

Multiscale Geographically Weighted Regression

Computation, Inference, and Application

by

Ziqi Li

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2020 by the
Graduate Supervisory Committee:

A. Stewart Fotheringham, Chair
Michael F. Goodchild
Wenwen Li

ARIZONA STATE UNIVERSITY

August 2020

ABSTRACT

Geographically Weighted Regression (GWR) has been broadly used in various fields to model spatially non-stationary relationships. Classic GWR is considered as a single-scale model that is based on one bandwidth parameter which controls the amount of distance-decay in weighting neighboring data around each location. The single bandwidth in GWR assumes that processes (relationships between the response variable and the predictor variables) all operate at the same scale. However, this posits a limitation in modeling potentially multi-scale processes which are more often seen in the real world. For example, the measured ambient temperature of a location is affected by the built environment, regional weather and 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. This results in each parameter surface being allowed to have a different degree of spatial variation, reflecting variation across covariate-specific processes. In this way, MGWR has the capability to differentiate local, regional and global processes by using varying bandwidths for covariates. Additionally, bandwidths in MGWR become explicit indicators of the scale at various processes operate. The proposed dissertation covers three perspectives centering on MGWR: computation; inference; and application. The first component focuses on addressing computational issues in MGWR to allow MGWR models to be calibrated more efficiently and to be applied on large datasets. The second component aims to statistically differentiate the spatial scales at which different processes operate by quantifying the uncertainty associated with each bandwidth obtained from MGWR. In the third component, an empirical study will be conducted to model the changing relationships between county-level socio-economic factors and voter preferences in the 2008-2016 United States presidential elections using MGWR.

DEDICATION

This dissertation is dedicated to my dearly loved wife, Liling Chang who has been giving me constant support and encouragement. I also dedicate this dissertation to my beloved newborn son Charlie Li. I hope you become a strong and happy boy. Last but not least, I would like to dedicate this dissertation to my parents, Jinhui Li and Lingjuan Niu, for their unconditional love and support.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to Dr. A. Stewart Fotheringham for all the guidance and help during my years of PhD study. You have always been encouraging and supportive in and beyond academia, and I am extremely grateful and fortunate to have you as my mentor. You really inspire me to pursue an academic career.

I would also like to extend my gratitude to the members of my dissertation committee: Dr. Michael F. Goodchild and Dr. Wenwen Li for generously offering their time and guidance in my dissertation research as well as for providing me with advice and support for academic career developments.

I must also thank my colleges and friends in the School of Geographical Sciences and Urban Planning who have been constantly supporting me during this adventure. It is a great pleasure to meet and work with you all.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES.....	vii
 CHAPTER	
1 INTRODUCTION	1
Background.....	1
Structure of the Dissertation.....	5
2 COMPUTATIONAL IMPROVEMENTS TO MULTISCALE GEOGRAPHICALLY WEIGHTED REGRESSION	8
Introduction.....	9
MGWR Basics.....	12
Computational Cost of MGWR	18
Solutions to Address the Computational Issues in MGWR	20
Testing Dataset and Environment	25
Results.....	27
Conclusions	35
3 MEASURING BANDWIDTH UNCERTAINTY IN MULTISCALE GEOGRAPHICALLY WEIGHTED REGRESSION USING AKAIKE WEIGHTS..	42
Introduction.....	43
Overview of the Procedure for Measuring Bandwidth Uncertainty.....	46
Background of MGWR Related to Akaike Weights	47
Investigating Bandwidth Uncertainty in a Controlled Experiment	50
An Empirical Example of Obesity Modeling in Phoenix.....	63
Conclusions and Future Work.....	72

CHAPTER	Page
4	THE DYNAMICS OF VOTER PREFERENCE INFLUENCES ACROSS THREE U.S. PRESIDENTIAL ELECTIONS, 2008, 2012 AND 2016: A MULTISCALE GEOGRAPHICALLY WEIGHTED REGRESSION ANALYSIS 80
	Introduction..... 81
	Data Used..... 84
	Methods 86
	Results..... 88
	Conclusions..... 119
5	CONCLUSIONS 127
	REFERENCES 133

LIST OF TABLES

Table	Page
2.1. Computational Complexity of MGWR Calibration	19
2.2. Runtime and Memory Comparison Between MGWR 2.0 and GWmodel	29
3.1. Summary Statistics for the Optimal Bandwidths Obtained from the Bootstrap Samples.....	54
3.2. An Example of Obtaining the 95% Confidence Set of Bandwidth for Process b_1 Using Akaike Weights	57
3.3. 95% Confidence Interval of Bandwidth Based on Akaike Weight and Bootstrapping	58
3.4. 95% Confidence Interval of Bandwidth Based on Akaike Weight for Each Covariate.....	67
3.5. Summary of OLS Parameter Estimates and MGWR Akaike-Weighted Parameter Estimates.....	67
4.1. List of Independent Variables Used in This Study.....	86
4.2. OLS Regression Coefficients for 2008, 2012, and 2016 Elections.....	90
4.3. Comparison of Model Performance Between OLS and MGWR.....	93
4.4. Optimized Covariate-Specific Bandwidth and 95% Confidence Interval in MGWR.....	95
4.5. Voting Statistics of Counties in Pennsylvania, Wisconsin, and Michigan by Group.....	113
4.6. Democratic Vote Share in the Two-party Race in the 2016 Presidential Election Under Two Hypothetical Scenarios	119

LIST OF FIGURES

Figure	Page
2.1. Illustration of Local, Regional and Global Spatial Processes in GWR and MGWR	10
2.2. Theoretical Memory Demand of MGWR Calibration with Varying Numbers of Observations and Covariates	20
2.3. Illustration of the Block Update Covariate-specific Matrix by Chunks (e.g. the Number of Chunks = 2)	24
2.4. Synthesized Known Parameter Surfaces with Varying Scales from Local to Global.	26
2.5. Runtime Comparison Based on Two Example Datasets from Fotheringham et al. (2017)	28
2.6. Memory Demand and Runtime Change with Varying Numbers of Chunks Used	30
2.7. Runtime and Speed-up Factor of the New MGWR Calibration for Parameter Estimation, Inference and Total Calibration on a 4-processor Machine Based on a Sample Size of 10,000	32
2.8. Speed-up Factor of the New MGWR Calibration for Parameter Estimation, Inference and Total Calibration on an 8-processor Machine Based on a Sample Size of 10,000	33
2.9. Approximate Runtime Needed to Complete an MGWR Calibration (with Parameter Estimation and Inference) under Three Different Convergence Rates and for Varying Numbers of Both Observations and Covariates.	34
2.S1. Synthesized Known Parameter Surfaces with Varying Scales from Local to Global	41
2.S2. Recovered Parameter Estimates Surfaces from MGWR Using the Final Testing Dataset.	41
3.1. Generated Spatial Processes with Low- and High- Heterogeneity.	51

Figure	Page
3.2. Recovered Parameter Surfaces Using MGWR with Optimal Bandwidths 70 and 300 for the High- and Low- Heterogeneity Processes, Respectively.	52
3.3. Optimal Bandwidth Distributions of Two Processes in Figure 3.1 under Bootstrapping (Smooth Lines are Plotted Based on Kernel Density Estimation)	54
3.4. Akaike Weights and Bootstrap Frequencies for Different Bandwidths (Smooth Lines are Plotted Based on Kernel Density Estimation)	57
3.5. Comparison of Conditional and Unconditional Standard Errors of Local Parameter Estimates for Processes $\mathbf{b}_1$ and $\mathbf{b}_2$	61
3.6. Comparison of Conditional and Unconditional Local Parameter Estimates for Processes $\mathbf{b}_1$ and $\mathbf{b}_2$	62
3.7. 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	65
3.8. 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.	66
3.9. Maps of Akaike Weighted Local Parameter Estimates with Insignificant Parameters Masked Out and Shaded in Grey	70
3.10. Comparison of Conditional Parameter Estimates (Based on the Optimal Bandwidths) with Unconditional Parameter Estimates (Calculated with Akaike Weights)	71
3.11. Comparison of Conditional Parameter Standard Errors (Based on the Optimal Bandwidths) with Unconditional Parameter Standard Errors (Calculated with Akaike Weights)	71

Figure	Page
3.A1. Bootstrapping Bandwidth Distributions of the Model Described in Equation (10) with a Distance-Based Fixed Kernel	77
3.A2. Akaike Weights for Two Processes Shown in Figure 3.1 with a Distance-Based Fixed Kernel	77
3.B1. Simulated Processes	78
3.B2. Optimal Bandwidth Distributions for Simulated Processes in Figure 3.B1 Under Bootstrapping	78
3.C1. Four Generated Spatial Processes with the Same Degree of Heterogeneity	79
3.C2. Optimal Bandwidth Distributions of Four Same Heterogeneity Processes in Figure 3.C1 under Bootstrapping	79
4.1. Observed County-Level Percentage of Vote for The Democratic Party in the Straight Fight against the Republican Party for 2008, 2012 and 2016.	82
4.2. Maps of OLS Residuals (left) and MGWR Residuals (right) for 2008, 2012 and 2016. .	95
4.3. Local Parameter Estimates of Median Income	97
4.4. Local Parameter Estimates of Sex Ratio.....	98
4.5. Local Parameter Estimates of Percentage of Population Employed in the Manufacturing Sector.....	99
4.6. Local Parameter Estimates of Percentage of Foreign Born Population.....	100
4.7. Local Parameter Estimates of the Gini Index	101
4.8. Local Parameter Estimates of the Percentage of Hispanic Population	102
4.9. Local Parameter Estimates of the Percentage of Population Aged Between 18 and 29	103
4.10. Local Parameter Estimates of the Percentage of Population Aged Above 65	104
4.11. Local Parameter Estimates of the Turnout Rate	105

Figure	Page
4.12. Local Parameter Estimates of the Percentage of the Third-party Vote.....	106
4.13. Local Parameter Estimates of the Percentage of People with Bachelor’s Degree or Higher.....	107
4.14. Local Parameter Estimates of the Log of Population Density	108
4.15. Local Parameter Estimates of the Percentage of People with Health Insurance	109
4.16. Local Parameter Estimates of the Percentages of Black Population.....	110
4.17. Local Parameter Estimates of the Intercept Obtained in MGWR.....	111
4.18. Map of Counties in Pennsylvania, Wisconsin and Michigan and Their Voted Party During 2008-2016 Period. “D-D-D” Denotes a County Voted for the Democratic Party in 2008, 2012 and 2016.	113
4.19. Socio-economic Changes During 2008-2016 for Each Group of Counties in PA-WI-MI	115
4.19. MGWR Local Parameter Estimates Changes During 2008-2016 for Each Group of Counties in PA-WI-MI.....	117

CHAPTER 1

INTRODUCTION

1.1 Background

The traditional approach of modeling spatial relationships is through global models assuming model parameters are constant over the space. For example, a global linear regression model produces a single regression equation from the data generated at different locations, assuming the data generating process is the same within the study area. Local models, on the other hand, allow model parameters to vary from location to location. Examples of local models include the earlier work of the expansion method (Casetti, 1972), Geographically Weighted Regression (Fotheringham et al., 2002), Bayesian Spatial Varying Coefficient models (Gelfand, 2003), and Eigenvector Spatial Filtering models (Griffith, 2008). Despite the varying specifications and fitting procedures of these different methods, they each produce parameter estimates specific to each sampled location for modeling potentially spatially varying relationships. The flexibility of estimating model parameters and the ability to examine local processes have led to local models receiving increasing attention in spatial analysis and they have been applied in a wide range of disciplines in the recent decades (Zhang et al., 2005; Shi et al., 2006; Cahill and Mulligan, 2007; Griffith and Chun, 2014; Zou et al., 2016; Xu et al., 2017).

Among the local models mentioned above, Geographically Weighted Regression (GWR) has probably received most attention because of its background in the well-known regression framework, its ease of interpretation and expandability, and the accessibility of popular software such as the Spatial Statistics Toolbox in ArcGIS, GWmodel library in the R environment, MGWR and its predecessor, GWR4. GWR fits a local regression model at each sampled location by borrowing the data from nearby to determine the local relationships. Data at the local regression point are given a weight of one, and data borrowed from nearby locations are given a

discounted weight less than one depending on the distance away from the focal regression point. The weighting is controlled by a kernel function, such as Gaussian, Bi-square, Tri-cube, among other potential functional forms that contain a continuous measure of distance-decay. The bandwidth, which is a parameter of the kernel function (e.g., standardization of the Gaussian function) is used to control the width of the kernel, and hence the amount of distance decay in each local regression. When the bandwidth is large, there are more data points included in the local fitting, thus producing local parameter estimates with low variance at the cost of high bias (under-fitting). When the bandwidth is small, this yields parameter estimates with low bias but high variance (over-fitting). Consequently, a large bandwidth produces a smoothed pattern of parameter estimates representing a spatial process with low or no variation over space. Using a small bandwidth produces a highly localized pattern of parameter estimates denoting a spatial process with a high degree of spatial heterogeneity.

Choosing the right bandwidth is critical in GWR because the model parameters are conditional on the selected bandwidth. The bandwidth selection follows a model selection procedure by examining the bias-variance tradeoff in the model. The overarching goal is to choose an optimal bandwidth which gives us a parsimonious model that best approximates the spatial processes that generated the data we observe. The Akaike Information Criterion (AIC) is a commonly used statistic that serves this purpose, which balances the goodness-of-fit of the model and the complexity of the model. GWR estimates a single bandwidth regardless of the number of covariates, assuming the amount of spatial variation in the different processes being modeled is the same. Consequently, in GWR local parameter estimates associated with different covariates exhibit the same degree of spatial heterogeneity. However, in real-world scenarios, data are often generated from spatial processes operating at different spatial scales. For example, air temperature measures may be determined by the local built environment, regional climates, and

global warming. Similarly, people's voting decisions are the consequence of a wide range of processes that operate at multiple scales from the household to the neighborhood to the country (Miller 1994; Forest, 2017). Thus, the use of single bandwidth in GWR, assuming the spatial outcome is a function of processes operating at the same spatial scale, seems overly rigid. To overcome this limitation, Fotheringham et al. (2017) developed Multiscale Geographically Weighted Regression (MGWR) which relaxes the single bandwidth assumption and allows different optimal bandwidths to be estimated for different covariates, thus reflecting different degrees of spatial heterogeneity in the associated processes. These multi-scale spatial processes can be captured by determining covariate-specific optimal bandwidths. Consequently, MGWR is a more accurate and flexible model than GWR, the latter being seen as a particular case of MGWR when the covariates-specific bandwidths are identical.

However, the increased flexibility of MGWR to model multi-scale spatial processes comes at additional computation costs. MGWR uses an iterative back-fitting algorithm following the additive model fitting procedure, which involves calibrating many GWR models repeatedly. The time complexity of MGWR algorithm is $O(kdN^2 \log N)$ where k is the number of covariates, d is the number of iterations before convergence is reached, and N is the number of observations in the model. As a result, this makes MGWR calibration challenging even for relatively small datasets. For example, the original implementation in Fotheringham et al. (2017) took more than two days for a model calibration with a dataset of around 2,000 locations. Considering model construction is a recursive trial and error procedure, this would inhibit the use of MGWR in empirical work. Moreover, the computation of the recently developed inferential statistics for MGWR by Yu et al. (2019) brings additional challenges both in time and memory, which limits the problem size that a standard MGWR calibration can handle to be smaller than 5,000 samples on an average computer. As large-scale fine-resolution spatial data are collected at increasing

speed from various sensors and sources ranging from remote sensing, GPS devices, to social media, such a computational bottleneck would place a severe limitation on the use of the MGWR model. Subsampling or grouping of spatial units may be an alternative workaround, but this compromises the quality and resolution of the outputs. Moreover, GWR and MGWR local estimates are more robust with large sample sizes (Páez et al., 2011). Therefore, there is an urge to develop a scalable MGWR implementation to enable large-scale and fine-resolution applications.

Another issue that needs to be addressed in MGWR is the lack of inference for the covariate-specific bandwidths which are obtained in the calibration of the model. 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. To make an inference about bandwidths as indicators of the relative spatial scales of different processes, however, it is crude to simply compare the covariate-specific bandwidths that are obtained based on the single set of observed samples and draw a conclusion that one process is more local or global than another. The covariate-specific bandwidths are subject to the sampling variation of the data and it is possible that any observed differences arise purely by chance. Therefore, understanding the uncertainty of the reported bandwidths is vital as it enables us to assess the significance of any difference in spatial scale across processes.

Furthermore, considering MGWR is a new model, more work is needed further to test the robustness of MGWR results for both bandwidth measures and local parameter estimates. So far, there have been a few published articles using MGWR in applied studies. Shabrina et al. (2019) studied the relationship between the food industry and the local variations of temporary accommodation in London. Fotheringham et al. (2019) evaluated the influences of air pollution in

Chinese cities. Yang et al. (2019) proposed a variant of MGWR for downscaling satellite-derived land surface temperature (LST). Cupido et al. (2019) investigated the county-level spatial variability in the determinants of mortality rates in the United States. Oshan et al. (2020) studied the spatial context of obesity determinants in the city of Phoenix, Arizona. Liu et al. (2020) analyzed the effects of socioeconomic factors of urbanization on various landscape patterns for Shenzhen, China. Notably, a recent paper by Fotheringham et al. (2020) applied MGWR to the 2016 presidential election data to examine the spatial variations in the influences of county-level socioeconomic factors on voter preference. The MGWR model achieves very impressive performance in terms of replicating voter preferences across US counties and the results show significant spatial variations in the determinants of voter preferences which occur over different spatial scales. Of particular note is the determination, through the local intercept estimates in MGWR, of a measure of the impact of geographical context on voting behavior. However, none of these studies has examined the temporally stability of the spatial variations in the reported local parameter estimates. An important validation of local models is the stability of their results over time. Here I extend the work of Fotheringham et al (2020) to cover the last three US Presidential elections to examine whether spatial variations in the determinants of voter preferences are temporally stable.

1.2 Structure of the dissertation

The scope of this dissertation hence focuses on the MGWR model addressing the issues or research directions mentioned above. The structure of the dissertation follows an integrated style with three separate studies constituting Chapters 2, 3, and 4, which focus on computation, inference, and application, respectively.

To be more specific, Chapter 2 develops a scalable MGWR algorithm to address the computational issues in MGWR. This development uses a parallelization technique following the

work of Li et al. (2019) as well as a chunk-wise partitioning method to remove the time and memory limitations in the MGWR calibration procedure. Consequently, the new MGWR software is hundreds of times faster than the original version and allows MGWR to be applied to moderate-to-large datasets with up to 100,000 observations. These improvements are integrated into the *mgwr* python package and the MGWR 2.1 software, both of which are freely available to download. This chapter is published in the International Journal of Geographical Information Science with co-author A. Stewart Fotheringham (Li and Fotheringham, 2020).

Chapter 3 discusses the bandwidth uncertainty under the (M)GWR framework and proposes methods to measure and account for this uncertainty. Based on simulation studies and an empirical example of obesity rates in Phoenix, the uncertainty of the bandwidths obtained in MGWR can be quantitatively measured by Akaike weights. Confidence intervals for the bandwidths can then be obtained in a computationally efficient manner. These bandwidth confidence intervals provide essential information about the scales over which different processes operate, especially when comparing covariate-specific bandwidths. Moreover, the unconditional inference can be computed following a model averaging approach based on Akaike weights to account for bandwidth uncertainty. This chapter is published in the Annals of the American Association of Geographers with co-authors A. Stewart Fotheringham, Taylor M. Oshan, and Levi John Wolf (Li et al., 2020).

Chapter 4 applies the MGWR model and methods developed in Chapters 2 and 3 to a political geography application examining the determinants of voter preferences across three U.S. Presidential Elections, 2008, 2012, and 2016. This chapter examines the temporal stability of the spatially varying relationships between socioeconomic factors and voter preference for the three elections. Particular interest is given to investigate local determinants of voter preference in

crucial counties in Pennsylvania, Michigan, and Wisconsin, which were critical in influencing the result of the election in 2016.

This dissertation concludes in Chapter 5 with a summary of the main developments, findings, and directions for future research of MGWR.

CHAPTER 2

COMPUTATIONAL IMPROVEMENTS TO MULTISCALE GEOGRAPHICALLY WEIGHTED REGRESSION

Abstract

Geographically Weighted Regression (GWR) has been broadly used in various fields to model spatially non-stationary relationships. Multiscale Geographically Weighted Regression (MGWR) is a recent advancement to the classic GWR model. MGWR is superior in capturing multi-scale processes over the traditional single-scale GWR model by using different bandwidths for each covariate. However, the multi-scale property of MGWR brings additional computation costs. The calibration process of MGWR involves iterative back-fitting under the additive model (AM) framework. Currently, MGWR can only be applied on small datasets within a tolerable time and is prohibitively time-consuming to run with moderately large datasets (greater than 5,000 observations). In this paper, we propose a parallel implementation that has crucial computational improvements to the MGWR calibration. This improved computational method reduces both memory footprint and runtime to allow MGWR modelling to be applied to moderate-to-large datasets (up to 100,000 observations). These improvements are integrated into the *mgwr* python package and the MGWR 2.0 software, both of which are freely available to download.

Keywords: geographically weighted regression, multiscale, local modelling, spatial analysis, parallel computing.

1 Introduction

In recent decades, spatially-varying coefficient (SVC) models have seen increased use in investigating potentially spatially non-stationary relationships. Geographically weighted regression (GWR) is one of the most popular and widely used SVC techniques (Fotheringham et al., 2002). It incorporates a data-borrowing mechanism at each location to obtain location-specific parameter estimates. Other alternative SVC frameworks include Bayesian spatially varying coefficient models (Gelfand et al., 2003; Finley 2011) and spatial eigenvector filtering (Griffith, 2007; Murakami and Griffith, 2015). Despite different model formulations and calibrations, all three methods are reported in the literature to yield similar local parameter estimates (Finley, 2011; Oshan and Fotheringham, 2017; Wolf et al., 2018). However, GWR has been the most widely adopted method due to its interpretability and accessibility. Various applications of GWR can be found in a wide range of fields, such as house price modelling (Bitter et al., 2007), atmospheric science (Hu et al., 2013), criminology (Troy et al., 2012), and public health (Comber et al., 2011).

Classic GWR is considered as a single-scale model that is based on one bandwidth parameter which controls the amount of distance-decay in weighting neighboring data around each location. The single bandwidth in GWR assumes that processes (relationships between the response variable and the predictor variables) all operate at the same scale. The optimal bandwidth in GWR can be considered as the best “average” bandwidth across all the covariate-specific processes. However, this posits a limitation in modeling potentially multi-scale processes which are likely to occur in the real world. For example, the measured ambient temperature of a location is affected by the built environment, regional weather and 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 covariate-specific bandwidths to be

optimized (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. In this way, MGWR has the capability to differentiate local, regional and global processes by optimizing a different bandwidth for each covariate. Additionally, bandwidths in MGWR become explicit indicators of the scale at various processes operate. An illustration of local, regional and global spatial processes can be seen in Figure 2.1. In GWR, because only one bandwidth is used for all covariates, all processes are assumed to have similar spatial variation and GWR will over-smooth local process and under-smooth global processes. However, in MGWR, with a correctly specified bandwidth for each covariate, the local (global) processes can be accurately captured by small (large) bandwidths.

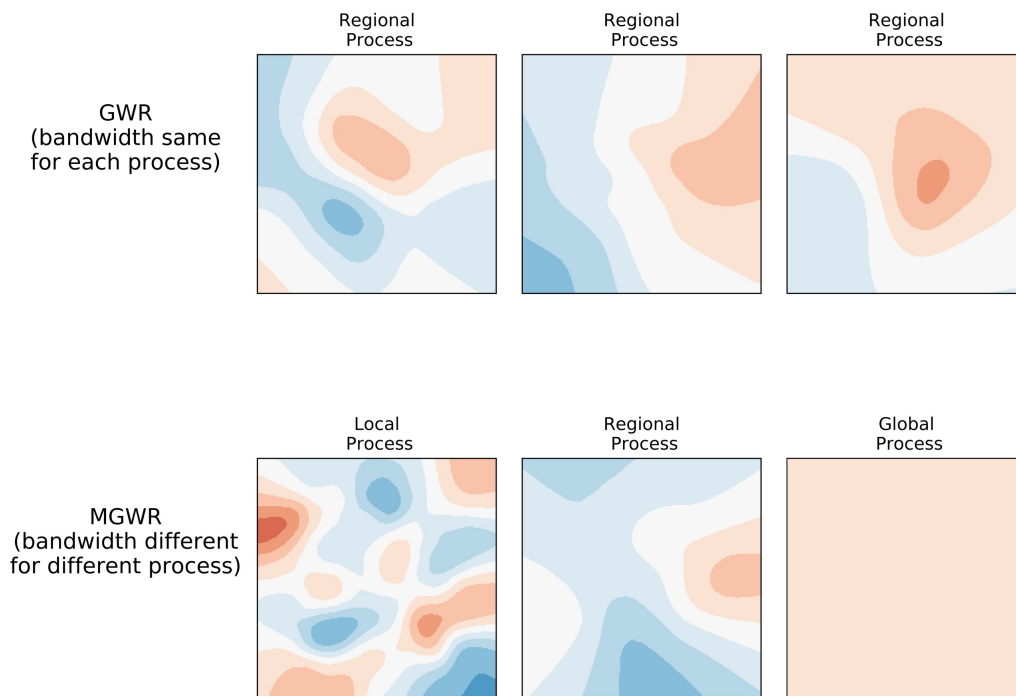


Figure 2.1. Illustration of local, regional and global spatial processes in GWR and MGWR.

The advance from a single scale local model (GWR) to a multiscale local model (MGWR) brings flexibility and better model fit at a cost of additional computation. The latter arises because multiscale models have more parameters to be estimated than single-scale models. MGWR, which is formulated as a Generalized Additive Model (Buja et al., 1989; Hastie and Tibshirani, 1990), utilizes a back-fitting algorithm for calibration which involves many univariate GWR calibrations. GWR calibration is not cheap in terms of computational cost since locally weighed regressions need to be calibrated at all locations. From the existing literature, studies using GWR often have fewer than 15,000 units of analysis (Li et al., 2019). Two major reasons why GWR has not generally been applied to large scale data are: (1) memory limitations of the hardware; and (2) inefficiency of the available software. Though there have been previous efforts to improve GWR calibration efficiency (Harris et al., 2010; Zhang, 2015; Tran et al., 2016), few of these have obtained drastic improvements or developed accessible operational software. Recently, however, a parallel GWR computational improvement algorithm called FastGWR has been developed to solve various computational issues in GWR calibration (Li et al., 2019). Based on an example dataset with 15,000 spatial locations, FastGWR is around 350 times faster than the widely-used R package *GWmodel* on a normal desktop without losing accuracy in parameter estimation. Additionally, the FastGWR parallel implementation addresses memory limitation and enables calibration of a GWR model with hundreds of thousands of locations within a reasonable time on a desktop machine. This brings new opportunities for geographically weighted regression models to be applied to large datasets. Because MGWR involves many GWR calibrations, this development can naturally benefit MGWR calibration by lowering runtime and memory.

In this study, our objective is to develop an efficient implementation of the MGWR model and to provide the spatial local modelling community with fast, scalable, reliable, and accessible software that can be used in applied studies. This paper is organized as follows. First,

we review the basics of the MGWR model which includes its back-fitting calibration process and statistical inference. Second, the computational challenges of MGWR are presented. Third, improvements to both the speed and memory of MGWR are described. Fourth, we propose a data-generating process to synthesize a dataset used for evaluating MGWR computation. Fifth, comparisons and evaluations of MGWR computation are presented. The paper concludes with possible future directions.

2 MGWR basics

2.1 MGWR formulation under GAM framework

A Generalized Additive Model (GAM) 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 variable, $\mathbf{f}_1(\mathbf{X}_1), \cdots, \mathbf{f}_k(\mathbf{X}_k)$ are additive components which are smooth functions of covariates, and $\boldsymbol{\varepsilon}$ is a column vector of i.i.d. error (Hastie and Tibshirani, 1990). The formulation of MGWR falls under this GAM framework (Fotheringham et al., 2017). The response variable $\mathbf{y}$ is the data observed over a spatial surface and $\mathbf{f}_1, \cdots, \mathbf{f}_k$ are spatial additive components estimated with covariate-specific bandwidths, so MGWR can be formulated as

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

and

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

where x_{ij} is j^{th} covariate $\mathbf{X}_j$ at location i , and β_{ij} is MGWR local parameter of the j^{th} covariate location i . The response variable can be viewed as the sum of multiple spatial layers. To estimate

local parameters β , we could re-arrange Equation (2) and knowing ε is independent of $\mathbf{X}_j$ gives the following component-wise conditional expectation

$$\mathbf{f}_{bw_j} = E\left(\mathbf{y} - \sum_{p \neq j} \mathbf{f}_{bw_p} - \varepsilon \mid \mathbf{X}_j\right) = E\left(\mathbf{y} - \sum_{p \neq j} \mathbf{f}_{bw_p} \mid \mathbf{X}_j\right) = \mathbf{A}_j(\mathbf{y} - \sum_{p \neq j} \mathbf{f}_{bw_p}) \quad (4)$$

where $\mathbf{A}_j$ is $E(\cdot \mid \mathbf{X}_j)$ which can be viewed as a projection (hat) matrix from a univariate GWR model of $\mathbf{X}_j$ that maps $\mathbf{y} - \sum_{p \neq j} \mathbf{f}_{bw_p}$ to $\hat{\mathbf{f}}_{bw_j}$. This GWR hat matrix $\mathbf{A}_j$ is expressed as

$$\mathbf{A}_j = \begin{pmatrix} x_{1j}(\mathbf{X}_j^T \mathbf{W}_1 \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{W}_1 & \dots \\ \vdots & \ddots \\ x_{nj}(\mathbf{X}_j^T \mathbf{W}_n \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{W}_n \end{pmatrix}_{n \times n} \quad (5)$$

where $\mathbf{W}_i$ is a diagonal spatial weight matrix computed with covariate-specific bandwidth and a kernel function (e.g. bi-square or Gaussian). Putting all the additive components in Equation (4) into a matrix form gives the following normal equations:

$$\begin{bmatrix} \mathbf{I} & \mathbf{A}_1 & \dots & \mathbf{A}_1 \\ \mathbf{A}_2 & \mathbf{I} & \dots & \mathbf{A}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_k & \mathbf{A}_k & \dots & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{f}_{bw_1} \\ \mathbf{f}_{bw_2} \\ \vdots \\ \mathbf{f}_{bw_k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_k \end{bmatrix} \mathbf{y} \quad (6)$$

Equation (6) can be written in abbreviated form as

$$\mathbf{P}\mathbf{f} = \mathbf{Q}\mathbf{y} \quad (7)$$

where $\mathbf{P}$ is an nk by nk matrix and $\mathbf{Q}$ matrix is an nk by n matrix. On rearranging, this gives additive components $\mathbf{f}_{bw_1}, \dots, \mathbf{f}_{bw_k}$ as

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_{bw_1} \\ \mathbf{f}_{bw_2} \\ \vdots \\ \mathbf{f}_{bw_k} \end{bmatrix} = \mathbf{P}^{-1} \mathbf{Q}\mathbf{y} \quad (8)$$

provided that matrix $\mathbf{P}$ is invertible. Then, covariate-specific hat matrices can be obtained by

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \vdots \\ \mathbf{R}_k \end{bmatrix} = \mathbf{P}^{-1} \mathbf{Q} \quad (9)$$

The normal equations in Equation (6) are not described in the MGWR literature but are useful as they provide a closed form of local parameter estimates and their variances as derived in MGWR. From a practical perspective, directly solving Equation (8) takes a time cost of $O(n^3 k^3)$ which is feasible for small datasets but is often prohibitive for even moderately large datasets. Another issue is that the solution from Equation (8) assumes that the smoothing matrices $\mathbf{A}_j$ are known *a priori*. In the context of MGWR, this means that the bandwidth of each covariate is known. However, this piece of information is usually unknown and MGWR model calibration relies on a data-driven bandwidth search routine to find an optimal fitted model that minimizes a goodness-of-fit criterion (e.g. AICc, cross-validation, etc.).

2.2 MGWR calibration

The solution to solve Equation (8) is to use the back-fitting process introduced in Buja et al. (1989) and explained in Hastie and Tibshirani (1990). Back-fitting estimators will converge to the solution in equation (8), provided that the inverse of matrix $\mathbf{P}$ exists. Fotheringham et al. (2017) incorporated this back-fitting algorithm into the calibration process for MGWR with automatic bandwidth searching. The detailed calibration process is as follows (see also Algorithm 1 below). Firstly, the parameter estimates at all locations are initialized from a GWR model containing the response variable and all covariates. Fitted additive terms $\hat{\mathbf{f}}_{1...k}$ are computed accordingly by element-wise multiplying parameter estimates with covariates. The current model residuals $\hat{\boldsymbol{\varepsilon}}$ can be computed by $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}}$. Then a univariate GWR model is calibrated using the residuals $\hat{\boldsymbol{\varepsilon}}$ plus the first additive term $\hat{\mathbf{f}}_1$ as a response variable and first covariate $\mathbf{X}_1$.

$$\text{Univariate GWR: } \hat{f}_1 + \hat{\varepsilon} \sim X_1 \quad (10)$$

The optimal bandwidth found in the univariate GWR model then temporarily become the optimal bandwidth bw_1 for the first covariate. With that bandwidth bw_1 , $\hat{f}_1$ and residuals $\hat{\varepsilon}$ are updated. This procedure is then repeated for covariate X_2 to update the second additive term $\hat{f}_2$ and residuals $\hat{\varepsilon}$. The process continues in this way through to the final covariate X_k when $\hat{f}_k$ and $\hat{\varepsilon}$ are updated. This finishes the first-round iteration and the second-round starts from the first additive term using the new values for the estimated $\hat{f}_k$ and $\hat{\varepsilon}$. The iterations continue until the change in a convergence indicator (such as the residual sum of squares) becomes sufficiently small between successive iterations. When the convergence threshold is met, parameter estimates and the final set of bandwidths $bw_{1...k}$ can be obtained from the last iteration of back-fitting.

Algorithm 1: MGWR Back-fitting Calibration

- 1: Calibrate a GWR model of $y \sim X$ and obtain fitted additive terms $\hat{f}_{1...k}$ and model residual $\hat{\varepsilon}$.
 - 2: Do until $\hat{f}_{1...k}$ converge:
 - 3: For each term j from 1 to k :

Calibrate univariate GWR model of $\hat{f}_j + \hat{\varepsilon} \sim X_j$ and get optimal bandwidth bw_j , new fitted term $\hat{f}_j^*$, and residual $\hat{\varepsilon}^*$.

Update $\hat{f}_j \leftarrow \hat{f}_j^*$ and $\hat{\varepsilon} \leftarrow \hat{\varepsilon}^*$.
 - 4: End for
 - 5: End do
-

2.3 MGWR inference

Yu et al. (2019) analytically derive the hat matrix for MGWR and the standard errors of the MGWR parameter estimates, which provides an inferential framework for MGWR under a

GAM framework. In this section, we will follow the notations used in Yu et al. (2019) to describe the computation of inference. The principal is to express each fitted component as follows

$$\hat{f}_j = \mathbf{R}_j \mathbf{y} \quad (11)$$

where $\mathbf{R}_j$ is an n by n covariate-specific hat matrix that maps response variable $\mathbf{y}$ to each fitted additive component $\hat{f}_j$. Then, the covariance of $\hat{f}_j$ can be obtained by $\mathbf{R}_j \mathbf{R}_j^T \sigma^2$ where σ^2 is the error variance of the MGWR model. The covariate-specific hat matrix $\mathbf{R}_j$ is computed within the back-fitting process after each univariate GWR by updating

$$\mathbf{R}_j \leftarrow \mathbf{A}_j \left(\mathbf{I} - \sum_{p \neq j}^k \mathbf{R}_p \right) \quad (12)$$

where $\mathbf{A}_j$ is the hat matrix of each univariate GWR model (Equation 4). Conveniently, let $\mathbf{I} - \sum_{p \neq j}^k \mathbf{R}_p$ be denoted as $\mathbf{R}_j^-$ so that Equation (12) becomes

$$\mathbf{R}_j \leftarrow \mathbf{A}_j \mathbf{R}_j^- \quad (13)$$

Then the MGWR hat matrix $\mathbf{S}$ that maps $\mathbf{y}$ onto $\hat{\mathbf{y}}$ can be obtained by summing the $\mathbf{R}$ matrices from the last iteration in the back-fitting once the convergence threshold is satisfied.

$$\mathbf{S} = \sum_1^k \mathbf{R}_j \quad (14)$$

With the $\mathbf{R}$ and $\mathbf{S}$ matrices computed, important model diagnostics can then be obtained as follows:

- 1) Covariate-specific Effective Number of Parameters (ENP):

$$ENP_j = tr(\mathbf{R}_j) \quad (15)$$

- 2) Model Effective Number of Parameters:

$$ENP_{model} = tr(\mathbf{S}) \quad (16)$$

- 3) Model AIC_c :

$$AIC_c = 2n \ln \left(\frac{RSS}{n} \right) + n \ln(2\pi) + n \left(\frac{n + tr(\mathbf{S})}{n - 2 - tr(\mathbf{S})} \right) \quad (17)$$

where RSS is the residual sum of squares.

- 4) The standard errors of the parameters can be obtained by:

$$SE(\hat{\beta}_j) = \sqrt{diag(\mathbf{C}\mathbf{C}^T \hat{\sigma}^2)} \quad (18)$$

where $\mathbf{C} = [diag(\mathbf{X}_j)]^{-1} \mathbf{R}_j$ and $[diag(\mathbf{X}_j)]^{-1}$ is the inverse of a diagonal matrix with $\mathbf{X}_j$ being its diagonal elements.

Detailed descriptions can be found in Yu et al. (2019).

2.4 Available software for calibrating MGWR model

Currently, there are two software packages that are available to calibrate an MGWR model, and both are actively maintained. One is *GWmodel* in the R environment (Gollini et al., 2015)¹ which provides an array of geographically weighted models. With the latest update in February 2019 (version 2.0-8), *GWmodel* is able to perform an MGWR calibration with inference based on the work of Fotheringham et al. (2017) and Yu et al. (2019). It also has the capability of calibrating Parameter-specific Distance Metrics (PSDM) models (Lu et al., 2017), which is a special use case of MGWR when different distance metrics are desired. A detailed comment on PSDM and MGWR can be found in Oshan et al. (2019). The second is MGWR 1.0 software

¹ The GWmodel R library is available at <https://cran.r-project.org/web/packages/GWmodel/>.

released in November, 2018. It is built on the *mgwr* python package² (Oshan et al., 2019) and has a user-friendly graphical interface and is available for both MacOS and Windows platforms. The computational improvements developed in this paper are integrated into the *mgwr* python package (version 2.1.0) and built into the MGWR 2.0 software³ which supersedes the previous MGWR 1.0 version.

3 Computational cost of MGWR

MGWR is computationally intensive due to the iterative back-fitting algorithm that involves many univariate GWR calibrations. The total number of univariate GWR calibrations needed is $k \times d$, where k is the number of covariates and d is the number of iterations before convergence is reached. Though d is hard to predict, there are three general rules of thumb from experiments which show that:

- 1) The greater the number of covariates, the more iterations are needed for convergence.
- 2) Higher levels of multicollinearity among covariates increase the number of iterations needed before convergence.
- 3) Local processes need more iterations to converge than regional or global processes.

Further analysis is needed to provide confirmatory evidence of these rules. In addition, standardization of both response and covariates is suggested before running the MGWR model (Fotheringham et al., 2017). This is not only for better comparison of parameter estimates and associated bandwidths, but it also speeds up convergence. Li et al. (2019) shows that for a multivariate GWR model with k covariates, the time complexity of the GWR algorithm is

² The *mgwr* python library is hosted at <https://github.com/pysal/mgwr>.

³ The MGWR 2.0 software is freely available at <https://sgsup.asu.edu/sparc/mgwr>.

$O(k^3 n^2 \log n)$ where the $\log(n)$ comes from golden-section search for the optimal bandwidth. Because MGWR only needs univariate GWR calibrations in its back-fitting, the time complexity of an MGWR calibration is approximately $O(kdn^2 \log n)$. The computational cost of the inference calculation is much more intensive because after each univariate GWR during the back-fitting, a multiplication of two n by n matrices $\mathbf{A}_j$ and $\mathbf{R}_j^-$ (Equation 13), is needed at a cost of $O(n^3)$ time complexity. In total, there would be $k \times d$ numbers of $O(n^3)$ operations needed. The time for updating $\mathbf{R}_j$ is negligible when n is relatively small (e.g. $n < 5,000$). However, as n increases, this operation does not scale well and will dominate runtime of the back-fitting more than the parameter estimation procedure. Further evidence of this statement will be shown in the results section. On the other hand, regarding the memory constraint, storing the k by n by n matrix $\mathbf{R}$ has a memory complexity of $O(kn^2)$. Table 2.1 summarizes the time and memory complexity of MGWR calibration for both back-fitting and inference. Figure 2.2 shows the theoretical memory allocation needed with varying n and k based on double-precision (64-bit for each number) matrices. Considering today's hardware, this will place a severe limitation in MGWR for calibrating a moderately large model ($>10,000$ locations). For example, a desktop computer with 16GB of available memory would fail to calibrate an MGWR model with more than 15,000 locations and with more than 5 covariates. What is more, in practice, because intermediate results are needed to be cached during the calibration, we often need more memory space than the theoretical memory.

Table 2.1. Computational complexity of MGWR calibration

Back-fitting		Inference	
Time	Memory	Time	Memory
$O(kdn^2 \log n)$	$O(kn)$	$O(kdn^3)$	$O(kn^2)$

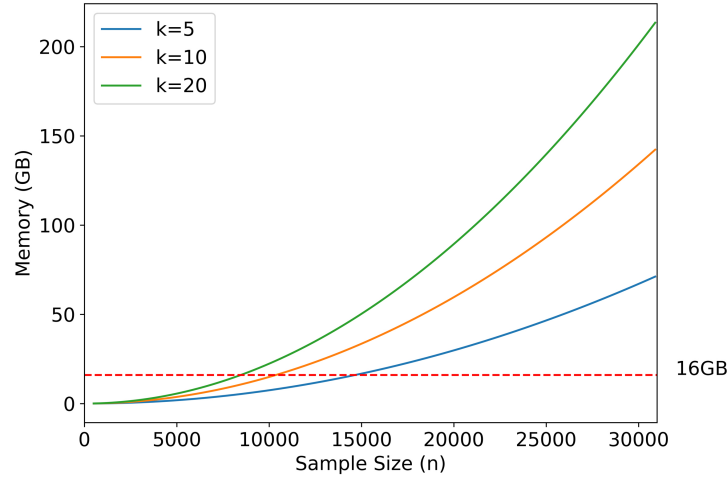


Figure 2.2. Theoretical memory demand of MGWR calibration with varying numbers of observations and covariates.

4 Solutions to address the computational issues in MGWR

MGWR calibration can be separated into two steps. The first is to estimate parameters and find the set of optimal bandwidths during a back-fitting procedure. The second is to compute MGWR inference to obtain parameter uncertainties, covariate-specific and overall model effective number of parameters (ENP), and model diagnostics (e.g. AICc). The first step of an MGWR calibration mainly involves univariate GWR calibrations; thus, here we can take advantage of the parallelization and optimization methods described in FastGWR (Li et al., 2019). For inference computation, we need to develop a novel parallel implementation that computes covariate-specific hat matrices $\mathbf{R}_j$ in column chunks to reduce the memory demand. We now describe these two developments.

4.1 MGWR back-fitting with FastGWR

Traditional GWR calibration consists of independent locally weighted least square regression at each location. Then local statistics are combined to compute model diagnostics such as AICc to assess model fit. This process is repeated for different bandwidths to find which bandwidth yields the best model fit using, for example, AICc. FastGWR utilizes the parallelizable nature of GWR and computes all local statistics in parallel without storing large matrices such as the weight matrix and hat matrix. Then AICc can be efficiently computed and compared across different bandwidths with the help of a golden-section search routine. The approach is described in detail in the context of MGWR. For a given bandwidth, there are two local statistics that are needed to compute AICc: the local influence value r_i (also the diagonal element of hat matrix) and the local residual $\hat{\varepsilon}_i$. Both can be computed as follows:

$$r_i = x_{ij}(\mathbf{X}_j^T \mathbf{W}_i \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{W}_i \quad (19)$$

$$\hat{\beta}_{ij}^* = \mathbf{X}_j(\mathbf{X}_j^T \mathbf{W}_i \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{W}_i (\hat{\mathbf{f}}_j + \hat{\boldsymbol{\varepsilon}}) \quad (20)$$

$$\hat{\varepsilon}_i^* = \hat{\beta}_{ij} x_{ij} - \hat{\beta}_{ij}^* x_{ij} + \hat{\varepsilon}_i \quad (21)$$

where x_{ij} is the measurement of the j^{th} covariate at location i , $\mathbf{X}_j$ is a column vector of the j^{th} covariate, $\mathbf{W}_i$ is a diagonal weight matrix (can be stored as a column vector to reduce memory) computed locally using the current bandwidth as a pre-specified kernel (e.g. bi-square or Gaussian), $\hat{\beta}_{ij}$ and $\hat{\varepsilon}_i$ are the parameter estimate and residual from the last iteration of back-fitting. The computation of three local statistics can be evenly assigned to multiple processors for parallel computing. Then once all processors finish their assigned tasks, covariate-specific AICc values can be computed as

$$AIC_c = 2n \ln \left(\frac{\sum_1^n \hat{\varepsilon}_i^*}{n} \right) + n \ln(2\pi) + n \left(\frac{n + \sum_1^n r_i}{n - 2 - \sum_1^n r_i} \right) \quad (21)$$

This procedure is repeated along with the golden section routine as detailed in Algorithm 2.

Algorithm 2: MGWR Back-fitting Calibration using parallel implementation from FastGWR

- 1: Calibrate a GWR model of $\mathbf{y} \sim \mathbf{X}$ and obtain fitted additive terms $\hat{\mathbf{f}}_{1\dots k}$ and model residual $\hat{\boldsymbol{\varepsilon}}$.
 - 2: Do until $\hat{\mathbf{f}}_{1\dots k}$ converge:
 - 3: For each term j from 1 to k :
 - 4: For bw in golden-section search:
 - 5: Parallel computing local statistics $\hat{\varepsilon}_i^*$, r_i , and $\hat{\beta}_{ij}^*$ for all locations i .
 - 6: Compute and assess covariate-specific $AICc$.
 - 7: Obtain the optimal bandwidth bw_j with minimum $AICc$ and update $\hat{\mathbf{f}}_j \leftarrow \hat{\boldsymbol{\beta}}_j^* \mathbf{X}_j$ and $\hat{\boldsymbol{\varepsilon}} \leftarrow \hat{\boldsymbol{\varepsilon}}^*$
 - 4: End for
 - 5: End do
-

4.2 Chunk-wise parallel computing of MGWR inference

From the previous discussion, it is known that the inference computation in MGWR requires a large memory allocation for the computation of the covariate-specific hat matrices $\mathbf{R}$ (dimension n by n by k), which can easily exceed the memory limit of a standard desktop computer; even for a moderately sized dataset. In this section, we propose a partitioning method that can compute the $\mathbf{R}$ matrix in small chunks of columns, each of which can easily fit into the memory of even a modest computer. From Equation (13) we can see that each covariate-specific hat matrix $\mathbf{R}_j$ is computed as $\mathbf{R}_j \leftarrow \mathbf{A}_j \mathbf{R}_j^-$ during back-fitting, where $\mathbf{A}_j$ is a univariate GWR hat matrix that projects $\mathbf{R}_j^-$ to $\mathbf{R}_j$ with the dimension n by n , and $\mathbf{R}_j^-$ is $\mathbf{I} - \sum_{p \neq j}^k \mathbf{R}_p$ of shape n by n . From the property of a projection matrix, it is known that updating each column of $\mathbf{R}_j$ is

independent of updating other columns of $\mathbf{R}_j$. This provides the opportunity to compute $\mathbf{R}_j$ by chunks of columns. Since both $\mathbf{A}_j$ and $\mathbf{R}_j^-$ are n by n matrices, in the case when both cannot fit into memory, we partition $\mathbf{A}_j$ by rows and $\mathbf{R}_j^-$ by columns to update a block of $\mathbf{R}_j$. The updating is illustrated in Figure 2.3 where we evenly split $\mathbf{A}_j$ and $\mathbf{R}_j^-$ into two chunks. This procedure is repeated in the back-fitting algorithm, and when the back-fitting converges we get a column chunk of $\mathbf{R}_j$, which can be denoted as $\mathbf{R}_{jc}$. Partial inference can be computed based on the chunk of $\mathbf{R}_{jc}$, such as the calculation of the effective number of parameters and the matrix $\mathbf{C}\mathbf{C}^T$ in Equation (18) to further compute parameter uncertainties. After that, we iterate to the second chunk of $\mathbf{R}_j$, and obtain the corresponding partial inference. Each chunk can be computed independently in parallel to fully utilize a multi-processor system. Once all chunks are computed, full MGWR inference can be combined using partial inference obtained from each chunk, and the formulas are given as follows:

$$ENP_{jc} = \sum_{c=1}^q \text{partial trace}(\mathbf{R}_{jc}) \quad (22)$$

$$\mathbf{C}\mathbf{C}_{jc}^T = \sum_{c=1}^q \left(\frac{\mathbf{R}_{jc}}{\mathbf{X}_j} \right)_{axis=1}^2 \quad (23)$$

where c is the current chunk and q is the total number of chunks and the subscript $axis=1$ indicates that the summation is applied to columns and reduces to a column vector. Details are shown in Algorithm 3.

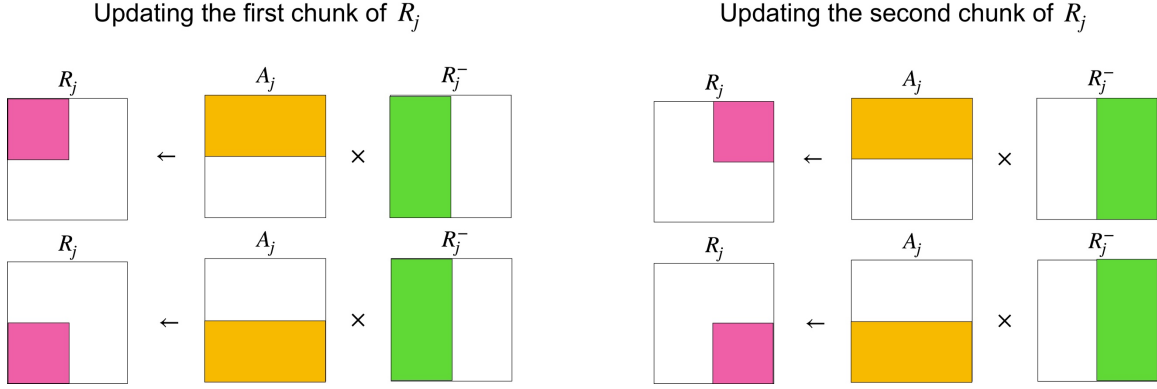


Figure 2.3. Illustration of the block update covariate-specific matrix $\mathbf{R}_j$ by chunks (e.g. the number of chunks = 2).

Algorithm 3: Chunk-wise parallel computing of MGWR inference

- 1: Initialize $\mathbf{R}_j$ from GWR, $ENP_j = 0$, and $\mathbf{C}\mathbf{C}_j^T = \mathbf{0}_{n \times 1}$
 - 2: Parallel computing each chunk c in the total number of q chunks
 - 3: Given bandwidths history in back-fitting iterations from parameter estimation
 - 4: For each term j from l to k :
 - 5: Block-update $\mathbf{R}_j$ and get a column chunk of $\mathbf{R}_j$
 - 6: Update column chunk of $\mathbf{R}_j^-$
 - 7: For each term j from l to k :
 - 8: $ENP_j = ENP_j + ENP_{jc}$
 - 9: $\mathbf{C}\mathbf{C}_j^T = \mathbf{C}\mathbf{C}_j^T + \mathbf{C}\mathbf{C}_{jc}^T$
 - 10: End for
 - 11: End do
-

Splitting covariate-specific hat matrices $\mathbf{R}$ into two chunks ($q=2$) should reduce memory needed by approximately half. For example, if the allocation of $\mathbf{R}$ needs 32GB of memory, if we use the chunk-wise computation described above and split $\mathbf{R}$ into two chunks, then memory

allocation needed will be reduced to around 16GB. Splitting $\mathbf{R}$ into q chunks is expected to reduce memory by a factor of q . However, it worth noting that since $\mathbf{A}_j$ is not stored and will be computed repeatedly for each chunk of $\mathbf{R}_j$, it will add additional cost if we specify the number of chunks greater than 1. In this way, we need to decide the minimum number of chunks that can make the chunk of $\mathbf{R}$ fit into the available memory. A simple calculation to decide q would be

$$q = \frac{8 \times k \times n \times n}{m \times 1024^3} \quad (24)$$

where m is the available memory in the hardware, and we assume we use double precision matrices (64-bit). For example, if the available memory is 16GB on a machine, and we have an MGWR model with 40,000 data points and 10 covariates, q will need to be at least 15 chunks to avoid memory problems.

5 Testing dataset and environment

5.1 A data generating process for deriving a test dataset

For testing the performance of the improved MGWR software, we develop a Data Generating Process (DGP) that can automatically synthesize multi-scale spatial processes for a varying number of covariates and observations. Detailed steps can be found in the supplementary materials. In this study, we generated 15 parameter surface grids with dimension of 200 by 200 in total of 40,000 locations. The resulting parameter surfaces $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{15}$ are spatial random fields with different heterogeneity, which can be seen in Figure 2.4. The response variable $\mathbf{y}^*$ is constructed by $\mathbf{y}^* = \sum_{j=1}^{15} \mathbf{b}_j \mathbf{x}_j + \boldsymbol{\varepsilon}$ where each covariate $\mathbf{x}_j$ and the error term $\boldsymbol{\varepsilon}$ are randomly drawn from the normal distribution $N(0, I)$.

5.2 Testing environment

The single desktop used in comparing performance across MGWR software packages is equipped with an Intel i7-4790 4-core CPU at 3.60Ghz and 16GB RAM. The Python environment used is version 3.6. The R environment used is version R 3.5.2. Runtimes are measured based on Python and R's built-in *time* modules. Memory usages are measured based on the *memory_profiler*⁴ module for Python and the built-in *memory* module for R. To enhance reproducibility, all testing codes, datasets used and results can be found publicly in a Github repository⁵.

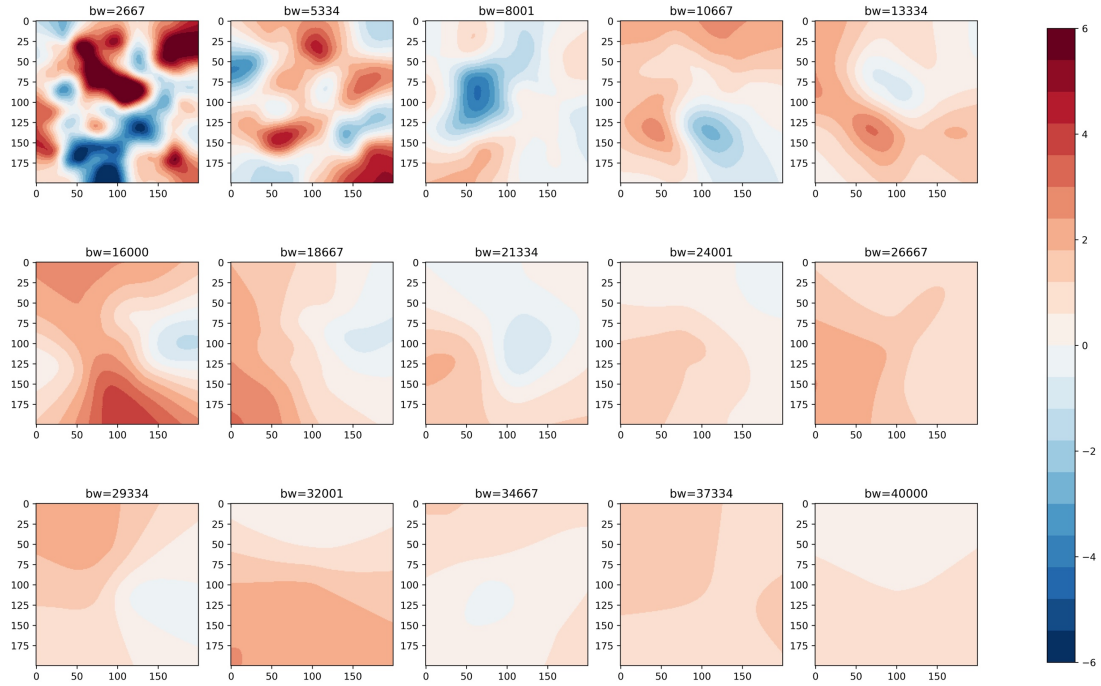


Figure 2.4. Synthesized known parameter surfaces with varying scales from local to global.

⁴ https://github.com/pythonprofilers/memory_profiler

⁵ https://github.com/c040120/com_improv_mgwr_data_codes

6 Results

6.1 Runtime comparison with the original MGWR implementation

First, we benchmark our improved algorithm with the original implementation of MGWR used in Fotheringham et al. (2017) to show that the improved method reduces various issues of computation. To demonstrate this, we reproduce MGWR models and record runtimes using the example datasets in Fotheringham et al. (2017)⁶. Figure 2.5 shows the comparison of runtimes from MGWR 2.0 and runtimes explicitly reported in Fotheringham et al. (2017) using two example datasets. The simulation 1 dataset used in Fotheringham et al. (2017) consists of three covariates with 625 observations. For this synthetic dataset, the calibration takes 4.8 minutes in Fotheringham et al. (2017, pg.1257), while it only takes 7.4 seconds after the improvements described here. For the other example used in that paper, the Irish potato famine dataset with 2317 locations and 9 covariates, the paper reported that the calibration process took 51 hours (Fotheringham et al., 2017 pg. 1260). In comparison, using the improved method proposed in this paper, the runtime is reduced to around 6 minutes, which is 510 times faster than the original implementation. For both examples, our improved method shows a crucial performance increase. Also, the speed-up is projected to be greater for more complicated models which contain more observations and covariates.

⁶ The computer used in Fotheringham et al. (2017) has dual 2.6 Ghz 8-core CPUs and 64 GB of memory.

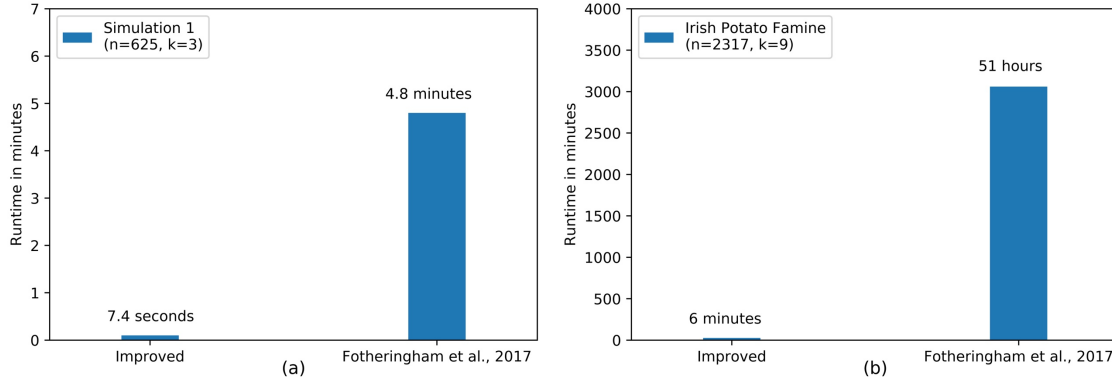


Figure 2.5. Runtime comparison based on two example datasets from Fotheringham et al. (2017).

6.2 Comparison against *GWmodel*

Second, we compare our proposed improvement against *GWmodel* package's MGWR implementation. We randomly draw a subset of data points from the synthetic dataset described in Section 5.1 and calibrate MGWR models using both software packages to record the runtimes and maximum memory usage. Results are presented in Table 2.2. With a sample size of $N=1,000$, MGWR 2.0 is around 16 times faster than *GWmodel*. When $N=2,000$, MGWR is around 30 times faster than *GWmodel* and when N increases to 5,000, MGWR 2.0 is around 140 times faster than *GWmodel*. When the sample size increases to 10,000, *GWmodel* generates a memory allocation error on the testing machine but MGWR 2.0 is able to complete the calibration within an hour. For sample size $N=20,000$, MGWR 2.0 takes around 6.5 hours to complete and for $N=40,000$, MGWR 2.0 takes around 4 days to finish. This is because memory allocation exceeds the memory limit of the testing machine so the software has to use more chunks to split the inference computation in order to fit in the memory, which slows the calibration. If a machine with more memory were used, the calibration by MGWR 2.0 would be accelerated by several orders of magnitude. From a memory perspective, MGWR 2.0 caps out the memory usage at around 11GB (approximate available memory on the test machine) when N is great than around 15,000 and the

software starts to bring the chunk parameter into action to avoid memory allocation error.

Table 2.2. Runtime and memory comparison between MGWR 2.0 and *GWmodel*

MGWR calibration with inference (10 covariates)				
Sample Size	<i>MGWR 2.0</i>		<i>GWmodel</i>	
N	Time (seconds)	Memory	Time	Memory
1000	50	1.3 GB	802	0.5 GB
2000	127	2.0 GB	3818	0.8 GB
5000	370	3.2 GB	52,119 (~15hrs)	5.5 GB
10,000	2652	10 GB	Memory error	NA
15,000	10,068 (~3hr)	11 GB	Memory error	NA
20,000	23,236 (~6.5hr)	11 GB	Memory error	NA
40,000	301,126 (~ 4d)	11 GB	Memory error	NA

6.3 Effectiveness of memory reduction in inference computation

With the improved implementation borrowed from FastGWR, memory is not an issue in the parameter estimation component of MGWR. In the inference component, we need to use the method described in Section 3.4 to separately compute covariate-specific hat matrices in chunks of columns. To demonstrate the effectiveness of partitioning this computation into chunks, we calibrate two MGWR models with sample sizes of $N=10,000$ and $N=20,000$ and 10 covariates, and record the memory usage with varying numbers of chunks. Figure 2.6 (a) shows how memory is reduced when the number of chunks is increased. When the number of chunks is 1, the inference computation does not involve any partitioning and covariate-specific matrices are computed as described in Yu et al (2019). When the number of chunks increases to 2, memory is reduced by approximately half, and this behavior is consistent for both sample sizes. For example, when $N=20,000$, the memory demand is reduced from around 40GB to around 20GB. When the number of chunks is increased to 4, the memory requirement is reduced further to

around 10GB. This implies that if a machine has a RAM of 16GB, it is not able to compute MGWR inference for a sample size of 20,000 due to the memory needed but if we use the improved method presented in this paper, we could increase the number of chunks and compute inference chunk-by-chunk to get around the memory constraint. However, increasing the number of chunks comes with the cost of increasing calibration time. Figure 2.6 (b) shows how runtime is increased when the number of chunks is increased. As can be seen, when 2 chunks are being used, it only adds a small amount of extra runtime. When more chunks are used, runtime goes up but only slowly. Figures 2.6 (a) and (b) combined show that using more chunks does not add too much extra computation time but does decrease memory demand drastically.

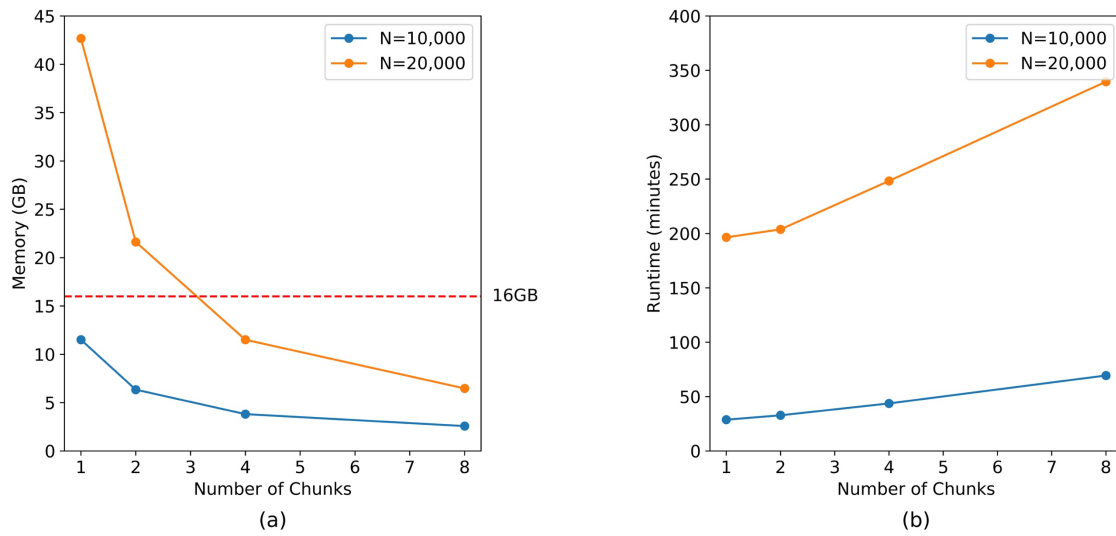


Figure 2.6. Memory demand and runtime change with varying numbers of chunks used.

6.4 On the scalability of MGWR parallelization

In this section, we examine the scalability of the parallelization used in the MGWR calibration. First, we calibrate MGWR models with sample sizes of $N=5,000$ and $N=10,000$ with varying numbers of cores. Since the parallelization mechanism is different for parameter

estimation and inference, we separately record the runtimes for the two components. In Figure 2.7, stacked bar plots (a) and (c) show the actual runtimes respectively for parameter estimation and inference computation for the two sample sizes using 1 to 4 processors. It can be seen that when $N=5,000$, parameter estimation takes more time than inference computation, while when $N=10,000$, inference computation takes more time than parameter estimation. As more processors are used, runtimes decrease logarithmically for both sample sizes. The line plots in Figure 2.7 (b) and (d) show the speed-up factors for parameter estimation, inference and a combination of both (total) based on runtime using a single processor so that the speed-up factor is 1 when using only 1 processor. Parameter estimation has strong scalability when using more processors, and using 4 processors generates a speed-up factor close to 3. Inference computation scales less well than parameter estimation but still the use of 4 processors generates a speed-up factor around 2. When combined, the MGWR calibration overall is around 2.4 times faster when using 4 processors than when using only a single processor. Since the testing machine only has 4 physical cores, it is hard to infer the scalability if a machine has more processors. Thus, we used another desktop machine with 8 physical cores to examine both the compatibility as well as scalability of the MGWR parallelization. The results can be seen in Figure 2.8 where it can be seen that using 8 processors speeds-up the overall MGWR algorithm by a factor of 4.8. The speed-up factor of parameter estimation is around 6 and speed-up factor of inference computation is around 4. Generally speaking, the scalability of MGWR parallelization is reasonable. There are two reasons for not getting perfect linear scalability. First, based on Amdahl's law about parallel computing that the speed-up of a program is limited by its serial part, the back-fitting procedure of the MGWR calibration involves sequential updating of parameter estimates and residuals which reduces performance. Second, parallelization can be implemented by either multi-threading or multi-processing. Multi-processing is more desirable in this situation because multi-threading is limited by the Global Interpreter Lock (GIL) of *Python*. However, the downside of using multi-

processing is that there is a noticeable amount of time needed for data copying to different processors, even though we limit this by minimizing the amount of data being copied.

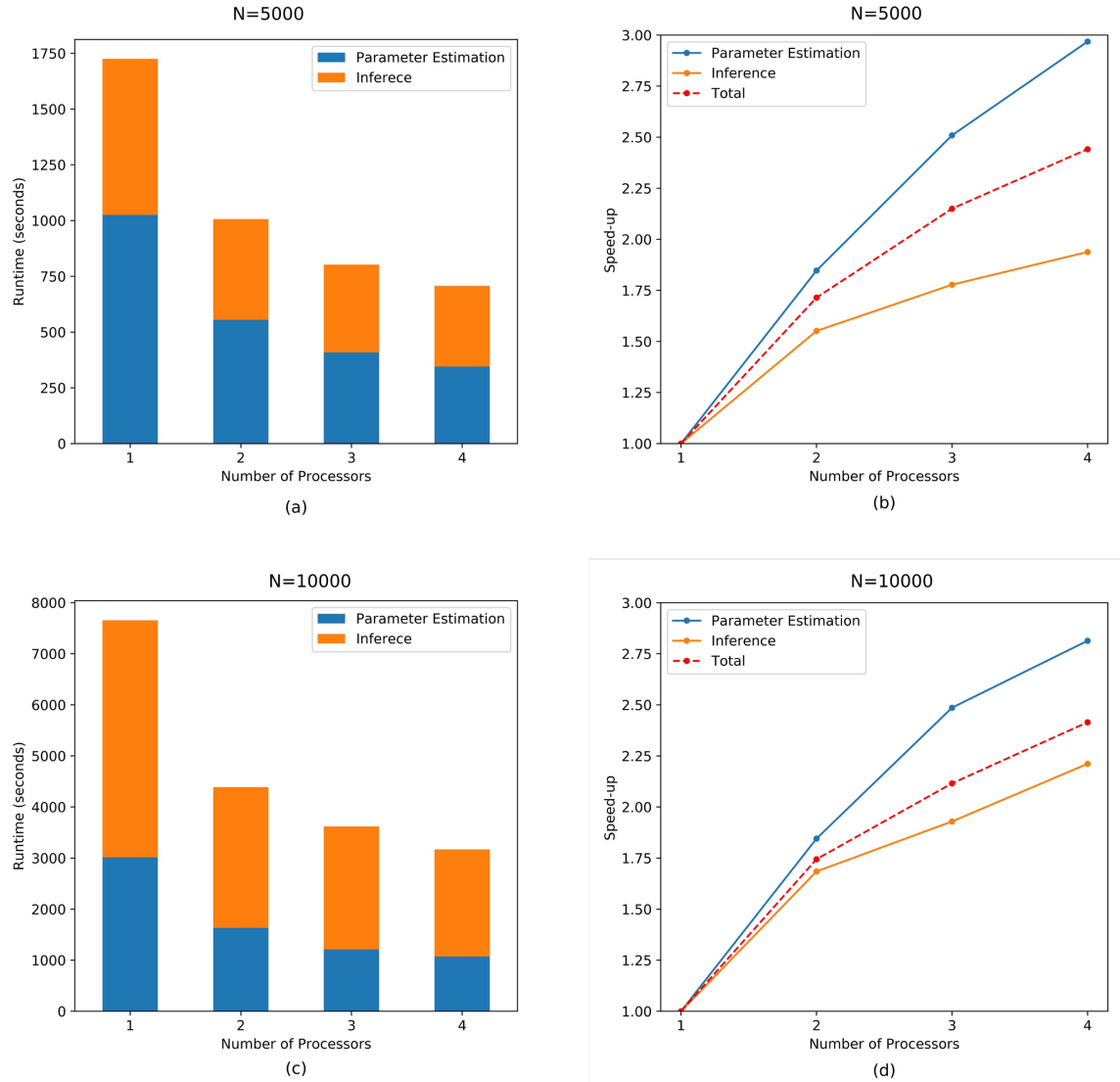


Figure 2.7. Runtime and speed-up factor of the new MGWR calibration for parameter estimation, inference and total calibration on a 4-processor machine based on a sample size of 10,000.

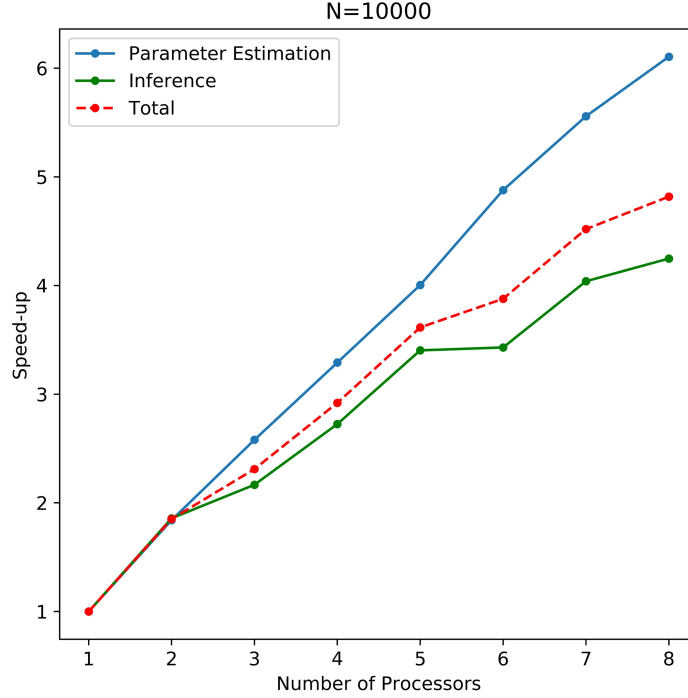


Figure 2.8. Speed-up factor of the new MGWR calibration for parameter estimation, inference and total calibration on an 8-processor machine based on a sample size of 10,000.

6.5 Computational evaluation of MGWR

In this section, we evaluate the computation of MGWR using the synthesized dataset described in Section 5.1. We calibrate MGWR models including inference for a varying number of sample sizes and with 5, 10 and 15 covariates. Given that convergence rates may vary in empirical studies, we also include three categories of convergence rate which are: fast convergence (10 back-fitting iterations); moderate convergence (50 back-fitting iterations); and slow convergence (100 back-fitting iterations). Figure 2.9 shows the approximate time needed to complete an MGWR calibration (with parameter estimation and inference) under the three different convergence rates and for varying numbers of both observations and covariates⁷. The

⁷ Results shown are based on testing machine described in Section 5.2, actual runtime may vary on different machines.

blue areas on each figure show what model size can be fitted in under 10 minutes; the green areas show the model size that can be calibrated within one hour; the yellow areas show the model size that can be calibrated within 6 hours; the red areas show the model size that can be calibrated within one day; and the purple areas show the model size needing more than one day for calibration. For example, for 10 covariates, a model with sample size of 2000 can be fitted in under 10 minutes if it converges at a moderate rate. In general, we can see that MGWR is able to be calibrated on models with fewer than 20,000 observations within a reasonable time (less than a day) if it converges moderately. When the sample size is larger than 20,000 or more covariates are involved, it might take several days to calibrate an MGWR model on a normal desktop computer. However, using a workstation with more processors and memory will accelerate MGWR calibration several times so the task would probably be completed within one day. The ability of MGWR 2.0 to utilize more computing resources opens the possibility for handling up-to 100,000 observations in an MGWR calibration.

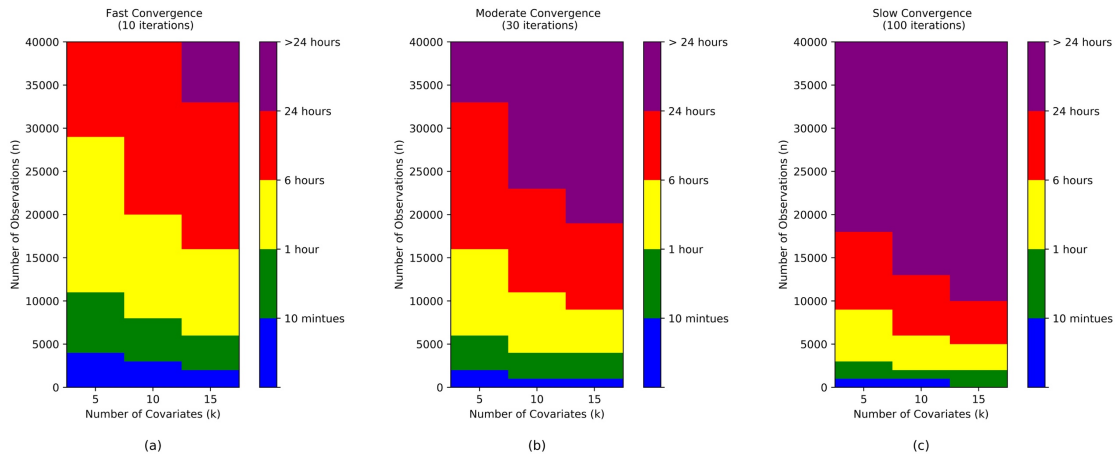


Figure 2.9. Approximate runtime needed to complete an MGWR calibration (with parameter estimation and inference) under three different convergence rates and for varying numbers of both observations and covariates.

7 Conclusions

In this paper, we make critical computational improvements to the newly developed Multiscale Geographically Weighted Regression (MGWR) model (Fotheringham et al., 2017). We extend the parallelization framework introduced in the FastGWR algorithm (Li et al., 2019) to speed-up each univariate GWR model within the MGWR back-fitting parameter estimation. Also, using the recently developed MGWR inference framework (Yu et al., 2019), we introduce a parallel chunk-wise partitioning method for computing MGWR inference to solve memory problems. We compare our improved calibration method with the original implementation used in Fotheringham et al. (2017) and find it is around 500 times faster. Also, we compare our new method with the *GWmodel* package in the R environment (Gollini, 2016) and find it to be 30 to 140 times faster than *GWmodel* for sample sizes ranges from $N=1000$ to $N=5000$. For sample sizes larger than 5000, only our improved method is able to calibrate an MGWR model. We further demonstrate that our new algorithm scales well on multi-processor machines with a speed-up ratio of 2.4 on a 4-processor machine and 4.8 on an 8-processor machine. With the current improvement, an MGWR model with around 20,000 observations can be calibrated within a day on a normal desktop machine with 4 processors and 16GB of memory. With a workstation with more processors and memory, it is expected that it is feasible to calibrate an MGWR model with up to 100,000 data points.

The methodology introduced in this paper has been integrated into the *mgwr* python package and the MGWR 2.0 GUI software, both of which are freely available to download. This not only enhances the accessibility of the MGWR model for new applications to explore multi-scale spatial heterogeneity but also brings the possibility of much larger scale local multi-scale analysis. In addition, the computational improvements described here can also benefit some recent spatial analysis models developed based on the MGWR framework such as the Parameter

Specific Distance Metric Model (Lu et al., 2017) and the Multiscale Geographically and Temporally Weighted Regression model (Wu et al., 2019) both of which have severe computational restrictions.

Data and codes availability statement

The data and codes that support the findings of this study are available in figshare.com with the identifier doi:10.6084/m9.figshare.10260881.v1.

References

- Bitter, C., Mulligan, G. F., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*, 9(1), 7-27.
- Buja, A., Hastie, T., & Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, 453-510.
- Cardozo, O. D., García-Palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*, 34, 548-558.
- Comber, A. J., Brunson, C., & Radburn, R. (2011). A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. *International journal of health geographics*, 10(1), 44.
- Fotheringham, A. S., Brunson, 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.
- Finley, A. O. (2011). Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, 2(2), 143-154.
- 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.
- Gollini, I., Lu, B., Charlton, M., Brunson, C., & Harris, P. (2015). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, 63(17).
- Harris, R., Singleton, A., Grose, D., Brunson, C., & Longley, P. (2010). Grid-enabling Geographically Weighted Regression: A Case Study of Participation in Higher Education in England. *Transactions in GIS*, 14(1), 43-61.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models* (Vol. 43). CRC Press.

- Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes Jr, M. G., Estes, S. M., ... & Liu, Y. (2013). Estimating ground-level PM_{2.5} concentrations in the southeastern US using geographically weighted regression. *Environmental Research*, 121, 1-10.
- Li, Z., Fotheringham, A. S., Li, W., & Oshan, T. (2018). 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.
- Lu, B., Yang, W., Ge, Y., & Harris, P. (2018). Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths. *Computers, Environment and Urban Systems*, 71, 41-57.
- Murakami, D., Lu, B., Harris, P., Brunsdon, C., Charlton, M., Nakaya, T., & Griffith, D. A. (2018). The importance of scale in spatially varying coefficient modeling. *Annals of the American Association of Geographers*, 1-21.
- Murakami, D., & Griffith, D. A. (2015). Random effects specifications in eigenvector spatial filtering: a simulation study. *Journal of Geographical Systems*, 17(4), 311-331.
- Murakami, D., & Griffith, D. A. (2019). Spatially varying coefficient modeling for large datasets: Eliminating N from spatial regressions. *Spatial Statistics*.
- 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., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (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.
- Oshan, T., Wolf, L. J., Fotheringham, A. S., Kang, W., Li, Z., & Yu, H. (2019). A comment on geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*. doi:10.1080/13658816.2019.1572895
- Tran, H. T., Nguyen, H. T., & Tran, V. T. (2016, October). Large-scale geographically weighted regression on Spark. In *Knowledge and Systems Engineering (KSE), 2016 Eighth International Conference on* (pp. 127-132). IEEE.

- Troy, A., Grove, J. M., & O'Neil-Dunne, J. (2012). The relationship between tree canopy and crime rates across an urban–rural gradient in the greater Baltimore region. *Landscape and Urban Planning*, 106(3), 262-270.
- Wolf, L. J., Oshan, T. M., & Fotheringham, A. S. (2018). Single and multiscale models of process spatial heterogeneity. *Geographical Analysis*, 50(3), 223-246.
- Wu, C., Ren, F., Hu, W., & Du, Q. (2019). Multiscale geographically and temporally weighted regression: exploring the spatiotemporal determinants of housing prices. *International Journal of Geographical Information Science*, 33(3), 489-511.
- Yu, H., Fotheringham, S., Li, Z., Oshan, T., Kang, W., & Wolf, L. (2019). Inference in multiscale geographically weighted regression. *Geographical Analysis*. doi: 10.1111/gean.12189
- Zhang, J. (2010). Towards personal high-performance geospatial computing (HPC-G): perspectives and a case study. In *Proceedings of the ACM SIGSPATIAL international workshop on high performance and distributed geographic information systems* (pp. 3-10). ACM.

Supplementary Material

Data Generating Process in Section 5.1

With a given set of bandwidths from small to large, we calibrate an MGWR model using randomly drawn response variable and covariates. The resulting parameter estimates will be used as known processes to reconstruct the final response variable used for performance testing. The detailed steps used in this study are described as follows:

- 1) Make a gridded coordinate system of size 200 by 200, which has a total of 40,000 locations.
- 2) Randomly draw a temporary fake response variable $\mathbf{y}$ for each of the grid cells from a normal distribution with 0 as its mean and 1 as its variance.
- 3) Draw k covariates $\mathbf{X}_{1...k}$ from a multivariate normal distribution with 0 as the mean and 1 as the variance.
- 4) Run an MGWR model with an adaptive (nearest neighbors) bi-square kernel using $\mathbf{y}$ as the response variable, $\mathbf{X}_{1...k}$ as the covariates and with a set of predefined adaptive bandwidths to be $[n/k, 2n/k, 3n/k \dots nk/k]$. A set of parameter surfaces can then be obtained. These sets of parameter estimates will have varying degrees of smoothness but their values will be close to zero. To make the estimates have a reasonable range, we multiply them by 100 and add a constant of 3 to get to the set of the parameter estimates $\boldsymbol{\beta}_{1...k}$. Figure 2.S1 shows the parameter surfaces for 15 generated known parameter surfaces. We can see that spatial variation in processes gradually decreases as the bandwidth increases.
- 5) Redraw $\mathbf{X}^*$ but this time we add a small amount of multicollinearity to the covariates pairs (a correlation coefficient of 0.3). This is simply to more accurately represent real-world situations.
- 6) Generate a final response variable $\mathbf{y}^*$ from the covariates in 5) and parameter surfaces in 4) with added random noise drawn from $N(0, 0.5)$.
- 7) The resulting $\mathbf{y}^*$ and $\mathbf{X}_{1...k}^*$ are the final synthetic dataset. Figure 2.S2 shows the recovered parameter estimates surfaces from MGWR using this final testing dataset.

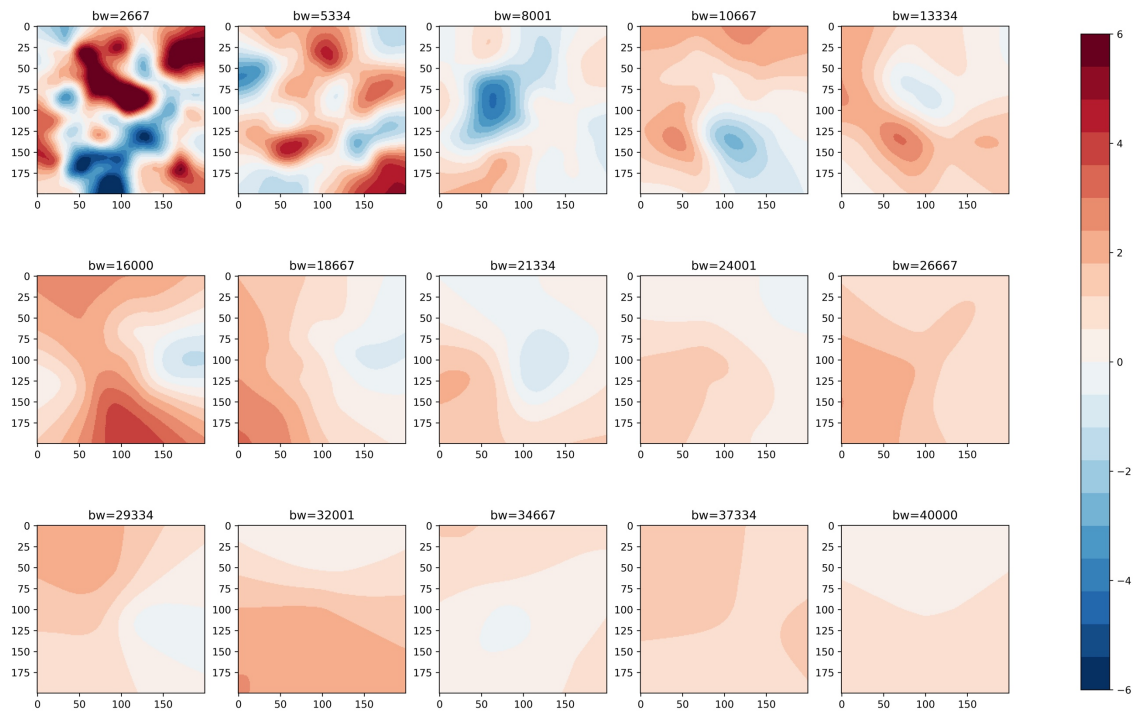


Figure 2.S1. Synthesized known parameter surfaces with varying scales from local to global.

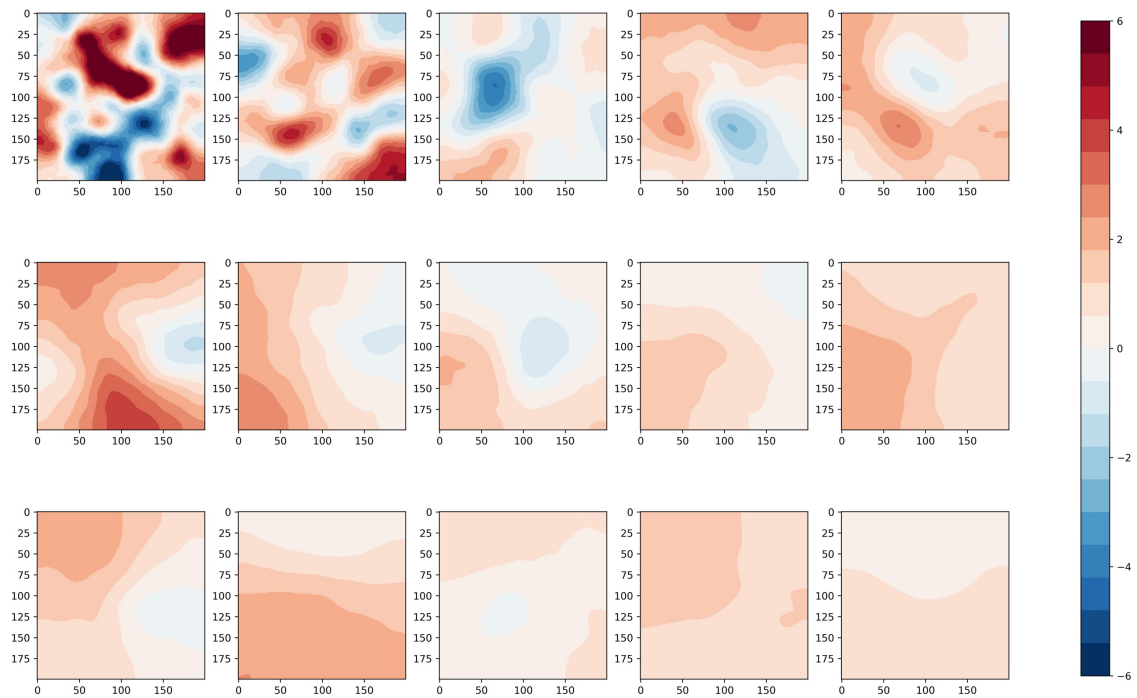


Figure 2.S2. Recovered parameter estimates surfaces from MGWR using the final testing dataset.

CHAPTER 3

MEASURING BANDWIDTH UNCERTAINTY IN MULTISCALE GEOGRAPHICALLY WEIGHTED REGRESSION USING AKAIKE WEIGHTS

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 criteria 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 operational 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 Multiscale 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 heterogeneity, 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 multiscale 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 more spatially smoothed process which 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 et al., 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 say 50 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 nearest neighbors. Data at the local regression point are given a weight of 1 and data borrowed from neighboring locations are given a discounted 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. Firstly, the background of MGWR is reviewed. Secondly, bandwidth uncertainties are examined with a simulation dataset using bootstrapping and Akaike weights. Thirdly, an empirical example of obesity rate modeling for the city of Phoenix is presented. Lastly, the paper concludes with remarks.

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, $\mathbf{f}_1(\mathbf{x}_1), \cdots, \mathbf{f}_k(\mathbf{x}_k)$ 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 $\mathbf{f}_1, \cdots, \mathbf{f}_k$ 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 j^{th} additive component $\mathbf{f}_{bw_j}(\mathbf{x}_j)$ is a product of element-wise multiplication ($\odot$) of local parameters $\mathbf{b}_j$ and covariate $\mathbf{x}_j$

$$\mathbf{f}_{bw_j}(\mathbf{x}_j) = \mathbf{b}_j \circ \mathbf{x}_j = \begin{pmatrix} b_{1j}x_{1j} \\ b_{2j}x_{2j} \\ \vdots \\ b_{nj}x_{nj} \end{pmatrix} \quad (3)$$

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 to the number of parameters is below 40, which is frequently the case in GWR if 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-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 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

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

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, 2002). The denominator is used to normalize the Akaike weights so that all values lie between 0 and 1 with the sum being 1 ($\sum w_k = 1$) (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. The x and y coordinates of location $i \in \{1 \dots, n\}$ are denoted as u_i and v_i and are constructed with the following formulas (6) and (7)

$$u_i = 12.5 + 12.5\sqrt{r_i} \cos \theta_i \quad (6)$$

$$v_i = 12.5 + 12.5\sqrt{r_i} \sin \theta_i \quad (7)$$

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 formulas (8) and (9), and these true parameter surfaces, $\mathbf{b}_1$ and $\mathbf{b}_2$, are shown in Figure 3.1.

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

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

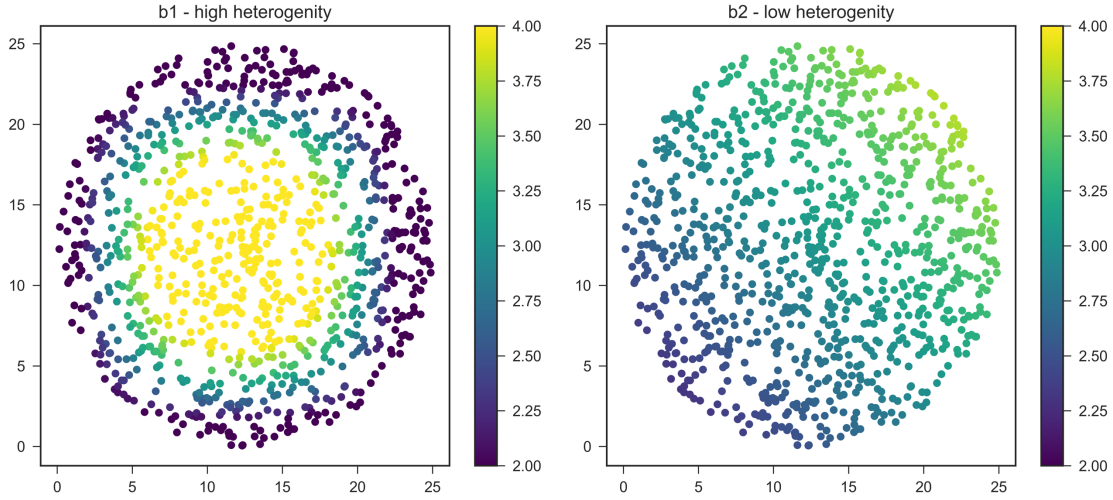


Figure 3.1. Generated spatial processes with low- and high- heterogeneity.

where b_{1i} and b_{2i} are the parameters of location $i \in \{1 \dots, n\}$; 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 column 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 column vectors of 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:

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

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 3.2.

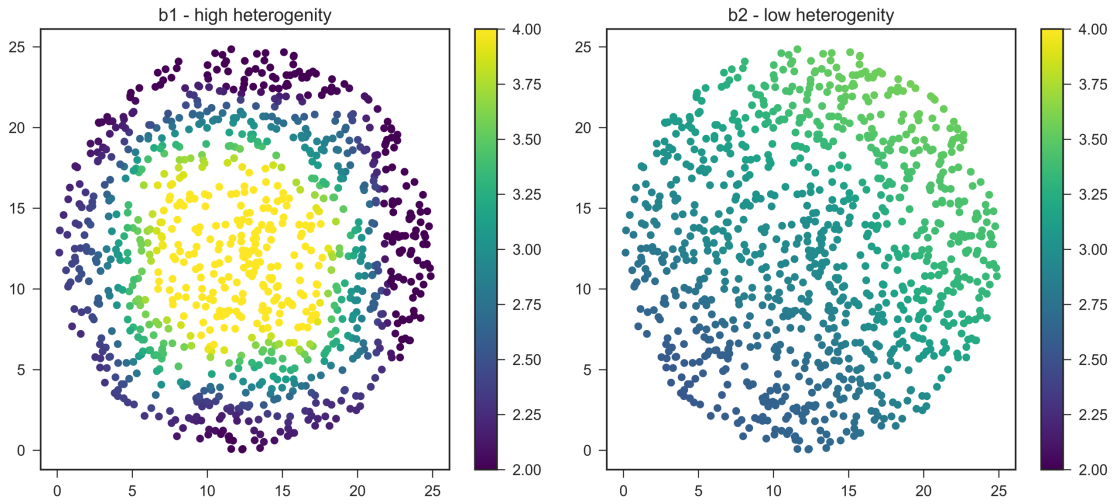


Figure 3.2. Recovered parameter surfaces using MGWR with optimal bandwidths 70 and 300 for the high- and low- heterogeneity processes, respectively.

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 $\hat{\boldsymbol{\varepsilon}}$. Then we generate B ($B=1,000$) bootstrap samples with each sample containing a new response vector $\mathbf{y}^* = \hat{\mathbf{y}} + \hat{\boldsymbol{\varepsilon}}_{rand}$ where $\hat{\boldsymbol{\varepsilon}}_{rand}$ is a residual vector randomly drawn with replacement from $\hat{\boldsymbol{\varepsilon}}$. 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 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 3.3 and Table 3.1.

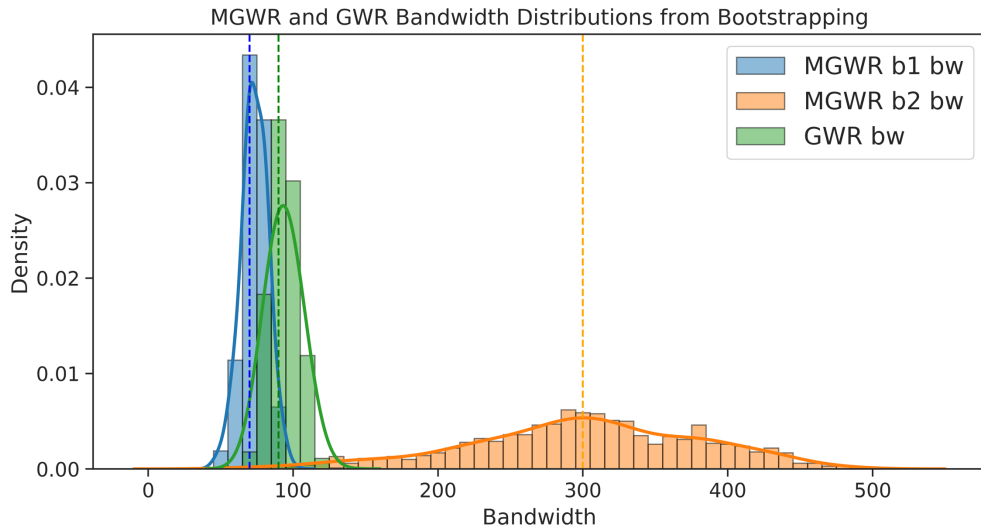


Figure 3.3. Optimal bandwidth distributions of two processes in Figure 3.1 under bootstrapping (smooth lines are plotted based on kernel density estimation).

Table 3.1. Summary statistics for the optimal bandwidths obtained from the bootstrap samples.

	Optimal	Mean	STD	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

Figure 3.3 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 3.1. 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, 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 model described in Equation (10) (see Appendix B); and 3) changing the orientation of the process $\mathbf{b}_2$ holding the level of heterogeneity constant (see Appendix C). Results all indicates 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.

4.3 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 1000 nearest neighbors with a 10-neighbor interval⁸. Each candidate bandwidth has a corresponding Akaike weight representing the probability of it being the optimal bandwidth (Figure 3.4). 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 3.4. 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, the inclusion to the bandwidth CI should stop when the cumulative Akaike weight is just above 0.95. Table 3.2 shows how this procedure operates for process $\mathbf{b}_1$. 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 the addition of the bandwidth 90 because at this point the cumulative Akaike weight equals 0.95. Consequently, we can state that the 95% CI of bandwidths for process $\mathbf{b}_1$ is [60, 90]. The bandwidth CI is not necessarily symmetrical, and if bandwidth 50 is included this would create a 99% CI.

⁸ Using a smaller step-size rather than 10 will produce more detailed Akaike weight curve but with additional computation.

Table 3.2. An example of obtaining the 95% confidence set of bandwidth for process $\mathbf{b}_1$ 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
...	...	...	...

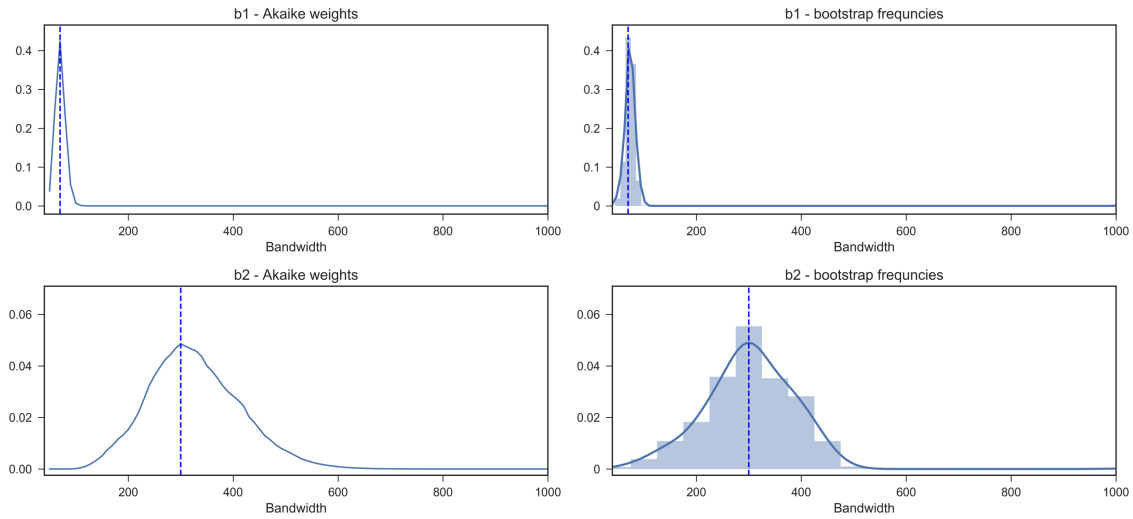


Figure 3.4. Akaike weights and bootstrap frequencies for different bandwidths (smooth lines are plotted based on kernel density estimation).

The Akaike weight-based 95% bandwidth CIs can be compared with the 95% CIs obtained from bootstrapping (Table 3.3). 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 3.3. 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 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 j^{th} covariate, $\mathbf{b}_j$, unconditional parameter estimates, $\widehat{\mathbf{b}}_j$ can be computed based on Akaike weights:

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

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

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

Note that there are two components in $\widehat{var}(\widehat{\mathbf{b}}_j)$: (i) the variance from parameter estimation, $\widehat{var}(\widehat{\mathbf{b}}_{rj})$, and (ii) the variance from bandwidth selection uncertainty, $(\widehat{\mathbf{b}}_{rj} - \widehat{\mathbf{b}}_j)^2$. 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, 2010). However, it is 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, the 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).

In Figure 3.5 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 $\mathbf{b}_1$ 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.

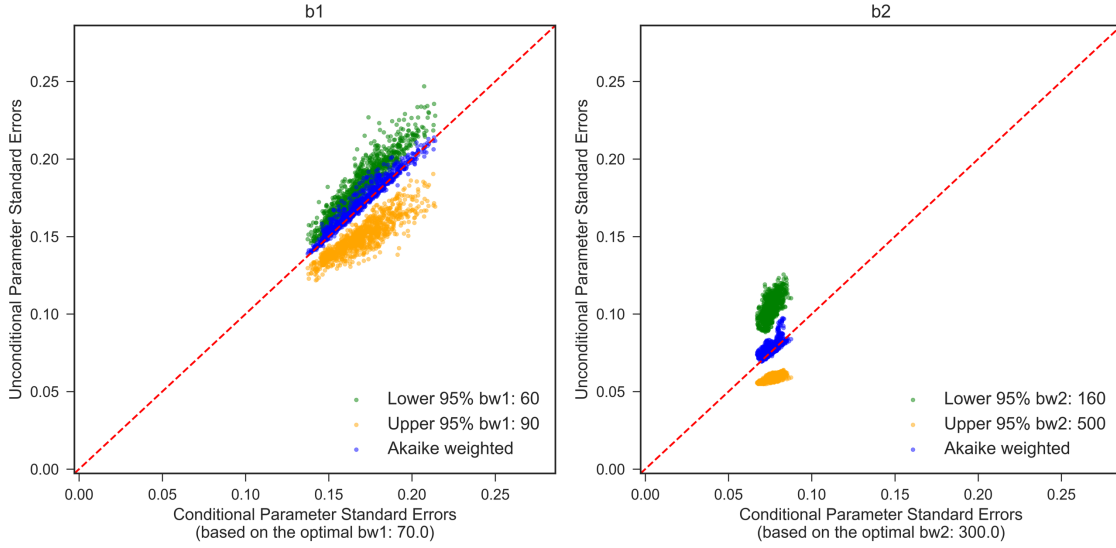


Figure 3.5. Comparison of conditional and unconditional standard errors of local parameter estimates for processes $\mathbf{b}_1$ and $\mathbf{b}_2$.

This is also seen in Figure 3.6 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 95% CI is computed by $\hat{\mathbf{b}}_j \pm 1.96\sqrt{\hat{var}(\hat{\mathbf{b}}_j)}$. 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 change 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.

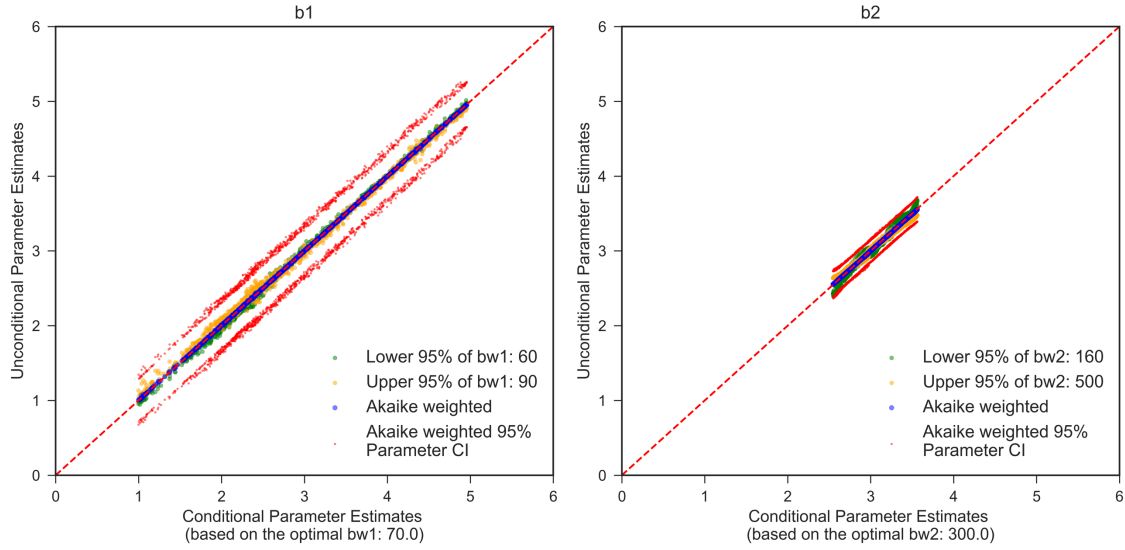


Figure 3.6. Comparison of conditional and unconditional local parameter estimates for processes b_1 and b_2 .

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 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 b_1 and 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 b_1) and (0.89 vs. 0.87 for 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) compared MGWR parameter standard errors with those obtained by Bayesian Spatially Varying Coefficients model and found that the MGWR parameter confidence intervals 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 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 ≥ 18) defined as obese⁹ (body mass index (BMI) ≥ 30.0 kg/m²) by The Centers for Disease Control and Prevention (CDC) in each of the 815 census tracts in the Phoenix metropolitan area¹⁰ 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 CDC¹¹, and the socio-demographic covariates are retrieved from the American Community Survey (ACS) 2015 5-year estimates dataset¹². The regression model is formulated as follows:

⁹ <https://www.cdc.gov/500cities/definitions/unhealthy-behaviors.htm#OBESITY>

¹⁰ Six sparsely populated tracts are removed in this example.

¹¹ <https://www.cdc.gov/500cities>

¹² <https://data.census.gov>

$$\begin{aligned}
Pct_Obesity_i = & b_{0i} + b_{1i}Pct_Checkup_i + b_{2i}Median_Age_i + b_{3i}Pct_Hispanic_i \\
& + b_{4i}Pct_SNAP_i + b_{5i}Pct_College_i + \varepsilon_i
\end{aligned} \tag{13}$$

where obesity percentage (*Pct_Obesity*) is regressed on percentage of population undergoing yearly 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*); $b_{0i} \dots, b_{5i}$ 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 et al. (2017) so that bandwidths and local parameters are comparable and invariant to the scale of the data.

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 3.7 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.

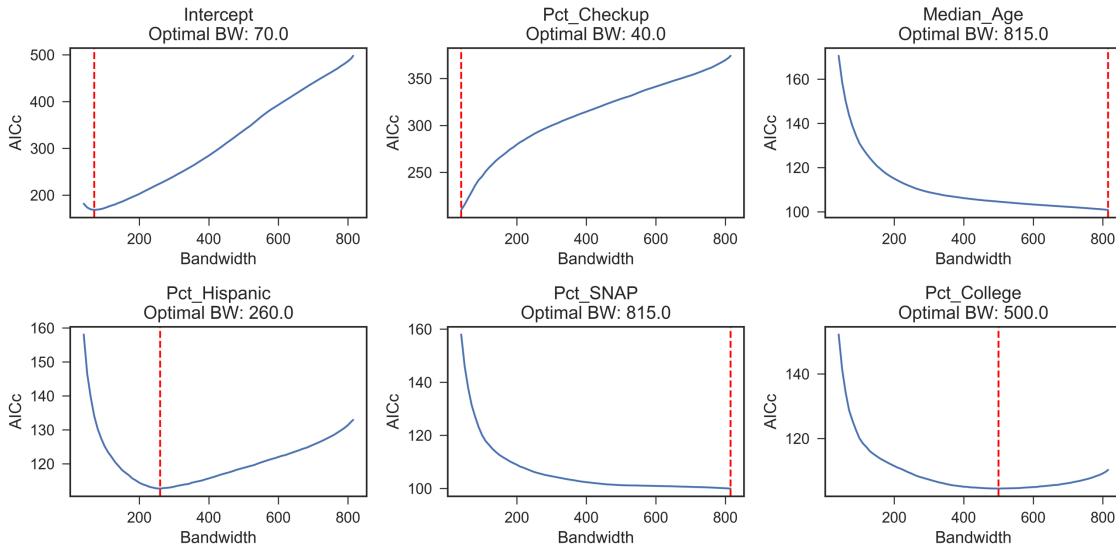


Figure 3.7. AICc values against bandwidth for each of the five covariates and the intercept. The red line is the optimal covariate-specific bandwidth based on the minimum AICc value.

Akaike weights and bandwidth CIs are computed for these five covariates plus the intercept as shown in Figure 3.8. In each case the red line shows the optimal bandwidth and the two green lines show the upper and lower bounds of the 95% 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 3.4. 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*.

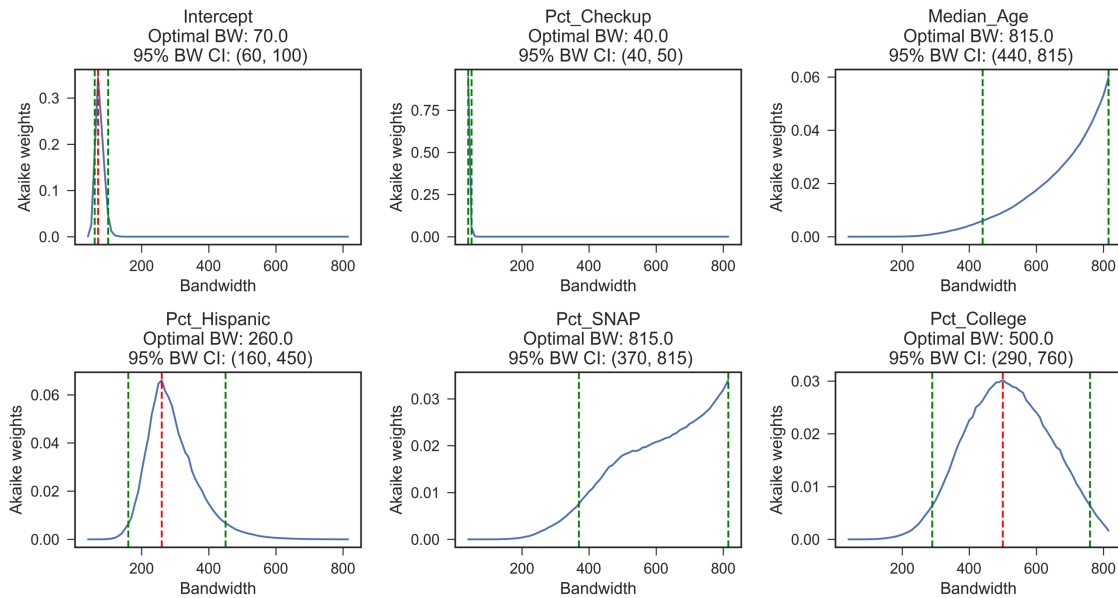


Figure 3.8. Akaike weights computed for different bandwidths for each set of local parameter estimates. The red line is the optimal bandwidth found by MGWR. The green lines depict the lower and upper bounds of the 95% confidence interval for each optimal bandwidth.

Table 3.4. 95% confidence interval of bandwidth based on Akaike weight for each covariate.

Covariates	Optimal bandwidth	95% confidence Interval
Intercept	70	[60, 100]
% of Check up	40	[40, 50]
Median age	815	[440, 815]
% of Hispanic	260	[160, 450]
% of SNAP	815	[370, 815]
% of College	500	[290, 760]

Table 3.5. Summary of OLS parameter estimates and MGWR Akaike-weighted parameter estimates

		OLS		MGWR Akaike weighted	
Covariates	Est.	SE.	t-value	Est. min*	Est. max*
Intercept	0.000	0.012	0.000	-0.297	0.503
% of Check Up	-0.235	0.030	-7.756	-0.610	-0.144
Median Age	0.186	0.027	6.755	0.119	0.136
% of Hispanic	0.136	0.018	7.397	0.056	0.133
% of Snap	0.536	0.021	25.230	0.323	0.332
% of College	-0.286	0.020	-14.439	-0.450	-0.362
R²		0.88		0.93	
AICc		606.0		287.3	

*Minimum and maximum of MGWR parameter estimates shown in the table are calculated for the significant (0.05) estimates only.

For context, we also report the OLS (Ordinary Least Squares) parameter estimates along with the minimum and maximum of the significant MGWR Akaike-weighted parameter estimates in Table 3.5. The MGWR model has a better fit than the OLS model with a higher R^2 value (0.93 vs. 0.88) and lower AICc (287.3 vs. 606.0). Maps of each of the four sets of local Akaike-weighted parameter estimates are shown in Figure 3.9. These are 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 3.10 and 3.11. Figure 3.9 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, 2015; Yu et al., 2019), are masked out and shaded in grey. 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 indicates 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.

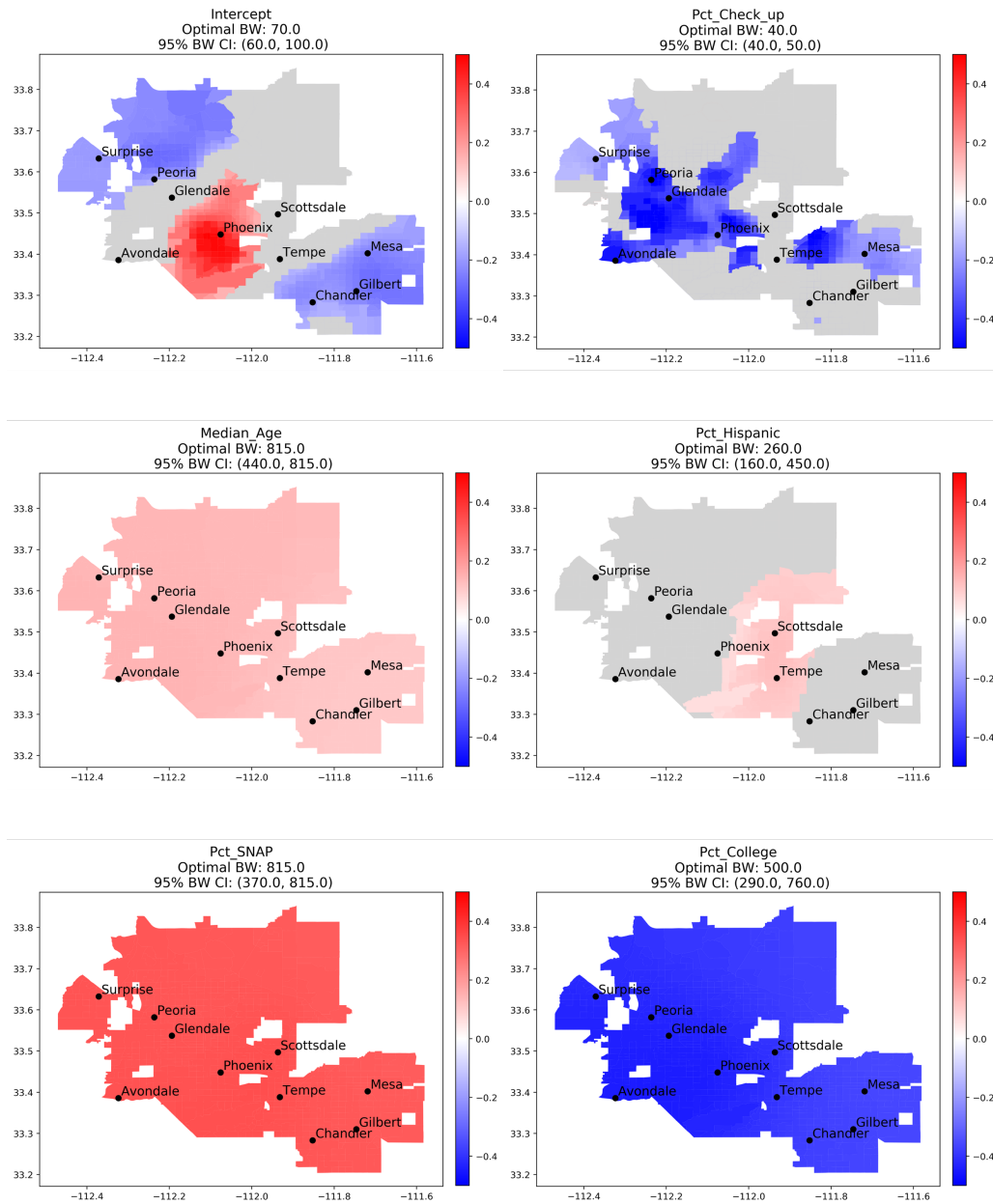


Figure 3.9. Maps of Akaike weighted local parameter estimates with insignificant parameters masked out and shaded in grey.

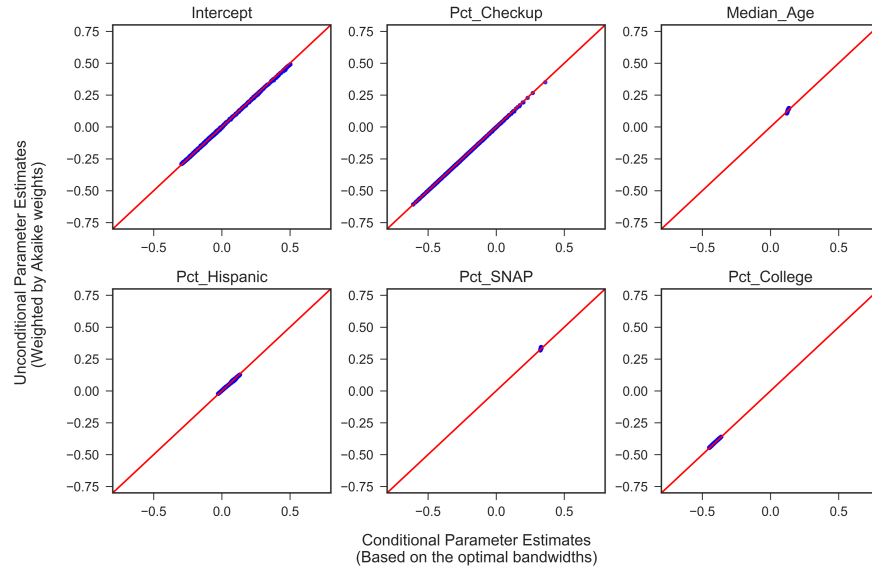


Figure 3.10. Comparison of conditional parameter estimates (based on the optimal bandwidths) with unconditional parameter estimates (calculated with Akaike weights).

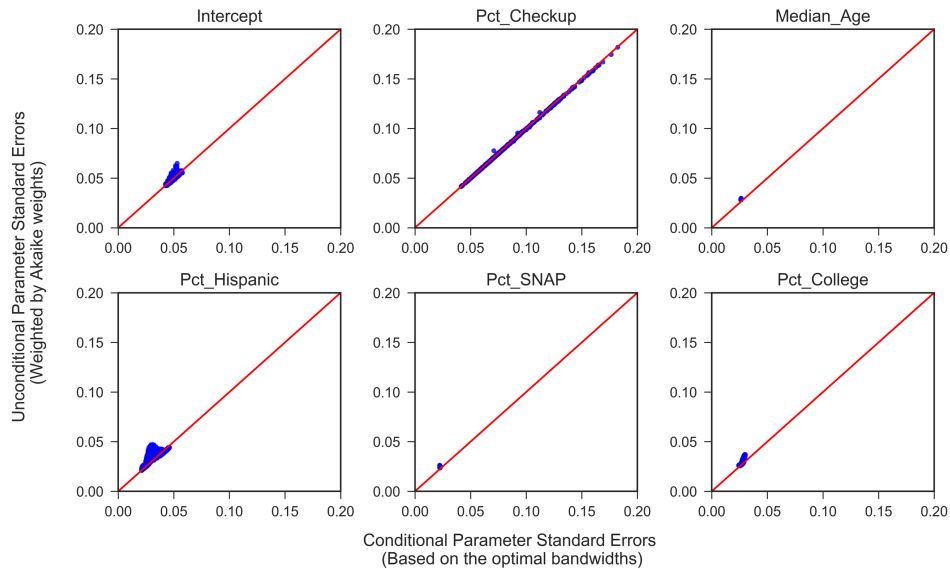


Figure 3.11. Comparison of conditional parameter standard errors (based on the optimal bandwidths) with unconditional parameter standard errors (calculated with Akaike weights).

6 Conclusions 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 multiscale 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 addition, 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 of the 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. It remains to be seen

whether the bandwidth uncertainty has generally little impact on parameter estimate uncertainty. 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. *International Journal of Geographical Information Science*.
- 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.
- Murakami, D., Lu, B., Harris, P., Brunsdon, C., Charlton, M., Nakaya, T., & Griffith, D. A. (2019). The importance of scale in spatially varying coefficient modeling. *Annals of the American Association of Geographers*, 109(1), 50-70.
- 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

Appendix A

This simulation example used the model described in Equation (10) (in the main 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 Figure 3.A1 and 3.A2).

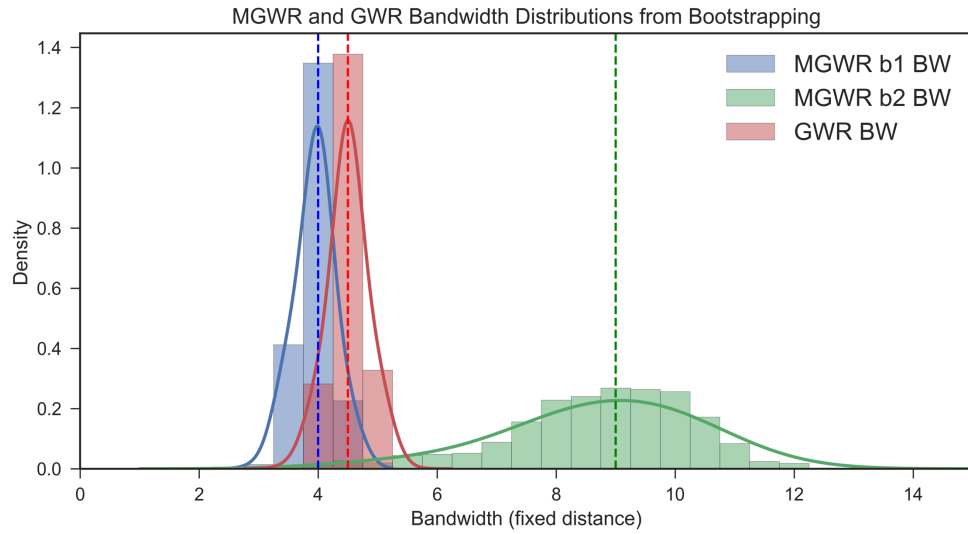


Figure 3.A1. Bootstrapping bandwidth distributions of the model described in Equation (10) with a distance-based fixed kernel.

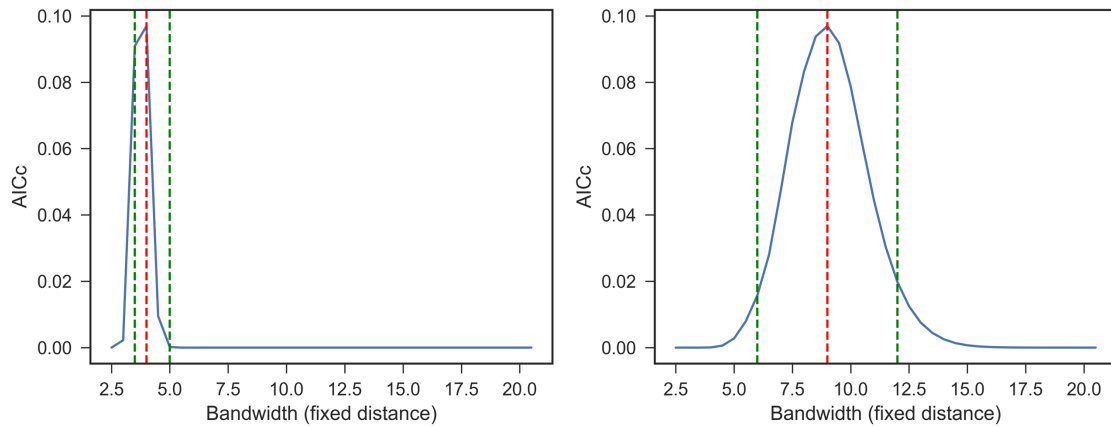


Figure 3.A2. Akaike weights for two processes shown in Figure 3.1 with a distance-based fixed kernel.

Appendix B

In this simulation example, we added a spatially varying intercept $\mathbf{b}_0$ (with heterogeneity between $\mathbf{b}_1$ and $\mathbf{b}_2$) to the model described in Equation (10) (in the main text), and the simulated processes can be seen in Figure 3.B1. The MGWR bandwidth distributions can be seen in Figure 3.B2 accordingly.

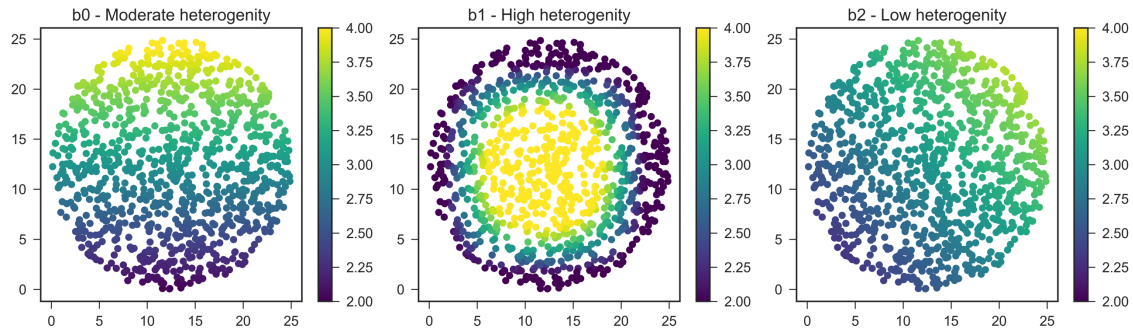


Figure 3.B1. Simulated processes.

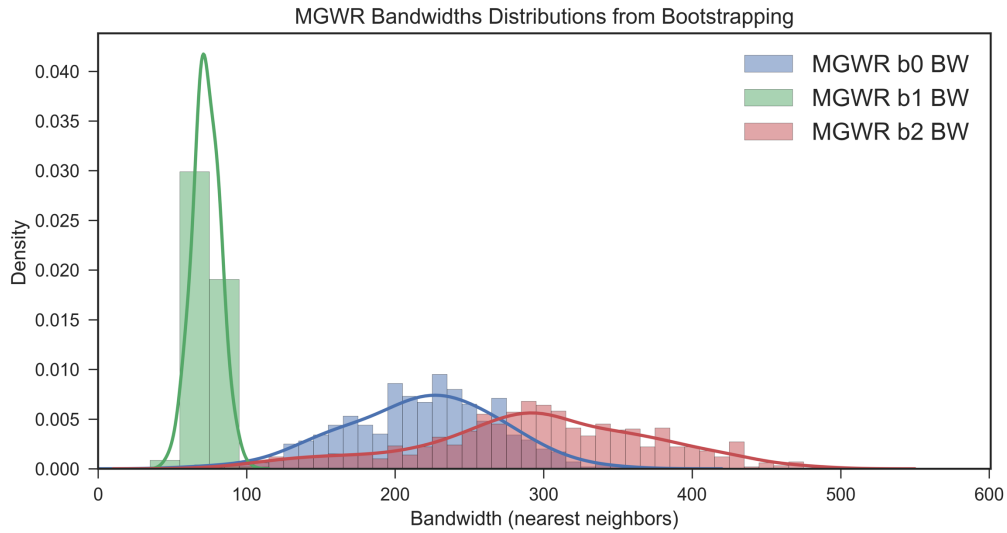


Figure 3.B2. Optimal bandwidth distributions for simulated processes in Figure 3.B1 under bootstrapping.

Appendix C

In this simulation example, we rotate the spatial process $\mathbf{b}_2$ (Figure 3.2 in the main text) to reproduce four spatial processes with different orientations but each having the same degree of heterogeneity, as illustrated in Figure 3.C1. 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 3.C2); bandwidth variation is a function solely of the scale over which the processes operate.

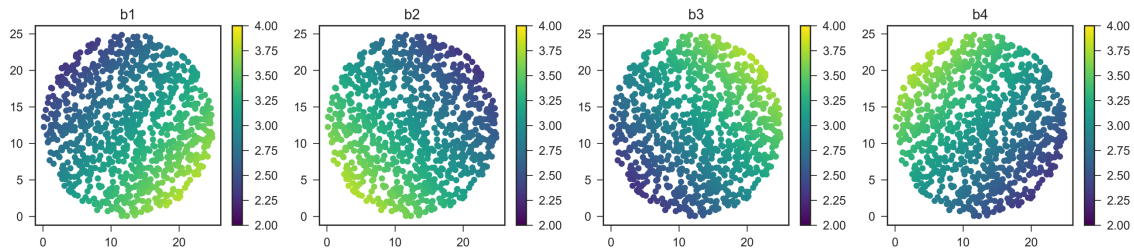


Figure 3.C1. Four generated spatial processes with the same degree of heterogeneity.

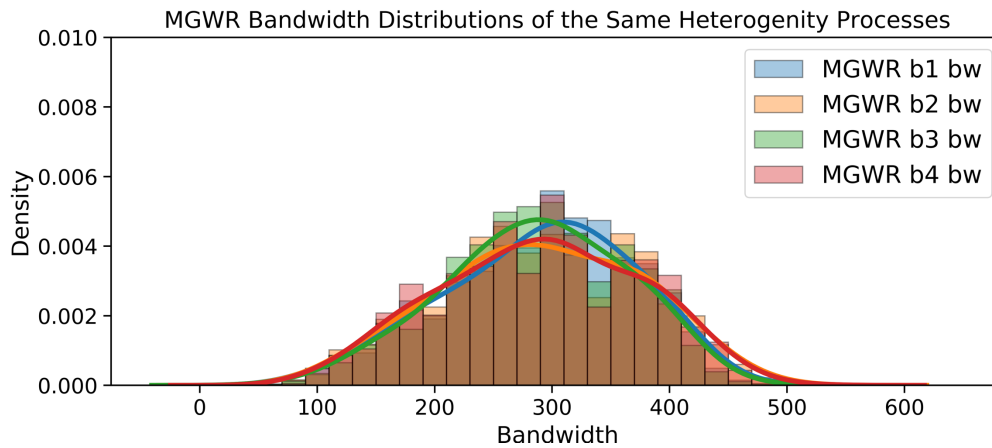


Figure 3.C2. Optimal bandwidth distributions of four same heterogeneity processes in Figure 3.C1 under bootstrapping.

CHAPTER 4
THE DYNAMICS OF VOTER PREFERENCE INFLUENCES ACROSS THREE U.S.
PRESIDENTIAL ELECTIONS, 2008, 2012 AND 2016: A MULTISCALE
GEOGRAPHICALLY WEIGHTED REGRESSION ANALYSIS

Abstract

Political and social processes that shape people's voting preferences might be linked with geographical location, varying from place to place, and operating at local, regional and national scales. If this is the case then traditional global models which try to identify the determinants of voter preferences will fail to capture such variations and will be mis-specified, severely in some cases. Here, we use a local modeling technique, multiscale geographically weighted regression (MGWR), to examine spatial variations in the influences of county-level socioeconomic factors on voter preference during the 2008-2016 US Presidential elections. A previous study (Fotheringham et al. 2020) has investigated such variations for one election – that of the 2016 U.S. Presidential election – and reported significant spatial variations in the determinants of voter preferences. A primary goal of this paper is to look into the temporal stability of these spatial variations over three Presidential elections to see if there is stability in the spatial patterns in the way in which socio-economic factors affect voter preferences. This is important because if there is little stability over time, the results in any one election may have limited meaning. Of particular interest is the temporal stability in the patterns of the local intercepts which measure the strength of the geographic context in affecting voter preferences. A secondary focus is on examining the determinants of voter preferences over time in several key counties in Pennsylvania, Wisconsin, and Michigan which played a vital role in influencing the outcome of the election in 2016.

Keywords: *Election, voter preference; multiscale geographically weighted regression, scale*

1 Introduction

Electoral geography is primarily concerned with mapping and explaining the uneven distribution of election outcomes. Examination of the social and political processes that shape voters' political preferences has been carried out systematically since the 1960s (e.g. Cox 1969), and there is a substantial body of literature focused on the spatial nature of elections from both compositional (characteristics of individuals) and contextual (place) effects (Taylor 1973; Taylor and Johnson 1979; Johnson et al., 1990; Warf and Leib; Agnew, 1994; Agnew, 2014). Despite King (1996)'s view that political behavior can be solely explained by individual factors, political geographers often believe that location and place play a vital role in shaping voters' preferences, and that political and social relationships are not uniform over space. The latter view has been reinforced by analyses using survey and polling data which suggest that individual characteristics are not sufficient to fully explain voting preferences, and there exists a certain amount of dependency between voting behavior and location and place. (O'Loughlin et al., 1994; Flint, 1996; Agnew, 1996; Sui and Hugil, 2002).

Recognizing that spatial variations in the determinants of voting behavior might exist, the use of classic global regression models, such as Ordinary Least Squares (OLS), seems overly simplistic because such models provide no insights into possible spatially varying behavior. To account for such spatial heterogeneity in processes, Geographically Weighted Regression (GWR) was developed (Brunsdon et al., 1996; Fotheringham et al., 2002) and has been shown to be a useful technique in electoral geography to explore localized relationships and geographic variations in the determinants of voting behavior. For example, Shin and Agnew (2008) applied GWR to electoral and census data for a case study of the Italian Lega Nord Party and stated that it is crucial to take spatial structures into consideration when conducting statistical analysis of voting patterns. Darmofal (2008) used GWR to account for the spatially varying effects of voting

behavior and demographics during the 1928-1936 Democratic realignment period. Cho and Gimpel (2009) looked at the local impact of economic hardship in the 2008 presidential election from a GWR model. Warf (2011) reported the non-stationary influence of class, ethnicity, and region on Obama's support in the 2008 presidential election. Additionally, Manley and Demšar (2015) analyzed the spatial relationship between turnout and socio-economic and political variables in the 2012 London mayoral election. All these studies support the notion that electoral behavior is shaped by localized political and social processes.

Despite the fact that GWR is able to capture localized relationships, the model is restricted by assuming that spatial variation in relationships operates at the same spatial scale for all processes by employing only one spatial weight matrix to account for the spatial structure. This is probably unrealistic in many real-world scenarios. What motivates and influences an individual to cast a particular vote is the consequence of a wide range of processes that may vary over space and operate at different geographic scales, from the very local to the regional or the global (Miller 1994). More specifically, Johnston and Pattie (2006) point out eight spatial scales, from the household to the neighborhood to the country that may have a compositional effect on people's voting behavior (Forest, 2017). This issue has been recognized in the spatial modeling literature and a new model developed by Fotheringham et al. (2017) termed, multiscale geographically weighted regression (MGWR), removes such an assumption and allows the estimation of multiscale processes in a single model. This is achieved by using a vector of covariate-specific bandwidths in MGWR in contrast to only one single bandwidth for all covariates in GWR. In this way, MGWR is able to account for the potentially different spatial structure of each process in the model. The covariate-specific bandwidths control the amount of bias and variance in the parameter estimates of each covariate separately. The estimated optimized bandwidths are explicit indicators of how the spatial processes operate over space

(Fotheringham et al., 2017). A covariate with a small optimized bandwidth indicates that the associated spatial process varies over relatively short distances, while a covariate with a large bandwidth suggests the associated process exhibits variation over large distances or may be constant over space. Consequently, the optimized covariate-specific bandwidths obtained in the calibration of an MGWR model measure and quantify the relative spatial scale over which each process operates. Recent research in different fields ranging from public health to environmental analysis supports the view that MGWR is more accurate than GWR (Fotheringham et al., 2019; Cupido et al., 2019; Yang et al., 2019; Oshan et al., 2020).

A recent study by Fotheringham et al. (2020) applied MGWR to voter preferences in the 2016 US Presidential election and found that the model performs very well in explaining the spatially varying socio-economic processes that operate at different spatial scales. Also, they demonstrate that the influence of spatial context in shaping voter preferences can be quantitatively separated from the influences of other socio-economic factors and is shown to be a significant factor in determining voter preference. It remains uncertain, however, whether the identified spatial variations in the determinants of voter preference are relatively stable over time. Therefore, in this paper, we extend the study of Fotheringham et al. (2020) and apply MGWR to the most recent three presidential elections in the period of 2008-2016 with a focus on the spatiotemporal changes of the factors in determining voters' preferences. Assessing the temporal stability of the determinants of voter preferences helps us understand the historical geography of presidential elections and also offers useful insights into future elections. If the spatial variation in the determinants of voter preferences is relatively stable over time, this generates greater confidence in the conclusion that the spatial patterns we observe reflect real processes and are unlikely to be artifacts of the data in any one-time period.

Further, by using MGWR we are able to understand the processes operating around each county. Trump's victory over Clinton in 2016 has been largely attributed to his marginal win in three states: Pennsylvania, Michigan, and Wisconsin. Specifically, Trump won Pennsylvania by 0.7 percent points (44,292 votes), Wisconsin by 0.7 points (22,748 votes), and Michigan by 0.2 points (10,704 votes). If Clinton had won all three states, she would have won the Electoral College votes 278 to 260. Therefore, counties in Pennsylvania, Wisconsin, and Michigan are critical in deciding the 2016 election outcome. Another focus of this paper is to examine the changes in both socio-economic make-ups and in which socio-economic factors affect voter preference across the three elections.

Accordingly, the paper proceeds as follows. In Sections 2 and 3, the data and the methods used are introduced. Section 4 discusses the comparison of the OLS model results with the MGWR model results nationwide, and then we focus on the local MGWR results in Pennsylvania, Wisconsin, and Michigan. The paper concludes in Section 5.

2 Data Used

The county-level election data from 2008-2016 were retrieved from the MIT Election lab website¹³. The dependent variable in the regression models is defined as the percentage share of the vote that went to the Democratic Party out of the votes that went to either the Democratic Party or the Republican Party in each county. That is, any votes for the third-party candidates are ignored in the calculation of the percentage of Democratic vote used in this study. Maps of this two-way percentage of Democratic vote for the contiguous United States (48 states plus the District of Columbia) are shown in Figure 4.1, and counties with blue (red) colors are those where the majority of voters in the county voted for the Democratic (Republican) Party. Over the three

¹³ The data were retrieved from: <https://electionlab.mit.edu>

time periods, the patterns are very similar but there was a steady erosion of the overall two-way percentage vote for the Democratic Party from 53.7% in 2008, 52.0% in 2012 to 51.1% in 2016.

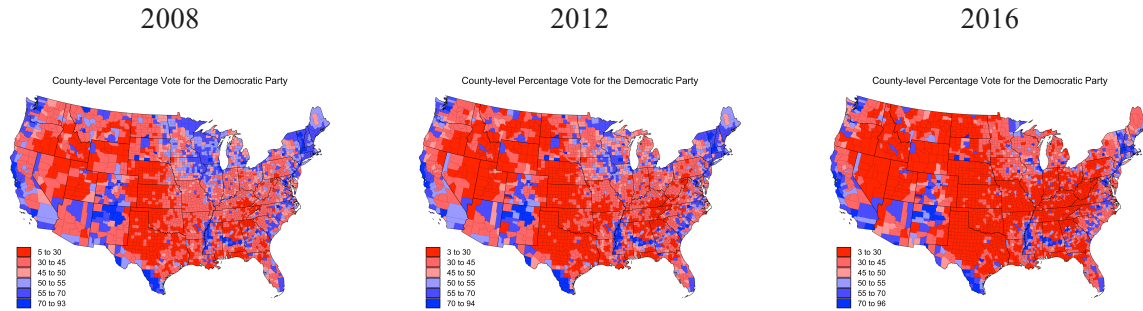


Figure 4.1. Observed county-level percentage of vote for the Democratic Party in the straight fight against the Republican Party for 2008, 2012 and 2016.

From the literature and mainstream media, certain socio-demographic and economic variables such as gender, race/ethnicity, age, income and education, are generally considered to have a direct impact on voting preferences (Degan and Merlo, 2011; Leighley and Nagler, 2013; Scala et al., 2015; Tyson and Maniam, 2016). We obtained the county-level independent variables from the United States Census Bureau's American Community Survey (ACS) 5-year estimate datasets (2005-2009¹⁴, 2008-2012, 2012-2016) for the corresponding three presidential election years. The descriptions of the independent variables are shown in Table 4.1. To mitigate the sampling bias of small counties in the survey, counties with fewer than 5,000 population are removed, and this analysis only considers the contiguous United States. Both dependent and independent variables are standardized to have mean of zero and variance of 1 for each year. This allows the regression coefficients to be comparable across variables and over time, and the standardization is recommended in MGWR analysis to produce comparable estimated covariate-

¹⁴ The earliest available ACS 5-year estimate dataset (2005-2009) was used for the 2008 election.

specific bandwidths (Fotheringham et al., 2017).

Table 4.1. The list of independent variables used in this study.

Variables	Descriptions
Sex_ratio	Sex ratio (the ratio of males to females)
Pct_age_18_29	Percentage of population aged 18 to 29
Pct_age_65	Percentage of population aged 65 and over
Pct_Black	Percentage of Black or African American alone
Pct_Hispanic	Percentage of Hispanic
Median_income	Median household income
Pct_Bachelor	Percentage of population with Bachelor's degree or higher
Gini	Gini Index
Pct_Manuf	Percentage of population employed in manufacturing industry
Ln(pop_den)	Ln of population density (persons per square mile)
Pct_3rd_Party	Percentage of third-party vote
Turnout	Voter turnout
Pct_FB	Percentage of foreign-born population
Pct_insured	Percentage of population with health insurance coverage

3 Methods

For each of the three election years, we build two regression models: (i) a global OLS model; and (ii) a local MGWR model to examine the county-level associations between voting preferences and socio-demographic and socio-economic factors. Each OLS model is specified as:

$$y_{it} = \beta_{0t} + \sum_j \beta_{jt} x_{ijt} + \varepsilon_{it}$$

where y_{it} is the percentage of Democratic vote for county i at year $t \in \{2008, 2012, 2016\}$, β_{jt} is the parameter for the j^{th} covariate x_{ijt} for county i at year t , and ε_{it} is the random error for county i at year t . Because the variables are all standardized, the intercept β_{0t} is zero.

The MGWR model has location-specific parameters and is formulated as

$$y_{it} = \beta_{bw_0t}(u_i, v_i) + \sum_j \beta_{bw_jt}(u_i, v_i)x_{ijt} + \varepsilon_{it}$$

where y_{it} is the dependent variable for county i at year t , (u_i, v_i) are the centroid coordinates of the county i , $\beta_{bw_0t}(u_i, v_i)$ is the local intercept for county i at year t , $\beta_{bw_jt}(u_i, v_i)$ is the parameter for the j^{th} independent variable x_{ijt} for county i at year t , and ε_{it} is the random error for county i at year t . The covariate-specific bandwidths are noted as $bw_0, bw_1, \dots, bw_j$.

MGWR local parameters are estimated along with the covariate-specific bandwidths $bw_0, bw_1, \dots, bw_j$, which allows more flexibility than only single bandwidth used in GWR. The estimation process of MGWR involves a back-fitting procedure following the Generalized Additive Model (GAM) framework (Hastie and Tibshirani, 2002; Fotheringham, et al., 2017). We first initialize the parameter estimates from a classic GWR model, then back-fitting starts from the first to the last covariate by calibrating univariate GWR against the partial additive components plus the current residuals as the dependent variable. In each univariate GWR, a covariate-specific optimal bandwidth is estimated. The back-fitting stops when the relative change in the parameter estimates is smaller than 1×10^{-5} . Then we can obtain the vector of covariate-specific bandwidths and parameter estimates from the last iteration of the back-fitting. The inference is computed following the details in Yu et al. (2019) to obtain the standard errors of the parameter estimates and other model diagnostics such as model AICc. We also computed the Akaike weights-based 95% bandwidth confidence intervals (CIs) following the approaches

introduced in Li et al. (2020). These CIs for the bandwidths account for the uncertainties in the model and make bandwidths comparable to infer the significance of any differences in spatial scales across processes. In this study, we use the adaptive bi-square kernel which uses the number of nearest neighbors as the unit of the bandwidth, which means that the spatial weight equals 1 at each regression location and decays to zero at the bandwidth (Fotheringham et al., 2002). Counties outside the bandwidth have zero weights. Both the OLS and MGWR models are calibrated by the MGWR 2.2 software¹⁵ which was developed by the Spatial Research Analysis Center at Arizona State University (Oshan et al., 2019; Li and Fotheringham 2020).

4 Results

4.1 Global OLS model

The global model calibration results are shown in Table 4.2 and depict the relationships between Democratic support and socio-economic variables for 2008, 2012 and 2016, respectively. These results are essentially averages across counties but we cannot see the individual county-level results. The R^2 value measures the goodness of fit of each model. The same model specification performs relatively better for 2016 with an R^2 value of 0.66, and performs modestly for 2008 and 2012 with R^2 values equal to 0.40 and 0.48 respectively. Residuals from the OLS models are mapped and shown in the left column of Figure 4.2, which depict significant (0.05 level) degrees of positive spatial autocorrelation based on the global Moran's I test as shown in Table 4.3. This raises caution when interpreting OLS results because the spatially correlated residuals violate the key assumption in OLS that they need to be independent and identically distributed (i.i.d). As a result, OLS regression coefficients may be biased and the type I error rates are inflated (Anselin and Bera, 1998, Dormann, 2007; Kuhn, 2007). To partially remedy this, we set a high significance threshold to be 0.001 in Table 4.2.

¹⁵ This software is freely available at <https://sgsup.asu.edu/sparc/mgwr>.

Since the variables are standardized, the estimates of the intercept are zero in each year and the parameter estimates (standardized coefficients) shown in the table can be interpreted as the relative importance of the associated variable affecting voter preference for the Democratic Party. Race/ethnicity (Pct_Black and Pct_Hispanic) and education (Pct_Bachelor) are the most positive variables in all three years, indicating that the county level vote for the Democratic Party increases as the percentage of Black and Hispanic population and college graduates increase in a county. Median household income, on the other hand, has the most negative association with the Democratic votes, which indicates increasing median household income in a county is associated with increasing proportions of votes for the Republican Party. Other consistently positive relationships with the Democratic Party votes are population density, percentage of third-party votes, and the turnout rate. The remaining variables are either insignificant (at 0.001 level) (e.g. percentage of people working in the manufacturing sector and percentage of young population) in all three years or only marginally significant (Gini, gender, and percentage of elderly) in some years and insignificant in other years.

Table 4.2. OLS regression coefficients for 2008, 2012, and 2016 elections.

Variables	2008	2012	2016
Intercept	0.00	0.00	0.00
Sex_Ratio	0.03	0.04	0.01
Pct_Black	0.36*	0.41*	0.53*
Pct_Hispanic	0.27*	0.28*	0.28*
Pct_Bachelor	0.29*	0.25*	0.43*
Median_Income	-0.48*	-0.46*	-0.31*
Pct_Age_65	-0.06	-0.06	-0.01
Pct_Age_18_29	-0.04	-0.02	-0.02
Gini	-0.06	-0.04	0.03
Pct_Manuf	-0.01	0.00	0.02
Ln(pop_den)	0.28*	0.22*	0.17*
Pct_3rd_Party	0.19*	0.23*	0.16*
Turnout	0.29*	0.34*	0.17*
Pct_FB	0.18*	0.20*	0.19*
Pct_Insured	0.27*	0.23*	0.17*
R ²	0.40	0.48	0.66

Significance code *0.001

By examining the magnitudes of the standardized coefficients across three years, we can see the shifting relative importance of each factor on voting preferences from 2008 to 2016. For example, the impact of the percentage of black population within each county on voting for the Democratic Party increases from 2008 to 2016 with coefficients increasing from 0.36 in 2008 and 0.41 in 2012 to 0.53 in 2016. The impact of education jumps significantly in 2016 from the previous two years (0.43 in 2016 vs. 0.29 in 2008 and 0.25 in 2012), indicating that there is a wider educational gap in voting preferences between voters with and without a college degree. The impact of income drops in 2016 (-0.31 in 2016 vs. -0.46 in 2008 and -0.48 in 2012)

indicating that the voting preferences between low-income and high-income population groups is narrowing. Similarly, the impact of living in a populated area on voting for the Democratic Party, from the evidence of the estimates of the population density, drops in 2016. Moreover, a higher turnout rate favors the Democratic Party in 2008, and 2012, but this positive impact drops in 2016, which is also supported by the fact that Democratic turnout rate is decreasing from 33.7% in 2008, 30.6% in 2012, down to 26.5% in the 2016 election¹⁶. Other factors remain quite stable over the three periods.

MGWR models are calibrated for all three Presidential elections and diagnostics of each of the models are shown in Table 4.3 for comparison with the OLS results. The MGWR models outperform the OLS models for all three years with R^2 values exceeding 0.90. The superiority of the MGWR models is also confirmed by their much smaller AICc values. Maps of residuals from MGWR models are compared with these from OLS models in Figure 4.2. By using the same color scheme, it is clearly seen that the magnitude of the residuals is generally much smaller for the three MGWR models. In addition, MGWR residuals have random patterns for all three years. This suggests that there is a great deal of non-stationarity in the processes affecting voter preference that are well captured by the MGWR model but which are inadequately modeled by the OLS model.

¹⁶ CNN Politics: <https://www.cnn.com/2016/11/11/politics/popular-vote-turnout-2016/index.html>

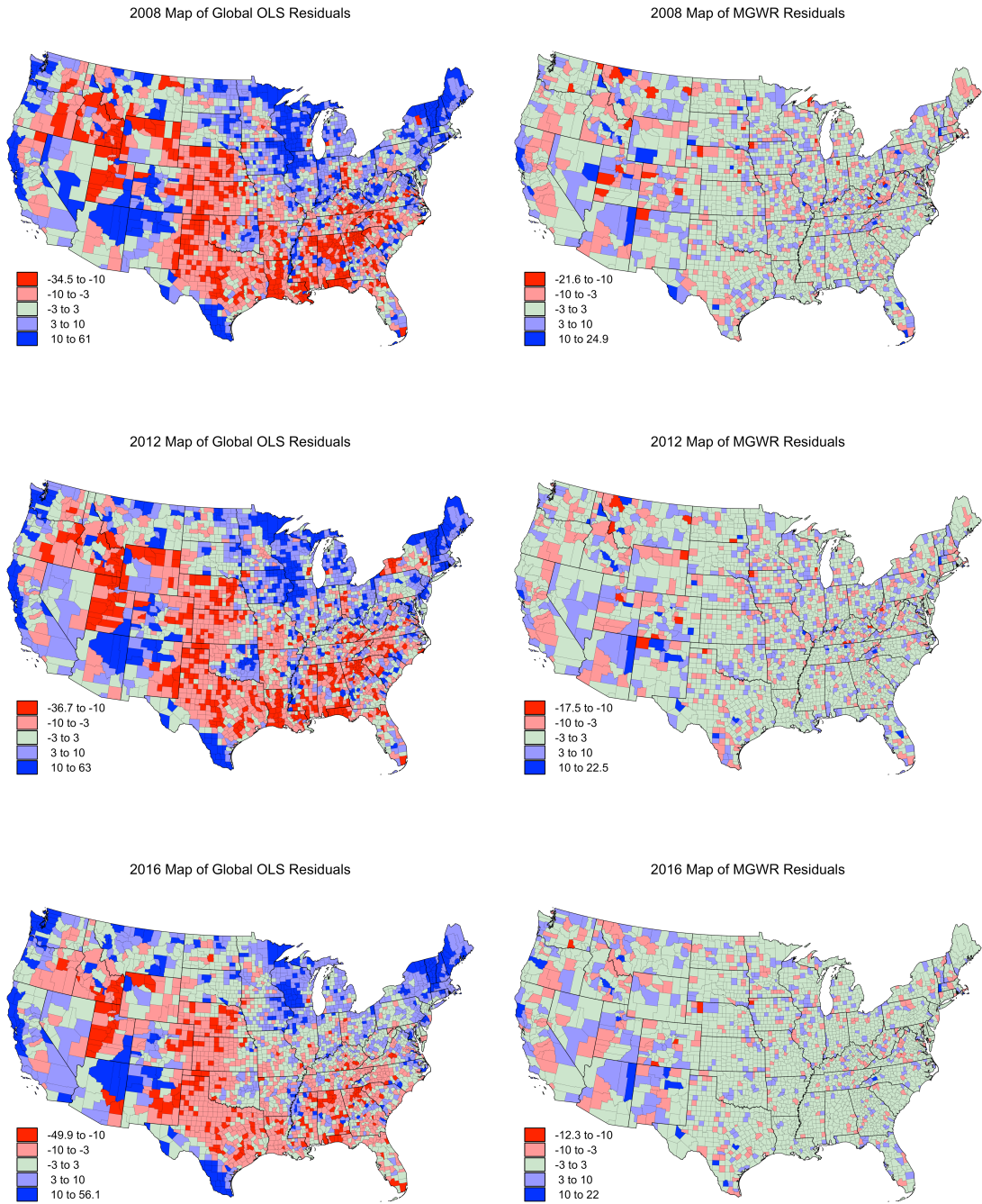


Figure 4.2. Maps of OLS residuals (left) and MGWR residuals (right) for 2008, 2012 and 2016.

Table 4.3. Comparison of model performance between OLS and MGWR.

	2008		2012		2016	
	OLS	MGWR	OLS	MGWR	OLS	MGWR
R²	0.40	0.91	0.48	0.92	0.66	0.95
AICc	6577.95	2447.99	6188.01	2139.44	5008.39	1044.59
Moran's I	0.03*	0.00	0.02*	0.00	0.03*	0.00
Spatial Pattern	Clustered	Random	Clustered	Random	Clustered	Random

* $p < 0.05$.

4.3 MGWR Bandwidths

As well as being a more accurate model of voter preference determinants, MGWR also provides optimized covariate-specific bandwidths, which measures the spatial scales at which the underlying processes vary. Table 4.4 summarizes the optimal bandwidths in terms of the number of nearest counties from which data are weighted along with the 95% confidence intervals (CI) of these optimized bandwidths. The bandwidth CIs are computed based on the methodology developed by Li et al. (2020), which uses Akaike weights to quantify the amount of bandwidth selection uncertainty taking account of the sampling variation of the data. Consequently, in this study, examining the covariate-specific bandwidths and CIs can help understand: (i) if there is any significant difference in the spatial scales over which different processes operate when comparing covariate-specific bandwidths for a particular election, and (ii) if there is any significant change in the spatial variations in the parameter estimates when comparing bandwidths for a particular covariate across the three elections. If the two bandwidth CIs being compared do not overlap, this indicates a strong evidence that the difference in the covariate-specific bandwidths is statistically significant. For example, in 2016, the bandwidth for median income is significantly larger than the bandwidth for the percentage of black population with non-

overlapping confidence intervals ((2158, 2717) vs. (43, 45)), suggesting that the influence of median income on voter preferences operates at a significantly larger spatial scale than that of the percentage of black population. Temporally, median income has consistently large bandwidths in all three elections with similar CIs, indicating that income has a consistent global impact on voting preference. The relationship between the proportion of the black population in a county and the Democratic votes within the county also has consistently small bandwidths in all three elections with similar CIs. This evidence shows that the influence of African American voters on the Democratic vote is highly variable over space but consistently localized over time. Moreover, the intercept, percentage of college graduates, Gini coefficient, population density also have similar bandwidths in all three elections, suggesting that the spatial variations in these determinants of voter preferences are stable over time. The rest of the variables have significant changes in bandwidths across the three elections.

Table 4.4. Optimized covariate-specific bandwidth and 95% confidence interval in MGWR.

	2008		2012		2016	
Variables	Opt.	CI	Opt.	CI	Opt.	CI
Intercept	43	(43,44)	43	(43,44)	43	(43,45)
Sex_Ratio	2807	(2156,2808)	2813	(2409,2813)	603	(446,850)
Pct_Black	43	(43,45)	43	(43,45)	43	(43,45)
Pct_Hispanic	43	(43,47)	1314	(1101,1505)	543	(446,601)
Pct_Bachelor	195	(174,233)	210	(151,197)	208	(174,210)
Median_Income	2368	(1753,2560)	2090	(1755,2409)	2659	(2158,2717)
Pct_Age_65	85	(78,101)	101	(78,106)	656	(600,850)
Pct_Age_18_29	460	(387,542)	494	(387,542)	58	(51,65)
Gini	1363	(1099,1753)	1246	(946,1350)	763	(696,1100)
Pct_Manuf	2809	(2156,2809)	424	(387,542)	2809	(2158,2810)
Ln(pop_den)	410	(351,542)	383	(351,447)	387	(292,446)
Pct_3rd_Party	312	(292,446)	318	(292,388)	160	(137,173)
Turnout	2765	(2156,2787)	415	(387,542)	117	(110,137)
Pct_FB	2809	(2406,2809)	1448	(1101,1755)	1424	(1100,1754)
Pct_Insured	87	(78,101)	43	(43,47)	43	(43,45)

4.4 MGWR parameter estimates

For each covariate, we can map the local parameter estimates obtained from MGWR for each of the three elections to examine the stability of the spatially varying effects of each covariate on voter preference. There are 15 sets of local parameter estimates maps, and we present them in terms of the magnitude of their bandwidth from processes that are global (with large bandwidths) to processes that are highly local scale (with small bandwidths). Counties with positive or negative local parameter estimates are colored in blue and red accordingly, indicating

that an increase in a given variable would increase the share of the vote for the Democratic (blue) or Republican (red) Party in that county. The same color scale is employed for all the parameter estimates maps, and the parameter estimates with higher absolute magnitudes have a darker color, indicating they have relatively more influence over the voting preference. Local parameter estimates for counties with fewer than 5,000 population, which are not included in the model, are interpolated using the Inverse Distance Weighting (IDW) technique with a power of 2. Counties with insignificant parameter estimates (at the 95% level) are shaded in grey, and the significance has been adjusted to account for multiple hypothesis testing (da Silva and Fotheringham, 2016; Yu et al., 2019).

Figure 4.3 depicts the local parameter estimates for household median income. MGWR results are aligned with the OLS results. That is, across all three elections, median income is consistently, globally, significant, and negatively associated with voter preference for the Democratic Party with bandwidths above 2,000. This indicates that when median income increases, the proportion of votes cast for the Republican Party within a county increases, with virtually the same degree of sensitivity for all the counties in the United States. Across the three elections, the influence of income on voter preference remains relatively the same.

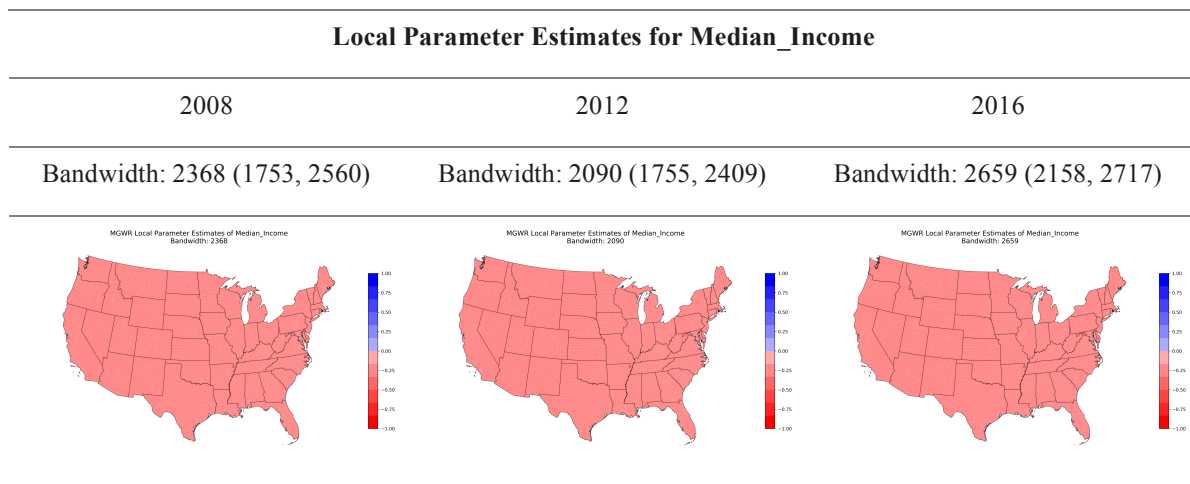


Figure 4.3. Local parameter estimates for median household income.

The local parameter estimates for sex ratio are shown in Figure 4.4. In 2008 and 2012, sex ratio has a global bandwidth indicating a national-level significant negative association on voting for the Democratic Party. That is, when there are more males than females in a county, voter share for the Democratic Party decreases in that county. However, in 2016, the influence of sex ratio is only significant in the northeast plus Michigan, Ohio, part of Illinois and Missouri, and is insignificant elsewhere. Local results show contradictory findings to that of the global model. In the OLS model, parameter estimates for sex ratio are positive for all three elections, whereas there are no significantly positive local estimates and in 2008 and 2012 all the local estimates were significantly negative. This is a clear indication of Simpson's Paradox in that a reverse relationship is seen in the analysis of the spatially aggregated data compared to the local modeling. The global results likely arise because in the parts of the country that generally favor the Republican Party, there are typically lower proportions of males and in parts of the country that favor the Democratic Party there are relatively higher proportions of males. This national trend is ignored in the calibration of local models which show that, locally, counties with higher proportions of males tend to vote Democrat.

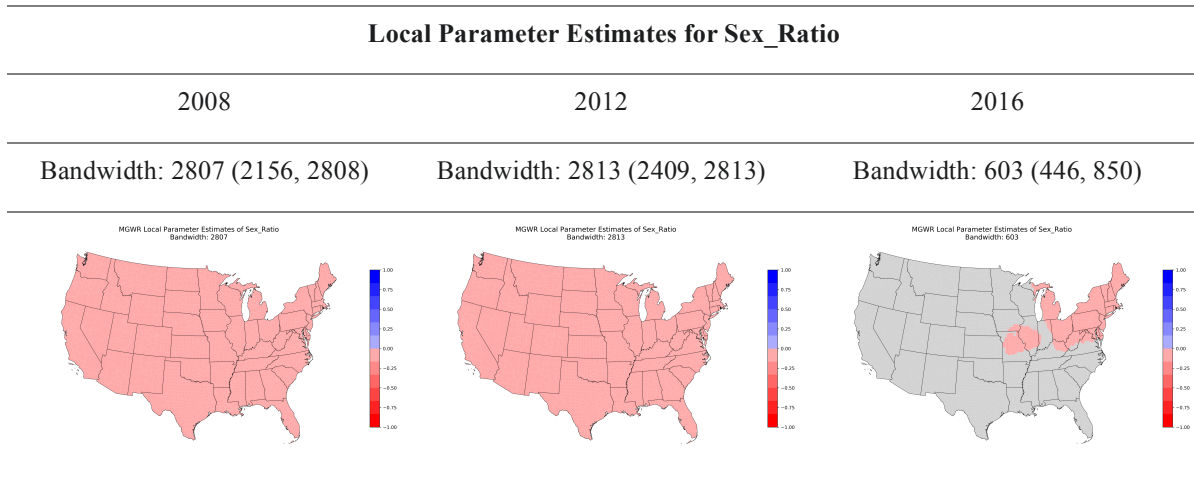


Figure 4.4. Local parameter estimates for sex ratio.

The local parameter estimates for the percentage of the population employed in the manufacturing industry are shown in Figure 4.5. Across the three elections, the local parameter estimates for the manufacturing employment are consistently negative, which indicates that as the proportion of manufacturing employment in a county increases, the share of the Republican vote increases. This influence is virtually constant across the US in 2008 and 2016 but in 2012 it was only found West and in Florida. One possible explanation for the spatial pattern in 2012 is that the insignificant counties are largely manufacturing dependent counties which experienced a great deal of distress due to the 2008 economic crisis. Another interesting aspect is that manufacturing workers are commonly perceived as having a preference for voting for the Democratic Party historically. However, the MGWR results indicate that when taking other factors into account at the local level, manufacturing employment has a negative association with Democratic vote share in the recent three elections.

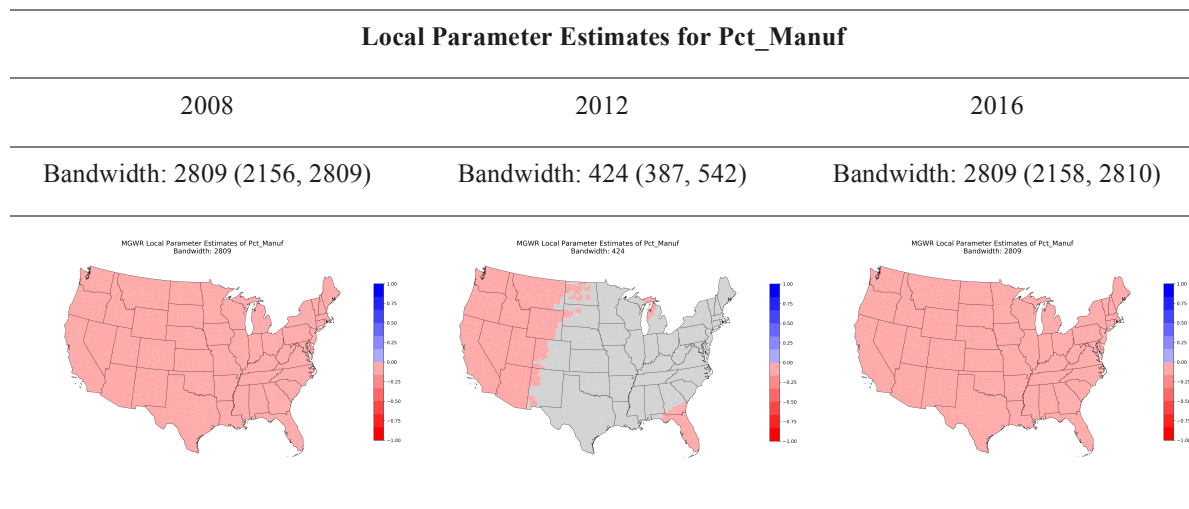


Figure 4.5. Local parameter estimates for the percentage of population employed in the manufacturing sector.

Figure 4.6 shows maps of local parameter estimates for the percentage of the foreign-born population in a county. In 2008, the proportion of foreign-born population within a county was not a significant predictor of the Democratic vote in any part of the country. In 2012 and 2016, however, the parameter estimates exhibit a regional pattern with the optimal bandwidth in both cases being close to 1,500 with a similar division between eastern and western parts of the country. In the 2012 election, counties in the west with higher proportions of foreign-born residents were significantly more likely to vote Republican with no significant relationship found in counties in the eastern half of the country. In the 2016 election, counties in the eastern half of the country with higher proportions of foreign-born residents were significantly more likely to vote for the Democratic Party whereas there was no significant relationship for the western counties. Consequently, the results from the last two elections are similar in this regard, foreign-born residents in the eastern part of the US appear to be more inclined to vote Democrat than in the west. It is unclear why this is so but one possibility is that the countries of origin are somewhat different for foreign-born residents of counties in the east, a greater proportion of

whom might be from Europe where many countries have decidedly more left-wing political leanings than in the US.

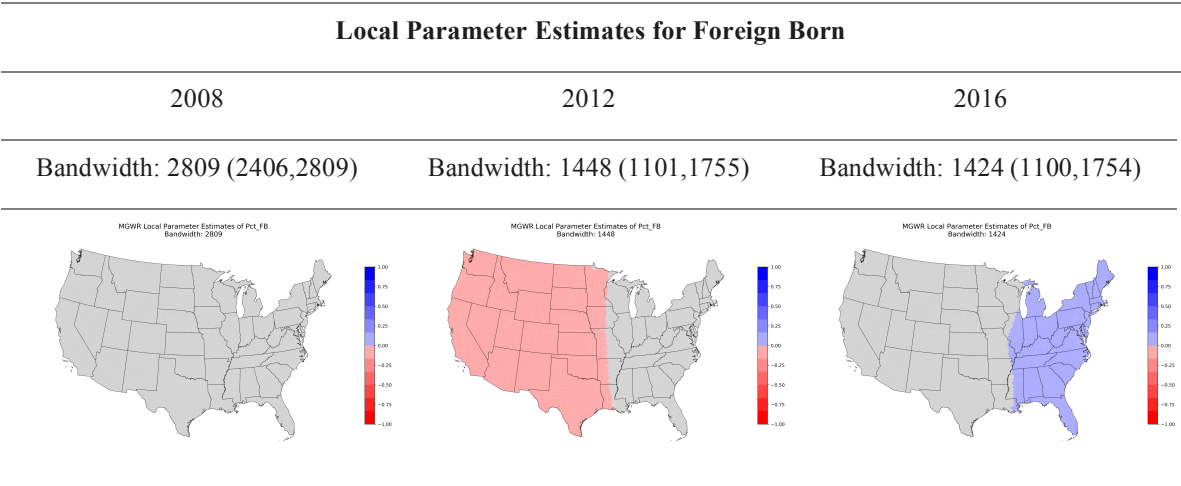


Figure 4.6. Local parameter estimates for the percentage of foreign-born population.

In Figure 4.7, the Gini coefficient, which is a measure of the degree of income disparity within each county, shows a regional relationship to voting preference with a bandwidth of approximately around 1,000. In all three elections, there is a significant negative association between income disparities and voting for the Democratic Party for the south-east. Interestingly this region of significant negative associations has been steadily shrinking whereas a region of significant positive association between income disparities and voting for the Democratic Party appeared in the West in 2012 and expanded in 2016. This could be a function of the general psychologies of residents in the south-east where there is a history of laissez-faire stemming from the original settlers from Europe and which may enhance a leaning to the Republican Party by voters with wealth in counties where there is also poverty. Conversely, residents of the western counties, those along the Pacific seaboard may have a more sympathetic view of the ‘have-nots’ within a county and this may strengthen their Democratic leanings. An alternative explanation

lies in the issue of voter registration laws within various states. The states in the south-east have typically made it more difficult for minorities to register to vote and a consequence of this would be that in counties with large income disparities, the portion of the population with the lower incomes, who would tend to vote Democrat, would be under-represented in the list of eligible voters. Consequently, counties in such states with high income disparities would have a relatively higher vote for the Republican Party. This is all speculation but provides an interesting avenue for further research.

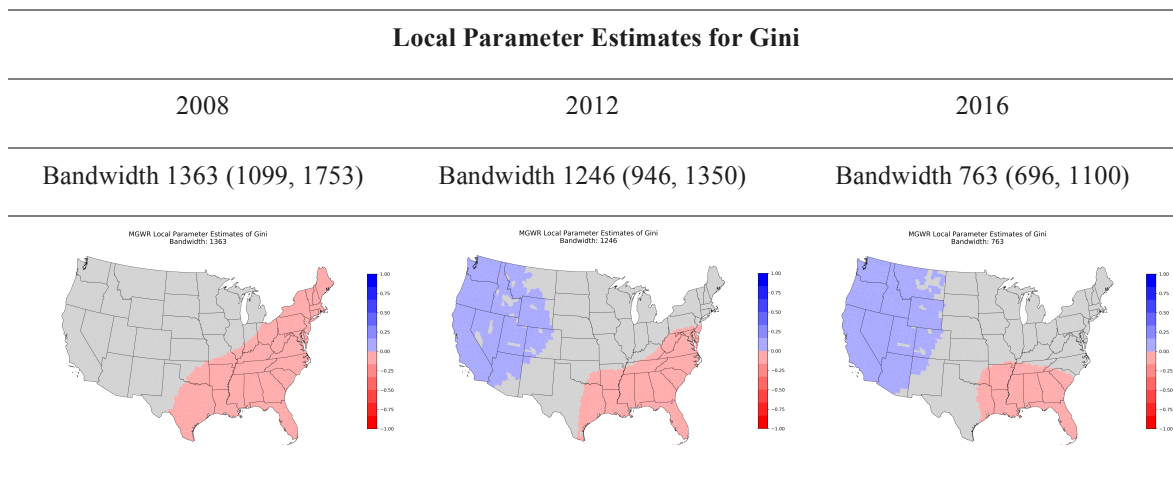


Figure 4.7. Local parameter estimates for the Gini index.

Figure 4.8 shows the local parameter estimates maps for the percentage of Hispanic population. In the last two elections, the impact of the Hispanic population on voting preference has been strong and consistent across the country with significant positive local parameter estimates being found in every county. The association of the proportion of Hispanics in each county with votes for the Democratic Party is stronger in the south and west where significant and long-term Hispanic populations can be found. Interestingly, in 2008, the relationship between the Hispanic population and Democratic vote was much more localized with a small bandwidth (43).

Significant positive associations between Hispanic population percentage and share of votes for the Democratic Party are only found in the southwestern states, including Arizona, New Mexico, Texas, which had large proportions of Hispanics as well as in counties in New York and New Jersey. There is also a significant positive relationship in some counties in the Deep South from Louisiana to South Carolina.

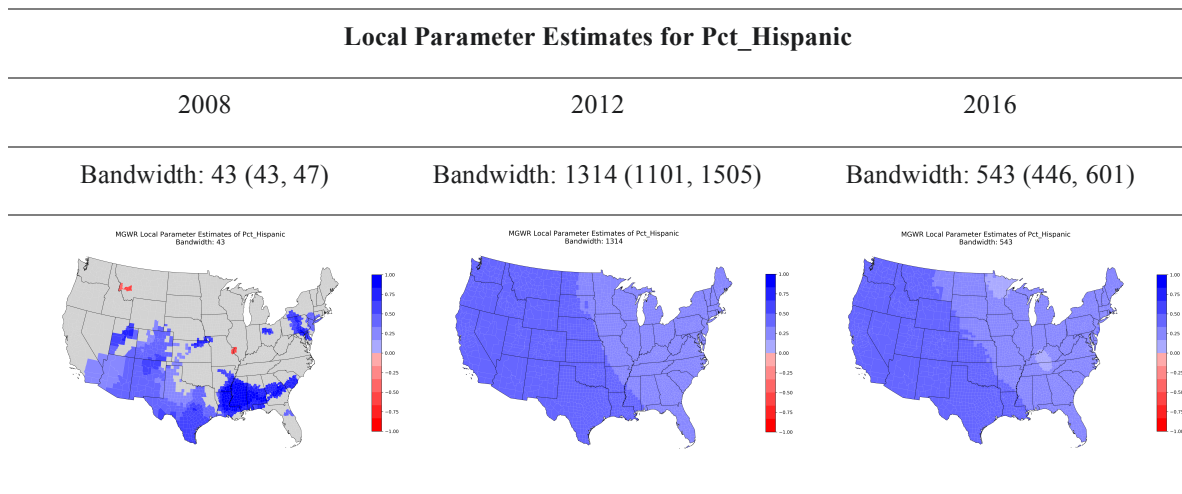


Figure 4.8. Local parameter estimates for the percentage of Hispanic population.

Figure 4.9 and Figure 4.10 show the local parameter estimates for the percentage of young (age 18-29) and elderly (age 65+) population, respectively. From both sets of maps, we can see the impact of young and elderly age groups on the voting for the Democratic Party shows largely regional variations. Everything else being equal, higher percentages of younger-age voters in western counties are associated with lower vote percentages for the Democratic Party. There is no significant impact of younger-age voters in much of the rest of the country except for a couple of interesting outliers in 2016 around Indiana University and Ohio State University where there is a significant positive relationship between younger voters and votes for the Democratic Party. This effect can also be seen in 2008 on a broader scale. It is interesting to note that the finding of

a significant negative relationship between younger voters and votes for the Democratic Party in many western counties is at odds with the media reporting of younger votes favoring the Democratic Party but here we have conditioned for many other factors such as education, income and ethnicity whereas many poles used by the media report largely unconditioned figures. It is further interesting to note the decline in the counties where a significant negative relationship was reported in 2016 compared to 2012 and 2008. The relationship between elderly voters and the Democratic share of votes is very consistent across all three elections and only exhibits a significant negative relationship for the western states. In counties in these states, increasing proportions of elderly voters are associated with decreasing shares of the vote for the Democratic Party. There is no significant relationship in the rest of the country.

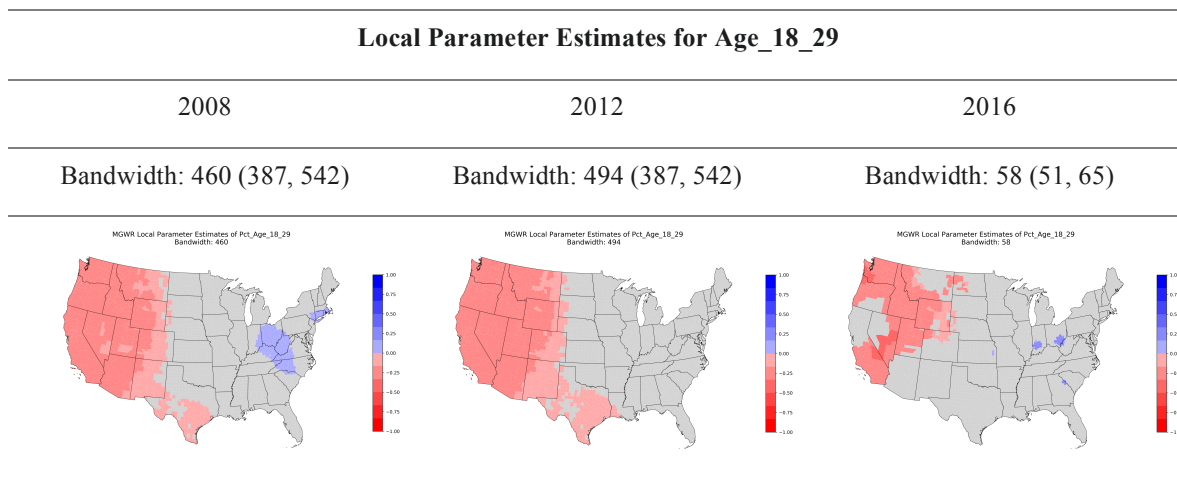


Figure 4.9. Local parameter estimates for the percentage of population aged between 18 and 29.

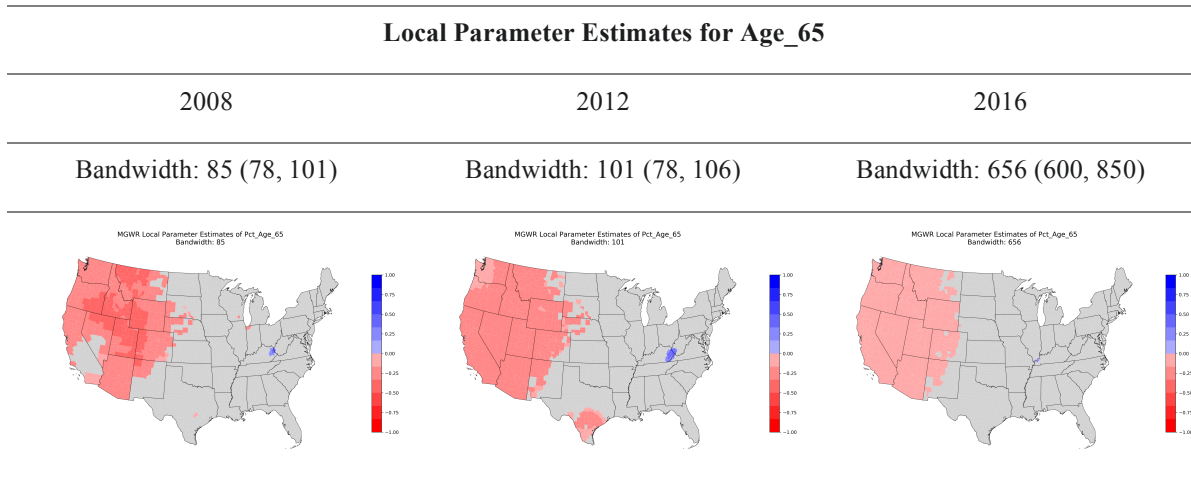


Figure 4.10. Local parameter estimates for the percentage of population aged above 65.

Figure 4.11 shows the association of the turnout rate (the proportion of those eligible to vote who actually voted) with people's voting preferences. In 2008, the Obama bandwagon effect meant that across the country, high turnout rates were associated with higher percentage votes for the Democratic Party. This effect was uniform as shown by the very large bandwidth. In 2012 enthusiasm for Obama had diminished and turnout had very little effect on voting preferences anywhere in the country, the only exceptions being in southern Florida and a few counties on the Illinois/Iowa/Missouri border where higher turnouts were beneficial to the Democratic Party. In 2016, the influence of voter turnout is different again with no significant impact being felt in the vast majority of counties but increasing voter turnout favored the Republican Party in counties in Montana, Idaho, and Utah but favored the Democratic Party in Florida, Louisiana, and south Texas.

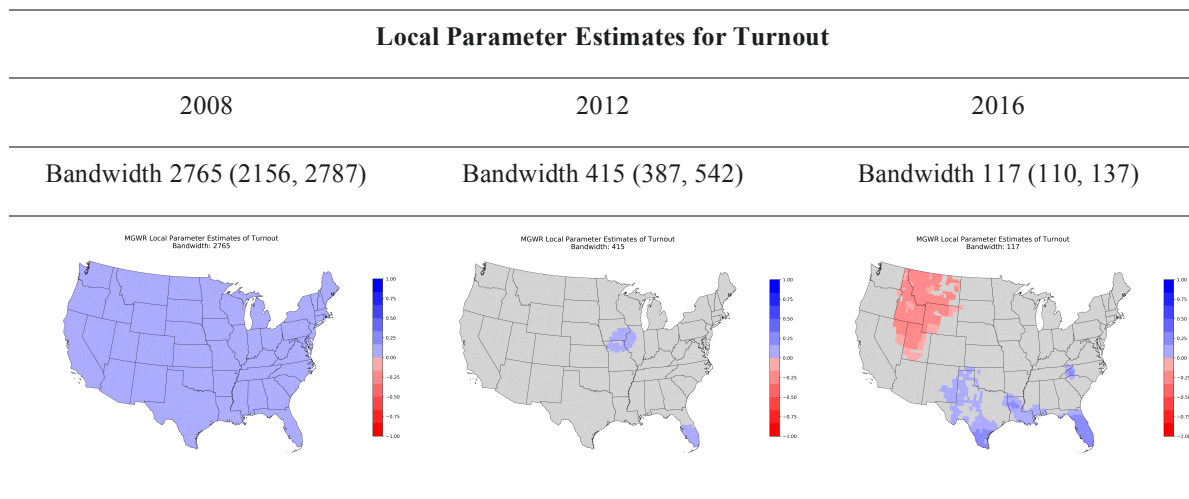


Figure 4.11. Local parameter estimates for the turnout rate.

Third-party voting is an interesting aspect of American elections. Essentially the vote is split between just two parties, the Democratic Party and the Republican Party. However, in some elections, a third-party candidate appears on the ballot who generates what can be a significant share of the votes. The spatial pattern of voting for third-party candidates is expected to be very election-specific as third-party candidates are generally ‘one-offs’. The local parameter estimates for the percentage of third-party votes cast in each county are shown in Figure 4.12. The covariate-specific bandwidth ranges approximately from 150 to 300, indicating a local-to-regional impact on voting preferences. In 2008, there is a significant positive association between the third-party vote and Democratic vote in the Northeast and the Southeast of the US. Most third-party votes were for either the independent candidate Ralph Nadar and the Libertarian candidate Bob Barr who received approximately 0.5% and 0.4% of the national vote, respectively. Interestingly, in no county in 2008 did third-party voting favor the Republican Party. In 2012, positive associations between third-party voting and votes for the Democratic Party exist in the West and the Southeast of the US when the primary third-party candidate was Gary Johnson of the Libertarian Party who received 1% of the total national vote. In 2016, the country

divided along third-party preferences with counties in New England, the Southeast, and much of the central US having a significant positive association between the third-party vote and the democratic vote, while in the counties within and surrounding Utah, there is a significant negative association. The third-party voting in counties within and near to Utah is interesting. Although the third-party candidate here was Evan McMullin, a conservative independent, it would appear that he drew votes away from the Democratic Party. Very likely this was an instance of Democrats voting strategically and for the Independent candidate who had a good chance of defeating Donald Trump (McMullin gained a 21.5% share of the votes in Utah). Consequently, the share of votes going to the Democratic Party decreased in counties where Evan McMullin was deemed to have a good chance of defeating Trump.

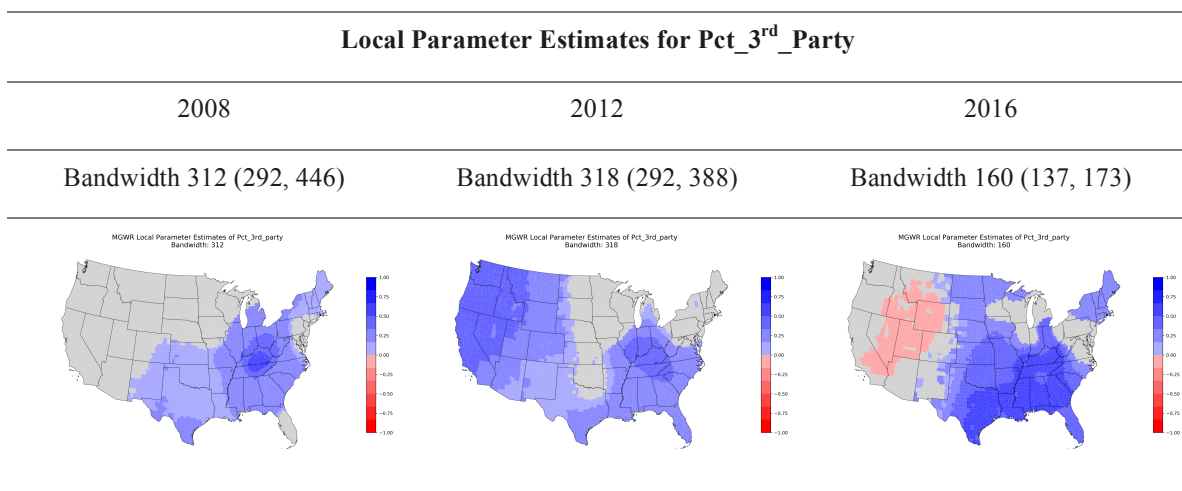


Figure 4.12. Local parameter estimates for the percentage of the third-party vote.

Increasing levels of education have a significantly positive influence on the Democratic vote across the nation, as shown in Figure 4.13: there is a strong positive relationship between the proportion of people with a Bachelor's degree in a county and the Democrat share of the vote in that county. This is a consistent result across all three elections but the 'education divide' in voter

preferences is getting stronger over time and varies across the country in magnitude, being stronger in counties in the Northeast, the West, and in the state of Florida.

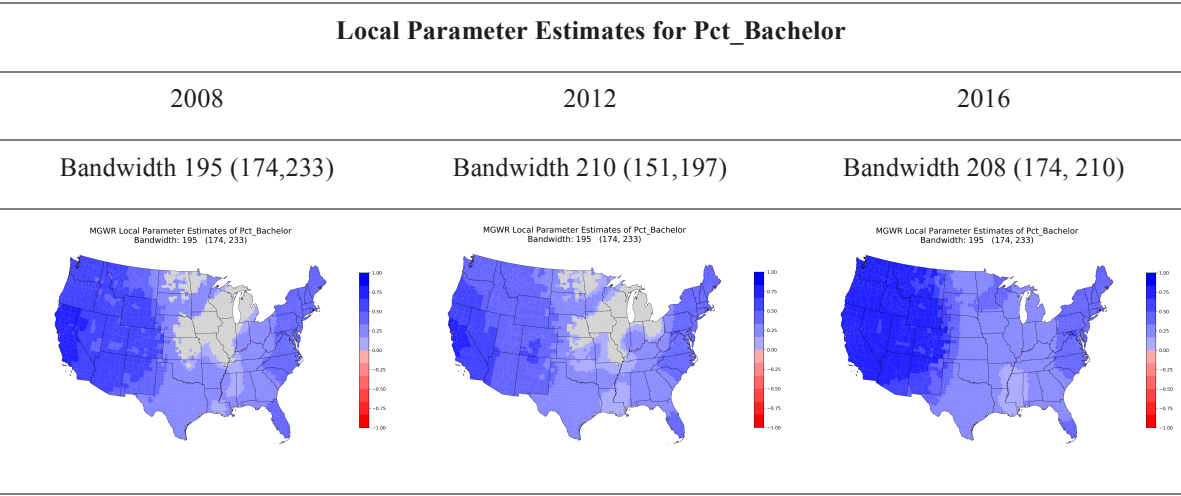


Figure 4.13. Local parameter estimates for the percentage of people with Bachelor’s degree or higher.

Figure 4.14 shows the local parameter estimates of log of the population density which are indicative of a rural/urban divide in voting preferences. With bandwidths close to 400 with similar confidence intervals, the association between population density and Democratic vote shows a broad and consistent spatial variation across the three elections. The urban/rural divide in voter preference is significant in counties in roughly half the country with counties in the north and west exhibiting a significant increased preference for the Democratic Party in cities compared to rural areas, *ceteris paribus*. In counties in the north-east and south-east, there is no significant difference in voter preference between urban and rural areas. The pattern is consistent across all three elections.

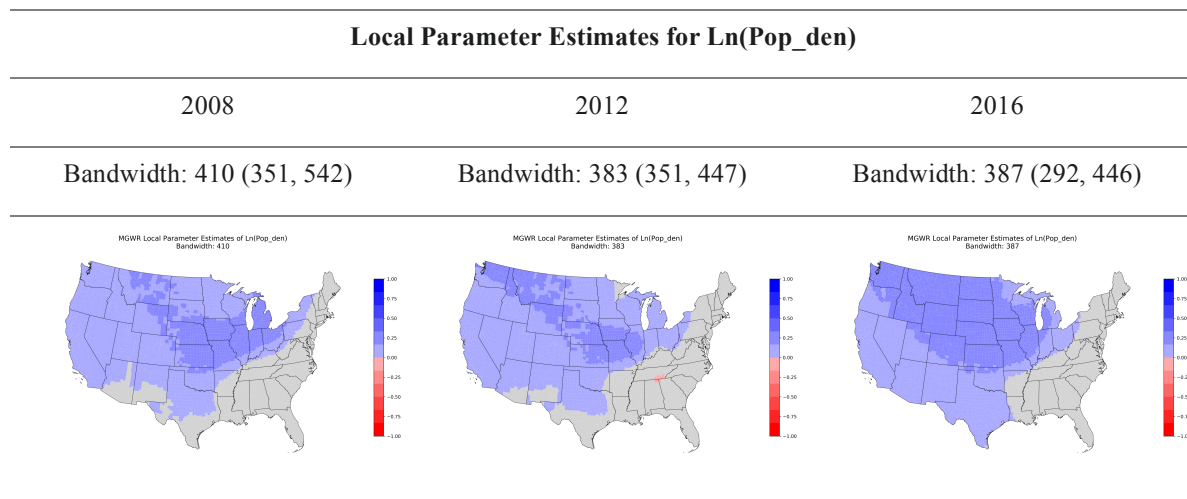


Figure 4.14. Local parameter estimates for the log of population density.

Local parameter estimates of percentage of people with health insurance are shown in Figure 4.15. Across the three years, the association of having health insurance with voting preference are localized with bandwidths under a 100. In 2008, voters in most western counties tended to have a preference for the Republican Party as the share of people with health insurance increased. An alternative view is that in counties with more uninsured people, there was a tendency to vote for the Democratic Party, probably Obama’s plan to reform health insurance in his campaign. In 2012 and 2016, the impact of health insurance is more limited and found only in counties in Arizona, New Mexico, the southern tip of Texas and areas around Montana. It is clear that having health insurance is a major predictor of voting Republican in Arizona, possibly because of the high proportion of retirees who have government-supplied insurance.

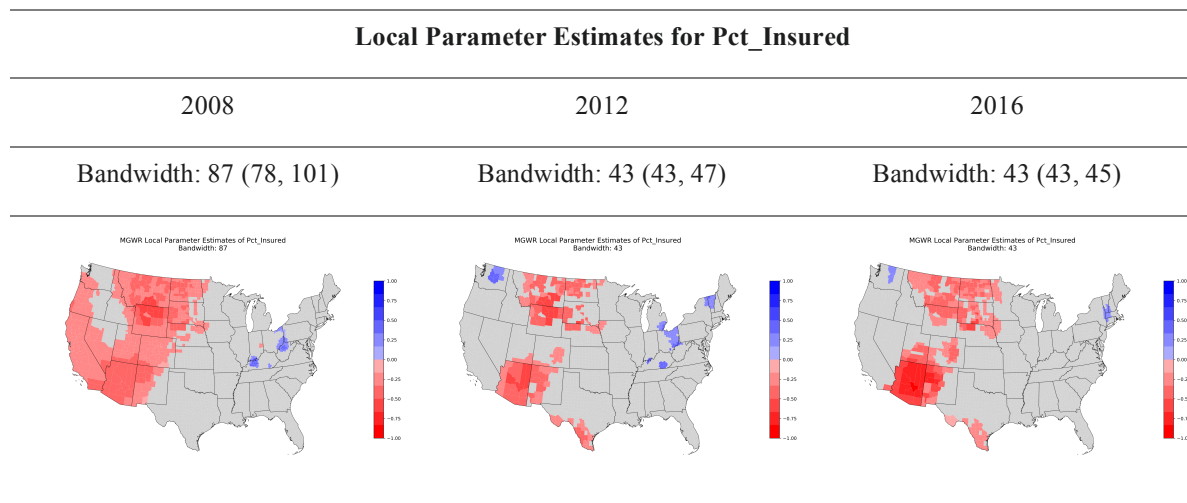


Figure 4.15. Local parameter estimates for the percentage of people with health insurance.

Figure 4.16 shows the impact of the percentage of African-Americans within each county on voter preferences. Across the nation, there is a strong and positive relationship between the percentage of African-Americans and the Democratic vote share. However, the intensity of this relationship is spatially varying, indicated by the small bandwidth suggesting a high degree of heterogeneity. The only counties where the percentage of African Americans does not have a significant impact on voter preferences are those where the proportion of African-Americans is very low such as those in Washington, Arizona, New Mexico, Colorado, Wisconsin, and New England. Again, the results are reassuringly consistent across all three elections.

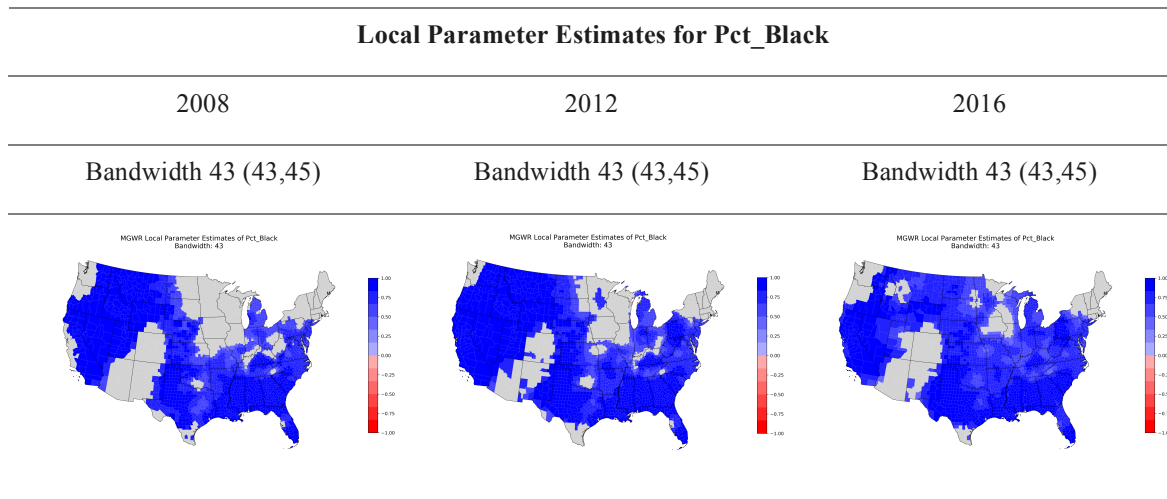


Figure 4.16. Local parameter estimates for the percentage of black population.

Local intercepts obtained in MGWR are important indicators of geographic context as they measure the intrinsic level of voter preference for the Democratic Party independent of all other contributing factors in the model. Fotheringham et al. (2020) interpret the local intercept as the intrinsic preference of voting for a political Party based on people’s location, and it quantitatively measures the contextual effects on voting for each county. In essence, the local intercept measures the preference for one Party over the other if each county had the ‘average’ population composition. Counties provided the geographic setting in which people are exposed to different social and physical influences that affect their voting preferences (Agnew and Shin, 2020). Some of the contextual influences are expected to be relatively stable over time because many factors are historically related. To examine whether the pattern of the spatial context remains stable over different elections, we compare the maps of local intercepts across the three elections. From Figure 4.17, we can see that the spatial patterns of local intercepts show a strong degree of consistency among the three elections, which means that the role of context in voting remains relatively similar. In all three elections, the West, Central North and the Northeast of the US are intrinsically blue counties (counties with an intrinsic leaning towards the Democratic

Party) whereas counties in the Central South and the South are red counties (counties with an intrinsic leaning towards the Republican Party). Counties with insignificant local intercepts do not show any intrinsic leaning towards one Party although they are not necessarily neutral in that they may have leanings towards either Party based on their population composition. That said, many of the counties in this ‘band of neutrality’, from Arizona in the West to Virginia in the East, forming a belt of swing battlegrounds that may lean to either Party in an election (Fotheringham et al., 2020). It is interesting to see that Florida, at least Southern Florida, does not show any intrinsic inclination towards the Democratic or the Republican Party in all three elections and this has long been a very key swing state in US elections. Despite the overall consistent nature of the spatial pattern of geographical context in voter preference across the US, there are several local and regional changes across the three elections. California and Oregon, for example, which are intrinsically blue states in 2008 and 2016 became neutral in 2012 and there is a contraction of the intrinsically Republican counties over time in Utah and Colorado.

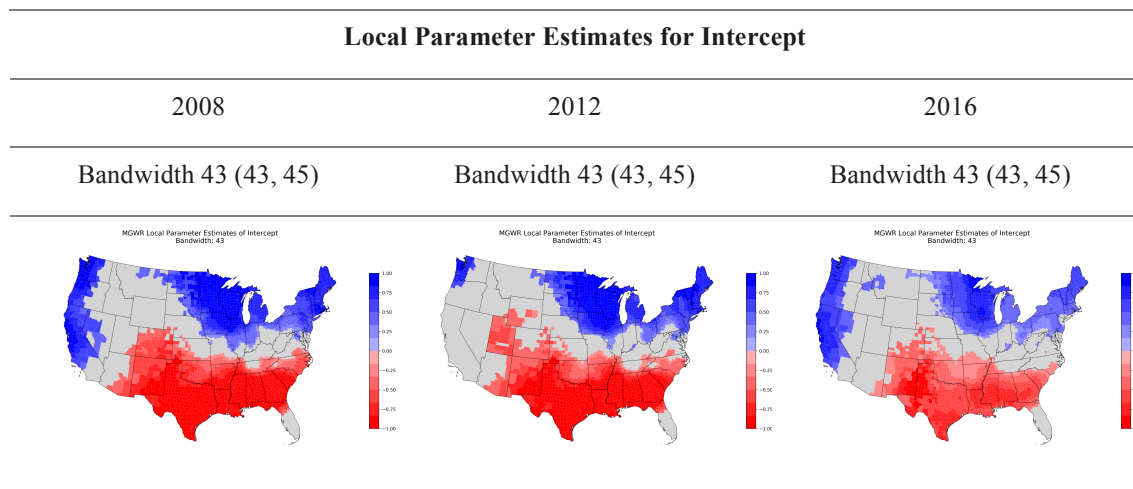


Figure 4.17. Local parameter estimates for the local intercept obtained in MGWR.

4.5 MGWR local results in Pennsylvania, Wisconsin, and Michigan

As a secondary interest, we can use the MGWR results to better inform us on what happened in certain critical counties which appeared to be very important in deciding the outcome of the 2016 election in favor of the Republican Party. These counties are in the important swing states of Pennsylvania, Wisconsin, and Michigan, and have previously been identified by Agnew and Shin (2020). We group counties in these three states based on their past three Presidential election results. There are 30 D-D-D counties that voted for the Democratic Party in 2008, 2012 and 2016. There are 38 D-D-R counties voted for the Democratic Party in 2008, 2012 and pivoted to the Republican Party in 2016. There is only one D-R-D county (Chester County, PA), 54 D-R-R counties, and 99 consistently Republican R-R-R counties. The locations of these five types of counties are shown in Figure 4.18 and Table 4.5 summarizes the voting results for the five groups. From Table 4.5, we can see that there has been a steady decrease in the percentage of Democratic votes from 2008, 2012 to 2016 for all groups of counties except for Chester County, PA where the Democrats gained 17,371 votes between 2012 and 2016. In total, from 2008 to 2012, the Democratic Party lost 640,485 votes across the three states, and from 2012 to 2016 they lost a further 598,725 votes. The decrease in votes for the Democratic Party became critical in the 2016 election when the Republican Party won each state by a tiny margin.

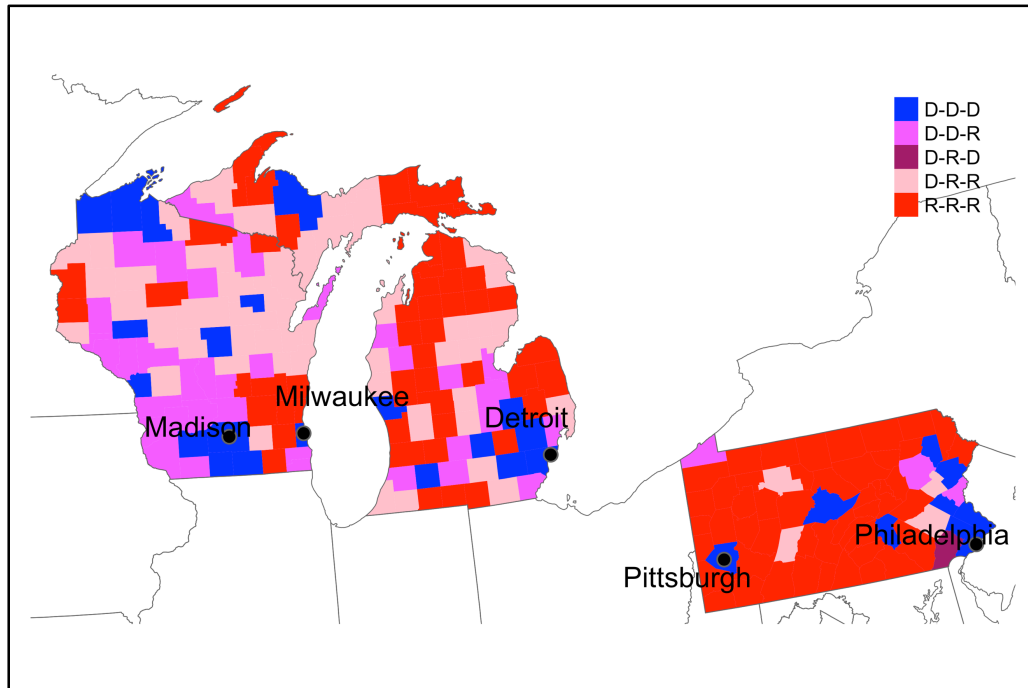


Figure 4.18. Locations of counties in Pennsylvania, Wisconsin and Michigan and their voted Party during 2008-2016 period. D-D-D denotes a county voted for Democratic Party in 2008, 2012 and 2016.

Table 4.5. Voting statistics of counties in Pennsylvania, Wisconsin and Michigan by group.

Group	No. of counties	% of Dem. votes in 2008	% of Dem. votes in 2012	Dem. votes +/- in 2012 from 2008	% of Dem. votes in 2016	Dem. votes +/- in 2016 from 2012
D-D-D	30	65.3%	63.8%	-226,647	60.6%	-147,004
D-D-R	38	55.3%	52.6%	-82,876	42.7%	-165,103
D-R-D	1	54.2%	49.3%	-13,522	52.7%	+17,371
D-R-R	54	51.8%	46.4%	-121,092	38.2%	-136,290
R-R-R	99	42.4%	38.1%	-196,348	31.6%	-167,699
Total	222	56.0%	53.0%	-640,485	47.3%	-598,725

To examine why this happened, we analyze the socioeconomic compositions of the five groups of counties over the three elections. These are shown in Figure 4.19. Because the socioeconomic variables are standardized for each year, they reflect the relative level of that variable compared with national averages. It is easy to characterize the D-D-D counties as having relatively a younger population, more people with a college degree, higher household income, a lower percentage of workers in the manufacturing industry, and higher population density. The D-D-R counties have relatively lower black and Hispanic populations, which indirectly shows that they are primarily white. They also have a more elderly population, a higher proportion of manufacturing employment, low population density, fewer foreign-born citizens, and a relatively low turnout rate compared with other counties within the three states. Chester County, PA, which is the only county that falls under the D-R-D category shows a quite distinct socio-economic profile compared with other counties. It is located west of Philadelphia, mostly urbanized, with high median household income, high education level, and more minority and foreign born population.

Looking across the three elections, some socioeconomic factors remain relatively stable over time. However, some interesting temporal trends can be identified:

- (i) The male to female ratios of R-R-R counties are higher in 2016 than in the previous two election years;
- (ii) D-D-R counties have an increasing trend of more elderly people;
- (iii) The percentage of manufacturing employment for all counties increased during the 2008-2016 period except for the D-R-D county and with the D-D-R counties having experienced the greatest increase;
- (iv) The third-party vote considerably decreased in 2012 but increased in 2016; and

- (v) The turnout rate of D-D-D and D-D-R counties dropped noticeably in 2016, while the turnout of R-R-R counties increased greatly.

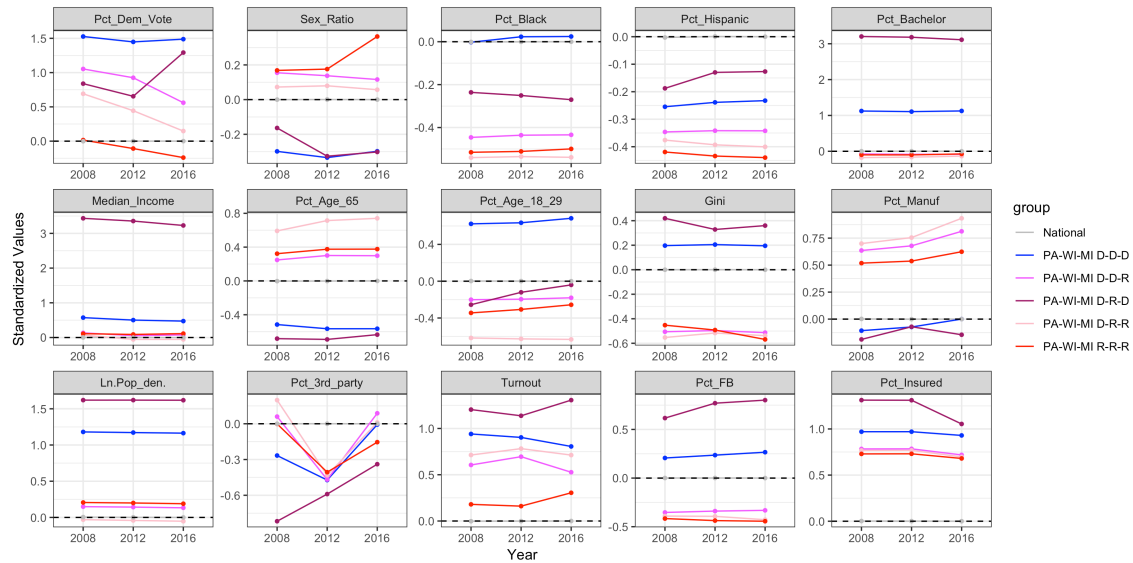


Figure 4.19. Socio-economic changes during 2008-2016 for each group of counties in PA-WI-MI.

Moreover, with MGWR, we are able to look at the processes associated with these socioeconomic factors at the local level and to examine whether these associations changed over the 2008-2016 period. The significant local parameter estimates obtained in the MGWR models are averaged for counties within each group. Figure 4.20 shows the counterpart of Figure 4.19 but for the local parameter estimates. The intercept measures the contextual effects of geography on voting preferences, which remains at a similar level in 2008 and 2012 but dropped in 2016. This indicates that in 2016, the intrinsic preference for voting for the Democratic Party dropped, and the socioeconomic influences were relatively more prominent than in the previous two elections. This is possibly because the Democratic candidate, Hilary Clinton, was not viewed with great enthusiasm by many traditional Democrat voters. This can also be seen by the reduction in voter turnout in the D-D-R counties. Clearly, Trump engaged his base support in these counties far

better than Clinton did and this resulted in him winning the overall election. The graphs of the changes in the local parameter estimates across the three elections also suggest that R-R-R and D-R-D counties had a lower intrinsic preference for voting for the Democratic Party compared with counties in the D-D-D, D-D-R and D-R-R groups. For race and ethnicity, in D-R-R and R-R-R counties, the tendency of Black voters to favor the Democratic Party decreased from 2012 to 2016, possibly due to both candidates being White. The impact of the Hispanic population on voting for the Democratic Party decreased noticeably from 2008 to 2012 and remained at a similar level between 2012 and 2016. The influence of education increased in 2016 and people with college degrees were more likely to vote for the Democratic Party in the 2016 election. Another notable change is that the influence of income on voting for the Republican Party (with a negative sign) decreased over time. Manufacturing employment shows a consistently small negative significant impact on voting for the Democratic Party. Despite there is a noticeable drop in 2012, but because the magnitude is quite small, the variation among the three elections is not noteworthy. For both population density and the third-party vote, there is a V-shape trend that the respective influence on voting preference dropped in 2012 but increased in the 2016 election. The rest of the variables do not show any consistent patterns in their associated local parameter estimates.

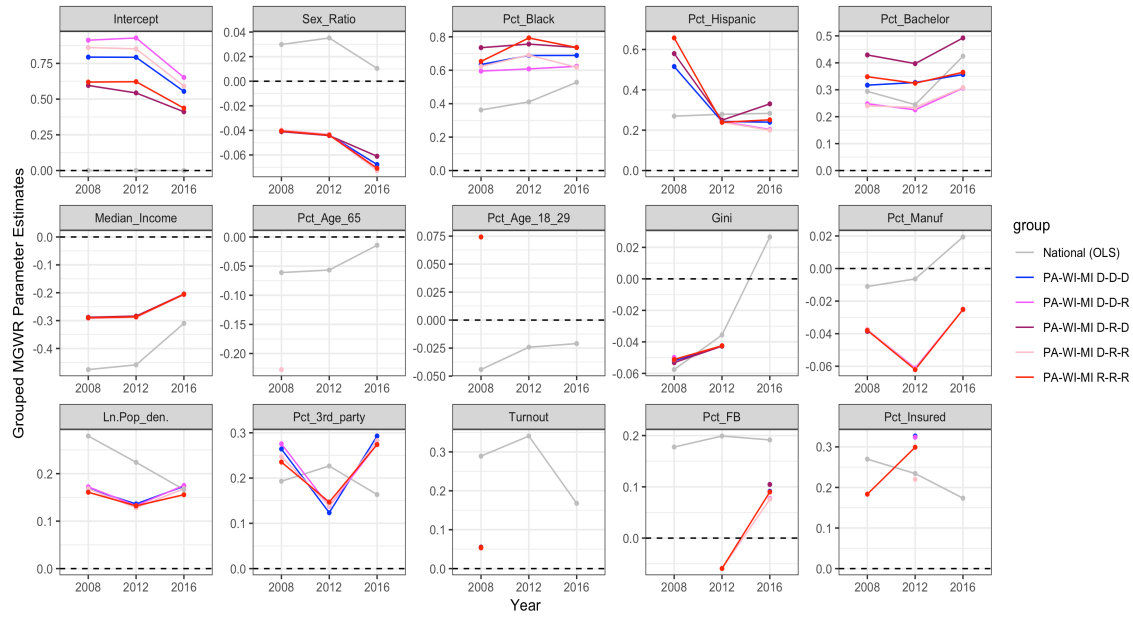


Figure 4.20. MGWR local parameter estimates changes during 2008-2016 for each group of counties in PA-WI-MI.

Overall, from the evidence in Figure 4.19 and Figure 4.20, we find that there were changes in some socio-economic factors but also in the degree to which these factors affected people’s voting preferences. Neither of these aspects should be omitted when predicting future elections. To reinforce this statement, we propose two hypothetical questions relating to the outcome of the 2016 Presidential election:

- (i) What would have happened in the 2016 presidential election if the socioeconomic make-up of each of the three key states, Pennsylvania, Wisconsin, and Michigan, had remained the same as in the 2012 election but voters’ behavior changed, as represented by the local parameters in the 2016 model?

- (ii) What would have happened if voter preferences had remained the same from 2012 to 2016, given the changing sociodemographic composition of each county?

The first question can be answered by plugging the 2012 socio-economic data into the 2016 MGWR model and calculating the predicted Democratic vote share for each of the three states. The second question can be answered by plugging the 2016 socio-economic data into the 2012 MGWR model. Table 4.6 summarizes the state voting results under these two scenarios. The actual Democratic vote share in the two-party race is very close to 50% for all three states. The 2016 MGWR model fits well and captures the tight margins in the three states, and only Michigan is misclassified as voting for the Democratic Party instead of the Republican Party. For the two hypothetical questions, when assuming socioeconomic factors stay the same between 2012 and 2016, the share of the Democratic vote in Pennsylvania, Wisconsin and Michigan would have been 46.1%, 47.3%, and 44.2%, respectively so Trump would have still won these three states and with an increased margin. Interestingly, if people's voting behavior had remained constant from 2012 to 2016 (that is, the effect of each socioeconomic factor on voting for the Democratic Party had remained constant) all three states would have been won by the Democrats resoundingly (53.4%, 55.2% and 60.4% respectively) and Hilary Clinton would now be President. Therefore, change in people's voting behavior was a more dominant factor in deciding the 2016 election outcome than the shifting in socio-economic. These results point to the importance of understanding and modeling voter preferences through local modeling.

Table 4.6 Democratic vote share in the two-party race in 2016 presidential election under two hypothetical scenarios.

Democratic vote share in the two-party race in 2016 presidential election				
State	Actual	2016 model with 2016 socio-economics	Scenario (i) 2016 model with 2012 socio-economics	Scenario (ii) 2012 model with 2016 socio-economics
Pennsylvania	49.6% (R)	48.6% (R)	46.1% (R)	53.4% (D)
Wisconsin	49.6% (R)	49.6% (R)	47.3% (R)	55.2% (D)
Michigan	49.9% (R)	50.4% (D)	44.2% (R)	60.4% (D)

5 Conclusions

In this study, we apply a local modeling approach, multiscale geographically weighted regression (MGWR), to examine the spatially and temporally varying nature of determinants of voter preferences in the 2008, 2012 and 2016 U.S. Presidential elections. The particular interest of this study is to investigate the spatiotemporal changes in the factors in determining voters' preferences. We extend the earlier work of Fotheringham et al. (2020) in which they applied the MGWR model to the 2016 Presidential election. They found that the MGWR model performed well in capturing the spatially varying relationships between socio-economic attributes and voter preference, also in demonstrating that these relationships varied at different spatial scales. Here we apply the same MGWR model to two additional sets of election results and the results indicate that the MGWR model produces consistent results across the three elections with very high degrees of fit.

To assess the spatiotemporal changes in the determinants of voting preferences, we compare the local parameter estimates produced from MGWR across the three elections. Results show that the spatial patterns of how socioeconomic factors affect voter preferences remain overall temporally stable despite some local and regional changes. The scales of spatial variations in the determinants can be measured by the optimal covariate-specific bandwidths. With a recent work by Li et al. (2020), the uncertainty of these bandwidths can also be quantified, which allows us to compare the bandwidths to infer if there is any significant difference in the spatial scales over which different processes operate. We also compare the bandwidth confidence intervals for a particular variable among the three elections to assess whether there are any significant changes of the spatial variations in the parameter estimates. The results show that the operational scales of the socio-economic processes remain stable over time. For example, the influence of median household income on voter preference in a county has been consistently global in the recent elections. Similarly, how the percentage of African American population affects the voter preference remains local over the elections. Furthermore, the spatial patterns of the local intercept, identified by Fotheringham et al. (2020) as the contextual effect to the election, show a consistently significant north-south divide in terms of the intrinsic partisan preferences in all three elections.

Additionally, recent work by Agnew and Shin (2020) claimed that a set of counties in Pennsylvania, Wisconsin, and Michigan (PA-WI-MI) played a critical role in the 2016 election. With the ability to look at local influences, we synthesize socioeconomic data and MGWR results for these counties based on their partisan preferences in the recent three elections. We found that the socioeconomic profile of different groups of counties are well stratified in PA-WI-MI. For example, consistent Democratic counties D-D-D, which voted the Democratic Party in all recent three elections, are well characterized with more minority population, more people with college

degrees, more young and fewer elderly population, and higher population density. The pivotal counties D-D-R, which voted Democratic in 2008 and 2012 but voted Republican in 2016, have lower Black and Hispanic populations, a more elderly population, a higher proportion of manufacturing employment, low population density, fewer foreign-born citizens, and a relatively low turnout rate compared with other counties in the three states. During the three elections, some socioeconomics have changed, but most of them remain relatively stable in PA-WI-MI. However, the impact of socioeconomic composition on voting preferences, which is measured by the local parameter estimates of MGWR, is showing more noticeable temporal trends. Specifically, the influence of Hispanic population on voting for the Democratic Party decreases significantly from 2008 to 2012, but stays relatively the same from 2012 to 2016. The tendency of Black voters to favor the Democratic Party decreased from 2012 to 2016 only in the Republican counties. Moreover, education exhibits a stronger positive influence, and income exhibits a weaker negative impact on the Democratic vote share from 2012 to 2016. The influence of the third-party vote declines in 2012 but increases in 2016. Manufacturing employment remains a consistently negative impact on voting preference during all three elections.

To support the view that the 2016 election is affected by both socioeconomic changes and voter preference changes, and to examine if either of these effects is dominant, we propose two hypothetical scenarios to the counties in PA-WI-MI which are vital in the election. The first scenario is what would happen if the socioeconomic composition of each county in PA-WI-MI remained the same from 2012 to 2016, but the determinants of voter preferences changed. By using the 2012 data with the local parameter estimates from the 2016 MGWR model, we can predict what would have happened under scenario 1. That is, Trump would have won each of the three states with even greater margins. The second scenario is what would happen if the determinants of people's voting preferences had remained the same between 2012 and 2016 but

the socioeconomic composition of each county changed as it did in reality. We can answer the question by using the 2016 sociodemographic data for each county with the 2012 MGWR parameter estimates to forecast the Democratic vote share in each county. In this case, we find that all three states would have been won by the Democratic Party, and Clinton would have won the election. Therefore, the shifting in the voters' preferences is a more dominant factor in deciding the outcome of the 2016 Presidential election.

In this study, because of the temporal availability of the county-level American Community Survey datasets, we apply the MGWR model only to the recent three elections. It would be fascinating to look at the stability or variability of how socioeconomic processes and spatial context affected voting preference over a much more extended period by utilizing the decennial census data. An even longer analysis of the dynamics of the determinants of voter preferences would identify which of the socio-economics have relatively consistent influences on voter preference, and which change significantly over time. However, how to match each election, which happened every four years, to the decennial census may need further experimentation. Moreover, to better assess the significance the temporal variability over a longer period, a panel extension of the MGWR model could be developed. With extended data and model formulation, we could assess the spatiotemporal trends of contextual and socioeconomic impacts, which will be very useful for understanding the historical electoral geography of the US and for voter targeting at local and national scales in the future election campaigns.

Data and codes availability statement

The data and codes that support the findings of this study are available in figshare.com with the identifier 10.6084/m9.figshare.12209660.

References

- Agnew, J. (1996). Mapping politics: how context counts in electoral geography. *Political geography*, 15(2), 129-146.
- Agnew, J. (2014). *Place and politics: The geographical mediation of state and society*. Routledge.
- Agnew, J., & Shin, M. (2020). The Counties that Counted: Could 2020 Repeat 2016 in the US Electoral College?. *The Forum* (Vol. 17, No. 4, pp. 675-692).
- Anselin, L., & Bera, A. K. (1998). Introduction to spatial econometrics. *Handbook of applied economic statistics*, 237.
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4), 281-298.
- Buja, A., Hastie, T., & Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, 17(2), 453-510.
- Cho, W. K. T., & Gimpel, J. G. (2009). Presidential voting and the local variability of economic hardship. In *The Forum* (Vol. 7, No. 1). De Gruyter.
- Cox, K. R. (1969). The voting decision in a spatial context. *Progress in geography*, 1, 81-117.
- Cupido, K., Fotheringham, A. S., & Jevtic, P. (2019). Local Modeling of US Mortality Rates: A Multiscale Geographically Weighted Regression Approach. *Available at SSRN 3472830*.
- Darmofal, D. (2008). The political geography of the new deal realignment. *American Politics Research*, 36(6), 934-961.
- Degan, A., & Merlo, A. (2011). A structural model of turnout and voting in multiple elections. *Journal of the European Economic Association*, 9(2), 209-245.
- Delaney, D., & Leitner, H. (1997). The political construction of scale. *Political geography*, 16(2), 93-97.
- Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., ... & Kühn, I. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5), 609-628.
- Duncan, J. S., & Ley, D. (2013). *Place/culture/representation*. Routledge.
- Flint, C. (1996). Whither the individual, whither the context?. *Political Geography*, 15(2), 147-151.
- Forest, B. (2018). Electoral geography: From mapping votes to representing power. *Geography*

- Compass*, 12(1), e12352.
- 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.
- Fotheringham, A. S., Yue, H., & Li, Z. (2019). Examining the influences of air quality in China's cities using multi-scale geographically weighted regression. *Transactions in GIS*.
- Fotheringham, A. S., Li, Z. & Wolf, L. J. (2020). Scale, Context and Heterogeneity: A Spatial Analytical Perspective on the 2016 US Presidential Election.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models* (Vol. 43). CRC Press.
- Johnston, R., Shelley, F. M., & Taylor, P. J. (1990). *Developments in electoral geography*. Routledge.
- Johnston, R., & Pattie, C. (2006). *Putting voters in their place: Geography and elections in Great Britain*. Oxford University Press.
- King, G. (1996). Why context should not count. *Political geography*, 15, 159-164.
- Kühn, I. (2007). Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, 13(1), 66-69.
- Leighley, J. E., & Nagler, J. (2013). *Who votes now?: Demographics, issues, inequality, and turnout in the United States*. Princeton University Press.
- 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
- Li, Z., & Fotheringham, A. S. (2020). Computational improvements to multi-scale geographically weighted regression. *International Journal of Geographical Information Science*, doi:10.1080/13658816.2020.1720692.
- Li, Z., Fotheringham, A. S., Oshan, T. M., & Wolf, L. J. (2020). Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights. *Annals of the American Association of Geographers*, doi:10.1080/24694452.2019.1704680.
- Miller, B. (1994). Political empowerment, local—central state relations, and geographically shifting political opportunity structures: Strategies of the Cambridge, Massachusetts, Peace Movement. *Political Geography*, 13(5), 393-406.

- O'Loughlin, J., Flint, C., & Anselin, L. (1994). The geography of the Nazi vote: Context, confession, and class in the Reichstag election of 1930. *Annals of the Association of American Geographers*, 84(3), 351-380.
- Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (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.
- Oshan, T. M., Smith, J. P., & Fotheringham, A. S. (2020). Targeting the spatial context of obesity determinants via multiscale geographically weighted regression. *International Journal of Health Geographics*, 19(1), 1-17.
- Scala, D. J., Johnson, K. M., & Rogers, L. T. (2015). Red rural, blue rural? Presidential voting patterns in a changing rural America. *Political Geography*, 48, 108-118.
- Shin, M. E., & Agnew, J. A. (2008). *Berlusconi's Italy: mapping contemporary Italian politics*. Temple University Press.
- Sui, D. Z., & Hugill, P. J. (2002). A GIS-based spatial analysis on neighborhood effects and voter turn-out: a case study in College Station, Texas. *Political Geography*, 21(2), 159-173.
- Taylor, P. J. (1973). Some implications of the spatial organization of elections. *Transactions of the Institute of British Geographers*, 121-136.
- Tyson, A. & Maniam, S (2016). Behind Trump's victory: Divisions by race, gender, education. Retrieved from <https://www.pewresearch.org/fact-tank/2016/11/09/behind-trumps-victory-divisions-by-race-gender-education>.
- Warf, B., & Leib, J. (2011). *Revitalizing electoral geography*. Burlington, VT: Ashgate.
- Yang, C., Zhan, Q., Lv, Y., & Liu, H. (2019). Downscaling Land Surface Temperature Using Multiscale Geographically Weighted Regression Over Heterogeneous Landscapes in Wuhan, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., Kang, W., & Wolf, L. J. (2019). Inference in multiscale geographically weighted regression. *Geographical Analysis*.

CHAPTER 5

CONCLUSIONS

5.1 Summary of the dissertation

Since its introduction by Fotheringham et al. (2017), Multiscale Geographically Weighted Regression (MGWR) has received considerable attention in the spatial analysis community (Shabrina et al., 2019; Cupido et al., 2019; Fotheringham et al., 2019; Murakami et al., 2019; Yang et al., 2019; Comber et al., 2020; Fotheringham et al., 2020; Liu et al., 2020; Oshan et al., 2020). Undoubtedly, MGWR is a more accurate model than classic GWR, and covariate-specific bandwidths offer great insight into the spatial scale over which different processes occur. Centering on MGWR, this dissertation presents three separate novel studies that improve the model from aspects covering computation, inference, and application.

The first study, which constitutes Chapter 2 of the dissertation, improves the computation efficiency of MGWR in terms of both time and memory. We extend the FastGWR parallel algorithm, which is detailed in Li et al. (2019), to estimate MGWR parameters efficiently. Additionally, a novel parallel chunk-wise partitioning method was introduced to save the memory footprint in MGWR inference computation. With these two strategies, the improved MGWR algorithm is hundreds of times faster than implementations in both Fotheringham et al. (2017) and the GWmodel R library (Gollini, 2015). Additionally, the new algorithm described here can automatically partition large matrices to fit in the available memory of the user's computing system. For accessibility, the algorithm is integrated into the mgwr python package and the MGWR 2.0 software, and it scales well on both personal computers and High-Performance Computing (HPC) infrastructures. These developments not only save users time during the model construction procedure but, more importantly, allow the application of MGWR to big spatial data, such as those collected from social media, remote sensing, and other sensors. Consequently, this

study opens up a new area of research to examine spatial process heterogeneity and scale for a large study area with finer geographic units.

The second study, which constitutes Chapter 3 of the dissertation, provides evidence and measures of bandwidth uncertainties in MGWR. The optimal covariate-specific bandwidths are closely related to how each underlying spatial process varies over space, which can be used to potentially differentiate local, regional, and global processes. However, no measures of the bandwidth uncertainty exist, so conclusion drawn based purely on comparing the optimal bandwidths may be flawed. Understanding the uncertainty attached to the optimal bandwidth is vitally important to make inferences about the different spatial scales over which processes operate. Based on simulated and empirical examples, results from both bootstrapping and a novel use of Akaike weights show that the optimal bandwidths selected in MGWR are subject to the sampling variation of the data. Akaike weights are shown to be the preferred statistic to quantify the uncertainty of optimal bandwidths because of their computational efficiency. Also, the weights can be interpreted as the prior probability of a model being selected as the true model. Bandwidth confidence intervals can be computed based on cumulative Akaike weights. Additionally, following a model averaging approach, MGWR parameter estimates can be weighted according to Akaike weights obtained from a set of candidate bandwidths. Accordingly, the bandwidth uncertainty can be accounted for, and the averaged parameter estimates are more robust than the ones based on a pre-set optimal bandwidth. From examples presented in the chapter, we find that the unconditional estimates and inference are marginally better than the conditional ones. An important next step is to investigate the coverage probability of the Akaike weights-derived bandwidth confidence intervals and to see in what situations model averaging is strongly preferred. Furthermore, the robustness of optimal bandwidths and uncertainties needs to be systematically examined in both controlled experiments and empirical studies to understand

the behavior of the optimal bandwidths, for example, in situations when covariates are correlated and when omitted variable bias presents. This chapter also shows implications when spatial weighting matrices are employed in other spatial models. Usually, a pre-defined spatial weight matrix is incorporated so model results are conditional on how the weight matrix is specified. The sensitivity or uncertainty of the specification to the results is often omitted. Here, we present a study to look at the uncertainty of bandwidth, which controls spatial weight matrix, in the GWR context. We hope this study can encourage more similar work in this direction.

The third study, which constitutes Chapter 4 of the dissertation, applies MGWR and the developments in Chapter 2 and 3, to examine the spatiotemporal changes of voter preference determinants across the recent three US Presidential elections, 2008, 2012, and 2016. The MGWR model achieves great fit with R^2 values greater than 0.9 for all three years and with a much smaller AICc compared to OLS. Covariate-specific bandwidths and local parameter estimates obtained from MGWR reveal that there is a great degree of consistency in spatial variations of how socioeconomic factors influence voter preferences across the three elections, which happened at different scales. For example, the impact of median household income has a consistently similar impact on voting preferences for all the counties in the US. We put special interest in the local intercepts, which are interpreted by Fotheringham et al. (2020) as measurements of the contextual effects on voting preferences, and these are shown to exhibit a consistent spatial pattern which clearly divides the US into three: counties that are intrinsically Democrat; counties that are intrinsically Republican; and counties that are intrinsically neutral. Additionally, with the capability to understand local relationships, we synthesize the MGWR results for counties in three critical swing states: Pennsylvania, Wisconsin, and Michigan. Counties in these states are grouped based on the voting results in the recent three elections. We found that the socioeconomics of different groups of counties are well stratified in the three

states. Further, results show that the outcome of the 2016 election in Pennsylvania, Wisconsin, and Michigan can be explained by an entwined effect of changes in socioeconomics, context, and their associated impacts on the voter preference. If socioeconomics stay the same from 2012 to 2016, we found that Pennsylvania, Wisconsin, and Michigan would have still voted for the Republican Party but Trump would have won with an increased margin. More interestingly, if the influences on the voter preference from socioeconomics and context remain the same, all three states would have voted for the Democratic Party, and Clinton would have won the election by winning the electoral seats from these three states. Therefore, both socioeconomics changes and shifts in the voting preferences need to be considered when look at future elections.

5.2 Future research directions

This section focuses on the future model developments of MGWR. There is a rich body of literature on developing extensions of the classic GWR model, which are designed to make GWR more flexible. A selected list of studies includes GWR with non-Gaussian errors (Nakaya, et al., 2005), with anisotropic kernels (Páez, 2004), with locally optimized bandwidths (Páez et al., 2002; Comber et al., 2018), using non-Euclidean distance metrics (Lu et al., 2015), incorporating temporal effects (Huang et al., 2010; Fotheringham et al., 2015), and with variable selection (Wheeler et al., 2009; Li et al., 2018). Further, MGWR is introduced, and the essence of MGWR to allow bandwidth to differ among covariates can be combined with the above-mentioned developments. There is a couple of studies already on this path, including the Parameter-specific Distance Metrics (PSDM) model of Lu et al. (2017), which is based on Lu et al. (2015), and Multiscale geographically and temporally weighted regression (MGTWR) of Wu et al. (2019), which is based prior work of Huang et al. (2010) and Fotheringham et al. (2015).

Along this line, it is feasible to extend MGWR with localized bandwidths and/or with anisotropic kernels based on the work of Páez et al. (2002; 2004) for situations when the degree of spatial variation in the processes is not uniform over space. Moreover, Generalized MGWR is needed to model non-Gaussian outcomes such as count (Poisson) and binary (Logistic) data. These developments are natural combinations of MGWR with previous GWR extensions, and they would be useful additions to the existing spatial modeling toolset. Further, some pioneering work has started to integrate Machine Learning (ML) techniques, such as random forest (Georganos et al., 2019) and neural networks (Wu et al., 2020; Du et al., 2020), with GWR to model locally non-linear relationships. Though I do see the potential of combining local models with Machine Learning for prediction purposes, how to interpret the model parameters and make an inference of the underlying spatial processes needs further exploration, which would be a challenging but exciting avenue. Additionally, it worth noting that as models becoming more flexible, the computational cost is inevitably increasing. Nearly all of the research, as mentioned above, have found computation burdensome. The improvements introduced in Chapter 2 and FastGWR (Li et al., 2019) would be very beneficial to these new models. Moreover, as bandwidth is an inevitable key parameter in local models, understanding the uncertainty of the bandwidth and how sensitive the results are regarding the use of different bandwidths are crucial. The measures of bandwidth uncertainty developed in Chapter 3 will be very useful here.

The (M)GWR models are not only inferential models but also predictive models. They are useful in making spatial prediction after accounting for the heterogeneity in the spatial processes generating the data that are being predicted. Some comparison studies showed that the classic GWR outperforms OLS and achieves comparable performance to kriging-based and machine learning methods (Propastin, 2009; Harris et al., 2010; Selby and Kockelman

, 2013; Zou et al., 2016; Hong et al., 2017). It is expected that MGWR achieves better prediction accuracy than GWR because it allows the bandwidths to vary so the estimation of the spatial processes is more accurate. However, the prediction function of MGWR hasn't been operationalized yet because of the complexity in the estimation procedure. Nevertheless, this is a feasible short term goal by following GWR and GAM frameworks and related literature. Simulated and empirical experiments and comparisons to GWR and other popular methods are needed to demonstrate the predictability of MGWR.

Lastly, I would like to comment on the evolution of the regression models in general and shed light on the future possible direction for local models. The development of regression models is a process of successively removing assumptions so that models are more flexible and accurate. It starts with OLS, to Generalized Linear Model, to Generalized Additive Model, then to complex ML algorithms such as Neural Networks, moving from linear to non-linear, from Gaussian to not only Gaussian, from parametric to non-parametric. Similarly, this is what I would see where local models are going. It starts from GWR with various assumptions, to the further extensions of GWR in the literature, to MGWR with multiple bandwidths, to above-mentioned models that combine MGWR with further developments and extensions, to a future local model that can capture the underlying spatial processes without strong assumptions of data distributions, spatial structures, and model forms. To Integrate ML algorithms into local modeling framework or to embody explicitly spatial structures into ML would be possible routes. However, the interpretability and inference of these models should not be comprised when moving forward towards these directions.

REFERENCES

- Agnew, J. (1996). Mapping politics: how context counts in electoral geography. *Political geography*, 15(2), 129-146.
- Agnew, J. (2014). *Place and politics: The geographical mediation of state and society*. Routledge.
- Agnew, J., & Shin, M. (2020). The Counties that Counted: Could 2020 Repeat 2016 in the US Electoral College?. *The Forum* (Vol. 17, No. 4, pp. 675-692).
- Anselin, L., & Bera, A. K. (1998). Introduction to spatial econometrics. *Handbook of applied economic statistics*, 237.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370.
- Bitter, C., Mulligan, G. F., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*, 9(1), 7-27.
- 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.
- Cardozo, O. D., García-Palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*, 34, 548-558.
- Casetti, E. (1972). Generating models by the expansion method: applications to geographical research. *Geographical analysis*, 4(1), 81-91.
- Comber, A. J., Brunsdon, C., & Radburn, R. (2011). A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. *International journal of health geographics*, 10(1), 44.
- Cahill, M., & Mulligan, G. (2007). Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review*, 25(2), 174-193.
- Cho, W. K. T., & Gimpel, J. G. (2009). Presidential voting and the local variability of economic hardship. In *The Forum* (Vol. 7, No. 1). De Gruyter.

- Comber, Alexis, Chris Brunsdon, Martin Charlton, Guanpeng Dong, Rich Harris, Binbin Lu, Yihe Lü et al. "The GWR route map: a guide to the informed application of Geographically Weighted Regression." *arXiv preprint arXiv:2004.06070* (2020).
- Cox, K. R. (1969). The voting decision in a spatial context. *Progress in geography*, 1, 81-117.
- Cupido, K., Fotheringham, A. S., & Jevtic, P. (2019). Local Modeling of US Mortality Rates: A Multiscale Geographically Weighted Regression Approach. *Available at SSRN 3472830*.
- da Silva, A. R., & Fotheringham, A. S. (2016). The multiple testing issue in geographically weighted regression. *Geographical Analysis*, 48(3), 233-247.
- Darmofal, D. (2008). The political geography of the new deal realignment. *American Politics Research*, 36(6), 934-961.
- Degan, A., & Merlo, A. (2011). A structural model of turnout and voting in multiple elections. *Journal of the European Economic Association*, 9(2), 209-245.
- Delaney, D., & Leitner, H. (1997). The political construction of scale. *Political geography*, 16(2), 93-97.
- Darmofal, D. (2008). The political geography of the new deal realignment. *American Politics Research*, 36(6), 934-961.
- Duncan, J. S., & Ley, D. (2013). *Place/culture/representation*. Routledge.
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- Flint, C. (1996). Whither the individual, whither the context?. *Political Geography*, 15(2), 147-151.
- Forest, B. (2018). Electoral geography: From mapping votes to representing power. *Geography Compass*, 12(1), e12352.
- 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.
- Fotheringham, A. S., Yue, H., & Li, Z. (2019). Examining the influences of air quality in China's cities using multi-scale geographically weighted regression. *Transactions in GIS*.
- Fotheringham, A. S., Li, Z. & Wolf, L. J. (2020). Scale, Context and Heterogeneity: A Spatial Analytical Perspective on the 2016 US Presidential Election.

- Finley, A. O. (2011). Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, 2(2), 143-154.
- 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.
- Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, 63(17).
- Griffith, D. A., & Chun, Y. (2014). An eigenvector spatial filtering contribution to short range regional population forecasting. *Economics and Business Letters*, 3(4), 208-217.
- 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., Fotheringham, A. S., Crespo, R., & Charlton, M. (2010). The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Mathematical Geosciences*, 42(6), 657-680.
- 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.
- Harris, R., Singleton, A., Grose, D., Brunsdon, C., & Longley, P. (2010). Grid-enabling Geographically Weighted Regression: A Case Study of Participation in Higher Education in England. *Transactions in GIS*, 14(1), 43-61.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models* (Vol. 43). CRC Press.
- Hong, H., Pradhan, B., Sameen, M. I., Chen, W., & Xu, C. (2017). Spatial prediction of rotational landslide using geographically weighted regression, logistic regression, and support vector machine models in Xing Guo area (China). *Geomatics, Natural Hazards and Risk*, 8(2), 1997-2022.

- Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes Jr, M. G., Estes, S. M., ... & Liu, Y. (2013). Estimating ground-level PM_{2.5} concentrations in the southeastern US using geographically weighted regression. *Environmental Research*, 121, 1-10.
- 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.
- Johnston, R., Shelley, F. M., & Taylor, P. J. (1990). *Developments in electoral geography*. Routledge.
- Johnston, R., & Pattie, C. (2006). *Putting voters in their place: Geography and elections in Great Britain*. Oxford University Press.
- King, G. (1996). Why context should not count. *Political geography*, 15, 159-164.
- 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.
- Kühn, I. (2007). Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, 13(1), 66-69.
- Leighley, J. E., & Nagler, J. (2013). *Who votes now?: Demographics, issues, inequality, and turnout in the United States*. Princeton University Press.
- 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.
- Li, Z., & Fotheringham, A. S. (2020). Computational improvements to multi-scale geographically weighted regression. *International Journal of Geographical Information Science*, doi:10.1080/13658816.2020.1720692.
- Li, Z., Fotheringham, A. S., Oshan, T. M., & Wolf, L. J. (2020). Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights. *Annals of the American Association of Geographers*, doi:10.1080/24694452.2019.1704680.
- Liu, P., Wu, C., Chen, M., Ye, X., Peng, Y., & Li, S. (2020). A Spatiotemporal Analysis of the Effects of Urbanization's Socio-Economic Factors on Landscape Patterns Considering Operational Scales. *Sustainability*, 12(6), 2543.

- Lu, B., Yang, W., Ge, Y., & Harris, P. (2018). Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths. *Computers, Environment and Urban Systems*, 71, 41-57.
- 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.
- Miller, B. (1994). Political empowerment, local—central state relations, and geographically shifting political opportunity structures: Strategies of the Cambridge, Massachusetts, Peace Movement. *Political Geography*, 13(5), 393-406.
- Murakami, D., Lu, B., Harris, P., Brunsdon, C., Charlton, M., Nakaya, T., & Griffith, D. A. (2019). The importance of scale in spatially varying coefficient modeling. *Annals of the American Association of Geographers*, 109(1), 50-70.
- O'Loughlin, J., Flint, C., & Anselin, L. (1994). The geography of the Nazi vote: Context, confession, and class in the Reichstag election of 1930. *Annals of the Association of American Geographers*, 84(3), 351-380.
- 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. P., & Fotheringham, A. S. (2020). Targeting the spatial context of obesity determinants via multiscale geographically weighted regression. *International Journal of Health Geographics*, 19(1), 1-17.
- Sui, D. Z., & Hugill, P. J. (2002). A GIS-based spatial analysis on neighborhood effects and voter turn-out: a case study in College Station, Texas. *Political Geography*, 21(2), 159-173.
- Páez, A., Farber, S., & Wheeler, D. (2011). A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A*, 43(12), 2992-3010.
- 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.

- 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.
- Propastin, P. A. (2009). Spatial non-stationarity and scale-dependency of prediction accuracy in the remote estimation of LAI over a tropical rainforest in Sulawesi, Indonesia. *Remote Sensing of Environment*, 113(10), 2234-2242.
- Scala, D. J., Johnson, K. M., & Rogers, L. T. (2015). Red rural, blue rural? Presidential voting patterns in a changing rural America. *Political Geography*, 48, 108-118.
- Selby, B., & Kockelman, K. M. (2013). Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression. *Journal of Transport Geography*, 29, 24-32.
- Shabrina, Z., Buyuklieva, B., & Ming, M. N. K. (2019). Airbnb, hotels, and saturation of the food industry: A multi-scale GWR approach. arXiv preprint arXiv:1905.12543.
- Shi, Haijin, Edward J. Laurent, Joseph LeBouton, Laila Racevskis, Kimberly R. Hall, Michael Donovan, Robert V. Doepker, Michael B. Walters, Frank Lupi, and Jianguo Liu. "Local spatial modeling of white-tailed deer distribution." *Ecological Modelling* 190, no. 1-2 (2006): 171-189.
- Shin, M. E., & Agnew, J. A. (2008). *Berlusconi's Italy: mapping contemporary Italian politics*. Temple University Press.
- Taylor, P. J. (1973). Some implications of the spatial organization of elections. *Transactions of the Institute of British Geographers*, 121-136.
- Tran, H. T., Nguyen, H. T., & Tran, V. T. (2016, October). Large-scale geographically weighted regression on Spark. In *Knowledge and Systems Engineering (KSE), 2016 Eighth International Conference on* (pp. 127-132). IEEE.
- Troy, A., Grove, J. M., & O'Neil-Dunne, J. (2012). The relationship between tree canopy and crime rates across an urban–rural gradient in the greater Baltimore region. *Landscape and Urban Planning*, 106(3), 262-270.
- Tyson, A. & Maniam, S (2016). Behind Trump's victory: Divisions by race, gender, education. Retrieved from <https://www.pewresearch.org/fact-tank/2016/11/09/behind-trumps-victory-divisions-by-race-gender-education>.
- Warf, B., & Leib, J. (2011). *Revitalizing electoral geography*. Burlington, VT: Ashgate.
- Wolf, L. J., Oshan, T. M., & Fotheringham, A. S. (2018). Single and multiscale models of process spatial heterogeneity. *Geographical Analysis*, 50(3), 223-246.

- Wu, C., Ren, F., Hu, W., & Du, Q. (2019). Multiscale geographically and temporally weighted regression: exploring the spatiotemporal determinants of housing prices. *International Journal of Geographical Information Science*, 33(3), 489-511.
- Xu, P., Huang, H., Dong, N., & Wong, S. C. (2017). Revisiting crash spatial heterogeneity: a Bayesian spatially varying coefficients approach. *Accident Analysis & Prevention*, 98, 330-337.
- Yang, C., Zhan, Q., Lv, Y., & Liu, H. (2019). Downscaling Land Surface Temperature Using Multiscale Geographically Weighted Regression Over Heterogeneous Landscapes in Wuhan, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Yu, H., Fotheringham, S., Li, Z., Oshan, T., Kang, W., & Wolf, L. (2019). Inference in multiscale geographically weighted regression. *Geographical Analysis*. doi: 10.1111/gean.12189
- Zhang, J. (2010). Towards personal high-performance geospatial computing (HPC-G): perspectives and a case study. In *Proceedings of the ACM SIGSPATIAL international workshop on high performance and distributed geographic information systems* (pp. 3-10). ACM.
- Zhang, Y., Brooks, D. H., Franceschini, M. A., & Boas, D. A. (2005). Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *Journal of biomedical optics*, 10(1), 011014.
- Zou, B., Pu, Q., Bilal, M., Weng, Q., Zhai, L., & Nichol, J. E. (2016). High-resolution satellite mapping of fine particulates based on geographically weighted regression. *IEEE Geoscience and Remote Sensing Letters*, 13(4), 495-499.