Methods in Ecology and Evolution

Methods in Ecology and Evolution 2011, **2**, 143–154 doi: 10.1111/j.2041-210X.2010.00060.x

Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence

Andrew O. Finley*

Departments of Forestry and Geography, Michigan State University, East Lansing, MI 48824-1222, USA

Summary

1. When exploring spatially complex ecological phenomena using regression models it is often unreasonable to assume a single set of regression coefficients can capture space-varying and scale-dependent relationships between covariates and the outcome variable. This is especially true when conducting analysis across large spatial domains, where there is an increased propensity for anisotropic dependence structures and non-stationarity in the underlying spatial processes.
2. Geographically weighted regression (GWR) and Bayesian spatially-varying coefficients (SVC) are the most common methods for modelling such data. This paper compares these methods for modelling data generated from non-stationary processes. The comparison highlights some strengths and limitations of each method and aims to assist those who seek appropriate methods to better understand spatially complex ecological systems. Both synthetic and ecological data sets are used to facilitate the comparison.
3. Results underscored the need for the postulated model to approximate the underlying mechanism generating the data. Further, results show GWR and SVC can produce very different regression coefficient surfaces and hence dramatically different conclusions can be drawn regarding the impact of covariates. The trade-off between the richer inferential framework of SVC models and computational demands is also discussed.

Key-words: Bayesian, Gaussian process, geographically weighted regression, geostatistics, Markov chain Monte Carlo, spatially-varying coefficients

Introduction

Ecologists often employ regression models to explore relationships between covariates and ecological phenomena of interest. A non-spatial regression model is adequate in the absence of extraneous structured variation, beyond what is explained by the covariates. However, when observations are spatially indexed, we might expect similar outcomes in proximate locations, possibly resulting from similar environmental conditions or disturbance regimes. Ignoring this spatial dependence may result in falsely precise estimates of model parameters and erroneous predictions. Hoeting (2009) offers a nice discussion on consequences of not meeting the assumption of uncorrelated model residuals. A common solution to spatial dependence among the residuals is to add a spatially-varying model intercept that accounts for spatial association through a decreasing function of distance and perhaps direction between observed locations, see e.g. the seminal paper by Diggle *et al.* (1998). Beyond helping to ensure the statistical validity of the model, the addition of a random spatial effect to the intercept allows for partitioning of residual uncertainty into a spatial and non-spatial component. Accounting for sources of uncertainty when modelling ecological phenomena is the central focus of the recent paper by Cressie *et al.* (2009). Recognizing and correctly partitioning residual uncertainty can improve inference, reveal missing covariates and increase prediction accuracy and precision.

In addition to addressing the basic assumptions of the linear model through, for example a spatially-varying intercept, other sources of uncertainty can be partitioned into spatial and non-spatial components. For example, given the complexity of ecological phenomena it is often unreasonable to assume a single set of regression coefficients can adequately capture the potential space-varying and scale-dependent relationships between covariates and the outcome variable across the domain of interest. Geographically weighted regression (GWR) detailed by Fotheringham *et al.* (2002) and the hierarchical modelling

*Correspondence author. E-mail: finleya@msu.edu
Correspondence site: http://www.respond2articles.com/MEE/

approach using spatially-varying coefficients (SVC) described by Gelfand *et al.* (2003) are the most common methods for analysing such data.

The objective of this paper is to compare these two methods for modelling data generated from non-stationary processes. The comparison highlights some strengths and limitations of each method and aims to assist those who seek appropriate methods to better understand spatially complex ecological systems. Similar comparisons of GWR and SVC have been made using public health data. Waller *et al.* (2007) compared GWR to a conditionally autoregressive (CAR) spatially-varying coefficients model. CAR spatial effects are specified for data observed over areal measurement units which are less frequently encountered in ecological data sets – the exception being, perhaps, outcomes summarized over ecoregions, forest stands or watersheds. Rather, the models considered here are for variables recorded at georeferenced locations, referred to as point-referenced data. Using point-referenced public health data, Wheeler & Calder (2007) and Wheeler & Waller (2009) compared GWR and a separable SVC model. These comparisons focused on the models' robustness to collinearity among covariates. Their results showed that in the presence of moderate and strong collinearity, SVC produced more accurate estimates of the regression coefficients than GWR. The comparison presented in this paper expands on these previous investigations by exploring model performance given more complex spatial processes.

The organization of the paper is as follows. 'Spatially-varying coefficients' offers the details of the Bayesian SVC model and a criterion for model selection. The fundamentals of GWR are reviewed in 'Geographically weighted regression'. 'Analyses' details the five data sets used to compare the candidate models. Three synthetic data sets are used to validate and explore model properties. Two ecological data sets that were previously analysed using GWR are reanalysed using the candidate models. The results and discussion of these analyses are provided in 'Results and discussion'. Summary remarks are offered in 'Conclusion'.

Models

SPATIALLY-VARYING COEFFICIENTS

Given the common Gaussian regression model, the SVC model is constructed by adding a continuous multivariate spatial process to account for *spatially-varying regression coefficients* and accommodate spatial dependence in the outcome, as is customary in geostatistics (Cressie 1993). When observations are spatially indexed by $\mathbf{s}$ (e.g. longitude and latitude) the outcome $y(\mathbf{s})$ is modelled using spatially-varying regression coefficients:

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + \tilde{\mathbf{x}}(\mathbf{s})'\mathbf{w}(\mathbf{s}) + \epsilon(\mathbf{s}), \qquad \text{eqn 1}$$

where $\mathbf{x(s)}$ is the set of p covariates, $\boldsymbol{\beta}$ is the associated column vector of regression coefficients, $\tilde{\mathbf{x}}(\mathbf{s})$ includes those covariates from $\mathbf{x(s)}$ whose regression coefficients are posited to vary spatially, and $\epsilon(\mathbf{s}) \overset{iid}{\sim} N(0,\tau^2)$ with residual variance parameter τ^2 which captures measurement error or micro-scale variation, often referred to as the nugget in the geostatistical literature. The vector $\mathbf{w(s)} \sim \text{MVGP}(\mathbf{0}, C_\mathbf{w}(\mathbf{s},\mathbf{s}^*))$ follows a *multivariate* Gaussian process which is completely characterized by its mean and a cross-covariance (matrix) function, $C_\mathbf{w}(\mathbf{s},\mathbf{s}^*;\boldsymbol{\theta}) = \text{cov}\{\mathbf{w(s)},\mathbf{w}(\mathbf{s}^*)\}$. The spatially-varying coefficients are then defined as $\tilde{\boldsymbol{\beta}}(\mathbf{s}) = \boldsymbol{\beta} + \mathbf{w}(\mathbf{s})$. To reduce clutter the $\mathbf{s}$ on $\tilde{\boldsymbol{\beta}}$ will be implicit.

Cross-covariance functions are not routine to specify because they require that for any number and choice of locations the resulting covariance matrix must be positive definite. Details on development and computational considerations for the MVGP specification of $\mathbf{w(s)}$ can be found in Gelfand *et al.* (2004) and Banerjee *et al.* (2004). Here, in place of the MVGP, two specifications that assume independence between the elements of $\mathbf{w(s)}$ within $\mathbf{s}$ are considered; the richer non-separable form $C_\mathbf{w}(\mathbf{s},\mathbf{s}^*) = \text{Diag}\left[\sigma_k^2\rho(\mathbf{s},\mathbf{s}^*;\boldsymbol{\theta}_k)\right]_{k=1}^{p}$ and the more restrictive separable form $C_\mathbf{w}(\mathbf{s}, \mathbf{s}^*) = \text{Diag}[\sigma^2\rho(\mathbf{s}, \mathbf{s}^*;\boldsymbol{\theta})]$, where ρ is a spatial correlation function and σ^2 is the associated variance parameter. This simplification was made because the independent process approach more closely approximates the GWR set up and is computationally less demanding.

Spatial correlation functions of varying complexity are available for defining $\rho(\mathbf{s}, \mathbf{s}^*;\cdot)$, see e.g. Cressie (1993). The two correlation functions considered here are an isotropic exponential, $\rho(\mathbf{s}, \mathbf{s}^*;\boldsymbol{\theta}) = \exp(-\phi||\mathbf{s}_i - \mathbf{s}_j||)$, where ϕ is the spatial range parameter, and an anisotropic form $\rho(\mathbf{s}, \mathbf{s}^*;\boldsymbol{\theta}) = \exp[(\mathbf{s} - \mathbf{s}^*)'[G(\psi)\Lambda^2G'(\psi)]^{-1}(\mathbf{s} - \mathbf{s}^*)]$, where the rotation matrix $G(\psi)$ controls directional dependence and the positive diagonal matrix Λ defines the range of spatial dependence (Paciorek & Schervish 2006).

When describing a spatial process it is useful to report the distance at which spatial dependence is negligible. This distance is referred to as the effective spatial range, and is defined here as the distance at which the spatial correlation drops to 0·05 (Banerjee *et al.* 2004, p. 26). For the isotropic exponential correlation function this is $-\log(0{\cdot}05)/\phi$ and for the anisotropic form this is $-\log(0{\cdot}05)\lambda$, where λ is a diagonal element in Λ. Note, there are two effective spatial ranges for each anisotropic process, one oriented with ψ, and the other perpendicular to ψ.

For n observations, $\mathbf{w} = (\mathbf{w}(\mathbf{s}_1)',\ldots,\mathbf{w}(\mathbf{s}_n)')'$ follows the Multivariate Normal distribution $\text{MVN}(\mathbf{0},\Sigma_\mathbf{w})$, where $\Sigma_\mathbf{w}$ is a np × np covariance matrix partitioned into $p \times p$ blocks with $C_\mathbf{w}(\mathbf{s}_i,\mathbf{s}_j)$ forming the (i,j)-th block [assuming $\tilde{\mathbf{x}}(\mathbf{s}) = \mathbf{x}(\mathbf{s})$]. For a large number of spatial locations, fitting this model becomes computationally prohibitive. Without further specification, fitting (1) will involve computing the inverse and determinant of $\Sigma_\mathbf{w}$. Such computations invoke linear solvers or Cholesky decompositions of complexity $O(n^3p^3)$, once every Markov chain Monte Carlo (MCMC) iteration, to produce estimates of $\boldsymbol{\theta}$. Modelling large spatial data sets has received much attention in the recent past, see e.g. review in Finley *et al.* (2009). If, as proposed above, we are willing to sacrifice inference on the off diagonal elements of $C_\mathbf{w}$ and focus on the diagonal elements, the variance of the spatial processes, then we can use efficient routines to operate on the now sparse $\Sigma_{\mathbf{w}}$. Moving from the non-separable to the separable specification of $C_{\mathbf{w}}$ affords additional computational gains. Specifically, by assuming the processes associated with $\boldsymbol{\beta}$ share common parameters, i.e. σ^2 and $\boldsymbol{\theta}$, then $\Sigma_{\mathbf{w}}^{-1} = \mathbf{R}(\boldsymbol{\theta})^{-1} \otimes \Sigma^{-1}$ and $|\Sigma_{\mathbf{w}}| = |\mathbf{R}(\boldsymbol{\theta})|^p|\Sigma|^n$, where $\otimes$ is the Kronecker product operator, $\mathbf{R}(\boldsymbol{\theta}) = \left[\rho(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})\right]_{i=1,j=1}^{n,n}$ is the $n \times n$ spatial correlation matrix, and $\Sigma = \sigma^2 I_p$ is the $p \times p$ spatial variance matrix. The non-separable $\Sigma_{\mathbf{w}}$ does not enjoy these useful identities.

To complete the SVC model Bayesian specification, a prior must be assigned to each parameter. For the analyses detailed in 'Analyses', the $\boldsymbol{\beta}$'s received *flat* prior distributions and the non-spatial and spatial variance parameters were assigned inverse-Gamma (IG) with hyperparameters IG(2, ·). This is considered a non-informative prior; with a shape value of 2, the IG distribution has infinite variance and is centred on the scale value, which were data set specific. The spatial range parameters ϕ and λ followed Uniform priors, which were chosen to support a spatial range from 0 to the maximum inter-location distance in the data set. The rotation parameters ψ's, in the anisotropic models, also received Uniform priors that were data set specific.

Letting $\gamma = \{\tau^2, \sigma^2, \boldsymbol{\theta}\}$, the joint posterior distribution for (1) is

$$\pi(\boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\gamma} \mid \mathbf{y}) \propto \pi(\boldsymbol{\gamma}) \times N(\mathbf{w} \mid 0, \Sigma_{\mathbf{w}}) \times \pi(\boldsymbol{\beta}) \times \prod_{i=1}^{n} N(y(\mathbf{s}_i) \mid \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + \tilde{\mathbf{x}}(\mathbf{s}_i)'\mathbf{w}(\mathbf{s}_i), \tau^2), \qquad \text{eqn 2}$$

where $\mathbf{y} = (y(\mathbf{s}_1),\ldots,y(\mathbf{s}_n))'$ and the prior distributions are $\pi(\gamma)$ and $\pi(\boldsymbol{\beta})$. The proposed MCMC samplers and associated algorithms for computing fitted $\bar{y}(\mathbf{s}_i)$ and predicted $\tilde{y}(\mathbf{s}_0)$, values are offered in Supporting Information Appendix S1. In addition, for completeness, details for fitting the full MVGP specification are offered in Appendix S2.

The results presented in 'Results and discussion' are based on three MCMC chains, with unique starting values, that were run for 25 000 iterations. The CODA package in R (http://www.r-project.org) was used to diagnose convergence by monitoring mixing using Gelman-Rubin diagnostics and autocorrelations (see e.g. Gelman *et al.* 2004, section 11·6). For all analyses, acceptable convergence was diagnosed within 10 000 iterations (which were discarded as burn-in). The sampler was coded in C++ and FORTRAN and leveraged Intel's Math Kernel Library threaded BLAS and LAPACK routines for matrix computations. This code will be available through the spSVC function in version 0·2–2 of the spBayes R package and also submitted along with the illustrative data sets to this article's on-line supporting information.

Deviance information criterion (DIC) (Spiegelhalter *et al.* 2002) was used to assess the Bayesian models' fit. Let Ω be the set of parameters being estimated for each model, we compute the expected posterior deviance $D(\Omega) = E_{\Omega|\mathbf{Y}}\{-2\log L$ (Data | Ω)}, where L(Data | Ω) is the first stage Gaussian likelihood from the respective model and the effective number of parameters (as a penalty) as $p_D = \overline{D(\Omega)} - D(\bar{\Omega})$, where $\bar{\Omega}$ is the posterior mean of the model parameters. The DIC is then given by $D(\Omega) + p_D$ and is easily computed from the posterior samples with lower values indicating better models.

GEOGRAPHICALLY WEIGHTED REGRESSION

A thorough description of GWR is offered in Fotheringham *et al.* (2002); however, a basic overview is provided here for completeness. Using GWR the outcome $y(\mathbf{s})$ is now modelled as:

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\ddot{\boldsymbol{\beta}}(\mathbf{s}) + \epsilon(\mathbf{s}), \qquad \text{eqn 3}$$

where $\mathbf{x(s)}$ is the set of p covariates, $\ddot{\boldsymbol{\beta}}(\mathbf{s})$ is the associated column vector of spatially-varying regression coefficients, and $\epsilon(\mathbf{s}) \overset{iid}{\sim} N(0, \tau^2)$. The regression coefficients at location $\mathbf{s}$ are estimated by $\hat{\ddot{\boldsymbol{\beta}}}(\mathbf{s}) = [X'\mathbf{H}(\mathbf{s})X]^{-1}X'\mathbf{H}(\mathbf{s})\mathbf{y}$, where X is an $n \times p$ matrix with $\mathbf{x}(\mathbf{s}_i)'$ as the ith row, $\mathbf{H}(\mathbf{s}) = \text{Diag}[h(\mathbf{s},\mathbf{s}_1), h(\mathbf{s},\mathbf{s}_2),\ldots,h(\mathbf{s},\mathbf{s}_n)]$ is a diagonal weight matrix, and $\mathbf{y}$ were previously defined. This location specific weight matrix places greater weight on observations proximate to $\mathbf{s}$. Refer to Fotheringham *et al.* (2002) for a set of popular kernel functions for defining the diagonal of $\mathbf{H}(\mathbf{s})$. Here the exponential kernel is again employed, $h(\mathbf{s}, \mathbf{s}^*) = \exp(-\phi||\mathbf{s}-\mathbf{s}^*||)$, and ϕ is again viewed as the spatial range parameter.

The value of ϕ is found by minimizing a specified objective function. This is commonly a brute force search using cross-validation. Here, the optimal ϕ was estimated using a leave-one-out cross-validation that minimizes the squared difference between observed and predicted values of the $y(\mathbf{s}_i)$'s. Given this optimal ϕ the fitted value for a given observed location is $\bar{y}(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\hat{\ddot{\boldsymbol{\beta}}}(\mathbf{s})$ and the predicted value at a new location is $\tilde{y}(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)'\hat{\ddot{\boldsymbol{\beta}}}(\mathbf{s}_0)$. The parameters for the GWR models were estimated using version 0.6–2 of the spgwr R package.

Analyses

Five data sets were used to compare non-spatial regression, i.e. $\mathbf{w}(\mathbf{s}) = 0$ in (1), GWR, and SVC models. These data were used to illustrate the candidate models' ability to estimate parameters of interest and predict at new locations. Mean squared error of prediction (MSEP) calculated as $\sum_{i=1}^{m}[y(s_i) - \tilde{y}(\mathbf{s}_i)]^2/m$, where m is the number of locations in the holdout set, was used to compare prediction. Lower values of MSEP suggest better predictive performance. The first three data sets are synthetic, each exhibiting increasing levels of spatial complexity. The fourth comprises forest inventory data previously analysed using GWR by Guo *et al.* (2008). The fifth was originally published by Jetz & Rahbek (2002) then subsequently reanalysed using GWR by Foody (2004). These data sets are detailed in the following subsections.

SYNTHETIC

The synthetic data sets are composed of 500 locations selected randomly from a unit square. The outcome value at each location was generated using the SVC model (1). The data sets use a 500×3 X, where the first column is the intercept and the values in the subsequent columns were generated independently from $N(0,1)$. The regression coefficients were set to $\boldsymbol{\beta} = (5,1,5)'$ and $\tau^2 = 1$ for the nugget. The data sets, referred to as Syn-1, Syn-2 and Syn-3, differ in the spatial processes associated with the $\boldsymbol{\beta}$. Syn-1 was generated using a separable spatial process for $\mathbf{w}$ with $\phi = 6$ and $\sigma^2 = 5$ for the spatial range and variance. A non-separable isotropic process with $\phi_0 = 12$, $\phi_1 = 6$, $\phi_2 = 4$ and $\sigma_0^2 = 1$, $\sigma_1^2 = 5$, $\sigma_2^2 = 10$ was used to generate $\mathbf{w}$ for Syn-2. Whereas Syn-3 was generated using a non-separable anisotropic process for $\mathbf{w}$ using $\pi/4 = 0{\cdot}79$ for ψ_0, ψ_1, ψ_2, corresponding to a 45° rotation and associated $\Lambda_0 = \mathrm{Diag}[0{\cdot}17, 0{\cdot}17]$, $\Lambda_1 = \mathrm{Diag}[0{\cdot}03, 0{\cdot}3]$ and $\Lambda_2 = \mathrm{Diag}[0{\cdot}3, 0{\cdot}03]$. Note, although an anisotropic correlation function was used for $\tilde{\boldsymbol{\beta}}_0$ the fact that $\lambda_{0,0} = \lambda_{0,1}$ makes it isotropic and the corresponding ψ_0 unidentifiable (here subscripts on λ indicate the row and column element of Λ). The values 0·03, 0·17 and 0·3 along the diagonal of Λ correspond to effective spatial ranges of 0·1, 0·5 and 0·9 distance units. The associated spatial variances σ_0^2, σ_1^2, σ_2^2 were 1, 5 and 10 respectively.

The three data sets were replicated 25 times using a different random number to generate each replicate. Then, from each replicate, a subset of 250 observations were selected randomly to serve as a holdout set to assess models' predictive performance. Therefore candidate models' parameter estimates were based on the remaining 250 observations.

TREE HEIGHT AND DIAMETER

The next data set used for illustration was recently analysed by Guo *et al.* (2008) using GWR. In that study they show improved model fit of GWR over OLS and also explored the impact of different values of ϕ on parameter and predictive inference. The analysis was based on 2391 mature trees measured on a 200 × 200 m portion of an uneven-aged softwood stand located near Sault Ste. Marie, Ontario (Ek 1969). The major tree species are balsam fir (*Abies balsamea*) and black spruce (*Picea mariana*). Location, diameter at breast height (DBH), and total height (TH) were recorded for each tree. See Guo *et al.* (2008) for a full description of these data. Here, the candidate models use log (DBH) to explain the variability in the outcome variable log (TH).

Again, to assess models' predictive performance, 598 observations (i.e. 25%) were selected randomly to serve as a holdout set. The remaining 1793 observations were used to fit the candidate models.

BIRD SPECIES RICHNESS

The final data set was originally described by Jetz & Rahbek (2002) and used to explore geographical patterns in sub-Saharan Africa bird species richness. In that study conclusions were based on simple stationary spatial regression models that related environmental covariates to the species richness outcome variable. Recognizing the potential for spatial non-stationarity in the relationship between species richness and the covariates, Foody (2004) reanalysed these data using GWR. The data set comprises 1599 measures of bird species richness at a 1° spatial resolution across sub-Saharan Africa. The environmental covariates included total annual precipitation (PREC), mean annual temperature (TEMP), and normalized difference vegetation index (NDVI) which is an indicator of vegetative land cover. Foody (2004) offers a full description of these variables. For the analysis presented here, covariates were scaled to have a mean of 0 and standard deviation of 1. The outcome variable was scaled 0–100 with 100 being the largest observed bird species richness.

A subset of 400 observations selected randomly were used as a holdout set, again to examine candidate models' predictive ability. Therefore, the remaining 1199 observations were used to estimate the models' parameters.

Results and discussion

SYNTHETIC

For the analysis of the synthetic data, the Bayesian non-spatial and SVC models' variance parameters were assumed to follow an IG(2,1), i.e. a vague prior with infinite variance and mean of 1. For Syn-3, the anisotropic rotation parameters ψ_0, ψ_1 and ψ_2 each received a mildly informative Uniform prior, $U(0,\pi/2)$. The other prior distributions were specified in 'Spatially-varying coefficients'.

Candidate models' parameter estimates for the first replicate of each of the three synthetic data sets are offered in Appendix S3. Across all synthetic data sets, GWR and SVC produced similar estimates of $\hat{\tilde{\boldsymbol{\beta}}}$ and $\tilde{\boldsymbol{\beta}}$ respectively. Further, compared to the τ^2 of the non-spatial model, GWR and SVC effectively apportion residual variance into spatial and non-spatial components. For example, using the first replicate of Syn-1 data set, the non-spatial model estimate for τ^2 is 12·31 vs. 1·53 and 1·20 for GWR and SVC respectively. This trend is repeated for Syn-2 and Syn-3 results (Appendix S3).

The surfaces from the first replicate of the Syn-3 outcome variable $\mathbf{y}$, $\tilde{\boldsymbol{\beta}}_0$, $\tilde{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{\beta}}_2$ are illustrated in Fig. 1(a), (d), (g) and (j) respectively. Here too the corresponding surfaces of fitted values for the GWR and SVC models are plotted in the second and third columns. Again, despite the simple separable isotropic parameterization of GWR, directional dependence is clearly seen in (h) and (k) – although to lesser degree than seen in the SVC estimates (i), (l).

The models differ substantially in predictive ability. The values in Table 1 are the mean and associated standard deviation over each replicates' MSEP (see Appendix S3 for replicate specific MSEP). For Syn-1, there is a relatively small difference in the predictive performance between GWR and SVC, 7·96 vs. 7·00 respectively. It is not surprising that the models perform similarly; like the SVC separable model, GWR assumes a single spatial range parameter ϕ. However, by explicitly accommodating each β's spatial process for the Syn-2 and Syn-3 data sets, the SVC model produces MSEP lower than GWR.

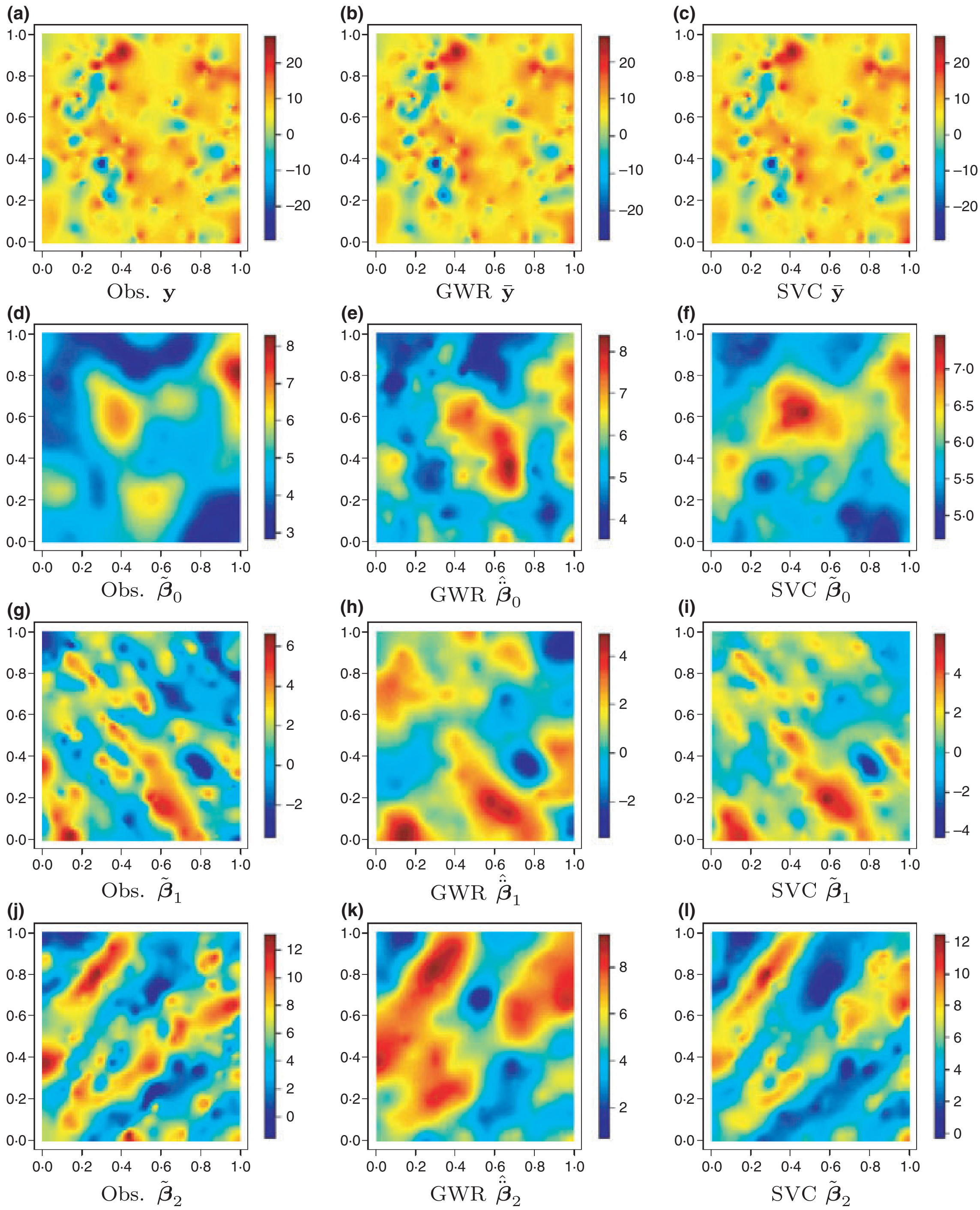

Fig. 1. Interpolated surfaces of model fitted values $\bar{\mathbf{y}}$ and spatially-varying regression coefficients for Syn-3 (anisotropic synthetic data set). SVC non-separable anisotropic model surfaces are based on each location's posterior predictive mean.

TREE HEIGHT AND DIAMETER

The relationship between tree DBH and TH is influenced by several individual and environmental factors. Individual factors include age, species and genetics, whereas, environmental factors include quality of soil, quantity of water and light, and competition for these resources. These factors can cause local spatial dependence in TH and varying relationships between DBH and TH across the domain. For example a cohort of trees of the same species, age and parentage and hence similar DBH and TH relationships might be established in gaps created by some form of disturbance e.g. fire, wind or insect infestation. An example of an environmental factor influencing tree growth characteristics is proximate similarities in soil productivity due to parent material or disturbance history. Given these unobserved covariates it is reasonable to allow the

Table 1. Mean squared error of prediction (MSEP) for the Syn-1 (separable synthetic), Syn-2 (non-separable isotropic), and Syn-3 (non-separable anisotropic) synthetic data sets

Data set	Non-spatial	GWR	SVC
Syn-1	14·9 (3·91)	7·96 (1·47)	7·00 (0·95)
Syn-2	15·02 (3·32)	7·31 (1·06)	6·02 (0·84)
Syn-3	16·56 (2·19)	14·24 (2·65)	7·4 (1·73)

Values are the mean (standard deviation) over 25 replicates.
GWR, geographically weighted regression; SVC, spatially-varying coefficients.

coefficient associated with DBH to vary spatially. Further, we might expect additional spatial dependence among the model residuals, which can be accommodated by a spatially-varying intercept.

Table 2 offers parameter estimates from the non-spatial, GWR and SVC separable and non-separable isotropic and anisotropic models. The variance parameters in both the non-spatial and SVC models were assigned IG(2,1) priors. All other parameter priors were defined in 'Spatially-varying coefficients'. The non-spatial, GWR, and SVC models produce very similar estimates of the median $\boldsymbol{\beta}$, $\hat{\hat{\boldsymbol{\beta}}}$ and $\tilde{\boldsymbol{\beta}}$, respectively; however, GWR produces a broader range of coefficient values vs. those produced by the SVC models. As expected, the spatial range ϕ_0 for the GWR and separable SVC models are similar and suggest an effective range of 15–20 m on the spatial components of β_0 and β_{DBH}. Further, GWR and separable SVC models produce comparable τ^2. Relaxing the assumption that regression coefficients share a common spatial range and variance, the non-separable isotropic model's ϕ_0 and ϕ_{DBH} suggest the residual spatial process, w_0, has a significantly shorter effective range *c.* 7·44 m than the spatial process associated with the DBH regression coefficient, w_{DBH}, of *c.* 211·27 m. The regression coefficients' spatial process variance, σ_0^2 and σ_{DBH}^2, are also significantly different. These disparities in the spatial processes suggest the non-separable model is warranted.

It is also instructive to create surfaces of the spatially-varying coefficients $\hat{\hat{\boldsymbol{\beta}}}(\mathbf{s})$ and $\tilde{\boldsymbol{\beta}}(\mathbf{s})$ by interpolating their fitted values (Fig. 2). Despite the similarity in the GWR and non-separable SVC fitted value surfaces (a) and (b) the models produce very different representations of the spatially-varying nature of the regression coefficients. Specifically, the disparity in spatial range is clear when comparing (d) and (g) to (e) and (h). From an inferential perspective these surfaces could suggest very different conclusions regarding the covariates' influence.

The patterns in Fig. 2(h) also suggest the possibility of directional dependence in the process associated with β_{DBH}. Therefore, the next candidate model was constructed using the anisotropic correlation function defined in 'Spatially-varying coefficients'. Given the approximately horizontal pattern orientation in Fig. 2(h), a uniform prior $U(-\pi/4,\pi/4)$ was chosen for ψ_{DBH}. Fig. 2(e) does not suggest any directional dependence and therefore an isotropic correlation function for w_0 could be used. However, for exploratory reasons, the anisotropic correlation function with the same prior support as w_{DBH} was used for the β_0 process. The priors on $\boldsymbol{\beta}$ and the variance components are the same as those used for the isotropic SVC models. Referring again to Table 2, the broad posterior distribution for ψ_0 and overlapping effective ranges $\lambda_{0,0}$ and $\lambda_{0,1}$ suggest an anisotropic correlation function was indeed not needed for the model intercept. However, the narrow posterior distribution of ψ_{DBH} and associated non-overlapping spatial ranges suggest there is statistical support for the use of an anisotropic process on β_{DBH}.

Compared to the GWR and separable SVC models the non-separable isotropic and anisotropic models produced signifi-

Table 2. Summary of candidate models' parameter estimates and mean squared error of prediction (MSEP) for the tree height and diameter data set. Non-spatial and SVC parameter posterior credible intervals, 50 (2·5, 97·5) percentiles. GWR intervals 50 (2·5, 97·5) percentiles of $\hat{\hat{\boldsymbol{\beta}}}$

	Non-spatial	GWR	SVC sep.	SVC non-sep. iso.	SVC non-sep. aniso.
$\tilde{\boldsymbol{\beta}}_0$	0·82 (0·76, 0·88)	0·86 (0·13, 1·48)	0·86 (0·77, 0·94)	0·89 (0·69, 1·08)	0·89 (0·69, 1·07)
$\tilde{\boldsymbol{\beta}}_{\mathrm{DBH}}$	0·61 (0·59, 0·63)	0·60 (0·39, 0·85)	0·60 (0·53, 0·66)	0·59 (0·54, 0·64)	0·59 (0·54, 0·64)
ψ_0	–	–	–	–	−0·041 (−0·68, 0·78)
ψ_{DBH}	–	–	–	–	−0·0049 (−0·21, 0·20)
ϕ_0	–	0·2	0·14 (0·09, 0·18)	0·40 (0·25, 0·76)	–
ϕ_{DBH}	–	–	–	0·0142 (0·0101, 0·0899)	–
$\lambda_{0,0}$	–	–	–	–	2·36 (1·41, 4·07)
$\lambda_{0,1}$	–	–	–	–	3·09 (2·01, 4·99)
$\lambda_{\mathrm{DBH},0}$	–	–	–	–	91·03 (64·51, 99·78)
$\lambda_{\mathrm{DBH},1}$	–	–	–	–	27·19 (16·79, 48·78)
σ_0^2	–	–	0·0011 (0·0009, 0·0013)	0·0089 (0·0058, 0·0116)	0·0083 (0·0061, 0·0112)
σ_{DBH}^2	–	–	–	0·0012 (0·0007, 0·0021)	0·0010 (0·0006, 0·0016)
τ^2	0·017 (0·016, 0·019)	0·01	0·0083 (0·0068, 0·0098)	0·0049 (0·0027, 0·0074)	0·0055 (0·0028, 0·0076)
Eff. $\mathrm{Range}_{0,0}$	–	15	21·76 (16·30, 33·34)	7·44 (3·97, 12·09)	7·09 (4·23, 12·22)
Eff. $\mathrm{Range}_{0,1}$	–	–	–	–	9·28 (6·02, 14·98)
Eff. $\mathrm{Range}_{\mathrm{DBH},0}$	–	–	–	211·27 (33·39, 297·02)	273·10 (193·53, 299·34)
Eff. $\mathrm{Range}_{\mathrm{DBH},1}$	–	–	–	–	81·57 (50·36, 146·33)
MSEP	0·019	0·015	0·014	0·014	0·014

GWR, geographically weighted regression; SVC, spatially-varying coefficients.

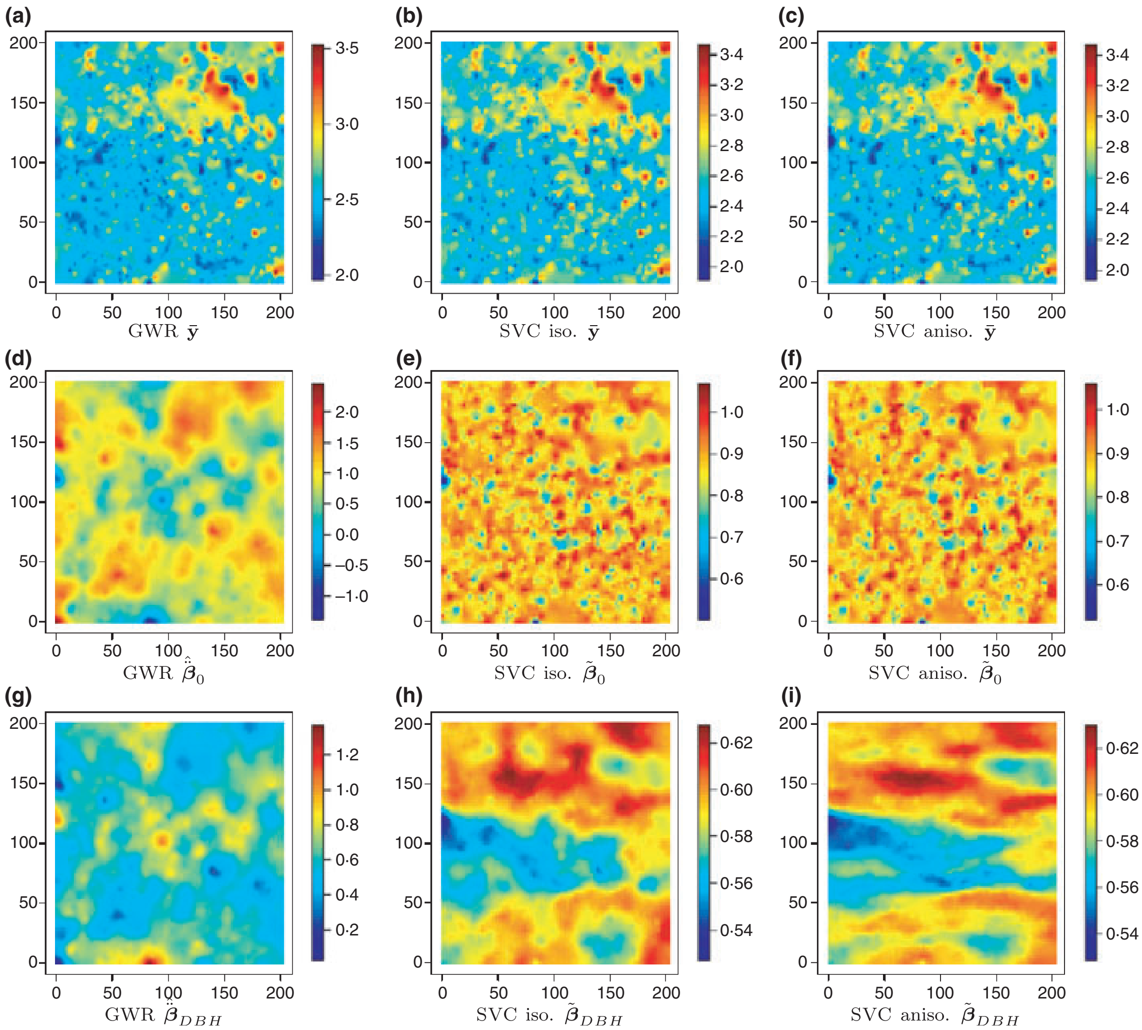

Fig. 2. Interpolated surfaces of model fitted values $\bar{\mathbf{y}}$ and spatially-varying regression coefficients for the tree height and diameter data set. SVC non-separable isotropic and anisotropic model surfaces are based on each location's posterior predictive mean.

cantly lower τ^2. However, based on the holdout set MSEP, there was not a clear advantage among the SVC models and only marginal improvement over the GWR model.

In addition to assessing parameter posterior distributions, we might look to DIC as a tool to evaluate Bayesian model fit. DIC for the non-spatial, separable and non-separable isotropic and anisotropic models were −5458, −6144, −6749 and −6621 respectively. Despite the support for anisotropic model, the non-separable isotropic model had a marginally lower DIC which, based on this criteria, makes it the preferred model.

BIRD SPECIES RICHNESS

Foody (2004) advocated the use of GWR to better understand the relationships between species richness and spatially-varying scale-dependent covariates over large geographic domains. He illustrated this point by reanalyzing bird species richness data observed across sub-Saharan Africa which was originally compiled by Jetz & Rahbek (2002). Foody (2004) accurately states the need to move beyond potentially misleading global regression models which can obscure the space-varying nature of relationships between the outcome variable of interest and covariates.

Recently, GWR has been shown to produce erroneous results in the presence of collinearity among covariates (Wheeler & Calder 2007; Wheeler & Waller 2009). This data set does in fact present several high simple correlations: NDVI and TEMP $\rho = -0{\cdot}83$; NDVI and PREC $\rho = 0{\cdot}79$; TEMP and PREC $\rho = 0{\cdot}84$. Further, inference based on the GWR and separable SVC model might be affected by the use of a common ϕ. Given these considerations the data set is re-examined using the separable and non-separable SVC models.

For this analysis the variance parameters in both the non-spatial and SVC models were assigned IG(2,1) priors. All other parameter priors were detailed in 'Spatially-varying coefficients'.

Candidate models' parameter estimates are given in Table 3. The table shows that GWR produces broader $\hat{\ddot{\boldsymbol{\beta}}}$ 95% intervals compared to those associated with $\tilde{\boldsymbol{\beta}}$. This might be partially attributed to a larger value of ϕ equal to 1·57 which results in shorter effective range and hence greater potential variation in the spatial portion of $\boldsymbol{\beta}$. Dispite the similar median estimates of $\hat{\hat{\boldsymbol{\beta}}}$ and $\tilde{\boldsymbol{\beta}}$ (Table 3) GWR and separable SVC models produced very different regression coefficient surfaces, columns 1 and 2 of Fig. 3, and hence different conclusions regarding the relationship between species richness and the covariates.

The environmental covariates are significant at the 0·05 level in the non-spatial model. In addition, the GWR and separable SVC models suggest covariates contribute significantly to explaining variability in the outcome. However, there likely remains a large portion of unexplained residual variability that might be more effectively apportioned into spatial and non-spatial components using the non-separable model. This redistribution is seen in the last column of Table 3, where σ_0^2 and ϕ_0 are now able to exclusively model the residual spatial dependence. Further, due to the strong collinearity among the covariates it is not surprising that the processes for β_{NDVI}, β_{TEMP} and β_{PREC} exhibit similar estimates of their respective ϕ and σ^2. The extra flexibility of β specific processes significantly reduce the model τ^2 compared to the separable SVC estimates. Again returning to Fig. 3 there are substantial differences in the $\tilde{\boldsymbol{\beta}}$ surfaces generated using SVC separable and non-separable models. These differences play out in both descriptive and statistical inference about the spatial influence of the covariates. The plots in Fig. 4 illustrate regions where $\tilde{\boldsymbol{\beta}}$ are significantly different than zero at the 0·01 level. These plots show the separable and non-separable models can produce very different $\tilde{\boldsymbol{\beta}}$ posterior distributions and hence statistical conclusions. Noting the high correlation between NDVI and PREC and the lack of significance in the $\tilde{\beta}_{\text{PREC}}$, one might consider dropping NDVI or PREC from the model; however, exhaustive analysis of the data is beyond the scope of this illustration. Given the strong correlation among the covariates the full MVGP non-separable model would be the next logical candidate model (see Appendix S4 for results).

Deviance information criterion for the non-spatial, separable and non-separable models were 7562, 2624 and 871 respectively. The lower DIC, τ^2, and MSEP suggest that the flexibility in the β-specific processes of the non-separable is warranted.

COMPUTATION TIME

As described in 'Spatially-varying coefficients' the SVC models are computationally onerous. The most demanding portion of the model fitting process is computing the inverse and determinant of the np $\times$ npΣ_w for each MCMC iteration. For analyses presented here, the three chains of 25 000 MCMC iterations took approximately: 10 min for the $n = 250$ and $P = 3$ synthetic data set; 1 h for the $n = 1793$ and $P = 2$ tree height and diameter data set, and; 2·5 h for the $n = 1199$ and $P = 4$ bird species richness data set.

Conclusion

Cressie *et al.* (2009) provide an excellent discussion on the need to correctly apportion sources of uncertainty when modelling ecological phenomena. It is this partitioning that can help reveal the true nature of the ecological processes of interest. Hoeting (2009) extends this discussion by detailing disadvantages of ignoring spatially structured residuals when using regression to explore ecological data. From a statistical standpoint it is important we accommodate spatial dependence among residuals and non-stationarity of regression coefficients. From a utilitarian perspective, and as seen in results presented here, addressing spatially structured dependence can improve model fit and prediction.

Analysis across large spatial domains is becoming more common due, in part, to increased access to large-scale

Table 3. Summary of candidate models' parameter estimates and mean squared error of prediction (MSEP) for the bird species richness data set. Non-spatial and SVC parameter posterior credible intervals, 50 (2·5, 97·5) percentiles. GWR intervals 50 (2·5, 97·5) percentiles of $\ddot{\boldsymbol{\beta}}$

	Non-spatial	GWR	SVC sep.	SVC non-sep.
$\tilde{\boldsymbol{\beta}}_0$	15·96 (13·08, 18·58)	37·09 (−36·95, 115·90)	35·10 (18·24, 54·59)	31·40 (8·75, 58·96)
$\tilde{\boldsymbol{\beta}}_{\text{NDVI}}$	44·80 (38·99, 50·41)	11·26 (−119·44, 133·44)	8·28 (−7·82, 24·34)	9·85 (2·95, 16·55)
$\tilde{\boldsymbol{\beta}}_{\text{TEMP}}$	−7·56 (−8·97, −6·25)	−4·70 (−42·88, 17·61)	−4·85 (−18·52, 6·39)	−4·36 (−6·81, −1·94)
$\tilde{\boldsymbol{\beta}}_{\text{PREC}}$	−3·20 (−4·46, −2·00)	2·52 (−25·46, 45·49)	2·71 (−9·54, 17·33)	1·02 (−1·35, 3·37)
ϕ_0	–	1·57	0·12 (0·07, 0·16)	0·07 (0·04, 0·12)
ϕ_{NDVI}	–	–	–	1·09 (0·06, 2·91)
ϕ_{TEMP}	–	–	–	0·20 (0·04, 2·73)
ϕ_{PREC}	–	–	–	0·32 (0·04, 2·71)
σ_0^2	–	–	62·45 (47·57, 97·25)	269·94 (159·35, 485·16)
σ_{NDVI}^2	–	–	–	0·55 (0·17, 2·28)
σ_{TEMP}^2	–	–	–	0·49 (0·15, 4·73)
σ_{PREC}^2	–	–	–	0·41 (0·14, 3·61)
τ^2	125·65 (116·64, 134·79)	2·18	1·46 (0·33, 3·59)	0·29 (0·13, 0·70)
Eff. Range$_0$	–	1·91	25·98 (19·30, 40·55)	42·98 (24·73, 77·30)
Eff. Range$_{\text{NDVI}}$	–	–	–	2·75 (1·03, 51·34)
Eff. Range$_{\text{TEMP}}$	–	–	–	14·90 (1·10, 84·94)
Eff. Range$_{\text{PREC}}$	–	–	–	9·29 (1·11, 75·64)
MSEP	128·6	22·07	16·01	15·97

GWR, geographically weighted regression; SVC, spatially-varying coefficients.

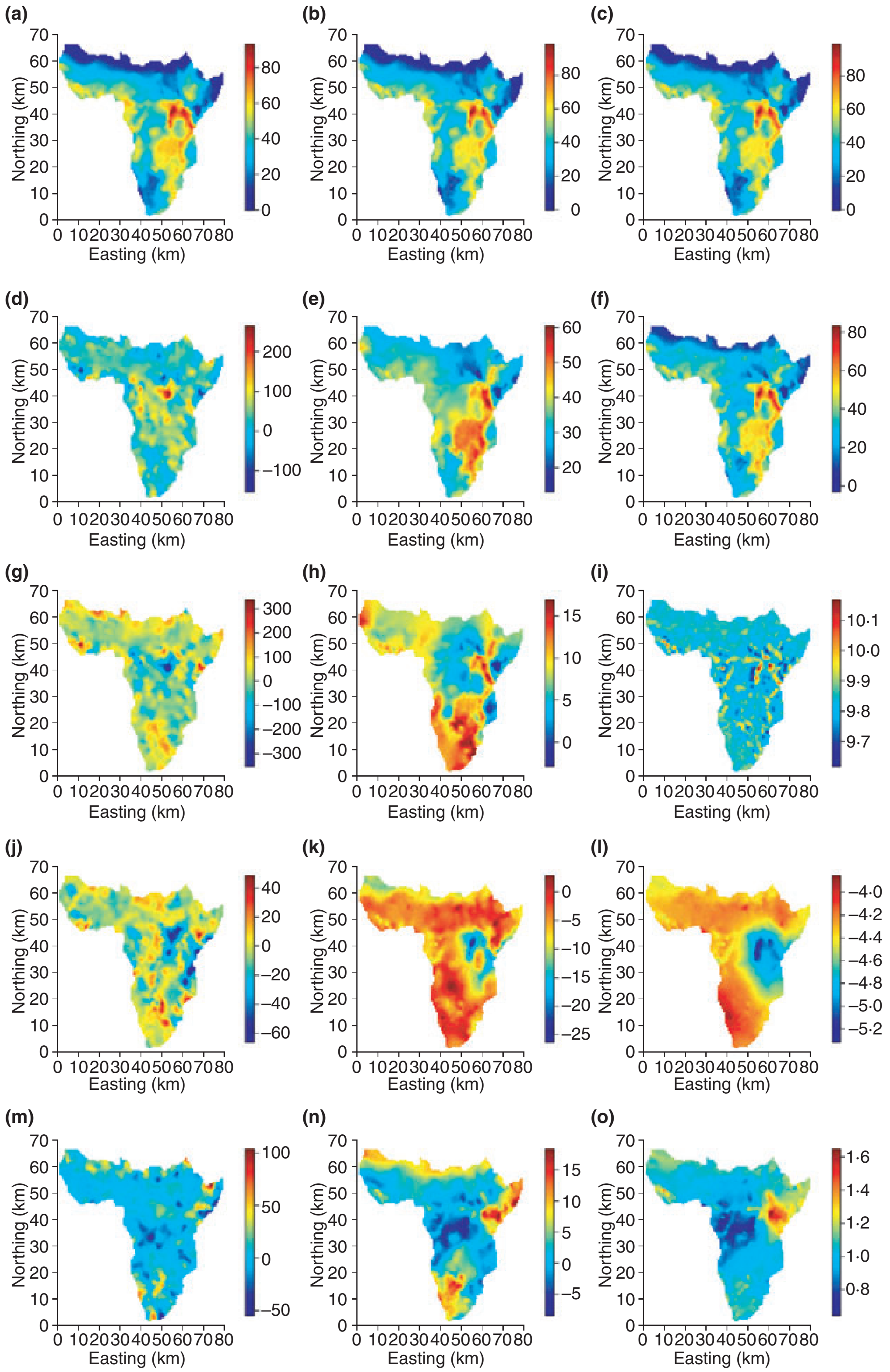

Fig. 3. Interpolated surfaces of model fitted values $\bar{y}$ and spatially-varying regression coefficients for the bird species richness data set. SVC separable and non-separable isotropic model surfaces are based on each location's posterior predictive mean.

ecological data sets and the need to address questions at landscape and regional scales. With expanding domains of interest comes an increased propensity for non-stationarity in the underlying spatial process. In these cases it is not reasonable to assume simple slope coefficients in a regression model can adequately capture the space-varying and scale-dependent relationships between the outcome and covariates.

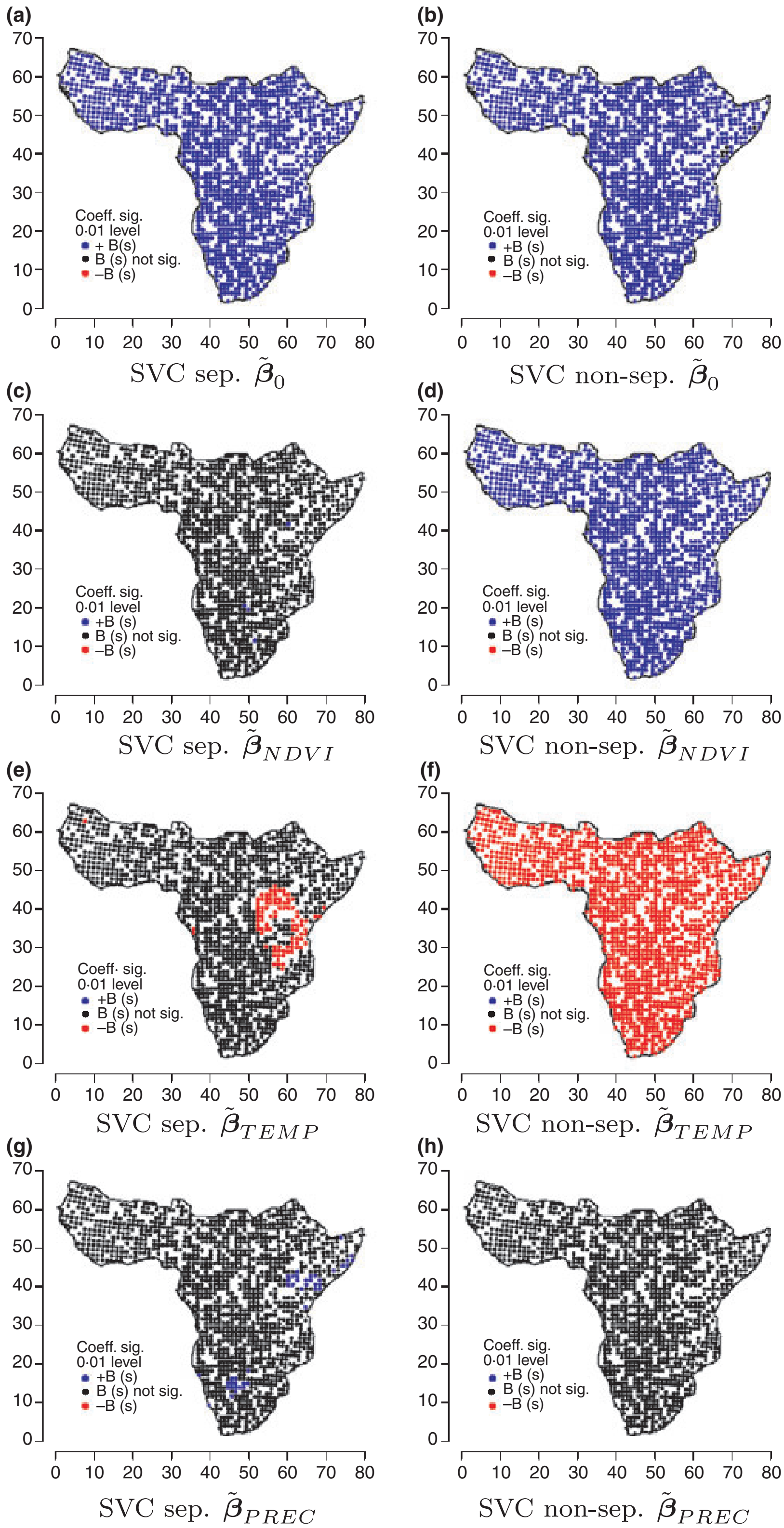

Fig. 4. Points represent bird species richness observations used to fit the SVC separable and non-separable isotropic models. Colours indicate significance and sign of the regression coefficients' posterior distribution at the 0·01 level.

This paper compared two regression approaches that accommodate non-stationarity. GWR is a descriptive approach that uses spatial weights to estimate spatially adaptive coefficients whereas SVC places either a univariate or multivariate spatial process on those regression coefficients that are thought to vary spatially. The advantage of SVC is that it provides a valid probability model from which full posterior inference for all model parameters and subsequent predictions can be made. When interest is in testing hypotheses concerning model parameters or assessing predictive uncertainty the SVC models offer a richer inferential framework.

Analysis of the synthetic data underscored the need for postulated models to approximate the underlying mechanism generating the data. Specifically, in the presence of non-stationarity with different spatial structures on the covariates, GWR models were inferior to the non-separable SVC models. Further, the use of complex correlation structures to define the second order properties of the regression coefficient processes can improve inference as illustrated using the anisotropic processes in the synthetic and forest inventory data analyses. The analysis of the bird species richness data set suggested GWR may produce very different regression coefficient surfaces from those of SVC and hence dramatically different conclusions can be drawn regarding the impact of covariates. Here too, the separable and non-separable SVC models produce quite different regression coefficient surfaces and subsequent inference. Lastly, the bird species richness analysis might contribute to the evidence that GWR is not robust to collinearity among the covariates.

The increased flexibility of SVC comes a cost. As noted by Bolker (2009), from an analytical and computational perspective these are often challenging models to fit. They require the analyst to have training and experience in specifying hierarchical Bayesian models. Specifically, choice of defensible prior distributions for the model parameters is important. Also because several parameters require a Metropolis MCMC step, reasonable starting values and proposal distributions must be specified to ensure chain convergence. Fortunately, there are now many references on these topics, see e.g. Gelman *et al*. (2004) for general Bayesian modelling and Banerjee *et al*. (2004) for guidance specific to hierarchical models for spatial data. Another limitation is the lack of available software to fit SVC models. This however is now partially addressed through new functions in the R package spBayes. By far the greatest challenge to widespread adoption of these models is computational. For even a moderately sized data set, e.g. greater than say 1000 observations and three spatially-varying coefficients, these models will require several days to collect sufficient MCMC samples for valid posterior inference. This is often referred to as the 'big-N' problem in spatial statistics and remains an active area of research. Fortunately, several recent advances have made SVC and similar hierarchical models more accessible to practitioners. Finley *et al*. (2009) offer a review of these methods and details about one in particular called the *predictive process*.

Ecologists are proposing increasingly complex hierarchical models to explore ecological phenomena – beyond simple single level regression. As a result, flexible frameworks are needed to incorporate spatial dependence and non-stationarity across multiple model levels. The hierarchical specification of SVC can be extended to many of these settings. In comparison, GWR should generally be viewed as a tool for descriptive and exploratory data analysis due to its limited inferential scope.

Acknowledgements

The author thanks Dr. Giles M. Foody and Dr. Lianjun Zhang for providing their data sets and suggestions about model comparison. This work was supported by the National Science Foundation DMS-0706870 and the USDA Forest Inventory and Analysis National Program and Forest Health Technology Enterprise Team.

References

Banerjee, S., Carlin, B.P. & Gelfand, A.E. (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, Boca Raton, FL.

Bolker, B. (2009) Learning hierarchical models: advice for the rest of us. *Ecological Applications*, **19**, 588–592.

Cressie, N.A.C. (1993) *Statistics for Spatial Data*, 2nd edn. Wiley, New York, NY.

Cressie, N.A.C., Calder, C.A., Clark, J.S., Ver Hoef, J.M. & Wikle, C.K. (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, **19**, 553–570.

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

Ek, A.R. (1969) Stem map data for three forest stands in northern Ontario. Inf. Rep. O-X-113. For. Res. Lab., Sault Ste. Marie, Ontario.

Finley, A.O., Banerjee, S. & McRoberts, R.E. (2009) Hierarchical spatial models for predicting tree species assemblages across large domains. *Annals of Applied Statistics*, **3**, 1052–1079.

Foody, G. (2004) Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology and Biogeography*, **13**, 315–320.

Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, Chichester.

Gelfand, A.E., Kim, H., Sirmans, C.F. & Banerjee, S. (2003) Spatial modelling with spatially varying coefficient processes. *Journal of the American Statistical Association*, **98**, 387–396.

Gelfand, A.E., Schmidt, A.M., Banerjee, S. & Sirmans, C.F. (2004) Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion). *Test*, **13**, 263–312.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004) *Bayesian Data Analysis*. 2nd edn. Chapman and Hall/CRC Press, Boca Raton, FL.

Guo, L., Ma, Z. & Zhang, L. (2008) Comparison of bandwidth selection in application of geographically weighted regression: a case study. *Canadian Journal of Forest Research*, **38**, 2526–2534.

Hoeting, J.A. (2009) The importance of accounting for spatial and temporal correlation in analyses of ecological data. *Ecological Applications*, **19**, 584–588.

Jetz, W. & Rahbek, C. (2002) Geographic range size and determinants of avian species richness. *Science*, **297**, 1548–1551.

Paciorek, C.J. & Schervish, M.J. (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, **17**, 483–506.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.

Waller L., Zhu, L., Gotway, C., Gorman, D. & Gruenewald, P. (2007) Quantifying geographic variations in associations between alcohol distribution and violence: a comparison of geographically weighted regression and spatially varying coefficient models. *Stochastic Environmental Research and Risk Assessment*, **21**, 573–588.

Wheeler, D. & Calder, C. (2007) An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems*, **9**, 145–166.

Wheeler, D. & Waller, L. (2009) Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *Journal of Geographical Systems*, **11**, 1–22.

Received 20 April 2010; accepted 26 July 2010
Handling Editor: Robert P. Freckleton

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. MCMC algorithms for the non-separable spatially-varying coeffcients model parameters and prediction.

Appendix S2. Multivariate Gaussian process spatially-varying coeffecients model.

Appendix S3. Synthetic data sets parameter estimates and MSEP.

Appendix S4. Supplemental analysis of the bird species richness data set.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.