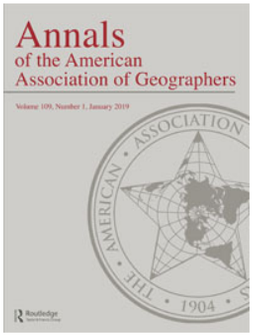

Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: https://www.tandfonline.com/loi/raag21

The Importance of Scale in Spatially Varying Coefficient Modeling

Daisuke Murakami, Binbin Lu, Paul Harris, Chris Brunsdon, Martin Charlton, Tomoki Nakaya & Daniel A. Griffith

To cite this article: Daisuke Murakami, Binbin Lu, Paul Harris, Chris Brunsdon, Martin Charlton, Tomoki Nakaya & Daniel A. Griffith (2019) The Importance of Scale in Spatially Varying Coefficient Modeling, Annals of the American Association of Geographers, 109:1, 50-70, DOI: 10.1080/24694452.2018.1462691

To link to this article: https://doi.org/10.1080/24694452.2018.1462691

The Importance of Scale in Spatially Varying Coefficient Modeling

Daisuke Murakami,[*] Binbin Lu,[†] Paul Harris,[‡] Chris Brunsdon,[§] Martin Charlton,[§] Tomoki Nakaya,[¶] and Daniel A. Griffith[#]

[*]*Department of Data Science, The Institute of Statistical Mathematics*
[†]*School of Remote Sensing and Information Engineering, Wuhan University*
[‡]*Sustainable Agricultural Sciences, Rothamsted Research*
[§]*National Centre for Geocomputation, Maynooth University*
[¶]*Graduate School of Environmental Studies, Tohoku University*
[#]*School of Economic, Political and Policy Sciences, The University of Texas at Dallas*

Although spatially varying coefficient (SVC) models have attracted considerable attention in applied science, they have been criticized as being unstable. The objective of this study is to show that capturing the "spatial scale" of each data relationship is crucially important to make SVC modeling more stable and, in doing so, adds flexibility. Here, the analytical properties of six SVC models are summarized in terms of their characterization of scale. Models are examined through a series of Monte Carlo simulation experiments to assess the extent to which spatial scale influences model stability and the accuracy of their SVC estimates. The following models are studied: (1) geographically weighted regression (GWR) with a fixed distance or (2) an adaptive distance bandwidth (GWRa); (3) flexible bandwidth GWR (FB-GWR) with fixed distance or (4) adaptive distance bandwidths (FB-GWRa); (5) eigenvector spatial filtering (ESF); and (6) random effects ESF (RE-ESF). Results reveal that the SVC models designed to capture scale dependencies in local relationships (FB-GWR, FB-GWRa, and RE-ESF) most accurately estimate the simulated SVCs, where RE-ESF is the most computationally efficient. Conversely, GWR and ESF, where SVC estimates are naïvely assumed to operate at the same spatial scale for each relationship, perform poorly. Results also confirm that the adaptive bandwidth GWR models (GWRa and FB-GWRa) are superior to their fixed bandwidth counterparts (GWR and FB-GWR). *Key Words: flexible bandwidth geographically weighted regression, Monte Carlo simulation, nonstationarity, random effects eigenvector spatial filtering, spatial scale.*

尽管空间变异係数（SVC）模型已吸引了应用科学的大量关注，但却仍被批评不够平稳。本研究的目标在于展现，捕捉各数据关系的"空间尺度"，是让SVC模式化更为平稳的重要关键，这麼做并可增加弹性。本研究于此摘要六大SVC模型在尺度特徵化上的分析属性。本研究通过一系列的蒙特卡罗模拟实验检视模型，以评估空间尺度影响模型稳定度的程度，及其SVC评估的精确度。本文研究下列模型：（1）距离固定下的地理加权迴归（GWR）；（2）自适应距离的带宽（GWRa）；（3）固定距离下的弹性带宽GWR（FB-GWR）；或（4）自适应距离带宽（FB-GWRa）；（5）特徵质空间过滤法（ESF）；以及（6）随机效应 ESF （RE-ESF）。研究结果揭露，设计用来捕捉地方关系中的尺度依赖之SVC模型（FB-GWR、FB-GWRa和RE-ESF），最正确地估计模拟的SVCs，而REESF则在计算上最有效率。反之，SVC估计被天真地假设在各关系中皆在相同的空间尺度上操作的GWR和ESF表现差劲。研究结果同时确认自适应带宽GWR模型（GWRa和FB-GWRa）较其固定带宽的对照组（GWR和FB-GWR）而言更为优越。*关键词：弹性带宽地理加权迴归，蒙特卡罗模拟，非平稳性，随机效应特徵质空间过滤法，空间尺度。*

Aunque los modelos de coeficiente espacialmente cambiante (SVC) han atraído mucha atención en ciencia aplicada, se les critica de ser inestables. El objetivo del presente estudio es mostrar que la captura de la "escala espacial" de cada relación de datos es crucialmente importante para hacer el modelado del SVC más estable y de ese modo agregarle flexibilidad. Se resumen en este artículo las propiedades analíticas de seis modelos de SVC en términos de su caracterización por escala. Los modelos son examinados a través de una serie de experimentos de simulación de Monte Carlo para evaluar el alcance con que la escala espacial influye sobre la estabilidad del modelo y en la exactitud de las estimaciones SVC. Se estudian los siguientes modelos: (1) regresión geográficamente ponderada (GWR) con una distancia fija, o (2) un ancho de banda de distancia adaptable (GWRa); (3) ancho de banda flexible GWR (FB-GWR) con distancia fija, o (4) anchuras de banda de distancia adaptable (FB-GWRa); (5) filtrado espacial eigenvector (ESF); y (6) efectos

Annals of the American Association of Geographers, 109(1) 2019, pp. 50–70
Initial submission, September 2017; revised submission, December 2017; final acceptance, February 2018
Published by Taylor & Francis, LLC.

aleatorios ESF (RE-ESF). Los resultados revelan que los modelos SVC diseñados para captar las dependencias de escala en las relaciones locales (FB-GWR, FB-GWRa, y RE-ESF) calculan con la mayor exactitud los SVC simulados, donde REESF es el de mayor eficiencia computacional. Por el contrario, GWR y ESF, donde ingenuamente se asume que las estimaciones operan a la misma escala espacial para cada relación, registran un desempeño pobre. Los resultados también confirman que los modelos GWR del ancho de banda adaptable (GWRa y FB-GWRa) son superiores a sus contrapartes de ancho de banda fija (GWR y FB-GWR). *Palabras clave: escala espacial, filtrado espacial eigenvector de efectos aleatorios, no-estacionalidad, regresión geográficamente ponderada con ancho de banda flexible, simulación Monte Carlo.*

Spatially varying coefficient (SVC) models are used to investigate nonstationarity in response to predictor data relationships in regression models. Provided that relationship heterogeneity exists, models output regression coefficient estimates that vary across space. Estimated SVCs can be mapped along with associated inference diagnostics and thus provide a deeper understanding of a study's spatial relationships. As with spatial autocorrelation, relationship spatial heterogeneity is a common property of many geographical processes (see Goodchild 2004; Anselin 2010), although differentiating one effect from the other can be difficult (e.g., Harris et al. 2017). Various approaches have been developed for SVC regression modeling, the most notable of which include (1) the spatial expansion method (Casetti 1972; Jones and Casetti 1992), (2) geographically weighted regression (GWR; Brunsdon, Fotheringham, and Charlton 1996, 1998; Fotheringham, Brunsdon, and Charlton 2002), (3) Bayesian SVC models (Besag, York, and Mollie 1991; Assunção 2003; Gamerman, Moreira, and Rue 2003; Gelfand et al. 2003; Wheeler and Calder 2007; Wheeler and Waller 2009), and (4) eigenvector spatial filtering (ESF)-based approaches (Griffith 2003, 2008; Murakami et al. 2017).

Among them, GWR has proven the most popular, including case studies in hedonic house price modeling (e.g., Bitter, Mulligan, and Dall'erba 2007; Páez, Long, and Farber 2008; Lu, Charlton, et al. 2014), environmental analysis (e.g., Brunsdon, McClatchey, and Unwin 2001; Harris, Fotheringham, and Juggins 2010; Jaimes et al. 2010), and disease mapping (e.g., Nakaya et al. 2005; Hu et al. 2012; Ndiath et al. 2015). Much of this popularity stems from its relative simplicity and readily available software like GWR4 (Nakaya 2015; see http://gwr.maynoothuniversity.ie/gwr4-software/) and the GWmodel R package (Lu, Harris, et al. 2014; Gollini et al. 2015). Despite the widespread uptake of basic GWR, it suffers from (at least) two severe limitations: (1) instability where local predictor variable collinearity can create spurious nonstationarities (Wheeler and Tiefelsdorf 2005; Páez, Farber, and Wheeler 2011) and (2) inflexibility where basic GWR assumes the same scale of spatial variation across each set of estimated SVCs (Brunsdon, Fotheringham, and Charlton 1999).

Relating to the first model limitation, demonstrations exist showing that SVCs estimated from GWR can be collinear with each other, detect unrealistically smooth map patterns, or take extreme values. Various collinearity diagnostics can be calculated to provide a better understanding of potential problems (Wheeler and Tiefelsdorf 2005; Wheeler 2007; Gollini et al. 2015), together with the implementation of some regularized GWR models (Wheeler 2007, 2009; Gollini et al. 2015) or GWR via an empirical Bayes approach (Bárcena et al. 2014)—all specifically designed to address collinearity. GWR purportedly is fairly robust to local collinear effects (Fotheringham and Oshan 2016), but on balance, evidence suggests otherwise (see, e.g., Harris et al. 2017). Observe that instability in estimated SVCs from GWR might arise for reasons other than collinearity, including the existence of outliers (Farber and Páez 2007; Harris, Fotheringham, and Juggins 2010) and the presence of spatial autocorrelation (Cho, Lambert, and Chen 2010).

For the second model limitation, basic GWR uses a single kernel bandwidth for its calibration, which is somewhat flawed in that it implicitly assumes the same degree of spatial smoothness for each set of estimated SVCs, which is unrealistic. Thus, when some relationships tend to operate at a larger scale whereas other relationships operate at a smaller scale, basic GWR will nullify these differences and only find a "best-on-average" scale of relationship nonstationarity (as using only a single bandwidth). To address this limitation, mixed (semiparametric) GWR can be implemented in which some relationships are assumed stationary (globally fixed) and others are assumed nonstationary (locally varying; Brunsdon, Fotheringham, and Charlton 1999; Fotheringham, Brunsdon, and Charlton 2002; Nakaya et al. 2005; Mei, Wang, and

Zhang 2006; Mei, Xu, and Wang 2016). A mixed GWR model only in part addresses this limitation, however, because the subset of locally varying relationships is still assumed to operate at the same spatial scale. Instead, flexible bandwidth GWR (FB-GWR) can be used, in which each relationship is specified using its own bandwidth and thus provides a true multiscale GWR model, where the scale of relationship nonstationarity may vary for each response to predictor variable relationship.

The development of FB-GWR follows that of Yang, Fotheringham, and Harris (2011, 2012), Yang (2014), Lu et al. (2017), Leong and Yue (2017; who renamed it conditional GWR), and Fotheringham, Yang, and Kang (2017; who renamed it multiscale GWR), all of whom implement the idea of a vector of bandwidths for GWR, as first set out by Brunsdon, Fotheringham, and Charlton (1999). The study of Lu et al. (2017) provides an extension of FB-GWR, where each relationship also can be specified with its own distance metric, as well as its own bandwidth. In this study, we implement the model of Lu et al. (2017) but where it is specified using only Euclidean distances, thus directly providing a FB-GWR model.

A Bayesian SVC model (specifically, the Gaussian process-based approach of Gelfand et al. 2003) can be viewed as a regularized alternative to GWR (i.e., it can address collinearity) and also is directly able to identify the spatial scale of each relationship through its specification of priors. Thus, both model limitations stated for GWR are implicitly addressed. Although increased estimation accuracy for the Bayesian SVC approach has been reported (see Wheeler and Calder 2007; Wheeler and Waller 2009), this improvement is computationally expensive, especially when scale is estimated for every set of SVCs (Finley 2011).

Unlike GWR, the ESF-based approach allows control of the number of parameters (i.e., model complexity) through variable selection. Helbich and Griffith (2016), Murakami et al. (2017), and Oshan and Fotheringham (2017), however, demonstrated instability of the ESF-based approach in cases where it can suffer just as basic GWR does with respect to Limitations 1 and 2. In this respect, Murakami et al. (2017) proposed an extended ESF-based approach that directly addresses these limitations (i.e., it is robust to local collinearity and allows for the possibility of each set of SVCs having a different degree of spatial smoothness). Furthermore, this random effects ESF model (RE-ESF) is shown to be computationally efficient, thus providing a practical alternative to the Bayesian SVC model that often is computationally intractable.

The study of Murakami et al. (2017), employing a Monte Carlo simulation experiment similar in design to that used here, not only demonstrated the advantage of the RE-ESF model over the Bayesian SVC model but also demonstrated its advantages over both basic and regularized GWR forms (following Gollini et al. 2015). The latter was not surprising given that neither single-bandwidth GWR form deals with the second model limitation. Hence, this study addresses this important gap by introducing FB-GWR to the same model comparison exercise. Because repeating all model comparisons of Murakami et al. (2017) is unnecessary, only (1) GWR with fixed distance or (2) adaptive distance bandwidth (GWRa), (3) FB-GWR with fixed distance or (4) adaptive distance bandwidths (FB-GWRa), (5) ESF, and (6) RE-ESF models are compared here. Thus, a Bayesian SVC model is not included here, but this study coupled with that by Murakami et al. (2017) provides a comprehensive comparison of major multiscale SVC models (i.e., FB-GWR and RE-ESF).

In summary, the aim of this study is to continue to demonstrate the importance of spatial scale in SVC models through FB-GWR and RE-ESF, the output of which should be more stable and flexible in comparison to their basic counterparts (GWR and ESF, respectively). Through this demonstration, this study simultaneously quantifies the influence from Limitation 1 on the scale of spatial variations in the estimated SVCs and from Limitation 2 on the scale of spatial variations in predictor variables. Although recent studies (e.g., Murakami et al. 2017; Oshan and Fotheringham 2017) have investigated the influence from the first limitation on the SVC estimates for various SVC models, they have ignored the second limitation. Because the balance between the two influences estimation accuracy of the coefficients (in the case of a stationary regression model; see Paciorek 2009), the robustness of an SVC model must be examined considering both spatial scales 1 and 2. Besides, this study identifies the source of instability in GWR and ESF models. The source is unclear especially considering that Fotheringham and Oshan (2016) showed a certain robustness of GWR to collinearity. We show that scales 1 and 2 are the trigger of the instability, by using an exhaustive simulation experiment.

The remaining sections are organized as follows. The next section outlines the GWR- and ESF-based

models. We then perform a Monte Carlo simulation experiment to quantify the impact of spatial scale on model stability. After that, we summarize a second Monte Carlo simulation experiment to evaluate the impact of spatial scale on SVC estimates and then we provide a concluding discussion. Study GWR and FB-GWR models are fitted using GWmodel (Version 2.0.4; see https://cran.r-project.org/package=GWmodel), the RE-ESF model is fitted using the R package spmoran (Version 0.1.2; Murakami 2017; see https://cran.r-project.org/web/packages/spmoran/index.html), and the ESF model is fitted by newly written R code.

Spatially Varying Coefficient Modeling

The Overarching SVC Model

A linear SVC model is formulated as follows:

$$y_i = \sum_{k=1}^{K} x_{i,k}\beta_k(s_i) + \varepsilon_i, \quad E[\varepsilon_i] = 0, \quad Var[\varepsilon_i] = \sigma^2, \tag{1}$$

where y_i represents the response variable at the ith sample site, with $i \in \{1, ..., N\}$, $x_{i,k}$ represents the kth predictor variable, with $k \in \{1, ..., K\}$, ε_i represents the disturbance, and σ^2 represents a variance parameter. $\beta_k(s_i)$ denotes the kth SVC for site i. *Local* and *global* approaches exist to estimate Equation 1, as detailed later, where, in general, a global approach to nonstationary modeling is preferred because it is more statistically coherent (e.g., Sampson, Damian, and Guttorp 2001).

Local Estimation (GWR and FB-GWR)

A local approach estimates coefficients at the ith site, $\{\beta_1(s_i), \ldots, \beta_k(s_i), \ldots, \beta_K(s_i)\}$, using only neighboring subsamples. Moving window regression (MWR; see Lloyd 2010) applies ordinary least squares estimation to neighboring subsamples at site i, whereas GWR applies weighted least squares estimation to neighboring subsamples that are weighted via a distance-decay scheme at site i. MWR is a special case of GWR when a boxcar kernel weighting scheme is specified (weights equal unity within the kernel and zero otherwise). Distance-decay weighting provides added flexibility to local regression modeling, allowing more data to have an influence locally, and tends to yield more smoothly varying coefficient surfaces. Suppose that $\boldsymbol{\beta}(s_i) = [\beta_1(s_i), \ldots, \beta_k(s_i), \ldots, \beta_K(s_i)]'$, where the prime represents matrix transpose; then the GWR estimator yields

$$\hat{\boldsymbol{\beta}}(s_i) = \left[\mathbf{X}'\mathbf{G}(s_i)\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{G}(s_i)\mathbf{y}, \tag{2}$$

where $\mathbf{X}$ is an $N \times K$ matrix of predictor variables, $\mathbf{y}$ is an $N \times 1$ vector of continuous response variables, and $\mathbf{G}(s_i)$ is an $N \times N$ diagonal matrix whose jth element $g(s_i, s_j)$ represents the weight assigned to the jth sample. Here, $g(s_i, s_j)$ is calculated with some kernel weighting function (see Gollini et al. 2015). For instance, the exponential kernel is defined as follows:

$$g(s_i, s_j) = \exp\left(-\frac{d(s_i, s_j)}{b}\right), \tag{3}$$

where $d(s_i, s_j)$ is the distance between locations s_i and s_j, and b denotes the bandwidth parameter. The resultant SVCs tend to be the global coefficients of a standard regression, if the bandwidth parameter, b, is set sufficiently large enough; otherwise, the SVCs are local. Here the bandwidth can be specified as a fixed distance, but for irregular sample configurations, the kernel window tends to include too few samples in sparsely sampled areas and too many samples in densely sampled areas. To counter this, an adaptive distance bandwidth can be specified, where the bandwidth varies according to a fixed local density of subsamples. An adaptive exponential kernel is defined as follows:

$$g^{ad}(s_i, s_j) = \exp\left(-\frac{d(s_i, s_j)}{b(s_i)^{ad}}\right), \tag{4}$$

where $b(s_i)^{ad}$ is the adaptive bandwidth for the ith site and is given by the distance between the ith site and the jth nearest neighbor. Note that this kernel is only adaptive to local sample density (and is used with GWRa); it is not adaptive in a predictor-specific sense. FB-GWR, described next, is one way to specify such a kernel.

Standard GWR, as described earlier, ignores differences of spatial scale across the SVCs, because the same (single, fixed or adaptive) bandwidth is specified for all data relationships. To counter this, each set of SVCs can be found using its own bandwidth, providing an extension of GWR with multiple bandwidths, one for each relationship (i.e., FB-GWR). Here the fixed bandwidth, an exponential kernel for FB-GWR, is defined as

$$g_k(s_i, s_j) = \exp\left(-\frac{d(s_i, s_j)}{b_k}\right), \tag{5}$$

where b_k is the fixed bandwidth for the kth parameter. The kth coefficient estimates might have global-scale spatial variations if b_k is set sufficiently large and local-scale spatial variations if b_k is set sufficiently small. The corresponding adaptive bandwidth version for FB-GWR (i.e., for FB-GWRa) is defined as

$$g_k^{ad}(s_i, s_j) = \exp\left(-\frac{d(s_i, s_j)}{b(s_i)_k^{ad}}\right), \tag{6}$$

where $b(s_i)_k{}^{ad}$ is the kth adaptive bandwidth.

Standard GWR is estimated as follows: (1) the bandwidth parameter is calibrated by minimizing the mean squared error (MSE) applying a leave-one-out cross-validation (CV) procedure (Brunsdon, Fotheringham, and Charlton 1996) and (2) the SVCs are estimated by substituting the calibrated bandwidth into Equation 2. FB-GWR is estimated in a similar fashion, except that in Step 1 a back-fitting approach is adopted (for details, see Lu et al. 2017), which sequentially iterates the calibration of b_k (or $b(s_i)_k{}^{ad}$) assuming that all bandwidth parameters are known (see also Yang 2014). The MSE minimization in Step 1 for GWR or FB-GWR can be replaced with maximization of the corrected Akaike information criterion (AICc) or some other information criterion. Observe that because $N \times K$ coefficients are estimated using only N samples, it is necessary to enhance model accuracy while avoiding overfitting. A CV or AICc approach is reasonable because it minimizes the generalization error (see Bishop 2006). In this study, the AICc approach is chosen for all GWR and FB-GWR fits and, as detailed earlier, only bandwidths corresponding to exponential kernels are found.

Global Estimation (ESF and RE-ESF)

This global approach estimates the SVCs by fitting spatial process models. The spatial expansion and ESF-based approaches are representative of such methods, where the former fits trend surface models, whereas the latter fits ESF models describing spatially structured SVC map patterns. The ESF-based approach is built on the Moran coefficient (MC; see Cliff and Ord 1973),[1] which is a diagnostic statistic for spatial dependence. The MC is formulated as follows:

$$\mathrm{MC}[\mathbf{y}] = \frac{N}{\mathbf{1}'\mathbf{C}\mathbf{1}}\frac{\mathbf{y}'\mathbf{M}\mathbf{C}\mathbf{M}\mathbf{y}}{\mathbf{y}'\mathbf{M}\mathbf{y}}, \tag{7}$$

where $\mathbf{1}$ is an $N \times 1$ vector of ones, $\mathbf{C}$ is an $N \times N$ connectivity matrix whose diagonal elements are zero, and $\mathbf{M} = \mathbf{I} - \mathbf{1}\mathbf{1}'/N$ is an $N \times N$ centering matrix. The MC is greater than $-1/(N-1) \approx 0$, which is the expectation of the MC in the absence of spatial dependence; is greater than $-1/(N-1)$ if the samples are positively spatially dependent; and is less than $-1/(N-1)$ if they are negatively dependent. Let us eigen-decompose the matrix $\mathbf{MCM}$ to $\mathbf{E}_{full}\mathbf{\Lambda}_{full}\mathbf{E}_{full}'$, where $\mathbf{E}_{full}$ is an $N \times N$ matrix with its lth column being the lth eigenvector $\mathbf{e}_l$, and $\mathbf{\Lambda}_{full}$ is an $N \times N$ diagonal matrix whose lth element is the l—the eigenvalue, λ_l. The eigenvectors have the following feature:

$$\begin{aligned}\mathrm{MC}[\mathbf{e}_l] &= \frac{N}{\mathbf{1}'\mathbf{C}\mathbf{1}}\frac{\mathbf{e}_l'\mathbf{M}\mathbf{C}\mathbf{M}\mathbf{e}_l}{\mathbf{e}_l'\mathbf{M}\mathbf{e}_l} \\ &= \frac{N}{\mathbf{1}'\mathbf{C}\mathbf{1}}\frac{\mathbf{e}_l'\mathbf{E}_{full}\mathbf{\Lambda}_{full}\mathbf{E}_{full}'\mathbf{e}_l}{\mathbf{e}_l'\mathbf{e}_l}, \\ &= \frac{N}{\mathbf{1}'\mathbf{C}\mathbf{1}}\lambda_l.\end{aligned} \tag{8}$$

Here Equation 8 suggests that the eigenvectors corresponding to positive eigenvalues are orthogonal basis functions describing positive spatial dependence, with each magnitude being indexed by its corresponding eigenvalue. Likewise, eigenvectors corresponding to negative eigenvalues explain negative spatial dependence. For details about Moran eigenvectors, see Griffith (2003).

The ESF-based SVC model of Griffith (2008) is formulated as

$$\begin{aligned}\mathbf{y} &= \sum_{k=1}^{K}\mathbf{x}_k \circ \boldsymbol{\beta}_k^{ESF} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \\ \boldsymbol{\beta}_k^{ESF} &= \beta_k\mathbf{1} + \mathbf{E}_k\boldsymbol{\gamma}_k,\end{aligned} \tag{9}$$

where $\mathbf{x}_k$ is an $N \times 1$ vector of the kth predictor variable (i.e., the kth column of matrix $\mathbf{X}$), $\mathbf{E}_k$ is an $N \times L_k$ matrix composed of L_k eigenvectors ($L_k < N$), $\boldsymbol{\gamma}_k$ is an $L_k \times 1$ coefficient vector, and "$\circ$" denotes the element-wise (Hadamard) product operator. Here $\boldsymbol{\beta}_k^{ESF} = \beta_k\mathbf{1} + \mathbf{E}_k\boldsymbol{\gamma}_k$ yields a vector of SVCs in which $\beta_k\mathbf{1}$ and $\mathbf{E}_k\boldsymbol{\gamma}_k$ represent the constant component and the spatially varying component, respectively. Equation 9 becomes the linear regression model with predictor variables $\{\mathbf{X}, \mathbf{x}_1 \circ \mathbf{E}_1, \ldots, \mathbf{x}_K \circ \mathbf{E}_K\}$ once the second equation is substituted into the first equation. Accordingly, the adjusted R^2 and other diagnostic statistics for the linear regression model are applicable for Equation 9.

The parameters of this model are estimated through the following steps: (1) eigenvectors, which are not of interest, are removed a priori from $\{\mathbf{E}_1, \ldots, \mathbf{E}_K\}$ (see later); (2) significant predictor variables are selected among $\{\mathbf{X}, \mathbf{x}_1 \circ \mathbf{E}_1, \ldots, \mathbf{x}_K \circ \mathbf{E}_K\}$ by

applying a forward variable selection technique; (3) $\{\beta_1, \ldots, \beta_K, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_K\}$ are estimated using the model after the variable selection; and (4) $\hat{\boldsymbol{\beta}}_k^{ESF} = \hat{\beta}_k \mathbf{1} + \mathbf{E}_k \hat{\boldsymbol{\gamma}}_k$ is calculated. In our analysis, $\mathbf{E}_k$ is defined by the eigenvectors corresponding to positive eigenvalues in Step 1 (see Murakami et al. 2017). Thus, all eigenvectors describing positive spatial dependence are taken into account. The adjusted R^2 is maximized in the variable selection Step 2.

The ESF-based approach, which estimates *deterministic* map patterns, has been extended to a random effects ESF-based approach (RE-ESF; Murakami and Griffith 2015), which models stochastic spatial processes. The RE-ESF-based SVC model (Murakami et al. 2017) is formulated as follows:

$$\begin{aligned} \mathbf{y} &= \sum_{k=1}^{K} \mathbf{x}_k \circ \boldsymbol{\beta}_k^{RE-ESF} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \\ \boldsymbol{\beta}_k^{RE-ESF} &= \beta_k \mathbf{1} + \mathbf{E}_k \boldsymbol{\gamma}_k, \quad \boldsymbol{\gamma}_k \sim N\left(\mathbf{0}_L, \sigma_{\gamma,k}^2 \boldsymbol{\Lambda}(\alpha_k)\right), \end{aligned} \tag{10}$$

where $\mathbf{0}_L$ is an $L \times 1$ vector of zeros, $\mathbf{E}$ is a matrix of L eigenvectors corresponding to positive eigenvalues, ${\sigma_{\gamma,k}}^2$ is a variance parameter, and $\boldsymbol{\Lambda}(\alpha_k)$ is an $L \times L$ diagonal matrix whose lth element is $\lambda_l(\alpha_k) = (\sum_l \lambda_l / \sum_l \lambda_l^{\alpha_k}) \lambda_l^{\alpha_k}$, where α_k is the key parameter. When α_k is large, coefficient estimates of the nonprincipal eigenvectors are strongly shrunk toward 0, and the kth SVCs, ${\boldsymbol{\beta}_k}^{RE\text{-}ESF}$, provide a large-scale spatial pattern. By contrast, ${\boldsymbol{\beta}_k}^{RE\text{-}ESF}$ has a small-scale spatial pattern when α_k is small. Thus, α_k is a scale parameter for the SVCs, and its effects for RE-ESF are analogous to the multiple bandwidths of FB-GWR.

Furthermore, Equation 10 has the following expression:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})\widetilde{\mathbf{u}} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \\ \widetilde{\mathbf{E}} &= \begin{bmatrix} \mathbf{x}_1 \circ \mathbf{E} & \ldots & \mathbf{x}_K \circ \mathbf{E} \end{bmatrix}, \\ \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) &= \begin{bmatrix} \sigma_{\gamma,1}^2 \boldsymbol{\Lambda}(\alpha_1) & & \\ & \ddots & \\ & & \sigma_{\gamma,K}^2 \boldsymbol{\Lambda}(\alpha_K) \end{bmatrix}, \\ \widetilde{\mathbf{u}} &= \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_K \end{bmatrix}. \end{aligned} \tag{11}$$

$\boldsymbol{\theta} \in \{\alpha_1, \ldots, \alpha_K, \sigma_{\gamma,1}^2, \ldots, \sigma_{\gamma,K}^2\}$, and $\mathbf{u}_k \sim N(\mathbf{0}_L, \sigma^2 \mathbf{I}_L)$, where $\mathbf{I}_L$ is an $L \times L$ identity matrix. Note that $\boldsymbol{\gamma}_k = {\sigma_{k(\gamma)}}^2 \boldsymbol{\Lambda}(\alpha_k)\mathbf{u}_k$, where Equation 11 suggests that the RE-ESF model is a linear mixed effects model. Furthermore, $\boldsymbol{\beta}$ and $\widetilde{\mathbf{u}}$ have the following best linear unbiased estimators:

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\widetilde{\mathbf{u}}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) \\ \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})\widetilde{\mathbf{E}}'\mathbf{X} & \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})\widetilde{\mathbf{E}}'\widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) + \mathbf{I}_{KL} \end{bmatrix}^{-1} \times \begin{bmatrix} \mathbf{X}\mathbf{y}' \\ \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})\widetilde{\mathbf{E}}\mathbf{y}' \end{bmatrix}, \tag{12}$$

where $\boldsymbol{\theta}$ is estimated by numerically maximizing the following Type II restricted likelihood (empirical Bayes/*h*-likelihood) function:

$$\begin{aligned} \log lik_R(\boldsymbol{\theta}) = &-\frac{1}{2}\log \begin{vmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) \\ \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})\widetilde{\mathbf{E}}'\mathbf{X} & \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})\widetilde{\mathbf{E}}'\widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) + \mathbf{I}_{KL} \end{vmatrix} \\ &-\frac{N-K}{2}\left[1 + \log\left(\frac{2\pi}{N-K}\left(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} + \hat{\widetilde{\mathbf{u}}}'\hat{\widetilde{\mathbf{u}}}\right)\right)\right], \end{aligned} \tag{13}$$

where $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})\hat{\widetilde{\mathbf{u}}}$. Given $\mathbf{X}'\mathbf{X}$, $\mathbf{X}'\widetilde{\mathbf{E}}$, and $\widetilde{\mathbf{E}}'\widetilde{\mathbf{E}}$, the computational complexity of Equation 13 is $O((K+KL)^3)$, which is independent of N. This ensures that, once these matrix products are evaluated a priori, the numerical optimization of $\boldsymbol{\theta}$ is fast, even for large samples.

Degrees of Freedom for the SVC Models

This section defines the degrees of freedom (*df*) for the SVC models, which is a measure of model stability. For a linear model, *df* is defined by $N - tr[\mathbf{H}]$, where $tr[\cdot]$ is the trace operator and $\mathbf{H}$ is the hat matrix such that $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$. For instance, *df* for the standard linear regression model is $df_{LM} = N - tr[\mathbf{H}_{LM}]$, which equals the sample size minus the number of regression coefficients, where $\mathbf{H}_{LM} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. A large *df* is desirable to avoid overfitting.

Table 1 summarizes *df* for the study SVC models. Here, $df_{GWR} = N - tr[\mathbf{H}_{GWR}]$ deflates when $\mathbf{X}'\mathbf{G}(s_i)\mathbf{X}$ is nearly singular. Singularity happens when the bandwidth is small and most elements of $\mathbf{G}(s_i)$ take near-zero values. In other words, small bandwidths introduce overfitting. The problem is serious if subsamples are sparsely distributed around the site s_i. Thus, GWR specified with an adaptive bandwidth (GWRa), which changes the kernel window size in

Table 1. The hat matrix, **H**, of the study spatially varying coefficient models, where $df = N - tr[\mathbf{H}]$

Model	Hat matrix, **H**	Parameters in **H**
GWR, GWRa	$\mathbf{H}_{GWR}$ = a matrix with its ith row being $\mathbf{x}(s_i)'[\mathbf{X}'\mathbf{G}(s_i)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{G}(s_i)$	b[a]
FB-GWR, FB-GWRa	Single hat matrix is not available (Fotheringham, Yang, and Kang 2017). The following hat matrix for the kth SVC appears in each iteration of the backfitting: $\mathbf{x}(s_i)'\ [\mathbf{x}_k'\ \mathbf{G}_k(s_i)\mathbf{x}_k]^{-1}\mathbf{x}_k'\ \mathbf{G}_k(s_i)$, where $\mathbf{G}_k(s_i)$ equals $\mathbf{G}(s_i)$ whose b is replaced with b_k.	$b_1, \ldots, b_K$[b]
ESF	$\mathbf{H}_{ESF} = [\mathbf{X} \quad \widetilde{\mathbf{E}}_{ESF}] \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\widetilde{\mathbf{E}}_{ESF} \\ \widetilde{\mathbf{E}}'_{ESF}\mathbf{X} & \widetilde{\mathbf{E}}'_{ESF}\widetilde{\mathbf{E}}_{ESF} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}' \\ \widetilde{\mathbf{E}}'_{ESF} \end{bmatrix}$, where $\widetilde{\mathbf{E}}_{ESF} = [\mathbf{x}_1 \circ \mathbf{E}_1 \quad \ldots \quad \mathbf{x}_K \circ \mathbf{E}_K]$	Selection of eigenvectors
RE-ESF	$\mathbf{H}_{RE-ESF} = [\mathbf{X} \quad \widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})] \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) \\ \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})\widetilde{\mathbf{E}}'\mathbf{X} & \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})\widetilde{\mathbf{E}}'\widetilde{\mathbf{E}}\widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta}) + \mathbf{I}_{KL} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}' \\ \widetilde{\boldsymbol{\Lambda}}(\boldsymbol{\theta})\widetilde{\mathbf{E}}' \end{bmatrix}$	$\boldsymbol{\theta} \in \{\alpha_1, \ldots, \alpha_K, \sigma^2{}_{1,\gamma}, \ldots, \sigma^2{}_{K,\gamma}\}$

Note: GWR = geographically weighted regression; GWRa = geographically weighted regression with adaptive distance bandwidths; FB-GWR = flexible bandwidth geographically weighted regression; FB-GRWa = flexible bandwidth geographically weighted regression with adaptive distance bandwidths; ESF = eigenvector spatial filtering; RE-ESF = random effects eigenvector spatial filtering.
[a]In case of GWRa, b is replaced with $b(s_i)^{ad}$.
[b]In case of FB-GWRa, $\{b_1, \ldots, b_K\}$ is replaced with $\{b(s_i)^{ad}{}_1, \ldots, b(s_i)_K{}^{ad}\}$.

accordance with sample density, would be an effective tool to mitigate this problem, where df_{GWRa} is likely to be larger than df_{GWR} in many cases, even though they are both given by $N - tr[\mathbf{H}_{GWR}]$ with different kernel specifications. In GWR specifications, $\mathbf{X}'\mathbf{G}(s_i)\mathbf{X}$ approaches $\mathbf{X}'\mathbf{X}$ as their bandwidth increases and their df increase up to $N - K$. Thus, GWR approaches are stable if the underlying associations ($\boldsymbol{\beta}_k$s) have large-scale spatial variations or no variations (i.e., are constant).

For GWR, the singularity of $\mathbf{X}'\mathbf{G}(s_i)\mathbf{X}$ also changes, depending on the spatial scale of the predictor variables. If $\mathbf{x}_k$ suggests small-scale spatial variations, $\mathbf{x}_k$ would have considerable variation within each kernel window. By contrast, if $\mathbf{x}_k$ suggests large-scale spatial variations, the variation of the predictors within each kernel window can be small. In the extreme case, if $\mathbf{x}_k$ has uniform values across the window around site s_i, it is exactly collinear with the intercept term within the window (i.e., suppose that $\mathbf{x}_1$ represents an intercept, the entries of the kth row and column of $\mathbf{X}'\mathbf{G}(s_i)\mathbf{X}$ take exactly the same values with the entries of the first row and column; the resulting $\mathbf{X}'\mathbf{G}(s_i)\mathbf{X}$ becomes singular).[2] The FB-GWR model, which calibrates the bandwidths implicitly considering the scale of each $\mathbf{x}_k$, is valuable not only to control the varying scales of the SVCs but also to stabilize the SVC estimates (e.g., in the presence of collinearity).

Regarding ESF, forward eigenvector selection implicitly identifies the model that maximizes accuracy, where $df_{ESF} = N - tr[\mathbf{H}_{ESF}]$. Given that Moran eigenvectors describe coefficient patterns at different spatial scales, eigenvector selection identifies the scale of spatial variation in each SVC set. The df_{ESF} decreases as the number of selected eigenvectors increases; it happens if the response and predictor variables have spatially varying associations at every scale.

Unlike all of the preceding models, the df for the RE-ESF model, $df_{RE\text{-}ESF} = N - tr[\mathbf{H}_{RE\text{-}ESF}]$, includes not just the scale parameters $\{\alpha_1, \ldots, \alpha_K\}$ but also the variance parameters $\{\sigma^2{}_{1,\gamma}, \ldots, \sigma^2{}_{K,\gamma}\}$. Different from GWR, FB-GWR, and ESF models whose df always decrease in the presence of small-scale spatially varying associations (i.e., $\boldsymbol{\beta}_k$s), the RE-ESF model is capable of maintaining a high $df_{RE\text{-}ESF}$ value by decreasing the variance parameters even if small-scale associations exist.

In short, both the scale of associations (i.e., $\boldsymbol{\beta}_k$s) and the scale of predictors influence model df or model stability. The section Simulation 1: Scales of x_k and β_k versus Model Stability analyzes how these two scales influence model df employing a simulation experiment. Section 4 then quantifies the influences of these two scales on SVC estimation accuracy through a second simulation experiment.

Simulation 1: Scales of x_k and $\boldsymbol{\beta}_k$ versus Model Stability

Overview

This section objectively evaluates model complexity with df values, while varying the predictor

variables and the scale parameters for the SVCs, and tests for which cases the SVC models are unstable (i.e., it investigates the first earlier model limitation). For simplicity, we evaluate only cases where the spatial scale of variation for each set of SVCs is the same. In other words, regression relationships are set to vary from small scales to large scales but always in the same fashion for each regression relationship in the model. Thus, FB-GWR is not analyzed here, because it simply defaults to standard GWR in this instance.

For the Monte Carlo simulation, we assume an SVC model, $\mathbf{y} = \boldsymbol{\beta}_0 + \mathbf{x}_1 \circ \boldsymbol{\beta}_1 + \mathbf{x}_2 \circ \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, where the predictor variables, $\mathbf{x}_k$s, are generated from[3]

$$\mathbf{x}_k = (1 - r_x)\boldsymbol{\varepsilon}_{x(ns)} + r_x \mathbf{C}(b_x)\boldsymbol{\varepsilon}_{x(s)}, \qquad (14)$$

where $\boldsymbol{\varepsilon}_{x(ns)} \sim N(\mathbf{0}, \mathbf{I})$ and $\boldsymbol{\varepsilon}_{x(s)} \sim N(\mathbf{0}, \mathbf{I})$. Here $\mathbf{C}(b_x)$ is a matrix that row-standardizes a symmetric spatial proximity matrix whose (i, j)th element equals $\exp(-d(s_i, s_j)/b_x)$ and where $d(s_i, s_j)$ is the Euclidean distance between sample sites s_i and s_j. Spatial coordinates of the sample sites also are allowed to vary and are generated from standard normal distributions. Thus, $\mathbf{C}(b_x)\boldsymbol{\varepsilon}_k$ is a spatial moving average (SMA) process, and r_x is the ratio of spatially dependent variation to total variation in $\mathbf{x}_k$. The bandwidth b_x determines the kernel radius in the SMA process. A small b_x assigns much greater weights to neighborhoods, and the resulting SMA process describes small-scale spatial variations. Likewise, a large b_x describes large-scale spatial variations.

The *df* values for GWR, GWRa, ESF, and RE-ESF are evaluated while varying the parameters for the predictor variables (see Table 2) and those for the SVCs (see Table 3). In each case, the *df* values for each model are evaluated 200 times. For ESF, $[\mathbf{x}_1 \circ \mathbf{E}_1, \ldots, \mathbf{x}_K \circ \mathbf{E}_K]$, where $\mathbf{E}_k$ consists of eigenvectors corresponding to positive eigenvalues, are candidates for variable selection. These eigenvectors also are used in RE-ESF (i.e., $\mathbf{E} = \mathbf{E}_k$). The ratio of the selected eigenvectors in ESF is as described in Table 3.

Table 2. Parameter settings for predictor variables: $\mathbf{x}_k$

Parameter	Notation	Case
Sample size	N	400
Bandwidth	b_x	{0.0, 0.2, 0.6, 1.0}
Ratio of spatial variation	r_x	{0.2, 0.6, 1.0, 2.0}

This section does not discuss estimating SVCs through model fitting but rather discusses calculating *df* by simply substituting known scale parameters b_x for the predictors and the same parameter b for SVCs into $df = N - tr[\mathbf{H}]$ (see Table 1). SVC estimation accuracy is discussed in Simulation 2.

Results

Figure 1 plots the mean estimated *df* values for the SVC models arising from this first Monte Carlo simulation experiment. Here the mean df_{GWR} results suggest that GWR (with a fixed bandwidth) is unstable when $\mathbf{y}$ and $\mathbf{x}_k$s have small-scale spatially varying associations ($b = 0.2$).

Unlike the mean df_{GWR} results, the mean df_{GWRa} results for GWRa (with its adaptive bandwidth) are always relatively large across all values of b. At least from this result, GWRa seems relatively stable compared to GWR. This finding is not surprising, given that numerous empirical studies have suggested as much (e.g., Harris, Fotheringham, and Juggins 2010). Only for highly regular sample configurations is fixed bandwidth GWR usually recommended.

In Figure 1, the mean $df_{RE\text{-}ESF}$ results are evaluated for cases with $\sigma_k = 0.1$ and $\sigma_k = 1.0$, respectively. On the one hand, when $\sigma_k = 1.0$, which implies weaker shrinkage, $df_{RE\text{-}ESF}$ takes small values. On the other hand, $df_{RE\text{-}ESF}$ values are large across cases when stronger shrinkage is imposed by $\sigma_k = 0.1$. Thus, the RE-ESF estimates are relatively stable even when the SVCs have local variation but with a proviso that the σ_k parameter

Table 3. Parameter settings for spatially varying coefficients (SVCs): $\boldsymbol{\beta}_k$

Model	Parameter	Notation	Case
GWR	Bandwidth	b	{0.2, 0.6, 1.0, 2.0}
GWRa	Adaptive bandwidth	$b(s_i)^{ad}$	{0.1, 0.3, 0.5, 1.0}
ESF	Ratio of predictor variables being selected	q	{0.2, 0.4, 0.6, 0.8}
RE-ESF	Scale	α_k	{0.2, 0.6, 1.0, 2.0}
	Variance	σ_k	{0.1, 1.0}

Note: GWR = geographically weighted regression; GWRa = geographically weighted regression with adaptive distance bandwidths; ESF = eigenvector spatial filtering; RE-ESF = random effects eigenvector spatial filtering.

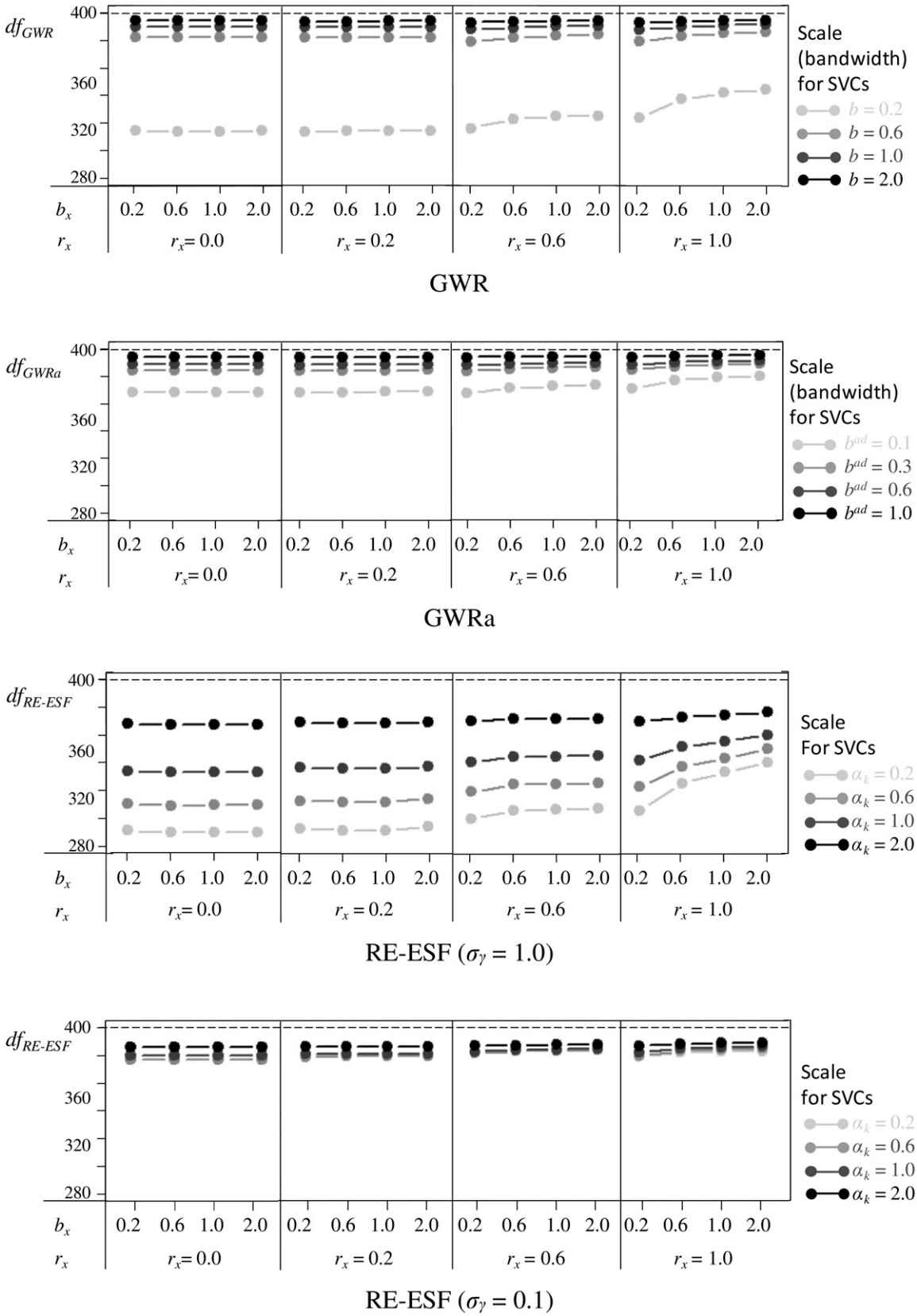

Figure 1. Mean degrees of freedom (*df*), with respect to the scale of the SVCs. In each panel, lighter lines represent *df*s evaluated with more locally tending SVCs, whereas darker lines represent *df*s evaluated with more globally tending SVCs. *Note:* GWR = geographically weighted regression; GWRa = geographically weighted regression with adapted distance bandwidth; RE-ESF = random effects eigenvector spatial filtering; SVCs = spatially varying coefficients.

is estimated appropriately. Furthermore, the mean df_{ESF} values, which equal the sample size minus the number of selected predictor variables, become {373.2, 349.3, 324.5, 301.7} in cases where the ratio of selected eigenvectors equals {0.2, 0.4, 0.6, 0.8}, respectively. Considering the usefulness of the shrinkage parameter, σ_k, in RE-ESF, regularized ESF (e.g., Seya et al. 2011) might be useful to increase df_{ESF}.

In summary, fixed bandwidth GWR can be very unstable when the bandwidth of local parameter estimation is inappropriately small, and ESF tends to be unstable as the number of selected eigenvectors increases. Conversely, GWRa is stable across both small- and large-scale SVC processes, and RE-ESF is similarly stable, provided that σ_k is estimated appropriately.

Also observable from Figure 1 is that predictor variable $\mathbf{x}_k$, with small-scale spatial variation universally, makes all SVC models unstable. In contrast, in the context of global spatial regression (e.g., spatial error model; LeSage and Pace 2009), Paciorek (2010) analytically showed that the coefficient estimates tend to be unstable if the spatial scales of the predictor variables are larger than the scale of the residual spatial process.[4] Thus, both small-scale $\mathbf{x}_k$ and large-scale $\mathbf{x}_k$ influence the reliability of the SVC estimates for different reasons. The next section discusses the influence of all of these instabilities on the accuracy of the SVC estimates themselves and tries to determine whether small-scale $\mathbf{x}_k$ or large-scale $\mathbf{x}_k$ is more harmful in SVC estimation. Besides, the scale of the true $\boldsymbol{\beta}_k$s also might influence their estimation accuracy. In this respect, the next section summarizes analyses about how the scales of $\mathbf{x}_k$ and $\boldsymbol{\beta}_k$ and their mismatches influence SVC estimation accuracy.

Simulation 2: Scales of $\mathbf{x}_k$ and $\boldsymbol{\beta}_k$ versus SVC Estimates

Overview

This section presents comparisons of all six study SVC models (GWR, GWRa, FB-GWR, FB-GWRa, ESF, and RE-ESF) through another Monte Carlo simulation experiment, where we now assess the accuracy of the estimated SVCs in relation to the (known) simulated SVCs. The synthetic data are generated with the following SVCs model:

$$\begin{aligned}
&\mathbf{y} = \boldsymbol{\beta}_0 + \mathbf{x}_1 \circ \boldsymbol{\beta}_1 + \mathbf{x}_2 \circ \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \quad 2^2\mathbf{I}),\\
&\boldsymbol{\beta}_0 = \mathbf{1} + \mathbf{C}(b_0)\boldsymbol{\varepsilon}_0, \quad \boldsymbol{\beta}_1 = (-2)\mathbf{1} + 3\mathbf{C}(b_1)\boldsymbol{\varepsilon}_1,\\
&\boldsymbol{\beta}_2 = (0.5)\mathbf{1} + \mathbf{C}(b_2)\boldsymbol{\varepsilon}_2,
\end{aligned} \tag{15}$$

where $\boldsymbol{\varepsilon}_k \sim N(\mathbf{0}, \mathbf{I})$. The spatial variation in $\boldsymbol{\beta}_1$ is three times greater than the spatial variation in $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_2$. We refer to $\boldsymbol{\beta}_1$ as a significant SVC process, whereas $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_2$ are considered to be insignificant SVC processes. This simulation experiment thus specifically investigates the second model limitation from earlier. Following the previous section, the predictor variables are generated from $\mathbf{x}_k = (1-r_x)\boldsymbol{\varepsilon}_{x(ns)} + r_x \mathbf{C}(b_x)\boldsymbol{\varepsilon}_{x(s)}$. Parameters are estimated 200 times while varying parameter values, as summarized in Table 4.

Results

The accuracy of each model's SVC estimates is evaluated using root mean squared error (RMSE) and bias diagnostics. The results and explanations of the bias diagnostics are given in the Supplemental Material, and only the RMSE results are presented here.

The RMSE for estimated $\hat{\boldsymbol{\beta}}_\mathbf{k}$ is formulated as follows:

$$RMSE\left[\hat{\boldsymbol{\beta}}_k\right] = \sqrt{\frac{1}{200N}\sum_{iter=1}^{200}\sum_{i=1}^{N}\left(\beta_k(s_i) - \hat{\beta}_k(s_i)^{iter}\right)^2}, \tag{16}$$

where $\beta_k(s_i)$ is the true SVC value generated from Equation 15 and $\hat{\beta}(s_i)_k^{iter}$ is the estimate obtained in the *iter*th iteration. To visualize the simulation results effectively, we use the two-dimensional plots presented in Figures 2 through 6. Here the horizontal axis always denotes the RMSEs for RE-ESF, whose SVC estimation accuracy has been shown to

Table 4. Parameter settings in 144 ($3 \times 4 \times 3 \times 4$) cases

Parameter	Notation	Case
Sample size	N	{50, 150, 400}
Bandwidth for $\{\boldsymbol{\beta}_0. \boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$	(b_0, b_1, b_2)	{(0.2, 0.2, 0.2), (1.0, 0.2, 1.0), (0.2, 1.0, 0.2), (1.0, 1.0, 1.0)}
Bandwidth for $\mathbf{x}_k$	b_x	{0.2, 0.6, 1.0}
Ratio of spatial variation in $\mathbf{x}_k$	r_x	{0.0, 0.4, 0.8, 1.0}

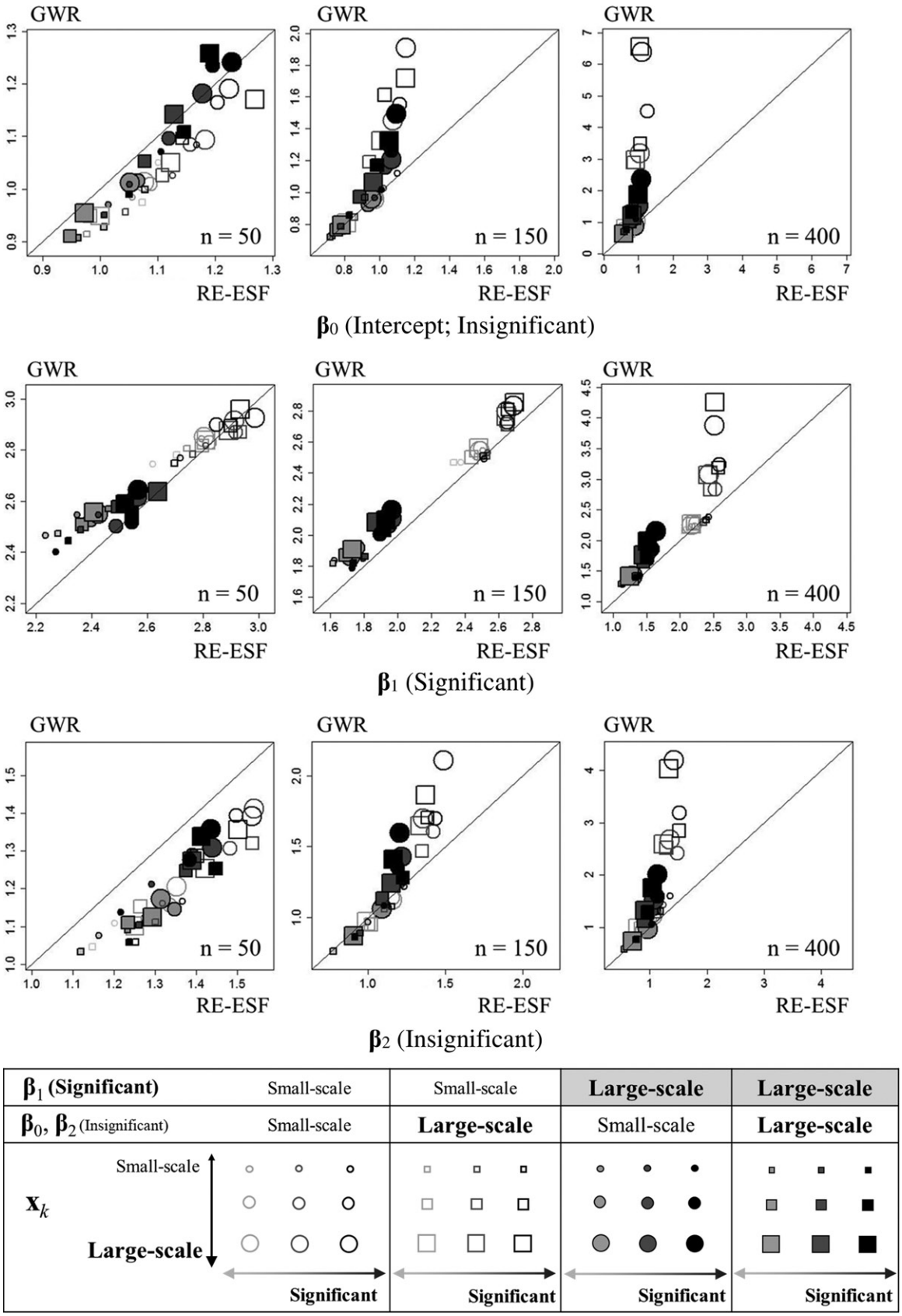

Figure 2. Root mean square error: Random effects eigenvector spatial filtering (x-axis) versus geographically weighted regresssion (y-axis). Large-scale means the large-scale (b and b_x). The lighter end of the Significant line means a small variance of the spatially dependent component (s_x), and the darker end means a large variance. GWR = geographically weighted regression; RE-ESF = random effects eigenvector spatial filtering.

be relatively good across all cases in the companion study of Murakami et al. (2017), and the vertical axis denotes the RMSEs for one of the models (GWR, ESF, FB-GWR, GWRa, and FB-GWRa).

Figure 2 compares (fixed bandwidth) GWR with RE-ESF for SVC estimation accuracy via RMSE. Here, GWR provides more accurate SVC estimation than RE-ESF if the plot outputs are concentrated in the bottom right triangle of each panel, whereas the estimated SVCs from RE-ESF are more accurate if the plot outputs are in the top left triangle. Results clearly demonstrate that the RMSEs from GWR are generally greater than those from RE-ESF and thus RE-ESF tends to be more accurate. This tendency is most conspicuous when the significant SVC ($\boldsymbol{\beta}_1$) has small-scale variations and $\mathbf{x}_k$ has strong large-scale variations, verifying that different scales of relationship nonstationarity need to be accounted for in SVC models (which RE-ESF does but GWR does not). This tendency is also substantial for the largest sample size, when $N = 400$. By contrast, GWR can perform equally as well as, or better than, RE-ESF when $N = 50$. This is interesting and might suggest that the smaller the sample size, the more difficult it is to detect relationships varying locally and across different spatial scales. Furthermore, Páez, Farber, and Wheeler (2011) recommended using GWR only for relatively large samples ($N > 160$), but these results suggest some value in GWR for small samples. Figure 3 compares the RMSE of estimated coefficients for ESF with RE-ESF, where it is clear that ESF provides poorer levels of SVC estimation accuracy than RE-ESF across all nine scenarios. As with GWR (and as would be expected), ESF provides relatively inaccurate SVC estimates in cases with small-scale variations in the significant SVC ($\boldsymbol{\beta}_1$) and strong large-scale variations in $\mathbf{x}_s$. Although not shown graphically, comparing **Figures 2** and 3 implies strongly that GWR tends to perform better than ESF.

Figure 4 compares FB-GWR and RE-ESF, where, interestingly, unlike GWR, no singular estimates appear from FB-GWR. Thus, GWR with multiple bandwidths (in this FB-GWR form) appears to stabilize SVC estimates and tentatively might provide a useful alternative to a regularized GWR model in addressing local collinearity issues. Because most plots follow the 45° line in Figure 4, FB-GWR provides SVC estimates that tend to be just as accurate as those from RE-ESF. Moreover, FB-GWR SVC estimates are more accurate than RE-ESF when $N = 50$. This is because RE-ESF is a likelihood approach relying on the law of large numbers. Conversely, the SVC estimates for FB-GWR tend to be marginally less accurate than those from RE-ESF when $\mathbf{x}_k$ have strong large-scale variation, and the significant SVC ($\boldsymbol{\beta}_1$) has small-scale variation (but for $N = 150$ and for $N = 400$ only).

Figure 5 compares GWRa with RE-ESF for SVC estimation accuracy. Somewhat surprisingly, GWRa does not suffer from any singular fit, and RMSE values are greatly reduced compared to the fixed bandwidth GWR results in Figure 2. The use of an adaptive bandwidth appears to be a simple and efficient solution to stabilize GWR modeling, although, in this case, stability might relate more to the effects of sample configurations than to other influences. Conversely, GWRa provides much poorer levels of SVC estimation accuracy (than GWR and RE-ESF) for the significant small-scale SVC, $\boldsymbol{\beta}_1$. Furthermore, the GWRa estimates for insignificant SVCs, $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_2$, tend to be more accurate than that found for GWR (see Figure 2), whereas the GWRa coefficient estimates provide broadly similar levels of accuracy to that found for RE-ESF for such cases. Figure 6 compares FB-GWRa with RE-ESF for SVC estimation accuracy. Here FB-GWRa appears to have accuracy tendencies similar to those found with both FB-GWR (Figure 4) and GWRa (Figure 5), in relation to RE-ESF. As would be expected, FB-GWRa is more accurate than GWRa for the significant small-scale SVC, $\boldsymbol{\beta}_1$, where the FB-GWRa results are more compatible with those from RE-ESF. Overall, FB-GWRa is found to estimate both weak and strong SVC processes relatively accurately.

The drawback to GWRa and FB-GWRa, however, is that the use of adaptive bandwidths implies that nonstationary relationships are operating within their own local region of dependence, whereas fixed bandwidth GWR and FB-GWR ensures that these regions are the same size everywhere and thus provides more generalized interpretations of the geographical process under study. For example, reporting that the nature of the relationship between crime and unemployment depends only on incident characteristics within a 2-km radius of the crime scene is intuitively more informative than reporting that this relationship depends only on the characteristics of the nearest thirty incidents of the crime scene.

Finally, Table 5 compares average computational times for all six SVC models for the three sample

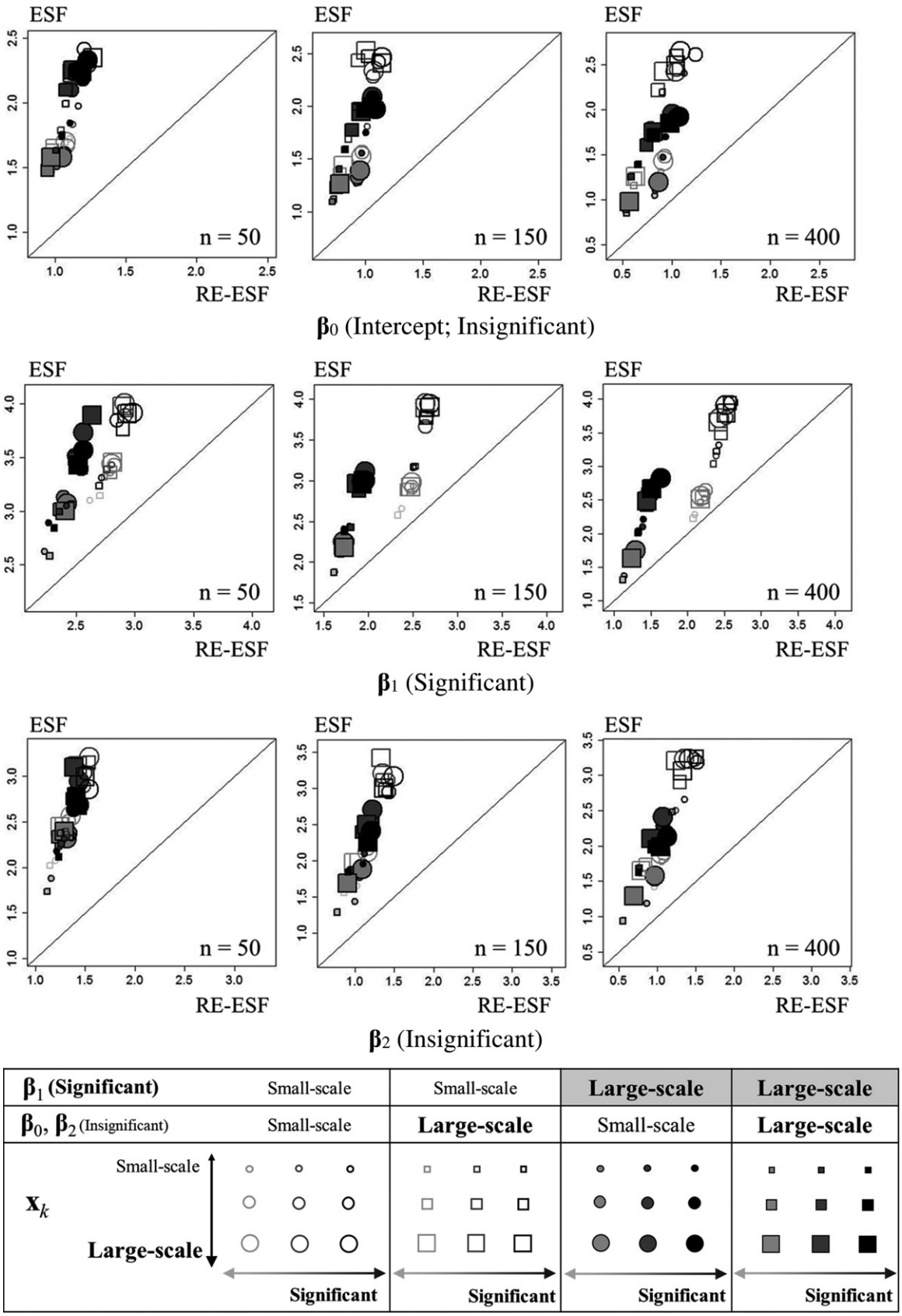

Figure 3. Root mean square error: Random effects eigenvector spatial filtering (*x*-axis) versus eigenvector spatial filtering (*y*-axis). Large-scale means the large scale (b and b_x). The lighter end of the Significant line means a small variance of the spatially dependent component (s_x), and the darker end means a large variance. ESF = eigenvector spatial filtering; RE-ESF = random effects eigenvector spatial filtering.

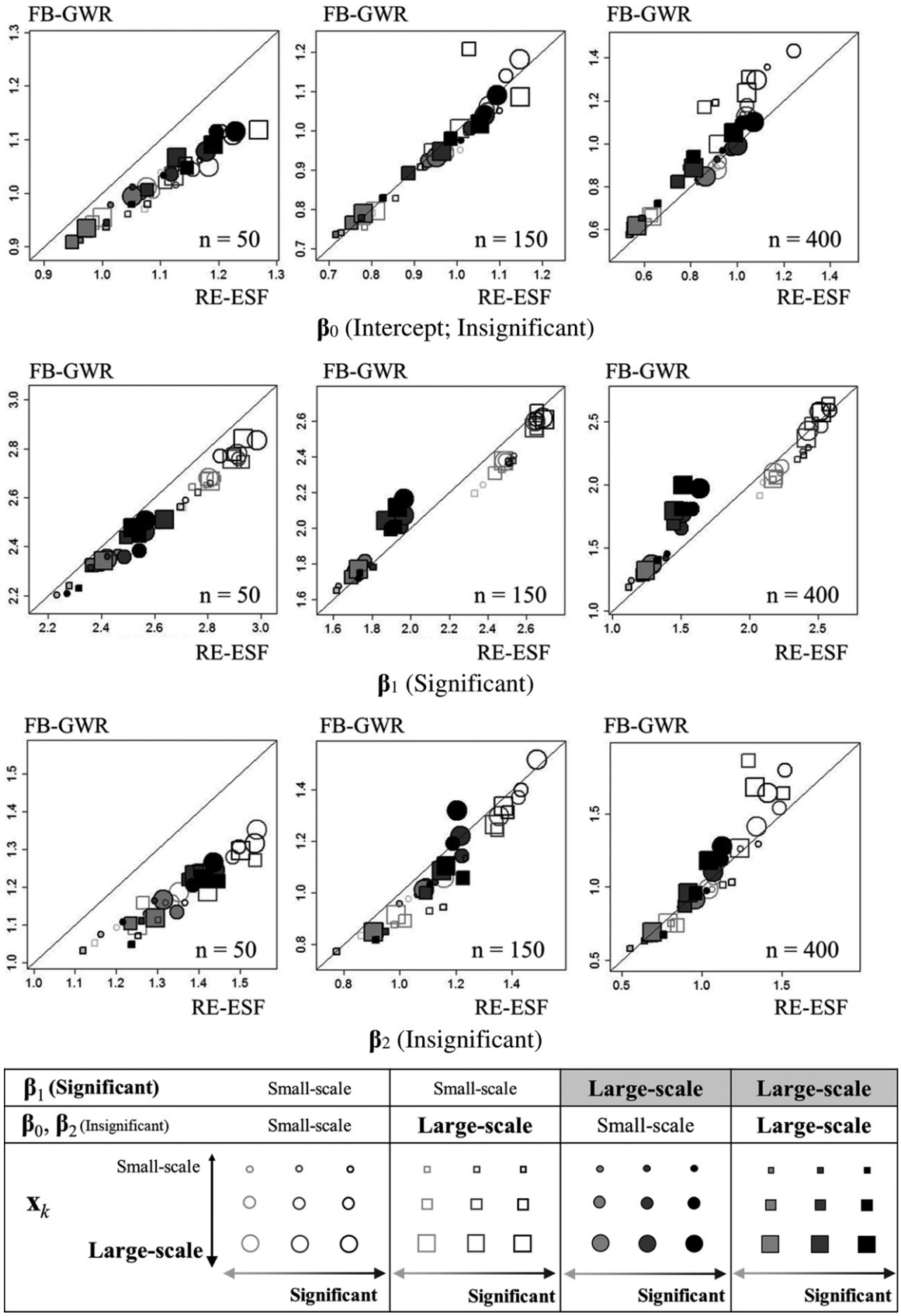

Figure 4. Root mean square error: Random effects eigenvector spatial filtering (*x*-axis) versus flexible bandwidth geographically weighted regression (*y*-axis). Large-scale means the large scale (b and b_x). The lighter end of the Significant line means a small variance of the spatially dependent component (s_x), and the darker end means a large variance. RE-ESF = random effects eigenvector spatial filtering; FB-GWR = flexible bandwidth geographically weighted regression.

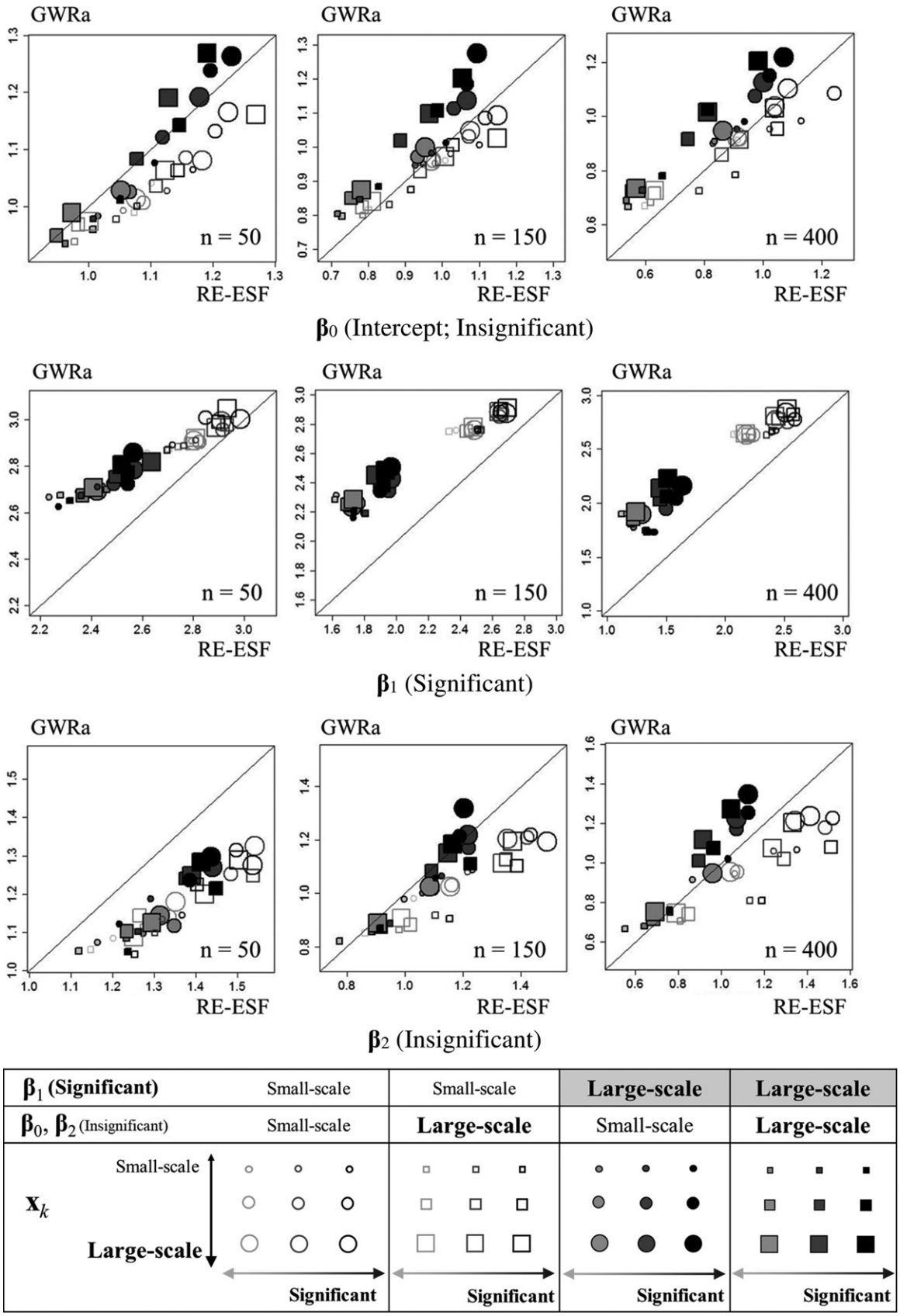

Figure 5. Root mean square error: Random effects eigenvector spatial filtering (x-axis) versus geographically weighted regression with adapted distance bandwidth (y-axis). Large-scale means the large scale (b and b_x). The lighter end of the Significant line means a small variance of the spatially dependent component (s_x), and the darker end means a large variance. RE-ESF = random effects eigenvector spatial filtering; GWRa = geographically weighted regression with adaptive distance bandwidths.

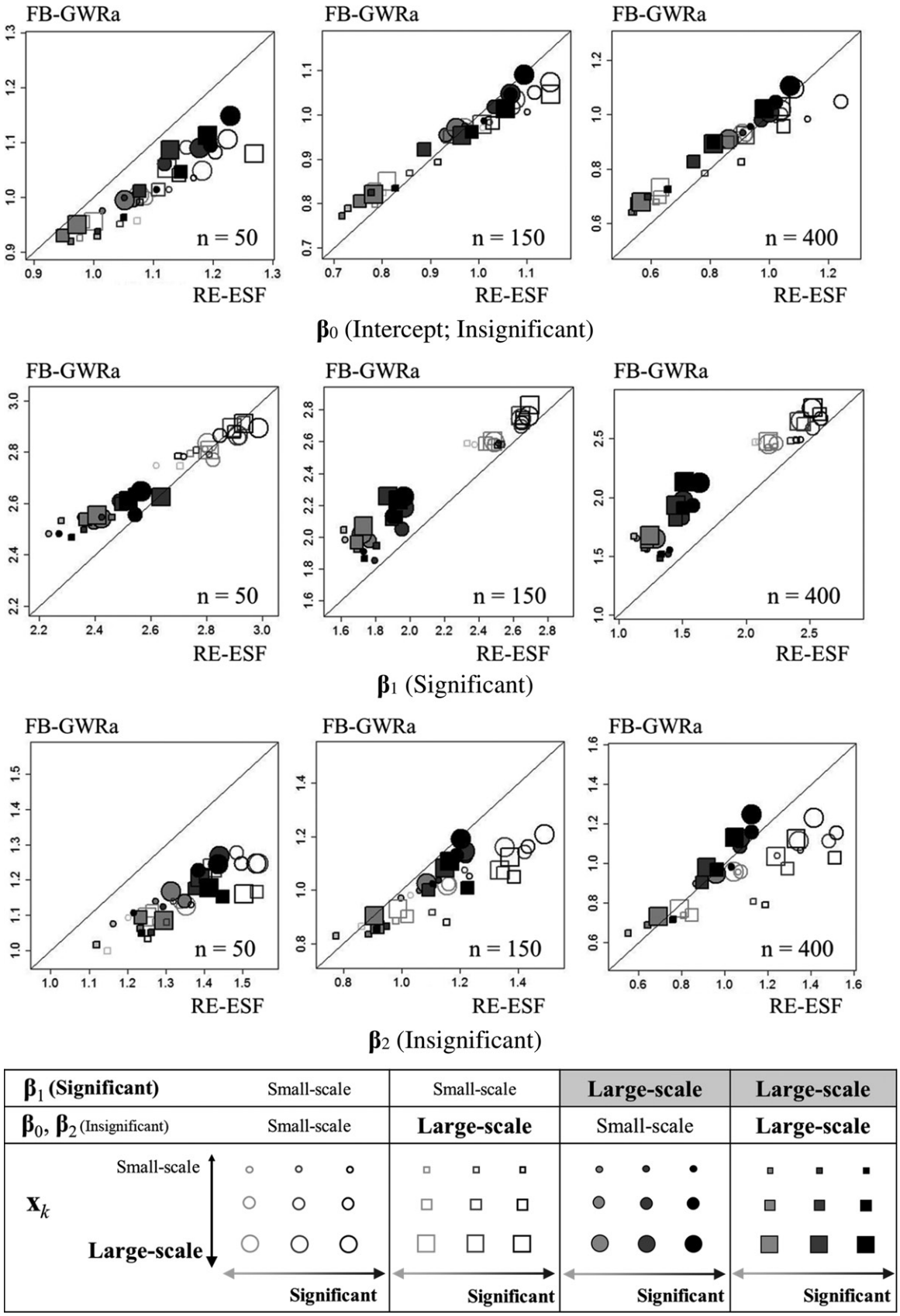

Figure 6. Root mean square error: Random effects eigenvector spatial filtering (*x*-axis) versus flexible bandwidth geographically weighted regression with adaptive distance bandwidths (*y*-axis). Large-scale means the large-scale (b and b_x). The lighter end of the Significant line means a small variance of the spatially dependent component (s_x), and the darker end means a large variance. RE-ESF = random effects eigenvector spatial filtering; FB-GWRa = flexible bandwidth geographically weighted regression with adaptive distance bandwidths.

Table 5. Average computational time in seconds

N	GWR	GWRa	FB-GWR	FB-GWRa	ESF	RE-ESF
50	0.13	0.18	1.50	10.31	1.49	0.29
150	0.54	0.72	12.02	12.96	10.52	0.77
400	2.44	2.63	93.52	65.41	72.56	3.38

Note: GWR = geographically weighted regression; GWRa = geographically weighted regression with adaptive distance bandwidths; FB-GWR = flexible bandwidth geographically weighted regression; FB-GRWa = flexible bandwidth geographically weighted regression with adaptive distance bandwidths; ESF = eigenvector spatial filtering; RE-ESF = random effects eigenvector spatial filtering.

sizes of $N = 50$, 150, and 400. As expected, GWR and GWRa run the fastest, because they are relatively simple. Considering the RMSE coefficient accuracy results earlier, one implied recommendation is that GWRa would often be a sensible and pragmatic choice for very large data sets. By contrast, FB-GWR and FB-GWRa are relatively slow due to their usage of the back-fitting algorithm in their calibration. Acceleration of these multiscale GWR models would be an important research topic in the future, although some work in this area is currently in progress (Lu et al. 2018). The ESF model is also slow because it requires stepwise eigenvector selection. By contrast, RE-ESF is as fast as GWR and GWRa, despite the fact that it estimates each spatial scale of each set of SVCs (i.e., RE-ESF is multiscale). This is because the computational complexity for optimizing the scale parameters is only $O(L^3K^3)$, which is independent of sample size. Note also that the cost for eigen-decomposition for RE-ESF, which is severe when N is large, can be lightened dramatically by an approximation proposed by Griffith (2000), which is for regular lattice data, or Murakami and Griffith (2017). Thus, RE-ESF also is recommended for very large data sets and should be preferred to GWRa when relationships are expected to vary not only locally but also across different spatial scales.

Concluding Remarks

This study investigated the influence of scale on SVC modeling, where relationships between the response and predictor operate not only locally but also at varying spatial scales. Results from simulation experiments suggest that standard GWR provides poor SVC estimates, when some SVCs vary at a small scale whereas others vary at a large scale. By contrast, a multiscale GWR model in FB-GWR provides SVC estimates that are relatively accurate for such processes. Differences in SVC estimation accuracy, and also SVC model stability, further depend on whether fixed distance or adaptive distance kernel bandwidths are specified for GWR or for FB-GWR, where, in general, adaptive ones should be preferred.

GWR and FB-GWR are examples of local approaches to SVC modeling, whereas ESF and RE-ESF models are both global approaches. Here RE-ESF is a regularized ESF model that is designed to capture scale dependencies in local relationships, just as FB-GWR is. In this study, RE-ESF is shown to more accurately estimate such multiscale SVC processes in comparison to not only ESF but also GWR. RE-ESF is also shown to be a more stable model than ESF or GWR. Both FB-GWR and RE-ESF are found to provide the most accurate estimates of the SVC processes generated in the simulation experiments but where RE-ESF is shown to be the most computationally efficient and thus more suitable for very large data sets. Overall, the results indicate that any future SVC study should pay keen attention to two important types of spatial scale—that of the scale of the associations and that of the scale of the predictors—and investigate with an FB-GWR or RE-ESF model accordingly. This is especially important considering any true set of local regression relationships does not operate at the same spatial scale (as naïvely assumed in the standard GWR model).

Still, there are remaining issues, where future work about the analytic properties of spatial scale and SVC estimates could follow that of Paciorek (2010), where only the effects on global (stationary) regression coefficients were investigated. Such studies would improve understanding about the scale problem and possibly enable the establishment of a local–global indicator of scale dependence for SVCs. In addition, to better understand differences in SVC models in practice, a wide variety of empirical SVC model studies is required. Further simulation work also might prove worthwhile; for example, using different simulation designs. We conducted some preliminary work in this area, where SVC model performance was assessed for retrieving smoothly varying deterministic coefficient processes rather than the random effect coefficient processes generated in this study. The results are given in the Supplemental Material.

Whether through empirical or simulation studies, the influence of a kernel function on FB-GWR

performance also should be assessed. Although this study used an exponential kernel, which decays slowly, the slow decay might help induce stability and accuracy of FB-GWR performance; a bi-square or tri-cube kernel that decay quickly might have a different effect in this respect. Gaussian, Cauchy (see Nakaya 2001), and anisotropic kernels (see Finley et al. 2009) also could be assessed.

Further comparison with Bayesian SVC models also is needed, especially as their computational efficiency continues to improve. For example, Finley et al. (2010) approximated a Bayesian SVC model with a low-rank approach, replacing the computationally heavy Markov chain Monte Carlo with an integrated nested Laplace approximation (Blangiardo and Cameletti 2015), which appears to be a likely avenue of model development. The RE-ESF can be viewed as a maximum a posteriori estimation of a low-rank Bayesian SVC model (see Murakami et al. 2017).

Finally, this and related studies have only considered basic multiscale SVC models; many extensions are possible, all of which warrant study, including (1) non-Gaussian SVC models (Atkinson et al. 2003; Griffith 2002, 2004; Nakaya et al. 2005), (2) spatiotemporal varying coefficient models (Huang, Wu, and Barry 2010; Griffith 2012; Fotheringham, Crespo, and Yao 2015), (3) spatial prediction SVC models (Harris et al. 2010; Harris, Brunsdon, and Fotheringham 2011; Griffith 2013), (4) spatial interaction models with SVCs (Nakaya 2001; Kordi and Fotheringham 2016; Griffith, Fischer, and LeSage 2017), and (5) the mitigation of the modifiable areal unit problem through SVC models (Fotheringham, Brunsdon, and Charlton 2002; Murakami and Tsutsumi 2015).

Funding

This work was supported by the National Natural Science Foundation of China (41401455, U1533102), the Japan Society for the Promotion of Science (17K12974, 17K14738, 15H04054), and the Biotechnology and Biological Sciences Research Council grants – BBS/E/C/000J0100, BBS/E/C/000I03320 and BBS/E/C/000I0330. The contribution of Science Foundation Ireland (Investigators Programme Grant 15/IA/3090 – Building City Dashboards) is gratefully acknowledged.

Supplemental Material

Supplemental material for this article is available on the publisher's website.

ORCID

Binbin Lu http://orcid.org/0000-0001-7847-7560
Paul Harris http://orcid.org/0000-0003-0259-4079
Chris Brunsdon http://orcid.org/0000-0003-4254-1780
Martin Charlton http://orcid.org/0000-0002-0622-393X
Tomoki Nakaya http://orcid.org/0000-0002-3827-1012
Daniel A. Griffith http://orcid.org/0000-0001-5125-6450

Notes

1. Griffith (2017) shows that the MC-based ESF is superior to the Geary's ratio-based ESF (Geary 1954), which could be used.
2. This problem, which is purely due to the local collinearity between $\mathbf{x}_1$ and $\mathbf{x}_k$, appears irrespective of the scale of the true spatially varying associations, $\boldsymbol{\beta}_k$s.
3. The predictor variables $\mathbf{x}_1$ and $\mathbf{x}_2$ are generated mutually independently. It would be an interesting topic for future work to evaluate SVC estimation accuracy by varying scales and the degree of multicollinearity simultaneously (e.g., Páez, Farber, and Wheeler 2011; Fotheringham and Oshan 2016; Oshan and Fotheringham 2017).
4. Estimation instability does not appear in this section because SVCs are implicitly assumed known and not estimated.

References

Anselin, L. 2010. Thirty years of spatial econometrics. *Papers in Regional Science* 89 (1):3–25. doi:10.1111/j.1435-5957.2010.00279.x.

Anselin, L., and S. Rey. 1991. Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23 (2):112–31. doi:10.1111/j.1538-4632.1991.tb00228.x.

Assunção, R. M. 2003. Space varying coefficient models for small area data. *Environmetrics* 14:453–73. doi:10.1002/env.599.

Atkinson, P. M., S. E. German, D. A. Sear, and M. J. Clark. 2003. Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geographical Analysis* 35 (1):58–82. doi:10.1111/j.1538-4632.2003.tb01101.x.

Bárcena, M. J., P. Menéndez, M. B. Palacios, and F. Tusell. 2014. Alleviating the effect of collinearity in geographically weighted regression. *Journal of Geographical Systems* 16 (4):441–66. doi:10.1007/s10109-014-0199-6.

Besag, J., J. York, and A. Mollie. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43:1–59. doi:10.1007/BF00116466.

Bishop, C. M. 2006. *Pattern recognition and machine learning*. New York: Springer.

Bitter, C., G. F. Mulligan, and S. Dall'erba. 2007. Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems* 9 (1):7–27. doi:10.1007/s10109-006-0028-7.

Brunsdon, C., A. S. Fotheringham, and M. E. Charlton. 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28 (4):281–98. doi:10.1111/j.1538-4632.1996.tb00936.x.

———. 1998. Geographically weighted regression—Modelling spatial non-stationarity. *Journal of the Royal Statistical Society, Series D: The Statistician* 47 (3):431–43. doi:10.1111/1467-9884.00145.

———. 1999. Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science* 39 (3):497–524. doi:10.1111/0022-4146.00146.

Brunsdon, C., J. McClatchey, and D. J. Unwin. 2001. Spatial variations in the average rainfall–altitude relationship in Great Britain: An approach using geographically weighted regression. *International Journal of Climatology* 21 (4):455–66. doi:10.1002/joc.614.

Casetti, E. 1972. Generating models by the expansion method: Applications to geographical research. *Geographical Analysis* 4 (1):81–91. doi:10.1111/j.1538-4632.1972.tb00458.x.

Cho, S.-H., D. M. Lambert, and Z. Chen. 2010. Geographically weighted regression bandwidth selection and spatial autocorrelation: An empirical example using Chinese agriculture data. *Applied Economics Letters* 17:767–72. doi:10.1080/13504850802314452.

Cliff, A. D., and J. K. Ord. 1973. *Spatial autocorrelation*. London: Pion.

Farber, S., and A. Páez. 2007. A systematic investigation of cross-validation in GWR model estimation: Empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems* 9 (4):371–96. doi:10.1007/s10109-007-0051-3.

Finley, A. O. 2011. Comparing spatially varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution* 2 (2):143–54. doi:10.1111/j.2041-210X.2010.00060.x.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically weighted regression: The analysis of spatially varying relationships*. West Sussex, UK: Wiley.

Fotheringham, A. S., R. Crespo, and J. Yao. 2015. Geographical and temporal weighted regression (GTWR). *Geographical Analysis* 47 (4):431–52. doi:10.1111/gean.12071.

Fotheringham, A. S., and T. M. Oshan. 2016. Geographically weighted regression and multicollinearity: Dispelling the myth. *Journal of Geographical Systems* 18 (4):303–29. doi:10.1007/s10109-016-0239-5.

Fotheringham, A. S., W. Yang, and W. Kang. 2017. Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*. 6: 1247–65. doi:10.1080/24694452.2017.1352480.

Gamerman, D., A. R. B. Moreira, and H. Rue. 2003. Space-varying regression models: Specifications and simulation. *Computational Statistics and Data Analysis* 42:513–33. doi:10.1016/S0167-9473(02)00211-6.

Geary, R. C. 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician* 5 (3):115–46. doi:10.2307/2986645.

Gelfand, A. E., H. J. Kim, C. F. Sirmans, and S. Banerjee. 2003. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association* 98 (462):387–96. doi:10.1198/016214503000170.

Gollini, I., B. Lu, M. Charlton, C. Brunsdon, and P. Harris. 2015. GWmodel: An R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software* 63 (17):1–50. doi:10.18637/jss.v063.i17.

Goodchild, M. F. 2004. The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers* 94 (2):300–303. doi:10.1111/j.1467-8306.2004.09402008.x.

Griffith, D. A. 2000. Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra & Its Applications* 321 (1–3):95–112. doi:10.1016/S0024-3795(00)00031-8.

———. 2002. A spatial filtering specification for the auto-Poisson model. *Statistics & Probability Letters* 58 (3):245–51. doi:10.1016/S0167-7152(02)00099-8.

———. 2003. *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization*. Berlin: Springer Science & Business Media.

———. 2004. A spatial filtering specification for the auto-logistic model. *Environment and Planning* A 36 (10):1791–1811. doi:10.1068/a36247.

———. 2008. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning* A 40 (11):2751–69. doi:10.1068/a38218.

———. 2012. Space, time, and space–time eigenvector filter specifications that account for autocorrelation. *Estadística Española* 54 (177):7–34.

———. 2013. Estimating missing data values for georeferenced Poisson counts. *Geographical Analysis* 45 (3):259–84. doi:10.1111/gean.12015.

Griffith, D. A.. 2017. Some robustness assessments of Moran eigenvector spatial filtering. *Spatial Statistics* 22 (2):155–79. doi:10.1016/j.spasta.2017.09.001.

Griffith, D. A., D. M. Fischer, and J. P. LeSage. 2017. The spatial autocorrelation problem in spatial interaction modelling: A comparison of two common solutions. *Letters in Spatial and Resource Sciences* 10 (1):75–86. doi:10.1007/s12076-016-0172-8.

Harris, P., C. Brunsdon, and A. S. Fotheringham. 2011. Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. *Stochastic Environmental Research and Risk Assessment* 25 (2):123–38. doi:10.1007/s00477-010-0444-6.

Harris, P., C. Brunsdon, B. Lu, T. Nakaya, and M. Charlton. 2017. Introducing bootstrap methods to investigate coefficient non-stationarity in spatial regression models. *Spatial Statistics* 21:241–61. doi:10.1016/j.spasta.2017.07.006.

Harris, P., A. S. Fotheringham, R. Crespo, and M. Charlton. 2010. The use of geographically weighted regression for spatial prediction: An evaluation of models using simulated data sets. *Mathematical Geosciences* 42 (6):657–80. doi:10.1007/s11004-010-9284-7.

Harris, P., A. S. Fotheringham, and S. Juggins. 2010. Robust geographically weighted regression: A technique for quantifying spatial relationships between freshwater acidification critical loads and catchment attributes. *Annals of the Association of American Geographers* 100 (2):286–306. doi:10.1080/00045600903550378.

Helbich, M., and D. A. Griffith. 2016. Spatially varying coefficient models in real estate: Eigenvector spatial filtering and alternative approaches. *Computers, Environment and Urban Systems* 57:1–11. doi:10.1016/j.compenvurbsys.2015.12.002.

Hu, M, Z. Li, J. Wang, L. Jia, S. Lai, Y. Guo, D. Zhao, and W. Yang. 2012. Determinants of the incidence of hand, foot and mouth disease in China using geographically weighted regression models. *PLoS ONE* 7 (6):e38978. doi:10.1371/journal.pone.0038978.

Huang, B., B. Wu, and M. Barry. 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science* 24 (3):383–401. doi:10.1080/13658810802672469.

Jaimes, N. B. P., J. B. Sendra, M. G. Delgado, and R. F. Plata. 2010. Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. *Applied Geography* 30 (4):576–91. doi:10.1016/j.apgeog.2010.05.004.

Jones, J. P., III, and E. Casetti. 1992. *Applications of the expansion method*. London and New York: Routledge.

Kordi, M., and A. S. Fotheringham. 2016. Spatially weighted interaction models (SWIM). *Annals of the American Association of Geographers* 106 (5):990–1012. doi:10.1080/24694452.2016.1191990.

Leong, Y. Y., and J. C. Yue. 2017. A modification to geographically weighted regression. *International Journal of Health Geographics* 16 (1):11. doi:10.1186/s12942-017-0085-9.

LeSage, J. P., and R. K. Pace. 2009. *Introduction to spatial econometrics*. Boca Raton, FL: CRC.

Lloyd, C. D. 2010. *Local models for spatial analysis*. Boca Raton, FL: CRC.

Lu, B., C. Brunsdon, M. Charlton, and P. Harris. 2017. Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science* 31 (5):982–98. doi:10.1080/13658816.2016.1263731.

Lu, B., M. Charlton, P. Harris, and A. S. Fotheringham. 2014. Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. *International Journal of Geographical Information Science* 28 (4):660–81. doi:10.1080/13658816.2013.865739.

Lu, B., P. Harris, M. Charlton, and C. Brunsdon. 2014. The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science* 17 (2):85–101. doi:10.1080/10095020.2014.917453.

Lu, B., W. Yang, Y. Ge, and P. Harris. 2018. Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths. *Computers, Environment and Urban Systems* 71: 41–57. doi:10.1016/j.compenvurbsys.2018.03.012.

Mei, C.-L., N. Wang, and W.-X. Zhang. 2006. Testing the importance of the explanatory variables in a mixed geographically weighted regression model. *Environment and Planning A* 38:587–98. doi:10.1068/a3768.

Mei, C.-L., M. Xu, and N. Wang. 2016. A bootstrap test for constant coefficients in geographically weighted regression models. *International Journal of Geographical Information Science* 30:1622–43. doi:10.1080/13658816.2016.1149181.

Murakami, D. 2017. Spmoran: An R package for Moran's eigenvector-based spatial regression analysis. *ArXiv* 1703.04467.

Murakami, D., and D. A. Griffith. 2015. Random effects specifications in eigenvector spatial filtering: A simulation study. *Journal of Geographical Systems* 17 (4) 311–31. doi:10.1007/s10109-015-0213-7.

———. 2018. Eigenvector spatial filtering for large data sets: Fixed and random effects approaches. *Geographical Analysis*. doi: 10.1111/gean.12156.

Murakami, D., and M. Tsutsumi. 2015. Area-to–point parameter estimation with geographically weighted regression. *Journal of Geographical Systems* 17 (3):207–25. doi:10.1007/s10109-015-0212-8.

Murakami, D., T. Yoshida, H. Seya, D. A. Griffith, and Y. Yamagata. 2017. A Moran coefficient–based mixed effects approach to investigate spatially varying relationships. *Spatial Statistics* 19:68–89. doi:10.1016/j.spasta.2016.12.001.

Nakaya, T. 2001. Local spatial interaction modelling based on the geographically weighted regression approach, *GeoJournal* 53:347–58. doi:10.1023/A:1020149315435.

———. 2015. Geographically weighted generalized linear modeling. In *Geocomputation: A practical primer*, ed. C. Brunsdon and A. Singleton, 201–20. London: Sage.

Nakaya, T., A. S. Fotheringham, C. Brunsdon, and M. Charlton. 2005. Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine* 24 (17):2695–2717. doi:10.1002/sim.2129.

Ndiath, M. N., B. Cisse, J. L. Ndiaye, J. F. Gomis, O. Bathiery, A. T. Dia, O. Gaye, and B. Faye. 2015. Application of geographically-weighted regression analysis to assess risk factors for malaria hotspots in Keur Soce health and demographic surveillance site. *Malaria Journal* 14:463. doi:10.1186/s12936-015-0976-9.

Oshan, T. M., and A. S. Fotheringham. 2017. Comparison of spatially varying regression coefficient estimates using geographically weighted and spatial-filter-based techniques. *Geographical Analysis*. 50:53–75. doi:10.1111/gean.12133.

Paciorek, C. J. 2010. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 25 (1):107. doi:10.1214/10-STS326.

Páez, A., S. Farber, and D. Wheeler. 2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A* 43 (12):2992–3010. doi:10.1068/a44111.

Páez, A., F. Long, and S. Farber. 2008. Moving window approaches for hedonic price estimation: An empirical comparison of modelling techniques. *Urban Studies* 45 (8):1565–81. doi:10.1177/0042098008091491.

Sampson, P. D., D. Damian, and P. Guttorp. 2001. Advances in modeling and inference for environmental processes with nonstationary spatial covariance. National Research Centre for Statistics and the Environment Technical Report Series 61:2001.

Seya, H., D. Murakami, M. Tsutsumi, and Y. Yamagata. 2015. Application of LASSO to the eigenvector selection problem in eigenvector-based spatial filtering. *Geographical Analysis* 47 (3):284–99. doi:10.1111/gean.12054.

Wheeler, D. C. 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A* 39 (10):2464–81. doi:10.1068/a38325.

———. 2009. Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso. *Environment and Planning A* 41 (3):722–42. doi:10.1068/a40256.

Wheeler, D. C., and C. A. Calder. 2007. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems* 9 (2):145–66. doi:10.1007/s10109-006-0040-y.

Wheeler, D. C., and M. Tiefelsdorf. 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems* 7 (2):161–87. doi:10.1007/s10109-005-0155-6.

Wheeler, D. C., and L. 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. *Journal of Geographical Systems* 11 (1):1–22. doi:10.1007/s10109-008-0073-5.

Yang, W. 2014. An extension of geographically weighted regression with flexible bandwidths. PhD dissertation, University of St. Andrews.

Yang, W., A. S. Fotheringham, and P. Harris. 2011. Model selection in GWR: The development of a flexible bandwidth GWR. In *The 11th International Conference on GeoComputation*, 20–22. London, UK.

———. 2012. An extension of geographically weighted regression with flexible bandwidths. In *GISRUK 2012*, Paper no. 79. Lancaster, UK.

DAISUKE MURAKAMI is an Assistant Professor in the Department of Data Science, Institute of Statistical Mathematics, Tachikawa, Tokyo 190–8562, Japan. E-mail: dmuraka@ism.ac.jp. His research interests include spatial and temporal statistics, quantitative geography, and regional science.

BINBIN LU is a Lecturer in the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, China. E-mail: binbinlu@whu.edu.cn. His research interests include spatial statistics, geographically weighted modeling, open-source geographic information systems, and R coding.

PAUL HARRIS is a Senior Research Scientist at Rothamsted Research, North Wyke, Okehampton, Devon EX20 2SB, UK. E-mail: paul.harris@rothamsted.ac.uk. His research focuses on the development and application of spatial statistics to agricultural, ecological, and environmental data.

CHRIS BRUNSDON is Professor of Geocomputation and Director of the National Centre for Geocomputation, National University of Ireland Maynooth, Maynooth, Kildare, Ireland. E-mail: christopher.brunsdon@mu.ie. His research interests include spatial data analysis and statistical modeling, computational data science, spatial data visualization, and the analysis of social and economic data.

MARTIN CHARLTON is Senior Lecturer in the National Centre for Geocomputation, Maynooth University, Maynooth, County Kildare, Ireland. E-mail: martin.charlton@nuim.ie. His research interests include spatial modeling, quantitative geography, geographic information science, and spatial epidemiology.

TOMOKI NAKAYA is Professor of Environmental Geography, Graduate School of Environmental Sciences, Tohoku University, 468-1, Aoba, Aramaki, Aoba-ku, Sendai, 980-0845, Japan. E-mail: tomoki.nakaya.c8@tohoku.ac.jp. His research interests include quantitative geography, spatial statistics, and spatial epidemiology.

DANIEL A. GRIFFITH is Ashbel Smith Professor of Geospatial Information Sciences in the School of Economic, Political, and Policy Sciences at the University of Texas at Dallas, Richardson, TX 75080. E-mail: dagriffith@utdallas.edu. His research interests include quantitative geography, spatial statistics, urban economics, and urban public health.