IDENTIFICATION OF URBAN TRAFFIC ACCIDENT HOT SPOTS AND THEIR CHARACTERISTICS BY USING GIS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

EZGİ KUNDAKCI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER SCIENCE
IN
GEODETIC AND GEOGRAPHIC INFORMATION TECHNOLOGIES

JANUARY 2014

Approval of the thesis:

IDENTIFICATION OF URBAN TRAFFIC ACCIDENT HOT SPOTS AND THEIR CHARACTERISTICS BY USING GIS

Submitted by **EZGİ KUNDAKCI** in partial fulfillment of the requirements for the degree of **Master of Science in Geodetic and Geographic Information Technologies Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen ____________
Dean, Graduate School of **Natural and Applied Sciences**

Assoc. Prof. Dr. Ahmet Coşar ____________
Head of Department, **Geodetic and Geographic Information Tech.**

Asst. Prof. Dr. Hediye Tüydeş Yaman ____________
Supervisor, **Geodetic and Geographic Information Tech. Dept.**, METU

Examining Committee Members:

Prof. Dr. Zuhal Akyürek ____________
Civil Engineering Dept., METU

Asst. Prof. Dr. Hediye Tüydeş Yaman ____________
Civil Engineering Dept., METU

Prof. Dr. Nebi Sümer ____________
Psychology Dept., METU

Prof. Dr. H. Şebnem Düzgün ____________
Mining Engineering Dept., METU

Assoc. Prof. Dr. Ela Babalık Sutcliffe ____________
City and Regional Planning Dept., METU

Date:30.01.2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: EZGİ KUNDAKCI

Signature:

ABSTRACT

IDENTIFICATION OF TRAFFIC ACCIDENT HOT SPOTS AND THEIR CHARACTERISTICS IN URBAN AREA BY USING GIS

Kundakcı, Ezgi

M.S., Department of Geodetic and Geographic Information Technologies
Supervisor: Assist. Prof. Dr. Hediye Tüydeş Yaman

January 2014, 135 pages

A major problem in the world and in Turkey is traffic accidents and consequent loss of life or property. In addition to the literature on highway black spots, it is necessary to identify urban locations with high traffic accident occurrence rates in order to shed light to underlying problems and develop preventive measures. Despite the more strict black spot definition for highways, accident prone locations in urban areas are regarded as hot spots, which can be simply defined as clusters representing dense accident occurrence areas. This thesis presents "Nearest Neighborhood Hierarchical (NNH) Clustering" method to find hot spots in urban region. Hot spot distribution is evaluated for different urban zones and based on accident type (pedestrian versus non-pedestrian) and occurrence time (morning peak, off-peak, evening peak and nighttime). Moreover, distances between hot spots and major intersections are examined to understand the relationship between them. Study also performs "Kernel Density Estimation (KDE) method", which is widely used in traffic safety analysis. Different levels of distance threshold are used to evaluate its impact on hot spot definition. Numerical results were obtained for Ankara using traffic accidents which occurred between 2008 and 2010.Since fatality accidents are rare events, it was possible to detect hot spots only when a threshold distance of 1km was assumed. Injury accidents create a significant number of hot spots with even 100 m threshold, which are located at or near major intersections. Pedestrian accident hot spots were condensed at or closer to Central Business District in Ankara, as expected from the literature.

Keywords: Traffic Accidents, Hot Spot Detection, Traffic Safety Analysis, Nearest Neighborhood Hierarchical (NNH) Clustering Method, Kernel Density Estimation, Geographic Information System (GIS).

ÖZ

KENTSEL ALANDA TRAFİK KAZA SICAK NOKTALARININ VE ÖZELLİKLERİNİN CBS İLE BELİRLENMESİ

Kundakcı, Ezgi

Y. Lisans, Jeodezi ve Coğrafi Bilgi Teknolojileri Bölümü
Tez Yöneticisi: Yard. Doç. Dr. Hediye Tüydeş Yaman

Ocak 2014, 135 sayfa

Dünyanın ve Türkiye'nin en önemli problemlerinden biri olan trafik kazaları, can veya mal kaybı ile sonuçlanmaktadır. Literatürde geçen karayollarındaki kara nokta tanımına ek olarak, kentsel alanlarda problemlerin belirlenmesi ve önlem alınmasına ışık tutması için trafik kazalarının yoğun olarak görüldüğü yerlerin tanımlanması gereklidir. Karayolları trafik güvenliğinde kullanılan kara nokta tanımına karşılık, kentsel bölgelerde kazaların yoğun olarak görüldüğü yerler sıcak nokta olarak adlandırılmaktadır. Bu tez çalışmasında, kentsel bölgelerde kaza sıcak noktalarının tespiti için "En Yakın Komşuluk Hiyerarşik Kümeleme" metodu sunulmaktadır. Sıcak nokta dağılımları, farklı kentsel bölge alanları, farklı kaza tipleri (yaya ve yaya olmayan) ve farklı kaza oluşum zamanlarına (sabah, öğle, akşam, gece) göre incelenmiştir. Ayrıca, kaza sıcak noktalarının ana kavşaklarla olan ilişkisini anlamak için birbirleri arasındaki mesafeler incelenmiştir. Çalışmada trafik kaza sıcak noktalarının analizinde sıkça kullanılan yöntemlerden biri olan "Çekirdek Yoğunluk Kestirim Yöntemi" de kullanılmıştır. Seçilen eşik mesafesinin sıcak not oluşumuna etkisi, farklı eşik mesafe değerleri kullanılarak incelenmiştir. Sayısal sonuçlar, Ankara 2008 ve 2010 yılları arasında meydana gelen trafik kazalarını kullanılarak elde edilmiştir. Ölümlü kazalar nadir olaylar olduğundan, 1km eşik mesafesinde sıcak nokta oluşturmak mümkün olmuştur. Yaralanma ile sonuçlanan kazalar, çoğunluğu kavşak ve ya kavşaklara yakın olmak üzere 100 metre eşik mesafesinde bile anlamlı sayıda sıcak nokta oluşturmaktadır. Yaya kaza sıcak noktaları literatürden de beklendiği üzere Ankara kent merkezi ve yakın çevresindeki alanlarda daha yoğun olarak gözlenmiştir.

Anahtar Kelimeler: Trafik Kazaları, Sıcak Nokta Tespiti, Trafik Güvenliği Analizi, En Yakın Komşuluk Hiyerarşik Kümeleme Metodu, Çekirdek Kestirim Yöntemi, Coğrafi Bilgi Sistemleri (CBS).

ACKNOWLEDGEMENTS

I would like to offer my special thanks to my supervisor Assist. Prof. Dr. Hediye Tüydeş Yaman for her support, encouragement and guidance throughout the development of this study and for great intelligence. Secondly, I would like to express many thanks to World Health Organization (WHO) Turkey office that kept the traffic safety problem in Turkey in the headlines diligently and provided the Ankara data used in this study.

I would like to show my greatest appreciation to Prof. Dr. Şebnem Düzgün, who enlightened me about the spatial analysis methods within her course. I've learned many things about the spatial data analysis from her lectures. I also would like to thank Prof. Dr. Nebi Sümer who gave me constructive comments and warm encouragement. Advice and comments given by my other committee members; Prof. Dr. Zuhal Akyürek and Assoc. Prof. Dr. Ela Babalık Sutcliffe, have been a great help in completing this thesis with their comments and suggestions.

I owe my deepest gratitude to my friends Barış Cav, Berk Yaylım and Ceren Ceylan, who listened and understood me every time in every situation throughout my Ankara life. Also, to my oldest friends Murat Mat, Erdem Marmara, Selin Tüfek and Hasan Höl, I want to express my deepest appreciation for being my best friends and to be exceptionally supporting in listening, understanding and calming me every time.

Finally, and most importantly, my special thanks to my all family for supporting and encouraging me during my study. Especially, my mother, my best friend in my life; Nalan Kundakcı, I could not achieve this thesis and most probably everything without her support, encourage and your endless love.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

CHAPTER 1

INTRODUCTION

1.1. Problem Definition

According to the recent reports of World Health Organization (WHO, 2013), one of the major problem in the world, as well as in Turkey, has been traffic accidents, which results in approximately 1.24 million of deaths and tens of millions injuries and disables every year. Moreover, 80% of all traffic accidents in the world have happened in developing countries. According to the WHO report of Turkey, nearly 10,000 people die and 200,000 people get injured as a result of the traffic accidents annually.

Associated with the growing technology and increasing population in Turkey, number of vehicles in traffic is increasing day by day. As a result, traffic exposure, and thus, the number of accidents, is increasing. However, the normalized rates of accidents per population or total number of vehicles still show very high fatality and injury rates in Turkey which signal a significant traffic safety problem that has to be addressed and analyzed to find contributing factors.

Many studies have been conducted for analyzing traffic accidents in Turkey as well as in the world. But, to develop effective countermeasures, accidents should be analyzed spatially in Geographic Information Systems (GIS) environment with their occurrence time, type, results and occurrence coordinates. Ultimately, identifying high probability accident areas is an important issue of traffic safety programs, because precautions and infrastructure assessments can be prioritized at these locations to make more efficient source allocations. Such areas are called “black spots” in highway safety literature and generally have some clearly defined methods to detect. For example, for black spot detection, intercity highways in Turkey are studied in 1-km unit segments in a “rate - quality method” that takes into account of mainly number and type of accidents, severity of accidents, type of roadway segment, traffic volume on the segment.

On the other hand, identification of such problematic areas in urban locations is more complicated, and it is loosely defined as “hot spot” in the urban traffic safety literature, mainly because traffic conditions, accident mechanisms and preventive measures differ greatly in urban traffic networks. Also, “hot spot” definition can vary based on the focus of the analysis. For one reason, it is not easy to define a unit segment in a complex urban network; a road segment, such as intersection, can be easily affected by the nearby segments or the built-environment components (pedestrian zones, shopping areas, bus stops, etc.). Secondly, any solution or improvement option may require designs and precautions beyond the changes in the road segment, alone. Based on the scope of a hot spot detection analysis, criteria for detection can be changed; minimum number of accidents (n_{min}), accident type, occurrence time, occurrence area, etc. Accidents occurring spatially close may be due to higher levels of traffic volumes, conflicting movement (e.g., at intersections), or inappropriate infrastructure designs (e.g., lack of pedestrian crossings).

After detection, it is also important to analyze the distribution of hot spots in the cities, which not always distributed uniformly over space or time. Urban built environment and transportation have always been in an active interaction that has both spatial and temporal dimensions, directly. Land use patterns affect the travel behavior in both trip generation and trip distribution, such that more trips are generated and distributed in some regions than the others, such as Central Business Districts (CBDs). Traffic volume is one of the basic parameters in accident analysis, since there is a strong relationship between the volume and number of accidents. Similarly, the more pedestrian activity in the region is, the higher the portion of pedestrian is expected in that area. For this reason, hot spots should be evaluated together with traffic volume or urban built environment characteristics and in time. In the absence of traffic volume data for the cities, other exposure measures such as geographical area and road network length can be used. Analyzing accident hot spots together with intersections is important to draw a conclusion about which intersection has a priority in terms of infrastructural regulations. Since most accidents occur on or near to the intersection, relationship between them should be analyzed and evaluated.

1.2. Objective of the thesis

The main objective of this thesis is to develop a framework that would lead to detection and understanding of the distribution of traffic accident hot spots in urban regions using GIS coupled with spatial statistical methods. The framework includes basically the following steps;

- Identify hot spots using Nearest Neighborhood Hierarchical (NNH) clustering method
- Severity mapping of hot spots
- Distribution of hot spots over urban zones
- Distribution of hot spots over different time periods (peaks and off-peaks)
- Distance analysis to the intersections
- Comparison of NNH with Kernel Density Estimation (KDE)

The first step is use of the NNH clustering to identify hot spots. While finding hot spots, accidents can be used with/without groupings by type/time. To detect hot spots, NNH clustering method is used with different criteria for fatality and injury accidents due to the great differences between them in terms of number of accidents. Hot spot severity is defined as the number of accidents found within the cluster, which is also taken into consideration in the analysis of the hot spot distances to the nearest intersection. The lower number of fatality accidents do not create many hot spots; consequently it is the injury accident hot spots that provided further distribution or distance analyses. This study also offers visualization technique to show hot spots with their severity values. Moreover, NNH clustering results are compared with KDE that is the most widely used method in accident analysis studies. Finally, the relation between hot spots and built environment are checked with statistical analysis tools to find whether analysis results are statistically significant or not.

1.3. Organization of the thesis

This thesis is outlined as follows; to provide a background, traffic safety analysis concepts and methods are presented in Chapter 2. Accident analysis with GIS, pedestrian safety analyses and NNH clustering algorithm and CrimeStat are discussed in the subsections of this chapter. The proposed methodology for the detection of accident hot spot in the urban area is presented in Chapter 3, which includes specific discussion of the issues of spatial and temporal distribution of hot spots and their relations with urban built environment. Chapter 4 presents the traffic safety data and characteristics of the case study, which is City of Ankara with 3-year fatality and injury accident data from 2008-2010. GIS data preparation is also included in this chapter. Hot spots detected in Ankara urban region as a case study are presented in Chapter 5, with a special section dedicated to the comparison of NNH clustering algorithm with KDE. Chapter 6 includes evaluation of distribution of identified hot spots over time and space. Furthermore, distance analyses of detected hot spots are explained in details with their statistical tests, as well. Final chapter is a discussion and conclusion part of this thesis. Detecting accident hot spot with different two methods, comparison of them, statistical analysis of distance analysis and contribution of this study are discussed, and future studies based on this work are recommended.

CHAPTER 2

LITERATURE REVIEW

In the following sections, traffic safety analysis methods, pedestrian safety analysis methods and traffic accident hot spot analysis methods used for traffic safety research are explained briefly. After the representation of traffic safety studies, NNH clustering method which is used for this study is explained in details. Furthermore, Kernel Density Estimation method, which most widely used method in traffic safety studied is represented. Also, comparison technique of different spatial analysis methods is explained. Finally, statistical analyses method used in this thesis are explained.

2.1. Traffic Safety Analysis Methods

Traffic safety analyses of the highways and urban roads show differences, due to the differences in infrastructural design and traffic flow characteristic (e.g. speed); thus, highway safety and urban traffic safety should be studied in a different manner. There are multiple factors that affect the traffic safety in highways as driver's behavior, traffic conditions and environmental factors such as weather, time of the day, road lighting, etc. Highways are less complex networks when compared to the urban roads, where traffic flow and speed limit are different from the urban ones. In most studies, traffic safety analysis on intercity highways and urban freeways are focused on mainly vehicle-to-vehicle crashes.

However, in urban traffic flow, pedestrian and cyclists have a risk of involvement in the traffic accidents. On the other hand, urban road networks are more complex, and there are more vulnerable road users. For these reasons, the traffic safety analyses used in highways and urban networks are different from each other.

In highway traffic safety, “black spots” are defined as unit road segments that show similar and dense accident rates. In Turkey, the General Directorate of Highways (GDH) uses the rate-quality control method in the evaluation of highway traffic safety. This statistical method analyzes the highways by dividing it into 1 km of segments. For each segment, three parameters are calculated; accident rate, accident frequency, severity index; and each of these parameters are compared with the different critical values. When all of these three values of a segment exceed the critical values, the segment is considered as a black spot. (Sjolinder et al., 2001)

However, accident prone locations in urban regions are regarded as “hot spots”, which can be defined differently based on the focus of the analysis. One reason is that it is not easy to define a unit segment in a complex urban network; a road segment, such as intersection, can be easily affected by the nearby segments or the built-environment components (pedestrian zones, shopping areas, bus stops, etc.). Also, any solution or improvement option may require designs and precautions beyond the changes in the road segment, alone. Based on the scope of a hot spot identification analysis, criteria for the identification can be changed; minimum number of accidents, accident type, occurrence time, occurrence area, etc.

2.1.1. GIS in Traffic Safety Analysis

Geographic or spatial data usage plays an important role on many fields of daily life. Many data used in life directly or indirectly have spatial information. Usage of spatial data required selection and manipulation of them correctly. In this step, GIS has an important role. GIS is a tool which provides collection, storage, manipulation, query, analyzing and visualization of the spatial data (Lloyd, 2010).

Geographic location gives information about accidents more than their coordinates. Occurrence of accidents at the same place or nearby places could be the indicator of causes of accidents. For this reason, traffic accidents should be analyzed with their coordinates. With the help of spatial analysis tools, high crash occurrence areas could be clearly identified. In order to understand causes of accidents and improve traffic safety, hot spot analyses have an important role. In the simplest term, hot spots can be explained as the highly incident occurrence areas.

One of the most important problems that traffic officials face is where and how to implement precautionary measures and provisions so that they can have the most significant impact for traffic safety. GIS is a very important and comprehensive management tool for traffic safety. Since 1990's, GIS technologies have been used more frequently for such studies due to the availability of low cost GIS with user-friendly interfaces. In the traffic safety analysis, GIS can visualize the locations of accidents and store the attributes of accidents like time of the accident, number of injuries, number of fatalities, characteristic of the roads that accidents occur, land-use characteristic, etc. GIS is explanatory in order to find the reason behind the occurrence of accidents since it relates the locations with many attributes.

Recently, many studies related with traffic safety analysis have been performed with the help of GIS. The type of analysis, applicable for accident analysis, include intersection analysis, segment analysis, cluster analysis, density analysis, pattern analysis, proximity analysis, spatial query analysis and spatial accident analysis modeling techniques. Kim and Levine (1996) described the traffic safety GIS prototype which performed spatial analysis of traffic accidents that is developed for Honolulu, Hawaii. Different type of spatial analysis methods based on point, segments and zones analyses had been developed. Affum and Taylor (1998) introduced a Safety Evaluation Method for a Local Area Traffic Management (LATM), which was also a GIS-based program for analyzing accident patterns over time. Hirasawa and Asano (2003) proposed a traffic accident analysis system based on GIS, which was linked with digital maps showing accident locations, and also provided information on road structure and accessory facilities, and weather information.

In order to identify hot spot (accident or crime related with roads, for example, vehicle thefts) on roads, network based spatial analysis can be used. Software package Spatial Analysis on a NETwork (SANET), can be used within the ArcGIS software. Network KDE, which is available in Sanet, can be used for identifying traffic accident hot spot segments on a network. Different from KDE measure the bandwidth based on Euclidian distance, network KDE provides measurement of bandwidth based on a network distances. (Okabe et al., 2006)

2.1.2. Spatial Analysis Methods for Traffic Safety

GIS tools have the ability to do spatial analysis; but there are many other software packages performing spatial statistical analysis. In some situations, researcher needing more than a GIS program, couple different analyses to GIS fully or loosely. Loose coupling means exporting data from the GIS package to use in a statistical spatial analysis framework. Close coupling involves calling a spatial analysis routine from within GIS (Bailey and Gatrell, 1995).

Levine et al. (1995a) analyzed spatial patterns of motor vehicle crashes assigned each of them to the nearest intersection in Honolulu, Hawaii. Study mainly focused on spatial variation of accidents with respect to the hour of the day, weekday or weekend, accident type, etc. In addition to this, population, employment and built environment characteristics were examined. Results showed that as employment density increased, so would the accident concentration be. In the suburban and rural areas, there were more alcohol related crashes occurred especially in night time. In another study by Levine et al. (1995b), the spatial variations of accident were studied as aggregated into small geographic areas. Study focused on characteristics of neighborhoods and areas, not just on the road system. They developed zonal relationship of accidents to population, employment and road characteristic. As a result, they generated expected accident maps which identify areas had higher than expected levels in terms of accident occurrence.

More recently, Xie and Yan (2008) used and compared planar and network KDE in terms of lixel size (linear pixel, same as raster cell but it is in a network), bandwidth and density visualization while identifying accident hot spots. As a result of whole study, both methods show that major corridors in the study area have more density values and major intersections can be identified as hot spot. However, results also show that network KDE is more convenient than planar KDE in terms of calculating density estimation for traffic accidents since planar KDE estimate density beyond accident data context.

Parasannakumar et al. (2011) studied clustering analysis with respect to accident type and occurrence time, using Moran's I Index to perform spatial autocorrelation. Clustering analyses were performed with KDE and Getis-Ord GI* statistics. As a result of the analyses, accident hot and cold spots were determined. Moreira et al.(2012) studied three different methods to identify hazardous road location (HRL) in the City of Vila Real, Portugese: NNH clustering algorithm, KDE and Point Density. As a result of the study, high speed was found as an indicator factor of single vehicle crashes (run-off crash). Eckley and Curtin (2012) searched clustering of traffic accidents on a network using Sanet software. Spatio-temporal clustering was performed with Knox method.

In Turkey, Erdogan et al. (2008) studied accidents hot spots and detected safety deficient areas on the highways in the city of Afyon, Turkey. KDE and repeatability analysis were conducted to explore accident hot spots. As a result of both analyses, almost same locations were founded as hot spots where most of them located junction points, cross roads and access roads to the villages and towns.. They also showed that accident density increase in summer and winter especially in August and December. Also weekends have higher frequencies. Moreover, they found that fatality accidents mostly occurred in midnight. Keskin et al. (2011) studied accident clusters within the boundary of Middle East Technical University Campus. Traffic accident for different seasons, days and time periods were determined. KDE, K-means clustering and NNH clustering were used to show accident hot spots. Hot spot zones were identified according to seasons, day and time periods in terms of occurrence. Results clearly showed that, accidents were mostly seen between 12 and 19 in terms of time period. Also, Monday have the highest frequency. Most recently, Kaygısız et al. (2012) studied the spatio-temporal analyses of traffic accidents occurred on the motorways in Turkey. Due to the data quality, only South Anatolian Motorway was analyzed in detail numerically. Both KDE and Network KDE methods were used to identify hot spots, which were detected mostly in the summer time. In the study area, most commonly seen accident types were rear-end, colliding with stationary object and run-off.

2.1.3. Traffic Safety and Built Environment

There is a strong relationship between the occurrence of accident and built environment. Other than human factors, road characteristics and environmental factors such as weather and time of the day, urban land use characteristics have an influence on the occurrence of accidents. There are certain land uses which generate traffic as employment area, commercial area, cultural activity area, etc. For this reason, built environment characteristics should be taken into consideration while performing traffic safety analysis.

Kim and Yamashita (2002) focused on the relationship between land use and vehicle accidents in Hawaii, simply by creating accident frequency maps, which showed that traffic accidents were highly clustered in the areas where people lived and worked. Furthermore, study calculated the number of accidents and number of injury accident depending on the per unit area taken into consideration the land use. Results showed that commercial and visitor lodging land uses produce the highest number of accidents in terms of per area unit. At the same time, these two land uses also generated the highest frequency of injury accidents. Later, Kim et al. (2006) studied the relationship between land use, population, employment and motor vehicle crashes in Honolulu, among which statistically significant relationship were identified. Negative binomial model was used to investigate the absolute and relative effects of these factors on the number of pedestrian, vehicle-to-vehicle, bicycle and total accidents. According to results, pedestrian accidents were mostly occurred in areas with high levels of commercial activities and parks. Overall accidents and vehicle-to-vehicle accidents were related with commercial areas. Moreover, areas around schools were associated with higher frequencies of overall accidents.

Selvi (2009) proposed traffic accident prediction model based on fuzzy logic algorithm. As a case study, he studied the city of İzmir metropolitan area. After the detection of traffic accident hot spot areas by using KDE, he searched the causing or affecting factors of the traffic accidents. In the prediction model, he used traffic variables (traffic flow, traffic density, average speed and average gap between vehicles), geometric variables (road width, number of lanes, number of minor accesses and percent of medians) and environmental variables (number of bus stops and weather conditions). As a result, he classified streets in relation to the time zones as high safety level, low safety level, low risk level and high risk level.

More recently, Kaygısız (2012) studied the relationship between land use and traffic accident occurrence in Eskişehir, Turkey via prediction models constructed by using traffic accident data and land use data. Two databases were created related to junctions and road segments by using GIS and he developed prediction model for both junctions and road segments. According to the results, more accidents were occurred at the four way intersections compared to the other type of intersections. Also, percentage of building is found to increase the number of accidents. Relationship between occurrence of accidents and land use type is found statistically significant. According to the model result, most effective land use type is found as mixed use whereas less effective is found as non-urban areas.

2.2. Pedestrian Safety Analysis

Pedestrian safety is another important issue of traffic safety officials. Since pedestrians and cyclists are the most vulnerable and defenseless road users, specific attention must be carried on by professions in terms of traffic safety precautions. GIS and spatial analysis methods that were explained previous sections are also used in pedestrian safety analyses studies. In recent years, there has been an increasing amount of literature on pedestrian safety analysis performed with spatial statistics and GIS.

2.2.1. GIS in Pedestrian Safety

GIS based identification of high risk areas for pedestrian accidents was performed by several studies (Lightstone et al. (2001); Krishnakumar et al. (2005); PBCAT (2014)). High Pedestrian Crash Locator (HPCL), a customized tool used in ArcGIS, computed the crash concentrations and crash rates of the selected high crash zones and rank the zones sing crash score method (Krishnakumar et al., 2005). Pedestrian and Bicycle Crash Analysis Tool (PBCAT) is a software application designed to assist State and local pedestrian and bicycle coordinators, planners, and engineers in addressing pedestrian and bicyclist crash problems. PBCAT accomplishes this goal by enabling users to develop a database of details associated with crashes between motor vehicles and pedestrians or bicyclists. One of these details is *crash type*, which describes the pre-crash actions of the involved parties. After developing a database of crash information, PBCAT users can analyze the data, produce reports, and select countermeasures to address the problems identified by the software (Web 1).

2.2.2. Spatial Analysis Methods for Pedestrian Safety

Different spatial analysis methods have been used in recent years for pedestrian safety studies, as well. Pulugurtha et al. (2007) studied the spatial patterns of pedestrian accidents with the help of Simple and Kernel Methods in order to identify high pedestrian accident zones. Later, Pulugurtha and Vanapalli (2008) studied also hazardous bus stop identification using KDE to create crash hot spot maps. Different from the former study, traffic volumes, bus stop coverage, transit ridership data and street centerline coverage were used in these ranking hazardous bus stops.

Troung and Somenahalli (2011) used Getis-Ord GI* to identify pedestrian hotspot maps while; Ha and Thill (2011) used KDE to create hot spot maps. Besides KDE, Blazquez and Celis (2012) Moran's I Index test was used to identify a positive spatial autocorrelation on accident contributing factors, times of day, straight road sections and intersections. More recently, Dai (2012) evaluated pedestrian safety on roadway segments. Highly pedestrian accident occurrence areas were detected with the help of the Bernoulli model available in SatScan statistical program.

2.2.3. Pedestrian Safety and Built Environment

Wedagama et al. (2006) specified urban zones, and analysis of accidents within a zone showed that pedestrian accidents highly occur in the city centre, especially in retail and community areas during working hours. Increase in retail land-use like clubs and bars also resulted with increasing in pedestrian casualties as well as cyclist causalities especially in nighttime hours More on the effect of built environment, a study showed that density of pedestrian accidents was found higher in commercial and residential areas, whereas it was quite low in the industrial areas (Loukaitou-Sideris et al. (2007)) while another showed that pedestrian connectivity and transit access have positive effects on injury accidents (Clifton et al. (2009)).

The relationship between the accident occurrence and severity of injuries and the built environment characteristics around school showed that the presence of recreational facilities on the school site increased the accident occurrence and injury severity (Clifton and Kreamer-Fults (2007)). Pedestrian accidents around bus stops (Troung and Somenahalli (2011); Pulugurtha and Vanapalli (2008)) were analyzed in more details. The relationship between pedestrian accidents and age has been widely investigated (Blazquez and Celis (2012); Dai (2012); Ha and Thill (2011)). Blazquez and Celis (2012) shows that there is no dependence found on accidents due to age, gender, weekday and month of the year; however positive spatial autocorrelation was found between child accidents and roads without traffic signs. On the other hand, study conducted by Dai (2012) showed that age, walking under influence, crossing the street not a cross walk and darting into traffic are contributors of pedestrian crashes. Another study performed by Ha and Thill showed the downtown and entertainment areas were found to be high locations for adult and intersection crashes.

2.3. Traffic Accident Hot Spot Analysis

Crashes occurring spatially close together may be the products of locations with higher levels of conflicting volumes, such as intersections or inappropriate infrastructure designs, such as lack of pedestrian crossings, etc. Identifying high probability crash areas is an important issue of traffic safety programs. Despite a more strict "black spot" definition in highway safety, as mentioned in the first section of this chapter, accident prone locations are regarded as "hot spots", which can be defined differently based on the focus of the analysis.

A considerable amount of studies is available on the detection of traffic accident hot spots which have been widely discussed in the literature. There are many accident hot spot analysis techniques used such as KDE, Network KDE, NNH Clustering Algorithm, Black Spot Analysis, Moran's I Index and Getis-Ord GI* statistics. There are distinctions between these methods; a method uses the actual location of an accident (point in GIS) and one that uses the number or density of accidents for a small geographical area such as zone or grid cell. Moran's I, Getis-Ord GI* statistics and KDE all use zone data, on the other hand, NNH and Black Spot methods use point data. The methodology of spatial analysis can change according to the aim of the research and expected results.

For this study, NNH clustering algorithm is used which is available in CrimeStat software. On the other hand, one of the most widely used methods in traffic safety, KDE method is also performed. Two types of KDE, normal and uniform available in CrimeStat, are used. To show the differences between these two methods; NNH and KDE, comparison will be done between them with the help of Prediction Accuracy Index (PAI). In the following sub-sections, these hot spot detection methods, CrimeStat software and PAI are explained in details.

2.3.1. CrimeStat Software and Nearest Neighborhood Hierarchical (NNH) Clustering

There are several software packages which extract information and perform analysis from the data based on spatial location of incidents. Crimestat is a free downloadable statistical package that mainly analyzes the point data. CrimeStat is a Windows-based spatial statistical program for the analysis of crime incident locations. It is developed by Ned Levine & Associates, with funding by the National Institute of Justice, an agency of the United States Department of Justice. It is designed to be coupled with GIS. Program performs many spatial analyses on point data (crime location, accident location etc.), zone data (blocks, traffic analysis zones, neighborhood/county/city districts etc.) or line (e.g. road segments). Program provides many analyses under the sections of spatial description (e.g. spatial distribution, distance analysis, hot spot analysis), spatial modeling (e.g. Interpolation, Space-Time Analysis) and Crime Travel Demand (Levine, 2010).

Program offers several statistical methods to detect hot spot locations such as; mode, fuzzy mode, NNH clustering, risk adjusted NNH clustering, K-means clustering, Moran statistic, Spatial and Temporal Analysis of crime (STAC) and KDE (Levine, 2010). Crimestat is mostly used in crime analysis as the name suggests, however, Moreira et al. (2012) used Crimestat to identify hazardous road locations with the tools of KDE, mode, fuzzy mode and NNH clustering methods. Also, Levine (2006 and 2009), Kaufman (2008) and Keskin et al.

(2011), also used Crimestat and NNH clustering method to detect traffic accident hot spot locations.

As mentioned in the literature, there are many hot spots techniques in which the main idea is “computing the number of cases within a limited area determined by the researcher”. In this study, NNH clustering analysis method is used to detect accident hot spots, which is focused on the identification of groups of data that are spatially close. In NNH clustering, the Euclidean distance between every data point pair is checked and used as a criterion for clustering. If a threshold distance (d) is selected, the point pairs with smaller distances are clustered together (see Figure 2.1). If desired, a second criterion can be specified as the minimum number of points (n_{min}) to be in a cluster; then, data points satisfying both distance and n_{min} criteria would be labeled a “cluster”, as defined in this study. If defined numbers of observations are nearer than the threshold value, a new cluster is generated. After the calculation of first order clusters, the second and high order clusters are formed with the same manner until only one cluster is left or the threshold criteria fails. For this reason, NNH clustering algorithm does not generate cluster from all observations in the study area. Since it depends on the criteria with regards to distance and n_{min}, only points satisfy the criteria are clustered.

NNH Clustering in this study is obtained using CrimeStat software that outputs clusters in two formats; convex hull and ellipse (see Figure 2.1). The user can visualize the clusters in these two different formats. There are both advantages and disadvantages of each format. Convex hull is a polygon that covers exactly all the clustered points. For a detailed analysis (e.g. neighborhood level analysis) convex hulls are preferable to the ellipses since convex hull is more precise and define the actual area where the hot spot occurs. On the other hand, ellipse is more symbolic representation of the cluster, which generally looks better on a map and is easily understood by users. The biggest disadvantage to an ellipse is that it forces a certain shape on the data, whether there are incidents in every part of it or not. There are “false positives” with convex hulls than with ellipses as well as fewer “false negatives”. As can be seen from the Figure 2.1, the ellipses can stick out beyond the actual locations but also can cut off part of the hot spot area; they are a mathematical abstraction (Levine, 2010).

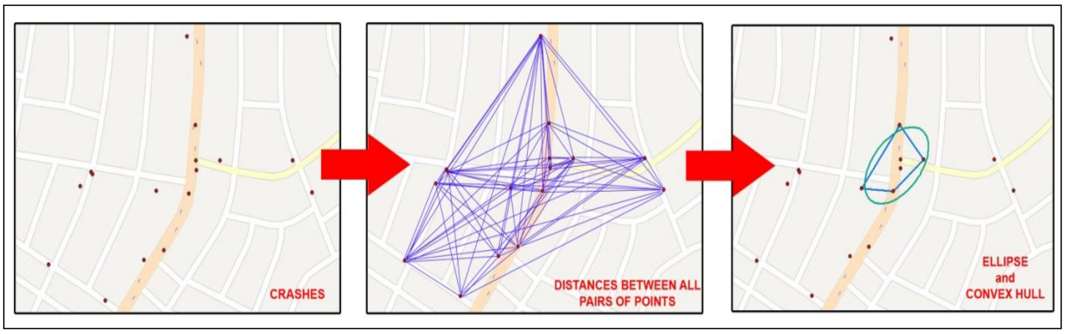

Figure 2.1: NNH clustering algorithm

2.3.2. Kernel Density Estimation

KDE is an interpolation technique that used in the identification of hot spots. Method is based on a point density function, which is calculated for all cells of a grid area (study area). KDE calculation in CrimeStat is needed the following steps;

- Study area extension should be generated (based on a lower left and upper right coordinates),
- Spatial resolution of grid cell is defined (either by cell spacing or number of columns),
- Centred on each incident location point, symmetrical surface is placed. The surface of density function changes according to the defined bandwidth value,
- For each cell on the defined grid area (study area), method sums the value of all surfaces for that reference cell (Levine, 2010).

This procedure is repeated for all reference locations. In Figure 2.2, schematic summary of KDE calculation procedure can be seen.

In Crimestat, there are five different mathematical function is available for calculating KDE; normal, uniform, quartic, triangular and negative exponential. All these five different KDE methods have different mathematical formula. In this study, normal and uniform KDE method are used, which are explained mathematically in details below (Levine, 2010).

Normal Interpolation Method

Normal distribution function has the following functional form (Eq. 2.1):

$$g(x_j) = \sum \left\{ [W_i * I_j] * \frac{1}{h^2 * 2\pi} * e^{-[\frac{d_{ij}^2}{2*h^2}]} \right\}, \qquad \text{Eq (2.1)}$$

where d_{ij} is the distance between an incident location and any reference point in the region, h is the standard deviation of the normal distribution (the bandwidth), W_i is the weight at point location and I_i is an intensity at a point location. This function extends to infinity in all directions and, thus, will be applied to any location in the region (Levine, 2010).

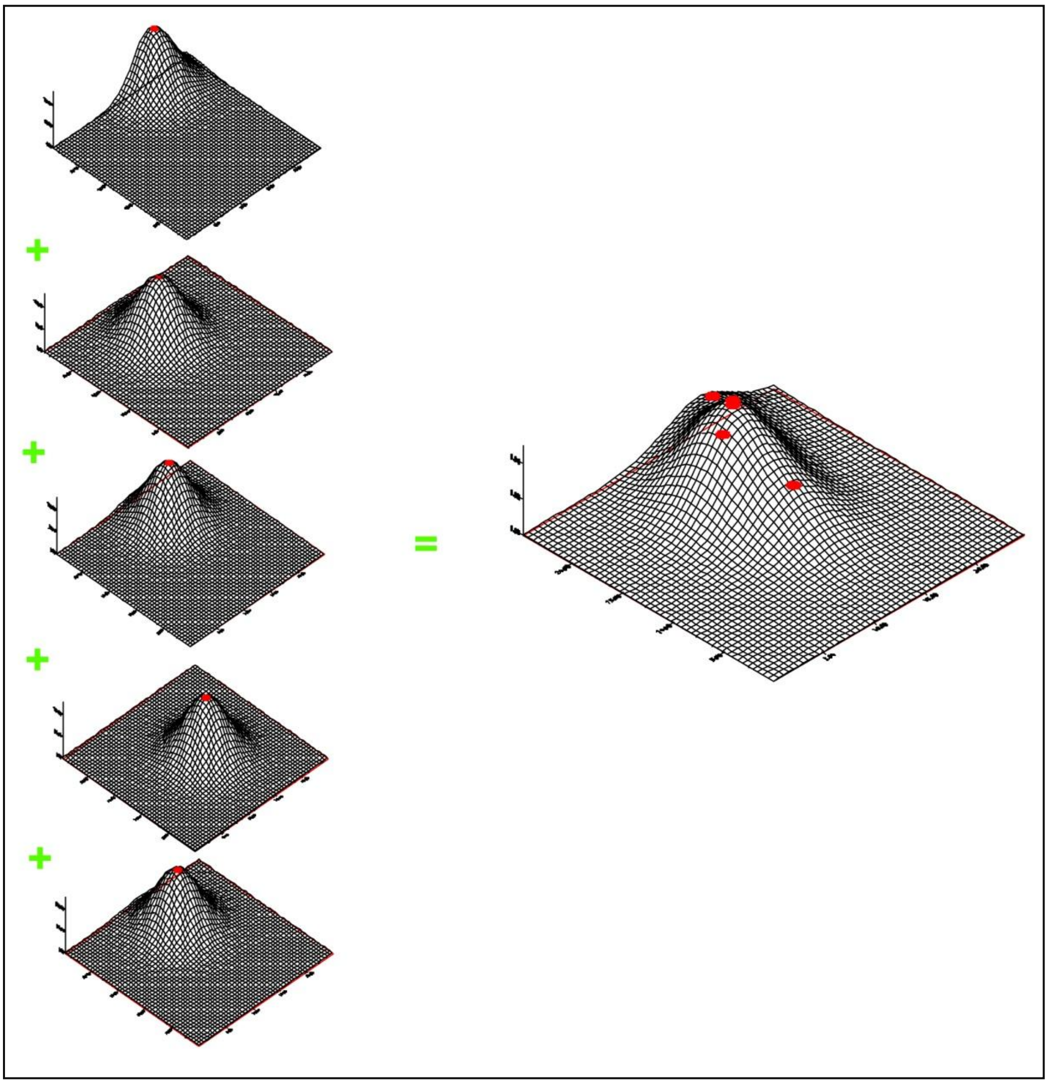

Figure 2.2: Summing of normal kernel surfaces for 5 points (Levine,2010)

Uniform Interpolation Method

Uniform distribution weights all points within the circle equally. Its functional form is:

Outside the specified radius, h: $g(x_j)=0$

Within the specified radius, h: $g(x_j)=\sum K$

where K is a constant. Initially, K is set to 0.01 but when re-scaled to ensure that either the densities or probabilities sum to their appropriate values (i.e., N for densities and 1.00 for probabilities) (Levine, 2010).

As an output, Crimestat offers three options; absolute densities, relative densities and probabilities. Absolute densities give density estimates using formulas depending on what type of method is used. Relative densities give normalized values that absolute densities are divided by the area of the grid cell. Probabilities give the density at any one cell which is divided by the total number of points in study area (Levine, 2010).

2.3.3. Comparison of Hot Spot Techniques

As explained in the literature, there are many hot spot techniques to identify high crash occurrence areas. However, problem is arisen when deciding of which technique should be used as a hot spot. Different techniques produce different hot spots in terms of shape, size and location. For this reason, hot spot techniques' results must be compared with each other.

Prediction Accuracy Index (PAI)

Different hot spot techniques create different results in terms of size, shape and location. However, there is no exact knowledge or definition about which hot spot technique best for detecting high accident occurrence areas. In light of these, PAI was developed by Chainey et.al. (2008) to determine if there are differences in terms of capturing or predicting hot spots. This method has been used in crime hot spot mapping for both comparing the different techniques and prediction ability. Methodology depends on the ratio of the percentage of points found in hot spots over all points in the entire study area to the area percentage of the identified hot spots over all study area. Percentage of points in hot spots over all points in the study area is called as Hit Rate. In short, Pattern et al. (2009) described the PAI as "PAI is the ratio of the hit rate percentage to the area percentage for the identified hot spot" (p.14). Mathematical functions of hit rate (Eq. 2.2) and PAI is given below (Eq. 2.3):

$$HR=\left(\frac{n}{N}\right)*100 , \qquad \text{Eq. (2.2)}$$

Where n is the total number of accidents that found in clusters and N is the total number of accidents in the study area.

$$PAI=\frac{\left(\frac{n}{N}\right)*100}{\left(\frac{a}{A}\right)*100} , \qquad \text{Eq. (2.3)}$$

Where n is the number of accidents in the hot spot and N is the total number of accidents in the study region, a is the area of hot spot and A is the area of the study region. A measured PAI is based on the historical data and is utility metric by which that particular hot spot technique can be compared to other.

There are some studies performed related with the comparison of different hot spot techniques. Prediction Accuracy Index (PAI) developed by Chainey et al. (2008), is the method to compare different hot spot techniques and measure the accuracy of them. They compare different methods by crime types. In their research, as a result of comparing different hot spot techniques in terms of prediction of crime and measurement of techniques, KDE hot spot technique gave best result. On the other hand, commentary about their research, written by Levine (2008), shows that convex hulls output of NNH clustering algorithm gave better results than KDE method. Studies, conducted by Pattern et al. (2009) and Hart et al. (2012) also show that NNH clustering algorithm give better PAI results than KDE and other methods in terms of both prediction and measurement.

2.4. Statistical Significance Analysis

To draw conclusion on the significance of analysis results, some statistical tests should be performed in order to understand, if and how strong a relationship existed. Testing for significance is comprised of; stating research hypothesis and null hypothesis, selecting confidence interval, computing the test for statistical significance and interpreting the results. Research hypothesis states the expected relationship between variables. On the other hand, null hypothesis, which mostly claims that there was no relationship between the variables, is called as “zero” hypothesis (H_0).

There are several analysis methods available to measure the significancy of results. According to the data characteristics, parametric or non-parametric tests can be performed. Chi-Square (χ^2), a well known non-parametric test, is used to measure differences between observed and expected values between two or more groups. It measures the number of objects found in each group differ significantly from the number of observed that would be expected. In order to perform Chi-Square test, expected number of objects is calculated. Then, difference between observed and expected values is calculated. Ratio of expected value to the square of differences gives the χ^2 value which is compared with χ^2 getting from table of χ^2 with n-1 degrees of freedom (n is the number of group). If the calculated χ^2 value exceeds the original χ^2 value read from table, it can be concluded that there are significant difference between observed and expected values (IBM, 2011).

The One-Way “Analysis of Variances (ANOVA)” is a parametric statistical technique producing one-way analysis of variance for a dependent variable by a single factor (independent), and it depends on the F ratio. It is mainly used to compare the significant differences between two or more groups of the selected dependent. Against the null hypothesis that all population means are equal, one can claim that at least mean of one group is different. If significantly different, those groups can be determined via the help of post-hoc range tests and pair wise multiple comparisons. In the SPSS software, a commonly used statistical tool, the range tests identify the homogenous subsets not different from each other, and significantly differently groups at an alpha level 0.05 are indicate significantly different group means (IBM, 2011).

CHAPTER 3

METHODOLOGY FOR URBAN HOT SPOT ANALYSIS

In urban accident hot spot analysis, for the identification of hot spots, there is no strict definition is available in the literature, similar to that of “black spot” for highways. Mainly because urban networks are complicated and complex with many types of transportation modes; there are many origins and destinations. As the level of traffic volumes is not spatially random over urban regions, the traffic accidents are not expected to be spatially random either. Furthermore, existence of pedestrian movements creates an additional complexity that is also very closely related to the built environment. Hence, it is not easy to talk about unit segments and normalization while searching accident hot spots over urban regions.

3.1. Framework for Detecting Traffic Accident Hot Spots in Urban Regions

The following methodology has been developed to detect traffic accident hot spots in the urban regions (Figure 3.1) via NNH clustering algorithm:

- Studying the traffic safety in urban regions is started with data preparation. Accident data can be studied with or without groupings. In other words, while performing traffic safety analysis, all accidents can be taken as only point events or accidents can be grouped by type or time. Mostly, studying with grouped data is more appropriate to evaluate the characteristics of accidents. Since it is well known that accidents and built environment parameters are in reaction with each other, it is important to relate them to find causality of accidents. For this study, detected accident hot spots are related with both identified urban zones via distribution analysis and identified major intersections via distance analysis.

- Hot spot detection process can be performed with “all accident data” or grouped data “by type” or “by time”.
- Detected urban hot spots can be further visualized thematically with their severity, which is an output of a NNH clustering method in addition to the boundary of the hot spot.
- Since KDE is one of the most widely used methods in the detection of hot spot studies, results of NNH clustering algorithm should be compared with it. This comparison can be performed with both visual and mathematical aspects. In order to compare two methods mathematically, PAI, which is explained briefly in the literature review, should be used.
- Detected hot spots could be related with urban built environment components. While studying the relationship between hot spots and built environment, severity of hot spots can be considered as dependent variable. Distribution of accident hot spots should be studied taken into consideration the distribution over time and space.
- After the completion of distribution and distance analyses, they must be concluded with the statistical test to measure significance. In order to checking the significance of the analysis results, parametric or non-parametric tests can be used. While testing the significance of distribution of hot spots over space or time, Chi-Square, non-parametric test, can be used. On the other hand, distance analysis results’ significance can be checked via parametric test, namely One – Way ANOVA. After the statistical test results, evaluation could be performed.

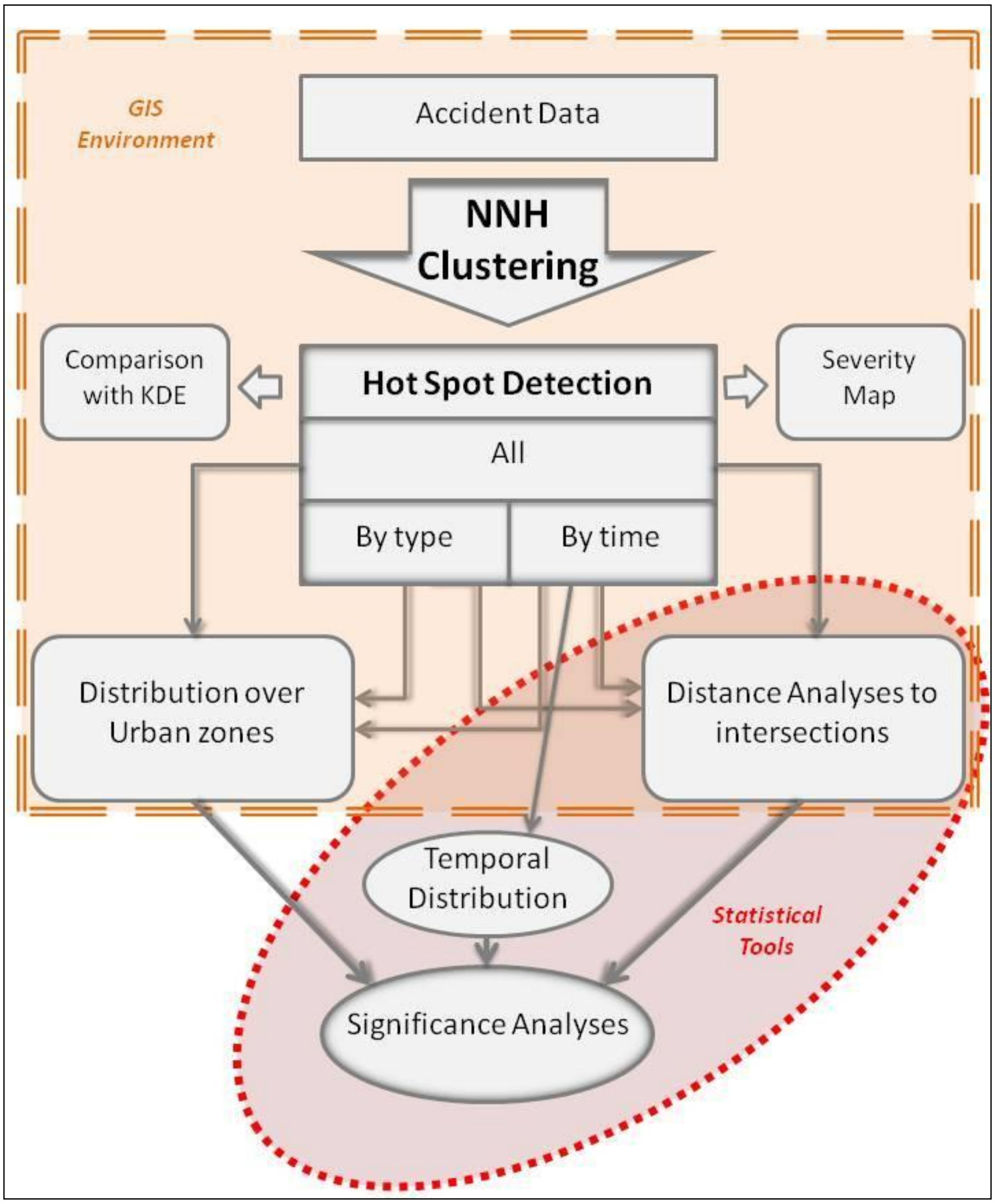

Figure 3.1: Framework of the proposed methodology for the detection of the traffic accident hot spots in urban regions

3.2. Hot Spot Analysis Using NNH

In CrimeStat, it is possible to define the threshold values randomly, which means defined by the software itself, or they can be predefined by the user. Random distance is generated by the size of the study area and number of observations in it; this may generate big clusters especially in big study areas. Instead, it is suggested to try different predefined distance thresholds to get the more robust results.

Based on the scope of the analysis, criteria can be changed. Since every city has their own network characteristics, the threshold distance, and n_{min} could be changed. For instance, if the aim is to find most priority locations where more than 50 accidents occur, researcher could specify it as n_{min}=50. On the other hand, since NNH results are given with their severity value, which is the number of accidents located in the boundary of a cluster, user should visualize clusters with regards to severity values to identify most important areas.

3.3. Hot Spot Analysis by Accident Type

Fatality accidents may be rare, which would not be analyzed with the same NNH criteria. For this reason, detection of fatality accident hot spots in urban regions, larger threshold distances with the smaller number of cluster severity would be needed. On the other hand, injury traffic accidents are more common compared to the fatality accidents in urban regions. Detection of injury accident hot spots in urban regions can be studied with different NNH criteria in terms of the threshold distance and n_{min} required in the cluster.

In urban regions, due to the existence of a significant number of vulnerable road users that include pedestrians, and their impact on the severity of the accidents, it is meaningful to study pedestrian and non-pedestrian accidents separately. In general, the density of the pedestrian accident would be high where the pedestrian activity is high. Mobility of pedestrian is high due to the built environment characteristics, and it is directly related with the occurrence of pedestrian accidents. On the other hand, occurrence of non-pedestrian accident is generally due to high speed, traffic volumes and built environment as conflict points. Speed limits of urban arterials affect both occurrences of both pedestrian and non-pedestrian accidents.

Besides the type of accidents, time of day is also related with the occurrence of accidents. Traffic volumes and speed profile are directly related with peak and off-peak hours of the day. During peak hours, a higher traffic volume that means higher exposure is expected in urban regions. On the other hand, less traffic volume is seen mostly in off-peak hours which create higher speed. As a result, accident hot spots should be studied taken into consideration the period of the day to take appropriate precautions in urban regions.

3.4. Spatial and Temporal Distributions

After detecting traffic accident hot spots in an urban region, to understand the distribution of these clusters, a) spatial distribution and b) time-based distribution of hot spots should be analyzed. If the traffic accident hot spots were not affected by these factors, we would have expected uniform distribution of them over the urban region and over 24 hour periods. Comparing the observed number of hot spots in different urban zones and time slots with expected values from uniform distribution, we can decide about the validity of the hypothesis. Furthermore, these analyses can be performed for pedestrian and non-pedestrian accident hot spots separately to understand the distribution of these accident type prone locations.

Spatial distribution of accident hot spots

Distribution of traffic accidents is not spatially random over region. Land use that effects directly the trip generation and trip distribution as well as vehicle kilometers traveled. For this reason, it is expected to be higher traffic volume in CBD and in densely urbanized areas. If there were no impact of land use on the formation of hot spots, dividing the urban region into zones would result in a distribution of hot spots proportional to geographical area of the zones. If the observed numbers of hot spots in zones are more or less than the expected values, it suggests that factors related to the zones may affect the traffic safety in that zone. Also, pedestrian activities are expected more in CBD and urban zones, where cultural, commercial, governmental and residential facilities are mostly seen; as a result more pedestrian accidents are expected in these regions. In the city center, due to traffic signalization, speed limit and traffic congestion, vehicles are expected to have lower speeds contrary to outer urban zones. For this reason, non-pedestrian accidents are expected to be seen more in the outer urban areas.

Time-based distribution of accident hot spots

High speed is a major factor causing serious accidents. The expected congestion during peak hours decrease the mean speeds significantly, which may reduce serious accidents (note: the bumper-to-bumper traffic conditions may increase property damage only accidents, on the other hand). Lack of traffic cause more speeding at night time, which in combination with improper lightening and visual issues may increase serious accidents. While it is technically possible to do an hourly traffic accident analysis, if there is not enough number of accidents for every hour, it may be more meaningful to divide the accidents (and hot spots) to basic categories such as, morning and evening peak periods, noon off-peak period, nighttime period. The cut off limits for these periods must be selected based on the local traffic flow characteristics in an urban region. The expected number of hot spots would be proportional to the total duration of the defined period.

3.5. Relation of Hot Spots with Built Environment

Understanding the relation between hot spots and urban built environments can be studied in two dimensions. First dimension, distribution of hot spot over urban zones is important to determine priority areas where precautions firstly take place. Since different urban land use creates different population density, speed profile, traffic volume and pedestrian activity, it is important to divide urban regions into different zones.

In the second dimension, accident hot spots should be analyzed and studied taken into consideration the urban network components, mainly intersections. Since intersections are the most dangerous conflict points of urban regions in terms of occurrence of traffic accidents, distances of hot spots to the intersections is important. To define where precautions are taken place to warn drivers and pedestrians, distance analysis is crucial.

3.5.1. Urban Zoning

As stated in the literature, different land use areas such as CBD make geographic distribution of accident locations (and hot spots) non-uniform over the space. To associate traffic accident hot spots (also called "clusters") with the urban land use, a simple way is to divide the urban region into zones with different transportation and land use patterns. Traditionally grown urban regions include, but not limited to, zones classified as;

- Central Business District (CBD),
- Urban Zones (or inner city) and
- Outside Zones (or transition zone).

Studying the analyses hot spots for traffic safety in urban regions, it is important to identify urban areas in terms of land use. Since CBD is an attraction center of the city, it is important to define it. Besides CBD, defining the inner city, which is can be considered as highly populated by settlement areas, should be defined. Mostly, high speed corridors are not located in the inner cities. For this reason, settlement areas should be separated from the high speed corridor. Furthermore, urban transition zones, where high speed corridors are seen should be identified. Moreover, if there exist, less dense areas like suburbs should be defined.

To draw conclusions on the distribution of the hot spots among different zones and time period, it is important to perform non-parametric test for the hypothesis. Statistical significance is a probabilistic assessment of the situation. In this approach, the real observed distribution of the hot spots in regard to spatial or time-based subgroups in an urban location should be checked against a null hypothesis (H_0) that such sub-grouping has no effect on the occurrences of the hot spots. This can be simply checked by performing a Chi-square (χ^2) test for the observed distribution of the hot spots over the null hypothesis.

3.5.2. Intersection Distance Analysis

Defining intersections in urban regions is important and challenging. First of all, on the contrary to strict geometric design codes for highways, there are no specific standards for road design in urban regions, in Turkey. That is why; intersection types and sizes vary greatly. For complex interchanges (Figure 3.2a) the borders of the intersection may be defined by the interchange ramps resulting in a non-simple shape. However, border of a simple at-grade intersection may be represented by a circular area with an appropriate radius.

Secondly, the intersections cannot be considered totally isolated from the traffic on the legs of it. So, the concept of intersection has to include an impact zone which would cover a certain amount of the legs, as well. The width of this impact zone should be selected such that stopping sight distances were included. The stopping sight distance is defined as the total of i) distance required to stop the vehicle with the braking system and ii) distance traveled during the perception-reaction time of an emergency situation that would require use of brakes. Assuming a rather conservative perception-reaction time of 2.5 seconds, the American Association of State Highways and Transportation Officials (AASHTO) guidebook (2004) suggests the use of total breaking distances shown in Table 3.1. The upper limits of 70 km/hr, 50 km/hr and 30 km/hr can be assumed for the average speeds on major arterials, minor arterials and collector streets in urban networks, respectively. As a result, we can suggest the use of 50m, 75m and 125m additional thickness around the intersection borders that would generate final area limit of intersections.

Table 3.1: Stopping sight distance on level (AASHTO,2004)

Design Speed (km/h)	Brake reaction distance (m)	Braking distance on level (m)	Stopping sight distance On Level	
			Calculated (m)	Design (m)
20	13.9	4.6	18.5	20
30	**20.9**	**10.3**	**31.2**	**35**
40	27.8	18.4	46.2	50
50	**34.8**	**28.7**	**63.5**	**65**
60	41.7	41.3	83	85
70	**48.7**	**56.2**	**140.9**	**105**
80	55.6	73.4	129	130
90	62.6	92.9	155.5	160
100	69.5	114.7	184.2	185
110	76.5	138.8	215.3	220
120	83.4	165.2	248.6	250
130	90.4	193.8	284.2	285
Note: Perception reaction distance predicated on a time of 2.5 s; deceleration rate of 3.4 m/s^2 used to determine braking distance.				

For distance analysis, the distance between the defined intersection limits and a hot spot area has to be determined. If the hot spot, partially or fully overlap with the intersection limits, the distance of the hot spot to the intersection is assumed zero (Figure 3.2a). Otherwise, the shortest Euclidean distance between the nearest points of the intersection limits and the hot spot border is calculated (Figure 3.2b). After calculating distances of each hot spot to the nearest intersection, these distances should be analyzed and evaluated. In the analyzing process, statistical significancy should be calculated. To do this, One-Way ANOVA test is a suitable to test the significance of distances. Significancy of distances should be analyzed with the dependent variables as intersection type, severity of hot spots etc. Analyzing distances with different variables help officials to take precautions.

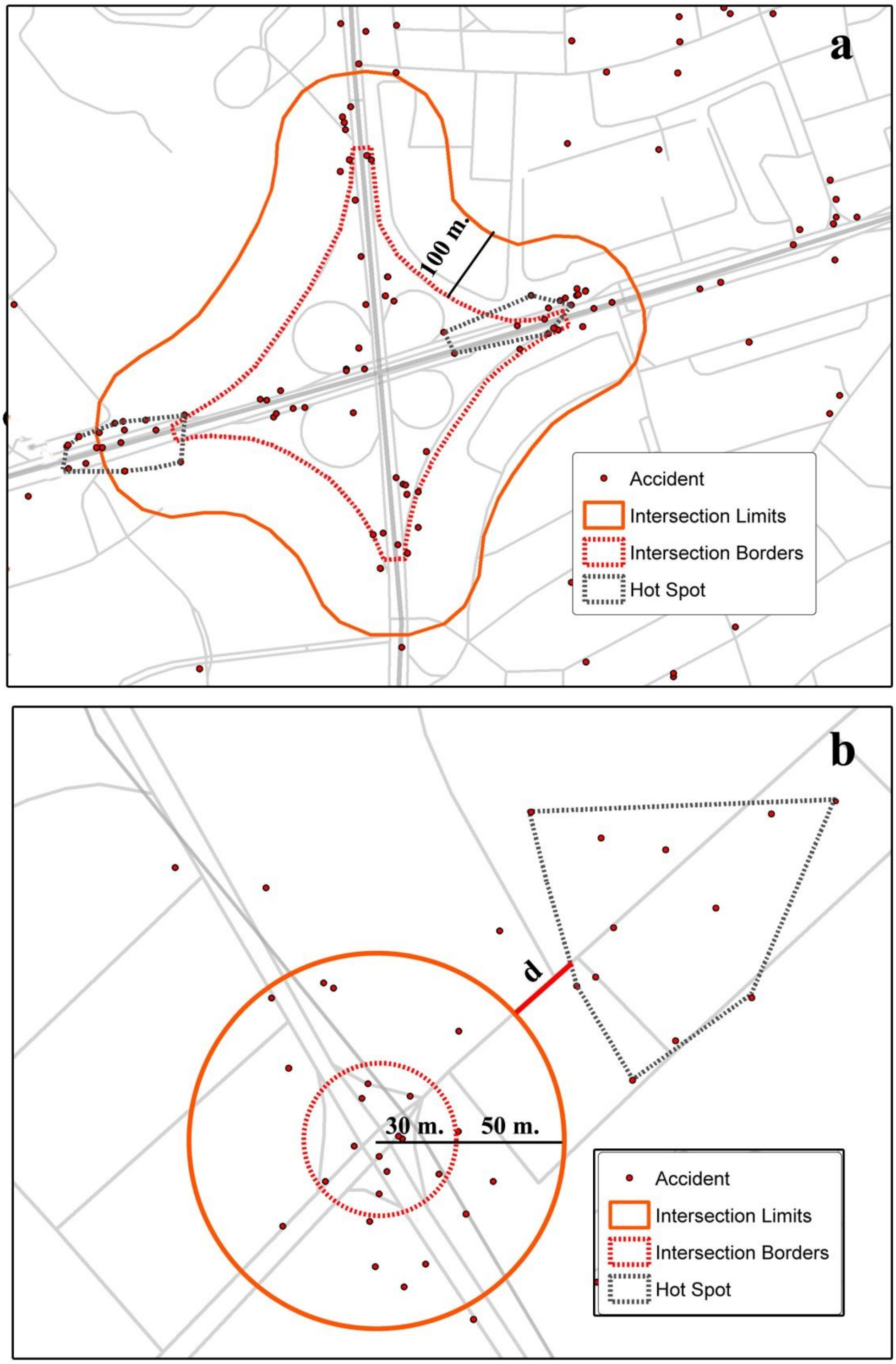

Figure 3.2: Example of limits of intersection definition at a) a grade-separated and b)an at-grade intersection

Distance intervals should be kept in the minimum intersection definition. Distance very close by data point is not captured in the definition can be captured by defining a short distance interval (such as 0-10m). Multiple distance intervals could be defined, but be careful not to relate further data points with the intersection. For example, 500+m can be group in one category. This mainly depends on city network and built-environment. It is important to do some preprocessing of the data and distance values to finalize the distance categories.

3.6. Comparison of Hot Spot Analysis by KDE

As mentioned in the Chapter 2, KDE is one of the most widely used methods to identify accident hot spot areas in traffic safety studies. For this reason, it is crucial to compare detected hot spots with the help of NNH clustering algorithm by KDE results. As explained before, KDE is an interpolation technique that generally used in the identification of hot spots. KDE calculates density estimate for each cell that are generated over the study area covered by incidents' locations.

In order to approximate the NNH clustering analyses results, total area of convex hulls are taken into consideration. Accordingly, average areas of convex hulls are calculated. This average can be considered as KDE cell size's area. According to this;

Average area of convex hulls = a^2

Cell size = a

Selection of bandwidth is an important issue of KDE. Bandwidth which refers to the width of Kernel is varied. Bandwidth affects directly the smoothness of resulting KDE map. A narrower bandwidth interval will lead to a finer mesh density estimate with all little peaks and valleys. Great differentiation between areas can be provided with smaller bandwidths. On the other hand, larger bandwidth provides smoother distributions. For this reason, less variability observed between areas with larger bandwidths.

Visualization of KDE maps is another issue over which thematic range to choose the represent different threshold values. It is difficult to decide which range should be chosen as hot spot. Showing the hot spots is needed to select threshold value of density estimation. In order to determine threshold values of this study's analyses, incremental multiply of the grid cell's mean is used. This visualization technique depends on the mean value of grid cells except cells that have zero values. Firstly, the mean value of cells have greater than zero value are calculated, then according to this mean value, categories are specified (Chainey et al., 2002; Eck et al., 2005). Accordingly, cells that have greater than the fivefold of the mean value is accepted as hot spot.

3.6.1. Visual Comparison of Methods

To compare NNH clustering analysis results and KDE results, resulted map of both methods are overlay to show how these methods differentiate in terms of a visual aspect. Since different types of KDE methods (normal, uniform, quartic etc.) have different mathematical formula, visual differences are expected from each type of KDE method. Moreover, as mentioned in the earlier chapter, NNH clustering algorithm has also different outputs; convex hull and ellipse. In the visual comparison step, all results and outputs should be compared with each other.

3.6.2. Mathematical Comparison of Methods

Other than comparing KDE and NNH clustering methods results visually, some mathematical calculations should be done. PAI helps to compare the ability of capturing hot spots between different methods. As mentioned earlier, PAI depends on the number of incidents and total area in terms of both hot spots and study area. In other words, PAI measured the methods' capturing ability of hot spots taken into consideration all the data and all the study area. The PAI is derived from the ratio of hit rate (the number of points in hot spots compared to the region) to the area percentage (proportion of the region represented by hot spot).

In the PAI calculation step, it is meaningful to compare all methods' results and outputs to understand and decide which method and type of output gives more accrurate result to detect hot spot.

3.7. Strength and Weakness of the Proposed Approach

As mentioned above, there are several spatial analysis methods that can be used traffic safety studies. This study specifically focused on utilization of, NNH clustering algorithm that depended on the point pattern analysis and constituted hot spots only from the defined criteria. Another well known point pattern analysis method, K-Means clustering divides data into K groups which are defined by the user. Method created K number of clusters from all data; all points are assigned to only one group and all points are clustered. This method can be useful when a researcher wants to control the grouping. For example, if 5 best ambulance deposit areas are tried to found, K-Means clustering method with the definition of 5 clusters gives an idea about the densest areas. Contrary to the K-Means clustering method, NNH clustering algorithm provides controlling the size of grouping either with defining of threshold distance and n_{min} in cluster. NNH clustering gives chance to identify dense small geographic environment. This advantage of NNH algorithm provides the identification of hot spots for the aim of this study. Since the main aim of this study is to identify hot spots indicating the specific location in subscale, NNH clustering algorithm is preferred to detect hot spot locations.

For an urban wide analysis, criteria can be defined based on the characteristic of the urban environment. Since the method gives the actual area of hot spot as a convex hull, where defined criteria are provided, it is easier to relate them with built environment. Most safety engineering interventions require identifying the specific location where accidents occur and specific factor causing them. This method that focuses on the location of accidents will be more useful for traffic safety engineers than one that the phenomenon over a larger area. For example, KDE smooth accident data over small grid cells. However, by smoothing the data, it will distort relationships making it appear that there are higher accident likelihoods in locations than actually exists. For instance, if there is an accident hot spot at the intersection of two arterials, the actual concentration of the accidents will be at the intersection or on the approaches to the intersection from all sided. Yet, the KDE will smooth the data to make it appears that adjacent areas have a substantial likelihood of having accidents, when, most likely, there are very few in the residential neighborhoods.

When performing NNH clustering algorithm, choice of a threshold distance is an important issue. In this method, defining threshold distance is subjective. Determining the appropriate threshold distance depends on experience and characteristics of study area. After the trial period of different threshold distances, analyzer can determine suitable threshold distance based on the scope of the study. On the other hand, characteristic of study area, in terms of network, distance can be increased or decreased. Parallel roads which are close to each other, closer intersections or narrower intersections may need smaller threshold distance definition.

For the aim of interventions for specific locations, the use of a smaller threshold distances may actually be appropriate. This definition creates micro-neighborhoods or almost specific hot spot locations. Same situation is valid for the definition of n_{min} in cluster. A criterion, definition of n_{min}, is also subjective and depends on the data features. Without a definition of n_{min}, program could identify clusters of two or three points. However, a hot spot of this size is usually not very meaningful. Consequently, the researcher should increase the number, and tried several numbers to be sure that the identified cluster represents a meaningful number of points. Like determination of threshold distance, specifying the n_{min} value depends on experience. After a trial period, user can decide optimum value.

Besides advantages, NNH clustering method has also limitations. One of them is, method only clusters points; a weighting or intensity variable have no effect on clustering results (Levine, 2010). Furthermore, since there is no information about the severity of injuries in terms of slight, serious etc., accidents are analyzed only as point events. As the method works with the real data points, any unreliability on the data side would be directly translated to hot spot detection, as either misdetection or un-detection of a potential hot spot.

CHAPTER 4

CASE STUDY: TRAFFIC SAFETY IN ANKARA

4.1. Province of Ankara

As a case study, Ankara was selected, as it is one of the pilot cities of Road Safety-10 (RS-10) project, conducted by WHO in Turkey and nine other countries simultaneously (Hyder and Bishai, 2012).

City of Ankara, which is the capital city of Turkey, located in the Central Anatolia Region. Ankara had a population of 4.9 million in 2012 according to the data from Turkish Statistical Institute (Web 2). Ankara has 25501 km^2 areas including 25 counties namely; Altındağ, Ayaş, Bala, Beypazarı, Çamlıdere, Çankaya, Çubuk, Elmadağ, Güdül, Haymana, Kalecik, Kızılcahamam, Nallıhan, Polatlı, Şereflikoçhisar, Yenimahalle, Gölbaşı, Keçiören, Mamak, Sincan, Kazan, Akyurt, Etimesgut, Evren and Pursaklar.

Ankara is a junction point of important highways connecting cities that Konya Highway is in the south, Kırıkkale Highway is in the east, Çankırı Highway is in the north-east, İstanbul highway and motorway is in the north-east, Eskişehir Highway is in the east. In order to provide transit access without entering urban area, beltway was constructed around the city center (Figure 4.1).

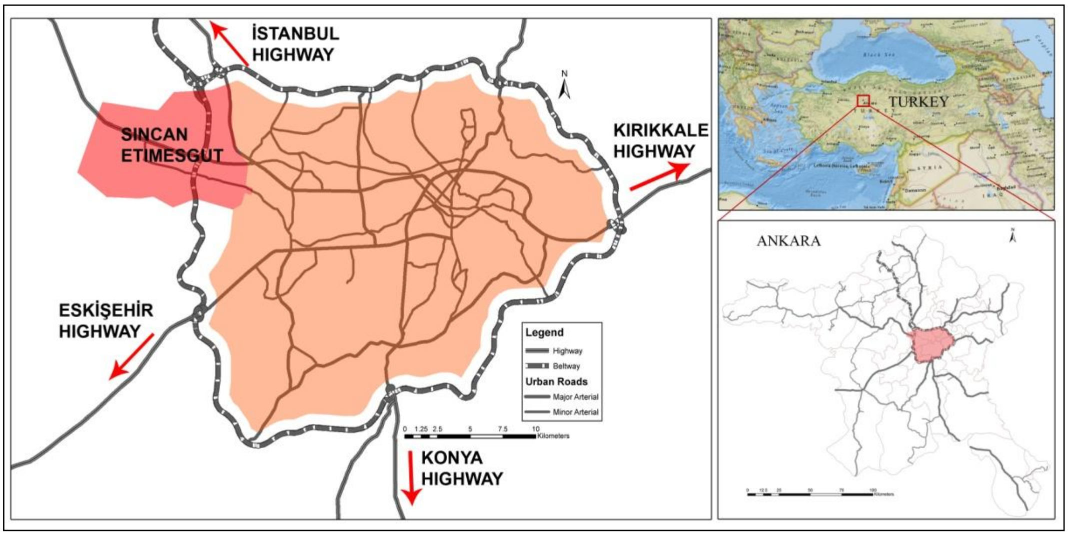

Figure 4.1: Location of Ankara and study area

4.2. Traffic Accident Data for Ankara

Traffic accident data with fatalities or injuries in Turkey had been collected by police at crash site. The reports are later digitized by local branches of police force and sent to the traffic accident database held by the General Directorate of Security Affairs. Traffic accident data provided to the WHO Turkey office include accident ID, date, time, location, type of accident, number of vehicles involved in an accident, road type, road geometry, intersection type, geographical coordinates of accident, age and sex of driver, weather condition, lighting condition, vehicle type and number of persons injured/killed.

Like many other geographic data, some errors can be seen in traffic accident data due to random or systematic errors. It has been mentioned in different reports that incorrect datum selection by the users causes some errors. Moreover, inconsistent data entering while transforming coordinates from GPS to report or report to the database resulted in mis-location of accidents in the GIS environment. In the study area, in certain locations, some accidents look as if they were shifted. Such a problematic area can be seen from the Figure 4.2. Where there was no such a straight road network, there is a sharp line as data border. Furthermore, complex street network is seen within the boundary of ellipse.

Such problems can be solved with the help of the address information of traffic accident data, which was missing for the data obtained from WHO Turkey. Thus, in a complex urban network area such Ankara, it was hard to correct the location of accidents visually or manually. For this reason, geographic location of traffic accident data for this study cannot be checked and corrected.

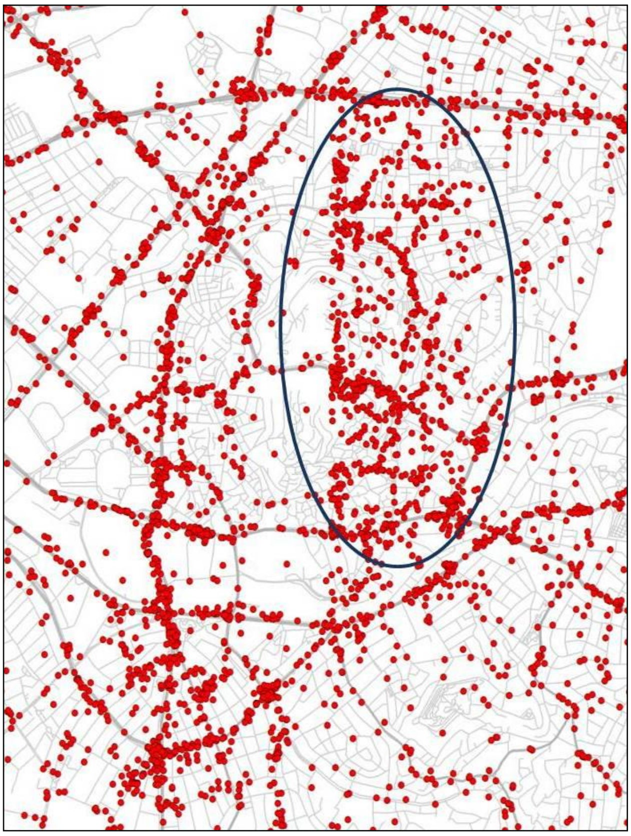

Figure 4.2: Shifted accidents in the case study area

4.3. Traffic Safety Statistics for Ankara

Traffic accident data of years 2008, 2009 and 2010 in the study area are used in the analysis of hot spots. Data were taken from World Health Organization. Between the years 2008-2010, a total of 24873 traffic accidents occurred in all Ankara; 17149 of them were within the boundary of urban (study) area. Out of these accidents, 24473 of them were injury accidents; on the other hand 400 of them resulted in fatality. Out of 400 fatality accidents, 156 of them were occurred within the boundary of urban area. For both type of accidents, most of them were seen in urban region. Descriptive statistics of 3 years accident data in Ankara is summarized in the Table 4.1.

Table 4.1: Descriptive statistics of 3-yr accidents in Ankara

	# of accidents	# of injury accidents	# of injury	# of fatality accidents	# of fatality
Ankara Urban Region	20876	20665	31170	211	261
Urban Area(Study Area)	17315	17149	25327	156	196
Sincan-Etimesgut	2679	2660	4256	29	34
Beltway	882	856	1587	26	31
Highway	2266	2114	4977	152	205
Rural Areas	1731	1694	3220	37	45
TOTAL	***24873***	***24473***	***39367***	***400***	***511***

As it can be seen from the Figure 4.3 fatality accidents are rarer events comparing to the injury accidents. For this reason, it is appropriate to study and analyze fatality and injury accidents separately.

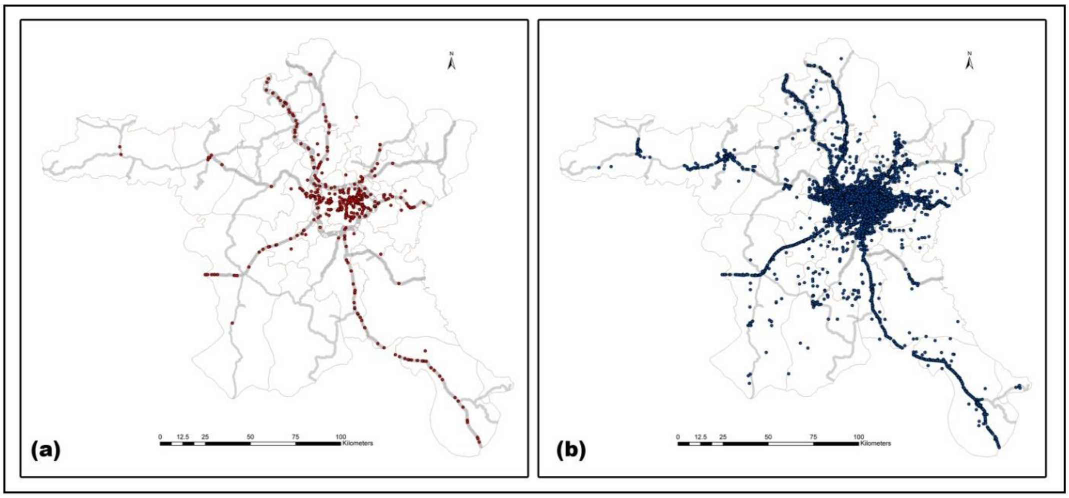

Figure 4.3: Fatality (a) and Injury (b) accidents in Ankara

According to the annual report of Ankara published by Turkish Statistical Institute (Web 3), with 209 vehicles per thousand population car ownership level, the city of Ankara ranks the first among other metropolitan cities of Turkey (Figure 4.3). Taken into consideration the mean value of car ownership of Turkey (114 vehicle/1000 people), Ankara has almost the double of mean values of the country. According to the number of accidents per ten thousand vehicle statistics among metropolitan cities of Turkey, Ankara has 82 accidents (Figure 4.4). This value is low compared to the other metropolitan cities and the mean value of Turkey. Ankara ranks the 72nd among all cities of Turkey according to the statistical data of number of traffic accident per 10000 vehicles (Web 3).

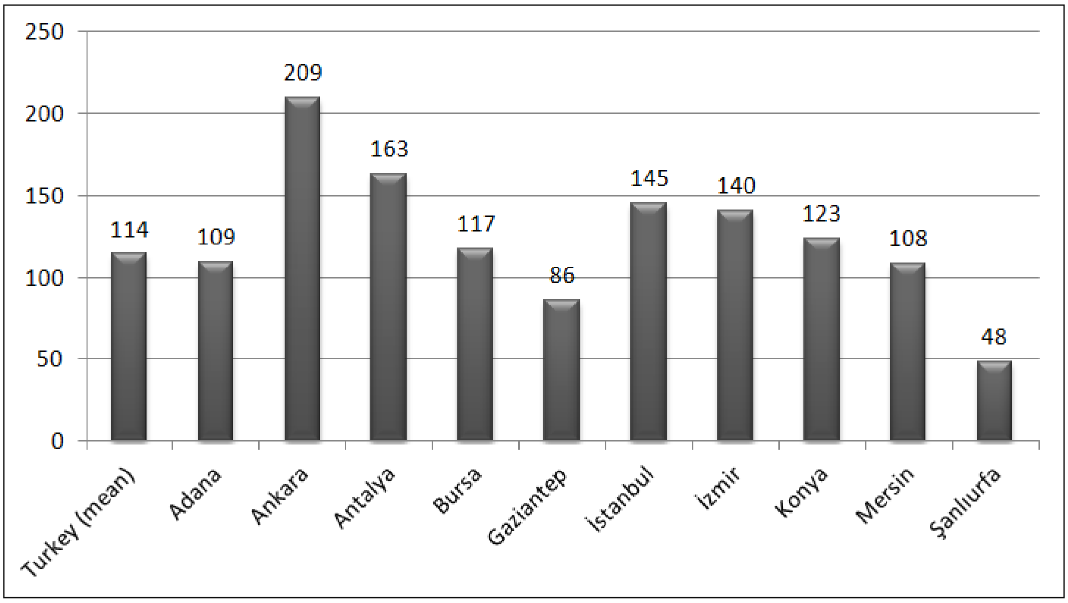

Figure 4.4: Number of motor vehicle per 1000 people

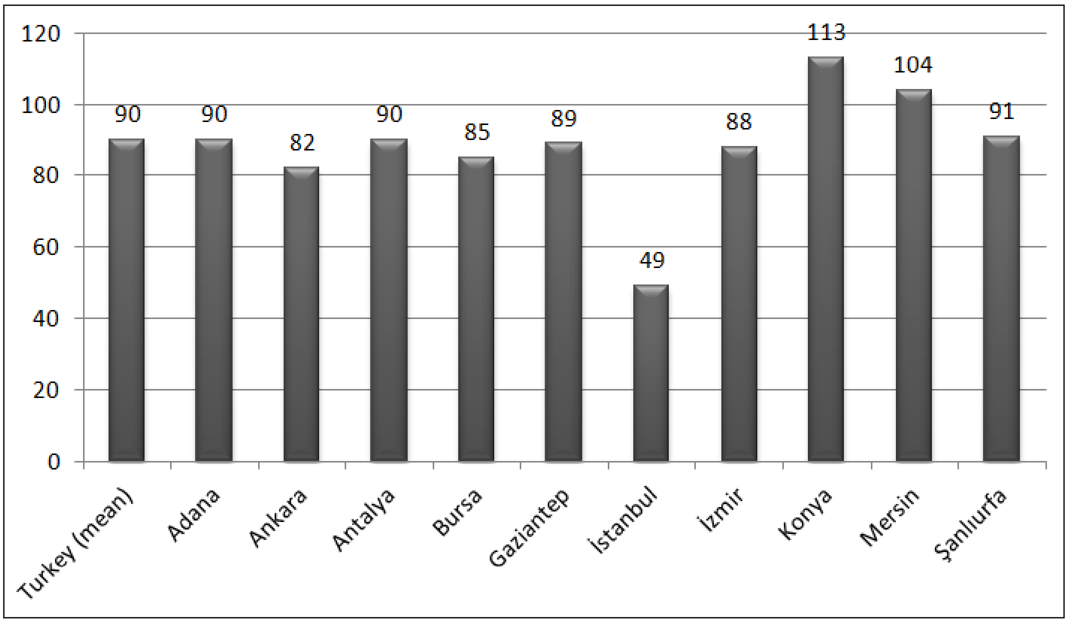

Figure 4.5: Number of traffic accidents per 10000 vehicles

4.4. Description of the Study Area

The study area included the urban region inside beltway (assuming 1 km buffer around it), as the traffic conditions and speed limits on the beltway are closer to highway levels than urban arterials (Figure 4.2). Areas of seven counties, Altındağ, Çankaya, Yenimahalle, Gölbaşı, Keçiören, Mamak and Etimesgut, are included within the beltway. Sincan and Etimesgut, western counties of Ankara, are excluded from the analyses since they have their own center of attraction. Because density of pedestrian and non-pedestrian traffic is different from Ankara city center, accidents occurred in Sincan and Etimesgut have been excluded from the analyses.

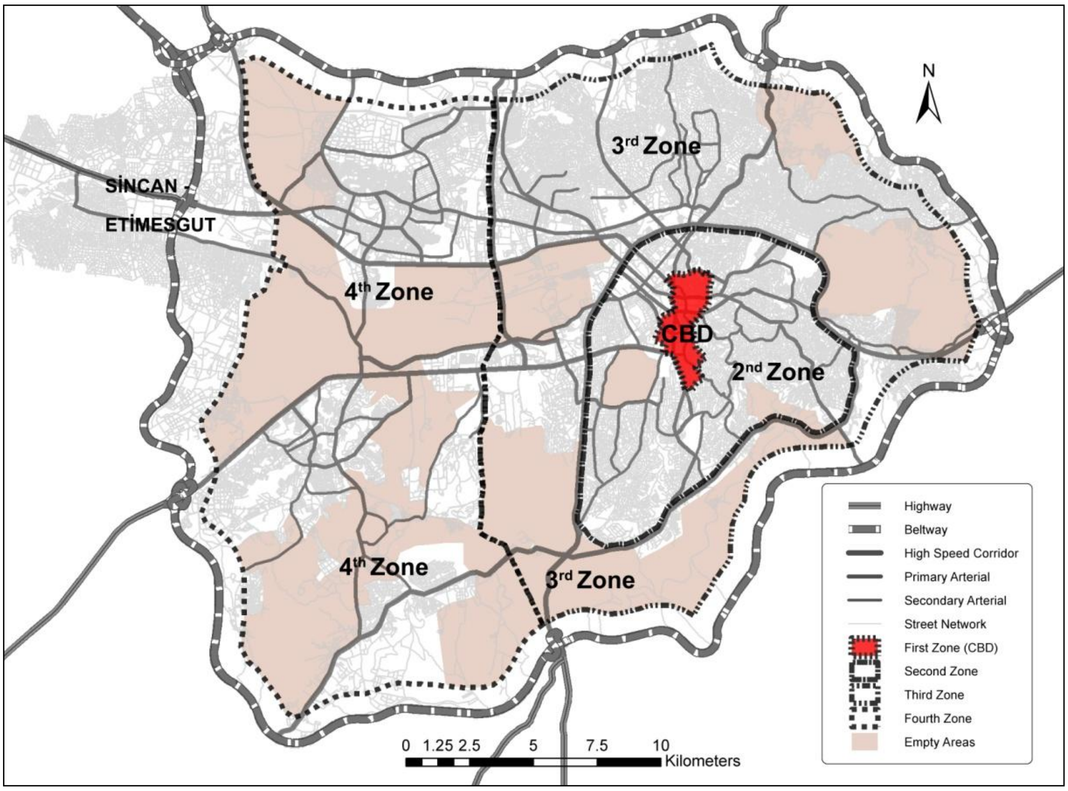

Figure 4.6: Study area

4.5. Spatial and Temporal Distribution Criteria

Urban Zoning

The study area is separated into four zones (Figure 4.6); first zone also called as CBD shows the old city which has been the center of shopping, business and governmental use. The second zone is included the early residential settlement areas surrounded by the old interstate ring road which is now an important urban arterial corridor. Third zone is also included residential areas but not much business area and shopping areas. The last, fourth zone is included less dense urbanized areas.

Kızılay and Ulus, the center of shopping, business and governmental use, are defined as the first zone called as Central Business District (CBD). This area has various types of land-use areas from residential, commercial, business, parks, public buildings, hospitals and cultural activity areas, etc. Also, this area is the hub center of public transportation where two metro rail-lines namely Ankaray and Batıkent metro rail lines are centered in Kızılay. Moreover, almost all the bus transit lines are distributed from CBD and they execute a u-turn to complete their journeys, which end at the point of departure (Yetiskul and Senbil, 2012). For these reasons, it is most probable to observe pedestrian activities high in this zone, and accordingly pedestrian accident hot spot are expected to be seen densely in this area.

Outside of the CBD, second zone is constituted from also mainly residential, commercial, business and governmental buildings. The second zone can be defined as the early residential settlement areas surrounded by old interstate roads; Mevlana and Turgut Özal Boulevard, Kırıkkale Highway, Doğukent and Turan Güneş Boulevard. These roads are not counted within this zone. Besides the residential usage, military areas, commercial, business and public buildings can be seen in this zone. Also, many of representative offices of countries are located in this zone.

Third zone is almost constituted from residential and commercial facilities. Different from the former two zones, this zone has big open areas like Atatürk Orman Çiftliği, forestry and military zones. Also, many big shopping malls are seen in this zone, which create traffic.

The region defined as the fourth zone covers the west side of the study area. This area is recently being populated and has less dense urbanization comparing the other zones. Different from the other three zones, industrial usages are seen in this zone namely Ostim and İvedik in the north part. Also, important university campuses can be seen namely Bilkent, Hacettepe, Başkent and Çankaya Universities. In the south part, recently developed residential areas can be seen namely Ümitköy, Yaşamkent, Konutkent etc. Besides these neighborhoods, there are suburban areas where detached houses are seen in the south part of the fourth zone.

Time Periods

As mentioned in the previous chapter, different time of day has different traffic volume. For this reason, time periods are created taken into consideration the peak and off-peak hours of Ankara. According to this, 4 time periods are specified. Two peak hours periods are, namely morning peak and evening peak, identified taken into consideration the activity of people like school, working, shopping, etc. Accordingly, hours between 7am and 9am are considered as the morning peak; while hours between 4pm and 8pm are considered as evening peak. Noon-off peak period is taken between the hours 10am and 3pm. Finally between the hours 9pm and 6am are identified as nighttime period.

Intersection Definitions

Road network data of Ankara is available for this study. With the help of road network data and open street map available in ArcGIS program, urban arterials and intersections are identified. Urban arterials are classified into three groups;

high speed corridor, primary arterial and secondary arterial. After defining urban arterials by type, intersections are identified. Intersections are also classified into three groups; interchange, major and minor intersections. Every crossing point of defined urban arterials whether on high speed corridor or major arterial are specified as major intersections. Besides this, intersections, which divide both high speed corridor and major arterial, are also defined as major intersections. Intersections, which do not divide arterials, are not counted in the intersection definition. Other than these, every crossing point of secondary arterials are classified as minor arterials. Accordingly, total number of 33 interchanges is found in the study area. Besides this, total number of 124 intersections is identified as major while 47 intersections are identified as minor. Road network and identified intersections can be seen in the Figure 4.7.

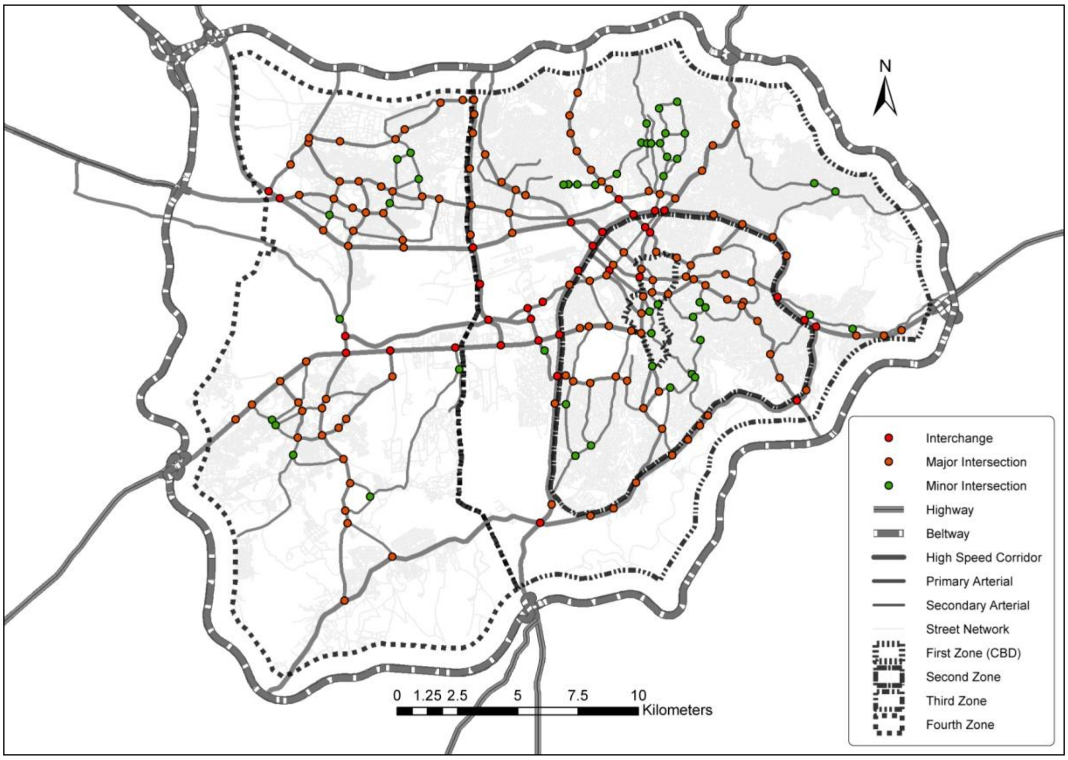

Figure 4.7: Road network and intersections in study area

In order to define impact zone of intersections, boundaries of multilevel junctions and the middle point of intersections are defined. For each multilevel junction, 100 meter buffer zones are created from their boundaries to identify the impact zone. For each at-level junctions, intersections' sizes are defined. Then, according to the type of speed level, impact zones are identified from the boundary of intersections. Speed level is considered as 70 km/h for high speed corridors, 50 km/h for primary arterials and 30 km/h for secondary arterials. Impact zones for each speed levels are calculated as 125 meters for 70 km/h speed, 75 meters for 50 km/h and 50 meters for 30 km/h. Then these impact zones are added to the size of the intersections and created circular area.

4.6. Fatality Accident Analysis for the Study Area

As mentioned previously, fatality accidents are much rare events comparing to the injury accidents. In the study area, total numbers of 156 accidents were occurred resulted with fatality. As can be seen from the Figure 4.8, fatality accidents were located mainly on old interstate ring roads and main arterials. Given information with accident database, fatality accidents are examined in terms of occurrence type and time.

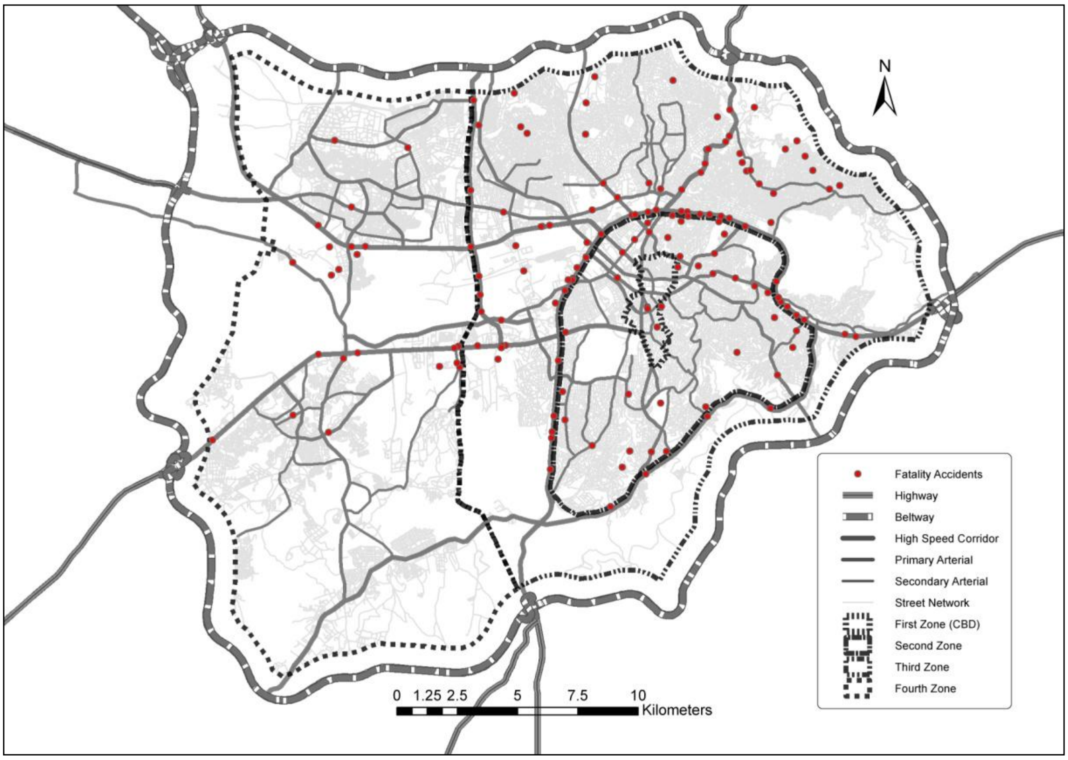

Figure 4.8: Fatality accidents in the study area

Occurrence type of fatality accidents shows differences, as well. According to the Figure 4.9, out of 156 fatality accidents, 71 of them were pedestrian accidents. It is expected to see dense pedestrian accidents especially in urban areas because of the high pedestrian movements. On the other hand, fatality accidents were occurred as a result of a collision with stationary object has high value comparing with other occurrence types.

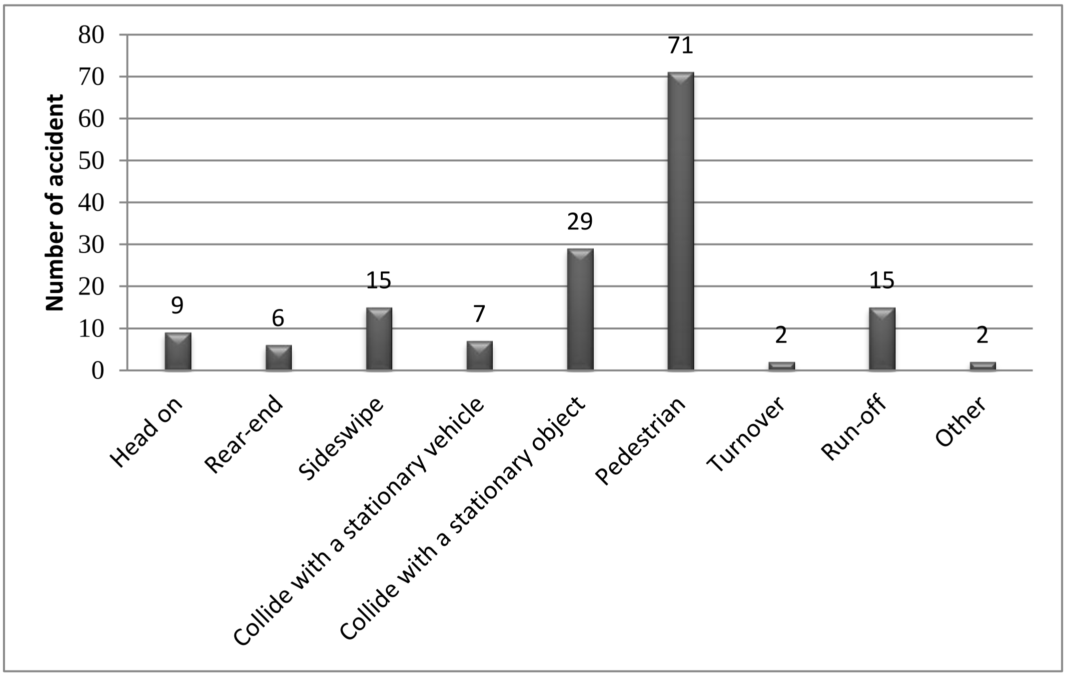

Figure 4.9: Number of fatality accidents by type in Ankara Urban region

Since fatality accidents are rare events in terms of numbers, occurrence times of accident are examined taken into consideration the peak hours of traffic. Relationship between number of accident and occurrence time of the accident is shown in Figure 4.7. Since visibility range is lower in nighttime periods comparing to daytime hours, high numbers of accidents occurred in nighttime periods are expected. The results are explained by merely the possibility that due to the low traffic volume in the off-peak periods, generally more speed is seen, and as a result more accidents occurred.

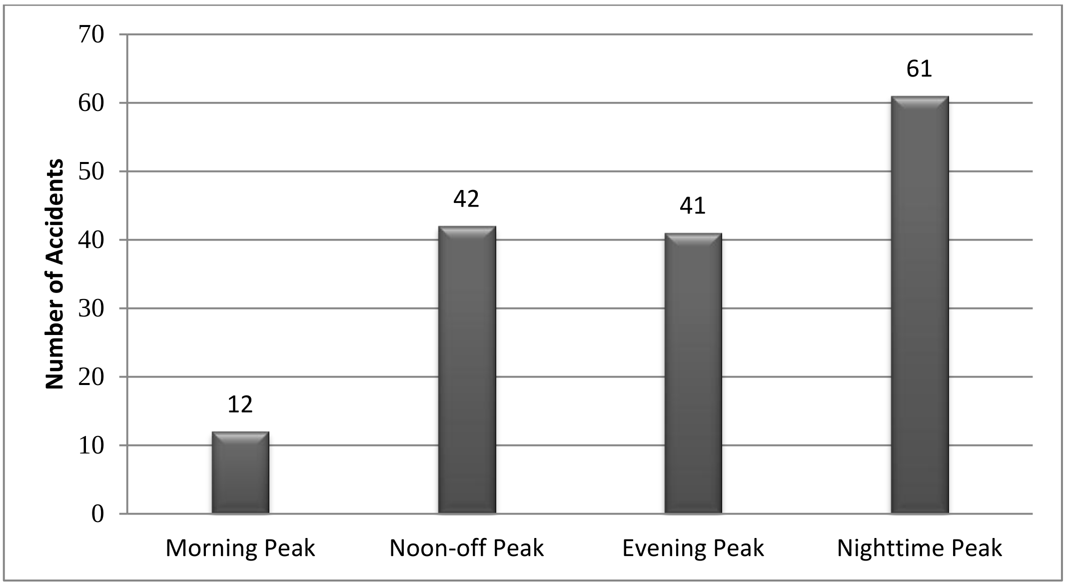

Figure 4.10: Fatality accidents in Ankara urban region by time periods

Descriptive statistics and spatial distribution of fatality accidents in terms of type and time periods is given in the Table 4.2. According to this, it is apparent that most of the fatality accidents were occurred in the third zone where high speed corridors and major arterials are located in. In both types of fatality accidents; pedestrian and non-pedestrian, also are highly seen in the third zone. Same situation is also seen in timely grouped accidents. For all time groups, number of fatality accidents is high also in the third zone.

Table 4.2: Descriptive statistics of 3-yr fatality accidents in study area

			1st Zone (CBD)	2nd Zone	3rd Zone	4th Zone	TOTAL
Network Statistics							
		Area (km^2)	5.51	75.8	117.25	106.66	**305.22**
		Total road length (km)	121.1	1470.3	2619	2051.2	**6261.6**
		Main Arterial road length (km)	10.1	60.2	132.2	102.3	**304.8**
Network Statistics (by percent)							
		Area (%)	1.8	24.8	38.4	34.9	100.0
		Road length_total (%)	1.9	23.5	41.8	32.8	100.0
		Road length_ main arterials (%)	3.3	19.8	43.4	33.6	100.0
Accident Data Set		**Fatality Accidents Statistics (by Type)**					
	All	# of accidents	3	36	94	23	**156**
		Accidents/km^2	0.54	0.47	0.8	0.21	**0.51**
		Accident/km arterial	0.2	0.44	0.62	0.18	**0.42**
		Accident/km road	0.02	0.02	0.04	0.01	**0.02**
	Pedestrian	# of accidents	1	19	45	6	**71**
		Accidents/km^2	0.18	0.25	0.38	0.05	**0.23**
		Accident/km arterial	0.07	0.23	0.29	0.05	**0.19**
		Accident/km road	0.01	0.01	0.02	0	**0.01**
	Non-Pedestrian	# of accidents	2	17	49	17	**85**
		Accidents/km^2	0.36	0.22	0.41	0.15	**0.27**
		Accident/km arterial	0.13	0.21	0.32	0.14	**0.23**
		Accident/km road	0.02	0.01	0.02	0.01	**0.01**
Accident Data Set		**Injury Accidents Statistics (Time-based)**					
Daytime	**Morning Peak (7am-9am)**	# of accidents	0	2	8	2	**12**
		Accidents/km^2	0	0.02	0.06	0.01	**0.03**
		Accident/km arterial	0	0.02	0.05	0	**0.03**
		Accident/km road	0	0	0	0	**0**
	Noon off-peak (10am-3pm)	# of accidents	0	12	26	4	**42**
		Accidents/km^2	0	0.15	0.22	0.03	**0.13**
		Accident/km arterial	0	0.15	0.17	0.03	**0.11**
		Accident/km road	0	0.01	0.01	0	**0.01**
	Evening Peak (4pm-8pm)	# of accidents	0	8	26	7	**41**
		Accidents/km^2	0	0.1	0.22	0.06	**0.13**
		Accident/km arterial	0	0.1	0.17	0.06	**0.11**
		Accident/km road	0	0.01	0.01	0	**0.01**
	Nighttime (9pm-6am)	# of accidents	3	14	34	10	**61**
		Accidents/km^2	0.54	0.18	0.28	0.09	**0.2**
		Accident/km arterial	0	0.17	0.22	0.08	**0.16**
		Accident/km road	0	0.01	0.01	0	**0.01**

4.7. Injury Accident Analysis for the Study Area

Traffic accident resulted with injury in the study area frequent events comparing to the fatality accidents. As can be seen from the Figure 4.11, there are many injury accidents over the study area. However, most of the injury accidents were located on high speed corridors and urban arterials. Visually, accidents are decreased gradually from the center to outer urban areas.

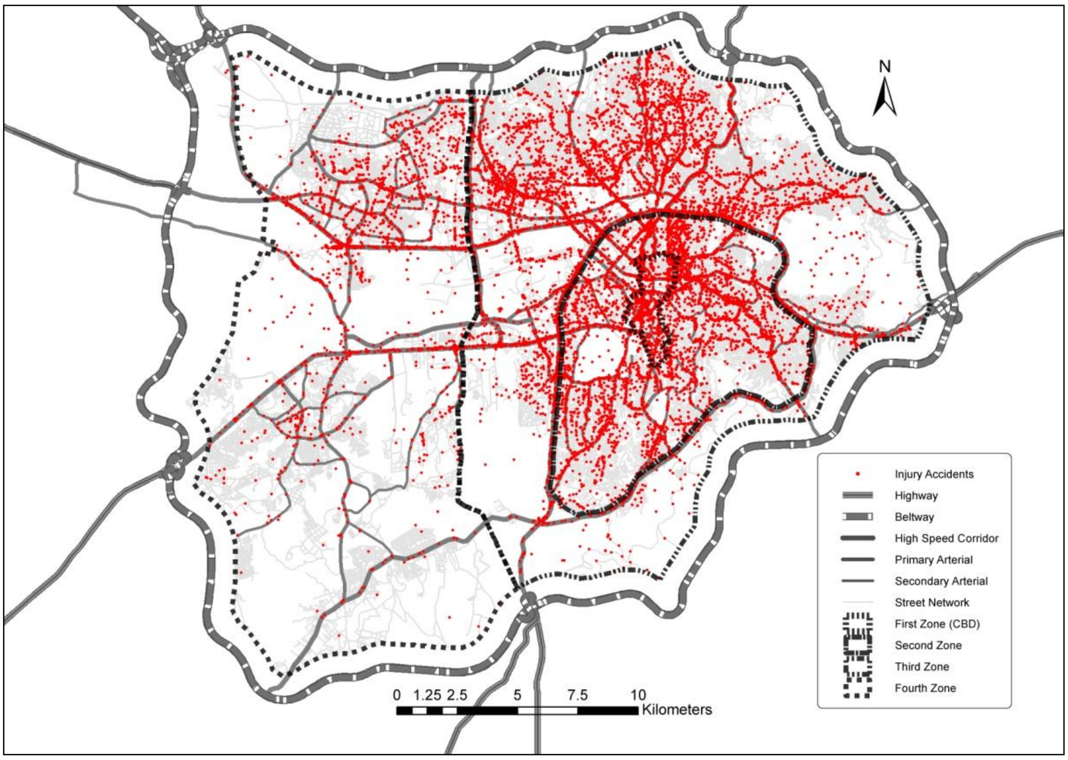

Figure 4.11: Injury accidents in the study area

A number of occurrence types of injury accidents seen from the Figure 4.12. According to the figure, pedestrian accident is high in number comparing to the other types of accidents. Collision with a stationary object, sideswipe and rear-end accidents are also had high numbers in urban area. Although there are 11 different types of accidents available in traffic accident database, for this study accidents were grouped into two types; pedestrian and non-pedestrian. Taking into consideration the uniform distribution of accidents in terms of accident types, this difference between accident types is statistically significant at p=0.05 value.

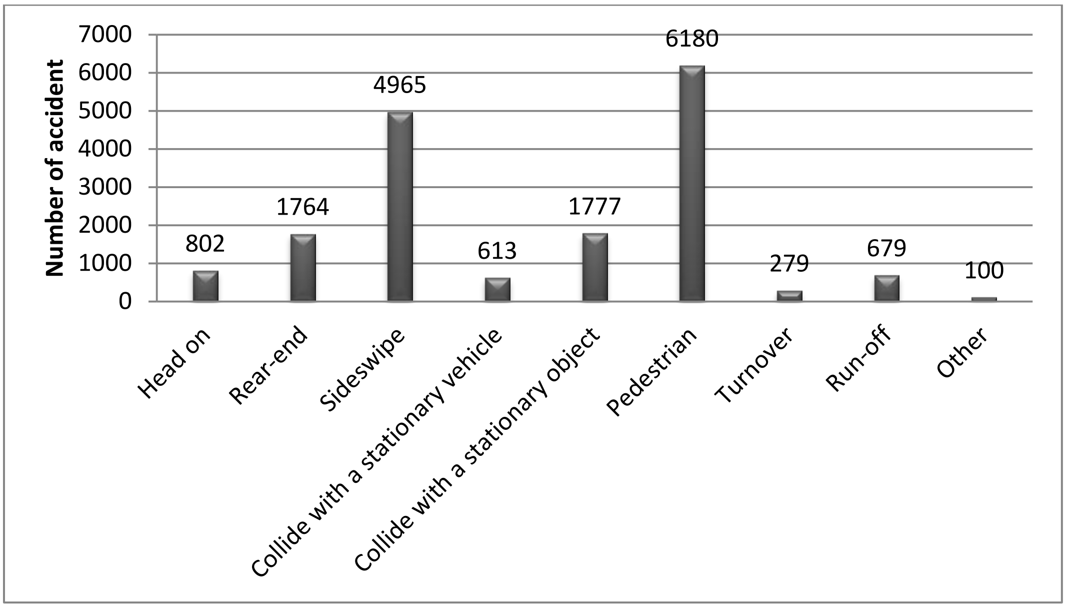

Figure 4.12: Number of injury accidents by type

Besides the event types of data, occurrence time is also taken into consideration. In the Figure 4.13, hourly distribution of injury accident data is shown. According to this, numbers of accidents have a peak especially in evening hours. Since it is difficult to study accidents by hours, accidents are calculated taken into consideration the time periods according to peak and off-peak hours.

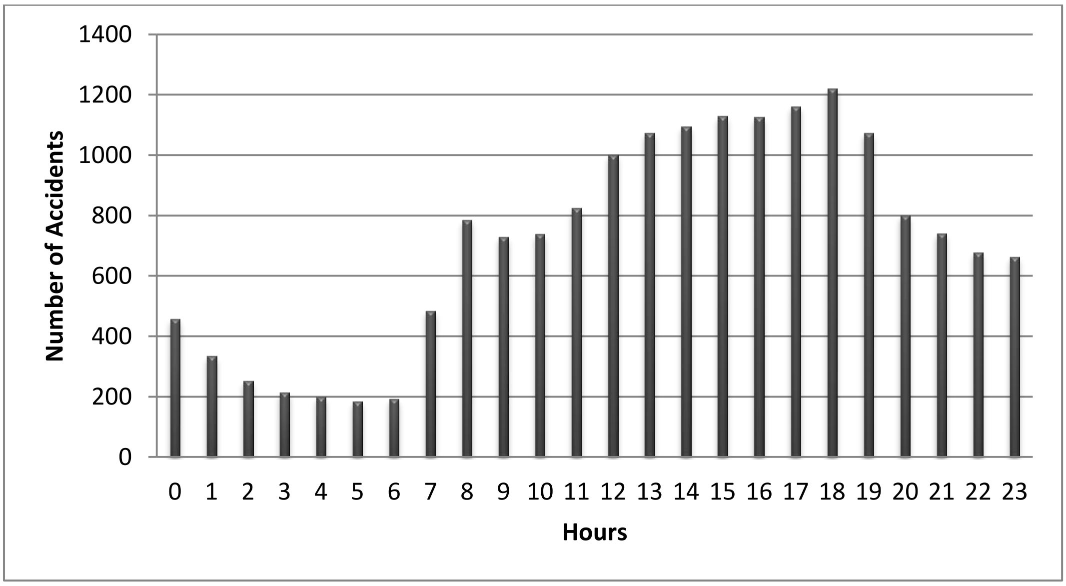

Figure 4.13: Number of injury accidents by hours

Regarding with occurrence time of injury accidents (Figure 4.14), out of 17149 them; 1999(11.7%) were seen between 7am and 9am, 5858(34.2%) were seen between 10am-3pm, 5378(%31.4) were seen between 4pm-8pm and 3914(22.8%) were seen between 9pm-6am. Frequencies of accidents with respect to time periods differ from each other. If all crashes were uniformly distributed in accordance with hours, expected number of accidents would be higher for 7am-9am and 9pm-6am and would be lower for 10am-3pm and 4pm-8pm. This difference was tested with the help of Chi-Square test. According to result, distribution of accidents regarding with time periods is statistically significant at p=0.05 value.

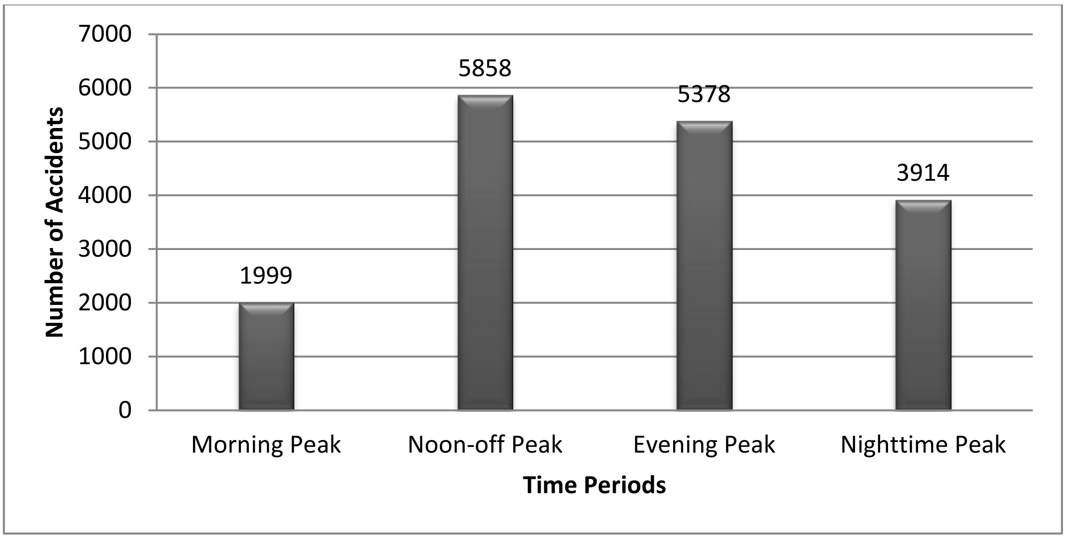

Figure 4.14: Number of injury accidents by time periods

In three years, a total of 24473 traffic injury accidents occurred in all Ankara; 17149 (%70) of them were within the limits of the study area. Even this share suggests that most of the injury accidents were seen in the urban area. Out of 17149 accidents in the study area, 6178 (%36) of them are pedestrian, and 10968 (%64) of them are non-pedestrian accidents. Since study area is an urban area, a significant proportion of pedestrian accident is expected due to the high mobility of pedestrian activities in study area.

In total 17149 injury accident, 25327 people were injured. In total 6178 pedestrian accidents, 6727 pedestrians were injured. On the other hand, in total 10968 non-pedestrian accidents, 18613 people were injured.

Descriptive statistics and spatial distribution of injury accidents in terms of type and time periods are given in the Table 4.3. According to this, it is apparent that most of the accidents were occurred in the third zone where major arterials are located in. This situation is also valid for another group of accidents.

When accidents are considered as occurrence density regarding with zones' urbanized areas, the most striking result to emerge from the accidents is that they gradually decreased from CBD to outer urban zones in terms of both groups of accidents. Pedestrian accidents have a great number of densities in terms of the geographical area compared to the non-pedestrian accidents. Besides this, pedestrian accidents have a very low density in the fourth zone compared to the other groups of accidents. Differences in distributions of both pedestrian and non-pedestrian accidents were tested with the help of Chi-Square test, and results were significant at the p=0.05 level.

Temporal analysis shows that, most of the injury accidents occurred in noon-off peak and evening peak periods. Like previous findings, these groups of accidents also were seen the highest percentage in the CBD in terms of accidents per square kilometers. Different from the accidents distribution in terms of type, accident densities for all time periods almost show the same results in second and third zones. The geographical distribution of hourly grouped data shows differences. These differences were tested by Chi-Square test, and results were significant at the p=0.05 level for all four groups.

Table 4.3: Descriptive statistics of 3-yr injury accidents in study area

			1st Zone (CBD)	2nd Zone	3rd Zone	4th Zone	TOTAL
Network Statistics							
		Area (km^2)	5.51	75.8	117.25	106.66	**305.22**
		Total road length (km)	121.1	1470.3	2619	2051.2	**6261.6**
		Main Arterial road length (km)	10.1	60.2	132.2	102.3	**304.8**
Network Statistics (by percent)							
		Area (%)	1.8	24.8	38.4	34.9	100.0
		Road length_total (%)	1.9	23.5	41.8	32.8	100.0
		Road length_ main arterials (%)	3.3	19.8	43.4	33.6	100.0
Injury Accidents Statistics (by Type)							
Accident Data Set							
	All	# of accidents	1126	5403	8161	2459	**17149**
		Accidents/km^2	204.3	71.2	69.6	23	**56.2**
		Accident/km arterial	111.5	89.8	61.7	24	**56.3**
		Accident/km road	9.3	3.7	3.1	1.2	**2.7**
	Pedestrian	# of accidents	675	2329	2820	357	**6181**
		Accidents/km^2	122.5	30.7	24	3.3	**20.2**
		Accident/km arterial	66.8	38.7	21.3	3.5	**20.3**
		Accident/km road	5.6	1.6	1.1	0.2	**1**
	Non-Pedestrian	# of accidents	451	3074	5343	2100	**10968**
		Accidents/km^2	81.8	40.5	45.5	19.6	**35.9**
		Accident/km arterial	44.7	51.1	41.9	20.5	**36.6**
		Accident/km road	3.7	2.1	2.1	1	**1.8**
Injury Accidents Statistics (Time-based)							
Accident Data Set							
Daytime	**Morning Peak (7am-9am)**	# of accidents	133	562	973	331	**1999**
		Accidents/km^2	24.1	7.4	8.2	3.1	**6.5**
		Accident/km arterial	13.2	9.3	7.4	3.2	**6.6**
		Accident/km road	1.1	0.4	0.4	0.2	**0.3**
	Noon off-peak (10am-3pm)	# of accidents	434	1901	2707	816	**5858**
		Accidents/km^2	78.7	25	23	7.6	**19.2**
		Accident/km arterial	43	31.6	20.5	8	**19.2**
		Accident/km road	3.6	1.3	1	0.4	**0.9**
	Evening Peak (4pm-8pm)	# of accidents	289	1672	2604	813	**5378**
		Accidents/km^2	52.4	22	22.2	7.6	**17.6**
		Accident/km arterial	28.6	27.8	19.7	7.9	**17.6**
		Accident/km road	2.4	1.1	1	0.4	**0.9**
	Nighttime (9pm-6am)	# of accidents	270	1268	1877	499	**3914**
		Accidents/km^2	49	16.7	16	4.6	**12.8**
		Accident/km arterial	26.7	21.1	14.2	4.9	**12.8**
		Accident/km road	2.2	0.9	0.7	0.2	**0.6**

CHAPTER 5

IDENTIFICATION OF TRAFFIC ACCIDENT HOT SPOTS IN ANKARA

The main aim of this chapter is to identify urban accident hot spot with the NNH clustering method in the urban area of Ankara. As mentioned in the earlier chapter, fatality and injury accidents are analyzed separately while identifying hot spots. For injury accident hot spots, different types and different time periods are taken into consideration. Besides the hot spot calculation with NNH clustering algorithm, KDE, most widely used method in traffic safety, also calculated to compare with the result of NNH clustering method.

5.1. NNH Clustering Method for Fatality Accidents

In order to identify fatality traffic accident hot spots, threshold distance of 1 km that is used to calculate black spot in highway traffic safety as well, is used. Since the number of fatality accidents is few in terms of numbers, n_{min} is taken as 3 for 3-yr accident data. According to the calculation result, total numbers of 15 hot spots are found. These hot spots are mostly located on high speed corridors and main arterials. As can be seen from the Figure 5.1, some of them are found at the intersections.

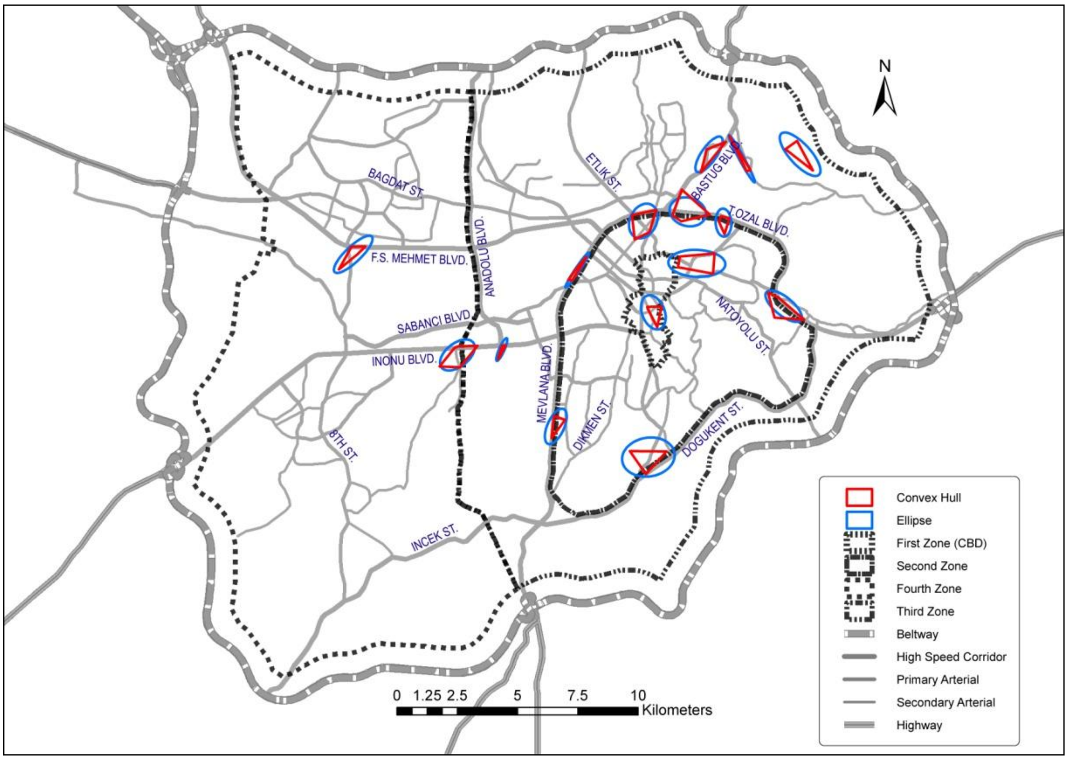

Figure 5.1: Fatality accident hot spots in study area

Severities of hot spot are change between 3 and 8. In the threshold distance of 1 km, maximum number of 8 accidents constitutes hot spot. There are two hot spots are found with the severity of 8 accidents; both are located on the high speed corridor namely Turgut Özal Boulevard. Also there are two hot spots calculated which have 7 accidents. They are located on high speed corridors as well. One of them is on the İnönü Boulevard and the other is on the Turgut Özal Boulevard. Hot spots, severities with 6 fatality accidents are located on the high speed corridors; one is on the Mevlana Boulevard and the other is on the İrfan Baştuğ Boulevard. There are also two hot spots constituted from 5 accidents, both are located on primary arterials in urban area. Besides these, there are 7 hot spots are calculated with the severities of 4 accidents. Out of 7 hot spots, 5 of them are found on the high speed corridor and other two are located on the primary arterials. All these fatality hot spots with respect to their severity values are also visualized with thematic mapping techniques which can be seen from the Figure

5.2. With the help of the thematic map, traffic safety officials can give priority to the areas where hot spot severities with high values are observed.

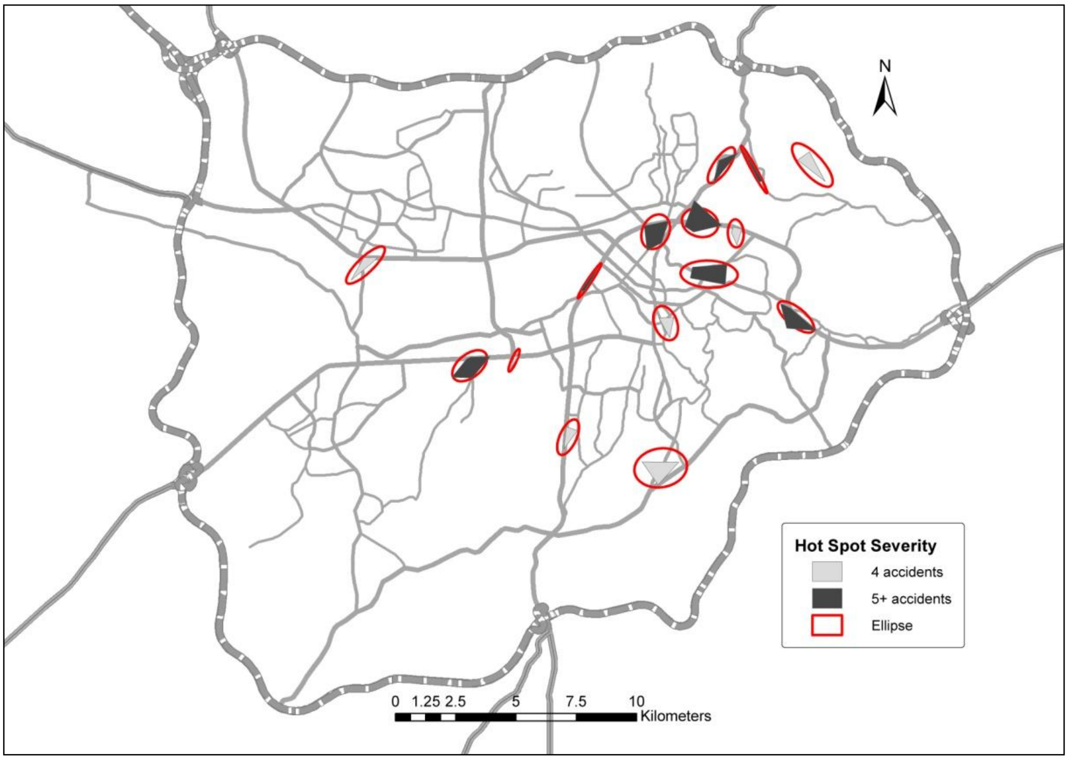

Figure 5.2: Thematic map of fatality accident hot spots with regards to hot spot severity

5.2. NNH Clustering Method for Injury Accidents

Injury accident hot spot detection procedure is started with the threshold distance of 1 km which is used in fatality accident hot spot analysis methods as well. Taken into consideration the high number of injury accidents, n_{min} that can be found in cluster is selected as 20. While visualizing the result of clustering method, thematic mapping technique is used with regards to the hot spots' severity values. As it can be seen from the Figure 5.3, NHH clustering method performing with given criteria give very big clusters which cover more than one roads and intersection. In order to relate them with urban built environment components, identification of smaller areas is more appropriate and meaningful.

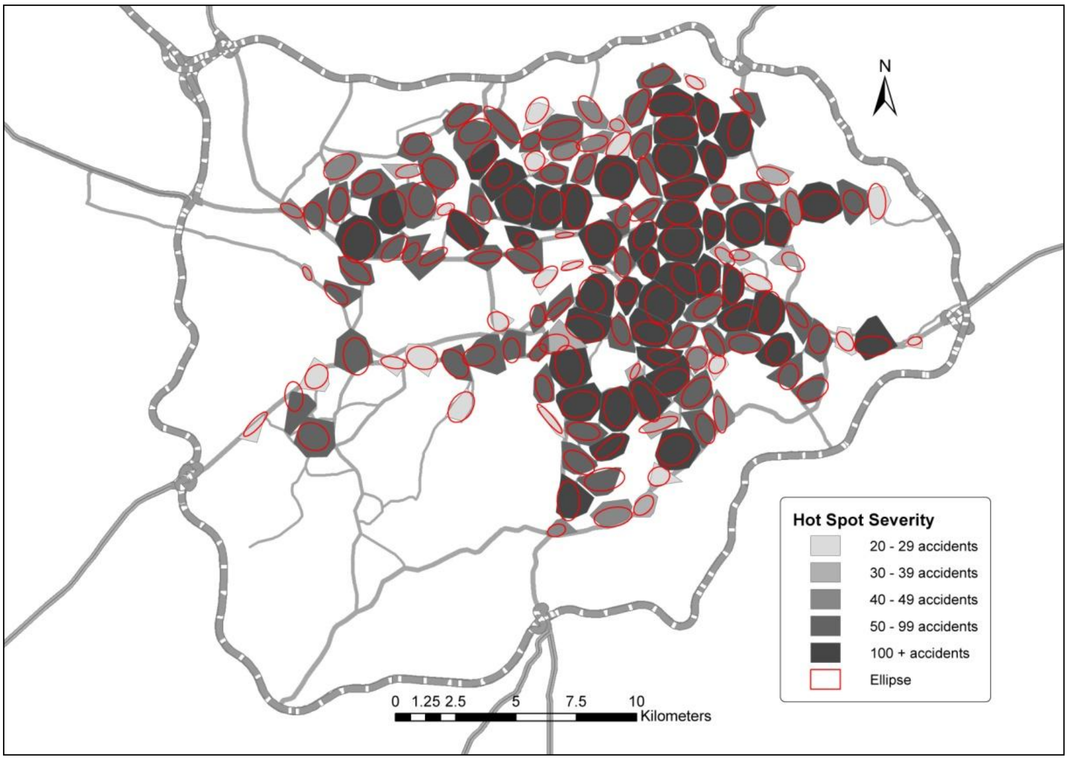

Figure 5.3: Thematic map of fatality accident hot spots with regards to hot spot severity (threshold distance=1 km, n_{min}=20)

While studying with injury accident to detect hot spots in urban area, it is important to find appropriate threshold distance. Since the main aim is to understand the characteristics of injury accident hot spot with regards to urban built environment facilities, size of hot spot is critical. Firstly, threshold distance of 1000 meters is calculated as used in fatality accidents with n_{min}=20. As can be seen from the Figure 5.4, hot spots calculated with 1000 meters threshold distance are too big and covers more than one road. For this reason, distance is decreased to 500 meters. Hot spots detected by threshold distance with 500 meters also cover big areas. Accordingly, threshold distance of 200 meters is calculated which gives better results than the previous ones. However, hot spots cover more than one road and bigger than the intersections.

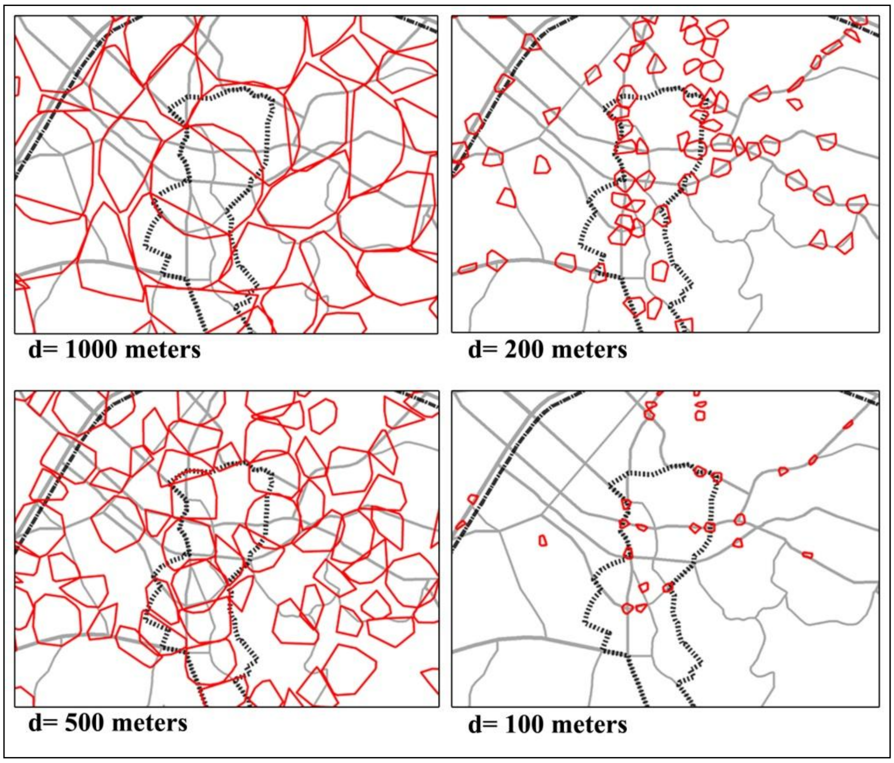

Figure 5.4: Different threshold distances for calculating cluster

As a result, after a trial period to find appropriate threshold value for clustering, 100 meters is found sufficiently to demonstrate traffic accident hot spots in urban region. Choice of threshold limits is important, which basically defines the scope of the clustering analysis. For highways with high speed limits, distance limit of 1 kilometer may be acceptable, but for urban locations; while 100 meters limits, generally speaking, is good to capture hot spots located at a major urban intersections (Figure 5.5). Taking smaller distances, such as 50 meters or 20 meters result in detection of hot spots at the approaching intersection or none, based on the n_{min} criterion.

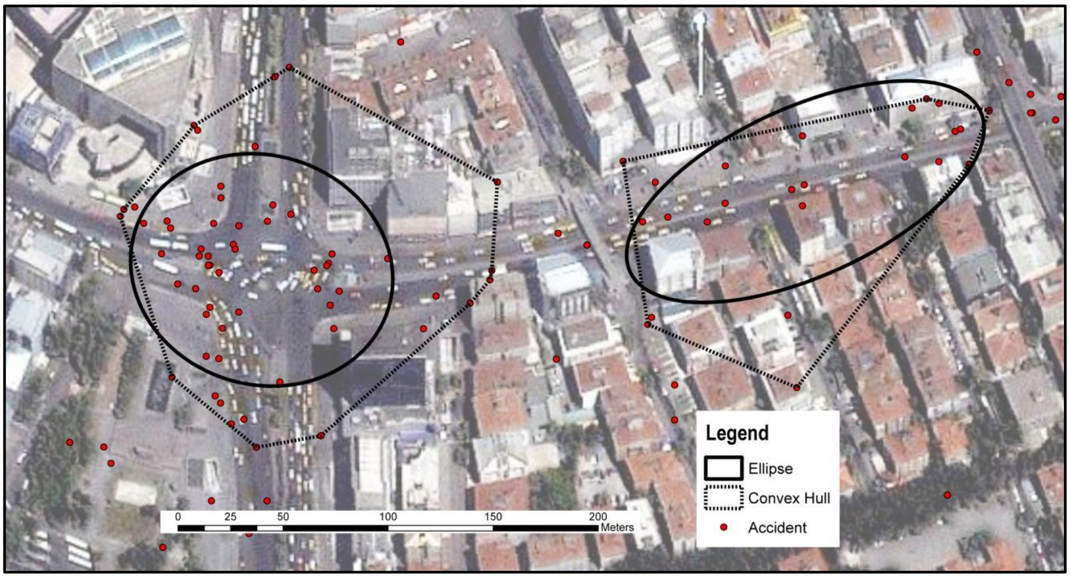

Figure 5.5: Two hot spots located over and urban intersection and a corridor in Kızılay Square (d=100 meters)

In CrimeStat, a NNH clustering analysis for all the injury accident data using a threshold distance of 100m and a minimum of 20 accidents in a cluster detected 64 hot spots (Figure 5.6). Most of them were found on or near the urban main arterials, in the CBD and the second zone supporting the fact that hot spots are in relation to the land use and transportation characteristics of a region.

NNH clustering analyses results' map performed with all injury accidents can be shown in the Figure 5.6. As can be seen from the figure, many of the hot spot are located on the major arterials like İrfan Baştuğ, Turgut Özal, Mevlana Boulevard. Besides this, some hot spots are located on the minor arterials. In outer urban zones, fewer hot spots are found compared to the center of the urban area.

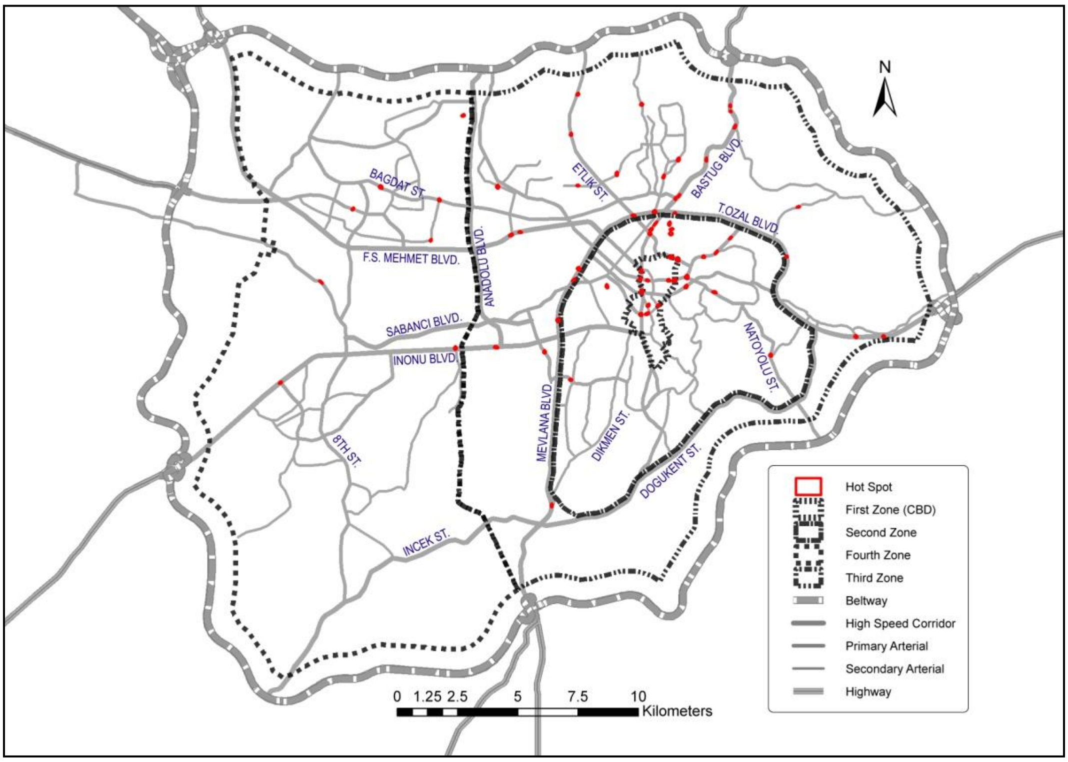

Figure 5.6: Distribution of hot spots considering all injury accidents (n_{min} =20)

Thematic mapping of NNH clustering analysis result performed with a threshold distance of 100m and a minimum of 20 accidents can be seen in the Figure 5.7. Like previous analyses results, hot spot severity is used while performing thematic mapping. Since hot spots are small, only the center of urban area is seen in the figure to show severity of hot spots. Thematic mapping shows clearly the more severe hot spots. It is useful traffic safety officials to determine locations of more severe areas identified as hot spot.

Since it is difficult to show thematic maps based on hot spot severity over all study area clearly, other analyses results are visualized without showing the hot spots' severities. However, hot spot severity is used in the distance analysis of hot spots to the nearest intersection. For this reason, in the following NNH clustering analysis maps, only the locations of hot spots will be shown.

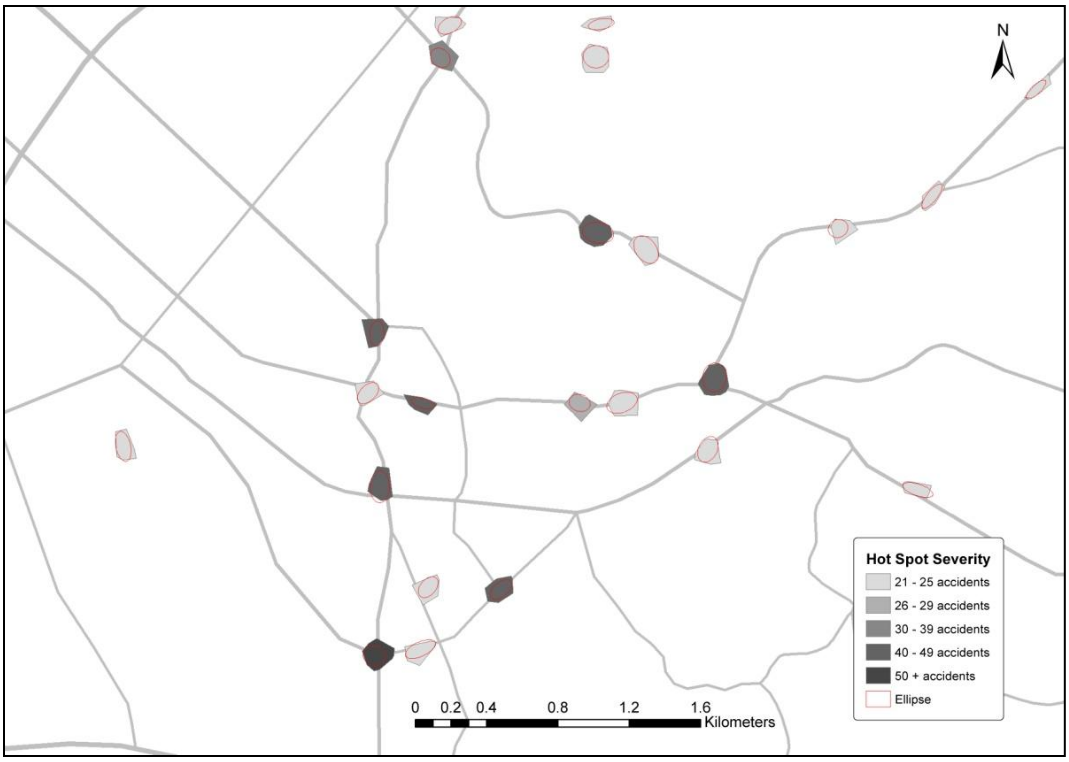

Figure 5.7: Thematic map of injury accident hot spots with regards to hot spot severity (n_{min} =20)

NHH clustering results performed with threshold distance of 100 meters and different n_{min} values (20, 10 and 5) are summarized in the Table 5.1. According to this, analysis result performed with n_{min}=20 found total numbers of 64 hot spots in the study area, out of 12 are located in the CBD. In the second zone, there are total numbers of 15 hot spots are calculated, on the other hand there are 29 hot spots are identified in the third zone. Finally, 8 hot spots are detected in the fourth zone.

Table 5.1: Number of hot spots in urban zones obtained by all accidents

# of Hot spots Location in Urban Zones						
Accident Data Set	**n_{min}**	**CBD**	**2^{nd} Zone**	**3^{rd} Zone**	**4^{th}Zone**	**TOTAL**
All	20	12(%18.8)	15(%23.4)	29(%45.3)	8(%12.5)	**64(%100)**
All	10	42(%12.2)	102(%29.7)	169(%49.3)	30(%8.8)	**343(%100)**
All	5	85(%8.9)	303(%31.6)	460(%47.9)	111(%11.6)	**959(%100)**

NNH clustering method performed with n_{min}=10 found total number of 343 hot spots in the study area. Out of 343 hot spots, 42 of them are located within the boundary of CBD, which are the %12.2 of all hot spots. In the second zone there are 102 hot spots. Most of the hot spots are found in the third zone, where major arterials and high speed corridors are located. According to this, there are 169 hot spots, which are nearly the half of the total number, are detected in this zone. On the other hand, there are total numbers of 20 hot spots calculated in the fourth zone, which have a percentage of 8.8 of all hot spots (Figure 5.8a).

NNH clustering method performed with n_{min}=5 found 959 hot spot in all study area; 85 of them are located within the boundary of CBD. In the second zone, there are total numbers of 303 hot spots are found. Like previous results performed with different n_{min} values, nearly half of the total numbers of hot spots are found in the third zone. There are 460 hot spots are identified in this zone. Out of 959 hot spots, 111 of them are found in the fourth zone (Figure 5.8b).

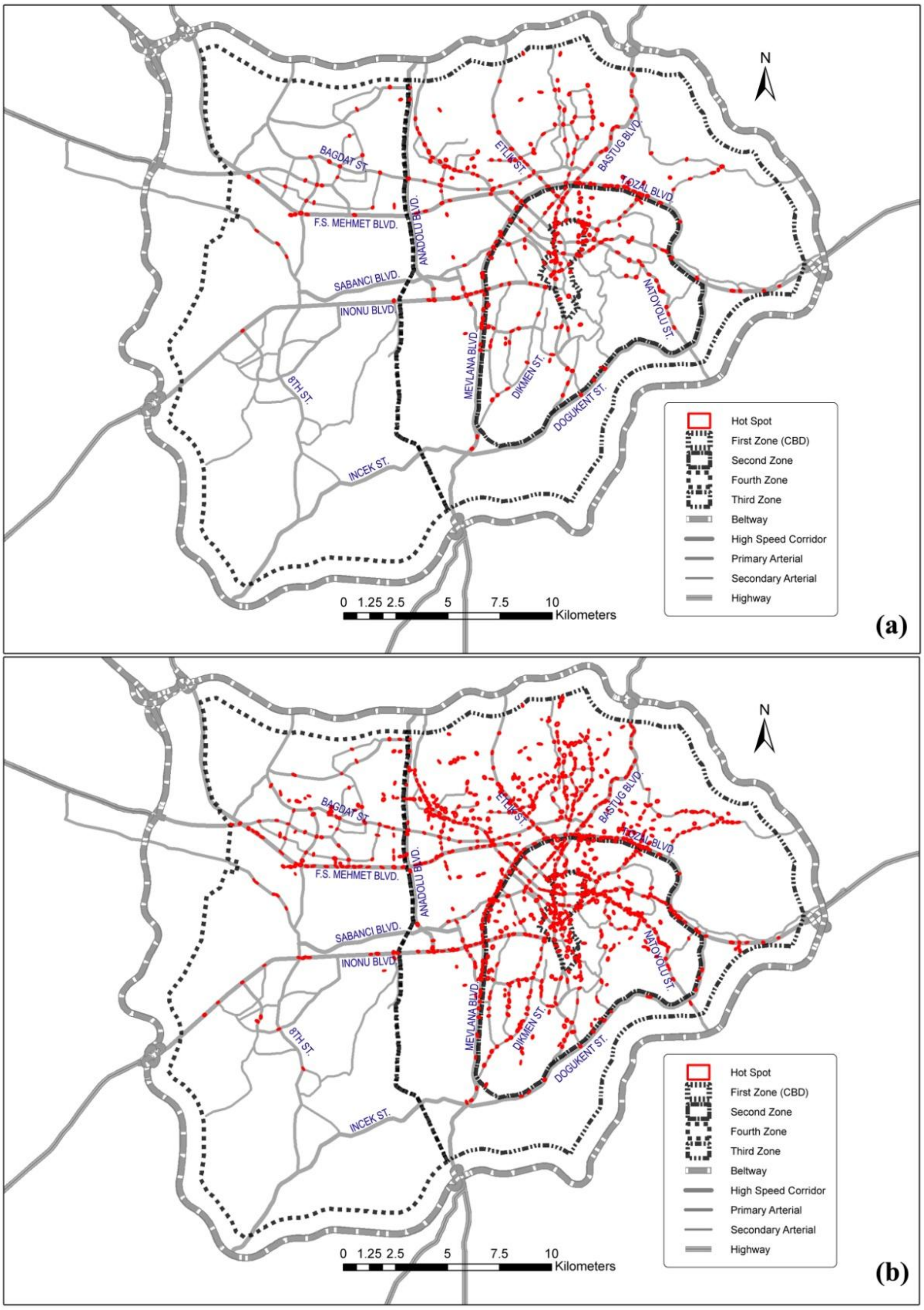

Figure 5.8: Distribution of hot spots considering all injury accidents with a) n_{min}=10 and b) n_{min}=5 accidents

Before studying pedestrian and non-pedestrian accidents separately, it is important to choose a proper n_{min} limit, as the active number of accidents for clustering would be reduced by categorization. Figure 5.9 displays the number of pedestrian and non-pedestrian accidents found in 64 hot spots. The distribution showed that majority of the hot spots had a total of 20-30 accidents in them; where more pedestrian accidents were clustered in the hot spots in the CBD and the second zone, and more non-pedestrian accidents were clustered in the third and the fourth zones. There were few clusters with more than 30 accidents, which also shows the appropriateness of the n_{min}= 20 for Ankara study. Seeing the high number of clusters with 10 pedestrian (or non-pedestrian) hot spots, n_{min}=10 is assumed for the separate analyses of these accidents.

Similarly, to choose a good n_{min} limit for time-based hot spot detection, the hot spots of all accidents are studied based on the distribution of accidents in four time periods. But, as it is difficult to display four time periods, a graph is constructed as accidents in all 3 daytime periods combined together as one 14-hr period versus 10-hr night period (Figure 5.10). The distribution of the time-based characteristics of the accidents in hot spots showed that that in most of the cases, the ratio of nighttime accidents were quarter or a third of the daytime accidents. Thus, for searching hot spots in 4 time periods, the thresholds were revised as d=100m, and n_{min}= 5 accidents.

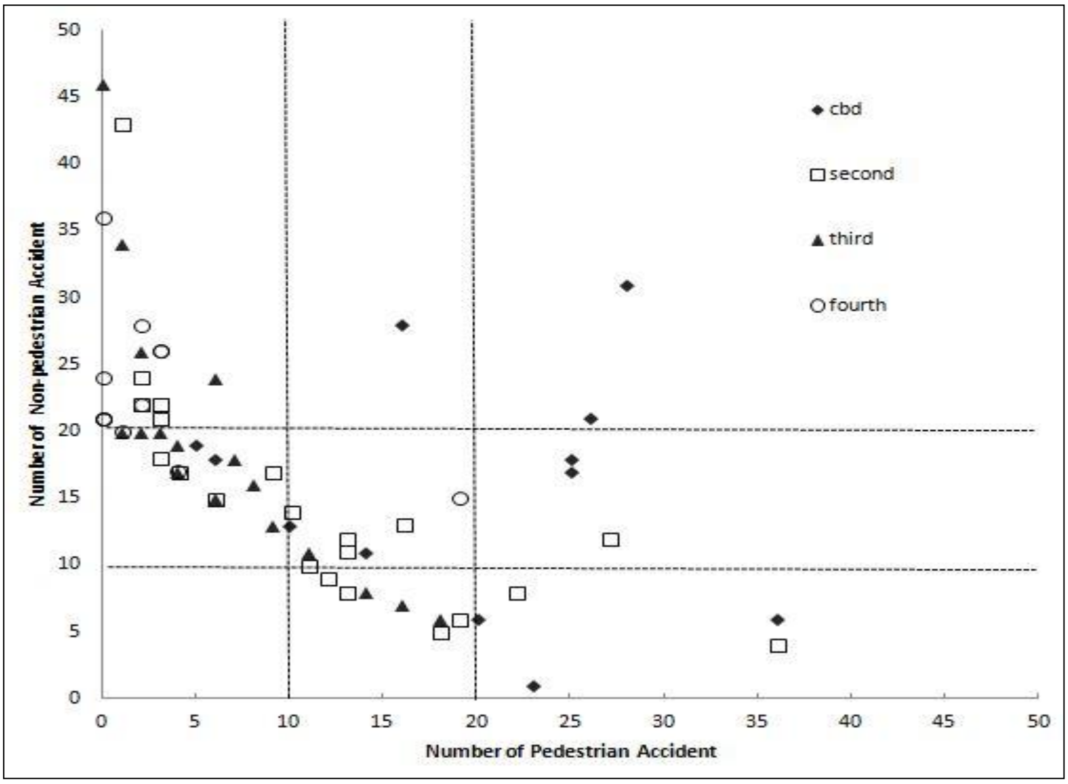

Figure 5.9: Relationship between pedestrian and non-pedestrian accidents for identified hot spots (n_{min}=20)

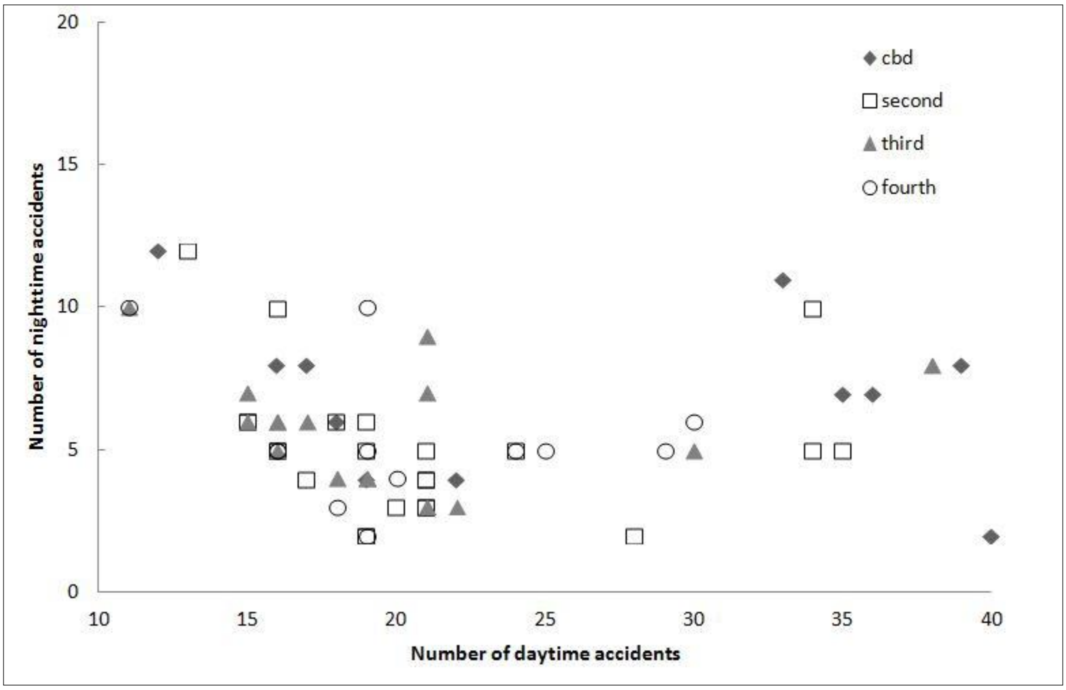

Figure 5.10: Relationship between daytime and nighttime accidents for identified hot spots (n_{min}=20)

5.3. NNH Clustering Method for Injury Accidents by Type and Time

Besides the identification of traffic accident hot spots obtained by all injury accidents without grouping, it is meaningful to analyze hot spots according to type and time occurrence. Since causes of different type or time of accident may differ, interventions for different type or time of accidents may be different as well. For this reason, accident should be

5.3.1. Hot Spots for Pedestrian versus Non-Pedestrian Accidents

With the same selection criteria (n_{min}=20 and d=100m), there were only 9 pedestrian accident hot spots, 6 of which were in the CBD and 3 in the first zone, whereas 22 non-pedestrian accident hot spots were scattered among all the zones (Table 5.2). But forcing minimum of 20 accidents in a subset of the data points is very restrictive, as discussed above. Thus, reducing the n_{min} to 10 in the clustering (with d=100m), more pedestrian (and non-pedestrian) accident hot spots were located in the urban zones (Table 5.2). Reducing the n_{min} 5 in clustering for the analyses of pedestrian and non-pedestrian hot spots, more hot spots were calculated as expected. According to this, total numbers of 262 hot spots detected for pedestrian accidents, on the other hand, there are 509 hot spots derived from the non-pedestrian accidents.

Table 5.2: Number of hot spots in urban zones obtained by pedestrian and non-pedestrian accidents

# of Hot spots Location in Urban Zones						
Accident Data Set	**n_{min}**	**CBD**	**2^{nd} Zone**	**3^{rd} Zone**	**4^{th}Zone**	**TOTAL**
Pedestrian	20	6(%66.7)	2(%22.2)	1(%11.1)	0	**9(%100)**
	10	21(%29.6)	28(%39.4)	21(%29.6)	1(%1.4)	**71(%100)**
	5	53(%20.2)	94(%35.9)	113(%43.2)	2(%0.7)	**262(%100)**
Non-Pedestrian	20	3(%13.6)	2(%9.1)	11(%50.0)	6(%27.3)	**22(%100)**
	10	11(%8.0)	30(%21.9)	73(%53.3)	23(%16.8)	**137(%100)**
	5	29(%5.7)	126(%24.7)	262(%51.5)	92(%18.1)	**509(%100)**

As can be seen from the Figure 5.11a and 5.12a, pedestrian accident hot spots are mostly located in the first and second zones since these zones attracts more pedestrian in terms of activities such as commerce, health, business, government, etc. In the third zone, pedestrian accident hot spots are seen in especially in Keçiören where residential usages are seen. On the Etlik Street, also pedestrian accident hot spots can be seen. In the fourth zone, there is only one pedestrian accident hot spot where Batıkent metro station is located.

Non-pedestrian accident hot spots are mostly located on the major and minor arterials. It can be seen from the figure that, hot spots are found on İnönü, Turgut Özal, İrfan Baştuğ and Mevlana Boulevard which are the old interstate ring roads of Ankara. These ring roads have more speed limits compared to the other arterials. Also, in the city center, traffic flow is slower compared to these old interstate ring roads, as a result in the first and second zones less non-pedestrian accidents' hot spots are found. Compared to the pedestrian accident hot spots, non-pedestrian accidents' are more dispersed over the study area (Figure 5.11b and 5.12b).

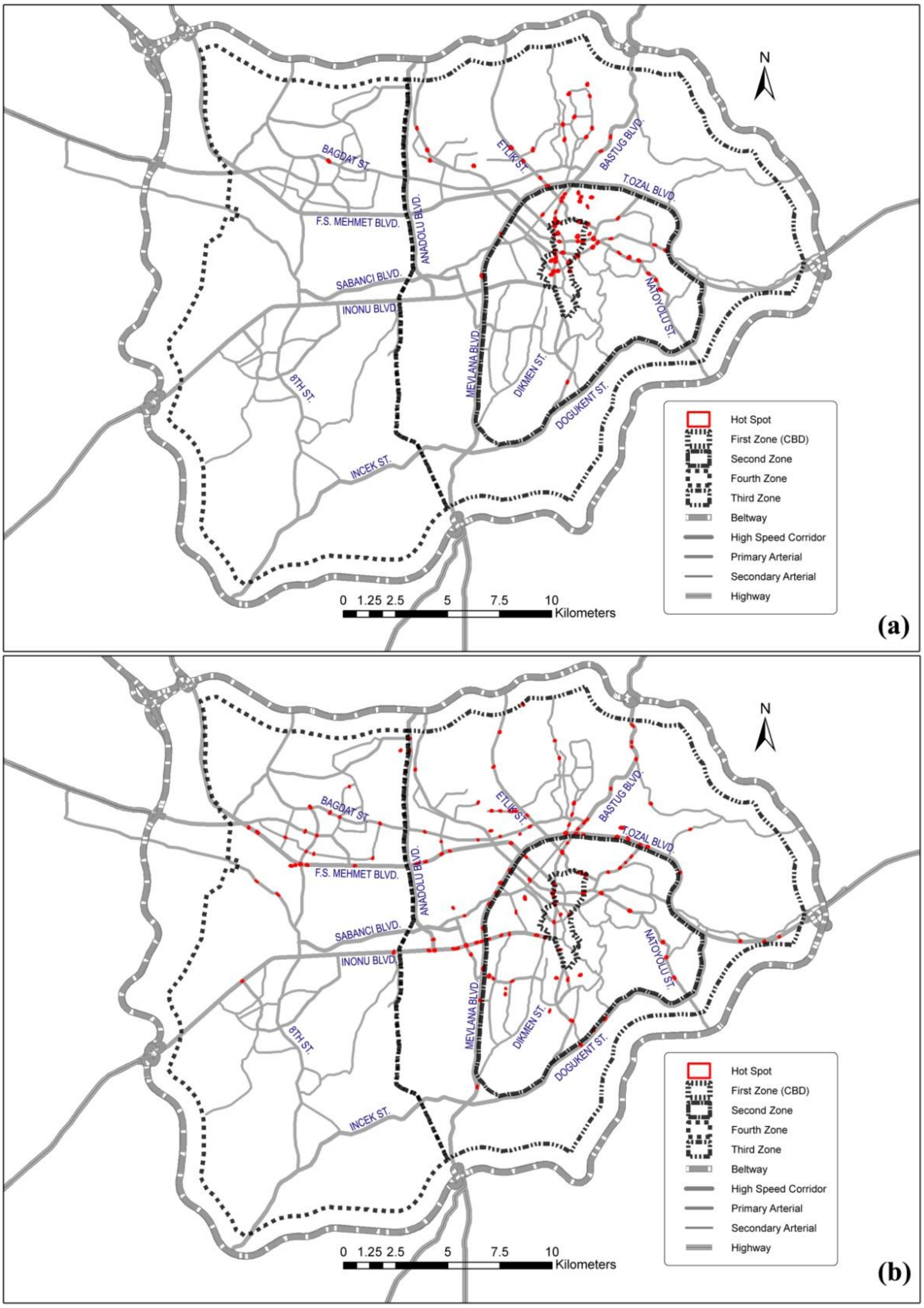

Figure 5.11: Hot spots of injury accidents for a) Pedestrian and b) Non-pedestrian injury accident (n_{min}=10)

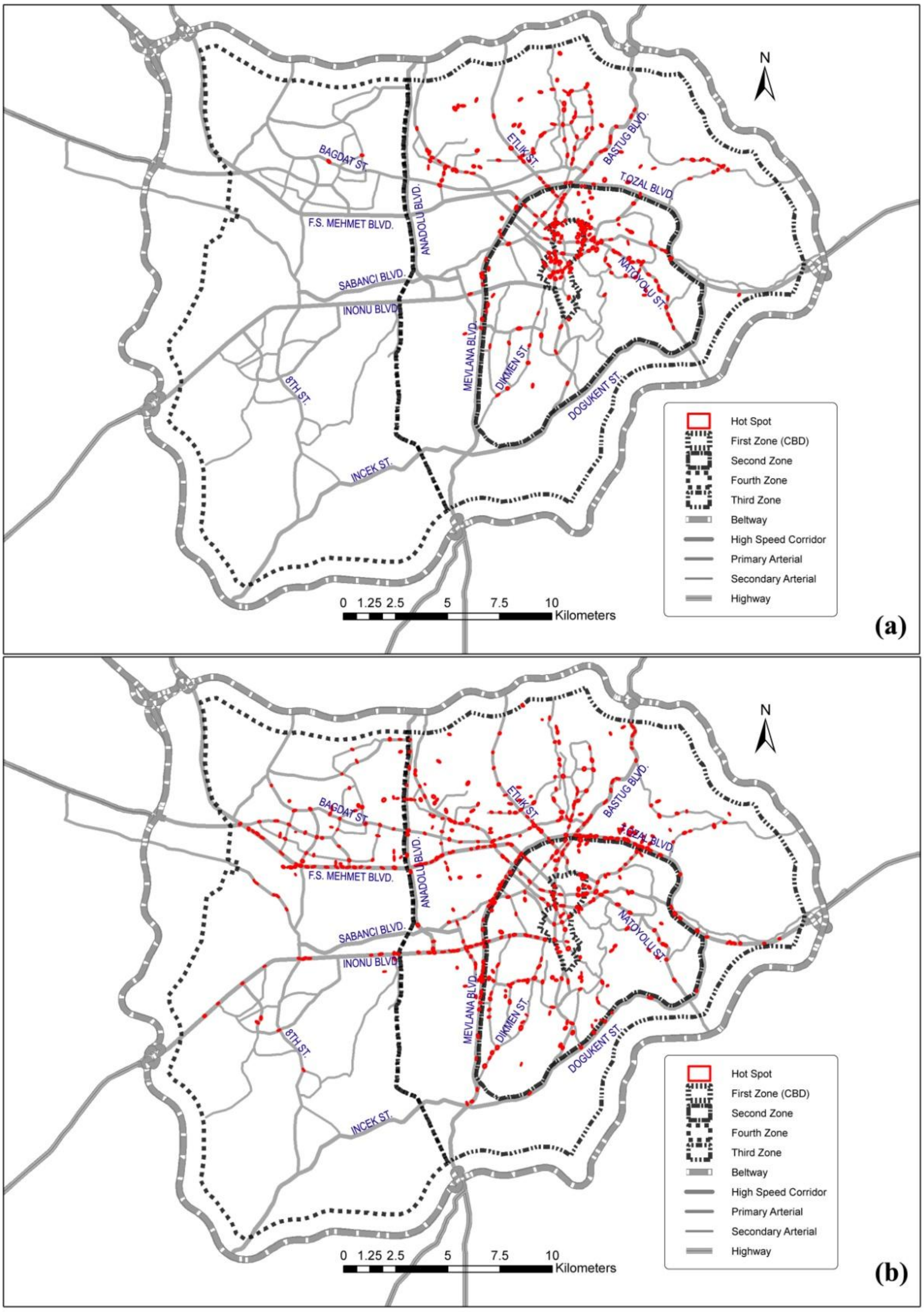

Figure 5.12: Hot spots of injury accidents for a) Pedestrian and b) Non-pedestrian injury accident (n_{min}=5)

5.3.2. Hot Spots for different Time Periods

Time-based hot spot locations also dispersed over the study area. For all time periods, hot spots are gradually decreased from CBD to outer urban zones. Morning peak hot spots (Figure 5.13a) are few in numbers compared to other time periods and they are located mostly in CBD. Evening peak hot spots (Figure 5.13b) are also few compared to the noon-off peak. Noon off-peak hot spots (Figure 5.14a) are mostly located in major and minor arterials especially in the first, second and third zones. Nighttime accident hot spots are (Figure 5.14b) located also on the major and minor arterials. Number of hot spots over identified urban zones is summarized in Table 5.3. According to the results, for more hot spots are calculated for noon off-peak hours. Compared to the other time periods, fewer hot spots are detected from morning peak accidents.

Table 5.3: Number of hot spots in urban zones obtained by time based data

# of Hot spots Location in Urban Zones						
Accident Data Set	**n_{min}**	**CBD**	**2nd Zone**	**3rd Zone**	**4thZone**	**TOTAL**
Morning Peak	5	5(%31.3)	5(%31.3)	5(%31.3)	1(%6.3)	**16(%100)**
Noon off-peak	5	26(%13.8)	61(%32.4)	90(%47.9)	11(%5.9)	**188(%100)**
Evening Peak	5	15(%11.6)	39(%30.2)	62(%48.1)	13(%10.1)	**129(%100)**
Nighttime	5	17(%16.8)	28(%27.7)	46(%45.5)	10(%9.9)	**101(%100)**

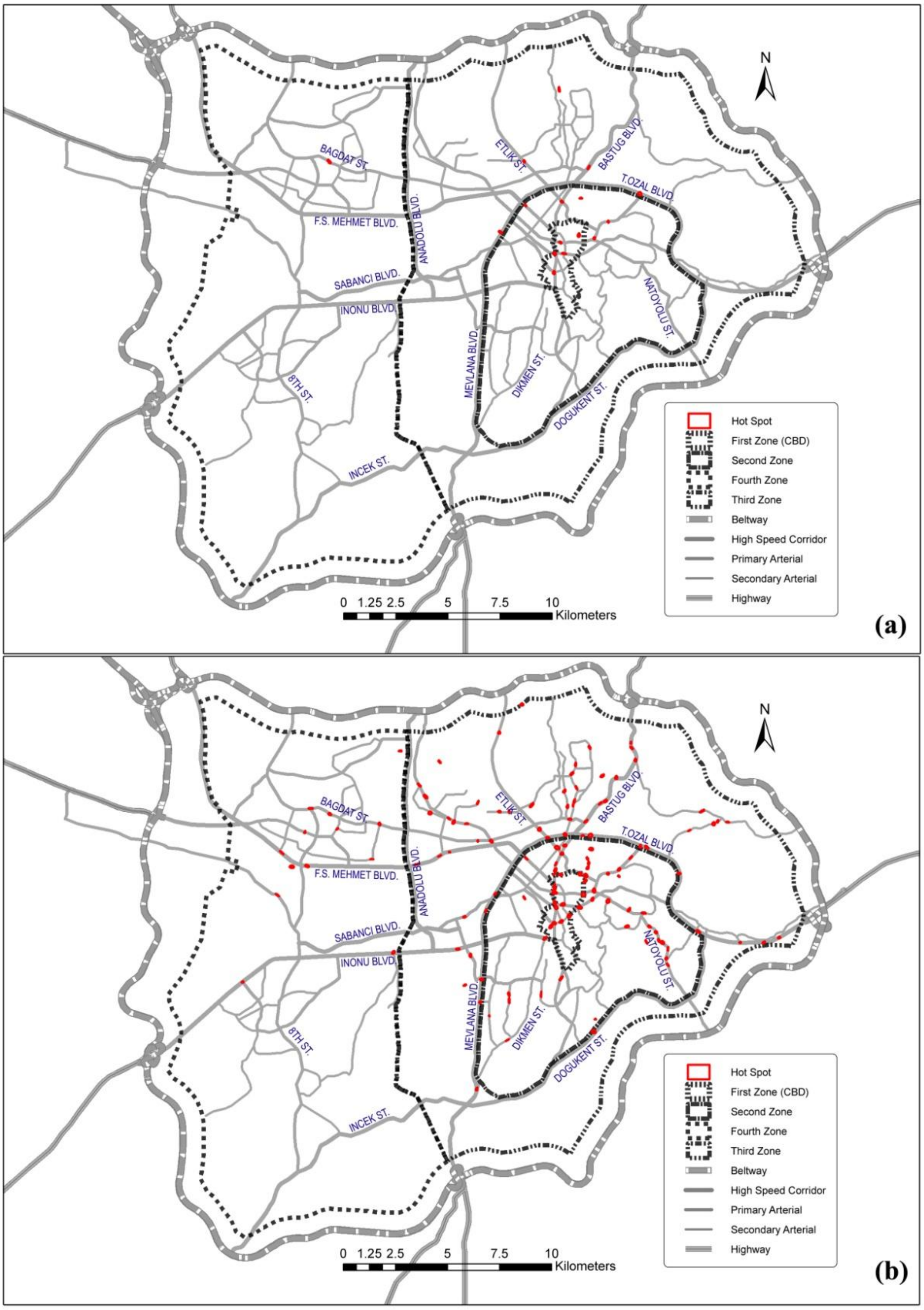

Figure 5.13: Hot spots of injury accidents for a) morning and b) evening peak (n_{min}=5)

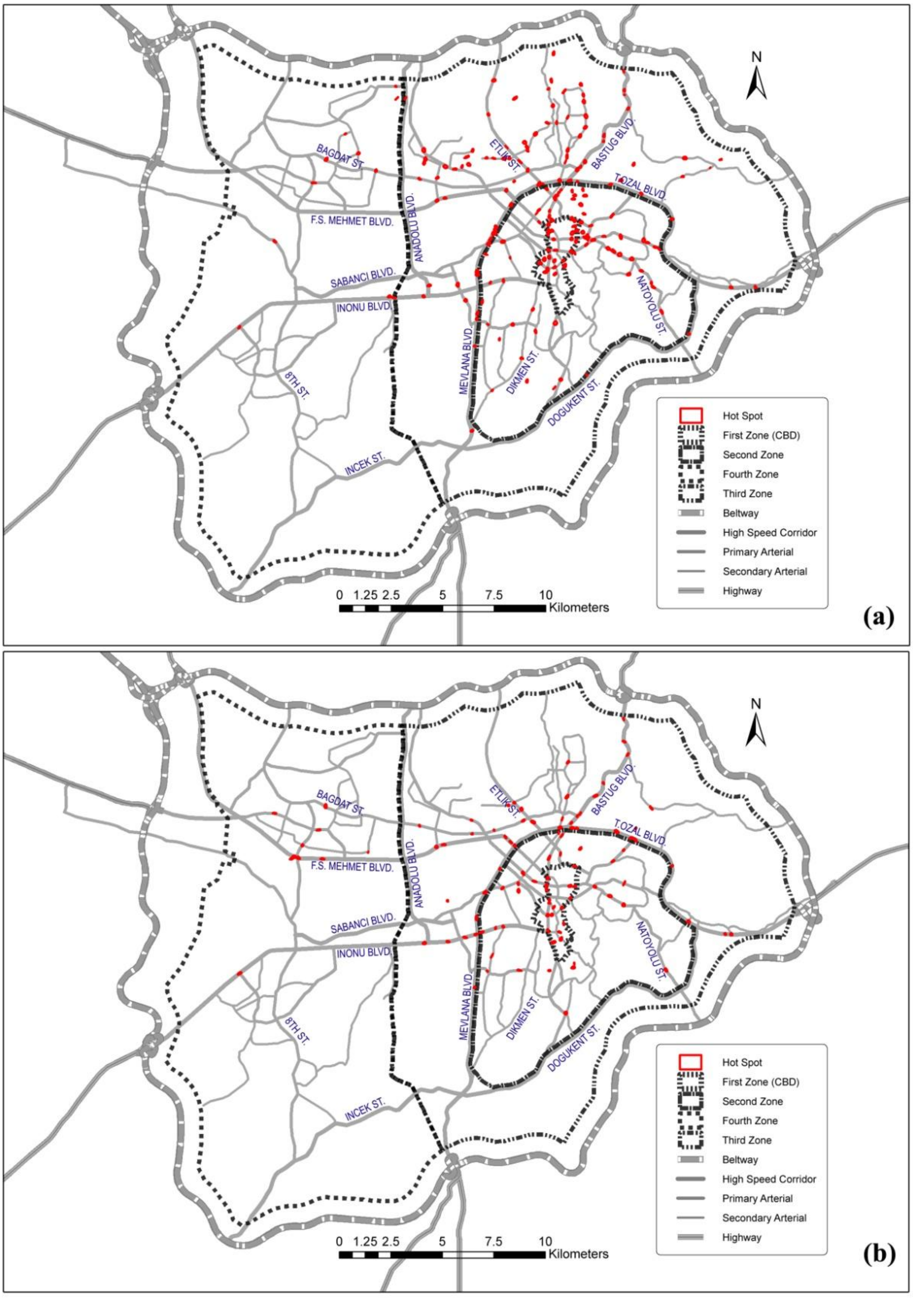

Figure 5.14: Hot spots of injury accidents for a)noon off-peak and b) nighttime (n_{min}=5)

5.4. Hot Spots with KDE

As mentioned in the methodology, cell sizes and bandwidth criteria for calculating KDE maps are derived from the calculation of detected hot spots by NNH clustering method. According to this, total area of convex hulls and maximum distance between points found in hot spot are taken into consideration.

After calculations of areas of convex hulls, cell sizes were found as 115 meters for all accidents' hot spots, 104 meters for pedestrian accidents' hot spots and 95 meters for non-pedestrian accidents' hot spots. Taken into consideration of mean sizes of all three calculated cell sizes, 100 meters is determined as cell size for three types of (all, pedestrian and non-pedestrian) accidents' analyses.

For time based accident analyses, cell sizes were calculated as 74 meters for morning, 79 meters for noon-off, 76 meters for evening and 70 meters for nighttime accidents' hot spots. Mean of these sizes were calculated as 74.75 meters, as a result in time based KDE analyses, cell sizes were determined as 75 meters.

For this study, bandwidth is also selected taking into consideration the NNH clustering results. To do this, maximum distances of incidents found in clusters are measured. According to this, maximum distance between accidents is found as 200 meters. For this reason, for both type of the KDE, 200 meters is used as a search bandwidth.

5.4.1. Normal KDE Results

As mentioned in the literature, bandwidth of normal KDE is the standard deviation of the normal distribution. That's why, smoother result are expected. In normal KDE method for all injury accidents (Figure 5.15), there are a big areas detected as hot spot in urban center. The main underlying reason is that accidents per area are high at these areas. Also, on major arterials, highly dense areas are calculated. These areas are not smooth, in other words they look like as peaks.

Locations of these peaks are mostly indicators of intersections and road connections. Besides these, on the some minor arterials, dense areas are almost observed.

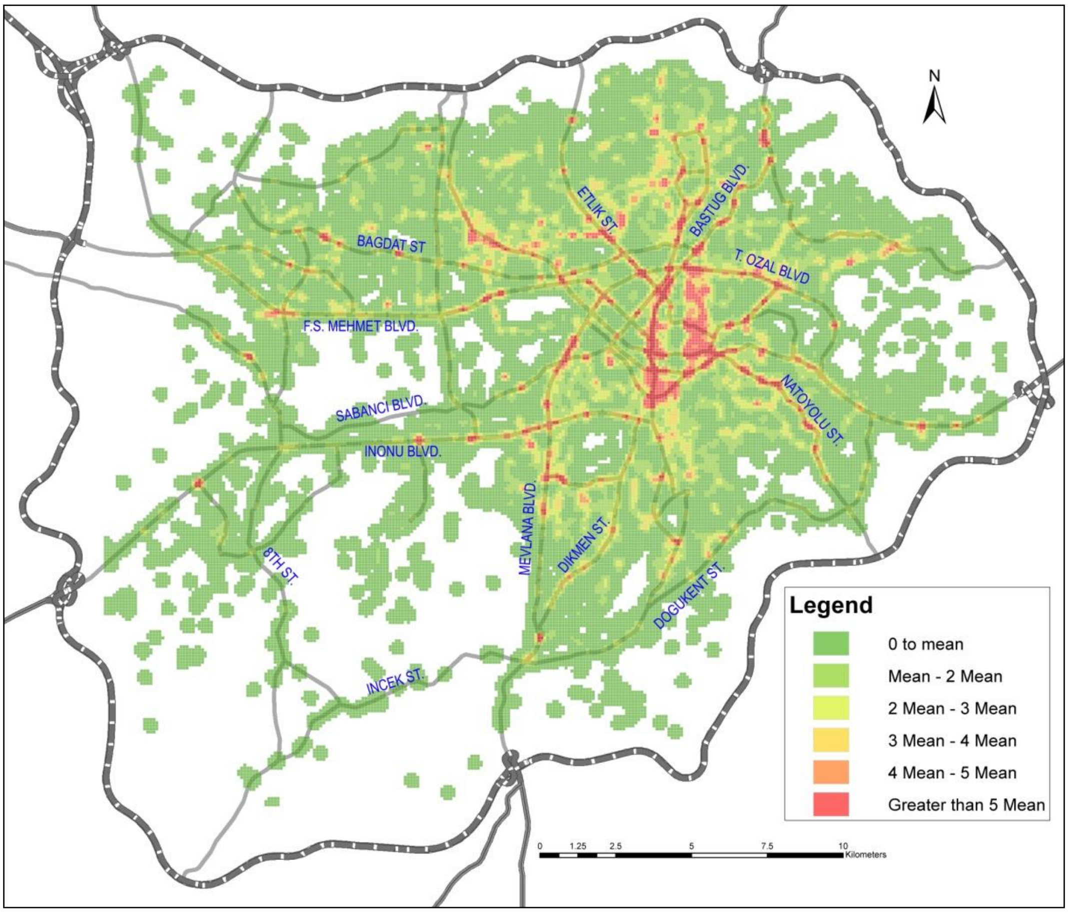

Figure 5.15: Normal KDE maps for all injury accidents

Normal KDE method generated from pedestrian accident is shown in the Figure 5.16a. According to this map, majority of dense areas identified as hot spots are mostly seen in city center. Since pedestrian activities are densely seen in the center, this result is expected. One interesting point is that, nearly no highly dense areas are detected on major arterials. On the other hand, on some minor arterials, dense areas can be seen where pedestrian activities are highly seen due to the land use characteristic like residential, commercial, etc.

KDE maps created from non-pedestrian accidents show different characteristic compared to the pedestrian accident KDE map (Figure 5.16b). According to this map, dense areas are located on the arterials. There are not big areas as observed in all and pedestrian accidents' KDE results. The main underlying reason is that the bigger part of the non-pedestrian accident is occurred on urban arterials. For this reason, it is expected to see highly dense areas over arterials. Besides the continuous dense areas over arterials, some peaks are almost observed. These peaks are mostly the location of intersections and road connections.

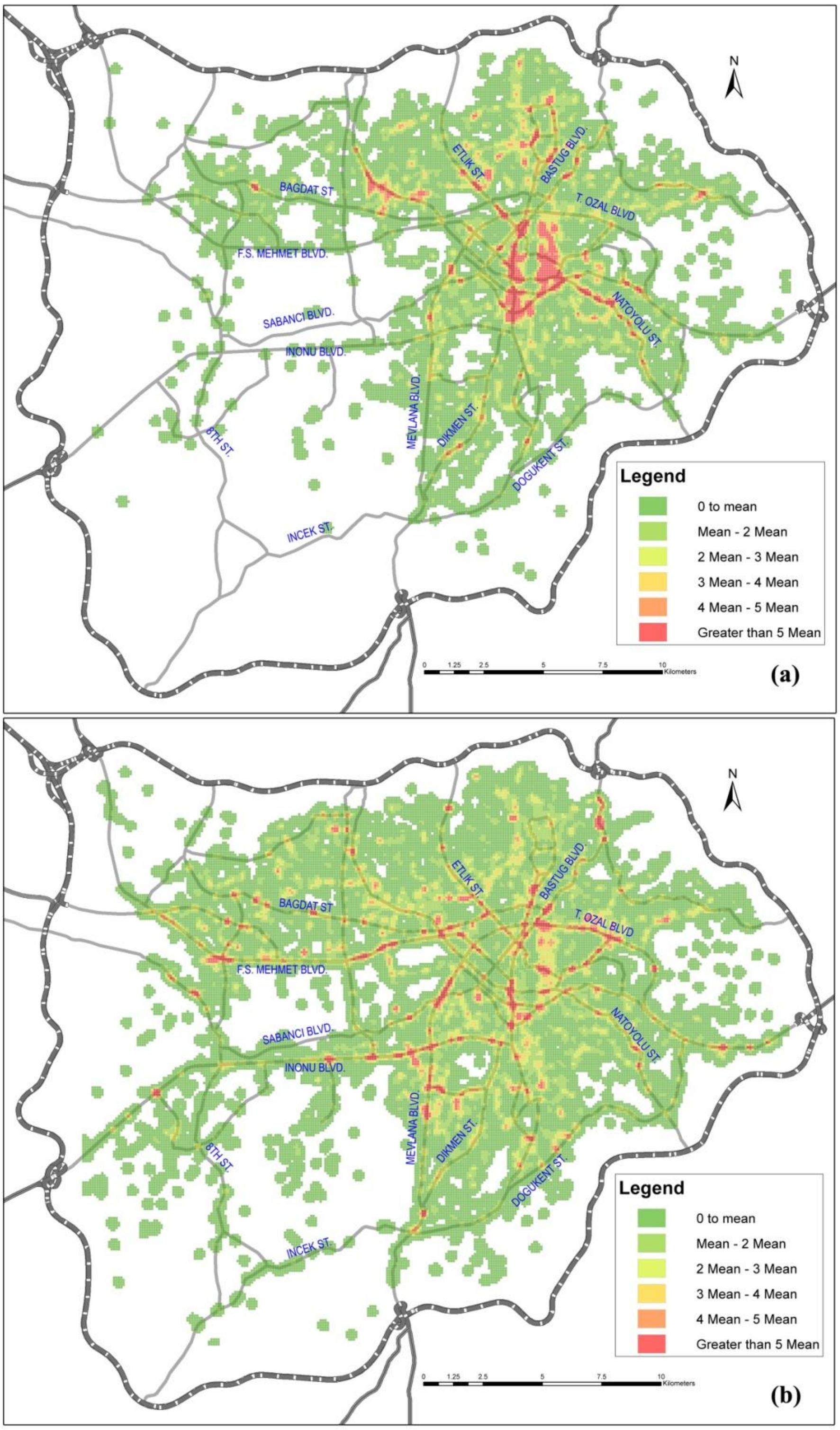

Figure 5.16: Normal KDE map for a)pedestrian b)non-pedestrian injury accidents

Time-based calculated normal KDE maps show similarities. According to maps, for four time periods, there are highly dense areas are calculated in the city center.

Beside city center, on the arterials, there are also densely identified peaks can be observed. Like former KDE results, these small areas observed on the arterials are the location of intersections and connection points.

KDE analysis result obtained from morning peak accidents shows that most of hot spots are located on urban arterials these hot spots are mostly located in or near the intersections (Figure 5.17a). Taking into consideration the distribution of hot spots, most of them are seen in the CBD and second zone. Evening peak accidents' hot spots are also located mostly on the CBD and second zone. Besides these, some hot spots are observed on the urban arterials either on the intersections or road sections (Figure 5.17b). Noon off-peak accident hot spots calculated by normal KDE are mostly located in the CBD and second zone (Figure 5.18a). Comparing with morning and evening peak accident normal KDE results, noon off-peak hot spots cover more area. Like previous findings, areas defined as hot spot are also located on or near the intersections or connection points of local streets. Normal KDE analysis performed with nighttime accidents (Figure 5.18b) also shows same result that most of the hot spots are located in the CBD. Total hot spot area is less compared to the noon-off peak accident hot spots as expected. Other than detected hot spot in CBD, on major arterials and minor arterials especially on and between the intersections, hot spots are also detected.

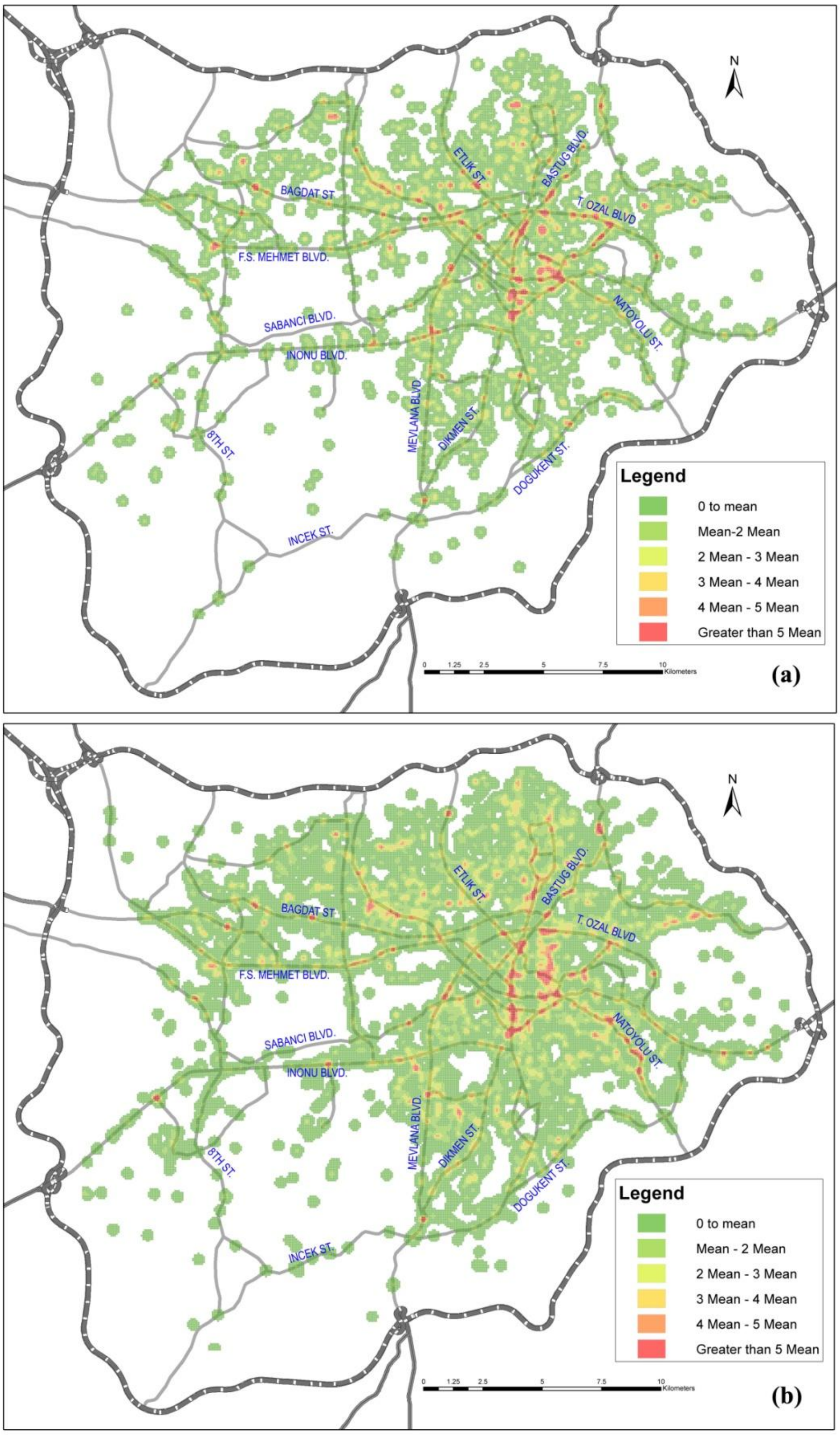

Figure 5.17: Normal KDE map for a)morning b)evening peak injury accidents

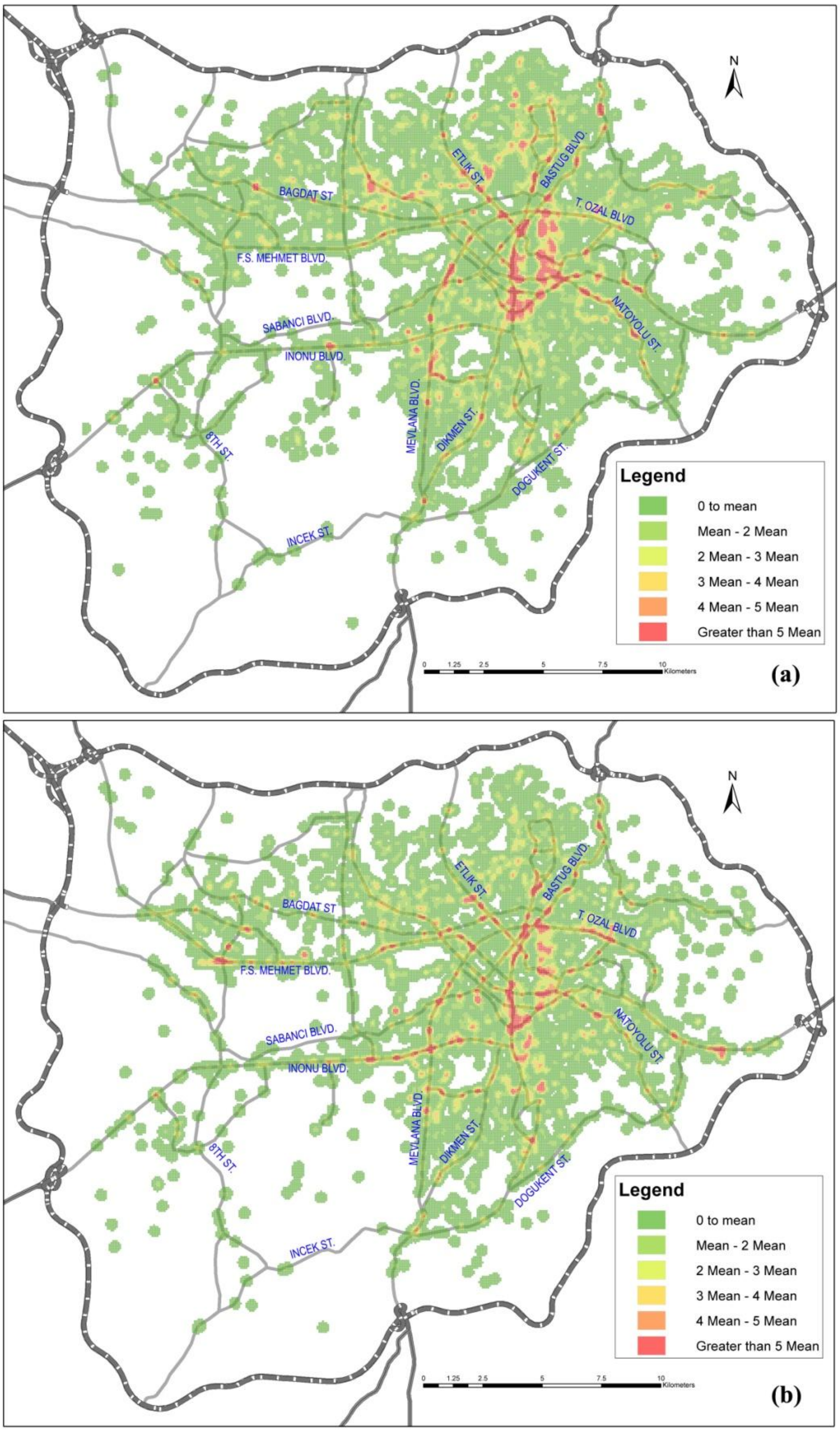

Figure 5.18: Normal KDE map for a)noon off-peak b)nighttime injury accidents

5.4.2. Uniform KDE Results

As mentioned in the literature, uniform KDE function weights all point equally which are in the boundary of search bandwidth. Outside the specified radius, weight is taken as zero. In other words, uniform KDE produce estimate only for the identified bandwidth. For this reason, it is expected that uniform KDE results are more spiky impressions compared to the normal KDE. It is most probably observe the local peaks regarded as hot spot over study region.

Uniform KDE result calculated by all injury accidents can be seen in the Figure 5.19. According to the map, variability between densities can be seen. In the city center, on the arterials, highly dense areas are seen. Along Atatürk Boulevard, dense areas are observed. Besides this, some little peaks are observed on the arterials. Most of them are located in or near the intersections or connection points of local roads to the arterials.

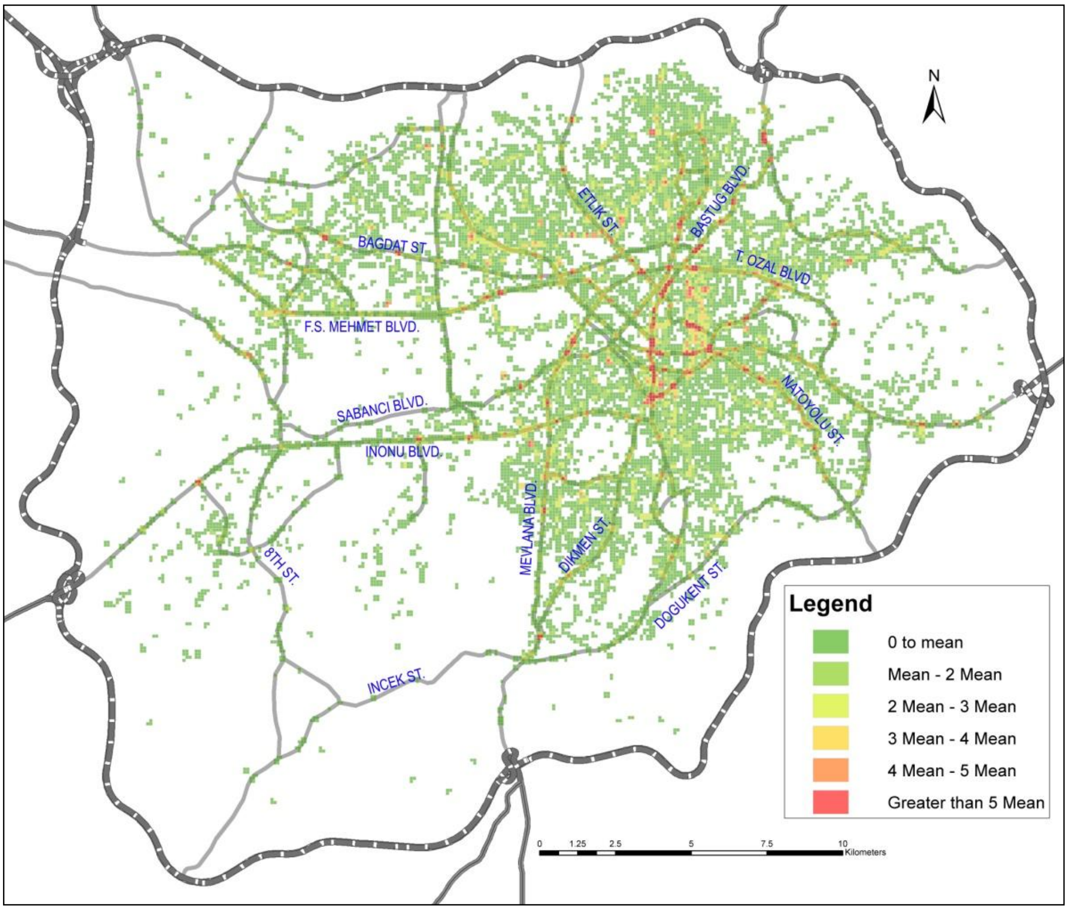

Figure5.19: Uniform KDE map for all injury accidents

According to the result of KDE map produced from pedestrian accident, most of the dense cells are located in the CBD and second zone. Especially along the Atatürk Boulevard, many cells regarded as hot spot are occurred. Since many commercial areas exist on both side of the road, it is most likely to see pedestrian accidents higher than other areas. Beside this, CBD and second zone are the attraction center of pedestrian activities. For this reason, high density areas are expected to see in these zones. In addition to these high density areas observed in the CBD and second zone, some small areas consisted from a few cells located on arterials. Most of them are in or near the intersections (Figure 5.20a).

KDE map obtained from the analysis of non-pedestrian accidents can be seen in the Figure 5.20b. Hot spot areas are mostly observed on the arterials. Different from pedestrian accident hot spots on arterials, non-pedestrian accidents hot spots spread out over the study area. These hot spots are smaller than the pedestrian accident hot spots in terms of size and number of cells aggregated together. More variations can be seen in figure that non-pedestrian accident hot spots are located on arterials. Besides the hot spots in or near the intersections, many hot spots are observed on the section of the arterials.

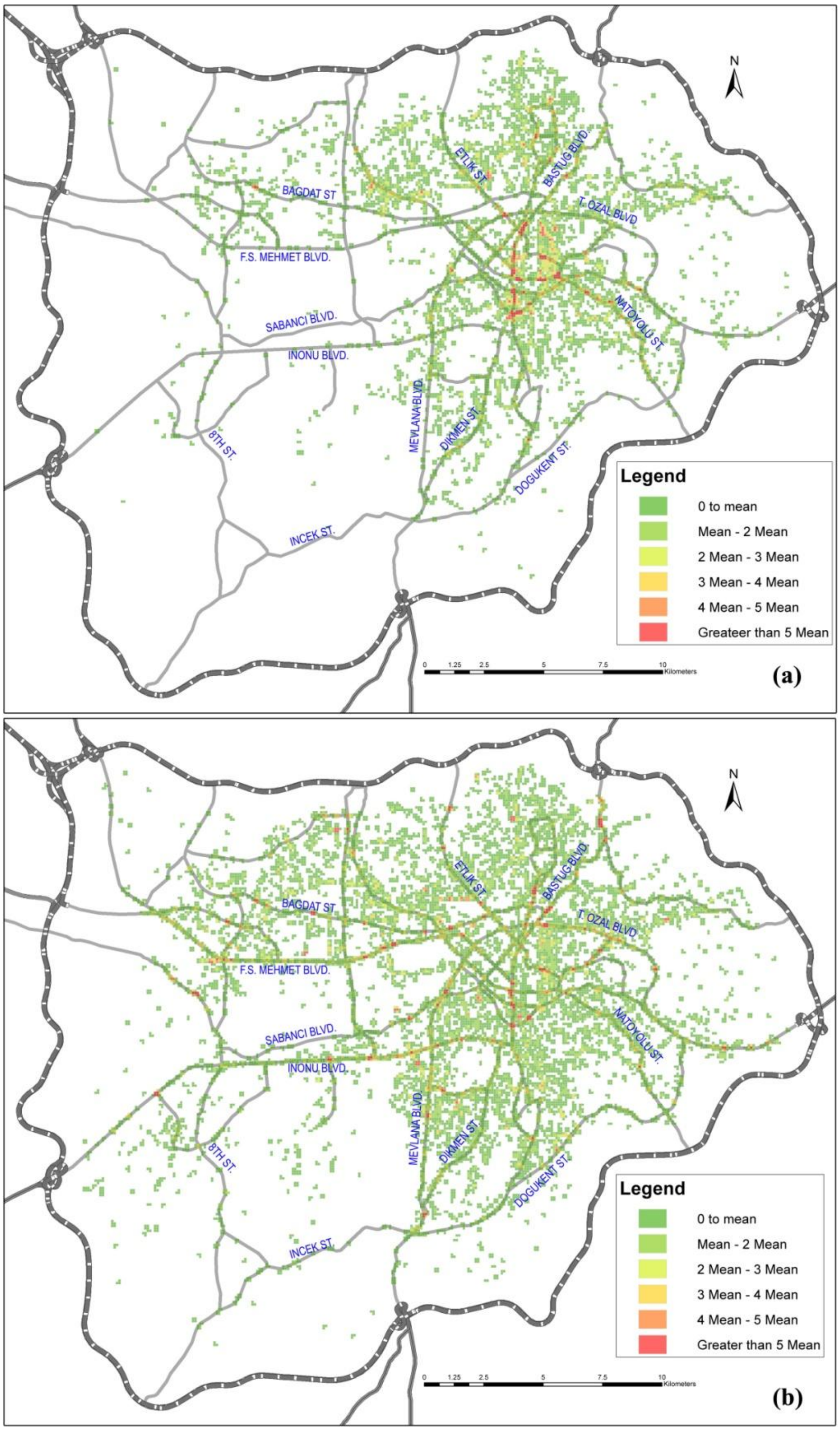

Figure 5.20: Uniform KDE map for a)pedestrian b)non-pedestrian injury accidents

KDE analysis result obtained from morning peak accidents shows that most of hot spots are located on urban arterials (Figure 5.21a). Also it is observed that, these hot spots are mostly located in or near the intersections. Taking into consideration the distribution of hot spots, most of them are seen in CBD and second zone. Since number of accidents observed in morning peak is less than the other time periods, number of hot spots observed in this peak also less than the other time periods. Evening peak accidents' hot spots are also located mostly on the CBD and second zone. Besides these, some hot spots are observed on the urban arterials either on the intersections or road sections (Figure 5.21b). Noon off-peak accident hot spots calculated by uniform KDE are mostly located in the CBD and second zone. Also, there are hot spots detected on the urban arterials. Like previous findings, these hot spots are also located on or near the intersections or connection points of local streets (Figure 5.22a). Uniform KDE analysis performed with nighttime accidents also shows same result that most of the hot spots are located in the CBD. Other than detected hot spot in CBD, on major arterials and minor arterials especially on the intersections, some hot spots are also detected (Figure 5.22b).

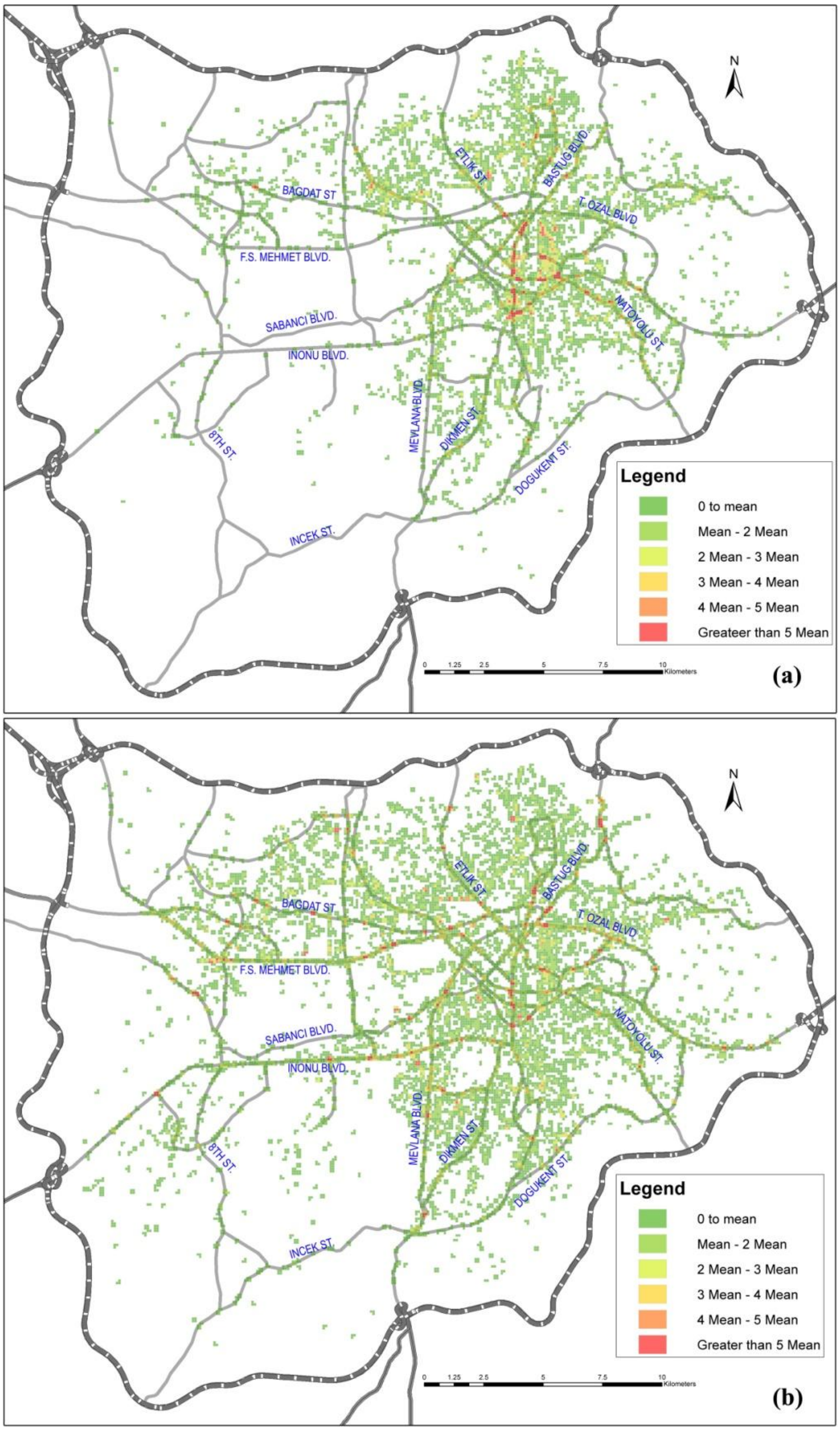

Figure 5.21: Uniform KDE map for a)morning b)evening peak injury accidents

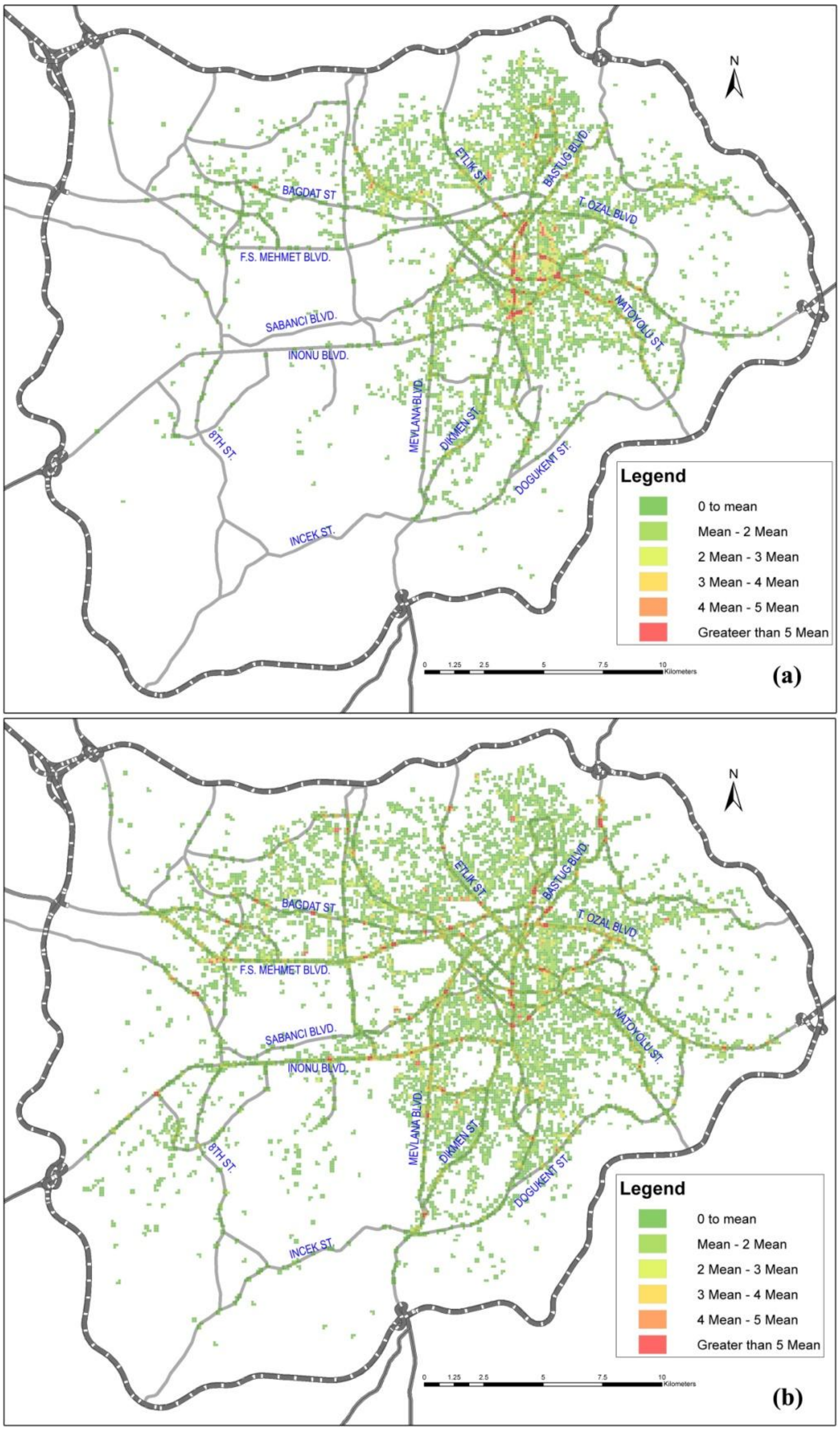

Figure 5.22: Uniform KDE map for a)noon off-peak b)nighttime injury accidents

5.5. Comparison of Hot Spot Methods

After performing different methods, it is important to show which method should be used for evaluation and further analyses of traffic accident hot spots. To do this, both visual and mathematical comparisons are done between results which are created from different groups of data and methods. Firstly, visual comparisons of methods are explained in details and figures. Then, mathematical comparison and numerical results are represented to show NNH clustering algorithm gives more accurate results compared to KDE.

5.5.1. Visual Comparison of Hot Spot Methods

Comparison between the outputs of NNH clustering algorithm visually (Figure 5.23) shows that convex hulls cover exactly all clustered points. In other words, boundary of convex hulls is specified from the points calculated in hot spot. On the other hand, ellipse, which is a symbolic representation of the cluster, is a certain shape formed by force. Ellipse does not have to contain all the points calculated as hot spot. For this reason, convex hulls give the actual area of hot spot. Especially, for the detailed calculations, convex hulls are better to use to get better results. However, ellipses look better and more understandable by user, for this reason, ellipses can be used for visualization step.

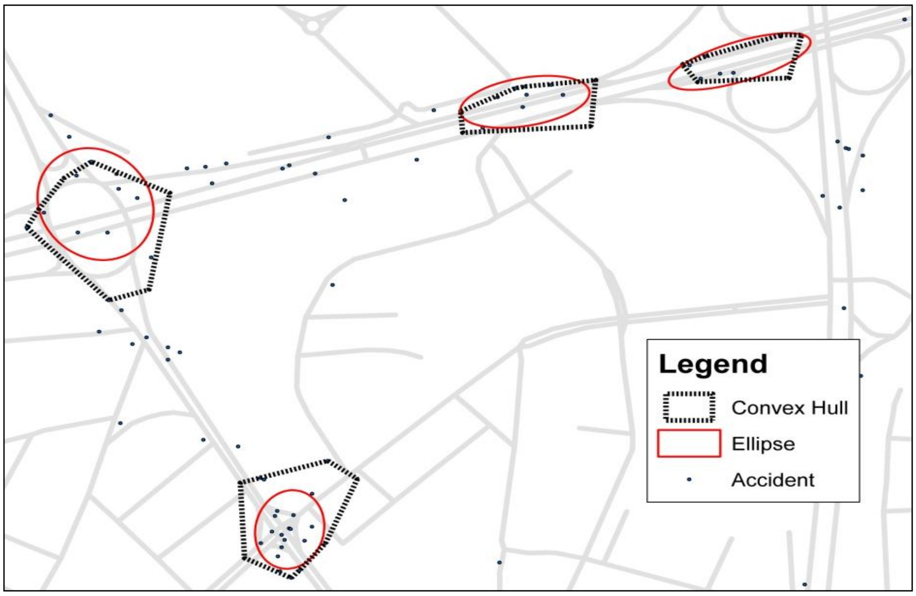

Figure 5.23: Visual comparison of convex hull and ellipse

There are visual differences between uniform and normal KDE analyses results' maps (Figure 5.24). It is clearly seen that normal KDE detect more area compared to the result of uniform KDE. It is occurred because normal KDE uses standard deviation of normal distribution. That's why smoother results can be observed in normal KDE maps. On the other hand, uniform KDE gives fewer hot spots in terms of area due to the mathematical calculation. In calculation step, bandwidth is the specified radius, which weight only point found within this radius. Uniform KDE provides more variations over study area. According to the results, uniform KDE is more appropriate for the analysis of traffic accident hot spots.

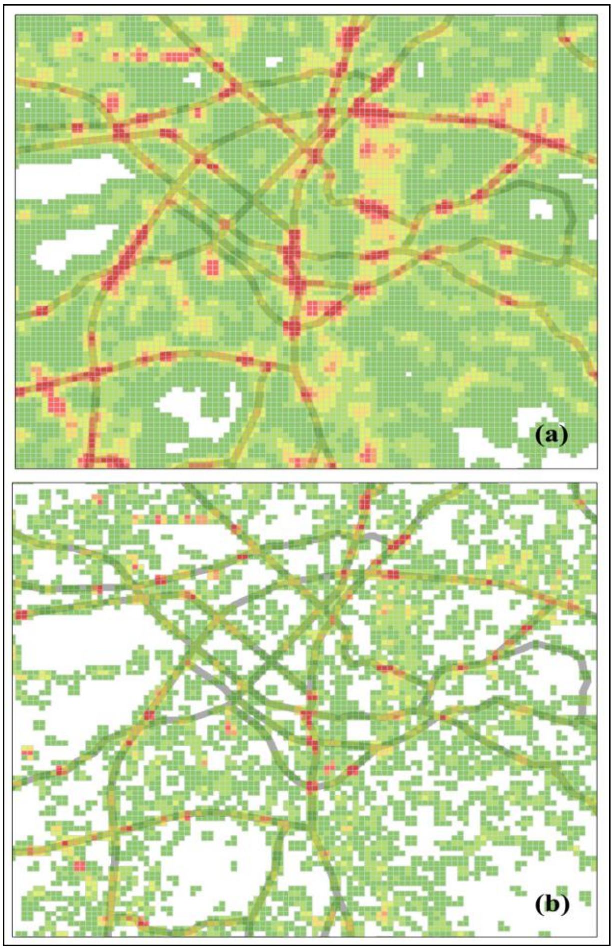

Figure 5.24: Visual comparison of a)normal and b)uniform KDE map

Visual comparison between NNH convex hulls and KDE maps for both types (normal and uniform) shows that, in most cases, hot spots detected by both methods are coincide. Comparison of the convex hull with normal KDE (Figure 5.25a), normal KDE identifies larger areas than convex hulls. On the other hand, convex hulls are coincide better in uniform KDE map (Figure 5.25b) since uniform KDE identifies smaller areas as a hot spot. As a result, NNH clustering algorithm with the output of convex hulls considered as more suitable to analyze traffic accident hot spot in the visual aspect.

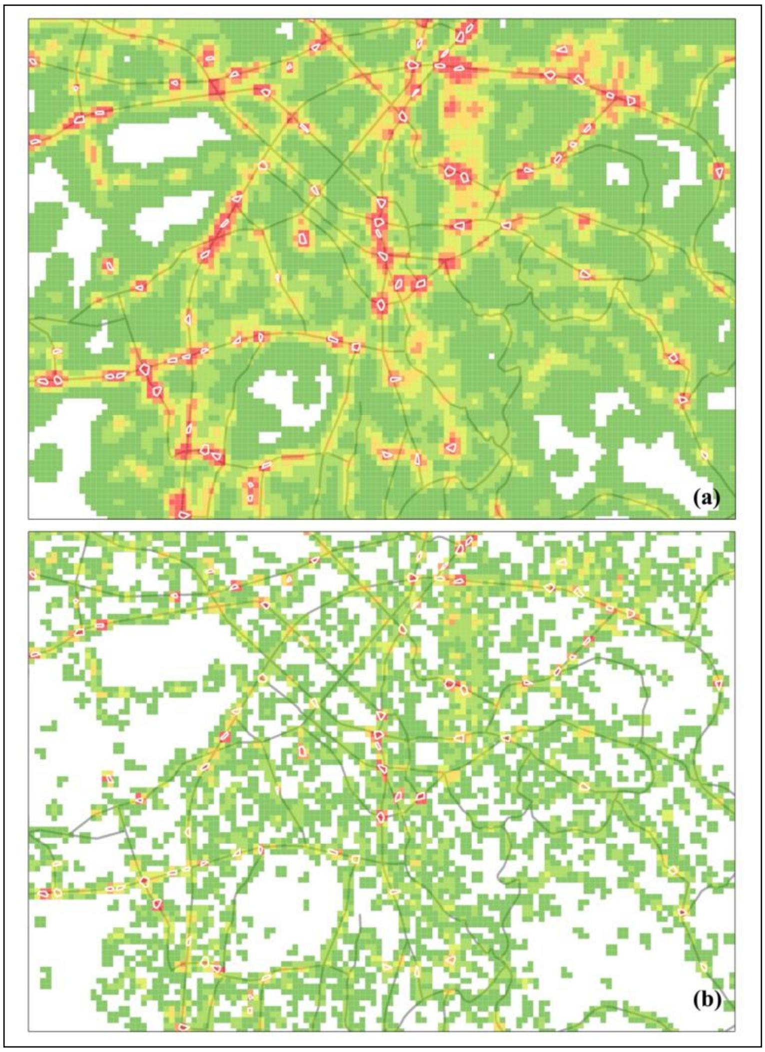

Figure 5.25: Comparison of convex hull with a)normal b)uniform KDE map

5.5.2. Mathematical Comparison of Hot Spot Methods

In this study, two different hot spot techniques were used. However, in the PAI calculation step, four different outputs are compared. Convex hulls and ellipses which are the outputs of NNH clustering algorithm and two different KDE methods, normal and uniform methods are compared with each other in terms of their PAI indexes. According to the results (Table 5.4), calculated PAI values for all accident data set with the help of two different clustering methods and four different outputs, convex hulls have the highest values compared with the other outputs. In all cases, ellipses also have high values compared to the KDE results.

In addition, when we look at the results of normal and uniform KDE, uniform KDE have higher PAI values for all accident data set. It shows us that, uniform KDE have better ability to capture hot spots. Since same criteria (cell size and bandwidth) were used in both types of KDE, it can be said that uniform KDE is more appropriate for identifying traffic accident hot spot.

Table 5.4: PAI result for hot spot methods

Method	Area (sq meter)	# of point	Density (n/A)	Hit Rate	Area Percentage	PAI
ALL ACCIDENT HOT SPOT						
NNH convex	845148.2	1764	0.0021	10.28	0.12	89.32
NNH ellipse	577998.9	1167	0.0020	6.80	0.08	86.40
Uniform KDE	1920000	1763	0.0009	10.27	0.26	39.30
Normal KDE	7130000	3720	0.0005	21.68	0.97	22.33
PEDESTRIAN ACCIDENT HOT SPOT						
NNH convex	777044.5	1072	0.0014	17.34	0.11	163.90
NNH ellipse	741560	709	0.0010	11.47	0.10	113.58
Uniform KDE	1060000	637	0.0006	10.31	0.14	71.39
Normal KDE	6260000	1767	0.0003	28.59	0.85	33.53
NON-PEDESTRIAN ACCIDENT HOT SPOT						
NNH convex	1228105	2179	0.0018	12.70	0.17	75.93
NNH ellipse	974656.3	1376	0.0014	8.02	0.13	60.42
Uniform KDE	1570000	1114	0.0007	6.49	0.21	30.36
Normal KDE	5800000	2261	0.0004	13.18	0.79	16.68
MORNING PEAK ACCIDENT HOT SPOT						
NNH convex	87974.51	114	0.0013	5.70	0.01	476.01
NNH ellipse	142903.6	85	0.0006	4.25	0.02	218.50
Uniform KDE	118125	43	0.0004	2.15	0.02	133.72
Normal KDE	2334375	339	0.0001	16.96	0.32	53.35
NOON OFF-PEAK ACCIDENT HOT SPOT						
NNH convex	1191586	1567	0.0013	26.75	0.16	164.82
NNH ellipse	1775548	1142	0.0006	19.49	0.24	80.61
Uniform KDE	781875	389	0.0005	6.64	0.11	62.35
Normal KDE	5371875	1267	0.0002	21.62	0.73	29.56
EVENING PEAK ACCIDENT HOT SPOT						
NNH convex	761669.1	981	0.0013	18.24	0.10	175.86
NNH ellipse	1206491	710	0.0006	13.20	0.16	80.35
Uniform KDE	315000	153	0.0005	2.84	0.04	66.32
Normal KDE	2795625	678	0.0002	12.61	0.38	33.11
NIGHTTIME ACCIDENT HOT SPOT						
NNH convex	497223.6	755	0.0015	19.29	0.07	284.87
NNH ellipse	787359.3	545	0.0007	13.92	0.11	129.86
Uniform KDE	343125	153	0.0004	3.91	0.05	83.66
Normal KDE	4275000	858	0.0002	21.92	0.58	37.65

CHAPTER 6

EVALUATION OF TRAFFIC ACCIDENT HOT SPOTS IN ANKARA

In this chapter, identified hot spots in the study area are analyzed with different parameters. Firstly, spatial and temporal distributions of urban hot spots are examined. After, distribution results are checked with the statistical test to find whether the results are significant or not. Then, distance analysis of detected urban hot spots to the nearest intersection is studied and explained. After the distance analysis step, results are also tested with the parametric test to search significancy.

6.1. Spatial and Temporal Distribution of Accident Hot Spots

As mentioned before, hot spots in the cities are not always distributed uniformly over space or time. Urban built environment and transportation have always been in an active interaction that has both spatial and temporal dimensions, directly. Land use patterns affect the travel behavior in both trip generation and trip distribution, such that more trips are generated and distributed in some regions than the others, such as commercial areas, residential areas. Also, different time periods directly affect the traffic volume in urban areas.

Traffic volume is one of the basic parameters in accident analysis, since there is a strong relationship between the volume and number of accidents. Similarly, the more pedestrian activity in the region is, the higher the portion of the pedestrian is expected in that region. For this reason, hot spots should be evaluated together with traffic volume or urban built environment characteristics and in time. In the absence of traffic volume data for the city of Ankara, other exposure measures such as geographical area and road network length are used.

The spatial distribution of injury accident hotspots among the four zones were obtained by simply checking in which zone a hotspot is located in. For hot spots obtained from all accidents' analysis, out of 64 hotspots, 12 (%18.8) of them are located in CBD. Besides this, in the second zone, there are 15 (%23.4) hot spots are found. 29 of all hot spots calculated by all accident are found in the third zone, which is nearly the half (%45.3) of the hot spots. Since major arterials and high speed corridors are located in this zone, this striking result is expected. In the fourth zone, which is the second large zone in terms of total urbanized area, there are 8 (%12.5) hot spots are found. As explained earlier, this result is not surprising since this zone is mostly covered by less dense residential areas.

With the same selection criteria (n_{min}=20 and d=100m), there were only 9 pedestrian accident hot spots, 6 of which were in the CBD, 2 in the first zone and 1 in the third zone, whereas 22 non-pedestrian accident hot spots were scattered among all the zones (Table 6.1). But forcing minimum of 20 accidents in a subset of the data points is very restrictive, as discussed before. Thus, reducing the n_{min} to 10 in clustering (with d=100m), more pedestrian and non-pedestrian accident hot spots were found in the urban zones (Table 6.1)

The most striking result emerged from the distribution of pedestrian only accidents hot spots that out of 71, 21 (%29.6) of them were found in the boundary of CBD. This is a very significant number taken into consideration the proportional area of CBD compared to other zones. As calculated before, CBD has %1.8 of all study area. Besides this, in the second zone, there were the total number of 28 pedestrian only accident hot spot were calculated. This is also an impressive result taken into consideration the proportional area of this zone. While most of the pedestrian only accidents hot spots mostly were seen in the CBD and second zone, 21 (%29.6) of them were in the third zone. Another interesting result emerged from the distribution of hot spots, only 1 hot spot was found in the fourth zone.

Table 6.1: Distribution of observed hot spot numbers (N_o) in urban zones

# of Hot spots Location in Urban Zones						
Accident Data Set	**n_{min}**	**CBD**	**2nd Zone**	**3rd Zone**	**4thZone**	**TOTAL**
All	20	12(%18.8)	15(%23.4)	29(%45.3)	8(%12.5)	**64(%100)**
Accidents by type						
Pedestrian	20	6(%66.7)	2(%22.2)	1(%11.1)	0	**9(%100)**
	10	21(%29.6)	28(%39.4)	21(%29.6)	1(%1.4)	**71(%100)**
Non- Pedestrian	20	3(%13.6)	2(%9.1)	11(%50.0)	6(%27.3)	**22(%100)**
	10	11(%8.0)	30(%21.9)	73(%53.3)	23(%16.8)	**137(%100)**
Accidents by time periods						
Morning Peak	5	5(%31.3)	5(%31.3)	5(%31.3)	1(%6.3)	**16(%100)**
Noon off-peak	5	26(%13.8)	61(%32.4)	90(%47.9)	11(%5.9)	**188(%100)**
Evening Peak	5	15(%11.6)	39(%30.2)	62(%48.1)	13(%10.1)	**129(%100)**
Nighttime	5	17(%16.8)	28(%27.7)	46(%45.5)	10(%9.9)	**101(%100)**

Out of 137 non-pedestrian accident hotspots, most of them were seen outside the CBD as expected. 11 (%8) hot spot were calculated in the boundary of CBD. In the second zone, there were 30 (%21.9) hot spots were found. More than half of all hot spots obtained from non-pedestrian accidents were seen in the third zone, where old interstate ring roads and high speed arterials are located. In this zone, there were 73 (%53.3) hot spot were found. As examined before, most of them were located on the arterials. Other than these, 23 (%16.8) hot spots were found in the fourth zone, mostly located in the southern part of the zone.

High speed, which is the one of the major factor causing the serious accidents, is seen off peak hours. The congestion is mostly seen in peak hours of the day. People activities like working, school, shopping, etc. are seen in daytime hours, which in turn generate traffic, in other words congestion. The congestion during peak hours decreases the mean speeds significantly, which in turn may reduce the occurrence of serious accidents. On the other hand, lack of traffic causes more speeding especially in night time. In combination with high speed, improper lightening and decreasing visibility range may increase the serious accidents.

Hot spots calculated by accident groups identified by time periods also show differences in terms of distribution over zones. There were total numbers of 16 hot spots identified by morning peak accidents. Interestingly, 5 hot spots were calculated in CBD, second and third zones. In other words, in these three zones, numbers of hot spots are the same. Since the transportation activities are high in the morning peak due to the working and school hours, number of hot spots observed in these zones could be higher from the expected numbers with regards to the proportional areas of each zone. Besides this, there was only one hot spot in the fourth zone which is located on the intersection.

Hot spots obtained from noon off-peak accidents were mostly located in the third zone. Nearly half of them, 90 of these (%47.9) were seen in this zone. Out of 188 hot spots, 26 (%13.8) of them were found in CBD; 61 (%32.4) of them were found in the second zone. Like previous distribution results, there were more hot spots than expected. In the fourth zone, 11 (%5.9) hot spots were seen, mostly in the northern part of the area.

Hot spots calculated by evening peak accidents show same attributes like noon off-peak accident hot spot results in terms of percentage. According to this, 15 (%11.6) hot spots were seen in CBD, and 39 (%30.2) hot spots were seen in the second zone. Out of the total number of 129 hot spots, 62 (%48.1) of them were located in the third zone. 13 hot spots were seen in the fourth zone, which the % 10.1 of all calculated hot spots is obtained from evening peak accidents. Compared to the morning and noon off-peak accident results, this percentage is higher from them.

Nighttime accident hot spots were mostly seen in the third zone as expected, where major arterials are located. According to this, 46 (%45.5) of 101 hot spot were located in the third zone. 17 (%16.8) hot spots were found in CBD. Its percentage is higher than noon-off and evening peak. In the second zone, there were 28 (%27.7) hot spots seen. Lastly, 10 (%9.9) hot spots were located in the fourth zone, mostly located on the southern part of the area as previous findings.

Since the traffic condition is changing during the hour of the day, number of accident occurrence is also not uniform over the time. However it is possible to do analysis of an hourly distribution of accident hot spot for every hour, it is more meaningful to divide accident hot spots to hourly periods. Accordingly, expected number of hot spots would be proportional to the total duration of defined periods.

6.2. Significance Analysis of Distribution of Hot Spots

To draw conclusions on the distribution of the hot spots among different zones and time period, it is important to perform non-parametric test for hypothesis. Statistical significance is a probabilistic assessment of a situation. In this approach, the real observed distribution of the hot spots with respect to spatial or time-based subgroups in an urban location should be checked against a null hypothesis (H_0) that such sub-grouping has no effect on the occurrences of the hot spots. This can be simply checked by performing a Chi-square (χ^2) test for the observed distribution of the hot spots over the null hypothesis.

After obtaining the distribution of the hot spots over zones and time period, it is important to know if these values are statistically significant compared to a uniform distribution in space or time, which would be the null hypothesis (H_0). In Crimestat, Monte Carlo simulation algorithm takes the data and assigns it randomly to the area. It then runs the NNH algorithm using randomly assigned data. By repeating the experiment many times (e.g. 100, 500, 1000), a confidence interval can be established for the number of clusters expected under random distribution. For Ankara injury data, Monte Carlo simulation algorithm was performed to produce expected number of hotspots, however with a given threshold distance of 100 meters and n_{min}, no clusters found; with this big geographic area, if distributed randomly; the accidents would not have produced any clusters.

Secondly, the numbers of detected hot spots were studied according to the areas of the zones. To calculate the expected numbers of hot spots (N_e), geographical area is used as a baseline. Total numbers of observed hot spots (N_o) for each accident data set was distributed with regard to area of each zone assuming as uniform distribution of hot spots. Performing a Chi-square (χ^2) analysis of the real and expected number of hot spots (in a zone or a time period), which had a critical value of $\chi^2_{3,0.05}$=7.81 for distributions with 4 intervals at p=0.05 level (Table 6.2), it was observed that distribution of;

- all injury hot spots over the four urban zones,
- pedestrian hot spots over the four urban zones,
- non-pedestrian hot spots over the four urban zones
- timely accident hot spot over four urban zones and
- hot spots over the four time periods

were statistically significantly different than uniform distribution over space and time. Analysis of the expected and observed numbers of hotspots, N_e and N_o, respectively can be seen in the Table 6.2.

Table 6.2: Chi-square test summaries for spatial and time-based distribution of injury accident hot spots

Spatial distribution of hot spots						
	CBD	**2^{nd} Zone**	**3^{rd} Zone**	**4^{th} Zone**	**Total**	**Hypothesis Testing**
All Accident Hot spots n_{min}=20						
Observed (N_o)	12	15	29	8	64	H_1
Expected (N_e)	1	16	25	22	64	H_0
$\Delta\chi^2=(N_o - N_e)^2/N_e$	116.84	0.06	0.64	8.91	**126.45**	(Reject H_0= H_1)
Pedestrian Accident Hot spots n_{min} =10						
Observed (N_o)	21	28	21	1	71	H_1
Expected (N_e)	1	18	27	25	71	H_0
$\Delta\chi^2$	360.01	5.56	1.33	23.04	**389.94**	(Reject H_0= H_1)
Non-Pedestrian Accident Hot spots n_{min} =10						
Observed (N_o)	11	30	73	23	137	H_1
Expected (N_e)	2	34	53	48	137	H_0
$\Delta\chi^2$	40.50	0.47	7.55	13.02	**61.54**	(Reject H_0= H_1)
Morning Peak Accident Hot Spots n_{min} =5						
Observed (N_o)	5	5	5	1	16	H_1
Expected (N_e)	0	4	6	6	16	H_0
$\Delta\chi^2$	73.63	0.06	0.17	4.17	**78.03**	(Reject H_0= H_1)
Noon off Peak Accident Hot Spots n_{min} =5						
Observed (N_o)	26	61	90	11	188	H_1
Expected (N_e)	3	47	72	66	188	H_0
$\Delta\chi^2$	171.98	4.17	4.50	45.83	**226.48**	(Reject H_0= H_1)
Evening Peak Accident Hot Spots n_{min} =5						
Observed (N_o)	15	39	62	13	129	H_1
Expected (N_e)	2	32	50	45	129	H_0
$\Delta\chi^2$	84.50	1.53	2.88	22.76	**111.67**	(Reject H_0= H_1)
Nighttime Accident Hot Spots n_{min} =5						
Observed (N_o)	17	28	46	10	101	H_1
Expected (N_e)	2	25	39	35	101	H_0
$\Delta\chi^2$	112.50	0.36	1.26	17.86	**131.97**	(Reject H_0= H_1)
Time-based distribution of Hot spots n_{min} =5						
	Morning Peak	**Noon Off-peak**	**Evening Peak**	**Night time**	**Total**	
Observed (N_o)	16	188	129	101	434	H_1
Expected (N_e)	54	109	90	181	434	H_0
$\Delta\chi^2$	26.74	57.26	16.90	35.36	**136.26**	(Reject H_0= H_1)

Distribution of all accidents hot spots over four zones show significant differences. According to this, in CBD, and third zone, observed hot spots are higher than the expected number of hot spots. On the other hand, in the second and fourth zone observed hot spots are lower than the expected number of hot spots.

Pedestrian only accident hot spots distributed differently according to the zone. In the CBD, observed number of hot spots is very high compared to the expected number of hot spots. Taking into consideration the area of CBD, expected number of hot spot is calculated as 1. However, total number of 21 pedestrian only accident hot spots is observed in the CBD. Other than CBD, in the second zone, where residential land usage is highly seen, observed number of hot spot is also higher that the expected. In the third and fourth zones, observed number of hot spot is lower than the expected. Especially in the fourth zone, difference between the observed and expected number of hot spot is very high and statistically significant.

Distribution of non-pedestrian accident hot spots among four zones shows statistically significant differences. According to the results, observed number of hot spots in CBD and third zone is higher than the expected. In the third zone, difference between observed and expected number of hot spots is very high since the most of the major arterials and high speed corridors are located in this zone. In the second and fourth zone, observed number of hot spot is lower than the expected number of hot spot.

Morning peak accident hot spots also show statistically significant differences in terms of distribution over four zones. According to this, in the CBD and second zone, observed number of hot spot is higher than the expected. On the other hand, in the third and fourth zones, observed number of hot spot is lower than the expected number of hot spots.

Distribution of noon off-peak accident hot spot over four zones shows differences. According to this, observed number of hot spot in CBD, second zone and third zone are higher than the expected number of hot spot. On the other hand, observed number of hot spot in fourth zone is lower than the expected. These differences are all statistically significant at $p=0.05$ value.

Evening peak accident hot spot also show differences in terms of observed and expected numbers according to the urbanized area. In the CBD, second zone and third zone, observed number of hot spot is higher than the expected number of hot spot.

Distribution of nighttime accident over four urban zone shows differences in terms of observed and expected numbers with regards to the total urbanized areas of zones. In CBD, second zone and third zone, observed number of hot spot is higher than the expected number of hot spot. Contrary to this, observed number of hot spot is lower than the expected in fourth zone. All these differences are statistically significant at $p=0.05$ value.

6.3. Distance Analysis of Hot Spots to the Nearest Intersections

Besides the importance of distribution of accident hot spots over urban areas, distances of hot spots to the major intersections is also important since the intersections are the major conflict points of the urban traffic flow. The hot spots in urban areas are mostly in relation with the intersections. Distances between hot spots and intersections can be the indicator of preventive measurements; where and what precautions should be implemented. Moreover, intersections that are the location of hot spots can be an indicator of wrong design.

In order to search relationship between accident hot spots and intersections, distances between them are measured. In this process, nearest intersection of each hot spot is calculated with their intersection type. In addition to this, frequency (number of accidents found in hot spot) information of each hot spot is available. With this information, relationship between the severity and distance to the intersection can also be evaluated.

Distances of each hot spot to the nearest intersections were calculated with the help of the "Near tool" within the subsection of Proximity Tool of Analysis

Toolbox of ArcGIS. After the near distances are calculated, graphs showing the relationship between distance and number of hot spots are generated.

In the Figure 6.1, the relationship between the number of hot spot obtained by all accident for n_{min}=20 and their distances to the nearest intersection can be seen. Graph show the skewed distribution which highlighted that most of the hot spots is located near to the intersection in the distance of 0-5 meters. In this interval, more than half of the hot spots (%56) are located. Getting away from the intersections, it is observed that number of hot spot is decreasing. It indicates that, accident prone locations are seen mostly in or near the major intersections of the urban area.

In order to compare the hot spots obtained by all accidents with the pedestrian and non-pedestrian accident hot spots which are calculated with n_{min}=10, hot spots of all accident are calculated by n_{min}=10. According to the Figure 6.2, like result obtained by n_{min}=20, most of the hot spots are seen between the interval 0-5 meters in terms of distances to the intersections. Different from previous one, hot spots are seen in the interval of 1200-1399 and 1400-1599.

In the Figure 6.3, distances of hot spots that are calculated from all accidents with n_{min}=5 to the intersections can be seen. According to the figure, like previous results, most of the hot spots are located at or near to the intersections. Different from previous results (n_{min}=20 and n_{min}=10), distances decrease gradually while getting away from the intersections.

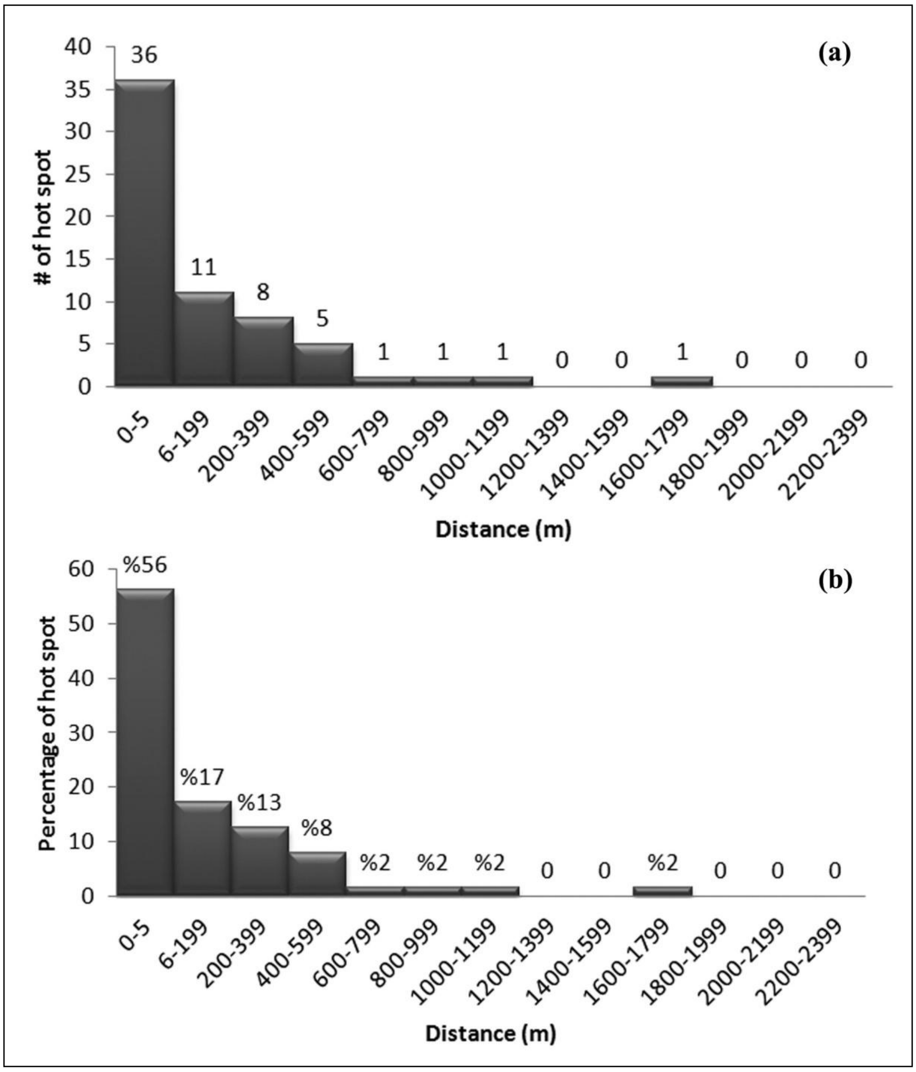

Figure 6.1: Relationship between a)number b)percentage of hot spot and their distances to nearest intersection for all accident hot spot (n_{min}=20)

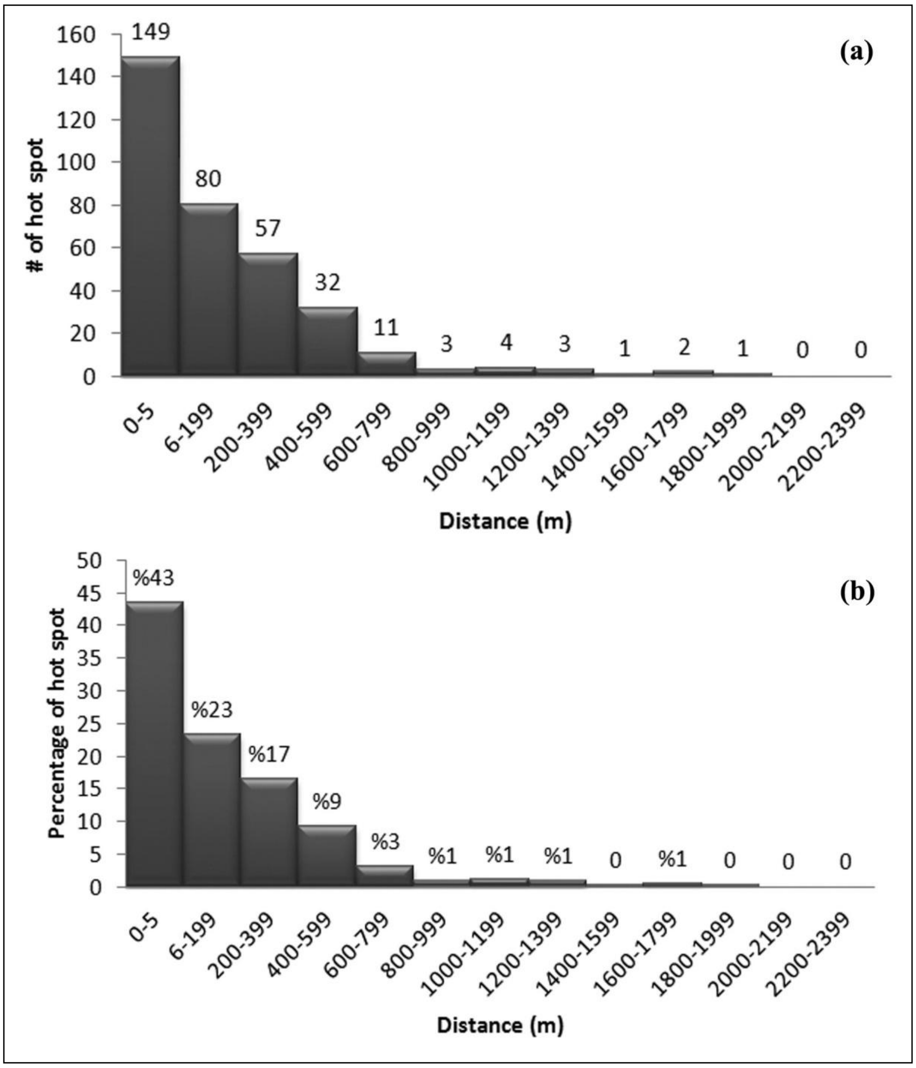

Figure 6.2: Relationship between a)number b)percentage of hot spot and their distances to nearest intersection for all accident hot spot (n_{min}=10)

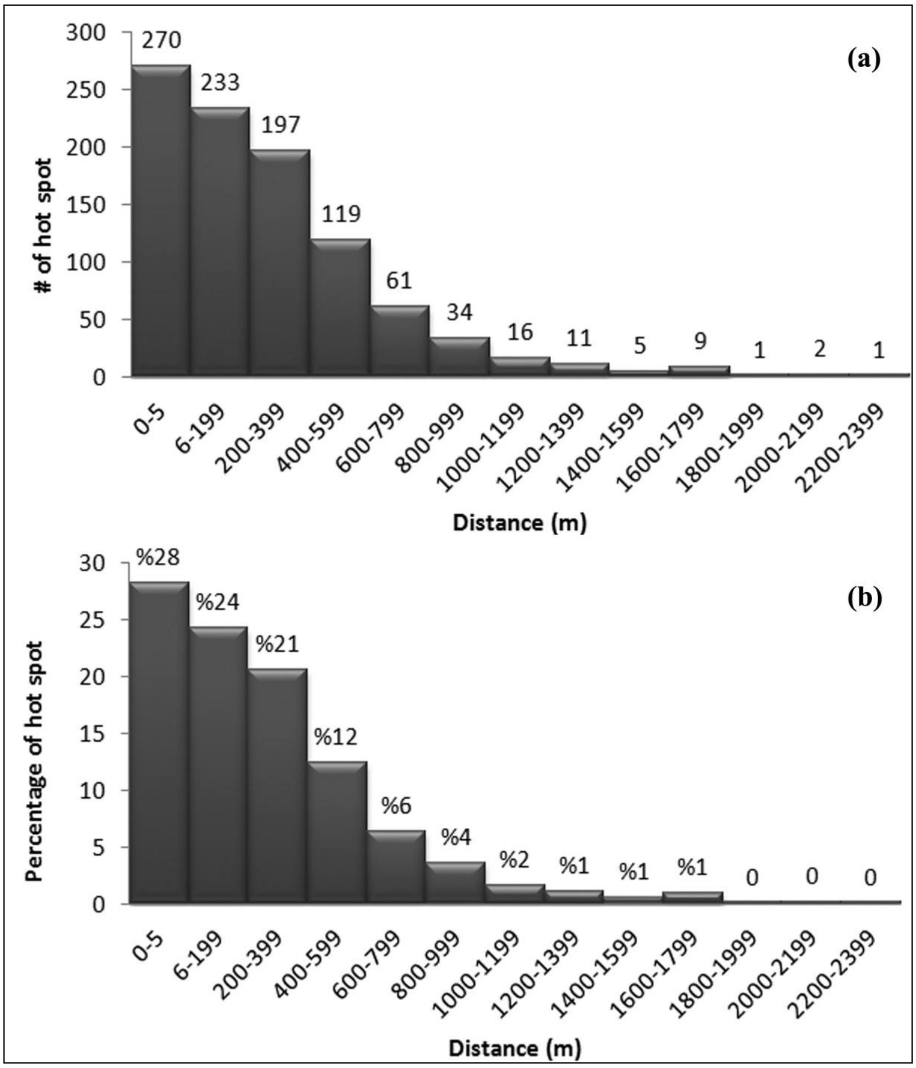

Figure 6.3: Relationship between a)number b)percentage of hot spot and their distances to nearest intersection for all accident hot spot (n_{min}=5)

Relationship between the numbers of hot spot obtained by non-pedestrian accidents and their distance to the intersection show same characteristic with all accident hot spots. According to the Figure 6.4 and 6.5, distances of hot spots to the intersections are decreased while getting away from the intersections. Also, most of the hot spots are located between the 0 and 5 meters, which are accepted as at the intersection.

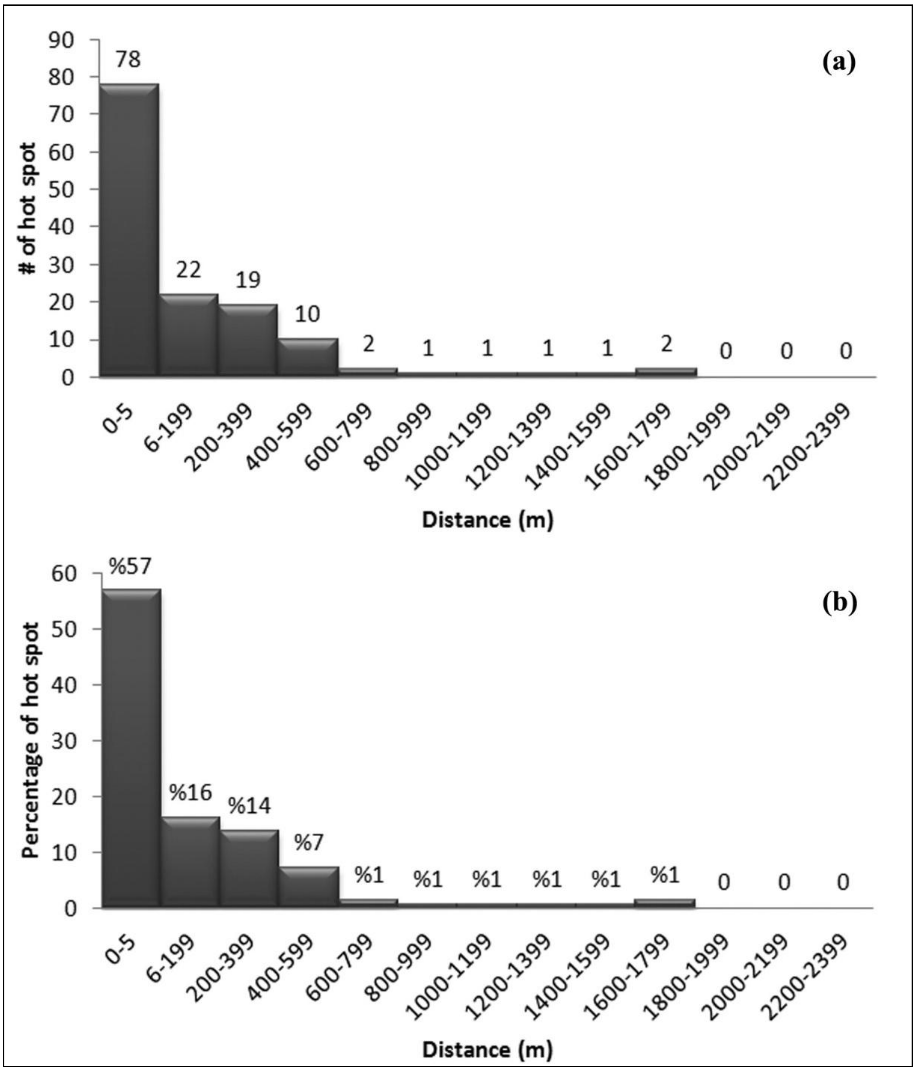

Figure 6.4: Relationship between a)number b)percentage of hot spot and their distance to nearest intersection for non-pedestrian accident hot spot (n_{min}=10)

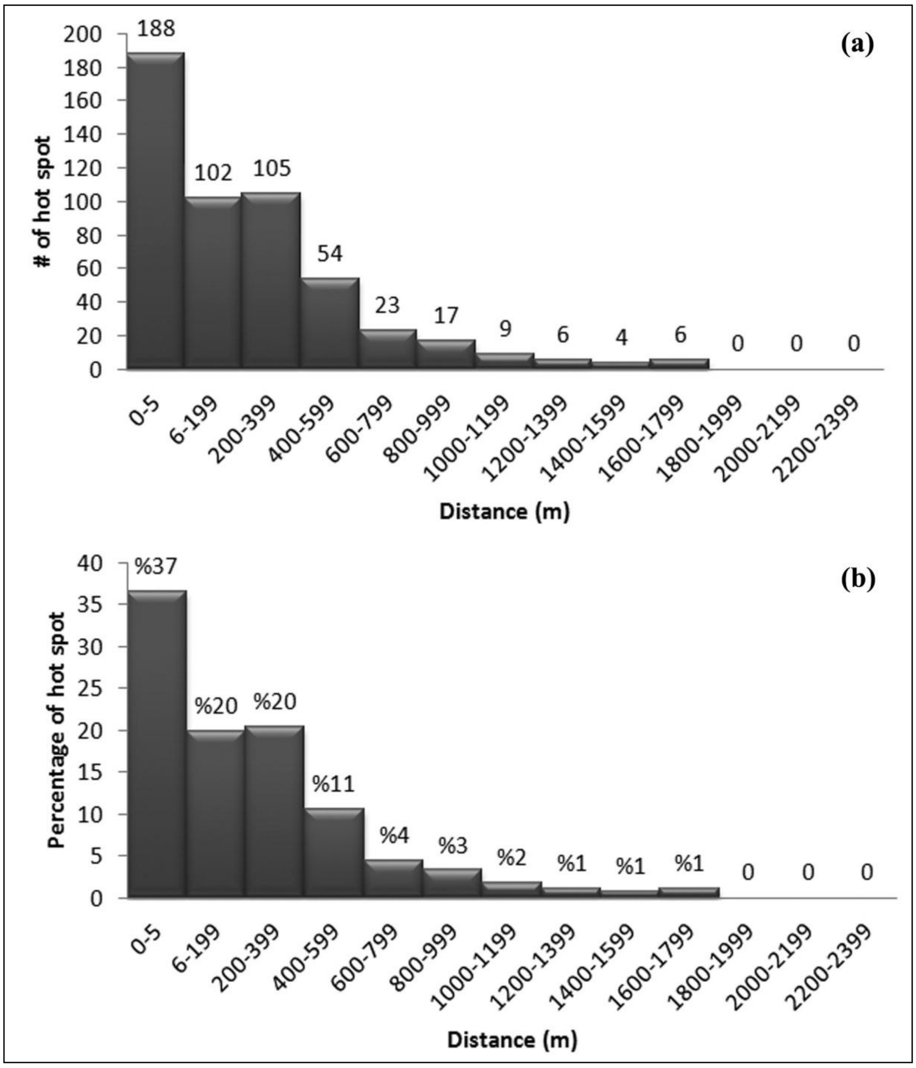

Figure 6.5: Relationship between a)number b)percentage of hot spot and their distance to nearest intersection for non-pedestrian accident hot spot (n_{min}=5)

Relationship between the numbers of hot spot obtained by pedestrian accidents and their distance to the intersection show same characteristic with previous results. According to the Figure 6.6 and 6.7, distances of hot spots to the intersections are decreased while getting away from the intersections; but different from previous results, pedestrian accident hot spots are decreased smoother. In other words, differences between intervals are decreased gradually compared the results of all and non-pedestrian accidents' hot spots. Most of the hot spots are located between the 0 and 5 meters, which are accepted as they are in the intersection.

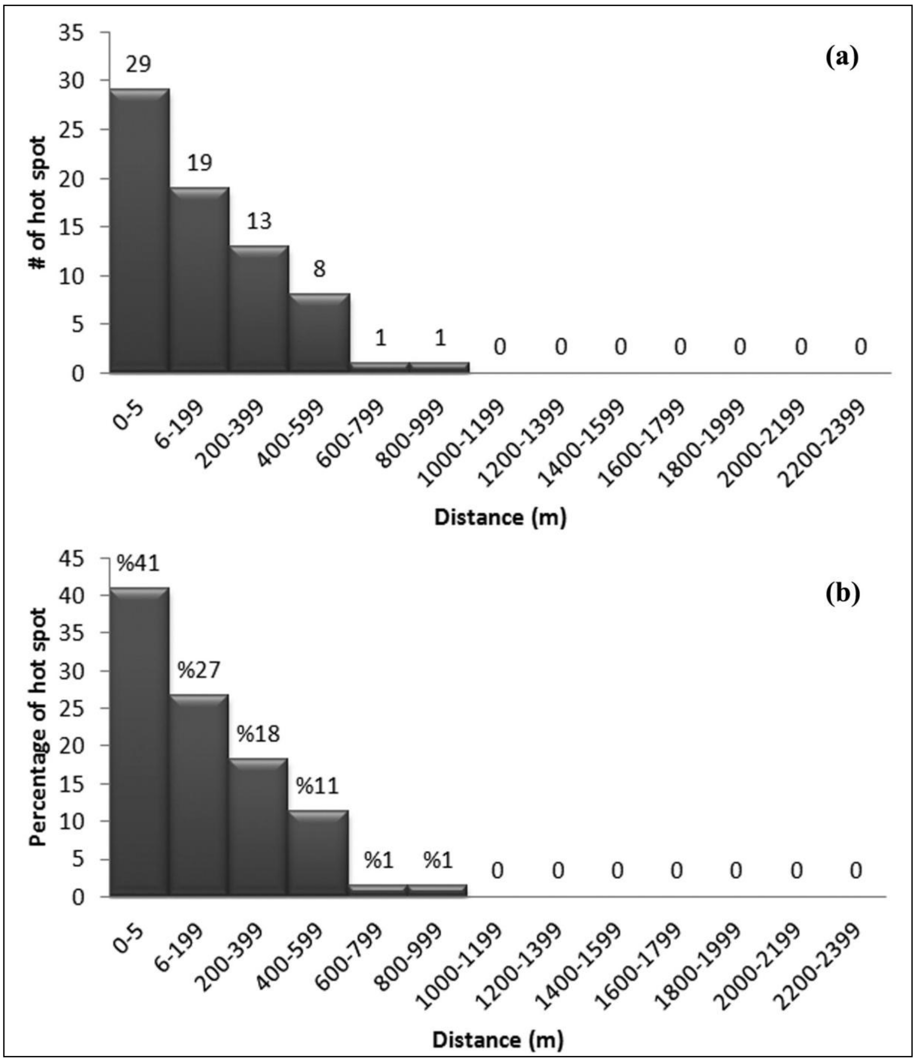

Figure 6.6: Relationship between a)number b)percentage of hot spot and their distances to nearest intersection for pedestrian accident hot spot (n_{min}=10)

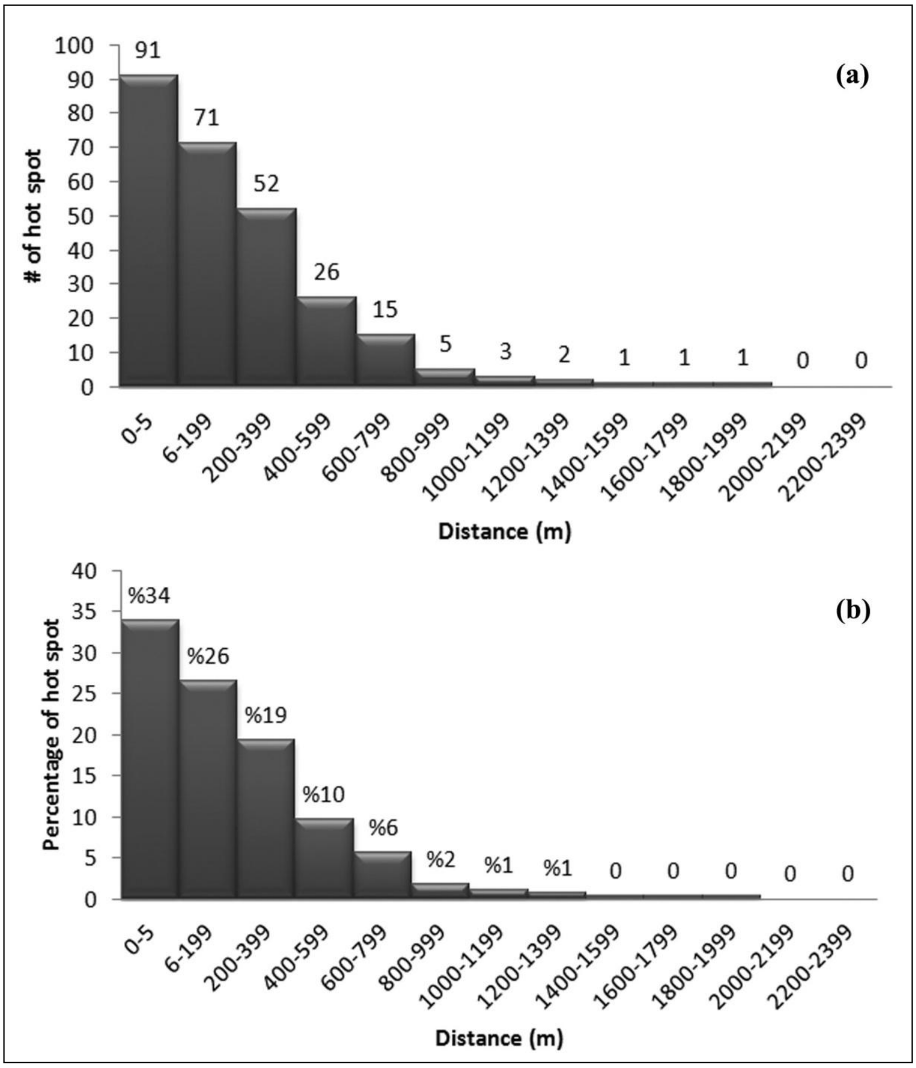

Figure 6.7: Relationship between a)number b)percentage of hot spot and their distances to nearest intersection for pedestrian accident hot spot (n_{min}=5)

6.4. Significance Analysis of Distances of Hot Spots to the Nearest Intersections

To conclude distances of hot spots to the nearest intersections, it is important to perform statistical tests whether distances are significant or not. In order to understand the significancy, parametric tests can be performed. Statistical significance of distance analysis results can be checked with the help of One-Way ANOVA test which depends on the comparison of means of different groups. To compare the differences of means, groups should be created. Furthermore, method measures the means with regards to dependent variable.

For this study, distances of hot spots to the nearest intersections are calculated. After the calculation, distances are grouped into distance intervals to describe the relationship between intersections and hot spots. According to this, between the distances of 0-5 meters are accepted as the hot spot is located on the intersection. After the distance of 5 meters, groups are generated at intervals of 200 meters distances. The other group is created between the distances of 6-199 meters that are considered that hot spots found in this interval are in close relationship with intersections. Third group is generated between the distances of 200-399 meters that are considered as occurrence of hot spot is related with intersections. Hot spot located far away from the 400 meters from the intersections are determined as there is weak relationship between intersections and hot spots.

While performing One-Way ANOVA, dependent variable is selected as the cluster severity. Aim is to search significancy that if there are more severe clusters on or near to the intersections. As can be understood from the clustering results, hot spots are mostly located on or near to the major intersection. Rather than the analyzing the occurrence location of hot spots, severity of them related to the intersections is more meaningful to take important precautions. As a result; hypothesis;

> **H_0:** There is no difference between the mean values of hot spot severity among different distance intervals from the intersections.
>
> **H_1:** There is significant difference between the mean values of hot spot severity among distance intervals through the intersections.

Descriptive statistics of analysis results are summarized in the Table 6.3. Table shows that hot spots obtained by all injury accidents are mostly located on the intersections. There are total numbers of 270 hot spots found on the intersections with the mean severity value of 13.04. Within the interval of 6-199 meters, there are 112 hot spots are calculated of which mean severity value is calculated as 10.73. The other two intervals have different numbers of hot spots; however mean values of severities are almost same. Hot spots calculated by non-pedestrian accidents show differences in terms of the distribution of distance intervals to the intersections. Accordingly, most of the hot spots are observed on the intersections. Numbers of hot spots that fall into other three intervals are nearly same. Again, mean value of severity of hot spots that located on the intersections is greater than the others. Pedestrian accident hot spots almost have the same situation that number of hot spot and mean value of severity on the intersection are greater than the other three intervals. Mean value of severities of other three intervals are almost the same.

Table 6.3: Descriptive Statistics

Injury Accident Data Set	**Distance Interval**	**# of hot spot**	**Mean**	**Std. Dev.**	**Min. Value of Severity**	**Max. Value of Severity**
All	0-5 m.	270	13,04	7,793	5	59
	6-199 m.	112	10,73	7,241	5	47
	200-399 m.	122	9,38	4,076	5	24
	400 + m.	455	9,28	4,578	5	46
	TOTAL	**959**	**10,52**	**6,153**	**5**	**59**
Non-pedestrian	0-5 m.	188	10,99	5,481	5	36
	6-199 m.	102	9,20	5,130	5	43
	200-399 m.	105	8,64	4,868	5	46
	400 + m.	119	8,47	4,054	5	34
	TOTAL	**514**	**9,57**	**5,093**	**5**	**46**
Pedestrian	0-5 m.	91	10,37	6,267	5	37
	6-199 m.	71	8,76	4,100	5	26
	200-399 m.	52	8,46	3,293	5	17
	400 + m.	54	8,46	3,260	5	20
	TOTAL	**268**	**9,19**	**4,750**	**5**	**37**

After the descriptive statistics, test of homogeneity of variances are calculated to search whether equal variances are assumed or not. Homogeneity of variances of each accident data set is calculated by Levene’s Test for Equal Variances and summarized in the Table 6.4. Since the F values are large, and the corresponding significance levels are below 0.05, equal variances cannot be assumed. As a result of non equal variances, it is needed to perform Robust Test for Equality of Means. To do this, Welch Statistics of each accident groups are calculated.

Table 6.4: Test of homogeneity of variances

Injury Accident Data Set	Levene Statistic	df1	df2	Sig.
All	21.715	3	955	,000
Non-pedestrian	4.374	3	510	,005
Pedestrian	4.511	3	264	,004

As a result of non equal variances, we need to be careful with interpreting the ANOVA table. Because the variances are not equal, the resulting F value and significance value might be off enough to sway the output of the test. The result of ANOVA test (Table 6.5) imply that we would reject H_0 for all groups, because of the low significance level (lower than 0.05), however, we need to verify this with the robust test for equality of means.

Table 6.5: ANOVA test results

Injury Accident Data Set		Sum of Squares	df	Mean Square	F	Sig.
All	Between Groups	2,588	3	862,720	24,462	**,000**
	Within Groups	33,681	955	35,268		
	Total	36,269	958			
Non-pedestrian	Between Groups	628,027	3	209,342	8,420	**,000**
	Within Groups	12,680	510	24,863		
	Total	13,308	513			
Pedestrian	Between Groups	196.719	3	65,573	2,970	**,032**
	Within Groups	5828.575	264	22,078		
	Total	6025.295	267			

Robust test of equality of means calculated by Welch Statistics is summarized in the Table 6.6. According to the results, all and non-pedestrian accident hot spots' statistic results are significant since significance values of them are below the 0.05. On the other hand, pedestrian accident hot spots Welch Statistic's significance is above the 0.05 which means that we fail to reject our H_0.

Table 6.6: Robust test of equality of means

Injury Accident Data Set	Welch Statistic	df1	df2	Sig.
All	18.521	3	305.717	,000
Non-pedestrian	8.209	3	259.093	,000
Pedestrian	2.268	3	143.272	,083

Since ANOVA and Welch Statistic results give sufficient evidence for rejecting H_0 for hot spots of all and non-pedestrian accident data set, it should be determined that interval groups are different from each other in terms of their mean severity values. To understand this, Duncan Post-Hoc test is performed for both data sets. What the Duncan test also does is put the variables into groups which have a similar mean score on the dependent variable. These are called 'homogeneous subsets'. According to the result of all accident data set (Table 6.7), mean severity value (13.04) of hot spots located on the statistically significant from the other three interval groups. Furthermore, hot spots located between the distances of 6-199 meters also statistically significant from the other three groups that have the mean severity value of 10.73.

Table 6.7: Duncan Post-Hoc test result for all accident hot spot

Distance Interval	# of hot spots	Subset for alpha = 0.05		
		1	2	3
400+ m.	455	9,28		
200-399m.	122	9,38		
6-199m.	112		10,73	
0-5 m.	270			13,04
Sig.		,875	1,000	1,000

Hot spots obtained from non-pedestrian accidents located on the intersection which are fall into interval of 0-5 meters have statistically significant differences in terms of their hot spot severity values (Table 6.8). According to this, mean severity value (10.99) of hot spot observed on the intersections is statistically greater than the other three interval groups. This shows that, non-pedestrian accident hot spots which are located on the intersections have more accidents.

Table 6.8: Duncan Post-Hoc test result for non-pedestrian accident hot spot

Distance Interval	# of hot spots	Subset for alpha = 0.05	
		1	2
400+ m.	119	8.47	
200-399m.	105	8.64	
6-199m.	102	9.20	
0-5 m.	188		10.99
Sig.		,875	1,000

6.5. Limitations in the Generalization of the Findings

Limitations of the findings in this study can be summarized as follows:

- Study area is limited within the urban regions of Ankara, hot spot formation may be due to infrastructure and driver behaviors in this regions; the results would be valid only and within this study area.
- As the traffic data obtained from WHO Turkey, did not have any address information to correct the geocoded accident locations, any possible data errors could not be corrected or omitted. Data quality is limited for this study.

  A major error is suspected at a region almost east of the CBD where a dense accident accumulation is detected (Figure 6.8) but, no parallel street network is observed. This situation may be occurred due to;
  - Incorrect geodetic datum selection of user which can cause error from 1 to hundreds of meters,
  - Entering inconsistent values to some extent in terms of X and Y coordinates while carrying read values from GPS device to the report or from report to the database.

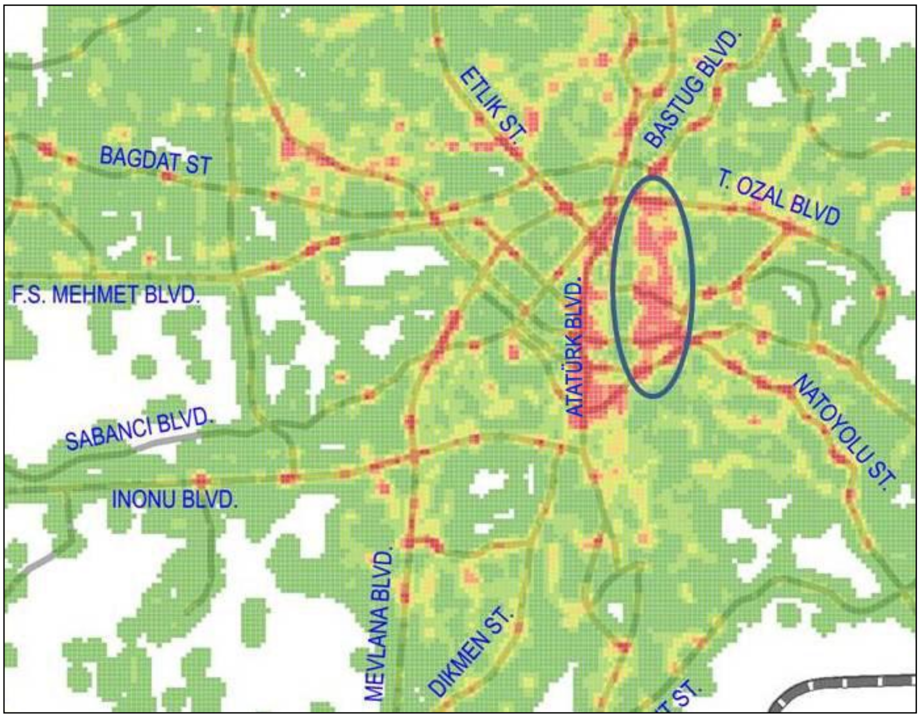

Figure 6.8: KDE result of mis-located accident data

- Road network data is available; however there is no intersection data or information is available for this study. Only major intersections are identified with the help of road network data and open street map available in ArcGIS program. Due to the changes in terms of regulations about intersections geometry or design, it is not possible to determine intersection type (number of legs, rotary intersection or roundabout) and specify as signalized or not.
- Traffic volume data is not available for the City of Ankara. Due to the time and budget limits, it is not possible to collect and calculate traffic volume data for a citywide analysis. Any normalization with the traffic volume could not be included in the studies.
- Besides these, land use data is not available for this study. Study is limited to the specified urban zones in terms CBD, inner urban zones and outer (urban transition) zones. Since no land use information about sublevel, accidents cannot be related with land use information in details.

CHAPTER 7

CONCLUSIONS AND FURTHER RECOMMENDATIONS

The distribution of traffic accidents over an urban region is generally not uniform, as expected. So, it is important to detect accident prone locations, which are called hot spots. Even though, there is a well-defined “black spot” definition for highway safety, it is not that easy or straightforward to adopt it for urban regions. Mainly because of the complex road network of urban regions, it is not easy to define a unit segment in urban areas. For this reason, there is not strict definition to find traffic accident hot spots. Also, definition of hot spot in urban areas can vary based on the scope of the study.

To address these complications in urban traffic safety analyses, it is important to develop a framework to study hot spots in urban regions methodologically, which is the main focus of this study. The proposed methodology focused on not only detection of urban traffic accident hot spots, but assessment of their distribution with regards to the built environment features. The hot spot detection process is carried by CrimeStat which produces clusters based on NNH technique. The built-environment relations are analyzed at the levels of urban zones and distance to nearest intersections. While all the calculations are carried in a GIS environment, the significance of the relations is tested via statistical tools.

7.1. Discussion about NNH Clustering Algorithm

NNH clustering algorithm results revealed the following insights;

- As the distance threshold value is decreased, minimum number of data points in a cluster n_{min} should be reduced to create reasonable cluster definitions.
- Assuming a large maximum distance in the NNH clustering result detection of larger hot spots that cannot be related to the network features such as intersections.
- A good limit for NNH clustering to study the relationship between hot spots and intersections require much smaller threshold value, which is found as100m for Ankara network.
- If sub-data groups are studied, the threshold distance can be kept the same to compare with all data hot spots, but n_{min} has to be reduced to reflect the expected loss of severity in the hot spot definitions. For example, n_{min}=20 and 100m criteria for all accident data set are replaced by n_{min}=10 and 100m for pedestrian versus non-pedestrian subgroups for the case study area. But, when the data was studied in four different time periods, n_{min}=5 was appropriate for detect hot spots.
- Comparison of NNH clustering with KDE showed that, both visually and mathematically, NNH created more accurate hot spots. Since KDE algorithm will smooth accident data over small grid cells, it will distort relationships making it appear that there are higher accidents likelihoods in locations than actually exist. In other words, KDE averages more over cells and produces less specific hot spot definition (small grid cells). For this reason, use of KDE would not be appropriate while analyzing the distances between hot spots and built environment components.

7.2. Discussion about Ankara case study results

The number of fatality accident hot spots is created with the threshold distance of 1 km that is also used in highway traffic safety, as well. Taken into consideration the number of fatality accidents is few, n_{min} is selected as 3. According to the results, total number of 15 hot spots is detected. Number of accidents found in clusters varies between 3 and 8. Results show that, fatality accident hot spots are mostly located on the high speed corridors and main arterials. Besides, some of them are located at the intersections. Visualization of fatality accident hot spots according to their number of accidents found in cluster provides traffic safety officials to identify priority areas. How well and accurate clusters are argumentative since they are created with 1 km threshold distance. For urban region studies, since network is more complicated compared to highways, identified hot spots cover more than one road in some areas.

The number of hot spots obtained from all injury accidents shows differences in terms of different n_{min} values as expected. For all results, it can be said that most of the hot spots are located on the third zone where high speed corridors and main arterials are mostly located. Traffic safety officials should be aware of this situation since nearly half of the total calculated hot spots are seen in this zone.

Pedestrian accident hot spots are also calculated with different n_{min} values. The most striking result emerged from pedestrian accident hot spot analysis is that most of the hot spots are located in the CBD and second zone where settlement area is mostly seen. More than half of the total number of hot spots is concentrated on the CBD and second zone as expected. On the other hand, in the fourth zone, number of pedestrian hot spot is one for n_{min}=10 and 2 for n_{min}=5 although there is settlement area located in this zone. This situation could be occurred due to the low mobility of pedestrian activities compared to the other zones.

Non-pedestrian injury accident hot spots are mostly located on the major and minor arterials as expected. More than half of the total number of calculated hot spots is seen in the third zone where high speed corridors and main arterials are located in. More non-pedestrian accident hot spots are seen in the fourth zone compared to the pedestrian accident hot spots. On the other hand, less non-pedestrian accident hot spots are seen in the CBD and the second zone compared to the pedestrian accident hot spots.

Time-based hot spot locations also disperse over the study area. For all periods, number of observed hot spots in the CBD and the second zone are higher than the expected values taken into consideration the uniform distribution of hot spots over urban zones. Contrary to this, number of observed hot spot is less than the expected in the fourth zone for all time periods as expected.

Number of hot spots over the four time periods show differences. According to this, observed number of hot spots for morning and nighttime periods is lower than the expected. Contrary to this, observed number of hot spots is higher than the expected number of hot spots for noon off and nighttime periods taken into consideration the uniform distribution of hot spots over time periods. Temporal distributions of hot spots show differences. Accordingly, fewer hot spots are observed when compared to the expected values for morning peak and nighttime accidents. On the other hand, there are more hot spots calculated than expected for noon off and evening peaks.

Distance analysis showed that hot spots are mostly located at the intersections. Besides, numbers of hot spots are decreased while getting further away from the intersections. Non-pedestrian accident hot spots show more sudden decrease from the intersection in terms of distance compared to the pedestrian accident hot spots. This situation shows that intersections have priority to take precautions for preventing accident especially for the non-pedestrian accident occurrence. Furthermore, distance intervals of hot spots obtained from all and non-pedestrian injury accidents to the nearest intersections show differences in terms of mean severity values of hot spots.

Evaluation of the results shows that, intersections are very dangerous for both pedestrian and non-pedestrian accidents. This situation could be occurred due to the wrong geometric design or wrong signalization regulations. In order to provide traffic safety, more effort could be made about the design and regulation of intersections. While designing intersections or improving the geometry of intersections, traffic safety issue should be taken into consideration. Rather than the automobile priority, pedestrian priority should be considered. On the other hand, especially in Ankara, speed profile is higher than the design speeds. This situation is valid for all zones of the study area. In the CBD and residential areas, average speed should be 30km/h; however, speed profile is mostly seen 50km/h and more. This problem is arisen from the wrong design of roads. In Ankara, transportation is planned taken into consideration the vehicle priority. It can be seen clearly from the nonstop road design of the arterials. Even in the second zone, where residential areas are mostly seen in, speed is seen over 50 km/h. In the off-peak hours, speed is reached 70-90 km/h. High speed on the arterials especially in the residential areas turned out great number of accident occurrence. To sum up, it is obvious that, to provide traffic safety in Ankara, traffic speed and traffic volume should be brought under the control by traffic safety officials.

Distribution of traffic accident hot spots shows that, in the CBD, number of observed hot spots is higher than the expected when taken into consideration the total urbanized area in the zones. Especially, pedestrian accident hot spots are overestimated. Since CBD is an employment center and attracts far more trips per square kilometer than any of the other zones, mobility of pedestrian activities and traffic volume is higher when compared to the other zones. Furthermore, there are more conflict points in the CBD than in zone 2, which, in turn has more conflict points than in zone 3 which, finally, has more conflict points than in zone 4. In order to provide pedestrian safety, regulations could be performed in CBD. Planning decisions should be determined and improved taken into consideration the pedestrian priority in the traffic. To do this, pedestrianization of CBD, Kızılay and Ulus, should be considered. Outer the CBD, some regulations could be done to reduce speed such as decelerator or rumble strip.

7.3. Further Recommendations

Further Recommendations to this study can be summarized as follows:

- The most important and missing point of this study is the lack of exposure data, the traffic counts at the intersections and along the road corridors. Because of this, it was never possible to calculate normalized values (per one million entering vehicles or per 100-million vehicle-km) and rank the accident risk at different regions. In this study, distribution of accident and their hot spots are related to the geographical area. The best test would have been to relate the accidents and hot spots Vehicle-Kilometer-Traveled (VKT) as one would expect there to be much higher traffic volume in the CBD than in the periphery.
- Urban zones can be divided into sub-zones to understand the relation between accident occurrence and land use. Residential areas, commercial areas, educational areas, industrial areas, non-urban areas etc. can be defined more specifically to understand the relationship between traffic accident hot spots and land use. Furthermore, if available, land use information should be used. Analyzing hot spots with land use information could be more informative for traffic safety officials.
- Intersections can be grouped into design type (roundabout, four-leg intersection, three-way intersection, multi-way intersection etc.) to understand the relationship between accident hot spots and geometrical design of intersections. Moreover, if available, information of the intersections in terms of signalization could be used to understand the effect of traffic lights on the occurrence of traffic accidents. Road type, number of lanes could be used to understand the relationship between the

hot spots and roads. Speed status of the road segments can be used when analyzing the hot spots.

- Other than intersections, different point of interests can be used to find causality between hot spots and urban built environment components – school locations, hospitals, bus stops, ambulance locations, etc.
- Urban hot spot detection can be further expanded to study other types of accidents more specifically. It is possible to detect hot spots with NNH clustering algorithm for each type of accidents.
- Severity of hot spots can also be defined by considering not only the number of accidents in it but also the total number of injuries and fatalities.

REFERENCES

Affum, J. K., & Taylor, M. A. (1998). “Selatm—a GIS based program for evaluating the safety benefits of local area traffic management schemes”. *Transportation Planning and Technology*, 21(1-2), 93-119.

AASHTO (2004), *A policy on geometric design of highways and streets*, Washington , DC .

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. Essex: Longman Scientific & Technical.

Blazquez CA, Celis MS. 2013. “A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile”. *Accident Analysis and Prevention,* 50: 304-311.

Chainey, S., Tompson, L. and Uhlig, S. (2008). “The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime”. *Security Journal.* Vol.21, 4-28.

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,* 14, 425-436.

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

Priyantha Wedagama, D. M., Bird, R. N., & Metcalfe, A. V. (2006). The influence of urban land-use on non-motorised transport casualties. *Accident Analysis & Prevention,* 38(6), 1049-1057.

Dai, D., (2012).”Identifying clusters and risk factors of injuries in pedestrian – vehicle crashes in a GIS environment”, *Journal of Transport Geography,* 24, 206-214.

Eck, J., Chainey, S., Cameron, J., & Wilson, R. (2005). Mapping crime: Understanding hotspots.

Eckley, D. C., & Curtin, K. M. (2012). “Evaluating the spatiotemporal clustering of traffic incidents”. *Computers, Environment and Urban Systems*.

Erdogan, S., Yilmaz, I., Baybura, T., Gullu, M.,(2008), “Geographical Information Systems Aided Traffic Accident Analysis System Case Study City of Afyonkarahisar”, *Accident Analysis and Prevention*, Vol. 40/1, 174-81.

Ha, H. H., & Thill, J. C. (2011). “Analysis of traffic hazard intensity: A spatial epidemiology case study of urban pedestrians”. *Computers, Environment and Urban Systems*, *35*(3), 230-240.

Hart, T. C. & Zandbergen, P. A. (2012). “Effects of Data Quality on Predictive Hotspot Mapping”. Publication 239861, National Criminal Justice Research Service: Washington, DC.

Hirasawa, M., & Asano, M. (2003, October). “Development of traffic accident analysis system using GIS”. *In Proceedings of the Eastern Asia Society for Transportation Studies*, (Vol. 4), 1193-1199.

Hyder, A.A. and D. Bishai (2012) Road Safety in 10 Countries: A Global Opportunity, *Traffic Injury Prevention*, 13(S1):1–2.

IBM (ed.) (2011). *IBM SPSS Statistics Base 20*.

Kaufman, J. (2008). "Creating a safer Houston through Crash Mapping". *Geography and Public Safety Bulletin*, 1 (2), 2-5.

Kaygısız, Ö. “Kentsel Arazi Kullanımı Trafik Kazas İlişkisi:Eskişehir Örneği”, Phd Thesis, Gazi University, Ankara, 2012

Kaygısız, Ö., Düzgün, H.Ş., Akın, S. and Çelik, Y. “Coğrafi Bilgi Sistemleri Kullanarak Trafik Kazalarının Zamansal ve Mekansal Analizi”. EGM and ODTÜ, Ankara, 2012

Keskin, F., Yenilmez, F., Colak, M., Yavuzer, I., Düzgün, H.S., (2011)."Analysis of traffic incidents in METU campus", *Procedia - Social and Behavioral Sciences*, 19, 61-70.

Kim, K., & Levine, N. (1996). “Using GIS to improve highway safety”. *Computers, environment and urban systems*, *20*(4), 289-302.

Kim, K., & Yamashita, E. (2002). “Motor vehicle crashes and land use: empirical analysis from Hawaii”. *Transportation Research Record: Journal of the Transportation Research Board*, *1784*(1), 73-79.

Kim, K., Brunner, I. M., & Yamashita, E. Y. (2006). “Influence of land use, population, employment, and economic activity on accidents”. *Transportation Research Record: Journal of the Transportation Research Board*, *1953*(1), 56-64.

Krishnakumar, V. K., Pulugurtha, S. S., & Nambisan, S. S. (2005, July). “Identification and ranking of high pedestrian crash zones using gis”. In *2005 Annual ESRI International User Conference, San Diego, California*.

Levine, N. (2006). Houston, “Texas, metropolitan traffic safety planning program”. *Transportation Research Record: Journal of the Transportation Research Board*, *1969*(1), 92-100.

Levine, N. (2009). “A Motor Vehicle Safety Planning Support System: The Houston Experience”. In *Planning Support Systems Best Practice and New Methods* (pp. 93-111). Springer Netherlands.

Levine, N.(2008) The “Hottest” part of a hotspot: Comments on “The Utility of hotspot mapping for predicting spatial patterns of crime.” *Security Journal*,21, 295-302

Levine, N., 2010. CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 3.3). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC.

Levine, N., Kim, K., Nitz, L., (1995a). "Spatial Analysis of Honolulu motor vehicle crashes: I. Spatial Patterns", *Accident Analysis and Prevention*, 27 (5), 663-674.

Levine, N., Kim, K., Nitz, L., (1995b). "Spatial Analysis of Honolulu motor vehicle crashes: II. Zonal Generators", *Accident Analysis and Prevention*, 27 (5), 675-685

Lightstone, A. S., Dhillon, P. K., Peek-Asa, C., & Kraus, J. F. (2001). A geographic analysis of motor vehicle collisions with child pedestrians in Long Beach, California: comparing intersection and midblock incident locations. *Injury prevention*, *7*(2), 155-160.

Lloyd, C. D. (2010). Spatial data analysis: an introduction for GIS users. Oxford University Press.

Loukaitou-Sideris, A., Liggett, R., Sung, H., (2007). "Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles", *Journal of Planning Education and Research*, 26, 338-351.

Moreira, I., Gonçalves, A., Shrusbsall, S. (2012). “Planning on safer urban mobility through spatial analysis: Application to Vila Real, Portugal”, AESOP 26th Annual Congress, METU, Ankara.

Okabe, A., Okunuki, K., & Shiode, S. (2006). “SANET: A toolbox for spatial analysis on a network”. *Geographical Analysis*, 38, 57–66.

Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011), “Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment”, *Procedia - Social and Behavioral Sciences* 21, 317-325.

Pulugurtha, S., and Vanapalli V. (2008). “Hazardous bus stops identification: An illustration using GIS”. *Journal of Public Transportation* 11, 65-83.

Pulugurtha, S.S., Krishnakumar, K.V., Nambisan, S.S., (2007). “New methods to identify and rank high pedestrian crash zones: An Illustration”, *Accident Analysis and Prevention*, 39(4), 800-811

Selvi, Ö. “Traffic Accident Predictions Based on Fuzzy Logic Approach for Safer Urban Environments, Case Study: İzmir Metropolitan Area”, Phd thesis, İzmir Institute of Technology, Turkey, 2009.

Sjolinder, K., Ek, H.,(2001). *Black Spot Manual*, Ankara: General Directorate of Highways

Troung, L.T.and Somenhalli, S. (2011).“Using GIS to identify pedestrian-vehicle crash hot spots and unsafe bus stops”. *Journal of Public Transportation,* 14, 99-114.

Van Pattern, J. McKeldin-Coner, D.Cox (2009). “A microspatial analysis of robbery: Prospective hot spotting in a small city”. *Crime Mapping: A Journal of Research and Practice,* 1(1), 7-32.

World Health Organization.(2013) “*Global Status Report On Road Safety 2013: Supporting a Decade for Action*,”, Luxembourg

Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems, 32*(5), 396-406.

Yetiskul, E., & Senbil, M. (2012). “Public bus transit travel-time variability in Ankara (Turkey)”. *Transport Policy*, *23,* 50-59.

Web 1: *Pedestrian and Bicycle Crash Analysis Tool (PBCAT).* Retrieved from http://www.pedbikeinfo.org/pbcat_us/index.cfm Accessed December, 2013

Web 2 : Turkish Statistical Institute. *Population and Demography; Population of Provinces by Years*. Retrieved from http://www.turkstat.gov.tr/UstMenu.do?metod=temelist Accessed July, 2013

Web 3: Turkish Statistical Institute. *Seçilmiş Göstergelerle Ankara 2012*
Retrieved from http://www.tuik.gov.tr/ilGostergeleri/iller/ANKARA.pdf
Accessed December 2013