공학석사 학위논문

Influences of Pedestrian Network on Pedestrian Crashes Considering Spatial Interactions

공간적 특성을 고려한 보행 네트워크의 보행 교통 사고에 대한 영향 분석

2016 년 2 월

서울대학교 대학원

건설환경공학부

박 세 현

Influences of Pedestrian Network on Pedestrian Crashes Considering Spatial Interactions

공간적 특성을 고려한 보행 네트워크의 보행 교통 사고에 대한 영향 분석

지도교수 고 승 영

이 논문을 공학석사 학위논문으로 제출함
2015 년 10 월

서울대학교 대학원
건설환경공학부
박 세 현

박 세 현의 석사 학위논문을 인준함
2015 년 12 월

위 원 장 이 청 원 (인)

부위원장 고 승 영 (인)

위 원 김 동 규 (인)

Abstract

Heading to the human-oriented society, walking is promoted trip mode. Walking is most fundamental travel behavior of human that occur most frequently in our daily lives. However, pedestrians are most vulnerable road users among others. Paying attentions to pedestrian-vehicular crashes, physical factors effecting the pedestrian behavior are considered in this study. A pedestrian start from an origin, walk on a path, and end up at a destination. This study focused on the geographic characteristics of these factors that leads to exposure to crashes. Using Geographically Weighted Poisson Regression instead of the traditional regression model, crash frequencies in Dongs of Seoul are analyzed based on the network characteristics in each Dong. Ratio of high-order roads connected to intersections, ratio of high-order roads length, ratio of crosswalks, and average block length were considered as independent variables. The built model revealed that it better fits than traditional model. Signs of four coefficients provided different relationship and result for each Dong. Dongs with same coefficient signs were grouped and gave possibility of interpretation. It is suggested that stakeholders to seek for countermeasures based on varying coefficients of independent variables.

Keyword : Pedestrian safety, Pedestrian-Vehicular crashes, Geographically Weighted Regression, Pedestrian network, Pedestrian trip generation
Student Number : 2014-20562

Table of Contents

List of Figures

List of Figures

Chapter 1. Introduction

1.1. Study Background

Walking is the oldest transportation mode in the history. Due to the tremendous growth in automobile ownership, cars displaced walking as the means of transportation. However, the statistic shows lots of people still walk to travel.

Table 1 Transportation Mode Use in 2010

Walking	Car	Bus	Subway/Rail	Taxi	Freight	Others
31.6%	37.2%	17.3%	6.8%	1.0%	2.7%	3.4%

Table 1 shows that walking is the second most used mode of transport. About one third of the total are traveled only on foot. In reality, walking is one of the process in the use of other modes. Apparently, it becomes the most frequent behavior in the process of travel.

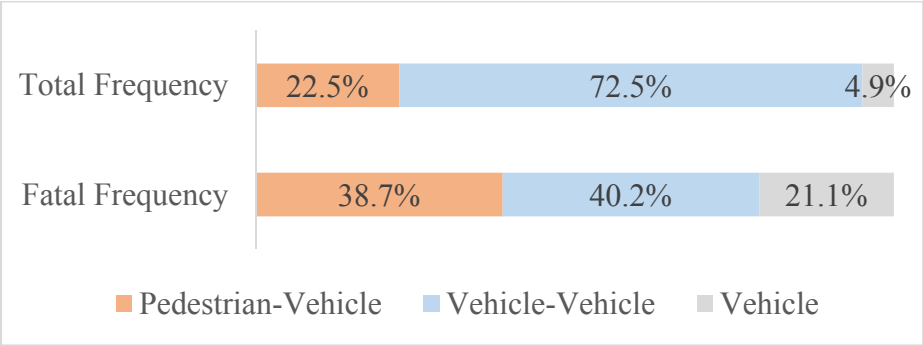


Figure 1 Fatality of Pedestrian-Vehicle Crashes

In the contrast, pedestrians are the most vulnerable road users of all as shown in Figure 1. Pedestrians took place 22.5% of the total accident. However, rate of the fatality of pedestrian was 38.7%. Once accident occurs, people are more likely to be killed while they are walking. As people are being encouraged to walk more, it is important to make pedestrian friendly environment.

The fundamental cause of a crash is the trip making itself which is the exposure to risk. Exposure is defined as the rate of contact with a potentially harmful agent or event. Pedestrian exposure refers to the rate of contact that a pedestrian has with vehicular traffic. Pedestrians are exposed to crash risk whenever they are walking in the vicinity of vehicular traffic.

Walking behavior is composed of physical factors; origin, destination, and walking path. These factors impact pedestrian safety by affecting the characteristics of traffic flow and travel behavior for both vehicles and pedestrian. The structure determines how direct a route is for drivers to follow and the number and types of turns encountered along the way. The connectivity, continuity, and shape of the route can affect vehicle speed and maneuvers, as well as driver visibility, thus impacting traffic safety (Quddus, 2008).

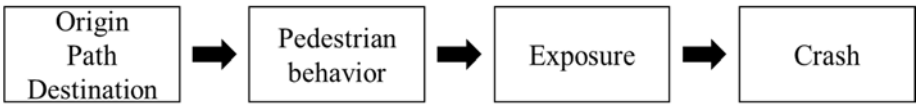


Figure 2 Pedestrian Crash Occurrence Process

According to Tobler's first law of geography, "Everything is related everything else, but near things are more related than distant things." The same concept is brought to this study. Trip generation is the first step of 4-Step-Model in transportation planning field. Production and attraction of a zone is known to be affected by various factors such as land use, demographic, income level and so on. These factors are mostly explained by the agglomeration effect. Agglomeration economies occur, for example, where lower transport costs bring firms closer together, resulting in lower unit costs and higher productivity (Venables, 2004). Interaction between these kinds of activities produces agglomeration forces that causes local concentration of activities. A point where a trip activity starts or ends are spatially distributed depending on the agglomeration effects of trip generating factors. A pedestrian network is composed of paths pedestrians walk on. Paths are physically continuous and pedestrian activities on paths are also continuously made. Route choices from origin to destination are made dependently on the built network.

1.2. Purpose of Research

Pedestrian-vehicular crashes are directly influenced by the pedestrian exposures to risk. Walking itself produces risks and depending on pedestrian behavior, the exposure to risk can be high or low. Pedestrian activities are highly related to the physical factors: origin, destination, and path. This study considers the spatial dependence of these factors to seek for better solutions for pedestrian safety.

Pedestrian network is composed of nodes which are origins and destinations, and links which are part of paths. This study search for variables that represent both pedestrian network and exposure to risk. Relationship between crash frequency and these variables are examined using geographically weighted regression technique.

Following steps are outline of this document that addresses these purposes:

- In chapter 2, previous studies related to this study are reviewed;
- In chapter 3, methods used in this study are described;
- In chapter 4, data descriptions are provided;
- In chapter 5, model structure and a synthesis of the results are presented;
- In chapter 6, conclusion and future studies are discussed.

Chapter 2. Literature Review

2.1. Safety Studies Related to Environmental Causes

Clifton and Kreamer-Fults (2007) examined pedestrian–vehicular crashes near public schools in Baltimore City of Maryland. Multivariate models of severity and crash risk exposure were developed as a function of social and physical characteristics of the area nearby schools. The result revealed that driveways and turning bays decreased both crash frequency and injury severity. The presence of recreational facilities increased them both. Variables such as transit access, commercial access, and population density came out to be significant.

Depaire et al. (2008) employed latent class clustering technique in order to identify homogenous traffic accident types. Total seven cluster were identified based on variables including road environments and accident information. For each cluster, multinomial model was built. The result showed that applying latent class clustering as a preliminary analysis helps to reveal hidden relationships and segment traffic accidents efficiently.

Dissanayake et al. (2009) investigated the research to establish a link between land use and child pedestrian casualties. Crash data and GIS data in Newcastle upon Tyne are analyzed to develop Generalized Linear Models (GLMs). Models differ dependent variable including total crashes, fatal crashes, and considering temporal variations. The result revealed that secondary retail and high density of residential area leads to higher total casualties. Primary retail and low density residential area are also associated with casualties at different time of the day and week.

Wier et al. (2009) proposed multivariate model to predict vehicle-pedestrian collisions based on environmental and demographic characteristics in San Francisco, California. Street features, land use, and population characteristics were selected as explanatory variables. The model explained about 72% of the systematic variation in vehicle-pedestrian collisions and included measures of traffic volume, arterial streets without transit, land area, proportion of neighborhood commercial and residential-neighborhood commercial uses, employee and resident populations, proportion of poverty population and proportion aged 65 and older.

Clifton et al. (2009) examined the impact of personal and environmental characteristics on severity of pedestrian-vehicle collisions injuries using a generalized ordered probit model. Pedestrian–vehicle crash data for Baltimore City were used and land use, urban form and transportation information specific to the individual crash locations were also considered. The results revealed that women pedestrians were less involved in crashes than males; children were more likely to

be in sustaining injuries and elderly persons are more likely to be fatally injured. Intersections without a crosswalk and at night were associated with higher injury risk.

Zahabi et al. (2011) proposed to identify the impacts of road design, built environment, and other factors on severity of pedestrians and cyclists involved collisions. Dataset from police reports in city of Montreal, Canada were provided as well as road information and built environment information. Lighting condition, vehicle movement, presence of an intersection, vehicle type were highly associated with severity of non-motorist collisions.

Ukkusuri et al. (2012) developed pedestrian accident frequency models based on land use and road design in New York City, US. Dependent variables were total collisions and fatally injured collisions. Land use, demographics, transit infrastructure, road network and travel behavior were provided as independent variables. Greater proportion of industrial, commercial, and open land use contributed to more frequent crashes. Number of schools and transit stops also affected frequency.

Mohamed et al. (2013) combined data mining technique and regression methods to identify the main factors associated with the severity of pedestrian injuries. This data from New York City, US (2002–2006) and the City of Montreal, Canada (2003–2006) were used in this research. General injury severity models were developed for sub-group of populations obtained by clustering method. The technique used in this paper reveals how the segmentation of the accident contribute to efficiently showing the relationship between severity outcomes and built environment and socio-demographic features. For data from New York City, a latent class with ordered probit model came out with best results, however, K-means with a multinomial logit model provided best results for City of Montreal. Independent variables such as pedestrian age, location type, driver age, vehicle type, alcohol involvement, light conditions, and others contributed to fatal crashes.

Park (2014) identified the physical environmental characteristics that affect pedestrian-vehicular crashes through investigating partial patterns of the crashes with GIS. Negative binomial regression method had been employed to reveal the relationship between physical environment and the pedestrian-vehicular collisions in Seattle, US, from 2000 to 2004. Traffic circle density and traffic signal density among street design, the percentage of single-family housing areas, the number of bus residential areas found to be associated with the crashes.

Park (2014) investigated the study to explore environmental correlations of pedestrian-vehicular collisions within the neighborhood schools. Negative binomial regression model was developed using data from 2000 to 2004. The result indicated that sidewalk density, traffic signal density, cul-de-sac density, the number of fast-

food restaurants, the number of parking lots, and the number of residential units are positively affects the crashes.

2.2. Spatial Studies on Crashes

Cloutier et al. (2007) integrated socio-economic and environmental data into a geographic information system to develop a pedestrian safety model in school environment. Geographically weighted regression was selected as the model. Results demonstrated that the average distance between accident points and schools were less than 500 meters. Each risk factors spatially varied suggesting different countermeasures to different neighborhoods.

Li et al. (2011) employed the GWR model to predict intersection crashes in City of Chicago. Crash data from 2001 to 2008 were used to develop the model. Independent variables such as major and minor road daily traffic, number of major and minor road through and left-turn lanes, and household income level affected intersection crashes. Through the analysis of variance (ANOVA) test, GWR model improved the previous model. The Monte Carlo test revealed significance of spatial variability of independent variables in the model.

Zhang et al. (2011) reported that the road network connectivity affects the non-motorist traffic safety since the network connectivity impact the non-motorist travel behavior. Four measures including block density, intersection density, street density, and mean block length are calculated based on 321 census tracts in Alameda County, California. Including four network connectivity measures, other variables such as traffic behavior, land use, transportation facility, and demographic features are added to GWR model. The result demonstrated that higher connectivity contributes to the decrease of pedestrian-bicyclist accident. Among the network connectivity measures, street density were more stable and explainable than other measures.

Li et al. (2013) built Geographically Weighted Poisson Regression (GWPR) to explain the impact of spatial heterogeneity on county-level crash. Traffic patterns, road network attributes, and socio-demographic features of 58 counties in California were concerned as explanatory variables in the model. The result of GWPR was compared with that of GLM. The GWPR outperformed the GLM in the prediction of crashes.

Pirdavani et al. (2014) simulated the impact of teleworking on the traffic safety in Flanders of Belgium. Since the spatial interaction exist in many spatial variables affecting crashes, GWR model is adopted to capture the spatial correlation. Crashes from 2004 to 2007 were used and network and socio-demographic variables of 2200 traffic analysis zones were selected as the independent variable. Crash prediction models were developed for null scenario and teleworking scenario. The later

scenario revealed that the policy decreases crashes in the study area.

Shariat-Mohaymany et al. (2015) argued that trip generation as a function of land use, socio-economic, and demographic characteristics, network characteristics, and traffic volume affect crashes. Generalized Linear Model (GLM) and GWPR model are employed based on 253 traffic analysis zones in Mashhad, Iran. GWPR outperformed in showing the non-stationary of the phenomenon.

Table 2 Methods in Previous Studies

Author	Kernel Function	Optimization Method	Model	Goodness-of-fit
Cloutier et al. (2007)	Gaussian Kernel	CV	GWR	Moran's I ANOVA
Li et al. (2011)	-	-	GWR (lognormal)	ANOVA Monte Carlo test
Zhang et al. (2011)	-	-	GWPR	AICc
Li et al. (2013)	Adaptive Kernel	AICc	GWPR	MAD MSPE Moran's I AICc
Pirdavani et al. (2014)	-	-	GWPR	Moran's I
Shariat-Mohaymany et al. (2015)	Adaptive Kernel	AICc	GWPR	Adjusted Pearson chi-square Adjusted deviance MAD MSPE

2.3. Network Performance Measures and Indices

The role of a network is to allow users to move conveniently and safely to the destination. It is hard to measure a network performance in a direct way. Various measures and indices are suggested by several works.

Dill (2004) defined and described measures of network connectivity drawn from multiple fields, including transportation, urban planning, geography, and landscape ecology.

Table 3 Network Connectivity Measures

Measure	Value Type	Description
Block length	Mean	From an intersection curb or centerline to another
Block size	Mean	Perimeter or area
Block density	Density	Number of census blocks divided by unit area
Intersection density	Density	Number of intersection divided by unit area
Street density	Density	Total street length divided by unit area
Connected node ratio	Ratio	Number of real node divided by number of total nodes
Link-node ratio	Ratio	Number of link divided by number of node
Grid pattern	Binary	Grid pattern or others
Pedestrian route directness	Ratio	Route distance divided by Euclidean distance
Effective Walking Area	Ratio	Number of parcel within a walking distance divided by number of parcel within a circle area
Gamma index	Ratio	Number of link divided by number of maximum possible link
Alpha index	Ratio	Number of circuit divided by number of maximum circuit

Rodrige et al. (2013) mentioned several measures and indices from field of graph theory.

Table 4 Measures in Graph Theory

Measure or indices	Description	Definition or measurement
Diameter	The extent of a graph	The length of the shortest path between the most distanced nodes of a graph
Number of Cycles	The level of development and complexity of a transport system	Number of links minus number of nodes plus number of subgraphs
Order of a node	The importance of a node	The number of attached links in a graph
Detour index	The efficiency of a transport network	Transport distance, real distance divided by straight distance
Network density	The territorial occupation of a transport network	Length of links divided by area of surface
Pi index	The level of development of a transport system	Total length of a graph divided by distance along its diameter
Eta index	The relationship between a network as a whole and its edges	Average length per link
Theta index	The function of a node	Average amount of traffic per intersection
Beta index	The level of connectivity in a graph	Number of links divided by number of nodes

Kim et al. (2014) presented a systematic method of evaluation with the premise that satisfaction rate scale of the walking will vary according to the characteristics of land use by footpath types. Pedestrian safety elements are applied to evaluate the pedestrian environment. Separation from driveway, separation from bikeway, and etc. are considered.

2.4. Summary and Research Needs

Papers on pedestrian-vehicular crashes are reviewed based on built environmental causes and spatial analysis. Network measures and indices are also reviewed. Presence of facilities, land use features, and network characteristics were considered as the influencing factors on the crashes. Factors mentioned lead to exposure to crashes. Variables related to exposure are noted in previous studies reviewed.

In this study, exposures are considered as previous studies but in a quite different ways. Intersections are noticed to be a remarkable element that determine form and characteristics of network which also affect the measures and indices. Intersections are points where pedestrians and vehicles meet directly. Thus, exposure to crashes increases near intersecting points. Crosswalks which are a part of a pedestrian network are also a place where direct collisions occur. Network connectivity measures reveal accessibility of these network elements that are exposed to risks.

Chapter 3. Methods

3.1. Spatial Autocorrelation and Spatial Heterogeneity

As Tobler (1970) established the first law of geography, data possess not only the data itself but also geographical data near it. The phenomenon that spatial characteristics of objects affecting each other is called spatial interaction. Spatial autocorrelation is spatially dependent distribution of objects. Anselin and Bera (1998) defined spatial autocorrelation as "Spatial autocorrelation can be loosely defined as the coincidence of value similarity with locational similarity. In other words, high or low values for a random variable tend to cluster in space (positive spatial autocorrelation) or locations tend to be surrounded by neighbors with very dissimilar values (negative spatial autocorrelation). Of the two types of spatial autocorrelation, positive autocorrelation is by far the more intuitive. Negative spatial autocorrelation implies a checkerboard pattern of values and does not always have a meaningful substantive interpretation." Positive and negative spatial autocorrelation are two types and each indicates that similar values are spatially more clustered and more dispersed.

Spatial heterogeneity is also a phenomenon that occurs in geography. It is usually defined generally as the complexity and variability of a system property in space (Li, 1994). It can also be refer as a mix of concentrations of multiple elements within a area.

Figure 4 depicts the agglomeration effects of origin and destination. As mentioned in Chapter 1, zones with similar land use, similar demographics, and similar incomes characteristics are closer to each other. These factors affect trip generation of a zone. Thus, it is clear that trip generation of zones are correlated to each other. Figure 5 shows that closer the zones are, productions and attractions trips are likely to be similar. The figure also reveals that distant zones contains different characteristics.

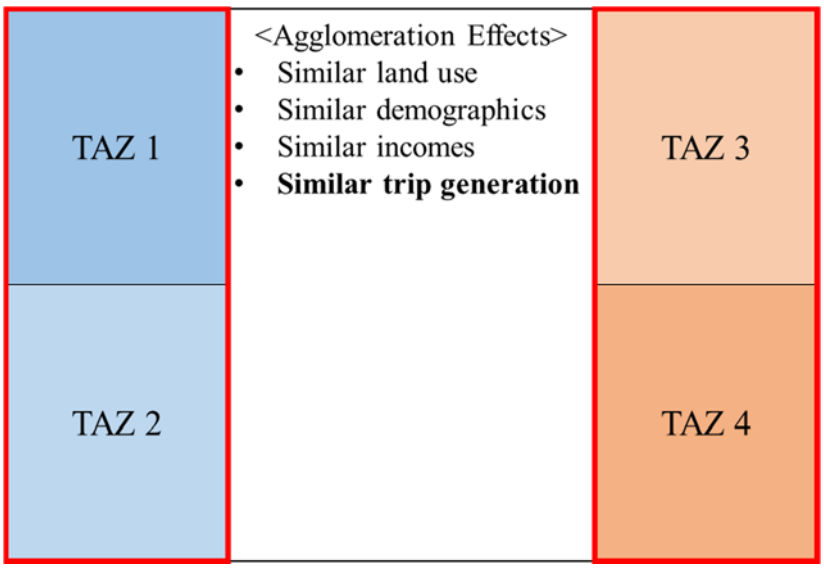


Figure 3 Agglomeration Effects of O/D

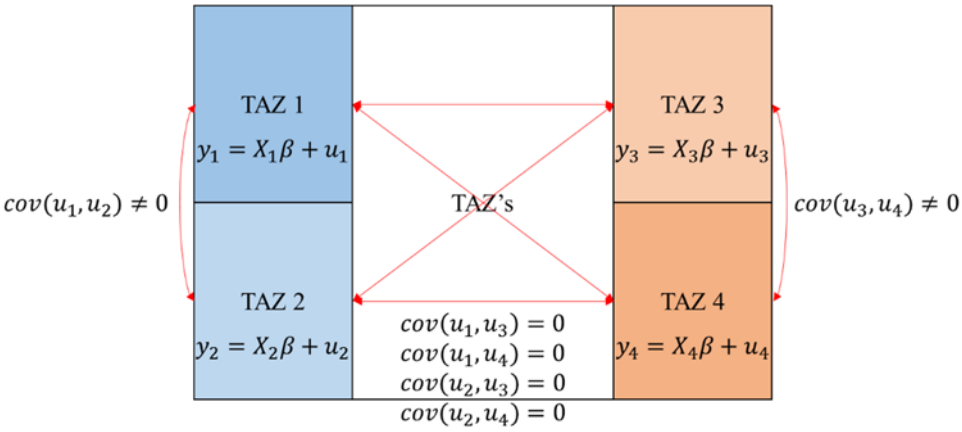


Figure 4 Spatial Autocorrelation of O/D

Figure 6 depicts the spatial characteristics of a network. Trip behavior is a continuous and spatially dependent activity. Travelers walk on the network and closer network sections have spatial autocorrelation and distant sections are less dependent and have spatial heterogeneity.

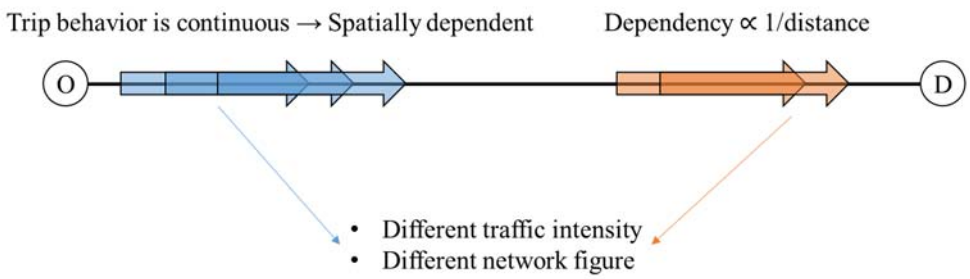


Figure 5 Spatial Autocorrelation and Heterogeneity of N/W

3.2. Geographically Weighted Regression (GWR)

3.2.1. Geographically Weighted Regression (GWR)

Traditional regression model cannot capture the spatial phenomenon explained in the previous section. Geographically weighted regression is the model developed to overcome this problem. Geographically weighted regression (GWR) was developed to allow relationships between dependent and independent variables to vary across space (Fotheringham et al., 2002). A traditional regression model is written as

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i$$

where y_i is the dependent variable in the location i; β_0 is the intercept of the model; k is the total number of independent variables; β_k is the parameter of the kth independent variable; $x_{ik}$is the kth independent variable observed in location i; and ε_i is the error term in the location i. The parameter β_k is a global one that does not change over space.

GWR is a local model rather than global model that does not vary over space.

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

where (u_i, v_i) is the coordinates of the ith location usually a centroid of the census tract; $\beta_k(u_i, v_i)$ is a local parameter at location point i which is the continuous function. Thus, GWR allows spatial variations of parameters which is a reasonable way to capture the spatial interactions between objects and spatial heterogeneity.

Table 5 Comparison between Two Models

Global Model	Local Model
• Summarize data for whole region • Single-valued statistic • Non-mappable • GIS-unfriendly • Spatially limited • Emphasize similarities across space • Search for regularities • Classic Regression	• Local disaggregations of global statistics • Multi-valued statistic • Mappable • GIS-friendly • Spatial • Emphasize differences across space • Search for exceptions • Geographically Weighted Regression

Source: Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression*. John Wiley & Sons, Limited.

The basic GWR assumes a normally distributed error structure in the model, however in the case of the count data such as crash data, a Poisson distribution is more appropriate. Although a negative binomial distribution is known to be more effective than a Poisson distribution due to the over-dispersion, Poisson regression does not cause inaccurate estimates (Hadayeghi et al., 2010). Poisson type of GWR is written as

$$\ln(y_i) = \ln(\beta_0(u_i, v_i)) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

The parameter $\beta_k(u_i, v_i)$ can be expressed as following matrix form

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0(u_1, v_1) & \beta_1(u_1, v_1) & \cdots & \beta_k(u_1, v_1) \\ \vdots & & \ddots & \vdots \\ \beta_0(u_n, v_n) & \beta_1(u_n, v_n) & \cdots & \beta_k(u_n, v_n) \end{bmatrix}$$

In a vector form, it is written as

$$\widehat{\boldsymbol{\beta}}(i) = (\boldsymbol{X}'\boldsymbol{W}(i)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(i)\boldsymbol{Y}$$

where $W(i)$ denotes spatial weight matrix that is convenient expressed from of $W(u_i, v_i)$ and the matrix form follows

$$\boldsymbol{W}(i) = \begin{pmatrix} w_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{in} \end{pmatrix}$$

where $w_{ij}(j = 1,2, \dots, n)$ is the weight given to point j in the calibration of model for point i. For each point i, a model is calibrated based on observed points near the point. Weights are inversely proportional to the distance between two points. Thus, points closer to the regression point are weighted more than ones that are farther from the point.

Most commonly used weighting schemes are the Gaussian and bi-square functions.

$$w_{ij} = \exp[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2]$$

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2, & if\ d_{ij} < b \\ 0, & otherwise \end{cases}$$

The parameter b denotes the bandwidth. The bandwidth is constant in the Gaussian function which is called fixed kernel then the weight functions are same for all points. A problem might arise in the model with Gaussian function where points are sparse which causes parameter estimates with large standard errors. An alternative weight function is the bi-square function also known as adaptive kernel that allows weights to vary over space according to the density of the data point: large bandwidths where the data are sparse and small bandwidths where the data are dense.

Methods of selecting the optimal bandwidth are Cross-Validation and the corrected Akaike Information Criterion (AICc). The basic idea of Cross-Validation method is by using part of the data to evaluate the model instead of the entire data set, rest of the data is used to test the predicting performance of the model. AICc is a model selection method that minimizes Kullback-Leiber distance between the models. Lower the value of two values indicate bandwidth with better performance.

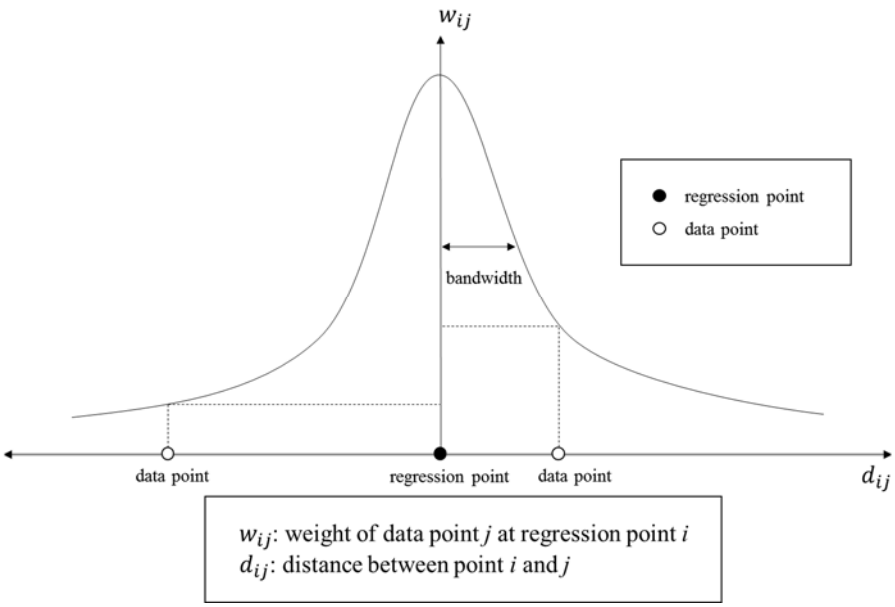


Figure 6 Bandwidth

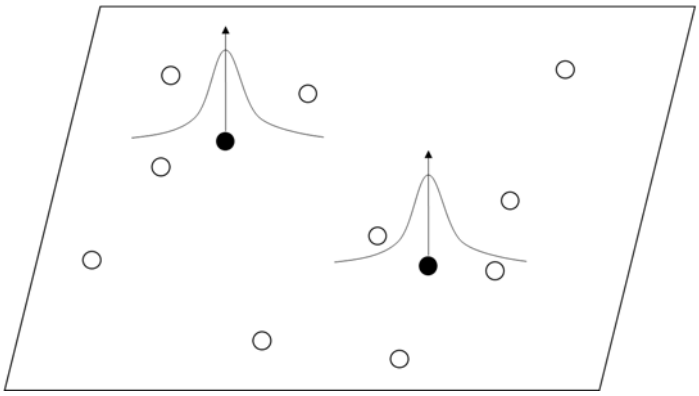


Figure 7 Fixed Bandwidth

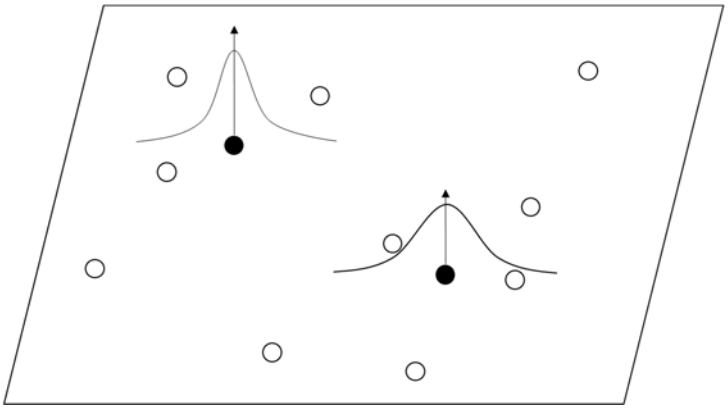


Figure 8 Adaptive Bandwidth

Table 6 Bandwidth Optimization Method

Method	Description
Cross-Validation (CV)	$\mathrm{CV} = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(b)]^2$ $\hat{y}_{\neq i}(b)$: fitted value of y_i with the observations for point i omitted from the calibration process
Corrected Akaike Information Criterion (AICc)	$\mathrm{AICc} = 2nln(\hat{\sigma}) + nln(2\pi) + n\{\frac{n + tr(\boldsymbol{S})}{n - 2 - tr(\boldsymbol{S})}\}$ n: sample size $\hat{\sigma}$: estimated standard deviation of the error term $tr(\boldsymbol{S})$: trace of hat matrix which is a function of the bandwidth ($\hat{y} = \boldsymbol{S}y$)

3.2.2 Goodness of Fit

The first goodness of fit measure is Mean Absolute Deviation (MAD). The MAD provides a measure of the average misprediction of the model.

$$\mathrm{MAD} = \frac{\sum_{i=1}^{N} |\hat{Y}_i - Y_i|}{N}$$

$\hat{Y}_i$: Predicted number of crashes at region i

Y_i: Observed number of crashes

For a fitted Poisson regression the deviance equals to

$$D = 2\sum_{i=1}^{n}\{Y_i \log(Y_i/\mu_i) - (Y_i - \mu_i)\}$$

Where, $\mu_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k)$, predicted mean for observation i based on the estimated model parameters.

The observed values Y_i will be close to their predicted means μ_i, causing both terms in D to be small, so the deviance to be small.

The value of the AIC itself has no meaning, however it is useful when models are compared to reveal the best one. The model with the lowest AIC value is the best model among others.

Chapter 4. Data

4.1. Data resource

Pedestrian-vehicular crash data was provided by the Traffic Accident Analysis System (TAAS) of the Korea Road Traffic Authority (KoROAD). A three-year frame data, from 2009 to 2013, were collected from the 467 Dongs in the city of Seoul. Road network attributes were provided by the Ministry of the Interior through the road name address digital map in format of shape file. Origin and destination (O/D) patterns were obtained from the Korean Transport Database (KTDB) Center. The trip volume data based on trip mode in 2013 was used in this study.

4.2. Data description

The frequency data is aggregated based on 467 legal Dongs in Seoul. Crash time, climate, driver characteristics, pedestrian characteristics, and other features are included in the data.

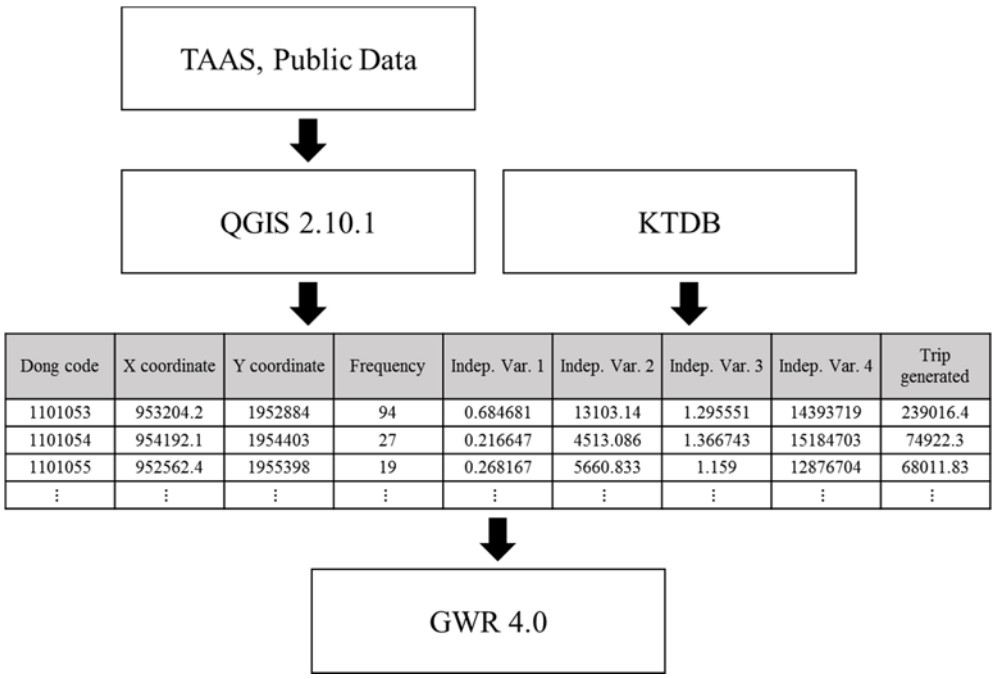
The road name address digital map contains variety of shape files including buildings, their entrances and exits, parks, rivers, intersections, roads, subways, and so on. Based on Dong of administration shape file, data are aggregated.

The O/D pattern data is consist of several transportation modes including walking, vehicle, various bus types, subway, and others. Trips from a zone to all the other zone including the zone itself are provided in matrix form.

Table 7 Dong Types

Legal Dong (467 Dongs)	Dong of Administration (423 Dongs)
TAAS Crash data	KTDB O/D data MOI Digital map

Two types of Dong exist in Korea: legal Dong and Dong of administration. The legal Dong which is more segmented than Dong of administration, often provides with small sample data within a Dong. Thus, in this study, Dong of administration is the basic TAZ. All data are modified based on Dong of administration using QGIS 2.10.1, a powerful tool in geography field.



Dong code	X coordinate	Y coordinate	Frequency	Indep. Var. 1	Indep. Var. 2	Indep. Var. 3	Indep. Var. 4	Trip generated
1101053	953204.2	1952884	94	0.684681	13103.14	1.295551	14393719	239016.4
1101054	954192.1	1954403	27	0.216647	4513.086	1.366743	15184703	74922.3
1101055	952562.4	1955398	19	0.268167	5660.833	1.159	12876704	68011.83
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Figure 9 Data Process

Chapter 5. Model

4.1. Model Structure

Dependent variable is a number of crashes in each Dong from 2009 to 2013. Exposure variable for the GWPR model is trips made by pedestrians, buses, and subways. Since travelers using transits transfer to pedestrians from origin to stops or from stops to destination, they are included as an exposure variable.

One of the four independent variable is ratio of number of intersections with higher-order roads divided by total number of intersections. Hierarchy of roads are classified by width in 4 levels: highways, Daero, Ro, Gil. Intersections are classified based on hierarchy of connected roads. Gil is mostly local roads with one or two lanes where vehicles cannot speed up so that accidents are hardly expected to occur at points where only Gils intersect. Rest of the road hierarchies are regarded as higher-order since classification is based on drivers' view and above Ro are high enough to pedestrians.

Table 8 Hierarchy of Roads

Hierarchy of Roads	Width	Number of Lanes
Highway	-	-
Daero	Over 40m	Over 7 lanes
Ro	12m ~ 40m	2 ~ 7 lanes
Gil	Else	Else

Ratio of high-order roads length divided by total roads length are also considered. Roads with low-order hierarchy tend to contain mixed traffic which increase exposure. However, vehicles in local roads are limited in speed so that crashes might not occur.

Ratio of crosswalk length to total road length is considered. Crosswalk is a road facility that aim to protect pedestrian's right to cross roads safely. Depending on the environment, crosswalks might increase or decrease collisions.

Average block length is a network connectivity measure. A higher value of the measure provide more choice of routes to vehicles and pedestrians. Table 9 shows descriptive statistics of four independent variables of 423 Dongs in Seoul. Independent variable 1 to 4 refer to high-order intersection ratio, high-order road ratio, crosswalk ratio, and average block length.

Table 9 Descriptive Statistics

	Frequency	Total Trip	Indep. Var. 1	Indep. Var. 2	Indep. Var. 3	Indep. Var. 4
Min	4	13131.59	0.02	0	0.002	30.80
Max	188	475994.80	1.22	0.84	0.081	93.82
Mean	36.17	80825.77	0.29	0.23	0.023	53.83
S.T.D..	23.56	58318.35	0.19	0.13	0.014	12.29

GWPR model is developed using the software GWR 4.0. For the kernel function in the model, adaptive kernel with bi-square function is used as following reasons.

- Fixed kernel (Gaussian function): with few points given, parameter estimates result in large standard errors.
- Adaptive kernel (Bi-square function): allows spatially varied weight according to the density of the data. With sparse data points, it has larger bandwidth and with densed points, it has smaller bandwidth.

Golden section method is applied to find the optimum bandwidth. It is to find the best bandwidth that satisfies both biasness and variance using the AICc value.

4.2. Analysis and Result

The frequency of each Dong is depicted in Figure 10.

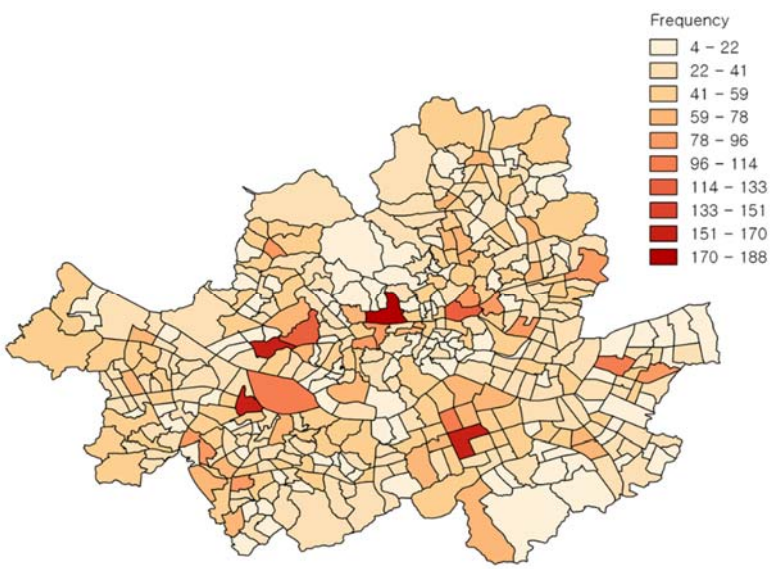


Figure 10 Frequency of Each Dong

The output of the GWPR model is different location-specific estimates for each zone. All variables estimates are functions of each zone. Table 10 summarizes statistics for coefficients of the model.

Table 10 Summary Statistics for Coefficient

Independent Variables	Mean	S.T.D.	Min	Max	Range
Intercept	-7.67	0.20	-8.46	-7.20	1.26
High-order intersection ratio	0.14	0.27	-0.45	1.13	1.58
High-order road ratio	-0.12	0.25	-0.75	0.63	1.39
Crosswalk ratio	0.06	0.23	-0.39	0.69	1.08
Average block length	-0.10	0.17	-0.62	0.34	0.96

Signs of the coefficients are depicted as follows. Blue colored Dongs in the map below indicates positive coefficients, red colored Dongs are negative coefficients, and black colored Dongs are not significant. Figure 11 shows that majority of Dongs are positive signs. Crashes are more likely to occur in these Dongs when intersections connected to Daero and Ro. In Dongs with negative signs, conversely, as intersections connected to Gils increase, crashes occur more frequently. Figure 12 reveals that majority of Dongs are negative signs. Collisions increase where a Dong contains high percentage of local roads and decreases where the percentage is low. As in Figure 13, half of Dongs show as crosswalk ratio increases crashes and inversely effects in rest of Dongs. Figure 14 shows that well-connected networks (shorter average block length) increases crashes in majority of Dongs which are colored in red.

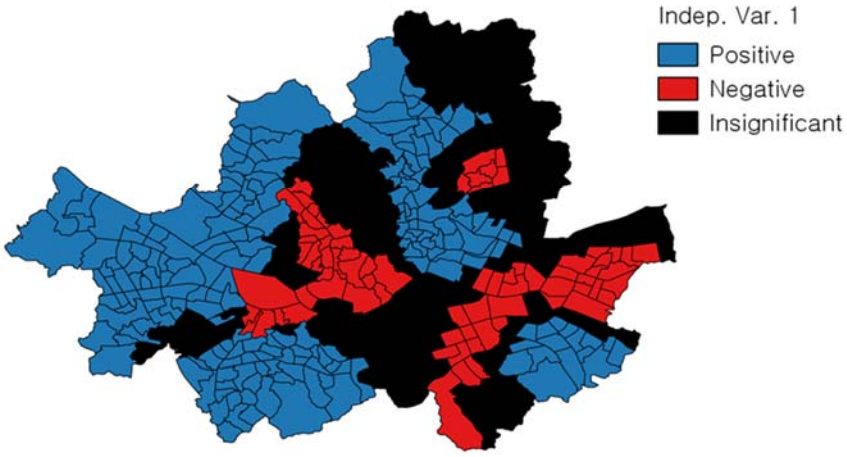


Figure 11 High-order Intersection Ratio

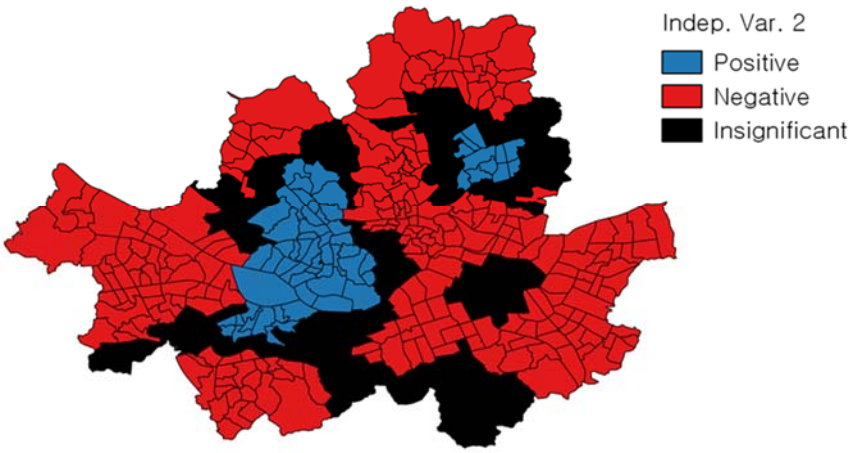


Figure 12 High-order Road Ratio

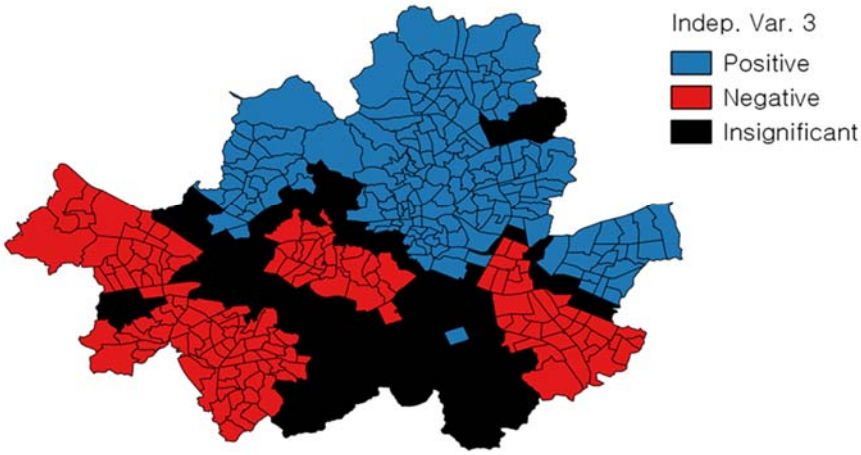


Figure 13 Crosswalk Ratio

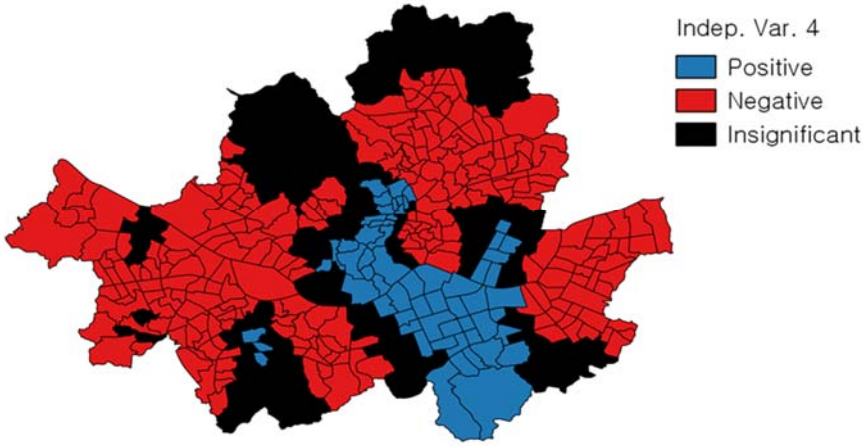


Figure 14 Average Block Length

Four independent variables showed different patterns of coefficient signs. Table 11 shows cases that are grouped with same signs of four variables. Cases are ordered in the number of Dongs each contain. In Dongs of case 1, crashes occur more in networks with high ratio of high-order roads connected to intersections, low ratio of high-order roads length, low ratio of crosswalks, and well connected roads. In case 2, only difference with case 1 is that more crosswalks increase crashes. Each cases can be explained with the same process.

Table 11 Cases with Same Coefficient Signs

Case	Indep. Var. 1	Indep. Var. 2	Indep. Var. 3	Indep. Var. 4	Accum. Num. of Dongs	Accum. %
1	+	-	-	-	45	11%
2	+	-	+	-	80	19%
3	+	-	-	x	107	25%
4	x	-	+	x	132	31%
5	+	x	+	-	152	36%
6	x	x	+	-	172	41%
7	+	-	x	-	189	45%
8	+	-	+	x	204	48%
9	-	-	+	-	218	52%
10	+	x	x	-	231	55%
11	x	-	+	-	244	58%
12	x	-	x	+	254	60%
13	x	x	x	x	264	62%
14	-	+	-	x	273	65%
15	x	+	-	-	282	67%
16	x	x	-	-	291	69%
17	-	+	-	+	298	70%

The MAD, AICc, and Deviance showed that the local model fits better than the global one. Varying effects of different explanatory variables on crash occurrence in different parts of the study area. Each Dong has its own set of coefficients and they can explained by the signs. It is suggested to apply countermeasures based on these data.

Table 12 Goodness of Fit

	MAD	AICc	Deviance
Global model	3763.14	418.00	9.00
GWPR model	2232.24	301.67	7.40
Difference	1530.89	116.33	13.16

Chapter 6. Discussion

6.1. Conclusion

This study describes the impact of pedestrian network characteristics on pedestrian-vehicular crashes using geographical weighted regressing technique. Mentioned little in previous studies, this study included intersections which are important element in network and also critical in crash analysis. The hierarchy of roads and the ratio of crosswalks are considered in order to apply exposure to risk in the model. Average block length, one of the network connectivity measure, provide information on complexity of pedestrian routes.

Signs of independent variables provide network features that increase exposure to crashes and also countermeasures to reduce them. High-order intersection ratio suggest types of intersections each Dong need to care. High-order road ratio also provide types of roads each Dong need to focus. The coefficient of crosswalk ratio reveals the appropriateness of location and design of the crosswalk. Average block length as a network connectivity measure, let stakeholders determine whether pedestrian routes are related to crash frequencies. Each dongs are given with different set of coefficient signs which lead to different interpretation and countermeasures. The remarkable conclusion of this paper is that impact of pedestrian network on pedestrian-vehicular crashes vary over districts and stakeholders need to be aware of the differences and apply the countermeasures.

6.2. Further Research

This study is based on data that is aggregated in a TAZ level. Crash data with coordinates provide rich information of network and crash itself. Variables can be more accurate leading to the development of more realistic model. Also, building information can be combined with trip generating weights for trip generation will give specific origins and destinations.

References

Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *statistics textbooks and monographs*, 155, 237-290.

Clifton, K. J., & Kreamer-Fults, K. (2007). An examination of the environmental attributes associated with pedestrian–vehicular crashes near public schools.*Accident Analysis & Prevention*, *39*(4), 708-715.

Clifton, K. J., Burnier, C. V., & Akar, G. (2009). Severity of injury resulting from pedestrian–vehicle crashes: What can we learn from examining the built environment?. *Transportation research part D: transport and environment*,*14*(6), 425-436.

Cloutier, M. S., Apparicio, P., & Thouez, J. P. (2007). GIS-based spatial analysis of child pedestrian accidents near primary schools in Montréal, Canada. *Applied GIS*, *3*(4), 1-18.

Dill, J. (2004, January). Measuring network connectivity for bicycling

Dissanayake, D., Aryaija, J., & Wedagama, D. P. (2009). Modelling the effects of land use and temporal factors on child pedestrian casualties. *Accident Analysis & Prevention*, *41*(5), 1016-1024.

Depaire, B., Wets, G., & Vanhoof, K. (2008). Traffic accident segmentation by means of latent class clustering. *Accident Analysis & Prevention*, *40*(4), 1257-1266.

and walking. In *83rd Annual Meeting of the Transportation Research Board, Washington, DC* (pp. 11-15).

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

Hadayeghi, A., Shalaby, A. S., & Persaud, B. N. (2010). Development of planning level transportation safety tools using geographically weighted Poisson regression. Accident Analysis & Prevention, 42(2), 676-688.

Li, H., & Reynolds, J. F. (1994). A simulation experiment to quantify spatial heterogeneity in categorical maps. *Ecology*, *75*(8), 2446-2455.

Li, Z., Wang, W., Liu, P., Bigham, J. M., & Ragland, D. R. (2013). Using geographically weighted Poisson regression for county-level crash modeling in

California. *Safety science*, *58*, 89-97.

Matkan, A. A., Mohaymany, A. S., Mirbagheri, B., & Shahri, M. (2011). Explorative Spatial Analysis of Traffic Accidents Using GWPR Model for Urban Safety Planning. In *3rd International Conference on Road Safety and Simulation*.

Mohamed, M. G., Saunier, N., Miranda-Moreno, L. F., & Ukkusuri, S. V. (2013). A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada. *Safety science*, *54*, 27-37.

Park, S. H., Kim, D. K., Kho, S. Y., & Kang, S. (2012). Spatial Analysis Methods for Identifying Hazardous Locations on Expressways in Korea. In *Transportation Research Board 91st Annual Meeting* (No. 12-3162).

Pirdavani, A., Bellemans, T., Brijs, T., Kochan, B., & Wets, G. (2014). Assessing the road safety impacts of a teleworking policy by means of geographically weighted regression method. *Journal of transport geography*,*39*, 96-110.

Peponis, J., Allen, D., French, S., Scoppa, M., & Brown, J. (2007). Street connectivity and urban density. In *6th International Space Syntax Symposium Istanbul: ITU Faculty of Architecture*.

Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accident Analysis & Prevention, 40(4), 1486-1497.

Rodrigue, J. P., Comtois, C., & Slack, B. (2013). *The geography of transport systems*. Routledge.

Shariat-Mohaymany, A., Shahri, M., Mirbagheri, B., & Matkan, A. A. (2015). Exploring Spatial Non-Stationarity and Varying Relationships between Crash Data and Related Factors Using Geographically Weighted Poisson Regression.*Transactions in GIS*, *19*(2), 321-337.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography, 234-240.

Ukkusuri, S., Miranda-Moreno, L. F., Ramadurai, G., & Isa-Tavarez, J. (2012). The role of built environment on pedestrian crash frequency. *Safety science*,*50*(4), 1141-1151.

Venables, A. J. (2007). Evaluating urban transport improvements: cost-benefit analysis in the presence of agglomeration and income taxation. Journal of transport

economics and policy, 173-188.

Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention*, *41*(1), 137-145.

Zhang, Y., Bigham, J., Li, Z., Ragland, D., & Chen, X. (2012, January). Associations between road network connectivity and pedestrian-bicyclist accidents. In *91st Annual Meeting of the Transportation Research Board, Washington DC, Washington, DC*.

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

김숙희, 이규진, & 최기주. (2014). 토지이용을 고려한 보행환경 평가지표 개발 및 적용에 관한 연구. *대한토목학회논문집*, *34*(3), 931-938.

박승훈. (2014). 근린환경이 보행자-차량 충돌사고에 미치는 영향. *국토계획*,*49*(3), 143-157.

박승훈. (2014). 안전한 통학로 조성을 위한 보행자-차량간 교통사고와 학교주변의 물리적 환경과의 연관성 연구. *대한건축학회 논문집-계획계*, *30*(8), 181-189.

국문 초록

최근 차대사람 사고의 원인 규명과 그에 대한 많은 대책이 제시되고 있다. 차대사람 사고는 차와 사람의 행위간에 발생한다. 본 연구에서는 물리환경적 특성이 보행자 이동행태에 미치는 영향에 주목하고 사고와의 연관성을 도출하고자 하였다. 물리환경적 특성은 네트워크 특성으로 나타내었으며, 네트워크 특성 중 보행 사고에의 노출이 반영된 도로 위계에 따른 교차로 비율, 도로 위계별 연장 길이 비율, 횡단보도 비율, 평균 블록 길이를 반영하였다. 이러한 네트워크 특성이 지니는 공간적 상호작용을 공간 회귀 가중 모형을 통해 반영하여 보다 현실적이고 적합한 모형을 구축하였다. 서울시 423개 행정동의 공간 자료 및 사고 자료를 이용하여 공간 가중 포아송 회귀 모형을 구축하였다. 그 결과, 전통적인 회귀 모형에 비해 적합함이 나타났다. 각 계수의 부호에 따라 해석이 가능하며, 모든 계수의 부호가 같은 행정동들의 군집별 특성을 살펴볼 수 있었다. 각 경우에 따라, 관리가 필요한 교차로의 특성, 도로의 위계 등의 네트워크 특성이 다르게 나타났다.

주요어 : 보행 사고, 차대사람, 보행 네트워크, 공간 가중 회귀
학번 : 2014-20562