ROADWAY SAFETY INSTITUTE

Human-centered solutions to advanced roadway safety

Safety in Numbers: Pedestrian and Bicyclist Activity and Safety in Minneapolis

Kristin Carlson
Brendan Murphy
Alireza Ermagun
David Levinson
Andrew Owen

Department of Civil, Environmental, and Geo- Engineering
University of Minnesota

Final Report

REGION 5

CTS 18-05

University of Minnesota
Driven to Discover

The University of Akron

ILLINOIS
University of Illinois at Urbana-Champaign

Southern Illinois University Edwardsville

Western Michigan University

Technical Report Documentation Page

1. Report No.	2.	3. Recipients Accession No.
CTS 18-05		

4. Title and Subtitle	5. Report Date
Safety in Numbers: Pedestrian and Bicyclist Activity and Safety in Minneapolis	March 2018
	6.
7. Author(s)	**8. Performing Organization Report No.**
Kristin Carlson, Brendan Murphy, Alireza Ermagun, David Levinson, and Andrew Owen	
9. Performing Organization Name and Address	**10. Project/Task/Work Unit No.**
Department of Civil, Environmental, and Geo- Engineering University of Minnesota 500 Pillsbury Drive SE Minneapolis, MN 555455	CTS #2015038 11. Contract (C) or Grant (G) No. DTRT13-G-UTC35
12. Sponsoring Organization Name and Address	**13. Type of Report and Period Covered**
Roadway Safety Institute Center for Transportation Studies University of Minnesota 200 Transportation and Safety Building 511 Washington Ave. SE Minneapolis, MN 55455	Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

http://www.roadwaysafety.umn.edu/publications/

16. Abstract (Limit: 250 words)

This investigation aims to evaluate whether the Safety in Numbers phenomenon is observable in the midwestern U.S. city of Minneapolis, Minnesota. Safety in Numbers (SIN) refers to the phenomenon that pedestrian safety is positively correlated with increased pedestrian traffic in a given area. Walking and bicycling are increasingly becoming important transportation modes in modern cities. Proper placement of non-motorized facilities and improvements has implications for safety, accessibility, and mode choice, but proper information regarding estimated non-motorized traffic levels is needed to locate areas where investments can have the greatest impact. Assessment of collision risk between automobiles and non-motorized travelers offers a tool that can help inform investments to improve non-motorized traveler safety. Models of non-motorized crash risk typically require detailed historical multimodal crash and traffic volume data, but many cities do not have dense datasets of non-motorized transport flow levels. Methods of estimating pedestrian and bicycle behavior that do not rely heavily on high-resolution count data are applied in this study. Pedestrian and cyclist traffic counts, average automobile traffic, and crash data from the city of Minneapolis are used to build models of crash frequencies at the intersection level as a function of modal traffic inputs. These models determine whether the SIN effect is observable within the available datasets for pedestrians, cyclists, and cars, as well as determine specific locations within Minneapolis where non-motorized travelers experience elevated levels of risk of crashes with automobiles.

17. Document Analysis/Descriptors	18. Availability Statement
Walking, Pedestrians, Bicycling, Cyclists, Safety, Accessibility, Nonmotorized transportation, Pedestrian safety	No restrictions. Document available from: National Technical Information Services, Alexandria, Virginia 22312

19. Security Class (this report)	20. Security Class (this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	72	

Safety in Numbers: Pedestrian and Bicyclist Activity and Safety in Minneapolis

FINAL REPORT

Prepared by:

Kristin Carlson
Brendan Murphy
Alireza Ermagun
David Levinson
Andrew Owen
Department of Civil, Environmental, and Geo- Engineering
University of Minnesota

March 2018

Published by:

Roadway Safety Institute
Center for Transportation Studies
University of Minnesota
200 Transportation and Safety Building
511 Washington Ave. SE
Minneapolis, MN 55455

ACKNOWLEDGMENTS

The funding for this project was provided by the United States Department of Transportation's Office of the Assistant Secretary for Research and Technology for the Roadway Safety Institute, the University Transportation Center for USDOT Region 5 under the Moving Ahead for Progress in the 21st Century Act (MAP-21) federal transportation bill passed in 2012.

Contents

List of Figures

List of Tables

1 Introduction

This investigation aims to evaluate whether the Safety In Numbers phenomenon is observable in originally collected pedestrian and crash data in the midwestern U.S. city of Minneapolis, Minnesota. Safety In Numbers (SIN) refers to the phenomenon that pedestrian safety is positively correlated with increased pedestrian traffic in a given area, e.g., that the per-pedestrian risk of injurious interaction with motorized vehicles decreases as a function of the increasing flow of pedestrian traffic. SIN is well-supported by pedestrian crash data across a number of studies in various urban environments and reviews [18, 23, 3]. The concept has seen relatively widespread adoption in urban planning schools of thought, though its temporal causality is not clear-cut [3], and it is commonly discussed only in the context of pedestrian risk depending on pedestrian flow levels.

Walking and bicycling are increasingly becoming important transportation modes in modern cities, for a wide variety of reasons. Rates of walking and bicycling to work in the United States hover around 2.8% and 0.6%, respectively, with public transit use barely higher at 5% nationally [30]. Proper placement of non-motorized facilities and improvements has implications for safety [38], accessibility, and mode choice [17], but proper information regarding estimated non-motorized traffic levels is needed to locate areas where investments can have the greatest impact. Assessment of collision risk between automobiles and non-motorized travelers offers a powerful and informative tool in urban planning, and can help inform proper placement of improvements and treatment projects to improve non-motorized traveler safety.

Many of the issues with the collection of standardized non-motorized transportation data have to do with the factors that influence pedestrian and bicycle behavior. A model of non-motorized transport risk assessment is uninformative if the pedestrian and vehicular flows do not accurately represent real levels, and many cities do not have dense datasets of non-motorized transport flow levels, instead favoring counts of vehicle traffic. As such, active transport flow levels must be extrapolated from sparse datasets using comprehensive methodologies. For these reasons, as well as the overall lack in non-motorized travel counts for many communities, methods of estimating pedestrian and bicycle behavior that do not rely heavily on high-resolution count data are applied in this study.

Pedestrian and cyclist traffic counts, average automobile traffic, and crash data from the city of Minneapolis are used to build models of crash frequencies at the intersection level as a function of modal traffic inputs. These models determine whether the SIN effect is observable within the available datasets both pedestrians, cyclists, and cars, as well as determine specific spatial locations within Minneapolis where non-motorized travelers experience elevated levels of risk of crashes with automobiles, relative to intersections elsewhere in the city. The ability to identify specific unsafe locations based upon aggregated count and crash data offers an additional tool for city planners to implement in multimodal planning.

2 Pedestrian Activity Estimation

Walking and bicycling are increasingly becoming important transportation modes in modern cities, for reasons including individual and societal wellness, avoiding negative environmental externalities associated with motorized modes, and the costs and availability shortages sometimes associated with fossil fuels [21]. Planning for biking and walking, and creating societal programs to increase their levels, has been cited as a targeted health need in urban planning going forward [27, 33, 4]. Resource limitations, particularly in high-population and developing countries, impose constraints on the maximum level of personal motorized travel allowed, and as a result, there is a greater need for viable alternatives. In addressing the viability and availability of alternative modes, high-resolution spatial data on non-motorized transportation behavior patterns is needed.

Rates of walking and bicycling to work in the United States hover around 2.8% and 0.6%, respectively, with public transit use barely higher at 5% nationally [30]. Proper placement of pedestrian treatments and improvements has implications for safety [38], accessibility, and mode choice [17], but proper information regarding estimated non-motorized traffic levels is needed to locate areas in need of improvement. In determining salient locations for non-motorized improvements, it is important to have accurate records of both existing and potential travel demand (e.g., current levels of walking in a neighborhood, as well as good models of increased demand due to potential treatments); however, good quality, high-granularity datasets for non-motorized travel can be difficult to obtain, especially standardized for national spatial inventories [29]. For this reason, practitioners and researchers must frequently rely on estimation models for non-motorized traffic, and various methods can suffer from issues of data quality, granularity, and the presence of location-specific variables [26].

Many of the issues with the collection of standardized non-motorized transportation data have to do with the factors that influence pedestrian and bicycle behavior. A model of active transport risk assessment is uninformative if the pedestrian and vehicular flows do not accurately represent corresponding levels *in situ*, and many cities do not have dense datasets of active transport flow levels, instead favoring counts of vehicle traffic. As such, active transport flow levels must be extrapolated from sparse datasets using comprehensive methodologies. Land use data are well-documented by the U.S. Census Bureau to the Census Block level of resolution; general socioeconomic characteristics are maintained as well, and can have significant influence [36]. However, more specific socioeconomic characteristics are salient in non-motorized travel beyond just adjusted income levels, as well as weather variables [31] and latent, subjective variables such as visibility and perceptions of lighting, which can be more difficult to obtain at high spatial resolution [20], and can complicate inter-city comparisons. For these reasons, as well as the overall lack in non-motorized travel counts for many communities, methods of estimating pedestrian and bicycle behavior that do not rely heavily on high-resolution count data are applied in this study.

Aggregate travel behavior studies typically involve analysis at the level of Transportation Analysis Zones (TAZs), which are too coarse to allow robust analysis of non-motorized travel [36, 17]; regional travel surveys at the TAZ level consider many trip purposes but are similarly coarse and typically have too small of sample sizes to allow for robust city-to-city comparison. Census block-level information regarding economic accessibility (access to jobs) via both strictly walking and via the net accessibility benefit of public transportation will first be used to explain observed pedestrian traffic at a subset of intersections in the city of Minneapolis, Minnesota. Road network betweenness centrality will also be used as an explanatory variable, as a proxy of the underlying network structure. A framework for comprehensive pedestrian risk assessment modeling, using pedestrian volume, vehicle volume, and an environmental factor (crosswalk length) on a university campus is provided by Schneider et al. [38]. The motivation for constructing models of pedestrian traffic is in supplementing the sparse data currently available and deriving a reusable framework to provide a more complete picture of pedestrian activity throughout the city at the level of individual intersections, based on non-location-specific available data. This study aims to extensively use accessibility to jobs as its primary metric of modeling and estimating pedestrian activity in an urban area, supplemented by other variables which describe vehicular traffic and the built network environment.

2.1 Methodology

2.1.1 Data

This section briefly describes the data sources used in the pedestrian estimation models, and the data preparation process.

To serve the project's mapping and geometry needs, U.S. Census TIGER 2010 datasets for blocks and core-based statistical areas (CBSAs) were used, specifically for the Minneapolis-St. Paul, Minnesota region. U.S. Census Longitudinal Employer-Household Dynamics (LEHD) 2011 Origin-Destination Employment Statistics (LODES) data were used to quantify job accessibility and land use. A North America extract of OpenStreetMap (OSM), retrieved April 2014, was used as an input to custom-developed software for calculating job accessibility at the block level. Turning movement counts (TMC) for the years 2000-2013 in Minneapolis provided the targeted pedestrian count data, and GTFS data provided by Metro Transit allowed for use of the Minneapolis-St. Paul metropolitan region's transit operators' current schedules in performing the necessary job accessibility computations.

A pedestrian travel network graph for Minneapolis was then constructed, to begin the data preparation steps. The datasets for TMCs and AADT figures were not geocoded; this process was performed manually, linking the pedestrian and automobile count data to the intersections of interest. The count data were spread across multiple years, and were processed in PostgreSQL to obtain daily figures, averaged across the years for which count data were available. TMC data were processed by summing the total number of pedestrians passing through each intersection during the count day-long count periods. The count numbers used in the regression models were raw, and not upscaled to create 24-hour estimates.

For each Census block in Minneapolis, the travel time to all other blocks within a 3.1 mile radius (5 km) was calculated for a single departure time, as walking speed was taken to be constant and independent of the time of day. Using this travel time matrix, the cumulative opportunity accessibility to jobs was computed for each census block, using time thresholds of 5, 10, …, 30 minutes. Performing these same calculations with transit as the transportation mode instead of walking allows calculation of the "net transit" benefit experienced by transportation system users.

Betweenness centrality was calculated for the Minneapolis OSM road network, with a radius of 1 km (approximately 0.7 miles), to represent typical, short-length walking trips [46].

Once all data were collated and prepared, negative binomial regression was performed to describe pedestrian behavior with variables of walking accessibility, net transit accessibility, network betweenness centrality, and accessibility to job opportunities by sector. The time-thresholds for walking accessibility and time of day for the pedestrian counts to use in the modeling were determined with single-variable regression models, to compare variable explanatory power. The model formulation appropriateness was tested using a repeated k-fold cross-validation process, withholding random 20% test samples.

Regarding software used, intersection locations were determined from OSM road centerline data for the Minneapolis-St. Paul CBSA in QGIS. The subset of intersections for which count data were available is displayed in Figure 1; these intersections were used to construct the regression models. Accessibility calculations were performed using OpenTripPlanner (OTP) open-source routing software; GIS work performed in QGIS and PostGIS; network centrality measures computed in ArcMap GIS with the Urban Network Analysis Tools toolbox; statistical work done in SQL, Python, and R.

2.1.2 Accessibility

The first type of explanatory variable used in the model of Minneapolis pedestrian count data is cumulative opportunity accessibility. Using OTP software, walking travel times along the network are calculated from each Census block centroid in Minneapolis, to each other block centroid within the travel-time thresholds of 5, 10, …, 30 minutes. Travel times by transit were calculated using General Transit Feed Specification (GTFS) data in conjunction with the OTP software framework. GTFS data are included wherever feeds are made available by transit agencies, and the targeted service date for analysis was January 22nd, 2014, to reflect non-holiday, normal weekday service schedules. Job opportunities are summed from each block centroid reachable within a given time threshold, yielding an X-minute accessibility measure. Job opportunities are broken down by economic sector, as defined by the North American Industry Classification System. There are two accessibility calculations used in this study: accessibility to jobs from Census block centroids by walking, and accessibility to jobs from Census block centroids by both transit and walking.

The walking mode is included in transit accessibility calculations, to account for access to, and egress from, transit stations, as well as mid-trip transfers. Pedestrian counts are often taken at intersections in either gross counts, or divided by turning movement type. This study uses Turning Movement Count (TMC) data from approximately 750 intersections in Minneapolis; intersection counts were calculated by adding the various TMC types for each intersection in the analysis group, to yield a gross figure of pedestrian activity within an intersection. Two-hour counts for pedestrian activity were used for morning peak (7 a.m. – 9 a.m.), midday (11 a.m. – 1 p.m.), and evening peak (4 p.m. – 6 p.m.). Accessibility calculations were performed using the following formulation of a gravity-based model:

$$A_i = \sum_j O_j f\left(C_{ij}\right) \tag{1}$$

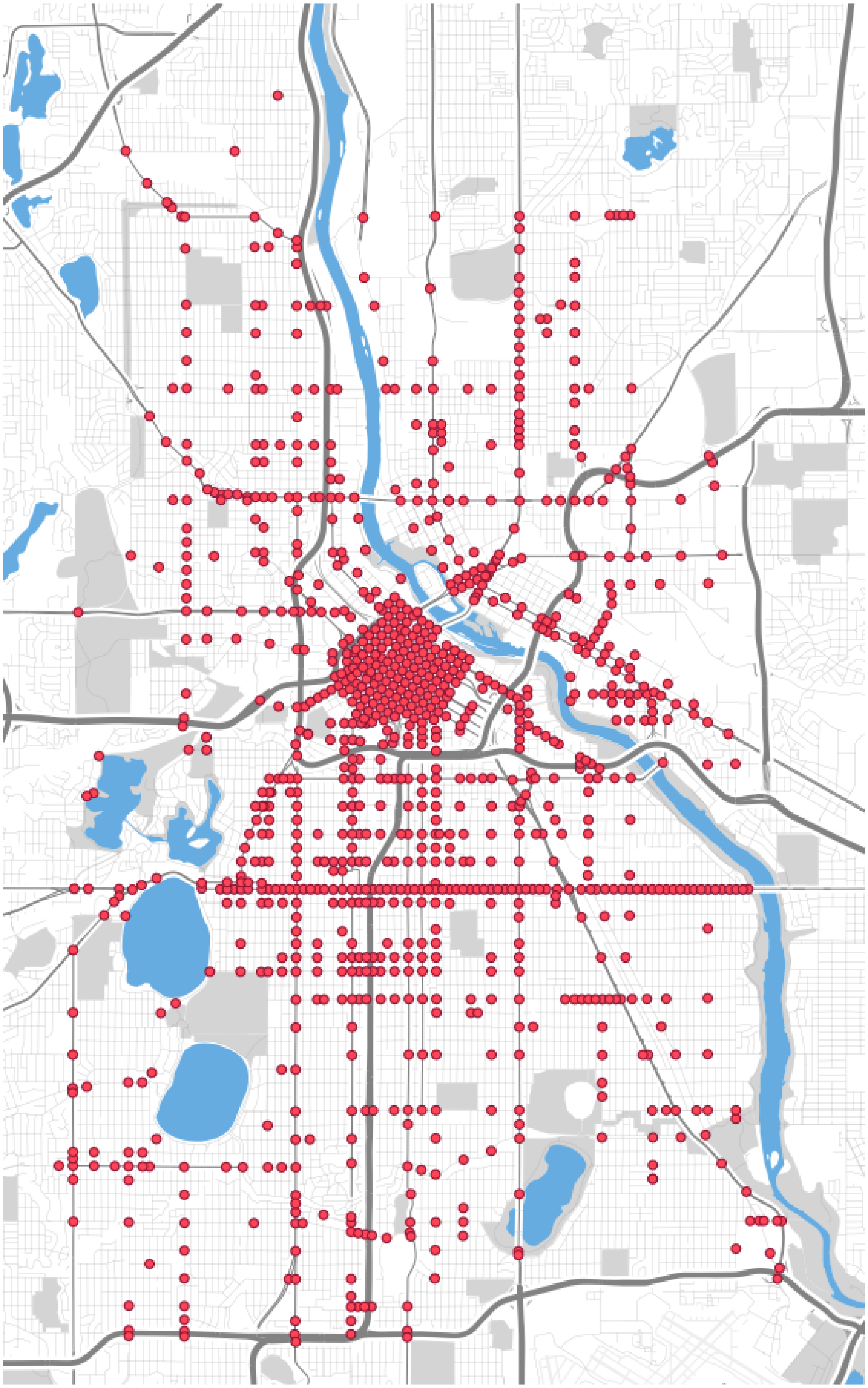

Figure 1: Locations of intersections in Minneapolis with raw pedestrian count data.

$$A_i = \text{accessibility for location } i \tag{2}$$
$$O_j = \text{number of opportunities at location } j \tag{3}$$
$$C_{ij} = \text{time cost of travel from } i \text{ to } j \tag{4}$$
$$f\left(C_{ij}\right) = \text{weighting function} \tag{5}$$
$$\tag{6}$$

The choice of weighting function has a large impact on the resulting Accessibility calculations; however, one of the simplest interpretations of cumulative opportunities is an integer count, using the following weighting function:

$$f\left(C_{ij}\right) = \begin{cases} 1 & \text{if } C_{ij} \leq t \\ 0 & \text{if } C_{ij} > t \end{cases} \tag{7}$$

$$t = \text{travel time threshold}$$

This intuitively makes sense when applied to opportunities such as jobs, number of restaurants, transit route departures, and other discrete integer variables in the surrounding environment, and the simplicity of counting destinations within on-network travel radii avoids complexities introduced with impedance variables such as demographics and differential monetary costs. It is predicted that origins exhibiting higher accessibility values would see greater pedestrian activity throughout the day. It should be noted that this formulation relies on a simplifying assumption: as all job opportunities within a Census block are assumed to be located at its centroid, intra-block walking trips, which are a real use-case, are not directly captured by this framework. However, if a person is traversing an intersection and thus would be counted within the Turning Movement Counts, it is likely they are moving from one Census block to another.

Accessibility for both walking, and walking + transit modes, are used in the estimation models; subtracting walking accessibility from the multimodal walking + transit accessibility yields the net transit benefit, and including walking and net transit separately in the regression models allows for explicit evaluation of how important transit benefits are to influencing pedestrian activity. Multiple regression was then performed to determine the explanatory power of the accessibility measures in predicting pedestrian and vehicular traffic in the AM, midday, PM peaks, as well as for a 6-hour summed count. These additional tables are omitted here. It was expected that origins exhibiting higher walking-accessibility values, and higher centrality values, would see greater pedestrian activity throughout the day.

Sector-specific significant factors were then added to the model, to examine which job sector categories may be more strongly correlated with walking as a commute mode. To determine which variables should be included, a stepwise negative binomial regression process was implemented: evening pedestrian count data were modeled with only accessibility data corresponding to the 20 LEHD job sectors as inputs. In each round of stepwise modeling, the least-significant sector factor was removed from the model; this process continued until only factors significant at the $p < 0.05$ level remained. These sector factors were then included in the parsimonious negative binomial model. A 5-minute threshold was chosen for these sector factors, to assess the ability of certain types of jobs to be walking-attractants in the more immediate walk-shed; this process was performed across the entire population of 788 intersections with pedestrian count data.

2.1.3 Centrality

In an attempt to reflect pedestrian activity on the underlying topology of the transportation network, a centrality measure was computed in ArcGIS with the Urban Network Analysis Toolbox, and added to the regression models. Various types of network measures of centrality have been proposed in their applicability to estimation of non-motorized activity levels [29, 2, 9], and safety and collision rates [47, 7]. One of the most common measures of centrality is "betweenness" centrality, or how "between" other nodes or links a given node or link is. When considering route choice and estimating modal traffic flows, link betweenness centrality is often considered, and consists of the proportion of shortest paths between all node pairs that pass through a link or node [28]. Relatedly, stress centrality consists of counting the number of times each link in a given network is utilized among the set of shortest paths between all node pairs, and is given by:

$$C_s(k) = \sum_{i,j \in V} \sigma_{ij}(k) \tag{8}$$

where σ_{ij} is either 1 if link k is used in shortest path σ_{ij}, and 0 otherwise. This form of stress centrality has been used to spatially assess transportation systems ([8]).

Instead of assessing centrality at the link level, this study adapts the above formulation to calculate betweenness centrality of intersections themselves. This allows a direct correspondence between the centrality metric and the pedestrian count data, which are intersection-based and not link-based. For the walking mode, it is not reasonable to include the entire set of road network intersections as possible destinations for a given intersection-origin, due to the lower speed of the walking mode - an assumed 5 km/h. Thus, for the centrality calculations for the walking mode, an on-network radius of 1 km, to represent 12 minutes of walking at average speed, was implemented to increase the saliency and relevance of centrality to actual walking behavior. Additionally, similar modifications to the above for bicycle modes may be implemented for walking, in particular modifying O/D frequency to reflect that a certain subset of nodal origins and destinations exhibit much higher activity levels than others; for simplicity, such modifications were not attempted in this study.

To reflect typical work trips, [29] chose O/D pairs such that origins were strictly residential parcels, and non-residential parcels were destinations in the morning, and the order was reversed in the evening. However, the authors speculated that allowing for non-residential destinations in the evening to reflect more complex after-work tours could increase model explanatory power [29]. Additionally, O/D pairs were limited by a network distance threshold of 5 miles, per the *National Household Travel Survey* [13]. O/D multipliers specified relative magnitude of trip generation, since parcels are heterogeneous in their trip generation capacity; these included density of dwelling units within residential parcels, and square footage density for all other parcels.

These modifications constitute potentially salient areas for further investigation in our model of pedestrian traffic. O/D pairs can be tailored to favor walking trips from residential parcels to commercial destinations, as well as modeled in a distribution based on trip length instead of using a single figure. Stress centrality is first used to evaluate preliminary explanatory power, and feasibility of applying centrality metrics to this model.

2.1.4 Pedestrian Activity Estimation

Multivariate negative binomial regression over the explanatory variables was performed in R for the walking mode. Different time-thresholds of accessibility were compared for explanatory power of pedestrian activity, of which the strongest threshold was chosen for a final parsimonious model to estimate pedestrian traffic throughout the sampled intersections. Iterative stepwise regression was performed using the economic sector accessibility variables, in an attempt to account for the possible differential walking trip generation levels of different job sectors. The parsimonious model is then applied to the entire sample of intersections within Minneapolis with count data, and the estimated pedestrian levels are compared to actual counts for comparison. The model formulation used is as follows:

$$ln(Z) = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \ldots + \beta_m * x_m \tag{9}$$

Z is the pedestrian count, $x_1, x_2, \ldots, x_m$ are known predictive variables, and $\beta_0, \beta_1, \beta_2, \ldots, \beta_m$ are coefficients to be estimated. Use of the negative binomial model for non-motorized traffic count data is well-supported in the literature [37, 16, 38].

A k-fold cross-validation process was performed to assess the robustness of the model and its sensitivity to input sampling. Specific spatial areas of underestimation and overestimation are discussed.

2.2 Results

Full tabulation of all single-variable regression models, to determine which time thresholds and peak-hour periods to use for greatest explanatory power in modeling pedestrian traffic levels, are omitted for brevity. It was found that the 15-minute threshold of total accessibility, combined with the PM-peak period pedestrian counts and other variables, yielded the best explanatory power for walking accessibility. A parsimonious model for walking activity, in terms of the strongest explanatory variables, is reported in Table 3. Net transit accessibility benefit was included as an explanatory variable in the pedestrian activity estimation model, to account for the effect of transit in urban cores of increasing pedestrian activity by attracting additional users beyond pure foot traffic. Table 1 lists summary statistics for the pedestrian count datasets used in the following analysis—pedestrian turning movement counts between 2000 and 2013 for Minneapolis.

A total of 788 intersections were included in the study, distributed across the entire city, and correspond to locations where pedestrian count data were available. This population was split into two subpopulations of intersections—those at which AADT data were available, and those at which AADT data were not available. This yielded two separate, parallel sampled populations and models, and allowed for comparison between the two models to determine what, if any, explanatory power AADT information offered for pedestrian activity. Each subpopulation was modeled across

Table 1: Pedestrian Activity Dataset Summary Statistics

Description	Value
Intersections with evening ped counts	788
Intersections with AADT counts	470
Intersections without AADT counts	318
80% training sample size for AADT models	376
80% training sample size for non-AADT models	256
20% testing sample size for AADT models	94
20% testing sample size for non-AADT models	64
Average & standard deviation 6-hour peds per day, AADT	812.99; 1825.90
Average & standard deviation 6-hour peds per day, no AADT	515.23; 1160.458
Maximum 6-hour peds, AADT	14, 793
Maximum 6-hour peds, no AADT	11, 470
Average & standard deviation morning peds per day, AADT	209.43; 466.99
Average & standard deviation morning peds per day , no AADT	123.29; 239.94
Maximum morning peds, AADT	3, 968
Maximum morning peds, no AADT	1, 856
Average & standard deviation midday peds per day, AADT	291.20; 751.86
Average & standard deviation midday peds per day, no AADT	182.08; 523.59
Maximum midday peds, AADT	6, 057
Maximum midday peds, no AADT	6, 266
Average & standard deviation evening peds per day , AADT	306.10; 636.26
Average & standard deviation evening peds per day, no AADT	208.32; 435.74
Maximum evening peds, AADT	4, 951
Maximum evening peds, no AADT	3, 577

Note: Summary statistics for datasets used in pedestrian activity analysis: pedestrian turning movements between 2000 and 2013 for the City of Minneapolis.

the entire sample for the parsimonious negative binomial models listed in Table 3, and then independently sampled for k-fold cross-validation at an 80% training level. Table 1 also gives the specific sample sizes for the cross-validation procedure, as well as basic statistics of the pedestrian count dataset, broken down by time of day and by model.

First, the pedestrian counts were modeled in terms of only walking accessibility, for different thresholds and times of day. From this, the strongest explanatory power was determined for PM peak period counts, at a 15-minute accessibility threshold. Pedestrian counts were then modeled in terms of transit & walking accessibility (bimodal accessibility), for different times of day. A 30-minute transit threshold was used, in accordance with the reported data available in the Access Across America: Transit 2014 report [32]. Net transit accessibility, a measure which looks at the contribution to accessibility from transit service, was also investigated as a potential explanatory variable for walking activity. A threshold of 30-minutes was again used. Betweenness stress centrality was included to relate walking activity to the underlying network structure. Accessibility and betweenness centrality are mapped in Figure 2 and Figure 3, respectively.

Table 2 displays the most significant factors obtained in the stepwise negative binomial regression process, to determine which LEHD sector job categories are more strongly correlated with walking behavior in their immediate walk-sheds. The categories of jobs found to positively correlate with walking activity were Finance, Education, and Arts & Entertainment; the categories of jobs found to negatively correlate with walking activity were Transportation, Real Estate, Administrative, and Hospitality & Food.

Table 2: LEHD job sector stepwise negative binomial regression results: remaining significant factors at the 5-minute travel time threshold.

Explanatory variable	Coefficient estimate
Transportation jobs 5min	$-0.0141^{**}(0.0043)$
Finance jobs 5min	$0.0081^{***}(0.0023)$
Real Estate jobs 5min	$-0.0037^{*}(0.0016)$
Administrative jobs 5min	$-0.0105^{***}(0.0013)$
Education jobs 5min	$0.0107^{***}(0.0010)$
Arts & Entertainment jobs 5min	$0.0146^{**}(0.0047)$
Hospitality & Food jobs 5min	$-0.0090^{***}(0.0010)$
Constant	$5.4026^{***}(0.0556)$
Observations	782
Log Likelihood	-4,893.324
θ	0.579^{***} (0.026)
Akaike Inf. Crit.	9,802.65
Note:	*p<0.1; **p<0.05; ***p<0.01

Regression results for the two negative binomial models for walking activity, with and without AADT included, are in Table 3. For the model which included AADT data, the significant explanatory factors were: accessibility by walking, net transit benefit to accessibility, betweenness centrality, AADT, and accessibility to finance, education, and hospitality jobs. Of these, the aggregate accessibility factors were positively correlated with walking activity, as were AADT and accessibility to finance and education jobs; betweenness and accessibility to hospitality jobs were negative correlates, though betweenness was less strongly correlated at the $p < 0.1$ level. For the model of intersections at which AADT data were not available, fewer significant factors emerged: accessibility by walking, and accessibility to administrative, eduction, and hospitality jobs. Of these, walking accessibility and accessibility to education jobs were strongly positively correlated with pedestrian activity; accessibility to administrative and hospitality jobs were each negatively correlated with pedestrian activity. In this model, betweenness centrality was not found to be a significant predictor, but was weakly positively correlated with walking.

A series of maps shows additional views of the data used in the modeling process; Figure 2 shows accessibility to jobs within 30 minutes by walking in Minneapolis, and Figure 3 shows the normalized betweenness centrality of all intersections in Minneapolis calculated with a 1 km radius. Accessibility by walking, given the walking mode's uniform nature, shows where economic activity is most concentrated in the region. Centrality gives a sense of the most important nodes in the street network of Minneapolis — that is, the nodes that would affect the highest number of shortest paths between origins and destinations, were they to be rendered impassible. Walking accessibility showed a positive correlation with pedestrian activity in both models, as shown in Table 3. Figure 4 shows the raw levels of daily pedestrian activity, aggregated from manual pedestrian counts between 2000 and 2013, while Figure 5 shows the estimated levels of evening peak pedestrian activity in Minneapolis at intersections which had AADT data, calculated using the coefficients outlined in Table 3 for the AADT model. Figure 6 shows estimated pedestrian activity at intersections without AADT data, calculated using the coefficients outlined in Table 3 for the no-AADT model.

To examine the accuracy of the estimated model, the differences between actual and estimated pedestrian activity, for the AADT and no-AADT models, are mapped in Figure 7 and Figure 8, respectively. Additionally, example spatial distributions of jobs in categories of Finance and Education are shown in Figure 9 and Figure 10, respectively.

Table 3: Parsimonious Negative Binomial Pedestrian Model Regression Results: With & Without AADT

Explanatory variable:	*Dependent variable:*	
	Average PM pedestrians	
	With AADT	Without AADT
Walking accessibility (15-minute)	$7.027e-06^{**}$	$3.312e-05^{***}$
	$(2.536e-06)$	$(4.057e-06)$
Net transit accessibility (30-minute)	$1.140e-05^{***}$	$2.418e-06$
	$(1.388e-06)$	$(1.642e-06)$
Betweenness	$-1.025e-05^{*}$	$9.414e-06^{\cdot}$
	$(4.172e-06)$	$(5.085e-06)$
AADT	$2.746e-05^{**}$	
	$(9.930e-06)$	
Transportation jobs 5min	$-8.833e-03^{\cdot}$	$-1.117e-02^{\cdot}$
	$(5.233e-03)$	$(6.074e-03)$
Finance jobs 5min	$6.656e-03^{**}$	$9.389e-04$
	$(2.507e-03)$	$(3.568e-03)$
Real estate jobs 5min	$-9.182e-04$	$-1.554e-03$
	$(2.488e-03)$	$(1.917e-03)$
Administrative jobs 5min	$-4.962e-04$	$-1.088e-02^{***}$
	$(1.702e-03)$	$(1.720e-03)$
Education jobs 5min	$9.140e-03^{***}$	$6.664e-03^{***}$
	$(1.114e-03)$	$(1.715e-03)$
Arts entertainment jobs 5min	$1.096e-02^{*}$	$1.072e-02$
	$(5.575e-03)$	$(7.009e-03)$
Hospitality food jobs 5min	$-7.644e-03^{***}$	$-5.282e-03^{**}$
	$(1.162e-03)$	$(1.665e-03)$
Constant	4.054^{***}	4.469^{***}
	$(1.564e-01)$	$(1.521e-01)$
Observations	452	303
Log Likelihood	-2,857.964	-1,781.407
θ	0.719*** (0.044)	0.704*** (0.054)
Akaike Inf. Crit.	5,739.927	3,584.814
F Statistic	23.970*** (df = 8; 477)	42.139*** (df = 7; 1008)
Note:	*p<0.1; **p<0.05; ***p<0.01; (standard error)	

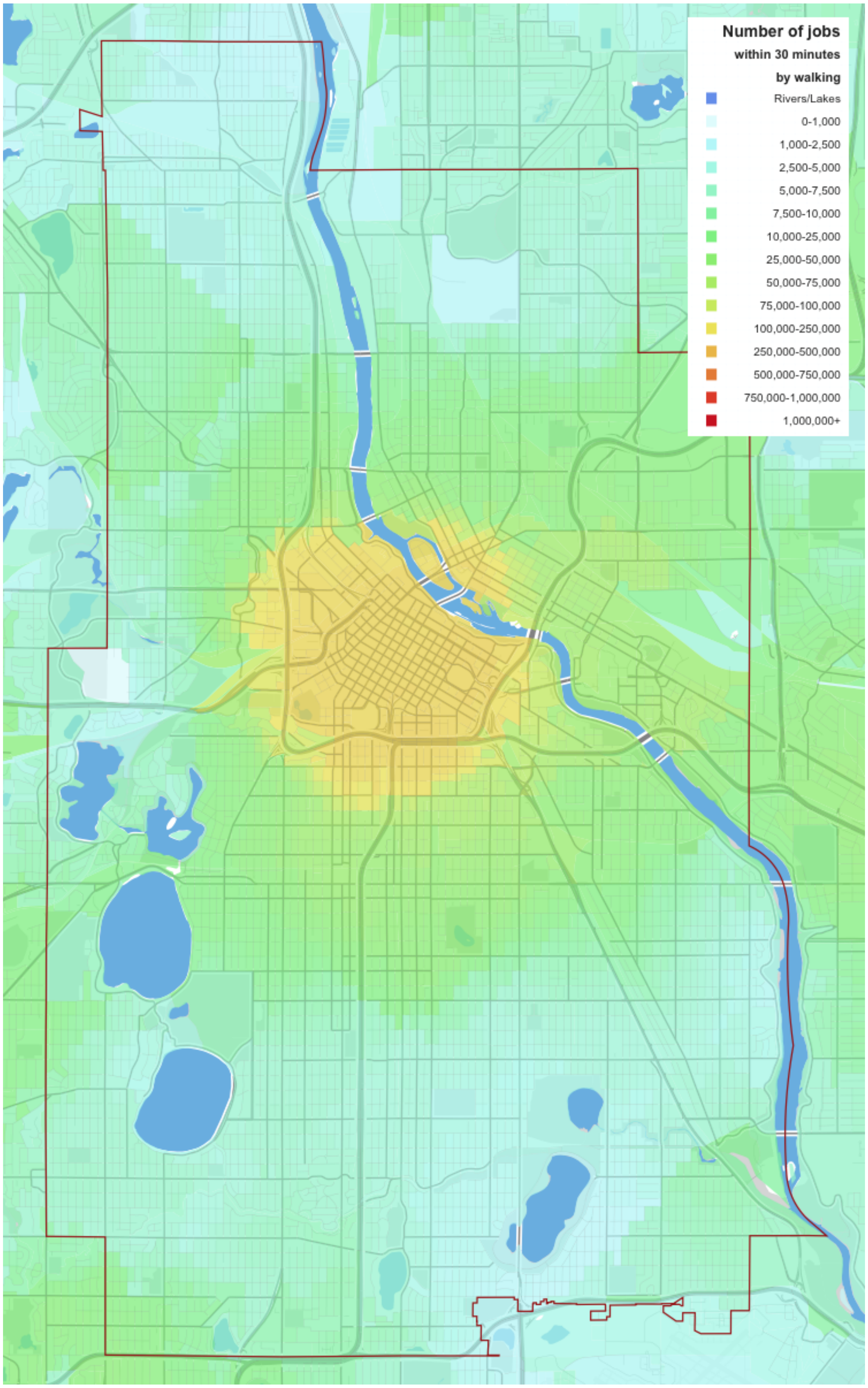

Figure 2: Accessibility to jobs within 30 minutes by walking in Minneapolis.

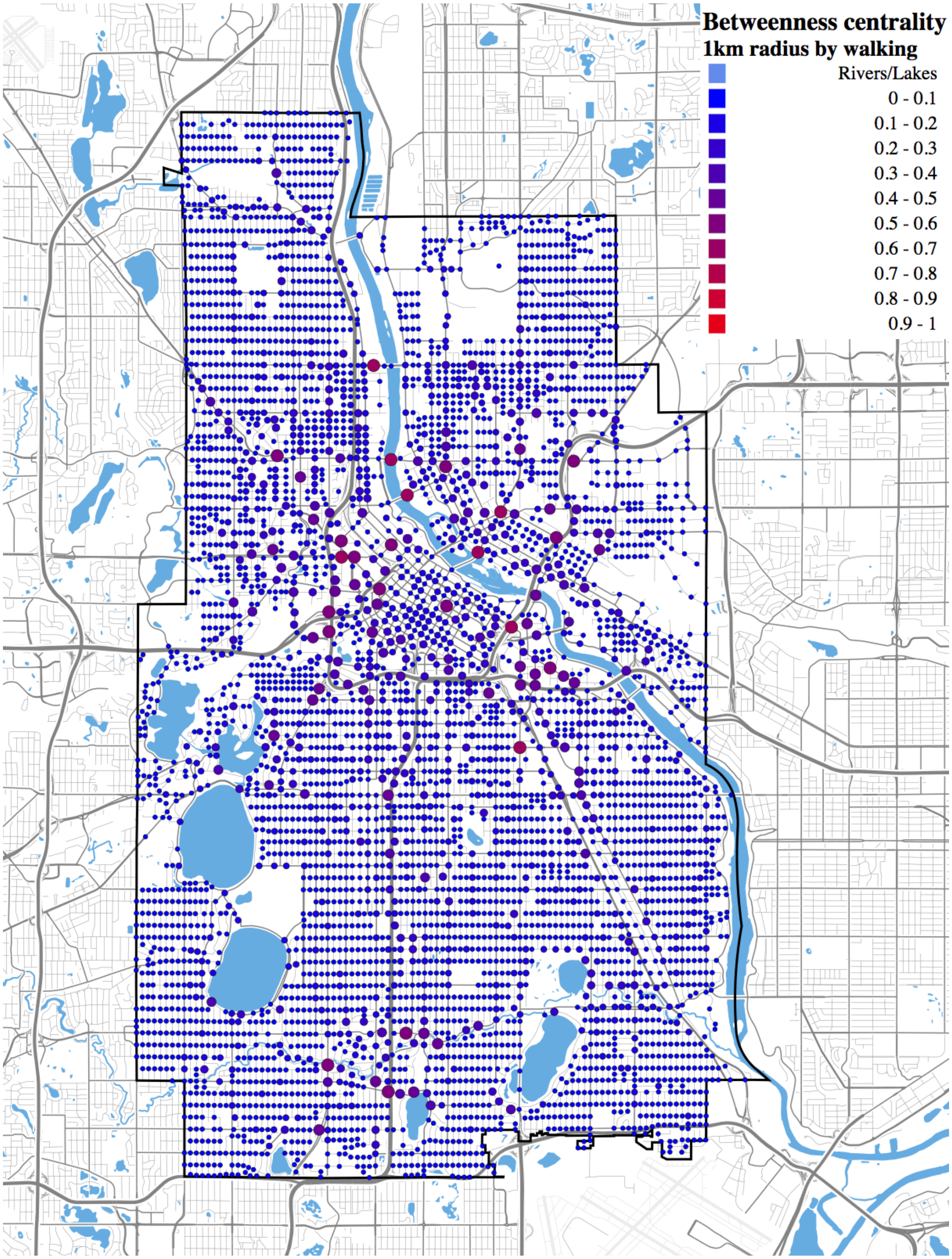

Figure 3: Normalized betweenness centrality of all intersections in Minneapolis; radius of 1km.

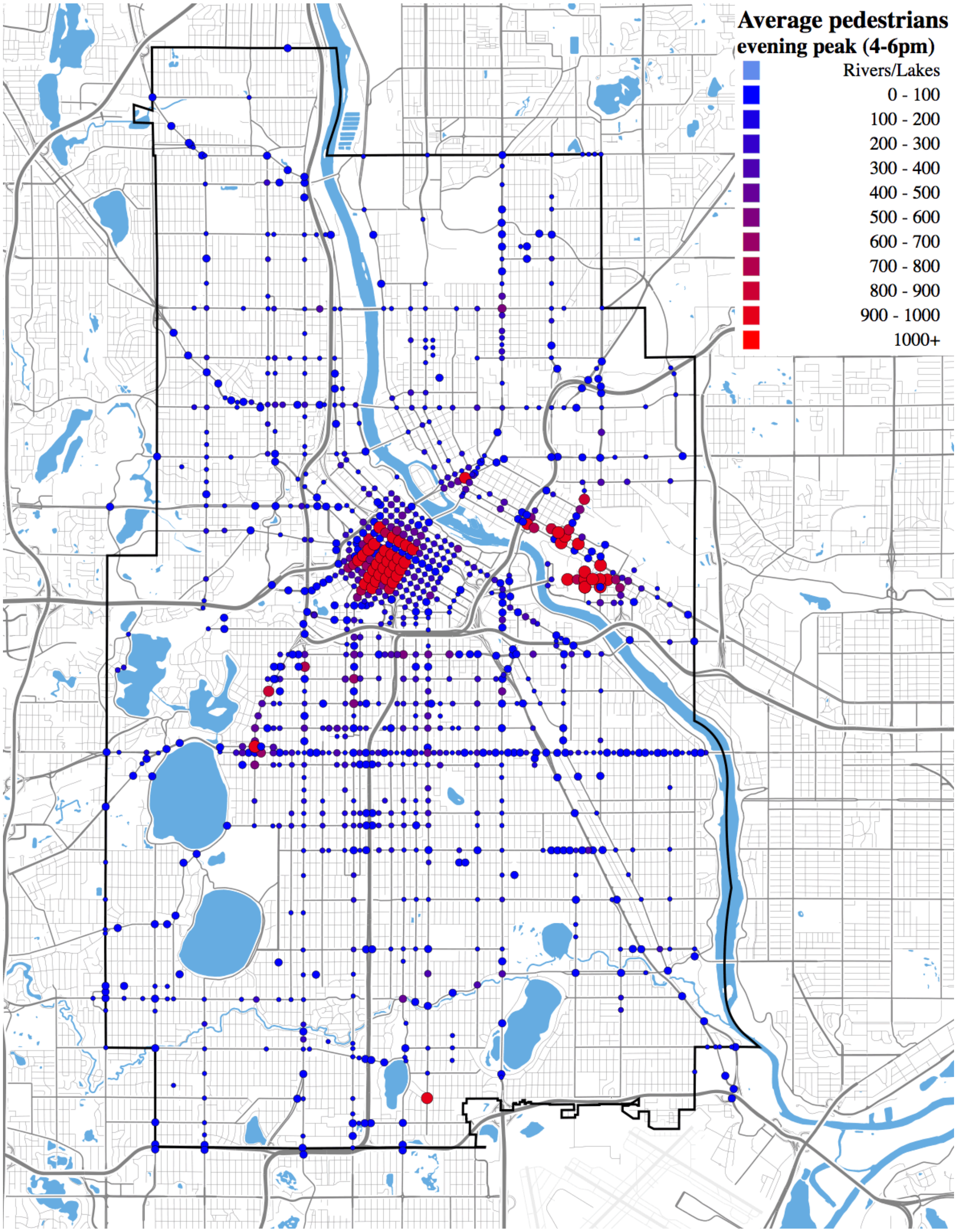

Figure 4: Raw levels of evening peak (4-6pm) pedestrian activity in Minneapolis, 2000-2013.

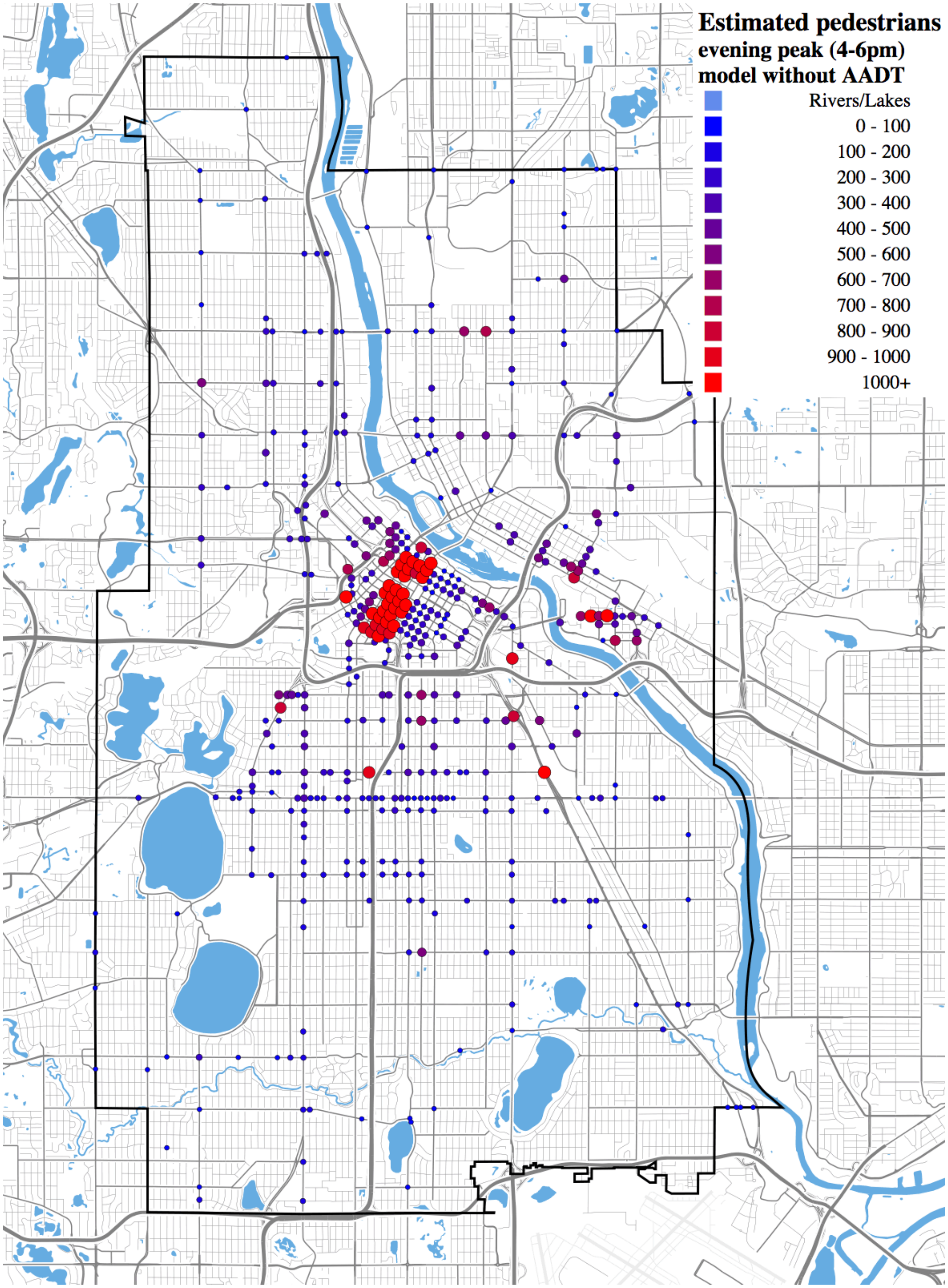

Figure 5: Estimated levels of evening peak pedestrian activity in Minneapolis, at intersections included in the AADT model.

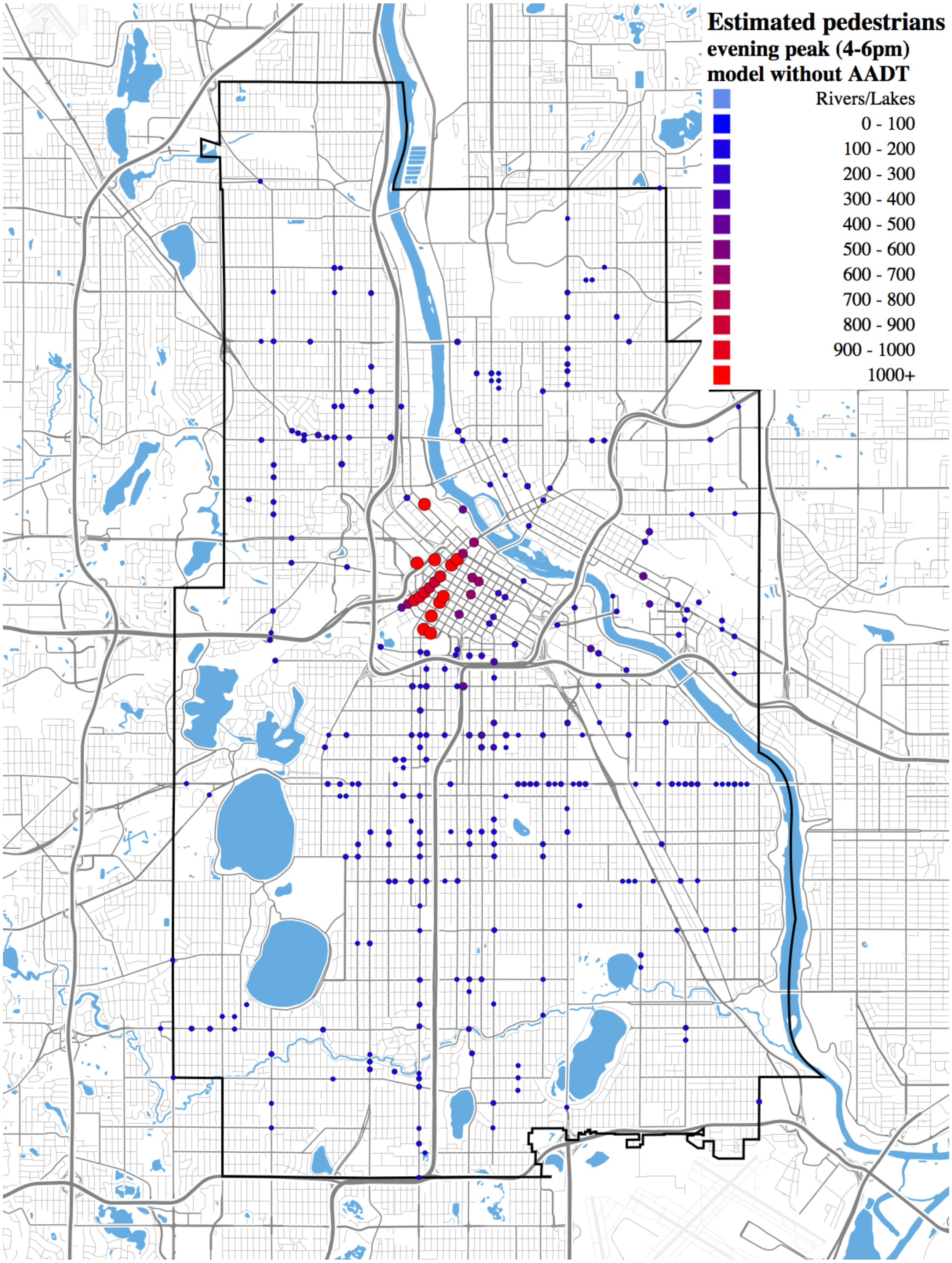

Figure 6: Estimated levels of evening peak pedestrian activity in Minneapolis, at intersections included in the no-AADT model.

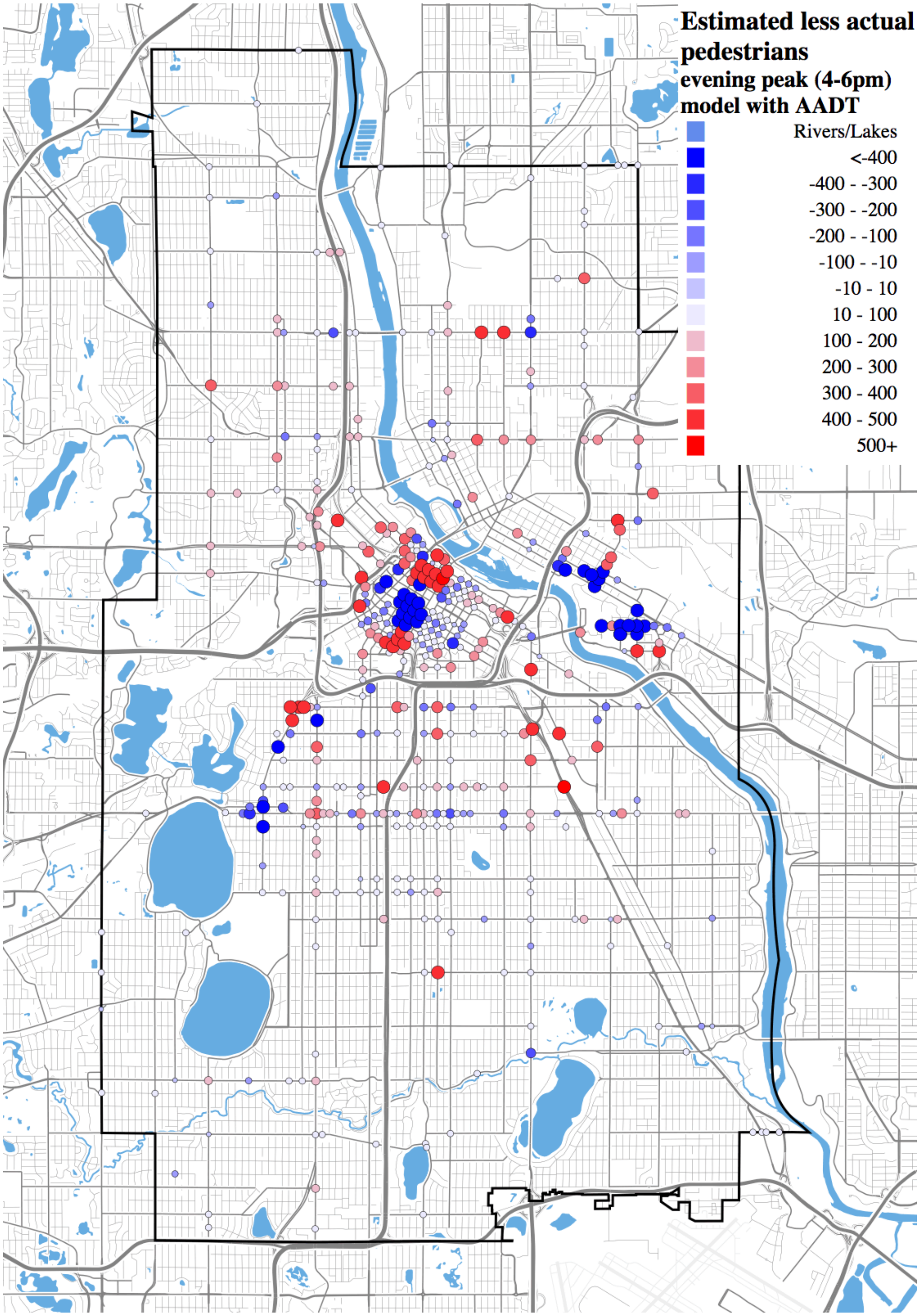

Figure 7: Estimated minus actual pedestrian activity, PM peak period, at intersections included in the AADT model. Reds are areas of overestimation; blues are areas of underestimation.

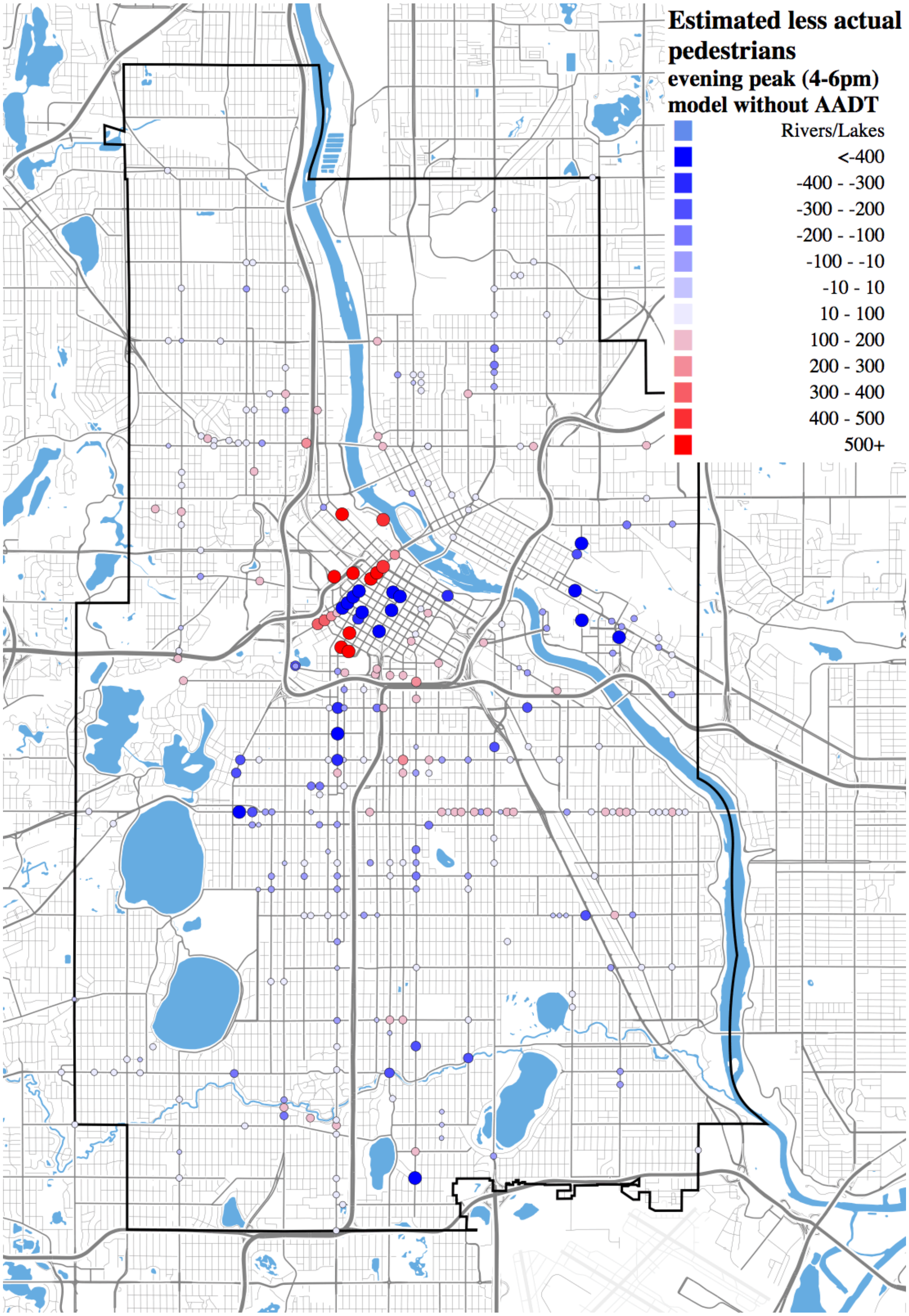

Figure 8: Estimated minus actual pedestrian activity, PM peak period, at intersections included in the no-AADT model. Reds are areas of overestimation; blues are areas of underestimation.

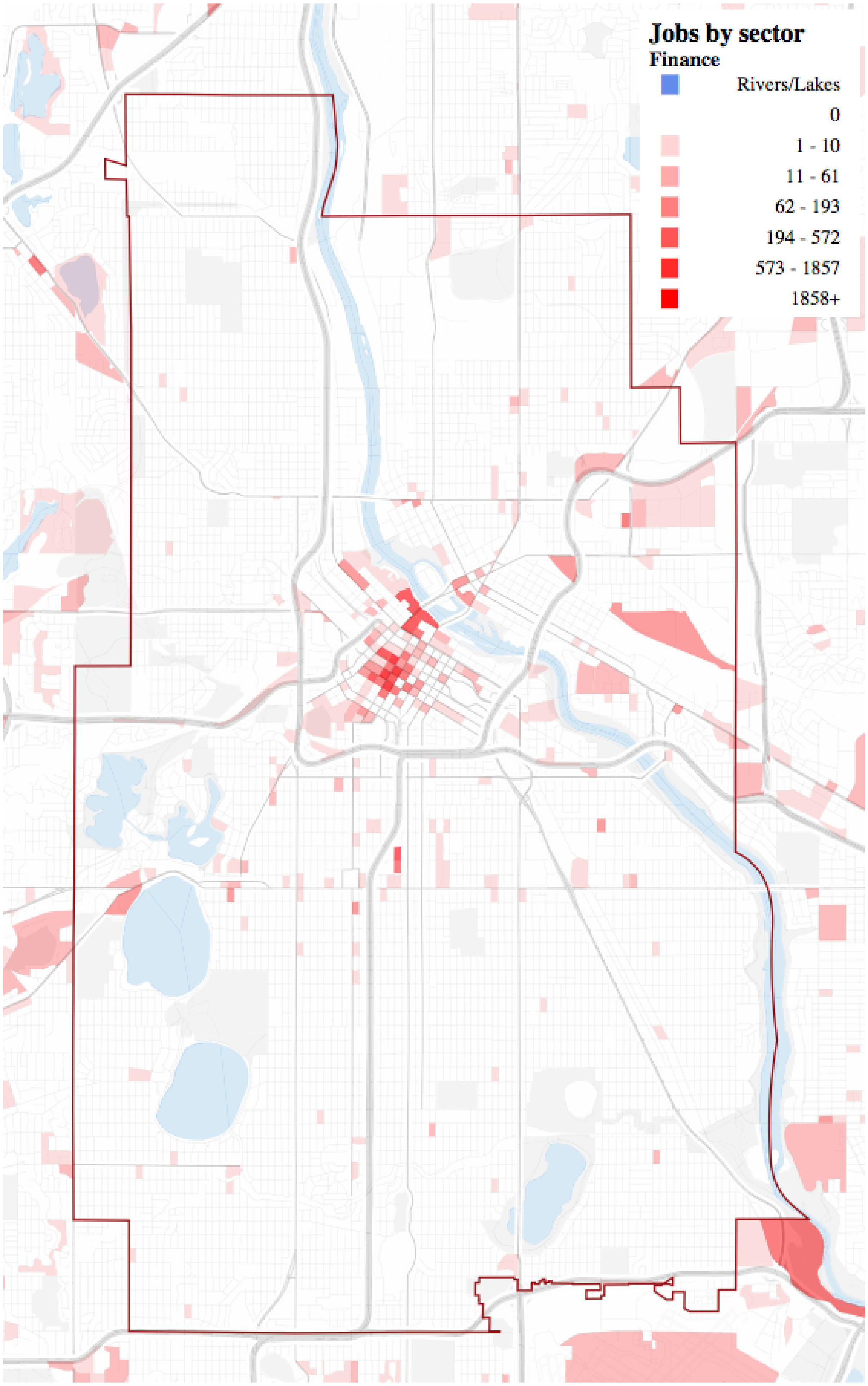

Figure 9: Spatial distribution of finance jobs in Minneapolis, based on LEHD data.

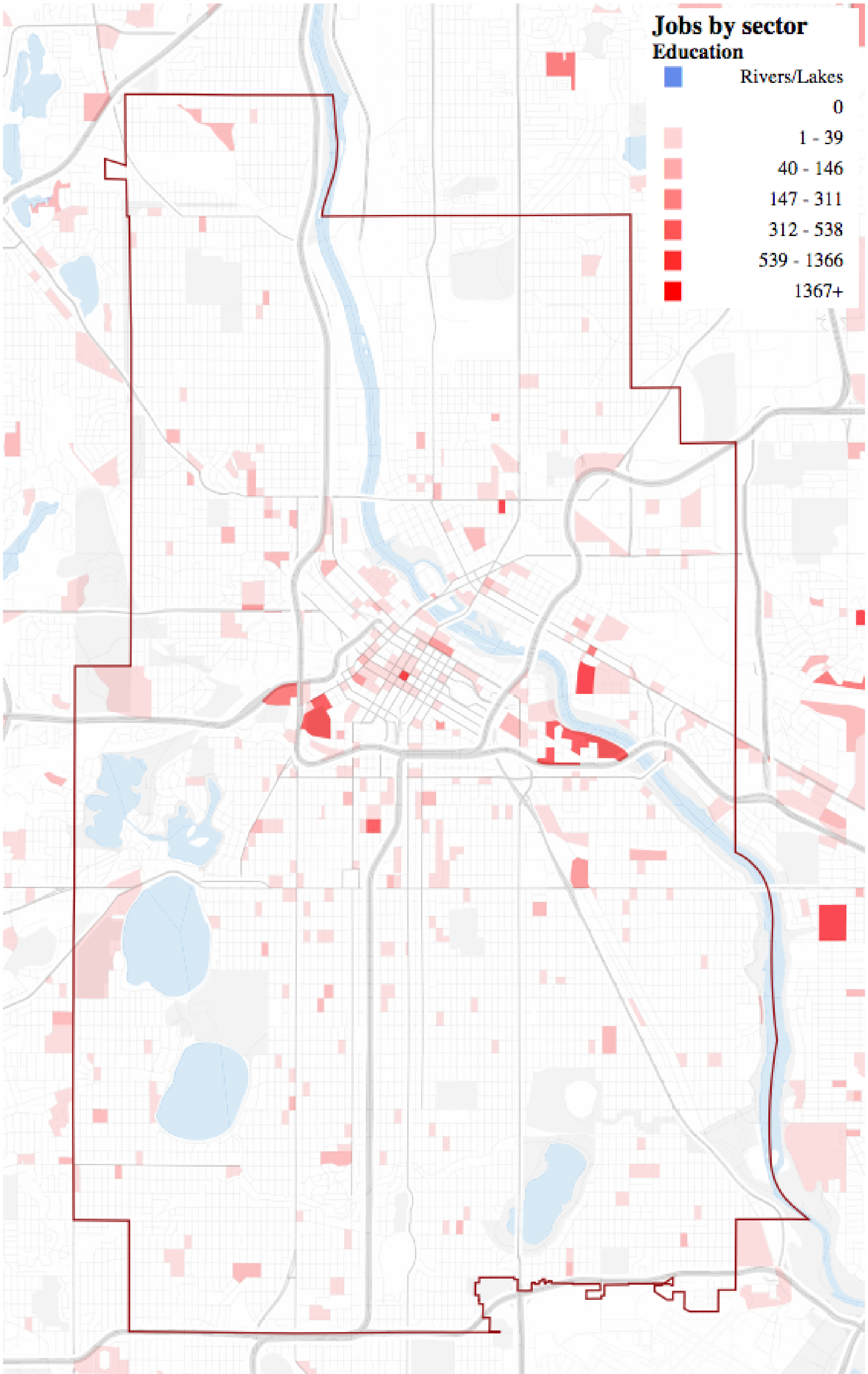

Figure 10: Spatial distribution of education jobs in Minneapolis, based on LEHD data.

2.2.1 Data Validation

A repeated k-fold cross validation process was used to test the robustness of the model formulation and its ability to predict pedestrian traffic at intersections not included in the modeling. The full dataset used previously was split in an 80% training, 20% testing regime, as referenced in the dataset statistics in Table 1. The dataset was divided into 5 folds 10 separate times, and the mean standard error and standard deviation metrics were computed for the results compared to the testing subset, to assess the predictive accuracy for both models using AADT data and without using AADT data. These metrics, along with the Grand Mean and Pooled Variance across all validation trials, are listed in Table 4.

Table 4: K-fold cross-validation results: mean standard errors and standard deviations across 10 trials of 80% training, 20% testing sample split.

Trial number	MSE (AADT)	MSE (no AADT)	SD (AADT)	SD (no AADT)
1	549.93	961.48	117.63	728.89
2	514.12	803.85	248.29	389.76
3	708.68	766.85	354.83	245.79
4	525.83	781.60	213.11	428.40
5	549.62	830.44	123.46	349.28
6	543.69	860.92	130.39	476.30
7	552.63	768.29	125.24	287.04
8	557.71	826.65	94.92	113.69
9	563.99	679.43	102.90	421.67
10	549.71	906.26	120.21	808.95
Grand Mean (AADT)				561.59
Grand Mean (no AADT)				818.58
Pooled Standard Deviation (AADT)				181.32
Pooled Standard Deviation (no AADT)				469.42

As listed in Table 1, the average and standard deviation for evening pedestrian counts for AADT intersections were 306.10 and 636.26, respectively; these figures for the non-AADT intersection population were 208.32 and 435.74, respectively. The Grand Mean Standard Errors for the k-fold cross-validation process, after 10 folds, as percentages of the average PM pedestrian counts, were 183.47% for the AADT population, and 392.94% for the non-AADT population. In both population cases, the MSE of prediction in the k-fold cross-validation process was larger than the actual average evening pedestrian activity, with this gap significantly larger for the non-AADT intersection population. The Pooled Standard Deviations for the cross-validation process, as percentages of the actual population standard deviations, were 28.50% for the AADT population, and 107.73%.

2.3 Discussion & Conclusion

For the single-variable models of pedestrian activity in terms of census block centroid accessibility to jobs via walking, the evening peak period provided the best explanatory power. For all three time periods, as well as the 6-hour total count, log-likelihood values peaked near 15-minute thresholds, and dropped off in either direction. The correlation between walking accessibility and walking activity is positive. Walking is commonly thought of as a 15-minute-mode, in that the majority of people walking in urban areas will be on trips of duration 15 minutes or less. Further, in dense urban areas, distance matters - a high-threshold measurement of walking accessibility will tend to blur the results and differences between origin points, thus potentially failing to reflect local variabilities in walking patterns. Additionally, aggregate accessibility data at the 5-minute threshold level was found to be a consistently less significant predictor of pedestrian activity than higher thresholds. And while other, non-work walking trips do occur throughout the day, the choice of an evening peak commute time for analysis and modeling links the work-based accessibility metrics and the walking count data.

It was found that pedestrian counts in the evenings exhibited the strongest correlations with the accessibility variables tested, and midday counts exhibited the weakest correlation strengths. It is possible that midday pedestrian traffic is more dispersed in both nature of trip-making and timing, due to variable work schedules. Both the morning and evening periods exhibited stronger correlations with job-based accessibility metrics, in accordance with traditional work commute timings. The subtle difference between the two periods could be explained in part through analysis of individual trip diaries - specifically the distributions of departure and arrival times for morning and evening trips.

As was hypothesized, the aggregate walking and net transit accessibility measures exhibited positive correlations with pedestrian activity levels for both AADT and non-AADT models, and walking accessibility was a significant correlate in both models. However, betweenness at a radius of 1 km was found to be a negative correlate and weakly significant in the AADT model, and positive but not significant in the non-AADT model. Its influence on pedestrian activity in the model was overshadowed by that of other factors, and perhaps was too positively correlated with AADT in the first model for both factors to exhibit positive correlation with walking. However, the accessibility metrics give a reasonable framework through which to estimate modal traffic levels at every intersection in Minneapolis and, by extension of the broader framework, in other cities as well. Betweenness centrality did not exhibit as strong a positive correlation as was predicted. This may have resulted from the specific methodology used - that is, a centrality calculation that takes into account heterogeneous trip generation within an urban area due to varying land use patterns may lead to higher predictive power of centrality measures toward actual pedestrian behavior patterns. Pedestrian behavior in urban areas does not exhibit uniform all-to-all trip generation distribution; rather, there are major sources and attractors, which would shift the distribution of route choices, and thus link and intersection centrality, to favor routes between those origin-destination pairs. Applying techniques analogous to those in [29] to the walking model may yield more accurate pedestrian behavior estimation based on the centrality metric.

Accessibility to Education and Finance jobs was found to be significantly predictive of increased pedestrian activity in the first model, while accessibility to Hospitality & food jobs was found to be significantly predictive of decreased pedestrian activity, relative to other categories. Only accessibility to Education jobs was found to be a significant positive predictor in the second model, while accessibility to Administrative and Hospitality & food jobs were found to be significant negative predictors. Spatial maps of the distributions of Finance and Education jobs are shown in Figure 9 and Figure 10, respectively. Finance jobs are heavily concentrated in the downtown core area, and education jobs are concentrated on walkable campuses. Further, it is plausible that certain categories of jobs attract greater or lesser levels of walking among their workers, dependent on such factors as dress requirements, vehicle needs (e.g. construction and contract workers), and typical density of jobs within each category. Additional cross-comparison analysis among economic job categories is needed to investigate these effects, but initial analysis indicates these spatial distributions correlate to the regression coefficients in Table 2.

A significant and pervasive challenge with analysis dependent on pedestrian, bicycle, and vehicular count data is the issue of data quality and format. Methodologies and data standards can vary from city to city and jurisdiction to jurisdiction; this study used a combination of national (Census, LEHD) datasets and local (Minneapolis traffic) data. Some cities, such as Boston, do not have robust pedestrian and bicycle counting programs throughout the city; others, such as Philadelphia, may have varying data release and non-disclosure agreements between MPOs, cities, and police departments; still other cities may have inconsistent data tracking and release practices, such as Washington, D.C. Such hurdles can make the collection and processing of pedestrian and bicycle spatial use data on a national scale exceedingly difficult. Better standards of practice in data collection, management, and distribution are needed.

However, with pedestrian activity estimation based on sampling existing counts, accessibility analysis, and betweenness centrality of the underlying network, it becomes possible to predict the landscape of pedestrian activity within the urban area. Such techniques may prove important in informing urban planning processes and decisions, pedestrian safety programs, and highlighting areas of the city that experience higher pedestrian activity as salient areas for fine-grained attention to built environment details. An important extension of the identification of intersections with higher potential pedestrian traffic is the visualization of such areas - e.g. downtown. It is reasonable to expect certain levels of pedestrian traffic, even where counts may not exist.

Data validation was performed using a k-fold cross-validation process, with 10 trials of 5 folds each. The AADT model showed both a lower grand MSE and lower pooled pooled variance across all 10 trials than the non-AADT model, indicating that the model which included AADT as a variable was more reliable and less inaccurate in its predictions of pedestrian activity. The loss of AADT information produces more variability in the predictions, indicating that AADT offers some amount of explanatory power for pedestrian activity—indeed this is the case, as shown in Table 3. Generally within a downtown area this makes sense—traffic across modes would be positively correlated.

There are a few caveats to mention regarding the ability of simply accessibility and centrality to accurately predict pedestrian behavior. Figure 7 and Figure 8 highlight sections of the urban area where the model differed significantly from the actual pedestrian counts in each model. Within the AADT model, pedestrian activity at 306 intersections was overpredicted, and activity at 146 intersections was underpredicted. Within the non-AADT model, 199 intersections were overpredicted, and 104 intersections were underpredicted. The distributions of prediction differences for AADT and non-AADT models had means $\mu = -10.36$ and $\mu = 65.25$, and standard deviations $\sigma = 548.65$ and $\sigma = 706.18$, respectively. 94.25% of the sampled intersections had $estimated - actual$ differences within 1 standard deviation from the mean for the AADT model; 94.06% of sampled intersections had prediction differences within 1 standard deviation from the mean for the non-AADT model. The cases of underestimation and overestimation are geographically interesting to note; the two major areas of underestimation are the inner downtown core, and the East Bank Campus of the University of Minnesota, just east of the Mississippi River, while the major area of overestimation is the north-central business district downtown. The downtown core and the campus of the University are characterized by significant pedestrian activity and are considered walkable areas, while streets become less walkable closer to the I-94 and I-394 corridors in the western portion of downtown. While the road network structure and proximity to downtown would predict significant pedestrian activity, physical barriers exist within the built environment. These cases highlight the limitations of centrality and accessibility in capturing elements of the built environment relevant to pedestrian activity where local and hyper-local factors may play significant roles.

3 Pedestrian Safety Analysis

3.1 Introduction

Assessment of collision risk between pedestrians and automobiles offers a powerful and informative tool in urban planning regimes, and can be leveraged to inform proper placement of improvements and treatment projects to improve pedestrian safety. Existing, available datasets of crashes, pedestrian counts, and automobile traffic flows can be combined to identify intersections, corridors, or other urban areas that exhibit elevated collision risks to pedestrians. As the availability of count data gradually increases due to automation techniques in counting and crash detection, the process of leveraging these data to determine areas of cities in need of intermodal conflict mitigation will become easier and more problem areas can be identified [14, 5]. Many cities and urban areas do not have access to good-quality count data for traffic other than vehicles, and must instead use estimation techniques to model active transport flow levels, a technique commonly used in planning applications in Europe and Asia, but not yet in the United States [33]. Example methodologies for modeling pedestrian and bicycle non-motorized traffic levels are readily available [25, 9, 34, 40, 37, 16]. However, the City of Minneapolis has a sufficiently rich dataset of pedestrian counts available to allow for meaningful pedestrian safety analysis without implementing complex traffic estimation models.

Safety levels associated with transportation in cities continue to be a problem, with 1.24 million road users being killed in on-road accidents in 2010 globally, and another 20-50 million injured [45]. Further, a full 22% of traffic deaths worldwide are pedestrians, which is quite a high figure considering the transportation mode of walking harbors little danger unto itself; additionally, only 79 countries worldwide have implemented policies to physically separate vulnerable road users (pedestrians, cyclists, and motorcyclists) from autos [45]. Non-motorized transportation modes tend to be some degree of unsafe in most average developed urban areas, except where specific programs and treatments have been employed to address the safety concerns, such as in Copenhagen, Denmark [19].

This investigation aims to evaluate whether the Safety In Numbers phenomenon is observable in originally collected pedestrian and crash data in the midwestern U.S. city of Minneapolis, Minnesota. Safety In Numbers (SIN) refers to the phenomenon that pedestrian safety is positively correlated with increased pedestrian traffic in a given area, e.g. that the per-pedestrian risk of injurious interaction with motorized vehicles decreases as a function of the increasing flow of pedestrian traffic. SIN is well-supported by pedestrian crash data across a number of studies in various urban environments and reviews [18, 23, 3]. The concept has seen relatively widespread adoption in urban planning schools of thought, though its temporal causality is not clear-cut [3], and it is commonly discussed only in the context of pedestrian risk depending on pedestrian flow levels. The United States Department of Transportation (USDOT) Strategic Plan for Fiscal Years 2012-2016 aims to reduce non-vehicle-occupant fatalities to 0.15 per 100 million vehicle-miles-traveled (VMT) by 2016. However, such a goal does not account for risk dependence on pedestrian flow levels, and thus the federal guidelines ignore the SIN effect.

By necessity, data informing placement of improvements and projects for walking and bicycling safety, such as pedestrian bump-outs [44] and traffic-calming measures [6], must be sufficiently granular; travel behavior studies are typically performed at the Transportation Analysis Zone (TAZ) level, which is insufficiently fine-grained to allow for analysis of the shorter-distance travel modes of bicycling and walking. [38] provides a relevant framework for building a comprehensive pedestrian risk assessment model, with a granular focus on a specific university campus and a model which included factors of pedestrian flow, vehicle flow, and an environmental factor (crosswalk length). [43] provide precedent for area-level modeling of pedestrian risk incorporating zoning and land use characteristics. These levels of detail correlate well with the realities of implementation of pedestrian safety investments, which occur not on the city-wide level, but within specific intersections and road segments.

Pedestrian traffic counts, Average Annual Daily Traffic (AADT), and crash data from the city of Minneapolis are used to build a model of crash frequencies at the intersection level as a function of modal traffic inputs. This model determines whether the SIN effect is observable within the available datasets for both pedestrians and cars, as well as determine specific spatial locations within Minneapolis where pedestrians experience elevated levels of risk of automobile crashes, relative to intersections elsewhere in the city. The ability to identify specific unsafe locations based upon aggregated count and crash data offers an additional tool for city planners to implement in multimodal and pedestrian-specific planning.

3.2 Methodology

The existence of the SIN effect was examined within collected data for the city of Minneapolis, at the intersection level. This framework was chosen over other possible areas of analysis, such as mid-block or a link-based framework, due to intersections being the predominant location where pedestrians interact with cars. Turning movement counts (TMCs) for the years 2000-2013 provided pedestrian count data at the intersection level; AADT measurements from 2000-2013 provided vehicle traffic flow levels on street links; traffic crash records from 2000-2013 yielded crash data with location-specific metadata to allow geocoding; an ESRI shapefile of intersection geolocations in Minneapolis provided by the city allowed for spatial analysis and geocoding.

A few steps of data processing occurred prior to building models of crash counts, and crashes per pedestrian. The TMC data identified independent pedestrians passing through an intersection, defined by their directional heading, across 12-hour counts (6AM to 6PM); to account for this, pedestrian counts for each direction (e.g. northbound, eastbound) were summed together to yield a total count. The time windows extracted from the count data were the three peak periods of the day: morning peak (7-9AM), midday (11AM-1PM), and evening peak (4-6PM); these were summed together to produce a 6-hour count, representing pedestrian traffic when a higher number of cars are on the road. The counts took place sporadically across the 14-year window, and most intersections were only counted once or twice due to the rotating schedule on which counts occurred. An "annual average daily 6-hour count" was obtained by averaging the intersection-counts over the number of years for which that intersection was counted. AADT measurements were associated with street links, not to intersections. To associate AADT numbers with intersections, the AADT for each unique street at an intersection (typically two) was summed together. Crashes were tabulated to include both non-fatal and fatal crashes. Finally, manual geocoding of the three datasets (TMC, AADT, and crash counts) to the intersection spatial layer took place, to allow spatial analysis.

At a given intersection, "pedestrian risk" is defined as the number of crashes, between cars and pedestrians, that occurred across the 14-year period, divided by the 6-hour count of pedestrians; this gives a metric to assess the risk of being hit by a car that an individual pedestrian may experience at such an intersection. "Car risk" is defined as the number of crashes that occurred across the analysis period, divided by AADT, which gives a metric to assess the risk of being involved in a car-pedestrian collision which a driver may experience. How these risks vary from intersections with low to high traffic flows determines whether pedestrians or cars experience SIN.

The sample of intersections to use within the analysis was created by identifying intersections with both nonzero pedestrian counts and nonzero AADT data; there were 448 such intersections, and a summary of the sample data can be found in Table 5 in Section 3.3. A map of the intersections included in the analysis is shown in Figure 11. Of note, the 6-hour pedestrian counts comprised a full 84.29% of the 12-hour count totals, averaged across the the 448 included intersections. Of the 448 in the sample, 105 intersections had 6-hour counts which constituted 100% of the 12-hour count totals.

Table 5: Pedestrian Safety Dataset Summary Statistics

Description	Value
Intersections with pedestrian counts & AADT	448
Minimum 6-hour pedestrians	1
Maximum 6-hour pedestrians	14,793
Average 6-hour pedestrians (standard deviation)	832.96 (1,843.72)
6-hour peds average percent of 12-hour count	82.49%
Minimum AADT	252
Maximum AADT	40,623
Average AADT (standard deviation)	8893.33 (5613.00)
Total crashes at sampled intersections	1192 (1180 injuries, 12 deaths)
Minimum crashes at sampled intersections	0
Maximum crashes at sampled intersections	27
Average crashes at sampled intersections (standard deviation)	2.66 (3.86)

Note: Summary statistics for datasets used in pedestrian safety analysis: pedestrian turning movements between 2000-2013, AADT between 2000-2013, and aggregate crash reports between 2000-2013, for the City of Minneapolis.

The model form used is as follows:

$$R = C * Q_p^{b_p} * Q_c^{b_c} \tag{10}$$

where R is the risk factor (either crashes per pedestrian, or crashes per car), Q_p is 6-hour pedestrian traffic flow, Q_c is auto traffic flow (AADT), and C is a constant. This model form allows for log-linear regression analysis, since

$$log(R) = b_p * log(Q_p) + b_c * log(Q_c) + log(C) \tag{11}$$

Such a log-linear model form is commonly used within crash modeling frameworks involving vehicular flows; Abdel-Aty et al. [1] give a log-linear framework for modeling crash frequency using demographic and environmental variables, and Lee et al. [24] give a model of crash exposure in terms of vehicular flows and environmental variables.

Single-variable models including only pedestrian traffic or vehicle traffic, as well as a model with both traffic modes, are included in the analysis. Table 6 outlines the hypotheses for the signs of the exponents in both single-variable and multivariable models of pedestrian and car risk factors.

Table 6: Hypotheses for signs of variable exponents in per-pedestrian and per-car crash risk models.

Model	6-hour peds (b_p)	AADT (b_c)
Single-variable, pedestrian risk	-	n/a
Single-variable, car risk	n/a	-
Multivariable, pedestrian risk	-	+
Multivariable, car risk	+	-

In general, it is hypothesized that increased traffic of a mode has a *negative* effect on rate of crashes per vehicle or user of that mode (e.g. increased pedestrians correlate with lower per-pedestrian crash risk); and, that increased traffic of a mode has a *positive* effect on rate of crashes per vehicle or user of the *other* mode (e.g. increased car traffic correlates with higher per-pedestrian crash risk). Thus, the SIN effect is predicted for both pedestrian and auto modes, and it is predicted that safety for a mode decreases with an increase in traffic of other modes.

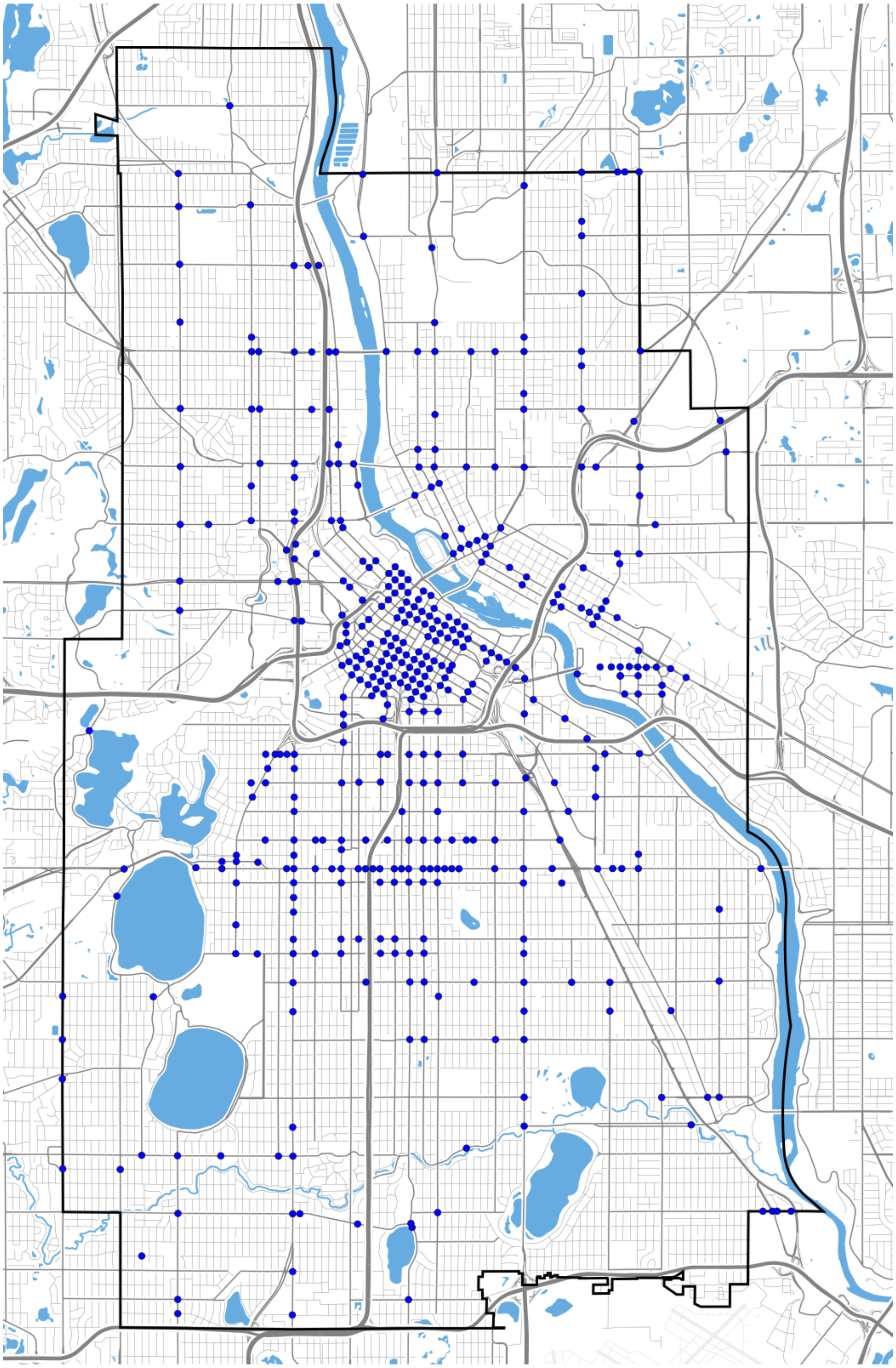

Figure 11: Locations of intersections in Minneapolis with both pedestrian counts and AADT data.

3.3 Data analysis

Table 5 lists summary statistics for the datasets used in the safety analysis: automobile-pedestrian crashes between 2000 and 2013; pedestrian turning movement counts (TMC) between 2000 and 2013; and automobile AADT figures between 2000 and 2013.

As mentioned in Section 3.2, a total of 448 intersections were identified as having both nonzero pedestrian counts and AADT data. The average 6-hour pedestrian count across these intersections was 832.96, and the average AADT was 8893.33. The number of crashes counted at individual intersections, across the 14-year window, ranged from 0 to 27, with an average of 2.66. A series of figures shows visual representations of the datasets included in the analysis. Figs. 12 to 14 show histograms of the pedestrian counts, AADT data, and crash counts, respectively. Figs. 15 to 17 show maps of the pedestrian counts, AADT data, and crash counts, respectively.

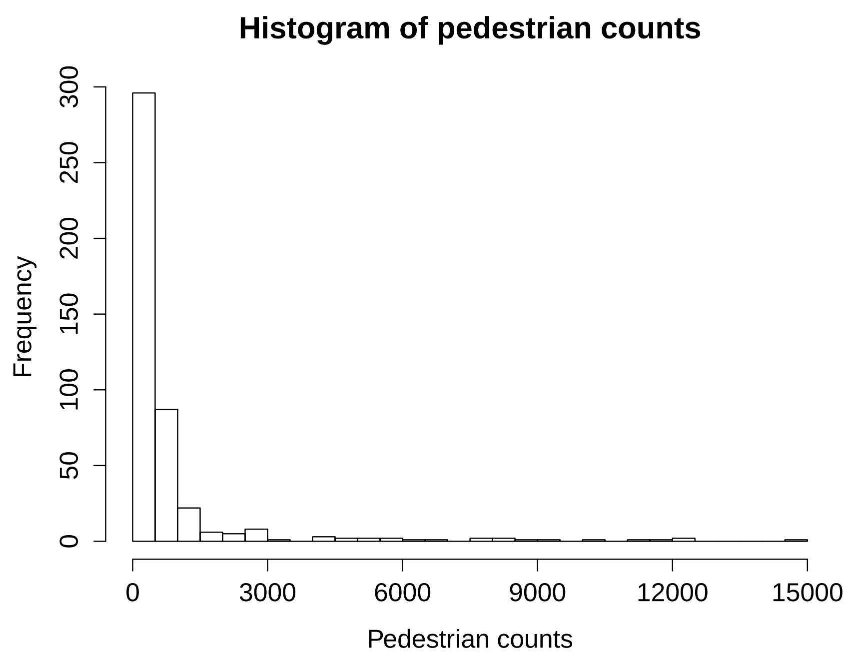

Figure 12: Histogram of pedestrian count data.

As shown in the histograms in Figs. 12 to 14, intersections trend towards having low pedestrian counts, middling AADT, and few crashes. Pedestrian counts and crash counts show decreasing exponential distributions, while AADT values show a more unimodular distribution. Figure 15 gives a spatial representation of pedestrian counts, showing significant activity within the downtown core, as well as areas east of the Mississippi River, corresponding to the University of Minnesota campus. Figure 16 shows the spatial AADT distribution across the city; AADT is more uniformly distributed across a broader area than the pedestrian activity. Finally, Figure 17 shows the spatial distribution of crash counts, with the most occurring along busy corridors to the south and immediate northwest of the downtown core.

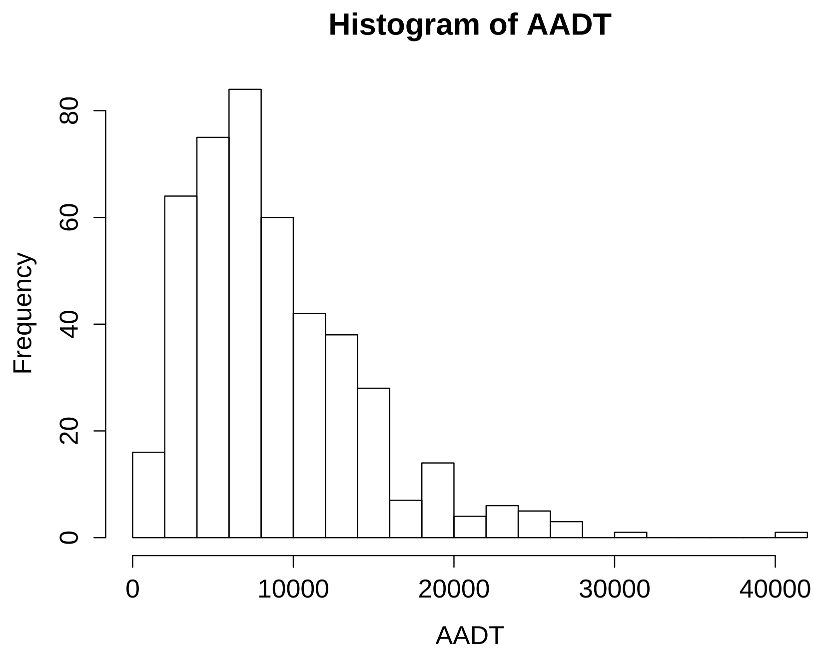

Figure 13: Histogram of AADT data.

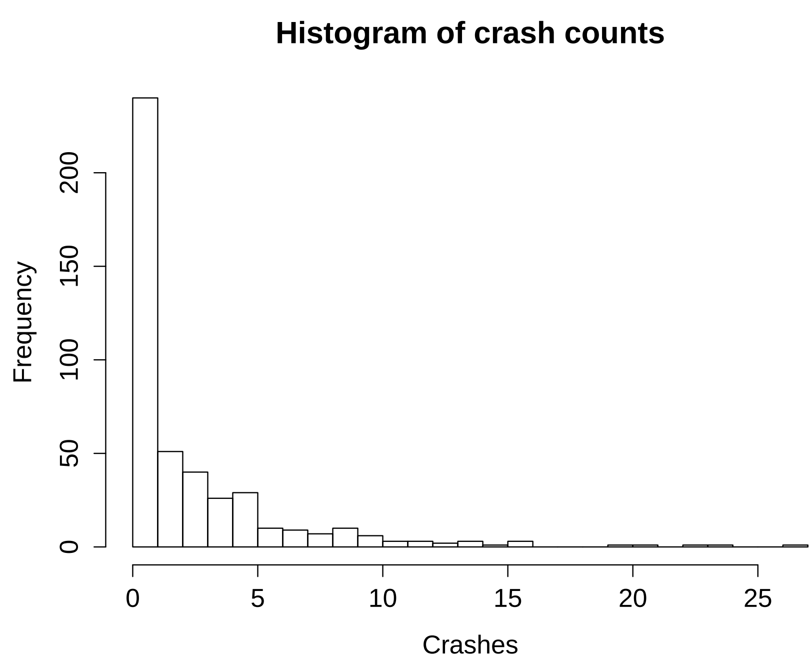

Figure 14: Histogram of crash count data.

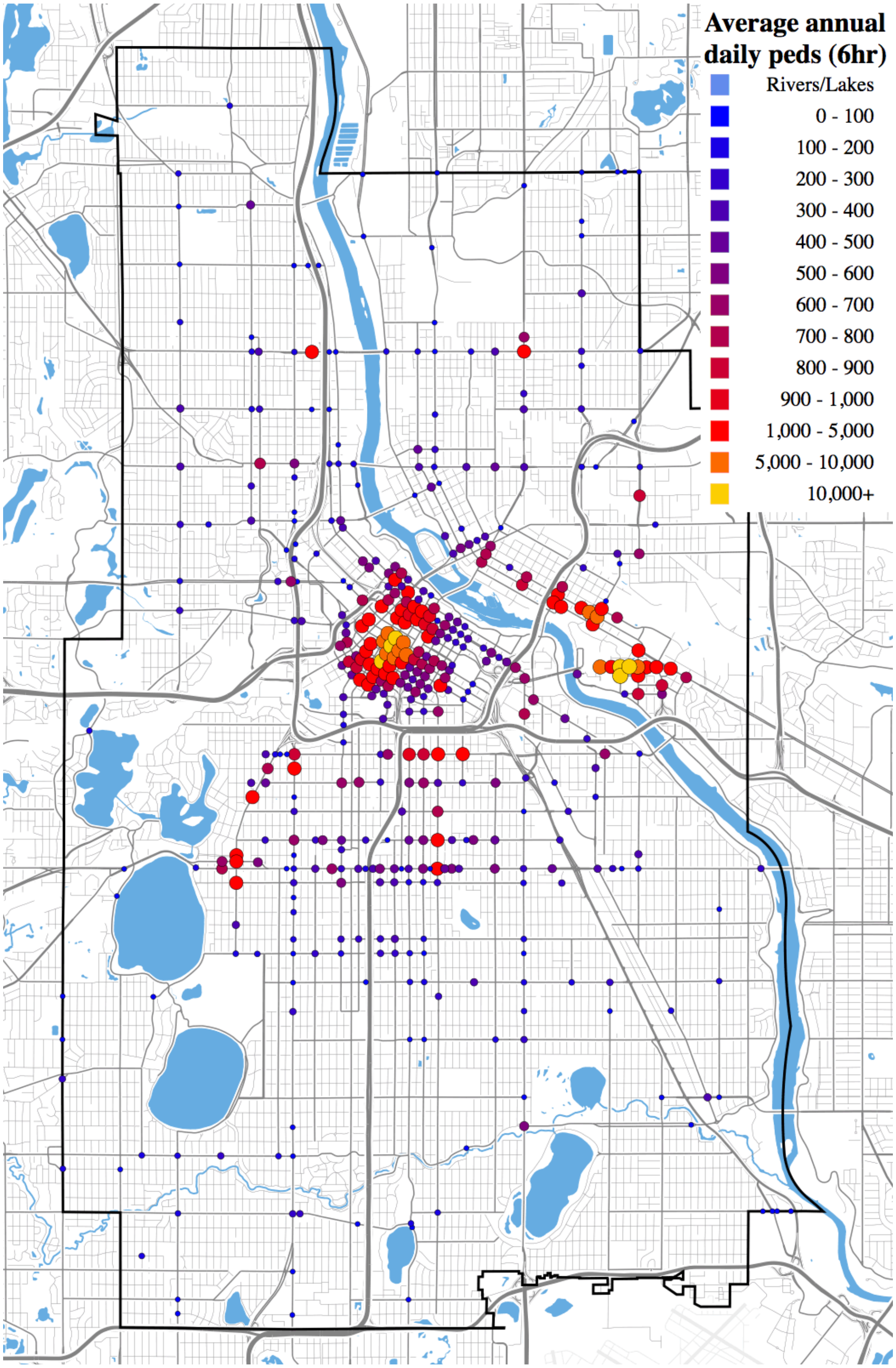

Figure 15: Average annual 6-hour pedestrian counts. Both dot size and color scale correlate with pedestrian count levels.

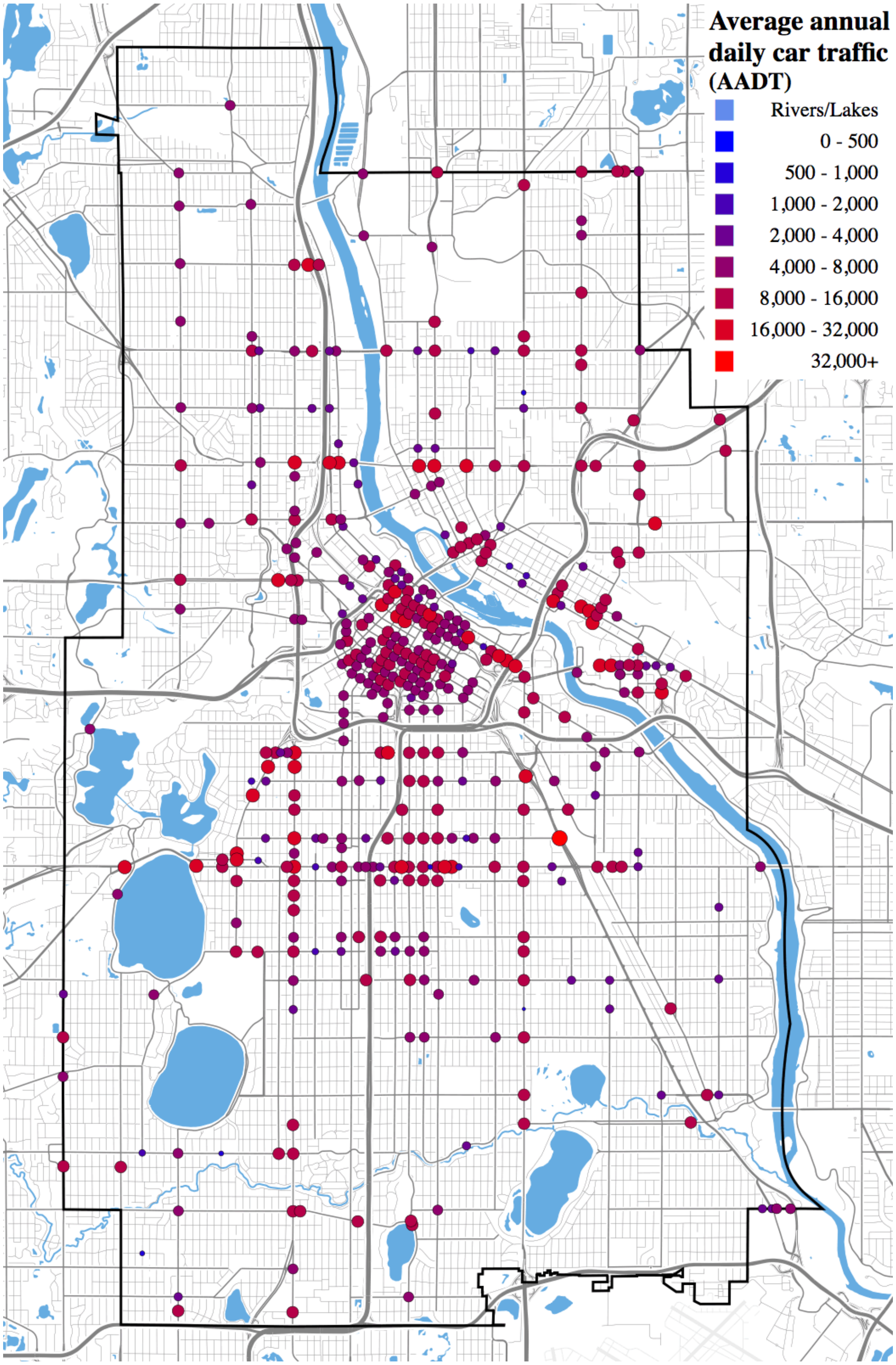

Figure 16: Average annual daily car traffic (AADT). Both dot size and color scale correlate with AADT levels.

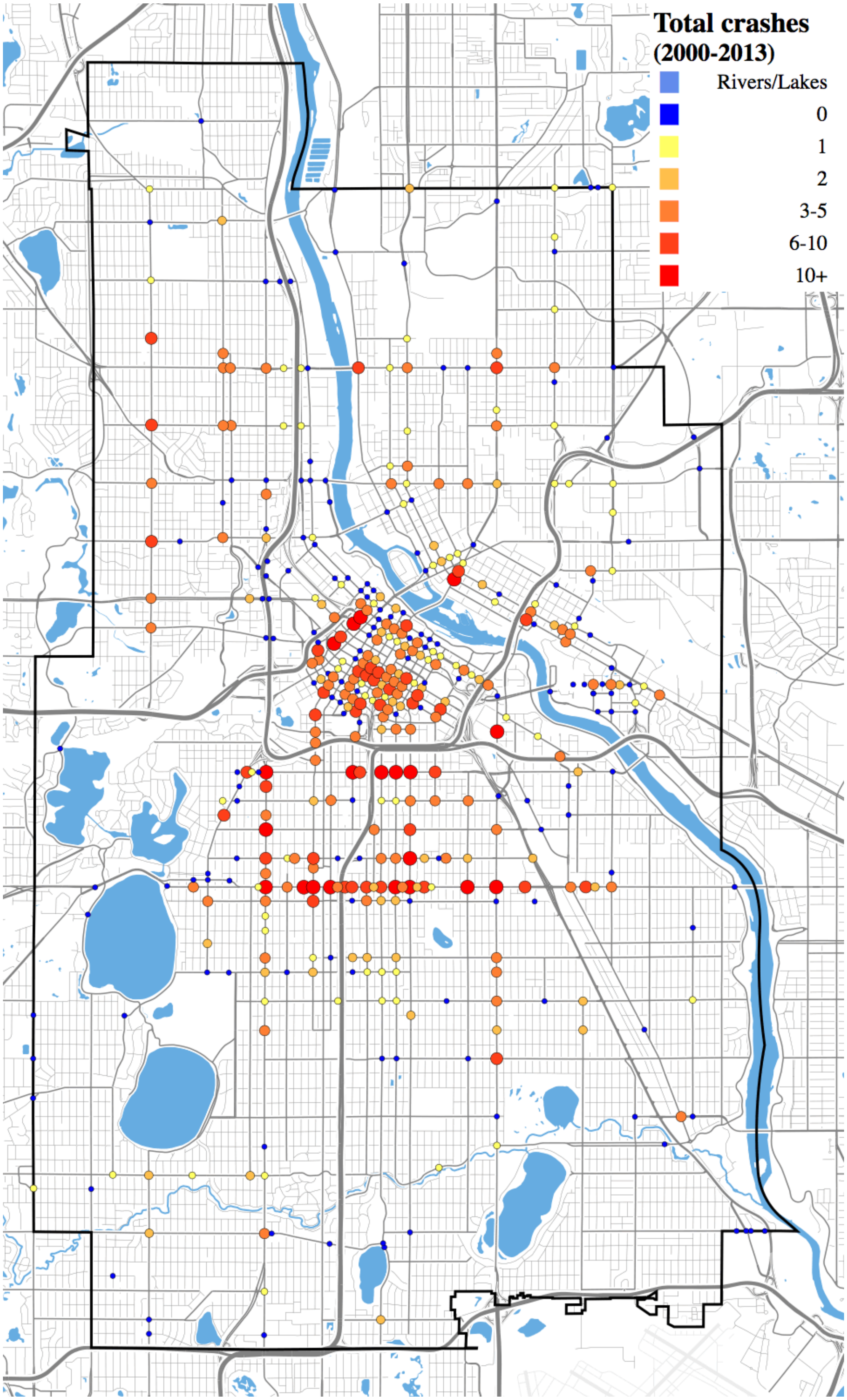

Figure 17: Crash counts at sample intersections across the analysis time window of 2000-2013. Both dot size and color scale correlate with crash count levels.

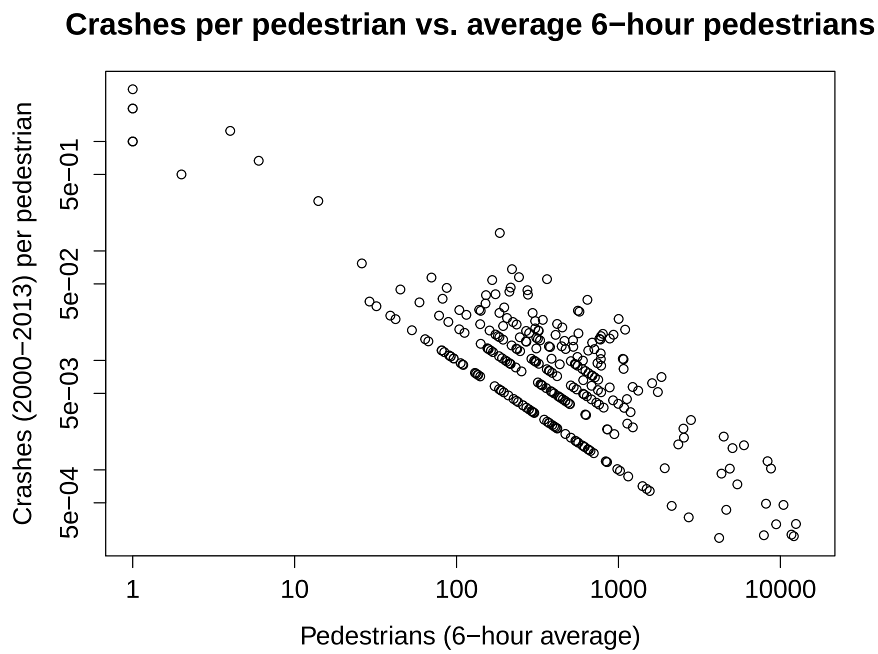

Figure 18: Scatter plot of 14-year crashes (2000-2013) per pedestrian (6-hour average daily) vs. 6-hour pedestrian counts, log-log scale.

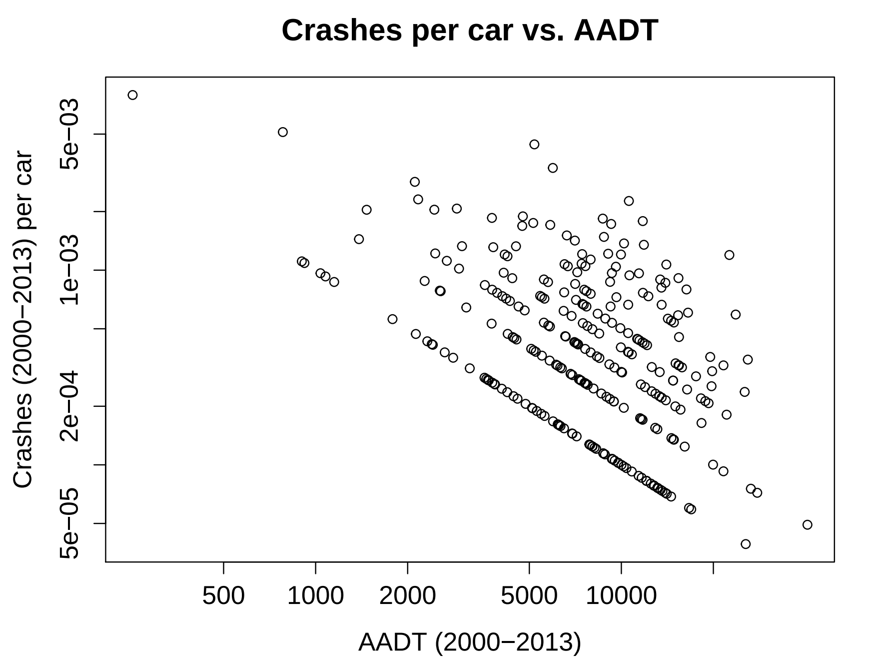

Figure 19: Scatter plot of 14-year crashes per car (AADT) vs. AADT, log-log scale.

Figure 18 and Figure 19 show log-log scatter plots of the relationships between crashes per pedestrian and pedestrian traffic, and crashes per car and car traffic, respectively. These plots suggest a negative relationship between crashes per pedestrian and pedestrian traffic, and between crashes per car and car traffic, and give a view of the dependent variables for the single-variable and multivariable regression analysis. The parallel linear patterns in these plots correspond to isoclines of intersections with the same crash counts.

Figure 20 shows a map of crash totals at intersections, divided by the number of pedestrians counted across the three two-hour peak periods in a day. This view of the data gives spatial representation to intersections, or clusters of intersections, characterized by different levels of risk associated with crossing the intersection. Focus area *A* corresponds to the north-south corridor of Penn Avenue in Minneapolis; focus area *B* is the University of Minnesota campus; focus area *C* is the downtown Central Business District (CBD); focus area *D* constitutes the area surrounding the east-west corridor of Lake Street and the north-south corridor of Lyndale Avenue. All four of these focus areas show a change in pedestrian risk, from considering only raw crash counts (Figure 17) to accounting for pedestrian counts (Figure 20). Areas *A* and *D* show elevated pedestrian risk when accounting for pedestrian counts, and areas *B* and *C* show lower levels of per-pedestrian crash risk.

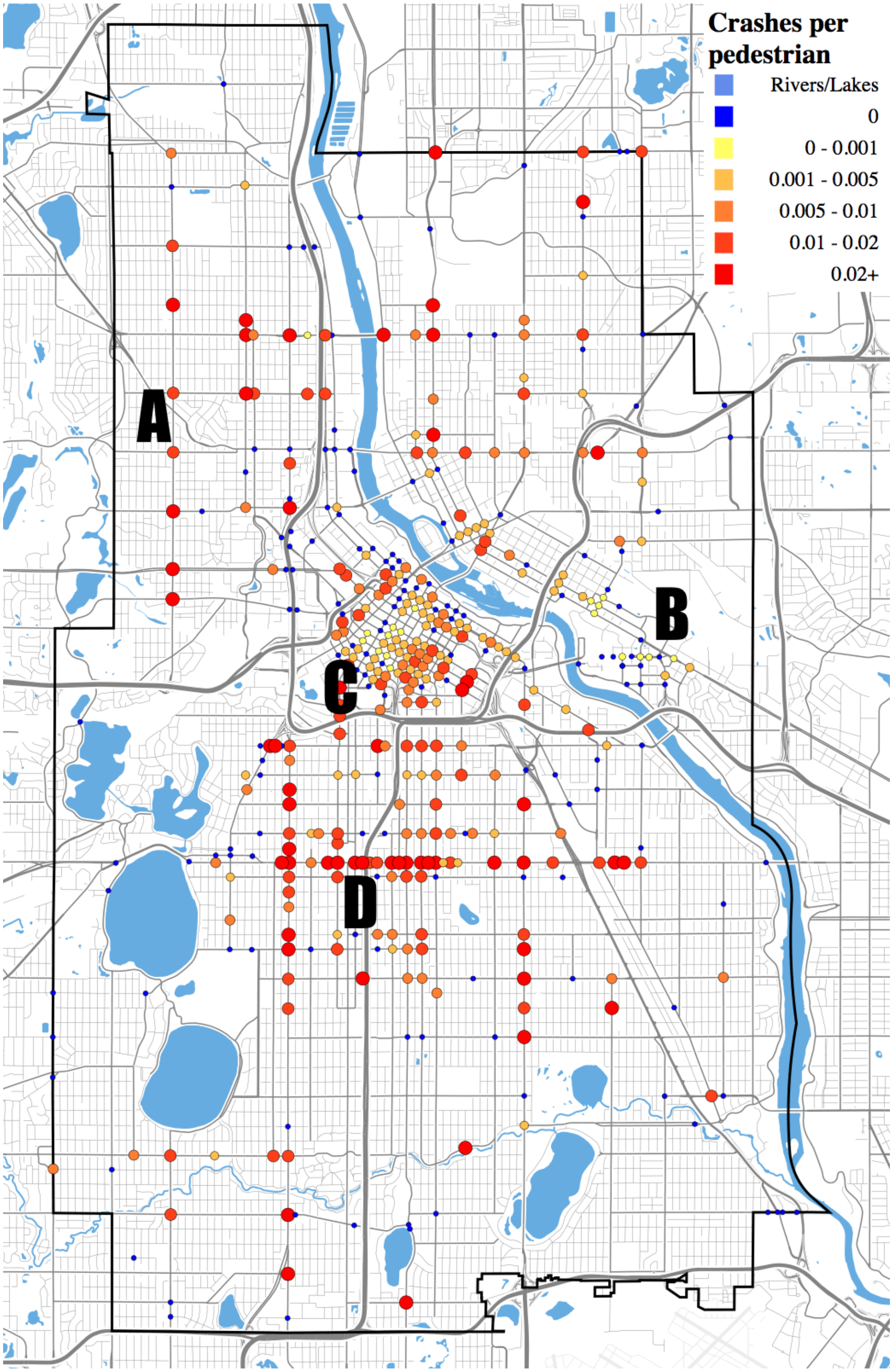

Figure 20: Total crashes (2000-2013) per pedestrian (6-hour daily average) at intersections in Minneapolis. Letter labels indicate focus areas for discussion.

3.4 Regression results

Table 7 shows the log-linear regression results for the single-variable models, and Table 8 shows the log-linear regression results for the multivariate models. For the single-variable log-linear model for SIN for pedestrians, the coefficient of log(6-hour peds) was found to be -0.035 and strongly significant. This is also the exponent b_p in (10), indicating a negative exponential relationship between pedestrian traffic levels and the per-pedestrian risk of a crash. For the single-variable log-linear model for SIN for automobiles, the coefficient of $log(\text{AADT})$ was found to be -0.0003 and strongly significant. This is the exponent b_c in (10), suggesting a negative exponential relationship between auto traffic levels and the per-automobile risk of hitting a pedestrian. However, the effect is two orders of magnitude less for automobiles than it is for pedestrians. Additionally, an R^2 of 0.213 was observed for the single-variable pedestrian risk model, while an R^2 of only 0.098 was observed for the single-variable car risk model.

Table 7: Single-variable log-linear pedestrian safety regression results

	log(crashes/ped)	log(crashes/car)
log(6-hour peds)	-0.035^{***}	
	(0.003)	
$log(\text{AADT})$		-0.0003^{***}
		(0.00004)
Constant	0.223^{***}	0.003^{***}
	(0.019)	(0.0004)
Observations	448	448
R^2	0.213	0.098
Adjusted R^2	0.211	0.096
Residual Std. Error (df = 446)	0.109	0.001
F Statistic (df = 1; 446)	120.392^{***}	48.536^{***}
Note:	(standard error); *p<0.1; **p<0.05; ***p<0.01	

For the multivariable log-linear model describing per-pedestrian crash risk, the exponent b_p was again found to be negative (-0.036) and strongly significant, indicating a negative relationship between increasing pedestrian traffic and per-pedestrian risk of a crash. The exponent b_c was found to be positive and weakly significant, indicating that increased automobile traffic has a positive relationship with the per-pedestrian risk of a crash at intersections.

Similar relationships were found in the multivariable log-linear model describing per-car crash risk. Exponent b_p was positive and strongly significant, indicating a positive relationship between pedestrian traffic and per-car risk of hitting a pedestrian; exponent b_c was negative and strongly significant, indicating a negative relationship between auto traffic and per-car risk of hitting a pedestrian. Again, the model for per-pedestrian crash risk showed a higher R^2 (0.219) than that of the per-car crash risk model (0.117).

In both single-variable and multivariable models, the SIN effect appeared to be stronger for pedestrians than for cars, indicated by the coefficient for log(6-hour peds) in the crashes-per-pedestrian models being larger in magnitude than the coefficient for $log(\text{AADT})$ in the crashes-

Table 8: Multivariable log-linear pedestrian safety regression results

	log(crashes/ped)	log(crashes/car)
log(6-hour peds)	-0.036^{***}	0.0001^{***}
	(0.003)	(0.00002)
log(AADT)	0.014^{*}	-0.0003^{***}
	(0.007)	(0.00004)
Constant	0.102	0.003^{***}
	(0.067)	(0.0004)
Observations	448	448
R^2	0.219	0.117
Adjusted R^2	0.215	0.113
Residual Std. Error (df = 445)	0.109	0.001
F Statistic (df = 2; 445)	62.284^{***}	29.462^{***}
Note:	(standard error); *p<0.1; **p<0.05; ***p<0.01	

per-car models. This is visible in Figure 18, in which the relationship for pedestrians appears to be more coherent than that shown in Figure 19 for cars.

3.5 Discussion

The SIN effect was observed in the analysis of the available pedestrian, auto, and crash data at 448 sampled intersections in Minneapolis. At intersections characterized by higher levels of pedestrian traffic, lower per-pedestrian rates of crashes involving automobiles and pedestrians were observed; at intersections characterized by higher levels of auto traffic, lower per-car rates of crashes involving automobiles and pedestrians were observed. The SIN effect for cars was found to be a few orders of magnitude weaker than the SIN effect for pedestrians; SIN pertaining to pedestrian safety is a well-documented phenomenon [18, 23, 3]. The precise reasons behind this effect are not definitively known; however, the aforementioned studies have hypothesized psychological effects on drivers, in that when driving in environments characterized by greater average levels of pedestrians, drivers may tend to act with more caution. And while these various relationships were statistically observed within the traffic data, it is important to note that the causal directionality of the SIN effect cannot be inferred directly.

The per-pedestrian crash rate was found to increase with increasing automobile traffic, and the per-car crash rate was found to increase with increasing pedestrian traffic, as shown in Table 8. These effects, along with the SIN phenomenon observed for both pedestrians and cars, were consistent with the hypotheses outlined in Table 6. Holding pedestrians constant and increasing car traffic increases the risk of a pedestrian being hit by a car by simple probability of interaction, and the same holds true for increasing pedestrian traffic for a given AADT value. However, the SIN effect was stronger and more coherent for pedestrians than for cars, indicated by their disparate R^2 values. This suggests that there may be more factors relevant to the number of auto-pedestrian crashes per vehicle than just the intersection's AADT, such as intersection geometry, speed limit, number of lanes, or other environmental variables.

An ongoing challenge with pedestrian safety analysis dependent on count and crash data is the issue of data quality and availability. Data practices vary from city to city and state to state, with implications to investigations intending to aggregate safety data for cross-jurisdiction comparison. Additionally, a large amount of city data collection pertaining to street utilization is still performed manually, and such processes are error-prone and inconsistent between jurisdictions. This study used an available, fairly robust dataset provided by the City of Minneapolis, covering pedestrian counts, AADT, and crash data for 14 years from 2000–2013. Some cities, such as Boston, do not have robust pedestrian and bicycle counting programs throughout the city; others, such as Philadelphia, may have varying data release and non-disclosure agreements between Metropolitan Planning Organizations (MPOs), cities, and police departments; still other cities may have inconsistent data tracking and release practices. The collection and processing of pedestrian and bicycle spatial safety data on an aggregate scale becomes exceedingly difficult. Better standards of practice in data collection, management, and distribution are needed.

This analysis took place over a 14-year time period, during which there may have been changes to the on-street pedestrian network facilities which influenced both pedestrian and auto traffic, as well as crash frequency. For instance, construction on the Metro Transit Green Line light rail system, which runs through the University of Minnesota campus (area *B* in Figure 20), began in 2010. Auto and pedestrian traffic patterns in the area were greatly altered, with part of the area being permanently closed to car traffic, and many pedestrian improvements being implemented with the light rail line. This study does not account for area-specific network changes, as the focus is city-wide aggregate pedestrian safety; more detailed investigation into network changes in specific areas, and resulting traffic and accident patterns, would be needed.

Visualizing unsafe intersections, or groups of intersections, within an urban area is an important angle of analysis to undertake with the types of datasets used in this investigation. Problematic areas within the city environment become readily apparent; when multiple intersections with relatively high pedestrian injury risk burden lie in the same corridor, such as Lake Street in Minneapolis (see area *D* in Figure 20), a discussion of pedestrian safety and the surrounding built environment should occur. The entire Lake Street corridor stands out as an area with elevated pedestrian risk burdens, given the number of pedestrians walking there, compared to the relatively walk-friendly downtown district (area *C*) and the University of Minnesota campus (area *B*). Similarly, the Penn Avenue corridor in North Minneapolis (area *A*) shows a series of intersections with elevated pedestrian risk.

Most of the corridors displaying elevated pedestrian risks in Figure 20 may be classified as urban arterials, with more than one lane of traffic in each direction. That such roadways would be unsafe for pedestrians may seem obvious, but a visualization tool can be more powerful in informing planners the scope and extent of dangerous streets within an urban area than simple crash counts alone.

3.6 Conclusion

Through the pedestrian risk burden analysis framework outlined in this study, it is possible to observe the SIN effect at the individual intersection level for both pedestrians and cars, as well as identify intersections with disproportionately high rates of crashes for its level of pedestrian activity. Pedestrians were found to be at a lower risk of being hit by a car at intersections with higher pedestrian traffic, and individual cars were found to be at a lower risk of hitting pedestrians at intersections with more car traffic. The causality of the SIN effect is not understood, and more research should be conducted to understand its causes, but it is still a justification for improving the walkability and pedestrian safety of urban environments. Assessing the per-pedestrian crash rates at spatial locations, as opposed to crash counts alone, allows practitioners and planners to more readily identify target areas where improvements to pedestrian infrastructure may be warranted.

All associated datasets and appropriate metadata used within the pedestrian study have been uploaded to the Data Repository for the University of Minnesota (DRUM).

4 Bicyclist Activity Estimation and Safety Analysis

4.1 Introduction

Active modes of travel, including walking and bicycling, are increasingly becoming important transportation modes in modern cities. Individual and societal wellness, environmental impacts, and resource availability are among multifarious reasons that have drawn the attention of transportation and urban planners toward active modes of travel. Planning for biking and walking and creating societal programs to increase their levels have been cited as a targeted health need in urban planning going forward [27, 33, 4].

Resource limitations, particularly in high-population and developing countries, impose constraints on the maximum level of personal motorized travel allowed, and as a result, there is a greater need for viable alternatives. While active modes of travel have been shown to be positively correlated with curbing air pollution and promoting health, over half of the global annually reported 1.25 million vehicle accidents involve a pedestrian, bicyclist or motorcyclist. The number of bicyclists compared to motorized vehicles would suggest nearly negligible annual traffic accidents, yet bicyclists make up 2% of all traffic related deaths [42]. Active modes of travel as a set of modes tends to be less safe than motor vehicles on a per-kilometer basis. This holds true in most average developed urban areas, except where specific programs and treatments have been employed to address safety concerns, such as in Copenhagen, Denmark [19].

Rates of walking and bicycling to work in the United States hover around 2.8% and 0.6%, respectively, with public transit use higher at 5% nationally [42]. Although proper placement of bicyclist treatments and improvements has implications to both safety [36] and accessibility and mode choice [17], proper information regarding estimated non-motorized traffic levels is needed to locate areas in need of improvement. In determining salient locations for non-motorized improvements, it is important to have accurate records of both existing and potential travel demand (e.g., current levels of biking in a neighborhood, as well as good models of increased demand due to potential treatments); however, good quality, high-granularity datasets for non-motorized travel can be difficult to obtain, especially standardized for national spatial inventories [29]. Hence, practitioners and researchers must frequently rely on estimation models for non-motorized traffic, and various methods can suffer from issues of data quality, granularity, and the presence of location-specific variables [26].

Many of the issues with the collection of standardized non-motorized transportation data have to do with the factors that influence pedestrian and bicyclist behavior. A model of active transport risk assessment is uninformative if the bicyclist and vehicular flows do not accurately represent corresponding levels *in situ*, and many cities do not have dense datasets of active transport flow levels, instead favoring counts of vehicle traffic. As such, active transport flow levels must be extrapolated from sparse datasets using comprehensive methodologies. Population and employment data are well-documented by the U.S. Census Bureau to the Census Block level of resolution, and general socioeconomic characteristics are maintained as well and can have significant influence [36]. However, more specific socioeconomic characteristics are salient in non-motorized travel beyond just adjusted income levels, as well as weather variables [31] and latent, subjective variables, such as visibility and perceptions of lighting, which can be more difficult to obtain at high spatial resolution [20] and can complicate inter-city comparisons. For these reasons, as well as the overall lack in non-motorized travel counts for many communities, methods of estimating pedestrian and bicyclist behavior that do not rely heavily on high-resolution count data are applied in this study.

A reduced-form core facility demand model gets around the issue of data quality and granularity by using easily retrievable datasets to predict bicyclist counts at the intersection level [15]. This model was developed using existing pedestrian and bicyclist counts, land-use variables, and transportation network variables extracted from the state of Minnesota. The facility demand model estimates peak-period traffic volumes at intersections where counts are unavailable. The outcome was an estimated comprehensive pedestrian and bicyclist count dataset for the city of Minneapolis, which can be used to examine trends related to bicyclist activity.

This investigation aims to further understanding of the "safety in numbers" (SIN) phenomenon and its dependence on both bicyclist and vehicular flow levels. SIN refers to the phenomenon that bicyclists as road users become safer when there are more riders present in a given locale or area. To date, SIN is well-supported by bicyclist crash data across a number of studies in various urban environments and reviews [18, 23, 3]. The most frequently cited hypothesized cause of the SIN effect is that motorists adapt their driving behavior when traversing a roadway that frequently carries pedestrian and bicyclist traffic [41, 18]. The SIN concept has seen relatively widespread adoption in urban planning schools of thought, though its temporal causality is not clear-cut [3], and it is commonly discussed only in the context of bicyclist risk depending on bicyclist flow levels.

A lesser-studied occurrence is the "safety in congestion" (SIC) effect. The hypothesis states that greater volumes of cars will reduce the per-car crash rate. The reasoning is that greater congestion reduces vehicle speeds thereby giving drivers greater reaction time to reduce the severity of a crash or avoid one altogether. This study tests the SIC effect in two ways. By assessing the estimated crashes per bicyclist and per car outcomes as a function of bicyclist and vehicle traffic, the SIC effect can be supported by either modes of transport.

Aggregate travel behavior studies typically involve analysis at the level of Traffic Analysis Zones (TAZs), which are too coarse to allow robust analysis of non-motorized travel [36, 17]. Regional travel surveys consider many trip purposes but are similarly coarse, and typically have sample sizes too small to allow for robust city-to-city comparison. A framework for comprehensive bicyclist risk assessment modeling, using estimated bicyclist volume per intersection, observed vehicle volume and crash records is presented in this report. The motivation for using models of bicyclist traffic is in supplementing the sparse data currently available to assess bicyclist risk burdens of collisions at every intersection in Minneapolis. Bicyclist risk burdens — the risk of an individual bicyclist being struck by a vehicle — are calculated and compared for both the raw and predicted crash per bicyclist datasets. This process allows us to construct a more complete spatial picture of how bicyclist collision risk varies throughout an urban area at the level of individual intersections, based on data widely available to practitioners, transportation authorities and the public.

The remainder of this report is organized as follows. Section 4.2 reviews the studies that make up the background of the current state of SIN research. Section 4.3 introduces the data used in this research along with the data extraction and preparation process. Following the discussion of the methodology in Section 4.5, the results of the modeling procedure are provided along with an in-depth interpretation of interest variables both qualitatively and quantitatively. Finally, Section 4.6 concludes the report with summarizing the key findings and opening new research avenues.

4.2 Background

The notion of a safety in numbers effect has been around since the 1940s, when Smeed [39] showed that road fatalities per vehicle were lower in countries with more driving. He demonstrated that an exponential curve describes the relationship between fatal vehicular crashes and vehicle kilometers traveled (VKT). In the last two decades, several studies from around the world have shown that the SIN effect exists. A variety of methodologies were employed to try to capture the magnitude and the contributing factors to the SIN effect while controlling for environment and human behavior. One such study took place in Hamilton, Ontario, Canada where pedestrian flow was compared to the crash rates [23]. Data collected from 300 signalized intersections from 1983-1986 contained pedestrian crashes and estimated pedestrian and vehicular flows. When risk per pedestrian was calculated using the expected pedestrian flows, decreasing risk was associated with increasing pedestrian flows. Conversely, increasing vehicle flows was associated with increased pedestrian risk. Models were estimated for different times of the day. The crash counts at each intersection were considered as a Poisson random variable. This study found that drivers seem to expect pedestrians when the pedestrian flow is over 30 pedestrians per hour. It was also found that the level of bicyclist flow is more important for bicyclist safety than the level of vehicular exposure. A similar study was conducted in Sweden in 1996 which compared bicyclist volumes against crashes at 95 intersections. Once again, an inverse relationship was found between bicyclist volume and the number of bicyclist-auto crashes [11].

A study in 2003 used five datasets which included three population level and two time series datasets. It was found that the SIN effect is "consistent across communities of varying size, from specific intersections to cities and countries, and across time periods" [18]. This study used a dataset that linked the number of crashes with the amount of walking and bicycling, however vehicle flow was not an explanatory variable. The model was estimated as a power curve. It was found that the number of pedestrians and bicyclists struck by vehicle vary by the 0.4 power of the pedestrian or bicyclist traffic. Years earlier, researchers in Australia had tested the power model on a dataset that contained over 100 years of crash information [22]. Another more recent Australian study [12] used three types of pedestrian/bicyclist injury datasets to recreate the negative exponential curve. Safety in numbers was found to exist in Australia with a similar exponential relationship compared to the American studies. If cycling doubles, the risk per kilometer falls by about 34%.

After reviewing several years of studies that were conducted around the world to verify the SIN effect, Rune Elvik found that transferring trips from motorized vehicles to walking and biking will reduce the number of crashes [12]. This study changed functional form based on the type of crash involving a pedestrian/bicyclist (multi-vehicle, single-vehicle). The parameters that were varied included number of motor vehicles, pedestrians, bicyclists and the coefficient values for pedestrian and bicyclist crashes. The exponential form was used and the risk calculated for different AADT values (2,000-30,000). It was found that, in theory, the total number of crashes could go down if a substantial share of trips by motorist transport is transferred to walking or cycling [12].

Studies that show evidence of the safety in numbers effect have suggested that policies that encourage walking or biking may improve safety for these vulnerable road users. Bhatia and Wier [3] asserts that the non-linear association between pedestrian and bicyclist volumes and the rates of crashes is too simplistic of a model to draw conclusions from and basing urban planning guidelines on the SIN effect would be unwise.

4.3 Data

We employed two types of data to create the crash prediction models: (1) Estimated Bicyclist Activity Data and (2) Bicyclist-Auto Crash Prediction Model Data. The former is borrowed from a previously developed land use and transportation network regression model. The latter is raw observations taken from local records. This section discusses these data sources and the data preparation process.

4.3.1 Estimated Bicyclist Activity Data

Facility-demand bicyclist estimates developed by Hankey and Lindsey [15] are derived from a reduced-form core model, which allows practitioners to use easily retrievable datasets to predict bicyclist counts at the intersection level. The model used to develop the estimated bicyclist traffic was derived from counts taken in September during the peak-period (4 p.m. - 6 p.m.). Independent variables were selected based on their known likelihood to affect a citizen's propensity to bike. The 2014 employment accessibility for the Twin Cities region was included along with land use variables (i.e., industrial area, population density, retail area, open space) and the number of bicyclist facilities. Temporal variables such as temperature and precipitation were used to account for the weather shifts in Minnesota and the resulting bicyclist activity. The 2010 U.S. Census core-based statistical areas for Minneapolis-St. Paul were used to cordon bicyclist facility counts, population, and demographic information. A complete discussion is provided by Hankey and Lindsey [15]. The estimated counts were chosen for this study to expand the dataset available for analysis. Figure 21 depicts the number of daily bicyclists observed to pass through each intersection shown. In comparison, Figure 22 shows the estimated bicyclist counts that were made available through the facility-demand model.

- PM peak period bicyclist counts observed in September from 2007-2015, conducted by the City of Minneapolis Department of Public Works (DPW) and Transit for Livable Communities (TLC)
- Employment accessibility within 5-60 minutes of walking in 2014, University of Minnesota Accessibility Observatory
- Land use statistics, Metropolitan Council 2015
- U.S. Census TIGER 2010 datasets: blocks, core-based statistical area (CBSA) for Minneapolis-St. Paul
- Tabulation of yearly bicyclist facilities, (DPW) and (TLC) 2007-2015

- Weather parameters, (DPW) and (TLC) 2007-2015

4.3.2 Bicyclist-Auto Crash Prediction Model Data

The crash prediction model presented in this report uses four independent variables to predict the number of crashes at a given intersection in Minneapolis over a 14 year period. The independent variables include the estimated bicyclist TMC and the observed average annual daily traffic (AADT) and their quadratic forms for 489 intersection in Minneapolis. The training dataset included 80% of these for a total of 383 intersections. The remaining 20% were used as the test dataset. For comparative purposes, raw bicyclist turning movement counts (TMC) were assessed and plotted to verify that using estimated bicyclist TMCs would be an improvement compared with the raw data. The number of bicyclist-auto crashes that were recorded from 2000 to 2013 at each of these intersection was included as the dependent variable. A 2014 OpenStreetMap extract of the Twin Cities region was used to geocode crash records to intersections. OpenStreetMap is an open-source platform of free and reusable geospatial data.

- OpenStreetMap (OSM) North America extract, retrieved July 2016
- Raw bicyclist Turning Movement Counts (TMC) 2007-2014, City of Minneapolis
- Estimated bicyclist Turning Movement Counts September 4-6 PM, City of Minneapolis
- Average Annual Daily Traffic (AADT) measurements 2000-2013, City of Minneapolis
- Traffic crash records 2000-2013, City of Minneapolis

4.4 Data Preparation

Intersection locations were determined from OSM road centerline data for the Minneapolis-St. Paul CBSA (Core-Based Statistical Area). To get a sense for the magnitude of bicyclist traffic throughout Minneapolis, Figure 23 was developed to visualize the distribution of bicyclist activity. The estimated bicyclist TMC was geocoded to intersections for a single value of bicyclist traffic at each intersection from 4 p.m. - 6 p.m. Estimated peak-hour bicyclist volumes were expanded to 24-hour counts using bicyclist traffic count factors to extrapolate the estimated counts [35]. The AADT records from 2000 to 2013 were averaged over those years and assigned to intersections by applying a mid-block buffer around each intersection in QGIS and summing the cross streets for a single value of vehicle traffic. Crash records were geocoded to intersections using the OSM extract and QGIS. Figure 23 shows the locations and levels of bicyclist-auto crashes that occurred at each intersection in the test dataset. GIS work was performed in QGIS and PostGIS; statistical work done in Stata and NLOGIT. Table 9 gives the description and statistics of the variables used in this study for both parts of the modeling procedure.

1. Construct bicyclist travel network graph for Minneapolis

2. Geocode bicyclist Turning Movement Count (TMC), estimated bicyclist activity, Average Annual Daily Traffic (AADT), and crash data to spatial locations

Table 9: Bike Activity & Safety Dataset summary statistics

Variable	Description	Min	Max	Mean	Std. Dev.
	Training Data (n=383)				
Crashes	Cumulative crashes from 2000-2013	0	16	1.50	2.4
Vehicle Traffic	Mean daily traffic per intersection 2000-2013	252	30798	8584.9	5693.0
Bicyclist Traffic	24 hour bicyclist count per intersection	37	3,935	793.1	562.6
	Test Data (n=106)				
Crashes	Cumulative crashes from 2000-2013	0	11	1.54	2.3
Vehicle Traffic	Mean daily traffic per intersection 2000-2013	440	25927	8424.8	4920.1
Bicyclist Traffic	24 hour bicyclist count per intersection	57	2,163	877.2	539.7

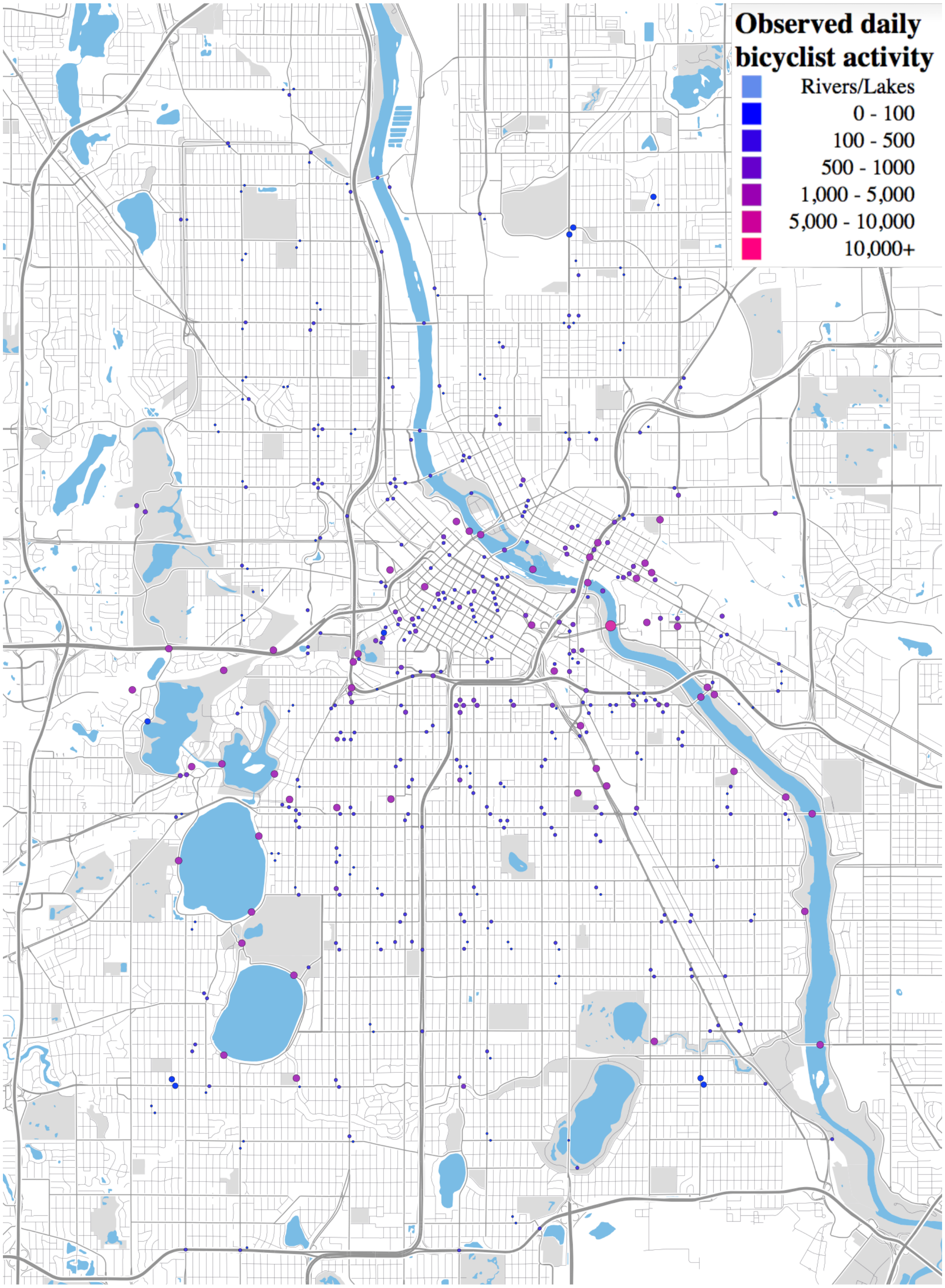

Figure 21: Observed levels of daily bicyclist activity in Minneapolis, 2007-2014

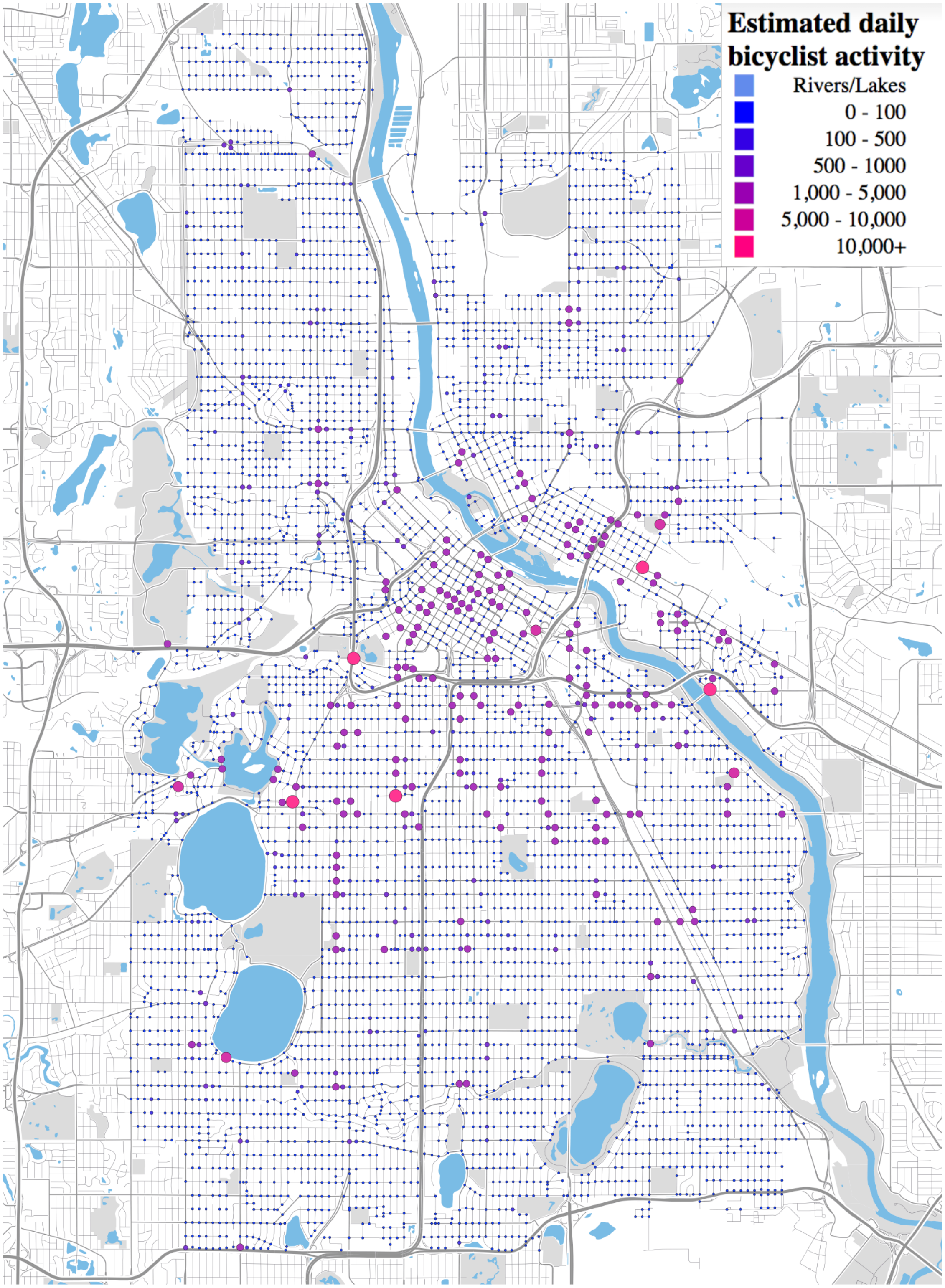

Figure 22: Estimated levels of daily bicyclist activity in Minneapolis.

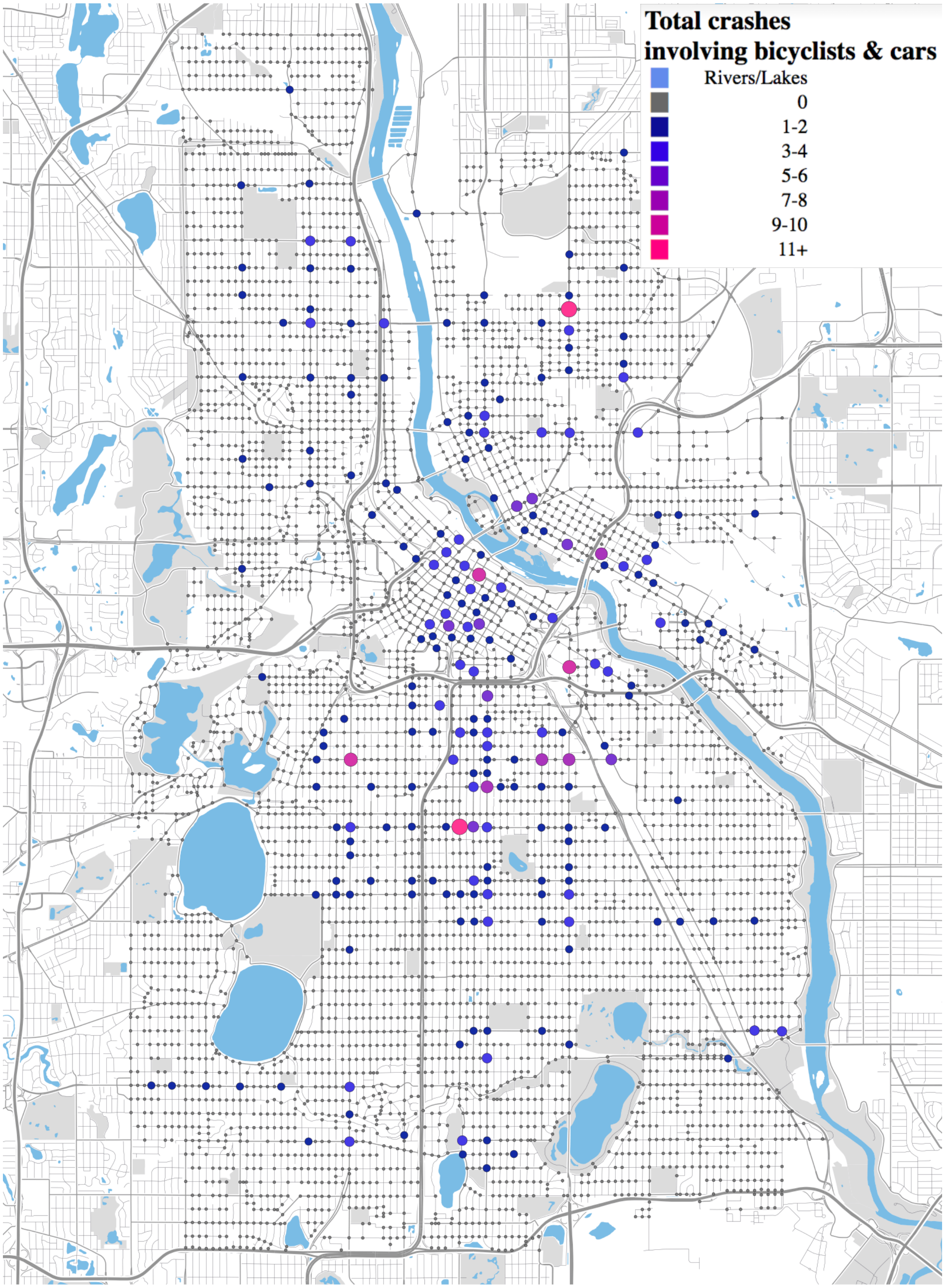

Figure 23: Raw levels of bicyclist-auto crashes in Minneapolis, 2000-2013.

4.5 Modeling: The Evidence of Safety in Numbers

To investigate the association between number of crashes and interest variables, two distinct models were used: (1) Ordinary Least Squares and (2) Two-part Model of Crashes. Student's t-statistic and Adjusted R^2 are measured to test the statistical significance of variables and the general fit of models, respectively. The data is randomly divided into two portions: An 80% part for training the models and a 20% part for testing the prediction power of the models. These models are based on two primary hypotheses:

- The number of crashes has a diminishing return to scale with respect to average daily motor vehicle traffic.
- The number of crashes has a diminishing return to scale with respect to daily bicyclist traffic.

To test the preceding hypotheses, the interest variable is used along with its quadratic form in the models. Doing so reflecects a more realistic association between number of crashes and the interest variables. The following subsections delve into the modeling process and testing the preceding hypotheses.

4.5.1 Ordinary Least Squares Model of Crashes

Ordinary Least Squares (OLS) regression is used at the outset of this study to get a preliminary understanding of the relationship between bicyclist and vehicle traffic and the resulting number of crashes over a 14 year period. The summary statistics for the predicted and predictor variables is given in Table 9. The linear regression analysis is performed in STATA V.14, and the results are shown in Table 10. The results indicate that the explanatory variables are significant at the 5% significance level. Both the SIN and SIC effects are supported by the direction of the linear regression results. Vehicle and bicyclist activity are estimated to have a positive effect on the number of crashes. The overall fit of the model is ($R^2 = 0.10$). Figure 24 is based on the test set of intersections in Minneapolis and shows that intersections with low observed crashes tend to be overestimated while more dangerous intersections are have an underestimated number of crashes.

Table 10: Ordinary Least Squares regression results—bike crashes

Variable	*Coefficient*	*Std. Err.*	*t-test*	*p-value*
Vehicle Traffic	2.46×10^{-4}	6.22×10^{-5}	3.95	0.000
Vehicle Traffic^2	-6.14×10^{-9}	2.45×10^{-9}	-2.51	0.013
bicyclist Traffic	1.74×10^{-3}	5.07×10^{-4}	3.44	0.001
bicyclist Traffic^2	-5.06×10^{-7}	1.91×10^{-7}	-2.65	0.008
Constant	-0.84	0.39	-2.13	0.034
Number of observations		383		
Adjusted R^2		0.1032		

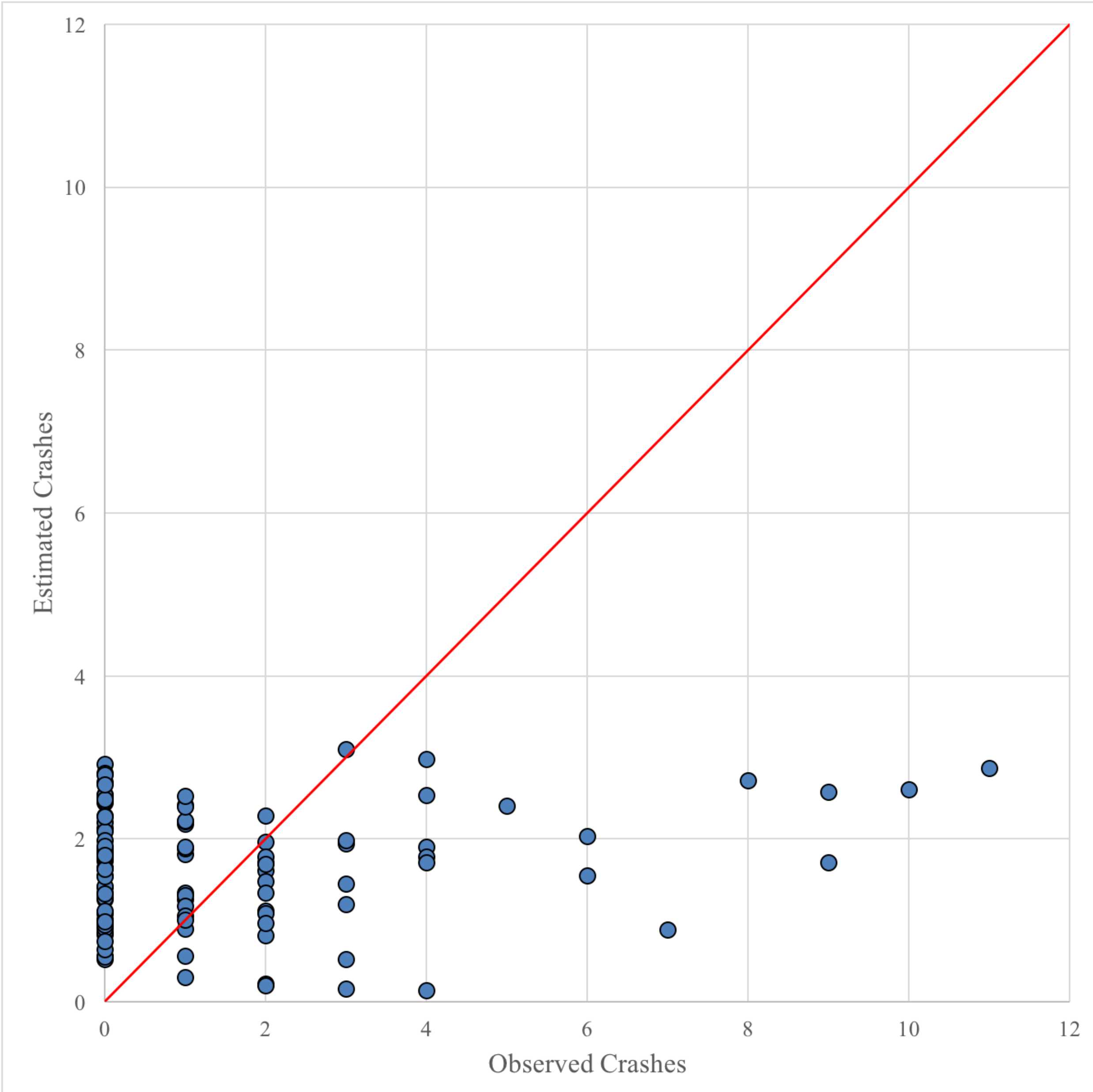

Figure 24: OLS regression estimated crashes against observed crashes with 45 degree divider.

4.5.2 Two-part Model of Crashes

Modeling the number of bicyclist-auto crashes requires a different procedure than traditional count outcomes or linear models, as the majority of intersections have no cumulative crashes over time. A high proportion of zeros in the distribution of the number of crashes variable means that standard approaches such as least squares regression misleads the results. To overcome this challenge, statisticians have introduced a host of methods that account for such infrequent distributions including the Heckit, latent Heckit, and Two-part methods. These models are generally applied to model distributions of continuous and nonnegative data which contain a large proportion of zero observations. Dow and Norton [10] discuss the advantages and disadvantages of each method broadly.

First cut analysis demonstrates that nearly half of the intersections used in this study have no reported crashes between 2000 and 2013. To represent the marginal effects of exogenous variables accurately, a Two-part model of crashes is used. This model comprises Probit regression for the first part and Poisson regression for the second part of the model. The former predicts the probability of zero versus non-zero crashes at a given intersection, and the latter model is applied to intersections with one or more than one crash. It was assumed that the crash data were not over-dispersed and the excess zeros assumption was addressed in the first-part of the modeling procedure. Figure 25 depicts the framework of the Two-part model of crashes. The modeling results are outlined in Table 11.

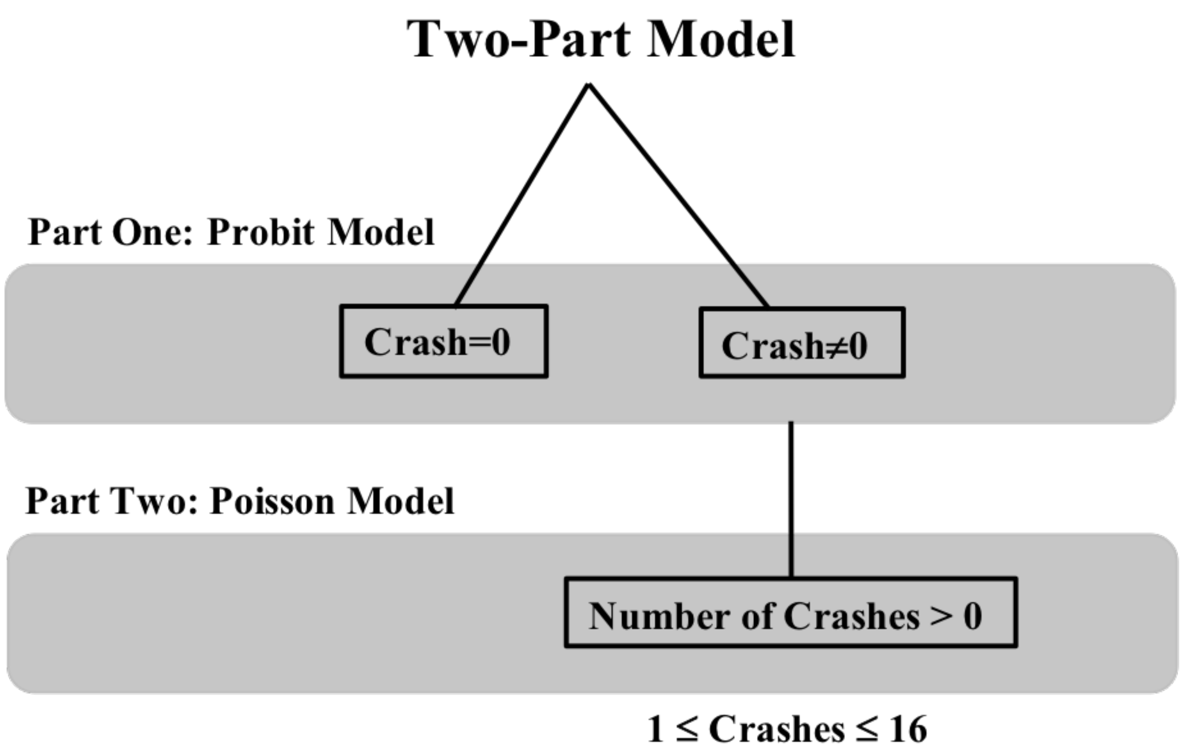

Figure 25: Two-part model tree diagram.

4.5.3 General Discussion

The student's t-statistic measurement indicates that all variables are significant at the 90% confidence interval. Looking at the first-part of the model, the downward parabola form of the *Vehicle*

Traffic variable demonstrates that the probability of crashes has a diminishing return to scale considering the average annual daily traffic as an input. This confirms our hypothesis that increasing the number of vehicles reduces the rate of crashes. The results also show that the probability of crashes starts declining beyond the vertex of a parabola, where the parabola crosses its axis of symmetry. In our specific case, the vertex point equals 13,861.25. This means the probability for a crash to occur begins to decline by increasing the AADT beyond the 13,861.25 value. Congestion causes roads to operate at a fundamentally different level compared to pre-congestion traffic volumes. The SIC effect may be a result of the characteristics of highly congested roads. The same trend exists for the bicyclist traffic regressor. An increase in traffic of bicyclists beyond the 1,314.22 vertex point decelerates the probability for a crash to occur. Bicyclists experience the SIN effect after such volumes have been reached.

Looking at the second part of the model, like the first part, the downward parabola form of the *Vehicle Traffic* variable shows that the number of crashes has a diminishing return to scale considering the average annual daily traffic as an input. This is also true for the bicyclist traffic. Extracting the vertex points of both exogenous variables, the number of crashes starts decreasing when vehicle traffic and bicyclist traffic per intersection exceed 29,568 and 1,532 respectively.

Table 11: Two-part model results—bike crashes

Variable	Part One		Part Two	
Model Specification	Y_1 No Crash: 0, Crash: 1		Y_2 Number of Crashes	
Description	*Coefficient*	*t-test*	*Coefficient*	*t-test*
Vehicle Traffic	1.05×10^{-4}	3.03	9.58×10^{-5}	3.70
(Vehicle Traffic)^2	-3.82×10^{-9}	-2.74	-1.62×10^{-9}	-1.66
Bicyclist Traffic	6.65×10^{-4}	2.12	9.90×10^{-4}	3.61
(Bicyclist Traffic)^2	-2.53×10^{-7}	-1.91	-3.23×10^{-7}	-2.57
Constant	-0.80	-3.57	-0.14	-0.75
Number of observations	383		190	
Pseudo R^2	0.03		0.352	

4.5.4 Sensitivity Analysis

To quantify the association between number of crashes and interest variables, the elasticity of each independent variable was calculated for both the OLS and the Two-part model of crashes. The elasticity results are outlined in Table 12. By definition, the elasticity is the ratio of the percentage change in number of crashes to the percentage change in the interest variable. In line with the above hypotheses, both *Vehicle Traffic* and *Bicyclist Traffic* variables have an inelastic effect. This result is compatible among both the OLS and the Two-part models. The magnitude of effect, however, is varied.

- **Vehicle Traffic**: A 1% increase in the average annual daily motor vehicle traffic increases the probability of crashes by 0.14% and the number of crashes, given there is a crash, by 0.80%.

- **Bicyclist Traffic**: A 1% increase in the average daily bicyclist traffic increases the probability of crashes by 0.09% and the number of crashes, given there is a crash, by 0.50%.

To segments of data are used in order to compare the elasticities of the Two-Part model with the OLS model. One set of OLS elasticities is based on the entire test dataset and includes observations with zero crashes. The second set of elasticity values for the OLS model are calculated for observations with more than zero crashes. Several comparisons can be made between the effects of motor vehicle and bicyclist traffic among the OLS and Two-Part models.

In general, there are marginal differences between elasticity values in the OLS model while the Two-Part model indicates that vehicle and bicyclist traffic differ in their contribution to the probability and later the number of crashes. The OLS model elasticities that were calculated using the observations that contained one or more crashes can be compared to the second part of the Two-Part model due to the modeling procedure. By screening out the observations with zero crashes in the OLS elasticity calculation, the respective elasticity values become closer in magnitude to the Poisson Model elasticity values. Nonetheless, the Two-Part model is the more reliable of the two models presented for analysis. It appears that once an intersection has been predicted to have a crash by the Probit model, the effect of motor vehicle traffic on the number of crashes is greater than the effect of bicyclist traffic. This would indicate that prescribing motor vehicle facility improvements would have a greater likelihood of reducing bicyclist-auto crashes than would bicyclist facility improvements. It cannot be said what this means for policy makers and practitioners due to the uncertain nature of predicting where crashes may occur.

A key takeaway from the sensitivity analysis is that increasing the presence of motor vehicle traffic has a greater impact on the number of crashes than bicyclist traffic as indicated by the Two-Part model. It may be justified by the positive correlation between the bicyclist demand and bicyclist facility, which may result in more awareness of drivers.

Table 12: Elasticity of bicyclist-auto crashes with respect to motor vehicle and bicyclist traffic.

Variable	**OLS Regression**		**Two-Part Model**	
	All Crashes	*Crashes >0*	*Part One: Probit Model*	*Part Two: Poisson Model*
Vehicle Traffic	0.65	0.65	0.14	0.80
Bicyclist Traffic	0.70	0.61	0.09	0.50

4.5.5 Prediction Accuracy

Out of 106 test data, 55 intersections were observed to have between one and 11 crashes over the 14 year period. The first part of the model predicted the 51.8% of the crashes accurately with the probability of 50%. The accuracy of the model is slightly better than random with a pseudo R^2 value of 0.03. A low R^2 value is acceptable in this case because bicyclist-auto crash occurrences are highly random events. The mean relative percentage error (MRPE) measurement (Equation 12) is used to measure the prediction accuracy of the second-part of the model. In this equation, C_i is the observed number of crashes at intersection i, C_i' is the predicted number of crashes at intersection i, and n stands for the number of observations.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\left|C_i' - C_i\right|}{C_i} \times 100 \tag{12}$$

The MRPE results show that the second-part of the model predict the number of crashes with a 82.6% error on average.

Figure 26 and Figure 27 graphically represent the prediction of crashes. These plots were generated by using the test dataset and applying the Two-part model to predict the number of bicyclist-auto crashes across a range of bicyclist and vehicle traffic levels. These graphs can be used to assess the vehicle and bicyclist risk associated with an intersection of recorded traffic volumes.

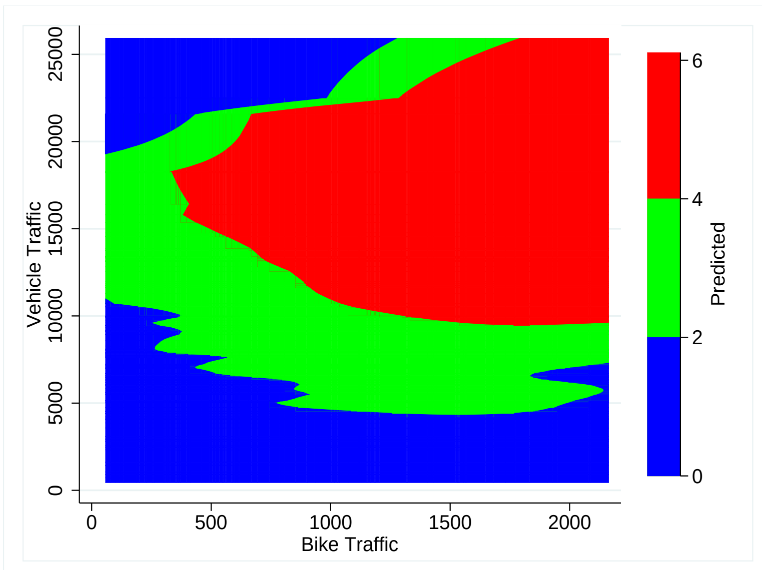

Figure 26: The contour plot of the predicted crashes

Figure 28 depicts the method used to determine if one or more crashes are estimated to occur at a given intersection. Intersections that were initially estimated to have a crash probability less than 0.5 were taken as zero-crash intersections while estimated probabilities greater than or equal to 0.5 were taken to be intersections with one or more crashes. Figure 29 represents the estimated probability for a bicyclist-auto crash to occur at every intersection within the test data set. The intersections with one or more estimated crashes were passed up to the second part of the Two-Part model. Figure 30 plots the estimated number of crashes (1 or greater) against the observed number of crashes for every intersection within the test data set. Intersections with one or two observed crashes tend to be overestimated while intersections that have been observed to be more dangerous are underestimated.

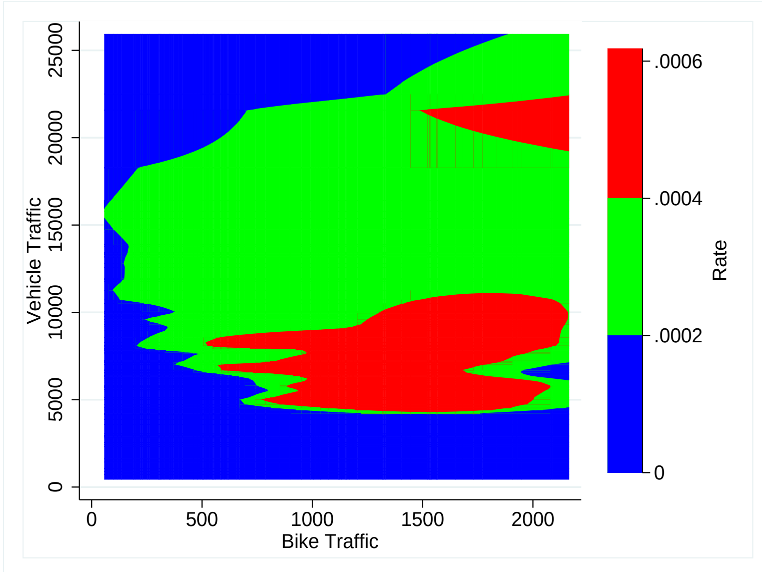

Figure 27: The contour plot of the rate of number of crashes to traffic volume

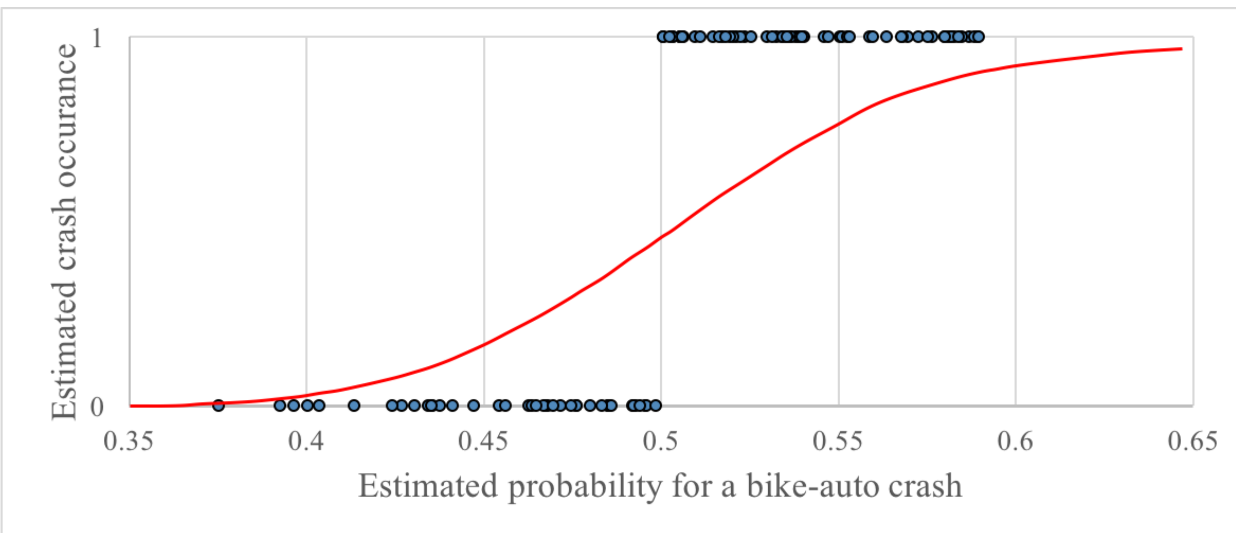

Figure 28: Two-part model estimated probability for a bicyclist-auto crash.

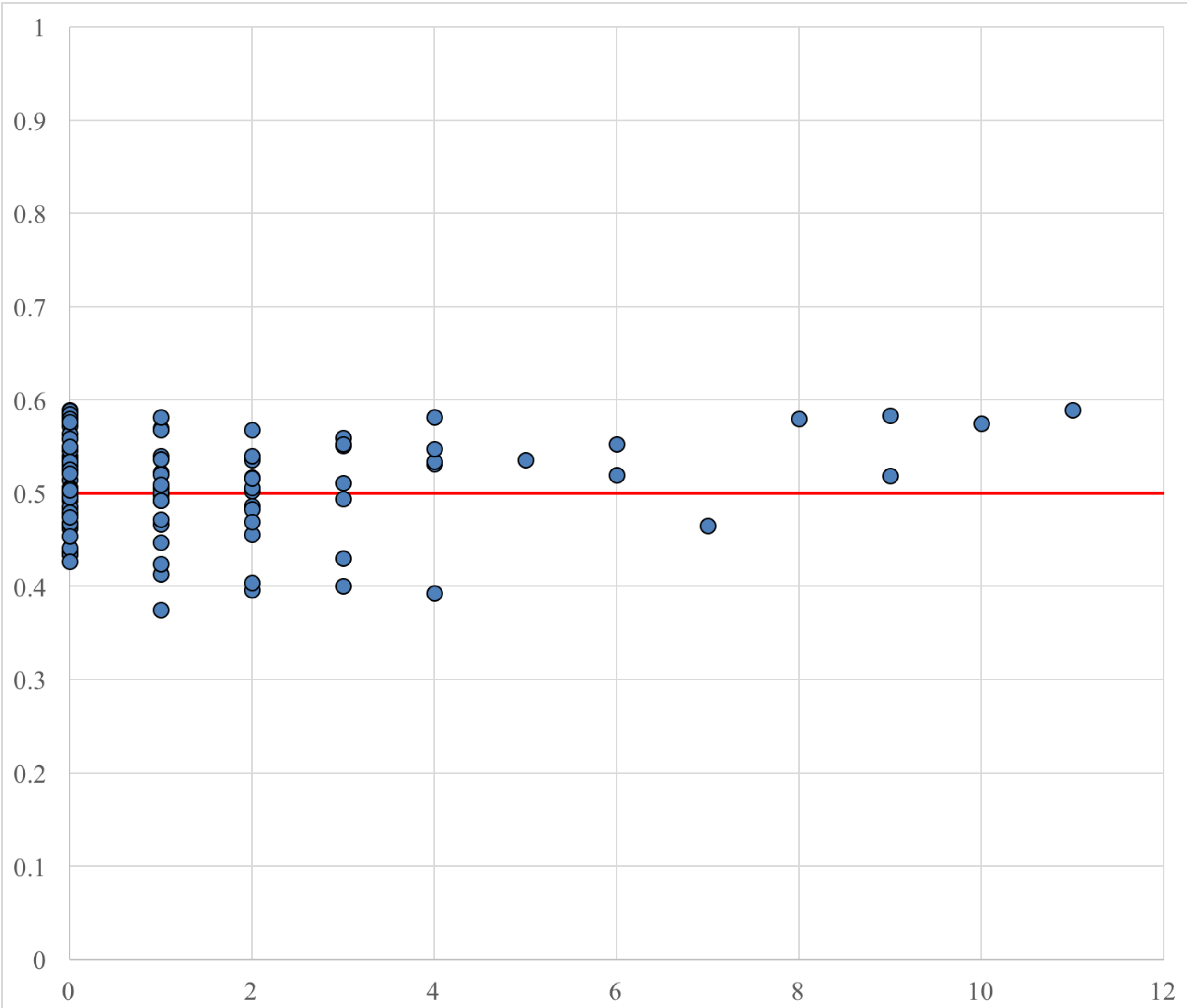

Figure 29: Two-part model estimated probability for a crash to occur against observed crashes with p=0.5 divider.

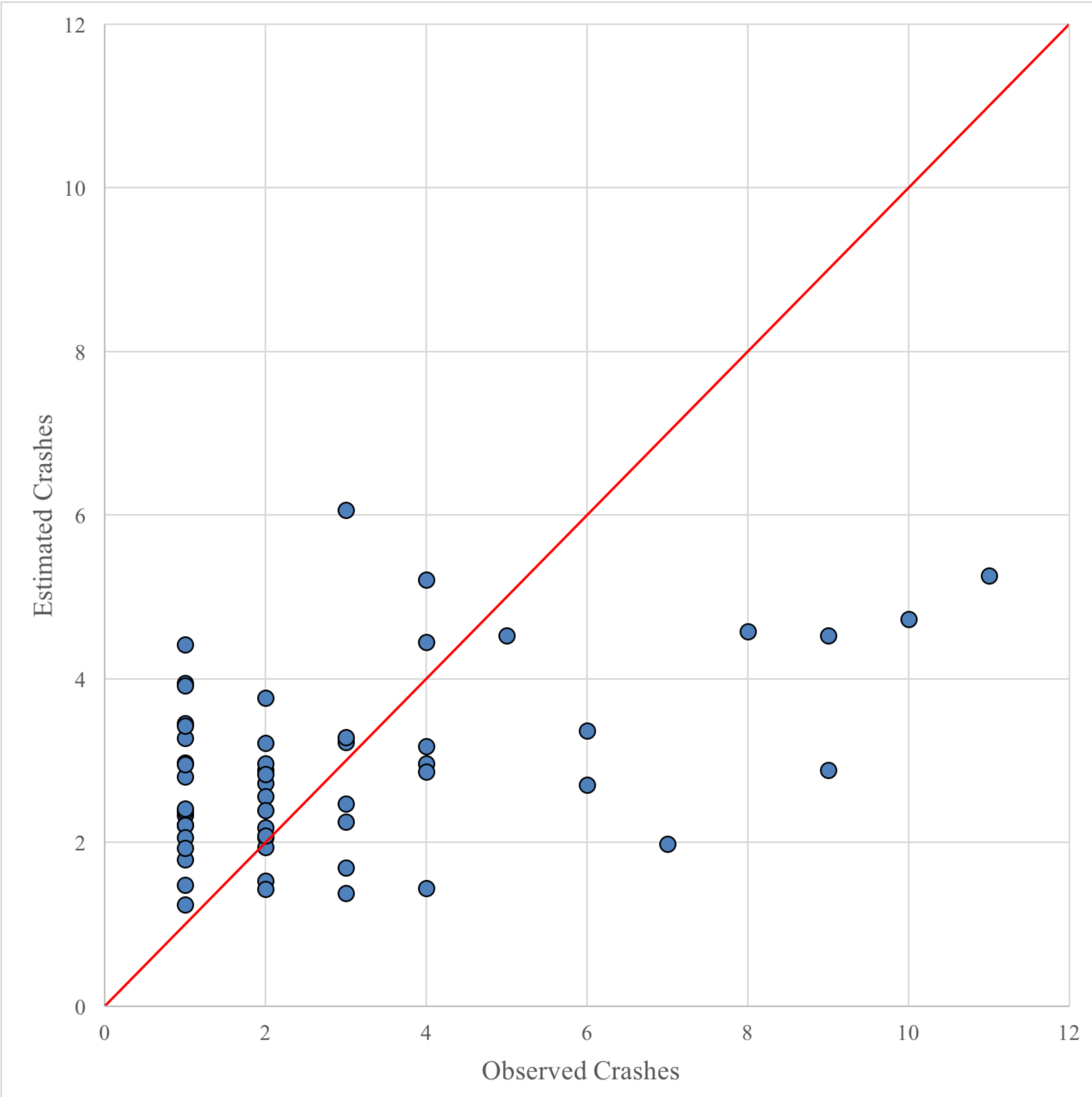

Figure 30: Two-part model estimated crashes, given there is a crash, against observed crashes with 45 degree divider.

4.6 Conclusion

The concept of SIN in transportation planning reflects that there is a non-linear statistical correlation between the number of pedestrians and cyclists and the number of crashes. Studies used longitudinal and cross-sectional data at different level of aggregation to examine whether and to what extent the SIN phenomenon is legitimate. The results appear mixed. The current study is an attempt to delve into this research realm by applying a two-part model of crashes on traffic data for 489 intersections in the Minneapolis - St. Paul metropolitan area between 2000 and 2013.

The data were randomly into two pieces to not only calibrate the model for number of crashes against the average daily vehicle traffic and the daily bicyclist traffic (DBT), but also to measure the accuracy of the model. To understand the association function between the number of crashes and both average daily vehicle traffic and the daily bicyclist traffic, the quadratic functional form of the interest variables was used in the model. This enables us to shed light on the accuracy of the SIN phenomenon and to quantify the safety returns to scale.

Both ADT and DBT were found to have diminishing returns to scale. This accentuates the positive role of SIN. Increasing the number of vehicles and cyclists decelerates not only the probability of crashes, but the number of crashes as well. However, their impacts are unequal. Measuring the elasticity of the variables, it is found that a 1% increase in the average daily motor vehicle traffic increases the probability of crashes by 0.14% and the number of crashes by 0.80%. While, a 1% increase in the average daily bicyclist traffic increases the probability of crashes by 0.09% and the number of crashes by 0.50%. Additionally, the saturation point of the safety in numbers for bicyclists is markedly less than motor vehicles. Extracting the vertex point of the parabola functions reveals that the number of crashes starts decreasing when vehicle traffic and bicyclist traffic per intersection exceed 29,568 and 1,532, respectively.

From the prediction side, this study indicates that unlike the emphasis on the effectiveness of ADT and DBT, these variables do not have a significant impact on describing the number of crashes. The results of the Nagelkerke Pseudo R^2 demonstrate that ADT and DBT improve the estimate of probability of crashes by 3% compared with a null model and improve the estimate of the number of crashes by 35% compared with the null model.

Because this study contemplated whether and to what extent the vehicle and bicyclist traffic affects the number of crashes, it provides insights for future research avenues. The following suggestions are made for further research:

- The use of additional road geometry features, such as signalization and the number of approach lanes, may improve the model, which could in getting a more accurate safety in numbers impact.

- By accounting for variables that may influence vehicle and bicyclist traffic, it may be possible explain a greater percentage of the variation in the number of crashes for a given intersection configuration and activity level.

- One caveat with the bicyclist-auto crash dataset is that bicyclists tend to report only the more severe crashes. This means the crash records underreport the actual number of crashes that occur on a yearly basis which masks the true risk level at an intersection. The prevailing limitation to this and other bicyclist behavior studies is the lack of consistent bicyclist TMC data collected annually.

References

[1] Abdel-Aty, M. A., Chen, C. L., and Schott, J. R. (1998). An assessment of the effect of driver age on traffic accident involvement using log-linear models. *Accident Analysis and Prevention*, 30(6):851–861.

[2] Anciães, P. R. (2011). *Urban transport, pedestrian mobility and social justice: A GIS analysis of the case of the Lisbon metropolitan area*. PhD thesis, London School of Economics and Political Science.

[3] Bhatia, R. and Wier, M. (2011). "Safety in numbers" re-examined: Can we make valid or practical inferences from available evidence? *Accident Analysis and Prevention*, 43(1):235–240.

[4] Brownstone, D. (2008). Key relationships between the built environment and VMT. Special Report 298, Transportation Research Board, Division on Engineering and Physical Sciences.

[5] Bu, F., Green-Roesel, R., Diogenes, M. C., and Ragland, D. R. (2007). Estimating pedestrian accident exposure: Automated pedestrian counting devices report. Technical report, UC Berkeley Traffic Safety Center.

[6] Campbell, B. J., Zegeer, C. V., Huang, H. H., and Cynecki, M. J. (2004). A review of pedestrian safety research in the united states and abroad. Technical Report FHWA-RD-03-042, U. S. Department of Transportation, Federal Highway Administration.

[7] Dai, D., Taquechel, E., Steward, J., and Strasser, S. (2010). The impact of built environment on pedestrian crashes and the identification of crash clusters on an urban university campus. *The Western Journal of Emergency Medicine*, 11(3):294–301.

[8] Derrible, S. (2012). Using network centrality to determine key transfer stations in public transportation systems. In *Transportation Research Board 91st Annual Meeting*, number 12-0021.

[9] Do, M. T., Grembek, O., Ragland, D., and Chan, C.-Y. (2013). Weighting integration by block heterogeneity to evaluate pedestrian activity. In *Transportation Research Board 92nd Annual Meeting*.

[10] Dow, W. H. and Norton, E. C. (2003). Choosing between and interpreting the heckit and two-part models for corner solutions. *Health Services and Outcomes Research Methodology*, 4(1):5–18.

[11] Ekman, L. (1996). On the treatment of flow in traffic safety analysis-a non-parametric approach applied on vulnerable road users. *Lunds Tekniska Hogskola Bulletin*, 136.

[12] Elvik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis and Prevention*, 41(4):849–855.

[13] Federal Highway Administration (2010). 2009 national household travel survey.

[14] Greene-Roesel, R., Diogenes, M. C., Ragland, D. R., and Lindau, L. A. (2008). Effectiveness of a commercially available automated pedestrian counting device in urban environments: Comparison with manual counts. Technical Report UCB-ITS-TSC-2008-5, UC Berkeley Traffic Safety Center.

[15] Hankey, S. and Lindsey, G. (2016). Facility-demand models of peak-period pedestrian and bicycle traffic: A comparison of fully-specified and reduced-form models. *Transportation Research Record*, (2586):48–58.

[16] Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B., and Xu, Z. (2012). Estimating use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis, MN. *Landscape and Urban Planning*, 107(3):307–316.

[17] Iacono, M., Krizek, K. J., and El-Geneidy, A. (2010). Measuring non-motorized accessibility: Issues, alternatives, and execution. *Journal of Transport Geography*, 18(1):133–140.

[18] Jacobsen, P. (2003). Safety in numbers: More walkers and bicyclists, safer walking and bicycling. *Injury Prevention*, 9(3):205–209.

[19] Jensen, S. U. (2008). Safety effects of blue cycle crossings: A before-after study. *Accident Analysis and Prevention*, 40(2):742–750.

[20] Kamargianni, M. and Polydoropoulou, A. (2014). Generation Y's travel behavior and perceptions towards walkability constraints. In *Transportation Research Board 93rd Annual Meeting*.

[21] Kharecha, P. A. and Hansen, J. E. (2008). Implications of "peak oil" for atmospheric co2 and climate. *Global Biogeochemical Cycles*, 22(3):1–10.

[22] Knott, J. W. (1994). Road traffic accidents in New South Wales, 1881-1991. *Australian Economic History Review*, 34:80–116.

[23] Leden, L. (2002). Pedestrian risk decrease with pedestrian flow: A case study based on data from signalized intersections in Hamilton, Ontario. *Accident Analysis and Prevention*, 34(4):457–464.

[24] Lee, C., Hellinga, B., and Saccomanno, F. (2003). Real-time crash prediction model for the application to crash prevention in freeway traffic. In *Transportation Research Board 82nd Annual Meeting*.

[25] Liu, X. and Griswold, J. (2009). Pedestrian volume modeling: A case study of san francisco. *Yearbook of the Association of Pacific Coast Geographers*, 71(1):164–181.

[26] Lowry, M. (2014). Spatial interpolation of traffic counts based on origin-destination centrality. *Journal of Transport Geography*, 36:98–105.

[27] Lumsdon, L. and Mitchell, J. (1999). Walking, transport and health: Do we have the right prescription? *Health Promotion International*, 14(3):271–279.

[28] McCahil, C. and Garrick, N. W. (2008). The Applicability of Space Syntax to Bicycle Facility Planning. *Transportation Research Record: Journal of the Transportation Research Board*, 2074(-1):46–51.

[29] McDaniel, S., Lowry, M. B., and Dixon, M. (2014). Using origin-destination centrality to estimate directional bicycle volumes. *Transportation Research Record*, (2430):12–19.

[30] McKenzie, B. (2014). Modes less traveled: Bicycling and walking to work in the United States, 2008–2012. American Community Survey Report ACS-26, U. S. Census Bureau.

[31] Miranda-Moreno, L. F., Morency, P., and El-Geneidy, A. M. (2011). The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections. *Accident Analysis and Prevention*, 43(5):1624–1634.

[32] Owen, A. and Levinson, D. (2014). Access Across America: Transit 2014. Technical Report CTS 14-11, University of Minnesota, Center for Transportation Studies.

[33] Raford, N. and Ragland, D. (2004). Space syntax: Innovative pedestrian volume modeling tool for pedestrian safety. *Transportation Research Record*, 1878(1):66–74.

[34] Raford, N. and Ragland, D. R. (2005). Pedestrian volume modeling for traffic safety and exposure analysis: The case of Boston, Massachusetts. In *Transportation Research Board 85th Annual Meeting*.

[35] Roll, J. F. (2013). Bicycle Traffic Count Factoring: An Examination of National, State and Locally Derived Daily Extrapolation Factors. Master's thesis, Portland State University.

[36] Schneider, R. J., Arnold, L. S., and Ragland, D. R. (2010). Pilot model for estimating pedestrian intersection crossing volumes. *Transportation Research Record*, 2140:13–26.

[37] Schneider, R. J., Henry, T., Mitman, M. F., Stonehill, L., and Koehler, J. (2013). Development and application of the san francisco pedestrian intersection volume model. Technical Report WP-2013-4, UC Berkeley: Safe Transportation Research & Education Center.

[38] Schneider, R. J., Ryznar, R. M., and Khattak, A. J. (2003). An accident waiting to happen: A spatial approach to proactive pedestrian planning. *Accident Analysis and Prevention*, 36(2):193–211.

[39] Smeed, R. J. (1949). Some statistical aspects of road safety research. *Journal of the Royal Statistical Society, Series A (General)*, 112(1):1–34.

[40] Tabeshian, M. and Kattan, L. (2014). Modeling nonmotorized travel demand at intersections based on traffic counts and gis data in Calgary, Canada. *Transportation Research Record*, (2430):38–46.

[41] Todd, K. (1992). Pedestrian regulations in the United States: A critical review. *Transportation Quarterly*, 46:541–559.

[42] Toroyan, T., Peden, M. M., and Iaych, K. (2013). WHO launches second global status report on road safety. *Injury Prevention*, 19(2):150–150.

[43] Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., and Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention*, 41(1):137–45.

[44] Wolshon, B. and Wahl, J. (1999). Novi's main street: Neotraditional neighborhood planning and design. *Journal of Urban Planning and Development*, 125(1):2–16.

[45] World Health Organization (2013). Global status report on road safety 2013: Supporting a decade of action. Technical report, World Health Organization.

[46] Yang, Y. and Diez-Roux, A. V. (2012). Walking distance by trip purpose and population subgroups. *American Journal of Preventive Medicine*, 43(1):11–19.

[47] Zhang, Y., Bigham, J., Ragland, D., and Chen, X. (2015). Investigating the associations between road network structure and non-motorist accidents. *Journal of Transport Geography*, 42:34–47.