

 Open access • Journal Article • DOI:10.1080/24694452.2019.1704680

## Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights — [Source link](#)

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

**Institutions:** Arizona State University, University of Maryland, College Park, University of Bristol

**Published on:** 11 Feb 2020 - Annals of the American Association of Geographers (Routledge)

**Topics:** Akaike information criterion

Related papers:

- [Multiscale Geographically Weighted Regression \(MGWR\)](#)
- [mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale](#)
- [Geographically Weighted Elastic Net: A Variable-Selection and Modeling Method under the Spatially Nonstationary Condition](#)
- [Scalable GWR: A Linear-Time Algorithm for Large-Scale Geographically Weighted Regression with Polynomial Kernels](#)
- [Semi-parametric Geographically Weighted Regression Modelling using Linear Model of Coregionalization](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/measuring-bandwidth-uncertainty-in-multiscale-geographically-19i8slvpi3>

# Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights

Ziqi Li<sup>1</sup>, A. Stewart Fotheringham<sup>1\*</sup>, Taylor M. Oshan<sup>2</sup>, Levi John Wolf<sup>3</sup>

1. Spatial Analysis Research Center, School of Geographical Sciences and Urban Planning, Arizona State University, USA

2. Center for Geospatial Information Science, Department of Geographical Sciences, University of Maryland, USA

3. School of Geographical Sciences, University of Bristol, UK

*\*Corresponding Author: A. Stewart Fotheringham*

Funding: This research was funded by U.S. National Science Foundation (NSF) grant number 1758786.

## Abstract

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 paper, we find that bandwidth selection is subject to uncertainty in both single-scale and multi-scale 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.

Keyword: *multiscale geographically weighted regression, bandwidth, spatial processes scale, Akaike weight, model selection uncertainty*

## 1. Introduction

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 operating over different spatial scales. Classic global models ignore both 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 process but inadequately incorporate multi-scale 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 et al., 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 Multi-scale GWR (MGWR) removes the single bandwidth assumption and allows the bandwidths for each covariate to vary (Fotheringham et al., 2017). This results in each parameter surface being allowed to have a different degree of spatial variation, reflecting variation across covariate-specific processes.

Comparisons of MGWR with other single-scale and multi-scale spatially varying coefficients (SVC) models are available, such as the comparison with classic GWR (Fotheringham et al., 2017; Harris, 2019; Murakami et al., 2019), with Bayesian SVC (Wolf et al., 2017), with eigenvector spatial filtering (ESF) (Oshan et al., 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., 2019), computational improvements by parallelization (Li et al., 2019; Li and Fotheringham, 2019), and accessible software (Oshan et al., 2019a) have been developed for MGWR, all of which greatly enhances the utility of MGWR in modeling multi-scale spatially heterogeneous processes.

Bandwidth is a key parameter in the geographically weighted regression 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 an Akaike Information Criterion (AIC), 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 tradeoff between bias and variance for parameter estimates. A large bandwidth produces local parameter estimates with low variance at the cost of high bias (under-fitting), whereas a small bandwidth yields parameter estimates with low bias but high variance (over-fitting). 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 spatially smoothed process which has regional to global trend, 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 et al., 2017). The bandwidth also has an interpretable real-world meaning. There are two popular definitions of bandwidth that are used in the GWR framework: (i) a distance-based measure and (ii) the number of nearest neighbors used in each local regression. The interpretation of say a 50-km bandwidth (or 300-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 50 km (or 300-nearest neighbors). In either case, this means that the data at the focal point of the local regression (if they exist) have a weight of 1 and all other data used in the local calibration have a weight less than 1, 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. However, if the analyst wants to make an inference about bandwidths as indicators of the relative spatial scales of different processes, it is naïve to simply compare the covariate-specific bandwidths which are obtained based on the single observed dataset (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 et al., 2009; Wood et al., 2016). Additionally, previous studies show that GWR and MGWR bandwidths are not fixed from the evidence of Monte Carlo simulations and sub-sampling results (Fotheringham et al., 2017; Wolf et al., 2018; Oshan et al., 2019b). Nevertheless, there are no methods in the context of 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 et al., 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’.

## **2. Overview of the procedure for measuring bandwidth uncertainty**

Burnham and Anderson (2002) discuss extensively the uncertainty regarding model selection and propose 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 (Posada and Buckley, 2004; Johnson and Omland, 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% 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 multi-model 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, 2002; Wagenmakers and Farrell, 2004; Symonds and Moussalli, 2011; Burnham and Anderson, 2011).

In this paper, 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 (CI) 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 paper proceeds as follows. In Section 3, the background of MGWR is reviewed. In Section 4 bandwidth uncertainties are examined with a simulation dataset using bootstrapping and Akaike weights. In Section 5, an empirical example of obesity rate modeling for the city of Phoenix is presented. The paper concludes with remarks in Section 6.

### 3. Background of MGWR related to Akaike weights

Multiscale geographically weighted regression (MGWR) was developed under the generalized additive model framework of Hastie and Tibshirani (1990) by Fotheringham et al. (2017). MGWR is formulated as

$$\mathbf{y} = \mathbf{f}_1(\mathbf{x}_1) + \mathbf{f}_2(\mathbf{x}_2) + \cdots + \mathbf{f}_k(\mathbf{x}_k) + \boldsymbol{\varepsilon} \quad (1)$$

where  $\mathbf{y}$  is a column vector of response variables,  $\mathbf{f}_1(\mathbf{x}_1), \dots, \mathbf{f}_k(\mathbf{x}_k)$  are additive components which are smooth functions of covariates, and  $\boldsymbol{\varepsilon}$  is a column vector of i.i.d. error terms. The response variable  $\mathbf{y}$  is the data observed over a spatial surface and  $\mathbf{f}_1, \dots, \mathbf{f}_k$  are spatial additive components estimated with covariate-specific bandwidths, so MGWR can be formulated as

$$\mathbf{y} = \mathbf{f}_{bw_1}(\mathbf{x}_1) + \mathbf{f}_{bw_2}(\mathbf{x}_2) + \cdots + \mathbf{f}_{bw_k}(\mathbf{x}_k) + \boldsymbol{\varepsilon} \quad (2)$$

where each additive component  $\mathbf{f}_{bw_j}(\mathbf{x}_j)$  is a product of element-wise multiplication ( $\otimes$ ) of local parameters  $\boldsymbol{\beta}_j$  and covariate  $\mathbf{x}_j$

$$\mathbf{f}_{bw_j}(\mathbf{x}_j) = \boldsymbol{\beta}_j \otimes \mathbf{x}_j = \begin{pmatrix} \beta_{1j}x_{1j} \\ \beta_{2j}x_{2j} \\ \vdots \\ \beta_{nj}x_{nj} \end{pmatrix} \quad (3)$$

The calibration of each smooth function  $\mathbf{f}_{bw_j}$  in MGWR model uses the back-fitting algorithm developed by Buja et al. (1989). The estimation procedure can be initialized in various ways (e.g. using GWR estimates, Fotheringham et al. 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 ( $\hat{\mathbf{f}}_j + \hat{\boldsymbol{\varepsilon}} \sim \mathbf{x}_j$ ) successively. The back-fitting converge when parameter estimates are unchanging within a pre-defined threshold. A detailed description and implementation of the back-fitting algorithm can be found in Fotheringham et al. (2017) and Oshan et al. (2019a). Covariate-specific optimal bandwidths are estimated within univariate GWRs by minimizing the corrected AIC (AICc), which is formulated as

$$AIC_c = -2\log(\mathcal{L}) + 2n \left( \frac{tr(\mathbf{S}) + 1}{n - tr(\mathbf{S}) - 2} \right) \quad (4)$$

where  $\mathcal{L}$  is the model likelihood given the data and  $tr(\mathbf{S})$  is the trace of the hat matrix. AICc is a small sample bias adjustment to the classic AIC and should be used when the ratio of data points and number of

parameters is below 40, which is often seen in GWR when processes have high-heterogeneity. When the ratio is greater than 40, the AICc rapidly approaches the classic AIC (Hurvich and Tsai, 1989; Burnham and Anderson, 2002). Therefore, in general, this corrected version of AIC is often suggested in GWR for bandwidth selection (Fotheringham et al., 2002). The use of AIC or its variant AICc for model selection is based on the information-theory that there is a relationship between expected Kullback-Leibler distance (information lost) and the maximized log-likelihood. The best model represents the process that generated 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 AICc. For simplicity, we will use AIC here (rather than AICc) as a generic term for all AIC variants. A candidate set of  $R$  bandwidths can be defined as  $\{bw_1, bw_2, bw_3 \dots, bw_R\}$  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  $k \in 1 \dots R$  can be obtained by

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

The numerator  $\exp\left(-\frac{1}{2}\Delta_k\right)$  denotes the likelihood of the bandwidth given the maximum likelihood estimators based on the same data, which also measure the relative strength of evidence for each bandwidth (Akaike, 1983; Burnham and Anderson, 2002). The denominator is used to normalize the Akaike weights so that all values lie between 0 and 1 with the sum being 1 ( $\sum_1^R w_k = 1$ ) (Burnham and Anderson, 2002). 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% chance of being the bandwidth that best approximates the corresponding underlying process. 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 will examine the use of Akaike weights for measuring bandwidth uncertainty based on a simulated dataset.

## 4. Investigating bandwidth uncertainty in a controlled experiment

### 4.1 Construction of a simulated dataset with two different spatially heterogeneous processes

We simulate a study area with 1000 locations that are randomly distributed in a circular coordinate space. A circular coordinate space is used because in a later section we rotate the coordinate space to see if the optimal bandwidth is sensitive to the orientation of the heterogeneity in the processes. The  $x$  and  $y$  coordinates of location  $i$  are denoted as  $u_i$  and  $v_i$  and are constructed with the following formulas (6) and (7)

$$u_{1i} = 12.5 + 12.5\sqrt{r} \cos \theta \quad (6)$$

$$v_{1i} = 12.5 + 12.5\sqrt{r} \sin \theta \quad (7)$$

where a radius  $r$  and angle  $\theta$  are randomly selected from uniform distributions. Then we synthesize two spatial processes using formulas (8) and (9), and these true parameter surfaces,  $\mathbf{b}_1$  and  $\mathbf{b}_2$ , are shown in Figure 1.

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

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

[Figure 1 about here]

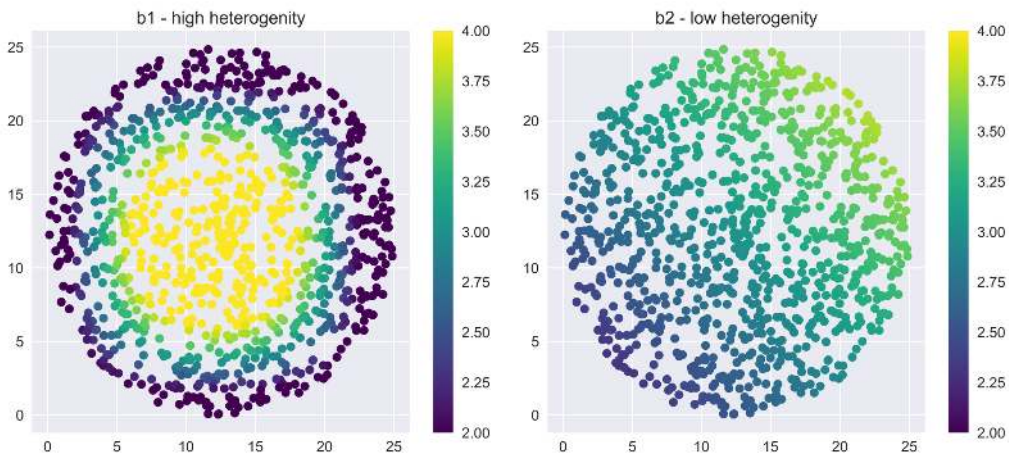


Figure 1. Generated spatial processes with low- and high- heterogeneity.



where  $b_{1i}$  and  $b_{2i}$  are the parameters of location  $i$  ( $i = 1 \dots n$ );  $u_i$  and  $v_i$  are the  $x$  and  $y$  coordinates of location  $i$ . The resulting process  $\mathbf{b}_1$  has high spatial heterogeneity, with high values in the center of the map and low values at the periphery. The parameters range between 0 to 5. The process  $\mathbf{b}_2$  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 et al. (2017).

Covariate vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are randomly drawn from a normal distribution with mean of 0 and variance of 1. A spatially random noise vector  $\boldsymbol{\varepsilon}$  is added to the surface with mean of 0 and variance of 1. Those two processes,  $\mathbf{b}_1$  and  $\mathbf{b}_2$ , along with two covariates,  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , and a random noise are used to generate the synthetic response variable  $\mathbf{y}$  in the following manner:

$$\mathbf{y} = \mathbf{b}_1\mathbf{x}_1 + \mathbf{b}_2\mathbf{x}_2 + \boldsymbol{\varepsilon} \quad (10)$$

An MGWR model is calibrated on this synthetic dataset using the *mgwr* python package (Oshan et al., 2019a). For consistency, in the following simulation and empirical studies we use an adaptive bi-square kernel (Fotheringham et al., 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 dataset. For this model, the two covariate-specific optimal bandwidths are 70 and 300 for the local process  $\mathbf{b}_1$  and the regional process  $\mathbf{b}_2$  respectively. Recovered parameter estimate surfaces for this synthetic dataset are shown in Figure 2.

[Figure 2 about here]

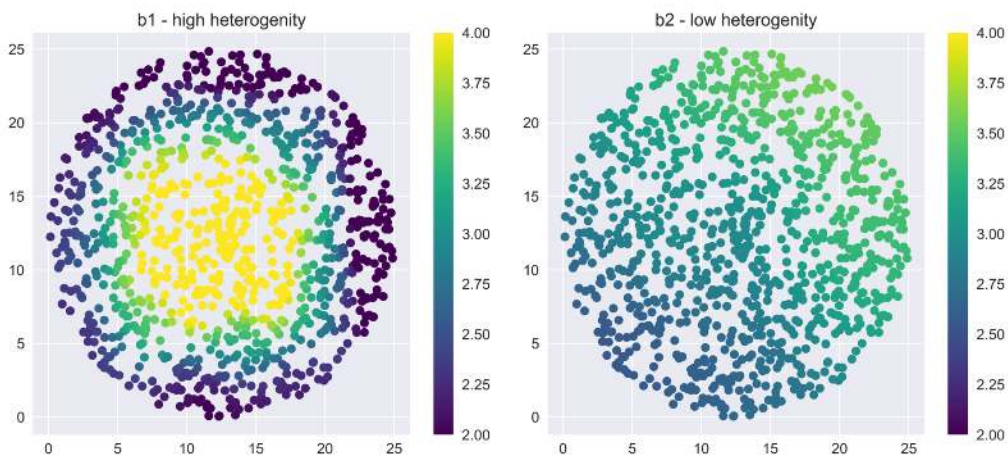


Figure 2. Recovered parameter surfaces using MGWR with optimal bandwidths 70 and 300 for the high- and low- heterogeneity processes, respectively.

#### 4.2 Bandwidth selection uncertainty: evidence from bootstrapping

Bootstrapping is a variant of Monte Carlo estimation for confidence intervals based on random sampling with *replacement* (Efron and Tibshirani, 1993). The technique is popularly used for deriving empirical confidence intervals of a statistic when the analytical solution is unavailable. In this paper, we are interested in assessing how the selected optimal bandwidths vary across each bootstrap sample and from this the empirical distribution of optimal bandwidths can be obtained. We use a non-parametric residual-based bootstrap method by randomly resampling residuals which 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 et al., 2016; Harris et al., 2017).

In the context presented here we first calibrate an MGWR model with the original response vector  $\mathbf{y}$  and two covariates  $\mathbf{x}_1, \mathbf{x}_2$  and save the fitted values  $\hat{\mathbf{y}}$  and residuals  $\hat{\boldsymbol{\epsilon}}$ . Then we generate  $B$  ( $B=1,000$ ) bootstrap samples with each sample containing a new response vector  $\mathbf{y}^* = \hat{\mathbf{y}} + \hat{\boldsymbol{\epsilon}}_{rand}$  where  $\hat{\boldsymbol{\epsilon}}_{rand}$  is a residual vector randomly drawn with replacement from  $\hat{\boldsymbol{\epsilon}}$ . For each bootstrap sample, we regress  $\mathbf{y}^*$  onto  $\mathbf{X}$  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 certain bandwidth  $j$  being selected as the optimal bandwidth. A 95% empirical confidence interval can be obtained by using the 2.5% and 97.5% percentiles of the bandwidths from the bootstrap samples (Efron and Tibshirani, 1993). This bootstrap method is applied to GWR for comparison with MGWR. The bootstrap results are presented in Figure 3 and Table 1.

**[Figure 3 about here]**

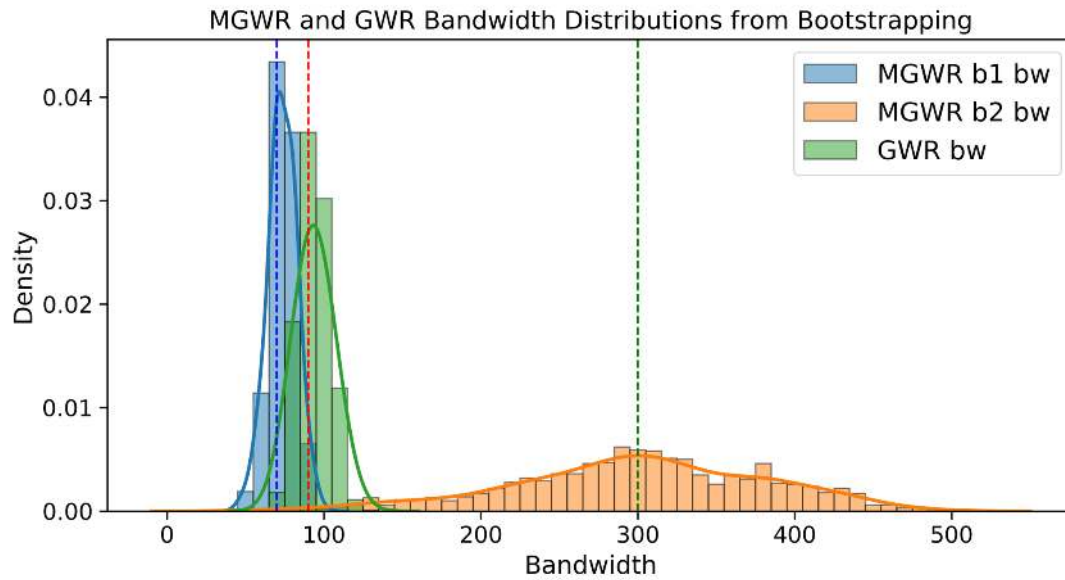


Figure 3. Optimal bandwidth distributions of two processes in Figure 1 under bootstrapping.

[Table 1 about here]

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

	OPTIMAL	MEAN	STD	95% CONFIDENCE INTERVAL	WIDTH
<b>GWR BW</b>	90	93.6	10.1	[80, 110]	30
<b>MGWR <math>b_1</math> BW</b>	70	73.5	8.4	[60, 90]	30
<b><math>b_2</math> BW</b>	300	300.5	80.1	[130, 440]	310

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 dataset 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 describe adequately 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 1000 bootstrap samples. The empirical 95% confidence interval is [80, 110]. For MGWR, the optimal bandwidth for the high-heterogeneity process  $\mathbf{b}_1$  has a mean of 73.5, standard deviation of 8.4, and the empirical 95% confidence interval is [60, 90]. The optimal bandwidth for the low-heterogeneity process  $\mathbf{b}_2$  has a mean of 300.5, standard deviation of 80.1 and an empirical confidence interval of [130, 440]. By comparing the two covariate-specific bandwidths in MGWR, we can see that the selection of the bandwidth for the low heterogeneity process  $\mathbf{b}_2$  tends to be more uncertain than that of the high heterogeneity process  $\mathbf{b}_1$  and has a much wider confidence interval (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 confidence intervals for bandwidths. The mutually exclusive confidence intervals for the bandwidths associated with processes  $\mathbf{b}_1$  and  $\mathbf{b}_2$  imply that the two processes have significantly different optimal bandwidths at the 95% confidence level and we can thus claim that process  $\mathbf{b}_1$  operates at a relatively more local scale than process  $\mathbf{b}_2$ .

Following the same bootstrapping method described above, another simple simulation is conducted to explore how the optimal bandwidths and associated uncertainties respond to variations in the orientation of the process holding the level of heterogeneity constant. To do this, we rotate the spatial process  $\mathbf{b}_2$  (in Figure 2) to reproduce four spatial processes with different orientations but each having the same degree of heterogeneity, as illustrated in Figure 4.

[Figure 4 about here]

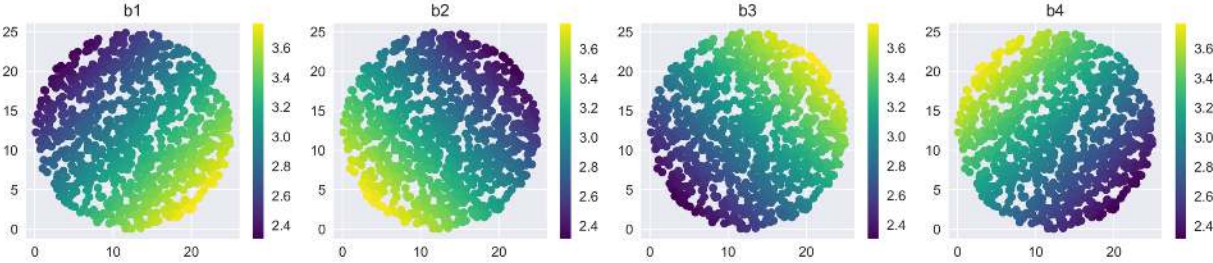


Figure 4. Four generated spatial processes with the same degree of heterogeneity.

Figure 5 shows the optimal bandwidth distributions for these four processes obtained from 1000 bootstrap samples and it is clear that the bandwidth distributions for each of the four processes are roughly the same. Consequently, process orientation has no effect on the optimal bandwidths obtained from MGWR; bandwidth variation is a function solely of the scale over which the processes operate.

[Figure 5 about here]

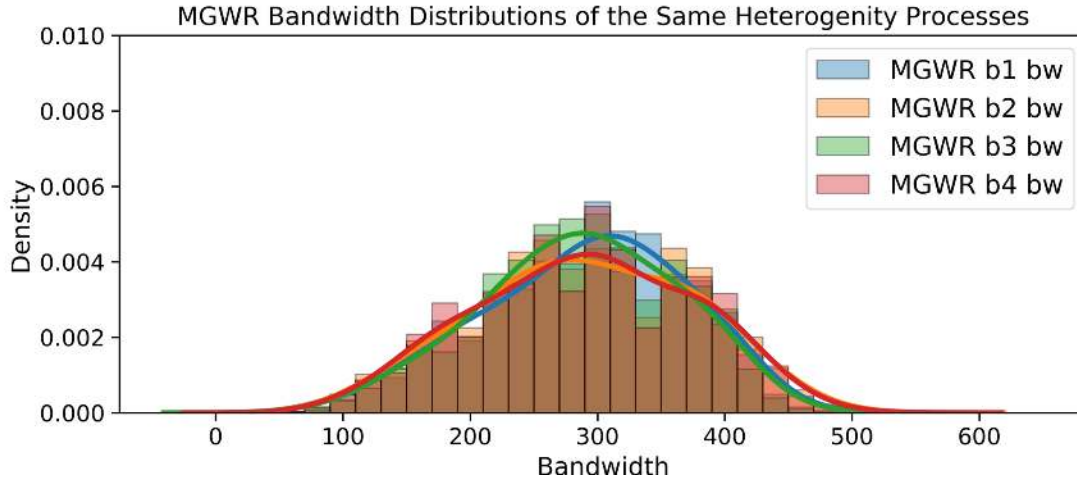


Figure 5. Optimal bandwidth distributions of four same heterogeneity processes in Figure 4 under bootstrapping.

### 4.3 Using Akaike weights for measuring bandwidth selection uncertainty

Following equation (5) in Section 3, we compute Akaike weights for the simulation dataset described in Section 4.1 which involves two processes with different degrees of spatial heterogeneity. For each covariate, we evaluate candidate bandwidths from 50 to 1000 nearest neighbors with a 10-neighbor interval<sup>1</sup>. Each candidate bandwidth has a corresponding Akaike weight representing the probability of it being the optimal bandwidth (Figure 6). 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 6. Following Burnham and Anderson (2002) and Symonds and Moussalli (2011), a 95% confidence interval (CI) of bandwidths can be obtained by ranking the Akaike weights in *descending* order and including bandwidths into the CI until the cumulative Akaike weight equals to 0.95. In some cases, where a coarse searching interval is used, the cumulative Akaike weight might not be exactly 0.95; therefore,

<sup>1</sup> Using a smaller step-size rather than 10 will produce more detailed Akaike weight curve but with additional computation.

the inclusion to 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 into the 95% CI stops after adding the bandwidth 90 because at this point the cumulative Akaike weight equals 0.95. Consequently, we can conclude that the 95% 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% CI.

[Table 2 about here]

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

BANDWIDTH	AICC	AKAIKE WEIGHT	CUM. AKAIKE WEIGHT
70	2907.87	0.42	0.42
60	2908.94	0.25	0.67
80	2909.13	0.23	0.90
90	2911.98	0.05	0.95
50	2912.69	0.04	0.99
...	...	...	...

[Figure 6 about here]

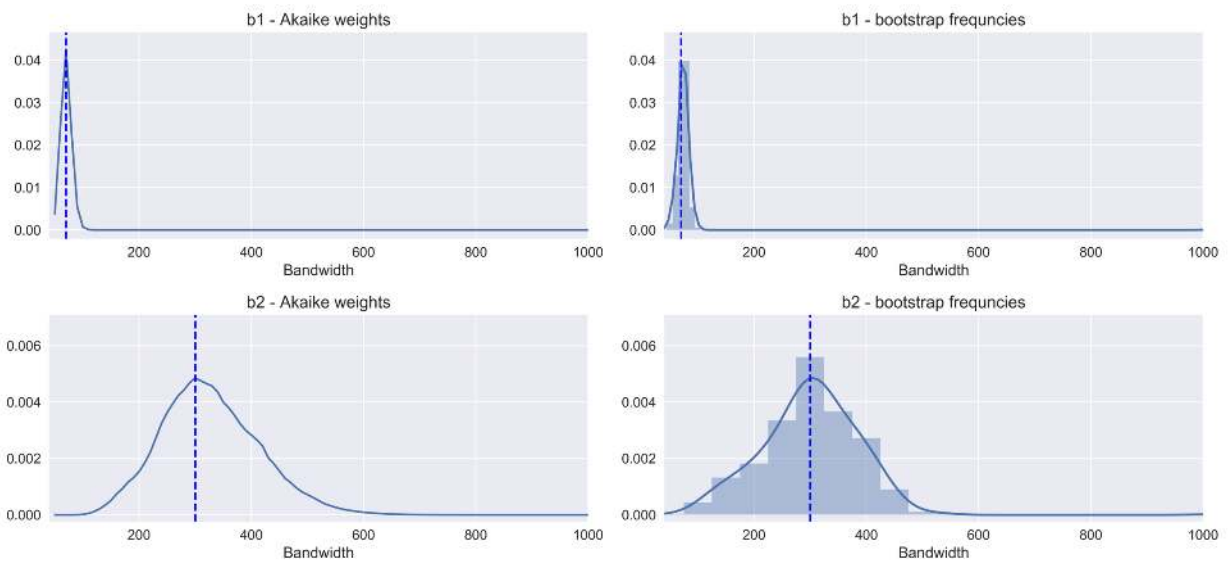


Figure 6. Akaike weights and bootstrap frequencies for different bandwidths.

The Akaike weight-based 95% bandwidth CIs can be compared with the 95% CIs obtained from bootstrapping (Table 3). The 95% CI computed based on Akaike weights is [60, 90] for process  $\mathbf{b}_1$ , which is exactly the same as the 95% confidence interval obtained from bootstrapping. For process  $\mathbf{b}_2$ , the Akaike weights-based 95% CI is [160, 500] which is similar to the 95% confidence interval obtained from bootstrapping [130, 440]. Based on the Akaike weights-based bandwidth confidence interval, we also arrive at the same conclusion that processes  $\mathbf{b}_1$  and  $\mathbf{b}_2$  have significantly different optimal bandwidths at the 95% confidence level. This result is particularly robust, since Akaike weights and bootstrapping embed two distinct concepts of measuring model selection uncertainty. The former uses a data-based weight-of-evidence while 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, 2002).

[Table 3 about here]

Table 3. 95% confidence interval of bandwidth based on Akaike weight and bootstrapping.

PROCESS	OPTIMAL BANDWIDTH	95% AKAIKE WEIGHT CI	95% BOOTSTRAP CI
$\mathbf{b}_1$	70	[60, 90]	[60, 90]
$\mathbf{b}_2$	300	[160, 500]	[130, 440]

#### 4.4 Unconditional inference: accounting for bandwidth selection uncertainty in local parameter estimation

Following the multi-model inference approach of Burnham and Anderson (2002), 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 unconditional inference with MGWR inference. For each set of local parameters for the  $j^{\text{th}}$  covariate,  $\boldsymbol{\beta}_j$ , unconditional parameter estimates,  $\widehat{\boldsymbol{\beta}}_j$  can be computed based on Akaike weights:

$$\widehat{\boldsymbol{\beta}}_j = \sum_{r=1}^R w_{rj} \widehat{\boldsymbol{\beta}}_{rj} \quad (11)$$

where there are  $R$  candidate bandwidths being evaluated during the bandwidth selection;  $w_{rj}$  is the Akaike weight for bandwidth  $r$  and  $\widehat{\boldsymbol{\beta}}_{rj}$  is the set of local parameter estimates obtained when using bandwidth  $r$ . The variance of parameter estimates  $\boldsymbol{\beta}_j$  can be calculated using the following equation (Burnham and Anderson, 2002):

$$\widehat{var}(\widehat{\boldsymbol{\beta}}_j) = \left[ \sum w_{rj} \sqrt{\widehat{var}(\widehat{\boldsymbol{\beta}}_{rj}) + (\widehat{\boldsymbol{\beta}}_{rj} - \widehat{\boldsymbol{\beta}}_j)^2} \right]^2 \quad (12)$$

Note that there are two components in  $\widehat{var}(\widehat{\boldsymbol{\beta}}_j)$ : (i) the variance from parameter estimation,  $\widehat{var}(\widehat{\boldsymbol{\beta}}_{rj})$ , and (ii) the variance from bandwidth selection uncertainty,  $(\widehat{\boldsymbol{\beta}}_{rj} - \widehat{\boldsymbol{\beta}}_j)^2$ . Therefore, usually the variance of the unconditional parameter estimates will be greater than that of the conditional parameter estimates because the latter are based purely on a single set of covariate-specific bandwidths (i.e. deterministic) and therefore neglects bandwidth selection uncertainty (Symonds and Moussalli, 2010). However, it worth noting that this might not always be true in (M)GWR because using a smaller bandwidth will yield larger parameter variances than using a larger bandwidth, *ceteris paribus*. However, weighted parameter variance is expected to yield more accurate confidence intervals for the local parameter estimates 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 (Bozdogan, 1987; Akaike, 1989).



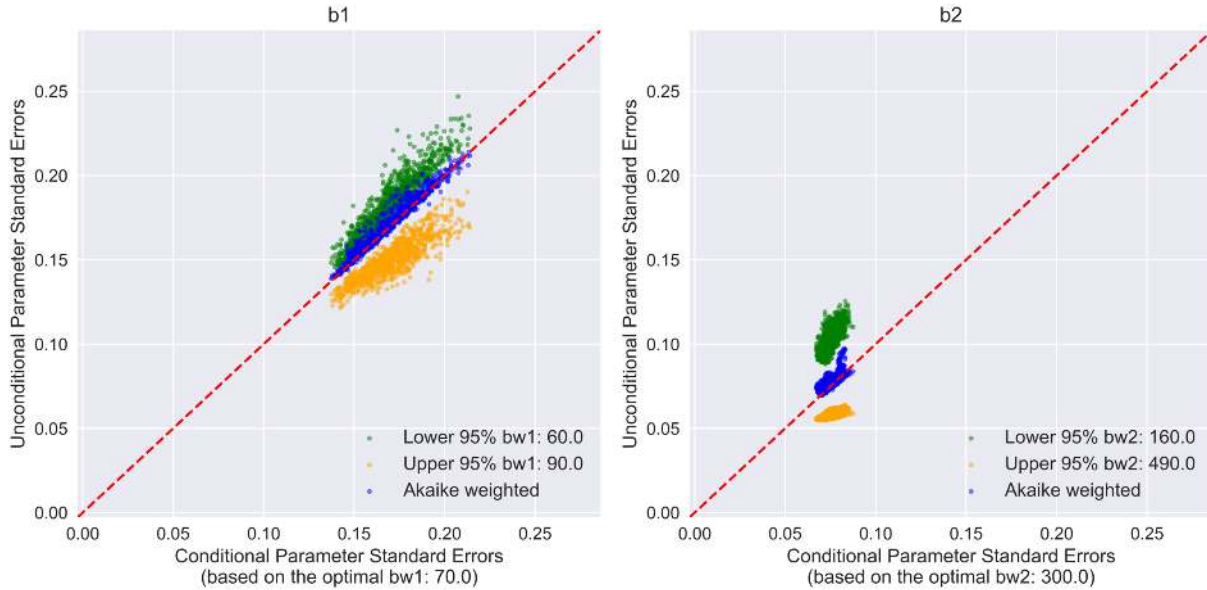


Figure 7. Comparison of conditional and unconditional standard errors of local parameter estimates for processes  $\mathbf{b}_1$  and  $\mathbf{b}_2$ .

In Figure 7 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  $\mathbf{b}_1$  and  $\mathbf{b}_2$ . 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: (i) using the lower bound of the 95% bandwidth CI (in green); (ii) using the upper bound of the 95% bandwidth CI (in orange); and (iii) using the Akaike-weighted standard errors (in blue). For process  $\mathbf{b}_1$  the upper and lower bounds of the 95% bandwidth confidence interval 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 tradeoff 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% 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  $\mathbf{b}_2$  the results are similar to those for  $\mathbf{b}_1$  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  $\mathbf{b}_2$  whereas it is between 60 and 90 nearest neighbors for  $\mathbf{b}_1$  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% CI.

[Figure 7 about here]

This is also seen in Figure 8 where the results of the unconditional and conditional local estimates of  $\mathbf{b}_1$  and  $\mathbf{b}_2$  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 above. We also represent the Akaike-weighted 95% CI (red) and the upper (green) and lower (orange) 95% CI using the bandwidth CI based on the cumulative Akaike weights. The Akaike weighted parameter estimate CI is computed by  $\hat{\beta}_j \pm 1.96 \sqrt{\hat{var}(\hat{\beta}_j)}$ . All the local parameter 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% bandwidth CI will produce parameter estimates within the Akaike-weighted parameter CI. This helps in interpreting how bandwidth changes will impact 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% confidence interval of the optimal bandwidth does not have a significant impact on the local parameter estimates.

[Figure 8 about here]

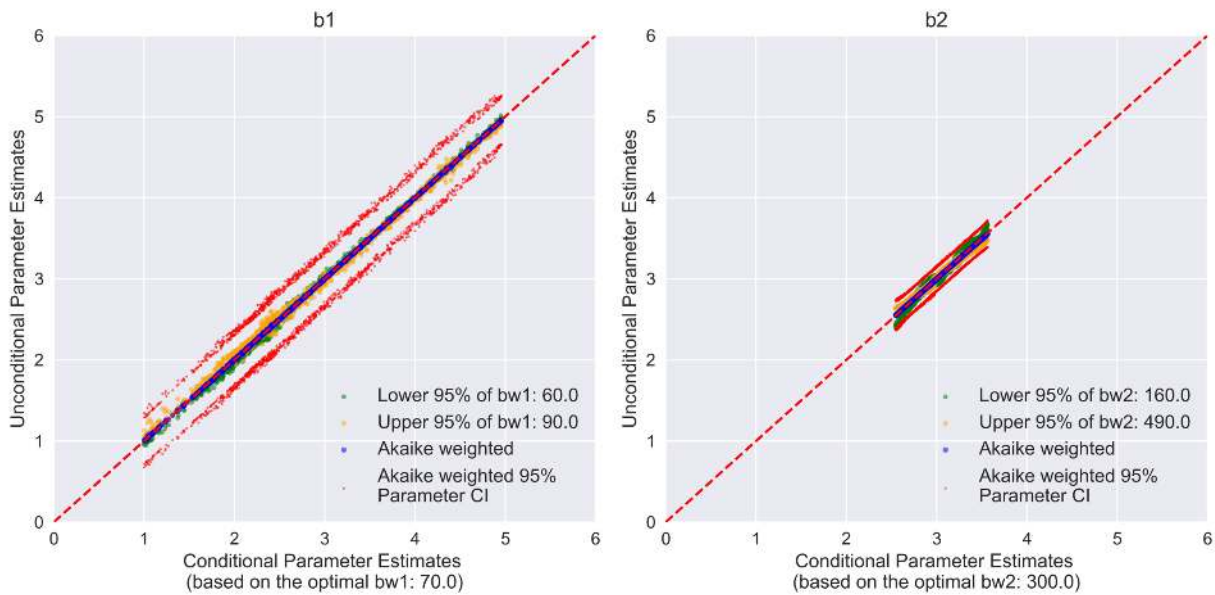


Figure 8: Comparison of conditional and unconditional local parameter estimates for processes  $\mathbf{b}_1$  and  $\mathbf{b}_2$

To examine whether unconditional local parameters which account for bandwidth uncertainties are more accurate than conditional parameters, we conduct a Monte Carlo simulation using the model described in Section 4.1 equation (9). The model was run for 1000 realizations using fixed covariates and randomly drawn errors. Within each realization we compute conditional and unconditional parameter estimates and confidence intervals and count how many times each type of confidence interval contains the true parameters  $\mathbf{b}_1$  and  $\mathbf{b}_2$ . A good 95% confidence interval should contain true parameters 95% of the time. Results indicate that the Akaike-weighted parameter estimate confidence intervals have marginally better coverage probabilities than the conditional confidence intervals (0.81 vs. 0.80 for  $\mathbf{b}_1$ ) and (0.89 vs. 0.87 for  $\mathbf{b}_2$ ) although 94.1% and 97.2% of total locations, respectively, have increased confidence interval coverage probabilities of more than 1%. Wolf et al. (2018) compare MGWR parameter standard errors with those obtained by Bayesian Spatially Varying Coefficients model and find that the MGWR parameter confidence intervals are 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 confidence interval coverage probability is still less than the nominal 95% level, which reveals that there may exist other sources of parameter estimate uncertainty that are neglected.

## 5. 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 (age  $\geq 18$ ) defined as obese<sup>2</sup> (body mass index (BMI)  $\geq 30.0$  kg/m<sup>2</sup>) by The Centers for Disease Control and Prevention (CDC) in each of the 815 census tracts in the Phoenix metropolitan area<sup>3</sup> as the response variable, and the five most influential covariates that determine obesity rates identified by the study of Oshan et al. (2019c): percentage of population visiting to doctor for a routine check-up 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 the percentage of people with a college degree. The obesity and the annual checkup data are downloaded from the 2014 “500 Cities project” from the CDC<sup>4</sup>, and the socio-demographic covariates are retrieved from the American Community Survey (ACS) 2015 5-year estimates dataset<sup>5</sup>. The regression model is formulated as follows:

---

<sup>2</sup> <https://www.cdc.gov/500cities/definitions/unhealthy-behaviors.htm#OBESITY>

<sup>3</sup> Six sparsely populated tracts are removed in this example.

<sup>4</sup> <https://www.cdc.gov/500cities>

<sup>5</sup> <https://data.census.gov>

$$\begin{aligned}
Pct\_Obesity = \beta_0 + \beta_1 Pct\_Checkup + \beta_2 Median\_Age + \beta_3 Pct\_Hispanic \\
+ \beta_4 Pct\_SNAP + \beta_5 Pct\_College + \varepsilon
\end{aligned}
\tag{13}$$

where obesity percentage (*Pct\_Obesity*) is regressed on percentage of population undergoing yearly check-ups (*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*). 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 et al. (2017) so that bandwidths and local parameters are comparable and invariant to the scale of the data. The overall  $R^2$  value for this MGWR model is 0.93.

For each of the covariates and the intercept, we determine optimal bandwidths using AICc within a range of 40 to 815 nearest neighbors with a step size of 10. Figure 9 shows the bandwidth search history where the x-axis is the bandwidth being evaluated and the y-axis is the AICc value. Covariate-specific optimal bandwidths with minimum AICc 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 the *Intercept* and *Pct\_Checkup*, with small optimal bandwidths, AICc values increase dramatically with increasing bandwidths, which suggests the data generating processes in both cases are very local. For the covariates *Median\_Age*, and *Pct\_SNAP*, the AICc value decrease with increasing bandwidths suggesting the data generating processes for these relationships are global and 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.

**[Figure 9 about here]**

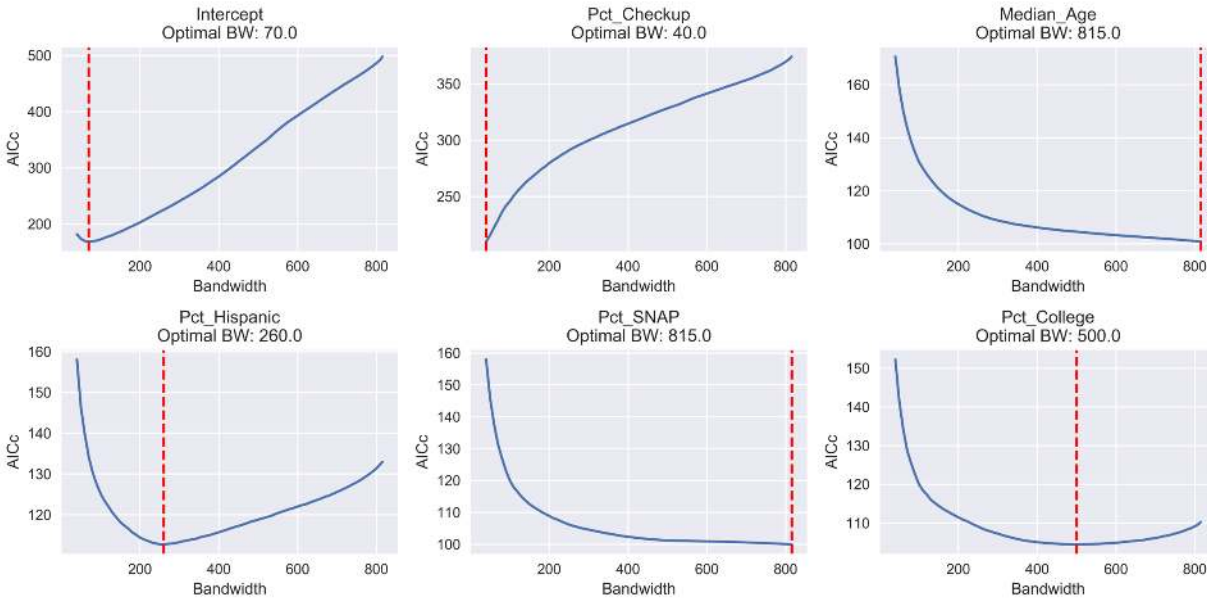


Figure 9. AICc values against bandwidth for each of the five covariates and the intercept. The red line is the optimal covariate-specific bandwidth based on the minimum AICc value.

Akaike weights and bandwidth CIs are computed for these five covariates plus the intercept as shown in Figure 10. In each case the red line shows the optimal bandwidth and the two green lines show the upper and lower bounds of the 95% 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 confidence intervals 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 the 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 that the relationships between obesity rates and *Pct\_Hispanic* and *Pct\_College* exhibit significant spatial heterogeneity but this heterogeneity is significantly less than for the *Intercept* and *Pct\_Checkup*.

[Figure 10 about here]

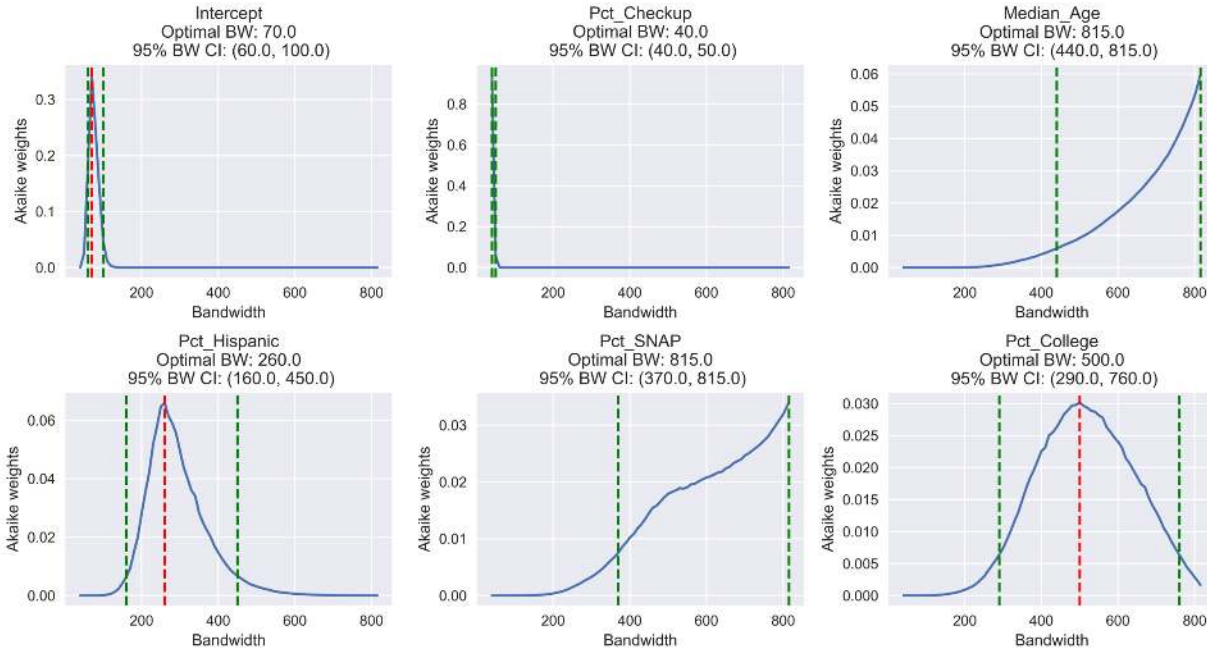


Figure 10. Akaike weights computed for different bandwidths for each set of local parameter estimates. The red line is the optimal bandwidth found by MGWR. The green lines depict the lower and upper bounds of the 95% confidence interval for each optimal bandwidth

[Table 4 about here]

Table 4. 95% confidence interval of bandwidth based on Akaike weight for each covariate.

COVARIATES	OPTIMAL BANDWIDTH	95% CONFIDENCE INTERVAL
Intercept	70	[60, 100]
Pct_Check_up	40	[40, 50]
Median_Age	815	[440, 815]
Pct_Hispanic	260	[160, 450]
Pct_SNAP	500	[370, 815]
Pct_College	815	[290, 760]

Maps of each of the four sets of local Akaike-weighted parameter estimates are shown in Figure 11. These are shown to be very similar to the conditional parameter estimates obtained from MGWR and comparisons

between the two sets of parameter estimates and standard errors (unconditional and conditional) can be found in Figures 12 and 13. Figure 11 shows the significant Akaike weighted local parameters for each covariate using the same color scheme. Insignificant parameters are masked out and shaded in grey, based on the critical t-values with multiple hypothesis testing adjustment (da Silva and Fotheringham, 2015; Yu et al., 2019). 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 the *Intercept* and *Pct\_Checkup* are visually much more heterogeneous than the local estimates associated with the other covariates which 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 indicate that here obesity rates are significantly higher than expected given the socio-economic conditions modeled in this part of the city. Conversely, in the northwest and southeast 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 to 0.35 (including insignificant estimates) and significant negative relationships between obesity and having regular check-ups 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.

**[Figure 11 about here]**



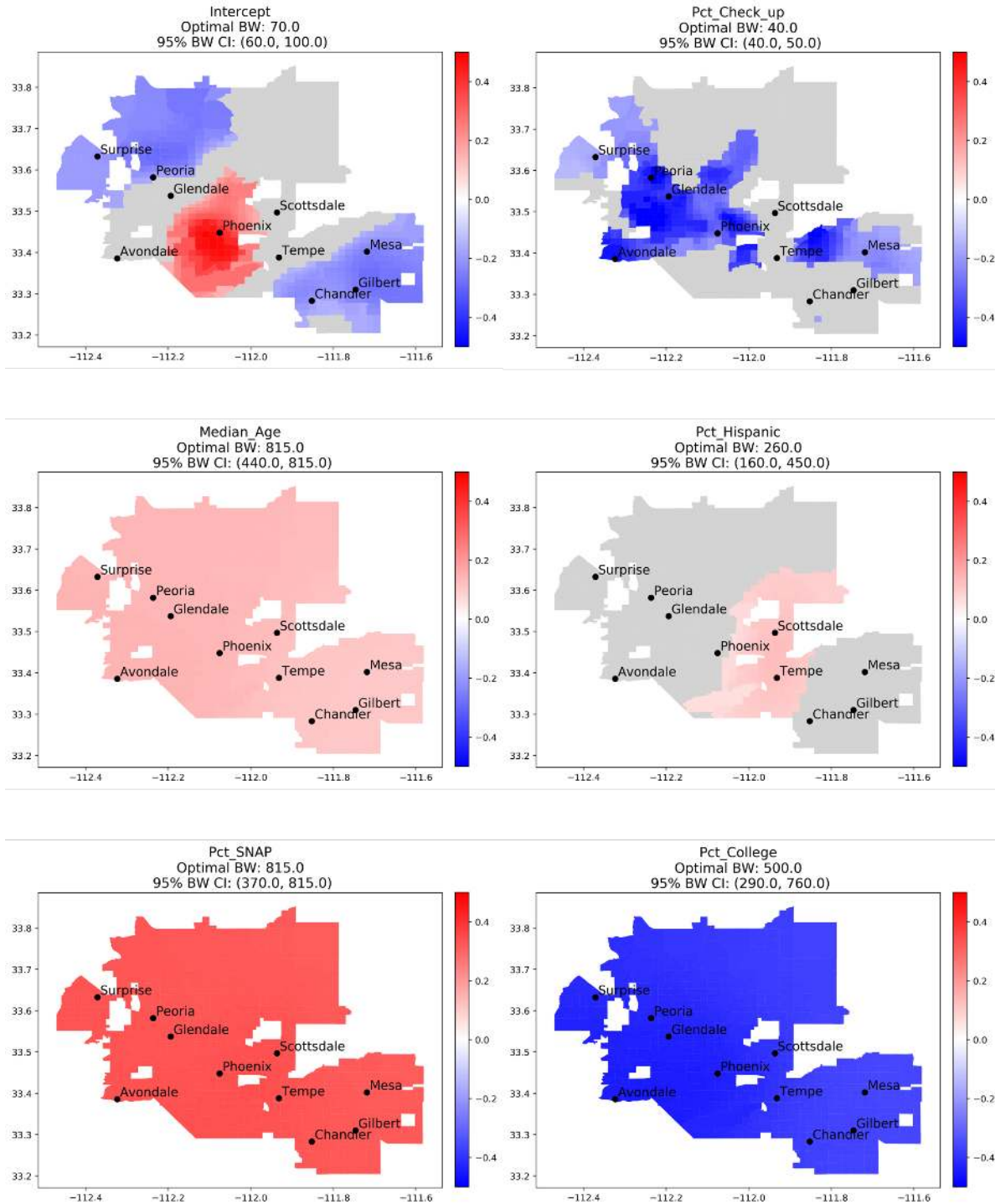


Figure 11. Maps of Akaike weighted local parameter estimates with insignificant parameters masked out and shaded in grey.



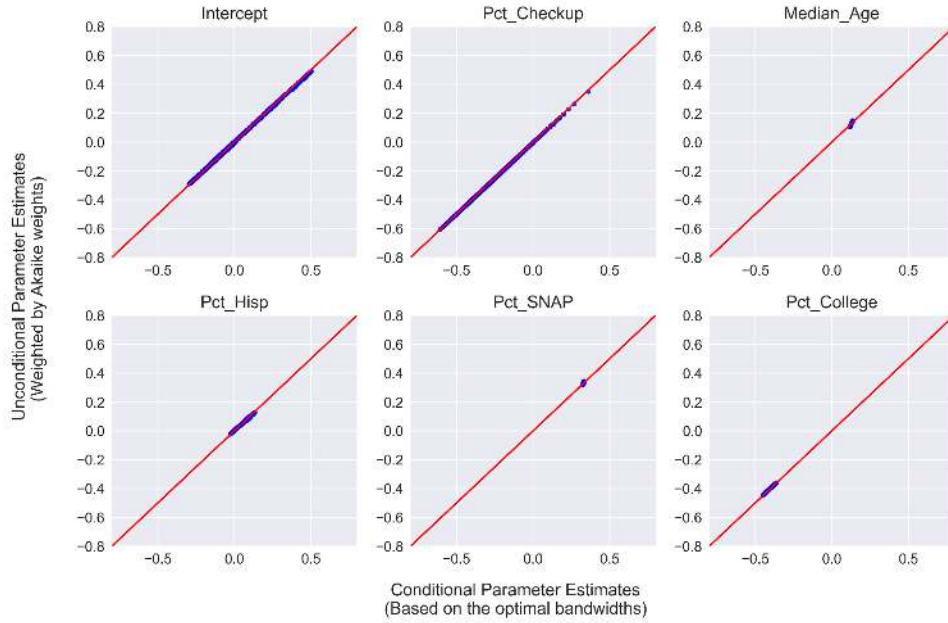


Figure 12. Comparison of conditional parameter estimates (based on the optimal bandwidths) with unconditional parameter estimates (calculated with Akaike weights).

[Figure 12 about here]

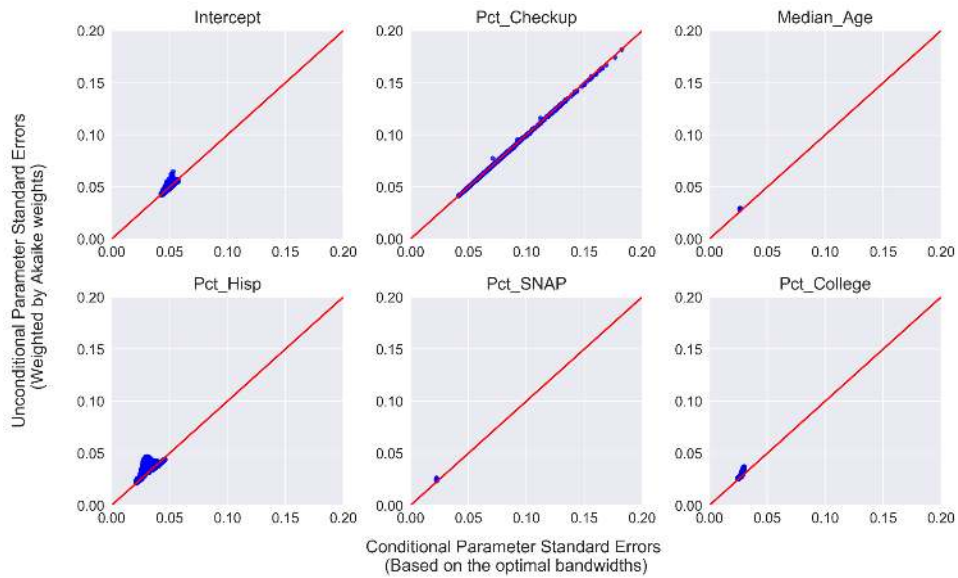


Figure 13. Comparison of conditional parameter standard errors (based on the optimal bandwidths) with unconditional parameter standard errors (calculated with Akaike weights).

[Figure 13 about here]

## 6. Conclusion and future work

Bandwidths in geographically weighted regression (GWR) are essentially related to the scale over which the underlying spatial processes operates. The recent advancement of multi-scale geographically weighted regression (MGWR) allows covariate-specific optimal bandwidths to be determined, allowing comparisons of the spatial scales over which different processes operate to be made (Fotheringham et al., 2017). However, bandwidth selection based on optimizing a goodness-of-fit criterion such as AICc contains an intrinsic uncertainty associated with it because bandwidth uncertainty is not accounted for. In this paper, 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 this paper, we examine the use of Akaike weights, an information-theory statistic, to measure covariate-specific bandwidth uncertainty and to obtain bandwidth confidence intervals. Akaike weights quantify the probability that a given bandwidth is optimal, which is a natural extension from AIC-based model selection (Burnham and Anderson, 2002). For example, a bandwidth with an Akaike weight of 0.6 has probability of 0.6 of being selected as the optimal bandwidth given the data. Based on both simulated data and an empirical example of modelling Phoenix obesity rates, we find that Akaike weights-derived bandwidth confidence intervals 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 confidence intervals although the improvements are marginal in the examples used here. The computation of Akaike weights is implemented in the *mgwr* python package (Oshan et al., 2019a) to increase accessibility of the methodology introduced in this paper.

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 geographically weighted regression framework, bandwidth selection is essentially a model selection problem. In this paper, 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 pre-selected bandwidth nor a set of pre-selected variables. These tasks and the work presented here provide great potential to enhance the quantification of process heterogeneity and scale.

## References

- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370.
- Buja, A., Hastie, T., & Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, 17(2), 453-510.
- Burnham, K. P., & Anderson, D. R. (1998). Practical use of the information-theoretic approach. In *Model Selection and Inference*. Springer, New York, NY.
- Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral ecology and sociobiology*, 65(1), 23-35.
- da Silva, A. R., & Fotheringham, A. S. (2016). The multiple testing issue in geographically weighted regression. *Geographical Analysis*, 48(3), 233-247.
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.
- Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 107(6), 1247-1265.
- Gelfand, A. E., Kim, H. J., Sirmans, C. F., & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462), 387-396.
- Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: challenges and solutions. *Journal of evolutionary biology*, 24(4), 699-711.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models* (Vol. 43). CRC Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*, Springer Series in Statistics.
- Harris, P., Brunsdon, C., Lu, B., Nakaya, T., & Charlton, M. (2017). Introducing bootstrap methods to investigate coefficient non-stationarity in spatial regression models. *Spatial Statistics*, 21, 241-261.
- Harris, P. (2019). A simulation study on specifying a regression model for spatial data: choosing between autocorrelation and heterogeneity effects. *Geographical Analysis*, 51(2), 151-181.
- Hurvich, C. M., & Tsai, C. L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *Journal of time series analysis*, 14(3), 271-279.
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in ecology & evolution*, 19(2), 101-108.

- Koh, L. P. (2008). Can oil palm plantations be made more hospitable for forest butterflies and birds?. *Journal of Applied Ecology*, 45(4), 1002-1009.
- Li, Z., & Fotheringham A.S. (2019). Computational Improvements to multiscale geographically weighted regression. doi:osf.io/dm5s4.
- Li, Z., Fotheringham, A. S., Li, W., & Oshan, T. (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), 155-175.
- Mei, C. L., Xu, M., & Wang, N. (2016). A bootstrap test for constant coefficients in geographically weighted regression models. *International Journal of Geographical Information Science*, 30(8), 1622-1643.
- Oshan, T. M., & Fotheringham, A. S. (2018). A comparison of spatially varying regression coefficient estimates using geographically weighted and spatial-filter-based techniques. *Geographical Analysis*, 50(1), 53-75.
- Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019a). 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.
- Oshan, T., Wolf, L. J., Fotheringham, A. S., Kang, W., Li, Z., & Yu, H. (2019b). A comment on geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, 33(7), 1289-1299.
- Oshan, T. M., Smith, J., & Fotheringham, A. S. (2019c). A multiscale spatial analysis of obesity determinants in Phoenix, Arizona. <https://doi.org/10.31219/osf.io/unfwj>
- Posada, D., & Buckley, T. R. (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.
- Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., & Levin, S. A. (2013). Marine taxa track local climate velocities. *Science*, 341(6151), 1239-1242.
- Symonds, M. R., & Moussalli, A. (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.
- Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review*, 11(1), 192-196.
- Wolf, L. J., Oshan, T. M., & Fotheringham, A. S. (2018). Single and multiscale models of process spatial heterogeneity. *Geographical Analysis*, 50(3), 223-246.
- Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), 1548-1563.

- Wheeler, D. C., & Calder, C. A. (2007). An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems*, 9(2), 145-166.
- Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., Kang, W., & Wolf, L. J. (2019). Inference in multiscale geographically weighted regression. *Geographical Analysis*. doi:10.1111/gean.12189
- Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., & Wolf, L. J. (2019). On bias of geographically weighted regression models. DOI:10.31219/osf.io/etb42.