

# Efficient Statistical Mapping of Avian Count Data

**J. Andrew Royle\***

U.S. Fish and Wildlife Service

Division of Migratory Bird Management

**Christopher K. Wikle**

Dept. of Statistics

University of Missouri

August 2, 2004

**Abstract.** We develop a spatial modeling framework for count data that is efficient to implement in high-dimensional prediction problems. We consider spectral parameterizations for the spatially varying mean of a Poisson model. The spectral parameterization of the spatial process is very computationally efficient, enabling effective estimation and prediction in large problems using Markov chain Monte Carlo techniques.

---

\*11510 American Holly Drive; Laurel, MD 20708; Phone: (301) 497-5673; fax: (301) 497-5706 email: andy\_royle@fws.gov

We apply this model to creating avian relative abundance maps from North American Breeding Bird Survey (BBS) data. Variation in the ability of observers to count birds is modeled as spatially independent noise, resulting in over-dispersion relative to the Poisson assumption. This approach represents an improvement over existing approaches used for spatial modeling of BBS data which are either inefficient for continental scale modeling and prediction or fail to accommodate important distributional features of count data thus leading to inaccurate accounting of prediction uncertainty.

**Key Words:** Breeding Bird Survey, mapping count data, Poisson model, random effects, spatial prediction, spatial modeling, spatial statistics.

# 1 Introduction

The North American Breeding Bird Survey (BBS; Robbins et al. 1986) is conducted during the breeding season of each year by volunteer observers. The sampling unit in the BBS is a roadside route 24.5 miles (39.2 *km*) in length, containing 50 stops. At each stop along a route, birds are counted by sight and sound for a period of 3 minutes. Over 4,000 routes have been surveyed in North America. Data from the BBS are widely used by ecologists to investigate problems in theoretical ecology, management, and conservation. Data are often used for estimation of temporal trends (Link and Sauer 1997a, 1997b), and describing spatial variation in relative abundance (Sauer et al. 1995; Flather and Sauer 1996), for example, to map the relative abundance of a species across its range. Although much effort has been directed toward rigorous statistical modeling of temporal structure, little has been done to develop a similar and consistent framework for assessing questions of spatial variation.

In particular, various procedures have been employed for spatial modeling in BBS investigations including splines (Yang et al. 1995) and kriging (Villard and Maurer 1996). Inverse-distance interpolation (Sauer et al. 1995) is the current approach used to produce relative abundance maps on the BBS analysis website (see <http://www.mbr-pwrc.usgs.gov/bbs/>). However, we feel that spatial modeling of these data is best carried out within a model-based context which properly exploits the properties of count data, and facilitates characterization of uncertainty associated with the resulting maps. For example, BBS data exhibit distinct mean-variance structure such as might be explained by a Poisson assumption and a formal model-based framework for prediction allows maps to be qualified with estimates of uncertainty that are consistent with the mean-variance linkage. This seems especially relevant

for BBS data as resulting maps are often used as “data” in further analysis, such as in the assessment of landscape and habitat influences on bird abundance.

Spatial modeling using the Poisson distribution is common in disease mapping (Best et al. 1998), and the geostatistical framework was laid out by Diggle et al. (1998). Royle, Link and Sauer (2001) investigated this approach for creating relative abundance maps from BBS data. Although this general modeling approach is appealing and relatively straightforward to implement in moderately-sized problems, it is limiting in high-dimensional problems. For example, continental-scale mapping of data from a large number of BBS routes, requiring prediction at a relatively fine resolution is not practically feasible, let alone in an operational context carried out for all species detected by the BBS. Recent developments of so-called spectral approaches for modeling spatial processes (Higdon 1998, Wikle and Cressie 1999, Wikle et al. 2001, Fuentes 2001) have been successfully applied in Gaussian settings, and lead to great efficiencies in estimation and prediction in large problems. The goal of this paper is to exploit spectral representations to develop a framework for spatial modeling of count data that is practical to implement in large problems, such as for the BBS data.

In Section 2, we describe a spatial Poisson model which is suitable for mapping data collected from the BBS. In Section 3 we develop the basic ideas behind the spectral representation of Gaussian spatial processes. In Section 4 we describe how to incorporate the spectral parameterization within the Poisson modeling framework, thus enabling rigorous modeling in large problems. A brief description of estimation using Markov chain Monte Carlo (MCMC) techniques is given in Section 5, and the model is applied to BBS counts in Section 6. Conclusions and discussion of some key advantages and limitations are given in Section 7.

## 2 The Model for BBS Counts

Let  $y(s)$  be the count of birds made on a sample route centered at location  $s$ . We assume that  $y(s)$  has a Poisson distribution with mean  $\lambda(s)$ , denoted as:

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

Thus, two observations, say  $y(s)$  and  $y(s')$ , are independent *conditional on* their means  $\lambda(s)$  and  $\lambda(s')$ . Dependence among the observations will be parameterized as a random spatial effect in the mean. This was the key idea behind the generalized linear prediction approach proposed by Diggle et al. (1998). We model variation in the mean count on the log-scale. Assume that:

$$\log(\lambda(s)) = \mu + z(s) + \eta(s). \quad (1)$$

Here,  $\mu$  is a constant, and  $z(s)$  and  $\eta(s)$  are random effects. Generally,  $\mu$  could be replaced with an arbitrary regression function such as  $\mu(s) = \sum_{j=1}^p \beta_j x_j(s)$  where the  $x_j(s)$  are spatially indexed covariates, such as habitat. This model is more generally useful, but the added generality is not necessary in our development.

We assume that  $z(s)$  is a spatially correlated random variable,

$$z(s) \sim \text{Normal}(0, \sigma_z^2) \quad (2)$$

with  $\text{Corr}(z(s), z(s')) = c_\theta(\|s - s'\|)$  being the correlation between the random effect at locations  $s$  and  $s'$ . Many correlation functions are possible. For example, the Matérn class of correlation models has gained widespread popularity (Stein 1999, pg. 48). We focus on this class of models in Section 3. See Cressie (1991) for other common choices. An important

aspect of this model is that the route effects are assumed to be correlated, allowing for prediction of  $z(s)$  (and  $\lambda(s)$ ) at arbitrary, unsampled, locations.

A critical consideration in modeling BBS data is that observer abilities differ substantially and it is important to accommodate variation in observer bias in modeling BBS data (Sauer et al. 1994). The important consequence of observer variation is that it leads to over-dispersion in the data relative to the Poisson assumption. Failure to adequately model important sources of over-dispersion will yield improper estimates of uncertainty for model parameters, and in a spatial setting, inadequate prediction variances. Link and Sauer (1997a) address modeling over-dispersion that arises as a consequence of observer variation in the context of estimating temporal trends.

In (1),  $\eta(s)$  represents these observer biases, which are assumed to be independent  $\text{Normal}(0, \sigma_\eta^2)$  random variables. The independence assumption may be justified on the grounds that each route has a different observer. Unfortunately, this is not universally so – some observers run multiple routes. However, more detailed information is not generally available. One consequence of the spatial aspect of this problem is that  $\eta(s)$  also contains “small-scale” spatial variation, i.e., from uncorrelated processes, analogous to the *nugget effect* in geostatistics. Here, observer error is only one source of this small-scale variation, but likely the most important source. The measurement error effectively serves as an over-dispersion term and is commonly employed in disease mapping (e.g. Best et al. 1998).

To reduce clutter in formulae given below, we will define  $u(s) \equiv \log(\lambda(s))$ . Let  $\mathbf{u}$  be the  $n \times 1$  vector of observed counts. We may therefore express (1) as:

$$\mathbf{u}|\mathbf{z} \sim \text{Normal}(\mu\mathbf{1} + \mathbf{z}, \sigma_\eta^2\mathbf{I}) \quad (3)$$

where vectors and matrices are of dimension  $n \times 1$  and  $n \times n$ .

## 2.1 Gridding the Data

As expressed in (3), the spatial process  $z(s)$  is assumed to be a continuous space process defined for all  $s$  in the sampling domain. It is useful to consider a subtle generalization of the model which allows for the spatial process to be discrete. For example, defined by counties or other geopolitical strata as is common in disease mapping applications, or physiographic strata in ecological applications. Data might therefore correspond to areal totals over the strata, or perhaps one or more point observations within each stratum.

To accommodate this possibility, we express the model for the  $n$ -vector  $\mathbf{u}$  as

$$\mathbf{u} = \mu \mathbf{1} + \mathbf{Kz} + \boldsymbol{\eta}$$

where  $\mathbf{1}$  is  $n \times 1$ ,  $\mathbf{z}$  is the  $N \times 1$  spatial process,  $\boldsymbol{\eta}$  is an  $n \times 1$  vector of *iid* normal noise, and  $\mathbf{K}$  is an  $n \times N$  matrix describing the association between data and process. Each row of  $\mathbf{K}$  relates the observation at  $s$  to one or more of the elements of  $\mathbf{z}$ .

There is great flexibility in the choice of models for  $\mathbf{z}$  and specification of  $\mathbf{K}$ . Note that (3) is the case where  $\mathbf{K} = \mathbf{I}$  with  $N = n$  and we may retain the continuous-space Gaussian parameterization as specified by (2). Alternatively,  $\mathbf{z}$  may have discrete support and assumed to be a spatial autoregressive process (He and Sun 2000), or general Markov random field. In contrast to modeling spatial structure in  $\mathbf{z}$ , one may incorporate complicated spatial structure into  $\mathbf{K}$  and assume relatively simple structure on  $\mathbf{z}$ . For example, Higdon (1998) describes process convolution models where columns of  $\mathbf{K}$  are weights chosen according to some class of convolution kernels and  $\mathbf{z}$  are assumed to be white noise – that is, uncorrelated

in space.

Estimation and prediction under this model can be carried out using MCMC techniques as described by Diggle et al. (1998). The practical limitation is that updating the random spatial effects requires manipulation of the variance-covariance matrix of  $z(s_1), z(s_2), \dots, z(s_N)$ . Moreover, for prediction, one requires the  $N \times n$  covariance matrix between  $z$  at the prediction and data locations. Within the MCMC framework, this is especially problematic since computations involving these matrices must be carried out at each step of the algorithm, often for many thousands of iterations. However, the multivariate normal representation (2) may be replaced by a much more economical representation, known as a *spectral representation*. This allows for efficient computation under such models. We now turn to development of spectral representations for the spatial process  $\mathbf{z}$ .

### 3 Spectral Representations

The basis of many statistical procedures, including the one outlined above, is a multivariate normal distribution. For illustration, consider some spatial process  $z(\mathbf{s}_i)$  for  $i = 1, \dots, N$ . Assume  $\mathbf{z} \equiv [z(\mathbf{s}_1), \dots, z(\mathbf{s}_N)]'$  has multivariate normal (MVN) distribution:

$$\mathbf{z} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_z). \quad (4)$$

The key to efficient spatial modeling in large problems is parameterization of this multivariate normal in an efficient manner.

Rewrite  $\mathbf{z}$  in terms of a spectral decomposition,

$$\mathbf{z} = \boldsymbol{\Psi}\boldsymbol{\alpha}, \quad (5)$$



where  $\Psi$  is an  $N \times N$  matrix of basis functions  $\Psi \equiv [\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N]$ , where  $\boldsymbol{\psi}_i \equiv [\psi_i(\mathbf{s}_1), \dots, \psi_i(\mathbf{s}_N)]'$ , and  $\boldsymbol{\alpha}$  is an  $N \times 1$  vector of spectral coefficients (projections of the process  $z$  onto the basis functions), with distribution,

$$\boldsymbol{\alpha} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha). \quad (6)$$

If the basis functions are orthogonal, then

$$\boldsymbol{\alpha} = \Psi' \mathbf{z}, \quad (7)$$

and  $\boldsymbol{\Sigma}_\alpha = \Psi' \boldsymbol{\Sigma}_z \Psi$ . Thus, in place of explicitly making assumption (4), we make use of the spectral parameterization (5) and (6).

There are several advantages to writing the spatial process  $\mathbf{z}$  in terms of the spectral process  $\boldsymbol{\alpha}$ . First, the spectral operator often acts as a decorrelator so that  $\boldsymbol{\Sigma}_\alpha$  is relatively sparse, as compared to  $\boldsymbol{\Sigma}_z$ . For example, if  $\Psi$  represents the eigenvectors of the singular-value decomposition of  $\boldsymbol{\Sigma}_z$  (e.g., as in principal components analysis (PCA), or empirical orthogonal function (EOF) (e.g., see Obled and Creutin 1986) analysis in spatial statistics), then  $\boldsymbol{\Sigma}_\alpha$  is diagonal with the diagonal elements proportional to the variance of the spectral component. Similarly, as described in the Appendix, when  $\Psi$  represents Fourier basis functions (sines and cosines in two-dimensions), then  $\boldsymbol{\Sigma}_\alpha$  is approximately diagonal (see results on the distribution of discrete Fourier transform coefficients, e.g., Shumway and Stoffer, 2000, Section T3.12).

Another benefit of the spectral representation is its efficient computation. As discussed in Section 5, this is critical for the implementation of large spatial fields in Bayesian models. At this point, we note that (5) and (7) are matrix representations of the inverse spectral transform and spectral transform, respectively. For example, if  $\Psi$  is the Fourier basis set,

then (7) may be computed using a Fast Fourier Transform (FFT) rather than performing the matrix multiplication; in fact, one need not ever construct  $\Psi$  in this case. Similarly, (5) corresponds to the inverse FFT in this case. The Appendix contains more detail on the implementation of the FFT in this context. One can also use fast spectral transform algorithms for other basis functions such as wavelets.

Finally, although not used explicitly here, one can truncate the basis formulation in some cases to achieve dimension reduction. For example, in the case of PCA or EOF basis functions, many (if not most) of the spectral coefficients are negligible and thus one can truncate the matrix  $\Psi$  to  $N \times m$  by discarding the last  $N - m$  columns of  $\Psi$ , in which case  $\alpha$  is an  $m \times 1$  vector, where  $m < N$ . One may need to include an error term in (5) to account for this truncation. For an example, see Wikle and Cressie (1999).

In a Bayesian implementation, one pays particular attention to the prior distribution on the spectral process,  $\alpha$ . In this example, this process is assumed to be multivariate normal (6). Thus, the specification of the covariance matrix  $\Sigma_\alpha$  is critical. We note that if the original process covariance matrix is parameterized in terms of parameters  $\theta$ ,  $\Sigma_z(\theta)$ , then so must the spectral process,  $\Sigma_\alpha(\theta)$ . In some physical and biological settings, one may have scientific justification for specifying prior distributions for  $\theta$ . For example, in atmospheric science applications one might know *a priori* that there should be a scaling behavior in the variance at different spatial scales (e.g., Wikle et al. 2001). Alternatively, without strong scientific guidance on the covariance structure, one might choose a parametric family for  $\Sigma_z(\theta)$  and thus  $\Sigma_\alpha(\theta)$ . That is, if we choose a family of spatial covariance functions for  $\mathbf{z}$  in terms of parameters  $\theta$  then for specific basis functions we can express the covariance function for  $\alpha$  explicitly in terms of the same parameters  $\theta$ . For example, consider the Matérn class

of covariance functions,

$$C(r) = \phi(ar)^\nu K_\nu(ar), \quad \phi > 0, a > 0, \nu > 0, \quad (8)$$

where  $K_\nu$  is the modified Bessel function,  $\nu$  is related to the degree of smoothness of the spatial process,  $a$  is related to the correlation range, and  $\phi$  is proportional to the variance of the process (e.g., Stein 1999, pg. 48). The corresponding spectral density function at frequency  $\omega$  is,

$$f(\omega) = \frac{2^{\nu-1} \phi \Gamma(\nu + d/2) a^{2\nu}}{\pi^{d/2} (a^2 + \omega^2)^{\nu+d/2}}, \quad (9)$$

where  $d$  is the dimensionality of the spatial process (e.g., Stein 1999, pg. 49). Thus, if one chooses  $\Psi$  to be Fourier basis functions, then (9) suggests the form of  $\Sigma_\alpha(\boldsymbol{\theta})$  (diagonal, with the diagonal element corresponding to frequency  $\omega$  given by (9)) as discussed in the Appendix.

## 4 Spectral Representation of the Poisson Model

For implementation of the spectral model we have nested the sample region into a grid consisting of 6800 equal-area cells with spacing of approximately 20 miles. The grid was generated using the Lambert-Azimuthal equal area projection, but maps presented subsequently have been projected back to longitude/latitude coordinates. For BBS data, it is important to keep in mind that the sample “sites” actually consist of the centers of 24.5 mile roadside routes. Thus, in some sense, there is a fairly coarse canonical resolution to these data. The present grid size was chosen so as to approximate the resolution of the sample route. We will index the prediction grid cells as  $\{r_1, r_2, \dots, r_N\}$  where  $r_j$  is the center point of cell  $j$ .

We assume that a count made at any particular route located within grid-cell  $r_j$  (determined by proximity from the center of the route) has a mean associated with that grid-cell. The grid-cell mean and route-specific deviations from this mean are assumed additive on the log scale. We modify our notation slightly to acknowledge the discrete model grid and association of each observation with grid-cells, letting  $y_{ij}$  be the observer count on route  $i$ , located within grid-cell  $j$ . Thus,  $y_{ij} \sim \text{Poisson}(\lambda_{ij})$ . As before,  $u_{ij} = \log(\lambda_{ij})$ , and  $\mathbf{u}$  is the  $n$ -vector of observed log-means. As in Section 2, we specify a normal prior distribution for  $\mathbf{u}$ :

$$\mathbf{u}|\mathbf{z}, \mu, \sigma_\eta^2 \sim \text{Normal}(\mu\mathbf{1} + \gamma\mathbf{K}\mathbf{z}, \sigma_\eta^2\mathbf{I}) \quad (10)$$

where  $\mathbf{z}$  is the  $6800 \times 1$  grid process,  $\mathbf{K}$  is an  $n \times 6800$  incidence matrix. Thus, the  $i^{\text{th}}$  row of  $\mathbf{K}$ , say  $\mathbf{k}_i$  contains a single 1 corresponding to the grid cell which observation  $i$  is associated. For BBS routes, a better choice of  $\mathbf{K}$  might be based on an allocation of the length of the BBS route into grid-cells that the route traverses so that rows of  $\mathbf{K}$  are weights that sum to 1. However accurate digitized route information is not presently available. The scale parameter  $\gamma$  is incorporated here to facilitate estimation of the variance components, as described below.

We parameterize the spatial process according to the discussion of Section 3:

$$\mathbf{z}|\boldsymbol{\alpha}, \sigma_\epsilon^2 \sim \text{MVN}(\boldsymbol{\Psi}\boldsymbol{\alpha}, \sigma_\epsilon^2\mathbf{I}_{N \times N}). \quad (11)$$

Note that due to the discretization, there may be multiple observations within a grid-cell. Consequently,  $\sigma_\epsilon^2$  is uncorrelated small-scale “among grid-cell” variance – analogous to a *nugget effect* in classical geostatistics. On the other hand,  $\sigma_\eta^2$  represents within grid-cell variation which is primarily that among observers. Both variance components are identifiable

as long as there are some replicate (within grid-cell) samples, though the posterior correlation between the variance components is likely to be high if there are few replicate observations.

For the spectral coefficients, we assume

$$\boldsymbol{\alpha} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta})),$$

where  $\boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta})$  is the spectral correlation matrix implied by assuming that the spatial process follows an exponential covariance with spatial dependence parameter  $\boldsymbol{\theta}$ . Note, the exponential covariance function is a special case of the Matérn class with  $\nu = 1/2$ . Thus, the correlation parameter  $\theta$  is univariate ( $\boldsymbol{\theta} = a$ ) from (8). Note that because of the scaling parameter  $\gamma$  in (10), diagonal elements of  $\boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta})$  are determined by the spectral density in (9), which depends on  $\theta$ . Thus, there is no scale parameter associated with  $\boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta})$  (i.e.  $\phi = 1$ ). This scaling of  $\mathbf{z}$  leads to more stability in estimation of the variance components by MCMC (described in the following section).

Of the remaining parameters,  $\sigma_\epsilon^2$  and  $\sigma_\eta^2$  were assigned inverse-gamma priors:  $\text{IG}(1, 10)$  and  $\text{IG}(.5, 2)$ , respectively, and  $\mu$  was assigned a diffuse normal prior. The prior distribution for  $\theta$  merits some discussion because the use of certain ostensibly “non-informative” priors for  $\theta$  can yield improper posterior distributions (Berger et al. 2001). The solution suggested by Berger et al. (2001), in the Gaussian random field case, is to use the reference prior. In the present context, calculation of the marginal likelihood necessary to obtain the reference prior seems prohibitive, because the likelihood is Poisson and the spatial process model is nested one level lower in the hierarchy. Consequently, we made the pragmatic decision to adopt the reference prior given by Berger et al. (2001), recognizing that this is not the reference prior for the Poisson-Gaussian spatial model under consideration, but that it seems a sensible

solution in the present context because it is proper and does arise under the Gaussian random field model described in Section 3.

## 4.1 Prediction

Our primary goal is prediction of several of the spatially indexed quantities in the model. Of course, the primary benefit of a statistical approach to spatial modeling is that measures of uncertainty (e.g., prediction variances) are useful to quantify map precision. For these purposes, we obtain estimates of posterior means and corresponding posterior variances using MCMC methods as discussed in the following Section.

Since the spatial structure in this model is contained in the spatial process  $z$ , a map of this quantity provides a suitable summary of the spatial structure in the data. Note however that prediction of  $z$  in grid-cells containing data may include a non-zero contribution from the uncorrelated noise component due to  $\sigma_\epsilon^2$ . Alternatively, one might consider predicting the “noiseless” version of  $z_j$ , i.e. just the mean  $\Psi\alpha$ .

It is also useful to predict the expected mean count for a single hypothetical route within *each* grid-cell (whether containing data, or not), i.e.  $\lambda_j$ . Clearly, interest is in an estimate of this quantity *free of observer effects*. Thus for grid-cells containing data we do not wish to predict the  $u_{ij}$  corresponding to a particular route,  $i$ . (in other words, we want the mean of a “typical” observation made within the grid-cell). Since the distribution of the logarithm of the conditional mean of a single count made within grid-cell  $j$  is

$$\log(\lambda_j) \sim \text{Normal}(\mu + z_j, \sigma_\eta^2),$$

the mean of  $y_j$ , unconditional on the observer effect, is:

$$\lambda_j = e^{\mu + z_j + \frac{1}{2}\sigma_\eta^2}, \quad (12)$$

which is the mean averaged over the normal mixing distribution of observer effects – i.e. the usual log-normal result. This quantity, and its variance, are easily estimated using MCMC, as described in the following Section, by generating a Markov chain sequence for a single *unobserved*  $u$  for each grid-cell, and exponentiating those simulated values.

## 5 Model Fitting by Markov Chain Monte Carlo

Posterior analysis of non-normal models containing random effects is relatively straightforward using MCMC techniques. Given the Bayesian formulation in Section 4, our goal is to draw samples from the posterior distribution of unknown quantities (parameters and predictands) in the model. To do this, we set up a sequence of full-conditional distributions for each of the unknowns, and sample repeatedly from these. To be brief, we neglect discussion of this and MCMC in general, instead focusing on the most relevant components of the MCMC algorithm. Gilks et al. (1995) and Gelman et al. (1995) provide general descriptions of MCMC. Clayton (1996) is a good reference on Bayesian analysis of Generalized Linear Mixed Models, and Diggle et al. (1998) provide the formulation of generalized linear prediction in a spatial context.

### 1. Full-conditional for $u$ –

The full-conditional for a particular  $u_{ij}$  (route  $i$  within grid-cell  $j$ ) is the product of the

Poisson likelihood and the normal prior for  $u_{ij}$ :

$$[u_{ij}|\cdot] \propto [y_{ij}|u_{ij}][u_{ij}|\mu, \gamma, z_j, \sigma_\eta^2]$$

which may be conveniently sampled using Metropolis-Hastings algorithm. Because the  $u$ 's are conditionally independent of one another, sampling the whole collection of  $u$ 's may be vectorized. For prediction free of observer effects, it is necessary to sample a single  $u$  for each grid-cell from the prior distribution (i.e. without the likelihood component), under the current parameter values.

## 2. Full-conditional for $z$ –

Since the model components for both  $\mathbf{u}$  and  $\mathbf{z}$  are normal distributions, the full-conditional for  $\mathbf{z}$  is multivariate normal:

$$[\mathbf{z}|\cdot] = \text{MVN}(\mathbf{V}^{-1}\mathbf{f}, \mathbf{V}^{-1}) \quad (13)$$

where

$$\mathbf{V}^{-1} = \left( \frac{\gamma^2}{\sigma_\eta^2} \mathbf{K}'\mathbf{K} + \frac{1}{\sigma_\epsilon^2} \mathbf{I} \right)^{-1}$$

and

$$\mathbf{f} = \frac{\gamma}{\sigma_\eta^2} \mathbf{K}'(\mathbf{u} - \mu\mathbf{1}) + \frac{1}{\sigma_\epsilon^2} \boldsymbol{\Psi}\boldsymbol{\alpha}.$$

The benefit of parameterizing the model such that  $\mathbf{K}$  is an incidence matrix and the  $z$ 's independent, conditional on  $\boldsymbol{\alpha}$  (implied by (11)) is evident. In this case,  $\mathbf{K}'\mathbf{K}$  is diagonal. Consequently,  $\mathbf{V}$  is diagonal and so the  $z$ 's may be sampled directly from *univariate* normal full-conditional distributions. Note grid-cells that do not contain data correspond to all-zero columns in  $\mathbf{K}$  which serve here to zero-out the likelihood contribution in the full-conditional, causing those  $z$ 's to be sampled directly from the prior distribution.



### 3. Full-conditional for $\alpha$ –

The full-conditional for  $\alpha$  is also multivariate normal with  $\mathbf{V}$  and  $\mathbf{f}$  in (13) replaced by:

$$\mathbf{V} = (\Psi' \Psi / \sigma_\epsilon^2 + \Sigma_\alpha(\theta))^{-1} \quad (14)$$

$$= (\mathbf{I} / \sigma_\epsilon^2 + \Sigma_\alpha(\theta))^{-1} \quad (15)$$

and

$$\mathbf{f} = \Psi' \mathbf{z} / \sigma_\epsilon^2. \quad (16)$$

The simplification from (14) to (15) is due to the orthogonality of the spectral basis functions  $\Psi$ . Similarly, (16) is computed by taking the Fourier transform of  $\mathbf{z} / \sigma_\epsilon^2$ , which is a fast operation. Thus, if the prior covariance  $\Sigma_\alpha(\theta)$  is diagonal, then the full-conditional distribution of  $\alpha$  is updated trivially as independent univariate normal distributions.

### 4. Full-conditionals for remaining parameters –

The remaining parameters  $\mu$ ,  $\sigma_\epsilon^2$ , and  $\sigma_\eta^2$  may be sampled from the usual Normal and Gamma full-conditional distributions. Samples of  $\log(\theta)$  were drawn using a Metropolis-Hastings algorithm with a normal proposal distribution for  $\log(\theta)$  with standard-deviation 0.25.

The MCMC algorithm is run for a large number of iterations, say  $M$ , generating a sequence of values for each unknown in the model (parameter, or predictand) from the posterior distribution once the Markov chain has converged (this is discussed in the following Section). For example, to quantify features of the posterior predictive distribution of  $u_j$ , a hypothetical value in grid cell  $j$ , the MCMC output is  $\{u_j^{(1)}, u_j^{(2)}, \dots, u_j^{(M)}\}$ . We may estimate features of the posterior distribution from these sampled data – the posterior mean

is estimated as the sample mean of the MCMC output, the posterior variance (the prediction variance) is the sample variance, etc.. A 95% prediction interval can be computed in several ways; a convenient interval is the lower and upper 2.5<sup>th</sup> percentiles. Posterior characteristics of *functions* of parameters may be estimated by summarizing the appropriate function of MCMC output. For example, the the posterior of the Poisson mean,  $\lambda_j$  may be characterized by simply exponentiating the sampled values of  $u$ .

## 6 Application

We illustrate application of the spectral model using Carolina Wren counts from the BBS during 1998. The 921 BBS routes used in this analysis are shown in (Fig. 1). For these data, eight Markov chains of length 520,000 were generated using the MCMC algorithm described in the previous Section, with the first 20,000 samples used for “burn-in”. We retained every 50<sup>th</sup> sample for analysis (10,000 samples per chain). Each chain was initialized with random draws of the prior parameters  $(\mu, \gamma, \theta, \tau_\eta, \tau_\epsilon)$  from uniform distributions. Convergence was assessed using the Gelman-Rubin diagnostic (Gelman and Rubin 1992). Convergence was indicated for all parameters, the “potential scale rediction” was near 1 for all parameters (1.0037, 1.0027, 1.0031, 1.0004, 1.0034 for  $\mu, \gamma$ , etc..). Post-burnin samples from all 8 chains were used in computing posterior summaries (80,000 samples total). The MCMC history for the parameter  $\theta$  is shown in Figure 2(a) which, although indicating high autocorrelation, does appear stable and well-mixing.

Posterior means and standard deviations for the model parameters are given in Table 1. The salient points of interest are the high indicated spatial correlation – the range parameter

of the exponential correlation function was estimated to be approximately 600 (miles) and its posterior distribution (Fig. 2(b)) is concentrated away from 0. Under the parameterization described in Section 4, the variance among grid-cells is equal to  $\gamma^2\sigma_\epsilon^2$  and the within grid-cell variance (i.e. observer) is  $\sigma_\eta^2$ . Due to the replication at this grid resolution, both variance components should be well-identified; their estimated posterior correlation was  $-0.506$ . We see from the summaries of Table 1 that the within grid-cell variance (that taken to represent observer effects) is considerably larger than the among grid-cell variance component, consistent with our belief that variation among observers is a dominant source of over-dispersion in BBS counts.

The important *predictable* spatial structure in the model was parameterized rather arbitrarily at the resolution of the grid imposed over the survey region. Ideally, observations within grid cells are not correlated beyond that implied by the spectral model, which would be indicated by a small  $\sigma_\epsilon^2$  variance component. From the estimated variance components we can characterize the intra-grid correlation, say  $\rho$ , which is the ratio of the measurement error variance  $\gamma^2\sigma_\epsilon^2$  to the sum of that and the observer variance  $\sigma_\eta^2$ .

The posterior distribution of  $\rho$  is also summarized in Table 1. The posterior mean of 0.283 does suggest the presence of some within-grid structure not modeled as “process” which is a consequence of the discretization. We are primarily interested in characterizing spatial structure in the counts. Point-wise posterior means of the log-scale  $Z$  process (Fig. 3) are useful in this regard, but the expected route count free of observer effects, i.e. the posterior means of  $\lambda(s)$  in (12), is more directly interpretable. This is shown in Fig. 4.

Table 1: Posterior means and standard deviations of model structural parameters

Parameter	Posterior mean	Posterior std. deviation
$\mu$	1.009	0.174
$\sigma_\eta$	0.435	0.044
$\gamma$	1.831	0.137
$\theta$	599.371	77.875
$\sigma_\epsilon$	0.199	0.031
$\rho$	0.283	0.087

## 7 Conclusions and Discussion

Spectral representations of spatial processes provide a computationally efficient means for spatial modeling in high dimensional problems. We have combined this idea with the generalized linear prediction framework of Diggle et al. (1998) to yield an effective method of producing relative abundance maps from the North American Breeding Bird Survey. Figs. 3 and 4 illustrate the very prominent spatial structure in BBS count data indicative of underlying structure in habitat and other processes dictating variation in the distribution of the species within its range. Thus, the notion of improving spatially explicit estimation by combining data within the context of a spatial model appears to be well founded.

The possibility exists to explicitly account for certain factors responsible for spatial variation in relative abundance of a species, such as habitat. However, habitat is not static, and development of suitable resolution data in both space and time is difficult. Nevertheless, such information is easily incorporated into our modeling framework. Even conditional on such information, biogeographic and population dynamic considerations in addition to a host of

other factors (climate, weather, etc..) will induce substantial spatial structure beyond that which one is able to control for. Consequently, we believe the ability to model this residual spatial structure at the continental scale will prove beneficial in ecological investigations involving BBS data.

We believe that a statistical approach to mapping is preferred to more common interpolation procedures such as inverse-distance weighting, spline smoothing, and (Gaussian) kriging which ignore structure inherent in count data, because the assessment of uncertainty is based on a sensible model for count data, one that admits the positive, integer-valued support of counts data, and the inherent mean-variance relationship that is evident in survey data of organisms. Importantly, statistical approaches permit characterization of predictive uncertainty. For the BBS data analyzed here, the relative posterior standard deviation is shown in Fig. (5) where we note the strong relationship between predictive precision and station density. In regions of low sample density, relative precision is much lower, as expected. The ability to characterize predictive uncertainty can be useful for guiding decisions regarding survey design as is commonly done in many environmental and atmospheric science monitoring contexts (e.g., Haas 1992; Cox et al. 1996; Oehlert 1996; Bueso et al. 1998; Nychka and Saltzman 1998; Berliner et al. 1999). Wikle and Royle (2004) consider the problem of model-based design for a waterfowl survey under a spatial Poisson model similar to that considered here.

While the spectral model was implemented in discrete space, at a fairly coarse spatial resolution, the efficiency of spectral models enables prediction at an arbitrarily fine resolution. However, the nature of any particular application may suggest otherwise. In our case, BBS counts are not point observations, but rather some ill-defined spatial aggregation. Thus, we

have chosen a rather coarse model grid. One consequence of this is that it induces replication, or multiple observations within a grid-cell. In doing so, there is some loss of ability to model variation due to small-scale spatial processes. There is a trade-off between modeling this variation as process, by imposing an arbitrarily small grid such that each grid-cell contains a single observation, and the reality that the support of BBS counts is not arbitrarily small.

Because implementation of the spectral Poisson model is very efficient, the framework may be implemented on a larger scale, even in an operational context. Moreover, useful generalizations may be considered including development of a space-time model that unifies the existing trend assessment (Link and Sauer 1997a) with a model for spatial structure.

## **ACKNOWLEDGMENTS**

We thank an anonymous referee for many helpful comments on an earlier draft of this manuscript, and Alan Davenport (U.S. FWS) for GIS support. This research has been supported by a grant from the U.S. Environmental Protection Agency's Science to Achieve Results (STAR) program, Assistance Agreement No. R827257-01-0. Wikle's research was supported by National Science Foundation grants DMS-0139903 and CMG 0222057.

## **APPENDIX: Modeling Spatial Processes in the Spectral Domain**

It is well-known that efficient simulation of spatial processes can be obtained by spectral methods (e.g., Shinozuka and Jan 1972; Mejia and Rodriguez-Iturbe 1974). However, such

methods have not seen extensive use in the modeling of spatial random effects in the lower stages of a hierarchical model for spatial data. The essence of this approach is described below, with particular focus on the case with Fourier spectral basis functions.

Consider a Gaussian spatial process  $z(\mathbf{s}_i)$  for  $i = 1, \dots, N$ . Thus,  $\mathbf{z} \equiv [z(\mathbf{s}_1), \dots, z(\mathbf{s}_N)]'$  has a multivariate normal (MVN) distribution  $\mathbf{z} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_z)$ . We can rewrite this process in terms of a spectral decomposition,  $\mathbf{z} = \boldsymbol{\Psi} \boldsymbol{\alpha}$ , where  $\boldsymbol{\Psi}$  is an  $N \times N$  matrix of basis functions  $\boldsymbol{\Psi} \equiv [\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_N]$ , where  $\boldsymbol{\psi}_i \equiv [\psi_i(\mathbf{s}_1), \dots, \psi_i(\mathbf{s}_N)]'$ , and  $\boldsymbol{\alpha}$  is an  $N \times 1$  vector of spectral coefficients. We assume that the spectral basis functions are complete and orthonormal (i.e.,  $\boldsymbol{\Psi}'\boldsymbol{\Psi} = \boldsymbol{\Psi}\boldsymbol{\Psi}' = \mathbf{I}$ ). It then follows that  $\boldsymbol{\alpha}$  has the distribution  $\boldsymbol{\alpha} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$ , where  $\boldsymbol{\Sigma}_\alpha = \boldsymbol{\Psi}'\boldsymbol{\Sigma}_z\boldsymbol{\Psi}$ .

Consider the case where we expand the  $z(\mathbf{s})$  process in terms of Fourier basis functions (i.e., sines and cosines). Furthermore, we assume that the spatial process is defined on a regular lattice and, for simplicity of illustration, that the spatial domain is defined in one-dimension (analogous results hold for higher spatial dimensions). Then, for locations  $s_i$ ,  $i = 1, \dots, N$  and spatial frequencies  $\omega_k = k/N$  for  $k = 0, \dots, N/2$  ( $N$  even) we have

$$\begin{aligned} z(s_i) &= \sum_{k=0}^{N/2} \alpha^{(1)}(k) \cos(2\pi s_i \omega_k) + \sum_{k=1}^{N/2-1} \alpha^{(2)}(k) \sin(2\pi s_i \omega_k) \\ &= \boldsymbol{\psi}_i^{(1)'} \boldsymbol{\alpha}^{(1)} + \boldsymbol{\psi}_i^{(2)'} \boldsymbol{\alpha}^{(2)} \\ &= \boldsymbol{\psi}_i' \boldsymbol{\alpha} \end{aligned}$$

where  $\boldsymbol{\psi}_i = [\boldsymbol{\psi}_i^{(1)'} \quad \boldsymbol{\psi}_i^{(2)'}]'$ ,  $\boldsymbol{\psi}_i^{(1)} = [\psi_i^{(1)}(\omega_0), \dots, \psi_i^{(1)}(\omega_{N/2})]'$ ,  $\boldsymbol{\psi}_i^{(2)} = [\psi_i^{(2)}(\omega_1), \dots, \psi_i^{(2)}(\omega_{N/2-1})]'$ ,  $\psi_i^{(1)}(\omega_k) = \cos(2\pi s_i \omega_k)$ ,  $\psi_i^{(2)}(\omega_k) = \sin(2\pi s_i \omega_k)$ ,  $\boldsymbol{\alpha} = [\boldsymbol{\alpha}^{(1)'} \quad \boldsymbol{\alpha}^{(2)'}]'$ ,  $\boldsymbol{\alpha}^{(1)} = [\alpha^{(1)}(0), \dots, \alpha^{(1)}(N/2)]'$ , and  $\boldsymbol{\alpha}^{(2)} = [\alpha^{(2)}(1), \dots, \alpha^{(2)}(N/2 - 1)]'$ .

It is well-known that for second-order stationary random processes, the  $\alpha$  coefficients

are nearly uncorrelated and their variances at a given frequency are approximately equal to one-half the power spectral density at that frequency (except for frequency  $\omega_0$  and  $\omega_{N/2}$  at which the variance is equal to the associated power spectral density) (e.g., Shumway 1988). Thus, assuming  $\mathbf{z}$  is second-order stationary with covariance matrix  $\mathbf{\Sigma} = \sigma^2 \mathbf{R}$ , where  $\mathbf{R}$  is the correlation matrix, then  $\text{cov}(\boldsymbol{\alpha}) \approx \sigma^2 \mathbf{D}$ , where  $\mathbf{D}$  is a diagonal matrix with the diagonal elements given by  $[f(\omega_0), 1/2f(\omega_1), \dots, 1/2f(\omega_{N/2}), 1/2f(\omega_1), \dots, 1/2f(\omega_{N/2-1})]'$ , where  $f(\omega_k)$  is the spectral density at frequency  $\omega_k$  corresponding to the correlation function used to construct  $\mathbf{R}$ .

In our case, we parameterize the correlation matrix  $\mathbf{R}(\theta)$  in terms of a spatial dependence parameter  $\theta$ . Thus, the diagonal matrix  $\mathbf{D}(\theta)$  is also a function of  $\theta$ . In the analysis presented here, we specified a Matern covariance matrix with the associated spectral density given by (9).

Although one could create the  $N \times N$  basis function matrix  $\boldsymbol{\Psi}$ , it is not practical computationally for large  $n$ . Rather, we recognize that the operation  $\boldsymbol{\Psi}'\mathbf{z}$  is simply the discrete Fourier transform operation (and  $\boldsymbol{\Psi}\boldsymbol{\alpha}$  is the inverse discrete transform operation.) Such operations are traditionally implemented very efficiently via the Fast Fourier Transform (FFT), in which case one never need actually construct the matrix  $\boldsymbol{\Psi}$ . The use of such an efficient transform algorithm is the key to the implementation of the hierarchical spatial model on very large domains. One complicating issue is that most statistical software packages implement the FFT in its complex representation (e.g., see Shumway 1988). To follow the matrix notation used in the model development, it is convenient to convert the FFT results to a vector of real numbers. Such a conversion is simply a matter of bookkeeping and is fairly simple to implement. We have made available our code (in R and Matlab) to imple-



ment the spectral FFT with Matérn correlation matrices on the second author's website: <http://www.stat.missouri.edu/~wikle> under the link "Spatial Spectral Code".

Finally, as discussed in Dietrich and Newsam (1993) the spectral approach to simulating spectral random fields for a given correlation function is approximate. Recall that the assumption that the covariance matrix for the spectral coefficients is diagonal is based on approximate results. However, although the correlation implied by the spectral approach may not be exactly that of the functional form on which it was based, the differences are typically minor in practice and, regardless, the implied correlation function is a valid correlation model.

## REFERENCES

- Berger, J.O., V. De Oliveira and B. Sansó. 2001. Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, **96**, 1361-1374.
- Berliner, L.M., Lu, Z.-Q., and Snyder, C.. 1999. Statistical Design for Adaptive Weather Observations, *Journal of the Atmospheric Sciences*, **56**, 2536-2552
- Best, N.G., L.A. Waller, A. Thomas, E.M. Conlon, and R.A. Arnold. 1998. Bayesian models for spatially correlated disease and exposure data. *Bayesian Statistics 6*, pp. 1-18. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.
- Bueso, J.M., Angulo, J.M., and Alonso, F.J.. 1998, A state-space model approach to optimum spatial sampling design based on entropy, *Environmental and Ecological Statistics*, **5**, 29-44.

- Clayton, D.G. 1996. Generalized linear mixed models. Pages 275-301 *in* W.R. Gilks, S. Richardson and D.J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
- Cox, D.D, Cox, L.H., and Ensor, K.B.. 1996. Spatial Sampling for the Environment, *Environmental and Ecological Statistics*, **4**, 219-233.
- Cressie, N.A.C.. 1991. *Statistics for Spatial Data*. John Wiley & Sons, New York.
- Dietrich, C.R., and G.N. Newsam. 1993. A fast and exact method for multidimensional Gaussian stochastic simulations. *Water Resources Research*, **29**, 2861-2869.
- Diggle, P.J., J.A. Tawn, and R.A. Moyeed. 1998. Model-based geostatistics (with discussion). *Applied Statistics* **47**, 299-350.
- Flather, C.H. and J.R. Sauer. 1996. Using landscape ecology to test hypotheses about large-scale abundance patterns in migratory birds. *Ecology*, **77**(1), 28-35.
- Fuentes, M. 2001. A high frequency kriging approach. *Environmetrics* **12**, 469-483.
- Gelman, A and D.B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-511.
- Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 1995. *Bayesian Data Analysis*. Chapman & Hall, London.
- Gilks, W.R., S. Richardson, and D.J. Spiegelhalter. 1995. *Markov Chain Monte Carlo in Practice*. Chapman & Hall. London.
- Haas, T.C.. 1992. Redesigning continental-scale monitoring networks, *Atmospheric Envi-*

- ronment*, **26A**, 3323-3333.
- He, Z., and D. Sun. 2000. Hierarchical Bayes estimation of hunting success rates with spatial correlations. *Biometrics* **46**, 192-199.
- Higdon, D. 1998. A process-convolution approach to modeling temperatures in the North Atlantic Ocean. *Journal of Environmental and Ecological Statistics* **5**, 173-190.
- Link, W.A., and J.R. Sauer. 1997a. Estimation of population trajectories from count data. *Biometrics* **53**, 63-72.
- Link, W.A., and J.R. Sauer. 1997b. New Approaches to the Analysis of Population Trends in Land Birds: A comment on statistical methods. *Ecology* **78**, 2632-2634.
- Mejia, J.M. and I. Rodriguez-Iturbe. 1974. On the synthesis of random field sampling from the spectrum: An application to the generation of hydrological spatial processes. *Water Resources Research*, **10**, 705-711.
- Nychka, D., C.K. Wikle and J.A. Royle. 2001. Large spatial prediction problems and nonstationary random fields. In review.
- Obled, C. and J.D. Creutin. 1986. Some developments in the use of empirical orthogonal functions for mapping meteorological fields. *J. Climate and Applied Meteorology* **25**, 1189-1204.
- Oehlert, G.W.. 1996. Shrinking a wet deposition network, *Atmospheric Environment*, **30**, 1347-1357.
- Nychka, D. and Saltzman, N.. 1998. Design of air-quality monitoring networks, in *Case*

- Studies in Environmental Statistics*, eds. D. Nychka, W. Piegorisch, and L. Cox, New York.
- Robbins, C.S., D.A. Bystrak, and P.H. Geissler. 1986. The Breeding Bird Survey: its first fifteen years, 1965-1979. USDOI, Fish and Wildlife Service Resource Publication 157. Washington, D.C.
- Royle, J.A., W.A. Link, and J.R. Sauer. 2001. Statistical mapping of count survey data. In *Predicting Species Occurrences: Issues of Scale and Accuracy*, (Scott, J. M., P. J. Heglund, M. Morrison, M. Raphael, J. Haufler, B. Wall, editors). Island Press. Covello, CA.
- Sauer, J.R., B.G. Peterjohn, and W.A. Link. 1994. Observer differences in the North American Breeding Bird Survey. *Auk* **111**, 50-62.
- Sauer, J.R., G.W. Pendleton, and S. Orsillo. 1995. Mapping of bird distributions from point count surveys. Pages 151-160 in C.J. Ralph, J.R. Sauer, and S. Droege, eds. *Monitoring Bird Populations by Point Counts*, USDA Forest Service, Pacific Southwest Research Station, General Technical Report PSW-GTR-149.
- Shinozuka, M. and C.M. Jan. 1972. Digital simulation of random processes and its applications. *J. Sound Vibration*, **25**, 111-128.
- Shumway, R.H. 1988. *Applied Statistical Time Series Analysis*, Prentice-Hall: New Jersey.
- Shumway, R.H. and D.S. Stoffer. 2000. *Time Series Analysis and Its Applications*. Springer-Verlag, New York.
- Stein, M. 1999. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag: New York.

- Villard M.A. and B.A. Maurer. 1996. Geostatistics as a tool for examining hypothesized declines in migratory songbirds. *Ecology* **77**(1), 59-68.
- Wikle, C.K. and N. Cressie. 1999. A dimension reduction approach to space-time Kalman filtering. *Biometrika* **86**, 815-829.
- Wikle, C.K., R.F. Milliff, D. Nychka and L.M. Berliner. 2001. Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association* **96**, 382-397.
- Wikle, C.K., and J.A. Royle. 2004. Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. *Environmetrics* (in review).
- Yang, K., D. B. Carr and R. J. O'Connor. 1995. Smoothing of Breeding Bird Survey data to produce national biodiversity estimates. *Computing Science and Statistics: Proceeding of the 27th Symposium on the Interface*, pp. 405-409.

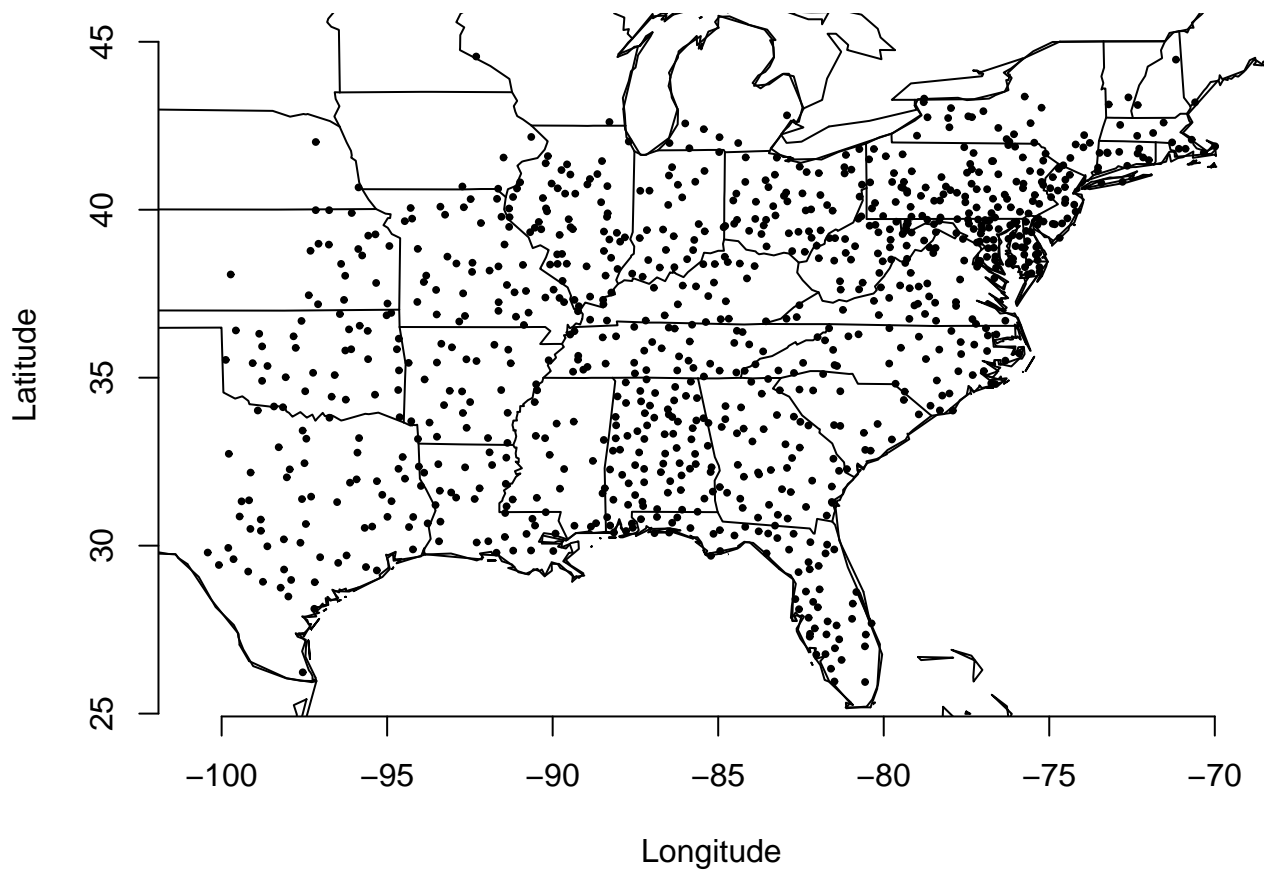


Figure 1: Locations of BBS routes used in Carolina Wren analysis.

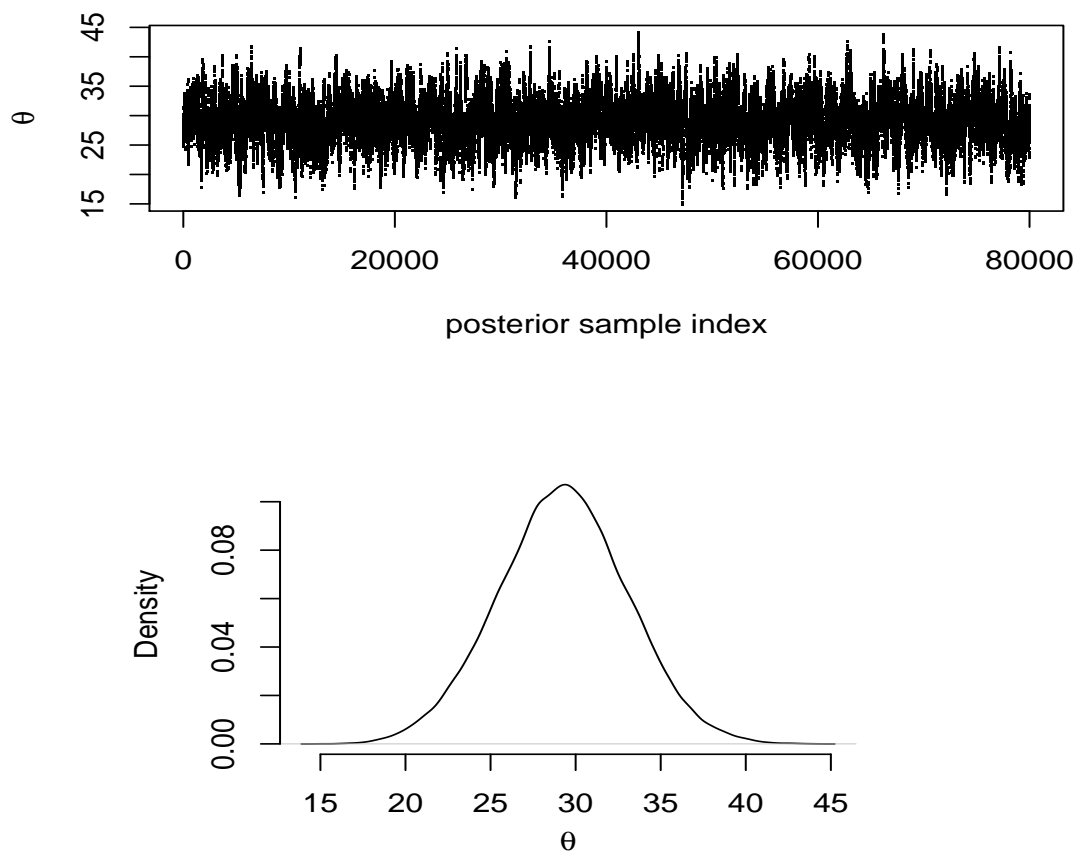


Figure 2: MCMC history for parameter  $\theta$  and its estimated posterior distribution.

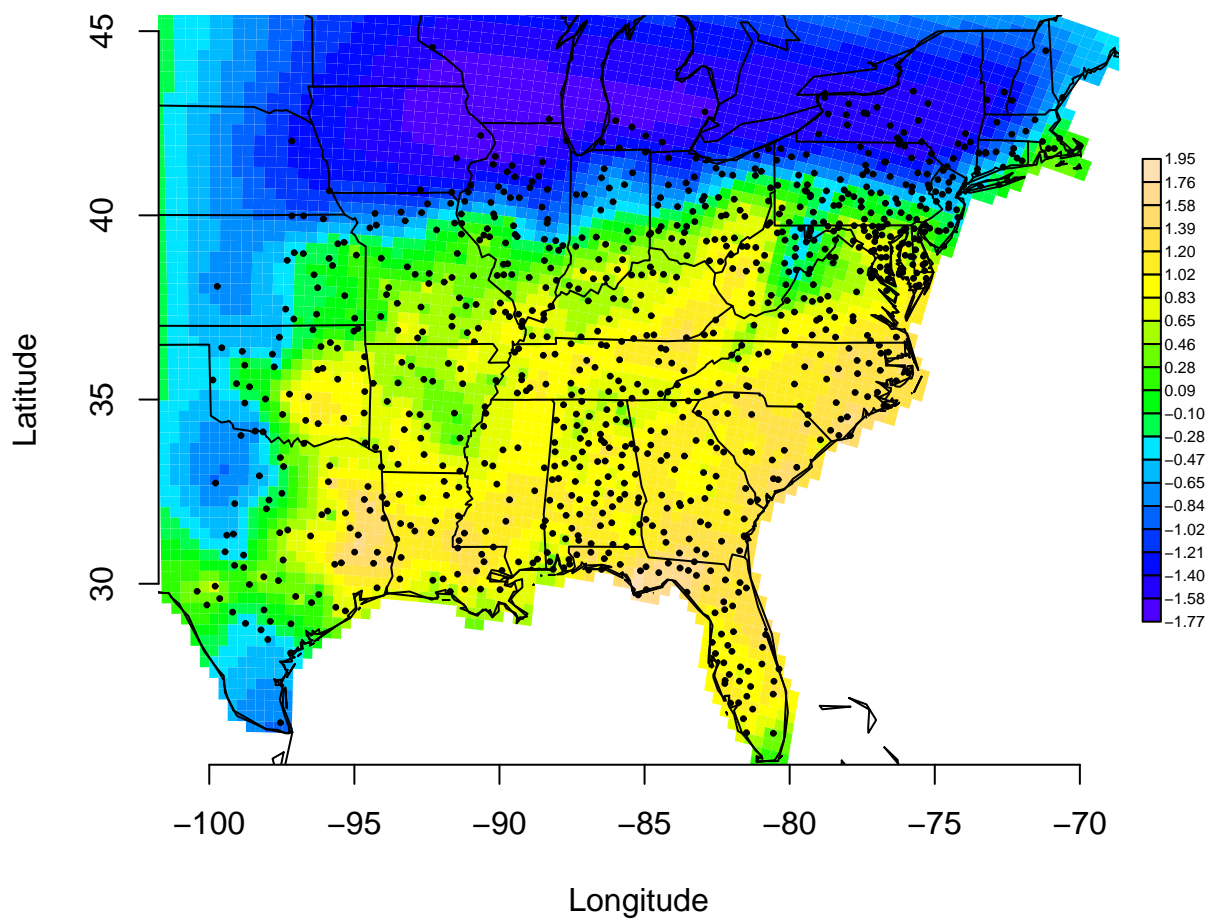


Figure 3: Estimated spatial process (point-wise posterior means)



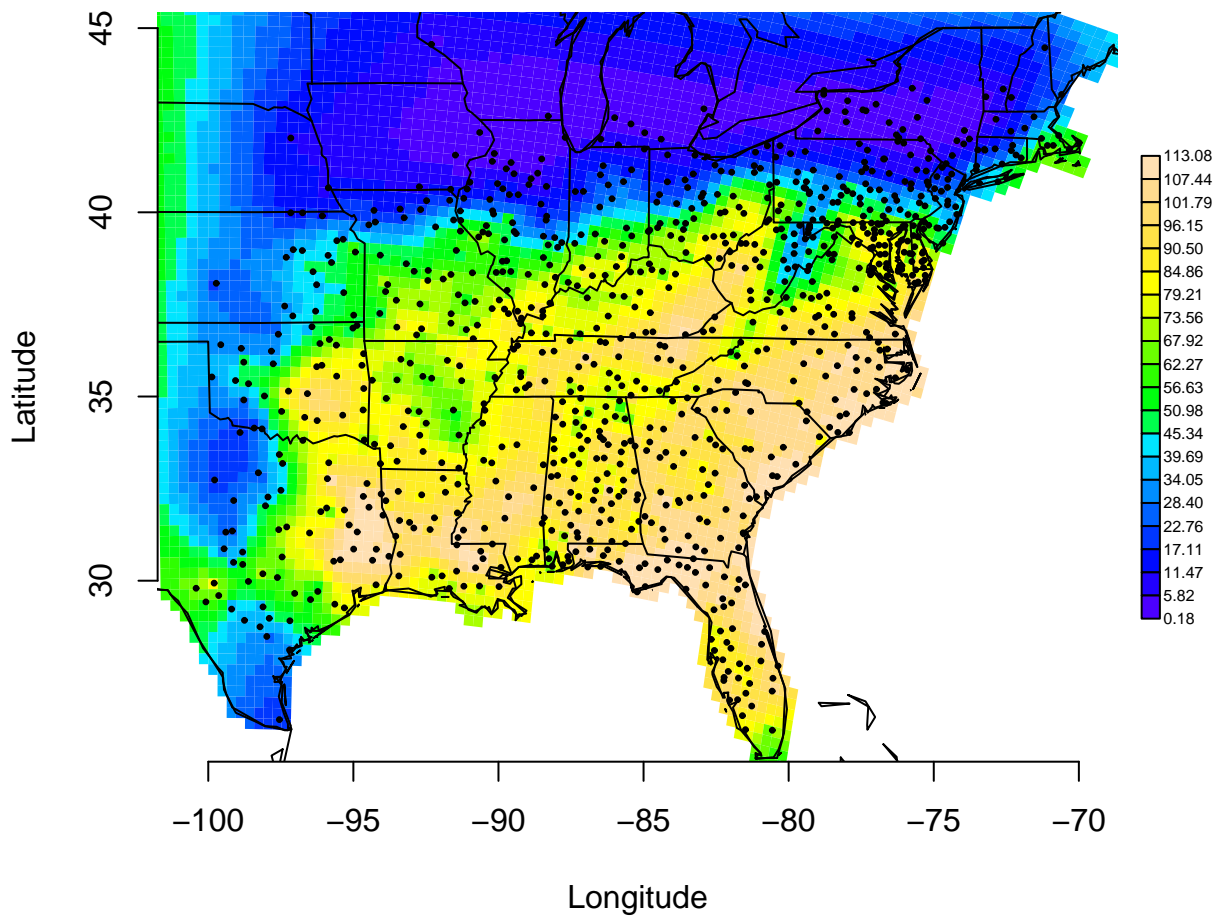


Figure 4: Point-wise posterior means of expected route count.

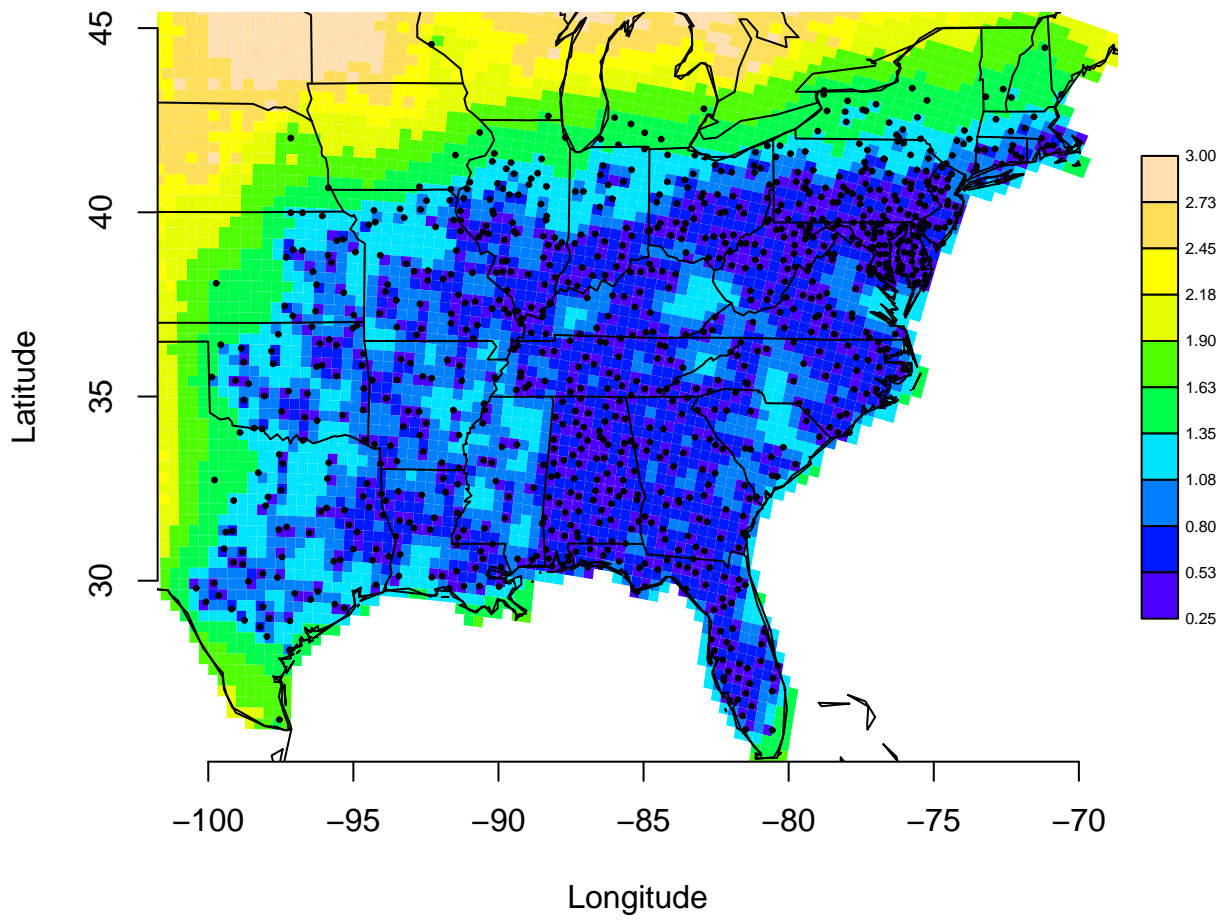


Figure 5: Relative posterior standard deviation of expected route count predictions.