Revisiting Traffic theories for different models in lights of Pre-Crash Traffic Condition

MD. MAHMUD HOSSAIN

Department of Civil & Environmental Engineering

Islamic University of Technology

2015

Revisiting Traffic theories for different models in lights of Pre-Crash Traffic Condition

MD. MAHMUD HOSSAIN
Student ID: 115431

A THESIS SUBMITTED FOR THE DEGREE OF SCIENCE IN CIVIL ENGINEERING

Department of Civil & Environmental Engineering

Islamic University of Technology

2015

APPROVAL

This is to certify that the thesis submitted by Md. Mahmud Hossain entitled as "**Revisiting Traffic theories for different models in lights of Pre-Crash Traffic Condition**" has been approved fulfilling the requirements for the Bachelor of Science Degree in Civil Engineering.

____

Moinul Hossain

Assistant Professor,
Department of Civil and Environmental Engineering (CEE)
Islamic University of Technology (IUT)
Board Bazar, Gazipur, Bangladesh.

DECLARATION

I declare that the undergraduate research work reported in this thesis has been performed by me under the supervision of Assistant Professor Moinul Hossain. I have exercised reasonable care to ensure that the work is original and has not taken from the work of others.

____

Md. Mahmud Hossain
Student ID: 115431
November, 2015

ACKNOWLEDGEMENTS

Writing a significant scientific thesis is hard work and it would not be possible without support. All praises to Allah (SWT) for giving me the opportunity to complete this report. I would especially like to thank my mother and father who have provided me with the strength and dedication needed to complete this thesis.

It is the proud privilege to acknowledge my sincere and deepest sense of gratitude to my respected Supervisor Moinul Hossain, Assistant Professor of Civil and Environmental Engineering for his untiring efforts, inspiration, invaluable guidance and constructive criticisms throughout the period.

I would like to express gratitude to all of the Departmental faculty members for their help and support.

I am also thankful to my friends and well-wishers for their valuable suggestion and cordial assistance.

ABSTRACT

The concept of predicting the probability of a road crash on an access controlled road is gaining momentum in recent years. Considerable research has been carried out to establish several statistical and artificial intelligence based models which can take real-time traffic data as input to predict the short time crash probability. In general, the underlying methodology builds knowledge about traffic condition during the time periods when there is no crash taking place from the historical detector data and based on that, if they find the real-time traffic data to be differing significantly from the normal data, they classify it as crash prone. Interestingly, although these models heavily deal with the basic traffic flow variables, such as, speed, flow, density, and/or their descriptive statistics or surrogate measures to build and operate such models, very few studies have been focused on whether and how the fundamental relationships of traffic flow theories differ for pre-crash and normal traffic conditions. The main purpose of this research is to revisit classical two-phase traffic flow theories in lights of pre-crash and normal traffic conditions and identify the variations in traffic conditions using pre-crash and normal traffic data.

The data for the analysis was extracted from 210 detectors from March 2014 to August 2014 on the Shibuya 3 and Shinjuku 4 routes of Tokyo Metropolitan Expressway in Japan. The detector data consisted of information on speed, vehicle count, occupancy and number of heavy vehicles aggregated for each minute for each lane of the study area. During this time 620 crashes took place. Matching the crash and the detector data, pre-crash and normal traffic condition datasets were prepared. For pre-crash data, for each crash point, data for each minute for the ten minutes leading to crash were collected from the nearest detector, nearest upstream and nearest downstream detectors. For normal traffic conditions, detector data was collected through random sampling with the filter that ensured that none of the data belonged to any timestamp where a crash took place on that direction of the route within three hour before or after that time.

Afterwards, four classical two-phase traffic flow models: Greenshields, Greenberg, Underwood and Bell-shaped models, were separately built for each detector location and

each timestamp using the pre-crash dataset. Similarly, the four relationships were generated using the normal traffic condition dataset.

The results show that all four classical two-phase traffic flow models are held true for both normal and pre-crash traffic conditions. Downstream detectors and nearest detectors could explain crash data better than upstream detectors. Free flow speed and jam density values of classical models which are depended on model accuracy can explain the variations between normal and hazardous traffic condition.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background

Traffic flow theories can be called the foundation of Traffic Science building the backbone of transportation engineering. They have various applications in transportation and urban planning (Velasco and Saavedra, 2008). Traffic flow theories mathematically represent a system in which many vehicles are interacting with each other. They, in microscopic level, provide understanding about the phenomena related with the movement of individual vehicles along a roadway and in macroscopic level explains the behavior of a traffic stream. These theories are essential in determining the fundamental characteristics of roadways such as their ability to sustain in various traffic flow level and for given set of capacities.

In traffic flow theory a basic distinction is made between microscopic and macroscopic traffic flow models and their variables. Macroscopic model characterizes the traffic as a whole and considers the relationships of traffic characteristics like speed, flow, density of traffic stream. Microscopic traffic flow model considers each of the individual vehicles in road network and how they interact with each other. Traffic flow models can be used to establish traffic simulation based on characteristics of the flow (macroscopic) or individual vehicles (microscopic).

The basic idea of traffic flow theory is based on the concept that flow of fairly heavy traffic appears like a stream of a fluid. A macroscopic theory of traffic can be developed considering traffic stream as an effectively one-dimensional compressible fluid by following hydrodynamic theory of fluids. The behavior of sizable aggregated vehicles is taken under consideration in this analogy. All traffic stream models and theories have to satisfy the rules of conservation of the vehicle numbers on the road.

Researchers have been involved in developing traffic flow models since the first half of the last century. As discussed earlier, macroscopic traffic stream models represent how the behavior of traffic flow parameters (speed, flow, density) changes with respect to one another. Around 80 years ago, the first traffic flow theory was proposed by Greenshield (1933) observing traffic on a highway. He performed tests associated with determination of traffic flow, traffic density and speed using photographic measurement methods for the first time and the derived model established relationships between traffic flow variables (speed, flow, density). The model was simplistic in nature and he assumed a linear relationship between traffic speed and density. However, in reality, it is very hard to find such a relationship between speed and density which saw the emergence of more complex models.

Greenberg's logarithmic model (1959) was established to answer the drawback of Greenshield's model which assumed a logarithmic relation between speed and density. This model has achieved much popularity because of its simplistic analytical derivation. Main drawbacks of this model is when density tends to zero the speed tends to become infinity. For this reason, Greenberg's model does not perform well in predicting speed at low density. To overcome the limitation of Greenberg's model, Underwood (1961) assumed an exponential relationship between speed and density. The limitation of this model is when speed becomes zero, density reaches infinity. Hence this cannot be used under high density traffic conditions. Drake et al. (1967) relate traffic flow model with bell-shaped curve model and proposed that speed-flow specification associated with the normal distribution curve. Dagnazo (1994) proposed that fundamental relation is triangular in the density– flow plane. Hence, since 1933, several traffic flow theories have evolved each with its own strength and drawbacks.

Apart from being the essential parts of the traffic flow theories, the basic traffic flow variables – speed, flow and density, are used for many activities related to traffic engineering. One such emerging field is road safety. Road crashes have adverse effects on both the health of the people and can be harmful to the economy. High amount of money has been needed to recover the conditions created by road accidents. Due to road accidents, high income countries spent around 518 billion USD per year and medium/ low income countries spent 65 billion USD per year (WHO, 2004). Authorities of different countries have taken various steps to reduce road crashes. The classical approaches have divided causes of crash based on human, vehicle and road and environment related factors as presented with Figure 1.1. Among these factors, human factors are responsible for most of the crash cases. Several studies related to road safety have identified different factors responsible for road crashes and minimize the impacts of the selected factors on road crashes. However, most of the initiatives have been geared towards vehicle and road and environment related factors. In recent times, some studies are investigating the possibility to involve traffic flow variables for predicting road crashes in real time. Such models are known as real-time crash prediction models. These models are based on the hypothesis that some combinations of speed, flow and density can result in traffic conditions causing discomfort to drivers leading to mistakes that may result in as crashes. Now-a-days, especially in the developed world, a large number of roads, especially access controlled roads are equipped with number of sensors cameras, magnetic sensors, infrared sensors, microwave sensors, laser sensors, and inductive loop detectors (Lawrence A. Klein, 2006). These devices are making it easier to obtain data of traffic flow variables in real time.

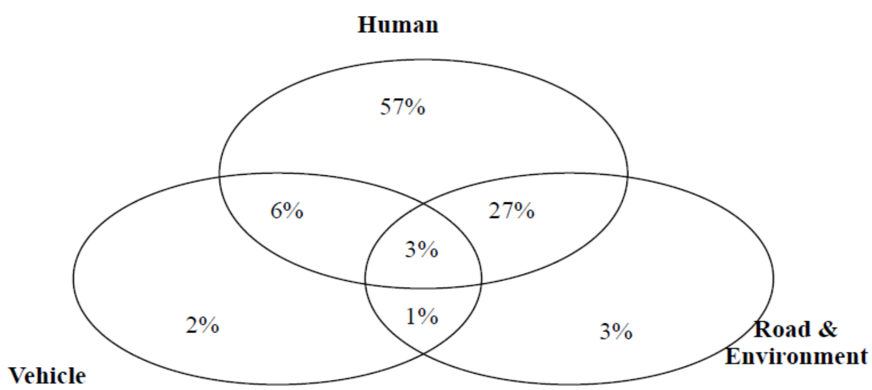

Figure 1.1: Factors Contributing to Road Crashes in the USA (Treat et al., 1977)

1.2 Problem Statement:

The research development in real-time crash prediction has so far been focused on distinguishing pre-crash data from normal traffic data through various kinds of prediction models.

Likelihood of crash occurrence was developed based on real time traffic data where statistical relationships showed that speed variation was an important variable (Oh et al., 2001). Golob and Recker (2001) investigated relationships of the flow of traffic, and weather and ambient lighting conditions with different crashes. The results proved that median traffic speed and temporal speed variations are strongly related to the type of collisions. Lee et al. (2003) developed a log-linear model to predict crash occurrence by establishing relationships between coefficient of variation in speed, traffic density and speed difference to identify crash likelihood. Abdel-Aty et al. (2005) developed real-time crash risk prediction models considering different speeds of traffic conditions based on logistic regression because mechanisms of multi-vehicle crashes were changing with different speed regimes. Relationships were building up between severe crash occurrence and the chosen variables. Separate real-time crash prediction model was developed for freeway and ramp (Hossain and Muromachi, 2011). The level of congestion and speed difference between upstream and downstream were used as the most important predictors of crash occurrence. Xu et al. (2012) used clustering analysis to establish connection between different traffic states and crash risks on freeways. Crash prediction models were developed using crash reports, real-time traffic and weather data (Yu and Abdel-Aty, 2014). Random forest model was firstly performed to identify the most important variables associated with crash occurrence.

However, very few studies have focused on whether and how the fundamental relationships of traffic flow theories differ for pre-crash and normal traffic conditions.

1.3 Purpose and Objectives

The purpose of this research is to revisit traffic flow theories in lights of pre-crash and normal traffic conditions. The specific objectives are:

Objective 1: Derive and compare classical two-phase traffic flow relationships for pre-crash and normal traffic condition data. This will include deriving four classical traffic flow theories – Greenshields, Greenberg, Underwood and Bell-shaped models, separately for pre-crash and normal traffic data.

Objective 2: Evaluate longitudinal variation of traffic flow relationships both in time and space scale for pre-crash traffic data, and,

Objective 3: Discuss the variations in traffic conditions for pre-crash and normal traffic data, if any.

1.4 Scope

The main scope of this study is to investigate two phase traffic flow theories using macroscopic traffic flow data for different models with different traffic conditions (Pre-Crash and Normal). Identifying the pattern of change occurrence during pre-crash traffic flow condition from normal traffic flow condition is also an important scope in this study.

1.5 Limitations

This Study has several limitations. Only speed-density relationship has been used for performing analyzes. Any variables are not considered that are related to weather though human factors are sometimes depended on certain weather condition. Statistical Analyses, through hypothesis testing are not performed evaluating the significance of differences between normal and hazardous traffic conditions. Random sampling has been performed only one time for establishing normal traffic dataset.

CHAPTER 2
LITERATURE REVIEW

2.0 General

This chapter provides a self-contained summary of all the major concepts and developments needed to be known for understanding the state of the art of this thesis topic.

2.1 Road crash and its causes

A Road crash is the most unwanted thing to happen to a road user, though they happen quite often. According to WHO (World Health Organization), road crashes are 9th major cause of death globally. Crash occurrence is not only the waste of valuable time but also increasing many consequence activities such as repair, maintenance and road clearance. Road crash prediction is quite difficult because causes of road crashes are related with various factors like road conditions, driver's behavior as human factors, environment conditions, rules and regulations on roads for road safety. Many researchers have suggested that crashes can occur even if the vehicle, environment or weather and road geometry are favorable to safe driving. This happens due to sudden formation of hazardous traffic condition causing driving uncertain behavior and discomfort. Human factors such as over speeding, drunken driving, distractions to driver are directly or indirectly responsible for road crashes compared to road, environment and vehicle related factors (Sabey et al., 1975). A similar study by Treat et al. (1977) on US data discovered that road/environment and vehicle related factors are very much less responsible compared to human factors for road crashes (Figure 1.1). To understand crash phenomena, addressing human factors are the most important task. But Human factors like different types of human errors related with road crashes are hard to measure because these factors can't be captured through experimental setup due to associated risk.

2.2 Crash phenomena and its determinants

Studies to find out relationships between crashes and traffic flow can be classified into two types: (a) aggregate studies, which is based on number of crashes or crash rates for a certain time periods (typically months or years) and for specific roads or networks and statistical distributions of parameters or variables are used to understand of traffic flow for similar time and roads or networks (b) disaggregate studies based on details of each of the crashes and parameters or variables of traffic flow are used to represent traffic flow for each of the crash time and place (Golob et al., 2004). In this work they established relationship between different types of crashes and real-time traffic flow using 30-second interval volume and occupancy data. Among two types of studies, it is found that Aggregated traffic flow variables (speed, flow, density) can be considered as a surrogate measure of human factors (Hossain and Muromachi, 2013). Macroscopic studies use crash data in terms of Vehicle Miles of Travel (VMT) which is mostly related with long time periods. Relationships between human factor variables (speed, flow, density) and traffic flow theory are important for understanding crash phenomena. Traffic flow theory is depended on the relationships between speed, flow, occupancy and space mean speed (Lighthill and Whitham, 1955; Richards, 1956). For thinking more deeply the relationships between speeds, flow, and density in different models are important.

2.3 Classical traffic flow theories

Around 80 years ago, traffic flow relationships had been developed. Different traffic models generate different relationships considering different sort of conditions. Every traffic model has some limitation and different models have been established to minimize the limitations. It has been found that different models represent different traffic flow condition under some considerations. Different models have been developed to establish relationships of variables like speed-flow-density. Display of the appropriate models can provide the traffic manages the traffic flow characters of different roads which can help them to do more research on the traveler’s travel behaviors, traffic ramp control, and congestion analysis. Besides availability of different models in describing traffic flow theory is an important application for developing modern traffic control theory by finding adequate solutions of problems. Four models- Greenshields model(1935), Greenberg model(1959), Underwood model(1961), Bell shaped curve model(1967) have been used widely to characterize traffic flows as well as traffic conditions. Greenshields(1935) succeeded to develop a model of uninterrupted traffic flow that explains the sequences that are observed in real traffic flows and assumed a linear relationship between speed and density. It is mostly accurate and relatively simple. Greenberg (1959) has achieved a great success because this model can be derived analytically and established a non-linear relationship between speed-density. When density tends to zero, speed tends to infinity. Underwood (1961) tried to minimize the limitations on Greenberg model and established that speed becomes zero only when density reaches infinity which is the major drawback of this model. Drake (1967) proposed relationships that are based on normal statistical distributions such as bell shape.

2.4 Traffic flow variables and crash prediction

To understand the comparison of different models relationships investigate these relationships with real time traffic flow data is essential. Many researchers have been worked on traffic flow variables and their relationships for different models. Examples are: Verification of Flow-density relationship using strong data set collected from the Queen Elizabeth Way in Ontario (Hall et al., 1986), Analyzing the speed-flow relationship based on the traffic flow and travel time data in Singapore (Lum et al., 1998), speed-flow and flow-density relationships in normal traffic condition based on large data set (Nielsen and Jorgensen, 2008), Determine level of safety based on the speed-flow relationship of different lanes considering a certain traffic condition using Chengdu expressway data of China (Xie et al., 2011). For proper understanding crash phenomena connection between crash prediction and normal traffic flow condition has to be established.

Prediction of crash occurrence based on normal traffic flow condition is the most heavily studied topic in the development of real time crash prediction model as well as road safety. Using normal traffic flow condition that changes with time and different situations for identifying the likelihood of crash occurrence is a significant concept in road safety development. But there is no clear understanding about how traffic flow variables and their relationships effect on road safety. Many researchers have done various types of analysis associated with different models to identify what the changes of normal traffic flow relationships occur in pre-crash traffic flow conditions. But the approaches are not sufficient to predict crash occurrence using traffic flow variables. Using cluster analysis, it is found that hazardous traffic condition can be great extent distinguished from the normal traffic condition (Hossain and Muromachi. 2011). Question may be asked which relationships are more suitable for representing traffic flow conditions. Also important to identify the methods to compare normal traffic flow relationships with pre-crash traffic flow conditions.

As discussed earlier, different relationships can be generated to understand traffic flow conditions. These relationships can be flow-speed, speed-density and flow-density. For any type of research, it will be very much difficult to work with every possible relationship. To avoid the problem, most preferable relationship has to be identified. The relationship of speed-flow or speed-density is depended on only road type and free flow speed. In most of the research work (mentioned earlier) speed-density proves to be good representative of traffic flow condition (Normal or Hazardous). Lee et al. (2002) established the concept of pre-crash condition and determine the likelihood of crash is significantly depended on short-term turbulence of traffic flow. They have identified two important factors of road crashes- speed and density.

Relationships are established based on data. Accuracy of data related with proper data collection method is significant to finding the actual result. Using loop detector data in road crash prediction for road safety is still in preliminary stage. Modern sensor technologies have the ability to identify vehicle presence by vehicle signature and measure number of vehicle count as flow/occupancy and speed. Traffic flow parameters such as density, travel time and origin-destination pairs can be collected by installing image processors along roadway section (Mimbela and Klein, 2007). Detector availability is also essential to get the required data for analysis. For pre-crash traffic flow condition, this importance is much higher. Obtaining downstream/ upstream/ nearest detector data is required to find out the comparison of different detector data before crash occurrence. Availability and accuracy of real-time traffic flow data are important to identify correct the relationships of speed, flow, density. Inductive loop detectors, magnetic sensors, video image processors, microwave radar sensors, infrared sensors, laser radar sensors have been used to determine traffic flow variables (speed, flow, density, space mean speed, travel time etc). These technologies have been chosen based on required accuracy/ size of database, cost and specifications for authorities (Mills and Gibson, 2006). Using loop detector data for establishing flow-occupancy relationship on aggregated studies has been done in previous studies (Coifman et al., 2000).

Classical equations for different models are to be generated based on relationships. Many researchers have focus on the characteristic of traffic flow during period of 1950-1970. A number of mathematical models were established to describe speed-density relationship and calibrated by fitting curves or relationship with respect to normal traffic data (May, 1991). Using traffic flow data classical equations for different fundamental models are generated to understand the significance or importance of variables.

Method of analysis must be suitable and reliable based on objectives and purpose of research work. Comparing traffic flow relationships for different models can be done to identify the difference between normal and pre-crash traffic flow conditions and identify which model represent or goodness fit for normal and pre-crash traffic flow. Besides this, it is very much important to identify the time period of change occurrence for developing road safety. A developed regression model has been discovered using traffic flow variable relationships for Japanese cities (Ohta and Harata, 1989). Regression model analysis on normal traffic condition for different types of roads has been done earlier (LU and MENG, 2013). But thinking not only for normal traffic condition but also for pre-crash traffic flow condition is important now-a-days. The time difference between beginning of changes occurrence and crash time helps to find out in taking some safety measurement for avoiding the occurrence of crushes.

The aim of the work is to compare speed-density classical curves to estimate functional relationships of traffic flow data on normal traffic condition and pre-crash traffic condition. Shibuya 3 and Shinjuku 4 routes of Tokyo Metropolitan Expressway have been used for data collection because of their availability of detector and generating different types of traffic conditions. Single loop detectors are distributed throughout the freeway system. A large number of recorded speeds, flow, density measurements are observed to estimate the consistency of relations. Based on traffic flow data, several regression curves are compared to find which models are fit better in normal traffic condition and hazardous traffic condition using correlation coefficient (R^2) values.

Flow, speed, occupancy data are available for different detectors. For this finding out the upstream/ downstream/ nearest detector data is much easier. Each of the detector data has been compared with normal traffic flow condition data to identify which represent mostly the pre-crash traffic flow condition. So another aim of this paper is to identify the appropriate representative models and detectors (from previous crash points) for pre-crash traffic flow condition using normal traffic flow relationships.

Chapter 3
Methodology

3.1 Introduction

A series of tasks have been performed to fulfill the research needs and stated objectives. Classical equations for different models (Greenshields, Greenberg, Underwood and Bell shape) have been derived from detector data. Then the equations for various traffic conditions were derived using simple linear regressions. The overall workflow is explained with Figure 3.1.

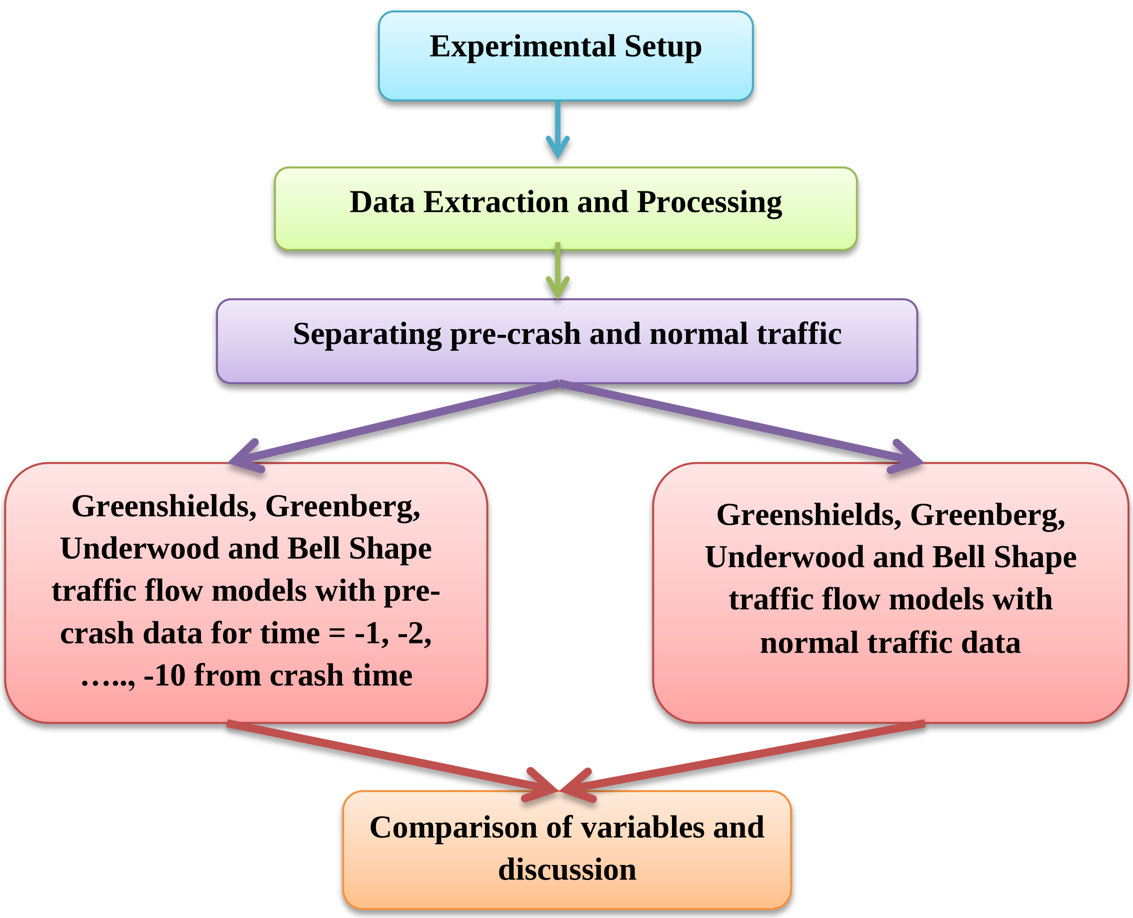

Figure 3.1: Overall workflow of the methodology

3.2 Study area and the data

For the proposed study, it was required to have a suitable site where substantial number of crashes take place and at the same time, the location has sections which are access controlled and heavily instrumented, which helps in collecting high resolution traffic flow data as well as crash data with excellent accuracy. Moreover, the sensor spacing is also expected to be roughly homogeneous in order to derive conclusion from special variations. Shibuya 3 and Shinjuku 4 routes of Tokyo Metropolitan Expressway are chosen as the study area as they fulfill all these criteria. They are situated in the center of Tokyo Metropolitan area connecting and covering some important business and residential areas. The length of Shibuya 3 and Shinjuku 4 routes are respectively 11.9 kilometers and 13.5 kilometers. A schematic diagram of the routes is provided in Figure 3.2 (not drawn to scale). This study area contains uniformly spaced detectors having average distance of 250 meters from each other. Total number of detector placed in two routes is 210. The routes sustain high number of crushes every year. Crash data and Normal data were collected for 6 months (March 2014 to August 2014) and 620 crash samples had complete data.

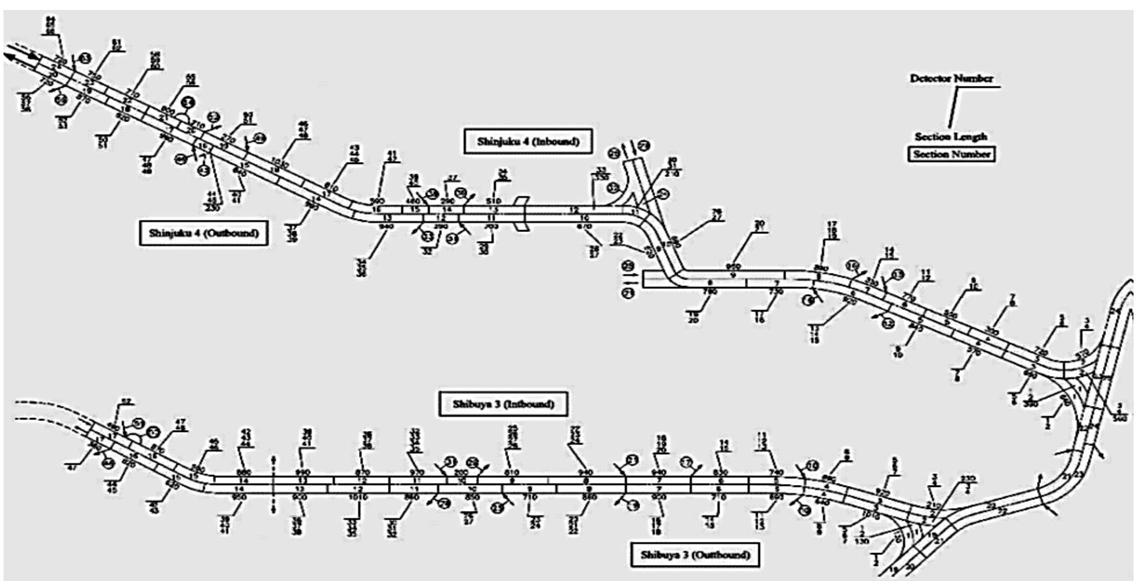

Figure 3.2: Shibuya 3 and Shinjuku 4 routes of Tokyo Metropolitan Expressway

Source: provided by Tokyo Metropolitan Expressway Company Limited

3.3 Experimental Setup

In Tokyo Metropolitan Expressways, data are stored for each detector for every eight milliseconds. The data is also stored for each lane covering information on speed, flow, occupancy and number of heavy vehicles. The authority aggregated the data for each detector for every one minute and provided us to conduct the research. Apart from detector data, they also provided us crash data which included information on date, time, location, vehicles involved, and types of lane. Also, a map was provided which included information on location of ramps, detector positions, and section length.

The pre-crash data was defined as crash data within the vicinity of the crash just before the crash time. Hence, it was important to decide upon the term 'near' and 'just before crash time'. Three locations were chosen – the closest detector (ND) location from crash, the nearest in the upstream (UD) and the nearest in the downstream (DD) as presented with Figure 3.3.

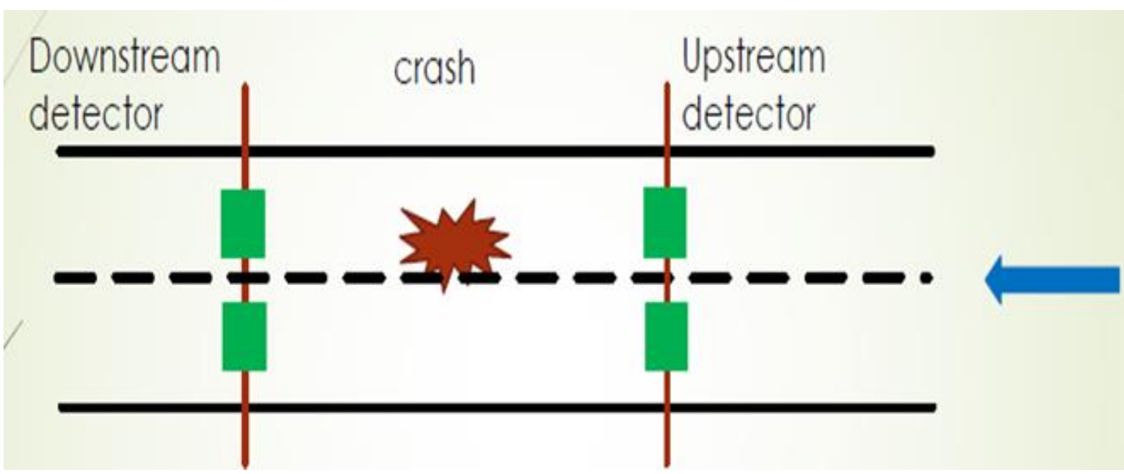

Figure 3.3: Detector location with respect to crash location

3.4 Data Preparation

As mentioned earlier, a total 210 are detectors are placed in the two routes, considering both directions. Hence, the primary task is to identify nearest detector, upstream detector and downstream detector for each crash point. Initially, the scope of the study was expected to cover up to two detectors both downstream and upstream. Hence, as it can be seen, DD2 suggests the second detector in the downstream from the crash location and so on. ND means nearest detector and NDB and NDA stand for Nearest Detector Before (NDB) and Nearest Detector After (NDA) as presented in Table 3.1.

Table 3.1: Identifying different type of detector for each crash point

Crash ID	DD2	DD1	ND	UD1	UD2	NDB	NDA
481	04	01	03	-	-	03	03
482	02	04	01	03	-	03	01
483	02	04	01	03	-	03	01
484	02	04	01	03	-	01	01
485	02	04	01	03	-	01	01
486	06	05	02	04	01	02	05
487	06	05	02	04	01	02	05
488	08	07	06	05	02	05	06
489	08	07	06	05	02	05	06
490	08	07	06	05	02	05	06
491	09	08	07	06	05	07	08
492	10	09	08	07	06	08	09
493	13	10	09	08	07	09	10
494	13	10	09	08	07	09	10

After identifying detectors, pre-crash data (speed, flow, density) from the corresponding crashes were extracted from the database. For each crash and for each detector location (say, ND) ten separate datasets were retrieved, each for each minute before crash. Hence, the pre-crash dataset considers traffic conditions for up to ten minutes before crash. As the data contained information on speed, volume, occupancy and number of heavy vehicles. As number of heavy vehicles is not a variable considered as a fundamental variable in four classical traffic flow theories selected in this study, it was dropped from the extracted dataset. Also, volume for each lane was merged into one number and the speed and occupancy represented the average for each detector. Table 3.2 presents the sample pre-crash data for the 5th minute before crash for the nearest detector (ND) for the first six crash data. The flow is in number of vehicles unit, speed is in kilometers per hour and occupancy is in percentage value.

Table 3.2: Pre-crash data set (for 5 min, ND)

Crash ID	flow	Speed	Occupancy
1	8	31.8	2.9
2	10	30	3.05
3	5	28.85	1.7
4	4	31.6	1.1
5	21	10.8	17.4
6	13	10.5	20.6

Normal dataset was compiled by taking random selecting detector data for various detectors at various locations. However, it was taken care of that no data is collected from any direction of any route 3 hours before and 3 hours after a crash. The total number of samples collected for normal data was 1289. A sample of the normal dataset is provided with Table 3.3

Table 3.3: Normal Dataset (Sample)

Flow	Speed	Occupancy
7	1.5	29.1
47	1.9	33.65
35	58.9	12
49	45.05	14.1
30	22.85	27.35

9	91.65	1.9
19	8.7	42.75
50	53.7	17.25

3.5 Classical Speed-density Models:

Fundamental relationship among three major variables of traffic flow theory: flow (q), speed (V) and density (K) is given below:

$$q = V * K \qquad (1)$$

Equation (1) is important to generate relationship between speed and density and other relationships as density-flow and speed-flow will be got automatically.

Around 80 years ago, first traffic flow model (Greenshields model) established to find out the relationships of major variables representing traffic flow condition. Greenshields (1935) suggested a linear relationship between speed and density.

$$K = K_j \left(1- \frac{V}{V_f}\right) \qquad (2)$$

Where K_j = jam density (when V = 0); V_f = free flow speed

May (1991) established classical speed-density relationships for some other important models using fundamental relationship.

Greenberg:

$$K = K_j e^{\frac{-V}{V_0}} \qquad (3)$$

Underwood:

$$K = K_0 \ln\left(\frac{V_f}{V}\right) \qquad (4)$$

Bell-shape:

$$K = K_0 (2\ln\frac{V_f}{V})^{0.5} \quad (5)$$

Where K_0 is the density at maximum flow level; V_0 is the speed at maximum flow level.

To get speed-flow model (if necessary), common suggestion proposed by many researchers was using density and speed data for establishing the mathematical models and then covert them using equation (1). From these equation, only K_j or K_0 and one of V_f or V_0 are need to know to get a speed-density relationship curve.

3.6 Simple Linear Regression Model:

Simple regression analysis always involves a single independent or predictor variable and a single dependent or outcome variable. The purpose of simple linear regression analysis is to make predictions about the value of the outcome or dependent variable in giving certain values of the predictor or independent variable. For example: in speed-flow relationship if density is dependent variable and speed is an independent variable, for a certain value of speed a predicted density value will be found.

Simple linear regression analysis produces a linear regression line:

$$\mathbf{Y = bX + a} \quad (6)$$

Where,

Y = predicted value of dependent variable;

X = value of dependent variable;

b = unstandardized regression coefficient or slope;

a = intercept (means the point where the regression line intercepts the Y axis)

Square of distance between data point and drawn line using data is called squared deviation for that point. Adding all of squared deviation for each of the data point together is called sum of squared deviations, or sum of squares. The sum of the squared deviations sometimes called sum of squares will be different depending on where line has been drawn in scatter plot. In any scatterplot, there is only one line that produces the smallest sum of squares and this line is the regression line. In short, the linear regression line represents the straight line that produces the smallest sum of squared deviations from the line.

From regression analysis, correlation coefficient (R^2) is also found which represents the percent of the data is the closest of the line of the best fit. It indicates how well the regression line represents the data. For example if R^2 = 0.65, it means that 65% of the total variation in Y can be explained by linear relationship between X and Y.

3.7 Deriving linear forms for classical traffic flow theories

Equation (2), (3), (4), (5) represents classical equations of speed-density relationship for different models. In order to run regression analysis, these non-linear models were transformed into linear equations as presented with Table 3.4

Table 3.4: Transformation from speed-density function

Model	Speed-density Function	Transformation
Greenshields	$K = K_j\,(1-\frac{V}{V_f})$	$K = K_j - K_j * \frac{V}{V_f}$
Greenberg	$K = K_j e^{\frac{-V}{V_0}}$	$\ln K = \ln(K_j) + (-\frac{1}{V_0})V$
Underwood	$K = K_0 \ln(\frac{V_f}{V})$	$K = K_0 \ln(V_f) - K_0 \ln(V)$
Bell Shape	$K = K_0 (2\ln\frac{V_f}{V})^{0.5}$	$K^2 = 2K_0^2 \ln(V_f) - 2K_0^2 \ln(V)$

CHAPTER 4
RESULTS AND ANALYSIS

4.1 Introduction

The results obtained through the previously outlined methodologies are presented in this chapter. In order to predict crash phenomena using traffic flow theory, Simple linear regression analyses are performed by using a popular open source statistical analysis program named “R”.

4.2 Normal Traffic condition

Figure 4.1 shows matrix scatter plot for normal traffic flow data. These kind of plot helps to identify the relationships between traffic flow variables such as speed, density and occupancy. Generally when density increases, speed will decrease but when flow increases, speed will increases to a certain limit and then start declining.

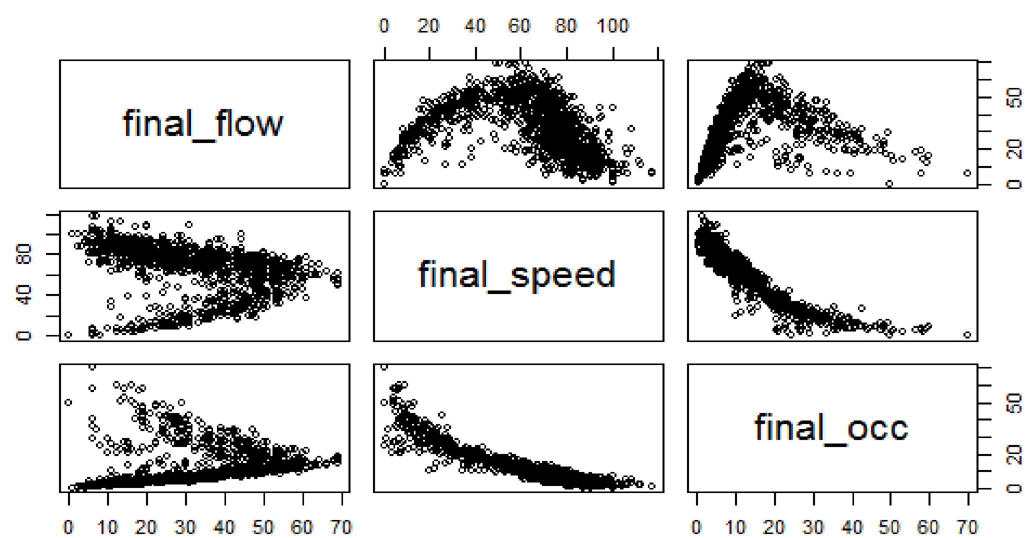

Figure 4.1: Matrix of scatter plots for normal traffic flow data

Figure 4.2 shows regression lines for four models (Greenshields, Greenberg, Underwood, Bell Shape) using R. Table 4.1 represents four regression line equations for normal traffic flow condition. Regression line equations help in estimating parameters (K_0, V_f, K_j, V_0) of models.

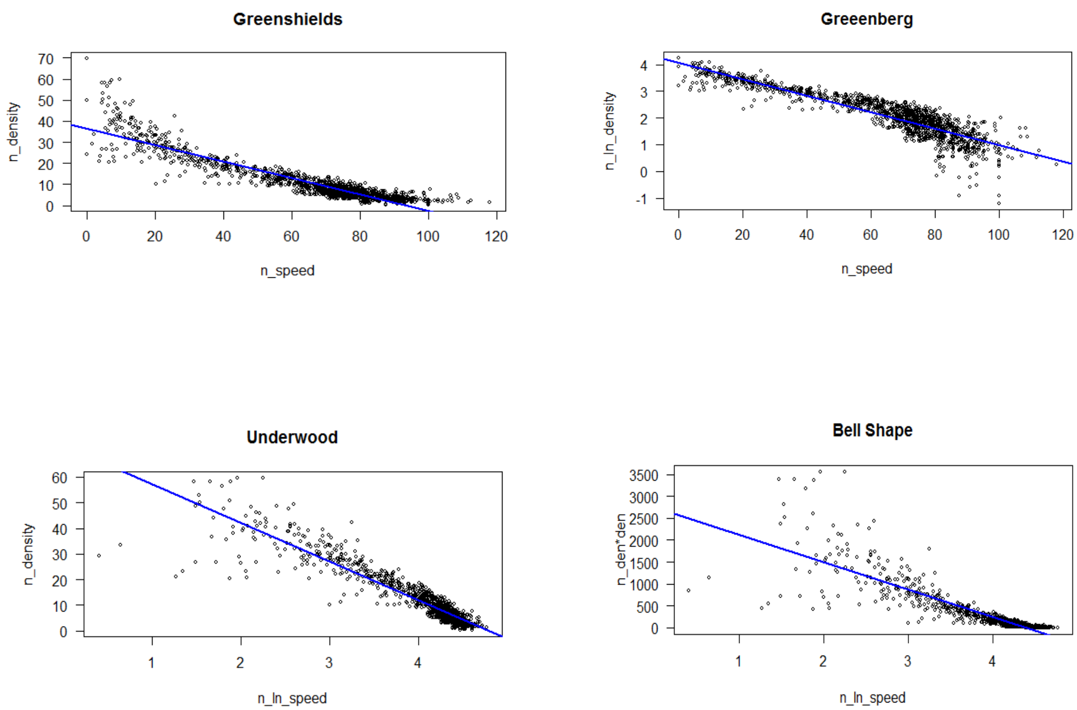

Figure 4.2: Regression lines for four models (Normal flow data)

Table 4.1: Regression line equations for normal traffic condition

Model	Transformation	Regression line equation
Greenshields	K = K_j - K_j* $\frac{V}{V_f}$	K = 36.585631 - 0.389404V
Greenberg	lnK = ln(K_j) + (- $\frac{1}{V_0}$)V	lnK = 4.0815583 - 0.0309193V
Underwood	K= K_0ln(V_f) - K_0ln(V)	K = 72.4114 - 15.1288lnV
Bell Shape	K^2 = $2K_0^2$ln(V_f) - $2K_0^2$ln(V)	K^2 = 2754.902 - 624.42lnV

From regression analysis, the correlation coefficients (R^2) and estimated parameters are found for each model.

Table 4.2: R^2 and estimated parameters value for normal traffic condition

Model	R^2	V_f	V_0	K_0	K_j
Greenshields	***0.8397***	93.95			36.586
Greenberg	*0.8044*		32.34		59.238
Underwood	***0.8562***	119.86		15.1288	
Bell Shape	*0.7544*	82.43		17.67	

Traffic density and speed relationship on the models of Greenshields, Greenberg, Underwood, and Bell-shape are much related with correlation coefficients of each model. The highest value of correlation coefficient is related to Underwood model with R^2 0.8562 and after that Greenshields model with R^2 is 0.8397 (Table 4.2).

Comparing the four model's parameters, the free flow speed V_f in Greenshields model is lower than Underwood model (93.95 and 119.86) because of the difference between R^2 values. As Bell Shape model's regression is not very accurate compared to other models($R^2 = 0.7544$), it's V_f only shows to be 82.43. V_0 can be assume to be 32.34 because there is only one estimation for this parameter presented in Greenberg Model in which the value of R^2 is 0.8044. K_0 is assumed to be 15.1288 as the underwood model is better than the Bell-shape model. K_j value is higher in Greenberg because of comparatively low R^2 value than Greenshields.

Table 4.3: Classical Equations with parameters (Normal Traffic Condition)

Model Name	Classical Equation
Greenshields	$K=36.586(1-\frac{V}{93.95})$
Greenberg	$K=59.238e^{(-\frac{V}{32.34})}$
Underwood	$K=15.1288\ln\frac{119.86}{V}$
Bell Shape	$K=17.67(2\ln\frac{82.43}{V})^{0.5}$

4.3 Comparison of R^2 values

4.3.1 Pre-crash traffic condition (UD1)

Regression Analysis has been performed in first upstream detector (UD1) up to 10 min before crash point. Table 4.4 shows the correlation coefficient values (R^2) for four models.

Table 4.4: Coefficient (R^2) values for UD1

Model Name	T=10	T=9	T=8	T=7	T=6
Greenshields	*0.5406*	*0.5377*	*0.525*	*0.5449*	*0.5403*
Greenberg	*0.6581*	*0.6948*	*0.6374*	*0.6426*	*0.6743*
Underwood	*0.733*	*0.6987*	*0.7264*	*0.7061*	*0.7226*
Bell Shape	*0.6493*	*0.5379*	*0.5518*	*0.5925*	*0.592*
	T=5	**T=4**	**T=3**	**T=2**	**T=1**
Greenshields	*0.5552*	*0.5337*	*0.5179*	*0.5472*	*0.549*
Greenberg	*0.6897*	*0.6054*	*0.6934*	*0.7005*	*0.6595*
Underwood	*0.7388*	*0.7455*	*0.7118*	*0.7747*	*0.7585*
Bell Shape	*0.5939*	*0.6086*	*0.4955*	*0.6165*	*0.6266*

The highest and lowest value of correlation coefficients (R^2) in different models are 0.552 and 0.5179; 0.7005 and 0.6054; 0.7747 and 0.6987; 0.6493 and 0.4955. Correlation coefficients (R^2) values are lower for pre-crash condition in each model compared to normal condition. Greenberg and Underwood model performs well for pre-crash condition. Each of the time periods, Underwood model performs better than Greenberg model. Greenshields model performance is the poorest. Bell shape performance does not follow any sequence.

4.3.2 Pre-crash traffic condition (DD1)

Regression analysis is performed in first downstream detectors (DD1) to identify whether the same principle of first upstream detectors (UD1) is applicable or not (Table 4.5).

Table 4.5: Coefficient (R^2) values for DD1

Model Name	T=10	T=9	T=8	T=7	T=6
Greenshields	*0.5933*	*0.6063*	*0.5748*	*0.6023*	*0.5909*
Greenberg	*0.6576*	*0.6703*	*0.7438*	*0.7571*	*0.7418*
Underwood	*0.7799*	*0.818*	*0.7931*	*0.7944*	*0.7673*
Bell Shape	*0.6336*	*0.7121*	*0.6355*	*0.6629*	*0.5959*
	T=5	**T=4**	**T=3**	**T=2**	**T=1**
Greenshields	*0.5985*	*0.5931*	*0.6227*	*0.5976*	*0.6025*
Greenberg	*0.7276*	*0.6986*	*0.7844*	*0.7388*	*0.7354*
Underwood	*0.7894*	*0.7835*	*0.7905*	*0.8085*	*0.7827*
Bell Shape	*0.6483*	*0.6634*	*0.6397*	*0.6811*	*0.5865*

The highest and lowest value of correlation coefficients (R^2) in different models are 0.6227 and 0.5748; 0.7844 and 0.6576; 0.818 and 0.7673; 0.7121 and 0.5865. Correlation coefficients (R^2) values are lower in each model compared to normal condition but comparatively higher than UD1. Greenberg and Underwood coefficient values are very close in some cases (T=6, T=3) which indicate hazardous traffic condition. Again, Greenshields model performance is the poorest. Bell shape performance does not follow any sequence.

4.3.3 Pre-crash traffic condition (ND)

The most important detectors in any crash phenomena are nearest detectors (ND) because these detectors are much closer to crash points. Regression analysis has been performed with great importance in ND (Table 4.6).

Table 4.6: Coefficient (R^2) values for ND

Model Name	T=10	T=9	T=8	T=7	T=6
Greenshields	*0.1962*	*0.2034*	*0.2203*	*0.2179*	*0.1968*
Greenberg	*0.611*	*0.6367*	*0.6342*	*0.6533*	*0.6236*
Underwood	*0.6733*	*0.6837*	*0.7165*	*0.6724*	*0.6532*
Bell Shape	*0.5608*	*0.5997*	*0.615*	*0.5468*	*0.5226*
	T=5	**T=4**	**T=3**	**T=2**	**T=1**
Greenshields	*0.2108*	*0.2182*	*0.2216*	*0.195*	*0.1951*
Greenberg	*0.6288*	***0.6921***	***0.6665***	***0.6594***	***0.6253***
Underwood	*0.6651*	***0.6842***	***0.6552***	***0.6101***	***0.5888***
Bell Shape	*0.5264*	*0.5093*	*0.4887*	*0.4374*	*0.3964*

Ranges of correlation coefficient (R^2) in different models are 0.195 to 0.2216; 0.611 to 0.6921; 0.6101 to 0.7165; 0.3964 to 0.615. Correlation coefficients (R^2) values are lower in each model compared to normal condition and pre-crash condition for DD1. Greenshields regressions are not so accurate for ND compared to UD1 and DD1. In some case Greenberg performs better compared to Underwood (T=4 to T=1). These two differences represent better pre-crash traffic condition.

4.3.4 Pre-crash traffic condition (NDB and NDA)

To understand the cause of change occurrence, regression analysis has to be performed for nearest detectors before (NDB) which are related with ND and UD1 and nearest detectors after (NDA) which are related with ND and DD1.

Table 4.7: Coefficient (R^2) values for NDB

Model Name	T=10	T=9	T=8	T=7	T=6
Greenshields	*0.3678*	*0.3671*	*0.3678*	*0.3655*	*0.3503*
Greenberg	*0.6465*	*0.634*	*0.6475*	*0.6316*	*0.628*
Underwood	*0.6972*	*0.6836*	*0.6968*	*0.6722*	*0.6637*
Bell Shape	*0.5995*	*0.5451*	*0.5326*	*0.5587*	*0.5212*
	T=5	**T=4**	**T=3**	**T=2**	**T=1**
Greenshields	*0.377*	*0.3722*	*0.3522*	*0.3643*	*0.3816*
Greenberg	*0.6363*	*0.6428*	*0.6457*	*0.6773*	*0.667*
Underwood	*0.6759*	*0.71*	*0.6499*	*0.6948*	*0.6854*
Bell Shape	*0.5301*	*0.5461*	*0.4301*	*0.5391*	*0.5338*

In table 4.7, Ranges of correlation coefficient (R^2) in different models are 0.3503 to 0.3816; 0.628 to 0.6773; 0.6499 to 0.71; 0.4301 to 0.5995. Greenshields R^2 values are slightly higher than ND but still lower than DD1, UD1. Greenberg and Underwood regression accuracy are very close in some cases (T=3 to T=1). These differences represent better hazardous condition as discussed earlier (section 4.3.2).

Table 4.8: Coefficient (R^2) values for NDA

Model Name	T=10	T=9	T=8	T=7	T=6
Greenshields	*0.2928*	*0.2974*	*0.3019*	*0.154*	*0.3028*
Greenberg	*0.61*	*0.6398*	*0.7053*	*0.7176*	*0.6857*
Underwood	*0.7492*	*0.7766*	*0.7852*	*0.7556*	*0.7637*
Bell Shape	*0.6345*	*0.683*	*0.6397*	*0.6327*	*0.6438*
	T=5	**T=4**	**T=3**	**T=2**	**T=1**
Greenshields	*0.3067*	*0.3009*	*0.3119*	*0.3087*	*0.2943*
Greenberg	*0.7238*	*0.697*	***0.7205***	***0.7327***	***0.6679***
Underwood	*0.7495*	*0.7317*	***0.7304***	***0.7183***	***0.6685***
Bell Shape	*0.603*	*0.5687*	*0.5702*	*0.552*	*0.4715*

Observing Table 4.8, it is found that correlation coefficients (R^2) in different models are varies between 0.154 and 0.3119; 0.61 and 0.7327; 0.6685 and 0.7852; 0.4715 and 0.683. Greenshields R^2 values are slightly higher than ND but lower than DD1, UD1, NDB. Greenberg performs similar (T=5) and better (T=3 to T=1) compared to Underwood in some cases. Still Bell shape performance does not follow any sequence.

Observing R^2 values of different models for each detector, it is found that ND, NDA, NDB and DD represent better hazardous traffic condition.

4.4 Comparison of estimated parameters

From regression analyzes, regression line equations have been found. These equations are used to estimate the parameters of each models and represented with classical equations.

Table 4.9: Classical Equations (speed-density) for UD1

Model Name	T=9	T=8	T=7
Greenshields	K=35.34(1- $\frac{V}{89.07}$)	36.25(1- $\frac{V}{87.42}$)	35.23(1- $\frac{V}{87.26}$)
Greenberg	K=44.95$e^{(-\frac{V}{36.57})}$	43.6$e^{(-\frac{V}{37.12})}$	43.53$e^{(-\frac{V}{36.57})}$
Underwood	K=14.1ln$\frac{126.42}{V}$	14.57ln$\frac{121.15}{V}$	14.13ln$\frac{123.72}{V}$
Bell Shape	K=19.16(2ln $\frac{71.92}{V}$)$^{0.5}$	67.31(2ln $\frac{67.31}{V}$)$^{0.5}$	18.62(2ln $\frac{73.03}{V}$)$^{0.5}$
	T=6	**T=5**	**T=4**
Greenshields	K=36.04(1- $\frac{V}{86.89}$)	35.35(1- $\frac{V}{86.81}$)	35.24(1- $\frac{V}{86.23}$)
Greenberg	K=44.04$e^{(-\frac{V}{36.52})}$	43.72$e^{(-\frac{V}{36.16})}$	42.06$e^{(-\frac{V}{36.08})}$
Underwood	K=14.37ln$\frac{120.67}{V}$	14.51ln$\frac{117.42}{V}$	14.97ln$\frac{112.29}{V}$
Bell Shape	K=19.15(2ln $\frac{71.16}{V}$)$^{0.5}$	19.1(2ln $\frac{70.61}{V}$)$^{0.5}$	19.63(2ln $\frac{68.62}{V}$)$^{0.5}$
	T=3	**T=2**	**T=1**
Greenshields	K=34.81(1- $\frac{V}{87.58}$)	35.72(1- $\frac{V}{86.18}$)	34.41(1- $\frac{V}{86.18}$)
Greenberg	K=43.39$e^{(-\frac{V}{35.57})}$	43.5$e^{(-\frac{V}{35.32})}$	41.97$e^{(-\frac{V}{35.67})}$
Underwood	K=14.55ln$\frac{115.68}{V}$	15.11ln$\frac{109.64}{V}$	14.63ln$\frac{112.04}{V}$
Bell Shape	K=18.84(2ln $\frac{72.72}{V}$)$^{0.5}$	19.84(2ln $\frac{67.36}{V}$)$^{0.5}$	18.74(2ln $\frac{69.97}{V}$)$^{0.5}$

From classical equations (Table 4.9), estimated parameters of each model can be found for UD1. V_f (85.15 to 91.112) and K_j (33.67 to 36.25) of Greenshields model are comparatively lower than normal condition (V_f=93.95, K_j=36.586) due to low R^2 values compared to normal data. V_0 (35.32 to 37.78) of Greenberg are slightly higher than normal data (V_0=32.34) and corresponding K_j (41.97 to 44.9491) are comparatively higher than Greenshields but lower than normal (K_j=59.238). Underwood V_f (109.646 to 126.42) are much fluctuating due to better performance for pre-crash condition and Bell

Shape V_f (67.31 to 74.053) can be assumed to be lower compared to normal condition (V_f=82.43). No considerable change has been found in parameter K_0.

Table 4.10: Classical Equations (speed-density) for DD1

Model Name	T=9	T=8	T=7
Greenshields	$K=36.8(1-\frac{V}{88.38})$	$36.82(1-\frac{V}{87.57})$	$37.23(1-\frac{V}{87.28})$
Greenberg	$K=45.73e^{(-\frac{V}{36.46})}$	$47.65e^{(-\frac{V}{35.41})}$	$48.02e^{(-\frac{V}{34.81})}$
Underwood	$K=15.67\ln\frac{114.24}{V}$	$15.88\ln\frac{112.11}{V}$	$15.88\ln\frac{111.7}{V}$
Bell Shape	$K=20.04(2\ln\frac{69.62}{V})^{0.5}$	$20.38(2\ln\frac{68.32}{V})^{0.5}$	$19.94(2\ln\frac{69.96}{V})^{0.5}$
	T=6	**T=5**	**T=4**
Greenshields	$K=38.45(1-\frac{V}{85.78})$	$38.65(1-\frac{V}{84.39})$	$37.45(1-\frac{V}{85.26})$
Greenberg	$K=48.86e^{(-\frac{V}{34.38})}$	$48.44e^{(-\frac{V}{33.69})}$	$45.91e^{(-\frac{V}{34.19})}$
Underwood	$K=15.76\ln\frac{113.38}{V}$	$15.78\ln\frac{111.02}{V}$	$15.53\ln\frac{109.13}{V}$
Bell Shape	$K=20.25(2\ln\frac{69.98}{V})^{0.5}$	$20.37(2\ln\frac{68.24}{V})^{0.5}$	$19.76(2\ln\frac{68.3}{V})^{0.5}$
	T=3	**T=2**	**T=1**
Greenshields	$K=38.52(1-\frac{V}{83.87})$	$38.24(1-\frac{V}{83.03})$	$38.92(1-\frac{V}{82.26})$
Greenberg	$K=49.17e^{(-\frac{V}{32.57})}$	$46.86e^{(-\frac{V}{33.39})}$	$47.99e^{(-\frac{V}{32.64})}$
Underwood	$K=15.97\ln\frac{107.06}{V}$	$15.53\ln\frac{107.26}{V}$	$16.21\ln\frac{103.85}{V}$
Bell Shape	$K=19.97(2\ln\frac{68.9}{V})^{0.5}$	$20.48(2\ln\frac{64.78}{V})^{0.5}$	$20.71(2\ln\frac{65.64}{V})^{0.5}$

In table 4.10, Greenshields model V_f values (82.257 to 88.56), Underwood V_f values (103.85 to 114.98) and Bell Shape V_f values (64.77 to 70.21) are comparatively lower than UD1. Greenshields K_j values (36.17 to 38.915) and Greenberg K_j values (45.65 to 49.17) are slightly higher than UD1. These two differences happen due to more regression accuracy compared to UD1. V_0 (32.57 to 36.08) of Greenberg are slightly higher than normal data (V_0=32.34) and UD1.

Table 4.11: Classical Equations (speed-density) for ND

Model Name	T=9	T=8	T=7
Greenshields	$K=25.24(1-\frac{V}{109.8})$	$26.18(1-\frac{V}{105.35})$	$26.74(1-\frac{V}{104.66})$
Greenberg	$K=40.34e^{(-\frac{V}{39.05})}$	$41.53e^{(-\frac{V}{37.48})}$	$42.11e^{(-\frac{V}{37.21})}$
Underwood	$K=13.43\ln\frac{128.44}{V}$	$13.96\ln\frac{122.59}{V}$	$13.47\ln\frac{128.72}{V}$
Bell Shape	$K=18.31(2\ln\frac{71.65}{V})^{0.5}$	$18.76(2\ln\frac{70.24}{V})^{0.5}$	$18.77(2\ln\frac{71.16}{V})^{0.5}$
	T=6	**T=5**	**T=4**
Greenshields	$K=25.82(1-\frac{V}{106.2})$	$27.21(1-\frac{V}{100.74})$	$27.69(1-\frac{V}{96.95})$
Greenberg	$K=40.53e^{(-\frac{V}{37.56})}$	$40.96e^{(-\frac{V}{34.49})}$	$43.24e^{(-\frac{V}{34.49})}$
Underwood	$K=13.92\ln\frac{120.41}{V}$	$15.03\ln\frac{107.45}{V}$	$15.03\ln\frac{107.45}{V}$
Bell Shape	$K=18.45(2\ln\frac{70.37}{V})^{0.5}$	$19.93(2\ln\frac{64.36}{V})^{0.5}$	$19.93(2\ln\frac{64.36}{V})^{0.5}$
	T=3	**T=2**	**T=1**
Greenshields	$K=27.5(1-\frac{V}{96.21})$	$26.72(1-\frac{V}{100.01})$	$27.27(1-\frac{V}{98.78})$
Greenberg	$K=41.15e^{(-\frac{V}{34.85})}$	$39.84e^{(-\frac{V}{36.21})}$	$40.26e^{(-\frac{V}{35.93})}$
Underwood	$K=14.32\ln\frac{109.93}{V}$	$13.67\ln\frac{116.65}{V}$	$13.76\ln\frac{117.04}{V}$
Bell Shape	$K=19.10(2\ln\frac{65.58}{V})^{0.5}$	$18.37(2\ln\frac{68.75}{V})^{0.5}$	$18.61(2\ln\frac{69.21}{V})^{0.5}$

In table 4.11, Greenshields V_f values (96.21 to 111) are comparatively higher than UD1 and UD2, K_j values (24.62 to 27.696) are simply the opposite due to low regression accuracy compared to other detectors. Underwood V_f values (107.45 to 128.72) are lower than DD1 and suddenly drop on T=4. Then increment of V_f is continued up to T=1.

Sudden V_f drop is also observed in bell shape model on T=5. As Underwood V_f values increases or decreases, Bell Shape V_f values (64.365 to 72.77) increases or decreases. The changes in Underwood and Bell Shape V_f values happens due to Greenberg model performs better than Underwood and Bell shape model in this time period which represent better hazardous traffic condition. Greenberg K_j (39.84 to 43.237) and V_0 (34.49 to 39.05) are relatively higher than DD1.

Table 4.12: Classical Equations (speed-density) for NDB

Model Name	T=9	T=8	T=7
Greenshields	$K=30.56(1-\frac{V}{95.62})$	$30.99(1-\frac{V}{94.8})$	$31.74(1-\frac{V}{93.12})$
Greenberg	$K=41.36e^{(-\frac{V}{38.46})}$	$43.22e^{(-\frac{V}{37.09})}$	$43.25e^{(-\frac{V}{36.7})}$
Underwood	$K=13.52\ln\frac{128.78}{V}$	$14.13\ln\frac{124.41}{V}$	$13.79\ln\frac{127.93}{V}$
Bell Shape	$K=18.73(2\ln\frac{70.72}{V})^{0.5}$	$18.78(2\ln\frac{72.59}{V})^{0.5}$	$19(2\ln\frac{71.89}{V})^{0.5}$
	T=6	**T=5**	**T=4**
Greenshields	$K=31.36(1-\frac{V}{92.84})$	$31.87(1-\frac{V}{90.78})$	$32.37(1-\frac{V}{88.26})$
Greenberg	$K=42.35e^{(-\frac{V}{36.91})}$	$42.38e^{(-\frac{V}{36.27})}$	$44.11e^{(-\frac{V}{34.26})}$
Underwood	$K=14.34\ln\frac{119.82}{V}$	$14.18\ln\frac{119.48}{V}$	$15.34\ln\frac{107.91}{V}$
Bell Shape	$K=18.94(2\ln\frac{70.88}{V})^{0.5}$	$18.97(2\ln\frac{70.91}{V})^{0.5}$	$20.26(2\ln\frac{65.39}{V})^{0.5}$
	T=3	**T=2**	**T=1**
Greenshields	$K=31.71(1-\frac{V}{90})$	$31.52(1-\frac{V}{89.02})$	$31.62(1-\frac{V}{88.75})$
Greenberg	$K=42.78e^{(-\frac{V}{35})}$	$42.26e^{(-\frac{V}{35.2})}$	$42.78e^{(-\frac{V}{34.72})}$
Underwood	$K=14.49\ln\frac{114.65}{V}$	$14.69\ln\frac{110.21}{V}$	$14.17\ln\frac{115.13}{V}$
Bell Shape	$K=19.01(2\ln\frac{71.07}{V})^{0.5}$	$19.45(2\ln\frac{66.53}{V})^{0.5}$	$18.62(2\ln\frac{70.26}{V})^{0.5}$

For NDB (Table 4.12), Greenshields V_f values (88.265 to 96.284) are comparatively lower than ND and K_j values (30 to 32.37) are simply the opposite due to high regression accuracy compared to ND. Sudden drops happen in Underwood V_f value (107.91 to 128.78) at T=4 and Bell Shape V_f value (65.59 to 73.35) at T=3 due to very close regression accuracy in some cases. Greenberg K_j (41.36 to 44.114) and V_0 (34.26 to 38.45) are relatively higher than DD1.

Table 4.13: Classical Equations (speed-density) for NDA

Model Name	T=9	T=8	T=7
Greenshields	K=28.71(1- $\frac{V}{100.86}$)	29.68(1- $\frac{V}{97.3}$)	29.97(1- $\frac{V}{97.43}$)
Greenberg	K=43.16$e^{(-\frac{V}{37.81})}$	45.77$e^{(-\frac{V}{35.85})}$	46.2$e^{(-\frac{V}{35.38})}$
Underwood	K=15.24ln$\frac{115.96}{V}$	15.92ln$\frac{109.97}{V}$	15.41ln$\frac{114.13}{V}$
Bell Shape	K=19.61(2ln$\frac{69.87}{V}$)$^{0.5}$	65.98(2ln$\frac{65.98}{V}$)$^{0.5}$	19.76(2ln$\frac{69.65}{V}$)$^{0.5}$
	T=6	**T=5**	**T=4**
Greenshields	K=30.1(1- $\frac{V}{96.55}$)	31.3(1- $\frac{V}{93.35}$)	30.55(1- $\frac{V}{93.5}$)
Greenberg	K=45.92$e^{(-\frac{V}{35.16})}$	47.03$e^{(-\frac{V}{33.95})}$	45.51$e^{(-\frac{V}{33.74})}$
Underwood	K=15.75ln$\frac{110.15}{V}$	15.96ln$\frac{107.55}{V}$	15.93ln$\frac{104.93}{V}$
Bell Shape	K=19.99(2ln$\frac{67.66}{V}$)$^{0.5}$	20.83(2ln$\frac{65.01}{V}$)$^{0.5}$	20.13(2ln$\frac{66.01}{V}$)$^{0.5}$
	T=3	**T=2**	**T=1**
Greenshields	K=30.29(1- $\frac{V}{92.13}$)	30.29(1- $\frac{V}{91.16}$)	30.13(1- $\frac{V}{90.98}$)
Greenberg	K=44.3$e^{(-\frac{V}{33.83})}$	43.6$e^{(-\frac{V}{34.15})}$	42.94$e^{(-\frac{V}{34.14})}$
Underwood	K=15.52ln$\frac{104.75}{V}$	14.92ln$\frac{107.56}{V}$	14.94ln$\frac{107.86}{V}$
Bell Shape	K=19.8(2ln$\frac{65.78}{V}$)$^{0.5}$	19.58(2ln$\frac{65.79}{V}$)$^{0.5}$	19.5(2ln$\frac{67.2}{V}$)$^{0.5}$

In NDA (Table 4.13), Greenshields V_f values (90.976 to 101.56) are comparatively higher than NDB and K_j values (28.18 to 31.29) are simply the opposite due to low regression accuracy for NDB. Underwood V_f values (104.746 to 116.9) are much stable from T=5 to T=1. As Underwood V_f values increases or decreases, Bell Shape V_f values (65 to 70.872) increases or decreases. The changes in Underwood and Bell Shape V_f values happens due to Greenberg model performs better than Underwood and Bell shape model in this time period which represent better hazardous traffic condition. . Greenberg K_j values (42.94 to 47.03) and are relatively higher than NDA due to better regression accuracy and and V_0 (33.74 to 37.809) does not follow any sequence.

CHAPTER 5
CONCLUTIONS

5.0 General

This study utilized sensor data placed on Tokyo Metropolitan Expressways to collect pre-crash and normal traffic condition data and compare their characteristics from the classical two phase traffic flow theory perspective. This chapter summarizes the key findings based on the pre-set objectives.

5.1 Derive and compare classical two-phase traffic flow theories

The first objective was to derive and compare classical two-phase traffic flow relationships for pre-crash and normal traffic condition data. This will include deriving four classical traffic flow theories – Greenshields, Greenberg, Underwood and Bell-shaped models, separately for pre-crash and normal traffic data. The following sub-sections discuss the findings.

5.1.1 Greenshield’s Theory

Greenshields model (R^2=0.8397) performed well in normal condition. Observing each of the pre-crash condition, it was found that the highest regression accuracy in Greenshields model was 0.6227. So, Greenshields performance was not accurate compared to normal condition.

Free flow speeds in Greenshield’s Theory for pre-crash condition varied from 82.257 to 111. Values were comparatively lower than normal (93.95) where accuracy was relatively higher. Opposite thing was also found. So, there is a relation between free flow speed and correlation coefficient (inversely proportional). Higher free flow speeds represent hazardous traffic condition.

Jam density values in Greenshield's Theory for pre-crash condition varied from 24.62 to 38.915. Values were comparatively lower than normal (36.586) where accuracy was relatively higher (0.6227). Proportional relationship has been found between jam density and correlation coefficient. Low jam density values represent hazardous traffic condition.

5.1.2 Greenberg's Theory

Greenberg model R^2 (0.6054 to 0.7844) performed well in pre-crash condition though accuracy was lower than normal (0.8044). Some considerable changes were found in model performance through comparing with other models (section 5.1.5). Better performance of Greenberg model represents hazardous traffic condition.

V_0 of Greenberg model was higher than normal data (32.34) due to better accuracy. Greenberg jam density in pre-crash condition was always lower than normal. As performance of model was increased, jam density value was increased. Increasing value of jam density may explain hazardous traffic condition.

5.1.3 Underwood Theory

Underwood model performed well in pre-crash condition (0.6101 to 0.7852) but comparatively lower than normal (0.8562). Some considerable changes were found in model performance through comparing with other models (section 5.1.5). Free flow speed values could be lower or higher than normal (119.86). Decreasing values of free flow speed can represent the comparison between pre-crash and normal data. No considerable change was found in parameter K_0.

5.1.4 Bell-shaped Theory

Bell shape model performance (0.3964 to 0.7121) was not very accurate compared to normal data (0.7544) and did not follow any pattern or sequence. Free flow speeds (64.365 to 74.05) were comparatively lower than normal (82.43). No considerable change was found in other parameter(K_0).

5.1.5 Comparison among classical two-phase theories

Underwood and Greenberg model performed well in pre-crash condition compared to Greenshields and Bell Shape. Generally in pre-crash traffic condition, Underwood model performed better than Greenberg model. But in some cases, it was found that Greenberg accuracy was closed to Underwood or exceeded Underwood accuracy (Table 4.5, 4.6, 4.7, and 4.8). These differences represent better hazardous traffic condition as well as better pre-crash data. Greenshields and Greenberg free flow speeds were better than Bell Shape to represent better pre-crash condition. When Greenberg model performs better than Underwood and Bell shape model, variation of free flow speed in Underwood is proportional to variation of free flow speed in Bell Shape.

5.2 Longitudinal variation for time and space:

5.2.1 Greenshield's Theory

As discussed earlier, Greenshields model was not accurate enough in pre-crash condition. Greenshields regression accuracy for UD1, DD1, ND, NDB, NDA were 0.552, 0.6227, 0.2216, 0.3816, 0.3119. Greenshields performed well for downstream detectors. Comparatively much lower values were observed for ND, NDA, NDB which indicates that nearest and downstream detectors represent better hazardous traffic condition.

Analyzing free flow speed for UD1(91.112), DD1(88.56), ND(111), NDA(101.56), NDB(96.284), it was clear that free flow speeds were higher than normal (93.95) for ND, NDA, NDB due to low regression accuracy (section 5.1.1).

Highest jam density values for UD1, DD1, ND, NDB, NDA were 36.25, 38.915, 27.696, 32.37, 31.29. Jam density values were lower than normal (36.586) for ND, NDB, NDA due to low regression accuracy and higher for DD1 due to high accuracy (relationship discussed in section 5.1.1). No particular time slice was found to be significantly superior.

5.2.2 Greenberg's Theory

Greenberg model performance was much better for DD1(0.7844), ND(0.6921) and NDA(0.7327) compared to other detectors. As better accuracy in Greenberg represents hazardous traffic condition, nearest detectors and downstream detectors may explain the comparison between pre-crash and normal traffic. Particular time slice was found to be significantly superior (section 5.2.5).

V_0 values were much higher for ND(39.05), NDA(37.809) and NDB(38.45) and Jam density values were higher in nearest and downstream detectors compared to upstream detectors because Greenberg performed on these detectors compared to others. Higher V_0 value and increasing jam density indicate hazardous traffic condition which may be explained using nearest and downstream detectors data.

5.2.3 Underwood Theory

Underwood model performed better for each of the detector in pre-crash condition. Underwood free flow speed values were lower for ND, NDA compared to other which represent better pre-crash condition. Free flow speed values were much stable for NDA (T=5 to T=1). Values were suddenly dropped for NDB, ND at T=4. Free flow speed values were continuously increasing in ND (T=4 to T=1). These changes in values happened due to some uncertain behavior of Underwood model performance (section 5.2.5). Nearest detectors may represent considerable the difference between pre-crash and normal data.

5.2.4 Bell-shaped Theory

Observing each of the detectors, it was found that Bell Shape model correlation coefficient values were lower for nearest detectors (0.3964 to 0.683) compared to others (0.4955 to 0.7121). Sudden drop of free flow speeds were observed in ND (T=4). Free flow speed values were continuously increasing in ND (T=4 to T=1). These happened due to some uncertain behavior of free flow speed which generally did not follow any changing pattern (section 5.2.5).

5.2.5 Comparison among classical two-phase theories

Greenberg and Underwood coefficient values were very close for DD1 (T=6, T=3), NDB (T=3 to T=1) and NDA (T=5). Greenberg performed better compared to Underwood for ND (T=4 to T=1) and NDA (T=3 to T=1). It means that nearest and downstream detectors represent better hazardous traffic condition.

Free flow value was suddenly dropped in Underwood for ND (T=4) and NDB (T=4). Sudden free flow drop was also observed in bell shape model for ND (T=5) and NDB (T=3). Variations of free flow speed in ND and NDA were proportional. Underwood free flow speed values were much stable for NDA (T=5 to T=1). These changes happened due to Greenberg model performed better than Underwood and Bell shape model at this time period.

5.3 Variation between normal and hazardous traffic conditions

Some important conclusions have been found from the analyses. All four classical two-phase traffic flow models held true for both normal and pre-crash traffic conditions. Underwood performs better in explaining both normal and pre-crash data. Both Underwood and Greenberg perform well for pre-crash data. Greenshields model performs the poorest and Bell-shape’s performance did not follow any specific pattern.

Nearest and downstream detectors explain pre-crash data better than upstream detectors. Estimated parameters are much depended on model accuracy. Substantial Changes has been found in model accuracy and free flow speed between 5 to 1 minutes before crash time. Decreasing jam density and increasing free flow speed represents better hazardous traffic condition

5.4 Limitations and future scope

There are some limitations in this study which can be improved in future work. Different crash types are not considered in this work. Only one sampling was done from normal traffic data to verify normal traffic condition for different models. Hypothetical testing can be done to identify normal and pre-crash graphs are statistically significant or not.

To improve the results, data collection period need to increased or use two different data period together. Three-phase traffic conditions need to be checked. Important future scope is to see all variables together and model things in three dimensions.

REFERENCES

- Abdel-Aty, M., Uddin, N., Pande, A., 2005. Split models for predicting multi-vehicle crashes during high-speed and low-speed operating conditions on freeways. Transportation Research Record 1908, 51–58.
- Coifman, B., Lyddy, D., Sabardonis, A., 2000. The Berkeley Highway Laboratory-Building on the I-880 field experiment. In: Proc. IEEE ITS Council Annual Meeting, IEEE, pp. 5–10.
- Daganzo, 1994. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transp Res Part B Methodol 28(4), 269–287.
- Drake, L.S., Schofer, J.L., May, A.D., 1967. A statistical analysis of speed-density hypotheses. Highway Research Record 154, 53-87.
- Golob, T., Recker, W., 2001. Relationships among urban freeway accidents, traffic flow,weather and lighting conditions. Berkeley: California PATH program, institute of transportation studies, ISSN 1055-1417, UCB-ITS-PWP-19.
- Golob, T., Recker, W., Alvarez, V., 2004. Freeway safety as a function of traffic flow. Accident Analysis and Prevention 36, 933–946.
- Greenberg, H., 1959. An analysis of traffic flow. Operations Research 7, 79-85.
- Greenshields, B. D., J. T. Thompson, H. C. Dickinson, R. S. Swinton, 1933. The Photographic Method of Studying Traffic Behavior. Highway Research Board Proceedings, Vol. 13, pp. 382–399.
- Greenshields, B.D., 1935. A study of traffic capacity. Highway Research Board 14, 448-477.
- Hall, F.L., Allen, B.L., Gunter, M.A., 1986. Empirical analysis of freeway flow-density relationships. Transportation Research Part A: General 20(3), 197-210.
- Hossain, M., Muromachi, Y., 2011. Understanding crash mechanism and selecting appropriate interventions for real-time hazard mitigation on urban expressways. Transportation Research Record 2213, 53–62.

References

- Hossain, M., Muromachi, Y., 2013. Understanding crash mechanism on urban expressways using high-resolution traffic data. Accident Analysis and Prevention 57, 17– 29.
- Lawrence A. Klein, M. K. M., David R.P. Gibson, 2006. Traffic Detector Handbook, Third ed., Vol. 1, pp. 288.
- Lee, C., Saccomanno, F., Hellinga, B., 2003. Real-time crash prediction model for the application to crash prevention in freeway traffic. Transportation Research Record 1840, 67–77.
- Lee, C., Saccomanno, F.,Hellinga, B., 2002. Analysis of crash precursors on instrumented freeways. Presented at the 81st annual meeting of Transportation Research Board, Washington, D.C.
- Lighthill, M., Whitham, G., 1955. On kinematic waves II. A theory of traffic flow on long crowded roads. Proc Royal Society of London Part A 229 (1178), 317–345.
- Lu, Z., Meng, Q., 2013. Analysis of Traffic Speed-Density Regression Models-A Case Study of Two Roadway Traffic Flows in China. In Proceedings of the Eastern Asia Society for Transportation Studies, 9.
- Lum, K., Fan, H., Lam, S., Olszewski, P., 1998. Speed-Flow Modeling of Arterial Roads in Singapore. Journal of Transportation Engineering 124(3), 213-222.
- May, A.D., 1991. Traffic flow dundamentals. Prentice-Hall, Inc., Englewood Cliffs, N.J.
- Mills, M.K., Gibson, D., 2006. Traffic Detector Handbook – Third Edition Volume II, FHWAHRT-06-139, Federal Highway Administration, US Department of Transportation, Washington, DC.
- Mimbela, L. E. Y., Klein, L. A., 2007. A Summary of Vehicle Detection and Surveillance Technologies used in Intelligent Transportation Systems. Report prepared for Federal Highway Administration's (FHWA) Intelligent Transportation Systems Program Office. Sponsored in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

- Nielsen, O.A., Jorgensen, R.M., 2008. Estimation of speed-flow and flow-density relations on the motorway network in the greater Copenhagen region. Intelligent Transport Systems, IET 2(2), 120-131.
- Oh, C., Oh, J., Ritchie, S., 2001. Real-time estimation of freeway accident likelihood. In: Presented at 80th Annual Meeting of the Transportation Research Board, CD-ROM, Washington, D.C.
- Ohta, K., Harata, N., 1989. Properties of aggregate speed-flow relationship for road networks, Proc. 5-th World Conference on Transport Research, Yokohama, Japan, pp.451-462.
- Richards, P., 1956. Shock waves on the highway. Operations Research 4 (1), 42–51.
- Sabey, B. E., Staughton, G. C., 1975. Interacting Roles of Road Environment, Vehicle and Road User in Accidents. Paper presented at the 5th International Conference on the International Association for Accident and Traffic Medicine, London.
- Treat, J. R., Tumbas, N. S., McDonald, S .T., Shinar, D., Hume, R. D., Mayer, R. E., Stanisfer, R. L., Castillan, N. J., 1977. Tri-level Study of the Causes of Traffic Accidents. Report No. D.
- Underwood, R.T., 1961. Speed, volume, and density relationships: Quality and theory of traffic flow. Yale Bureau of Highway Traffic, 141-188.
- Velasco and Saavedra, 2008. Macroscopic Models in Traffic Flow,Qual.Th.Dyn.Syst. 7, 237–252.
- WHO, 2004. World report on road traffic injury prevention. World Health Organization
- Xie, H., Jiang, Y., Zhang, L., 2011. Empirical Study on Relationship among the Lane-Changing, Speed and Traffic Flow. Systems Engineering Procedia 2(0), 287-294.
- Xu, C., Liu, P., Wang, W., Li, Z., 2012. Evaluation of the impacts of traffic states on crash risks on freeways. Accident Analysis and Prevention 47, 162–171.

- Yu, R., Abdel-Aty, M., 2013. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Safety Science 63, 50–56.