A Report from the University of Vermont Transportation Research Center

Spatial Analysis of Intersection Bicycle and Pedestrian Counts

TRC Report 13-013 | Sullivan, Lu, Troy | February 2015

Spatial Analysis of Intersection Bicycle and Pedestrian Counts

February 5, 2015

Prepared by:

Jim Sullivan
George Lu
Austin Troy

Transportation Research Center
Farrell Hall
210 Colchester Avenue
Burlington, VT 05405

Phone: (802) 656-1312
Website: www.uvm.edu/trc

Acknowledgements

The authors would like to acknowledge the USDOT for funding this work through the University Transportation Centers (UTC) at the University of Vermont. Data and technical input was provided by the teams at the Chittenden County Regional Planning Commission, Local Motion and the Vermont Agency of Transportation. The contributions to this body of work by Lisa Aultman-Hall and Jonathan Dowds are greatly appreciated.

Table of Contents

List of Tables

List of Figures

1 Introduction and Objectives

Encouraging travelers to walk and bike in lieu of motorized modes of travel benefits both the traveler and the community at large. Maximizing these system benefits is critically important for the state and municipalities, especially when funding for transportation is scarce. In order to make better funding decisions for non-motorized transportation infrastructure, it is first necessary to understand comprehensively the current walking and biking behavior of a region's inhabitants.

This study investigates the linkage between non-motorized traffic volumes and the built environment by focusing on a larger set of road intersection-based counts of the PM-peak hours. The dearth of effective methods to address the spatial dependencies present in these comprehensive data sets motivated this geospatial study to determine (a) whether spatial dependency exists for non-motorized traffic volumes, and (b) whether a significant spatial relationship could be identified between non-motorized traffic volumes and specific built-environment characteristics once the spatial dependency was accounted for. Addressing this non-random factor in spatial based counts is an essential step to attaining a robust understanding of bicycle and pedestrian travel throughout a region. Some of the technical information covered in this report was also compiled in a conference paper (Lu et. al., 2012).

For a better prediction of motorized and non-motorized travel on multimodal facilities, spatial dependency must be considered because traffic volume at one monitoring station is related to the volume at neighboring stations due to the routing and the continuity in the network due to area-wide traffic circulation and common origins and destinations. A few studies have acknowledged this spatial dependency. Of them, Eom et al. (2006) researched annual average daily traffic (AADT) using spatial Kriging estimation. The spatial model outperforms that of the ordinary least-square (OLS) model. Zhao and Park (2004) analyzed AADT in grid-like networks utilizing geographically weighted regression (GWR) that 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. A smaller group of studies have conducted geospatial analyses of walking and bicycling with appropriate recognition of spatial dependency. Zahran et al. (2008) acknowledged spatial dependency in their study of nationwide county-based data.

2 Study Area

The study area for this project is Chittenden County, Vermont, the planning region for the Chittenden County Regional Planning Commission (CCRPC) (see Figure 1).

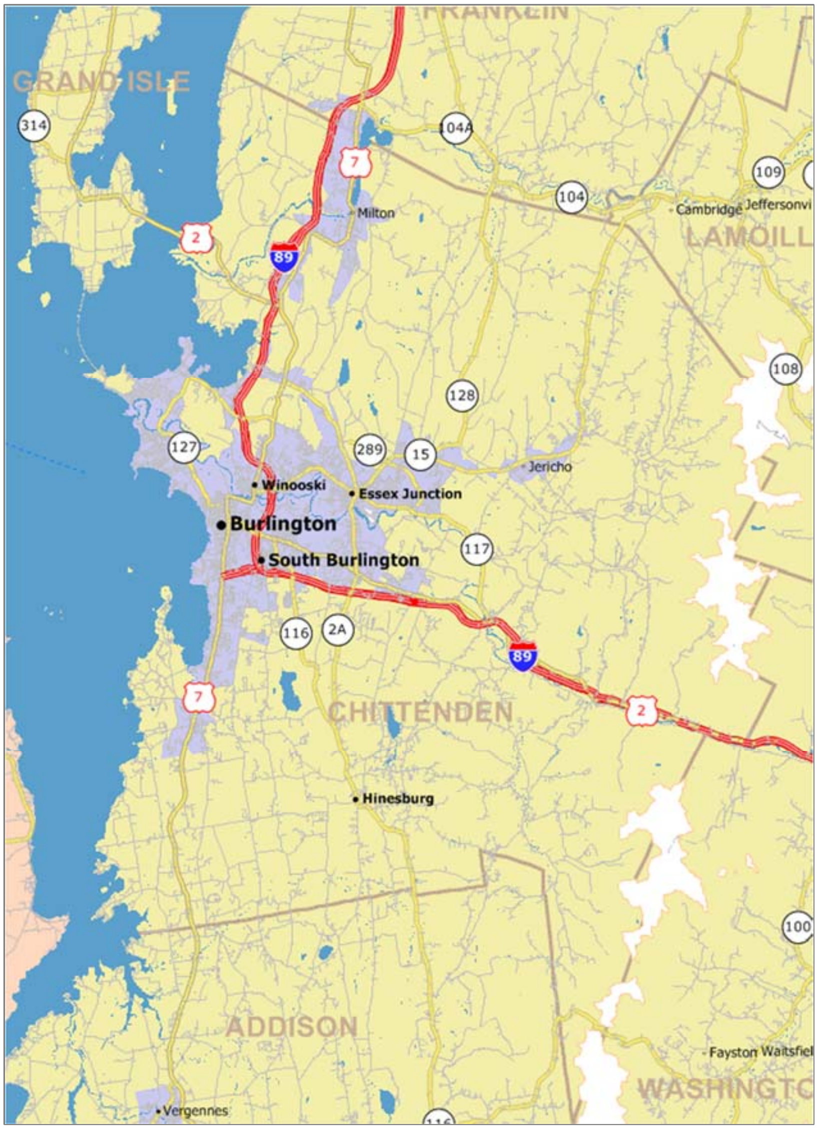


Figure 1 Project Study Area

The CCRPC area includes a 62-square-mile urban area that contains Burlington, the largest city in Vermont. It is bounded to the west by Lake Champlain and to the east by public lands in the Green Mountains. Chittenden County has the largest population and employment in the state, with approximately 150,000 residents (of approximately 620,000 in Vermont) and more than 100,000 jobs. Like most regions in the country, the urban core has spread into neighboring municipalities and now includes a suburban development pattern around Burlington.

3 Data

3.1 Intersection-Based Non-Motorized Traffic Count Data

Intersection-based counts were manually collected by CCRPC using traffic-count boards from each inbound approach of 428 intersections throughout the study area between 2000 to 2009 (black and red dots in Figure 2).

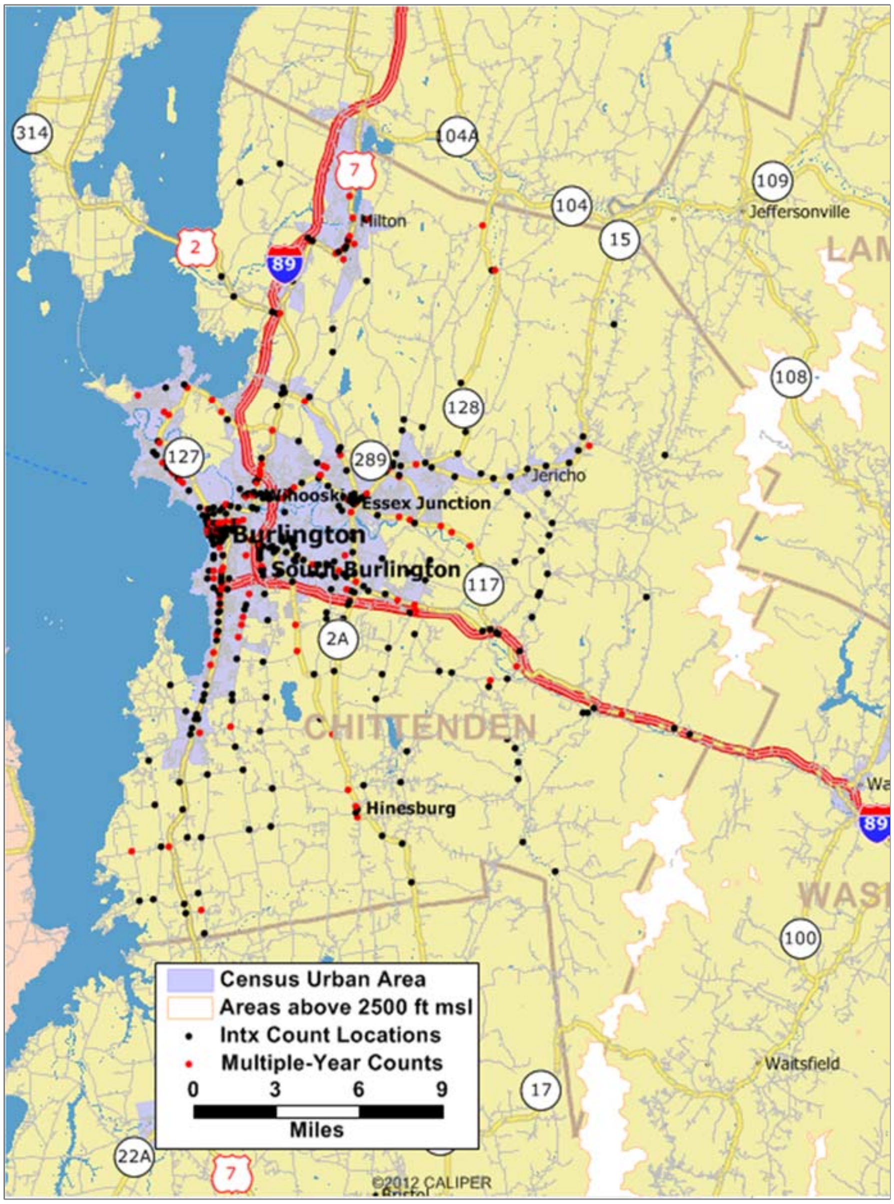


Figure 2 Intersection-Based Non-Motorized Traffic Count Locations, 2000 - 2009

Bicycling and walking volumes present at each intersection were recorded only as a total for each approach – the turning movements or outbound approach was neglected. At some locations, the counts represented hours from multiple years in the 10-year period (red dots in Figure 2). Aggregated by hourly-total count, the initial dataset consisted of 3,541 records, or an average of 8 hourly totals per location. Almost all of count locations encompassed the 2-hour PM-peak period (4:00-6:00PM), and many of them included more hours in the day, up to a maximum of 12 hours between 7:00am and 7:00pm. Growth factors were applied to normalize data.

3.2 Land-Use and Infrastructure Data

Land uses in the study area were taken from the Vermont E911 database and geographical information system (GIS), which consists of the location and functional classification of each habitable structure in the state. The Vermont E911 data includes residential locations (single-family, multi-family, seasonal, and mobile homes) and non-residential locations (commercial, industrial, educational, governmental, health-care and public gathering). Vermont is unique in that this E911 database is publicly available to support emergency-response personnel statewide via the Vermont Center for Geographic Information (VCGI).

Ambient land-use and infrastructural attributes that are commonly associated with non-motorized travel in the literature were selected as independent variables, (Owens et. al., 2010; Cervero and Kockelman, 1997). Most of these attributes require the identification of a buffer area within which the attribute is measured around a count at one of the 346 intersection-based count locations in the reduced data set. A 1,000-ft buffer area was selected as a rough approximation of a common median walking-trip distance, and a 2,500-ft buffer area was selected as a rough approximation of a common median bike-trip distance, and a maximum walking-trip distance. The descriptive statistics for each of these independent variables are compiled in Table 1.

Table 1 Descriptive Statistics for Independent Variables to be Modeled

Independent Variable	Buffer	Min.	Max.	Mean	Std Dev.
Count of All Buildings	1,000 feet	0	525	93.7	109.9
	2,500 feet	4	2,345	502.7	535.2
Count of Commercial Buildings	1,000 feet	0	228	15.5	31.9
	2,500 feet	0	464	63.7	102.6
Count of Educational Buildings	1,000 feet	0	31	1.0	3.5
	2,500 feet	0	98	5.0	13.0
Count of Public Buildings	1,000 feet	0	15	2.0	2.9
	2,500 feet	0	54	8.3	9.4
Count of Residential Buildings	1,000 feet	0	499	75.0	96.6
	2,500 feet	3	2,124	425.1	456.3
Count of All Intersections	1,000 feet	1	30	9.1	7.0
	2,500 feet	1	142	41.6	33.5
Total Roadway Length (miles)	1,000 feet	0.2	1.2	0.6	0.3
	2,500 feet	0.4	5.9	2.5	1.3
Neighborhood Connectivity Density (NCD)*	1,000 feet	0.6	11.1	5.3	2.4
	2,500 feet	0.5	10.8	4.9	2.0
Distance from Burlington Urban Area Centroid (mi.)	NA	0.03	5.8	1.7	1.3

Notes: *NCD is the number of intersections divided by total road length

4 Data Preparation

4.1 Initial Reduction

Outliers in the initial dataset of hourly total volumes were eliminated. A total of five records were removed because the hourly totals exceeded 500 travelers, reducing the size of the data set to 3,536 records. It was then necessary to map the 3,536 records to intersections so that spatial analyses would be possible. The statistical treatment of spatial dependency required one record per observation site. First, to eliminate the need for a weekly correction, only observations from Tuesdays, Wednesdays, or Thursdays were kept. Weekends and Fridays were eliminated because it was assumed that intersection-based non-motorized travel behavior varied between weekdays and weekends. Mondays were eliminated due to the common occurrence of holidays and the associated influence of the weekend activities. Then, only summer observations (during June, July, and August) were used to eliminate the need for seasonal correction. Finally, only PM-peak 2-hour volumes were considered to exclude the need for a daily correction. Once these filtering steps had been executed, it was assumed that a more homogeneous data set would result. No daily, weekly, or seasonal adjustments would be needed for the final set of 964 2-hour count totals at 346 intersections.

4.2 Temporal Corrections

Obviously, some intersections had more than one 2-hour count total. This occurred when the PM-peak period for an intersection was counted in multiple years during the 10-year study period. 115 of the 346 intersections had multiple records representing repeated PM-peak 2-hour counts in separate years. Initially it was assumed that corrections should be made to compensate for the temporal variation in the age of the reduced dataset in the event that annual growth or decline in non-motorized traffic had occurred. OLS regression was applied to estimate temporal growth trends for: (i) all data from 2000 to 2009; (ii) data for only those locations where more than one year was represented. OLS regression results for all data indicated a very low coefficient of determination (R^2) of 0.004. This meant that no general growth trend existed in the study area during the 10-year study period. For the 115 intersections where more than one year was represented, simple linear regression was used to calculate location-specific growth rates. Figure 3 contains a histogram of the growth rates calculated for each of these intersections in bins at 0.1% increments.

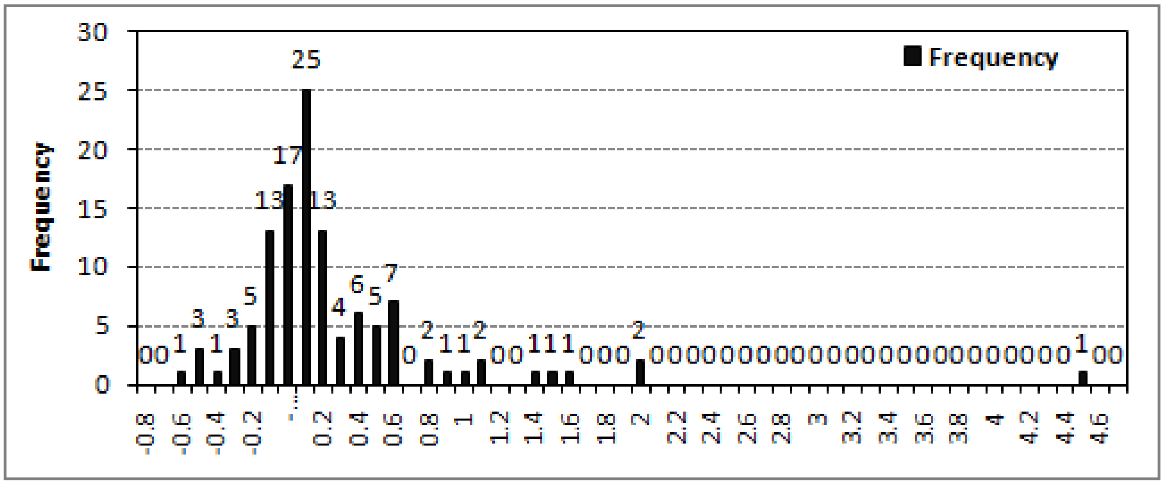


Figure 3 Histogram of Growth Rates for Intersections with Multiple Years of Counts

A total of 37% of the growth rates are negative and 63% are positive. The minimum and maximum are -0.7% and +4.5% respectively. The mean and median are 0.2% and 0.1% respectively. The “0.0%-0.1%” bin contains 25 out of 115 growth rates, accounting for the largest portion. Based on these characteristics, it was assumed that the data could be used confidently without temporal corrections for all 10 years of study period. This finding is consistent with motorized traffic volumes in Chittenden County, which have seen no significant changes through that same period (VTrans, 2010).

Supported by this finding, the intersections represented by multiple years were aggregated into a single volume by averaging the count records from each year and rounding them to an integer. The resultant values are summarized in Table 2. The histogram for these station-based counts resembles a typical Poisson or Negative Binomial distribution (Figure 4).

Table 2 Characteristics of Final Intersection-Based Set of PM-Peak Traffic Counts

Characteristic	No. of Observations or PM-Peak 2-Hour Volume
Number of Observations	346
Minimum	0
Maximum	316
Mean	23
Median	8
Standard error	2

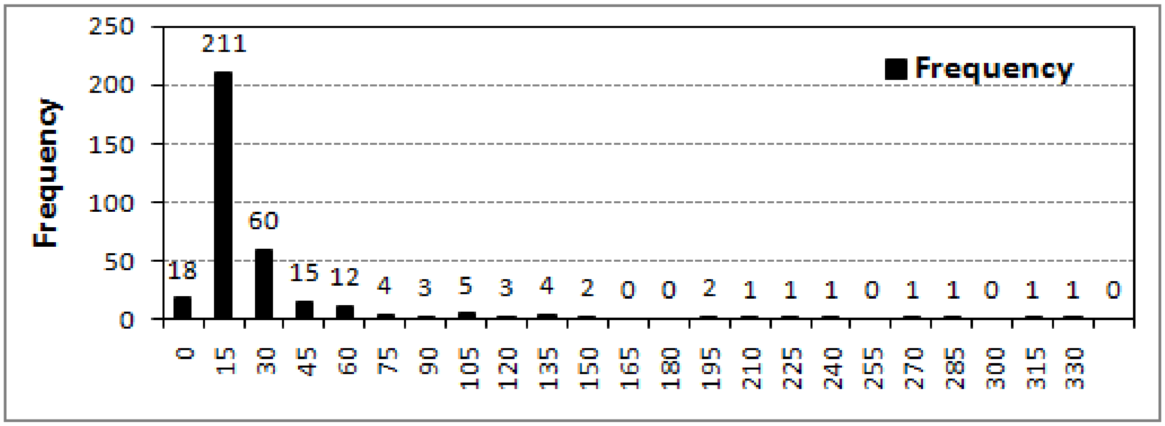


Figure 4 Histogram of PM-Peak, 2-Hour Intersection-Based BP Counts for 2000 - 2009

5 Results

The entities at nearby locations often share more similarities than the entities far apart. Often, 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" (Tobler, 1970). 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" in which 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. Spatial regression models capture spatial relationships to avoid these weaknesses. This study was intended to: (a) Identify SA for the dependent variable and each of independent variables; (b) Use a spatial regression model to find the significant spatially-dependent relationships between the non-motorized volume and the independent variables of interest.

5.1 Measurement of Spatial Dependency

Spatial dependency can be assumed to exist among multimodal traffic at neighboring intersections. Firstly, the neighboring intersections share a large portion of through-volumes. Secondly, traffic signal coordination promotes the group formation of traversing travelers. Moreover, neighboring intersections are surrounded by similar land-use characteristics and infrastructural elements. For instance, two or more intersections can lie within the walking distance of one building. Hence, traffic volumes likely exhibit SA among close intersections across two-dimensional domain. With spatial coordinates, 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 require a spatial weights matrix which reflects the intensity of the geographic nexus among neighboring observations, being global in the sense that they estimate the overall degree of SA. Moran's I was calculated for n observations on a variable x at locations i, j, in terms of cross-products of the deviations from the mean (Moran, 1950):

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \tag{3}$$

Where:

$\bar{x}$ – the mean of the x variable;
w_{ij} – the elements of the weight matrix;
S_0 – the sum of the elements of the weight matrix: $S_0=\Sigma_i \Sigma_j w_{ij}$; and

I – changing from -1 (perfect dispersion) to +1 (perfect correlation); 0 indicates a random spatial pattern.

Positive SA happens when similar values exist nearby, while negative SA happens for dissimilar values. Geary's c is based on the deviations in responses of each observation, as the equation below shows (Geary, 1954).

$$c = \frac{n-1}{2S_0} \frac{\sum_i \sum_j w_{ij} \left(x_i - x_j\right)^2}{\sum_i \left(x_i - \bar{x}\right)^2} \tag{4}$$

The expectation of c is 1 in the absence of SA, regardless of the specified weight matrix (Sokal and Oden, 1978). Geary's c index ranges from 0 to 2. A value of 1 means there is no SA, whereas values at the range edges indicate a positive (0 to 1) or negative (1 to 2) SA. Both indices are inversely related to each other. Moran's I is found consistently more powerful than Geary's c (Cliff and Ord, 1975). The former is a more global measurement and sensitive to extreme values, whereas the latter is more sensitive to differences in close proximity. A natural logarithm transformation was applied to the count volumes to facilitate the SA analysis by making it closer to normality. The W_{ij} parameter is used as a distance-based weight, which denotes the inverse of the distance between two intersections.

Moran's I has an expected value of [-1/(n-1)], which approaches zero as the sample size expands or in the absence of SA. 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 (Moran, 1950). Table 3 shows all observed Moran's I indices and Geary's c indices for all of the variables to be modeled in this study.

Table 3 Moran's I and Geary's c Calculation Results

Variable	Observed		Z		Pr > |Z|	
	I[1]	c[2]	I	c	I	c
Dependent variable[3]	0.23	0.88	76.19	-4.49	<.0001	<.0001
Distance to Burlington downtown centroid	0.19	0.69	63.1	-11.6	<.0001	<.0001
Predictors within the 1,000-ft scope (Buffer Area A)						
Number of all buildings	0.26	1.01	86.71	0.37	<.0001	0.710
Number of all intersections	0.20	0.93	66.15	-2.55	<.0001	<.011
Number of commercial buildings	0.16	1.29	52.20	10.70	<.0001	<.0001
Number of educational buildings	0.08	1.38	25.70	14.00	<.0001	<.0001
Number of public buildings	0.09	1.07	28.65	2.67	<.0001	0.008
Number of residential buildings	0.21	1.06	67.94	2.23	<.0001	0.026
Total roadway length	0.18	0.90	59.78	-3.56	<.0001	0.0004
NCD	0.11	0.90	35.66	-3.91	<.0001	<.0001
Predictors within the 2,500-ft scope (Buffer Area B)						
Number of all buildings	0.34	0.92	110.3	-3.17	<.0001	0.002
Number of all intersections	0.30	0.85	99.6	-5.67	<.0001	<.0001
Number of commercial buildings	0.27	1.12	87.7	4.25	<.0001	<.0001
Number of educational buildings	0.18	1.28	59.8	10.30	<.0001	<.0001
Number of public buildings	0.23	0.95	75.9	-1.85	<.0001	0.065
Number of residential buildings	0.31	0.92	100.5	-2.88	<.0001	0.004
Total roadway length	0.30	0.81	98.7	-7.13	<.0001	<.0001
NCD	0.20	0.77	67.0	-8.55	<.0001	<.0001

Notes:
1. I – Moran's I index (expected values were -0.003)
2. c – Geary's c index (expected values were 1.00)
3. After logarithmic transformation for normality approximation.

All of the Moran's I indices exceed the expectation -0.003 with Z-scores larger than 1.96, which means a significant positive SA. The Geary's c indices also indicate positive SA (O'Sullivan and Unwin, 2003). The independent variable "Number of All Buildings" is illustrated in Figure 5 for each buffer area to demonstrate the presence of autocorrelation.

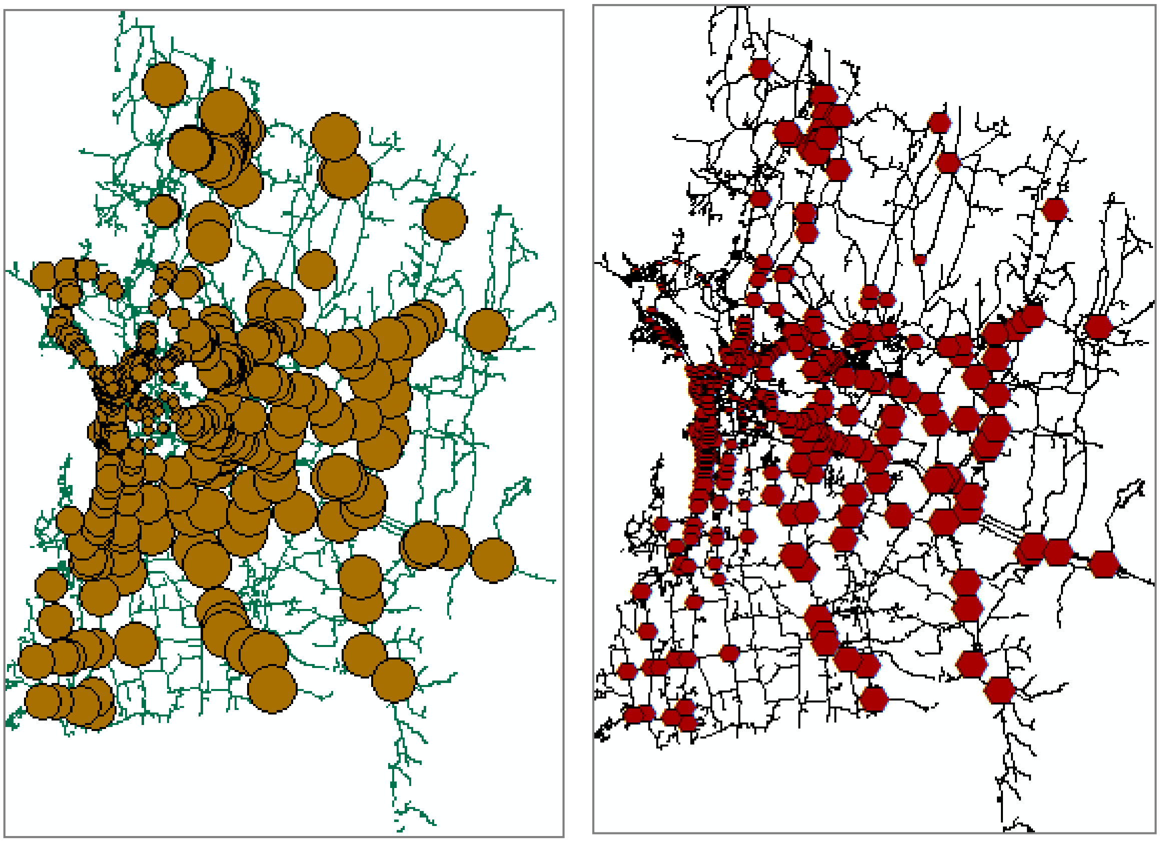


Figure 5 Number of All Buildings within 1,000 feet (left) and 2,500 feet (right) of each Intersection-Based Count Location

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. The test results suggest that most of these variables are significantly correlated with each other in each buffer area. Expectedly, number of all buildings has very 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 assumed that the number of educational buildings is a variable that should be treated separately in regression modeling. Based on the bivariate analysis results for both buffer areas, GWR procedures were separately applied to two buffer areas. The multicollinearity was tested for each buffer area separately.

5.2 Spatial Regression

SA may operate in twofold forms: spatial dependency and spatial heterogeneity both of which are principal challenges in spatial analysis. Depending on specific statistical techniques, spatial dependency can enter 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 between the dependent variable and the independent variables. Generally, a model with autocorrelated errors could produce better estimators and predictors than an OLS model but may be outperformed by universal Kriging for the purpose of producing optimal predictors (Vichiensan et. al., 2006). None of these models, however, explicitly 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 assumedly homogeneous everywhere, removing the possibility that the process operates differently in varied locations (Vichiensan et. al., 2006).

5.2.1 GEOGRAPHICALLY-WEIGHTED REGRESSIONS

GWR models spatially heterogeneous processes in various areas, with the underlying philosophy that parameters may be estimated anywhere in study area given a dependent variable and a set of independent variables measured at locations whose spatial coordinates exist (Brunsdon et. al., 1996; Fotheringham et. al., 2002). GWR gives relatively more weight to geographically close observations and less (or zero) to distant ones. This weighting scheme assumes that using geographically close observations is essential to estimating local coefficients. Traditionally, a linear regression model may be written as:

$$y_i = \mathrm{X}_i^T \mathrm{B} + \varepsilon_i \qquad (5)$$

Where:

$\mathrm{X}_i = \{1, x_{i1}, ..., x_{ik}, ..., x_{iK}\}^T$ – The $(K+1)$-dimensional vector of the ith independent observation ($i = 1, \ldots, N$; $k = 1, \ldots, K$);
$\mathrm{B} = \{\beta_0, \beta_1, ..., \beta_k, ..., \beta_K\}^T$ – The $(K+1)$-dimensional vector of coefficients for the intercept and independent variables;
x_{ik} – The kth independent variable for the ith observation;
N – Total number of independent observations;
K – Total number of independent variables;
ε_i – Random error term.

When the traditional linear model is applied, one global parameter is estimated for each independent variable to represent its relationship to the dependent variable. The OLS estimates for parameters take the form:

$$\hat{\mathrm{B}} = (\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T\mathrm{y} \qquad (6)$$

Where $\hat{\mathrm{B}}$ represents the vector of global parameters estimated, X is a matrix of intercept and independent variables, and **y** represents a vector of observations on the dependent variable. $\hat{\mathrm{B}}$ is constant irrelevant to the spatial locations of N observations. GWR extends the framework by estimating local parameters as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{K} \{\beta_k(u_i, v_i) x_{ik}\} + \varepsilon_i \tag{7}$$

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 $B(u, v)$ 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 on the surface (Brunsdon et. al., 1996; Fotheringham et. al., 2002). Algebraically, the GWR estimator is:

$$\hat{\mathrm{B}}\,(u, v) = (\mathrm{X}^{\mathrm{T}}\mathrm{W}(u_i, v_i)\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{W}(u_i, v_i)\mathrm{y} \tag{8}$$

Where W (u_i, v_i) is a $n \times n$ matrix whose off-diagonal and diagonal elements respectively 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 the intercept but also it addresses spatial heterogeneity by incorporating coordinates in parameter estimates (Fotheringham et. al., 2002). There is evidence that GWR can reduce the *residuals* more substantially, compared with models 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 (Fotheringham et. al., 2002).

The 1,000-ft buffer area is enclosed within the 2,500-ft buffer area, so each variable of the same category (e.g., number of educational buildings) in the former is numerically a subset of the latter. Statistically, this situation creates a multicollinearity issue that must be addressed for 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. A positive or negative correlation means respectively a perfectly linear relationship in an ascending or descending 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.

Within a modeling process with the philosophy of "model parsimony", all 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 (Akaike, 1973). When the AIC values for two models differ by more than 3, the models are considered significantly different (Fotheringham et. al., 2002). A model with smaller AIC is considered closer to the unknown true model (Burnham and Anderson, 2002).

Table 4 displays the results from GWR application to the 1,000-ft buffer area.

Table 4 Results of the GWR Model for the 1,000-foot Buffer Area

Independent Variable	*p*-values for significance test						
	A-1	A-2	A-3	A-4	A-5	A-6	A-7
Land-use							
Number of all buildings	0.29	0.00	0.00	**0.00**	0.00		0.00
Number of commercial buildings	0.40						
Number of educational buildings	0.95	0.00	0.00	**0.00**		0.01	0.00
Number of public buildings	0.53						
Number of residential buildings	0.33						
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 attribute							
Intercept	0.11	0.08	0.00	**0.00**	0.00	0.00	0.00
Measures of fit							
AIC	2486	2492	2481	**2469**	2470	2505	2464
R-squared	0.561	0.632	0.625	**0.636**	0.623	0.597	0.629
Adjusted R-squared	0.542	0.581	0.582	**0.595**	0.587	0.551	0.594

All p-values from Monte Carlo heterogeneity tests less than 0.05 mean the relevant parameters significantly vary from place to place. Both Bi-square and Gaussian models were fitted and the former was found to have the better results. Model A-1 encompasses all independent variables whose parameters vary insignificantly given the p-values above 0.05. The foregoing Pearson test revealed that the number of total buildings is strongly correlated with number of intersections and each of the land-use variables except for number of educational buildings, so Model A-2 excluded these three land-use variables and that infrastructure-based variable. Model A-2 reveals that number of total 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 excluding total roadway length, Model A-3 shows number of total buildings and number of educational buildings have significant heterogeneity, and the model parsimony has been increased by lowering the AIC from 2492 to 2481. Model A-4 omits NCD and leads to an AIC reduction from 2481 to 2469. Unfortunately, Models A-5 and A-6 reverse the trend if either of significant land-use variables in A-4 is included, increasing the AIC (from 2469 to 2470 and 2505 respectively) and decreasing the number of significant variables from three to two. Model A-7 re-includes both of the significant land-use variables but omits distance to downtown centroid, yielding the lowest AIC and significant p-values for both variables. Model A-7 has a lower AIC (2464 vs. 2469) but a slightly lower R-square (0.629 vs. 0.636), compared with Model A-4. Given the similar parsimony and model validity, Model A-4 importantly unveils more underlying information in shaping spatial relationship. Therefore, Model A-4 is used as the final model on this buffer area.

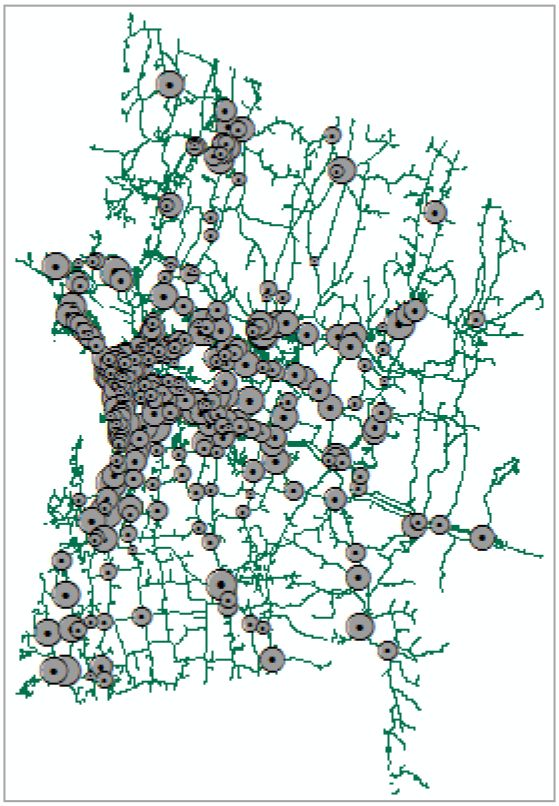
Table 5 demonstrates the results from GWR application to the 2,500-ft buffer area.

Table 5 Results of the GWR Model for the 2,500-foot Buffer Area

Independent Variable	*p*-values for significance test						
	B-1	B-2	B-3	B-4	**B-5**	B-6	B-7
Land-use							
Number of all buildings	0.36	0.03	0.00	0.00	**0.00**	0.00	
Number of commercial buildings	0.35						
Number of educational buildings	0.36						
Number of public buildings	0.34						
Number of residential buildings	0.36						
Network							
Number of intersections	0.16	0.04	0.05	0.05	**0.02**		0.00
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 attribute							
Intercept	0.00	0.00	0.00	0.00	**0.00**	0.05	0.01
Measures of fit							
AIC	2463	2455	2449	2452	**2442**	2450	2463
R-squared	0.625	0.629	0.617	0.649	**0.643**	0.644	0.639
Adjusted R-squared	0.592	0.598	0.595	0.612	**0.613**	0.610	0.600

Model B-1 includes all independent variables. The multicollinearity identified on this buffer area makes it necessary to exclude other land-use variables in further modeling. Then, Model B-2 yields three variables with significant heterogeneity, including number of all buildings, number of intersections, and NCD. Total roadway length and distance to downtown centroid have insignificant parametric heterogeneity based on p-values. 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. Continually, Model B-5 reduces the AIC from 2,452 in

Model B-4 to 2,442 and also results in a significant intercept term in local parameter variation. Although each of Model B-6 and Model B-7 reveals a significant parameter variation, they increase the AIC and the p-values for the 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 (2,442). Figure 6 provides corresponding maps of the residuals from the estimates for each model at each data point in the study area.



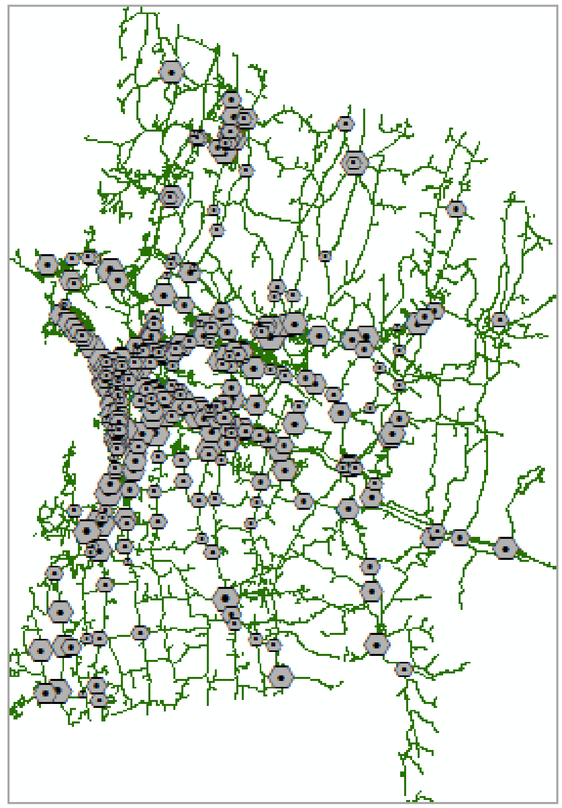


Figure 6 Residuals for Estimates from GWR Model A-4 (left) and GWR Model B-5 (right)

A comparison of Figure 6 and Figure 5 indicates the effect of the GWR procedure on accounting for spatial autocorrelation. The lack of spatial trends in the residuals shown in Figure 6 indicates that the GWR models successfully accounts for the spatial dependency that was present in the independent variables, an example of which is shown in Figure 5.

The independent variables share one common characteristic in parametric heterogeneity - the majority or large number of smaller values are scattered in proximity of Burlington area, and these larger values are spread 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 the 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 a limited increase in non-motorized demand. Comparatively, when the buildings are sparsely located where non-motorized travel demand is low, the introduction of a new building may have a more drastic influence upon additional non-motorized travels. Note that for these three variables the patterns are somewhat different from 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 disparate from that of the former due to the consideration for educational coverage of local populations.

5.2.1.1 GLOBAL MODELS

Two global models, ignoring geographical weighting for spatial heterogeneity, were fitted for comparison with the two final GWR models, as Table 6 shows.

Table 6 Final Global Models for the 1,000-foot (A-4) and 2,500-foot (B-5) Buffer Areas

Global Parameter	AIC	R^2	Adj. R^2	ß	Std Err	T
Global Model A-4						
Intercept	2504	0.506	0.500	19.68	1.159	16.98
Distance to downtown centroid				-0.0004	8.2E-05	-5.00
Number of all buildings				0.06	0.005	12.45
Number of educational buildings				0.728	0.139	5.25
Global Model B-5						
Intercept	2461	0.561	0.558	12.95	0.74	17.45
Number of intersections				0.088	0.03	2.90
Number of all buildings				0.013	0.002	6.63

Comparison of the fit parameters of each set of models (local (GWR) and global) reveals the extent to which the use of GWR benefited this project. 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, the 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 the AIC down from 2,461 to 2,441.

6 Conclusions

This study developed intersection-based GWR models to better understand how built-environment factors affect non-motorized travel. Strong spatial autocorrelations were confirmed in the count dataset, meaning that counts for intersections closer to one another were more likely to be similar. Models were fitted to each of two buffer areas to account for the differing influence of built environment on cycling and walking. GWR models were used to reflect the inherent spatial heterogeneity in the independent variables used to represent attributes of the built-environment.

For the 1,000-ft buffer area, the number of all buildings and number of educational buildings were significant, relating positively to non-motorized volumes. Distance from the Burlington urban area centroid was also significant with an increasing distance correlated to lower volumes. These results suggest that efforts to promote safe walking and bicycling routes to schools are an effective policy measure to promote these modes of travel, since these destinations are already strongly correlated with non-motorized travel.

For the 2,500-ft buffer area, total building density and intersection density were significant, relating positively to non-motorized volumes. These findings are consistent with previous studies that have concluded that non-motorized travel is more common where destinations are closer together and street connectivity is higher, typically in downtown urban-centers.

The strength of these relationships was estimated by investigating the parameters of the global regression model. Comparison of the fit parameters of each set of models (local (GWR) and global) reveals the extent to which the use of GWR benefited the study. Although the global model does not correct for the effects of spatial autocorrelation, it can be used to estimate the relative influence of each of the independent variables found to be significant. In this way, the global model parameters continue to help us better understand the nature of the relationships between non-motorized travel and the built environment. The results of the global regression model indicate that the largest magnitude of effect comes from the total number of educational buildings near the intersection and intersection density.

7 References

Akaike, Hirotogu. 1998. “Information Theory and an Extension of the Maximum Likelihood Principle.” In Selected Papers of Hirotugu Akaike, edited by Emanuel Parzen, Kunio Tanabe, and Genshiro Kitagawa, 199–213. New York, NY: Springer New York.

Brunsdon, Chris, A. Stewart Fotheringham, and Martin E. Charlton. 1996. “Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity.” Geographical Analysis 28 (4): 281–298.

Burnham, K., and D. Anderson. 2002. “Model Selection and Multi-Model Inference - A Practical Information-Theoretic Approach.” Springer-Verlag, New York, 2002.

Cervero, Robert, and Kara Kockelman. 1997. “Travel Demand and the 3Ds: Density, Diversity, and Design.” Transportation Research Part D: Transport and Environment 2 (3) (September): 199–219.

Cliff, A. D., and J. K. Ord, 1975. The choice of a test for spatial autocorrelation. In Davis, John C., North Atlantic Treaty Organization Scientific Affairs Division, Kansas Geological Survey, and University of Nottingham Dept of Geography. Display and Analysis of Spatial Data. Wiley., pp. 54-77, 1975.

Eom, Jin, Man Park, Tae-Young Heo, and Leta Huntsinger. 2006. “Improving the Prediction of Annual Average Daily Traffic for Nonfreeway Facilities by Applying a Spatial Statistical Method.” Transportation Research Record 1968: 20–29.

Fotheringham, A. Stewart, Chris Brunsdon, and Martin Charlton. 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. 1st ed. Wiley, New York.

Geary, R. C. 1954. “The Contiguity Ratio and Statistical Mapping.” The Incorporated Statistician 5 (3) (November 1): 115–146.

Lu, George X., James L. Sullivan, and Austin Troy, 2012. Impact of Ambient Built-Environment Attributes on Non-Motorized Travel: Geospatial Analysis of Chittenden County, Vermont. Prepared for presentation at the 91st Annual Meeting of the Transportation Research Board, Washington, D.C., January 22-26, 2012. TRB 91st Annual Meeting Compendium of Papers, No. 12-2561.

Moran, P. A. P. 1950. “Notes on Continuous Stochastic Phenomena.” Biometrika 37 (1/2) (June 1): 17–23.

O’Sullivan, David, and David Unwin. 2010. Geographic Information Analysis. 2nd ed. Wiley, 2010. pp. 180-205.

Owens, Peter M., Linda Titus-Ernstoff, Lucinda Gibson, Michael L. Beach, Sandy Beauregard, and Madeline A. Dalton. 2010. “Smart Density: a More Accurate Method of Measuring Rural Residential Density for Health-related Research.” International Journal of Health Geographics 9 (1) (December 1): 1–8.

Sokal, R. R. and N. L. Oden, and Barbara A. Thomson, 1998. “Local Spatial Autocorrelation in Biological Variables.” Biological Journal of the Linnean Society 65 (1) (September): 41–62.

Tobler, W. R. 1970. “A Computer Movie Simulating Urban Growth in the Detroit Region.” Economic Geography 46 (June 1): 234–240.

Vichiensan, Varameth, Antonio Páez, Kenji Kawai, and Kazuaki Miyamoto. 2006. “Nonstationary Spatial Interpolation Method for Urban Model Development.” Transportation Research Record: Journal of the Transportation Research Board 1977: 103–111.

VTrans, 2010. 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.

Zahran, Sammy, Samuel D. Brody, Praveen Maghelal, Andrew Prelog, and Michael Lacy. 2008. “Cycling and Walking: Explaining the Spatial Distribution of Healthy Modes of Transportation in the United States.” Transportation Research Part D: Transport and Environment 13 (7) (October).

Zhao, Fang, and Nokil Park. 2004. “Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic.” Transportation Research Record: Journal of the Transportation Research Board 1879: 99–107.