Impact of Ambient Built-Environment Attributes on Sustainable Travel Modes: A Spatial Analysis in Chittenden County, Vermont

By

George X. Lu*, PhD
Research Analyst
Transportation Research Center
University of Vermont
Farrell Hall, 210 Colchester Road
Burlington, VT 05401
Tele: (802) 656-9035
Fax: (802) 656-9892
E-Mail: xlu@uvm.edu
* Corresponding Author

James Sullivan, P.E.
Research Analyst
Transportation Research Center
University of Vermont
Farrell Hall, 210 Colchester Road
Burlington, VT 05401
Tele: (802) 656- 9679
Fax: (802) 656-9892
E-Mail: jlsulliv@uvm.edu

Austin Troy, PhD
Associate Professor
Director, Transportation Research Center
University of Vermont
Farrell Hall 110, 210 Colchester Road
Burlington, VT 05401
Tele: (802)-656-8336
Fax: (802) 656-9892
E-Mail: atroy@uvm.edu

Prepared for the 91st Annual Meeting of the Transportation Research Board
Washington, D.C., January 22-26, 2012

Length of Paper (TRB12-2561):
5192 words, 8 tables & figures @ 250 words each
7192 equivalent words

ABSTRACT:

Non-motorized travel in terms of walking or bicycling plays a critical role in promoting healthy living style, offering sustainable alternatives to environmental impacts, energy consumption, and societal costs of motorized travel modes, and allocating limited resources for constructing pedestrian-orientated transportation infrastructure facilities. Historically, much research has been focused on the nexus between ambient built-environment attributes and travel mode choices or shares or non-motorized travel prediction. For a better understanding of non-motorized travels at multimodal facilities, spatial dependency should be considered since traffic volume at one monitoring station is correlated with that at neighboring sites due to the continuity in area-wide traffic circulation. However, few studies have been conducted in spatial analysis of walking and bicycling traffic at intersections.

Utilizing a 10-year assembly of non-motorized traffic counts and a geographical information system (GIS) which contains intersection-based location data and functional classifications of habitable infrastructure in Chittenden County of Vermont, this study determined whether spatial autocorrelation exists for non-motorized volumes and ambient built-environment attributes, and geographically weighted regression (GWR) was applied on two different data-collection scales to identify whether spatially varying relationships operate significantly between non-motorized volumes and specific surrounding characteristics on each scale. Some variables are found significant in spatially influencing non-automobile travel. The resultant models can estimate walking and bicycling volumes at countywide intersections. Better estimation of non-motorized travel locally facilitates transportation planning, facility design, safety enhancement, and operational analysis.

Key Words: Non-motorized travel, built environment, GIS, spatial analysis

INTRODUCTION

A critical objective in transportation engineering studies advocating "New Urbanism and Smart Growth" is to enhance the level of non-motorized-mode travels dominantly characterized by walking and bicycling activities (*1*). There is a contention from new urbanists and others that to create more compact, diverse, and pedestrian-orientated neighborhoods can meaningfully shape how Americans travel in everyday life. Not only do non-motorized travel modes play a key role in promoting healthy living style but also they provide sustainable alternatives to environmental impacts, energy consumption, and societal costs of motorized travel modes. Additionally, non-motorized travel reinforces the ridership level of public transit systems, since most public transit travels necessitate a walking or bicycling short trip at both terminals of a route. Furthermore, the allocation of limited resources for building infrastructural facilities amicable to non-motorized travel modes is essentially associated with the accurate measurement of pedestrian and bicyclist activities at the location or area under construction.

In this study, temporal and spatial features of walking and bicycling travels were examined via utilizing the data widely collected in Chittenden County of Vermont State. The overall objective was to develop an intersection-location-based model which quantitatively characterizes the spatially-dependent relationship between non-motorized traffic volumes and a medley of factors pertinent to ambient built-environment elements. The specific intention was to better understand what surrounding attributes are significant contributors to non-motorized travels with the presence and influence of spatial dependency. Given this deepened understanding, it may be possible to fill the voids in traditionally sparse collections of non-motorized travel counts and to determine the true fraction of our travel activities which occur locally on non-motorized travel modes.

HISTORICAL RESEARCH

Historically, much research explored the plausible nexus between ambient built-environment attributes and travel mode choices or shares. Of them, Rodriguez and Joo (*2*) examined the connection between non-motorized mode choices and miscellaneous built-environment variables, considering typical modal characteristics. Particularly local topography and sidewalk availability are found significantly important to the attractiveness of non-motorized modes. Guo et al. (*3*) assessed the built-environment-related effects on motorized and non-motorized traffic. Few factors were identified as key contributors to a transition from motorized to non-motorized modes. Arguably, the increases in bikeway density or roadway-network connectivity have the best potential for supplementing existing motorized travel with non-motorized alternative.

Some studies investigated how various ambient features influence non-motorized travels in the United States. Previous studies indicated that built-environment factors are significant predictors of non-motorized travels. For example, Cervero and Kockelman (*4*) found that statistically density, land-use diversity, and pedestrian-oriented design significantly encourage non-motorized travels, although the impacts appear fairly marginal. Frank and Engelke (*5*) showed that gridded street networks can promote bicycling and walking activities by reducing trip distances, offering alternative pathways, and slowing motorized travel. Cervero and Duncan (*6*) found, although personal and household factors were more significant, land use and street connectivity in San Francisco had moderate effects on promoting non-motorized short trips. Frank et al. (*7*) found that measures of land-use mix, residential density, and intersection density in Atlanta were positively associated with daily minutes of moderate physical activity. Aytur et al. (*8*) made discoveries in North Carolina that communities designed for "active transportation" had the strongest influence on non-motorized travels among lower-income individuals. Cervero et al. (*9*) found that street density, connectivity, and proximity to cycling lanes are essential to physical activity while land-use mixtures not. Past studies also addressed the issue of predicting non-motorized volumes using built-environment-based variables. Liu and Griswold (*10*) demonstrated that pedestrian volumes can be reasonably estimated by environmental factors given appropriate measurement and geographical scale.

Aultman-Hall et al. (*11*) investigated pedestrian counts in Vermont to identify the influential factors for volume variability, and their predictive models indicate bad weathers diminish aggregate walking levels by a moderate amount. Some studies specifically examined non-motorized crossing volumes at intersections. Of them, Pulugurtha and Repaka (*12*) developed predictive models for the pedestrian activities at signalized intersections, finding urban residential density is the most significant contributor to pedestrian activity level. Using pedestrian crossing volumes at intersections, Schneider et al. (*13*) created a pilot model which shows number of jobs and commercial retail properties, total population, and presence of a regional transit station close to an intersection are significant factors.

For a better prediction of motorized and non-motorized travels at various multimodal facilities, spatial dependency should be considered because of the theory that traffic volume at one monitoring station is correlated with the volumes at neighboring stations due to the continuity in area-wide traffic circulation. A few studies have acknowledged the spatial dependency between travels of different modes and built-environment-related factors. Of them, Eom et al. (*14*) researched annual average daily traffic (AADT) using Kriging estimation. The predictive capability of this spatial model outperforms that of ordinary least-square (OLS) model. Zhao and Park (*15*) analyzed AADT in grid-like networks utilizing geographically weighted regression (GWR) which compensates for spatial dependency by estimating local model parameters. They found GWR models were more accurate than OLS models and useful for studying the effects of the variables at different locations. Even fewer studies have conducted the geospatial analyses of walking and bicycling travels which treat spatial dependency. Zahran et al. (*16*) studied nationwide non-motorized commute trips to conclude that the spatial distribution of travel volumes is positively associated with population density, natural amenities, education, wealth and estimates of local civic concerns. Pucher et al. (*17*) reviewed trends in bicycling levels, safety, and policies in North America, using national aggregate and city-specific data. They found there is much spatial variation and socioeconomic inequality in bicycling travel rates.

The issue, "How do spatial patterns in non-motorized travel level change with distinct ambient built-environment attributes", remains inadequately researched in Vermont and a spatial dataset was not extensively collected before. This dearth of effective analyses motivated this geospatial study which is to determine (a) whether a spatial autocorrelation exists for the non-motorized traffic volumes and ambient built-environment attributes (b) whether spatially varying relationships significantly operate between non-motorized volumes and some specific built-environment characteristics.

DATA PREPARATION

This study is performed using two comprehensive and robust datasets. The first is an assembly of non-motorized traffic counts at many intersections located in Chittenden County, Vermont. These counts have been systematically gathered during 10 years by the Chittenden County Metropolitan Transportation Organization (CCMPO). The second is a geographical information system (GIS) which contains the point location data and the functional classification of each habitable infrastructure in Chittenden County. The building locations come from the Vermont E911 database which houses all residence locations in terms of single-family, multi-family, seasonal, and mobile homes and all non-residence locations in terms of commercial, industrial, educational, governmental, health-care and public gathering. Vermont is unique in that the E911 database is publicly available to support emergency-response personnel statewide via the Vermont Center for Geographic Information (VCGI) which provides the GIS data source, information, and events.

Data Sources

This study analyzed the counts of pedestrians and bicyclists which were manually collected using traffic-count boards at countywide intersections from 2000 to 2009 (Figure 1(a)). Geo-coded land-use data and a street network were offered by the VCGI, and the ArcGIS program quantified the number of buildings and total roadway length within two circular areas concentrically surrounding each intersection. Chittenden County is the most densely inhabited county out of 14 counties in Vermont, with a population of approximately 150,000 and a geographical coverage of more than 620 square miles. This county owns sustainable urban transportation systems which are composed of an extensive network of bike lanes and set-aside street space for pedestrians and bicyclists. The collection of traffic volumes was oriented toward counts from each inbound approach of 428 sites. Walking and bicycling volumes observed at an intersection was recorded only on its original approach – the outbound approach was intentionally neglected. At some sites, the counts spanned multiple years in the 10-year period. Aggregated by hourly-total count, the initial dataset consisted of 3,541 records, or on average 8 hours per location, totally 5 outliers with hourly volumes exceeding 500 non-motorized travelers was eliminated, reducing the size to 3,536 records. Almost all of count locations encompassed the 2-hour PM-peak period (4:00-6:00PM), and many of them included up to a maximum of 12 hours of data collection.

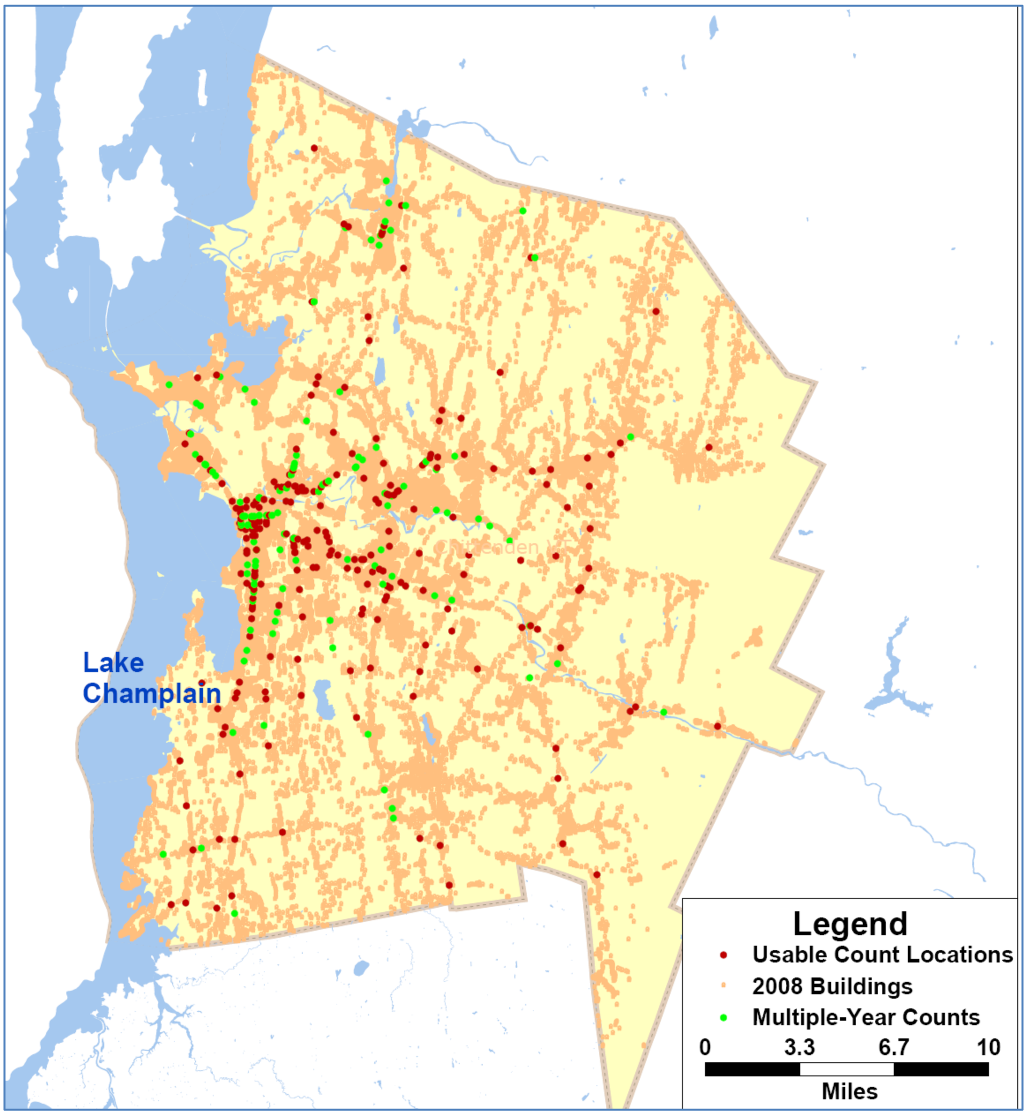


Figure 1 115 count locations of multiple-year observations.

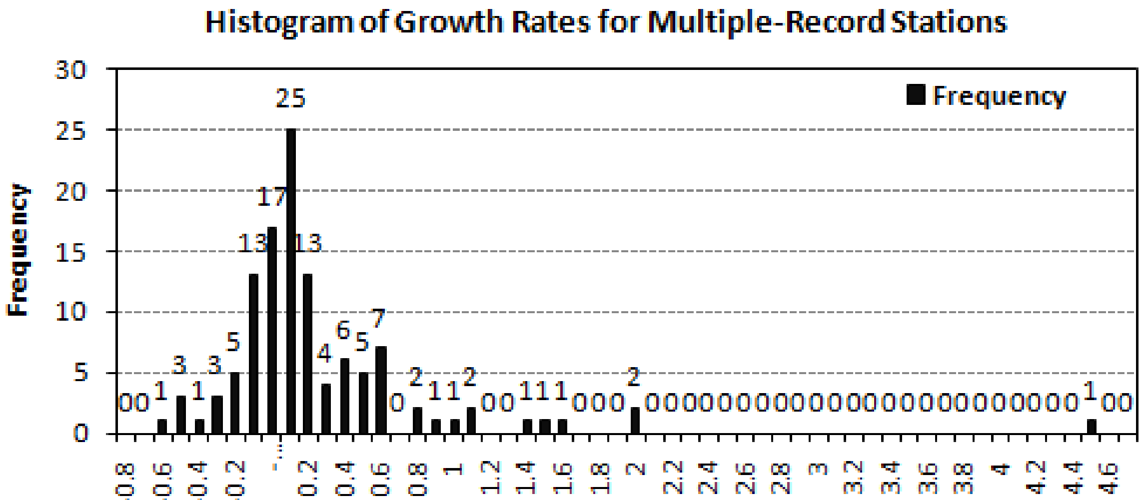
Initial Reductions

It was necessary to reduce 3,536 records to those each of which is singularly mapped to one intersection positioned in a spatial coordinate system, since the statistical processing of spatial dependency required there was one record per observation site. To eliminate the need for weekly correction, only these observations from Tuesdays, Wednesdays, and Thursdays were considered. Due to the common occurrence of holidays and potentially the behavioral influence latently residual from personal activities in weekends, Mondays were omitted for data-noise reduction. Then, only summer-related observations during June, July, and August were used for eradicating seasonal correction needs. Additionally, only PM-peak 2-hour volumes were considered to exclude the daily correction. Once these filtering steps had been executed, it was assumed no daily, weekly, and seasonal corrections would be entailed for 964 records at 346 intersections. Totally 115 intersections had records which span multiple years, representing repeated PM-peak counts in separate years.

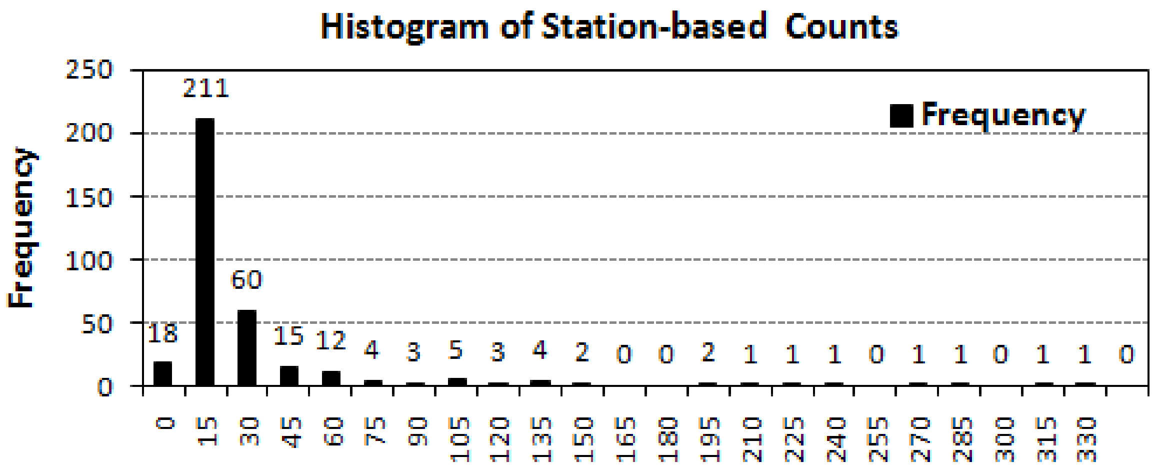
Temporal Corrections

It was initially assumed that corrections should be made to compensate for the temporal variation in the age of the reduced data in the event that annual growth or decline in non-motorized traffic demand had occurred. To account for the data age, the OLS regression was applied to test if statistical trends existed for: (i) All data from 2000 to 2009; (ii) Data grouped by those count locations where more than one count was implemented in different years but fixed for same days and hours. OLS regression results for (i) indicated a very low R-square (0.004). This meant that there is not a general linear trend in non-motorized volumes in 10 years. The individual locations were expected to have experienced temporal growth in non-motorized traffic demand. For these 115 intersections (Figure 1(b)) where more than one count was conducted in separate years, simple linear regression was employed to roughly estimate general intersection-specific growth rates. Of 115 growth rates obtained, 43 (37.4%) are negative and 72 (62.6%) are positive. The minimum and maximum are -0.7% and 4.5%. Mean and median constitute useful measures for the central tendency in observations (*18*), which are 0.2% and 0.1% respectively. In order that the measurement precision does not preclude the determination of a mode, all growth rates were assigned into a series of exclusive bins which were defined in the incremental magnitude unit of 0.1%. The "0.0%-0.1%" bin contains 25 out of 115 rates, accounting for the largest portion (Figure 2(a)).

The descriptive analysis led to following summaries: (a) For all non-motorized volumes, there is not a statistically significant 10-year trend; (b) Of 115 growth rates, a considerable proportion (37.4%) is negative; (c) The median and the bin-based mode are close to zero. Accordingly, it is believed that no significant temporal effects may exist for different years at all intersections, and the data could be used confidently without temporal corrections for 10 years. This finding is consistent with motorized traffic volumes which are generally in monotonously positive relation with counts of pedestrians and bicyclists: no significant local changes during the same period (*19*). The multiple-count records for each intersection were aggregated into a single-intersection-based volume by averaging the count records and rounding them to an integer. The results are descriptively characterized by: 346 observations; mean/median, 23/8 travelers; standard error, 2 travelers; minimum/maximum travelers, 0/316 travelers. The histogram for these station-based counts resembles, to a large extent, typical discrete distributions such as Poisson or Negative Binomial distributions (Figure 2(b)).



(a)



(b)

Figure 2 Two histograms: (*a*) non-motorized traffic growth rates for count stations, and (*b*) station-based non-motorized traffic volumes.

Built-Environment Attributes

Some factors regarding ambient land-use and infrastructural structures were selected as independent variables (also termed "predictors"), which are commonly associated with non-motorized travels in the literature (*4,20*). These factors were identified within two buffer areas. One buffer area refers to the circular scope which has an intersection as the center and 1,000 feet as the radius, and the other has 2,500 feet as the radius. These buffer areas correspond roughly to the mid-point and the edge of the 0.5-mile distance typically representative of walking trips. These factors are related to manifold facets of the built environments in intersection-centered buffer areas. The descriptive statistics for each of these predictors were compiled in Table 1. For illustrative brevity, the variables corresponding to the 2,500-ft buffer area are exhibited in parentheses.

Table 1 Non-motorized Traffic Volume and Predictor Variables for Analysis

Independent Variables	Type	Descriptive statistics in 1,000-ft (2,500-ft) buffer area				
		$\underline{x}$	$\overline{x}$	R	$\hat{\mu}$	S.D.
Count of All Buildings	I	0.0 (4.0)	525.0 (2345.0)	525.0 (2341.0)	93.7 (502.7)	109.9 (535.2)
Count of Commercial Buildings	I	0.0 (0.0)	228.0 (464.0)	228.0 (464.0)	15.5 (63.7)	31.9 (102.6)
Count of Educational Buildings	I	0.0 (0.0)	31.0 (98.0)	31.0 (98.0)	1.0 (5.0)	3.5 (13.0)
Count of Public Buildings	I	0.0 (0.0)	15.0 (54.0)	15.0 (54.0)	2.0 (8.3)	2.9 (9.4)
Count of Residential Buildings	I	0.0 (3.0)	499.0 (2124.0)	499.0 (2121.0)	75.0 (425.1)	96.6 (456.3)
Count of All Intersections	I	1.0 (1.0)	30.0 (142.0)	29.0 (141.0)	9.1 (41.6)	7.0 (33.5)
Total Roadway Length	C	912.2 (2292.5)	6345.2 (31149.0)	5433.0 (28856.5)	2880.5 (13054.3)	1324.5 (6818.7)
NCD [a]	C	0.6 (0.5)	11.1 (10.8)	10.5 (10.3)	5.3 (4.9)	2.4 (2.0)
Distance to Burlington Centroid	C	181.7	30341.5	30159.8	8782.1	6759.0

NOTE: C – continuous variable; I – integer variable; $\underline{x}$ – Minimum; $\overline{x}$ – Maximum; R – Range; $\hat{\mu}$ – Mean; S.D.– Standard Deviation; [a] Neighborhood Connectivity Density: Number of intersections divided by total road length (*20*)

STUDY METHODOLOGY

The entities at nearby locations often share more similarities than the entities far apart. Usually, this notion is termed "Tobler's first law of geography": "everything is related to everything else, but near things are more related than distant things" (*21*). Spatial dependency produces spatial autocorrelation (SA) in statistics when it conflicts with the assumption of independent observations required for most standard statistical techniques. Hence, regression analyses without compensating for spatial dependency can yield "spatial heterogeneity": overall parameters estimated for the entire system inadequately describe the process at any given location and the estimated degree of autocorrelation varies significantly across geographic domain. This study was intended to: (a) Identify SA; (b) Use a spatial regression model to capture the spatially-dependent relationship between non-motorized volumes and the predictors.

Spatial Autocorrelation

Spatial dependency exists among multimodal traffic volumes at neighboring intersections. Firstly, the neighboring intersections share a large portion of through volumes. Secondly, traffic signal coordination promotes the clustering of traversing travelers. Moreover, neighboring intersections are surrounded by similar land-use characteristics and infrastructural elements. For instance, two or more intersections can lie in the walking distance of one building nearby. With spatial coordinates known, the first step is to utilize classic statistics to measure the degree of spatial dependency in dataset. Moran's I and Geary's c indices estimate globally the overall degree of SA (*22*, *23*). Positive SA happens when similar values exist nearby, while negative SA happens for dissimilar values. Geary's c expectation equals 1 in the absence of SA, regardless of the specified weight matrix (*24*). Geary's c ranges from 0 to 2 (1 means no SA), whereas values at the range edges indicate a positive (0 to 1) or negative (1 to 2) SA. Moran's I is consistently more powerful than Geary's c (*25*). A natural logarithm transformation was applied to the integer volumes to facilitate closeness to normality.

Spatial Regression

Specific statistical techniques determine how spatial dependency enters spatial regression models: (i) in the error terms; (ii) as the relationship between the dependent variable and a spatial lag of itself; or (iii) as the relationship of the dependent variable to the predictors. Generally, a model with autocorrelated error terms produces better estimators and predictions than an OLS model but may be outperformed by universal Kriging for the purpose of producing optimal predictions (*26*). None of these models, however, address the issue of spatial heterogeneity. Although they are appropriate for describing a process with a non-constant mean, the nature of relationships itself is assumed to be ubiquitously homogeneous, removing the possibility that the process operates differently in varied locations (*26*).

Geographically Weighted Regression

GWR models spatially heterogeneous processes, with the underlying philosophy that parameters may be estimated anywhere in study area given a dependent variable and a set of predictors measured at locations where spatial coordinates are known (*27,28*). The GWR weighting scheme assumes that using geographically approximate observations is essential to estimating local parameters. Traditionally, a linear regression model may be written as:

$$y_i = \mathrm{X}_i^T \mathbf{B} + \varepsilon_i \tag{1}$$

Where: $\mathrm{X}_i = \{1, x_{i1}, \ldots, x_{ik}, \ldots, x_{iK}\}^T$ – The (*K+1*)-dimensional vector of the i th independent observation ($i = 1, \ldots, N$; $k = 1, \ldots, K$);

$\mathrm{B} = \{\beta_0, \beta_1, \ldots, \beta_k, \ldots, \beta_K\}^T$ – The (*K+1*)-dimensional parameter vector for intercept and predictors;

x_{ik} – The k th predictor for the i th observation;

N – Total number of independent observations;

K – Total number of predictors;

ε_i – Random error term.

One global parameter is estimated for the relationship between each predictor and the dependent variable, in forms of:

$$\hat{\mathrm{B}} = (\mathbf{X}^\mathbf{T}\mathbf{X})^{\mathbf{-1}}\mathbf{X}^\mathbf{T}\mathbf{y} \tag{2}$$

Where $\hat{\mathrm{B}}$ represents the vector of global parameters estimated, **X** is a matrix of intercept and predictors, and **y** represents a vector of observations. $\hat{\mathrm{B}}$ is constant regardless of the spatial locations of *N* observations. GWR extends the framework as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{K} \{\beta_k(u_i, v_i) x_{ik}\} + \varepsilon_i \tag{3}$$

Where (u_i, v_i) denotes the spatial coordinates of point i and $\beta_k(u_i, v_i)$ is a realization of the continuous function $B(\boldsymbol{u}, \boldsymbol{v})$ at point i. Hence, there is a continuous surface of parameter values and measurements of this surface are taken at certain points to denote the spatial variability (*27*, *28*). Algebraically, the GWR estimates are:

$$\hat{\mathrm{B}}(\boldsymbol{u}, \boldsymbol{v}) = (\mathbf{X}^\mathbf{T}\mathbf{W}(u_i, v_i)\mathbf{X})^{\mathbf{-1}}\mathbf{X}^\mathbf{T}\mathbf{W}(u_i, v_i)\mathbf{y} \tag{4}$$

Where $\mathbf{W}(u_i, v_i)$ is a $n \times n$ matrix whose off-diagonal and diagonal elements denote zero and the geographical weighting of observed data for point i.

Importantly, not only does GWR deal with spatial dependency by embodying geographical location in intercept terms but also it addresses spatial heterogeneity by incorporating spatial coordinates in parameter estimates (*28*). There is evidence that GWR model can reduce the *residuals* more substantially than these with an autoregressive term, because of the way in which the spatially-dependent relationship is modeled through geographically varying parameter estimates rather than through the error term (*28*).

Multicollinearity Test

The 1,000-ft buffer area is enclosed into the 2,500-ft one, so each variable of the same category (e.g., number of educational buildings) in the former is numerically a subset of that in the latter. Statistically, this situation perfectly induces the multicollinearity issue harmful to the validity in statistical inference**.** Pearson correlation coefficient measures the direction and degree of linearity relationship between two variables, useful to test general presence of multicollinearity relationship. A positive or negative one means respectively a perfectly linear relationship in increasing or decreasing fashion. A zero denotes the absence of the linear relationship. Generally, a strong correlation is indicated by a coefficient whose absolute value exceeds 0.80; statistically a p-value (<0.05) means a significant linear relationship.

Model Parsimony

The modeling process adhered to the philosophy of "model parsimony", which means, other things being equal and given any two models with equal log likelihood values, the model with fewer parameters is better. The Akaike Information Criterion (AIC) is useful for evaluating models (*29*). When the AIC values for two models differ by more than 3, the models are considered significantly different (*28*). A model with smaller AIC is considered closer to the unknown true model (*30*).

STUDY RESULTS

For statistical hypothesis testing, Moran's I values can be transformed to Z-scores whose absolute value greater than 1.96 indicates significant SA at the 5% level (*22*). As Table 2 shows, all Moran's I indices observed exceed the expectation (-0.003) and the Z-scores larger than 1.96, which means a significant positive SA. The observed Geary's c indices also indicate positive SA (*31*).

Table 2 Moran's I and Geary's c Calculation Results

	Observed		Expected		Z		Pr > |Z|	
Dependent and predictors	I	c	I	c	I	c	I	c
Dependent Variable [1]	0.23	0.88	-0.003	1.00	76.19	-4.49	<.0001	<.0001
Distance to Burlington downtown centroid	0.19	0.69	-0.003	1.00	63.10	-11.6	<.0001	<.0001
Predictors within the 1,000-ft buffer area								
Number of all buildings	0.26	1.01	-0.003	1.00	86.71	0.37	<.0001	0.710
Number of all intersections	0.20	0.93	-0.003	1.00	66.15	-2.55	<.0001	<.011
Number of commercial buildings	0.16	1.29	-0.003	1.00	52.20	10.70	<.0001	<.0001
Number of educational buildings	0.08	1.38	-0.003	1.00	25.70	14.00	<.0001	<.0001
Number of public buildings	0.09	1.07	-0.003	1.00	28.65	2.67	<.0001	0.008
Number of residential buildings	0.21	1.06	-0.003	1.00	67.94	2.23	<.0001	0.026
Total roadway length	0.18	0.90	-0.003	1.00	59.78	-3.56	<.0001	0.0004
NCD	0.11	0.90	-0.003	1.00	35.66	-3.91	<.0001	<.0001
Predictors within the 2,500-ft buffer area								
Number of all buildings	0.34	0.92	-0.003	1.00	110.29	-3.17	<.0001	0.002
Number of all intersections	0.30	0.85	-0.003	1.00	99.63	-5.67	<.0001	<.0001
Number of commercial buildings	0.27	1.12	-0.003	1.00	87.65	4.25	<.0001	<.0001
Number of educational buildings	0.18	1.28	-0.003	1.00	59.80	10.30	<.0001	<.0001
Number of public buildings	0.23	0.95	-0.003	1.00	75.89	-1.85	<.0001	0.065
Number of residential buildings	0.31	0.92	-0.003	1.00	100.53	-2.88	<.0001	0.004
Total roadway length	0.30	0.81	-0.003	1.00	98.68	-7.13	<.0001	<.0001
NCD	0.20	0.77	-0.003	1.00	66.98	-8.55	<.0001	<.0001

NOTE: I – Moran's I index; c – Geary's c index; [1] – After Logarithm transformation for normality approximation

The Pearson test treats two *interval* variables well-approximated by a normal distribution. Therefore, distance to downtown centroid, total roadway length, and NCD were excluded. The test for the variables of same type demonstrates that there are very strong (> 0.80) correlations for almost all variables from 1,000-ft and 2,500-ft buffer areas, due to the presence of spatial containment. This result indicates that the spatial regression should avoid including simultaneously the variables from both buffer areas. In each buffer area, the test exhibits that most of these variables are significantly correlated one another. Expectedly, number of all buildings has strong correlation with both numbers of commercial and residential buildings. Due to the recurring low correlations associated with the educational buildings in the 1,000-ft buffer area, it is believed that number of educational buildings is a variable which should be treated separately. Based on these bivariate analysis results, GWR procedures were separately applied to two buffer areas which represent changed data-collection scales. The multicollinearity on each scale was addressed.

GWR for 1,000-ft Buffer Area

Table 3 (Part I) displays the GWR results from the 1,000-ft buffer area. All *p*-values from Monte Carlo heterogeneity tests, when less than 0.05, mean the parameters of correspondent predictors vary significantly from place to place. Bi-square and Gaussian models were experimented and the former was found to have the better results. Model A-1 involves all predictors whose parameters vary insignificantly in spatial sense due to their *p*-values above 0.05. The foregoing Pearson test revealed that number of total buildings is strongly correlated with number of intersections and each of the land-use predictors except for number of educational buildings, so Model A-2 discarded three land-use predictors and that infrastructure-based predictor. Model A-2 reveals that the parameters for number of all buildings and number of educational buildings vary significantly at different intersections, and total roadway length is found to have an unacceptably high *p*-value. After disregarding total roadway length, Model A-3 shows number of all buildings and number of educational buildings have significantly changing parameters with locations varied, and model parsimony has been increased by lowering AIC from 2491.53 to 2481.42. Model A-4 omits NCD and makes an AIC reduction from 2481.42 to 2468.66. However, Models A-5 and A-6 reverse the trend if either of significant land-use predictors in A-4 is included, increasing AIC (from 2468.66 to 2469.59 and 2504.65 respectively) and decreasing the number of significant predictors from three to two. Model A-7 re-includes both significant land-use predictors but omits distance to downtown centroid, yielding the lowest AIC and significant p-values for both predictors. Model A-7 has a lower AIC (2463.57 vs. 2468.66) but a slightly lower R-square (0.629 vs. 0.636), compared with Model A-4. Given similar model parsimony and validity, Model A-4 importantly unveils more underlying information in shaping spatial relationship. Therefore, Model A-4 is used as the final model for the 1,000-ft buffer area.

Table 3 GWR Analysis (Part I): Modeling Results for the 1,000-ft Buffer Area

	p-values [a] for significance test						
Predictors	A-1 (*K*=9)	A-2 (*K*=5)	A-3 (*K*=4)	**A-4 (*K*=3)**	A-5 (*K*=2)	A-6 (*K*=2)	A-7 (*K*=2)
Land-use							
Number of all buildings	0.29	0.00[c]	0.00[c]	**0.00[c]**	0.00[c]		0.00[c]
Number of commercial buildings	0.40						
Number of educational buildings	0.95	0.00[c]	0.00[c]	**0.00[c]**		0.01	0.00[c]
Number of public buildings	0.53						
Number of residential buildings	0.33						
Infrastructure-network							
Number of intersections	0.98						
Total roadway length	0.62	0.75					
Distance to downtown centroid	0.13	0.02	0.01	**0.02**	0.02	0.01	
NCD	0.94	0.69	0.75				
Model-specific attributes							
Intercept	0.11	0.08	0.00[c]	**0.00[c]**	0.00[c]	0.00[c]	0.00[c]
AIC	2485.50	2491.53	2481.42	**2468.66**	2469.59	2504.65	2463.57
R-square [b]	0.561	0.632	0.625	**0.636**	0.623	0.597	0.629
Adjusted R-square	0.542	0.581	0.582	**0.595**	0.587	0.551	0.594

Table 3 GWR Analysis (Part II): Modeling Results for the 2,500-ft Buffer Area

Predictors	p-values[a] for significance test B-1 (*K*=9)	B-2 (*K*=5)	B-3 (*K*=4)	B-4 (*K*=3)	**B-5 (*K*=2)**	B-6 (*K*=2)	B-7 (*K*=2)
Land-use							
Number of all buildings	0.36	0.03	0.00[c]	0.00[c]	**0.00[c]**	0.00[c]	
Number of commercial buildings	0.35						
Number of educational buildings	0.36						
Number of public buildings	0.34						
Number of residential buildings	0.36						
Infrastructure-network							
Number of intersections	0.16	0.04	0.05	0.05	**0.02**		0.00[c]
Total roadway length	0.28	0.32					
Distance to downtown centroid	0.08	0.26	0.18				
NCD	0.24	0.02	0.02	0.23		0.11	0.36
Model-specific attributes							
Intercept	0.00[c]	0.00[c]	0.00[c]	0.00[c]	**0.00[c]**	0.05	0.01
AIC	2463.05	2454.98	2448.92	2452.11	**2441.70**	2449.69	2463.23
R-square[b]	0.625	0.629	0.617	0.649	**0.643**	0.644	0.639
Adjusted R-square	0.592	0.598	0.595	0.612	**0.613**	0.610	0.600

NOTE: [a] Tests via Monte Carlo significance test procedure and values less than 0.05 are significant; [b] Coefficient of Determination; [c] Very small values less than 0.000001

GWR for 2,500-ft Buffer Area

Table 3 (Part II) demonstrates the GWR results from the 2,500-ft buffer area. Model B-1 includes all predictors, and Model B-2 yields three predictors with significant p-values, including number of all buildings, number of intersections, and NCD. Total roadway length and distance to downtown centroid have insignificant parametric heterogeneity. The absence of total roadway length from Model B-3 makes no difference in uncovering significance, and distance to downtown centroid is still insignificant in local parameter variation. After neglecting distance to downtown centroid, Model B-4 reveals that number of all buildings and number of intersections are significant in local parameter variation. Continually, Model B-5 reduces AIC from 2452.11 in Model B-4 to 2441.70 and also makes the intercept term spatially significant. Although each of Model B-6 and Model B-7 reveals a significant local parameter variation, they increase AIC and p-values for intercept terms. The adjusted R-square values are also reduced. Therefore, Model B-5 is treated as the final model due to its high R-Square (0.643), the highest adjusted R-square (0.613), and the lowest AIC (2441.70).

Table 4 Global Regression Parameters without Geographical Weighting

Global Parameter	AIC	R^2	Adj. R^2	β	Std Err	T
Model A-4 Related						
Intercept	2504.24	0.506	0.500	19.677	1.159	16.984
Distance to downtown centroid				-0.0004	8.23E-05	-5.000
Number of all buildings in the 1,000-ft buffer area				0.062	0.005	12.450
Number of educational buildings in the 1,000-ft buffer area				0.7281	0.139	5.250
Model B-5 Related						
Intercept	2461.13	0.561	0.558	12.950	0.742	17.453
Number of intersections				0.088	0.030	2.903
Number of all buildings in the 2,500-ft buffer area				0.013	0.002	6.634

Two global models, ignoring geographical weighting for spatial heterogeneity, were fitted for comparison with the two final GWR models, as Table 4 shows. Comparison between the fitted parameters of each set of models (local (GWR) and global) reveals the extent to which the use of GWR benefited the fit of the model. However, the global model parameters continue to help us better understand the nature of the relationships between non-motorized travel and the built environment. Model A-4 improves R-squared (Adjusted R-squared) from 0.506 to 0.636 (0.500 to 0.595), although an increase is to be expected given the difference in degrees of freedom. However, AIC reduction from the global model to the local model (2504.24 to 2468.66) suggests that the local model is truly a better fit to the data even accounting for differences in degrees of freedom. For Model B-5, the GWR model enhances R-squared (Adjusted R-squared) from 0.561 to 0.643 (0.558 to 0.613) and brings AIC down from 2,461.13 to 2,441.70. The global results suggest that non-motorized traffic volume is positively related to number of nearby buildings (especially educational ones within 1,000-ft scope) and negatively associated, but to a much lesser degree, with distance from downtown. However, the relationship to nearby buildings is much stronger in the 1,000-ft buffer area. The global results also suggest that non-motorized traffic volume is more strongly related to number of intersections for the 2,500-ft buffer area.

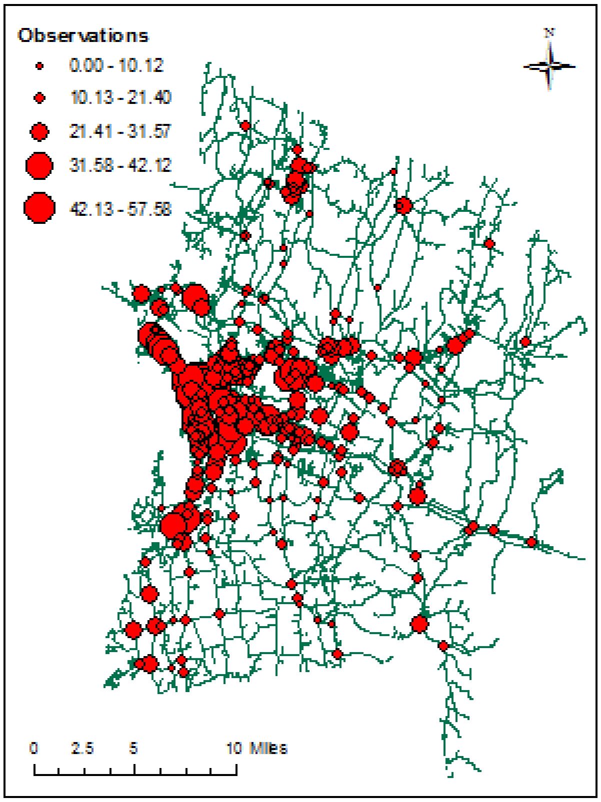
GWR Mappings

The main GWR outputs consist of a set of localized estimates and associated diagnostics. Different from the singular global values traditionally estimated, these local values lend themselves to being mapped for spatial patterns. With large datasets, certain forms of visualization, such as mapping, is the only effective way to make sense of the large volume of outputs. Although the displays of the local parameter estimates are of primary interest, it is instructive and informative to plot local R-square statistics or local standard deviations.

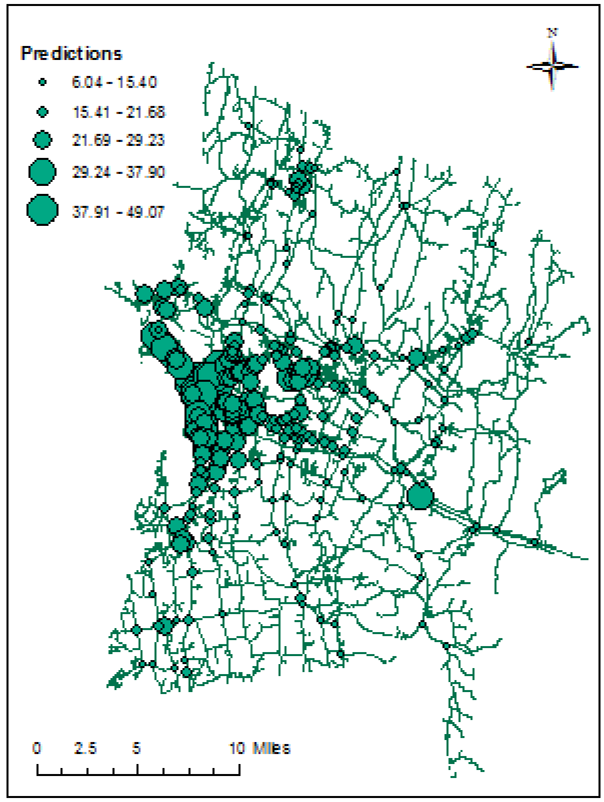
Figure 3(a), (b) presents a display of graduated circles regarding the dependent variable observed and predicted by Model A-4. The resultant pattern suggests that in this case there is a very close spatial pattern between the observed dependent and the predicted one. The values around Burlington area are larger and denser than those in other areas around. Figure 3(c-h) exhibits the mappings of the estimates for intercept term, three variables with significant p-values, and two local diagnostics on 1000-ft scale. Figure 3(c) shows, there is a broad regional pattern with higher values of the intercept term in Burlington area and lower values elsewhere. This means the count stations in Burlington area generate more non-motorized volumes than outer areas, verifying GWR models spatial dependency through heterogeneous intercept parameters at different locations.

Figure 3(d-f) shows, three predictors share one common characteristic in spatial parametric variation: the majority or a huge amount of smaller values are scattered in proximity of Burlington area, and these larger values are spread over outwards from Burlington area. This could be interpreted by the limited marginal effect of a unit increase in the three variables on the generation of non-motorized travel demand. Take number of all buildings as an example: since there are already a substantial number of all buildings which generate a large amount of non-motorized traffic, it is highly likely that an additional new building brings forth limited increase in non-motorized demand. Comparatively, when the buildings are sparsely located at somewhere with non-motorized travel demand low, the introduction of a new building will have more drastic influence upon additional non-motorized travels. Note that for these three variables the patterns of their larger estimates are somewhat different one another. This difference between number of all buildings and number of educational buildings could be attributed to the spatial placement of the latter which is definitely disparate from that of the former due to the consideration for educational coverage of local populations.

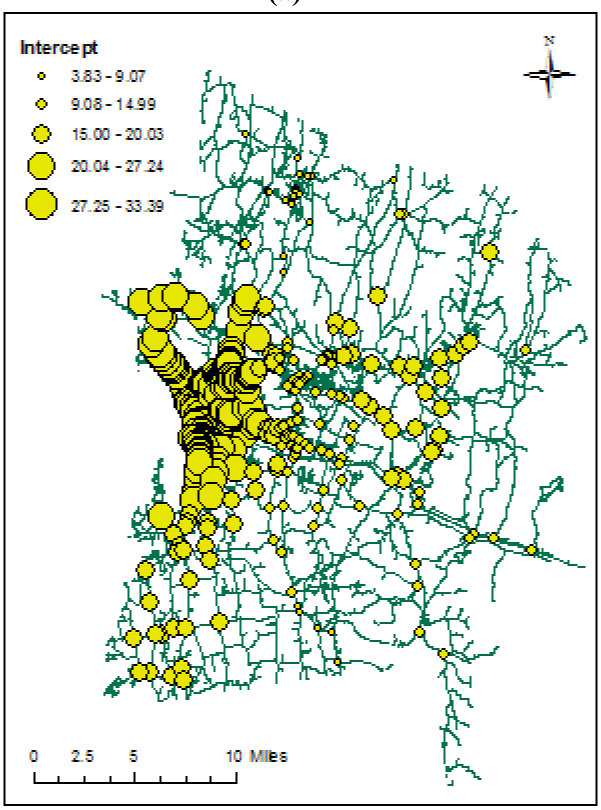
Figure 3(g) and (h) demonstrates the residuals and local R-square values. The countywide and irregular pattern for size differences among neighboring residuals well implies the spatial randomness of error terms after the GWR modeling. Simultaneously, the high proportion and rather evenly widespread distribution of larger R-square values indicates that Model A-4 is fitted with sound statistical validity.


Observations
0.00 - 10.12
10.13 - 21.40
21.41 - 31.57
31.58 - 42.12
42.13 - 57.58
N
0 2.5 5 10 Miles

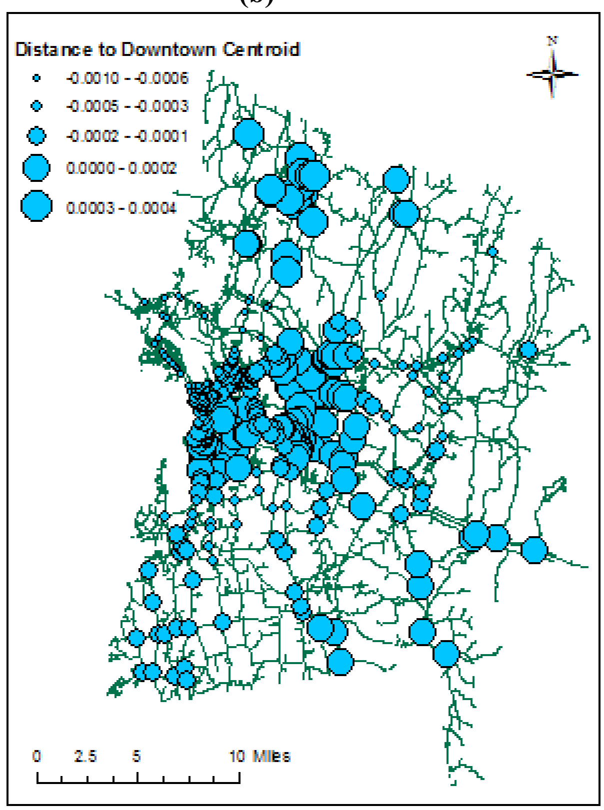
(a)


Predictions
6.04 - 15.40
15.41 - 21.68
21.69 - 29.23
29.24 - 37.90
37.91 - 49.07
N
0 2.5 5 10 Miles

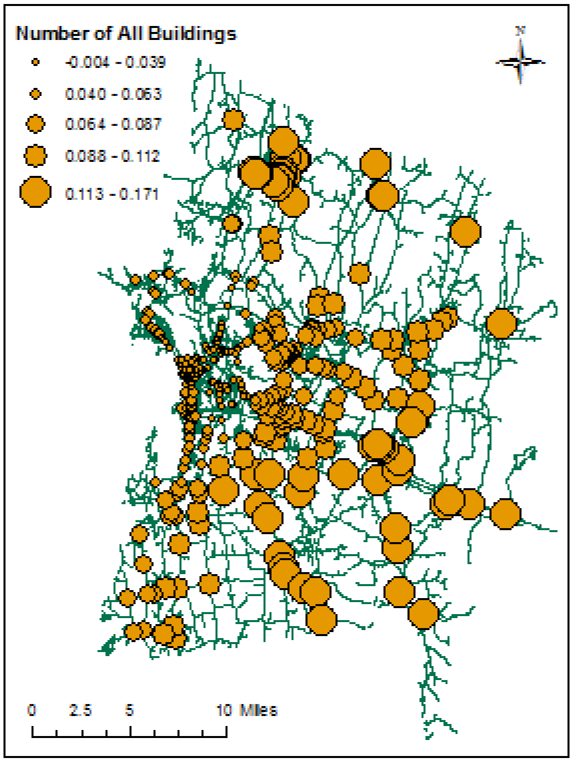
(b)


Intercept
3.83 - 9.07
9.08 - 14.99
15.00 - 20.03
20.04 - 27.24
27.25 - 33.39
N
0 2.5 5 10 Miles

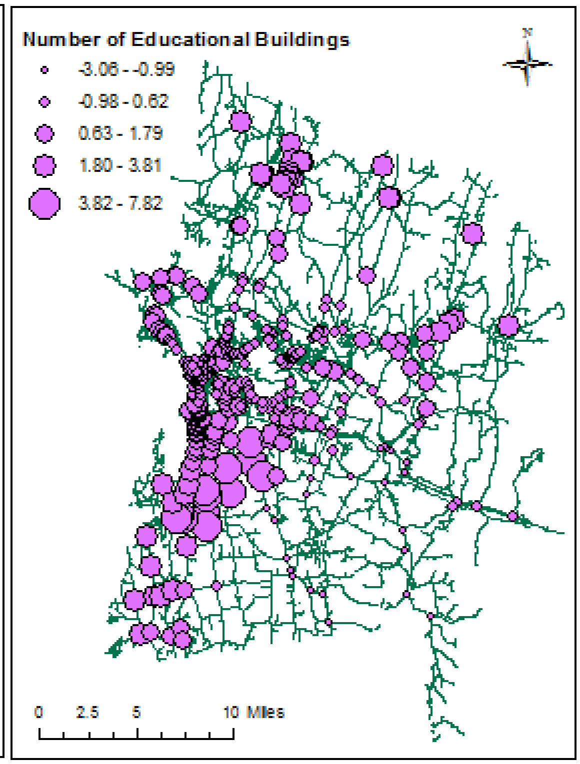
(c)


Distance to Downtown Centroid
-0.0010 - -0.0006
-0.0005 - -0.0003
-0.0002 - -0.0001
0.0000 - 0.0002
0.0003 - 0.0004
N
0 2.5 5 10 Miles

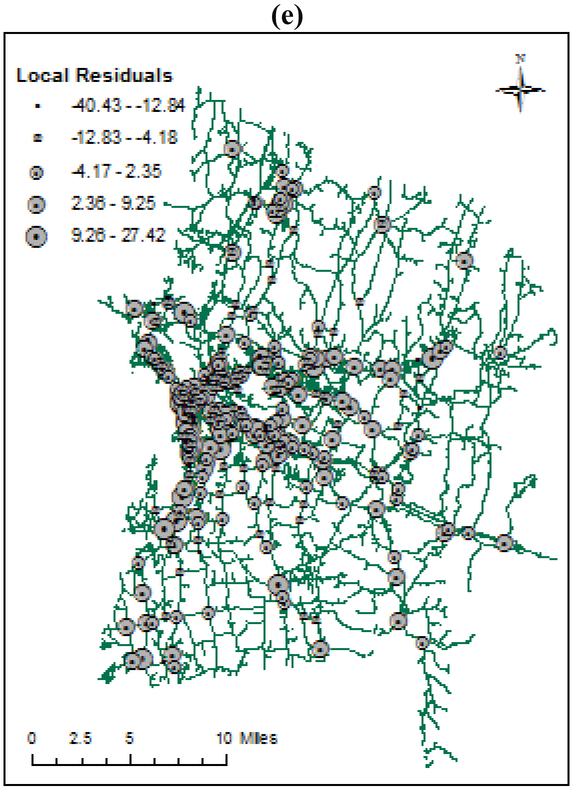
(d)



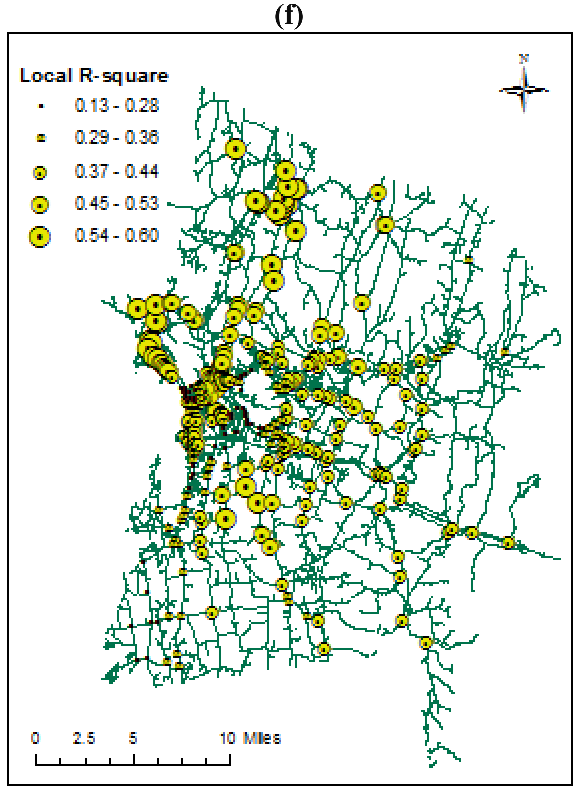
(e)



(f)



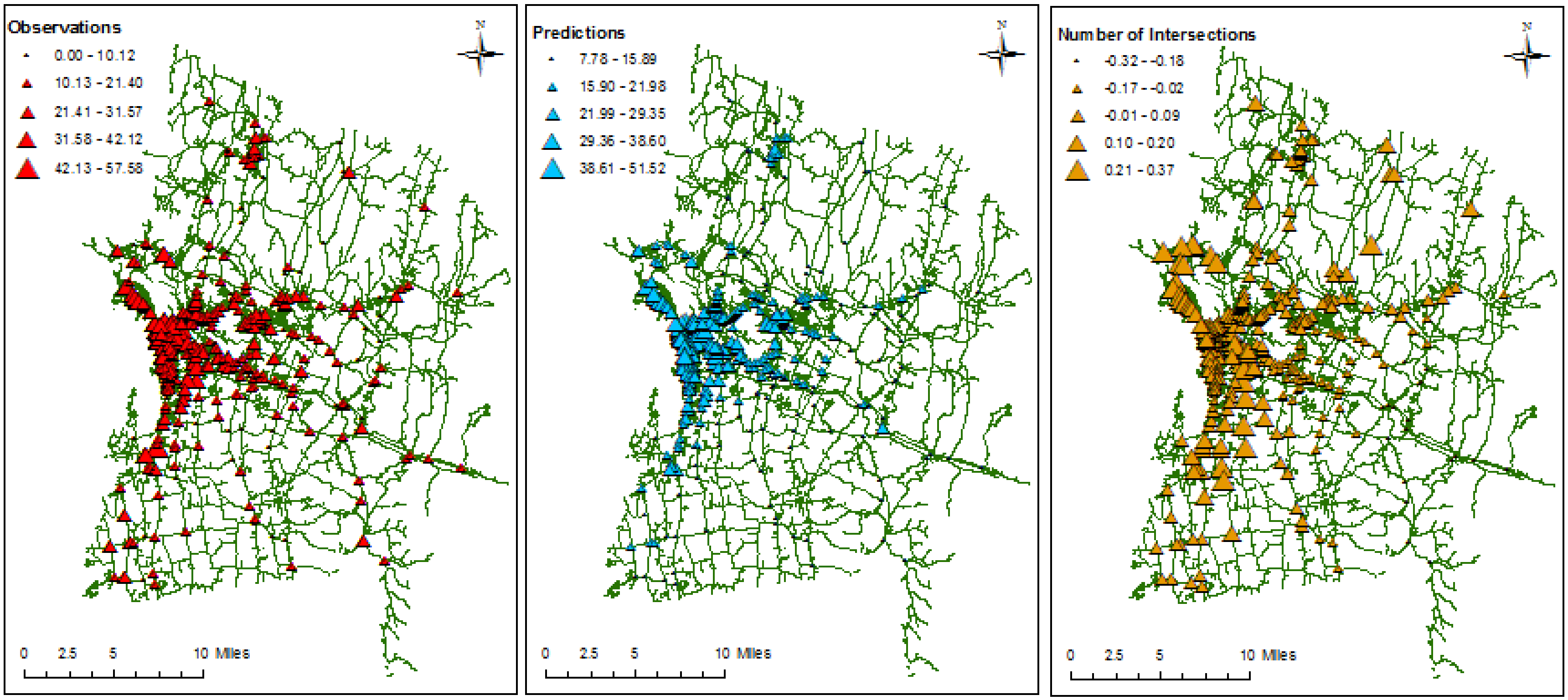
(g)



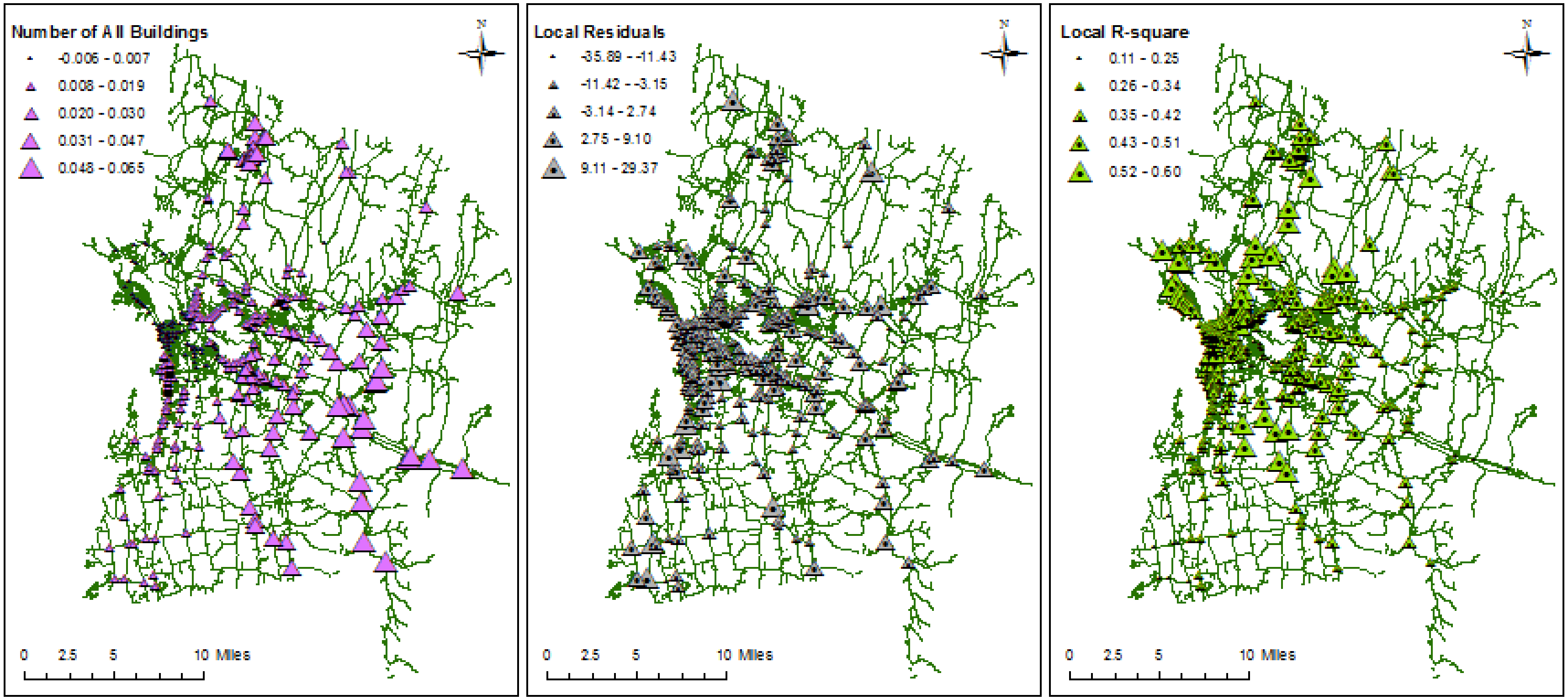
(h)

Figure 3 Mapping results for local estimates (Model A-4): (*a*) observations, (*b*) predictions, (*c*) intercept term, (*d*) distance to downtown centroid, (*e*) number of all buildings, (*f*) number of educational buildings, (*g*) local residuals, and (*h*) local R-square.

Figure 4(a-f) illustrates a mapping display of the GWR results from Model B-5, which suggests that the spatial pattern of the observed dependent variable much resembles that of the predicted one. The values around Burlington area are higher than those in surrounding rural areas. Additionally, as to the estimates for two predictors, there is a regional pattern with larger values of number of intersections in Burlington area and lower values elsewhere. The spatial characteristic of the estimates for the other variable, number of all buildings, is very similar to that for the counterpart in Model A-4. As to the residuals and local R-square values, again, there is a widespread and irregular pattern for size differences among neighboring residuals, and the considerable proportion and rather even distribution of bigger R-square imlies good fittness of Model B-5.


Observations
0.00 - 10.12
10.13 - 21.40
21.41 - 31.57
31.58 - 42.12
42.13 - 57.58
0 2.5 5 10 Miles
Predictions
7.78 - 15.89
15.90 - 21.98
21.99 - 29.35
29.36 - 38.60
38.61 - 51.52
0 2.5 5 10 Miles
Number of Intersections
-0.32 - -0.18
-0.17 - -0.02
-0.01 - 0.09
0.10 - 0.20
0.21 - 0.37
0 2.5 5 10 Miles

(a) **(b)** **(c)**



(d) **(e)** **(f)**

Figure 4 Mapping results for dependent variable and local estimates (Model B-5): (*a*) observations, (*b*) predictions, (*c*) number of intersections, (*d*) number of all buildings, (*e*) local residuals, (*f*) local R-square.

CONCLUDING REMARKS

This spatial study developed intersection-based models to understand how ambient built-environment elements spatially influence non-motorized travel in a county. Two standard statistics verified that strong spatial autocorrelation exists in the dataset, and bivariate analyses confirmed multicollinearity with most of the built-environment factors on two specific scales. GWR models were fitted on two data-collection scales to account for spatial heterogeneity.

On the smaller scale, number of all buildings and number of educational buildings were significant in local parametric variation to positively change walking and bicycling volumes. Distance from downtown centroid was also significant in such a sense. These results suggest that efforts for promoting safe walking and bicycling routes to schools may be an important way to promote these modes of travel, since these destinations are already strongly correlated with non-motorized travel. On the larger scale, total building density and intersection density were identified as spatially varying contributors significant in shaping non-motorized traffic volumes. The GIS mappings of GWR outputs clearly reveal the spatial relationships of these spatially significant contributors with non-motorized travel demand. These findings are consistent with previous studies which understood that non-motorized travel is more common where destinations are closer together and street connectivity is higher, typically in downtown urban-centers. However, this study has quantified precisely the extent to which these factors play a role in spatial dependency. Evidently, the scale for data collection is critical to spatial regression analysis.

The results can facilitate the estimation of bicycling and walking volumes at intersections across Chittenden County. Better estimation of non-motorized travel activity can be conducive to local transportation planning, infrastructure design, safety enhancement, and operational analysis. Given the count-type dependent variable which cannot be measured continuously in spatial domain, it is worthwhile to perform similar investigation via other geospatial methods such as point pattern analysis (*32*). It is generally accepted that to collect more demographic, socioeconomic, and infrastructural data is instrumental to unveiling more findings in spatial dependency.

ACKNOWLEDGEMENTS

This paper was derived from Signature Project 4C (Assessing the Impact of Weather, Season and Built Environment on Pedestrian Traffic) under the University Transportation Center program sponsorship from the United States Department of Transportation. The Chittenden County Metropolitan Planning Organization staff is much appreciated for data collection assistance. Andrew Weeks is acknowledged for his early work in project development. The authors deliver their deep gratitude to Professor Lisa Aultman-Hall of Civil & Environmental Engineering at the University of Vermont for helpful advices.

REFERENCES

1. Katz, P. *The New Urbanism: Toward an Architecture of Community*. New York, McGraw-Hill, Inc, 2004.
2. Rodrıguez, D. A. and J. Joo. The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D*, Vol. 9, 2004, pp. 151–173.
3. Guo, J. Y., C. Bhat, and R. B. Copperman. Effect of the Built Environment on Motorized and Non-Motorized Trip Making. *Transportation Research Record, No. 2010*, National Academies, Washington, D.C., 2007, pp. 1–11.
4. Cervero, R., and K. Kockelman. Travel Demand and the 3 Ds: Density, Diversity, and Design. *Transportation Research, Part D*, 2(3), 1997, pp. 199-219.
5. Frank, L., and P. Engelke. The built environment and human activity patterns: Exploring the impacts of urban form on public health. *Journal of Planning Literature*, 16, 2001, pp. 202–218.
6. Cervero, R., and M. Duncan. Walking, bicycling, and urban landscapes: Evidence from the San Francisco Bay Area. *American Journal of Public Health*, 93(9), 2003, pp. 1478–1483.

7. Frank, L., T. Schmid, J. Sallis, J. Chapman, B. Saelens. Linking objectively measured physical activity with objectively measured urban form: Findings from SMARTRAQ. *American Journal of Preventive Medicine*, 28, 2005, pp. 117–125.
8. Aytur S., D. Rodrıquez, K. Evenson, D. Catellier, and W. Rosamond. Promoting active community environments through land use and transportation planning. *American Journal of Health Promotion*, 21(4), 2007, pp. 397–407.
9. Cervero, R., O. L. Sarmiento, E. Jacoby, L. F. Gomez, and A. Neiman. Influences of Built Environments on Walking and Cycling: Lessons from Bogotá. *International Journal of Sustainable Transportation*, 3:203–226, 2009.
10. Liu, X-H, and J. Griswold. A Method for Estimating Pedestrian Volume for Street Intersections in San Francisco. Proceeding of the 86th Annual Transportation Research Board Meeting, January, Washington, D.C., 2007.
11. Aultman-Hall, L., D. Lane, and R. R. Lambert. Assessing the Impact of Weather and Season on Pedestrian Traffic Volumes, *Transportation Research Record, No. 2140*, National Academies, Washington, D.C., 2009, pp. 35–43.
12. Pulugurtha, S. S., and S. R. Repaka. Assessment of Models to Measure Pedestrian Activity at Signalized Intersections. In *Transportation Research Record, No. 2073,* National Academies, Washington, D.C., 2008, pp. 39–48.
13. Schneider, R. J., L. S. Arnold, and D. R. Ragland. Pilot Model for Estimating Pedestrian Intersection Crossing Volumes. In *Transportation Research Record, No. 2140,* National Academies, Washington, D.C., 2009, pp. 13–26.
14. Eom, J. K., M. S. Park, T-Y. Heo, and L. F. Huntsinger. Improving the Prediction of Annual Average Daily Traffic for Nonfreeway Facilities by Applying a Spatial Statistical Method. *Transportation Research Record, No. 1968*, National Academies, Washington, D.C., 2006, pp. 20–29.
15. Zhao, F., and N. Park. Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. In *Transportation Research Record, No. 1879,* National Research Council, Washington, D.C., 2004, pp. 99–107.
16. Zahran, S., S. D. Brody, P. Maghelal, A. Prelog, and M. Lacy. Cycling and walking: explaining the spatial distribution of healthy modes of transportation in the United States. *Transportation Research, Part D*, 13: (7), 2008, pp. 462-470.
17. Pucher, J., R. Buehler, and M. Seinen. Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research, Part A*, 45 (6), 2011, pp. 451-475.
18. Washington, S., M. Karlaftis, and F. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis.* Chapman & Hall/CRC, Boca Raton, Fla, 2003.
19. VTrans. Continuous Traffic Counter Grouping Study and Regression Analysis Based on 2009 Traffic Data. Prepared by the Vermont Agency of Transportation, Planning, Outreach & Community Affairs: Traffic Research Unit. Dated March 2010.
20. Owens, P. M., L. Titus-Ernstoff, L. Gibson, M. L. Beach, S. Beauregard, and M. A. Dalton. Smart density: A more accurate method of measuring rural residential density for health-related research. *International Journal of Health Geographics*, 9:8, 2010.
21. Tobler, W. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46, 1970, 234-240.
22. Moran, P. A. P. Notes on Continuous Stochastic Phenomena. *Biometrika*, 37, 1950, pp. 17–33.
23. Geary, R. C. The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5 (3), 1954, pp. 115–145.
24. Sokal, R. R. and N. L. Oden. Spatial autocorrelation in biology. Methodology. *Biological Journal of the Linnean Society*, 10, 1978, pp. 199-228.
25. Cliff, A. D., and J. K. Ord. The choice of a test for spatial autocorrelation. In J. C. *Davies and M. J. McCullagh (eds) Display and Analysis of Spatial Data*. John Wiley and Sons, London, 54-77, 1975.
26. Vichiensan, V., A. Páez, K. Kawai, and K. Miyamoto. Nonstationary Spatial Interpolation Method for Urban Model Development. In *Transportation Research Record, No. 1977,* National Academies, Washington, D.C., 2006, pp. 103–111.
27. Brunsdon, C., A. S. Fotheringham, and M. Charlton. Geographically weighted regression: a method for exploring spatial non-stationarity. *Geographical Analysis*, 28(4), 1996, pp. 281-298.
28. Fotheringham, A. S., C. Brunsdon, and M. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationship.* John Wiley & Sons, New York, 2002.

29. Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. *Proc., Second Symposium on Information Theory,* Akademiai Kiado, Budapest, Hungary, 1973, pp. 267–281.
30. Burnham, K., and D. Anderson. *Model Selection and Multi-Modal Inference.* Springer-Verlag, New York, 2002.
31. O'Sullivan, D., and D. Unwin. *Geographic Information Analysis*. John Wiley & Sons Inc., pp. 180-205, 2003.
32. Fingleton, B., G. J. G. Upton. *Spatial Data Analysis by Example: Point Pattern and Quantitative Data*. Wiley Series in Probability and Statistics, John Wiley & Sons Inc, 1985.
33. Chow, L-F., F. Zhao, X. Liu, M-T. Li, and I. Ubaka. Transit Ridership Model Based on Geographically Weighted Regression. *Transportation Research Record, No. 1972,* National Academies, Washington, D.C., 1972, pp. 105–114.