A Moran coefficient-based mixed effects approach to investigate spatially varying relationships

Murakami, Daisuke

Yoshida, Takahiro

Seya, Hajime

Griffith, Daniel A.

Yamagata, Yoshiki

(Citation)
Spatial Statistics,19:68-89

(Issue Date)
2017-02

(Resource Type)
journal article

(Version)
Accepted Manuscript

A Moran coefficient-based mixed effects approach to investigate spatially varying relationships

Daisuke Murakami[1,*], Takahiro Yoshida[1,2], Hajime Seya[3], Daniel A. Griffith[4], and Yoshiki Yamagata[1]

[1]Center for Global Environmental Research, National Institute for Environmental Studies,
16-2 Onogawa, Tsukuba, Ibaraki, 305-8506, Japan
Email: murakami.daisuke@nies.go.jp; yamagata@nies.go.jp

[2]Graduate School of Systems and Information Engineering, University of Tsukuba,
1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573, Japan
Email: yoshida.takahiro@sk.tsukuba.ac.jp

[3] Department of Civil Engineering, Graduate School of Engineering, Kobe University,
1-1 Rokkodai, Nada, Kobe, 657-8501, Japan
Email: hseya@people.kobe-u.ac.jp

[4]School of Economic, Political and Policy Science, The University of Texas, Dallas,
800 W Campbell Rd, Richardson, TX, 75080, USA
Email: dagriffith@utdallas.edu

*: Corresponding author

Abstract

This study develops a spatially varying coefficient model by extending the random effects eigenvector spatial filtering model. The developed model has the following properties: its spatially varying coefficients are defined by a linear combination of the eigenvectors describing the Moran coefficient; each of its coefficients can have a different degree of spatial smoothness; and it yields a variant of a Bayesian spatially varying coefficient model. Moreover, parameter estimation of the model can be executed with a relatively small computational burden. Results of a Monte Carlo simulation reveal that our model outperforms a conventional eigenvector spatial filtering (ESF) model and geographically weighted regression (GWR) models in terms of the accuracy of the coefficient estimates and computational time. We empirically apply our model to the hedonic land price analysis of flood hazards in Japan.

Keywords

Random effects, eigenvector spatial filtering, spatially varying coefficient, geographically

weighted regression, Moran coefficient, hedonic price analysis

1. Introduction

Spatial heterogeneity is one of the important characteristics of spatial data (Anselin, 1988). Geographically weighted regression (GWR) (Fotheringham et al., 2002; Wheeler and Páez, 2009; Fotheringham and Oshan, 2016) is one useful approach for explicitly accounting for spatial heterogeneity of the model structure through spatially varying coefficients (SVCs). GWR has been widely applied in socioeconomic studies (e.g., Bitter et al., 2007; Huang et al., 2010), ecological studies (e.g., Wang et al., 2005; Austin, 2007), health studies (e.g., Nakaya et al., 2005; Hu et al., 2012), and many others.

Despite the wide-ranging set of applications, existing studies have shown that the basic (original) GWR specification has several drawbacks. First, the coefficients of the basic GWR typically suffer from multicollinearity (Páez et al., 2011; Wheeler and Tiefelsdorf, 2005). Second, the basic GWR assumes the same degree of spatial smoothness for each coefficient, which is a rather strong assumption that fails to hold in most empirical applications. Fortunately, several extended GWRs have been proposed to address these problems. With regard to the first problem, Wheeler (2007; 2009) proposes regularized GWR, by combining ridge and/or lasso regression with GWR, and its robustness in terms of the multicollinearity problem has been demonstrated. The limitations of regularized GWR specifications are its bias in coefficient estimates, just like conventional ridge and/or lasso regression. With regard to the second problem concerning uniform smoothers, Yang et al. (2014) and Lu et al. (2015) attempted to overcome this limitation.

The Bayesian spatially varying coefficients (B-SVC) model, based on a geostatistical (Gelfand et al., 2003) or lattice autoregressive approach (Assunçao, 2003), is another form of the spatially varying coefficients model that requires Markov chain Monte Carlo (MCMC). Wheeler and Calder (2007) and Wheeler and Waller (2009) suggest that the coefficient estimates for the B-SVC model of Gelfand et al. (2003) are robust in terms of multicollinearity. In contrast to the GWR model, the B-SVC model allows differential spatial smoothness across coefficients. However, this differential makes computational costs prohibitive if a sample size is moderate to large (Finley, 2011).

Although Integrated Nested Laplace Approximation (INLA)[1] based SVC estimations are becoming available now (Congdon, 2014)[2], their estimation accuracy and computational efficiency are largely unexplored.

Hence, a SVC model with the following properties still needs to be developed: (a) robust to multicollinearity; (b) the possibility for each coefficient to have a different degree of spatial smoothness; and, (c) computational efficiency. This study develops a model satisfying these requirements by combining an eigenvector spatial filtering (ESF; Griffith 2003; Chun and Griffith, 2014) based SVC model (Griffith, 2008) and a random effects ESF (RE-ESF: Murakami and Griffith, 2015) model.

The following sections are organized as follows. Sections 2 and 3 introduce the GWR model and ESF-based SVC model of Griffith (2008), respectively. Section 4 introduces the RE-ESF model, and extends it to a SVC model. Section 5 compares the properties of our model with those of other SVC models. Section 6 summarizes results

---

[1] See Rue et al. (2009) for details on the INLA approach and Blangiardo and Cameletti (2015) for its R programming.

[2] Congdon (2014) publishes an R code of an INLA to estimate a conditional autoregressive model-based SVC model (see Gamerman et al., 2003).

from a comparative Monte Carlo simulation experiment, and section 7 uses our model in a hedonic analysis. Section 8 concludes our discussion.

2. GWR specifications

The basic GWR model for a site $s_i \in D \subset \Re^2$ is formulated as follows:

$$\mathbf{G}(s_i)^{1/2}\mathbf{y} = \mathbf{G}(s_i)^{1/2}\mathbf{X}\boldsymbol{\beta}(s_i) + \mathbf{u}, \qquad E[\mathbf{u}] = \mathbf{0}, \qquad Var[\mathbf{u}] = \sigma^2\mathbf{I}, \tag{1}$$

where $\mathbf{y}$ is an $N \times 1$ vector of continuous response variables, $\mathbf{X}$ is an $N \times K$ matrix of explanatory variables, $\boldsymbol{\beta}(s_i)$ is a $K \times 1$ vector of geographically varying coefficients, $\mathbf{u}$ is a $N \times 1$ vector of disturbances, $\mathbf{0}$ is an $N \times 1$ vector of zeros, $\mathbf{I}$ is an $N \times N$ identity matrix, σ^2 is a variance parameter, and $\mathbf{G}(s_i)$ is an $N \times N$ diagonal matrix whose j-th element is given by a geographically weighting function, $g(s_i, s_j)$. Eq.(1) is a regression linear model local weighted by $g(s_i, s_j)$. The weighted least squares (WLS) estimator of $\boldsymbol{\beta}(s_i)$ yields

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

where $'$ denotes the matrix transpose.

Stone (1980) and Fan (1993) show that locally weighted regression, including

GWR, maximizes the rate of asymptotic convergence to a true function that is given by a local linear smoother of **y**, and the smoothness of $g(s_i, s_j)$ is required to identify the true function. Wheeler and Calder (2007) and Wheeler and Waller (2009) applied the following exponential function form, which weighs more heavily for neighboring samples than distant samples:

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

where $d(s_i, s_j)$ is the Euclidean distance between locations s_i and s_j, and r is the bandwidth parameter, which is large if coefficients have global scale spatial variation, and small if they have local scale spatial variation.

A standard estimation procedure for the basic GWR is as follows: (1) the bandwidth is calculated based on the leave-one-out cross-validation or a corrected AIC minimization (see Fotheringham et al., 2002), and (2) $\boldsymbol{\beta}(s_i)$ is estimated by substituting the estimated bandwidth into Eqs. (2) and (3).

After Wheeler and Tiefelsdorf (2005) demonstrate that GWR coefficients essentially are collinear, active discussion shifted to regularized GWR. For example, Wheeler (2007) proposes a ridge regularization-based GWR that replaces Eq. (2) with the following equation:

$$\hat{\boldsymbol{\beta}}(s_i) = (\mathbf{X}'\mathbf{G}(s_i)\mathbf{X} + \eta\mathbf{I}_K)^{-1}\mathbf{X}'\mathbf{G}(s_i)\mathbf{y}, \tag{4}$$

where η is the ridge regularization parameter, and $\mathbf{I}_K$ is a $K \times K$ identity matrix. Wheeler (2009) and Gollini et al. (2015) extended the ridge GWR to vary η locally. Specifically, they propose the locally compensated ridge GWR (LCR-GWR) estimator, which is formulated as follows:

$$\hat{\boldsymbol{\beta}}(s_i) = (\mathbf{X}'\mathbf{G}(s_i)\mathbf{X} + \eta(s_i)\mathbf{I}_K)^{-1}\mathbf{X}'\mathbf{G}(s_i)\mathbf{y}, \tag{5}$$

where $\eta(s_i)$ is the ridge parameter for location s_i, and LCR-GWR calibrates $\eta(s_i)$ based on the degree of multicollinearity in the corresponding local model. Because $\eta(s_i)$ increases bias of the coefficient estimator, just like the standard ridge estimator, Gollini et al. (2015) suggest introducing $\eta(s_i)$ only for local models whose multicollinearity is excessive. The estimation procedure for LCR-GWR is as follows (Gollini et al., 2015): (1) the bandwidth and ridge parameters are estimated by the leave-one-out cross-validation, and (2) $\boldsymbol{\beta}(s_i)$ is estimated by substituting them into Eq. (5).

3. ESF-based SVC specifications

Section 3.1 introduces the ESF approach 1, and section 3.2 presents an ESF-based SVC model, which we extend to a RE-ESF-based model in section 4.

3.1. The ESF approach

Moran ESF is based on the Moran coefficient (MC; see, Anselin and Rey, 1991), which is a spatial dependence diagnostic statistic formulated as follows[3]:

$$MC = \frac{N}{\mathbf{1}'\mathbf{C}\mathbf{1}} \frac{\mathbf{y}'\mathbf{MCMy}}{\mathbf{y}'\mathbf{My}}. \tag{6}$$

where $\mathbf{1}$ is an $N \times 1$ vector of ones, $\mathbf{y}$ is an $N \times 1$ vector of variable values, $\mathbf{C}$ is an $N \times N$ connectivity matrix whose diagonal elements are zero, and $\mathbf{M} = \mathbf{I} - \mathbf{11}'/N$ is an $N \times N$ matrix for double centering. MC is positive if the sample values in $\mathbf{y}$ display positive spatial dependence and negative if they display negative spatial dependence. Based on Griffith (2003) and Griffith and Chun (2014), the 1st eigenvector, $\mathbf{e}_1$, is the set of real

[3] ESF also could be based on other indices, such as the Geary ratio (Geary, 1954).

numbers that has the largest MC value achievable by any set of real numbers for the spatial attunement defined by **C**; $\mathbf{e}_2$, is the set of real numbers that has the largest achievable MC value by any set that is orthogonal with $\mathbf{e}_1$; and so forth, the *l*-th eigenvector, $\mathbf{e}_l$, is the set of real numbers that has the largest achievable MC value by any set that is orthogonal with $\{\mathbf{e}_1, ..., \mathbf{e}_{l\text{-}1}\}$. Thus, $\mathbf{E}_{full} = \{\mathbf{e}_1, ..., \mathbf{e}_N\}$, provides all the possible distinct map pattern descriptions of latent spatial dependence, with each magnitude being indexed by its corresponding eigenvalue in $\{\lambda_1, ..., \lambda_N\}$ (Griffith, 2003).

M is replaced with $\mathbf{M}_X = \mathbf{I} – \mathbf{X}(\mathbf{X}'\,\mathbf{X})^{-1}\mathbf{X}'$ if **y** is a residual vector of a linear regression model. In that case, MC is positive if sample values in **y** have variations that are positively spatially dependent and orthogonal with **X**. The reverse is true for negative spatial dependence. The eigenvectors of $\mathbf{M}_X\mathbf{C}\mathbf{M}_X$ are defined as those for **MCM** except that they are orthogonal with **X**. In other words, $\mathbf{e}_l$, is the set of real numbers that has the largest achievable MC value by any set that is orthogonal with $\{\mathbf{e}_1, ..., \mathbf{e}_{l\text{-}1}\}$ and **X**.

ESF describes the latent map pattern in a georeferenced response variable **y**, using a linear combination of eigenvectors, $\mathbf{E}\boldsymbol{\gamma}$, where **E** is a matrix composed of *L* eigenvectors in $\mathbf{E}_{full}$ ($L < N$) that is given either from **MCM** or $\mathbf{M}_X\mathbf{C}\mathbf{M}_X$, and $\boldsymbol{\gamma} = [\gamma_1, ..., \gamma_L]'$ is an $L \times 1$ coefficient vector. The linear ESF model is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \tag{7}$$

where $\boldsymbol{\varepsilon}$ is a $N \times 1$ vector of disturbances. Because Eq. (7) is in the form of the standard linear regression model, ordinary least squares (OLS) estimation is applicable for its parameter estimation[4]. The L eigenvectors in **E** may be selected as follows (see, Chun et al., 2016): (a) eigenvectors corresponding to small eigenvalues, which explain the inconsequential level of spatial dependence, are removed[5], and (b) Eigenvectors are chosen by applying an accuracy maximization (e.g., adjusted R^2 maximization)–based or a residual MC minimization–based stepwise variable selection process to the candidate set prepared in (a).

Many studies demonstrate the effectiveness of ESF in estimation and inference for $\boldsymbol{\beta}$ in the presence of spatial dependence (e.g., Chun, 2014; Griffith and Chun, 2014;

---

[4] Another approach includes the model selection procedure based on LASSO (Seya et al., 2015).

[5] $\lambda_l / \lambda_1 > 0.25$ and $\lambda_l > 0$ are commonly used criteria (e.g., Griffith, 2003; Tiefelsdorf and Griffith, 2007; Drey, 2006; Hughes and Haran, 2013).

Margaretic et al., 2015). For more details about ESF, see Griffith (2003), Griffith and Paelinck (2011), Griffith and Chun (2014), and Griffith and Chun (2016).

3.2. ESF-based SVC specifications

To capture possible spatially varying influences from explanatory variables, Griffith (2008) extended ESF to the following SVC model:

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{x}_k \circ (\beta_k 1) + \sum_{k=1}^{K} \mathbf{x}_k \circ \mathbf{E}_k \boldsymbol{\gamma}_k + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{8}$$

which also is expressed as

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{x}_k \circ \boldsymbol{\beta}_k^{ESF} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{9}$$

$$\boldsymbol{\beta}_k^{ESF} = \beta_k \mathbf{1} + \mathbf{E}_k \boldsymbol{\gamma}_k,$$

where $\mathbf{x}_k$ is an $N \times 1$ vector of the k-th explanatory variable (i.e., k-th column of $\mathbf{X}$), $\mathbf{E}_k$ 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. Note that $\sum_{k=1}^{K} \mathbf{x}_k \circ (\beta_k 1)$ in Eq.(8) equals $\mathbf{X}\boldsymbol{\beta}$. $\boldsymbol{\beta}_k^{ESF} = \beta_k \mathbf{1} + \mathbf{E}_k \boldsymbol{\gamma}_k$ yields a vector of spatially varying coefficients, in which $\beta_k \mathbf{1}$ and $\mathbf{E}_k \boldsymbol{\gamma}_k$ represent the constant component and spatially varying component, respectively. The parameters can be estimated, as for the standard ESF specification, as follows: (a) eigenvectors corresponding to small eigenvalues are removed from each $\mathbf{E}_k$; (b) significant variables in $\mathbf{X}$, $\mathbf{x}_1 \circ \mathbf{E}_1$, ..., $\mathbf{x}_K \circ \mathbf{E}_K$ are selected by applying an OLS-based forward variable selection technique, and $\boldsymbol{\beta}= [\beta_1, ..., \beta_K]'$ and $\boldsymbol{\gamma}_k$ are then estimated; and, (c) $\hat{\boldsymbol{\beta}}_k^{ESF}$ = $\hat{\beta}_k \mathbf{1} + \mathbf{E}\hat{\boldsymbol{\gamma}}_k$ is calculated. Helbich and Griffith (2016) empirically demonstrated that spatial variation of the ESF-based coefficients can be significantly different from those for GWR.

4. RE-ESF-based SVC specifications

4.1. The RE-ESF approach

While the conventional ESF model is a fixed effects model, Murakami and Griffith (2015) show that random effects versions of ESF increase the estimation accuracy of regression coefficients and their standard errors with shorter computational time. This section extends RE-ESF to a SVC model as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\gamma} \sim N(\mathbf{0}_L, \sigma_\gamma^2 \boldsymbol{\Lambda}(\alpha)), \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \qquad (10)$$

where $\mathbf{0}_L$ is an $L \times 1$ vector of zeros, $\mathbf{E}$ is given by the subset of L eigenvectors corresponding to positive eigenvalues, which capture positive spatial dependence [6] (without applying the stepwise variable selection process), and $\mathbf{\Lambda}(\alpha)$ is an $L \times L$ diagonal matrix whose l-th element is $\lambda_l(\alpha) = (\Sigma_l\lambda_l/\Sigma_l\lambda_l^{\alpha})\lambda_l^{\alpha}$, where α and σ_γ^2 are parameters. A large α shrinks the coefficients of the non-principal eigenvectors strongly toward 0, and the resulting $\mathbf{E\gamma}$ describes a global map pattern. By contrast, when α is small, $\mathbf{E\gamma}$ describes a local map pattern. Thus, α controls the spatial smoothness of the underlying map pattern. RE-ESF has two interpretations (Murakami and Griffith, 2015): it describes a map pattern explained by MC (see Section 5.1, for further details), and it describes a Gaussian process after a rank reduction (see, Appendix 1).

Eq. (10) can be rewritten as follows:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{EV}(\mathbf{\theta})\mathbf{u} + \mathbf{\varepsilon}, \qquad \mathbf{u} \sim N(\mathbf{0}_L, \sigma^2\mathbf{I}_L), \qquad \mathbf{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \tag{11}$$

where $\mathbf{\theta} = \{\sigma_\gamma^2/\sigma^2, \alpha\}$, $\mathbf{I}_L$ is an $L \times L$ identity matrix, and $\mathbf{V}(\mathbf{\theta})$ is a diagonal matrix whose

[6] Since **MCM** and $\mathbf{MC}_X\mathbf{M}$ have N - 1 and N - K eigenvectors corresponding to non-zero eigenvectors respectively, keeping all eigenvectors, which drastically consumes degrees of freedom, is not sensible. Many of (RE-)ESF studies consider eigenvectors corresponding to positive eigenvalue because positive spatial dependence is dominant in most social-economic and natural science data (Griffith, 2003; Griffith and Peres-Neto, 2006).

l-th element is $(\sigma_\gamma/\sigma)\lambda_l(\alpha)^{1/2}$. Note that $\mathbf{V}(\boldsymbol{\theta})\mathbf{u}$ in Eq. (11) equals $\boldsymbol{\gamma}$.

The parameters in Eq. (11) (or Eq. (10)) are estimated by using the residual maximum likelihood (REML) method of Bates (2010). Following his specification, the likelihood function is defined by $loglik(\boldsymbol{\beta}, \boldsymbol{\theta}) = \int p(\mathbf{y}, \mathbf{u} \mid \boldsymbol{\beta}, \boldsymbol{\theta}) d\mathbf{u}$, and the restricted log-likelihood by $loglik_R(\boldsymbol{\theta}) = \int loglik(\boldsymbol{\beta}, \boldsymbol{\theta}) d\boldsymbol{\beta}$.

The estimation procedure is summarized as follows: (a) $\boldsymbol{\theta}$ is estimated by maximizing the restricted log-likelihood Eq. (12) with the plugins of Eqs. (13) and (14); (b) $\boldsymbol{\beta}$ and $\boldsymbol{\gamma} = \mathbf{V}(\boldsymbol{\theta})\mathbf{u}$ are estimated by substituting the estimated $\boldsymbol{\theta}$ into Eq. (14); and, (c) σ^2 is estimated by substituting the estimated parameters into Eq. (15). In other words,

$$loglik_R(\boldsymbol{\theta}) = -\frac{1}{2}\log\left|\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{E}\mathbf{V}(\boldsymbol{\theta}) \\ \mathbf{V}(\boldsymbol{\theta})\mathbf{E}'\mathbf{X} & \mathbf{V}(\boldsymbol{\theta})^2 + \mathbf{I}_L \end{bmatrix}\right| - \frac{N-K}{2}\left(1 + \log\left(\frac{2\pi d(\boldsymbol{\theta})}{N-K}\right)\right), \quad (12)$$

$$d(\boldsymbol{\theta}) = \min_{\boldsymbol{\beta}, \mathbf{u}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{E}\mathbf{V}(\boldsymbol{\theta})\mathbf{u}\|^2 + \|\mathbf{u}\|^2, \quad (13)$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{E}\mathbf{V}(\boldsymbol{\theta}) \\ \mathbf{V}(\boldsymbol{\theta})\mathbf{E}'\mathbf{X} & \mathbf{V}(\boldsymbol{\theta})^2 + \mathbf{I}_L \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{V}(\boldsymbol{\theta})\mathbf{E}'\mathbf{y} \end{bmatrix}, \text{ and} \quad (14)$$

$$\hat{\sigma}^2 = \frac{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{E}\mathbf{V}(\boldsymbol{\theta})\mathbf{u}\|^2}{N-K}, \quad (15)$$

where $\|\bullet\|^2$ denotes the L_2-norm of a vector $\bullet$, and $\mathbf{V}(\boldsymbol{\theta})^2 = \mathbf{V}(\boldsymbol{\theta})\mathbf{V}(\boldsymbol{\theta}) = \mathbf{V}(\boldsymbol{\theta})\mathbf{E}'\mathbf{E}\mathbf{V}(\boldsymbol{\theta})$.

Based on Murakami and Griffith (2015), the computational complexity of Eq. (12) is $O((K+L)^3)$, which is smaller than the complexity of the likelihood maximization in standard spatial statistical models ($O(N^3)$). They also reveal that RE-ESF estimates $\mathbf{\beta}$ with smaller estimation error and a shorter computation time than ESF.

Similar models have been used in the statistics literature (e.g., Hughes and Haran, 2013; Johnson et al., 2013; Lee and Barran, 2015). They use $\mathbf{E}$ generated from $\mathbf{M}_X\mathbf{C}\mathbf{M}_X$. This is because this specification eliminates confounders between $\mathbf{X}$ and $\mathbf{E}$ and stabilizes the parameter estimates. However, Murakami and Griffith (2015) show that the elimination leads to biased standard errors of $\mathbf{\beta}$ and recommend using $\mathbf{MCM}$. Section 6 examines which specification is more appropriate for SVC modeling.

4.2. RE-ESF-based SVC models

As with the basic RE-ESF, the RE-ESF-based SVC model is expected to outperform the ESF-based one in terms of estimation accuracy and computational time. Thus, we combine the RE-ESF model and the ESF-based SVC model (Eq. (9)):

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{x}_k \circ \boldsymbol{\beta}_k^{R-ESF} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{16}$$

$$\boldsymbol{\beta}_k^{R-ESF} = \beta_k \mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_k, \qquad \boldsymbol{\gamma}_k \sim N(\mathbf{0}_L, \sigma^2_{k(\gamma)} \boldsymbol{\Lambda}(\alpha_k)),$$

where α_k is a parameter that controls the spatial smoothness of the k-th coefficients, and $\sigma^2_{k(\gamma)}$ controls the variance. The k-th coefficients consist of the fixed constant, $\beta_k \mathbf{1}$, and random spatially varying components, $\mathbf{E}\boldsymbol{\gamma}_k$.

Eq. (16) can be expressed as (see Eqs. (8) and (9))

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{E}}\tilde{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{17}$$

where

$$\tilde{\mathbf{E}} = \left[\mathbf{x}_1 \circ \mathbf{E} \;\; \cdots \;\; \mathbf{x}_K \circ \mathbf{E}\right], \quad \tilde{\boldsymbol{\gamma}} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \vdots \\ \boldsymbol{\gamma}_K \end{bmatrix} \sim N\left[\begin{pmatrix} \mathbf{0}_L \\ \vdots \\ \mathbf{0}_L \end{pmatrix}, \begin{pmatrix} \sigma^2_{1(\gamma)} \boldsymbol{\Lambda}(\alpha_1) & & \\ & \ddots & \\ & & \sigma^2_{K(\gamma)} \boldsymbol{\Lambda}(\alpha_K) \end{pmatrix}\right].$$

Eq. (17) essentially is identical to Eq. (10). Hence, it is further rewritten similar to the rewriting from Eq. (10) to Eq. (11):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{E}}\tilde{\mathbf{V}}(\boldsymbol{\Theta})\tilde{\mathbf{u}} + \boldsymbol{\varepsilon}, \qquad \tilde{\mathbf{u}} \sim N(\mathbf{0}_{LK}, \sigma^2 \mathbf{I}_{LK}), \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{18}$$

$$\tilde{\mathbf{V}}(\boldsymbol{\Theta}) = \begin{bmatrix} \mathbf{V}(\boldsymbol{\theta}_1) & & \\ & \ddots & \\ & & \mathbf{V}(\boldsymbol{\theta}_K) \end{bmatrix}, \qquad \tilde{\mathbf{u}} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_K \end{bmatrix},$$

where $\mathbf{\Theta} = \{\mathbf{\theta}_1, ..., \mathbf{\theta}_K\}$, $\mathbf{\theta}_k = \{\sigma_{k(\gamma)}{}^2/\sigma^2, \alpha_k\}$, $\mathbf{0}_{LN}$ is an $L_K \times 1$ vector of zeros, $\mathbf{I}_{LN}$ is an $L_K \times L_K$ identity matrix, and $\mathbf{V}(\mathbf{\theta}_k)$ is a diagonal matrix whose l-th element is $(\sigma_{\gamma(\gamma)}/\sigma)\lambda_l(\alpha_k)^{1/2}$.

Because Eq. (17) is identical to the RE-ESF model, Eq. (10), the REML estimation for RE-ESF is readily applicable to the proposed model. The estimation procedure is summarized as follows: (a) $\mathbf{\Theta}$ is estimated by maximizing the profile restricted log-likelihood, Eq. (19), with the plugins of Eqs. (20) and (21); (b) $\mathbf{\beta}$ and $\tilde{\gamma} = \tilde{\mathbf{V}}(\mathbf{\Theta})\tilde{\mathbf{u}}$ are estimated by substituting the estimated $\mathbf{\Theta}$ into Eq. (21); and, (c) σ^2 is estimated by substituting the estimated parameters into Eq. (22). In other words,

$$loglik_R(\mathbf{\Theta}) = -\frac{1}{2}\log\left\| \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\tilde{\mathbf{E}}\tilde{\mathbf{V}}(\mathbf{\Theta}) \\ \tilde{\mathbf{V}}(\mathbf{\Theta})\tilde{\mathbf{E}}'\mathbf{X} & \tilde{\mathbf{V}}(\mathbf{\Theta})\tilde{\mathbf{E}}'\tilde{\mathbf{E}}\tilde{\mathbf{V}}(\mathbf{\Theta}) + \mathbf{I}_{LK} \end{bmatrix} \right\| - \frac{N-K}{2}\left(1 + \log\left(\frac{2\pi\tilde{d}(\mathbf{\Theta})}{N-K}\right)\right) \text{ ,(19)}$$

$$\tilde{d}(\mathbf{\Theta}) = \min_{\mathbf{\beta},\tilde{\mathbf{u}}} \; \| \mathbf{y} - \mathbf{X}\mathbf{\beta} - \tilde{\mathbf{E}}\tilde{\mathbf{V}}(\mathbf{\Theta})\tilde{\mathbf{u}} \|^2 + \| \tilde{\mathbf{u}} \|^2, \tag{20}$$

$$\begin{bmatrix} \hat{\mathbf{\beta}} \\ \hat{\tilde{\mathbf{u}}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\tilde{\mathbf{E}}\tilde{\mathbf{V}}(\mathbf{\Theta}) \\ \tilde{\mathbf{V}}(\mathbf{\Theta})\tilde{\mathbf{E}}'\mathbf{X} & \tilde{\mathbf{V}}(\mathbf{\Theta})\tilde{\mathbf{E}}'\tilde{\mathbf{E}}\tilde{\mathbf{V}}(\mathbf{\Theta}) + \mathbf{I}_{LK} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \tilde{\mathbf{V}}(\mathbf{\Theta})\tilde{\mathbf{E}}'\mathbf{y} \end{bmatrix}, \text{ and} \tag{21}$$

$$\hat{\sigma}^2 = \frac{\| \mathbf{y} - \mathbf{X}\mathbf{\beta} - \tilde{\mathbf{E}}\tilde{\mathbf{V}}(\mathbf{\Theta})\tilde{\mathbf{u}} \|^2}{N-K}. \tag{22}$$

Although the REML estimation requires a determinant calculation, computational complexity is only $O((K+KL)^3)$, which can be decreased by reducing the number of eigenvectors in **E**. The computational burden also can be reduced by replacing some $\boldsymbol{\beta}_k^{R-ESF} = \beta_k\mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_k$ with $\beta_k\mathbf{1}$, which means restricting some coefficients to be constants across a given geographic landscape.

The variance-covariance matrices of the coefficients are

$$Cov\begin{bmatrix}\boldsymbol{\beta}\\ \tilde{\boldsymbol{\gamma}}\end{bmatrix} = Cov\begin{bmatrix}\boldsymbol{\beta}\\ \tilde{\mathbf{V}}(\boldsymbol{\Theta})\tilde{\mathbf{u}}_k\end{bmatrix} = \sigma^2\begin{bmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\tilde{\mathbf{E}}\\ \tilde{\mathbf{E}}'\mathbf{X} & \tilde{\mathbf{E}}'\tilde{\mathbf{E}} + \tilde{\mathbf{V}}(\boldsymbol{\Theta})^{-2}\end{bmatrix}^{-1}, \tag{23}$$

where is the inverse of . Because is a diagonal matrix, its inverse is easily calculated. As for $= \beta_k\mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_k$, the variance of the constant component, $\beta_k\mathbf{1}$, is estimated in Eq. (23). The covariance matrix of the spatially varying components, $\mathbf{E}\boldsymbol{\gamma}_k$, is estimated as follows:

$$, \tag{24}$$

where *Cov*[], which is the covariance matrix of $\boldsymbol{\gamma}_k$, is a sub-matrix of *Cov*[] in Eq.(23). The diagonals of $Cov[\mathbf{E}\boldsymbol{\gamma}_k]$ are useful to test if $= \beta_k\mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_k$ has statistically significant spatial variation, whereas the diagonals of $Cov[\boldsymbol{\gamma}_k]$ indicate which eigenvectors are statistically significant.

A problem is how to estimate $\boldsymbol{\Theta}$ efficiently. For example, when five explanatory variables are considered, we need to optimize 10 parameters in $\{\sigma_{1}{}^{2}{}_{(\gamma)}, ..., \sigma_{5}{}^{2}{}_{(\gamma)}, \alpha_1, ..., \alpha_5\}$

simultaneously, which can be computationally expensive. Hence, in addition to simultaneous estimation, we apply an approximation that estimates the coefficient's variance parameters, $\sigma_k{}^2{}_{(\gamma)}$s, first, and the spatial smoothness parameters, α_ks, thereafter. In the first step, we impose $\alpha_k = 1$, which implicitly has been assumed in RE-ESF-type models (e.g., Hughes and Haran, 2013). Assuming a unique value for each α_k, which implies the same degree of spatial smoothness for each coefficient, is another way to increase computational efficiency. Section 6 compares the effectiveness of these simplifications.

5. Properties of RE-ESF-based SVC model

This section clarifies advantages and disadvantages of our SVC model by comparing it with the ESF-based SVC specification (section 5.1), GWR specifications (section 5.2), and the B-SVC model of Gelfand (2003) (section 5.3).

5.1. A comparison with the ESF-based specification

Both the ESF-based model and our model describe their k-th coefficients using $\beta_k \mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_k$. The ESF approach regards $\mathbf{E}\boldsymbol{\gamma}_k$ as fixed effects, whereas ours considers it as random effects, where $\boldsymbol{\gamma}_k \sim N(\mathbf{0}_L, \sigma_{k(\gamma)}^2\boldsymbol{\Lambda}(\alpha_k))$. Our specification has additional variance parameters, $\sigma_{k(\gamma)}^2$ and α_k. They shrink $\mathbf{E}\boldsymbol{\gamma}_k$ strongly toward zero when $\sigma_{k(\gamma)}^2$ is small and α_k is large. Owing to these parameters, our estimator might be more robust to multicollinearity than the estimator of ESF, which is a fundamental problem in SVC models (Wheeler and Tiefelsdorf, 2005).

The parameter α_k also controls the spatial smoothness of each varying coefficient. A large α_k shrinks the coefficients $\gamma_{k,l}$ corresponding to the non-principal eigenvectors strongly toward zero, where $\gamma_{k,l}$ is the l-th element of $\boldsymbol{\gamma}_k$. As a result, $\mathbf{E}\boldsymbol{\gamma}_k$ has a global (smoother) map pattern. Interestingly, α_k is interpretable in terms of MC. $MC[\mathbf{E}\boldsymbol{\gamma}_k]$ can be calculated by substituting $\mathbf{E}\boldsymbol{\gamma}_k$ into Eq. (6) as follows (see Griffith, 2003):

. (25)

$MC[\mathbf{E}\boldsymbol{\gamma}_k]$ is proportional to the average of the L eigenvalues, which are weighted by $\gamma^2_{k,l} = Var[\gamma_{k,l}]$. As α_k grows, the weights $\gamma^2_{k,l}$ on greater eigenvalues are inflated, along with

$MC[\mathbf{E}\boldsymbol{\gamma}_k]$. In particular, $MC[\mathbf{E}\boldsymbol{\gamma}_k]$ takes its maximum value if $\alpha_k = \infty$. By contrast, if $\alpha_k = 0$, $\sigma_{k(\gamma)}{}^2$ shrinks all coefficients equally. In short, α_k is an MC-based shrinkage parameter that intensifies the underlying spatial dependence of $= \beta_k\mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_k$.

Computational efficiency is another advantage of our approach. Unlike the ESF-based SVC model, ours does not require the stepwise variable selection, which can be very slow especially for large datasets.

5.2. A comparison with GWR specifications

A major advantage of our model relative to GWR is its capability of allowing different spatial smoothness of SVCs. GWR studies usually assume the same degree of spatial smoothness for each coefficient, which is unlikely in many real-world situations. Moreover, our approach estimates coefficients based on a global estimation, whereas GWR iterates with local estimations. The global estimation that considers all observations might be more robust than local estimations that consider nearby observations only. Indeed, the efficiency of local estimations depends on the rank sufficiency and collinearity of the (geographically weighted) explanatory variables around each site. Our global estimation is not compromised by such problems.

By contrast, GWR is simpler and easier to extend for non-Gaussian data modeling, spatial interpolation, and other purposes (Fotheringham et al., 2002; Nakaya et al., 2005). Besides, GWR is applicable to a large data set, and can be made faster with parallel computing (Harris et al., 2010), whereas our model is not parallelizable because it requires an eigen-decomposition. Furthermore, GWR approaches are easily implemented (e.g., using the Spatial Statistics Toolbox in ArcGIS (http://www.esri.com/), or spgwr (Bevand et al., 2006), gwrr (Wheeler, 2013), and GWmodel (Gollini et al., 2015) in the R packages). Our model needs to be extended to overcome these disadvantages.

5.3. A comparison with the B-SVC models

5.3.1. Model

The B-SVC model is formulated as follows:

$$, \quad , \qquad (26)$$

, , (27)

where δ^2_k and τ^2_k are variance parameters. Here, **C** is assumed to be known. B-SVC describes both SVCs with [a constant term: $\beta_k\mathbf{1}$] + [a centered Gaussian process, $\mathbf{Me}_k$], and residuals with another Gaussian process.

As described in Appendix 1, $\mathbf{Me}_k \sim N(\mathbf{0}_N, \delta_k^2\mathbf{MCM}+\tau^2_k\mathbf{M})$ can be expanded as follows, after reducing eigen-functions corresponding to [7]:

, , . (28)

Eqs. (27) and (28) indicate that = $\beta_k\mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_k + \boldsymbol{\varepsilon}_k$ (after a rank reduction), whereas our model yields =$\beta_k\mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_k$ (see Eq. (16)). , which does not include $\boldsymbol{\varepsilon}_k$, captures a smoother map pattern than . The difference between and arises because our model is based on the MC, which does not consider variances within each sample, whereas the B-SVC model describes Gaussian processes, which capture within sample variance with δ_k^2 and τ_k^2.

Let us assume that $\mathbf{x}_1$ is a constant. Then, our model, Eq. (16), can be expanded

[7] Here, $\mathbf{MM}' = \mathbf{M}$ is used. It holds because **M** is a symmetric and idempotent matrix.

using Eqs. (27) and (28), as follows:

, .

, .

, , (29)

Thus, our model is a variant of the B-SVC model whose is replaced with , and the Gaussian process, $\mathbf{e}_1$, with a centered Gaussian process, $\mathbf{Me}_1$.

An important distinction between these two models is that ours approximates SVCs with a linear equation, $\mathbf{E}\boldsymbol{\gamma}_k$, whereas the B-SVC model usually does not. The linear specification allows us to apply the computationally efficient REML estimation (see section 5.3.2).

5.3.2. Estimation method

While our model is estimated by the REML method, the B-SVC model must be estimated with MCMC. Because MCMC is robust, even if a sample size is small, the B-SVC model is preferable for small-to-medium size samples. However, MCMC is computationally expensive, particularly when different degrees of spatial smoothness are allowed for each coefficient (Finley, 2011). Therefore, our model is more suitable for medium-to-large size samples. Because our method does not require iterative sampling, unlike MCMC, it is preferable to B-SVC in terms of simplicity, too.

6. Results from a Monte Carlo simulation experiment

This section summarizes a Monte Carlo simulation experiment comparing our model with GWRs and the ESF-based model in terms of SVCs estimation accuracy and computational efficiency.

6.1. Outline

This section compares the conventional GWR, LCR-GWR, and ESF-based SVC models with $\mathbf{M}$ and $\mathbf{M}_X$, respectively (ESF and ESF$_X$), to our RE-ESF-based models with $\mathbf{M}$ and $\mathbf{M}_X$ (RE-ESF and RE-ESF$_X$), respectively. We also compare the following approximations of RE-ESF with $\mathbf{M}$: the RE-ESF that estimates $\sigma^2_{k(\gamma)}$s first and α_ks thereafter (RE-ESF (A1)), and the RE-ESF whose α_ks are assumed to be uniform (RE-ESF (A2)).

The exponential model, Eq. (3), is used to evaluate the geographical weights in the GWR and LCR-GWR. Regarding RE-ESF, a similar exponential model, Eq.(30), is used to evaluate the (i, j)-th element of the proximity matrix **C**, c_i:

(30)

.

Following Dray et al. (2006), the range parameter r is given by the maximum distance in the minimum spanning tree connecting all sample sites. **E** in RE-ESF consists of the eigenvectors corresponding to positive eigenvalues. The same eigenvectors are regarded as candidates to be entered into the ESF model, and they are selected by the adjusted-R^2 based forward variable selection technique. This distance-based ESF often is called Moran’s eigenvector maps, a popular approach in ecology (see, Dray et al., 2006; Griffith and Peres-Neto, 2006; Legendre and Legendre, 2012). Regarding ESF, to cope with multicollinearity, variables with variance inflation factors (VIFs) above 10 are excluded from the candidates in each variable selection step. As for LCR-GWR, following Gollini et al. (2015), the ridge term is introduced only for local models whose condition number exceeds 30.

We generate data using Eq.(31):

, , (31)

, ,

, ,

, .

$\mathbf{x}_1$, whose coefficients take –2 on average, accounts for more of the variation in $\mathbf{y}$, whereas $\mathbf{x}_2$, whose coefficients take 0.5 on average, accounts for less variation.

The covariates in Eq. (31) are generated from Eq. (32):

, (32)

, ,

Eq. (32) assumes that $\mathbf{x}_k$ equals [the centered disturbance, $\mathbf{Me}_{k(ns)}$] + [the centered spatially dependent component, $\mathbf{E}\boldsymbol{\gamma}_0$ (=$\mathbf{ME}\boldsymbol{\gamma}_0$)], whose contribution ratios are 1w_s and w_s, respectively. $\mathbf{x}_k$ has strong spatial dependence when w_s is near 1. Some studies (e.g., Hughes and Haran, 2013) reveal that coefficient estimates tend to be unstable when explanatory variables are spatially dependent. This is because spatially dependent explanatory variables can confound with spatially dependent errors. However, no study has examined the extent to which such spatial confounding influences the spatially varying coefficient estimates. We examine it by varying the intensity of spatial dependence in $\mathbf{x}_k$ with w_s.

Table 1 summarizes DGPs employed in SVC-related simulation studies. This table shows that multicollinearity has been considered. By contrast, spatial confounding has never been analyzed in the context of SVC estimation as far as the authors know. Because we do not know how to control the degrees of multicollinearity and spatial confounding simultaneously, this simulation focuses on only the latter.

[Table 1 around here]

The response variable and covariates are generated on N sample sites whose two geocoded coordinates are given by two random samples from $N(\mathbf{0}, \mathbf{I})$[8]. Then, SVC models are fitted to these variables, and $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are estimated iteratively while varying the sample size N{50, 150, 400}, the ratio of the spatial dependence component in $\mathbf{x}_k$, w_s{0.0, 0.4, 0.8}, and the spatial smoothness of the coefficients: $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$; (α_1, α_2) ={(0.5, 1.0), (1.0, 1.0), (2.0, 1.0)}. In each case, estimations are iterated 200 times.

In addition to the RE-ESF-based data generating process (DGP), which can be too optimistic for our model, a spatial moving average (SMA)-based DGP is also tested. The latter generates data from the SVC model, Eq.(31), whose spatially varying components, $\mathbf{E}\boldsymbol{\gamma}_k$, are replaced with the SMA process, , where $\boldsymbol{\varepsilon}_0 \sim N(\mathbf{0}, \mathbf{I})$ and is a matrix that row-standardizes $\mathbf{I} + \mathbf{C}(r_k)$. Estimations are iterated 200 times while varying N{50, 150, 400}, w_s{0.0, 0.4, 0.8} and (r_1, r_2) ={(0.5, 1.0), (1.0, 1.0), (2.0, 1.0)}. Unlike RE-ESF, which describes a reduced rank spatial process, SMA describes a spatial process without approximation; the SMA-based simulation is needed to examine the coefficient

[8] An assumption of $N(\mathbf{0}, \mathbf{I})$ implies fewer samples near periphery areas. It is likely for many socioeconomic data including land price data, which typically have fewer samples in suburban areas.

estimation accuracy for a non-approximated spatial process. Although we do not discuss it, simulation with GWR-based DGP would be an interesting future topic.

These simulations are performed using R version 3.1.1 (https://cran.r-project.org/) on a 64 bit PC whose memory is 48 GB.

6.2. Results

The estimation accuracy is evaluated by the root mean squared error (RMSE),

$$, \qquad (33)$$

where $\beta_{k,i}$ is the i-th element of the true $\boldsymbol{\beta}_k$, and is the estimate. Tables 2 and 3 summarize the RMSEs in cases of RE-ESF-based DGP and SMA-based DGP, respectively.

[Table 2 around here]

[Table 3 around here]

When SMA is used for data generation, the estimates of RE-ESF models are more accurate than those of GWR and LCR-GWR for a medium-to-large sample size (N = 150 or 400). This tendency is significant if the explanatory variables are spatially dependent (i.e., w_s is large). By contrast, when N = 50, although the RE-ESF is still better than GWR specifications, their gaps are relatively small because RE-ESF relies on an REML estimation, which is less efficient for small samples. On the other hand, if RE-ESF is used for data generation, the estimates of RE-ESF models are more accurate than GWR and LCR-GWR across cases.

Even though use of $\mathbf{M}_X$ is recommended in Hughes and Haran (2013) and Johnson et al. (2013), among others, ESF$_X$ and RE-ESF$_X$ are worse than ESF and RE-ESF, respectively. This is because SVCs estimated by ESF$_X$ and RE-ESF$_X$ are always uncorrelated with (centered) $\mathbf{X}$ even if true SVCs are strongly correlated with $\mathbf{X}$. The result clearly suggests that using models with $\mathbf{M}_X$ is not appropriate for SVC estimation.

Tables 2 and 3 also show the large RMSEs of the ESF coefficients. This may be because ESF does not consider eigenvalues, which act as deflators for coefficients on eigenvectors corresponding to small (in absolute value) eigenvalues in our model.

Among RE-ESF models with **M**, which indicate small RMSEs, the RE-ESF without an approximation and RE-ESF(A1) outperform the opponents in many cases. RE-ESF (A1) would be a good alternative.

$\boldsymbol{\beta}_2$ conveys relatively minor effects. tends to be small in RE-ESF (A2), which assumes constant α_ks, rather than RE-ESF and RE-ESF (A1), which assume non-constant α_ks, especially when SMA-based DGP is assumed. In other words, the estimation of the coefficient smoothness parameters (α_ks) can fail to capture the spatial variation of the SVCs, accounting for a small portion of variations in **y**. Nevertheless, the gaps in their RMSEs are marginal, and their RMSEs are smaller than those of the GWR and LCR-GWR.

$\boldsymbol{\beta}_1$ describes relatively strong impacts. The of RE-ESF and RE-ESF (A1) are smaller than those of RE-ESF (A2). This tendency is substantial when the covariates have strong spatial dependence (i.e., w_s is large). This result suggests that non-uniform smoothness parameters, α_ks, in RE-ESF and RE-ESF (A1) play an important role in appropriately capturing SVCs, accounting for a large portion of variations in **y**.

In each model, RMSE increases in the presence of strong spatial dependence in the covariates, which can confound with spatial dependence in residuals. This result reveals the importance of considering the confounding factor typically ignored in SVC-related studies. Increases in the RMSEs are relatively small in RE-ESF and RE-ESF (A1), including the coefficient smoothness parameter, α_k, which thus might be helpful in mitigating this problem.

We then compare mean bias, which is defined as follows:

$$. \tag{34}$$

Table 4 summarizes mean bias estimated in cases with RE-ESF-based DGP and e_1=2. In each model, mean biases of $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ are small relative to their true mean values (-2 and 0.5). It is verified that estimators of these SVC models are nearly unbiased. While it is suggested that use of $\mathbf{M}_X$ reduces bias in regression coefficients, such a reduction is not conceivable in our result probably because the bias is sufficiently small even if $\mathbf{M}$ is used.

[Table 4 around here]

Finally, Table 5 summarizes average computational times. RE-ESF (A2), RE-ESF (A1), and RE-ESF are the first, second and third fastest, respectively. The computational efficiency of RE-ESF does not hold when either the sample size, N, or the number of SVCs, K, is large because RE-ESF requires optimizing the $2K$ parameters simultaneously. Base on Table 5, RE-ESF is slower than GWR if N 5000. Still, RE-ESF (A1), whose coefficient estimates are as accurate as those for RE-ESF, is faster than GWR. Use of RE-ESF (A1), which allows spatial variation only for several focused coefficients, is a sensible option to reduce computational cost. Note that although ESF involves the computing slowest because of the eigenvector selection step, this step can be replaced with computationally more efficient approaches, such as lasso estimation (Seya et al., 2015).

[Table 5 around here]

7. An application to a land price analysis

This section empirically compares SVC models. Results show that ESF-based and RE-ESF-based SVC models are robust to multicollinearity, and they furnish reasonable SVC estimates for actual data.

7.1. Outline

This section presents an application of GWR, LCR-GWR, the ESF-based SVC model, and the RE-ESF-based SVC model to analyze land price and flood hazard in Ibaraki prefecture, Japan. The western part of Ibaraki was seriously damaged by a river flood in September 2015 (see Figure 1). By December 21, 2015, 54 residences were totally destroyed, 3,752 suffered large-scale partial destruction, and 208 were partially destroyed, while about 10,390 people were in shelters at the peak of the disaster.

[Figure 1 around here]

Our goal here is to assess whether high hazard areas were appropriately recognized as less attractive areas before the flood. To examine this concern, we analyze the relationship between flood hazards and land prices. Specifically, logged officially assessed land prices in 2015 (sample size: 647; see Figure 1 and Table 6) are described using the aforementioned SVC models. The response variables are flood depth (*Flood*), distance to the nearest railway station in km (*Station_D*), and railway distance between the nearest station and Tokyo station (*Tokyo_D*), which is located about 30 km from the southwestern border of Ibaraki. All of these variables measures are available from the National Land Numerical Information download service provided by the Ministry of Land, Infrastructure, Transport and Tourism (http://nlftp.mlit.go.jp/ksj-e/index.html). The VIFs of these variables for an OLS model with all covariates included are 1.09, 1.02 and 1.07, respectively. Thus, serious multicollinearity is not present among them. Since the main objective of this analysis is to compare the SVC models, including GWR approaches, which loses degrees of freedom drastically as the number of explanatory variables increases (Griffith, 2008), we restricted the number of explanatory variables to three.

[Table 6 around here]

This empirical analysis is performed by employing R version 3.1.1 for computation purposes, and ArcGIS version 10.3 (http://www.esri.com/) for visualization. R and ArcGIS were executed on a 64 bit PC whose memory is 48 GB. The 'GWmodel' package in R was used to estimate GWR and LCR-GWR parameters.

7.2. Results

Hereafter, the vector of the spatially varying intercepts is denoted by $\boldsymbol{\beta}_0$ and those of the spatially varying coefficients for *Tokyo_D*, *Station_D*, and *Flood* are denoted by $\boldsymbol{\beta}_{\mathrm{Tk}}$, $\boldsymbol{\beta}_{\mathrm{St}}$, and $\boldsymbol{\beta}_{\mathrm{Fl}}$, respectively.

Table 7 summarizes the variance parameters ($\sigma^2_{k\,(\gamma)}$ and α_k) estimated by RE-ESF. $\sigma^2_{k\,(\gamma)} = 0$ regarding $\boldsymbol{\beta}_{\mathrm{Tk}}$ shows that the impact of *Tokyo_D* is constant across the target area.[9] The positive $\sigma^2_{k(\gamma)}$ values for $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_{\mathrm{St}}$, and $\boldsymbol{\beta}_{\mathrm{Fl}}$ suggest that each has spatial variation.

[Table 7 around here]

The spatial smoothness (or scale) of $\boldsymbol{\beta}_{\mathrm{Fl}}$ is strongly intensified by a large α_k value. In contrast, the spatial smoothness of $\boldsymbol{\beta}_{\mathrm{St}}$, whose α_k equals zero, is not intensified. Although the bandwidths estimated by GWR and LCR-GWR (1.53 km and 2.77 km, respectively) suggest the existence of local spatial variations in each coefficient; based on the α_k values, bandwidths might actually differ across coefficients. More specifically, the bandwidths of $\boldsymbol{\beta}_{\mathrm{St}}$, $\boldsymbol{\beta}_{\mathrm{Fl}}$, and $\boldsymbol{\beta}_{\mathrm{Tk}}$ are likely to be small, moderate and very large,[10] respectively.

Figure 2 displays the boxplots of the estimated coefficients. While the boxplots of $\boldsymbol{\beta}_{\mathrm{St}}$ are similar across the models, the variance of $\boldsymbol{\beta}_{\mathrm{Fl}}$ is inflated in GWR, and those of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_{\mathrm{Tk}}$ are highly inflated in GWR and LCR-GWR. For example, while logged land

[9] The variance becomes zero even when we apply RE-ESF (A1).
[10] The coefficients of GWR are constant when the bandwidth is extremely large.

prices take values between 8.57 and 12.58, $\boldsymbol{\beta}_0$ estimated by GWR ranges between –5.76 and 26.15.

[Figure 2 around here]

The variance inflation might be because GWR and LCR-GWR rely on local estimations. Because *Tokyo_D* has a global map pattern, its variations tend to be small in each local subsample. As a result, GWR might fail to differentiate influences from *Tokyo_D*, with small variations, and intercepts with no variation. Wheeler (2010) also reports a similar problem. Although Fotheringham and Oshan (2016) report the robustness of GWR to multicollinearity, it might not be true when explanatory variables have global map patterns. Because ESF and RE-ESF consider all samples in their estimation, their coefficients are more stable, even if some of the covariates have global patterns. Interestingly, the boxplots of the ESF coefficients are similar to those of the RE-ESF coefficients.

Although the variance of $\boldsymbol{\beta}_{\mathrm{Fl}}$ in GWR also is inflated, it is moderated for LCR-GWR. Effectiveness of the regularized GWR approach is verified. ESF and RE-ESF also provide stable coefficient estimates.

Table 8 summarizes correlation coefficients among SVCs. $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_{\mathrm{Tk}}$ have strong negative correlations for the GWR and LCR-GWR. The greater variations of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_{\mathrm{Tk}}$, portrayed in Figure 2, are attributable to their multicollinearity. By contrast, correlation coefficients for the ESF and RE-ESF models are reasonably small, and no serious multicollinearity was found. The result is consistent with a suggestion by Griffith (2008) that the ESF-based specification is robust to multicollinearity.

[Table 8 around here]

Figure 3 plots the estimated coefficients. In each model, the estimated $\boldsymbol{\beta}_0$ demonstrates greater land prices in the nearby Tokyo area and around Mito city, which is the prefectural capital. The spatial distributions of $\boldsymbol{\beta}_{\mathrm{St}}$ suggest that land prices decline rapidly as distance to the nearest station increases in nearby station areas, whereas this reduction is moderated in suburban areas. The estimated $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_{\mathrm{St}}$ are similar across models.

[Figure 3 around here]

Consistent with the expected negative sign of $\boldsymbol{\beta}_{\mathrm{Tk}}$, 643/648 of its elements for ESF, and all of its elements for RE-ESF are negative. In contrast, 465/648 and 10/648 elements are positive in the GWR and LCR-GWR, respectively, probably because of the variance inflation previously discussed. Another notable difference is that RE-ESF $\boldsymbol{\beta}_{\mathrm{Tk}}$ estimates have no spatial variation (i.e., $\sigma^2_{k\,(\gamma)} = 0$), whereas the other $\boldsymbol{\beta}_{\mathrm{Tk}}$ estimates that have significant spatial variation.

The elements of $\boldsymbol{\beta}_{\mathrm{Fl}}$ are negative if flood-prone areas have lower land prices. $\boldsymbol{\beta}_{\mathrm{Fl}}$ obtained from RE-ESF displays a smoother map pattern than for the other models because of the large α_k value (3.02). The $\boldsymbol{\beta}_{\mathrm{Fl}}$ for RE-ESF is negative around Mito, where high hazard areas are appropriately recognized as less attractive. In contrast, $\boldsymbol{\beta}_{\mathrm{Fl}}$ is positive in the western area, including the area flooded in September 2015. In other words, high hazard areas are recognized as attractive areas. This result implies that benefits of rivers (e.g., natural environment, landscape) are emphasized more than flood hazard. This situation may have increased the resulting damage from the 2015 flood. In contrast, the $\boldsymbol{\beta}_{\mathrm{Fl}}$ estimated by the other models takes both positive and negative values in the flooded area.

Finally, Table 9 summarizes the computational times. For reference, the computational time of RE-ESF (A1) also is calculated and included. This table shows that RE-ESF is computationally more efficient than LCR-GWR and ESF in the case of N = 647 and three covariates. Furthermore, computation of estimates for RE-ESF (A1) is more than three times faster than for GWR in this case. Of note is that GWR calculations are faster than RE-ESF(A1) if sample size is large. This timing difference is because of the requirement of an eigen-decomposition.

[Table 9 around here]

In summary, we empirically verified that each SVC model can provide different results, and that the estimates of RE-ESF seem reliable (i.e., interpretable and displaying smaller variance).

8. Concluding remarks

This study proposes an RE-ESF-based SVC model whose coefficients are interpretable based on the MC. A simulation analysis and an empirical analysis involving land prices suggest advantages of our model in terms of estimation accuracy, computational time, and interpretability of coefficient estimates.

Unlike GWR models and the typical B-SVC model, RE-ESF estimates the smoothness of each SVC in a computationally efficient manner. Although coefficient smoothness parameters also can be introduced into the B-SVC model, their estimation is computationally prohibitive. Thus, our approach is useful as a flexible and relatively simple procedure. Meanwhile, computationally efficient and flexible alternatives, including the integrated nested Laplace approximation (INLA: Rue et al., 2009)-based SVC model (Congdon, 2014), have been proposed recently. Therefore, the comparison of our model with these is therefore an important future research topic.

Another remaining issue is to compare our model with other SVC models from the viewpoint of statistical inference for the $\boldsymbol{\beta}_k$s. We also need to examine the validity of our model in cases with many covariates for which multicollinearity among SVCs can be serious. Furthermore, extension of our model to a wide range of applications would be an interesting next step. These extensions might include change of support problems (e.g., Murakami and Tsutsumi, 2012; 2015), interaction data modeling (Chun and Griffith, 2011), non-Gaussian data modeling (Fotheringham et al., 2002; Nakaya et al., 2005; Griffith, 2011), multilevel modeling (e.g., Dong et al., 2016), spatiotemporal data modeling (Fotheringham et al., 2015; Huang et al., 2010; Griffith, 2012).

Acknowledgements

This work was supported by Grants-in-Aid for Scientific Research from JSPS (Research Project Number: 15H04054). We thank Professor Morito Tsutsumi (University of Tsukuba) and Professor Takahisa Yokoi (Tohoku University) for their suggestive advices.

Appendix 1: Relationship between the RE-ESF model and the geostatistical model.

The standard Gaussian process model is formulated as follows:

, , (A1)

where $\mathbf{X}_{-1}$ is a $(K\text{-}1) \times N$ matrix of explanatory variables without intercept term (i.e., $\mathbf{X} = [\mathbf{1}, \mathbf{X}_{-1}]$), $\boldsymbol{\beta}_{-1}$ is a $(K\text{-}1) \times 1$ vector of regression coefficients, and β_0 is a parameter. **e** can be expanded as follows:

, , (A2)

where is the mean of **e**. Murakami and Griffith (2015) reveal~~s~~ the following relationship:

, (A3)

where $\mathbf{I}(\lambda_l \neq 0)$ is a $N \times N$ diagonal matrix whose l-th entry is 1 if $\lambda_l \neq 0$, and 0 otherwise. Eq.(A3) becomes $\mathbf{E}(\delta^2\boldsymbol{\Lambda}+\tau^2\mathbf{I}_L)\mathbf{E}' = \delta^2\mathbf{E}\boldsymbol{\Lambda}\mathbf{E}' + \tau^2\mathbf{I}$ after reducing eigen-functions corresponding to . Thus, **Me** with the rank reduction, $\mathbf{Me}_{red}$, behaves as

, (A4)

which equals

, , . (A5)

By substituting Eq.(A5) into **Me** in Eq.(A2), Eq.(A1) yields

,

, (A6)

where $\boldsymbol{\beta} = [\beta_0 +, \boldsymbol{\beta}_{-1}']'$. Thus, our model, which is identical with Eq.(A6), is a low rank approximation of Eq.(A1). Similar discussion holds even if $\mathbf{M}_X$ is used (see, Murakami and Griffith, 2015).

References

1) Anselin L (1988) Spatial Econometrics, Methods and Models. Kluwer Academic, Dordrecht.

2) Anselin L and Rey S (1991) Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23(2): 112–131.

3) Assunçao, R. M. (2003). Space varying coefficient models for small area data. Environmetrics, 14(5), 453-473.

4) Austin M (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modeling* 200(1–2): 1–19.

5) Bates DM (2010) *lme4: Mixed-effects modeling with R*. http://lme4.r-forge.r-project.org/book.

6) Bivand R., Yu D, Nakaya T and Garcia-Lopez M-A (2006) Package ‘spgwr’. https://cran.r-project.org/web/packages/spgwr/spgwr.pdf.

7) Bitter C, Mulligan GF and Dall'erba S (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.

8) Blangiardo, M., & Cameletti, M. (2015). Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley & Sons.

9) Chun Y (2014) Analyzing space-time crime incidents using eigenvector spatial filtering: an application to vehicle burglary. *Geographical Analysis* 46(2): 165–184.

10) Chun Y and Griffith DA (2011) Modeling network autocorrelation in space-time migration flow data: an eigenvector spatial filtering approach. *Annals of the Association of American Geographers* 101(3): 523–536.

11) Chun Y and Griffith DA (2014) A quality assessment of eigenvector spatial filtering based parameter estimates for the normal probability model. *Spatial Statistics*, 10, 1–11.

12) Chun Y, Griffith DA, Lee M and Sinha P (2016) Eigenvector selection with stepwise regression techniques to construct eigenvector spatial filters. *Journal of Geographical Systems*. 18 (1): 67–85.

13) Congdon P (2014) *Applied Bayesian modelling*. Chichester, UK: John Wiley & Sons.

14) Cressie N and Wikle CK (2011) *Statistics for spatio-temporal data*. New York: John Wiley & Sons.

15) Dong G, Ma J, Harris R and Pryce G (2016) Spatial random slope multilevel modeling using multivariate conditional autoregressive models: A case study of subjective travel satisfaction in Beijing. *Annals of the American Association of Geographers* 106(1): 19–35.

16) Dray S, Legendre P and Peres–Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modeling* 196(3–4): 483–493.

17) Fan J (1993) Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, 21(1): 196–216.

18) Farber S and Páez A (2007) A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems*, 9(4): 371–396.

19) Finley AO (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): 143–154.

20) Fotheringham AS, Brunsdon C and Charlton M (2002) *Geographically weighted regression: the analysis of spatially varying relationships*. Chichester, UK: John Wiley & Sons.

21) Fotheringham AS, Crespo R and Yao J (2015) Geographical and Temporal Weighted Regression (GTWR). *Geographical Analysis* 47(4): 431–452.

22) Fotheringham AS and Oshan TM (2016) Geographically weighted regression and multicollinearity: Dispelling the myth. *Journal of Geographical Systems* 18(4): 303–329.

23) Gamerman D, Moreira AR, and Rue H (2003) Space-varying regression models: specifications and simulation. *Computational Statistics and Data Analysis*, 42(3): 513–533.

24) Geary RC (1954) The contiguity ratio and statistical mapping. *The incorporated*

statistician 5(3): 115–146.

25) Gelfand AE, Kim HJ, Sirmans CF and Banerjee S (2003) Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association* 98(462): 378–396.

26) Gollini I, Lu B, Charlton M, Brunsdon C and Harris P (2015) GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software* 63: 17.

27) Gomez-Rubio V, Bivand RS and Rue H (2014) Spatial models using Laplace approximation methods. In: Fischer MM and Nijkamp P (eds). *Handbook of Regional Science*. Berlin, Springer.

28) Griffith DA (2003) *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. Berlin: Springer.

29) Griffith DA (2008) Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A* 40(11): 2751–2769.

30) Griffith DA (2011) Positive spatial autocorrelation impacts on attribute variable frequency distributions. *Chilean Journal of Statistics* 2(2): 3–28.

31) Griffith DA (2012) Space, time, and space-time eigenvector filter specifications that account for autocorrelation. *Estadística española* 54(177): 7–34.

32) Griffith DA and Chun Y (2014) Spatial autocorrelation and spatial filtering. In: Fischer MM and Nijkamp P (eds) *Handbook of Regional Science*. Berlin: Springer, pp.1435–1459.

33) Griffith DA and Chun Y (2016) Evaluating eigenvector spatial filter corrections for omitted georeferenced variables. *Econometrics*, 4(2), 29.

34) Griffith DA and Paelinck JHP (2011) *Non–standard spatial statistics and spatial econometrics*. Berlin: Springer.

35) Griffith DA and Peres-Neto PR (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses in exploiting relative location information. *Ecology* 87(10): 2603–2613.

36) Harris R, Singleton A, Grose D, Brunsdon C and Longley P (2010) Grid-enabling

geographically weighted regression: A case study of participation in higher education in England. *Transactions in GIS*. 14 (1): 43–61.

37) Helbich, M and Griffith DA (2016) Spatially varying coefficient models in real estate: Eigenvector spatial filtering and alternative approaches. *Computers, Environment and Urban Systems*, 57, 1–11.

38) Hu M, Li Z, Wang J, Jia L, Lian Y et al. (2012) Determinants of the incidence of hand, foot and mouth disease in China using geographically weighted regression models. *Plos One* 7(6): e38978.

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

40) Hughes J and Haran M (2013) Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75(1): 139–159.

41) Johnson DS, Conn PB, Hooten MB, Ray JC and Pond BA (2013) Spatial occupancy

models for large data sets. *Ecology*. 94 (4): 801–808.

42) Lee D and Sarran C (2015) Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. *Environmetrics*. 26 (7): 477–487.

43) Legendre P and Legendre L (2012) *Numerical Ecology (Third Edition)*. Amsterdam: Elsevier.

44) Lu B, Harris P, Charlton M and Brunsdon C (2015) Calibrating a geographically weighted regression model with parameter-specific distance metrics. *Procedia Environmental Sciences*, 26, 109–114.

45) Margaretic P, Thomas–Agnan C, Doucet R and Villotta Q (2015) Spatial dependence in (origin–destination) air passenger flows. *Papers in Regional Science*. DOI: 10.1111/pirs.12189.

46) Murakami D and Griffith DA (2015) Random effects specifications in eigenvector spatial filtering: a simulation study. *Journal of Geographical Systems* 17(4): 311–331.

47) Murakami D and Tsutsumi M (2012) Practical spatial statistics for areal interpolation.

Environment and Planning B: Planning and Design, 39(6): 1016–1033.

48) Murakami D and Tsutsumi M (2015) Area-to-point parameter estimation with geographically weighted regression. *Journal of Geographical Systems* 17(3): 207–225.

49) Nakaya T, Fotheringham AS, Charlton M and Brunsdon C (2005) Geographically weighted Poisson regression for disease associative mapping. *Statistics in Medicine* 24(17): 2695–2717.

50) Páez A, Farber S and Wheeler DC (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.

51) Rue H, Martino S and Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2): 319–392.

52) Stone CJ (1980) Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, 8(6): 1348–1360.

53) Seya H, Murakami D, Tsutsumi M and Yamagata Y (2015) Application of LASSO to the eigenvector selection problem in eigenvector based spatial filtering. *Geographical Analysis* 47(3): 284–299.

54) Tiefelsdorf M and Griffith DA (2007) Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A* 39(5): 1193.

55) Wang Q, Ni J and Tenhunen J (2005) Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Global Ecology and Biogeography* 14(4): 379–393.

56) Wheeler DC (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A* 39(10): 2464–2481.

57) Wheeler DC (2009) Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environment and Planning A* 41(3): 722–742.

58) Wheeler DC (2010) Visualizing and diagnosing coefficients from geographically

weighted regression. In: Jiang B and Yao X (eds) *Geospatial Analysis and Modeling of Urban Structure and Dynamics*, Springer.

59) Wheeler DC and Calder CA (2007) An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems* 9(2): 145–166.

60) Wheeler DC and Páez A (2009). Geographically Weighted Regression. In: Fischer MM and Getis A (eds) *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Berlin: Springer, pp.461–484.

61) Wheeler DC and Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems* 7(2): 161–187.

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

63) Wheeler DC (2013) Package ‘gwrr’. https://cran.r-

project.org/web/packages/gwrr/gwrr.pdf.

64) Yang W, Fotheringham AS and Harris P (2014) An extension of geographically weighted regression with flexible bandwidths. Proceedings of GISRUK 20th Annual Conference.

Table 1. Summary of SVC-related simulation studies

<table>
<tr><th>Study</th><th>Model for SVC generation</th><th>Spatial dependence in X</th><th>Multi-collinearity in X</th><th>Model</th></tr>
<tr><td>Farber and Paez (2007)</td><td>Trend surface</td><td rowspan="5"></td><td></td><td>GWR</td></tr>
<tr><td>Wheeler and Calder (2007)</td><td>Gaussian process and trend surface</td><td>×</td><td rowspan="2">GWR and B-SVC</td></tr>
<tr><td>Finley et al. (2009)</td><td>Gaussian process</td><td></td></tr>
<tr><td>Paez et al. (2010)</td><td>Spatial eigenvector (e_l)</td><td rowspan="2">×</td><td rowspan="2">GWR</td></tr>
<tr><td>Fotheringham and Oshan (2016)</td><td>SMA with white noise</td></tr>
<tr><td>Our study</td><td>RE-ESF ($E\gamma$) and SMA</td><td>×</td><td></td><td>GWR, ESF, and RE-ESF</td></tr>
</table>

Table 2. RMSEs of the estimated coefficients (DGP: RE-ESF)

N	Coef.	r_1	w_s	GWR	LCR-GWR	ESF	RE-ESF	RE-ESF (A1)	RE-ESF (A2)	ESF_X	RE-ESF_X
50	$\boldsymbol{\beta}_1$	0.5	0.0	1.70	1.69	1.43	**1.16**	**1.15**	1.58	1.62	1.41
			0.4	1.99	1.97	2.01	**1.51**	**1.52**	2.11	2.28	1.90
			0.8	2.46	2.32	2.70	**1.77**	**1.76**	2.47	2.87	2.35
		2.0	0.0	1.33	1.34	1.31	**0.89**	**0.90**	1.53	1.51	1.11
			0.4	1.62	1.63	1.82	**1.19**	**1.20**	1.87	2.09	1.64
			0.8	2.26	2.18	2.62	**1.65**	**1.66**	2.36	3.03	2.36
	$\boldsymbol{\beta}_2$	0.5	0.0	1.15	1.11	1.34	0.94	**0.92**	**0.91**	1.37	0.94
			0.4	1.30	1.28	1.91	1.14	**1.13**	**1.10**	1.98	1.22
			0.8	2.00	1.86	2.46	**1.29**	**1.32**	1.34	2.42	1.54
		2.0	0.0	0.97	0.94	1.24	0.82	**0.81**	**0.81**	1.29	0.87
			0.4	1.34	1.32	1.88	1.09	**1.07**	**1.06**	1.92	1.18
			0.8	1.74	1.62	2.31	**1.28**	1.28	**1.23**	2.32	1.51
150	$\boldsymbol{\beta}_1$	0.5	0.0	1.36	1.37	1.08	**0.86**	**0.87**	1.33	1.15	0.97
			0.4	1.64	1.63	1.55	**1.17**	**1.18**	1.72	1.74	1.41
			0.8	2.10	2.07	2.38	**1.45**	**1.45**	2.21	2.55	1.86
		2.0	0.0	1.02	1.02	0.91	**0.62**	**0.61**	0.86	1.01	0.71
			0.4	1.25	1.25	1.42	**0.90**	**0.89**	1.19	1.55	1.11
			0.8	1.56	1.55	2.13	**1.08**	**1.08**	1.41	2.37	1.58
	$\boldsymbol{\beta}_2$	0.5	0.0	0.91	0.87	1.02	**0.61**	**0.61**	0.65	1.03	0.65
			0.4	1.11	1.08	1.42	0.82	**0.81**	**0.82**	1.49	0.89
			0.8	1.65	1.58	2.12	**1.00**	**0.99**	1.08	2.13	1.13
		2.0	0.0	0.78	0.77	0.96	**0.63**	0.63	**0.62**	1.00	0.65
			0.4	0.96	0.95	1.39	**0.79**	0.79	**0.76**	1.43	0.83
			0.8	1.30	1.27	2.03	**0.93**	**0.93**	0.94	2.00	1.10
400	$\boldsymbol{\beta}_1$	0.5	0.0	1.22	1.22	0.85	**0.72**	**0.72**	1.20	0.88	0.76
			0.4	1.47	1.46	1.27	**1.00**	**1.00**	1.55	1.38	1.13
			0.8	1.85	1.79	1.94	**1.20**	**1.20**	1.75	2.10	1.48
		2.0	0.0	0.80	0.79	0.77	**0.48**	**0.48**	0.54	0.81	0.53
			0.4	0.98	0.97	1.16	**0.65**	**0.65**	0.76	1.23	0.79
			0.8	1.23	1.21	1.71	**0.78**	**0.78**	0.92	1.87	1.11
	$\boldsymbol{\beta}_2$	0.5	0.0	0.83	0.79	0.81	**0.51**	**0.51**	0.54	0.82	0.52
			0.4	0.99	0.97	1.16	**0.65**	**0.65**	0.68	1.18	0.69
			0.8	1.46	1.35	1.73	**0.77**	**0.77**	0.82	1.79	0.88
		2.0	0.0	0.64	0.63	0.78	**0.48**	**0.48**	0.48	0.80	0.49
			0.4	0.79	0.79	1.15	**0.63**	0.63	**0.62**	1.16	0.66
			0.8	1.03	1.00	1.64	**0.73**	**0.73**	0.71	1.72	0.81

Note: w_s intensifies the spatial dependence in $\mathbf{X}$; r_1 determines the spatial scale of $\boldsymbol{\beta}_1$. Dark gray denotes the minimum RMSE in each case, and light gray denotes the second minimum RMSE.

Table 3. RMSEs of the estimated coefficients (DGP: SMA)

N	Coef.	r_1	w_s	GWR	LCR-GWR	ESF	RE-ESF	RE-ESF (A1)	RE-ESF (A2)	ESF$_X$	RE-ESF$_X$
50	$\boldsymbol{\beta}_1$	0.5	0.0	2.32	2.32	2.65	**2.23**	**2.25**	2.58	2.73	2.37
			0.4	2.45	2.45	3.12	**2.40**	**2.40**	2.70	3.20	2.74
			0.8	2.83	2.78	4.06	**2.63**	**2.64**	3.05	3.83	2.96
		2.0	0.0	**2.43**	**2.44**	2.87	2.44	2.46	2.66	2.84	2.48
			0.4	**2.56**	**2.56**	3.35	2.59	2.60	2.80	3.34	2.77
			0.8	2.71	**2.68**	4.07	**2.68**	2.71	2.96	3.91	3.11
	$\boldsymbol{\beta}_2$	0.5	0.0	1.13	1.12	1.92	**1.11**	1.14	**1.12**	1.80	1.12
			0.4	1.25	1.24	2.49	**1.23**	1.26	**1.18**	2.25	1.32
			0.8	1.57	1.53	3.42	**1.48**	1.52	**1.42**	3.00	1.56
		2.0	0.0	**1.09**	**1.09**	1.94	1.12	1.14	1.12	1.84	1.14
			0.4	**1.22**	1.22	2.58	1.26	1.29	**1.18**	2.64	1.36
			0.8	1.40	**1.36**	3.35	1.37	1.39	**1.30**	2.66	1.42
150	$\boldsymbol{\beta}_1$	0.5	0.0	1.87	1.87	2.00	**1.77**	1.78	1.99	1.98	**1.75**
			0.4	2.04	2.04	2.40	**1.92**	**1.93**	2.23	2.62	2.15
			0.8	2.44	2.41	3.63	**2.11**	**2.13**	2.75	3.69	2.69
		2.0	0.0	**1.75**	1.76	2.05	**1.75**	1.76	1.91	2.08	1.82
			0.4	1.85	1.85	2.28	**1.81**	**1.81**	2.03	2.54	2.06
			0.8	2.15	2.15	3.33	**1.98**	**2.00**	2.51	3.49	2.75
	$\boldsymbol{\beta}_2$	0.5	0.0	0.88	0.87	1.33	**0.80**	0.81	**0.78**	1.35	0.82
			0.4	1.05	1.04	1.84	**0.95**	0.96	**0.91**	1.90	1.03
			0.8	1.49	1.44	3.24	**1.20**	**1.23**	1.24	3.07	1.51
		2.0	0.0	0.81	0.81	1.29	0.80	0.80	**0.79**	1.34	**0.79**
			0.4	0.94	0.93	1.70	**0.93**	0.93	**0.88**	1.84	1.05
			0.8	1.27	1.27	2.86	**1.10**	**1.12**	1.16	2.55	1.34
400	$\boldsymbol{\beta}_1$	0.5	0.0	1.44	1.45	1.53	**1.36**	**1.36**	1.55	1.54	1.39
			0.4	1.63	1.61	1.85	**1.49**	**1.49**	1.76	2.09	1.75
			0.8	2.25	2.15	3.21	**1.74**	**1.75**	2.31	3.29	2.35
		2.0	0.0	1.23	1.24	1.40	**1.20**	**1.19**	1.25	1.44	1.21
			0.4	1.37	1.37	1.71	**1.31**	**1.31**	1.42	2.12	1.71
			0.8	1.72	1.71	2.63	**1.46**	**1.46**	1.90	3.16	2.45
	$\boldsymbol{\beta}_2$	0.5	0.0	0.75	0.74	1.01	**0.58**	**0.59**	0.59	1.06	0.62
			0.4	0.92	0.90	1.40	**0.71**	0.71	**0.68**	1.50	0.84
			0.8	1.51	1.43	2.84	**0.99**	**0.99**	1.03	2.55	1.22
		2.0	0.0	0.61	0.61	0.93	**0.55**	0.56	**0.55**	0.96	0.57
			0.4	0.79	0.79	1.32	**0.69**	0.70	**0.67**	1.54	0.86
			0.8	1.28	1.26	2.41	**0.92**	**0.93**	0.95	2.35	1.21

Note: See Table 2.

Table 4. Bias of the estimated coefficients (r_1 = 2; DGP: SMA)

N	Coef.	w_s	GWR	LCR-GWR	ESF	RE-ESF	RE-ESF (A1)	RE-ESF (A2)	ESF_X	$RE\text{-}ESF_X$
50	$\boldsymbol{\beta}_1$	0.0	0.03	0.06	**0.02**	0.04	0.04	**0.02**	0.06	0.05
		0.4	-0.02	**0.01**	0.04	0.04	0.05	**0.00**	0.09	0.08
		0.8	0.11	0.21	0.05	0.06	**0.04**	**0.01**	0.27	0.22
	$\boldsymbol{\beta}_2$	0.0	-0.03	-0.04	0.01	**0.00**	**0.01**	-0.04	0.02	0.00
		0.4	**0.04**	**0.03**	0.07	0.05	0.05	0.04	0.09	0.09
		0.8	0.02	-0.01	0.01	-0.02	**0.01**	0.07	**0.01**	-0.07
150	$\boldsymbol{\beta}_1$	0.0	0.01	0.04	0.00	**0.00**	**0.00**	0.00	-0.01	0.00
		0.4	-0.01	**0.00**	-0.02	**0.00**	0.01	0.03	0.01	0.02
		0.8	-0.01	0.02	-0.07	**0.01**	**0.00**	0.04	-0.02	0.04
	$\boldsymbol{\beta}_2$	0.0	**0.01**	0.01	0.01	0.01	0.01	0.01	**0.01**	0.01
		0.4	-0.04	-0.04	-0.05	-0.03	**-0.02**	**-0.02**	-0.05	-0.02
		0.8	**0.04**	0.04	0.08	0.07	0.07	**0.04**	0.06	0.06
400	$\boldsymbol{\beta}_1$	0.0	**-0.01**	**0.00**	-0.02	-0.02	-0.02	-0.02	-0.03	-0.02
		0.4	-0.06	-0.05	**-0.04**	**-0.04**	-0.04	-0.04	-0.05	-0.05
		0.8	-0.06	-0.04	**0.00**	-0.03	**-0.03**	-0.04	-0.03	-0.04
	$\boldsymbol{\beta}_2$	0.0	-0.01	-0.01	0.00	**0.00**	0.00	0.00	0.00	**0.00**
		0.4	**0.00**	-0.01	**0.00**	0.00	-0.01	0.00	-0.02	-0.01
		0.8	-0.05	-0.06	0.05	-0.04	**-0.03**	-0.05	**-0.01**	-0.04

Note: w_s intensifies the spatial dependence in **X**. Dark gray denotes the minimum bias in each case and light gray denotes the second minimum bias.

Table 5. Mean computational time in seconds (DGP: RE-ESF; r_1 = 2; r_2 = 1).

N	GWR	LCR-GWR	ESF	RE-ESF	RE-ESF (A1)	RE-ESF (A2)
50	0.30	0.80	0.82	0.20	0.10	0.10
150	2.05	3.88	5.15	0.79	0.39	0.35
400	13.18	20.40	32.39	5.04	2.49	2.12
1,000	79.00	115.21	275.48	24.40	15.81	12.78
2,000	326.42	495.61	2056.01	75.33	122.06	65.96
5,000	1984.99	3465.37	56324.66	2241.90	1110.97	883.26

Note: Because computational times were very similar across iterations, regarding cases with *N* = {1,000, 2,000, 5000}, we performed five replicates, and averaged the resulting five computational times.

Table 6. Summary statistics for land prices (100 JPY/m^2).

Statistics	Value
Mean	35.68
Median	29.50
Standard error	27.14
Maximum	290.00
Minimum	5.28
Sample size	647

Table 7. Estimates of the variance parameters in RE-ESF: $\sigma^2_{k\,(\gamma)}$ controls the variance of each coefficient, and α_k controls the spatial scale of their variations.

	$\boldsymbol{\beta}_0$	$\boldsymbol{\beta}_{\mathrm{Tk}}$	$\boldsymbol{\beta}_{\mathrm{St}}$	$\boldsymbol{\beta}_{\mathrm{Fl}}$
$\sigma^2_{k\,(\gamma)}$	1.71	0.00	0.35	0.61
α_k	0.27	N.A.[1)]	0.00	3.02

[1)] Because $\boldsymbol{\beta}_{\mathrm{St}}$ lacks spatial variation (i.e., $\sigma^2_{k\,(\gamma)}$ = 0.00), α_k for $\boldsymbol{\beta}_{\mathrm{St}}$ is undefined.

Table 8. Correlation coefficients among SVCs.

GWR

	β_0	β_{Tk}	β_{St}	β_{Fl}
β_0		-0.87	-0.27	0.03
β_{Tk}			0.23	-0.03
β_{St}				-0.08

LCR-GWR

	β_0	β_{Tk}	β_{St}	β_{Fl}
β_0		-0.74	-0.45	-0.18
β_{Tk}			0.24	0.26
β_{St}				0.34

ESF

	β_0	β_{Tk}	β_{St}	β_{Fl}
β_0		0.01	-0.44	0.08
β_{Tk}			-0.24	-0.23
β_{St}				-0.01

RE-ESF

	β_0	β_{Tk}	β_{St}	β_{Fl}
β_0		NA [1)]	-0.41	-0.29
β_{Tk}			NA	NA
β_{St}				-0.10

[1)] Regarding RE-ESF, correlation coefficients between β_{Tk} and other coefficients cannot be calculated because it lacks spatial variations (i.e., $\sigma^2_{k\,(\gamma)} = 0$).

Table 9. Computational time in seconds.

GWR	LCR-GWR	ESF	RE-ESF	RE-ESF (A1)
36.2	62.3	107	51.7	11.8

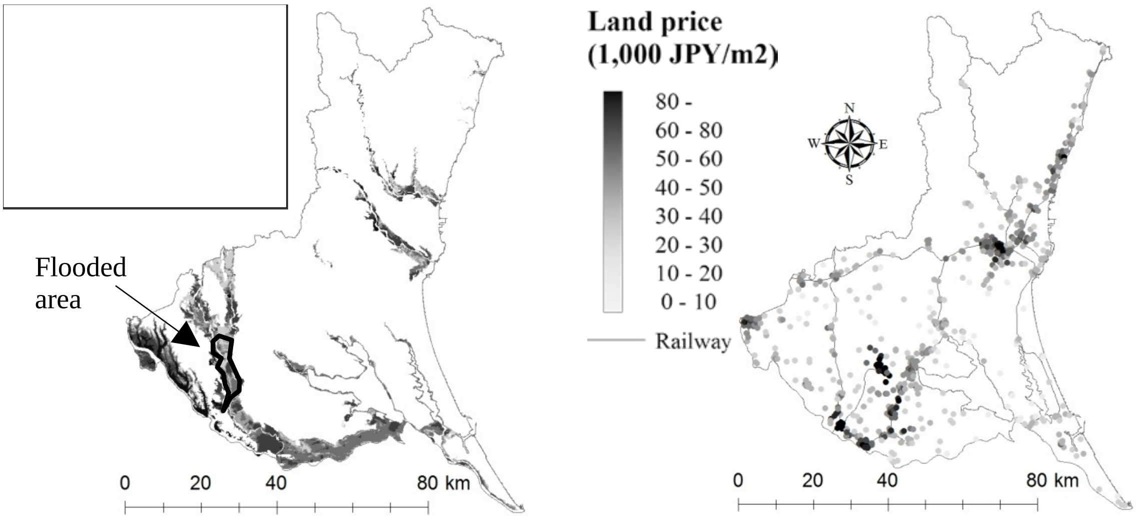

Figure 1. Anticipated flood depth (left) and officially assessed land prices in 2015 (right) in the Ibaraki prefecture.

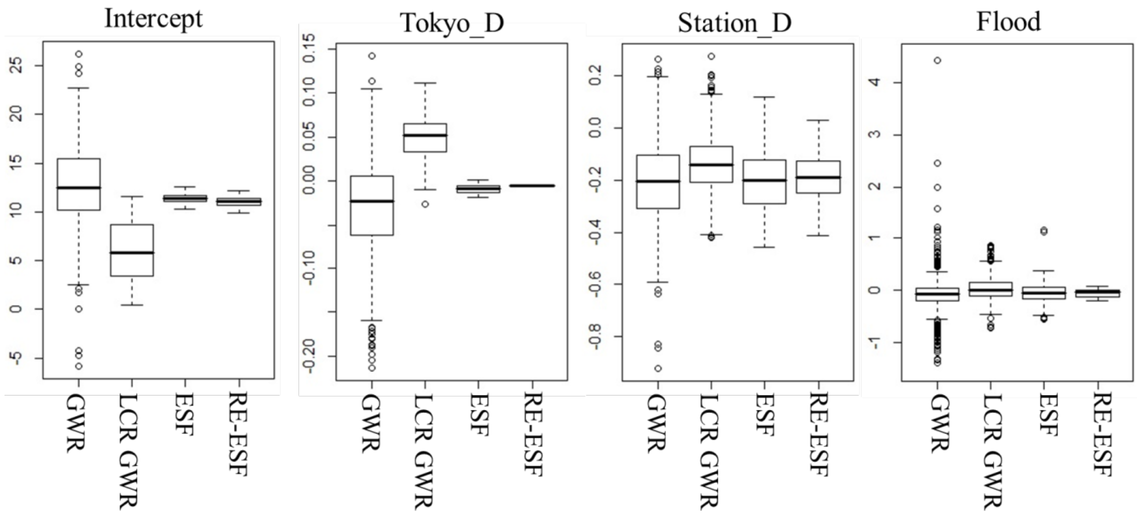

Figure 2. Boxplots of the estimated spatially varying coefficients.

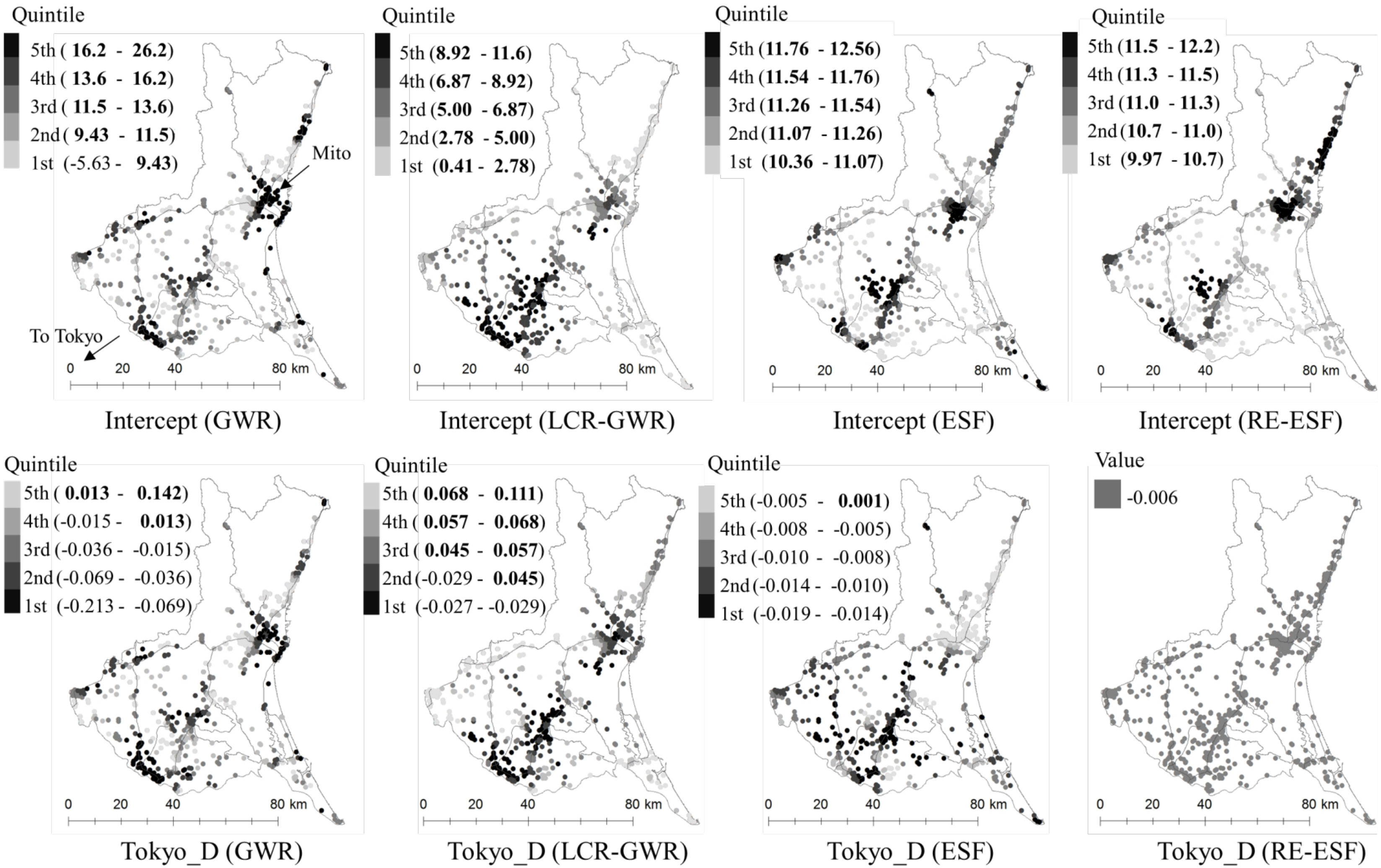

Figure 3. Spatial plots of the estimated coefficients. In each legend, positive values are denoted by bold text.

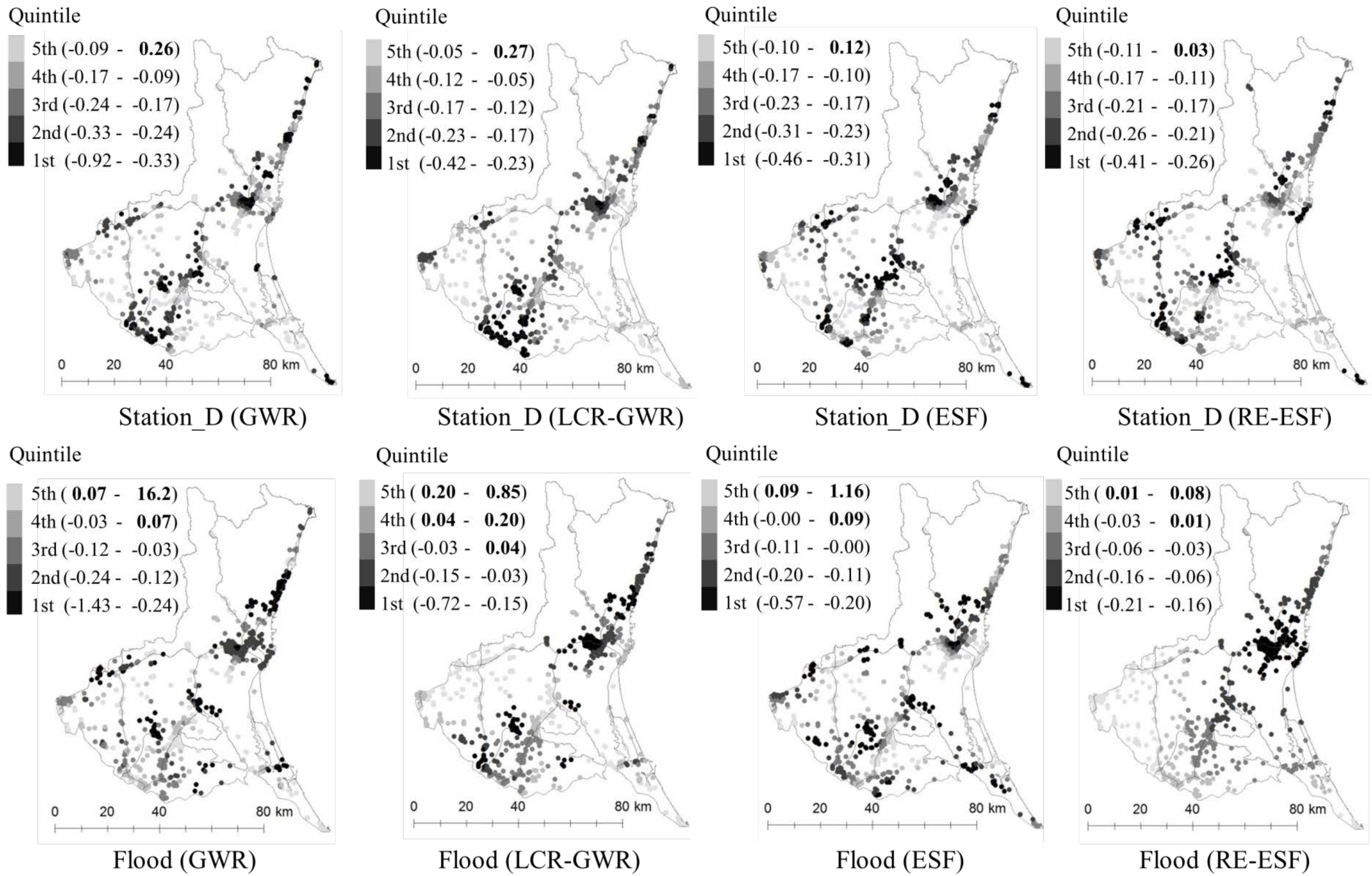

Figure 3. Spatial plots of the estimated coefficients (continued). In each legend, positive values are denoted by bold text.