geographical analysis

Geographical Analysis (2018) **50,** 53–75

A Comparison of Spatially Varying Regression Coefficient Estimates Using Geographically Weighted and Spatial-Filter-Based Techniques

Taylor M. Oshan, A. Stewart Fotheringham

School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, USA

Geographically weighted regression (GWR) is a technique that explores spatial nonstationarity in data-generating processes by allowing regression coefficients to vary spatially. It is a widely applied technique across domains because it is intuitive and conforms to the well-understood framework of regression. An alternative method to GWR that has been suggested is spatial filtering, which it has been argued provides a superior alternative to GWR by producing spatially varying regression coefficients that are not correlated with each other and which display less spatial autocorrelation. It is, therefore, worthwhile to examine these claims by comparing the output from both methods. We do this by using simulated data that represent two sets of spatially varying processes and examining how well both techniques replicate the known local parameter values. The article finds no support that spatial filtering produces local parameter estimates with superior properties. The results indicate that the original spatial filtering specification is prone to overfitting and is generally inferior to GWR, while an alternative specification that minimizes the mean square error (MSE) of coefficient estimates produces results that are similar to GWR. However, since we generally do not know the true coefficients, the MSE minimizing specification is impractical for applied research.

Introduction

Nonstationarity in data-generating processes goes largely undetected in traditional global models. Hence, local models are necessary to capture the effects of spatially heterogenous processes. While many local statistical techniques exist (Fotheringham and Brunsdon 1999), one method that has become particularly popular is geographically weighted regression (GWR), which explicitly allows regression coefficients to vary over space using a relatively simple extension of a global ordinary least-squares estimator. Location-specific coefficients are particularly useful because they can be mapped so that nonstationary processes can be explored visually and analytically (Fotheringham, Brunsdon, and Charlton 2002). Furthermore, GWR

Correspondence: Taylor M. Oshan, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ
e-mail: toshan@asu.edu

Submitted: May 11, 2016. Revised version accepted: May 14, 2017.

doi: 10.1111/gean.12133

inherits traditional regression diagnostics, such as standard errors and *t*-values, making it simple to understand and apply. As a result, GWR has been used to model many different phenomena, including crime patterns (Cahill and Mulligan 2007), health risks (Gilbert and Chakraborty 2011), house prices (Fotheringham, Crespo, and Yao 2015), climate change (Brown et al. 2011), and species distribution (Miller and Hanham 2011).

One critique of GWR is that it is highly susceptible to multicollinearity (Wheeler and Tiefelsdorf 2005), whereby it is suggested that multicollinearity among explanatory variables causes intolerable levels of correlation among GWR coefficients. While Wheeler and Tiefelsdorf make the point that there may be locally induced multicollinearity, even when none or little is detected in the global data set, their assertion that low-to-moderate levels of global multicollinearity implies problematic levels of local multicollinearity may be highly exaggerated. First, in their initial experiments either the true coefficient values are unknown and cannot be assessed or their synthetic data and models are inappropriately specified. Second, subsequent results show that GWR is robust to high levels of multicollinearity, when the sample size is large (Páez, Farber, and Wheeler 2011; Fotheringham and Oshan 2016). While a method for explicitly determining when a sample size is sufficiently large or insufficiently small has yet to be developed, several global regression modeling tools, such as multicollinearity measures, visualizations, and regularized regression have been extended to diagnose or mitigate the effects that may arise from small sample sizes and high multicollinearity (Wheeler 2007, 2010). However, these tools can be relatively complex compared to their global regression counterparts.

A general critique of the work of Wheeler and Tiefelsdorf (2005) is that they use the term multicollinearity to refer to several different correlation phenomena and do not always clearly distinguish between them. Following Fotheringham and Oshan (2016), we distinguish between three types of correlation that can be observed in GWR: (1) between explanatory variables; (2) between location-specific coefficient estimates of different coefficient surfaces; and (3) between coefficient estimates within a single coefficient surface (i.e., spatial autocorrelation). It is well known that extreme levels of multicollinearity among explanatory variables in global regression can cause issues such as unstable coefficients, counter-intuitive coefficient signs, artificially high R^2 diagnostics despite few or no significant estimates, and inflated standard errors of the coefficient estimates (Belsey, Kuh, and Welsch 1980; Obrien 2007). An outcome of this scenario is a type-II error problem, which has also been demonstrated in GWR (Fotheringham and Oshan 2016).

However, attributing the correlation between coefficient estimate surfaces solely to correlation between explanatory variables is an overly simplistic characterization of relationships in local models. Correlation between coefficient estimate surfaces may also be caused by concurvity, a concept from the generalized additive modeling literature, which describes correlation between functionally transformed variables, and has been shown to be an issue more generally within local modeling (Ramsay, Burnett, and Krewski 2003a,b). In GWR, conurvity may be likely due to the fact that the same kernel function and spatial dimensions are often used to weight several variables. Furthermore, when this kernel is parameterized with a single bandwidth, the risk of concurvity may be greater since the kernel functions will be more similar than if they were each parameterized with a unique bandwidth. Specifications proposing different distance metrics and multiple bandwidths (Lu et al. 2015; Fotheringham Yang, and Kang, in press) may be useful for mitigating concurvity within a GWR framework, though concurvity in GWR remains an under-explored issue. Finally, correlation across sets of local coefficient

estimates may be due to the fact the data-generating processes that are represented by the true local coefficients may themselves be correlated.

Spatial autocorrelation within a set of GWR coefficient estimates may be caused by the smoothing nature of the continuous kernel functions used. However, it is possible to explain much of the spatial autocorrelation observed in GWR parameter estimates as desirable properties and *not* as undesirable properties. The essence of GWR is to leverage Tobler's first law of geography, which states that nearer things will be more related, such that autocorrelated data can been seen as the product of autocorrelated processes, which are represented by spatially autocorrelated model coefficients. Hence, there is nothing inherently wrong with spatial autocorrelation among estimated local coefficients.

An eigenvector spatial-filter-based local regression (SFLR) has been suggested (Griffith 2008) as an alternative to GWR to reduce or remove multicollinearity effects described by Wheeler and Tiefelsdorf (2005). This technique creates synthetic variables by interacting the explanatory variables with the eigenvectors of a spatial weight matrix. A subset of the synthetic variables is then selected using a stepwise regression procedure, whereby local coefficients can be obtained by combining the global explanatory variable coefficients with the coefficients for the selected interaction terms that correspond to each explanatory variable. The SFLR technique is touted as superior to GWR in one empirical application in that it accounts for more autocorrelation in the residuals, produces less spatial autocorrelation within sets of coefficients, and displays less multicollinearity across sets of location-specific coefficients (Griffith 2008). However, the true properties of the local coefficients were unknown and without knowledge of the true local coefficients it is impossible to claim superiority of one technique over another by comparing predicted coefficients. In addition, it is alarming that GWR and SFLR, which are claimed to be approximately equivalent methods, seemingly produce contrasting coefficient estimate values (Griffith 2008). The SFLR method has been applied infrequently (Griffith 2008), while GWR has produced agreeable results in many studies. Therefore, the primarily goal of this article is to explore this incongruence by employing simulated data in order to test which of the techniques can more reliably estimate the true coefficients of nonstationary processes. The results of the two approaches are compared with particular attention to the levels of correlation within and between the true coefficient surfaces and the estimated coefficient surfaces. An alternative computational routine is then investigated which illuminates the true relationship between the GWR and SFLR frameworks. Furthermore, several issues of generalizability and replicability of the SFLR framework are discussed.

Background

Geographically weighted regression

A basic GWR model may be specified as

$$y_i = \beta_{i0} + \sum_{k=1}^{p} \beta_{ik} x_{ik} + \epsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where y_i is the dependent variable at location i, β_{i0} is the intercept coefficient at location i, x_{ik} is the kth explanatory variable at location i, β_{ik} is the kth local regression coefficient for the kth explanatory variable at location i, and ϵ_i is the random error term associated with location i.

Using a weighted least squares estimation routine, this specification produces a set of coefficients for each explanatory variable at every location. In matrix form this is given by

$$\hat{\boldsymbol{\beta}}(i)=[\boldsymbol{X}'\boldsymbol{W}(i)\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{W}(i)\boldsymbol{y} \tag{2}$$

where $\boldsymbol{X}$ is a matrix of explanatory variables, $\boldsymbol{W}(i)=\text{diag}[w_1(i),\ldots,w_n(i)]$ is the diagonal weights matrix that weights each observation based on its distance from location i, $\hat{\boldsymbol{\beta}}(i)$ is a vector of coefficients, $\boldsymbol{y}$ is a vector of observations of the dependent variable, and $'$ denotes the matrix transpose operation. A kernel function is applied to the distances between observations and calibration locations to calculate the weights matrix. The bi-square function is a common kernel function for placing emphasis on observations that are closer in space to calibrations points and is given by

$$w_j(i)=\begin{cases}[1-(\frac{d_{ij}}{b})^2]^2, & \text{if } j\in\{N_i\},\\ 0, & \text{if } j\notin\{N_i\}\end{cases} \tag{3}$$

where d_{ij} is the distance between i and j, b is the distance to the Nth nearest neighbor, and $\{N_i\}$ is the set of all observations that are no more than b distance from i. An observation at location i would have a weight of 1 whereas those observations which are farther than distance d will have a weight of zero. Prior to fitting a GWR model, an optimal number of number of nearest neighbors must be selected, which in turn defines b in the above bi-square kernel, also called the bandwidth. We opt to use an AIC minimization routine to select the kernel bandwidth throughout this research as it provides a useful trade-off between goodness-of-fit and degrees of freedom or model complexity (Fotheringham, Brunsdon, and Charlton 2002). A GWR-specific AIC is calculated as

$$AIC_c=2n\log_e(\hat{\sigma})+n\log_e(2\pi)+n\left\{\frac{n+tr(\boldsymbol{S})}{n-2-tr(\boldsymbol{S})}\right\} \tag{4}$$

with n denoting the sample size, $\hat{\sigma}$ defined as the standard deviation of the error term, and $tr(\boldsymbol{S})$ being the trace of the hat matrix. After the bandwidth is selected, the weights can be computed using equation (3) and finally, the GWR model can be fitted at each calibration location using equation (2) to obtain a set of local coefficients. An overall R^2 for a GWR model can be obtained by taking the average of the local R^2 values for each calibration location.

Spatial-filter-based local regression

Eigenvector spatial filters are based on the interpretation that the eigenvectors of a modified connectivity matrix are the set of possible orthogonal and uncorrelated map patterns (i.e., degree of spatial autocorrelation) (Griffith 1996, 2011). Further, the first eigenvector, $\boldsymbol{E}_1$, is the set of real numbers that produces the map pattern with the largest achievable Moran's I correlation coefficient (MC), the second eigenvector, $\boldsymbol{E}_2$, is the set of real numbers that produces the map pattern with the largest achievable MC while remaining uncorrelated with $\boldsymbol{E}_1$, and continues on such that $\boldsymbol{E}_n$, achieves the largest negative MC and is uncorrelated with the preceding $(n-1)$ eigenvectors. The modified connectivity matrix,[1] $\boldsymbol{W}$, is most frequently defined as

$$(\boldsymbol{I}-\mathbf{1}\mathbf{1}'/n)\boldsymbol{C}_n(\boldsymbol{I}-\mathbf{1}\mathbf{1}'/n) \tag{5}$$

where $\boldsymbol{I}$ is an $n \times n$ identity matrix, 1 is an $n\times 1$ vector of 1′s, and $\boldsymbol{C}_n$ is the original binary connectivity matrix for n mutually exclusive and exhaustive spatial units that partition the study space.

By selecting a subset of the eigenvectors derived from $\boldsymbol{W}$ and creating a linear combination, it is possible to produce a synthetic variable that accounts for spatial autocorrelation in the residuals, which can be incorporated into a linear regression model as

$$\boldsymbol{y}=\beta_0\mathbf{1}+\sum_{p=1}^{P}\boldsymbol{X}_p\beta_p+\sum_{k=1}^{K}\boldsymbol{E}_k\beta_{E_k}+\boldsymbol{\epsilon} \tag{6}$$

where the first two terms give a traditional linear regression specification (i.e., the intercept and a combination of coefficient and explanatory variable pairs), and the third term represents the spatial filter. Here, $\boldsymbol{E}_k$ is a potential eigenvector in the total set of K selected eigenvectors. Griffith posits that the specification in (6) should also include an interaction term between each explanatory variable and each selected eigenvector in order to create local coefficients. He further postulates that adding the new terms will provide a specification that is approximately equal to a GWR model such that

$$\hat{\boldsymbol{y}}_{gwr} \approx (\beta_0\mathbf{1}+\sum_{k_0=1}^{K_0}\boldsymbol{E}_{k_0}\beta_{k_0})+\sum_{p=1}^{P}(\beta_p\mathbf{1}+\sum_{k_p=1}^{K_p}\boldsymbol{E}_{k_p}\beta_{k_p})\cdot\boldsymbol{X}_p+\boldsymbol{\epsilon} \tag{7}$$

where $\hat{\boldsymbol{y}}$ is the vector of predicted values of a GWR model, $\cdot$ denotes element-wise matrix multiplication (i.e., Hadamard matrix multiplication), and k_p represents an eigenvector from the set of K_p eigenvectors that describe explanatory variable p. According to Griffith (2008), when the first and third term in (6) are combined in (7) they become the GWR intercept coefficients, which will be explored in more detail in subsequent sections. Finally, by distributing the $\boldsymbol{X}_p$ across their premultiplied sums in equation (7) and re-arranging the terms, the following specification is achieved

$$\boldsymbol{y}=\beta_0\mathbf{1}+\sum_{p=1}^{P}\boldsymbol{X}_p\cdot 1\beta_p+\sum_{k=1}^{K}\boldsymbol{E}_k\beta_{E_k}+\sum_{p=1}^{P}\sum_{k=1}^{K}\boldsymbol{X}_p\cdot\boldsymbol{E}_k\beta_{pE_k}+\boldsymbol{\epsilon} \tag{8}$$

where the first term is the regression intercept, the second term represents the explanatory variables with their global coefficients, and the third and fourth terms denote the global eigenvectors and the interaction terms between explanatory variables and eigenvectors, respectively. Importantly, the global explanatory variables are required in this specification in order to construct the interaction terms, which ultimately allow the construction of local coefficients. Once eigenvectors and interaction terms are selected and equation (8) is estimated, the local coefficients for each explanatory variable can be obtained through an additive combination of the coefficient associated with each global explanatory variable, p and those coefficients associated with the interaction terms based on the corresponding global explanatory variable, unlike the GWR framework. However, it is not clear how hypothesis testing regarding local coefficients from spatial filtering is performed, nor how a correction for multiple hypothesis testing can be applied.

An important step in the spatial-filter-based framework is selecting a subset of the eigenvectors and interaction terms. The methodology entails a forward stepwise regression variable selection algorithm among all interaction terms and global eigenvectors based on statistical significance. To begin, only eigenvectors displaying moderate to high spatial autocorrelation (i.e., Moran's I coefficient (MC) approximately greater than 0.25) are selected. Then, for each iteration of the stepwise routine, each candidate variable is tested in the model. The variable that produces the smallest P-value is selected, and if any other variables become insignificant (P-value greater than 0.1), they are removed from the model. The algorithm continues until no variable can be added to the model that also produces an associated P-value of less than 0.1001, at which point the routine terminates and the local coefficients can be processed (Griffith 2008). In a later section, this specification is compared to two alternative mean-square-error-minimization (MSE) selection variations of this stepwise selection routine developed as part of this research.

Simulating nonstationarity and multicollinearity

Two processes were used to simulate data with known properties in order to reliably compare local modeling methods. Process (I) relies on neither the GWR or SFLR specification while process (II) explicitly relies on the SFLR specification.

Agnostic simulated data—process I

To begin, 2,500 observations were generated for the cells of a 50 by 50 grid, where each cell is considered a spatial unit. The observations within the cells were generated using the following equation

$$Y_i = \beta_{0i} + \beta_{1i} X_{1i} + \beta_{2i} X_{2i} + \epsilon_i \tag{9}$$

where Y is the generated observation, β_0, β_1, and β_2 are *known* locally varying coefficients, X_1 and X_2 are variables drawn from random normal distributions (mean of 50 and variance of 75), ϵ is a random normal error term (mean of 0 and variance of 15), and i is the index for each of the 2,500 locations. The spatial distributions of β_0, β_1, and β_2 for process I are illustrated in Fig. 1. β_0 has a gradual regional trend, with values generally increasing from the north–west to the south–east; β_1 is distributed in a checker-board manner generally with the lowest values in the north–west, higher values in the south–east, and medium values in the north–east and south–west; and β_2 is also distributed with a checker-board pattern but a more complex one, where smaller values form a cross and larger values are found in each of the four corners. β_1 and β_2 were derived by sampling from different multivariate normal distributions depending on discrete localities. These surfaces were then smoothed using a uniform filter to remove rigid boundaries between locales and subsequently resampled from a normal distribution with the smoothed value as the mean and the original locale-specific variance in order to generate the speckled patterns observed in Fig. 1. Since these surfaces are based on local processes, they result in MC values that signify very strong spatial autocorrelation (0.982, 0.947, and 0.912 for β_0, β_1, and β_2, respectively).[2] There is strong correlation between β_0 and β_1 (Pearson's $r = 0.88$), though there is virtually no correlation among β_0 and β_2 or β_1 and β_2 (Pearson's $r < 0.1$). Moderate collinearity between X_1 and X_2 was generated using the formula

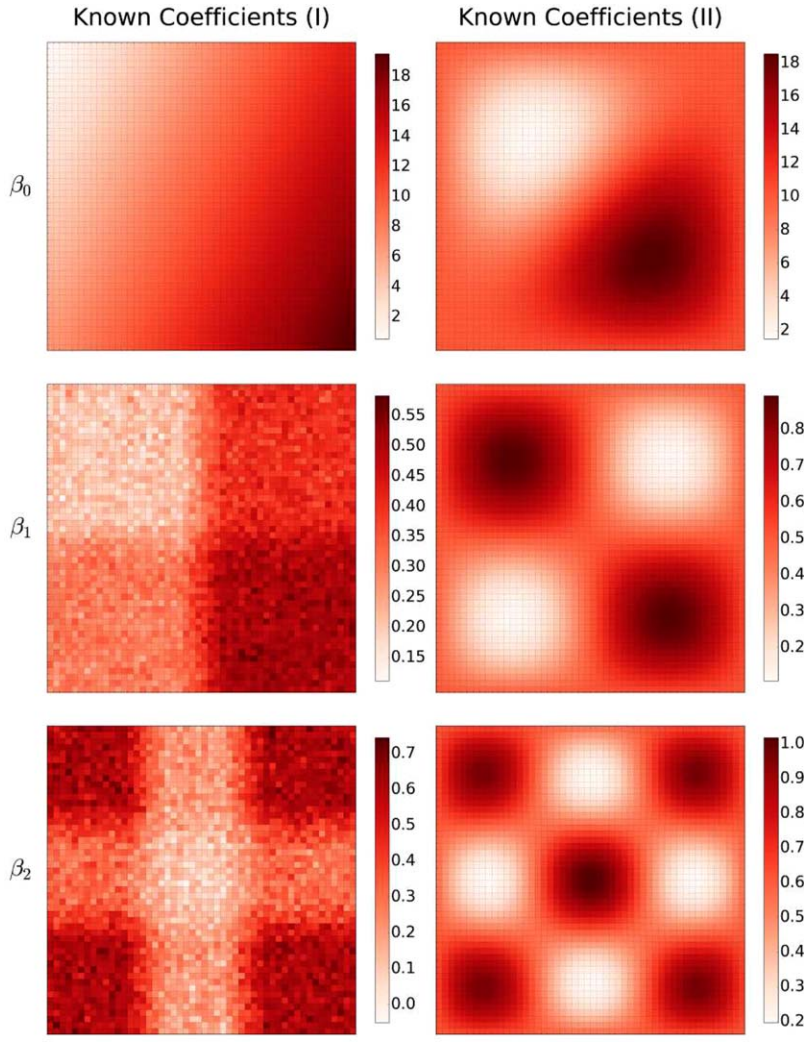

Figure 1. Simulated coefficient surfaces generated using normally distributed random variables for process I (left) and process II (right). β_0 represents the coefficients associated with the intercept, while β_1 and β_2 represent the coefficients associated with the simulated explanatory variables X_1 and X_2, respectively. Low coefficient values are shaded lighter while high values are shaded darker.

$$X_2 = r * X_1 + \sqrt{1-r^2} * X_2 \tag{10}$$

where r is the approximate level of correlation desired between X_1 and X_2 and was specified at 0.5 to created the observed values denoted by Y_i. Even though this simulated data set is not based upon either GWR or SFLR, it is impossible to know if the data may favor one of the specifications over the other. Therefore, a second set of simulated data is generated that is in accordance with the SFLR methodology, and is, therefore, certainly not biased against the SFLR methodology. This removes any potential ambiguity from whether or not the observed results are due to the methodology or from an unfavorable data-generating process.

SFLR-based simulated data—process II

The same 50 by 50 tessellation was used to derive a first order binary contiguity spatial weight matrix, which was then transformed according to equation (5) and subsequently decomposed into eigenvalues and corresponding eigenvectors. Observations were then generated using a simplified version of equation (8) where one eigenvector was selected to interact with each of the variables. This yields the following formula

$$Y_i = \beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + E_{2i}\beta_{E_2} + (X_{1i}E_{3i})\beta_{E_{1,3}} + (X_{2i}E_{10i})\beta_{E_{2,10}} + \epsilon_i \quad (11)$$

where X_1, X_2, and ϵ_i are as defined in process I, E_{2i}, E_{3i}, and E_{10i} are the eigenvectors associated with eigenvalues 2, 3, and 10, and β_0, β_1, β_2, β_{E_2}, $\beta_{E_{1,3}}$, $\beta_{E_{2,10}}$ are scalar coefficients given by 10, 0.5, 0.6, −200.0, −10.0, and 10.0, respectively. The first three terms of equation (11) are the main effects that form a basic global regression, while the second three terms are interaction terms that allow spatially local deviations from the global model. Summing the global coefficients from each of the main effects with any associated coefficients from an interaction term multiplied by its corresponding eigenvectors yields local coefficient surfaces (Fig. 1). Coefficient values and eigenvectors were chosen by trial and error in order to create surfaces which were similar in their general pattern and magnitude to those from process I. It should be noted that these coefficient surfaces for process II are smooth by nature, unlike those from process I, which is evident in their higher MC values of 1.016, 1.013, and 1.003 for β_0, β_1, and β_2, respectively. Since the coefficients are derived from eigenvectors that are orthogonal, it follows that there is no correlation between any of the three coefficient surfaces for process II.

Exploring the spatial-filter-based local regression framework

Since there is no existing software for carrying out the SFLR method, custom code was developed.[3] SFLR models were then calibrated using data from process I and process II, where the candidate eigenvector terms were limited to the first 750 eigenvectors and their interaction with each of the explanatory variables, X_1 and X_2, for a total set of 2, 225 synthetic variable candidates. A cut-off of the first 750 eigenvectors was chosen because these eigenvectors showed moderate to high levels of positive spatial autocorrelation (MC greater than 0.25), which approximately corresponds to the original methodology (Griffith 2008). A subset of the candidates was selected using Griffith's (2008) forward stepwise routine, whereby model fit statistics (adjusted R^2 and AIC) and accuracy statistics (MSE of $\hat{\beta}_1$ and $\hat{\beta}_2$) were collected at each iteration of the algorithm.

Overall, the SFLR results for process I and process II (Fig. 2) show similar trends where model fit increases with more iterations. More specifically, the model fit in terms of both adjusted R^2 and AIC for both sets of variables increases rapidly at first, as the number of iterations increases, but then levels off, showing only modest gains for additional iterations. Interestingly, the MSE associated with the local estimated coefficient surfaces quickly reaches a minimum before increasing as more model terms are selected at higher iterations. This trend is less apparent for process II where the minimum MSE is reached within the first few iterations, though in both cases this implies that as model fit increases the accuracy of the coefficients decreases. It should be noted that while the SLFR model for process II terminated after 350 iterations (315 selected candidates, implying 37 selected terms were removed from the model after becoming insignificant), the SFLR model for process I did not terminate naturally.

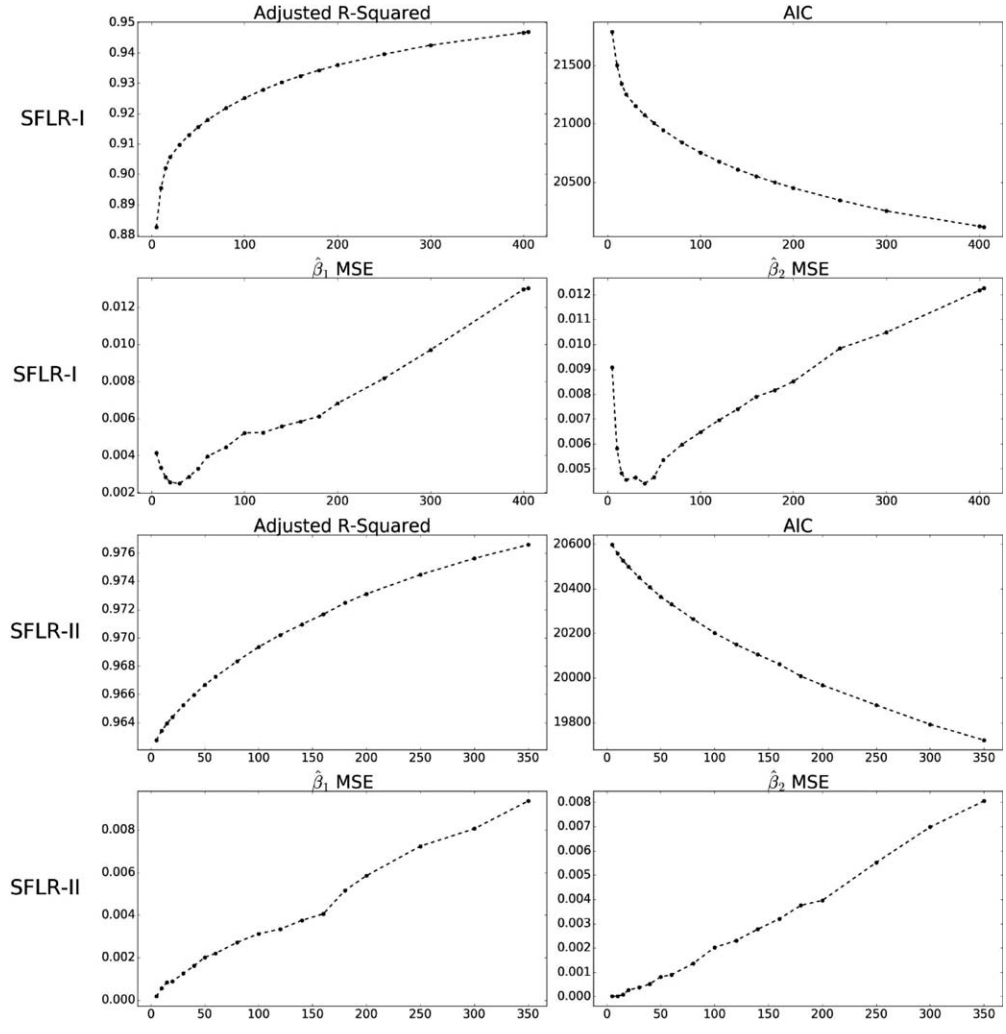

Figure 2. Adjusted *R*-squared, AIC, and MSE for $\hat{\beta}_1$ and $\hat{\beta}_2$ estimated via SFLR for various numbers of stepwise regression iterations for both process I and process II.

Instead, the routine stopped after 405 iterations (355 selected candidates with 50 being removed for insignificance) due to the inability of the regression calibration to converge to a solution for the coefficient estimates.[4]

The local β_1 and β_2 coefficients estimated for both process I and process II (Fig. 3) display similar spatial distributions in which the overall patterns of the known coefficients are discernible, though they are obfuscated by what appears to be random perturbations. This is especially striking for the coefficient estimates from process II because the true coefficients are known to be smooth. An important effect of this is that the local coefficient estimate distributions are all wider than the corresponding known coefficients, often with negative coefficient estimates where none should exist. This signals that the noisy pattern that is observed is likely due to overfitting. Essentially, the model fit is increased at the expense of coefficient accuracy and considerable computation time.

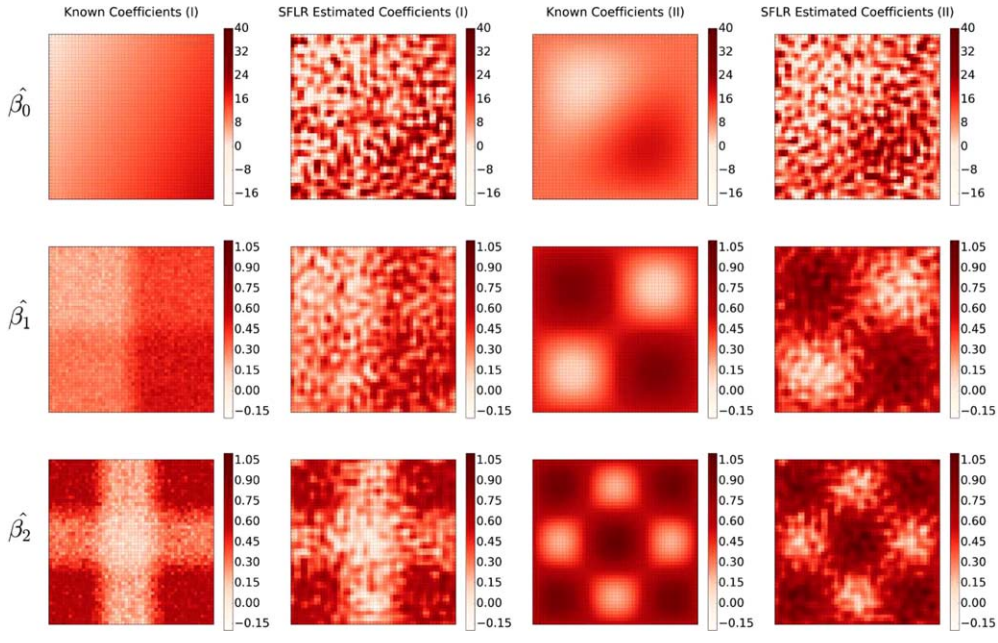

Figure 3. Known coefficients and estimated coefficients using SFLR for process I and process II. Low coefficient values are shaded lighter while high values are shaded darker.

Comparing geographically weighted regression and spatial-filter-based local regression

Griffith (2008, 2,768) claims that the SFLR framework is superior to GWR because the estimated coefficients from SFLR "remain unbiased, yield a better global model fit, are polluted with considerably lower levels of spatial autocorrelation, and, for the most part, display little relationship to the GWR coefficients." Of note is that these conclusions are reached without using known coefficients to assess the results and are drawn based on only a single example. That means that it is simply not possible to comment on the appropriate level of spatial autocorrelation to expect in the coefficient estimates, on whether or not the coefficient estimates are unbiased, or on which of the methods produces more accurate coefficient estimates. Therefore, we provide a comparison of the ability of SLFR and GWR frameworks to replicate local parameters, where the true values are known.

A GWR model was calibrated using a bi-square nearest neighbor bandwidth by minimizing the AICc measure given in (4) for both process I and process II. The resulting coefficient estimates are presented in Fig. 4. It is apparent that the surfaces of GWR coefficient estimates are similar to those of the known coefficients. This is further demonstrated in Fig. 5, which shows a strong linear correlation between the GWR coefficient estimates and the known values for both processes showing that GWR satisfactorily replicates the known local coefficients. Of course, the GWR model results are not perfect. The coefficient estimate surfaces for process I are more smooth than the known coefficients while the coefficient estimates for process II are slightly less smooth than the known coefficients. Despite this, the GWR coefficient estimates (Fig. 5) for both processes show stronger linear relationships with the known coefficients than the SFLR coefficient estimates.

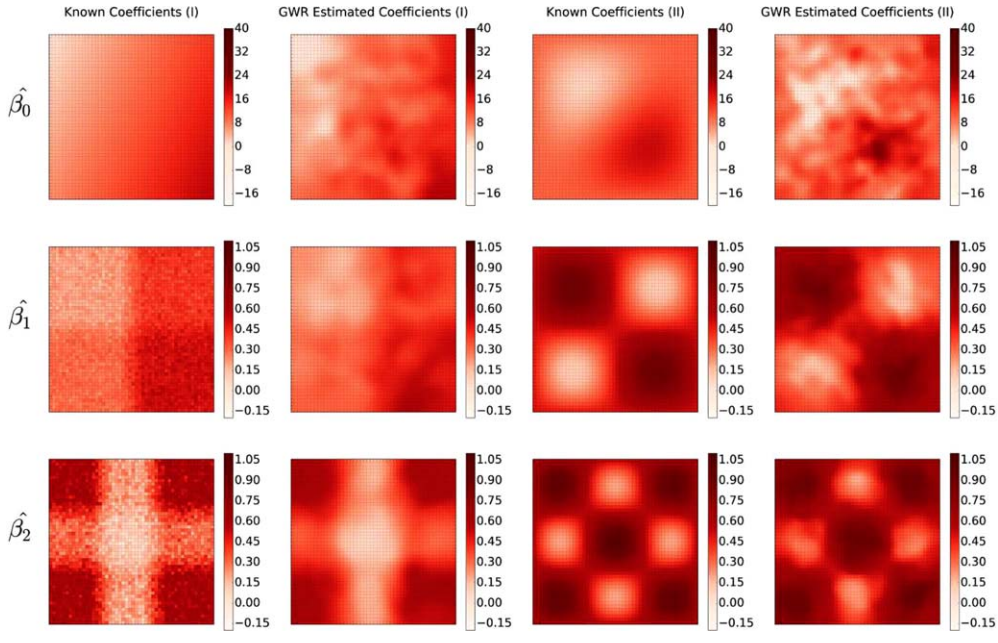

Figure 4. Known coefficients and estimated coefficients using GWR for process I and process II. Low coefficient values are shaded lighter while high values are shaded darker.

The GWR coefficient estimates and the SFLR coefficient estimates from the previous section are compared in Fig. 6, whereby Griffith's postulation that the two methods produce different results is not supported. The local coefficient estimates for β_1 and β_2 from both methods are clearly positively correlated for both process I and process II. The β_0 estimates have much less of a trend for both processes, but this is expected since the error structure for GWR and SFLR are inherently different. A similar conclusion is reached by comparing the MSE values of the estimated coefficient surfaces (Table 1), where GWR consistently achieves more accurate coefficient estimates (i.e., lower MSE values) than SFLR for both process I and process II. The SFLR methodology does produce lower levels of spatial autocorrelation among the estimated parameter surfaces and better model fit, as Griffith (2008) claimed, though this is because of overfitting and should be seen as undesirable properties of the model. Clearly, these lower levels of spatial autocorrelation are departures from the known coefficients. The spatial autocorrelation as indicated by the MC values for the GWR estimates are consistently closer to the known values than for the SFLR estimates.

One final observation is the Pearson's correlation coefficients between coefficient estimate surfaces (Table 1). For GWR, these are all as expected with strong correlation between β_0 and β_1 for process I and very low correlation between all other coefficient surfaces. For SFLR, the correlation coefficients are all as expected, except $r(\hat{\beta}_0, \hat{\beta}_1) = 0.14$ for process I, which is substantially lower than its true value of 0.88. This indicates that it may not necessarily be desirable to always have coefficient surfaces that display less correlation between themselves and is a potential drawback to SFLR. In practical applications, it is possible to employ local correlation coefficients to explore potential spatial variation of correlation between coefficient estimates (Wheeler 2007), though this does not necessarily indicate whether variation is naturally occurring or due to other factors.

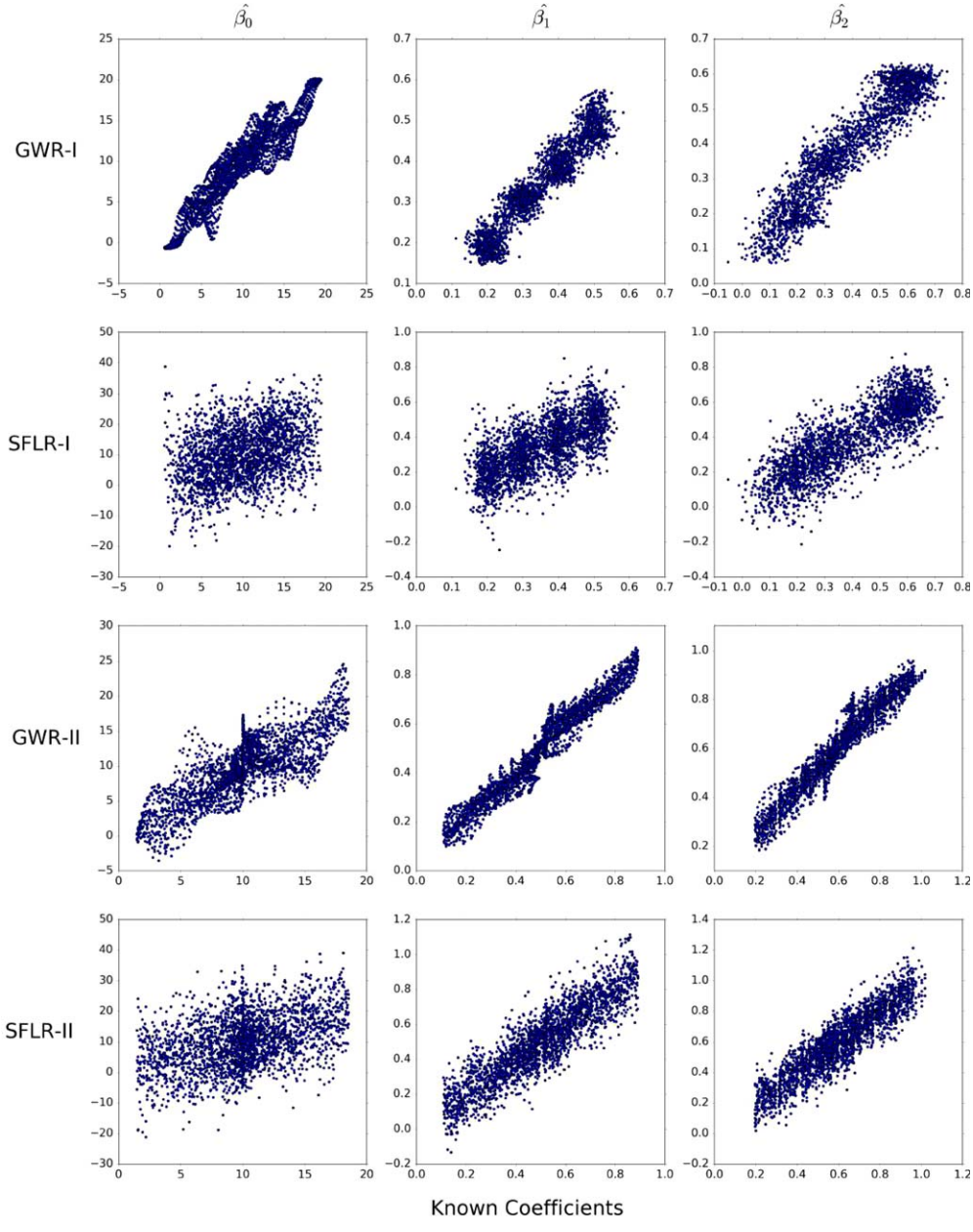

Figure 5. Scatter plots of GWR and SFLR coefficient estimates for process I and process II against corresponding known coefficient values.

Optimal stepwise routines for spatial-filter-based local regressions

Given the results showing that the original SFLR routine hits minimum MSE values well before the significance-based stepwise routine terminates, two new stepwise routines which incorporate a MSE-minimization criterion are now introduced to explore what the results would look like if the SFLR methodology was accurately replicating the coefficient surfaces. In the first routine (A), the variable that achieves the lowest additive MSE between $\hat{\beta}_1$ and $\hat{\beta}_2$

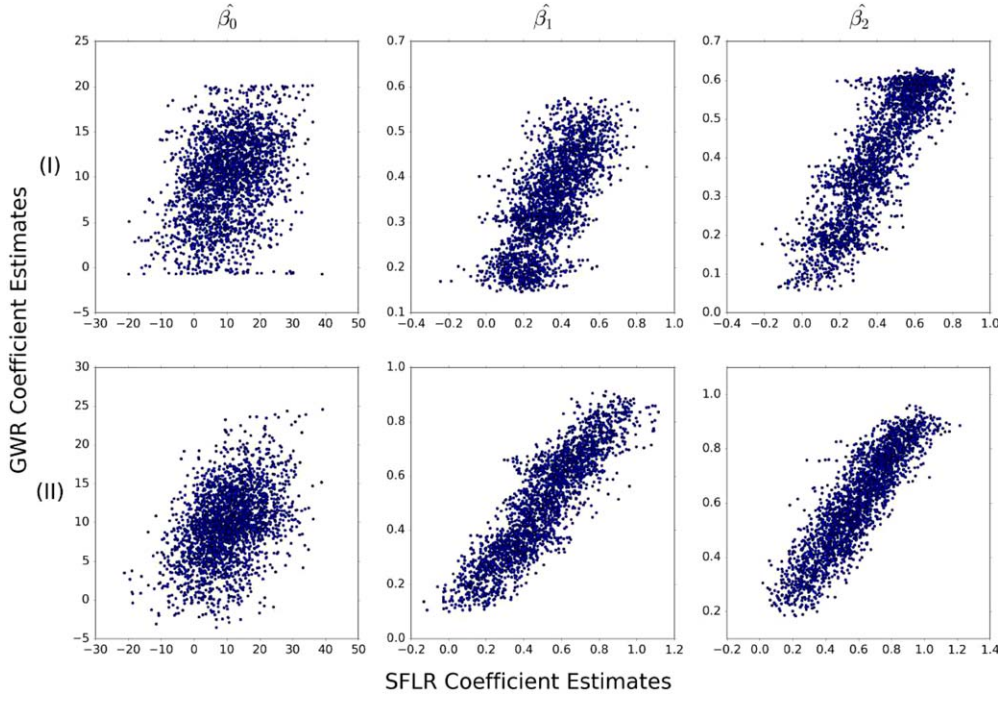

Figure 6. Scatter plots of SFLR coefficient estimates against the GWR coefficient estimates for process I and process II.

is selected at each iteration. The algorithm continues until no further reduction in MSE can be achieved[5]. In contrast, the second routine (B) selects the variable with the lowest *P*-value, similar to the original routine, though the algorithm is terminated when new variables no longer decrease the MSE, similarly to algorithm one. In both new algorithms, variables that become insignificant (i.e., *P*-value greater than 0.1001) at each iteration are still removed.

Each algorithm was used to carry out the estimation of a model for process I and process II to assess the results and compare them to GWR. Figs. 7 and 8 provide the resulting estimated coefficient surfaces alongside their respective known coefficient surfaces. It is clear that these surfaces are much more representative of the spatial patterns associated with the known coefficient surfaces than those from the original SFLR routine. Unsurprisingly, the MSE-minimizing SFLR algorithms almost perfectly reproduce the known coefficients for process II, which are generated using an SFLR specification. The only exception is a small amount of noise in algorithm A where it appears that the spatially patterned β_2 has been confounded with the random error inherent in the data. This may occur because in SFLR the error term is specified globally in conjunction with the mean effect of each explanatory variable. In contrast, a GWR model assumes a global error term, although in practice it utilizes an independent error term for each estimation location. The superiority of the MSE-minimizing algorithms are supported in Fig. 9, which indicates a stronger linear trend between the estimated coefficients from both new algorithms and the known coefficients than there is between the estimated coefficients from the original SFLR algorithm and the known coefficients (Fig. 5). Again, it is not surprising that there is a perfect linear trend for the majority of the scatter plots for the MSE-minimized SFLR coefficient estimates and similar results would be expected for GWR if the data were simulated

Table 1. Comparison of MSE, Spatial Autocorrelation, Model Fit, and Pearson's Correlation Coefficients for GWR and SFLR Models Using Data from Process I and Process II

	MSE β_0	MSE β_1	MSE β_2	MC β_0	MC β_1	MC β_2	Adj. R^2	AIC	$r(\beta_0, \beta_1)$	$r(\beta_0, \beta_2)$	$r(\beta_1, \beta_2)$
Known-I	–	–	–	0.982	0.947	0.912	–	–	0.884	0.0496	0.0940
GWR-I	3.239	0.00129	0.00401	0.980	0.994	0.994	0.888	21149.363	0.787	−0.0497	0.116
SFLR-I	72.725	0.0130	0.0123	0.651	0.809	0.905	0.947	20116.961	0.145	−0.0870	0.00542
Known-II	–	–	–	1.016	1.013	1.003	–	–	0.0	0.0	0.0
GWR-II	16.039	0.0686	0.0757	0.988	1.006	0.996	0.953	20971.251	−0.0234	0.0335	0.0208
SFLR-II	72.523	0.0826	0.0860	0.654	0.937	0.930	0.977	19721.121	−0.0656	−0.109	−0.0571

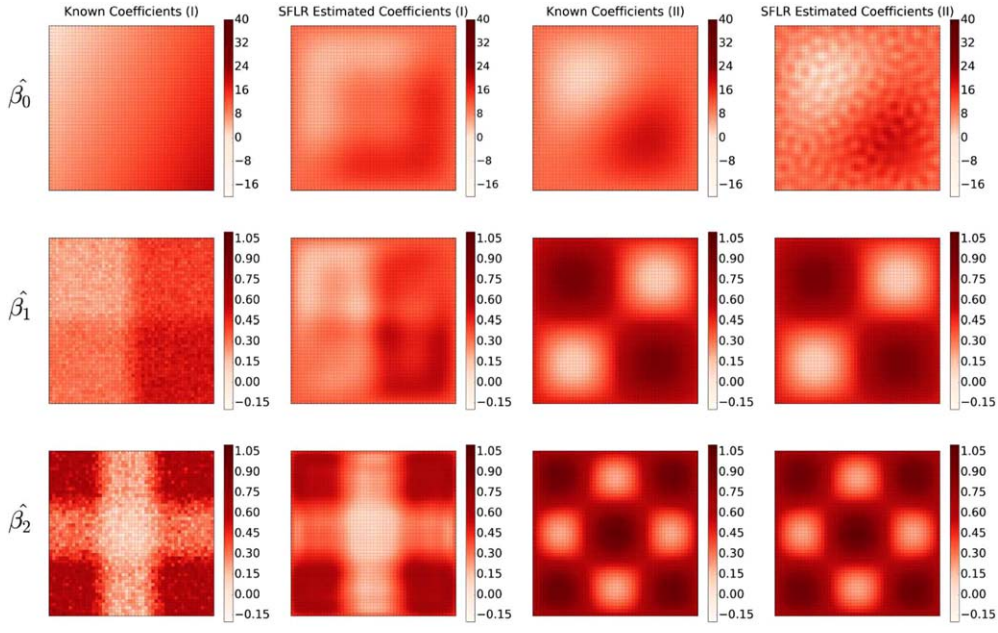

Figure 7. Known coefficients and coefficients estimated using MSE-minimizing SFLR algorithm A for process I and process II. Low coefficient values are shaded lighter while high values are shaded darker.

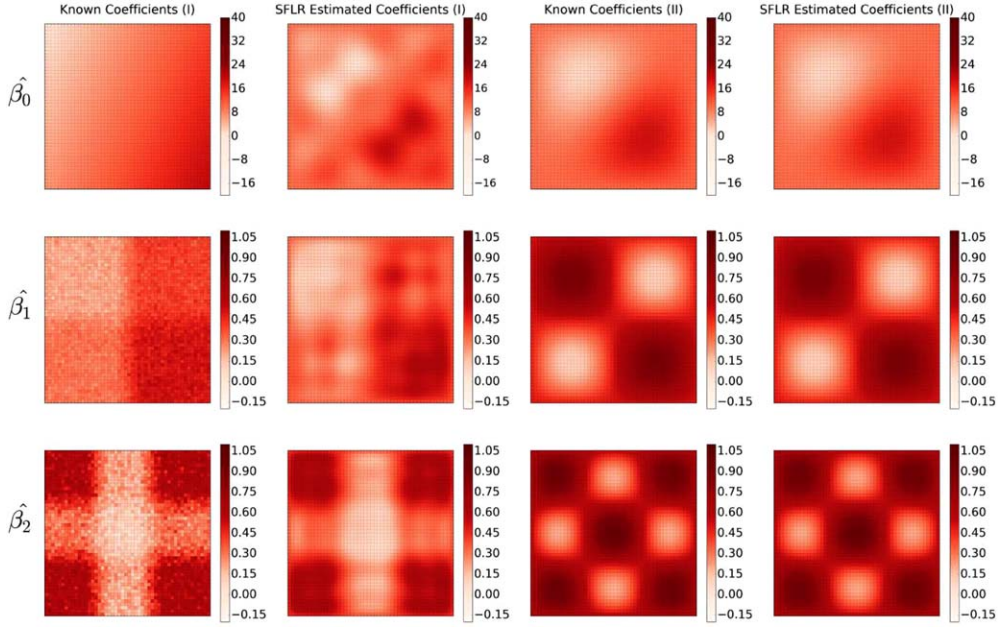

Figure 8. Known coefficients and coefficients estimated using MSE-minimizing SFLR algorithm B for process I and process II. Low coefficient values are shaded lighter while high values are shaded darker.

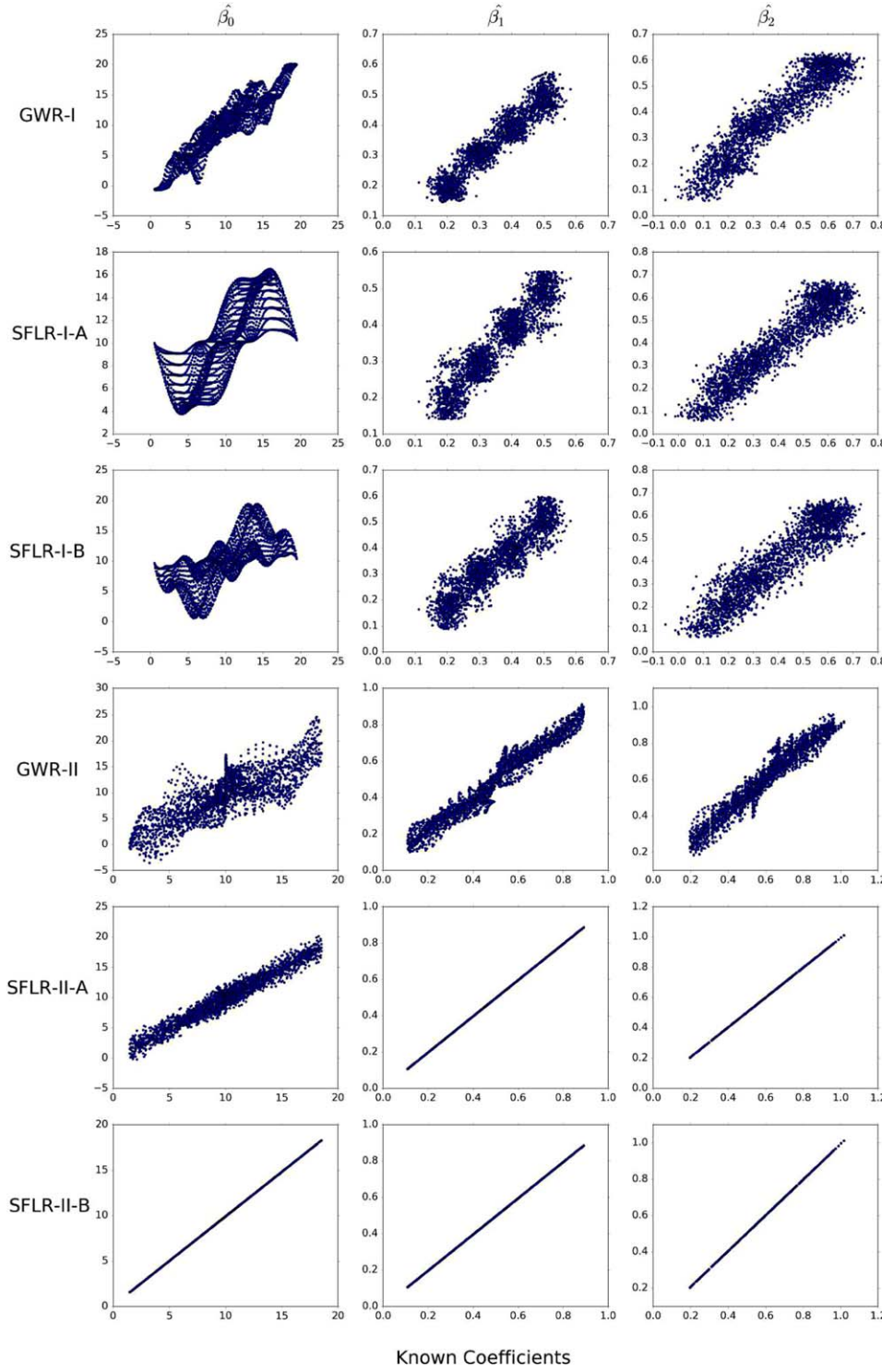

Figure 9. Scatter plots of MSE-minimized SFLR coefficient estimates against known coefficients for algorithm A and B on process I and process II.

using an explicit GWR specification. Interestingly, both SFLR-A and SFLR-B create "ribbons" in the coefficient estimated for β_0, which is theorized to be a problematic feature of GWR indicative of multicollinearity issues by Wheeler and Tiefelsdorf (2005). In contrast, Fotheringham and Oshan (2016) suggest these patterns are due to spatial patterns inherent in the true coefficients and this study provides additional evidence toward this hypothesis and extends it to show that it is can occur irrespective of the local modeling method being used.

Another unexpected outcome is that the new MSE-minimizing algorithms produce coefficient estimate surfaces that are smoother than the originally proposed SFLR algorithm. The coefficient estimate surfaces for process II are generally as smooth as the known coefficient surfaces, while the coefficient estimate surface for process I are more smooth than the known coefficient surfaces, which is similar to the results from GWR (Fig. 4). This is corroborated by the linear trends between the MSE-minimized SFLR coefficient estimates and the GWR coefficient estimates (Fig. 10). The only exception to this pattern of smooth surfaces is the coefficient estimate surface for β_2 for process II using SFLR-A, though this textured pattern indicates a deviation from the known coefficients and is, therefore, undesirable.

Overall, the model fit, coefficient accuracy, and spatial autocorrelation statistics (Table 2) from the SFLR coefficients using the MSE-minimized stepwise selection routines are all similar to those produced by GWR. The optimal values for these metrics are varied across different models, data sets, and selection routines. Therefore, it is likely that when an MSE-minimizing algorithm, or one that approximates it, is used, then SFLR and GWR produce similar results. Finally, according to Table 2, both GWR and MSE-minimizing SFLR yield coefficient estimate surfaces with similar levels of between-set correlation. For process I, little to no correlation exists between the estimated coefficient surfaces between β_0 and β_2 and β_1 and β_2 for any of the models. In addition, high correlation is maintained between β_0 and β_1 across the models, though the correlation is generally underestimated rather than overestimated. For process II, the estimated coefficient surfaces for all of the models maintain that there is virtually no correlation between any of the coefficient estimate surfaces.

Issues with the spatial-filter-based methodology

These new results strongly indicate that the original SFLR method is prone to overfitting. This observation mirrors additional results where the SFLR method was tested using an AIC selection criterion within a genetic algorithm optimization framework (Helbich and Griffith 2016). Though SFLR achieved a very low in-sample prediction error in this study, its out-of-sample prediction error was very high, which is another indicator of overfitting. Therefore, overfitting appears to be a general problem of the SFLR method that is not particular to the simulated data used here and raises the important question, "why is the spatial-filter-based regression method prone to overfitting?".

Adding interaction terms to a linear model, as in SFLR, and local regression, such as GWR, are both more flexible than a basic linear model because they can approximate more complex relationships. More flexible models are known to be more susceptible to overfitting, especially for higher dimensional model data sets (i.e., when there are a large number of variables). Increasing the dimensionality typically increases the chance of overfitting because there are more opportunities for one or more of the dimensions to be statistically significant by random chance even though they may actually just be "noise" (Gareth et al. 2013). For example, if an additional interaction term increases the model fit even slightly, then the SFLR method may

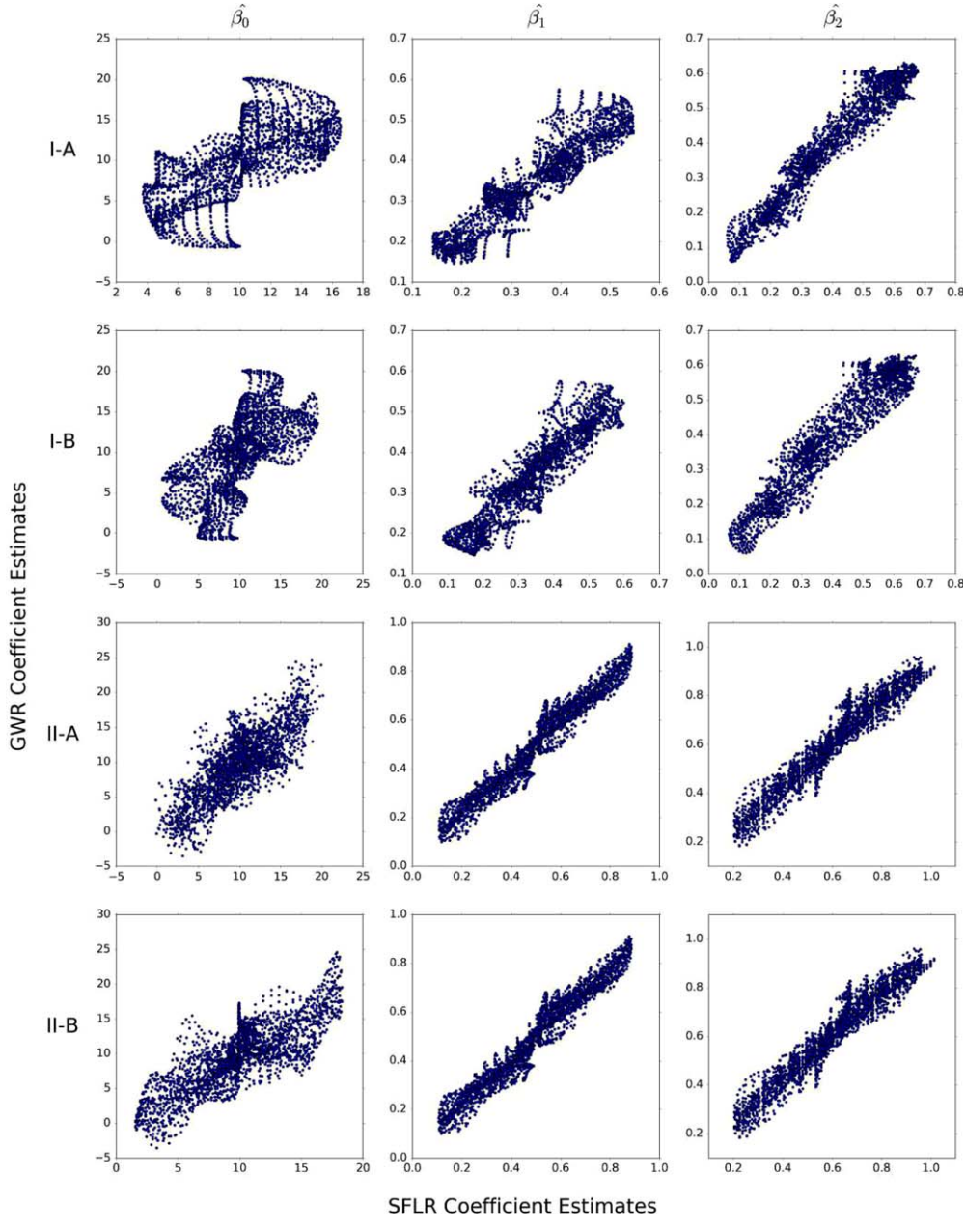

Figure 10. Scatter plots of MSE-minimized SFLR coefficient estimates against GWR coefficient estimates for algorithm A and B on process I and process II.

select it despite the errors that could be induced in the local coefficients. Therefore, a solution to the overfitting issue might be obtained by either reducing the dimensionality or restricting the structure of the model to be less flexible.

A simple method for decreasing the dimensionality might be to decrease the number of interaction terms. The SFLR method only arbitrarily filters the interaction terms to include those eigenvectors that have moderate-to-high spatial autocorrelation. The cut-off for

Table 2. Comparison of MSE, Spatial Autocorrelation, Model Fit, and Pearson's Correlation Coefficients for GWR and SFLR Using MSE-Minimization Routines A and B on Data from Process I and Process II

	MSE β_0	MSE β_1	MSE β_2	MC β_0	MC β_1	MC β_2	Adj. R^2	AIC	$r(\beta_0, \beta_1)$	$r(\beta_0, \beta_2)$	$r(\beta_1, \beta_2)$
Known-I	–	–	–	0.982	0.947	0.912	–	–	0.884	0.0496	0.0940
GWR-I	3.239	0.00129	0.00401	0.980	0.994	0.994	0.888	21149.363	0.787	−0.0497	0.116
SFLR-I-A	7.498	0.00170	0.00380	1.013	1.010	1.002	0.906	21250.258	0.697	0.0362	0.108
SFLR-I-B	11.086	0.00254	0.00458	1.011	1.008	1.002	0.906	21239.453	0.721	0.0	0.0
Known-II	–	–	–	1.016	1.013	1.003	–	–	0.0	0.0	0.0
GWR-II	16.039	0.0686	0.0757	0.988	1.006	0.996	0.953	20971.251	−0.0234	0.0335	0.0208
SFLR-II-A	11.396	0.0727	0.0800	0.990	1.013	1.003	0.962	20614.742	0.0	0.0	0.0
SFLR-II-B	10.524	0.0724	0.0802	1.016	1.013	1.003	0.962	20618.327	0.0	0.0	0.0

eigenvector selection used in this research was a Moran's I correlation coefficient value of 0.25 or greater, which roughly corresponds to the value used in the initial SFLR application (Griffith 2008). Certainly, a different cut-off could be used, though if this is an important aspect of the SFLR method, it will likely vary for different spatial systems because each system will yield a different set of eigenvectors. For instance, as the number of spatial units, n increases, so does the number of eigenvectors and hence, so does the number of interaction terms, p. This is problematic because when there is more than one explanatory variable, p will always grow faster than n, guaranteeing very high dimensionality, unless a substantially reduced set of the eigenvectors are selected. However, it seems unwarranted that SFLR essentially filters the data twice. First to subset the pool of eigenvectors and then subsequently to select an even smaller subset of interaction terms. Rather, an alternative method for reducing dimensionality may be to employ a method that progressively narrows the entire candidate pool of interaction terms based on an exogenous factor. For example, restricted regression, which limits new variables to those that are orthogonal to previously included explanatory variables (Hodges and Reich 2010), may offer an efficient strategy for reducing the dimensionality.

SFLR could also potentially be made less prone to overfitting by enforcing a "strong hierarchy" structure. The interaction terms in SFLR have a hierarchical structure where the interaction effects are interpreted as local deviations from main effects. Despite this, SFLR currently requires that only the explanatory variables be forced into the model as main effects and not the corresponding eigenvectors. It is well known that interaction terms should be based on main effects, therefore, a method that enforces this hierarchy, which has been developed in a LASSO framework, could be useful for the SFLR methodology (Bien, Taylor, and Tibshirani 2013).

Using a selection criterion that can better balance model fit and model complexity would certainly be beneficial for the SFLR method. One way of doing this has been demonstrated that considers the mean square error of the local coefficients. However, this method is not practical for applied work, since we generally do not know what the true coefficients are and therefore cannot compute the mean square error between the true coefficients and the coefficient estimates. In GWR there is the AICc, which penalizes the fit statistic based on the effective degrees of freedom and is expected to be larger than the degrees of freedom for a corresponding global model. SFLR may benefit from a similar correction.

Another serious issue of the SFLR framework is that of replicability. Since eigenvectors are not necessarily unique to eigenvalues, different algorithms may produce different eigenvectors so it is necessary to be aware of which software was employed for a given study. One example of this limitation was discovered first hand when trying to validate the SFLR methodology. Two different algorithms for extracting eigenvectors produced eigenvectors with opposite signs. While these are equivalent and do not effect the stepwise regression routine, it may affect future researchers' ability to reproduce results if they are unaware of this issue. More seriously, when testing the methodology on another data set not reported here, some of the eigenvectors were completely different depending on whether R or Python code was used, which can lead to different results.

Finally, there is the issue of regression diagnostics and hypothesis testing. While GWR extends diagnostics found in global models, no such extensions have yet been produced for the SFLR method. In particular, it is unclear how a significance test for a local parameter estimate from SFLR may be computed or how one can deal with the issue of multiple hypothesis testing, which has been addressed in GWR (da Silva and Fotheringham 2015). A significance test

customized for the SFLR method will be necessary in order to evaluate the robustness of the estimated coefficients produced by the method.

Conclusion

In its original conception, SFLR does not appear to produce superior results to GWR. Using Griffith's (2008) stepwise selection routine results in overfitted models where there is a severe loss of estimated coefficient accuracy in return for modest gains in model fit and inaccurately low levels of spatial autocorrelation. For the simulated data presented in this research, GWR is able to better capture the non-stationarity in local parameter estimates, and more accurately replicate the known coefficients when compared to Griffith's SFLR framework. Furthermore, GWR does so in a fraction of the time. Despite, Griffith's claim that SFLR and GWR produce substantially different outputs, the two sets of local parameter estimates are generally similar, albeit that the SFLR ones are generally not as accurate. This discrepancy seems to be due to an inherent flaw in the original SFLR methodology that causes overfitting and can require lengthy compute times of up to several days on a standard notebook computer.

It turns out that the SFLR framework *can* be approximately equivalent to GWR, however, when a MSE-minimizing stepwise selection routine is employed to select a subset of candidate synthetic variables. By doing so, the SFLR framework then produces local coefficient estimates that are similar to those from GWR in magnitude, overall accuracy, spatial autocorrelation, and that yield a similar model fit. The computing time is also drastically reduced from days to only minutes, which is comparable to GWR. It seems then that the SFLR framework can provide a sort of discrete spatial weighting mechanism in the form of a subset of eigenvectors and interaction terms, somewhat akin to GWR's continuously defined spatial weighting mechanism. However, GWR's spatial weighting mechanism produces a parameter representative of the scale at which processes occur and no such feature is available through the SFLR method. Given the potential of both GWR and SFLR to produce such similar results, this provides strong evidence that there is no a priori disadvantage to displaying spatial autocorrelation within a local coefficient surface or multicollinearity between surfaces. Even where local coefficient estimates may be artificially smoothed, such as in GWR, it does not necessarily greatly impact their accuracy. Further explorations of these conclusions are needed using simulated coefficients from different types of local processes that arise from more complex spatial relationships.

The analysis in this article has demonstrated the importance of differentiating between different types and sources of collinearity. In the case of SFLR, a narrow view of collinearity made it difficult to notice that the method could be severely susceptible to overfitting. Future investigations of local modeling techniques should be sure to more clearly define collinearity, why it arises, and how it effects different aspects of the modeling framework. It is possible that SFLR may be useful when the sample size is small and there is very high multicollinearity or there are very influential outliers. However, this will remain unknown unless the issue of overfitting can be remedied, perhaps with some of the suggestions proposed in this article. Despite the encouraging results using the MSE minimization criterion, we typically do not have access to the true coefficients that are needed to compute the MSE, so this criterion is only useful for theoretical work. Therefore, until SFLR can be specified with a feasible selection criterion that is robust to overfitting, it is not a viable framework for applied research, whereas GWR remains an effective method for capturing nonstationarity.

Notes

1 While equation (5) is the most commonly found transformed connectivity matrix, others have been defined (Tiefelsdorf and Griffith 2007) and employed (Chun 2008). In addition, $\boldsymbol{C}$ may be standardized using coding schemes, such as the W-coding scheme, which is also known as the row-sum standardization (Boots 1999).
2 For all MC values reported throughout this article, the null hypothesis that there is no spatial autocorrelation could be rejected based on significance tests using random permutations. The resulting z-scores were always greater than 50.
3 Validation of the routine was carried out by replicating the results presented in Griffith (2008) using the Georiga data set originally described by Fotheringham, Brunsdon, and Charlton (2002). Code was written in the Python programming language (version 2.7) using the numerical computation libraries, numpy (version 1.11.1) and scipy (0.18.0).
4 To be exact, it was the singular value decomposition operation that could not converge, which underlies the ordinary least squares solver in the statsmodels package of the Python programming language.
5 The algorithm terminates after two iterations without a reduction in the MSE, which helps reduce the chance of getting stuck in a local minimum, but does not guarantee a global minimum.

References

Belsey, D. A., E. Kuh, and R. E. Welsch. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Bien, J., J. Taylor, and R. Tibshirani. (2013). "A Lasso for Hierarchical Interactions." *The Annals of Statistics* 41(3), 1111–41.

Boots, B. (1999). "A Variance-Stabilizing Coding Scheme for Spatial Link Matrices." *Environment and Planning A* 31, 165–80.

Brown, S., V. L. Versace, L. Laurenson, D. Ierodiaconou, J. Fawcett, and S. Salzman. (2011). "Assessment of Spatiotemporal Varying Relationships Between Rainfall, Land Cover and Surface Water Area Using Geographically Weighted Regression." *Environmental Modeling & Assessment* 17(3), 241–54.

Cahill, M., and G. Mulligan. (2007). "Using Geographically Weighted Regression to Explore Local Crime Patterns." *Social Science Computer Review* 25(2), 174–93.

Chun, Y. (2008). "Modeling Network Autocorrelation Within Migration Flows by Eigenvector Spatial Filtering." *Journal of Geographical Systems* 10(4), 317–44.

da Silva, A. R., and A. S. Fotheringham. (2015). The Multiple Testing Issue in Geographically Weighted Regression: The Multiple Testing Issue in GWR. *Geographical Analysis* 48(3), 233–47.

Fotheringham, A. S., and C. Brunsdon. (1999). "Local Forms of Spatial Analysis." *Geographical Analysis* 31(4), 340–58.

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

Fotheringham, A. S., R. Crespo, and J. Yao. (2015). "Geographical and Temporal Weighted Regression (GTWR)." *Geographical Analysis* 47(4), 431–52.

Fotheringham, A. S., and T. M. Oshan. (2016). "Geographically Weighted Regression and Multiocollinearity: Dispelling the Myth." *Journal of Geograpical Systems* 18, 303–29.

Fotheringham, A. S., W. Yang, and W. Kang. (in press). Mutliscale GWR: A Model to Quantify Scale in Spatial Processes. *The Annals of the Association of American Geographers*.

Gareth, J., D. Witten, T. Hastie, and R. Tibshirani. (2013). *An Introduction to Statistical Learning: With Applications in R*. New York: Springer.

Gilbert, A., and J. Chakraborty. (2011). "Using Geographically Weighted Regression for Environmental Justice Analysis: Cumulative Cancer Risks from Air Toxics in Florida." *Social Science Research* 40(1), 273–86.

Griffith, D. A. (1996). "Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Geo-Referenced Data." *Canadian Geographer/Le Gographe Canadien* 40(4), 351–67.

Griffith, D. A. (2008). "Spatial-Filtering-Based Contributions to a Critique of Geographically Weighted Regression (GWR)." *Environment and Planning A* 40(11), 2751–69.

Griffith, D. A. (2011). "Visualizing Analytical Spatial Autocorrelation Components Latent in Spatial Interaction Data: An Eigenvector Spatial Filter Approach." *Computers, Environment and Urban Systems* 35(2), 140–9.

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.

Hodges, J. S., and B. J. Reich. (2010). "Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love." *The American Statistician* 64(4), 325–34.

Lu, B., P. Harris, M. Charlton, and C. Brunsdon. (2015). "Calibrating a Geographically Weighted Regression Model with Parameter-specific Distance Metrics." *Procedia Environmental Sciences* 26, 109–14.

Miller, J. A., and R. Q. Hanham. (2011). "Spatial Nonstationarity and the Scale of Species-Environment Relationships in the Mojave Desert, California, USA." *International Journal of Geographical Information Science* 25(3), 423–38.

Obrien, R. M.(2007). "A Caution Regarding Rules of Thumb for Variance Inflation Factors." *Quality & Quantity* 41(5), 673–90.

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.

Ramsay, T., R. Burnett, and D. Krewski. (2003a). "Exploring Bias in a Generalized Additive Model for Spatial Air Pollution Data." *Environmental Health Perspectives* 111(10), 1283–8.

Ramsay, T. O., R. T. Burnett, and D. Krewski. (2003b). "The Effect of Concurvity in Generalized Additive Models Linking Mortality to Ambient Particulate Matter." *Epidemiology* 14(1), 18–23.

Tiefelsdorf, M., and D. A. Griffith. (2007). "Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach." *Environment and Planning A* 39(5), 1193–221.

Wheeler, D. C. (2007). "Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression." *Environment and Planning A* 39(10), 2464–81.

Wheeler, D. C. (2010). "Visualizing and Diagnosing Coefficients from Geographically Weighted Regression Models." In *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, Vol. 99, 415–36, edited by B. Jiang and X. Yao. Dordrecht: Springer Netherlands.

Wheeler, D., and M. Tiefelsdorf. (2005). "Multicollinearity and Correlation Among Local Regression Coefficients in Geographically Weighted Regression." *Journal of Geographical Systems* 7(2), 161–87.