Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: https://www.tandfonline.com/loi/raag21

Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights

Ziqi Li, A. Stewart Fotheringham, Taylor M. Oshan & Levi John Wolf

Check for updates

Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights

Ziqi Li,[*] A. Stewart Fotheringham,[*] Taylor M. Oshan,[†] and Levi John Wolf[‡]

[*]*Spatial Analysis Research Center, School of Geographical Sciences and Urban Planning, Arizona State University*
[†]*Center for Geospatial Information Science, Department of Geographical Sciences, University of Maryland*
[‡]*School of Geographical Sciences, University of Bristol*

Bandwidth, a key parameter in geographically weighted regression models, is closely related to the spatial scale at which the underlying spatially heterogeneous processes being examined take place. Generally, a single optimal bandwidth (geographically weighted regression) or a set of covariate-specific optimal bandwidths (multiscale geographically weighted regression) is chosen based on some criterion, such as the Akaike information criterion (AIC), and then parameter estimation and inference are conditional on the choice of this bandwidth. In this article, we find that bandwidth selection is subject to uncertainty in both single-scale and multiscale geographically weighted regression models and demonstrate that this uncertainty can be measured and accounted for. Based on simulation studies and an empirical example of obesity rates in Phoenix, we show that bandwidth uncertainties can be quantitatively measured by Akaike weights and confidence intervals for bandwidths can be obtained. Understanding bandwidth uncertainty offers important insights about the scales over which different processes operate, especially when comparing covariate-specific bandwidths. Additionally, unconditional parameter estimates can be computed based on Akaike weights accounts for bandwidth selection uncertainty. *Key Words: Akaike weight, bandwidth, model selection uncertainty, multiscale geographically weighted regression, spatial processes scale.*

带宽是地理加权回归模型中的一个关键参数，此参数与所研究潜在空间异构过程中所发生的空间尺度密切相关。在此过程中，通常会根据某些准则（例如赤池信息准则（AIC））选择单一最佳带宽（地理加权回归）或一组根据特定于协变量的最佳带宽（多尺度地理加权回归），然后以该带宽选择为条件，进行参数估计和推断。本文的作者发现，带宽选择在单尺度和多尺度地理加权回归模型中均受到不确定性的影响。作者还证明了这种不确定性可以被测量和解释。基于凤凰城关于肥胖率的模拟研究和实证举例，作者表明可以通过赤池权重对带宽不确定性进行定量测量，可以获得带宽的置信区间。理解带宽不确定性为不同进程的运行尺度提供了重要见解，尤其是在比较特定于协变量的带宽时更是如此。另外，赤池权重所揭示的带宽选择的不确定性，还可以用于计算无条件参数估值。*关键词: 赤池权重，带宽，模型选择不确定性，多尺度地理加权回归，空间过程规模。*

La amplitud de banda, un parámetro clave en los modelos de regresión geográficamente ponderada, está estrechamente relacionada con la escala espacial en la cual ocurren los procesos subyacentes con heterogeneidad espacial, bajo escrutinio. En general, una amplitud de banda óptima individual (regresión geográficamente ponderada) o un conjunto óptimo de amplitudes de banda con covariaciones específicas (regresión geográficamente ponderada a multiescala) son escogidas a partir de un criterio determinado, tal como el criterio de información Akaike (AIC), y desde ahí el estimativo e inferencia del parámetro quedan condicionados por la escogencia de esta amplitud de banda. En este artículo, encontramos que la selección de amplitud de banda está sujeta a incertidumbre en los modelos de regresión geográficamente ponderada tanto a escala sencilla como a multiescala, y demostramos que esta incertidumbre puede medirse y explicarse. Con base en estudios de simulación y en un ejemplo empírico de tasas de obesidad en Phoenix, mostramos que las incertidumbres de amplitud de banda pueden medirse cuantitativamente con pesos Akaike, y se pueden derivar los intervalos de confianza para las amplitudes de banda. Entender la incertidumbre de amplitud de banda ofrece perspectivas importantes acerca de las escalas a que operan diferentes procesos,

Annals of the American Association of Geographers, 0(0) 2020, pp. 1–
Initial submission, August 2019; revised submission, October 2019; final acceptance, October 2019
Color versions of one or more figures in the article can be found online at www.tandfonline.com/raag
Published by Taylor & Francis, LLC.

especialmente cuando se comparan amplitudes de banda de covariación específica. Por otro lado, los cálculos incondicionales de parámetro pueden computarse tomando en cuenta los pesos Akaike para la incertidumbre en la selección de amplitud de banda. *Palabras clave: amplitud de banda, escala de procesos espaciales, incertidumbre en la selección del modelo, peso Akaike, regresión geográficamente ponderada a multiescala.*

Investigating spatial processes through associations between a response variable and a set of explanatory variables has been one of the most important and fertile research areas in geography and related fields. Spatial processes, however, have the intrinsic properties of potentially being both heterogeneous and operational over different spatial scales. Classic global models ignore both of these properties and return only stationary (single) parameter estimates and provide no information on spatial scale. Local models such as geographically weighted regression (GWR) can capture the heterogeneity of processes but inadequately incorporate multiscale properties of processes into the modeling. Their major limitation is the use of only a single kernel bandwidth across the set of covariates, which is the equivalent of assuming that the different processes being modeled all operate at the same scale (Fotheringham, Yang, and Kang 2017; Murakami et al. 2019). Such an assumption seems unrealistic in the real world. For example, the measured ambient temperature of a location is affected by the local built environment, regional weather patterns, and trends in global warming, all of which operate at different scales. A recent advancement to GWR termed multiscale GWR (MGWR) removes the single bandwidth assumption and allows the bandwidths for each covariate to vary (Fotheringham, Yang, and Kang 2017). This results in each parameter surface being allowed to have a different degree of spatial heterogeneity, reflecting variation across covariate-specific processes.

Comparisons of MGWR with other single-scale and multiscale spatially varying coefficient (SVC) models are available, such as the comparison with classic GWR (Fotheringham, Yang, and Kang 2017; Harris 2019; Murakami et al. 2019), with Bayesian SVC (Wolf et al. 2018), with eigenvector spatial filtering (ESF; Oshan and Fotheringham 2018), and with random-effects eigenvector spatial filtering (RE-ESF; Murakami et al. 2019). All comparisons agree that in terms of parameter estimation accuracies, MGWR is superior to the classic GWR and comparable to much more complicated models such as Bayesian SVC and RE-ESF. Additionally, an analytical inferential framework (Yu et al. 2020), computational improvements by parallelization (Li and Fotheringham 2019; Li et al. 2019), and accessible software (Oshan, Li, et al. 2019) have been developed for MGWR, all of which greatly enhance the utility of MGWR in modeling multiscale spatially heterogeneous processes.

Bandwidth is a key parameter in the GWR framework. The major advance of MGWR over GWR is that covariate-specific bandwidths are obtained rather than a single "average" bandwidth. Consequently, GWR can be considered as a special case of MGWR when all bandwidths are the same. Often the optimal bandwidth selection is data driven based on model selection statistics such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC), generalized cross-validation, or some other panelized fit score. The fundamental goal of choosing the optimal bandwidth is to find the best trade-off between bias and variance for parameter estimates. A large bandwidth produces local parameter estimates with low variance at the cost of high bias (underfitting), whereas a small bandwidth yields parameter estimates with low bias but high variance (overfitting). Once the optimal bandwidth is determined, it can be interpreted in terms of the spatial scale of the underlying data generating process. In other words, a larger bandwidth determined by the data indicates a more spatially smoothed process that has regional or no variability, whereas a smaller bandwidth reveals spatially localized relationships. Therefore, covariate-specific optimal bandwidths obtained in MGWR can be used as explicit indicators of the scale at which various processes operate (Fotheringham, Yang, and Kang 2017). The bandwidth also has an interpretable real-world meaning and can be either a distance-based measure or the number of nearest neighbors used in each local regression. The interpretation of a bandwidth of, for example, fifty nearest neighbors is that the process being estimated at the current location is affected by other neighboring locations in a spatially discounted manner up to a radius of fifty nearest neighbors. Data at the local regression point are given a weight of one and data borrowed from neighboring locations are given a discounted weight less than one depending on how far they are from the regression point.

The use of multiple bandwidths gives MGWR the capability to potentially differentiate local, regional, and global processes by comparing the optimal bandwidths for different covariates. If the analyst wants to make an inference about bandwidths as indicators of the relative spatial scales of different processes, however, it is naive to simply compare the covariate-specific bandwidths that are obtained based on the single observed data set (i.e., deterministic) and draw a conclusion that one process is more local or global than another. It is quite possible that the covariate-specific bandwidths are different by chance and subject to the sampling variation of the uncaptured noise (i.e., stochastic). Understanding the covariate-specific bandwidth uncertainty is thus crucial to being able to make inferences about the different spatial scales over which processes operate. Smoothing parameter (e.g., bandwidth, in the context of GWR) uncertainties have been recognized in the statistical literature for spline-based generalized additive models (Hastie, Tibshirani, and Friedman 2009; Wood, Pya, and Säfken 2016). Additionally, previous studies show that GWR and MGWR bandwidths are not fixed from the evidence of Monte Carlo simulations and subsampling results (Fotheringham, Yang, and Kang 2017; Wolf, Oshan, and Fotheringham 2018; Oshan, Wolf, et al. 2019). Nevertheless, there are no methods in the context of a GWR framework to quantitatively measure such bandwidth uncertainties. Another issue of neglecting bandwidth uncertainty is that parameter estimation and inference are conditional on the optimal bandwidths and therefore incorporation of bandwidth uncertainty into parameter uncertainty is important (Wolf, Oshan, and Fotheringham 2018). By doing so, parameter estimates would be unconditional of the chosen bandwidth, thus making inference more robust. Ignoring the uncertainties from bandwidth selection will lead to underestimates of the variances for local parameter estimates because the variance of the component stemming from bandwidth selection uncertainty is missing and, consequently, this will make it easier to declare local results as interesting or significant.

Overview of the Procedure for Measuring Bandwidth Uncertainty

Burnham and Anderson (1998) extensively discussed the uncertainty regarding model selection and proposed using the Akaike weight, an information theory–based statistic, to measure model selection uncertainties. This method has drawn great attention and is widely used in applied statistics (Johnson and Omland 2004; Posada and Buckley 2004; Koh 2008; Pinsky et al. 2013). Akaike weights are computed based on AIC and can be interpreted as the relative likelihood of a certain model being selected given the data. For instance, a model with an Akaike weight of 0.6 indicates that given the data at hand, it has a 60 percent chance of being selected as the optimal model among other candidate models. In the context of GWR, bandwidth selection is a type of model selection. Candidate bandwidths are evaluated based on AIC (or an equivalent measure) and the bandwidth with minimum AIC is selected as the optimal one. It is a natural extension, therefore, to use the Akaike weights to measure the relative likelihood of a bandwidth being selected as optimal and to quantify the bandwidth selection uncertainty. Moreover, Akaike weights can be used to average parameter estimates and variances following a multimodel inference framework with the advantage that the resulting parameters are unconditional of the selected model. This approach has been shown to be useful in the statistical literature and in applied studies (Burnham and Anderson 1998; Wagenmakers and Farrell 2004; Burnham, Anderson, and Huyvaert 2011; Symonds and Moussalli 2011).

In this article, we demonstrate that bandwidth has intrinsic uncertainty from the evidence of bootstrapping and from Akaike weights, which can both be used to obtain confidence intervals (CIs) for bandwidths in MGWR. We also examine the use of Akaike weights to compute unconditional parameter estimates and variances and compare them with their conditional equivalents. The article proceeds as follows. First, the background of MGWR is reviewed. Second, bandwidth uncertainties are examined with a simulation data set using bootstrapping and Akaike weights. Third, an empirical example of obesity rate modeling for the city of Phoenix is presented. Finally, the article concludes with remarks.

Background of MGWR Related to Akaike Weights

MGWR was developed under the generalized additive model framework of Hastie and Tibshirani

(1990) by Fotheringham, Yang, and Kang (2017). MGWR is formulated as

$$y = f_1(x_1) + f_2(x_2) + \ldots + f_k(x_k) + \varepsilon, \tag{1}$$

where y is a column vector of response variables, $f_1(x_1), \ldots, f_k(x_k)$ are additive components that are smooth functions of covariates, and ε is a column vector of independent and identically distributed error terms. The response variable y is the data observed over a spatial surface and $f_1, \ldots, f_k$ are spatial additive components estimated with covariate-specific bandwidths, so MGWR can be formulated as

$$y = f_{bw_1}(x_1) + f_{bw_2}(x_2) + \ldots + f_{bw_k}(x_k) + \varepsilon, \tag{2}$$

where each jth additive component $f_{bw_j}(x_j)$ is a product of element-wise multiplication ($\circ$) of local parameters b_j and covariate x_j:

$$f_{bw_j}(x_j) = b_j \circ x_j = \begin{pmatrix} b_{1j}x_{1j} \\ b_{2j}x_{2j} \\ \vdots \\ b_{nj}x_{nj} \end{pmatrix}. \tag{3}$$

The calibration of each smoothing function f_{bw_j} in the MGWR model uses the back-fitting algorithm developed by Buja, Hastie, and Tibshirani (1989). The estimation procedure can be initialized in various ways (e.g., using GWR estimates; Fotheringham, Yang, and Kang 2017) and then parameters are estimated and updated by calibrating univariate GWR models that regress the current estimated additive component plus partial residual on each covariate ($\hat{f}_j + \hat{\varepsilon} \sim x_j$) successively. The back-fitting converges when parameter estimates are unchanging within a predefined threshold. A detailed description and implementation of the back-fitting algorithm can be found in Fotheringham, Yang, and Kang (2017) and Oshan, Li, et al. (2019). Covariate-specific optimal bandwidths are estimated within univariate GWRs by minimizing the corrected AIC ($\mathrm{AIC_c}$), which is formulated as

$$AIC_c = -2\log(\ell) + 2n\left(\frac{tr(\mathbf{S}) + 1}{n - tr(\mathbf{S}) - 2}\right), \tag{4}$$

where ℓ is the model likelihood given the data and $tr(S)$ is the trace of the hat matrix. $\mathrm{AIC_c}$ is a small sample bias adjustment to the classic AIC and should be used when the ratio of data points to the number of parameters is below forty, which is frequently the case in GWR if processes have high heterogeneity. When the ratio is greater than forty, the $\mathrm{AIC_c}$ rapidly approaches the classic AIC (Hurvich and Tsai 1993; Burnham and Anderson 1998). Therefore, in general, this corrected version of AIC is often suggested in GWR for bandwidth selection (Fotheringham, Brunsdon, and Charlton 2002). The use of AIC or its variant $\mathrm{AIC_c}$ for model selection is based on the information-theoretic relationship between expected Kullback–Leibler distance (information lost) and the maximized log-likelihood. The best model represents the process that generates the data with minimum information lost.

To obtain Akaike weights, we first need to define the "comparison set" of models. In the context of MGWR, of interest is the set of covariate-specific bandwidths that best approximate the underlying spatial processes. During the bandwidth selection of each univariate GWR within the back-fitting of MGWR, the optimal bandwidth is selected based on the minimum $\mathrm{AIC_c}$. For simplicity, we will use AIC here (rather than $\mathrm{AIC_c}$) as a generic term for all AIC variants. A candidate set of R bandwidths can be defined as $\{bw_1, bw_2, bw_3, \ldots, bw_R\}$ beforehand or within the bandwidth search routine. In either case, we consider a set of candidate bandwidths to evaluate AIC and the minimum AIC obtained within the R bandwidths is denoted as $AIC_{\min}$. For bandwidth k within the candidate set R, AIC differences can be computed as $\Delta_k = AIC_k - AIC_{\min}$. Then, the Akaike weight of a candidate bandwidth $k \in \{1, \ldots, R\}$ can be obtained by

$$w_k = \frac{\exp\left(-\frac{1}{2}\Delta_k\right)}{\sum_{r=1}^{R} \exp\left(-\frac{1}{2}\Delta_r\right)}. \tag{5}$$

The numerator $\exp(-\frac{1}{2}\Delta_k)$ denotes the likelihood of the bandwidth k given the maximum likelihood estimators based on the same data, which also measures the relative strength of evidence for each bandwidth (Akaike 1983; Burnham and Anderson 1998). The denominator is used to normalize the Akaike weights so that all values lie between zero and one with the sum being one ($\sum w_k = 1$; Burnham and Anderson 1998). The resulting Akaike weight is the likelihood of a given bandwidth being the optimal one. For instance, an Akaike weight of 0.75 for a bandwidth indicates that this bandwidth has a 75 percent chance of being the bandwidth that best approximates the corresponding underlying process.

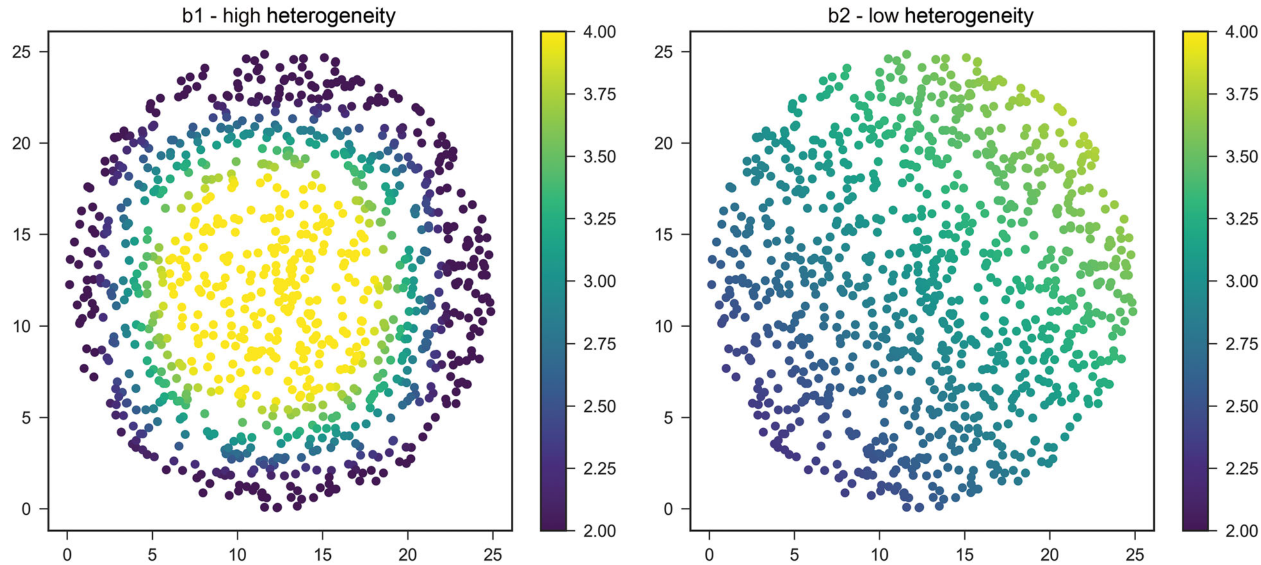

Figure 1. Generated spatial processes with low and high heterogeneity.

Following this approach, the Akaike weight curve can be plotted against a set of candidate bandwidths for each covariate in the MGWR model and the bandwidth probability distribution can be obtained. In the following section, we examine the use of Akaike weights for measuring bandwidth uncertainty based on a simulated data set.

Investigating Bandwidth Uncertainty in a Controlled Experiment

Construction of a Simulated Data Set with Two Different Spatially Heterogeneous Processes

We simulate a study area with 1,000 locations that are randomly distributed in a circular coordinate space. The x and y coordinates of location $i \in \{1, \dots n\}$ are denoted as u_i and v_i and are constructed with Equations 6 and 7:

$$u_i = 12.5 + 12.5\sqrt{r_i}\cos\theta_i \tag{6}$$

$$v_i = 12.5 + 12.5\sqrt{r_i}\sin\theta_i, \tag{7}$$

where a radius r_i and an angle θ_i are randomly selected from uniform distributions $U(0,\ 1)$ and $U(0,\ 2\pi)$, respectively. Then we synthesize two spatial processes using Equations 8 and 9; these true parameter surfaces, b_1 and b_2, are shown in Figure 1.

$$b_{1i} = 1 + 1/324*(36-(6-u_i/2)^2)*(36-(6-v_i/2)^2) \tag{8}$$

$$b_{2i} = 2 + 1/24*(u_i + v_i), \tag{9}$$

where b_{1i} and b_{2i} are the parameters of location $i \in \{1, \dots, n\}$, and u_i and v_i are the x and y coordinates of location i. The resulting process b_1 has high spatial heterogeneity, with high values in the center of the map and low values at the periphery. The parameters range between 0 and 5. The process b_2 has relative low heterogeneity, with a positive trend from southeast to northwest ranging from 2.4 to 3.7. A similar data generating process can be seen in Fotheringham, Yang, and Kang (2017).

Covariate column vectors x_1 and x_2 are randomly drawn from a normal distribution with mean of 0 and variance of 1. A spatially random noise vector ε is added to the surface with mean of 0 and variance of 1. Those two column vectors of processes, b_1 and b_2, along with two covariates, x_1 and x_2, and a random noise are used to generate the synthetic response variable y in the following manner:

$$y = b_1 \circ x_1 + b_2 \circ x_2 + \varepsilon. \tag{10}$$

An MGWR model is calibrated on this synthetic data set using the *mgwr* Python package (Oshan, Li, et al. 2019). For consistency, in the following simulation and empirical studies we use an adaptive bi-square kernel (Fotheringham, Brunsdon, and Charlton 2002) where the bandwidth is interpreted as the number of nearest neighbors and the largest possible bandwidth is the total number of locations in the data set. For this model, the two covariate-

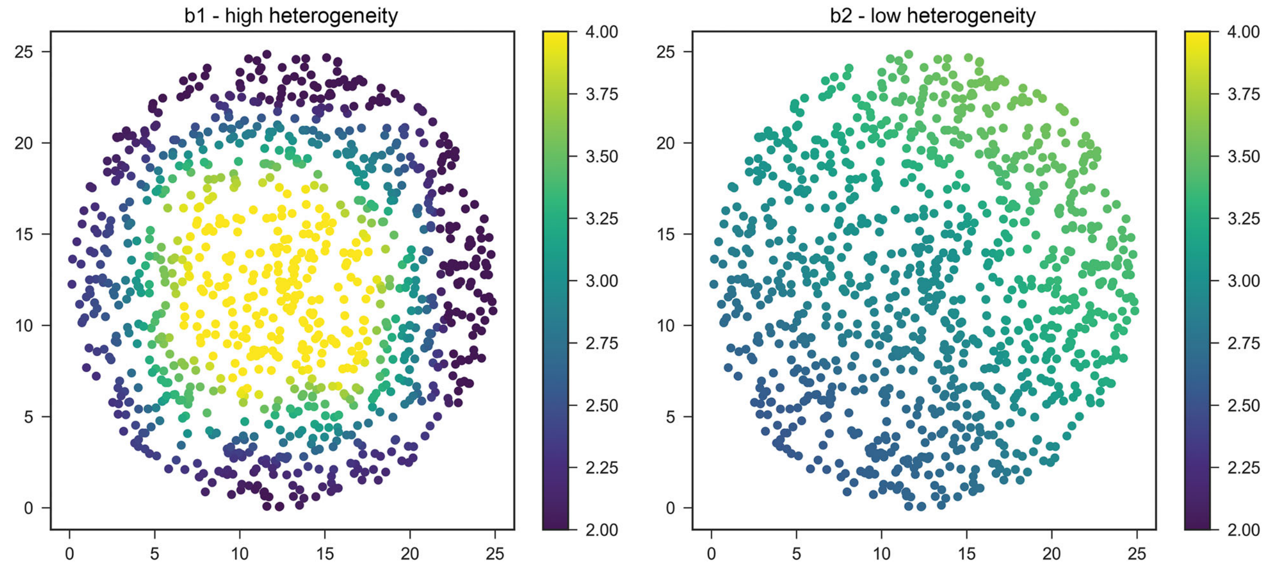

Figure 2. Recovered parameter surfaces using multiscale geographically weighted regression with optimal bandwidths 70 and 300 for the high- and low-heterogeneity processes, respectively.

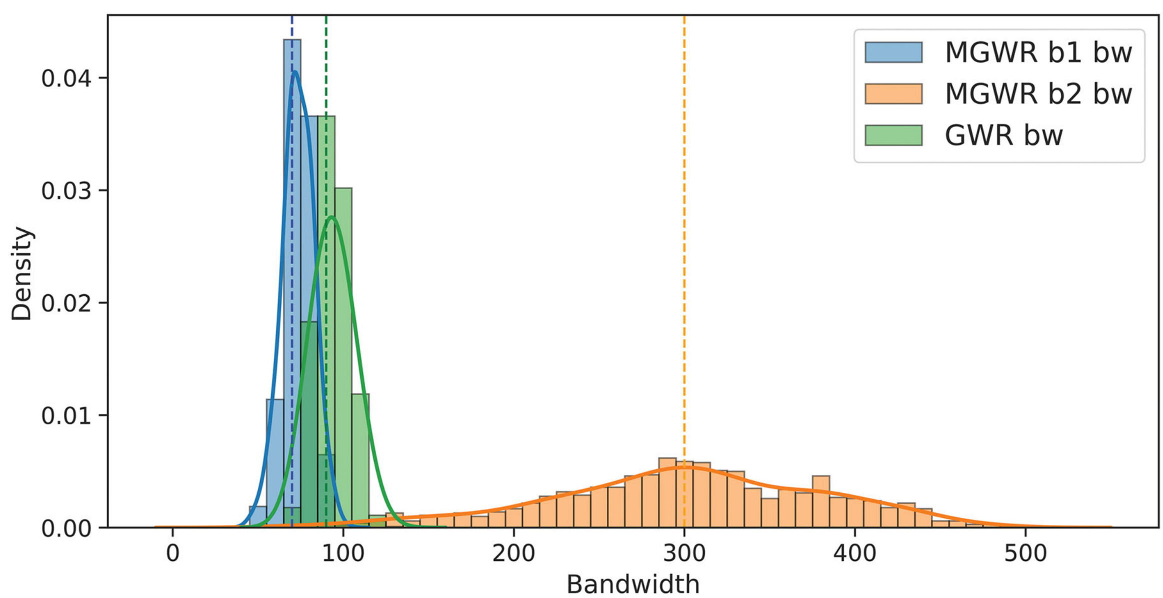

Figure 3. Optimal bandwidth distributions of two processes in Figure 1 under bootstrapping (smooth lines are plotted based on kernel density estimation). MGWR = multiscale geographically weighted regression; bw = bandwidth; GWR = geographically weighted regression.

specific optimal bandwidths are 70 and 300 for the local process b_1 and the regional process b_2, respectively. Recovered parameter estimate surfaces for this synthetic data set are shown in Figure 2.

Bandwidth Selection Uncertainty: Evidence from Bootstrapping

Bootstrapping is a variant of Monte Carlo estimation for CIs based on random sampling with replacement (Efron and Tibshirani 1994). The technique is popularly used for deriving empirical CIs of a statistic when the analytical solution is unavailable. In this article, we are interested in assessing how the selected optimal bandwidths vary across each bootstrap sample, from which the empirical distribution of optimal bandwidths can be obtained. We use a nonparametric residual-based bootstrap method by randomly resampling residuals that are added to the fitted values. This specific type of bootstrap method has been widely used in regression problems and can also be found in the GWR literature (Mei, Xu, and Wang 2016; Harris et al. 2017).

In the context presented here, we first calibrate an MGWR model with the original response vector y and two covariates x_1, x_2 and save the fitted values $\hat{y}$ and residuals $\hat{\varepsilon}$. Then we generate B ($B = 1{,}000$) bootstrap samples with each sample containing a new response vector $y^* = \hat{y} + \hat{\varepsilon}_{\text{rand}}$ where $\hat{\varepsilon}_{\text{rand}}$ is a residual vector randomly drawn with replacement from $\hat{\varepsilon}$. For each bootstrap sample, we regress y^* onto X using MGWR and save the optimal covariate-specific bandwidths from MGWR. Finally, the estimated bootstrap optimal bandwidth distribution can be acquired from these B bootstrap samples, and these bandwidth selection frequencies represent the uncertainty of a bandwidth j being selected as the optimal bandwidth. A 95 percent empirical CI can be obtained by using the 2.5th and 97.5th percentiles of the bandwidths from the bootstrap samples (Efron and Tibshirani 1994). This bootstrap method is applied to GWR for comparison with MGWR. The bootstrap results are presented in Figure 3 and Table 1.

Table 1. Summary statistics for the optimal bandwidths obtained from the bootstrap samples

	Optimal	*M*	*SD*	95% Confidence interval	Width
GWR BW	90	93.6	10.1	[80, 110]	30
MGWR b_1 BW	70	73.5	8.4	[60, 90]	30
b_2 BW	300	300.5	80.1	[130, 440]	310

Notes: GWR = geographically weighted regression; BW = bandwidth; MGWR = multiscale geographically weighted regression.

Figure 3 shows the distribution of the optimal bandwidths from the bootstrap samples for GWR (in green) and MGWR (in blue and orange). The optimal bandwidth selected for the original data set is 90 (number of nearest neighbors) for GWR and the optimal covariate-specific bandwidths are 70 and 300 for MGWR. It is clear that the MGWR covariate-specific bandwidths adequately describe the relative amounts of heterogeneity in the underlying data generating processes b_1 and b_2. In contrast, the single optimal bandwidth in GWR lies between the two covariate-specific bandwidths in MGWR and does not differentiate between the two processes or represent either particularly accurately. Bandwidth statistics are summarized in Table 1. For GWR, the single optimal bandwidths have a mean of 93.6 and standard deviation of 10.1 across the 1,000 bootstrap samples. The empirical 95 percent CI is [80, 110]. For MGWR, the optimal bandwidth for the high-heterogeneity process b_1 has a mean of 73.5 and standard deviation of 8.4, and the empirical 95 percent CI is [60, 90]. The optimal bandwidth for the low-heterogeneity process b_2 has a mean of 300.5, standard deviation of 80.1, and empirical CI of [130, 440]. By comparing the two covariate-specific bandwidths in

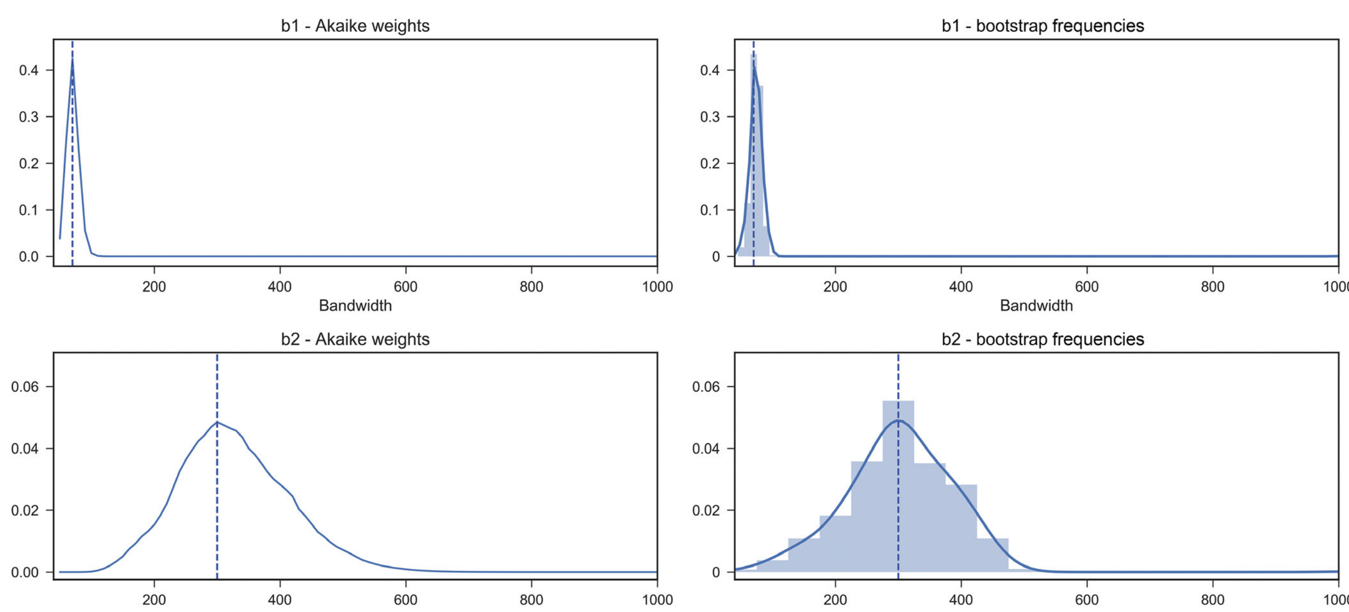

Figure 4. Akaike weights and bootstrap frequencies for different bandwidths (smooth lines are plotted based on kernel density estimation).

Table 2. An example of obtaining the 95 percent confidence set of bandwidth for process b_1 using Akaike weights

Bandwidth	AIC_c	Akaike weight	Cumulative Akaike weight
70	2,907.87	0.42	0.42
60	2,908.94	0.25	0.67
80	2,909.13	0.23	0.90
90	2,911.98	0.05	0.95
50	2,912.69	0.04	0.99
...	...	...	...

Notes: AIC_c = corrected Akaike information criterion.

MGWR, we can see that the selection of the bandwidth for the low-heterogeneity process b_2 tends to be more uncertain than that of the high-heterogeneity process b_1 and has a much wider CI (310 vs. 30).

Given the evidence from bootstrapping, it is clear that bandwidth selection in both GWR and MGWR is subject to the sampling variation of the random noise contained in the data. Bootstrapping provides a useful way for quantifying this bandwidth selection uncertainty and for deriving empirical CIs for bandwidths. The mutually exclusive CIs for the bandwidths associated with processes b_1 and b_2 imply that the two processes have significantly different optimal bandwidths at the 95 percent confidence level; we can thus claim that process b_1 operates at a relatively more local scale than process b_2. Following the same bootstrapping method described previously, we conducted another three simulations: (1) using the same model described in Equation 10 but calibrated with a fixed Gaussian kernel instead of an adaptive bi-square kernel (see Appendix A); (2) adding a spatially varying intercept to the model described in Equation 10 (see Appendix B); and (3) changing the orientation of the process b_2 holding the level of heterogeneity constant (see Appendix C). Results all indicate that the optimal bandwidths in MGWR are subject to the sampling variation of the noise, and bandwidth variation is a function solely of the scale over which the processes operate.

Using Akaike Weights for Measuring Bandwidth Selection Uncertainty

Following Equation 5, we compute Akaike weights for the simulation model described in Equation 10, which involves two processes with different degrees of spatial heterogeneity. For each covariate, we evaluate candidate bandwidths from 50 to 1,000 nearest neighbors with a ten-neighbor interval.[1] Each candidate bandwidth has a corresponding Akaike weight representing the probability of it being the optimal bandwidth (Figure 4). These Akaike weights can be compared with the bootstrap relative frequencies described in the previous section. There is a clear similarity between the two approaches for the processes shown in Figure 4. Following Burnham and Anderson (1998) and Symonds and Moussalli (2011), a 95 percent CI of bandwidths can be obtained by ranking the Akaike weights in descending order and including bandwidths in the CI until the cumulative Akaike weight equals 0.95. In some cases, where a coarse searching interval is used, the cumulative Akaike weight might not be exactly 0.95; therefore, the inclusion in the bandwidth CI should stop when the cumulative Akaike weight is just above 0.95. Table 2 shows how this procedure operates for process b_1. The optimal bandwidth of 70 has an Akaike weight of 0.42. Bandwidths of 60, 80, and 90 have descending Akaike weights of 0.25, 0.23, and 0.05, respectively. Inclusion in the 95 percent CI stops after the addition of the bandwidth 90 because at this point the cumulative Akaike weight equals 0.95. Consequently, we can state that the 95 percent CI of bandwidths for process b_1 is [60, 90]. The bandwidth CI is not necessarily symmetrical, and if bandwidth 50 is included this would create a 99 percent CI.

Table 3. Ninety-five percent CI of bandwidth based on Akaike weight and bootstrapping

Process	Optimal bandwidth	95% Akaike weight CI	95% Bootstrap CI
b_1	70	[60, 90]	[60, 90]
b_2	300	[160, 500]	[130, 440]

Notes: CI = confidence interval.

The Akaike weight–based 95 percent bandwidth CIs can be compared with the 95 percent CIs obtained from bootstrapping (Table 3). The 95 percent CI computed based on Akaike weights is [60, 90] for process b_1, which is exactly the same as the 95 percent CI obtained from bootstrapping. For process b_2, the Akaike weight–based 95 percent CI is [160, 500], which is similar to the 95 percent CI obtained from bootstrapping [130, 440]. Based on the Akaike weight–based bandwidth CI, we also arrive at the same conclusion that processes b_1 and b_2 have significantly different optimal bandwidths at

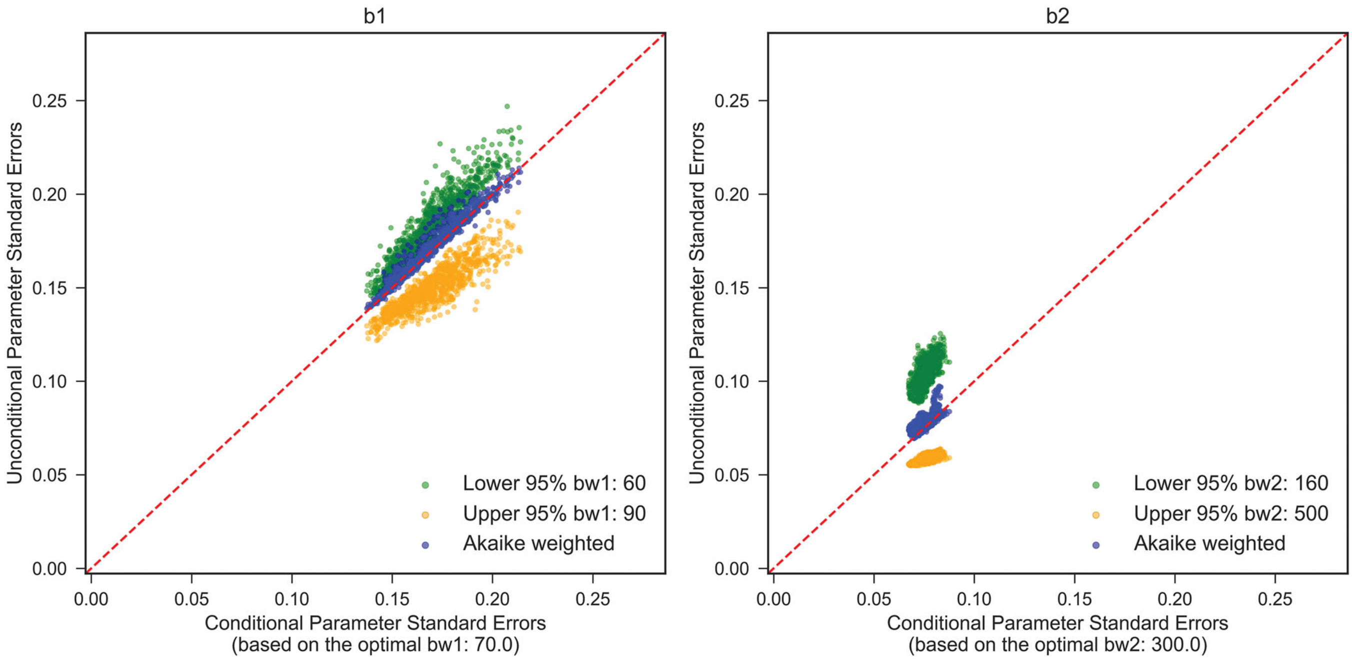

Figure 5. Comparison of conditional and unconditional standard errors of local parameter estimates for processes b_1 and b_2. bw = bandwidth.

the 95 percent confidence level. This result is particularly robust because Akaike weights and bootstrapping embed two distinct concepts of measuring model selection uncertainty. The former uses a data-based weight of evidence, whereas the latter models the sampling distribution of the bandwidth parameter when each bandwidth is still estimated without uncertainty. Using the Akaike weights is preferred because not only is it much less computer intensive but it also employs the statistically grounded concept of model likelihood given a set of candidate models and data (Burnham and Anderson 1998).

Unconditional Inference: Accounting for Bandwidth Selection Uncertainty in Local Parameter Estimation

Following the multimodel inference approach of Burnham and Anderson (1998), averaging models based on a spectrum of bandwidths can give parameters that are unconditional on the choice of the bandwidths. In this section, we compare the inference based on unconditional local parameter estimates with that based on the MGWR results, which are conditional on the optimized selected bandwidth. For each set of local parameters for the jth covariate, b_j, unconditional parameter estimates $\hat{\bar{b}}_j$ can be computed based on Akaike weights:

$$\hat{\bar{b}}_j = \sum\nolimits_{r=1}^{R} w_{rj}\hat{b}_{rj}, \tag{11}$$

where there are R candidate bandwidths being evaluated during the bandwidth selection; w_{rj} is the Akaike weight for bandwidth r; and $\hat{b}_{rj}$ is the set of local parameter estimates obtained when using bandwidth r. The variance of the parameter estimates in column vector b_j can be calculated using the following equation (Burnham and Anderson 1998):

$$\hat{v}ar\left(\hat{\bar{b}}_j\right) = \left[\sum w_{rj}\sqrt{\hat{v}ar\left(\hat{b}_{rj}\right) + \left(\hat{b}_{rj} - \hat{\bar{b}}_j\right)^2}\right]^2. \tag{12}$$

Note that there are two components in $\hat{v}ar\left(\hat{\bar{b}}_j\right)$: (1) the variance from parameter estimation, $\hat{v}ar\left(\hat{b}_{rj}\right)$, and (2) the variance from bandwidth selection uncertainty, $\left(\hat{b}_{rj} - \hat{\bar{b}}_j\right)^2$. Therefore, usually the variance of the unconditional parameter estimates will be greater than that of the conditional parameter estimates because the latter is based purely on a single set of covariate-specific bandwidths (i.e., deterministic) and therefore neglects bandwidth selection uncertainty (Symonds and Moussalli 2011). It is worth noting, though, that this might not always be true in (M)GWR because using a smaller bandwidth will yield larger parameter variances than

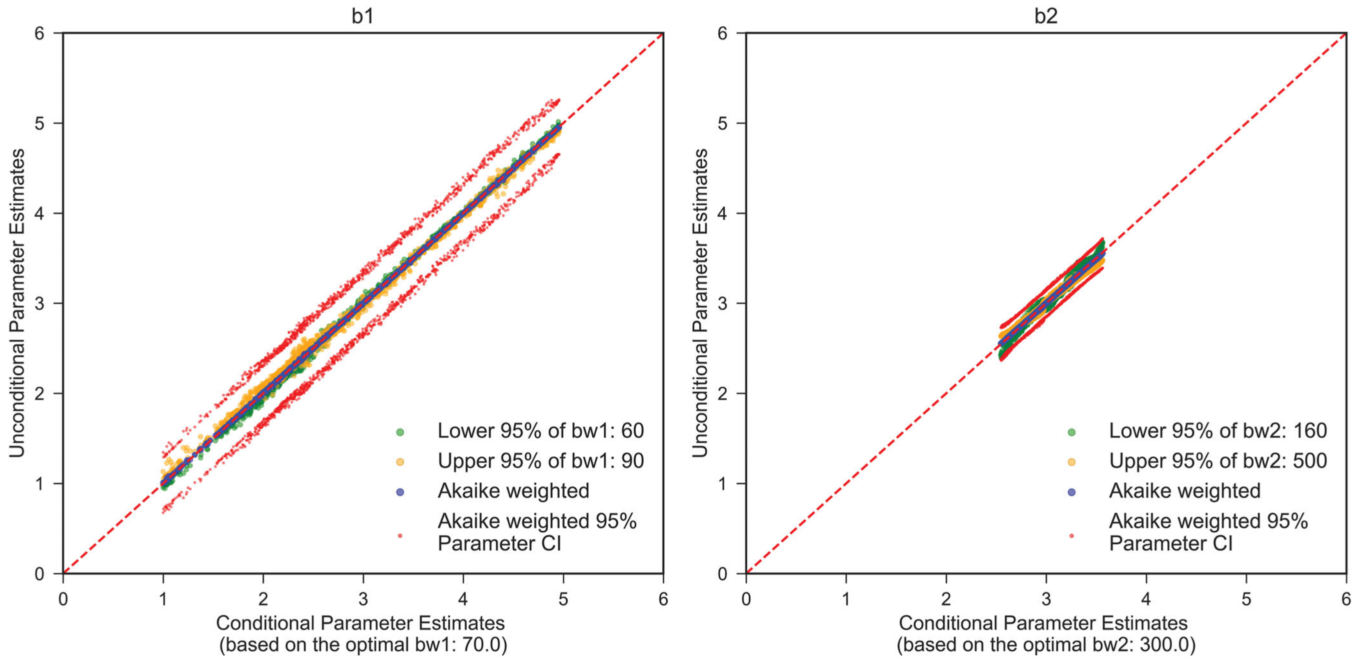

Figure 6. Comparison of conditional and unconditional local parameter estimates for processes b_1 and b_2. bw = bandwidth.

using a larger bandwidth, ceteris paribus. The weighted parameter variance is expected to yield more accurate CIs for the local parameter estimates, however, because bandwidth uncertainty is accounted for. Additionally, the averaged parameter estimates have a Bayesian interpretation if the Akaike weights are specified as prior probabilities on the bandwidths (Akaike 1979; Bozdogan 1987).

In Figure 5 we compare the uncertainties associated with the conditional parameter estimates obtained from MGWR with covariate-specific optimal bandwidths with their unconditional equivalents obtained using Akaike weights for the two processes b_1 and b_2. For both processes we plot three sets of unconditional local parameter standard errors (on the y axis) against their equivalent conditional standard errors. These are obtained by (1) using the lower bound of the 95 percent bandwidth CI (in green); (2) using the upper bound of the 95 percent bandwidth CI (in orange); and (3) using the Akaike-weighted standard errors (in blue). For process b_1 the upper and lower bounds of the 95 percent bandwidth CI are 60 and 90, respectively. Using a bandwidth of 60 (90) for the calibration gives higher (lower) parameter standard errors than those obtained using the optimal bandwidth of 70. This is due to the bandwidth's bias–variance trade-off property: Using a smaller bandwidth results in greater parameter uncertainty but reduced bias. Akaike-weighted parameter standard errors are computed based on bandwidths from 60 to 90 (95 percent bandwidth CI) with weights derived from Equation 12 and, as expected, these generally have slightly larger uncertainty than the conditional parameter estimates because they include bandwidth uncertainty. For process b_2 the results are similar to those for b_1, although the range of both the conditional and unconditional standard errors is much lower. This is because the bandwidth CI lies between 160 and 490 nearest neighbors for process b_2, whereas it is between 60 and 90 nearest neighbors for b_1, creating more variability in the results. When the optimal bandwidth is large and the level of process heterogeneity is low, the results are relatively insensitive to bandwidth variation within the 95 percent CI.

This is also seen in Figure 6, where the results of the unconditional and conditional local estimates of b_1 and b_2 are shown and where the horizontal axis depicts the conditional local parameter estimates from MGWR and the vertical axis depicts the unconditional local parameter estimates using the Akaike weights procedure described earlier. We also represent the Akaike-weighted 95 percent CI (red) and the upper (green) and lower (orange) 95 percent CI using the bandwidth CI based on the cumulative Akaike weights. The Akaike-weighted parameter estimate 95 percent CI is computed by $\hat{\bar{b}}_j \pm 1.96\sqrt{\hat{v}ar\left(\hat{\bar{b}}_j\right)}$. All of the local parameter

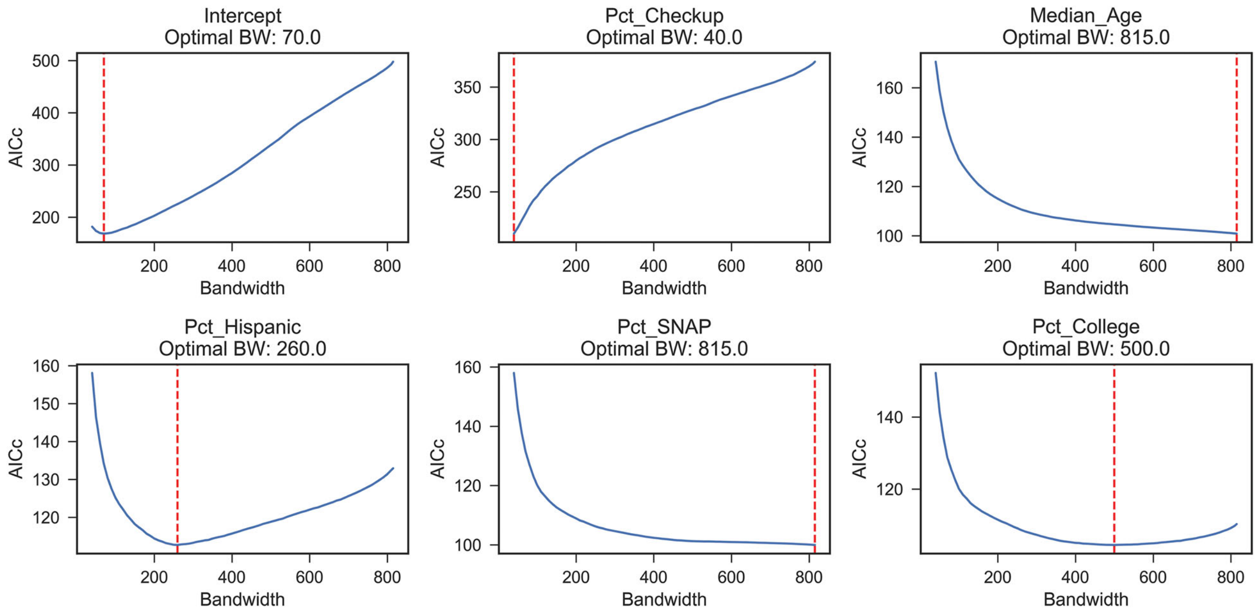

Figure 7. Corrected AIC_c values against bandwidth for each of the five covariates and the intercept. The red line is the optimal covariate-specific bandwidth based on the minimum AIC_c value. AIC_c = corrected Akaike information criterion; BW = bandwidth.

estimates using the lower and upper bounds of the bandwidth CI are within the Akaike-weighted parameter CI, indicating that using any bandwidth within the 95 percent bandwidth CI will produce parameter estimates within the Akaike-weighted parameter CI. This helps in interpreting how bandwidth change will affect the local parameter estimates; that is, we can now determine what degree of bandwidth change will generate significantly different local parameter estimates. Our results indicate that using any bandwidth within the 95 percent CI of the optimal bandwidth does not have a significant impact on the local parameter estimates.

To examine whether unconditional local parameter estimates, which account for bandwidth uncertainty, are more accurate than conditional parameters, which do not, we conduct a Monte Carlo simulation using the model described in Equation 9. The model was run for 1,000 realizations using fixed covariates and randomly drawn errors. Within each realization we compute conditional and unconditional parameter estimates and CIs and count how many times each type of CI contains the true parameters b_1 and b_2. A good 95 percent CI should contain true parameters 95 percent of the time. Results indicate that the Akaike-weighted parameter estimate CIs have marginally better coverage probabilities than the conditional CIs (0.81 vs. 0.80 for b_1) and (0.89 vs. 0.87 for b_2), although 94.1 percent and 97.2 percent of total locations, respectively, have increased CI coverage probabilities of more than 1 percent. Wolf, Oshan, and Fotheringham (2018) compared MGWR parameter standard errors with those obtained by a Bayesian spatially varying coefficients model and found that the MGWR parameter CIs were much smaller than the Bayesian counterparts, potentially because of neglecting bandwidth uncertainty in MGWR. In this study, however, we show that even after taking bandwidth uncertainty into account, parameter CI coverage probability is still less than the nominal 95 percent level, which reveals that there might exist other sources of parameter estimate uncertainty that are neglected.

An Empirical Example of Obesity Modeling in Phoenix

A real-world study is used to illustrate the use of Akaike weights in quantifying bandwidth uncertainty through an MGWR analysis of obesity rate determinants. The example uses the percentage of adults (aged eighteen and older) defined as obese (body mass index $\geq 30.0\,\mathrm{kg/m^2}$) by the Centers for Disease Control and Prevention in each of the 815 census tracts in the Phoenix metropolitan area[2] as the response variable and the five most influential covariates that determine obesity rates identified by the study of Oshan, Smith, and Fotheringham

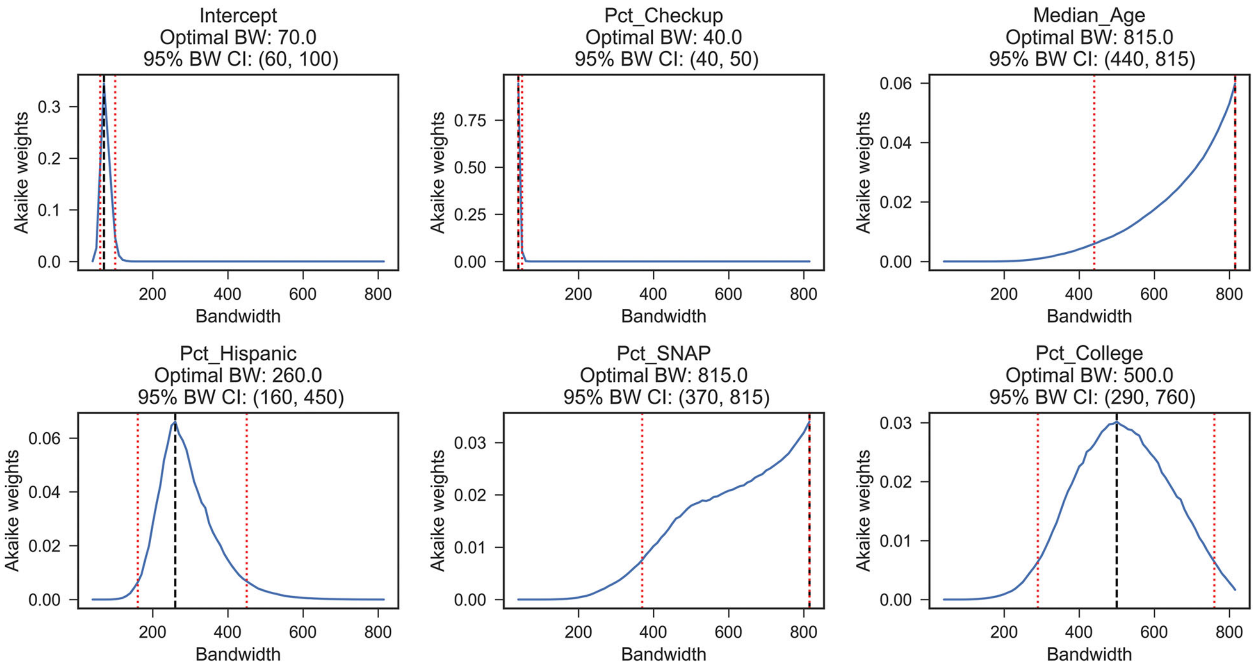

Figure 8. Akaike weights computed for different bandwidths for each set of local parameter estimates. The dashed line is the optimal bandwidth found by MGWR. The dotted lines depict the lower and upper bounds of the 95 percent confidence interval for each optimal bandwidth. MGWR = multiscale geographically weighted regression; BW = bandwidth; CI = confidence interval.

(2019): percentage of population visiting a doctor for a routine checkup within the past year, median age of people living in the census tract, percentage of population of Hispanic origin, percentage of households receiving Supplemental Nutrition Assistance Program (SNAP) benefits, and percentage of people with a college degree. The obesity and annual checkup data were downloaded from the 2014 500 Cities Project from the Centers for Disease Control and Prevention (2016), and the sociodemographic covariates were retrieved from the American Community Survey 2015 five-year estimates data set (U.S. Census Bureau 2015). The regression model is formulated as follows:

$$
\begin{aligned}
Pct\_Obesity_i = {} & b_{0i} + b_{1i}Pct\_Checkup_i \\
& + b_{2i}Median\_Age_i + b_{3i}Pct\_Hispanic_i \\
& + b_{4i}Pct\_SNAP_i + b_{5i}Pct\_College_i + \varepsilon_i,
\end{aligned} \tag{13}
$$

where obesity percentage (*Pct_Obesity*) is regressed on percentage of population undergoing yearly checkups (*Pct_Checkup*), median age of people in the census tract (*Median_Age*), percentage of Hispanic population, percentage of SNAP recipients (*Pct_SNAP*), and percentage of population with college degrees (*Pct_College*); $b_{0i}, \ldots, b_{5i}$ are the local parameter estimates of MGWR. We apply an adaptive bi-square kernel using nearest neighbors to calibrate the MGWR model. The largest possible bandwidth in this setting is the total number of census tracts in the study area, which is 815. Both the covariates and the response variable are standardized as suggested by Fotheringham, Yang, and Kang (2017) so that bandwidths and local parameters are comparable and invariant to the scale of the data.

Table 4. Ninety-five percent confidence interval of bandwidth based on Akaike weight for each covariate

Covariates	Optimal bandwidth	95% Confidence interval
Intercept	70	[60, 100]
Pct_Checkup	40	[40, 50]
Median_Age	815	[440, 815]
Pct_Hispanic	260	[160, 450]
Pct_SNAP	500	[370, 815]
Pct_College	815	[290, 760]

For each of the covariates and the intercept, we determine optimal bandwidths using AIC_c within a range of 40 to 815 nearest neighbors with a step size of ten. Figure 7 shows the bandwidth search history where the x axis is the bandwidth being evaluated and the y axis is the AIC_c value. Covariate-specific

Table 5. Summary of OLS parameter estimates and MGWR Akaike-weighted parameter estimates

	OLS			MGWR Akaike weighted	
Covariates	Est.	*SE*	*t* Value	Est. min[a]	Est. max[a]
Intercept	0.000	0.012	0.000	−0.297	0.503
Pct_Checkup	−0.235	0.030	−7.756	−0.610	−0.144
Median_Age	0.186	0.027	6.755	0.119	0.136
Pct_Hispanic	0.136	0.018	7.397	0.056	0.133
Pct_SNAP	0.536	0.021	25.230	0.323	0.332
Pct_College	−0.286	0.020	−14.439	−0.450	−0.362
R^2		0.88		0.93	
AIC_c		606.0		287.3	

Notes: OLS = ordinary least squares; MGWR = multiscale geographically weighted regression; AIC_c = corrected Akaike information criterion.
[a]Minimum and maximum of MGWR parameter estimates shown in the table are calculated for the significant (0.05) estimates only.

optimal bandwidths with minimum AIC_c values are 70, 40, 815, 260, 815, and 500 for *Intercept*, *Pct_Checkup*, *Median_Age*, *Pct_Hispanic*, *Pct_SNAP*, and *Pct_College*, respectively. We can see that for *Intercept* and *Pct_Checkup*, with small optimal bandwidths, AIC_c values increase dramatically with increasing bandwidths, which suggests that the data generating processes in both cases are very local. For the covariates *Median_Age* and *Pct_SNAP*, the AIC_c value decreases with increasing bandwidths, suggesting that the data generating processes for these relationships are global. For the covariates *Pct_Hispanic* and *Pct_College*, the optimal bandwidths suggest processes that exhibit some degree of spatial heterogeneity, although this is not very pronounced.

Akaike weights and bandwidth CIs are computed for these five covariates plus the intercept as shown in Figure 8. In each case the red line shows the optimal bandwidth and the two green lines show the upper and lower bounds of the 95 percent bandwidth CI. The sum of the area under the Akaike weight curve is equal to 1 and the sum of the area bounded by green dashed lines is approximately equal to 0.95. The optimal bandwidth and CIs are summarized in Table 4. The *Intercept* has an optimal bandwidth of 70 with a bandwidth CI from 60 to 100. The optimal bandwidth of the covariate *Pct_Checkup* is 40 with a bandwidth CI from 40 to 50, indicating a very locally heterogeneous relationship with obesity rate. The covariates *Median_Age* and *Pct_SNAP* both have global bandwidths of 815 with similar bandwidth CIs of [440, 815] and [370, 815], respectively. The covariate *Pct_Hispanic* has an optimal bandwidth of 260 with a CI from 160 to 450, indicating that the associated process is heterogeneous over a moderate spatial scale. Finally, the covariate *Pct_College* has an optimal bandwidth of 500 with a CI from 290 to 760, again indicating a process that exhibits a low degree of spatial heterogeneity. Consequently, we can conclude that the local parameter estimates for the *Intercept* and *Pct_Checkup* are significantly more heterogeneous than the estimates of the other covariates; the local parameter estimates for *Median_Age* and *Pct_SNAP* exhibit no significant spatial heterogeneity and the relationships between these variables and obesity rates are constant across Phoenix; and the relationships between obesity rates and *Pct_Hispanic* and *Pct_College* exhibit significant spatial heterogeneity but this heterogeneity is significantly less than that for the *Intercept* and *Pct_Checkup*.

For context, we also report the ordinary least squares (OLS) parameter estimates, along with the minimum and maximum of the significant MGWR Akaike-weighted parameter estimates in Table 5. The MGWR model has a better fit than the OLS model, with a higher R^2 value (0.93 vs. 0.88) and lower AIC_c (287.3 vs. 606.0). Maps of each of the four sets of local Akaike-weighted parameter estimates are shown in Figure 9. These are very similar to the conditional parameter estimates obtained from MGWR; comparisons between the two sets of parameter estimates and standard errors (unconditional and conditional) can be found in Figures 10 and 11. Figure 9 shows the significant Akaike-weighted local parameters for each covariate using the same color scheme. Insignificant parameters, based on the critical *t* values with multiple hypothesis testing adjustment (da Silva and Fotheringham 2016; Yu et al. 2020), are masked out and shaded in gray. The different degrees of heterogeneity observed from the local parameter maps are closely related to the variations in the optimal bandwidths previously discussed. The local parameter estimates for *Intercept* and *Pct_Checkup* are visually much more heterogeneous than the local estimates associated with the other covariates that have larger bandwidths. The estimates for the local *Intercept* include both significantly positive and significantly negative estimates, both of which are locally clustered. The cluster of significantly positive estimates in central Phoenix indicates that here obesity rates are significantly

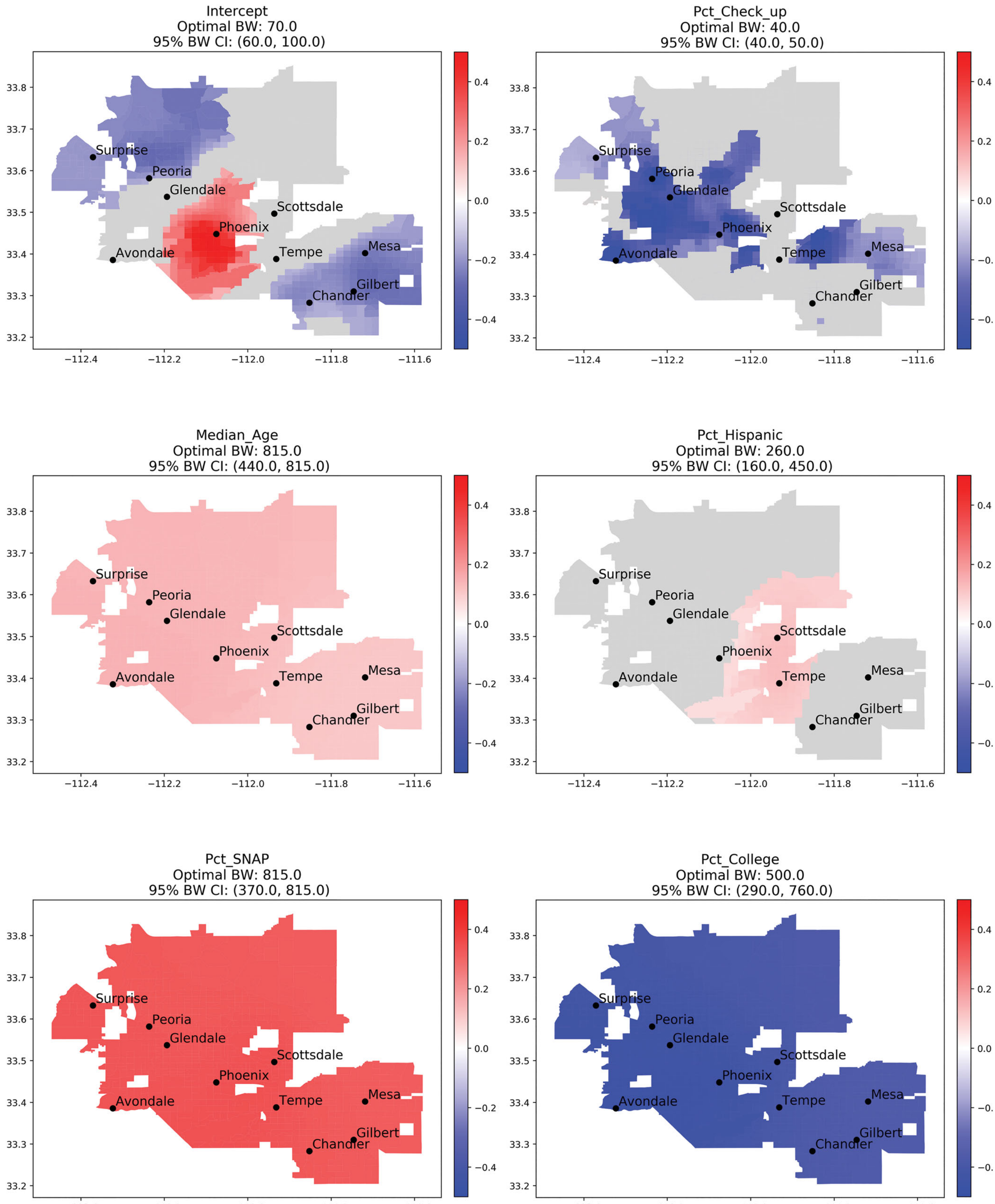

Figure 9. Maps of Akaike-weighted local parameter estimates with insignificant parameters masked out and shaded in gray. BW = bandwidth; CI = confidence interval.

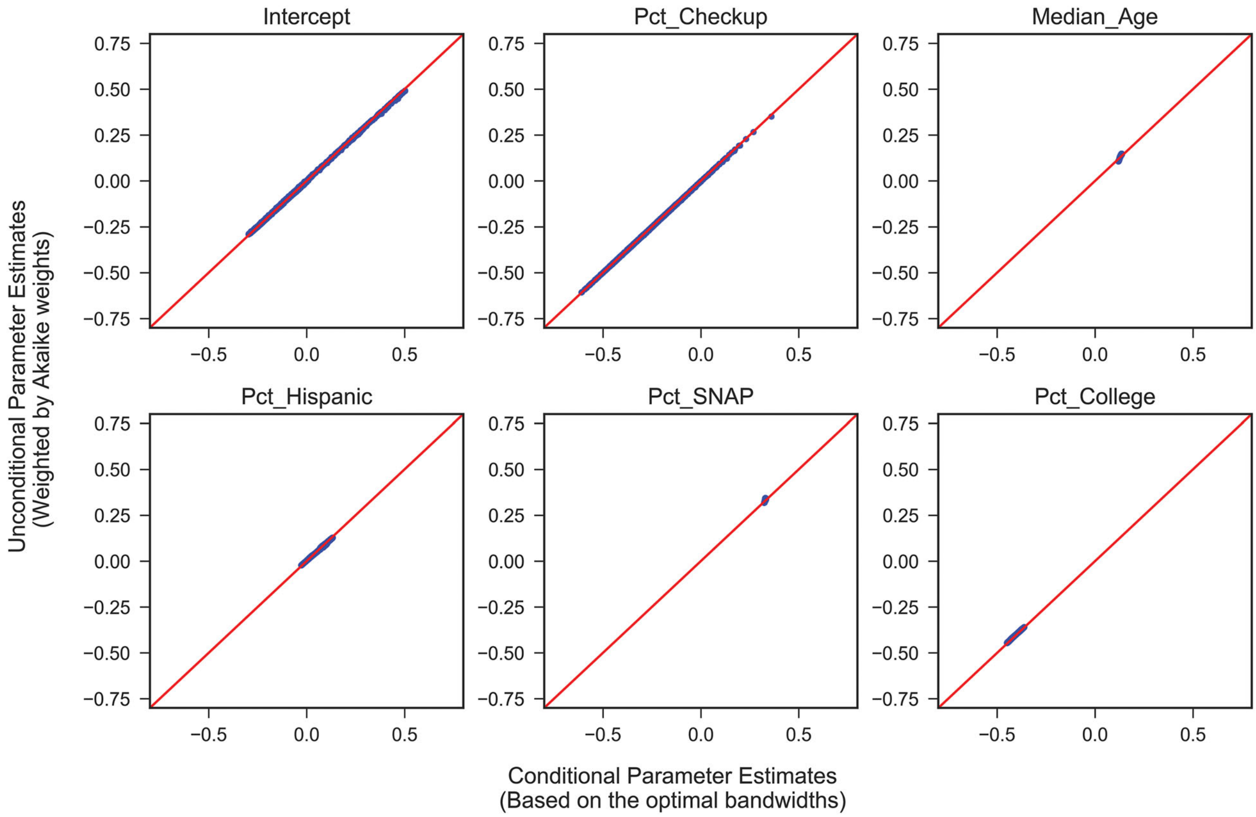

Figure 10. Comparison of conditional parameter estimates (based on the optimal bandwidths) with unconditional parameter estimates (calculated with Akaike weights).

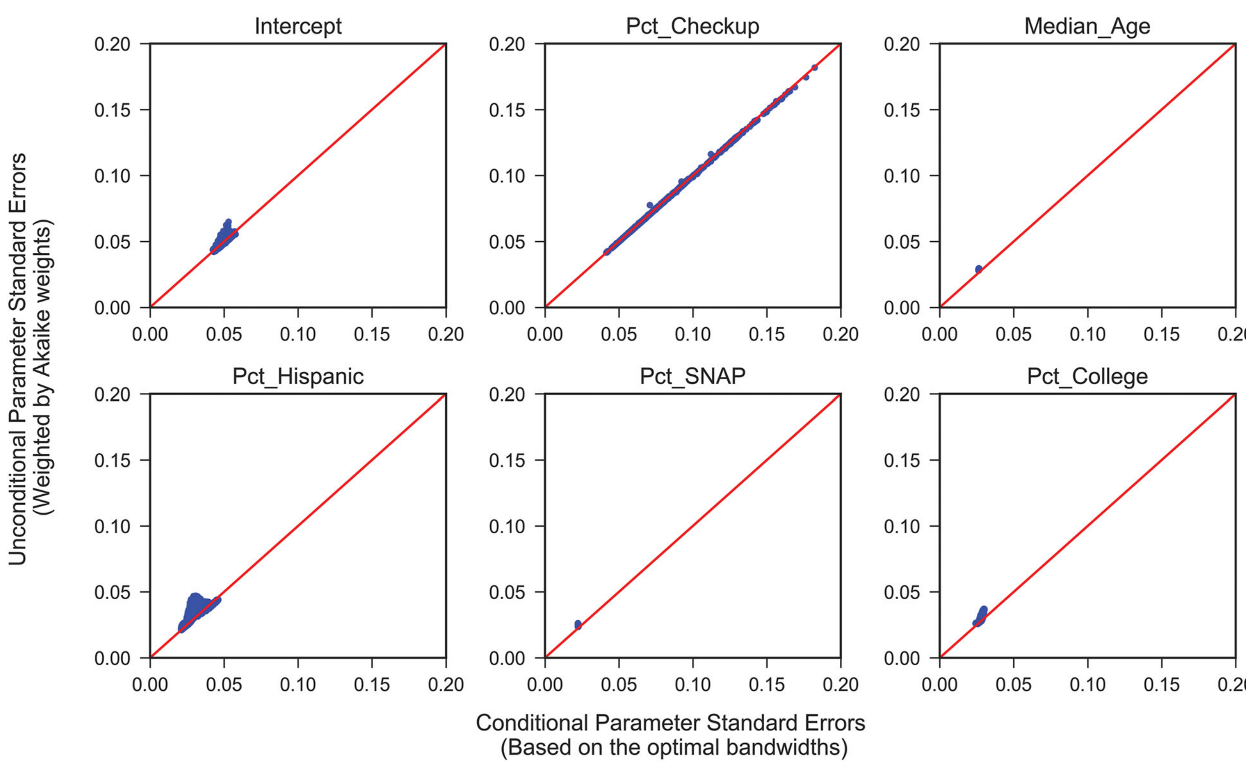

Figure 11. Comparison of conditional parameter standard errors (based on the optimal bandwidths) with unconditional parameter standard errors (calculated with Akaike weights).

higher than expected given the socioeconomic conditions modeled in this part of the city. Conversely, in the northwest and southeast areas of the greater Phoenix region, there are areas where obesity rates are significantly lower than expected given the modeled conditions in those areas. In the remaining parts of the city, the local intercept estimates are not significantly different from zero. The local estimates for *Pct_Checkup* have the greatest spatial variability with a range between –0.61 and 0.35 (including insignificant estimates), and significant negative relationships between obesity and having regular checkups are found across wide parts of central Phoenix in a band that stretches from the western to the most eastern parts of the city. No significant relationships are found in the rest of the city. The local parameter estimates for *Median_Age* are significantly positive everywhere—as we age, we have a tendency to become obese—and they vary only gradually across the city. Similarly, there are only minor variations in the local estimates for the parameters associated with the covariate *Pct_Hispanic*, and these are generally insignificant except in Scottsdale and Tempe. The local parameter estimates for *Pct_SNAP* and *Pct_College* are virtually identical everywhere, indicating again the global nature of these relationships. As expected, obesity rates are higher in areas with high percentages of families on SNAP and lower in areas where higher proportions of the population have a college degree.

Conclusion and Future Work

Bandwidths in GWR are essentially related to the scale over which the underlying spatial processes operate. The recent advancement of MGWR allows covariate-specific optimal bandwidths to be determined, allowing comparisons of the spatial scales over which different processes operate to be made (Fotheringham, Yang, and Kang 2017). Bandwidth selection based on optimizing a goodness-of-fit criterion such as AIC_c, however, contains an intrinsic uncertainty associated with it because bandwidth uncertainty is not accounted for. In this article, we correct this omission and show from the evidence of bootstrapping that the selection of the optimal bandwidth is subject to random sampling variation and that it is important to account for this uncertainty when relating bandwidths to the scales of spatial processes.

In addition, we examine the use of Akaike weights, an information theory statistic, to measure covariate-specific bandwidth uncertainty and to obtain bandwidth CIs. Akaike weights quantify the probability that a given bandwidth is optimal, which is a natural extension of AIC-based model selection (Burnham and Anderson 1998). For example, a bandwidth with an Akaike weight of 0.6 has a probability of 0.6 of being selected as the optimal bandwidth given the data. Based on both simulated data and an empirical example of modeling Phoenix obesity rates, we find that Akaike weight–derived bandwidth CIs can provide useful insights into the spatial scale over which different processes operate. Akaike weights can also be perceived as the prior probability of a model being the true model and therefore can be used to weight parameter estimates across candidate bandwidths and obtain unconditional local parameters that are independent of a single bandwidth. We find that unconditional local parameter estimates generally have more accurate CIs, although the improvements are marginal in the examples used here. It remains to be seen whether bandwidth uncertainty has generally little impact on parameter estimate uncertainty. The computation of Akaike weights is implemented in the *mgwr* Python package (Oshan, Li, et al. 2019) to increase accessibility of the methodology introduced in this article.

Model selection uncertainty has gained increasing attention in statistical modeling, although the majority of the literature about model selection relates to variable selection. Within the GWR framework, bandwidth selection is essentially a model selection problem. In this article, we therefore pay attention to the issue of model uncertainty from the aspect of bandwidth selection to help understand how spatial processes operate. An important future task is to investigate variable selection uncertainty and bandwidth uncertainty simultaneously in terms of their impact on local parameter estimates. It would be useful to investigate how the optimal bandwidth(s) and their associated uncertainties behave under the presence of omitted variable bias, which is an often-encountered scenario in real-world applications. It is expected that model parameter estimates will be more robust when simultaneously accounting for both bandwidth uncertainty and variable selection uncertainty so that they are not conditional on a preselected bandwidth or a set of preselected variables. These tasks and the work presented here provide great potential to enhance the quantification of process heterogeneity and scale.

Notes

1. Using a step size smaller than ten will produce more detailed Akaike weight curve but with additional computation.
2. Six sparsely populated tracts are removed in this example.

References

Akaike, H. 1979. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika* 66 (2):237–42.

Akaike, H. 1983. Information measures and model selection. *International Statistical Institute* 44:277–91.

Bozdogan, H. 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52 (3):345–70. doi: 10.1007/BF02294361.

Buja, A., T. Hastie, and R. Tibshirani. 1989. Linear smoothers and additive models. *The Annals of Statistics* 17 (2):453–510. doi: 10.1214/aos/1176347115.

Burnham, K. P., and D. R. Anderson. 1998. *Model selection and inference*. New York: Springer.

Burnham, K. P., D. R. Anderson, and K. P. Huyvaert. 2011. AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* 65 (1):23–35. doi: 10.1007/s00265-010-1029-6.

Centers for Disease Control and Prevention. 2016. 500 Cities: Local data for better health. https://www.cdc.gov/500cities

da Silva, A. R., and A. S. Fotheringham. 2016. The multiple testing issue in geographically weighted regression. *Geographical Analysis* 48 (3):233–47. doi: 10.1111/gean.12084.

Efron, B., and R. J. Tibshirani. 1994. *An introduction to the bootstrap*. Boca Raton, FL: CRC.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically weighted regression: The analysis of spatially varying relationships*. Hoboken, NJ: Wiley.

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. doi: 10.1080/24694452.2017.1352480.

Hastie, T. J., and R. J. Tibshirani. 1990. *Generalized additive models*. Vol. 43. Boca Raton, FL: CRC.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Harris, P. 2019. A simulation study on specifying a regression model for spatial data: Choosing between autocorrelation and heterogeneity effects. *Geographical Analysis* 51 (2):151–81. doi: 10.1111/gean.12163.

Harris, P., C. Brunsdon, B. Lu, T. Nakaya, and M. Charlton. 2017. Introducing bootstrap methods to investigate coefficient non-stationarity in spatial regression models. *Spatial Statistics* 21:241–61. doi: 10.1016/j.spasta.2017.07.006.

Hurvich, C. M., and C. L. Tsai. 1993. A corrected Akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis* 14 (3):271–79. doi: 10.1111/j.1467-9892.1993.tb00144.x.

Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. *Trends in Ecology & Evolution* 19 (2):101–8. doi: 10.1016/j.tree.2003.10.013.

Koh, L. P. 2008. Can oil palm plantations be made more hospitable for forest butterflies and birds? *Journal of Applied Ecology* 45 (4):1002–9. doi: 10.1111/j.1365-2664.2008.01491.x.

Li, Z., and A. S. Fotheringham. Forthcoming. Computational improvements to multiscale geographically weighted regression. *International Journal of Geographical Information Science*. Advance online publication. doi: 10.1080/13658816.2020.1720692.

Li, Z., A. S. Fotheringham, W. Li, and T. Oshan. 2019. Fast geographically weighted regression (FastGWR): A scalable algorithm to investigate spatial process heterogeneity in millions of observations. *International Journal of Geographical Information Science* 33 (1):155–75.

Mei, C. L., M. Xu, and N. Wang. 2016. A bootstrap test for constant coefficients in geographically weighted regression models. *International Journal of Geographical Information Science* 30 (8):1622–43. doi: 10.1080/13658816.2016.1149181.

Murakami, D., B. Lu, P. Harris, C. Brunsdon, M. Charlton, T. Nakaya, and D. A. Griffith. 2019. The importance of scale in spatially varying coefficient modeling. *Annals of the American Association of Geographers* 109 (1):50–70. doi: 10.1080/24694452.2018.1462691.

Oshan, T. M., and A. S. Fotheringham. 2018. A comparison of spatially varying regression coefficient estimates using geographically weighted and spatial-filter-based techniques. *Geographical Analysis* 50 (1):53–75. doi: 10.1111/gean.12133.

Oshan, T. M., Z. Li, W. Kang, L. J. Wolf, and A. S. Fotheringham. 2019. mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information* 8 (6):269. doi: 10.3390/ijgi8060269.

Oshan, T. M., J. Smith, and A. S. Fotheringham. 2019. A multiscale spatial analysis of obesity determinants in Phoenix, Arizona. Open Science Framework. doi: 10.31219/osf.io/unfwj.

Oshan, T. M., L. J. Wolf, A. S. Fotheringham, W. Kang, Z. Li, and H. Yu. 2019. A comment on geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science* 33 (7):1289–99. doi: 10.1080/13658816.2019.1572895.

Pinsky, M. L., B. Worm, M. J. Fogarty, J. L. Sarmiento, and S. A. Levin. 2013. Marine taxa track local climate velocities. *Science* 341 (6151):1239–42. doi: 10.1126/science.1239352.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53 (5):793–808. doi: 10.1080/10635150490522304.

Symonds, M. R., and A. Moussalli. 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology* 65 (1):13–21. doi: 10.1007/s00265-010-1037-6.

U.S. Census Bureau. 2015. 2011–2015 American Community Survey 5-year estimates. https://factfinder.census.gov.

Wagenmakers, E. J., and S. Farrell. 2004. AIC model selection using Akaike weights. *Psychonomic Bulletin & Review* 11 (1):192–96. doi: 10.3758/BF03206482.

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. doi: 10.1111/gean.12147.

Wood, S. N., N. Pya, and B. Säfken. 2016. Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* 111 (516):1548–63. doi: 10.1080/01621459.2016.1180986.

Yu, H., A. S. Fotheringham, Z. Li, T. Oshan, W. Kang, and L. J. Wolf. 2020. Inference in multiscale geographically weighted regression. *Geographical Analysis* 52 (1):87–106. doi: 10.1111/gean.12189.

ZIQI LI is a PhD Candidate in the School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85281. E-mail: Li.Ziqi@asu.edu. His research interests include GIScience and spatial data science. He is currently working on the MGWR model, covering aspects from computation and inference to applications.

A. STEWART FOTHERINGHAM is a Regents' Professor of Computational Spatial Science in the School of Geographical Sciences and Urban Planning at Arizona State University, Tempe, AZ 85281. E-mail: stewart.fotheringham@asu.edu. His research interests include spatial data analytics, spatial processes, and spatial interaction modeling.

TAYLOR M. OSHAN is an Assistant Professor in the Center for Geospatial Information Science within the Department of Geographical Sciences at the University of Maryland, College Park, MD 20740. E-mail: toshan@umd.edu. His research interests are centered on developing methods to analyze multiscale spatial and temporal processes and applying them in the context of urban health and transportation, as well as building open source tools.

LEVI JOHN WOLF is a Senior Lecturer (Assistant Professor) in the School of Geographical Sciences at the University of Bristol, Bristol BS8 1QU UK. E-mail: levi.john.wolf@bristol.ac.uk. His research interests include probabilistic programming and machine learning for applications in urban geography, politics, sociology, and economics.

Appendix A

This simulation example used the model described in Equation 10 in the text but calibrated with a fixed Gaussian kernel (distance based) instead of an adaptive bi-square kernel. We found that the results show similar bandwidth distributions (see Figures A.1 and A.2).

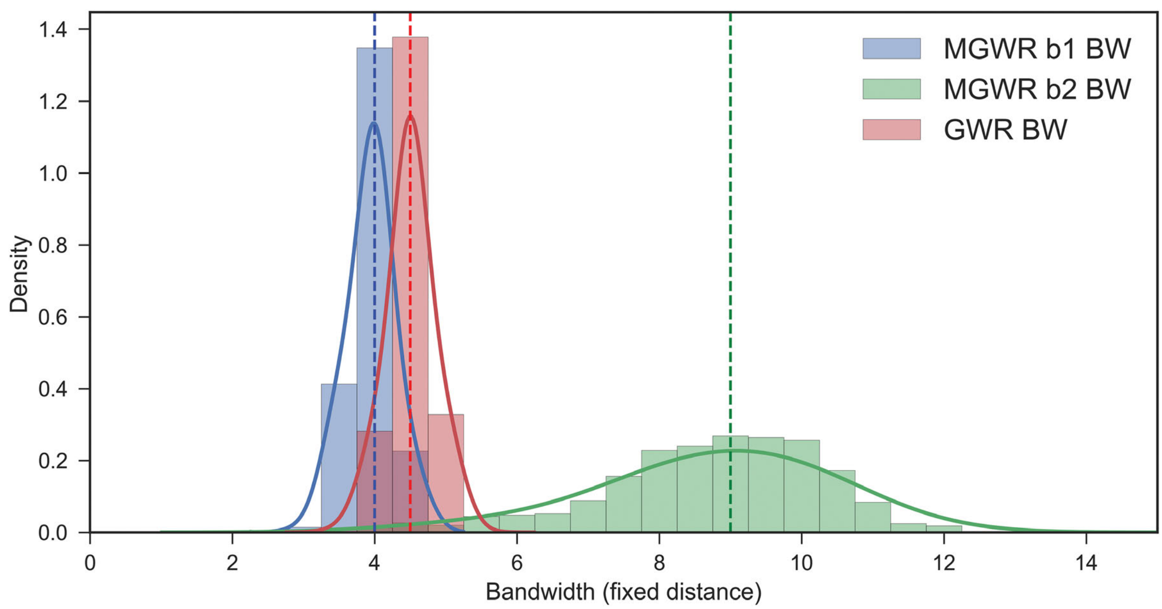

Figure A.1. Bootstrapping bandwidth distributions of the model described in Equation 10 with a distance-based fixed kernel. MGWR = multiscale geographically weighted regression; BW = bandwidth; GWR = geographically weighted regression.

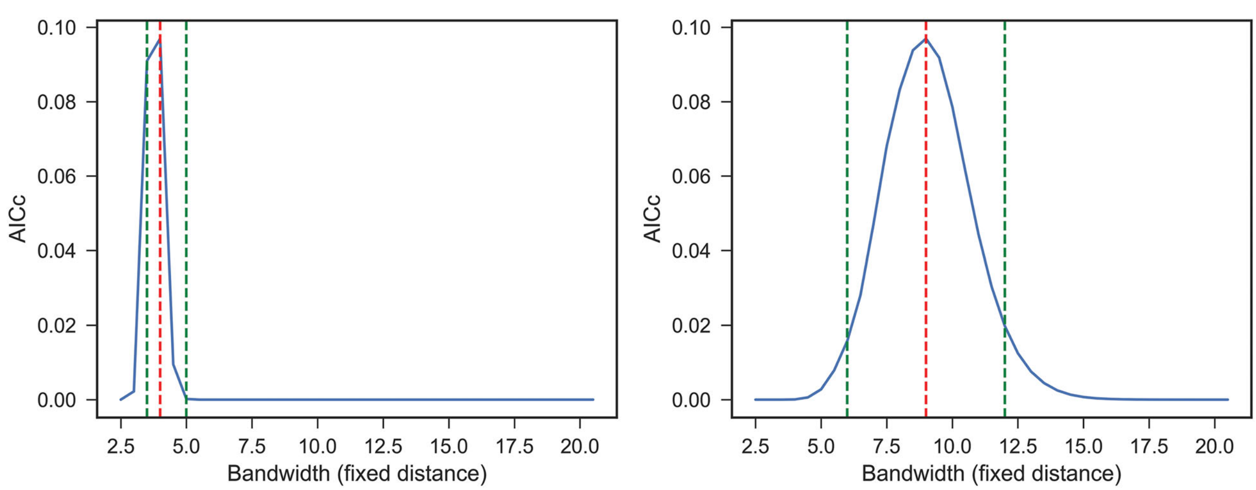

Figure A.2. Akaike weights for two processes shown in Figure 1 with a distance-based fixed kernel. AICc = corrected Akaike information criterion.

Appendix B

In this simulation example, we added a spatially varying intercept b_0 (with heterogeneity between b_1 and b_2) to the model described in Equation 10 in the text, and the simulated processes can be seen in Figure B.1. The MGWR bandwidth distributions can be seen in Figure B.2 accordingly.

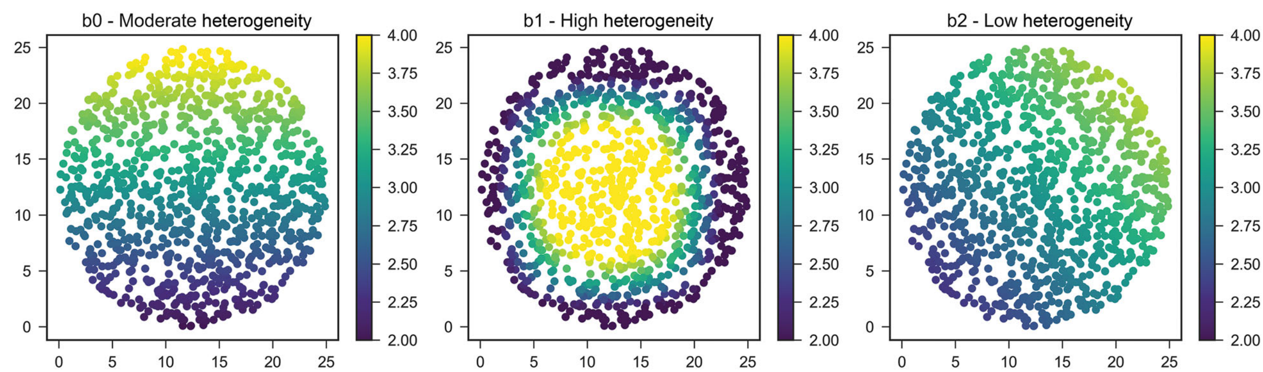

Figure B.1. Simulated processes.

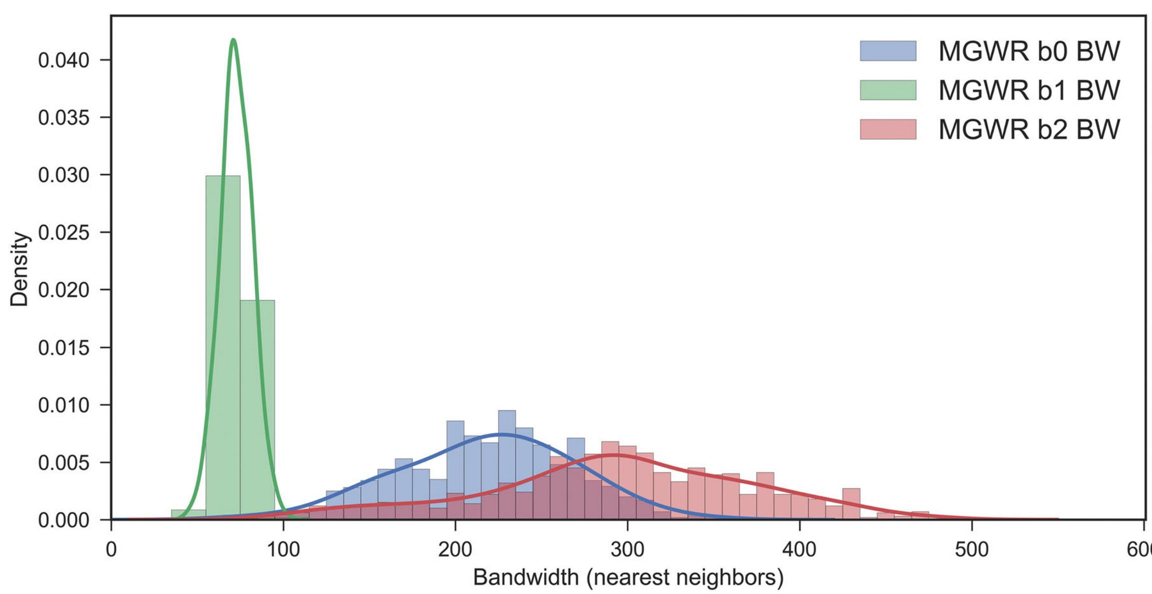

Figure B.2. Optimal bandwidth distributions for simulated processes in Figure B.1 under bootstrapping. MGWR = multiscale geographically weighted regression; BW = bandwidth.

Appendix

In this simulation example, we rotate the spatial process b_2 (Figure 2 in the text) to reproduce four spatial processes with different orientations but each having the same degree of heterogeneity, as illustrated in Figure C.1. The objective is to explore how the optimal bandwidths and associated uncertainties respond to variations in the orientation of the process holding the level of heterogeneity constant. We found that process orientation has no effect on the optimal bandwidths obtained from MGWR (see Figure C.2); bandwidth variation is a function solely of the scale over which the processes operate.

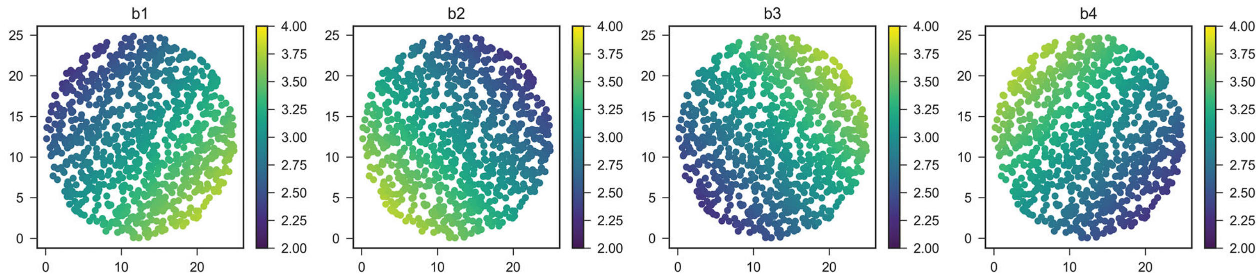

Figure C.1. Four generated spatial processes with the same degree of heterogeneity.

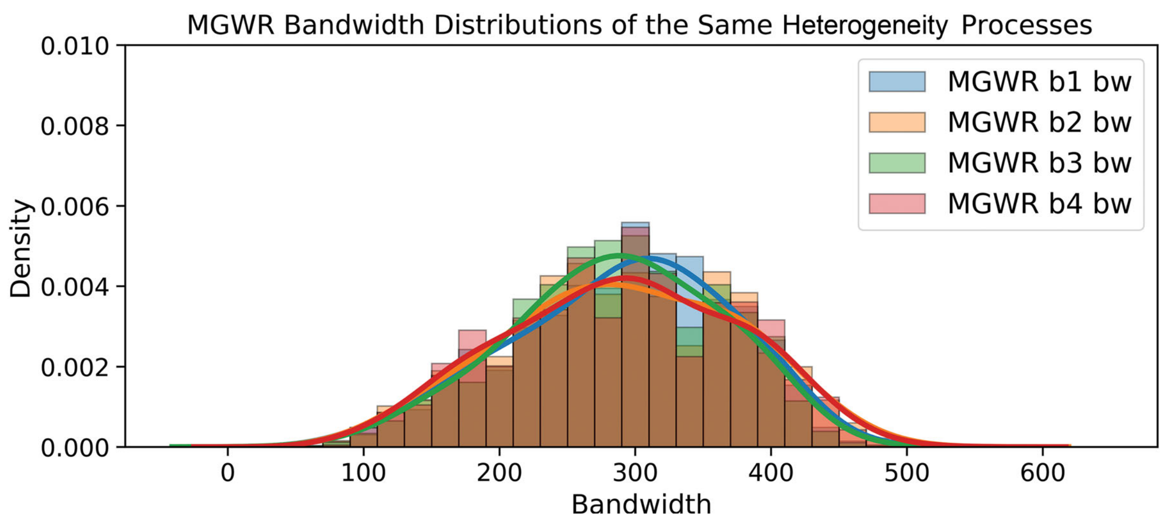

Figure C.2. Optimal bandwidth distributions of four same heterogeneity processes in Figure C.1 under bootstrapping. MGWR = multiscale geographically weighted regression; bw = bandwidth.