# INVESTIGATION OF TRAFFIC FLOW CHARACTERISTICS: CASE STUDY OF D100 HIGHWAY

by

Muhammed Ali Çevik

B.S., Civil Engineering, Bogazici University, 2017

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Civil Engineering
Boğaziçi University
2019

INVESTIGATION OF TRAFFIC FLOW CHARACTERISTICS:
CASE STUDY OF D100 HIGHWAY

APPROVED BY:

Assoc. Prof. Ilgın Gökaşar .....

(Thesis Supervisor)

Assist. Prof. Gürkan Günay. .....

Assist. Prof. Mehtap Işık .....

DATE OF APPROVAL: 22.07.2019

# ACKNOWLEDGEMENTS

I want to express my deepest gratitude to my thesis supervisor Assoc. Prof. Ilgın Gökaşar, who guided and helped me through my graduate education and through my thesis. Without her guidance, suggestions, patience and encouragement, this thesis couldn't be finished.

I want to thank Assist. Prof. Mehtap Işık and Assist. Prof. Gürkan Günay for taking part in my thesis defense jury and for their valuable comments and insights.

I want to thank Alperen Timuroğulları for his valuable assistance in the image processing stage and for his help with the data preparation.

I also want to thank Department of Traffic in Istanbul Metropolitan Municipality for providing the video footages used in this thesis.

# ABSTRACT

## INVESTIGATION OF TRAFFIC FLOW CHARACTERISTICS: CASE STUDY OF D100 HIGHWAY

Bridges located on the Bosphorus create some of the most problematic locations in Istanbul traffic network where congestion is a big problem. In this thesis, using the video footages provided by the Department of Traffic in Istanbul Metropolitan Municipality the traffic characteristics on a short segment of road located on the D100 Highway upstream of the 15 July Martyrs Bridge and just upstream of the Metrobus on-ramp and Yildiz on-ramp is investigated. The video footages are from 10.10.2018 to 16.10.2018 and a total of approximately 12 hours of data is contained in the final data. The road is divided into regions, and the speed and flow values inside these regions are calculated after the image processing stage. It is found that speed and flow rate on the left lane is generally the highest while speed and flow rate on the right lane is generally the lowest in uncongested traffic conditions while they balance each other when traffic congestion is introduced. Regression models are created regarding the traffic parameters inside these regions using IBM SPSS. With these regression models, it is found that left lane and middle lane have similar flow and speed characteristics and it is also found that speed on a lane is positively correlated with speed on other lanes.

# ÖZET

## TRAFİK KARAKTERİSTİKLERİNİN İNCELENMESİ: D100 KARAYOLU VAKA ÇALIŞMASI

İstanbul Boğazı'nın üstündeki köprüler, İstanbul gibi trafik sıkışıklıklarının sık yaşandığı bir şehirdeki en problemli noktalardan bazıları. Bu tezde, D100 Karayolunda bulunan, 15 Temmuz Şehitler Köprüsünden önce ve Yıldız katılımından hemen önce yer alan kısa bir yol segmenti, İstanbul Büyükşehir Belediyesi Trafik Müdürlüğü'nün sağladığı kamera görüntüleri kullanılarak, trafik karakteristikleri açısından inclelenmiştir. Video kamera görüntülerinde 10 Ekim'den 16 Ekim'e toplam 7 günden yaklaşık olarak 12 saatlik veri yer almakta. Verinin oluşturulması işlemi için, yol bölgelere ayrıştırıldı ve bu bölgelerdeki görüntü işleme algoritmasından sonra elde edilen hız ve akım değerleri incelendi. Akıcı trafikte sol şeritte genel olarak en yüksek akım ve hız verileri gözlenmişken, sağ şeritte genel olarak en düşük akım ve hız verileri gözlendi. Ancak yoğun trafik koşullarında şeritler arasındaki akım ve hız verileri arasındaki farkın kapandığı görüldü. Trafik parametreleri kullanılarak IBM SPSS aracılığıyla regresyon modelleri oluşturuldu. Bu regresyon modellerinde sol ve orta şeritin benzer hız ve akım karakteristiklerine sahip olduğu ve bir şeritteki hızın diğer şeritlerdeki hızlarla doğru orantılı olduğu bulundu.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... iii

ABSTRACT ..... iv

ÖZET ..... v

LIST OF FIGURES ..... viii

LIST OF TABLES ..... xii

1. INTRODUCTION ..... 1

1.1. Problem Statement ..... 1

1.2. Goals and Objectives ..... 2

1.3. Thesis Outline ..... 2

2. LITERATURE REVIEW ..... 3

2.1. Traffic Characteristics of Freeways ..... 4

2.2. Traffic Detection Systems ..... 7

2.2.1. Detector Types ..... 8

2.2.1.1. Pavement Invasive Detectors ..... 8

2.2.1.2. Non-Pavement Invasive Detectors ..... 8

2.2.2. Traffic Studies Done with Video Image Processing ..... 10

3. THEORY ..... 12

3.1. Traffic Flow Theory ..... 12

3.1.1. Speed ..... 12

3.1.1.1. Microscopic Speed ..... 13

3.1.1.2. Macroscopic Speed ..... 13

3.1.2. Flow ..... 15

3.1.3. Density ..... 17

3.2. Traffic Stream Models ..... 18

3.3. Regression ..... 25
3.3.1. Simple Linear Regression ..... 25
3.3.2 Multiple Linear Regression ..... 30
3.3.3 Factor Analysis ..... 32
4. METHODOLOGY ..... 34
4.1. Study Area Description ..... 34
4.2. Data Preparation ..... 41
4.3. Data Analysis ..... 46
4.3.1. Speed and Flow Calculations ..... 46
4.3.2. Regression Analysis ..... 52
4.3.3. Complications Regarding Obtaining Sound Data ..... 59
4.4. Results ..... 59
4.4.1. Speed-Flow Results ..... 59
4.4.2. Heat Maps of Regions ..... 68
4.4.2.1. Flow Heat Maps ..... 68
4.4.2.2. Speed Heat Maps ..... 72
4.4.3. Regression Results ..... 75
4.4.3.1. Regression Model for the Date 12.10.2018 ..... 76
4.4.3.2. Regression Model for the Date 16.10.2018 ..... 77
5. CONCLUSIONS ..... 80
REFERENCES ..... 83

# LIST OF FIGURES

Figure 2.1. Flowchart of the Literature Review.....4

Figure 3.1. A generalized speed flow curve proposed by Hall et al. .....17

Figure 3.2. Speed-density diagram. .....19

Figure 3.3. Speed-flow diagram. .....20

Figure 3.4. Flow-density diagram.....21

Figure 3.5a. (Speed vs density) 3.5b (Speed vs flow) 3.5c (Flow vs density) Comparison of Greenshields' models (in blue) and sample empirical data. .....22

Figure 3.6. Fundamental diagrams in the case of triangular speed-flow relationship. .....25

Figure 3.7. Simple linear relationship between y and x.....26

Figure 3.8. Sample simple regression with error terms. .....27

Figure 4.1. Location of the study area (Yandex Maps). .....34

Figure 4.2. The satellite image of the study area (Yandex Maps). .....35

Figure 4.3. A sketch of the study area. .....35

Figure 4.4. Still image from the video footages. .....37

Figure 4.5. Study area with the original regions. .....37

Figure 4.6. Regions on the main arterial with information regarding their lengths. .....38

Figure 4.7. A sample image from vehicle tracking mechanism. ..... 41

Figure 4.8. Pixels to meters conversion sketch. ..... 44

Figure 4.9. Factor analysis setup (SPSS). ..... 52

Figure 4.10. Factor analysis-extraction (SPSS). ..... 53

Figure 4.11. Factor analysis-rotation (SPSS) ..... 53

Figure 4.12. KMO and Bartlett's test results (SPSS). ..... 53

Figure 4.13. Explained total variances of the components (SPSS). ..... 54

Figure 4.14. Scree plot of the components (SPSS). ..... 55

Figure 4.15. Rotated component matrix (SPSS). ..... 55

Figure 4.16. Setup of the linear regression (SPSS) ..... 56

Figure 4.17. Model summary of the initial model (SPSS) ..... 57

Figure 4.18. Anova of the initial model (SPSS). ..... 57

Figure 4.19. Coefficients of the initial model (SPSS). ..... 57

Figure 4.20. Model summary of the final model (SPSS) ..... 58

Figure 4.21. Anova of the final model (SPSS). ..... 58

Figure 4.22. Coefficients of the final model (SPSS). ..... 58

Figure 4.23. Speed-flow graph of region 4 (16.10.2018). ..... 61

Figure 4.24. Speed-flow graph of region 4 (10.10.2018). ..... 61

Figure 4.25. Speed-flow graph of region 5 (16.10.2018). ..... 62

Figure 4.26. Speed-Flow graph of region 5 (10.10.2018). ..... 63

Figure 4.27. Speed-flow graph of region 3 (16.10.2018). ..... 64

Figure 4.28. Speed-flow graph of region 3(10.10.2018). ..... 65

Figure 4.29. Comparison of flow rates of regions 3,4 and 5 (16.10.2018). ..... 66

Figure 4.30. Comparison of speeds of regions 3,4 and 5 (16.10.2018). ..... 66

Figure 4.31. Comparison of flow rates of regions 3,4 and 5 (10.10.2018). ..... 67

Figure 4.32. Comparison of speeds of regions 3, 4 and 5 (10.10.2018). ..... 68

Figure 4.33. Flow heat map of the right lane – 15.10.2018. ..... 69

Figure 4.34. Flow heat map of the middle lane – 16.10.2018. ..... 70

Figure 4.35. Flow heat map of the left lane – 12.10.2018. ..... 71

Figure 4.36. Speed heat map of the right lane – 10.10.2018. ..... 72

Figure 4.37. Speed heat map of the middle lane – 13.10.2018. ..... 74

Figure 4.38. Speed heat map of the left lane – 16.10.2018. ..... 75

Figure 4.39. Correlations between principal components-12.10.2018 (SPSS). ..... 76

Figure 4.40. Summary of the model-12.10.2018 (SPSS). ..... 76

Figure 4.41. Anova of the model-12.10.2018 (SPSS). ..... 77

Figure 4.42. Coefficients of the model-12.10.2018 (SPSS). ..... 77

Figure 4.43. Correlations between principal components-16.10.2018 (SPSS). ..... 78

Figure 4.44. Summary of the model-16.10.2018 (SPSS). ..... 79

Figure 4.45. Anova of the model-16.10.2018 (SPSS). ..... 79

Figure 4.46. Coefficients of the model-16.10.2018 (SPSS). ..... 79

# LIST OF TABLES

Table 3.1. A summary of single regime speed-density models. ..... 23

Table 3.2. A summary of multi regime speed-density models. ..... 24

Table 4.1. Region coordinates. ..... 39

Table 4.2. Node coordinates. ..... 40

Table 4.3. Sample output obtained from image processing. ..... 42

Table 4.4. Dates and times of video footages. ..... 45

Table 4.5. Dates and times of successfully processed video footages. ..... 45

Table 4.6. The time intervals that are excluded from final analysis. ..... 45

Table 4.7. Sample data with pixels converted to distance. ..... 47

Table 4.8. Speed-flow data for 11.10.2018 (11:30-12:00). ..... 50

Table 4.9. Components to be used in the regression model. ..... 56

Table 4.10. Some traffic parameters found on region 3, region 4 and region 5 according to Greenshields equation. ..... 65

# 1. INTRODUCTION

Highways have a very important role in sustainability of a transportation network. They connect communities with each other whether it is across a city or across a country. They have very important implications on the economy of a city. More advanced and well maintained highway systems mean more developed communities. An urban freeway is a type of highway. They are also called express highways. Urban freeways are the most vital parts of a city's road network especially in a city like Istanbul, where the study area is located.

Traffic congestion is an issue associated with urban freeways all around the world, especially for metropolitans. Some negative aspects of traffic congestion are increased amount of fuel consumption, higher rate of traffic accidents, impact on psychological state of the drivers and deterioration in road lifetime [1]. Istanbul, being one of the most crowded cities in the world, has congestion prone urban freeways with lots of problematic locations such as bridges located on the Bosphorus.

For this thesis, an important urban freeway segment which is located upstream of the 15 July Martyrs Bridge on the European side is analyzed in detail regarding traffic flow characteristics, especially speed and flow rate. There are also two on-ramps connecting to the main freeway segment downstream which makes the area even more prone to congestion problems. Video camera footages overseeing the entire region are used for the analysis of traffic flow with the help of image processing algorithms. Then, the obtained data is analyzed on Microsoft Excel to provide the final results. With the findings of this study, it will be possible to have an idea about the traffic flow characteristics in this area.

## 1.1. Problem Statement

Urban freeways play a crucial role in transportation network of a city. Entering and exiting a freeway is established via on-ramps and off-ramps. The study site in question is a part of O-1 freeway upstream of the 15 July Martyrs Bridge and there are two on-ramps connecting to the main arterial, one for regular vehicles and one for metrobuses. The

merging point is an important location as it is a potential bottleneck for the traffic. So, it is of importance to analyze the traffic flow characteristics on the main arterial along with the effects of on-ramps to have a general idea about the traffic flow patterns observed in this area.

## 1.2. Goals & Objectives

The freeway segment that is focused on in this study is one of the problematic locations in Istanbul road network in terms of congestion. The main goal of this thesis is to have a detailed idea about the traffic flow characteristics in this area. To achieve this goal, the following objectives are aimed:

(i) Examining speed and flow trends on the regions on the main arterial.
(ii) Investigating the effect of on-ramps on the upstream freeway main arterial.
(iii) Using regression models to further analyze the parameters of traffic flow.

## 1.3. Thesis Outline

The rest of the thesis is organized as the following: Firstly, in the Literature Review section, the studies about traffic studies on freeways are reviewed and traffic detection methods and traffic studies about video camera detection is touched upon. Then, in the Theory section, theoretical knowledge about traffic flow theory, traffic stream models and regression analysis are introduced. In the Methodology section, the steps taken to get to the results are explained and these results are interpreted. The thesis is finalized with the conclusions section where the general findings and future recommendations acquired from this study are provided.

# 2. LITERATURE REVIEW

Transportation is an essential part in everyone's lives as everyone needs means of travel in order to carry out fundamental tasks such as working and studying, or for other reasons such as shopping. Since a big chunk of trips is done via roadways, investigation and analysis of traffic characteristics on roadways is a very important task that needs to be done. Understanding the relationships between traffic parameters allows responsible parties to better design and maintain roadway facilities which results in higher quality of traffic operations, improved roadway performances and higher quality of life for commuters.

The roadway section that is investigated in this thesis is from an urban freeway. Traffic characteristics on urban freeways are of utmost importance as urban freeways are vital parts of transportation networks of cities. The first part of this literature review focuses on the traffic studies done on freeways.

The traffic characteristics on road networks can be obtained with the help of traffic detectors. There are many detector types available. The detection type that is used for acquiring the data of this thesis is video image processing with the footages from the camera overseeing the road. The second part of this literature review focuses on the general types of detectors and the traffic studies that are done with video image processing algorithms.

The general structure of this literature review is given on the flowchart which is presented in Figure 2.1.

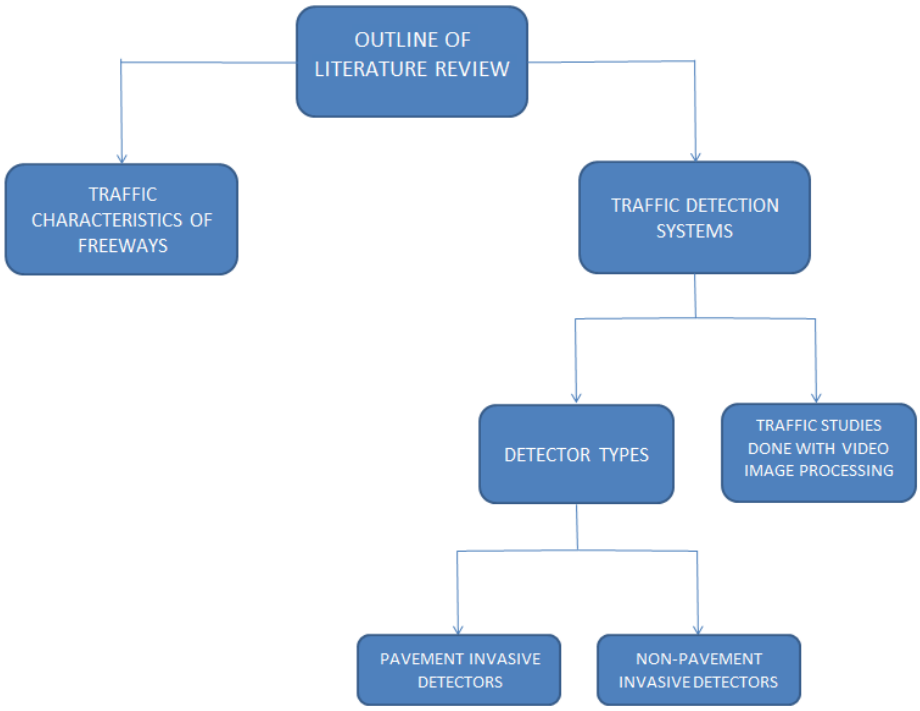


Figure 2.1. Flowchart of the Literature Review.

## 2.1. Traffic Characteristics of Freeways

The first comprehensive traffic study on a highway was conducted by Greenshields in 1930's. Greenshields was the first engineer to use visuals obtained from a camera to measure flow, density and speed of a highway. With this study, Greenshields was able to postulate a relationship between speed and density and thus the first traffic stream model. The details of this model are mentioned in the theory section of this thesis.

Until the 1950's, traffic engineering as a discipline was more about finding solutions to problems that are encountered in traffic. But with studies done by mathematicians and engineers such as Wardrop [2], it has evolved into a field with more scientific nature. With the introduction of computers in the 1960's, the field evolved further. With these advancements in the field of traffic engineering and in technology, more and more empirical studies are conducted on road segments in the world for number of reasons such as gathering information about the roadways in question and testing the adequacy of the scientific model that is examined. In the rest of this section studies that are done on certain aspects of traffic flow on freeways are investigated.

Hall *et al.* [3] used the data obtained from Queen Elizabeth Way in Ontario, Canada to conclude that the relationship between flow and density is a continuous one, rather than a discontinuous one and it has a shape of a reverse lambda. In order to uncover the true nature of traffic flow, only data from days with ideal conditions, meaning clear weather and no incidents, were used in the final analysis. Banks [4] used data from loop detectors on a freeway segment in San Diego to analyze the traffic in terms of occupancy (density), flow rate and speed. Some findings are that the average speed drops very little until peak flow is reached and the relationship between the flow and the occupancy (density) is linear up to peak flow and inversely linear after peak flow just like Hall *et al.* [3] found. Chin and May [5] studied the speed-flow relationship on Caldecott Tunnel which is located in California. The results are similar to the ones that Banks had found [4]. It is found that speed is not dependent on flow when the flow rate is small (up to 800 vehicles per hour (vph)). Then, there is a slight drop of speed of 0.7 mph for every 100 vph increment of increase which is a little higher than what Banks had found.

Traffic characteristics might differ from a country to another country due to some factors such as the difference of behavior between the drivers, the difference of climate and the general difference in topography. Highway Capacity Manual (HCM), which is accepted as the guide to traffic analysis and control by many countries in the world including Turkey, is developed for United States. Andrade and Setti [6] obtained speed flow curves from the data from 24 traffic-counting stations in Brazil where HCM is used. The main goal of this study is to come up with definitive speed flow curves representative of Brazil and to compare them with HCM 2010. It is found that compared to HCM 2010, density is lower at capacity and the speed is higher at capacity.

The main traffic flow characteristics on a freeway are different among lanes. Hurdle *et al.* [7] studied on the flow-speed relationship for individual lanes on two freeway sections in Toronto. One finding from this study is that the average speed is higher on the left lane even in low flow conditions because faster drivers prefer to use the left lane when the slower drivers are using the other lanes. One other interesting finding is that the average speed on a freeway as a whole increases until the flow rate is increased up to a certain level and then it starts to decrease with increased flow while individual lane speeds decrease the whole time. May [8] and Busch [9] conducted separate studies about lane

volume distribution with altering flow rates in 2 three-lane freeways United States and Germany respectively. May [10] then combined the results of these two studies to compare them in more detail. In both of the studies, with increased total directional flow, the percentage of drivers that use the left lane increases whereas the percentage in the right lane decreases. The left lane percentages were higher in Germany because of German regulations that prohibits passing from the right, the higher difference in speed between cars and trucks and the larger space between on-ramps and off-ramps.

Sasahara *et al.* [11] argued that it is healthier to investigate the conditions of traffic on a lane by lane basis. It is concluded in this study that the methodology presented in the Highway Capacity Manual, which takes average speed of all the lanes in operation, is not adequate in explaining the flow of traffic in certain conditions such as a lane being blocked. It is found in this study that the flow rate distribution is uneven between lanes and it is dependent on a number of factors such as the volume/capacity ratio. The flow rate percentage in the right lane drops if an on-ramp or an off-ramp is present downstream which is the case for the study area investigated in this thesis. The speed-flow curves found in HCM are still useful to forecast individual lane performance if they are adjusted with lane by lane field data.

Daganzo [12] argued that drivers can simply be classified into two categories which are rabbits (aggressive drivers) and slugs (regular drivers). Rabbits always prefer the lane with the highest speed aka the left lane and their primary ambition is to maximize their speed. The outside lanes are composed of a combination of both slugs and rabbits where rabbits try to reach to the left lane if possible. However, this is only assumed for uncongested conditions. For congested traffic conditions, rabbits and slugs are distributed randomly. Banks and Amin [13] tested Daganzo's theory with data obtained from several sites. It is argued in this study that Daganzo's logic is oversimplified by showing that the aggressive drivers are not only keen on choosing the lane with the fastest speed. It is concluded that a better alternative theory to Daganzo's theory is that rabbits choose lanes not only with relative speed of lanes but also with relative densities between lanes as well.

Knoop *et al.* [14] obtained traffic data from a motorway segment in Netherlands and found that the distribution of flow is lower on the outside lanes and higher on the left lanes

on freeway segments upstream of on-ramps compared to regular freeway segments for densities between 10 veh/km and 130 veh/km. This can be explained by the fact that the drivers avoid using the rightmost lane to avoid the merging vehicles coming from the on-ramp. Also it is found that as speed limit on a roadway is increased, the fraction of vehicles using the outside lane decreases. Duret *et al.* [15] found that the density of the left lane increases more than the other lanes with increased flow rates and it is also found that a driving ban on trucks increases the flow rates observed on the right lane significantly.

External conditions such as weather have an effect on the characteristics of freeway flow. The impact of rain is negative as shown by past research. The research done by Wang and Luo [16] is such an example. In this study, rain is classified into light, moderate and heavy rain and it is found that maximum flow rate of the freeway examined dropped by 15.7%, 19.1% and 32.5% respectively and free flow speed is decreased by 4.4%, 7.3% and 10.6% respectively with respect to clear weather conditions on the same freeway. Akin *et al.* [17] has studied on the impact of weather on traffic flow characteristics on the freeways of Istanbul. In this study, it is found that the average speed of vehicles dropped by 8-12% and the capacity dropped by 7-8% with rainy weather and the volume of traffic dropped by 65% with light snow.

In this section, some studies which focus on traffic flow characteristics are touched upon. The theory section of this thesis contains the basic theoretical background of these studies and is useful in understanding more about them.

## 2.2. Traffic Detection Systems

Detectors are essential in obtaining traffic data from a roadway. The data is then used in many traffic studies and applications. Therefore, using a proper and reliable detector is very important in obtaining sound results. According to the Federal Highway Administration (FHWA), traffic detectors are divided into two in general which are: pavement invasive detectors and non-pavement invasive detectors [18]. In the first part, the types of detectors for each group are briefly explained. In the second part, some studies that focus on obtaining traffic data with image processing algorithms, which is the type of detection used in this thesis, are mentioned.

### 2.2.1. Detector Types

2.2.1.1. Pavement Invasive Detectors. Detectors that are located inside the pavement are called pavement invasive detectors. They give accurate results and are not affected by weather effects but they are susceptible to stresses caused by the traversing vehicles which require them to undergo maintenances on a regular basis. Improper installation of pavement invasive detectors decreases the lifespan of the pavement so they need to be installed with caution. The main types of pavement invasive detectors are inductive loop detectors, magnetometers and magnetic detectors.

- Inductive Loop Detectors: It is the traffic detector type that is used the most around the world. It is composed of an electrically conducting loop, which is insulated and it is installed in the pavement. It can come in many sizes and shapes depending on the area of detection, vehicle types that are wanted and the objective of detection (speed, counting etc.). It has the best accuracy compared with the other frequently used detector types. However this accuracy might decrease if a large variety of vehicles are wanted to be detected [18].
- Magnetometer: It uses earth's magnetic field to measure changes in vertical and horizontal dimensions. With this, magnetometers can detect passing vehicles and stationary vehicles above. They are useful on bridge decks, viaducts and temporary installations in construction zones. They are not as widely used as inductive loop detectors. One advantage that magnetometers have against loop detectors is that they are less susceptible to stresses [18].
- Magnetic Detectors: They measure the changes in the lines of flux of earth's magnetic field similar to magnetometers. They are limited by the fact that they cannot detect vehicles below a certain minimum speed and this means that they shouldn't be used as presence detectors as they might not detect stationary vehicles. They can detect speeds of vehicles directly if the speeds are above this threshold. [18].

2.2.1.2. Non-Pavement Invasive Detectors. Non-pavement invasive detectors are traffic detectors that are not located inside the pavement and don't require the pavement to be cut, unlike pavement invasive detectors. All of them give sound results in clear weather

conditions but performances of some of them may degrade with rainy, foggy or snowy weather. The main types of non-pavement detectors are microwave radar, active infrared, passive infrared, ultrasonic, acoustic and video image processors.

- Microwave Radar Detectors: Uses microwave energy to detect flow and speed. Some versions that include frequency modulated continuous waves can work as a presence detector too. They are also not as sensitive to external conditions such as weather compared to other non-pavement invasive detectors [18].
- Active Infrared Detectors: They function by transmitting infrared energy to the vehicles. This energy is then reflected back to the detector, which allows the detector to obtain traffic data. Their performances reduce considerably if there is a thick fog or blowing snow in the vicinity [18].
- Passive Infrared Detectors: Unlike active infrared detectors, they don't transmit any energy. They rather function by detecting the energy originating from vehicles, road and other objects. They can detect speed well. Just like active infrared detectors, their performances reduce with external weather effects such as rain, snow and thick fog [18].
- Ultrasonic Detectors: They use ultrasonic sound energy waves to operate. They can count vehicles, detect presence and determine occupancy. They are affected negatively by environmental circumstances such as air turbulence and change in temperature [18].
- Acoustic Detectors: They detect sounds produced by the vehicles in the traffic to determine speed and presence. They are affected negatively from cold weather and prominent background noise [18].
- Video Image Processors: Cameras are used to record the traffic. These recordings are then used to obtain traffic data by processing and converting. They can be used to obtain traffic data from an area instead of other detectors which only gather data from a point. Thus a single camera can replace multiple loop detectors if it is set up well. If multiple cameras are used in a synchronized manner, the data that can be obtained gets even richer. However their performances degrade significantly with weather conditions such as fog, snow and rain. Other visual elements such as shadows, occlusion, night-time conditions and water on camera lens are also problems that are encountered [18].

### 2.2.2. Traffic Studies Performed with Video Image Processing

Video camera footages are very useful in a field like traffic engineering where most studies are based on observation. Despite having the limitations mentioned in the previous section, it is still one of the most effective ways of analyzing traffic flow. It is becoming increasingly popular since the beginning of the 21st century. With new technological advancements, it will be more feasible to use cameras to monitor and analyze traffic flow. There are numerous studies and research done on traffic flow analysis using video processing algorithms. Zhang *et al.* [19] proposed a method of video based measurements to calculate traffic flow data. The validity of the method has been tested on 3 different test sites, first site having ideal conditions, second site having issues with shadows and the third site having issues with weather and lighting conditions. All the results were satisfactory with site 3 performing the worst. It is acknowledged that further improvements are necessary to address visual issues like light reflections and longitudinal occlusions.

Ai *et al.* [20] used video trajectory data from an interstate freeway in Columbus, Ohio to study time headway characteristics and speed flow diagrams under both congested and uncongested traffic conditions. A video footage of around 100 minutes which includes both congested and uncongested traffic conditions was used in the analysis. In the study, headway is divided into 4 types which are car-car, car-truck, truck-car and truck-truck where the first one is the leading vehicle and the second one is the following vehicle. Under uncongested traffic conditions, truck-car headways turned out smaller than car-truck headways which means that headways get smaller when trucks are in front. Under congested traffic conditions the headway is found to be bigger in mixed traffic than all-car traffic. It is acknowledged that the study period needed to be longer in order to come up with better and sounder results.

Cho *et al.* [21] proposed an algorithm to determine velocity fields from video footages with poor resolution and demonstrated the integrity of the algorithm with an empirical study done on a freeway segment of 1 mile length that is located in Berkeley, California. A detecting and tracking algorithm was not possible for the low resolution footages that are obtained. Because of this, another method, namely intensity profiles is used to calculate speeds of vehicles. The results were very similar to the loop detector data

obtained from the same location. However it is not possible to obtain flow and density of traffic with this algorithm.

Park *et al.* [22] proposed a video based framework which can be used to obtain the flow and speed characteristics of a roadway. The framework was tested on the footages taken from a highway in Atlanta. Speeds were calculated with the help of the coordinates of vehicles and their corresponding timestamps. The results from the video footages are compared with the GPS measurements and it is found that the two results are fairly similar with little amount of error. However, just like most other video based applications, this method suffers when the visibility is limited e.g. night time, poor weather conditions.

Guo *et al.* [23] used video camera footages for sections on three freeways in China to demonstrate the impact that lane changing has on speed and headway of vehicles. Lane changing had a big impact on flow characteristics in the target lane as it decreases the time headway between the lane changing vehicle and the following vehicle on the target lane. The most driving factors to change lanes in a traffic stream were the speed difference between a following and leading vehicle, the space between vehicles, varying traffic conditions between lanes and the combined effect of all three. However it is acknowledged that aggressive drivers change their lanes regardless of these factors. It is also found that lane changing rates increase when there is an off-ramp (attraction) or on-ramp (repulsion) downstream.

Zheng *et al.* [24] used video cameras to obtain speed, flow and headway data from two road sections from urban expressways in Shanghai and Zhengzhou. In this study, a wide variety of environmental conditions are included, such as snowy days. It is concluded that traffic flow is a non-linear phenomenon and it has many components like location and climate. It is found in this study that for short average space headways, the mean velocity on Shanghai Yan'an Viaduct is 25% lower on sunny days compared to snowy days but for long average headways it is 36% higher on sunny days compared to snowy days. In the study, drivers are classified as aggressive and careful drivers where aggressive drivers have the traits of high speed and short headway. The number of aggressive drivers is seemed to be higher with bigger flow rate but no definitive conclusion is made in this study.

# 3. THEORY

## 3.1. Traffic Flow Theory

For this thesis, the main theoretical consideration is the traffic flow theory, particularly the speed and flow characteristics of the road section and the relationship between these two parameters. Speed, flow and density are the main parameters of traffic stream and they have been a research area in traffic engineering since early 20th century. The relationship between these parameters, mostly speed and density, are first modeled by Greenshields [25]. Since then, a lot of modifications and adjustments have been made on the relationships between these parameters.

Speed, flow and density are directly related to each other and form the fundamental equation of traffic flow which is given in Equation 3.1. With two of these at hand, the third parameter can easily be calculated.

$$q = k * u \tag{3.1}$$

where

q= flow rate (vehicles/hour),
k = density (vehicles/kilometers),
u = speed (kilometers/hour).

### 3.1.1. Speed

Speed is by definition the distance travelled by an object, which in our case a vehicle, in unit time. Speed is an important tool to assess the performance of a roadway. Speed is essential in fuel consumption and emission studies and it is needed to determine the level of service of a roadway.

Speed can be investigated in terms of both microscopic and macroscopic means. Microscopic speed characteristics include the analysis of speeds of individual vehicles whereas macroscopic speed characteristics are about the speed trends of all the vehicles combined, rather than focusing on individual vehicles.

3.1.1.1. Microscopic Speed. Microscopic speed can be defined as the rate of movement of a single vehicle, in units of distance over time. It can be represented as:

$$u_i = \frac{dx}{dt} \tag{3.2}$$

where

$u_i$ = Speed of the $i^{th}$ vehicle (meters/second),
$d_x$ = the distance travelled by vehicle I (meters),
$d_t$ = the length of time (seconds).

Microscopic speed characteristics are examined in terms of time mean speed. Time mean speed is the arithmetic average of all the vehicles passing a certain section of a road. Time mean speed can be calculated by the following equation:

$$\bar{u}_t = \frac{1}{N}\sum_{i=1}^{N} u_i \tag{3.3}$$

where

$\bar{u}_t$ = Time mean speed (meters/second),
$u_i$ = Speed of the $i^{th}$ vehicle (meters/second),
N = total number of vehicles.

3.1.1.2. Macroscopic Speed. Macroscopic speed, i.e., space mean speed, is another way of calculating the speed average on a roadway. It differs from time mean speed in the sense that calculating space mean speed doesn't involve taking the average of speeds, but it

rather involves taking the average time a vehicle spends on a particular length of a roadway. Space mean speed can be calculated by the equation:

$$\bar{u}_s = \frac{D}{\frac{1}{N}\sum_i t_i} \tag{3.4}$$

where

$\bar{u}_s$ = Space mean speed (meters/second),
D= The distance that vehicles are investigated for (meters),
$t_i$ = The time it takes for vehicle i to travel distance *D* (seconds).

Microscopic speed parameter time mean speed and macroscopic speed parameter space mean speed are closely related to each other. Equation 3.5 summarizes this relationship. As it can be understood from this equation, time mean speed is always higher than space mean speed.

$$\bar{u}_t = \bar{u}_s + \frac{S_s^2}{\bar{u}_s} \tag{3.5}$$

where

$S_s^2$= Variance of space mean speed,
$\bar{u}_t$ and $\bar{u}_s$ are previously defined.

Space mean speed is seen more accurate compared to time mean speed by most engineers. Knoop *et al.* [26] used data from individual cars on a motorway stretch to compare space mean and time mean speed and found that space mean speed is a better fit for a fundamental diagram of traffic flow.

Free flow speed is a widely used parameter in traffic applications. It is defined as the average speed that vehicles would travel with if the density and flow rate is low and all external conditions, such as weather, are ideal.

Optimum speed is the speed of traffic when the flow is at capacity level.

### 3.1.2. Flow

Flow rate is defined as the number of cars passing a point in a specified unit time interval such as an hour. It can be found by:

$$q = \frac{N}{T} \tag{3.6}$$

where:

N=Number of vehicles counted,
T= Vehicle counting duration.

Flow rate can also be represented as the reciprocal of mean time headway by equating the total time to the sum of time headways between all the vehicles.

$$q = \frac{N}{T} = \frac{N}{\sum_{i=1}^{N} h_i} = \frac{N}{N * \bar{h}} = \frac{1}{\bar{h}} \tag{3.7}$$

Time headway can be defined as the time interval between the front bumper (or rear bumper) of a leading vehicle is at a point on the road and the front bumper (or rear bumper) of the following vehicle arrives at the same point on the road. It is in terms of seconds.

Space headway is very similar to time headway. It can be defined as the distance between a leading vehicle's front bumper (or rear bumper) and the front bumper (or rear bumper) of the following vehicle. It is in terms of meters.

Capacity is defined as the maximum amount of flow rate that the road can sustain under prevailing roadway, traffic and control conditions [27].

Traffic flow on a roadway can be in three forms which are: free flow, flow at capacity and congested flow. The definitions of these three forms are given as follows:

- Free Flow Traffic: Under free flow conditions, drivers have the freedom to travel with the speed they desire. The interaction between vehicles is at minimum. Free flow traffic is largely associated with low density and high speed conditions.
- Flow at capacity: The drivers are much more restricted in their freedom compared to free flow conditions. Vehicle interactions are much more prevalent too as the time headways between vehicles decrease significantly. At flow at capacity, the road is at its most efficient state.
- Congested Flow: The traffic is at its worst state. Density is at its maximum and the flow rate and speed are at their minimum. It may result in stop and go traffic. When the vehicles are stopped, space headway reaches its minimum as all vehicles are located bumper to bumper.

Hall *et al.* [28] proposed a speed-flow curve that is based on real life measurements from various highways which is shown in Figure 3.1. On this curve, the three forms of traffic flow which is mentioned above are visualized by the three segments. Segment 1 is equivalent to free flow traffic, segment 2 is equivalent to capacity flow and segment 3 is equivalent to congested flow. On segment 1, the speed is constant and is equal to the free flow speed up until to a certain flow rate from where it experiences a slight drop. On this segment free flow conditions are prevalent and therefore free flow traffic is experienced. On segment 2, there is a big drop in speed for the same flow rate which is based on the studies done by Persaud and Hurdle [29]. This drop reflects the highly varied speeds by the vehicles at capacity. For this entire segment, the flow rate is equal to the capacity flow. On segment 3, the speeds observed are very low because of the congested nature of the traffic. At its extreme, both the flow rate and speed can get the value of 0.

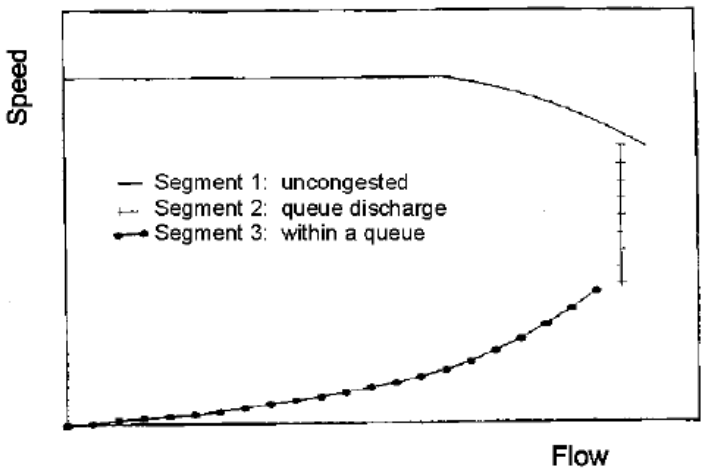


Figure 3.1. A generalized speed flow curve proposed by Hall et al.

### 3.1.3. Density

Density is the number of vehicles on a unit length of road, such as a km. It is a measure of how crowded a road section is. It is an important traffic characteristic as it is a measure of traffic performance according to users of the road and traffic operators. Density reflects the maneuvering freedom that the drivers have. It is closely related to space headway. This relationship is shown as in the following equation:

$$k = \frac{1}{\bar{h}_s} \tag{3.8}$$

where

$\bar{h}_s$= average space headway (meters).

Density can also be represented in terms of occupancy such as in Equation 3.9:

$$Density = \frac{occupancy}{average\ vehicle\ length + detection\ zone\ length} \tag{3.9}$$

Occupancy is defined as the percentage of time a detector is covered. It can have values between 0% and 100%. The former means that there are no vehicles on the road and the latter means that the traffic is stopped. Occupancy is highly correlated with density.

Hall et al. (1986) found that occupancy, like density, can be used in classifying traffic conditions which are congested, free flow and transitional states. Occupancy can simply be measured by the following equation:

$$occupancy = \frac{\sum_{i=1}^{n} t_i}{t} \tag{3.10}$$

where

$t_i$: the time an individual ($i^{th}$) vehicle spends on top of a detector,
n: total number of vehicles to pass through the detector,
t: total duration of observation.

Optimum density and jam density are two important parameters. Optimum density is defined as the density which exists when the traffic is at its capacity. Jam density on the other hand is the density at which flow and speed is equal to zero and when no vehicle is moving. It is the density observed on traffic jams. According to the Greenshields' model the jam density is two times the optimum density. But, it is known from field studies that the jam density is more than two times the optimum density. A typical value of optimum density is around 36 to 48 vehicles/mile whereas a typical value of jam density is around 185 to 250 vehicles/mile [30].

## 3.2. Traffic Stream Models

Traffic stream models can be divided into two in general which are single regime and multi regime models. In single regime models, only a single relationship between the traffic parameters is assumed for the entirety of flow conditions, from free flow to congested flow. Multi-regime models tried to improve on this by separating relationships are assumed for differing flow conditions.

The first single regime model and also the first comprehensive model that relates traffic flow parameters with each other is proposed by Greenshields [25]. Greenshields assumed a linear relationship between speed and density just as shown in Figure 3.2. However, this equation was established by only seven measurement points, 6 of them

being in free flow conditions. Despite this, it is not that far off the actual speed-density relationships that are observed, so it is still seen as an accurate representation of the relationship between speed and density. Until then, this relationship has been improved time and time again. Speed can be calculated using the parameters shown in Figure 3.2 as shown in Equation 3.11:

$$u = u_f(1 - \frac{k}{k_j}) \tag{3.11}$$

where

$u_f$: free flow speed (kilometers/hour),
$k_j$: jam density (vehicles(kilometer).

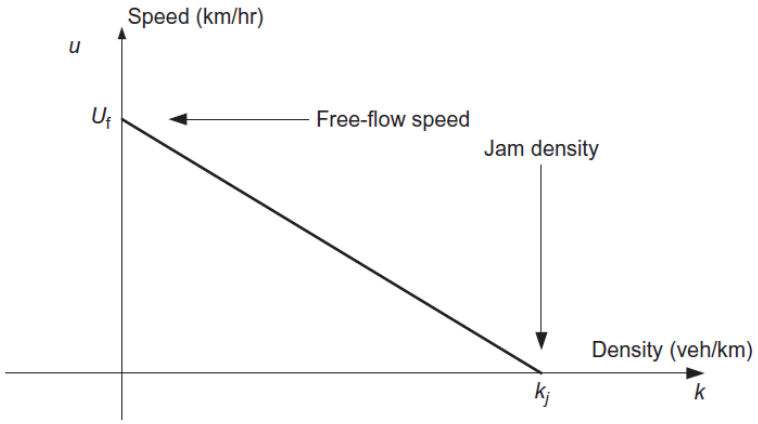


Figure 3.2. Speed-density diagram [31].

Keeping density on the left side and all the others on the right, we get:

$$k = k_j(1 - \frac{u}{u_f}) \tag{3.12}$$

Using the fundamental relationship of traffic flow:

$$q = u * k = u * k_j\left(1 - \frac{u}{u_f}\right) = k_j(u - \frac{u^2}{u_f}) \tag{3.13}$$

Equation 3.13 shows the basic relationship between flow and speed according to Greenshields' model. Figure 3.3 shows the plot form of this equation. Here, $u_m$ is the mean speed and $u_f$ is the free flow speed. The graph can be divided into two sections which are stable and unstable traffic conditions. The part above $u_m$ denotes the stable traffic conditions where the speed average is high and the traffic is fluent. The part below $u_m$ shows the unstable traffic conditions which is apprehensible from the low speed values in this region.

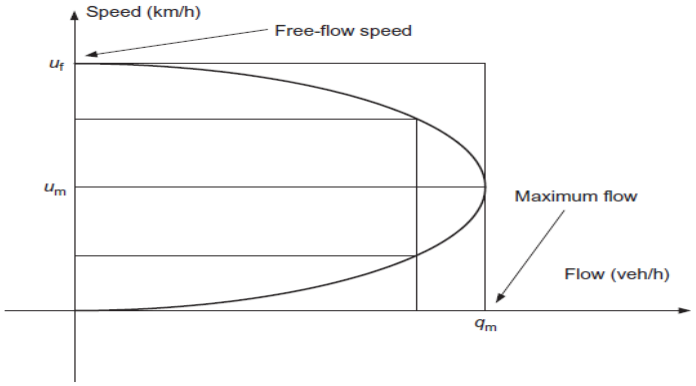


Figure 3.3. Speed-flow diagram [31].

Flow-density relationship can be derived by using the fundamental equation and speed-density relationship just like flow-speed relationship:

$$q = u * k = u_f \left(1 - \frac{k}{k_j}\right) k = u_f \left(k - \frac{k^2}{k_j}\right) \tag{3.14}$$

Figure 3.4 shows the plot form of this equation. In this figure, $q_m$ denotes the flow at capacity, $k_m$ denotes the optimum density and $k_j$ denotes the jam density. It is a simple parabolic relationship that is similar to the speed-flow curve. The left side of $k_m$ is the free flow region whereas the right side of $k_m$ is the congested flow region.

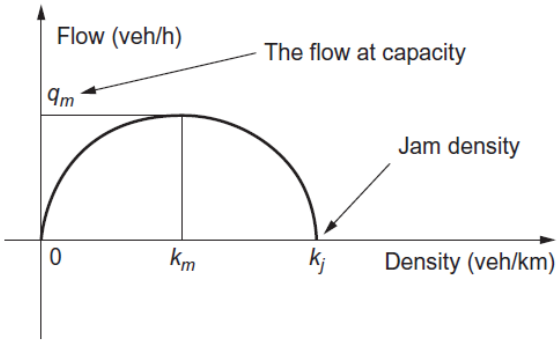


Figure 3.4. Flow-density diagram [31].

Figures 3.5.a, 3.5.b and 3.5.c show the differences between a sample empirical data and Greenshields' speed-density model and the corresponding speed-flow and flow-density models. As it is clearly seen from these graphs, Greenshields' model is off from empirical measurements. Speed-density graph shows a negative exponential trend in real life measurements whereas it is linear in Greenshields' model. On the empirical speed-flow graph, the speed remains constant with little drop until peak flow rate is reached. After this point there is a drop of speed just as mentioned in the model proposed by Hall *et al.* [28]. After the drop the speed and flow diminishes to zero. On the speed-flow version of the Greenshields' model, a parabolic shape is observed. On this graph, some features that are present on the empirical measurements are lacking such as the speed drop-off at peak flow or the constant speed from zero flow to peak flow rate. The flow-density graph obtained from empirical measurements show an inverted v pattern where the flow increases linearly until optimum density and then decreases linearly until jam density is reached. Greendhields' flow-density graph is parabolic just like the speed-flow graph. Greenshields' model overestimates the optimal density and underestimates the jam density compared to the empirical measurements. Despite its deficiencies, Greenshields' speed-density model, and its corresponding speed-flow and flow-density models are still used today thanks to their simplicity and its sufficiency in explaining the relationships between fundamental traffic parameters.

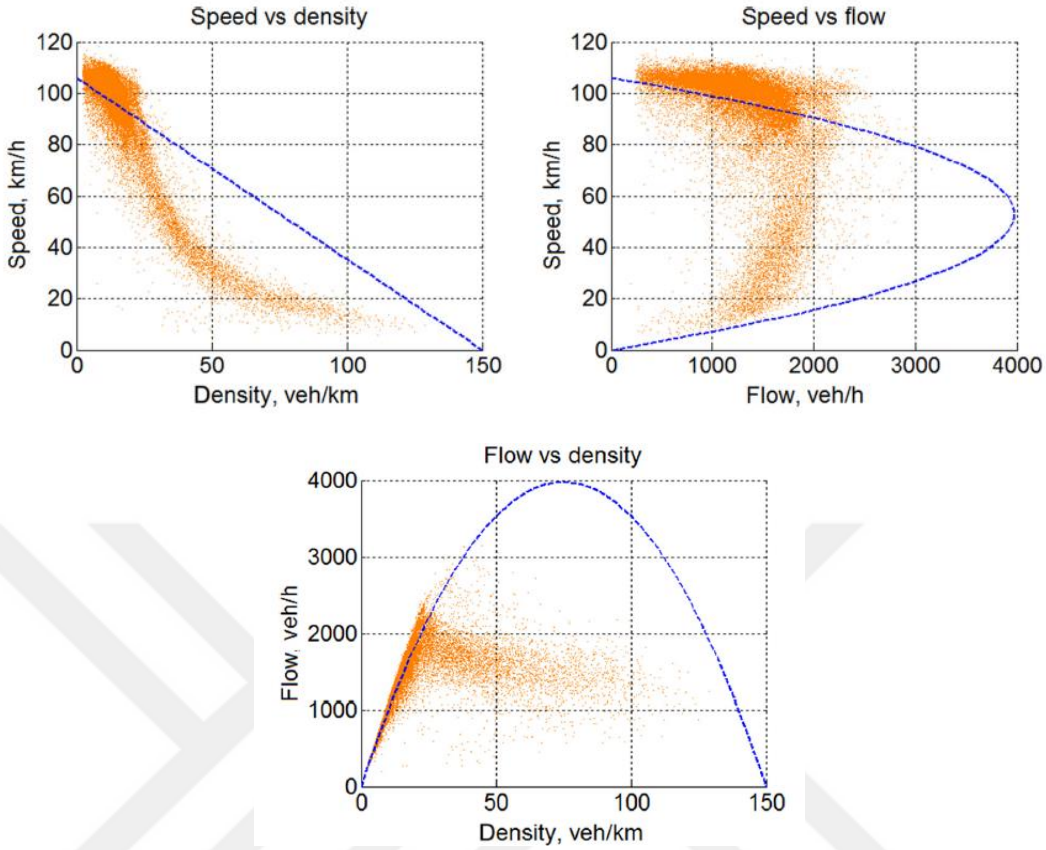


Figure 3.5a (Speed vs density) 3.5b (Speed vs flow) 3.5c (Flow vs density) Comparison of Greenshields' models (in blue) and sample empirical data [32].

After Greenshields, other engineers such as Greenberg, Underwood, Drake, Drew and Pipes came up with their own single-regime traffic flow models. These models are more accurate than the Greenshields' model, but still are lacking in explaining the real life traffic flow characteristics. A summary of these models is given in Table 3.1.

Greenberg's model [33] is the second single regime model in the literature. It is concluded in this model that a nonlinear model would be more suitable than a linear model, such as Greenshields' model. This model requires the jam density, like Greenshields' model, and the optimum speed, which is difficult to measure. A big drawback of this model is the fact that the free flow speed diverges to infinity. Edie addresses this issue by using Greenberg's model only in the congested region in his multi-regime model [34].

Underwood's model [35] requires the free flow speed and the optimum density to be known. The problem with this model is that the speed is never zero and the jam density is

infinity. For this reason, Edie used Underwood's model in the free flow region of his model [34].

Drake's model [36] is a slight improvement of Underwood's model [35]. It shares the same issues with Underwood's model which were speed never being equal to zero and jam density diverging to infinity.

Drew's model [37] introduced a parameter *n* which can take the values -1, 0 and 1 to create a family of models. When n is equal to 1, the equation turns into Greenshields' model. When it is zero, it becomes parabolic and when it is -1, it becomes exponential. Pipes-Munjal model [38] is quite similar to Drew's model. Again for an n value of 1, the equation reduces to Greenshields' model.

Table 3.1. A summary of single regime speed-density models.

| Author | Model |
|---|---|
| Greenshields | $u = u_f(1 - \frac{k}{k_j})$ |
| Greenberg | $u = u_o \ln(\frac{k_j}{k})$ |
| Underwood | $u = u_f e^{-\frac{k}{k_0}}$ |
| Drake | $u = u_f e^{-\frac{1}{2}(\frac{k}{k_o})^2}$ |
| Drew | $u = u_f[1 - \left(\frac{k}{k_j}\right)^{(n+1)/2}]$ |
| Pipes-Munjal | $u = u_f[1 - \left(\frac{k}{k_j}\right)^{n}]$ |

Because the single regime models presented above are unable to fit the empirical observations for the entirety of density range, it was thought that it would be better to use multiple equations for different ranges of density. Multi-regime models were proposed with this set of mind. Edie's model [34] was the first multi-regime model. As mentioned before, it combined the Underwood model for free flow region and Greenberg model for the congested flow region.

Three other multi-regime models were proposed by the researchers from Northwestern University [36]. The first model used 2 distinct equations which originates from Greenshields' equation for the free flow and congested flow regions. The second model accepts a constant free flow speed for the free flow region and a Greenberg model for the congested flow region. The third model is a three regime model which incorporates 3 Greenshields' type equation to free flow, transitional flow and congested flow regions. A summary of all these multiregime models are presented in Table 3.2.

Table 3.2. A summary of multi regime speed-density models.

| Model Type | Free-Flow Region | Transitional-Flow Region | Congested-Flow Region |
|---|---|---|---|
| Edie Model | $u = 54.9e^{-k/163.9}$ (k≤50) | -------- | $u = 26.8 * \ln(\frac{162.5}{k})$ (k≥50) |
| Two-Regime Model | $u = 60.9 - 0.515k$ (k≤65) | --------- | $u = 40 - 0.265$k (k≥65) |
| Modified Greenberg Model | $u = 48$ (k≤35) | | $u = 32\ln(\frac{145.5}{k})$ (k≥35) |
| Three-regime Model | $u = 50 - 0.098k$ (k≤40) | $u = 81.4 - 0.913k$ (40≤k≤65) | $u = 40 - 0.265k$ (k≤65) |

A better alternative to the Greenshields' model (and the other single regime models) is the triangular speed-density model which is presented in Figure 3.6. This model is seen as a much better representation of traffic flow by researchers and engineers and is also often used in academia. In the flow-density graph, it is assumed that:

- Flow rate is zero at jam density and zero density just like Greenshields' model.
- Flow rate increases linearly until $k_c$, optimum density. The slope here is equal to the free flow speed.
- Flow rate decreases linearly starting with optimum density until jam density with slope *w* which is the shockwave speed.

With these assumptions, the following conclusions can be arrived:

- On the speed-density graph, speed is constant and is equal to free flow speed until optimum density and then it exponentially diminishes to 0 until jam density.
- On the speed-flow graph, speed is constant and is equal to free flow speed until capacity is reached and then diminishes to zero at flow=0.

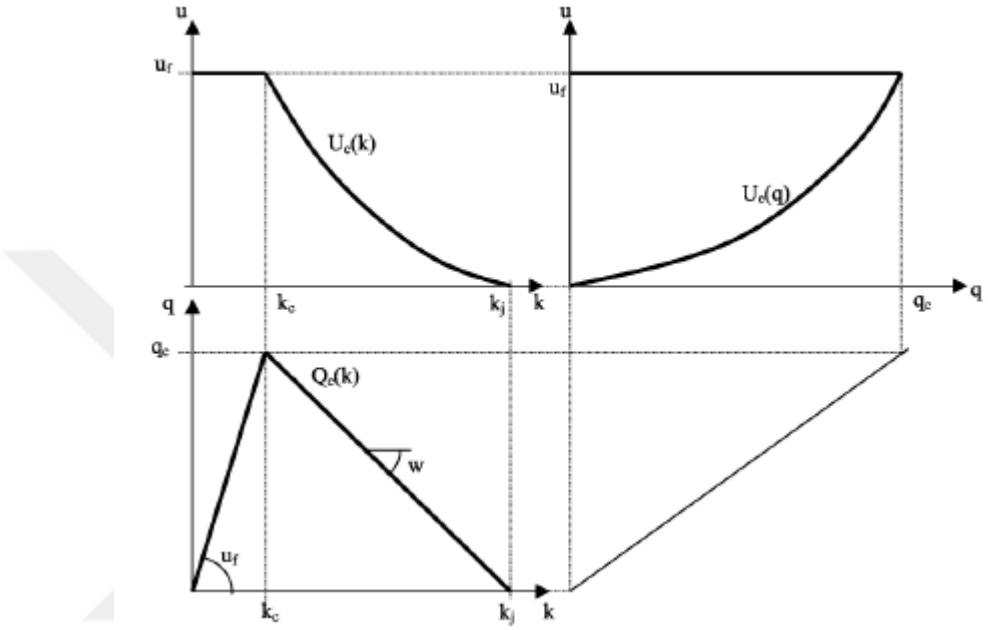


Figure 3.6. Fundamental diagrams in the case of triangular speed-flow relationship [39].

## 3.3. Regression

### 3.3.1. Simple Linear Regression

In this thesis, regression analysis is used to come up with meaningful relationships between the traffic parameters. Regression analysis is a technique which is used to find the best relationship between a dependent variable *Y* and an independent variable *x* [40]. The independent variable *x* is used to explain the dependent variable *Y*. The simplest form of this relationship is a linear relationship which is given in the following equation:

$$Y = \beta_0 + \beta_1 x \tag{3.15}$$

where

$\beta_0$= intercept,

$\beta_1$= slope.

This equation, which is also called the simple regression equation, is visualized in Figure 3.7.

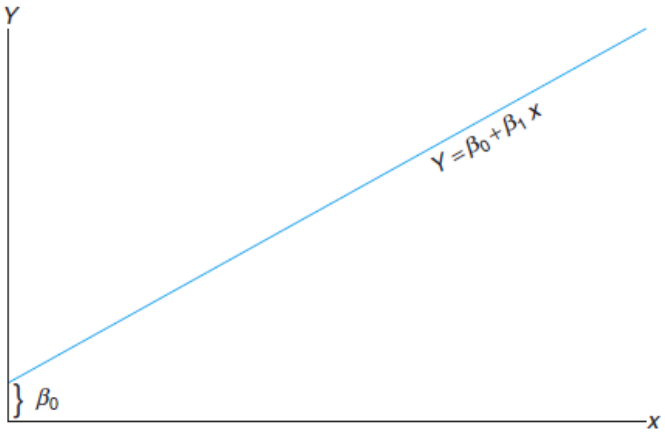


Figure 3.7. Simple linear relationship between y and x [40].

In reality however, most of the data points that are gathered from observations do not fit this line exactly and will deviate from this line by some margin. This margin is called the error. With the addition of error, Equation 3.15 changes into:

$$Y = \beta_0 + \beta_1 x + \varepsilon \tag{3.16}$$

where

ε = error,
$\beta_0$ and $\beta_1$ previously mentioned.

On Figure 3.8, a sample regression line with 5 data points is shown. Because these data point are not linear, a linear line that is representative of the general trend of all these points is created between these data points. In doing so, 5 error terms are created.

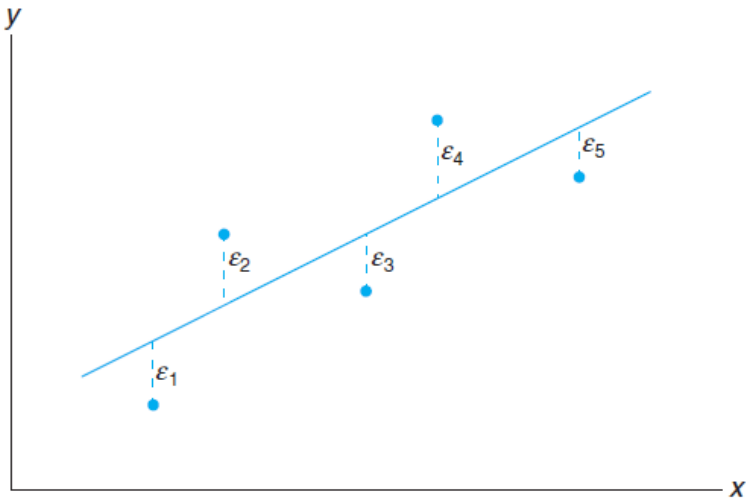


Figure 3.8. Sample simple regression with error terms [40].

Every individual observation *i* can be calculated using the following equation:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{3.17}$$

where

$y_i$ = value of the dependent variable *i*,
$x_i$ = value of the independent variable *i*.

Estimating the regression coefficients $\beta_0$ and $\beta_1$ is an important task of regression analysis. To this estimators $b_0$ and $b_1$ are used in their place to create the estimated regression line which can be denoted as:

$$\hat{y} = b_0 + b_1 x \tag{3.18}$$

where

$\hat{y}$ = predicted value.

Error terms ($e_i$) are defined as the difference between the observed and estimated values. Error, which is also called the residual, can be formulated as follows:

$$e_i = y_i - \hat{y}_i \tag{3.19}$$

It is possible to compare model parameters $b_0$ and $b_1$ with certain numerical values such as $\beta_0$ and $\beta_1$. For this goal, t-statistics are used. If the values $\beta_0$ and $\beta_1$ are selected as 0, the t-statistics will give answer to whether the parameters are different than zero or not. The t-statistics can be calculated using the following equarions:

$$t_{b_0} = \frac{b_0 - \beta_0}{\sqrt{\frac{\sum x_i^2 \sum e_i^2}{n(n-2)(\sum x_i^2 - n\bar{x}^2)}}} \tag{3.20}$$

$$t_{b_1} = \frac{b_1 - \beta_1}{\sqrt{\frac{\sum e_i^2}{(n-2)(\sum x_i^2 - n\bar{x}^2)}}} \tag{3.21}$$

where

$t_{b_0}$= t-statistics of parameter $b_0$,

$t_{b_1}$= t-statistics of parameter $b_1$,

$x_i$= individual independent variable terms,

$e_{i=}$ individual error terms,

n= sample size,

$\bar{x}$= mean of independent variable.

After calculating the t-statistics using the equations above, the found values are compared with the table values for t-statistics. To find the correct table value, several information such as, confidence interval, degree of freedom and type of distribution (1 tailed or 2 tailed) are needed. The parameters can be tested to see if they are significantly different than zero. In such a case, if the calculated t-statistics value is bigger than table values, it is concluded that the parameters are not equal to zero and are significant.

The best fit for the model is obtained when the sum of squares of errors (SSE) are minimized because a lower SSE means that the deviations from the regression line are minimized. This procedure is called the method of least squares. The sum of squares of errors can be calculated as follows:

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (3.22)$$

Regression sum of squares (SSR) is another important parameter and it is defined as the variation of dependent variable explained by the model. It is formulated as follows:

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \quad (3.23)$$

The summation of regression sum of squares and error sum of squares give the total sum of squares (SST) as is shown in the following equation:

$$SST = SSR + SSE \quad (3.24)$$

Testing the simple linear model in terms of the relationship between Y and X is done with F-statistics using the following equation:

$$F = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i e_i^2/(n-2)} = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (\hat{y}_i - y_i)^2/(n-2)} = \frac{Mean\ Regression\ Sum\ of\ Squares}{Mean\ Error\ Sum\ of\ Squares} \quad (3.25)$$

where

$\hat{y}_i$ = i$^{th}$ predicted value,
$\bar{y}$ = mean of values of the dependent variable,
$e_i^2$= sum of squares of errors,
n= sample size,
$y_i$ = value of the dependent variable *i*.

Mean sum of squares can be found by dividing sum of squares with degrees of freedom. In this case the degree of freedom is one for regression because there is only one independent variable and (n-2) for the error term. If the F-value obtained from the equation above is bigger than the table value which is determined with respect to a certain level of significance and degrees of freedom of errors and regressors, it can be concluded that there is a relationship between the independent variable x and the dependent variable y, i.e. $\beta \neq 0$.

Testing the adequacy of the fitted line is done by a parameter called coefficient of determination, which is denoted by $R^2$. Coefficient of determination tells us how much of the variability is explained by the fitted model. It has values between 0 and 1, former indicating no fit and latter indicating perfect fit. It is formulated by:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{3.26}$$

As it is clearly seen by this equation, $R^2$ becomes equal to 1 if SSE is zero. This means that if all the data points are situated on the fitted line, the fitted line is perfectly adequate.

### 3.3.2. Multiple Linear Regression

In the previous section, regression models with only one independent variable were described. However, most regression models require multiple independent variables. For this reason, a more sophisticated technique known as multiple linear regression is used for models with multiple independent variables. A general form of a multiple linear regression model can be represented as:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i \tag{3.27}$$

where

$\beta_i$'s = coefficients where $\beta_0$ is the intercept,
$y_i$ = $i^{th}$ value of the dependent variable,
$x_i$'s = $i^{th}$ value of the k independent variables,
$\varepsilon_i$'s = individual error terms.

Multiple linear regression models can be generalized using matrices. This model is called the general linear model and they can be denoted as:

$$y = X\beta + \varepsilon \tag{3.28}$$

where

$$y=\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, X=\begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \ldots & x_{kn} \end{bmatrix}, \beta=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Testing the general linear model is done with F tests similar to simple linear models. To test the null hypothesis that all the coefficients are equal to zero i.e.:

$$H_0=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The F-statistics can be formulated as:

$$F = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MER} = \frac{R^2/(k)}{(1-R^2)/(n-k-1)} \tag{3.29}$$

where

SSR= Regression sum of squares,
SSE= Error sum of squares,
MSR= Mean regression sum of squares,
MER= Mean error sum of squares,
n= number of equations,
k= number of independent variables,
$R^2$ = coefficient of determination.

If the calculated F value using Equation 3.29 is greater than the corresponding table F-statistics value, the null hypothesis that states that all the model coefficients are zero can be rejected.

For testing whether individual coefficients are equal to zero or not, t-statistics are used. The t-statistics can be calculated with the following equation:

$$t_{b_i} = \frac{\beta_i - b_i}{\sqrt{e'e/(n-k-1)c_{ii}}} \quad (3.30)$$

where

$c_{ii}$= the value from the i$^{th}$ row and i$^{th}$ column of the matrix (X'X)$^{-1}$.

Alternatively, t-statistics can be calculated with:

$$t = \hat{\beta}/(standard\ error\ of\ \hat{\beta}) \quad (3.31)$$

Standard error is the square root of the i$^{th}$ diagonal element of the covariance matrix $\hat{\sigma}^2(X'X)^{-1}$. It is very similar to standard deviation, the main difference being that standard error is about a sample and standard deviation is about a population. It is a standard output in many statistical softwares such as SPSS and Minitab. For both of the equations above, if the t-value found is bigger than the table t-value, the null hypothesis that the coefficient is equal to zero can be rejected. With such as a result, it can be concluded that this coefficient is not equal to zero and it is significant.

### 3.3.3. Factor Analysis

Factor analysis is the procedure of identifying relationships between the large amount of variables to see how they are related. It is a method that is used in reduction of data . Factor analysis is a part of general linear model and it can be divided into two types [41]:

- Exploratory Factor Analysis: In exploratory factor analysis, any variable may be associated with any of the factors. This is the most favored type of factor analysis by researchers and it is also the type that is used in this thesis.
- Confirmatory Factor Analysis: In confirmatory factorial analysis, unlike exploratory factor analysis, the factor structure is pre-determined and then tested to see if it is true. It is assumed that each factor is related with stated subset of measured variables.

There are several methods of factor analysis and some of these are [41]:

- Principal component analysis: The most well-known and most popular method. Principal component analysis is done with the following methodology: First the maximum variance observed is extracted and put into the first factor. Then the variance that is explained by the first factors are removed and the same process is done for the second factor. It continues until the last factor is reached.
- Common factor analysis: It is the second most used method after principal component analysis. It is done by extracting the common variance and putting them into factors. The main goal of common factor analysis is to find the least number of factors that can attribute for the common variance of the variables.
- Maximum likelihood method: It is a method which determines the values of the parameters inside the model. The determination of these parameter values are done in a way that the likelihood of the process designated by the model is maximized [42].

Assumptions of factor analysis are [41]:

- No outliers
- Sample size should be adequate.
- Multicollinearity doesn’t necessarily have to be perfect
- Homoscedasticity, which is the equivalance of variance, is not required
- Linearity

# 4. METHODOLOGY

## 4.1. Study Area Description

For this study, a freeway segment located upstream of the 15 July Martyrs Bridge on the D100 Highway is selected. This location is selected because it is an interesting location that is prone to traffic congestions because of the bridge and the two on-ramps, one of which are for private vehicles, and the other is for metrobuses. The main arterial is 3 lanes wide, the vehicle on-ramp is 2 lanes wide but it narrows to a single lane just before the merge with the main arterial and the metrobus on-ramp is a single lane road. The location of the study area is shown in Figure 4.1. The satellite image of the study area is shown in Figure 4.2. In this figure, the study area is disclosed by the red lines. A sketch of the study area is presented in Figure 4.3.

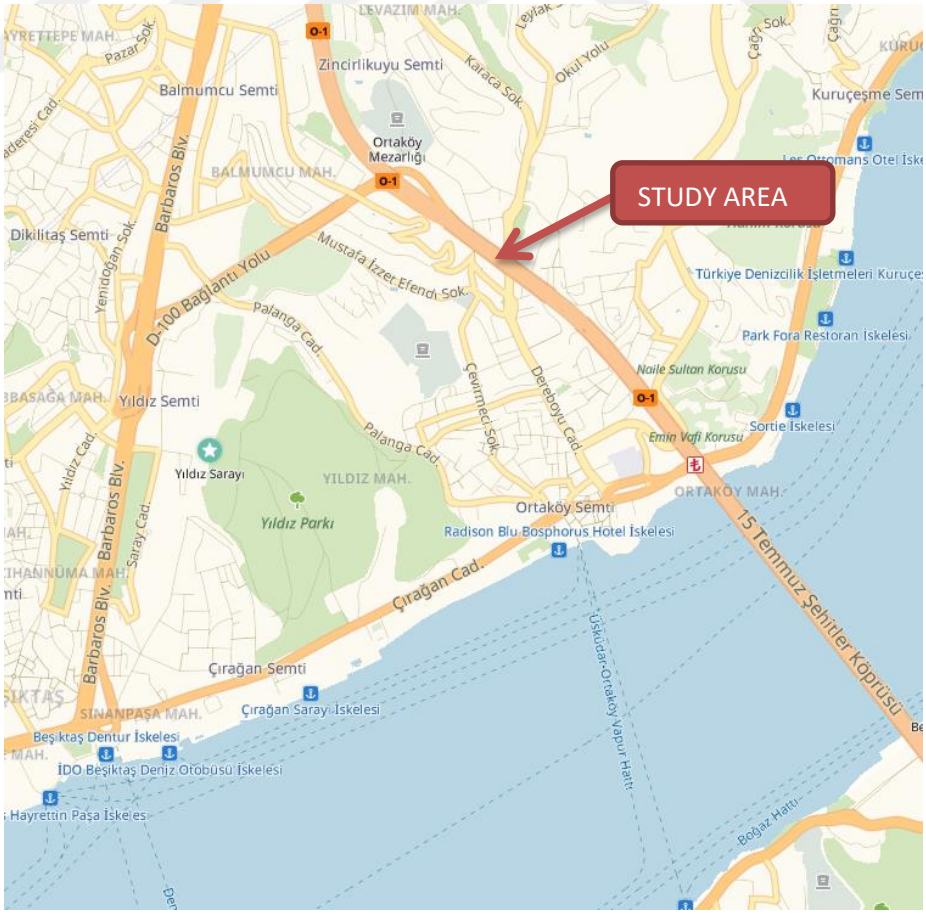


Figure 4.1. Location of the study area (Yandex Maps).

Figure 4.2. The satellite image of the study area (Yandex Maps).

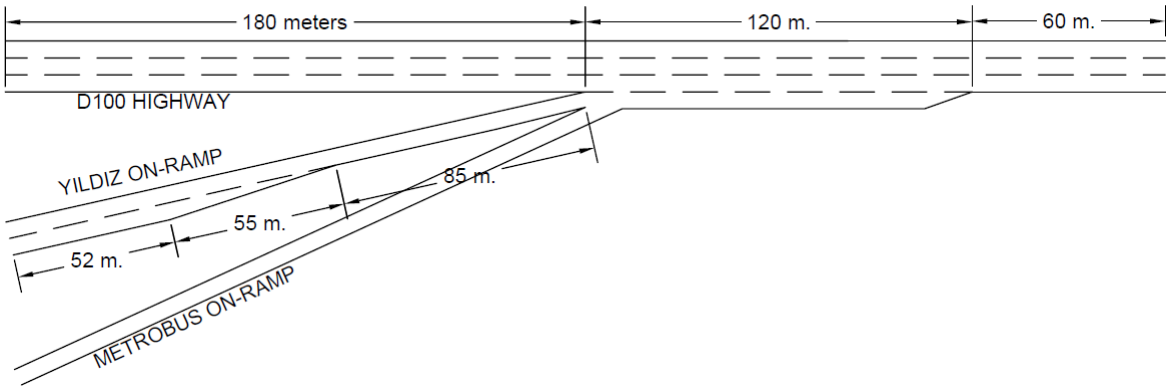


Figure 4.3. A sketch of the study area.

A video camera overseeing the region is used in the analysis of traffic flow. Figure 4.4 shows a still image from the video footages obtained from this camera. For the analysis of traffic and for convenience, the road is divided into regions as shown in Figure 4.5.

Figure 4.6 shows a clearer picture of the regions on the main arterial compared to Figure 4.5. In this figure, it is shown that regions 0, 3, 6, 9, 12, 15, 18, 21 are on the right lane, regions 1, 4, 7, 10, 13, 16, 19, 22 are on the middle lane and regions 2, 5, 8, 11, 14, 17, 20, 23 are on the left lane. The lengths of the initial regions on the main arterial are selected in reference to the lane lines and the gaps between the lane lines. The lengths of a single lane line and a single spacing between lane lines is found to be 15 meters using the satellite images provided by Yandex Maps. The length of the first few regions on the main arterial is selected as 2 lane lines and 2 gaps between lane lines. This means that for these regions, the length is equal to 30 meters. The lengths of further regions on the main arterial and the regions in the off-ramp can't be determined this way because of issues related to visibility and unavailability of lane lines. For this reason, the lengths of these regions are measured manually with the help of Google Maps.

The lengths and locations, i.e., the beginning and ending points, of the regions are given in Table 4.1. The corner points of the regions, which are called nodes, are also numbered as seen in Figure 4.5. The upper leftmost point in Figure 4.5 is the origin and has the pixel coordinates (0,0). Coordinates of all the nodes on the road are determined with respect to this origin point. Table 4.2 shows the pixel locations of every node in the area.

The main areas of interest in this thesis are the traffic characteristics of the regions located on the main arterial. Thus, the regions on the on-ramp will not be considered in the analysis.

It should also be said that the regions that are shown both on Figure 4.5 and Figure 4.6 were the original regions that were created before the data gathering stage. Due to issues related to visuals, detection quality and inadequacy of data, only regions up to and including 15, 16 and 17 are present in the final data.


Figure 4.4. Still image from the video footages.

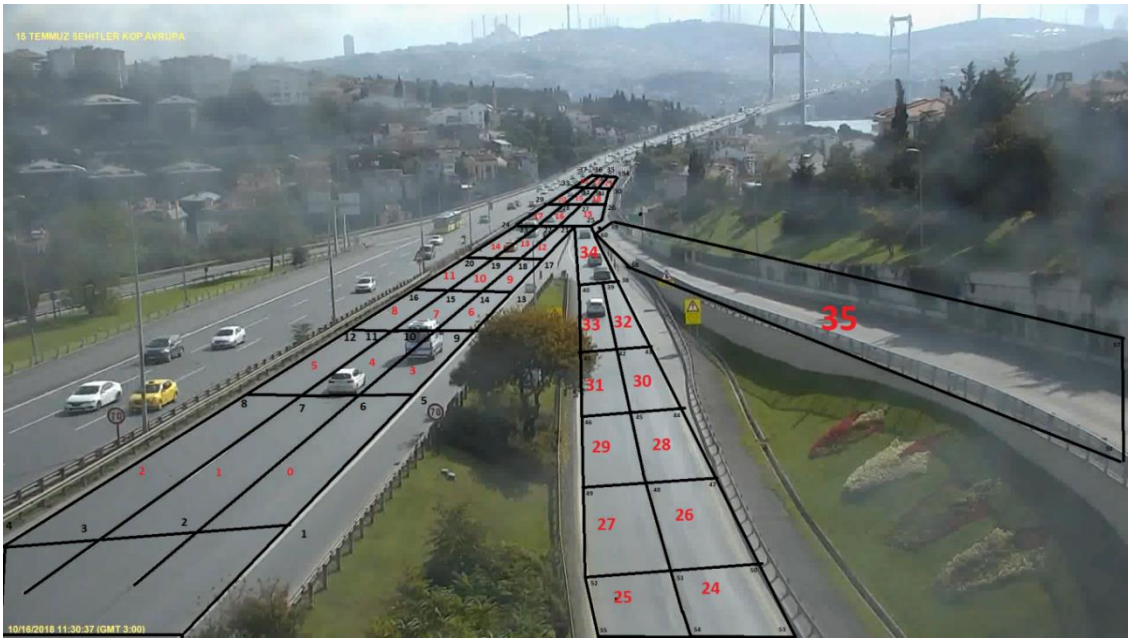


Figure 4.5. Study area with the original regions.

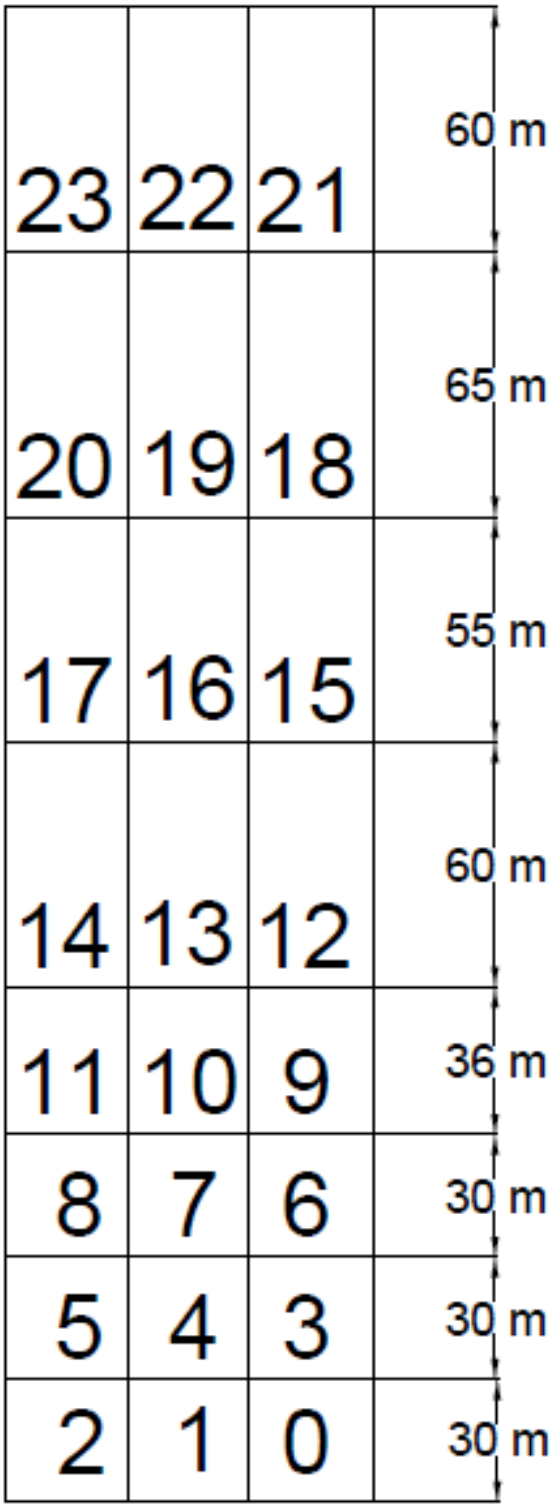


Figure 4.6. Regions on the main arterial with information regarding their lengths.

Table 4.1. Region coordinates.

| Region | Start(m.) | End(m.) | Length(m.) | Region | Beginning(m.) | End(m.) | Length(m.) |
|---|---|---|---|---|---|---|---|
| 0 | 11.8 | 41.8 | 30 | 18 | 252.8 | 317.8 | 65 |
| 1 | 11.8 | 41.8 | 30 | 19 | 252.8 | 317.8 | 65 |
| 2 | 11.8 | 41.8 | 30 | 20 | 252.8 | 317.8 | 65 |
| 3 | 41.8 | 71.8 | 30 | 21 | 317.8 | 377.8 | 60 |
| 4 | 41.8 | 71.8 | 30 | 22 | 317.8 | 377.8 | 60 |
| 5 | 41.8 | 71.8 | 30 | 23 | 317.8 | 377.8 | 60 |
| 6 | 71.8 | 101.8 | 30 | 24 | 0 | 8.21 | 8.21 |
| 7 | 71.8 | 101.8 | 30 | 25 | 0 | 8.21 | 8.21 |
| 8 | 71.8 | 101.8 | 30 | 26 | 8.21 | 22.81 | 14.6 |
| 9 | 101.8 | 137.8 | 36 | 27 | 8.21 | 22.81 | 14.6 |
| 10 | 101.8 | 137.8 | 36 | 28 | 22.81 | 32.63 | 9.82 |
| 11 | 101.8 | 137.8 | 36 | 29 | 22.81 | 32.63 | 9.82 |
| 12 | 137.8 | 197.8 | 60 | 30 | 32.63 | 51.86 | 19.23 |
| 13 | 137.8 | 197.8 | 60 | 31 | 32.63 | 51.86 | 19.23 |
| 14 | 137.8 | 197.8 | 60 | 32 | 51.86 | 109.16 | 57.3 |
| 15 | 197.8 | 252.8 | 55 | 33 | 51.86 | 109.16 | 57.3 |
| 16 | 197.8 | 252.8 | 55 | 34 | 109.16 | 197.01 | 87.85 |
| 17 | 197.8 | 252.8 | 55 | | | | |

Table 4.2. Node coordinates.

| Node | x(pixels) | y(pixels) | Node | x(pixels) | y(pixels) |
|---|---|---|---|---|---|
| 1 | 501 | 887 | 31 | 1,014 | 312 |
| 2 | 333 | 898 | 32 | 991 | 311 |
| 3 | 166 | 910 | 33 | 970 | 309 |
| 4 | 1 | 923 | 34 | 1,051 | 293 |
| 5 | 710 | 665 | 35 | 1,032 | 293 |
| 6 | 606 | 666 | 36 | 1,016 | 292 |
| 7 | 507 | 667 | 37 | 1,001 | 291 |
| 8 | 410 | 664 | 38 | 1,055 | 471 |
| 9 | 808 | 555 | 39 | 1,024 | 473 |
| 10 | 736 | 555 | 40 | 982 | 477 |
| 11 | 662 | 557 | 41 | 1,107 | 581 |
| 12 | 588 | 552 | 42 | 1,048 | 586 |
| 13 | 873 | 490 | 43 | 984 | 592 |
| 14 | 817 | 487 | 44 | 1,159 | 688 |
| 15 | 758 | 487 | 45 | 1,071 | 693 |
| 16 | 700 | 484 | 46 | 988 | 702 |
| 17 | 925 | 434 | 47 | 1,218 | 806 |
| 18 | 881 | 433 | 48 | 1,096 | 816 |
| 19 | 834 | 433 | 49 | 990 | 823 |
| 20 | 785 | 430 | 50 | 1,290 | 952 |
| 21 | 974 | 376 | 51 | 1,140 | 964 |
| 22 | 948 | 376 | 52 | 993 | 977 |
| 23 | 912 | 375 | 53 | 1,356 | 1075 |
| 24 | 867 | 376 | 54 | 1,173 | 1073 |
| 25 | 1,006 | 375 | 55 | 1,016 | 1076 |
| 26 | 1,030 | 341 | 56 | 1,912 | 805 |
| 27 | 987 | 341 | 57 | 1,912 | 498 |
| 28 | 957 | 341 | 58 | 1,066 | 450 |
| 29 | 923 | 341 | 59 | 1,062 | 357 |
| 30 | 1,040 | 313 | 60 | 1,013 | 392 |

## 4.2. Data Preparation

The output obtained from image processing algorithms contains the following elements: Vehicle id, time, x and y coordinates of the vehicles and the region of the vehicles which is dependent on the x and y coordinates. The coordinates in the outputs are in terms of pixels, meaning that it shows the location of the vehicles in terms of the pixels that vehicles appear on the computer screen. This coordinate is selected as the middle point of the center of the rectangle and the center of the bottom edge of the rectangle. The rectangle here is used to track a vehicle which is visible in Figure 4.7. The selection of the coordinate was done this way because it was established that this selection gave the best representation of a vehicle's position on the roadway for early regions on the main arterial. However the detection of vehicles' regions proved to be problematic in the further regions starting from regions 9, 10, 11. For these regions, and some other regions that had off results, correction factors are used to come up with more accurate results. A sample output from video image processing is shown in Table 4.3.

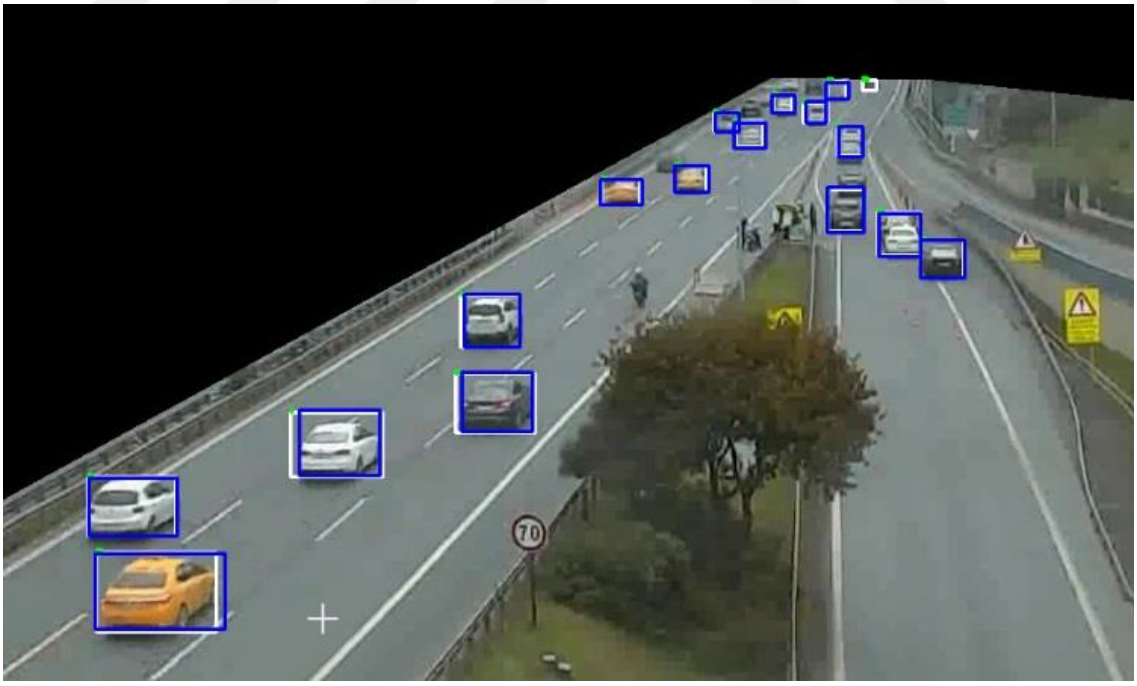


Figure 4.7. A sample image from vehicle tracking mechanism.

Table 4.3. Sample output obtained from image processing.

| ID | FRAME | TIME | XMIN | YMIN | XMAX | YMAX | CX | CY | REGION |
|---|---|---|---|---|---|---|---|---|---|
| 2332 | 1814 | 217.68 | 247.141 | 702.721 | 347.297 | 769.9838 | 297.2192 | 753.1681 | 2 |
| 2332 | 1815 | 217.8 | 281.65 | 682.03 | 377.965 | 746.7585 | 329.8073 | 730.5764 | 2 |
| 2332 | 1816 | 217.92 | 315.071 | 665.971 | 403.94 | 725.536 | 359.5054 | 710.6448 | 2 |
| 2332 | 1817 | 218.04 | 346.559 | 655.105 | 424.02 | 706.4335 | 385.2894 | 693.6015 | 2 |
| 2332 | 1818 | 218.16 | 372.832 | 642.023 | 446.634 | 690.4256 | 409.7334 | 678.3249 | 2 |
| 2332 | 1819 | 218.28 | 395.433 | 628.247 | 467.691 | 675.2326 | 431.5619 | 663.4863 | 5 |
| 2332 | 1820 | 218.4 | 416.671 | 613.547 | 490.851 | 661.7337 | 453.7609 | 649.6869 | 5 |
| 2332 | 1821 | 218.52 | 437.428 | 600.006 | 511.3 | 647.9659 | 474.3639 | 635.9758 | 5 |
| 2332 | 1822 | 218.64 | 454.607 | 587.539 | 532.075 | 638.3435 | 493.3409 | 625.6425 | 5 |
| 2332 | 1823 | 218.76 | 475.639 | 577.421 | 548.88 | 625.8867 | 512.2591 | 613.7703 | 5 |
| 2332 | 1824 | 218.88 | 494.8 | 568.306 | 561.63 | 612.5282 | 528.2151 | 601.4726 | 5 |
| 2332 | 1825 | 219 | 509.893 | 555.276 | 578.434 | 600.9048 | 544.1632 | 589.4977 | 5 |
| 2332 | 1826 | 219.12 | 527.645 | 549.035 | 590.561 | 591.0718 | 559.1034 | 580.5627 | 5 |
| 2332 | 1827 | 219.24 | 542.843 | 540.638 | 603.783 | 581.4342 | 573.3128 | 571.2351 | 5 |
| 2332 | 1828 | 219.36 | 556.39 | 534.7 | 615.239 | 574.4381 | 585.8144 | 564.5035 | 5 |
| 2332 | 1829 | 219.48 | 570.57 | 526.859 | 627.611 | 565.6037 | 599.0904 | 555.9176 | 5 |
| 2332 | 1830 | 219.6 | 581.876 | 520.396 | 638.632 | 559.5461 | 610.254 | 549.7586 | 8 |
| 2332 | 1831 | 219.72 | 594.642 | 513.671 | 647.622 | 550.2925 | 621.1318 | 541.137 | 8 |
| 2332 | 1832 | 219.84 | 606.526 | 504.854 | 656.923 | 539.9233 | 631.7245 | 531.156 | 8 |
| 2332 | 1833 | 219.96 | 615.178 | 499.395 | 666.837 | 536.0669 | 641.0076 | 526.899 | 8 |
| 2332 | 1834 | 220.08 | 626.457 | 494.282 | 676.303 | 530.1763 | 651.3799 | 521.2028 | 8 |
| 2332 | 1835 | 220.2 | 636.968 | 488.083 | 684.923 | 523.0611 | 660.9457 | 514.3167 | 8 |
| 2332 | 1836 | 220.32 | 647.826 | 481.424 | 694.109 | 515.4303 | 670.9673 | 506.9287 | 8 |
| 2332 | 1837 | 220.44 | 658.156 | 474.584 | 703.527 | 508.2816 | 680.8413 | 499.8573 | 8 |
| 2332 | 1838 | 220.56 | 665.673 | 470.378 | 710.308 | 503.998 | 687.9902 | 495.5929 | 8 |
| 2332 | 1839 | 220.68 | 674.262 | 466.434 | 717.458 | 499.3841 | 695.86 | 491.1466 | 8 |
| 2332 | 1840 | 220.8 | 681.37 | 461.921 | 724.566 | 495.3363 | 702.9677 | 486.9824 | 8 |
| 2332 | 1841 | 220.92 | 688.514 | 459.15 | 729.259 | 490.7494 | 708.8863 | 482.8495 | 11 |
| 2332 | 1842 | 221.04 | 696.402 | 455.769 | 735.459 | 486.0364 | 715.9305 | 478.4696 | 11 |
| 2332 | 1843 | 221.16 | 702.072 | 452.075 | 739.442 | 481.2137 | 720.7569 | 473.929 | 11 |
| 2332 | 1844 | 221.28 | 709.337 | 446.195 | 747.837 | 476.3413 | 728.587 | 468.8047 | 11 |
| 2332 | 1845 | 221.4 | 714.985 | 442.899 | 752.692 | 472.7427 | 733.8386 | 465.2817 | 11 |
| 2332 | 1846 | 221.52 | 721.985 | 441.258 | 757.511 | 468.9988 | 739.7479 | 462.0635 | 11 |
| 2332 | 1847 | 221.64 | 726.805 | 435.454 | 764.302 | 465.2114 | 745.5534 | 457.7721 | 11 |

The image processing algorithm that is used for is called YOLO (You Only Look Once) more specifically, YOLO V3 algorithm. The main reason why YOLO is chosen is the fact that it is faster than other methods and it gives accurate results when used [43]. The detection algorithm of Yolo is as follows: First, a single neural network is applied to the

entirety of the image. Then, the image is divided into regions by the neural network. And finally boxes are created around the objects inside these regions. In this project, we are only interested in the movement of 4 wheel vehicles. Thus, adjustments have been made to eliminate data from other moving objects such as grass and data from motorcycles by defining a minimum box size for regions.

Having pixel coordinates at hand is not enough for the purposes of this thesis. To calculate speeds of vehicles in meaningful terms, i.e. km/hr, a conversion must be done. A sketch which contains 6 regions and their boundaries are given in Figure 4.7. With this figure, the basic idea behind the conversion from pixel coordinates to real coordinates will be told. The conversion is based on a vehicle's relative perpendicular distance to upper and lower region boundaries. First, the boundaries of the regions are formulized with respect to their slopes. As it can be seen from Figure 4.8, this equation can be generalized as:

$$0 = y + m_1 x + k \tag{4.1}$$

The regions' distances are given as 30, 60 and 90 meters in Figure 4.8. The black dot that is shown in the lower middle region denotes a vehicle's position in a time instant and its coordinate is ($x_0$,$y_0$). It was already possible to calculate this vehicle's pixel coordinates. To convert it to distance in terms of meters first the lengths a and b are calculated using the following equation which is the equation for the length between a dot and a line. The length a is calculated as the following:

$$a = \frac{|m_1 * x_0 + y_0 + k|}{\sqrt{m_1^2 + 1}} \tag{4.2}$$

The length b can be calculated in a similar way:

$$b = \frac{|m_2 * x_0 + y_0 + n|}{\sqrt{m_2^2 + 1}} \tag{4.3}$$

With these two parameters and the distances of the upper and lower boundaries of the region that the vehicle is in at hand, the location of the vehicle can be calculated as in the following equation:

$$Vehicle\ Location = x + \left(\frac{a}{a+b}\right) * (y - x) \quad (4.4)$$

where

x= Distance of region's lower boundary,

y= Distance of region's upper boundary,

a= Vehicle's perpendicular distance to lower boundary,

b= Vehicle's perpendicular distance to lower boundary.

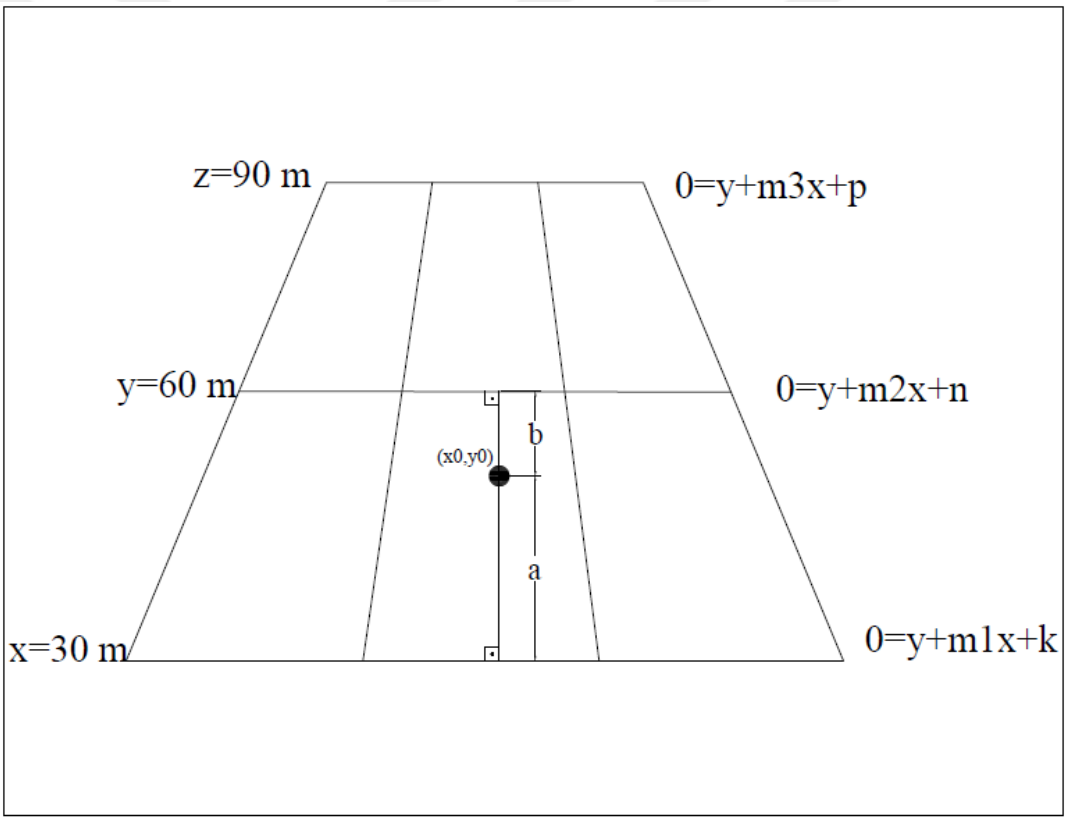


Figure 4.8. Pixels to meters conversion sketch.

The study area is analyzed for 7 days from 10.10.2018 to around 16.10.2018 around noon hours. The dates and the time interval of the footages are shown on Table 4.4. Due to complications during the image processing stage, it wasn't possible to process all the footages. The time intervals of the successfully processed footages which are used in the data analysis are shown in Table 4.5.

Table 4.4. Dates and times of video footages.

| Dates of Footages | Time Interval |
|---|---|
| 10.10.2018 | 11:30-14:00 |
| 11.10.2018 | 11:30-14:00 |
| 12.10.2018 | 11:46-14:00 |
| 13.10.2018 | 11:30-14:00 |
| 14.10.2018 | 11:30-13:00 |
| 15.10.2018 | 11:30-14:00 |
| 16.10.2018 | 11:30-14:00 |

Table 4.5. Dates and times of successfully processed video footages.

| Dates | Time Interval |
|---|---|
| 10.10.2018 | 11:30-13:56 |
| 11.10.2018 | 11:30-12:41 |
| 12.10.2018 | 11:46-14:00 |
| 13.10.2018 | 11:30-12:37 |
| 14.10.2018 | 11:30-13:00 |
| 15.10.2018 | 11:30-12:28 |
| 16.10.2018 | 11:30-14:00 |

Not all of the data from the time intervals given in Table 4.5 were usable because of the change of angle of the camera in certain instances. These instances were detected both with the unavailability of data in the output and in the process of watching the video footages. Thus, these time intervals, which are shown in Table 4.6, need to be excluded from the final data which is used in the analysis.

Table 4.6. The time intervals that are excluded from final analysis.

| Date | Deleted Minutes |
|---|---|
| 10.10.2018 | 23,55,56,57,58 |
| 12.10.2018 | 123 |
| 15.10.2018 | 53,54 |
| 16.10.2018 | 66,67,68,69,70 |

## 4.3. Data Analysis

### 4.3.1. Speed and Flow Calculations

The calculations of flow rate and speed values are conducted on Microsoft Excel. In order to do this, average flow rates and average speed are calculated on regions separately within one minute intervals. With the final data after calculating average speeds and flow rates, the relationship between flow and speed with time and with each other can be derived.

For the calculation of flow rate within a region in a time interval, number of unique vehicles that are present in the region during this time interval are counted. This data alone gives us the volume of traffic inside this region during this time interval. By multiplying this value by 60, the 1-minute volume can be converted into hourly flow rate. The volumes are calculated through a single formula in Excel. In this formula, several checks are done to determine whether the vehicles are satisfying all the conditions. The first check is to see whether the vehicle is in a certain region or not. If it is the next step is to check the timestamp. If it is inside the desired time interval, the vehicle is added to the volume counter for this case. If any of these 2 checks fail or the vehicle was already added to the volume counter before, it won't be taken into account in the volume of this region in this certain time interval.

In calculation of speed; however, the output obtained from image processing algorithms wasn't adequate because originally the coordinates were obtained as pixel coordinates as seen in Table 4.3. The process of converting pixel coordinates to meters is mentioned in the data preparation section of this chapter. A sample data with pixels coordinated to meters is shown in Table 4.7.

Table 4.7. Sample data with pixels converted to distance.

| ID | Frame | CX | CY | REGION | Distance | Time(sec) | Speed(m/s) | Speed(km/hr) |
|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 630,8916 | 590,29 | 4 | 60,9146 | 0,36015 | 16,86149 | 60,70138 |
| 1 | 4 | 639,797 | 582,819 | 4 | 62,93882 | 0,4802 | 17,36242 | 62,5047 |
| 1 | 5 | 648,7384 | 575,137 | 4 | 65,02318 | 0,60025 | 12,38847 | 44,59849 |
| 1 | 6 | 658,1608 | 569,7 | 4 | 66,51042 | 0,7203 | 10,0382 | 36,13752 |
| 1 | 7 | 666,6845 | 565,312 | 4 | 67,7155 | 0,84035 | 16,85831 | 60,68991 |
| 1 | 8 | 675,2241 | 557,883 | 4 | 69,73934 | 0,9604 | 12,9941 | 46,77875 |
| 1 | 9 | 681,664 | 552,58 | 7 | 71,29928 | 1,08045 | 27,26176 | 98,14233 |
| 1 | 10 | 690,4934 | 545,099 | 7 | 74,57206 | 1,2005 | 23,96081 | 86,25893 |
| 1 | 11 | 697,0854 | 538,535 | 7 | 77,44855 | 1,32055 | 25,28185 | 91,01464 |
| 1 | 12 | 704,5913 | 531,649 | 7 | 80,48364 | 1,4406 | 20,17494 | 72,62977 |
| 1 | 13 | 710,7283 | 526,179 | 7 | 82,90564 | 1,56065 | 19,05317 | 68,5914 |
| 1 | 14 | 717,7211 | 521,061 | 7 | 85,19297 | 1,6807 | 22,41941 | 80,70988 |
| 1 | 15 | 725,0393 | 515,043 | 7 | 87,88442 | 1,80075 | 14,85036 | 53,46128 |
| 1 | 16 | 730,424 | 511,086 | 7 | 89,66721 | 1,9208 | 13,76158 | 49,54167 |
| 1 | 17 | 738,1951 | 507,509 | 7 | 91,31929 | 2,04085 | 14,49717 | 52,18981 |
| 1 | 18 | 744,1414 | 503,694 | 7 | 93,05967 | 2,1609 | 12,7015 | 45,72538 |
| 1 | 19 | 750,0313 | 500,384 | 7 | 94,58449 | 2,28095 | 17,90709 | 64,46551 |
| 1 | 20 | 755,2911 | 495,643 | 7 | 96,73423 | 2,401 | 17,74183 | 63,8706 |
| 1 | 21 | 758,9145 | 490,907 | 7 | 98,86414 | 2,52105 | 19,93094 | 71,75138 |
| 1 | 22 | 764,3793 | 485,827 | 10 | 101,2568 | 2,6411 | 20,35786 | 73,28829 |
| 1 | 23 | 769,5103 | 482,145 | 10 | 103,7008 | 2,76115 | 24,46648 | 88,07933 |
| 1 | 24 | 773,4175 | 477,637 | 10 | 106,638 | 2,8812 | 37,01167 | 83,24199 |

After the conversion of pixel coordinates to 1-D distance coordinates, speeds of vehicles can be calculated in metric terms. To calculate the speed of a vehicle, the distance between the vehicle's final position is subtracted from its previous position and the result is then divided into the amount of time passed between these two instances. To summarize, speed is calculated using the following equation:

$$u = \frac{X_{i+1} - X_i}{\Delta t} \tag{4.5}$$

where

$X_{i+1}$= vehicle's last position,

$X_i$= vehicle's previous position,

$\Delta t$= time interval between vehicle's last and previous position.

The speed calculation can be demonstrated from Table 4.9. For example the vehicle with ID=1 is at x=67.7155 m in the 7$^{th}$ frame which corresponds to 0.84035 seconds. In the next frame (frame 8) at 0.9604 seconds, the vehicle is located at x=69.73934 m. With these numbers the speed of the vehicle in this instance can be calculated as:

$$u = \frac{69.73934 - 67.7155}{0.9604 - 0.84035} = 16.8583 \text{ m/s}$$

, which is the result also found in Table 4.7. This result can then be converted into km/hr by multiplying it with 3.6 which makes this speed value found above equal to 60.69 km/hr.

To calculate the average speed within a region during a time interval, the speed values of vehicles that are in the certain region in the certain time interval are summed using SUMIFS function in Excel. In this formula, it is first checked if the speed is bigger than zero or not. In some instances, the speeds turned out below zero due to conversion of pixels to meters not being 100% accurate. Also because the speed calculation is automated in Excel, the speed is designated as 0 when the speed is calculated between the data from two vehicles to eliminate these cases. The next check is to eliminate speed values bigger than 120 km/hr as speed values larger than 120 km/hr are most likely due to an error in estimation of speed. The next check is to see if the vehicle is in the wanted time interval and the final check is done to see if the vehicle is in a certain region. After all the speed values which satisfy these conditions are summed up, the aggregate speed value which is used in the calculation of the average speed for a certain region and for a certain 1 minute time interval is determined.

The next step in calculating the average speed in a region is to calculate the number of cases of speeds. It is done using the COUNTIFS function in Excel. The checks in this formula are done to see if the speed is recorded in a certain time interval and if the vehicle is in a certain region and if the speed is between 0 and 120. If all these checks satisfy the count increases by one.

After both summing the speeds and counting the cases where the speeds are summed, the average speed in a region within a time interval can be calculated using the equation:

$$Average\ Speed = \frac{Aggregate\ Speed}{Number\ of\ Satisfying\ Cases} \tag{4.6}$$

Note that this average speed is equivalent to the time mean speed.

With this methodology, the average speed and flow rates are calculated for every minute and every region for the time intervals and dates given in Table 4.5. With the use of these results, the speed and flow characteristics of the study area can be interpreted. Table 4.8 shows the output of the speed and flow readings in regions 0-8 for the time interval 11:30-12:00 in 11.10.2018 via the use of the methods presented above. As it is seen from this table, regions 0, 3 and 6 have the lowest readings for almost all minutes. This is not a surprising result as right lanes tend to have the lowest demands and lowest speeds in uncongested traffic conditions. Regions 1, 4, 7, which are on the middle lane and regions 2, 5, 8, which are on the right lane, have similar flow rates but regions 2, 5 and 8 generally have higher flow rates compared to regions 1, 4 and 7. Other meaningful interpretations regarding this data are done with visuals in the Results section of this chapter.

Table 4.8. Speed-flow data for 11.10.2018 (11:30-12:00).

| | Region 0 | | Region 1 | | Region 2 | | Region 3 | | Region 4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Minutes | Flow | Speed | Flow | Speed | Flow | Speed | Flow | Speed | Flow | Speed |
| 0 | 480 | 80 | 1329 | 89 | 1170 | 91 | 300 | 99 | 1260 | 90 |
| 1 | 540 | 76 | 1595 | 81 | 2340 | 88 | 540 | 83 | 1380 | 86 |
| 2 | 480 | 78 | 1861 | 85 | 2172 | 84 | 540 | 75 | 1920 | 86 |
| 3 | 720 | 88 | 1329 | 85 | 1671 | 84 | 780 | 90 | 1320 | 85 |
| 4 | 480 | 85 | 1595 | 83 | 1671 | 86 | 480 | 94 | 1620 | 84 |
| 5 | 600 | 82 | 1529 | 84 | 1671 | 84 | 660 | 85 | 1800 | 87 |
| 6 | 600 | 82 | 1595 | 88 | 1588 | 83 | 600 | 81 | 1620 | 87 |
| 7 | 660 | 81 | 1595 | 85 | 1504 | 90 | 720 | 82 | 1560 | 82 |
| 8 | 480 | 84 | 1396 | 89 | 919 | 93 | 540 | 90 | 1380 | 88 |
| 9 | 420 | 85 | 1329 | 89 | 1755 | 87 | 540 | 92 | 1440 | 91 |
| 10 | 240 | 83 | 1329 | 90 | 1755 | 89 | 300 | 94 | 1380 | 89 |
| 11 | 360 | 80 | 1329 | 85 | 1337 | 93 | 360 | 85 | 1380 | 84 |
| 12 | 420 | 90 | 1529 | 87 | 1504 | 95 | 420 | 93 | 1740 | 89 |
| 13 | 600 | 85 | 1396 | 87 | 1003 | 90 | 480 | 84 | 1380 | 92 |
| 14 | 480 | 86 | 1595 | 89 | 1838 | 93 | 360 | 88 | 1560 | 91 |
| 15 | 600 | 83 | 1861 | 86 | 836 | 95 | 480 | 85 | 1680 | 86 |
| 16 | 900 | 85 | 1994 | 82 | 2256 | 85 | 1020 | 89 | 1920 | 83 |
| 17 | 720 | 80 | 1462 | 85 | 1755 | 87 | 660 | 87 | 1380 | 86 |
| 18 | 660 | 88 | 1795 | 88 | 2089 | 86 | 720 | 92 | 1920 | 89 |
| 19 | 660 | 81 | 1529 | 88 | 1922 | 92 | 720 | 82 | 1260 | 91 |
| 20 | 780 | 85 | 1861 | 85 | 2340 | 86 | 720 | 88 | 1800 | 84 |
| 21 | 900 | 78 | 2061 | 85 | 2507 | 87 | 1020 | 83 | 2160 | 84 |
| 22 | 660 | 85 | 1396 | 87 | 1588 | 89 | 720 | 89 | 1440 | 87 |
| 23 | 960 | 82 | 2061 | 83 | 3092 | 86 | 1140 | 87 | 1920 | 83 |
| 24 | 660 | 79 | 1662 | 88 | 1922 | 92 | 600 | 83 | 1560 | 88 |
| 25 | 180 | 89 | 1064 | 90 | 1337 | 90 | 300 | 94 | 1200 | 90 |
| 26 | 720 | 79 | 1928 | 87 | 1337 | 93 | 720 | 85 | 1800 | 88 |
| 27 | 540 | 83 | 1728 | 87 | 3092 | 88 | 600 | 80 | 1860 | 87 |
| 28 | 600 | 89 | 1462 | 86 | 2089 | 91 | 720 | 89 | 1380 | 86 |
| 29 | 600 | 76 | 1662 | 82 | 1671 | 88 | 660 | 78 | 1620 | 82 |
| 30 | 480 | 82 | 1928 | 87 | 1922 | 88 | 600 | 81 | 1920 | 88 |

Table 4.8. Speed-flow data for 11.10.2018 (11:30-12:00) (cont.).

| | Region 5 | | Region 6 | | Region 7 | | Region 8 | |
|---|---|---|---|---|---|---|---|---|
| Minutes | Flow | Speed | Flow | Speed | Flow | Speed | Flow | Speed |
| 0 | 1680 | 92 | 120 | 83 | 1200 | 89 | 1755 | 94 |
| 1 | 1800 | 92 | 420 | 69 | 1140 | 83 | 1865 | 89 |
| 2 | 1920 | 88 | 360 | 67 | 1980 | 82 | 1975 | 83 |
| 3 | 1860 | 87 | 480 | 80 | 1200 | 84 | 2029 | 86 |
| 4 | 1800 | 88 | 300 | 85 | 1200 | 82 | 1755 | 87 |
| 5 | 1500 | 91 | 540 | 83 | 1800 | 84 | 1536 | 85 |
| 6 | 1500 | 92 | 540 | 77 | 1680 | 86 | 1700 | 91 |
| 7 | 1380 | 93 | 720 | 79 | 1860 | 80 | 1810 | 84 |
| 8 | 1140 | 94 | 360 | 85 | 1500 | 86 | 1316 | 92 |
| 9 | 1560 | 91 | 420 | 89 | 1380 | 87 | 1591 | 88 |
| 10 | 1560 | 94 | 180 | 94 | 1320 | 86 | 1810 | 92 |
| 11 | 1740 | 92 | 240 | 81 | 1260 | 82 | 1920 | 89 |
| 12 | 1560 | 95 | 300 | 88 | 1560 | 88 | 1865 | 91 |
| 13 | 1320 | 93 | 420 | 76 | 1440 | 85 | 1262 | 91 |
| 14 | 1740 | 91 | 300 | 88 | 1440 | 92 | 1920 | 89 |
| 15 | 1560 | 97 | 480 | 84 | 1560 | 84 | 1755 | 93 |
| 16 | 2040 | 87 | 1080 | 86 | 1860 | 82 | 1755 | 85 |
| 17 | 1680 | 89 | 660 | 81 | 1560 | 84 | 1810 | 87 |
| 18 | 2280 | 85 | 840 | 86 | 1800 | 87 | 1920 | 85 |
| 19 | 1680 | 94 | 600 | 81 | 1500 | 85 | 1591 | 92 |
| 20 | 1860 | 91 | 600 | 80 | 1740 | 82 | 1810 | 88 |
| 21 | 2220 | 88 | 1080 | 78 | 2040 | 82 | 2359 | 84 |
| 22 | 1500 | 91 | 660 | 81 | 1500 | 86 | 1591 | 89 |
| 23 | 2520 | 88 | 1020 | 80 | 1980 | 81 | 2359 | 83 |
| 24 | 1740 | 93 | 480 | 71 | 1860 | 85 | 1810 | 89 |
| 25 | 1560 | 95 | 120 | 81 | 1200 | 87 | 1591 | 94 |
| 26 | 1500 | 93 | 600 | 78 | 1980 | 85 | 1646 | 90 |
| 27 | 2520 | 89 | 420 | 76 | 1920 | 82 | 2304 | 87 |
| 28 | 1800 | 93 | 660 | 89 | 1380 | 86 | 1810 | 91 |
| 29 | 1740 | 89 | 600 | 74 | 1680 | 81 | 1865 | 91 |
| 30 | 1980 | 90 | 420 | 80 | 1920 | 87 | 2084 | 91 |

### 4.3.2. Regression Analysis

One of the goals of this thesis is to obtain regression models with the help of traffic parameters. In creation of these regression models, IBM SPSS is used. Because the number of variables is high and it is not apparent which is related to which, exploratory factor analysis is used to reduce the number of variables and come up with meaningful relationships. After the factors are created, linear regression analysis can be performed and the regression model can be obtained. In this section, the regression procedure is explained with the use of flow and speed data from 16.10.2018.

Figure 4.9 shows the initial setup of the factor analysis. As it can be seen from this figure, all the speed and flow parameters are entered into the factor analysis. The factor analysis type that is used in this analysis is the principal component analysis. Thus, as it is seen from Figure 4.10, principal component analysis is selected in the extraction tab. One other useful tool in SPSS in doing the factor analysis is the rotation. Rotation helps the factor loadings to be redistributed. This helps the user to obtain more meaningful models. The type of rotation that is used here is varimax rotation as it is seen from Figure 4.11. It is short for variable maximization. With varimax rotation it is possible for each variable to measure one factor which is ideal for our purposes.

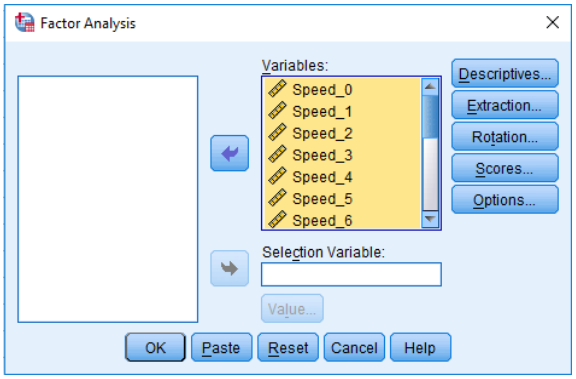


Figure 4.9. Factor analysis setup (SPSS).

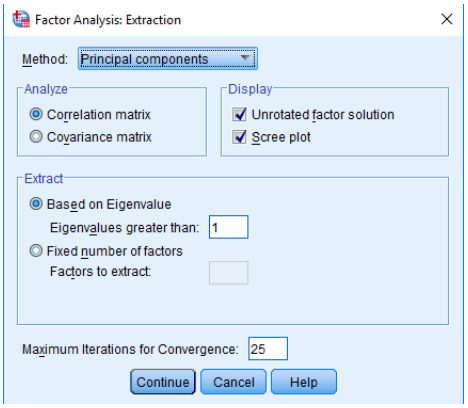


Figure 4.10. Factor analysis-extraction (SPSS).

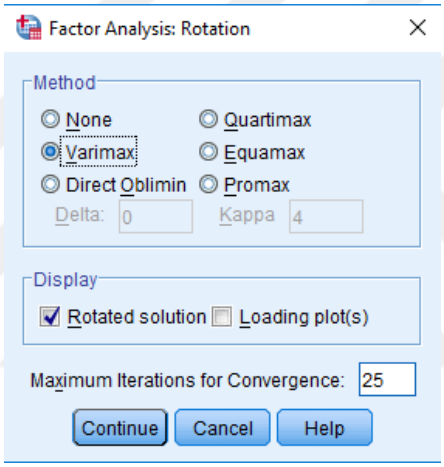


Figure 4.11. Factor analysis-rotation (SPSS).

After the factor analysis is run, the results are interpreted and the factors that will be used in the regression analysis are determined.

**KMO and Bartlett's Test**

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | ,857 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 3092,288 |
| | df | 153 |
| | Sig. | ,000 |

Figure 4.12. KMO and Bartlett's test results (SPSS).

Figure 4.12 shows the KMO and Bartlett's test results. Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy measures how much the used data is suited for factor

analysis. Values closer to 1 are preferred. In this case, a KMO value of 0,857 is observed which means that the data used is fit to be used in a factor analysis. Bartlett's test of sphericity is used to test the null hypothesis that the correlation matrix is equivalent to the identity matrix. This null hypothesis needs to be rejected for a healthy factor analysis which is the case here because the significance value is 0.000.

Figure 4.13 shows the total variance explained output from SPSS. Here, the components that are significant and will be used in the model are the ones having initial eigenvalues bigger than 1. There are 5 components which satisfy this and only these five components are included in the later stages of the analysis as seen from extraction sums and rotation sums of squared loadings tabs in this table. Component 1 is the most effective component as 47.1% of the variance is explained by it. The amount of total variance explained decreases as you go from the first component to the last component. The fifth and last component that is included in the final model explains 6,8% of the total variance. The scree plot given in Figure 4.14 visualizes the eigenvalues of principal components.

**Total Variance Explained**

| | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| Component | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 8,474 | 47,079 | 47,079 | 8,474 | 47,079 | 47,079 | 4,843 | 26,906 | 26,906 |
| 2 | 2,937 | 16,314 | 63,393 | 2,937 | 16,314 | 63,393 | 2,942 | 16,344 | 43,250 |
| 3 | 1,661 | 9,226 | 72,619 | 1,661 | 9,226 | 72,619 | 2,805 | 15,585 | 58,835 |
| 4 | 1,331 | 7,393 | 80,012 | 1,331 | 7,393 | 80,012 | 2,607 | 14,481 | 73,316 |
| 5 | 1,233 | 6,848 | 86,860 | 1,233 | 6,848 | 86,860 | 2,438 | 13,543 | 86,860 |
| 6 | ,693 | 3,850 | 90,710 | | | | | | |
| 7 | ,343 | 1,907 | 92,617 | | | | | | |
| 8 | ,279 | 1,552 | 94,169 | | | | | | |
| 9 | ,191 | 1,063 | 95,233 | | | | | | |
| 10 | ,157 | ,872 | 96,105 | | | | | | |
| 11 | ,148 | ,820 | 96,925 | | | | | | |
| 12 | ,127 | ,706 | 97,631 | | | | | | |
| 13 | ,121 | ,673 | 98,304 | | | | | | |
| 14 | ,093 | ,518 | 98,822 | | | | | | |
| 15 | ,090 | ,498 | 99,320 | | | | | | |
| 16 | ,049 | ,275 | 99,595 | | | | | | |
| 17 | ,044 | ,245 | 99,840 | | | | | | |
| 18 | ,029 | ,160 | 100,000 | | | | | | |

Extraction Method: Principal Component Analysis.

Figure 4.13. Explained total variances of the components (SPSS).

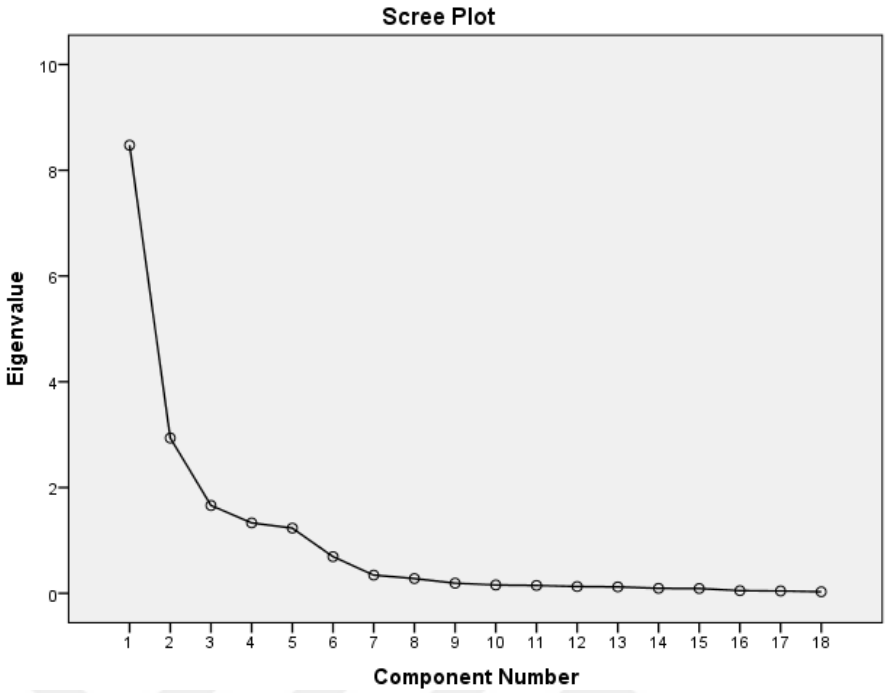


Figure 4.14. Scree plot of the components (SPSS).

Figure 4.15 shows the rotated component matrix, which are used in the interpretation of the components. The numbers here give the correlations of variables where a higher value is preferred. In Table 4.9, the interpretations of these rotated components are given.

**Rotated Component Matrix[a]**

| | Component | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Speed_5 | ,844 | | | | |
| Speed_8 | ,840 | | | | |
| Speed_4 | ,833 | | | | |
| Speed_7 | ,823 | ,372 | | | |
| Speed_1 | ,803 | ,306 | | | |
| Speed_2 | ,746 | | | | |
| Speed_3 | | ,935 | | | |
| Speed_0 | | ,909 | | | |
| Speed_6 | ,305 | ,908 | | | |
| Flow_3 | | | ,918 | | |
| Flow_6 | | | ,892 | | |
| Flow_0 | | | ,890 | | |
| Flow_4 | | | | ,895 | |
| Flow_7 | | | | ,867 | |
| Flow_1 | -,305 | | | ,829 | |
| Flow_8 | | | | | ,898 |
| Flow_5 | | | | | ,894 |
| Flow_2 | -,531 | | | | ,685 |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 6 iterations.

Figure 4.15. Rotated component matrix (SPSS).

Table 4.9. Components to be used in the regression model.

| Components | Component Name |
|---|---|
| 1 | Speed in left and middle lanes |
| 2 | Speed in the right lane |
| 3 | Flow rate in the right lane |
| 4 | Flow rate in the middle lane |
| 5 | Flow rate in the left lane |

With these 5 components, 5 factors can be created. In creation of these factors, the mean values of all the variables inside a component are computed. These 5 factors are used in the regression analysis. With these factors, multiple models can be created. In this section, a single model that assumes flow rate in the middle lane as the dependent variable and all others as the independent variables is created. The setup of this linear regression is shown in Figure 4.16.

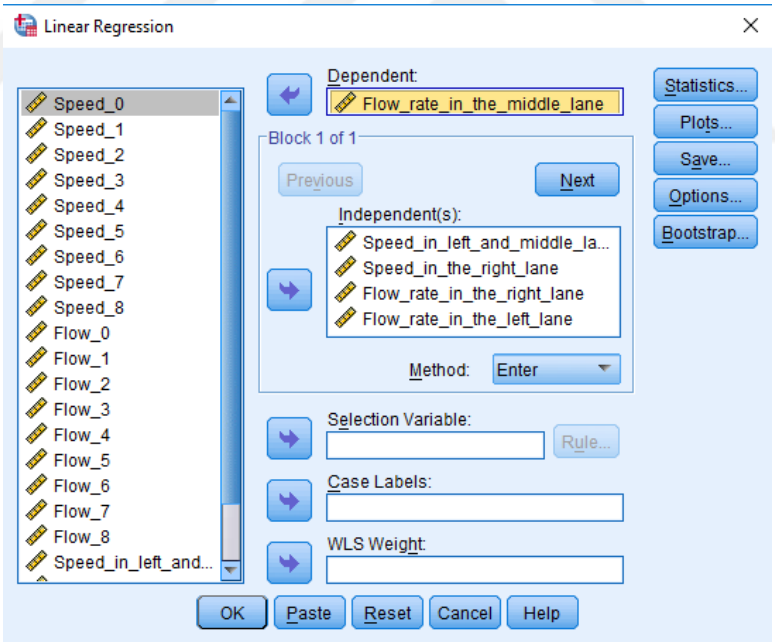


Figure 4.16. Setup of the linear regression (SPSS).

Figures 4.17, 4.18 and 4.19 show the Model Summary, ANOVA and Coefficients tables of the initial model respectively. As it is seen in Figure 4.17, $R^2$ value is equal to 0,315 which is an ok value. The model is significant with an F value of 16,114 and a significance value of 0,000 as seen in Figure 4.18. In Figure 4.19, the coefficients and the significance of independent variables are shown. A significance value of 0,05 is accepted

as the threshold meaning that variables with significance value lower than 0,05 are significant. It is observed in Figure 4.19 that only Speed_in_left_and_middle_lanes and Flow_rate_in_the_left_lane are significant. Thus, these two variables are used in the final model.

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,561[a] | ,315 | ,296 | 198,37333 |

a. Predictors: (Constant), Flow_rate_in_the_left_lane, Speed_in_the_right_lane, Flow_rate_in_the_right_lane, Speed_in_left_and_middle_lanes

Figure 4.17. Model summary of the initial model (SPSS).

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 2536537,641 | 4 | 634134,410 | 16,114 | ,000[b] |
| | Residual | 5509276,769 | 140 | 39351,977 | | |
| | Total | 8045814,410 | 144 | | | |

a. Dependent Variable: Flow_rate_in_the_middle_lane

b. Predictors: (Constant), Flow_rate_in_the_left_lane, Speed_in_the_right_lane, Flow_rate_in_the_right_lane, Speed_in_left_and_middle_lanes

Figure 4.18. Anova of the initial model (SPSS).

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | | Correlations | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 2161,490 | 523,442 | | 4,129 | ,000 | | | | | |
| | Speed_in_left_and_middle_lanes | -13,646 | 6,114 | -,228 | -2,232 | ,027 | -,454 | -,185 | -,156 | ,467 | 2,141 |
| | Speed_in_the_right_lane | 2,266 | 3,810 | ,049 | ,595 | ,553 | -,168 | ,050 | ,042 | ,713 | 1,403 |
| | Flow_rate_in_the_right_lane | ,157 | ,111 | ,117 | 1,411 | ,160 | ,378 | ,118 | ,099 | ,712 | 1,405 |
| | Flow_rate_in_the_left_lane | ,205 | ,054 | ,340 | 3,768 | ,000 | ,517 | ,303 | ,264 | ,602 | 1,660 |

a. Dependent Variable: Flow_rate_in_the_middle_lane

Figure 4.19. Coefficients of the initial model (SPSS).

Figure 4.20, 4.21 and 4.22 show the Model Summary, ANOVA and Coefficients results of the final model respectively. The model has an $R^2$ value of 0,303 as seen in Figure 4.20 and is significant with an F value of 30,928 and a significance value of 0,000 as seen in Figure 4.21. In Figure 4.22, it is observed that both the independent variables are significant in explaining the dependent variable with significance values 0,008 and 0,000.

The constant is also significant with a significance value of 0,000. Thus with the coefficients given in Figure 4.22, the regression model equation can be written as:

$$\text{Flow in the middle lane} = 2422{,}599 + 0{,}231 * \text{Flow in the left lane} - 13{,}907 * \text{Speed in left and middle lanes} \quad (4.7)$$

After Equation 4.7 is examined, it can be said that, on 16.10.2018, the flow rate in the middle lane is positively correlated with the flow rate in the left lane and negatively correlated with the speed in left and middle lanes. The positive correlation with flow in the left lane makes sense because in the case of a flow increase in the left lane, you would expect more vehicles to use the middle lane both to make way for the incoming vehicles on the left lane and to divert from the left lane if it gets too crowded.

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,551[a] | ,303 | ,294 | 198,66557 |

a. Predictors: (Constant), Speed_in_left_and_middle_lanes, Flow_rate_in_the_left_lane

Figure 4.20. Model summary of the final model (SPSS).

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 2441357,117 | 2 | 1220678,559 | 30,928 | ,000[b] |
| | Residual | 5604457,293 | 142 | 39468,009 | | |
| | Total | 8045814,410 | 144 | | | |

a. Dependent Variable: Flow_rate_in_the_middle_lane

b. Predictors: (Constant), Speed_in_left_and_middle_lanes, Flow_rate_in_the_left_lane

Figure 4.21. Anova of the final model (SPSS).

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 2422,599 | 495,650 | | 4,888 | ,000 | | | | | |
| | Flow_rate_in_the_left_lane | ,231 | ,052 | ,382 | 4,451 | ,000 | ,517 | ,350 | ,312 | ,665 | 1,504 |
| | Speed_in_left_and_middle_lanes | -13,907 | 5,132 | -,233 | -2,710 | ,008 | -,454 | -,222 | -,190 | ,665 | 1,504 |

a. Dependent Variable: Flow_rate_in_the_middle_lane

Figure 4.22. Coefficients of the final model (SPSS).

### 4.3.3. Complications Regarding Obtaining Sound Data

There were some complications with the data. Due to the lack of clear visuals in the further regions, starting with regions 18, 19, 20, vehicle detection proved to be problematic. Therefore the original study area had to be shortened to only include regions up to regions 15, 16 and 17.

Another problem is the inaccurate estimation of a vehicle's region in further regions starting from regions 9, 10 and 11. This problem is caused by the angle of the camera with respected to the main arterial. As you get further from the camera, it becomes harder to determine which lane that a vehicle is in. Because of this problem a vehicle on the middle lane may show up on the left lane. This problem was detected with consistent abnormally high flow rates on the left lane and the individual data that indicates a vehicle has changed its lane but in fact didn't. To solve this problem, correction factors are introduced for these problematic regions. Also the original flow rate and speed results for some other regions such as region 2 didn't turn out as well as expected due to issues regarding detection. Correction factors are introduced for these regions as well. For the estimation of the correction factors, the flow rates of the regions on the main arterial are manually noted down by watching a total of 30 minutes of footages. Then the flow rates found this way are compared to the flow rates originally found by the image processing algorithms for the same time intervals. With this comparison, it is possible to deduct how far the estimated flow rates are off from the actual flow rates.

## 4.4. Results

### 4.4.1. Speed-Flow Results

Speed-flow relationship in this study area is an essential part of this thesis. In this section, speed and flow data from 10.10.2018 and 16.10.2018 are investigated with graphs on three regions which are region 3, region 4 and region 5. Regions 3, 4 and 5 are located right next to each other as it is seen from Figure 4.5 and Figure 4.6. 16.10.2018 had free flow conditions for the entirety of the footages whereas both congested flow and free flow conditions were observed on the footages for 10.10.2018. Greenshields speed-flow

diagram is fitted on top of the speed flow graphs to generalize the traffic patterns observed on these regions and to obtain some useful parameters such as optimum speed, free flow speed and capacity. Then, speed and flow values in these regions are compared with respect to time and meaningful interpretations are made between them. In preparation of all the graphs and the fitted curves, Microsoft Excel is used.

Figure 4.23 shows the speed-flow graph of region 4 for 16.10.2018. Region 4 is located on the middle lane in the early section of the study area. As it can be seen from Figure 4.23, free flow conditions are prevalent in region 4, just as it is for all the regions, during the entirety of the study period (11:30-14:00) in 16.10.2018 which in turn means that no traffic congestion has been experienced. The average speed is around 85 km/hr and the maximum flow rate is found as 2220 vehicles/hour (veh/hr).

Figure 4.24 shows the speed-flow graph of the same region as in Figure 4.23 (region 4) on 10.10.2018 within 11:30 and 13:56. It is clear that the pattern of this speed-flow graph is quite different than the one shown in Figure 4.23. The most important difference is that the graph in Figure 4.24 contains both the congested region and the uncongested region. Speed values drop down to around 10 km/hr in the congested region. A maximum flow rate of 2280 veh/hr is observed in this region during this time interval.

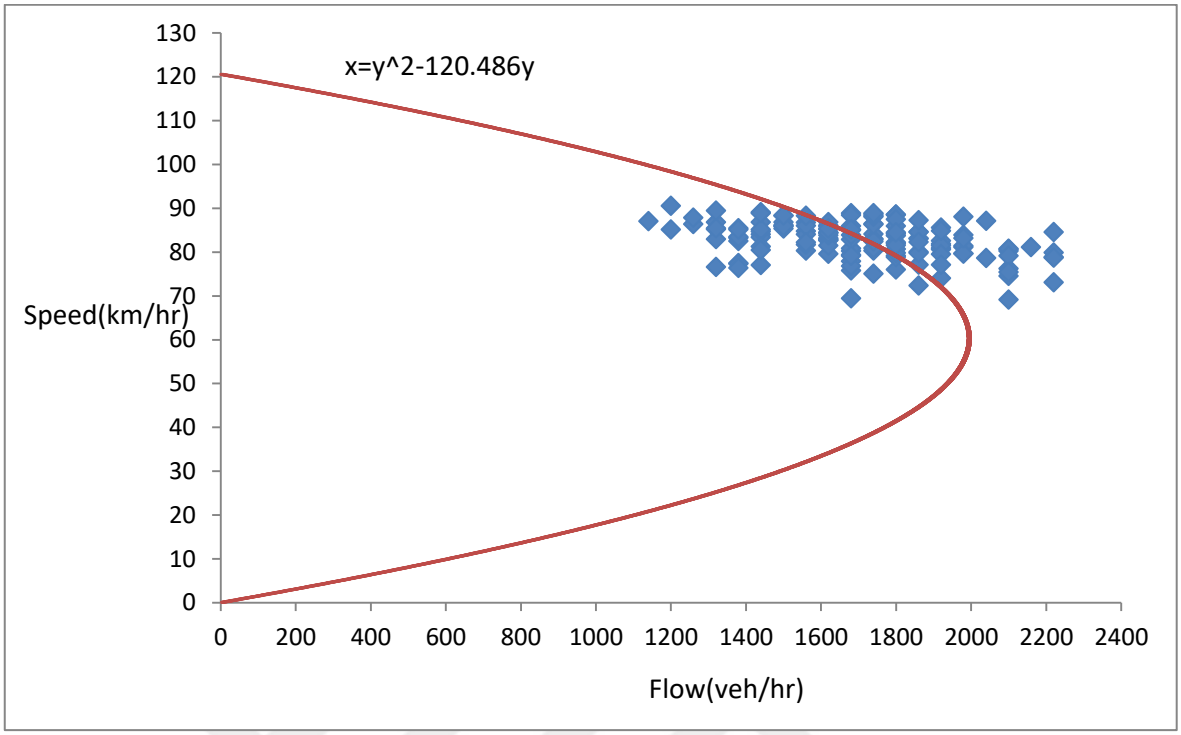


Figure 4.23. Speed-flow graph of region 4 (16.10.2018).

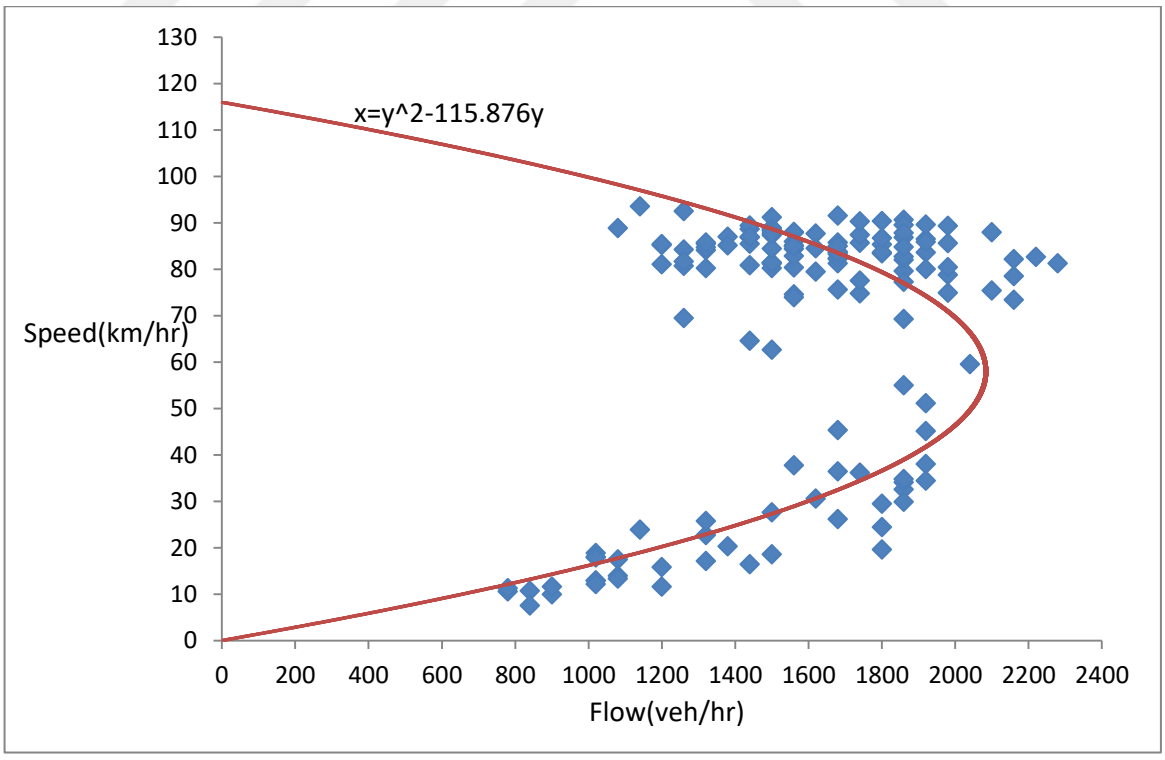


Figure 4.24. Speed-flow graph of region 4 (10.10.2018).

Figure 4.25 shows the speed-flow graph of region 5 from 16.10.2018 within 11:30 and 14:00. This region is located on the left lane, right to the left of region 4. In this figure, it is observed that the results are quite similar to the results obtained from region 4 in the same time interval, which are shown in Figure 4.23. Both regions experience uncongested conditions near capacity. However, it can be said that the flow rates reached in region 5 are higher than the flow rates in region 4. A maximum flow rate of 2600 veh/hr is reached in this region during this time interval.

In Figure 4.26, the speed-flow graph of region 5 on 10.10.2018 is shown. The maximum flow rate observed here is 2640 veh/hr which is higher than the maximum flow rate in region 4 (2280 veh/hr) on the same time interval. The presence of congested flow on this region for this time interval can easily interpreted by the very low speed values observed on Figure 4.26.

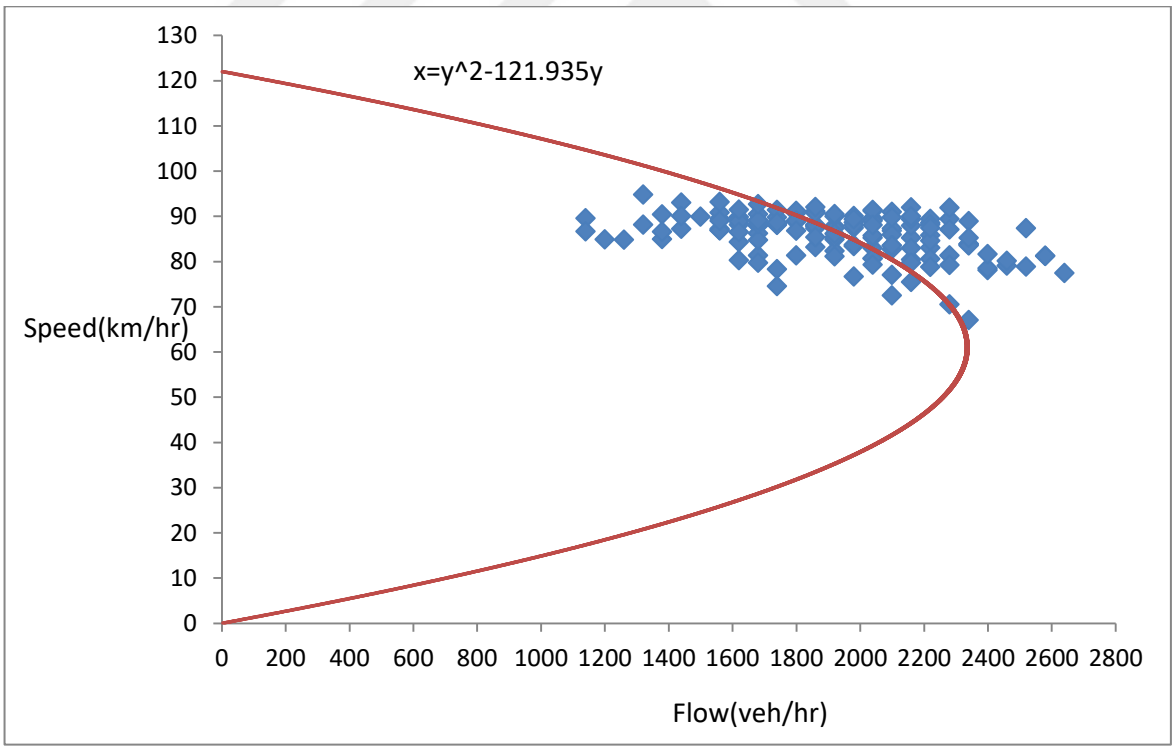


Figure 4.25. Speed-flow graph of region 5 (16.10.2018).

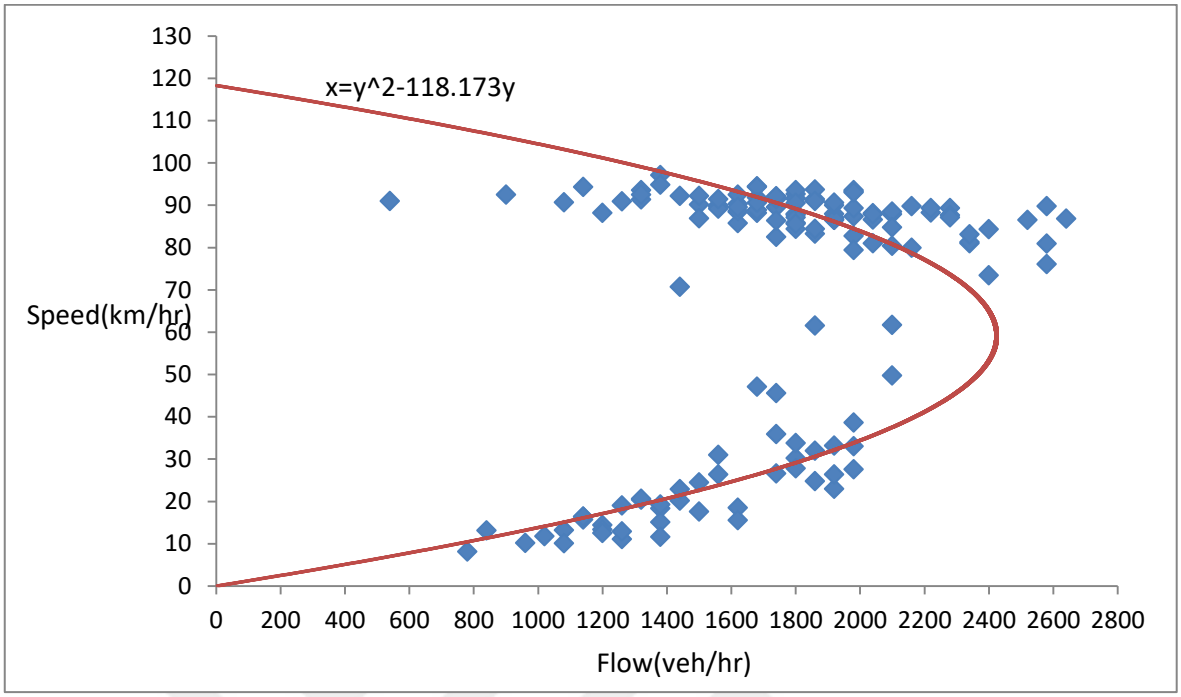


Figure 4.26. Speed-Flow graph of region 5 (10.10.2018).

In Figure 4.27, the speed-flow graph of region 3, which is located at the right side of region 4, on 16.10.2018 is shown. It is seen in this figure that free flow conditions are dominant in region 3 on 16.10.2018, just like regions 4 and 5. However, the flow rates observed are noticeably smaller compared to regions 4 and 5. Some potential reasons for this result are:

- Drivers' general tendency to not use right lanes on freeways and multilane highways.
- The diversion effect from the upcoming on-ramps that scare away drivers from using the right lanes.

The average speed values are around 80 km/hr for all flow rates observed. This is a lower value compared to the other 2 regions which have average speeds around 85 km/hr.

In Figure 4.28, the speed-flow graph of region 3 on 10.10.2018 is shown. Just like regions 4 and 5, region 3 is affected by the congestion during some portion of the video footages. Thus, the speed-flow graph is divided into uncongested and congested regions with the black line like the other two regions. The flow rates in the congested part of region

3 is close to the flow rates in the congested parts of the other two regions, much more so than what is observed in the uncongested parts. This can be explained by the uniformity of traffic flow in the case of congestion.

In Table 4.10, optimum speed, free flow speed and capacity findings with respect to the Greenshields equation fits on the speed-flow graphs are shown. As it is seen from Figure 4.27 and Table 4.10, this process wasn't performed for region 3 on 16.10.2018. The reason for this is the poor fit due to the spread of the data. Region 4 and region 5 have very similar optimum speed (around 60 km/hr) and free flow speed (around 120 km/hr) with region 5 having slightly higher values. The optimal speed and free flow speed of Region 3 are found as 49 km/hr and 98 km/hr which are 10 to 20 km/hr lower than their counterparts in region 4 and region 5. On 10.10.2018, the capacity is found as 1421 veh/hr for region 3, 2080 veh/hr for region 4 and 2420 veh/hr for region 5 and on 16.10.2018 capacity is 1991 veh/hr for region 4 and 2331 veh/hr for region 5. With these results, it can be said that region 5 has the biggest theoretical traffic capacity among the three regions and region 3 has the least capacity among these regions.

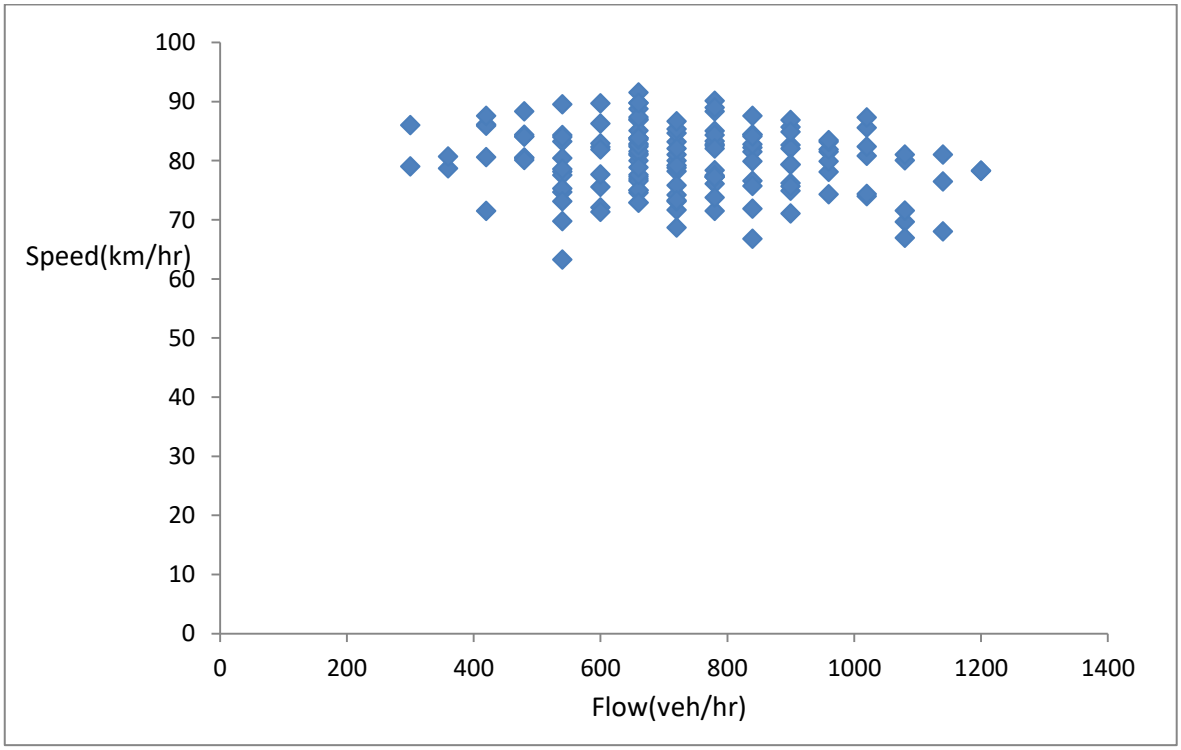


Figure 4.27. Speed-flow graph of region 3 (16.10.2018).

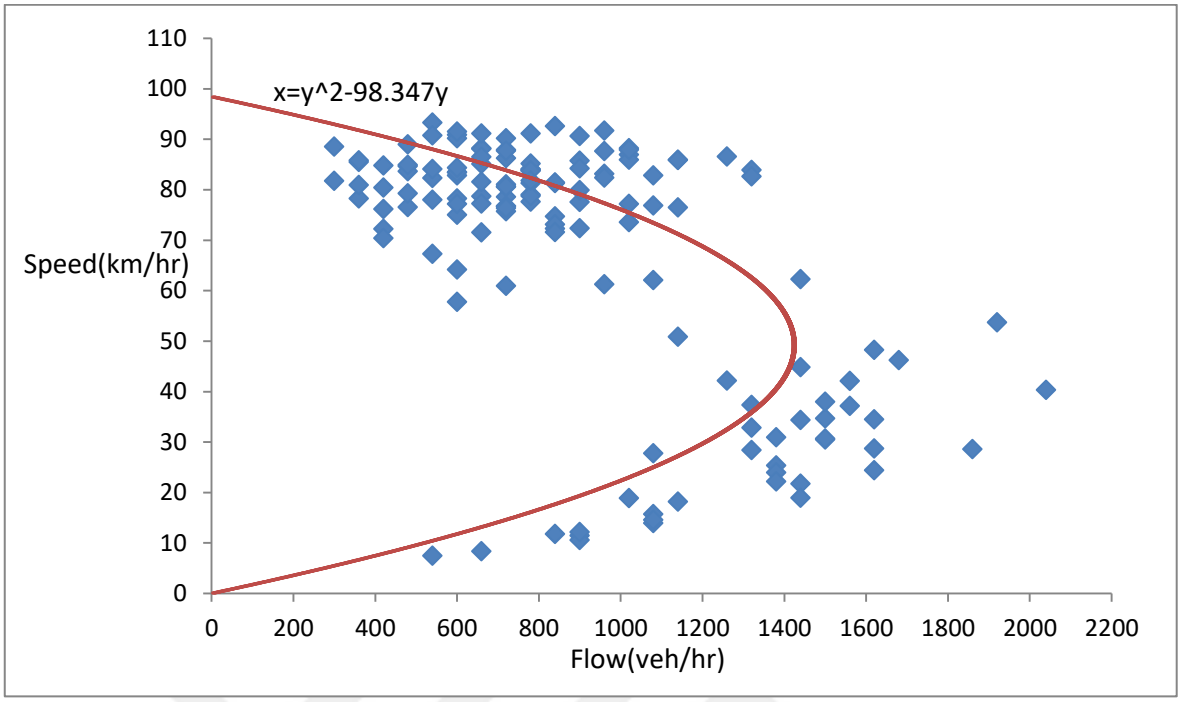


Figure 4.28. Speed-flow graph of region 3(10.10.2018).

Table 4.10. Some traffic parameters found on region 3, region 4 and region 5 according to Greenshields equation.

| Traffic Parameters | 10.10.2018 | | | 16.10.2018 | | |
|---|---|---|---|---|---|---|
| | Region 3 | Region 4 | Region 5 | Region 3 | Region 4 | Region 5 |
| Optimum Speed | 49,175 | 57,95 | 59,075 | NA | 60,25 | 60,975 |
| Free Flow Speed | 98,347 | 115,876 | 118,173 | NA | 120,486 | 121,935 |
| Capacity | 1421,1011 | 2079,8939 | 2420,138 | NA | 1991,3725 | 2331,347 |

The speed and flow rates of regions 3, 4, 5 on 16.10.2018 and 10.10.2018 are compared with respect to time on Figures 4.29, 4.30 , 4.31 and 4.32.

In Figure 4.29, the flowrates of regions 3, 4 and 5 are compared with data from 16.10.2018. region 5 seems to be the region with the highest flowrate for almost every minute and it fluctuates between 1200 and 2500 veh/hr. Region 4 has similar but generally lower flowrates than region 5 with some visual correlation between the two regions. Region 3, which is from the right lane, has the smallest flowrates for the entire study period on 16.10.2018 and it fluctuates between 400 and 1200 veh/hr. In Figure 4.30, the speed values of regions 3, 4 and 5 are compared with data from 16.10.2018. Region 5, the left lane, has the highest speed values for almost the entire duration of footages from this

day. Region 3 generally has the smallest speeds everywhere with a few exceptions with region 4 being in between region 3 and region 5. The average speed values fluctuate between 70 km/hr and 90 km/hr for all regions with region 3 dipping more than the others.

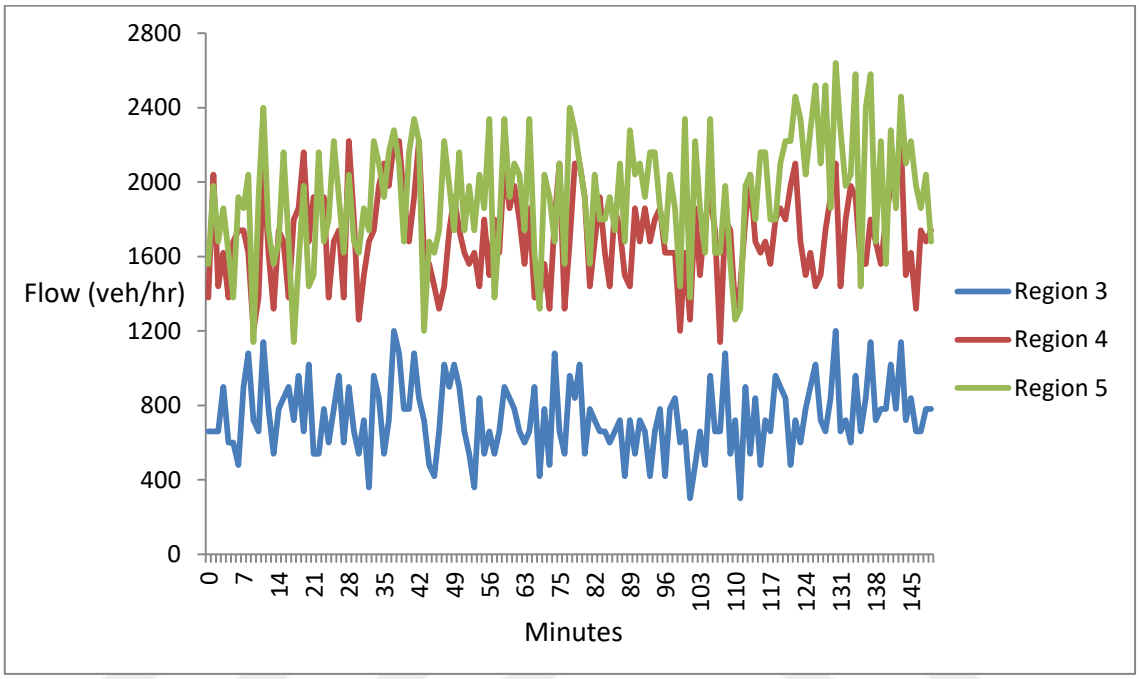


Figure 4.29. Comparison of flow rates of regions 3,4 and 5 (16.10.2018).

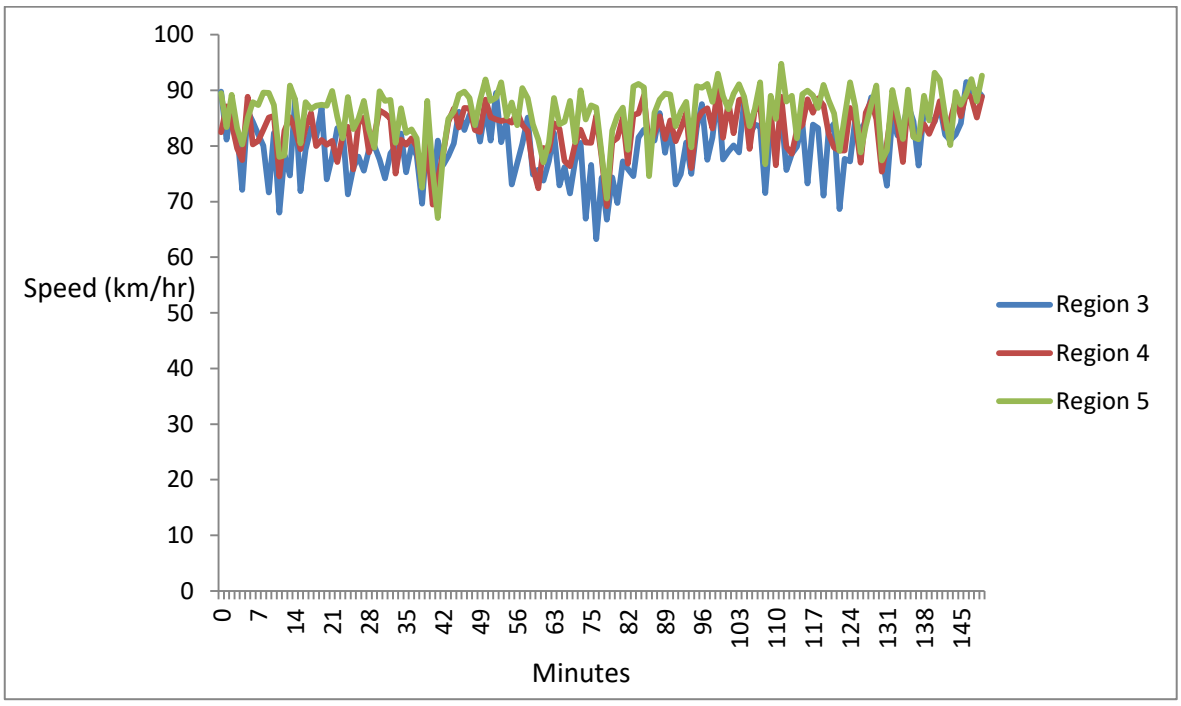


Figure 4.30. Comparison of speeds of regions 3,4 and 5 (16.10.2018).

In Figure 4.31, flowrate comparison of regions 3,4 and 5 are visualized with the data from 10.10.2018. The results here are more interesting than the results found in Figure 4.29. For the first 60 minutes of the footages, regions 4 and 5 have higher flowrates which are between 1000 and 2500 veh/hr compared to region 3 which generally have flowrates between 500 and 1000 veh/hr. But, with the introduction of congestion to the freeway, the flowrates are equalized in the middle of the footages for around 40 minutes as they fluctuate between 500 and 2000 veh/hr. Then, with transition from congested flow to uncongested flow, the flow rates return to their original pattern with region 5 having the highest, region 4 having the second highest and region 3 having the smallest.

In Figure 4.32, the comparison of speed values of regions 3,4 and 5 are visualized with the data from 10.10.2018. Just as in Figure 4.31, the effect of congestion is quite apparent. Until around 60th minute of the video footages, region 5 generally has the highest speeds with region 4 having second highest and region 3 having lowest speeds. At the 60-minute mark, the average speeds of all regions decrease significantly down to 20 km/hr. Then, the speed values remain low with fairly similar patterns in all the regions until the congestion is cleared around the 100 minute mark. After this point, speed values follow similar patterns to the first 60 minutes of the footages.

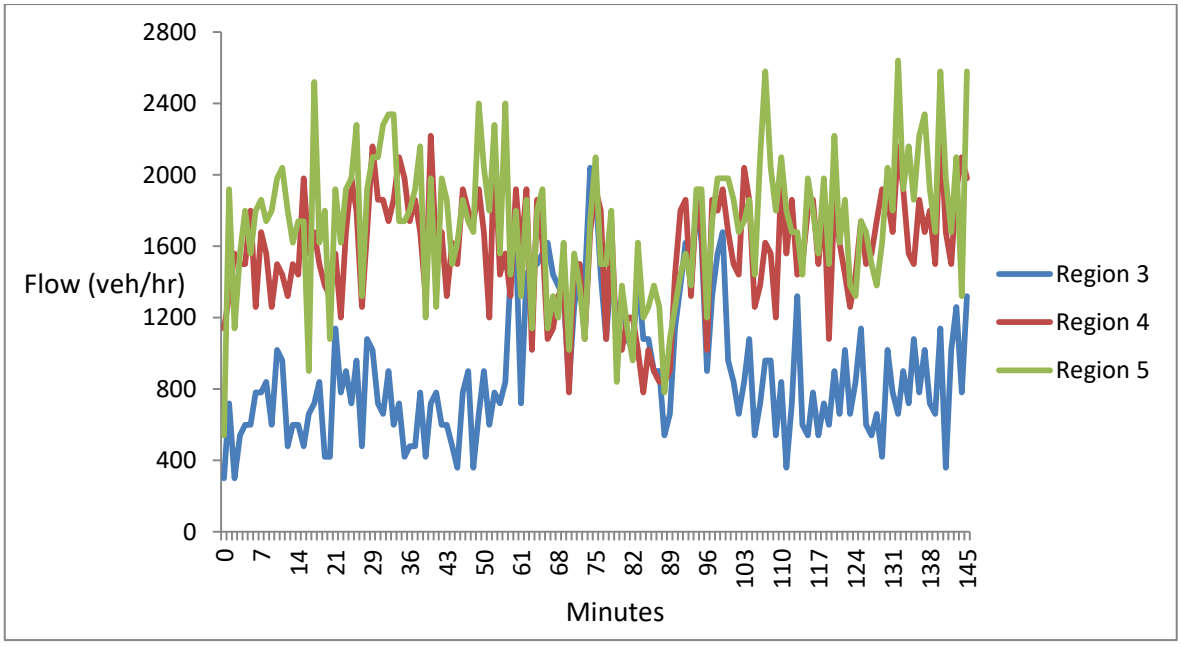


Figure 4.31. Comparison of flow rates of regions 3,4 and 5 (10.10.2018).

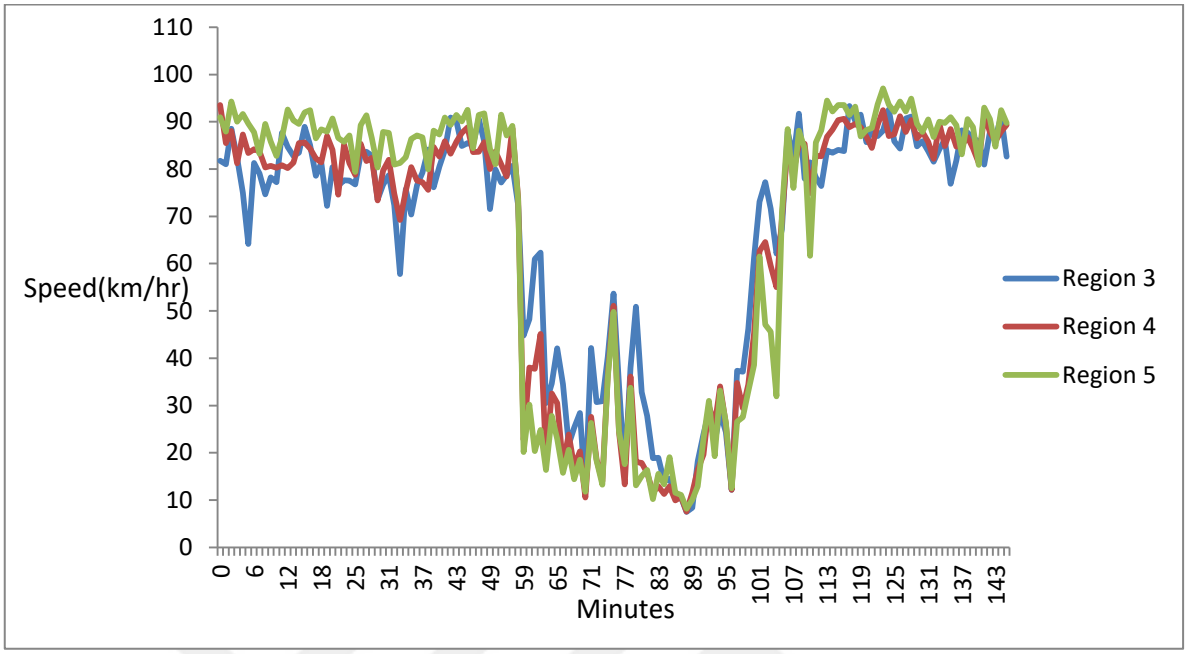


Figure 4.32. Comparison of speeds of regions 3, 4 and 5 (10.10.2018).

### 4.4.2. Heat Maps of Regions

Heat mapping is a useful technique in explaining data graphically where the values are presented as colors [44]. With heat mapping, many things can be understood at once. It is a more effective way of presenting data compared to regular plots. In this section, some heat maps of the regions are shown in a lane by lane basis with respect to the speed and flow data acquired.

4.4.2.1. Flow Heat Maps. In this section, 1 flow heat map for each lane is selected and examined. In these heat maps, consecutive regions are placed from left to right, starting from the lowest region. The y axis contains the time variable which indicates when the average flow value is recorded within the footages. The colors are the third dimension and they indicate the flow rate in that particular time. As the color within the cells get darker, the flow rate gets bigger as is visible from the scale at the right side of the heat map on the following figures.

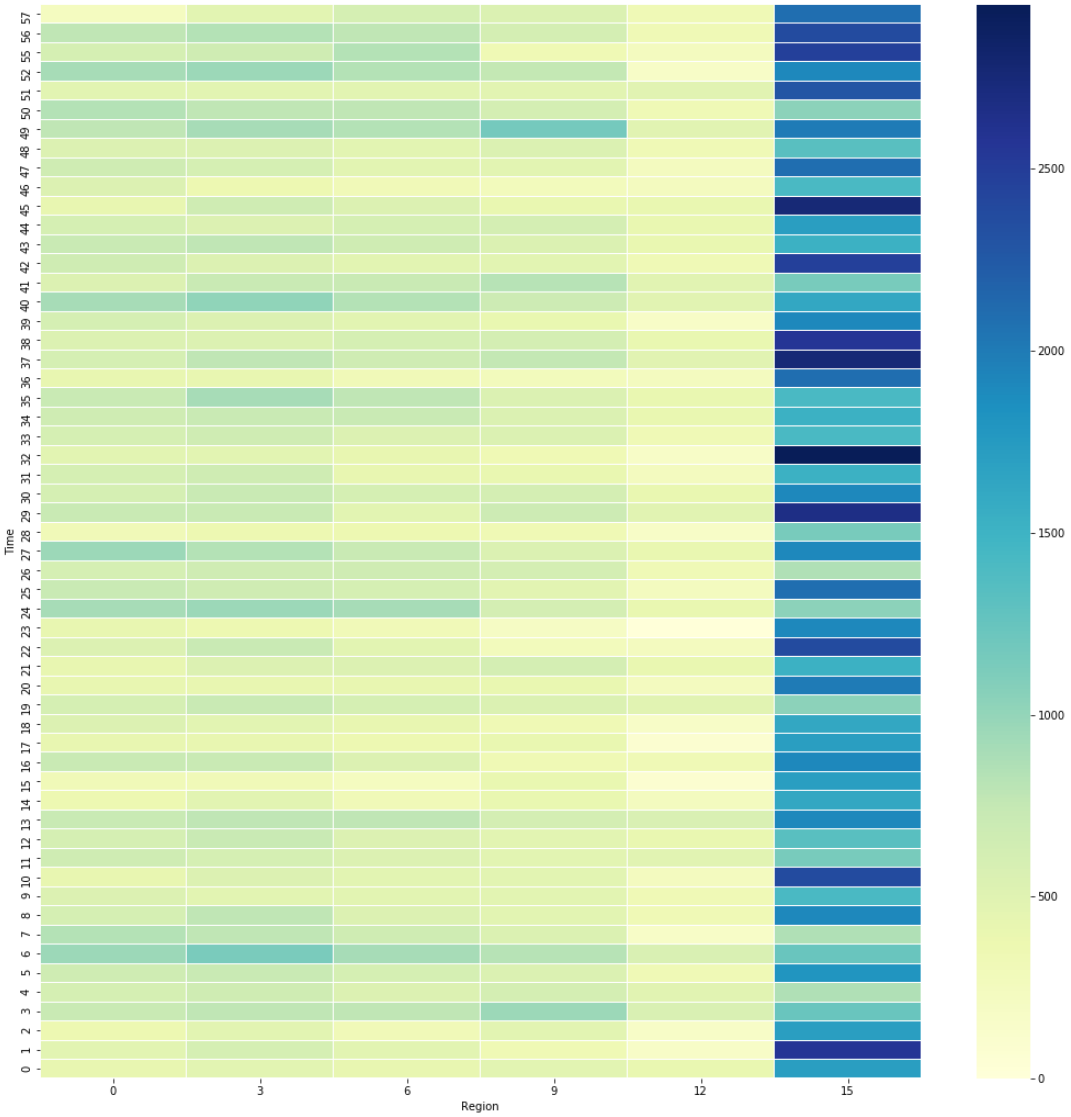


Figure 4.33. Flow heat map of the right lane – 15.10.2018.

In Figure 4.33, the heat map of the right lane for 15.10.2018 is given. Regions 0, 3, 6, 9 and 12 have average flow rates of 600-800 veh/hr with some increases and decreases here and there. There is also a noticeable decrease in flow rate on region 12 compared to the regions upstream. This is most likely related to the diversion of vehicles from the right lane because of the on-ramps. Region 15 is a special region. It contains vehicles coming in from region 12 as well as the vehicles from the on-ramps. All the vehicles that come in from the on-ramps must pass through region 15. Thus, flow values on region 15 can get as big as 2500 veh/hr at certain minutes.

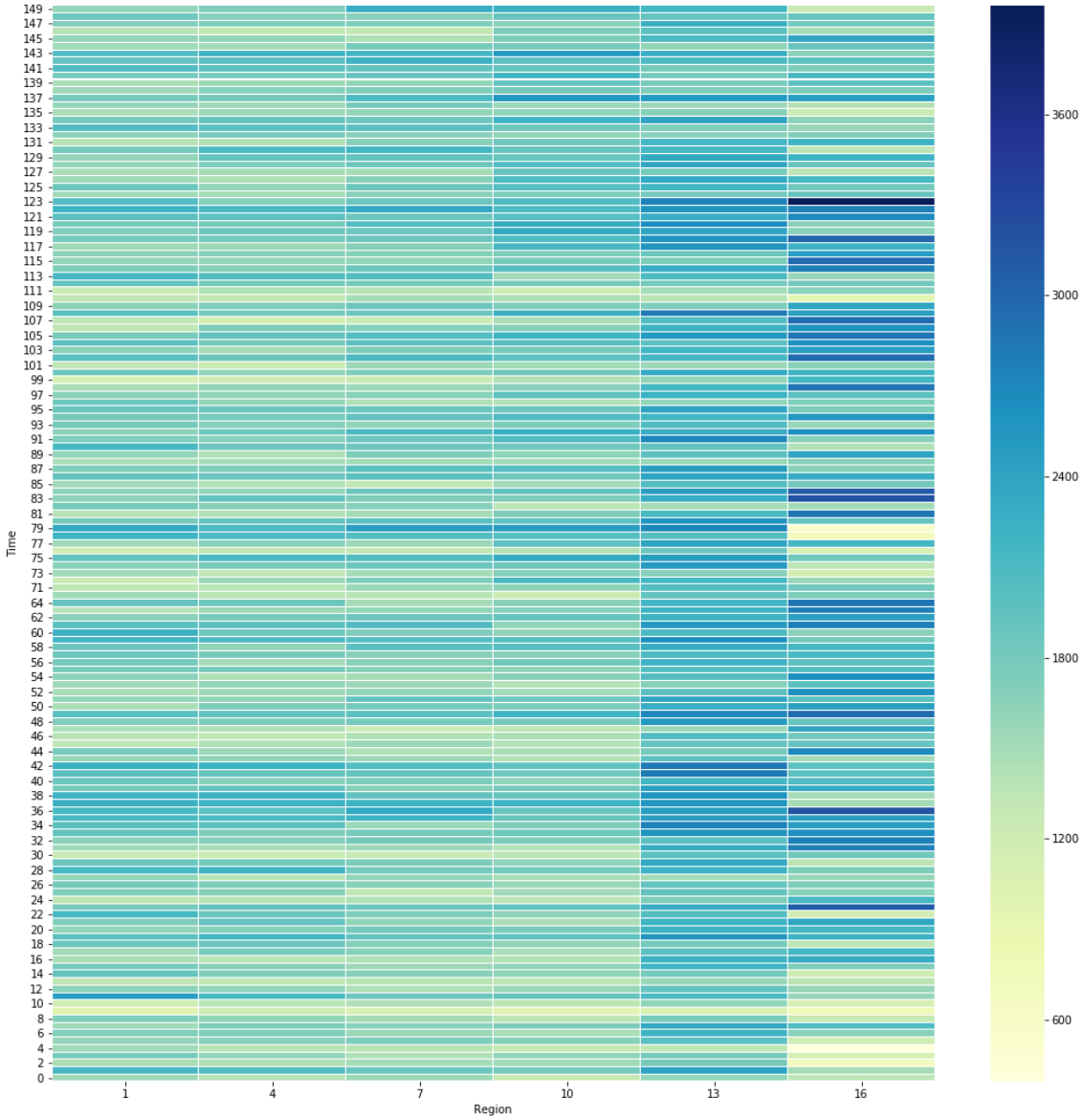


Figure 4.34. Flow heat map of the middle lane – 16.10.2018.

In Figure 4.34, the heat map of the middle lane with respect to the flow rates observed on 15.10.2018 is given. The flow rates have an average of 1800 veh/hr for all the regions with regions 13 and 16 having higher flow rates at most times. This increase in flow rate can be explained by the lane changing vehicles from other two lanes. The reason that there are light cells on region 16 is related to the inability of the image processing algorithm to capture these vehicles in this region within these minutes.

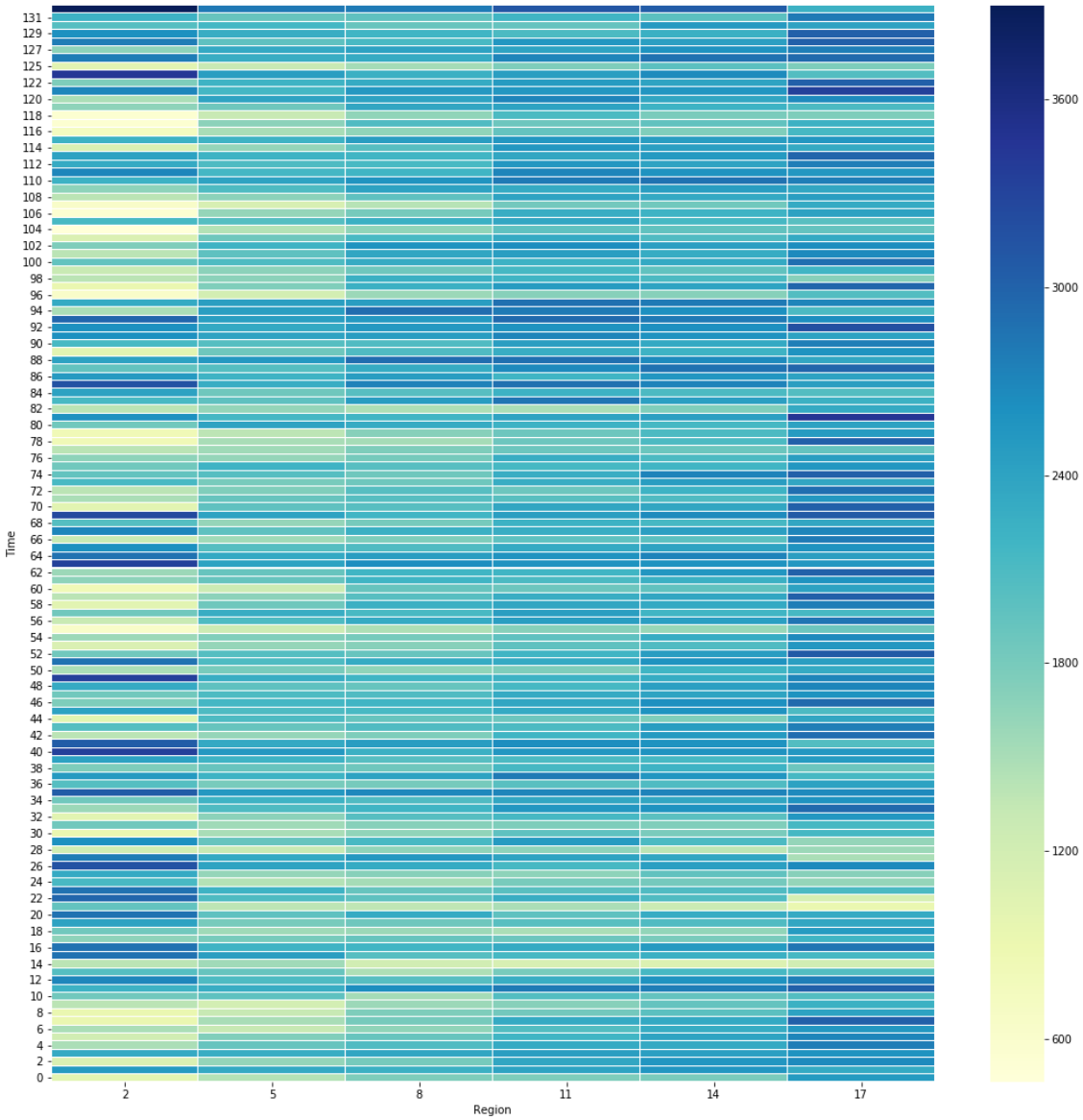


Figure 4.35. Flow heat map of the left lane – 12.10.2018.

In Figure 4.35, flow rate heat map of the left lane on 12.10.2018 is shown. The average flow rate is around 2000 veh/hr for regions 2, 5 and 8 but it gets higher and higher as it reaches 2400 veh/hr on region 17. The flow rate seems stable for all minutes with some increase in regions 8, 11 and 14 between 80th and 95th minutes. The blank cells found on region 2 is also due to the inability of the image processing algorithm to detect, just as it was the case for region 16 on Figure 4.34.

4.4.2.2. Speed Heat Maps. In this section, one speed heat map for each lane is selected and examined. These speed heat maps have the same setup with the flow heat maps that are examined before. The only difference is that the colors represent the speed of a region instead of the flow rate of a region.

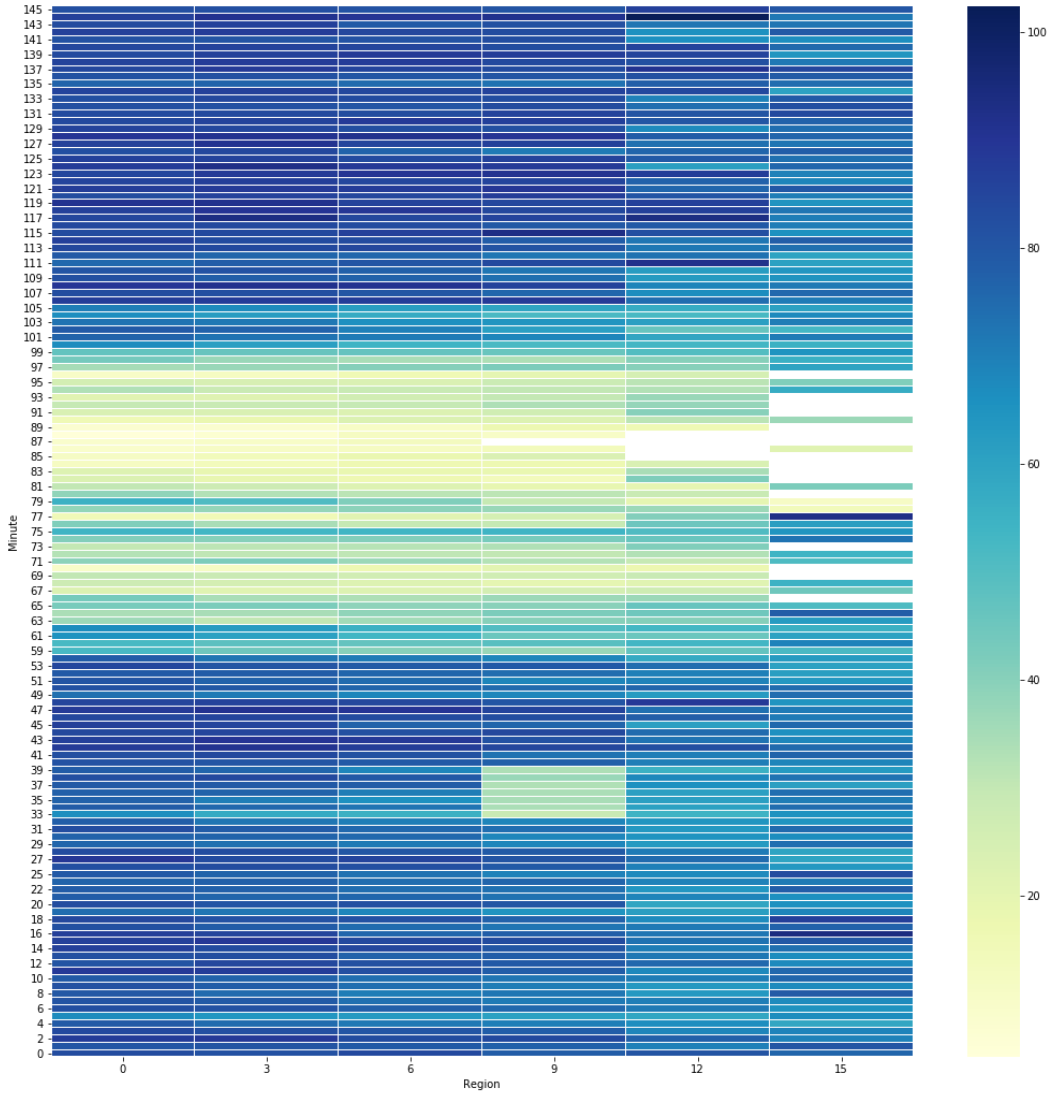


Figure 4.36. Speed heat map of the right lane – 10.10.2018.

In Figure 4.36, the speed heat map of 10.10.2018 is shown. It is clearly visible that the traffic experiences uncongested traffic conditions for the first 1 hour and the last 45 minutes of the 145 minutes data. During this uncongested conditions, the speed values are near 100 km/hr for regions 0, 3, 6 and 9 but there is a noticeable decrease in speed 12 at

most cases which is most likely related to the on-ramps. There are unusual speed readings on region 9 between 33th and 39th minutes where speeds drop to 30-40 km/hr. After investigating this by watching the footages, it is found out that this is related to a car stopping on the emergency lane while invading some portion of the right lane. This must have caused the drivers to be cautious while passing through this location. There is also a 40 minute congested traffic period between the 60th and 100th minutes. In this interval, speed values drop to 10 km/hr at some instances because of the congestion. This result is also found in Figure 4.32 which compares the speed values of three regions in the main arterial on the same day.

In Figure 4.37, the speed heat map of the middle lane on 13.10.2018 is given. It is observed from this figure that speed is around 80-90 km/hr on all the regions for the first 42 minutes of the footages. On the later 25 minutes of the footages, congested flow is experienced by the main arterial. Thus speed values drop to around 20 km/hr for all regions.

In Figure 4.38, the speed heat map of the left lane on 16.10.2018 is given. It is clearly observed from this figure that free flow conditions are present for the whole duration of the footages. Average speeds exceed 80 km/hr for the vast majority of the footages. This result can also be checked by the speed measurements of region 5, which is located on the left lane, on the same day in Figure 4.30. No noticeable speed increases are noted among regions with the exception of region 17 during some instances.

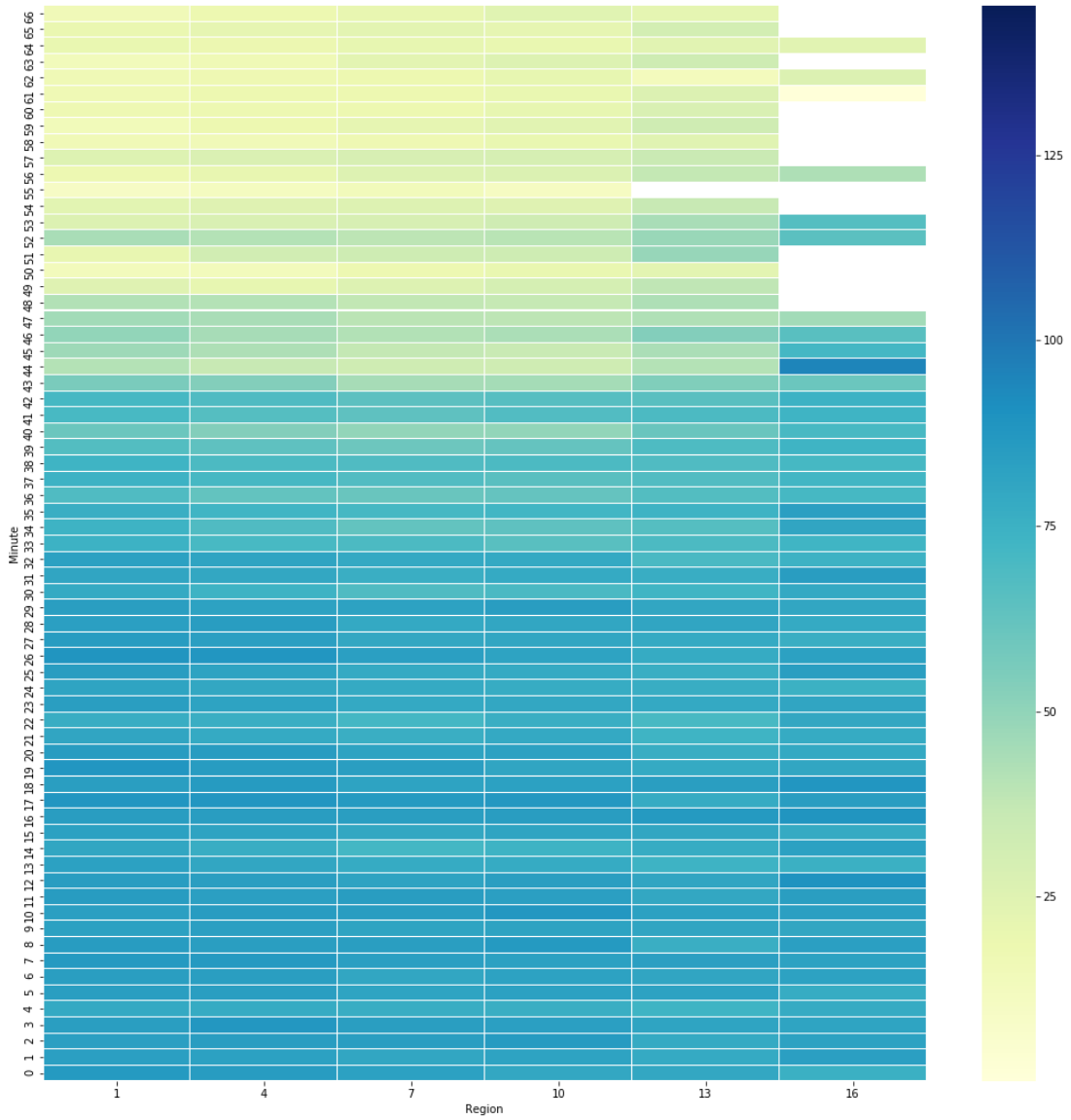


Figure 4.37. Speed heat map of the middle lane – 13.10.2018.

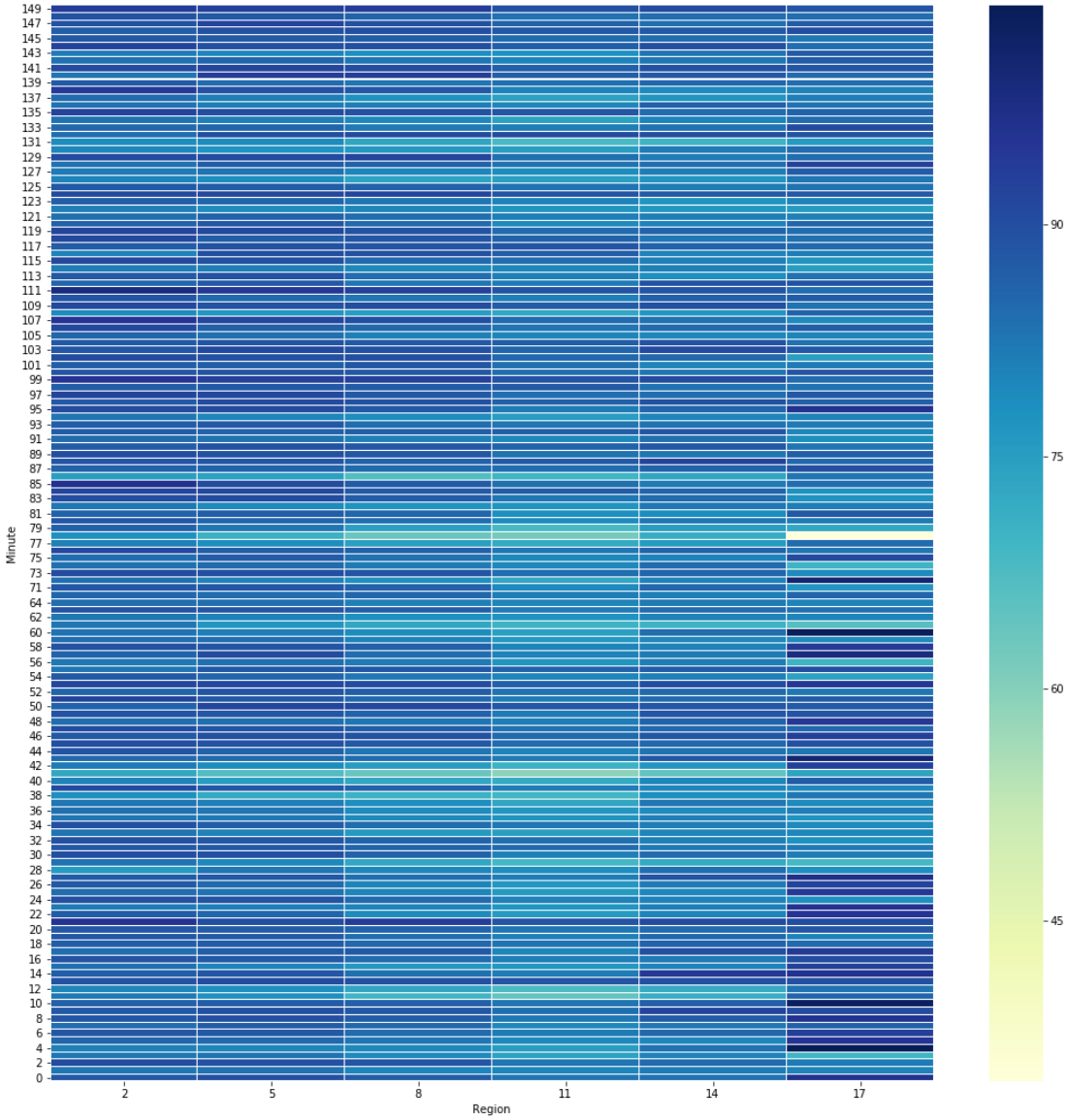


Figure 4.38. Speed heat map of the left lane – 16.10.2018.

### 4.4.3. Regression Results

In this section, some regression models that can be created with the data at hand are examined. In creation of these regression models, the methodology presented in the data analysis section is used. It is found that only the data of 12.10.2018 and 16.10.2018 have meaningful principal components. Thus, one regression model from the traffic data of 12.10.2018 and one new regression model from the traffic data of 16.10.2018 is introduced.

4.4.3.1. Regression Model for the Date 12.10.2018. For 12.10.2018, 4 principal components are created which are: Speed in middle and left lanes, flow in the right lane, speed in the right lane and flow in middle and left lanes. The correlations between these components are given in Figure 4.39. In this figure, it is observed that speed in the middle and left lanes is positively correlated with the speed on the right lane and negatively correlated with the flow in the middle and left lanes and flow in the right lane, flow in middle and left lane is negatively correlated with speed in the right lane and positively correlated with flow in the right lane. With these facts in mind, a regression model which assumes speed in middle and left lanes as the dependent variable is created. The results are given in Figure 4.40, Figure 4.41 and Figure 4.42.

**Correlations**

| | | Speed_in_middle_and_left_lane | Flow_in_middle_and_left_lane | Speed_in_right_lane | Flow_in_right_lane |
|---|---|---|---|---|---|
| Pearson Correlation | Speed_in_middle_and_left_lane | 1,000 | -,570 | ,611 | -,331 |
| | Flow_in_middle_and_left_lane | -,570 | 1,000 | -,378 | ,517 |
| | Speed_in_right_lane | ,611 | -,378 | 1,000 | -,022 |
| | Flow_in_right_lane | -,331 | ,517 | -,022 | 1,000 |
| Sig. (1-tailed) | Speed_in_middle_and_left_lane | . | ,000 | ,000 | ,000 |
| | Flow_in_middle_and_left_lane | ,000 | . | ,000 | ,000 |
| | Speed_in_right_lane | ,000 | ,000 | . | ,402 |
| | Flow_in_right_lane | ,000 | ,000 | ,402 | . |
| N | Speed_in_middle_and_left_lane | 132 | 132 | 132 | 132 |
| | Flow_in_middle_and_left_lane | 132 | 132 | 132 | 132 |
| | Speed_in_right_lane | 132 | 132 | 132 | 132 |
| | Flow_in_right_lane | 132 | 132 | 132 | 132 |

Figure 4.39. Correlations between principal components-12.10.2018 (SPSS).

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,726[a] | ,527 | ,516 | 3,33010 |

a. Predictors: (Constant), Flow_in_right_lane, Speed_in_right_lane, Flow_in_middle_and_left_lane

Figure 4.40. Summary of the model-12.10.2018 (SPSS).

**ANOVA**[a]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 1579,601 | 3 | 526,534 | 47,480 | ,000[b] |
| | Residual | 1419,466 | 128 | 11,090 | | |
| | Total | 2999,067 | 131 | | | |

a. Dependent Variable: Speed_in_middle_and_left_lane

b. Predictors: (Constant), Flow_in_right_lane, Speed_in_right_lane, Flow_in_middle_and_left_lane

Figure 4.41. Anova of the model-12.10.2018 (SPSS).

**Coefficients**[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 64,485 | 5,423 | | 11,891 | ,000 | | | | | |
| | Flow_in_middle_and_left_lane | -,004 | ,001 | -,296 | -3,764 | ,000 | -,570 | -,316 | -,229 | ,598 | 1,672 |
| | Speed_in_right_lane | ,399 | ,054 | ,495 | 7,359 | ,000 | ,611 | ,545 | ,447 | ,816 | 1,226 |
| | Flow_in_right_lane | -,004 | ,002 | -,167 | -2,295 | ,023 | -,331 | -,199 | -,140 | ,698 | 1,433 |

a. Dependent Variable: Speed_in_middle_and_left_lane

Figure 4.42. Coefficients of the model-12.10.2018 (SPSS).

The model has an $R^2$ value of 0,527 as seen in Figure 4.40 which is pretty good. The model is significant with an F-value of 47,480 and a significance value of 0,000 as seen in Figure 4.41. All the independent variables are significant as seen in Figure 4.42 so the regression equation can be written as:

$$\text{Speed\_in\_middle\_and\_left\_lane} = 64{,}485 - 0{,}004 * \text{Flow\_in\_middle\_and\_left\_lane} + 0{,}399 * \text{Speed\_in\_right\_lane} - 0{,}004 * \text{Flow\_in\_right\_lane} \quad (4.8)$$

With Equation 4.8, it can be said that the speed in middle and left lanes is negatively correlated with flow in middle, left and right lanes and positively correlated with speed in the right lane. This was also the result obtained from the correlations table. This is a logical result as an increase in flow rate results in the decrease in speed in the uncongested region where most of the data is obtained from. It also makes sense that the speed in the middle and left lanes increase in the case where the speed in the right lane increases.

4.4.3.2. Regression Model for the Date 16.10.2018. For 16.10.2018, a model has already been created in the Data Analysis section as demonstration. In this section, another model with another dependent variable is created. The principal components for this model are:

Speed in left and middle lanes, speed in right lane, flow in right lane, flow in middle lane and flow in left lane, as shown in Table 4.9. In the demonstration model, flow in the middle lane was selected as the dependent variable. In this model, the dependent variable is selected as speed in the right lane.

Figure 4.43 shows the correlations between the principal components. It is observed from this figure that the speed in the right lane is positively correlated with speed in left and middle lanes and negatively correlated with all flow components, speed in the left and middle lanes is also negatively correlated with all flow components and the flow components are positively correlated with each other. This is the same result with the components of the model for 12.10.2018.

**Correlations**

| | | Speed_in_the_right_lane | Speed_in_left_and_middle_lanes | Flow_rate_in_the_right_lane | Flow_rate_in_the_middle_lane | Flow_rate_in_the_left_lane |
|---|---|---|---|---|---|---|
| Pearson Correlation | Speed_in_the_right_lane | 1,000 | ,525 | -,192 | -,168 | -,222 |
| | Speed_in_left_and_middle_lanes | ,525 | 1,000 | -,470 | -,454 | -,579 |
| | Flow_rate_in_the_right_lane | -,192 | -,470 | 1,000 | ,378 | ,481 |
| | Flow_rate_in_the_middle_lane | -,168 | -,454 | ,378 | 1,000 | ,517 |
| | Flow_rate_in_the_left_lane | -,222 | -,579 | ,481 | ,517 | 1,000 |
| Sig. (1-tailed) | Speed_in_the_right_lane | . | ,000 | ,010 | ,021 | ,004 |
| | Speed_in_left_and_middle_lanes | ,000 | . | ,000 | ,000 | ,000 |
| | Flow_rate_in_the_right_lane | ,010 | ,000 | . | ,000 | ,000 |
| | Flow_rate_in_the_middle_lane | ,021 | ,000 | ,000 | . | ,000 |
| | Flow_rate_in_the_left_lane | ,004 | ,000 | ,000 | ,000 | . |
| N | Speed_in_the_right_lane | 145 | 145 | 145 | 145 | 145 |
| | Speed_in_left_and_middle_lanes | 145 | 145 | 145 | 145 | 145 |
| | Flow_rate_in_the_right_lane | 145 | 145 | 145 | 145 | 145 |
| | Flow_rate_in_the_middle_lane | 145 | 145 | 145 | 145 | 145 |
| | Flow_rate_in_the_left_lane | 145 | 145 | 145 | 145 | 145 |

Figure 4.43. Correlations between principal components-16.10.2018 (SPSS).

In Figures 4.44, 4.45 and 4.46, the regression analysis results are given. As seen in Figure 4.44, the model has an $R^2$ value of 0,276 which is passable. The model is significant with an F value of 54,490 and a significance value of 0,000 as it is observed from Figure 4.45. On the final model, the only significant variable in explaining speed in the right lane

are speed in the left and middle lanes, as seen in Figure 4.46. So, the regression equation can be written as:

$$\text{Speed\_in\_the\_right\_lane} = 22{,}556 + 0{,}682*\text{Speed\_in\_left\_and\_middle\_lanes} \quad (4.9)$$

From this equation, it can be said that the speed in the right lane is positively correlated with speed in the other lanes, just as it was interpreted from Figure 4.43. This result makes sense because a trend in speed increase on the middle and left lanes most likely means that the speed is also increased on the right lane.

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,525[a] | ,276 | ,271 | 4,38860 |

a. Predictors: (Constant), Speed_in_left_and_middle_lanes

Figure 4.44. Summary of the model-16.10.2018 (SPSS).

**ANOVA[a]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 1049,461 | 1 | 1049,461 | 54,490 | ,000[b] |
| | Residual | 2754,158 | 143 | 19,260 | | |
| | Total | 3803,619 | 144 | | | |

a. Dependent Variable: Speed_in_the_right_lane

b. Predictors: (Constant), Speed_in_left_and_middle_lanes

Figure 4.45. Anova of the model-16.10.2018 (SPSS).

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Zero-order | Partial | Part | Tolerance | VIF |
| 1 | (Constant) | 22,556 | 7,728 | | 2,919 | ,004 | | | | | |
| | Speed_in_left_and_middle_lanes | ,682 | ,092 | ,525 | 7,382 | ,000 | ,525 | ,525 | ,525 | 1,000 | 1,000 |

a. Dependent Variable: Speed_in_the_right_lane

Figure 4.46. Coefficients of the model-16.10.2018 (SPSS).

# 5. CONCLUSIONS

In this study, the traffic flow characteristics of a highway segment upstream of the 15 July Martyrs Bridge are investigated using the video footages obtained from the seven day period which begins at 10.10.2018 and ends at 16.10.2018. The road is divided into 35 regions in the beginning but only regions 0 to 17, which are located on the main arterial, are examined in detail due to reasons such as unavailability of data from further regions and insignificance of the characteristics of regions on the on-ramps for the purposes of this study.

First, the video footages are converted to real data using image processing algorithms. But this data contained pixel coordinates of the vehicles instead of coordinates in metric terms. So, the pixel coordinates are converted to metric coordinates using the method mentioned in the Methodology section. Then, the data is used and manipulated to obtain the speed and flow rate values within regions.

After the speed and flow rates inside the regions are determined, the relationships between these parameters are examined on regions located on the right, middle and left lanes. With the final speed and flow data at hand, it is found that the right lane had the smallest flow averages among all lanes most of the time. This flow pattern on the right lane is most likely related to the fact that the right lane on a three lane highway is generally preferred less and it is also related to the presence of the on-ramps downstream. Speed values are also smallest on the right lane most of the time. This is most probably related to the fact that slower vehicles use the right lane more than the other lanes. Left and middle lanes had comparable speed and flow readings with left lane generally having higher flow values and slightly higher speed values. It is also found that in the case of congested flow; however, the speed and flow values become almost equal among all the regions. These results are in line with most traffic studies done on highways.

According to the Greenshields fits for the speed-flow data for 10.10.2018 and 16.10.2018, region 5, which is located on the left lane, has the highest capacity with 2400 veh/hr, region 4, which is located on the middle lane, has the second highest capacity with

2000 veh/hr and region 3, which is located on the right lane, has a capacity of 1400 veh/hr. Region 5 and region 4 have the highest free flow speeds which are around 120 km/hr with region 3 having a free flow speed of 100 km/hr. Similar results are obtained for other regions located on the study area so these results are also representative of the lanes that the regions are located on. Left lane have identical theoretical speed and flow values with the theoretical values given in Highway Capacity Manual [27] where the theoretical free flow speed on a freeway is 120 km/hr and the theoretical capacity flow of a lane is 2400 veh/hr/ln. Middle lane also has the same theoretical speed with HCM but its capacity is lower. For the right lane, the free flow speed and capacity values are considerably lower than the values given in HCM.

Consecutive regions on lanes are found to have similar speed and flow characteristics. The main difference of speed and flow readings are observed between regions in different lanes. This is a logical result since flow rate of consecutive regions might only differ with lane changing which wasn't encountered very frequently. The same thing can be said for speed values as drivers tend to travel with constant speeds with little to no acceleration or deceleration on freeways. Thus, unless there is a special circumstance such as a slow moving vehicle, it is very normal to observe similar speed readings in regions on the same lane.

Region 15 is a special region which contains vehicles coming in from both region 12 and the on-ramps. For this reason, much higher flow rates are observed on this region compared to other regions located on the right lane.

Three linear regression models are created with the traffic data at hand using the principal components analysis approach. In these models, the principal components, which are based on the speed and flow values inside regions, are related to each other and one of the principal components is predicted with the other components. One regression model is created for 12.10.2018 and it is found that the dependent variable speed in middle and left lanes is positively correlated with speed in the right lane and negatively correlated with flow rates of all lanes. For 16.10.2018, two regression models are created. In the first model, which was presented in the Data Analysis section, the dependent variable flow in the middle lane is found to be positively correlated with flow in the left lane and negatively

correlated with speed in the middle and left lane. In the second model, dependent variable speed in the right lane is found to be positively correlated with speed in middle and left lanes with no flow factor being significant. With these results, it can be said that speed on a lane is positively correlated with speed values on other lanes and negatively correlated with flow rates of all lanes. It is also found that left lane and middle lane have comparable speed and flow values with each other as seen by the principal components that group these two lanes together.

# REFERENCES


1. National Programme on Technology Enhanced Learning, *Congestion Studies*, https://nptel.ac.in/courses/105101008/584_Congestion/point2/point.html, accessed at July 2019.

2. J. G. Wardrop, "Some Theoretical Aspects of Road Traffic Research", *Proceedings of the Institution of Civil Engineers*, Vol. 1(3), pp. 325-362, 1952.

3. F. Hall, B. Allen, M. Gunter, "Empirical Analysis of Freeway Flow-Density Relationships", *Transportation Research Part A: General*, Vol. 20(3), pp. 197-210, 1986.

4. J. H. Banks, "Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations", *Transportation Research Record*, No: 1225, pp. 53-60, 1989.

5. H. C. Chin, A. D. May, "Examination of the Speed-Flow Relationship at the Caldecott Tunnel", *Transportation Research Record*, No: 1320, pp. 75-82, 1991.

6. G. R. Andrade, J. R. Setti, "Speed-Flow Relationship and Capacity for Expressways in Brazil", *Transportation Research Board Annual Meeting*, Washington D.C., 2014.

7. V. F. Hurdle, M. I. Merlo, D. Robertson, "Study of Speed-Flow Relationships on Individual Freeway Lanes", *Transportation Research Record*, No: 1591, pp. 7-13, 1997.

8. A. D. May, "Traffic Characteristics and Phenomena on High Density Controlled Access Facilities", *Traffic Engineering*, Vol. 31(6), 1961.

9. F. Busch, “Spurbelastungen und Haufigkeit von Spurwechseln auf einer dreispurigen BAB-Richtungsfahrbahn”, *Strassenverkehrstechnik*, Vol. 28(6), pp. 228-231, 1984.

10. A. D. May, *Traffic Flow Fundamentals*, Prentice Hall, Englewood Cliffs, New Jersey, pp. 60-62, 1990.

11. F. Sasahara, L. Elefteriadou, S. Dong, “Lane-by-Lane Analysis Framework for Conducting Highway Capacity Analyses at Freeway Segments”, *Transportation Research Record*, 2019.

12. C. F. Daganzo, “A Behavioral Theory of Multi-Lane Traffic Flow”, *Transportation Research Part B: Methodological*, Vol. 36(2), pp. 131-158, 2002.

13. M. R. Amin, J. H. Banks, “Test of a Behavioral Theory of Multi-Lane Traffic Flow: Queue and Queue Discharge Flows”, *Transportation Research Record*, No: 1852, 2003.

14. V. L. Knoop, A. Duret, C. Buissoni, B. v. Arem, “Lane Distribution of Traffic near Merging Zones Influence of Variable Speed Limits”, *13th International IEEE Conference on Intelligent Transportation Systems*, Madeira, Portugal, pp. 485-490, 2010.

15. A. Duret, S. Ahn, C. Buisson, “Lane Flow Distribution on a Three-Lane freeway: General features and the effects of traffic controls”, *Transportation Research Part C : Emerging Technologies*, Vol. 24, pp. 157-167, 2012.

16. Wang Y., Luo J., “Study of Rainfall Impacts on Freeway Traffic Flow Characteristics”, *Transportation Research Procedia*, Vol. 25, pp. 1533-1543, 2017.

17. D. Akin, V. Sisiopiku, A. Skabardonis, “Impacts of Weather on Traffic Flow Characteristics of Urban Freeways in Istanbul”, *Procedia – Social and Behavioral Sciences*, Vol. 16, pp. 89-99, 2011.

18. Dunn Engineering Associates, *Traffic Control Systems Handbook,* Federal Highway Administration, New Jersey Ave, Washington DC, 2005.

19. G. Zhang, R. P. Avery, Y. Wang, "Video-Based Vehicle Detection and Classification System for Real-Time Traffic Data Collection Using Uncalibrated Video Cameras", *Transportation Research Board Annual Meeting*, Washington D.C., 2007.

20. Q. Ai, Z. Yao, H. Wei, Z. Li, "Identifying Characteristics of Freeway Traffic Headway by Vehicle Types Using Video Trajectory Data", *10th International Conference of Chinese Transportation Professionals*, pp. 1984-1992, Beijing, 2010.

21. Y. Cho, J. Rice, "Estimating Velocity Fields on a Freeway from Low-Resolution Videos", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 7(4), pp. 463-469, 2006.

22. M.-W. Park, E. Palinginis, I. Brilakis, M. Hunter, R. Guensler, "Unsupervised Framework for Traffic Counting: Speed Estimation Based on Camera Network Data", American Society of Civil Engineers, *International Conference on Computing in Civil and Building Engineering*, Orlando, Florida, pp. 1594-1601, 2014.

23. M. Guo, Z. Wu, H. Zhu, "Empirical Study of Lane-changing Behavior on Three Chinese Freeways", *PLOS ONE*, Vol. 13(1), 2018.

24. X.-Q. Zheng, Z. Wu, S.-X. Xu, M.-M. Guo, Z.-X. Lin, Y.-Y. Zhang, "Video-based Measurement and Data Analysis of Traffic Flow on Urban Expressways", *Acta Mechanica Sinica*, Vol. 27(3), pp. 346-353, 2011 .

25. B. D. Greenshields, "A Study on Traffic Capacity", *Highway Research Board, Proceedings*, Vol. 14, 1935.

26. V. Knoop, S. P. Hoogendorn, H. van Zuylen, “Empirical Differences Between Time Mean Speed and Space Mean Speed”, *Traffic and Granular Flow ‘07,* pp 351-356, 2009 .

27. Transportation Research Board, *Highway Capacity Manual*, Fourth Edition, Transportation Research Board, 1997.

28. F. L. Hall, V. F. Hurdle, J. H. Banks, “Synthesis of Recent Work on the Nature of Speed-Flow and Flow-Occupancy (Or Density) Relationships on Freeways”, *Transportation Research Record*, No: 1365, pp. 12-18, 1992.

29. B. N. Persaud, V. F. Hurdle, “Some New Data That Challenge Some Old Ideas about Speed-Flow Relationships”, *Transportation Research Record,* No: 1194, pp. 191-198, 1988.

30. Massachusetts Institute of Technology, *Flow, Speed and Density*, https://lostcontact.mit.edu/afs/eos.ncsu.edu/info/ce400_info/www2/flow1.html, accessed at March 2019.

31. D. Teodorovic, M. Janic, *Transportation Engineering: Theory Practice and Modeling*, Butterworth-Heinemann, Oxford, United Kingdom, 2016.

32. D. Ni, *Traffic Flow Theory Characteristics Experimental Methods, and Numerical Techniques*, Butterworth-Heinemann, Oxford, United Kingdom, 2016.

33. H. Greenberg, “An Analysis of Traffic Flow”, *Operations Research,* Vol. 7, pp:79-85, 1959.

34. L. C. Edie, “Car Following and Steady-State Theory for Non-congested Traffic”, *Operations Research,* Vol:9 - No:1, pp. 66-76, 1961.

35. R. T. Underwood, *Speed, Volume and Density Relationships, Quality and Theory of Traffic Flow*, Yale Bureau of Highway Traffic, New Haven, Connecticut, 1961.

36. J. L. Drake, J. L. Schofer, A. D. May, "A Statistical Analysis of Speed Density Hypotheses", *Highway Research Record 154*, pp. 53-87, 1967.

37. D. R. Drew, *Traffic Flow Theory and Control*, McGraw-Hill Book Company, New York, 1968.

38. L. A. Pipes, "Car-Following Models and the Fundamental Diagram of Road Traffic", *Transportation Research 1(1),* pp. 21-29, 1967.

39. L. H. Immers, S. Logghe, "Traffic Flow Theory", *Basics of Traffic Engineering*, Heverlee, Belgium, 2002.

40. R. E. Walpole, R. H. Myers, S. L. Myers, K. Ye, *Probability & Statistics for Engineers & Scientists*, Ninth Edition, Prentice Hall, 2011.

41. Statistics Solutions, *Factor Analysis*, https://www.statisticssolutions.com/factor-analysis-sem-factor-analysis/, accessed at June 2019.

42. Towards Data Science, *Probability concepts explained: Maximum likelihood estimation*, https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1, accessed at July 2019.

43. Pjreddie, *Yolo: Real-Time Object Detection*, https://pjreddie.com/darknet/yolo/, accessed at June 2019.

44. Optimizely, *Heatmap*, https://www.optimizely.com/optimization-glossary/heatmap/, accessed at June 2019.