different variances (Roberts and Rosenthal, 2001); this is certainly the case for the random effects $\mathbf{w}_U, \mathbf{w}_V$. To improve the mixing of the Markov chain we follow Christensen et al. (2006) and transform the vector of random effects into approximately (a posteriori) uncorrelated components with homogeneous variance. For convenience, let $\mathbf{w} = (\mathbf{w}_U^T, \mathbf{w}_V^T)^T$ be the vector of spatial random effects and $\mathbf{Y} = (\mathbf{U}^T, \mathbf{V}^T)^T$. The covariance matrix for $\mathbf{w} \mid \mathbf{Y}$ is approximately $\tilde{\Sigma} = (\Sigma^{-1} + \Lambda(\hat{\mathbf{w}}))^{-1}$, where $\Lambda(\hat{\mathbf{w}})$ is a diagonal matrix with entries $\frac{\partial^2}{(\partial w_j)^2} \log \pi(Y_j \mid w_j)$, and w_j and Y_j are the j th elements of \mathbf{w} and \mathbf{Y} respectively. Also, $\hat{\mathbf{w}}$ is assumed to be a typical value of \mathbf{w} , such as the posterior mode of \mathbf{w} . Let $\tilde{\mathbf{w}}$ be such that $\mathbf{w} = \tilde{\Sigma}^{1/2} \tilde{\mathbf{w}}$. Christensen et al. (2006) suggest updating $\tilde{\mathbf{w}}$ instead of \mathbf{w} , since $\tilde{\mathbf{w}}$ has approximately uncorrelated components with homogeneous variance, simplifying the construction of an efficient MCMC algorithm. For our application, setting $\Lambda(w(x)) = 0$ for all x is a convenient choice and appears to be adequate, though there are alternative approaches (see Christensen et al., 2006). An efficient Metropolis-Hastings algorithm is obtained by updating the transformed parameter vector $\tilde{\mathbf{w}}$ via Langevin-Hastings. The remaining parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \Theta$ may then be updated using simple Metropolis random walk updates. This worked reasonably well in our examples, but certain reparameterizations may also be helpful in cases where mixing is poor.

The above Langevin-Hastings algorithm was found in Recta et al. (2009) to be efficient in a number of real data and simulated examples. Similar efficiencies were seen for SGLMs for count data in Christensen et al. (2006). For prediction at a new set of locations, say $\mathbf{s}_1^*, \dots, \mathbf{s}_m^*$, we would first predict the spatial random effect vectors $\mathbf{w}_U^* = (w_U(\mathbf{s}_1^*), \dots, w_U(\mathbf{s}_m^*))^T$ and $\mathbf{w}_V^* = (w_V(\mathbf{s}_1^*), \dots, w_V(\mathbf{s}_m^*))^T$. Again, sample-based inference provides a simple and effective way to obtain these predictions. Given a sampled vector of $(\mathbf{w}_U, \mathbf{w}_V, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Theta)$ from above, we can easily sample the vectors $\mathbf{w}_U^*, \mathbf{w}_V^* \mid \mathbf{w}_U, \mathbf{w}_V, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Theta$

from the posterior predictive distribution as it is a multivariate normal, similar in form to (1.6). Once these vectors are obtained, the corresponding predictions for the incidence and prevalence (U and V) processes at the new locations by simulating from the corresponding Bernoulli and Truncated Poisson distributions. Many scientific question related to prediction or inference may be easily answered based on the samples produced from the posterior distribution of the regression parameters and spatial dependence parameters, along with the samples from the posterior predictive distribution.

Approximating an SGLM by a linear spatial model

Another approach for constructing efficient MCMC algorithms involves approximating an SGLM by a linear spatial model. This can be done by using an appropriate normal approximation to the non-Gaussian model. Consider an SGLM of the form described in Section 1.3.1. A linear spatial model approximation may be obtained as follows:

$$\begin{aligned} M(\mathbf{s}_i) \mid \boldsymbol{\beta}, w(\mathbf{s}_i) &\sim N(X(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i), c(\mathbf{s}_i)), \quad i = 1, \dots, n, \\ \mathbf{w} \mid \Theta &\sim N(0, \Sigma(\Theta)), \end{aligned} \tag{1.18}$$

with $M(\mathbf{s}_i)$ representing the observation or some transformation of the observation at location \mathbf{s}_i , and $c(\mathbf{s}_i)$ an approximation to the variance of $M(\mathbf{s}_i)$. It is clear that an approximation of the above form results in a joint normal specification for the model. Hence, the approximate model is a linear spatial model of the form described in Section 1.2.1 and the resulting full conditional distribution for the spatial random effects, $\pi(\mathbf{w} \mid \Theta, \boldsymbol{\beta}, \mathbf{Z})$, is multivariate normal. Note that an approximation of this form can also be obtained for SGLMs that have underlying GMRFs. We consider, as an example, the following version of the well known Poisson-GMRF model (Besag et al., 1991),

$$\begin{aligned} Z(\mathbf{s}_i) \mid w(\mathbf{s}_i) &\sim \text{Poi}(E_i \exp\{w(\mathbf{s}_i)\}), \quad i = 1, \dots, n, \\ f(\mathbf{w} \mid \tau) &\propto \tau^{(N-1)/2} \exp\{-\mathbf{w}^T Q(\tau) \mathbf{w}\}, \end{aligned} \tag{1.19}$$

with a proper inverse gamma prior for τ . An approximation to the Poisson likelihood above may be obtained by following Haran (2003) (also see Haran et al., 2003; Haran and Tierney,

2009). By using the transformation $M(\mathbf{s}_i) = \log\left(\frac{Y_i}{E_i}\right)$, and a delta method approximation to obtain $c(\mathbf{s}_i) = \min(1/Y_i, 1/0.5)$, we derive the approximation $M(\mathbf{s}_i) \sim N(w(\mathbf{s}_i), c(\mathbf{s}_i))$. Other accurate approximations of the form (1.18), including versions of the Laplace approximation, have also been studied (cf. Knorr-Held and Rue, 2002; Rue and Held, 2005).

The linear spatial model approximation to an SGLM has been pursued in constructing block MCMC algorithms where the approximate conditional distribution of the spatial random effects can be used as a proposal for a block Metropolis-Hastings update (Haran et al., 2003; Knorr-Held and Rue, 2002). The spatial linear model approximation above also allows for the random effects to be integrated out analytically, resulting in low dimensional approximate marginal distributions for the remaining parameters of the model. The approximate marginal and conditional may be obtained as follows:

1. The linear spatial model approximation of the form (1.18) results in a posterior distribution $\hat{\pi}(\Theta, \boldsymbol{\beta}, \mathbf{w} \mid \mathbf{Z})$. This can be used as an approximation to the posterior distribution of the SGLM ($\pi(\Theta, \boldsymbol{\beta}, \mathbf{w} \mid \mathbf{Z})$.)
2. The approximate distribution $\hat{\pi}(\Theta, \boldsymbol{\beta}, \mathbf{w} \mid \mathbf{Z})$ can be analytically integrated with respect to \mathbf{w} to obtain a low dimensional approximate marginal posterior $\hat{\pi}(\Theta, \boldsymbol{\beta} \mid \mathbf{Z})$. The approximate conditional distribution of \mathbf{w} is also easily obtained in closed form, $\hat{\pi}(\mathbf{w} \mid \Theta, \boldsymbol{\beta}, \mathbf{Z})$.

The general approach above has been explored in the development of heavy-tailed proposal distributions for rejection samplers, perfect samplers and efficient MCMC block samplers (Haran, 2003; Haran and Tierney, 2009). Separate heavy-tailed approximations to the marginal $\hat{\pi}(\Theta, \boldsymbol{\beta} \mid \mathbf{Z})$ and the conditional $\hat{\pi}(\mathbf{w} \mid \Theta, \boldsymbol{\beta}, \mathbf{Z})$ can be used to obtain a joint distribution, which may then be used as a proposal for an independence Metropolis-Hastings algorithm that proposes from the approximation at every iteration (see Haran and Tierney, 2009, for details). This algorithm is uniformly ergodic in some cases (Haran and Tierney, 2009) so rigorous ways to determine MCMC standard errors and the length of the Markov chain (Flegal et al., 2008; Jones et al., 2006) are available. The general framework described above for obtaining a linear spatial model approximation and integrating out the random effects (Haran, 2003; Haran and Tierney, 2009), has been extended in order to obtain fast,

fully analytical approximations for SGLMs and related latent Gaussian models (Rue et al., 2009). While their fully analytical approximation may not have the same degree of flexibility offered by Monte Carlo-based inference, the approach in Rue et al. (2009) completely avoids MCMC and is therefore a promising approach for very efficient fitting of such models especially when model comparisons are of interest or large data sets are involved.

Both the Langevin-Hastings algorithm and MCMC based on ‘linearizing’ an SGLM, along with their variants, result in efficient MCMC algorithms in many cases. An advantage of algorithms that use proposals that depend on the current state of the Markov chain (like the Langevin-Hastings algorithm or other block sampling algorithms discussed here), over fully-blocked independence Metropolis-Hastings approaches is that they take into account local properties of the target distribution when proposing updates. This may result in a better algorithm when a single approximation to the entire distribution is inaccurate. However, if the approximation is reasonably accurate, the independence Metropolis-Hastings algorithm using this approximation can explore the posterior distribution very quickly, does not get stuck in local modes, and can be easily parallelized as all the proposals for the algorithm can be generated independently of each other. Since massively parallel computing is becoming increasingly affordable, this may be a useful feature.

1.3.4 Maximum likelihood inference for SGLMs

It is important to note that even in a non-Bayesian framework, computation for SGLMs is non-trivial. For the SGLMs described in the previous section, the maximum likelihood estimator maximizes the integrated likelihood. Hence, the MLE for Θ, β maximizes

$$\int \mathcal{L}(\Theta, \beta, \mathbf{w}; \mathbf{Z}) d\mathbf{w}, \quad (1.20)$$

with respect to Θ, β . Evaluating the likelihood requires high dimensional integration and the most rigorous approach to solving this problem uses Markov chain Monte Carlo maximum likelihood (Geyer, 1996; Geyer and Thompson, 1992). Alternatives include Monte Carlo Expectation-Maximization (MCEM) (Wei and Tanner, 1990) as explored by Zhang (2002) for

SGLMs though in some cases, fast approximate approaches such as composite likelihood may be useful, as discussed for binary data by Heagerty and Lele (1998). In general, computation for maximum likelihood-based inference for SGLMs may often be at least as demanding as in the Bayesian formulation. On the other hand, the Bayesian approach also provides a natural way to incorporate the uncertainties (variability) in each of the parameter estimates when obtaining predictions and estimates of other parameters in the model.

1.3.5 Section summary

SGLMs provide a very flexible approach for modeling dependent data when there is a known non-Gaussian sampling mechanism at work. Either GP or GMRF models can be used to specify the dependence in a hierarchical framework. Constructing MCMC algorithms for SGLMs can be challenging due to the high dimensional posterior distributions, strong dependence among the parameters, and expensive matrix operations involved in the updates. Recent work suggests that MCMC algorithms that involve block updates of the spatial random effects can result in improved mixing in the resulting Markov chains. Constructing efficient block updating algorithms can be challenging, but finding accurate linear spatial model approximations to SGLMs or using Langevin-Hastings-based approaches may improve the efficiency of the MCMC algorithm in many situations. Matrix operations can be greatly speeded up by using similar tools to those used for linear spatial models. These will be discussed again in Section 1.5 in the context of spatial modeling for large data sets. Appropriate SGLMs and efficient sample-based inference allow for statistical inference for a very wide range of interesting scientific problems.

1.4 Non-Gaussian Markov random field models

Non-Gaussian Markov random field models (NMRFs) provide an alternative to SGLM approaches for modeling non-Gaussian lattice/areal data. These models were first proposed as ‘auto-models’ in Besag (1974) and involve specifying dependence among spatial random variables conditionally, rather than jointly. NMRFs may be useful alternatives to SGLMs,

especially when used to build space-time models, since they can model some interactions in a more direct and interpretable fashion, for example when modeling the spread of contagious diseases from one region to its neighbor, thereby capturing some of the dynamics of a process. GMRFs, as described by equation (1.7), are special cases of Markov random field models. A more general formulation is provided as follows:

$$p(Z_i \mid \mathbf{Z}_{-i}) \propto \exp \left(X_i \boldsymbol{\beta} + \psi \sum_{j \neq i} c_{ij} Z_j \right),$$

with $\psi > 0$. This conditional specification results in a valid joint specification (Besag, 1974; Cressie, 1993) and belongs to the exponential family. Consider a specific example of this for binary data, the autologistic model (Besag, 1974; Heikkinen and Högmänder, 1994):

$$\log \frac{p(Z_i = 1)}{p(Z_i = 0)} = X_i \boldsymbol{\beta} + \psi \sum_{j \neq i} w_{ij} Z_j, \quad \text{where } w_{ij} = 1 \text{ if } i, j \text{ are neighbors, else } 0.$$

For a fixed value of the parameters $\boldsymbol{\beta}$ and ψ , the conditional specification above leads to an obvious univariate Metropolis-Hastings algorithm that cycles through all the full conditional distributions in turn. However, when inference for the parameters is of interest, as is often the case, the joint distribution can be derived via Brook's Lemma (Brook, 1964); also see Cressie (1993, Chapter 6), to obtain

$$p(Z_1, \dots, Z_n) = c(\psi, \boldsymbol{\beta})^{-1} \exp \left(\boldsymbol{\beta} \sum_i X_i Z_i + \psi \sum_{i,j} w_{ij} Z_i Z_j \right),$$

where $c(\psi, \boldsymbol{\beta})$ is the intractable normalizing constant, which is actually a normalizing *function* of the parameters $\psi, \boldsymbol{\beta}$. Other autoexponential models can be specified in similar fashion to the autologistic above, for example the auto-Poisson model for count data (see Ferrándiz et al., 1995), or the centered autologistic model (Caragea and Kaiser, 2009). Specifying conditionals such that they lead to a valid joint specification involves satisfying mathematical constraints like the positivity condition (see, for e.g. Besag, 1974; Kaiser and Cressie, 2000) and deriving the joint distribution for sound likelihood based analysis can be challenging. Also, the resulting dependence can be non-intuitive. For example, Kaiser and Cressie (1997) propose a “Winsorized” Poisson automodel since it is not possible to model positive depen-

dence with a regular Poisson auto-model. In addition, it is non-trivial to extend these models to other scenarios, say to accommodate other sources of information or data types such as zero-inflated data. These challenges, along with the considerable computational burden involved with full likelihood-based inference for such models (as we will see below) have made non-Gaussian Markov random fields more difficult to use routinely than SGLMs.

Since the joint distributions for non-Gaussian MRFs contain intractable normalizing functions involving the parameters of interest, Besag (1975) proposed the ‘pseudolikelihood’ approximation to the likelihood, which involves multiplying the full conditional distributions together. The pseudolikelihood is maximized to provide an approximation to the MLE. This approximation works well in spatial models when the dependence is weak, but under strong dependence, the maximum pseudolikelihood estimate may be a very poor approximation to the maximum likelihood estimate (see Gumpertz et al., 1997). Markov chain Monte Carlo maximum likelihood (Geyer, 1996; Geyer and Thompson, 1992) provides a sound methodology for estimating the MLE via a combination of Markov chain Monte Carlo and importance sampling. This is yet another instance of the enormous flexibility in model specification due to the availability of MCMC-based algorithms. MCMC maximum likelihood is a very general approach for maximum likelihoods involving intractable normalizing functions that also automatically provides sample-based estimates of standard errors for the parameters, though the choice of importance function plays a critical role in determining the quality of the estimates.

Now consider a Bayesian model obtained by placing a prior on the parameters (say $(\boldsymbol{\beta}, \psi)$ for the autologistic model). Since the normalizing function is intractable, the Metropolis-Hastings acceptance ratio cannot be evaluated and constructing an MCMC algorithm for the model is therefore non-trivial. Approximate algorithms replacing the likelihood by pseudolikelihood (Heikkinen and Högmänder, 1994) or by using estimated ratios of normalizing functions have been proposed, but these do not have a sound theoretical basis (though recent work by Atchade et al. (2008) is a first attempt at providing some theory for the latter algorithm). A recent auxiliary variables approach (Møller et al., 2006) has opened up possibilities for constructing a Markov chain with the desired stationary distribution though it requires samples from the *exact distribution* of the auto-model at a fixed value of the pa-

parameter, which is typically very difficult. Perfect sampling algorithms that produce samples from the stationary distribution of the Markov chains do exist for some models such as the autologistic model (Møller, 1999; Propp and Wilson, 1996). Perfect samplers are attractive alternatives to regular MCMC algorithms but are typically computationally very expensive relative to MCMC algorithms; Bayesian inference for non-Gaussian Markov random fields, however, is one area where perfect sampling has potential to be useful. Zheng and Zhu (2008) describe how the Møller et al. (2006) approach can be used to construct an MCMC algorithm for Bayesian inference for a space-time autologistic model. While there has been some recent activity in this area, Bayesian inference and computation for auto-models is still a relatively open area for research.

1.4.1 Section summary

Non-Gaussian MRFs are an alternative to modeling dependent non-Gaussian data. The specification of dependence does not involve link functions and can therefore provide a more direct or intuitive model than SGLMs for some problems. Unfortunately, the mathematical constraints that allow conditional specifications to lead to valid joint specifications of non-Gaussian Markov random field can be complicated and non-intuitive. Also, such models are not easily extended to more complicated scenarios (as discussed in Section 1.5.) Sound inference for non-Gaussian MRFs has been a major hurdle due to intractable normalizing functions that appear in the likelihood. Maximum-likelihood based inference for such models can be done via Markov chain Monte Carlo Maximum Likelihood; Bayesian inference for such models has been an even greater challenge, but recent research in MCMC methods has opened up some promising possibilities. Potential advantages and disadvantages of non-Gaussian MRFs over SGLMs are yet to be fully explored.

1.5 Extensions

The classes of models described in the previous three sections, while very rich, are relatively simple and only have two or three (hierarchical) levels each. Using random field models as

building blocks, a very large number of more flexible models can be developed for tackling an array of important scientific problems. In particular, when there is interest in incorporating mechanistic models and physical constraints, spatial models can be specified via a series of conditional models, capturing the physical characteristics while still accounting for spatial and temporal dependence and various sources of error. For instance, Wikle et al. (2001) describe a series of conditionally specified models to obtain a very flexible space-time model for tropical ocean surface winds.

In similar fashion, joint models for data that are both point-level and areal can be easily specified in a hierarchical framework, providing a model-based approach to dealing with data available at different levels of aggregation (see the discussion of spatial misalignment in Banerjee et al., 2004, Ch.6). Methods for spatial processes that either occur or are observed at multiple scales (Ferreira and Lee, 2007) take advantage of much of the same basic machinery described here. Models for spatiotemporal processes are particularly important since many problems involve space-time data, and it is critical to jointly model both sources of dependence. Spatiotemporal processes are particularly useful when mechanistic models are of interest and when there are interesting dynamics to be captured. Assuming space-time ‘separability’, where it is assumed that the dependencies across time and space do not interact (mathematically, the covariance is multiplicative in the spatial and temporal dimensions), allows the use of Kronecker products and dramatic increases in the speed of matrix computations. However, nonseparability is often not a tenable assumption. Classes of computationally tractable spatial models with stationary, nonseparable covariances have been proposed (Cressie and Huang, 1999; Gneiting, 2002) to address this issue, but in many cases computation can quickly become very challenging with increase in space-time data, particularly for more flexible models. While MCMC-based approaches are feasible in some cases (see Wikle et al., 1998), approximate approaches based on dimension reduction and empirical Bayes estimation combined with Kalman filtering (Wikle and Cressie, 1999) or sequential Monte Carlo-based approaches (Doucet et al., 2001) may be more computationally efficient, for example in fitting multiresolution space-time models (Johannesson et al., 2007). It is also often of interest to model dependencies among multiple space-time variables, along with various sources of missing data and covariate information; multiple variables, covariates and missing data are very common in many studies, particularly in public health and

social science related research. These variables and missing information can just be treated as additional random variables in the model, and added into the sampler, thereby accounting for any uncertainties associated with them. Multivariate models can be explored via both multivariate GMRFs (Carlin and Banerjee, 2003; Gelfand and Vounatsou, 2003) and multivariate or hierarchical Gaussian process models (Royle and Berliner, 1999).

As with many other areas of statistics, a major challenge for spatial modelers is dealing with massive data sets. This is particularly problematic for Gaussian-process based models since matrix operations involving very large matrices can be computationally prohibitive. One set of approaches centers around fast matrix computations that exploit the sparsity of matrices in GMRF models. Rue and Tjelmeland (2002) attempt to approximate GPs by GMRFs to exploit these computational advantages in the GP model case as well, but discover that the approximation does not always work as desired, particularly when the dependence is strong. However, utilizing sparsity does seem to be among the more promising general strategies, as shown in recent work by Cornford et al. (2005) who describe a framework to first impose sparsity and then exploit it in order to speed up computations for large data sets. Furrer et al. (2006) and Kaufman et al. (2008) use covariance tapering while Cressie and Johannesson (2008) use a fixed number of basis functions to construct a non-stationary covariance and exploits the special structure in the resulting covariance matrix to drastically reduce computations. Other approaches that have been investigated in recent years include using a Fourier basis representation of a Gaussian process (see Fuentes, 2007; Paciorek, 2007), and fast likelihood approximations for a Gaussian process model based on products of conditionals (Caragea, 2003; Stein et al., 2004; Vecchia, 1988). Wikle (2002) presents an approach for modeling large scale spatial count data using an SGLM where he uses a spectral-domain representation of the spatial random effects to model dependence. Higdon (1998) describes a kernel mixing approach by utilizing the fact that a dependent process can be created by convolving a continuous white noise process with a convolution kernel. By using a discrete version of this process, for instance with a relatively small set of independent normal random variates, it is possible to model a very large spatial data set. The recently developed ‘predictive process’ approach due to Banerjee et al. (2008) involves working with a low dimensional projection of the original process, thereby greatly reducing the computational burden.

Stationarity and isotropy may be restrictive assumptions for the spatial process, particularly when there are strong reasons to suspect that dependence may be different in different directions or regions. Including anisotropy in GP models is fairly standard (Cressie, 1993); however, it is more difficult to obtain valid nonstationary processes that are also computationally tractable. Several such models for non-stationary processes have been proposed, including spatially-varying kernel based approaches for Gaussian processes (Higdon et al., 1999; Paciorek and Schervish, 2006) and Gaussian Markov random fields, or by convolving a fixed kernel over independent spatial processes with different kernels (Fuentes and Smith, 2001). MCMC plays a central role in fitting these flexible models.

1.5.1 Section summary

Linear Gaussian random field models and SGLMs provide nice building blocks for constructing much more complicated models in a hierarchical Bayesian framework. Much of the benefit of Bayesian modeling and sample-based inference is realized in spatial modeling in situations where a flexible and potentially complicated model is desired, but is much more easily specified via a series of relatively simple conditional models, where one or more of the conditional models are spatial models.

1.6 Summary

The linear Gaussian random fields discussed in Section 1.2 are enormously powerful and flexible modeling tools when viewed in a maximum likelihood or Bayesian perspective, as are some of the non-Gaussian random fields discussed briefly in Section 1.4. Although the discussion here has focused on spatial data, the methods are useful in a very wide array of non-spatial problems, including machine learning and classification (primarily using GP models, see Rasmussen and Williams, 2006), time series, analysis of longitudinal data, and image analysis (primarily GMRF models, see references in Rue and Held, 2005). The major advantage of using a Bayesian approach accrues from the ability to easily specify coherent joint models for a variety of complicated scientific questions and data sets with multiple

sources of variability. Bayesian approaches are particularly useful for the broad classes of models and problems discussed under Sections 1.3 and 1.5.

MCMC and sample-based inference in general greatly expands the set of questions that can be answered when studying spatial data, where for complicated conditionally specified models, asymptotic approximations may be difficult to obtain and hard to justify. Sample-based inference allows for easy assessment of the variability of all estimates, joint and marginal distributions of any subset of parameters, incorporation of non-linearities, multiple sources of variability, allowing for missing data, and the use of scientific information in the model as well as via prior distributions. It is also typically much harder to study likelihood surfaces in maximum likelihood inference, but one can routinely study posterior surfaces — this provides a much more detailed picture regarding parameters in complicated models, which is important when likelihood or posterior surfaces are multimodal, flat or highly skewed. In principle, once the computational problem is solved (a good sampler is implemented), essentially all such questions can be based on the estimated posterior distribution. Since this is a very concise overview of Gaussian random field models for spatial data, we have neither discussed the theory underlying Gaussian random fields nor important principles of exploratory data analysis and model checking, but these can be found in many texts (Banerjee et al., 2004; Cressie, 1993; Diggle and Ribeiro, 2007; Schabenberger and Gotway, 2005).

It is perhaps best to end with some words of caution: While MCMC algorithms and powerful computing have made it possible to fit increasingly flexible spatial models, it is not always clear that there is enough information to learn about the parameters in such models. Zhang (2004) shows that not all parameters are consistently estimable in maximum likelihood-based inference for Gaussian Process SGLMs; however, one quantity is consistently estimable (the ratio of the scale and range parameters in a Matérn covariance), and this is the quantity that drives prediction. It has been noted that the likelihood surface for the covariance parameters in a linear Gaussian process model is relatively flat, which accounts for the large standard errors of the estimates in a maximum likelihood setting (see Handcock and Stein, 1993; Li and Sudjianto, 2005). In our experience, this can also lead to large posterior standard deviations for the parameters in a Bayesian framework, both in

GMRF and GP-based models. Remarkably, prediction based on these models often works extremely well (Stein, 1999, provides a theoretical discussion of this in the context of linear GP models). This can be viewed as both a positive and as a negative: a positive because often researchers are most concerned with prediction, and a negative because inference about the parameters is unreliable, and model validation techniques such as cross-validation cannot detect problems with inference about these parameters, as noted by Zhang (2002). The large uncertainties about parameter values can also be a problem in other situations. For example, the inferred dependence (range parameter) may not differ significantly even when differences in dependence exist, say at different time points or across different subregions. While the discussion above has centered on GP-based models and there has been less theoretical work on studying properties of GMRF-based models, many of the same practical concerns appear to exist, including identifiability issues, especially for the variance components (see, for e.g., Bernardinelli et al., 1995). Prior sensitivity analysis should be a critical component of a careful spatial analysis.

As discussed at length here, Gaussian random field models are very useful but present considerable computational challenges. Some software packages to meet this challenge include the R (Ihaka and Gentleman, 1996) packages **geoR** (Ribeiro and Diggle, 2001), **spBayes** (Finley et al., 2007) for various linear GP models, **geoRglm** (Christensen and Ribeiro, 2002) for GP models for count data, **ramps** (Smith et al., 2008) for joint linear models for point-level and areal data, as well as the **GeoBugs** module in **WinBUGS** (Lunn et al., 2000) for various GMRF models. However, although these software packages are a great advance, there still remains a lot of room for development of new algorithms and software for the classes of spatial models discussed here.

Acknowledgments: The author thanks Tse-Chuan Yang and Sham Bhat for help with drawing maps and Sudipto Banerjee, Brad Carlin, Victor de Oliveira, Don Richards, Margaret Short, Lance Waller, and Tse-Chuan Yang for helpful discussions. Comments from the editor and an anonymous referee also helped improve this article.

Bibliography

- Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions*. Dover, New York.
- Agarwal, D. K., Gelfand, A. E., and Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics*, 9(4):341–355.
- Amit, Y., Grenander, U., and Piccioni, M. (1991). Structural image restoration through deformable templates. *Journal of the American Statistical Association*, 86:376–387.
- Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics.
- Atchade, Y., Lartillot, N., and Robert, C. (2008). Bayesian computation for statistical models with intractable normalizing constants. Technical report, University of Michigan, Department of Statistics.
- Banerjee, S. (2005). On geodetic distance computations in spatial modeling. *Biometrics*, 61(2):617–625.
- Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical modeling and analysis for spatial data*. Chapman & Hall Ltd.
- Banerjee, S., Gelfand, A., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 70:825–848.
- Berger, J. O., De Oliveira, V., and Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456):1361–1374.

- Bernardinelli, L., Clayton, D., and Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine*, 14:2411–2431.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B: Methodological*, 36:192–236.
- Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician: Journal of the Institute of Statisticians*, 24:179–196.
- Besag, J. (1989). Towards Bayesian image analysis. *Journal of Applied Statistics*, 16:395–407.
- Besag, J. (2002). Discussion of “What is a Statistical Model?” by Peter McCullagh. *Annals of Statistics*, 30:1267–1277.
- Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems (Disc: P41-66). *Statistical Science*, 10:3–41.
- Besag, J. and Green, P. J. (1993). Spatial statistics and Bayesian computation (Disc: P53-102). *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 55:25–37.
- Besag, J. and Kempton, R. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics*, 42:231–251.
- Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82:733–746.
- Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (Disc: P21-59). *Annals of the Institute of Statistical Mathematics*, 43:1–20.
- Brook, D. (1964). On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, 51:481–483.
- Brooks, R., Wardrop, D., Thornton, K., Whigham, D., Hershner, C., Brinson, M., and Shortle, J. (2006). Integration of Ecological and Socioeconomic Indicators for Estuaries and Watersheds of the Atlantic Slope. Final Report to US EPA STAR Program, Agreement R-82868401, Washington, DC. Prepared by the Atlantic Slope Consortium, University Park, PA. Technical report, Pennsylvania State University, <http://www.asc.psu.edu>.

- Caragea, P. C. (2003). Approximate likelihoods for spatial processes, Ph.D. Dissertation. Technical report, University of North Carolina, Chapel Hill, NC, Department of Statistics.
- Caragea, P. C. and Kaiser, M. (2009). Autologistic models with interpretable parameters. *Journal of Agricultural, Biological and Environmental Statistics*, 14:281–300.
- Carlin, B. P. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data. In *Bayesian Statistics 7*, pages 45–64. Oxford University Press.
- Christensen, O., Møller, J., and Waagepetersen, R. (2001). Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models. *Methodology and Computing in Applied Probability*, 3(3):309–327.
- Christensen, O. F. and Ribeiro, P. J. (2002). GeoRglm: A package for generalised linear spatial models. *R News*, 2(2):26–28.
- Christensen, O. F., Roberts, G. O., and Sköld, M. (2006). Robust Markov Chain Monte Carlo methods for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15(1):1–17.
- Christensen, O. F. and Waagepetersen, R. (2002). Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics*, 58(2):280–286.
- Clifford, P. (1990). Markov random fields in statistics. In Grimmett, G. R. and Welsh, D. J. A., editors, *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley on His 70th Birthday*, pages 19–32. Clarendon Press [Oxford University Press].
- Cornford, D., Csato, L., and Oppel, M. (2005). Sequential, Bayesian geostatistics: a principled method for large data sets. *Geographical Analysis*, 37(2):183–199.
- Cowles, M., Yan, J., and Smith, B. (2009). Reparameterized and marginalized posterior and predictive sampling for complex bayesian geostatistical models. *Journal of Computational and Graphical Statistics*, 18:262–282.
- Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94:1330–1340.

- Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.
- Cressie, N. A. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York, 2nd. edition.
- Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86:953–963.
- David, F. and Johnson, N. (1952). The truncated Poisson. *Biometrics*, 8(4):275–285.
- De Oliveira, V. (2000). Bayesian prediction of clipped Gaussian random fields. *Computational Statistics & Data Analysis*, 34(3):299–314.
- De Oliveira, V. (2007). Objective Bayesian analysis of spatial data with measurement error. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 35(2):283–301.
- Diggle, P. and Ribeiro, P. (2007). *Model-based Geostatistics*. Springer-Verlag Inc.
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics (Disc: P326–350). *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 47:299–326.
- Doucet, A., De Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag Inc.
- Ferrándiz, J., López, A., Llopis, A., Morales, M., and Tejerizo, M. L. (1995). Spatial interaction between neighbouring counties: Cancer mortality data in Valencia (Spain). *Biometrics*, 51:665–678.
- Ferreira, M. A. R. and De Oliveira, V. (2007). Bayesian reference analysis for Gaussian Markov random fields. *Journal of Multivariate Analysis*, 98(4):789–812.
- Ferreira, M. A. R. and Lee, H. (2007). *Multiscale modeling: a Bayesian perspective*. Springer-Verlag Inc.
- Finley, A., Banerjee, S., and Carlin, B. (2007). spBayes: an R package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software*, 19:1–20.

- Flegal, J., Haran, M., and Jones, G. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23:250–260.
- Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, 102:321–331.
- Fuentes, M. and Smith, R. (2001). A new class of nonstationary spatial models. *North Carolina State University Institute of Statistics Mimeo Series*, 2534.
- Furrer, R., Genton, M., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 5:502–523.
- Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics (Oxford)*, 4(1):11–15.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions in Pattern Analysis and Mach. Intelligence*, 6:721–741.
- Geyer, C. (1996). Estimation and optimization of functions. In *Markov chain Monte Carlo in practice* Eds. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J., pages 241–258. London: Chapman and Hall.
- Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society, Series B*, 54:657–683.
- Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal (Disc: P76-93). *Statistical Science*, 9:55–76.
- Gneiting, T. (2002). Nonseparable, Stationary Covariance Functions for Space-Time Data. *Journal of the American Statistical Association*, 97(458):590–601.
- Golub, G. and Van Loan, C. (1996). *Matrix Computation*. The Johns Hopkins University Press Baltimore, third. edition.
- Grenander, U. and Keenan, D. M. (1989). Towards automated image understanding. *Journal of Applied Statistics*, 16:207–221.

- Gumpertz, M. L., Graham, J. M., and Ristaino, J. B. (1997). Autologistic model of spatial pattern of *Phytophthora* epidemic in bell pepper: Effects of soil variables on disease presence. *Journal of Agricultural, Biological, and Environmental Statistics*, 2:131–156.
- Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics*, 35:403–410.
- Haran, M. (2003). Efficient Perfect and MCMC Sampling Methods for Bayesian Spatial and Components of Variance Models, Ph.D. Dissertation. Technical report, University of Minnesota, Minneapolis, MN, Department of Statistics.
- Haran, M., Bhat, K., Molineros, J., and De Wolf, E. (2009). Estimating the risk of a crop epidemic from coincident spatiotemporal processes. accepted for publication in *The Journal of Agricultural, Biological & Environmental Statistics*.
- Haran, M., Hodges, J. S., and Carlin, B. P. (2003). Accelerating computation in Markov random field models for spatial data via structured MCMC. *Journal of Computational and Graphical Statistics*, 12:249–264.
- Haran, M. and Tierney, L. (2009). Automated Markov chain Monte Carlo for a class of spatial models. Technical report, Pennsylvania State University, Department of Statistics.
- Heagerty, P. J. and Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, 93:1099–1111.
- Health Resources and Services Administration (2003). Health professions, area resource file (arf) system. Technical report, Quality Resource Systems, Inc., Fairfax, VA.
- Heikkinen, J. and Högmänder, H. (1994). Fully Bayesian approach to image restoration with an application in biogeography. *Applied Statistics*, 43:569–582.
- Higdon, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean (Disc: P191-192). *Environmental and Ecological Statistics*, 5:173–190.
- Higdon, D., Swall, J., and Kern, J. (1999). Non-stationary spatial modeling. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A., editors, *Bayesian Statistics 6 – Proceedings of the Sixth Valencia International Meeting*, pages 761–768. Clarendon Press [Oxford University Press].

- Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5:299–314.
- Johannesson, G., Cressie, N., and Huang, H.-C. (2007). Dynamic multi-resolution spatial models. *Environmental and Ecological Statistics*, 14(1):5–25.
- Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). Fixed-width output analysis for Markov Chain Monte Carlo. *Journal of the American Statistical Association*, 101(476):1537–1547.
- Kaiser, M. S. and Cressie, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statistics & Probability Letters*, 35:423–432.
- Kaiser, M. S. and Cressie, N. (2000). The construction of multivariate distributions from Markov random fields. *Journal of Multivariate Analysis*, 73(2):199–220.
- Kaufman, C., Schervish, M., and Nychka, D. (2008). Covariance Tapering for Likelihood-Based Estimation in Large Spatial Data Sets. *Journal of the American Statistical Association*, 103(484):1545–1555.
- Knorr-Held, L. and Rue, H. (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics*, 29(4):597–614.
- Lawson, A. (2008). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Chapman & Hall/CRC.
- Le, N. and Zidek, J. (2006). *Statistical analysis of environmental space-time processes*. Springer Verlag.
- Li, R. and Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in Gaussian kriging models. *Technometrics*, 47(2):111–120.
- Liu, J. S., Wong, W. H., and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 81:27–40.
- Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337.

- Marks, M., Lapin, B., and Randall, J. (1994). Phragmites australis(P. communis): Threats, management, and monitoring. *Natural Areas Journal*, 14(4):285–294.
- Matheron, G. (1971). The theory of regionalized variables and its applications. *Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau*, 5:211.
- McCullagh, P. (2002). What is a statistical model? *The Annals of Statistics*, 30(5):1225–1310.
- McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*. Chapman & Hall Ltd.
- Metropolis, N., Rosenbluth, M., Rosenbluth, A., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Chemical Physics*, 21:1087–1092.
- Møller, J. (1999). Perfect simulation of conditionally specified models. *Journal of the Royal Statistical Society, Series B, Methodological*, 61:251–264.
- Møller, J., Pettitt, A., Berthelsen, K., and Reeves, R. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*, 93(2):451–458.
- Møller, J. and Waagepetersen, R. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall Ltd.
- Mollié, A. (1996). Bayesian mapping of disease. In *Markov chain Monte Carlo in practice* Eds. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J., pages 359–379. London: Chapman and Hall.
- Paciorek, C. (2007). Bayesian Smoothing with Gaussian Processes Using Fourier Basis Functions in the spectralGP Package. *Journal of statistical software*, 19(2).
- Paciorek, C. and Schervish, M. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17:483–506.
- Paulo, R. (2005). Default priors for Gaussian processes. *The Annals of Statistics*, 33(2):556–582.
- Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1-2):223–252.

- Rasmussen, C. and Williams, C. (2006). *Gaussian processes for machine learning*. Springer.
- Rathbun, S. L. and Fei, S. (2006). A spatial zero-inflated Poisson regression model for oak regeneration. *Environmental and Ecological Statistics*, 13(4):409–426.
- Recta, V., Haran, M., and Rosenberger, J. L. (2009). A two-stage model for incidence and prevalence in point-level spatial count data. Technical report, Penn State University, Department of Statistics.
- Ribeiro, P. and Diggle, P. (2001). geoR: A package for geostatistical analysis. *R News*, 1(2):14–18.
- Roberts, G. and Rosenthal, J. (2001). Optimal Scaling for Various Metropolis-Hastings Algorithms. *Statistical Science*, 16(4):351–367.
- Roberts, G. O. and Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 60:255–268.
- Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2:341–363.
- Royle, J. and Berliner, L. (1999). A Hierarchical Approach to Multivariate Spatial Modeling and Prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 4(1):29–56.
- Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 63(2):325–338.
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall.
- Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.
- Rue, H. and Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, 29(1):31–49.

- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments (C/R: P423-435). *Statistical Science*, 4:409–423.
- Saltonstall, K. (2002). Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. *Proceedings of the National Academy of Sciences*, 99(4):2445.
- Schabenberger, O. and Gotway, C. (2005). *Statistical Methods For Spatial Data Analysis*. CRC Press.
- Short, M., Carlin, B., and Gelfand, A. (2005). Bivariate Spatial Process Modeling for Constructing Indicator or Intensity Weighted Spatial CDFs. *Journal of Agricultural, Biological & Environmental Statistics*, 10(3):259–275.
- Smith, B. J., Yan, J., and Cowles, M. K. (2008). Unified geostatistical modeling for data fusion and spatial heteroskedasticity with R package RAMPS. *Journal of Statistical Software*, 25:91–110.
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag Inc.
- Stein, M. L., Chi, Z., and Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 66(2):275–296.
- Sun, D., Tsutakawa, R. K., and Speckman, P. L. (1999). Posterior distribution of hierarchical models using CAR(1) distributions. *Biometrika*, 86:341–350.
- Tibbits, M. M., Haran, M., and Liechty, J. C. (2009). Parallel multivariate slice sampling. Technical report, Pennsylvania State University, Department of Statistics.
- Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society, Series B, Methodological*, 50:297–312.
- Wall, M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121(2):311–324.
- Waller, L. and Gotway, C. (2004). *Applied spatial statistics for public health data*. John Wiley & Sons, New York.

- Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85:699–704.
- Welsh, A., Cunningham, R., Donnelly, C., and Lindenmayer, D. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, 88(1-3):297–308.
- Wikle, C. (2002). Spatial modeling of count data: A case study in modelling breeding bird survey data on large spatial domains. *Spatial Cluster Modelling*, pages 199–209.
- Wikle, C. K., Berliner, L. M., and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5:117–154.
- Wikle, C. K. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86:815–829.
- Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001). Spatiotemporal hierarchical Bayesian modeling tropical ocean surface winds. *Journal of the American Statistical Association*, 96(454):382–397.
- Yan, J., Cowles, M. K., Wang, S., and Armstrong, M. P. (2007). Parallelizing MCMC for Bayesian spatiotemporal geostatistical models. *Statistics and Computing*, 17(4):323–335.
- Yang, T.-C., Teng, H.-W., and Haran, M. (2009). A spatial investigation of the socio-economic predictors of infant mortality in the United States. to appear in *Applied Spatial Analysis and Policy*.
- Yue, Y. and Speckman, P. L. (2009). Nonstationary spatial Gaussian Markov random. Technical report, University of Missouri-Columbia, Department of Statistics.
- Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, 58(1):129–136.
- Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.
- Zheng, Y. and Zhu, J. (2008). Markov Chain Monte Carlo for a Spatial-Temporal Autologistic Regression Model. *Journal of Computational and Graphical Statistics*, 17(1):123–137.

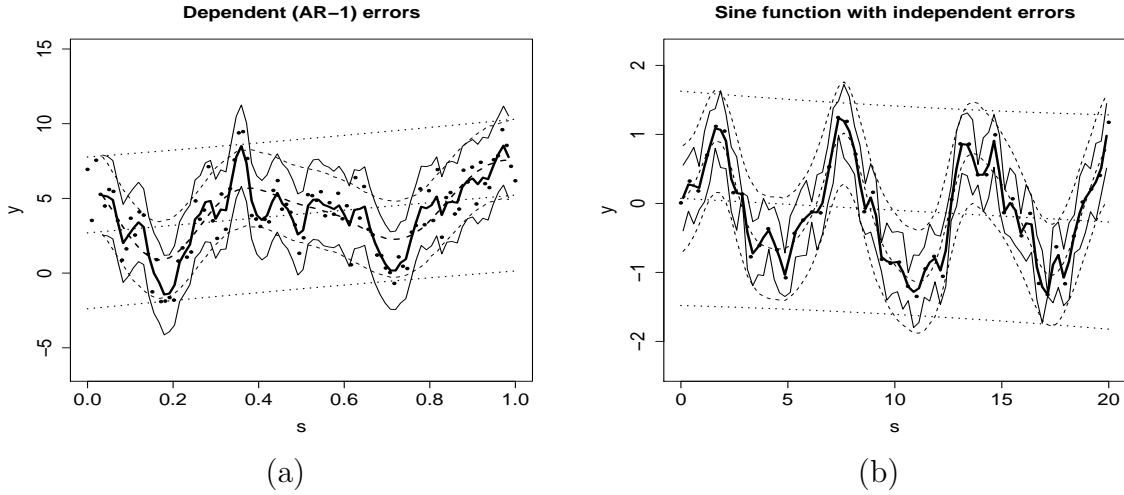


Figure 1.1: Black dots: simulated data. Solid curves: Gaussian process with exponential covariance. Dashed curves: Gaussian process with gaussian covariance. Dotted lines: independent error model. In all cases, the mean and 95% prediction intervals are provided.

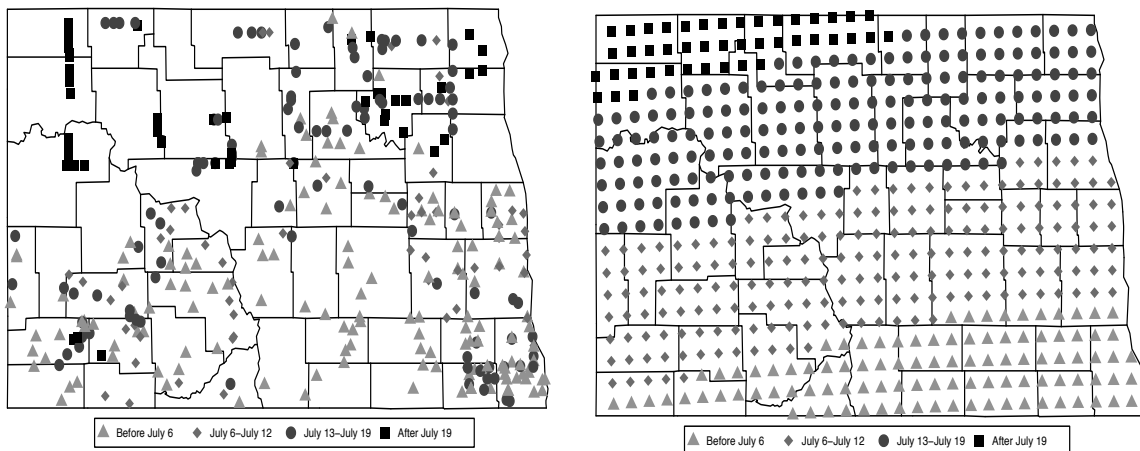


Figure 1.2: (a) raw flowering date. (b) interpolated flowering dates at desired grid locations, using means from posterior predictive distribution from linear Gaussian Process model.

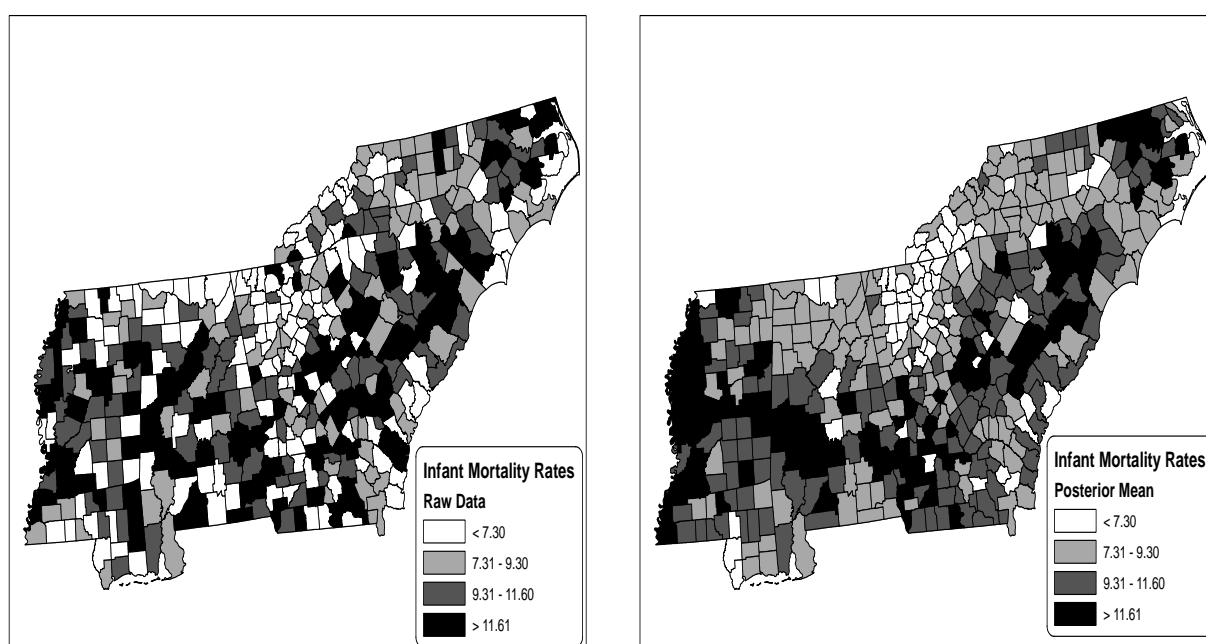


Figure 1.3: Left: raw infant mortality rates. Right: posterior mean infant mortality rates