geographical analysis

Geographical Analysis (2019) **0**, 1–20

Inference in Multiscale Geographically Weighted Regression

Hanchen Yu[1], A. Stewart Fotheringham[2], Ziqi Li[2], Taylor Oshan[3], Wei Kang[4], Levi John Wolf[5]

[1]School of Government, Peking University, Beijing, PRC, [2]Spatial Analysis Research Center, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ USA, [3]Center for Geospatial Information Science, Department of Geographical Sciences, University of Maryland, College Park, MD USA, [4]Center for Geospatial Sciences, School of Public policy, University of California, Riverside, CA USA, [5]School of Geographical Sciences, University of Bristol, Bristol, UK

A recent paper expands the well-known geographically weighted regression (GWR) framework significantly by allowing the bandwidth or smoothing factor in GWR to be derived separately for each covariate in the model—a framework referred to as multiscale GWR (MGWR). However, one limitation of the MGWR framework is that, until now, no inference about the local parameter estimates was possible. Formally, the so-called "hat matrix," which projects the observed response vector into the predicted response vector, was available in GWR but not in MGWR. This paper addresses this limitation by reframing GWR as a Generalized Additive Model, extending this framework to MGWR and then deriving standard errors for the local parameters in MGWR. In addition, we also demonstrate how the effective number of parameters can be obtained for the overall fit of an MGWR model and for each of the covariates within the model. This statistic is essential for comparing model fit between MGWR, GWR, and traditional global models, as well as for adjusting multiple hypothesis tests. We demonstrate these advances to the MGWR framework with both a simulated data set and a real-world data set and provide a link to new software for MGWR (MGWR1.0) which includes the novel inferential framework for MGWR described here.

Introduction

There is a great deal of interest in both the geographical and statistical literature on modeling frameworks that allow the estimation of spatially varying parameters (*inter alia* Casetti 1972; Fotheringham, Brunsdon, and Charlton 2002; Fotheringham, Charlton, and Brunsdon 1998; Gelfand et al. 2003; Griffith 2008; LeSage 2004; Oshan and Fotheringham 2017). Two main frameworks dominate the literature—geographically weighted regression (GWR) and spatially varying coefficients models (SVC). The former is a purely local model and based on a spatial

Correspondence: A. Stewart Fotheringham, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, USA
e-mail: stewart.fotheringham@asu.edu

Submitted: 3 May 2018; Revised version accepted: 18 December 2018.

doi: 10.1111/gean.12189

disaggregation of a classical regression model in which the local parameter estimates are generated by "borrowing data" around locations and then calibrating separate models for each location (Fotheringham, Brunsdon, and Charlton 2002; Fotheringham et al. 1996). The latter is framed in a Bayesian context and although it produces local parameter estimates, it uses all the data to do so rather than being a purely local model (Wolf, Oshan, and Fotheringham 2018). Until recently, the Bayesian SVC framework, while usually estimated using many more free parameters (such as prior specifications and Markov chain Monte Carlo tuning parameters), had an advantage over the GWR framework: a variant existed that allowed separate bandwidths to be estimated for each covariate in the model. GWR, however, was limited to fit a single optimal bandwidth, which typically reflected an "average" of the best bandwidths for each process.

Recently, this limitation was removed by the development of multiscale GWR (Fotheringham, Yang, and Kang 2017) which allows covariate-specific bandwidths to be optimized. This is a significant development for three reasons. First, allowing each covariate its own level of spatial smoothing places MGWR on the same footing as the non-separable spatially varying coefficients model in the Bayesian framework (Finley 2011) and in so doing addresses a limitation of the GWR framework. Second, these covariate-specific bandwidths can be used as indicators of the spatial scale over which each process operates. Third, the multi-bandwidth approach produces a more accurate and useful model of real-world processes than GWR and is arguably more more intuitive and easier to calibrate than the Bayesian SVC models. A comparison and fuller discussion of MGWR and Bayesian SVC models is given by Wolf, Oshan, and Fotheringham (2017).

However, although MGWR is a considerable improvement over GWR, a limitation of the MGWR model is that until now no formal inferential framework existed for the local parameter estimates.[1] Formally, the so-called "hat matrix," which projects the observed response vector into the predicted response vector, was available in GWR but not obtained for MGWR. This paper addresses this limitation by reframing GWR as a generalized additive model (GAM) (Buja, Hastie, and Tibshirani 1989; Hastie and Tibshirani 1990), and extending this framework to MGWR to derive standard errors for the local parameter estimates. The paper also derives covariate-specific and overall model formulations for the effective number of parameters (ENP) to be computed for MGWR. These diagnostics are essential for comparing model fit between MGWR, GWR, and traditional global models as well as for adjusting multiple comparisons during parameter hypothesis testing. We demonstrate these advances to the MGWR framework with both a simulated data set and a real-world data set. We first reframe GWR as a GAM and then extend this framework to MGWR to derive the standard errors of the local parameter estimates. We provide a link to a user-friendly software package for MGWR which includes the inferential framework described here.

GWR as a GAM

Standard models such as those calibrated by OLS regression assume that the processes generating the data we observe are the same across space. GWR removes this assumption and allows processes to vary over space. It does this by calibrating a separate regression model at each location by borrowing data from nearby locations and weighting these data by distance from the regression point such that data from locations nearer to the regression point are weighted more than data from more distant locations. This approach is a generalization of kernel regression concepts (Cleveland 1979), and has significant and powerful meaning in the context of spatial statistics. The rate at which the weights of locations decline as distance increases is controlled by a bandwidth, which is optimized in the GWR calibration procedure. GWR is formulated as

$$y_i = \sum_{j=1}^{k} \beta_{ij} x_{ij} + \varepsilon_i \tag{1}$$

where for the observation at location $i \in \{1, 2, \ldots, n\}$, y_i is the response variable, x_{ij} is the jth predictor variable, j $\in \{1, 2, \ldots, \mathrm{k}\}$, β_{ij} is the jth parameter estimate, and ε_i is the error term. GWR calibration for the $(k,1)$ coefficients at location i in matrix form is given by

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{i}} = \left(\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}_{\boldsymbol{i}} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}_{\boldsymbol{i}} \boldsymbol{y}, \quad i \in \{1, 2, \ldots, n\} \tag{2}$$

where $\boldsymbol{X}$ is the (n, k) matrix of predictor variables (including a column of 1 s for the intercept), $\boldsymbol{y}$ is the $(n,1)$ response variable, and $\boldsymbol{W}_{\boldsymbol{i}}$ is the (n, n) diagonal spatial weighting matrix for location i with the diagonal elements representing the weights attached to each location and is calculated based on a specified kernel function and bandwidth. The bandwidth is assumed to be constant across all the covariates implying that the processes producing the observed values of each covariate vary at the same spatial scale.

MGWR (Fotheringham, Yang, and Kang 2017) is a more generalized GWR that loosens the constant bandwidth assumption and allows each predictor to have a different optimized bandwidth. Both GWR and MGWR can be thought of as GAMs. A GAM represents the response variable as the sum of smoothed predictor variables. A typical GAM can be expressed as:

$$\boldsymbol{y} = \sum_{\mathrm{j}=1}^{\mathrm{k}} \boldsymbol{f}_{\boldsymbol{j}} + \boldsymbol{\varepsilon} \tag{3}$$

where $\boldsymbol{f}_{\boldsymbol{j}}$ is a smooth function applied on the jth predictor variable (Wood 2017). In the context of (M)GWR, each smooth function $\boldsymbol{f}_{\boldsymbol{j}}$ is a spatial GWR parameter surface calibrated using a known bandwidth. This bandwidth is constant for all predictor variables in GWR but can be varying over j in MGWR.

The hat matrix in GWR

The derivation of the hat matrix, also referred as a projection matrix or influence matrix, is crucial to deriving inferential statistics in GWR. The hat matrix maps the response values directly to the fitted values and is necessary for calculating model diagnostics such as the corrected Akaike Information Criterion (AICc) and the equivalent/effective number of parameters as well as inferential statistics such as local standard errors and local t values. Each row i in the GWR hat matrix $\boldsymbol{S}$ is given by:

$$\boldsymbol{s}_{\boldsymbol{i}} = \boldsymbol{X}_{\boldsymbol{i}} \left(\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}_{\boldsymbol{i}} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}_{\boldsymbol{i}} \tag{4}$$

where $\boldsymbol{X}_{\boldsymbol{i}}$ is the ith row of predictors $\boldsymbol{X}$. So, the full hat matrix $\boldsymbol{S}$ can be expressed as:

$$\boldsymbol{S}=\begin{pmatrix} \boldsymbol{X}_1\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}_1\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}_1 \\ \dots \\ \boldsymbol{X}_{\boldsymbol{n}}\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}_{\boldsymbol{n}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}_{\boldsymbol{n}} \end{pmatrix}_{n\times n} \tag{5}$$

The hat matrix $\boldsymbol{S}$ can be decomposed into covariate-specific additive hat matrices $\boldsymbol{R}_{1\dots\boldsymbol{k}}$ that map $\boldsymbol{y}$ to the fitted additive terms $\hat{\boldsymbol{f}}_{1\dots\boldsymbol{k}}$. The $\boldsymbol{R}_{1\dots\boldsymbol{k}}$ matrices have two properties:

$$(1).\ \hat{\boldsymbol{f}}_{\boldsymbol{j}}=\boldsymbol{R}_{\boldsymbol{j}}\boldsymbol{y}.\ \ \boldsymbol{R}_{\boldsymbol{j}} \text{ projects } \boldsymbol{y} \text{ onto each fitted additive term } \hat{\boldsymbol{f}}_{\boldsymbol{j}}. \tag{6}$$

$$(2).\ \hat{\boldsymbol{y}}=\boldsymbol{S}\boldsymbol{y}=\sum_{j=1}^{k}\boldsymbol{R}_{\boldsymbol{j}}\boldsymbol{y}, \text{ which implies } \sum_{j=1}^{k}\boldsymbol{R}_{\boldsymbol{j}}=\boldsymbol{S}. \tag{7}$$

To derive $\boldsymbol{R}_{\boldsymbol{j}}$ we first express $\hat{\boldsymbol{f}}_{\boldsymbol{j}}$ as a column vector

$$\hat{\boldsymbol{f}}_{\boldsymbol{j}}=\begin{pmatrix} x_{1j}\hat{\beta}_{1j} \\ \dots \\ x_{nj}\hat{\beta}_{nj} \end{pmatrix} \tag{8}$$

where x_{ij} and $\hat{\beta}_{ij}$ are the jth covariate and parameter estimate, respectively, at location i. Let $\boldsymbol{e}_{\boldsymbol{j}}$ be the jth row vector of a k by k identity matrix

$$\boldsymbol{e}_{\boldsymbol{j}}=\begin{pmatrix} \underbrace{0\dots0}_{j-1} & 1 & \underbrace{0\dots0}_{k-j} \end{pmatrix} \tag{9}$$

so

$$\hat{\beta}_{ij}=\boldsymbol{e}_{\boldsymbol{j}}\hat{\boldsymbol{\beta}}_{\boldsymbol{i}} \tag{10}$$

where $\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}$ is a column vector of k parameter estimates at location i. To proceed, we replace $\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}$ in Equation (10) with the GWR estimator and obtain a place-specific estimator:

$$\hat{\beta}_{ij}=\boldsymbol{e}_{\boldsymbol{j}}\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}_{\boldsymbol{i}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}_{\boldsymbol{i}}\boldsymbol{y} \tag{11}$$

Finally, replace each $\hat{\beta}_{ij}$ in Equation (8) to define the partial smooth for process j *as:*

$$\hat{\boldsymbol{f}}_{\boldsymbol{j}}=\begin{pmatrix} \mathrm{x}_{1\mathrm{j}}\boldsymbol{e}_{\boldsymbol{j}}\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}_1\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}_1 \\ \dots \\ \mathrm{x}_{\mathrm{nj}}\boldsymbol{e}_{\boldsymbol{j}}\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}_{\boldsymbol{n}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}_{\boldsymbol{n}} \end{pmatrix}\boldsymbol{y} \tag{12}$$

Then, the hat matrix $\boldsymbol{R_j}$ for the jth additive term $\hat{\boldsymbol{f}}_j$ is:

$$\boldsymbol{R_j} = \begin{pmatrix} \mathrm{x}_{1\mathrm{j}}\boldsymbol{e_j}\left(\boldsymbol{X^T W}_1\boldsymbol{X}\right)^{-1}\boldsymbol{X^T W}_1 \\ \dots \\ \mathrm{x}_{\mathrm{nj}}\boldsymbol{e_j}\left(\boldsymbol{X^T W_n X}\right)^{-1}\boldsymbol{X^T W_n} \end{pmatrix}_{n\times n} \tag{13}$$

Given that the above hat matrix is derived for GWR when represented as a GAM, we can now develop this same framework to derive the hat matrix equivalent for MGWR.

MGWR as a GAM

Fotheringham, Yang, and Kang (2017) proposed a backfitting algorithm for the calibration of MGWR models based on viewing the model as a GAM. Firstly, parameter estimates at all locations are initialized and fitted additive terms $\hat{\boldsymbol{f}}_{1\dots \boldsymbol{k}}$ are computed based on the initial parameter estimates and data on the covariates.[2] The model residuals $\hat{\boldsymbol{\varepsilon}}$ at this point are obtained from:

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \sum_{\mathrm{j}=1}^{\mathrm{k}} \hat{\boldsymbol{f}}_{\boldsymbol{j}} \tag{14}$$

Then, the residuals $\hat{\boldsymbol{\varepsilon}}$ plus the first additive term $\hat{\boldsymbol{f}}_1$ are regressed on $\boldsymbol{X}_1$ (first covariate) using GWR to find an optimal bandwidth bw_1 as well as a new set of parameter estimates in order to update the first additive term $\hat{\boldsymbol{f}}_1$ and residuals $\hat{\boldsymbol{\varepsilon}}$. This procedure is then repeated for covariate $\boldsymbol{X}_2$ to update the second additive term $\hat{\boldsymbol{f}}_2$ and residuals $\hat{\boldsymbol{\varepsilon}}$. The process continues in this way through to the final covariate $\boldsymbol{X_k}$ when $\hat{\boldsymbol{f}}_{\boldsymbol{k}}$ and $\hat{\boldsymbol{\varepsilon}}$ are updated. This finishes the first round iteration and the second round starts from the first additive term using the new values for the estimated $\mathbf{f}_{\mathrm{k}}$ and $\hat{\boldsymbol{\varepsilon}}$. The iterations continue until the change in a convergence indicator (such as the residual sum of squares) becomes sufficiently small between successive iterations.

The hat matrix in MGWR

In this section, we propose a solution to calculate the hat matrix for MGWR using the backfitting calibration procedure described above. We start with initializing the backfitting algorithm using GWR local parameter estimates. We then calculate the additive hat matrix $\boldsymbol{R_j}$ for each covariate using Equation (13). Then, we regress the jth covariate $\boldsymbol{X}_{\mathrm{j}}$ against $\hat{\boldsymbol{f}}_{\boldsymbol{j}} + \hat{\boldsymbol{\varepsilon}}$ and find the partial optimal model for the jth covariate yielding an optimized bw_{j}. Let $\boldsymbol{A_j}$ be the hat matrix of the partial optimal model so that

$$\hat{\boldsymbol{f}}_{\boldsymbol{j}}^{*} = \boldsymbol{A_j}\left(\hat{\boldsymbol{f}}_{\boldsymbol{j}} + \hat{\boldsymbol{\varepsilon}}\right) \tag{15}$$

where $\hat{\boldsymbol{f}}_{\boldsymbol{j}}^{*}$ is the updated $\hat{\boldsymbol{f}}_{\boldsymbol{j}}$ and $\hat{\boldsymbol{f}}_{\boldsymbol{j}}$ is the jth fitted additive term from the previous iteration. From Equation (14) and the properties of $\boldsymbol{R}_{1\dots\mathrm{k}}$ defined in Equations (6) and (7), we have

$$\hat{\boldsymbol{f}}_{\boldsymbol{j}} + \hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{f}}_{\boldsymbol{j}} + \boldsymbol{y} - \sum_{\mathrm{j}=1}^{\mathrm{k}} \hat{\boldsymbol{f}}_{\boldsymbol{j}} = \boldsymbol{R_j y} + \ \boldsymbol{y} - \boldsymbol{Sy} \tag{16}$$

Combining Equations (15) to (16), we then obtain

$$\hat{\boldsymbol{f}}_j^* = \boldsymbol{A}_j\left(\hat{\boldsymbol{f}}_j + \hat{\boldsymbol{\varepsilon}}\right) = \boldsymbol{A}_j\left(\hat{\boldsymbol{f}}_j + \boldsymbol{y} - \sum_{\mathrm{j}=1}^{\mathrm{k}} \hat{\boldsymbol{f}}_j\right) = \boldsymbol{A}_j\left(\boldsymbol{R}_j\boldsymbol{y} + \boldsymbol{y} - \boldsymbol{S}\boldsymbol{y}\right) = \left(\boldsymbol{A}_j\boldsymbol{R}_j + \boldsymbol{A}_j - \boldsymbol{A}_j\boldsymbol{S}\right)\boldsymbol{y} \tag{17}$$

Therefore, the additive hat matrix $\boldsymbol{R}_j^*$ for MGWR in the next iteration is:

$$\boldsymbol{R}_j^* = \boldsymbol{A}_j\boldsymbol{R}_j + \boldsymbol{A}_j - \boldsymbol{A}_j\boldsymbol{S} \tag{18}$$

The new hat matrix $\boldsymbol{S}^*$ is then updated accordingly

$$\boldsymbol{S}^* = \boldsymbol{S} - \boldsymbol{R}_j + \boldsymbol{R}_j^* \tag{19}$$

The pseudo-code for the hat matrix calculation is as follows:

MGWR hat matrix calculation

Initiate $\hat{\boldsymbol{f}}_{1\ldots\boldsymbol{k}}, \hat{\boldsymbol{\varepsilon}}, \boldsymbol{R}_{1\ldots\boldsymbol{k}}, \boldsymbol{S}$ from $GWR\,\{\boldsymbol{y} \sim \boldsymbol{X}\}$
Do until $\hat{\boldsymbol{f}}_{1\ldots\boldsymbol{k}}$ converge:
    For each term j from 1 to k:
        $\hat{\boldsymbol{f}}_j, \hat{\boldsymbol{\varepsilon}}, \boldsymbol{A}_j \leftarrow GWR\left\{\hat{\boldsymbol{f}}_j + \hat{\boldsymbol{\varepsilon}} \sim \boldsymbol{X}_j\right\}$ using optimal bw_j
        $\boldsymbol{R}_j \leftarrow \boldsymbol{A}_j\boldsymbol{R}_j + \boldsymbol{A}_j - \boldsymbol{A}_j\boldsymbol{S}$
        $\boldsymbol{S} \leftarrow \sum_1^k \boldsymbol{R}_j$
    End for
End do

Local standard errors of the parameter estimates in MGWR

Given we now have a hat matrix defined for MGWR, we can use this to derive standard errors of the local parameter estimates and other diagnostics. We rewrite Equation (8) into matrix multiplication form as:

$$\hat{\boldsymbol{f}}_j = diag\left(\boldsymbol{X}_j\right)\hat{\boldsymbol{\beta}}_j \tag{20}$$

where

$$diag\left(\boldsymbol{X}_j\right) = \begin{pmatrix} x_{1j} & 0 & 0 & 0 \\ 0 & x_{2j} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & x_{nj} \end{pmatrix}_{n\times n} \tag{21}$$

Then, $\hat{\boldsymbol{\beta}}_j$ can be given by

$$\hat{\boldsymbol{\beta}}_j = \left[diag\left(\boldsymbol{X}_j\right)\right]^{-1}\hat{\boldsymbol{f}}_j \tag{22}$$

and we replace $\hat{\boldsymbol{f}}_j$ with $\boldsymbol{R}_j\mathbf{y}$ to yield

$$\hat{\boldsymbol{\beta}}_j = \left[diag\left(\boldsymbol{X}_j\right)\right]^{-1}\hat{\boldsymbol{f}}_j = \left[diag\left(\boldsymbol{X}_j\right)\right]^{-1}\boldsymbol{R}_j\mathbf{y} = \boldsymbol{C}\boldsymbol{y} \tag{23}$$

where $\boldsymbol{C} = \left[diag\left(\boldsymbol{X}_j\right)\right]^{-1}\boldsymbol{R}_j$. The variances of the parameter estimates $\hat{\boldsymbol{\beta}}_j$ are then given by

$$var\left(\hat{\boldsymbol{\beta}}_j\right) = diag\left(\boldsymbol{C}\boldsymbol{C}^{\boldsymbol{T}}\hat{\sigma}^2\right) \tag{24}$$

where $\hat{\sigma}^2$ is the normalized residual sum of squares from MGWR defined by

$$\hat{\sigma}^2 = \frac{\sum\left(y_i - \hat{y}_i\right)^2}{n - v_1} \tag{25}$$

where

$$v_1 = trace\left(\boldsymbol{S}\right) \tag{26}$$

Once the variances of the *j*th parameter estimates for all n locations are obtained from Equation (24), the standard errors are available using:

$$SE\left(\hat{\boldsymbol{\beta}}_j\right) = \sqrt{var\left(\hat{\boldsymbol{\beta}}_j\right)} \tag{27}$$

and the local standard errors for all the MGWR parameter estimates are:

$$SE\left(\hat{\boldsymbol{\beta}}\right) = \left[SE\left(\hat{\boldsymbol{\beta}}_1\right), SE\left(\hat{\boldsymbol{\beta}}_2\right) \ldots SE\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}\right)\right]_{n\times k} \tag{28}$$

which means the pseudo- *t* test for the local parameter estimate in process *j* at location *i* is given by the typical *t* test:

$$\frac{\hat{\beta}_{ij}}{SE\left(\hat{\beta}_{ij}\right)} \sim \mathrm{t}_{n-v1} \tag{29}$$

Covariate-specific ENP values in MGWR

In the iterative procedure described above, to derive a hat matrix, $\boldsymbol{S}$, for MGWR, we recognize that $\sum_{j=1}^{k}\boldsymbol{R}_j = \boldsymbol{S}$. That is, the overall hat matrix for the MGWR is composed of *k* covariate-specific projection matrices $\boldsymbol{R}_j$ which are derived from an iterative process where for each of the k covariates the smoother $\hat{\boldsymbol{f}}_j$ is projected onto $\boldsymbol{y}$ as:

$$\hat{\boldsymbol{f}}_j = \left(\boldsymbol{A}_j\boldsymbol{R}_j + \boldsymbol{A}_j - \boldsymbol{A}_j\boldsymbol{S}\right)\boldsymbol{y} \tag{30}$$

and the additive projection matrix $\boldsymbol{R}_j^*$ in the next iteration is:

$$\boldsymbol{R}_j^* = \boldsymbol{A}_j\boldsymbol{R}_j + \boldsymbol{A}_j - \boldsymbol{A}_j\boldsymbol{S} \tag{31}$$

with the new hat matrix $\boldsymbol{S}^*$ updated accordingly

$$\boldsymbol{S}^* = \boldsymbol{S} - \boldsymbol{R_j} + \boldsymbol{R_j^*} \tag{32}$$

At the final iteration of this process, each $\boldsymbol{R_j}$ represents a covariate-specific hat matrix, which can then be used to derive the ENP for each covariate from:

$$ENP_j = trace\left(\boldsymbol{R_j}\right) \tag{33}$$

The value of ENP_j is useful because it is used to adjust the critical value of t to account for multiple testing as shown below. Note that the sum of the covariate-specific ENP_j values will equal the ENP for the whole model.

Covariate-specific adjustments to the critical value of *t* due to multiple hypothesis testing

The local t tests for GWR and MGWR suffer from false positives induced by multiple hypothesis testing (Miller 1981; Tukey 1991). The issue is further complicated by the potential dependence across local estimates (Benjamini and Yekutieli 2001). We follow da Silva and Fotheringham (2016) who propose an effective correction to the significance level α for the local t tests to maintain a proper family-wise error rate ξ (often set as 0.05) and to avoid false postitves in GWR.[3] The adjusted value of α for the local t tests is given by:

$$\alpha = \frac{\xi}{\frac{ENP}{P}} \tag{34}$$

where ENP is the effective number of parameters in GWR and P is the number of parameters in the global model. In the case of the calibration of a model by GWR, when the processes are all stationary, the bandwidth would tend to infinity, ENP would tend to P and α would tend to ξ; in all other situations, $\alpha < \xi$.

In MGWR, given that we are able to compute ENP_j, a separate value of ENP for each covariate, the adjustment to alpha should be undertaken separately for each covariate as:

$$\alpha_j = \frac{\xi}{ENP_j} \tag{35}$$

For those covariates where the optimized bandwidth tends to infinity, ENP_j tends to 1 and α_j tends to ξ; in all other situations, $\alpha_j < \xi$.

An example using simulated data

To demonstrate the validity of the analytical results provided above, we construct a simulation experiment as follows. Consider a model with two covariates of the following form:

$$y_i = \beta_{1i}x_{1i} + \beta_{2i}x_{2i} + \varepsilon_i \tag{36}$$

where values of the covariates are randomly drawn from a normal distribution (0, 2) for 625 locations on a regular 25×25 grid. The local parameter values, β_{1i} and β_{2i}, are obtained from pre-specified surfaces (see Fig. 1). The values of $\boldsymbol{y}$ are derived from Equation (36) with ε_i set to zero for all i. Given the values of y_i, x_{1i}, and x_{2i} for all locations, we calibrate the model in Equation (36) by MGWR to obtain covariate-specific optimized bandwidths and an estimate of σ^2 in Equation (25).[4] To simulate the standard errors of the local parameter estimates, we then undertake the following iterative process:

1. Draw a random error term, ε_i, from $N(0, \hat{\sigma}^2)$ for each i.
2. Use these values in Equation (40) to generate a new value of y_i for every i.
3. Regress the new $\boldsymbol{y}$ against $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ using the MGWR calibration procedure and store the local parameter estimates.

This process is repeated 1,000 times to generate 1,000 estimates for each local coefficient, the standard deviation of which yields an experimental standard error for each local coefficient estimator. Note that in step (i) it is important to draw the random error terms from the normal distribution with mean 0 and variance $\hat{\sigma}^2$ which is obtained by calibrating the model in Equation (36), where all ε_i are set to 0 by MGWR. Using some other prior value for the variance of the error term will result in simulated standard errors of the local parmater estimates that are only correct up to some scaling factor.

The resulting experimental local standard errors of the estimates of β_1 and β_2 are shown in Figs. 2 and 3, respectively, along with the equivalent analytically derived standard errors from Equation (24). The similarity in the two sets of local standard errors is reinforced in the scatterplots in Fig. 4 and confirms that the analytical expression in Equation (24) yields correct estimates of the local standard errors of the parameters obtained in the calibration of a model by MGWR. We now demonstrate the use of Equation (24) in a real-world data set on educational attainment across counties in Georgia, commonly used in GWR as a demonstrator. We use this to show that when the MGWR covariate-specific bandwidths are all similar to the single bandwidth obtained in GWR, the standard errors of the local parameter estimates from both models,

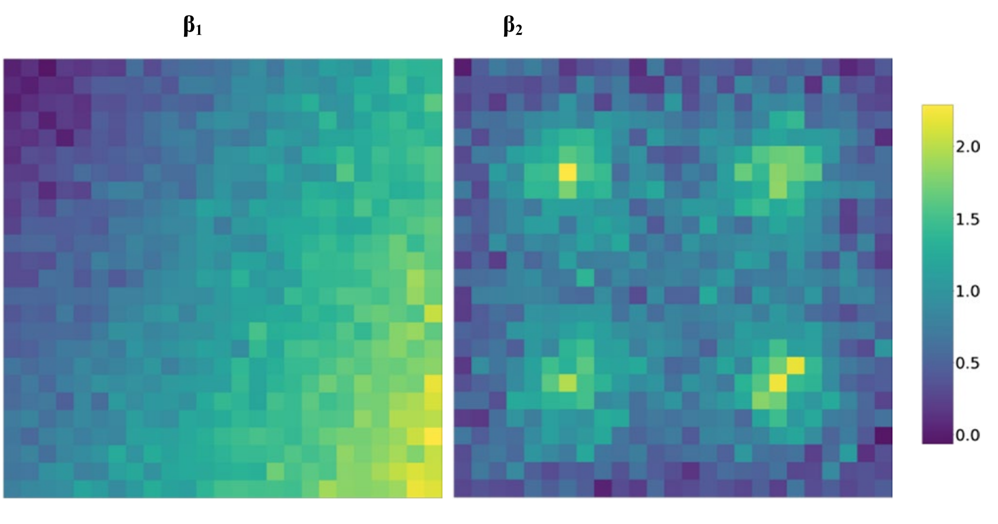

Figure 1. Parameter Surfaces for β_1 and β_2.

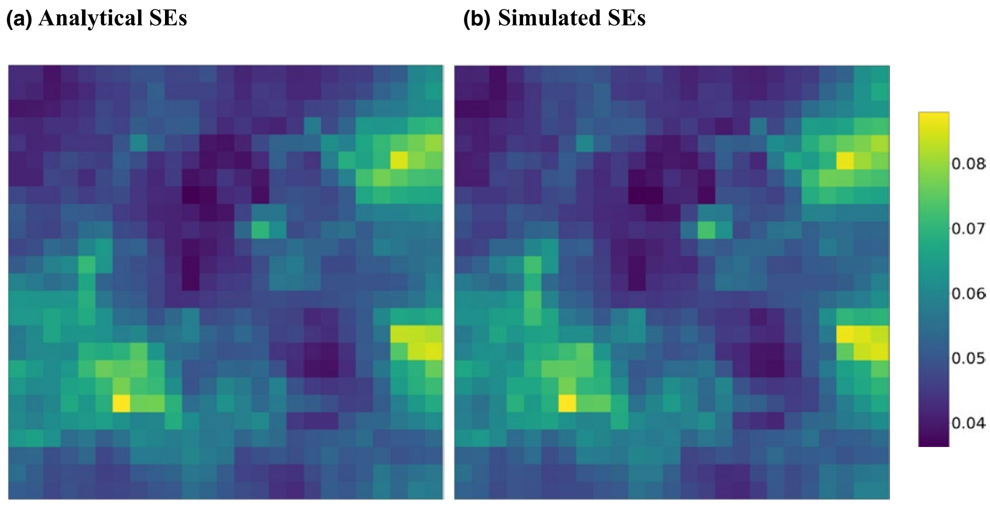

Figure 2. Analytical and simulated local standard errors for β_1.

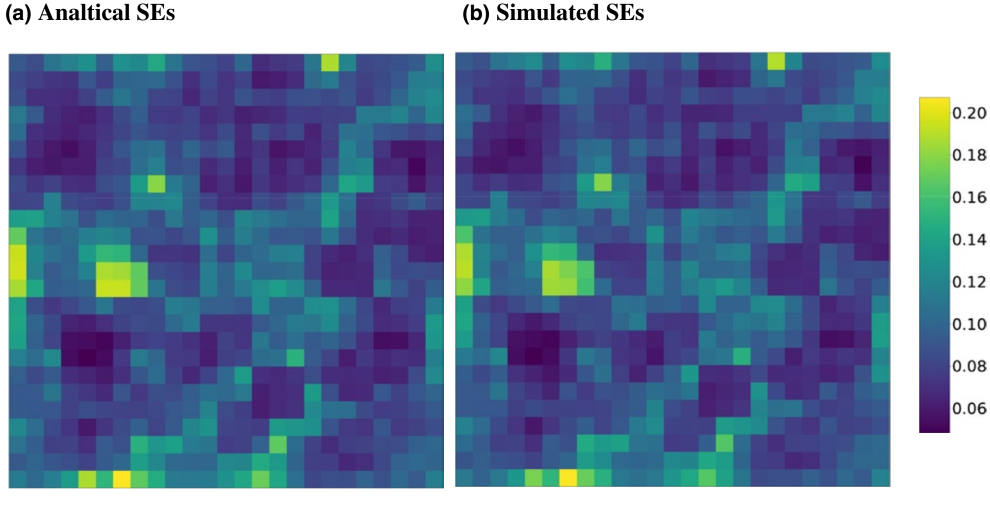

Figure 3. Analytical and simulated local standard errors for β_2.

and subsequent inferences drawn from the results, are similar, as expected. This provides further reinforcement of the methodology described above.

An example of educational attainment in Georgia

To demonstrate the calculation of local standard errors in a model calibrated by MGWR, we establish a simple model in which the percentage of people with a bachelor's degree is regressed on three covariates: the percentage of foreign-born residents, the percentage of people classified

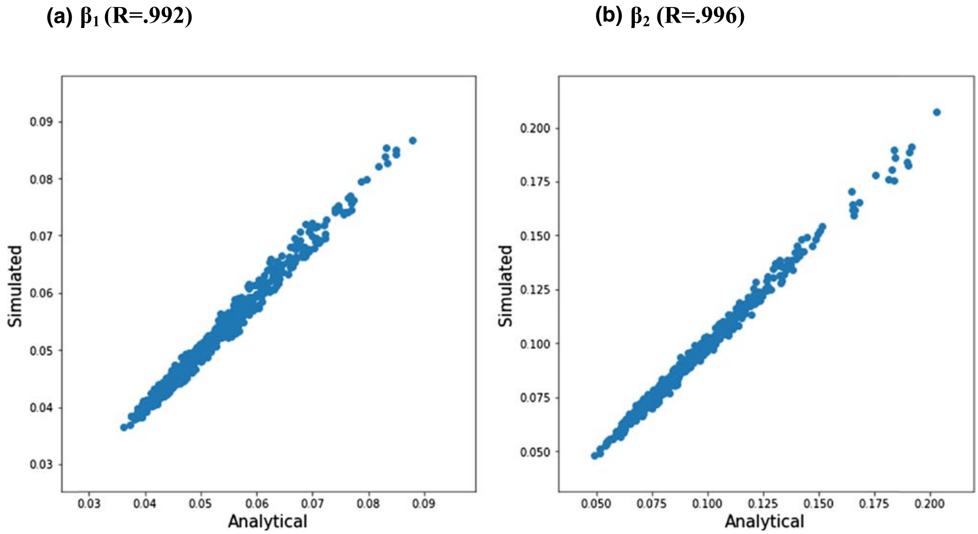

Figure 4. Scatterplots of the analytical and simulated standard errors for β_1 and β_2.

Table 1. Global OLS Calibration Results from the Georgia Data

Covariate	Est. Parameter	SE	t value
Intercept	0.0	0.054	0
Foreign-born	0.458	0.066	7.0
Rural	–0.374	0.065	–5.7
African American	–0.084	0.055	–1.5

AICc = 332
Adj. R^2 = 0.54
n = 159

as rural, and the percentage of African Americans. The data are drawn from the 1990 US census of population for the 159 counties in the State of Georgia and have been provided as a sample data set in successive releases of the GWR software, the current version being available at https://sgsup.asu.edu/sparc/software/mgwr. The covariates used here were checked extensively for local multicollinearity using local correlations, local VIFs, and local condition numbers from both GWR and MGWR and no problems were identified.[5]

The global OLS model calibration results are shown in Table 1 and indicate that higher proportions of people with a bachelor's degree are associated with counties having higher proportions of residents who are foreign-born and lower proportions of rural population. The proportion of African Americans in each county has no impact on the proportion of people with a bachelor's education. The AICc for this model is 885 and the adjusted r-squared value is 0.54.

In order to demonstrate the validity of Equation (24) in deriving the standard errors of the local parameter estimates from MGWR, the model described above was calibrated by GWR (where the standard error formulation is well-known) and MGWR. Both calibrations utilize an adaptive bandwidth using a bisquare weighting function (Fotheringham, Brunsdon, and Charlton 2002). Both produce four sets of local parameter estimates and for each we map three inference-related diagnostics: the local parameter estimates themselves; their estimated standard errors; and a map of the local parameter estimates associated with t values which exceed the critical threshold using a 95% confidence interval and adjusted for multiple comparisons during parameter hypothesis testing (da Silva and Fotheringham 2016). This latter map is useful in identifying just those locations with significant local parameter estimates to help focus attention on areas of interest. These are given in Figs. 5–8, respectively. Unlike in GWR, the adjusted critical value for the *t* statistics in MGWR is covariate-specific as the effective number of parameters for each covariate is a function of the covariate-specific bandwidth. The covariate-specific bandwidths, effective number of parameter estimates, adjusted alpha, and critical *t* values from the MGWR calibration along with their GWR equivalents are shown in Table 2.

The single optimal bandwidth obtained in the GWR calibration is 117. This is the number of locations (out of a possible 159) that have a non-zero weighting in the local model calibration. However, this is a weighted average across the covariates in the model which themselves may have different optimal weighting functions. The MGWR results demonstrate this. The optimal bandwidths for each of the four sets of parameter estimates are 92 for the intercept, 101 for foreign-born, 158 for rural, and 136 for African American. Conceptually, this means that the site-specific baseline for the model (reflected in the intercept) is *more local* than in GWR, as are the relationships between foreign-born population fractions and educational attainment. To counterbalance this, the relationships between the remaining covariates and educational attainment are more global than in GWR. So, although all four local sets of parameter estimates from MGWR exhibit only broad regional variations, they do differ slightly with the intercept exhibiting the most spatial variation and rural exhibiting the least. Indeed, the influence of the rural covariate is effectively stationary over space.

Because in this example the four sets of local parameter estimates vary at a similar spatial scale, the GWR and MGWR results should be similar, as seen by the ENP for both models (11.80 for GWR and 11.37 for MGWR). ENP is a measure of model complexity and represents the equivalent number of independent parameters estimated in a local model. For the whole model, if there is no spatial non-stationarity in the processes being examined, ENP tends to the number of parameters estimated in the global model. Because ENP is a function of the hat matrix, and because we can now obtain a hat matrix for each covariate in MGWR, as well as a "whole model" estimate of ENP, it is possible to obtain a separate ENP value for each covariate as shown in Table 2. These values indicate the equivalent number of independent local parameters required to model the smoother for each covariate. In the limit, as the local model tends to the global model, this value will tend to 1 as the global parameter would apply everywhere. ENP is used to adjust the value of alpha which determines the critical value of *t* in the presence of multiple tests (da Silva and Fotheringham 2016). The effect of this can be seen in Table 2 in the reported adjusted values of alpha and the associated critical *t* values for each covariate.

Figs. 5–8 demonstrate both the validity and the usefulness of the local standard error estimator for MGWR given in Equation (24). The local parameter estimates and associated standard errors are mapped in (a) and (b), respectively, for both GWR and MGWR while (c) indicates only the significant local parameter estimates defined as having an associated *t* value greater than the critical value given in Table 2 which has been adjusted for multiple tests. In the case of the local intercept, the parameter estimates from GWR and MGWR are similar, although the MGWR local standard errors tend to be slightly larger. The maps of the significant local estimates show

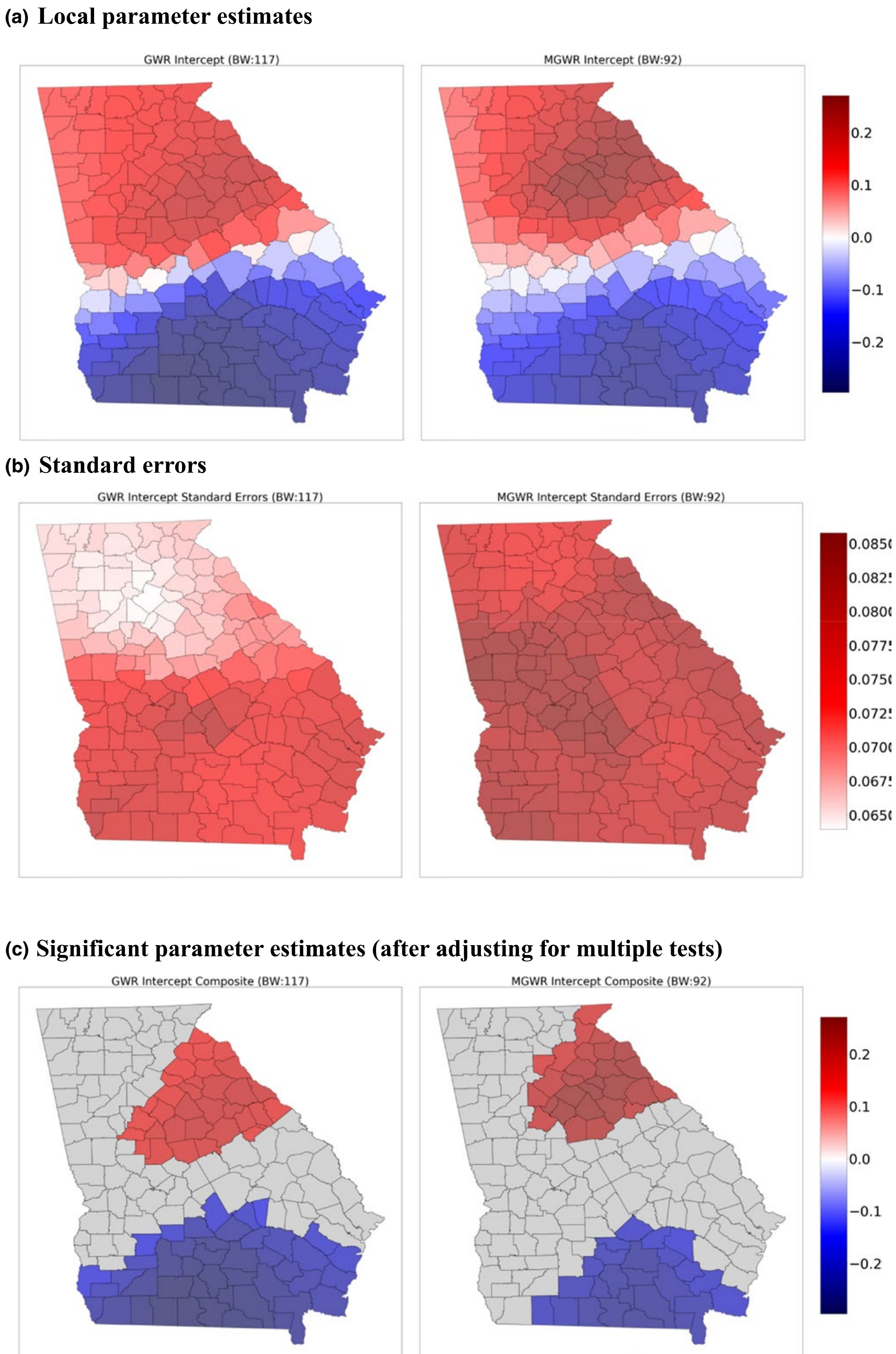

Figure 5. Intercept.

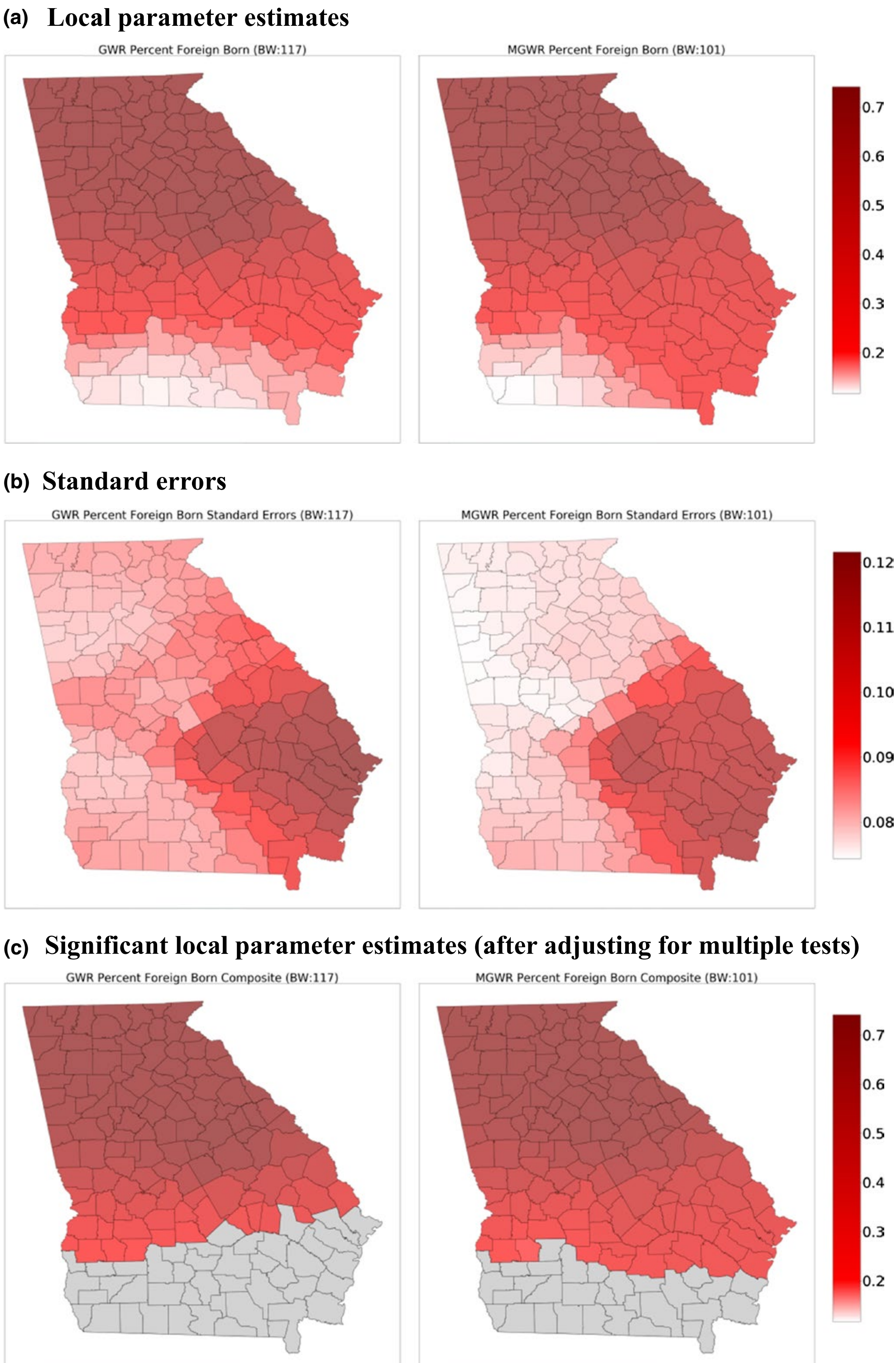

Figure 6. Foreign-born.

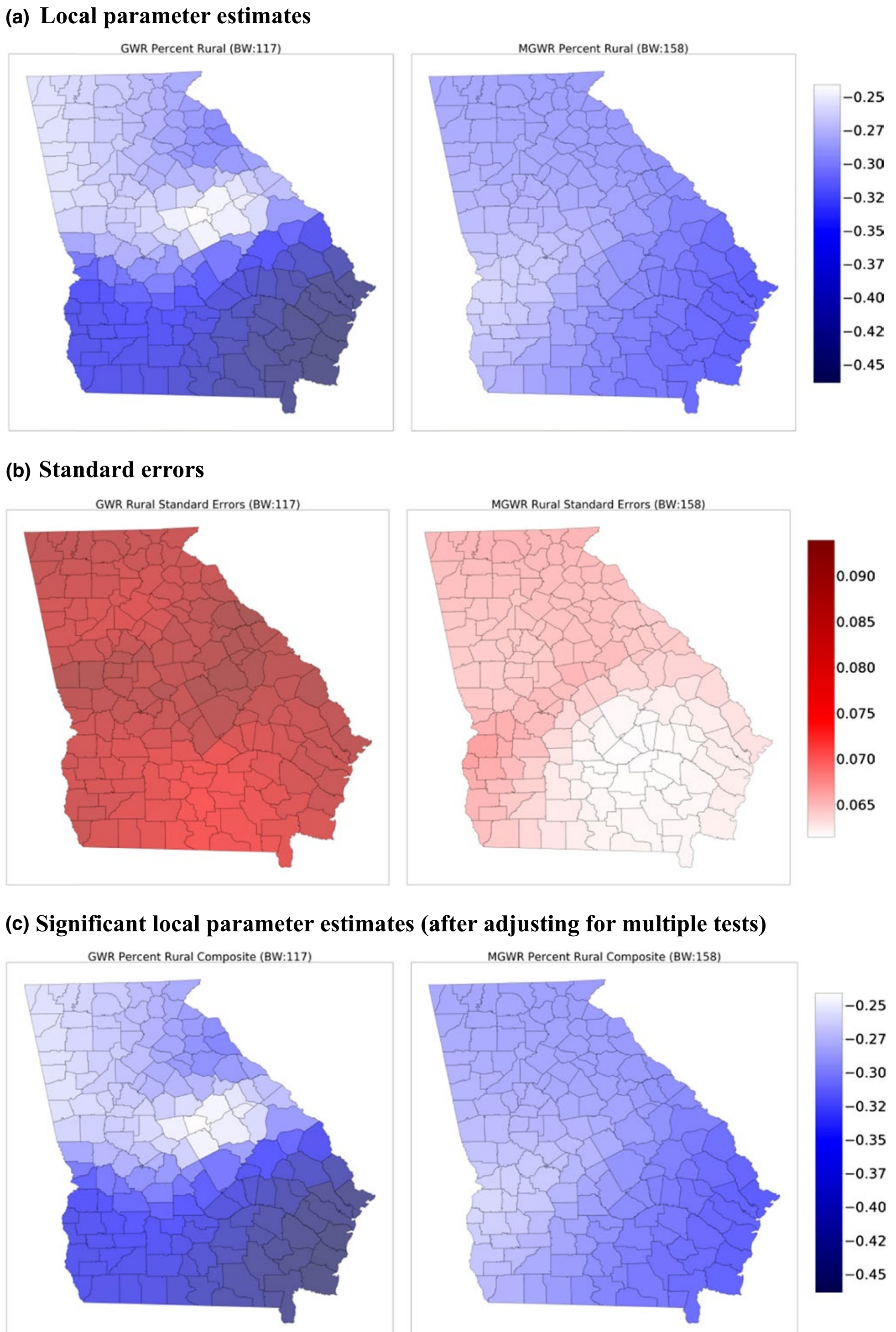

Figure 7. Rural.

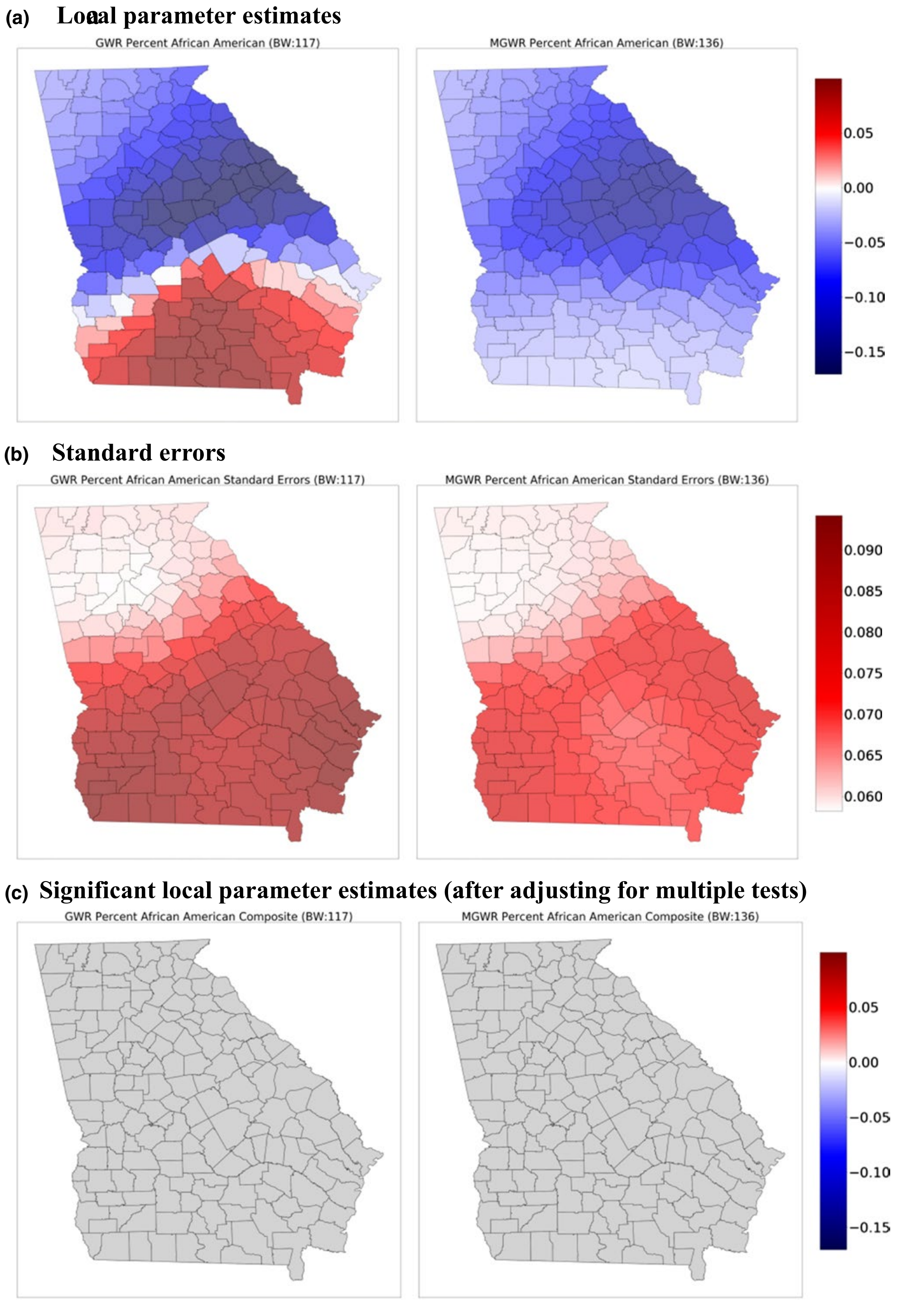

Figure 8. African American.

Table 2. GWR and MGWR Summary Statistics for the Georgia Data

	GWR	MGWR					
Diagnostic	Entire model	Entire model	Intercept	Foreign-born	Rural	African American	Sum of components
Bandwidth	117	n/a	92	101	158	136	
Effective No. of parameters	11.80	11.37	3.84	3.51	1.75	2.25	11.37
Adjusted α	0.0169	0.0176	0.0130	0.0143	0.0285	0.0221	
Critical t (95%)	2.41	2.40	2.51	2.48	2.21	2.31	
AICc	299	297					

$n = 159$.

very similar patterns with a raised percentage of people having a bachelor's education centered around Athens and the University of Georgia in the northeast and a lowered percentage in the southern part of the state. The extent of this latter region, however, is more curtailed for the MGWR results.

In Fig. 6, both the maps of the local parameter estimates and standard erros for the foreign-born covariate are very similar when derived from GWR and MGWR and consequently the resulting maps of significant local estimates are almost identical and both techniques would yield similar inferences about the role of foreign-born population and education levels which is that higher proportions of people with a bachelor's education are positively associated with higher levels of foreign-born population throughout the state except for the southern counties near the Florida border.

In the MGWR results from Fig. 7, it is clear that although a slight amount of variation in the local parameters exists, the effect of the rural population variable is effectively global. The optimized bandwidth is just about as large as it can get (158 out of 159) and the maps of the local parameter estimates, standard errors, and significant local estimates are very similar throughout the state.[6] The inference from the MGWR results is that the association between the percentage of people with a bachelor's education and the percentage of the population born abroad is significantly positive throughout the state. Although the inferential results from GWR are similar, some difference in the three sets of maps are clear—a result of the optimized GWR bandwidth being 117 instead of the MGWR value of 158. There is a greater variation across the state in all three of the GWR-based maps with a spurious "banding" of intensity of association apparent from north to south. The standard errors of the parameter estimates are consistently larger for GWR than for MGWR but especially so in the southeast of the state.

In Fig. 8, although there is some variation in the local estimates for the African American percentage, the local standard errors for each location are relatively high and no location in either the GWR or MGWR results exhibits any significant association between educational attainment levels and the percentage of African American population.

These results are informative because they demonstrate the similarity between the GWR and MGWR results when the optimal covariate-specific bandwidths are all similar to that of the single GWR bandwidth. This is reassuring because the MGWR standard errors derived from (24) yield, as they should, similar values to those obtained in GWR.

Summary

The recent development of MGWR by Fotheringham, Yang, and Kang (2017) significantly adds to the GWR framework by allowing the optimization of covariate-specific bandwidths which yield potentially valuable information on the spatial scales at which different processes opeate. This development also removes an advantage that the Bayesian non-separable spatially varying coefficients model had over the GWR framework. However, lacking in the MGWR framework was the ability to calculate standard errors for the local parameter estimates and the ENP for both the overall model and each covariate. This paper shows how both these diagnostic statsitics can be computed for MGWR and demonstrates this for both a simulated and an empirical data set. The key to this advance is the derivation of both GWR and MGWR as GAMs from which a hat matrix can be computed and which allows the computation of local standard errors and ENP values.

The derivation of MGWR with a full set of local regression diagnostics is a significant advance in spatial modeling not only because it offers the ability to examine the spatial scales over which different processes operate and to make inferences about local parameter estimates but also because the framework is eminently expandable and scable and therefore arguably offers greater potential for new developments and empirical applications than the equivalent Bayesian framework. A user-friendly suite of software for MGWR that includes the advances made to the model's inferential framework described here is available at https://sgsup.asu.edu/sparc/software/mgwr.

Notes

1 Wood(2017) provides a very useful general simulation method to approximate confidence intervals which involves iterative resampling and re-fitting of models. In each step, one first samples new local coefficients from the estimated slopes and parameter variance-covariance matrix. Then an artificial auxiliary response vector is constructed using these re-sampled local slopes and the model re-fit. This approximates the posterior distribution of the slopes and bandwidth parameters, assuming that the estimates obtained during backfitting reflect the true estimates, and whose sampling distributions are stable. While this approach is useful in its generality, we aim for an explicit, theory-driven formal statement of standard errors specific to the MGWR specification, not a generic Monte Carlo approach to approximate them.

2 The local parameters for the MGWR calibration can be initialized as 0, or as the parameter estimates from a global OLS model or as the local parameter estimates from a GWR calibration (Fotheringham et al. 2017).

3 A GWR-specific adjustment to a false discovery rate (FDR)-based correction (Benjamini and Hochberg 1995) for multiple hypothesis testing is not currently available but is the subject of active investigation. Even though there are debates (deCastro and Singer 2006) on the relative efficiacy of FWER and FDR-based corrections for multiple hypothesis testing, there is no doubt that the use of a GWR-specific FWER adjustment for multiple hypothesis tests is far superior to a FDR adjustment which does not account for the appropriate level of dependency in the tests.

4 The MGWR calibration is initialized with the GWR estimates and the calibration process is deemed to have converged once the score of change in the GWR smooth functions (SOC-f) between consecutive backfitting iterations is smaller than 10^{-5}, a threshold suggested by Fotheringham et al. (2011) based on some Monte Carlo experiments. The same initialization and termination criteria are applied to the analysis of the empirical data set. The software (MGWR 1.0) used to derive all the results in this paper can be downloaded at https://sgsup.asu.edu/sparc/software/mgwr where users can download a Windows (64 bit) and a MacOS (64 bit) version of the software plus a user manual and three sample data sets, one of which (Georgia) is used below. Users can also access and contribute to the Python source code for MGWR 1.0 at https:/github.com/pysal/mgwr.

5 These results are available from the second author.

6 Note that although the bandwidth is very close to its maximum value, there remains some spatial variation in the local parameter estimates because even though almost all the locations are used in the estimation of the local parameters, the data from these locations are still weighted. For the result to be truly global, the optimized bandwidth would need to be infinite which is operationally impossible. We therefore treat relationships as global if the optimized bandwidth is very close to the maximum possible operationally.

Acknowledgements

The authors would like to acknowledge the support of the US National Science Foundation (NSF Grant 1758786) and the National Social Foundation of China (Grants 71473008 and 13&ZD166).

[Corrections added on 23 March 2019, after first online publication: Acknowledgements section was added.]

References

Benjamini, Y., and Y. Hochberg. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society B* 57(1), 289–300.

Benjamini, Y., and D. Yekutieli. (2001). "The Control of the False Discovery Rate in Multiple Testing under Dependency." *The Annals of Statistics* 29(4), 1165–88.

Buja, A., T. Hastie, and R. Tibshirani. (1989). "Linear Smoothers and Additive Models." *The Annals of Statistics* 17(2), 453–510.

Casetti, E. (1972). "Generating Models by the Expansion Method: Applications to Geographical Research." *Geographical Analysis* 4(1), 81–91.

Cleveland, W. S. (1979). "Robust Locally weighted Regression on Smoothing Scatterplots." *Journal of the American Statistical Association* 74(368), 829–36.

da Silva, A., and A. Stewart Fotheringham. (2016). "The Multiple Testing Issue in Geographically Weighted Regression." *Geographical Analysis* 48(3), 233–47.

de Castro, M. C., and B. H. Singer. (2006). "Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Tests in Local Statistics of Spatial Association." *Geographical Analysis* 38, 180–208.

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

Fotheringham, A. S., M. E. Charlton, and C. Brunsdon (1996). "The Geography of Parameter Space: An Investigation into Spatial Non-stationarity." *International Journal of Geographic Information Science* 10, 605–27.

Fotheringham, A. S., M. E. Charlton, and C. Brunsdon. (1998). "Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis." *Environment and Planning A* 30(11): 1905–27.

Fotheringham, A. S., W. Yang, and W. Kang. (2017). "Multiscale Geographically Weighted Regression (MGWR)." *Annals of the American Association of Geographers* 107(6), 1247–65.

Finley, A. O. (2011). "Comparing Spatially-varying Coefficients Models for Analysis of Ecological Data with Non-stationary and Anisotropic Residual Dependence." *Methods in Ecology and Evolution* 2(2), 143–54.

Gelfand, A. E., H.-J. Kim, C. F. Sirmans, S. Banerjee. (2003). "Spatial Modeling with Spatially Varying Coefficient Processes." *Journal of the American Statistical Association* 98(462), 387–96.

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

Hastie, T., and R. Tibshirani. (1990). Generalized Additive Models. London: Chapman and Hall/CRC Press.

LeSage, J. P. (2004). "A Family of Geographically Weighted Regression Models." In *Advances in Spatial Econometrics*, 241–264, edited by L. Anselin, R. J. G. M. Florax and S. J. Rey. Berlin: Springer.

Miller, R. G. (1981). Simultaneous Statistical Inference. New York: Springer-Verlag.

Oshan, T., and A. S. Fotheringham (2018). "A Comparison of Spatially Varying Coefficient Estimates Using Geographically Weighted and Spatial-Filter-Based Techniques." *Geographical Analysis* 50(1), 53–75.

Tukey, J. W. (1991). "The Philosophy of Multiple Comparisons." *Statistical Science* 6(1), 100–16.

Wolf, L. J., T. M. Oshan, and A. S. Fotheringham. (2018). "Single and Multiscale Models of Process Spatial Heterogeneity." *Geographical Analysis* 50(3), 223–46. https://doi.org/10.1111/gean.12147

Wood, S. N. (2017) General Additive Models: An Introduction with R, 2nd ed. London: Chapman and Hall/CRC Press.