

Multiscale Geographically Weighted Regression (MGWR)

A. Stewart Fotheringham,* Wenbai Yang,[†] and Wei Kang*

*School of Geographical Sciences & Urban Planning, Arizona State University

[†]School of Geography & Geosciences, University of St. Andrews

Scale is a fundamental geographic concept, and a substantial literature exists discussing the various roles that scale plays in different geographical contexts. Relatively little work exists, though, that provides a means of measuring the geographic scale over which different processes operate. Here we demonstrate how geographically weighted regression (GWR) can be adapted to provide such measures. GWR explores the potential spatial nonstationarity of relationships and provides a measure of the spatial scale at which processes operate through the determination of an optimal bandwidth. Classical GWR assumes that all of the processes being modeled operate at the same spatial scale, however. The work here relaxes this assumption by allowing different processes to operate at different spatial scales. This is achieved by deriving an optimal bandwidth vector in which each element indicates the spatial scale at which a particular process takes place. This new version of GWR is termed multiscale geographically weighted regression (MGWR), which is similar in intent to Bayesian nonseparable spatially varying coefficients (SVC) models, although potentially providing a more flexible and scalable framework in which to examine multiscale processes. Model calibration and bandwidth vector selection in MGWR are conducted using a back-fitting algorithm. We compare the performance of GWR and MGWR by applying both frameworks to two simulated data sets with known properties and to an empirical data set on Irish famine. Results indicate that MGWR not only is superior in replicating parameter surfaces with different levels of spatial heterogeneity but provides valuable information on the scale at which different processes operate. *Key Words:* geographically weighted regression, multiscale, spatially varying coefficients, spatial nonstationarity.

尺度根本上是个地理的概念,且有众多的既有文献,探讨尺度在不同地理脉络中扮演的各种角色。但相对而言,少有研究提供不同过程操作的地理尺度之测量工具。我们于此展现如何采用地理加权回归(GWR)提供此般测量。GWR透过决定最适宽带,探讨各种关系的潜在空间非静止性,并提供各种过程操作的空间尺度测量。但古典的GWR,却预设所有进行模式化的过程,皆在相同的空间尺度中操作。本研究工作透果考量不同过程在不同空间尺度操作,为此一预设鬆绑。上述工作藉由导出最适宽带向量达成,其中各元素表明特定过程发生的空间尺度。此一崭新的GWR版本称为多重尺度地理加权回归(MGWR),其目的与贝叶斯不可分割的空间变异系数(SVC)模型相似,尽管可能提供了检视多重尺度过程的更为弹性且可进行尺度化的架构。本研究运用向后演算法,进行MGWR中的模式校正和宽带向量选择。我们将GWR和MGWR架构运用至两组具有已知属性的虚拟数据集和爱尔兰大飢荒的经验数据集,从而比较两者的表现。研究结果指出,MGWR不仅在複製具有不同层级空间变异性的参数表面上表现较佳,更提供了有关不同过程所进行的尺度的宝贵信息。 *关键词:* 地理加权回归, 多重尺度, 空间变异系数, 空间非静止性。

La escala es un concepto geográfico fundamental, y existe una literatura sustancial que discute los diversos roles que aquella juega en diferentes contextos geográficos. Con todo, es relativamente escaso el trabajo que suministre un medio para medir la escala geográfica a la cual operan diferentes procesos. Aquí demostramos cómo la regresión geográficamente ponderada (RGP) puede adaptarse para proveer tales medidas. La RGP explora el carácter espacial no estacionario de las relaciones y suministra una medida de la escala espacial a la cual operan los procesos mediante la determinación de un ancho de banda óptimo. Sin embargo, la RGP clásica asume que todos los procesos en proceso de modelación operan a la misma escala espacial. El trabajo desarrollado aquí flexibiliza esta suposición permitiendo que diferentes procesos operen a diferentes escalas espaciales. Esto se logra derivando un vector de ancho de banda óptimo en el que cada elemento indica la escala espacial a la que un proceso particular tiene lugar. Esta nueva versión de la RGP es denominada regresión geográficamente ponderada a multiescala (RGPM), la cual es de intento similar a los modelos bayesianos inseparables de coeficientes espacialmente variados (SVC), aunque potencialmente de lugar a un marco más flexible y escalable en el cual examinar procesos multiescalares. La calibración del modelo y la selección del vector del ancho de banda en la

RGPM se efectúan usando un algoritmo adaptado. Comparamos el desempeño de la RGP y la RGPM aplicando los dos marcos a dos conjuntos de datos simulados de propiedades conocidas y a un conjunto de datos empíricos de la hambruna irlandesa. Los resultados indican que la RGPM no solo es superior en replicar las superficies parámetro con diferentes niveles de heterogeneidad espacial, sino que provee información valiosa sobre la escala a la que operan diferentes procesos. *Palabras clave: regresión geográficamente ponderada, multiescala, coeficientes espacialmente variados, espacialidad no estacionaria.*

Scale is a fundamental geographic concept and is the focus of a huge and diverse literature that discusses the various roles that scale plays in different geographical contexts (e.g., Harvey 1968; Moellering and Tobler 1972; Brenner 2001; Tate and Atkinson 2001; Liverman 2004; Paasi 2004; Sheppard and McMaster 2004). Goodchild (2001) claimed, “Scale is perhaps the most important topic of geographical information science” (10), and McMaster and Sheppard (2004) stated, “Scale is intrinsic to nearly all geographical enquiry” (1). It is generally recognized that different processes can operate at different spatial scales and we often distinguish between micro- and macroprocesses or between local and global processes. We know that the weather and tides are determined by a multitude of processes operating at vastly different spatial scales. We talk about international, national, and local processes affecting our salaries and how much we pay for goods and services. We recognize that declining densities of some fish species might be a function of both global climate change and also local overfishing and that the impact of a disease on societies may be determined not only by international and national movement patterns but also by local supplies of medicines. Despite “geography” and “scale” being virtually synonymous, however, surprisingly little work exists that actually provides a means of measuring the geographic scale over which different processes operate.¹ Such information would be useful to better understand both the nature of geographic processes and our observations about the real world that are a product of these processes. One exception to this statement is the development of Bayesian nonseparable spatially varying coefficient (SVC) models (Gelfand, Kim, et al. 2003; Gelfand, Schimdt, et al. 2003; Finley 2011) that do identify the scale of each relationship separately. Further discussion of such models is left until later; the intention here is to develop the equivalent potential within a geographically weighted regression (GWR) framework, which is potentially more flexible and scalable.

GWR is one of the most widely applied methods for exploring the potential spatial nonstationarity of relationships (Fotheringham, Charlton, and Brunson 1996; Fotheringham, Brunson, and Charlton 2002; Atkinson et al. 2003; Foody 2003; Lloyd 2010, 2011; Fotheringham and Oshan 2016). Instead of producing an “average” global estimate of each relationship in the model, GWR allows that these relationships between the response variable and predictor variables might vary across space. To calibrate a GWR model at any one location, data are “borrowed” from nearby locations and weighted according to the distance each nearby location is from the regression point. This is in accordance with Tobler’s first law of geography that “everything is related to everything else, but near things are more related than distant things” (Tobler 1970). Hence, not only does GWR identify spatial heterogeneity in processes but it also takes advantage of the spatial dependence in data—so tying together the two main distinguishing features of spatial analysis.

In an early recognition that GWR could be made more flexible in terms of the spatial processes being investigated, semiparametric GWR (SGWR) was developed, extending the traditional GWR framework by allowing a subset of the parameters to be fixed over space and a subset to vary over space (Brunson, Fotheringham, and Charlton 1999; Fotheringham, Brunson, and Charlton 2002). SGWR still has the limitation, though, that all of the spatially varying parameters are assumed to arise from processes operating at the same spatial scale. This does not allow, for example, one process to operate over a local scale and another to operate over a broader, regional scale. It seems quite reasonable to imagine, however, that some relationships, such as the effect of rainfall on vegetation density, operate at one spatial scale, whereas others, such as competition from surrounding plants, operate at different spatial scales. By relaxing the assumption that all the spatially varying processes in a model operate at the same spatial scale, we generate the potential to produce a more powerful spatial model. This is the essence of multiscale geographically weighted regression (MGWR).

The remainder of the article is organized as follows. We first introduce the MGWR model formulation and expand on the methodological issues involved in model calibration and bandwidth selection. We then describe two simulated data set designs that are used to compare the performance of GWR and MGWR. Next an empirical application of MGWR is given in using a data set on the determinants of population decline during the Irish famine. We close the article with a summary of the key findings and thoughts on directions for future research.

MGWR

Model Formulations

GWR

Traditional or global regression assumes that the relationships being examined through the model's parameters are constant over space. This assumption is relaxed in GWR by allowing the parameters to vary spatially. The GWR model formulation can be described as follows. Assuming that there are n observations, for the observation $i \in \{1, 2, \dots, n\}$ at location $(\mathbf{u}_i, \mathbf{v}_i)$, the linear regression model is

$$y_i = \sum_{j=0}^m \beta_j(\mathbf{u}_i, \mathbf{v}_i)x_{ij} + \varepsilon_i, \quad (1)$$

where x_{ij} is the j th predictor variable, $\beta_j(\mathbf{u}_i, \mathbf{v}_i)$ is the j th coefficient, ε_i is the error term, and y_i is the response variable.

SGWR

SGWR is an extension of GWR that allows for the coexistence of local and global relationships. It can be considered as a special case of MGWR. For the observation $i \in \{1, 2, \dots, n\}$ at location $(\mathbf{u}_i, \mathbf{v}_i)$, the linear regression model is

$$y_i = \sum_{j=1}^{k_a} a_j x_{ij}(a) + \sum_{l=1}^{k_b} b_l(\mathbf{u}_i, \mathbf{v}_i) x_{il}(b) + \varepsilon_i, \quad (2)$$

where k_a is the number of global predictor variables, k_b is the number of local predictor variables, $x_{ij}(a)$ is the j th global predictor variable, $x_{il}(b)$ is the l th local predictor variable, a_j is the j th global coefficient, $b_l(\mathbf{u}_i, \mathbf{v}_i)$

is the l th local coefficient, ε_i is the error term, and y_i is the response variable.

MGWR

Whereas both GWR and SGWR constrain the local relationships within each model to vary at the same spatial scale, MGWR allows the conditional relationships between the response variable and the different predictor variables to vary at different spatial scales (Yang 2014). That is, the bandwidths indicating the data-borrowing range can vary across parameter surfaces.

$$y_i = \sum_{j=0}^m \beta_{bwj}(\mathbf{u}_i, \mathbf{v}_i)x_{ij} + \varepsilon_i, \quad (3)$$

where bwj in β_{bwj} indicates the bandwidth used for calibration of the j th conditional relationship.

Operationalizing MGWR

Bandwidth Selection

Bandwidth selection is relatively straightforward in GWR and SGWR because only a single bandwidth is required. The optimal bandwidth is selected through trials: In each trial, a bandwidth is selected, then either GWR or SGWR is fitted using the bandwidth, then a goodness-of-fit measure such as AICc is calculated where AICc is defined by

$$\text{AICc} = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \frac{n + \text{tr}(\mathbf{S})}{n - 2 - \text{tr}(\mathbf{S})}, \quad (4)$$

where $\hat{\sigma}$ is the estimated standard deviation of the error term and $\text{tr}(\mathbf{S})$ is the trace of the hat matrix \mathbf{S} . The optimal bandwidth is that which minimizes AICc. It is impracticable to use the same process for the bandwidth selection in MGWR, however, because the number of potential combinations of bandwidths for the different processes will often be very large. A different procedure is hence needed.

Model Calibration

Model calibration for a Gaussian GWR can be conducted using weighted least squares. The estimator for the coefficients at location $(\mathbf{u}_i, \mathbf{v}_i)$ is shown in Equation (5) where \mathbf{X} is the design matrix and \mathbf{W}

(u_i, v_i) is the spatial weighting matrix for location (u_i, v_i) . $W(u_i, v_i)$ is the same for each relationship due to the same bandwidth being used for all the relationships in the model.

$$\hat{\beta}(u_i, v_i) = [(X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y]. \quad (5)$$

Thus, each row of the hat matrix S_{GWR} for a basic Gaussian GWR calibration is defined as in Equation (6) where X_i is the i th row of the design matrix X :

$$r_i = X_i [(X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i)]. \quad (6)$$

The calibration for a Gaussian SGWR is less straightforward, involving $k_a + 2$ basic GWR calibrations and 1 ordinary least squares (OLS) calibration (Fotheringham, Brundson, and Charlton 2002). The procedures could be described as follows:

1. Regress each global predictor variable $x_j(a)$ ($j = 1, \dots, k_a$) against the local predictor variables $X(b)$ using basic GWR and obtain the residuals $\epsilon_{X(a)}$. Similarly, regress the response variable y against local predictor variables $X(b)$ using basic GWR and obtain the residuals ϵ_y .
2. Obtain the global coefficient vector estimate \hat{a} by regressing ϵ_y against $\epsilon_{X(a)}$ using OLS.
3. Obtain the local coefficient matrix estimate \hat{b} by regressing $y - X(a)\hat{a}$ against $X(b)$ using basic GWR.

The hat matrix for the SGWR calibration can be derived as shown in Equation (7):

$$S_{SGWR} = X(a)(X(a)WX(a))^{-1}X(a)^T W + S_{GWR}^*, \quad (7)$$

where $W = (I - S_{GWR}^*)^T (I - S_{GWR}^*)$ and S_{GWR}^* is the hat matrix for the last step.

In MGWR, however, different bandwidths imply that each relationship at the same location will have a different spatial weighting matrix. Thus, the GWR estimator is not applicable in MGWR.

Calibration of MGWR Models: A Back-Fitting Algorithm

Back-fitting algorithms, which maximize the expected log likelihood, are commonly used for the calibration of

generalized additive models (GAMs; Hastie and Tibshirani 1986; Buja, Hastie, and Tibshirani 1989; Everitt 2005) and provide the means of calibrating an MGWR model. Following the logic of GAM, $\beta_{bwj} x_j$ in MGWR is defined as the j th additive term f_j , resulting in the GAM-style MGWR:

$$y = \sum_{j=0}^m f_j + \epsilon. \quad (8)$$

The basic idea of back-fitting is to calibrate each term in the model with a smoother assuming that all the other terms are known. In the case of the Gaussian MGWR model, the smoother is the GWR estimator defined in Equation (5). The back-fitting algorithm for the calibration of an MGWR model is thus defined as shown in Figure 1.

First, all of the additive terms need to be initialized, which means that all of the local coefficients need to be assigned initial estimates. Using these initial values, an initial set of estimates of y is obtained and a set of residuals is calculated. These residuals plus the “current” value of the first term f_0 are then regressed on x_0 using GWR, which produces an optimal bandwidth bw_0 for the relationship between y and x_0 and also a new set of local estimates for the relationship

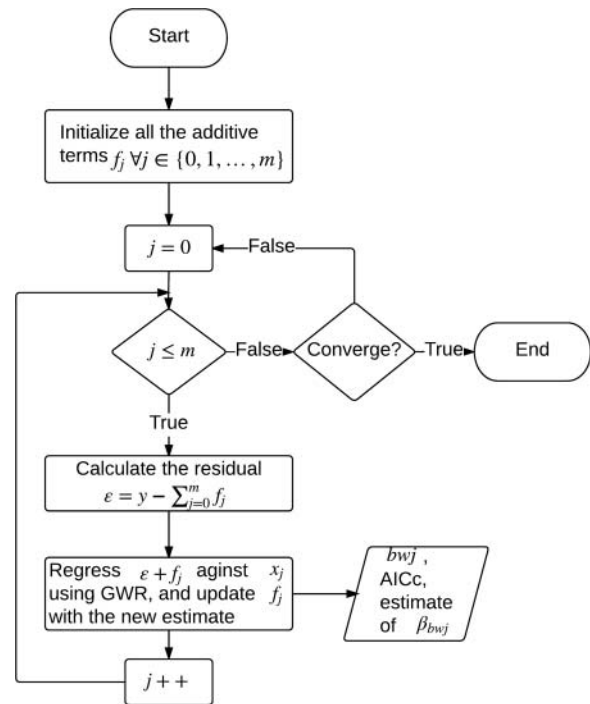


Figure 1. Back-fitting algorithm for a multiscale geographically weighted regression model calibration.

between \mathbf{y} and \mathbf{x}_0 that are used to update the value of the first term \mathbf{f}_0 . The process then moves on to the second variable \mathbf{x}_1 following the same procedure (a new set of residuals is computed using the updated \mathbf{f}_0 and these plus the value of \mathbf{f}_1 are regressed on \mathbf{x}_1 with GWR to create an optimal bandwidth $bw1$, etc.) and continues in this manner until the local parameters associated with the last variable \mathbf{x}_m are estimated, which completes the first iteration. The iterations continue until the changes of all the terms on successive iterations are sufficiently small to declare that convergence is reached.

Two decisions from the user are involved in the algorithm. The first concerns initializing the local coefficient estimates. Three options seem most obvious:

1. All of the estimates are set to zero.
2. Initial coefficient estimates are taken from the outcome of a global model.
3. Initial coefficient estimates are taken from the calibration of the model by GWR.

The choice of the initial guesses should not influence the selected optimal bandwidth vector but might affect the number of iterations needed to reach convergence. Here, we use the GWR estimates as the initial MGWR estimates; the justification for this is given next.

The second decision that the user needs to make is the choice of termination criterion—the value of the differential between successive iterations (score of change [SOC]) by which the process is deemed to have converged. Two types of SOC could be used:

1. SOC-RSS: proportional change in the residual sum of squares (RSS):

$$SOC_{RSS} = \frac{|RSS_{new} - RSS_{old}|}{RSS_{new}} \quad (9)$$

2. SOC-f: change in the GWR smoother:

$$SOC_f = \sqrt{\frac{\sum_{j=1}^p \sum_{i=1}^n (\hat{f}_{ij}^{new} - \hat{f}_{ij}^{old})^2}{n} \bigg/ \sum_{i=1}^n \left(\sum_{j=1}^p \hat{f}_{ij}^{new} \right)^2} \quad (10)$$

Both SOC-RSS and SOC-f are scale-free, although the latter has the advantage of being focused on the relative changes of the additive terms rather than on overall model fit. Consequently, in what follows we use SOC-f as the termination criterion.

One further issue in the operation of MGWR models is that the interpretation and comparison of the individual bandwidths is facilitated by standardizing all of the variables in the model to have mean = 0 and standard deviation = 1. This standardization allows the bandwidths to be direct indicators of the spatial scale at which the conditional relationship between \mathbf{y} and each predictor variable varies. Without standardization, the bandwidths will also reflect the scale and variation in each predictor variable.

Implementation in Python

The back-fitting algorithm for MGWR just described was programmed in the Python environment.² MGWR is organized as a Python class. Class attributes, including the starting values of the local parameter estimates and the convergence criterion, can be assigned by users.

Simulation Design

To demonstrate the performance of MGWR and to compare it with GWR, several questions need to be addressed:

1. Does MGWR produce reliable estimates of scale through the independent bandwidths for each covariate and how do these compare to the single optimized bandwidth for GWR?
2. Does MGWR produce more accurate estimates of local parameter surfaces than GWR?
3. Does MGWR produce more accurate estimates of the response variable than GWR?
4. MGWR is a big model—how much longer does it take to run than GWR?

These questions are examined by applying MGWR and GWR to a series of simulated data sets with known properties and varying levels of spatial heterogeneity across local parameter surfaces. To construct these simulated databases, the basis for the data generating process (DGP) is the GWR-like linear model in Equation (1). The spatial layout consists of a regular 25×25 lattice.³

We produced m parameter surfaces of different levels of spatial heterogeneity, generated normally distributed predictor variables x as well as a normally distributed error term ε , and then computed the response variable y based on the DGP. Because we have complete control over the DGP, the evaluation and comparison of the model calibration results is straightforward.

Simulation Design 1

This simulated data set is used to examine the relative performances of GWR and MGWR when the DGP is such that the local parameter surfaces exhibit varying degrees of spatial heterogeneity. Specifically, the DGP is

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \varepsilon_i. \quad (11)$$

We designed three different parameter surfaces for each of the covariates with zero, medium, and high spatial heterogeneity based on the following rules:

$$\beta_{zero} = 3 \quad (12)$$

$$\beta_{low} = 1 + \frac{1}{12}(u + v) \quad (13)$$

$$\beta_{high} = 1 + \frac{1}{324} \left[36 - \left(6 - \frac{u}{2} \right)^2 \right] \left[36 - \left(6 - \frac{v}{2} \right)^2 \right], \quad (14)$$

where v is the value of the vertical coordinate (with values increasing from north to south with an increment of 1) and u being the value of the horizontal coordinate (with values increasing from west to east with an increment of 1). Each of these parameter surfaces is visualized in Figure 2.

We assigned each of the three parameter surfaces to the coefficients of the three predictor variables as follows: $\beta_0 = \beta_{zero}$, $\beta_1 = \beta_{low}$, and $\beta_2 = \beta_{high}$. The values of x_1 and x_2 were generated randomly from a normal distribution $N(0, 1)$ and the error term was generated

from a normal distributed $\varepsilon \sim N(0, 0.5^2)$. The response variable y was then computed based on Equation (11). We generated 100 different data sets in this way to examine the robustness of the results to the random nature of the DGP.

Simulation Design 2

In Simulation Design 1, the parameter surfaces are generated in a manner such that they have unequal degrees of spatial heterogeneity and are used here to examine the relative performances of GWR and MGWR with the expectation that MGWR should provide more accurate information on the processes being modeled. To check that MGWR does not provide spurious results when the processes modeled have the same degree of spatial heterogeneity, we develop a second simulation. In this case, the DGP is

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \varepsilon_i \quad (15)$$

and the two local parameter surfaces have the same degree of spatial heterogeneity as shown in Figure 3. To achieve this, we use the same surface for β_0 as that derived in Simulation Design 1 in Equation (13) and then we rotate this surface 90 degrees to obtain the local values for β_1 . This procedure ensures that the two surfaces have equal degrees of spatial heterogeneity but are uncorrelated. In this case n is again 625.

The values of x_1 were generated from the standard normal distribution $N(0, 1)$ and the error term was simulated from a normal distribution with $\varepsilon \sim N(0, 0.5^2)$. The response variable y was computed from Equation (11) with known values for each of the local parameters and for each of the covariates. Again, 100 different data sets were generated to examine the consistency of the results and to

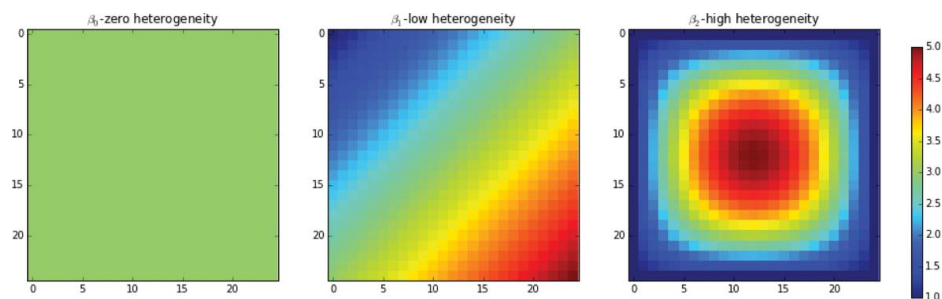


Figure 2. Simulation 1: True parameter surfaces of zero, low, and high heterogeneity. (Color figure available online.)

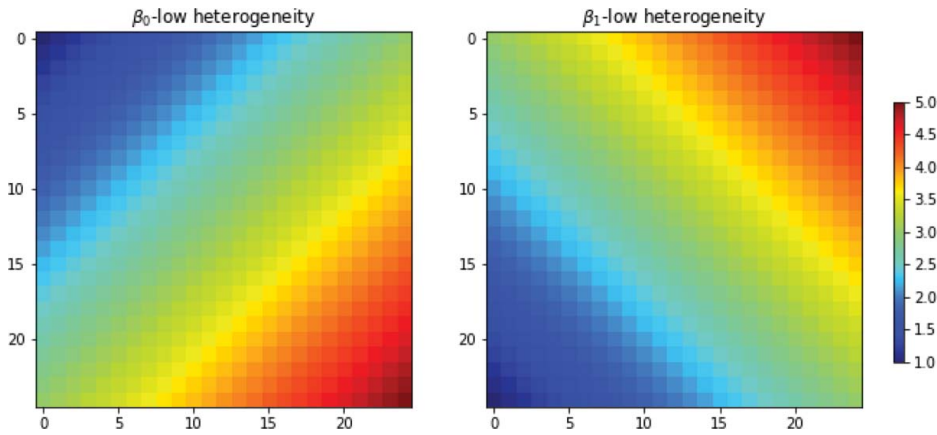


Figure 3. Simulation 2: True parameter surfaces of equal heterogeneity. (Color figure available online.)

safeguard against the results being specific to one particular data set.

Evaluation Methods

Several evaluation methods were employed to address the questions listed earlier regarding the relative performance of both MGWR and GWR using the data sets drawn from both simulation designs.

Bandwidth Comparison

Because MGWR relaxes the assumption of a single bandwidth for all the relationships being modeled, we expect MGWR to be able to differentiate between relationships that are relatively homogeneous and those that are relatively heterogeneous and to be able to differentiate between processes that operate at a local scale and those that operate at a more regional scale. Also of interest is to see to what extent the single bandwidth from GWR is an average of the individual bandwidths associated with each of the covariates in the model. To aid these comparisons, we adopt an adaptive bandwidth for all calibrations that is based on the optimal number of nearest neighbors (Fotheringham, Brundson, and Charlton 2002). This generates an intuitive optimal bandwidth that is the number of nearest neighbors that contribute to the regression results at i . For example, if the optimal bandwidth were 100, this indicates that the model weights the closest 100 data points around i between 1 and 0 with higher weights assigned to closer data points and the weight of the 100th furthest data point being zero. The adaptive

bisquare kernel is defined for this purpose as follows:

$$w_{ij} = \begin{cases} \left[1 - \left[\frac{d_{ij}}{G_i} \right]^2 \right]^2, & \text{if } d_{ij} < G_i, \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

where d_{ij} is the distance between point i and j , and G_i is the distance from focal point i to its M th nearest neighbor. M is the optimal number of nearest neighbors, determined by multiple AICc comparisons.

Local Parameter Estimation Accuracy

The ability of both GWR and MGWR to replicate the known parameter surfaces is measured by the root mean squared error (RMSE) of the coefficient β_j :

$$RMSE_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (\beta_j(u_i, v_i) - \hat{\beta}_j(u_i, v_i))^2}, \quad (17)$$

where $\hat{\beta}_j(u_i, v_i)$ is the estimated coefficient for location i from either GWR or MGWR. A smaller $RMSE_j$ value indicates a more accurate replication of the known set of local parameters $\beta_j(u_i, v_i)$, $i \in \{1, 2, \dots, n\}$.

Goodness of Fit

The RSS, as shown in Equation (18), is used to evaluate the goodness of fit of both models to the known set of \mathbf{y}_i values (with $\hat{\mathbf{y}}_i$ being the estimated response variable). Although RSS is not a perfect indicator of goodness of fit in that it is neither unit-free nor takes model complexity into account, it serves to give insight into the relative performances of the two local modeling

approaches. Although we would have preferred to use a goodness-of-fit criterion such as AICc that accounts for model complexity, the calculation of AICc remains a research challenge for MGWR because of the separate weight matrices used for each covariate.

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (18)$$

Computational Efficiency

Computational efficiency is measured by the time required for one model calibration run. As the back-fitting algorithm for MGWR is an iterative procedure, it is expected that the time cost in calibrating MGWR will be significantly higher than that for GWR, although if convergence can be reached quite quickly, run times might be tolerable.

Results

Initialization and Convergence Criteria

To compare the relative performances of GWR and MGWR we first established viable values for the stopping criterion (SOC-f) and the starting values of the local parameter estimates in the MGWR routine. Table 1 presents typical evidence from across the 100 calibrations of the MGWR model using the sets of simulated data in Design 1. The results are drawn from one of the 100 calibrations (data set 4 was selected arbitrarily as being representative of the results across all 100 simulations). Table 1 shows the individual AICc values and the optimized bandwidths at each iteration using zeros as the initial guesses for all local parameter estimates. The results converge after five iterations, suggesting a viable stopping point of 10^{-5} using the SOC-f criterion. Consequently, this value is used to derive all subsequent results.

The second decision to be made in the implementation of the MGWR algorithm is the choice of starting values for the local parameter estimates. Table 2 shows the representative results again for Simulation Design 1 where three options for the starting values of the local parameter estimates are explored: using zeros for all estimates, using the GWR estimates, and using the OLS estimates. For this particular data set, the optimized bandwidth values converge at 536, 62, and 44 for the local intercept, the local relationship between y and x_1 , and the local relationship between y and x_2 , respectively. This indicates that the local intercept varies at a much broader spatial scale than do the other two local parameters. There is little to choose among the three different sets of starting values as they lead to the same result, although the use of zeros and OLS estimates as starting values takes one iteration longer to converge. Henceforth, we use the GWR local estimates as the starting values of the MGWR.

Simulation Design 1

We now describe the results of calibrating the model in Equation (11) by OLS, GWR, and MGWR using the 100 random data sets resulting from applying the DGP described in Simulation Design 1 where the parameter surfaces have unequal degrees of spatial heterogeneity. In each case we use $SOC-f \leq 10^{-5}$ as the termination criterion and we use the GWR estimates as the starting point of the iterative optimal bandwidth calibration routine.

Optimal Bandwidth Vector

Figure 4 shows the resulting bandwidths from the calibration of a GWR model and an equivalent MGWR model on the 100 simulated data sets for Design Process 1. In Figure 4, \mathbf{b}^* is the single optimal

Table 1. Bandwidth vector selection process for Simulation 1 data set 4 (initial guesses: zeros)

Iterations	AICc ₀	AICc ₁	AICc ₂	b ₀	b ₁	b ₂	SOC_f
1	3,544.475178	3,107.084069	1,223.544247	622	310	46	0.039566412
2	1,118.033996	966.721046	984.2842977	444	62	44	0.002339227
3	904.1492707	954.050443	984.5833555	524	62	44	0.000155162
4	902.270467	953.5527079	984.7089459	536	62	44	1.70×10^{-5}
5	902.1610436	953.4854188	984.7075472	536	62	44	2.03×10^{-6}
6	902.1504341	953.4776078	984.7065979	536	62	44	3.29×10^{-7}
7	902.148657	953.4763536	984.7063954	536	62	44	5.60×10^{-8}

Table 2. Impact of initial guesses on Simulation 1 data set 4 ($\text{SOC-f} \leq 10^{-5}$)

Iterations	Zeros			GWR			OLS		
	b_0	b_1	b_2	b_0	b_1	b_2	b_0	b_1	b_2
1	622	310	46	260	62	44	413	146	46
2	444	62	44	524	62	44	512	62	44
3	524	62	44	536	62	44	536	62	44
4	536	62	44	536	62	44	536	62	44
5	536	62	44				536	62	44

Note: GWR = geographically weighted regression; OLS = ordinary least squares.

bandwidth obtained from the GWR calibration and b_0^* , b_1^* , and b_2^* are the optimal bandwidths for each of the three relationships obtained in MGWR. To see more clearly the differences between the estimates of the single bandwidth in GWR and the two separate bandwidths for b_1^* and b_2^* in MGWR, the three sets of optimal bandwidths are presented in Figure 5 without that for the intercept being displayed. From both Figure 4 and Figure 5 it is clear that MGWR correctly identifies the three different scales at which the local parameters vary. The optimal bandwidth for the local intercept is very large with the median being close to 625, the maximum value, indicating a global process, whereas the optimal bandwidths for the relationship between y and x_1 and between y and x_2 are relatively small, indicating more local processes. The optimal

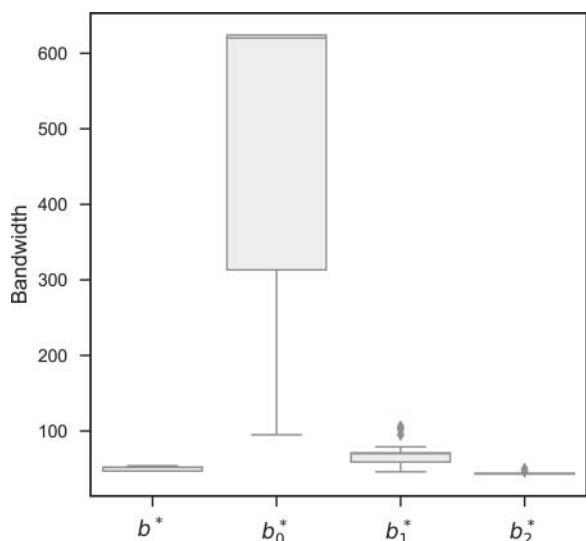


Figure 4. Simulation 1: Optimal bandwidth b^* from geographically weighted regression and optimal bandwidth vector $[b_0^*, b_1^*, b_2^*]$ from multiscale geographically weighted regression.

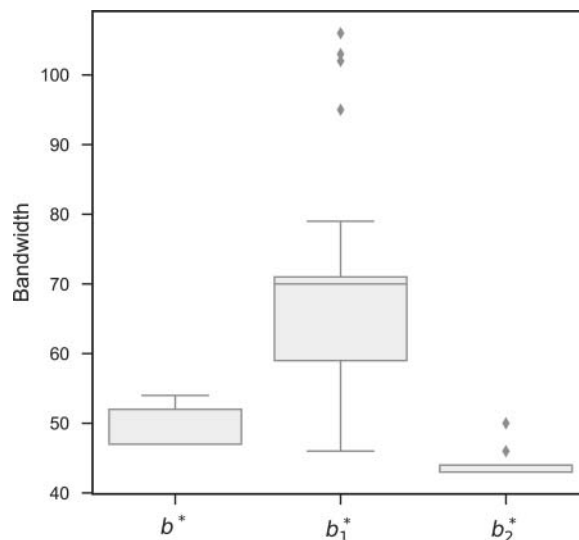


Figure 5. Simulation 1: Optimal bandwidth b^* from geographically weighted regression and optimal bandwidths b_1^* and b_2^* from multiscale geographically weighted regression.

bandwidth for the relationship between y and x_2 is consistently smaller than that between y and x_1 , indicating that the former process is slightly more localized—as evidenced in Figure 2. The single optimal bandwidth obtained in the GWR calibration can be seen as a weighted average of the different levels of spatial heterogeneity exhibited by the three separate processes with the weighting being a function of the explanatory power of each relationship in the local model.

Local Parameter Estimation Accuracy

The accuracy with which both MGWR and GWR replicate the three known parameter surfaces, β_0 , β_1 , and β_2 , for each of the 100 simulations is shown by the boxplots of RMSE values in Figure 6. Each RMSE value represents the average error in the ability of a model to replicate a known parameter surface. It is clear that, overall, MGWR replicates the three surfaces more accurately than does GWR, an assertion reinforced by the visualization of the actual and predicted surfaces for one representative simulation in Figure 7.

The results from Figures 6 and 7, however, show that the improvement in the replication of the parameter surfaces when using MGWR compared to GWR is not the same for the three surfaces. GWR replicates the surface for both β_1 and β_2 accurately but does a relatively poor job of replicating the surface for β_0 . MGWR, because of its relationship-specific bandwidth estimation, replicates all three surfaces accurately. The reason

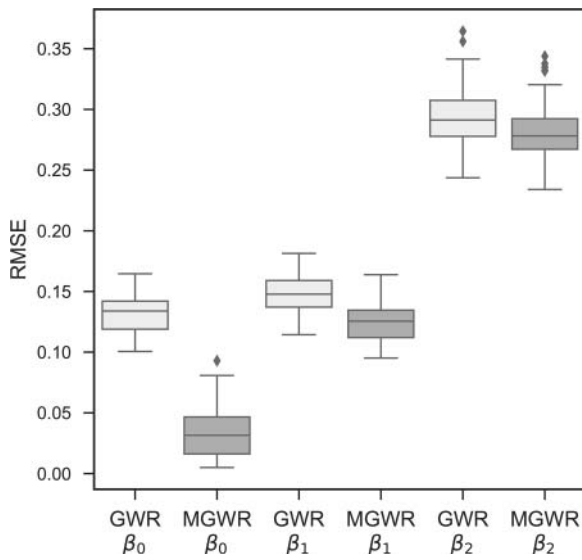


Figure 6. Simulation 1: Comparison of root mean squared error from geographically weighted regression and multiscale geographically weighted regression for each parameter surface. RMSE = root mean square error; GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression.

that GWR is able to provide accurate estimates of the surfaces of β_1 and β_2 is that the “average” bandwidth obtained in the GWR calibration is similar to the relationship-specific bandwidths obtained for both β_1 and β_2 in MGWR. The appropriate bandwidth for β_0 is much larger than the average bandwidth obtained in GWR. Had the surfaces been assigned differently with β_2 , say, exhibiting a broad regional trend, then GWR would have yielded a poor replication of this surface.

Goodness of Fit

Figure 8 displays the values of the RSS for the OLS, GWR, and MGWR models using each of the 100 simulated data sets. Also included are the values of the noise introduced into each data set ($\sum \varepsilon_i^2$). It is clear that OLS is a relatively poor model to apply to these data because of the inherent spatial nonstationarity but that both GWR and MGWR replicate the data accurately. To see the comparison between the GWR and MGWR results more clearly,

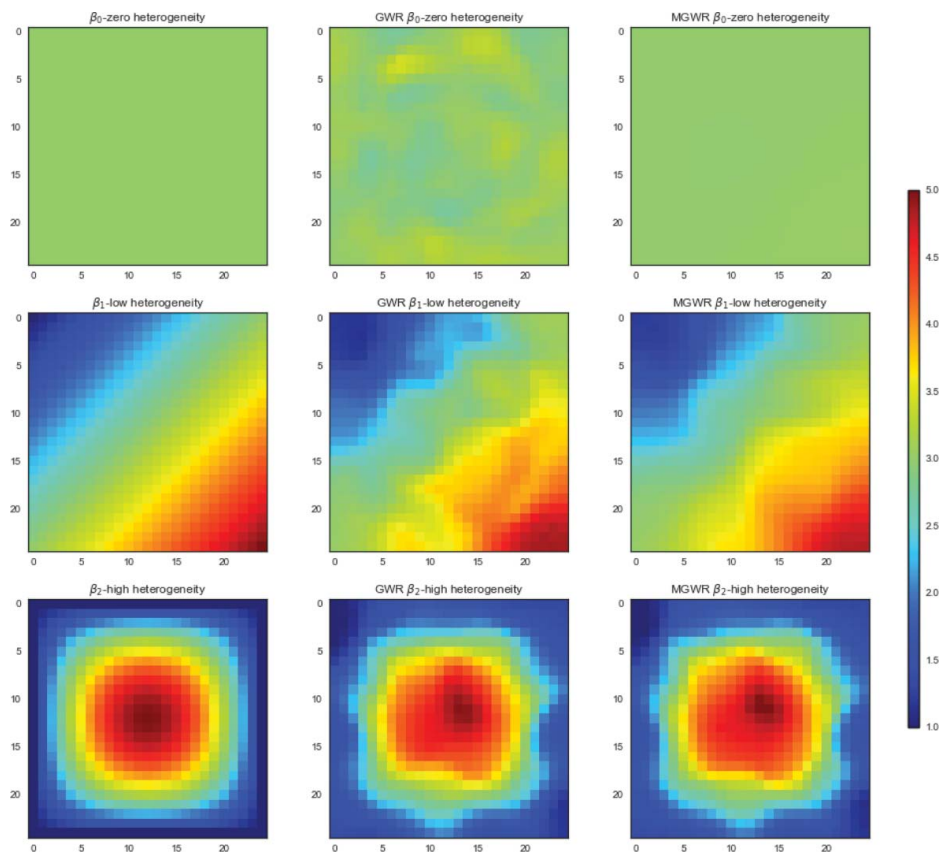


Figure 7. Simulation 1 data set 4, from left to right: true parameter surface, estimated surface from geographically weighted regression, and estimated surface from multiscale geographically weighted regression. GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression. (Color figure available online.)

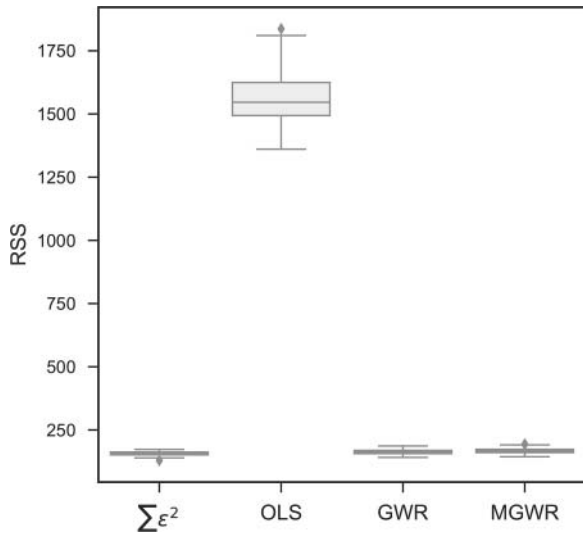


Figure 8. Simulation 1: Sum of squared random errors and residual sum of squares for the ordinary least squares, geographically weighted regression, and multiscale geographically weighted regression models. RSS = residual sum of squares; OLS = ordinary least squares; GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression.

the RSS values are repeated in Figure 9 omitting the OLS results. Although both models replicate the known y values accurately, it is interesting to note that the GWR results appear to be slightly superior to those of MGWR. The reason for this is presumably that in GWR the local intercept varies over space

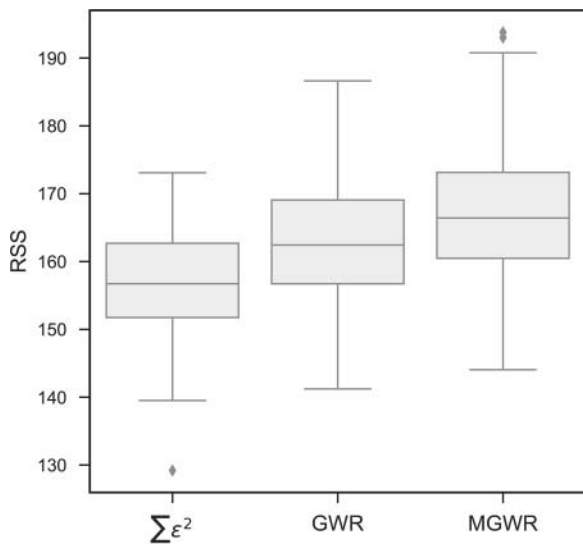


Figure 9. Simulation 1: Sum of squared random errors and residual sum of squares for the geographically weighted regression and multiscale geographically weighted regression models. RSS = residual sum of squares; GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression.

and “captures” some of the residual effect in the model, whereas in MGWR, the local intercept is essentially constant over space. In essence, the slightly better replication of the y variable from GWR is because of overfitting and not due to any better insights into the processes generating the data.

Computational Efficiency

MGWR is clearly a more complex model than GWR, and one element of comparison that is useful to note is the computational resources needed to fit both models. We show the time taken for each of the 100 runs for both models in Figure 10. Using the GWR estimates as the initial values of the local parameters and a convergence criterion of $SOC-f \leq 10^{-5}$, the average time needed for an MGWR calibration on data sets generated from Simulation 1 is about 0.08 hour or 4.8 minutes, which is not intolerable, although it is about eight times that needed for a GWR calibration.⁴

Simulation Design 2

We now describe the results of calibrating the model by OLS, GWR, and MGWR using the 100 random data sets resulting from applying the DGP described in Simulation Design 2 where the parameter surfaces have equal spatial heterogeneity. We again

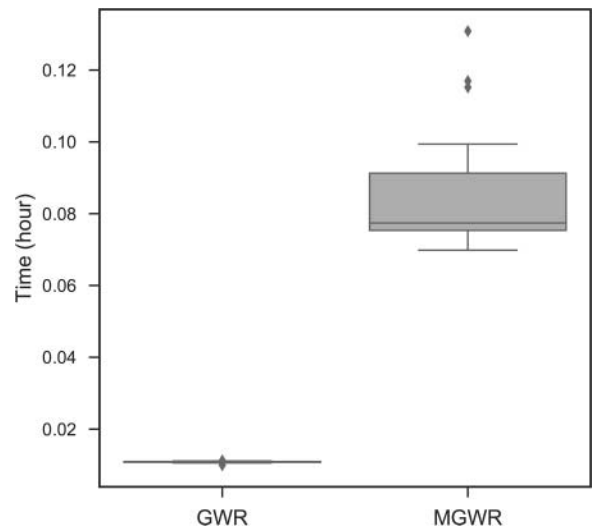


Figure 10. Simulation 1: Time for each of the 100 runs for geographically weighted regression and multiscale geographically weighted regression models. GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression.

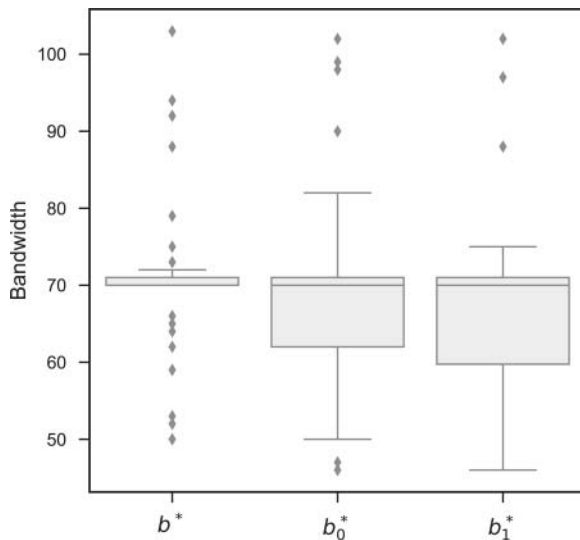


Figure 11. Simulation 2: Optimal bandwidth b^* from geographically weighted regression and optimal bandwidth vector $[b_0^*, b_1^*]$ from multiscale geographically weighted regression.

use $\text{SOC-}f \leq 10^{-5}$ as the termination criterion and the GWR estimates as the starting point of the iterative optimal bandwidth calibration routine.

Optimal Bandwidth Vector

Figure 11 displays the single optimal bandwidth for GWR and the two separate optimal bandwidths generated by MGWR for each of the 100 simulated data sets produced from Simulated Design 2. In this case, because

the two parameter surfaces have the same degree of spatial heterogeneity, the single bandwidths from GWR and the two separate bandwidths from MGWR are very similar with medians of seventy nearest neighbors. There is slightly more variability in the two separate bandwidths produced by MGWR across the 100 simulations probably because the GWR bandwidths are the average of the two MGWR bandwidths. The similarity in the results is borne out by a comparison of the known parameter surfaces and the predicted surfaces for one representative data set shown in Figure 12.

Local Parameter Estimation Accuracy

The ability of GWR and MGWR to replicate each of the two known parameter surfaces is shown in Figure 13, which depicts the RMSE values for both parameter surfaces for each of the 100 simulations. As expected, given the equal degree of heterogeneity in the two surfaces, the two models produce very similar results in terms of replicating the known surfaces of β_0 and β_1 . This is reinforced by the maps of the actual surfaces of the two sets of local parameters and their estimated values from both GWR and MGWR shown in Figure 12.

Goodness of Fit

Figure 14 shows the RSS values for the OLS, GWR, and MGWR calibrations against the sum of squared error terms added to the model. The results are very

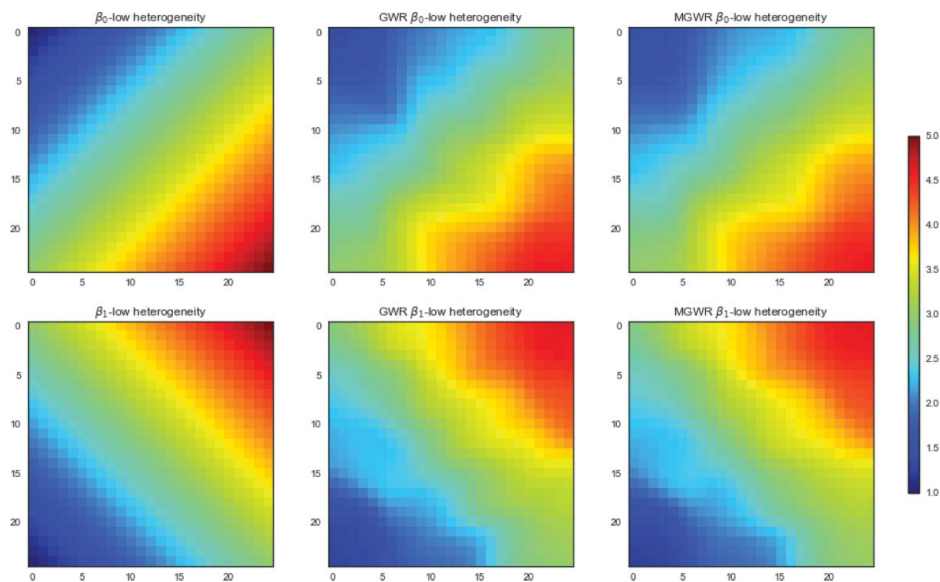


Figure 12. Simulation 2 data set 4, from left to right: true parameter surface, estimated surface from geographically weighted regression, and estimated surface from multiscale geographically weighted regression. GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression. (Color figure available online.)

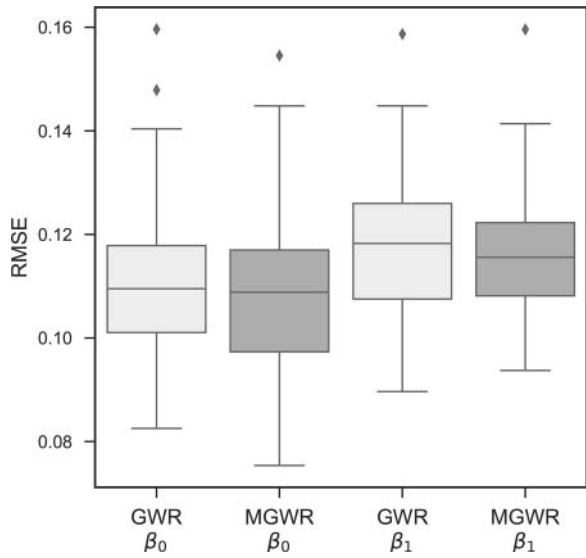


Figure 13. Simulation 2: Comparison of root mean squared error from geographically weighted regression and multiscale geographically weighted regression for each parameter surface. RMSE = root mean square error; GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression.

similar to those described for the earlier experiment—OLS is relatively poor at replicating the values of the dependent variable when those values are generated from a spatially varying process, whereas both GWR and MGWR replicate the surfaces with a very high degree of accuracy.

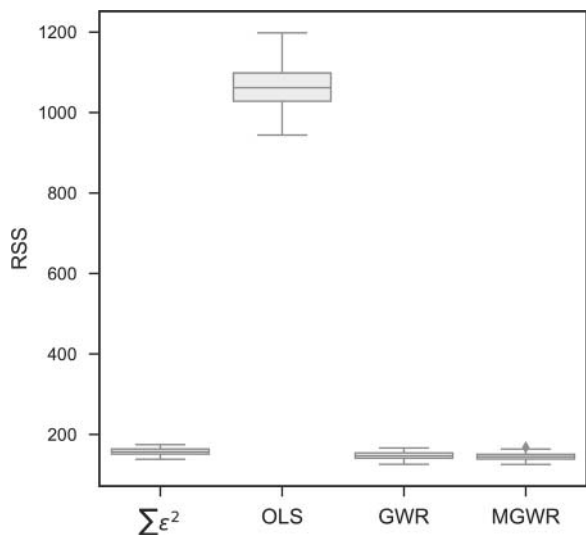


Figure 14. Simulation 2: Sum of squared random errors and residual sum of squares for the ordinary least squares, geographically weighted regression, and multiscale geographically weighted regression models. RSS = residual sum of squares; OLS = ordinary least squares; GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression.

Computational Efficiency

Figure 15 shows the run time for each of the 100 model calibrations for both GWR and MGWR. Because the model is very simple, the run times are lower than in the previous experiment. The median run time for the GWR calibrations is 0.6 minutes, whereas the median of the MGWR run times is around 3.15 minutes.

An Empirical Example of the Irish Famine

To provide an empirical application of MGWR to real data, we used data on population change between 1841 and 1851 for 2,317 electoral divisions (EDs) in the Republic of Ireland, which covers the period of the Great Famine (1845–1850), during which there was a large but spatially uneven decline in the population of Ireland due to death and emigration. To examine the spatial variation in population loss, for each ED we have data on eight predictor variables shown to be significantly related to population change in an earlier study by Fotheringham, Kelly, and Charlton (2013) and described in Table 3.⁵ An adaptive bandwidth was used for both the GWR and MGWR calibrations. The GWR calibration yielded an optimal bandwidth of 136 nearest neighbors, implying fairly localized processes. The MGWR model calibration converged after forty iterations and

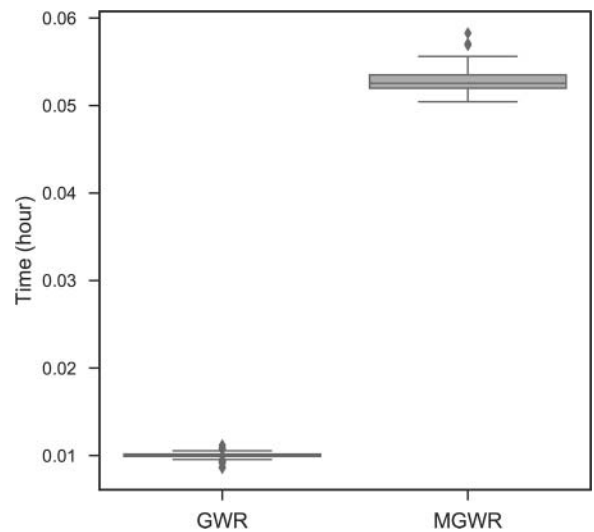


Figure 15. Simulation 2: Time for each of the 100 runs for geographically weighted regression and multiscale geographically weighted regression models. GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression.

Table 3. List of eight predictor variables for Irish famine case study

Index	Predictor variable
1	Ln population density on cropped land
2	% land under grain cultivation
3	Ln average holding size
4	Ln workhouse proximity
5	Ln urban proximity
6	Ln potato cultivation
7	Ln distance to coast
8	Mean elevation

a run time of fifty-one hours. The individual optimal bandwidths specific to each parameter in the MGWR model are given in Table 4, where it can be seen that the processes being modeled operate at different spatial scales. The parameter estimates associated with the variables percentage of each ED under grain cultivation and distance to the coast are global: The optimal bandwidth in each case is as large as it could be (2,316 nearest neighbors). The relationships between population change and average holding size and workhouse proximity exhibit spatial nonstationarity but the processes vary at a broad regional scale (optimal bandwidths of 365 and 210 nearest neighbors, respectively). The other variables in the model have impacts on population change that vary over relatively short distances, with the optimal bandwidths

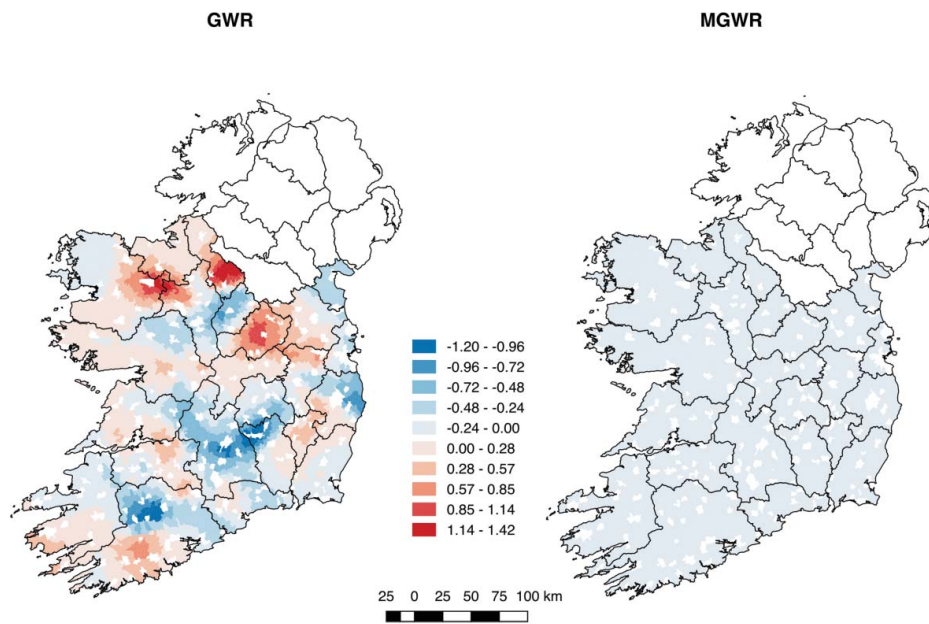
Table 4. Optimal bandwidth for each parameter surface obtained from geographically weighted regression and multiscale geographically weighted regression

Variable	Bandwidth	
	GWR	MGWR
Grain cultivation	136	2,316
Mean elevation	136	81
Average holding size	136	365
Potato cultivation	136	86
Urban proximity	136	66
Workhouse proximity	136	210
Population density on cropped land	136	91
Distance to coast	136	2,316

Note: GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression.

for mean elevation being eighty-one nearest neighbors; land under potato cultivation being eighty-six nearest neighbors; urban proximity being sixty-six nearest neighbors; and the percentage of cropped land being ninety-one nearest neighbors.

To see the variations in the spatial scales at which the different processes operate more clearly, in Figures 16 to 18 we present the local parameter estimates both from GWR and MGWR for three representative surfaces: the local parameter estimates associated with the variables “distance to the coast” (no variation), “workhouse proximity” (regional

**Figure 16.** Geographically weighted regression and multiscale geographically weighted regression local estimates for distance to coast. GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression. (Color figure available online.)

variation), and “mean elevation” (local variation), respectively. In Figure 16 it can be seen that the local parameter estimates for the variable “distance to the coast” derived from the MGWR model are uniformly negative across the country and are close to zero, indicating that proximity to the coast had little or no effect on population decline throughout the whole of the country. In contrast, the GWR estimates suggest that there is some spatial variation in the effect of distance to the coast, with it being positive in some parts of the country and negative in others. This is clearly the result of a single “average” bandwidth being derived in the GWR calibration, which for this relationship is a gross underestimate (136 nearest neighbors instead of 2,316).

In Figure 17, the local parameter estimates for the “workhouse proximity” variable are broadly similar from both the GWR and the MGWR calibrations because the bandwidth for the relationship between population decline and proximity to workhouses in the MGWR calibration is fairly close to that of the “average” bandwidth obtained in GWR (210 nearest neighbors for MGWR and 136 for GWR). The MGWR results, however, suggest a slightly more regional and weaker trend in the relationship with many of the local estimates being close to zero. There is some evidence, though, that workhouses had an ameliorating influence on population decline in parts of the west and the southwest but in the remainder of

the country the impact of workhouse proximity was either weakly negative or largely nonexistent.

Finally, Figure 18 describes the spatial variation in the local parameter estimates for mean elevation derived from both GWR and MGWR. The two patterns are similar because the MGWR bandwidth for this variable is close to that of the average GWR bandwidth (81 compared to 136) but there is a slightly clearer local effect in the MGWR parameter estimates. The general pattern is one where along the west coast, population decline was less severe at higher elevations, whereas in much of the rest of the country the effect was negligible or negative, the latter implying that population loss was more severe on relatively higher ground. The differential impact of elevation on population loss is most clearly seen in Roscommon, where in the far north of the county the relationship is strongly positive and in the rest of the county it is strongly negative. The positive relationships, mainly down the west of the country, tally with areas of generally higher relief and where the higher ground was less populated and more devoted to grazing rather than potato cultivation (Whelan 1997). Consequently, as the potato crop failed, causing massive levels of starvation and disease, the communities at higher elevations were somewhat more immune to the disaster. In the rest of the country where relief is generally lower, slightly higher ground might have had the advantage of better drainage for potato cultivation so that these areas were the hardest hit when the potato blight occurred.

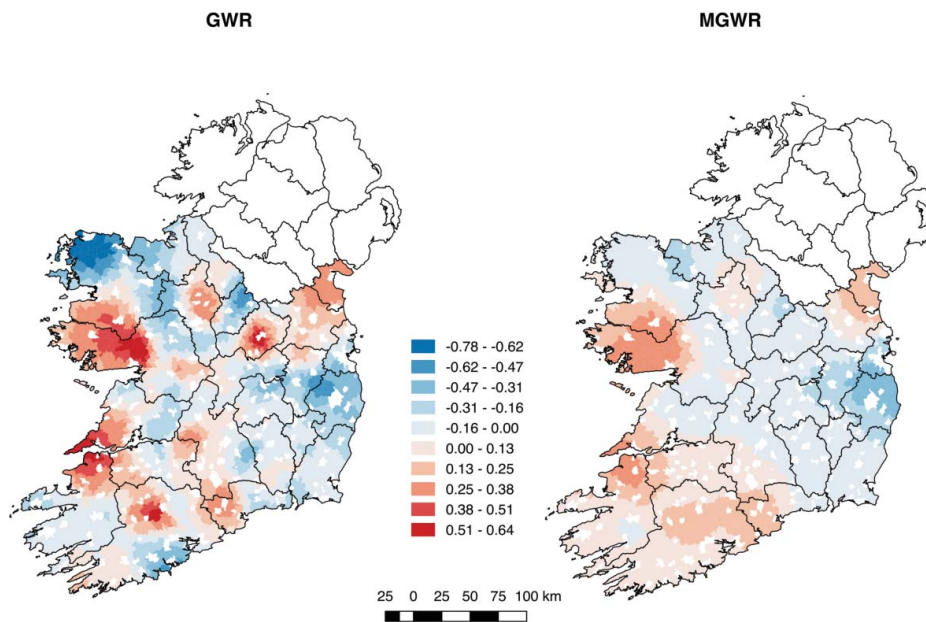


Figure 17. Geographically weighted regression and multiscale geographically weighted regression local estimates for workhouse proximity. GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression. (Color figure available online.)

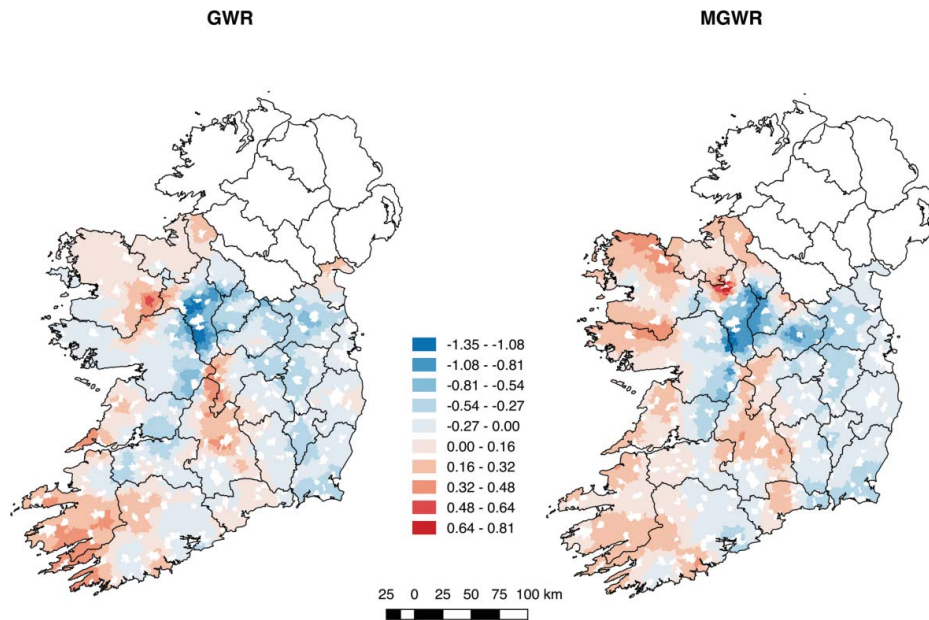


Figure 18. Geographically weighted regression and multiscale geographically weighted regression local estimates for mean elevation. GWR = geographically weighted regression; MGWR = multiscale geographically weighted regression. (Color figure available online.)

Discussion and Conclusions

GWR allows the investigation of spatial nonstationarity in spatial processes by relaxing the assumption that the parameters of a model are constant across space. The calibration of GWR also provides an optimized bandwidth that describes the spatial scale over which the processes being modeled vary. As only one bandwidth is used in the weighting function for GWR, however, an implicit assumption in the model is that the different processes being investigated vary at the same spatial scale. This prohibits, for example, the accurate interpretation of a situation where two processes exist, one operating at a regional scale and the other operating at a more local scale. Here we propose an extension to GWR, MGWR, to relax the implicit assumption within basic GWR that all the relationships operate at the same scale. A back-fitting algorithm is adopted for the model calibration and is implemented in a Python environment. This is demonstrated to be stable over different randomizations and the results do not appear to depend on the initial conditions. Calibration by MGWR produces a separate optimized bandwidth for each relationship in the model, thus indicating how different relationships operate over different spatial scales, and produces more accurate local parameter estimates. The separate bandwidths produced by MGWR have an intuitive interpretation in terms of geographical scale.

Using two different sets of simulated data with known properties, MGWR is shown to (1) accurately discriminate parameter surfaces with high spatial heterogeneity from those with low spatial heterogeneity and (2) produce similar bandwidths for processes that operate at the same spatial scale. Hence, the optimized bandwidths from MGWR provide valuable information on the scale at which different processes operate. As such, it provides an alternative to the nonseparable Bayesian SVC model in a model form that is arguably computationally less demanding, scales more easily, and, to some, is more intuitive. As Finley (2011) noted of SVC models: “By far the greatest challenge to widespread adoption of these models is computational” (153). Now that MGWR has been shown to be a viable model form that will generate different measurements of scale for different processes, the next step will be to compare its performance along a series of metrics with SVC models. Wheeler and Páez (2010) and Wheeler and Waller (2009) both compared GWR and SVC models but only as single-process models. Finley (2011) made a start in the direction of multiprocess models, but his comparisons were limited to a single-process GWR with both a single and multiprocess SVC model. Now that MGWR has been established as a multiprocess model, a more meaningful comparison can now be made with multiprocess SVC models.

Researchers interested in understanding the processes that generate the associations between data we

observe in the real world within the GWR framework now have four choices:

1. The standard modeling approach is that of global regression in which the implicit assumption is that the processes generating the associations being modeled do not vary over space and in which a single parameter is estimated for each relationship specified in the model.
2. GWR, which relaxes the assumption of spatial stationarity in the global model by allowing the parameters in the model to vary over space. The implicit assumption in this model is that all the local parameters in the model vary at the same spatial scale.
3. Semiparametric GWR (SGWR), which allows a subset of the relationships being modeled to be fixed over space and a subset to vary spatially. The implicit assumption in this model is that the subset of parameters that vary over space do so at the same spatial scale.
4. MGWR, which not only allows parameter estimates to vary over space but also allows the scale of this variation to vary across parameter surfaces.

Consequently, MGWR represents a significant advance in non-Bayesian regression modeling with spatial data. Not only can we investigate spatial heterogeneity in spatial processes, but we can identify the spatial scale over which different processes operate in an intuitive manner. Further research is now needed, though. In particular, it remains to be determined whether a goodness-of-fit statistic similar to AICc can be derived for a local model such as MGWR where there is no single weighting matrix. Allowing different bandwidths for each relationship in the model means that the nature of the weighting matrix varies across each set of local parameter estimates and hence the usual form of AICc is meaningless. Equivalently, the calculation of the equivalent number of independent parameters in the model, as is done, for example, in GWR, is challenging because there is no single hat matrix of which to take the trace. Future research will be focused on these issues and the construction of equivalents to the local R^2 , local standard errors, and local t statistics that need to be implemented within the algorithm. Three further directions for future research include increasing the efficiency of the code to reduce run times, undertaking a comparison of the performance of MGWR and SVC models along a series of metrics, and developing an inferential framework for MGWR. The latter can be developed in a fairly

straightforward manner via bootstrapping or Monte Carlo procedures, but this will add significantly to computing time. Consequently, research also needs to be directed to developing a more formal inferential procedure possibly through the established general additive modeling framework (Hastie and Tibshirani 1990).

This article establishes the merits of MGWR and demonstrates how it can be used to identify and measure the different spatial scales over which different processes operate. In doing so, it opens up many new possibilities for further empirical investigations into the role of multiscale processes. Geographers and other spatial scientists now have an alternative methodology to nonseparable Bayesian SVC models that provides an arguably more intuitive measure of the spatial scale over which different processes operate. An important question can now be investigated: How common are multiscale processes in generating the observations we measure in both our social and physical environments?

Notes

1. There is a considerable volume of research devoted to measuring the geographic scale over which data vary (e.g., spatial variograms and correlograms) and some has been extended to the spatial variability of joint relationships between variables, but the concentration in this article is on identifying the geographic scale at which processes occur.
2. See <https://www.python.org/>.
3. Both sets of simulated data employed in this analysis contain locations on a regular 25×25 grid for convenience. To check that the results were not an artefact of the regular design of locations, we repeated the experiments with a data set of locations drawn randomly from the grid. The results were virtually identical to those reported and so are not repeated here but are available from the authors. This finding that the results of the simulation studies are not dependent on the regularity of the sample design is in line with Haining (1986) and Anselin and Rey (1991).
4. The computer used throughout this research has dual 2.6 Ghz 8-core CPUs and 64 GB of memory.
5. The empirical application reported here differs from that reported by Fotheringham, Kelly, and Charlton (2013) in three ways. First, the original study included data from what is now Northern Ireland and hence had a larger sample size of 3,098 EDs rather than the 2,317 used here. Second, the original study employed nine predictor variables rather than the eight used here because one of the original variables, the proportion of population classed as urban, was zero for the vast majority of the reduced 2,317 ED data set. Third, here both the response variable and the set of predictor variables were standardized to have mean of 0 and variance of 1, which facilitates both the interpretation of the optimal bandwidths and the comparison of the local parameter estimates.

References

- Anselin, L., and S. Rey. 1991. Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23 (2): 112–31.
- Atkinson, P. M., S. E. German, D. A. Sear, and M. J. Clark. 2003. Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geographical Analysis* 35 (1): 58–82.
- Brenner, N. 2001. The limits to scale? Methodological reflections on scalar structuration. *Progress in Human Geography* 25 (4): 591–614.
- Brunsdon, C., A. S. Fotheringham, and M. Charlton. 1999. Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science* 39 (3): 497–524.
- Buja, A., T. Hastie, and R. Tibshirani. 1989. Linear smoothers and additive models. *The Annals of Statistics* 17 (2): 453–510.
- Everitt, B. S. 2005. Generalized additive model. In *Encyclopedia of statistics in behavioral science*, ed. B. S. Everitt and D. C. Howell, 719–21. Chichester, UK: Wiley.
- Finley, A. O. 2011. Comparing spatially-varying coefficients: Models for analysis of ecological data with non-stationarity and anisotropic residual dependence. *Methods in Ecology and Evolution* 2:143–54.
- Footy, G. 2003. Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI–rainfall relationship. *Remote Sensing of Environment* 88 (3): 283–93.
- Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester, UK: Wiley.
- Fotheringham, A. S., M. Charlton, and C. Brunsdon. 1996. The geography of parameter space: An investigation of spatial non-stationarity. *International Journal of Geographic Information Systems* 10 (5): 605–27.
- Fotheringham, A. S., M. Kelly, and M. Charlton. 2013. The demographic impacts of the Irish famine: Towards a greater geographical understanding. *Transactions of the Institute of British Geographers* 38 (2): 221–37.
- Fotheringham, A. S., and T. Oshan. 2016. GWR and Multicollinearity: Dispelling the myth. *Journal of Geographical Systems* 18 (4): 303–29.
- Gelfand, A. E., H. Kim, C. F. Sirmans, and S. Banerjee. 2003. Spatial modelling with spatially varying coefficient processes. *Journal of the American Statistical Association* 98:387–96.
- Gelfand, A. E., A. M. Schmidt, and C. F. Sirmans. 2003. Multivariate spatial process models: Conditional and unconditional Bayesian approaches using coregionalization. Technical Report 20, Institute of Statistics and Decision Sciences, Duke University, Durham, NC.
- Goodchild, M. 2001. Models of scale and scales of modelling. In *Modelling scale in geographic information science*, ed. N. Tate and P. M. Atkinson, 3–10. Chichester, UK: Wiley.
- Haining, R. 1986. Spatial models and regional science: A comment on Anselin's paper and research directions. *Journal of Regional Science* 26 (4): 793–98.
- Harvey, D. W. 1968. Pattern, process and the scale problem in geographical research. *Transactions of the Institute of British Geographers* 45:71–78.
- Hastie, T., and R. Tibshirani. 1986. Generalized additive models. *Statistical Science* 1 (3): 297–310.
- . 1990. *Generalized additive models*. London: Chapman and Hall/CRC.
- Liverman, D. 2004. Who governs, at what scale and at what price? Geography, environmental governance and the commodification of nature. *Annals of the Association of American Geographers* 94:734–38.
- Lloyd, C. D. 2010. Exploring population spatial concentrations in Northern Ireland by community background and other characteristics: An application of geographically weighted spatial statistics. *International Journal of Geographical Information Science* 24 (8): 1193–1221.
- . 2011. *Local models for spatial analysis*. Boca Raton, FL: CRC.
- McMaster, R. B., and E. Sheppard. 2004. Introduction: Scale and geographic inquiry. In *Scale and geographic inquiry: Nature, society and method*, ed. E. Sheppard and R. B. McMaster, 1–22. Oxford, UK: Blackwell.
- Moellering, H., and W. Tobler. 1972. Geographical variances. *Geographical Analysis* 4:34–50.
- Paasi, A. 2004. Place and region: Looking through the prism of scale. *Progress in Human Geography* 28:536–46.
- Sheppard, E., and R. B. McMaster, eds. 2004. *Scale and geographic inquiry: Nature, society and method*. Oxford, UK: Blackwell.
- Tate, N., and P. M. Atkinson, eds. 2001. *Modelling scale in geographic information science*. Chichester, UK: Wiley.
- Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234–40.
- Wheeler, D. C., and A. Páez. 2010. Geographically weighted regression. In *Handbook of applied spatial analysis: Software tools, methods and applications*, ed. M. M. Fischer and A. Getis, 461–68. Berlin: Springer-Verlag.
- Wheeler, D. C., and L. A. Waller. 2009. Comparing spatially varying coefficient models: A case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *Journal of Geosystems* 11:1–22.
- Whelan, K. 1997. *The atlas of the Irish rural landscape*. Cork, Ireland: Cork University Press.
- Yang, W. 2014. An extension of geographically weighted regression with flexible bandwidths. PhD thesis, School of Geography and Geosciences, University of St. Andrews, Fife, Scotland, UK. <http://hdl.handle.net/10023/7052> (last accessed 1 August 2017).

A. STEWART FOTHERINGHAM is Professor of computational spatial science in the School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85281. E-mail: Stewart.Fotheringham@asu.edu. He is also a Distinguished Scientist in the Julie Ann Wrigley Global Institute of Sustainability. His research interests are in the analysis of spatial data sets using statistical, mathematical, and computational methods. He is well known in the fields of spatial interaction modeling and local statistical analysis, the latter as one of the developers of geographically weighted regression (GWR). He has substantive interests in health data, crime patterns, retailing, and migration.

WENBAI YANG received her PhD degree in human geography from the School of Geography and Geosciences, University of St. Andrews, Fife, Scotland KY16 9AJ, UK. E-mail: yangwb97@sina.com. Her research interests are in geographic information systems and spatial analysis, statistical modeling, and qualitative and quantitative analysis.

WEI KANG is a PhD candidate in the School of Geographical Sciences & Urban Planning, Arizona State University, Tempe, AZ 85281. E-mail: wkang12@asu.edu. Her research interests include spatial statistics, spatial econometrics, temporal geographic information systems and GIScience, and regional economic growth.

Copyright of Annals of the American Association of Geographers is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.