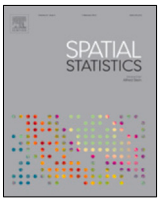
Spatial Statistics 38 (2020) 100453

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



On the measurement of bias in geographically weighted regression models☆

Hanchen Yu [a], A. Stewart Fotheringham [a,*], Ziqi Li [a], Taylor Oshan [b], Levi John Wolf [c]

[a] *Spatial Analysis Research Center, School of Geographical Sciences and Urban Planning, Arizona State University, USA*
[b] *Center for Geospatial Information Science, Department of Geographical Sciences, University of Maryland, USA*
[c] *School of Geographical Sciences, University of Bristol, UK*

ARTICLE INFO

Article history:
Received 14 April 2019
Received in revised form 10 December 2019
Accepted 14 May 2020
Available online 22 May 2020

Keywords:
Geographically weighted regression
Multiscale geographically weighted regression
Bias

ABSTRACT

Under the realization that Geographically Weighted Regression (GWR) is a data-borrowing technique, this paper derives expressions for the amount of bias introduced to local parameter estimates by borrowing data from locations where the processes might be different from those at the regression location. This is done for both GWR and Multiscale GWR (MGWR). We demonstrate the accuracy of our expressions for bias through a comparison with empirically derived estimates based on a simulated dataset with known local parameter values. By being able to compute the bias in both models we are able to demonstrate the superiority of MGWR. We then demonstrate the utility of a corrected Akaike Information Criterion statistic in finding optimal bandwidths in both GWR and MGWR as a trade-off between minimizing both bias and uncertainty. We further show how bias in one set of local parameter estimates can affect the bias in another set of local estimates. The bias derived from borrowing data from other locations appears to be very small.

☆ The authors would like to acknowledge the support of the U.S. National Science Foundation (NSF Grant 1758786) and the support of the National Social Science Foundation of China (Grants 17ZDA055).

* Corresponding author.
E-mail addresses: hanchenyu@pku.edu.cn (H. Yu), stewart.fotheringham@asu.edu (A.S. Fotheringham), lziqi@asu.edu (Z. Li), toshan@asu.edu (T. Oshan), levi.john.wolf@bristol.ac.uk (L.J. Wolf).

https://doi.org/10.1016/j.spasta.2020.100453
2211-6753/

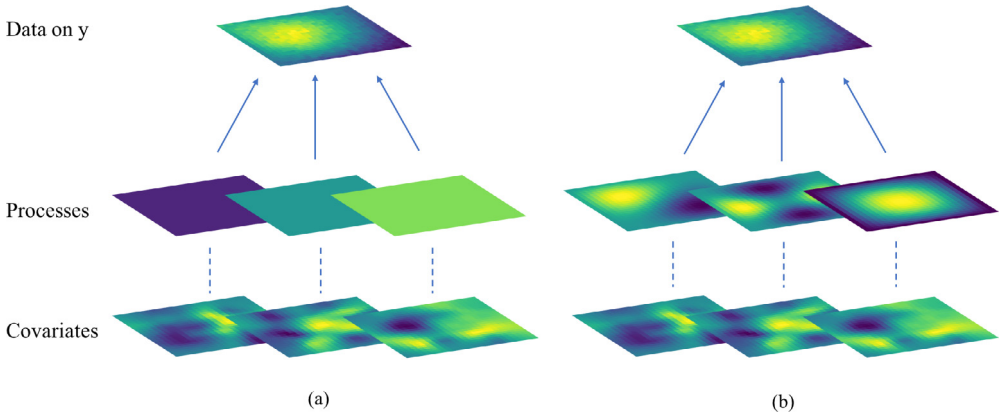


Fig. 1. Two different data generating processes.

1. Introduction

Although our understanding of the world can be advanced by purely theoretical constructs (the prediction of as-yet-unobserved sub-atomic particles, for example), perhaps a greater contribution to this understanding arises from observing aspects of the real world and advancing theories which support these observations (the motion of the planets, for example). In following this latter approach, an important distinction to be made is that between the data we observe and the processes that have produced these data. Formally, if we state the relationship

$$y_i = f_1(x_{1i}), f_2(x_{2i}), \ldots, f_m(x_{mi}) \tag{1}$$

the datum we observe at each location i is represented by y_i and these values are related to a set of covariates $x_1 \ldots x_m$ also observed at each location i through a set of processes $f_1 \ldots f_m$. Much effort is given to trying to identify the processes $f_1 \ldots f_m$, which are often unobservable, given measurements of y, $x_1 \ldots x_m$ for each location i. In traditional models of the real world, the processes $f_1 \ldots f_m$. are assumed to be stationary over space: that is, the same set of processes exists at all locations so that spatial variations in y_i arise solely from spatial variations in the covariates $x_{1i} \ldots x_{mi}$. This situation is described in Fig. 1a.

An alternative view of the real world, as described by Fig. 1b, is that spatial variation in y_i might be produced not only by spatial variation in the covariates $x_1 \ldots x_m$ but also by variations in the processes $f_1 \ldots f_m$. That is, the model of the real world described by Eq. (1) should be replaced with:

$$y_i = f_{1i}(x_{1i}), f_{2i}(x_{2i}), \ldots, f_{mi}(x_{mi}) \tag{2}$$

where the processes are allowed to vary across locations. Several types of what are known as local modeling frameworks have been proposed based on the view of the world represented in Eq. (2). Examples included Bayesian spatially varying coefficient models (Gelfand et al., 2003), eigenvector spatial filtering approaches (Griffith, 2008) and geographically weighted regression (Fotheringham et al., 2002). Here we concentrate on the latter which has seen widespread adoption across many disciplines (Cahill and Mulligan, 2007; Brown et al., 2012; Miller and Hanham, 2011; Fotheringham et al., 2015; Zou et al., 2016) and for which a more flexible approach, termed multiscale geographically weighted regression (MGWR), has recently been developed (Fotheringham et al., 2017).

2. Geographically weighted regression as a data-borrowing technique

Suppose we want to calibrate a model whose general formula is that given by Eq. (2) and where we believe the processes that generate our observations on y are spatially varying. If we had sufficient data on y and $x_1 \ldots x_m$ at each location, we could calibrate a separate model for each location and examine the spatial variability of the resulting local parameter estimates. Unfortunately, such a situation is rare; the more usual situation is one where only a single measurement of y and each of the covariates $x_1 \ldots x_m$ is recorded at each location. The traditional approach would then be to ignore any possible spatial variation in the processes represented by $f_{1i} \ldots f_{mi}$ and to calibrate a global model which would generate a single estimate of each f_j and which would be in effect an average of the set of the location-specific f_{ji} processes which are unobservable. An alternative, which is provided by the geographically weighted regression (GWR) framework, is to calibrate a separate model for each location by 'borrowing' data from surrounding locations.

Although borrowing data from nearby locations allows the measurement of location-specific processes (or, rather, proxies for these processes in terms of location-specific parameter estimates), it clearly introduces bias into the local estimates because the data that are borrowed from other locations are the product of processes which may be different from those acting at the focal local for the regression. This bias is mitigated by weighting the data that are borrowed from surrounding locations from 1 to 0 to reflect their proximity to the regression focal point with data from nearer locations having a greater weight than data from more distant locations. This weighting is typically achieved by adopting a continuous distance-based kernel function and determining an optimal bandwidth (degree of distance-decay) for this kernel from the data. This is the essence of GWR and its variants.

Despite this mitigation strategy, however, a bias in the resulting local parameter estimates still exists and the determination of this bias is the focus of this paper. We first describe the calculation of bias in local parameter estimates for GWR which has a single optimized bandwidth and then we turn to the more complex MGWR which has covariate-specific optimal bandwidths. Being able to calculate the bias in the local parameter estimates derived from both GWR and MGWR is important for several reasons:

(i) To determine the extent of this bias — if it were sufficiently large it would negate the utility of the data-borrowing framework;
(ii) To examine the extent to which the bias in GWR local parameter estimates is mitigated in MGWR which contains extra flexibility in allowing covariate-specific bandwidths;
(iii) To investigate the relationship between scale (bandwidth) misspecification and the bias contained in the local parameter estimates; and
(iv) To examine the role of the bias–variance trade-off in determining the optimal bandwidth. Currently, we typically use some variant of an information-criterion statistic to determine the optimal bandwidth which is assumed to measure a trade-off between bias and uncertainty in the local parameter estimates. By being able to measure the bias directly, we could both test the efficacy of various information criterion statistics in measuring this trade-off and possibly replace a single goodness-of-fit criterion with direct measures of both bias and uncertainty.

The paper proceeds as follows. First, the bias in a GWR model is identified and then this framework is extended to MGWR. In both cases, we derive analytical expressions for the bias and compare these to the measurement of bias obtained empirically from known local parameter surfaces. Finally, we compare the bias GWR to that in MGWR and make comments on the utility of our findings.

3. Derivation of the analytical expression for the data-borrowing bias in GWR parameter estimates

3.1. GWR formulation

Geographically Weighted Regression (GWR) allows parameter estimates from a linear regression model to vary locally. It calibrates a separate regression model at each location of interest by

borrowing data from nearby observations and weighting them by their distance from the regression point. GWR is formulated as

$$y_i = \sum_{j=1}^{m} \beta_{ij} x_{ij} + \varepsilon_i \tag{3}$$

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

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{i}} = \left(\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}_{\boldsymbol{i}} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}_{\boldsymbol{i}} \boldsymbol{y}, i \in \{1, 2, \ldots, n\} \tag{4}$$

where $\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}$ is an $m \times 1$ column vector of parameter estimates at location i, $\boldsymbol{X}$ is the $n \times m$ matrix form of the predictor variables, $\boldsymbol{y}$ is the $n \times 1$ vector of the response variable, and $\boldsymbol{W}_{\boldsymbol{i}}$ is a diagonal spatial weight matrix that weights each observation in terms of its distance from location i. The weighting scheme consists of selecting a kernel function and a bandwidth parameter that indicates the relationship between weight and proximity. A popular data-borrowing scheme is the adaptive (i.e., k-nearest neighbors) bi-square kernel function, which has the advantageous interpretation that the bandwidth parameter defines the number of neighbors which have non-zero weights in the local regression. This bandwidth parameter is typically selected by optimizing a corrected Akaike information criterion (AICc). The AICc penalizes smaller bandwidths to avoid over-fitting and maintains a balance between parameter estimate bias and variance and is defined as

$$\mathrm{AIC_c} = 2n \ln\left(\frac{RSS}{n}\right) + n \ln 2\pi + n\left\{\frac{n + tr(\boldsymbol{S})}{n - 2 - tr(\boldsymbol{S})}\right\} \tag{5}$$

where n is the number of observations, RSS is the residual sum of squares, and $tr(\boldsymbol{S})$ is the trace of the hat matrix and the Effective Number of Parameters (ENP) of the model. This nearest-neighbor bi-square kernel data-borrowing scheme parameterized via AICc optimization is utilized throughout this paper.

Yu et al. (2018) express the GWR formulation from Eq. (1) as a Generalized Additive Model (GAM) as

$$\boldsymbol{y} = \sum_{\boldsymbol{1}}^{\boldsymbol{m}} \boldsymbol{f}_{\boldsymbol{j}} + \boldsymbol{\varepsilon} \tag{6}$$

where the response variable $\boldsymbol{y}$ is expressed as a linear combination of m smooth terms $\boldsymbol{f}_{1\ldots\boldsymbol{m}}$ and an i.i.d error term, $\boldsymbol{\varepsilon}$. The jth smooth term $\boldsymbol{f}_{\boldsymbol{j}}$ is comprised of the GWR data-borrowing scheme, also known as a smoothing function, applied to the jth predictor variable. In GWR a single smoothing function is applied to all of the predictor variables, which means the same bandwidth parameter or data-borrowing range is associated with each covariate and associated local parameter estimate surface. Multi-scale GWR (MGWR) (Fotheringham et al., 2017) provides a more generalized extension of this framework in which each smoothing function, $\boldsymbol{f}_{\boldsymbol{j}}$, has its own data-borrowing scheme and bandwidth parameter. As a result, each parameter estimate surface is free to vary with a unique degree of spatial smoothness, and is therefore more appropriate for capturing processes operating at different spatial scales.

In the following sections, we derive analytical expressions for the bias in the local parameter estimates generated by GWR and MGWR and decompose this bias into covariate-specific contributions.

3.2. Data-borrowing bias in GWR parameter estimates

In this section, we derive the analytical form of the bias γ_{ij} for each GWR parameter estimate $\hat{\beta}_{ij}$ of the jth covariate at location i. For convenience, let $\boldsymbol{e}_{\boldsymbol{j}}$ be the jth row of the (m, m) identity

matrix, so that

$$\boldsymbol{e_j} = (\underbrace{0, 0, 0, \ldots, 0}_{\text{j}-1}, 1, \underbrace{0, 0, \ldots, 0}_{\text{m}-\text{j}}) \tag{7}$$

and

$$\hat{\beta}_{ij} = \boldsymbol{e_j}\hat{\boldsymbol{\beta}}_{\boldsymbol{i}} \tag{8}$$

where $\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}$ is a column vector of parameter estimates with dimensions *(m, 1)* at location *i*. Using Eqs. (7) and (8), the conditional expectation of the parameter estimate $\hat{\beta}_{ij}$ on bandwidths vector $\boldsymbol{bw} = (\text{bw}_1, \text{bw}_2, \ldots, \text{bw}_\text{m})$ can be written as

$$\begin{aligned} \text{E}\left(\hat{\beta}_{ij}|bw\right) &= \text{E}\left(\boldsymbol{e_j}\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}|bw\right) = \text{E}\left(\boldsymbol{e_j}\boldsymbol{P_i}\boldsymbol{y}|bw\right) \\ &= \text{E}\left(\boldsymbol{e_j}\boldsymbol{P_i}\left(\sum_{\mathbf{1}}^{\boldsymbol{m}}\boldsymbol{f_k} + \boldsymbol{\varepsilon}\right)|bw\right) = \text{E}\left(\boldsymbol{e_j}\boldsymbol{P_i}\sum_{\mathbf{1}}^{\boldsymbol{m}}\boldsymbol{f_k}|bw\right) + \text{E}\left(\boldsymbol{e_j}\boldsymbol{P_i}\boldsymbol{\varepsilon}|bw\right) \end{aligned} \tag{9}$$

where

$$\boldsymbol{P_i} = \left(\boldsymbol{X^T}\boldsymbol{W_i}\boldsymbol{X}\right)^{-1}\boldsymbol{X^T}\boldsymbol{W_i} \tag{10}$$

So, given the assumption that error $\boldsymbol{\varepsilon}$ is independent of $\boldsymbol{X}$, so $\text{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}] = 0$, and we have

$$\text{E}\left(\hat{\beta}_{ij}|bw\right) = \text{E}\left(\boldsymbol{e_j}\boldsymbol{P_i}\sum_{\mathbf{1}}^{\boldsymbol{m}}\boldsymbol{f_k}|bw\right) = \boldsymbol{e_j}\boldsymbol{P_i}\sum_{\mathbf{1}}^{\boldsymbol{m}}\boldsymbol{f_k} \tag{11}$$

Then we add and subtract the same β_{ij} to Eq. (11) to give

$$\text{E}\left(\hat{\beta}_{ij}|bw\right) = \boldsymbol{e_j}\boldsymbol{P_i}\sum_{\mathbf{1}}^{\boldsymbol{m}}\boldsymbol{f_k} = \boldsymbol{e_j}\boldsymbol{P_i}\sum_{\mathbf{1}}^{\boldsymbol{m}}\boldsymbol{f_k} + \beta_{ij} - \beta_{ij} \tag{12}$$

and

$$\text{E}\left(\hat{\beta}_{ij}|bw\right) = \beta_{ij} - \left(\beta_{ij} - \boldsymbol{e_j}\boldsymbol{P_i}\sum_{\mathbf{1}}^{\boldsymbol{m}}\boldsymbol{f_k}\right) = \beta_{ij} - \gamma_{ij} \tag{13}$$

where the bias of parameter estimate $\hat{\beta}_{ij}$ can be analytically expressed by

$$\gamma_{ij} = \beta_{ij} - \text{E}\left(\hat{\beta}_{ij}|bw\right) = \beta_{ij} - \boldsymbol{e_j}\boldsymbol{P_i}\sum_{\mathbf{1}}^{\boldsymbol{m}}\boldsymbol{f_k} \tag{14}$$

3.3. A covariate-specific decomposition of the data-borrowing bias in GWR parameter estimates

From Eq. (14) we can see that γ_{ij} is dependent on not only the *j*th term $\boldsymbol{f_j}$ but also on the other terms $\boldsymbol{f_{i\neq j}}$. Therefore, we would like to decompose γ_{ij} into covariate-specific contributions to parameter estimate bias. Rewriting equation (8), we have

$$\beta_{ij} = \boldsymbol{e_j}\boldsymbol{\beta_i} = \boldsymbol{e_j}\boldsymbol{I}\boldsymbol{\beta_i} = \boldsymbol{e_j}\left(\boldsymbol{X^T}\boldsymbol{W_i}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X^T}\boldsymbol{W_i}\boldsymbol{X}\right)\boldsymbol{\beta_i} = \boldsymbol{e_j}\boldsymbol{P_i}\boldsymbol{X}\boldsymbol{\beta_i} \tag{15}$$

and substituting (15) into (14) we then have the bias of the local parameter estimate $\hat{\beta}_{ij}$ as

$$\begin{aligned}
\gamma_{ij} &= \boldsymbol{e_j P_i}\left(\boldsymbol{X\beta_i} - \sum_{\mathbf{1}}^{\boldsymbol{m}} \boldsymbol{f_k}\right) \\
&= \boldsymbol{e_j P_i}\left(\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}\begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{im} \end{pmatrix} - \begin{pmatrix} \sum_{k=1}^{m} x_{1k}\beta_{1k} \\ \sum_{k=1}^{m} x_{2k}\beta_{2k} \\ \vdots \\ \sum_{k=1}^{m} x_{nk}\beta_{nk} \end{pmatrix}\right) \\
&= \boldsymbol{e_j P_i}\left(\sum_{k=1}^{m}\begin{pmatrix} x_{1k}\beta_{ik} \\ x_{2k}\beta_{ik} \\ \vdots \\ x_{nk}\beta_{ik} \end{pmatrix} - \sum_{k=1}^{m}\begin{pmatrix} x_{1k}\beta_{1k} \\ x_{2k}\beta_{2k} \\ \vdots \\ x_{nk}\beta_{nk} \end{pmatrix}\right) \\
&= \boldsymbol{e_j P_i}\sum_{k=1}^{m}\begin{pmatrix} x_{1k}(\beta_{ik} - \beta_{1k}) \\ x_{2k}(\beta_{ik} - \beta_{2k}) \\ \vdots \\ x_{nk}(\beta_{ik} - \beta_{nk}) \end{pmatrix} \\
&= \sum_{k=1}^{m}\boldsymbol{e_j P_i}\begin{pmatrix} x_{1k}(\beta_{ik} - \beta_{1k}) \\ x_{2k}(\beta_{ik} - \beta_{2k}) \\ \vdots \\ x_{n}(\beta_{ik} - \beta_{nk}) \end{pmatrix} = \sum_{k=1}^{m}\theta_{ijk}
\end{aligned} \tag{16}$$

where

$$\theta_{ijk} = \boldsymbol{e_j P_i}\begin{pmatrix} x_{1k}(\beta_{ik} - \beta_{1k}) \\ x_{2k}(\beta_{ik} - \beta_{2k}) \\ \vdots \\ x_{nk}(\beta_{ik} - \beta_{nk}) \end{pmatrix} \tag{17}$$

Here θ_{ijk} is the contribution to the bias γ_{ij} of parameter estimate $\hat{\beta}_{ij}$ from the parameter surface $\boldsymbol{\beta_k}$. If $\boldsymbol{\beta_k}$ were a flat surface where the true parameters are constant across space, then θ_{ijk} would be zero. This indicates that a flat parameter surface with an asymptotically infinite bandwidth does not bias estimates of the other parameters. However, when $\boldsymbol{\beta_k}$ is not constant, θ_{ijk} is non-zero indicating that $\boldsymbol{\beta_k}$ will cause bias in the estimates of the other parameters. Eq. (13) also demonstrates that the GWR parameter estimate $\hat{\beta}_{ij}$ is often biased, since γ_{ij} can only be zero if all covariates have asymptotically infinite bandwidths; a situation where GWR is equivalent to OLS.

4. An example of the data-borrowing bias in GWR parameter estimates

A simulated dataset is used to examine the data-borrowing bias in GWR based on the analytical solution given above. Three known parameter surfaces are generated from Eqs. (18)–(20).

$$\boldsymbol{\beta_0} = 3 \tag{18}$$

$$\boldsymbol{\beta_1} = 1 + \frac{1}{12}(u + v) \tag{19}$$

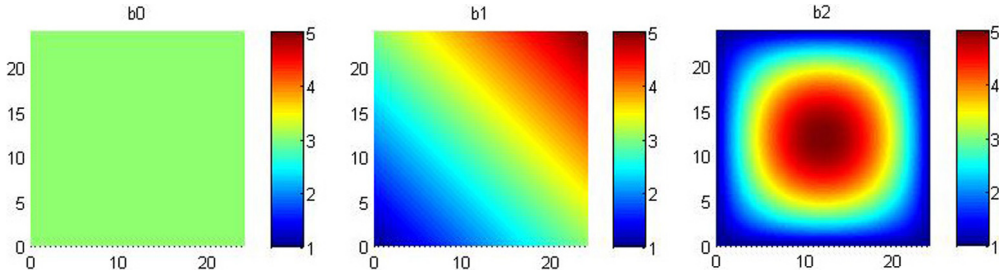


Fig. 2. Known parameter surfaces of three synthesized processes.

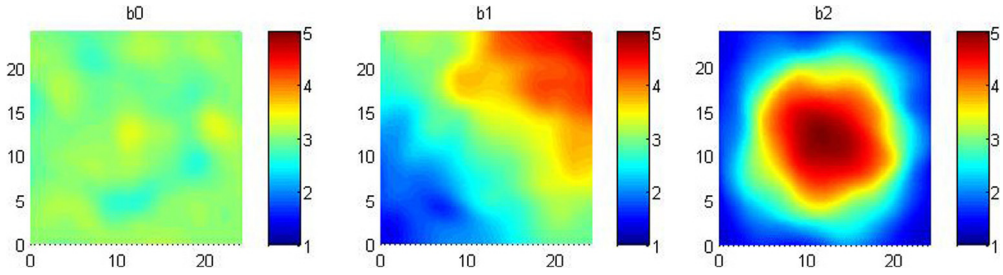


Fig. 3. GWR parameter estimates from model Eq. (21).

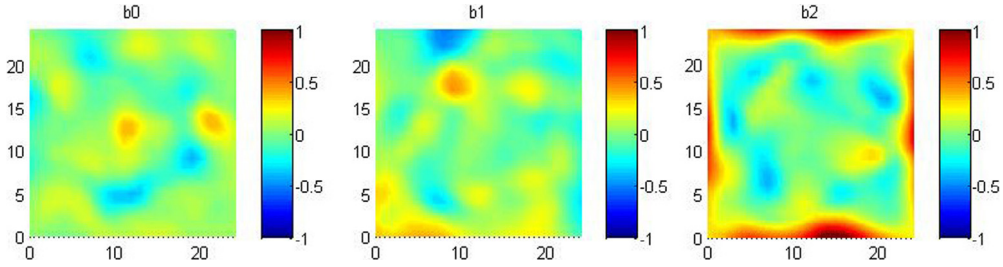


Fig. 4. Empirical bias in parameter estimates by taking difference of Figs. 2 and 3.

$$\boldsymbol{\beta_2} = 1 + \frac{1}{324}[36 - (6 - \frac{u}{2})^2][36 - (6 - \frac{v}{2})^2] \tag{20}$$

where u and v are the x and y coordinates of a 25 by 25 grid. These known parameter surfaces are displayed in Fig. 2.

Then a synthetic response variable $\boldsymbol{y}^*$ is constructed using known parameter surfaces $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$ with randomly drawn predictor variables $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ from normal distribution of $N\,(0,\,1\,)$ as well as a randomly drawn error from $N(0,\,0.5)$ as described in Eq. (21).

$$\boldsymbol{y}^* = \boldsymbol{\beta}_0 + \boldsymbol{\beta_1 X_1} + \boldsymbol{\beta_2 X_2} + \boldsymbol{\varepsilon} \tag{21}$$

A GWR model is calibrated based on this simulated dataset using an adaptive bi-square kernel which generates an optimal bandwidth of 50 nearest neighbors when using *AICc* as the optimization criterion. The local parameter estimates from Eq. (21) are shown in Fig. 3. By taking the difference between Figs. 2 and 3, the bias in the local parameter estimates is shown in Fig. 4 whereas Fig. 5 depicts the analytical bias computed from Eq. (14).

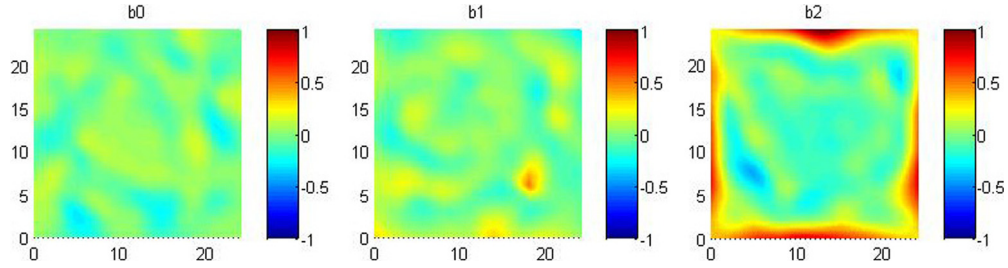


Fig. 5. Analytically-derived GWR bias computed from Eq. (14).

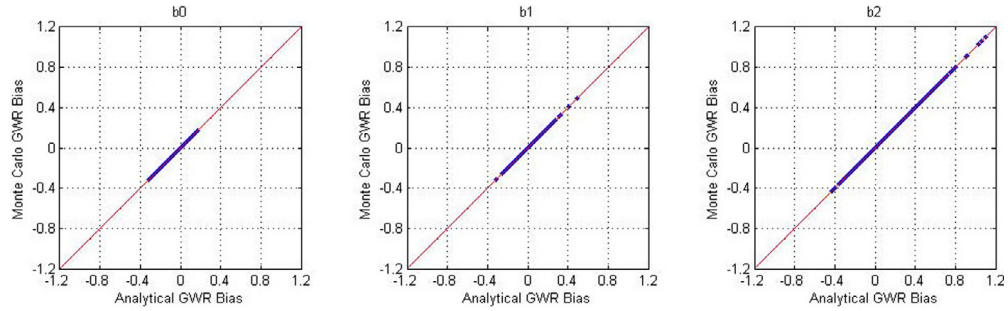


Fig. 6. Comparison of the analytically derived parameter bias from Eq. (21) and the mean of the errors in each local estimate from 10,000 GWR calibrations based on different drawings of y_i.

It can be seen from Figs. 4 and 5 that the bias is close to zero for most locations and is relatively random compared to the magnitude and spatial patterning of the true parameter values in Fig. 2. The one exception to this is on the $\boldsymbol{\beta_2}$ surface where there is clearly an edge effect that causes bias to be higher around the border of the study area. This analytical bias was confirmed via Monte Carlo simulations of 10,000 realizations of y_i values with an added random error drawn from *N(0,0.5)* to the model in Eq. (21). The results showing the relationship between the analytical bias from Eq. (21) and the mean of the 10,000 simulations for each of the 625 local estimates of the three parameters are shown in Fig. 6.

The results above are based on the optimal bandwidth of 50 nearest neighbors being selected by minimizing a corrected AIC statistic as shown in Fig. 7.

However, given that the data-borrowing bias in GWR local parameter estimates can be computed from Eq. (21) for any bandwidth value, we can use this to examine the sensitivity of the bias to variations in the bandwidth and also to examine the efficacy of using AICc as an optimizing criterion. In Fig. 8 we demonstrate the sensitivity of both bias and uncertainty (variance) in local parameter estimates to variations in the bandwidth. For each bandwidth, Fig. 8 displays the mean percentage bias in the local parameter estimates averaged over the three sets of estimates for all 625 locations (blue line) and the equivalent mean standard error as an indicator of the uncertainty attached to the local parameter estimates. (green line). As expected, as the bandwidth increases and more data are borrowed from increasingly distant locations, the bias increases and the variance decreases. However, while the bias increases steadily as the bandwidth increases, the variance flattens out relatively quickly. The optimal bandwidth selected on the basis of AICc minimization seems reasonable — it provides a set of local parameter estimates with both low bias and low variance. It is tempting to suggest that the optimal bandwidth should be where the lines cross (somewhere around 40 nearest neighbors) but this has no basis statistically and is a graphical nicety only. A bandwidth of 40 instead of 50 would produce less bias but at the cost of increasing uncertainty.

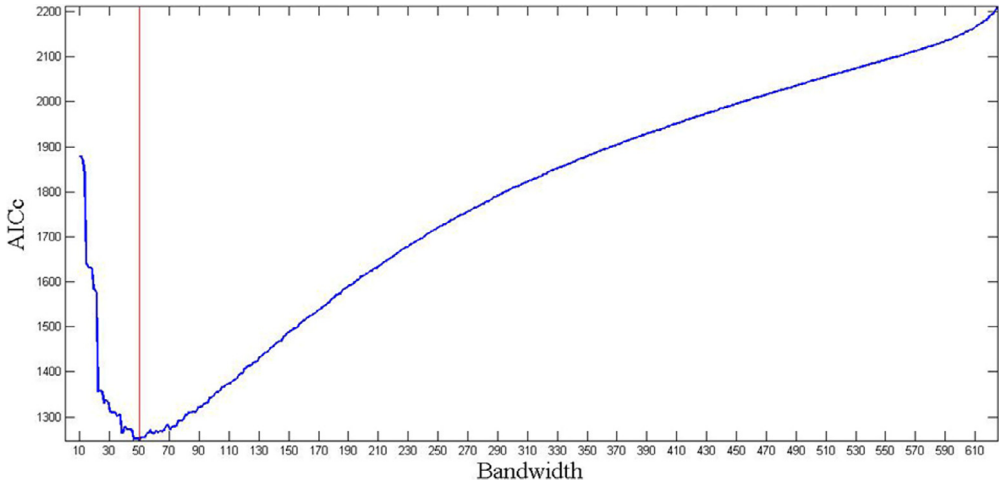


Fig. 7. The relationship between AICc and bandwidth. The optimal bandwidth that minimizes AICc in the model calibration is 50 nearest neighbors.

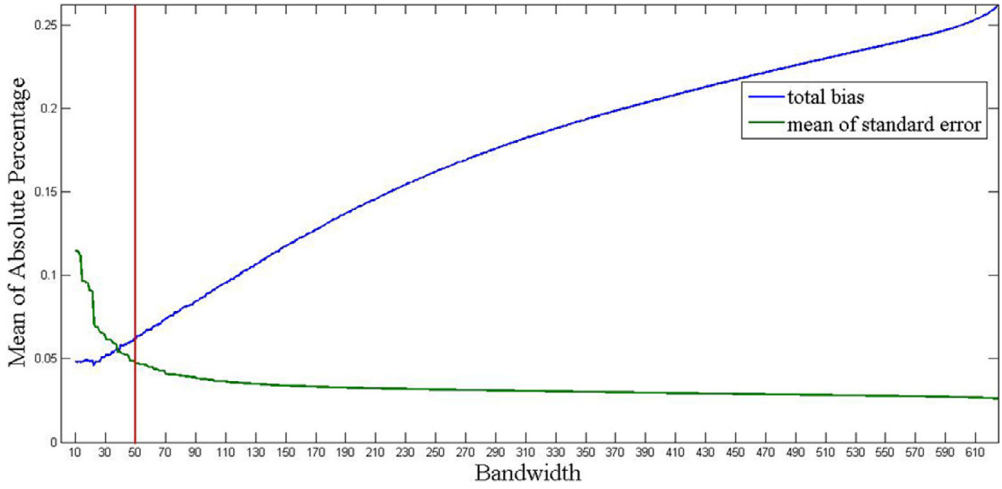


Fig. 8. Bias–Variance (Blue–Green) tradeoff at different bandwidths. The vertical red line shows the optimal bandwidth using AICc as the selection criterion. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, Eq. (17) shows how the bias in one set of local parameter estimates depends in part on the bias in the estimates of the other local parameters. Formally, from Eq. (17) θ_{ijk} is the contribution to the bias γ_{ij} of parameter estimate $\hat{\beta}_{ij}$ from the parameter surface β_k and these values are displayed in Fig. 9 for each of the three sets of parameters across a range of bandwidths. The results indicate that the bias in one set of local estimates derived from bias in the other sets of local parameter estimates is negligible and this is the case across all bandwidths. The bias in the local intercept is always very low whereas the bias in β_2 increases rapidly as the bandwidth increases because the pattern of the true values of β_2 exhibits relatively large spatial heterogeneity. The bias in the estimates of β_1 tends to increase less dramatically as the bandwidth increases because the pattern of the true values of β_1 exhibits less spatial heterogeneity than that of the true values of β_2.

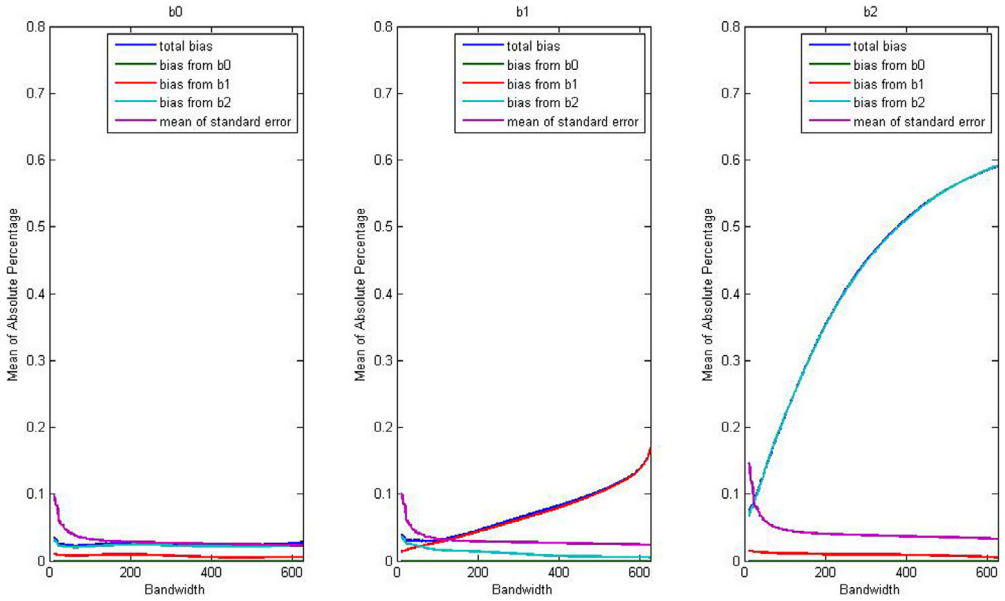


Fig. 9. The contributions to the bias in one set of local parameter estimates from the bias in the other local parameter estimates in the model at different bandwidths computed from Eq. (17).

5. Derivation of the analytical expression for the data-borrowing bias in MGWR parameter estimates

5.1. MGWR formulation

Compared to classic GWR model which assumes bandwidth to be the same across covariates, MGWR (Fotheringham et al., 2017) relaxes this assumption and allows covariate-specific bandwidths to be optimized. The formulation of MGWR is the same as the GAM formulation of GWR in Eq. (6) except that each component of the smoothing function $\boldsymbol{f_j}$ is calibrated with a covariate-specific bandwidth bw_j. Then the response variable y can be expressed as the sum of spatially varying surface components with different degrees of smoothness in each.

$$\boldsymbol{y} = \sum_{\mathbf{1}}^{\boldsymbol{m}} \boldsymbol{f}_{\boldsymbol{bwj}} + \boldsymbol{\varepsilon} \tag{22}$$

The calibration process of MGWR follows the traditional back-fitting in GAM (Hastie and Tibshirani, 1990; Buja et al., 1989). It starts by initializing the parameter estimates from a GWR model, then successive univariate GWR models $GWR\{\hat{\boldsymbol{f}}_{\boldsymbol{j}} + \hat{\boldsymbol{\varepsilon}} \sim \boldsymbol{X_j}\}$ are calibrated using the current fitted component plus the current partial residual regressed against the current covariate. This is done successively across each of the covariates to complete one iteration of the calibration procedure. The second iteration begins using the updated values of the fitted components and the back-fitting algorithm iterates in such a way until parameter estimates do not change between iterations.

5.2. Data-borrowing bias in MGWR parameter estimates

Because MGWR is calibrated by a back-fitting process, the derivation of an analytical expression for the data-borrowing bias in MGWR parameter estimates is more complex than in GWR. To begin

with, we express the expectation of the fitted term $\hat{\boldsymbol{f}}_{\boldsymbol{j}}$ as

$$\begin{aligned} \mathrm{E}\left(\hat{\boldsymbol{f}}_{\boldsymbol{j}}|bw\right) &= \mathrm{E}\left(\begin{pmatrix} x_{1j}\hat{\beta}_{1j} \\ x_{2j}\hat{\beta}_{2j} \\ \vdots \\ x_{nj}\hat{\beta}_{nj} \end{pmatrix} \middle| bw\right) = \mathrm{E}\left(\begin{pmatrix} x_{1j}(\beta_{1j}-\gamma_{1j}) \\ x_{2j}(\beta_{2j}-\gamma_{2j}) \\ \vdots \\ x_{nj}(\beta_{nj}-\gamma_{nj}) \end{pmatrix} \middle| bw\right) \\ &= \mathrm{E}\left(\begin{pmatrix} x_{1j}\beta_{1j}-x_{1j}\gamma_{1j} \\ x_{2j}\beta_{2j}-x_{2j}\gamma_{2j} \\ \vdots \\ x_{nj}\beta_{nj}-x_{nj}\gamma_{nj} \end{pmatrix} \middle| bw\right) = f_{\boldsymbol{j}} - \boldsymbol{\tau_j} \end{aligned} \tag{23}$$

where $\boldsymbol{\tau_j}$ is a column vector of covariate x_{ij} times the bias γ_{ij} for location $i \in \{1, 2, \ldots, n\}$

$$\boldsymbol{\tau_j} = \begin{pmatrix} x_{1j}\gamma_{1j} \\ x_{2j}\gamma_{2j} \\ \vdots \quad \vdots \\ x_{nj}\gamma_{nj} \end{pmatrix}_{nx1} \tag{24}$$

In the MGWR back-fitting procedure, each fitted term is updated successively as

$$\hat{\boldsymbol{f}}_{\boldsymbol{j}}^{*} = \boldsymbol{A_j}\left(\hat{\boldsymbol{f}}_{\boldsymbol{j}} + \hat{\boldsymbol{\varepsilon}}\right) \tag{25}$$

where $\hat{\boldsymbol{f}}_{\boldsymbol{j}}$ is the fitted term from the previous iteration, $\hat{\boldsymbol{\varepsilon}}$ is the current model residual, $\boldsymbol{A_j}$ is the GWR hat matrix of the model *GWR* $\{\hat{\boldsymbol{f}}_{\boldsymbol{j}} + \hat{\boldsymbol{\varepsilon}} \sim \boldsymbol{X_j}\}$. From Yu et al. (2018), the covariate-specific hat matrix $\boldsymbol{R_j}$ can be updated by

$$\boldsymbol{R_j^{*}} = \boldsymbol{A_j} - \boldsymbol{A_j S} + \boldsymbol{A_j R_j} \tag{26}$$

where $\boldsymbol{S}$ is the model hat matrix from MGWR. After obtaining an updated $\boldsymbol{R_j^{*}}$, the new hat matrix $\boldsymbol{S^{*}}$ can be updated accordingly by

$$\boldsymbol{S^{*}} = \boldsymbol{S} - \boldsymbol{R_j} + \boldsymbol{R_j^{*}} \tag{27}$$

and the updated fitted term $\hat{\boldsymbol{f}}_{\boldsymbol{j}}^{*}$ can be expressed by

$$\hat{\boldsymbol{f}}_{\boldsymbol{j}}^{*} = \boldsymbol{R_j^{*}}\boldsymbol{y} = \left(\boldsymbol{A_j} - \boldsymbol{A_j S} + \boldsymbol{A_j R_j}\right)\boldsymbol{y} \tag{28}$$

Then its expectation can be obtained from

$$\begin{aligned} \mathrm{E}\left(\hat{\boldsymbol{f}}_{\boldsymbol{j}}^{*}|bj\right) &= \mathrm{E}\left(\left(\boldsymbol{A_j} - \boldsymbol{A_j S} + \boldsymbol{A_j R_j}\right)\boldsymbol{y}|bj\right) \\ &= \mathrm{E}\left(\boldsymbol{A_j y}|bj\right) - \mathrm{E}\left(\boldsymbol{A_j S y}|bj\right) + \mathrm{E}\left(\boldsymbol{A_j R_j y}|bj\right) \\ &= \boldsymbol{A_j y} - \boldsymbol{A_j}\mathrm{E}\left(\boldsymbol{S y}|bj\right) + \boldsymbol{A_j}\mathrm{E}\left(\boldsymbol{R_j y}|bj\right) \end{aligned} \tag{29}$$

where

$$E\left(\boldsymbol{R_j y}|bj\right) = E\left(\hat{\boldsymbol{f}}_{\boldsymbol{j}}|bj\right) = \boldsymbol{f_j} - \boldsymbol{\tau_j} \tag{30}$$

and

$$E\left(\boldsymbol{S y}|bj\right) = E\left(\hat{\boldsymbol{y}}|bj\right) = E\left(\sum_{\boldsymbol{1}}^{\boldsymbol{m}}\hat{\boldsymbol{f}}_{\boldsymbol{j}}|bj\right) = \boldsymbol{y} - \sum_{\boldsymbol{1}}^{\boldsymbol{m}}\boldsymbol{\tau_j} \tag{31}$$

Substituting Eqs. (30) and (31) into (29), we get

$$\mathrm{E}\left(\hat{\boldsymbol{f}}_{\boldsymbol{j}}^{*}|bj\right) = \boldsymbol{A_j y} - \boldsymbol{A_j}\left(\boldsymbol{y} - \sum_{\boldsymbol{1}}^{\boldsymbol{m}}\boldsymbol{\tau_k}\right) + \boldsymbol{A_j}\left(\boldsymbol{f_j} - \boldsymbol{\tau_j}\right) = \boldsymbol{A_j f_j} + \boldsymbol{A_j}\sum_{\boldsymbol{1}}^{\boldsymbol{m}}\boldsymbol{\tau_k} - \boldsymbol{A_j \tau_j} \tag{32}$$

and the updated smoothing bias $\boldsymbol{\tau}_j^*$ of $\boldsymbol{f}_j$ is then

$$\begin{aligned}\boldsymbol{\tau}_j^* &= \boldsymbol{f}_j - \mathrm{E}\left(\hat{\boldsymbol{f}}_j^* | bj\right) = \boldsymbol{f}_j - \left(\boldsymbol{A}_j\boldsymbol{f}_j + \boldsymbol{A}_j\sum_{1}^{m}\boldsymbol{\tau}_k - \boldsymbol{A}_j\boldsymbol{\tau}_j\right) \\ &= \left(\boldsymbol{I} - \boldsymbol{A}_j\right)\boldsymbol{f}_j + \boldsymbol{A}_j(\boldsymbol{\tau}_j - \sum_{1}^{m}\boldsymbol{\tau}_k)\end{aligned} \tag{33}$$

Once the back-fitting calibration converges, Eq. (24) can be used to obtain $\boldsymbol{\gamma}_j$ by

$$\boldsymbol{\gamma}_j = \left[diag(\boldsymbol{X}_j)\right]^{-1}\boldsymbol{\tau}_j = \begin{pmatrix}\gamma_{1j}\\ \gamma_{2j}\\ \vdots\\ \gamma_{nj}\end{pmatrix}_{nx1} \tag{34}$$

where $\left[diag(\boldsymbol{X}_j)\right]^{-1}$ is an inverse of a diagonal matrix with the jth covariate $\boldsymbol{X}_j$ filling the diagonal and $\boldsymbol{\gamma}_j$ is a column vector of the bias in each of the local parameter estimates associated with covariate j so that

$$\hat{\boldsymbol{\beta}}_j = \boldsymbol{\beta}_j - \boldsymbol{\gamma}_j \tag{35}$$

5.3. Decomposing MGWR parameter bias into covariate-specific contributions

In a similar manner to decomposing the data-borrowing bias in local parameter estimates from GWR into covariate-specific contributions, we can decompose $\boldsymbol{\tau}_j$ into m components, so that

$$\boldsymbol{\tau}_j = \sum_{k=1}^{m}\boldsymbol{\alpha}_{jk} \tag{36}$$

where $\boldsymbol{\alpha}_{jk}$ is the contribution to the bias $\boldsymbol{\tau}_j$ from term $\hat{\boldsymbol{f}}_k$ and is expressed as

$$\boldsymbol{\alpha}_{jk} = \begin{pmatrix}x_{1j} & \theta_{1jk}\\ x_{2j} & \theta_{2jk}\\ \vdots & \vdots\\ x_{nj} & \theta_{njk}\end{pmatrix}_{nx1} \tag{37}$$

In the back-fitting procedure, instead of updating $\boldsymbol{\tau}_j$ as shown in Eq. (33), we update $\boldsymbol{\alpha}_{jk}$ by

$$\boldsymbol{\alpha}_{jk}^* = \begin{cases}\boldsymbol{A}_j(\boldsymbol{\alpha}_{jk} - \sum_{l=1}^{m}\boldsymbol{\alpha}_{jl}) & j \neq k\\ \left(\boldsymbol{I} - \boldsymbol{A}_j\right)\boldsymbol{f}_j + \boldsymbol{A}_j(\boldsymbol{\alpha}_{jk} - \sum_{l=1}^{m}\boldsymbol{\alpha}_{jl}) & j = k\end{cases} \tag{38}$$

and use Eq. (36) to obtain $\boldsymbol{\tau}_j$. In this way, we can decompose the bias in each term $\hat{\boldsymbol{f}}_j$ into covariate specific contributions. Once we obtain $\boldsymbol{\alpha}_{jk}^*$, we can use Eq. (37) to get $\boldsymbol{\theta}_{jk}$ so that

$$\boldsymbol{\theta}_{jk} = \left[diag(\boldsymbol{X}_j)\right]^{-1}\boldsymbol{\alpha}_{jk} = \begin{pmatrix}\theta_{1jk}\\ \theta_{2jk}\\ \vdots\\ \theta_{njk}\end{pmatrix}_{nx1} \tag{39}$$

where $\boldsymbol{\theta}_{jk}$ is a column vector of covariate-specific contributions from the kth term $\hat{\boldsymbol{\beta}}_k$ to the bias in the jth term $\hat{\boldsymbol{\beta}}_j$, and the total bias in $\hat{\boldsymbol{\beta}}_j$ can be expressed by the sum of each covariate-specific

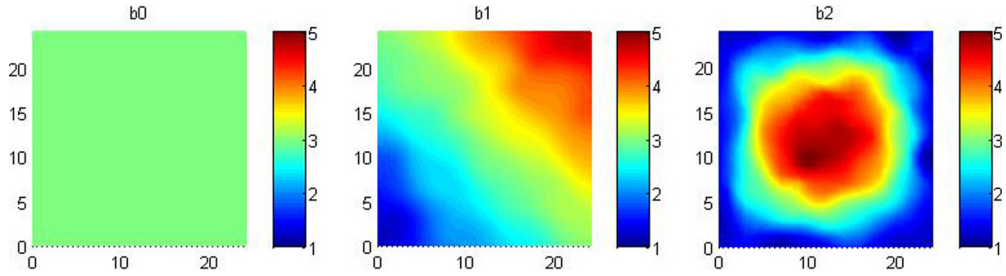


Fig. 10. MGWR parameter estimates from the model in Eq. (21).

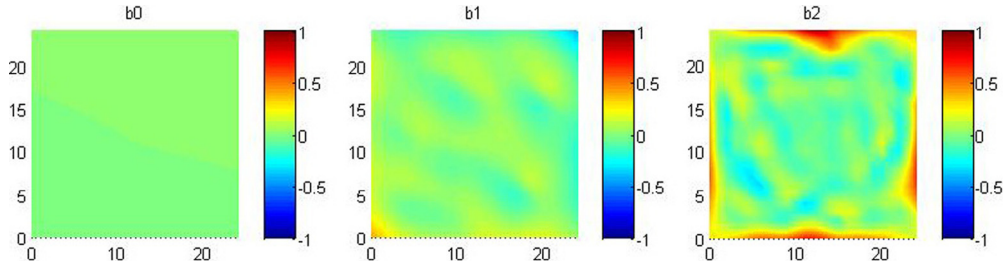


Fig. 11. Analytically-derived MGWR bias surfaces computed from Eq. (40).

contribution as

$$\boldsymbol{\gamma_j} = \boldsymbol{\beta_j} - \hat{\boldsymbol{\beta}}_j = \sum_{k=1}^{m} \boldsymbol{\theta_{jk}} \tag{40}$$

6. An example of data-borrowing bias in MGWR parameter estimates

In the following section, the same simulated dataset described above in the GWR discussion is used to examine the data borrowing-bias in MGWR and to compare the degree of parameter bias in GWR and MGWR. Using the same model as in Eq. (21) and the same data as described above for GWR, the parameter estimate surfaces from the MGWR calibration are shown in Fig. 10. Rather than obtaining a single bandwidth for all three sets of local parameters as in GWR, MGWR allows a covariate-specific bandwidth to be optimized. The surface $\boldsymbol{\beta_0}$ which is constant over space has an optimal bandwidth of 625; effectively suggesting this relationship is global. The surfaces for $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ have optimal bandwidths as 47 and 26, respectively, representing medium to high spatial heterogeneity.

The analytically-derived data-borrowing bias calculated from Eq. (40) is shown in Fig. 11 and the empirically derived bias is shown in Fig. 12. Both figures clearly demonstrate that the bias in the local estimates of β_0 has been reduced because in MGWR the covariate-specific bandwidth of 625 more accurately reflects the homogeneity of the parameter surface than does the single optimal bandwidth of 50 found in GWR.

Further evidence of the accuracy of the analytical expression for the bias in MGWR parameter estimates given in Eq. (40) is provided in Fig. 13 where the computed analytical bias is compared to the mean bias derived from 10,000 simulations of the set of y_i values described above for MGWR. For all three sets of local parameter estimates, the results are virtually identical implying that the analytical equation can be used with confidence to derive the bias in MGWR-derived local parameter estimates.

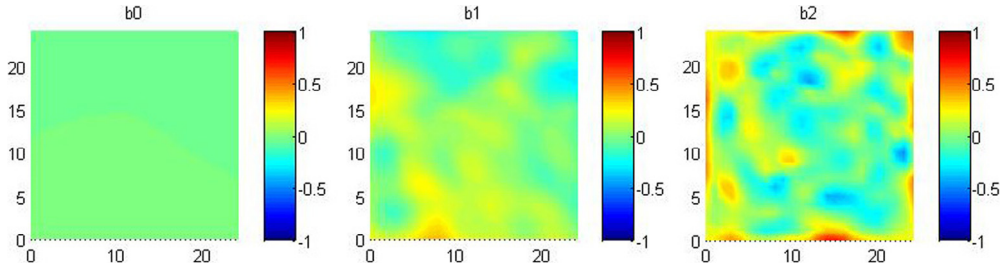


Fig. 12. Empirically-derived bias in MGWR parameter estimates.

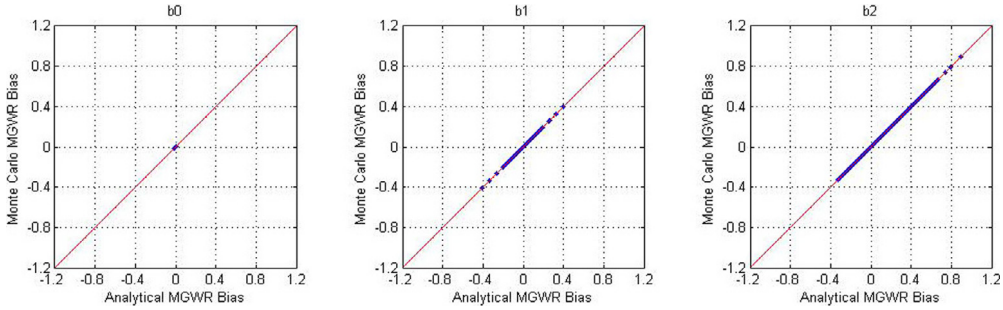


Fig. 13. Comparison of analytically-derived MGWR bias from Eq. (34) and the Monte Carlo simulated MGWR bias based on average of 10,000 realizations.

7. A comparison of bias in GWR and MGWR parameter estimates

Given we now have explicit analytical functions with which to calculate the bias of local parameter estimates obtained from both GWR and MGWR, it is of interest to compare the degree of bias in the local parameter estimates from the two methods, as shown in Fig. 14. For the local estimates of the intercept, β_0, shown in Fig. 14a, 90% of the estimates from GWR contain more bias than the equivalent estimates from MGWR and even when the GWR bias is less than the MGWR equivalent, the values are extremely close to zero for both models and the difference is just random noise. Indeed all the estimates from MGWR contain virtually no bias at all as the covariate-specific bandwidth is 625 and the relationship being modeled in MGWR is a spatially stationary one so borrowing data from other locations introduces no bias — the process being modeled is the same everywhere. Conversely, the GWR estimates of the local intercept contain bias because the single optimized bandwidth in GWR is 50, so that the set of predicted local estimates contains some degree of spatial heterogeneity whereas in reality the estimates are the same everywhere.

For the estimates of β_1, depicted in Fig. 14b, the bias in the MGWR estimates again tends to be less than in the GWR estimates — 72% of the points lie in the two triangles to the left and right of the figure forming a 'bow-tie' shaped region in which the GWR bias is greater than the equivalent MGWR bias. In this case, the superiority of the MGWR estimates in terms of bias is not as great as that in the local intercepts because the covariate-specific bandwidth in MGWR for β_1 is very similar to the single bandwidth obtained in GWR (47 vs. 50). The bias is still generally lower in the MGWR estimates because the bias contribution from the estimates of β_0 and β_2 will be less.

For β_2 the situation is described in Fig. 14c and extends the trend described above for β_1. The MGWR estimates contain less bias for 74% of the parameter estimates compared to their GWR counterparts but the bias is greater on average than that for the estimates of β_1 which in turn is greater than that for the estimates of β_0. The trend in bias magnitude across the three sets of parameters reflects the degree of spatial heterogeneity in the true parameters. When the

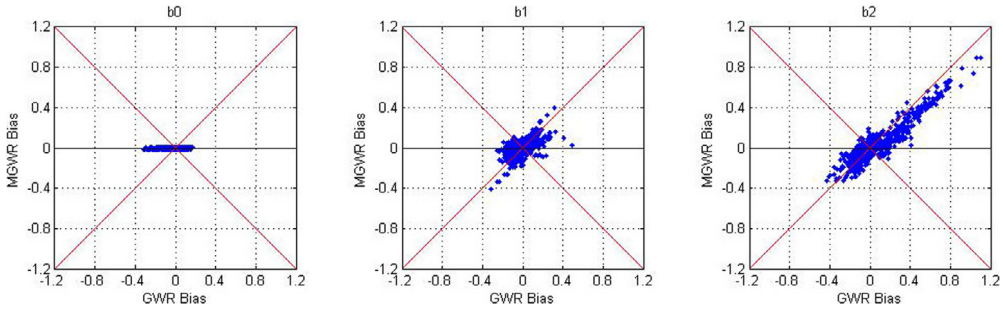


Fig. 14. Comparison of bias in MGWR and GWR parameter estimates. Points falling into the bow-tie shaped area are where the bias in the MGWR estimates is smaller than the bias in the GWR estimates.

parameters exhibit strong spatial heterogeneity, borrowing data will lead to greater bias than when the parameters exhibit relatively weak spatial heterogeneity.

The results shown in Fig. 14 can be decomposed to show the contributions to the results of bias in each of the three sets of local parameter estimates. This decomposition is shown in Fig. 15 where the sum across each of the three rows equates to the three graphs in Fig. 14. The bulk of each bias plot shown in Fig. 14 is a function of bias related to that parameter for each location (the diagonals of Fig. 15) but in some instances, there is a cross-bias from other parameters. This is most noticeable in the bias in the GWR-derived local estimates for β_1 which contain a bias from the biased estimates of β_2.

We can examine the spatial distribution of the bias for all three sets of local parameter estimates derived from both GWR and MGWR as shown in Fig. 16 in terms of a percentage bias for each local estimate which is simply the percentage error in estimating each local parameter. The difference in the two models is clearly seen in the spatial pattern of bias in the local intercepts: the MGWR estimates have virtually zero bias whereas the GWR estimates display quite large bias. The biases for the other two sets of local parameter estimates are somewhat similar with MGWR tending to produce lower bias. For b_1 the bias tends to be positive in the south-east and negative in the north-west which reflects the spatial pattern of the true values as shown in Fig. 10 where the south-west has lower-than-average values and the north-east has higher-than-average values. The effect of local data borrowing in the estimation of the parameters produces estimates which are less extreme and therefore the pattern of the bias reflects the pattern of the actual parameters. Similarly for the local estimates of b_2 the spatial pattern of the bias is circular with positive bias towards the periphery where the true parameters are relatively low and an inner ring of negative bias where the true parameters are relatively high. However, it is interesting that the bias tends to zero for the parameters having the highest values in the center of the area where there is presumably less mixing of disparate values in the data borrowed to estimate the local parameters.

Finally, we can compare the standard errors of the local estimates of β_0, β_1, and $\beta_2$derived from GWR and MGWR, as shown in Fig. 17 for all 625 locations. Where points lie to the right of the diagonal, the GWR estimates have greater uncertainty than their MGWR counterparts and where points lie to the left of the line, the uncertainty of the MGWR estimates is greater. It is immediately evident that whereas the uncertainty associated with all 625 local estimates of both β_0 and β_1 is greater when the estimates are obtained via GWR, the reverse is the case for the estimates of β_2. In the case of β_0, the covariate-specific bandwidth obtained in MGWR is 625 compared to the single bandwidth in GWR of 50. This means that in the local regressions for β_0 under MGWR there are more data points used than in the GWR model and so the uncertainty about each estimate is lower. In the case of β_2, the reverse is the case: the covariate-specific bandwidth in MGWR is 26 compared to the single bandwidth of 50 in GWR so that more data points are used in the GWR local regressions and hence the uncertainty in the resulting parameter estimates is lower in GWR than in MGWR. Note that although the precision of the estimates of β_2 might be greater in GWR, the estimates

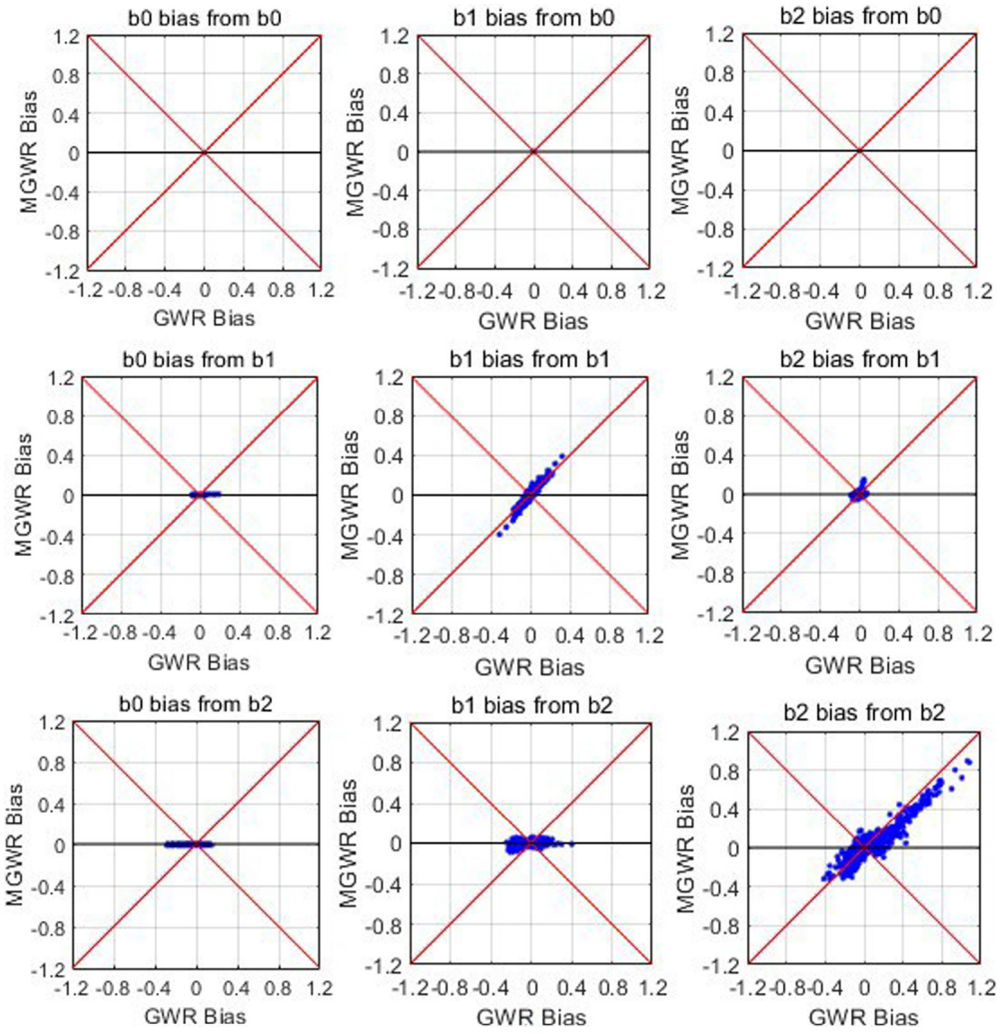


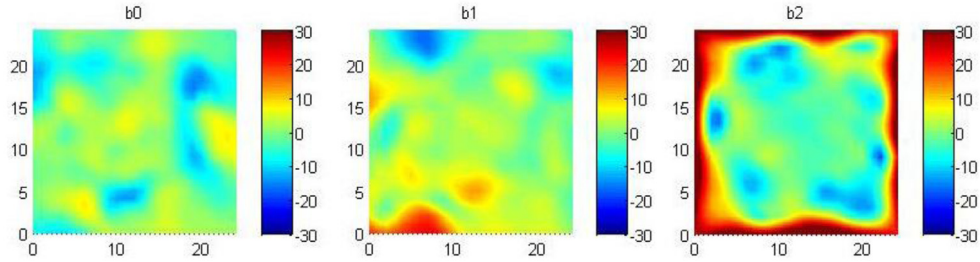
Fig. 15. Comparison of decomposed covariate-specific bias in MGWR (y-axis) and GWR (x-axis).

themselves contain more bias. For β_1 the situation is more complex. The single bandwidth obtained in GWR (50) is very close to that of the covariate-specific bandwidth for β_1 obtained in MGWR yet the standard errors for the GWR local estimates of β_1 are all larger than the corresponding values obtained through MGWR. The extra uncertainty in GWR estimates of the local β_1 parameters must therefore result from the extra uncertainty in the estimates of the local intercept in GWR.

8. Conclusions

GWR and its recent successor, MGWR, allow the estimation of local parameters by borrowing data from nearby locations and weighting these data according to how proximal they are to the location for which the local parameters are estimated. If the processes being estimated vary over spatial, borrowing data from nearby locations will introduce bias into the local parameter estimates. This is well-known and indeed the general method of finding an optimal bandwidth (degree of

(a) GWR



(b) MGWR

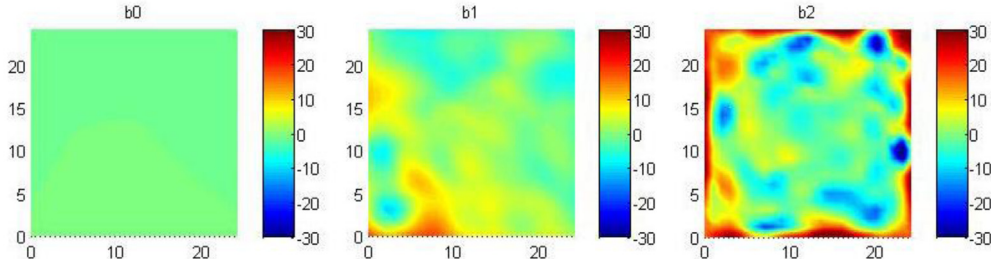


Fig. 16. Percentage of bias in each local estimate from (a) GWR and (b) MGWR.

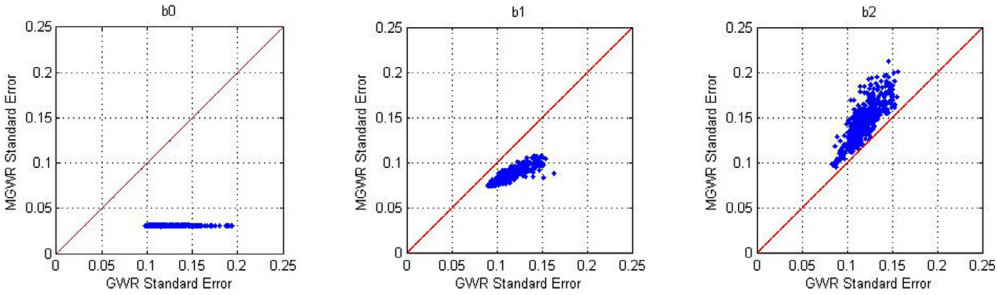


Fig. 17. Comparison of standard errors in MGWR parameter estimates with standard errors in GWR parameter estimates.

distance-decay in the weighting function) in both GWR and MGWR is to minimize a statistic that is generally held to be a trade-off between bias and variance. Choosing too small a bandwidth leads to a small bias but large variance; choosing too large a bandwidth leads to a low variance but large bias. Until now, however, it has not been possible to compute bias directly. This paper provides the analytical expressions for bias in local parameter estimates in both a GWR and an MGWR framework. The expressions are supported by an analysis of a simulated dataset with known local parameter surfaces so that an experimental bias can be calculated for each local parameter estimate and compared to the equivalent values derived from the analytical expressions. This simulated example demonstrates the viability of the analytical expressions.

The ability to be able to compute the data-borrowing bias in each local parameter estimate is important for several reasons. Firstly, it is useful to be able to measure the extent of the bias contained within each parameter estimate and here it is shown that this bias is relatively small.

Secondly, it is useful to quantify the degree to which bias in GWR local parameter estimates is mitigated by MGWR which allows covariate-specific bandwidths to be optimized. We demonstrate that generally the bias in MGWR-derived local parameter estimates is lower than that of the GWR counterparts. Thirdly, it is useful to be able to examine how sensitive the bias in local parameter estimates is to the optimized bandwidth and then to examine the classic bias–variance trade-off in the derivation of the optimal bandwidth. It would appear from the limited evidence presented here that the corrected AIC statistic is a reasonable criterion to minimize in order to find an optimal bandwidth that produces a good trade-off between bias and variance in local parameter estimates.

References

Brown, S., Versace, V.L., Laurenson, L., Ierodiaconou, D., Fawcett, J., Salzman, S., 2012. Assessment of spatiotemporal varying relationships between rainfall, land cover and surface water area using geographically weighted regression. Environ. Model. Assess. 17 (3), 241–254.

Buja, A., Hastie, T., Tibshirani, R., 1989. Linear smoothers and additive models. Ann. Statist. 45, 3–510.

Cahill, M., Mulligan, G., 2007. Using geographically weighted regression to explore local crime patterns. Soc. Sci. Comput. Rev. 25 (2), 174–193.

Fotheringham, A.S., Brundson, C., Charlton, M., 2002. Geographically weighted regression: the analysis of spatially varying relationships. Wiley, Chichester, West Sussex, England.

Fotheringham, A.S., Crespo, R., Yao, J., 2015. Geographical and temporal weighted regression (GTWR). Geogr. Anal. 47 (4), 431–452.

Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale geographically weighted regression (MGWR). Ann. Amer. Assoc. Geogr. 107 (6), 1247–1265.

Gelfand, A.E., Kim, H.J., Sirmans, C.F., Banerjee, S., 2003. Spatial modeling with spatially varying coefficient processes. J. Amer. Statist. Assoc. 98 (462), 387–396.

Griffith, D.A., 2008. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). Environ. Plan. A 40 (11), 2751–2769.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall, New York.

Miller, J.A., Hanham, R.Q., 2011. Spatial nonstationarity and the scale of species–environment relationships in the Mojave Desert, California, USA. Int. J. Geogr. Inf. Sci. 25 (3), 423–438.

Yu, H., Fotheringham, S., Li, Z., Oshan, T., Kang, W., Wolf, L.J., 2018. Inference in multiscale geographically weighted regression. http://dx.doi.org/10.31219/osf.io/4dksb.

Zou, B., Pu, Q., Bilal, M., Weng, Q., Zhai, L., Nichol, J.E., 2016. High-resolution satellite mapping of fine particulates based on geographically weighted regression. IEEE Geosci. Remote Sens. Lett. 13 (4), 495–499.