Abstract

Recent work in local spatial modeling has affirmed and broadened interest in multivariate local spatial analysis. Two broad approaches have emerged: Geographically Weighted Regression (GWR) which follows a frequentist perspective and Bayesian Spatially Varying Coefficients (SVC) models. Although several comparisons between the two approaches exist, recent developments, particularly in GWR, mean that these are incomplete and missing some important axes of comparison. Consequently, there is a need for a more thorough comparison of the two families of local estimators, including recent developments in multi-scale variants and their relative performance under controlled conditions. We find that while both types of local models generally perform similarly on a series of criteria, some interesting and important differences exist.

1 Introduction

Analytical methods for modeling spatial variation in geographic relationships (process spatial heterogeneity) have long been of interest to researchers from both the social sciences and the physical sciences. Techniques for investigating spatial variation in model parameters have been around for over forty years (Casetti, 1972) and the development and application of such models has become increasingly pervasive over the last two decades. Typically, these techniques use spatially defined subsets of the available sample to investigate how the relationships vary across a study area. Two categories of spatial analysis techniques for investigating spatial variation can be distinguished. The first group, regional models, require discrete local subsets to be identified by the analyst *a priori*. This group includes spatial fixed effects models, spatial regime models, and Bayesian hierarchical models that map nested probability models to varying spatial scales. In these cases, the data are partitioned into a finite number of subsets with well-defined membership, so the number of locales that can be compared is fixed at the outset. This makes these models more regional than local in nature and we do not concern ourselves with these models in this paper. In contrast, the second group of models estimate process heterogeneity directly from the data without pre-specified groups, typically providing estimates of a process at every location in a sample, and conceptually at locations not in the sample, which allows for a more flexible and in-depth analysis of the heterogeneity among processes. Examples of such models include eigenvector spatial filter-based local regression (SFLR) (Griffith, 2008; Oshan and Fotheringham, 2016; Oshan and Fotheringham, in press; Murakami et al., 2017), geographically weighted regression (GWR) (Fotheringham, Charlton, and Brunsdon, 1998; Brunsdon, Fotheringham, and Charlton, 1998; Fotheringham,

Brunsdon, and Charlton, 2002; LeSage, 2004), and some kinds of Bayesian spatially-varying coefficient (SVC) models (Gelfand et al., 2003; Banerjee, Carlin, and Gelfand, 2014). Of these three types of local models, the GWR and Bayesian SVC frameworks are more frequently encountered and have been the subject of recent comparisons (Wheeler and Calder, 2007; Waller et al., 2007; Wheeler and Waller, 2009; Finley, 2011).

However, many important features of GWR and Bayesian SVC models remain unexplored. First, many previous experiments either use empirical data where the true coefficients are unknown or use relatively small sample sizes, usually consisting of only a few hundred spatial units, or both. Focus on small sample properties is valid but limits understanding. GWR has been shown to be more robust in larger samples (Páez, Farber, and Wheeler, 2011; Fotheringham and Oshan, 2016), so explorations of small sample performance likely do not reflect the technique's large sample properties. Second, many comparisons provide a somewhat incomplete picture of the overall model performance. In a controlled experiment, where coefficient surfaces are known, models can be compared on many different dimensions, such as goodness of model fit, surface recovery, estimate accuracy, estimate efficiency, as well as other computational and procedural measures of ease of use. Existing empirical comparisons have generally been assessed by model fit, and thus are limited in the intuition they are able to build. Finally, the multi-scale GWR method of Fotheringham, Yang, and Kang, (in press) has not yet been compared Bayesian specifications, such as those previously discussed by Finley, (2011). As a result, an important aspect of this work is to extend the above-mentioned axes of comparison to multi-scale models.

The primary goal of this paper therefore is to explore the quality and meaning of single- and multi-scale specifications of GWR and Bayesian SVC models. As a result, a review of the models and their estimators is first provided. Next, the simulation design used to obtain known non-stationary processes is introduced. Results are then presented, which compare and contrast various features of GWR and SVC models. Specifically, the efficiency and accuracy of each specification's ability to estimate known non-stationary processes are assessed. Computation time and specification generalizability for each model are also considered in order to to better gauge the practical trade-offs of each model. In addition, how the bandwidth parameters change between model types and between sample size is examined. Following this, the overall differences and comparative advantages of each technique are summarized. Consequently, a more holistic comparison and evaluation of these local modeling frameworks is provided.

2 Local Analysis of Spatially-Varying Relationships

The subsequent discussion outlines the single and multi-scale model specifications of both GWR and SVC models and compares the structure of these local estimators.

2.1 Geographically Weighted Regression

Geographically weighted methods, (Fotheringham, Charlton, and Brunsdon, 1998; Brunsdon, Fotheringham, and Charlton, 1998; Fotheringham, Brunsdon, and Charlton, 2002), are a set of techniques to model potential parameter heterogeneity in spatial processes. When focused on the linear model, Geographically Weighted Regression (GWR) techniques model spatially non-stationary parameters using a data borrowing technique. Typically, a Geographically Weighted Regression is stated:

$$Y(s) = X(s)\beta(s) + \epsilon(s) \tag{1}$$

where $Y(s)$ denotes the relationship to a response specific to the area around site s. The data for this local process realization is often related to other locales using a smooth function of distance expressed in a local weighting matrix for each location, $W(s)$. Thus, given the form of $W(s)$, location-specific parameter estimates, $\beta(s)$, can be constructed using standard estimating techniques, essentially conducting a regression at each site:

$$\hat{\beta}(s) = \left(X'W(s)X\right)^{-1} X'W(s)Y \tag{2}$$

Here, $W(s)$ is a diagonal matrix containing the importance weight of the ith observation to the focal site, s. Constructing $W(s)$ can be done in many ways, but common choices of weighting functions are smooth kernel distance functions (Fotheringham, Brunsdon, and Charlton, 2002), parameterized by some measure of the range or scale of the process and which has a basis in the geographical 'law' that nearby things are more related than those that are further away (Tobler, 1970). The range value, often called the "bandwidth," is unknown in practice, so it is typically estimated by optimizing a model fit metric.

A novel development provided by Fotheringham, Yang, and Kang, (in press), multi-scale GWR or MGWR, allows each β_j, $j = 1, 2, \ldots, p$, to have a unique bandwidth. In this case, the estimator for parameter k becomes dependent on the bandwidth used for that parameter:

$$\hat{\beta}(s)_j = \left(X'W(s, \phi_j)X\right)^{-1} X'W(s, \phi_j)Y \tag{3}$$

This model is estimated using an iterative back-fitting algorithm developed from the theory of generalized additive models. First, a set of initial estimates is generated and an initial set of predictions for Y is obtained. Then, the residuals for these predictions are computed. These residuals plus the current set of estimate values for the first term, denoted f_0, are then regressed on X_0 using GWR as a subroutine. This provide an estimate of the bandwidth, ϕ_0 for the *partial* regression between X_0 and f_0. The process then moves on to the second variable, X_1, following the same procedure: a new set of residuals is computed using the updated f_0 and these new residuals plus the current value of f_1 are GW-regressed on X_1 to estimate the partial bandwidth ϕ_1. This continues until the last of the p partial GWRs are computed and each partial smoother f_j updated. This constitutes a single iteration of the backfitting, which iterates until f_j converge. Again, a full presentation of this technique is provided in Fotheringham, Yang, and Kang, (in press).

At this time, no analytical expression is available for standard errors of the MGWR estimates. Parameter uncertainty in β and ϕ_j will be assessed through Monte Carlo replications. The MGWR framework is flexible because it may be used to model processes that exhibit no spatial heterogeneity, have similar spatial heterogeneity, or where each process has a different level and nature of spatial heterogeneity. However, neither GWR nor MGWR treat bandwidth as a stochastic quantity and, therefore, do not provide an estimate of uncertainty associated with the bandwidth estimates.

2.2 Bayesian Spatially-Varying Coefficients Model

An alternative approach to modeling local spatial relationships is to specify spatial structure using a Bayesian hierarchical model. For Bayesian Spatially-Varying Coefficient models with endogenous spatial scales, a specification structure proposed by Gelfand et al., (2003) has become common. For a full discussion of the different possible generalizations of the SVC, Finley, Banerjee, and Carlin, (2007) or Banerjee, Carlin, and Gelfand, (2014) provides an excellent overview. At its core, the SVC specification is a correlated mixed effects model, where the random effect covariance is structured using a spatial correlation kernel.

For the SVC model, first consider a model predicting response Y, using p covariates observed at N sites, grouped into an $N \times p$ design matrix, X. An SVC specification models the response as a function

of a global effect μ and two random effects, ζ and ϵ. For site i, the response becomes:

$$y_i = \sum_{j}^{p} X_{ij}(\mu_j + \zeta_{ij}) + \epsilon_i \tag{4}$$

where ϵ_i is an independent and identically distributed (iid) normal error term with zero mean and variance τ^2, and ζ_{ij} is a spatially-correlated error term for each process j. In this model, μ represents the mean global effect for process j. Then, site-specific parameter estimates, β_{ij}, are recovered by computing $\mu_j + \zeta_{ij}$ for each site. The spatial structure in the local disturbance term is induced in ζ through its hierarchical prior covariance. To state this model in vector form, Gelfand et al., (2003) "stretches" X into an $N \times Np$ matrix, X_s, where each row corresponds to an observation and each set of p columns contains the p covariates observed in that row. This tiling places p-length covariate vectors on the diagonal and zeros elsewhere. Then, ζ can become the $Np \times 1$ collection of random effects where the ith block contains p random effects for each of the j processes, and μ is $Np \times 1$ and contains the p-length vector for the process means repeated N times. With this tiling, the model can be stated:

$$Y = X_s\mu + X_s\zeta + \epsilon \tag{5}$$

After this, the hierarchical prior for ζ is stuctured by a $p \times p$ *between-process* covariance matrix, T, and an $N \times N$ *between-site* correlation matrix, $H(\phi)$, that is a function of ϕ alone. The full covariance matrix for the $Np \times 1$ spatial random effects is:

$$\zeta \sim \mathcal{N}(0, H(\phi) \otimes T) \tag{6}$$

Typically, $H(\phi)$ may be the same kind of spatial kernel function as in a GWR. However, the kernel here affects the covariance of the random effects directly, rather than weighting the data used to estimate local models. This is because the full $H(\phi) \otimes T$ covariance matrix defines a correlation structure between each process at all sites N, embodying the full set of $Np \times Np$ site-process relationships. As in GWR, this kernel may also be more complex, modeling anisotropy or alternative distance metrics. However, the kernel must provide a valid covariance matrix to ensure the resulting covariance is still valid. This precludes the class of "adaptive" nearest-neighbor bandwidths often used in GWR because these result in an asymmetric H. For the remaining parameters, the SVC model's conventional priors are picked for

conditional conjugacy[1] and often estimated *via* Gibbs sampling (Finley, Banerjee, and Carlin, 2007).

To generalize this model to be comparable to an MGWR, a method to model multiple processes with potentially different scales is required. While a specification is suggested in Finley, (2011, Appendix 1), it must be elaborated to make it operational. The case where all covariates vary locally is derived here since this is the focus of the analysis, but including non-varying terms is a straightforward extension. First, we define a new tiling of X, called X_c, which differs from the tiling used in the single-bandwidth SVC model of Gelfand et al., (2003) where stacking occurs in N sets of p. For the multi-process model, we instead stack p sets of N.

Let X_c be an $N \times Np$ matrix that stacks the diagonalized covariates X_j horizontally. This means X_c is p $N \times N$ diagonal matrices:

$$X_c = \begin{bmatrix} Diag[X_1] & Diag[X_2] & \dots & Diag[X_p] \end{bmatrix} \tag{7}$$

Then, let ζ be an $Np \times 1$ column vector created by stacking each processes' N-vector of site-specific spatial effects for each of the p processes. Since each process has its own bandwidth, each ζ_j for $j \in \{1, 2, \dots, p\}$, has its own hierarchical prior. These are modeled separably to match the MGWR specification, which treats subprocesses as conditionally independent of one another.[2] Together, $X_c\zeta$ is the $N \times 1$ vector that matches the correct process-and-site specific random effect ζ_{ij} to the correct covariate x_{ij} in each site:

$$X_c\zeta = \begin{bmatrix} \sum_j^p x_{1j}\zeta_{1j} \\ \sum_j^p x_{2j}\zeta_{2j} \\ \vdots \\ \sum_j^p x_{nj}\zeta_{nj} \end{bmatrix} \tag{8}$$

With this, we can state a local model for the response in vector form:

$$Y = X\mu_\beta + X_c\zeta + \epsilon \tag{9}$$

$$\zeta \sim \mathcal{N}(0, Diag[H(\phi_j)\sigma_j^2]_j^p) \tag{10}$$

$$\epsilon \sim \mathcal{N}(0, \tau^2) \tag{11}$$

[1] It is important to note that the canonical SVC specification's inverse gamma/Wishart priors for scale parameters/covariance matrices in HLMs have been shown to be unintentionally informative in other spatial and aspatial HLM specifications (Gelman et al., 2006; Polson, Scott, et al., 2012; Alvarez, 2014; Simpson et al., 2014)

[2] One may consider the non-separable SVC specifications of Finley, (2011) a more robust model, but the discussion is restricted to separable models here to ensure that similar specifications are compared.

where $Diag[M]_i^k$ is a block-diagonal matrix with M repeated k times along the diagonal. The covariance matrix for ζ is a block-diagonal $Np \times Np$ matrix, so ζ can be constructed by collecting together the draws of each process's distribution of spatial effects separately:

$$\zeta_j \sim \mathcal{N}(0, H(\phi_j)\sigma_j^2) \tag{12}$$

Using this blocking, the local effects for each process can again be recovered by adding together the mean effect and the random perturbations for that process:

$$\beta_{ij} = (\mu_\beta)_j + \zeta_{ij} \tag{13}$$

The fact that each processes' spatial random effects decompose into their own distributions also ensures that, after marginalizing or conditioning on the global parameters, the posterior distribution for each process is independent. A Gibbs sampler that iterates between drawing global model parameters $\{\mu, \tau, \zeta\}$ and the process-specific parameters $\{\phi_j, \sigma_j^2\}$ can be defined. Unlike the blocking used in the single-bandwidth SVC, this allows the MSVC to process p of the $N \times N$ process-specific covariance matrices, rather than handling the full $Np \times Np$ random effects covariance of the single-bandwidth model. The exact conditional posteriors for this specification are provided as supplementary materials using the analogous prior forms to the single-bandwidth SVC.

2.3 Local Estimator Families

While GWR and SVC models enable similar analyses, their estimators are quite different and may be categorized as either having a kernel-smoothing term or a mixed effect term, respectively. A primary feature of the mixed effect specification that defines SVC models is a spatially-patterned interaction effect defined over the entire map to "localize" a global process estimate. Typically, this means that the estimator is divided into one process-specific mean estimate, μ_j, and a random component for each observation in the jth process, ζ_{ij}:

$$x_{ij}\beta_{ij} = x_{ij}\left(\mu_j + \zeta_{ij}\right) \tag{14}$$

This mixed-effect structure may be present regardless of the estimation method, and is often intrinsic to the specification. In the case of the SVC model, it follows directly from its structure as a specific instance

of the more general class of hierarchical linear models (as in Lindley and Smith, 1972). Eigenvector spatial filter specifications can also be recast as a kind of correlated random effects local model that uses spatially-patterned synthetic variables and interaction terms to reconstruct the spatially-patterned parameter surfaces (e.g. Murakami and Griffith, 2015). In addition, this mixed effects structure is also used to model non-stationarity in many hierarchical models that examine regional nonstationarity (Gelman and Hill, 2006). Even the Bayesian GWR specification of LeSage, (2004) yields an augmented GWR estimator that induces this random effect structure. All of these mixed effect specifications provide an explicit model that relates the local fluctuations to the global process structure. Typically, unless special structure is found in the data, the estimation of the random spatially-correlated components (ζ_{ij}) requires a full $N \times N$ covariance matrix for the effects. Since this $N \times N$ covariance matrix is frequently some distance-decay spatial correlation function, these techniques often characterize the correlation between observations over the entire map, and suggest how correlated any ζ_{ij} and ζ_{kj} are, given the bandwidth estimate.

Kernel-smoothed specifications typically used in GWR are fundamentally different from mixed effect local specifications used in the SVC model. First, site-specific estimates β_i are constructed directly from a local data set constructed using the geographic weights matrix $W(s)$. Second, kernel-smoothed GWR estimators do not exhibit a formal link between global process means and local correlated fluctuations around this mean, since each site estimate set comprises a unique regression on the derived data. Indeed, there is no special local GWR estimator – each local weighted regression is a distinct model and *is* the global regression when a kernel is infinite. Each local estimator, conditional on the process bandwidth estimate(s), only considers the data local to that site, which may overlap or coincide with other models' local data. Thus, a GWR estimation is not a "full-map" technique in the way that the mixed effect specifications are. The (M)GWR estimator conducts N regressions on potentially-overlapping sets of data where N_i, the sample size of the regression at site i, is often much smaller than N. In fact, in adaptive bandwidth specifications, the N_i may be estimated from the data directly and may embody a different notion of map distance at each site, since densely-packed areas may have a smaller effective kernel than areas where observations are sparse. This style of nearest-neighbors model is currently unavailable for mixed effects specifications, since the nearest-neighbor adaptive bandwidth specification may be asymmetric and may invalidate the covariance matrix.

The lack of distributional linking between the local effect estimator and an estimate of global process mean in GWR is noted offhand in many comparisons (e.g. Wheeler and Waller, 2009; Finley, 2011),

but its importance is often elided. This difference indicates GWR is more similar to local regression in parametric space (Cleveland, 1979; Cleveland and Devlin, 1988) than the correlated disturbance (i.e., mixed effects) concepts that inform SVC local model specifications. However, this locality comes at a cost, since GWR does not provide a full probabilistic model relating its bandwidth estimate, local coefficients, and the global process mean. While the Geographically Weighted Lasso (Wheeler, 2009) can introduce "local-to-global" regularization in GWR specifications, the free parameter governing shrinkage is arbitrary in contrast to the regularization in Bayesian hierarchical estimators underlying the SVC model (Draper and Van Nostrand, 1979). Therefore, this difference in basic estimator structure should be noted explicitly in comparisons of the model specifications.

3 Simulation Design

In order to compare the single- and multi-scale variants of GWR and SVC local models, a known set of patterns which are illustrative of non-stationary processes are needed. Each pattern or surface is a single *process* whose pattern is expected to be recovered by the models as a set of site-specific marginal effects β_{i0}, β_{i1} and β_{i2}. Many previous studies of local modeling techniques were computationally limited, using smaller samples sizes, simpler patterns, or both. For example, (Wheeler and Calder, 2007), estimated single-scale models on surfaces comprised of two lattices where $n = 100$, one with a vertical gradient and one with a horizontal gradient. Therefore, to obtain a controlled experimental data set with a significantly larger sample size and surface diversity, three distinct surfaces with various levels and types of heterogeneity are specified on regular lattices using sample sizes of $n = 625$ and $n = 2500$.

Shown in Figure 1, the surfaces are generated to ensure that the spatial heterogeneity for each surface is substantially different. That is, in the first surface, all sites have the same value, so the bandwidth is expected to be large. In the second surface, sites in the top left are most dissimilar from those in the bottom right, and the resulting estimated bandwidths from multi-scale models are expected to be significantly smaller than those for the first surface. Finally, in the last surface, four distinct peaks provide a pattern where nearby sites are even more likely to be similar, and is anticipated to produce the most "local" bandwidth estimates of the three patterns (i.e., the smallest bandwidth). Y is constructed from these surfaces using an $N \times p$ matrix of synthetic covariates $(X_{1,2} \sim \mathcal{N}(0, 2))$ and a random

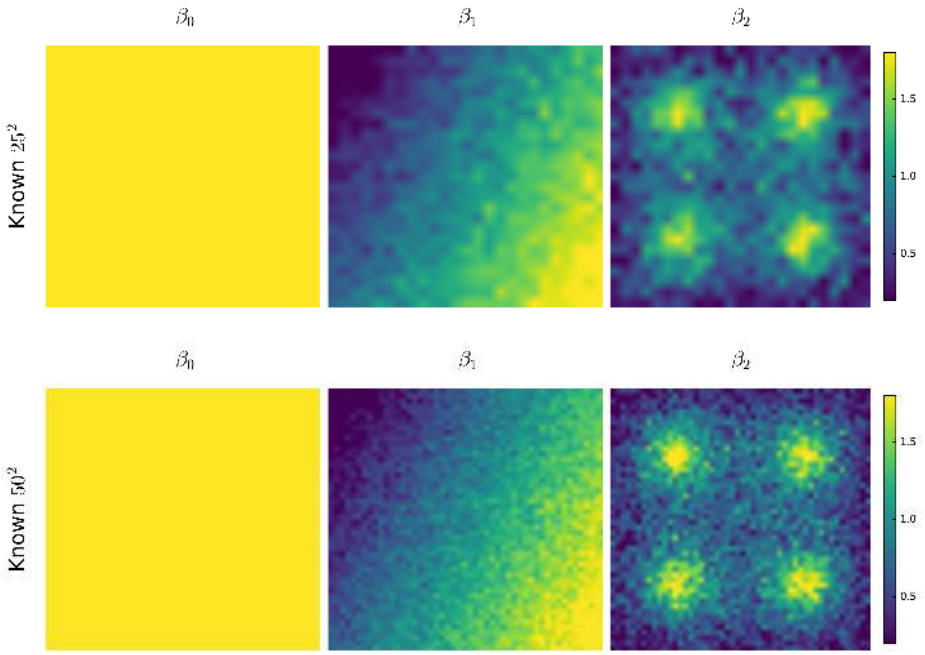


Figure 1: Three covariates in two differently-sized processes. These are the "known" patterns, with the side-length labeled. So, the $n = 625$ known pattern is labeled as "Known 25^2."

unobserved component, ϵ, with fixed deviation, $\sigma = 2$:

$$Y = X\beta + \epsilon \tag{15}$$

For each sample size, a GWR and MGWR model are estimated first using fixed bandwidth(s) and then using adaptive bandwidth(s). In addition, an SVC and multi-scale separable SVC (MSVC) model are estimated on this synthetic data. The SVC and MSVC are estimated using Metropolis-within-Gibbs sampling, as suggested by Gelfand et al., (2003), while the GWR and MGWR are fit by optimizing the corrected Akaike Information Criterion (Fotheringham, Brundson, and Charlton, 2004). The MGWR model also utilizes an iterative back-fitting algorithm described in Fotheringham, Yang, and Kang, (in press).

4 Results

Once single-scale and multi-scale variants of both GWR and SVC models are fit,[3] a comparison of several factors can be made between the two modeling frameworks. These include: general process surface recovery, process estimate accuracy, model fit, the magnitude and nature of bandwidth estimate values, calibration and diagnostics, efficiency of process estimates, scalability regarding sample size, flexibility of methodology, and accessibility. The adaptive bandwidth fits are omitted for brevity, because they are nearly identical to the fixed exponential distance-based kernel in this case due to the regularity in the density of sample points. It should be noted that in more realistic scenarios, the adaptive bandwidth formulation possible in GWR might prove to have superior properties to the Bayesian SVC models where an adaptive bandwidth framework does not exist.

4.1 Surface Recovery

All four of the techniques (GWR, MGWR, SVC, MSVC) are able to generally recover the pattern of heterogeneity in each process (Figure 2). That is, site-specific posterior estimates for the (M)SVC models and the point estimates for the (M)GWR models visually resemble the patterns apparent in the known process surfaces (Figure 1). However some differences in the ability of the four models to recover the parameter surfaces are evident. For (M)GWR, sample size is important and both GWR and MGWR perform better with larger sample sizes: for the 25x25 matrix GWR and MGWR produce overly smoothed local parameter estimates. As expected, MGWR outperforms GWR for both samples but the SVC model outperforms the MSVC model in both the 25x25 and the 50x50 data sets. The fact that the single-scale model outperforms the multi-scale model is unexpected, since the latter should (in theory) be able to conform better to local variation in each process. However, the differences across the four models, particularly for the large sample, are relatively minor compared to the surfaces which would be generated by global models — all four models are able to identify the nature of the spatial variation in the processes being modeled.

[3] Implementations in Python were used to estimate the models, which were all computed on a platform equipped with two Intel Xeon ES-2640, 2.6GHz and 64GB of available memory. The Python implementations leveraged Numpy for array computations and standard lower-level linear algebra routines were run in parallel using the Intel Math Kernel Library. For the Bayesian SVC routines the Markov chains were run for 22000 iterations, after which convergence was assessed. Convergence for the multi-scale Bayesian model was difficult to identify, and an alternative heuristic will be discussed in Section 4.5. The single-scale SVC converged rapidly to a stable interval estimate in all parameters. In all cases, the last 6000 observations from the Bayesian models will be used for estimate comparison without thinning.

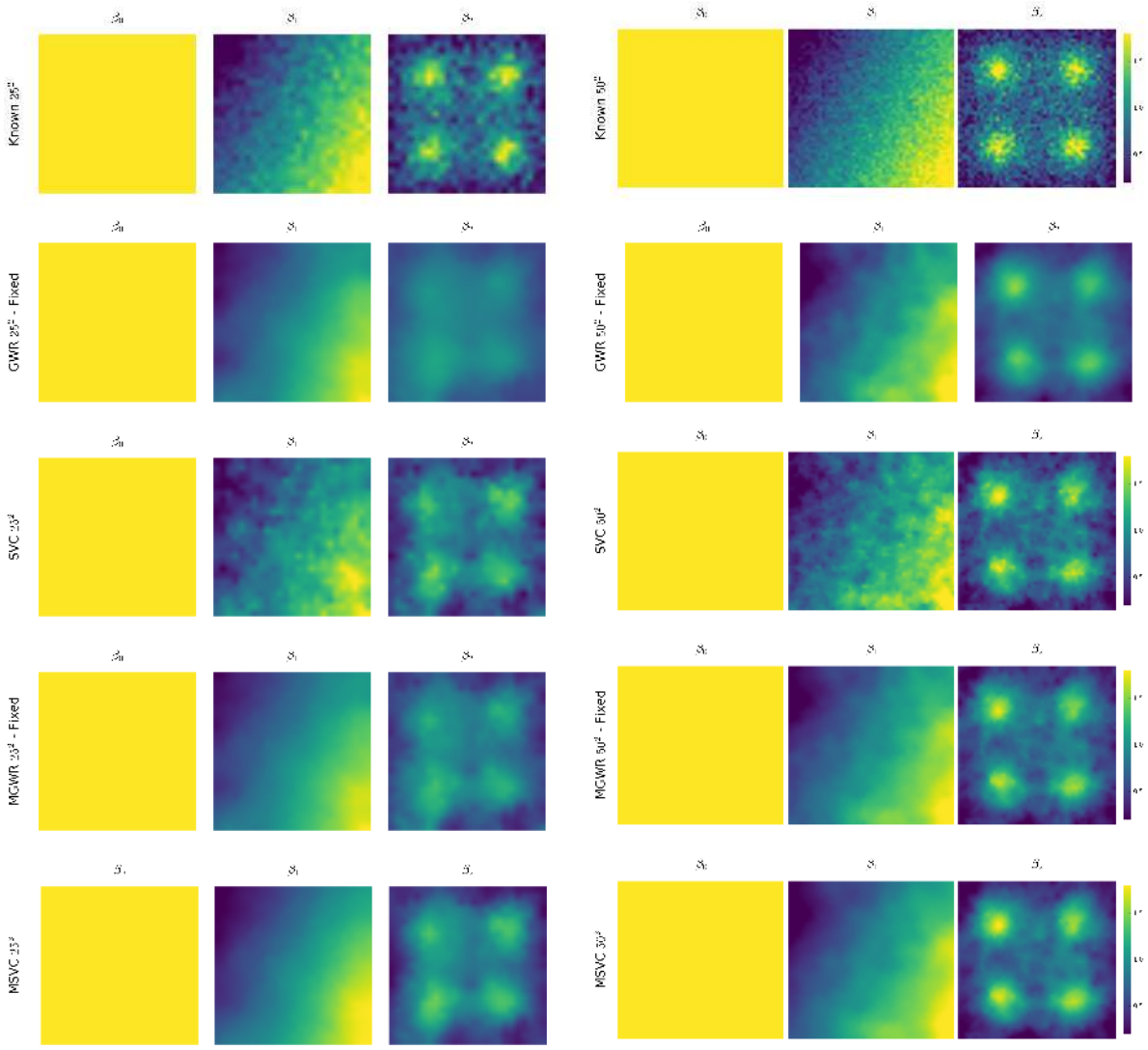


Figure 2: Maps of the known parameter surfaces and the recovered parameter surfaces for each specification & problem size. The first row is the known parameter surfaces. The next two rows contain the single-scale models and the bottom two rows contain the multi-scale models. On the left are the $n = 625$ models and on right the $n = 2500$. The row labels note each scenario's side length (e.g. $25^2 = n = 625$) and model name, (M)SVC or (M)GWR.

	GWR 25	SVC 25	GWR 50	SVC 50
MSE(β_0)	0.017398	0.019426	0.001106	0.009084
MSE(β_1)	0.019069	0.031364	0.013702	0.025584
MSE(β_2)	0.086998	0.046174	0.044762	0.046174

Table 1: Mean squared error for estimate recovery in single-scale models

4.2 Estimate Accuracy

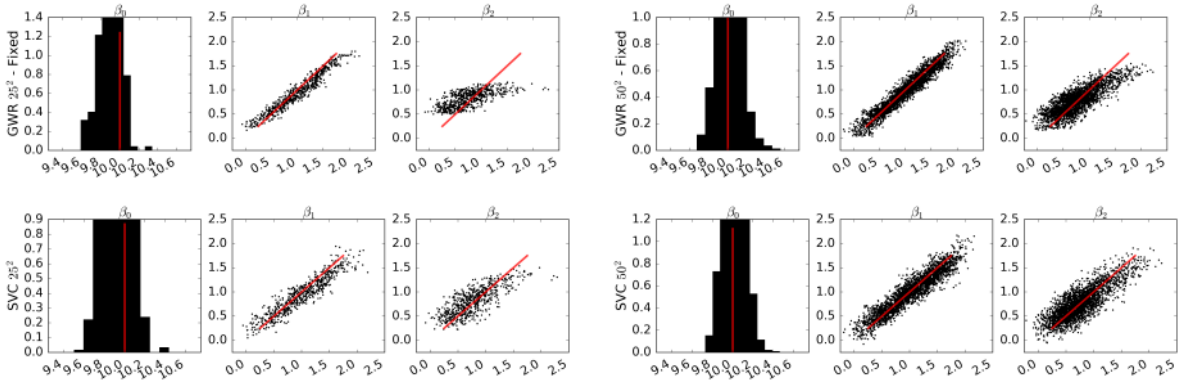
Quantifying the accuracy of the parameter estimate surfaces provides a more detailed analysis of process recovery. The mean squared error and the correlation between the known surfaces and the estimated surfaces for the single-scale models are shown in Table 1 and 2, respectively. At the smaller scale, GWR estimates β_0 and β_1 with lower error than the SVC, but estimates β_2 with almost twice the error as the SVC; it fails to pick up the complexity of the β_2 parameter surface, as well as the SVC model does. However, at the larger scale, GWR estimates all three of the surfaces with lower error than the SVC. For the global intercept, the MSE for the SVC model is over 8x larger than that of the GWR model and for the β_1 surface the MSE for the SVC model is almost twice that of the GWR model. This supports previous work that indicated that GWR can provide more accurate estimates with larger sample sizes (Páez, Farber, and Wheeler, 2011; Fotheringham and Oshan, 2016), especially for more complex surfaces such as β_2. The correlations between the predicted and actual surfaces of β_1 and β_2 suggest that GWR outperforms SVC in replicating the local values of β_1 for both sample sizes but SVC outperforms GWR in replicating the local values of β_2. However, the difference in performance narrows markedly as sample size increases and the performance of the two models is virtually the same for the 50x50 matrix. It is interesting to speculate on what would happen in even larger samples.

The source of estimate error in the GWR and SVC models appear to be different according to Figure 3a, which demonstrates the overall performance for each surface in each model against the scenario of perfect replication (the solid red line). GWR estimates for the $25x25$ matrix tend to err by *over-smoothing* the β_2 surface, under-estimating the larger values and over-estimating the smaller values. This can be seen as a form of *bias*, which is likely the reason GWR had almost double the mean squared error than the SVC for β_2 for this matrix. However, this over-smoothing is much less for the $50x50$ matrix, although there is still noticeable under-prediction of the larger β_2 values. In contrast, the SVC model appears to have noisier estimates than GWR, though the noise is consistent regardless of the magnitude of β_2.

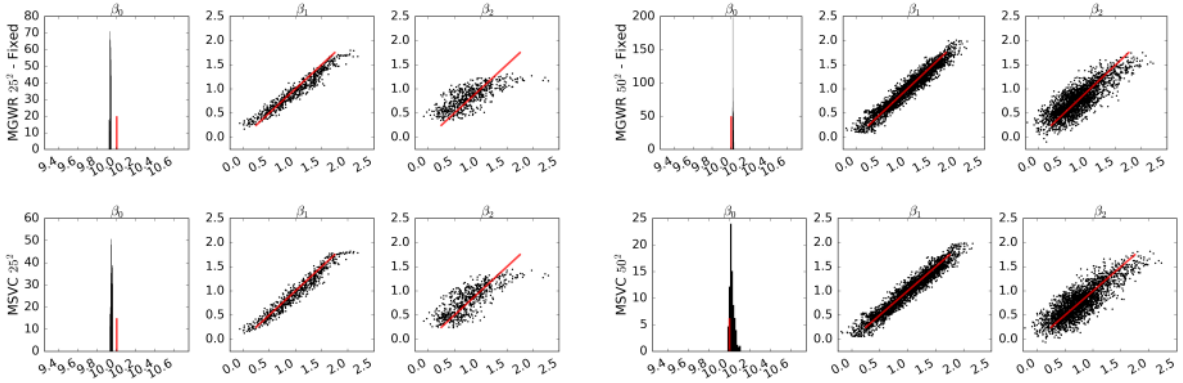
The multi-scale model estimates tend to be more consistent than the single-scale model estimates

	GWR 25	SVC 25	GWR 50	SVC 50
$\rho(\beta_1, \hat{\beta}_1)$	0.967274	0.933194	0.962017	0.935944
$\rho(\beta_2, \hat{\beta}_2)$	0.670550	0.780006	0.817833	0.819730

Table 2: Correlation between estimated and known value for single-scale models



(a) Single-scale Models



(b) Multi-scale Models

Figure 3: Scatterplots and histograms relating the estimated and known parameter values. Red lines indicate perfect estimate recovery, and black dots or bins reflect the values of the recovered estimates. β_0 is the same everywhere, so recovered parameters are presented as a histogram. Labeling is akin to that in Figure 2

	MGWR 25	MSVC 25	MGWR 50	MSVC 50
MSE(β_0)	0.012363	0.000452	0.004400	0.003012
MSE(β_1)	0.015673	0.014279	0.019564	0.013920
MSE(β_2)	0.047987	0.044703	0.065979	0.054315

Table 3: Mean squared error for estimate recovery in multi-scale models

	MGWR 25	MSVC 25	MGWR 50	MSVC 50
$\rho(\beta_1, \hat{\beta}_1)$	0.969841	0.971346	0.965528	0.966188
$\rho(\beta_2, \hat{\beta}_2)$	0.744311	0.784360	0.823949	0.827317

Table 4: Correlation between estimate and known value for multi-scale models

in that MSVC recovers all three known surfaces with lower error than MGWR at either scale (Table 3), although the visualization of the errors in Figure 3b indicate the results from both models are virtually identical. Interestingly, for the smaller scale the β_1 estimates from the MSVC and MGWR model appear to exhibit the same over-smoothing properties observed in the GWR model (Figure 3b), but this disappears in both models when sample size is increased. Table 4 illustrates that multi-scale models produce estimates that are more strongly correlated the the single scale models with the true values at both sample sizes, although the discrepancies are marginal in some cases. While the MSVC generally has marginally higher levels of correlation between estimates and known values for β_1 and β_2, the MGWR model for the smaller sample size provides a large improvement over the correlation yielded by the GWR model for β_2. It appears that MGWR may mitigate the over-smoothing present in the single-scale GWR estimates for β_2, which encourages the use of MGWR over GWR.

4.3 Model Fit

For single-scale models, the SVC outperforms the GWR model in terms of the root mean squared error between the predicted values, $\hat{Y}$, and the known dependent variable values, Y, for both samples, although the difference in fit decreases as sample size increases (Figure 4). In contrast, the prediction errors for the multi-scale models are much closer, with the MSVC outperforming the MGWR for the smaller sample size and the MGWR outperforming the MSVC for the larger sample size. However, this difference is marginal and the two multi-scale models have very similar goodness-of-fits. All the models appear to provide unbiased predictions for Y over the entire range of magnitudes, despite the previously discussed bias for the largest and smallest magnitudes in the coefficient surface estimates. The

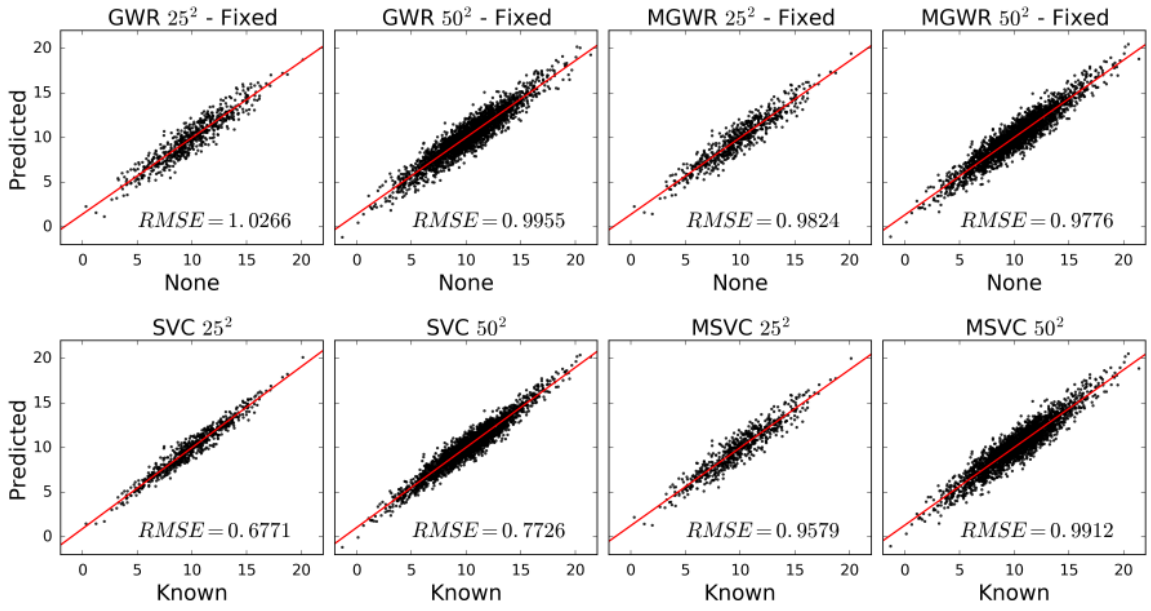


Figure 4: Scatterplots of predicted & fitted response.

spatial distribution of prediction errors are omitted since they exhibit no clear spatial pattern, indicating little-to-no spatial autocorrelation in the model residuals.

A surprising result is that the single-scale SVC has *the best fit overall* in terms of the root mean squared error. When moving from a single-scale to a multi-scale GWR, the accuracy improves slightly, which is expected due to the fact that each processes is being separately modeled rather than as one average process. However, when moving from a single-scale to a multi-scale SVC, accuracy degrades substantially, which is unexpected. This may be due to the fact that the SVC explicitly incorporates between-surface correlation, while the MSVC assumes each surface to be independent of the other surfaces, similarly to MGWR. It is possible that a MSVC with a non-separable covariance for β_j that explicitly models between-surface correlation like the SVC might perform even better, but would also likely be significantly more complex to estimate than the separable MSVC analogous to MGWR that is considered here.

4.4 Bandwidth Estimation

In endogenous-scale local models, bandwidth is interpreted as a proxy for the scale at which a process occurs. If estimates of bandwidth differ between the (M)GWR and (M)SVC models or between sample sizes for each specification, this may reveal differences in the substantive interpretation of the bandwidth

parameter.

In both model frameworks, the bandwidth represents a "real world" distance, though distance is used differently by each specification. For the SVC models, the bandwidth controls the spatial correlation kernel for the random effects. Thus, a large bandwidth parameter, ϕ, means that even distant spatial random effects are strongly correlated. For the GWR models, the bandwidth describes the width of the kernel used to include and weight observations for the local data in each site-specific model. Therefore, a large ϕ includes further observations than a smaller ϕ. Table 5 describes the bandwidth values for both single-scale model forms for the two sample sizes.

This provides an insight into what "bandwidth" implies in the two models. As can be seen in Figure 1 the underlying patterns of the three parameter surfaces are the same in both the small (25^2) and (50^2) samples. The surface of β_0 is constant for both samples but the 'granularity' of both β_1 and β_2 increases. That is, although the same overall spatial pattern of heterogeneity exists in both samples, the local variance of the parameters increases in the larger sample size. Conceptually, the bandwidth in GWR appears to describe the overall spatial pattern of heterogeneity in the local parameter surface whereas the bandwidth in the SVC model incorporates two effects – the overall spatial pattern of spatial heterogeneity and local variation. Neither of these interpretations is 'better' than the other; they are simply different. If one wants an overall measure of the degree of spatial heterogeneity, the GWR bandwidth is more appropriate as this is the same in both scales. If ones wants a measure that is a function of both overall heterogeneity and local variation, the SVC bandwidth is more appropriate. This important difference is highlighted in the multi-scale results presented in Table 6.

Both MGWR and MSVC models produce relationship-specific bandwidths estimates rather than a single 'average' value as in the GWR and SVC models. This is important if the different processes represented in the model vary at different spatial scales. The resulting bandwidth estimates for the MGWR and MSVC models in Table 6 highlight this situation. Our three sets of parameters (shown in Figure 1) vary at different scales: β_0 is invariant over space; β_1 exhibits a medium amount of spatial heterogeneity with a trend running north-west to south-east; and β_2 exhibits the largest amount of spatial heterogeneity with four evenly-spaced peaks of high values surrounded by lower values. A transect of values across the three surfaces from the north-west to the south-east as in Figure 5 characterize the complexity of the spatial heterogeneity in each surface; β_0's transect is a line with a slope of 0; β_1's transect is heterogenous with a single trend; β_2's transect is heterogenous with multiple trends. The estimated bandwidths for both the MGWR and MSVC models reflect the different patterns with the

smallest bandwidth being estimated for β_2 and the largest for β_0. Consequently, both models generate important information on the different scales of heterogeneity in the three different processes, which is obscured in the GWR and SVC models. The results for the MGWR models in comparison to the GWR model results in Table 5 exemplify this. For the $25x25$ sample, the GWR bandwidth is 2.03 where the three separate bandwidths for β_0, β_1, β_2, in the MGWR model are 17.06, 2.14, and 1.25, respectively, reflecting the different levels of spatial heterogeneity in the three surfaces. The GWR 'average' bandwidth is essentially a weighted average of the two 'behavioral' parameters, β_1 and β_2, where the weight is related to the relative influence of X_1 and X_2 on y.

These results are replicated for the applications of GWR and MGWR to the $50x50$ sample. Interestingly, the GWR estimate of ϕ is effectively the same when $n = 625$ or $n = 2500$ whereas the estimate of ϕ for the SVC model doubles when the grid grows from 25^2 to 50^2. The ϕ_0 estimate from MGWR doubles when the study area is increased from $n = 625$ to $n = 2500$, but the bandwidth estimate for β_1 and β_2 remain approximately the same. This doubling of magnitude of the bandwidth for the intercept (17.06 for the smaller sample size and 34.79 for the larger sample size) appear to be related to the study area size since both appear to be approximately half of the maximal distance in their respective study area (35.36 and 70.72). This indicates that MGWR seems to ascribe half of the maximal study area distance as an estimate of a "global" process (i.e. stationarity in β_0) in the case of a regular grid. As with the GWR results, the bandwidths for β_1 and β_2 appear to be invariant to sample size, reflecting the similarity of the general patterns of the spatial heterogeneity in the two parameter surfaces.

The bandwidth estimates for the MSVC models must be interpreted in a different manner than the MGWR bandwidths, since they are used directly in the covariance matrix, rather than in generating local data. Notably, the MVSC bandwidth estimates are much larger than that for the MGWR bandwidth estimates (or any other model). In fact, they are larger than any possible pairwise distance within the study area. For example, in the $n = 2500$ case, the largest distance between two points is along the diagonal of the grid, which is approximately 70 units in magnitude. Even β_2, the process that is theorized to have the smallest scale, has a bandwidth estimate that is over four times this maximal distance. To assist with an interpretation of the MSVC bandwidths, it is helpful to examine a correlation matrix of the MSVC instead of directly interpreting the values of ϕ_j. The correlation matrix for the random effects in an (M)SVC model takes its structure from the spatial kernel function, $H(\phi_j)$. Given a functional form for H, a negative exponential function in this case, the correlation in the spatial random effects depends on ϕ_j alone. For the MSVC, the distributions of site-to-site spatial correlations $\rho(i, j)$, in ζ_j resulting from $H(\phi_j)$

	GWR 25	SVC 25	GWR 50	SVC 50
ϕ	2.03	5.54	2.00	10.22

Table 5: Bandwidth estimates for single-scale models

	MGWR 25	MSVC 25	MGWR 50	MSVC 50
ϕ_0	17.06	489.25	34.79	45426.03
ϕ_1	2.14	324.74	2.24	896.61
ϕ_2	1.25	85.87	1.31	317.33

Table 6: Bandwidth estimates for multi-scale models

for many values of ϕ_j can be visualized (Figure 6). As ϕ_j increases, correlations in the random effects between sites rapidly increase to 1. The solid black line in Figure 6 represents the median correlation implied by ϕ_j, and the dotted black envelope indicates the 5th and 95th percentiles of between-site correlation at that value of ϕ_j, respectively. For the β_0 and β_1 surfaces, most of the random effects are strongly correlated, while the β_2 surfaces have moderate-to-high between-site correlations. Between the two sample sizes, the vastly different ϕ_j estimates result in nearly the same amounts of between-site correlation, though correlation between all sites is slightly higher in the larger lattice than in the smaller one.

4.5 Calibration & Diagnostics

Local models typically have a larger number of free parameters that need to be tuned than do traditional global models and this has implications for both computing time and the calculation of diagnostics. As is common for spatial parameters in hierarchical models, the bandwidth(s) in the (M)SVC models have non-standard distributions, and must be sampled using Markov Chain Monte Carlo techniques (Brooks et al., 2011). The other (M)SVC parameters have conditionally conjugate prior choices, yielding a single step within a Gibbs sampler. In the case of the SVC, standard concerns of Markov Chain Monte Carlo (MCMC) estimation apply and there are a wealth of methods to assess model convergence (Cowles and Carlin, 1996). However, the bandwidth in the MSVC model presents unique challenges for identifying chain convergence. Since one of the processes analyzed here has no spatial variation (i.e., global process), the bandwidth estimate may increase without bound due to the fact that marginal changes to ϕ provide effectively the same model fit. Thus, it is useful to employ the inter-site correlation displayed in Figure 6 and to consider a trace of ϕ_j stable when the correlations implied by ϕ_j remains constant. Since

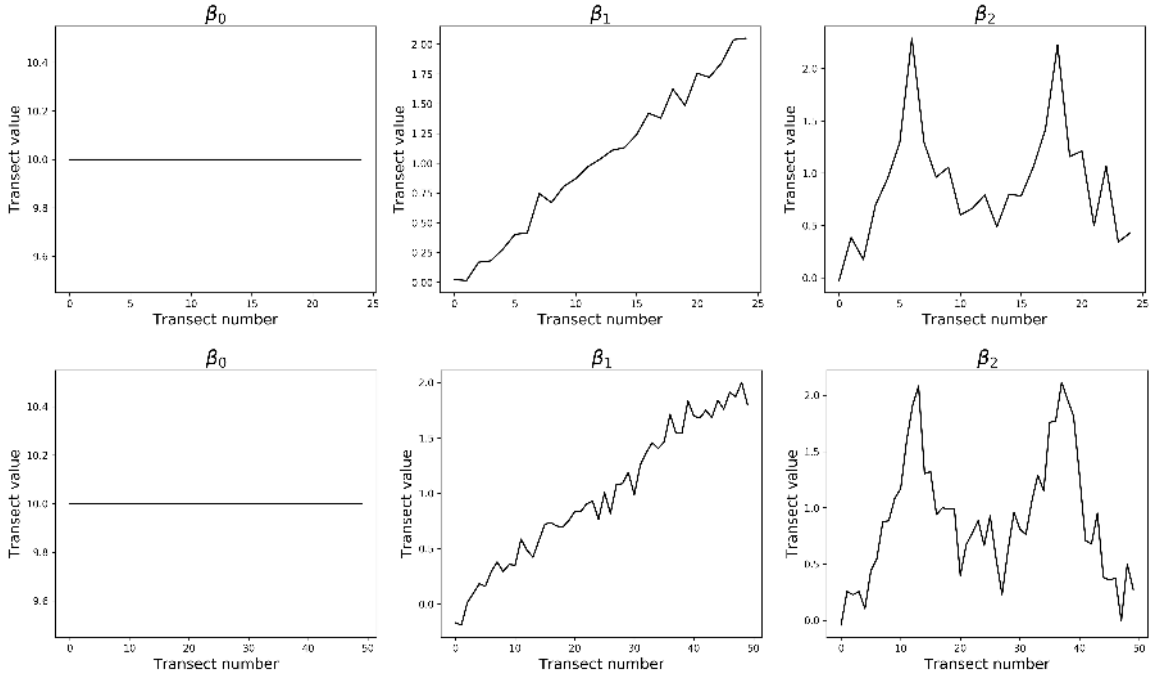


Figure 5: Plots of values for each known parameter surface sampled along a diagonal transect from the north-west to the south-east. These plots characterize the heterogeneity in the parameter surfaces: β_0's transect is a line with a slope of 0; β_1's transect is heterogenous with a single trend; β_2's transect is heterogenous with multiple trends. The top row is for the 25×25 sample and the bottom row is for the 50×50 sample.

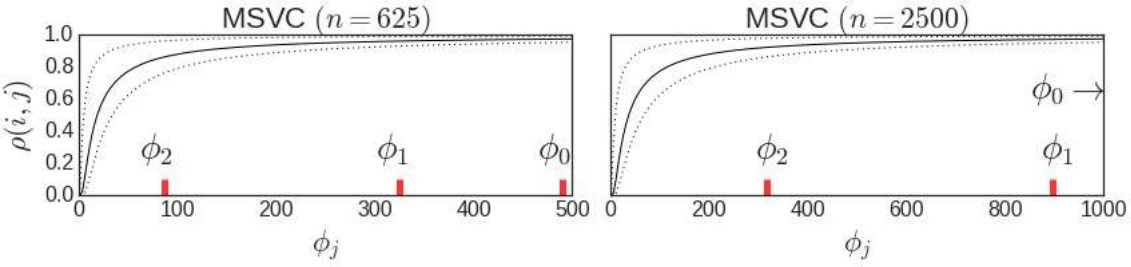


Figure 6: Correlations in spatial random effects implied by different values of ϕ_j in two problem sizes. The black line denotes the median inter-site correlation at a given ϕ_j, and the dotted black lines mark the 5th and 95th percentiles of the distribution of correlations at ϕ_j.

it is not possible to tell ahead of time whether a given surface will be "global" and produce unbounded ϕ_j, this heuristic is necessary. Alternatively, one may restrict the range of ϕ to an arbitrary fraction of the maximum map distance, as is common the GWR framework. However, the fact that these parameters have substantively-different behavior between the two specification types means an appropriate domain restriction for ϕ may be different for an SVC than for a GWR, so focusing instead on the convergence of between-site correlation should be preferred to arbitrary domain restrictions.

For GWR, model calibration is trivial since the bandwidth is first selected by optimizing a model fit or information criterion, in this case a GWR-specific version of the Akaike information criterion, and all other parameters are computed conditionally on the derived bandwidth. However, model calibration of MGWR is more complex since the bandwidths of each process are estimated conditionally on the other bandwidths. Therefore, an iterative bandwidth estimation routine, which uses a general additive modeling (GAM) framework, continues until there are little-to-no changes in some score function and the bandwidths are considered to have converged. Here, convergence in MGWR is assessed by a smooth "score of change" function outlined in Fotheringham, Yang, and Kang, (in press). This score is the sum of squared changes between the current estimates and the previous estimates, divided by a total sum of squares in the current iteration. First, let $X_{i,j}$ denote the observation of the jth covariate at the ith site. Likewise, let β_{ij} be the estimate of the jth process at the ith site. Then, the score of change for iteration t over all surfaces $X\beta$ is:

$$SOC_{X\beta} = \sqrt{\frac{\sum_j^p \frac{\sum_i^n (X_{ij}^{(t)}\beta_{ij}^{(t)} - X_{ij}^{(t-1)}\beta_{ij}^{(t-1)})}{n}}{\sum_i^n \left(\sum_j^p X_{ij}^{(t)}\beta_{ij}^{(t)}\right)^2}} \tag{16}$$

where t denotes the current iteration and $t-1$ the previous iteration of the fitting technique. This score function is based on results from the wider GAM literature and is further discussed by Fotheringham, Yang, and Kang, (in press). It is possible that other score functions may yield better model fit, though this is left for future research.

4.6 Efficiency of Estimation

Uncertainty about the estimated values of the β surfaces is analyzed using estimated standard errors. Standard errors for the single-scale GWR models are computed analytically using the same weight matrix that is used during parameter estimation (Fotheringham, Brunsdon, and Charlton, 2002). A Bayesian analogue to this standard error is constructed using the posterior standard deviation of the sampled co-

efficients at each site for the (M)SVC models. However, there exists no analytical expression for the standard error in a MGWR at this time. Therefore, we construct standard errors for MGWR using a Monte Carlo simulation technique. Since the distribution of X_1 and X_2 was controlled, independent, and random, we resample them and reconstruct a new Y vector using the same known surfaces shown in Figure1 and the random X and the idiosyncratic error ϵ. The distributions of the estimated standard errors for the (M)GWR models and the posterior standard deviations from the (M)SVC models are shown in Figure 7 and 8 for the single-scale and multi-scale models, respectively.

For the theoretical processes employed in this research, the M(GWR) models consistently produce surface estimates with less uncertainty. GWR standard errors are much smaller than the SVC posterior standard deviations. On average, the posterior standard deviation for β_2 and β_1 in the SVC is between 2 and 4 times as large as the analytically-derived GWR standard error. The Monte Carlo MGWR standard errors are also slightly smaller than the posterior standard deviation for the MSVC. The MSVC posterior standard deviations are approximately 1.5 times the Monte Carlo standard errors for β_1 and β_2 in the MGWR model. The MGWR standard error for β_0 is only slightly below the MSVC posterior standard deviation, though.

It is also possible to analyze the uncertainty associated with the individual bandwidth estimates. While no parametric model for ϕ_j was provided in this case, the estimates for ϕ_j in the MGWR Monte Carlo experiment yielded consistent estimates. These estimates also tend to cluster much more tightly around the mean estimate of ϕ_j than those for the posterior standard deviations from the MSVC model.[4] Since the magnitudes of ϕ_j are so strongly different between the two methods, the coefficient of variation provides a better measure of the relative noise associated with estimates. Here, we use an empirical estimate of the coefficient of variation:

$$\hat{CV} = \frac{\sigma(\hat{\phi_j})}{\hat{\phi_j}} \tag{17}$$

where $\sigma(\hat{\phi_j})$ is the posterior standard deviation for ϕ_j in an MSVC model and the standard deviation over Monte Carlo replications in the MGWR case. This coefficient of variation for the MGWR bandwidth estimates are nearly one tenth the coefficient of variation for the MVSC estimates for both ϕ_1 and ϕ_2. Thus, the bandwidth estimate over Monte Carlo replications is substantially more precise than that for ϕ_j in the MSVC model for the local surfaces.[5] This relationship reverses for ϕ_0, though, where the MSVC

[4] By itself, this does not indicate higher *precision* for this estimate, though: the relevant distributional theory for the bandwidth estimate in a (M)GWR over Monte Carlo replications is not defined.

[5] This comparison is approximate and only holds if the MSVC MCMC chains are geometrically ergodic, which is difficult to identify in practice and has no sufficient condition. Further, *replications* of an MGWR would be most succinctly compared to

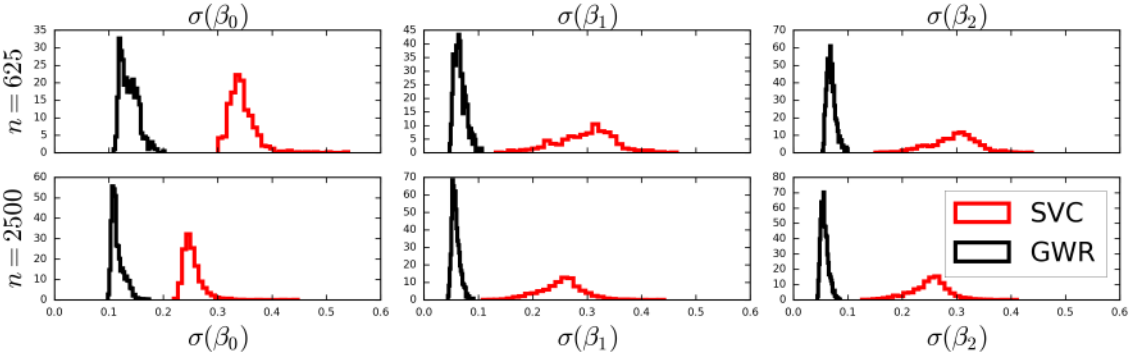


Figure 7: Standard Errors for GWR and posterior standard deviation for the SVC model

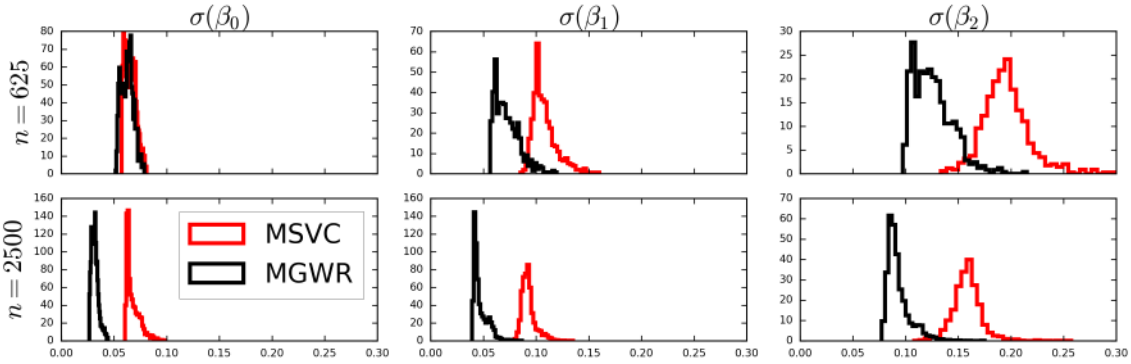


Figure 8: Posterior standard deviation for β in the MSVC and Monte Carlo estimates of standard error for MGWR

has a much smaller coefficient of variation than the GWR. Regardless, a Monte Carlo technique that uses known surfaces is not applicable where the data generating process is unknown. Therefore, no estimate of uncertainty of ϕ_j (or ϕ in a single-scale model) is available for (M)GWR in applied settings while the posterior standard deviations used to assess uncertainty in the (M)SVC models is still applicable.

4.7 Scalability

In terms of scalability, Banerjee et al., (2008a) note that efficient sampling of SVC-type models grows cubic in N, due to requiring the Cholesky factorization of the $Np \times Np$ covariance matrix. GWR estimators, however, only require N sets of $p \times p$ covariances, which avoids this particular problem. Instead, the computational burden of GWR techniques is that N models are fit iteratively during bandwidth estimation. The computational burden is further increased in MGWR as the algorithm essentially carries out

replications of the MSVC, not the posterior standard deviations in a single MSVC.

GWR 25	SVC 25	GWR 50	SVC 50	MGWR 25	MSVC 25	MGWR 50	MSVC 50
.4	186.63	3.76	4834.61	3.6	233.68	31.433	4324.131

Table 7: Computation time in minutes.

several GWR estimations until the score of change function converges. The computing times to calibrate all four models with both data sets are given in Table 7.

In practice, the single-scale models were generally faster than the multi-scale models (Table 7), but the savings were more pronounced for GWR models than they were for SVC models. Ultimately, the (M)GWR models require much less overhead in terms of computational time than the (M)SVC models. These timings demonstrate that GWR is *much* faster than 22000 iterations of a Gibbs sampler for either type of SVC model. While all 22000 iterations are required for the MSVC models to sample after convergence, the SVC converges in around 5000 iterations. Noting this, the single-scale SVC draws around 94 samples per minute when $n = 625$ and around 5 samples per minute when $n = 2500$. In the time a GWR finishes, the SVC would only draw 37 samples in the smaller scenario and around 19 samples in the larger scenario. A sampling of around 5000 draws would still require a significantly larger amount of time than that used by the GWR in either sample size. Regardless of the rate of convergence, it is highly unlikely that these Bayesian models will be time-competitive with GWR-derived methods unless special care is taken to increase the speed of the sampler or to approximate the problem.

Many different methods to address Markov Chain Monte Carlo estimation performance for SVC models are available (Gelfand et al., 2003; Banerjee et al., 2008b; Eidsvik et al., 2012a). Notably, estimating SVC models using integrated nested Laplace approximation may yield significant performance benefits over direct Gibbs sampling (Eidsvik et al., 2012b) and has also been demonstrated to be useful for other families of local spatial models (Lindgren and Rue, 2015). In addition, knot-based subsampling strategies, like those explored by Banerjee et al., (2008a) may make SVC's tractable for larger sample sizes, though they can also be employed in a GWR framework, and thus are unlikely to provide a significant *advantage* for (M)SVC models over the (M)GWR models. Another major advantage for the (M)GWR modeling framework is that it may be easily parallelized, since it essentially consists of an ensemble of distinct (and related) regressions. This means that even models with very large sample sizes may be estimated efficiently in terms of computation time if high-performance computing resources are available.

4.8 Flexibility

Both the GWR and SVC modeling frameworks draw from more general modeling paradigms, which allow them to accommodate various statistical extensions and to be applied in many different contexts. However, some popular extensions to GWR, such as adaptive kernel weighting (i.e., nearest-neighbors), are not currently available for the SVC framework. Furthermore, GWR is easily extended to non-Gaussian probability models, such as the Poisson and Binomial models, through the use of generalized linear models. These forms are also available for SVC-style models, such as the Poisson-SVC suggested by Waller et al., (2007). Thus, GWR is more flexible, since more exotic asymmetric kernel weighting functions can be used. However, developments in adaptive covariances for SVC models may narrow this gap.

4.9 Accessibility

Accessibility encompasses computation time, the software available for each technique, and the ease with which a model can be specified. A discussion of the main computational burden of each modeling frameworks has already been provided and it has been demonstrated that GWR requires significantly less computation time. High-quality GWR implementations exist under free software licenses for both command-line and graphical user-interface style software environments. In contrast, Gibbs sampling is typically used for the SVC models, which presents an obstacle for practitioners. While the general Gaussian process models underlying the SVC can be fit in `spBayes`(Finley, Banerjee, and Carlin, 2007), the actual SVC model that produces local process estimates like GWR is not supported. In addition, the SVC-family of models may be estimable in Bayesian frameworks, like Stan (Carpenter, 2015) or OpenBUGS (Thomas and O'Hara, 2004), but these lack the ability to exploit the sparsity inherent in the random effects covariance matrix. In terms of specifying models, GWR models may be formulated using the language of ordinary regression, and specifications can be understood in terms of the regression formulas that have become common in many statistical computing environments. Even with Bayesian frameworks like Stan, the SVC model (and Bayesian models in general) may require more knowledge of the underlying mathematics in order to state them and program them in Stan, OpenBUGS, or another programming language for estimation. Given that the SVC typically requires longer computation times and more expert knowledge to employ correctly, it is clear that GWR is a more accessible framework for applied work. However, as Bayesian modeling increases in popularity, perhaps the two frameworks may become equally accessible.

5 Discussion & Conclusion

By comparing the SVC and GWR modeling frameworks, several similarities become apparent. Although the (M)GWR and (M)SVC frameworks were developed independently, have their origins in different statistical traditions (frequentist and Bayesian), and are mathematically and logically distinct, they nevertheless share the common goal of accounting for process spatial heterogeneity and appear to produce very similar results in terms of model predictions and local parameter estimates. Both frameworks have at least two versions — one which assumes that all the processes being modelled exhibit the same degree of spatial heterogeneity (GWR and SVC) and one which allows the degree of spatial heterogeneity to vary by process (MGWR and MSVC).

However, this paper identifies several axes along which the two frameworks differ. First, the interpretation of the bandwidth parameter is *not* transferrable between the two frameworks because the correlation kernel used to specify spatial mixed effects in the SVC simply behaves differently from the data-borrowing kernel of a GWR. Importantly, the two frameworks displayed different bandwidth behavior when the sample size of the spatial processes was increased. While GWR bandwidth estimates are typically consistent regardless of the sample size, the SVC bandwidth estimates increase when sample size increases while the spatial pattern is held constant. This incongruity between bandwidth interpretations makes it difficult to assess what the appropriate behavior of bandwidth estimates should be for surfaces with different patterns, levels of spatial heterogeneity, and sample sizes. Defining the substantive meaning of bandwidth estimates is an under-explored avenue of research that deserves future attention. Specifically, this raises questions about how to interpret bandwidth and how hypotheses about process scale might be tested. Are bandwidth estimates specific to a single study area or statistical method? If not, can bandwidths from different surfaces, model specifications, or spatial processes be related at all? Exploring the substantive relationship between process, surface structure, model specification, and bandwidth would build a more robust understanding of spatial processes in the context of local statistical modeling. We have taken first steps in this direction, and believe further work in this domain could illuminate novel understandings of local models that go beyond surface examination of model fit.

Another difference between the two modeling frameworks stems from the technical specification of the bandwidth(s). GWR specifies bandwidths as *deterministic* quantities that are considered fixed when estimating the coefficient surface. In contrast, the SVC specifies the bandwidth as a random variable in a Bayesian context, using a prior and yielding a posterior distribution of plausible values. Therefore, the

SVC bandwidth point estimate has *intrinsic* uncertainty that may be quantified by the standard deviation of its posterior distribution, while GWR does not produce a measure of uncertainty or estimation error for its bandwidth estimates. This difference probably explains why GWR provides smaller standard errors for the local parameter estimates. In contrast, quantities that depend on ϕ, namely the local random effects and recovered coefficient surfaces, are expected to be noisier in the SVC because uncertainty about ϕ is present in the posterior sample. This uncertainty may be seen as a benefit, since the process estimates may be over-confident if uncertainty about bandwidth estimates does not propagate into the process estimates. Even though an estimate of uncertainty about ϕ in MGWR was constructed using Monte Carlo simulation in this case and indicated that the precision of the optimal bandwidth might be greater in GWR than in the SVC framework, this is not available in empirical work, where the distribution of X is not controlled.

A third difference between the two modeling frameworks is that MGWR lacks an analytical expression for the standard errors of the process estimates. In this research Monte Carlo replications were used in order to estimate the standard errors, but this is only possible for experimental research where the processes are specified. Therefore, developing a feasible estimate of the MGWR standard errors is another important topic for future research. Until one is developed, MSVC specifications retain this advantage over MGWR. Estimated standard errors on the local parameter estimates are available in both the GWR and SVC models.

Despite the lack of process inference in MGWR, the GWR framework still displayed strong merit in terms of efficiency of estimation, scalability, flexibility, and accessibility. Given these advantages, it is clearly important to continue building on the MGWR model of Fotheringham, Yang, and Kang, (in press). However, this does not imply that the SVC framework should not also be further developed. It was ultimately the single-scale SVC that achieved the best model fit even though both of the multi-scale models produced process estimate surfaces that more closely resembled the known surfaces. This indicates that further study of the advantages of multi-scale local models versus single-scale models deserves significantly more attention, regardless of the modeling framework. In particular, the development of a robust, computationally tractible muli-scale Bayesian SVC specification with full covariance both between sites and between processes is a clear next step.

Formal specification tests, like those available for spatial econometric models (Florax, Folmer, and Rey, 1998), could be helpful in assessing whether multi-scale or single-scale variants should be preferred, rather than relying solely on a model fit criteria. This is particularly important for applied settings

where the true processes are unknown and it is not possible to quantify the accuracy of the process estimates as was done here. Until such tests are available, it seems reasonable to prefer the single-scale SVC model for achieving the best in-sample predictions of the dependent variable. However, the model does not identify any differences in the spatial scale over which processes operate, which in some applications could be the primary interest. In this study, there was no analysis of out-of-sample prediction, which could be another direction of future research into multi-scale local models. A final extension of this work that might be useful is the investigation of a non-separable MSVC model that incorporates *between-process* and *between-site* correlation simultaneously, which is more closely related to the single-scale SVC and may also achieve a higher model fit. However, the cost of estimating such a model may be prohibitively expensive for all but the most modestly sized datasets.

Finally, other axes of dissimilarity between the two local modeling frameworks include flexibility and accessibility. The (M)GWR framework would appear to produce more intuitive distance measures of the scales over which processes operate and these can also be extended to an adaptive bandwidth framework where bandwidth is measured in terms of the number of nearest neighbors for which weights are non-zero and which has intuitive appeal as well as being efficient. The (M)GWR framework can also be extended to other non-regression frameworks such as spatial interaction modeling (Kordi and Fotheringham, 2016), discriminant analysis (Brunsdon, Fotheringham, and Charlton, 2007) and kriging (Harris, Charlton, and Fotheringham, 2010). The (M)SVC framework, on the other hand, has appeal in terms of its flexibility in being able to encompass a wide variety of error-covariance structures. The (M)GWR framework is more limited in this regard being based on the concept of 'data borrowing' and the fundamental geographical principle that "near things are more related than more distant ones"; a principle that also applies to processes. Ultimately, the choice of which model to employ to study process spatial heterogeneity might come down to personal preference, access to software and familiarity with the broad statistical framework in which each model sits.

This paper has presented the most comprehensive comparison of GWR and SVC models to date. While there are some significant divergences between the two modeling frameworks, overall they may be more alike than they differ, which provides strong evidence in favor of local statistical modeling and the investigation of non-stationary processes. A primary novelty of the comparisons made here is that each framework was extended to include processes that occur at multiple and potentially different scales. Though multi-scale local models are relatively recent, it appears that the differences between single-scale and multi-scale models may be just as great as the differences between the two modeling frame-

works. The discrepancies observed here seem to be driven by the differences in estimator structure and choice of how bandwidth is specified. In general, local models are more complex than global models, and the same is true of multi-scale local models compared to single-scale local models. Therefore, much work still needs to be done to understand multi-scale models and make their implementations simpler and widely available. As these local modeling techniques are applied in various new contexts, new insights about the models themselves may also be acquired.

References

Alvarez, Ignacio (2014). “Bayesian inference for a covariance matrix”. In: *arXiv preprint arXiv:1408.4050*.

Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand (2014). *Hierarchical modeling and analysis for spatial data*. Crc Press.

Banerjee, Sudipto et al. (2008a). “Gaussian predictive process models for large spatial data sets”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.4, pp. 825–848.

— (2008b). “Gaussian predictive process models for large spatial data sets”. In: *J. R. Stat. Soc. Series B Stat. Methodol.* 70.4, pp. 825–848.

Brooks, Steve et al. (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

Brunsdon, Chris, Stewart Fotheringham, and Martin Charlton (2007). “Geographically Weighted Discriminant Analysis”. In: *Geographical Analysis* 39.4, pp. 376–396. ISSN: 0016-7363. DOI: 10.1111/j.1538-4632.2007.00709.x. URL: http://dx.doi.org/10.1111/j.1538-4632.2007.00709.x.

Brunsdon, Christopher, A. Stewart Fotheringham, and Martin Charlton (1998). “Geographically Weighted Regression-modelling spatial nonstationarity”. In: *Journal of the Royal Statistical Society* 47.3, pp. 431–443.

Carpenter, Bob (2015). “Stan: A Probabilistic Programming Language”. In: *Journal of Statistical Software*.

Casetti, Emilio (1972). “Generating Models by the Expansion Method: Applications to Geographical Research*”. In: *Geographical analysis* 4.1, pp. 81–91.

Cleveland, William S (1979). “Robust locally weighted regression and smoothing scatterplots”. In: *Journal of the American statistical association* 74.368, pp. 829–836.

Cleveland, William S and Susan J Devlin (1988). “Locally weighted regression: an approach to regression analysis by local fitting”. In: *Journal of the American Statistical Association* 83.403, pp. 596–610.

Cowles, M.K. and B.P. Carlin (1996). “Markov chain Monte Carlo convergence diagnostics: a comparative review”. In: *Journal of the American Statistical Association* 91.434, pp. 883–904.

Draper, Norman R and R Craig Van Nostrand (1979). “Ridge regression and James-Stein estimation: review and comments”. In: *Technometrics* 21.4, pp. 451–466.

Eidsvik, Jo et al. (2012a). “Approximate Bayesian inference for large spatial datasets using predictive process models”. In: *Comput. Stat. Data Anal.* 56.6, pp. 1362–1380.

Eidsvik, Jo et al. (2012b). “Approximate Bayesian inference for large spatial datasets using predictive process models”. In: *Computational Statistics & Data Analysis* 56.6, pp. 1362–1380.

Finley, Andrew O. (2011). “Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence: Spatially-varying coefficients models”. en. In: *Methods in Ecology and Evolution* 2.2, pp. 143–154. ISSN: 2041210X. DOI: 10.1111/j.2041-210X.2010.00060.x. URL: http://doi.wiley.com/10.1111/j.2041-210X.2010.00060.x (visited on 10/29/2015).

Finley, Andrew O, Sudipto Banerjee, and Bradley P Carlin (2007). “spBayes: an R package for univariate and multivariate hierarchical point-referenced spatial models”. In: *Journal of Statistical Software* 19.4, p. 1.

Florax, Raymond, Henk Folmer, and Sergio J. Rey (1998). “The Relevance of Hendry’s Econometric Methodology: Experimental Simulation Results for Linear Spatial Process Models.” In: *rsue* under submission.

Fotheringham, A. S., C. Brundson, and M. Charlton (2004). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. New York: Wiley.

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

Fotheringham, A. S., M. E. Charlton, and C. Brunsdon (1998). “Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis”. In: *Environment and Planning A* 30.11, 1905ă–ă1927. DOI: 10.1068/a301905. URL: http://www.envplan.com.ezproxy1.lib.asu.edu/abstract.cgi?id=a301905 (visited on 09/17/2014).

Fotheringham, A. Stewart and Taylor M. Oshan (2016). “Geographically weighted regression and multicollinearity: dispelling the myth”. en. In: *Journal of Geographical Systems*. ISSN: 1435-5930, 1435-5949. DOI: 10.1007/s10109-016-0239-5. URL: http://link.springer.com/10.1007/s10109-016-0239-5 (visited on 09/21/2016).

Fotheringham, A. Stewart, Wenbai Yang, and Wei Kang (in press). “Multi-scale Geographically Weighted Regression”. In: *Annals of the American Association of Geographers*.

Gelfand, Alan E et al. (2003). “Spatial modeling with spatially varying coefficient processes”. In: *Journal of the American Statistical Association* 98.462, pp. 387–396.

Gelman, Andrew and Jennifer Hill (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, Andrew et al. (2006). “Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)”. In: *Bayesian analysis* 1.3, pp. 515–534.

Griffith, Daniel A (2008). “Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR)”. In: *Environment and Planning A* 40.11, pp. 2751–2769.

Harris, Paul, Martin Charlton, and A. Stewart Fotheringham (2010). “Moving window kriging with geographically weighted variograms”. en. In: *Stochastic Environmental Research and Risk Assessment* 24.8, pp. 1193–1209. ISSN: 1436-3240, 1436-3259. DOI: 10.1007/s00477-010-0391-2. URL: http://link.springer.com/10.1007/s00477-010-0391-2 (visited on 05/15/2017).

Kordi, Maryam and A. Stewart Fotheringham (2016). “Spatially Weighted Interaction Models (SWIM)”. en. In: *Annals of the American Association of Geographers* 106.5, pp. 990–1012. ISSN: 2469-4452, 2469-4460. DOI: 10.1080/24694452.2016.1191990. URL: http://www.tandfonline.com/doi/full/10.1080/24694452.2016.1191990 (visited on 05/15/2017).

LeSage, James P (2004). “A family of geographically weighted regression models”. In: *Advances in spatial econometrics*. Springer, pp. 241–264.

Lindgren, Finn and Håvard Rue (2015). “Bayesian spatial modelling with R-INLA”. In: *Journal of Statistical Software* 63.19.

Lindley, Dennis V and Adrian FM Smith (1972). “Bayes estimates for the linear model”. In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–41.

Murakami, Daisuke and Daniel A Griffith (2015). “Random effects specifications in eigenvector spatial filtering: a simulation study”. In: *Journal of Geographical Systems* 17.4, pp. 311–331.

Murakami, Daisuke et al. (2017). “A Moran coefficient-based mixed effects approach to investigate spatially varying relationships”. en. In: *Spatial Statistics* 19, pp. 68–89. ISSN: 22116753. DOI: 10.1016/j.spasta.2016.12.001. URL: http://linkinghub.elsevier.com/retrieve/pii/S2211675316301798 (visited on 02/08/2017).

Oshan, Taylor and A. Fotheringham Stewart (2016). “A Closer Examination of Spatial-Filter-Based Local Models”. In: *International Conference on GIScience Short Paper Proceedings* 1. ISSN: 2573-783X. DOI: 10.21433/B31104T0T6DS. URL: http://escholarship.org/uc/item/04t0t6ds (visited on 09/14/2017).

Oshan, Taylor M. and A. Stewart Fotheringham (in press). “A Comparison of Spatially Varying Regression Coefficient Estimates Using Geographically Weighted and Spatial-Filter-Based Techniques”. In: *Geographical Analysis*.

Páez, Antonio, Steven Farber, and David Wheeler (2011). “A Simulation-Based Study of Geographically Weighted Regression as a Method for Investigating Spatially Varying Relationships”. en. In: *Environment and Planning A* 43.12, pp. 2992–3010. ISSN: 0308-518X, 1472-3409. DOI: 10.1068/a44111. URL: http://epn.sagepub.com/content/43/12/2992 (visited on 10/28/2015).

Polson, Nicholas G, James G Scott, et al. (2012). “On the half-Cauchy prior for a global scale parameter”. In: *Bayesian Analysis* 7.4, pp. 887–902.

Simpson, Daniel P et al. (2014). “Penalising model component complexity: A principled, practical approach to constructing priors”. In: *arXiv preprint arXiv:1403.4630*.

Thomas, Andrew and Robert B O’Hara (2004). *OpenBUGS*.

Tobler, W. R. (1970). “A computer movie simulating urban growth in the Detroit region”. In: *Economic Geography* 46.2, pp. 234–240.

Waller, Lance A et al. (2007). “Quantifying geographic variations in associations between alcohol distribution and violence: a comparison of geographically weighted regression and spatially varying coefficient models”. In: *Stochastic Environmental Research and Risk Assessment* 21.5, pp. 573–588.

Wheeler, David C (2009). “Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso”. In: *Environment and planning A* 41.3, pp. 722–742.

Wheeler, David C and Catherine A Calder (2007). “An assessment of coefficient accuracy in linear regression models with spatially varying coefficients”. In: *Journal of Geographical Systems* 9.2, pp. 145–166.

Wheeler, David C and Lance A Waller (2009). “Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests”. In: *Journal of Geographical Systems* 11.1, pp. 1–22.