Supplementary materials for this article are available at 10.1007/s13253-016-0274-1.

The Bayesian Group Lasso for Confounded Spatial Data

Trevor J. HEFLEY, Mevin B. HOOTEN, Ephraim M. HANKS, Robin E. RUSSELL, and Daniel P. WALSH

Generalized linear mixed models for spatial processes are widely used in applied statistics. In many applications of the spatial generalized linear mixed model (SGLMM), the goal is to obtain inference about regression coefficients while achieving optimal predictive ability. When implementing the SGLMM, multicollinearity among covariates and the spatial random effects can make computation challenging and influence inference. We present a Bayesian group lasso prior with a single tuning parameter that can be chosen to optimize predictive ability of the SGLMM and jointly regularize the regression coefficients and spatial random effect. We implement the group lasso SGLMM using efficient Markov chain Monte Carlo (MCMC) algorithms and demonstrate how multicollinearity among covariates and the spatial random effect can be monitored as a derived quantity. To test our method, we compared several parameterizations of the SGLMM using simulated data and two examples from plant ecology and disease ecology. In all examples, problematic levels multicollinearity occurred and influenced sampling efficiency and inference. We found that the group lasso prior resulted in roughly twice the effective sample size for MCMC samples of regression coefficients and can have higher and less variable predictive accuracy based on out-of-sample data when compared to the standard SGLMM.
Supplementary materials accompanying this paper appear online.

Key Words: Collinearity; Dimension reduction; Generalized linear mixed model; Spatial confounding.

Trevor J. Hefley (✉), Department of Statistics, Kansas State University, Manhattan, KS 66502, USA (E-mail: *thefley@ksu.edu*). Mevin B. Hooten, U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Departments of Fish, Wildlife, and Conservation Biology, and Statistics, Colorado State University, Fort Collins, CO 80523, USA (E-mail: *mevin.hooten@colostate.edu*). Ephraim M. Hanks, Department of Statistics, Pennsylvania State University, State College, PA, USA (E-mail: *hanks@psu.edu*). Robin E. Russell (E-mail: *rerussell@usgs.gov*) and Daniel P. Walsh (E-mail: *dwalsh@usgs.gov*), U.S. Geological Survey National Wildlife Health Center, Madison, WI, USA.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 22, Number 1, Pages 42–59
DOI: 10.1007/s13253-016-0274-1

1. INTRODUCTION

The SGLMM is an important and widely used tool in fields such as agriculture, econometrics, epidemiology, ecology, and forestry (e.g., Schabenberger and Gotway 2004; Waller and Gotway 2004; Cressie and Wikle 2011; Stroup 2012). In many applications of the SGLMM, the goal is to obtain inference about the regression coefficients by accounting for spatial correlation while simultaneously achieving good predictive ability at locations that were not sampled. For example, researchers may be interested in understanding landscape level risk factors associated with the geographic variability in prevalence of a disease as well as predicting the probability of infection for individuals at locations that were not sampled. As another example, the geographic distribution of a species may depend on complex and difficult to observe ecological processes (e.g., seed dispersal) as well as easy to measure habitat characteristics (e.g., soil type), both of which can generate spatial patterns. In both examples, the SGLMM may be the preferred tool by scientists due to (1) the interpretability resulting from the linear structure of the model and (2) the ability to explicitly model the spatial dependence through the specification of a random effect. Although the SGLMM is widely used, difficulties can occur when covariates with spatial structure are of interest (Clayton et al. 1993; Reich et al. 2006; Hodges and Reich 2010; Paciorek 2010; Hughes and Haran 2013; Hanks et al. 2015; Murakami and Griffith 2015). The goal of our work is to evaluate an alternative specification of the SGLMM for use with spatially structured covariates.

The SGLMM for continuous spatial processes from Gotway and Stroup (1997) and Diggle et al. (1998) can be written as

$$\mathbf{y} \sim [\mathbf{y}|\boldsymbol{\mu}, \psi] \tag{1}$$

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} \tag{2}$$

$$\boldsymbol{\eta} \sim N\left(\mathbf{0}, \sigma_\eta^2 \mathbf{C}(\boldsymbol{\phi})\right), \tag{3}$$

where $\mathbf{y} \equiv (y_1, \ldots, y_n)'$ is a set of n observations at fixed spatial locations $\mathbf{s}_i (i = 1, 2, \ldots, n)$ that arise from an arbitrary distribution belonging to the exponential family denoted by $[\cdot]$. The parameters $\boldsymbol{\mu}$ and ψ are the conditional expected value and dispersion of $[\cdot]$, respectively. The function $g^-1(.)$ transforms the linear predictor $\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}$ so that it has the same support as $\boldsymbol{\mu}$. The linear predictor has three components: $\mathbf{X}$, an $n \times p$ matrix that contains an intercept and covariates, $\boldsymbol{\beta}$, a $p \times 1$ vector of regression coefficients, and $\boldsymbol{\eta} \equiv (\eta_1, \ldots, \eta_n)'$, a zero-mean Gaussian random effect with variance σ_η^2 and spatial correlation matrix $\mathbf{C}(\boldsymbol{\phi})$ with parameters $\boldsymbol{\phi}$.

Regularization methods are widely used with non-spatial regression models to improve predictive performance, computation, and inference (Hoerl and Kennard 1970; Tibshirani 1996). When implementing the SGLMM, regression coefficients are typically not regularized, while the spatial effects are regularized due to the natural shrinkage imposed by the hierarchical Gaussian random effect. Within a Bayesian framework, shrinkage priors have been used to regularize coefficients in non-spatial regression models (e.g., Park and Casella 2008; Kyung et al. 2010; Mallick and Yi 2013; Bhattacharya et al. 2015). In Bayesian

implementations of the SGLMM, regression coefficients $\boldsymbol{\beta}$ are often assigned a mean-zero Gaussian prior, but the prior variance is typically assumed to be a known hyperparameter and chosen to be large, leading to minimal shrinkage of the regression coefficients. The spatial random effects $\boldsymbol{\eta}$, however, are typically assigned a mean-zero Gaussian prior that induces stronger regularization through the assumption of spatial smoothness of the random field and estimation of the prior variance.

When implementing the SGLMM, covariates that are spatially indexed and structured are sometimes collinear with the spatial random effect (Clayton et al. 1993; Reich et al. 2006; Hodges and Reich 2010; Paciorek 2010; Hughes and Haran 2013; Hanks et al. 2015; Murakami and Griffith 2015). This phenomenon is known as spatial confounding and complicates regularization of the regression coefficients in the SGLMM. When covariates are collinear with the spatial random effect, both the regression coefficients and spatial random effect compete to explain variability in the response. In the presence of collinearity, regularizing the regression coefficients will influence the spatial random effect and vice-versa. To ameliorate the explanatory trade-off when regularizing the SGLMM, one could treat the variance of the spatial random effect σ_η^2 as a tuning parameter and regularize both the spatial random effect and regression coefficients. This independent regularization approach would result in two tuning parameters and require two-dimensional optimization for tuning. An alternative to regularizing the regression coefficients and random effects independently is grouped regularization (e.g., Hui et al. 2016). In this paper, we propose a joint shrinkage prior on both regression coefficients and spatial random effects in the SGLMM. This joint prior can be viewed as a Bayesian group lasso prior and provides many of the benefits of regularization to SGLMMs. In particular, we show that the Bayesian group lasso prior can result in greater computational efficiency and the possibility of greater predictive power when compared to the standard independent priors for SGLMMs.

The structure of this paper is outlined as follows. In Sect. 2, we present the Bayesian spatial group lasso (SGL). In Sect. 3, we review spatial confounding and suggest an informative diagnostic that is easy to monitor during MCMC sampling. In Sect. 4, we compare the SGL to the SGLMM using two examples. In Sect. 5, we compare the SGL to the SGLMM using simulated data. In Sect. 6, we conclude with a discussion of the advantages and disadvantages of the SGL.

2. BAYESIAN SPATIAL GROUP LASSO

Motivated by the Karhunen–Loéve expansion, the SGLMM presented in (1)–(3) can be parameterized as

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}(\boldsymbol{\phi})\boldsymbol{\alpha} \tag{4}$$

$$\boldsymbol{\alpha} \sim N\left(\mathbf{0}, \sigma_\eta^2 \boldsymbol{\Lambda}(\boldsymbol{\phi})\right), \tag{5}$$

where $\mathbf{Z}(\boldsymbol{\phi})$ is an $n \times n$ matrix of eigenvectors and $\boldsymbol{\Lambda}(\boldsymbol{\phi})$ is an $\times n$ diagonal matrix of corresponding eigenvalues obtained from a spectral decomposition of the spatial correlation matrix $\mathbf{C}(\boldsymbol{\phi}) = \mathbf{Z}(\boldsymbol{\phi})\boldsymbol{\Lambda}(\boldsymbol{\phi})\mathbf{Z}(\boldsymbol{\phi})'$, and $\boldsymbol{\alpha}$ is an $n \times 1$ vector of basis coefficients (Wikle 2010).

The SGL relies on the hierarchical group lasso from Kyung et al. (2010, pp. 378–379) and the eigen basis vector parametrization (4) of the SGLMM, but uses the priors

$$\boldsymbol{\beta}|\sigma_\beta^2 \sim N\left(\mathbf{0}, \sigma_\beta^2\mathbf{I}\right) \tag{6}$$

$$\boldsymbol{\alpha}|\sigma_\eta^2, \boldsymbol{\phi} \sim N\left(\mathbf{0}, \sigma_\eta^2\boldsymbol{\Lambda}(\boldsymbol{\phi})\right) \tag{7}$$

$$\sigma_\beta^2|\lambda \sim \text{gamma}\left(\tfrac{p+1}{2}, \tfrac{\lambda^2}{2}\right) \tag{8}$$

$$\sigma_\eta^2|\lambda \sim \text{gamma}\left(\tfrac{n+1}{2}, \tfrac{\lambda^2}{2}\right), \tag{9}$$

where λ is a tuning parameter that controls the joint shrinkage of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. When the data model (1) is a Gaussian distribution, $[\boldsymbol{\alpha}|\sigma_\eta^2, \boldsymbol{\phi}]$ and $[\boldsymbol{\beta}|\sigma_\beta^2]$ should also condition on the variance of the data model to guarantee a unimodal full posterior (Park and Casella 2008; Kyung et al. 2010). While the above specification of the group lasso implies two groups, there may be some regression coefficients that we do not shrink and should be grouped differently (e.g., the intercept and covariates without spatial indices and structure). In this case, we write the linear predictor as

$$g(\boldsymbol{\mu}) = \mathbf{W}\boldsymbol{\gamma} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}(\boldsymbol{\phi})\boldsymbol{\alpha} \tag{10}$$

where $\mathbf{W}$ is a $n \times q$ matrix of covariates and $\boldsymbol{\gamma}$ is a $q \times 1$ vector of regression coefficients. Throughout this paper, we assume the prior $\boldsymbol{\gamma} \sim N\left(\mathbf{0}, \sigma_\gamma^2\mathbf{I}\right)$ with the known hyperparameter σ_γ^2.

The tuning parameter λ can be selected using out-of-sample data or by cross-validation. Within the Bayesian context, λ can be estimated using marginal maximum likelihood or by assigning a prior on λ (Mallick and Yi 2013). A common prior is

$$\lambda^2 \sim \text{gamma}\,(r, s) \tag{11}$$

where r and s are known hyperparameters (Park and Casella 2008; Kyung et al. 2010). When the data model (1) is a Gaussian distribution or a probit model, full-conditional distributions for all quantities except parameters that enter into the correlation matrix nonlinearly are available in closed form (Web Appendix A).

Although we have specified the linear predictor of the SGL using eigen basis vectors in (4), alternative basis vectors could be used. For example, dimension reduction may be needed to implement the SGLMM for large data sets. When dimension reduction is desired, alternative basis functions or a subset of eigen basis vectors can be used in place of the full eigen basis expansion (see data example 2; Higdon 2002; Banerjee et al. 2008; Wikle 2010).

3. SPATIAL CONFOUNDING

The term spatial confounding has been used to describe multicollinearity among covariates in $\mathbf{X}$ and the spatial random effect $\boldsymbol{\eta}$ from (2). Using the eigen basis vector parameterization of the SGLMM (4), the potential for multicollinearity among the covariates in

$\mathbf{X}$ and the matrix of basis vectors $\mathbf{Z}(\boldsymbol{\phi})$ is clear. Restricted spatial regression (RSR) is one approach that has been used to alleviate multicollinearity among covariates and the spatial random effect (Reich et al. 2006; Hodges and Reich 2010). To demonstrate RSR, let $\mathbf{P}$ be the orthogonal projection onto the column space of $\mathbf{X}$

$$\mathbf{P} \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \tag{12}$$

Hanks et al. (2015) specified the linear predictor as

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \mathbf{P})\boldsymbol{\eta} + \mathbf{P}\boldsymbol{\eta}. \tag{13}$$

The linear predictor in (13) decomposes the spatial random effect into the two components that are orthogonal and the orthogonal complement to the covariates in $\mathbf{X}$. To alleviate multicollinearity, the RSR approach removes $\mathbf{P}\boldsymbol{\eta}$ from the model resulting in

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\delta} + (\mathbf{I} - \mathbf{P})\boldsymbol{\eta}. \tag{14}$$

By removing $\mathbf{P}\boldsymbol{\eta}$ in (13), all of the variability in the direction of $\mathbf{X}$ explained by $\boldsymbol{\eta}$ is now absorbed by the linear function $\mathbf{X}\boldsymbol{\delta}$ in (14), where

$$\boldsymbol{\delta} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\eta}\,. \tag{15}$$

As a result, the regression coefficients estimated under the SGLMM with RSR are $\boldsymbol{\delta}$ and not $\boldsymbol{\beta}$. Under RSR, all variability in the direction of $\mathbf{X}$ is explained by $\mathbf{X}\boldsymbol{\delta}$, whereas, under the standard model, $\mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ compete to explain variability. As a consequence, $\boldsymbol{\delta}$ can be interpreted as the unconditional relationship between the transformed mean of the response and the linear predictor (Hanks et al. 2015). Several studies have shown that RSR provides computational benefits, such as better mixing and increased effective sample size (ESS; Givens and Hoeting 2012) for $\boldsymbol{\delta}$, when implementing the SGLMM using MCMC sampling (Hughes and Haran 2013; Hanks et al. 2015). A possible alternative or addition to RSR is to use regularization to improve computational efficiency. If regularization improves computational efficiency, then the SGL can be applied to $\boldsymbol{\beta}$ or $\boldsymbol{\delta}$, depending on the component inference is desired for.

In practice, there is little guidance on which quantity, $\boldsymbol{\beta}$ or $\boldsymbol{\delta}$, is appropriate in applied contexts (Hodges and Reich 2010; Hooten et al. 2013; Hanks et al. 2015; Schmidt et al. 2015). In some situations, the specifics of the study might dictate whether $\boldsymbol{\beta}$ or $\boldsymbol{\delta}$ is most appropriate; thus, we proceed assuming that studies may desire inference on the conditional regression parameters $\boldsymbol{\beta}$ from the standard (potentially confounded) SGLMM.

Unlike multicollinearity among covariates, multicollinearity among covariates and the continuous spatial random effect $\boldsymbol{\eta}$ cannot be assessed prior to fitting the model (unless $\boldsymbol{\phi}$ in (3) is known). When using MCMC to fit the SGLMM in the Bayesian context, we have found it helpful to monitor the derived quantity

$$r(\mathbf{x}_j, \mathbf{z}_k)^{(t)} = \frac{\sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right)\left(z_{ik}^{(t)} - \bar{z}_k^{(t)}\right)}{\sqrt{\sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right)^2} \sqrt{\sum_{i=1}^{n} \left(z_{ik}^{(t)} - \bar{z}_k^{(t)}\right)^2}} \tag{16}$$

which is the correlation among the jth covariate of interest ($j = 1, \ldots, p$) and the kth eigenvector ($k = 1, \ldots, n$) obtained from the spectral decomposition of the spatial covariance matrix (3) at the tth MCMC iteration. Note that, if $\boldsymbol{\phi}$ in (3) is known, the eigenvectors associated with the spatial random effect are constant (for all MCMC samples t) and $r(\mathbf{x}_j, \mathbf{z}_k)$ can be calculated prior to fitting the model. Alternatively, when one wishes to assess and compare the magnitude of confounding among eigenvectors, the coefficient of determination may also be useful (i.e., $r(\mathbf{x}_j, \mathbf{z}_k)^2$ from (16)). Although any set of basis vectors could be used, eigen basis vectors represent a naturally ordered orthogonal basis expansion. For confounded data, previous studies have reported that confounding may occur at a single spatial scale (e.g., $|r(\mathbf{x}_j, \mathbf{z}_k)| \gg 0$ for $k = 1$; Hodges and Reich 2010). As an alternative to (16), one could monitor the correlation between $\mathbf{x}$ and $\boldsymbol{\eta}$ (i.e., a linear combination of the columns of $\mathbf{Z}(\boldsymbol{\phi})$); however, as shown by Hodges and Reich (2010) and in our experience (e.g., data examples), confounding is likely to occur at certain spatial scales and can be better diagnosed and understood using eigen basis vectors. When dimension reduction is needed and the basis vectors no longer capture distinct and orthogonal spatial scales, we recommend using the left singular vectors obtained from a singular value decomposition of $\mathbf{Z}(\boldsymbol{\phi})$ in (16; see data example 2).

4. DATA EXAMPLES

We demonstrate the SGLMM and SGL for two binary data examples using the standard (2) and RSR parameterizations (14) of the linear predictor. For both examples, we used a portion of the data to fit the model and the remaining portion to evaluate the predictive performance using the log score (Gelman et al. 2014; Hooten and Hobbs 2015). For each example, we report the mean of the posterior distribution and the equal-tail 95% credible intervals (CI) for regression coefficients and the out-of-sample log score. To assess the computational efficiency of the models, we report the ESS from 10,000 samples obtained from a single MCMC chain. For comparison purposes, we show predictive maps for the study areas and compare all spatial models to a model that did not include the spatial random effect (hereafter referred to as the NS model).

For both examples, we constructed an MCMC algorithm using the closed form full-conditional distributions for all variables with the exception of the range parameter in the covariance function (or basis function). For the range parameter, we used a random walk Metropolis–Hastings algorithm with an acceptance ratio tuned to approximately 40%. For both examples, we drew 10,000 MCMC samples after a burn-in interval of 2000. We include computational details for each example with a more detailed presentation of the material and the necessary computer code to reproduce all results and figures in R (R Core Team 2015; Web Appendix B, C, D, and E). All run times resulted from running the code contained in

the Web Appendix B, C, and D on a laptop computer with a 2.8-GHz quad-core processor, 16 GB of RAM, and optimized basic linear algebra subprograms.

4.1. EXAMPLE 1: PREDICTING THE DISTRIBUTION OF A SPECIES

Many ecological studies aim to understand how covariates influence the presence or abundance of a species at locations that were sampled as well as make predictions at locations that were not sampled (Hefley and Hooten 2016). Spatial regression models are a popular tool used to accomplish this goal. For example, Hooten et al. (2003) used a binary spatial regression model to predict the probability that a plant, pointed-leaved tick trefoil (*Desmodium glutinosum*), occurs in 10×10 m plots across a 328-ha area from presence–absence data collected at 216 plots (Fig. 1a). A common problem when predicting the occurrence of a species is that data are sparse relative to the prediction domain. For this example, only 0.66% of the plots were sampled within the prediction domain (Fig. 1a). In this application, Hooten et al. (2003) used a spatial random effect to increase the predictive ability of a binary regression model and to account for spatial correlation. A suitable SGLMM for presence–absence data is

$$\begin{aligned} \mathbf{y}|\gamma_0, \beta_1, \beta_2, \boldsymbol{\eta} &\sim \text{Bernoulli}\left(\Phi(\gamma_0 + \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \boldsymbol{\eta})\right) \\ \boldsymbol{\eta}|\sigma_\eta^2, \phi &\sim N\left(\mathbf{0}, \sigma_\eta^2\mathbf{C}(\phi)\right), \end{aligned} \tag{17}$$

where $\mathbf{y}$ is an $n \times 1$ vector with elements 1 if the species is present and 0 if the species is absent at a sampled location, Φ is the standard normal cumulative distribution function applied element-wise (i.e., probit link function), γ_0 is the intercept, β_1 and β_2 are regression coefficients, $\mathbf{x}_1$ and $\mathbf{x}_2$ are the centered and scaled covariates (aspect and relative elevation; Fig. 1b; Web Figure 1), and $\boldsymbol{\eta}$ is the spatial random effect assuming an exponential correlation function (i.e., $c_{ij}(\phi) = e^{-\frac{d_{ij}}{\phi}}$ where $c_{ij}(\phi)$ is the element in the ith row and jth column of $\mathbf{C}(\phi)$ in (17) and d_{ij} is the distance in km between locations i and j). The correlation between the two spatial covariates ($\mathbf{x}_1$ and $\mathbf{x}_2$) was -0.13. For the SGLMM and SGL, we used the following priors: $\gamma_0 \sim N(0, 10)$ and $\phi \sim \text{uniform}(0, 2.86)$ (note, 2.86 km is the diagonal distance across the study area). For the SGLMM, we used the priors: $\boldsymbol{\beta} \sim N(0, 10\mathbf{I})$ and $\frac{1}{\sigma_\eta^2} \sim \text{gamma}(2.1, 1.1)$ (note, $E(\sigma_\eta^2) = 1$ and $\text{Var}(\sigma_\eta^2) = 10$). Our motivation for the prior on σ_η^2 was to let the data determine the amount of regularization on $\boldsymbol{\eta}$ by using a diffuse prior. For the SGL, we used the prior $\lambda^2 \sim \text{gamma}\ (0.1, 0.1)$ (note, $E(\lambda^2) = 1$ and $\text{Var}(\lambda^2) = 10$). For the NS model, we used the priors $\gamma_0 \sim N(0, 10)$ and $\boldsymbol{\beta} \sim N(\mathbf{0}, 10\mathbf{I})$. We randomly selected 75% ($n = 162$) of the data to fit all models and used the remaining 25% ($n = 54$) of the data to evaluate the predictive performance.

The predicted probability of occurrence ranged from 0.04 to 0.84 using the SGLMM and 0.09–0.65 using the SGL (Fig. 1c, d). The smaller range of predicted probabilities of occurrence is a result of the regularization of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ (cf., Fig. 2a, b, Web Figures 2 and 3). The mean and 95% CI from the posterior distributions of the aspect covariate (β_1) were similar for the SGLMM and SGLMM with RSR (Fig. 2a). The posterior means of β_1 (aspect) obtained from the SGL and SGL with RSR were closer to zero, and the 95% CIs

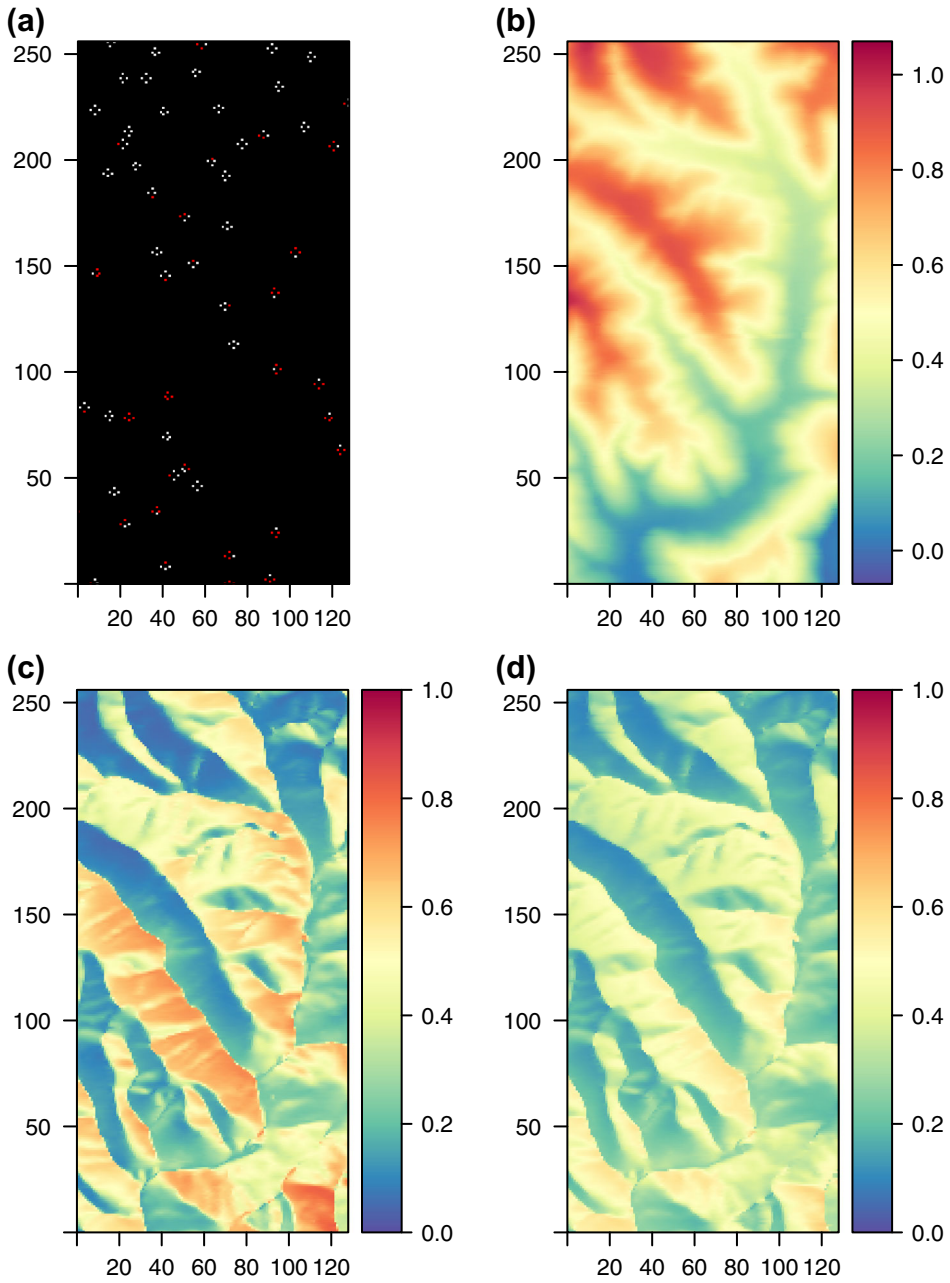

Figure 1. Prediction domain from the Missouri Ozark Forest Ecosystem Project presented in Hooten et al. (2003). *Red* and *white pixels* in **a** represent the plot locations that were sampled and whether the plant, pointed-leaved tick trefoil, was present (*red*) or absent (*white*). The map in **b** shows the relative elevation covariate. Maps in the *lower panels* show the predicted probability of presence from a spatial probit model (**c**; SGLMM in Fig. 2) and a spatial group lasso probit model (**d**; SGL in Fig. 2) (Color figure online).

were narrower when compared to the same quantities from the SGLMM and SGLMM with RSR (Fig. 2a). In contrast, the posterior mean and 95% CI for the elevation covariate (β_2) were substantially different under the SGLMM when compared to the SGLMM with RSR. If

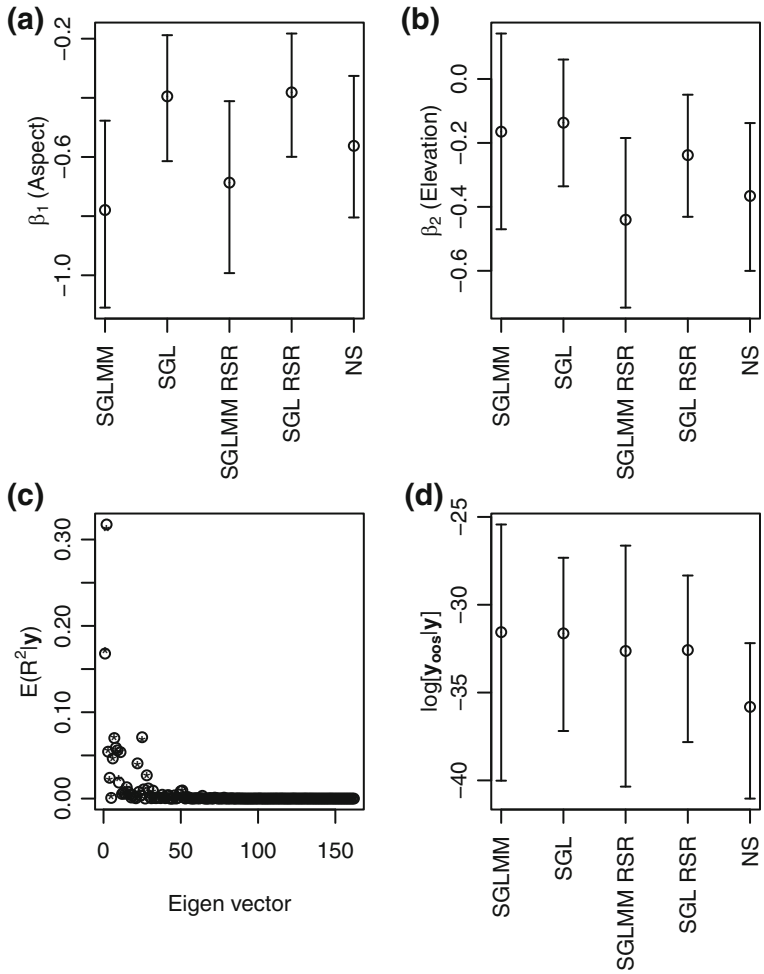

Figure 2. Posterior expectation (*open circles*) and 95% equal-tail credible intervals for two covariates (**a**, **b**) and the out-of-sample log score (**d**) estimated from a spatial probit model (SGLMM), a spatial group lasso probit model (SGL), spatial models with random effects restricted to be orthogonal to the covariates (SGLMM RSR and SGL RSR, respectively; note the estimated regression coefficient under RSR is δ_j from (15)), and a model that did not include a spatial random effect (NS). **c** The coefficient of determination between the elevation covariate and the 152 eigenvectors obtained from the spatial covariance matrix using the SGLMM (*open circles*) and SGL (*asterisk*).

researchers were to judge significance of the β_2 (elevation) covariate based on a 95% CI that did not contain zero, then choosing to use RSR would have altered the inference (Fig. 2b). For both the SGLMM and SGL, the 95% CI for β_2 (elevation) contained zero. In contrast, the 95% CI for β_2 (elevation) did not contain zero for both models when RSR was used (Fig. 2b). The discrepancy between summaries of the posterior distributions of β_2 (elevation) is likely a result of correlation (16) between the first and second leading eigenvectors and the elevation covariate; this demonstrates that confounding with the elevation covariate is occurring at a broad spatial scale (Fig. 2c). The posterior mean of the log score based on out-of-sample data was similar for the SGL and for the SGL with RSR when compared to

the SGLMM, SGLMM with RSR, and NS models (Fig. 2d). The 95% CI from the posterior distributions for the log score from the SGL and SGL with RSR was narrower than the 95% CIs obtained from the SGLMM and SGLMM with RSR (Fig. 2c). The ESS for β_1 (aspect) and β_2 (elevation) were: 784 and 942 for the SGLMM; 916 and 1701 for the SGLMM with RSR; 2708 and 2925 for the SGL; 3001 and 3886 for the SGL with RSR; and 2493 and 2684 for the NS model. The run time to obtain 12,000 MCMC samples for the ESS calculations was 389, 378, 393, 393, and 4 s for the SGLMM, SGL, SGLMM with RSR, SGL with RSR, and NS models, respectively.

4.2. EXAMPLE 2: DISEASE RISK FACTOR ANALYSIS

Chronic wasting disease (CWD) is a fatal transmissible spongiform encephalopathy that occurs in ungulates. CWD was first discovered in Colorado, USA in 1967. Since 1967, CWD has been detected in numerous species and in other areas of North America (Williams et al. 2002). In the state of Wisconsin, USA, CWD was first detected in white-tailed deer (*Odocoileus virginianus*) in 2001, as a result of the Wisconsin Department of Natural Resources (DNR) surveillance efforts. One of the goals of continued surveillance of CWD by the DNR is to identify risk factors associated with the disease. Previous research has identified sex and age as risk factors; however, wildlife biologists are also interested in understanding how features of the landscape influence prevalence (e.g., Walter et al. 2011; Evans et al. 2016).

Data collected by the Wisconsin DNR include information for deer that were sampled using various methods (e.g., hunter-harvested deer, fatal vehicle collisions) and of various locational accuracy. CWD was first detected in the southwestern portion of Wisconsin, and the disease has increased in prevalence in this area, but has much lower prevalence in other areas of the state. Because of the high and increasing prevalence in southwestern Wisconsin, we limit our analysis to hunter-harvested white-tailed deer collected in southwestern Wisconsin in 2002 during the initial outbreak. Furthermore, because of location accuracy, we focused our analysis on deer with location information that was collected at the section-level (i.e., the location of the harvested deer was recorded as the center of a public land survey system section of land). This resulted in a sample size of 10,263 tested deer (162 positive deer; Fig. 3a). We used the 2011 National Land Cover Dataset to calculate nine landscape covariates: the proportion of water, human development (open space, low intensity, medium intensity, high intensity), deciduous forest, grassland, pasture, and cultivated crops (land classes 11, 21, 22, 23, 24, 41, 71, 81, 82, respectively) within a 2.59 km^2 grid cell (Fig. 3b and Web Figure 6, Homer et al. 2015). A 2.59 km^2 grid cell represents the location accuracy of a section of land (a 1 mi^2 square area), is roughly the size of the home range of a white-tailed deer, and is likely representative of the habitat that the tested deer would have spent a majority of its time within. We randomly selected 50% ($n = 5132$) of the data to fit all models and used the remaining 50% of the data to evaluate the predictive performance.

Given the relatively large size of our training sample ($n = 5132$), the full-rank spatial model would be computationally demanding. To facilitate computation, we used a process convolution with an exponential kernel and $m = 48$ equally spaced knots (Web Figure 7). Briefly, the process convolution constructs a continuous spatial model based on a Gaussian

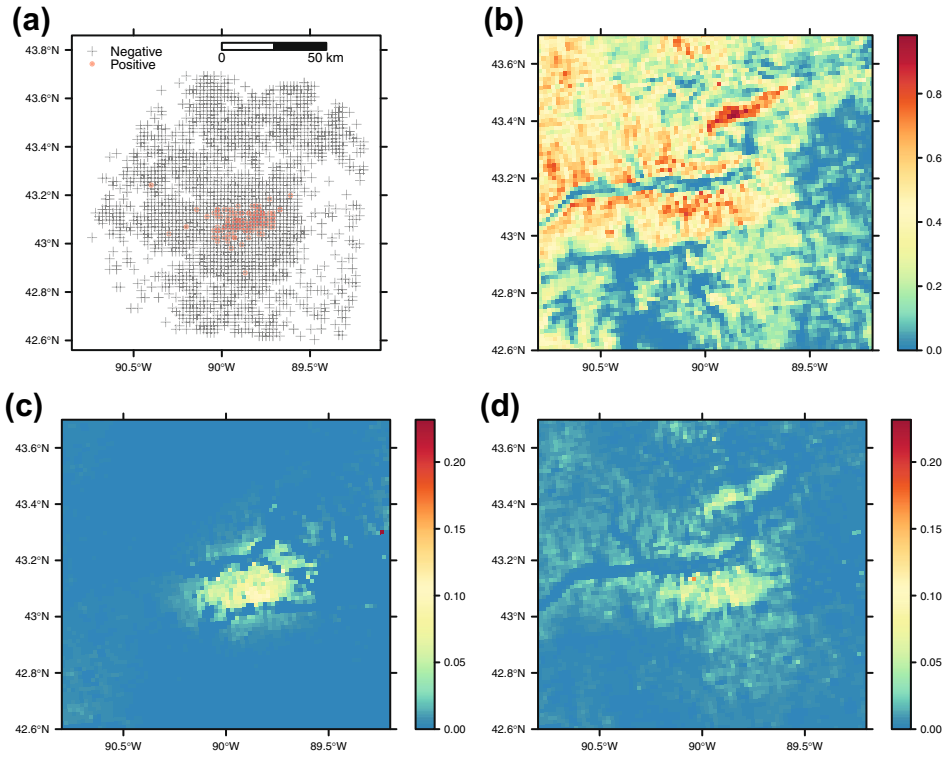

Figure 3. Prediction domain for 10,263 white-tailed deer tested for chronic wasting disease in southwest Wisconsin, USA (**a**; 162 positive deer). The map in **b** shows the percent forest covariate. Maps in the *lower panels* show the predicted probability of infection from a spatial probit model (**c**; SGLMM in Fig. 4) and a spatial group lasso probit model (**d**; SGL in Fig. 4). Note the predicted prevalence is for male white-tail deer in age-class 4.

process by convolving independent and identically distributed Gaussian random variables on a lattice with an arbitrary kernel (Higdon 2002; Hefley et al. 2016). Using process convolution, the spatial binary model is

$$\begin{aligned} \mathbf{y}|\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \phi &\sim \text{Bernoulli}\left(\Phi(\mathbf{W}\boldsymbol{\gamma} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}(\phi)\boldsymbol{\alpha})\right) \\ \boldsymbol{\alpha} &\sim N\left(\mathbf{0}, \sigma_\alpha^2 \mathbf{I}\right), \end{aligned} \tag{18}$$

where $\mathbf{y}$ is an $n \times 1$ vector with elements 1 if a tested deer is positive for CWD and 0 otherwise, Φ is the standard normal cumulative distribution function, $\mathbf{W}$ and $\mathbf{X}$ are matrices of covariates, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are the associated regression coefficients, $\mathbf{Z}(\phi)$ is an $n \times m$ matrix of exponential kernel basis vectors with range parameter ϕ (i.e., $z_{ij}(\phi) = e^{-\frac{d_{ij}}{\phi}}$ where $z_{ij}(\phi)$ is the element in the ith row and jth column of $\mathbf{Z}(\phi)$ in (18), d_{ij} is the distance in km between locations i and j), and $\boldsymbol{\alpha}$ is an $m \times 1$ vector of basis coefficients. We centered and scaled the landscape covariates and also included the sex and age of the tested deer. The covariates sex and age were not spatially indexed or structured, whereas the landscape covariates exhibited spatial structure (Fig. 3b; Web Figure 6). As a result, there is potential for multicollinearity with the landscape covariates and the basis vectors

in $\mathbf{Z}(\phi)$. For this example, we placed all nine landscape covariates in $\mathbf{X}$ and applied the group lasso prior, whereas the sex and age covariates are included in $\mathbf{W}$. For the landscape covariates, we monitored the correlation among the landscape covariates and the left singular vectors obtained from a singular value decomposition of $\mathbf{Z}(\phi)$. For the SGLMM and SGL, we used the following priors: $\boldsymbol{\gamma} \sim N(\mathbf{0}, 10\mathbf{I})$ and $\phi \sim$ uniform$(0, 200)$ (note, 200 km is approximately the diagonal distance across the study area). For the SGLMM, we used the priors: $\boldsymbol{\beta} \sim N(0, 10\mathbf{I})$ and $\frac{1}{\sigma_\eta^2} \sim$ gamma$(2.1, 1.1)$ (note, $E(\sigma_\eta^2) = 1$ and $\text{Var}(\sigma_\eta^2) = 10$). As in the first example, our motivation for the prior on σ_η^2 was to let data determine the amount of regularization on $\boldsymbol{\eta}$ by using a diffuse prior. For the SGL, we used the prior $\lambda^2 \sim$ gamma $(1, 1)$ (note, $E(\lambda^2) = 1$ and $\text{Var}(\lambda^2) = 1$). For the NS model, we used the priors $\boldsymbol{\gamma} \sim N(\mathbf{0}, 10\mathbf{I})$ and $\boldsymbol{\beta} \sim N(\mathbf{0}, 10\mathbf{I})$.

The predicted prevalence was highest in the center of the study area and ranged from 0.00 to 0.23 when using the SGLMM and from 0.00 to 0.17 when using the SGL (Fig. 3c, d). As in the first example, the smaller range of prevalence is a result of the regularization of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ (cf. Fig. 4a, b, Web Figures 8 and 9). The mean and 95% CIs from the posterior distributions of the coefficients corresponding to the sex and age covariates were similar for all models (Web Appendix D). For illustrative purposes, we present results for one of the nine landscape covariates (forest). The 95% CI for the coefficient corresponding to the forest covariate (β_1) contained zero for all models, but the 95% CIs were much wider for the SGLMM, SGLMM with RSR, and NS models, which may be inflated due to correlation (16) among the basis vectors and the forest covariate (Fig. 4). The posterior mean of the log score based on out-of-sample data was highest for the SGL and SGL with RSR (Fig. 4c). The ESS for β_1 (forest) was 35 for the SGLMM, 41 for the SGLMM with RSR, 463 for the SGL, 495 for the SGL with RSR, and 75 for the NS model. The required time to obtain 12,000 MCMC samples for the ESS calculations was 562, 617, 5703, 5759, and 56 s for the SGLMM, SGL, SGLMM with RSR, SGL with RSR, and NS models, respectively (Web Appendix D).

5. SIMULATED DATA EXAMPLES

Based on the two examples above, we made three general observations: (1) confounding, and the scale at which it occurs, can be detected by monitoring the correlation between eigen basis vectors and covariates as in (16); (2) in the presence of confounding, the SGL shrinks regression coefficients, and in some cases, the posterior distribution of the log score; and (3) the SGL increases the ESS of regression coefficients relative to the SGLMM and SGLMM with RSR. To support these three findings, we simulated data that are representative of our examples. We simulated data from the model:

$$\begin{aligned} \mathbf{y}|\gamma_0, \boldsymbol{\beta}, \boldsymbol{\eta} &\sim \text{Bernoulli}\left(\Phi(\gamma_0 + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta})\right) \\ \boldsymbol{\eta} &\sim N\left(\mathbf{0}, \sigma_\eta^2 \mathbf{C}(\phi)\right), \end{aligned} \tag{19}$$

where $\mathbf{y}$ is an $n \times 1$ vector of binary observations, Φ is the standard normal cumulative distribution function, $\mathbf{X}$ is an $n \times 3$ vector of covariates that is generated by adding Gaussian

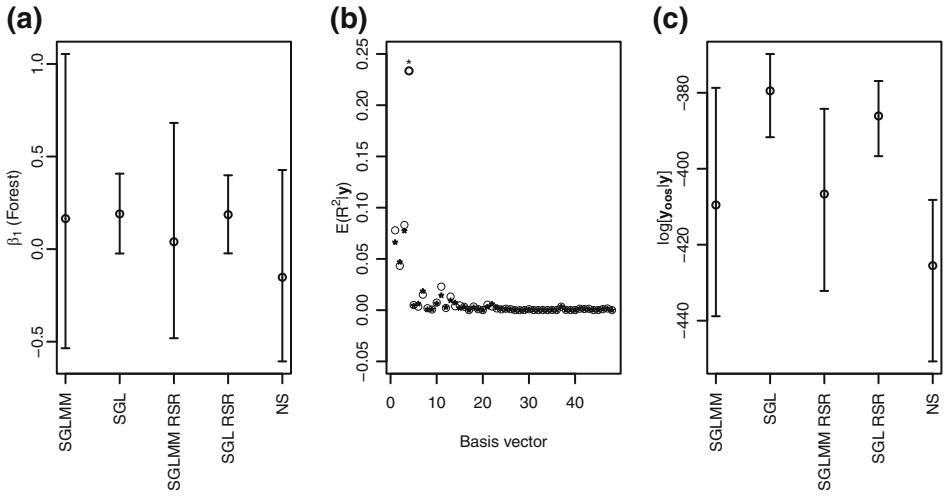

Figure 4. Posterior expectation (*open circles*) and 95% equal-tail credible intervals for the forest covariate (**a**) estimated from a spatial probit model (SGLMM), spatial group lasso probit model (SGL), spatial models with random effects restricted to be orthogonal to the covariates (SGLMM RSR and SGL RSR, respectively; note the estimated regression coefficient under RSR is δ_j from (15)), and a model that did not include a spatial random effect (NS). **b** The coefficient of determination between the forest covariate and 48 basis vectors using the SGLMM (*open circles*) and SGL (*asterisk*). **c** The out-of-sample log score for each method.

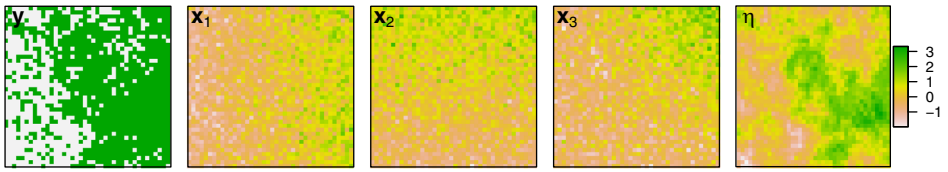

Figure 5. Simulated binary response ($\mathbf{y}$) that depends on three covariates ($\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$), and a spatial random effect $\boldsymbol{\eta}$ at $n = 1600$ location on the unit square (note that $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = 0$).

distributed errors to the "east–west" gradient ($\mathbf{x}_1$), "north–south" gradient ($\mathbf{x}_2$), and the "northeast–southwest" gradient ($\mathbf{x}_3$). The vector $\boldsymbol{\eta}$ is the spatial random effect generated from a Gaussian distribution with an exponential correlation function (i.e., $c_{ij}(\phi) = e^{-\frac{d_{ij}}{\phi}}$ where $c_{ij}(\phi)$ is the element in the ith row and jth column of $\mathbf{C}(\phi)$ in (19) and d_{ij} is the Euclidean distance in between locations i and j). The covariates were designed to result in confounding with the spatial random effect at a broad spatial scale; thus, we expect that the leading eigen vector(s) will be correlated with $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$ (Fig. 5). We set $\gamma_0 = 0, \beta_1 = 1$, $\beta_2 = 0$, $\beta_3 = 0$, $\sigma^2_{\mathrm{x}} = 0.25$, $\sigma^2_\eta = 1$, and $\phi = 0.25$ and generated data at 1600 equally spaced locations on the unit square (Fig. 5). We randomly selected 500 observations that were used to fit the models and the remaining 1100 observations to evaluate the predictive ability of the models.

For the SGLMM and SGL, we used the following priors: $\gamma_0 \sim N(0, 10)$ and $\phi \sim \text{uniform}(0, 1)$. For the SGLMM, we used the priors: $\boldsymbol{\beta} \sim N(0, 10\mathbf{I})$ and $\frac{1}{\sigma^2_\eta} \sim \text{gamma}(2.1, 1.1)$ (as in previous examples, $E(\sigma^2_\eta) = 1$ and $\text{Var}(\sigma^2_\eta) = 10$). For the SGL, we used the prior $\lambda^2 \sim \text{gamma}\ (0.1, 0.1)$ (note, $E(\lambda^2) = 1$ and $\text{Var}(\lambda^2) = 10$). For the NS

model, we used the priors $\gamma_0 \sim N(0, 10)$ and $\boldsymbol{\beta} \sim N(0, 10\mathbf{I})$. For our simulation experiment, we used a single data set to present detailed results and an additional 10 simulated data sets to assess sampling variability. Using a single simulated data set, we fit the SGLMM, the SGLMM with RSR, the SGL, the SGL with RSR, and the NS model. We report the posterior mean and 95% CIs for β_1, β_2, β_3 and the out-of-sample log score, the coefficient of determination (16) between the spatial covariate and 20 eigenvectors with the largest eigenvalues, and the ESS for β_1. Using the 10 additional simulated data sets, we fit the SGLMM and the SGL models. For the 10 simulated data sets, we report the percentage of simulations for which the mean log score was greatest for each model, the average ESS for β_1, β_2, and β_3, and the average maximum coefficient of determination between the spatial covariate and all eigenvectors. For all models, we obtained 12,000 MCMC samples and discarded the first 2000. More details related to the simulated data and the code required to reproduce our results can be found in Web Appendix E.

Our simulated data resulted in a spatial random effect that was smooth and had leading eigenvectors that were correlated with all three covariates (Fig. 6a). For the single data example, the posterior mean for β_1 was larger than the true value ($\beta_1 = 1$) and the 95% CI contained the true value for the SGLMM and the NS model (Fig. 6a). The posterior mean of β_1 obtained from the SGL and SGL with RSR was always closer to zero, and the 95% CIs were narrower when compared to the SGLMM and SGLMM with RSR (Fig. 6a). The posterior means of β_2 and β_3 were close to zero for all methods, and the 95% CI contained the true values ($\beta_2 = 0$ and $\beta_3 = 0$; Fig. 6c, e). The posterior mean of the out-of-sample log score was higher and the 95% CI were narrower for the SGL and SGL with RSR, when compared to the other methods. The ESS for β_1 was highest for the SGL and SGL with RSR (Fig. 6d). For the 10 additional simulated data sets, the SGL had the highest mean log score in 20% of the simulated data sets. The average ESS for β_1, β_2, and β_3 was 450, 855, and 977 for the SGLMM and 2327, 3447, and 4003 for the SGL model. The average maximum coefficient of determination between the spatial covariate and all eigenvectors was 0.39 for both the SGLMM and SGL. In general, results from the simulation support our findings from the two data examples.

6. DISCUSSION

Several studies have developed variants of the lasso for spatial and spatio-temporal data, but these methods were implemented under a penalized likelihood paradigm for the linear mixed model and focused on covariance estimation (Zhu and Liu 2009; Hsu et al. 2012) or model selection (Huang et al. 2010; Zhu et al. 2010). We demonstrated that the group lasso can be easily implemented for binary data within a Bayesian framework, but the group lasso prior can be used with any SGLMM. Our results show that the group lasso prior can be useful for regularizing spatial models when multicollinearity among spatially indexed and structured covariates and random effects occurs. In many applications, the SGLMM can be computationally expensive to implement when the sample size is large (Wikle 2010). When implementing the SGLMM using MCMC methods, efficient sampling that results in a large ESS can linearly decrease the computational burden of spatial models. Several

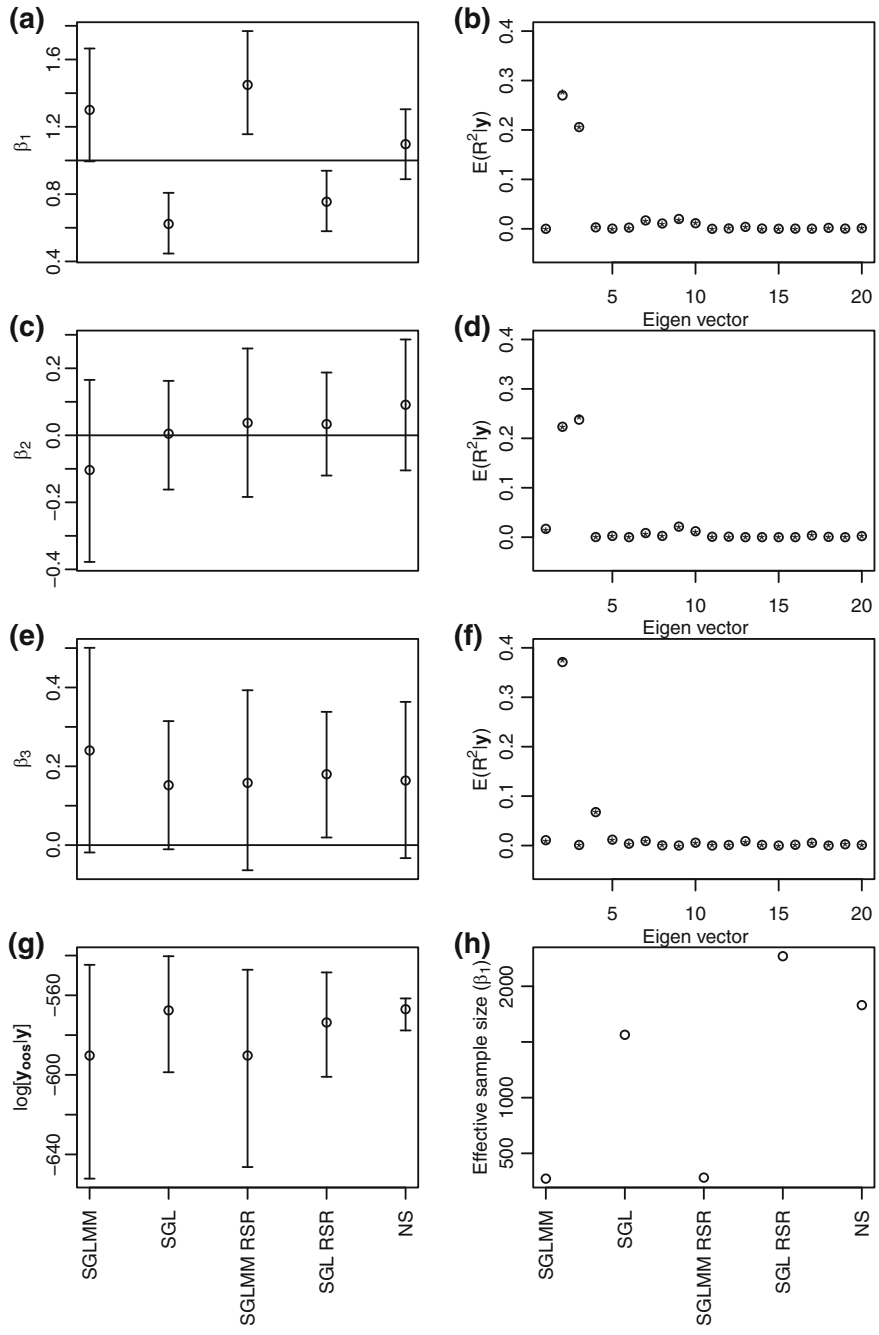

Figure 6. Results from five models fit to simulated data. Shown are the posterior expectations (*open circles*) and 95% equal-tail credible intervals for the three regression coefficient (**a**, **c**, **e**; the true value is represented by the *horizontal line*) and the out-of-sample log score (**g**). Also shown is the posterior mean of the coefficient of determination between the covariates ($\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$) and the first 20 eigen basis vectors (**b**, **d**, **e**) for the spatial probit model (SGLMM; *open circles*) and the spatial group lasso probit model (SGL; *asterisk*) and the effective sample size for β_1 from 10,000 MCMC samples (**h**) estimated from the SGLMM, SGL, spatial models with random effects restricted to be orthogonal to the covariates (SGLMM RSR and SGL RSR, respectively), and a model that did not include a spatial random effect (NS). Note the estimated regression coefficient under RSR is δ_j from (15).

studies have demonstrated that RSR results in more efficient MCMC sampling and our results corroborate this finding. In addition, our results show that regularization via the SGL results in computational gains in terms of ESS for both the standard and RSR models. In some situations, the SGL results in an ESS for regression coefficients that is as large as the model that did not have a spatial random effect.

In general, the SGL shrinks the regression coefficients, the spatial random effects, and predicted quantities (e.g., compare Fig. 1c, d). As a consequence, the SGL may result in a less diffuse posterior distribution of the log score with higher posterior expectation (i.e., better predictive accuracy), when compared to the SGLMM (e.g., Fig. 4c). Although not demonstrated in our examples, the parameter λ can be tuned to yield the optimal predictive model, which could result in the SGL having a higher predictive ability when compared to the SGLMM (e.g., using n-fold cross-validation; Park and Casella 2008; Kyung et al. 2010). In some situations, based solely on the predictive score, the SGLMM and SGL perform similar to the model that did not have a spatial random effect (e.g., example 1 in Fig. 2d). When collinearity among covariates and the spatial random effect occurs, the observed increase in predictive ability by including a spatial random effect (or the converse, spatial covariates) is attenuated; in extreme cases, the spatial covariate (or random effect) may explain most (or all) of the variability in the response. As with extreme cases of multicollinearity in traditional regression models, it is unlikely that regression coefficients are identifiable without prior knowledge of which covariates or basis vectors to exclude from the SGLMM.

Our results show that confounding can be detected by monitoring the correlation between covariates and the basis vectors associated with the spatial effect as in (16) and can help identify when the assumptions of the SGLMM are satisfied (i.e., $|r(\mathbf{x}_j, \mathbf{z}_k)| \approx 0, \ \forall \ j, k$ in (16)). Obtaining reliable inference from the SGLMM is challenging when the covariates are correlated with the spatial random effect (Clayton et al. 1993; Reich et al. 2006; Hodges and Reich 2010; Paciorek 2010; Hughes and Haran 2013; Hanks et al. 2015; Murakami and Griffith 2015). In applied contexts, there is little guidance on when inference from the standard SGLMM and SGLMM with RSR is desirable (Hodges and Reich 2010; Hooten et al. 2013; Hanks et al. 2015; Schmidt et al. 2015). If researchers were to judge the significance of regression coefficients based on 95% CIs that did not contain zero, then choosing to use RSR could alter inference (Fig. 2b). We showed that monitoring the correlation between covariates and basis vectors associated with the spatial random effect can help identify when multicollinearity may be a problem, but making reliable inference remains an open question. Results from our simulated data set demonstrate the issue related to inference (Fig. 6b). In a separate simulation study, we did not find generalizable advice for when one prior might be preferred over others with respect to inference about regression coefficients. In future work, an examination of the tuning parameter for the SGL may have desirable inferential properties for regression coefficients, such as nominal coverage probability (e.g., Gunes and Bondell 2012).

ACKNOWLEDGEMENTS

We would like to acknowledge Dennis Heisey for his early contributions to development of this research endeavor. We thank Jun Zhu and two anonymous reviewers for valuable insight and discussions about this

work. We thank the staff of the Wisconsin Department of Natural Resources for their collaboration in obtaining deer tissue samples and the Wisconsin hunters who provided them. In particular, we thank Erin Larson for maintaining the CWD sample data base. Funding for this project was provided by the USGS National Wildlife Health Center via Grant G14AC00366. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

[Received April 2016. Accepted December 2016. Published Online January 2017.]

REFERENCES

Banerjee, S., Gelfand, A., Finley, A., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society, Series B* 70, 825–848.

Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2015). Dirichlet-laplace priors for optimal shrinkage. *Journal of the American Statistical Association* 110, 1479–1490.

Clayton, D. G., Bernardinelli, L., and Montomoli, C. (1993). Spatial correlation in ecological analysis. *International Journal of Epidemiology* 22, 1193–1202.

Cressie, N., and Wikle, C. (2011). *Statistics for Spatio-Temporal Data*. Hoboken, New Jersey: John Wiley & Sons.

Diggle, P. J., Tawn, J., and Moyeed, R. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society, Series C* 47, 299–350.

Evans, T. S., Kirchgessner, M. S., Eyler, B., Ryan, C. W., and Walter, W. D. (2016). Habitat influences distribution of chronic wasting disease in white-tailed deer. *The Journal of Wildlife Management* 80, 284–291.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 24, 997–1016.

Givens, G. H., and Hoeting, J. A. (2012). *Computational Statistics*. Hoboken, New Jersey: John Wiley & Sons.

Gotway, C. A., and Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics* 2, 157–178.

Gunes, F., and Bondell, H. D. (2012). A confidence region approach to tuning for variable selection. *Journal of Computational and Graphical Statistics* 21, 295–314.

Hanks, E. M., Schliep, E. M., Hooten, M. B., and Hoeting, J. A. (2015). Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics* 26, 243–254.

Hefley, T. J., Broms K. M., Brost B. M., Buderman, F. E., Kay, S. L., Scharf J. R., Williams, P. J. and Hooten, M. B. (2016). The basis function approach for modeling autocorrelation in ecological data. *Ecology*. doi:10.1002/ecy.1674

Hefley, T. J., and Hooten, M. B. (2016). Hierarchical species distribution models. *Current Landscape Ecology Reports* 1, 87–97.

Higdon, D. (2002). Space and space-time modeling using process convolutions. *Quantitative Methods for Current Environmental Issues* 3754.

Hodges, J. S., and Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician* 64, 325–334.

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67.

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G.,*et al.* (2015). Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing* 81, 345–354.

Hooten, M. B.,, Hanks, E. M., Johnson, D. S., and Alldredge, M. W. (2013). Reconciling resource utilization and resource selection functions. *Journal of Animal Ecology* 82, 1146–1154.

Hooten, M. B., and Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs* 83, 3–28.

Hooten, M. B., Larsen, D. R., and Wikle, C. K. (2003). Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology* 18, 487–502.

Hsu, N., Chang, Y., and Huang, H. (2012). A group lasso approach for non-stationary spatial–temporal covariance estimation. *Environmetrics* 23, 12–23.

Huang, H., Hsu, N., Theobald, D.M., and Breidt, F.J. (2010). Spatial lasso with applications to GIS model selection. *Journal of Computational and Graphical Statistics* 19, 963–983.

Hughes, J. and Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society, Series B* 75, 139–159.

Hui, F., Müller, S., and Welsh, A. (2016). Joint selection in mixed models using regularized PQL. *Journal of the American Statistical Association* doi:10.1080/01621459.2016.1215989

Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis* 5, 369–411.

Mallick, H. and Yi, N. (2013). Bayesian methods for high dimensional linear models. *Journal of Biometrics & Biostatistics* S1, 005.

Murakami, D., and Griffith, D. A. (2015). Random effects specifications in eigenvector spatial filtering: a simulation study. *Journal of Geographical Systems* 17, 311–331.

Paciorek, C. (2010). The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science* 25, 107–125.

Park, T., and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association* 103, 681–686.

R Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* 62, 1197–1206.

Schabenberger, O., and Gotway, C. A. (2004). *Statistical Methods for Spatial Data Analysis*. Boca Raton, Florida: Chapman & Hall/CRC Press.

Schmidt, A. M., Rodríguez, M. A., and Capistrano, E. S. (2015). Population counts along elliptical habitat contours: hierarchical modeling using poisson-lognormal mixtures with nonstationary spatial structure. *Annals of Applied Statistics* 9, 1372–1393.

Stroup, W. W. (2012). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Boca Raton, Florida: CRC Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58, 267–288.

Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data.* Hoboken, New Jersey: John Wiley & Sons.

Walter, D. W., Walsh, D. P., Farnsworth, M. L., Winkelman, D. L., and Miller, M. W. (2011). Soil clay content underlies prion infection odds. *Nature Communicaitons* 2, 200.

Wikle, C. K. (2010). Low Rank Representations for Spatial Processes in *Handbook of Spatial Statistics*, pgs. 107–118. Boca Raton, Florida: CRC Press.

Williams, E. S., Miller, M. W., Kreeger, T. J., Kahn, R. H., and Thorne, E. T. (2002). Chronic wasting disease of deer and elk: a review with recommendations for management. *The Journal of Wildlife Management* 3, 551–563.

Zhu, J., Huang, H., and Reyes, P. (2010). On selection of spatial linear models for lattice data. *Journal of the Royal Statistical Society, Series B* 72, 389–402.

Zhu, Z. and Liu, Y. (2009). Estimating spatial covariance using penalised likelihood with weighted L_1 penalty. *Journal of Nonparametric Statistics* 21, 925–942.