

Exploring equity using geographically weighted regression

**Final Version of** Kim, J. and Nicholls, S. (2016). Using Geographically Weighted Regression to Explore the Equity of Public Open Space Distributions. *Journal of Leisure Research*, 48(2), 105-133.

## **Using Geographically Weighted Regression to Explore the Equity of Public Open Space Distributions**

### **Abstract**

Assessing levels of equity inherent in the distributions of the public open spaces that they manage is an important responsibility of park and recreation agencies. Multivariate regression offers one way of conducting such assessments. However, traditional ordinary least squares (OLS) techniques fail to explore important local variations in relationships among variables. This study explored the utility of geographically weighted regression (GWR) in an equity analysis of public beaches in the Detroit Metropolitan Area. The GWR models exhibited substantial improvements in model performance over the OLS models. GWR offers public leisure agencies a powerful technique via which to better understand local patterns of access and equity, ultimately leading to the formulation of more effective and efficient recreation planning and management policies.

**KEYWORDS:** Access, equity, geographically weighted regression, public open spaces

## Introduction

Green and blue spaces such as parks, playgrounds, trails, golf courses and lakes are public open spaces (POSs) that can provide local communities with recreation settings in addition to various other environmental, social, health, and economic benefits (Porter, 2001; Taylor, Floyd, Whitt-Glover, & Brooks, 2007). Concerns regarding inequities in the distribution of POSs have risen over the last few decades (Byrne, Wolch, & Zhang, 2009; Deng, Walker, & Strager, 2008; Tarrant & Cordell, 1999; Taylor et al., 2007). As a result, multiple studies have attempted to determine levels of equity across various demographic and socioeconomic groups for parks (Byrne et al., 2009; Maroko, Maantay, Sohler, Grady, & Arno, 2009; Moore, Diez Roux, Evenson, McGinn, & Brines, 2008; Nicholls, 2001; Nicholls & Shafer, 2001; Omer, 2006; Talen, 1997; 1998), trails (Estabrooks, Lee, & Gyurcsik, 2003), playgrounds (Smoyer-Tomic, Hewko, & Hodgson, 2004), golf courses (Deng et al., 2008), recreational forests (Tarrant & Cordell, 1999), and campsites (Porter & Tarrant, 2001).

To measure the degree of equity inherent in the distribution of POSs, multivariate linear regression using the ordinary least squares (OLS) method has recently been employed. OLS regression uses a global predictive model to capture the strength and significance of the statistical relationship between dependent and independent variables over an entire study area (Gilbert & Chakarabarty, 2011). However, spatial data such as the geographic locations of POSs, measures of access to POSs (e.g., distance or travel time between origin and destination), and spatially referenced census data, may exhibit spatial effects such as spatial dependence and spatial heterogeneity that can lead to biased estimation results using traditional multivariate techniques (Fotheringham, Brunson, & Charlton, 2002). The equity of POSs, as represented by the relationship between level of access and spatially referenced census data, should ideally

## Exploring equity using geographically weighted regression

be examined using specialized research methods that explicitly account for spatial location and therefore differ from those used to analyze non-spatial data. To date, however, this has not typically been the case.

The purpose of this study is to demonstrate the value of geographically weighted regression (GWR) as an equity analysis tool. Specifically, the relative benefits of GWR techniques relative to traditional OLS methods are demonstrated via a case study of public beaches in the Detroit Metropolitan Area.

### **Literature review**

#### *Previous Approaches to the Measurement of POS Equity*

To measure the degree of equity inherent in the distribution of parks and recreation facilities, previous studies have investigated the existence and extent of relationships between levels of access to these facilities and residents' demographic and socioeconomic status. A variety of different methods such as non-parametric difference of means tests (Nicholls, 2001; Nicholls & Shafer, 2001), linear correlation (Omer, 2006; Smoyer-Tomic et al., 2004), equity mapping (Talen, 1997; 1998), and multivariate linear regression (Deng et al., 2008; Porter & Tarrant, 2001; Tarrant & Cordell, 1999) have been utilized. Among these methods, multivariate linear regression using the OLS method has been recognized as the most powerful. Those studies that have employed multivariate techniques have tended to utilize a logistic approach, which categorizes level of access to POSs as a dichotomous outcome (e.g., 1: has access; 0: does not have access). Deng et al. (2008), for example, used logistic regression to examine the distributional equity of golf courses relative to Chinese residents in Calgary, Canada, over a 10-year time span. Results indicated that Chinese residents were concentrated in several parts of Calgary during this time, and that they were more likely than Anglo-Canadians to reside in

## Exploring equity using geographically weighted regression

census tracts that did not contain, or were not near to, golf courses. However, distributional inequity decreased during the study period, primarily due to the construction of new golf courses in or near Chinese communities. Tarrant and Cordell (1999) determined the relationships between the distribution of outdoor recreation sites and census variables in northern Georgia, finding inequity with regard to household income, but no evidence of any inequity with respect to race, occupation or ethnic heritage. Porter and Tarrant (2001) investigated socioeconomic and racial inequities with respect to the distribution of federal tourism sites and campsites in southern Appalachia; findings showed that the distribution of these sites was advantageous to White populations and disadvantageous to minority populations. This study uses a two-pronged, finer approach – based on the number of public beaches within a specified distance of each census unit and the distance between each census unit and the closest public beach – thereby providing a far more comprehensive portrayal of the extent to which access varies across the study area than a dichotomous (access versus no access) analysis can provide.

### *Ordinary Least Squares (OLS) Regression and Spatial Effects*

OLS is the most widely known and used regression method to model a dependent variable's association with a set of independent variables. OLS is based on two critical assumptions: (1) the observations are independent of one another; and (2) there is a stationary relationship among variables, meaning a spatially constant relationship between dependent and independent variables that can be interpreted by average (global) parameter estimates across an entire study area (Fotheringham et al., 2002). However, “spatial is special” (Longley, Goodchild, Maguire, & Rhind, 2005, p. 5); spatial data exhibits spatial dependence (also known as spatial autocorrelation) and spatial heterogeneity (spatial non-stationarity) that make

## Exploring equity using geographically weighted regression

it difficult to meet the assumptions and requirements of traditional OLS regression and can bias OLS results (Fotheringham, Charlton, & Brunson, 1998; Fotheringham et al., 2002).

Spatial dependence is the extent to which the value of an attribute in one location is more likely to be similar to the value of the attribute in a nearby location than in a distant location (Mennis & Jordan, 2005). Spatial dependence is a function of Tobler's (1970) First Law of Geography, which stated that "everything is related to everything else, but near things are more related than distant things" (p. 236). Spatial dependence "is determined both by similarities in position, and by similarities in attributes" (Longley et al., 2005, p. 517). According to Anselin (1988), large residuals are likely to occur if geographic features are spatially autocorrelated when using non-spatial statistical methods such as OLS regression.

Spatial heterogeneity or non-stationarity refers to the tendency for "the relationships among the independent and dependent variables [to] vary over space" (Mennis & Jordan, 2005, p. 249). In other words, every location has an intrinsic level of uniqueness with regard to the causal relationship between variables that may not be described by constant global parameter estimates (Gilbert & Chakraborty, 2011; Fotheringham et al., 2002). When a lack of spatial uniformity or homogeneity is caused by the effects of spatial dependence and/or varying relationships between variables, spatial heterogeneity is likely to occur (Anselin, 1988).

Spatial heterogeneity can thus be regarded as a special case of spatial dependence, and spatial dependence and heterogeneity often occur jointly (Longley et al., 2005). Ignoring spatial heterogeneity gives rise to inaccurate regression results, such as biased parameter estimates and misleading significance tests (Anselin, 1988). Equity research based on linear statistical analyses has failed to account for these spatial effects, leading to violation of the basic assumptions of OLS, including linearity, homoscedasticity, and independence and

## Exploring equity using geographically weighted regression

normality of residuals. Meanwhile, research methods that address these spatial effects have remained underexploited by POS researchers and practitioners. This study provides a powerful demonstration of the improvements possible using spatially explicit regression techniques.

### *Geographically Weighted Regression (GWR)*

GWR has recently become a popular means of modeling local spatial heterogeneity between variables. GWR assumes that relationships between variables may differ from location to location (Fotheringham et al., 2002). In other words, GWR generates a set of local regression coefficients for each observation point in the study area.

The traditional multiple linear regression model can be expressed as follows:

$$y_i = a_0 + \sum_{k=1}^k a_k x_{ik} + e_i, k = 1, \dots, k,$$

where  $y_i$  is the vector of the estimated parameter for observation  $i$ ,  $a_0$  is the intercept parameter,  $a_k$  is the regression coefficient for the  $k_{th}$  independent variable,  $x_{ik}$  is the value of the  $k_{th}$  independent variable for observation  $i$ , and  $e_i$  is a random error term for observation  $i$ . As noted above, this model is based on assumptions of independence and homogeneity such that the residuals should be both independent and drawn identically from a normal distribution with a mean of zero (Fotheringham et al., 1998). GWR extends the traditional multiple linear regression framework by allowing local parameters to be estimated as follows:

$$y_i = a_{i0}(u_i, v_i) + \sum_{k=1}^k a_{ik}(u_i, v_i)x_{ik} + e_i, k = 1, \dots, k,$$

where  $(u_i, v_i)$  is the coordinate of the  $i_{th}$  point in the study area,  $a_{i0}(u_i, v_i)$  is the intercept parameter at point  $i$ ,  $a_{ik}(u_i, v_i)$  is the local regression coefficient for the  $k_{th}$  independent variable at point  $i$ , and  $x_{ik}$  is the value of the  $k_{th}$  independent variable at point  $i$ . Thus, unlike linear multiple regression models, GWR considers important local variations in relationships.

## Exploring equity using geographically weighted regression

Based on Tobler's (1970) First Law of Geography, all observed data points in GWR are weighted by their spatial proximity to the regression point, with observed data points closer to the regression point weighted more heavily than those located farther away (Fotheringham et al., 2002). The weight of an observed data point is thus at a maximum when it shares the same location as the regression point, and decreases as the distance between the two points increases.

In GWR, the weights of observed data points depend on the kernel chosen and that kernel's bandwidth (Fotheringham et al., 2002). A kernel can be defined as a circle of influence or circular area with a given radius around one particular regression point; the given radius is called the bandwidth (Zhang & Shi, 2004). The Gaussian and bi-square kernel functions are commonly used in GWR. The Gaussian kernel function is also referred to as a kernel with a fixed bandwidth because it is based on the assumption that the bandwidth at each regression point is consistent across the study area, and is applied when the observed data points are regularly spaced in the study area (Fotheringham et al., 2002). The weight for the Gaussian kernel function is estimated as follows:

$$w_{ij} = \exp [-(d_{ij}/ b)^2],$$

where  $d_{ij}$  is the Euclidean distance between the regression point  $i$  and the data point  $j$ , and  $b$  is the bandwidth. At the regression point, the weight of a data point is unity; weights decrease as the distance from the regression point increases. However, the weights of all the data points are non-zero, no matter how far they are from the regression point.

The bi-square kernel function is called a kernel with adaptive bandwidth because it permits use of variable bandwidth, and is used when the observed data points are clustered in the study area (Fotheringham et al., 2002). For example, the size of the bandwidth increases

## Exploring equity using geographically weighted regression

when the observed data points are widely spaced and decreases when they are closer. The weight for the bi-square kernel function is estimated as follows:

$$w_{ij} = [1 - (d_{ij} / b)^2] \text{ when } d_{ij} \leq b, w_{ij} = 0 \text{ when } d_{ij} > b$$

At the regression point  $i$ , the weight of the data point is unity and falls to zero when the distance between  $i$  and  $j$  equals the bandwidth. When the distance is greater than the bandwidth, the weight of the data point is zero. The bandwidth is selected so that there is the same number of data points with non-zero weights at each regression point.

Bandwidth has a substantial influence on GWR results (Gilbert & Charkraborty, 2011). Bandwidth can be thought of as a smoothing parameter; a larger bandwidth can cause greater smoothing. If the estimated parameters are similar in value across the study area, an over-smoothed model is applied, and if the estimated parameters include much local variation, an under-smoothed model is adopted. Somewhere between these two extremes is regarded as the best bandwidth (Fotheringham et al., 1998, 2002).

Three methods have commonly been used to determine the best bandwidth: (1) providing a user-supplied bandwidth; (2) selecting a bandwidth that minimizes a cross-validation (CV) function, and (3) selecting a bandwidth that minimizes the Akaike Information Criterion (AIC). The latter has most commonly been employed (Fotheringham et al., 2002). The AIC is a measure of relative model performance and is helpful for comparing different regression models.  $AIC_c$  is AIC with a correction for finite sample sizes (Bozdogan, 1987). This takes the following form:

$$AIC_c = 2n \log_e(\hat{\sigma}) + n \log_e(2\pi) + n[(n + \text{tr}(S))/(n-2 - \text{tr}(S))]$$

where  $n$  is the number of observations in the dataset,  $\hat{\sigma}$  is the estimate of the standard deviation of the residuals, and  $\text{tr}(S)$  is the trace of the hat matrix.  $AIC_c$  values can be used not only to



## Exploring equity using geographically weighted regression

compare models with different independent variables but also to compare the global model with a local GWR model (Bozdogan, 1987).

Compared to traditional OLS models, GWR offers two important benefits: (i) it yields error terms (residuals) that are considerably smaller and less spatially dependent than residuals from corresponding OLS models; and (ii) the ability to visualize spatial variations in regression diagnostics and model parameters (Gilbert & Charkraborty, 2011). Mapping regression diagnostics such as standardized residuals, local r-square, and parameter estimates can play an important role in exploring how statistical relationships and their significance vary over space.

### *GWR in the Context of Equity*

GWR has been employed to analyze environmental inequities in the distribution of a variety of undesirable land uses and their outcomes, including toxic air releases (Gilbert & Chakraborty, 2011; Mennis & Jordan, 2005) and air pollution (Jephcote & Chen, 2012). To date, however, only one study has used GWR to explore inequities in the distribution of desirable land uses such as POSs. Maroko et al. (2009) used both OLS and GWR to examine the statistical relationship between level of access to parks and residents' racial and ethnic status in New York City, US. The results indicated that the OLS model found a weak relationship with lower  $R^2$  and higher AIC, while GWR suggested spatial non-stationarity, indicating disparities in accessibility that vary over space with higher  $R^2$  and lower AIC.

## **Method**

### *Study Area: Detroit Metropolitan Area (DMA), Michigan*

The Detroit Metropolitan Area (DMA), also referred to as Metro Detroit, is located in southeast Michigan and includes three counties (Oakland, Wayne, and Macomb). The 12<sup>th</sup> largest metropolitan area in the US, the DMA had a population of 3,863,924 and an area of

## Exploring equity using geographically weighted regression

1,958.96 square miles (3,463.2 km<sup>2</sup>) in 2010 (U.S. Bureau of the Census, 2010). The DMA was chosen as the study area for two reasons. First, the DMA contains a high number and density of public beaches. According to the Michigan Department of Environmental Quality (MDEQ, 2013), almost 14.5% (n=178) of all public beaches in Michigan (n=1,224) are located in the DMA. Second, the DMA is home to the highest population density and most diverse population in Michigan. Whereas the population density of Michigan is 174.8 inhabitants per square mile (67.5/ km<sup>2</sup>), the population density of the DMA is 2,792.5 inhabitants per square mile (1,078.2/km<sup>2</sup>). The DMA's racial and ethnic composition is as follows: White (70.1%), African American (22.8%), Hispanic (6.2%), Asian (3.3%), Native American (0.3%), and Pacific Islander (0.02%) (U.S. Bureau of the Census, 2010).

### *Unit of Analysis*

The choice of areal unit is critical in any spatial analysis; this study employed the census tract (CT). A CT is defined as a subdivision of a county with “a mean population of approximately 4,000 people that are relatively homogeneous in socioeconomic characteristics” (Moore et al., 2008, p. 17). There are 1,164 CTs in the DMA. Figure 1 shows the locations of the 178 public beaches and the CT boundaries within the study area.

Figure 1 about here

### *Variable Definitions and Data Acquisition*

Level of access to public beaches served as the dependent variable. Access was measured in two manners: (1) the number of public beaches within 20 miles of each CT centroid, and (2) the shortest road network distance from each CT centroid to the nearest public beach. These two measures reflect the container and minimum distance approaches as explained by Talen and Anselin (1998). The container approach is simple and efficient. Haas

## Exploring equity using geographically weighted regression

(2009) estimated that residents were willing to travel 20 miles for beach-based recreation activities such as boating, fishing, and swimming. The number of public beaches within 20 network-distance miles of each CT centroid was therefore utilized as the container measure. Use of the minimum distance approach recognizes that, although an individual could theoretically interact with all the POSs in his or her local environment, most POSs such as parks are, in reality, mainly used by nearby residents. Use of two approaches enabled the equity findings to be compared and contrasted at each step of subsequent analysis. Due to its far superior representation of the actual landscape, only network distance was employed.

Multiple conceptualizations of equity exist, e.g., Wicks and Crompton (1986) identified the four equity models – equality, compensatory (or need), demand (or preferences), and market (or willingness to pay) – that have most commonly been employed in the parks and recreation profession. As described above, a compensatory or need-based model of equity has typically been employed to measure the equity of LDLUs, based on the assumption that in the public realm disadvantaged residents or the most needy groups or areas should be awarded (compensated with) extra services. A need-based definition of equity was therefore adopted. A variety of demographic and socioeconomic variables were considered to represent residents' need with regard to access to public beaches: (1) population density; (2) age; (3) race/ethnicity; (4) income; (5) housing value; (6) educational attainment; (7) language; (8) vehicle ownership; (9) housing occupancy; and (10) economic status. Groups considered most likely to be in need of better than average access to public beaches were those residing in more densely populated areas, the young and elderly, non-Whites, those earning low incomes and living in lower value housing, those having lower educational attainment, those with non-English spoken at home, those without a vehicle, and those residing in areas with lower proportions of occupied housing

## Exploring equity using geographically weighted regression

and higher poverty rates. Table 1 summarizes the variables and their operational definitions; it also indicates how an increase in the value of each dependent variable should be interpreted with respect to the need-based definition of equity employed.

Insert Table 1 about here

Geographic data such as CT boundaries and the street network were gathered from the Michigan GIS data library (<http://www.mcgi.state.mi.us/mgdl/>). Public beach locations were acquired from the MDEQ (<http://www.deq.state.mi.us/beach/>). Racial/ethnic and socioeconomic data for 2010 were obtained from the U.S. Bureau of the Census.

### *Data Analysis*

Data analysis was conducted using ArcGIS (version 10.0), the ArcGIS Network Analyst extension, SPSS (version 20.0), and GWR (version 4.0). Network analysis was employed to calculate the two dependent variables for each CT. Next, multivariate regression analysis using OLS was conducted to investigate the relationship between level of public beach access and residents' demographic and socioeconomic status. GWR was then conducted to explore spatial variations using the same dependent and independent variables. A bi-square kernel function was used due to the varying size and shape of CTs as well as varying density of public beaches in the DMA. The optimal kernel size was determined through an iterative statistical optimization process to minimize the  $AIC_c$ . Statistical diagnostics (e.g., local parameter estimates and local  $R^2$ ) from GWR were mapped to explore spatially varying relationships among variables;  $R^2$ ,  $AIC_c$ , and Moran's I of regression residuals were compared to quantify any improvement in model fit of GWR over OLS.

## **Results**

### *Estimated OLS Parameters*

## Exploring equity using geographically weighted regression

Two separate OLS regression analyses were performed to examine the effects of residents' demographic and socioeconomic status on the number of public beaches accessible within a 20-mile journey of each CT centroid (container approach, Model 1), and the minimum distance to the nearest public beach from each CT centroid (minimum distance approach, Model 2). Results of the two OLS models are presented in Table 2. Because the VIF values associated with MHI were greater than 7.5 (Model 1: 10.25; Model 2: 10.22), MHI was removed from the pool of independent variables due to the existence of collinearity.

For Model 1 (container approach), both the Joint F- and Joint Wald statistics indicated statistical significance for the overall model (Joint F: 55.59,  $p < 0.01$ ; Joint Wald: 1,008.19,  $p < 0.01$ ). The value of adjusted  $R^2$  (0.379) indicated a moderate goodness-of-fit. Five of thirteen independent variables (BLACK, ASIAN, POPD, EDU, and VEHIC) were statistically significant at the 0.05 level, suggesting equitable access to public beaches with respect to proportions of Black and Asian population but inequitable access with respect to population density, educational attainment, and vehicle ownership. These interpretations are due to the positive sign on the coefficients BLACK (0.190) and ASIAN (0.951) indicating an increase in proportion Black or Asian with the number of parks within 20 miles, the positive sign on the education coefficient (1.247) indicating an increase in the proportion of the population holding a four-year university degree or higher with an increasing number of parks, and the negative signs on the population density (-0.005) and vehicle ownership (-0.435) coefficients indicating a decrease in population density and proportion of households without a vehicle with an increasing number of parks. In all other cases the lack of significance associated with the coefficient indicated that no statistically meaningful relationship existed between the level of each independent variable and level of public beach access. The Koenker (BP) statistic (163.46,

## Exploring equity using geographically weighted regression

$p < 0.01$ ) indicated that Model 1 exhibited spatial non-stationarity, thus warranting GWR analysis.

For Model 2 (minimum distance approach), both the Joint F and Joint Wald statistics indicated statistical significance for the overall model (Joint F: 45.17,  $p < 0.01$ ; Joint Wald: 365.42,  $p < 0.01$ ) while the value of adjusted  $R^2$  (0.185) indicated a lower level of model performance than that of Model 1. Three of thirteen independent variables (POPD, AGE64, and EDU) were statistically significant at the 0.05 level, suggesting inequitable access to public beaches respect to population density, proportion of elderly population, and educational attainment, i.e., that as population density and proportion elderly increase, minimum distance to the nearest public beach also increases, whereas as proportion of the population holding a four-year university degree or higher increases, minimum distance to the nearest public beach declines. The Koenker (BP) statistic (97.63,  $p < 0.01$ ) indicated that Model 2 exhibited spatial non-stationarity, again suggesting additional GWR analysis.

Insert Table 2 about here

### *Estimated GWR Parameters*

Results of the two GWR models are presented in Table 3. For GWR Model 1 (container), the local adjusted  $R^2$  varied over the study area from a minimum of 0.02 to a maximum of 0.92 (mean: 0.69). The local condition index ranged from a minimum of 9.7 to a maximum of 24.8, indicating the absence of local collinearity among the independent variables. The ranges of the local coefficients for the variables significant in the OLS model were -126.40 to 67.72 with a mean of -1.98 (BLACK), -21.79 to 27.46 (mean:-1.39, ASIAN), -18.55 to 26.81 (mean: -1.36, POPD), -8.09 to 58.92 (mean: 4.87, EDU), and -25.34 to 19.55 (mean: -1.12, VEHIC), respectively. This variability in the local coefficients suggests that the

## Exploring equity using geographically weighted regression

relationships between the number of public beaches accessible within a 20-mile journey from each CT centroid, and residents' demographic and socioeconomic status, are not stationary.

For GWR Model 2 (minimum distance), the local adjusted  $R^2$  varied over the study area from a minimum of 0.27 to a maximum of 0.92 (mean: 0.70). The local condition index (which ranged from 8.6 to 24.4) indicated the absence of local collinearity among the independent variables. The ranges of the local coefficients for the variables significant in the OLS model were -1.29 to 1.40 (mean: 0.14, POPD), -1.01 to 2.85 (mean: 0.12, AGE64), and -3.25 to 2.73 (mean: -0.02, EDU), respectively, again suggesting non-stationary relationships between the variables.

Insert Table 3 about here

### *Spatially Varying Relationships Explored by GWR*

Although Table 3 suggests the existence of spatial variations in the local coefficients and goodness-of-fit of the two GWR models, it does not show how the relationships between level of access to public beaches and residents' demographic and socioeconomic status vary across the study area. Figures 2-11 map the spatial distribution of local coefficients and local  $R^2$  for those independent variables that were statistically significant in the two OLS models; lighter colors indicate negative values, whereas darker colors indicate positive values. These maps are also summarized in Table 4.

Insert Table 4 about here

*Model 1 BLACK (Figure 2).* The OLS coefficient for BLACK was 0.145 ( $p < 0.05$ ), indicating equitable access to public beaches with regard to Black population across the study area (Table 3). However, Figure 2 and Table 4 show that both positive ( $n = 523$ , 44.9%) and negative ( $n = 641$ , 55.0%) correlations occur. The local coefficients for BLACK ranged from -

## Exploring equity using geographically weighted regression

126.39 to 67.72 (mean: -1.98). Strong positive correlations (local coefficient  $> 31.7$  [2 standard deviations above the mean]), indicating equitable access to public beaches with respect to Black population, were observed in parts of Oakland and Macomb counties. Strong negative correlations (local coefficient  $< -35.66$  [2 standard deviations below the mean]), indicating inequitable access, emerged in parts of Macomb county. While 492 (42.2%) of the CTs had local coefficients greater than the OLS coefficient, 672 (57.7%) had lower local coefficients. This variability in the model parameters suggests that the relationship between number of public beaches accessible within a 20-mile journey and proportion of Black population is not stationary.

Insert Figure 2 about here

*Model 1 ASIAN (Figure 3).* The OLS coefficient for ASIAN was 0.092 ( $p < 0.05$ ), indicating equitable access to public beaches with regard to Asian population (Table 3). However, Figure 3 and Table 4 show that both positive ( $n = 678, 58.2\%$ ) and negative ( $n = 486, 41.7\%$ ) correlations occur. The local coefficients for ASIAN ranged from -21.79 to 27.46 (mean: -1.39). Strong positive correlations (local coefficient  $> 10.55$ ), indicating equitable access to public beaches with respect to Asian population, were observed in parts of Oakland and Macomb counties. Strong negative correlations (local coefficient  $< -13.33$ ), indicating inequitable access, emerged in parts of Oakland and Wayne counties. While 411 (35.3%) of the CTs had local coefficients greater than the OLS coefficient, 488 (41.9%) had lower local coefficients, indicating a non-stationary relationship between variables.

Insert Figure 3 about here

*Model 1 POPD (Figure 4).* The OLS coefficient for POPD was -0.270 ( $p < 0.05$ ), indicating inequitable access to public beaches with regard to population density (Table 3).



## Exploring equity using geographically weighted regression

However, Figure 4 and Table 4 show that both positive ( $n = 446$ , 38.3%) and negative ( $n = 718$ , 61.6%) correlations occur. The local coefficients for POPD ranged from -18.55 to 26.81 (mean: -1.36). Strong positive correlations (local coefficient  $> 9.12$ ), indicating equitable access to public beaches with respect to population density, were observed in parts of Oakland county. Strong negative correlations (local coefficient  $< -11.84$ ), indicating inequitable access, emerged in parts of Oakland, Macomb, and Wayne counties. While 447 (38.4%) of the CTs had local coefficients greater than the OLS coefficient, 717 (61.5%) had lower local coefficients, indicating a non-stationary relationship between variables.

Insert Figure 4 about here

*Model 1 EDU (Figure 5).* The OLS coefficient for EDU was 1.247 ( $p < 0.01$ ), indicating inequitable access to public beaches with regard to level of educational attainment (Table 3). However, Figure 5 and Table 4 show that both positive ( $n = 749$ , 64.3%) and negative ( $n = 415$ , 35.6%) correlations occur. The local coefficients for EDU ranged from -8.09 to 58.92 (mean: 4.87). Strong positive correlations (local coefficient  $> 15.95$ ), indicating equitable access to public beaches with respect to educational attainment, were observed in parts of Oakland and Macomb counties. Strong negative correlations (local coefficient  $< -6.21$ ), indicating inequitable access, emerged in parts of Oakland, Macomb, and Wayne counties. While 598 (51.3%) of the CTs had local coefficients greater than the OLS coefficient, 566 (46.6%) had lower local coefficients, indicating a non-stationary relationship between variables.

Insert Figure 5 about here

*Model 1 VEHIC (Figure 6).* The OLS coefficient for VEHIC was -0.101 ( $p < 0.05$ ), indicating inequitable access to public beaches with regard to vehicle ownership (Table 3). However, Figure 6 and Table 4 show that both positive ( $n = 480$ , 41.2%) and negative ( $n = 684$ ,

## Exploring equity using geographically weighted regression

58.7%) correlations occur. The local coefficients for VEHIC ranged from -29.34 to 58.92 (mean: 19.55). Strong positive correlations (local coefficient  $> 8.86$ ), indicating equitable access to public beaches with respect to vehicle ownership, were observed in parts of Oakland and Macomb counties. Strong negative correlations (local coefficient  $< -11.1$ ), indicating inequitable access, emerged in parts of Oakland, Macomb, and Wayne counties. While 630 (54.1%) of the CTs had local coefficients greater than the OLS coefficient, 534 (45.8%) had lower local coefficients, indicating a non-stationary relationship between variables.

Insert Figure 6 about here

*Model 1  $R^2$  (Figure 7).* The global value of  $R^2$  was 0.379 but the local value of  $R^2$  varied over the study area from 0.2 to 0.92 (mean: 0.690). The majority of the CTs ( $n = 1,120$ , 96.2%) had local  $R^2$  values greater than the global value of  $R^2$  while only 44 (3.7%) had local  $R^2$  values lower than the global value (Table 4). The local model had the best explanatory power across the study area (in excess of 80.0%). However, the local model had very low explanatory power in parts of Macomb and Wayne counties (as low as 20.0%), indicating that level of access to public beaches in these areas is not explained adequately by the set of explanatory variables. These findings indicate that the explanatory power of the local model is not stationary, i.e., that model performance is spatially heterogeneous across the study area.

Insert Figure 7 about here

*Model 2 POPD (Figure 8).* The OLS coefficient for POPD was 0.180 ( $p < 0.05$ ), indicating inequitable access to public beaches with regard to population density (Table 3). However, Figure 8 and Table 4 show that both positive ( $n=771$ , 66.2%) and negative ( $n=393$ , 33.7%) correlations occur. The local coefficients for POPD ranged from -1.29 to 1.40 (mean: 0.14). Strong positive correlations (local coefficient  $> 1.04$ ), indicating inequitable access to

## Exploring equity using geographically weighted regression

public beaches with respect to population density, were observed in parts of Oakland and Macomb counties. Strong negative correlations (local coefficient  $< -0.76$ ), indicating equitable access, emerged in parts of Oakland, Macomb, and Wayne counties. While 770 (66.1%) of the CT had local coefficients greater than the OLS coefficient, 394 (33.8%) had lower local coefficients, indicating a non-stationary relationship between variables.

Insert Figure 8 about here

*Model 2AGE64 (Figure 9).* The OLS coefficient for AGE64 was 0.084 ( $p < 0.05$ ), indicating inequitable access to public beaches with respect to elderly population (Table 3). However, Figure 9 and Table 4 show that both positive ( $n=628$ , 53.9%) and negative ( $n=536$ , 46.0%) correlations occur. The local coefficients for AGE64 ranged from -1.01 to 2.85 (mean: 0.12). Strong positive correlations (local coefficient  $> 1.06$ ), indicating equitable access to public beaches with regard to elderly population, were observed in parts of Oakland county. Strong negative correlations (local coefficient  $< -0.82$ ), indicating inequitable access, emerged in parts of Macomb, Oakland, and Wayne counties. While 550 (67.5%) of the CTs had local coefficients greater than the OLS coefficient, 614 (52.7%) had lower local coefficients, indicating a non-stationary relationship between variables.

Insert Figure 9 about here

*Model 2EDU (Figure 10).* The OLS coefficient for EDU was -0.257 ( $p < 0.05$ ), indicating inequitable access to public beaches with regard to educational attainment (Table 3). However, Figure 10 and Table 4 show that both positive ( $n=536$ , 46.0%) and negative ( $n=628$ , 53.9%) correlations occur. The local coefficients for EDU ranged from -3.25 to 2.73 (mean: -0.02). Strong positive correlations (local coefficient  $> 1.82$ ), indicating equitable access to public beaches with respect to educational attainment, were observed in parts of Macomb and

## Exploring equity using geographically weighted regression

Wayne counties. Strong negative correlations (local coefficient  $< -1.86$ ), indicating inequitable access, emerged in parts of Macomb and Wayne counties. While 566 (48.6%) of the CTs had local coefficients greater than the OLS coefficient, 598 (51.3%) had lower local coefficients, indicating a non-stationary relationship between variables.

Insert Figure 10 about here

*Model 2R<sup>2</sup> (Figure 11).* The global value of  $R^2$  was 0.185 but the local value of  $R^2$  varied over the study area from 0.27 to 0.92 (mean: 0.70). All CTs (n=1,164, 100.0%) had local  $R^2$  values greater than the global value. The local model had the best explanatory power in parts of Wayne, Oakland, and Macomb counties (in excess of 80.0%), though it performed less well in parts of Oakland county (as low as 27.0%)

Insert Figure 11 about here

### *Comparison of Spatial Autocorrelations of Residuals between OLS and GWR*

Given the statistically significant spatial clustering of high and low residuals, global Moran's I of residuals from each of the OLS and GWR models were computed to compared the degree of spatial autocorrelation between them (Table 5). Although significant positive spatial autocorrelation was found for both OLS models (Moran's I statistic [Model 1: 0.36; Model 2: 0.61] and p-value [Model 1:  $p < 0.05$ ; Model 2:  $p < 0.05$ ]), and both GWR models (Moran's I statistic [Model 1: 0.10; Model 2: 0.15] and p-value [Model 1:  $p < 0.05$ ; Model 2:  $p < 0.05$ ]), the global Moran's I statistics for the two GWR models were much lower than those for the OLS models. These findings show that GWR models can improve model fit by reducing the spatial autocorrelation in the residuals.

Insert Table 5 about here

### *Comparison of Model Performance between OLS and GWR*

Model performance was evaluated by comparing the  $R^2$  and the  $AIC_c$  values for the OLS and GWR models. The lower the  $AIC_c$  and higher the  $R^2$  value the better (Gilbert & Chakraborty, 2011). If the adjusted  $R^2$  value of the GWR model is higher and the  $AIC_c$  value is at least three points lower than that of the OLS, the GWR model is considered to significantly improve upon its corresponding OLS model. For Model 1, the adjusted  $R^2$  value dramatically increased from 0.379 (OLS) to 0.693 (GWR).  $AIC_c$  decreased from 11,839.75 (OLS) to 8,679.89 (GWR). For Model 2, the adjusted  $R^2$  value dramatically increased from 0.185 (OLS) to 0.702 (GWR).  $AIC_c$  decreased from 6,300.11 (OLS) to 4,085.73 (GWR). These findings indicate that GWR models provide significantly better goodness-of-fit than OLS models when assessing the spatial distribution of access to public beaches in the DMA.

### **Discussion and Implications**

This study has demonstrated the utility and feasibility of GWR when measuring the degree of equity inherent in the distribution of access to POSs. It is one of the first papers in the recreation/parks field to employ GWR, thereby making both methodological and practical contributions to the literature. As seen in Table 3, the two GWR models produced great improvements in model performance (as measured by  $R^2$ ,  $AIC_c$ , and Moran's I statistics of standardized residuals) over the corresponding OLS models. Although the OLS  $R^2$  values (Model 1: 0.379; Model 2: 0.185) were generally on par with those of previous POS equity studies (Deng et al., 2008 [ $R^2$ : 0.28]; Maroko et al., 2009 [ $R^2$ : 0.23]; Porter & Tarrant, 2001 [ $R^2$ : 0.18]; Tarrant & Cordell, 1999 [ $R^2$ : 0.27]), those relatively low levels of explanatory power imply that the OLS models may not have been properly specified due to (i) model misspecification and/or (ii) spatial effects. First, there may be some missing determinants of level of access to POSs that could improve model performance. Second, local variations might exist

## Exploring equity using geographically weighted regression

in the relationships between level of access and residents' demographic and socioeconomic status that reduce the explanatory power of the global model. Several authors such as Anselin (1988) and Fotheringham et al. (2002) have shown that local variations between variables can reduce the explanatory power of models when employing traditional multivariate techniques. However, as anticipated, the GWR models in this study provided more desirable statistical results, including higher  $R^2$ , lower standardized residuals, and lower  $AIC_c$ , than the OLS models (Table 3). Thus, this study provides strong evidence in support of the suggestion that GWR models can provide better goodness-of-fit than OLS models when assessing the spatial distribution of access to POSs such as public beaches in the DMA. This statement is consistent with previous equity studies of locally unwanted land uses (Gilbert & Charkraborty, 2011; Mennis & Jordan, 2005) and urban parks (Maroko et al., 2009). These findings not only indicate the need for researchers to realize the utility of GWR, but also suggest the desirability of additional data collection at the individual level, e.g., via a resident survey or qualitative methods, to identify missing explanatory variables that might even further improve model performance (whether using OLS or GWR).

The GWR models identified spatially varying relationships between level of access to public beaches and residents' demographic and socioeconomic status, highlighting the intricate patterns of access and equity that simply cannot be identified using global OLS techniques (Figures 2-11). This finding is consistent with those of Maroko et al. (2009), the only other known POS equity study to employ GWR, which indicated local variations between level of access to POSs and residents' demographic and socioeconomic status across New York City. As noted by Fotheringham et al. (1998), "there are spatial variations in people's tastes or attitudes or there are different administrative, political, or other contextual issues that produce

## Exploring equity using geographically weighted regression

different responses to the same stimuli across space” (p.1906). While this study clearly demonstrates both the variations in statistical relationships between the level of public beach access and residents' demographic and socioeconomic status across the DMA, and the utility of GWR as an exploratory spatial data technique, the findings also represent a starting point for future quantitative or qualitative investigations into the various social, political, economic, and historical factors associated with, i.e., that might help explain, the inequities of access to POSs observed in specific areas. The study suggests that a more detailed analysis of the interrelationships between residents' characteristics and attitudes, the layout of road networks, and land use and settlement patterns, should be conducted to understand how and why analytical results for variables differ across a study area.

The GWR models also provided insight with respect to the sign and magnitude of the parameter estimates. As shown in Table 2, OLS Model 1 indicated that equitable access to public beaches exists with respect to the Black and Asian populations. These findings were unexpected in this study area and are inconsistent with previous studies (Abercrombie et al., 2008; Bryne et al., 2009; Deng et al., 2008; Moore et al., 2008; Talen, 1998); further analysis using GWR indicated the influence of local variations between the variables caused by spatial dependence and spatial heterogeneity. Specifically, GWR Model 1 indicated equitable access to public beaches with respect to Black population in parts of Oakland and Macomb counties, but inequitable access in parts of Macomb county (Figure 2). Similarly, though equitable access to public beaches with respect to Asian population was observed in parts of Oakland and Macomb counties, inequitable access emerged in parts of Wayne county (Figure 3). Ignoring local variations between variables can lead to biased estimation results (Anselin, 1988). OLS Model 1 failed to explore important local variations between variables. As a result, the positive

## Exploring equity using geographically weighted regression

global coefficients of BLACK (0.190) and ASIAN (0.951) were obtained through a linear combination of the independent variables without any consideration of spatial effects. However, the mean GWR coefficients of BLACK (-1.98) and ASIAN (-1.39) for Model 1 indicated inequitable access to public beaches among the Black and Asian populations, by exploring local variations between the variables (Table 3). These results are consistent with those of previous POS equity studies and clearly demonstrate the additional insight and detail provided when using GWR. Though neither method allows for cause-and-effect relationships to be established, the findings can be considered in the context of several relevant theories. First, the market-based equity approach (Wicks & Crompton, 1986) suggests that an inequity in goods and services distribution occurs if minority groups cannot afford the necessary market price. The median household income (MHI) of Oakland county (\$65,636) is substantially greater than those of Wayne (\$41,504) and Macomb (\$53,628); similarly, the median housing value (MHV) of Oakland county (\$177,600) exceeds those of Wayne (\$97,100) and Macomb (\$134,700). Not only do the residents of Oakland county exhibit higher levels of purchasing power (e.g., higher incomes and housing values), but they are able to use that purchasing power to acquire properties in more attractive areas close to desirable amenities. Authors such as Nicholls and Crompton (2005a, 2005b, 2007) have demonstrated the premiums associated with properties adjacent to or nearby a variety of land- and water-based recreation opportunities. Also of relevance is MacIntyre's (2000) model of "deprivation amplification," which refers to a pattern of diminished opportunities related to the features of the local environment. As noted by Taylor et al. (2007, p. 55), "deprivation amplification indicates that in places where people have limited resources (e.g., money, private transportation), there are fewer safe, open green spaces where people can walk, jog, or take their children to play." Lastly, the theory of "marginality,"



## Exploring equity using geographically weighted regression

which identified a variety of socio-cultural, political, and economic constraints that tend to influence disadvantaged groups' difficulties in gaining access to resources (Park, 1928), may also be implied. As noted by West (1989, p. 11), "because of lower incomes, minorities are seen as having constraints on their ability to afford the cost of participation, or of transportation to recreation sites."

The findings of this study also suggest significant methodological and practical implications for community recreation planning and management. Methodologically, the GWR approach described here constitutes a substantial advance over the use of traditional OLS methods to measure the equity of POSs. Specifically, the GWR approach dealt with spatial effects such as spatial dependence and spatial heterogeneity that can lead to biased estimation results, thereby providing more accurate estimation results with better model performance compared to the traditional OLS approach.

The application of GWR also enables broadening of the scope of the research question. Traditionally, the fundamental goal of equity-related research in the urban service delivery literature has been limited to identifying "who gets what" in the context of environmental or territorial justice (Talen, 1998, p. 22). This study, however, widened the focus from "who gets what" to "who gets, what, where, and to what extent/how significantly," allowing identification of neighborhoods with inequitable access to public beaches specific to particular demographic and socioeconomic variables (Table 6 lists these locations). Such results can guide those state and local leisure agencies whose missions include concern for the provision of equitable access, by identifying the people and places most in need of increased public service delivery. This information can also assist local advocacy groups, community organizations, and minority populations in their attempts to provide or gain equitable access to POS-based recreation

## Exploring equity using geographically weighted regression

opportunities. Besides methodological development of an improved approach to the identification and measurement of equity, this study also offers parks and recreation agencies a tool via which they can better understand local patterns of access and equity and thus facilitate the formulation of locally appropriate policy solutions as and where needed, i.e., such findings may be used by leisure agencies to allocate limited budgets more efficiently by accurately pinpointing the most disadvantaged or needy areas and populations. Given that the existence of a natural beach is dependent on the presence of a water body, and that the construction of new water bodies is likely unrealistic, more feasible options in the Detroit case are the installation of spray parks at existing public park facilities, or the consideration of partnerships with local transportation providers to facilitate access to existing beaches. Moreover, the results of this study may facilitate a more informed decision making process because active stakeholder involvement, an essential part of the participatory approach, can be influenced positively by increased access to and interaction with information, especially when it is provided in visual, e.g., map, form (Yang, Madden, Kim, & Jordan, 2012). Information regarding spatial patterns of access to public beaches, residents' demographic and socioeconomic characteristics, and knowledge of the local variations in relationships among these variables could contribute to a spatial decision support system through the integration of Web-based GIS for more open, effective and efficient community-based leisure planning. Such systems also allow for improved accountability and openness on the part of public agencies.

### **Limitations and Future Studies**

Despite the many promising aspects of GWR, several limitations should be acknowledged. First, when measuring the level of access to public beaches, this study did not consider other objective and subjective factors, such as awareness of the location of POSs, POS

## Exploring equity using geographically weighted regression

size, environmental quality, and perceived or actual levels of crowding and safety, all of which can impact residents' recreation destination choice. To provide more comprehensive assessments of overall accessibility, future studies should incorporate one or more of these variables into their analyses. Second, findings are limited to a single POS type and geographic location (public beaches in the DMA) and are likely not generalizable. Additional studies of other geographic regions and POS types should be conducted to further demonstrate the utility and applicability of GWR, and to provide useful access/equity data to the POS providers in those communities. Third, this study does not consider the modifiable areal unit problem, a statistical bias that can radically affect the results of statistical tests due to the choice of district boundaries (Longley et al., 2005). Future studies should identify the sensitivity of multiple scales when measuring the accessibility and equity of public beaches. Lastly, while the GWR models do better capture spatial autocorrelation patterns in the dataset than their OLS counterparts, they do not control for all of it, as shown in Table 5. Better diagnostic tools and remedial methods to address this limitation are still required and should be integrated into future investigations; alternatively, the impacts of using different weighting systems could be explored.

## References

- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Boston, MA: Kluwer Academic Publishers.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370.
- Byrne, J., Wolch, J., & Zhang, J. (2009). Planning for environmental justice in an urban national park. *Journal of Environmental Planning and Management*, 52(3), 365-392.
- Deng, J., Walker, G., & Strager, M. (2008). Assessment of territorial justice using geographic information systems: A case study of distributional equity of golf courses in Calgary, Canada. *Leisure/Loisir*, 32(1), 203-230.
- Estabrooks, P. A., Lee, R. E., & Gyurcsik, N. C. (2003). Resources for physical activity participation: Does availability and accessibility differ by neighborhood socioeconomic status? *Annals of Behavioral Medicine*, 25(2), 100-104.
- Fotheringham, S. A., Charlton, M., & Brunson, C. (1998). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30, 1905-1927.
- Fotheringham, S. A., Brunson, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. New York, NY: Wiley.
- Gilbert, A., & Chakraborty, J. (2011). Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. *Social Science Research*, 40(1), 273-286.
- Hass, K. (2009). *Measuring accessibility of regional parks: A comparison of three GIS techniques* (Unpublished master's thesis). San Jose State University, San Jose, CA.

## Exploring equity using geographically weighted regression

- Jephcote, C., & Chen, H. (2012). Environmental injustices of children's exposure to air pollution from road-transport within the model British multicultural city of Leicester: 2000-09. *Science of the Total Environment*, 414, 140-151.
- Lindsey, G., Maraj, M., & Kuan, S. (2001). Access, equity, and urban greenways: An exploratory investigation. *The Professional Geographer*, 53(3), 332-346.
- Longley, P., Goodchild, M. F., Maguire, D., & Rhind, D. (2005). *Geographic information systems and science*. New York, NY: Wiley.
- Lotfi, S., & Koohsari, M. J. (2009). Measuring objective accessibility to neighborhood facilities in the city (A case study: Zone 6 in Tehran, Iran). *Cities*, 26(3), 133-140.
- MacIntyre, S. (2000). The social patterning of exercise behaviours: The role of personal and local resources. *British Journal of Sports Medicine*, 34(1), 6-6.
- Maroko, A. R., Manntay, J. A., Sohler, N. L., Grady, K. L., & Arno, P. S. (2009). The complexities of measuring access to parks and physical activity sites in New York City: A quantitative and qualitative approach. *International Journal of Health Geographies*, 8(1), 1-23.
- Mennis, J. L., & Jordan, L. (2005). The distribution of environmental equity: Exploring spatial nonstationarity in multivariate models of air toxic releases. *Annals of the Association of American Geographers*, 95(2), 249-268.
- Michigan Department of Environmental Quality (2013). *Michigan beaches*. Retrieved May 27, 2014, from <http://www.dep.state.mi.us/beach/>
- Moore, L. V., Diez Roux, A. V., Evenson, K. R., McGinn, A. P., & Brines, S. J. (2008). Availability of recreational resources in minority and low socioeconomic status areas. *American Journal of Preventive Medicine*, 34(1), 16-22.

## Exploring equity using geographically weighted regression

Nicholls, S. (2001). Measuring the accessibility and equity of public parks: A case study using GIS. *Managing Leisure*, 6(4), 201-219.

Nicholls, S., & Crompton, J. L. (2005a). Impacts of regional parks on property values in Texas. *Journal of Park and Recreation Administration*, 23(2), 87-108.

Nicholls, S., & Crompton, J. L. (2005b). The impact of greenways on property values: Evidence from Austin, Texas. *Journal of Leisure Research*, 37(3), 321-341.

Nicholls, S., & Crompton, J. L. (2007). The impact of a golf course on residential property values. *Journal of Sport Management*, 21(4), 555-570.

Nicholls, S., & Shafer, C. S. (2001). Measuring accessibility and equity in a local park system: The utility of geospatial technologies to park and recreation professionals. *Journal of Park and Recreation Administration*, 19(4), 102-124.

Porter, R. (2001). *Environmental justice and north Georgia wilderness areas: A GIS based analysis* (Unpublished doctoral dissertation). University of Georgia, Athens, GA.

Porter, R., & Tarrant, M. A. (2001). A case study of environmental justice and federal tourism sites in southern Appalachia: A GIS application. *Journal of Travel Research*, 40(1), 27-40.

Smoyer-Tomic, K. E., Hewko, J. N., & Hodgson, M. J. (2004). Spatial accessibility and equity of playgrounds in Edmonton, Canada. *The Canadian Geographer/Le Geographe Canadien*, 48(3), 287-302.

Talen, E. (1997). The social equity of urban service distribution: An exploration of park access in Pueblo, Colorado, and Macon, Georgia. *Urban Geography*, 18(6), 521-541.

Talen, E. (1998). Visualizing fairness: Equity maps for planners. *Journal of the American Planning Association*, 64(1), 22-38.

Talen, E., & Anselin, L. (1998). Assessing spatial equity: An evaluation of measures of

## Exploring equity using geographically weighted regression

- accessibility to public playgrounds. *Environmental and Planning A*, 30(4), 595-613.
- Tarrant, M. A., & Cordell, H. K. (1999). Environmental justice and the spatial distribution of outdoor recreation sites: An application of geographic information systems. *Journal of Leisure Research*, 31(1), 18-34.
- Taylor, W., Floyd, M., Whitt-Glover, M., & Brooks, J. (2007). Environmental justice: A framework for collaboration between the public health and parks and recreation fields to study disparities in physical activity. *Journal of Physical Activity & Health*, 4(1), 50-63.
- Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2), 234-240.
- United States Bureau of the Census.(2010). American fact finder. Retrieved December 28, 2014, from [http:// factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml](http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml).
- Wicks, B. E., & Crompton, J. L. (1986). Citizen and administrator perspectives of equity in the delivery of park services. *Leisure Sciences*, 8, 341-365.
- Yang, B., Madden, M. Kim, J., & Jordan, T. R. (2012). Geospatial analysis of barrier island beach availability to tourists. *Tourism Management*, 33(4), 840-854.
- Zhang, L., & Shi, H. (2004). Local modeling of tree growth by geographically weighted regression. *Forest Science*, 50(2), 225-244.

Exploring equity using geographically weighted regression

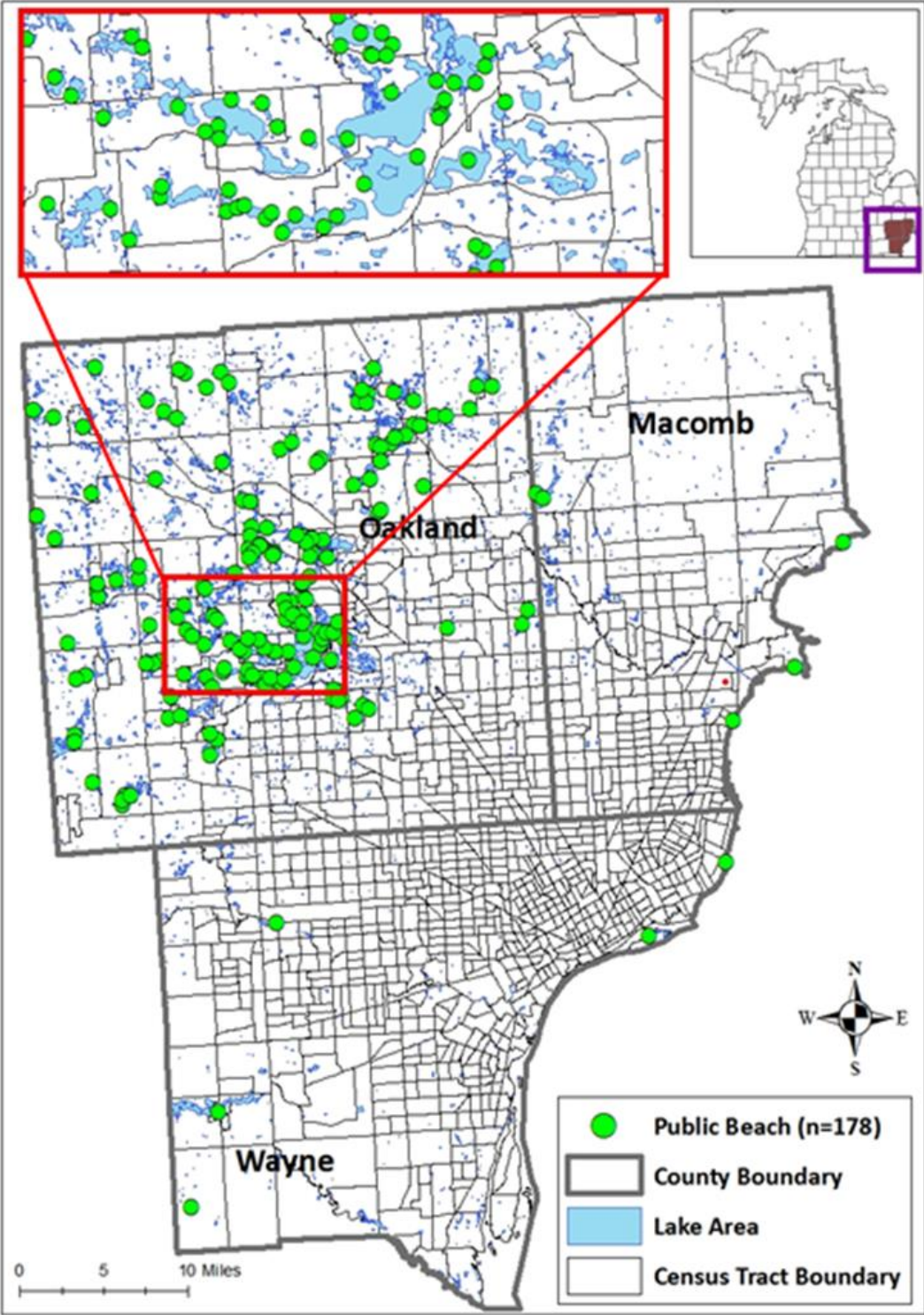


Figure 1. Study area



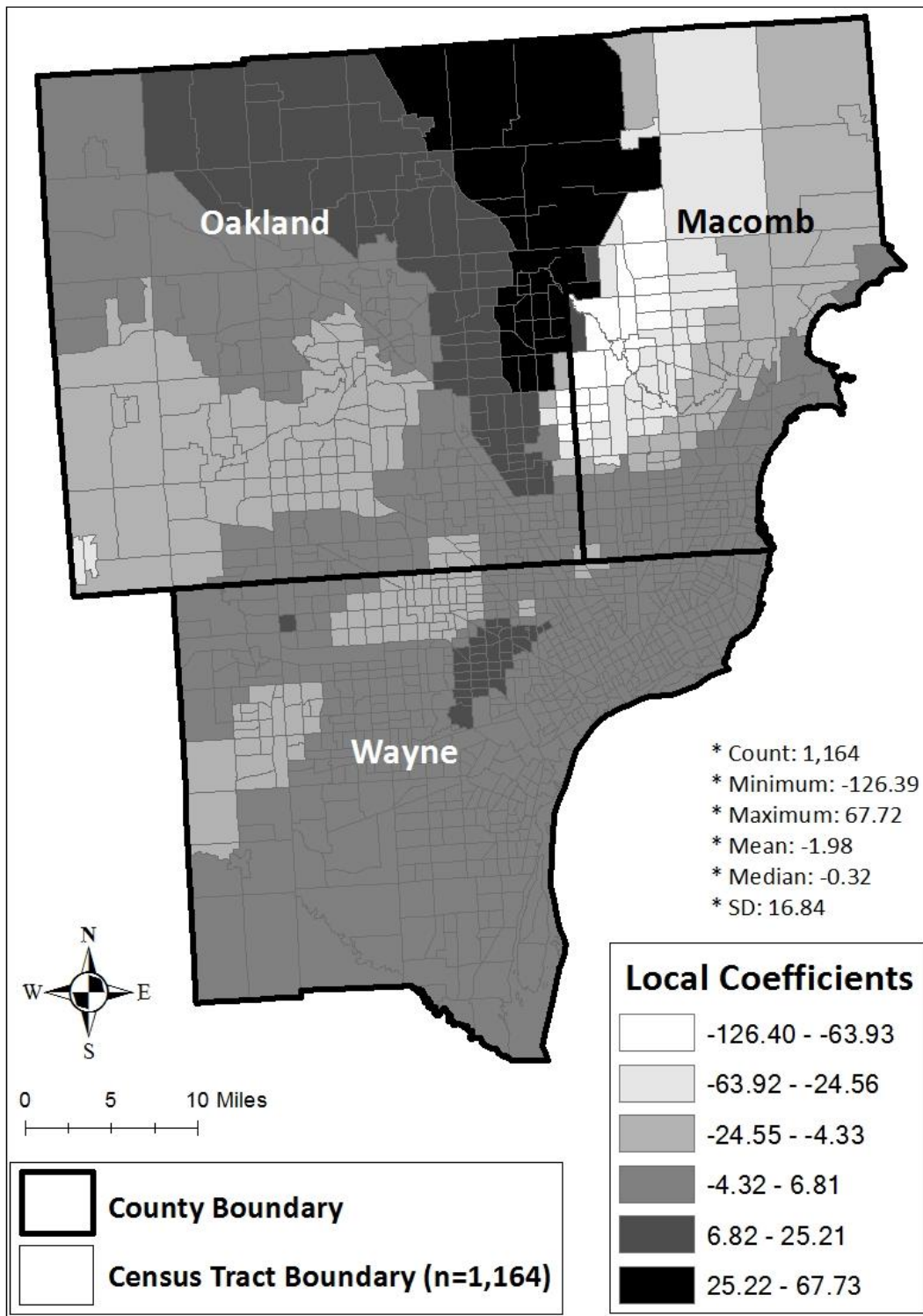


Figure 2. Spatial distribution of local parameter estimate for proportion (%) of Black population by census tract, DMA (Model 1)

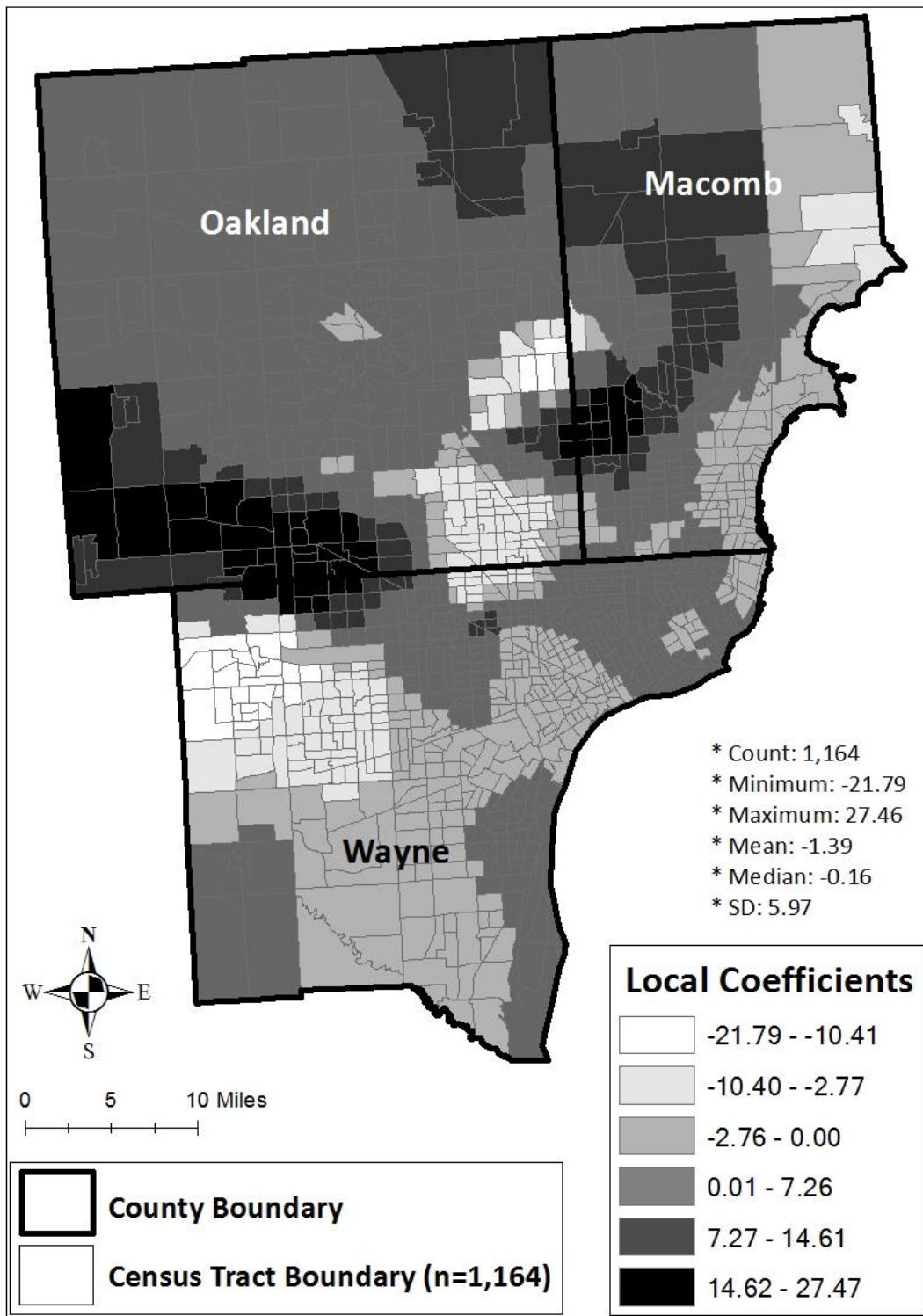


Figure 3. Spatial distribution of local parameter estimate for proportion (%) of Asian population by census tract, DMA (Model 1)

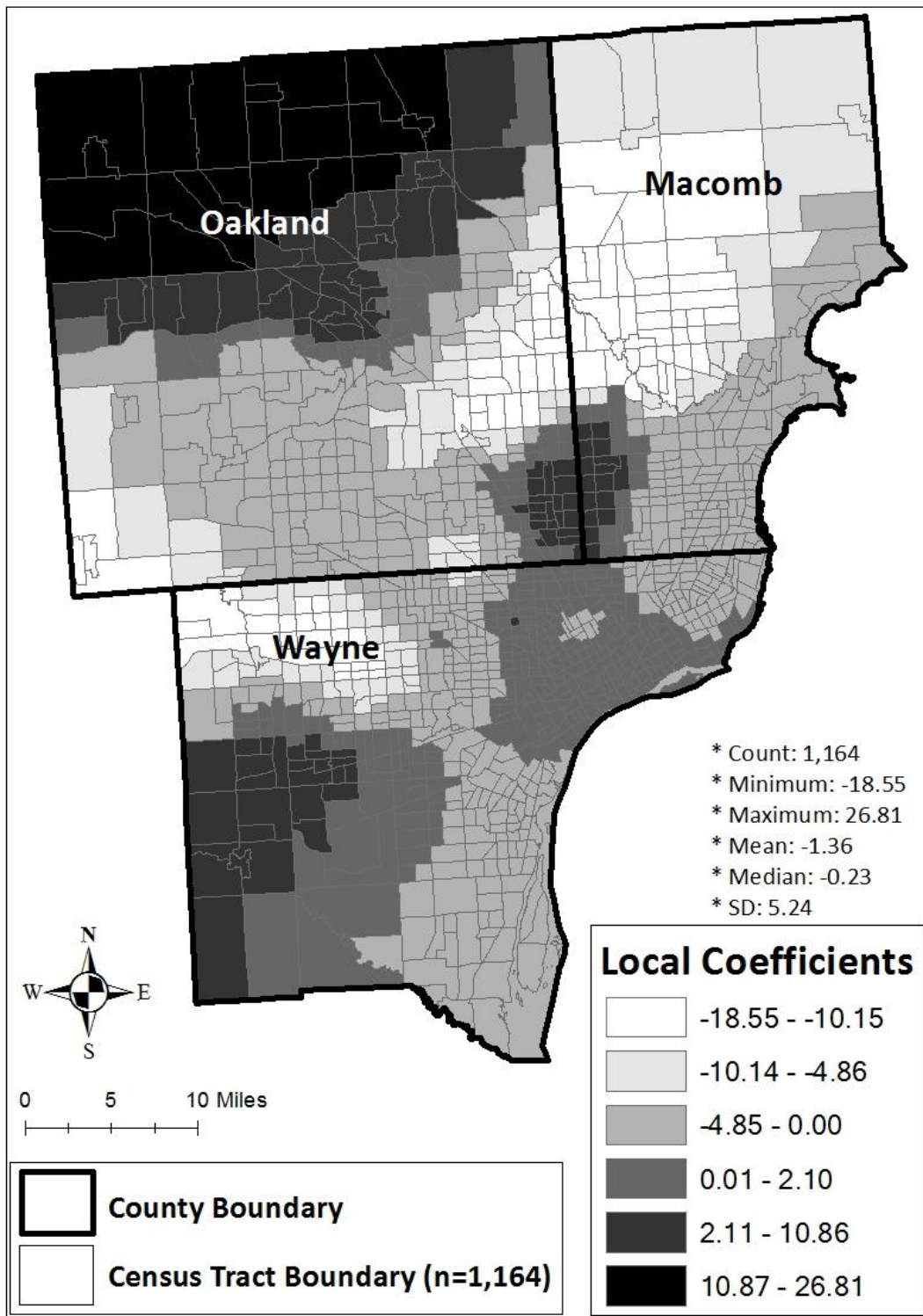


Figure 4. Spatial distribution of local parameter estimate for population per square mile by census tract, DMA (Model 1)

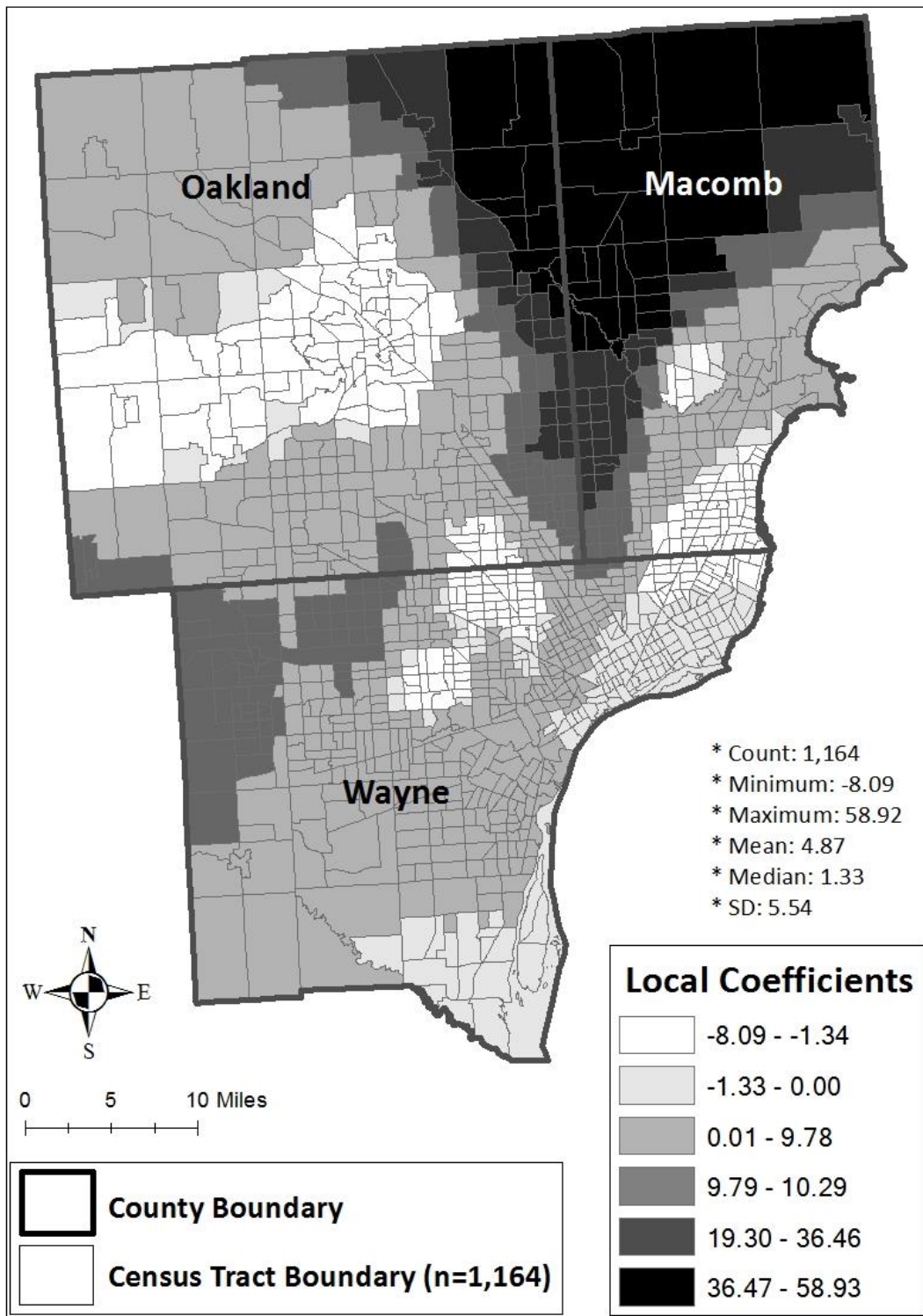


Figure 5. Spatial distribution of local parameter estimate for population with a four-year university degree or higher by census tract, DMA (Model 1)



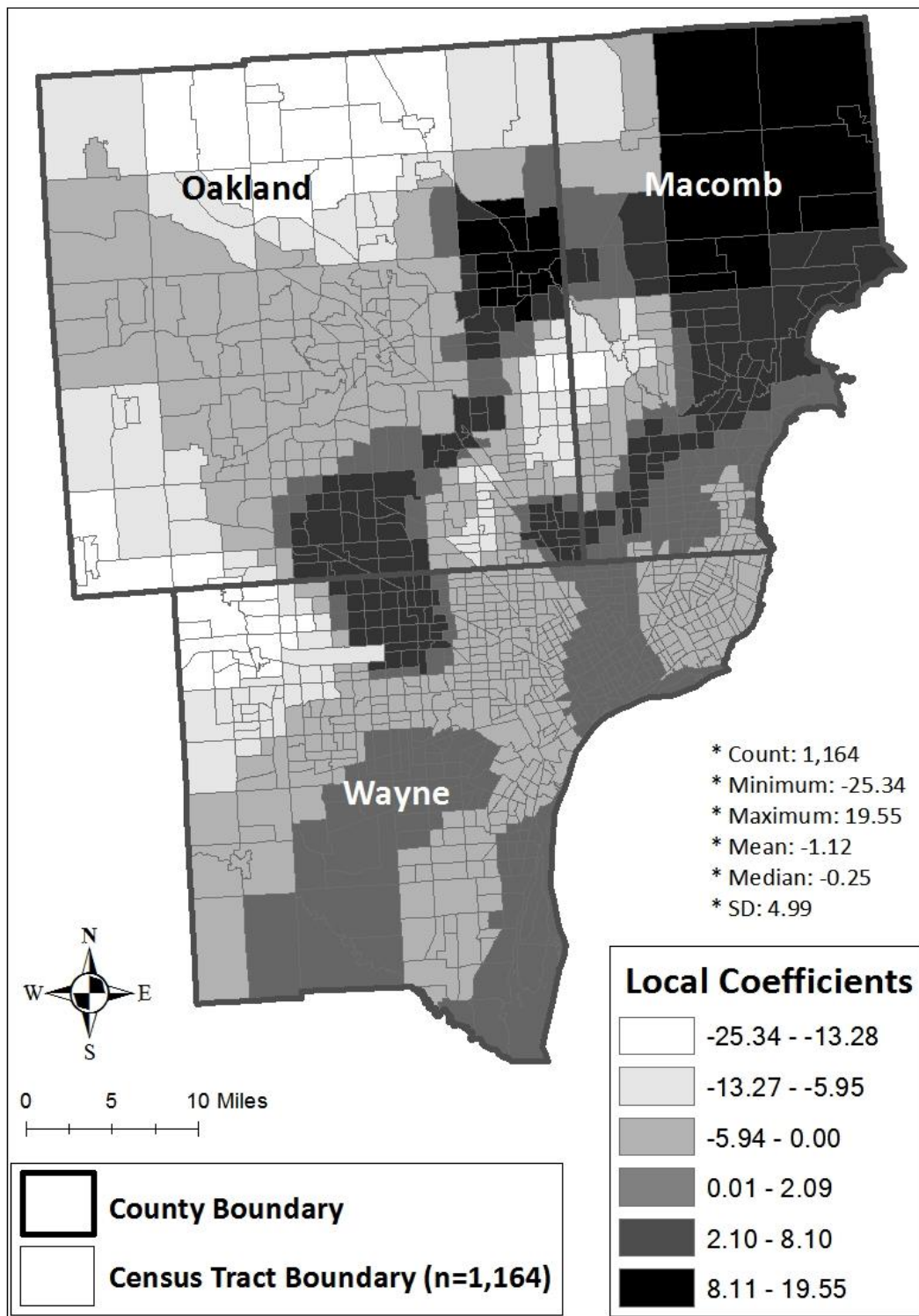


Figure 6. Spatial distribution of local parameter estimate for proportion (%) of households without a vehicle by census tract, DMA (Model 1)

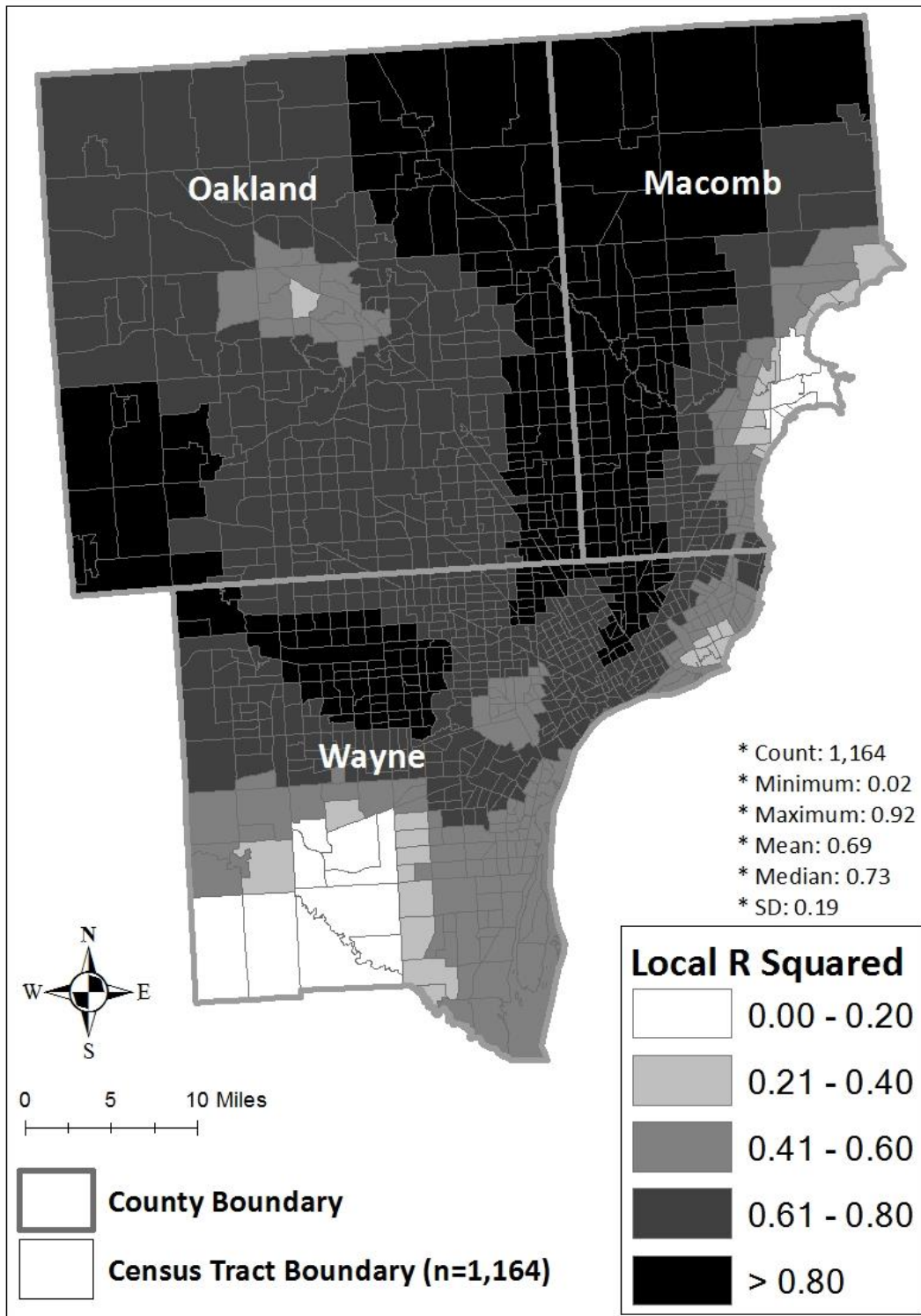


Figure 7. Spatial distribution of local  $R^2$  by census tract, DMA (Model 1)

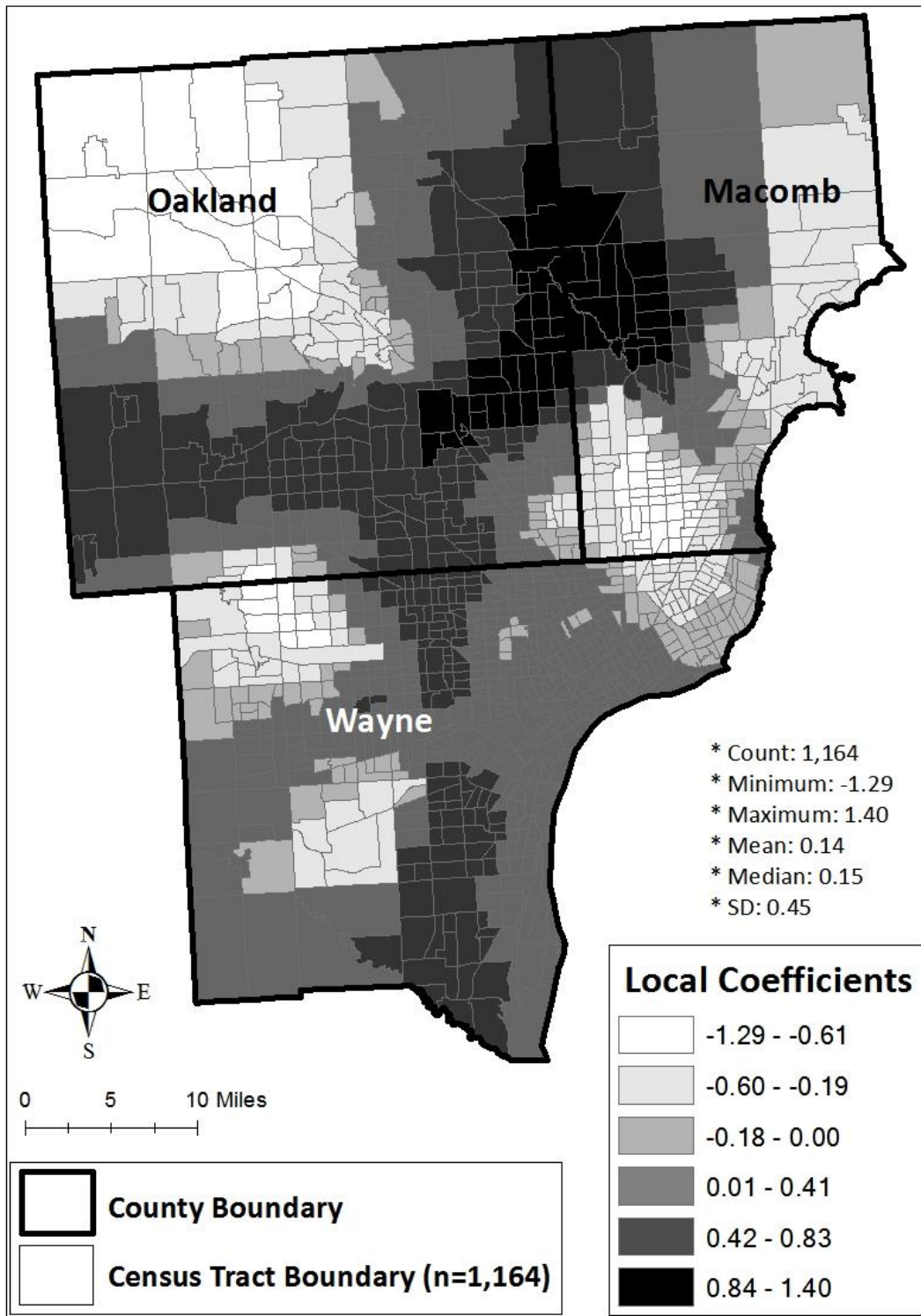


Figure 8. Spatial distribution of local parameter estimate for population per square mile by census tract, DMA (Model 2)

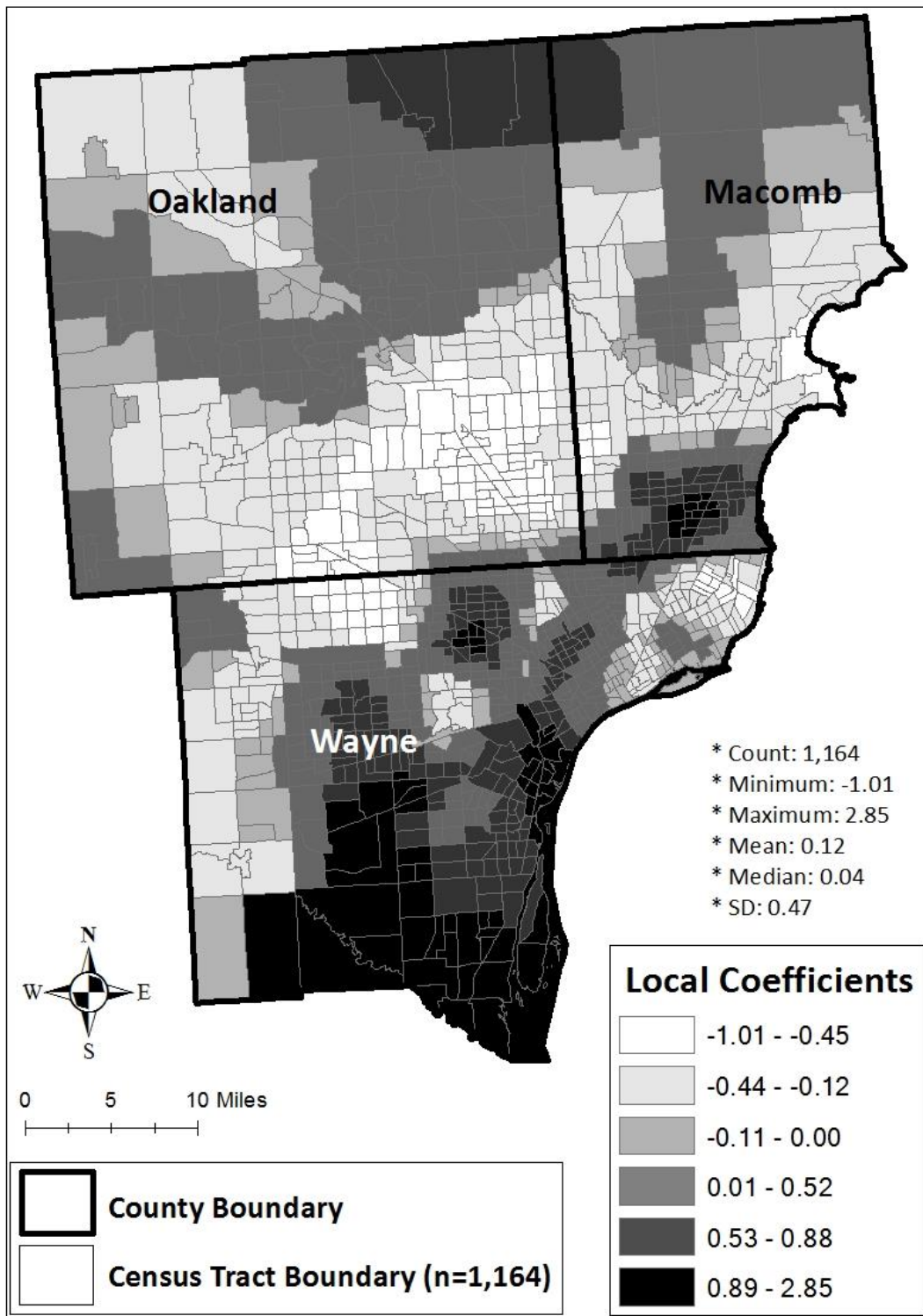


Figure 9. Spatial distribution of local parameter estimate for proportion (%) of population over age 64 by census tract, DMA (Model 2)



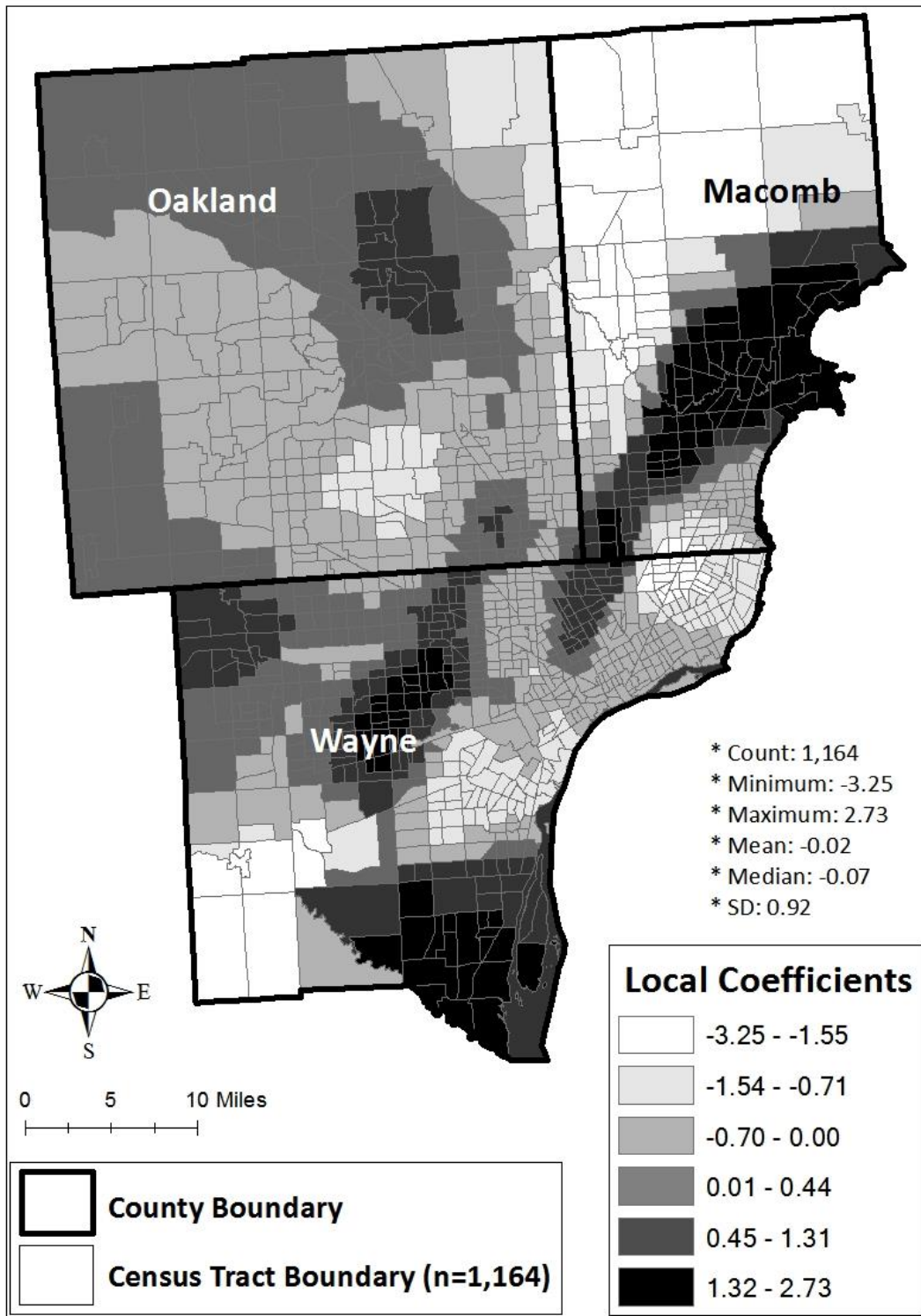


Figure 10. Spatial distribution of local parameter estimate for population with a four-year university degree or higher by census tract, DMA (Model 2)

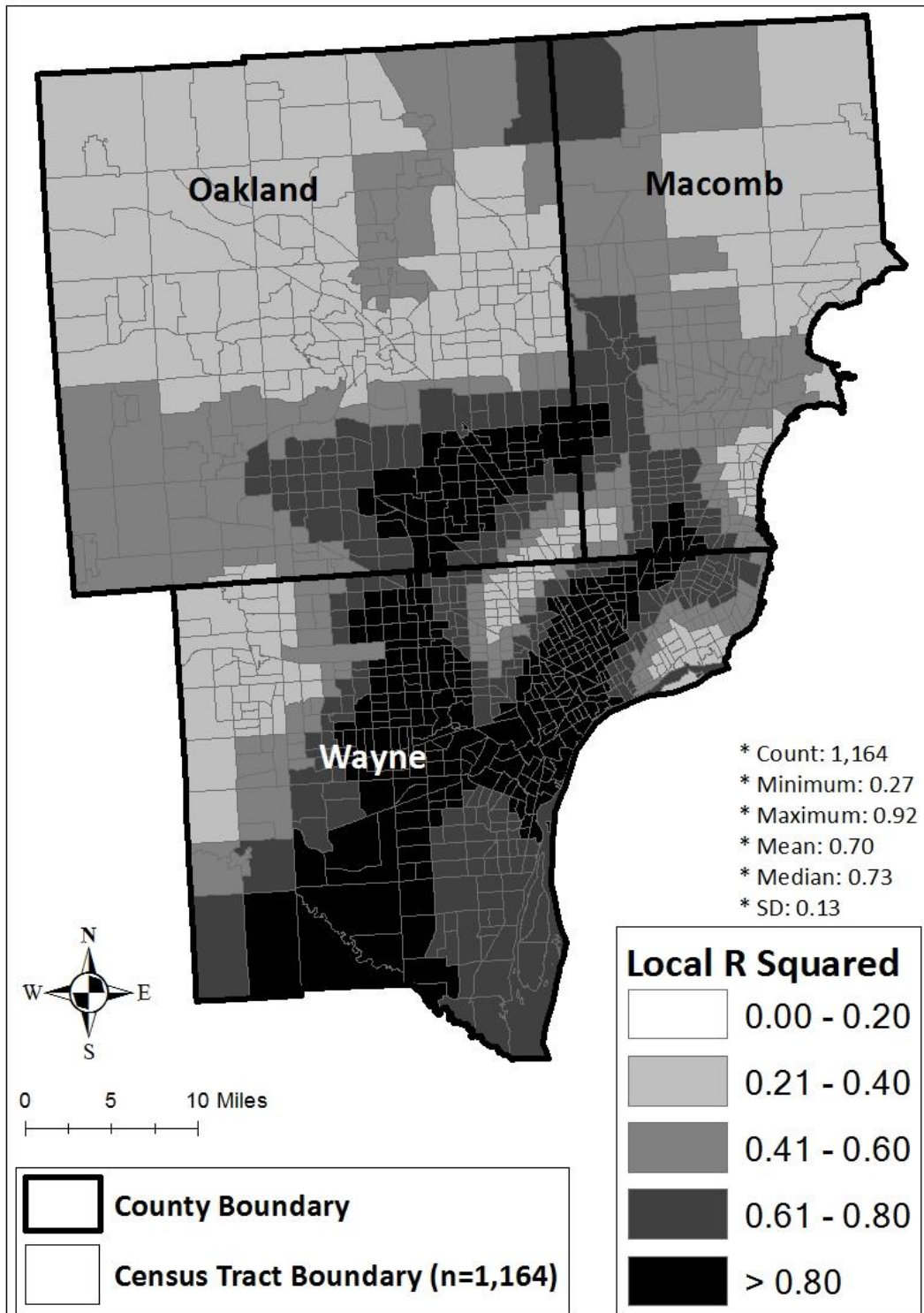


Figure 11. Spatial distribution of local  $R^2$  by census tract, DMA (Model 2)

Exploring equity using geographically weighted regression

Table 1. Dependent and independent variables

Variable	Operational definition	Abbreviation	Increasing level of equity associated with independent variable indicated when dependent variable....
Level of access to public beaches (DV)	(1) Number of public beaches within 20 miles of each CT	(1) NOPB	
	(2) Shortest road network distance from CT to the nearest public beach (in miles)	(2) DISTPB	
Population density (IV)	Population per square mile	POPD	Increases (NOPB); Decreases (DISTPB)
Age (IV)	(1) Proportion (%) of population under age 18	(1) AGE18	Increases (NOPB); Decreases (DISTPB)
	(2) Proportion (%) of population over age 64	(2) AGE64	Increases (NOPB); Decreases (DISTPB)
Race/ethnicity (IV)	(1) Proportion (%) of Black population	(1) BLACK	Increases (NOPB); Decreases (DISTPB)
	(2) Proportion (%) of Asian population	(2) ASIAN	Increases (NOPB); Decreases (DISTPB)
	(3) Proportion (%) of Hispanic population	(3) HISPAN	Increases (NOPB); Decreases (DISTPB)
Housing value (IV)	Median housing value (\$)	MHV	Decreases (NOPB); Increases (DISTPB)
Income (IV)	Median household income (\$)	MHI	Decreases (NOPB); Increases (DISTPB)
Educational attainment (IV)	Proportion (%) of population with a four-year university degree or higher	EDU	Decreases (NOPB); Increases (DISTPB)
Language (IV)	Proportion (%) of population with non-English spoken at home	LAN	Increases (NOPB); Decreases (DISTPB)
Vehicle ownership (IV)	Proportion (%) of households without a vehicle	VEHIC	Increases (NOPB); Decreases (DISTPB)
Housing occupancy (IV)	Proportion (%) of occupied housing units	HO	Decreases (NOPB); increases (DISTPB)
Economic status (IV)	Proportion (%) of population below the poverty line	ECON	Increases (NOPB); Decreases (DISTPB)

Note: DV (dependent variable), IV (independent variable)

Exploring equity using geographically weighted regression

Table 2. Results of two OLS regression models

Variable	Model 1 (container)						Model 2 (minimum distance)					
	Unstandardized Coefficient		Standardized Coefficient	t	p	VIF	Unstandardized Coefficient		Standardized Coefficient	t	p	VIF
	$\beta$	SE	$\beta$				$\beta$	$\beta$	SE			
Intercept	45.683	25.692		1.77	0.07		3.792	2.39		1.59	0.11	
BLACK	0.190	0.062	0.145	3.06	< 0.01	4.16	0.011	0.006	0.099	1.83	0.06	4.16
ASIAN	0.951	0.435	0.092	2.18	0.02	3.33	0.054	0.041	0.064	1.32	0.18	3.33
HISPAN	0.087	0.213	0.016	0.41	0.68	2.75	0.01	0.020	0.003	0.07	0.94	2.75
POPD	-0.005	0.000	-0.270	-	< 0.01	1.50	0.0002	0.000	0.180	5.55	< 0.01	1.50
MHV	0.000054	0.000	0.091	1.89	0.06	4.28	-0.000005	0.000	-0.098	-1.79	0.07	4.28
AGE18	-0.258	0.320	-0.029	-	0.42	2.40	-0.002	0.030	-0.003	-0.07	0.93	2.40
AGE64	-0.544	0.299	-0.057	-	0.06	1.85	0.065	0.028	0.084	2.32	0.02	1.85
EDU	1.247	0.124	0.471	10.08	< 0.01	4.07	-0.054	0.012	-0.251	-4.70	< 0.01	4.07
LAN	0.038	0.135	0.009	0.28	0.77	2.04	-0.003	0.013	-0.010	-0.27	0.78	2.04
ECON	0.055	0.170	0.018	0.32	0.74	5.92	-0.008	0.016	-0.033	-0.51	0.60	5.92
HO	-0.085	0.248	-0.015	-	0.72	3.57	0.036	0.023	0.079	1.57	0.11	3.57
VEHIC	-0.435	0.186	-0.101	-	0.01	3.50	-0.023	0.017	-0.066	-1.32	0.18	3.50

N = 1,164  
 $R^2 = 0.386$ , Adjusted  $R^2 = 0.379$   
 $AIC_c = 11,839.75$   
 Joint F-statistic = 55.59 (p-value<0.01)  
 Joint Wald statistic = 1,008.19 (p-value<0.01)  
 Koenker (BP) statistic = 163.46 (p-value<0.01)

N = 1,164  
 $R^2 = 0.194$ , Adjusted  $R^2 = 0.185$   
 $AIC_c = 6,300.11$   
 Joint F-statistic = 45.17 (p-value<0.01)  
 Joint Wald Statistic = 365.42 (p-value<0.01)  
 Koenker (BP) statistic = 97.63 (p-value<0.01)

Note:  $\beta$  (Beta): regression coefficient; SE: standard error; t: t-value; p: p-value; VIF: variance inflation factor;  $AIC_c$ : corrected Akaike's information criterion

Exploring equity using geographically weighted regression

Table 3. Results of two GWR models

Variable	Model 1 (container)					Model 2 (minimum distance)				
	OLS Coefficient	GWR Coefficients			Range	OLS Coefficient	GWR Coefficients			Range
	$\beta$	Minimum	Mean	Maximum		$\beta$	Minimum	Mean	Maximum	
Intercept		-36.64	41.68	151.21	187.85		1.29	6.90	16.13	14.84
BLACK	0.145	-126.40	-1.98	67.72	194.12	0.099	-5.55	0.31	7.77	13.32
ASIAN	0.092	-21.79	-1.39	27.46	49.25	0.064	-2.81	0.09	4.71	7.52
HISPAN	0.016	-104.82	-2.30	205.51	310.33	0.003	-7.54	0.17	8.64	16.18
POPD	-0.270	-18.55	-1.36	26.81	63.91	0.180	-1.29	0.14	1.40	2.69
MHV	0.091	-21.24	0.90	29.69	50.93	-0.098	-4.10	-0.17	2.84	6.94
AGE18	-0.029	-15.71	-1.33	8.53	24.24	-0.003	-1.57	0.04	4.58	6.15
AGE64	-0.057	-11.18	0.07	12.14	23.32	0.084	-1.01	0.12	2.85	3.86
EDU	0.471	-8.09	4.87	58.92	67.01	-0.251	-3.25	-0.02	2.73	5.98
LAN	0.009	-21.43	0.93	19.28	40.71	-0.010	-1.66	-0.09	4.30	5.96
ECON	0.018	-20.37	1.07	47.97	68.34	-0.033	-2.51	0.02	4.15	6.66
HO	-0.015	-29.58	-0.57	13.80	43.38	0.079	-1.61	0.21	4.89	6.50
VEHIC	-0.101	-25.34	-1.12	19.55	44.89	-0.066	-1.85	0.05	2.20	4.05
Adjusted R <sup>2</sup>	0.379	0.02	0.69	0.92	0.90	0.185	0.27	0.70	0.92	0.65
Condition Index		9.7	14.6	24.8	15.1		8.6	16.3	24.4	15.8
N = 1,164						N = 1,164				
AIC <sub>c</sub> (OLS) = 11,839.75						AIC <sub>c</sub> (OLS) = 6,300.11				
AIC <sub>c</sub> (GWR) = 8679.89						AIC <sub>c</sub> (GWR) = 4,085.73				
Neighbors = 147						Neighbors = 147				

Note:  $\beta$  (Beta): standardized OLS coefficient; AIC<sub>c</sub>: corrected Akaike's information criterion

Table 4. Classification of census tracts by values of local coefficient and local  $R^2$ 

	Variable	Number of census tracts (N = 1,164)			
		LC > 0 (%)	LC < 0 (%)	LC > GC (%)	LC < GC (%)
Model 1	BLACK	523 (44.9%)	641(55.0%)	492 (42.2%)	672 (57.7%)
	ASIAN	678 (58.2%)	486 (41.7%)	411 (35.3%)	488 (41.9%)
	POPD	446 (38.3%)	718 (61.6%)	447 (38.4%)	717 (61.5%)
	EDU	749 (64.3%)	415 (35.6%)	598 (51.3%)	566 (46.6%)
	VEHIC	480 (41.2%)	684 (58.7%)	630 (54.1%)	534 (45.8%)
	$R^2$	Adjusted $R^2$ (OLS): 0.379 Adjusted $R^2$ (GWR): 0.690		GWR > OLS (%) 1,120 (96.2)	GWR < OLS (%) 44 (3.7)
Model 2	POPD	771 (66.2%)	393 (33.7%)	770 (66.1%)	394 (33.8%)
	AGE64	628 (53.9%)	536 (46.0%)	550 (47.2%)	614 (52.7%)
	EDU	536 (46.0%)	628 (53.9%)	566 (48.6%)	598 (51.3%)
	$R^2$	Adjusted $R^2$ (OLS): 0.185 Adjusted $R^2$ (GWR): 0.700		GWR > OLS (%) 1,164 (100)	GWR < OLS (%) 0 (0.0)

*Note:* LC: local coefficient by GWR; GC: global coefficient by OLS; LC > GC: census tract in which the value of the local coefficient is greater than the value of the global coefficient; LC < GC: census tract in which the value of the local coefficient is less than the value of the global coefficient

Table 5. Comparison of spatial autocorrelations of residuals between OLS and GWR

	Model 1		Model 2	
	OLS	GWR	OLS	GWR
Moran's I (residual)	0.36	0.10	0.61	0.15
z-score	63.87	18.5	105.83	26.34
p-value	< 0.01	< 0.01	< 0.01	< 0.01

Exploring equity using geographically weighted regression

Table 6. Neighborhoods with inequitable access to public beaches by census variable

Variable	Inequitable Neighborhood		
	City (County)	Township (County)	
Model 1	BLACK	Sterling Heights (M)	Shelby (M), Washington (M)
	ASIAN	Troy (O)	Canton (W), Plymouth (W)
	POPD	Livonia (W), Rochester (O), South Lyon (O), Troy (O)	Macomb (M), Ray (M), Shelby (M), Washington (M)
	EDU	Rochester (O), Rochester Hills (O)	Addison (O), Armada (M), Bruce (M), Oakland (O),
	VEHIC	Novi (O), Sterling Heights (M), Troy (O)	Brandon (O), Groveland (O), Independence (O), Plymouth (W),
Model 2	POPD	Rochester Hills (O), Troy (O)	Bloomfield (O), Shelby (M), Washington (M)
	AGE64	Detroit (W), Ferndale (O), Livonia (W), Warren (M)	Addison (O), Armada (M), Bruce (M), Oakland (O),
	EDU	Detroit (W), Eastpointe (M), Romulus (W), Sterling Heights (M), Warren (M)	Armada (M), Bruce (M), Ray (M), Richmond (M), Shelby (M), Wahsington (M)

*Note:* O: Oakland county; M: Macomb county; W: Wayne county