GEOSIMULATIVE APPROACH TO ROAD ACCIDENT ANALYSIS

THESIS SUBMITTED TO

Symbiosis Institute of Geoinformatics

FOR PARTIAL FULFILLMENT OF THE M.Sc. DEGREE

BY

ANDRIE SINGH

PRN 11030342007

Batch 2011-2013

Symbiosis Institute of Geoinformatics

Symbiosis International University

5th & 6th Floor, Atur Centre, Gokhle Cross Road,

Model Colony,

Pune – 411016

CERTIFICATE

Certified that this thesis titled **'GEOSIMULATIVE APPROACH TO ROAD ACCIDENT ANALYSIS** is a bonafied work done by **Miss ANDRIE SINGH** '**NAVTEQ India Private Ltd., Mumbai'** and Symbiosis Institute of Geoinformatics, under our supervision.

Supervisor Internal   **Supervisor External**

Dr. T.P. Singh   Mr Sarvang Patel
Associate Professor   Project Quality Leader
Symbiosis Institute of Geoinformatics   NAVTEQ India Private Ltd.

Pune   Mumbai

DECLARATION

I hereby declare that I am the sole author of this thesis. I also certify that although I may have conferred with others in preparing this assignment, and have drawn upon the range of sources cited in this work, but the content of this project is my original work.

SIGNATURE:

DATE:

PLACE: **NAVTEQ India Private Ltd., Mumbai**

CONTENTS

ACKNOWLEDGEMENTS

I express my earnest thanks to Mr Kapil Chhabra, Director, India Production Centre, NAVTEQ India, Mumbai for allowing me to take up this project.

My sincere gratitude to Mr Rupesh Wadhwa and Mr Basant Rajput, Production Manager I, NAVTEQ India, Mumbai and also Mr Kevin Cardozo, Project Leader, NAVTEQ India, Mumbai for their valuable suggestions during the internship period.

I am grateful to Brig. K. K. V. Khanzode, Director Symbiosis Institute of Geoinformatics for providing me the support and guidance

I would also like to pay my gratitude to Mr Sarvang Patel (Project Quality Leader) and Anshu Khare (Project Leader) for their vital role during the course of the project.

I gratefully thank Dr. T.P. Singh for his continuous support and help provided by him for the successful completion of this project.

I would also like to thank all the staff members and the faculty members of Symbiosis Institute of Geoinformatics, Pune, for their support and friendly behaviour without which it would have been impossible to complete the project.

Andrie Singh

CORPORATE PROFILE

Nokia is a Finnish multinational communications corporation. It is primarily engaged in the manufacturing of mobile devices and in converging Internet and communications industries. They make a wide range of mobile devices with services and software that enable people to experience music, navigation, video, television, imaging, games, business mobility and more.

Nokia Location & Commerce business unit, which comprises its entire NAVTEQ digital mapping business, leads the development of a "new class of integrated social location products and services for consumers, as well as platform services and local commerce services for device manufacturers, application developers, internet services providers, merchants and advertisers".

NAVTEQ India (subsidiary of Nokia Location & Commerce) is a Chicago-based provider of Geographic Information Systems (GIS) data and is a dominant company in providing the base electronic navigable maps. The company is a wholly owned subsidiary of Nokia but operates independently

NAVTEQ India (subsidiary of Nokia Location & Commerce) is headquartered in Chicago, Illinois (USA) with more than 4,400 employees worldwide. With a regional European headquarters located in Veldhoven, Netherlands, a major production facility in Fargo, North Dakota, USA, and support centers in Seoul, South Korea, Mexico City, Mexico and in Yokohama, Japan

LIST OF FIGURES

• • •

LIST OF TABLES

LIST OF EQUATIONS

ABBREVIATION LIST

AOI: Area of Interest

CSS: Cascading Style Sheet

DEM: Digital Elevation Model

GDB: Geodatabase

GIS: Geographic Information System

GPS: Global Positioning System

GUI: Graphical User Interface

GWR: geographic weighted regression

HTML: Hyper Text Markup Language

IRF: International Road Federation

KDE: Kernel Density Estimation

LULC: Land Use and Land Cover

MUIP: Mumbai Urban Infrastructure Project

MUTP: Mumbai Urban Transport Project

OLS: ordinary least squares

OSPF: Open Shortest Path First

PARCLO: Partial Cloverleaf

POI: Point of Interest

SRTM: Shuttle Radar Topography Mission

TIN: Triangular Irregular Network

PREFACE

Accidents today are among the leading causes of death, in some cases the number one cause, particularly the more industrialises nations. The number of minor as well as serious injuries and the human suffering and economic loss due to disabilities caused by accidents is inestimable. Thus while medical science has conquered the ravages of many diseases, accidents have become a new "epidemic" of health importance, calling for equal effort for control and prevention

Road accidents, can be a big threat to the country, and has been a major concern for the population and the government to reduce its occurrence .Millions are spent in order to reduce the accident occurrence through several modes of campaign. Unfortunately, over the years it keeps on increasing. Thus, by the use of an intelligent and effective Road Navigation System, the police would be able to greatly control and manage the whole accident events. This system further has an advantage of providing comprehensive intelligent GIS-based solutions for accident analysis and management.

Road traffic accident is complicated to analyse as it crosses all the boundaries of engineering, geography or human behaviour. Therefore, there rises a need for a more systematic approach, which can automatically detect the surface topography (curvature and gradient) and the statistically significant spatial accident clusters. To implement countermeasures efficiently and effectively, it is important to identify the accident prone locations and analyse the accident patterns, so that the most appropriate measures can be taken for each of these specific location.

This system is solely based on GIS technologies. This intelligent and effective Road Navigation System is used to get better results from accident data, which will include the most effective and useful queries and reports. This work focuses and evaluates the performance of a GIS based solution in order to integrate the Navigation Technology and Safety.

OBJECTIVE

The objective of the project is to forecast the accident clusters with the aspect of road navigation based on its curvature and gradient. Also understand the underlying parameters for the increasing frequency of the same on the Area of Interest.

An analytical study will be conducted based on the width, gradient and curvature of the road, as on hilly terrains, the gap between the hill and the edge of the road becomes a pivotal factor for the identification of the blindspot areas and the accident clusters.

Although all the concerns are not directly addressed but few of the problems faced by the driver on the road has been identified. The list given below highlights the same:

- Unaware of the blindspots present along the course of the road
- Accident zones not known
- Are the Zones lying in the Range of the Service Areas of Hospitals and Police Stations.

The role of the Navigation System comes into play when the driver is assisted with the semi-automated navigator which gives a pop up as the slope and curvature of the road changes, and prompts the driver to control his speed and thus reduces further hassles.

ABSTRACT

Recent report by WHO indicate: (a) India having the highest number of fatal accidents, and (b) low and middle income countries accounting for 90% of accidents. The concern, regarding road safety in a developing country, like India, further amplifies in the absence of proper framework to analyze the accident related data using modern tools and techniques.

*This project undertaken at NAVTEQ, Mumbai and Symbiosis Institute of Geoinformatics, Pune has the potential to develop an API for drivers based on safe driving perspective. Furthermore, generation of map layout will be carried out which will highlight the blindspots and the accident prone zones. For this reason, a motorway located at **Thane-Ghodbunder Road (State Highway 42)** has been selected as our **AOI** (Area of Interest), which will be my case study to determine the feasibility of this system. This project uses ArcGIS to combine the road network data, the LANDSAT image, the SRTM data and the toposheet as a raster GIS plot to be displayed. ArcGIS has been further used for the generation of slope, aspect and gradient and ERDAS has been used for the process of layer stacking and the classification of the image. Visual Studio has even been used for developing the webform and the API.*

This project talks about the road accidents caused due to blindspots or greater road curvature present along the stretch. An effort has been made to study in the detail the road accidents that have taken place along this terrain. A ground truth validation has also been conducted which gives an insight to the described situation.

***Keywords:** Road Accidents, Blindspots, Curvature, GIS, Accident mapping*

CHAPTER 1 – INTRODUCTION

1.1 Problem Statement

"To realize the value of one second; ask a person who has survived an accident". In 2011, more than 20,800 people of Mumbai had lost their lives in road accidents. In Maharashtra, there is a highway related death every 15-20 minutes and an injury due to the same in every 5 minutes. Highway crash related deaths and injuries are a major concern for the government, despite the progress that has been made in the past 40-50 years to improve the transport safety through safer vehicles and roadsides and more responsible drivers. Although there has been a drop in this fatality rate but it still remains an on-going challenge for us to do more to improve this safety.

Over the years there has been a tremendous change in the mobility habits. Culture, economy and technology has built a new society based on a new pattern of accessibility, integration of various networks, faster and more efficient and effective transportation and has resulted in changes in the way we animate and move.

Automobiles have become a necessity from luxury and have thus resulted in a rapid increase in road accidents. The transience rate has reached unbelievable numbers in the recent time and is of utmost importance in terms of projects and research purpose.

An automotive navigation system is a satellite navigation system designed for use in automobiles. It typically uses a GPS navigation device to acquire position data to locate the user on a road in the unit's map database. Using the road database, the unit can give directions to other locations along roads also in its database.

As world is grooming, new and versatile technologies and techniques in navigation system are budding. The year 2012 has been marked very many novelty functionalities being incorporated in the existing navigation system, so that it can minimize the probable occurrence of these accidents.

1.2 Background

The current scenario of Mumbai has limitations when analysing crashes on multilane roads such as motorways (highways). The accomplishment of reducing these fatalities and injuries can greatly be achieved through a combination of efforts, including a proper use of road design, improving the driver's perception and their willingness to obey the safety norms. Despite these improved efforts the annual toll of loss of human life has not changed substantially in all these years. Vector GIS has been used in displaying these accident spots. Furthermore, raster GIS was used to analyse this spatially. It started with mapping and georeferencing the map within the limits of the Mumbai-Thane district.

The spatio-temporal analysis of road accidents aims for the identification of statistical relationships between accidents and their occurrence in our AOI, Thane-Ghodbunder Road. So, by using integrated statistical analysis techniques in GIS, the data has been analysed, as well as an insight has been obtained to how this are distributed in the specified region. Hence, by means of spatial statistics the creation of a pertinent data analysis was possible, to evaluate the tendencies and relationships between the existing data. The aim of this study is to outline the importance and efficiency of using GIS in collecting, processing and analysing data. It is an efficient tool for displaying the spatial distribution of accidents along the road network. GIS enables the rapid processing and display on the map of the information regarding accidents and its use in making decisions regarding the reduction and prevention of accidents.

CHAPTER 2 – REQUIREMENTS

2.1 Softwares

2.1.1 ArcGIS 10.0

ESRI's ArcGIS is a software used for working with maps and spatial information. It is used for creation of map, analysing those mapped information and discovering and managing these geographic information in the database. It provides a structure wherein these GIS resource, such as maps, imagery, geodatabases and tools, can be delivered as a web service.

Figure 1 ArcGIS Logo

ArcGIS is a platform for designing and managing solutions through the application of geographic knowledge. It enables you to perform deep analysis, gain a greater understanding of your data, and make more informed high-level decisions.

It is built around geodatabases, which use object-relational database approach for storing spatial data. This GDB contains topology information and can also model behaviour of features. ArcGIS is accessed using one or three software, each varying in their functionality.

- **ArcView**: Provide comprehensive mapping and analysis tool, with simple editing and geographic capabilities.
- **ArcEditor**: It provides all ArcView functionalities along with few advanced editing capabilities.
- **ArcInfo**: It not just provides functions mentioned in the above two, but also advanced geographic capabilities and data management as well
- **Map Appreciation**: It is used for organizing and managing the database. It also allows us to preview the dataset and view the metadata.
- **ArcMap**: It is most commonly used. It views and analyses spatial data and also helps in creating map layouts.
- **ArcReader**: It allows one to view and query maps created with the other ArcGIS products.
- **ArcScene**: It is a 3D visualization application that allows viewing GIS data in three dimensions. It permits overlaying many layers of data in 3D environment.

2.1.2 ERDAS Imagine 2011

ERDAS has been supplying image processing software since 1978. Their flagship desktop package, ERDAS IMAGINE, led the way in image processing and raster data handling. Over time this software has evolved from command line prompt to the ribbon interface in the latest version. Along the way, much additional functionality has been added that enables the user to extract geospatial information from the raw satellite or aerial imagery.

Figure 2 ERDAS Logo

Other products have been incorporated into the product suite over time. LPS is a complete suite of photogrammetric production tools while ERDAS ER Mapper complements the image processing functionality of ERDAS IMAGINE. It is built on v2010, the most significant release of the software since 1992 when it evolved from a command line interface. The new 2011 version has a Microsoft Ribbon style user interface, providing a modern look and feel, improving usability and productivity while reducing the learning curve. No other software tool on the market is as powerful when it comes to raster image processing.

There is also a range of classification, feature extraction and change detection options available, both from within the core software and from add-on modules. IMAGINE also comes equipped with inbuilt connectivity to support enterprise geospatial image processing, enabling client side interaction with ESRI ArcSDE and Oracle Spatial 10g.

ERDAS IMAGINE provides true value, consolidating remote sensing, photogrammetry, LiDAR analysis, and radar processing into a single product. We offer many solutions in one, incorporating the following standards, enterprise capabilities and products. It is a scalable solution that will grow with time. With three tiers and a multitude of add-on modules, we can easily tailor the system to meet the required specification.

It includes a complete set of tools to analyse data from the imagery via mosaicking, surface interpolation, advanced image interpretation and orthorectification. It includes analysis of multispectral and hyperspectral images and also includes graphical spatial data modeling.

2.1.3 HTML & JAVASCRIPT

HyperText Markup Language (HTML) is the main markup language for creating web pages and other information that can be displayed in a web browser. HTML is written in the form of HTML elements consisting of tags enclosed in angle brackets (like <html>), within the web page content. HTML tags most commonly come in pairs.

The purpose of a web browser is to read HTML documents and compose them into visible or audible web pages. The browser does not display the HTML tags, but uses the tags to interpret the content of the page. HTML elements form the building blocks of all websites. HTML allows images and objects to be embedded and can be used to create interactive forms. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. It can embed scripts written in languages such as JavaScript which affect the behaviour of HTML web pages.

JavaScript (JS) is an interpreted computer programming language. It was originally implemented as part of web browsers so that client-side scripts could interact with the user, control the browser, communicate asynchronously, and alter the document content

Figure 3 JavaScript & HTML Logo

that was displayed. JavaScript is a prototype-based scripting language that is dynamic, weakly typed, and has first-class functions. Its syntax was influenced by the language C. It is a multi-paradigm language, supporting object-oriented, imperative, and functional programming styles.

JavaScript is a scripting language and is dynamic in nature. It was originally developed in Netscape by Brendan Eich. It is a light-weight, interpreted, object-oriented language with first class functions but used in many non-browser environments. The JavaScript standard is ECMAScript. Objects are created programmatically in JavaScript, by attaching methods and properties. Once an object has been constructed it can be used as a blueprint (or prototype) for creating similar objects.

2.1.4 Microsoft Office 2010

Microsoft Office is an office suite of desktop applications, servers and services for the Microsoft Windows and OS X operating systems, introduced by Microsoft on August 1, 1989. Initially it was a marketing term for a

bundled set of applications. The first version of office

Figure 4 Microsoft Office Logo

contained Microsoft Word, Microsoft Excel and Microsoft PowerPoint. Over the years, Office applications have substantially grown. Presently it has been reported that Office is being used by over a billion people worldwide.

The components of Microsoft Office that is used for Desktop Applications are:

2.1.4.1 MSWord

It is a word processor and was previously considered as the main program in Office. Its proprietary DOC format is considered as *de facto* standard. Report Documentation is the part that is carried out best in MSWord and we have done it in the similar way.

2.1.4.2 MSExcel

Microsoft Excel is a spreadsheet program that originally competed with the dominant Lotus 1-2-3. It features calculation, graphing tools, pivot tables. For the purpose of this project, we have used excel to draw statistical diagram to represent the data obtained on accidents as well as the survey data.

2.1.4.3 MSAccess

It is a database management system that combines relational Microsoft Jet Database Engine with a graphical user interface and software development tools. It has been used by us for the process of database creation.

2.2 Tools

For the completion of this project we have conducted a ground survey for the purpose of obtaining the co-ordinates of the accident sites as well as the POIs that were selected for this (Police Stations and Hospitals).

For the very purpose, we have used a mid-level **GPS Navigation System – NAVON N485.**

It is a high-definition, true colour display with a touch input. It is a portable navigation system designed for a car electronics interface. It provides detailed map information of cities, voice navigation and automatic smooth switching. It has a high sensitivity GPS receiver and antenna inside and is faster and more accurate to receive signal and positioning. Its advantage is it takes 1 second to re-capture the satellite signal, with less than 10 m for positioning range and the time accuracy for that is 1 microsecond.

Figure 5 NAVON N485

Technical Parameters:

Processor: SiRF Atlas IV 500 MHz

OS: Microsoft Windows CE 6.0

Memory: 2GB ROM / 64 vagy 128 MB RAM

Display: 4.3'' TFT LCD, 480 x 272, 16 bit

Audio: built-in mono loudspeaker

Connectors: 3.5 mm ear-pieces, mini-USB

Battery: 1000mAh Lithium-ion

Size: 124 x 77 x 10 mm

Weight: 146 g

2.3 Development Platform

2.3.1 Hardware
Model: Dell T3400

Memory: 4 GB RAM

Storage: 250 GB Hard Disk

Processor: Intel(R) Core(TM) 2 Duo

2.3.2 Software
Operating System: Windows 7 Enterprise

Browser: Mozilla Firefox 20.0.1 and Google Chrome 26.0.1410.64

2.4 Development Technology
HTML5

CSS

JavaScript

developer.here.com JavaScript API explorer

The fact that Javascript is light, cross-browser, and plug-in-less and supports HTML5 made it a perfect choice for the project.

The system is aimed at being used by the commoner, so it should be able to handle large traffic and should run without any plug-in. Moreover, the additional features of HTML5 enabled to provide the user with a rich experience without requiring heavy plug-ins.

CHAPTER 3 – STUDY AREA

3.1 Area Description

Mumbai is the capital of Maharashtra and Maharashtra is the most populous state of the country. Along with the neighbouring areas of Navi Mumbai and Thane, it is the most populous region of the world. For the successful completion of this project, we have taken our AOI (Area of Interest) as **State Highway 42**, which is commonly known as the **Thane-Ghodbunder Road.** It is a state highway that runs through the entire district of Thane. It is one of the major roads connecting the Eastern Express.

It is approximately 20 km long and links the two main arterial roads of Mumbai, the Eastern Express. The road skirts the northern boundary of Salsette Island. The road is also considered of paramount significance not only because of its scenic beauty around Yeoor hills and Thane creek but also for the reason that it heads the list when it comes to the incidence of road accidents in/around Mumbai.
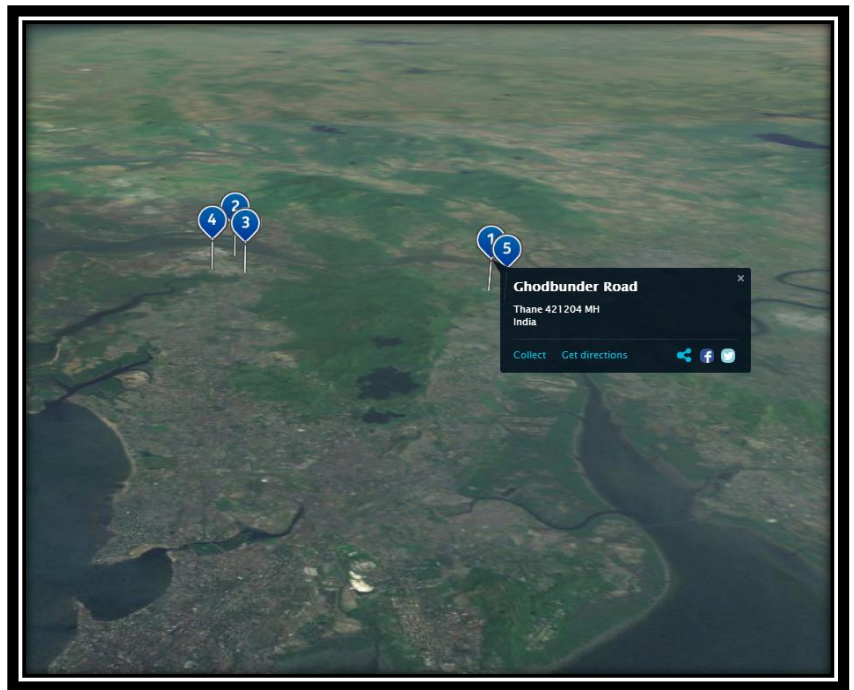
Figure 6 The AOI selected (Thane-Ghodbunder Road)

The road extents from 19° 17' 3.95" N, 72° 54' 16.05"E to 19° 13' 1.25" N, 72° 58' 41.55"E. We had selected the entire stretch which highlights the accident prone sites as well as the turns present along the same. The road is situated at an elevation of 180 m above MSL (mean sea level).

3.2 Traffic along Thane-Ghodbunder Road

Traffic congestion in the city has reached alarming levels. Absence of a good public transport system coupled with the increase in the number of cars on city roads has worsened the situation. Over the years, the state has taken various measures to regulate vehicular traffic in the city, but to no avail. To counter the traffic crisis in the city, the state government has launched the Mumbai Urban Infrastructure Project (MUIP) and the Mumbai Urban Transport Project (MUTP). Work is also underway on the metro rail and monorail projects. In addition to this, 50 skywalks are being built across the region. All these, once completed, will definitely help to curb the traffic congestion.

The vehicle population in Thane has reached 12.61 lakh, and is expected to increase every year. Expecting buoyancy in the sale of motor vehicles during 2011-12, the state transport department has estimated a 15.24% rise in tax collection. It has set a tax collection target of Rs. 4,000 crore for this fiscal. In 2010-11, the department garnered Rs. 3,471 crore in tax revenue.

The government statistics report reveal that the two-wheeler and car population still remains robust and constitutes more than 80% of the total motor vehicle population. As per this given statistics, it can be said that the city has witnessed an increase in the number of light motor vehicles. While the rising trend is a sign of increase in the buying power of citizens, it is likely to worsen the vehicular density on Thane roads.

3.3 Accidents along Thane-Ghodbunder Road

The Thane-Ghodbunder road also being a state highway is highly prone to any kind of road accidents. A survey conducted for this project give us a rough estimate to the actual statistics. The data for the last seven years (2007-2011) were gathered and based on that it can be concluded that the Road Accidents has had a decline of 0.77%, from 16.14% to 15.37%. Similarly, there has been a wan in the percentage of people killed and injured in these road accidents. The drop is a little more than 2% for the former and about 1.5% for the latter. But if we get to the depth of the situation, accident severity has increased from 19.9% to 26.9% (2002-2011). Accident severity is defined as the number of persons killed per 100 accidents.

EVIDENCES TO SUPPORT THE SELECTION OF THE AOI (THANE – GHODBUNDER ROAD)

THANE: A woman aged 33 was killed on Ghodbunder Road on Monday when a Thane Municipal Transport (TMT) bus ran over her.

Tejal Hadkar, a resident of Pride Park, Ghodbunder Road, was riding a scooty with her sister-in-law Arundhati as pillion. The two-wheeler skid, throwing off both. Before they could collect themselves, the TMT bus, which was close behind, ran over Hadkar. Arundhati, too, was hit, but escaped with injuries.

Ghodbunder Road is a part of the Mumbai-Ahemdabad highway, and therefore, experiences heavy traffic. Before its concretisation, motorists faced lot of difficulties while driving as the road was replete with potholes and experienced traffic jams frequently. Tatvdyan University, Gaimukh, Ovala, Kasarwadavli, Nagla-bunder are a few major accident-prone stretches on the road.

"The main cause of accidents is the carelessness on part of the drivers and the pedestrians," said DCP (Traffic) Dyaneshwar Chavan. Of the 180 cases of accidental death registered in Thane commissionerate this year till August — including Thane, Bhiwandi, Kalyan-Dombivli and Ulhasnagar zones, 60 victims were pedestrians.

Mumbai, September 26 Even three months after the concretisation of Ghodbunder Road by the Maharashtra State Road Development Corporation (MSRDC), there is no sign of decline in the number of road accidents. The road was open to traffic in June with much fanfare after concretisation. According to traffic police, there have been 24 cases of accidents between June and August resulting in 3 deaths and 35 injured persons. Also, seventy people were injured and 18 have died since January in accidents on Ghodbunder Road.

A Good Samaritan averted several likely accidents along Ghodbunder Road, Thane, on Tuesday evening. Though injured himself, Mandar Bedekar and his wife Shweta stood for two hours, signalling to bikers to avoid a slippery stretch. He returned home after dark, once he ensured that the authorities who arrived by then had cleaned the spot thoroughly.

Around 5 pm, a truck carrying slush passed through Ghodbunder, leaving a trail of its contents on the concrete road. Due to heavy rain, within minutes the road turned so slippery that several bikers skid on the kilometre-long stretch.

Ghodbunder Road is a very busy road as it also connects Thane to the Ahmedabad Highway. The road is a 14-kms stretch starting from Kapurbavdi. Most of the lanes are shut due to the ongoing flyover construction work. Due to this, the accident rate has gone up and many people have been knocked down by heavy vehicles while trying to cross the road.

CHAPTER 4 – ASSOCIATED TERMS

o **Area of Interest:** The extent used to define a focus area for either a map or database production.

o **Accident Sites:** The location of an unexpected occurrence, failure or loss, either at a plant or along a transportation route. To put in layman terms, it is an area which has faced an accident.

o **Autocorrelation:** The correlation or similarity of values, generally values that are nearby in a dataset. Temporal data is said to exhibit serial autocorrelation when values measured close together in time are more similar than values measured far apart in time. Spatial data is said to exhibit spatial autocorrelation when values measured nearby in space are more similar than values measured farther away from each other.

o **Base Map:** It is used for locational reference and often includes a geodetic control network as part of its structure. It is a map to which GIS data layers are registered and rescaled.

o **Blindspots:** It is defined as a part of an area that cannot be directly observed under existing circumstances. It is a portion of a field that cannot be inspected. In terms of telematics, it is an area to the side and slightly behind a driver's field of vision that is not reflected in the vehicle's rear-view mirror.

o **Contour:** Imaginary line on a map that connects points of equal elevation based on a vertical datum, usually sea level.

o **Crash:** To land in an abnormal manner, usually causing severe damage or destruction on impact.

o **Curvature:** Curvature is the second derivative of a surface, or the slope of the slope. This is most useful for identifying areas of rapid change in slope or aspect. The Curvature tool can be used to create three different curvature rasters – an output curvature raster, an optional profile curve raster, and an optional plan curve raster.

o **Data:** Any collection of related facts arranged in a particular format; often, the basic elements of information that are produced, stored, or processed by a computer.

o **Dataset:** Any collection of related data usually grouped or stored together.

- o **Digitise:** The process of converting the geographic features on an analog map into digital format using a digitizing tablet, or digitizer, which is connected to a computer.

- o **Dijkstra's Algorithm:** An algorithm that examines the connectivity of a network to find the shortest path between two points. Dijkstra's algorithm is named after the Dutch computer scientist Edsger Dijkstra.

- o **Geocode:** The process of assigning a street address to a location, or a code representing the location of an object, such as an address, a census tract, a postal code, or x ,y coordinates.

- o **Geodatabase:** A database or file structure used primarily to store, query, and manipulate spatial data. Geodatabases store geometry, a spatial reference system, attributes, and behavioural rules for data. Various types of geographic datasets can be collected within a geodatabase, including feature classes, attribute tables, raster datasets, network datasets, topologies, and many others.

- o **Georeference:** Aligning geographic data to a known coordinate system so it can be viewed, queried, and analyzed with other geographic data.

- o **Geostatistics:** A class of statistics used to analyse and predict the values associated with spatial or spatio-temporal phenomena. Geostatistics provides a means of exploring spatial data and generating continuous surfaces from selected sampled data points.

- o **GIS:** An integrated collection of computer software and data used to view and manage information about geographic places, analyze spatial relationships, and model spatial processes. A GIS provides a framework for gathering and organizing spatial data and related information so that it can be displayed and analyzed.

- o **Gradient:** The ratio between vertical distance (rise) and horizontal distance (run), often expressed as a percentage. A 10-percent gradient rises 10 feet for every 100 feet of horizontal distance. It is the rate of inclination of a surface in a given direction. The ascending or descending part of an incline. It is the degree to which something ascents.

- o **GUI:** Acronym for graphical user interface. It is defined as an interface between a user and a computer system that involves the use of a mouse- controlled screen cursor to select options from the menus, make choices with buttons, start programs by clicking icons etc. It is designed to standardize and simplify the use of computer programs.

- o **Hit-and-Run:** It is closely associated with road accidents, where the driver of the motor vehicle leaves the scene of the accident without stopping to give assistance. It is especially the case involving a pedestrian and another vehicle.

- o **Hotspots:** Hotspot analysis is the investigation of the location the most dangerous sites plausible to get affected by accidents. Hotspots adapt the geometrical characteristics and the environmental features of the problematic locations in the existing road network.

- o **HTML:** Acronym for Hypertext Markup Language. A markup language used to create Web pages for publication on the Internet. HTML is a system of tags that define the function of text, graphics, sound, and video within a document

- o **Incident:** It refers to something that takes place or comes to pass. It applies to a minor occurrence. It is an undesired event that may cause personal harm or other damage but on a small scale.

- o **Javascript:** A scripting language that runs within a web browser and interacts with HTML code to enable web developers to add functionality to their websites.

- o **Kernel:** On a raster, an analysis boundary or processing window within which cell values affect calculations and outside which they do not. Filters are used mainly in cell-based analysis where the value of a centre cell is changed to the mean, the sum, or some other function of all cell values inside the filter. A filter moves systematically across a raster until each cell has been processed.

- o **Kriging:** An interpolation technique in which the surrounding measured values are weighted to derive a predicted value for an unmeasured location. Weights are based on the distance between the measured points, the prediction locations, and the overall spatial arrangement among the measured points. Kriging is unique among the interpolation methods in that it provides an easy method for characterizing the variance, or the precision, of predictions.

- o **Lag:** In the creation of a semivariogram, the sample distance used to group or bin pairs of points. Using an appropriate lag distance can be helpful in revealing scale-dependent spatial correlation.

- o **LANDSAT:** Multispectral, earth-orbiting satellites developed by NASA that gather imagery for land-use inventory, geological and mineralogical exploration, crop and forestry assessment, and cartography.

o **Model Builder:** The interface used to build and edit geoprocessing models in ArcGIS.

o **Mortality:** It is the state or condition of being susceptible to death. It is also defined as the number of deaths in a given place, cause or time.

o **Mosaic:** A raster dataset composed of two or more merged raster datasets—for example, one image created by merging several individual images or photographs of adjacent areas.

o **Multimodal Network:** A network in which two or more types of transportation modes (such as walking, riding a train, or driving a car) are modeled.

o **Neighborhood:** On a raster, an analysis boundary or processing window within which cell values affect calculations and outside which they do not. Filters are used mainly in cell-based analysis where the value of a centre cell is changed to the mean, the sum, or some other function of all cell values inside the filter.

o **Network Analysis:** Network Analysis focuses on edge-node topology to represent real life networks of information, which may include: driving direction, multi-modal routing, service roads, information networks and few more. A series of nodes are connected to each other by edges. Each of these edges starts or ends with a node.

o **Nugget:** A parameter of a covariance or semivariogram model that represents independent error, measurement error, or microscale variation at spatial scales that is too fine to detect. The nugget effect is seen as a discontinuity at the origin of either the covariance or semivariogram model.

o **Overlay:** A spatial operation in which two or more maps or layers registered to a common coordinate system are superimposed, either digitally or on a transparent material, for the purpose of showing the relationships between features that occupy the same geographic space.

o **p-Value:** A probability resulting from a statistical test of the coefficient associated with each independent variable in a regression model. The null hypothesis for this statistical test states that the coefficient is not significantly different from zero. Small p-values reflect small probabilities. They suggest that the coefficient is significantly different from zero, and consequently, that the associated explanatory variable is helping to model or predict the dependent variable. Variables with coefficients near zero do not help predict or model the dependent variable.

- o **QQ Plot:** A scatter chart in which the quantiles of two distributions are plotted against each other.

- o **Risk:** It is the probability or threat of damage, injury or any negative occurrence that is caused by external/internal vulnerabilities, and that can be avoided through pre-emptive action.

- o **Routing:** Process of selecting paths in a network along which to send network traffic.

- o **Service Area:** A type of network analysis that determines the region that encompasses all accessible streets (streets that lie within a specified impedence).

- o **Shapefile:** A vector data storage format for storing the location, shape, and attributes of geographic features. A shapefile is stored in a set of related files and contains one feature class.

- o **Situational Awareness:** It is the perception involving being aware of what is happening in the vicinity, in order to understand one's actions and its impact on the goals and objectives, immediately or in the near future.

- o **Slope:** The incline, or steepness, of a surface. Slope can be measured in degrees from horizontal (0–90), or percent slope. A slope of 45 degrees equals 100 percent slope. As slope angle approaches vertical (90 degrees), the percent slope approaches infinity.

- o **SRTM:** The Shuttle Radar Topography Mission (SRTM) obtains elevation data on a near-global scale to generate the most complete high-resolution digital topographic database of Earth. SRTM consists of a specially modified radar system that flew on board the Space Shuttle Endeavour during an 11-day mission in February of 2000.

- o **Survey:** Measuring physical or geometric characteristics of the earth. Surveys are often classified by the type of data studied or by the instruments or methods used.

- o **Telematics:** It is an integrated use of telecommunications and informatics. Evolved to refer to systems used in automobiles that combine wireless communication with GPS tracking.

- o **Terrain:** It is a general term referring to the lay of the land. It is the vertical and horizontal dimension of land surface. Generally expressed in terms of elevation, slope and orientation of terrain features.

- o **Toposheet:** A rectangular map bounded by lines of latitude and longitude, often a map sheet in either the 7.5-minute or 15-minute series published by the U.S. Geological Survey. Quadrangles are also called toposheets.

- o **Validation:** It is a cognitive process of establishing a valid proof. An assessment of an action or decision plan to deduce that it is correct.

- o **Variogram:** A function of the distance and direction separating two locations that is used to quantify dependence. The variogram is defined as the variance of the difference between two variables at two locations. The variogram generally increases with distance and is described by nugget, sill, and range parameters. If the data is stationary, then the variogram and the covariance are theoretically related to each other.

- o **Verification:** It is the process of establishing the truth, accuracy or validity of something. It is like an additional proof attached to some hypothesis or theory.

- o **Voronoi Map:** A partition of space into areas, or cells, that surround a set of geometric objects (usually points). These cells, or polygons, must satisfy the criteria for Delaunay triangles. All locations within an area are closer to the object it surrounds than to any other object in the set. Voronoi diagrams are often used to delineate areas of influence around geographic features.

- o **Webpage:** A page of information stored on a website and viewed in a web browser. Web pages may contain text, graphics, animations, forms for data entry, and links to other webpages.

- o **Width:** It is the measurement or extent of something from side to side. In this project, it is in terms of the road width.

- o **z-Score:** A statistical measure of the spread of values from their mean, expressed in standard deviation units, where the z-score of the mean value is zero and the standard deviation is one. In a normal distribution, 68 percent of the values have a z-score of plus or minus 1, meaning they lie within one standard deviation of the mean. Z-scores are a common scale on which different distributions, with different means and standard deviations, can be compared.

CHAPTER 5 – ACCIDENT BASICS

5.1 What is an accident?

An accident is the final event in an unplanned process that results in injury or illness to an employee and possibly property damage. It is the final effect of multiple causes. An "event," occurs when one "actor" (one person/thing) performs an "action" (does something). In this definition, a person or thing will do something that result in a change of state. Accidents are processes that culminate in a final event that causes injury or illness.

5.2 Accidents and Incidents

Accidents are part of a broad group of events that adversely affect the completion of a task. These events are incidents. For simplicity, the procedures discussed in this course apply most appropriately to accidents, but they are also applicable to all incidents in general.

5.3 Accident Types

Accidents can broadly be classified under the following broad heads:

- Physical and Non-physical
- By activity
- By Vehicle

In the last few decades, road safety has become a high priority in many governments' policy-making, and it is now a critical item on the political agenda. Road mortality rates are very high around the world and its economic and social cost is a barrier to social development. In fact, road accidents are one of the main causes of death.
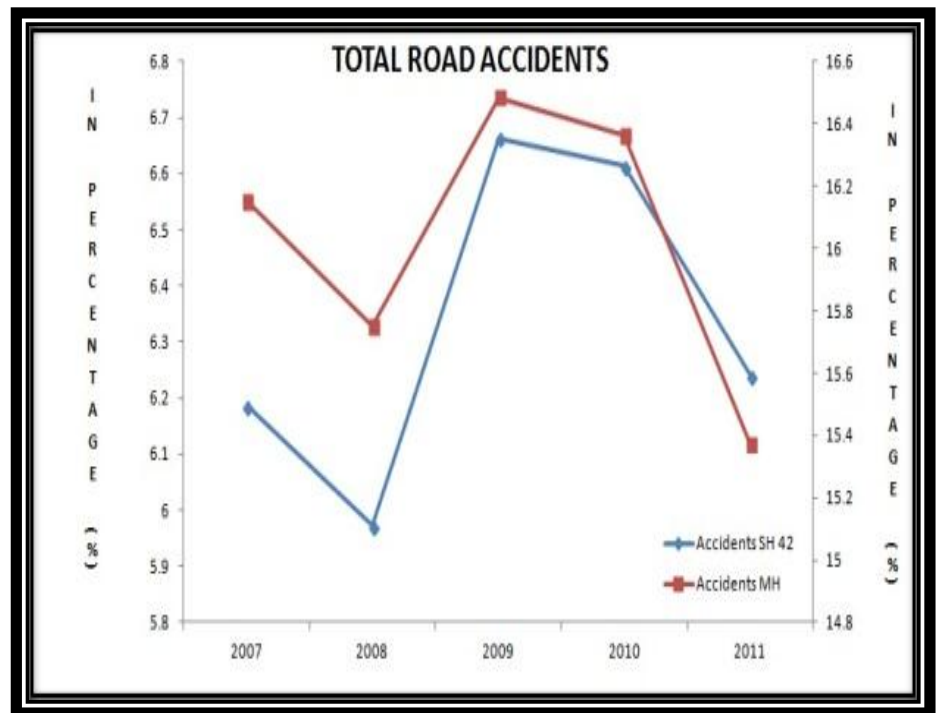
Figure 7 Total Accidents in Maharashtra and Thane-Ghodbunder Road

In 2011, India encountered 479216 road accidents of which 73661 had taken place in Maharashtra. The number of people killed and injured in these accidents was also large for both the country and the state.

Road accidents are complex to analyse as it crosses all the boundaries of geography. This project involves a presentation of a report with all accident spots along our AOI. The geographic information is a valuable asset but its usefulness can only be achieved when the data is gathered and analysed on a georeferenced perspective. But, the true meaning of the information comes to light only when it is displayed in the form of map layouts.

The problem of accident is a very acute in highway transportation due to complex flow pattern of vehicular traffic, presence of mixed traffic along with pedestrians. Traffic accident leads to loss of life and property. Thus the traffic engineers have to undertake a big responsibility of providing safe traffic movements to the road users and ensure their safety. Road accidents cannot be totally prevented but by suitable traffic engineering and management the accident rate can be reduced to a certain extent.

5.4 Objective of Accident Studies
Some of the objectives of accident studies are listed below:

- To study the causes of accident and suggest corrective measures at potential locations
- To evaluate the existing design
- To compute the financial losses incurred
- To support the proposed design and provide economic justification to improve the present scenario
- To carry out before and after studies and demonstrate the improvement in the problem

5.5 Accident Causes
Road accidents are a global tragedy with ever raising trend. They are common fare in our daily news as we see these accidents happen almost every day. Road accidents happen very quickly but usually leave a lasting and devastating effect. Due to these accidents, countless of individuals and families were left changing the direction that their lives are taking.

Personal injury sustained from a car accident is becoming a major tort case in the country. It is particularly dealt with by law to compensate the rising number of helpless victims who suffer injuries, disability and even death. Those who are involved may found themselves facing or pursuing civil or criminal suits.

The majority of road crashes are caused by human error. In 1985, a study by K. Rumar, using British and American crash reports as data, found that 57% of crashes were due solely to human factors, 27% combined both roadway and driver factors, 6% combined vehicle and driver factors, 3% solely to roadway factors, 3% to combined roadway, driver and vehicle factors, 2% solely to vehicle factors and 1% to combined roadway and vehicle factors. Later, these causes were generalized and were found to hold valid for any of the country.

The dictionary defines accident as "an unexpected and undesirable event, a mishap unforeseen and without apparent cause." Strictly speaking, most accidents are not accidents at all: they are collisions that could and should have been avoided. Accidents are caused by some person(s) failing to act safely, or to correct an unsafe condition. Negligence is a legal theory that is the basis for many car accident lawsuits, but often has more than one cause. Though there is no particular order to the different causes of accidents.

The main causes are mentioned below:

- **Human Factors:** Majority of road crashes are caused by human error. Research shows that driver error accounts for over 80% of all the fatal crashes on road.
    - Motor Vehicle Speed
    - Driver Impairment
        - Alcohol
        - Physical Impairment
        - Youth
        - Old Age
        - Sleep Deprivation
        - Drug Use
        - Distraction
        - Combination of Factors

- **Road Design**
    - Hazard Visibility
    - Road Surfaces
    - Traffic Control Devices and Traffic Flow
    - Weather
    - Behavioural Control Devices
- **Vehicle Design and Maintenance**
    - Design Defects
    - Tyre Blowouts
    - Mechanical Failure

5.6 Accident Data Collection

The accident data collection is the first step in the accident study. The data collection of the accidents is primarily done by the police. Motorist accident reports are secondary data which are filed by motorists themselves. The data to be collected should comprise all of these parameters:

- **General:** Date, time, person involved in accident, classification of the accident like fatal, senior, minor
- **Location:** Description and detail of location of accident
- **Details of Vehicle Involved:** Registration number, description of vehicle, loading detail, vehicular defects
- **Nature of Accident:** Detail of collision, damages, injury and casualty
- **Road and Traffic Condition:** Details of road geometry, surface characteristics, type of traffic, traffic density
- **Primary Cause:** Details of various possible cases which are the main causes of the accident
- **Accident Cost:** Financial losses incurred due to property damage, personal injury and casualty

CHAPTER 6 – LITERATURE REVIEW

6.1 Introduction

The purpose of the literature review is to identify and synthesize appropriate references to demonstrate and illustrate the presence and absence of knowledge and information regarding road accidents. These references include journal articles, conference papers, published technical reports and other readily available information on the World Wide Web and other sources. This literature review will, also present evidence that supports the need of this research.

6.2 Road Accidents: An International Perspective

The international traffic and highway system includes more than 22 million km of roads in service of the vehicular traffic and the road users, 470 million car passengers and 145 million users of different modes of transport. One fifth of the vehicles that use the roadway system belong to India.

It is illustrated that vehicular crashes represent one of the main causes of death having significant social extensions as well as substantial financial consequences. According to the *International Health Organization*, 600000 people are killed and 15 million are injured every year in crashes. In addition, the statistical analysis of the readily available crash data conducted by the *International Road Federation (IRF)* indicates that 320000 fatal crashes are occurring annually in 108 countries that belong to the five continents, which cooperate with the IRF. The countries that represent the highest fatality rate is China (>70000) followed by Russia (65000) and then **India (60000),** *Yves, 2001*.

6.3 Road Accidents and Statistics

Road crash analysis has been widely examined and various methods for identifying unsafe locations have been developed. Simple methods for identifying unsafe locations, where the number of crashes or the crash rate per unit exposure exceeds a given threshold, are routine and straightforward *(Taylor et al. 2000). Austroads (1988)* describes another method that uses critical crash rates to determine whether the crash record of each location is significantly greater than the system wide average. Empirical Bayes method, involve developing a statistical model based on reference population and comparing the expected number of crashes with the observed number *(Elvik 2008; Li and Zhang 2008).* In addition to the crash rates, unsafe locations are ranked according to their severity.

Severity Index for a particular location is calculated with the help of the formula given below:

$$SI = (3.0 * X_1) + (1.8*X_2) + (1.3*X_3) + X_4\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{ Equation 1}$$

Where,

X_1: total number of fatal crashes

X_2: total number of serious injury crashes

X_3: total number of other injury crashes

X_4: total number of property-damage-only crashes

However, these methods, along with other traditional methods, focus on road segments or specific locations and thus produce results that are partially dependent on the length of road segment *(Thomas 1996)* and might not be able to capture the wide crash hotspots *(Anderson 2009)*.

6.4 Road Accidents and GIS

In many developed countries, GIS has been widely used for mapping the accident locations. Many academicians and even government agencies are working on building new tools and improving the existing scenarios for road safety analysis. One such example can be found in one of the reports prepared by *Federal Highway Administration of US Department of Transportation* (*Smith et al. 2001*). Another such effort, to give an overview of road safety analysis methods, has been given by *Aylo (2010)*. The report contains an overview of statistical methods to analyze accident patterns. Even though statistical analysis has widely spread uses, safety analysis using only statistical method is not complete. This is mainly because, accident is a spatial phenomenon. Many spatial factors are taken into consideration for understanding the implication in the number of road accidents over the years. GIS is widely used to perform spatial data management and analysis operations.

Spatial patterns of accident crash data can also be analyzed by spatial autocorrelation, statistics that take into account simultaneously discrete event's locations and their values (*Long Tien Truong, Sekhar V.C. Somenahalli, 2011*).

Mandloi & Gupta (2003) has explained this in detail. Weights for each accident sites have been determined by giving weight to several parameters at a particular situation, with which the weight for the data point at that location can be determined. These methods were then compared to choose the one that best suits the requirements. One such comparative study was undertaken by *Anderson (2007). King Ivan and Ionel Haidu (2012)* greatly spoke on the spatio-temporal distribution of road accidents and also mentioned about their distribution with respect to the moments in time (hours, days and seasons).

Omer M. Soysal, Helmut Schneider (Louisiana State University, Baton Rouge, LA) discussed about the Zonal Statistics to identify hot regions of road accidents. They presented the use of ArcGIS zonal statistic tool for this hot region identification. This helped in identifying the areas with higher crash index.

6.5 Algorithms

6.5.1 Dijkstra's Algorithm

It was conceived by a Dutch computer scientist, Edsger Dijkstra in 1956 and was published in 1959, is a graph that solves the single-source shortest path problem for a graph with non-negative edge path costs, producing a shortest path tree. This algorithm is used in routing or in GPS Technology. It can be used for calculating the shortest route between two sets of points. The problem of finding shortest path plays a central role in the design and analysis of networks.

Finding the shortest route is one of the prime aims of our project, as it mainly deals with accidents and its important providing the shortest route from the accident spot to the nearest hospital or emergency centres. Dijkstra's algorithm is used in Internet's Open Shortest Path First (OSPF) routing procedure. The routing algorithm is present below in pseudo-code:

Given a network $G = (N, E)$, with a positive cost D_{ij} for all edges $(I, j \in N)$, start node S and a set P of permanently labelled nodes, the shortest path from the start node S to every node j is found as follows:

Initially $P = \{S\}$, $D_s = 0$, and $D_j = d_{sj}$ for $j \in N$
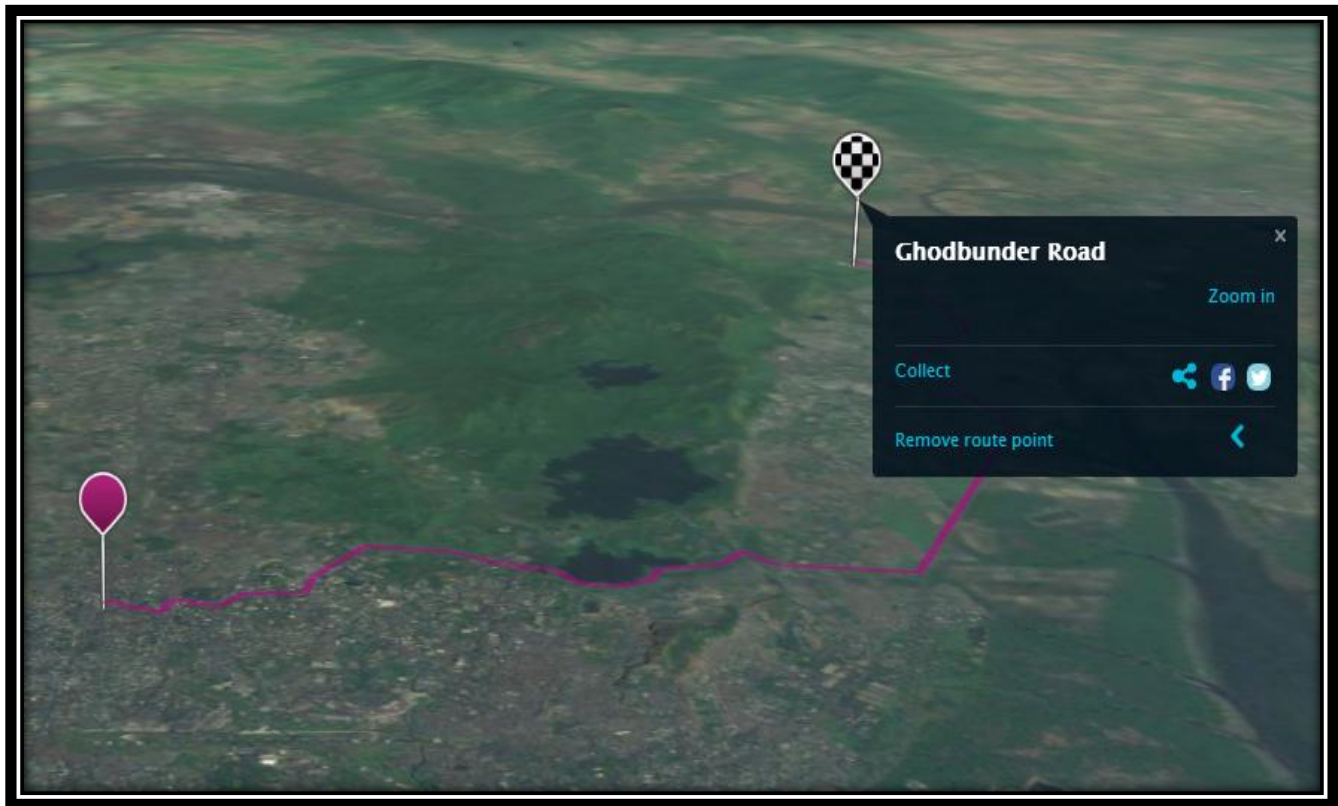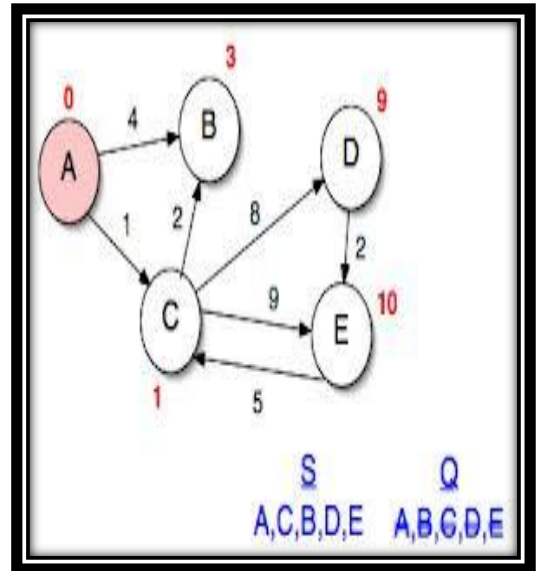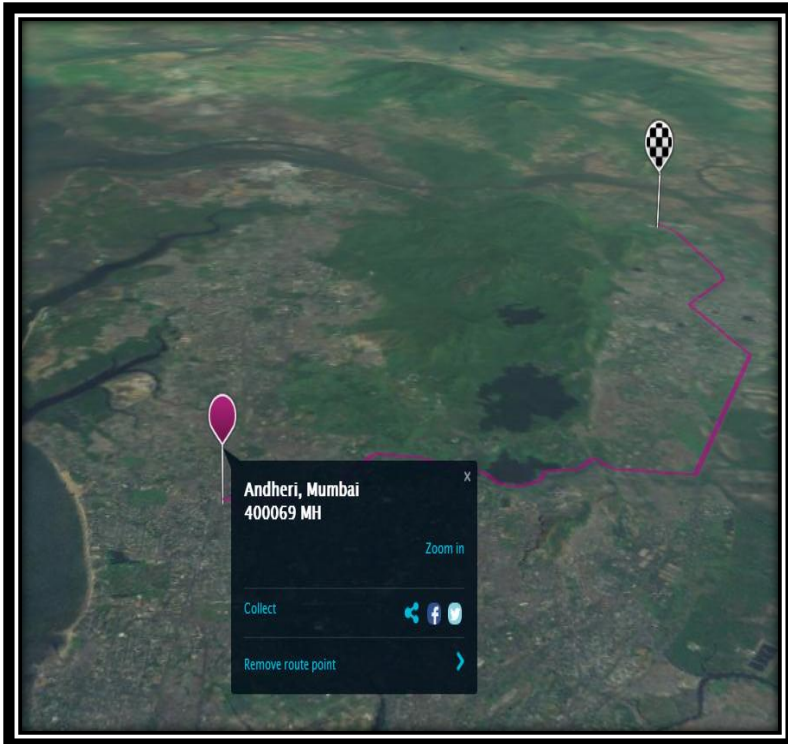
Figure 8 Dijktra's Algorithm

6.5.2 Horn's Algorithm

Slope measures the rate of change of elevation at a surface location. It may be expressed as percent slope or degree slope. Percent slope is 100 times the ratio of rise (vertical distance) over run (horizontal distance), whereas degree slope is the arc tangent of the ratio of rise over run.

The method of Horn (1981) is a modified version of the technique of Sharpnack and Akin (1969), using unequal weighting coefficients for the nearer elevation values. These weightings are proportional to the reciprocal of the square of the distance from the kernel centre. Horn's method is implemented in the "SLOPE" function of the ArcInfo GRID software package (ESRI, 1995). This method uses eight neighbouring cells and applies a weight of 2 to the four immediate neighbours and a weight of 1 to the four corner cells.

Horn's algorithm computes slope at C_0

$$S = \frac{\sqrt{[(e1+2e4+e6) - (e3+2e5+e8)]^2 + [(e6+2e7+e8) - (e1+2e2+e3)]^2}}{8d} \quad \ldots\ldots\ldots\ldots \text{ Equation 2}$$

And the D value at C_0 is computed by:

$$D = \arctan \frac{([(e6 + 2e7 + e8) - (e1 + 2e2 + e3)]}{([(e1 + 2e4 + e6) - (e3 + 2e5 + e8)]} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ Equation 3}$$

The accuracy of slope measures can influence the performance of models that use slope as inputs. Horn's algorithm is best for estimating slope in an area of moderate topography. Errors in slope estimates tend to be greater in areas of higher slopes. Data precision problems can be the cause for slope errors in areas of low relief.
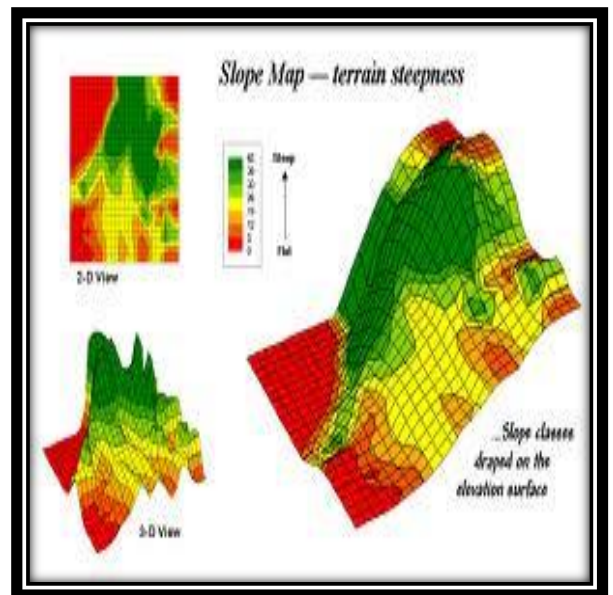
Figure 9 Slope Calculation

6.5.3 Zevenbergen and Thorn's Algorithm

Curvature is the second derivative of a surface, or the slope of the slope. This is most useful for identifying areas of rapid change in slope or aspect. The basic principle of curvature calculation, as with slope and aspect, is to pass a moving window over the elevation surface and fit the elevation values, the coefficients of which will yield the slope, aspect and curvature of the centre cell of the moving window.

A common algorithm is to fir a 3–by–3 window with a quadratic polynomial equation and is often referred to as the Zevenbergen and Thorne's Algorithm. It is given by:

$$z = Ax^2y^2 + Bx^2y + Cxy^2 + Dx^2 + Ey^2 + Fxy + Gx + Hy + I \ldots\ldots\ldots \text{Equation 4}$$

The coefficients *A-I* can be estimated by using the elevation values in the 3-by-3 window and the raster cell size. The measures of profile curvature can be computed from the coefficients. Profile Curvature is estimated along the direction of maximum slope. A positive curvature value at a cell means that the surface is upwardly convex at the cell location.
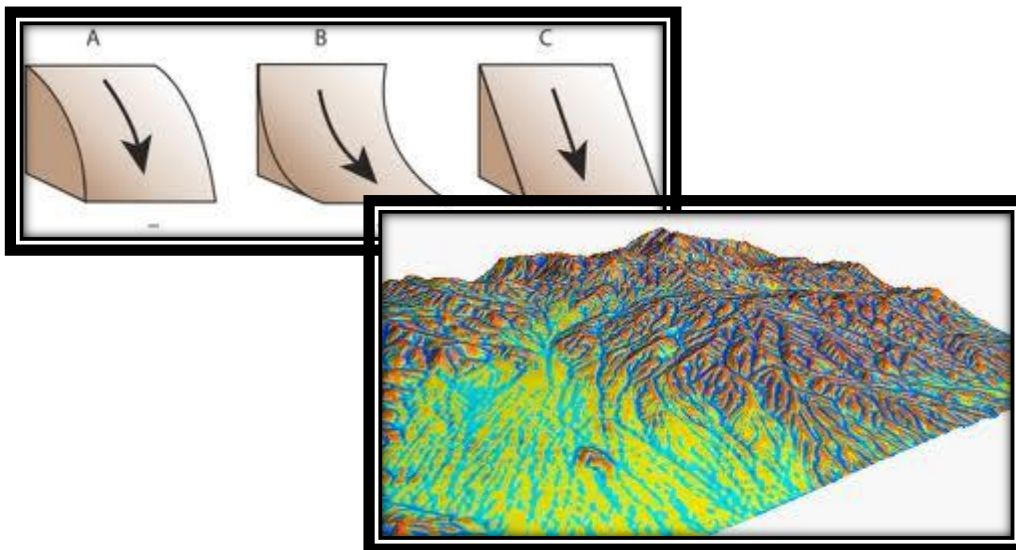
Figure 10 Curvature Calculation

CHAPTER 7 – ROAD GEOMETRY

7.1 Introduction

Geometric design for transportation facilities includes the design of geometric cross sections, horizontal alignment, vertical alignment, intersections, and various design details. These basic elements are common to all linear facilities, such as roadways, railways, airports etc.

Although the details of design standards vary with the mode and the class of facility, most of the issues involved in geometric design are similar for all modes. In all cases, the goals of geometric design are to maximize the comfort, safety, and economy of facilities, while minimizing their environmental impacts.

This chapter focuses on the fundamentals of geometric design, and presents standards and examples from different modes. This design of highway facilities deals with the proportion of physical elements of highways, lane widths, clearances, cross-section dimensions, etc.

Physical dimensions of geometric design elements are determined by:

- Characteristics of driver
- Characteristics of vehicle
- Characteristics of road

Factors Influencing Highway Design:

- Functional classification of highway being designed
  - ➢ Principal arterials
  - ➢ Minor arterials
  - ➢ Collectors
  - ➢ Local roads
- Expected traffic volume and vehicle mix
  - ➢ Volume: design for 30th highest hourly volume
  - ➢ Vehicle mix: proportion of passenger cars, heavy vehicles
- Design speed
- Topography of area in which highway will be located

- LOS to be provided

- Available funds

- Safety

- Social and environmental factors

7.2 The Geometric Cross Section

The primary consideration in the design of geometric cross sections for highways, runways, and taxiways is drainage. Details vary depending on the type of facility and agency. Highway cross sections consist of travelled way, shoulders (or parking lanes), and drainage channels.

- Shoulders are intended primarily as a safety feature. They provide for accommodation of stopped vehicles, emergency use, and lateral support of the pavement. Shoulders may be either paved or unpaved.

- Drainage channels may consist of ditches (usually grassed swales) or of paved shoulders

with berms or curbs and gutters.
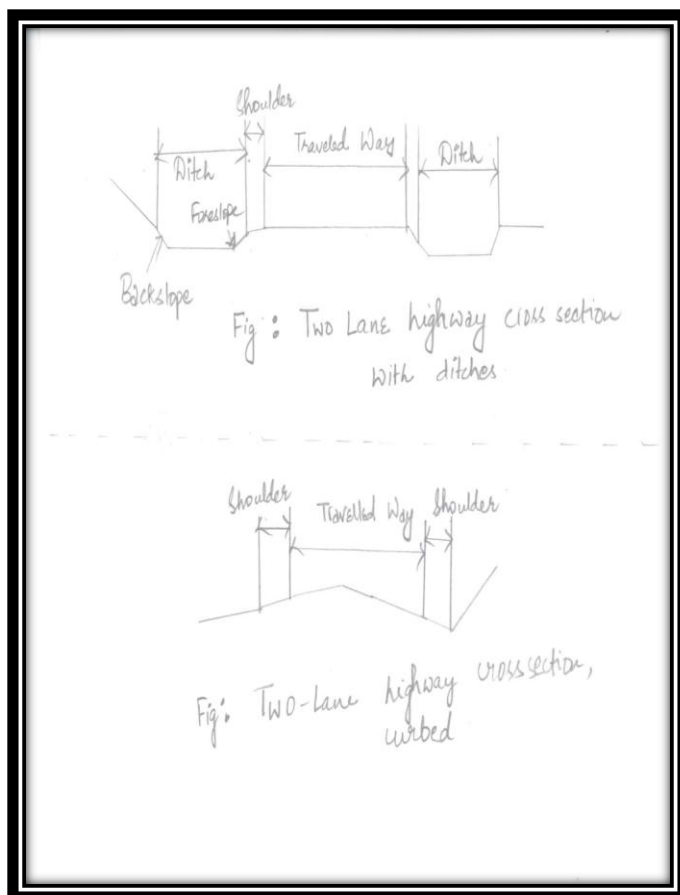
Figure 12 Two Lane Highway Cross Section   *49*   **Figure 11 Divided Highway Cross Section**
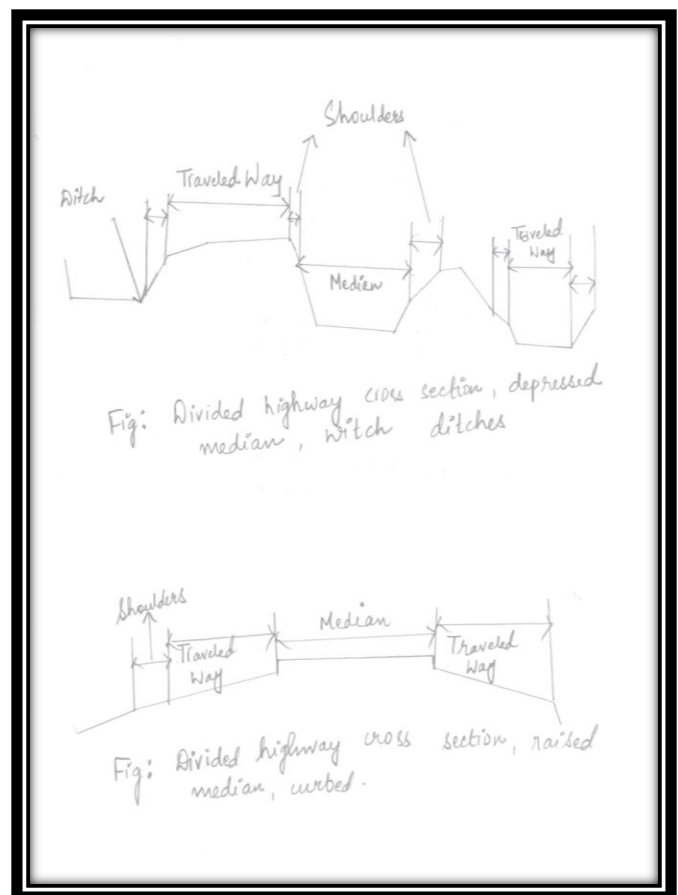
7.3 Vertical Alignment

The vertical alignment of a transportation facility consists of *tangent grades* (straight lines in the vertical plane) and *vertical curves*. Vertical alignment is documented by the **profile.** The **profile** is a graph that has elevation as its vertical axis and distance, measured in stations along the centre-line or other horizontal reference line of the facility, as its horizontal axis.

7.3.1 Tangent Grades

These are designated according to their slopes or grades. Maximum grades vary, depending on the type of facility, and usually do not constitute an absolute standard. The effect of a steep grade is to slow down the heavier vehicles (which typically have the lowest power/weight ratios) and increase operating costs. Furthermore, the extent to which any vehicle (with a given power/weight ratio) is slowed depends on both the steepness and length of the grade. The effect of the slowing of the heavier vehicles depends on the situation, and is often more a matter of traffic analysis than simple geometric design.

As a result, the maximum grade for a given facility is a matter of judgment, with the trade-offs usually being cost of construction versus speed. In the case of railroads, on the other hand, the trade-off is an economic one, involving travel time, construction cost, and minimum power/weight ratios for trains on various grades.
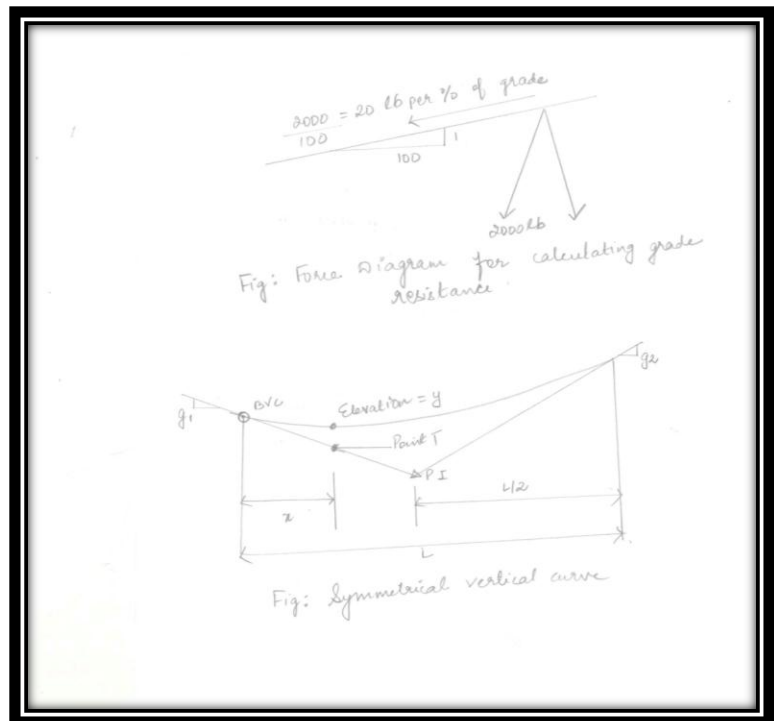
Figure 13 Vertical Tangents

7.3.2 Vertical Curves

Vertical curves are normally parabolas centred about the point of intersection

(P.I.) of the vertical tangents they join. Vertical curves are thus of the form

$$y = y_0 + g_1x + rx^2/2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Equation 5}$$

> y_ elevation of a point on the curve

> y_0 _ elevation of the beginning of the vertical curve (BVC)

> g_1 _ grade just prior to the curve

> x_ horizontal distance from the BVC to the point on the curve

> r_ rate of change of grade

The rate of change of grade, in turn, is given by

$$r = g_2 - g_1/2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Equation 6}$$

Where, g_2 is the grade just beyond the end of the vertical curve (EVC) and L is the length of the curve. Also, vertical curves are sometimes described by K, the reciprocal of r. K is the distance in meters required to achieve a 1 percent change in grade. Vertical curves are classified as *sags* where $g_2 > g_1$ and *crests* otherwise. Note that r (and hence the term$rx^2/2$) will be positive for sags and negative for crests.

7.3.3 Vertical distances in the vertical curve formulas are the product of grade times a horizontal distance

The parabola is selected as the vertical curve so that the rate of change of grade, which is the second derivative of the curve, will be constant with distance. Note that the first derivative is the grade itself, and since the rate of change of grade is constant, the grade of any point in the vertical curve is a linear function of the distance from the BVC to the point. That is,

$$g = dy/dx = g_1 + (Rx) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Equation 7}$$

The quantity $rx^2/2$ is the distance from the tangent to the curve and is known as the offset. If x is always measured from the BVC, the offset given by $rx^2/2$ will be measured from the g_1 tangent.

To determine offsets from the g_2 tangent, x should be measured backward from the EVC. Since the curve is symmetrical about its centre, the offsets from the g_1 and g_2 tangents, respectively, are also symmetrical about the centre of the curve, which occurs at the station of its P.I.

Fig: Stopping sight distance diagram for crest vertical curve

For crest vertical curves, the minimum length depends on the sight distance, the height of the driver's eye, and the height of the object to be seen over the crest of the curve. The minimum length is given by the formula

$$L_{min} = \begin{cases} \dfrac{AS^2}{200\left(\sqrt{h_1} + \sqrt{h_2}\right)^2} & \text{when } S \leq L \\[4mm] 2S - \dfrac{200\left(\sqrt{h_1} + \sqrt{h_2}\right)^2}{A} & \text{when } S > L \end{cases}$$
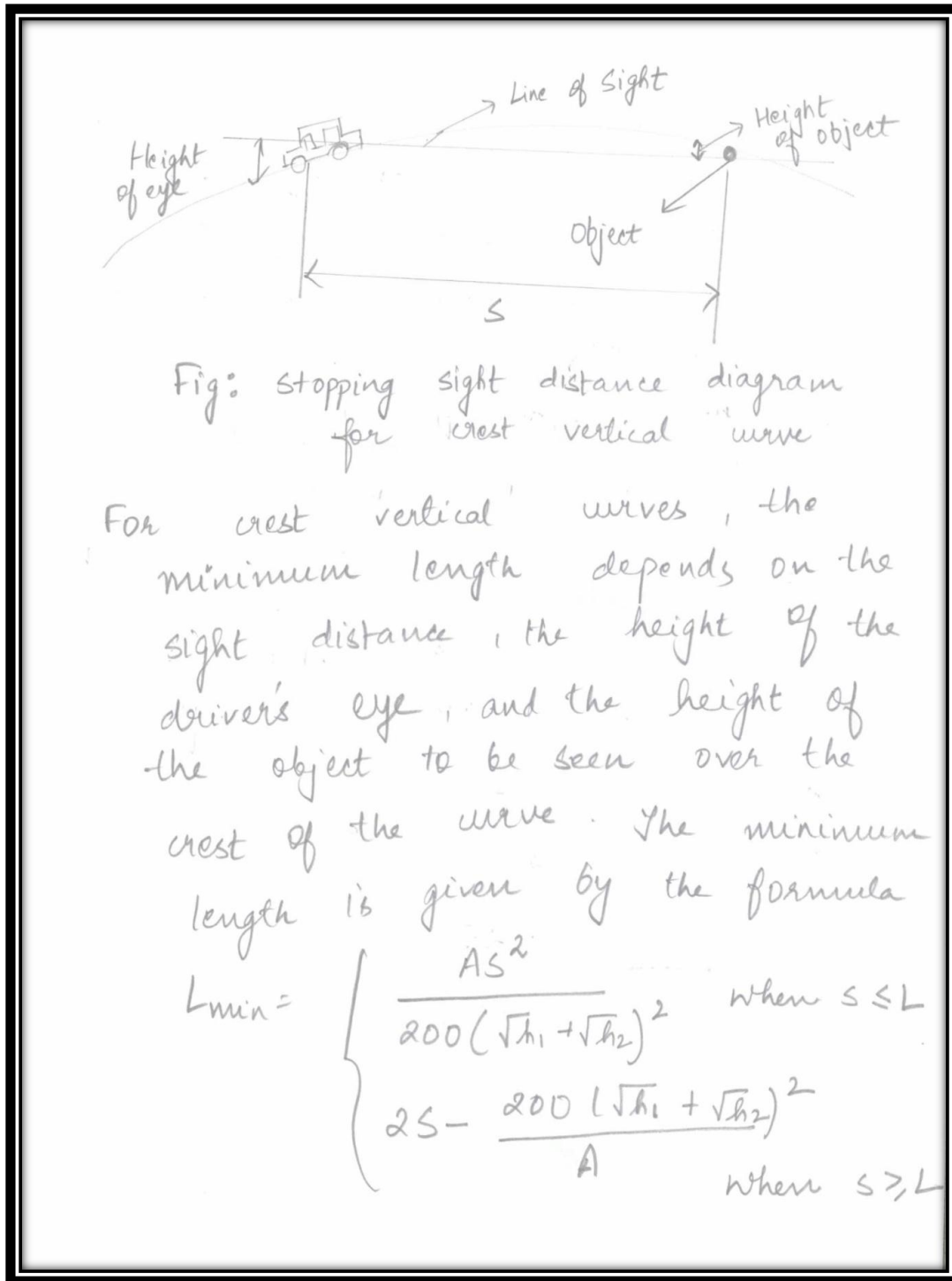
Figure 14 Stopping Sight Distance Diagram and its Derivation

Where $S =$ sight distance

   $L =$ vertical curve length

   $A =$ absolute value of the algebraic difference in grades, in percent $|g_1 - g_2|$

   $h_1 =$ height of eye

   $h_2 =$ height of object

Inserting the standard values for h_1 & h_2 i.e., 0.150 m for h_2 and 1.070 m for h_1. We get

$$L_{min} = \begin{cases} \dfrac{AS^2}{404} & \text{when } S \leq L \\ \\ 2S - \dfrac{404}{A} & \text{when } S > L \end{cases}$$

for stopping sight distance

$$L_{min} = \begin{cases} \dfrac{AS^2}{946} & \text{when } S \leq L \\ \\ 2S - \dfrac{946}{A} & \text{when } S > L \end{cases}$$

for passing sight distance.

Figure 15 Sight Distance Derivation
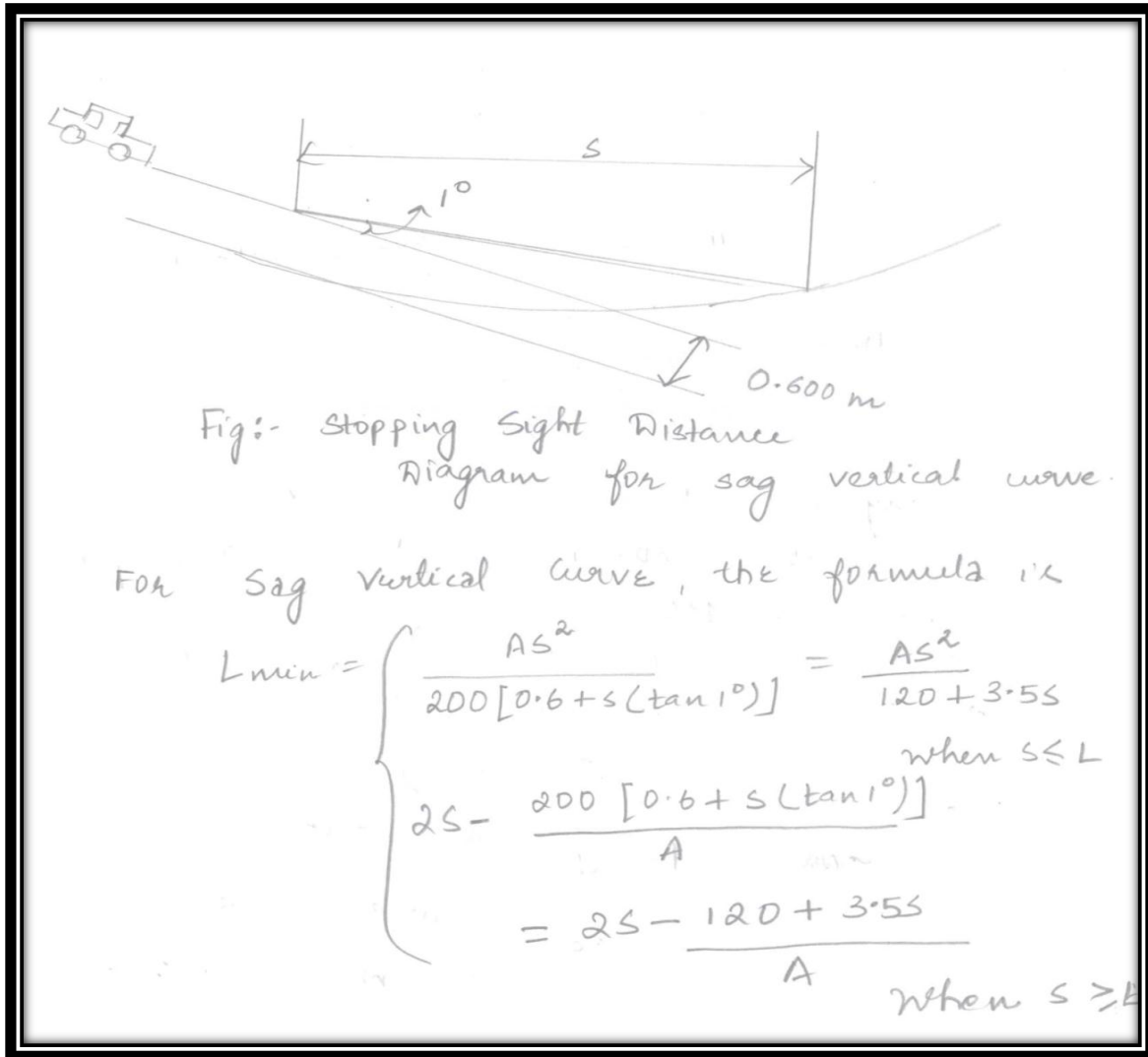
Figure 16 Stopping Sight Distance Diagram for Sag Vertical Curve

In some cases, sag vertical curves with a small total grade change can be sharp enough to cause discomfort without violating sight distance standards. In this case, it is necessary to establish a comfort criterion of the form

$$r \leq \frac{a}{v^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Equation 8}$$

Where, r is the rate of change of grade, 'a' is the maximum radial acceleration permitted, and y is speed. There is no general agreement as to the maximum value of radial acceleration that can be tolerated without producing discomfort.

Finally, vertical curve lengths may be limited by the need to provide clearances over or under objects such as overpasses or drainage structures. In the case of sag vertical curves passing over objects or crest vertical curves passing under them, the required clearances establish minimum lengths; in the case of crest vertical curves passing over objects or sags passing under them, the clearances establish maximum lengths. Where clearances limit vertical curve lengths, adequate sight distance should still be provided.

In either case, the maximum or minimum length of the vertical curve may be determined by assuming that the clearance is barely met and calculating the length of the vertical curve passing through the critical point thus established.

The equation for the offset

$$y' = \frac{rx^2}{2}$$

where r is

$$r = \frac{g_2 - g_1}{L} = \frac{A}{L}$$

$$x = \frac{L}{2} + z$$

Substituting these values in

$$r \leq \frac{a}{v^2} \quad, \text{ we get}$$

$$y' = \frac{A\left(L/2 + z\right)^2}{2L}$$

and on rearranging eqn

$$r = \frac{g_2 - g_1}{L} = \frac{A}{L}$$

we get

$$AL^2 + \left(4Az - 8y'\right)L + 4Az^2$$

Equation 9

On solving which we get two roots.

The smaller of these represents a vertical curve i.e., tangent between the P.I & the critical point.

Just letting $W = y'/A$ to simplify the notation, the solution in the larger root leads to

$$L = 4w - 2z + 4\sqrt{w^2 - wz}$$

as an expression for the maximum or minimum vertical curve length.
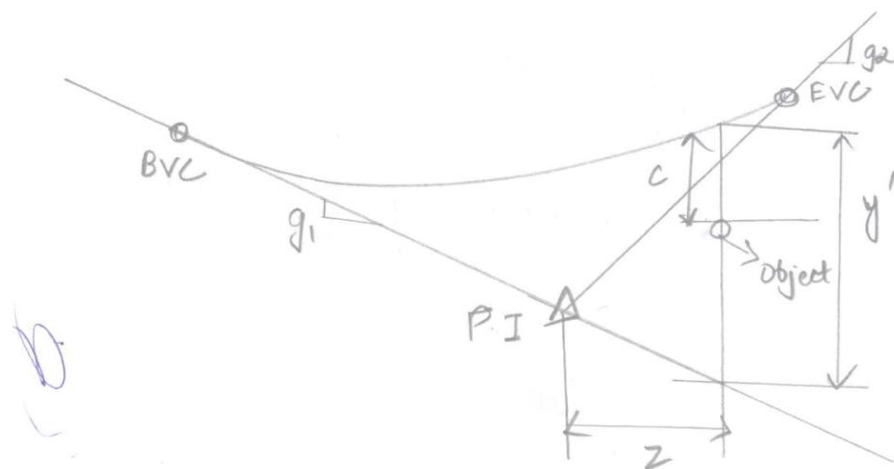
Fig: Calculation of lengths of vertical curves

Figure 17 Calculation of lengths of vertical curves

7.4 Horizontal Alignment

Horizontal alignment for linear transportation facilities such as highways and railways consists of horizontal tangents, circular curves, and possibly transition curves.

7.4.1 Horizontal tangents

Horizontal tangents are described in terms of their lengths (as expressed in the stationing of the job) and their directions. Directions may be either expressed as bearings or as azimuths and are always defined in the direction of increasing station. Azimuths are expressed as angles turned clockwise from due north; bearings are expressed as angles turned either clockwise or counterclockwise from either north or south. For instance, the azimuth 280 is equivalent to the bearing north 80west (or N80W).
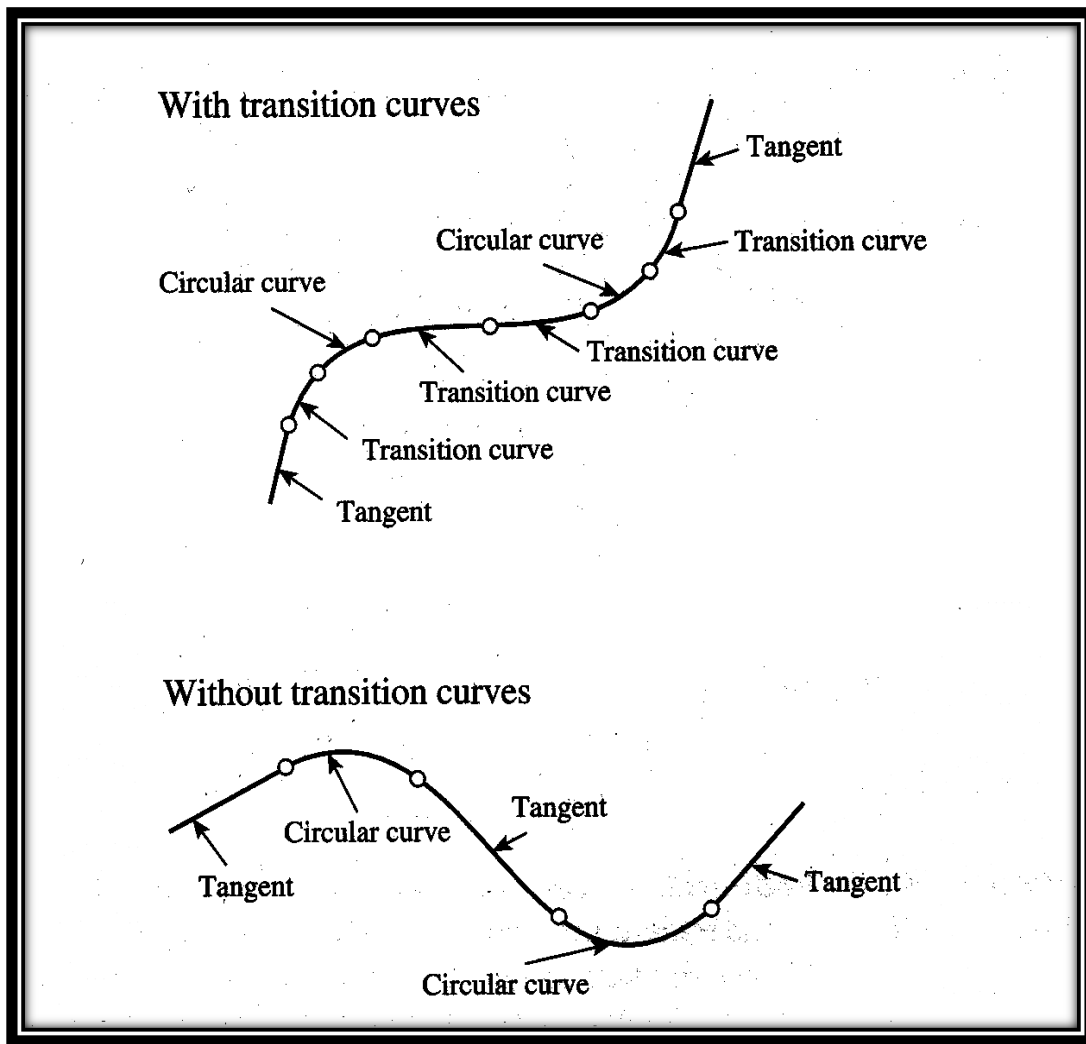
Figure 18 Representation of Transition Curve for Highways

• • •

7.4.2 Circular Curves

Horizontal curves are normally circular. Horizontal curves are described by radius (R), central angle (Δ) (which is equal to the deflection angle between the tangents), length (L), semi-tangent distance (T), middle ordinate (M), external distance (E), and chord (C). The curve begins at the tangent-to-curve point (TC) and ends at the curve-to-tangent point (CT). In the past, severity of curvature was sometimes expressed in degree of curvature. Although obsolete in the metric system, degree of curvature may still be encountered in some situations. Degree of curvature may be defined in two ways. The arc definition is the angle subtended by a 100 ft. arc. The chord definition is the angle subtended by a 100 ft. chord. The relationship between radius (in feet) and degree of curvature (arc definition) is

$$\textbf{D= 36,000/ 2πR = 5729.58/R}\ \dotfill \text{Equation 10}$$

Where D = degree of curvature and R = radius of curvature, in feet.

The length of a circular curve is given by

$$\textbf{L = 2πR/360}^{0} \textbf{= R } \Delta_{\textbf{rad}}\ \dotfill \text{Equation 11}$$

Where Δis the central angle of the curve; Δrad refers to Δmeasured in radians.

Figure 3.13 Elements of a simple circular horizontal curve.

In this figure,

R = radius, usually measured to the centerline of the road, in ft (m),

Δ = central angle of the curve in degrees,

PC= point of curve (the beginning point of the horizontal curve),

PI= point of tangent intersection,

PT= point of tangent (the ending point of the horizontal curve),

T = tangent length in ft (m),

M = middle ordinate in ft (m),

E = external distance in ft (m), and

L = length of curve in ft (m).

Figure 19 Elements of Simple Circular Horizontal Curve

The semi-tangent *T* of a circular curve is given by:

$$\mathbf{T = R * \tan\frac{\Delta}{2}}\text{................................} \text{Equation 12}$$

The external distance *E* is given by:

$$\mathbf{E = R\left[\frac{1}{\cos\left(\frac{\Delta}{2}\right)} - 1\right]}\text{................................} \text{Equation 13}$$

The middle ordinate *M* is given by:

$$\mathbf{M = R\left[1 - \cos\left(\frac{\Delta}{2}\right)\right]}\text{................................} \text{Equation 14}$$

The chord *C* is given by:

$$\mathbf{C = 2R\, Sin\left(\frac{\Delta}{2}\right), L = \left(\frac{\pi}{180}\right)R\Delta}\text{................................} \text{Equation 15}$$

Circular curves are usually laid out in the field by occupying the tangent-to curve point TC with a transit and then establishing successive points by turning deflection angles and measuring chords. The deflection angle in radians d_x to a point on the curve at a distance x from the TC is given by: $\boldsymbol{dx = \left(\frac{x}{2R}\right)rad}$ .... Equation 16

The chord cx to this point is given by:

$$\boldsymbol{Cx = 2R\sin(dx)}\text{................................} \text{Equation 17}$$

Minimum radius of horizontal curve is most commonly established by the relationship between design speed, maximum rate of super-elevation, and curve radius. The relationship between the radius of curvature R, the setback distance m, and the sight distance s is given by

$$\boldsymbol{m = R\left[1 - \cos\frac{28.65s}{R}\right]}\text{................................} \text{Equation 18}$$

$$\boldsymbol{s = \frac{R}{28.65}\left[\cos^{-1}\frac{(R-m)}{R}\right]}\text{................................} \text{Equation 19}$$

Where, the angles are measured in degrees

Since these formulas are hard to solve for R, design charts or tables are normally used to find the minimum radius of curvature that will provide stopping sight distance. Minimum lengths or radii of horizontal curves may also be based on appearance criteria. Where deflection angles are small, a short horizontal curve may give the appearance of a kink. To prevent this, minimum horizontal curve lengths may be prescribed for curves with small deflection angles.

Fig:- Horizontal curve features

$$D = \frac{36,000}{2\pi R} = \frac{5729.58}{R}$$

D = Degree of curvature
R = radius of curvature in feet
L = length of circular curve is given by

$$L = \frac{2\pi R\Delta}{360^\circ} = R\Delta_{rad}$$

Δ = central angle of the curve
Δ_{rad} = Δ measured in radians

$$T = R\tan\frac{\Delta}{2}$$

The middle ordinate,

$$M = R - R\cos\left(\frac{\Delta}{2}\right)$$
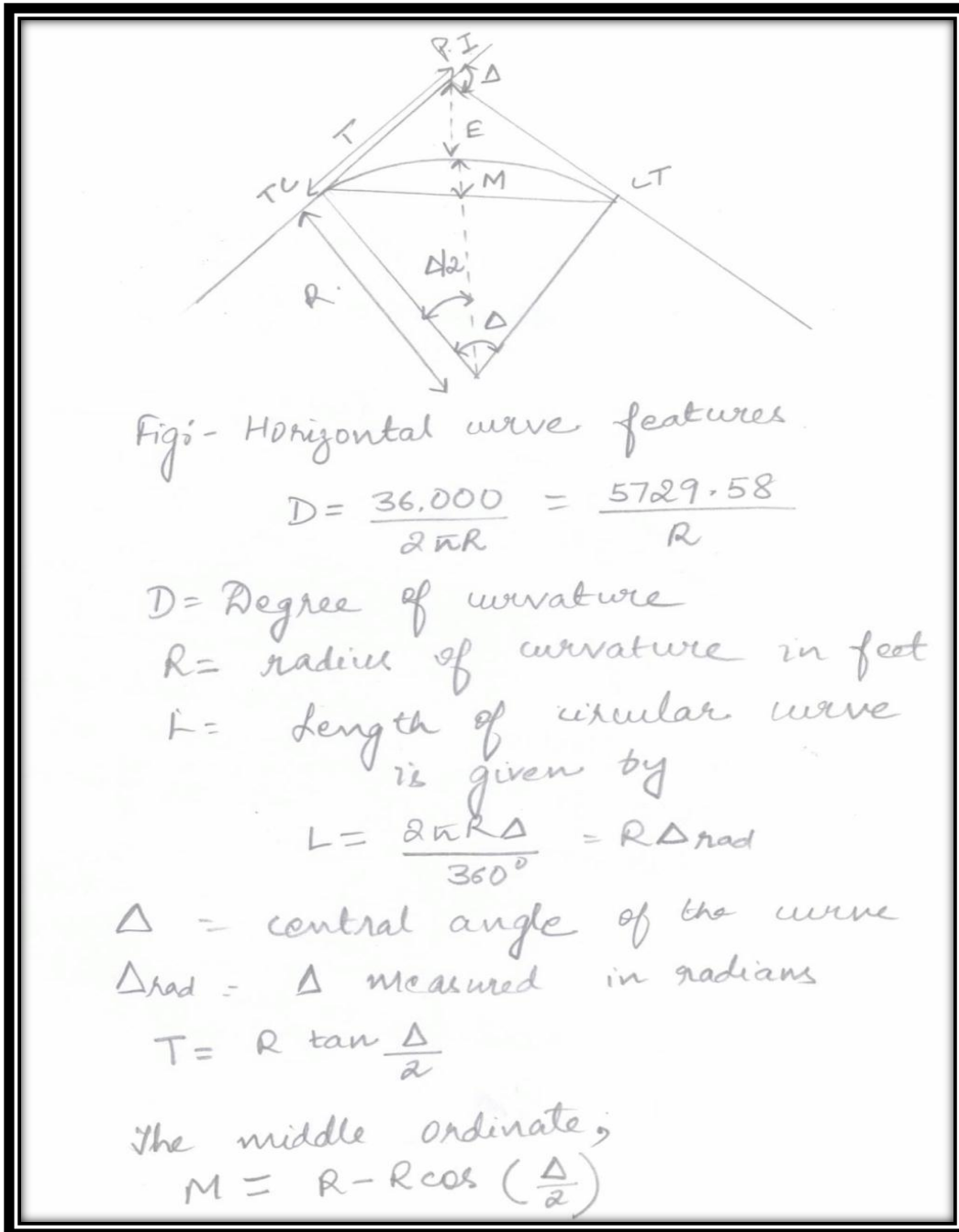
Figure 20 Horizontal Curve Features

• • •

The external distance E is:

$$E = \frac{R}{\cos(\Delta/2)} - R$$

The chord C is given by

$$C = 2R \sin\left(\frac{\Delta}{2}\right)$$
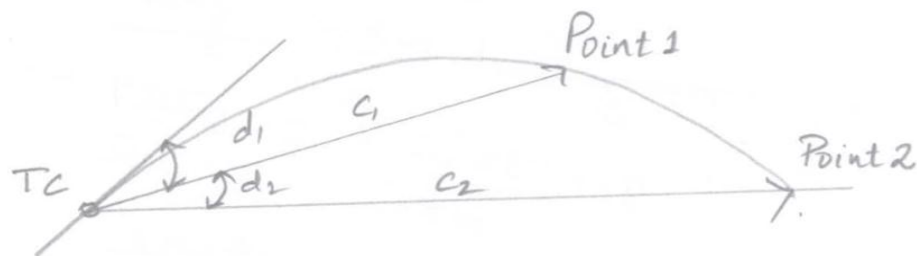
Fig: Layout of Circular curve

The deflection angle in radiance d to a point on the curve at a distance x from TC is given by

$$d_x = \left(\frac{x}{2R}\right) rad$$

The chord c_x to this point is given by

$$c_x = 2R \sin d_x$$

Figure 21 Circular Curve Layout

7.4.3 Transition Curves

Transition curves are used to connect tangents to circular curves. Several forms of curve have been used for this purpose. The most logical choice from a theoretical standpoint, and the only one discussed here, is the clothoid spiral, for which the radius of curvature varies as the inverse of the distance along the curve from its beginning. Spirals are used both for esthetic reasons and because they provide a "rational" super-elevation transition. In the case of railways, such rational super-elevation transitions are virtually necessary for reasons of vehicle dynamics. In the case of highways, spirals are used primarily for esthetic purposes.

They are most appropriate for roadways with relatively high design standards, where larger radius curves are used. Under these circumstances, drivers can often see a considerable way ahead on the roadway, and can detect the difference between the smoother, flowing lines provided by the transition curves and the more abrupt alignment that results in their absence. For roadways with lower design standards, some recent research reports indicate that use of spirals may increase accident rates. It is surmised that the reason is that drivers have a harder time judging the severity of curves where spirals are used. Since the spiral is defined as the curve such that the reciprocal of the radius varies linearly from zero at the TS to 1_Rc at the SC.

$$\frac{1}{R} = \left(\frac{1}{Rc}\right)\left(\frac{L}{Ls}\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \text{Equation 20}$$

OR

$$RL = (R_c) * (L_s) = A^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots \text{Equation 21}$$

The spiral angle θ is given by

$$\theta = \frac{L}{2R} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \text{Equation 22}$$

Spirals are laid out in the field in a manner similar to that for circular curves. In this case the TS are occupied by the transit, and successive points along the spiral are established by turning deflection angles and measuring chords. For a point on a spiral whose coordinates have previously been calculated, the deflection angle d is

$$d = \tan^{-1}\frac{Y}{X} \dots\dots\dots\dots\dots\dots\dots\dots\dots \text{Equation 23}$$

• • •

And the chord c is

$$c = \sqrt{X^2 + Y^2}$$ ....…..….. **Equation 24**

In order to determine the stations of the critical points on the curves, it is necessary to remember that stationing runs along the curves and not the tangents.

Fig: Spiral transition curve nomenclature

The spiral is defined as the curve such that the reciprocal of the radius varies linearly from zero at the TS to $1/R_c$ at the SC,

$$\frac{1}{R} = \left(\frac{1}{R_c}\right)\left(\frac{L}{L_s}\right)$$

$$\Rightarrow RL = R_c L_s = A^2$$
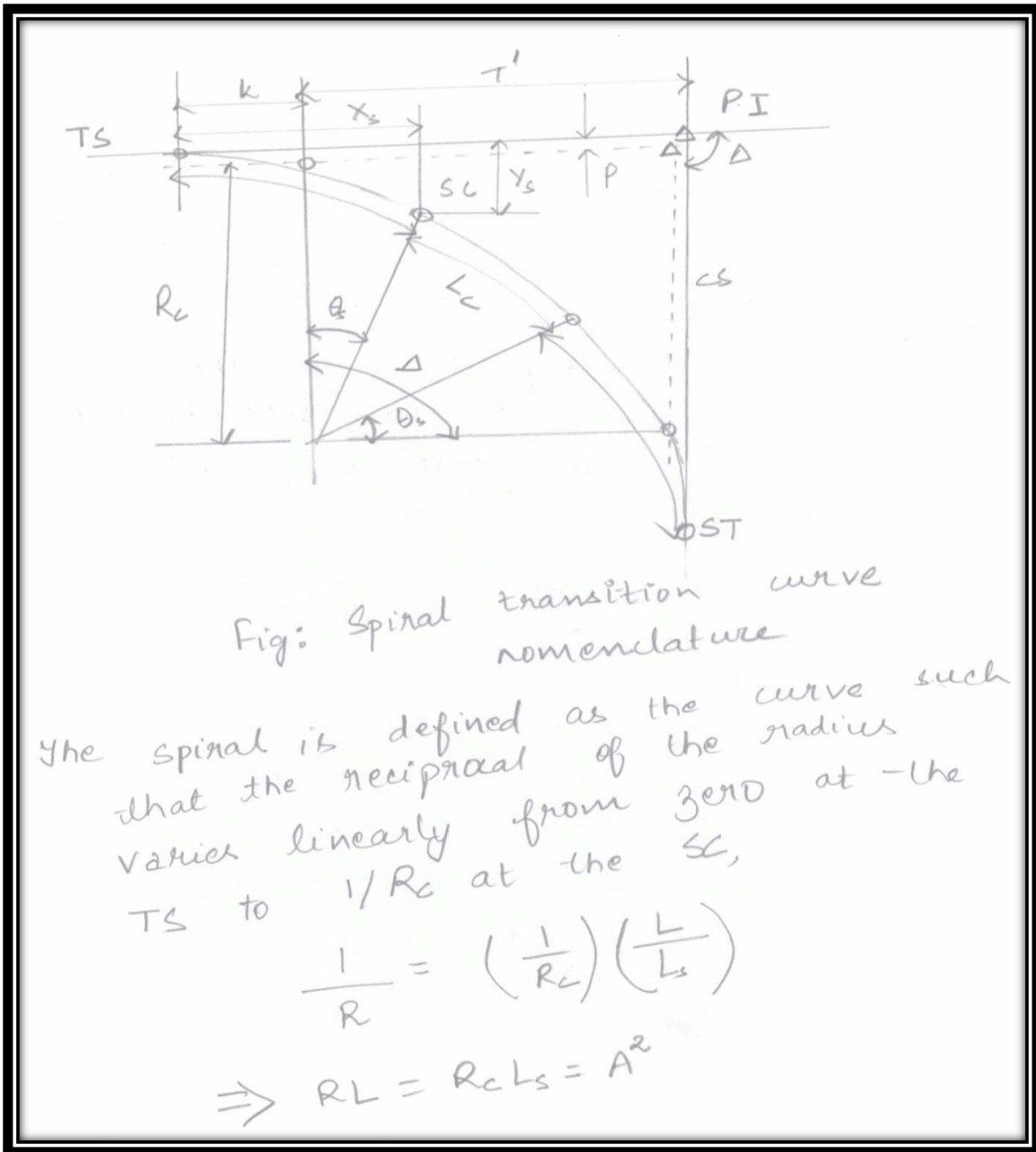
Figure 22 Nomenclature for Spiral Transition Curve

The spiral angle θ is given by

$$\theta = \frac{L}{2R}$$

$$\Rightarrow \quad \theta_s = \frac{L_s}{2R_c}$$

The coordinates x & y of any point on the spiral may be expressed as functions of L:

$$X = L - \frac{L^5}{40 A^4} + \frac{L^9}{3,456 A^8} + \cdots$$

$$Y = \frac{L^3}{6 A^2} - \frac{L^7}{336 A^6} + \frac{L^{11}}{42,240 A^{10}} + \cdots$$

other measurements of interest are

$$p = Y_s - R_c (1 - \cos\theta_c)$$

$$k = X_s - R_c \sin\theta_s$$

$$T' = (R_c + p) \tan\left(\frac{\Delta}{2}\right)$$

and

$$L_c = R_c (\Delta_{rad} - 2\theta_c) = R_c \Delta_{rad} - L_s$$

For a point on a spiral curve, whose coordinates have previously been calculated, the deflection angle d is

$$d = \tan^{-1}\left(\frac{Y}{X}\right)$$

and the chord c is

$$c = \sqrt{X^2 + Y^2}$$

Equation 25

7.4.4 Superelevation

The purpose of superelevation or banking of curves is to counteract the centripetal acceleration produced as a vehicle rounds a curve. The term itself comes from railroad practice, where the top of the rail is the profile grade. In curves, the profile grade line follows the lower rail, and the upper rail is said to be "superelevated." Since most railways are built to a standard gage, superelevations are given as the difference in elevation between the upper and lower rail. In the case of highways, somewhat more complicated modifications of the cross section are required, and, because widths vary, *superelevation is expressed as a slope.*

In this figure,

R_v = radius defined to the vehicle's traveled path in ft (m),

α = angle of incline in degrees,

e = number of vertical ft (m) of rise per 100 ft (m) of horizontal distance,

W = weight of the vehicle in lb (N),

W_n = vehicle weight normal to the roadway surface in lb (N),

W_p = vehicle weight parallel to the roadway surface in lb (N),

F_f = side frictional force [centripetal, in lb (N)],

F_c = centripetal force [lateral acceleration × mass, in lb (N)],

F_{cp} = centripetal force acting parallel to the roadway surface in lb (N), and,

F_{cn} = centripetal force acting normal to the roadway surface in lb (N).

Figure 23 Superelevation

• • •

65
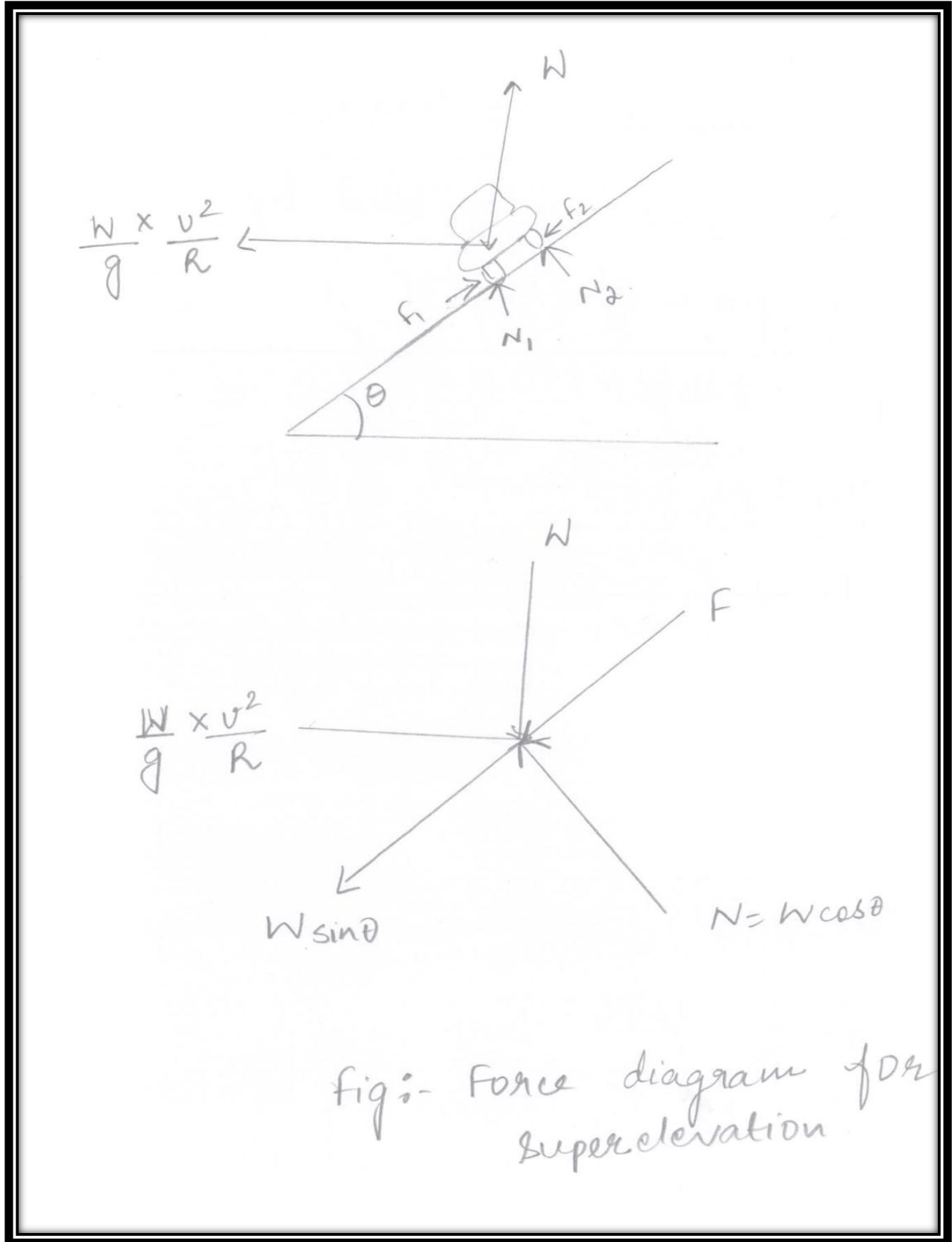
Figure 24 Force Diagram for Superelevation

R = radius of the curve

v = speed of the vehicle.

$\frac{v^2}{R}$ = radial acceleration.

which will be opposed by

a force of $\left(\frac{W}{g}\right)\left(\frac{v^2}{R}\right)$

W = Weight

N_1 & N_2 = Normal forces

F_1 & F_2 = lateral forces
also called
friction forces
for highway
vehicles.

$F_1 \leq \mu N_1$

$F_2 \leq \mu N_2$

where μ is coefficient of friction
between the tires and the
roadway.

Figure 25 Superelevation Diagram for Force

Summing forces parallel to the roadway gives

$$\cos\theta\left(\frac{W}{g}\right)\left(\frac{v^2}{R}\right) = F + W\sin\theta$$

Defining a factor "f" so that

$$f = \frac{F}{N}$$

and $N = W\cos\theta + \sin\theta\left(\frac{W}{g}\right)\left(\frac{v^2}{R}\right)$

So from

$$c = \sqrt{x^2 + y^2} \quad \text{we get}$$

$$\cos\theta\left(\frac{W}{g}\right)\left(\frac{v^2}{R}\right) = W\sin\theta + fW\cos\theta + f\sin\theta\left(\frac{W}{g}\right)\left(\frac{v^2}{R}\right)$$

Dividing by $W\cos\theta$

$$\Rightarrow \frac{v^2}{gR} = \tan\theta + f + f\tan\theta\left(\frac{v^2}{gR}\right)$$

here $\tan\theta$ is the cross slope of the roadway, which is the same as the superelevation rate e. Consequently,

$$\frac{v^2}{gR} = e + f + ef\left(\frac{v^2}{gR}\right)$$

or $\frac{v^2}{gR}(1 + ef) = e + f$

where $ef \approx 1$,

$$\Rightarrow \frac{v^2}{gR} = f + e$$

$$R = \frac{v^2}{g(f+e)}$$

Figure 26 Forces Parallel to Highway - Superelevation

• • •

Maximum rates of superelevation are limited by the need to prevent slow-moving vehicles from sliding to the inside of the curve and, in urban areas, by the need to keep parking lanes relatively level and to keep the difference in slope between the roadway and any streets or driveways that intersect it within reasonable bounds. In addition, there is a trade-off between the maximum rate of superelevation and the minimum curve radius permitted at any design speed.

Superelevation transitions involve modification of the roadway cross section from normal crown to full superelevation, at which point the entire roadway width has a cross-slope. The manner in which this transition is accomplished is expressed by a superelevation diagram, which is a graph of superelevation (cross-slope) versus distance measured in stations. As an alternative, the diagram may show the difference in elevation between the profile grade and the edge versus distance. Superelevation transitions involve modification of the roadway cross section from normal crown to full superelevation, at which point the entire roadway width has a cross-slope.

For highways, superelevation is usually applied by rotating the cross section about the profile grade line. This will be the centre-line in the case of two-lane highways, undivided multilane highways, and multilane divided highways with paved medians.

For multilane divided highways with wide medians, the inside edge of travelled way is often used as the axis of rotation. For freeway ramps, the axis of rotation will normally be the edge of travelled way closer to the freeway; however, it may be shifted to avoid sags in the edge grade of the ramp.
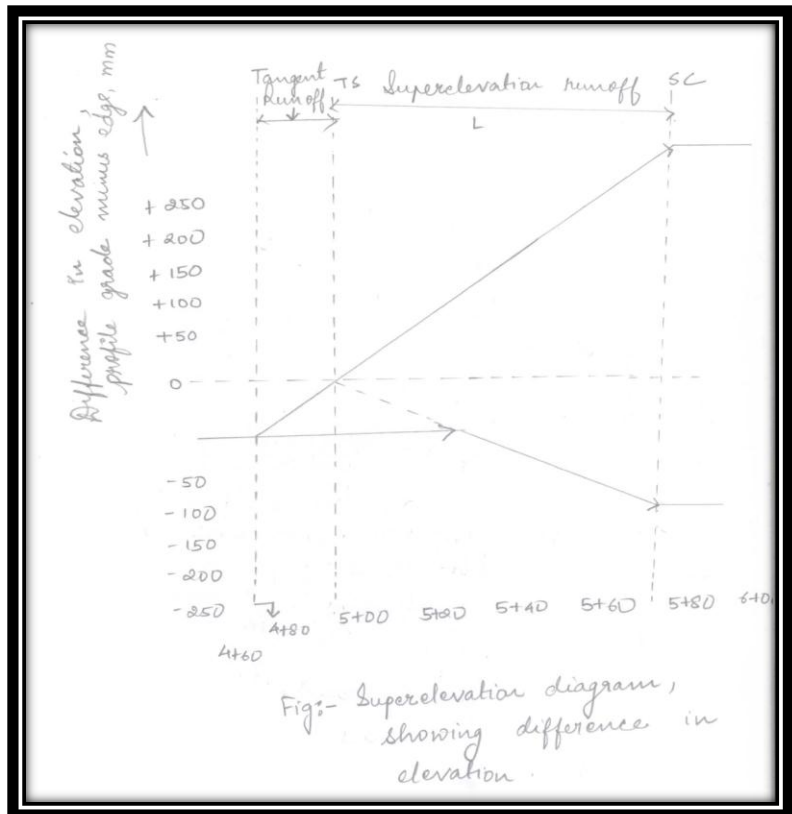
Figure 27 Differences in elevation shown by Superelevation
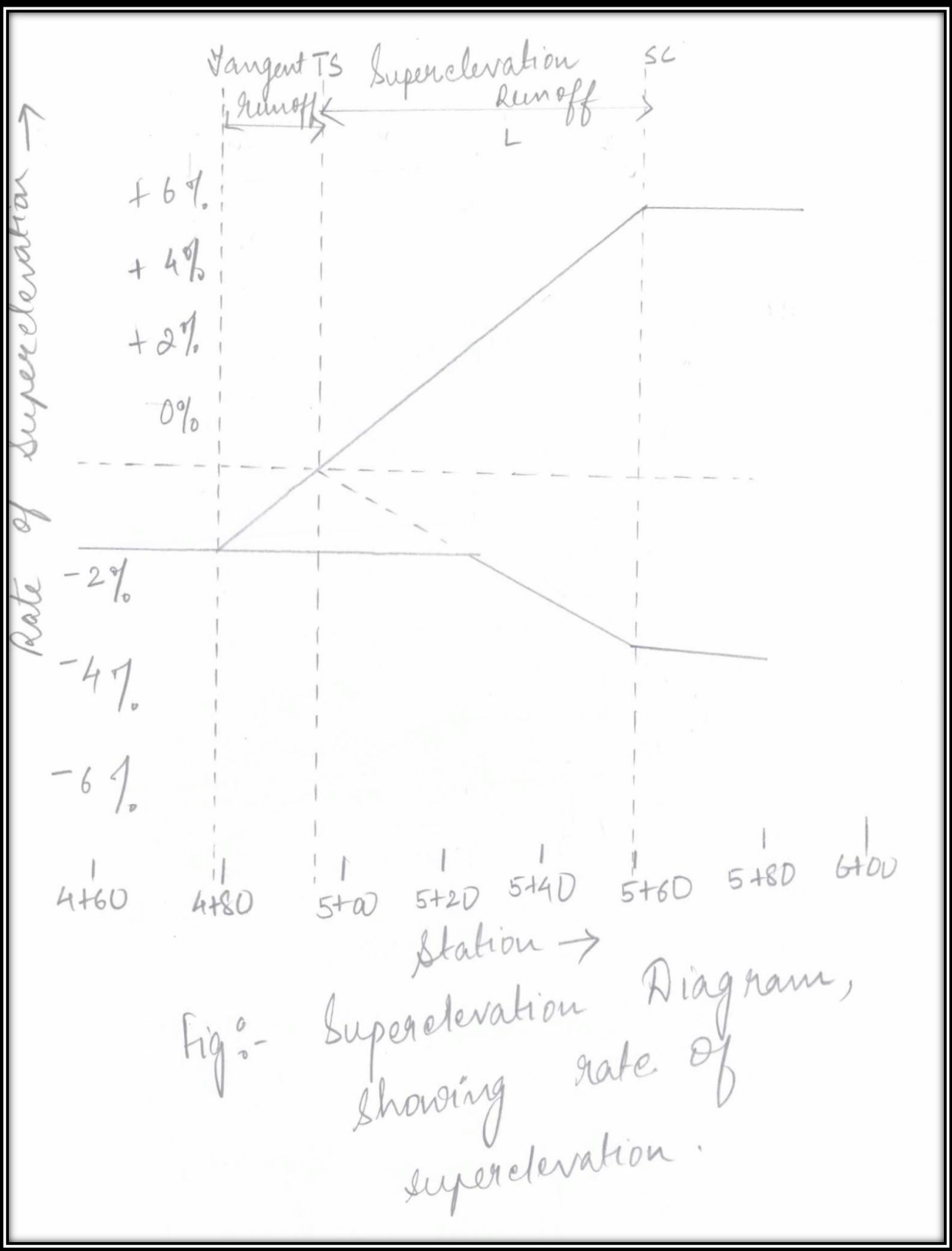
Figure 28 Rate of Superelevation
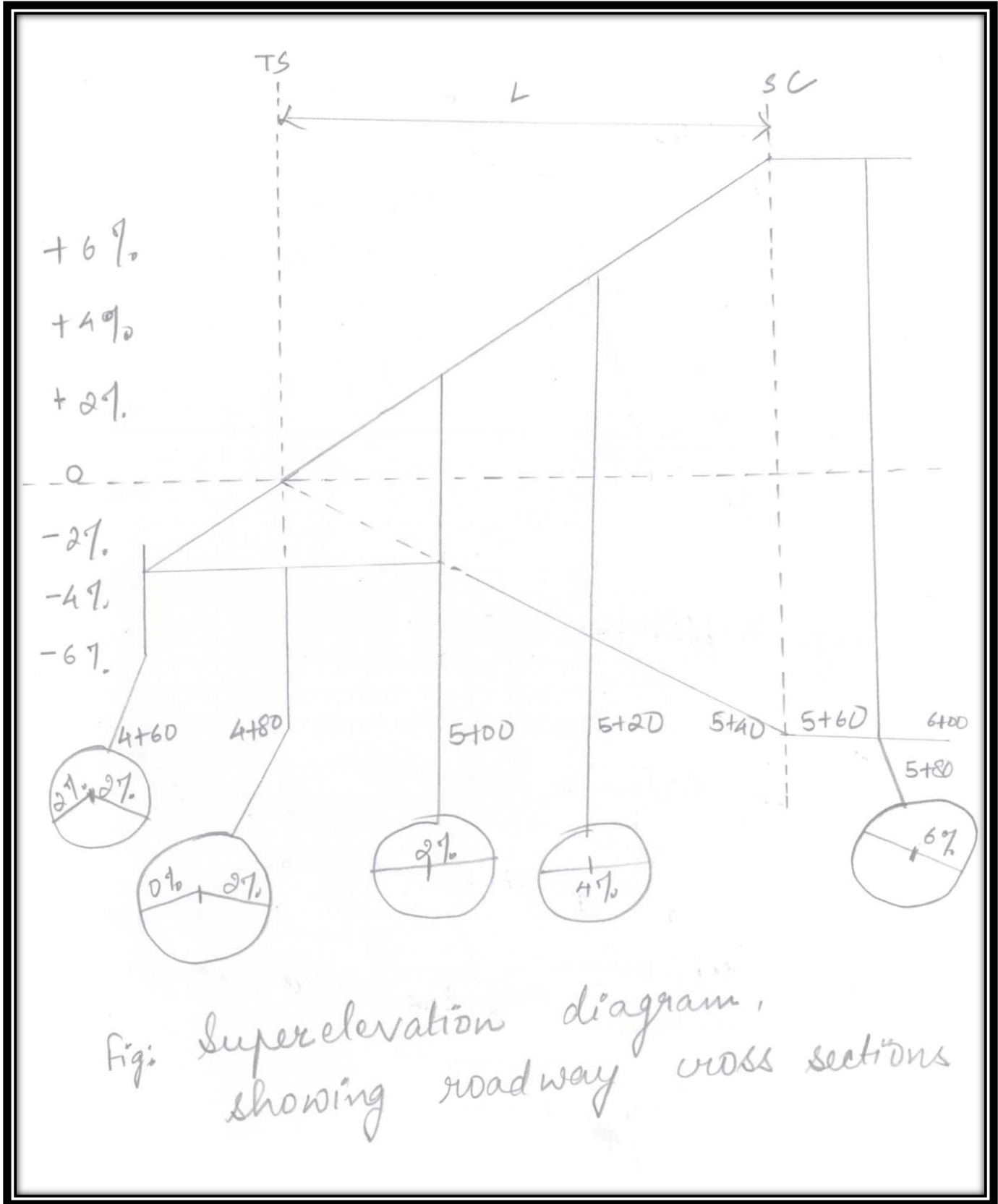
Figure 29 Roadway Cross Section in Superelevation
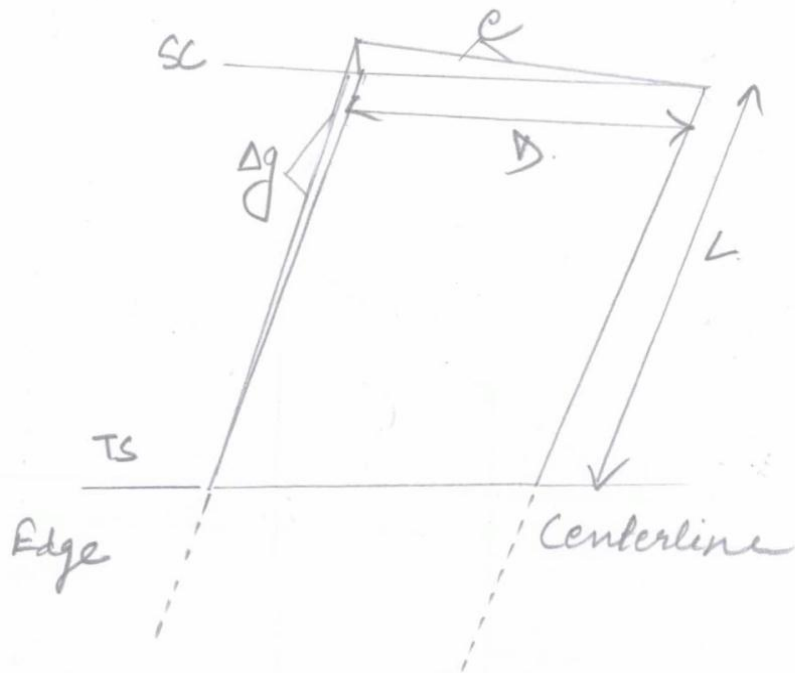
Fig: Difference in grade resulting from Superelevation transition.

L: The distance from the TS to the SL in the superelevation diagram

The difference in grade between the centerline & the edge is

$$\Delta g = \frac{De}{L}$$

Figure 30 Grade differences from Superelevation transition

7.5 Coordinates of Horizontal and Vertical alignment

Transportation facilities such as highways and railways are three-dimensional objects. Although many aspects of their design can be determined by considering horizontal and vertical alignment separately from one another, it is important to understand the relationship between them. Proper coordination of horizontal and vertical alignment is important for reasons related to the aesthetics, economics, and safety of the facility.

As a general rule, horizontal curvature and grades should be kept in balance. That is, the designer should avoid both the provision of minimal curvature at the expense of long, steep grades and the provision of level vertical alignment at the expense of excessive horizontal curvature. Where there is both horizontal and vertical curvature, it is normally best from an aesthetic standpoint to provide the impression of a single three-dimensional curve in both the horizontal and vertical planes.

This means that horizontal and vertical curves should normally coincide. In some cases, however, safety considerations may suggest that horizontal curves be extended beyond vertical curves in order to avoid hiding the beginning of the horizontal curve from drivers approaching it. This is especially important where horizontal curves coincide with rather sharp crest vertical curves. In addition, use of long, relatively gentle curves and short tangents will normally produce a more flowing line than will use of long tangents and short, sharp curves. This is especially true when the deflection angle of horizontal tangents or the difference between grades is small. At the same time, too much curvature may also pose problems.

On two-lane highways, for instance, it is also important to provide an adequate number of sections with passing sight distance, and these sections need to be of adequate length to prevent drivers from becoming impatient when following slow-moving vehicles. Finally, on intersection approaches, highway alignments should provide adequate sight distance and should be as flat and straight as possible. In addition to these rules, there are a number of guidelines related to the positioning of curves (either horizontal or vertical) relative to one another. These are based on aesthetics and in the case of horizontal curves, on the necessity of providing adequate superelevation transitions.

Reversing horizontal curves must be separated by a tangent or by transition curves to allow for the development of superelevation. The superelevated cross-slopes of the roadway will be in opposite directions in the two curves, so that some distance must be provided for rotation of the cross section. Where transition curves are used, however, the ST of the first curve may coincide with the TS of the second. Reversing vertical curves pose no problem.

Compound curves result when two curves of differing radius (or for vertical curves, different rates of change of grade) join one another. Such curves are normally avoided for centreline alignment, although use of three-centred compound curves for pavement edges in intersections is common. Problems with compound curves include aesthetics and, in the case of horizontal curves, difficulty in developing the necessary superelevation transition, as well as possible deception of the driver as to the severity of the curve. Compound curves may be acceptable if the difference in radius is small or if they occur on a one-way roadway and the radius of curvature increases in the direction of travel. Broken-back curves consist of two curves in the same direction separated by a short tangent. Such curves (whether horizontal or vertical) are objectionable on aesthetic grounds and should be replaced by a single, larger- radius curve.
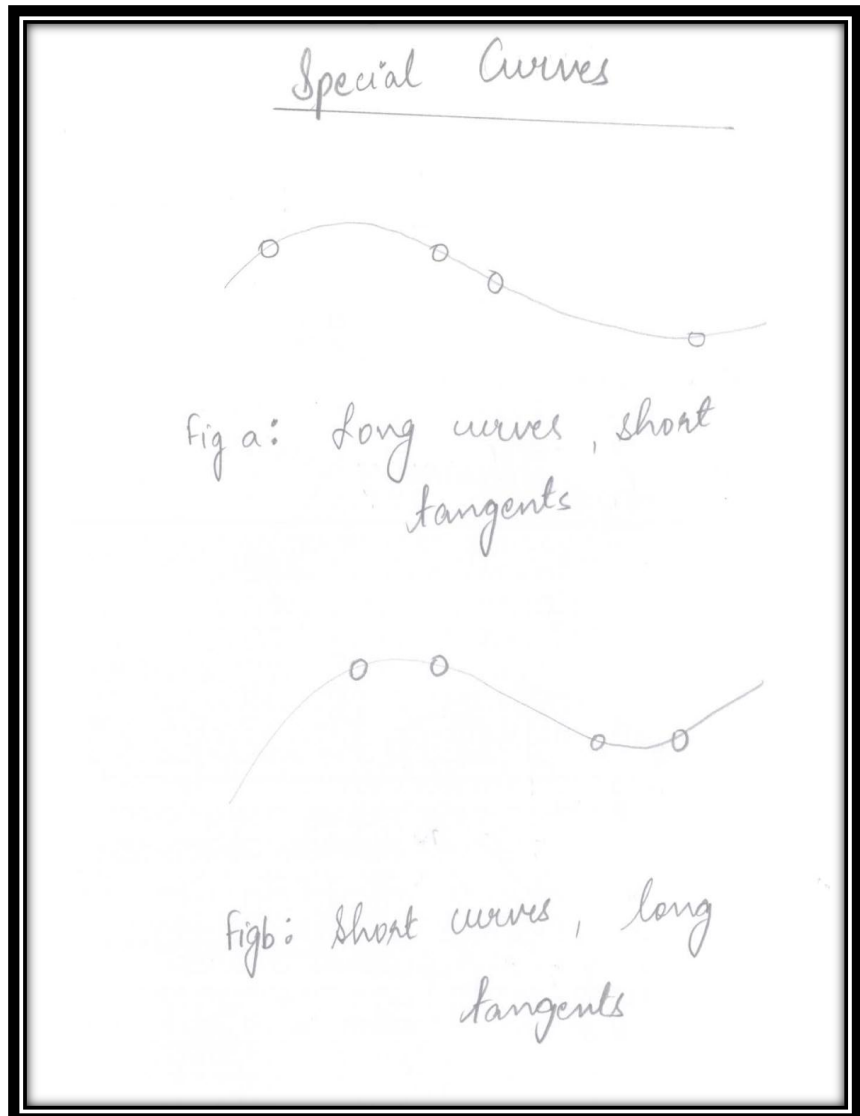
Figure 31 Special Curves (I)

Figure 32 Special Curves (II)
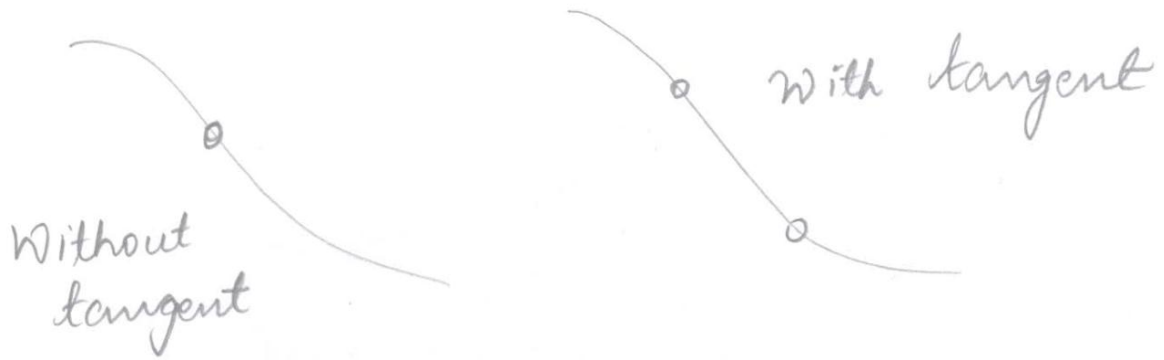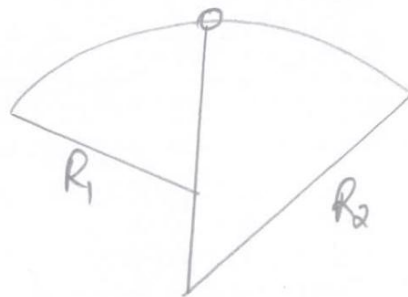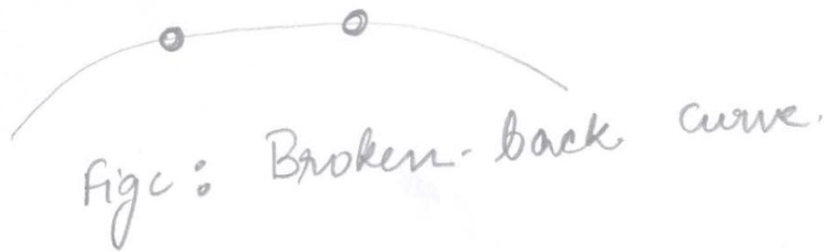
7.6 Intersections & Interchanges

Geometric design of transportation facilities must provide for the resolution of traffic conflicts. In general, these may be classified as merging, diverging, weaving, and crossing conflicts. Merging conflicts occur when vehicles enter a traffic stream; diverging conflicts occur when vehicles leave the traffic stream; weaving conflicts occur when vehicles cross paths by first merging and then diverging; and crossing conflicts occur when they cross paths directly. There are three basic ways of resolving crossing conflicts. Time-sharing solutions involve assignment of the right-of-way to particular movements for particular times. An example of this type of solution is the signalized intersection. Space-sharing solutions convert crossing conflicts into weaving conflicts.

An example of this is the traffic circle or rotary. Grade separation solutions eliminate the crossing conflict by placing the conflicting traffic streams at different elevations at their point of intersection. Examples of this solution are freeway interchanges and highway–railway grade separations. Freeway interchanges are classified primarily according to the way in which they handle left-turning traffic.

Diamond interchanges employ diamond ramps, which connect to the cross road by means of an at-grade intersection. Left turns are accomplished by having vehicles turn left across traffic on the cross road. Cloverleaf intersections employ loop ramps, in which vehicles turn left by turning 270° to the right. Partial cloverleaf (or parclo) interchanges involve various combinations of diamond and loop ramps. Direct interchanges employ direct ramps, in which vehicles turn left by means of a left-turning ramp that is grade-separated as it crosses both traffic streams. Frontage roads are used to provide access to adjacent property, and collector–distributor roads are used to intercept traffic from local streets that do not cross the freeway. Interchange configurations are selected on the basis of structural costs, right-of way costs, and ability to serve traffic.

In general, diamond ramps have the lowest structural and land use costs, but also have the lowest capacities. Loop ramps provide higher capacities at moderate structural costs, but have high right-of-way costs because they take up more space. Direct ramps have the highest capacities, and often have moderate right-of-way costs, but often have very high structural costs due to the extensive structures required.

Direct ramps are often used in freeway-to-freeway interchanges, where they are referred to as freeway-to-freeway connectors (or branch connectors) rather than ramps. Freeway ramp junctions are normally treated as standard details, but may differ from agency to agency. In general, there are two types: tapered and parallel. Highway intersections at grade also pose a number of special geometric problems. Pavement edges in intersections must be rounded to accommodate the wheel tracking paths of large vehicles. This is normally done by use of templates for particular design vehicles, In addition, care must be taken to ensure that building setbacks and landscaping provide adequate sight distance for vehicles approaching the intersection, and pavement crowns must be flattened and warped so as to allow for drainage while at the same time providing as smooth a crossing as possible. There are no set rules for design of vertical alignment and cross section through intersections; rather each case must be analysed individually in order to provide the best possible combination of smooth ride and drainage within whatever constraints may be present.
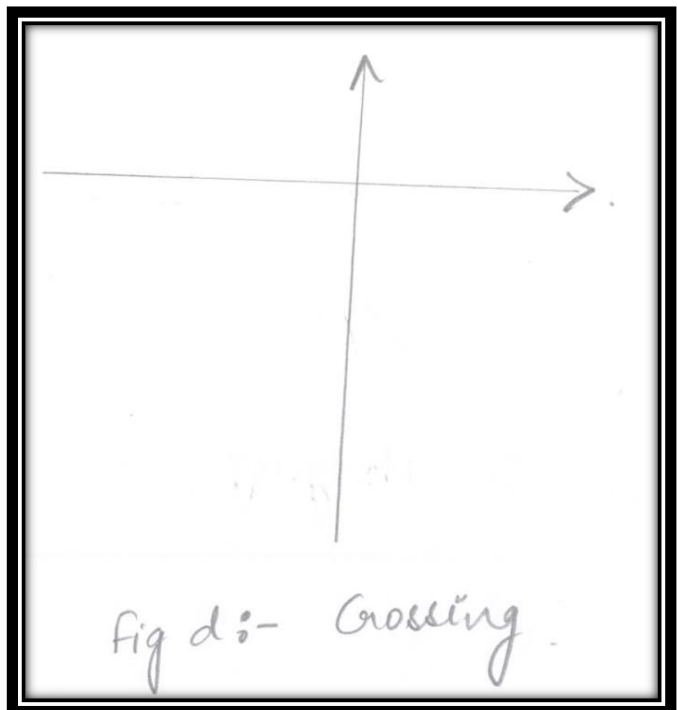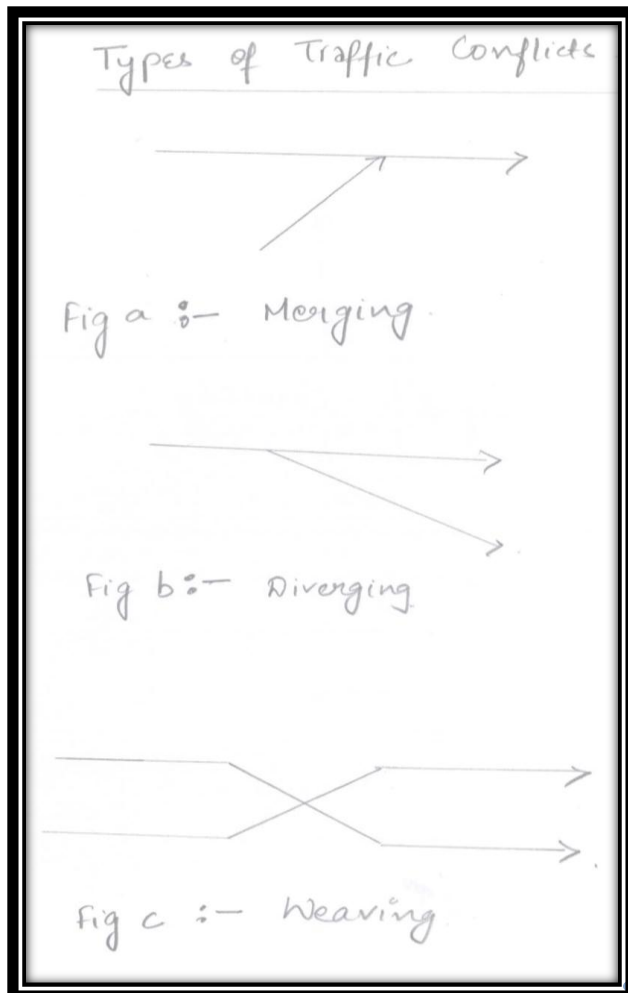
Figure 33 Traffic Conflicts

CHAPTER 8 - METHODOLOGY

AOI Selected → LANDSAT/SRTM Image Downloaded → Toposheets Obtained

Road Network Data Overlayed ← Toposheets Mosaiced to cover the AOI ← Toposheets Georeferenced

Real Time Accident Sites Overlayed (2007-2011) → Hotspots created for the Selected AOI → Service Area maps generated

Global Trends Predicted ← Kriging and Co-kriging Calculated ← Kernel Density Calculated

Cross Variance Clouds Generated → Slope and Curvature Calculated → Map Layout Generated

Web Interface Designed ← Survey Conducted ← Statistical Analysis Performed

CHAPTER 9 – INPUTS REQUIRED

Toposheets Used (NE 43-1 and NE 43 - 5)
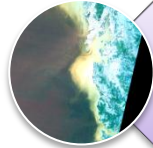
LANDSAT Imagery (7 bands, stacked together)

SRTM Imagery of the Western Coastal Region

Outline of Mumbai covering the AOI

Road Network Shapefiles

Point Data for POI location
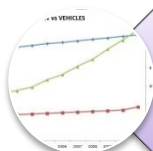
Statistical Data
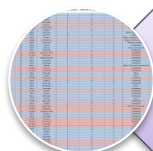
Survey Data

CHAPTER 10 –DATA CREATION

10.1 Introduction

In many ArcMap applications, users use basemaps on which operational information, observations and derived information from analytic models are displayed and are then used.

- Many cities use road network data as their basemaps, on which different layers are overlaid which shows occurrence of any event or incident
- Orthoimagery are generally used as base maps on which operational information are often overlaid

A basemap is mainly used for locational reference and provides a framework on which users overlay or pile up their operational layers, perform tasks and visualize geographic information. The basemap serves as a foundation for all subsequent operations and mapping. Many maps include multiple base layers.

10.2 Point Data Creation

Point data formation was one of the important chunks of the data creation. In this project, we have created the data for accident sites. These data was formed based on ground survey that was carried out by us as well as by studying the statistical data that was available to us. Along with these available data, logical reasoning was also applied so that it made it easier for us to identify and geocode the accident sites.

The accident site that was created was further divided into two sets. Since the data collected was huge, for that very purpose it was necessary for us to categorise them into these sections for a better representation.
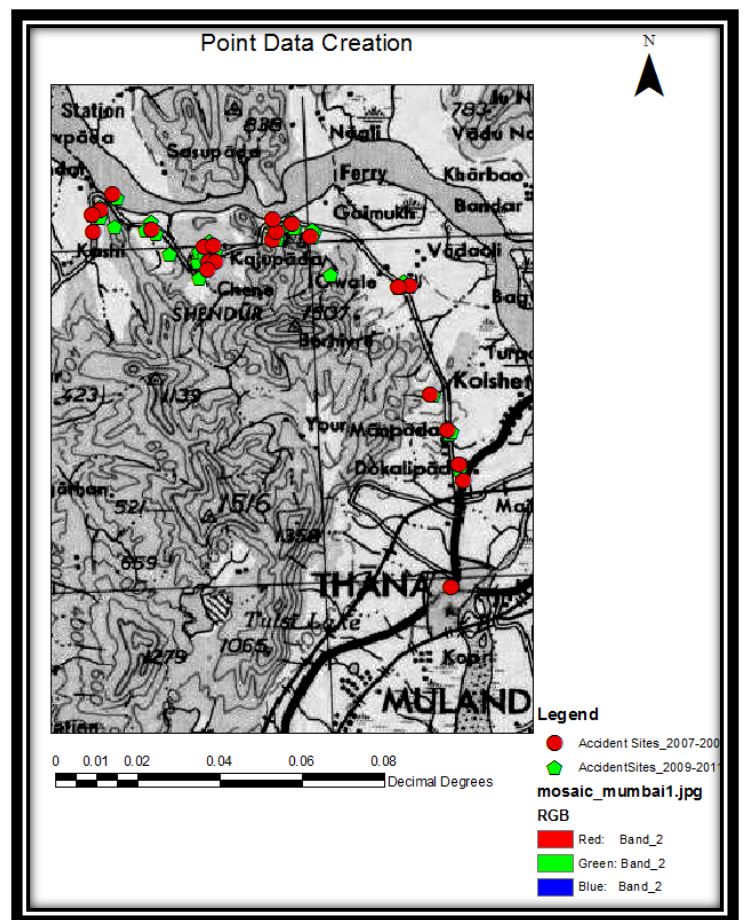
Figure 34 Point Data Creation on ArcGIS

10.3 AOI Digitization

Digitizing is the process of interpreting and converting paper map or image data to vector digital data. In manual digitizing you trace the lines or points from the source media. On-screen digitization involves digitization of a set of features from a scanned photo or image. Digitizing on screen is often called "heads-up" digitization.

For the completion of this project, we have digitized out AOI, which is the Thane-Ghodbunder Road from the georeferenced toposheet.

10.4 TIN Creation

It is an acronym for Triangular Irregular Network. It is a digital data structure used in a GIS for representation of a surface. A TIN is a vector based representation of the physical land surface, made from

Figure 35 AOI Digitization

irregularly distributed nodes and lines with 3D coordinates (x, y and z) that are arranged in a network of non-overlapping triangles. Since the nodes can be placed irregularly over a surface, TINs can have a higher resolution in areas where surface is highly variable or where more detail is desired and lower resolution in areas that are less variable. TINs were first invented by Phil Mellor whilst studying Sociology at University of Edinburgh.

A TIN comprises a triangular network of vertices, known as mass points, with associated coordinates in three dimensions connected by edges to form a triangular tessellation. Three-dimensional visualizations are readily created by rendering of the triangular facets. In regions where there is little variation in surface height, the points may be widely spaced whereas in areas of more intense variation in height the point density is increased. . ArcGIS supports the Delaunay triangulation method.

The resulting triangulation satisfies the Delaunay triangle criterion, which ensures that no vertex lies within the interior of any of the circumcircles of the triangles in the network. If the Delaunay criterion is satisfied everywhere on the TIN, the minimum interior angle of all triangles is maximized. The result is that long, thin triangles are avoided as much as possible.

Contours are a common source of digital elevation data. In general, all the vertices of the contour lines are used as mass points for triangulation. In many cases, flat triangles are created in the surface. Flat triangles are created whenever a triangle is formed from three nodes with the same elevation value. Flat triangles are frequently generated along contours when the sample points occur along the contour at a distance that is less than the distance between contours. When these "excess" vertices are not removed, the Delaunay triangulation discovers that the closest sample points are those along the same contour, causing the generation of flat triangles.

The flat triangles have a slope of 0 and do not have defined aspect. They might cause problems when the surface is used for modeling. The input features used to create a TIN remain in the same position as the nodes or edges in the TIN. This allows a TIN to preserve all the precision of the input data while simultaneously modeling the values between known points.

TINs are made from mass points, breaklines, and polygons. Mass points are point height measurements; they become nodes in the network. Mass points are the primary input into a TIN; they determine the overall shape of the surface.

Breaklines are lines with or without height measurements. They become sequences of one or more triangle edges. Breaklines typically represent either natural features, such as ridgelines or streams, or built features, such as roadways. Polygons represent surface features with area—such as lakes—or boundaries (also called hulls) of separately interpolated areas. Hulls could define the shores of individual islands in an archipelago, or the boundary of a study area.
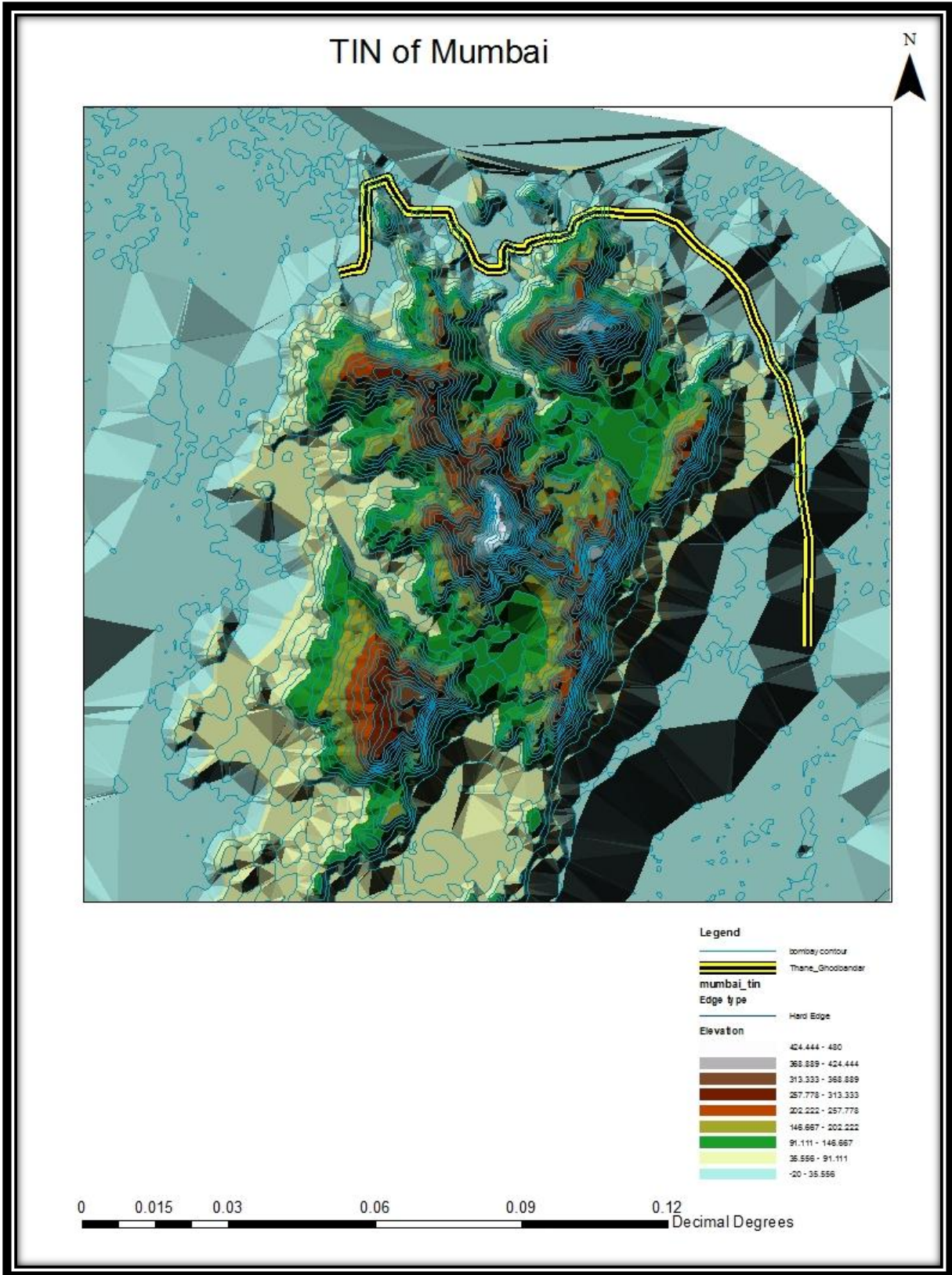
Figure 36 Representation of TIN

• • •

83

10.4.1 RASTER Creation from TIN

A raster surface can be converted to a TIN for use in surface modeling or to simplify the surface model for visualization. Conversion to a Triangular Irregular Network (TIN) can also allow us to enhance the surface model by adding features, like streams and roads that were earlier not present in the original raster.

When a raster is converted to a TIN, a certain number of raster mesh points become nodes in the TIN. (A mesh point is a location where four cell corners meet.)

The number of mesh points used to create the TIN is the smallest number that satisfies two conditions. First, the output TIN must cover the entire surface area of the input raster. Second, a user-specified z-tolerance must be met. The z-tolerance is a number that limits z-value differences between the input and output surfaces.

Raster to TIN first generates a candidate TIN using sufficient input raster points (cell centers) to fully cover the perimeter of the raster surface. It then incrementally improves the TIN surface until it meets the specified Z tolerance. It does so by adding more cell centers on an as-needed basis during an iterative process.

The interpolation zone of the input raster is revered in the output TIN. The polygonal representation of the interpolation zone is extracted from the raster, with z-values.

A large z-tolerance allows the TIN surface to conform less closely to the raster. The output TIN has fewer nodes and triangles and the conversion process is faster. A small z-tolerance makes the TIN conform more closely to the raster. The TIN has more nodes and triangles and takes longer to process.

The z-tolerance is the maximum allowed difference between the z-value of an input raster cell and the z-value of the output TIN at the location corresponding to the raster cell center.

To put it another way: suppose you convert a raster to a TIN, then identify a location on the surface to compare the input and output elevation values. If the location you click on corresponds to the center of a raster cell, you can be sure that the elevation difference between the two layers will not be greater than the z-tolerance. If the location you click on does not correspond to a cell center, the elevation difference may or may not be greater than the z-tolerance.
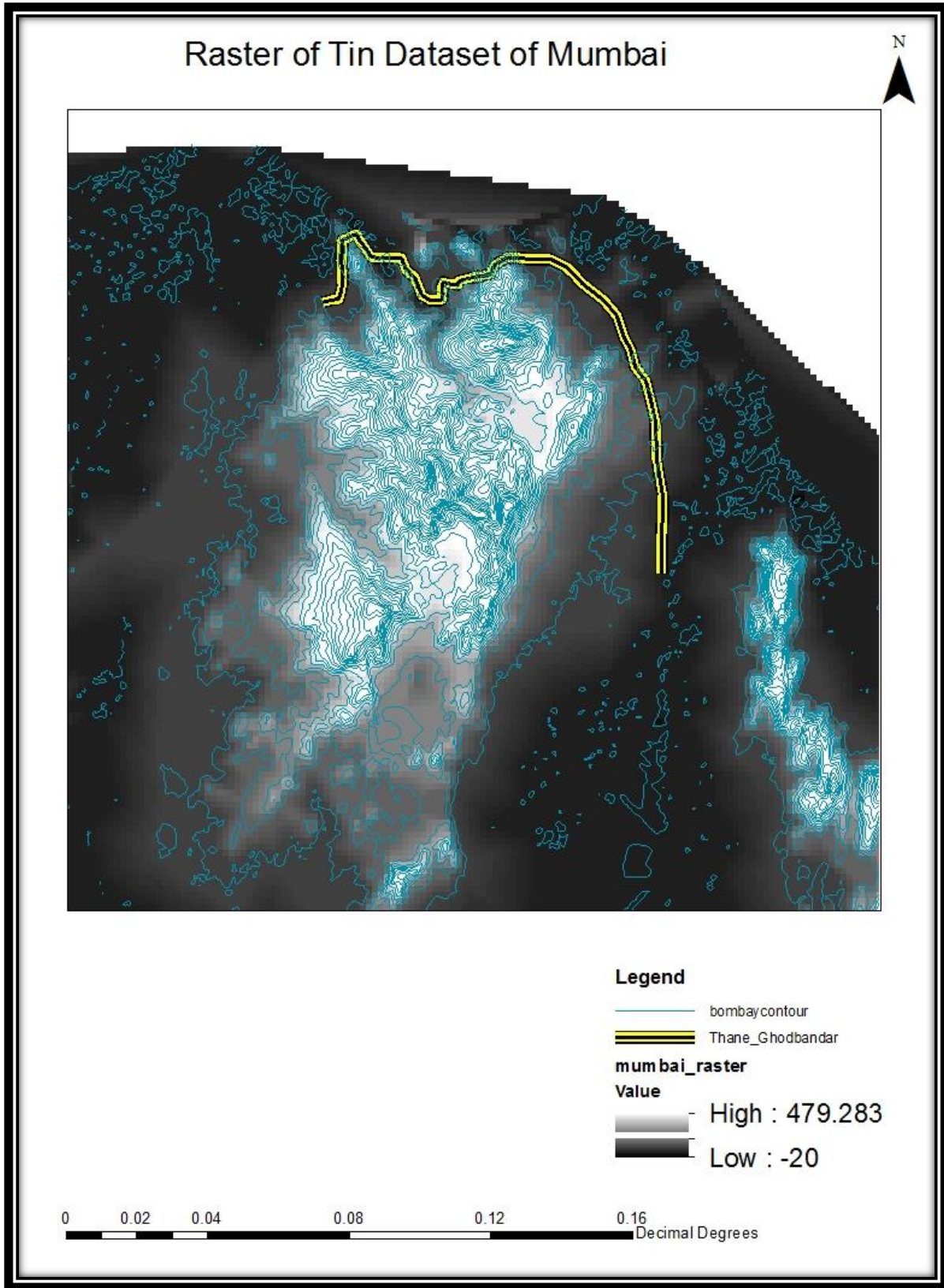
Figure 37 Raster Representation of TIN Dataset

10.5 DEM Formation

The generation of a DEM (Digital Elevation Model) from contour map data is a common Raster GIS application. The most common digital data of the shape of the earth's surface is cell-based DEM. This is used as input to quantify the characteristics of the land surface.

A digital elevation model (DEM) is based on a higher number of points with X-, Y- and Z- co-ordinates describing the bare soil. The DEM may be arranged in a raster or a random form. It is a representation of a continuous surface. The accuracy of this data is determined primarily by the resolution (the distance between sample points).

Errors in DEMs are usually classified as either sinks or peaks. A sink is an area surrounded by higher elevation values and is also referred to as a depression or pit.
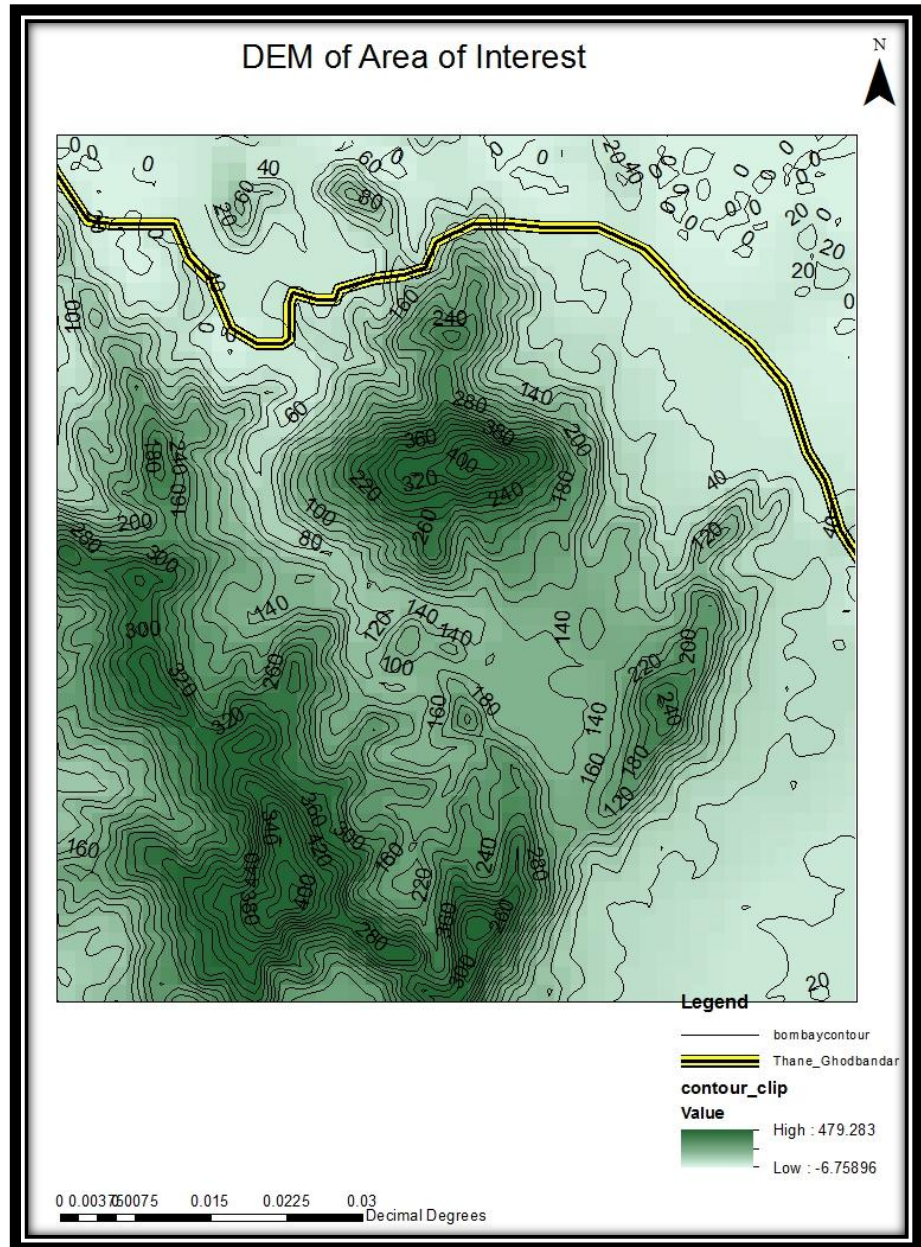
Figure 38 Representation of DEM along with AOI

CHAPTER 11 – ANALYSIS & DISCUSSION

11.1 Hotspots Analysis

Hotspot analysis is the investigation of the location the most dangerous sites plausible to get affected by accidents. It is carried out or is generally split into different phases. Firstly, the hazardous locations (hot spots) need to be identified. Next, the locations are ranked in order of their severity (which is calculated by Accident Rate).

Figure 39 Hotspot Analysis

Consequently, it is analyzed why few locations are hotspots and others are not.

Hotspots adapt the geometrical characteristics and the environmental features of the problematic locations in the existing road network. This task entails localizing and treating intersections and road segments with an unusual high accident rate $(\frac{\textbf{Accident Count}}{\textbf{Fleet Count}})$**,** the so-called hotspots. To reduce the number of accidents, it is important to identify the concentration of these accidents. As a result, the geographical aspect (slope and curvature) plays an important role to describe and tackle the problem of accidents. The importance of geographical aspect is worth acknowledging, but very often statistical – non-spatial regression models are used to replicate the number of accidents with respect to some variables. This section illustrates the use of Moran's index for accidents along State Highway-42 (Thane-Ghodbunder Road) in Maharashtra. An important issue for accident analysis is the variability, i.e., the fact that the yearly number of accidents on the road segment varies for each year. This can be explained by the inherent accident risk of a road segment. The randomness in the number of accidents depends on factors which cannot always be predicted. Therefore, it is of utmost importance to study the road network.

Hotspot analysis uses vectors and not rasters to identify the locations of statistically significant hotspots and coldspots in the data. It produces Z scores and P values. A high Z score and small P value for a feature indicates a significant hotspot. A higher (or lower) the Z score, the more intense the clustering. A Z score near zero means no spatial clustering.

11.1.1 Workflow for Hotspot Analysis

Individual Point Location

Hierarchichal Techniques (Nearest Neighbor)

Partioning Techniques (K-means)

Cluster Identification (Kernel Density)

Clumping Techniques (Grouping of clusters)

Hotspots Generation
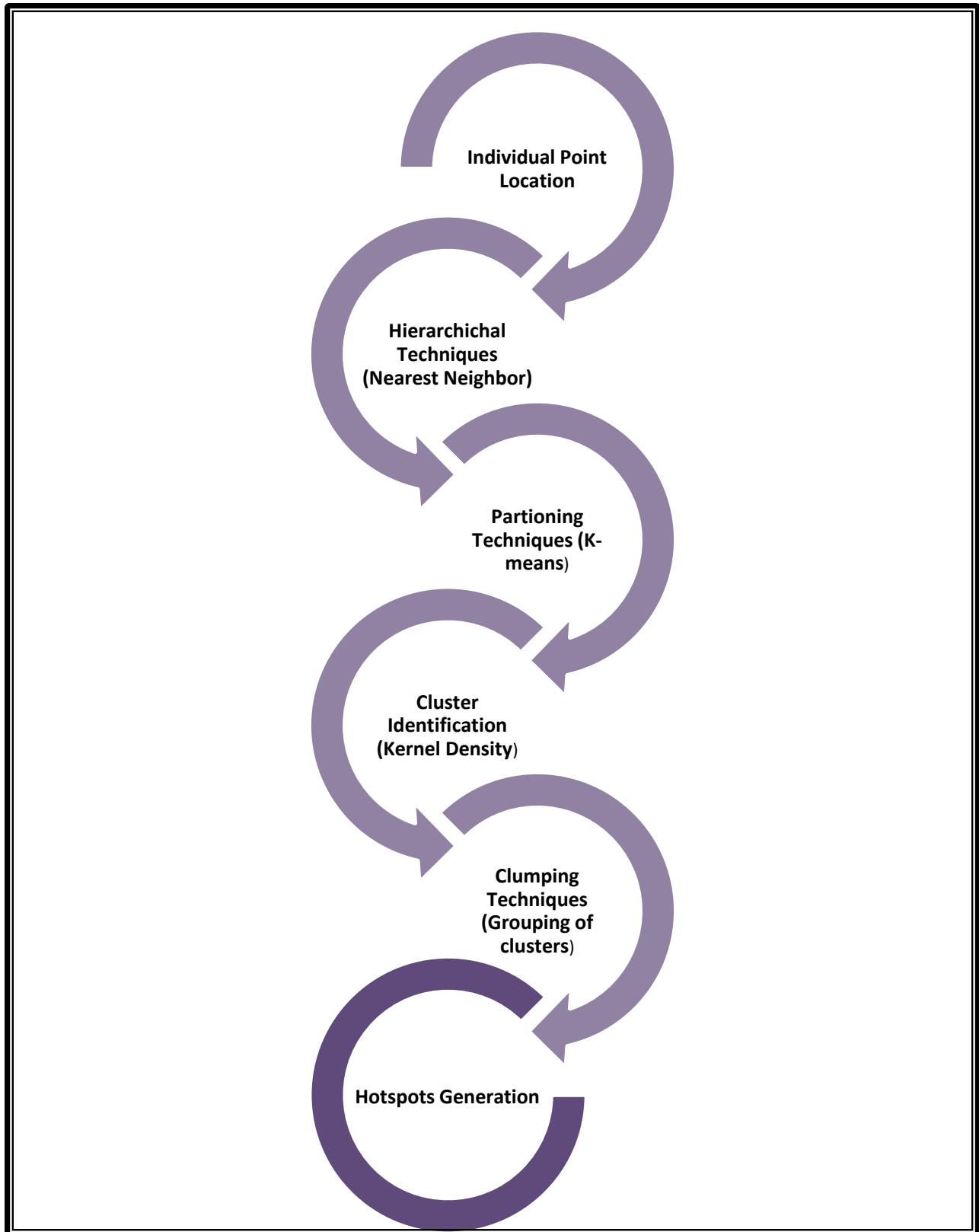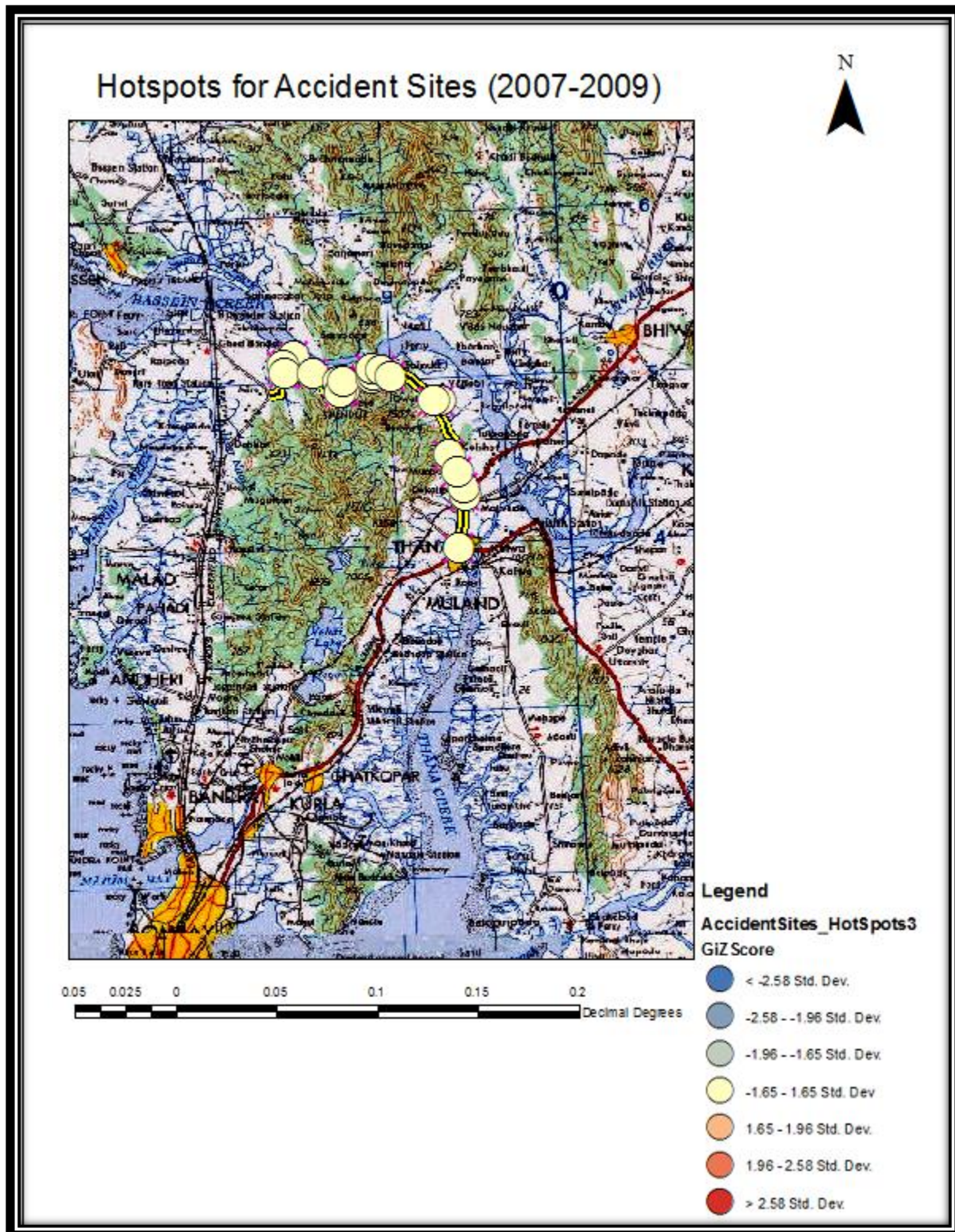
Figure 40 Hotspots around the Thane-Ghodbunder Road (2007-2009)

This shows the Average Nearest Neighbour along with the Dataset Information. It also gives the Z score and the p value for the hotspots generated for the year 2007-2009.

Nearest Neighbor Ratio: 0.640595
z-score: -3.224972
p-value: 0.001260

Significance Level (p-value)	Critical Value (z-score)
0.01	< -2.58
0.05	-2.58 – -1.96
0.10	-1.96 – -1.65
---	-1.65 – 1.65
0.10	1.65 – 1.96
0.05	1.96 – 2.58
0.01	> 2.58

(Random)

Significant   Significant

Clustered   Random   Dispersed

Given the z-score of -3.22, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Average Nearest Neighbor Summary

Observed Mean Distance:	556.078989
Expected Mean Distance:	868.065892
Nearest Neighbor Ratio:	0.640595
z-score:	-3.224972
p-value:	0.001260

Dataset Information

Input Feature Class:	Accident Sites_2007-2009
Distance Method:	EUCLIDEAN
Study Area:	66311378.561595

Figure 41 Average Nearest Neighbor along with the Dataset Information for 2007-2009
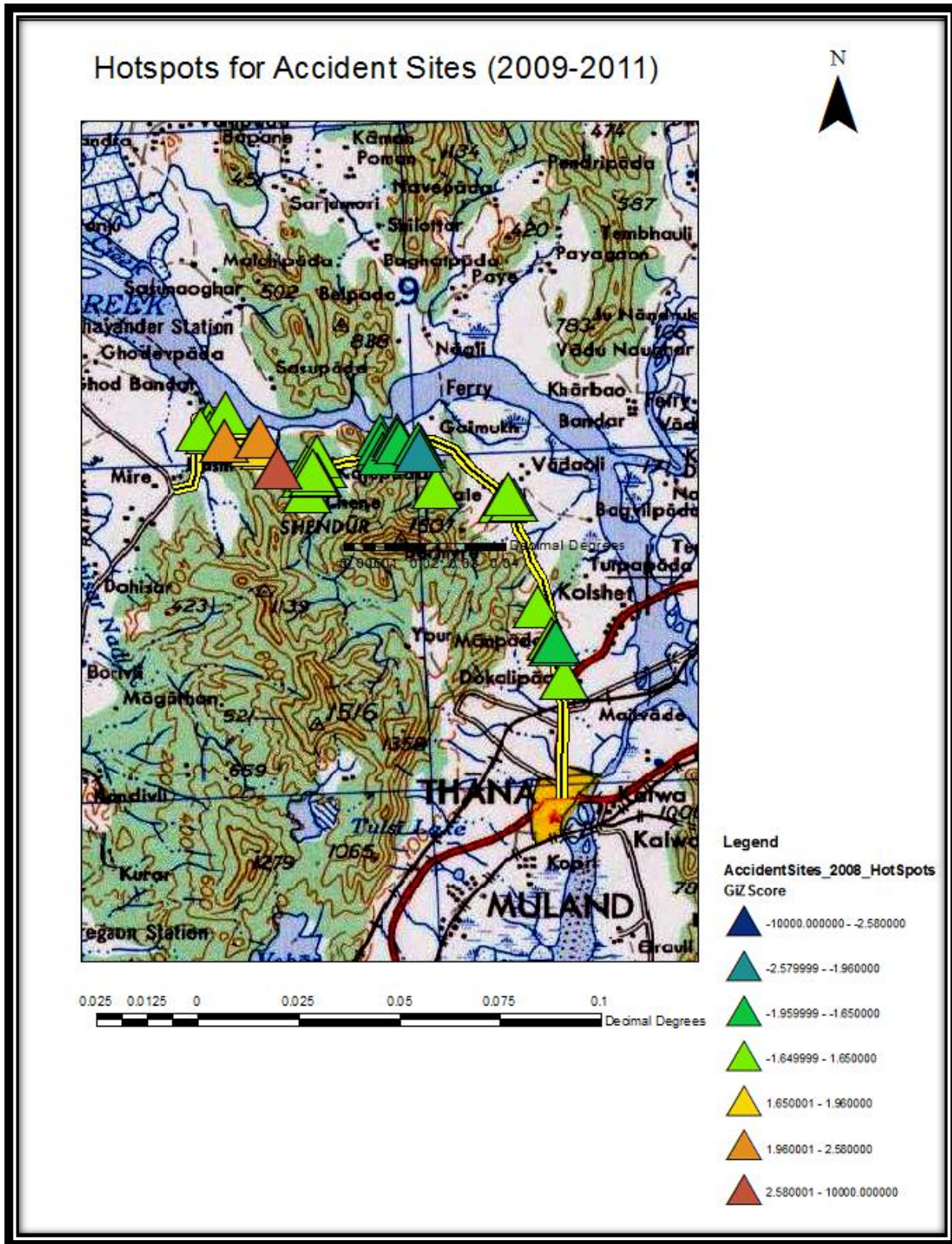
Figure 42 Hotspots along the Thane-Ghodbunder Road (2009-2011)

This report shows the Dataset information along with the summary of the average nearest neighbour. It gives us the Z score as well the p value. The observed and the expected mean distance are also mentioned in this report.
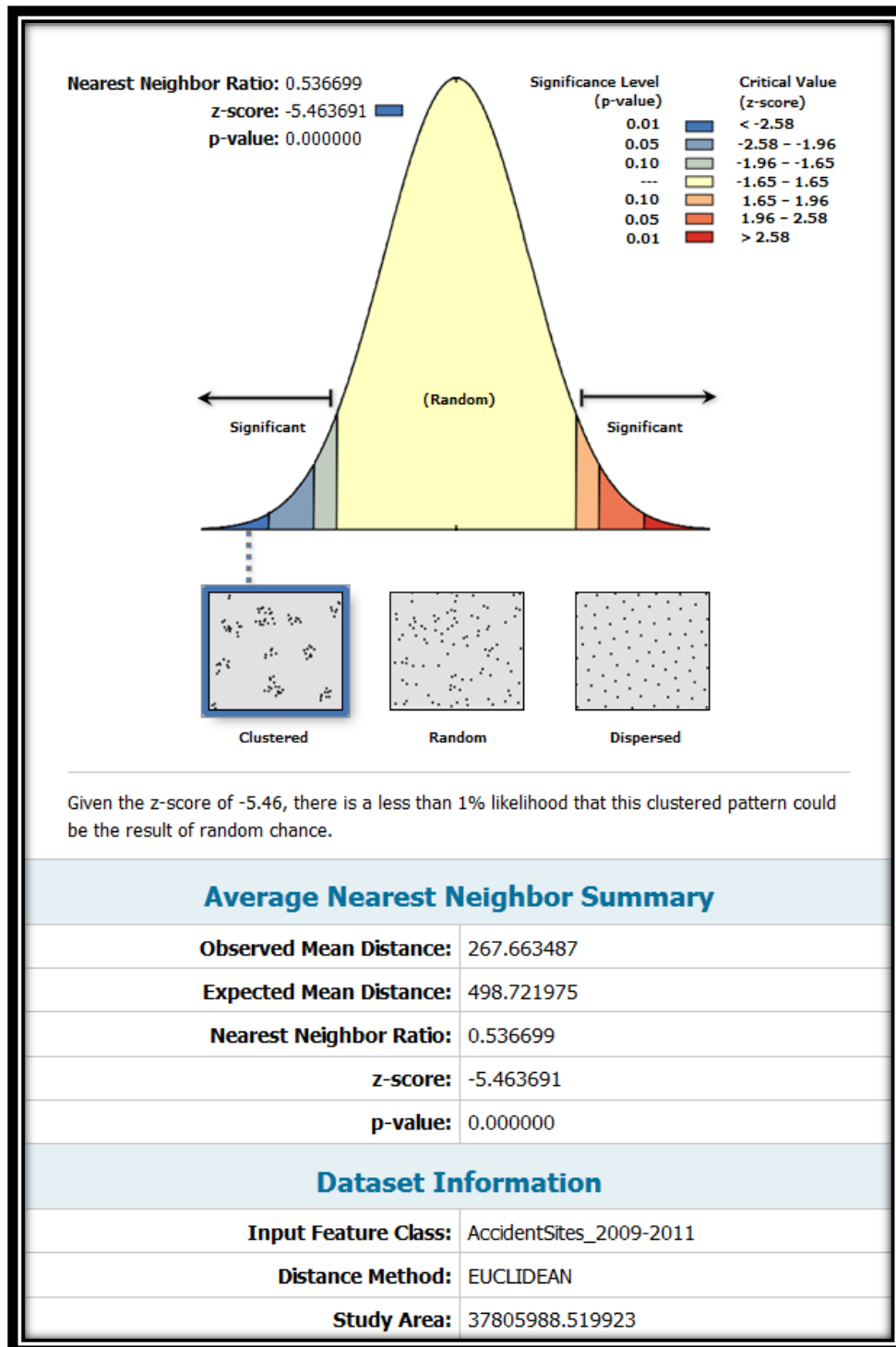
Figure 43 Average Nearest Neighbor along with the Dataset Information for 2007-2009

SYMBIOSIS INSTITUTE OF GEOINFORMATICS, PUNE

This attribute table gives us a view at the co-ordinates of the different hotspots along with its accident count as well as the fleet count. This in turn generates the Accident Rate, which has thus helped in the generation of the hotspot site for the year 2007-2009.

FID	Shape *	Id	Latitude	Longitude	Accident_C	Fleet_C	Accident_R
0	Point	0	19.281	72.894	26	130	0.2
1	Point	0	19.285	72.893	45	132	0.340909
2	Point	0	19.286	72.895	2	126	0.15873
3	Point	0	19.29	72.892	0	126	0
4	Point	0	19.281	72.908	33	136	0.242647
5	Point	0	19.277	72.921	25	126	0.198413
6	Point	0	19.273	72.922	23	136	0.169118
7	Point	0	19.273	72.923	0	136	0
8	Point	0	19.271	72.922	1	132	0.757576
9	Point	0	19.283	72.937	2	129	0.015504
10	Point	0	19.283	72.942	3	132	0.022727
11	Point	0	19.281	72.938	0	132	0
12	Point	0	19.279	72.937	1	130	0.769231
13	Point	0	19.281	72.947	2	136	0.147059
14	Point	0	19.279	72.947	5	132	0.378788
15	Point	0	19.267	72.968	0	131	0
16	Point	0	19.268	72.971	0	140	0
17	Point	0	19.241	72.976	6	132	0.454545
18	Point	0	19.232	72.98	1	128	0.78125
19	Point	0	19.224	72.983	4	132	0.30303
20	Point	0	19.22	72.984	0	132	0
21	Point	0	19.194	72.981	2	132	0.151515

(0 out of 22 Selected)

Accident Sites_2007-2009

Table 1 Attribute Table showing Accident Rate for 2007-2009

FID	Shape *	Id	Latitude	Longitude	Accident_C	Fleet_C	Accident_R
0	Point	0	19.285	72.893	59	158	0.373418
1	Point	0	19.287	72.895	20	159	0.125786
2	Point	0	19.284	72.896	42	150	0.28
3	Point	0	19.282	72.899	3	150	0.02
4	Point	0	19.283	72.908	5	159	0.031447
5	Point	0	19.281	72.906	19	159	0.119497
6	Point	0	19.28	72.909	35	160	0.21875
7	Point	0	19.275	72.912	65	160	0.40625
8	Point	0	19.269	72.92	50	158	0.316456
9	Point	0	19.273	72.919	0	149	0
10	Point	0	19.276	72.92	53	153	0.346405
11	Point	0	19.275	72.921	0	155	0
12	Point	0	19.274	72.921	12	50	0.24
13	Point	0	19.274	72.922	46	149	0.308725
14	Point	0	19.273	72.922	2	145	0.013793
15	Point	0	19.276	72.924	0	150	0
16	Point	0	19.283	72.938	0	126	0
17	Point	0	19.281	72.938	65	160	0.40625
18	Point	0	19.28	72.938	16	132	0.121212
19	Point	0	19.279	72.938	9	146	0.061644
20	Point	0	19.283	72.942	0	146	0
21	Point	0	19.282	72.942	6	146	0.041096
22	Point	0	19.282	72.943	13	159	0.081761
23	Point	0	19.281	72.947	22	150	0.146667
24	Point	0	19.281	72.948	2	150	0.013333
25	Point	0	19.279	72.947	0	159	0
26	Point	0	19.267	72.968	0	150	0
27	Point	0	19.269	72.969	1	159	0.006289
28	Point	0	19.267	72.971	4	142	0.028169
29	Point	0	19.241	72.977	0	39	0
30	Point	0	19.233	72.98	23	159	0.144654
31	Point	0	19.232	72.98	21	156	0.134615
32	Point	0	19.225	72.983	14	156	0.089744
33	Point	0	19.224	72.981	13	159	0.081761
34	Point	0	19.223	72.983	0	156	0
35	Point	0	19.221	72.982	1	160	0.00625
36	Point	0	19.221	72.983	0	159	0
37	Point	0	19.221	72.986	0	159	0

(0 out of 38 Selected)

AccidentSites_2009-2011

This attribute table gives us a view at the co-ordinates of the different hotspots along with its accident count as well as the fleet count. This in turn generates the Accident Rate, which has thus helped in the generation of the hotspot site for the year 2009-2011.

Table 2 Attribute Table showing Accident Rate for 2009-2011

The Hotspot Analysis (Getis-Ord Gi*) tool runs to display the hotspots shown above. The attribute table given below shows the different Z scores and P values of different accident sites. The Getis-Ord local is given by:

Figure 44 The Hotspot Analysis Tool

$$G_i^* = \frac{\sum\limits_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum\limits_{j=1}^{n} w_{i,j}}{S \sqrt{\dfrac{\left[n \sum\limits_{j=1}^{n} w_{i,j}^2 - \left(\sum\limits_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}}$$

Equation 26

OBJECTID *	Shape *	SOURCE_ID	GiZScore Fixed 23048	GiPValue Fixed 23048	Accident_R
1	Point	0	-0.460461	0.645185	0.2
2	Point	1	-0.460461	0.645185	0.340909
3	Point	2	-0.424177	0.671437	0.15873
4	Point	3	-0.460461	0.645185	0
5	Point	4	-0.384334	0.700731	0.242647
6	Point	5	0.644795	0.51906	0.198413
7	Point	6	0.874158	0.382032	0.169118
8	Point	7	0.874158	0.382032	0
9	Point	8	0.874158	0.382032	0.757576
10	Point	9	0.874158	0.382032	0.015504
11	Point	10	-0.809173	0.418416	0.022727
12	Point	11	0.546312	0.584851	0
13	Point	12	0.546312	0.584851	0.769231
14	Point	13	-1.201775	0.229451	0.147059
15	Point	14	-0.809173	0.418416	0.378788
16	Point	15	-0.809173	0.418416	0
17	Point	16	-0.909111	0.363292	0
18	Point	17	0.859116	0.390277	0.454545
19	Point	18	0.859116	0.390277	0.78125
20	Point	19	0.859116	0.390277	0.30303
21	Point	20	0.859116	0.390277	0
22	Point	21	1.368442	0.171174	0.151515

Table 3 z-Score and p-Values for 2007-2009

OBJECTID *	Shape *	SOURCE_ID	GiZScore Fixed 23048	GiPValue Fixed 23048	Accident_R
1	Point	0	1.025817	0.304978	0.373418
2	Point	1	1.025817	0.304978	0.125786
3	Point	2	0.984316	0.32496	0.28
4	Point	3	2.101208	0.035623	0.02
5	Point	4	1.025817	0.304978	0.031447
6	Point	5	2.365909	0.017986	0.119497
7	Point	6	1.933883	0.053128	0.21875
8	Point	7	1.933883	0.053128	0.40625
9	Point	8	2.799687	0.005115	0.316456
10	Point	9	1.492301	0.13562	0
11	Point	10	1.492301	0.13562	0.346405
12	Point	11	0.948921	0.342661	0
13	Point	12	1.492301	0.13562	0.24
14	Point	13	1.492301	0.13562	0.308725
15	Point	14	1.492301	0.13562	0.013793
16	Point	15	1.492301	0.13562	0
17	Point	16	1.492301	0.13562	0
18	Point	17	1.492301	0.13562	0.40625
19	Point	18	-1.906874	0.056537	0.121212
20	Point	19	-1.906874	0.056537	0.061644
21	Point	20	-1.906874	0.056537	0
22	Point	21	-1.906874	0.056537	0.041096
23	Point	22	-1.906874	0.056537	0.081761
24	Point	23	-1.906874	0.056537	0.146667
25	Point	24	-1.906874	0.056537	0.013333
26	Point	25	-1.906874	0.056537	0
27	Point	26	-1.906874	0.056537	0
28	Point	27	-1.906874	0.056537	0.006289
29	Point	28	-2.149567	0.03159	0.028169
30	Point	29	-1.048898	0.294225	0
31	Point	30	0.183178	0.854658	0.144654
32	Point	31	0.183178	0.854658	0.134615
33	Point	32	0.183178	0.854658	0.089744
34	Point	33	-1.123183	0.26136	0.081761
35	Point	34	-1.399805	0.161572	0
36	Point	35	-1.726522	0.084254	0.00625
37	Point	36	-1.668389	0.095239	0
38	Point	37	-1.466591	0.142487	0

These tables give us a relation between the Z scores and the P values along with the accident rates that have taken place along the Thane-Ghodbunder Road.

The statistics that have been taken into account has been for the last 5 years, i.e. from 2007-2011.

Table 4 z-Score and p-Value for 2009-2011

The report below shows the Spatial Autocorrelation where the Global Moran's Index is specified along with the expected index. It also denotes the variance along with the Z-score and the p-Value. Euclidean Distance has been used for this and the distance threshold has been given as 55.
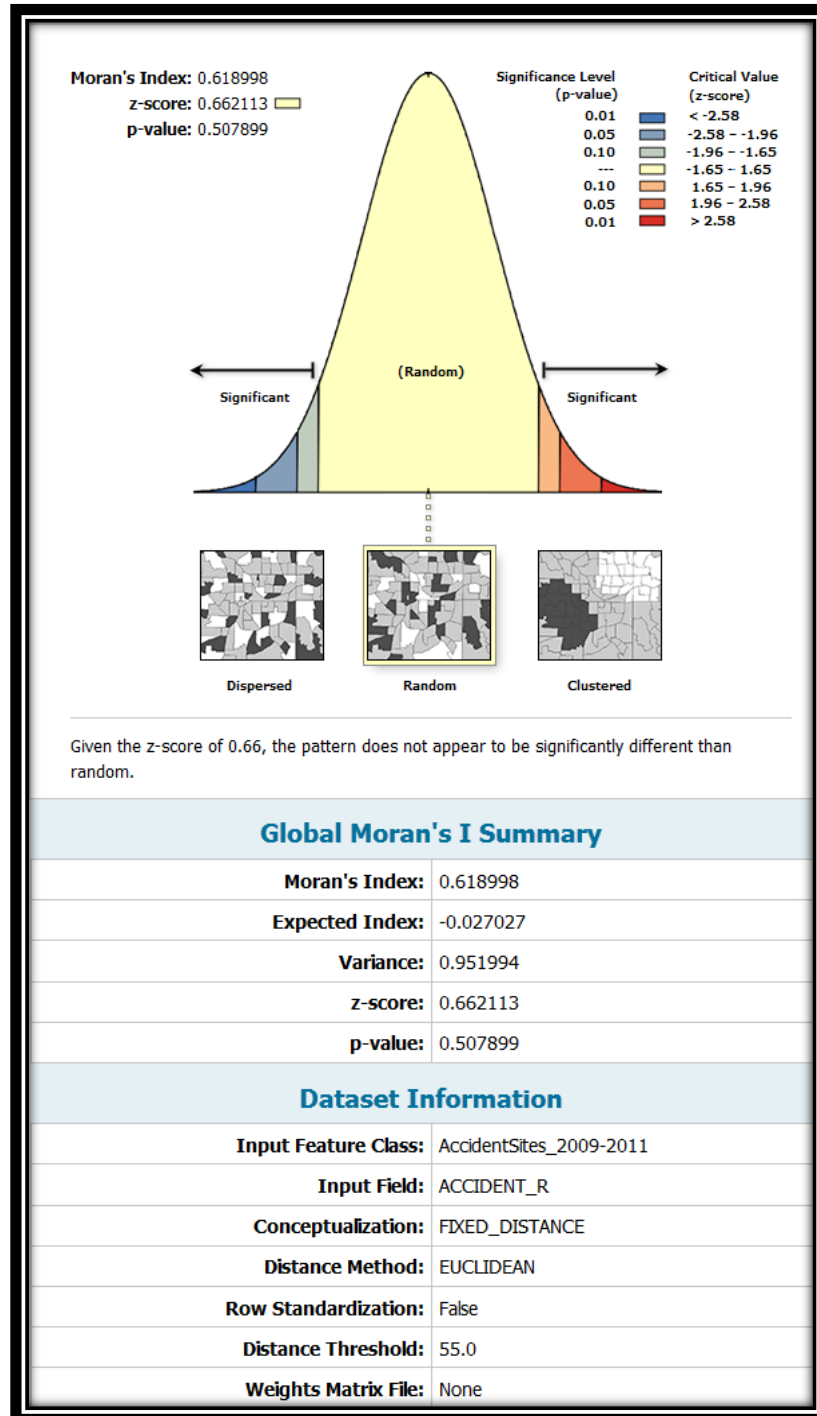
Figure 45 Moran's I Summary and Dataset Information

11.1.2 Kernel Density

Kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. In some fields such as signal processing and econometrics it is also termed the Parzen–Rosenblatt window method, after Emanuel Parzen and Murray Rosenblatt, who are usually credited with independently creating it in its current form. It has an advantage of simplicity but it also has disadvantages, such as lack of continuity.
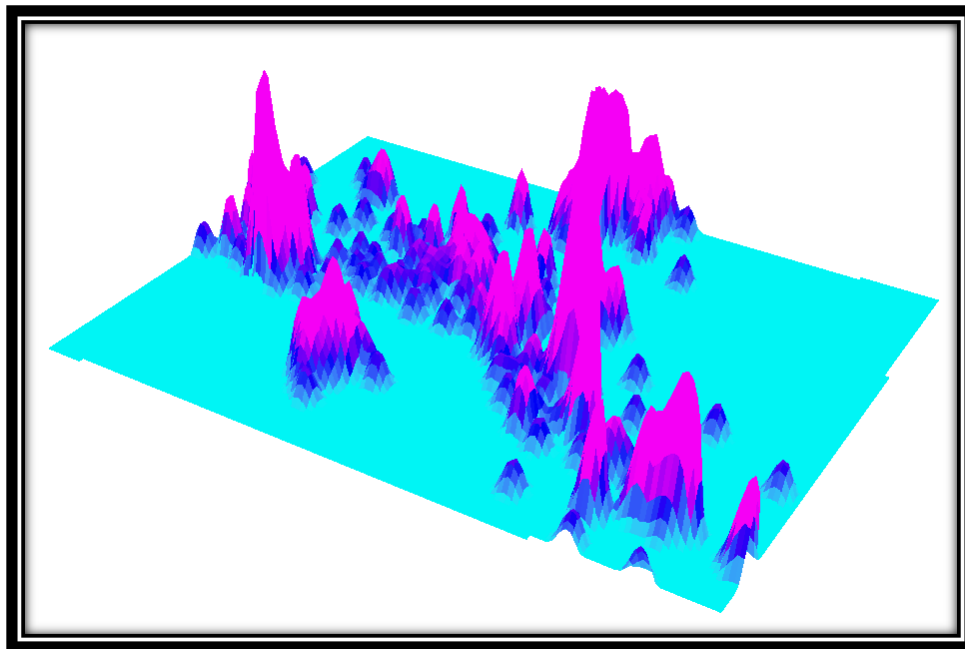
Figure 46 Kernel Density of Mumbai and its surroundings

The Kernel Density tool calculates the density of features in a neighbourhood around those features. It can be calculated for both line and point features. Conceptually, a smoothly curved surface is fitted over each point. The surface value is highest at the location of the point and diminishes with increasing distance from the point, reaching zero at the Search radius distance from the point. Only a circular neighbourhood is possible. The volume under the surface equals the Population field value for the point. The density at each output raster cell is calculated by

adding the values of all the kernel surfaces where they overlay the raster cell centre. The kernel function is based on the quadratic kernel function described in Silverman.

Increasing the radius will not greatly change the calculated density values. Although, more number of points fall inside the larger neighbourhood. This number will then be divided by a larger area when calculating density. This thus helps in generating a more generalized output raster
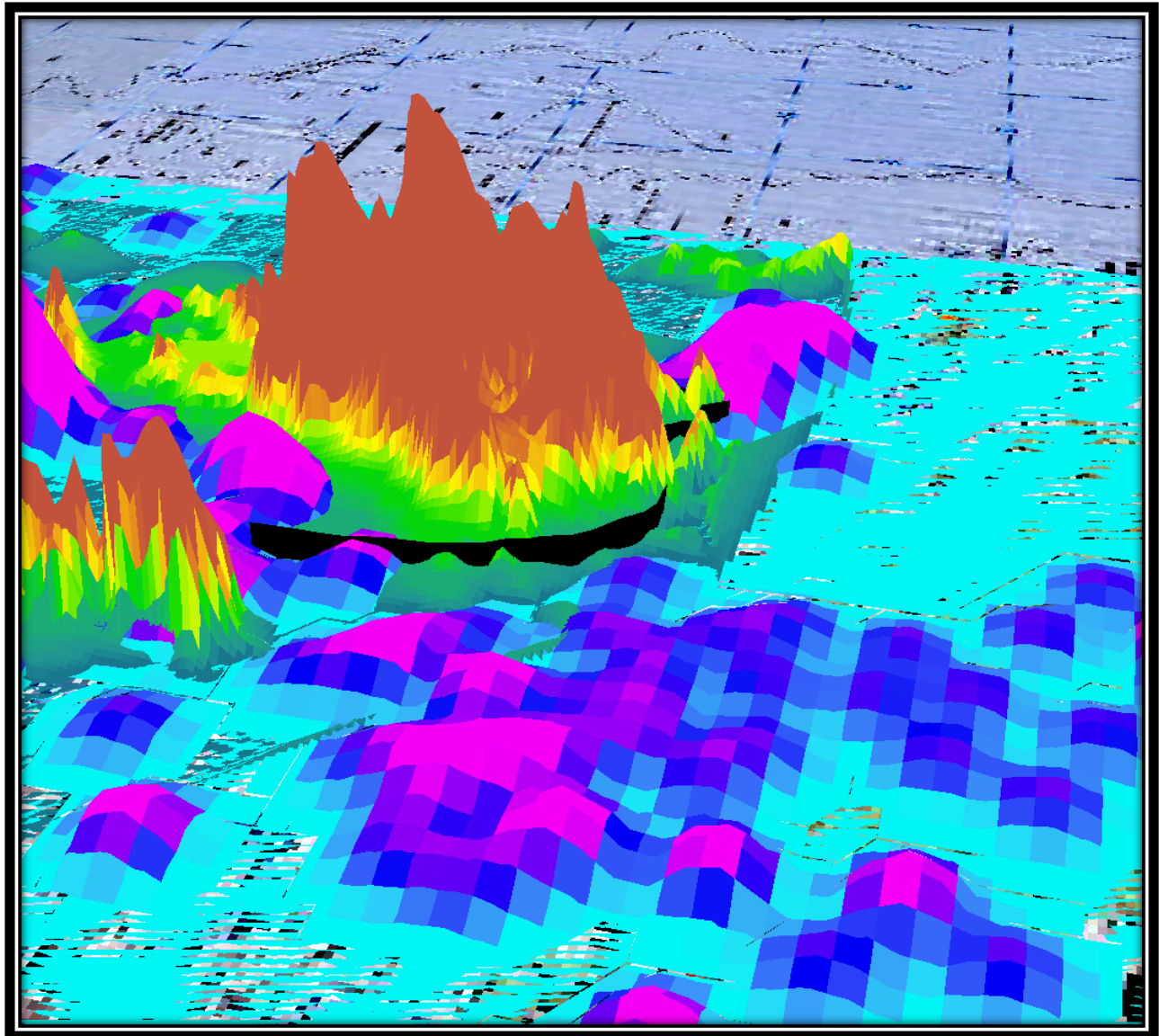
Figure 47 Kernel Density & Elevation along the AOI

11.2 Network Analysis

GIS deals with the interrelated nature of geographic data. Some of these data has a few specialized rules, parameters and restrictions on interaction. Measurements, relationship-testing and analysis, all these occur within this particular framework. Network Analysis focuses on edge-node topology to represent real life networks of information, which may include: driving direction, multi-
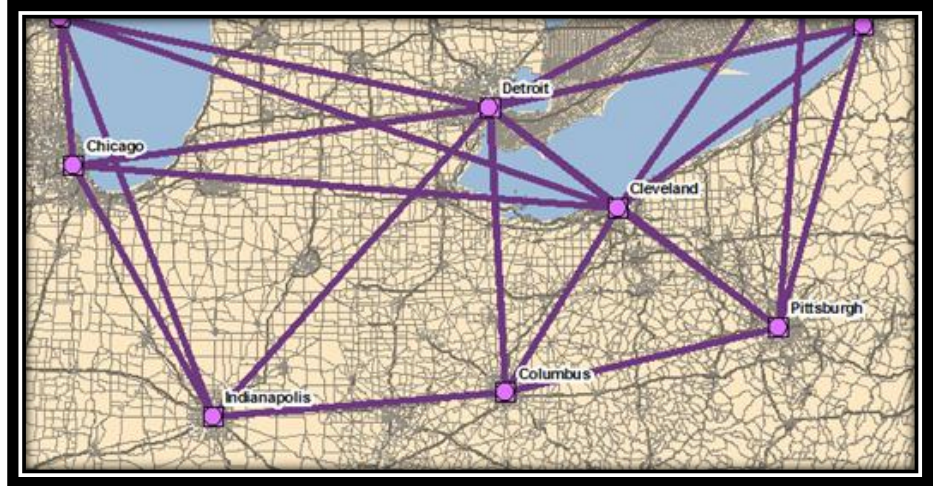
Figure 48 Networks being shown

modal routing, service roads, information networks and few more.

A series of nodes are connected to each other by edges. Each of these edges starts or ends with a node. Each node touches at least one edge and they have one or more edge connecting them. ArcGIS Network Analyst provides network based spatial analysis, such as routing, location allocation, service area, fleet routing, etc. It dynamically models realistic network conditions. Networks can easily be built using a sophisticated data model. Rules dictate how objects can move through the network. A set of nodes are connected by lines which represent some type of flow. This is in turn incorporated with flow rules.

For this project, we have created service areas for the important POIs (Hospitals and Police Stations) along the Thane-Ghodbunder Road. The purpose of the creation of the service area is because, any accident which takes place along this stretch, these POIs can cater to them earliest compared to the others present around. 9 hospitals and 4 police stations had been identified along this stretch. Furthermore, a 5 km radius has been specified for each of them and thus the service area maps have been thus, generated.

11.2.1 Workflow for Network Analysis

Police Stations near Area of Interest

N

Kashimira Police Station

Kasarwadavli Police Station

Patlipada Police Station

Text

Kapurbavdi Police Station

Police Stations
mumbai_network
Edges
Thane_Ghodbandar
mumbai.osm-mainroads_gen0

0.00306.012.018.024
Decimal Degrees
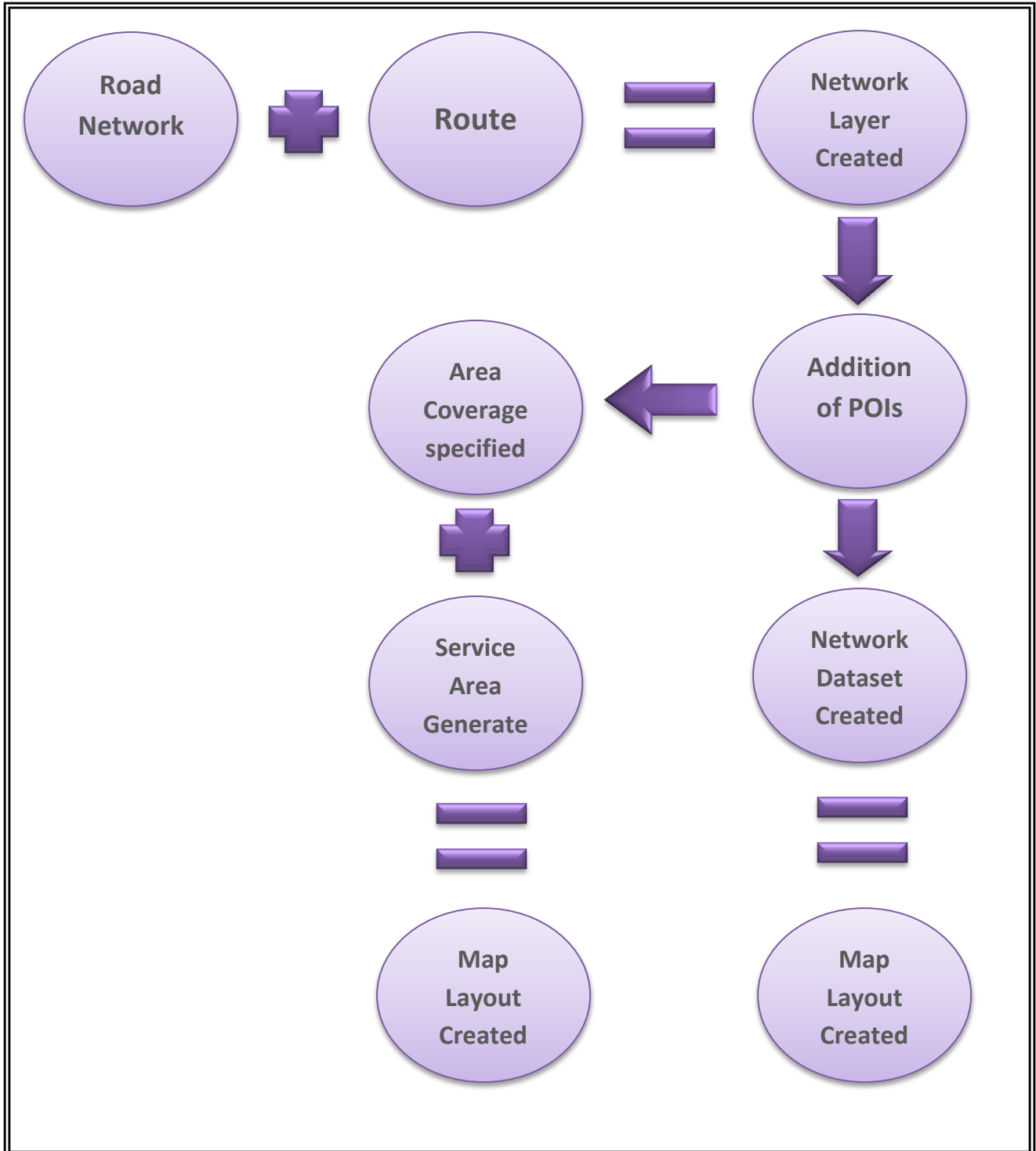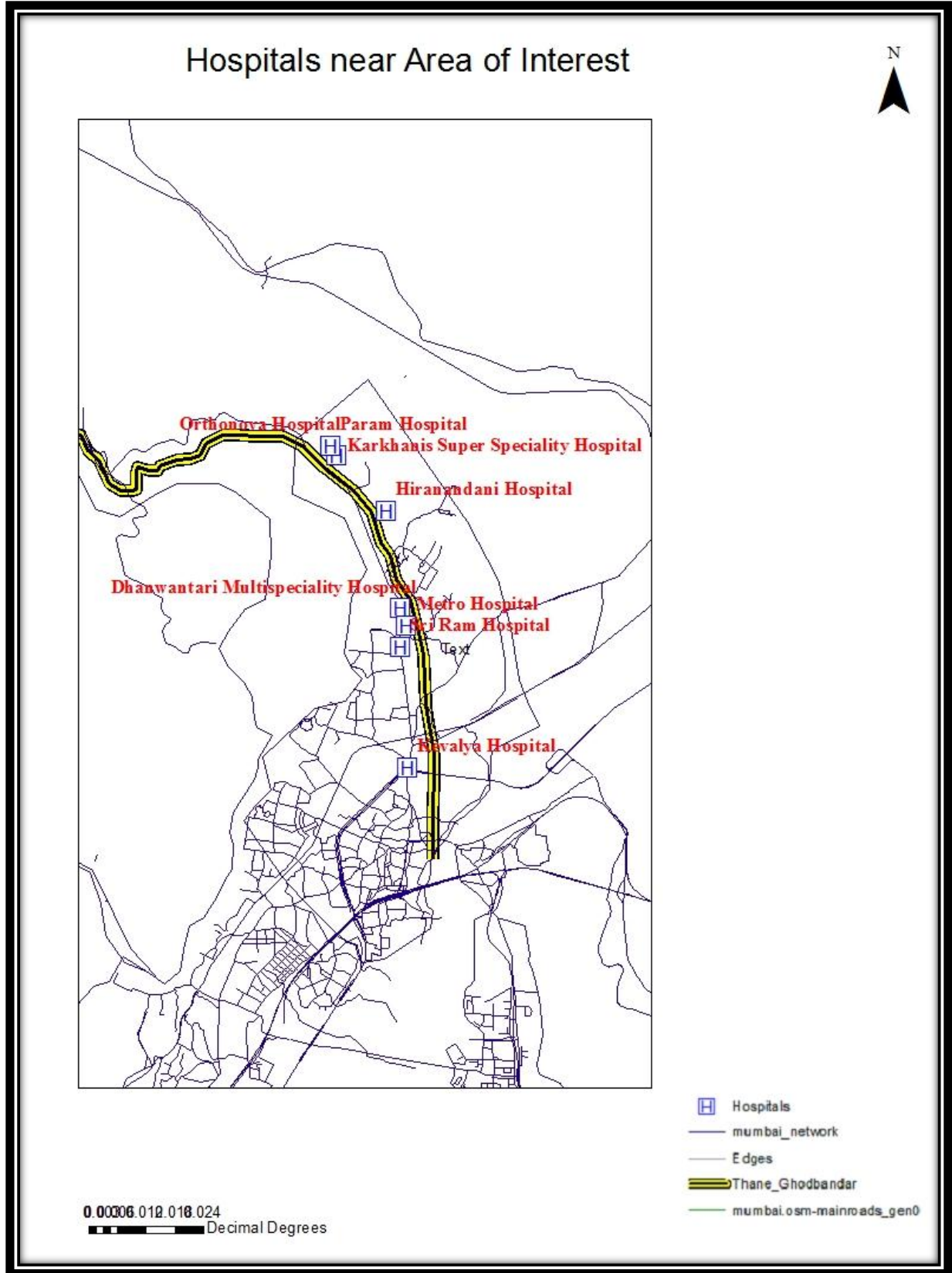
Fig 42: Representation of Important Police Stations in the Area of Interest

• • •
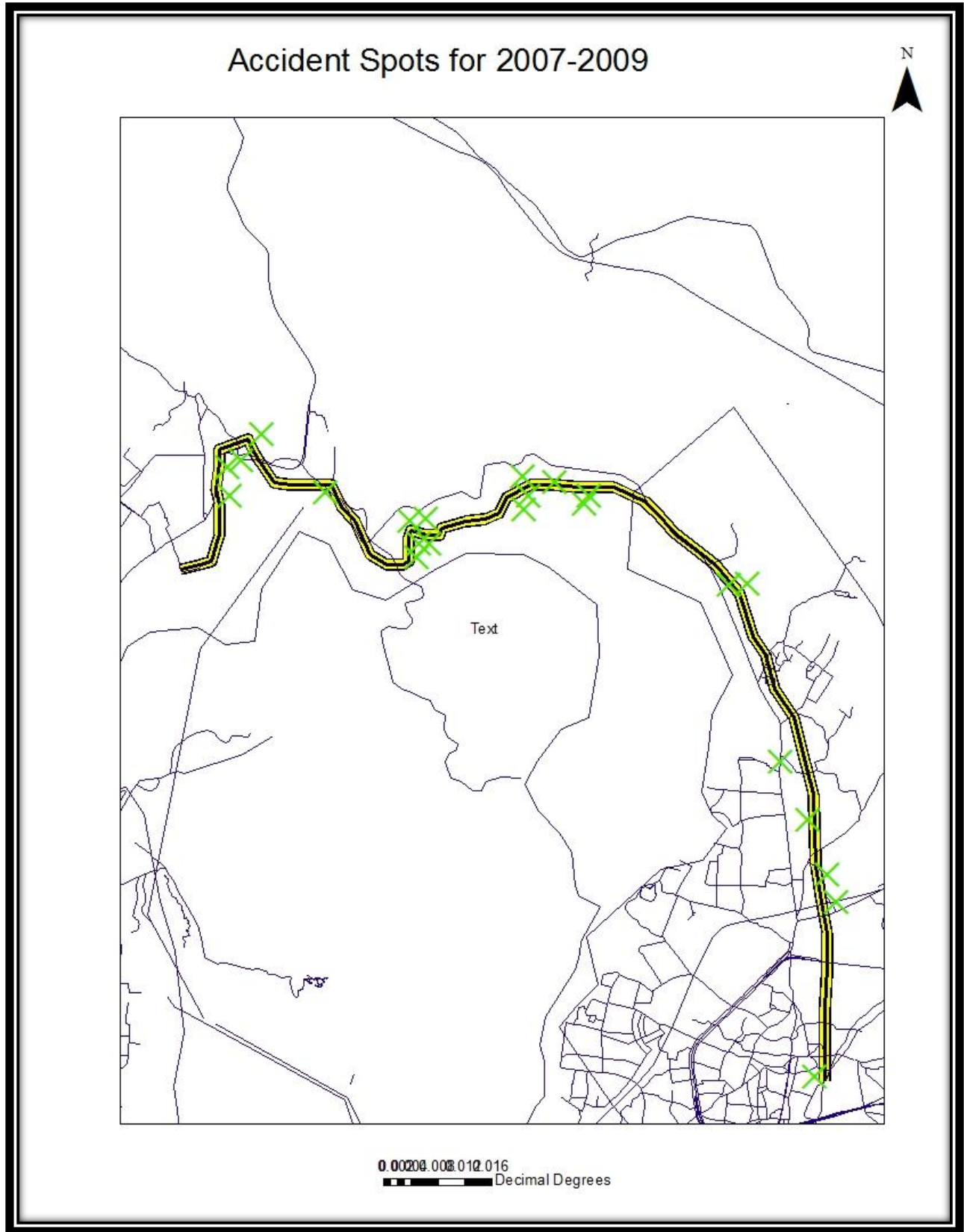
Fig 43: Representation of the Accident Spots for the year 2007-09
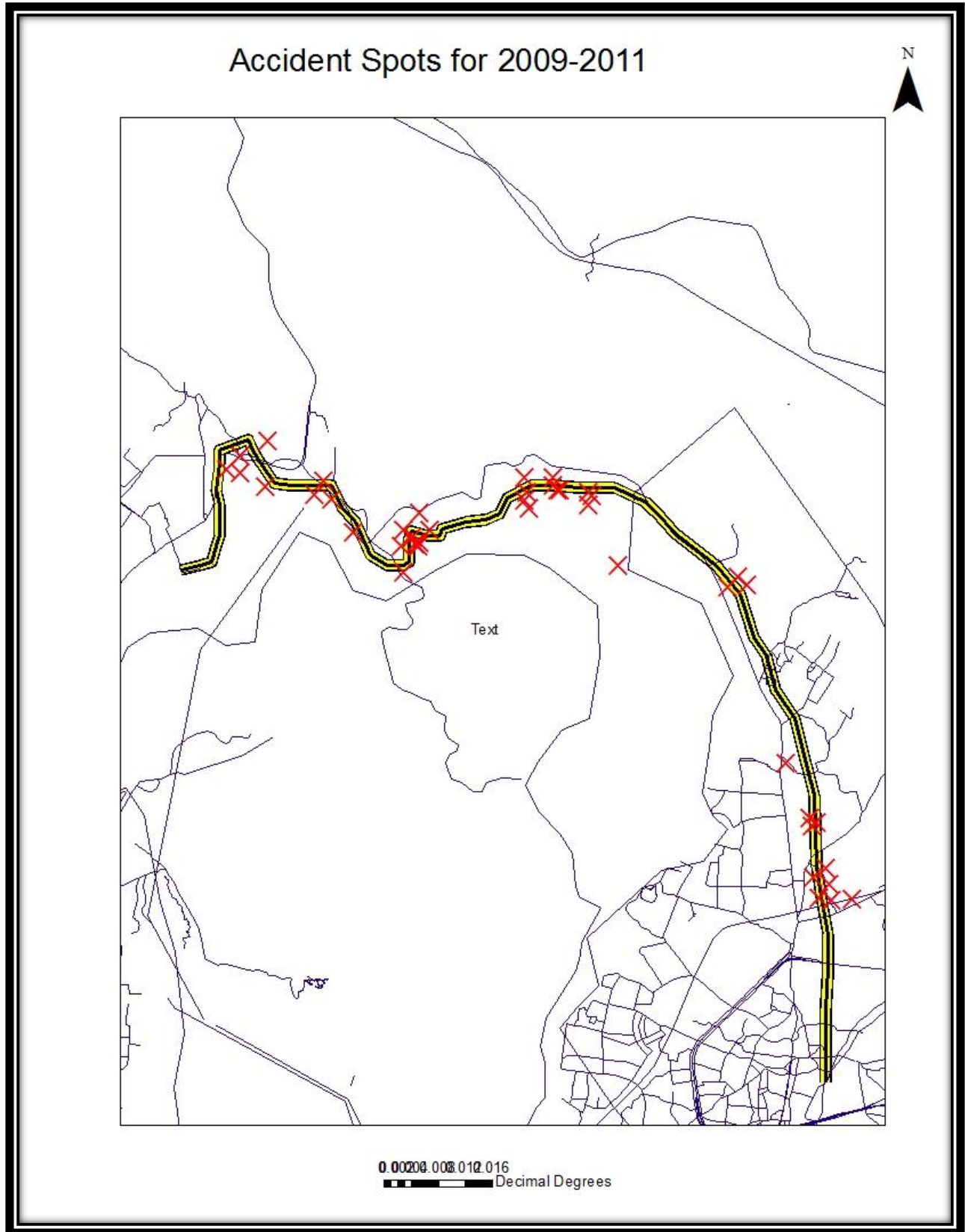
Fig 44: Representation of the Accident Spots for the year 2009-11
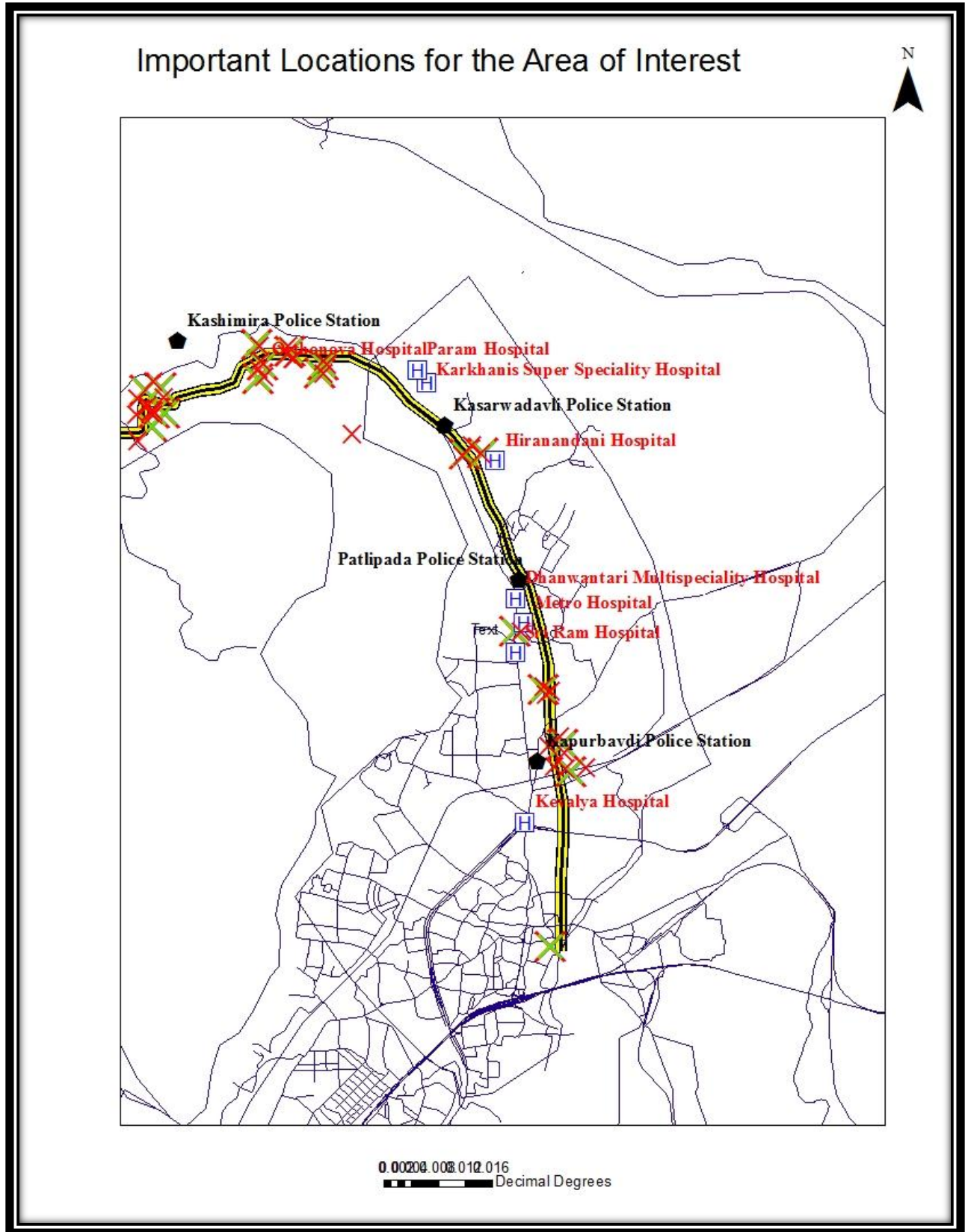
Figure 49 Locations of the POIs along the AOI
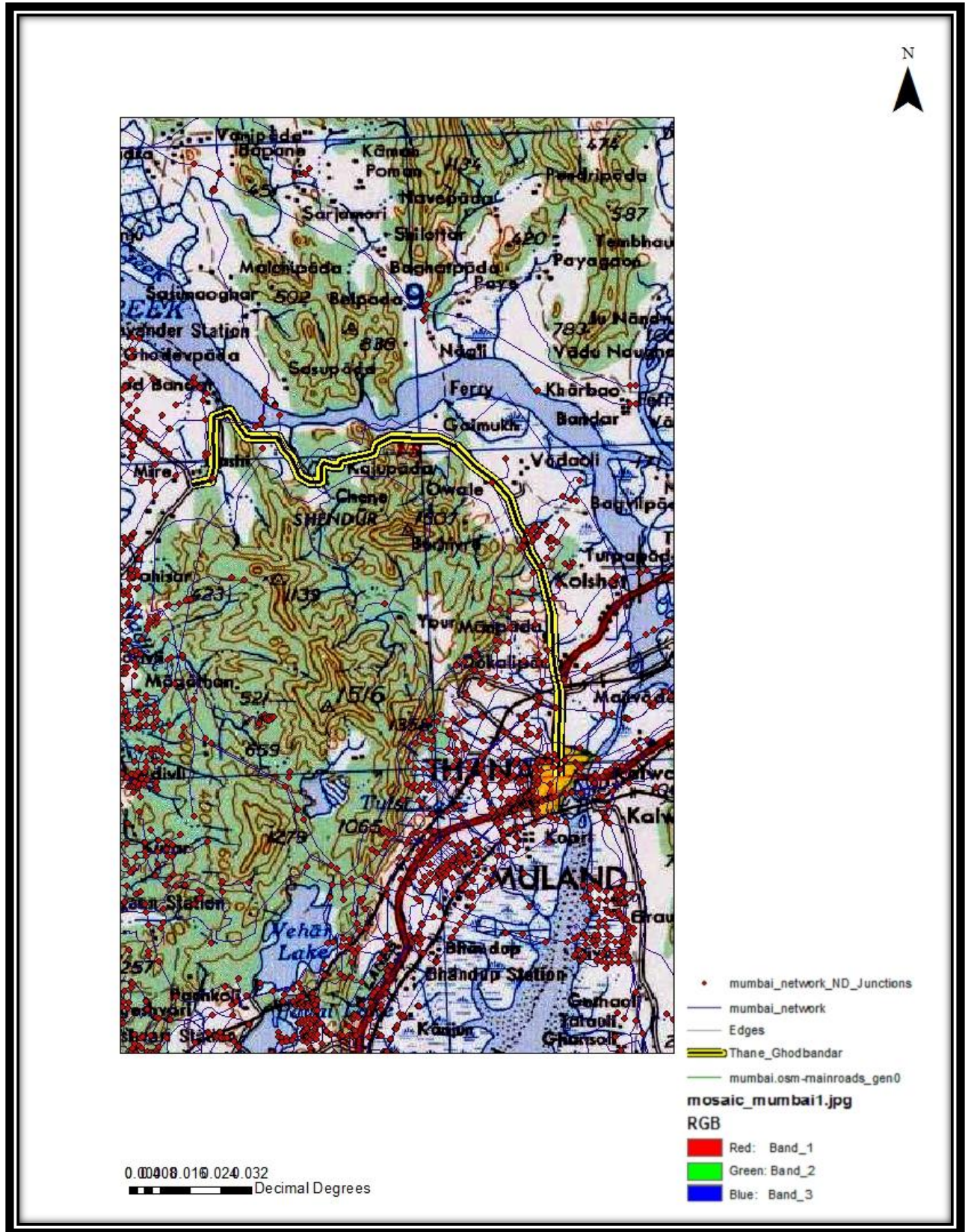
Figure 50 Overlay of the AOI along with the Network Dataset on the Toposheet
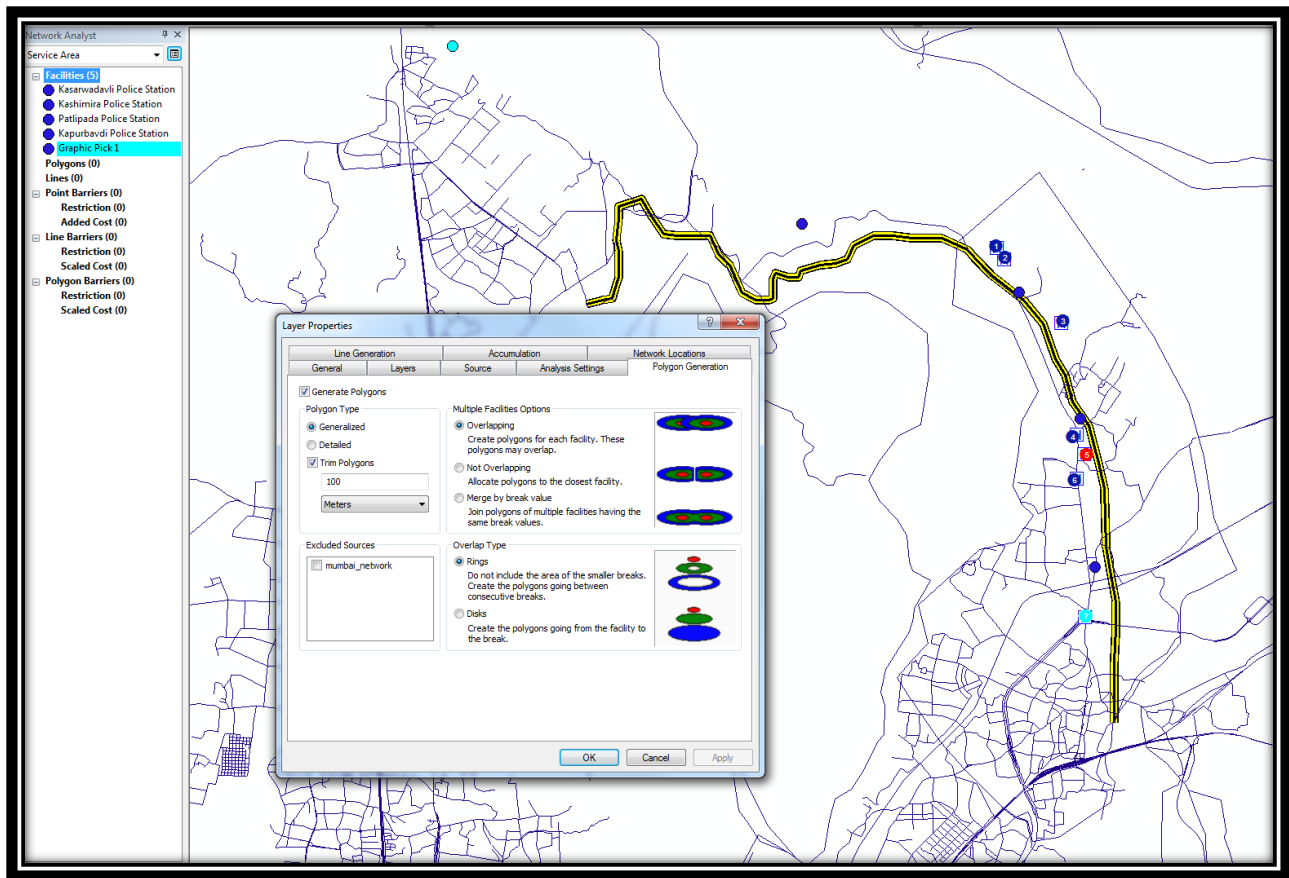
Figure 51 Parameters selected for Service Area generation
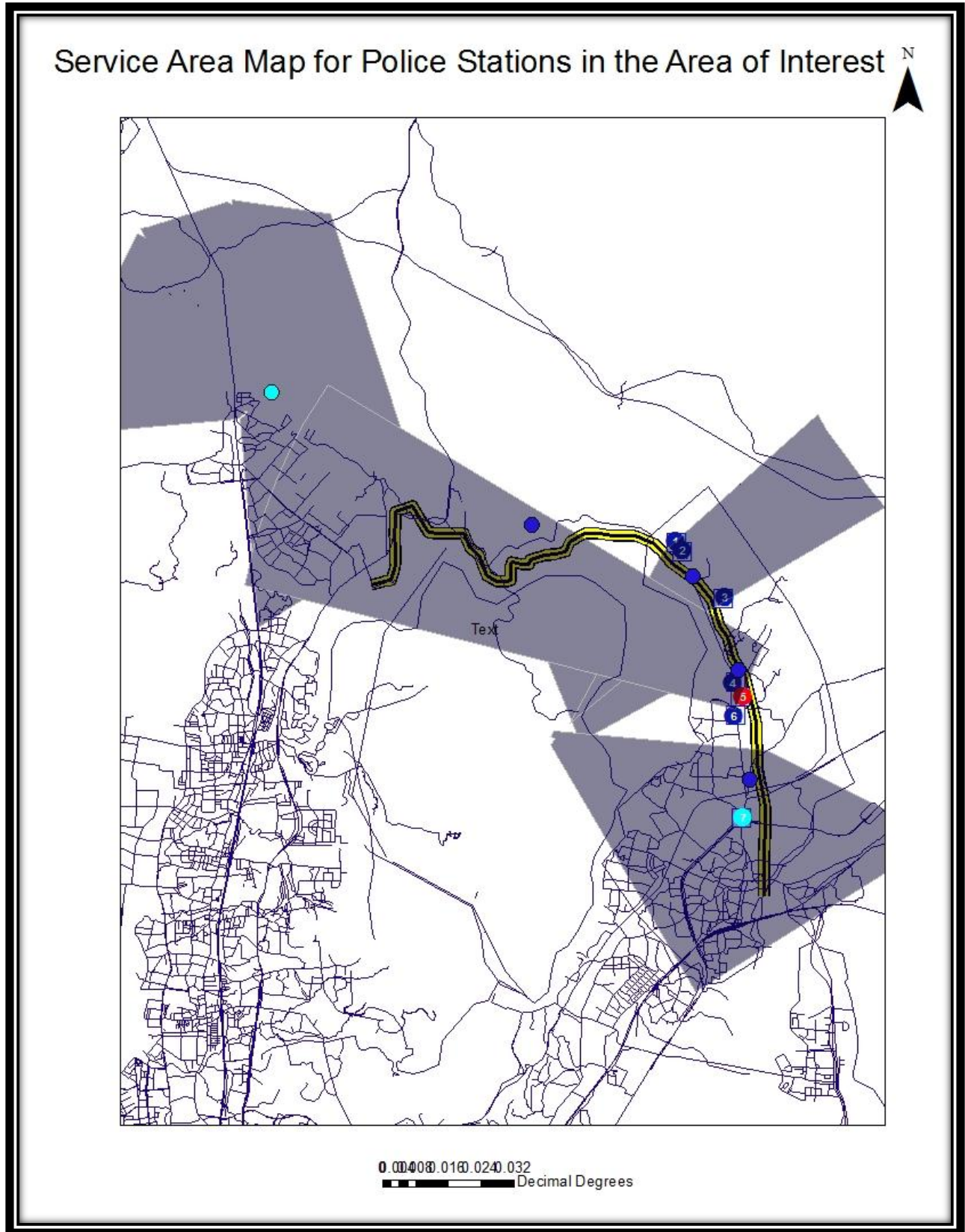
Figure 52 Service Area Map for Police Stations

11.3 Geostatistical Analysis

Any measurement we take in Earth and environmental sciences has a spatio-temporal reference. This reference is determined by four main parameters:

➢ Geographic location

➢ Elevation

➢ Time of measurement

➢ Spatio-temporal support

Spatial-temporal data analysis is a combination of geoinformation science and spatio-temporal statistics. As implied, geostatistics is a subset of statistics specialized in analysis and interpretation of geographically referenced data.
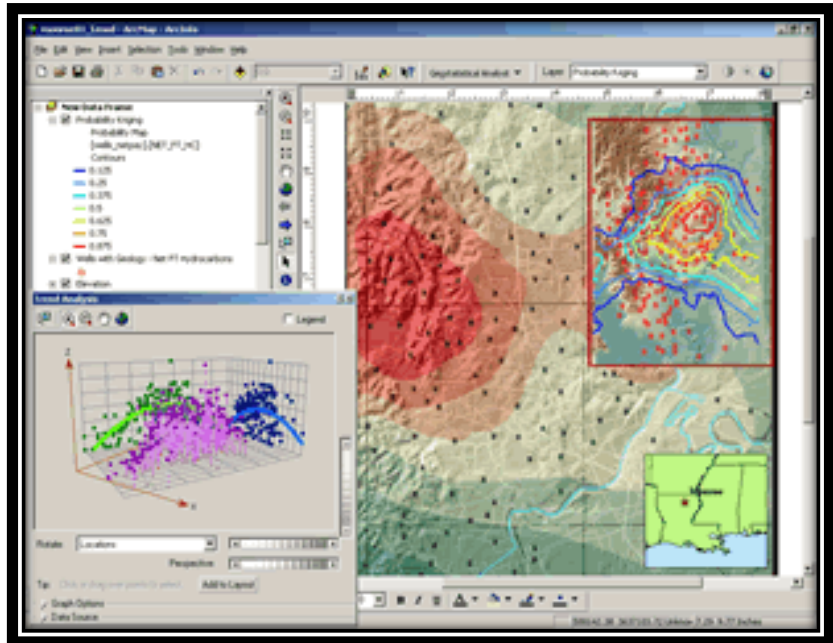
Figure 53 Geostatistical Analysis on ArcGIS

The main function of this geostatistics is that it determines the probability of certain variables occurring over an area where identifying every possible location would be impossible. It provides a multitude of powerful interpolation methods with advanced analytical tools for generating interpolated surfaces from discrete spatial data measurements. These surfaces represent a statistical estimation or prediction of where a certain phenomenon may occur.

This provides two groups of interpolation techniques: deterministic and Geostatistical. Deterministic techniques use mathematical functions for interpolation. Geostatistics relies on both statistical and mathematical methods, which can be used to create surfaces and assess the uncertainty of the predictions.

11.3.1 Geostatistical Workflow

```
Map and examine the data  →  Pre-process data, if necessary (transform, de-trend, de-cluster)
```

Map and examine the data

Pre-process data, if necessary (transform, de-trend, de-cluster)

Model spatial structure

Define search strategy

Predict values at unsampled locations

Quantify uncertainty of the predictions

Check for reasonable results produced by model for prediction and uncertainty

NO

YES

Use of information in decision making

11.3.2 Kriging

An interpolation technique where the surrounding values are measured and weighted to derive a predicted value for an unmeasured location. Weights are based on the distance between the measured points, the prediction locations, and the overall spatial arrangement among the measured points. Kriging is unique among the interpolation methods in that it provides an easy method for characterizing the variance, or the precision, of predictions. Kriging is based on regionalized variable theory, which assumes that the spatial variation in the data being modelled is homogeneous across the surface. That is, the same pattern of variation can be observed at all locations on the surface. The theory behind interpolating or extrapolating by kriging was developed by the French mathematician **Georges Matheron** based on the Master's thesis of **Danie G. Krige,** the pioneering plotter of distance-weighted average gold grades at the Witwatersrand reef complex in South Africa.

Kriging belongs to the family of linear least squares estimation algorithms. The aim of kriging is to estimate the value of an unknown real-valued function, f, at a point, x^*, given the values of the function at some other points, $x_1, \cdots, x_n$. A kriging estimator is said to be linear because the predicted value $\hat{f}(x^*)$ is a linear combination that may be written as:

$$\hat{f}(x^*) = \sum_{i=1}^{n} \lambda_i(x^*) f(x_i) \dots\dots\dots\dots\dots\dots\dots\dots\dots\text{ Equation 27}$$

Depending on the stochastic properties of the random fields, different types of kriging apply. The type of kriging determines the linear constraints on the weights $\dot{\omega}_i$ and hence the method for calculating the weights, depend upon the type of kriging. The classical types of kriging are:

- Ordinary
- Simple
- Universal
- Indicator
- Disjunctive
- Probability

Ordinary kriging is the most commonly used type. It assumes a constant but unknown mean.
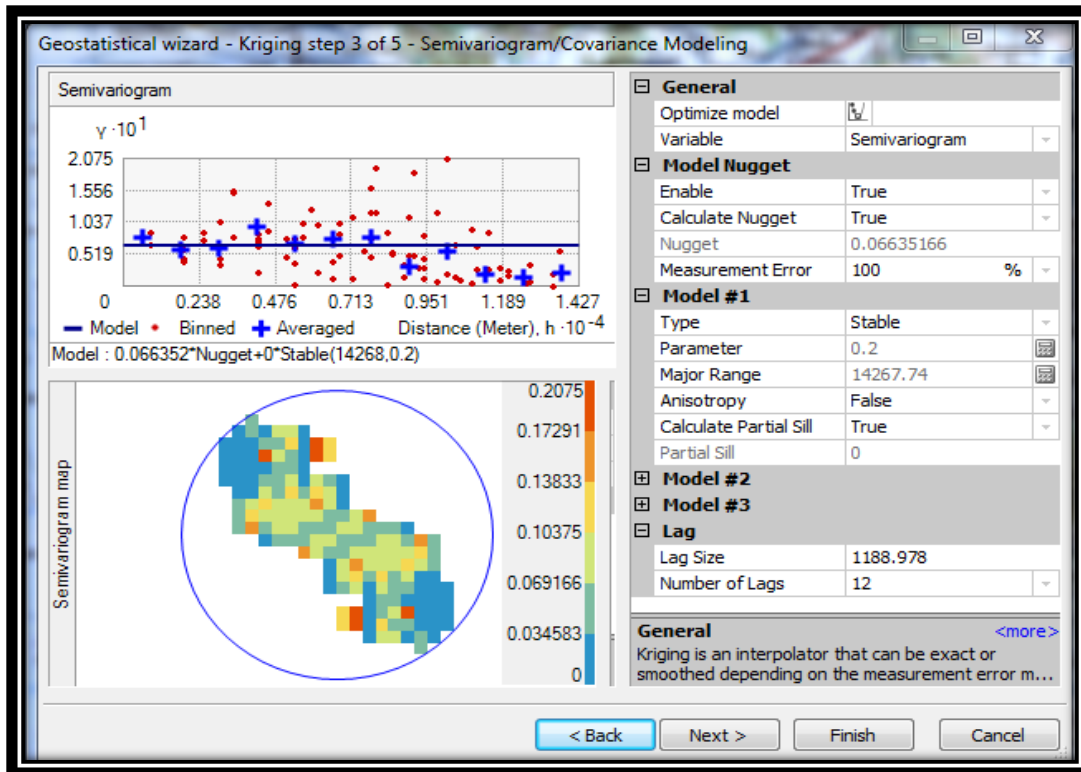
Figure 54 Semivariogram for Accident Sites for 2007-2009
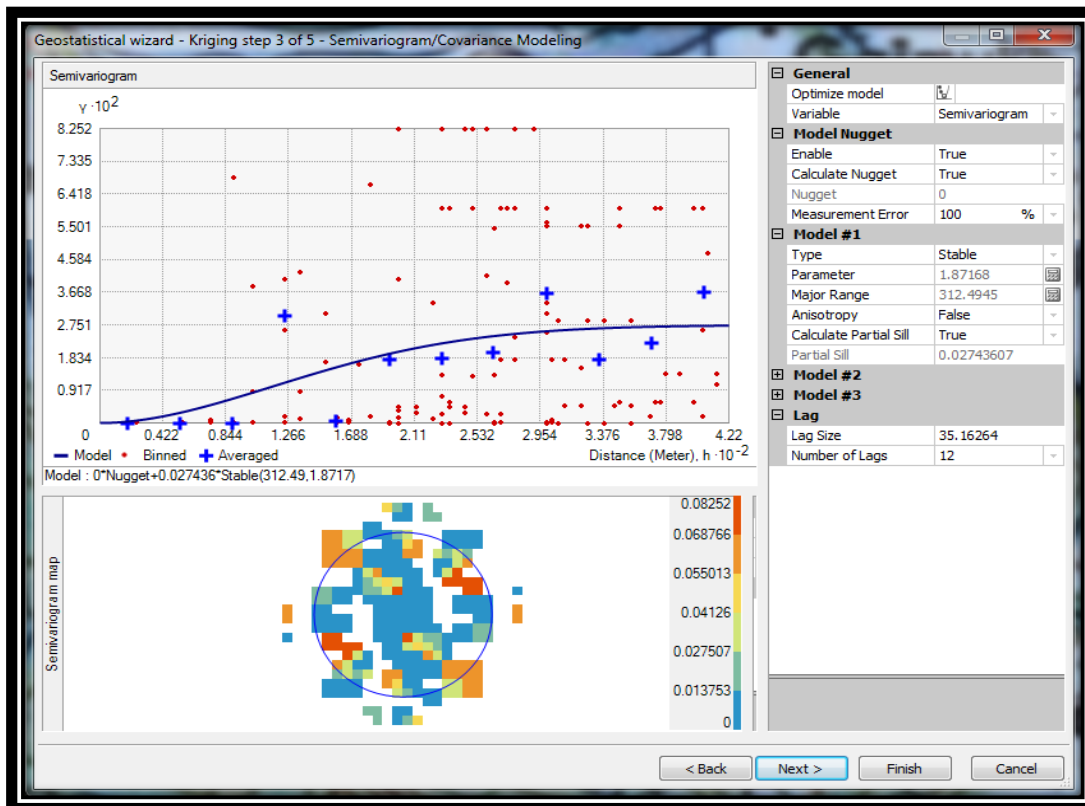
Figure 55 Semivariogram for Accident Sites for 2009-2011

The above figure displays the semivariogram/covariance model which allows us to examine the spatial relationships between the measured points (in this case, it's the Accident Sites) for 2007-2009 and 2009-2011. Things that are closer together are more alike than things that are farther apart.

A semivariogram gives us information about the nature and structure of the spatial dependency in a random field which must be estimated from the data. Estimating a semivariogram helps us in deriving an empirical estimate about the data. A traditional semivariogram plots one-half of the squares of the differences between samples versus their distance from one another.
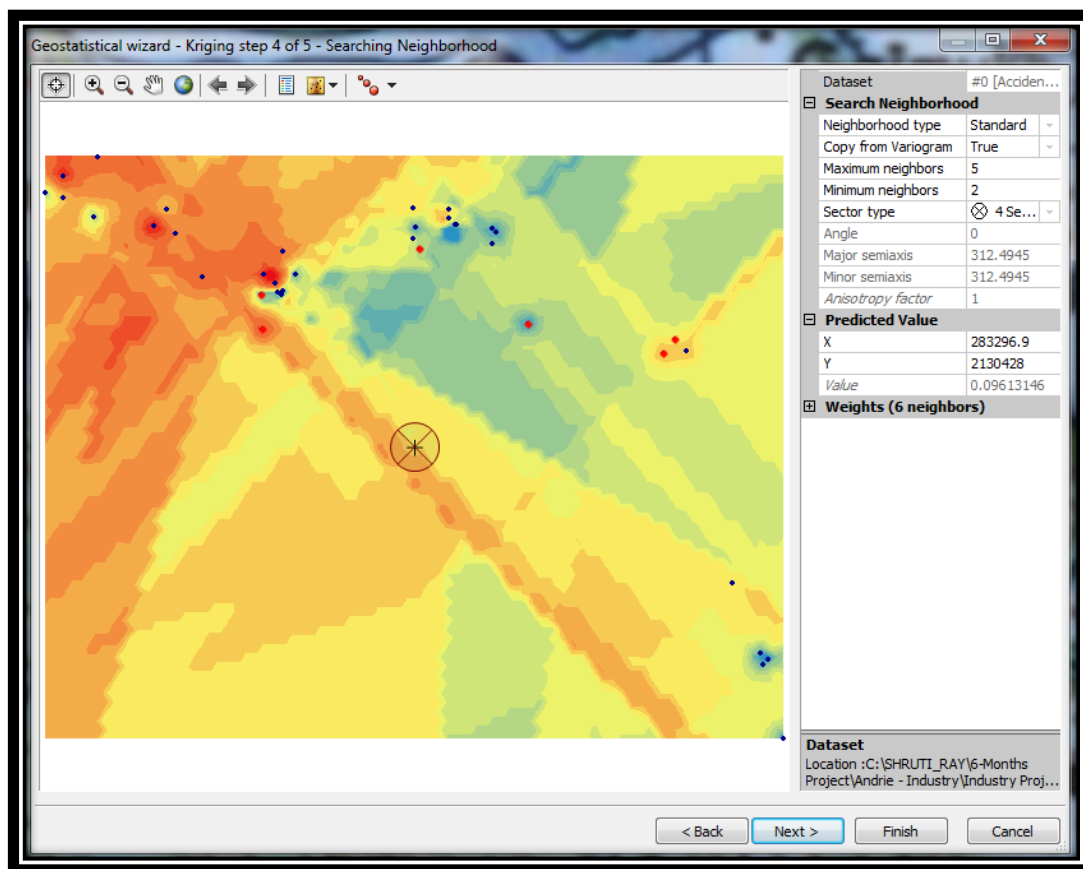
Figure 56 Cross Validation Diagram for Accident Sites (2009-2011)

The crosshair shows a location that has no measured value. We can predict a value by using the values of the measured locations. Using the surrounding points and the semivariogram or covariance model fitted previously, you can predict values for the unmeasured location. The cross-validation diagram gives you an idea of how well the model predicts the values at the unknown locations.

Figure 57 Prediction Graph for Cross-Validation

Table 5 Cross-Validation table showing Error
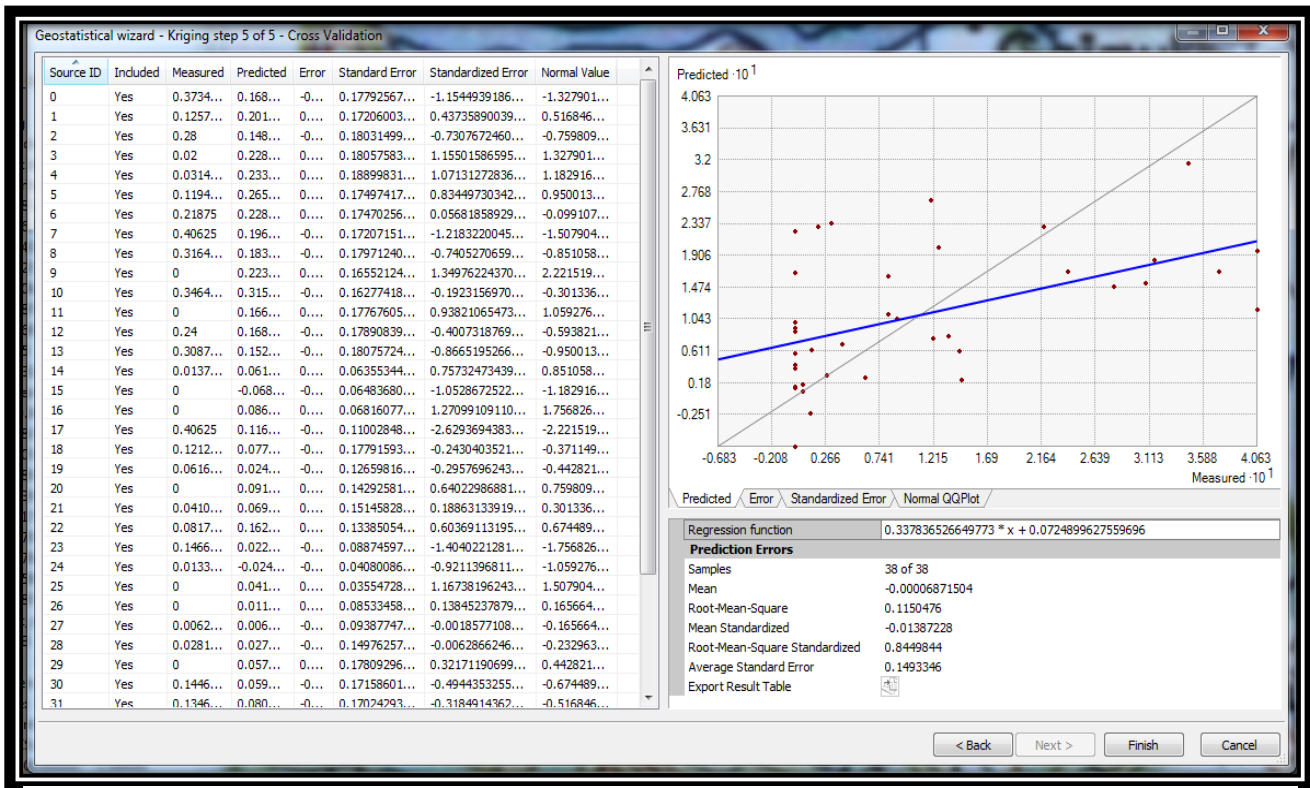
Figure 58 Prediction Graph with Sample Size for 2009-2011
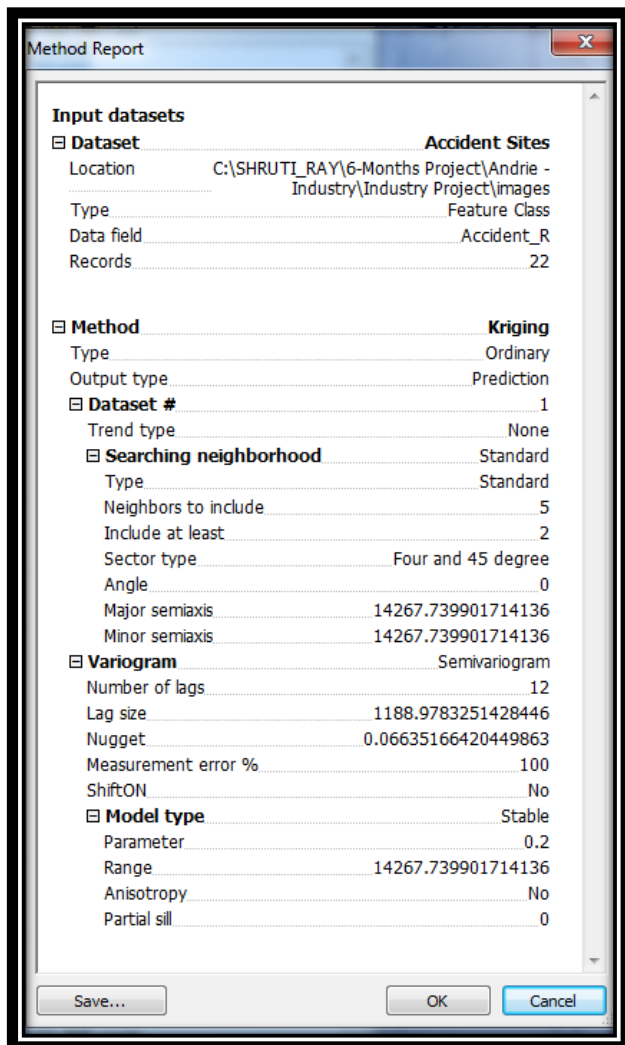
This is the Report that is generated after the steps involved in process of kriging get completed. It includes the Input Dataset which has the Dataset name as well as its Location, Type, Data Field and the Total number of Records that it contains.

It also specifies the method used (kriging in this case) along with its type as well as the output type. The Variogram along with the lag size and the nugget size is also mention. The important thing is that, it also contains the percentage of error.

This is a report generated for Accident Sites for 2007-2009. A similar report was generated for Accident Sites 2009-2011.

Figure 59 Report Generated for Accident Sites

• • •

114

11.3.3 Histogram

The Histogram tool provides a univariate description of the data. The tool displays the frequency distribution for the dataset of interest and calculates summary statistics.

The frequency distribution is a bar graph that displays how often observed values fall within certain intervals or classes. The number of classes of equal width that is used in the histogram is specified. The relative proportion of data that falls in each class is represented by the height of each bar.

The important features of a distribution can be summarized by statistics that describes its location, spread and shape.

11.3.3.1 Measures of Location

Measures of location provide us with an idea of where the centre and the other parts of the distribution lie. The mean is the arithmetic average of the data and provides a measure of the centre of the distribution. The median corresponds to a cumulative proportion of 0.5.

11.3.3.2 Measures of Spread

The spread of points around the mean value is another characteristic feature of the frequency distribution. The variance of the data is the average squared deviation of all values from the mean.

The standard deviation is the square root of the variance. It describes the spread of the data about the mean in the same units as its original measurements. The smaller the variance and standard deviation, the tighter is the cluster of measurements about the mean value.

11.3.3.3 Measures of Shape

The coefficient of skewness is a measure of the symmetry of a distribution. For symmetric distributions, the coefficient of skewness is zero. If a distribution has a long right tail of large values, it is positively skewed, and if it has a long left tail of small values, it is negatively skewed. The mean is larger than the median for positively skewed distributions and vice versa for negatively skewed distributions. The kurtosis is based on the size of the tails of a distribution and provides a measure of how likely the distribution will produce outliers. The kurtosis of a normal distribution is three.
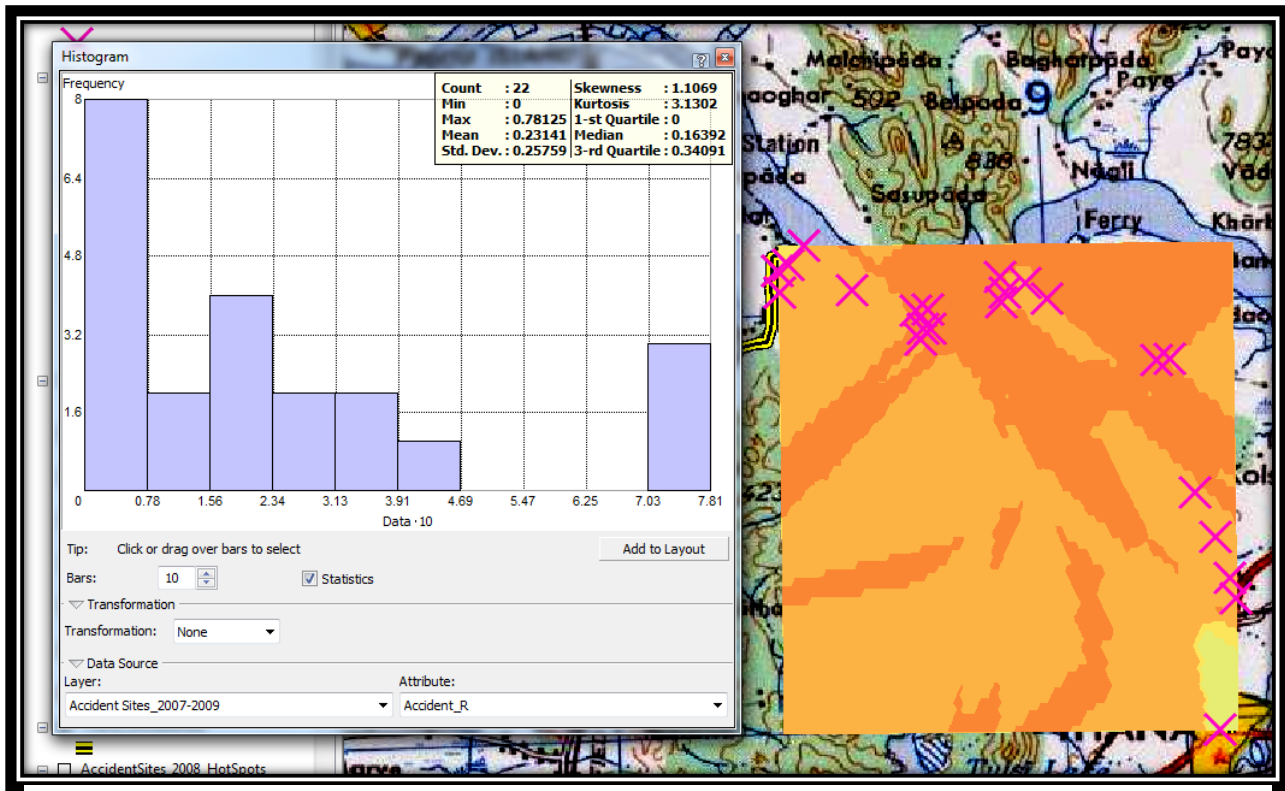
Figure 60 Histogram showing Mean, Median and Standard Deviation (2007-2009)
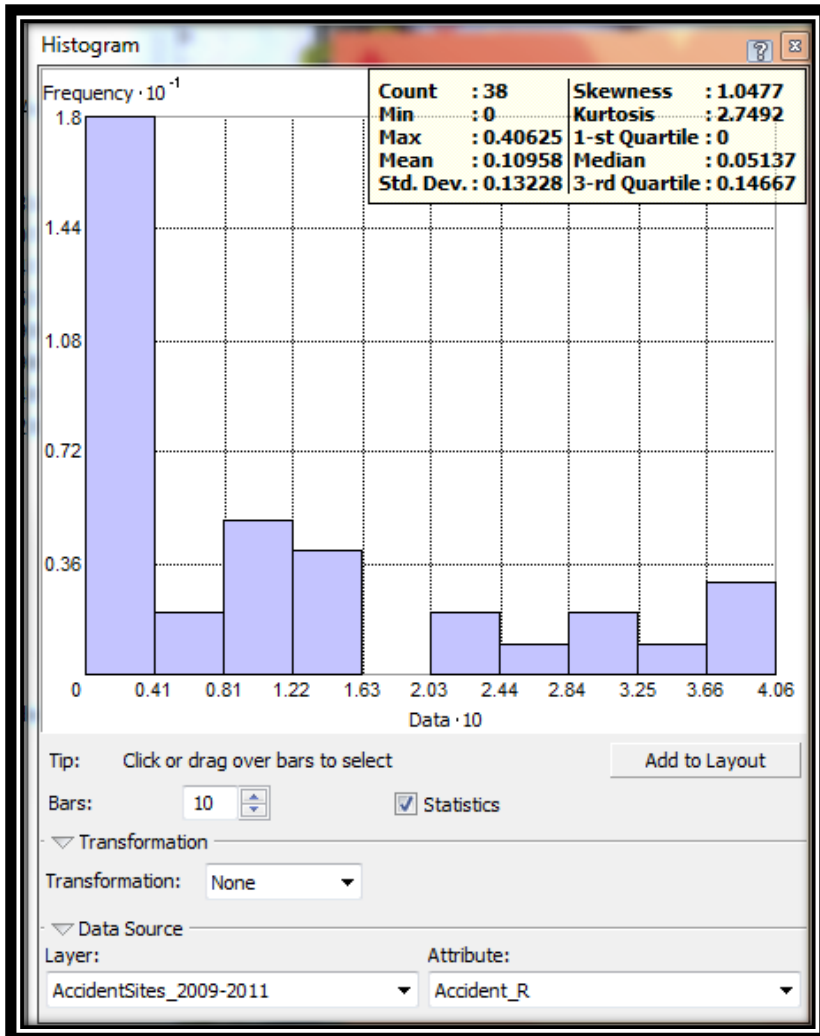
This distribution has been depicted in a histogram with a range of values split in 10 classes. The frequency of data within each class is represented by the height of each bar. Generally, the important features of a distribution are its central value, spread, and symmetry. If the mean and the median are approximately the same value, it implies that the distribution is a normal distribution.

Figure 61 Histogram showing Mean, Median and Standard Deviation (2009-2011)

11.3.4 Q-Q Plot

Q–Q plot ("Q" stands for quantile) is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. First, the set of intervals for the quantiles are chosen. A point (x, y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate). Thus the line is a parametric curve with the parameter which is the (number of the) interval for the quantile.

If the two distributions being compared are similar, the points in the Q–Q plot will approximately lie on the line y = x. If the distributions are linearly related, the points in the Q–Q plot will approximately lie on a line, but not necessarily on the line y = x. Q–Q plots can also be used as a graphical means of estimating parameters in a location-scale family of distributions.

A Q–Q plot is used to compare the shapes of distributions, providing a graphical view of how properties such as location, scale, and skewness are similar or different in the two distributions. Q–Q plots can be used to compare collections of data, or theoretical distributions. The use of Q–Q plots to compare two samples of data can be viewed as a non-parametric approach to comparing their underlying distributions.

A general QQ plot is a graph on which the quantiles from two distributions are plotted versus each other. For two identical distributions, the QQ plot will be a straight line. Therefore, it is possible to check the normality of the accident occurrence by plotting the quantiles of that data versus the quantiles of a standard normal distribution. From the normal QQ plot above, we can see that the plot is not very close to being a straight line. The main departure from this line occurs at low values.

A Q–Q plot is a plot of the quantiles of two distributions against each other, or a plot based on estimates of the quantiles. The pattern of points in the plot is used to compare the two distributions. The main step in constructing a Q–Q plot is calculating or estimating the quantiles to be plotted.

If the data does not exhibit a normal distribution in either the histogram or normal QQ plot, it may be necessary to transform the data to make it conform to a normal distribution before using certain kriging interpolation techniques.
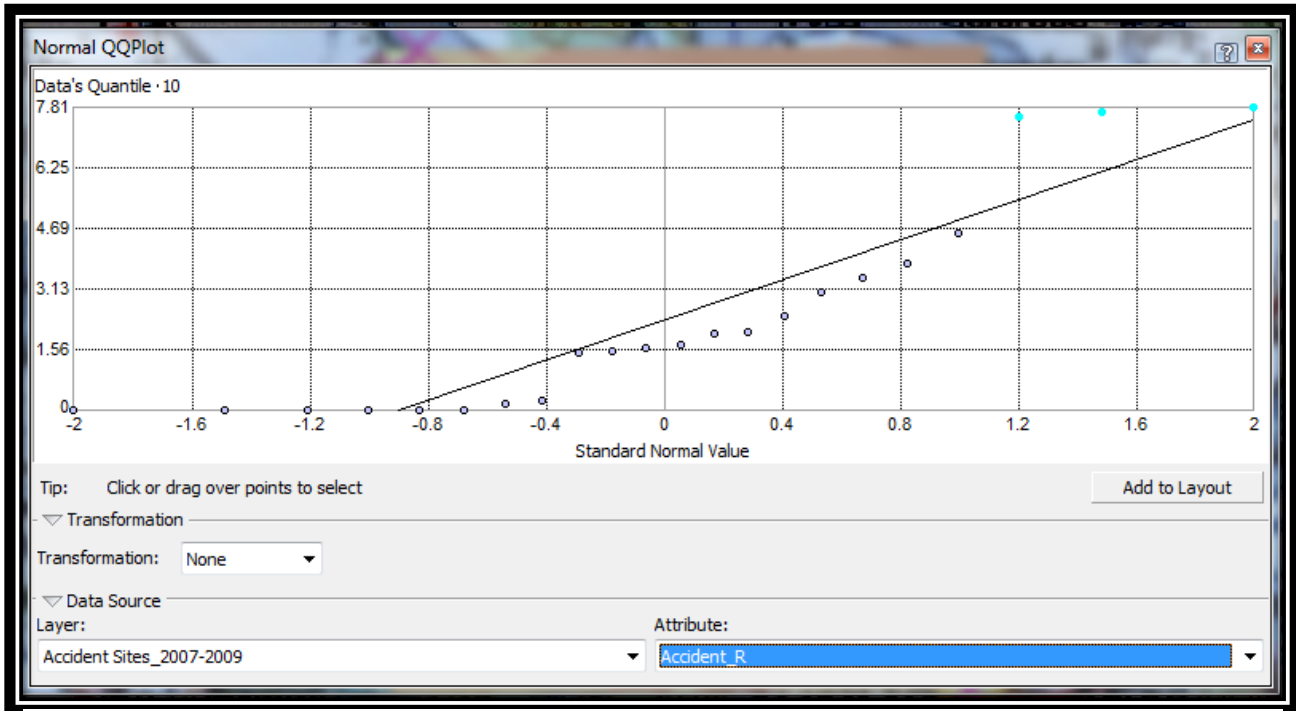
Figure 63 Normal QQ Plot for 2007-2009
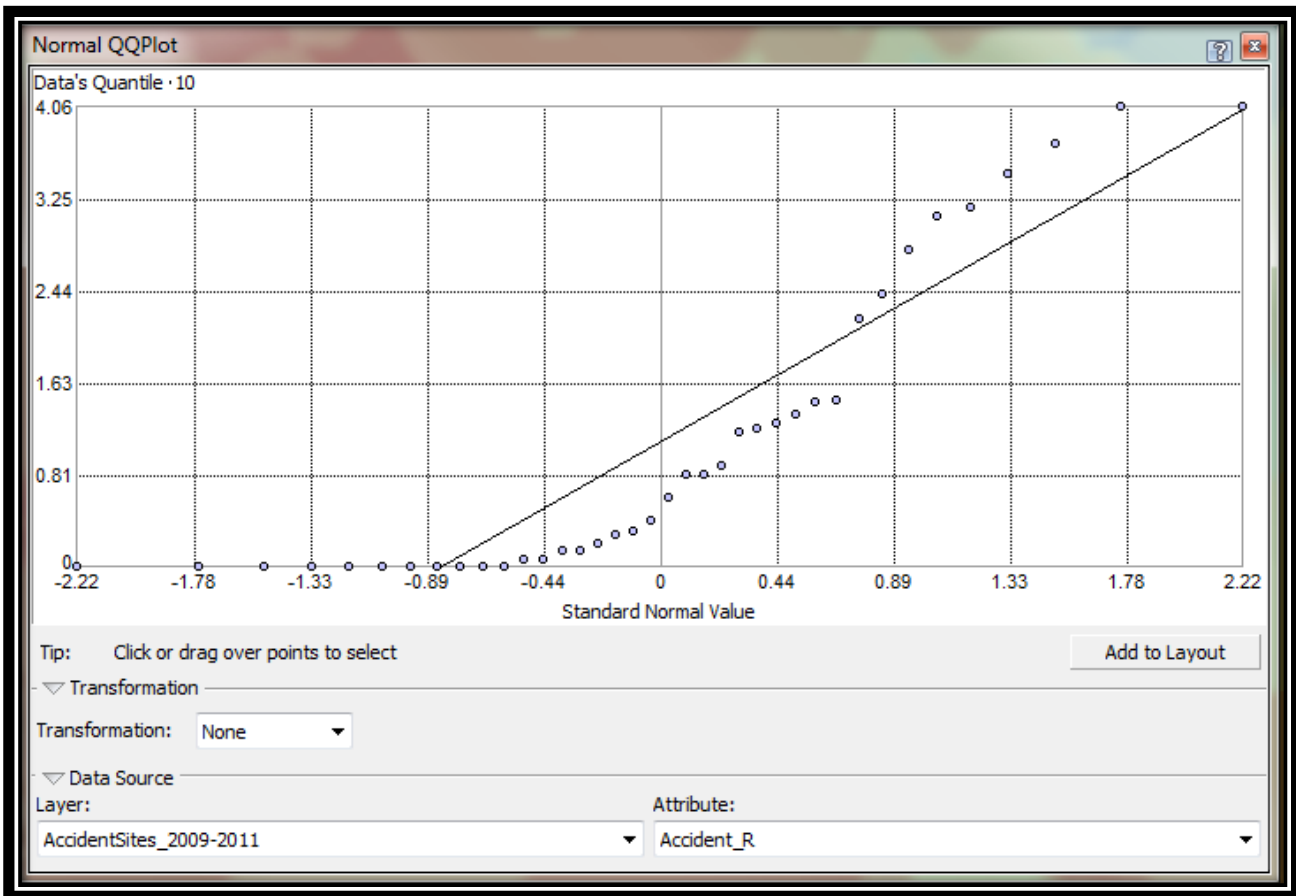
Figure 62 Normal QQ Plot for 2009-2011

11.3.5 Trend Analysis

If a trend exists in the data, it is a nonrandom (deterministic) component of a surface that can be represented by a mathematical formula. For instance, a gently sloping hillside can be represented by a plane. A valley would be represented by a more complex formula (a second-order polynomial) that creates a U shape. The Trend Analysis tool enables us to identify the presence/absence of trends in the input dataset and identify which order of polynomial fits the trend best.

Each vertical stick in the trend analysis plot represents the location and value (height) of each accident sites. The data points are projected onto the perpendicular planes, an east–west and a north–south plane. A best-fit line (a polynomial) is drawn through the projected points, showing trends in specific directions. If the line were flat, this would indicate that there is no trend.
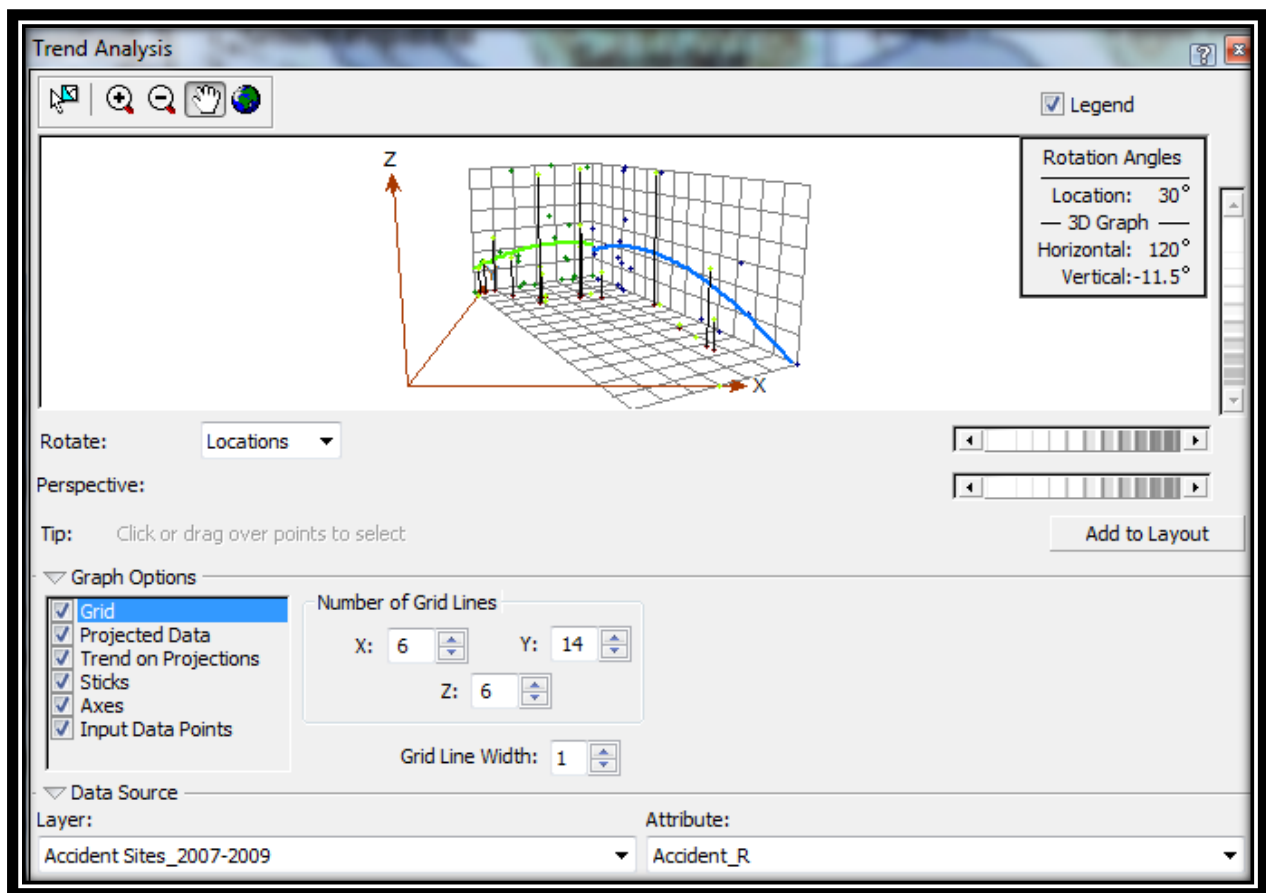
Figure 64 Trend Analysis for 2007-2009 based on Accident Rates

In the above figure, the green line starts from low values, and increases as it moves towards the centre of the X-axis, and then decreases slightly. Similarly, the blue line increases as it moves north and decreases gradually even before it reaches the centre. This demonstrates that the data seems to exhibit more or less a strong trend from the centre of the data domain in almost all directions.
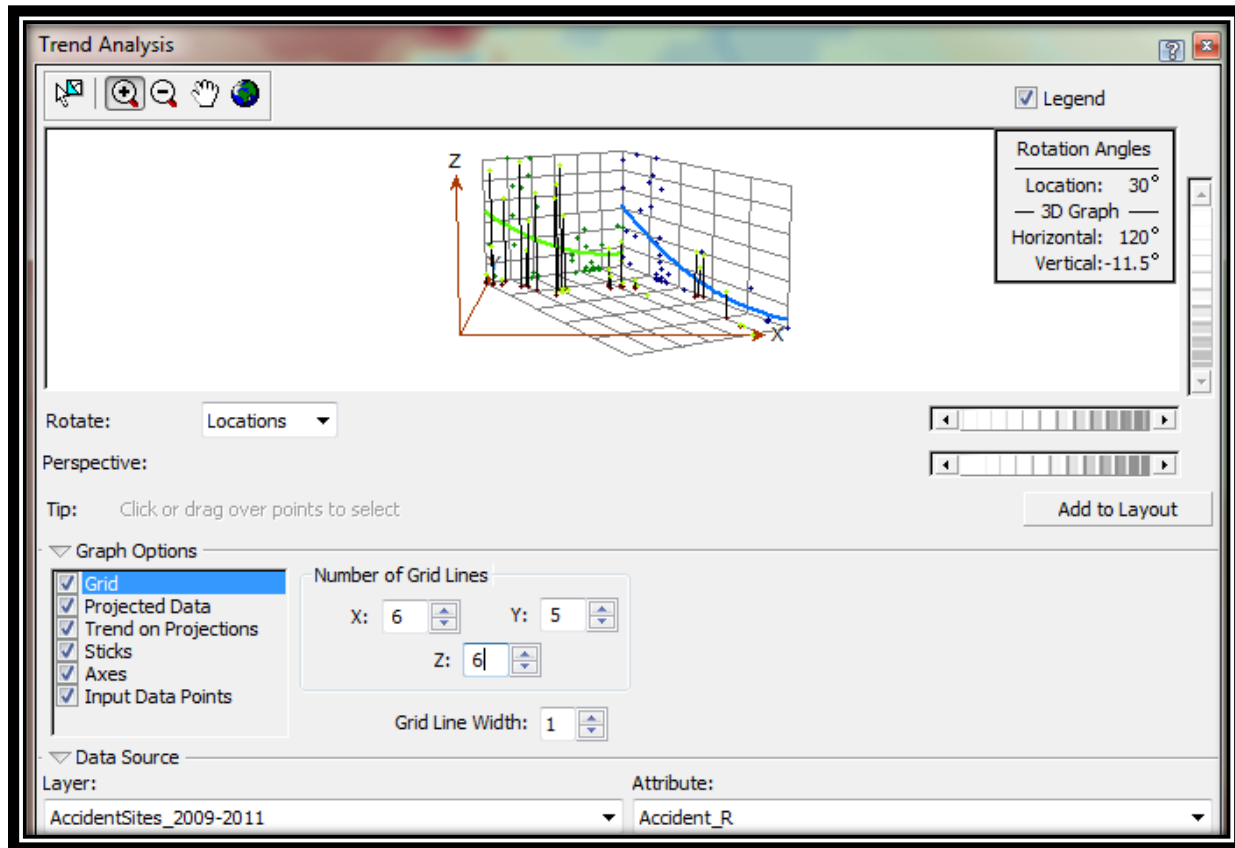
Figure 65 Trend Analysis for 2009-2011 based on Accident Rates

Similarly, in the above figure the green line starts from a high value and increases at a gradual rate as we move towards the centre and even further it decreases. Throughout, the green curve shows a decline till the end. The blue line also follows the trend similar to the green line. It starts at a higher value and as we move further down it decreases and almost touches the base. This reveals that the trend for this data (Accident Sites for 2009-2011) is not that strong as compared to the previous data (Accident Sites for 2007-2009). These points can be adjusted by rotating the locations and the trend that is demonstrated is stronger.

11.3.6 Voronoi Map

In mathematics, a Voronoi diagram is a special kind of decomposition of a metric space determined by distances to a specified discrete set of objects in the space. It is named after Georgy Voronoi.

In the simplest case, we are given a set of points S in the plane, which are the Voronoi sites. Each site s has a Voronoi cell, also called a Dirichlet cell, V(s) consisting of all points closer to s than to any other site. The segments of the Voronoi diagram are all the points in the plane that are equidistant to the two nearest sites. The Voronoi nodes are the points equidistant to three (or more) sites.

Voronoi maps are constructed from a series of polygons formed around the location of a sample point. Voronoi polygons are created so that every location within a polygon is closer to the sample point in that polygon than any other sample point. After the polygons are created, neighbours of a sample point are defined as any other sample point whose polygon shares a border with the chosen sample point.

The dual graph for a Voronoi Map corresponds to the Delaunay triangulation for the same set of points. The closest pair of points matches to the two adjacent cells in the Voronoi Map. Two points are adjacent on the convex hull if and only if their Voronoi cells share an infinitely long side. The Voronoi Map tool provides a number of methods for assigning or calculating values to polygons. These are:

- **Simple**: The value assigned to a cell is the value recorded at the sample point within that cell.
- **Mean**: The value assigned to a cell is the mean value that is calculated from the cell and its neighbors
- **Mode**: All cells are placed into five class intervals. The value assigned to a cell is the mode of the cell and its neighbors.
- **Cluster**: All cells are placed into five class intervals. If the class interval of a cell is different from each of its neighbors, the cell is colored gray to distinguish it from its neighbors.

- **Entropy**: All cells are placed into five classes based on a natural grouping of data values. The value assigned to a cell is the entropy that is calculated from the cell and its neighbors

$$Entropy = -\sum (p_i * \log p_i) \dots\dots\dots\dots\dots\dots\dots\dots. \text{ Equation 28}$$

Where p_i is the proportion of cells are assigned to each class.
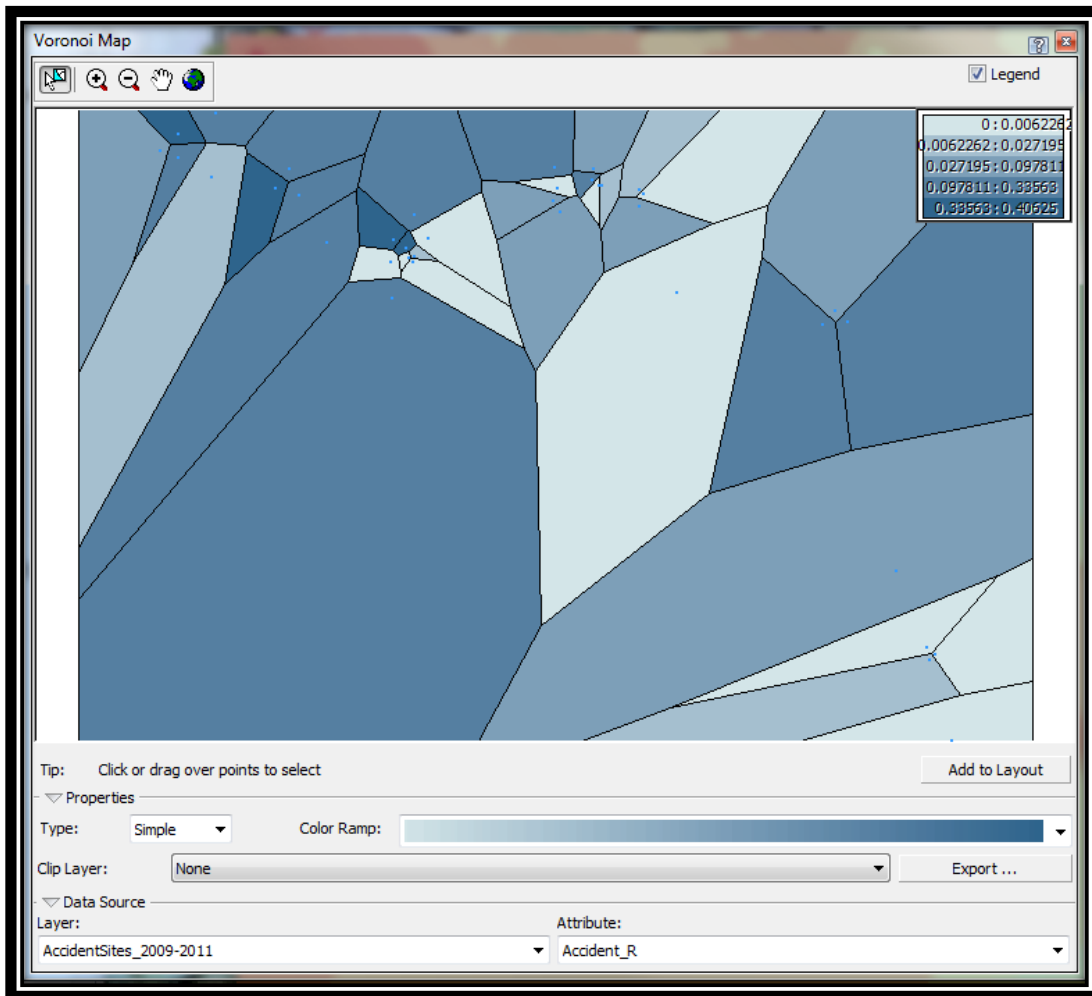
Figure 66 Voronoi Map for Accident Sites for 2009-2011

- **Median:** The value assigned to a cell is the median value calculated from the frequency distribution of the cell and its neighbors.

- **Standard Deviation:** The value assigned to a cell is the standard deviation that is calculated from the cell and its neighbors.

- **Inter-quartile Range:** The 1st and 3rd quartiles are calculated from the frequency distribution of a cell and its neighbors.

● ● ●

122

11.3.7 Semivariogram

The okapi variogram 2γ is a function describing the degree of spatial dependencies of a spatial random field or stochastic process $Z(x)$. It is defined as the expected squared increment of the values between locations *x* and *y* (Wackernagel 2003). It is given by:

$$2\gamma(x, y) = E\ (|Z(x) - Z(y)|^2)\ \dots\dots\dots\dots\dots\dots\dots\dots\text{Equation 29}$$

$\gamma\ (x,\ y)$ itself is called the semivariogram.

Semivariogram is non-negative as it is the expectation of a square. As a consequence, the semivariogram might be non-continuous only at the origin. The height of the jump at the origin is sometimes referred to as *nugget* or *nugget effect*. Limit of the variogram tending to infinity lag distances are termed as *sill*. The distance in which the difference of the variogram from the sill gets negligible is referred as *range.*

The variogram is the key function in geostatistics as it is used to fit a model of the temporal/spatial correlation of the observed phenomenon. A distinction can be made between the experimental variogram that is visualization of the possible spatial/temporal correlation and the variogram model is further used to define the weights of the kriging function. Experimental variogram is an empirical estimate of the covariance of a Gaussian process.

When a variogram is used to describe the correlation of different variables it is called cross-variogram. Cross-variograms are used in co-kriging. In case there are any binary values or values that represent classes, then indicator variograms are used. Indicator kriging uses indicator variogram.

In mathematical statistics, a set of n measured values gives $df = (n\text{-}1)$ Equation 30 degrees of freedom whereas the in situ or temporarily ordered set give $df\ (o) = 2(n\text{-}1)$ …….….. Equation 31 the degrees for the first variance term. It is claimed that the variogram and semivariogram are both invalid measures for variability, precision and risk because the sum of squared differences between *x* and *x+h* is divided by n, the number of data in the set rather than by $\mathbf{df\ (o) = 2(n\text{-}1)}$ …….….. Equation 32 the degrees of freedom for the first variance term of the ordered set.
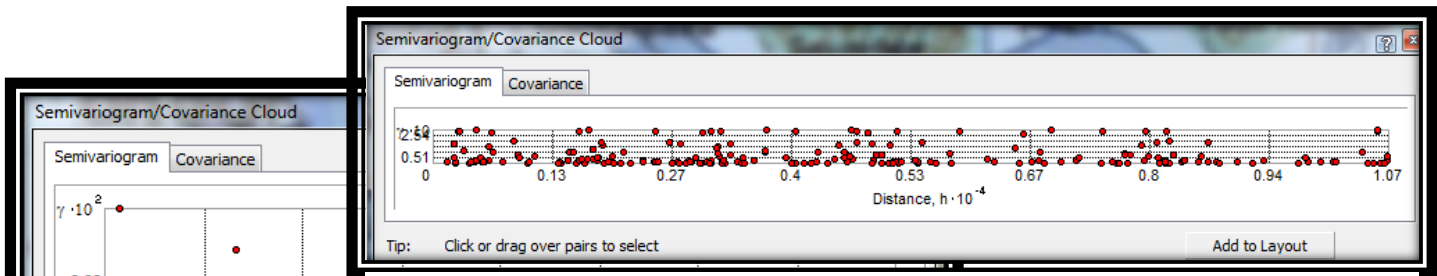
Figure 67 Semivariogram for Accident Sites for 2007-2009
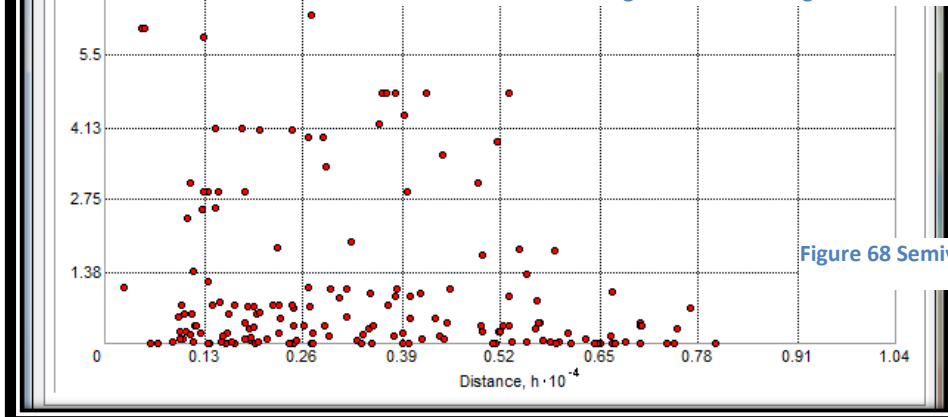
Figure 68 Semivariogram for Accident Sites for 2009-2011
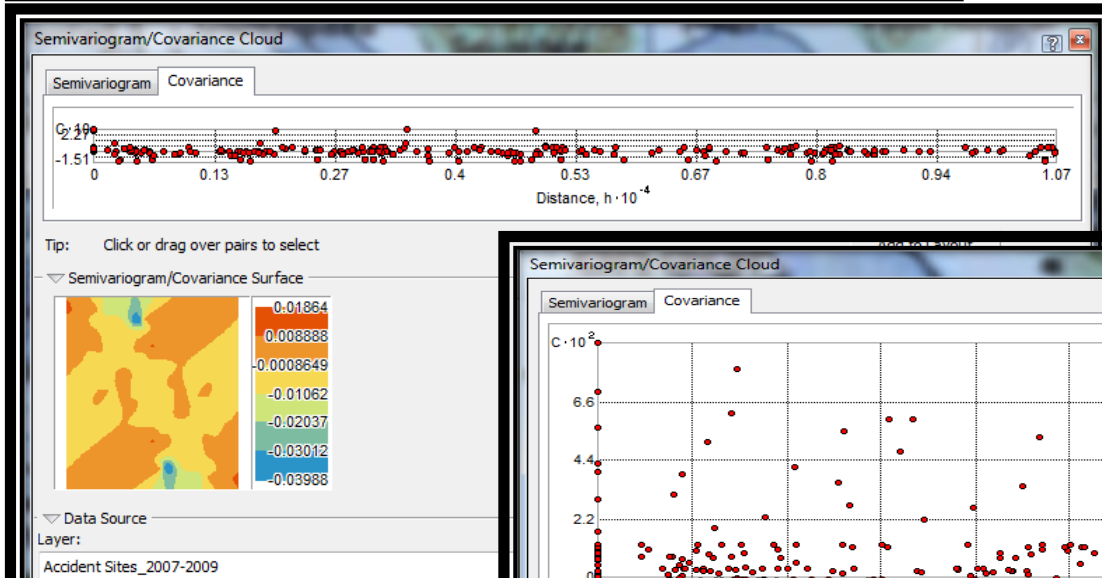
Figure 69 Covariance for Accident Sites (2007-2009)
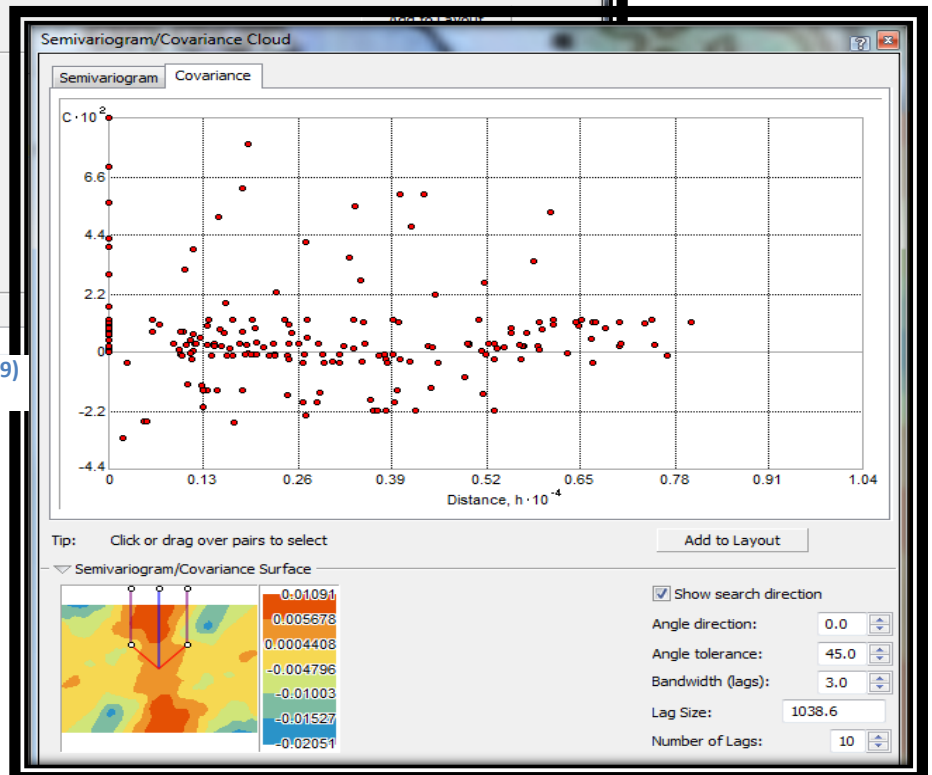
Figure 70 Covariance for Accident Site (2009-2011)

11.3.8 Cross-Covariance Cloud

The statistical tendency of variables of different types, attributes, names, and so on, to vary in ways that are related to each other. Positive cross covariance occurs when both variables tend to be above their respective means together, and negative cross covariance occurs if one variable tends to be above its mean when the other variable is below its mean.

The term cross-covariance is sometimes used to refer to the covariance *cov(X, Y)* between two random vectors X and Y, in order to distinguish that concept from the "covariance" of a random vector X, which is understood to be the matrix of co-variances between the scalar components of X.
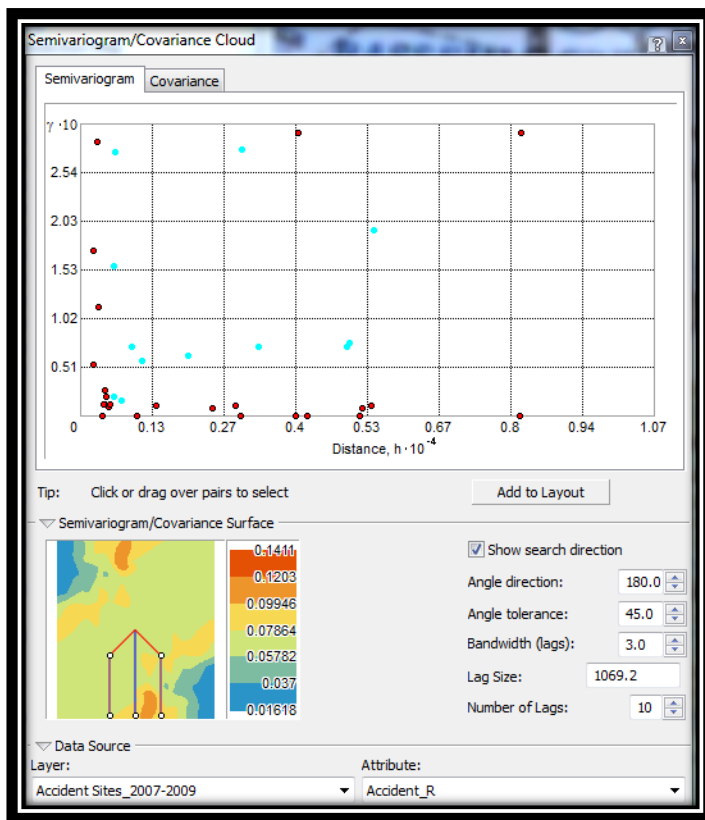
Figure 72 Covariance Cloud of Accident Sites for 2007-2009
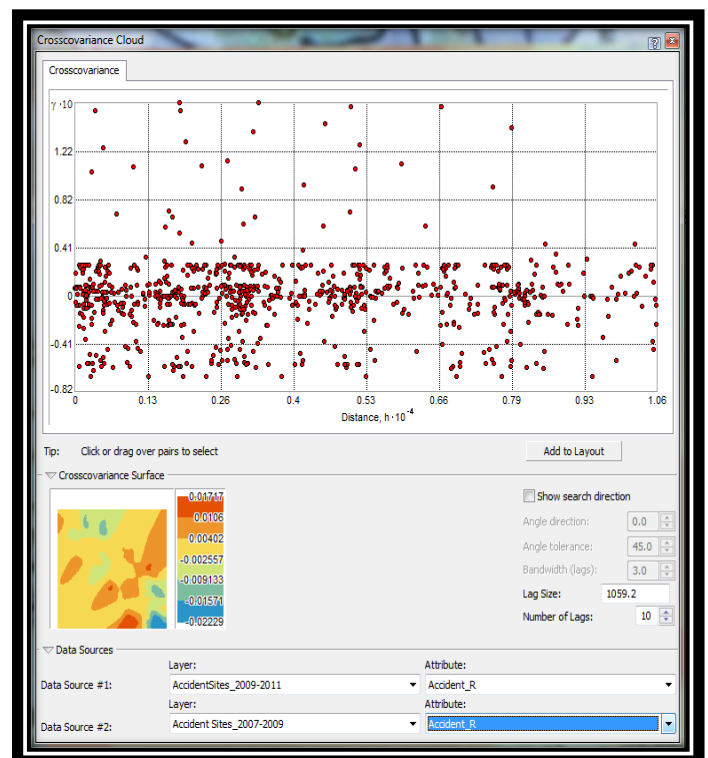
Figure 71 Covariance Cloud of Accident Sites for 2009-2011

● ● ●

125

The semivariogram/covariance cloud allows you to examine the spatial autocorrelation between the measured sample points. It is generally assumed that things that are close to one another are more alike. The semivariogram/covariance cloud lets you examine this relationship. A semivariogram value, which is the difference squared between the values of each pair of locations, is plotted on the y-axis relative to the distance separating each pair of measurements, which is plotted on the x-axis.

Each red dot in the semivariogram/covariance cloud represents a pair of locations. Since locations that are close to each other should be more alike. In the semivariogram plot the locations that are closest (on the far left on the x-axis) should have small semivariogram values (low values on the y-axis). As the distance between the pairs of locations increases (moving right on the x-axis), the semivariogram values should also increase (move up on the y-axis). However, a certain distance is reached where the cloud flattens out, indicating that the values of the pairs of points separated by more than this distance are no longer correlated.

Looking at the semivariogram, if it appears that some data locations that are close together (near zero on the x-axis) have a higher semivariogram value (high on the y-axis) than you would expect, you should investigate these pairs of locations to see if there is a possibility that the data is inaccurate.

For the covariance cloud for Accident Sites 2007-2009, we can find pair of sample locations that are selected in the semivariogram are highlighted in the light blue colour. The lines with high semivariogram values for a particular distance between the points in a pair are those that correspond to the largest gradient in the ozone values.

The semivariogram/covariance cloud illustrated that the unusually high semivariogram values are largely represented by the lines perpendicular to the coast. The analysis using this tool indicates that the interpolation model should account for anisotropy. The semivariogram surface indicates there is spatial autocorrelation in the data.

11.3.9 Semivariogram / Covariance Modeling

Semivariogram / Covariance cloud explores the overall spatial autocorrelation of the measured points. The goal of semivariogram/covariance modeling is to determine the best fit for a model that will pass through the points in the semivariogram. The semivariogram is a graphic representation used to provide a picture of the spatial correlation in the dataset.

The Semivariogram/Covariance Modeling dialog box allows you to fit models to the spatial relationships in the dataset. The lag size is the size of a distance class into which pairs of locations are grouped to reduce the large number of possible combination. In this process of modeling, our goal is to fit a curve through all the points those were earlier obtained. The empirical semivariogram values (red points) are grouped according to the separation distance they are associated with. The points are split into bins (or lags), and the lag size determines how wide each interval (bin) will be. This process is known as binning.
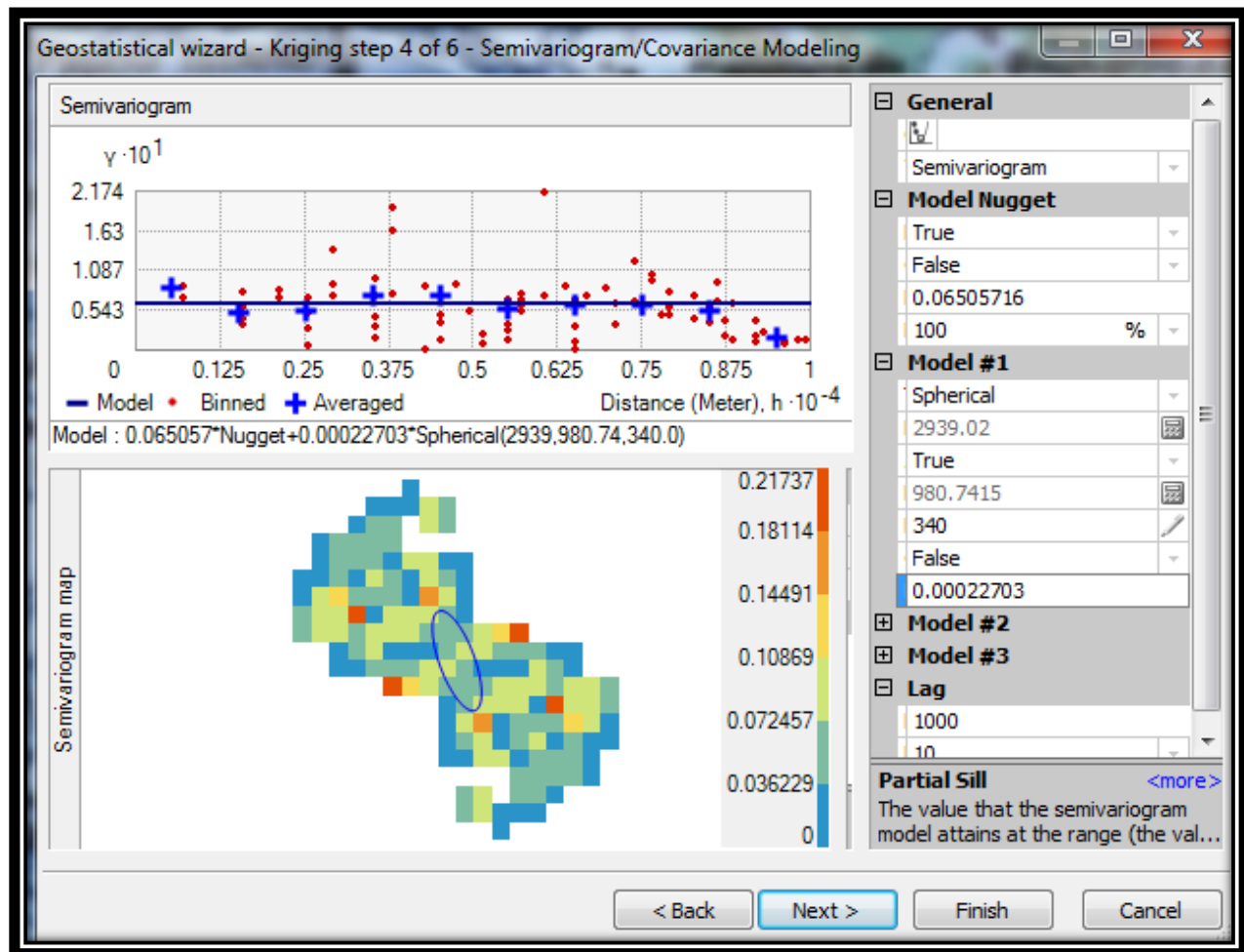
Figure 73 Semivariogram/ Covariance Modeling of Accident Sites (2007-2009)
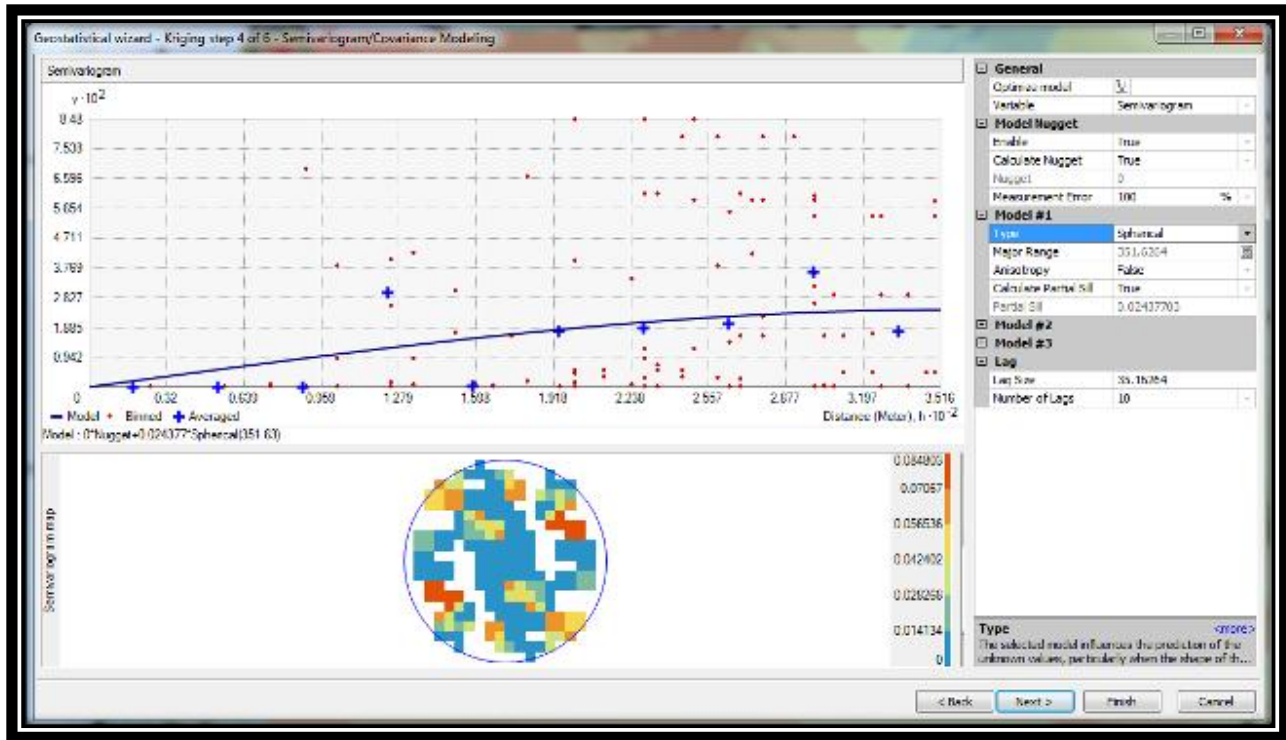
Figure 74 Semivariogram/ Covariance Modeling of Accident Sites (2009-2011)

The flattening of the semivariogram indicates that there is little autocorrelation in attribute (accident rate) values beyond the specified range. The fitting algorithm actually uses variable lag sizes, which allow spatial autocorrelation in accident sites which is to be captured, at short distances.

Thus, by removing the trend, the semivariogram will model the spatial autocorrelation among data points without having to consider the trend in the data. The trend will automatically be added back to the calculations before the final surface is produced.

The color scale, which represents the calculated semivariogram value, provides a direct link between the empirical semivariogram values on the graph and those on the semivariogram surface. The value of each cell in the semivariogram surface is color coded, with lower values shown in blue and green and higher values shown in orange and red. The average value for each cell of the semivariogram surface is plotted on the semivariogram graph and depicted as a red point.

The average value for each lag is plotted on the semivariogram graph as well, and is depicted as a blue cross. The x-axis on the semivariogram graph is the distance from the center of the cell to the center of the semivariogram surface.

11.3.10 Searching Neighborhood

A common practice followed to limit the data is by using a circle (or ellipse) to enclose the points that are used to predict a value at an unmeasured location. To avoid bias in a particular direction, the circle (or ellipse) can be divided into sectors from which an equal number of points are selected. So, by using the Searching Neighborhood dialog box, you can specify the number of points (a maximum of 200), the radius (or major/minor axis), and the number of sectors of the circle (or ellipse) to be used for prediction.

It can be assumed that as the locations get farther from the prediction location, the measured values will have less spatial autocorrelation with the prediction location. The points selected in the data view window indicate the weights that will be associated with each measured value to predict a value for the location marked by the crosshair.

It is possible to eliminate locations that are farther away that demonstrate little influence using search neighborhoods. Not only is there less relationship with locations that are farther away, but it is possible that the locations that are farther away may have a detrimental influence if they are located in an area much different than the prediction location.
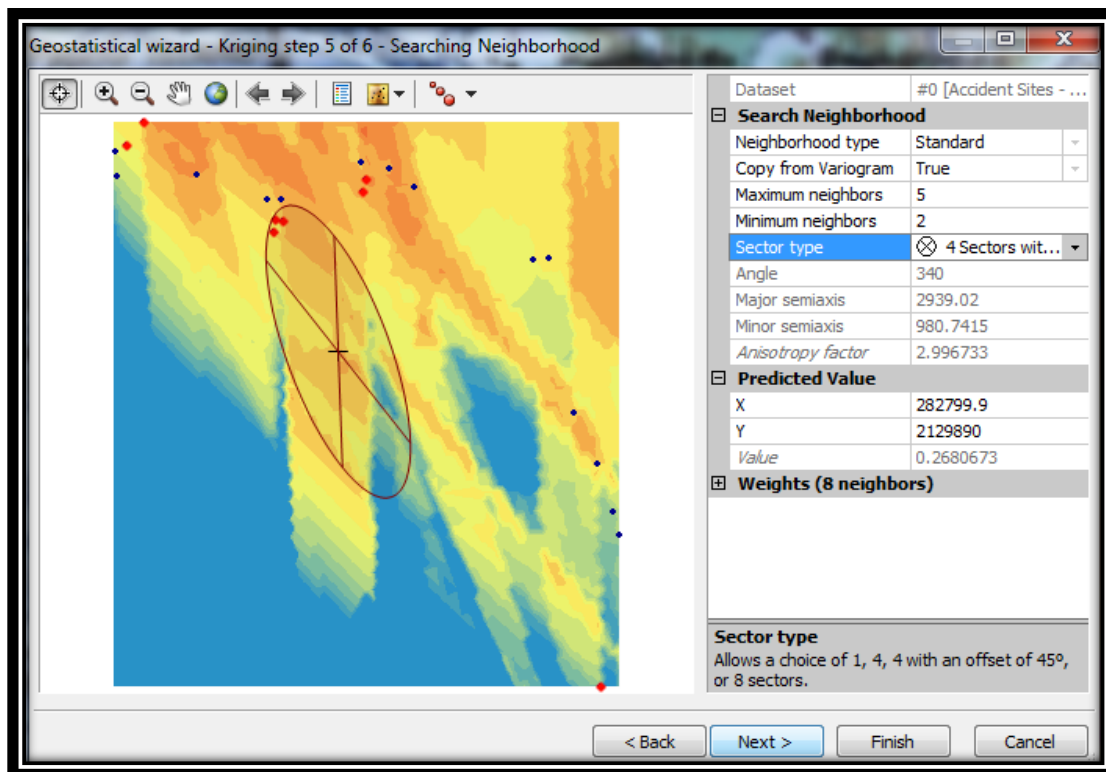
Figure 75 Tool for Searching Neighborhood of Accident Sites 2007-2009

The specified shape of the neighborhood restricts how far and where to look for the measured values to be used in the prediction. Other neighborhood parameters restrict the locations that will be used within that shape.

The shape of the neighborhood is influenced by the input data and the surface that you are trying to create. If there are no directional influences on the spatial autocorrelation of your data, then you will want to consider points equally in all directions. If there is directional autocorrelation in the data, then the shape of the neighborhood is an ellipse with the major axis parallel with the direction of long-range autocorrelation. Once a neighborhood is specified, then the location of the points within the space can be restricted.
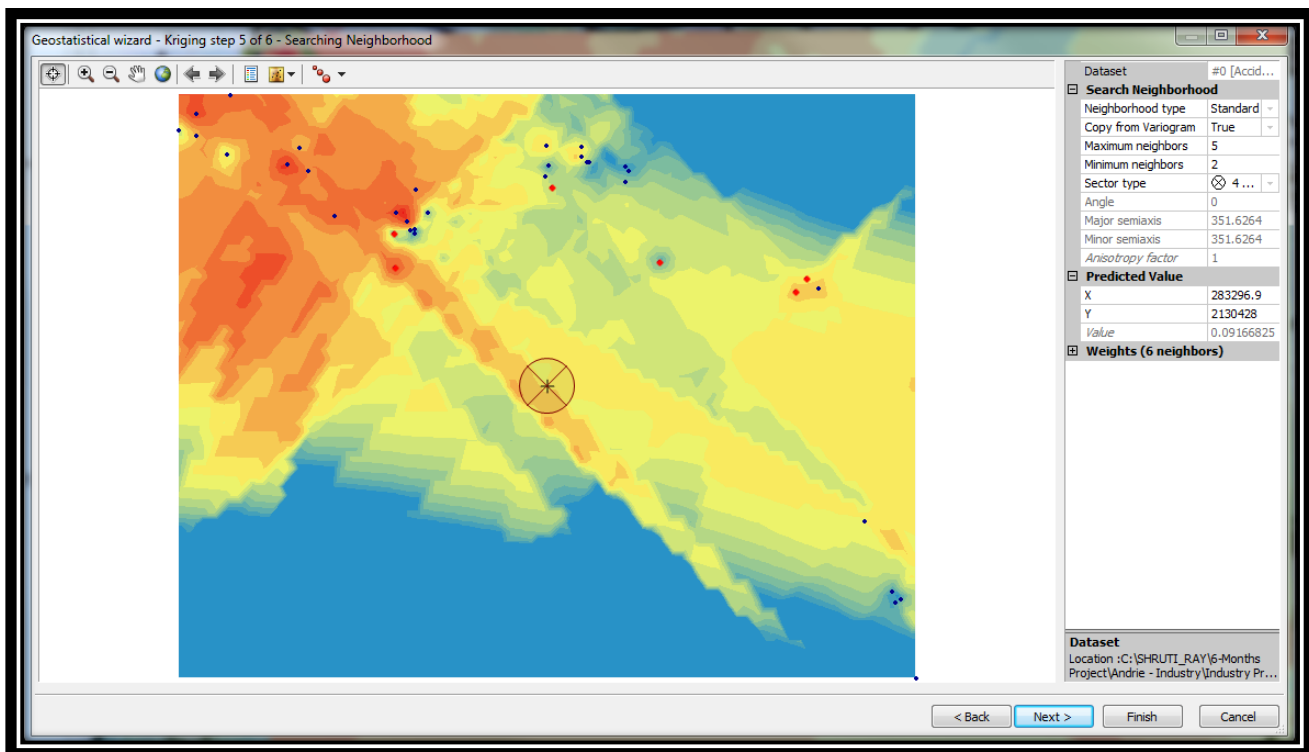
Figure 76 Tool for Searching Neighborhood of Accident Sites 2009-2011

The maximum and the minimum neighbor that we want to include can also be defined. The neighborhood can further be divided into sectors to ensure that we get values from all the directions. If the neighborhood is divided into sectors, the specified maximum and minimum number of neighbors will be applicable to each sector. Using the data configuration within the specified neighborhood, in conjunction with the fitted semivariogram model, the weights for the measured locations can be determined. From the weights and the values, a prediction can be made for the unknown value for the prediction location.

11.3.11 Cross Validation

It is a procedure for testing the quality of a predicted data distribution. In cross validation, a piece of data whose value is known independently is removed from the dataset and the rest of the data is used to predict its value. Full cross validation is done by removing, in turn, each piece of data from the dataset and using the rest of the data to predict its value.

Cross-validation uses all of the data to estimate the trend and autocorrelation models. It removes each data location, one at a time, and predicts the associated data value. For all points, cross-validation compares the measured and predicted values. In a sense, cross-validation "cheats" a little by using all of the data to estimate the trend and autocorrelation models. After completing cross-validation, some data locations may be set aside as unusual, requiring the trend and autocorrelation models to be refit.

Table 6 Cross Validation table for 2007-2009

Geostatistical wizard - Kriging step 6 of 6 - Cross Validation

Source ID	Included	Measured	Predicted	Error	Standard Error	Standardized Error	Normal Value
0	Yes	0.3734...	0.170...	-0...	0.16941100...	-1.1954200474...	-1.507904598662...
1	Yes	0.1257...	0.216...	0...	0.16403994...	0.55523853248...	0.759809563650...
2	Yes	0.28	0.110...	-0...	0.16953461...	-0.9973728526...	-1.18291683454748
3	Yes	0.02	0.244...	0...	0.16984595...	1.32108521229...	1.756826621770...
4	Yes	0.0314...	0.252...	0...	0.18099308...	1.22217705608...	1.507904598662...
5	Yes	0.1194...	0.263...	0...	0.16639441...	0.86442895444...	1.059276903083...
6	Yes	0.21875	0.230...	0...	0.16598560...	0.07125640830...	0.165664403981...
7	Yes	0.40625	0.198...	-0...	0.16366975...	-1.2672231437...	-1.756826621770...
8	Yes	0.3164...	0.163...	-0...	0.16917232...	-0.9059211996...	-0.950013740850...
9	Yes	0	0.237...	0...	0.15901163...	1.49054273908...	2.221519595441...
10	Yes	0.3464...	0.285...	-0...	0.15745831...	-0.3884779464...	-0.516846739281...
11	Yes	0	0.187...	0...	0.16915663...	1.10549106588...	1.327901869655...
12	Yes	0.24	0.165...	-0...	0.16971848...	-0.4408274899...	-0.674489749382...
13	Yes	0.3087...	0.113...	-0...	0.17020669...	-1.1495957243...	-1.32790186965507
14	Yes	0.0137...	0.056...	0...	0.08893772...	0.47934127472...	0.593821398982...
15	Yes	0	-0.013...	-0...	0.09379999...	-0.1454758643...	-0.232963854350...
16	Yes	0	0.092...	0...	0.09251341...	0.99537314665...	1.18291683454748
17	Yes	0.40625	0.113...	-0...	0.12578012...	-2.3285279778...	-2.221519595441...
18	Yes	0.1212...	0.047...	-0...	0.16834502...	-0.4404785715...	-0.593821398982...
19	Yes	0.0616...	0.031...	-0...	0.13053783...	-0.2335373843...	-0.371149337180...
20	Yes	0	0.082...	0...	0.14329992...	0.57813352961...	0.950013740850...
21	Yes	0.0410...	0.067...	0...	0.14898459...	0.17974870610...	0.442821672283...
22	Yes	0.0817...	0.095...	0...	0.14243322...	0.09612311320...	0.232963854350...
23	Yes	0.1466...	0.036...	-0...	0.11574826...	-0.9547433387...	-1.059276903083...
24	Yes	0.0133...	-0.000...	-0...	0.07840767...	-0.1715555810...	-0.301336529775...
25	Yes	0	0.041...	0...	0.07274462...	0.56986303735...	0.851058514467...
26	Yes	0	0.013...	0...	0.10668590...	0.12456233177...	0.301336529775...
27	Yes	0.0062...	0.005...	-0...	0.11420102...	-0.0104187026...	-0.099107844935...
28	Yes	0.0281...	0.027...	-0...	0.14847878...	-0.0013033125...	-0.032987929738...
29	Yes	0	0.086...	0...	0.16757859...	0.51627124020...	0.674489749382...
30	Yes	0.1446...	0.061...	-0...	0.16328548...	-0.5118414249...	-0.759809563650...
31	Yes	0.1346...	0.076...	-0...	0.16323445...	-0.3566669448...	-0.442821672283...
32	Yes	0.0897...	0.093...	0...	0.15483662...	0.02296809100...	0.099107844935...
33	Yes	0.0817...	0.085...	0...	0.17019765...	0.02280065027...	0.032987929738...
34	Yes	0	0.028...	0...	0.14015491...	0.20511877525...	0.516846739281...
35	Yes	0.00625	0.004...	-0...	0.12190474...	-0.0183512450...	-0.165664403981...
36	Yes	0	0.018...	0...	0.11955640...	0.15308670747...	0.371149337180...
37	Yes	0	-0.111...	-0...	0.17863703...	-0.6217709476...	-0.851058514467...

Predicted ·10¹

Measured ·10¹

Predicted / Error / Standardized Error / Normal QQPlot

Regression function	0.314098055731087 * x + 0.0676490663055182
Prediction Errors	
Samples	38 of 38
Mean	-0.00751083
Root-Mean-Square	0.1213329
Mean Standardized	-0.04120787
Root-Mean-Square Standardized	0.7937123
Average Standard Error	0.1476931
Export Result Table	

< Back   Next >   Finish   Cancel

Table 7 Cross Validation table of Accident Sites for 2009-2011

The calculated statistics serve as diagnostic that indicate whether the model is reasonable for map prediction. The objective of cross validation is to help us in making an informed decision model about which model provides the most accurate prediction. For a model that provides the most accurate prediction, the mean error should be close to 0, the root mean square error and average standard error should be as small as possible, and the root mean square standardized error should be close to 1.

The term prediction error is used for differences between the prediction and the actual measure value. The Cross Validation dialog box also displays scatter-plots that show Error, Standardized Error and QQ Plot for each data point.

The Prediction Standard Error Map for the Accident Sites for 2007-2009 and 2009-2011 calculated with respect to the Accident Rate for the same years have been given below:

The prediction standard error quantifies the uncertainty for each location in the surface that has been created.

Figure 77 Prediction Standard Error Map of Accident Sites for 2007-2009
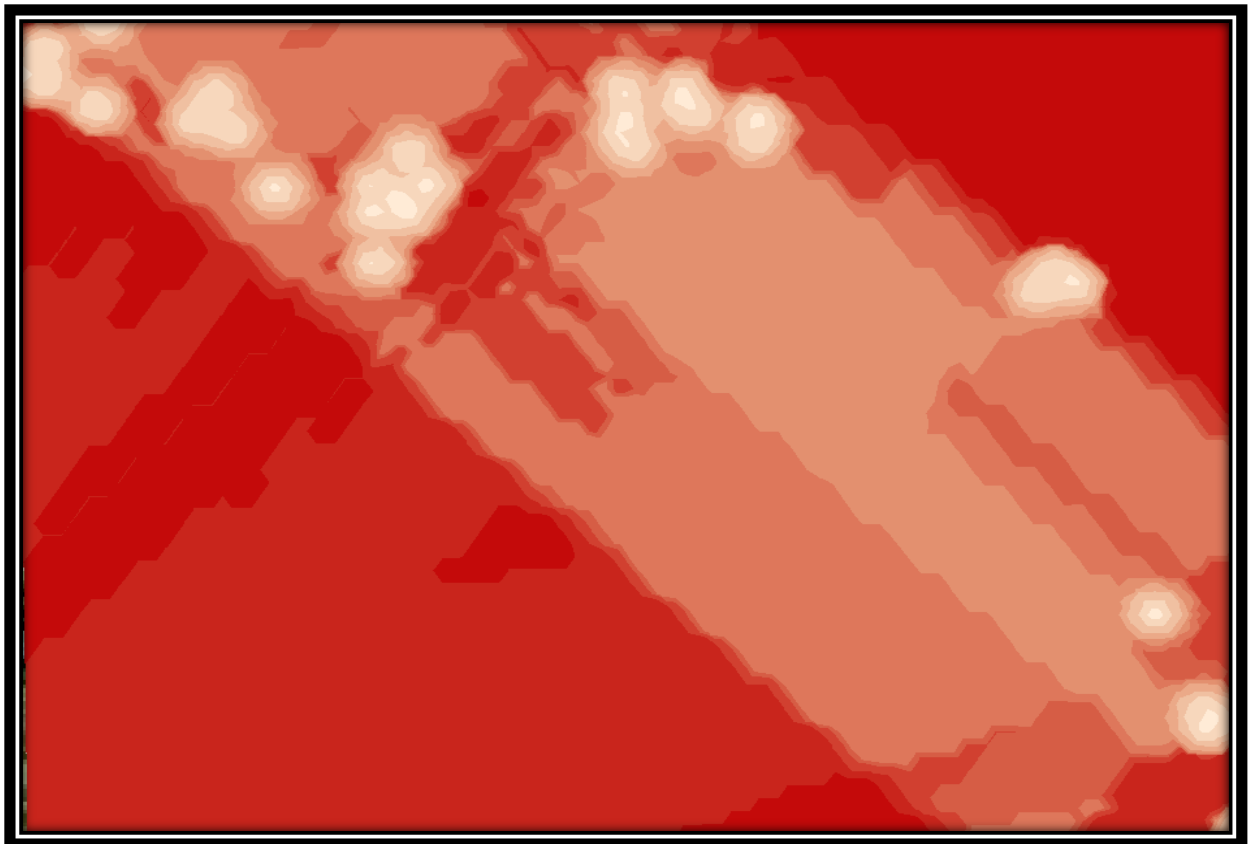
Figure 78 Prediction Standard Map for Accident Sites for 2009-2011

11.3.12 Comparing Models

Comparison helps you determine how good the model that created a geostatistical layer is relative to another model. The two geostatistical layers that you are comparing could have been created by two different models, or they could have been created with the same model but with different parameters. In the first case, you are comparing which method is best for your data, and in the second, you are examining the effects of different input parameters on a model when creating the output surface.

The Comparison dialog box uses the cross-validation statistics. However, it allows us to examine the statistics and the plots side by side. Generally, the best model is the one that has the standardized mean nearest to zero, the smallest root-mean-squared prediction error, the average standard error nearest the root-mean-squared prediction error, and the standardized root-mean-squared prediction error nearest to one.
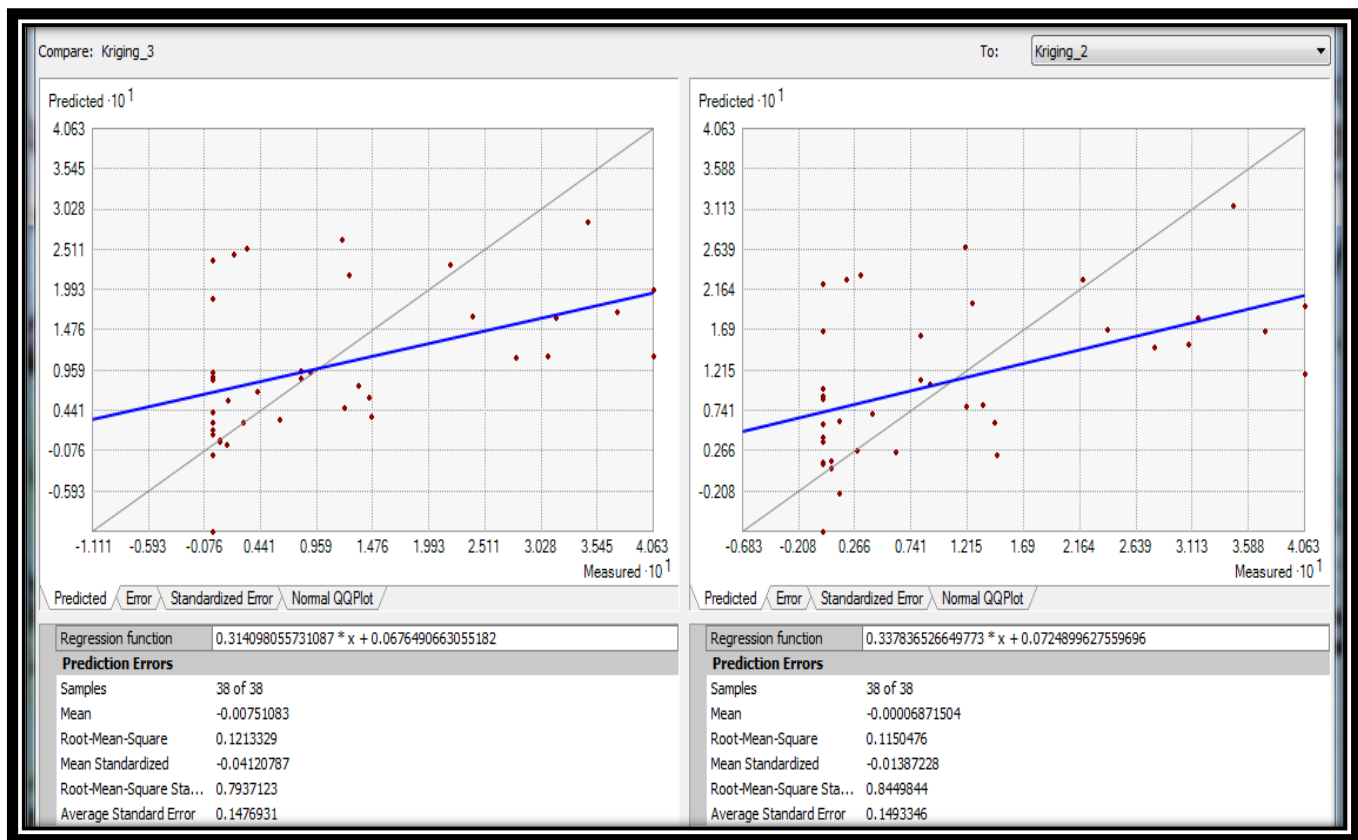
Figure 79 Cross Validation for 2009-2011 showing a comparison

11.4 Slope Analysis

The Slope tool calculates the maximum rate of change between each cell and its neighbors, for example, the steepest downhill descent for the cell (the maximum change in elevation over the distance between the cell and its eight neighbors). Every cell in the output raster has a slope value. The lower the slope value, the flatter is the terrain and the higher the slope value, the steeper the terrain. The output slope raster can be calculated as percent of slope or degree of slope.

When the slope angle equals 45 degrees, the rise is equal to the run. Expressed as a percentage, the slope of this angle is 100 percent. As the slope approaches vertical (90 degrees), the percentage slope approaches infinity.

The Slope tool is most frequently run on an elevation dataset, as the following diagrams show. Steeper slopes are shaded red on the output slope raster. However, the function can also be used with other types of continuous data, such as population, to identify sharp changes in value.
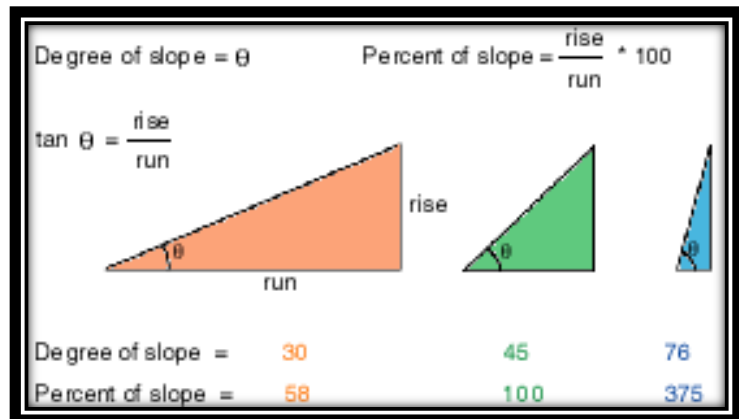
Figure 80 Slope Analysis
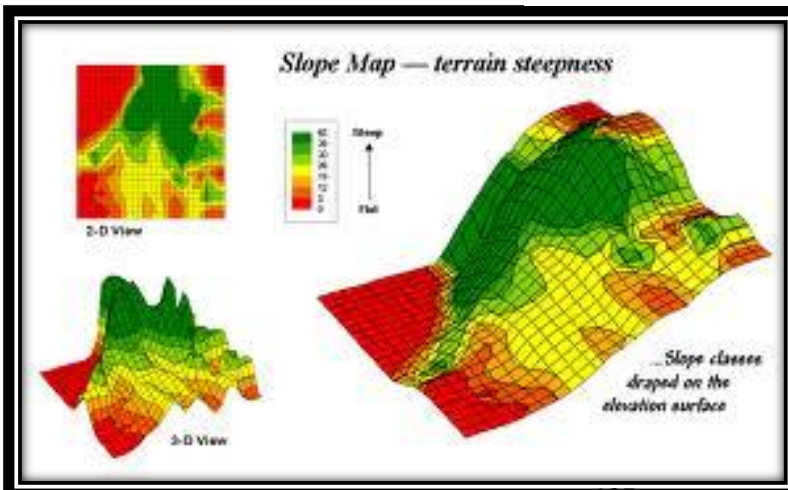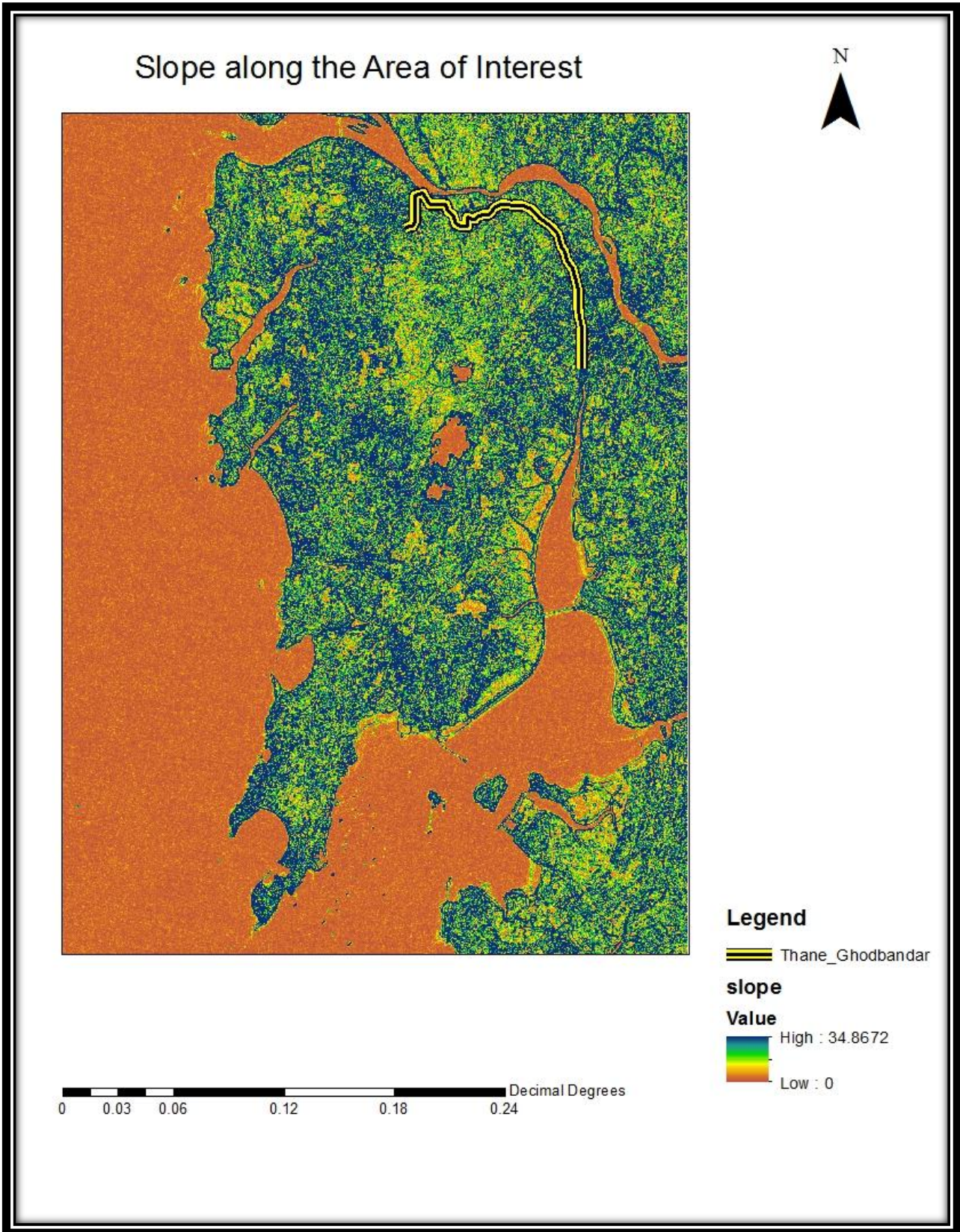
Figure 81 Slope shown along the selected AOI

11.5 Curvature

GIS applications in hydrological studies often require computation of surface curvature to determine if the surface at a cell location is upwardly convex or concave. We have a common algorithm, to calculate surface curvature; it fits a 3 by 3 window with a quadratic polynomial equation, known as Zevenbergen and Thorn. Profile curvature is estimated along the direction of maximum slope. Plane curvature is estimated across the direction of maximum slope. And curvature measures the difference between the two: profile curvature – plan curvature. A positive curvature value at a cell means that the surface is upwardly convex at the cell location. A negative curvature value means that the
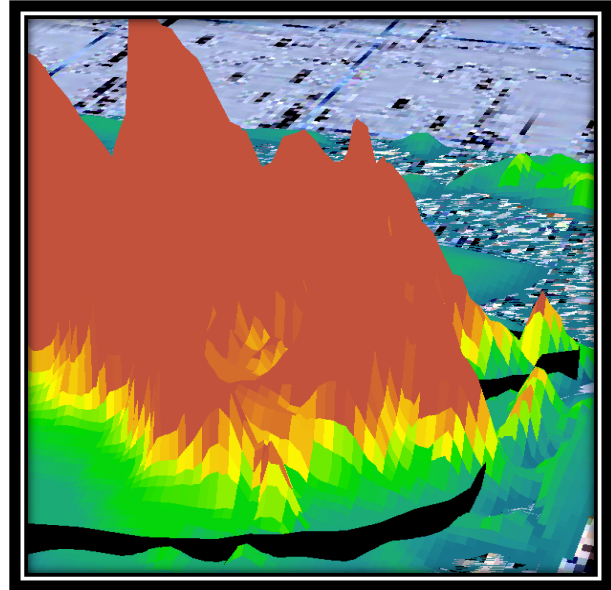
Figure 82 Curvature along the AOI

surface is upwardly concave. And a zero value means that the surface is flat.
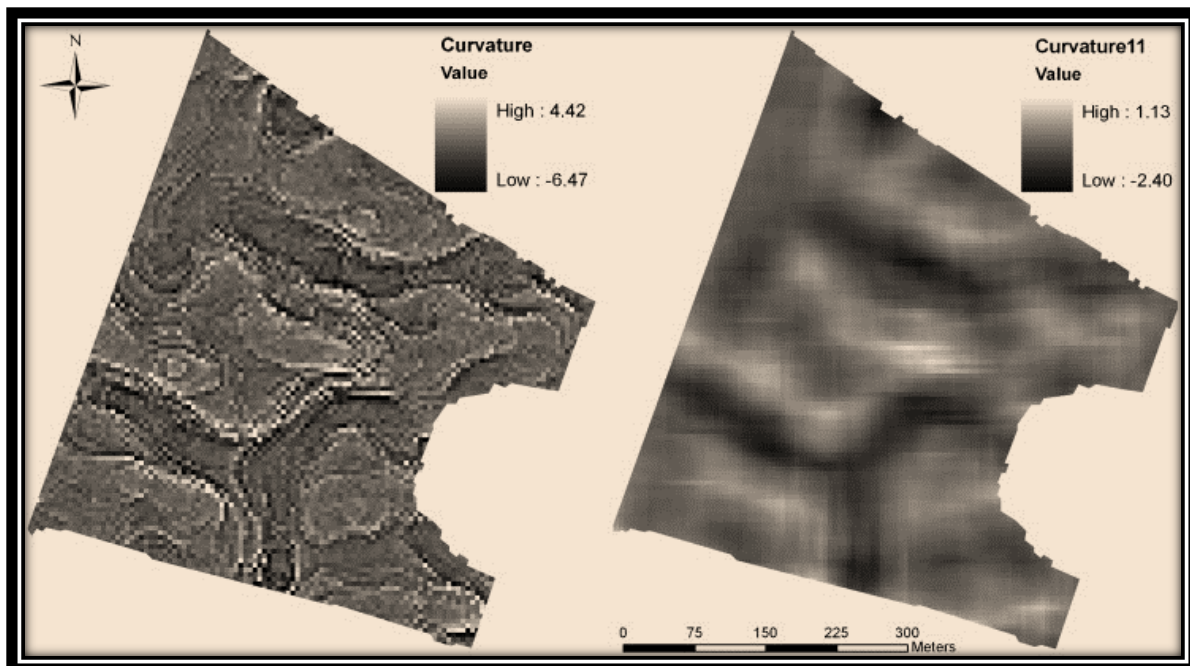
Figure 83 Different Values of Curvature being shown in the AOI
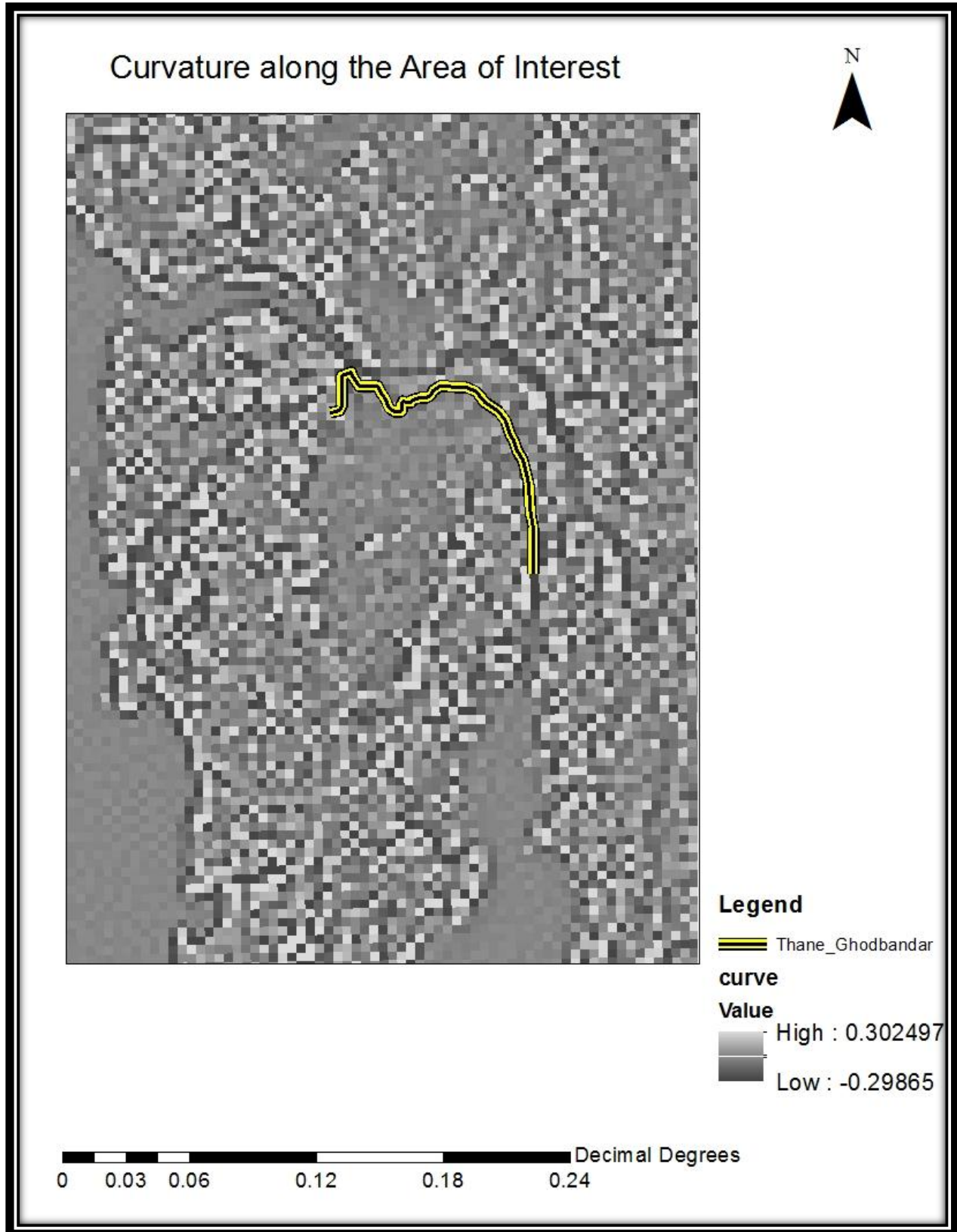
Figure 84 Curvature along the AOI

CHAPTER 12 – STATISTICAL ANALYSIS

Statistics is the study of collection, organization, analysis, interpretation and presentation of data. It deals with all aspects including the planning of data collection in terms of design of surveys and experiments. Statistics is a mathematical science involving the collection, interpretation, analysis, and presentation of data. It is often used to make predictions based on data.

Statistics can be described as the scientific, mathematical study of data. In very large datasets, it is impossible or impractical to quickly analyze every piece of data, so a sample of the data is studied and the rest of the data results can be extrapolated from the sample data. This gives mathematicians results faster and allows for other mathematical exercises such as trend analysis, frequency and distribution analysis and many other types of data interpretation.

The Statistical analysis of accident is carried out periodically at critical locations or road stretches which will help to arrive at suitable measures to effectively decrease accident rates. It is the measure (or estimates) of the number and severity of accident. Accident prone stretches of different roads may be assessed by finding the accident density per length of the road. The places of accidents are marked on the map and the points of their clustering **(black spot)** are determined. By statistical study of accident occurrence at a particular road or location or zone of study for a long period of time it is possible to predict with reasonable accuracy the probability of accident occurrence per day or relative safety of different classes of road user in that location.

The interpretation of the statistical data is very important to provide insight to the problem.

The position of India in the year 2011 in country-wise number of person killed per 100000 populations is shown and the increase in rate of accident from year 2005
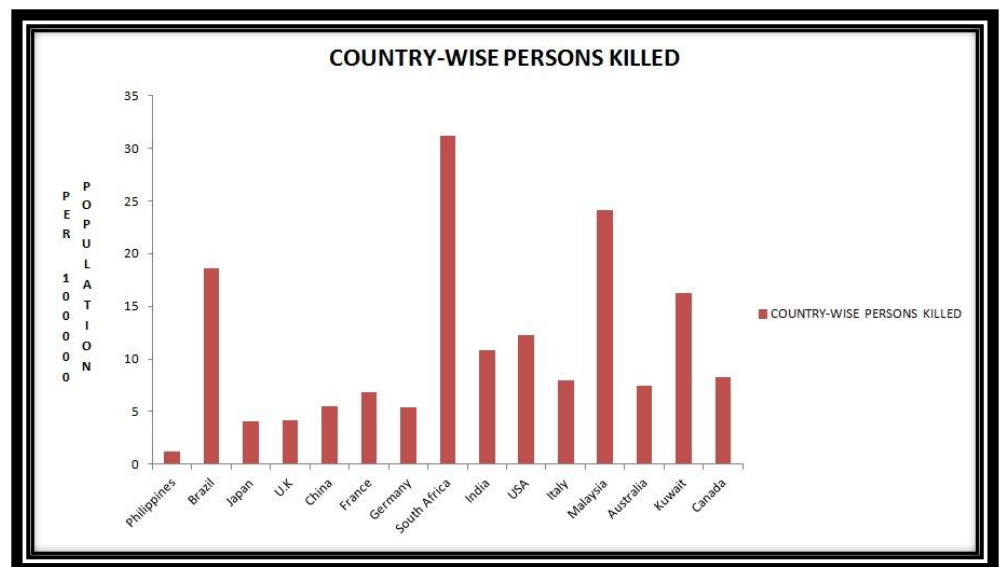

Figure 85 Country wise representation of total persons killed in accidents

to year 2011 is also shown. In 2011, 14 accidents occurred per hour.

Road accidents are not reliably and regularly published by all countries of the world but accidents of the most severe form, those involving a fatality, are normally more reliably recorded than the other types. It is the scientific discipline that provides methods to help us make sense of data. It is a formal process of making decisions based on collecting, analyzing, and drawing conclusions from data, which uses sound processes and scientific tools. A familiarity with the goals and method of statistics teaches us how to make intelligent judgments and informed decisions in the presence of uncertainty and variation. Statistics forms an essence of any good project.

There are namely two main branches of statistics:

- **Descriptive Statistic**s: Methods for organizing and summarizing data which provide increased understanding and an effective way to present data. Bar Charts, Line Graphs, Pie Diagrams are also examples of descriptive statistics.

- **Inferential Statistics**: Generalizing from a sample to the population from which it was selected and assessing the reliability of such generalizations. Trend Line is an example of inferential statistics.

For our project, descriptive statistics has been used because the statistical data was readily available with us.

There are different types of datasets that are used for the purpose of statistical representation. There are three main types of datasets that are used, namely:

- **Univariate Data:** A dataset consisting observations on a single attribute (or variables).

- **Multivariate Data:** A dataset consisting of observations on two or more attributes (or variables).

- **Bivariate Data:** A dataset consisting of observations on only two attributes (or variables). It is a special case of Multivariate Data.

In this case, we have mainly used Bivariate and Multivariate Data for the representation of our data.

YEAR	2007	2008	2009	2010	2011
MAHARASHTRA	16.14993878	15.75213417	16.48427451	16.36140762	15.37114787
STATE HIGHWAY 42	6.186474432	5.972645437	6.662417047	6.613729064	6.240609662

Table 8 Road Accidents Statistics for Maharashtra & SH – 42
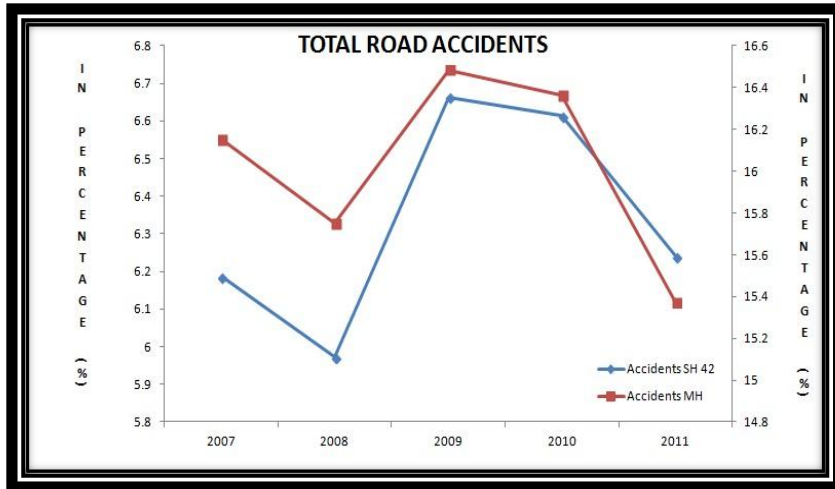
Figure 86 Representation of Total Road Accidents for Maharashtra & SH - 42

This graph illustrates the total number of accidents that has taken place in Maharashtra and also along our AOI. This has been calculated with respect to the total accident figures of the country and then the required percentage has been calculated. Also the data is for the period 2007-2011.

YEAR	2007	2008	2009	2010	2011
MAHARASHTRA	11.02700063	10.66207	10.80258613	10.72634257	9.796931
STATE HIGHWAY 42	0.45815019	0.576562	0.68865302	0.632630096	0.568837

Table 9 Persons killed in road accidents for Maharashtra & SH - 42

The graph on the right gives us a picture of the total number of persons who have been killed in road accidents. It displays the statistics of both Maharashtra and also of our area, which is the Thane-Ghodbunder Road. This data has been represented in percentage after being compared to the country's data, for a period of 5 years.
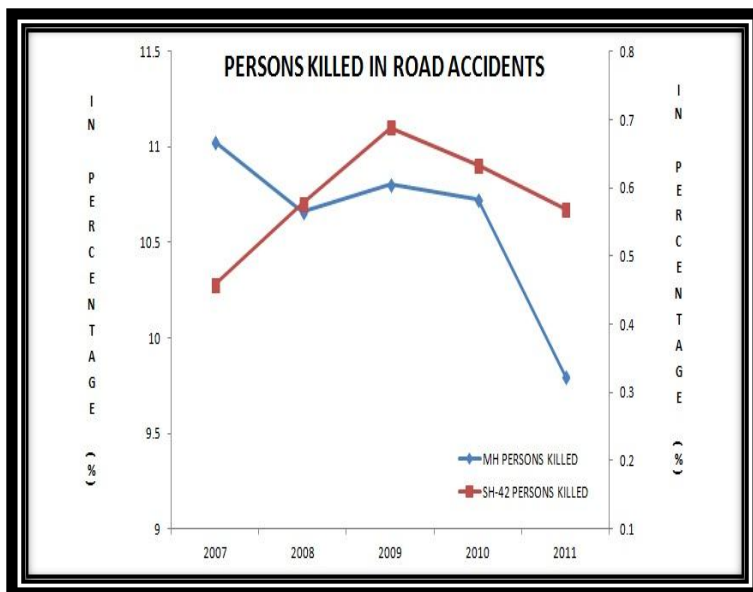
Figure 87 Representation of persons killed in road accidents in Maharashtra & SH - 42

YEAR	2007	2008	2009	2010	2011
MAHARASHTRA	10.73078355	10.20276801	8.77704274	10.27713044	9.222347762
STATE HIGHWAY 42	1.429484145	1.197362444	1.606767509	1.504790717	1.268165349

Table 10 Persons injured in road accidents for Maharashtra & SH - 42
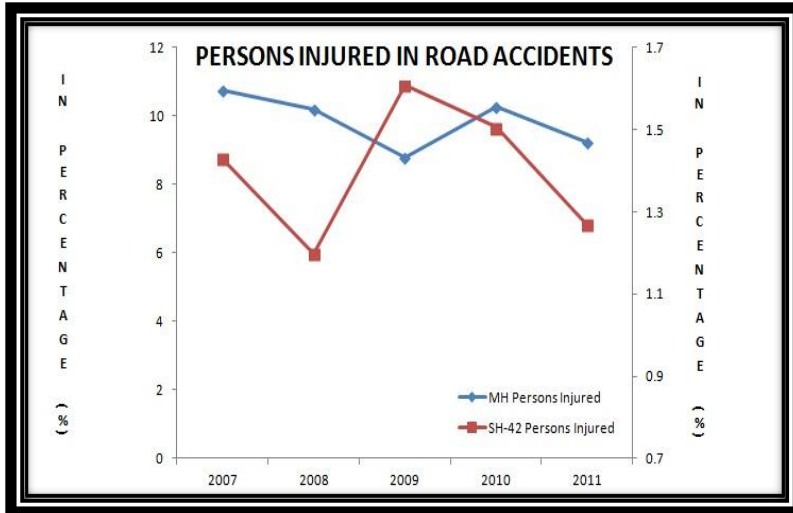
Figure 88 Persons killed in road accidents in Maharashtra & SH - 42

This graph represents the total number of persons injured in accidents for the last 5 years (2007-2011). This thus makes our work easy in comparison as the base for both of them is same. They have been calculated with reference to the total persons injured in India.

YEAR	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Tot. Pop. (thousands)	1016710	1028610	1045547	1067388	1079117	1095722	1112186	1128521	1144734	1160813
Urban Pop. (thousands)	281416	285748	294668	301224	307721	314145	321623	329112	336746	371460
Vehicle Pop. (Thousands)	52.37	54.99	58.92	67.01	72.72	81.5	89.61	100.7	110.52	121.63

Table 11 Statistics of total and urban population and vehicle population (2002-2011)

This graph helps us to compare the increase in the population and the vehicle population. It can be seen that after 2006, there has been a steep rise in the vehicle population and thus resulting in more number of road accidents.
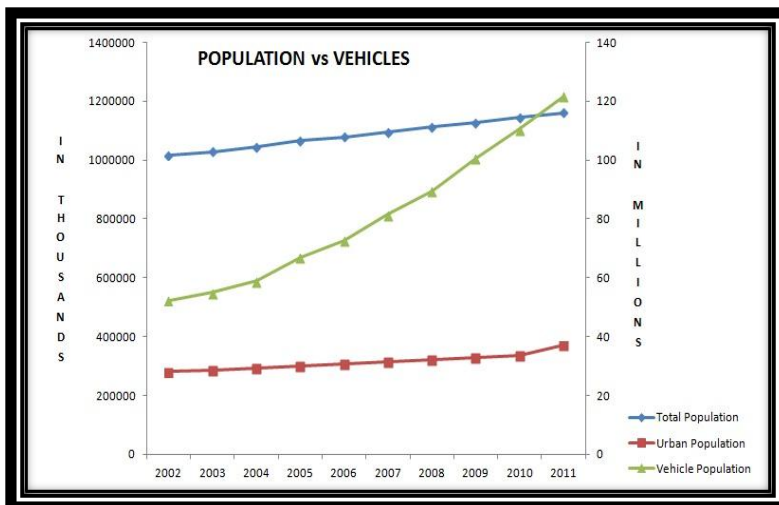
Figure 89 Trend showing the total and urban population along with the vehicle population

ROAD TYPES	SURFACED (in kms.)	UNSURFACED (in kms.)
MAHARASHTRA	13724	600
THANE	212001	60683

Table 12 Road types in Thane & Maharashtra

The bar graph represents the road surface in Maharashtra as well as Thane. Since, our AOI is a part of the Thane region so the road type associated is more associated with this area. It is obvious that Thane being a developed region has a larger percentage of surfaced roads as compared to the unsurfaced roads.
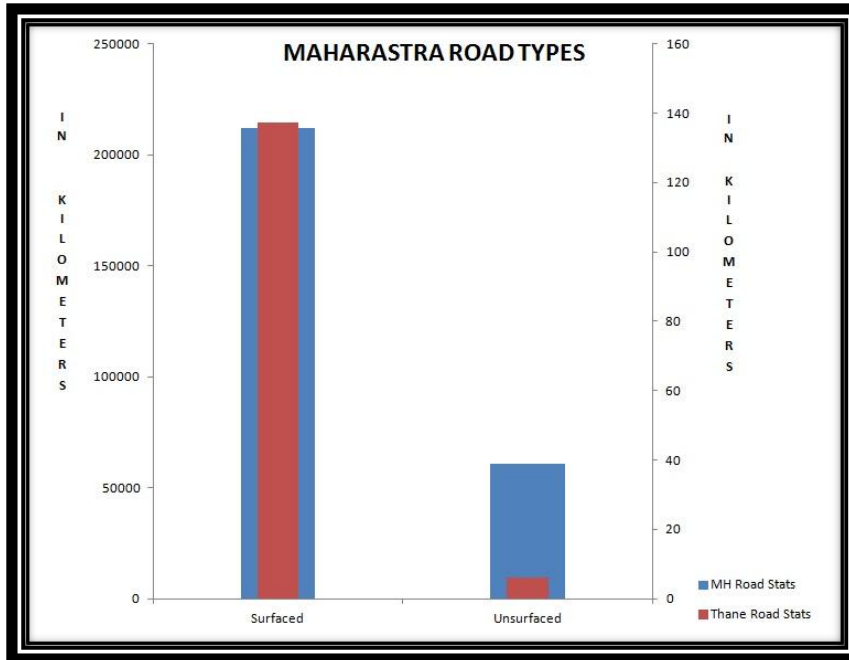
Figure 90 Road Types Representation

YEAR	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Accident Severity	19.9	20.8	21.1	21.5	21.6	22.9	23.9	24.7	25.8	26.9

Table 13 Accident Severity from 2002-2011

This graph shows the increasing accident severity over the years. Accident severity is defined as the total number of persons killed per 100 accidents. After 2006, we can observe a sharp increase in the curve which we can relate to the increase in the total number of vehicle population.
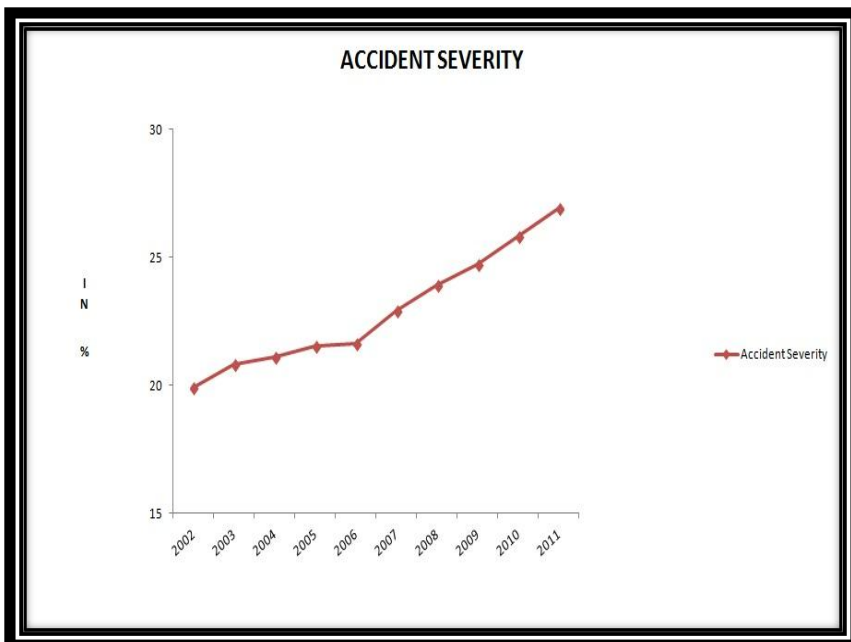
Figure 91 Accident Severity for 2002-2011

143

CATEGORY	VEHICLE POPULATION (SH 42)	VEHICLE POPULATION (MAHARASHTRA)
Two Wheelers	721705	12429011
Cars	312489	2027080
Jeeps	36884	394647
Taxi meter fitted	3442	57047
Tourist Cabs	24302	111449
Auto rickshaws	86668	640700
School Buses	280	7415
Private Services Vehicles	1615	8326
Ambulances	1210	9571
Trucks & Lorries	70460	334863
Others	2263	27188

Table 14 Vehicle Population of SH - 42 & Maharashtra
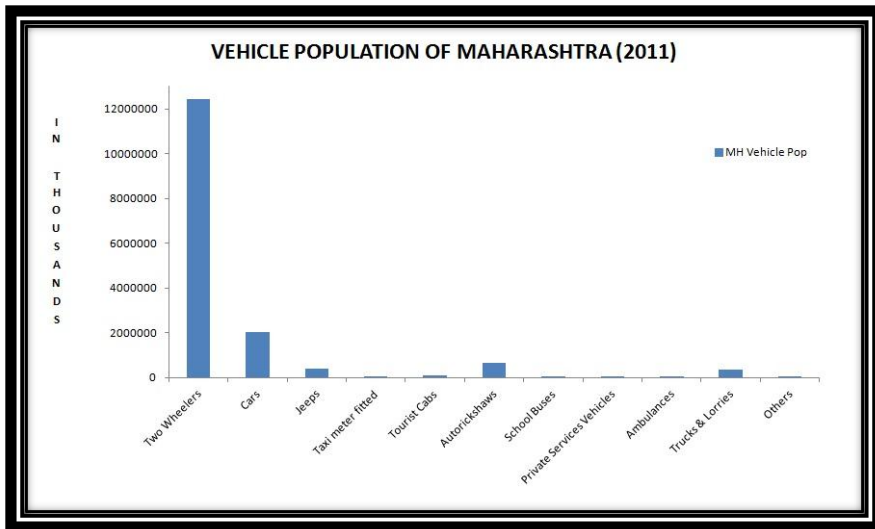
Figure 92 Vehicle Population of Maharashtra

This graph represents the total vehicle population for the entire state of Maharashtra as well as for our AOI, which is the Thane-Ghodbunder Road, for the year 2011. It can be seen from these column graphs that the number of two wheelers is highest in both cases and is then followed by cars. Our AOI, being a state highway does have considerable amount of truck and lorries travelling. The remaining occupies almost negligible amounts.

Figure 93 Vehicle Population of SH - 42

YEAR	ACCIDENTS ON NATIONAL HIGHWAY (in %)	ACCIDENTS ON STATE HIGHWAY (in %)
2002	28.6	22.5
2003	32.3	23.5
2004	31.4	22.4
2005	30.3	23.5
2006	29.6	23.6
2007	30.4	18.5
2008	29	24.4
2009	28.5	25.6
2010	29.3	23.8
2011	30	24.5

Table 15 Accidents occurring on National & State Highway
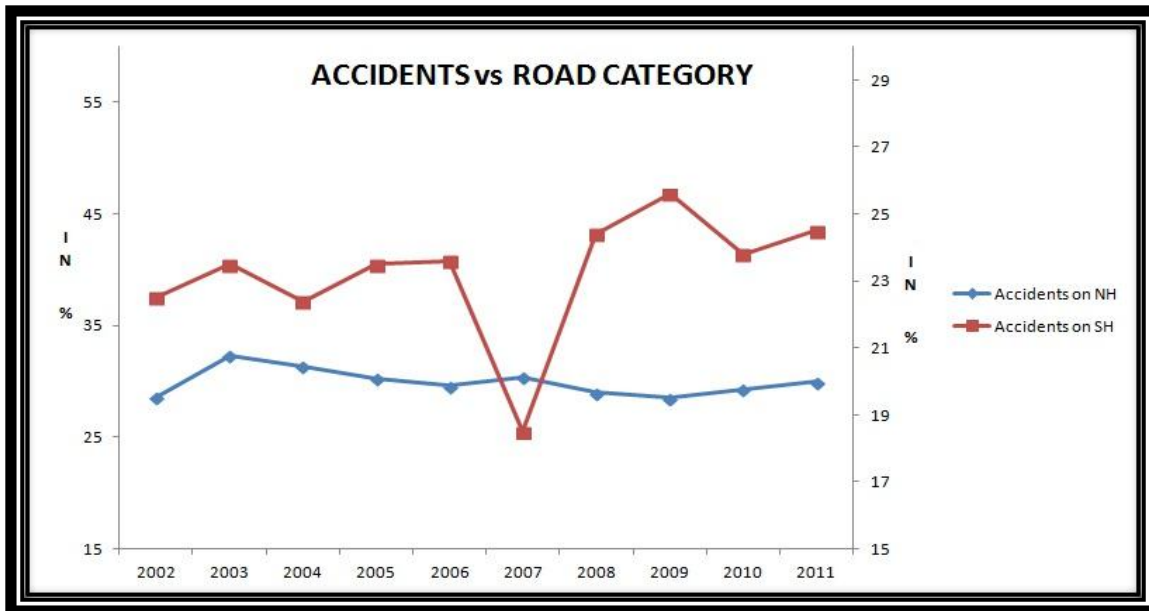
Figure 94 Correlation between accidents & road category

The above graph illustrates the relation between the road category and the rate of accidents. Road category refers to the division of the road type like, National Highway, State Highway and others. It can be seen that in the last 10 years, the accidents on National Highways have increased from 28.6% to 30% and it has been from 22.5% to 24.5 in case of State Highways. The primary cause of the increase in these accidents is over-speeding. It's the general human nature to over-speed while driving on the highways, which leads to these accidents.

• • •

CATEGORIES	TWO-WHEELERS	AUTO RICKSHAW	CARS	BUSES	TRUCKS & LORRIES	OTHERS
ACCIDENT %	23.7	7.3	21.8	9.5	23.3	14.4

Table 16 Vehicle contributing to accidents in Thane-Ghodbunder Road
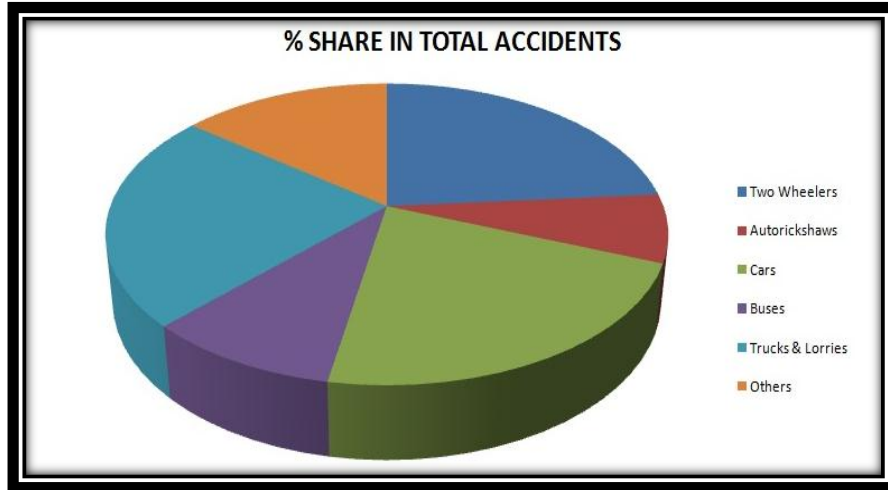
Figure 95 Vehicles contributing to accidents in SH - 42

The above pie chart shows the distribution of accidents in the Thane Ghodbunder Road with respect to the vehicle category. It can be seen from the chart that the major share is taken up by the two-wheelers and trucks & lorries. It is then followed by cars. Since, it is a state highway the prevalence of cars are quite high and thus explaining its contribution to road accidents.

TIME	06:00-09:00	09:00-12:00	12:00-15:00	15:00-18:00	18:00-21:00	21:00-00:00	00:00-03:00	03:00-06:00
% SHARE IN ACCIDENTS	11.8	16.2	14.9	16.7	16.1	10.7	6.6	7

Table 17 Date Time Modifier showing accident occurrences
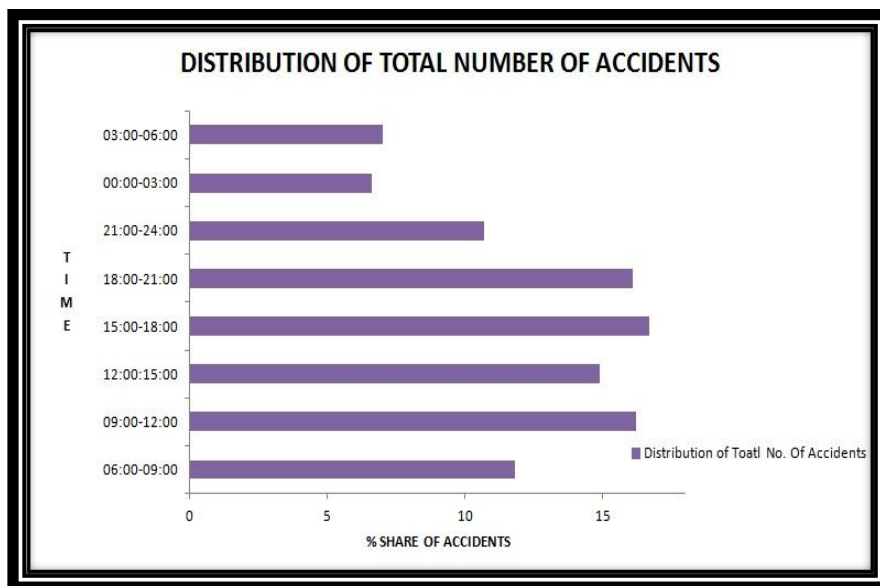
Figure 96 Date Time Modifier showing accident occurrences

From this given bar, it can be seen that the maximum percentage of accident occurs during 03:00PM-06:00PM. This is when the rush is at its peak and thus resulting in most number of road accidents. It is followed by 09:00AM-

• • •

12:00PM.

LULC Class	Thane Area (sq. km)	AOI Area (sq. km)
Urban - Built Up	295.19	9.37
Urban - Mining	2.2	0.02
Rural - Built Up	59.09	0.93
Plantation	650.31	4.34
Crop Land	2720.53	0.61
Fallow	789.68	4.74
Evergreen/Semi-Evergreen	2144.7	1.22
Deciduous	426.51	3.24
Swamp/Mangroves	10.26	
Scrub Forest	437.19	0.05
Barren/Wastelands	1245	1.13
Wetlands/Water Bodies	777.34	3.11

Table 18 LULC Classes for Thane & SH - 42

Figure 97 LULC representation of Thane
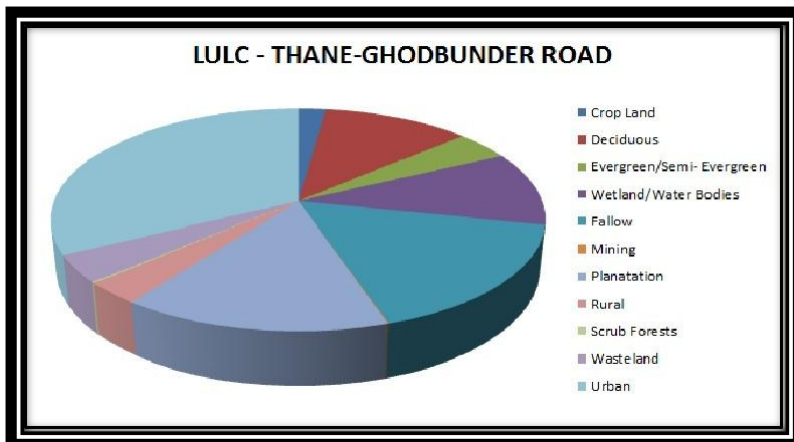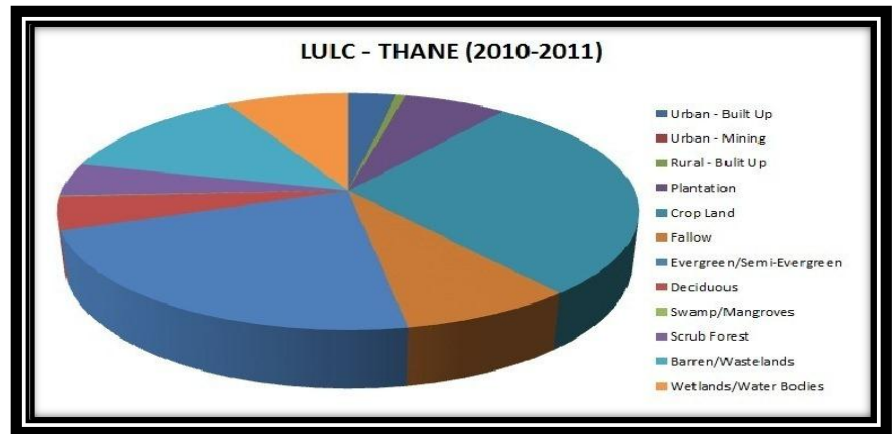
Figure 98 LULC representation of SH - 42
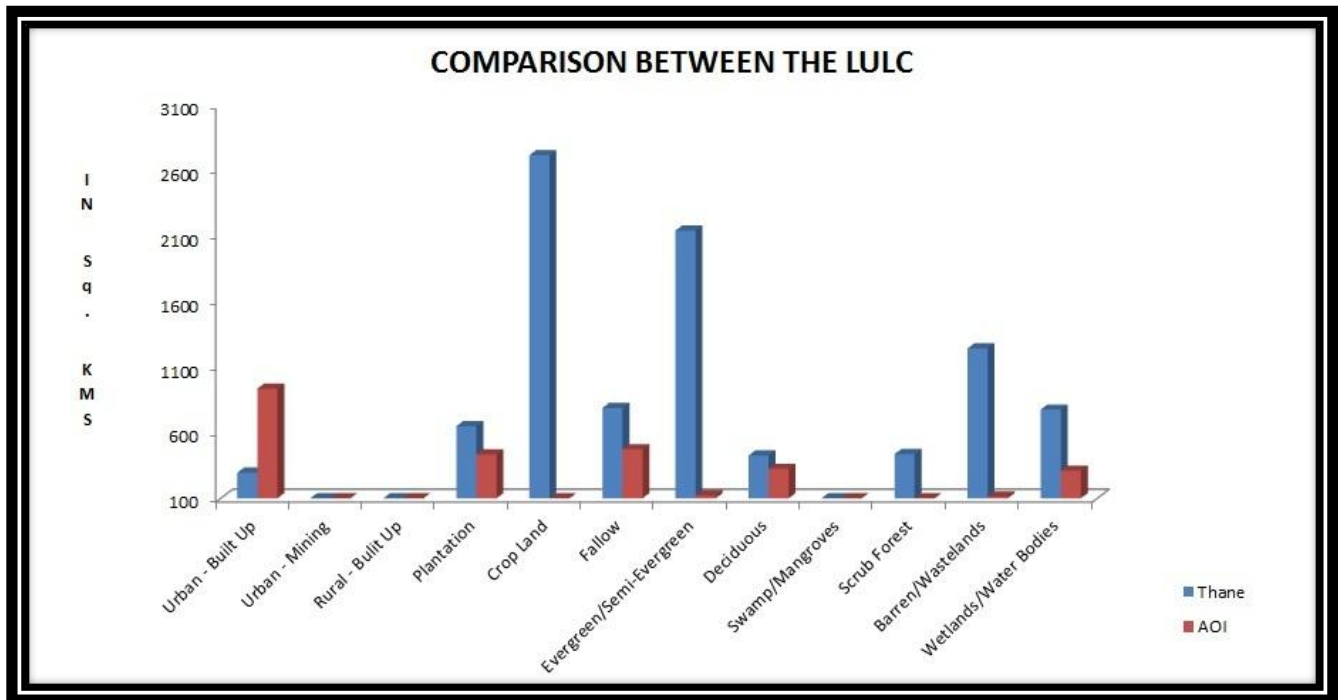
The above compound graph explains the different LULC (Land Use and Land Cover) classes. It covers almost all the broad classes ranging from settlements to plantation to crop land. In this diagram we can see that crop land covers the maximum area followed by evergreen/semi-evergreen forests. Since the region is situated along the Western Ghats so the percentage of forests is high. Also to add to this is the Sanjay Gandhi National Park which is located in close proximity. So the chance of having a high percentage of forests is noticeable.

It can also be added that the area has about 20% of the total area devoted to barren land or wasteland.

CHAPTER 13 – SURVEY ANALYSIS

Surveying is an important component of any report. It is like an evidence to support the analysis that has been carried out and data forms an integral part of this survey. Data can be defined as the quantitative or qualitative values of a variable. It is the smallest unit of information from which other measurements and analysis are carried out. Data sources are broadly classified as Primary and Secondary Data.

Primary data refers to something collected from the original first hand source. It is more reliable, authentic and objective. It cannot be changed by human so its validity is greater than secondary data.

Secondary data is the data that has already been collected by and is readily available from other sources. They are cheaper and can be obtained quickly as compared to the former.

In normal circumstances, secondary data is not given much importance but in a country like India, *a word of mouth is considered as the most reliable source*. Thus, our entire survey is based on a third party data, where we have visited our AOI in order to find out more information about that area.

A questionnaire had been drafted, which had a list of questions which were asked to people along this road. On the basis of that, a database was created which had a record of all the people who were surveyed. There replies were then represented in the form of statistical diagrams to further analysis accidents.

13.1 Accident Report Form

The main purpose of the database was to answer any kind of questions related to accident occurrence, its type, place and few more. For example, the question of "who" was involved for analysis purpose needs more than just the person's name. All the people involved in the accident, along with their ages, sexes needs to be recorded. The vehicle types involved in the collision must also be known along with its details.

The accident Report Form can be divided into three main section: General details, Vehicle (+Driver) details and Casualty Details. The relationship between these three areas can be one to many, and thus the accident data should be stored in relational database.

13.2 Location Coding

It is important that the accident data is recorded accurately and a sketch map can be provided showing the actual scene. This helps others to precisely know the spot. From this given sketch the accident location can be located on the map and will have more than one reference point. For the GIS systems, the geographical co-ordinates of that spot can also be acquired.
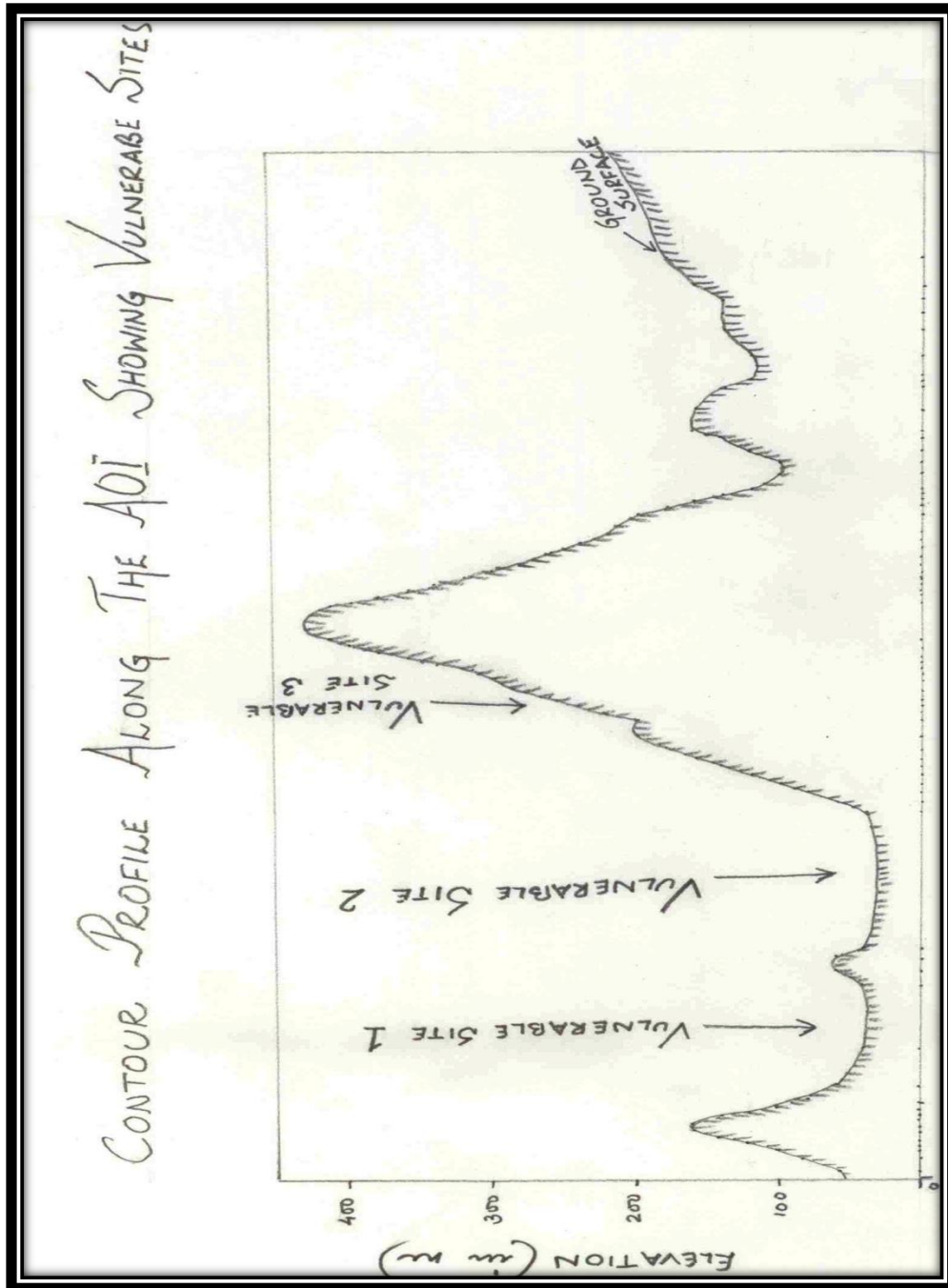
Figure 100 Snapshots showing the AOI along with its important location for accidents

Figure 101 A snapshot of the data collection process

13.3 Methodology for the Survey

```
┌─────────────────────────┐
│     QUESTIONNAIRE        │
│     PREPARATION          │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│     GROUND TRUTH         │
│     VALIDATION           │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│    SURVEY CONDUCTED      │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│     EVIDENCES            │
│     ANALYSIS             │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│     DATABASE             │
│     PREPARATION          │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│     STATISTICAL          │
│     REPRESENTATION       │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│     SURVEY REPORT        │
│     GENERATION           │
└─────────────────────────┘
```
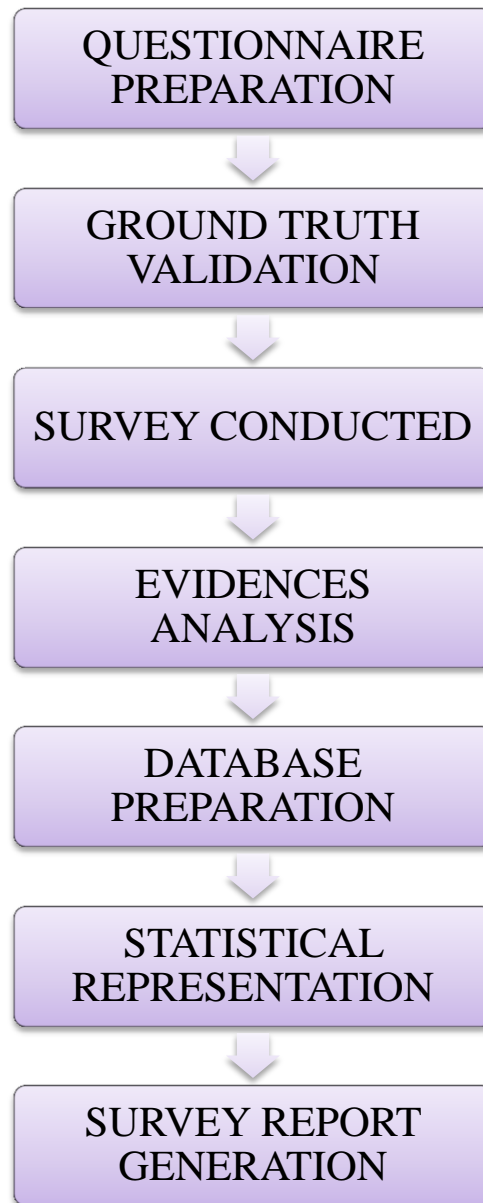
13.3 Database of the Survey Conducted

Sr. No.	NAME	AGE (yrs)	SEX	EMPLOYMENT STATUS	TOTAL FAMILY MEMBERS	AFFECTED BY ACCIDENT (Y/N)	MEMBERS AFFECTED	TYPE OF INJURY (Major/Minor)	ACCIDENT WITNESSED (Y/N)
1	Akash Lele	23	Male	Student	9	N	0	Unaffected	N
2	Anmol Divekar	36	Male	Worker	5	Y	1	Minor. Few Bruises	N
3	Gagan Shirke	20	Male	Worker	4	N	0	Unaffected	N
4	Sanjita Barwe	29	Female	Employed	3	N	0	Unaffected	Y
5	Parth Kale	31	Male	Labourer	5	N	0	Unaffected	N
6	Nupoor Soman	53	Female	Unemployed	6	Y	5	Major. Sole survivor	Y
7	Mehul Dagade	18	Male	Student	3	N	0	Unaffected	N
8	Haresh Bopde	40	Male	Employed	4	Y	1	Major. Head on Collison	Y
9	Ishan Pawgi	49	Male	Labourer	10	N	0	Unaffected	Y
10	Mahesh Pulekar	33	Male	Farmer	7	Y	0	Minor. The family has survived	Y
11	Shikha Muley	25	Female	Housewife	6	Y	2	Major. Collision with a truck	Y
12	Jayesh Sone	32	Male	Employed	3	Y	0	Members unaffected	Y
13	Balaji Mali	46	Male	Farmer	5	Y	0	Unaffected	Y
14	Sarang Naik	28	Male	Labourer	8	Y	2	Minor. Few Bruises	Y
15	Alok Raje	35	Male	Employed	5	N	0	Unaffected	N
16	Arzoo Balphatak	18	Female	Domestic Help	5	Y	0	Unaffected	Y
17	Shefali Nene	28	Female	Employed	3	N	0	Unaffected	Y
18	Pawan	25	Male	Dabba Supplier	5	N	0	Unaffected	N
19	Piyush Raja	26	Male	Employed	4	N	0	Unaffected	N
20	Shantanu Sinha	28	Male	Employed	1	N	0	Unaffected	N
21	Laxam Doke	21	Male	Student	3	Y	1	Minor. The family has survived	Y
22	Daksha Gajare	61	Female	Retired	6	N	0	Unaffected	Y
23	Sejal	54	Female	Housewife	4	N	0	Unaffected	Y
24	Tirth Vispute	32	Male	Employed	3	Y	2	Minor	Y
25	Yogesh Damle	22	Male	Student	5	Y	1	Minor	Y
26	Akhila Kaur	52	Female	Housewife	3	N	0	Unaffected	Y
27	Vishal Takle	14	Male	Student	9	N	0	Unaffected	N
28	Ajit Chitre	30	Male	Employed	4	Y	0	Unaffected	Y
29	Paritosh Bapat	46	Male	Employed	4	Y	1	Minor	Y
30	Gaurav Kulkarni	34	Male	Employed	3	N	0	Unaffected	N
31	Divya Kale	21	Female	Student	1	N	0	Unaffected	Y
32	Harsh Mishra	33	Male	Worker	3	Y	0	Unaffected	Y
33	Hemant Grover	45	Male	Employed	5	N	0	Unaffected	N
34	Mohit Sone	63	Male	Retired	2	Y	1	Minor	Y
35	Richa Bhunja	26	Female	Employed	1	N	0	Unaffected	N
36	Hemraj Nene	21	Male	Student	6	N	0	Unaffected	N
37	Karuna Mhatre	50	Female	Employed	9	Y	2	While crossing road	Y
38	Jyotika Aadekar	37	Female	Employed	4	N	0	Unaffected	Y
39	Shikha Sharma	49	Female	Domestic Help	5	N	0	Unaffected	N
40	Rakesh Chavan	21	Male	Student	4	Y	0	Unaffected	Y
41	Omkarnath Patil	36	Male	Employed	3	N	0	Unaffected	N
42	Chaya Lal	42	Female	Labourer	5	N	0	Unaffected	N
43	Bhakti Rane	56	Female	Housewife	5	N	0	Unaffected	N
44	Deepti Nayak	34	Female	Employed	2	Y	0	Minor	Y
45	Radhika Singh	19	Female	Student	3	N	0	Unaffected	N
46	Pranavesh Naik	32	Male	Labourer	4	N	0	Unaffected	Y
47	Adil	33	Male	Worker	4	Y	0	Unaffected	Y
48	Shashank Shetty	36	Male	Worker	5	Y	1	Minor cuts	Y
49	Darpak Marane	46	Male	Employed	9	Y	3	Major	Y
50	Manna Nimkar	50	Male	Employed	5	N	0	Unaffected	N
51	Sukriti Dhole	25	Female	Employed	1	N	0	Unaffected	N
52	Pratiti Patel	29	Female	Unemployed	3	N	0	Unaffected	Y
53	Nafi Wagh	35	Male	Labourer	6	Y	0	Crash	Y
54	Ram Prasad	41	Male	Worker	3	N	0	Unaffected	Y
55	Shakuntala Kulkarni	20	Female	Student	1	N	0	Unaffected	N

Table 19 Database of the survey conducted

13.4 Statistical Representation

The following statistical diagrams were prepared from the surveyed conducted by us.

13.4.1 Accident Ratio

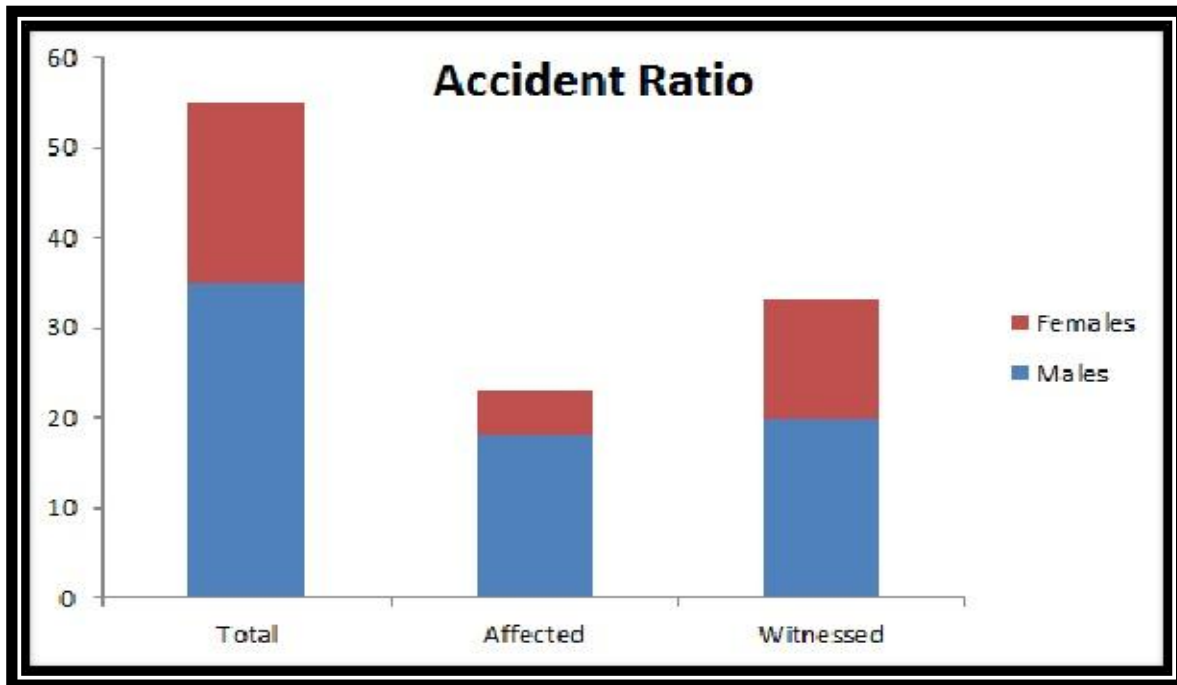|	**Males**	**Females**
Total Number	35	20
Affected Number	18	5
Witnessed Number	20	13

Table 20 Accident Ratio between the people surveyed

Figure 102 Accident Ratio Representation

The above table and graph gives us a relation between the ratio of the male and the female population, who were surveyed in that stretch of road. It gives us an idea about the total number of people who were affected by accidents and the number of people who witnessed accidents as well. The total column gives us the total number of people who were surveyed along the Thane-Ghodbunder Road. They have further been segmented into males and females which give us a visual effect in understanding the data.

13.4.2 Visibility on Road

Excellent	Good	Fair	Poor
5	13	20	17

Table 21 Accident occurrence depending on road visibility

Table 22 Accident occurrence representation based on road visibility

The above graph gives us an idea of the visibility condition of the road and according the accidents have occurred along the area surveyed – Thane Ghodbunder Road. From the graph we can see that, the maximum road accidents have occurred when the road visibility was fair and is closely followed by poor visibility conditions.

Fair visibility results in about 36% of the accidents (from the survey conducted) and the poor visibility condition accounts for almost 31% of the accidents. It has also been seen that even globally, most number of accidents occurs due to fair to moderately fair visibility. That very trend has also been seen along this Thane-Ghodbunder Road stretch.

13.4.3 Time of Day of Occurrence

Daylight	Dusk	Night
10	15	30

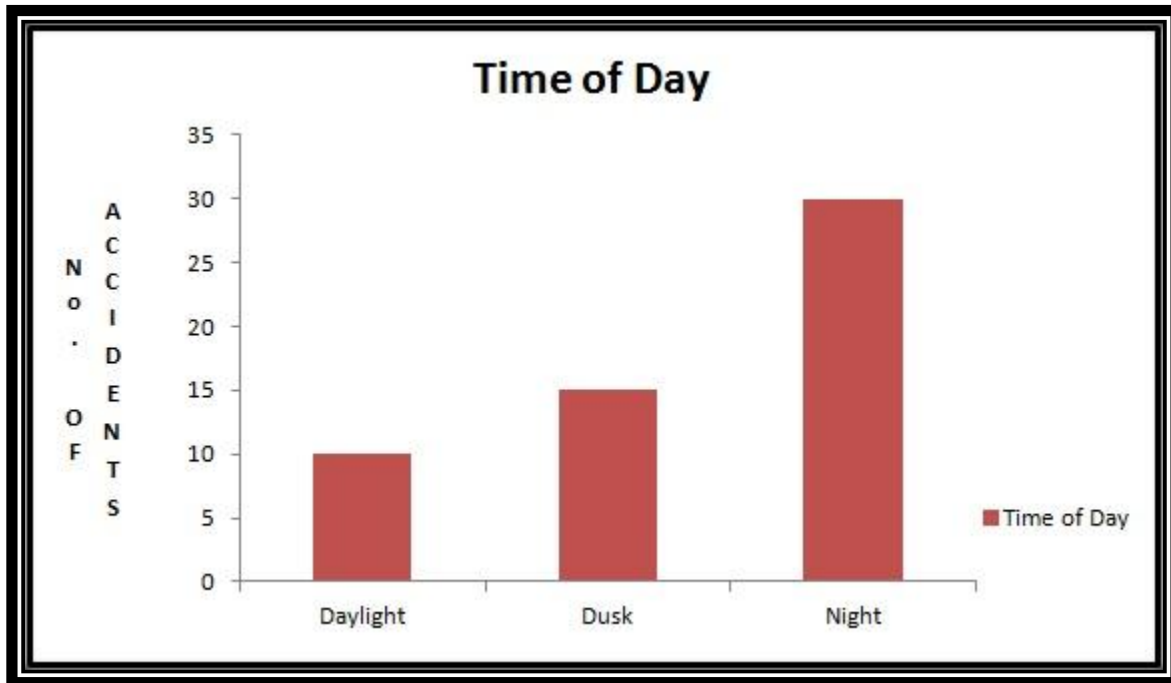Table 23 Accident occurrence on the basis of time of day

Figure 103 Time of day affecting the accident occurrence

It is a well-known fact that, most of the accidents occur during the night. It has been the same for this area as well. Almost 55% of the total accidents occur at night. It is indirectly related to the visibility conditions. Daylight provides with a better visibility as compared to dusk or night. Many times a foggy or a misty weather during the day or night may give rise to increasing number of accidents.

This graph shown above is based on the survey we had undertaken along the Thane-Ghodbunder Road. We had surveyed 55 people and these graphs are based on the answers that they had provided to out questionnaire.

13.4.4 Road Conditions

Dry	Damp	Wet
16	5	34

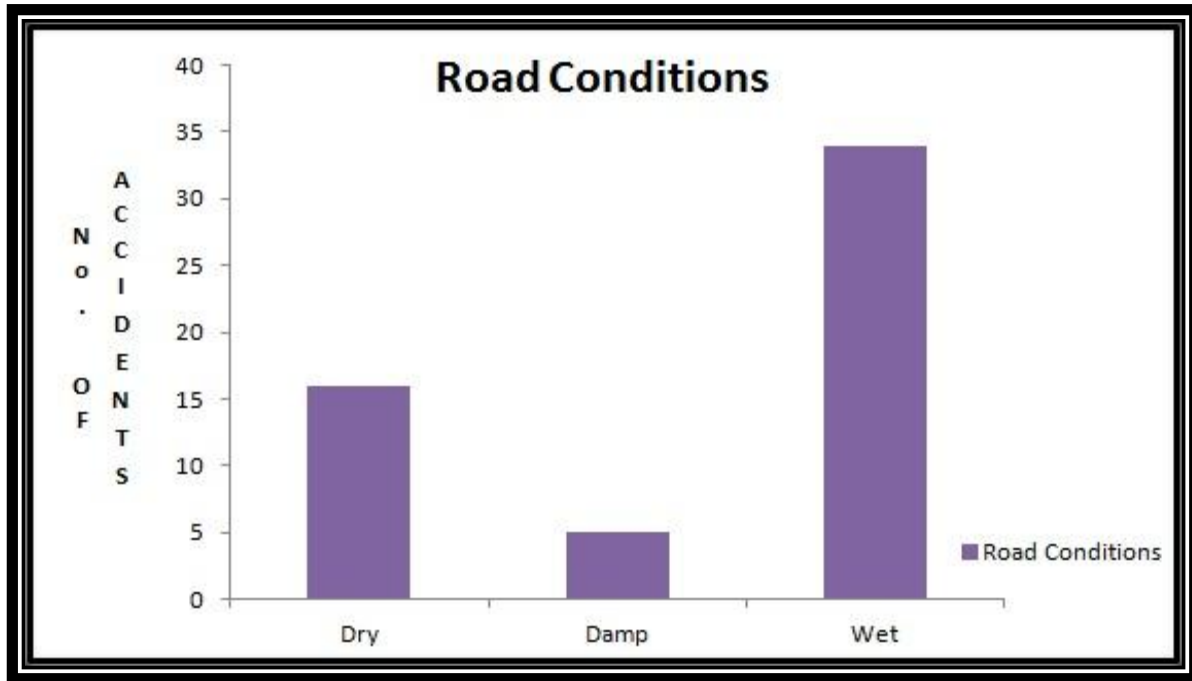Table 24 Road conditions affecting accidents

Figure 104 Accident occurrence based on road conditions

Road conditions also have a direct relation with the rate or number of occurrence of accidents. Monsoons or the rainy seasons, witness the maximum number of accidents.

In the survey undertaken, 55 people were surveyed and based on their response this column bar has been constructed. From the table and the graph it can be seen that, more than 60% of the people surveyed have said that most of the accidents which has occurred along the Thane-Ghodbunder Road is a result of wet road conditions. Damp road conditions also do result is accidents but the number is even less than 10%.

As seen, dry road conditions also results in accident occurrence which is about 30%. It is because of the uneven road surfaces or potholed and jaggy surfaces that cause most of these accidents. So, we can see that both the dry and wet road conditions do affect the occurrence or rate of accidents.

CHAPTER 14 – WEB INTERFACE CREATION

A website is a term that describes a collection of webpages, media files and other digital assets that are combined in the URL. The importance of a website is determined by many reasons. It is sometimes the desire and goal that can be achieved through a website. In our case, we have mainly been dealing with accident sites along the Thane-Ghodbunder Road.

The purpose of us designing this website is to provide knowledge to the people who have been travelling along this stretch of road. Not only that, this website pre-forecasts the occurrence of accidents and also suggests the areas mostly prone to accidents. The information provided is greatly supported by statistical data which has been gathered and also the process of surveying where the people travelling along this section of the road were reviewed.

The website also gives an idea to the hospitals and police stations that are provided around the accident sites at a close proximity with 5km as its radius. The first thing the needs to be done before creating a website are to plan what elements and features that needs to be incorporated. Not just that, this website also provides the information but also displays maps of the area that has been selected at the click of a button. It also displays the important POIs (Hospitals and Police Station) along that road with their area of coverage. The maps can be viewed in not just in the normal map layout view but the terrain and the satellite view can also be observed.

The map also displays the routing between two points selected on the map. Not just that, the map also displays information bubble about the location that has been selected. If desired, it can also zoom in to the point selected. This is a map which is also dynamic in nature and interacts quite easily with the user. To put it simply, the website is user-friendly in nature and can be quite easily used by the people.

The designing of the website, which displays the map, has been done using HTML, Javascript and CSS.
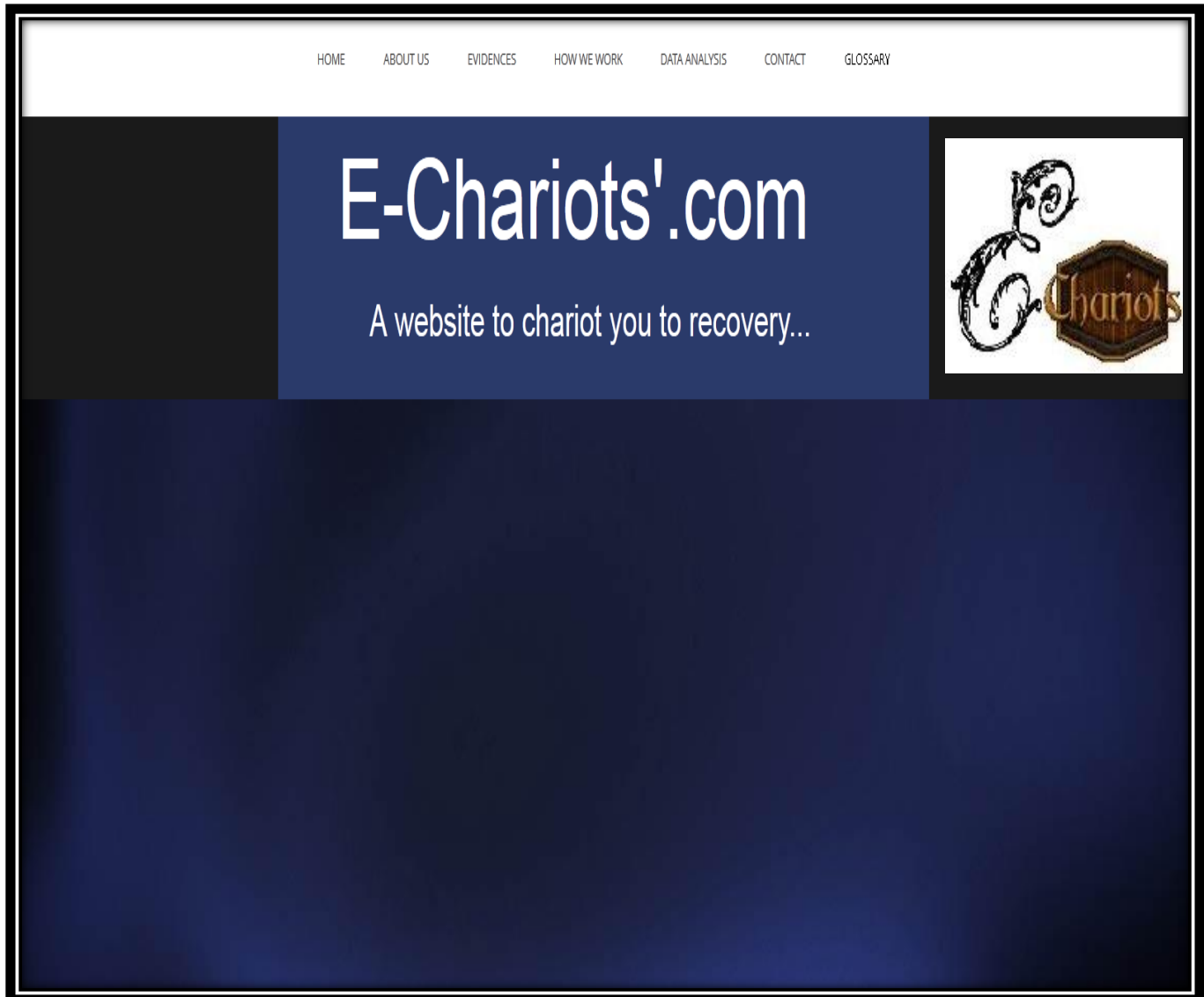
14.1 WIX

14.1.1 The MAIN Interface

Figure 105 The MAIN INTERFACE of the informative website
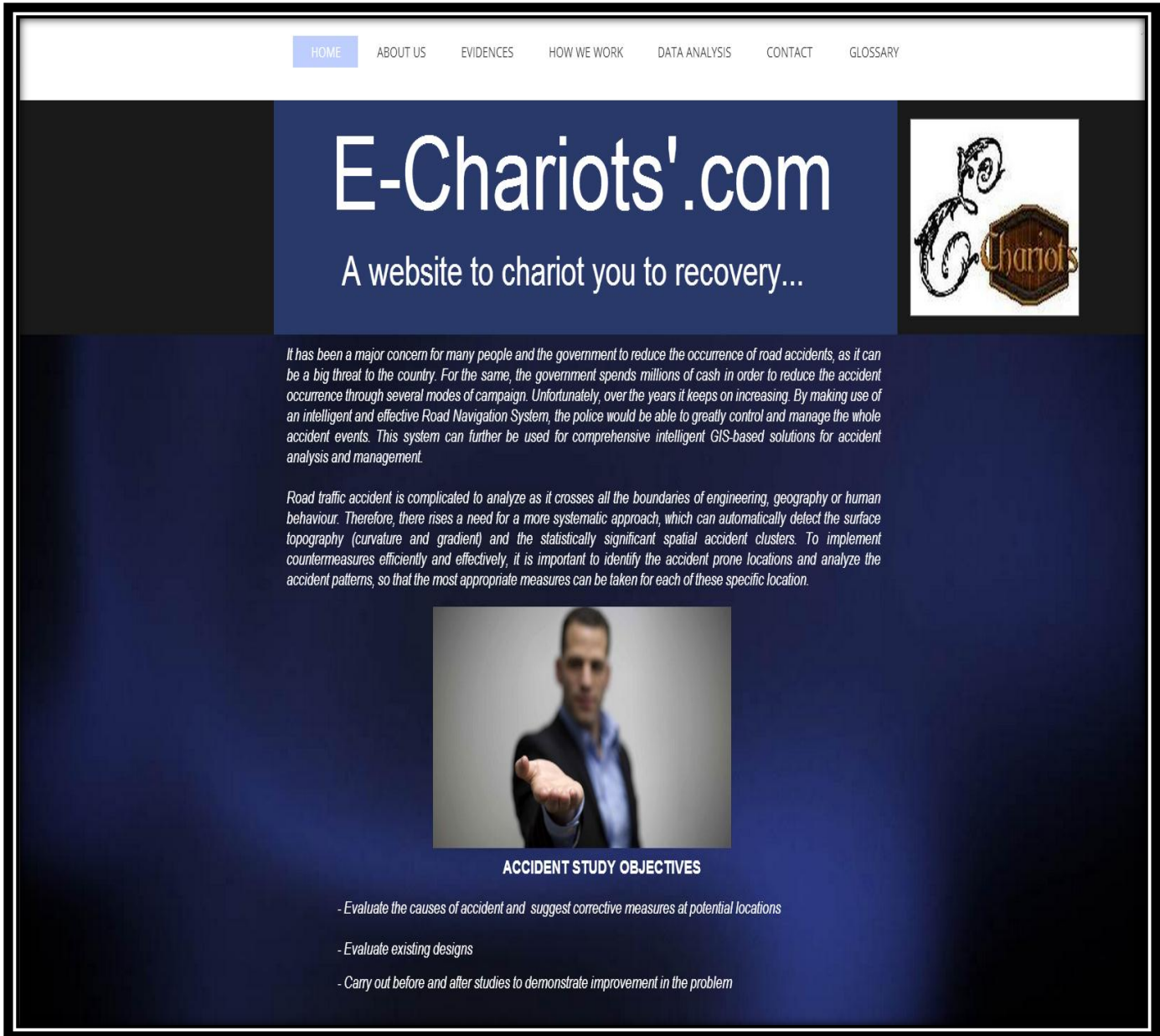
14.1.2 HOME

Figure 106 The HOME page

This is the main interface of our website, titled "E-Chariots'.com". This shows us the snapshot of all the tabs that are present on this page. Each of these tabs has further drop downs helping us to navigate to the page of our choice. This is the Home page and it mainly talks about accidents and the objectives that have been undertaken for this project on accident study.
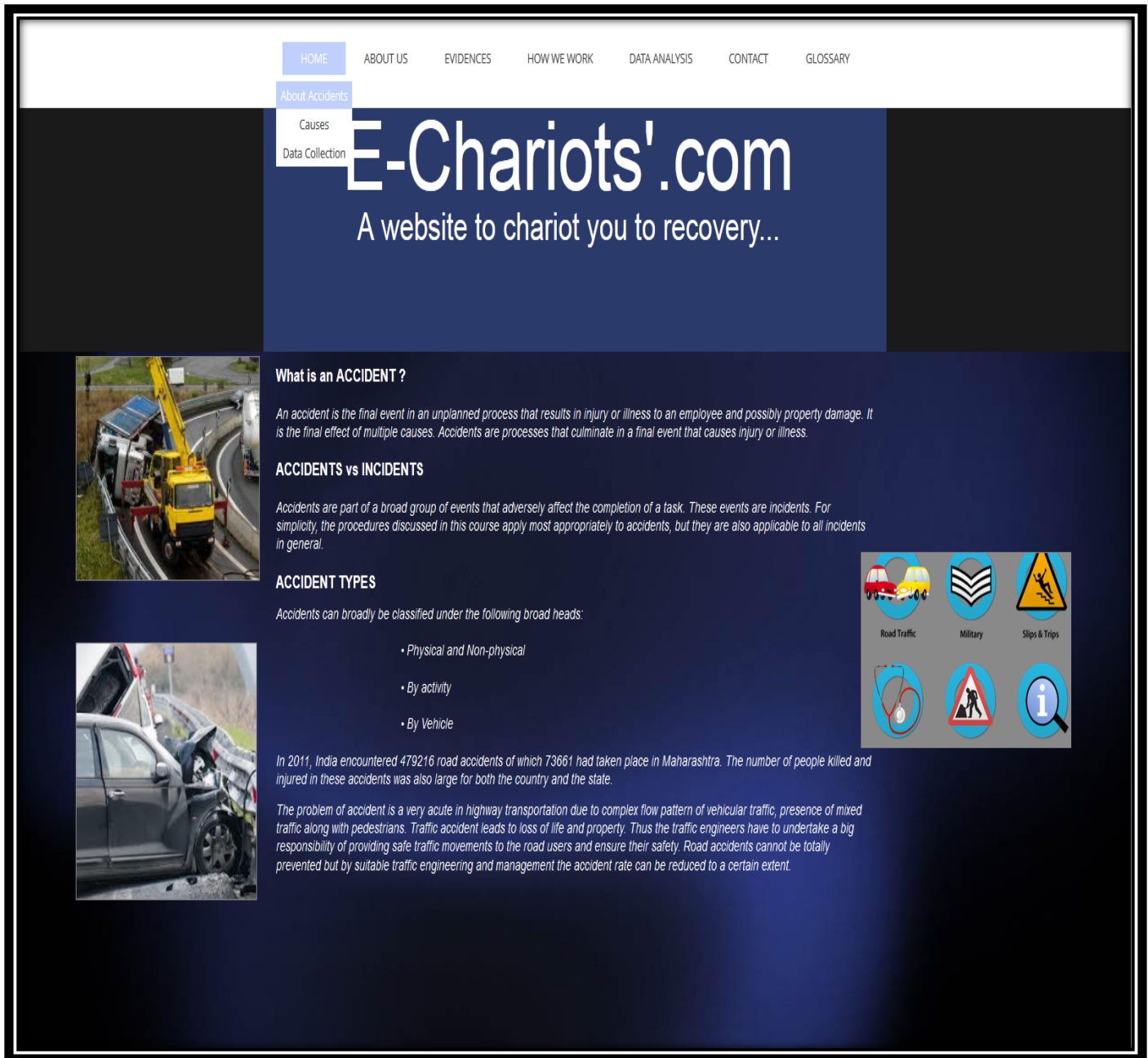
14.1.2.1 About Accidents

Figure 107 ABOUT ACCIDENTS page

The HOME tab is further sub-categorized into smaller parts. The first broad head under this tab is "About Accident". This gives us a detail about accidents, its definition and the different types of accidents that are commonly found. This also gives the accident statistics for the entire country and also gives its ratio with the state.

14.1.2.2 Causes

Figure 108 CAUSES page

The next thing under the HOME tab is the causes of accidents. It is pointless talking about accidents, if the factors or the causes of its occurrence are not mentioned. It is broadly classified in three different types and each of them has a further classification. They have been shown in different columns for a better understanding along with images.

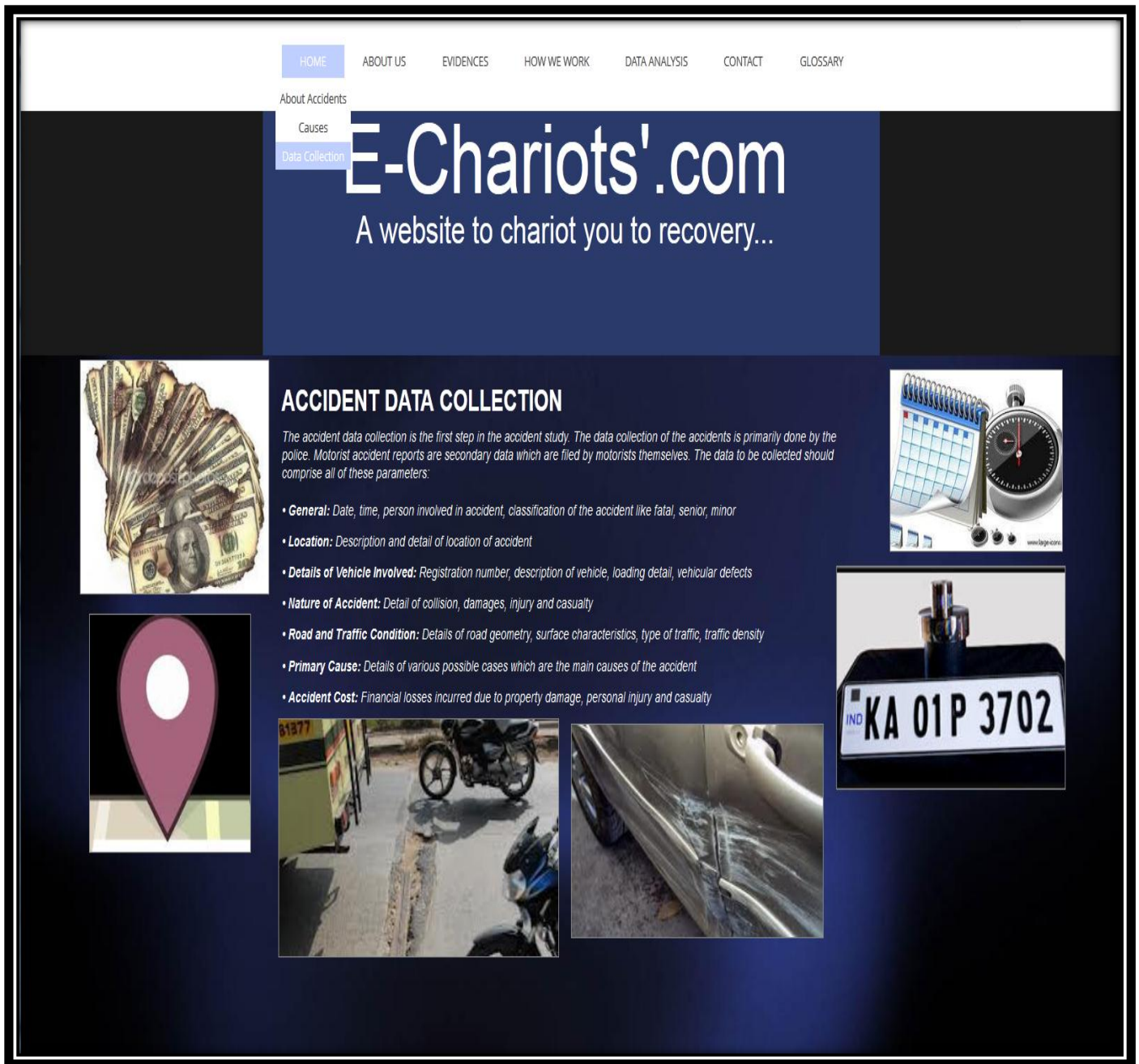*14.1.2.3 Data Collection*

The next thing that is of prime importance is the Accident Data Collection. After an accident occurs the important information that are required are provided in this tab. This talks about the importance data that needs to be collected / obtained from a person who has faced an accident.

14.1.3 ABOUT US

E-Chariots'.com
A website to chariot you to recovery...

"To realize the value of one second; ask a person who has survived an accident".

In 2011, more than 20,800 people of Mumbai had lost their lives in road accidents. In Maharashtra, there is a highway related death every 15-20 minutes and an injury due to the same in every 5 minutes. Highway crash related deaths and injuries are a major concern for the government, despite the progress that has been made in the past 40-50 years to improve the transport safety through safer vehicles and roadsides and more responsible drivers. Although there has been a drop in this fatality rate but it still remains an on-going challenge for us to do more to improve this safety.

Over the years there has been a tremendous change in the mobility habits. Culture, economy and technology has built a new society based on a new pattern of accessibility, integration of various networks, faster and more efficient and effective transportation and has resulted in changes in the way we animate and move.

Automobiles have become a necessity from luxury and have thus resulted in a rapid increase in road accidents. The transience rate has reached unbelievable numbers in the recent time and is of utmost importance in terms of projects and research purpose.

An automotive navigation system is a satellite navigation system designed for use in automobiles. It typically uses a GPS navigation device to acquire position data to locate the user on a road in the unit's map database. Using the road database, the unit can give directions to other locations along roads also in its database.

As world is grooming, new and versatile technologies and techniques in navigation system are budding. The year 2012 has been marked very many novelty functionalities being incorporated in the existing navigation system, so that it can minimize the probable occurrence of these accidents.

The spatio-temporal analysis of road accidents aims for the identification of statistical relationships between accidents and their occurrence in our AOI, Thane-Ghodbunder Road. So, by using integrated statistical analysis techniques in GIS, the data has been analysed, as well as an insight has been obtained to how this are distributed in the specified region.

Hence, by means of spatial statistics the creation of a pertinent data analysis was possible, to evaluate the tendencies and relationships between the existing data.
The aim of this study is to outline the importance and efficiency of using GIS in collecting, processing and analyzing data.

It is an efficient tool for displaying the spatial distribution of accidents along the road network. GIS enables the rapid processing and display on the map of the information regarding accidents and its use in making decisions regarding the reduction and prevention of accidents.

The current scenario of Mumbai has limitations when analysing crashes on multilane roads such as motorways (highways). The accomplishment of reducing these fatalities and injuries can greatly be achieved through a combination of efforts, including a proper use of road design, improving the driver's perception and their willingness to obey the safety norms.

Despite these improved efforts the annual toll of loss of human life has not changed substantially in all these years. Vector GIS has been used in displaying these accident spots. Furthermore, raster GIS was used to analyse this spatially. It started with mapping and georeferencing the map within the limits of the Mumbai-Thane district.

Figure 110 ABOUT US page

This is the page that opens on the click of ABOUT US. It talks all about accidents that have occurred in Maharashtra and also along our area selected. It also discusses the current scenario of Mumbai and its surrounding regions. It also discusses the use of GIS in this project and how it helps in dealing with accidents.

14.1.3.1 Geographical Location
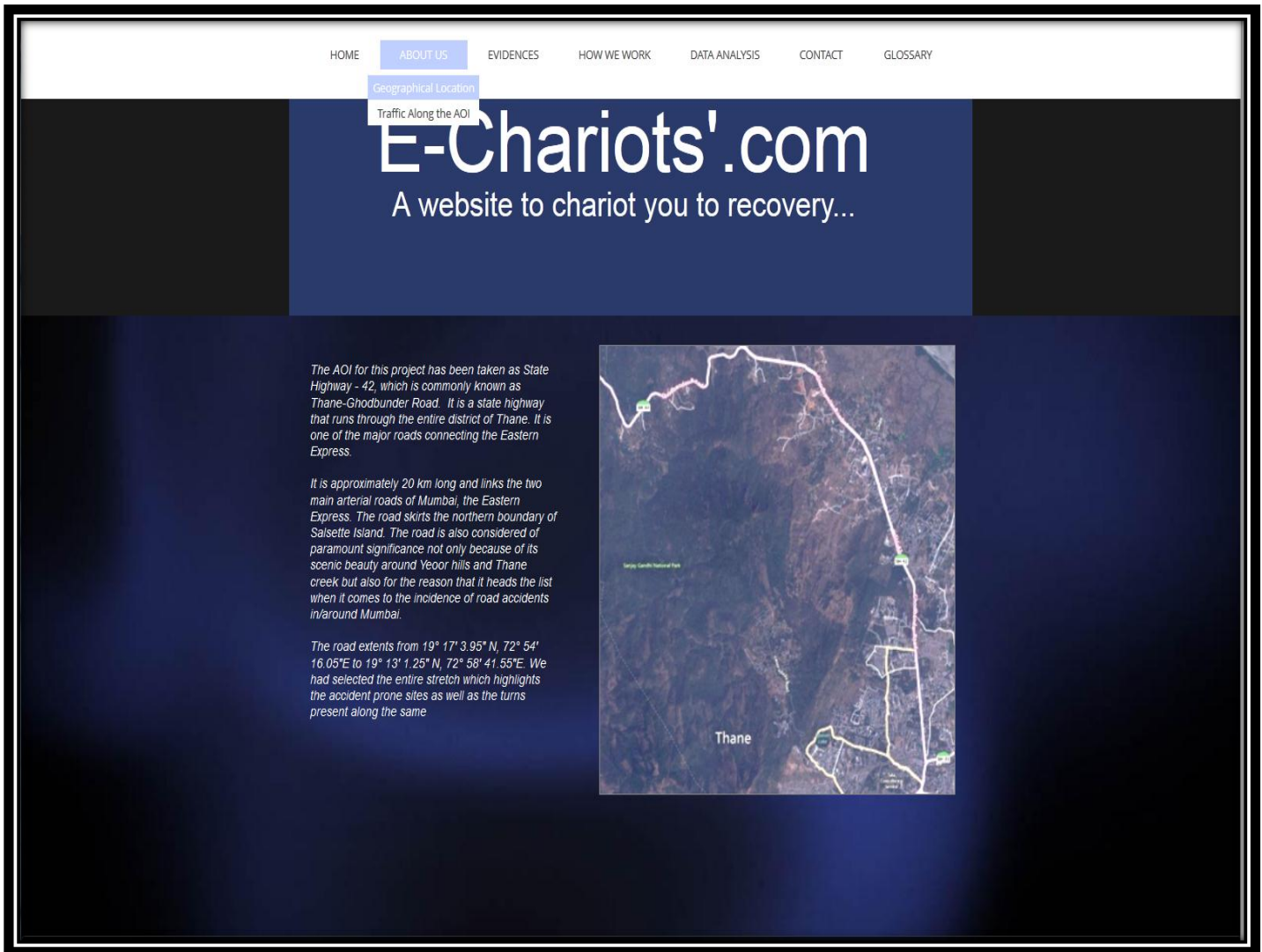
<div align="center">Figure 111 GEOGRAPHICAL LOCATION details</div>

The first sub-header tab that can be seen below the ABOUT US tab is the Geographical Location. On the click of this tab, it displays the locational co-ordinates of the AOI that we have selected. It also gives us the length of the entire road stretch that we have covered in this project. This road is also the State Highway and it traverses through the entire Thane district and is indeed an important network. It is also considered important because it connects NH-8 (Mumbai Ahmedabad Highway) to the Eastern Express Highway.

14.1.3.2 Traffic along the AOI

Figure 112 Page showing TRAFFIC ALONG THE AOI

The next sub-header is Traffic Along the chosen AOI. We are talking about accidents and it is quite important to mention the traffic along the specified network. This gives us the total vehicle population and the tax revenue that is collected from this source. It also tells us that the more than 80% of the total vehicle population consists of two-wheelers and cars. Over the last 5 years, these figures have been increasing greatly and soon the vehicular density will worsen the state of the Thane-Ghodbunder Roads.

14.1.4 EVIDENCES
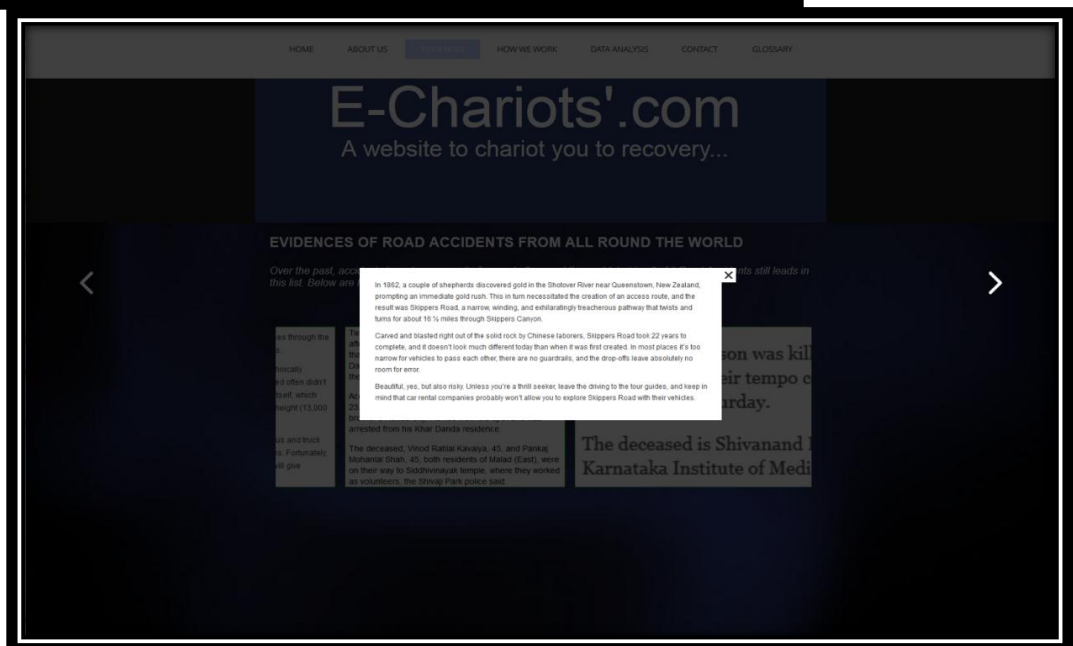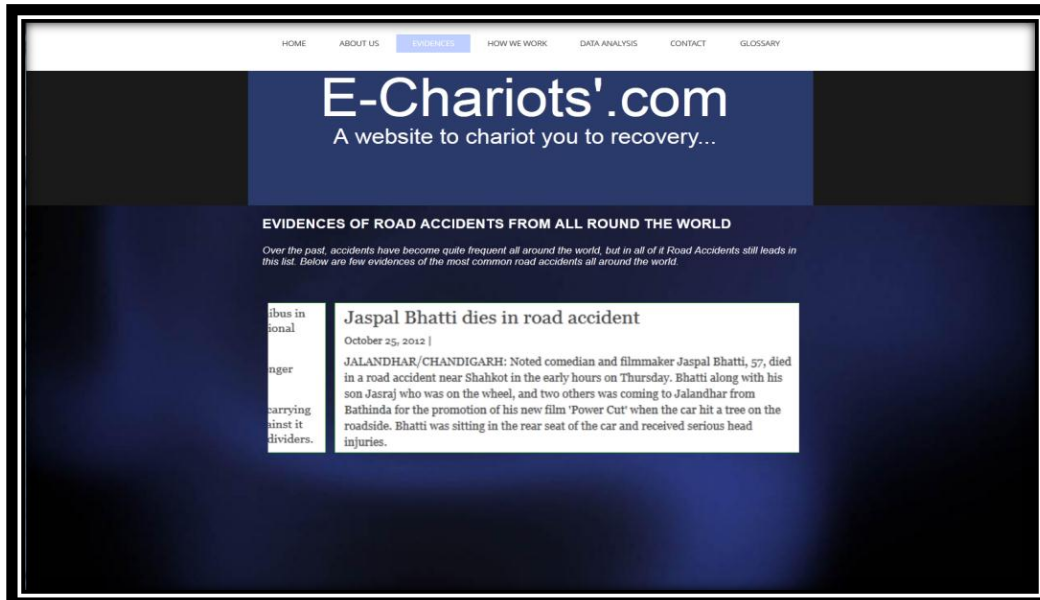
Figure 113 EVIDENCES shown in a slideshow

This is the evidence tab, which gives us the list of 18 road accidents that has taken place all over the world. It ranges from India to Pakistan to New Zealand, Nepal and also America.

It has also been seen that India has the highest number of road accidents in the world. With 130000 deaths annually, the country has overtaken China and is now has the worst road traffic rate worldwide.

14.1.4.1 Evidences to Support the Selection of AOI

Figure 114 EVIDENCES TO SUPPORT THE SELECTION OF AOI

This lists the accidents that has occurred along the Thane-Ghodbunder Road and hence providing us with indications for the selection of the AOI. To add, Thane-Ghodbunder Road is the top accident prone road in the Navi Mumbai region.

14.1.5 HOW WE WORK

Figure 115 SNAPSHOT SHOWING THE WORKING

14.1.6 DATA ANALYSIS

Figure 116 DATA ANALYSIS page

This tab states about the Data Analysis that has been carried out by us for this project. Analyzing of data refers to the process of inspecting, cleaning, transforming and modeling of data highlighting the useful information and conclusions that can be drawn from the given set.

In this case, we have performed four broad groups under the data analysis head. These are: Hotspot Analysis, Network Analysis, Geostatistical Analysis and Statistical Analysis.

14.1.6.1 Hotspot Analysis
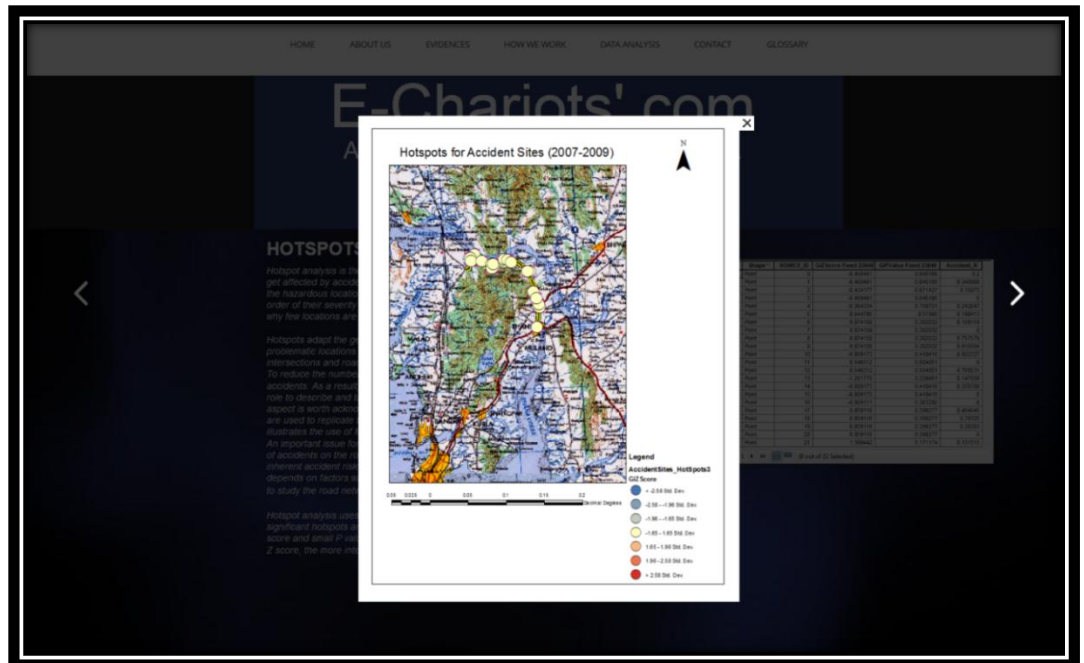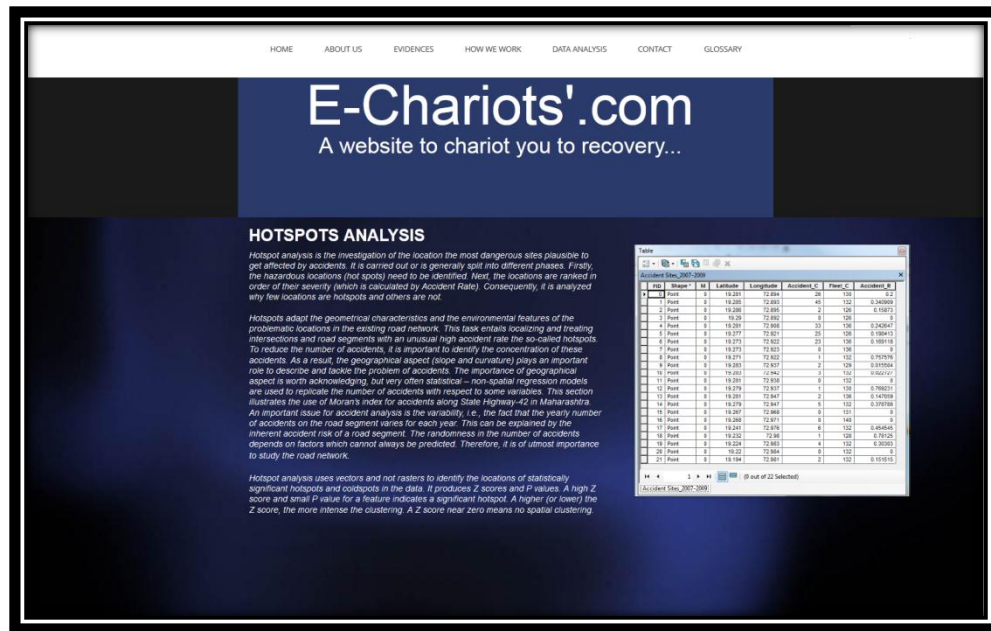
Figure 117 HOTSPOT ANALYSIS snapshots shown in the form of slideshow

This tab talks about the details of Hotspot Analysis along with the snapshots of the flow of it. It shows us the details of each of the attribute of the accident sites that was created for this. It also gives us the p-value and z-score of the selected data. Lastly, it displays a map layout of different hotspot regions.

14.1.6.2 Network Analysis

Figure 118 NETWORK Analysis Snapshot

The Network Analysis tab mentions about the entire important network that is present along the Thane-Ghodbunder Road. It also shows the location of the important POIs (Hospitals and Police Stations). The purpose of the creation of network datasets and service roads is to show the area covered by each of these POIs in case there is any accidents along the road stretch.

The service areas have been created with a radius of 5km and can cater to the needs of accidents occurring within that stretch.

14.1.6.3 Geostatistical Analysis

Figure 119 GEOSTATISTICAL ANALYSIS page

This is the tab which deals with Geostatistical Analysis. Geostatistical implies a subset of statistics that specializes in analysis and interpretation of geographically referenced data. Its main function is to determine the probability of certain variables occurring over an area and identifying the every possible location which in reality is impossible.

In this analysis, we have performed Kriging, Covariance Map, and Voronoi Diagram, identified the Global Trend Line and have also plotted the Cross-Covariance Cloud along with the Prediction Map.

14.1.6.4 Statistical Analysis
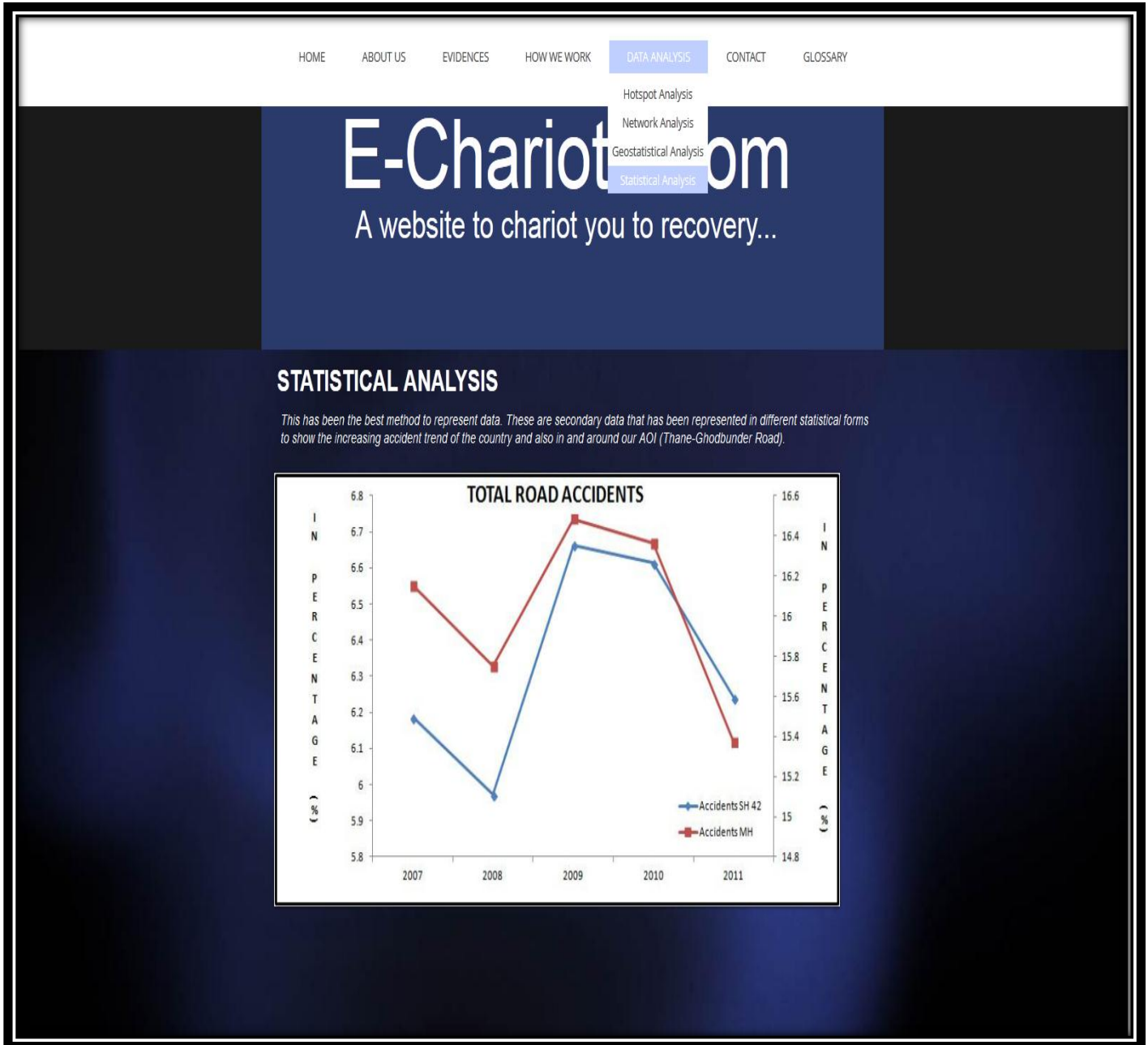
Statistical Analysis is considered as the best way to represent any kind of quantitative data. It ranges in the form of Line Graphs, Bar and Column Graphs and also Pie Diagrams. This helps us in comparing the data that is present with us. It also helps us in obtaining a desired trend of something that is unknown to us.

14.1.7 CONTACT
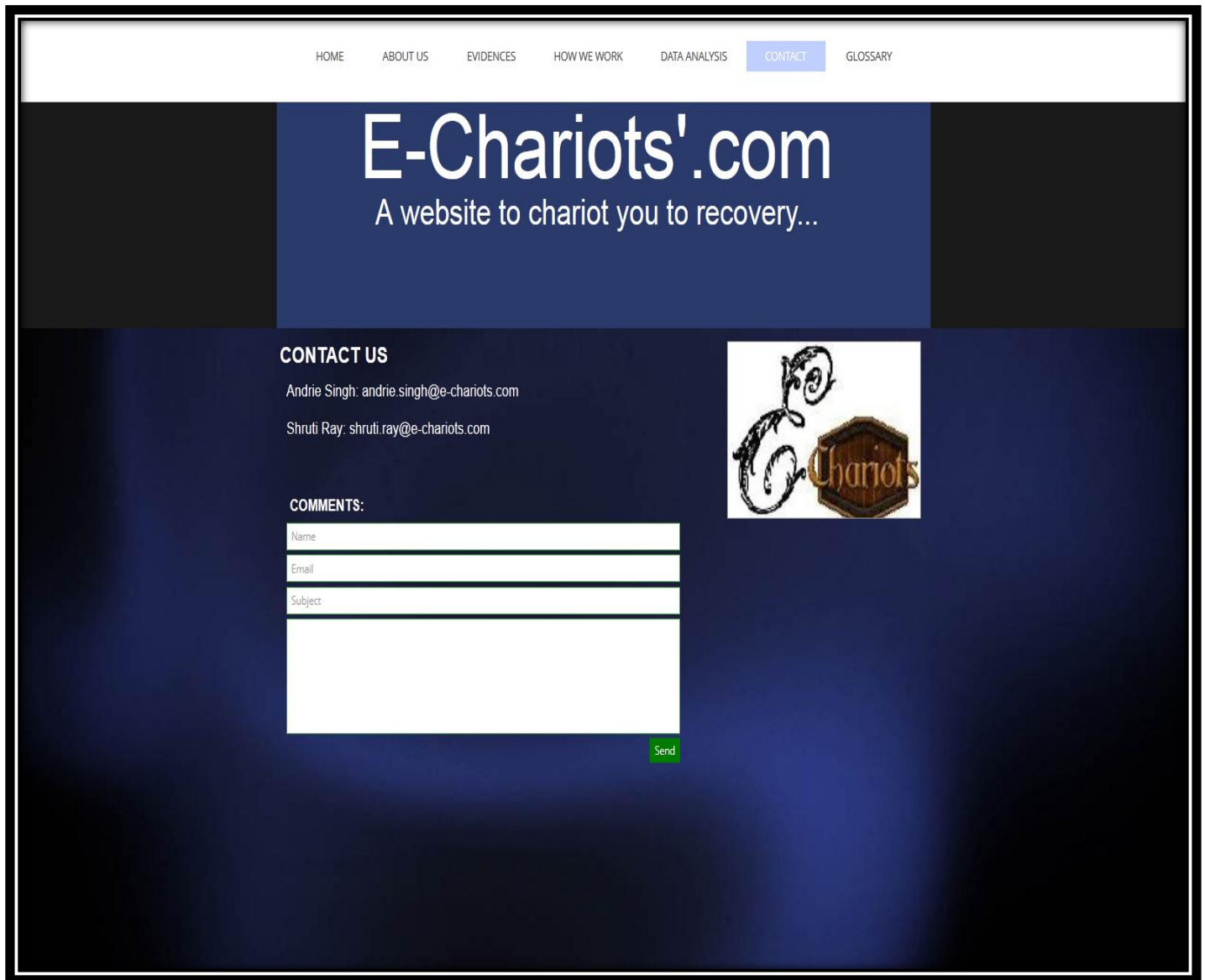
Figure 121 CONTACT page

This is the how the Contact Us page looks like. This provides the user with a means to connect to us. If he desires, he can talk about his/her problem and can easily get in touch with us.

14.1.8 GLOSSARY

It is sub-divided into two broad categories.

14.1.8.1 Software/Tools Used
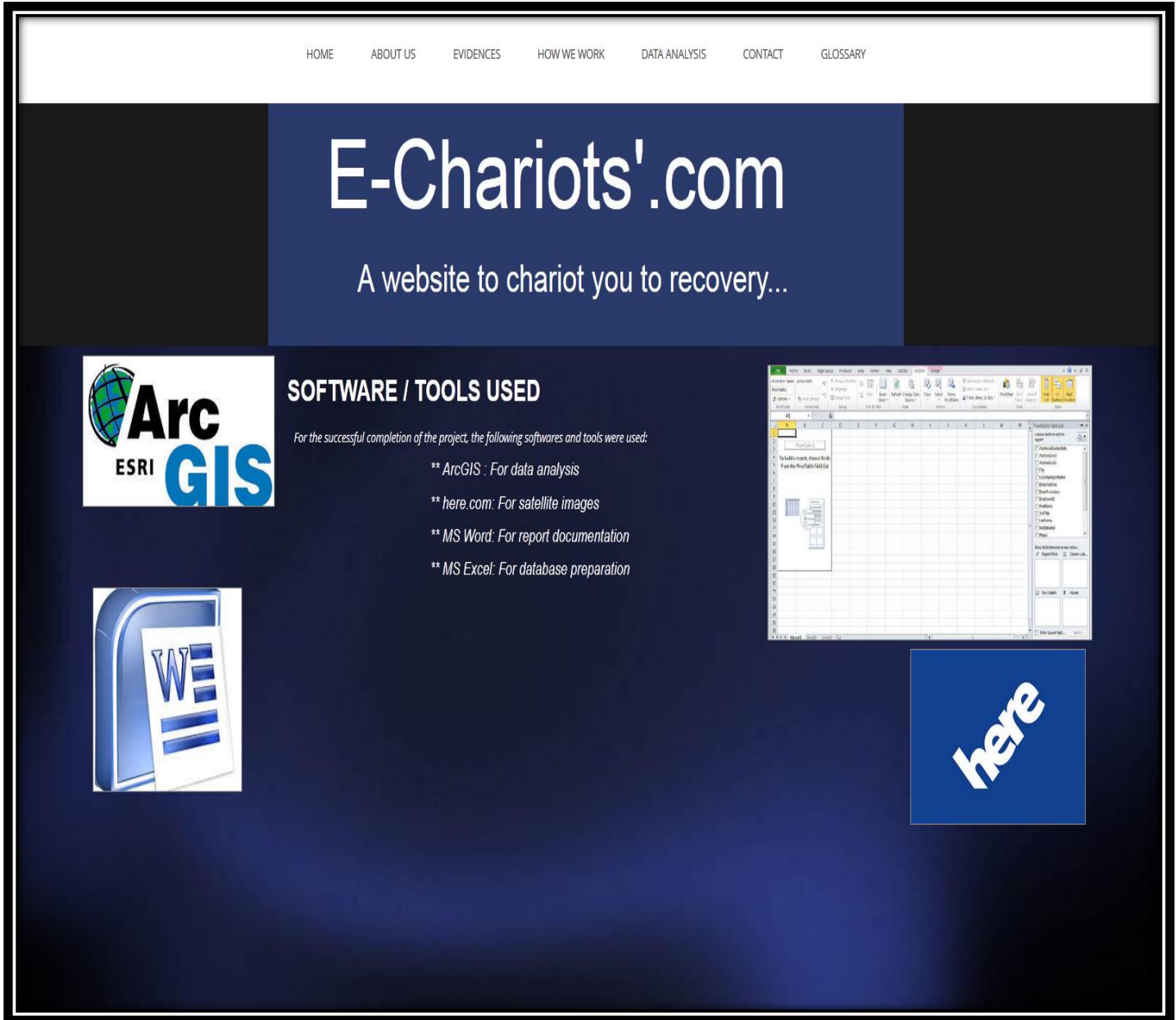
Figure 122 SOFTWARES/TOOLS used

For the successful completion of this project, the following softwares and tools has been used. ArcGIS has been used for accomplishing the data analysis. Maps that have been used are provided by "here.com". Creation of database has been carried out in MS Excel and documentation on MS Word.

14.1.8.2 Technical Terms

Figure 123 GLOSSARY elaborating the TECHNICAL TERMS

This talks about the technical terms which have been used by us for the completion of our project. It also gives us a basic idea about the algorithm that we have followed for displaying the map layouts.

14.2 HERE

14.2.1 The INTERFACE
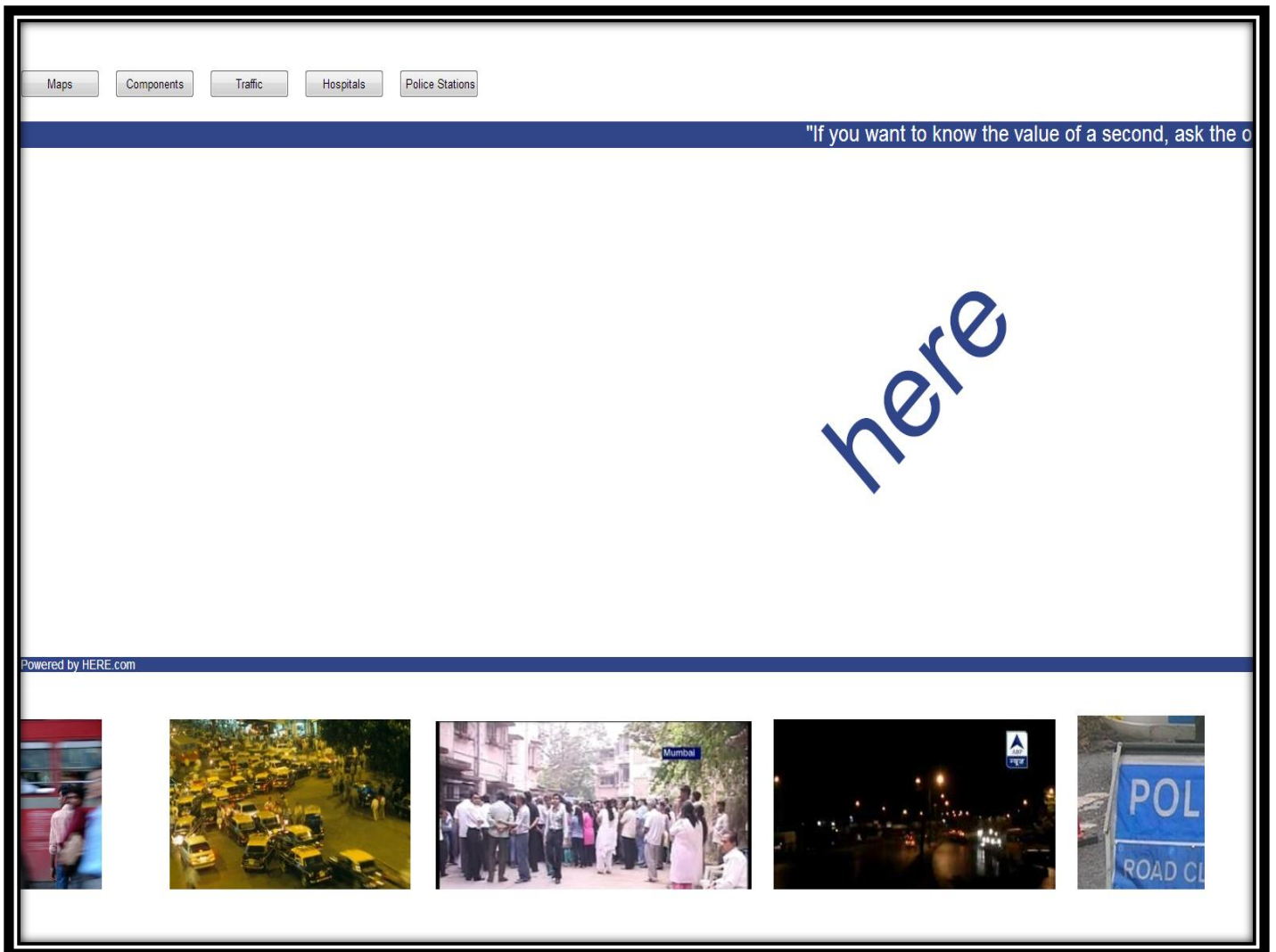
Figure 124 The Base Interface

This webpage displays the here maps and the all events related to that. Once this website is opened this is the display that can be seen on the screen. It displays five tabs. On the click of each of these tabs a different set of information is provided.

Also, there is a text marquee that runs through the top screen stating the tag line, *"If you want to know the value of a second; ask the one who has survived an accident"*. The bottom screen has an image marquee which displays images related to road accidents in Mumbai and its surrounding areas.
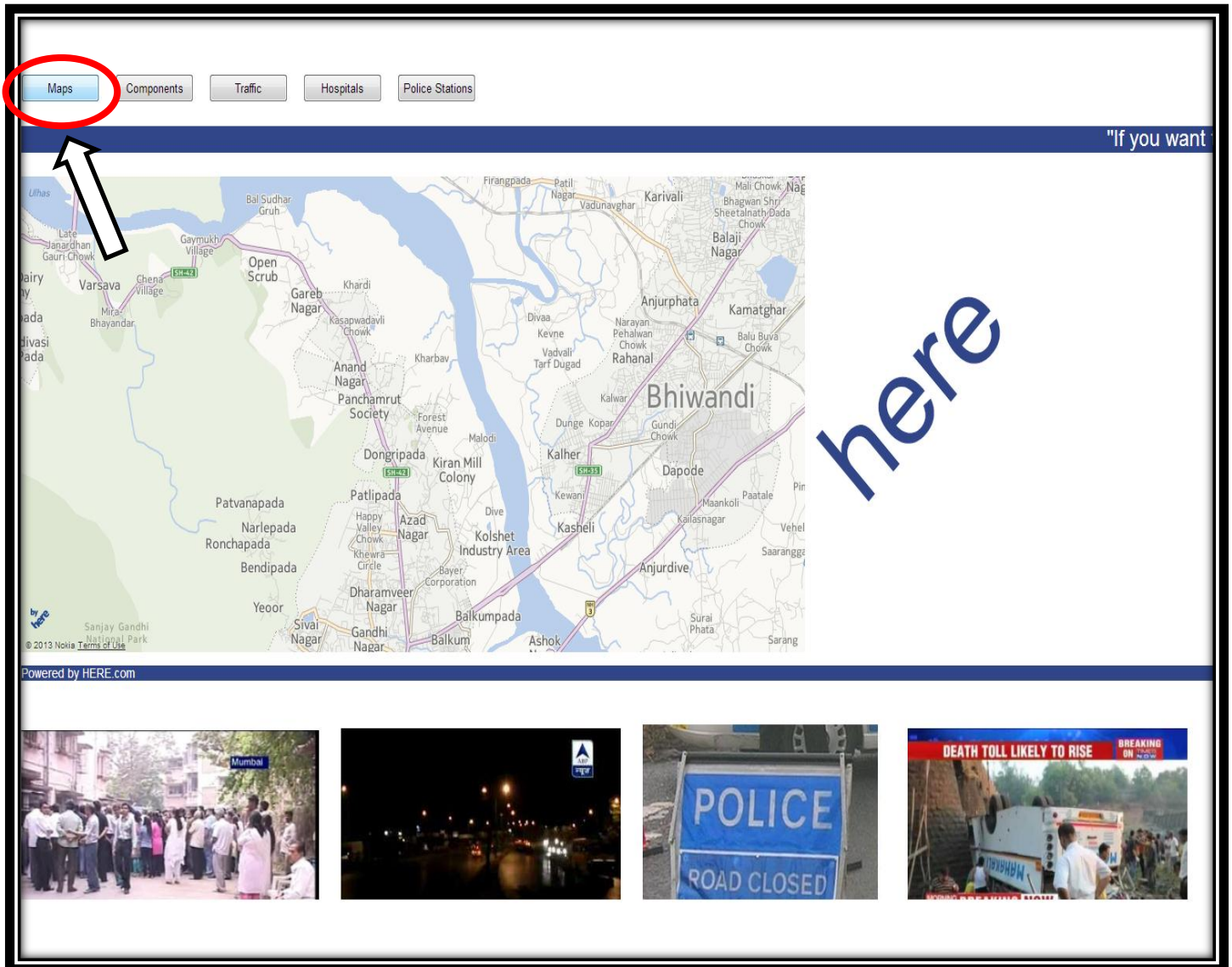
Figure 125 On the click of Maps

14.2.2 MAPS

The MAPS tab displays the map which shows the here map that directly displays AOI (Thane-Ghodbunder Road). The map can be zoomed in or zoomed out as per our needs. The map can also be panned depending on which areas we wish to view.

The marquees present (both the text and that of the images) can still be seen in this page.

14.2.3 COMPONENTS
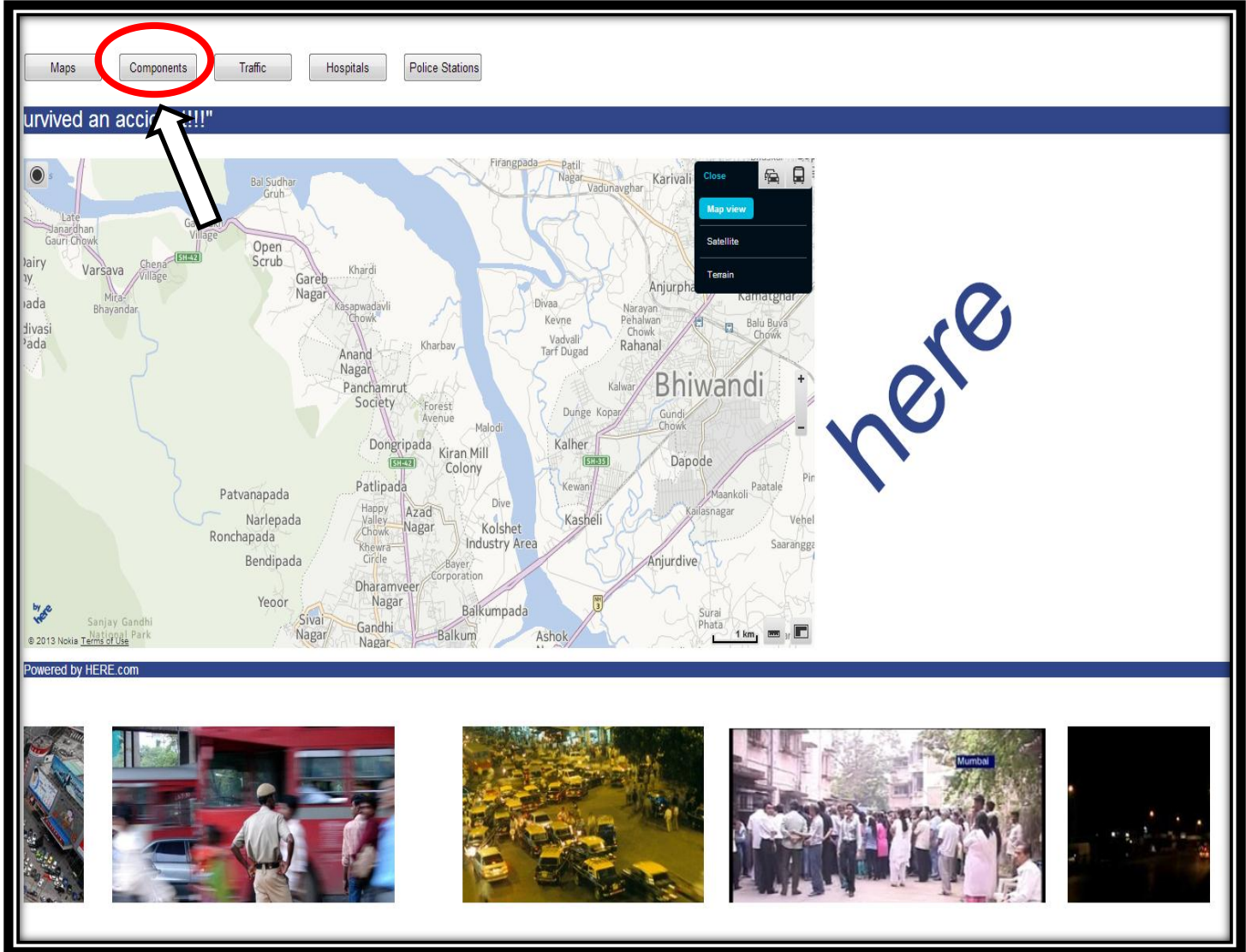
Figure 126 Click on Components

The COMPONENTS tab displays the different layers and switch to all these layers will be made possible. The layers include – Map View, Satellite View and Terrain View. This also displays the scale of the map and thus the distance can be calculated if and so desired.
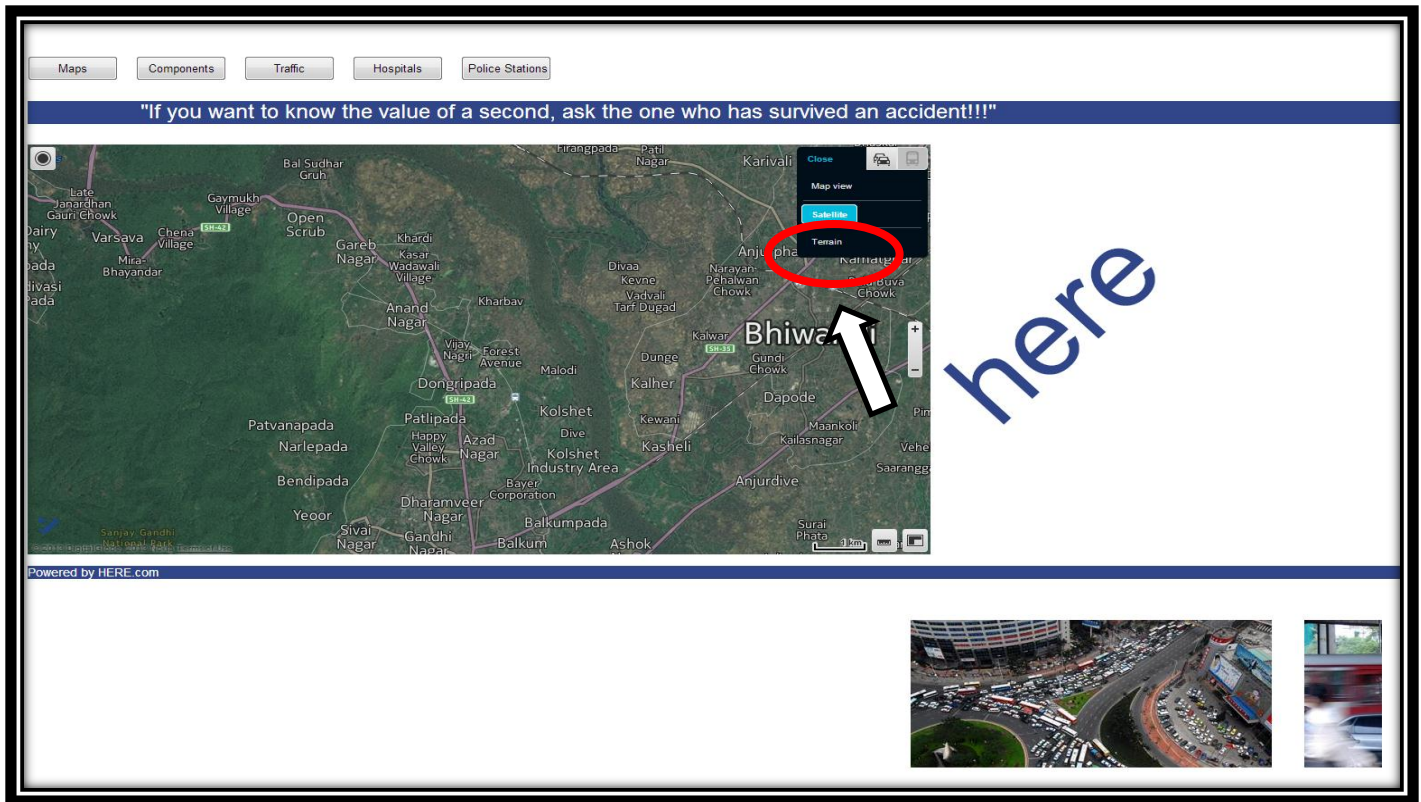
Figure 128 The Satellite View
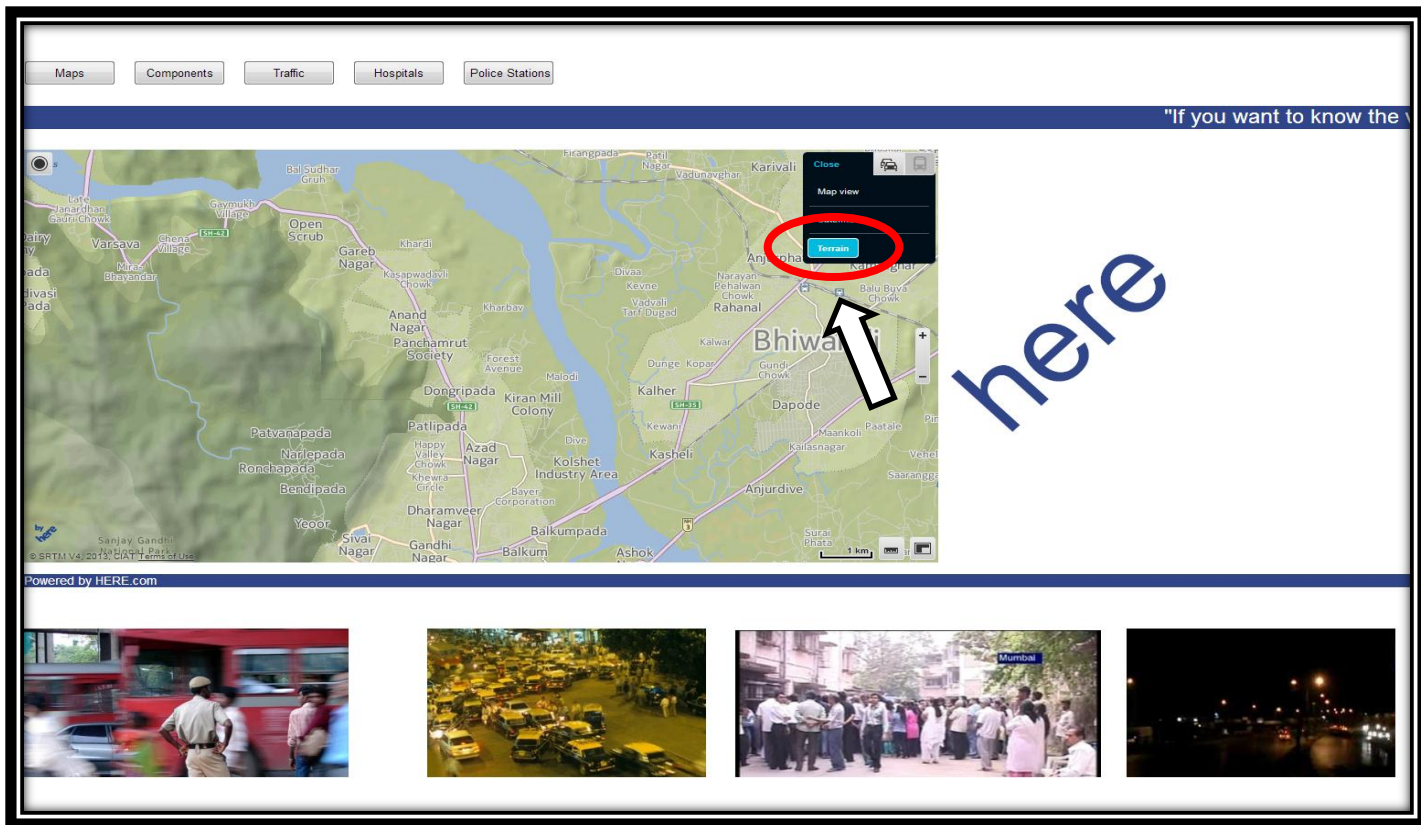
Figure 127 The Terrain View
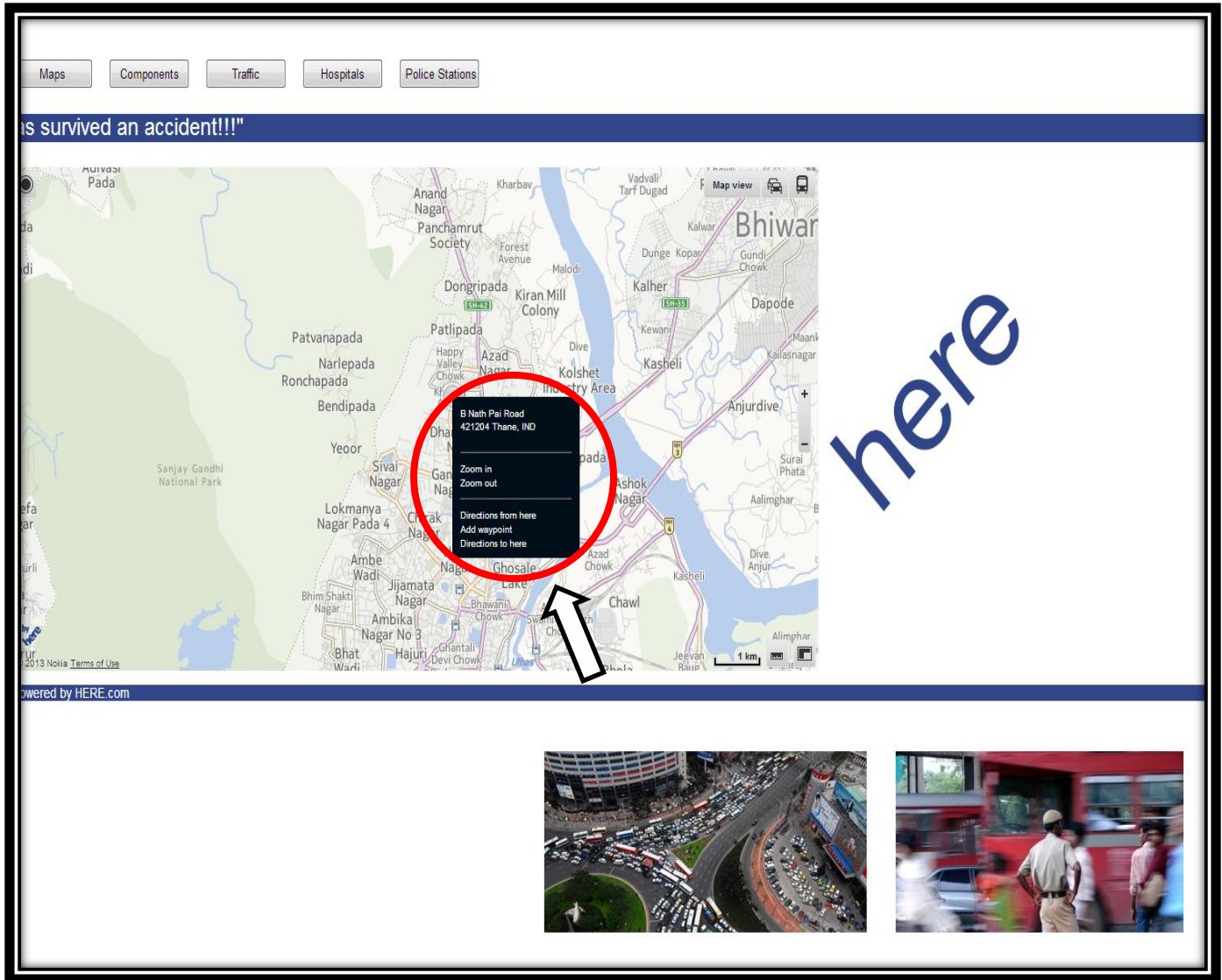
Figure 129 On Right Click

The above figure displays the information that can be seen on the right click. On right click a dialog box opens up which shows the exact location along with its zip code (if available). It also has the option of zooming in and zooming out.

There is also an option of adding a way point and calculating the route between two given point. Direction from here and direction to here, gives us a scope to show the shortest distance between two selected points.
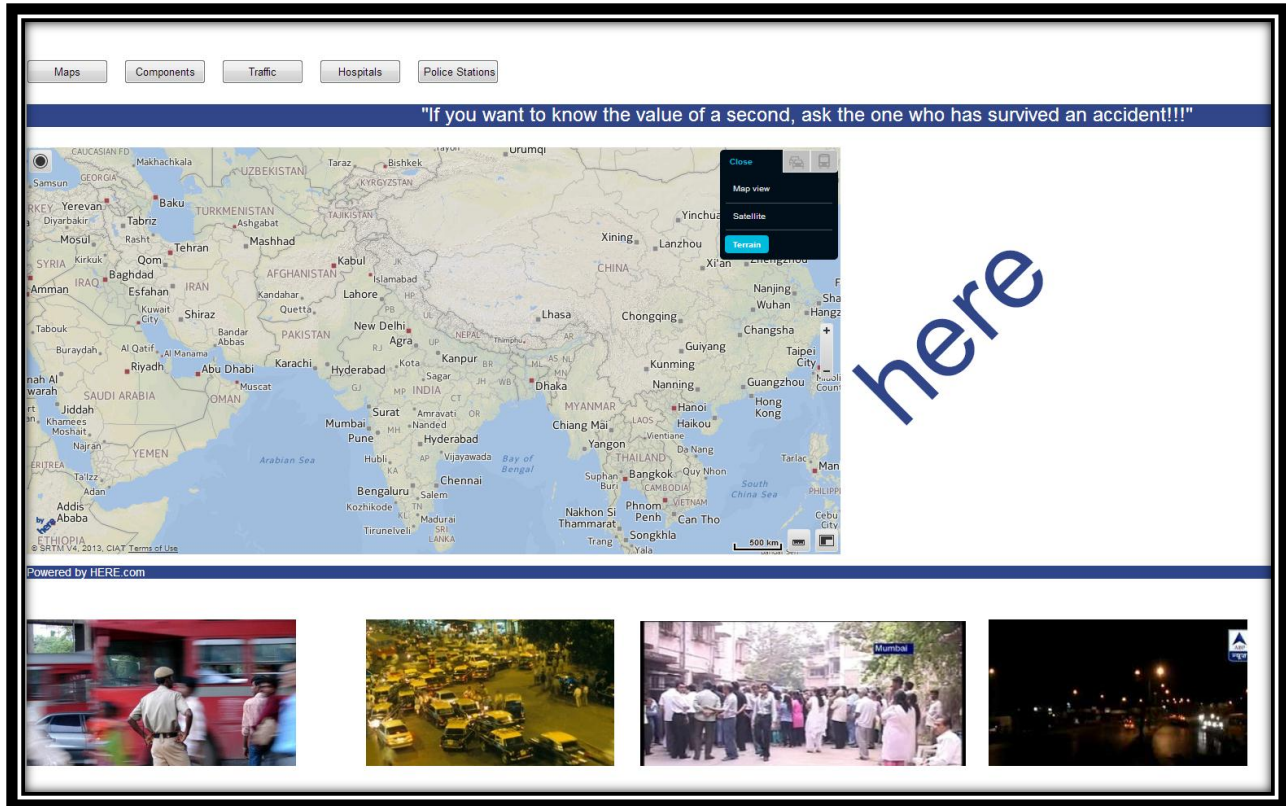
Figure 131 Components on Terrain View
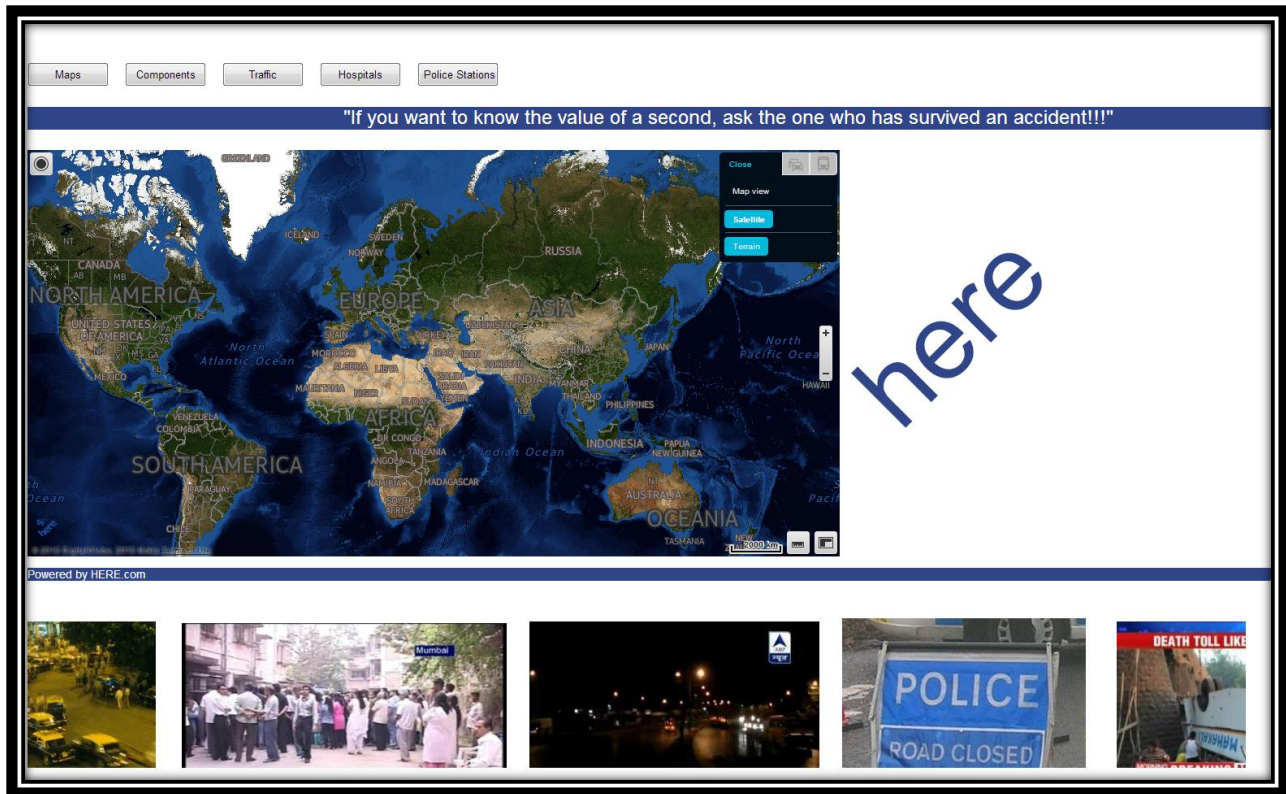
Figure 130 Components on Satellite View
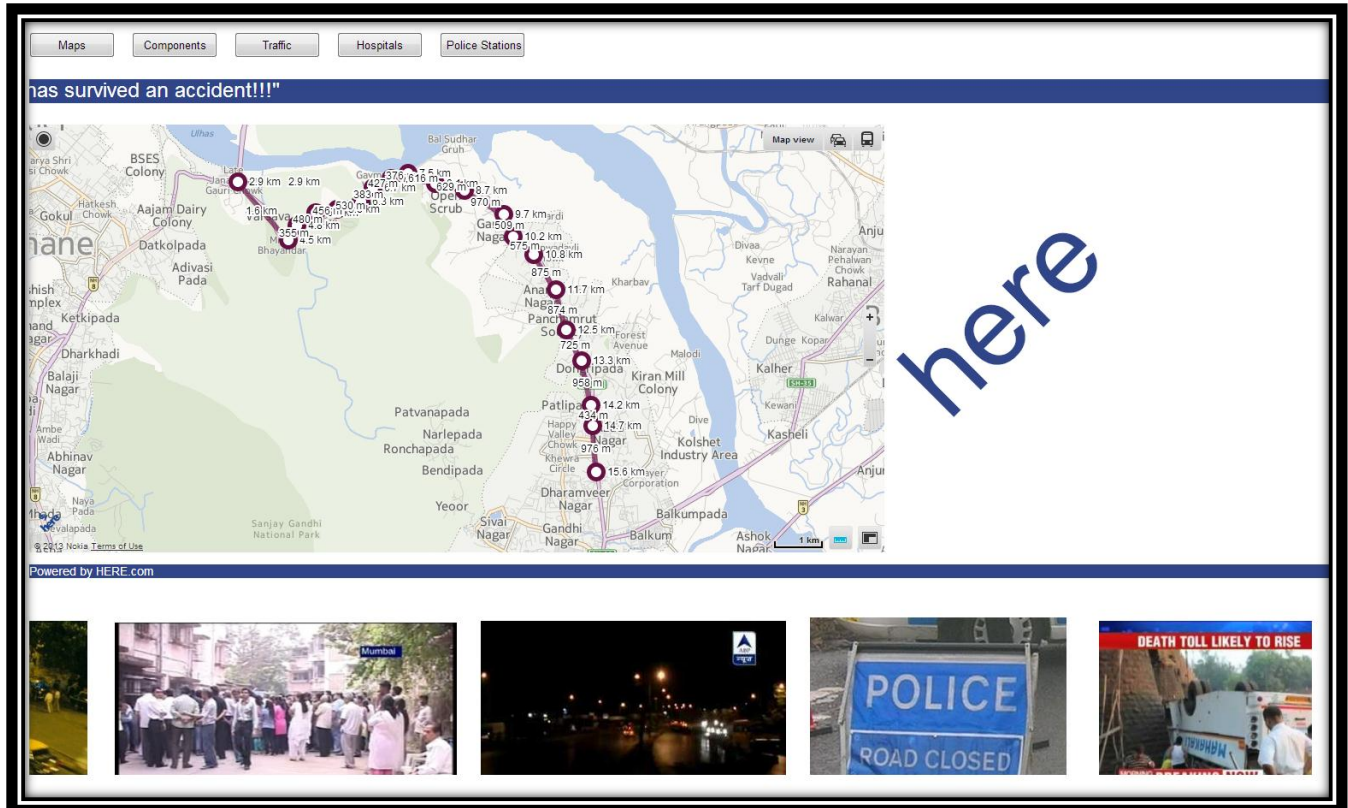
Figure 133 Length of AOI
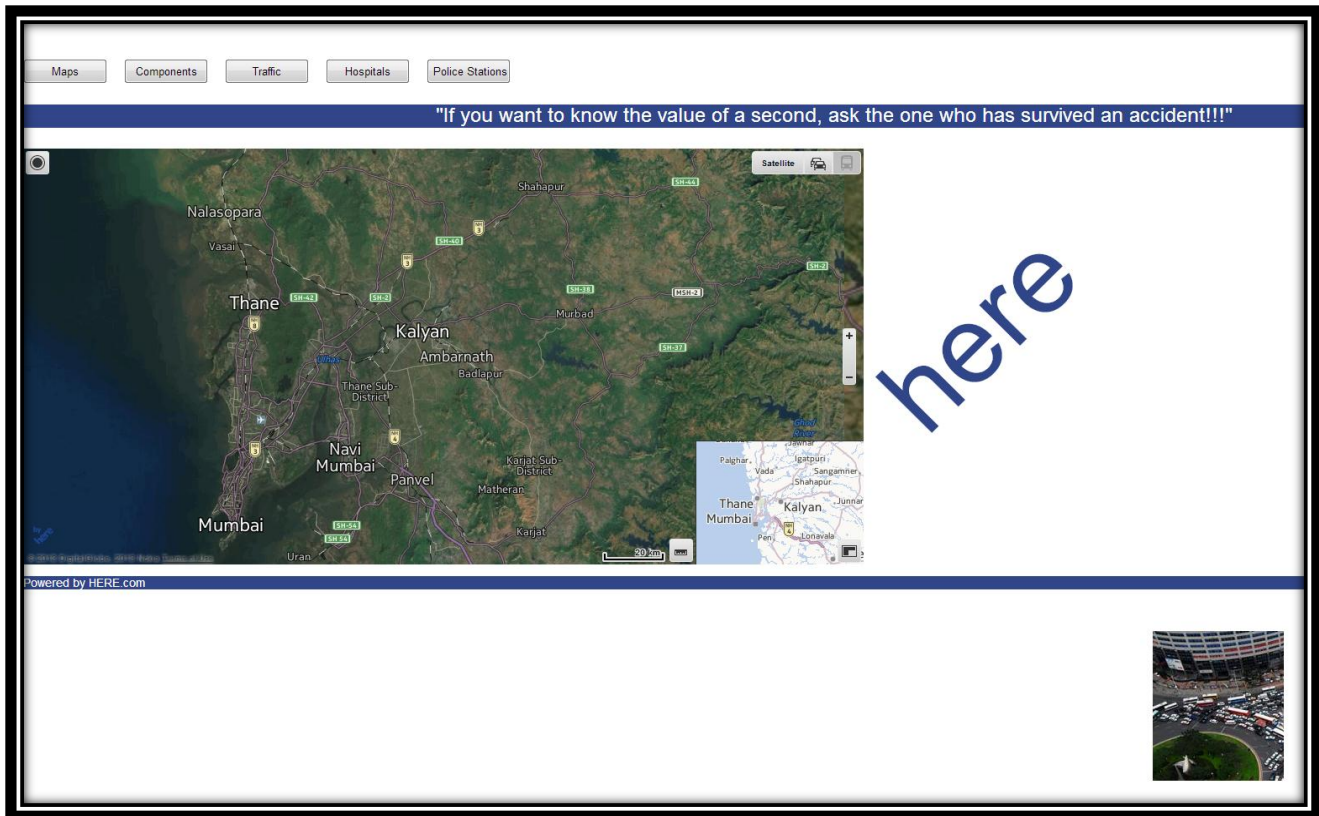
Figure 132 A Synoptic View

185
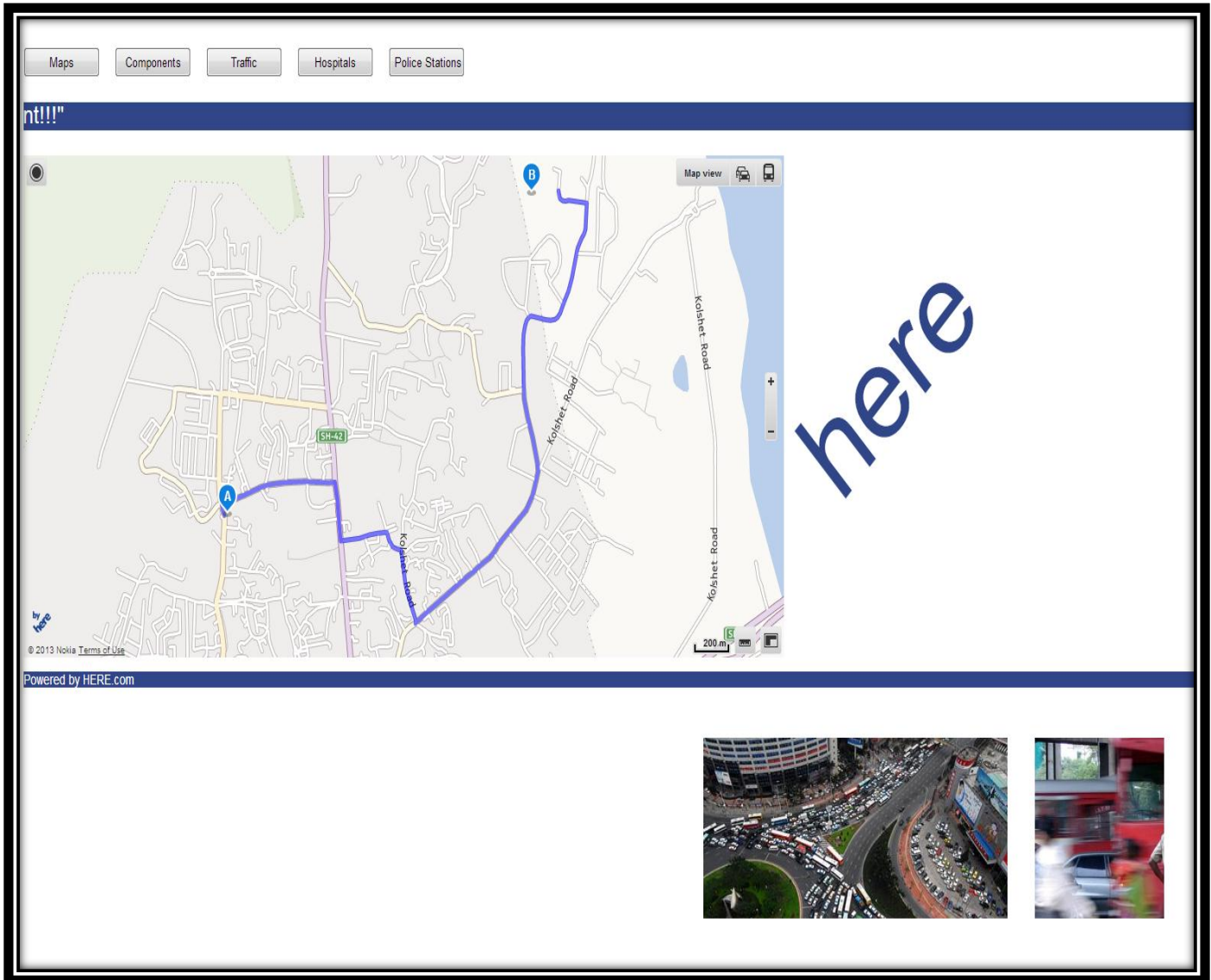
Figure 134 Routing on a Map View

This shows the routing between two selected points. For that purpose we just need to select, the source point and the destination point and the shortest route is calculated between them. This routing is carried out by using the Dijkstra's Algorithm which gives the route between two points.

The routing forms a part of the components tab.
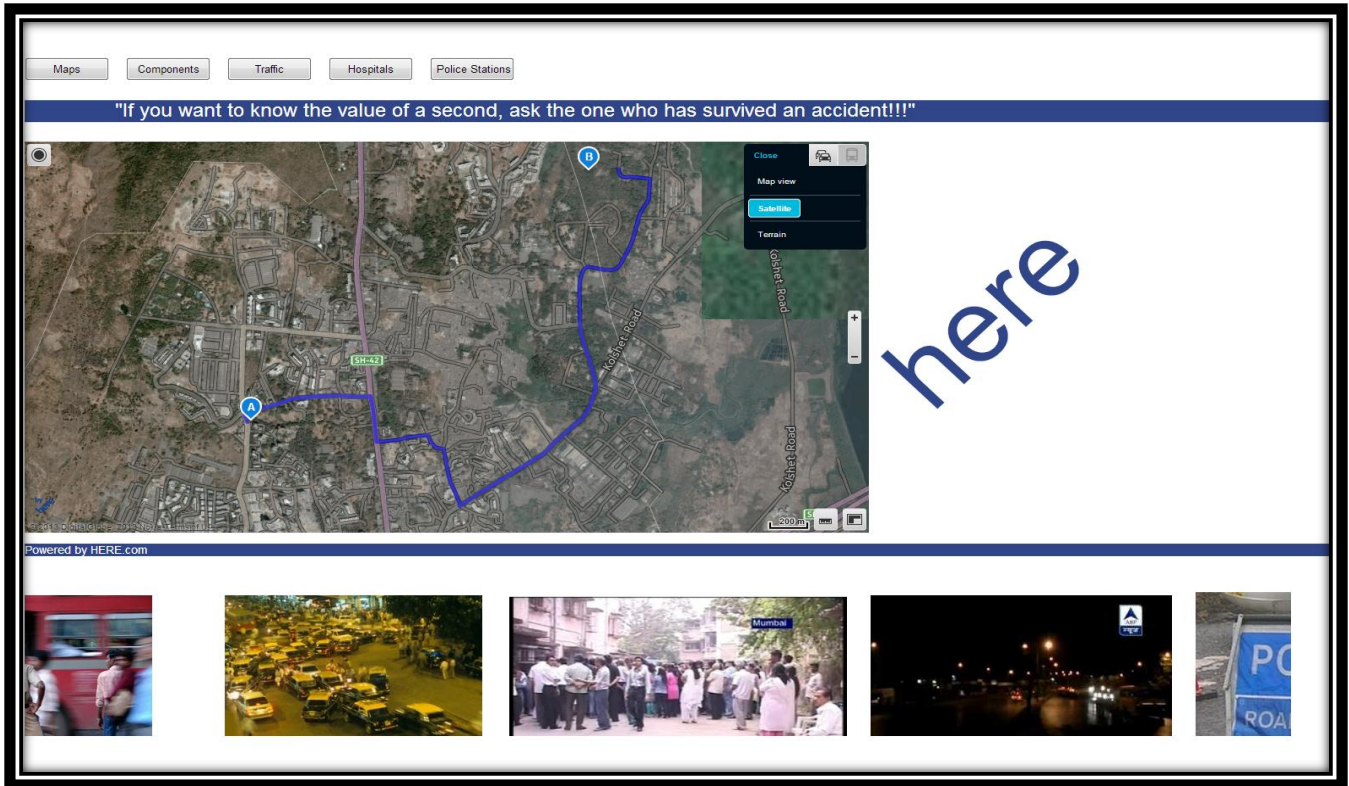
Figure 136 Routing on Satellite View
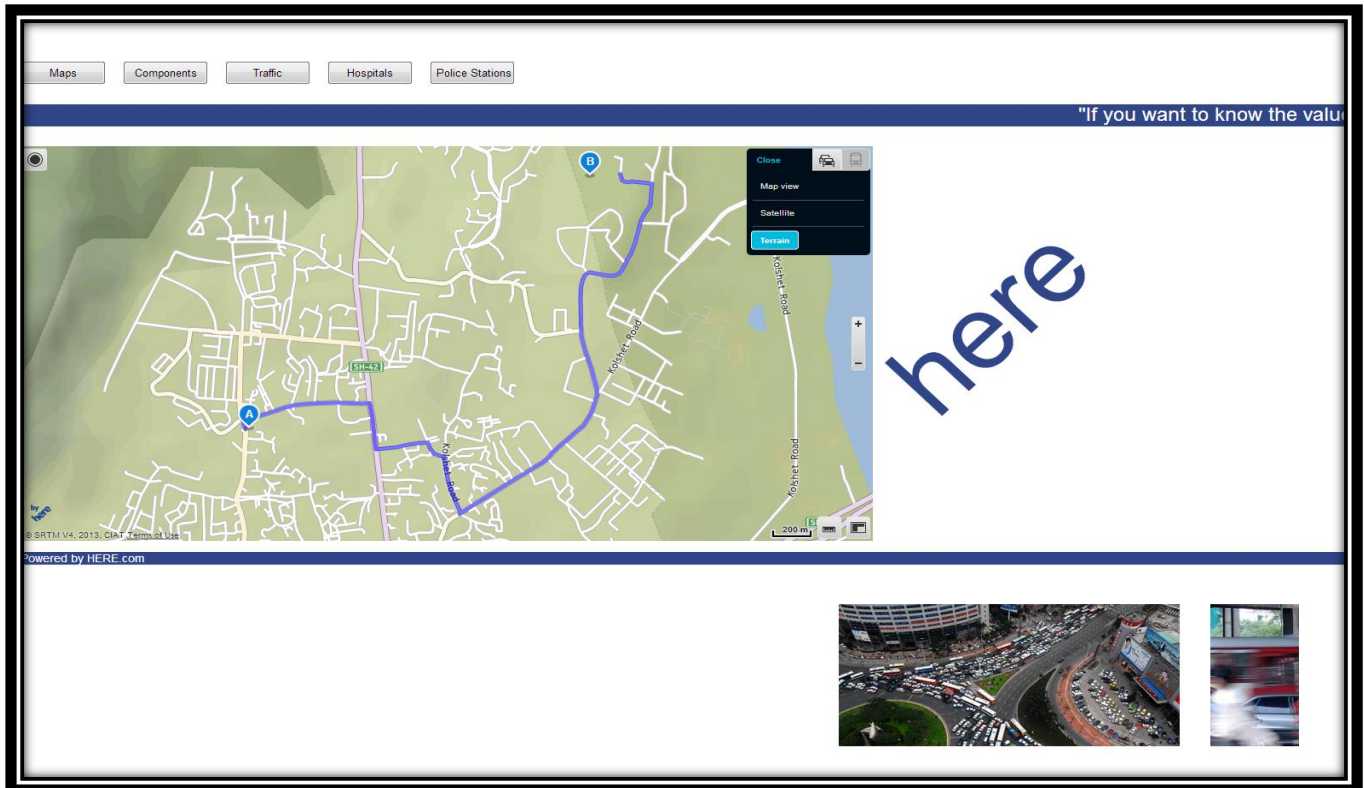
Figure 135 Routing on Terrain View

14.2.4 TRAFFIC
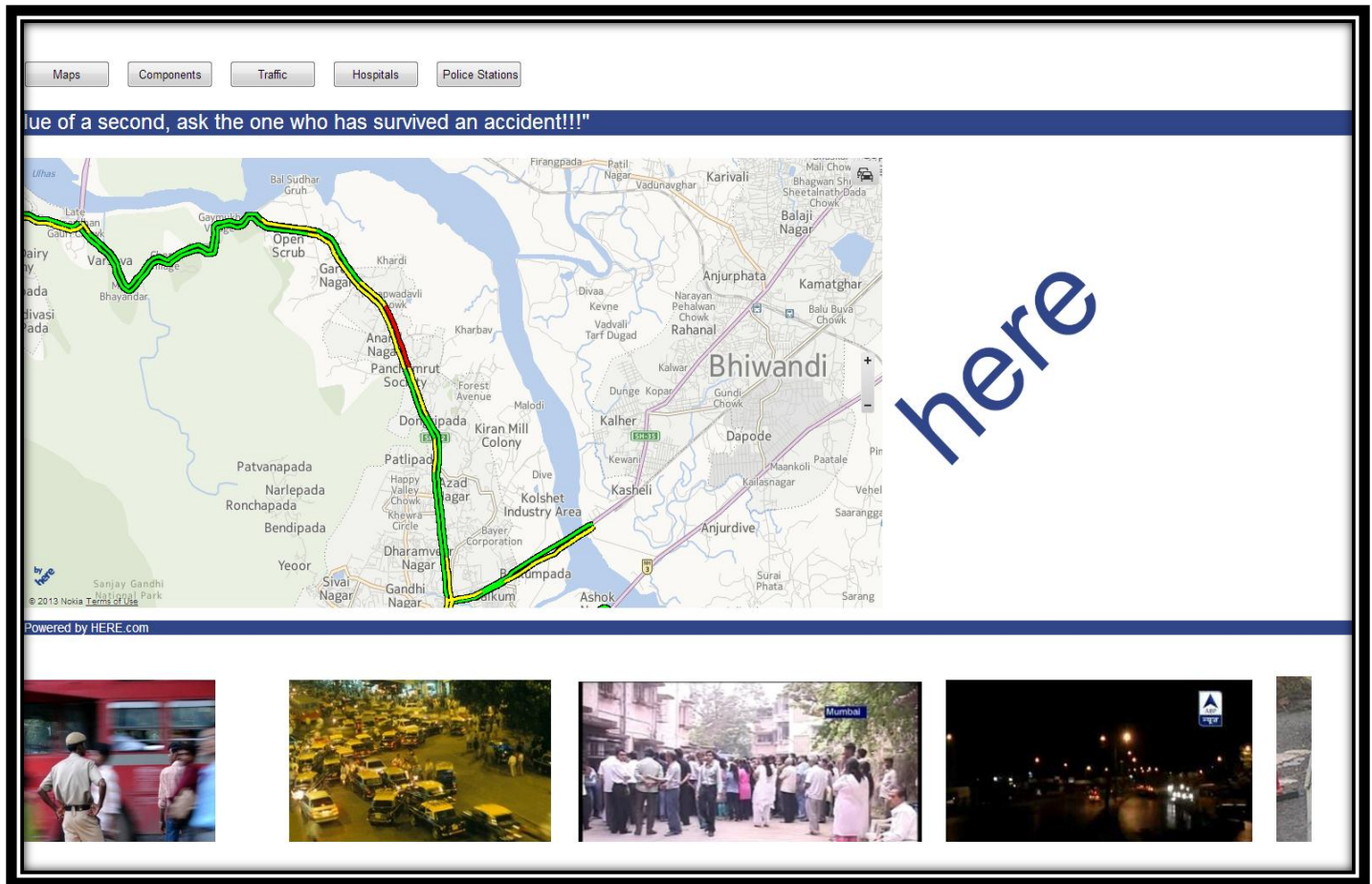
Figure 137 Traffic Data along the AOI

This tab directly links to the real time traffic data. In the figure above, we can see the traffic data along the Thane-Ghodbunder Road which was selected as our AOI. The real time traffic is available not just for our AOI but for any location on the map.

This also allows us to zoom in, zoom out and pan the map and detect for the desired location on the map.

14.2.5 HOSPITALS
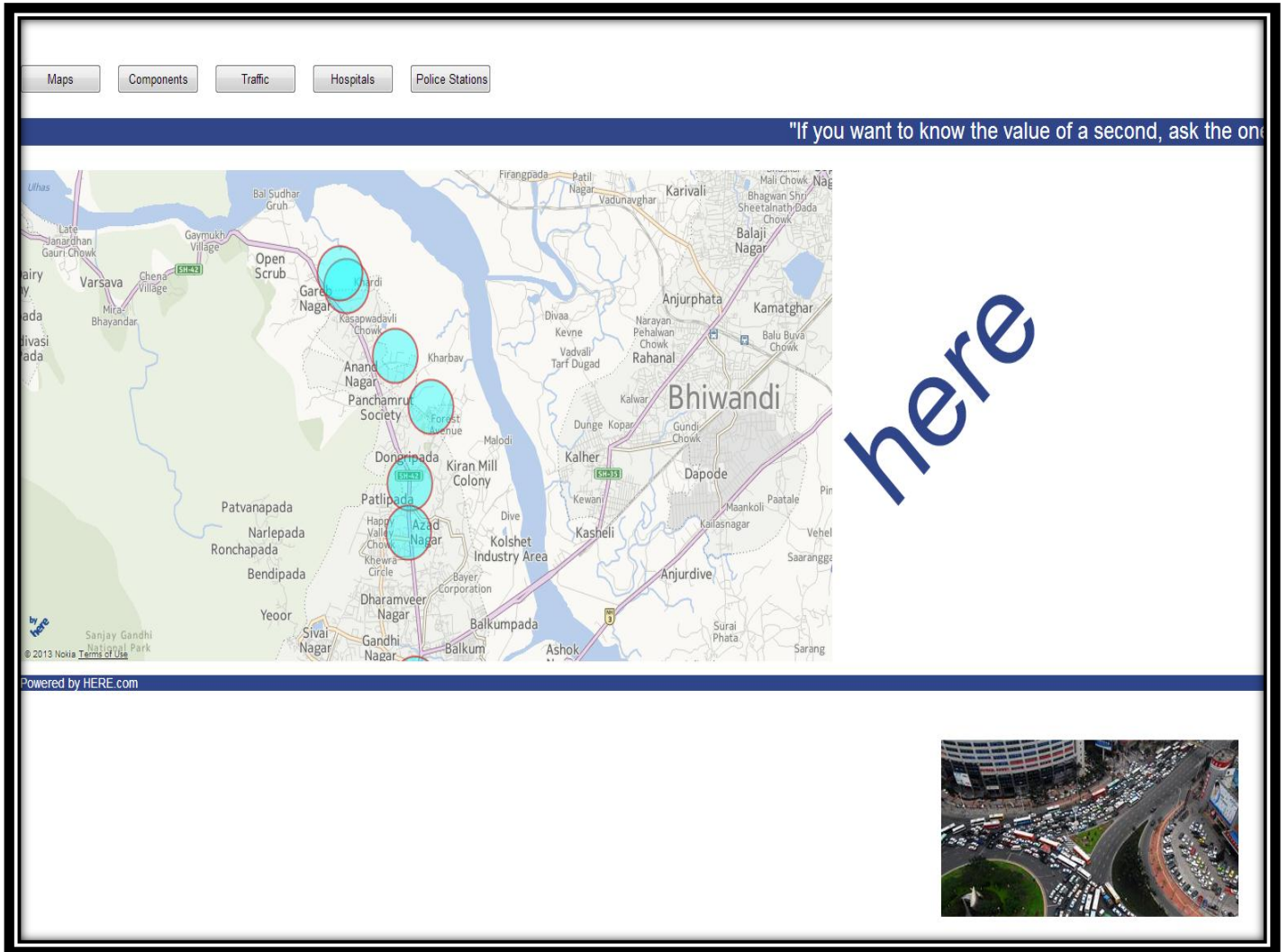
Figure 138 Overlay of Hospitals along the AOI

This Hospital tab displays and overlays all the hospitals that are present along the AOI selected. If desired it can show the distance of the hospitals from the accident sites

14.2.6 POLICE STATIONS
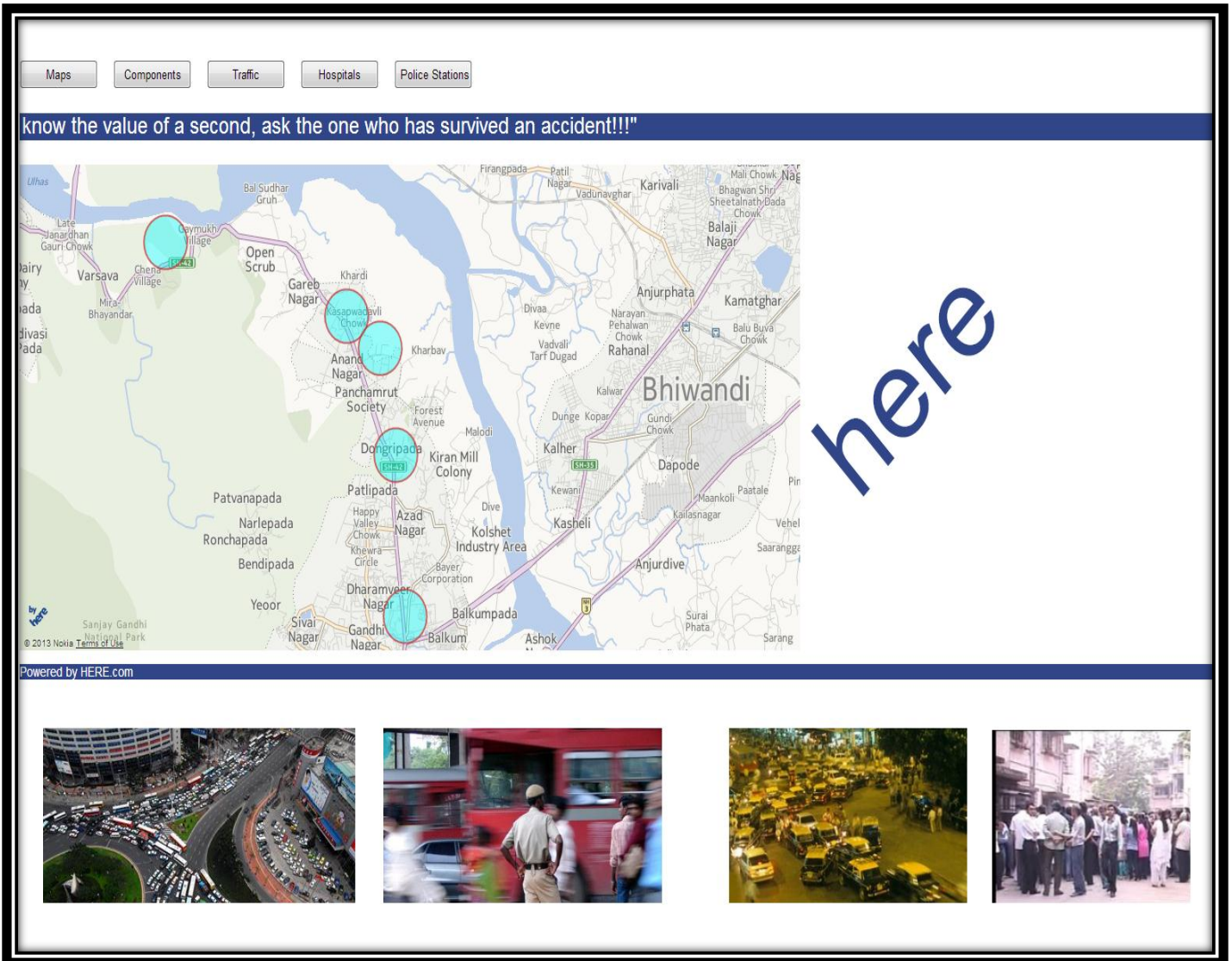
This Police Station tab displays and overlays all the police stations that are present along the AOI selected. If desired it can show the distance of the police station from the accident sites.

CHAPTER 15

SPATIAL MICRO-SIMULATION FOR ROAD SAFETY ANALYSIS

The Potential for Merging Different Micro-data for Simulative Study

Introduction

Transportation safety analysis aims at surveying events that can cause death, injury, and property damage. It also aims at understanding the causes and contributing factors to accident occurrence, severity of accident occurrence, risk and exposure to risk, and the actions required to mitigate all these. Geo-simulation and activity micro-simulation create virtual worlds of transportation systems and the agents living within these systems to assess land use and transportation policies at a very fine spatial and temporal resolutions. These relatively new methods can provide a variety of data for more informed safety analysis.

As one would expect surface transportation safety analysis examines a variety of "agents" moving within the transportation system. These agents are individual vehicles (e.g., passenger cars or commercial vehicles) operated by drivers of a wide range of training and experience in a multitude of environments and individual psychophysical conditions. Moreover, operational and policy analyses of the behaviour and performance of these agents and the performance of the system within which they act requires the use of a wide spectrum of spatial and temporal analytical fidelity and resolutions. At a somewhat coarser spatial resolution of a traffic system component, risk assessment uses a variety of tools to examine roadway intersections (e.g., analysis of conflicts among the different movements) and traffic and highway design principles to minimize accident occurrence at intersections and at different elements of roadways.

At higher levels of spatial abstraction such as an urban environment we also aim at identifying high risk locations in a city and even perform spatial analysis of much larger geographical regions . Based on these analyses preventive designs and policies can be developed that change by jurisdiction to match local circumstances and the resident population (e.g., the resident mix of very old and very young drivers). Similarly along the time dimension, we have analysis of time of day of accident occurrence that shows there are specific times during a day during which most accidents occur we also find weekly and monthly accident analysis showing seasonality and studies of accident involvement and severity in the life span of individuals but also the differential rates of changing/improving accident occurrence among different age groups. We also take into account the duration of driving before accident as many macro-temporal and daily rhythms are correlated and they also contain a variety of spatial and social context informants.

In fact, safety analysis also considers social determinants of accident occurrence and risk as well as perception and attitudes towards risk with many exogenous and endogenous determinants. All this motivates the increasing use of analytical methods that are more sophisticated, complex, but also more informative than simple descriptive statistics or basic regression techniques. The use of this type of models can be traced back to the 1950s when the freeway beltways were planned for Chicago, Detroit, and Pittsburgh in the US. At that time simulation models were mostly based on spatial interaction regional science theories aiming at linking land use to transportation system performance. Over time we experienced a gradual movement away from coarse level analyses and increasing acceptance of behavioural theories and analytical methods motivated by policy questions moving away from the construction of major projects and closer to policies that require understanding of markets.

The use of these simulation models is increasing among practitioners and it is reaching unprecedented sophistication in terms of theories, data collection, and modeling and simulation techniques. In this way we can also perform activity and travel demand analysis at any level of spatial and temporal aggregation.

Micro-simulation: An Introduction

Micro-analytic simulation or micro-simulation, attempts to re- produce or model the demographic, social or economic characteristics of human behaviour. Moreover, in terms of micro-simulation itself little of this work to date has had a spatial dimension. More recently there have been more applications of spatial micro-simulation, for example the national spatial micro-simulation model SVERIGE in Sweden. As there are limited sources of geographically detailed micro-data there is a need to create these datasets to answer certain research questions which can be done by using static geographical micro-simulation techniques. The use of micro-simulation can be justified as it is not only a novel method of answering "what-if" policy questions but also makes an original contribution to the existing literature in the field as well as allowing for interrogating the new micro-data created for applied policy analysis, answering previously unaddressed research questions.

It is said that micro-simulation models are concerned with the behaviour of micro-units such as persons within a family/ household, they are especially well suited to analyze the distributional impacts of policy changes. There is a wealth of just this type of analysis worldwide and examples of micro-simulation models already in use demonstrate this.

Micro-simulation is useful as a baseline can be constructed in the form of a mirco-dataset, from which the effects of changes (usually in the form of policy changes) to the population can be quantified or forecast into the future. In other words we can conduct a real-world experiment within a modeling environment without the need to actually conduct a real-world equivalent. The advantages of micro-simulation are outlined in brief namely; data linkage, re-aggregation, efficiency of storage and updating or forecasting. So overall, the significant advantages the micro-simulation modeling methodology can bring to empirical research with a focus on policy analysis in particular are numerous and justify the use of this technique for the analysis of accidents on Thane Ghodbunder road.
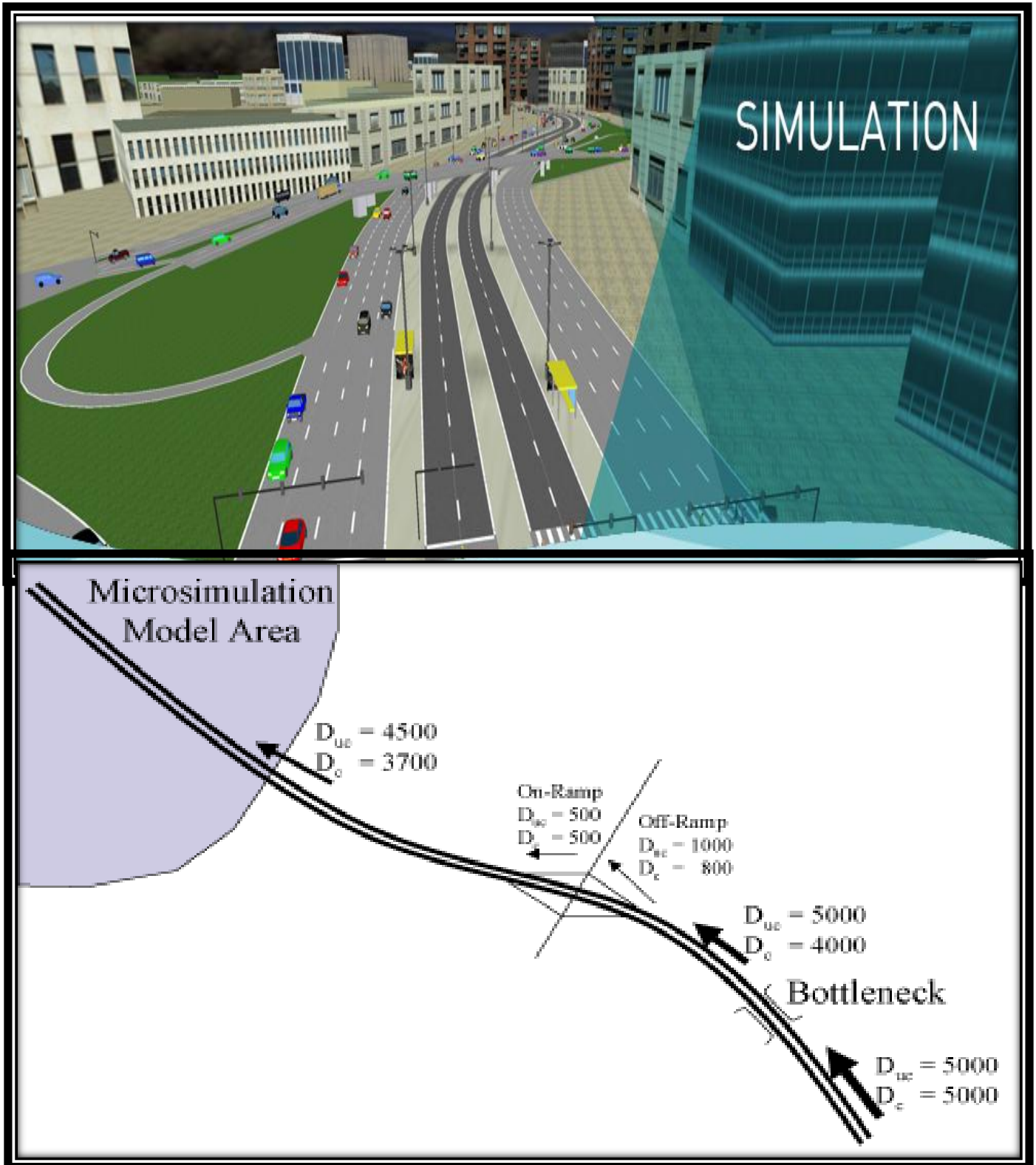
Figure 140 Simulation on Road Network

	Demand Forecasting Models	Micro-simulation Models	Operational Analysis Models
Charateristics	Uses land use, population and employment data to estimate future traffic demands for networks	Uses traffic information to model networks and determine measures of performance	Detailed capacity analysis of intersections and roundabouts, to get LOS, v/c and optimal settings
Examples	TRANSCAD EMME, VISSUM SYSTEM II	PARAMICS VISSIM, AIMSUN TRANSMODELLER	SYNCHRO, HCS, SIDRA, TRANSYT PASSER

Importance of model properties

Property	Value
Definition of standard methods	56%
Limited need for data acquisition	57%
Short lead time before use	65%
Easy integration with other databases and GIS	67%
High cost-effectiveness	71%
Possibility to run on a PC	73%
Provision of guidelines for use	88%
Models validated with real world data	93%
User-definition of key parameters possible	98%
Provision of default values for key parameters	100%

Figure 141 General micro-simulation models

The Research questions

What? Who? Where? When? Why?

The substantive research questions which have been formulated are designed to give this project a clear set of objectives, direction and focus. The research questions are outlined below:

The Real data analysis:

a) What was the accident cause?

b) Did the accident result in an injury?

c) Was the accident fatal?

d) What was the vehicle type?

e) What was the victim's position?

f) Who was the victim?

g) What was the collision type?

The Synthetic data analysis:

I. The Temporal Analysis:

a) What was the date of accident?

b) What was the day of accident?

c) What was the time of accident?

d) What was duration of driving before accident?

e) What was the direction of Travel?

II. The Physico-psychological Analysis:

a) What was the age of the driver?

b) What was the sex of the driver?

c) What is his tenure for holding the driving license?

d) What was the nature of accident?

Spatial Micro-simulation:

Spatial micro-simulation is a technique used to create simulated output data by combining or merging various datasets to populate and therefore to create a new synthetic population with an inbuilt geography. This means that the new synthetic data is created which is as close as possible to the real population for a given set of geographical areas. Once the data has been simulated it is possible to analyze and interpret it. For instance, the micro-simulated or synthetic data can be used in Geographical Information Systems (GIS) to draw maps of the new variables created or as a starting point for further dynamic micro-simulation modeling. The technique can be summarized as an up-scaling of data, for example, taking data from a survey (e.g. SHS), in order to match a series of constraint tables from a different source, e.g. the census output area totals for Scotland. The result is that the unknown census variables can be filled in from the variables which exist in the survey data. This can be achieved using spatial (or geographical) micro-simulation by iteratively merging census and survey data to create the data that will populate areas as opposed to creating just a national population which would ignore the geographical element of the data, which would be a spatial micro-simulation. This technique works by finding individuals in the survey who closely match the census constraints and then cloning these individuals so that there are enough to create a whole population for each small area. Furthermore, dynamic spatial micro-simulation, which involves forecasting key socio-economic variables into the future based on either current trends or on the consequences of different policy scenarios.

Micro-simulation Methodologies: Creating Micro-data

The various methods of creating the micro-data from the methodology include:

- Linear programming models,

- Discrete choice models

- Iterative proportional fitting (IPF)

- Simulated Annealing (SA)

- Combinatorial Optimization (CO)

- Reweighting approaches

Combinatorial Optimization is a method which can attempt to reduce the error, relative to the constraints, to acceptable levels of the output data. The method itself is a stochastic process, the initial position is selected at random, as is every candidate household subsequently .The basic principle is to select a combination of households that reproduces, in so far as is possible, the characteristics of a chosen area so that estimates are as close to the real world situation as possible.

The SA technique is an optimization method that attempts to introduce random components into a system. An SA procedure would select a combination of households that would have characteristics as close as possible to the actual data. The over-all aim of simulated annealing is to minimize the total absolute error

of the data. SA is a form of CO, but also can be used for reweighting records in a survey for example. SA also involves a stochastic or a random number element.

Deterministic reweighting as a spatial micro-simulation methodology is explained. What this entails is the reweighting of an existing micro-data sample (which is only available at coarse levels of geography), so that it would fit small area population statistics tables such that the result is a new micro-data sample for each small area. As is simply explained the weights represent the existence of the corresponding household in the corresponding area in other words how many of that household type exist in that area. This means that the original survey weights are recalculated to form a series of new weights that match a series of alternative constraints from a different source.

Some key types of Micro-simulation:

This section aims to define some key types of MSMs to give the reader an idea of the differences between alternative approaches and the varying levels of complexity of each of the different micro-simulation methodologies. It also alludes to the possible applications of each type of model which will differ slightly due to the type of model output. Describing the different approaches may also better illuminate which is the most appropriate to address the research questions.

- Static Micro-simulation - This is the most basic type of MSM. In essence what happens is that in static micro-simulation, certain behavioural relations and institutional conditions of a micro-data base of a certain time period are systematically varied. In other words, a micro-dataset is created and from this the policy analysis is performed for a given point in time. An example of this type of model is the TAXBEN model. This type of analysis (tax and benefit modeling) is particularly useful for `macro' or national level policy impacts, showing the potential impacts by particular subgroups of the population but doesn't take account of spatial impacts.

- Static Spatial Micro-simulation – A more advanced version of a static MSM with geography taken into account. This means that the spatial patterning of the model inputs and outputs are taken into consideration. In designing the model, provision is made for different areas to contain different numbers and types of individuals rather than a single national population. The main difference is that unlike a Static MSM, the Spatial MSM is estimated several times for each area separately. So, it is an extension of the Static MSM in that the national populations become the small area populations. In practice several separate MSMs are formed. This of course adds to the complexity of the model as it requires re-running an algorithm multiple times for the geography to be estimated. Static spatial micro-simulation has been effectively used for some time, for example this approach was used to forecast demand for water. Other examples include the Simulation Model for the Irish Local Economy (SMILE) which is a static spatial model as well as the SimCrime model, showing that there is room for application in areas outside economics.

- <u>Dynamic Micro-simulation</u> - A model is considered dynamic primarily when ageing is added to the model. It is stated that, the main difference between a dynamic micro-simulation model and a static one is the ageing procedure. So, the model is allowed to change over time, meaning that forecasts of present and future scenarios can be made. A static MSM, such as the two previous types, only allow for one point in time to be modeled, whereas a dynamic model will also allow the effects impacts of changes in policy over time to be modeled. The Dynamic Simulation of Income Model (DYNASIM) was one of the first models to begin the process of building a framework for this type of model. A further example of this type of model is CORSIM. This type of model would be particularly useful for policies which change over the life course; pensions analysis for example.

- <u>Dynamic Spatial Micro-simulation</u> – It is a dynamic MSM but with both a dynamic and a geographic element. In other words, this flavour is a more complex version of both the static spatial and dynamic micro-simulation models. Not only will the model change over time, but also across space. The task is to create geographically referenced micro-data and project them forward through time by simulating demographic processes. The SimBritain model is an example of a dynamic spatial micro-simulation model. Another example of a national dynamic spatial micro-simulation model is the SVERIGE model designed for analysis in Sweden. This type of model requires considerable resources in terms of model development and computing power. Most models of this type begin life as a simpler static MSM which is then further developed into a dynamic MSM. On the other hand, the power of the dynamic MSM to predict changes into the future or over time is a major feature of this type of model which may justify the considerable resource and time inputs. Dynamic spatial MSMs are therefore extensions of both static spatial and dynamic micro-simulation models which bring advantages in terms of forecasting geographically but also negatives with respect to the resource intensive nature of building such a model and the complexity involved in doing so.

The Different Simulation Models

In this section an overview of the variety of applications of micro-simulation is given as well as some of the MSMs that have been developed are discussed in brief. A disclaimer should be made at this point, in that this is in no way a comprehensive list of models, rather a selection of the breadth of applications.

- DYNASIM: This model is referred to in the literature as the first dynamic MSM and has been used as the basis for many more developments in micro-simulation modeling. The Dynamic Simulation of Income Model (DYNASIM) was aimed at addressing a key type of question, if policy makers are considering, one spending policy as compared to some other spending policy, they should know what the consequences of policies under consideration are likely to be. The model took around seven years to develop, showing the time and effort expended at the outset required to give consideration to individuals or households (micro units) rather than averages or aggregates (macro units) in a modeling context. The main issue would therefore appear to be the considerable resource required to build such a model. On the other hand, as this has been referred to as the first dynamic MSM, this is a reasonable explanation for the time taken to build a framework for any future analysis.

- CORSIM: The CORSIM model is another key model in the field of micro-simulation, a development that arguably made a series of crucial improvements to the DYNASIM model. This is because the model itself has been the starting point for many of the models that exist today, such as those in Sweden and Canada, via the sharing of the machine code used to create the micro-data. It is not explicitly a spatial MSM rather it generates a synthetic longitudinal dataset of families and persons, so is therefore a dynamic MSM. The main criticism is that there is no spatial element to the model which could provide additional useful information for policy analysis.

- SVERIGE: Micro-simulation has been used in Sweden to build a dynamic spatial MSM called SVERIGE (System for Visualizing Economic and Regional Influences Governing the Environment) which has then been the foundation for other analysis.

The type of analysis that has been performed includes the assessment of economic-demographic effects of immigration in Sweden. It was the first national level spatial model of its kind, examining the spatial consequences of policy in Sweden. The model itself is based on Swedish households and generates events for individuals based on household and regional socio-economic variables which determine the probabilities in the model of moving from one state to another state, for example into marriage. The model is separated into modules, like many other dynamic models, which contain a series of specific functions, for example a mortality module determines who is terminated from the model. The main feature of the model itself is that it has built upon previous models by adding a spatial element to the dynamic MSM. This makes the SVERIGE model a worthwhile example of the type of analysis that could be conducted in Scotland and demonstrates clearly the usefulness of the technique.

- MOSART: MOSART is a model developed by Statistics Norway and is a dynamic micro-simulation model with a cross-section of the Norwegian population and a comprehensive set of characteristics. The purpose of MOSART is to make projections of population, education, labour supply and public pension benefits for the Norwegian population. The MOSART model begins with a \1 per cent random sample of the population of Norway in 1989 ,40000 persons with actual information on marriage, birth histories, educational level and activities, pension status and pension entitlements. This shows the range of micro-data which already exists, providing a solid foundation for the model to be developed from. The model was developed to help assess the impacts of an ageing population on the expenditure on public pensions, among other things. It has been reported that by using historical data the model performs reasonably well and that its underlying assumptions turn out to be correct. The model was first developed in 1988 and completed in 1991, which shows the length of time such complex dynamic models take to develop, a significant disadvantage. As the model has been designed with sensitive data it is not widely available for use, mainly confined to Norwegian Government applications. On a positive note, the accuracy of the model and its usefulness have been demonstrated.

- <u>SMILE:</u> The SMILE (Simulation Model for the Irish Local Economy) is both a static and dynamic spatial MSM which has been designed to address the possible impacts of policy change and of economic development on the rural areas of Ireland. The main thrust of the model is a core which simulates the components of population change (mortality, migration, fertility) at the small area level which can be projected forward over time. It uses data from the UK Census of Population Small Area Population Statistics (SAPS) in Ireland. The static model creates a population at District Electoral Division level and assigns attributes to individuals which become the base population. This is then aged by adjusting the individuals' fertility, mortality and migration characteristics. The errors in the model are calculated to be reasonable averaging between 2 and 3 percent at the county level. The SMILE model produced a useful set of disaggregated spatial data for Ireland which could potentially be used for much more policy and spatial analysis. The model has also been used to estimate visitor numbers (among other variables) for an outdoor recreation site. Another use for the model was to examine the emissions of methane gases from Irish farms with micro-simulated data created using a simulated annealing methodology and combining data from the Irish Census of Agriculture and a National Farm Survey. This again shows the power of micro-simulation, which can be applied to a variety of settings as well as the potential to expand a static model to different fields or to forecast future policy scenarios.

The Data

A. <u>The Real Data:</u> As discussed in Chapter 12 (Statistical Analysis)
B. <u>The Synthetic Data</u>

The Geo-Simulation Schema

```
SHAPEFILES
   │
   ▼
GETIS-ORD GI*
   │
   ▼
ATTRIBUTES
```

```
Ordinary Least Squares (OLS)      Geographic Weighted Regression (GWR)
              │                            │
              └──────────┬─────────────────┘
                         ▼
                   ATTRIBUTES
                         │
                         ▼
                 Linear Regression  ◄──── DATA
                                            │
                                    Temporal Data
                                            │
                                  Physicopsychological Data
```

```
Simulation Model
      │
      ▼
  Final Output
```
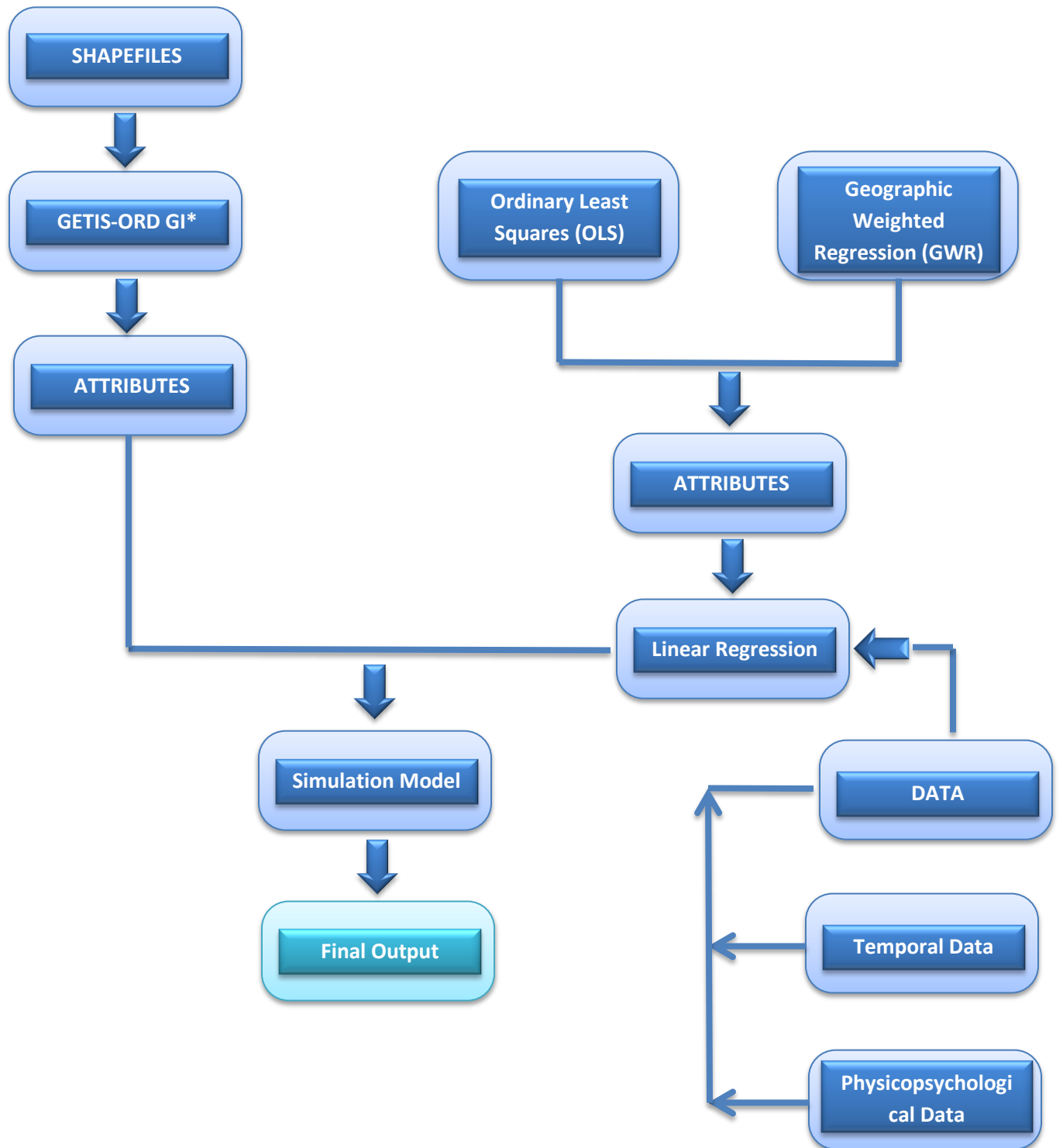
The RAYSIM
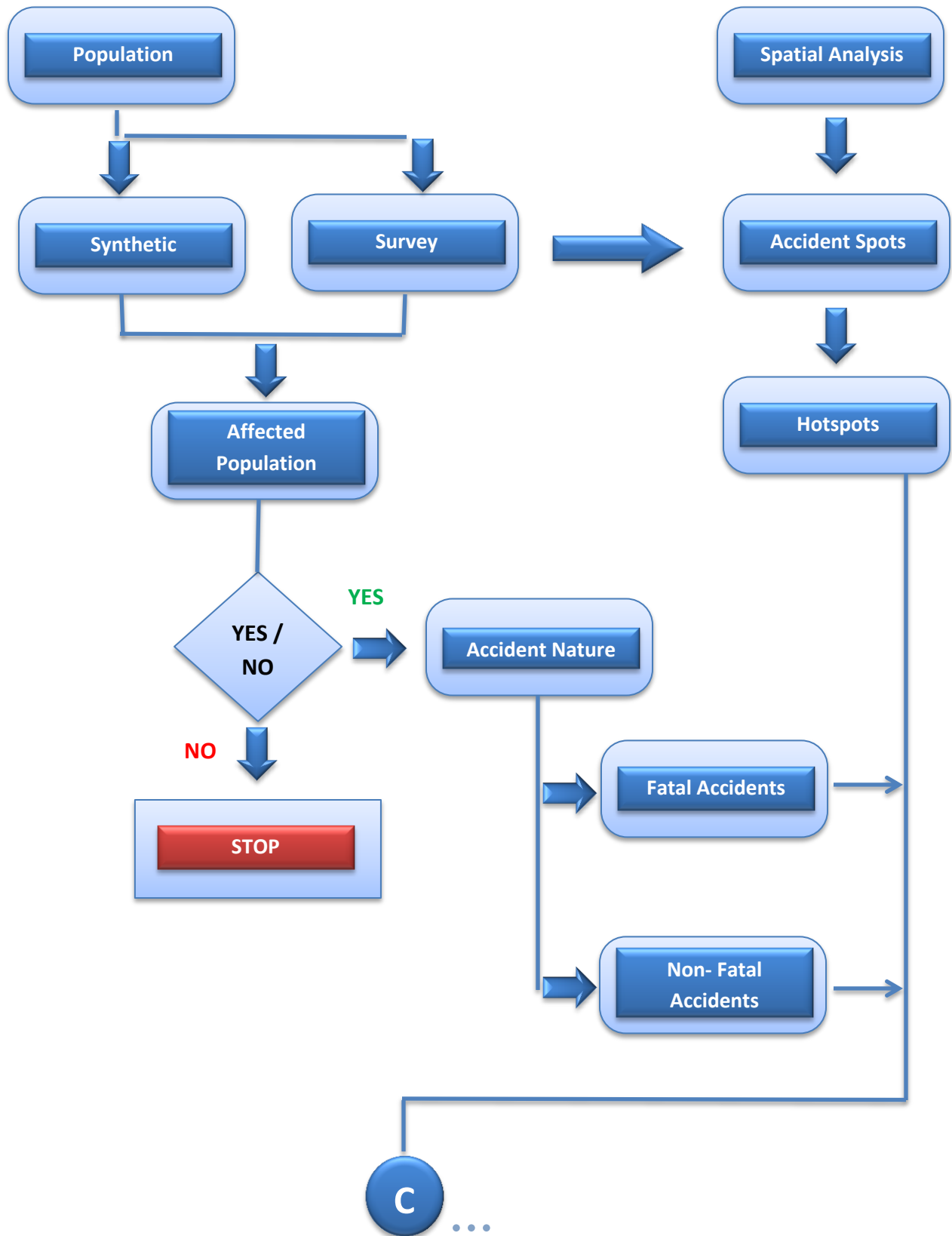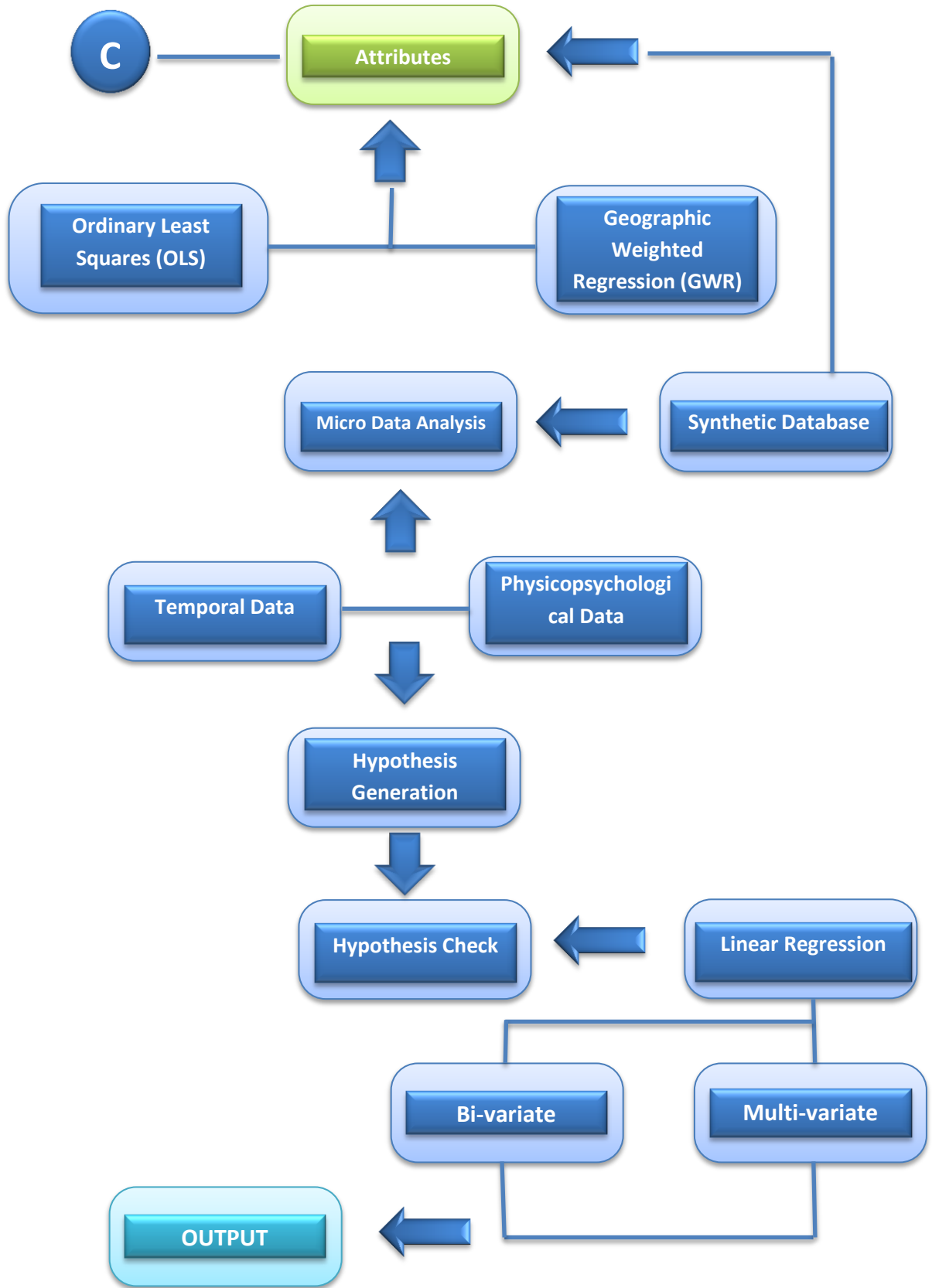
RESULTS AND DISCUSSION
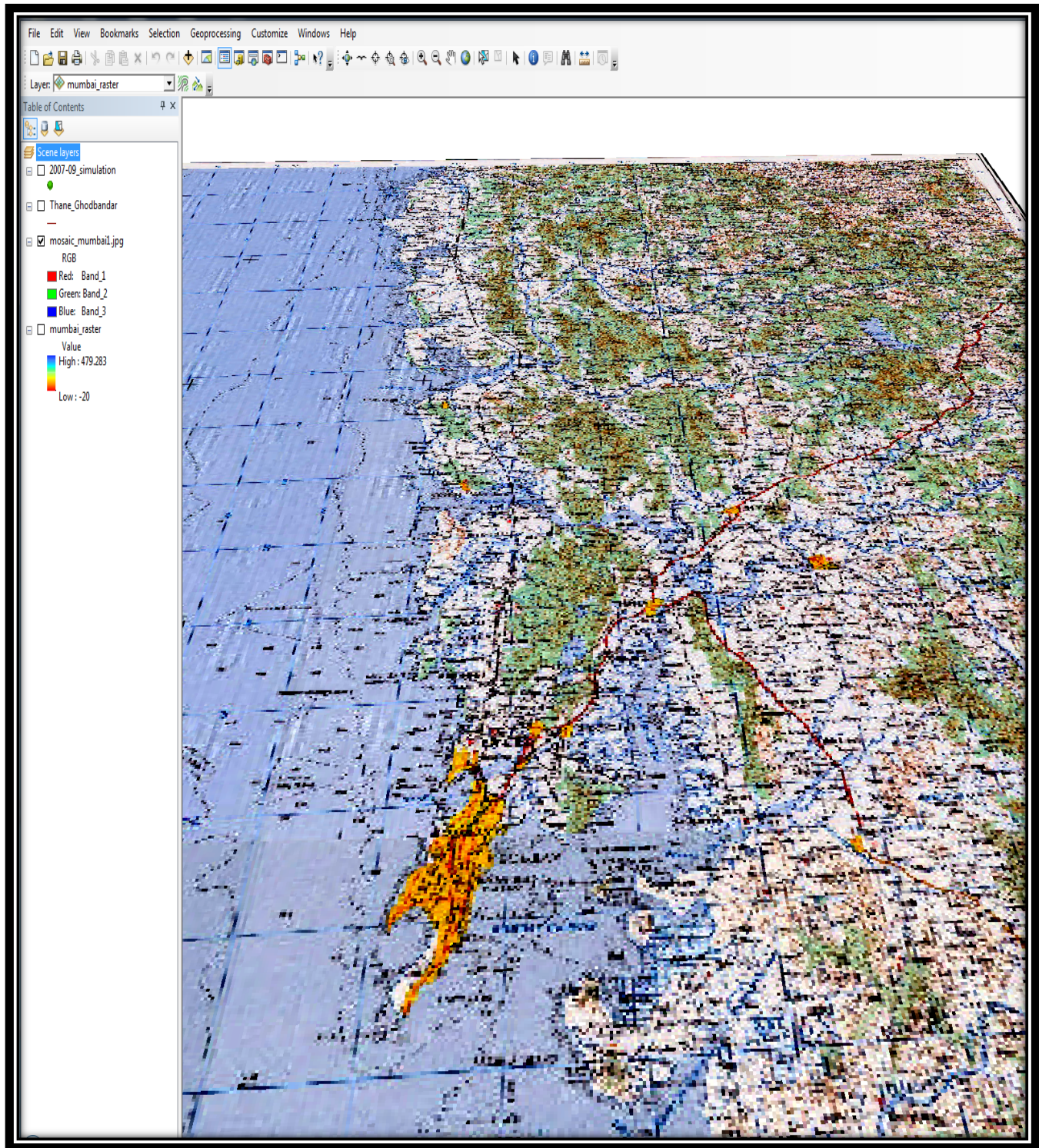
Figure 142 The georeferenced toposheet
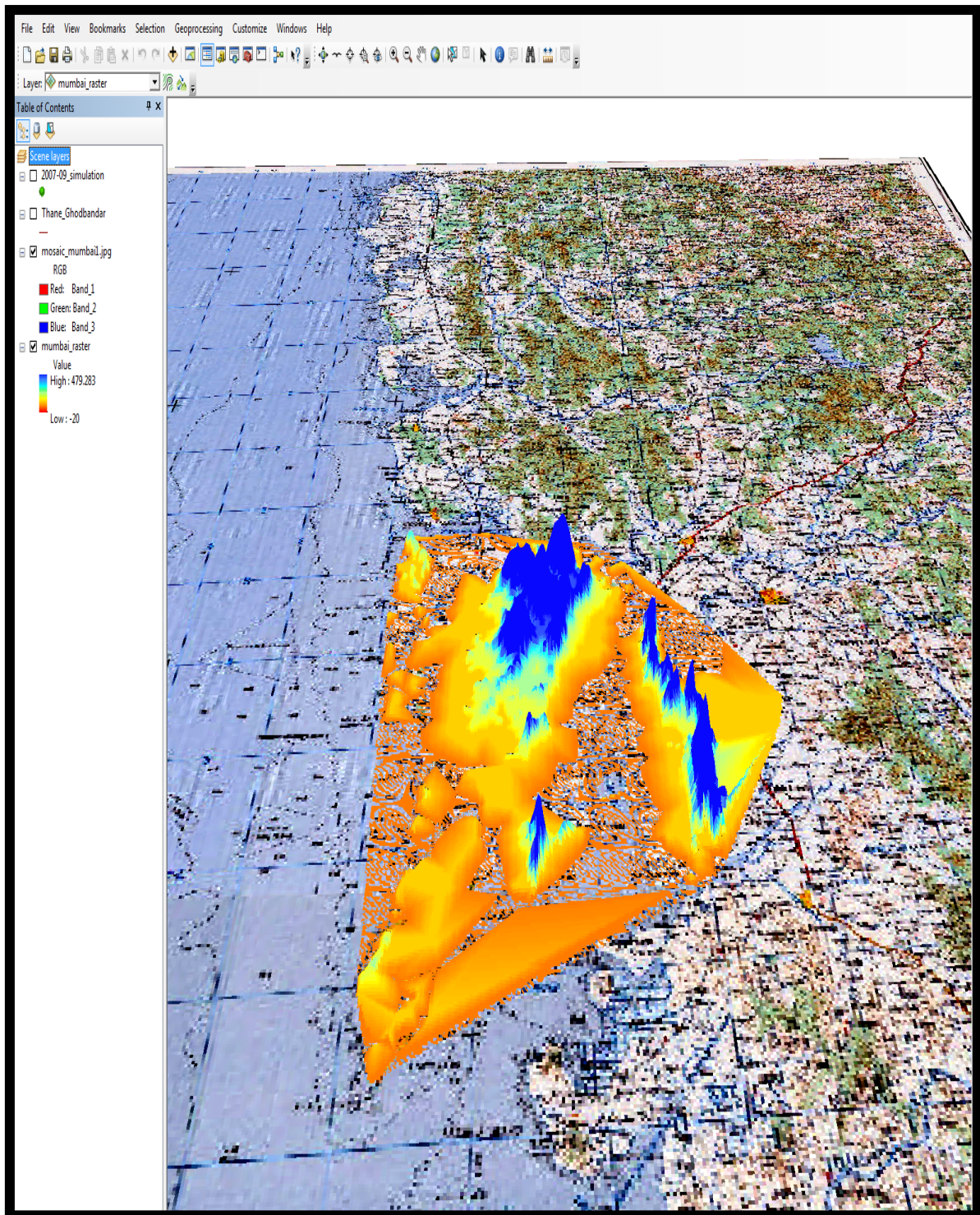
Figure 143 The georeferenced toposheet of the selected AOI
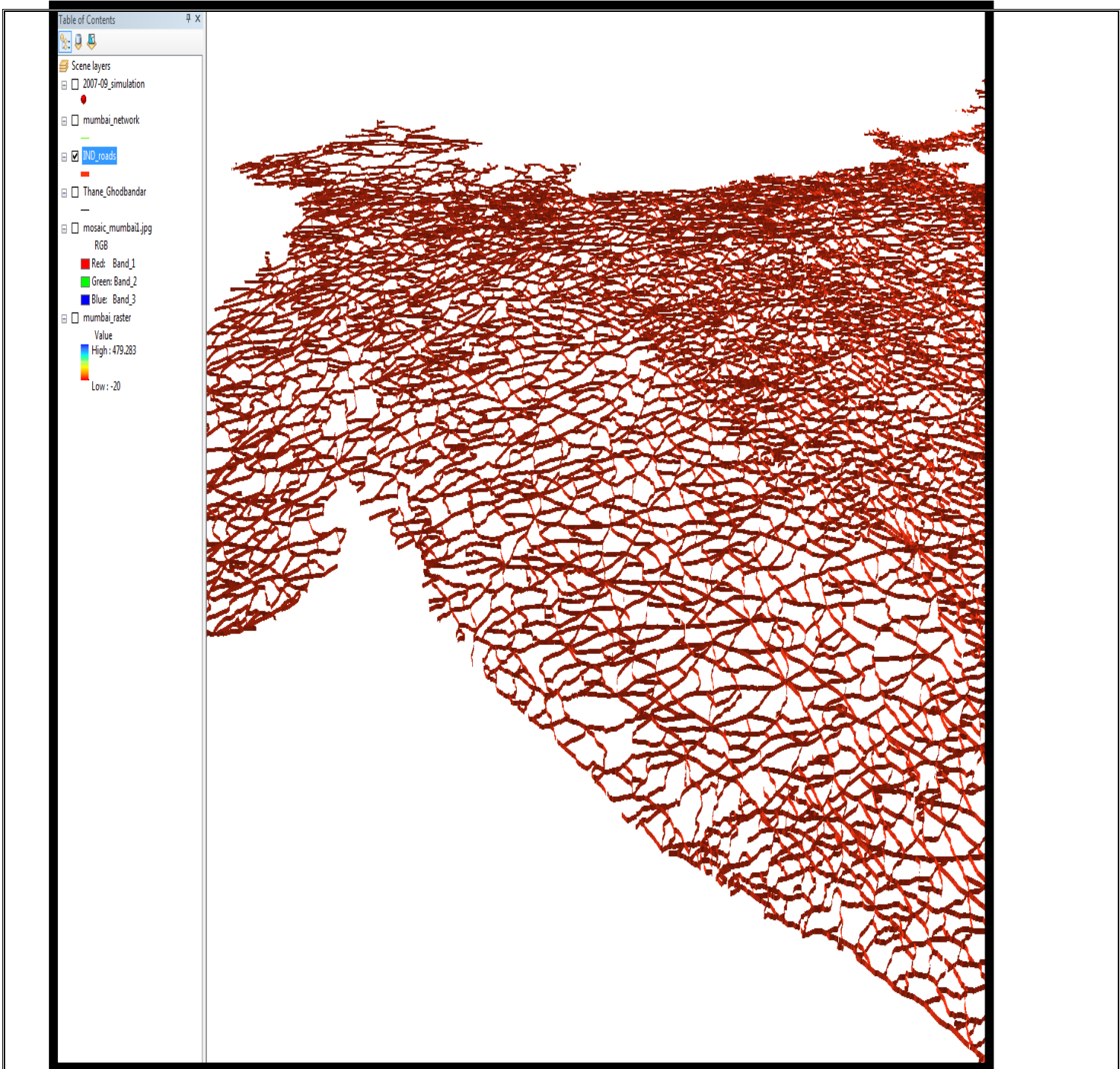
Figure 144 The 3D View of India Road
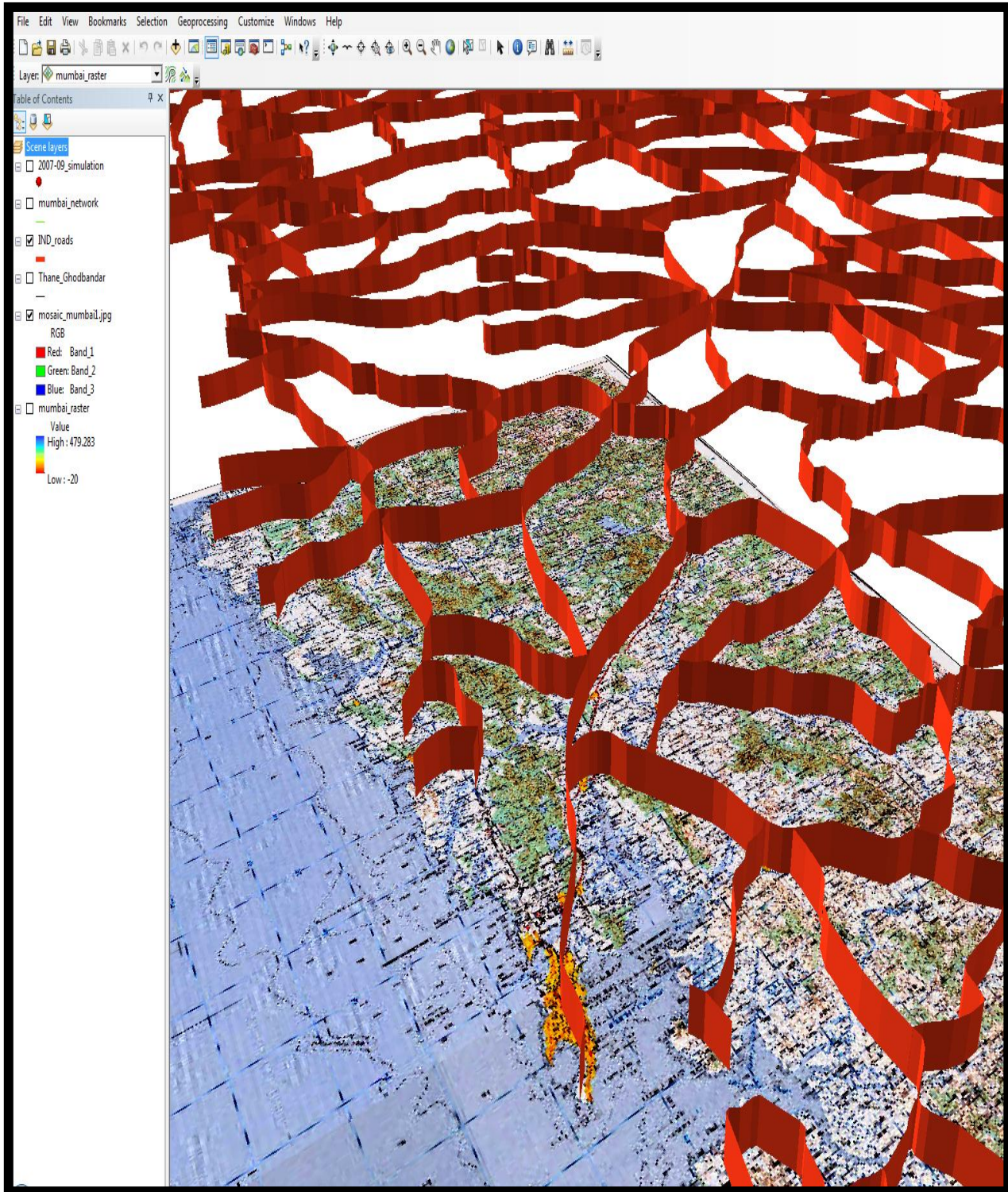
Figure 145 3D View of the road network overlaid on the toposheet
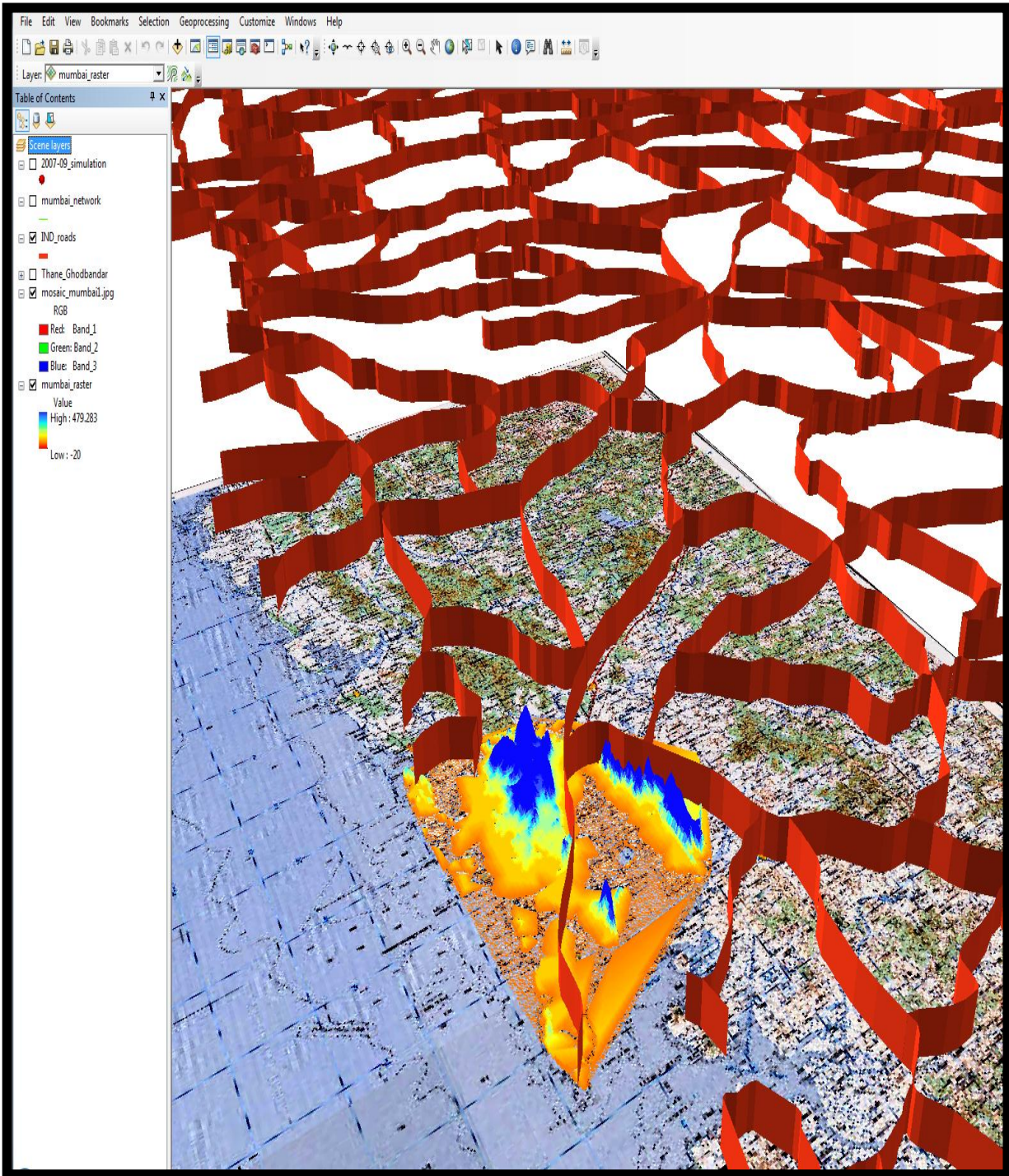
Figure 146 The 3D View of India Roads, shown in Mumbai
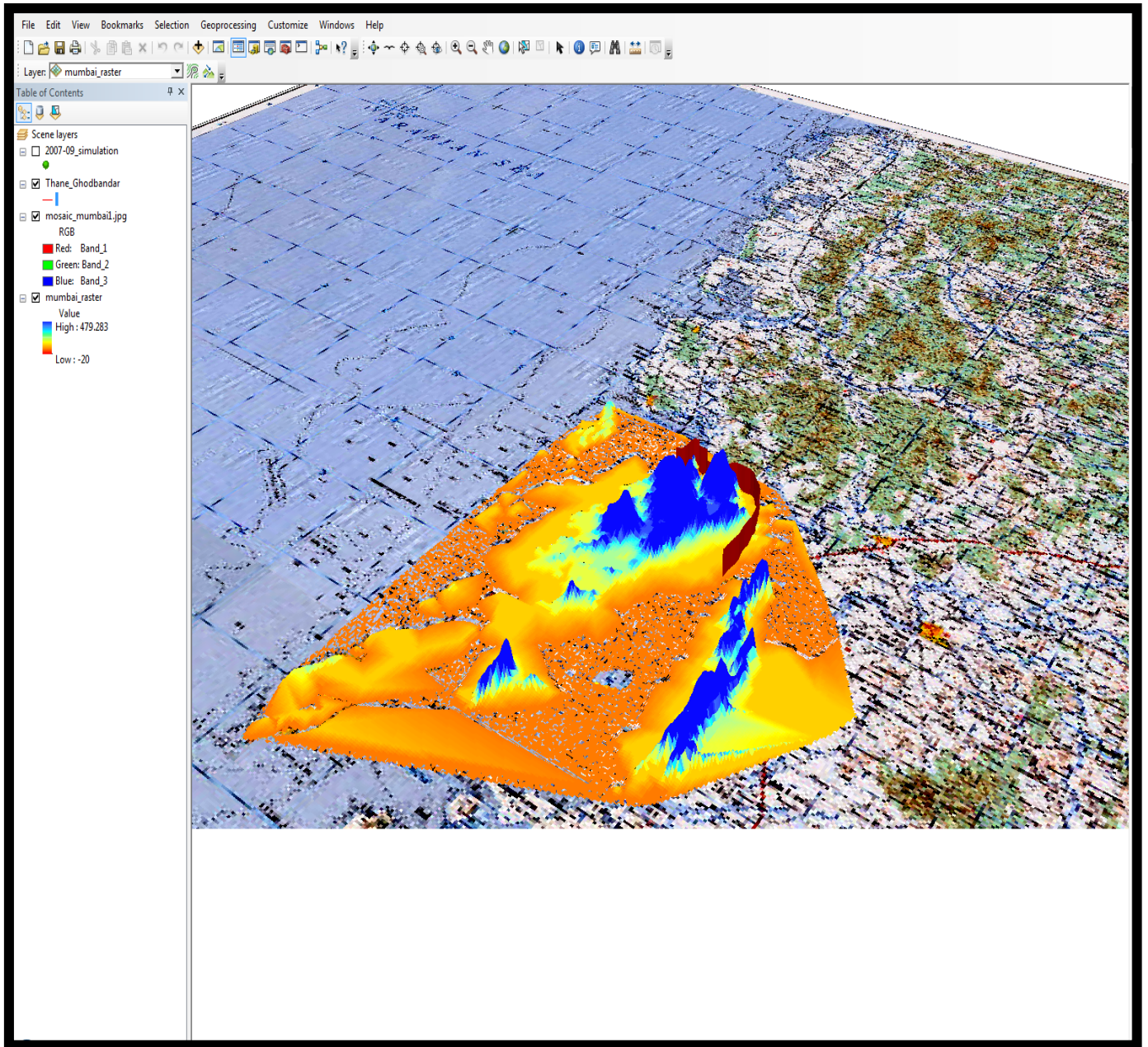
Figure 147 The 3D View of the Thane Ghodbandar Road
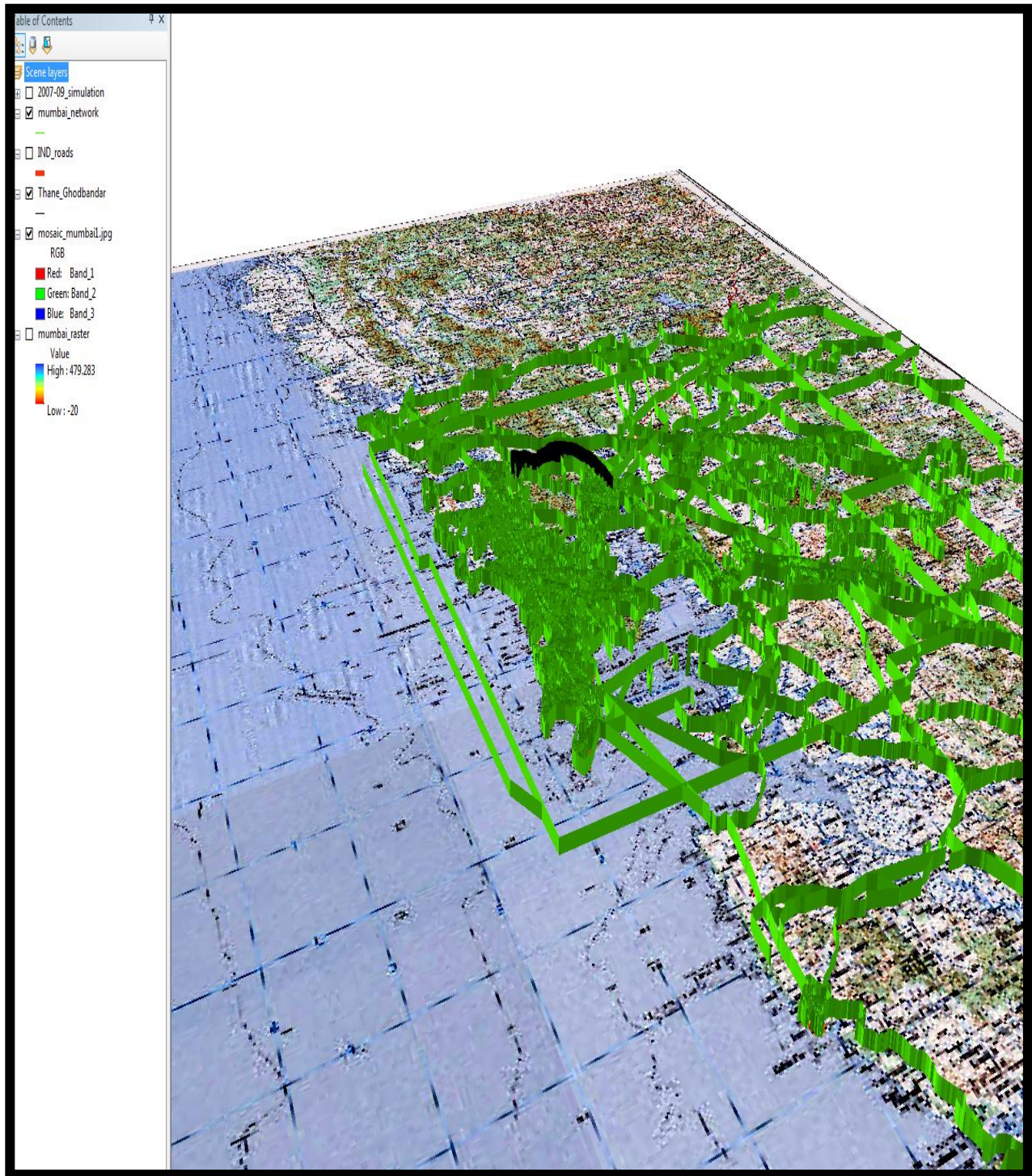
Figure 148 The Mumbai Network in 3D View
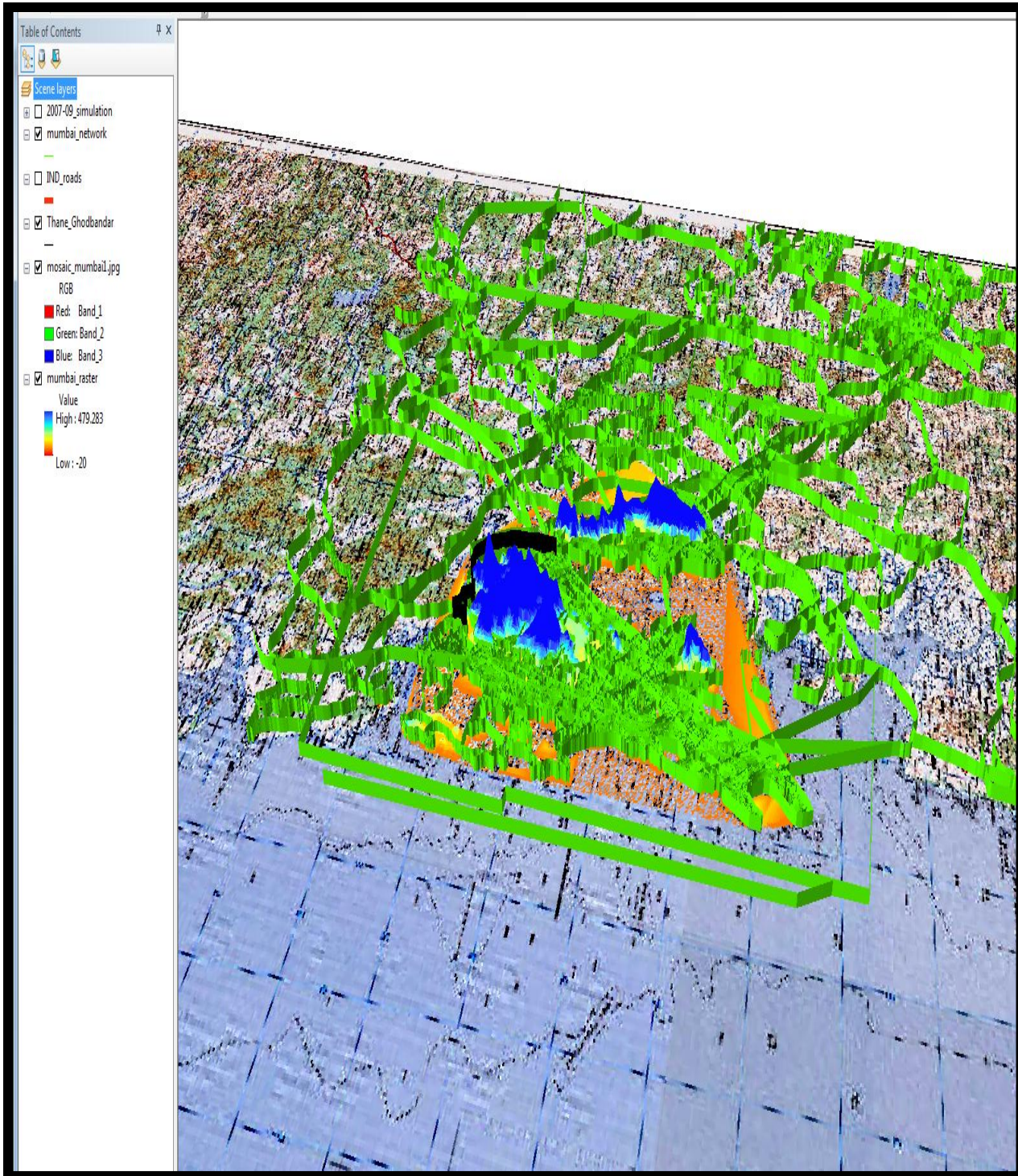
Figure 149 A 3D View of India & Mumbai Road Network along with the elevation
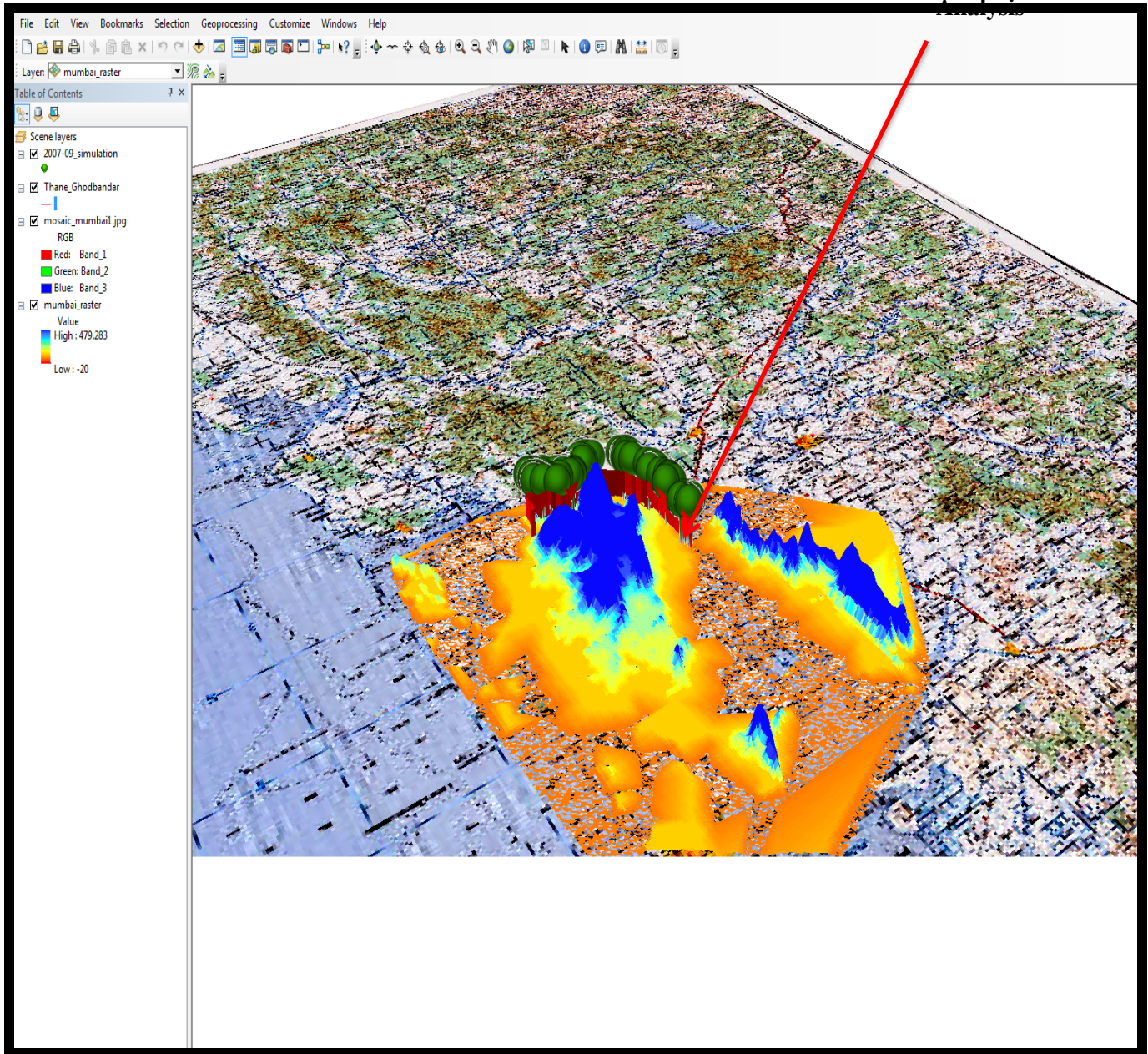
Figure 150 The 3D View of the Thane Ghodbandar Road with the Accident Sites for the Year 2007-09
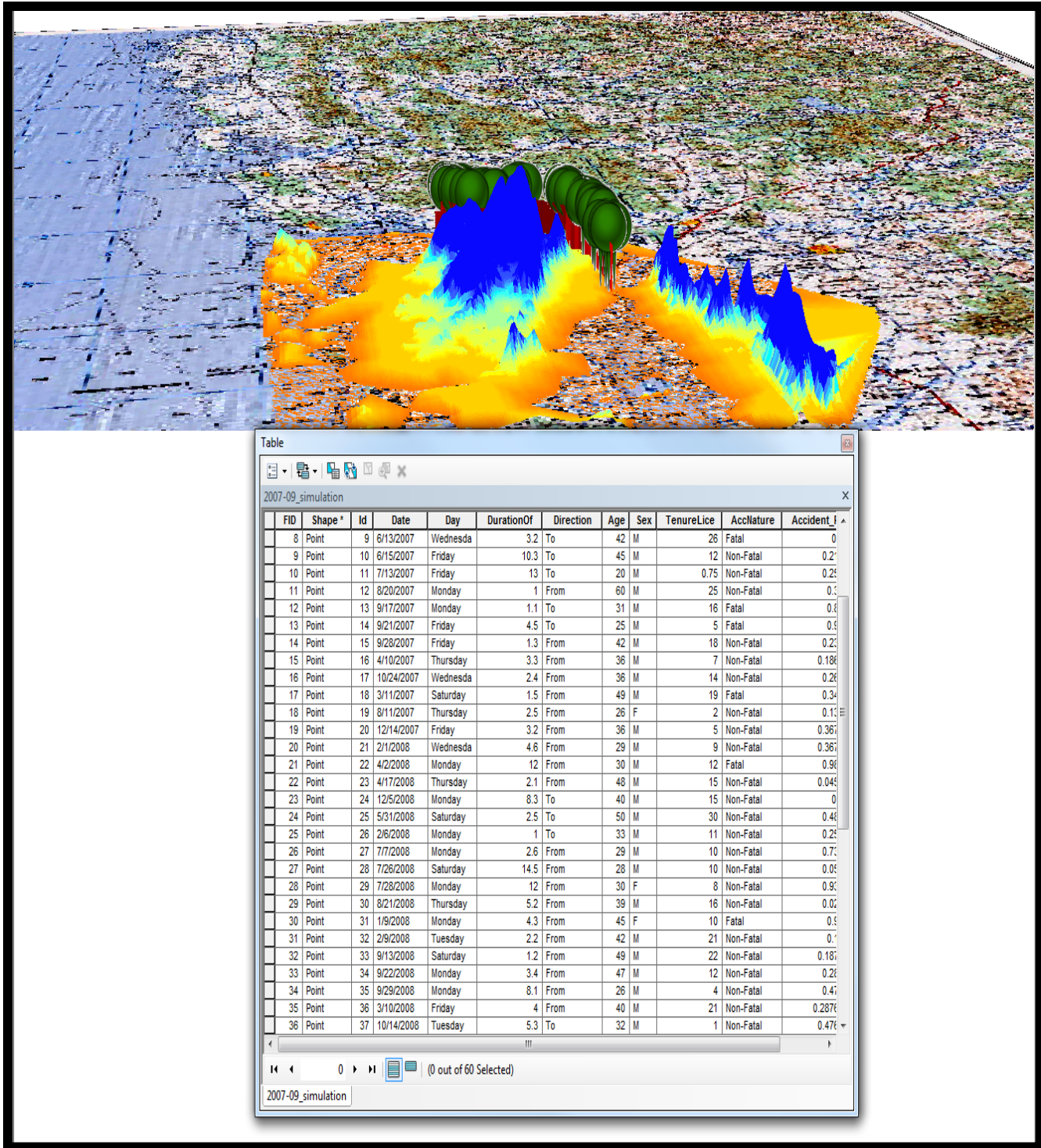
Figure 151 The Attribute Table of the Accident Sites for 2007-09 for the Microsimulative Analysis
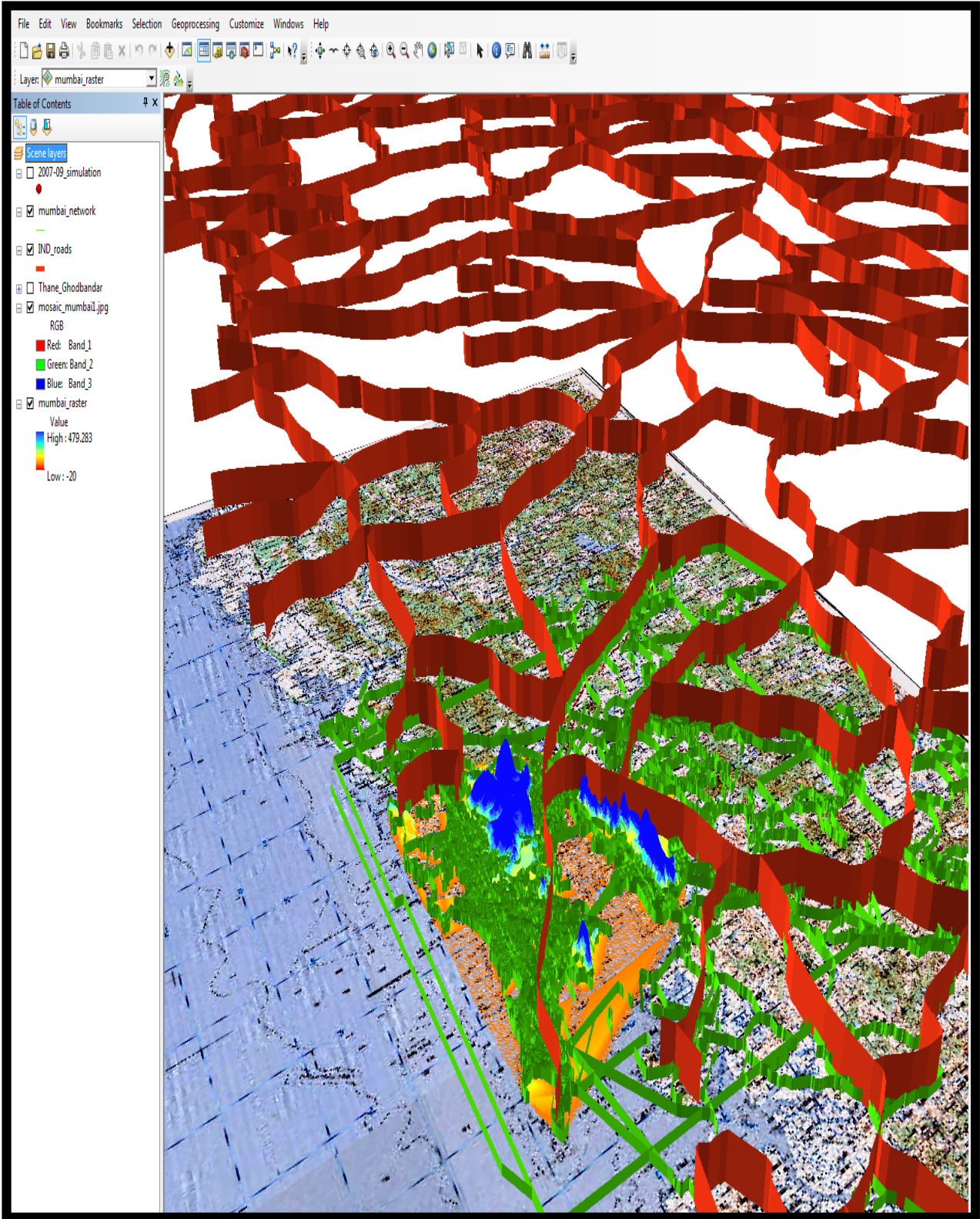
Figure 152 The 3D View of India Roads, shown in Mumbai along the Mumbai Network

The Geosimulation Analysis

A. Data & Methods
1. Introduction

In this section we will discuss the wide range of data and methods which were considered for this research project. It sets out a brief description of the data and the method and also the appropriate details the pros and cons of the usage of a particular dataset of method. The various different types and sources of data for this research project are dealt with to begin with. This is followed by a similar exercise for the methods that could be used.

We have used a wide range of data, to carry out the objective of this project work. Starting with the road, analysis, which included the gradient, curvature, bend, blind turns, and the associated spatial analysis. Further, the risk assessment factor associated with accidents and the nearby population distribution required a simulative analysis and study. This simulation when carried out at the microdata scale required a miscrosimulative study. The micro level data collection was a difficult task to achieve, so the Microsimulation has been carried out on a Synthetic Data Set. This Synthetic Data Set was achieved by, first categorizing the data, and understanding the different parameters, required to carry out the risk assessment.

The data were further grouped together to perform the

- Temporal Analysis
- Physicopsychological Analysis

The Database for year 2007-09 to carry out Microsimulation

2007-2009

Analysis of time to show the time of day, when accident occurs

Date	Day	Time of accident	Duration of Driving before accident	Direction of Travel	Geographical Location
1/2/2007	Friday	9:20:00 AM	2 Hours 15 Mins	To	72.894 19.281
13/04/2007	Friday	9:20:00 AM	1 Hours 5 Mins	From	72.892 19.29
7/5/2007	Monday	10:20:00 AM	3 Hours 25 Mins	From	72.922 19.273
21/5/2007	Monday	9:30:00 AM	1 Hours 15 Mins	From	72.923 19.273
22/05/2007	Tuesday	9:00:00 AM	1 Hours 00 Mins	From	72.938 19.281
5/6/2007	Tuesday	11:20:00 AM	2 Hours 15 Mins	From	72.922 19.271
5/6/2007	Tuesday	8:00:00 AM	00 Hours 45 Mins	To	72.976 19.241
8/6/2007	Friday	7:30:00 AM	1 Hours 05 Mins	From	72.98 19.232
13/06/2007	Wednesday	8:30:00 AM	2 Hours 35 Mins	To	72.983 19.224
15/06/2007	Friday	7:30:00 AM	1 Hours 45 Mins	To	72.894 19.281
13/07/2007	Friday	9:30:00 AM	3 Hours 15 Mins	To	72.893 19.285
20/08/2007	Monday	5:20:00 PM	10 Hours 45 Mins	From	72.895 19.286
17/09/2007	Monday	6:10:00 AM	10 Hours 5 Mins	To	72.892 19.29
21/09/2007	Friday	1:20:00 PM	00 Hours 10 Mins	To	72.922 19.273
28/09/2007	Friday	10:00:00 AM	1 Hours 00 Mins	From	72.923 19.273
4/10/2007	Thursday	11:00:00 AM	1 Hours 15 Mins	From	72.922 19.271
24/10/2007	Wednesday	12:20:00 PM	4 Hours 45 Mins	From	72.937 19.283
3/11/2007	Saturday	2:00:00 PM	1 Hours 45 Mins	From	72.942 19.283
8/11/2007	Thursday	9:00:00 PM	3 Hours 10 Mins	From	72.938 19.281
14/12/2007	Friday	3:40:00 PM	2 Hours 20 Mins	From	72.937 19.279
2/1/2008	Wednesday	10:30:00 AM	1 Hours 35 Mins	From	72.984 19.22
4/2/2008	Monday	10:20:00 AM	2 Hours 45 Mins	From	72.938 19.281
17/04/2008	Thursday	11:20:00 AM	0 Hours 15 Mins	From	72.947 19.279
12/5/2008	Monday	11:00:00 AM	1 Hours 00 Mins	To	72.976 19.241
31/05/2008	Saturday	10:10:00 AM	1 Hours 05 Mins	To	72.922 19.273
2/6/2008	Monday	7:00:00 AM	00 Hours 35 Mins	To	72.923 19.273
7/7/2008	Monday	6:30:00 AM	2 Hours 05 Mins	From	72.947 19.279
26/07/2008	Saturday	8:40:00 AM	00 Hours 40 Mins	From	72.971 19.268
28/07/2008	Monday	11:00:00 AM	4 Hours 5 Mins	From	72.976 19.241
21/08/2008	Thursday	9:30:00 AM	10 Hours 15 Mins	From	72.98 19.232
1/9/2008	Monday	10:20:00 AM	13 Hours 00 Mins	From	72.983 19.224
2/9/2008	Tuesday	8:40:00 AM	1 Hours 45 Mins	From	72.981 19.194
13/09/2008	Saturday	8:00:00 AM	1 Hours 25 Mins	From	72.971 19.268
22/09/2008	Monday	9:40:00 AM	2 Hours 35 Mins	From	72.976 19.241
29/09/2008	Monday	10:20:00 AM	3 Hours 20 Mins	From	72.983 19.224
3/10/2008	Friday	9:00:00 PM	1 Hours 05 Mins	From	72.984 19.22
14/10/2008	Tuesday	8:00:00 AM	5 Hours 00 Mins	To	72.981 19.194
27/10/2008	Friday	5:30:00 PM	1 Hours 5 Mins	From	72.892 19.29
12/11/2008	Wednesday	12:40:00 PM	00 Hours 15 Mins	To	72.947 19.281
17/11/2008	Monday	5:10:00 AM	2 Hours 50 Mins	From	72.984 19.22
12/12/2008	Friday	10:40:00 AM	1 Hours 00 Mins	From	72.922 19.271
29/12/2008	Monday	11:10:00 AM	3 Hours 15 Mins	To	72.942 19.283
4/1/2009	Sunday	12:02:00 PM	5 Hours 15 Mins	To	72.968 19.267
6/2/2009	Friday	8:00:00 AM	2 Hours 15 Mins	From	72.976 19.241
12/4/2009	Sunday	10:15:05 PM	2 Hours 35 Mins	From	72.984 19.22
13/4/2009	Monday	9:00:00 AM	1 Hours 15 Mins	From	72.981 19.194
2/5/2009	Saturday	6:20:09 PM	2 Hours 45 Mins	To	72.968 19.267
17/5/2009	Sunday	3:00:12 PM	00 Hours 35 Mins	To	72.971 19.268
23/05/2009	Saturday	11:10:00 AM	10 Hours 15 Mins	From	72.98 19.232
1/7/2009	Wednesday	1:20:00 PM	2 Hours 35 Mins	To	72.937 19.283
23/7/2009	Thursday	12:20:00 PM	3 Hours 20 Mins	From	72.937 19.279
17/8/2009	Monday	2:00:00 PM	00 Hours 30 Mins	To	72.968 19.267
18/8/2009	Tuesday	9:00:00 PM	4 Hours 00 Mins	To	72.984 19.22
26/8/2009	Wednesday	11:20:00 AM	1 Hours 20 Mins	From	72.894 19.281
23/09/2009	Wednesday	8:00:00 AM	2 Hours 03 Mins	From	72.892 19.29
30/10/2009	Friday	7:30:00 AM	1 Hours 05 Mins	From	72.921 19.277
26/11/2009	Thursday	11:00:00 AM	00 Hours 15 Mins	From	72.922 19.273
11/12/2009	Friday	12:20:00 PM	1 Hours 00 Mins	From	72.923 19.273
26/12/2009	Saturday	8:00:00 AM	4 Hours 45 Mins	From	72.922 19.271
30/12/2009	Wednesday	9:40:00 AM	1 Hours 45 Mins	From	72.942 19.283

Driver Analysis

Age	Sex	Tenure (in yrs)	Accident Nature
35	M	15	Fatal
40	M	11	Non-Fatal
55	M	22	Non-Fatal
30	F	5	Non-Fatal
32	M	7	Fatal
31	M	10	Non-Fatal
39	M	13	Non-Fatal
38	M	20	Non-Fatal
42	M	26	Fatal
45	M	12	Non-Fatal
20	M	0.75	Non-Fatal
60	M	25	Non-Fatal
31	M	16	Fatal
25	M	5	Fatal
42	M	18	Non-Fatal
36	M	7	Non-Fatal
36	M	14	Non-Fatal
49	M	19	Fatal
26	F	2	Non-Fatal
36	M	5	Non-Fatal
29	M	9	Non-Fatal
30	M	12	Fatal
48	M	15	Non-Fatal
40	M	15	Non-Fatal
50	M	30	Non-Fatal
33	M	11	Non-Fatal
29	M	10	Non-Fatal
28	M	10	Non-Fatal
30	F	8	Non-Fatal
39	M	16	Non-Fatal
45	F	10	Fatal
42	M	21	Non-Fatal
49	M	22	Non-Fatal
47	M	12	Non-Fatal
26	M	4	Non-Fatal
40	M	21	Non-Fatal
32	M	1	Non-Fatal
41	M	15	Non-Fatal
39	M	15	Non-Fatal
36	M	17	Non-Fatal
55	M	22	Non-Fatal
48	M	19	Fatal
26	M	3	Non-Fatal
31	M	6	Non-Fatal
24	F	3	Non-Fatal
35	F	0.5	Non-Fatal
40	M	17	Non-Fatal
36	M	12	Non-Fatal
45	M	13	Fatal
35	M	15	Non-Fatal
33	M	9	Non-Fatal
34	M	5	Non-Fatal
25	F	1	Non-Fatal
37	M	10	Non-Fatal
27	M	3	Non-Fatal
29	M	9	Non-Fatal
39	M	15	Fatal
38	M	14	Non-Fatal
36	M	6	Non-Fatal
40	F	6	Fatal

The Database for year 2009-11 to carry out Microsimulation

2009-2011

Analysis of time to show the time of day, when accident occurs

Date	Day	Time of accident	Duration of Driving before accident	Direction of Travel
16/1/2009	Friday	8:00:00 AM	3 Hours 45 Mins	From
29/1/2009	Thursday	7:30:00 AM	4 Hours 45 Mins	To
5/2/2009	Thursday	3:40:00 PM	5 Hours 45 Mins	From
19/2/2009	Thursday	10:30:00 AM	6 Hours 45 Mins	From
28/2/2009	Saturday	10:20:00 AM	7 Hours 45 Mins	From
3/4/2009	Friday	11:20:00 AM	8 Hours 45 Mins	From
16/4/2009	Thursday	11:00:00 AM	9 Hours 45 Mins	From
22/4/2009	Wednesday	10:10:00 AM	10 Hours 45 Mins	From
2/5/2009	Saturday	7:30:00 AM	1 Hours 05 Mins	From
11/5/2009	Monday	9:00:00 AM	00 Hours 35 Mins	From
13/5/2009	Wednesday	6:20:09 PM	5 Hours 15 Mins	From
19/5/2009	Tuesday	3:00:12 PM	2 Hours 15 Mins	To
28/5/2009	Thursday	11:10:00 AM	2 Hours 35 Mins	To
9/6/2009	Tuesday	1:20:00 PM	1 Hours 15 Mins	To
21/6/2009	Sunday	8:00:00 AM	13 Hours 00 Mins	From
13/8/2009	Thursday	9:40:00 AM	1 Hours 45 Mins	From
31/08/2009	Monday	10:20:00 AM	1 Hours 25 Mins	To
15/9/2009	Tuesday	9:00:00 AM	2 Hours 35 Mins	To
23/9/2009	Wednesday	8:00:00 AM	3 Hours 20 Mins	From
1/10/2009	Thursday	5:30:00 PM	1 Hours 05 Mins	From
9/10/2009	Friday	8:00:00 AM	00 Hours 15 Mins	To
15/10/2009	Thursday	9:40:00 AM	1 Hours 00 Mins	From
26/10/2009	Monday	10:40:00 AM	4 Hours 45 Mins	From
27/10/2009	Tuesday	11:40:00 AM	1 Hours 45 Mins	From
3/11/2009	Tuesday	8:40:00 AM	00 Hours 35 Mins	From
13/11/2009	Friday	11:00:00 AM	10 Hours 15 Mins	To
21/11/2009	Saturday	6:30:00 AM	2 Hours 05 Mins	To
29/11/2009	Sunday	8:00:00 PM	00 Hours 40 Mins	To
30/11/2009	Monday	11:35:00 AM	4 Hours 5 Mins	To
2/12/2009	Wednesday	10:00:00 AM	10 Hours 15 Mins	From
14/12/2009	Monday	11:00:00 AM	4 Hours 45 Mins	From
21/12/2009	Monday	12:20:00 PM	1 Hours 45 Mins	From
22/12/2009	Tuesday	11:10:00 AM	3 Hours 10 Mins	From
26/12/2009	Saturday	12:02:00 PM	1 Hours 5 Mins	From
27/12/2009	Sunday	8:00:00 AM	00 Hours 15 Mins	From
5/1/2010	Tuesday	5:25:00 AM	1 Hours 05 Mins	From
8/1/2010	Friday	6:00:00 PM	5 Hours 00 Mins	From
16/1/2010	Saturday	2:10:00 PM	10 Hours 5 Mins	From
23/1/2010	Saturday	12:40:00 PM	00 Hours 10 Mins	From
2/2/2010	Tuesday	5:10:00 AM	1 Hours 00 Mins	From
10/2/2010	Wednesday	10:00:00 PM	1 Hours 35 Mins	From
14/2/2010	Sunday	9:40:00 AM	1 Hours 15 Mins	From
20/2/2010	Saturday	10:35:00 PM	2 Hours 35 Mins	To
24/2/2010	Wednesday	11:40:00 AM	3 Hours 20 Mins	To
5/3/2010	Friday	10:15:05 PM	00 Hours 30 Mins	To
13/3/2010	Saturday	2:00:00 PM	1 Hours 20 Mins	To
16/3/2010	Tuesday	9:00:00 AM	2 Hours 45 Mins	From
19/3/2010	Friday	11:20:00 AM	3 Hours 00 Mins	From
25/3/2010	Thursday	5:20:00 PM	00 Hours 10 Mins	To
31/3/2010	Wednesday	1:20:00 PM	2 Hours 45 Mins	To
5/4/2010	Monday	12:20:00 PM	1 Hours 45 Mins	From
11/4/2010	Sunday	9:00:00 AM	2 Hours 35 Mins	From
19/4/2010	Monday	6:20:09 AM	4 Hours 00 Mins	From
30/4/2010	Friday	3:00:12 PM	1 Hours 20 Mins	To
5/5/2010	Wednesday	8:00:00 AM	2 Hours 03 Mins	To
16/5/2010	Sunday	7:30:00 AM	1 Hours 05 Mins	To
23/5/2010	Sunday	11:00:00 AM	00 Hours 35 Mins	To
7/6/2010	Monday	12:20:00 PM	5 Hours 15 Mins	From
18/6/2010	Friday	9:00:00 AM	6 Hours 00 Mins	From
27/6/2010	Sunday	8:00:00 AM	00 Hours 10 Mins	To
3/7/2010	Saturday	5:30:00 PM	2 Hours 45 Mins	To
11/7/2010	Sunday	8:00:00 AM	1 Hours 45 Mins	From
13/8/2010	Friday	2:50:00 PM	3 Hours 15 Mins	From
29/9/2010	Wednesday	11:10:00 AM	10 Hours 45 Mins	From
17/10/2010	Sunday	2:40:00 PM	4 Hours 45 Mins	To
23/10/2010	Saturday	3:40:00 PM	5 Hours 45 Mins	To
30/10/2010	Saturday	4:40:00 PM	6 Hours 45 Mins	To
5/11/2010	Friday	2:10:00 PM	7 Hours 35 Mins	From
24/11/2010	Wednesday	3:00:00 PM	2 Hours 45 Mins	From
25/11/2010	Thursday	7:40:00 PM	14 Hours 45 Mins	From
3/12/2010	Friday	8:15:00 PM	15 Hours 45 Mins	From
17/12/2010	Friday	3:40:00 PM	2 Hours 20 Mins	To
19/12/2010	Sunday	10:30:00 AM	2 Hours 50 Mins	To
23/12/2010	Thursday	11:15:00 AM	1 Hours 00 Mins	To
26/12/2010	Sunday	12:40:00 AM	3 Hours 15 Mins	From

Driver Analysis

Age	Sex	Tenure (in yrs)	Accident Nature
35	M	15	Non-Fatal
40	M	11	Fatal
55	M	22	Non-Fatal
30	F	5	Non-Fatal
32	M	7	Non-Fatal
31	M	10	Fatal
39	M	13	Non-Fatal
38	F	20	Non-Fatal
42	M	26	Non-Fatal
45	M	12	Fatal
20	F	0.75	Non-Fatal
60	M	25	Non-Fatal
31	M	16	Fatal
25	M	5	Non-Fatal
42	F	18	Non-Fatal
36	M	7	Non-Fatal
36	M	14	Fatal
49	M	19	Non-Fatal
26	F	2	Non-Fatal
36	M	5	Non-Fatal
29	M	9	Fatal
30	M	12	Non-Fatal
48	M	15	Non-Fatal
40	F	15	Non-Fatal
50	F	30	Non-Fatal
33	M	11	Non-Fatal
29	M	10	Non-Fatal
28	M	10	Fatal
30	F	8	Non-Fatal
39	M	16	Non-Fatal
45	F	10	Non-Fatal
42	M	21	Non-Fatal
49	M	22	Non-Fatal
47	M	12	Fatal
26	M	4	Non-Fatal
40	M	21	Fatal
32	M	1	Non-Fatal
41	F	15	Non-Fatal
39	F	15	Fatal
36	M	17	Non-Fatal
55	M	22	Non-Fatal
48	M	19	Non-Fatal
26	M	3	Fatal
31	M	6	Non-Fatal
24	F	3	Non-Fatal
35	F	0.5	Fatal
40	M	17	Non-Fatal
36	M	12	Non-Fatal
45	M	13	Fatal
35	F	15	Non-Fatal
33	M	9	Non-Fatal
34	M	5	Fatal
25	F	1	Non-Fatal
37	M	10	Non-Fatal
27	M	3	Fatal
29	M	9	Non-Fatal
39	M	15	Non-Fatal
38	M	14	Fatal
36	M	6	Non-Fatal
40	F	6	Fatal
35	F	0.5	Non-Fatal
40	M	17	Fatal
36	M	12	Non-Fatal
45	M	13	Fatal
35	F	15	Non-Fatal
33	M	9	Non-Fatal
34	M	5	Fatal
25	F	1	Non-Fatal
37	M	10	Non-Fatal
27	M	3	Fatal
29	F	9	Non-Fatal
39	M	15	Non-Fatal
38	M	14	Fatal
36	M	6	Non-Fatal
40	F	6	Non-Fatal

2. **Types of Data**

➤ Cross sectional Data: Cross sectional data is data that has been collected at a particular point in time. The fact that it is collected at a point in time means that it provides a `snapshot' of the population from which the data is derived. Each sub section deals with the potential sources of cross-sectional data that can be used to address the previously mentioned research questions. Cross sectional data is particularly useful for assessing metrics at a point in time such as the most frequent occurrence in road accidents. This can be extended to monitor the trends between several points in time by using repeated cross sectional measurements. The disadvantage of cross sectional data is that because it is a `snapshot' the temporality of relationships cannot be established, for example did accident get affected by the duration of travel before accident or vice versa. So causality is difficult to apportion with this type of data and this exercise is usually best avoided as the results are probably going to be misleading rather than informative and objective on such a question. There is also bias associated with this type of data as those who do not respond or are missed may be systematically different from those who do respond, for example respondents in a survey. On balance, cross sectional data usually provide a wealth of information which can be used for many aspects of public health research, road accident analysis etc.

➤ The Demographic Data: The Statistical analysis of accident is carried out periodically at critical locations or road stretches which will help to arrive at suitable measures to effectively decrease accident rates. It is the measure (or estimates) of the number and severity of accident. Accident prone stretches of different roads may be assessed by finding the accident density per length of the road. The places of accidents are marked on the map and the points of their clustering **(black spot)** are determined. By statistical study of accident occurrence at a particular road or location or zone of study for a long period of time it is possible to predict with reasonable accuracy the probability of accident occurrence per day or relative safety of different classes of road user in that location. The interpretation of the statistical data is very important to provide insight to the problem.

COUNTRY-WISE PERSONS KILLED

TOTAL ROAD ACCIDENTS

ACCIDENT SEVERITY

DISTRIBUTION OF TOTAL NUMBER OF ACCIDENTS

PERSONS KILLED IN ROAD ACCIDENTS

B. The Cross Sectional Data and Regression for the Microsimulative Study

In all our statistical work to date, we have been dealing with analyses of time-ordered data, or time series: the same variable or variables observed and measured at consecutive points of time. Usually but not necessarily, the points of time are equally spaced. Time-ordered data are very often pertinent for total quality; for example, we need to know whether our processes are in statistical control or whether they are being affected by, say, trends or special causes. We need also to evaluate the effectiveness of interventions aimed at improving our processes and to assure that we are holding the gains from effective interventions from the past. But not all data are time-ordered. There is also a type of data called cross-sectional data, where we are dealing with information about different individuals (or aggregates such as work teams, sales territories, stores, etc.) at the same point of time or during the same time period. For example, we might have data on total accidents per worker over the course of the last calendar year for all the workers in a given plant, or we might have questionnaire data on customer satisfaction for a sample of customers last month.

C. The RAYSIM Model:

➢ Introduction:

The Road Accident analysis Simulative Model (RAYSIM), works on the fundamental concepts of GIS, where the integration of GIS Data with microdata is done. The analysis of the GIS data is done with respect to the Hotspots and Coldspots for the accidents on the Thane Ghodbandar Road. The RAYSIM works on the principle of Linear Regression and Reweighted Methods.

RAYSIM takes into account the variables, and computes the results based on the weights assigned to them. The basic step for the successful running of the Model is conceptualization of a technique, which works on different hypothesis generated out of the various parameters considered.

It starts with the first Hypothesis, and after testing all the hypothesis, and integration of them all is done, to generate a common output, satiating the requirements of this project, to assess the risk factor associated.The RAYSIM in its initial stages, works on the Synthetic Data, which can anytime be substituted

with Real Data, to derive a real time result. The main drawback of this entire work is lack availability of data, of the required nature.

➢ How RAYSIM Works?

It works basically with the attributes of the Data, outputs from the Hotspots analysis and the Synthetic Database generated. The platform to carry out this is ArcGIS 10 and Microsoft Excel. The data in the form of Shapefiles are input into ArcGIS. There the Hotspots and network analysis are performed whose attributes are further input into the RAYSIM. The inputs are initially joined to the synthetic database, using the Joins option from the ArcGIS attribute database.

On these data the Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) are carried out. On the outputs obtained from these analyses, a weighted liner regression is carried out, for the Bivariate and Multivariate conditions.

These regression analyses are further analyzed for the Temporal Analysis and Physicopsychological Analysis. Further different hypothesis are generated, which are tested according to the regression formula, data availability and the analysis.

The results of these tests, for the different Hypothesis, will be associated together, to derive a correct Microsimulative approach and analysis to accident assessment on the Roads.

➢ The Working:

- Data Required:

  After the Hotspots analyses were carried out, the attribute from the layer generated was used to carry out the working of the RAYSIM. The attributes were chosen according to the layers, which were the best fit for the Bivariate and Multivariate Linear Regression Analysis.

- Methodologies Used:

  i. Hotspots Analysis: Discussed in Chapter ---------------

  ii. Ordinary Least Squares: In statistics, ordinary least squares (OLS) or linear least squares is a method for estimating the

unknown parameters in a linear regression model. This method minimizes the sum of squared vertical distances between the observed responses in the dataset and the responses predicted by the linear approximation. The resulting estimator can be expressed by a simple formula, especially in the case of a single regressor on the right-hand side. The OLS estimator is consistent when the regressors are exogenous and there is no perfect multicollinearity, and optimal in the class of linear unbiased estimators when the errors are homoscedastic and serially uncorrelated. Under these conditions, the method of OLS provides minimum-variance mean-unbiased estimation when the errors have finite variances. Under the additional assumption that the errors be normally distributed, OLS is the maximum likelihood estimator.
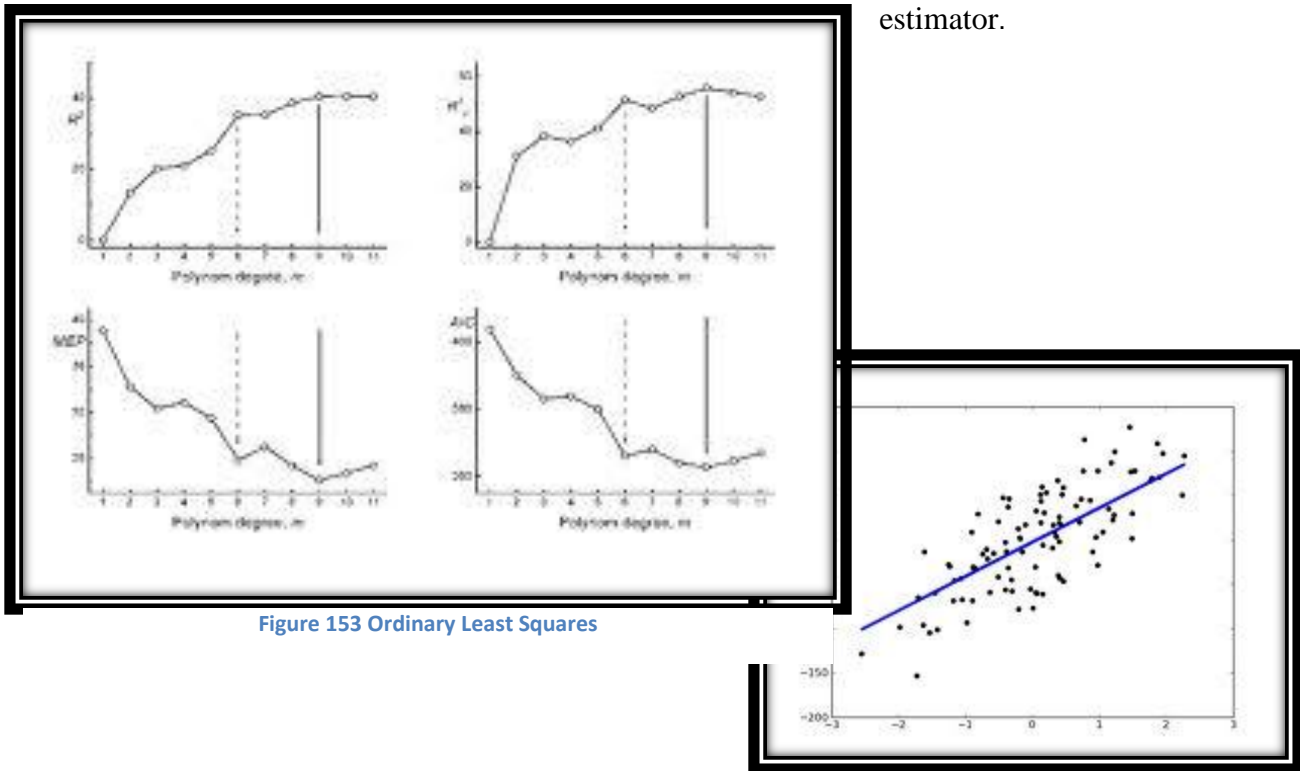
Figure 153 Ordinary Least Squares

iii. Geographically Weighted Regression: GWR is a relatively simple technique that extends the traditional regression framework of equation (1) by allowing local variations in rates of change so that the coefficients in the model rather than being global estimates are specific to a location i. The regression equation is then

$$ y_i = a_{i0} + \sum_{k=l,m} + \text{\euro} $$

Where, a_{ik} is the value of the kth parameter at location i. It is very easy with GWR to compute parameter estimates, for instance for locations that lie between data points, which makes it possible to produce detailed maps of spatial variations in relationships. In GWR, weighting an observation in accordance with its proximity to i would allow an estimation of aik to be made that meets the criterion of "closeness of calibration points."

iv. Bivariate Linear Regression: The simplest regression model is the bivariate one, in which there is one response or dependent variable, and one predictor or independent variable, and the relationship between the two is represented by a straight line. Building a bivariate linear regression model to represent the relationship between two variables by a straight line involves determining the coefficients of that line, a process known as "fitting" the regression line.

v. Multivariate Linear Regression: Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent variable y.

The population regression line for p explanatory variables x1, x2, ... , xp is defined to be

$$\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p.$$

This line describes how the mean response μ_y changes with the explanatory variables. The observed values for y vary about their means μ_y and are assumed to have the same standard deviation σ. The fitted values b0, b1, ..., bp estimate the parameters $\beta_0, \beta_1, ..., \beta_p$ of the population regression line.

Since the observed values for y vary about their means μ_y, the multiple regression model includes a term for this variation. In words, the model is expressed as DATA = FIT + RESIDUAL, where the "FIT" term represents the expression $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_p x_p$. The "RESIDUAL" term represents the deviations of the observed values y from their means μ_y, which are normally distributed with mean 0 and variance σ. The notation for the model deviations is ε.

Formally, the model for multiple linear regression, given n observations, is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... \beta_p x_{ip} + \varepsilon_i \text{ for } i = 1,2, ... n.$$

In the least-squares model, the best-fitting line for the observed data is calculated by minimizing the sum of the squares of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0). Because the deviations are first squared, then summed, there are no cancellations between positive and negative values. The least-

228

squares estimates b0, b1, ... bp are usually computed by statistical software.

The values fit by the equation $b0 + b1xi1 + ... + bpxip$ are denoted $\hat{y}$ i, and the residuals ei are equal to yi - $\hat{y}$ i, the difference between the observed and fitted values. The sum of the residuals is equal to zero.

The variance σ^2 may be estimated by

$$s^2 = \frac{\sum e_i^2}{n - p - 1}$$

also known as the mean-squared error (or MSE).
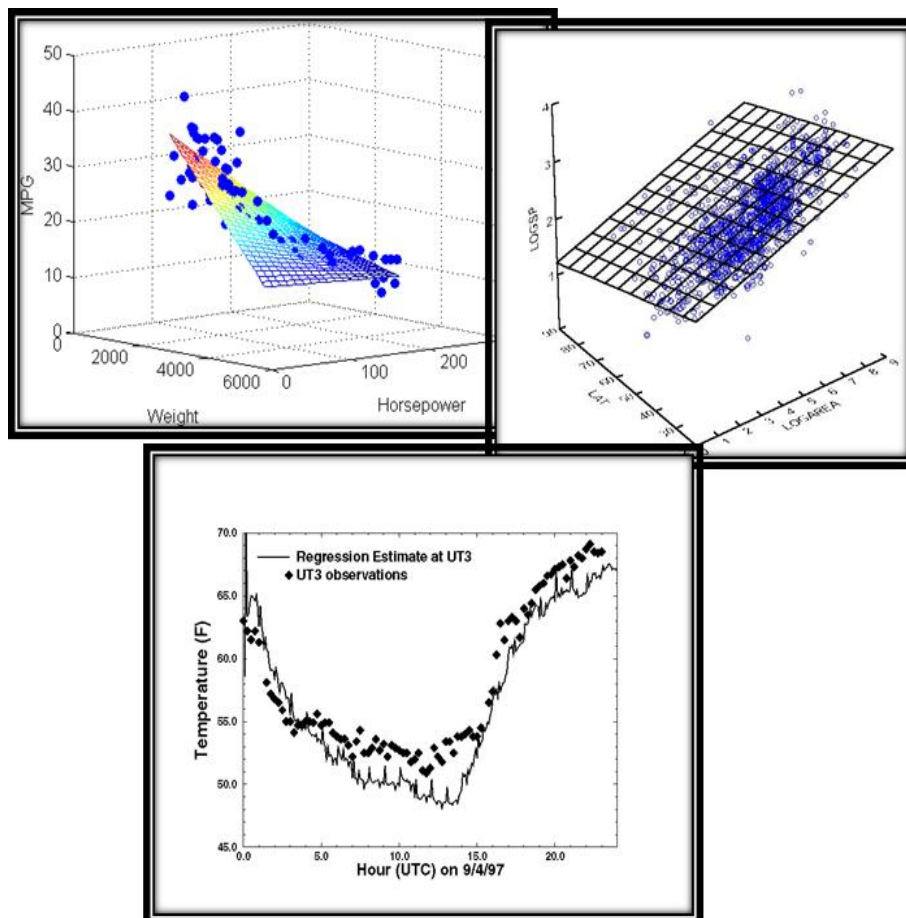
Figure 154 Multi-variate Regression

- <u>The Hypothesis Generation Task</u>: Problem structuring is a pre-decision process by which the decision maker develops the salient characteristics of the decision problem. The decision maker must first develop the objectives and constraints of the decision problem. Once the over-all objectives are formulated, various structural elements are supplied. Structural elements may include: possible acts which are specified by the decision maker, relevant states of the hypotheses, and possible outcomes. Outcomes are determined by both, the act that the decision maker chooses and the state of the world that obtains when that act occurs. The hypothesis set that the decision maker creates should contain plausible hypotheses. The construct of "plausibility" includes the notion that for a hypothesis to be included in the set of hypotheses it should be sufficiently probable to be worth further analysis. This does not necessarily involve an assessment process as detailed and thorough as is typically implied by the term "probability assessment." All that is logically necessary at the early stages of problem structuring is that the decision maker makes a rough "go/no go" decision in regard to each hypothesis. Hypotheses that pass this crude plausibility test may be more carefully assessed in later stages of decision analysis. While it is possible that plausibility assessment and probability assessment share common elements, there are a few clear differences. The first major difference is in the nature of the task requirements. In a probability assessment task, assessments are usually made about the relative likelihood of a set of specified hypotheses known to the decision maker. In a hypothesis generation task, hypotheses are evaluated with respect to whether or not they should be considered further. This evaluation is complicated by the fact that the evaluation should be relative to both previously-specified hypotheses that the decision maker may have and unspecified hypotheses that are yet to be generated by the decision maker. Hypothesis generation tasks also have the characteristic that generated hypotheses should be consistent with any available

information. This information may be specific data or knowledge about the task. Obviously, hypotheses that are inconsistent with the available evidence should not be considered. Information provided by data and the task has a second important role, since it serves as a basis for the memory search processes described in the next section. Although the emphasis will be on memory search processes, the importance of the data as constraints to the logical possibility of hypotheses should be kept in mind.
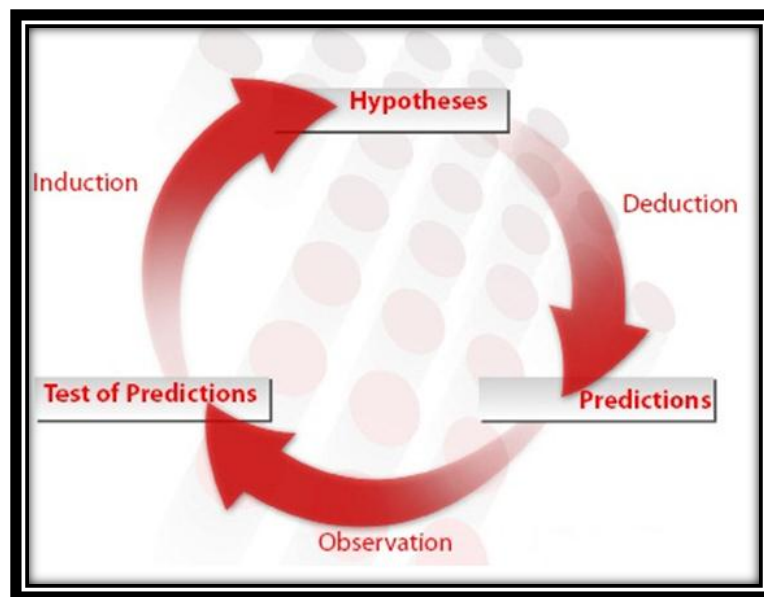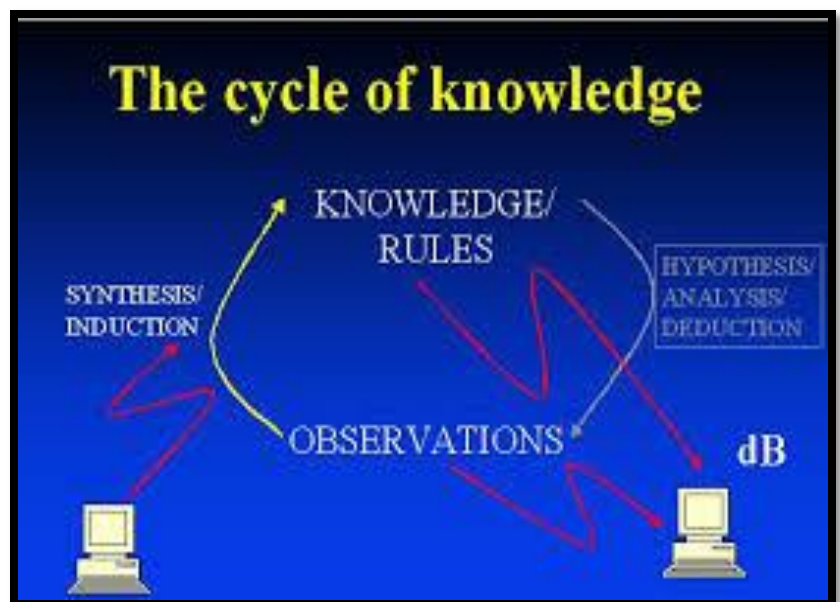
i. <u>Hypothesis Testing:</u> Hypothesis testing begins with an assumption called Hypothesis., that we make about a population parameter. A hypothesis is a supposition made as a basis for reasoning. According to Prof Morris Hamburg, " A hypothesis in statistics is simply a quantitative statement about a population." Palmer O Johnson beautifully describes a hypothesis as "islands in the uncharted seas of thought to be used as bases for consolidation as recuperation as we advance into the unknown."

ii. <u>Procedure for Testing Hypothesis:</u>
   - ➢ Set up a Hypothesis: The first thing to do is to set up a hypothesis about a population parameter. The two hypothesis in a statistical test are normally referred to as: Null Hypothesis & Alternative Hypothesis.
   - ➢ Set up a suitable significance level: The significance level is customarily expressed as a percentage, such as 5% is the probability of rejecting the null hypothesis, if it is true, The confidence with which an experimenter rejects or retains a hypothesis depends upon the significance level adopted.
   - ➢ Setting a test Criterion: This involves selecting an appropriate probability distribution for the particular test .i.e., probability distribution which can properly be applied. Test criteria must apply an appropriate probability distribution.
   - ➢ Doing Computations: Having taken the first three steps, we have completely designed a statistical test. Now we proceed to the performance of the various computations, from a random sample of size n, necessary for the test. These calculations include the testing statistic and the standard error of the testing statistic.
   - ➢ Making decisions: Finally, we may draw, statistical conclusions and take decisions. A statistical conclusion or statistical decision is a decision either to reject or to accept the null hypothesis. The decision will depend on whether the computed value of the test criterion falls in the region of rejection or the region of acceptance.

iii. <u>The Hypothesis :</u>
   a) *"**Longer Duration of Driving before an accident leads to higher accident rate.**"*
       The longer the duration of driving, greater is the fatigue, and this factor becomes important when we consider the Physicopsychological aspect of the microdata for

the simulative analysis. The decision making power of any human is also affected by his or her fatigue levels. We have understood, according to the analysis till now that, when considering the reasons and facts behind the road accidents, we are supposed to take into account, the decision making power of the driver as well as his mental state. This level of analysis goes more deep into the biasness towards the driver being solely responsible for the accidents, but as per this project and subject to the constraints of the data and the associated analysis, the mentioned hypothesis was drawn.

.
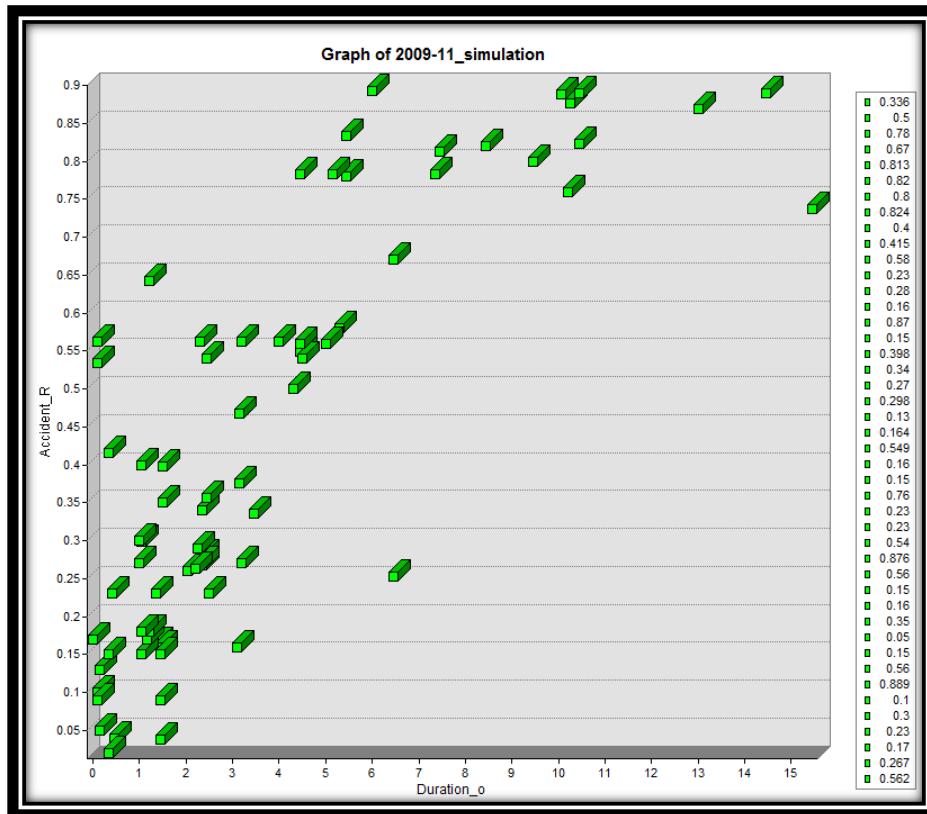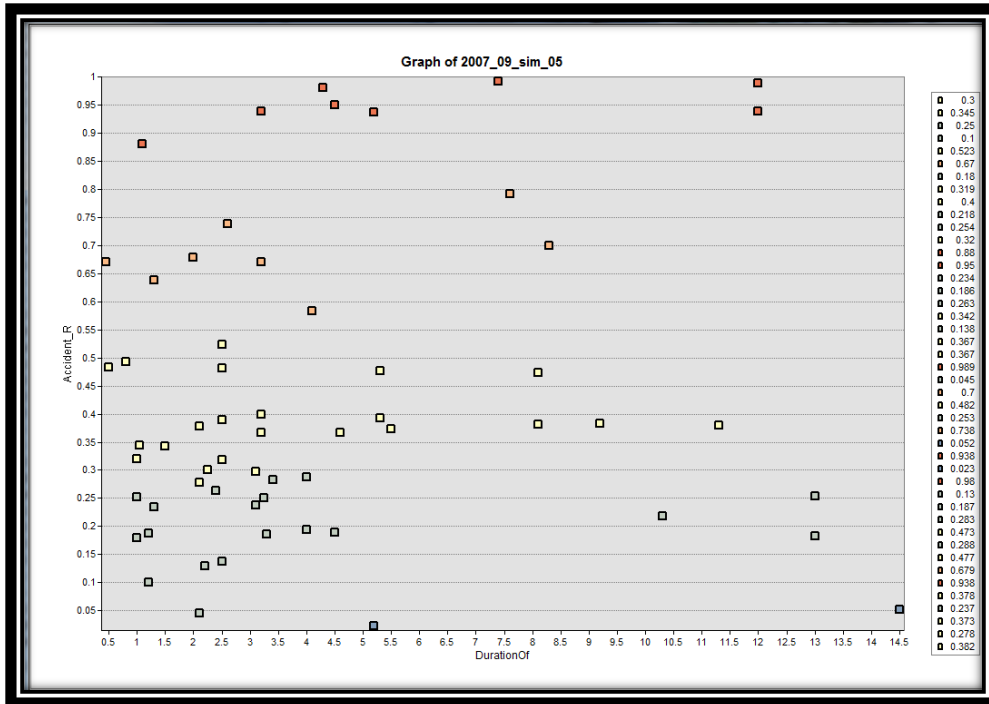
Figure 155: The Scatter Plot between the Duration of travel and the accident rate for 2007-11

.
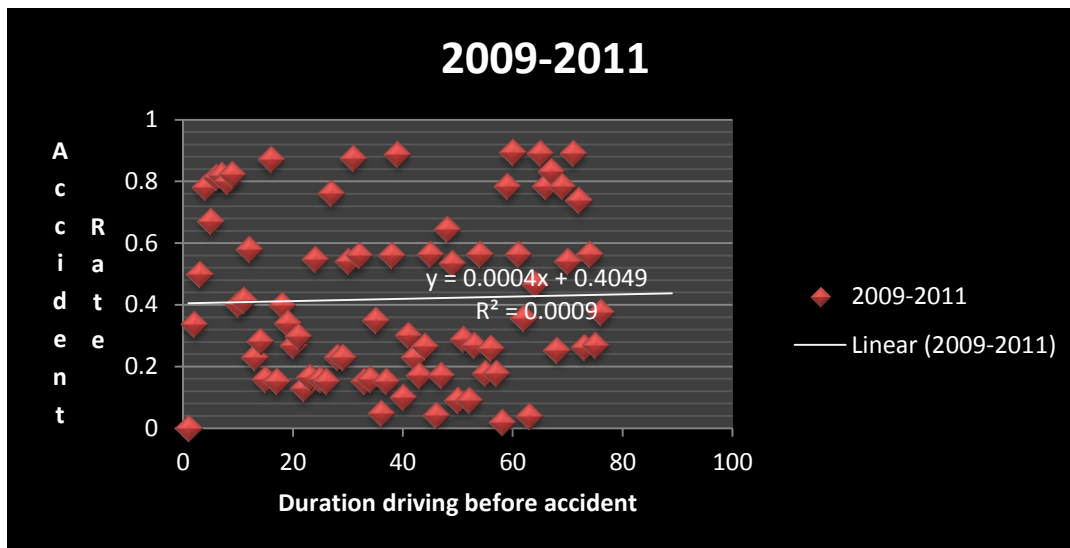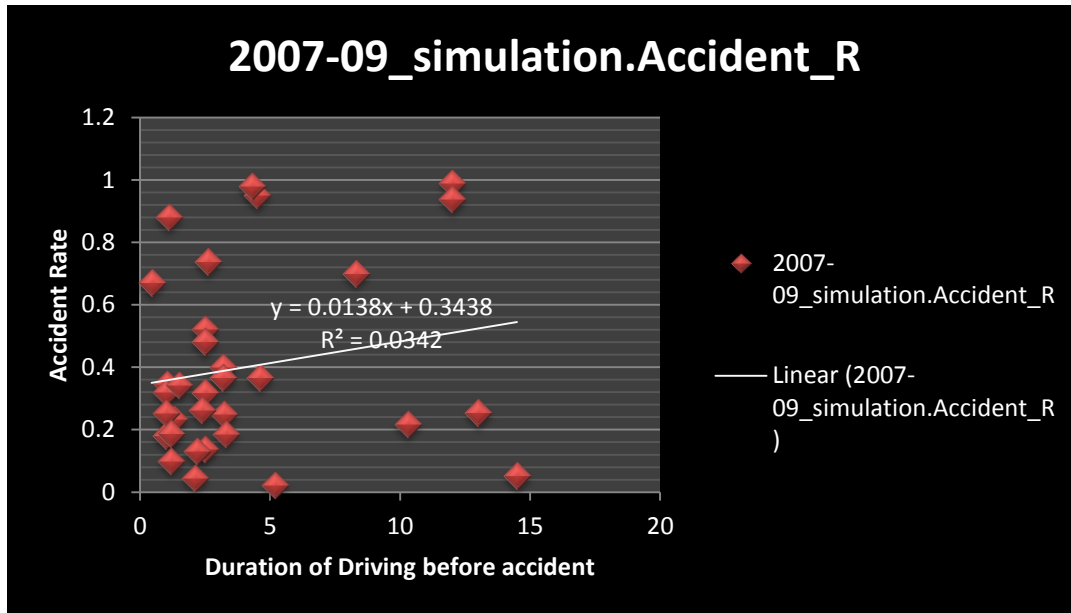
Figure 156 : The Bivariate Linear Regression between Duration of driving and the accident rate

b) *"Maximum accidents occur on Mondays & Fridays."*

A look at the statistical database shows an inclination towards more accidents on Mondays and Fridays, over a span of 4 years .i.e., 2007-11. This has been considered as one of the hypothesis, because of a trend observed in their distribution across the timespan. This also takes into account the effect of the two days on the traffic volume on the Thane Ghodbandar Road .The geographical location of the road explains the reason behind the traffic flow. After the statistical analysis of the duration of driving and the accident rate, we observe that it is 4 hours on an average for the accidents. But precisely when we look into the date, and understand the micro-level impact for the same, we observe that the duration of driving before meeting an accident is 2.5 hours on Mondays and Fridays. This further brings us to a conclusion that, the region within an area of 2.5 hrs*40kmph, i.e. 10 km from Thane Ghodbandar road lies in the risk zone during the weekends and Mondays. A survey study gives us the following reasons for the same:

> ➢ Rash driving on weekends for party
> ➢ Haste to reach home early
> ➢ Some plans for weekend which needs to be met.
> ➢ Lazy driving on Monday
> ➢ Sleepy State
> ➢ Other reasons.
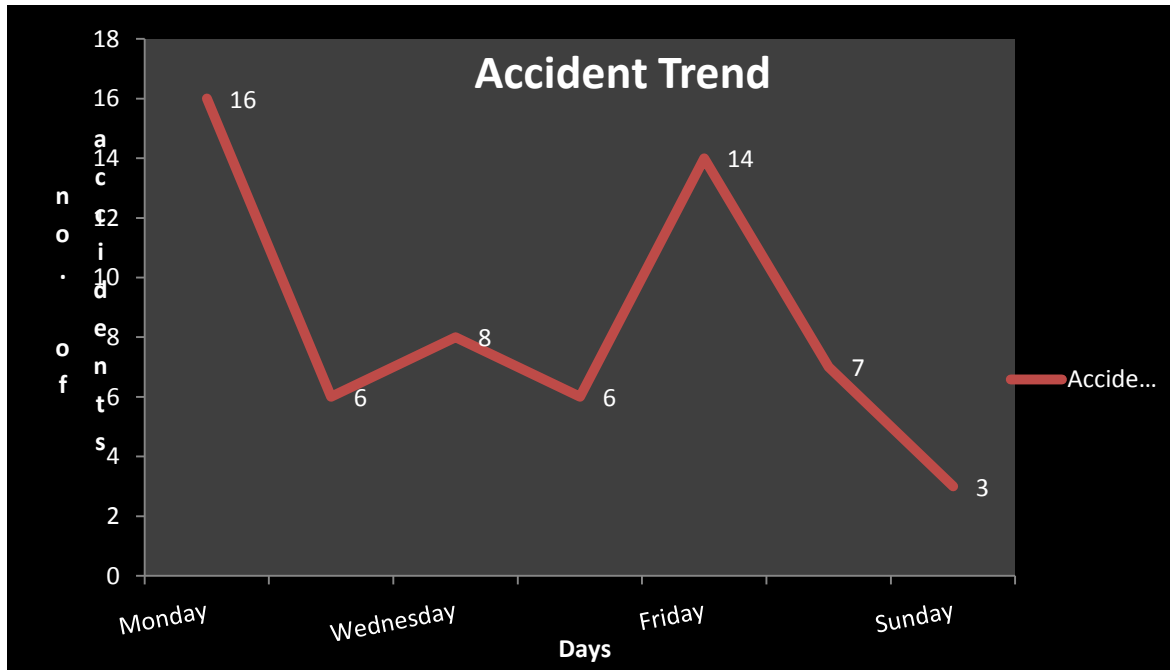
Figure 157 The Representation for the accident rates on weekdays
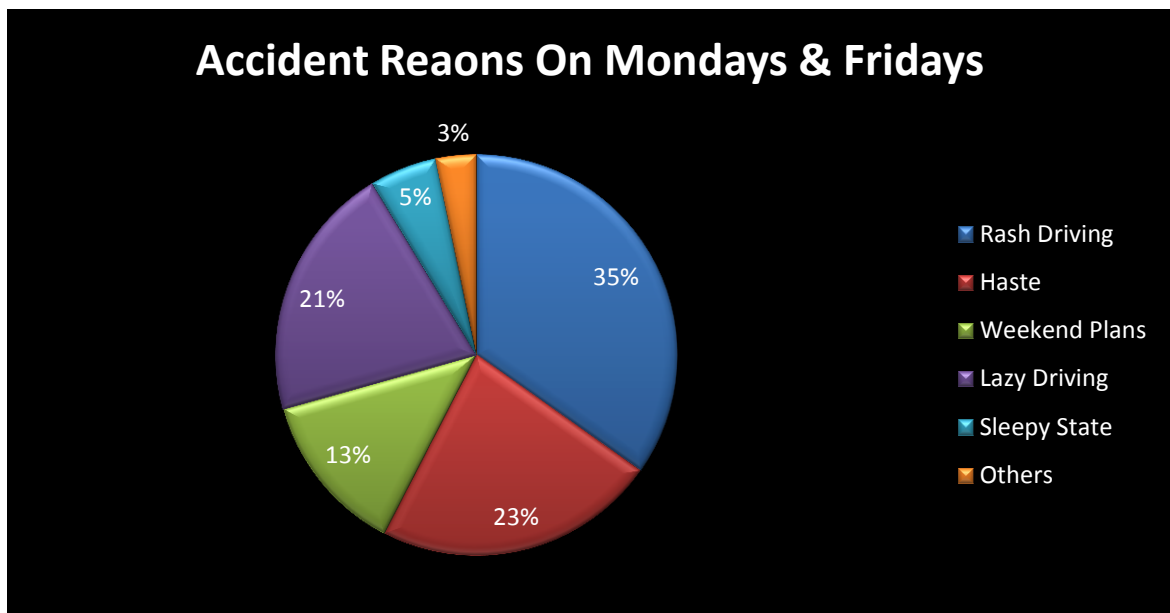
Figure 158 The Representation for the accident reason on Fridays and Mondays
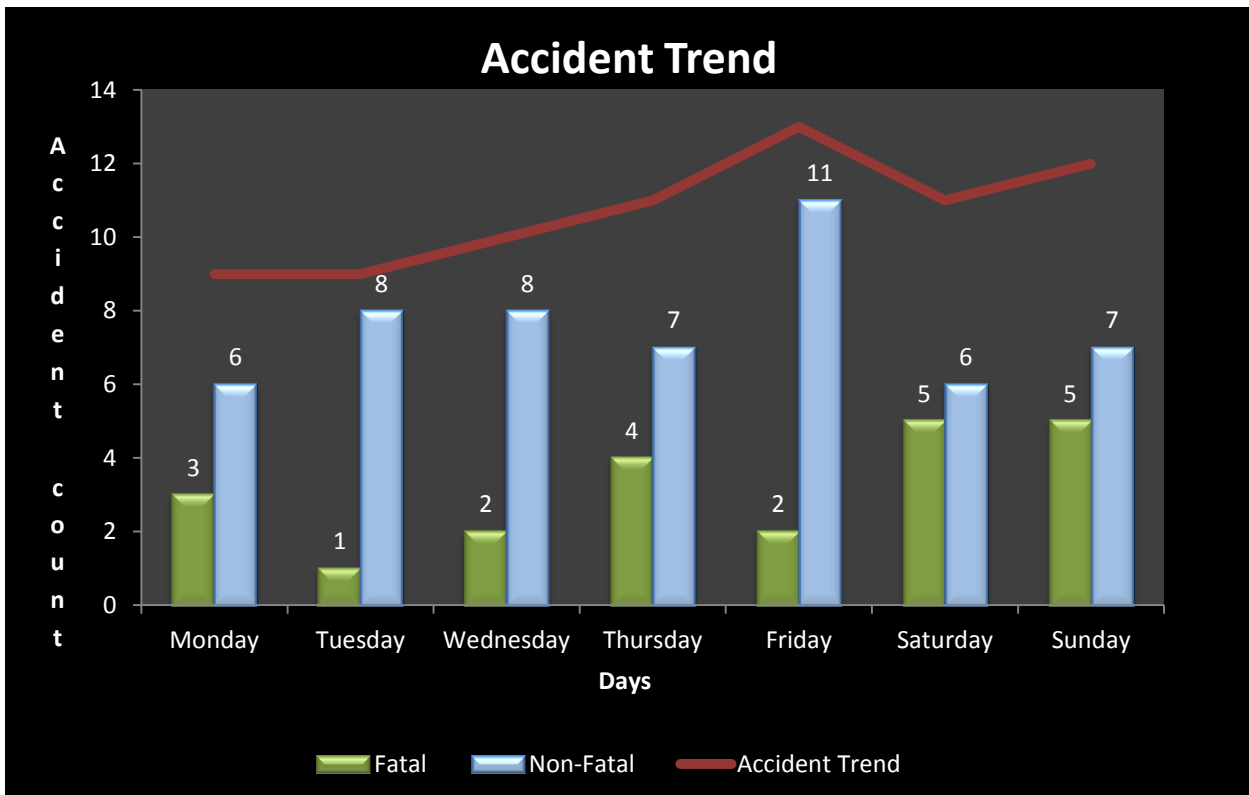
Figure 160 Accident trend and nature over the days for 2009-2011
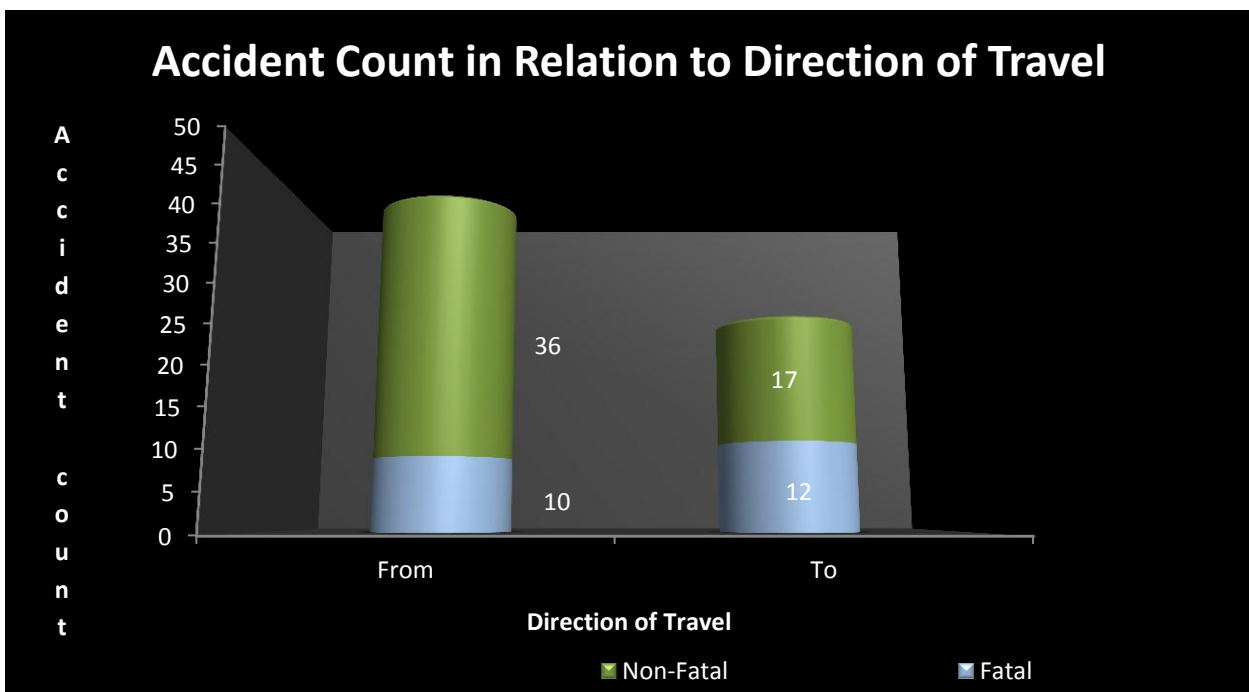
Figure 159 Accident count in relation to DOT
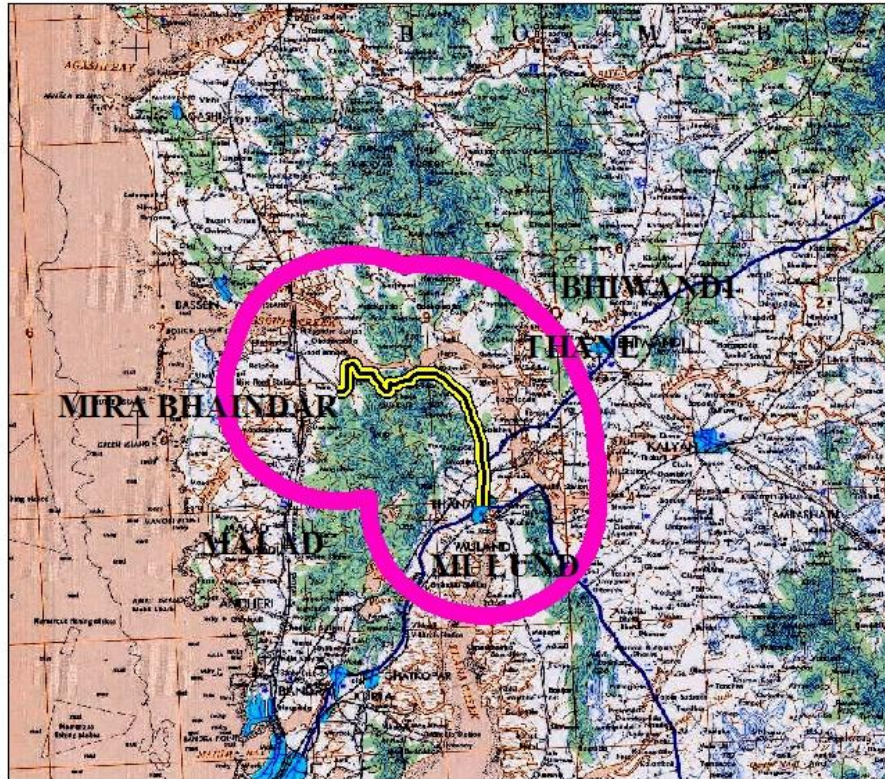
Figure 161 Map Layout for the Accident Prone Areas

c) *"Tenure of holding a driving license directly affects the rate of accidents."* This is again a Physicopsychological parameter, considered, which is relevant, if believing in "Experience is the best teacher".
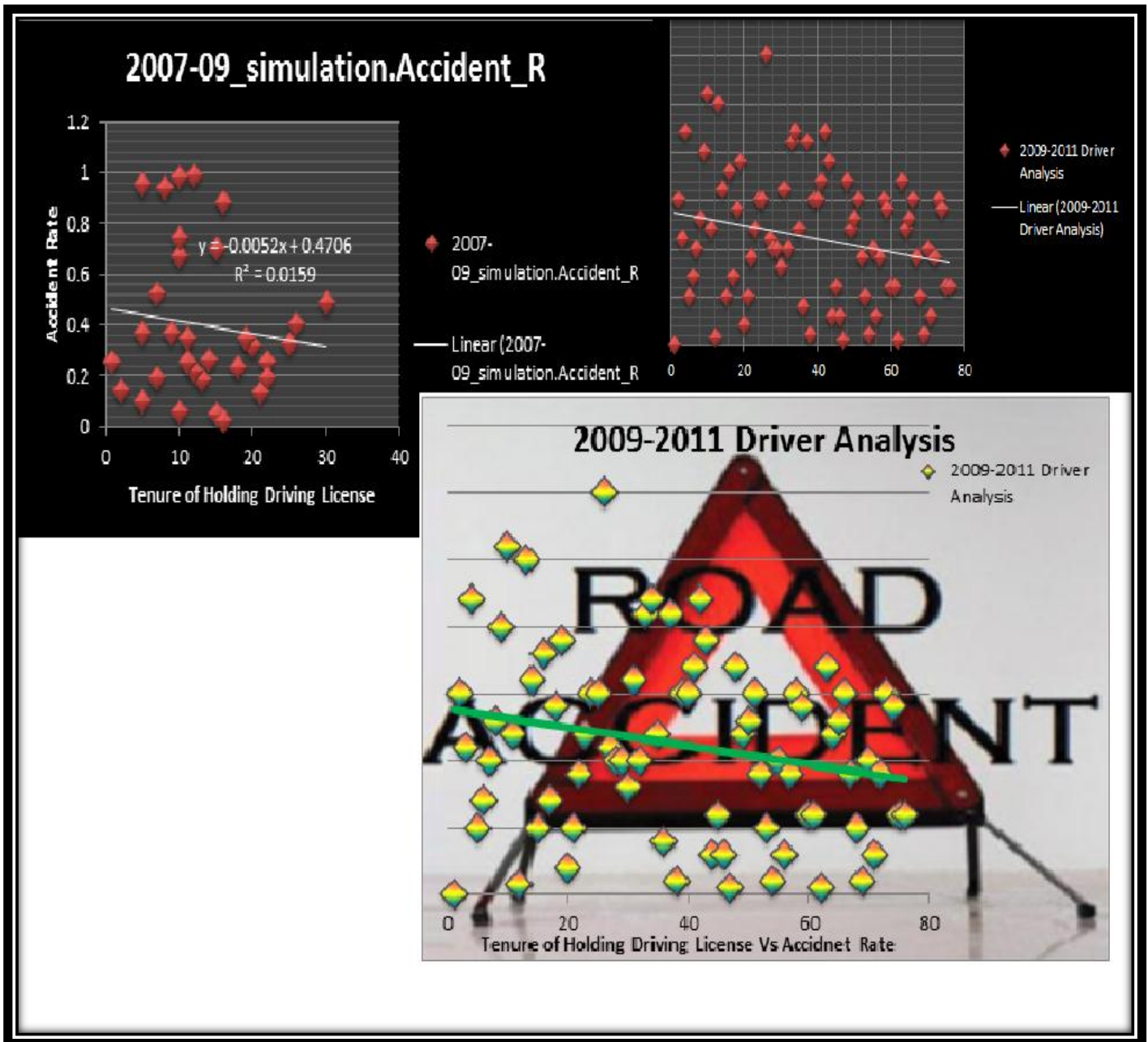
Figure 162 RepresentaEtion of bivariate regression between the Tenure of Holding a Driving License and the Accident Rate.

d) *"The nature of accidents, whether fatal or non- fatal is closely dependent on the date time modifiers."*

Here we see, whether the day of the week or the time of the accident, really play a strong role in the occurrence of accidents.

From the statistical analysis we observe that there is a slight increase in the number of non-fatal accidents on Mondays and Saturdays of the year 2007-09.

But, in the year 2009-11, we observe a different trend, which says, there is an increase in the number of accidents on Tuesdays and Saturdays. The shift observed although, based on the current data, could not be explained. The requisite to do is to carry out a survey for the same and understand this. However, in the absence of this level of microdata, we were not very successful in addressing this hypothesis, although sincere efforts were made for the same.

iv. <u>Evidence:</u> Evidence that will help validate or falsify the aforementioned hypotheses includes the following spatial and aspatial data.

- Pre-existing Data:
  a) Shapefiles
  b) Accident Data
  c) Raster Data
  d) State and Country Data on Accidents
  e) Synthesized Data
  f) Survey Data
  g) Statistical Data

- Derived aspatial Evidence:
  a) Graphs & Trend Lines

- Derived Spatial Evidence:
  a) Maps
  b) Spatial Database
  c) GPS points

- Unavailable data:
  a) The aspatial microdata
  b) The spatial microdata
  c) In-depth data for city wise accident data based on gender, age, time, etc. for a more comprehensive analysis
  d) Target Location Data
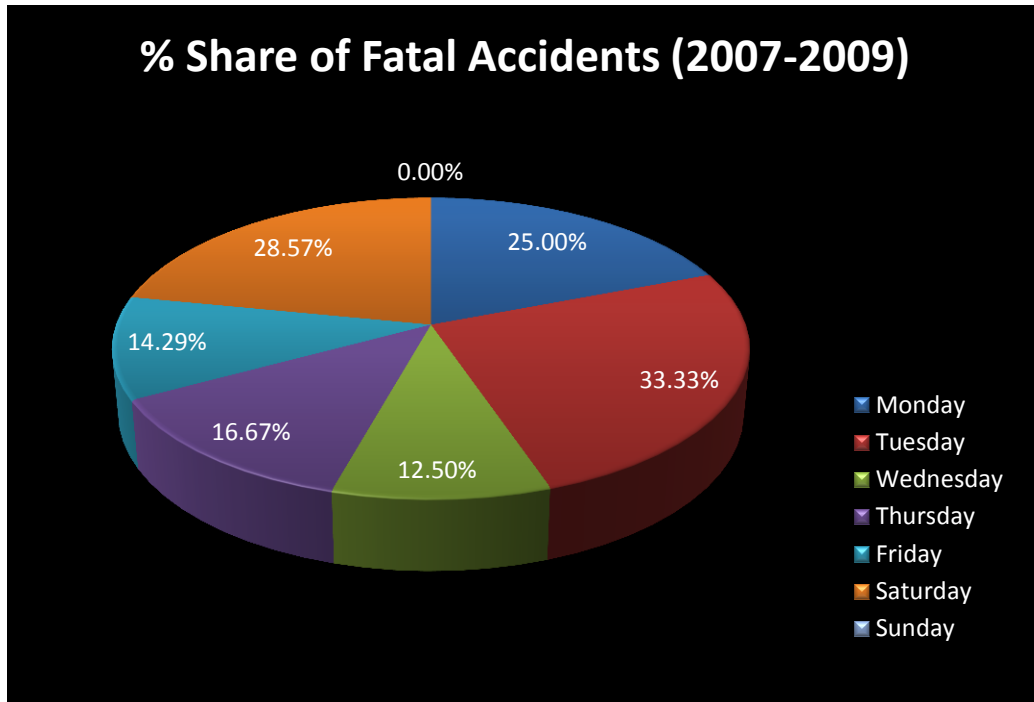
Figure 164 % share of fatal accidents for 2007-2011
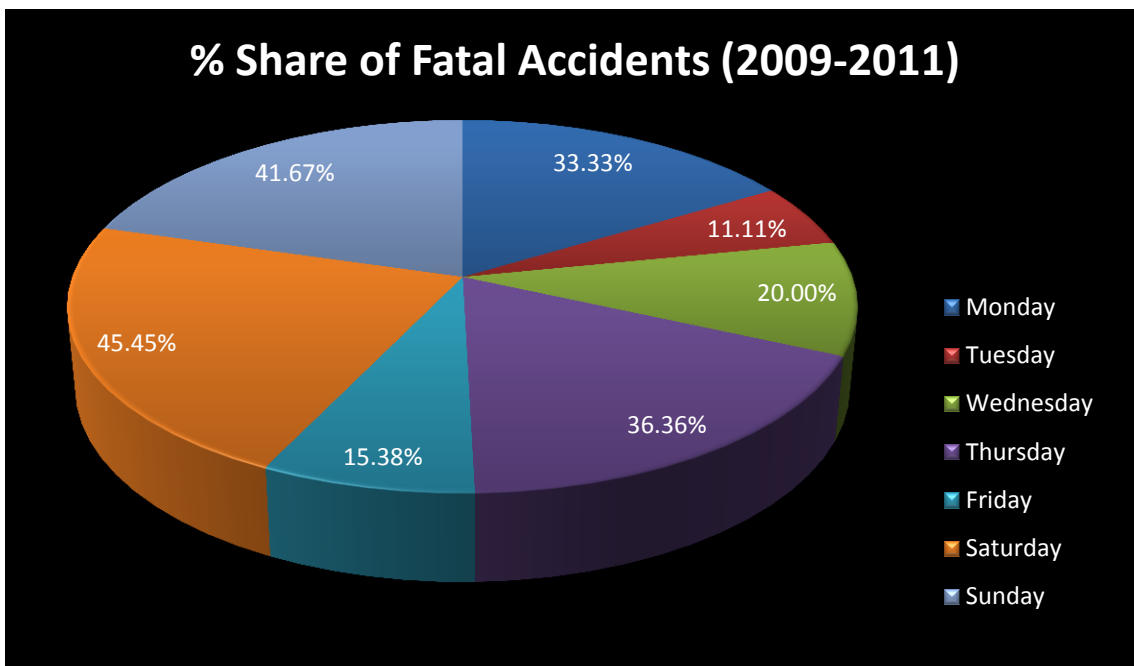
Figure 163 % share of fatal accidents for 2009-2011

➢ Results & Discussions

Accident Sites_2007-2009							
FID	Shape *	Id	Latitude	Longitude	Accident_C	Fleet_C	Accident_R
0	Point	0	19.281	72.894	26	130	0.2
1	Point	0	19.285	72.893	45	132	0.340909
2	Point	0	19.286	72.895	2	126	0.15873
3	Point	0	19.29	72.892	0	126	0
4	Point	0	19.281	72.908	33	136	0.242647
5	Point	0	19.277	72.921	25	126	0.198413
6	Point	0	19.273	72.922	23	136	0.169118
7	Point	0	19.273	72.923	0	136	0
8	Point	0	19.271	72.922	1	132	0.757576
9	Point	0	19.283	72.937	2	129	0.015504
10	Point	0	19.283	72.942	3	132	0.022727
11	Point	0	19.281	72.938	0	132	0
12	Point	0	19.279	72.937	1	130	0.769231
13	Point	0	19.281	72.947	2	136	0.147059
14	Point	0	19.279	72.947	5	132	0.378788
15	Point	0	19.267	72.968	0	131	0
16	Point	0	19.268	72.971	0	140	0
17	Point	0	19.241	72.976	6	132	0.454545
18	Point	0	19.232	72.98	1	128	0.78125
19	Point	0	19.224	72.983	4	132	0.30303
20	Point	0	19.22	72.984	0	132	0
21	Point	0	19.194	72.981	2	132	0.151515

Table 25 Attributes of the Dataset used for Accident Sites for 2007-09

OBJECTID *	Shape *	SOURCE_ID	GiZScore Fixed 23048	GiPValue Fixed 23048	Accident_R
1	Point	0	-0.460461	0.645185	0.2
2	Point	1	-0.460461	0.645185	0.340909
3	Point	2	-0.424177	0.671437	0.15873
4	Point	3	-0.460461	0.645185	0
5	Point	4	-0.384334	0.700731	0.242647
6	Point	5	0.644795	0.51906	0.198413
7	Point	6	0.874158	0.382032	0.169118
8	Point	7	0.874158	0.382032	0
9	Point	8	0.874158	0.382032	0.757576
10	Point	9	0.874158	0.382032	0.015504
11	Point	10	-0.809173	0.418416	0.022727
12	Point	11	0.546312	0.584851	0
13	Point	12	0.546312	0.584851	0.769231
14	Point	13	-1.201775	0.229451	0.147059
15	Point	14	-0.809173	0.418416	0.378788
16	Point	15	-0.809173	0.418416	0
17	Point	16	-0.909111	0.363292	0
18	Point	17	0.859116	0.390277	0.454545
19	Point	18	0.859116	0.390277	0.78125
20	Point	19	0.859116	0.390277	0.30303
21	Point	20	0.859116	0.390277	0
22	Point	21	1.368442	0.171174	0.151515

Table 26 Attributes of the Dataset from Hotspots Analyses for 2007-09

● ● ●

2009_11_ols

FID	Shape *	Estimated	Residual	StdResid	Id	Accident_R	Duration_o	Age	Tenure_for
0	Point	0.382437	-0.046437	-0.328750	1	0.336	3.45	35	15
1	Point	0.499646	0.000354	0.002504	2	0.5	4.3	40	11
2	Point	0.592792	0.187208	1.325330	3	0.78	5.45	55	22
3	Point	0.631795	0.038205	0.270470	4	0.67	6.45	30	5
4	Point	0.699579	0.113421	0.802960	5	0.813	7.45	32	7
5	Point	0.742418	0.077582	0.549237	6	0.82	8.45	31	10
6	Point	0.839121	-0.039121	-0.276957	7	0.8	9.45	39	13
7	Point	0.854001	-0.030001	-0.212388	8	0.824	10.45	38	20
8	Point	0.179935	0.220065	1.557940	9	0.4	1.05	42	26
9	Point	0.246894	0.168106	1.190100	10	0.415	0.35	45	12
10	Point	0.521391	0.058609	0.414921	11	0.58	5.3	20	0.75
11	Point	0.395854	-0.165854	-1.174160	12	0.23	2.5	60	25
12	Point	0.274734	0.005266	0.037279	13	0.28	2.35	31	16
13	Point	0.231962	-0.071962	-0.509452	14	0.16	1.15	25	5
14	Point	1.06989	-0.199895	-1.415150	15	0.87	13	42	18
15	Point	0.304754	-0.154754	-1.095570	16	0.15	1.45	36	7
16	Point	0.259313	0.138687	0.981826	17	0.398	1.5	36	14
17	Point	0.361491	-0.021491	-0.152148	18	0.34	2.35	49	19
18	Point	0.401995	-0.131995	-0.934452	19	0.27	3.2	26	2
19	Point	0.290816	0.007184	0.050860	20	0.298	1.05	36	5
20	Point	0.158147	-0.028147	-0.199268	21	0.13	0.15	29	9
21	Point	0.226915	-0.062915	-0.445406	22	0.164	1.35	30	12
22	Point	0.530034	0.018966	0.134268	23	0.549	4.45	48	15
23	Point	0.272773	-0.112773	-0.798372	24	0.16	1.45	40	15
24	Point	0.150998	-0.000998	-0.007067	25	0.15	0.35	50	30
25	Point	0.869538	0.109538	0.775468	26	0.76	10.2	33	11

OBJECTID *	Shape *	SOURCE_ID	GiZScore Fixed 23048	GiPValue Fixed 23048	Accident_R
1	Point	0	1.025817	0.304978	0.373418
2	Point	1	1.025817	0.304978	0.125786
3	Point	2	0.984316	0.32496	0.28
4	Point	3	2.101208	0.035623	0.02
5	Point	4	1.025817	0.304978	0.031447
6	Point	5	2.365909	0.017986	0.119497
7	Point	6	1.933883	0.053128	0.21875
8	Point	7	1.933883	0.053128	0.40625
9	Point	8	2.799687	0.005115	0.316456
10	Point	9	1.492301	0.13562	0
11	Point	10	1.492301	0.13562	0.346405
12	Point	11	0.948921	0.342661	0
13	Point	12	1.492301	0.13562	0.24
14	Point	13	1.492301	0.13562	0.308725
15	Point	14	1.492301	0.13562	0.013793
16	Point	15	1.492301	0.13562	0
17	Point	16	1.492301	0.13562	0
18	Point	17	1.492301	0.13562	0.40625
19	Point	18	-1.906874	0.056537	0.121212
20	Point	19	-1.906874	0.056537	0.061644
21	Point	20	-1.906874	0.056537	0
22	Point	21	-1.906874	0.056537	0.041096
23	Point	22	-1.906874	0.056537	0.081761
24	Point	23	-1.906874	0.056537	0.146667
25	Point	24	-1.906874	0.056537	0.013333
26	Point	25	-1.906874	0.056537	0
27	Point	26	-1.906874	0.056537	0
28	Point	27	-1.906874	0.056537	0.006289
29	Point	28	-2.149567	0.03159	0.028169
30	Point	29	-1.048898	0.294225	0
31	Point	30	0.183178	0.854658	0.144654
32	Point	31	0.183178	0.854658	0.134615
33	Point	32	0.183178	0.854658	0.089744
34	Point	33	-1.123183	0.26136	0.081761
35	Point	34	-1.399805	0.161572	0
36	Point	35	-1.726522	0.084254	0.00625
37	Point	36	-1.668389	0.095239	0
38	Point	37	-1.466591	0.142487	0

Figure 165 Attributes of the Dataset and from Hotspots Analyses for 2009-11
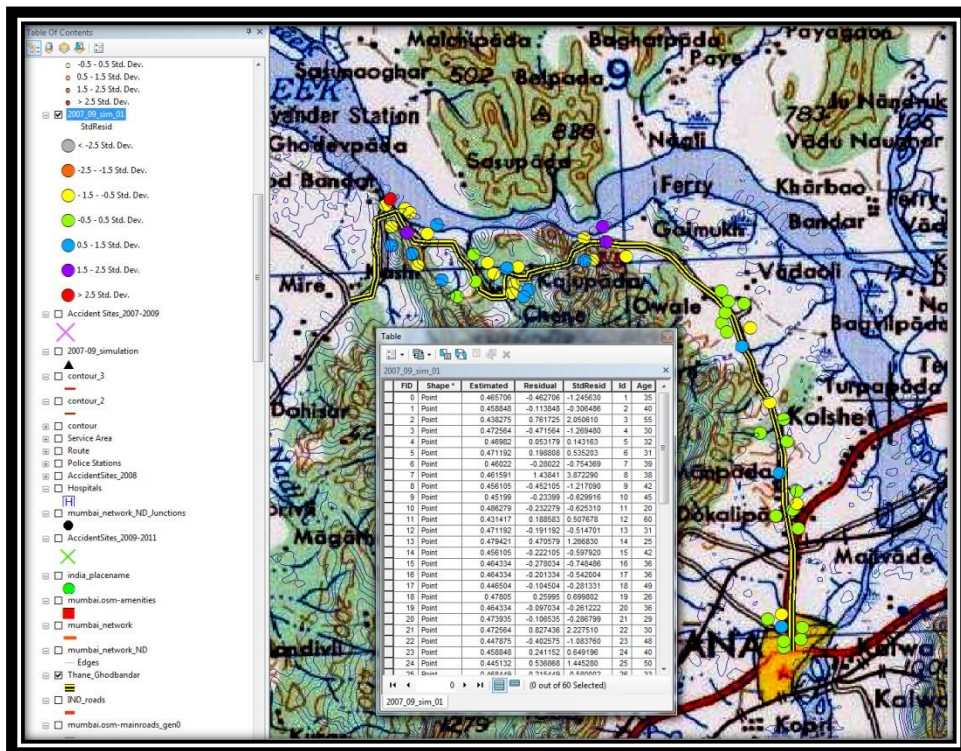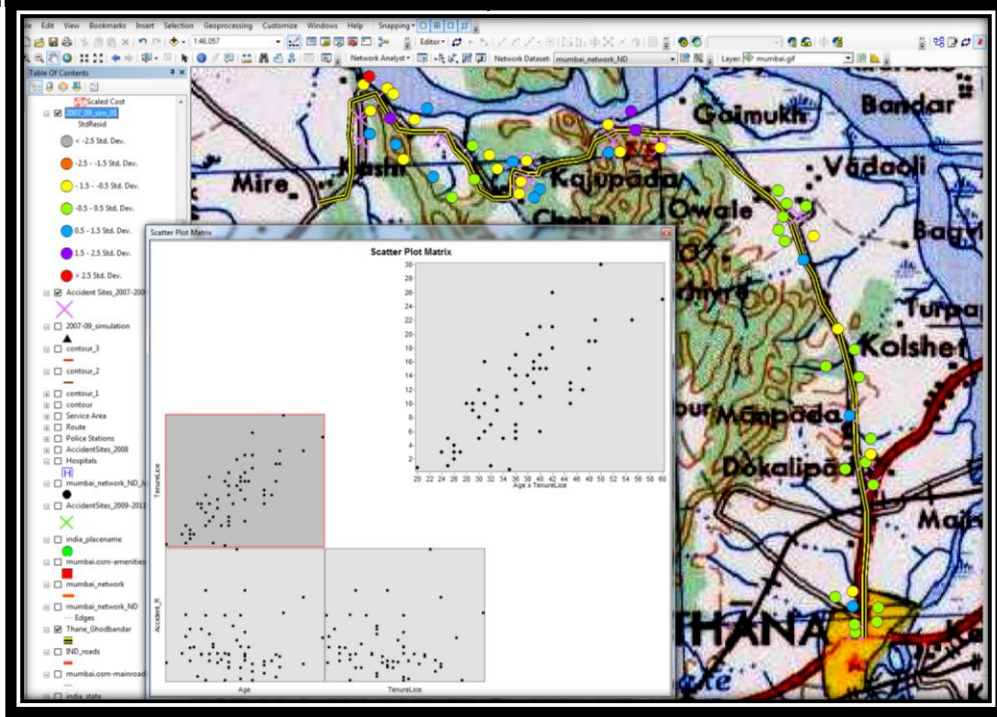
Figure 166 Attributes of the Dataset from OLS Analyses for 2007-09

2007_09_sim_05

Accident_R	DurationOf	Estimated	StdResid	Residual
0.3	2.25	0.413644	-0.422749	-0.113644
0.345	1.05	0.403484	-0.217556	-0.058484
0.25	3.25	0.42211	-0.640242	-0.17211
0.1	1.2	0.404754	-1.133670	-0.304754
0.523	2.5	0.41576	0.398926	0.10724
0.67	0.45	0.398404	1.010320	0.271596
0.18	1	0.40306	-0.829773	-0.22306
0.319	2.5	0.41576	-0.359944	-0.09676
0.4	3.2	0.421687	-0.080675	-0.021687
0.218	10.3	0.481801	-0.981327	-0.263801
0.254	13	0.504661	-0.932448	-0.250661
0.32	1	0.40306	-0.308980	-0.08306
0.88	1.1	0.403907	1.771040	0.476093
0.95	4.5	0.432694	1.924350	0.517306
0.234	1.3	0.4056	-0.638344	-0.1716
0.1863	3.3	0.422534	-0.878778	-0.236234
0.263	2.4	0.414914	-0.565111	-0.151914
0.342	1.5	0.407294	-0.242889	-0.065294
0.138	2.5	0.41576	-1.033250	-0.27776
0.3673	3.2	0.421687	-0.202317	-0.054387
0.3674	4.6	0.433541	-0.246040	-0.066141
0.989	12	0.496195	1.833210	0.492805
0.0453	2.1	0.412374	-1.365500	-0.367074
0.7	8.3	0.464868	0.874680	0.235132
0.482	2.5	0.41576	0.246408	0.06624
0.253	1	0.40306	-0.558217	-0.15006
0.738	2.6	0.416607	1.195570	0.321393
0.052	14.5	0.517362	-1.731120	-0.465362
0.938	12	0.496195	1.643490	0.441805
0.023	5.2	0.438621	-1.546090	-0.415621
0.98	4.3	0.431001	2.042250	0.548999
0.13	2.2	0.41322	-1.053570	-0.28322
0.1873	1.2	0.404754	-0.808916	-0.217454
0.283	3.4	0.423381	-0.522208	-0.140381
0.473	8.1	0.463174	0.036551	0.009826
0.28764	4	0.428461	-0.523845	-0.140821
0.4768	5.3	0.439467	0.138875	0.037333
0.6793	2	0.411527	0.996102	0.267773

Page 1 of 2

Accident_R	DurationOf	Estimated	StdResid	Residual
0.9383	3.2	0.421687	1.921770	0.516613
0.378	2.1	0.412374	-0.127868	-0.034374
0.237	3.1	0.42084	-0.683877	-0.18384
0.373	5.5	0.441161	-0.253554	-0.068161
0.2784	2.1	0.412374	-0.498375	-0.133974
0.3821	8.1	0.463174	-0.301592	-0.081074
0.3899	2.5	0.41576	-0.096199	-0.02586
0.937	5.2	0.438621	1.853950	0.498379
0.37992	11.3	0.490268	-0.410489	-0.110348
0.19383	4	0.428461	-0.872814	-0.234631
0.67	3.2	0.421687	0.923711	0.248313
0.393	5.3	0.439467	-0.172856	-0.046467
0.1892	4.5	0.432694	-0.905785	-0.243494
0.297	3.1	0.42084	-0.460680	-0.12384
0.6383	1.3	0.4056	0.865631	0.2327
0.792	7.6	0.458941	1.238960	0.333059
0.4838	0.5	0.398827	0.316096	0.084973
0.383	9.2	0.472488	-0.332890	-0.089488
0.493	0.8	0.401367	0.340870	0.091633
0.5839	4.1	0.429307	0.575077	0.154593
0.183	13	0.504661	-1.196560	-0.321661
0.992	7.4	0.457248	1.989250	0.534752

Table 31Attributes for simulation for accident sites 2007-2009 (i)

• • •

246

2007-09_simulation

FID	Id	Date	Day	Time	DurationOf	Direction	Age	Sex	TenureLice	AccNature	Accident_R
0	1	1/2/2007	Friday	0	2.25	To	35	M	15	Fatal	0.3
1	2	4/13/2007	Friday	0	1.05	From	40	M	11	Non-Fatal	0.345
2	3	7/5/2007	Monday	0	3.25	From	55	M	22	Non-Fatal	0.25
3	4	5/21/2007	Monday	0	1.2	From	30	F	5	Non-Fatal	0.1
4	5	5/22/2007	Tuesday	0	2.5	From	32	M	7	Fatal	0.523
5	6	5/6/2007	Tuesday	0	0.45	From	31	M	10	Non-Fatal	0.67
6	7	5/6/2007	Tuesday	0	1	To	39	M	13	Non-Fatal	0.18
7	8	8/6/2007	Friday	0	2.5	From	38	M	20	Non-Fatal	0.319
8	9	6/13/2007	Wednesday	0	3.2	To	42	M	26	Fatal	0.4
9	10	6/15/2007	Friday	0	10.3	To	45	M	12	Non-Fatal	0.218
10	11	7/13/2007	Friday	0	13	To	20	M	0.75	Non-Fatal	0.254
11	12	8/20/2007	Monday	0	1	From	60	M	25	Non-Fatal	0.32
12	13	9/17/2007	Monday	0	1.1	To	31	M	16	Fatal	0.88
13	14	9/21/2007	Friday	0	4.5	To	25	M	5	Fatal	0.95
14	15	9/28/2007	Friday	0	1.3	From	42	M	18	Non-Fatal	0.234
15	16	4/10/2007	Thursday	0	3.3	From	36	M	7	Non-Fatal	0.1863
16	17	10/24/2007	Wednesday	0	2.4	From	36	M	14	Non-Fatal	0.263
17	18	3/11/2007	Saturday	0	1.5	From	49	M	19	Fatal	0.342
18	19	8/11/2007	Thursday	0	2.5	From	26	F	2	Non-Fatal	0.138
19	20	12/14/2007	Friday	0	3.2	From	36	M	5	Non-Fatal	0.3673
20	21	2/1/2008	Wednesday	0	4.6	From	29	M	9	Non-Fatal	0.3674
21	22	4/2/2008	Monday	0	12	From	30	M	12	Fatal	0.989
22	23	4/17/2008	Thursday	0	2.1	From	48	M	15	Non-Fatal	0.0453
23	24	12/5/2008	Monday	0	8.3	To	40	M	15	Non-Fatal	0.7
24	25	5/31/2008	Saturday	0	2.5	To	50	M	30	Non-Fatal	0.482
25	26	2/6/2008	Monday	0	1	To	33	M	11	Non-Fatal	0.253
26	27	7/7/2008	Monday	0	2.6	From	29	M	10	Non-Fatal	0.738
27	28	7/26/2008	Saturday	0	14.5	From	28	M	10	Non-Fatal	0.052

Page 1 of 3

FID	Id	Date	Day	Time	DurationOf	Direction	Age	Sex	TenureLice	AccNature	Accident_R
28	29	7/28/2008	Monday	0	12	From	30	F	8	Non-Fatal	0.938
29	30	8/21/2008	Thursday	0	5.2	From	39	M	16	Non-Fatal	0.023
30	31	1/9/2008	Monday	0	4.3	From	45	F	10	Fatal	0.98
31	32	2/9/2008	Tuesday	0	2.2	From	42	M	21	Non-Fatal	0.13
32	33	9/13/2008	Saturday	0	1.2	From	49	M	22	Non-Fatal	0.1873
33	34	9/22/2008	Monday	0	3.4	From	47	M	12	Non-Fatal	0.283
34	35	9/29/2008	Monday	0	8.1	From	26	M	4	Non-Fatal	0.473
35	36	3/10/2008	Friday	0	4	From	40	M	21	Non-Fatal	0.28764
36	37	10/14/2008	Tuesday	0	5.3	To	32	M	1	Non-Fatal	0.4768
37	38	10/27/2008	Friday	0	2	From	41	M	15	Non-Fatal	0.6793
38	39	12/11/2008	Wednesday	0	3.2	To	39	M	15	Non-Fatal	0.9383
39	40	11/17/2008	Monday	0	2.1	From	36	M	17	Non-Fatal	0.378
40	41	12/12/2008	Friday	0	3.1	From	55	M	22	Non-Fatal	0.237
41	42	12/29/2008	Monday	0	5.5	To	48	M	19	Fatal	0.373
42	43	4/1/2009	Sunday	0	2.1	To	26	M	3	Non-Fatal	0.2784
43	44	6/2/2009	Friday	0	8.1	To	31	M	6	Non-Fatal	0.3821
44	45	12/4/2009	Sunday	0	2.5	From	24	F	3	Non-Fatal	0.3899
45	46	4/13/2009	Monday	0	5.2	From	35	F	0.5	Non-Fatal	0.937
46	47	2/5/2009	Saturday	0	11.3	To	40	M	17	Non-Fatal	0.37992
47	48	5/17/2009	Sunday	0	4	To	36	M	12	Non-Fatal	0.19383
48	49	5/23/2009	Saturday	0	3.2	From	45	M	13	Fatal	0.67
49	50	1/7/2009	Wednesday	0	5.3	To	35	M	15	Non-Fatal	0.393
50	51	7/23/2009	Thursday	0	4.5	From	33	M	9	Non-Fatal	0.1892
51	52	8/17/2009	Monday	0	3.1	To	34	M	5	Non-Fatal	0.297
52	53	8/18/2009	Tuesday	0	1.3	To	25	F	1	Non-Fatal	0.6383
53	54	8/26/2009	Wednesday	0	7.6	From	37	M	10	Non-Fatal	0.792
54	55	9/23/2009	Wednesday	0	0.5	From	27	M	3	Non-Fatal	0.4838
55	56	10/30/2009	Friday	0	9.2	From	29	M	9	Non-Fatal	0.383
56	57	11/26/2009	Thursday	0	0.8	From	39	M	15	Fatal	0.493
57	58	11/12/2009	Friday	0	4.1	From	38	M	14	Non-Fatal	0.5839
58	59	12/26/2009	Saturday	0	13	From	36	M	6	Non-Fatal	0.183

Page 2 of 3

Table 32 Attributes for simulation for accident sites 2009-2011 (i)

2009_11_ols

FID	Estimated	Residual	StdResid	Id	Accident_R	Duration_o	Age	Tenure_for
0	0.382437	-0.046437	-0.328750	1	0.336	3.45	35	15
1	0.499646	0.000354	0.002504	2	0.5	4.3	40	11
2	0.592792	0.187208	1.325330	3	0.78	5.45	55	22
3	0.631795	0.038205	0.270470	4	0.67	6.45	30	5
4	0.699579	0.113421	0.802960	5	0.813	7.45	32	7
5	0.742418	0.077582	0.549237	6	0.82	8.45	31	10
6	0.839121	-0.039121	-0.276957	7	0.8	9.45	39	13
7	0.854001	-0.030001	-0.212388	8	0.824	10.45	38	20
8	0.179935	0.220065	1.557940	9	0.4	1.05	42	26
9	0.246894	0.168106	1.190100	10	0.415	0.35	45	12
10	0.521391	0.058609	0.414921	11	0.58	5.3	20	0.75
11	0.395854	-0.165854	-1.174160	12	0.23	2.5	60	25
12	0.274734	0.005266	0.037279	13	0.28	2.35	31	16
13	0.231962	-0.071962	-0.509452	14	0.16	1.15	25	5
14	1.06989	-0.199895	-1.415150	15	0.87	13	42	18
15	0.304754	-0.154754	-1.095570	16	0.15	1.45	36	7
16	0.259313	0.138687	0.981826	17	0.398	1.5	36	14
17	0.361491	-0.021491	-0.152148	18	0.34	2.35	49	19
18	0.401995	-0.131995	-0.934452	19	0.27	3.2	26	2
19	0.290816	0.007184	0.050860	20	0.298	1.05	36	5
20	0.158147	-0.028147	-0.199268	21	0.13	0.15	29	9
21	0.226915	-0.062915	-0.445406	22	0.164	1.35	30	12
22	0.530034	0.018966	0.134268	23	0.549	4.45	48	15
23	0.272773	-0.112773	-0.798372	24	0.16	1.45	40	15
24	0.150998	-0.000998	-0.007067	25	0.15	0.35	50	30
25	0.869538	-0.109538	-0.775468	26	0.76	10.2	33	11
26	0.315174	-0.085174	-0.602983	27	0.23	2.5	29	10
27	0.162621	0.067379	0.477006	28	0.23	0.4	28	10
28	0.474727	0.065273	0.462100	29	0.54	4.5	30	8
29	0.873986	0.002013	0.014254	30	0.876	10.25	39	16
30	0.54703	0.01297	0.091824	31	0.56	4.45	45	10
31	0.242803	-0.092803	-0.656994	32	0.15	1.45	42	21
32	0.392867	-0.232867	-1.648570	33	0.16	3.1	49	22
33	0.339127	0.010873	0.076978	34	0.35	1.5	47	12
34	0.175143	-0.125143	-0.885943	35	0.05	0.15	26	4
35	0.202916	-0.052915	-0.374613	36	0.15	1.05	40	21
36	0.570523	-0.010523	-0.074499	37	0.56	5	32	1
37	0.878987	0.010013	0.070884	38	0.889	10.05	41	15

Page 1 of 2

FID	Estimated	Residual	StdResid	Id	Accident_R	Duration_o	Age	Tenure_for
38	0.172566	-0.072566	-0.513729	39	0.1	0.1	39	15
39	0.203446	0.096554	0.683547	40	0.3	1	36	17
40	0.306636	-0.076636	-0.542544	41	0.23	1.35	55	22
41	0.271754	-0.101754	-0.720361	42	0.17	1.15	48	19
42	0.33568	-0.06868	-0.486216	43	0.267	2.35	26	3
43	0.403959	0.158041	1.118840	44	0.562	3.2	31	6
44	0.191101	-0.151101	-1.069720	45	0.04	0.45	24	3
45	0.243002	-0.073002	-0.516816	46	0.17	0	35	0.5
46	0.241345	0.401655	2.843500	47	0.643	1.2	40	17
47	0.175582	0.359418	2.544490	48	0.535	0.1	36	12
48	0.222455	-0.132455	-0.937711	49	0.09	0.1	45	13
49	0.298684	-0.008684	-0.061479	50	0.29	2.25	35	15
50	0.272819	-0.182819	-1.294260	51	0.09	1.45	33	9
51	0.369579	-0.099579	-0.704962	52	0.27	2.35	34	5
52	0.458835	0.103165	0.730351	53	0.562	4	25	1
53	0.27232	-0.09232	-0.653575	54	0.18	1.2	37	10
54	0.319331	-0.059331	-0.420028	55	0.26	2.03	27	3
55	0.220962	-0.040962	-0.289989	56	0.18	1.05	29	9
56	0.190015	-0.170015	-1.203610	57	0.02	0.35	39	15
57	0.526032	0.256968	1.819200	58	0.783	5.15	38	14
58	0.629307	0.263693	1.866810	59	0.893	6	36	6
59	0.241461	0.321539	2.276320	60	0.563	0.1	40	6
60	0.413998	-0.057998	-0.410593	61	0.356	2.45	35	0.5

Table 32 Attributes for simulation for accident sites 2009-2011 (i)

2009_11_gwr

Observed	Cond	LocalR2	Predicted	Intercept	StdErrC1_I	StdErrC2_A	StdErrC3_I	StdResid
0.336	16.526299	0.721267	0.382407	0.036909	0.005781	0.003372	0.004006	-0.334144
0.5	16.526982	0.721246	0.499668	0.036938	0.005781	0.003372	0.004006	0.002381
0.78	16.526673	0.721256	0.592817	0.036928	0.005781	0.003372	0.004006	1.401543
0.67	16.527024	0.721245	0.631797	0.036944	0.005781	0.003372	0.004006	0.277673
0.813	16.526698	0.721255	0.699573	0.036932	0.005781	0.003372	0.004006	0.826479
0.82	16.526488	0.721262	0.742374	0.036922	0.005781	0.003372	0.004006	0.570985
0.8	16.526184	0.721272	0.839096	0.036913	0.005781	0.003372	0.004006	-0.289199
0.824	16.526129	0.721273	0.853897	0.036908	0.005781	0.003372	0.004006	-0.22889
0.4	16.526174	0.721272	0.179874	0.036907	0.005781	0.003372	0.004006	1.671095
0.415	16.526287	0.721268	0.246978	0.036911	0.005781	0.003372	0.004006	1.229169
0.58	16.527089	0.721243	0.521379	0.036948	0.005781	0.003372	0.004006	0.435658
0.23	16.526914	0.721249	0.395887	0.036943	0.005781	0.003372	0.004006	-1.271935
0.28	16.527256	0.721238	0.274691	0.036958	0.005781	0.003372	0.004006	0.039152
0.16	16.527457	0.721232	0.231961	0.036964	0.005781	0.003372	0.004006	-0.524824
0.87	16.527381	0.721234	1.069856	0.036961	0.005781	0.003372	0.004006	-1.57077
0.15	16.527668	0.721225	0.304777	0.036973	0.005781	0.003372	0.004006	-1.11518
0.398	16.52781	0.72122	0.259305	0.036977	0.005781	0.003372	0.004006	0.995996
0.34	16.527818	0.721221	0.361492	0.036982	0.005781	0.003372	0.004006	-0.156349
0.27	16.527532	0.72123	0.402009	0.036971	0.005781	0.003372	0.004006	-0.9593
0.298	16.527645	0.721226	0.29085	0.036974	0.005781	0.003372	0.004006	0.05202
0.13	16.527618	0.721227	0.158142	0.036973	0.005781	0.003372	0.004006	-0.204564
0.164	16.527818	0.721221	0.226897	0.036982	0.005781	0.003372	0.004006	-0.457257
0.549	16.527667	0.721225	0.53005	0.036977	0.005781	0.003372	0.004006	0.138484
0.16	16.527726	0.721223	0.27277	0.036977	0.005781	0.003372	0.004006	-0.808577
0.15	16.527915	0.721217	0.150959	0.036985	0.005781	0.003372	0.004006	-0.007373
0.76	16.528003	0.721215	0.869528	0.036993	0.005781	0.003372	0.004006	-0.819222
0.23	16.528084	0.721212	0.315161	0.036995	0.005781	0.003372	0.004006	-0.615649
0.23	16.528073	0.721213	0.162607	0.036996	0.005781	0.003372	0.004006	0.492586
0.54	16.528156	0.72121	0.474724	0.037	0.005781	0.003372	0.004006	0.469439
0.876	16.528294	0.721206	0.873978	0.037006	0.005781	0.003372	0.004006	0.015141
0.56	16.528169	0.72121	0.547045	0.037003	0.005781	0.003372	0.004006	0.095361
0.15	16.528216	0.721209	0.242772	0.037005	0.005781	0.003372	0.004006	-0.676562
0.16	16.528413	0.721203	0.392842	0.037015	0.005781	0.003372	0.004006	-1.697241
0.35	16.528658	0.721195	0.339115	0.037026	0.005781	0.003372	0.004006	0.080143
0.05	16.528795	0.721191	0.175138	0.037034	0.005781	0.003372	0.004006	-0.913756
0.15	16.529127	0.721181	0.202881	0.037048	0.005781	0.003372	0.004006	-0.388321
0.56	16.529431	0.721171	0.570527	0.037061	0.005781	0.003372	0.004006	-0.077545
0.889	16.531556	0.721102	0.879023	0.037134	0.005781	0.003372	0.004006	0.074477

Page 1 of 2

Observed	Cond	LocalR2	Predicted	Intercept	StdErrC1_I	StdErrC2_A	StdErrC3_I	StdResid
0.1	16.531556	0.721102	0.172516	0.037134	0.005781	0.003372	0.004006	-0.523112
0.3	16.53129	0.721111	0.203436	0.037124	0.005781	0.003372	0.004006	0.702643
0.23	16.531221	0.721113	0.306546	0.037122	0.005781	0.003372	0.004006	-0.571962
0.17	16.531374	0.721108	0.271684	0.037127	0.005781	0.003372	0.004006	-0.739788
0.267	16.53123	0.721113	0.335663	0.037125	0.005781	0.003372	0.004006	-0.498276
0.562	16.530879	0.721125	0.403942	0.037115	0.005781	0.003372	0.004006	1.13487
0.04	16.530827	0.721126	0.191082	0.037112	0.005781	0.003372	0.004006	-1.108209
0.17	16.530632	0.721132	0.242923	0.037104	0.005781	0.003372	0.004006	-0.547631
0.643	16.530355	0.721141	0.241312	0.037094	0.005781	0.003372	0.004006	2.89413
0.535	16.530621	0.721133	0.175537	0.037107	0.005781	0.003372	0.004006	2.587965
0.09	16.530163	0.721148	0.222385	0.037091	0.005781	0.003372	0.004006	-0.966153
0.29	16.529757	0.721161	0.298669	0.037073	0.005781	0.003372	0.004006	-0.062481
0.09	16.527377	0.721234	0.272827	0.036958	0.005781	0.003372	0.004006	-1.310688
0.27	16.528714	0.721193	0.36958	0.037026	0.005781	0.003372	0.004006	-0.718844
0.562	16.528972	0.721185	0.458848	0.037041	0.005781	0.003372	0.004006	0.753055
0.18	16.529357	0.721173	0.272297	0.037055	0.005781	0.003372	0.004006	-0.661726
0.26	16.531889	0.721091	0.319296	0.037143	0.005781	0.003372	0.004006	-0.429887
0.18	16.530228	0.721146	0.220948	0.037093	0.005781	0.003372	0.004006	-0.29642
0.02	16.529995	0.721153	0.189975	0.037083	0.005781	0.003372	0.004006	-1.224477
0.783	16.528382	0.721204	0.526023	0.037014	0.005781	0.003372	0.004006	1.842298
0.893	16.528294	0.721206	0.629321	0.037006	0.005781	0.003372	0.004006	1.918779
0.563	16.52659	0.721258	0.241553	0.03692	0.005781	0.003372	0.004006	2.372237
0.356	16.526638	0.721256	0.414083	0.03692	0.005781	0.003372	0.004006	-0.433057

Table 33 Attributes for simulation for accident sites 2009-2011 (ii)
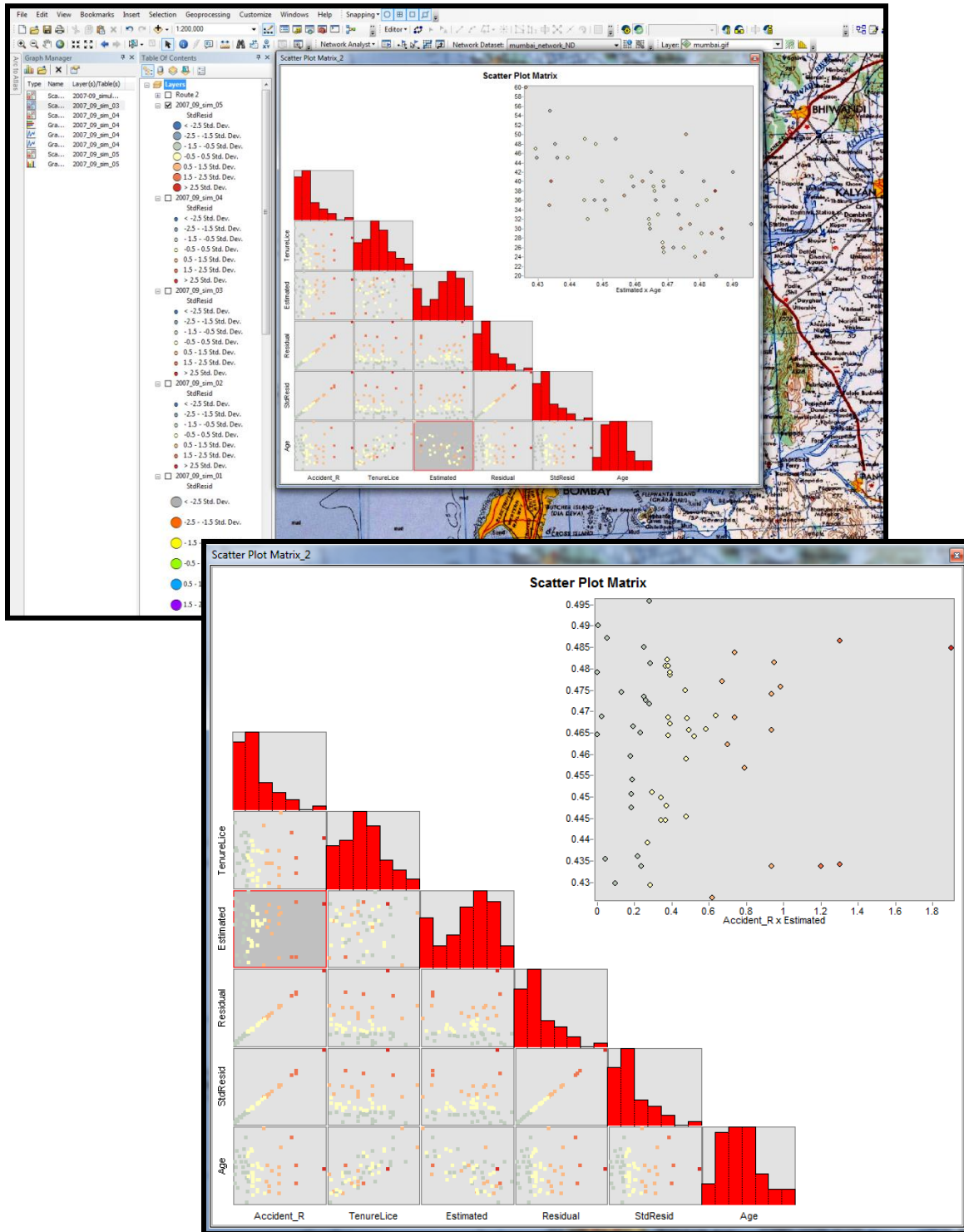
Figure 167 The scatterplot for the different Regression Analysis carried out on the OLS and GWR Database for 2007-2009
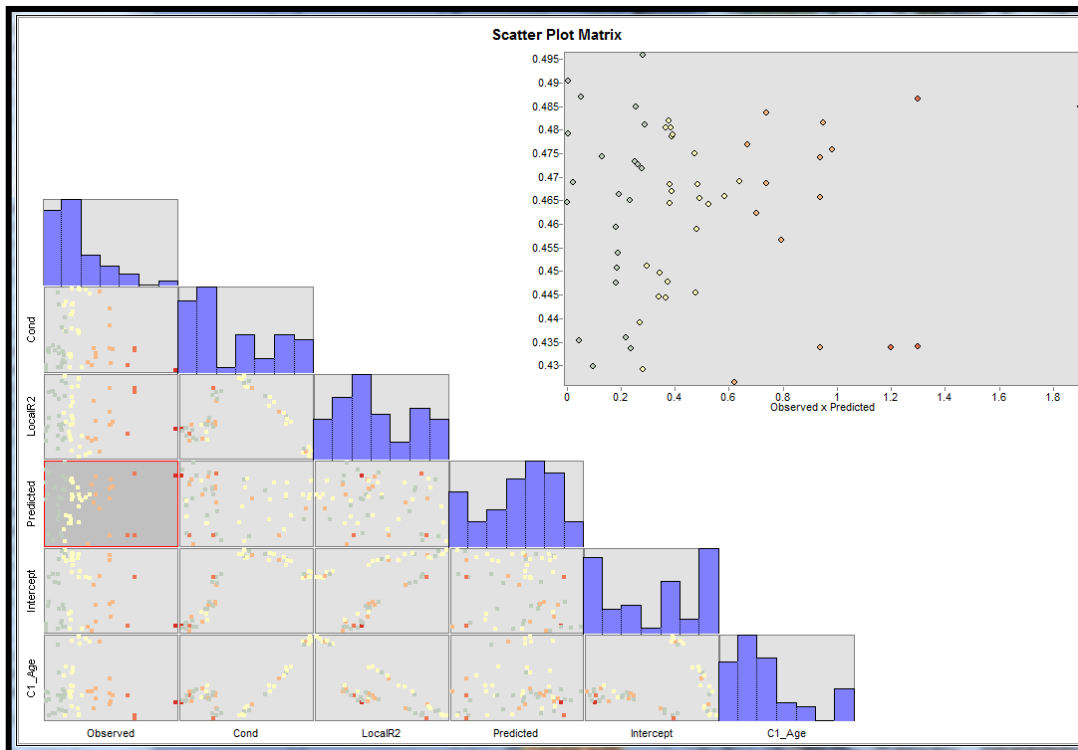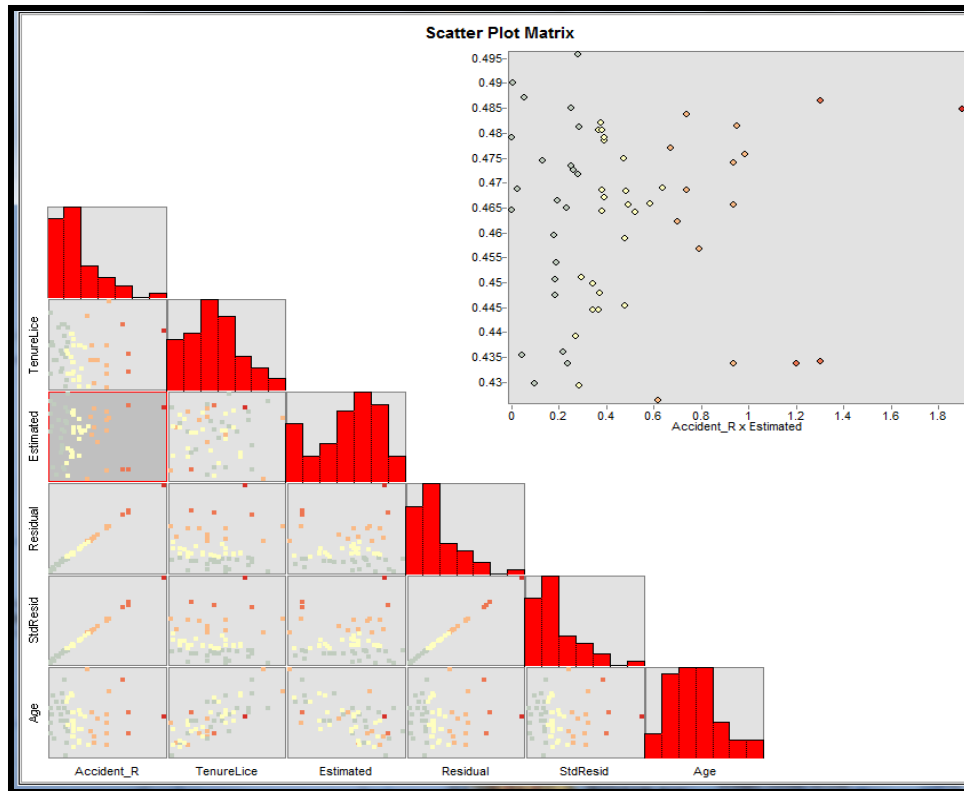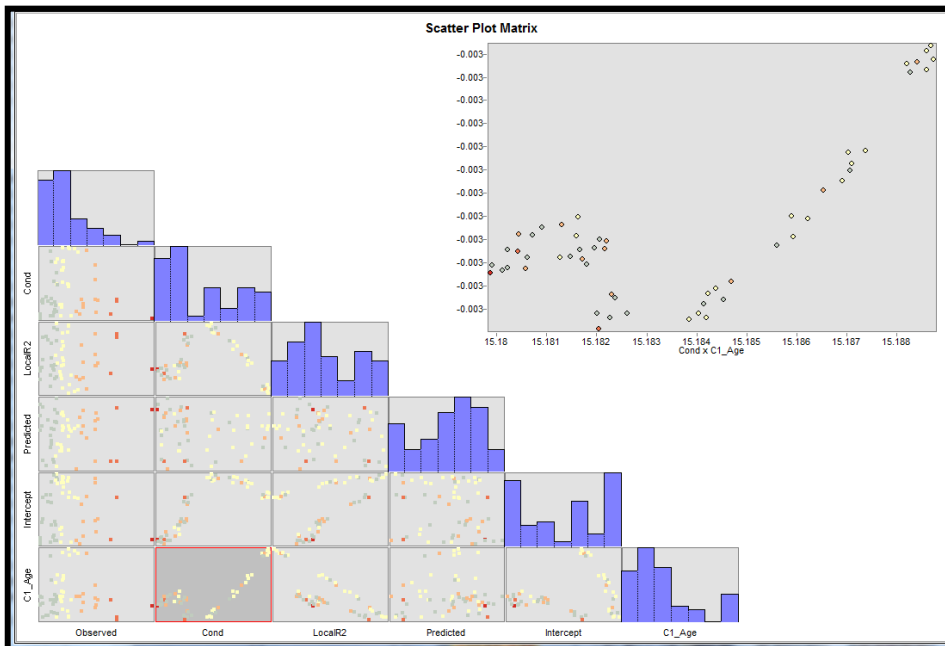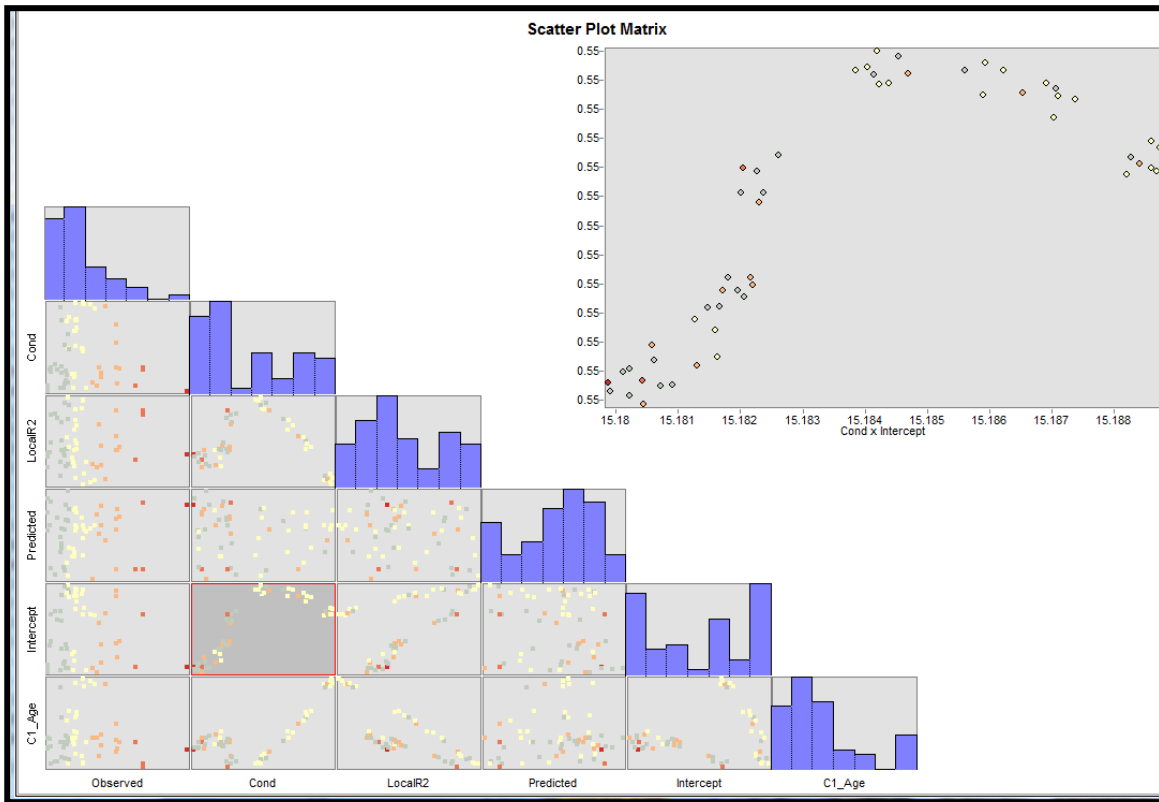
Figure 168 Fig: The scatterplot for the different Regression Analysis

169 Fig: The scatterplot for the different Regression Analysis carried out on the GWR Database
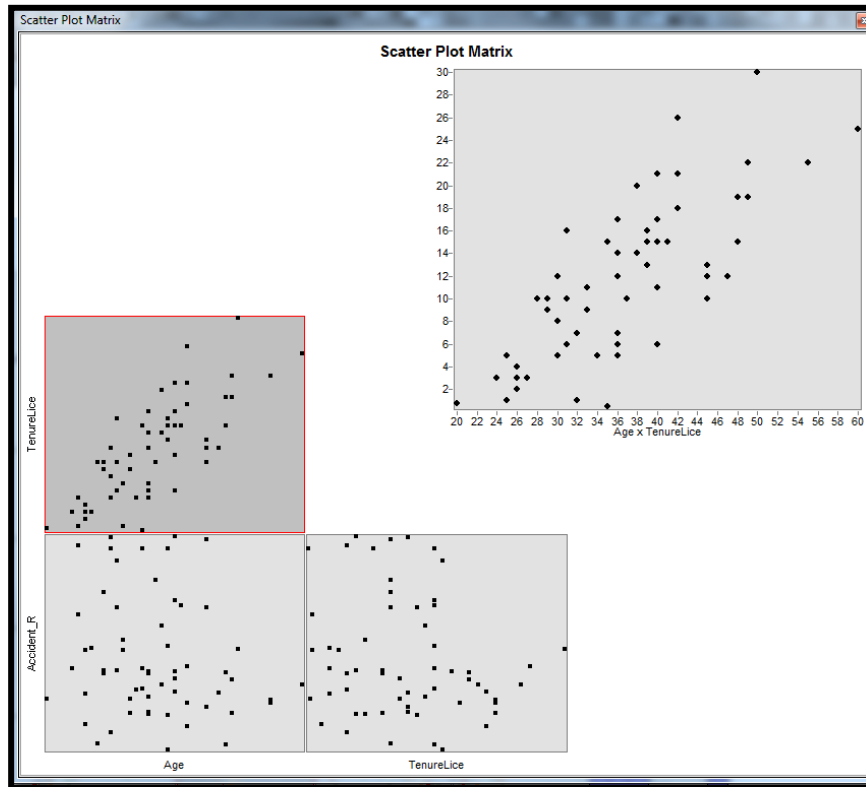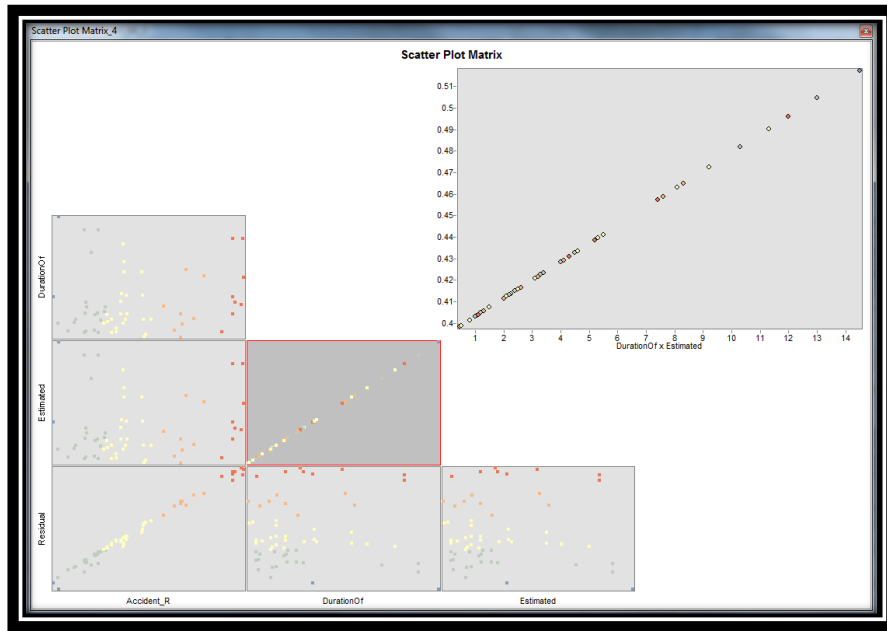
Figure 170 The scatter plot for the parameters taken for RAYSIM (2007-2009)
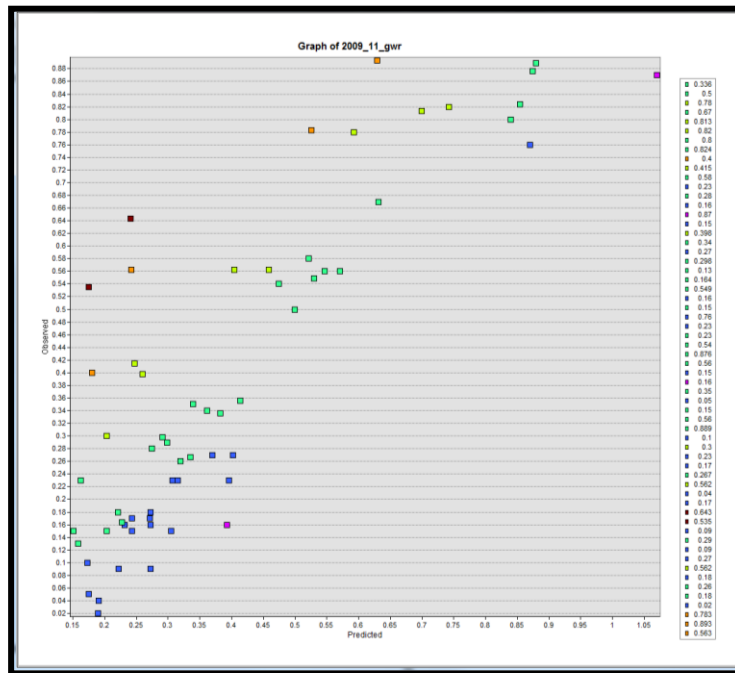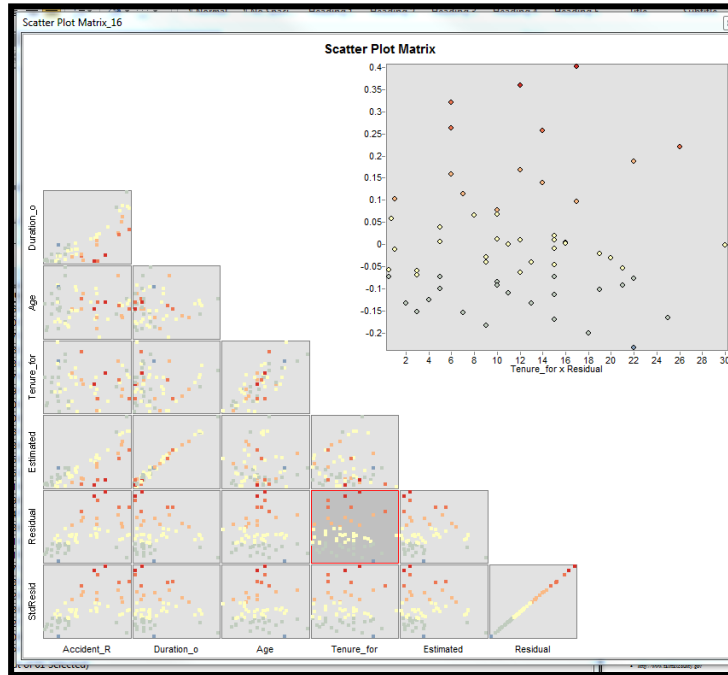
Figure 171 The scatter plot for the parameters taken for RAYSIM (2009-2011)
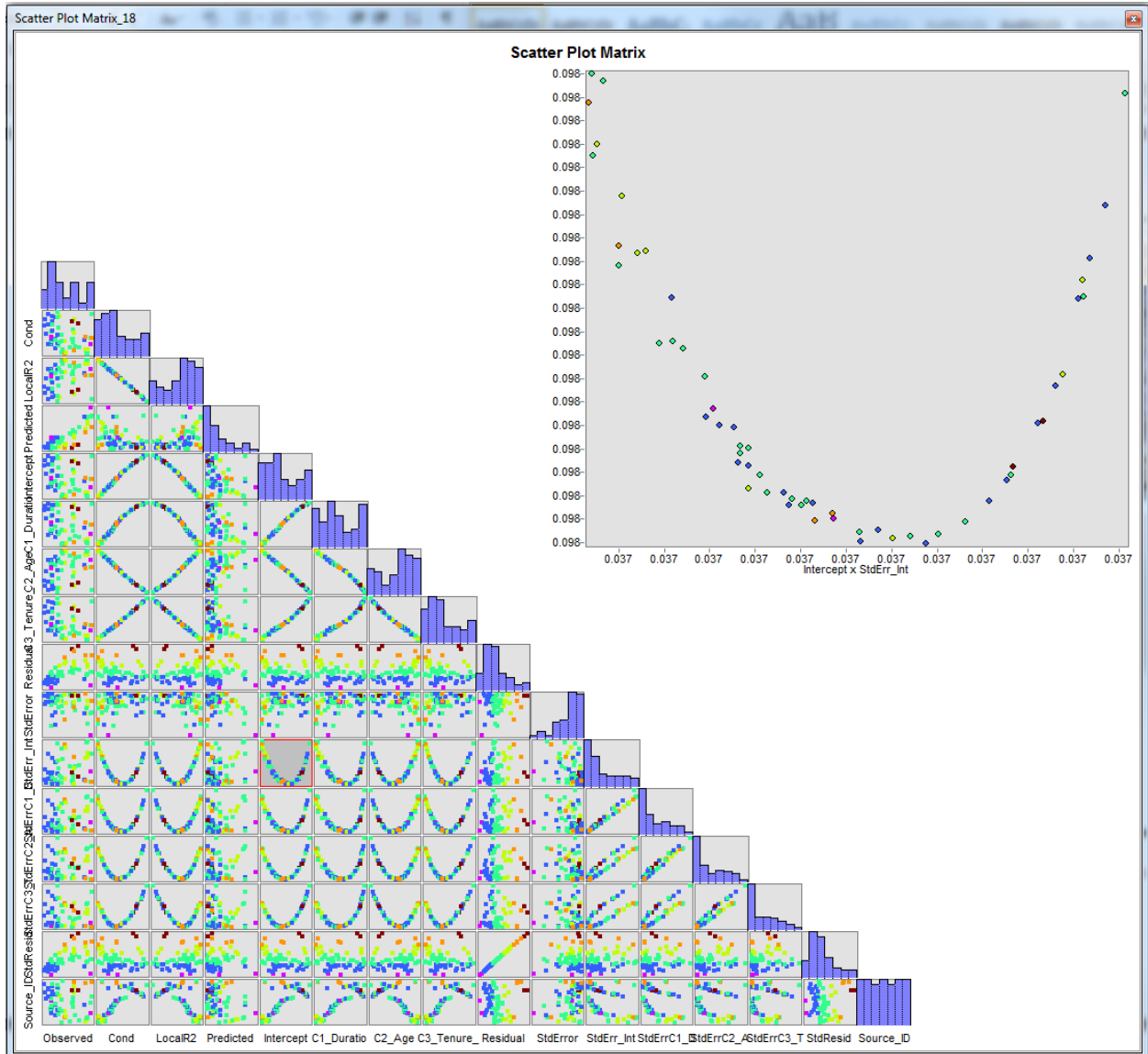
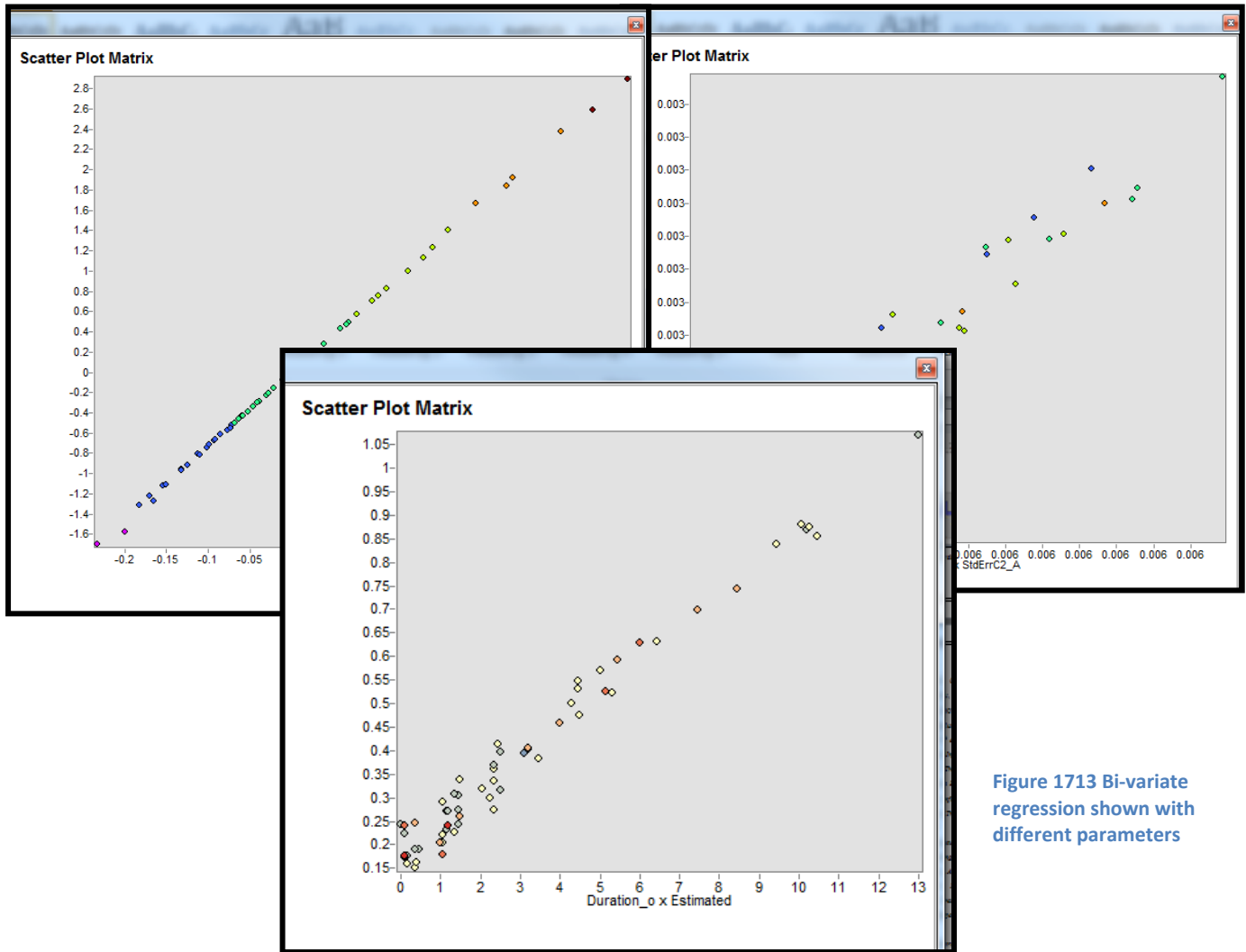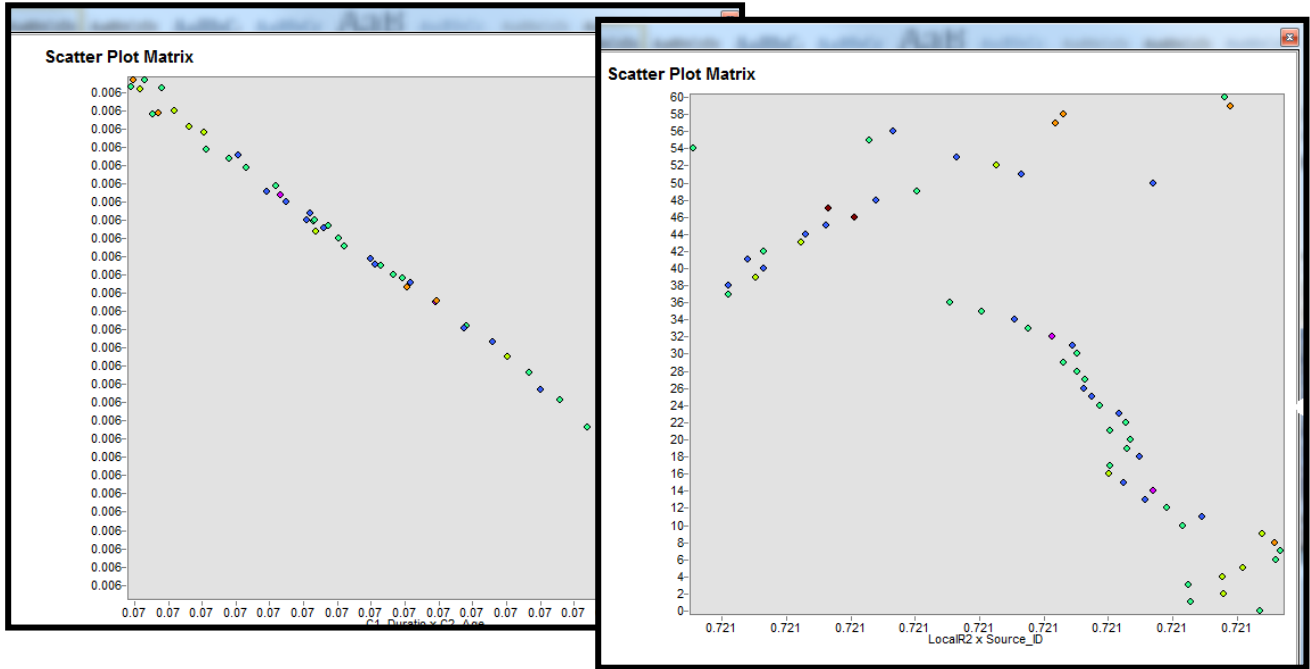**Figure 1702 The scatter plot for the parameters considered in RAYSIM for 2009-11**

Figure 1713 Bi-variate regression shown with different parameters

15.13 Conclusion

In the very beginning of this project, the impact of accidents was discussed and later they were correlated to the temporal, spatial and physicopsycological data and their impact on accidents. We further postulated the various ways which could determine or reasons that could alter the accident rate. This research has in essence examined a series of snapshots as an output to the microsimulation model, RAYSIM. It also has shown the potential for microsimulation modeling to be applied for the road accident analysis.

This approach has also given a more detailed picture based on the implications of the topography, the decision making power of the driver as well as the other temporal data on the accident rate.

The microsimulation model RAYSIM has also used relevant methods, or tools which can be applied to analyze and uncover the road accident reasons, ranging from the location to the time and day of accidents.

CHAPTER 16 – INFERENCE

The objective of this project was to forecast accident clusters based on the road geometry, gradient, width and curvature. Prediction of the blind-turns ahead was also one such aim of this project. For that very purpose, a decision support system for driver assistance aided to a navigator can be helpful in that case. **ADAS** (Advanced Driver Assistance System) are systems that help the driver in the driving process. It is designed with a safe Human-Machine Interface that increases car safety and delivers a more general road safety. It is rich, robust digital data which includes hundreds of road attributes.

ADAS Usage
The map attributes for ADAS has the following benefits and usage:

- Active Safety: Adaptive front lights to avoid collision
- Driver Assistance: Speed warning signs do pop up
- Energy Efficiency: Hybrid Power-train control and predictive cruise control
- Use-based Charging: Road User Charging and pay as you drive

ADAS Benefits
- Provide warnings on route
- Improves route calculation
- Route planning

NAVTEQ Maps include data which can specifically support ADAS, such as road geometry, slope and height. Relying on these ADAS capabilities; key customers of NAVTEQ Maps helps drivers in better detection and avoiding dangerous situations and which in turn will prevent accidents. This technology has greatly been put to use abroad but that isn't the case in India. If this ADAS technology is put to use the accident rates can be reduced to a great extent.

CHAPTER 17 – CONCLUSION

The project work started with the following points in mind:

- A Navigation Related Project
- Probable topics submitted
- Microsimulative approach selected, out of the given topics to NAVTEQ INDIA PVT LTD.

Once the selection was done for the topic, we needed data to carry out the work. This was followed by data gathering to carry out a GIS project. This data had various facets which initially started with Survey. The survey was conducted, firstly to evidence the reasons available for the selection of Thane Ghodbander Road as our area of interest, secondly to collect the sample points, using GPS Navigator NAVON N85. Thirdly, the Survey was conducted to validate the results obtained from the analysis of the road, spatially and aspatially.

This work led us to the following insights for the causes of the accidents on Thane Ghodbander Road

- The blind curves are the major reasons for accidents accounting to 56%
- The date time modifiers account for 23% of the total accidents
- The physicospychological factors pertaining to drivers account to 11% of total accidents
- The other causes like, drunken driving, ongoing construction projects, slipping, etc fall under accidental causes for accidents which account for 10% of the total causes.
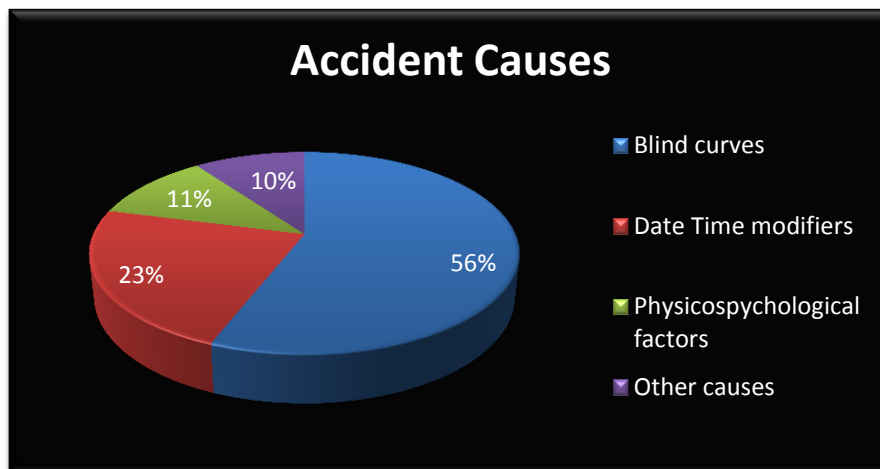

Figure 172 The Accident Reasons Percentage

Solutions were discussed based on the primary and secondary reasons. The spatial and aspatial analysis of the road gave us an insight as to why, was the accident rate alarming .i.e., 16% of the total road accidents in Maharashtra.

The primary solution discussed was:

- ADAS (Advance Driver Assisted System) assisted driving: This system has the potential to solve the problems leading to crash on the road. The reasons were well acknowledged by the experts, where also the road geometry was taken into account. This further exemplified the usage of ADAS, as to alert the drivers regarding the blind turns, sharp curvature, changing gradient, the traffic flow, etc. In high volume of traffic it helps maintain a safe distance from the potentially crashing vehicles.
- Placing of concave mirrors at appropriate locations, eg, Blind Turns.
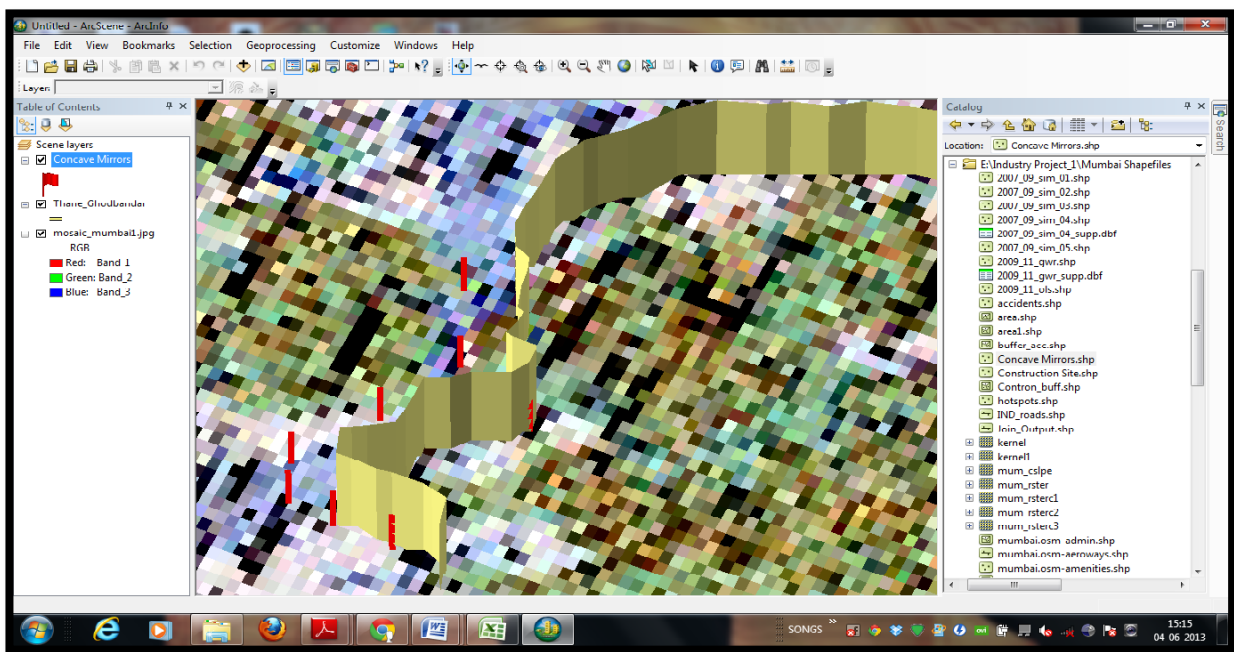
- Speed Check Signs at regular intervals.
- Safety Checks at locations where heavy constructions are taking place.
- With analysis we derived that, there were certain location eg. Gaimukh which was a hotspot for accident, but lacked any trauma centre nearby. From survey, it was also unveiled that, lack of such centres has also made a minor accident into a fatal crash and added to accident severity
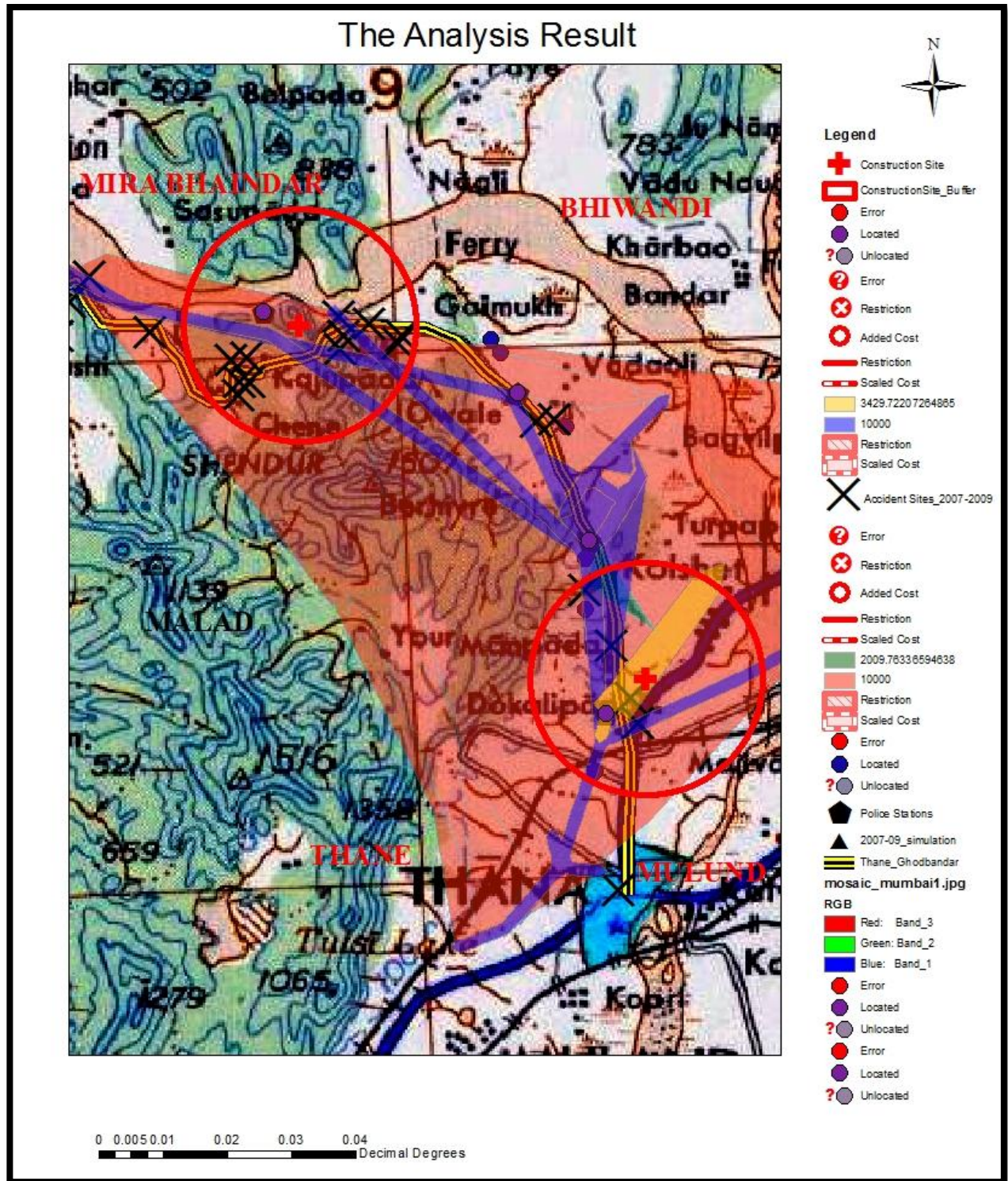
Figure 173 Analysis Result

BIBLIOGRAPHY

RESEARCH JOURNALS

- Affum, J. K. and M. A. P. Taylor, (1996). *"GIS-Based System for the Systematic Determination of Road Accident Problems and Causal Factors", Roads '96 Conference Proceeding–Part 5, ARRB Transport Research Ltd., Victoria, Australia,* 1996

- Affum, J. K. and M. A. P. Taylor, (1995). *"Integrated GIS Database for Road Safety Management", Technology Tools for Transportation Professionals–Moving into the 21st Century: Resource Papers for the 1995 International Conference, Institute of Transportation Engineers, Washington, D.C., and pp. 189-193.*

- Al-Ramadan, B., (2002). *Introduction to Geographic Information Systems Technology and Its Applications. Review 2002: A chapter in College of Environmental Design Book, KFUPM, Dhahran, 113-120*

- Alexandra Larsen, Melissa (2010). *Philadelphia Traffic Accident Cluster Analysis using GIS and SANET*

- Anderson, T., (2007). *Comparison of spatial methods for measuring road accident 'hotspots': a case study of London, Journal of Maps, v2007, 55-63*

- Anderson T., (2009), *Kernel Density Estimation and K-means Clustering to Profile Road Accident Hotspots, Accident Analysis Prevention, 41, 359-364*

- Anselin, L.: *Local indicators of spatial association – LISA. Geog. An. 27(2) (1995) 93–115*

- Austin, K., M. Tight and H. Kirby, *"The Use of Geographical Information Systems to Enhance Road Safety Analysis", Transportation Planning and Technology. Vol 20, 1997, pp. 249-266.*

- Aylo, A. (2010). *Developing A GIS Safety Analysis System: A Case Study*; *MS Thesis at University of Dayton, USA*

- Baker, S. P., A. Waller and J. Langlois, *"Motor Vehicle Deaths in Children: Geographic Variations", Accident Analysis and Prevention, Vol 23, Issue 1,February 1991, pp. 19-28.*

- Bureau of Transportation Statistics (1998), *Development of a Crash Referencing and Analysis System*

- CALIENDO, C., GUIDA, M. & PARISI, A. (2007). *A crash-prediction model for multilane roads. Accident Analysis & Prevention, 39, 657-670.*

- Choi, K. and I. Park (1996), *"Traffic Accident Analysis with GIS and Statistical Package", Ninth Symposium on Geographic Information Systems for Transportation. American Association of State Highway and Transportation Officials, Washington, D.C.*

- Deshpande, Nilesh, (2011). *Accident Mapping and Analysis Using Geographical Information Systems. International Journal of Earth Sciences and Engineering; Vol. 04, No 06 SPL, October 2011, pp 343-345*

- DVRPC. (2009). *Traffic Crash Analysis of the Delaware Valley. Philadelphia: Delaware Valley Regional Planning Commission*

- ERDOGAN, S. (2009). *Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. Journal of Safety Research, 40, 341-351.*

- 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 , 174-181*

- Geometric Design of Major/Minor Priority Junctions

- GIS-Based Crash Referencing and Analysis System: Summary Report, *Highway Safety Information System, Report No. FHWA-RD-99-081, U.S. Department of Transportation, February 1999*

- Goh, P. C., *"Traffic Accident Analysis Using Geoprocessing Techniques", Road and Transport Research. Vol 2, No. 2, 1993, pp. 76-85*

- GUNDOGDU, I. B. (2010). *Applying linear analysis methods to GIS-supported procedures for preventing traffic accidents: Case study of Konya. Safety Science, 48, 763-769.*

- Hooper, Rebecca and Goodier, Anthony, "*Geometry of Local Roads*"

- Hunt, Kevin. *New York State Accident Location Information System Analyzing the Past to Provide a Safer Future; NYS Department of Transportation, ESRI*

- Jorge Ricardo Ferreira. *Georeferencing Road Accidents with Google Earth: Transforming Information into Knowledge for Decision Support*

• • •

- Kuhn, Wolfgang (2010), *"Methodology for Checking shortcomings in the three-dimensional Alignment, Morgan State University, Department of Civil Engineering"*

- Libreros C, Ludwing and Garcia, Alfredo, (2007), *"Safety effect of the skew angle in right turn maneuvers", 3rd Urban Street Symposium, Seattle, Washington.*

- Mannering, *"Geometric Design" , pp. 45-72 (4th edition) pp. 45-75 (previous edition)*

- Meyer, M. D. W. and Sarasua, *"Geographic Information System-Based Transportation Program Management System for County Transportation Agency", Transportation Research Record 1364, Transportation Research Board, Washington, D.C., 1992.*

- Miller, J., *"What Value May Geographic Information Systems Add to the Art of Identifying Crash Countermeasures?" Virginia Transportation Research Council, VTRC 99-R13, Charlottesville, Virginia, April 1999.*

- Miller, S., T. Johnson, S. Smith and L. Raymond, *"Design and Development of a Crash Referencing and Analysis System", Proceedings of the 1995 Geographic Information Systems for Transportation (GIS-T) Symposium, American Association of State Highway and Transportation Officials, Washington, D.C., 1995.*

- Raub, R.A., *"Occurrence of Secondary Crashes on Urban Arterial Roadways", Transportation Research Record 1581, Transportation Research Board, Washington, D.C., 1997*

- Ross, Harper, and Joshua, Aldridge (2010). *Motorway Crash Analysis Using GIS*

- Saccomanno, F. F., K. C. Chong, and S. A. Nassar*, "Geographic Information System Platform for Road Accident Risk Modeling", Transportation Research Record 1581, Transportation Research Board, Washington, D.C., 1997*

- Shaw, S.; Rodrigue, J. (2006). *"Geographic Information Systems for Transportation (GIS-T", The Geography of Transport Systems, CC. Jean-Paul Rodrigue, Brian Slack (Ed.), Routledge, New York.*

- Sun, Z. (2003). *"A GIS System for Managing and Analyzing Urban Road Traffic Accident" Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 4, October 2003, pp 1442-1449.*

- Vrachnou, Amalia (2003). *An Analysis of Emergency Vehicle Crash Characteristics; Civil Engineering, Falls Church, Virginia*

- Yasunori, M.; (2000) *"GIS-based study on Road Traffic Assignement Models in Sendai City", Annual Report of the Engineering Research Institute, University of Tokyo*

- Yu, D.; Li, J.; Chen, H. & Yi, X. (2009). *A Research on Time Distribution of Traffic Accidents on Road, Proc Second Int. Conf. Intelligent Computation Technology and Automation ICICTA '09, 3, pp. 448-451.*

- Yves, P., *"A Statistical Model to Compare Road Mortality in OECD Countries", Accident Analysis and Prevention, Vol 33, 2001, pp371-385.*

- Xie, X and Yan, J. (2008). *"Kernel Density Estimation of traffic accidents in a network space," Computers, Environment and Urban Systems, 32 (5), pp. 396-406, 2008*

STATISTICAL JOURNALS

- Economic Profile of Maharashtra

- Dayal Sharma, R. , Jain, S. and Singh, K., *"Growth Rate of Motor Vehicles in India - Impact of Demographic and Economic Development"*

- Road Accidents in India, 2010, *"Government of India, Ministry of Road Transport and Highways, Transport Research Wing, New Delhi"*

- Basic Road Statistics of India, 2004-05, 2005-06, 2006-07 and 2007-08, *"Government of India, Ministry of Road Transport and Highways, Transport Research Wing, New Delhi", July 2010*

- Basic Road Statistics of India, 2008-09, 2009-10 and 2010-11, *"Government of India, Ministry of Road Transport and Highways, Transport Research Wing, New Delhi", August 2012*

- Motor Transport Statistics of Maharashtra, 2010-2011

WEBSITES

- http://www.indigeo.com/indiadata.html
- http://www.portal.gsi.gov.in/portal/page?_pageid=127,1213918&_dad=portal&_schema=PORTAL
- http://www.scribd.com/doc/23329517/GIS-TOPOSHEET
- http://www.caraccidentlawyerpros.com/Car-Accident-Causes.html
- http://www.smartmotorist.com/traffic-and-safety-guideline/what-causes-car-accidents.html
- http://www.lib.utexas.edu/maps/ams/india/
- http://metro.teczno.com/
- http://recepzihni.org/JECOSS/JOURNAL.OF.ECONOMIC.AND.SOCIAL.STUDIES-1-2%20indexed/JOURNAL.OF.ECONOMIC.AND.SOCIAL.STUDIES-1-2_p137-p153.pdf
- http://www.bricsindia.in/publication/chapter3.pdf
- http://www.mapsofindia.com/census2011/population.html
- https://www.cia.gov/library/publications/the-world-factbook/geos/in.html
- http://code.msdn.microsoft.com/ASPNET-MVC-4-with-Bing-Maps-947ae78c
- http://rbrundritt.wordpress.com/2012/01/05/using-bing-maps-in-winforms/
- https://developers.google.com/maps/documentation/javascript/
- http://www.bing.com/community/site_blogs/b/maps/archive/2012/11/05/getting-started-with-bing-maps-windows-store-apps-native.aspx
- http://www.atkearney.com/documents/10192/19079b53-8042-43ea-b870-ef42b1f033a6
- http://www.isuppli.com/Automotive-Infotainment-and-Telematics/Pages/Advanced-Driver-Assist-Systems.aspx
- http://analysis.telematicsupdate.com/insurance-telematics/video-insurance-telematics-and-driver-behavior
- http://www.automotive-fleet.com/channel/gps-telematics/news/list/tag/driver-assistance-systems.aspx?prestitial=1
- http://www.diva-gis.org/Data
- http://www.icpsr.umich.edu/CrimeStat/files/CrimeStatChapter.6.pdf

- http://giscollective.org/tutorials/gis-techniques/spatial-statistics/hot-spot-analysis/
- http://www.ucl.ac.uk/jdi/events/int-CIA-conf/ICIAC11_Slides/ICIAC11_3D_SChainey
- http://research.cs.tamu.edu/prism/lectures/pr/pr_l7.pdf
- http://isc.temple.edu/economics/Econ616/Kernel/ast_part1.pdf
- http://www.cdc.gov/dhdsp/maps/gisx/training/module3/files/2-module.pdf
- http://help.arcgis.com/en/arcgisdesktop/10.0/help/
- http://developer.here.com/javascript_api_explorer
- www.wix.com
- http://safety.fhwa.dot.gov/fourthlevel/interbriefing/01prob.htm
- http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/layerlist.html#CensusStatisticalData
- http://www.esri.com/data/free-data
- http://www.mapcruzin.com/download-free-arcgis-shapefiles.htm
- http://www.gerhardbechtold.com/LUPMIS/Manual/525_georeference_vector_map.html
- http://glcfapp.glcf.umd.edu:8080/esdi/index.jsp
- http://virginiadot.org/infoservice/vdot-welcome.asp#HighwaySystem
- http://www.aimsun.com/wp/
- http://www.tcs.com/resources/brochures/Pages/Mobile_Telematics_Solution.aspx
- http://www.theory.tifr.res.in/
- http://www.engineersgarage.com/
- http://www.mid-day.com/news/2012/apr/010412-Mumbais-10-Most-dangerous-Roads-part-2.htm
- http://www.vdstech.com/aspmap.aspx
- http://punetrafficpolice.gov.in/index.php/en/statistics/accident-statistics.html
- http://www.voiceofbharat.org/accidents/road.asp
- http://www.thaneweb.com/web/thane_basic.html
- http://www.thaneweb.com/web/thane-business-directory/hospitals-thane.html
- http://www.thanecity.gov.in/
- http://www.thanecity.in/thane-traffic-information
- http://www.cambridgeshire.gov.uk/transport/safety/methodologies/accident+sites.htm

• • •

- http://www.loudoun.gov/business/growth%20summary%201999/2-1.pdf
- http://www.ohio.com/mld/beaconjournal/news/state/3066787.htm
- http://www.usfa.fema.gov/dhtml/inside-usfa//ff_fat.cfm
- http://www.virginia.edu/~risk/guardrail/151_175.ppt
- http://www.virginiadot.org/quick/nova_quick.asp
- http://www.co.fairfax.va.us/comm/demogrph/popbut.htm
- http://www.co.prince-william.va.us/demographics/population.htm
- http://www.emsnetwork.org/ambulance_crashes.htm
- http://www.fairfaxcounty.gov
- http://www.jems.com/jems/e0212a.html

ANNEXURE

Questionnaire for the Survey

Name of interviewer: _____

Telephone: _____

I'm conducting a survey on accidental injury. This survey aims to grasp the accidental situation of the population along the Thane-Ghodbunder Road, Maharashtra, including pattern, cause and treatment of these accidents, and the preventive measures taken by the public. Information provided will be strictly kept confidential and will be solely used for statistical purpose. Thanks!

DATE	START TIME	END TIME	DURATION	No. OF HOUSEHOLDS
09.02.2013	10:00 AM	4:00 PM	6.0 hours	12
23.02.2013	01:00 PM	06:00 PM	4.0 hours	10
24.02.2013	11:30 AM	05:30 PM	6.0 hours	15
08.03.2013	12:00 PM	05:00 PM	5.0 hours	10
15.03.2013	12:00 PM	04:30 PM	4.30 hours	8

Table 27 Details of the Survey Conducted

DETAILS OF THE SURVEYEE

Name:

Age:

Sex:

Employment Status:

CONDITIONS AT THE TIME OF THE ACCIDENT

Time of day:

☐ Full daylight   ☐ Dusk   ☐ Night

Visibility:

☐ Excellent   ☐ Good   ☐ Fair   ☐ Poor

Road Conditions:

☐ Dry   ☐ Damp   ☐ Wet

<u>ACCIDENT RELATED QUESTIONS</u>

Accident Cause:

Did the accident result in an injury (Y/N):

Nature and Extent of Injury:

Action(s) Taken:

Eye witnesses around:

<u>THE FOLLOWING QUESTIONS PERTAINS TO THE VEHICLE</u>

Vehicle Type:

☐ Car   ☐ Truck

☐ Van   ☐ Bus

☐ Other: _____

Victim's Position:

☐ Driver

☐ Passenger --------- Location ----------   ☐ Left   ☐ Middle   ☐ Right

☐ Other ---------------- ☐ Front Passenger   ☐ Rear Passenger   ☐ Third Seat (rear)

Collision Type:

☐ Driver Side Impact   ☐ Head on Collision   ☐ Front Impact

☐ Passenger Side Impact   ☐ Rear Impact   ☐ Pedestrian Incident

<u>THE FOLLOWING QUESTIONS PERTAINING TO THE OTHER VEHICLE</u>

Vehicle Type:

☐ Car   ☐ Truck

☐ Van   ☐ Truck

☐ Bus   ☐ Other: _____

Victim's Position:

☐ Driver

☐ Passenger --------- Location ----------   ☐ Left   ☐ Middle   ☐ Right

☐ Other ---------------- ☐ Front Passenger   ☐ Rear Passenger   ☐ Third Seat (rear)

Collision Type:

☐ Driver Side Impact   ☐ Head on Collision   ☐ Front Impact

☐ Passenger Side Impact   ☐ Rear Impact   ☐ Pedestrian Incident

APPENDIX