A METHODOLOGY FOR EXPLORING THE RELATIONSHIP BETWEEN
INTERSECTION FORM FACTORS AND TRAFFIC CRASHES USING
GEOGRAPHICALLY WEIGHTED REGRESSION

By

REGINALD PIERRE-JEAN

A THESIS PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF
FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OR MASTER OF ARTS IN URBAN AND REGIONAL PLANNING

UNIVERSITY OF FLORIDA

2011

ACKNOWLEDGEMENTS

I would like to thank my committee members Dr. Ilir Bejleri, Dr. Paul Zwick and Dr. Andres Blanco who have guided me through this thesis process. I would also like my mother, wife and family for the encouragement they provided me to get where I am today. And finally I would like to thank the Geoplan crew, Kate, Crystal, Lex, Lance, Sam and Danny for knowledge and assistance provided throughout my graduate education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

9

Abstract of Thesis Presented to the Graduate School
Of the University Of Florida in Partial Fulfillment of the
Requirements for the Degree Master Of Arts In Urban And Regional Planning

A METHODOLOGY FOR EXPLORING THE RELATIONSHIP BETWEEN
INTERSECTION FORM FACTORS AND TRAFFIC CRASHES USING
GEOGRAPHICALLY WEIGHTED REGRESSION

By

Reginald Pierre-Jean

December 2011

Chair: Ilir Bejleri
Cochair: Paul Zwick
Major: Urban and Regional Planning

Crashes at intersections are a prominent and problematic traffic safety issue.

Crashes at intersections have been studied using global linear regression and before-and-after analytical methods. The Intersection form factors measured are; intersection Legs, traffic signals, traffic calming devices, corners preset, curbs present, sidewalks, percent slope, bridge intersection, park intersection, lane width, number of lanes, and traffic volume. Geographically Weighted Regression (GWR) is novel methodological approach in intersection analysis that models the relationships of these form factors to crash rates within a spatial context. GWR proves to be a more accurate modeling method overall than global linear regression. In addition to higher model performance that GWR exhibits, GWR shows the strength and variation in relationships along the data distribution for each observation. GWR also produces a visual representation of the relationships this allows for greater interpretation of the explanatory variables. In future GWR models with higher specification will produce crash rate prediction layers that will help aid in crash intersection analysis.

CHAPTER 1
INTRODUCTION

Background of the Problem

Traffic accidents (referred in this work as crashes) often occur at intersections due to multiple conflicts created by users travelling through. When studying intersection crashes, crash analysis on why they occur and how to reduce them is always a main priority. Traffic safety analysts use various analytical methods to assess the relationship of contributing causes on crashes. Contributing causes are the factors that may cause a crash. Understanding how contributing causes may lead to safer intersection design and policy decisions.

 Contributing causes can be organized in three groups. The first groups of factors are the "external contributing causes". These are situations resulting from external factors not related to driver error or intersection form. An example of external contributing causes is weather. The second group of contributing causes is driver behavior factors. For example driver used alcohol or improper driving maneuver. The third contributing cause and the focus for this thesis is the physical intersection form design. Intersection characteristics or intersection form is defined as the structural composition of traffic intersection components.

Problem Statement

"Before-and-after" studies or "with-and-without" studies are perhaps the most widespread methods used to evaluate system performance. (Levinson and Chen 2006). The (Levinson and Chen 2006) state:

"But this method (before-and-after analysis) will meet difficulties when the object of study is a long existing traffic management system. Firstly, it is usually impossible to

isolate the effects of traffic management from the effects of external variations. Before and after study is persuasive for evaluation of short-run impact under the condition that there is no significant variation in external circumstances, however, the evolution of traffic management from initialization to full operation usually covers decades."(p.4)

This statement addresses the main issue of system performance analysis, which can be applicable to intersection performance analysis. Before-and-after type analysis has certain disadvantages when analysis of the data needs to cover long and continuous periods of time. Other methods of analyzing relationships found in intersection performance analysis need to be explored.

Purpose of the study

This Study intends to explore the relationship of intersection form factors to intersection crashes using a geographically weighted regression model. What Intersection traffic form factors are the most important in reducing crashes at intersections? What factors increase the crash rate (number of crashes normalized by traffic volume) the most? How do spatial regression models compare to global regression models? Does spatial regression reveal relationships in a concise and useful manner? It is important to look at different ways of measuring the effectives of intersection design. Spatial regression may be an efficient method of analysis look at traffic intersections. This study explains the processes and techniques used to analyze this relationship using Portland Oregon data as a case study area.

Importance of the Study

Levinson & Chen state (2006) "Regression analysis is different from before and after study in that it tries to search out all the potential elements (including traffic management) that effect system performance, record their variation and use these

elements as the regression predictor variables to test the association between traffic system performance and traffic management."

This study analyzes form effect on intersection crashes using spatial regression. Evaluating the relationships of these variables on large and continuous timescales has been a deficiency in the before-and-after analysis. The findings may contribute to measuring the causational effect of physical intersection characteristics on crashes at intersections.

In order to determine the characteristics of intersection form and how they relate to

crashes this study will first look at supporting literature related to intersection crash

factors, intersection form and intersection analysis methods.

Intersection definition

An intersection is defined as "the general area where two or more roadways join or

cross, including the roadway and roadside facilities for traffic movements within the

area". (Highway Safety Manual 2010) An intersection is defined "by its functional and

physical areas. (Highway Safety Manual 2010) The physical area is shown in Figure 2-1

it consists of the area where the road segments occupy the same area. "The functional

area denoted by the Figure 2-2 contains the basic elements: Decision distance,

maneuver distance, and queue storage distance.
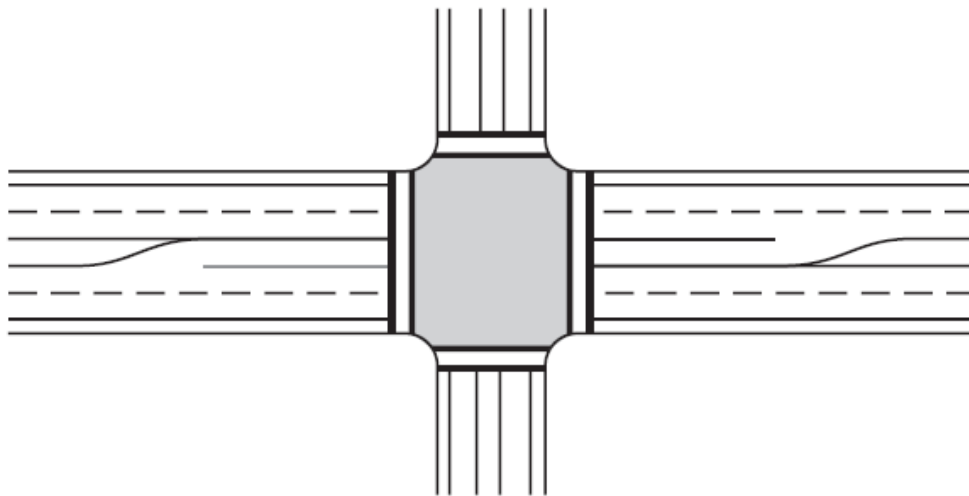
Figure 2-1. Physical Intersection. Highway Safety Manual. (2011)
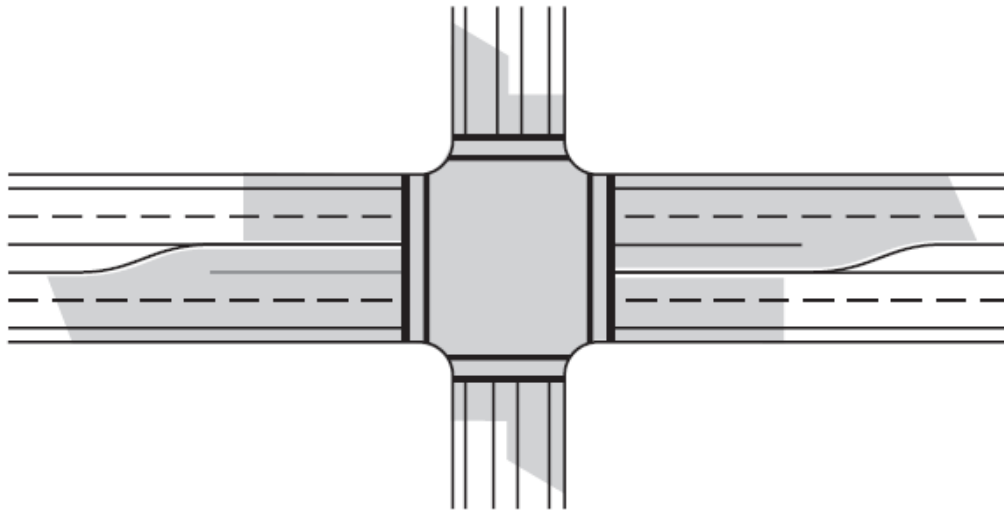
Figure 2-2. Functional Intersection. Highway Safety Manual (2011)
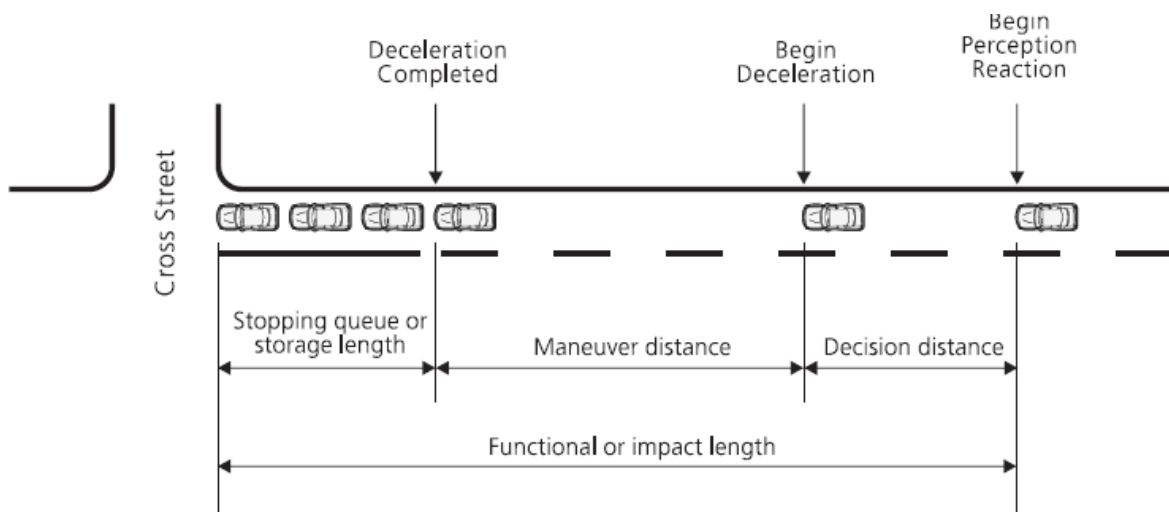
Figure 2-3. Elements of the Functional Area of an Intersection. Highway Safety Manual. (2011)

An intersection crash is somewhat harder to define as there is no absolute consensus about its definition. Many agencies define it as any crash that has occurred within the physical intersection area. Other agencies consider all crashes a certain distance from the physical intersection like 250 feet. Not every crash occurring within a 250 foot distance of an intersection is considered to be at an intersection crash.

(Highway Safety Manual 2010) In almost every case any crash that occurs within 100 ft. of the physical intersection is considered to be an at intersection crash. This is the default measurement used when relating crashes to intersection.

Intersection studies are often done in context of fatalities. Fatalities are usually the priority when researching automobile crashes. As we review traffic issues relating to intersections we will see this as a driving concern for improvement of intersection design and safety. There are two main categories of intersections, unsignalized intersections and signalized intersections. Unsignalized are defined as intersections with stop signs, or intersections where the driver must judge whether stopping is appropriate (such as roundabouts or yield signs). Unsignalized intersections are unique because traffic flow is directed by driver action. These actions vary and are unpredictable because of the lack of structured signaling. According to 2009 data from National Highway Traffic Safety Administration General Estimates System (NHTSA-GES) Fatality Analysis Reporting System (FARS) unsignalized intersections account for half of the fatalities observed in urban intersection crashes.

Table 2-1. Steps used by Drivers to Negotiate Unsignalized Intersections. National Highway Traffic Safety Administration (1994)

1. Detect the presence of the intersection during an approach
2. Correctly identify signage
3. Anticipate Sudden deceleration from lead vehicle(s)
4. Detect presence of cross traffic
5. Recognize crash hazards posed by cross traffic, perhaps by estimating the speed, acceleration, and distance of the approaching vehicles
6. Watch for and anticipate other traffic or pedestrians that may cause a cross traffic vehicle to suddenly stop in the SV travel lane.
7. Identify problems that might obstruct the driver's vision and attempt to overcome such problems
8. Stop the vehicle
9. Estimate when it is safe to proceed through the intersection

Traffic signals are by far the most common feature observed at signalized intersections. Traffic signals are automatic indication devices used facilitate the flow of vehicles, pedestrians and bicyclist automatically using signal timing techniques. Traffic signals play a significant role in achieving safer performance at intersections. Previous research has shown that in certain situations traffic lights will reduce the number and severity of crashes. (Rodegerdts, Nevers and Robinson 2004)

Table 2-2. Summary of motor Vehicle Crashes related to Junction severity In the United States during 2002

	Total Crashes		Fatalities/Injuries	
	Number	Percent	Number	Percent
Non-Intersection Crashes	3,599,000	57	1,022,549	52
Signalized Intersection Crashes	1,299,00	21	462,766	24
Non-Signalized Intersection Crashes	1,418,000	22	481,994	25
Total	6,316,000	100	1,967,309	100

Source: Adapted from table 28 of Traffic Safety Facts 2002 (p.50) National Highway Traffic Safety Administration (2004) http://www-nrd.nhtsa.dot.gov/Pubs/TSF2002.pdf

According to the Fatality Analysis Reporting system (FARS) and the National Automotive Sampling System-General Estimates System about 40% of the crashes were intersection-related crashes. (Choi 2010) Intersections constitute a small part of transportation networks, yet a large amount of crashes are concentrated at them. It is fairly obvious that there are conflicts of travel direction and many other factors that explain this concentration of crashes. Around 50% of total fatal crashes happen at intersections. The majority of these fatalities happen on unsignalized two lane roadways traveling at moderate speeds ~55MPH. Most crashes happen between 3pm and midnight. Automobile fatalities are the leading cause of death for people aged fifteen to forty four. (National Highway Traffic Safety Administration 2009) Intersections represent

a disproportionate share of the traffic safety issues and should be looked at as a priority

for analysis. (Federal Highway Administration 2009)

Table 2-3. 2007 National Intersection Crashes Federal Highway Administration Office
of Safety (2009)

	Number	Percentage
Total Fatal Crashes	37,435	
Total intersection and intersection related crashes	8,061	21.5%
Total Injury crashes	1,711,000	
Total Intersection and intersection related injury crashes	767,000	44.8%
Total property Damage Only (PDO) Crashes	4,275,000	
Total intersection and intersection related PDO crashes	1,617,000	37.8%
All Crashes	6,024,000	
Total intersection and intersection related crashes	2,392,061	39.7%

Intersection Safety

"Injury and fatality statistics for highway intersections and interchanges are ample

evidence that strategies to improve the safety of these crash-prone areas are urgently

needed. On average, there are five crashes at intersections every minute and one per

son dies every hour of every day at an intersection somewhere in the United States."

(AASHTO 2004)

Federal Highway Administration (FHWA) as identified 4 areas of primary focus in

improving safety and reducing crashes. These areas of focus are Intersections,

Roadway Departure, Pedestrians, and speeding. Intersections account for 21% of crash

fatalities and around 53% of all crashes from 2002-2006 are associated with

intersections. (Federal Highway Administration n.d.)

18

An intersection safety design has to fulfill dual objectives: mobility and safety.

These objectives conflict more often than not and some feature need to be

compromised for the sake of mobility. Balancing the efficient operation and safety of an

intersection is top priority. The most common types of crashes at intersections help

engineers prioritize safety elements to reduce crashes. Below are the most common

crash types at intersections for signalized and unsignalized intersections and their

respective suggested countermeasures according to a recent FHWA study 2004.

Figure 2-4. Angle Crashes. Federal Highway Administration

Angle crashes

Account for 42% of fatal crashes at signalized intersections potential

countermeasures are as follows:

- optimize change intervals
- improve sight distance
- restrict access
- provide targeted enforcement

- restrict parking

Account for 53% of fatal crashes at unsignalized intersections potential

countermeasures are as follows:

- clear sight triangles
- improve awareness of intersection
- apply access management
- improve turn lane design
- construct roundabouts
- construct acceleration lanes
- close/relocate intersections
- reduce/eliminate skew
- post appropriate speed limits
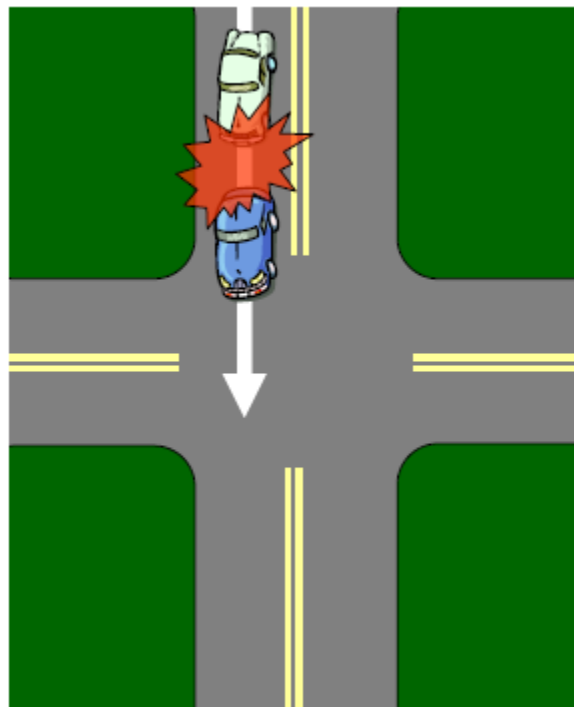
Figure 2-5. Rear end Crashes. Federal Highway Administration

Rear end crashes

Account for 8% of fatal crashes at signalized intersections potential

countermeasures are as follows:

- increase visibility of intersection and/or traffic signals
- increase awareness
- improve signal coordination
- install turn lanes
- control approach speeds
- optimize change intervals

Account for 6% of fatal crashes at unsignalized intersections potential

countermeasures are as follows:

- install turn lanes
- supplemental overhead signing
- provide shoulder bypass lanes
- provide pavement markings
- provide right-turn acceleration lanes
- provide left-turn acceleration lanes
- provide lighting
- post appropriate speed limits
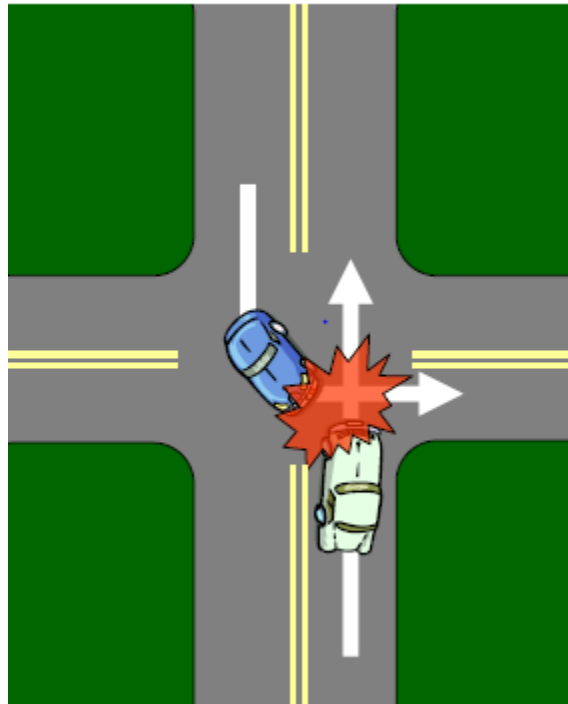
Figure 2-6. Left Turn Crashes Federal Highway Administration

Left Turn Crashes

Account for 21% of fatal crashes at signalized intersections potential

countermeasures are as follows:

- employ protected left turn phasing
- implement turn restrictions
- improve turning lane design
- reconstruct approaches
- improve sight distance
- improve signal coordination

Account for 8% of fatal crashes at unsignalized intersections potential

countermeasures are as follows:

- improve turn lane design
- implement turn restrictions
- use indirect left turn treatments
- provide lighting
- clear sight triangles
- provide left turn acceleration lanes
- construct roundabouts
- close/relocate high-risk intersections
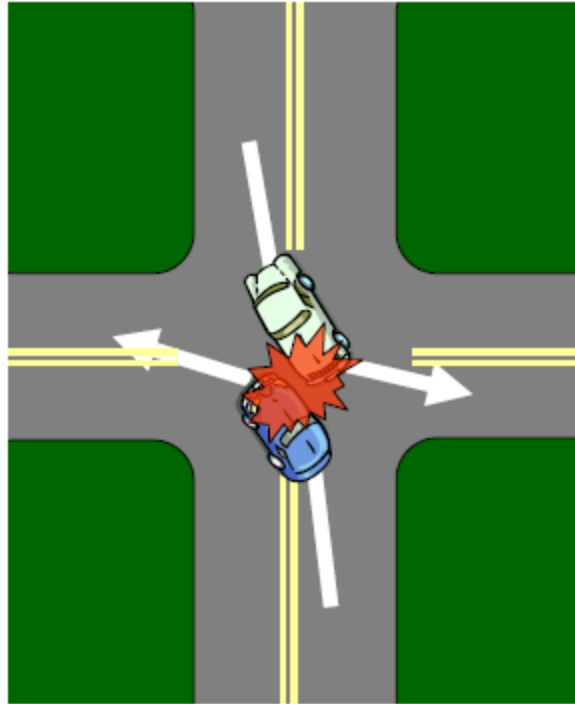
Figure 2-7. Sideswipe Crashes. Federal Highway Administration

Sideswipe Crashes

Account for 13% of fatal crashes at signalized intersections potential

countermeasures are as follows:

- install pavement markings
- provide protected left turn phasing

Account for 2% of fatal crashes at unsignalized intersections potential

countermeasures are as follows:

- install pavement markings
- provide lane assignment signing or marking
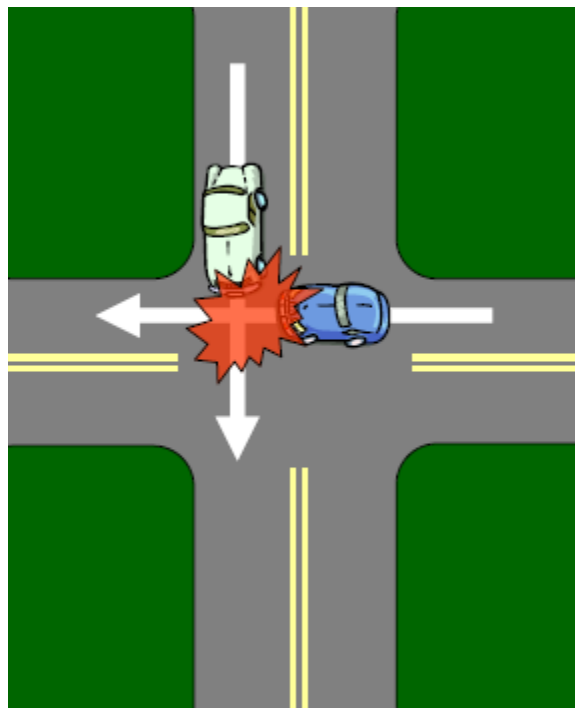- provide right-turn acceleration lanes
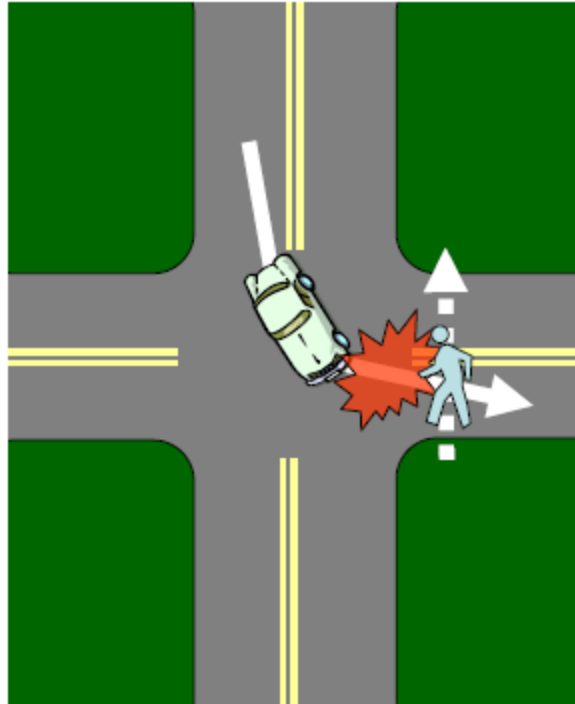
Figure 2-8. Pedestrian/Bicycle Crashes. Federal Highway Administration

Pedestrian Crashes

Account for 25% of fatal crashes at signalized intersections potential

countermeasures are as follows:

- improve signal hardware
- improve pedestrian/bicycle facilities
- provide information and education

Account for 14% of fatal crashes at unsignalized intersection crashes potential

countermeasures are as follows:

- improve pedestrian/bicycle facilities
- provide traffic calming
- information and education

Crash Reduction Factors

Crash reduction factors (CRF) are the percentage reduction that might be expected after implementing a given crash countermeasure. In some cases the CRF can be negative meaning that the counter measure leads to an increase in crashes. A CRF can be regarded as a general estimate of affect. As long steps are taken to ensure that the countermeasures apply to the condition being measure. (Hovey and Chowdhury 2005, 14) Before-and-after and cross-sectional study methods are two of the methods used to develop CRF's. In a before-and-after study, the safety effect of a countermeasure is determined by the difference in the number of crashes occurring before the after study; the cross-sectional approach usually uses regression methods to estimate crash frequencies from a large sample of roadway segments whose design attributes vary systematically. Such regression models estimate the marginal effects of changes in highway design attributes on crash frequencies.[1] (Shen and Gan 2007) Before and-After-Method. This more widely used approach to CRF development uses data before-and-after countermeasures are implemented. Crash count data for a 2 -3 year period is used. Data during the construction of the countermeasure is removed from that time period. A few factors must be considered between the periods also. Vehicles miles is needed to calculate exposure, traffic volumes should be equal, traffic composition should be similar and variation in crash data should not vary any more that 20 percent.

---

[1] The key difference between before-and-after and cross-sectional studies is not in the different methods used to analyze the data (as a matter of fact, they can be similar) but rather in the different concept of how to investigate the safety effect. In the before-and-after study, the idea is to investigate these locations where a given improvement has been applied within the period of analysis, while for the cross-sectional analysis; the investigated locations do not experience any major changes within the period of analysis. Thus, the before-and-after study focuses on the changes in safety over time, while the cross-sectional analysis focuses on the differences in safety between locations. (Tarko, Eranky and Sinha 1998)

The concept of the simple before-and-after study is straightforward. It is based on the assumption that if nothing has changed, the crash experience before improvement is a good estimate of what would have happened during the after period without improvement. The formula for deriving a CRF based on this method is:

$$\text{CRF} = \frac{(N_b - N_a)}{N_b} = 1 - \frac{N_a}{N_b}$$

Figure 2-9. Crash Reduction Factor Formula. Transportation Research Record 1840 (2007)

Nb and *Na* are the numbers of crashes at a treated site before and after the improvement took place, respectively.

According to Shen & Gan (2007) study;

Before and after method is subject to some pitfalls. These factors are the Regression to the mean issue, crash migration issue, maturation and external casual factors. Regression to the mean is a statistical phenomenon that occurs whenever a nonrandom sample is selected from a population. In this case the crash intersection. Since intersections with large crash counts are usually selected, these areas tend to have higher reductions than seen on average, even without any treatments, the crash frequencies would likely be reduced simply because the number of crashes at the sites tends to regress or return to the long-term mean number of crashes.

The Empirical Bayes (EB) method has been developed to adjust for the regression-to-the mean bias. The method is based on the following three assumptions:

- The number of crashes at any site follows a Poisson distribution.

- The means for a population of systems can be approximated by a gamma distribution.

- Changes from year to year from different factors are similar for all reference sites."(pp. 51)

Also noted by Shen & Gan; crash migration is the movement of the observed crashes to a new location or situation. This can be witnessed geographically and non-geographically. Crashes can geographically move to a new location. Or crashes can move to a new situation i.e. light pole installed reduced night time crashes but increased fixed object crashes. Maturation is the change in crash counts due to changing trends in the economy weather or traffic flow. Maturation is an issue that can have an effect on analysis if not accounted for. External Causal Factors are factors that affect the crash counts that are hard or cannot be measured for precipitation and economic conditions cannot be measured for their affect easily. There are some external factors that can be accounted for such as new development causing higher vehicle conflict exposure etc., but these situations are hard to measure nonetheless. Despite the potential problems the before-and-after method is the widely accepted form of CRF development. (Shen and Gan 2007)

The US Department of transportation provides a set of tables that show the CRF's for many crash countermeasures. The CRF's are displayed along with a standard error that is based on the standard deviation of the error in the estimate.
An example is shown in Table 2-9;

Table 2-4. Crash Reduction factor Adapted from Desktop Reference for Crash Reduction Factors (2007) Federal Highway Administration

Countermeasure	Crash type	Intersection type	CRF	CRF[std error]
Install right turn lane	Right turn	4 - Leg	50	5
Remove left turn lance	all	4 - leg	-45	10

In the Table 2-9 the first row shows us that for the countermeasure of installing a right turn at 4-legged intersections has a 50% reduction in right turn crashes with a standard error of 5 percent. Alternatively the second row tells us that if we remove a left turning lane from a 4-legged intersection of any crash type we will witness a 45% increase in all crash types with a standard error of 10 percent.

Intersection Form Elements

Intersections are designed with safety and ease of movement in mind. Designs of intersection vary from location to location based on the functional classification of intersection; the traffic volume observed by the intersection and land use characteristics it the vicinity. Intersection design should facilitate mobility all users of the intersection in a clear and safe manner. These are the principles of intersection design.

"Given that goal, it is sobering to realize that "in 2002, more than 9,000 Americans died and roughly 1.5 million Americans were injured in intersection related crashes. In economic terms, intersection related crashes in the year 2000 cost about $40 billion." (Institute of Transportation Engineers 2004)

Intersection Design Principles

At grade intersections are junctions where two or more axes cross at the same level. Like other highway feature safety cannot be achieved through design alone it requires collaboration between traffic controls, planning officials and traffic enforcement. The five general topics that are considered during traffic design according to the Toolbox on Intersection Safety and design (2004) are;

- Human factors: Drivers and Pedestrian actions like reaction time and interpretation

- Road way uses: This includes traffic volume and characteristics of users at intersections

- Physical elements: Topography and intersection angle or environmental factors.

- Economic factors: This includes factors like cost of intersection or effect of intersection on landuse.

- Functional Intersection areas: Includes intersection legs and width and size of intersection. Indicates complexity.

These elements compose the components of both physical intersection and the larger functional intersection. And lead to a decision on what design elements will be incorporated into the intersection. Some significant elements of intersection design are approach angle, traffic signals corners/curbs, gradients/percent slopes, lane widths, number of lanes, traffic calming devices, angle/skew of intersection and corner radius.

CHAPTER 3
METHODOLOGY

Research Design

In this section the methodology will be explained along with the reasoning behind the methods chosen. The objective of this study is to analyze the relationship between intersection form factors and intersection crashes. Portland Oregon is chosen as the study area due to data availability and quality data.

The strength and direction of the relationships are illustrated using spatial regression modeling. Eleven variables represent the intersection form factors; Traffic Volume, Intersection Legs, Traffic Signals, Sidewalks Curbs, Corners/Curb ramps, Traffic calming, percent slope, Bridges, Number of lanes and Lane width. The relationship of these explanatory variables is revealed in the form of coefficients created by the regression model.

Two regression models were used to model the characteristics of crash counts. The first model used is the Ordinary Least Squares (OLS) regression model that predicts global model dependent to independent variables. The second model, a geographically weighted regression (GWR) predicts the relationships of each observation using a local regression calculation. The modeling process uses the intersection node feature layer that has been processed to contain the explanatory variables that relate to each of the intersection characteristics. The OLS results are compared to the GWR results and independent-dependent relationships are interpreted from the results.
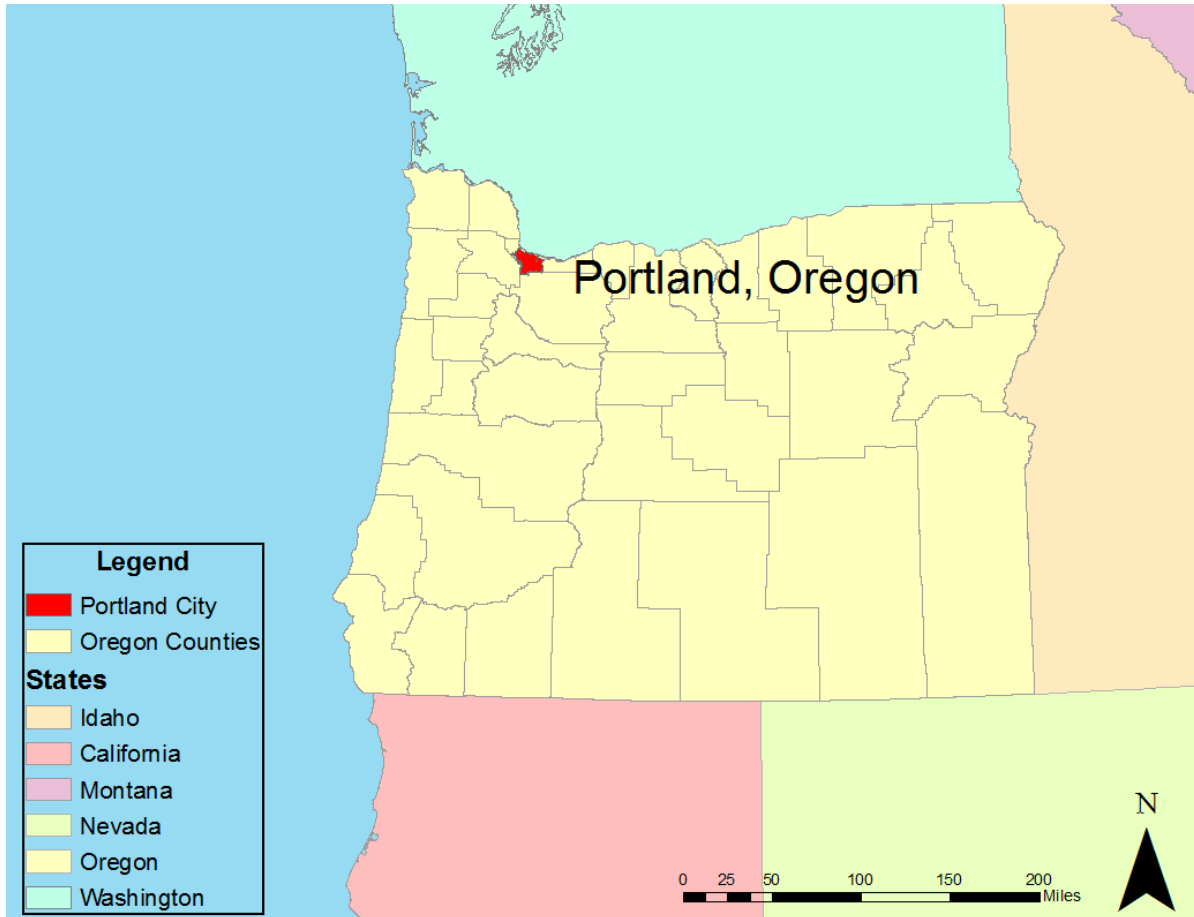
Figure 3-1. Portland Oregon

Geographic Setting

Portland, Oregon is a populous and dense city in the Pacific Northwest. It is a part of the greater Portland metropolitan area which also includes Vancouver, Gresham and Hillsboro. As of 2010 it is the most populated area in Oregon with an estimated population was 583,776. (U.S Census Bureau 2010) Portland is known for its strong land-use planning and dense urban environment. (Orski 2003) The Willamette River runs north through the city center. Major interstates I-85 and I-5 also join in Portland.

Transit Oriented development[1] plays a huge role in Portland's transportation network. Mixed use and high density development is promoted throughout the city. (Oregon Metro Council 2011) Portland is known for having many one way roads down town narrow roads this is known as having an alternating grid pattern. This pattern is not universal throughout the city but mainly it is the case. In TTI's 2011 Urban Mobility Report Port land was found to have the 13th worse travel time Index in the county, while having the fewest road network users per capita. This means that Portland's congestions makes commutes to 123%longer than expected. (Schrank, Lomax and Eisele 2011)

Portland maintains a very detailed GIS street network database. This includes data like roads, bridges, automobile, crash points and many other features that allow for a complex representation of the street network. This was the most important factor in choosing this location for the study. Also it is important to note Portland's variety in street layout which allows for a robust study of many different types of intersections.

Data was sourced from Portland's many online GIS data warehouses. The key data layers that were sourced were the Portland streets GIS layer and Portland automobile accidents for 2007 to 2008. These layers are the basis of the study and much of network dataset is calculated from spatial relationships between these two layers. Next any data relating to intersection infrastructure was added to the database for the data preparation phase.

---

[1] Transit-oriented development (TOD) is a mixed-use residential or commercial area designed to maximize access to public transport, and often incorporates features to encourage transit ridership.(Federal Transit Administration)
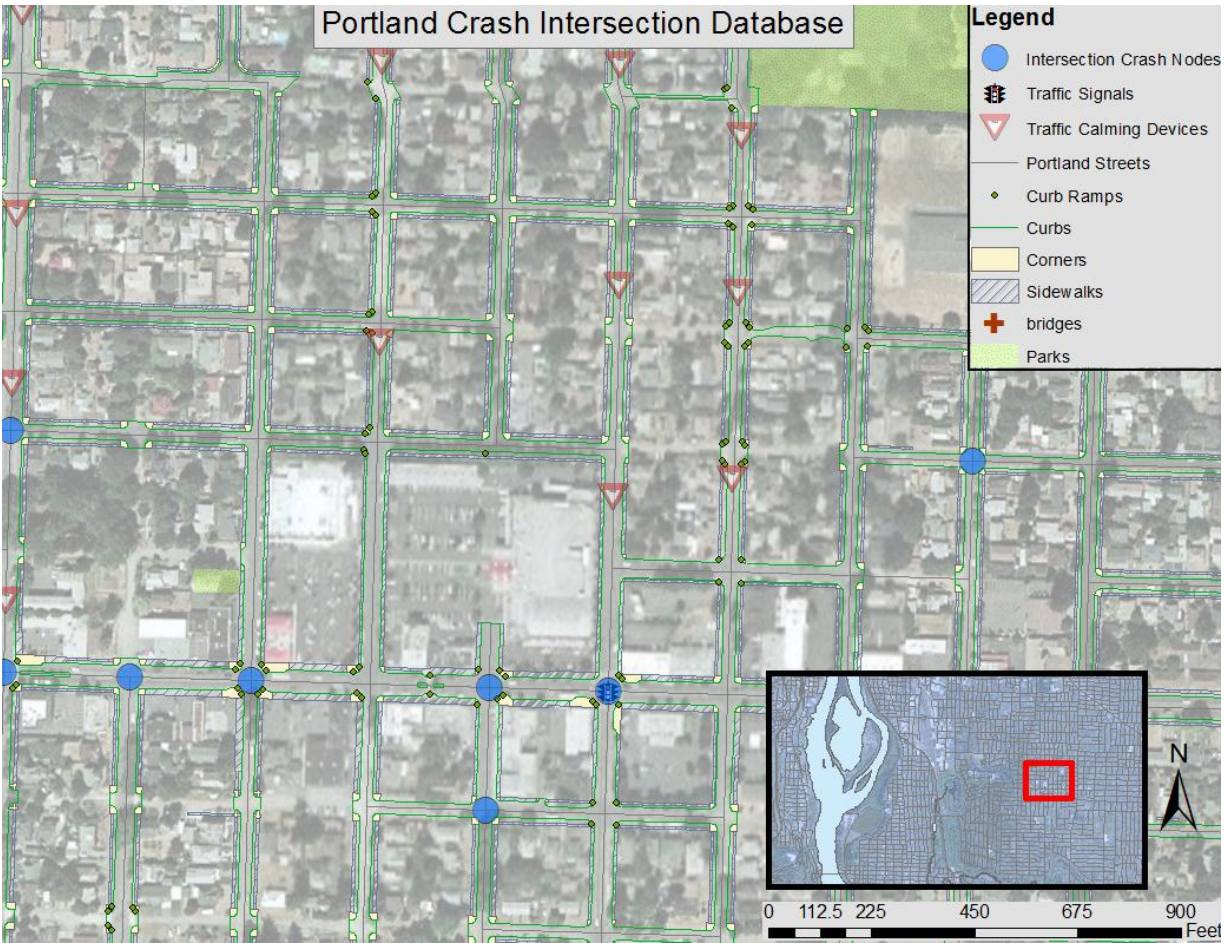
Figure 3-2. Intersection Database

Data Preparation

The goal of data preparation is to organize the data into a form suitable for analysis. This involves preparation of the intersection database and association of the relevant attribute with each intersection. Each variable used in the regression model is defined in this section.

Intersection Database Creation

In order to run the statistical model the data has to be aggregated into one single feature layer. This was done by first creating a network dataset out of the Portland streets segments. Nodes are then calculated for each segment and a node segment

association table is created. Intersections are defined where 3 or more segments share a node. The intersections layer is included into the GIS database along with all the related Portland transportation data.

Crash rate

It is known that crash counts increase with increasing traffic volume; however this relationship has only proven to be consistent for up to about 20,000 vehicles per day (Vpd).Volumes above 60,000 Vpd have a complex and erratic effect on crash count. The crash counts are high in some areas and much lower in others compare to crash counts taken at Vpd's lower than 20,000. (Joksch and Kostyniuk 1998)

This contradicts our perception that crash counts should vary simply with traffic volume. We also cannot fully assume that the deviations between crash points are due only to the intersection characteristics since traffic volume has a hard to measure relationship with intersection traffic crashes. A potential explanation for this phenomenon is that there might be intersection feature that are more common at high volumes. Another is that different crash types may have a higher potential of occurring at different traffic volumes.

How do you represent the relationship between crash counts and traffic volumes for the more mild traffic volumes? One way we can do this is to smooth the crash counts based on the traffic volume estimated at the intersection. This would require fitting a Gaussian kernel and a moving window bandwidth to smooth the various data points. (Joksch and Kostyniuk 1998) Another way to normalize the relationship between crash count and traffic volume is to calculate a crash rate. This method is the one used in most traffic safety studies. Crash rate is a measure of crashes in a given period of time against traffic volume observed; it is expressed in "crashes per million entering

Vehicles" (MEV).Crash rate is an important measure to identify locations that should be given priority for traffic safety improvement. According to MassHighway DOT,(2011) the formula for calculating the crash rate for an intersection is presented below, which is standard to the Traffic Engineering profession:
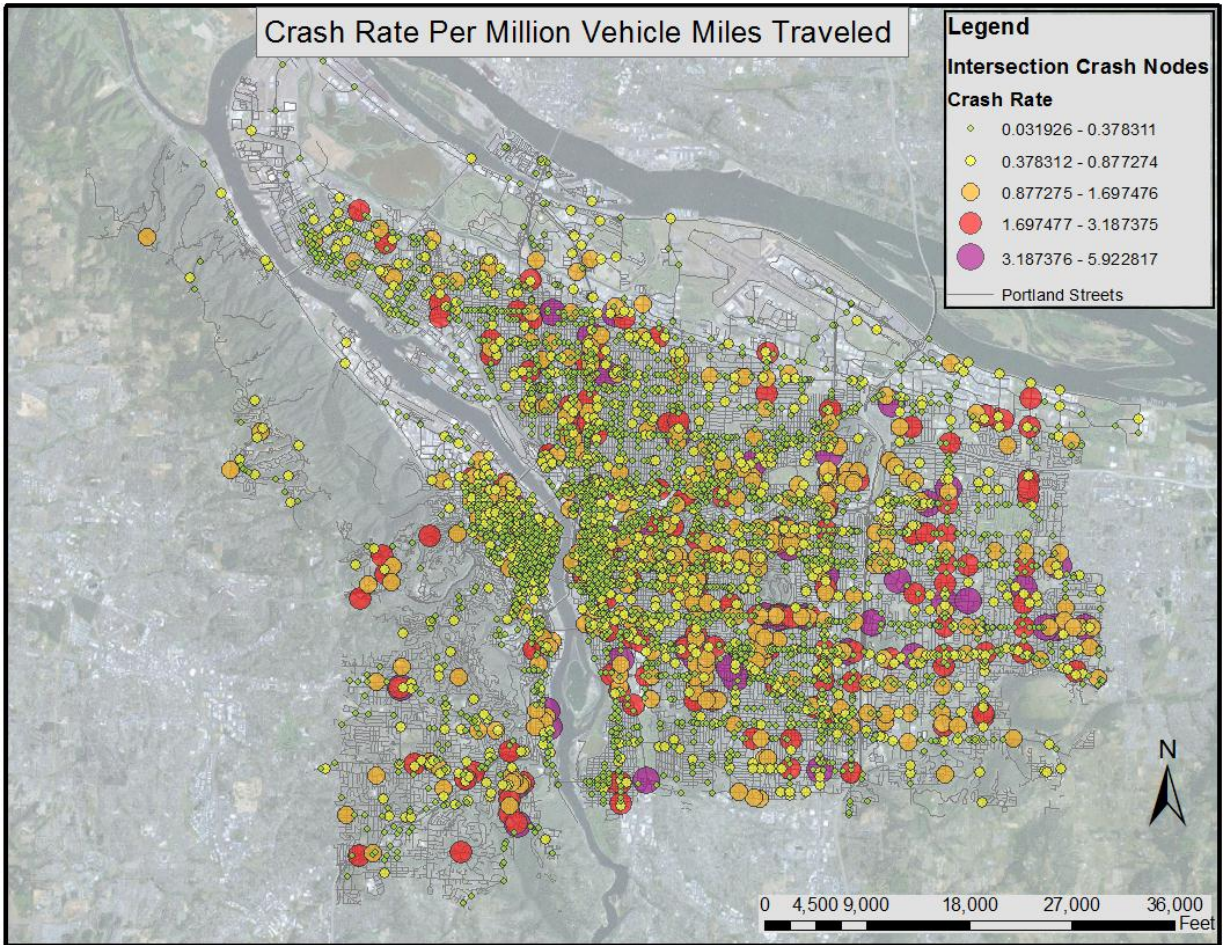
Figure 3-3. Crash Rate Portland

$$R = \frac{C * 10^6}{Tv * 365 * 2}$$

(3-1)

- R is crash rate calculated in MEV
- C is the number of crashes at the intersection
- Tv is the daily traffic volume at intersection

- 365 is the amount of days the crash data covers
- 2 is the amount of years the data covers

Intersection Form Explanatory Variables

Buffer zones were created around all intersection nodes using a radius 100 feet. These buffer zones were used to select the intersection form features associated with the physical and functional intersection. The ArcMap 9.3 Geoprocessing tools Spatial Join, Select by Buffer Distance and near tool were used to append the explanatory variables to the intersection nodes.
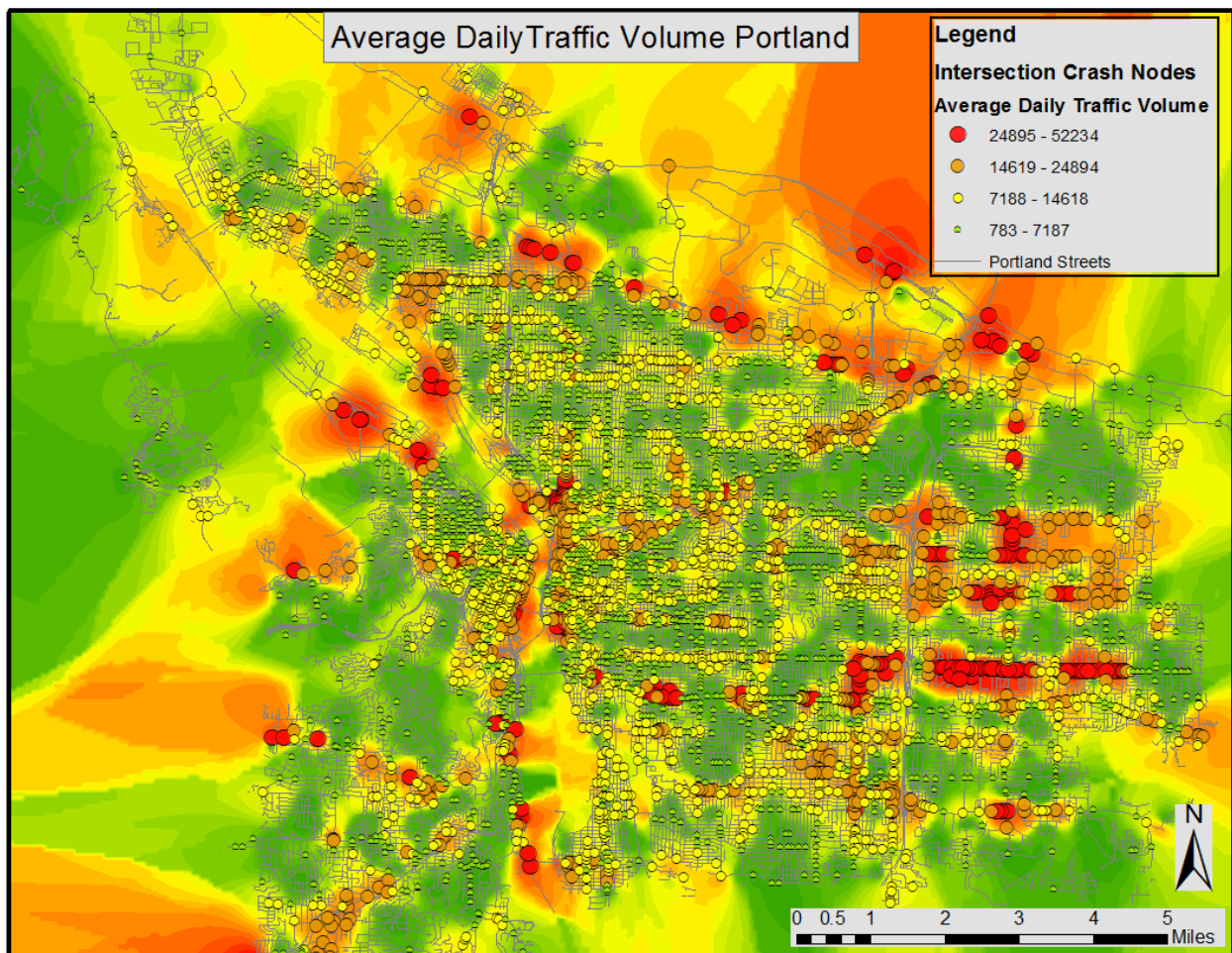
Figure 3-4. Traffic Volume

Traffic Volume at Intersection

Traffic volume is defined as the average daily combined bulk of directional travel coming into the intersection. This is the most common type of traffic count. In areas with a lot of multi-axle trucks the number recorded by the traffic counter can be larger than the actual number of vehicles. Volume counts are typically 24 hours in length. The traffic volumes given in the Portland dataset are given with the total volume already calculated. In order to calculate crash rate at each intersection a traffic volume surface was produced.

The traffic volume point's layer was recalculated into a surface by inverse distance weighted calculation. This multivariate interpolation process uses the known traffic volume points to calculate traffic volume for intersections that are not known. Considerations for transportation network relationships were taken into account to create the best possible surface calculation for traffic volume. Parameters like search radius and power were manipulated to get the best surface possible.

Figure 3-5. Intersection Legs Photograph Credit and Copyright:
www.portlandmaps.com, (2004)

Intersection Legs

Intersection legs are all the road segments that intersect the center of an

intersection.

> While the geometry of various types of intersections may vary, the
> complexity of an intersection increases with an increasing number of
> approach legs to the intersections. The number of potential conflicts for all
> users increases substantially at intersections with more than four legs.
> (Rodegerdts, Nevers and Robinson 2004)

Refer to Figure 3-6 for an example of the type on conflicts possible at 3 and 4-leg

intersections. These examples show how the conflict points increase as intersection leg

count increases. (Rodegerdts, Nevers, & Robinson, 2004)
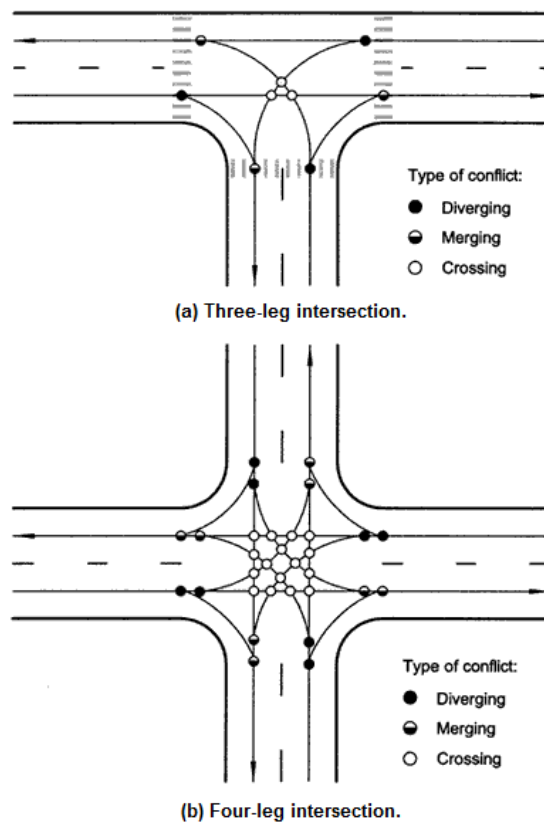
Figure 3-6. Potential Conflicts at Intersections. Rodegerdts, Nevers, & Robinson, 2004)

In intersection design five or more intersection leg implementations are to be

avoided where ever possible. (Fitzpatrick, Wooldridge and Blaschke 2005) Intersections

get more complex as the number of intersection approach legs increase. With an increase in the potential number of conflicts for all users this causes, an increase in crash count is expected.

Figure 3-7. Typical Traffic Signal Intersection Pattern (Rodegerdts, Nevers, & Robinson, 2004)

Traffic Signals

Traffic signals are used to assign vehicular and pedestrian right-of-way. They are used to promote the orderly movement of vehicular and pedestrian traffic and to prevent excessive delay to traffic. (Federal Highway Administration 2009) All signalized and unsignalized intersections in Portland are included in the GIS database for this study. In addition to promoting orderly movement of traffic, traffic signals also increase capacity of intersections, reduce the frequency of some types of crashes, and provide periodic continuous movement along a route. Also traffic signals interrupt heavy traffic to allow other to permit other traffic types. Traffic signals are expected to have reduced crashes at intersections compared to unsignalized intersections.
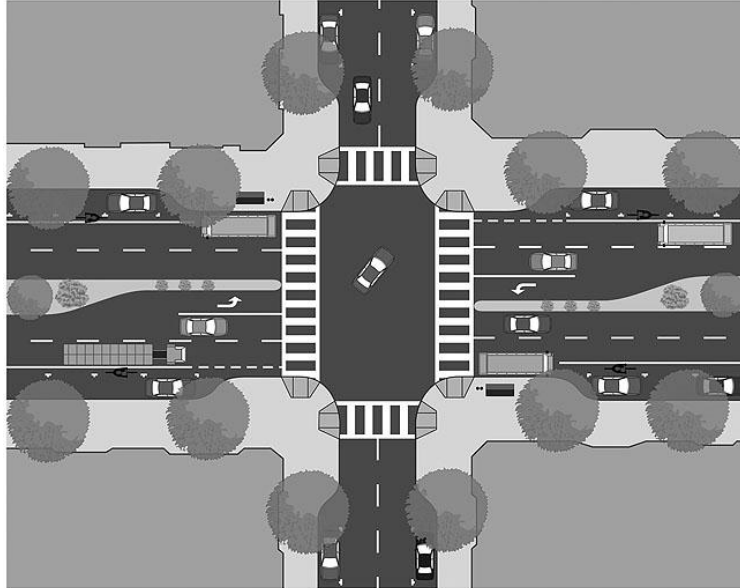
Figure 3-8. Sidewalk Crossing at Intersection. Institute of Transportation Engineers, 2010

Sidewalks

Sidewalks and paths are the primary travel way for pedestrians as they approach

intersections in developed suburban and urban districts. Properly designed sidewalks

provide mobility, accessibility and safety to all users. (Institute of Transportation

Engineers 2004) The sidewalks polygon feature layer is part of the Bureau of

Transportation's Pedestrian system. Although sidewalks are designed to provide

accessibility, mobility, and safety they also increase the complexity at intersections.

Therefore an increase in crashes is likely.
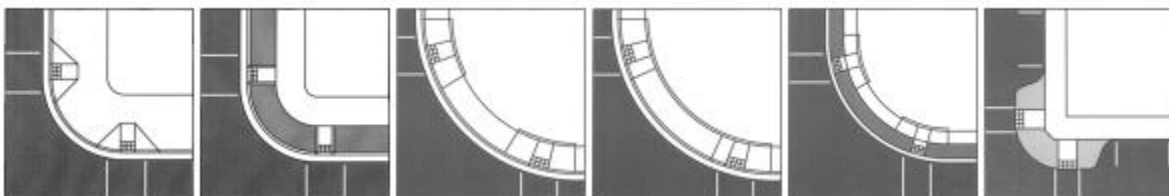
Figure 3-9. Types of Curbs (Rodegerdts, Nevers, & Robinson, 2004)

Curbs

This linear feature layer is part of the Portland bureau of transportation's sidewalk system. Curbs facilitate pedestrian movement throughout an intersection. Curbs also control where the conflict points are in left and right turning crashes. Curbs create delimitation between the roadway and sidewalk. Curbs are also part of the pedestrian and bicycle network which facilitates the traversal of the intersection. Curbs create an increase in complexity at intersections higher crash counts are plausible in the pedestrian and bicyclist crash types.
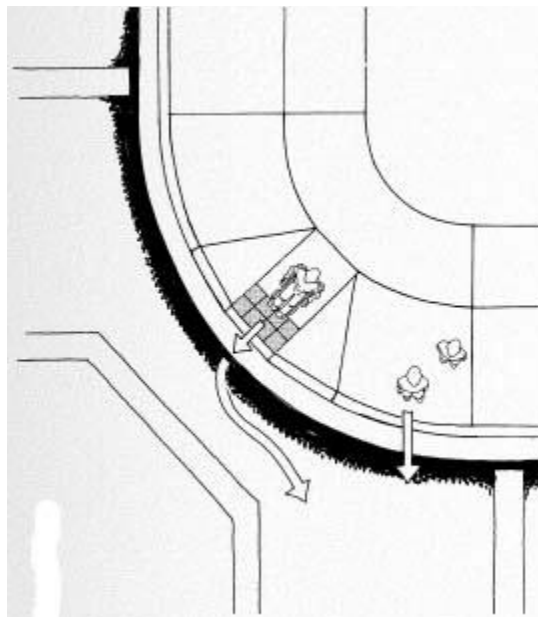
Figure 3-10. Curb Ramp. Designing Sidewalks and Trails for Access (2001)

Corners/Curb Ramps

Curbs are generally designed with a gutter to form a combination curb ramp and gutter section. They are used to provide drainage control and to improve delineation of the roadway. Curbs are used extensively on all types of urban highways with design speeds less than 50 mph. (Institute of Transportation Engineers 2004) This point feature layer is part of the Bureau of Transportation's pedestrian and road system. Curb ramps

(crosswalks) are the point of traversal for most pedestrians and bicyclist, because of this it is probable that an increase in crashes will be witnessed. This increase may be offset in the fact that gutters are also associated with this feature. Gutters increase the safety in wet road conditions by drainage and may offset the increased crash counts witnessed from curb ramps.
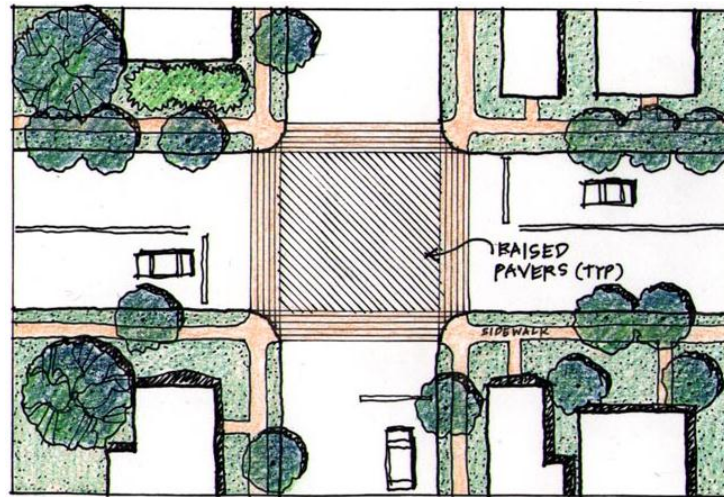
Figure 3-11. Raised intersection Drawing by B. Kim Erslev © (2003) (Beneficial Designs Inc. 2001)

Traffic Calming

Traffic calming is actions or devices to reduce vehicular traffic's intrusion into and its effects on urban life. Speed humps or textured road surfaces. This point feature layer includes a variety of devices whose purpose is to address the negative impacts of traffic in areas by reducing speeding, reducing traffic volumes and making the streets friendlier to pedestrians and bicyclists. (New York City Department of Transportation 2003) Traffic calming devices are specially implemented devices to safeguard an intersection by reducing vehicle speed in and increasing driver awareness at intersection. Traffic calming devices are expected to lower crash counts at intersections.

Figure 3-12. Bus stuck on Baxter Street drop-off (Photo credit: Ingrid Peterson)

Percent Slope

The recorded ground slope of intersection. (5% intervals) At-grade intersections

have a slope of 0%. The gradient of the intersection may reduce visibility at

intersections, which in turn may increase conflicts of the users.
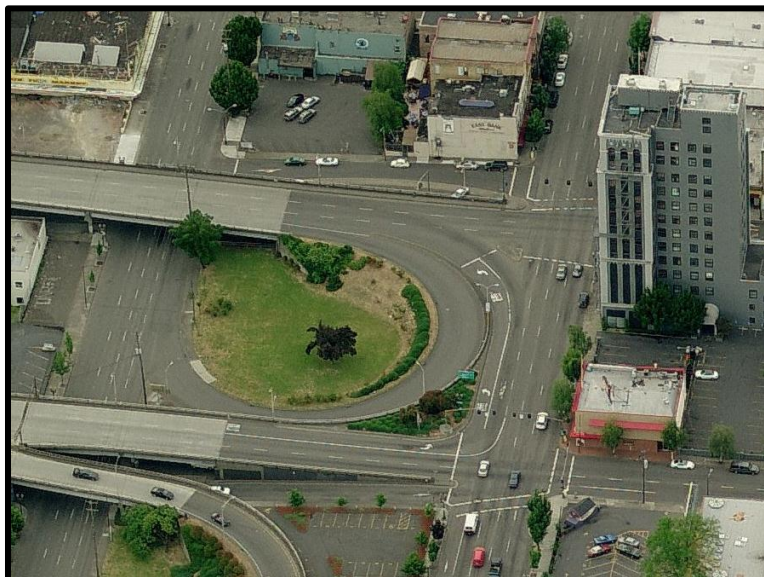
Figure 3-13. Bridges Association with Intersection (Photo credit: Google Maps Imagery)

Bridges

Bridges associated with intersections. This point feature layer includes bridge

structures which are part of an intersection leg. Bridge intersections are complex

features usually associated various signage and high traffic volume. These factors may increase crashes at intersections because of the confusion and complexity that may arise in the driver decision making process.
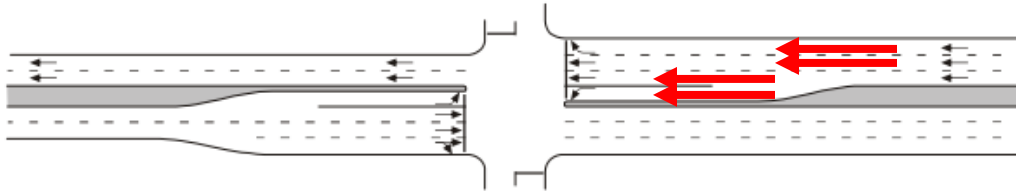
Figure 3-14. Red arrows Denote Number of Lanes. Adapted from Urban Intersection Design Guide Volume 1

Number of Lanes

Number of lanes of intersection legs section observed at intersection. Large number of lanes may be indicative of high vehicle traffic along with more vehicle lane switches; this may lead to an increase in crash count.
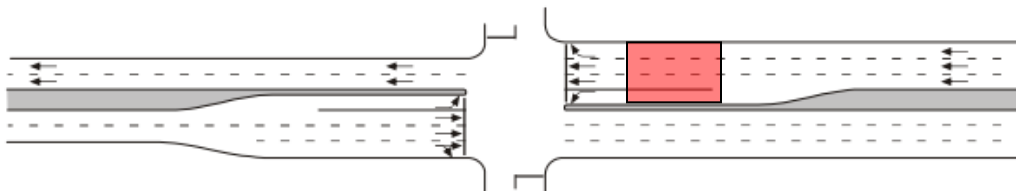
Figure 3-15. Red area denotes Lane Width. Adapted from Urban Intersection Design Guide Volume 1

Lane width

Width of intersection legs section observed at intersection. The width of the lanes can be expected to increase crash count just as number of lanes may. Lane width may also be an indicator of high traffic volume which is known to increase crashes up to a certain point.

Park Intersection

A park intersection is an intersection that has a park associated with one or more of its corners. Park intersections may present distractions, high pedestrian or bicycle conflicts which may inadvertently increase the amount of crashes witnessed.

Assigning attributes to intersections

Crash rates were selected and assigned an intersection node using buffers. Crash locations within a 100 foot buffer were aggregated to the intersection node table. Each intersection form layer was spatially joined to its nearest intersection and the attributes added to the intersection node database table. An example of an intersection form variable is "Number of lanes". This variable was created using a pavement type feature layer. Many different approaches were utilized in aggregating the variable data to the intersections nodes feature layer. Spatial joins and field calculations allow the data to be transferred over to intersections of analysis.

Figure 3-16. GIS Intersection Overview

The result of this process was a crash intersection node layer that represents the form of each intersection (explanatory variables) and contains intersection crashes. This feature layer is a subset of the intersection nodes layer except it contains only the intersection nodes that have crashes associated.

Strict criteria are used to maintain data consistency for the analysis. These criteria are used to trim and standardize the intersection database before any analysis is performed. The criteria are as follows.

1. Overpasses or other cases when the streets do not physically intersect are eliminated from the database using supporting data to locate and deduce whether the intersection is physical or not.

2. Intersections must not be located on an Interstate.

3. Highway ramps must not be associated with the intersections.

4. Intersection types that are not consistent with the scope of the study are removed. For example road intersections in parks.

Data Review

The purpose of preliminary data review is to ensure the data usability and feasibility for a linear regression analysis. The data must be checked for outliers, linearity and spatial clustering. These attributes are important to the successful operation of a linear regression model. For example OLS regression optimal attributes include removal of outliers, stationarity normal distributions and linear relationships, while GWR models work specifically well on data is spatially clustered.

Traffic street data has its own specific characteristics; nearness is measured in right angle distance also known as Manhattan distance[2] instead of Euclidean[3] distance.

---

[2] The Manhattan distance function computes the distance that would be traveled to get from one data point to the other if a grid-like path is followed. The Manhattan distance between two items is the sum of the differences of their corresponding components. (Deza and Deza 2009)

Access corridors also play a role in how the data is patterned. Traffic volume follows a logarithmic pattern along intersection distribution. In order to determine what the spatial distribution looks like, the ArcGIS 9.3 histogram tool was used to look at the data. The histogram tool displays a measure of the frequency distribution. The crash rate descriptive statistics of the histogram are found below in Figure 3-17

Count	: 3922	Skewness	: 3.4679
Min	: 0.031926	Kurtosis	: 18.817
Max	: 5.9228	1-st Quartile	: 0.15452
Mean	: 0.52355	Median	: 0.28877
Std. Dev.	: 0.67726	3-rd Quartile	: 0.58748

Figure 3-17. Crash Rate Descriptive statistics

The histogram has a large positive skew of 3.4679 with a mean of .52. The kurtosis of the dataset was 18.817. A normal dataset has a kurtosis[4] of around 3. Since it is clear that this dataset is non-normal it must be transformed to normal bell shaped distribution. The crash rate variable was transformed logarithmically to make its distribution normal for regression analysis.

$$\log(Crash\_rate) \tag{3-2}$$

Log Transformation is often used when the data has a positively skewed distribution and there are high values. The log transformation makes variance of the data range more constant by doing this. After the log transformation the histogram was reexamined for its normalcy. The results are found in Figure 3-18.

---

[3] The Euclidean distance or Euclidean metric is the "ordinary" distance between two points that one would measure with a ruler, and is given by the Pythagorean formula. (Deza and Deza 2009)

[4] Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case.

```
Count      : 3922      Skewness      : 0.398
Min        : -3.4443   Kurtosis      : 2.7673
Max        : 1.7788    1-st Quartile : -1.8675
Mean       : -1.1522   Median        : -1.2421
Std. Dev.  : 0.96164   3-rd Quartile : -0.53191
```

Figure 3-18. Normalized Crash Rate Descriptive statistics

The Skew is much nearer to zero. Also our kurtosis of 2.76 is much closer to three. Our mean and median are also similar. These results exhibit the aspects of a normal dataset.

A logarithmic transformation was also performed on traffic volume explanatory variable for similar reasons. This creates a linear relationship to crash rate allowing our results to be interpreted clearly. The before and after statistics of this transformation can be seen in Figure 3-19 and 3-20.

```
Count      : 3922      Skewness      : 1.3843
Min        : 783       Kurtosis      : 5.2433
Max        : 52234     1-st Quartile : 4878
Mean       : 10561     Median        : 8801
Std. Dev.  : 7702.9    3-rd Quartile : 13780
```

Figure 3-19. Traffic Volume Descriptive Statistics

```
Count      : 3922      Skewness      : -0.38543
Min        : 6.6631    Kurtosis      : 2.6722
Max        : 10.863    1-st Quartile : 8.4925
Mean       : 8.9907    Median        : 9.0826
Std. Dev.  : 0.78403   3-rd Quartile : 9.531
```

Figure 3-20. Normalized Traffic Volume Descriptive Statistics

Figure 3-21 shows the spatial clustering exhibited by the crash rates. Multiple areas of high positive z scores indicate a large clustering of high similar values at a 0.05 statistically significant level.
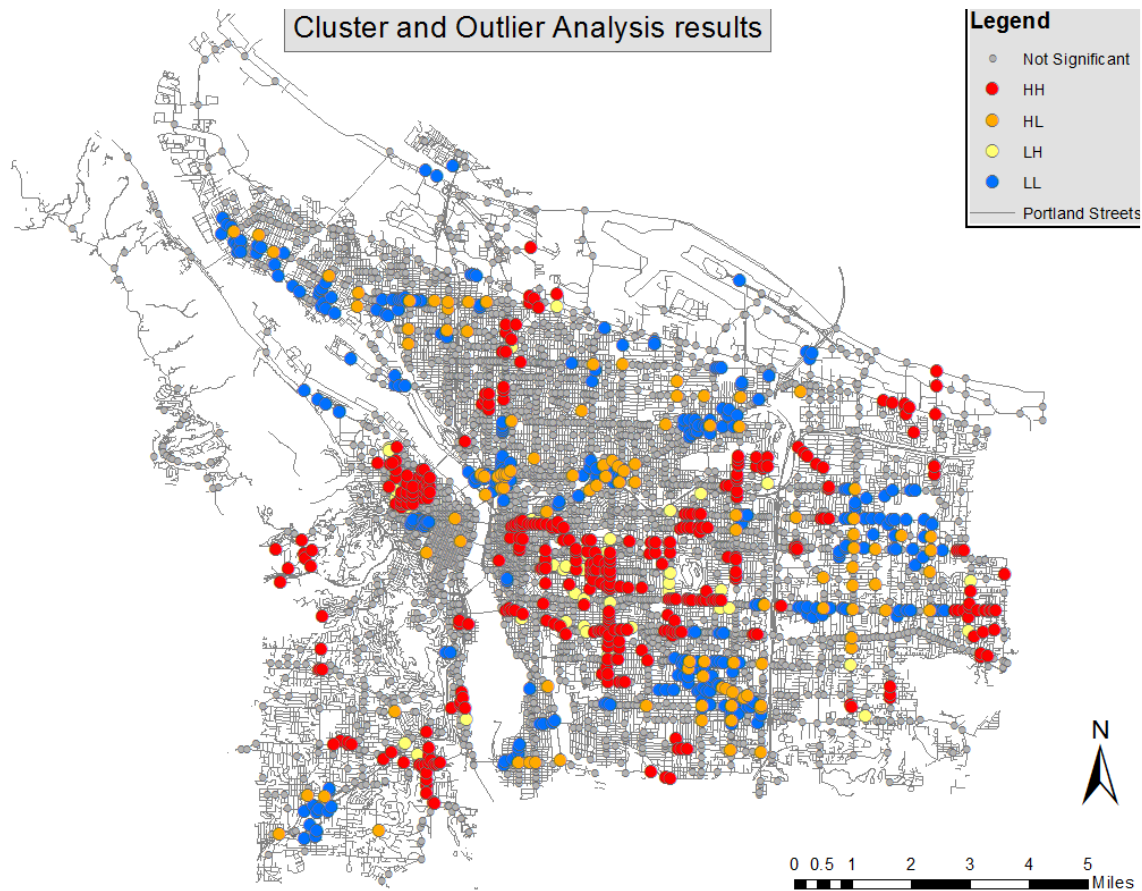
Figure 3-21. Spatial Clustering of Crash rates

Spatial clustering shows an interesting relationship. Spatial regression methods are used to analyze these relationships because it captures spatial dependency, meaning the spatial component is also a factor in the regression. The method of spatial regression selected is geographic weighted regression; this local version of spatial regression generates parameters disaggregated by spatial units of analysis. This allows for the assessment of spatial heterogeneity in estimated relationships. This is usefully when taking the "first law of geography" into account. Things that are near each other are likely to be similar. This spatial regression estimation model will output variables strength of correlation to accident count. We can use this to deduce whether this variable is a significant factor in accident count at intersections.

Assumptions of the Study

This study assumes that the relationships held between intersections are based on a Euclidean distance. This study also assumes the relationship between crash rate and traffic volume is linear for the distribution witnessed in the dataset.

Research method

Ordinary least squares

Ordinary least squares (OLS) regression is a common starting point for linear regression. Identifying and measuring relationships allows for a better understanding of what is taking place. Also OLS shows preliminary relationships that exist between independent variables and the dependent variable where. This is mainly the reason Ordinary least squares regression method is the first step we must take when modeling the data. OLS provides a global model of the variables we are trying to understand (ESRI 2009). After running OLS the general strength and significance of each variable is. OLS also tells whether each variable is a significant factor in the model. OLS gives the basic information needed to move on to geographically weighted regression.

Geographically Weighed Regression

Once OLS complete geographically weighted regression (GWR) is the next step in the analysis of the data. GWR is one of several spatial regression techniques, increasingly used in geography and other disciplines. GWR provides a local model of the variable or process you are trying to understand/predict by fitting a regression equation to every feature in the dataset. GWR constructs these separate equations by incorporating the dependent and explanatory variables of features falling within the bandwidth of each target feature. (ESRI 2009)

GWR is the significant analysis in this study. Through GWR we will answer the questions that this study raises and show relationships of intersection crash rates to intersection form factors. GWR is the analysis method of choice because of the tendencies of the data. Basic regression models assume that observations are independent of each other. This is not always the case with spatial data. Near things tend to be more related then distant things. (Tobler 1970) Not only might the variables in the model exhibit spatial dependence (that is, nearby locations will have similar values) but also the model's residuals might exhibit spatial dependence. The latter characteristic can be observed if the residuals from the basic regression are plotted on a map where commonly the residuals in neighboring spatial units will have a similar magnitude and sign. (Charlton 2009) These characteristics have negative implications on the variables that have possible spatial relationships. Geographically weighted regression does not assume the data is spatially homogenous. It takes into account that the relationships being modeled are spatially heterogeneous. The relationship of nearness differs as opposed to being the same everywhere in the model. The results of the GWR will reveal the relationships of intersection form factors to crash account and how they vary across space

CHAPTER 4
RESEARCH FINDINGS

The findings below show the output of the ordinary least square regression and geographically weighted regression. The results of the OLS show the global relationships of intersection form to crash rate. Comparison to the OLS model helps to clarify the significance in using a GWR model for this type of analysis. The relationships of each explanatory variable will be displayed using a coefficient map and coefficient variation raster to illustrate the relationships further. The coefficients maps are interpreted as follows: A negative coefficient lowers the crash rate by the respective amount per one unit of the explanatory variable. A positive coefficient raises the crash rate by the respective coefficient per one unit of the explanatory variable. The raster coefficient surface shows the regional variation of the explanatory variable. Regional variation illustrates the change in the of the coefficient variables over the data space.

Ordinary least squares

Ordinary Least Squares (OLS) regression is the first model used to analyze the relationships between the intersection features and crash rates. OLS creates a single regression equation that represents the relationship. The relationships we are modeling will reveal whether the relationships are strong positively, negatively or whether there is no relationship at all. A strong positive relationship will denote the number of crashes at intersection go down as the observed variable goes up.
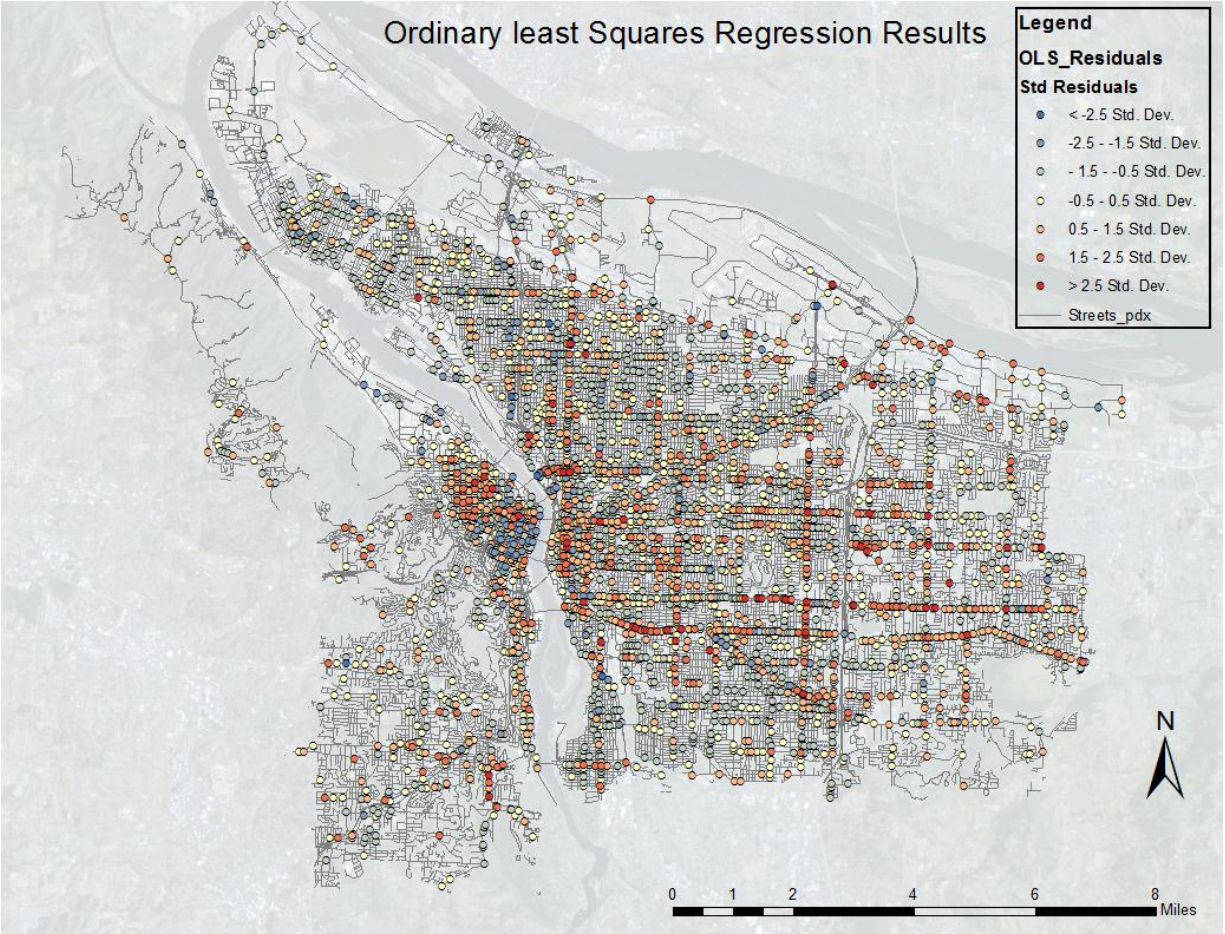
Figure 4-1. Ordinary Least Square Regression Results

To assess out models performance we look at the Multiple R-Squared and the Adjusted R-Squared. Values range from 0 – 1, higher values indicate more accurate models. We can see that for our multiple R-squared we get a result of 0.42. The chosen variables account for 42% of the variation in the dependent variable crash rate. The Adjusted R-Squared value of 0.42 represents model complexity as it relates to the data. The Summary of the OLS results are found below in Table 4-1

Table 4-1. OLS Results

Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr	VIF[1]
Intercept	-0.997591	0.195574	-5.100843	0.000001*	0.194614	-5.126001	0.000001*	------------
Intersection Legs	0.202754	0.023408	8.661564	0.000000*	0.023941	8.469032	0.000000*	1.174717
Traffic Signals	-0.667918	0.030405	-21.967123	0.000000*	0.034862	-19.159167	0.000000*	1.137924
Traffic Calming	-0.088019	0.028714	-3.065306	0.002202*	0.027569	-3.192694	0.001436*	1.023737
Corners	-0.007750	0.044919	-0.172542	0.863011	0.045018	-0.172165	0.863308	1.594663
Curbs	-0.090124	0.132507	-0.680144	0.496449	0.134491	-0.670107	0.502826	1.077899
Sidewalks	0.154196	0.042125	3.660402	0.000269*	0.045421	3.394821	0.000710*	1.469729
Percent Slope	0.005279	0.003787	1.393827	0.163462	0.003923	1.345406	0.178583	1.082866
Bridge	-0.198736	0.147899	-1.343722	0.179127	0.140483	-1.414658	0.157262	1.019422
Park	0.098880	0.041413	2.387658	0.016988*	0.038115	2.594262	0.009507*	1.020377
Lane Width	0.015061	0.000903	16.684694	0.000000*	0.000980	15.366648	0.000000*	1.344479
Number of lanes	-0.028595	0.018422	-1.552196	0.120711	0.022522	-1.269677	0.204282	1.144119
Traffic Volume	-0.000078	0.000002	-47.250741	0.000000*	0.000002	-36.747557	0.000000*	1.185923

Table 4-2. OLS Diagnostics

Number of Observations:	3922	Number of Variables:	13
Degrees of Freedom:	3909	Akaike's Information Criterion (AIC) [2]	8675.40818
Multiple R-Squared [2]	0.425375	Adjusted R-Squared [2]	0.423611
Joint F-Statistic [3]	241.141745	Prob(>F), (12,3909) degrees of freedom	0.000000*
Joint Wald Statistic[4]	2135.128270	Prob(>chi-squared), (12) degrees of freedom	0.000000*
Koenker (BP) Statistic [5]	308.128852	Prob(>chi-squared), (12) degrees of freedom	0.000000*
Jarque-Bera Statistic [6]	213.534112	Prob(>chi-squared), (2) degrees of freedom	0.000000*

Explanatory Variable Assessment

Each variable is assessed for its correlation and significance. The coefficient for each variable reflects the strength and direction of the relationship it holds with the dependent variable (crash rate). When the sign of the coefficient is *negative* it shares a negative relationship, this signifies that the independent variable is reducing the crash rate (e.g. a -0.66 decrease in crash rate is witnessed whenever traffic Signal is not present at an intersection). The Coefficient represents the expected change in the dependent variable for every one unit of change in the independent variable. The explanatory variables significance is tested within the model using a T-test, where the null hypothesis is the coefficient is equal to zero and not helping model. The smaller the probability or robust probability the more likely that the coefficient is not zero is high. An explanatory variable that is statistically significant is important to the model and signifies a relationship with the dependent variable.

Intersection legs: This variable proves to share a significant *positive* relationship with crash rate. As segment count increases the observed crash rate increases. A coefficient of 0.20 shows that the increasing complexity of intersection legs raises crash counts. A VIF of 1.17 shows that there is no redundancy in this variable.

Traffic signal: This variable proves to have significant *negative* relationship crash rate This is interpreted as meaning when traffic signals are present lower crash rates are witnessed. A coefficient of -0.66 is the strongest relationship in this study and proves to be the prominent variable affecting crash counts at intersections in OLS regression. A VIF of 1.17 shows that there is no redundancy in this variable.

Traffic calming: This variable proves to have significant *negative* relationship with crash rate. A coefficient of -0.8 shows a moderate decrease in crash rate, when traffic

55

calming devices are observed at intersections. A VIF of 1.02 shows that there is no redundancy in this variable.

Corners present: Non-significant negative relationship. There seems to be no relevant relationship between traffic corners devices and crash rates. A VIF of 1.59 verifies this variable is not redundant. This explanatory variable will be removed from the geographically weighted regression.

Curbs present: Non-significant negative relationship there seems to be no relationship between traffic curbs devices and crash counts. This explanatory variable will be removed from the geographically weighted regression.

Sidewalk present: Sidewalks exhibit a significant positive relationship with crash rate. A coefficient of 0.12 tells us as sidewalks are witness at intersections a larger crash rate will be observed. A VIF of 1.46 verifies this variable is not redundant.

Percent Slope: Non-significant positive relationship as it failed the T-test. There seems to be no relationship between the slope gradient and crash counts at intersections. This explanatory variable will be removed from the geographically weighted regression.

Bridge Intersection: Bridge intersections exhibit Non-significant negative relationship. The Probability and Robust probability of 0.155253 and 0.135205 prove to be too high for significance. This explanatory variable will be removed from the geographically weighted regression.

Park Intersection: Significant positive relationship. A coefficient of 0.09 shows that parks intersections are related with higher crash rates. A VIF of 1.46 verifies this variable is not redundant.

Lane width: Significant slight positive relationship coefficient of .014 shows that wider the lane width the larger the crash rate.

Number of lanes: Non-significant negative relationship. There seems to be no relationship between number of lanes and Crash rate. This explanatory variable will be removed from the geographically weighted regression.

Traffic Volume: Significant negative relationship. A coefficient -0.000078 proves a correlation with traffic volume at intersection with crash counts. As traffic volume goes up crash count is reduced. There are certain phenomenon witnessed with traffic volume that is discussed refer to literature on traffic volume for better understanding of traffic volumes.

Model significance

The Koenker (BP) statistic of 308.128852 and Joint Wald Statistic of 2135.128270 exhibit overall model significance. The null hypothesis for these tests is that the explanatory variables in the model are not effective.

Stationarity Assessment

The Koenker (BP) Statistic is also used to assess Stationarity. It is important to determine whether the explanatory variables behave the same everywhere in the study. When the same correlations between the explanatory variables and independent variables are witnessed everywhere in the study area the model exhibits stationarity. When the model is consistent in data space, the variation in the relationship between predicted values and each explanatory variable does not change with changes in explanatory variable magnitudes (there is no heteroscedasticity in the model). (ESRI 2011) The null hypothesis for this test is that the model is stationary. A p-value of smaller that 0.05 indicates significant heteroscedasticity and nonstationarity.

The p-value for the Koenker (BP) Statistic is 0.00000 so our model exhibits nonstationarity. Models that exhibit nonstationarity prove to be good candidates for geographically weighted regression.

Model Bias

Model bias The Jarque-Bera Statistic is used to assess whether the model is biased. A model is defined as being biased when the observed dependent values minus the estimated values are not normally distributed. These are known as residuals. When the p-value is less than 0.05 for a 95% confidence interval the residuals are not normally distributed. This can mean that the model is misspecified and one or more key variables are missing from the model.

A significant Jarque-Bera statistic is observed in the OLS model results. Further analysis of the residuals using a histogram is shown in Table 4-2
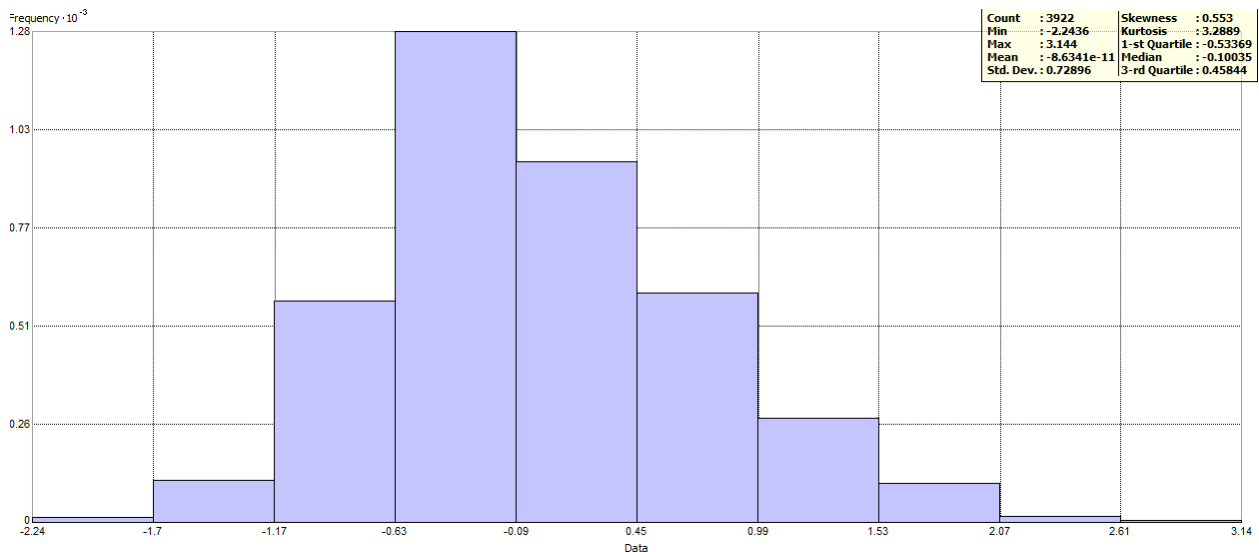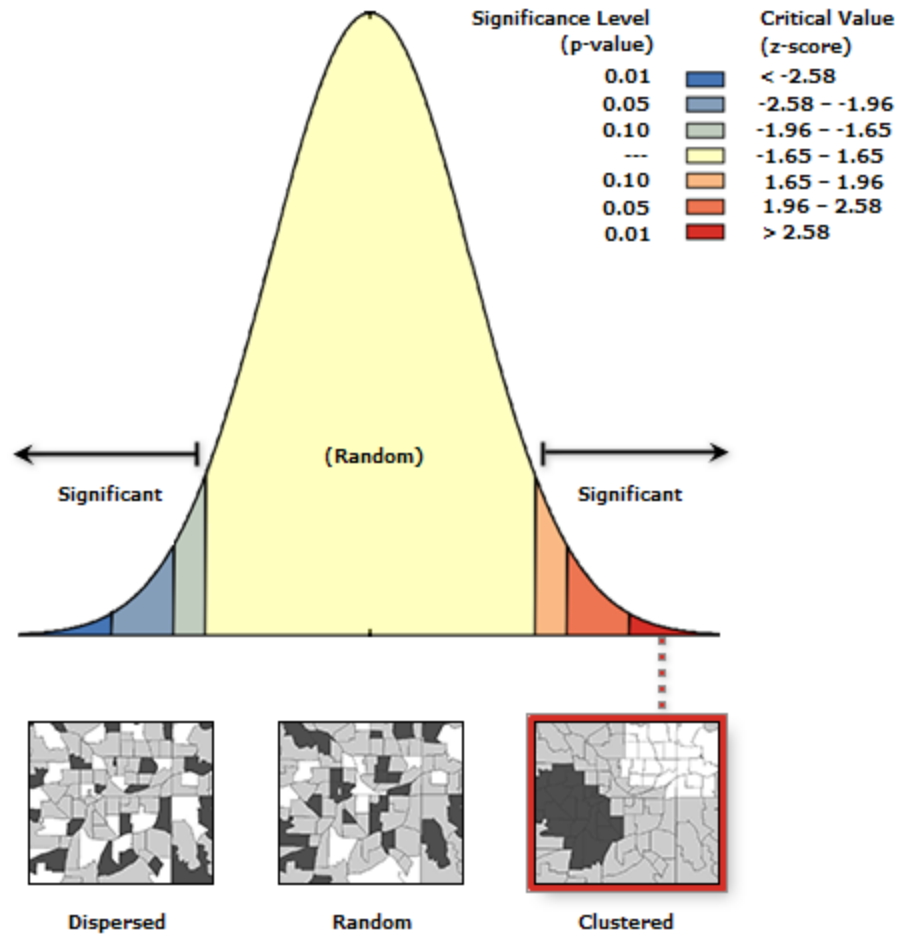
Figure 4-2. OLS Residual Histogram

Figure 4-3. Spatial Autocorrelation of OLS Residuals

The histogram of residuals reveals a near normal distribution. A positive skew of 0.55 and kurtosis of 3.2 are not too far off of what we would expect from a normal distribution. Also referring back to that strong heteroscedasticity witnessed in the stationarity assessment it is not surprising to see a small bias. Due to limitations of data key variables that may contribute to model strength are missing and this also contributes to the small model bias we are seeing. After running spatial autocorrelation (see Figure 4-3) we see that the residuals are clustered further emphasizing that fact high heteroscedasticity exists along with some variable misspecification. This is an expected result when the final R-squared is 0.42.

Geographically Weighted Regression

Geographically weighted regression (GWR) is the second and final model used to analyze the relationships exhibited by our explanatory variables. The results of the ordinary least squares regression showed the data exhibited strong heteroscedasticity. GWR provides a local model of the variable relation process. It fits a local regression equation to every feature in the dataset. It does this by incorporating the dependent and explanatory variables of features that lie within the bandwidth[1] of each target feature. This accommodates the spatial characteristics of the intersection data much better than Ordinary least square regression. The analysis will be able to model the relationship of crash rate to our intersection form variables spatially.

Table 4-3. GWR Results Output Table

VarName	Variable
Neighbors	140
ResidualSquares	575.068345
EffectiveNumber	306.568381
Sigma	0.627079
AIC	3574.863381
R^2	0.644784
R^2 Adjusted	0.570564

To assess the performance of the geographically weighted regression we look at the Adjusted R-Squared and AIC value to see how well our model preformed. As we see from Table 4-3 the adjusted R-Squared increased to.57 from .42 indicating this model has increased in complexity from OLS. In OLS our R-Squared was 42% and by

---

[1] Bandwidth or number of neighbors used for each local estimation and is perhaps the most important parameter for geographically weighted regression. It controls the degree of smoothing in the model. Typically, you will let the program choose a bandwidth or neighbor value for you by selecting either **aicc** (the corrected Akaike information criterion) or cross validation for the bandwidth method parameter. Both of these options try to identify an optimal fixed distance or optimal adaptive number of neighbors. Since the criteria for "optimal" are different for aicc than for cross validation, it is common to get a different optimal value. You may also provide an exact fixed distance or a particular number of neighbors by selecting bandwidth parameter for the bandwidth method.(Esri, 2011)

modeling our variables spatially across our study area the R-Squared has increased to 64 percent. Previously OLS reported an Akaike's Information Criterion (AIC) of 8675.4. The result for GWR is down to 3574.8. This allows for the models to be compared, the lower number of 3574.8 is much lower and therefore GWR is a much better model for the relationship of the variables. Residual-Squares record shown in Table 4-3 is the sum of the square residuals, the smaller this measure is the closer the fit of the GWR model to the observed data. The effective number record in shown Table 4-3 is a tradeoff between the variance of the fitted values and the bias of the coefficient estimates and is related to the choice of bandwidth. The effective number is used to compute a number of diagnostic measures. (ESRI 2009) Finally the sigma is the estimated standard deviation for the residuals smaller values of this number is preferred.
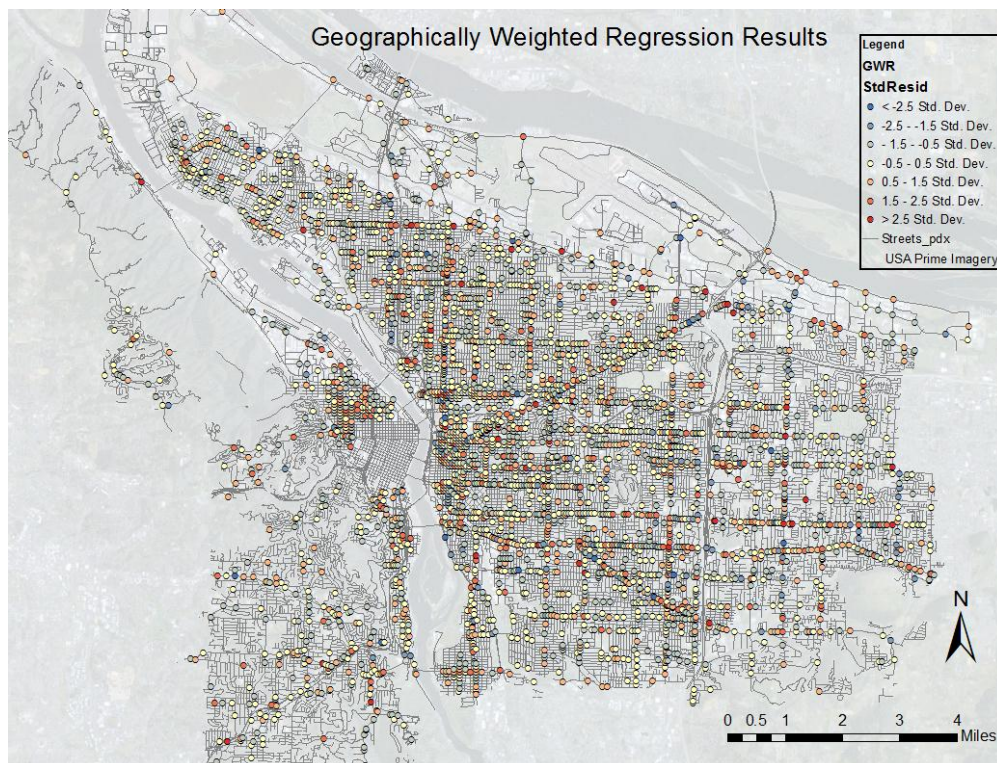
Figure 4-4. Geographically weighted Regression Results Standard Deviations of
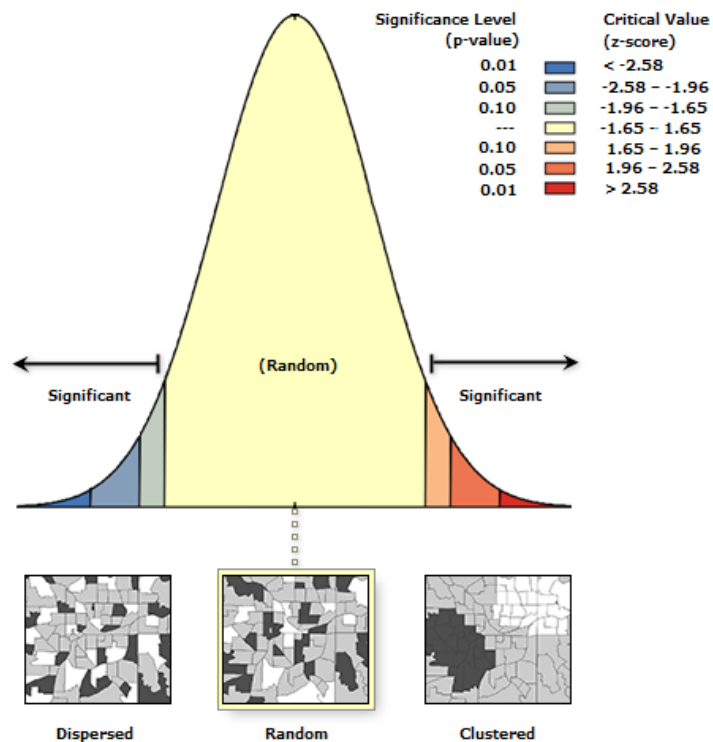Residuals

Figure 4-5. Spatial autocorrelation of GWR residuals

Another output of GWR is a map of residuals. It is important to quickly test for

spatial clustering that we run a spatial autocorrelation. In Figure 4-5 can see that the

residuals are not spatially clustered.

Local R^2

The Local Residual-Squared shown below in Figure 4-6 shows us how well the

model predicts the outcomes in each of the observations included in the local model. It

is a number from 0-1 denoting weakness to strength respectively. The model does

poorer in the more dense portions of central Portland. High model accuracy is pretty

much witnessed in the area surrounding the city center.
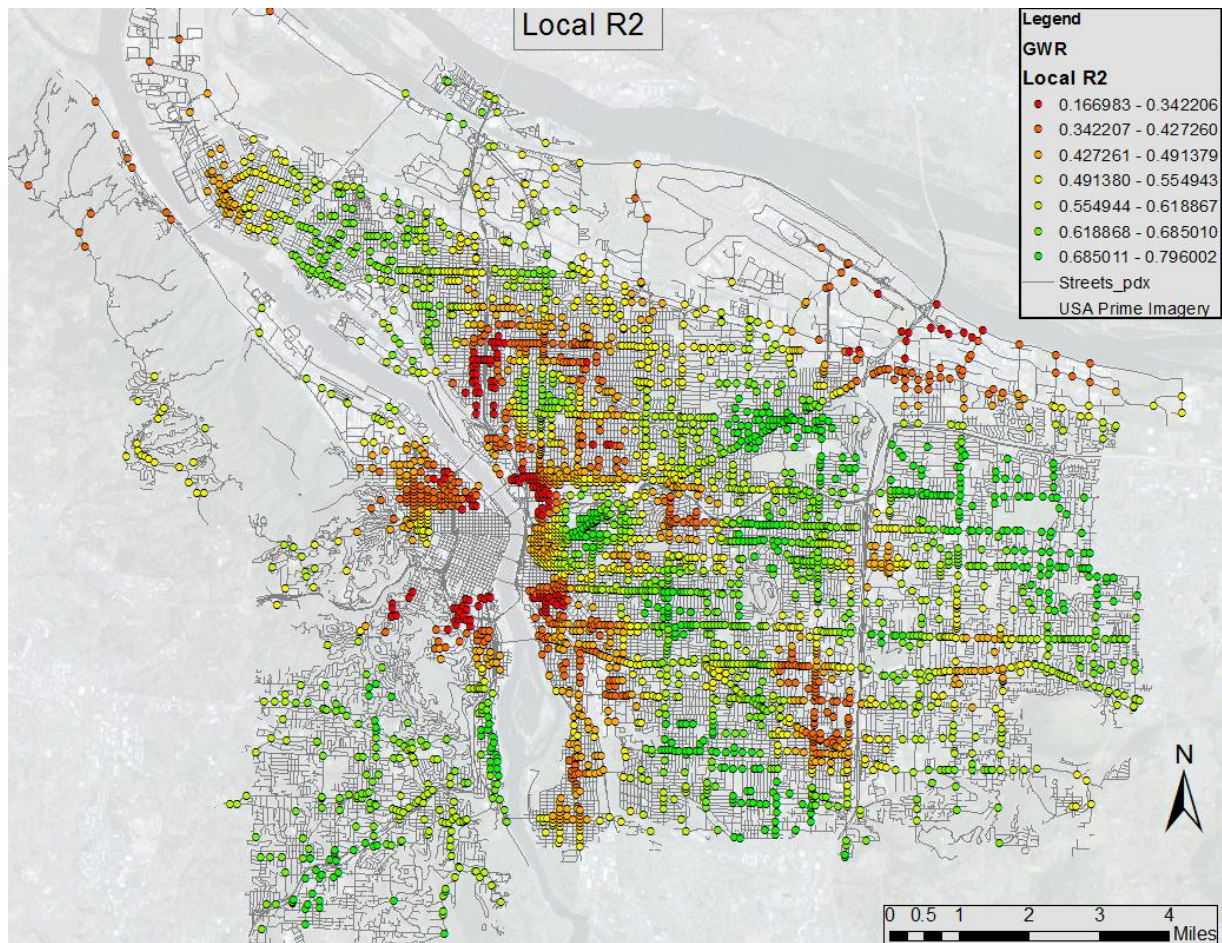
Figure 4-6. GWR Local R^2

GWR Variable Analysis

GWR outputs variable coefficients to show the strength and direction for each observation. GWR also creates a coefficient variability surface of each explanatory variable. This surface shows how the variables relationships vary across space; in effect testing how consistent is stationarity between the explanatory and dependent variable. Each variable is analyzed with its Coefficient this way it is shown where each variable varies spatially and its significance to crash rate.
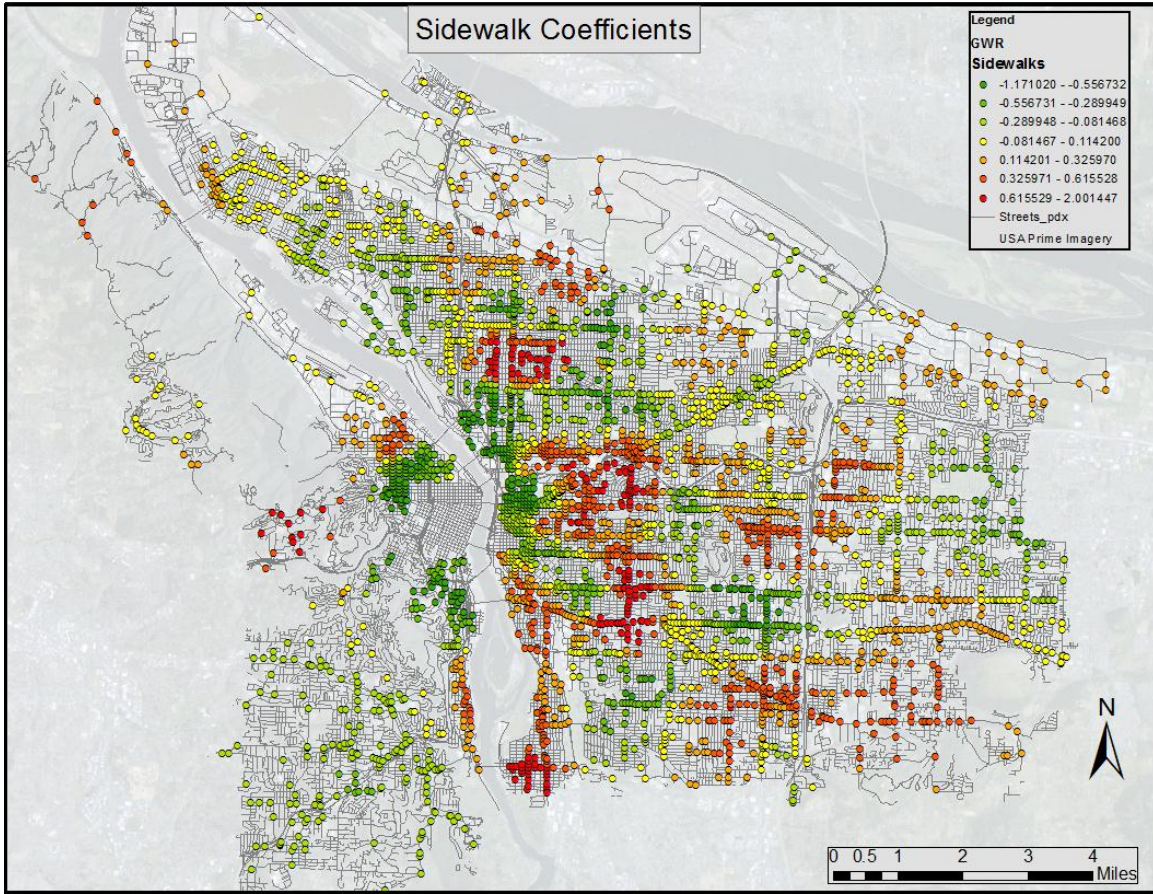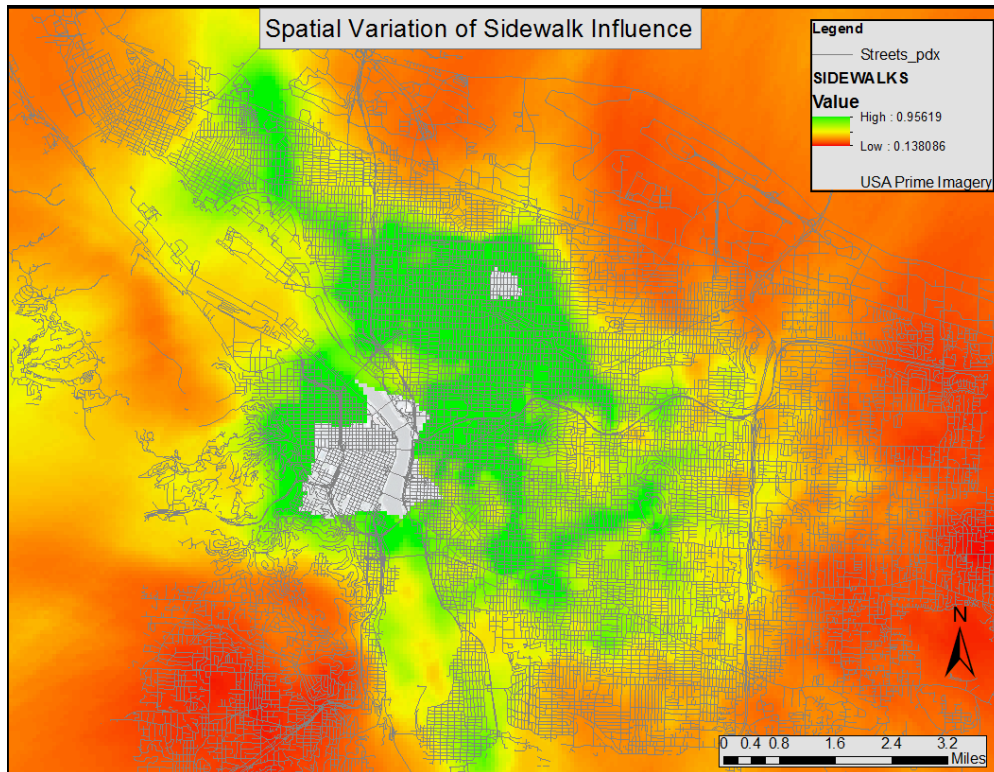
Figure 4-7. Sidewalk Coefficients GWR

Figure 4-8. Coefficient Variation surface

Sidewalks

The Figures 4-7 and 4-8 show the Coefficient map and coefficient variation surface for Sidewalks. Figure 4-9 shows the distribution of the coefficients. The distribution of coefficient map for sidewalks suggests high variation. The coefficients vary from a value of -1.17 and 2 this is a large spread meaning that in some areas sidewalks can influence crash rates positively or negatively. Further analysis of the large spread shows that there are large standard of errors on the records with large positive effects. These are areas where the model failed to perform well. The Sidewalk coefficient confirms that Sidewalks relationship with crash count varies highly through a large portion of the network. Because of its variation and large range is it hard to discern whether sidewalks has a strong relationship with crash can counts positively or negatively. This variable

65

has a general trend of having a negative coefficient meaning a reducing effect on crash

count. But overall high standard errors hinder the validity of this trend.
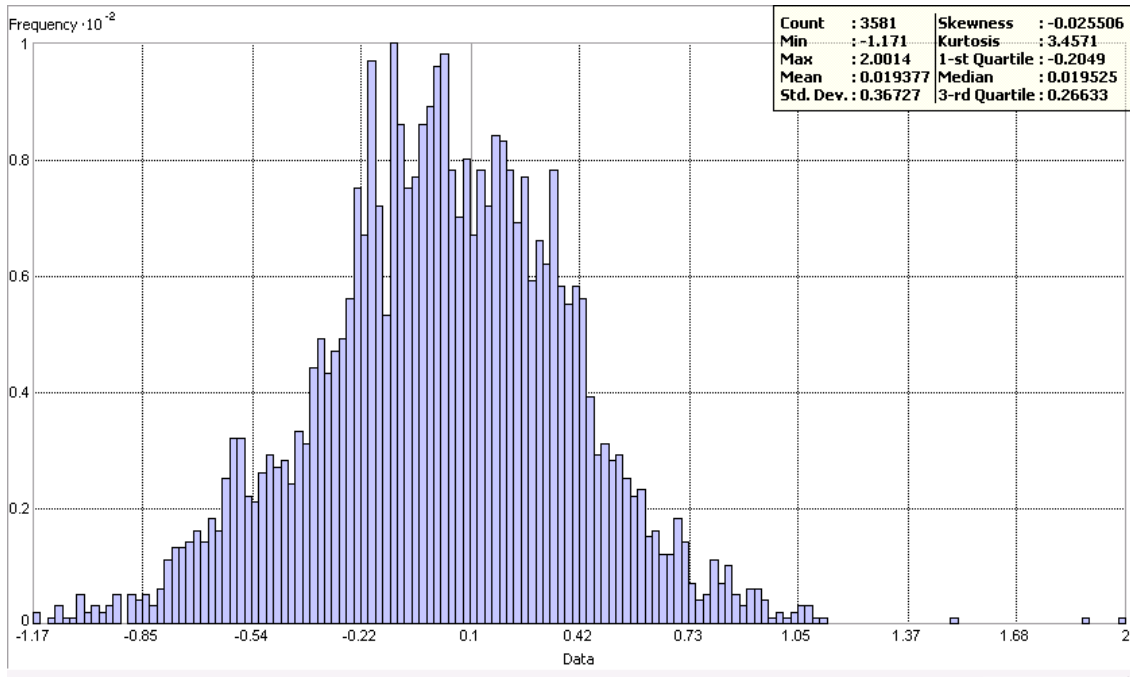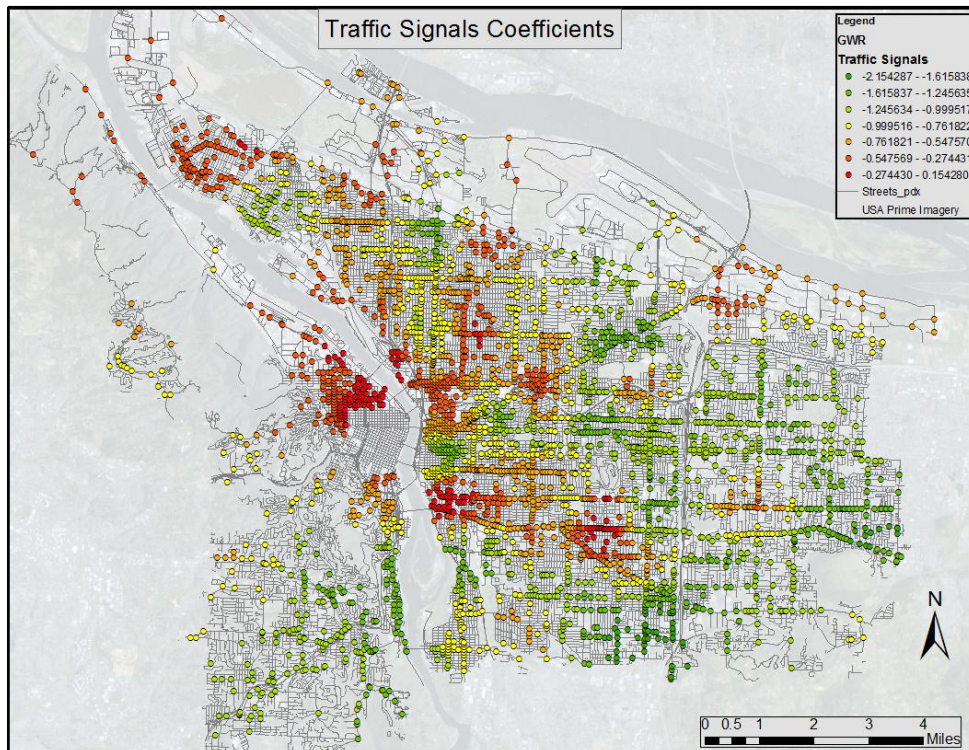

Figure 4-9. Sidewalk Coefficient Histogram
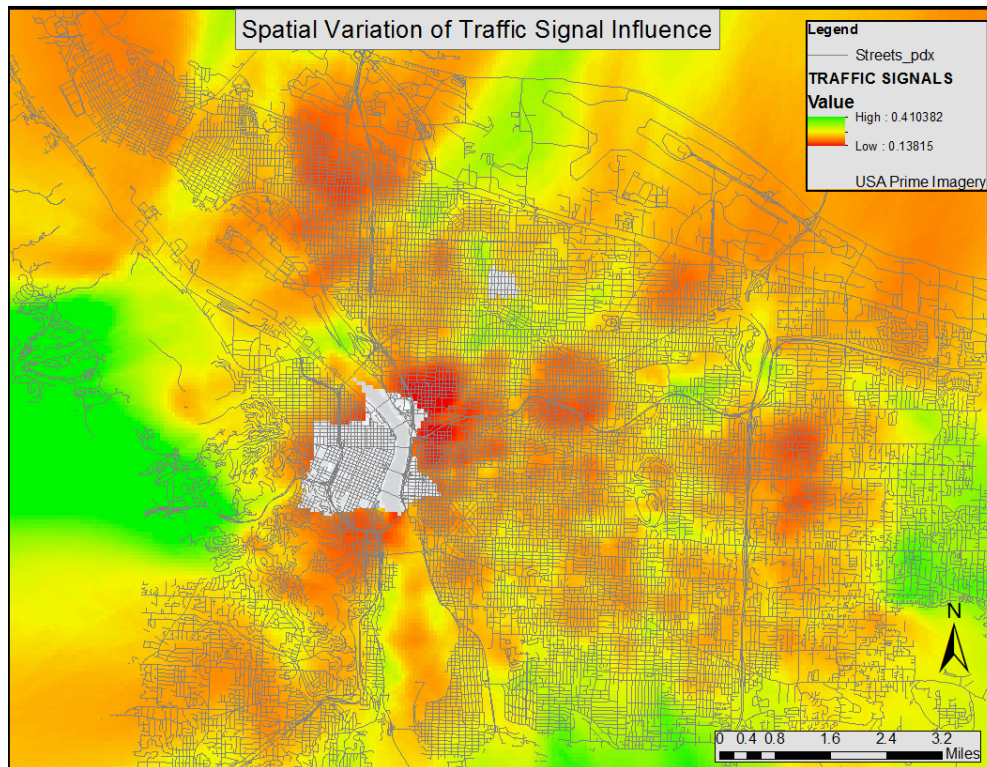

Figure 4-10. Traffic Signal Coefficients GWR

Figure 4-11. Traffic Signal Coefficient Variation surface

Traffic Signals

The Figures 4-10 and 4-11 show the Coefficient map and coefficient variation

surface for Traffic Signals. Figure 4-12 shows the distribution of the coefficients. The

distribution of the coefficient map for traffic signals suggests low variation. Traffic count

has a coefficient distribution of -2.1543 to 0.15 this spread shows an overall negative

direction. This means that traffic signals in most cases reduces the crash rates

witnessed throughout the city of Portland. The stand error distribution range is low

meaning that these coefficients are a reliable measure of correlation to the dependent

variable of crash count. The coefficient Variation surface is fairly consistent to showing

low traffic signal coefficient variability over the network. From this analysis we can

deduce that Traffic Signals are consistent factor in reducing crash rate by a large
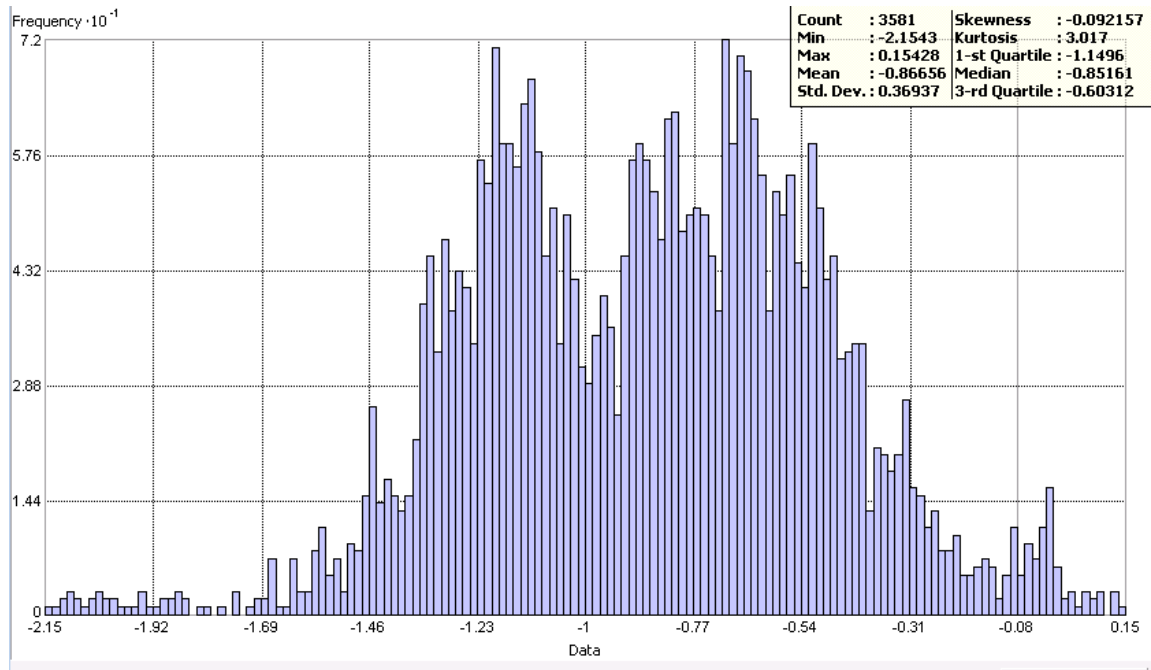
amount.
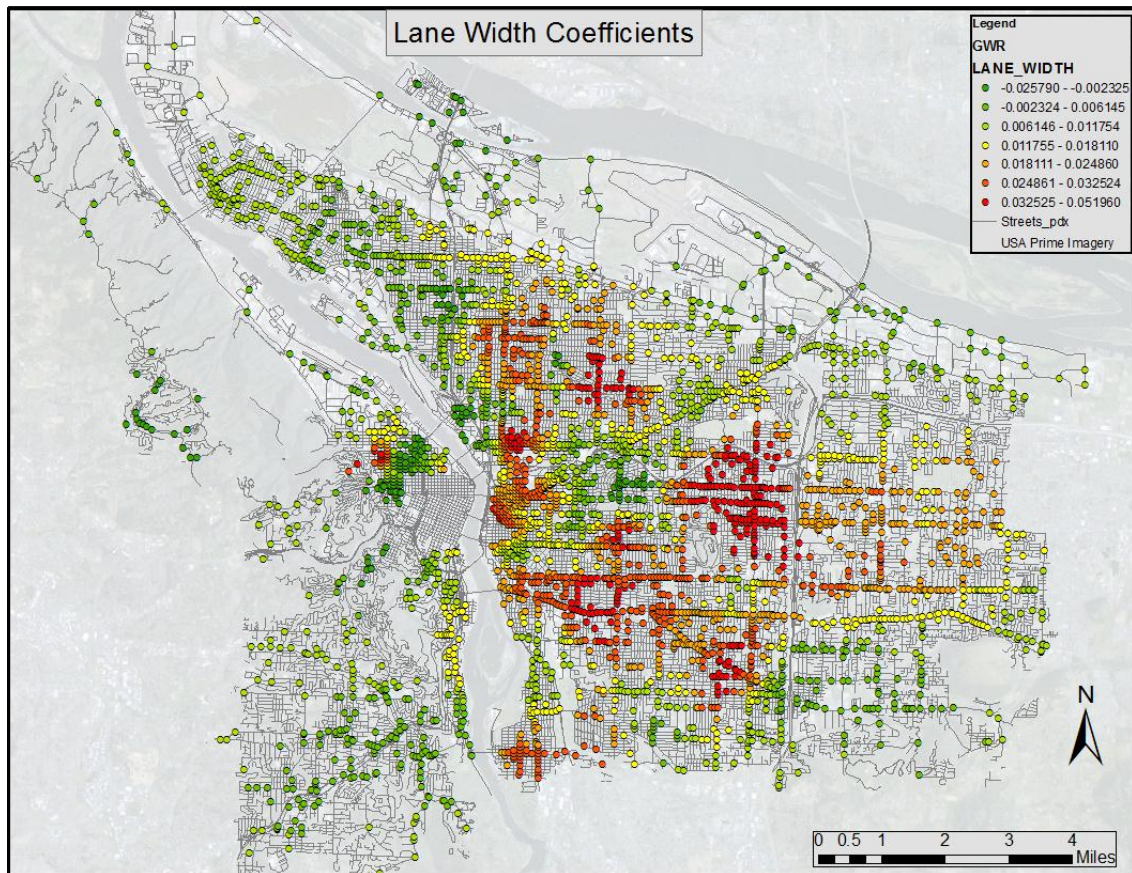
Figure 4-12. Traffic Signal Coefficient Histogram

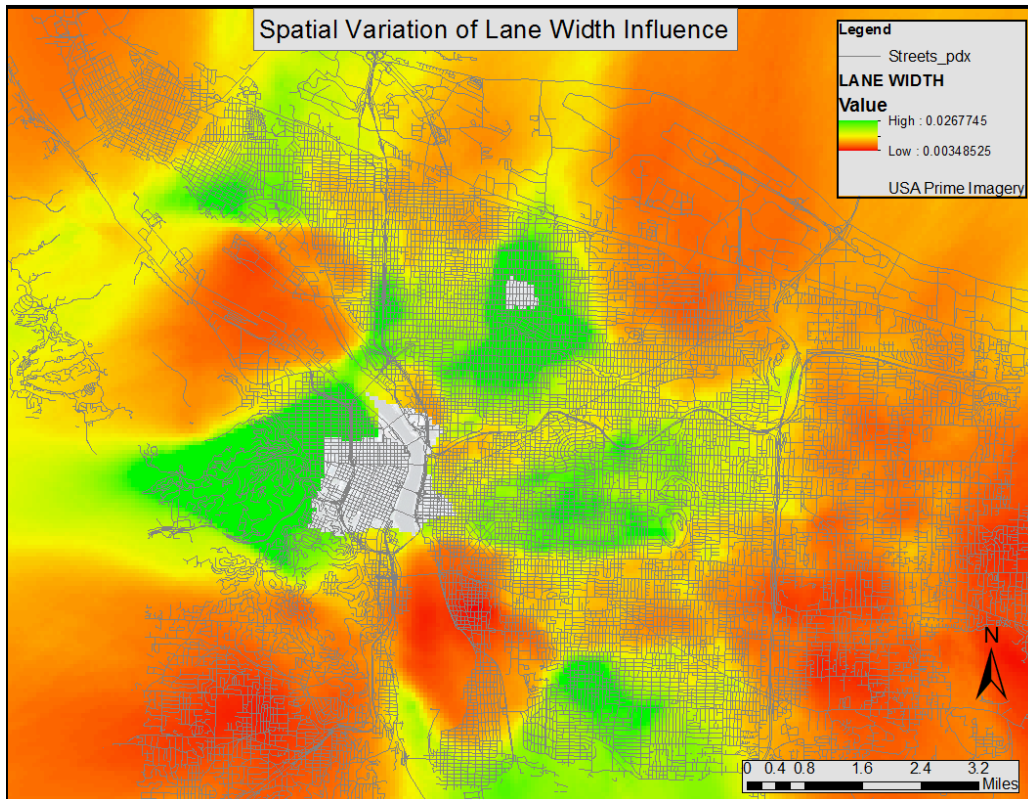Figure 4-13. Lane Width Coefficient GWR

Figure 4-14. Lane Width Coefficient Variation surface

Lane Width

The Figures 4-13 and 4-14 show the Coefficient map and coefficient variation

surface for Lane Width. Figure 4-15 shows the distribution of the coefficients. The

distribution of the coefficient map for travel signals shows a weak relationship with the

dependent variable. The Spread of the distribution is -0.025 to 0.051 with most of the

observations lying near the middle of that range. The coefficient variation surface

shows us that there is a large variation in the lane width coefficient variable within the

center of the city. Less variation in this variable is witness in the city periphery

suggesting that measuring lane may exhibit a stronger relationship to crash rate outside

the dense city center. The standard error for lane width exhibited a small variation with a

high frequency at the .006 mark. These results are fairly trustworthy when taking the

local R-Squared into account but still show a weak relationship and high variability with

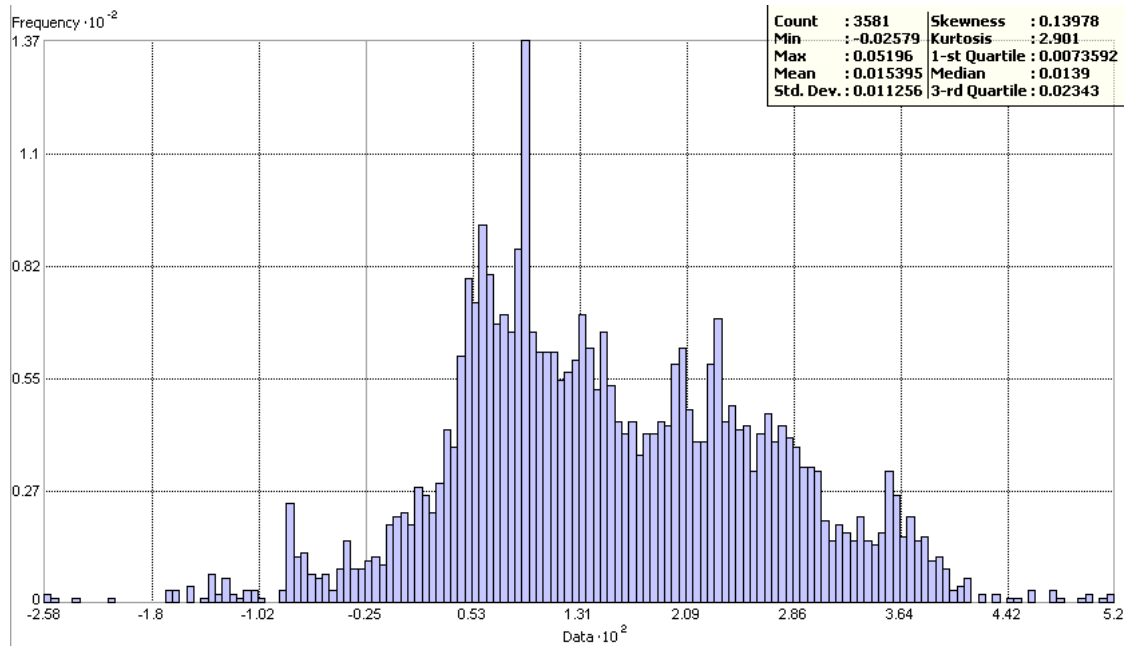crash rate throughout the city of Portland.

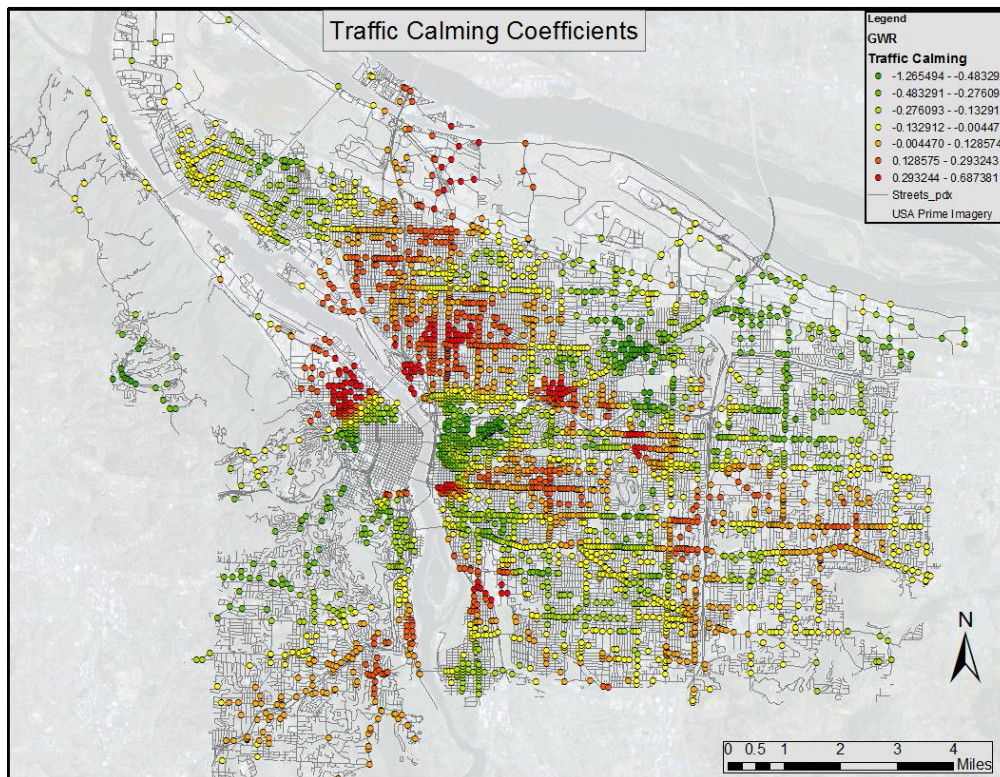

Figure 4-15. Lane Width Coefficient Histogram

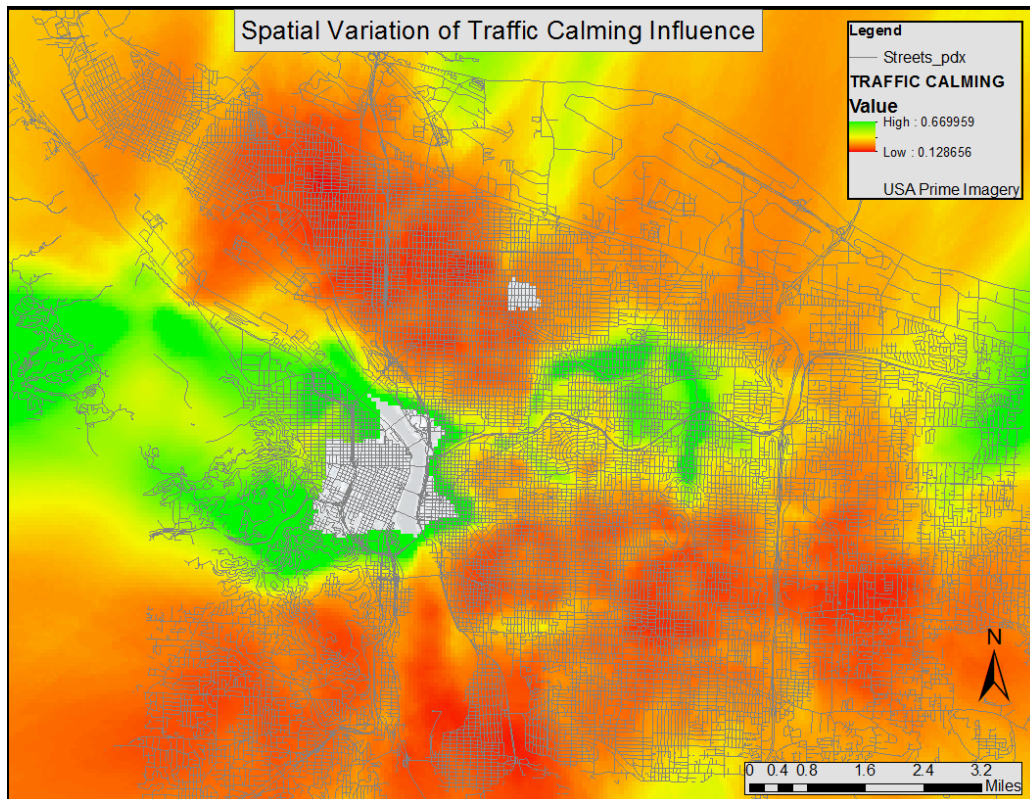Figure 4-16. Traffic Calming Coefficient GWR

Figure 4-17. Traffic Calming Coefficient Variation surface

Traffic Calming

The Figures 4-16 and 4-17 show the Coefficient map and coefficient variation

surface for Traffic Calming. Figure 4-18 shows the distribution of the coefficients. The

distribution of the coefficient map for traffic signals shows moderate spread of values

from -1.2655 to 0.68 this spread is misleading since the distribution has long "tails".

Most of the observations lie illustrates a negative direction meaning that traffic calming

has a reducing effect on crash rate. Standard Errors seem rather high. When looking

further into this issue the traffic calming variable seems only be represented in the areas

where low crash rate are witnessed. This tells us that traffic calming devices are

clustered in areas that do not see a lot of variable "traffic". This may be a definition issue

where traffic calming devices are, either not presents in the city core or they are just not

part of the database for the pedestrian districts. The traffic calming variation surface shows consistent model stationarity and low variation over the locations where traffic calming devices are witnessed. Traffic calming supposedly reduces crashes but the model does not show this correlation well. It is hard to discern whether the data or the model does not lend its self well to the analysis.
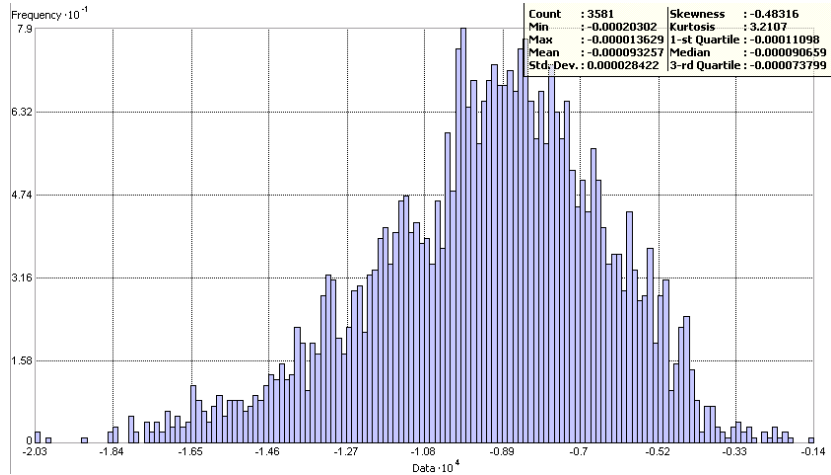

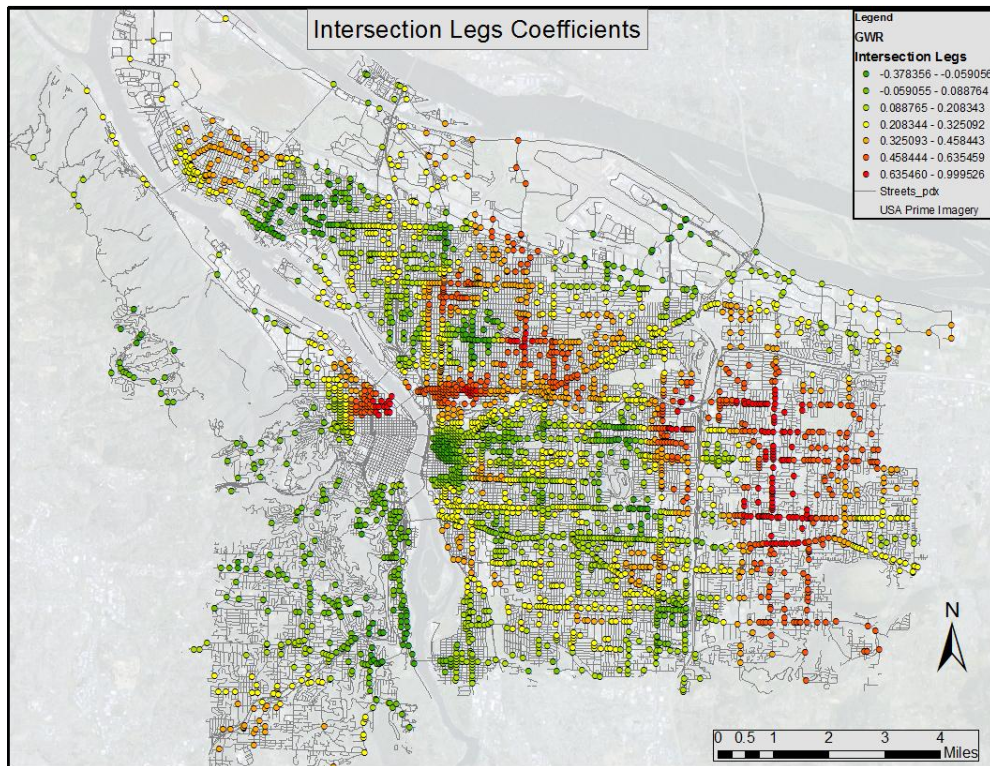Figure 4-18. Traffic calming Coefficient Histogram

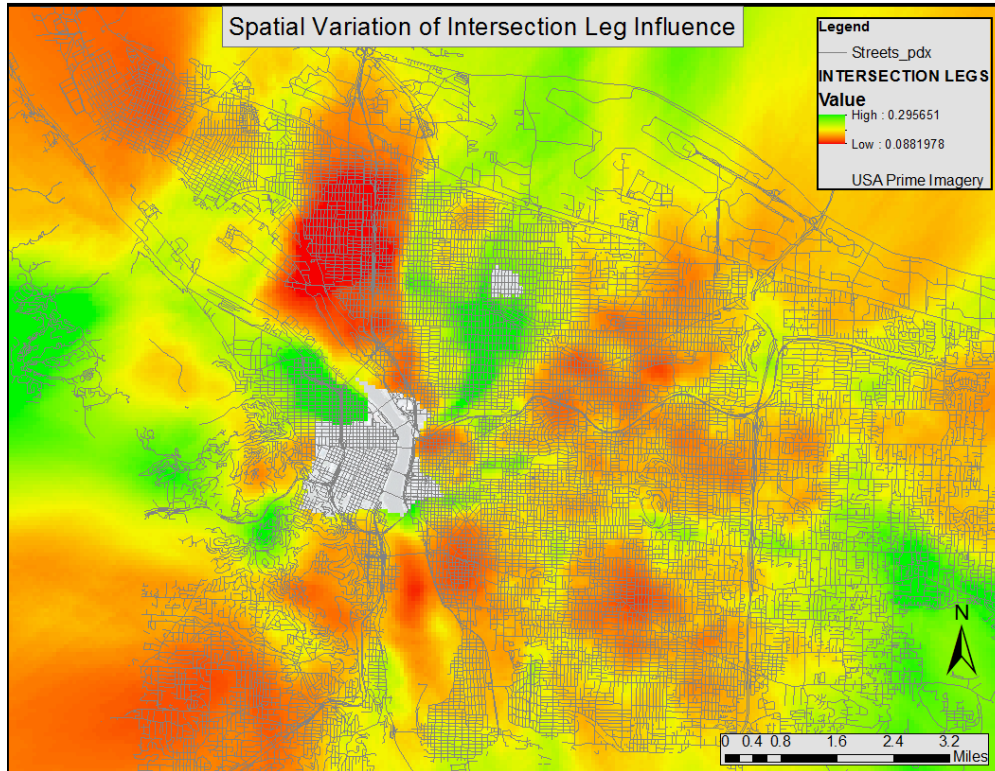Figure 4-19. Intersection Leg Coefficient Variation surface

Figure 4-20. Intersection Leg Coefficient Variation surface

Intersection Legs

The Figures 4-19 and 4-20 show the Coefficient map and coefficient variation surface for Intersection Legs. Figure 4-21 shows the distribution of the coefficients. The distribution of the coefficient map for shows a moderate positive correlation with traffic count. The spread of the distribution is from -0.37 to .99. This means as intersection legs increase and the intersection becomes more complex a rise in crash rate is witnessed. Although most of the observations lie with a positive correlation the GWR model shows us that for certain areas this does not remain true. In some cases higher amounts of intersection legs can lead to lower crash rate. The standard error is low indicating that observations modeled are trustworthy. The coefficient variation surface shows us that the intersection leg variable is subject to many variations throughout the network. It exhibits stationarity in the areas where high traffic volume is witnessed.

73

From this we can conclude that as the amount intersection legs increase so does the
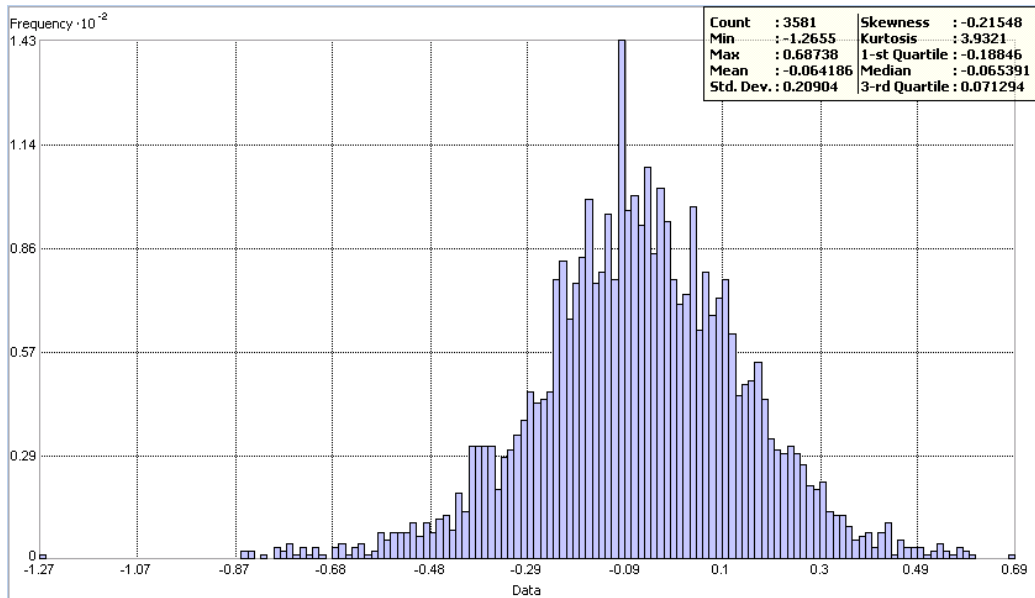
crash rate.
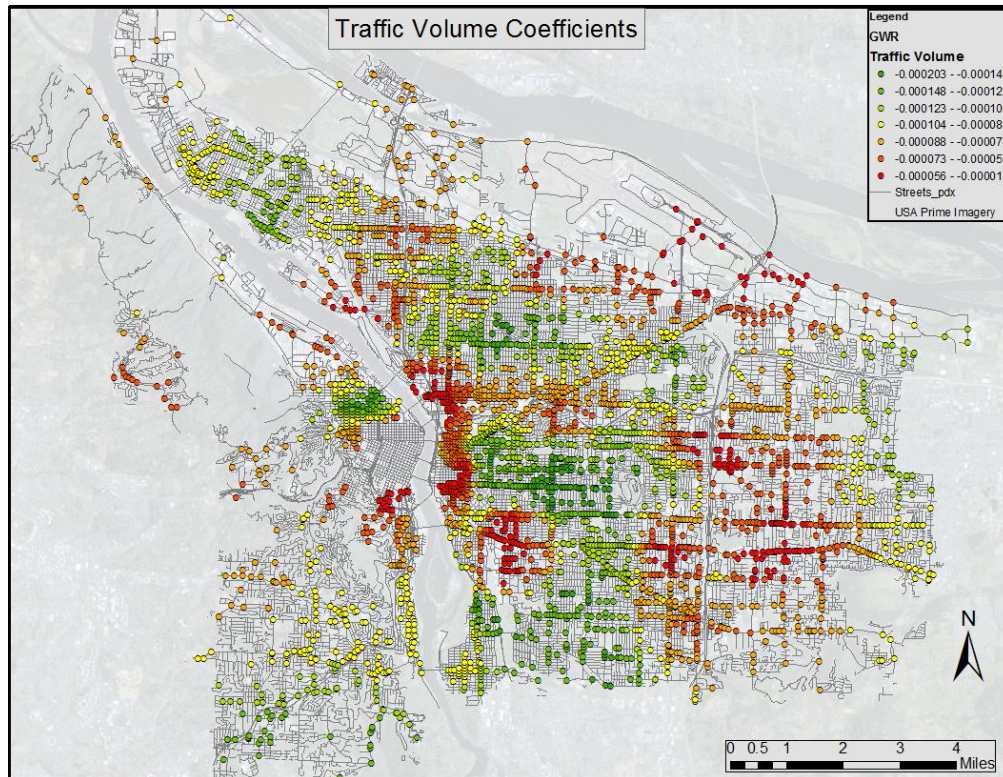

Figure 4-21. Intersection Legs Coefficient Histogram

Figure 4-22. Traffic Volume Coefficient surface
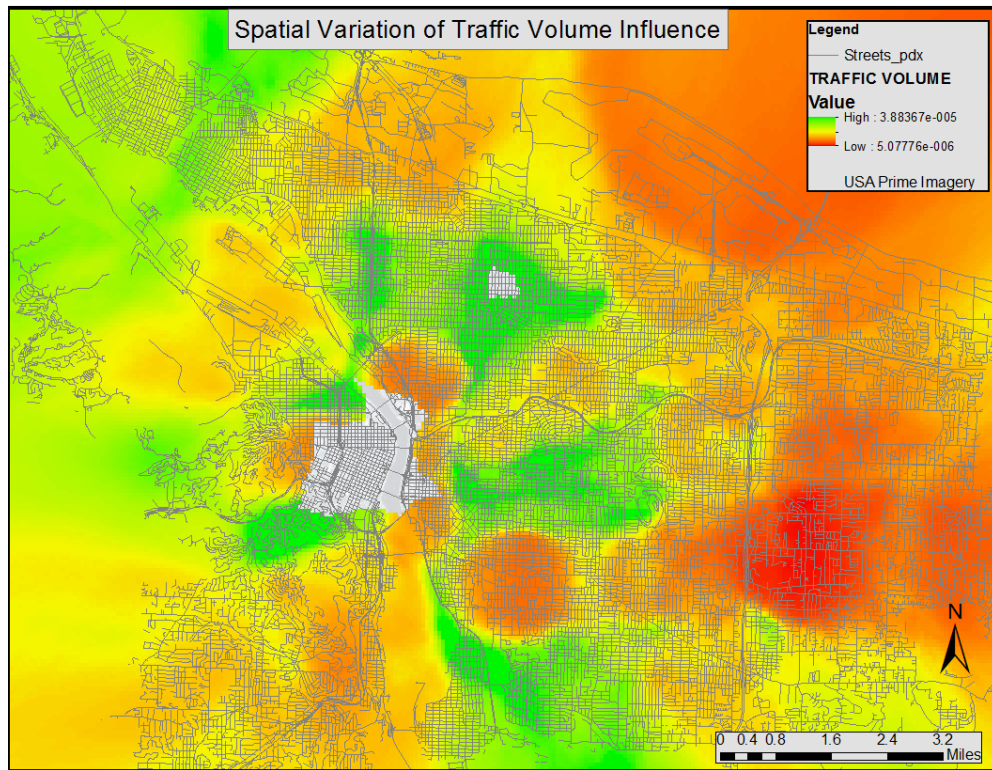
Figure 4-23. Traffic Volume Coefficient surface

Traffic Volume

The Figures 4-22 and 4-23 show the coefficient map and coefficient variation

surface for Traffic Volume. Figure 4-24 shows the distribution of the coefficients. The

Distribution of the coefficients for traffic volume suggests a really weak negative

correlation with crash rate. The spread of the observations is -0.0002 to -0.000013. The

direction suggests that as traffic column increases we see a decrease in crash rate.

This is an expected result because of after about 20,000vpd[2] the relationship that crash

rate increases with Traffic volume does not hold true. As traffic volume increases past

this level the crash rate witnessed regresses very quickly. The coefficient standard

errors are low and thus the results reliable. The coefficient variation surface shows that

traffic volume exhibits stationarity in areas where traffic volume is most dense. This is

---

[2] *Vehicles per day – Average Vehicles per day observed entering the intersection.*

75

expected because there is a large density variation in the traffic volumes witnessed throughout Portland's street network. Traffic volume is concentrated at highway exits and as traffic disperses through the city volume decreases rapidly.
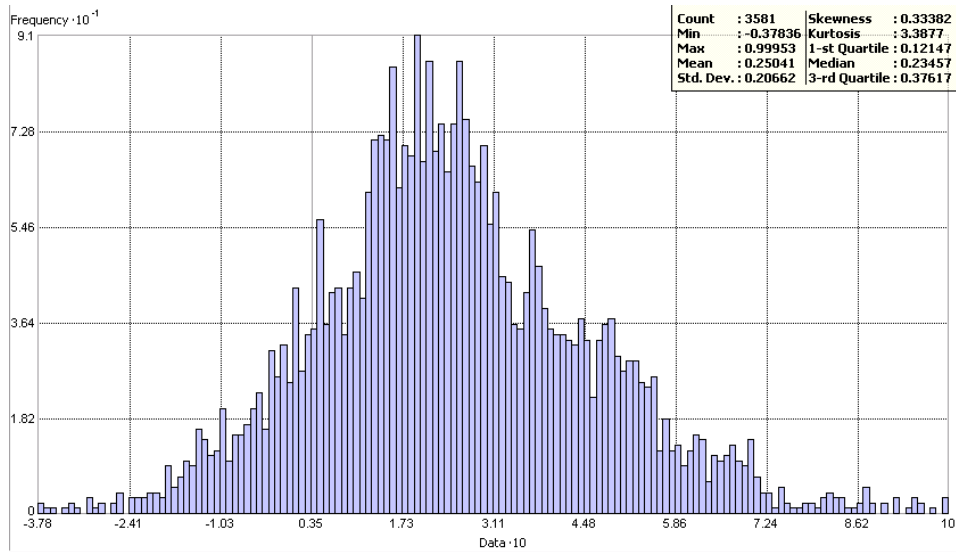
Figure 4-24. Traffic Volume Coefficient Histogram

The primary objective of this study is to analyze the relationships between traffic

intersection and intersection form factors in Portland Oregon using spatial regression.

The results show that GWR does provide significant and useful results. It also proves

that using GWR is significantly better at modeling intersection crashes than a global

regression method. The GWR model results show a visual and statistically accurate

relationship between intersection form and crashes. The results of the model can help

shape decisions concerning intersection design and policy. This chapter discusses the

findings from the result maps of GWR and how the variables relate. Then the study

looks at any external factors that may be weakening the model. It also discusses the

shortcomings of the spatial regression technique. The impact of geographic weighted

regression in intersection analysis is also discussed.

Figure 5-1. Model performance vs. OLS

Ordinary Least Squares

In doing the preliminary analysis it is discovered how weak using a global regression model like Ordinary Least Squares (OLS) is in modeling the relationship between traffic intersect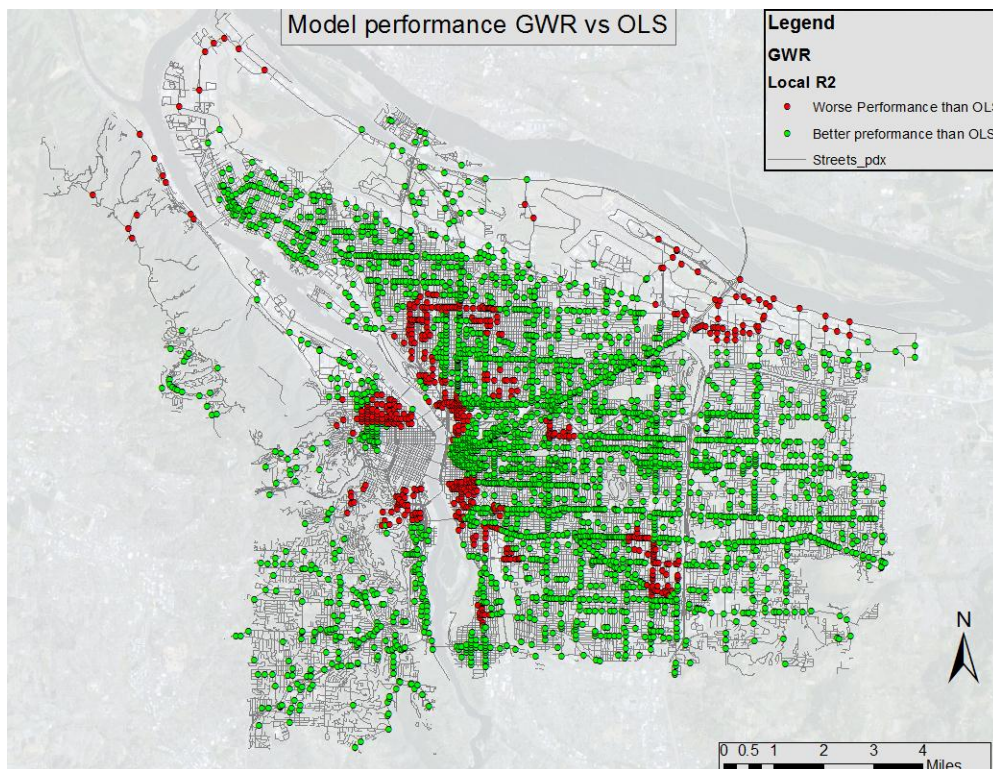ion crashes and intersection form. Figure 5-1 shows where the GWR model performed better than the OLS model. Since the model does not account for the inherent spatial relationships that exists in a transportation network it fails to produce relevant results in the direction and strength of correlations that exists between variables. Each observed intersection is an important and unique instance in the transportation network that using a global model to explain what is going on becomes irrelevant for our analysis.

Many of the OLS variables failed to exhibit significance. Corners, curbs, percent slope, bridges and number of lanes correlates with crash rate as expected, but did not affect the model as expected. Better variable specification and a larger dataset may show significance in future analysis but until more data is obtained these variables prove to be inconclusive. For the variables that exhibited significance; traffic signals, traffic calming devices, intersection legs, sidewalks, lane width and traffic volume the results followed expectations. Traffic signals shared the largest crash count with crash rate. Traffic volume relationship is not presented well because of its irregular relationship with crash rate. The larger traffic volumes mask the relationship that might be expected from the traffic volumes ranging from 0 to 20,000 Vpd.

Geographically Weighted Regression

This study shows that the geographically weighted regression models the relationships between crash rate at intersections and intersection form factors in a meaningful and conclusive manner. The witnessed relationships make sense on an

engineering level and expected results appear in many cases. The reason behind

GWR's success in modeling this relationship lies in the use of a local model for each

observation. The local model mirrors more accurately what we naturally see as

relationships that exist in transportation networks. An observation by observation model

taken into spatial context can be manipulated to reveal or predict effects that maybe the

result of future intersection form implementation. Also it is worth mentioning that the

results of the GWR model are independent of time because the data is continuous. This

is what sets GWR model apart from a "before and after" method of analysis which gets

weaker as the time in between observations gets larger and the external factors create

variations that cannot be accounted for. (Levinson and Chen 2006)

We will look at traffic signals for an example of the uses of this analysis. It is

apparent that traffic signals have the propensity to reduce crash rates over the greater

Portland area. These results are reinforced through the spatial stationarity[1] that this

variable exhibits over the range of the data. It is widely known that traffic signals are a

key factor in traffic safety management at intersections. The spatial regression results

correlate with the traffic safety management studies. To take it a step further the GWR

model models the strength of each traffic signal's impact on the crash rate at the local

level. In effect this study can directly model how strong an impact will be witnessed to a

location if a traffic signal is applied to that area.

<div align="center">**Weaknesses in the GWR model**</div>

Firstly GWR uses neighbors on a Euclidean distance measurement. Neighboring

intersections do not relate accurately on a transportation network using Euclidean

distance, although we see from the model that we get conclusive results, these may be

---

[1] Stationarity: the mean, variance and autocorrelation structure do not change over time and space.

further improved by a more accurate distance measure. Transportation network phenomenon is another area that weakens the model; throughout Portland there are bridges that funnel huge amounts of traffic. These bridges affect the intersection relationships making for lower a R^2 in the areas near bridges. This could be mitigated somewhat by using a distance to nearest bridge explanatory variable to analyze their effect on observations that are near them. The form of Portland's transportation network is just of a single type, alternating grid pattern. It is not known if transportation network was of another type, for example Radial-ringed[2] or contour forming[3] street patterns would affect the relationships in this model. The definition of nearness changes in each street pattern

There is much to be gained from analyzing intersection using spatial regression techniques. The benefits of and "Ad-hoc" algorithm that engineers can use on site to measure what is happening at intersections are significant. "Before-and-after" or "with-without" intersection analysis may be applying results that may not be relevant due to external factors. The continuousness of the data can diminish some of the effects of a "Before-and-After" study.

Limitations of the Study

It should be mentioned first that due to the limitation of data, it is not possible to test all the potential explanatory variables for the ideal regression model. This is a deficiency of this analysis. Missing landuse at intersection could be an important form

---

[2] Radial-ringed street pattern: A Street pattern of loops or rings that are surrounded by successively larger ones. Usually found in conjunctions with larger radial patterns. Radial rings incorporate the elements of radial and ring/concentric designs. (Defense Science Board 1996)

[3] Contour-forming street pattern: A Street pattern where pronounced relief influences construction of roadways along lines of elevation. Primary streets run parallel to contour lines, with intersection roads connecting them. (Defense Science Board 1996)

variable that may affect intersection crash rate. Visibility may also be a factor in intersection crash rate. This variable added to the model would most like have a large effect in increasing its accuracy. Errors in the data also can affect the model since traffic volume is arduous to measure at the intersection level city wide. Areas where the traffic volume is calculated for some intersections may deviate from what is actually happening. These weaknesses do not invalidate the results of GWR but they reduce its potential by a noticeable amount.

In many cases Intersections do not follow the rule that states "things that are near each other tend to be similar" absolutely. Access corridors, travel directions, landuse variations, weather all play a part in making the rule of nearness difficult to apply. Some intersections that are near each other may exhibit no relationships because of this. It is impossible to separate all factors of the traffic network. Each traffic system affects others and therefore is related to any model involving traffic analysis.

Because of the characteristics of traffic network data it is difficult to do an accurately model intersection crash rates using just intersection form features. To effectively model crash rate all the factors that contribute crash rate should be included into the model. This is not the case due to data limitations. This situation should be taken into consideration when viewing the regression results. This study is an exploratory use of the using spatial regression techniques to model the effects on crash count.

Conclusion

Before-and-after studies are the primary way of analyzing of the effect of intersection form factors on crashes at intersections. With this method there are limitations that exist affect the decision making process. This study implemented an

81

alternative measure - spatial regression to understand the relationship between intersections form factors and traffic crashes. It was revealed that through geographically weighted regression (GWR) it is possible to measure that relationship in accurate and visually insightful manner. This study also proves GWR is a much more accurate method of analysis when compared to global linear regression. This method of measurement may be useful in supporting the intersection design decisions of the future.

Suggestions for Future Research

In future research it is imperative to include additional variables relating to crash rate. This will greatly increase the model's accuracy. Then crash rate predication modeling would be the next step in this research. Predicting intersection crash rates from un-observed explanatory variables would be an interesting study to undertake. More data and mitigation of model design limitations would be required to produce accurate and practical prediction results.

Also incorporating network specific data analysis is the next step in improving the accuracy of the model. Modeling the relationships of intersection on traffic network at a network specific level will go a long way; developing an adaptive distance method for analysis of intersections. Like crash reduction factors, intersection form reduction factors could be developed from a geographically weighted regression model. This could help bring a new purpose to regression modeling at intersection by building a database of the effect of intersection form factors in varying situations

REFERENCES

AASHTO. *AASHTO Strategic Highway Safety Plan.* Washington, DC: American Association of State Highway and Transportation Officials, 2004.

Beneficial Designs Inc. *Designing Sidewalks and Trails for Access Part II of II: Best Practices Design Guide.* Federal Highway Admininstration, 2001.

Charlton, Martin. *Geographically Weighted Regression White Paper.* Maynooth, Co Kildare: National Centre for Geocomputation, 2009.

Choi, Eun-Ha. *Crash Factors in Intersection-Related Crashes: An On-Scene Perspective.* Washington, DC: Mathematical Analysis Division, National Center for Statistics and Analysis , 2010.

Defense Science Board. *Field Manual No. 3-06 of Combined Arms operations in urban terrain.* Fort Benning, Georgia: Headquarters Department of the Army, October 1996.

Deza, Marie Michel, and Elena Deza. *Encyclopedia of Distances.* Springer, 2009.

ESRI. *ArcGIS Resource Center.* April 27, 2011. http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Interpreting_OLS_results/005p00000030000000/ (accessed October 11, 2011).

—. *How GWR Regression works.* January 28, 2009. http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How%20GWR%20Regression%20works (accessed October 9, 2011).

—. *How OLS regression works.* January 28, 2009. http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How%20OLS%20regression%20works (accessed October 9, 2011).

Federal Highway Administration. "Intersection Safety Issue Briefs." *FHWA Safety.* November 2009. http://safety.fhwa.dot.gov/intersection/resources/#fhwasa10005 (accessed October 1, 2011).

—. "Safety at Unsignalized Intersections." *FHWA Safety.* n.d. http://safety.fhwa.dot.gov/intersection/unsignalized/presentations/unsig_pps_041409/long.cfm (accessed October 2, 2011).

Fitzpatrick, Kay, Mark D. Wooldridge, and Joseph D. Blaschke. *Urban Intersection Design Guide.* College Station: Texas Transportation Institute, 2005, 4-4.

Highway Safety Manual. *Highway Safety Manual.* Washington, D.C: American Association of State Highway and Transportation Officials, 2010.

Hovey, Peter, and Mashrur (Ronnie) Chowdhury. *Development of Crash Reduction Factors.* Columbus: Ohio Department of Transportation, 2005.

Institute of Transportation Engineers. *Designing Walkable Urban Thoroughfares: A Context Sensitive Approach.* Washington, DC: Institute of Transportation Engineers, 2010 .

Institute of Transportation Engineers. *Toolbox on Intersection Safety and Design.* Washington, DC: Institute of Transportation Engineers, 2004.

Joksch, Hans C., and Lidia P. Kostyniuk. *Modeling Intersection Crash Counts and Traffic Volume.* Ann Arbor: University of Michigan Transportation Reasearch Institute, 1998.

Levinson, David, and Wei Chen. "Traffic Management System Performance Using Regression Analysis." *California Path Program Institute Of Transportation Studies*, 2006: 4.

MassHighway DOT. "Crash Rate Information." *MassHighway DOT.* 2011. http://www.mhd.state.ma.us/downloads/footprint/roads/procedures09.pdf (accessed October 4, 2011).

National Highway Traffic Safety Administration. "Fatality Analysis Reporting System Encyclopedia ." 2009.

New York City Department of Transportation. *Downtown Brooklyn Traffic Calming Project.* New York: Arup Services New York Ltd, 2003.

Oregon Metro Council. "Transit-Oriented Development Program." *http://library.oregonmetro.gov.* May 2011. http://library.oregonmetro.gov/files//tod_strat_plan_execsum.pdf (accessed October 6, 2011).

Orski, C. Kenneth. "The "Smart Growth" Debate Continues." *Innovation Briefs* Vol. 14, no. No. 3 (2003): 1.

Rodegerdts, Lee, Brandon Nevers, and Bruce Robinson. *Signalized Intersections: Informational Guide.* Portland: Kittelson & Associates, Inc., 2004.

Schrank, David, Tim Lomax, and Bill Eisele. *Urban Mobility Report.* College Station: Texas Transportation Institute, 2011.

Shen, Joan, and Albert Gan. "Development of Crash Reduction Factors." *Transportation Research Record: Journal of the Transportation Research Board* (Transportation Research Board of the National Academies) 1840, no. Paper No. 03-4345 (2007): 50-56.

Tarko, A., S. Eranky, and K. Sinha. *Methodological Considerations in the Development and Use of Crash Reduction Factors.* Washington D.C.: Transportation Research Board, 1998.

Tobler, W.R. "A Computer Model Simulating Urban Growth in the Detroit Region." *Economic Geography* Volume 46 (June 1970): 237-240.

U.S Census Bureau . *American FactFinder.* 2010. http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml (accessed September 20, 2011).

BIOGRAPHICAL SKETCH

Reginald Pierre-Jean was born in Nyack, New York in 1986. Majored and graduated in Geography at the University of Florida in 2008. This is where he took up an interest in geospatial relationships. Solving spatial problems is one of his passions.