ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

MULTIVARIATE AND FUZZY CLUSTERING APPROACHES TO DYNAMIC CLASSIFICATION OF TRAFFIC FLOW STATES

M.Sc. THESIS

Mehmet Ali SİLGU

Department of Civil Engineering

Transportation Engineering Programme

JANUARY 2015

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

MULTIVARIATE AND FUZZY CLUSTERING APPROACHES TO DYNAMIC CLASSIFICATION OF TRAFFIC FLOW STATES

M.Sc. THESIS

Mehmet Ali SİLGU
(501111437)

Department of Civil Engineering

Transportation Engineering Programme

Thesis Advisor: Prof. Dr. Hilmi Berk ÇELİKOĞLU

JANUARY 2015

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

ÇOK DEĞİŞKENLİ VE BULANIK YAKLAŞIMLARLA TRAFİK AKIM KOŞULLARININ DİNAMİK SINIFLANDIRILMASI

YÜKSEK LİSANS TEZİ

Mehmet Ali SİLGU

İnşaat Mühendisliği Anabilim Dalı

Ulaştırma Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Hilmi Berk ÇELİKOĞLU

OCAK 2015

Mehmet Ali Silgu, a **M.Sc.** student of ITU **Graduate School Of Science Engineering And Technology** student ID 501111437, successfully defended the **thesis** entitled "**MULTIVARIATE AND FUZZY CLUSTERING APPROACHES TO DYNAMIC CLASSIFICATION OF TRAFFIC FLOW STATES**", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Hilmi Berk ÇELİKOĞLU**	....
	İstanbul Technical University	
Jury Members :	**Prof. Dr. Hikmet Kerem CIĞIZOĞLU**	....
	İstanbul Technical University	
	Doç.Dr.Serhan TANYEL	....
	Dokuz Eylül University	

Date of Submission : 12 December 2014
Date of Defense : 23 January 2015

Dedicated to my mom,

FOREWORD

I would like to express my sincere appreciation to Hilmi Berk Çelikoğlu, supervisor of this thesis, for his limitless support, precious guidance, and "being a brother more than a teacher" throughout the completion of this study. In every moment of my life, he will have a unique place, and having a chance to study with him is a special pleasure for me.

I also thank to Onur Deniz for his support, and I want him to know that I am going to stop my coding carrier with this study. I am grateful to Bora Genc, for making this study to look like a thesis.

Obviously, I am indebted to my parents and my brothers from Besiktas for their uncountable support. Absolutely, neither this thesis nor my education would have been possible without their encouragements and "sacrifices". I also thank to Cem Yakiskan, Tuncer Doger, Onder Abay and Emrah Serbes for their great patiences and believes about me to finish this thesis.

Research contained in this thesis is supported by The Scientific & Technological Research Council of Turkey (TUBITAK) under the project '111M415 Analysis of network flow propagation:model development and utilization of simulation models in calibration process' and benefited from participation in EU COST Action 'TU0903 MULTITUDE – Methods and tools for supporting the Use caLibration and validaTIon of Traffic simUlation moDEls'.

December 2014

Mehmet Ali SİLGU
(Civil and Geological Engineer)

TABLE OF CONTENTS

Page

ABBREVIATIONS

BIC	**:** Bayesian Information Criterion
CBD	**:** City Block Distance
CTM	: Cell Transmission Model
CW	: Continious Wace Doppler Waveforms
FCM	: Fuzzy C- Means
FHWA DOT	: Federal Highway Administration Department of Transportation
FMCW	: Frequency Modulated Continuous Waves
HCM	: Highway Capacity Manual
LOS	: Level of Service
LWR	: Lighthill- Whitham- Richards
MAE	: Mean Absolute Error
MAPE	: Mean Absolute Percentage Error
MSE	: Mean Square Error
NN	: Neural Network
PeMS	: Freeway Performance Measurement System
RMSE	: Root Mean Squared Error
RTMS	: Remote Traffic Microwave Sensor
SED	: Square Euclidean Distance
TPS	: Thin Plate Spline

LIST OF TABLES

Page

LIST OF FIGURES

<u>Page</u>

MULTIVARIATE AND FUZZY CLUSTERING APPROACHES TO DYNAMIC CLASSIFICATION OF TRAFFIC FLOW STATES

SUMMARY

In this thesis, performances of multivariate and fuzzy clustering methods in specifying flow state variations reconstructed by a macroscopic flow model are sought. In order to remove the noise in and the wide scatter of traffic data, raw flow measures are filtered prior to modeling process. Traffic flow is simulated by the cell transmission model adopting a two phase fundamental diagram. Flow dynamics specific to the selected freeway test stretch are used to determine prevailing traffic conditions. The classification of flow states over the fundamental diagram are sought utilizing the methods of multivariate and fuzzy cluster analyses by considering the stretch density. The fundamental diagram of speed-density is plotted to specify the current corresponding flow state. Both multivariate and fuzzy clustering analyses returned promising results on state classification which in turn helps to capture sudden changes on test stretch flow states.

The perfomance analyses are focused explicitly on the application of clustering for partitioning states of traffic flow, using three clustering algorithms: K-means clustering by Square Euclidean Distance, K-means clustering by City Block Distance and fuzzy C-means. Clustering algorithms have the flexibility to specify the number of clusters. Categorization has been based on similarities and dissimilarities of traffic flow variables without specifying arbitrary values to bound states. Clustering methods are processed in a time-varying fashion to partition the fundamental diagrams at selected temporal resolutions. In order to comparatively evaluate the clustering performances of multivariate and fuzzy methods on lane-based densities relative to deterministic clustering, a number of statistical criteria, including the root mean squared error, the mean square error , the mean absolute error , the mean absolute percentage error and the coefficient of determination.

The performances of K- means clustering and fuzzy c- means clustering, using both the Square Euclidean and City Block Distance measures, are evaluated in two cases, with two assumptions.The procedures followed by multivariate and fuzzy clustering methods are systematically dynamic that enables the partitions over the fundamental diagram match approximately with the flow states derived by the static partitioning method.

It is shown that the comprasions presented that the K-means clustering by Square Euclidean Distance and fuzzy c-means methods perform better and appear to yield to classifications consistent with two types of level of service, which are calculated by using HCM-defined level of service.

ÇOK DEĞİŞKENLİ VE BULANIK YAKLAŞIMLARLA TRAFİK AKIMININ DİNAMİK SINIFLANDIRILMASI

ÖZET

Bu tez kapsamında; makroskopik bir trafik akım modeliyle oluşturulmuş akım koşulu farklılaşmalarını belirlemek amacıyla çok değişkenli ve bulanık kümeleme yöntemlerinin başarımları araştırılmıştır. Veri toplama teknolojisindeki büyük gelişmelere rağmen ortaya çıkan gürültüyü gidermek ve geniş saçılımı kabul edilebilir düzeye getirmek amacıyla, ham trafik değişkenleri modelleme öncesi filtrelenmiştir. Trafik akımı, iki fazlı bir temel eğriyi baz alarak hesap yapan hücre geçişi modeliyle benzetilmiştir. Seçilen otoyol kesimindeki akım dinamikleri, varolan akım koşullarını belirlemek amacıyla irdelenmiştir. Temel eğri üzerinde akım koşullarının sınıflanması, kesim yoğunluk değişkeni gözetilerek kümeleme yöntemleriyle aranmıştır. Karar vericilerin kişisel görüşlerini yansıtan yoğunluk değerlerinin hizmet düzeyi belirlenmesinde ne derece doğru olduğu da tartışmalı bir konudur.

Kümeleme analizi ,çok değişkenli istatistiksel analiz yöntemlerinden sınıflandırma işlemine çok benzemekle beraber, sınıflandırma işleminde sınıflar önceden belli iken kümeleme analizinde sınıflar önceden belli değildir. Verilerin hangi kümelere, hatta kaç değişik kümeye ayrılacağı eldeki verilerin birbirlerine olan benzerliğine göre belirlenir. Kümeleme analizi antropolojiden telekominikasyona kadar geniş bir yelpazede kullanım sağlar. Veri setini oluşturan her bir veri kümelere ayrılırken, uzaklık ve benzerlik kavramlarından yararlanılır. Bu, veri setindeki her bir verinin diğer bir veri ile olan benzerliği ya da her bir verinin veri setindeki diğer verilerden uzaklığı olduğu gibi oluşturulan gerçek ve aday kümeler arasındaki mesafe ve benzerliği de içerir.

Akım koşullarının karşılaştırılmasında her bir hizmet düzeyi seviyesinin bir kümeyi temsil ettiği kabul edilmiştir. Buradan yola çıkarak hizmet düzeyi seviyesini belirleyen sınır değerleri aynı zamanda küme sınırlarını belirleyen sınır değerler olarak kabul edilmiştir. Çok değişkenli istatistiksel analiz yöntemlerinden biri olan kümeleme analizinde hiyerarşik olmayan küme yaklaşımı ele alınmıştır. Hiyerarşik olmayan kümeleme yöntemlerinde veri setini oluşturan veriler, önceden belirlenen küme sayısına göre ayrılır. Burada en önemli nokta, veri setini oluşturan veri sayısının belirlenen küme sayısından büyük olmasıdır. Hiyerarşik yöntemlerden en büyük farkı, küme sayısının önceden bilinmesidir. Bununla birlikte, kümeler arası en büyük ve en küçük mesafe ile benzerlik ölçütleri önceden tanımlanmalıdır.

Hiyerarşik olmayan kümeleme yöntemleri, hiyerarşik olanlara göre daha hızlı çalışırlar. Çünkü hiyerarşik olmayan yöntemlerde benzerlik/ mesafe matrisi kullanımına gerek yoktur.Bundan dolayı da büyük veri setlerine hiyerarşik yöntemlere kıyasla daha uygundurlar. Çalışma kapsamında hiyerarşik olmayan kümeleme yönteminin, hiyerarşik kümeleme metoduna kıyasla tercih edilmesinin temel sebebi veri setinin büyüklüğü ve hesap süresinden tasarruftur. Hiyerarşik

olmayan kümeleme yöntemlerinden biri olan ve tez kapsamında kullanılan K-ortalama yöntemi sürekli olarak kümelerin yenilendiği ve en uygun çözüme ulaşana kadar devam eden döngüsel bir yöntemdir. Çok değişkenli kümeleme analizinde, kümelerin merkezlerinin belirlenmesi için Manhattan Uzaklığı ve Öklidyen Kare uzaklığı kullanılmıştır. Aynı şekilde küme merkezi belirlenmesinde kullanılan Chebshyev uzaklığı literatürde benzer çalışma olmadığından tercih edilmemiştir. Bulanık c- ortalamalar yönteminde ise girdi ve çıktının nümerik olduğu varsayılmıştır. Bulanık c- ortalamalar yönteminde çok değişkenli istatistiksel analiz yöntemlerinin aksine küme merkezinin şekli konusunda kabul yapmak gerekmektedir.

Tez kapsamında tüm küme merkezlerinin yuvarlak olduğu kabulu yapılmıştır. Bulanık c- ortalamalar yönteminde ihtiyacı hasıl olan bulanıklaştırma parametresi, kümeleme işlemine son verme parametresi ve model belirleme matrisi gibi parametrelerde seçici davranılırak, diğer yöntemlere benzetimi sağlanmıştır. Küme merkezi kavramında, merkez kümenin ortasını temsil etmesine rağmen, aslında kümenin gerçekten tam ortasında böyle bir elemanın bulunmasına gerek yoktur. Elde edilen küme merkezlerinden yola çıkılarak herbir verinin hangi kümeye ait olduğu belirlenmiş ve hız-akım diyagramı oluşturulmuştur. Bahsi geçen hız-akım diyagramında değişikliği görmek amacıyla veri setine her 4 saatte bir yeni veri akışı sağlanmış ve “Yolların Kapasitesi El Kitabı”na, K- ortalama yöntemi Öklidyen Uzaklık yaklaşımı, K- ortalama yöntemi Manhattan yaklaşımı ve bulanık c- ortalama yaklaşımı kullanılarak kümeleme işlemi yapılmıştır. Kümeleme işlemleri sonucu elde edilen küme merkezlerinden yola çıkılarak küme sınırları belirlenmiştir. Küme sınırlarının belirlenmesinde veri setindeki en büyük yoğunluk değeri F hizmet düzeyinin son sınırı olarak kabul edilmiştir. A hizmet düzeyinin belirlenmesinde ise A hizmet düzeyini temsil eden küme merkezi iki ile çarpılmış ve elde edilen değer ile sıfır arasında kalan her veri A hizmet düzeyinde kabul edilmiştir. A ve F arasında kalan tüm diğer hizmet düzeylerinin sınırlarının belirlenmesi görselleştirilerek tez kapsamında sunulmuştur. Kümeleme işlemleri sonucunda özellikle E ve F hizmet düzeylerini temsil eden bölgelerde artma ve azalma gözlemlenmiştir.

Küme merkezlerinin başarısının sınanması için iki farklı durum oluşturulmuştur. Bunlardan ilki, küme merkezlerinin “Yolların Kapasitesi El Kitabında” sınırlandırılmış hizmet düzeyi sınırları kullanılarak hesaplanan statik yöntemdir. Statik yöntemde, her bir hizmet düzeyi bir küme olarak kabul edilmiş ve küme sınırları önceden bilindiği için, küme sınırını oluşturan değerler arası farkın ortalaması bulunmuştur. İkincisi ve dinamik olduğu düşünülen yöntemde ise “Yolların Kapasitesi El Kitabında”ki sınırlar arasında kalan verilerin aritmetik ortalaması hesaplanmaktadır.

Veri seti büyüdükçe her seferinde tekrar hesap yapılarak küme merkezlerinin değerleri tekrar tekrar bulunmuştur. Toplamda sistem 10 defa yüklenmiş ve küme merkezlerinin değişimlerini gösteren grafikler elde edlmiştir. Çok değişkenli ve bulanık kümeleme sistemlerinin, oluşturulan iki durum karşısında nasıl davrandığını görmek amacıyla hata değerleri hesaplanmıştır. Hem çok değişkenli hem de bulanık kümeleme yaklaşımları, örnek otoyol kesimi üzerindeki ani koşul değişimlerini tespit etmeye yarayan başarılı sınıflama sonuçları vermiştir.

Çok değişkenli ve bulanık kümeleme yöntemlerince izlenen prosedür, sistematik olarak dinamiktir ve temel eğri üzerinde statik bölütleme yöntemiyle elde edilen kümelere oldukça yaklaşık kümeler oluşturabilmektedir. K-ortalamalar ve bulanık c-ortalamalar yöntemiyle elde edilen sonuçlar üzerinden hesaplanmış belirlenim katsayıları, elde edilen sonuçları istatistik yönden karşılaştırmalı olarak değerlendirmek amacıyla kullanılmıştır. Çalışma kapsamında elde edilen bütün hata terimlerinde K- ortalama yöntemi Öklidyen yaklaşımı ile bulanık c- ortalamalar yönteminin birbirine yakın sonuçlar vermesinin sebebi olarak her iki yönteminde Öklidyen mesafeyi kullanarak çözüm yapması düşünülmektedir. Özellikle bulanık c-ortalamalar yönteminde bulanıklaştırıcı parametrenin uzaklık ölçütünü Öklidyen olarak ele alması için gerekli değer ortaya konmuştur. K- ortalama yöntemi Manhattan yaklaşımının bahsi geçen diğer kümeleme yaklaşımlarından özellikle ilk iki yükleme için farklılık göstermesinin temel sebebi olarak uzaklık ölçütü hesabında farklı davranması olduğu düşünülmektedir. Diğer yüklemelerde ise K- ortalama yöntemi Öklidyen yaklaşımına ve bulanık c- ortalama yöntemine benzer sonuçlar vermesinin veri seti büyüklüğüyle ilişkili olduğu düşünülmektedir.

İlerleyen çalışmalarda, veri toplanmasından kümeleme işleminin sonuna kadar olan bütün işlemlerin dinamik olarak yapılabileceği öngörüsünde bulunulmaktadır. Ardışık trafik ölçüm sensörlerinden elde edilen ölçümlerin dinamik olarak filtrelenmesi, trafik akımının yine dinamik olarak, iki fazlı bir temel eğriyi baz alan ve buna göre hesap yapan hücre geçişi modeliyle benzetilmesi ve son olarak kümelenmesi işlemi düşünülmektedir. Kümeleme işleminde küme sayısının dinamik olarak belirlenmesi ve "Yolların Kapasitesi El Kitabı"na göre kıyaslanması ilerleyen çalışmaların temelini oluşturmaktadır.

Küme sınırlarının belirlenmesinde farklı algoritmaların geliştirilerek geçerliliğinin belirlenmesi, kümeleme analizindeki en temel ölçüt olarak uzaklık parametrelerinin farklılaştırılması da gelecek çalışmalarda irdelenebilir.Bulanık kümeleme yönteminde girdinin sözel, çıktının ise nümerik olduğu yöntemlerin de incelenebileceği düşünülmektedir. Var olan hizmet düzeylerinin belirlenmesinde (A ve D hizmet düzeyleri dahil olmak üzere, A'dan D'ye) kişisel tecrübelere dayanıldığı göz önünde bulundurulursa, girdinin sözel olduğu bir durumun daha iyi sonuçlar verme ihtimali göz ardı edilmemelidir. Bütün bahsi geçen yöntemler dışında yapay sinir ağları ve genetik algoritmalar ile oluşturulacak dinamik bir sistemin performansının diğer yöntemlere kıyaslanması da gelecek çalışmaların konusu olabilir. Tez kapsamında kümeleme işleminde kullanılan MATLAB programının bu dinamik öngörüye uygun olup olmadığı da ilerleyen çalışmalarda cevabını bekleyen bir soru olarak karşımızda durmaktadır.

Sonuç olarak, yapılan çalışmanın tekrarsız kaza olaylarında ve kırılmalar sonucu oluşan akımdaki anlık değişimlerin bulunmasında ne derece etkili olduğunu zaman gösterecektir. Yükleme süresinin azaltılarak sistemin tekrar tekrar yüklenmesinin ve elde edilen küme merkezlerinin "Yolların Kapasitesi El Kitabı" değerlerine yakınlığı, benzerliğinin bulunması ve yorumlanması da oldukça önem arz etmektedir.

1. INTRODUCTION

1.1 Motivation

Traffic data is of the crucial inputs to realization of the activities performed for the planning, design, construction, maintenance and operation of roads. In transportation, traffic flow modeling aims generally to propose a series of analyses based on the measurements on traffic flow in order to both specify the variations in flow states and to realistically provide predictions for future.

It is well-known that instantaneous variations in traffic cause unstable flow conditions decreasing the volume and volume to capacity ratio all of which result in the traffic congestion (HCM, 2010). Congestion exists due to decreased ratio of capacity utilization, for example by a physical obstacle, yields to time and comfort related problems.

The specification of maximum densities, alternatively density boundaries, for LOS A through D is based on the collective professional judgment of the members of the Committee on Highway Capacity and Quality of Service of the Transportation Research Board. It is worthy to note that if the traffic parameters' boundaries are made only by human, these values will be subjectivity (Wei and Li, 2012). To reduce and finish up this subjectivity, it is sought to determine some approaches that are based on mathematical background. Cluster analysis, crisp or fuzzy, has mathematical background, aforesaid earlier, that contains in itself. That is why cluster analysis underlies this study.

1.2 Scope of the Thesis

The study proposes the integration of multivariate and fuzzy clustering methods to an overall process that seeks to capture flow state variations using traffic data of two adjacent freeway segments.

Incorporation of the multivariate clustering and fuzzy clustering methods to a flow state specification process helps to improve the detection and modeling performance on non-recurrent events, such as incidents, and contributes to the progressing literature as an up-to-date alternative. In order to present a comparative evaluation and derive accurate and realistic service level boundaries specific to analyzed freeway segment the level of service concept in Highway Capacity Manual 2010 is incorporated.

1.3 Thesis Contributions

The study aims to comparatively evaluate the dynamic classification performance of multivariate clustering analysis and fuzzy clustering methods in partitioning the fundamental diagram of flow vs. density, and hence flow vs. speed, that represent the flow states as resultants from a macroscopic cell transmission model simulation with the input of filtered data on actual traffic.

1.4 Thesis Organization

The thesis is organized into five chapters that motivation, scope of thesis and thesis contributions are explained in first chapter. The second chapter deals with the basic definitions of traffic flow characteristics and traffic stream models. Third section is related to basic freeway segment. This chapter contains the definition of level of service term. Explanation on the computation of the level of service and the literature overview on multivariate methods are presented in Chapter 3. The fourth chapter is composed of data analysis, modeling and classification issue. Chapter 5 concludes the thesis with findings and future research directions.

2. FUNDAMENTALS OF TRAFFIC FLOW DYNAMICS

In this chapter, a number of that are useful to describe the relationship among traffic flow characteristics are provided. Preceding the descriptions and definitions on traffic flow measures are given in Section 2.1. The following section presents variables of traffic flow. The next section progression in the measurement procedures and technologies are presented where the last section, on traffic models, focuses on relationship among speed, flow and density, either in two variables, or in those that attempt to deal simultaneously with the three variables.

2.1 Representing Traffic Flow Measures

When there is a point sensor to measure variables on the traffic flow, there are three types of data that can be collected. These data can be expressed as traffic flow measures. They are space, time and traffic unit which are denoted respectively by *X*, *T* and *N*. Before defining flow measures, it is important to define the term "trajectory". It can be expressed as "the location of the vehicle as a function of time". Figure 2.1 represents an example of a trajectory.

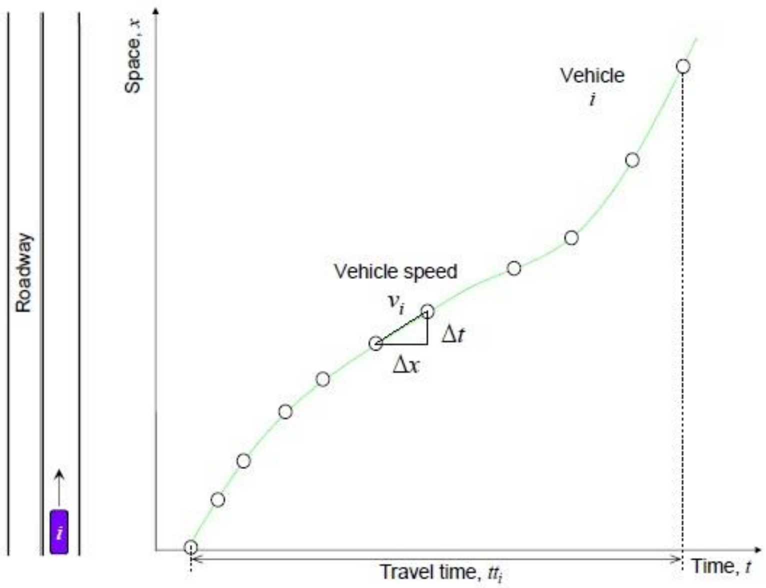

Figure 2.1 : A trajectory (Ni,2013).

Traffic flow measures are not meaningful on their own and they, solely, do not provide adequate information. It is difficult to construct mathematical models by using solely the traffic flow measures. Only, simulation models output data of which are flow measures can be calibrated. When two of the flow measures are considered at a time, the variables of traffic flow can be derived. Hence, it is straight to assume that a flow measure is infinite when deriving a variable considering the rest two measures (Çelikoglu and Gedizlioğlu, 2012). Figure 2.2 and Figure 2.3 show the situations respectively when distance and duration (time) are infinite. By using *N*-*T* relationship, in road section; ($X \rightarrow \Delta x$) $\rightarrow N/T$ = Flow;

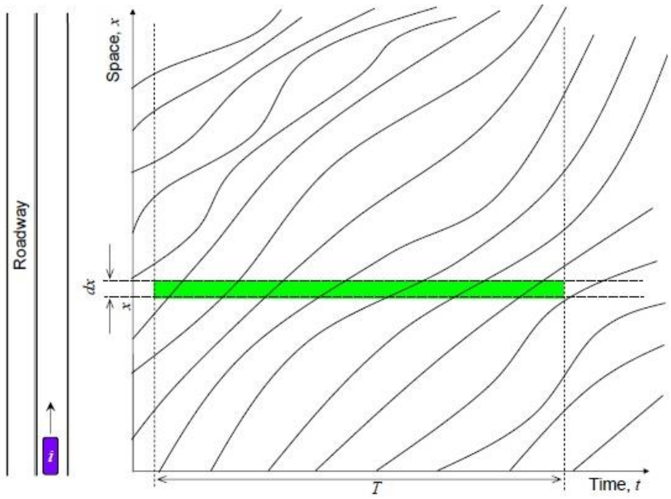

Figure 2.2 : Time- Space Diagram with infinite distance (Ni,2013).

By using *N*-*X* relationship, in an instant over a road segment; ($T \rightarrow \Delta t$)$\rightarrow$ N/X= Density;

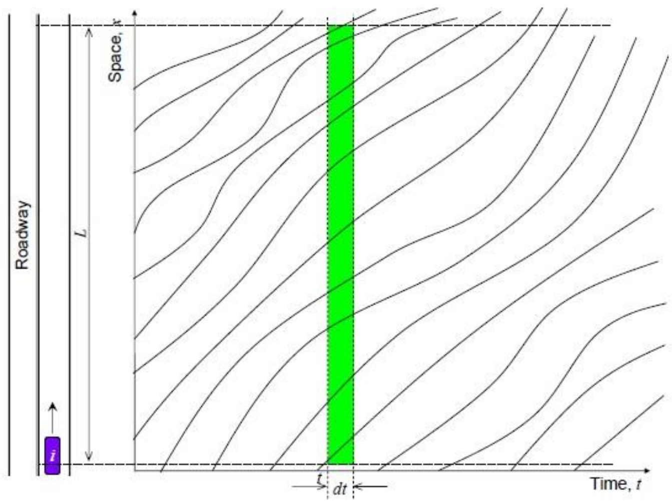

Figure 2.3 : Time- Space Diagram with infinite duration (Ni, 2013).

By using *X*-*T* relationship, for a traffic flow; (N=1)$\rightarrow X/T$ =Speed can be calculated.

2.2 Variables of Traffic Flow

Three traffic flow variables, i.e., volume or flow, speed, and density, can be used to describe traffic on any roadway. Volume or traffic flow is a parameter common to both uninterrupted and interrupted flow facilities, but speed and density apply primarily to uninterrupted flow (HCM, 2010). Some variables related to flow rate, such as spacing and headway, also are used for both types of facilities; other parameters, such as saturation flow or gap, are specific to interrupted flow.

2.2.1 Volume or flow

Volume can be determined as the total number of vehicles that pass over a given point or section of a lane or roadway during a given time interval. Volume can be expressed in terms of annual, daily, hourly, or sub hourly periods.

It is useful to define a variant of flow, the flow rate. Flow rate is "the hourly equivalent of flow at which vehicles pass over a given point or section of a lane or roadway during a given time interval of less than 1 h, usually 15 min" (HCM, 2010).

The distinction between volume and flow rate is important. Volume is the number of vehicles observed or predicted to pass a point during a time interval. Flow rate represents the number of vehicles passing a point during a time interval less than 1 h, but expressed as an equivalent hourly rate.

2.2.2 Speed

Speed is defined as the distance covered per unit of time, generally expressed as kilometers per hour (km/h). Average travel speed is computed by dividing the length of the highway to the average travel time of the vehicles travelling over it. If travel times t_1, t_2, $t_3, \ldots, t_n$ (in hours) are measured for n vehicles travelling over a segment of length L, the average travel speed is computed using Equation **(2.1)**;

$$S = \frac{nL}{\sum_{i=1}^{n} t_i} = \frac{L}{\frac{1}{n}\sum_{i=1}^{n} t_i} = \frac{L}{t_a}; \qquad \textbf{(2.1)}$$

Where S, L, t_i , n and t_a denotes respectively average travel speed (km/h), length of the highway segment (km), travel time of the i_{th} vehicle to travel over the section (h), number of travel times observed, and average travel time over L (h).

Several different speed parameters can be derived from traffic flow. These include the following:

Space mean speed is a statistical term denoting an average speed based on the average travel time of vehicles to travel over a segment of roadway.

Time mean speed is the arithmetic average of speeds of vehicles observed passing a point on a highway.

Free-flow speed is that speed which exists when flows approach zero under free-flow conditions while optimum speed is that speed which exists under maximum flow conditions (May, 1990).

Headway is the time between successive vehicles as they pass a point on a lane or roadway, also measured from the same point on each vehicle. Headway can be easily measured with observations that can make by using chronometer as vehicles pass a point on the roadway.

Flow can be alternatively derived using the headway measure by following Equation **(2.2)**.

$$Flow\ \left(\frac{veh}{h}\right) = \frac{3600}{headway\ (\frac{s}{veh})} \quad \textbf{(2.2)}$$

2.2.3 Density

Density is the number of vehicles occupying a given length of a lane or roadway at a particular instant of time.

Direct measurement of density in traffic engineering is difficult. It requires a vantage point for photographing, videotaping, or observing significant lengths of highway. Density can be computed, however, from the average travel speed and flow , which are measured more easily. Equation **(2.3)** is used for computing density.

$$k = \frac{q}{u} \quad \textbf{(2.3)}$$

Where *q*, *u* and *k* denotes respectively the flow (veh/h), average travel speed (km/h) and density (veh/km).

It is essential to explain two types of density. Jam density is that density that occurs when both speed and flow approach zero while optimum density occurs under maximum flow conditions. (May, 1990)

Roadway occupancy is frequently used as a surrogate for density in control systems because of its measurement convenience. Occupancy in space is the ratio of roadway length covered by vehicles. Also, occupancy in time identifies the ratio of time a roadway cross section is occupied by vehicles (HCM, 2000).

Spacing is the distance between successive vehicles in a traffic stream, measured from the same point on each vehicle. Spacing is a distance, measured in meters. Measuring the distance between common points on successive vehicles at a particular instant is given the spacing. It is required for aerial photographic techniques to obtain spacing.

Density can be alternatively derived using the spacing measure by following Equation **(2.4)**.

$$Density = \frac{1000}{spacing\ (\frac{m}{veh})} \qquad \textbf{(2.4)}$$

The relationship between average spacing and average headway in a traffic flow depends on speed, as indicated in Equation **(2.5)**;

$$Headway\ \left(\frac{s}{veh}\right) = \frac{spacing\ (\frac{m}{veh})}{speed\ (\frac{m}{s})} \qquad \textbf{(2.5)}$$

2.3 Measurement Procedures

Five measurement procedures are discussed in this section; i.e., measurement at a point; measurement over a short section; measurement over a length of road; the use of an observer moving in traffic stream; and wide-area samples obtained simultaneously from a number of vehicles (as a part of Intelligent Transportation System).

For each method, expect for wide-area samples obtained simultaneously from a number of vehicles, the section is aimed to contain an identification of the variables that the particular procedure measures, as contrasted with the variables that can only be estimated (FHWA DOT, 2006).

2.3.1 Measurement at point

Measurement at a point was the first procedure used for traffic data collection by hand tallies or pneumatic tubes. This method is easily suitable for procuring volume counts and for this reason flow rates directly, can provide time headways. Changing of technology to make measurements at a point on freeways can explain by an example. 30 years ago, most of measurements were made by pneumatic tubes; however, nowadays the measurements are made by microwave, radar, photocells, ultrasonic, and television cameras. Except for radar and microwave detectors, all those detectors which named on above cannot obtain speed at “a point”. Both microwave detectors and radar’s frequencies of operation mean that a vehicle needs to move only about one centimeter while speed is measuring. If there is not instrument for a moving vehicle, a second observation location is necessary to obtain speeds (FHWA DOT, 2006). The data which is used for this thesis was collected by radar techniques.

Radar is defined as “a device for transmitting electromagnetic signals and receiving echoes from objects of interest within its volume of coverage”. Moreover, the term, microwave, refers to the wavelength of transmitted energy, usually between 1 and 30 centimeters. This corresponds to a frequency range of 1 gigahertz (Ghz) to 30 gigahertz.

When a vehicle passes through the antenna beam, a portion of transmitted energy is reflected back towards to antenna. The energy then enters a receiver. Receiver is a tool where the detection is made and traffic flow data, such as volume, speed and vehicle length calculated.

In roadside applications; two types of microwave radar sensors are used. They are “continuous *wave (CW) Doppler waveforms*” and “*frequency modulated continuous waves (FMCW)*”. First of them is used to measure only number of vehicles and vehicular speed, but it cannot detect stopped vehicles. In contrast to *CW*, *FMCW* can detect stopped vehicles. Also occupancy, number of vehicles, speed and length of vehicles are measured by *FMCW*. Moreover, it can classify vehicles, too (Traffic Detector Handbook, 2006). *FMCW* is the radar technique to be used for this study.

2.3.2 Measurements over a short section

Early studies utilized a second pneumatic tube, used a second pneumatic tube, placed very close to the first, to acquire speeds. Nowadays, to obtain speeds, paired presence detectors such as inductive loops dispersed maybe five to six meters separated. If the distance between two lines is known, also video camera technology is useful to detect speeds.

Inductive loops or microwave beams take up space on the road. For this reason, they can evaluate as a short section measurements. In addition to this, these detectors produce a new variable; occupancy. Occupancy is defined as the percentage of time that the detection zone of the instrument is occupied by a vehicle. This variable is available because the loop gives a continuous reading (at 50 or 60 Hz usually), which pneumatic tubes or manual counts could not do. Because occupancy depends on the size of the detection zone of the instrument, the measured occupancy may differ from site to site for identical traffic, depending on the nature and construction of the detector.

As with point measurements, short-section data acquisition does not permit direct measurement of density (FHWA DOT,2006).

2.3.3 Measurement along a length of road

Measurements along a length of road come either from aerial photography, or from cameras mounted on tall buildings or poles. It is suggested that at least 0.5 kilometers (km) of road be observed. On the basis of a single frame from such sources, only density can be measured. The single frame gives no sense of time, so neither volumes nor speed can be measured. (FHWA DOT, 2006)

2.3.4 Moving observer method

There are two approaches to the moving observer method. The first is a simple floating car procedure in which speeds and travel times are recorded as a function of time and location along the road. While the intention in this method is that the floating car behaves as an average vehicle within the traffic stream, the method cannot give precise average speed data. It is, however, effective for obtaining qualitative information about freeway operations without the need for elaborate equipment or procedures. One form of this approach uses a second person in the car to record speeds and travel times. A second form uses a modified recording speedometer of the type regularly used in long-distance trucks or buses. One drawback of this approach is that it means there are usually significantly fewer speed observations than volume observations (FHWA DOT, 2006). Figure 2.4 represents the four methods to obtain traffic data.

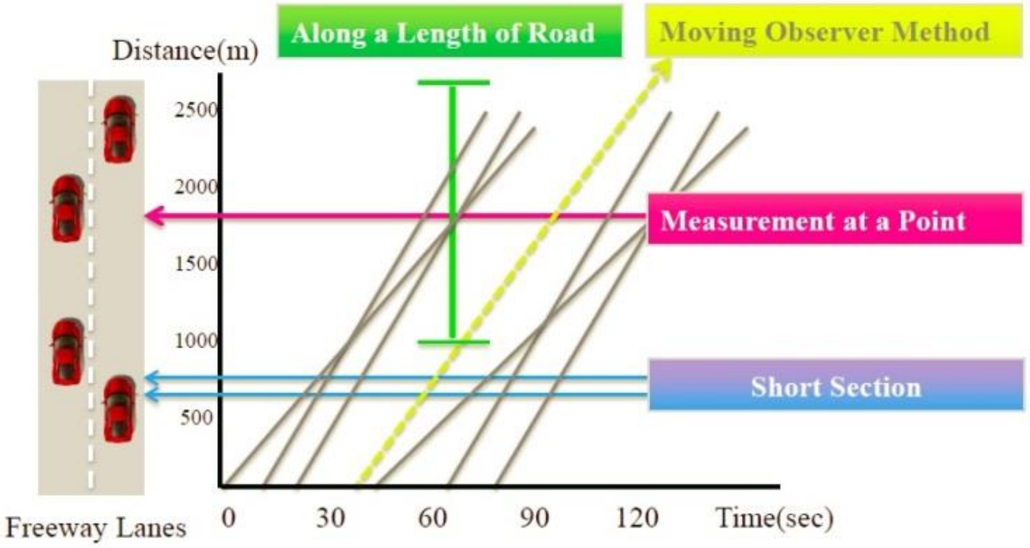

Figure 2.4 : Time- Space Diagram and 4 types of measurement.

2.3.5 Intelligent transportation system wide- area measurements

Wide-area samples obtained simultaneously from a number of vehicles method looks like the use of an observer moving a traffic stream.

2.4 Fundamental Relationships of Traffic Flow

In this section, the relationship between speed and density, is used as concentration in some works, is examined. It is worthy to note that Equation **(2.3)** is known as equation of fundemental relationship among traffic flow variables (May, 1990). Equation **(2.3)** can be also written as;

$$q = u * k \tag{2.6}$$

In this equation, if k is written as a function of u or the way around, i.e., u is written as a function of k, it can be seen that relationships between q-k and q-u are nonlinear. In spite of that, the relationship between u- k is easier.

Flow diagrams related to traffic flow variables are shown in Figure 2.5. A linear speed-density relationship is assumed to simplify this presentation of traffic flow fundamentals.

Figure 2.5 shows the relationship among traffic flow variables.

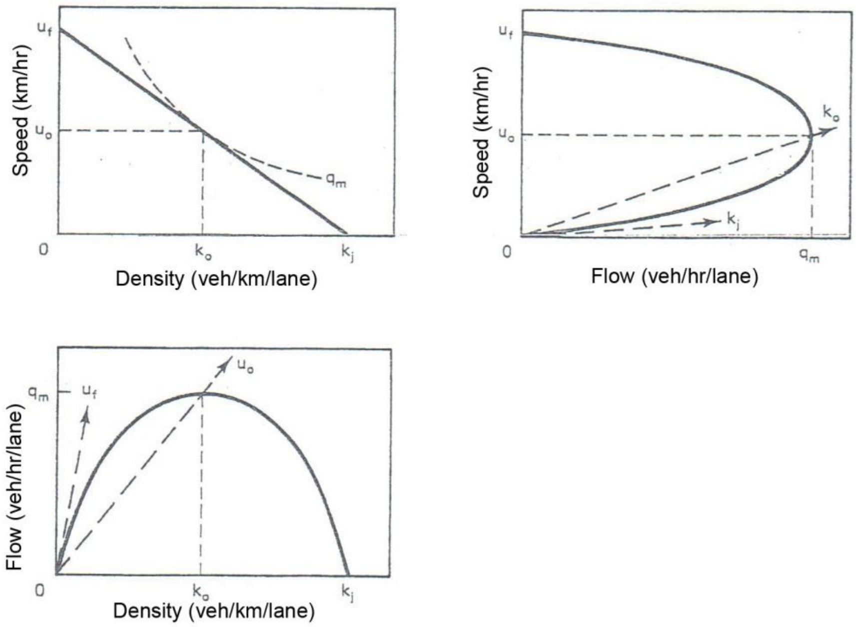

Figure 2.5 : Relationships between traffic flow variables (May, 1990).

Consider the speed-density relationship shown in the upper-left corner of Figure 2.5. As mentioned earlier a linear speed-density relationship is assumed to simplify the presentation. The equation for this relationship is given as Equation **(2.7)**:

$$u = u_f - \left(\frac{u_f}{k_j}\right)k \quad \textbf{(2.7)}$$

This relationship indicates that speed approaches free-flow speed (u_f) when density (and flow) approach zero ($k \rightarrow 0$ and $q \rightarrow 0$). As density (and flow) increases, speeds are reduced until flow is maximum (q_m), and speed and density approach their optimum values ($u \rightarrow u_o$ and $k \rightarrow k_o$). Further increases in density result in lower speeds (and lower flows) until density reaches its maximum value (k_j) and correspondingly speed (and flow) approach zero ($u \rightarrow 0$ and $q \rightarrow 0$). Note that flows can be represented on the speed-density diagram as contour lines with the maximum flow contour (q_m) just touching the speed-density line at optimum values of speed and density (u_o and k_o)(May, 1990).

The flow-density relationship is shown directly below the speed-density relationship in Figure 2.5 because of their common horizontal scales. The equation for this relationship can be derived by substituting *q/k* for *u* (based on Equation **(2.6)**) into Equation **(2.7)**, and solving for flow (*q*). The resulting flow-density equation is

$$q = u_f k - \left(\frac{u_f}{k_j}\right) k^2 \tag{2.8}$$

When density and flow approach zero, ($k \to 0$, $q \to 0$) and speed approaches free-flow speed ($u \to u_f$). As flow increases, density increases while speed is decreasing. When optimum density is reached, flow becomes maximum. Further increases in density result in decreased flow until finally, as jam density is reached, flow approaches zero. Note that speeds can be represented on the flow-density diagram as radial lines extending up to the right from the origin. Steeper- sloped lines represent higher speeds; that is, a vertical line represents a speed of infinity while a horizontal line represents a speed of zero. The slope of the flow-density curve is zero when maximum flow occurs (May, 1990). Therefore, the relationship between optimum and jam density can be determined in the following manner beginning with the Equation (2.8). Since $dq/dk=0$ when $k \to k_0$ and optimum density can be given as Equation (2.9):

$$k_o = \frac{k_j}{2} \tag{2.9}$$

It should be observed that this relationship is true only when the speed-density relationship is assumed linear. When nonlinear speed-density relationships are encountered, the procedures described above for relating optimum and jam density can be applied.

The speed-flow relationship is shown directly to the right of the speed-density relationship in Figure 2.5 because of their common vertical scales. The equation for this relationship can be derived by substituting *q/k* for *k* (based on Equation (2.6)) into Equation (2.7), and solving for speed (*u*). The resulting speed-flow equation is given as Equation (2.10):

$$u = u_f - \left(\frac{u_f}{k_j}\right)\frac{q}{u} \tag{2.10}$$

Solving for q, Equation **(2.10)** becomes as Equation **(2.11)**;

$$q = \frac{k_j}{u_f}\left(u_f u - u^2\right) \quad \textbf{(2.11)}$$

The upper part of the speed-flow curve is described as the free-flow regime and the lower part is mentioned as the congested flow regime. Under free-flow conditions, the speed decreases as the flow level increases up to the maximum flow. Further speed reductions coupled with flow reductions are encountered when density exceeds optimum density. The lower limb of the curve depicts this congested flow situation. Note that densities can be presented on the speed-flow diagram as radial lines extending up to the right from the origin. Steeper- sloped lines represent lower densities; that is, a vertical line represents a density of zero while a horizontal line represents a density of infinity (May, 1990). The relationship between optimum and free-flow speed can be determined by using Equation **(2.11)**. In Equation **(2.11)**, if optimum speed is written instead of speed, the equation becomes Equation **(2.12)**;

$$q_m = \frac{k_j}{u_f}\left(u_f u_o - u_o^2\right) \quad \textbf{(2.12)}$$

Substituting u_o, k_o for q_m and $2k_o$ for k_j, Equation **(2.12)** becomes as Equation **(2.13)**;

$$u_o k_o = \frac{2k_o}{u_f}\left(u_f u_o - u_o^2\right) \quad \textbf{(2.13)}$$

If Equation **(2.13)** is solved for u_o, optimum speed can be determined as Equation **(2.14)**;

$$u_o = \frac{u_f}{2} \quad \textbf{(2.14)}$$

Given Equation **(2.9)** and Equation **(2.14)**, the maximum flow (capacity) can be determined from free-flow speed and jam density as Equation **(2.15)**

$$q_m = \frac{u_f k_j}{4} \quad \textbf{(2.15)}$$

Again, it is important to remember that Equations **(2.9)**, **(2.14)**, and **(2.15)**are all based on a linear speed-density relationship.

The three diagrams shown in Figure 2.5 are redundant, for it is obvious that if any one relationship is known, the other two are uniquely defined. However, all three relationships are shown because each has a particular purpose and use. (May, 1990)

2.5 Traffic Stream Models

It is noticeable fact that drivers decrease their speeds when the number of cars around them increases. That is why early investigators tried to explore the relationship between speed and density. In addition to this, the relationship is classified into two types of models; Single Regime Models and Multi Regime Models. In the following Single Regime Models is initially explained.

The first single regime model was suggested by Greenshields. The model was based on measuring speed- density measurements derived from an aerial photographic study. Greenshields (1935) concluded that speed was a linear function of density. Equation **(2.16)** shows the relationship between speed and density.

$$u = u_f \left(1 - \frac{k}{k_j}\right) \qquad \textbf{(2.16)}$$

Where u_f is the free-flow speed and k_j is the jam density.

It is the second single regime model which was suggested by Greenberg. The model was based on measuring speed- density data for tunnels. In contrast to Greenshields, Greenberg (1959) concluded that a nonlinear model might be useful to explain the relationship among speed and density. Equation (2.17) shows that this nonlinear relationship.

$$u = u_o \ln\left(\frac{k_j}{k}\right) \qquad \textbf{(2.17)}$$

Where u_o is the optimum speed and k_j is the jam density.

The third model was suggested by Underwood (1961) by using the data of traffic studies on the Meritt Parkway in Connecticut. Equation (2.18) shows Underwood's speed density model:

$$u = u_f e^{-\frac{k}{k_0}} \qquad \textbf{(2.18)}$$

Where u_f is the free-flow speed and k_o is the optimum density.

The models were provided new perspective for describing the relationship among speed- density, especially for linear model (Pipes and Munjal, 1971). Equation **(2.19)** shows the relationship explained above.

$$u = u_f \left[1 - \frac{k}{k_j}\right]^n \tag{2.19}$$

Where n is a real number which is greater than zero. This model can be classified into three . These parts are illustrated in Figure 2.6. It is obvious that if n equals 1, the relationship turns into Greenshields' model.

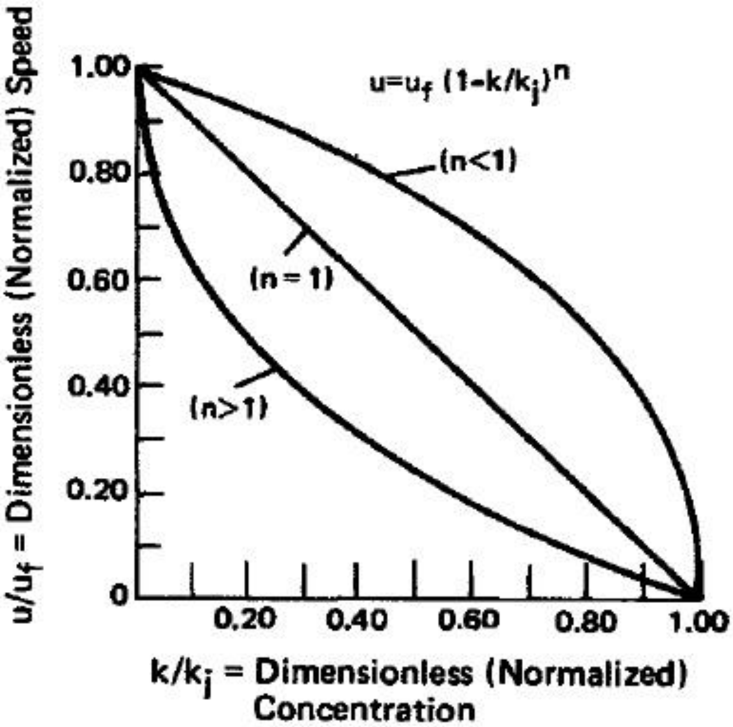

Figure 2.6 : Speed- Density models suggested by Pipes-Munjal (1971).

Drew (1965) suggested a family of models. It is formed in Equation (2.20)

$$\frac{du}{dk} = u_o k^{\frac{(n-1)}{2}} \tag{2.20}$$

Where n is a real number. Equation (2.20) can be solved to yield Greenberg's model. Figure 2.7 shows the results of three values of n, i.e.-1,0, and+1.

In Figure 2.7, u_m, k_j denotes respectively the optimum speed(km/h) and jam density (veh/km).

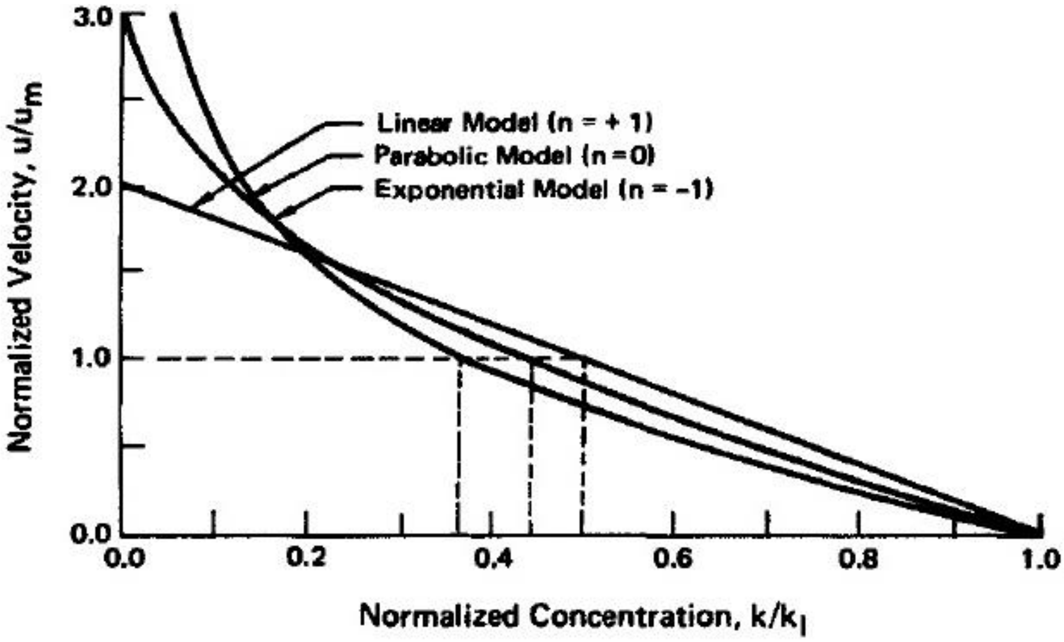

Figure 2.7 : Illustration of Drew's family of speed-density models (Drew, 1965).

Drake et al. (1967) suggested bell-shaped or normal curve as a model of speed-density using the form given in Equation **(2.21)**;

$$u = u_f e^{-\frac{1}{2}(\frac{k}{k_o})^2} \quad \textbf{(2.21)}$$

Equation **(2.21)** seems related to Underwood model. Like Underwood model, this model also requires free-flow speed and optimum density.

It is known that Greenberg's model is useful for representing traffic flow in high density (Gerlough and Huber, 1975). However, the model is not useful for low density. In contrast, Underwood model is useful for low density but not for high density. For this reason, in this part, there are some models to explain, which are called as Multi Regime models.

Edie (1963) suggested a model that is a mixture of Greenberg's and Underwood's model. If there is a high density, Greenberg's model is applicable. If there is a low density, Underwood's model is useful.

It is seen that in Figure 2.8, when normalized speed is plotted against normalized density or concentration, the two models become tangent in the midrange of density. Supporting the idea of the use of multi regime models, Northwestern University research team suggested three additional model formulations.

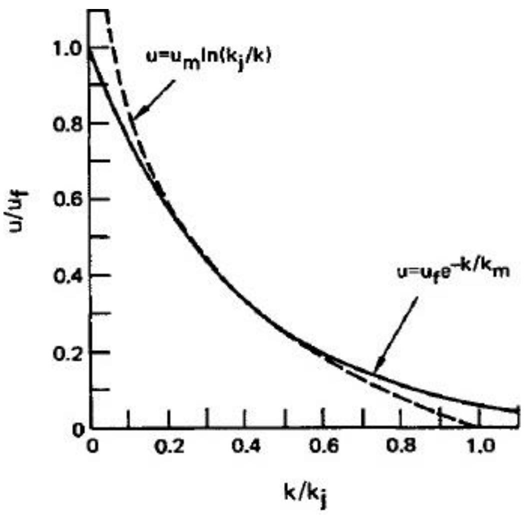

Figure 2.8 : Normalized speed vs. normalized concentration for Greenberg and Underwood model (Edie, 1961).

First of them was the use of Greenshields' model or free-flow regime and congested flow regime, separately. Second of them was to apply a constant-speed model for free flow regime and Greenberg's model for congested-flow regime. Third and last of them was to apply Greenshields' formulation to free-flow regime, transitional-flow regime and congested-flow regime, one by one (May, 1990).

Figure 2.9 shows the application of all single models on the *u*- *k* diagram. (May, 1990).

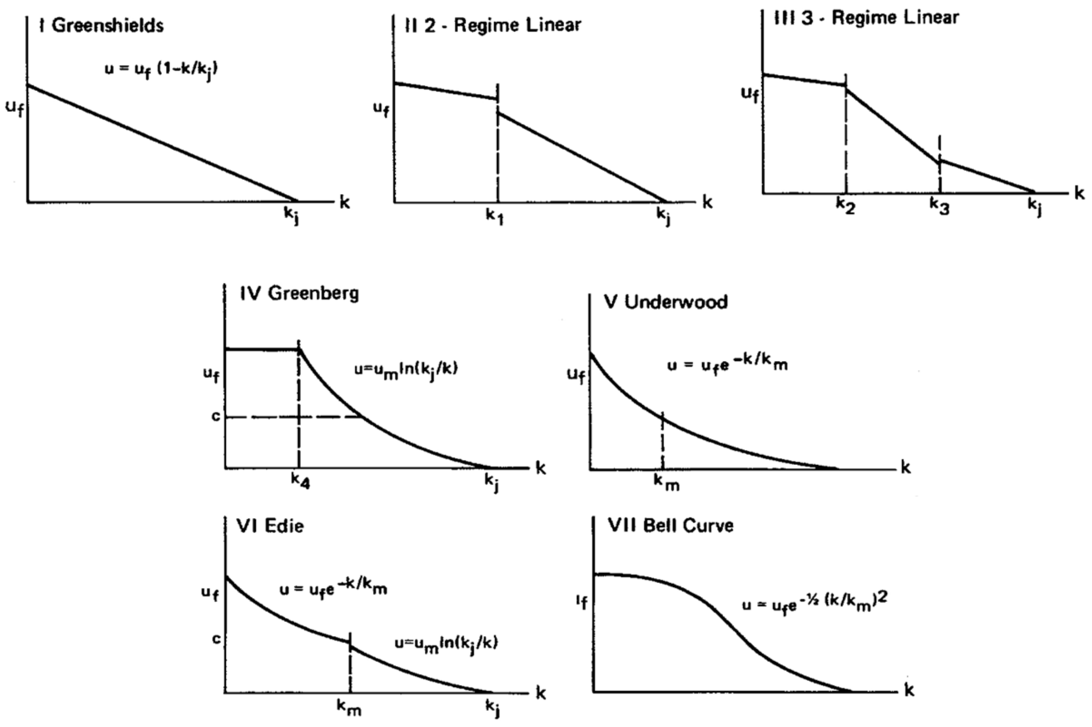

Figure 2.9 : All single regime models on u-k diagram (May, 1990).

Great usefulness of the flow- density curve in traffic control situations, Haight (1971) termed the flow- density curve "the basic diagram of traffic". Also most of flow-density models have been derived from speed- density models. For this reason, model which was derived from directly flow and density data is quite a little. (Gartner et al., 2001)

Some important features of flow- density diagram was defined by Gerlough and Huber (1975) as it is referred below;

1) If there is an absence of density, there can be no flow. Therefore, the curve must pass through the origin. If the space- mean speed is used as the ratio q/k, the slope of the curve at origin is the free-flow speed.

2) Observations showed that, if there is a high density, vehicles had strong tendency to stop. Also it made the value of flow zero. In this situation, the curve must have a point representing jam (maximum) density with zero flow.

3) Insomuch as there are noticeable flows at intermediate density values, there must be one or two more points of maximum flow between the two zero points.

4) It is not necessary for *q*-*k* curve to be continuous.

Table 2.1 represents the models which are derived from speed- density models. q_m, k_m, u_m, u_f, k_j denotes respectively maximum flow, maximum density, maximum speed, free-flow speed and jam density.

Table 2.1 : Flow- Density Models.

Flow – Density Model	Equation of Model	Special Points
Greenshields	$q = u_f k - \left(u_f k^2 / k_j\right)$	$k_m = \frac{k_j}{2}; u_m = \frac{u_f}{2}; q_m = u_f k_j / 4$
Greenberg	$q = k u_m \ln(\frac{k_j}{k})$	$k_m = \frac{k_j}{e}; u_m = u_m; q_m = u_m k_j / e$
Underwood	$q = k u_f e^{-(\frac{k}{k_m})}$	$k_m = k_m; u_m = u_f / e; q_m = u_f k_m / e$

2.6 Analogy of Fluid Dynamics and Kinematic Waves with the Traffic Flow

In the literature, different approaches are followed to study traffic flow propagation, depending on whether the link performances are expressed in an aggregate or a disaggregate way, and how vehicles' movements are traced (Celikoglu and Dell'Orco, 2007). The simulation is macroscopic if vehicle movements are traced implicitly, and the link performance is expressed in an aggregate way. In the following, the macroscopic theory of traffic flow modeling is presented referring to some mile-stone studies and considering widely accepted models, as well as the cell transmission model of Daganzo (1994) that is adopted to simulate the freeway test segment in the present study.

2.6.1 Conservation Law

In previous chapter, speed-density and flow-density relationships have been discussed. As it is known, the identity of traffic flow can be expressed in Equation (2.6). By deriving from (2.6), it can be said that it is an identity. It means it is self-guaranteed by generalized definition of traffic flow characteristics. Also, it is location and time specific. Equation (2.22) defines what the location and time specific means.

$$q(x,t) = k(x,t) * u(x,t), \tag{2.22}$$

The pair-wise relationship between speed-density and flow-density are expressed as given in **(2.23)**.

$$\begin{aligned} u &= f(k) \\ q &= f(u) \end{aligned} \tag{2.23}$$

A few comments on these relationships follow (Ni, 2013):

1- They define a fundamental diagram and therefore make different vehicular traffic flow from other kinds of flows;

2- They are location-specific;

3- They are equilibrium models;

4- They are deterministic, but such functions are only of statistical significance.

The main purposes of formulating a traffic flow theory are based on two benefits; to help better understand traffic flow and, by applying such knowledge, to control traffic for safer and more efficient operations. Therefore, a good theory must help to find answers the following questions:

- In the light of past information about traffic conditions, how do the road conditions change over time?
- Is there any bottleneck?
- In case of congestion, how long does it last?
- If an incident occurs, what is the best strategy to clean up in order that the impact on traffic is minimized?

Answers to these questions comprise analyzing dynamic change of traffic states over time and space. Unfortunately, the above relationships or models are only capable of describing traffic states without providing a mechanism to analyze how such states evolve. Starting from this chapter, dynamic models will be introduced to address these questions.

The derivation of a dynamic equation starts with analyzing a small volume of roadway traffic as a continuum. In here traffic flow is handled as a one-dimensional compressible fluid such as gas. Laws of conservation apply to this kind of fluid and the continuity equation, also known as first-order form of conservation is the mass conservation (Ni, 2013).

2.6.2 The Continuity equation

There are several ways to derive the continuity equation, each takes a different perspective on the small volume of roadway traffic.

2.6.2.1 Derivation I: finite difference

Gartner et al. (2001) utilized derivation on a scenario. Suppose a highway section is depicted by two observation stations at x_1 and x_2 and the segment length is $\Delta x = x_2 - x_1$. During time interval $\Delta t = t_2 - t_1$, N_1 vehicles passed x_1 and N_2 vehicles passed x_2. Figure 2.10 sketches a highway section which expressions are given above. It is known that inflow must be equal to outflow. The equations which are below are related to what it is tried to explain this part. Also the principle of not to destroy or lose vehicles in section is used in this equation.

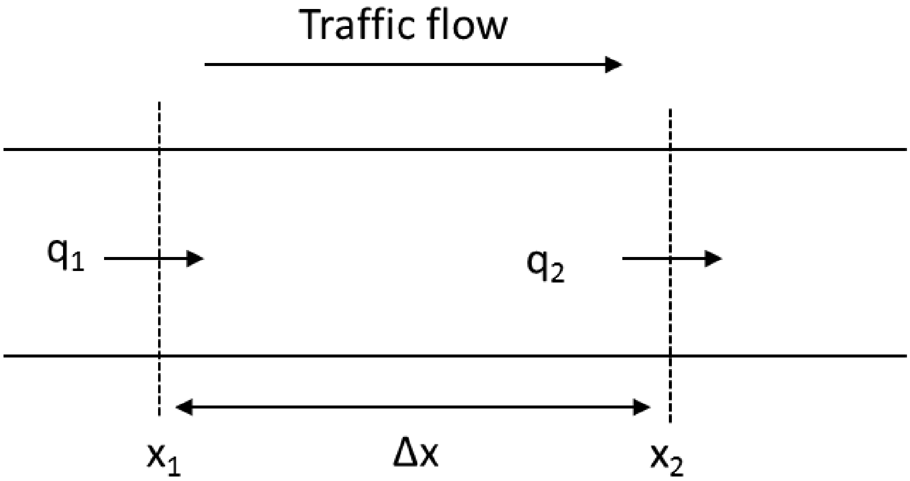

Figure 2.10: Deriving continuity equation I.

For this reason, the flow rates at these locations are:

$$q_1 = \frac{N_1}{\Delta t} \ and \ q_2 = \frac{N_2}{\Delta t} \tag{2.24}$$

The change of vehicles in the section is;

$$\Delta N = N_2 - N_1 = (q_2 - q_1)\Delta t = \Delta q \Delta t \tag{2.25}$$

Presume respectively traffic densities in the section at t_1 and t_2 are k_1 and k_2. On account of that, there are $M_1 = k_1\Delta x$ vehicles in the section at time t_1 and $M_2 = k_2\Delta x$ vehicles in the section at time t_2. The change of vehicles in the section can be expressed as:

$$\Delta M = k_1\Delta x - k_2\Delta x = (k_1 - k_2)\Delta x = -\Delta k \Delta x \tag{2.26}$$

Vehicles cannot be created or destroyed inside the section, the change of vehicles must be the same in the same section during the same time interval. For this reason;

$$\begin{aligned} \Delta N &= \Delta M \\ \Delta q \Delta t &= -\Delta k \Delta x \\ \Delta q \Delta t + \Delta k \Delta x &= 0 \end{aligned} \tag{2.27}$$

If both sides are divided by $\Delta x \Delta t$,

$$\frac{\Delta q}{\Delta x} + \frac{\Delta k}{\Delta t} = 0 \quad \textbf{(2.28)}$$

Let $\Delta x \rightarrow 0$ and $\Delta t \rightarrow 0$, the above difference equation becomes a partial differential equation:

$$\frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = 0 \quad \textbf{(2.29)}$$

The above equation can be abridged as

$$q_x + k_t = 0 \quad \textbf{(2.30)}$$

where $q_x = \frac{\partial q}{\partial x}$ and $k_t = \frac{\partial k}{\partial t}$.

2.6.2.2 Derivation II: finite difference

The derivation is basically the same as Derivation I. However, it is presented in a little bit different way. Figure 2.11 sketches a highway section $\Delta x = x_2 - x_1$ during time interval $\Delta t = t_2 - t_1$ At time t_1, there are N_1 vehicles in the section and at time t_2, there are N_2 vehicles in the section. In the course of the period, traffic keeps flowing into the section at rate q_1 and flowing out at rate q_2.

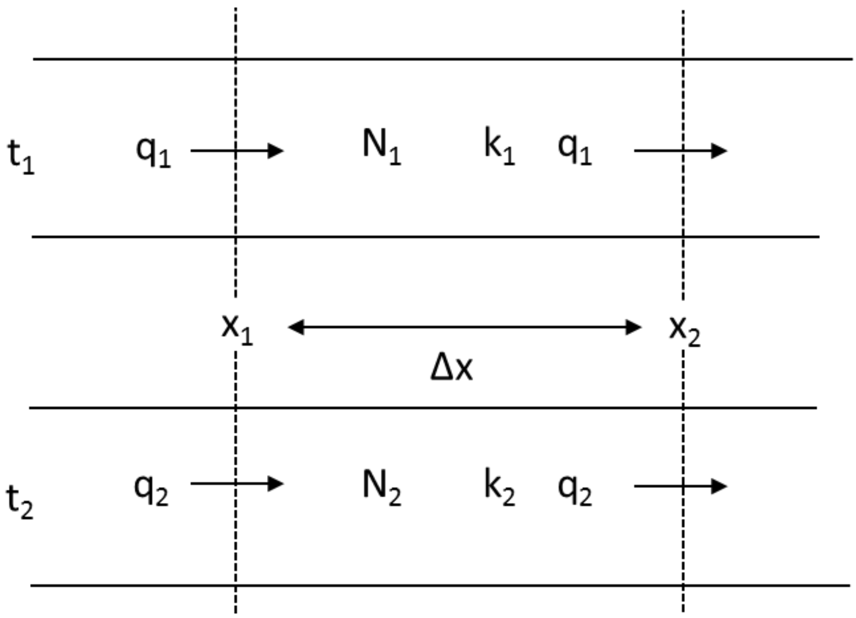

Figure 2.11 : Deriving contunity equation II.

Based on vehicle conservation, the following relationship holds:

Vehicles at $t_2 = vehicles\ at\ t_1 + inflow\ during\ \Delta t - outflow\ during\ \Delta t.$

This can be shown as;

$$N_2 = N_1 + q_1\Delta t - q_2\Delta t \quad \textbf{(2.31)}$$

It is essential to note that $N = k\Delta x$, so the Equation 2.22 becomes

$$k_2\Delta x = k_1\Delta x + q_1\Delta t - q_2\Delta t \quad \textbf{(2.32)}$$

After arranging terms and dividing both sides by $\Delta x\Delta t$

$$\frac{k_2 - k_1}{\Delta t} = -\frac{q_2 - q_1}{\Delta x} \quad \textbf{(2.33)}$$

Let $\Delta x \to 0$ and $\Delta t \to 0$,

$$q_x + k_t = 0 \quad \textbf{(2.34)}$$

2.6.2.3 Derivation III: fluid Dynamics

Figure 2.12 illustrates a small fluid cube of size $\delta x * \delta y * \delta z$. Fluid velocity v and density k at two sides of the cube also are shown.

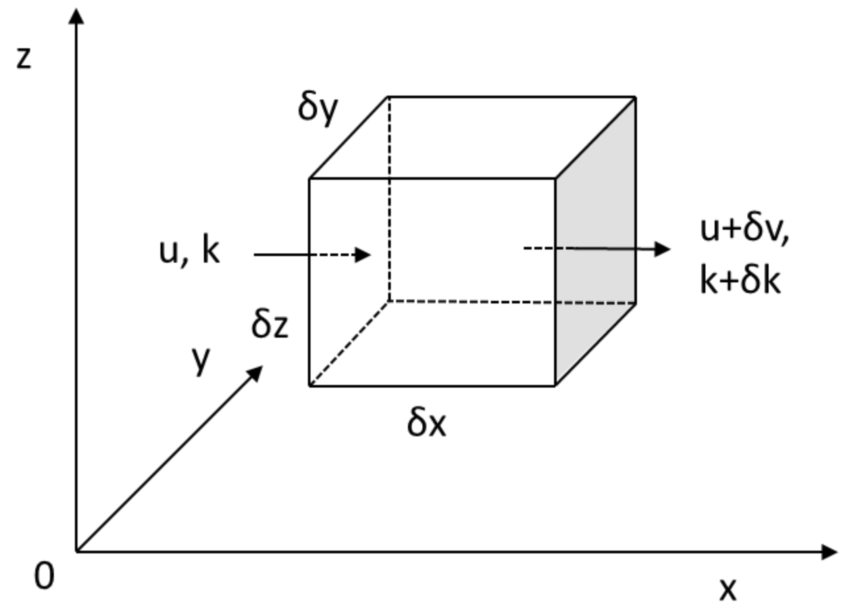

Figure 2.12 : Deriving continuity equation.

$$Mass\ flows\ into\ the\ cube\ = uk\delta y\delta z \quad \textbf{(2.35)}$$

$Mass\ flows\ out\ of\ the\ cube \quad = \quad (u+\delta u)(k+\delta k)\delta y\delta z = (u+\frac{\partial u}{\partial x}\delta x)(k+\frac{\partial k}{\partial x}\delta x)\delta y\delta z$=$(uk+u\frac{\partial k}{\partial x}\delta x+k\frac{\partial u}{\partial x}\delta x+\frac{\partial u}{\partial x}\frac{\partial k}{\partial x}\delta x\delta x)\delta y\delta z$

Mass stored in the cub= mass that flows in-mass that flows out **(2.36)**

$$u\frac{\partial k}{\partial x}\delta x+k\frac{\partial u}{\partial x}\delta x+\frac{\delta u}{\delta x}\frac{\partial k}{\partial x}\delta x\delta x)\delta y\delta z = (u\frac{\partial k}{\partial x}+k\frac{\partial u}{\partial x}+\frac{\partial u}{\partial x}\frac{\partial k}{\partial x}\delta x)\delta x\delta y\delta z \quad \textbf{(2.37)}$$

Take no notice of the high-order term, equation becomes;

$$= \left(u\frac{\partial k}{\partial x}+k\frac{\partial u}{\partial x}\right)\delta x\delta y\delta z = \frac{\partial(ku)}{\partial x}\delta x\delta y\delta z \quad \textbf{(2.38)}$$

Similar procedure applies to the other two directions of the cube. For this reason, the total mass stored in the cube is:

$$(\frac{\partial(ku)}{\partial x}+\frac{\partial(kv)}{\partial y}+\frac{\partial(k\omega)}{\partial z})\delta x\delta y\delta z \quad \textbf{(2.39)}$$

Mass stored in the cube must be balanced by the change of mass in the cube:

$$\frac{\partial k}{\partial t}\delta x\delta y\delta z \quad \textbf{(2.40)}$$

The law of mass conservation requires that

$$\left(\frac{\partial(ku)}{\partial x}+\frac{\partial(kv)}{\partial y}+\frac{\partial(k\omega)}{\partial z}\right)\delta x\delta y\delta z+\frac{\partial k}{\partial t}\delta x\delta y\delta z = 0 \quad \textbf{(2.41)}$$

Therefore

$$\frac{\partial k}{\partial t}+\left(\frac{\partial(ku)}{\partial x}+\frac{\partial(kv)}{\partial y}+\frac{\partial(k\omega)}{\partial z}\right) = 0 \quad \textbf{(2.42)}$$

Highway traffic constitutes a special case of the above with only one dimension, see Figure 2.13 Using the result derived above, one obtains:

$$\frac{\partial(ku)}{\partial x} + \frac{\partial k}{\partial t} = 0 \qquad \textbf{(2.43)}$$

Note that;

$$q = ku \qquad \textbf{(2.44)}$$

Therefore

$$q_x + k_t = 0 \qquad \textbf{(2.45)}$$

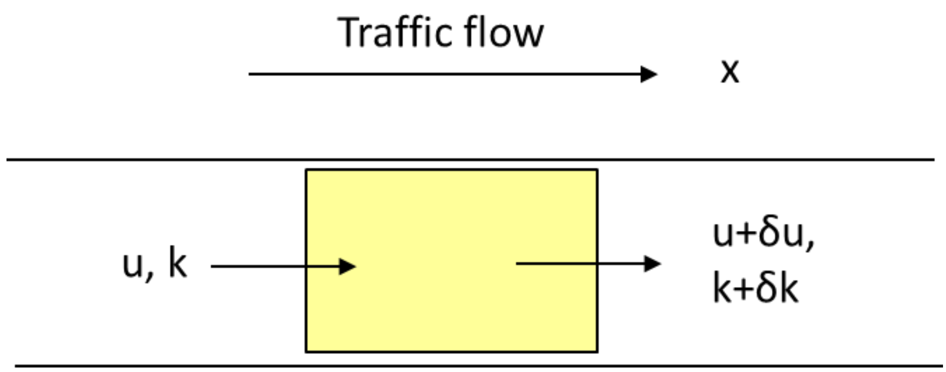

Figure 2.13 : Reducing 3D to 1D.

2.6.2.4 Derivation IV: scalar conservation law

This derivation is adopted from reference Illner et al. (2005). Consider a cell in the time-space area restricted by $(x_1, x_2) * (t_1, t_2)$. Figure 2.14 represents that area on time and space axis. Let traffic flow, speed, and density be functions of time and space, It means that, $q = q(x,t)$, $u = u(x,t)$, and $k = k(x,t)$. It is obvious that, the conservation of vehicles in the cell requires the following:

$$\int_{x_1}^{x_2} k(x,t_2)dx - \int_{x_1}^{x_2} k(x,t_1)dx = \int_{t_1}^{t_2} q(x_1,t)dt - \int_{t_1}^{t_2} q(x_2,t)dt \qquad \textbf{(2.46)}$$

$$\int_{x_1}^{x_2} [k(x,t_2) - k(x,t_1)]dx = \int_{t_1}^{t_2} [q(x_1,t) - q(x_2,t)]\, dt \qquad \textbf{(2.47)}$$

If *k(x,t)* and *q(x,t)* are differentiable in *x* and *t*, one obtains:

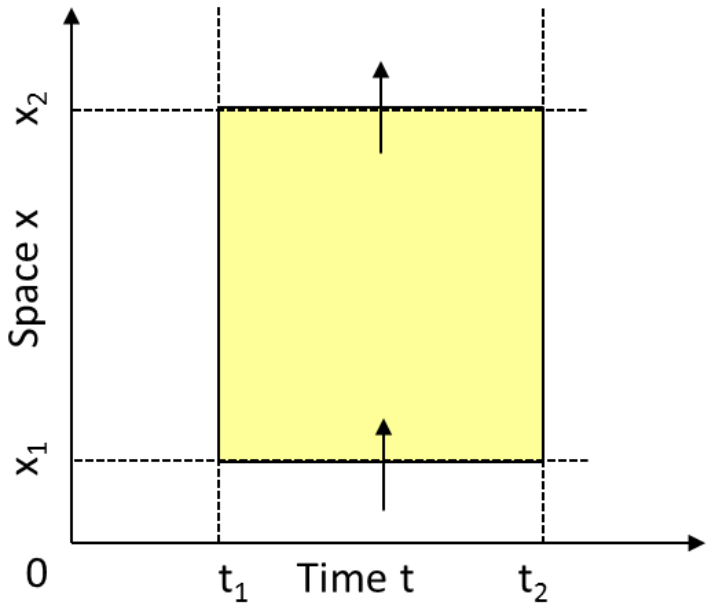

Figure 2.14 : Deriving continuity equation IV.

$$\int_{x_1}^{x_2}\int_{t_1}^{t_2}\frac{\partial k(x,t)}{\partial t}\,dt\,dx = -\int_{t_1}^{t_2}\int_{x_1}^{x_2}\frac{\partial q(x,t)}{\partial x}\,dx\,dt \qquad \textbf{(2.48)}$$

$$\int_{x_1}^{x_2}\int_{t_1}^{t_2}\left[\frac{\partial k(x,t)}{\partial x}+\frac{\partial q(x,t)}{\partial x}\right]dx\,dt = 0 \qquad \textbf{(2.49)}$$

According to the fundamental theorem of calculus of variables, it is acquired

$$\frac{\partial k(x,t)}{\partial t}+\frac{\partial q(x,t)}{\partial x} = 0 \qquad \textbf{(2.50)}$$

i.e.

$$q_x + k_t = 0 \qquad \textbf{(2.51)}$$

2.6.3 First-order dynamic model

Traffic evolution is about the process of how traffic flow characteristics evolve over time *t* and space *x* given some initial and limit conditions. It is obvious that traffic flow characteristics are functions of time *t* and space *x*. Moreover, the continuity equation have been derived to describe the dynamic relation between flow $q(x, t)$ and density $k(x, t)$:

$$q_x + k_t = 0 \quad \textbf{(2.52)}$$

This equation involves two independent variables x and t and two dependent variables $q(x,t)$ and $k(x,t)$. The equation cannot solve. Because the number of unknown variables are greater than the number of equations, i.e. if there are two unknown variables and there is only one equation, there is no way to solve equation. To provide another equation for solving equation is essential. In the present case, the identity turns up:

$$q(x,t) = k(x,t)u(x,t) \quad \textbf{(2.53)}$$

Unfortunately, by adding one equation, a third unknown variable is introduced, i.e. speed $u(x,t)$. Again, another equation is needed to supply. Considering that no better choice is available, an equilibrium traffic flow model, for example Greenshields model, has to be deployed. Such a model takes the form of:

$$u = U(k) \quad \textbf{(2.54)}$$

Putting everything together, a system of three equations including three unknown variables is obtained:

$$\begin{cases} q_x + k_t = 0 \\ q = ku \\ u = U(k) \end{cases} \quad \textbf{(2.55)}$$

Starting from initial and boundary conditions, the above system of equations is solvable. Hereby, traffic characteristics at an arbitrary time space point *(x,t)* can be found.

2.6.4 The Lighthill- Whitham-Richards model

End of first- order dynamical model , a dynamic traffic flow model was formulated based on conservation law:

$$\begin{cases} k_t + q_x = 0 \\ q = kv \\ u = U(k) \end{cases} \quad \textbf{(2.56)}$$

where $q = q(x,t)$ is flow, $k = k(x,t)$ is density, $u = u(x,t)$ is mean traffic speed. If the second and third equations are combined by removing u, a flow-density relationship $q = Q(k)$ is obtained and the dynamic model becomes:

$$\begin{cases} k_t + q_x = 0 \\ q = Q(k) \end{cases} \quad \textbf{(2.57)}$$

Equation 2.48 can be also written as;

$$k_t + Q'(k)k_x = 0 \quad \textbf{(2.58)}$$

where $Q'(k) = \frac{dQ(k)}{dk}$ .This is the so-called LWR model just to honor the three pioneers, Lighthill, Whitham, and Richards, who originally studied this problem. The LWR theory describes the model as a single partial differential equation in conservation form. Note that LWR model is basically a first-order, homogeneous, quasi-linear partial differential equation. In Figure 2.15 shows the Lighthill-Whitham- Richards equation with general flow- density relationship.

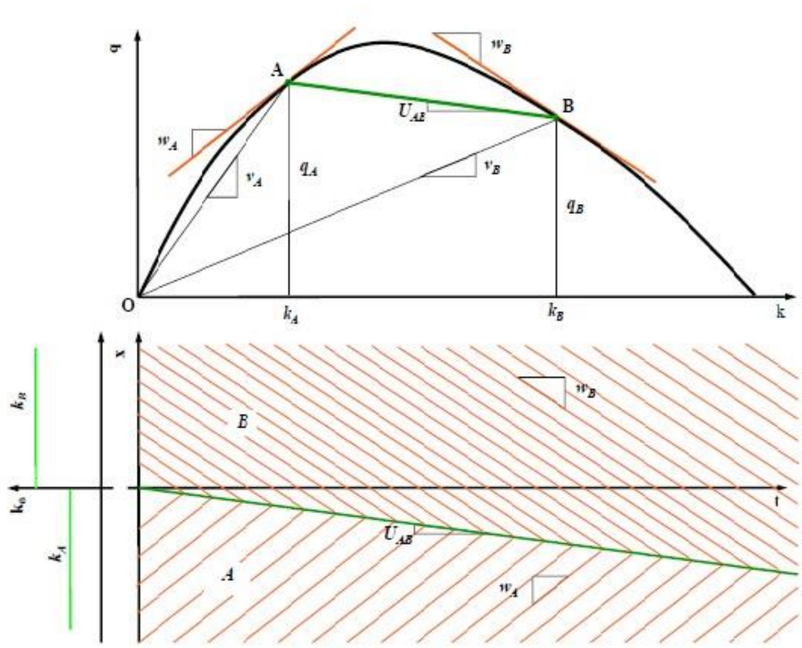

Figure 2.15 : LWR with general q-k relationship (Ni, 2013).

In the above examples, the underlying *q*-*k* relationship is explicitly given, e.g., Greenshields model. Hence, it is convenient to determine the speed of a kinematic wave from the initial condition. However, it is recognized that Greenshields model suffers inexactness and, often times, the underlying *q*-*k* relationship is graphically given by fitting from empirical data. In the circumstances, solution to the LWR model with general *q*-*k* relationship is typically determined graphically. Consider the following LWR model with general *q*-*k* relationship:

$$\begin{cases} k_t + q_x = 0 \\ q = Q(k) \\ k(t,0) = k_0(x) = \begin{cases} A \ if \ x \leq 0, \\ B \ if \ x > 0. \end{cases} \end{cases} \quad \textbf{(2.59)}$$

where the underlying *q*-*k* relationship is given in Figure 2.15 where A denote an operating point characterized by flow q_A, density k_A, and speed v_A and similar notation applies to point B. A time-space diagram is constructed below the *q* - *k* relationship with the initial condition aside.

2.6.5 Cell transmission model

The cell transmission model (CTM) was suggested by Daganzo (1994). The model was presented in two papers with the first addressing mainline traffic and the second network traffic.

2.6.5.1 Minimum principle

The triangular flow-density relationship presented in Figure 2.16. The relationship consists of three sections: uncongested (left) with free-flow speed U_f equal to forward wave kinematic speed ω_f, capacity (middle) q_m, and congested (right) with backward wave speed ω_b and jam density *K*.

A vertical line at any density *k* will intersect the three sections at height $k\omega_f$, q_m, and $(K-k)\omega_f$. Therefore, flow matching up to this density is found as the minimum of the three intersections:

$$q = min[k\omega_f, q_m, (K-k)\omega_b] \quad \textbf{(2.60)}$$

Corporeally, if the left section as conditions dictated by arrival traffic, the middle section as local capacity, and the right section as conditions dictated by downstream traffic is considered, the above equation basically says that traffic flowing through a point of highway should not outpace upstream arrival rate, local capacity, and the rate permitted by downstream conditions.

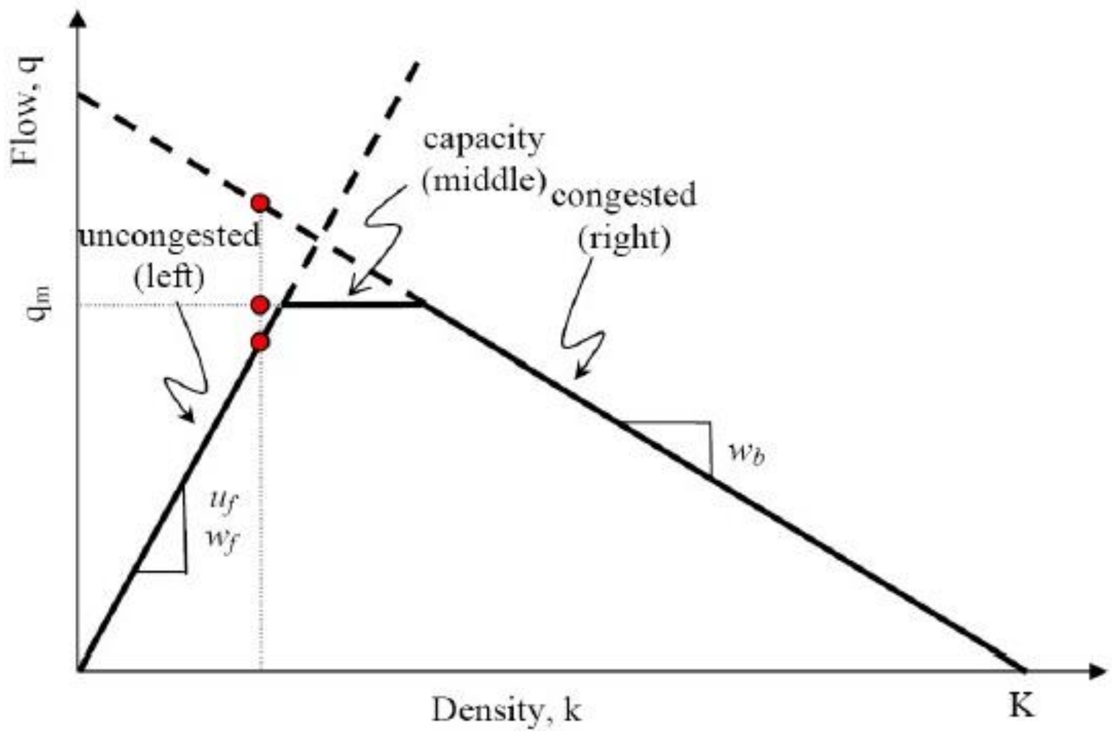

Figure 2.16 : Triangular flow-density relationship for CTM.

A point on the *flow-density* curve defines the operating condition of a stream of traffic. The speed of kinematic wave carried by the traffic, ω, is the tangent to the curve at this point. If the underlying flow-density relationship is triangular, finding kinematic wave speeds is relatively simplified. Primarily, there are only two kinematic wave speeds: a forward wave speed ω_f for all uncongested conditions (the left side of the triangle) and a backward wave speed ω_b for all congested conditions (the rightside). In addition, ω_f happens to be the same as free-flow speed U_f. As a special property of the triangular flow-density relationship, U_f applies to all uncongested conditions.

2.6.5.2 Mainline scenario

CTM uses the same discretization scheme like numerical solutions which are related to focus solving LWR problem. The scheme is demonstrated in Figure 2.17. for convenience.

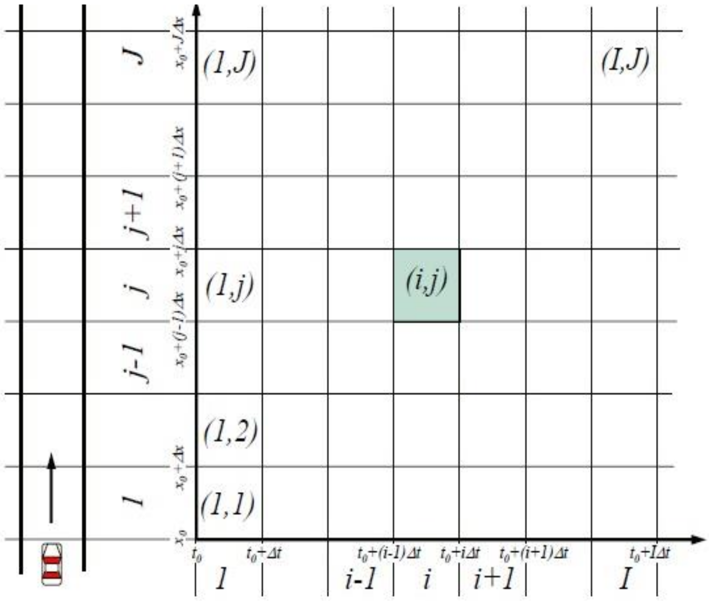

Figure 2.17 : Discretization scheme.

As it is aforementioned earlier, almost all the numerical solutions use the same discretization scheme. However, it is worthy to note that, CTM is seperated from other methods with a feature: the cell has a uniform length as the distance traveled by a vehicle at free-flow speed during one time step:

$$\Delta x = u_f \Delta t \qquad \textbf{(2.61)}$$

With regard to the minimum principle, traffic that can flow into segment *j*, $q_j\ (t_i)$, is constrained by the following:

$$q_j(t_i) = min\left[k_{j-1}(t_{i-1})\omega_f, q_m, (K - k_j(t_{i-1}))\omega_b\right] \qquad \textbf{(2.62)}$$

As a consequence of it, the amount of vehicles that can move into segment *j*, $y_j(t_i)$, is found by multiply both sides by Δt:

$$y_j(t_i) = q_j(t_i)\Delta t = min\left[k_{j-1}(t_{i-1})\omega_f \Delta t, (K - k_j(t_{i-1}))\omega_b \Delta t\right] \qquad \textbf{(2.63)}$$

Note that n = $k\Delta x$, $\Delta x = V_f \Delta t$, and $V_f = \omega_f$ due to triangular flow-density relationship. The above equation can be turned into the following form:

$$y_j(t_i) = min\left[k_{j-1}(t_{i-1})\Delta x, q_m\Delta t, \frac{\omega_b}{\omega_f}(K - k_j(t_{i-1}))\Delta x\right] \quad \textbf{(2.64)}$$

i.e.,

$$y_j(t_i) = min\left[n_{j-1}(t_{i-1}), q_m\Delta t, \frac{\omega_b}{\omega_f}(K\Delta x - n_j(t_{i-1}))\right] \quad \textbf{(2.65)}$$

The above equation provides that the amount of vehicles that can move into segment j, $y_j(t_i)$, is constrained by the following:

- vehicles in j — 1 previously: $n_{j-1}(t_{i-1})$
- the capacity of segment j, $q_m\Delta t$, and
- the empty space in j: j: $\frac{\omega_b}{\omega_f}\left(K\Delta x - n_j(t_{i-1})\right)$.

The equation can be further reduced to

$$y_j(t_i) = \min(S_{j-1}, R_j) \quad \textbf{(2.66)}$$

Where $S_{j-1} = min[n_{j-1}(t_{i-1}), q_m\Delta t]$ represents flow being sent from upstream and

$$R_j = min\left[q_m\Delta t, \frac{\omega_b}{\omega_f}(K\Delta x - n_j(t_{i-1}))\right] \quad \textbf{(2.67)}$$

is flow ready to be received by downstream. Therefore, the evolution of traffic on a freeway mainline can be stated as:

Storage in current cell=Storage in the cell previously+Vehicle flowed in-Vehicles flowed out **(2.68)**

Mathematically, this can be stated as

$$n_j(t_i) = n_j(t_{i-1}) + y_j(t_{i-1}) - y_{j+1}(t_{i-1}) \quad \textbf{(2.69)}$$

3. ISSUES RELEVANT TO TRAFFIC FLOW STATES

Traffic flow state, or alternatively pattern, specification can be described as estimating the traffic flow variables along a road stretch with an adequate spatial resolution at each time instant based on a limited amount of available measurements from detectors, where the sought state variables correspond to variable of traffic flow, such as the flows, space-mean speeds, and densities (Wang et al., 2007; Celikoglu, 2013)

Noting that we conducted the empirical investigations throughout the case over a basic freeway segment, characteristics of such freeway segments as well as the measure, the level of service concept, to express flow states over these types of freeways are explained in the following.

3.1 Basic Freeway Segments

Basic freeway segments can be defined as the freeway segments that are outside the influence of merging, diverging, or weaving maneuvers. In general, it means that lane-changing activity is not significantly influenced by the presence of ramps and weaving segments. Lane-changing activity generally reflects the normal desire of drivers to optimize their efficiency through lane-changing and passing maneuvers.

There exists a number of road geometry related, traffic flow related and environmental related parameters that affect the ultimate capacity of a basic freeway segment, including weather, visibility, and non-recurrent event occurrence. The deviation on the ideal conditions of these parameters generally results in the reduction on the speed, level of service (LOS), and capacity of the freeway segment.

The specific speed-flow-density relationship of a basic freeway segment depends on prevailing traffic and conditions. Roadway conditions can also be defined as a function of road. The ideal conditions that account for the theoretical capacity of a basic freeway segment should follow (HCM, 2010). As it is indicated in HCM (2010), minimum lane widths must be 3.6 m. Minimum right-shoulder lateral clearance between the edge of the travel lane and the nearest obstacle or object that influences traffic behavior must be 1.8 m. Minimum median lateral clearance must be 0.6 m. Interchange spacing must be 3 km or greater and level terrain, with grades must be no greater that 2 percent. Traffic stream must be composed entirely of passenger cars. A driver population must be composed driver population composed principally of regular users of the facility. These base conditions represent a high operating level, with a free-flow speed of 110 km/h or greater (HCM, 2010).

3.2 Level of Service Concept

"Level of service (LOS) is a qualitative measure that describes traffic conditions in terms of speed, travel time, freedom to maneuver, comfort, convenience, traffic interruptions, and safety". (HCM, 2010)

Density, speed and volume to capacity ratio are the characteristics of performance measurements of a basic freeway segment. Each of these characteristics is an indicator to understand traffic flow behaviour. Since density is accepted to appropriately present the within stream fluctations, it is used as the fundamental measure to quantify the level of service for freeways.

Density can be hardly obtained via measurement. However, it can be calculated using the fundamental relationship of traffic flow, where flow is strictly a function of both the density and the speed of a flow state.

Level of service boundaries in terms of density, veh/km/lane, for the traffic on a basic freeway segment, are given in Table 3.1 (HCM 1985, 1994, 1997, 2000, 2010).

Table 3.1 : Boundaries of existing HCM LOS for freeways.

Level of Service	HCM 1985	HCM 1994/ HCM 1997	HCM 2000/ HCM 2010
A	0-8	0-6	0-7
B	8-13	6-10	7-11
C	13-19	10-15	11-16
D	19-26	15-20	16-22
E	26-42	20-28	22-28
F	>42	Changeable/>28	>28

The specification of maximum densities for LOS A through D is based on the collective professional judgment of the members of the Committee on Highway Capacity and Quality of Service of the Transportation Research Board. The upper value shown for LOS E, 28 pc/km/ln, is the maximum density at which sustained flows at capacity are expected to occur.

3.2.1 Methodology to define level of service

The way to define level of service on a basic freeway segment is visualized in Figure 3.1. The process starts with the data collection and ends up with determining the LOS.

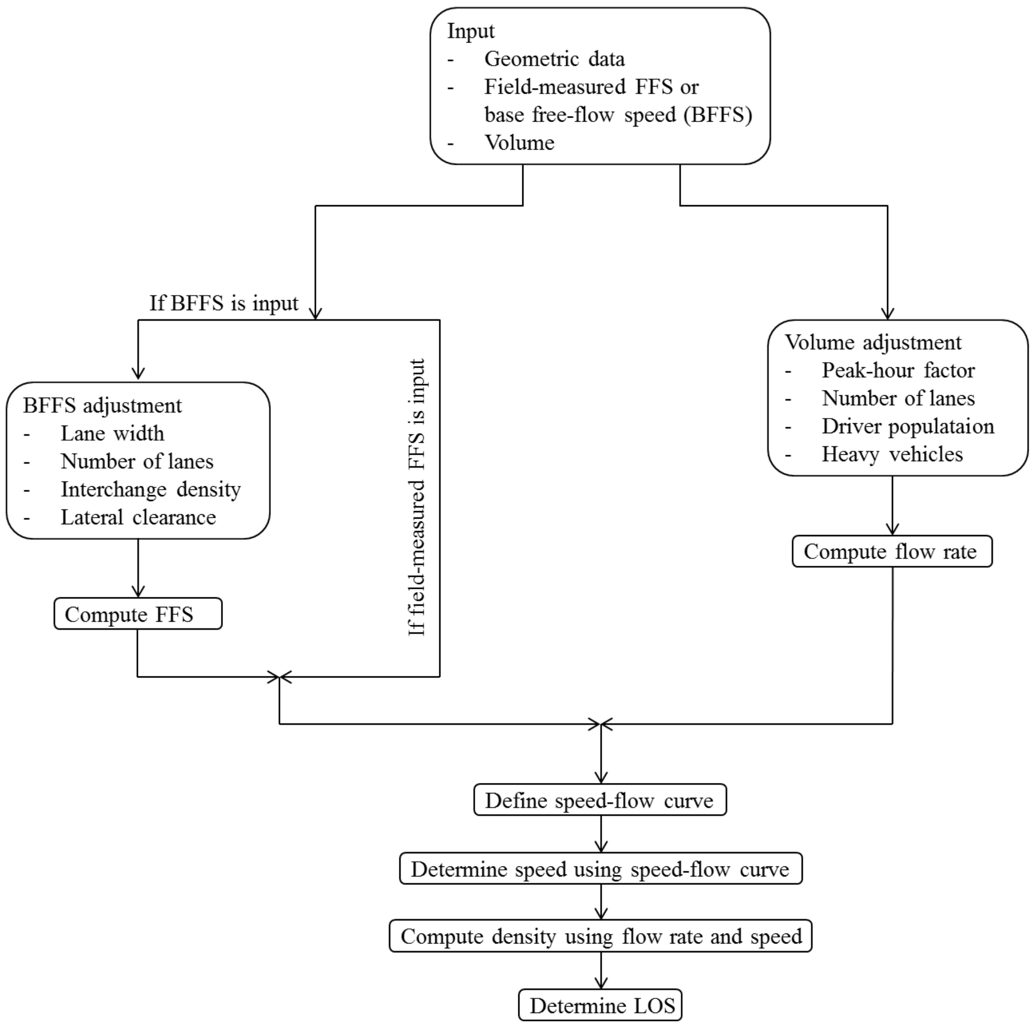

Figure 3.1 : Methodology of Defining LOS (HCM, 2000).

3.2.2 Level of service criteria for basic freeway segments

The LOS criteria,including density measure, for basic freeway segments are given in Table 3.2 with the corresponding free-flow speeds of 120 km/h or greater, 110 km/h, 100 km/h, and 90 km/h. In order to match a given LOS, the density criterion must be satisfied. In effect, under base conditions, these are the speeds and flow rates expected to occur at the density shown for each LOS. It is clear that when, even, the free flow speed values change, the LOS' boundaries do not change.

Table 3.2 : LOS criteria for Basic Freeway Segments (HCM ,2000).

Criteria	LOS				
	A	B	C	D	E
			FFS= 120 km/h		
Maximum density (pc/km/ln)	7	11	16	22	28
Minimum speed (km/h)	120.0	120.0	114.6	99.6	85.7
Maximum v/c	0.35	0.55	0.77	0.92	1.00
Maximum service flow rate (pc/h/ln)	840	1320	1840	2200	2400
			FFS= 110 km/h		
Maximum density (pc/km/ln)	7	11	16	22	28
Minimum speed (km/h)	110.0	110.0	108.5	97.2	83.9
Maximum v/c	0.33	0.51	0.74	0.91	1.00
Maximum service flow rate (pc/h/ln)	770	1210	1740	2135	2350
			FFS= 100 km/h		
Maximum density (pc/km/ln)	7	11	16	22	28
Minimum speed (km/h)	100.0	100.0	100.0	93.8	82.1
Maximum v/c	0.30	0.48	0.70	0.90	1.00
Maximum service flow rate (pc/h/ln)	700	1100	1600	2065	2300
			FFS= 100 km/h		
Maximum density (pc/km/ln)	7	11	16	22	28
Minimum speed (km/h)	90.0	90.0	90.0	89.1	80.4
Maximum v/c	0.28	0.44	0.64	0.87	1.00
Maximum service flow rate (pc/h/ln)	630	990	1440	1955	2250

Congestion and, hence, LOS F occurs when queques begin to form on freeways due to reasons, such as failure, breakdown, and etc. In case density increases instantaneously in the presence of quequing it may be considerably higher than the maximum value of 28 pc/km/ln for LOS E. Figure 3.2 shows the relationship between speed, flow, and density for basic freeway segments stressing various LOS on the basis of density boundaries.

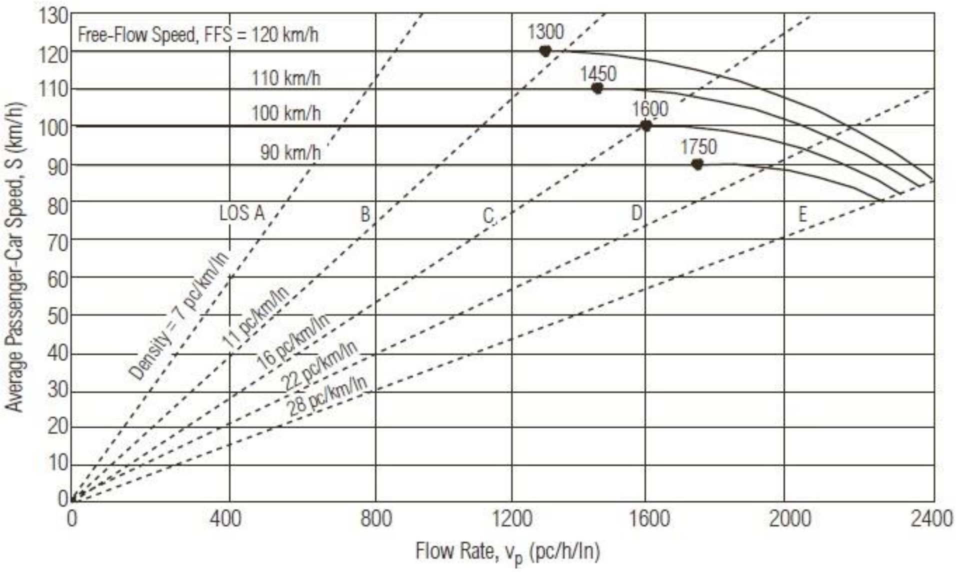

Figure 3.2 : Relationship between speed, density and flow rate (HCM, 2000).

Figure 3.3 visually demonstrates six LOS defined for basic freeway segments.

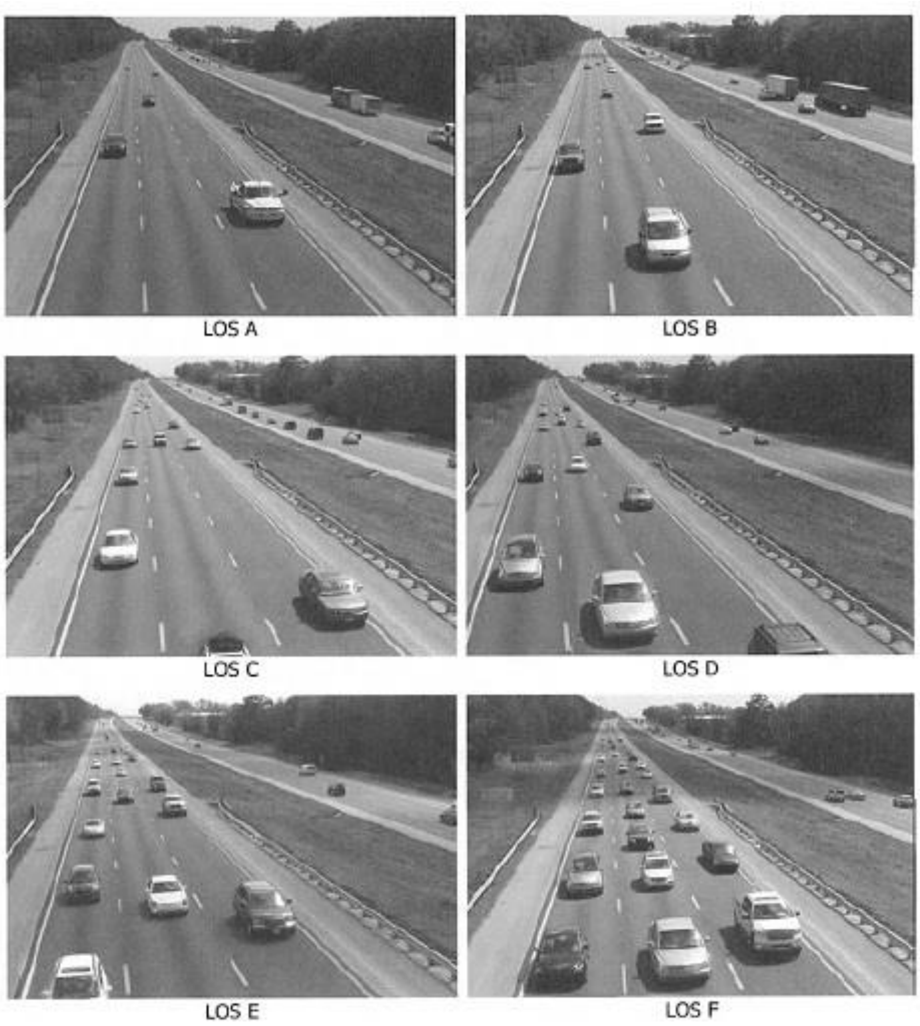

Figure 3.3 : LOS examples (HCM, 2010).

To conclude what varying LOS signify is summarized in the following, considering HCMs.

LOS A represents free-flow operations.

LOS B represents reasonably free-flow operations, and FFS on the freeway is maintained.

LOS C represents flow states speeds near the FFS of the freeway.

LOS D is the level at which speeds begin to lower with increasing flows, and with density increasing more quickly.

LOS E represents the operation at capacity. Operations on a freeway at this level are highly volatile due to the fact that there are virtually no usable gaps within the traffic stream, leaving little room to maneuver within the traffic stream.

LOS F represents breakdown, or unstable flow. Breakdowns occur when there are incidents, merge, weaving segments and lane drops. If the ratio of existing demand to actual capacity or forecast demand to estimated capacity exceeds 1.00, there becomes breakdowns, i.e., when the number of vehicles arriving at a point is greater than the number of vehicles moving through.

4. FLOW STATE CLASSIFICATION

Specification and classification of traffic flow state is a fundamental task for freeway traffic surveillance and control and has attracted considerable attention in the past three decades (Celikoglu, 2013). Although a great number of studies dealing solely with traffic flow performance modeling do exist, many of them ignore a solid motivation for the solution of real-life problems (Celikoglu, 2013).

The proposed process for flow state classification is composed of three sub-processes succeeding the pre-process of raw traffic data. The density measures are computed throughout a macroscopic traffic flow model simulation. The third sub-process is the classification of simulated density as a state indicator with the corresponding flow-rate and average speed measures considering a dynamic segmentation on flow states over the fundamental diagram of traffic flow. The overall flow state classification is shown in Figure 4.1.

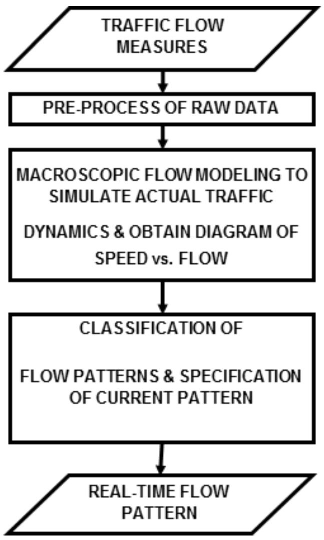

Figure 4.1 : Overall flow state classification process (Celikoglu, 2013).

The macroscopic flow modeling sub-processes run with the input filtered traffic flow measures that are collected by sensors mounted on a freeway stretch. The macroscopic flow model reconstructs section dynamics to obtain the fundamental diagram so that it can be partitioned in whether a dynamic or static manner. The traffic flow model is used to simulate and describe the dynamic behavior of flow propagation on the freeway. The resultant measures of density variable with the corresponding flow-rates and average speeds are used in the dynamic clustering analyses by multivariate and fuzzy methods, as well as the static clustering employing HCM 2010 boundaries (HCM, 2010). Finally, the current flow state is specified by the match-up of current value of the state indicator with the appropriate speed-flow diagram segment. It is worthy to note that the number of flow states to be classified is decided considering the level of service concept that is defined in the Highway Capacity Manual 2010 (HCM, 2010).

4.1 Brief Literature Review

It is essential to note that there is a need to differentiate the so-called short-term traffic flows forecasting studies (Vlahogianni et al., 2014) from the ones that traffic state is sought explicitly considering a traffic flow model besides an estimator (Wang et al., 2007; Celikoglu, 2013). Table 4.1 and Table 4.2 shows respectively the short-term studies chronological between 2004 and 2014.

Table 4.1 : Short-term traffic flow forecasting studies from 2004 to 2014 (Vlagohianni et al., 2014).

Author and Date	Area	Traffic Parameter	Prediction Step (min)	Prediction Horizon (steps)	Data Collection	Methodology Approach	Methodology Problem	Methodology Model
Yang et al.(2004)	Motorway	Speed	5	10	Detectors	Univariate	Time Series	Statistical
Rice and Van Zwet (2004)	Motorway	Speed	5	1	Detectors	Univariate	Pattern Recognation	Statistical
Ishak and Alecsandru(2004)	Motorway	Speed	5	4	Detectors	Univariate	Time Series	Hybrid*
Alecsandru and Ishak (2004)	Motorway	Speed	5	4	Detectors	Univariate	Function Approximation	Hybrid*
Shang et al. (2005)	Motorway	Speed	2	1	Detectors	Univariate	Time Series	Statistical
Kamarianakis et al. (2005)	Arterial	State	7.5	1	Detectors	Univariate	Time Series	Statistical
Quek et al. (2006)	Motorway	Density	1	60	Detectors	Univariate	Function Approximation	Neural Network
Cetin and Comert (2006)	Motorway	Speed	1	1	Detectors	Univariate	Time Series	Statistical
Vlahogianni (2007)	Arterial	State	1.5	1	Detectors	Univariate	Function Approximation	Neural Network
Vlahogianni (2008)	Arterial	State	1.5	1	Detectors	Univariate	Function Approximation	Neural Network
Hamad et al. (2009)	Motorway	Speed	5	5	Detectors	Univariate	Function Approximation	Neural Network
Chandra and Al- Deek (2009)	Motorway	Speed	5	1	Detectors	Multivariate	Time Series	Statistical
Kamarianakis et al. (2010)	Arterial	Volume/Speed/Occupancy	1.5	1	Detectors	Univariate	Time Series	Statistical
Guo and Williams (2010)	Motorway	Speed	5	1	Detectors	Univariate	Time Series	Hybrid*
Wang et al. (2011)	Motorway	Speed	5	1	Detectors	Univariate	Pattern Recognation	Hybrid*
Min and Wynter (2011)	Motorway	Volume/Speed	5	12	Detectors	Multivariate	Time Series	Statistical
Ishak et al. (2010)	Motorway	Speed	0.5	1	Detectors	Univariate	Function Approximation	Statistical
Djuric et al. (2011)	Motorway	Speed	5	6	Detectors	Univariate	Function Approximation	Statistical
Ye et al. (2012)	Motorway	Speed	0.01	60	GPS	Univariate	Function Approximation	Neural Network
Xia et al. (2012)	Motorway	State	5	1	Detectors	Multivariate	Clustering	Statistical
Tsirigotis et al. (2012)	Motorway	Speed	10	1	Detectors	Multivariate	Time Series	Statistical

Table 4.1 (continued) : Short-term traffic flow forecasting studies from 2004 to 2014 (Vlagohianni et al., 2014).

Author and Date	Area	Traffic Parameter	Prediction Step (min)	Prediction Horizon (steps)	Data Collection	Methodology Approach	Methodology Problem	Methodology Model
Kamarianakis et al. (2012)	Motorway	Speed	5	5	Detectors	Univariate	Time Series	Statistical
Dunne and Ghosh (2012)	Motorway	Volume /Speed	1	1	Detectors	Multivariate	Function Approximation	Neural Network
Wang and Shi (2012)	Motorway	Speed	5	1	Detectors	Univariate	Function Approximation	Neural Network
Vlahogianni and Karlaftis (2013)	Motorway	Speed	1	1	Detectors	Multivariate	Time Series	Hybrid*
Chan et al. (2013)	Motorway	Speed	1	5	Detectors	Univariate	Function Approximation	Hybrid*
Celikoglu (2013)	Motorway	Density	2	1	Detectors	Univariate	Function Approximation	Hybrid*

Table 4.2 : Short- term traffic flow forecasting studies only with multivariate methods from 2004 to 2014 (Vlagohianni et al., 2014).

Author and Date	Area	Traffic Parameter	Prediction Step (min)	Prediction Horizon (steps)	Data Collection	Methodology Approach	Methodology Problem	Methodology Model
Vlahogianni et al. (2005)	Arterial	Speed	3	5	Detectors	Multivariate	Function Approximation	Neural Network
Vlahogianni et al. (2007)	Arterial	Volume	1.5	1	Detectors	Multivariate	Time Series	Hybrid*
Vlahogianni (2008)	Arterial	State	1.5	1	Detectors	Multivariate	Function Approximation	Neural Network
Chandra and Al- Deek (2008)	Motorway	Speed	5	1	Detectors	Multivariate	Time Series	Statistical
Sheu et al. (2009)	Motorway	Volume	1	1	Detectors	Multivariate	Time Series	Neural Network
Innamaa(2009)	Motorway	Travel Time	5	4	Detectors	Multivariate	Pattern Recognation	Neural Network
Ghosh et al. (2009)	Arterial	Volume	15	50	Detectors	Multivariate	Time Series	Statistical
Chandra and Al- Deek (2009)	Motorway	Speed	5	1	Detectors	Multivariate	Time Series	Statistical
Tsekeris and Stathopoulos (2010)	Arterial	Volume	3	1	Detectors	Multivariate	Time Series	Statistical
Boto-Giralda et al. (2010)	Motorway	Volume	5	2	Detectors	Multivariate	Pattern Recognation	Hybrid*
Xia et al. (2011)	Motorway	Travel Time	15	1	Detectors	Multivariate	Time Series	Statistical

Table 4.2 (continued) : Short- term traffic flow forecasting studies only with multivariate methods from 2004 to 2014 (Vlagohianni et al., 2014).

Author and Date	Area	Traffic	Prediction		Data	Methodology		
		Parameter	Step (min)	Horizon (steps)	Collection	Approach	Problem	Model
Min and Wynter (2011)	Motorway	Volume/Speed	5	12	Detectors	Multivariate	Time Series	Statistical
Abu-Lebdeh and Singh (2011)	Arterial	Travel Time	5	1	Simulation	Multivariate	Function Approximation	Hybrid*
Xia et al. (2012)	Motorway	State	5	1	Detectors	Multivariate	Clustering	Statistical
Tsirigotis et al. (2012)	Motorway	Speed	10	1	Detectors	Multivariate	Time Series	Statistical
Sun et al. (2012)	Network	Volume	15	1	Detectors	Multivariate	Pattern Recognation	Neural Network
Dunne and Ghosh (2012)	Motorway	Volume /Speed	1	1	Detectors	Multivariate	Function Approximation	Neural Network
Vlahogianni and Karlaftis (2013)	Motorway	Speed	1	1	Detectors	Multivariate	Time Series	Hybrid*

Studies on flow state estimation, or equivalently pattern specification, show that both a flow modeling and an estimation component are to be processed in an integrated manner. In these studies, the flow state equation has been sought under different approaches. Some successful trials incorporating various approaches include the ones include the ones: by Treiber and Helbing (2002) in which the linear state equation through flow measure-based weighting scheme with interpolation is constructed; by Sun et al., (2003) in which linear measurement equations are utilized to employ flow measures under Eulerian framework; by Wang and Papageorgiou (2005) and Wang et al. (2007) in which filtering methods integrated to high-order traffic flow models are employed; by Munoz et al. (2003) and Sumalee et al. (2011) in which piecewise linear equations of biregime cell-based traffic flow model are used; and by Herrera and Bayen (2010) in which linear measure equations to process moving observer data under Lagrangian framework are established.

It is noted by Celikoglu (2013) that the above mentioned studies that deal with state specification necessitate a flow modeling component since the state estimator demands some critical characteristics of traffic, such as free-flow speed, jamming density, and capacity. Considering that such parameters makes it possible to determine flow states and specify shifts over the fundamental diagrams of flow vs. density and, hence, speed vs. flow, a dynamic approach to classify traffic flow patterns simulated by the original CTM has been introduced by Celikoglu (2013) utilizing the NN theory for density estimations in the overall process. Though important studies on explicit state specification are summarized above, a number of studies that somehow deal with flow state modeling are summarized in the following.

Hall et al. (1992) reviewed the data from earlier researchers and studied their compatibility with the revised picture of the speed-flow relationship. They believed that the congested branch of the speed-flow curve included two states. They classified the whole traffic condition into three major flow phases: uncongested, queue discharge, and congested.

Kerner and Rehborn (1996), studying German experimental data, investigated the complexity in traffic flow that was related to the space-time transition between free traffic flow and two other qualitatively different types of congested flow that they called "traffic jam" and "synchronized flow." Synchronized flow, the intermediate phase between uncongested and jams phases, was defined as the situation in which the highway becomes crowded, and all vehicles suddenly lose much of their freedom for maneuvering.In this flow condition, vehicles cannot change lanes and are forced to reduce their speeds.

Yang and Qiao (1998) used a neural-network pattern recognition method for categorizing traffic flow conditions in order to develop a method applicable to Chinese highway traffic.

Banks (1999) studied a data set collected in two locations in San Diego, California and concluded that, similar to Kerner and Rehborn's findings, there are qualitative differences in the appearance of flow-occupancy time sequences. He demonstrated that there might be two types of congested flow.

Helbing et al. (1999) presented a phase diagram of different kinds of congested conditions. They proposed four phases related to congestion in addition to free-flow and jam phases.

Park (2002) developed a hybrid neurofuzzy application to forecast short-term freeway traffic volume. The application consisted of two components. The first component, a fuzzy C-means method, was for clustering the traffic flow condition. The second component, a radial-basis function neural network, was for developing the forecasting model associated with each of those clusters.

Sun and Zhou (2005) proposed a K-means algorithm for categorizing a traffic data set (speed-density relationship) to two and three clusters in order to provide a natural tool for determination of breakpoints in multiregime traffic models.

Oh et al. (2005) demonstrated a method for development of LOS criteria based on section measures for real-time freeway analysis. Their study, conducted on a freeway section in Irvine, California, used reidentified median section speed as a measure of effectiveness. This measure was derived from analysis of inductive vehicle signatures and reidentification of vehicles traversing a major section of freeway in Irvine. The clustering algorithms, K-means and fuzzy approaches, were used to derive LOS criteria in terms of the reidentified median section speed.

Xia and Chen (2007a) studied a set of freeway data and applied a data clustering methodology. It was shown that between the two clustering variables, density and speed. They found that density had a more significant impact on the clustering results. The fundamental relationships between traffic parameters were then analyzed on the basis of the flow phase definition. They used K-means clustering method with Square Euclidean distance to define flow phases and suggested a five cluster model that utilized density and speed as input variables to categorize data into different clusters.

Xia and Chen (2007b) obtained data through California PeMS (Freeway Performance Measurement System) and developed a nested clustering technique to analyze the freeway operating condition. Their method defined the optimum number of clusters to show different flow phases. The flow characteristics they used were flow, speed, and occupancy. Their model presented an optimum fit of the statistical characteristics of the data based on the Bayesian information criterion (BIC) and the ratio of changes in dispersion measurement.

Bin et al. (2008) analyzed the traffic data of vessels in a certain water route by using K- means algorithm. In this study, for distance between data point and the cluster centroid, Euclidean distance was used.

Azimi and Zhang (2010) applied three different clustering methods to classify freeway traffic flow conditions based on the characteristics of the flow.

Antoniou et al. (2013) provided validation of a dynamic data driven framework that allowed for traffic state to apply two freeway data sets from Irvine, CA and Ayalon, Israel for estimation and prediction. In this study, BIC was chosen as optimization criteria and RMSN, RMSPE and MPE used as validation criteria.

The literature in which clustering methods are applied to transportation problems in Turkey is quite limited. Murat and Sekerler (2009) investigated the performance of K-means and fuzzy clustering methods on traffic accident data of Denizli city for the years of 2004, 2005 and 2006.

4.2 Modeling and Simulation

In the following, information on data set that is used to macroscopically model and simulate the traffic on the selected freeway segment is explained preceding the flow modeling sub-process. Following, descriptive statistics and empirical features of simulation results are presented.

4.2.1 Information on study area and data set

In this study the remote traffic microwave sensor (RTMS) data obtained within the scientific research project, with the number 108M299, funded by the Scientific and Technological Research Council of Turkey (Çelikoğlu et al., 2010) is processed to derive the macroscopic features of traffic flow. Field data had been collected in January 2009 from three successive microwave sensor units located with the spacing of 250 m and 750.50 m in side-fired position on a 4-lane freeway segment in Istanbul.

The test segment is on an approach from the European side of Istanbul to the Fatih Sultan Mehmet Bridge. Lane-specific measurements on traffic variables, i.e., long vehicle counts, $n_1(t)$, non-long vehicle counts, $n_2(t)$, and speed, $U(t)$, are obtained as aggregated in 2 minutes time interval. For each time interval, flow-rate is determined by the vehicle counts of 20 minutes, and the speed by the harmonic mean of the individual vehicles passing in this time period. Such a procedure is followed due to the fact that the arithmetic mean of spot speeds leads to systematic overestimations in congested conditions. In spite of the advances in data cllection technology, the problem of measurement noise on empirical data still exists, as in this case. It is shown that if the spacing of two adjacent detectors is no longer than about 1 km, it is appropriate to filter linearly the raw data (Treiber and Helbing, 2002). Further, it is noted in Treiber et al. (2010) that the wide scattering is partly a side effect of the measurement process. Therefore, in order to reconstruct observed dynamics at any point in a space-time plane from sensors positioned at discrete locations, the thin plate spline (TPS) interpolation method (Wahba, 1990) is applied to post-process raw data prior to conduct flow modeling and simulation following Celikoglu (2013).

The TPS is a commonly used basis function for representing coordinate mappings from $\mathfrak{R}^2$ to $\mathfrak{R}^2$ and the 2-D generalization of the cubic spline (Wahba, 1990). In the physical setting, the deflection is in the z-direction, orthogonal to the *xy* plane. While applying the TPS method in the problem of spatial and temporal coordinate transformation over raw data, it is assumed that any raw flow measure is represented by the third dimension, analogous to deflections of plate in the *z*-direction, as in Celikoglu (2013). The TPS filtering helped to tolerate the unrealistic empirical measures and measurement noise which led to fluctuations by affecting the underlying states.

Considering the parameters of the distribution of filtered measurements it is seen that variation on actual speeds increase from lane no. 1 to other lanes as lane no. 1 is more frequently used by heavy vehicles those tend to change lane relatively lesser than other vehicle classes. Note that this behaviour cannot be supported in terms of vehicle counts since the maximum of the vehicle counts is observed when the traffic is flowing at an optimum speed. Therefore, it is the main reason that density is derived as a function of both the flow and speed, as traffic flow theory dictates.

4.2.2 Flow modeling sub- process

In order to model the traffic flow on the slected freeway segment, the macroscopic flow modeling approach is followed. Macroscopic flow models, that are analogous enough to make the hydrodynamic theory useful in describing traffic dynamics, trace the collective vehicle dynamics in terms of aggregate variables such as density, flow and speed and are useful in reproducing freeway flow dynamics (Celikoglu, 2007). Considering the findings in the extensive research on the need to adopt fundamental diagram in flow modeling in case of wide scattering of empirical freeway data (Kerner and Rehborn, 1996; Kerner,2004; Treiber et al. 2010), a macroscopic flow model with a fundamental diagram has been chosen to be incorporated to flow modeling sub-process.

The model incorporated in flow modeling sub-process follows the fluid dynamics approach to theory of continuous vehicular traffic flow, defined upon the variables of flow-rate, q, density, k, and speed, u, and referred to as the LWR theory (Lighthill and Whitham, 1955; Richards, 1956), as denoted in Section 2.6. This theory assumes that flow is strictly a function of density, $q = Q(k)$, and consequently speed is strictly a function of density, $u = U(k)$. The LWR theory describes the model as a single partial differential equation in conservation form as given by Equation **(4.1)** or alternatively as given by Equation **(4.2)**,

$$\frac{\partial k}{\partial \mathrm{t}} + \frac{\partial\left(k \cdot \mathrm{U}k\right)}{\partial \mathrm{x}} = 0 \qquad \mathbf{(4.1)}$$

$$\frac{\partial k}{\partial \mathrm{t}} + \left(\mathrm{C}(k) \cdot \frac{\partial k}{\partial \mathrm{x}}\right) = 0 \qquad \mathbf{(4.2)}$$

where; $C(k) = (\partial Q(k))/\partial k$. The LWR theory expresses that slight fluctuations in flow are propagated upstream along kinematic waves, where the speed can be determined as the slope of flow-density curve, $C(k)$. Given the appropriate boundary conditions, solution to this model can be obtained by determining the function $k\ (x, t)$, where x and t represent respectively space and time. Different variations of the macroscopic model given by Equation **(4.2)** can be characterized by the speed-density relationship $u = U(k)$ and, consequently, by the adopted fundamental diagram (Celikoglu, 2013).

In order to obtain a convergent approximation to the continuous model constructed using the LWR theory, the discrete cell transmission approach of Daganzo (1994) that adopts a two-phase triangular fundamental diagram is utilized. The cell transmission model (CTM) is discrete both in space and time, and divides the freeway into ‘cells’. The traffic flow entering a cell bounded by points s and $s+1$, is considered to be constant between two successive times t and $t+\Delta t$ and can be determined by Equation **(4.3)** (Daganzo, 1994),

$$q^{s,s+1}(t) = min \begin{cases} u_f{}^{s-1,s}.\,k^{s-1,s}(t) \\ u_{cong}^{s,s+1}.\left(k_j^{s,s+1} - k^{s,s+1}(t)\right) \\ q_m^{s,s+1} \end{cases} \tag{4.3}$$

where; u_f is the free-flow speed, u_{cong} is the backward wave propagation speed in congestion, k_j is the jamming density, k_{opt} is the optimum density and q_m is the capacity, $k^{s,s+1}(t)$ is the average density of cell s, $s+1$ between times t and $t+\Delta t$, $k_j^{s,s+1}$ is the jamming density of cell s, $s+1$, $u_f^{s-1,s}$ is the free flow speed in cell s-1, s, $u_{cong}^{s,s+1}$ is the congestion wave speed in cell s, $s+1$, and $q_m^{s,s+1}$ is the capacity of cell s, $s+1$. A schematic representation of freeway test stretch divided into two cells that correspond to sections is provided in Figure 4.2.

Figure 4.2 shows the sensors and distances between sensors on the sketch.

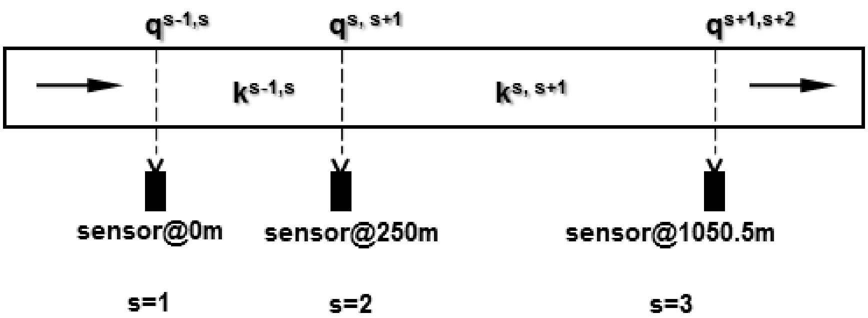

Figure 4.2 : Schematic represantation of freeway test stretch divided into sections (Celikoglu, 2013).

Note that the formulation of flow propagation via cell interfaces is useful for real-time applications, as flow variables are measured by point sensors and sensors may be mounted to match exact points.

Considering the prevailing phase speeds and densities, the flow on each section is determined by Equation **(4.3)** for each time interval Δt.

The flow modeling component follows the procedure explained above for the modeling and real-time simulation of actual traffic dynamics and the consequent reconstructions of section performances in terms of traffic variables. The real-time reconstruction of flow variables are sequentially used to update and partition the fundamental diagrams of flow-density and speed-flow at each or specified computation time intervals.

4.2.3 Empirical features of simulated data

Though the flow modeling process outputs, the variables of traffic flow, are used dynamically plot the fundamental diagram in selected time steps, a number of descriptive statistics on the entire output data set is provided in the following. The lane-based variation of mean, standard deviation, skewness, and kurtosis values of simulation model reconstructions on flow, speed and density are depicted in the following figures. Figure 4.3 shows the variation of means on flow data.

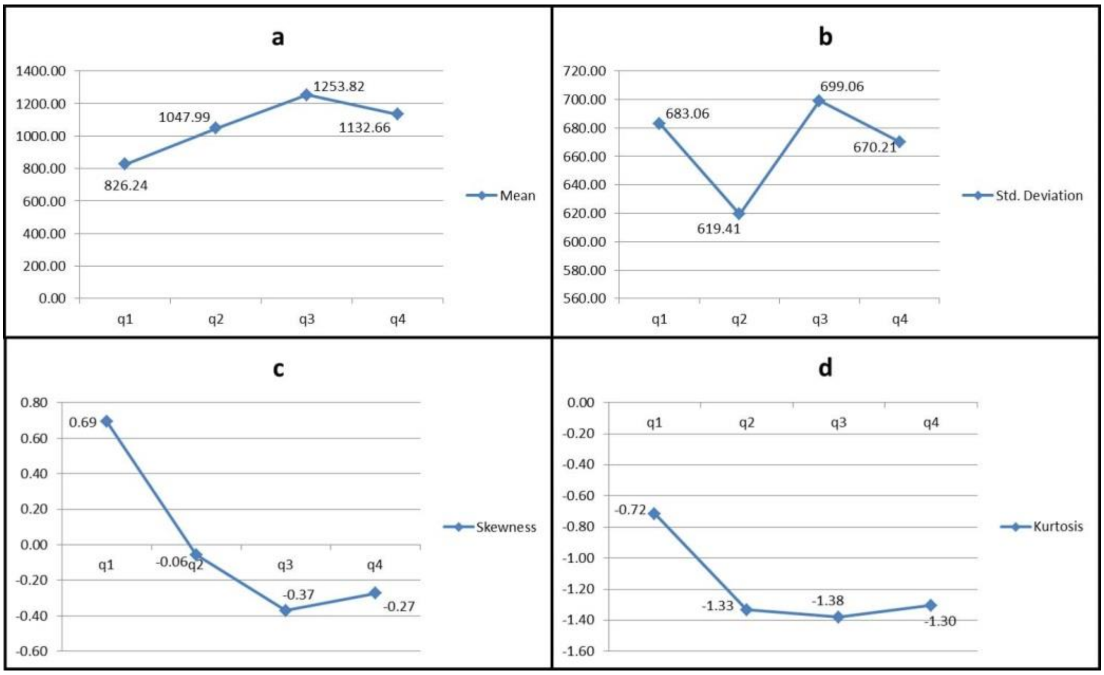

Figure 4.3 : Lane- Based Variations of Statistics on Flow Variable a) Mean, b) Standard Deviation, c) Skewness, and d) Kurtosis.

Figure 4.4 shows the variation of standard deviations on speed data.

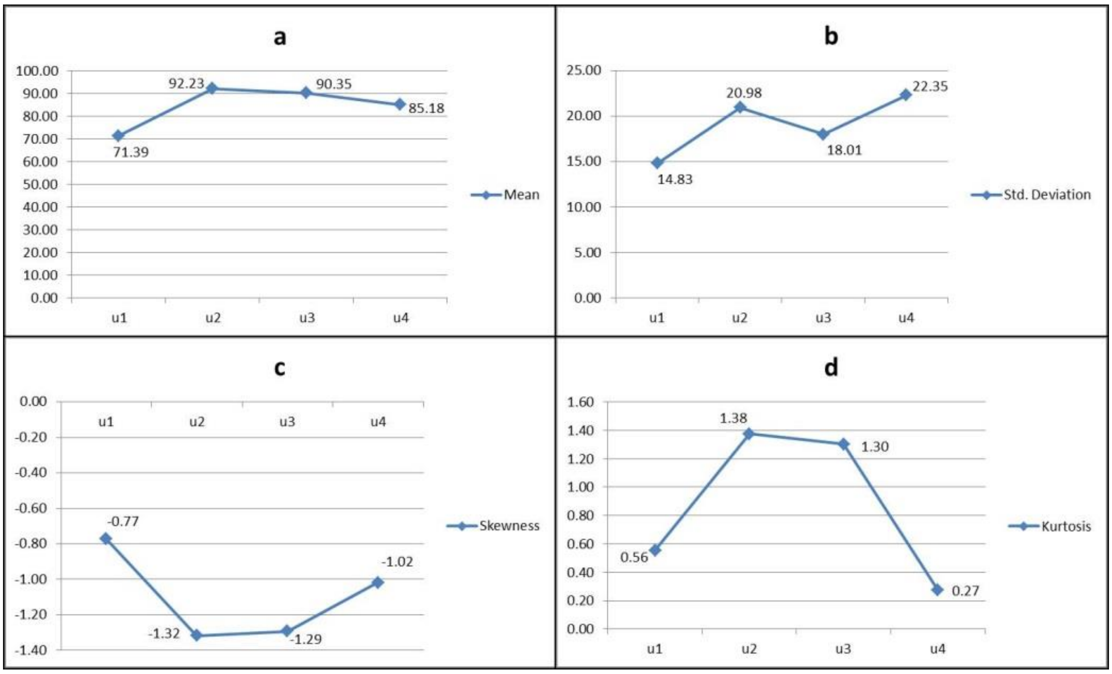

Figure 4.4 : Lane- Based Variations of Statistics on Speed Variable a) Mean, b) Standard Deviation, c) Skewness, and d) Kurtosis.

Figure 4.5 shows the variation of standard deviations on density data.

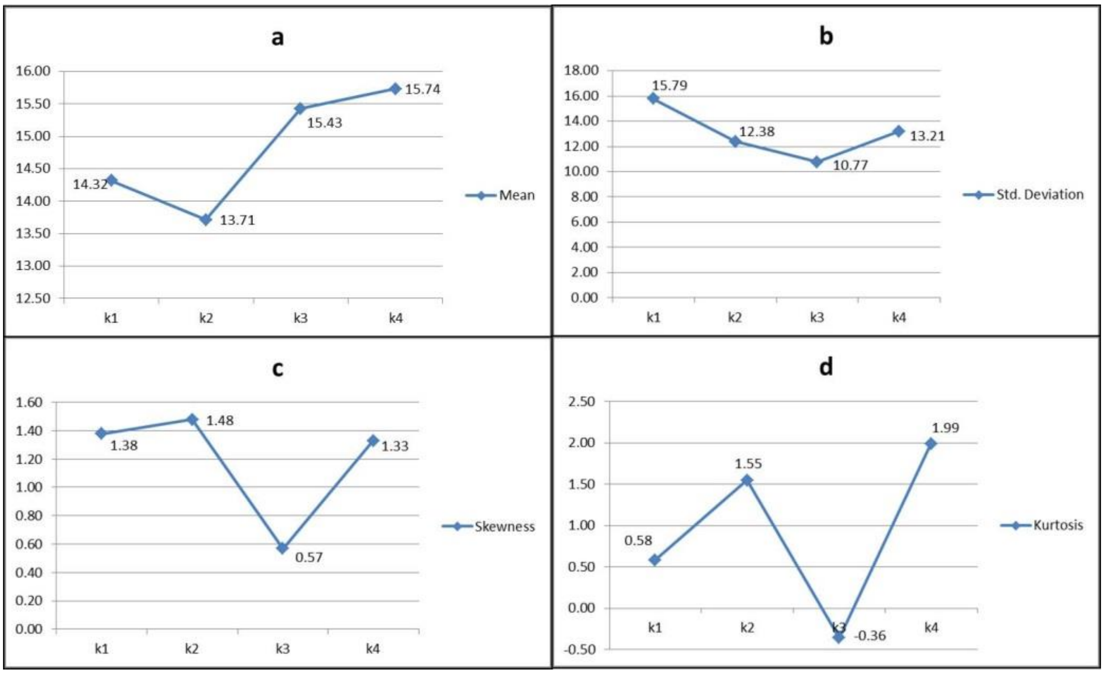

Figure 4.5 : Lane- Based Variations of Statistics on Density Variable a) Mean, b) Standard Deviation, c) Skewness, and d) Kurtosis.

In order to obtain the critical values on speed, flow, and density variables, the fundamental diagrams showing bivariate relationships are used.

4.3 Methods Applied

In the following, the methods that are incorporated partition the fundamental diagram, and hence, classify the traffic flow states are explained.

4.3.1 Deterministic clustering method

The LOS measures for multi-lane freeways in the Highway Capacity Manual (2010) are used to rule partitions. The LOS measures A through F for flow state (FS) representation in terms of vehicle per kilometer-lane is transferred from HCM (2010), as given by Equation **(4.4)**.

$$FS = \begin{cases} \text{A, if} & k \leq 7 \\ \text{B, if} & 7 < k \leq 11 \\ \text{C, if} & 11 < k \leq 17 \\ \text{D, if} & 17 < k \leq 22 \\ \text{E, if} & 22 < k \leq 28 \\ \text{F, otherwise} & \end{cases} \quad \textbf{(4.4)}$$

As it is aforementioned earlier, in HCM (2010) the boundary between LOS E and F represents capacity. The breakdown of flow, signified by the LOS F, is determined when demand flows are greater than the capacity. In cases where capacity is exceeded, the LOS method in HCM (2010) does not produce density estimates and, consequently, specific values within the LOS F. Therefore, in the dynamic classifications case a time-varying partitioning is performed, in contrast to static case in which the boundaries of partitions given by Equation **(4.4)** remain constant throughout the simulation horizon.

4.3.2 Multivariate clustering analysis

Jain et al. (1999) describes the cluster analysis as given below:

"Cluster analysis is the organization of a collection of patterns (usually represented as a vector of measurements, or a point in a multidimensional space) into clusters based on similarity. Intuitively, patterns within a valid cluster are more similar to each other that they are to a pattern belonging to a different cluster".

Even though, clustering analysis is counted on in the multivariate statistical methods, it is become dissimilar on the strength of its mathematical background. Moreover, it is useful to divide clustering analysis into two parts. Since clusters can formally be seen as subsets of the data set, one possible classification of clustering methods can be according to whether the subsets are fuzzy or crisp (hard).

Hard clustering methods are based on classical set theory, and require that an object either does or does not belong to a cluster. Hard clustering means partitioning the data into a specified number of mutually exclusive subsets.

Fuzzy clustering methods, however, allow the objects to belong to several clusters simultaneously, with different degrees of membership.

This section is aimed to show mathematical expressions of clustering analysis both k-means and fuzzy c-means. Before the giving expressions, the types of clusters are defined. For crisp clustering analysis, there are five types of cluster:

Well-Separated: A cluster is a set of objects in which each object is closer to every other object in the cluster than to any object not in the cluster. Sometimes a threshold is used to specify that all the objects in a cluster must be sufficiently close (or similar) to one another. It gives only goof result when the data contains natural clusters that are quite far from each other.

Prototype Based: A cluster is a set of objects in which each object is closer to the prototype that defines the cluster than to the prototype of any other cluster. For data with continuous attributes, the prototype of a cluster is often a centroid, i.e., the average (mean) of all the points in the cluster. For this study, it is made an assumption that the clusters are prototype- based.

Graph-Based: If the data is represented as a graph, where the nodes are objects and the links represent connections among objects, then a cluster can be defined as a connected component; i.e., a group of objects that are connected to one another, but that have no connection to objects outside the group.

Density-Based: A cluster is a dense region of objects that is surrounded by a region of low density. A density-based definition of a cluster is often employed when the clusters are irregular, and when noise and outliers are present

Shared-Property: More generally, it can be defined as a cluster as a set of objects that share some property. This definition encompasses all the previous definitions of a cluster.

4.3.2.1 K- means clustering

K-means clustering method has a long history, but is still subject of current researches. The method was originally proposed by MacQueen (1967). The K-means method and many of its variations are described in detail by Anderberg (1973) and Jain and Dubes (1988).

K- means clustering is counted on in prototype- based clustering techniques. In Prototype-based clustering techniques, one-level partitioning of data objects is created. There are a number of such techniques, such as K- means and K- medoids. In the following section, solely K-means will be focused on, which is one of the pioneering and most widely used clustering algorithms. K-means defines a prototype in terms of a centroid, which is usually the mean of a group of points, and is typically applied to objects in a continuous n-dimensional space.

4.3.2.2 K- means clustering algorithm

The basic version of K- means clustering can be arranged in four steps (Xu and Wunsch, 2008);

1) Initialize a K-partition randomly or based on some prior knowledge. Calculate the cluster prototype matrix;

$$M=[m_1,\ldots\ldots\ldots m_K) \qquad \textbf{(4.5)}$$

Where M is denoted as a prototype matrix and m is the sample mean for the each cluster.

2) Assign each object in the data set to the nearest cluster C_l, i.e,

$$x_j \in C_l,\ if\ \left\|x_j - m_l\right\| < \left\|x_j - m_i\right\| \text{ for } j=1,\ldots,N,\ i\neq l, \text{ and } i=1,\ldots\ldots\ldots,K \qquad \textbf{(4.6)}$$

Where x_j is a set of objects and C_l is denoted as clusters which are organized to objects.

3) Recalculate the cluster prototype matrix based on the current partition,

$$m_i = \frac{1}{N_1} \cdot \sum_{x_j \in C_i} \left(x_j\right) \qquad \textbf{(4.7)}$$

4) Repeat steps 2 and 3 until there is no change for each cluster.

The "distance" concept has to be defined while using K-means clustering. Distance represents a member of data sets range between cluster's centroids. In this study, two types of distance are used for finding the centroids' of clusters. First is the Square Euclidian distance (SED), which needs to be explained by Euclidian distance. It is simply the geometric distance in a multidimensional space and computed as given by Equation **(4.8)** (Mooi and Sarstedt, 2011)

$$\text{distance}(x, y) = \left[\sum_{i \in n} (x_i - y_i)^2 \right]^{1/2} \qquad \textbf{(4.8)}$$

Where x is denoted the coordinate of a point's value on the x axis in Cartesian system, y is denoted of a point's value on the y axis in Cartesian system, and n is the number of observations. In Figure 4.6 it is demostrated what is explained above.

The SED is computed as given by Equation **(4.9)** (Mooi and Sarstedt, 2011)

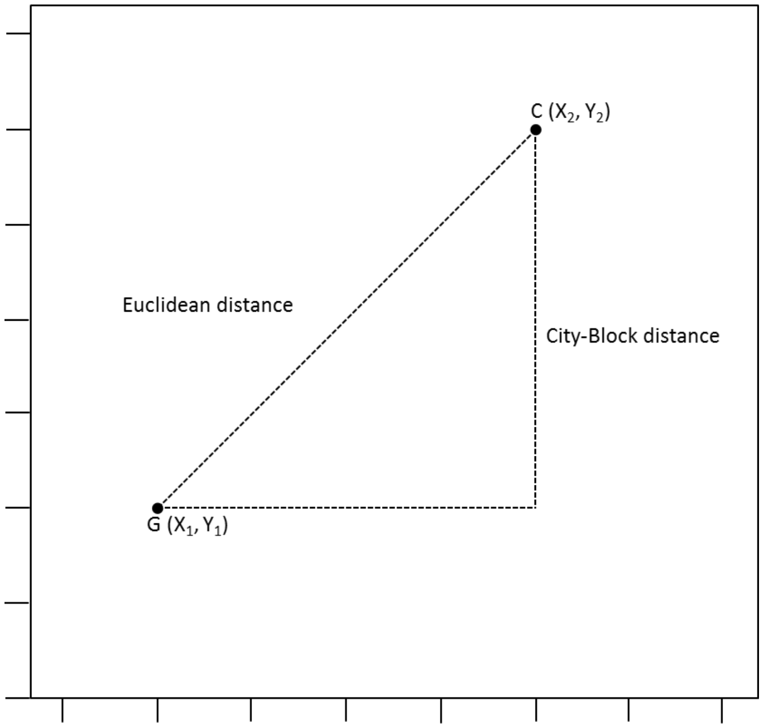

Figure 4.6 : Visualition of Euclidean and City-Block distance.

$$\text{distance}(x, y) = \sum_{i \in n} (x_i - y_i)^2 \qquad \textbf{(4.9)}$$

The distance between clusters is distance between centroids.

The City- Block Distance (CBD) uses the sum of the variables' absolute differences. This is generally called the Manhattan metric as it is alike to the walking distance between two points in a city like New York's Manhattan district, where the distance equals the number of block in the study directions North- South and East- West. In Figure 4.6 it is demostrated what is explained above CBD can be obtained as given by Equation **(4.10)** (Mooi and Sarstedt, 2011).

$$\text{distance}(x, y) = |x_1 - x_2| + |y_1 - y_2| \tag{4.10}$$

4.3.3 Fuzzy c- means clustering method

Most of the analytical fuzzy clustering algorithms are based on optimization of the basic c-means objective function, or some of its modification. Therefore, in the following fuzzy c- means functional is explained.

4.3.3.1 Fuzzy c- means functional

A large family of fuzzy clustering algorithms is based on minimization of the fuzzy c-means functional that is formulated as given by Equation **(4.11)** (Dunn, 1974; Bezdek, 1981):

$$J(\mathbf{Z};\mathbf{U},\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m \|\mathbf{z}_k - \mathbf{v}_i\|_{\mathbf{A}}^2 \tag{4.11}$$

Where

$$U = [\mu_{ik}] \in M_{fc} \tag{4.12}$$

is a fuzzy partition matrix of Z,

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c], \mathbf{v}_i \in \mathbb{R}^n \tag{4.13}$$

is a vector of cluster prototypes, which have to be determined,

$$D_{ik\mathbf{A}}^2 = \|\mathbf{z}_k - \mathbf{v}_i\|_{\mathbf{A}}^2 = (\mathbf{z}_k - \mathbf{v}_i)^T A\, (\mathbf{z}_k - \mathbf{v}_i) \tag{4.14}$$

is a squared inner-product distance norm, and

$$m \in [1, \infty) \tag{4.15}$$

is a parameter which determines the fuzziness of the resulting clusters.

4.3.3.2 Fuzzy c-means algorithm

The basic version of fuzzy c-means clustering can be arranged in four steps.

Given the data set **Z**, choose the number of clusters $1 < c < N$, the weighting exponent $m > 1$, the termination tolerance $\varepsilon > 0$ and the norm-inducing matrix **A**. Initialize the partition matrix randomly, such that $\mathbf{U}^{(0)} \in M_{fc}$.

1) Compute the cluster prototypes (means):

$$\mathbf{v}_i{}^{(l)} = \frac{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m \mathbf{z}_k}{\sum_{k=1}^{N} (\mu_{ik}{}^{(l-1)})^m}, 1 \leq i \leq c. \quad \textbf{(4.16)}$$

2) Compute the distances:

$$D_{ik\mathbf{A}}^2 = \left(\mathbf{z}_k - \mathbf{v}_i{}^{(l)}\right)^T \mathbf{A}\left(\mathbf{z}_k - \mathbf{v}_i{}^{(l)}\right), \quad 1 \leq i \leq c, 1 \leq k \leq N. \quad \textbf{(4.17)}$$

3) Update the partition matrix:

For $1 \leq k \leq N$;

if $D_{ik\mathbf{A}} > 0$ for all $i = 1,2,\ldots,c$

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c} (D_{ik\mathbf{A}}/D_{jk\mathbf{A}})^{2/(m-1)}} \quad \textbf{(4.18)}$$

Otherwise;

$$\mu_{ik}^{(l)} = 0 \text{ if } D_{ik\mathbf{A}} > 0, \text{and } \mu_{ik}^{(l)} \in [0,1] \text{ with } \sum_{i=1}^{c} \mu_{ik}^{(l)} = 1 \quad \textbf{(4.19)}$$

4) Repeat until

$$\left\|\mathbf{U}^{(l)} - \mathbf{U}^{(l-1)}\right\| < \epsilon \quad \textbf{(4.20)}$$

The parameters, i.e., the number of clusters, c, the 'fuzziness' exponent, m, the termination tolerance, ε, and the norm-inducing matrix, A. Moreover, the fuzzy partition matrix, **U**, has to be initialized prior to employing the FCM algorithm.

Number of Clusters: The number of clusters c is the most important parameter, in the sense that the remaining parameters have less influence on the resulting partition. When clustering real data without any a priori information about the structures in the data, one usually has to make assumptions about the number of underlying clusters.

Fuzziness Parameter: The weighting exponent m is a rather important parameter as well, because it significantly influences the fuzziness of the resulting partition.

Termination Criterion: The FCM algorithm stops iterating when the norm of the difference between **U** in two successive iterations is smaller than the termination parameter ε.

Norm-Inducing Matrix: The shape of the clusters is determined by the choice of the matrix **A** in the distance measure (Equation **(4.14)**). A common choice is **A = I**, that is the standard Euclidean norms given by Equation **(4.21)**;

$$D_{ik\mathbf{A}}^2 = (\mathbf{z}_k - \mathbf{v}_i)^T\,(\mathbf{z}_k - \mathbf{v}_i) \qquad \textbf{(4.21)}$$

4.4 Numerical Implementations

Before starting to refer applications on data clustering with different techniques and it is essential to clarify some points. As it is known that, dynamic classification approach is similar to the LOS concept defined for the static case. This approach slightly differs in terms of bounding density measures all of which change temporally throughout the simulation. Considering such a possible onsite implementation of the proposed approach supported by information dissemination hardware, the variation on road users' perception should stay in certain limits for appropriate behavior adaptations. This requires the dissemination of a reasonable number of information messages that are signified by captured shifts on states. Moreover, to obtain a consistent comparative evaluation with the static case, the user- defined class number is set to six. As it is understood, to choose class number as six is not an arbitrary decision.

In each time step, it represents adding the new values to starting data set each 4 hours, dynamic segmentation on the fundamental diagram is updated considering the critical value of maximum density and partitioning the current density range into classes upon user-defined rules. The flow state is specified by considering the match of the current density prediction to the appropriate class on the updated diagram. Therefore, the transitions and jumps on states between successive times are captured.

4.4.1 Clustering by Highway Capacity Manual approach

Despite the fact that it is employed for clustering purpose , the density boundaries in HCM is exactly used to define the level of service. For each time interval, the pre-specified limits that are defined in Equation **(4.4)** indicate level of service. Figure Figure 4.7, Figure 4.8, Figure 4.9, Figure 4.10 and Figure 4.11 show each 4 hours time interval that starts from 7440, and ends in 8520.

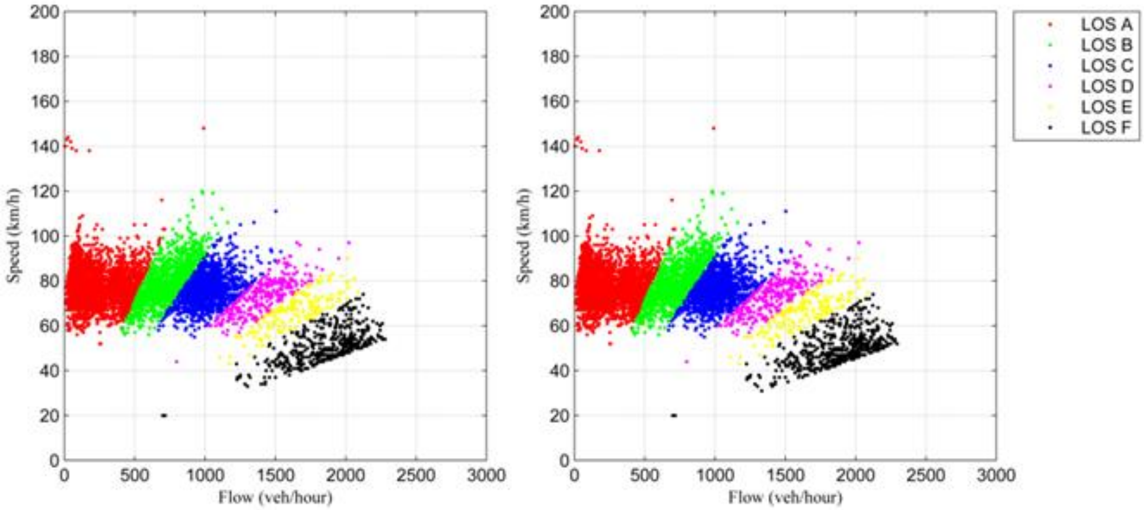

Figure 4.7 : HCM clustering result for 7440 and 7560.

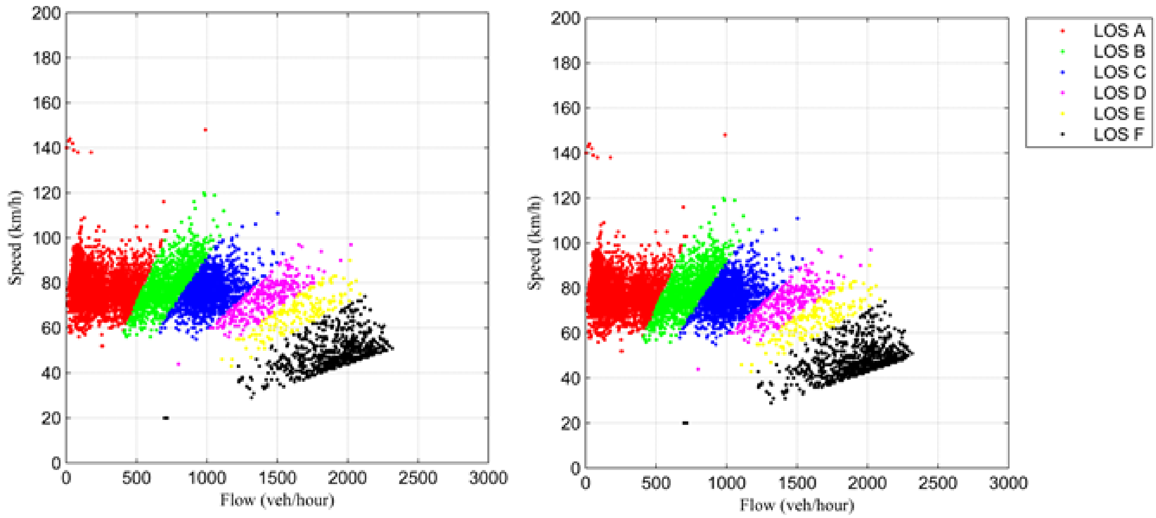

Figure 4.8 : HCM clustering result for 7680 and 7800.

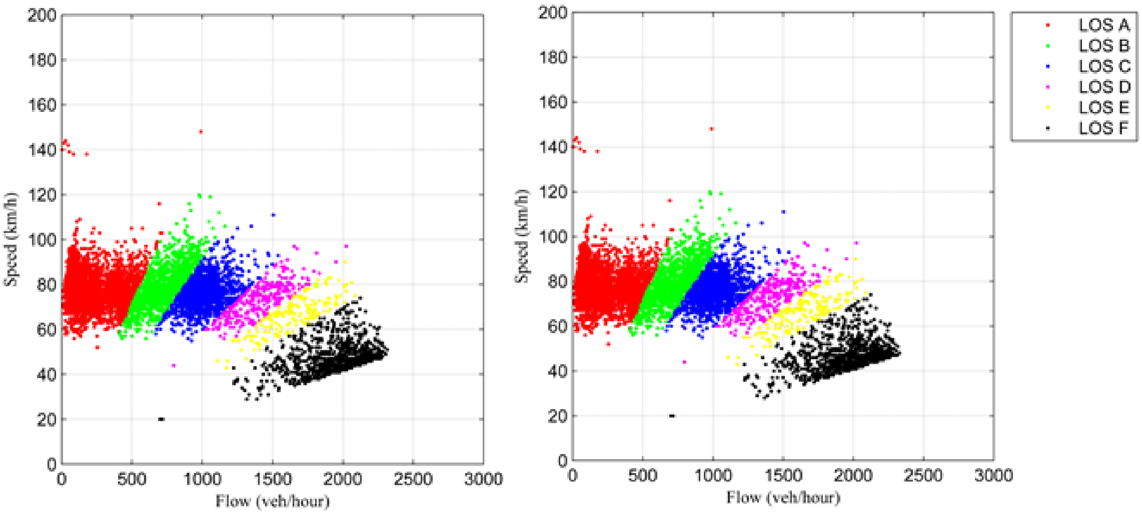

Figure 4.9 : HCM clustering result for 7920 and 8040.

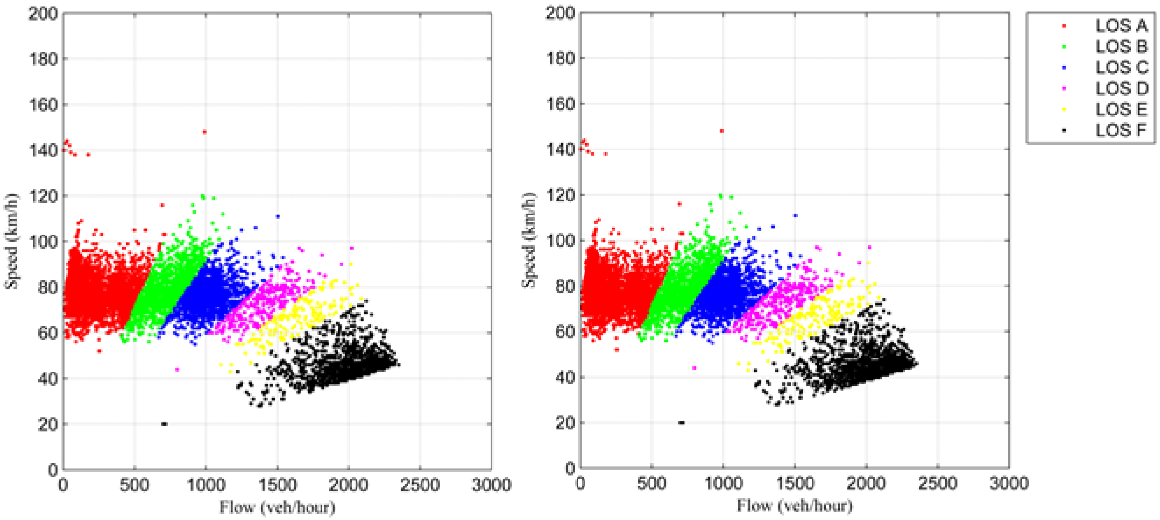

Figure 4.10 : HCM clustering result for 8160 and 8280.

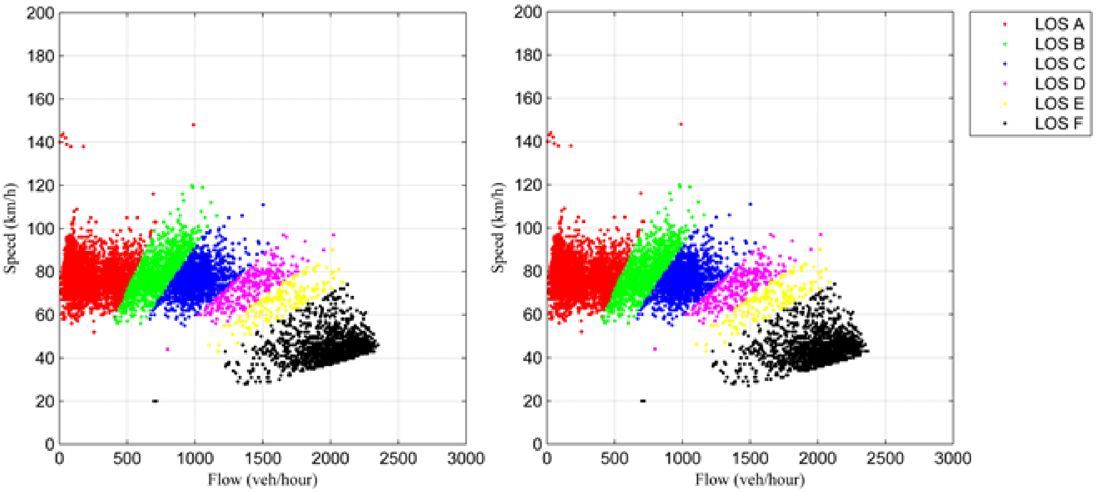

Figure 4.11 : HCM clustering result for 8400 and 8520.

4.4.2 Clustering by k-means

Figure 4.12 is provided to present the process followed in applying the K-means algorithm on data set. That sketch shows of K-means algorithm.

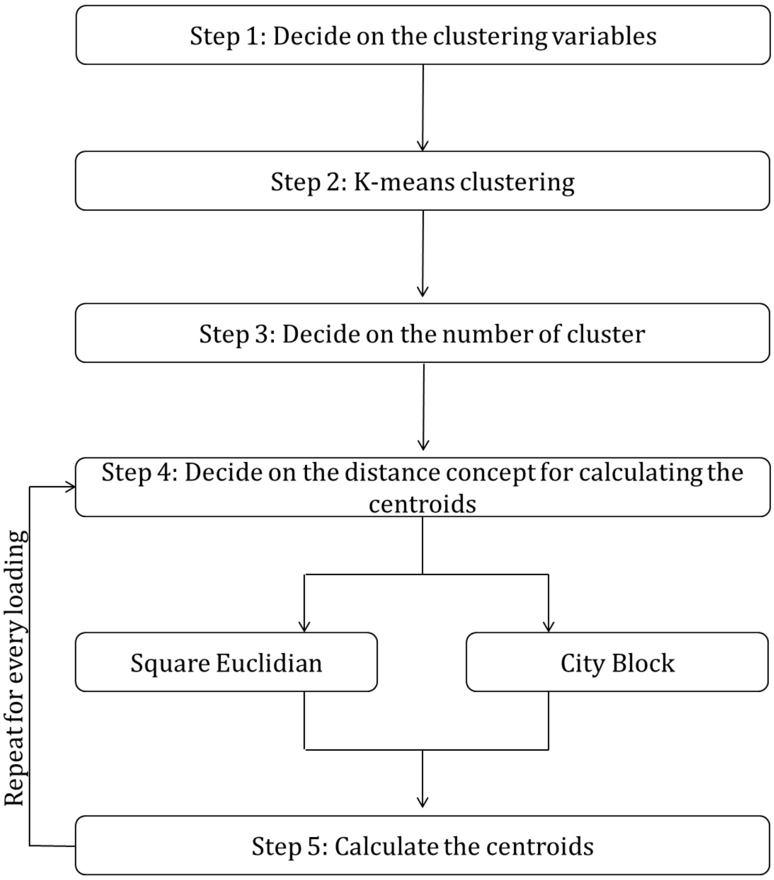

Figure 4.12 : Flow chart of K-means clustering.

For the first step in the Figure 4.7, it is necessary to define what the clustering variables are. In this study, flow and density are the clustering variables, that are further used to derive speed-flow diagrams. Third step of the sketch is to define number of clusters. As it aforementioned in Section 4.3, in order to make comparison considering the HCM boundaries, the number of clusters is set to six. Fourth and the prominent part of K- means clustering is to define distance concept and define the distances for calculating centroids.

4.4.2.1 Square Euclidean method

In order to obtain distances using Square Euclidean Method, Equation **(4.9)**is used. Figure 4.13, Figure 4.14, Figure 4.15, Figure 4.16 and Figure 4.17 show each 4 hours time interval that starts from 7440, and ends in 8520

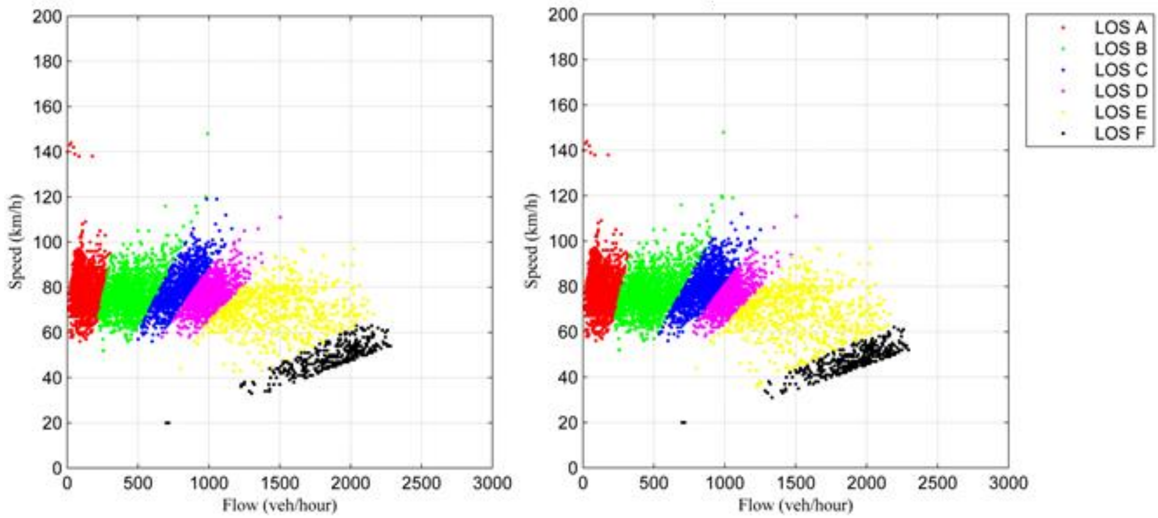

Figure 4.13 : Square Euclidean Method result for 7440 and 7560.

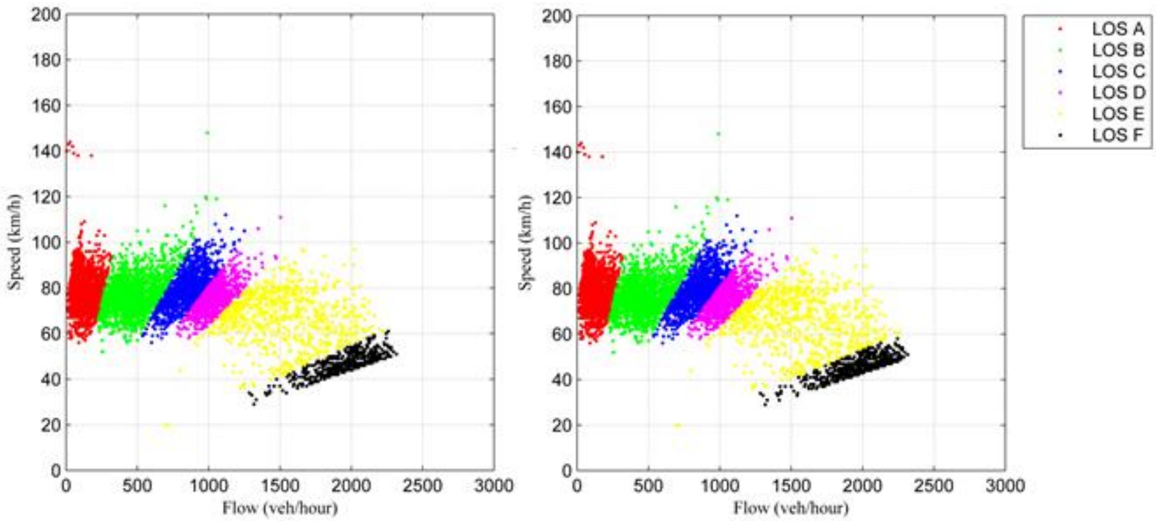

Figure 4.14 : Square Euclidean Method result for 7680 and 7800.

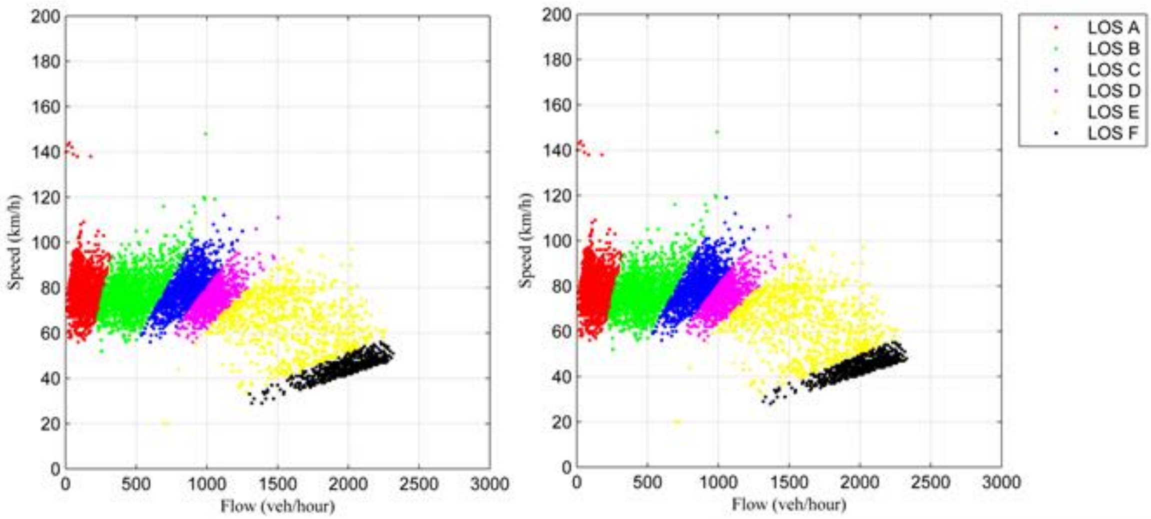

Figure 4.15 : Square Euclidean Method result for 7920 and 8040.

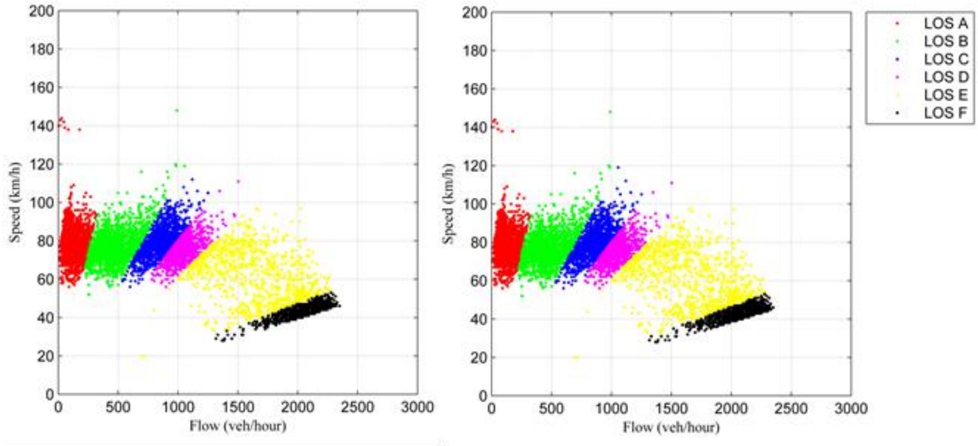

Figure 4.16 : Square Euclidean Method result for 8160 and 8280.

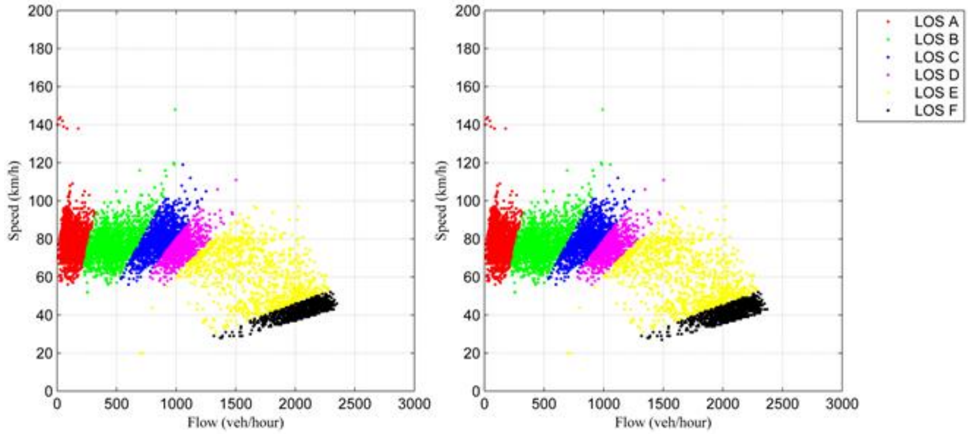

Figure 4.17 : Square Euclidean Method result for 8400 and 8520.

4.4.2.2 City Block distance method

In order to obtain distance measures in City Block Method, Equation **(4.10)** is used. Figure 4.18, Figure 4.19, Figure 4.20, Figure 4.21 and Figure 4.22 show each 4 hours time interval that starts from 7440, and ends in 8520.

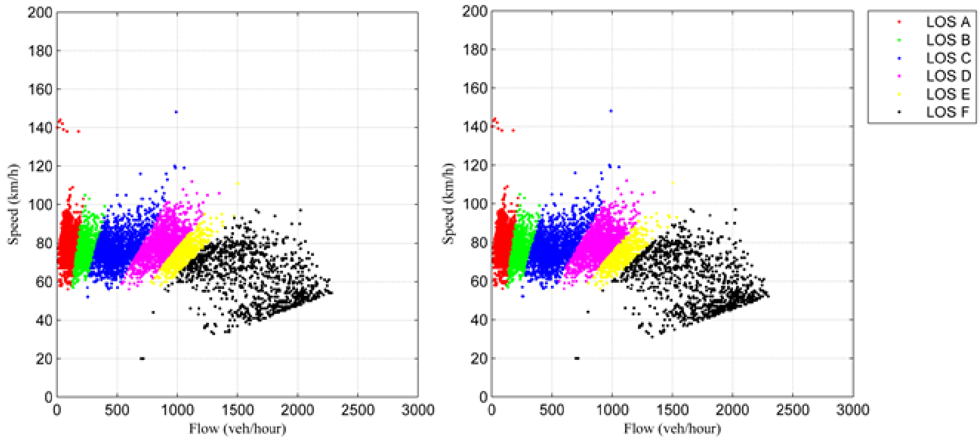

Figure 4.18 : City Block Method result for 7440 and 7560.

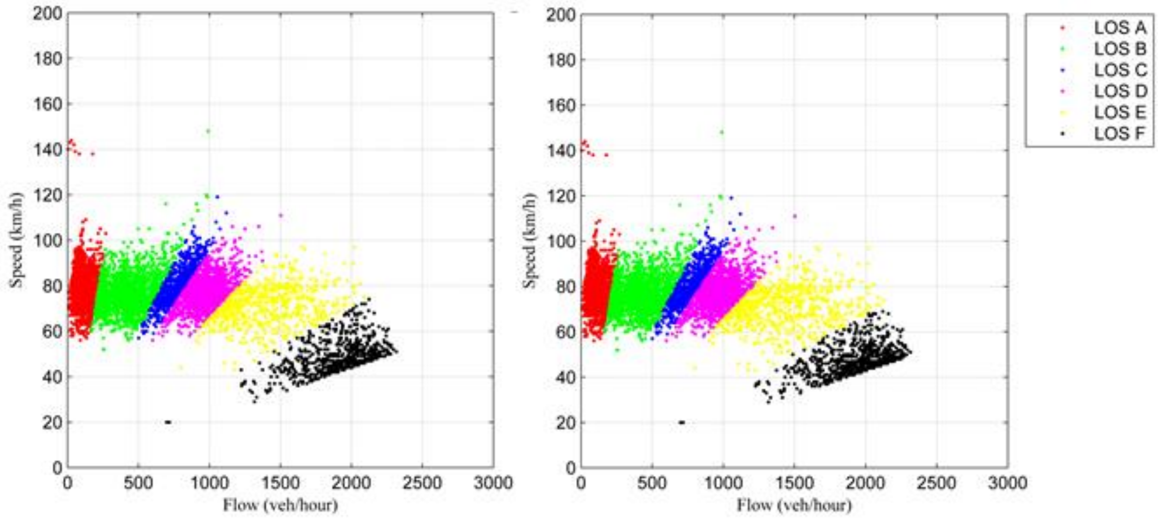

Figure 4.19 : City Block Method result for 7680 and 7800.

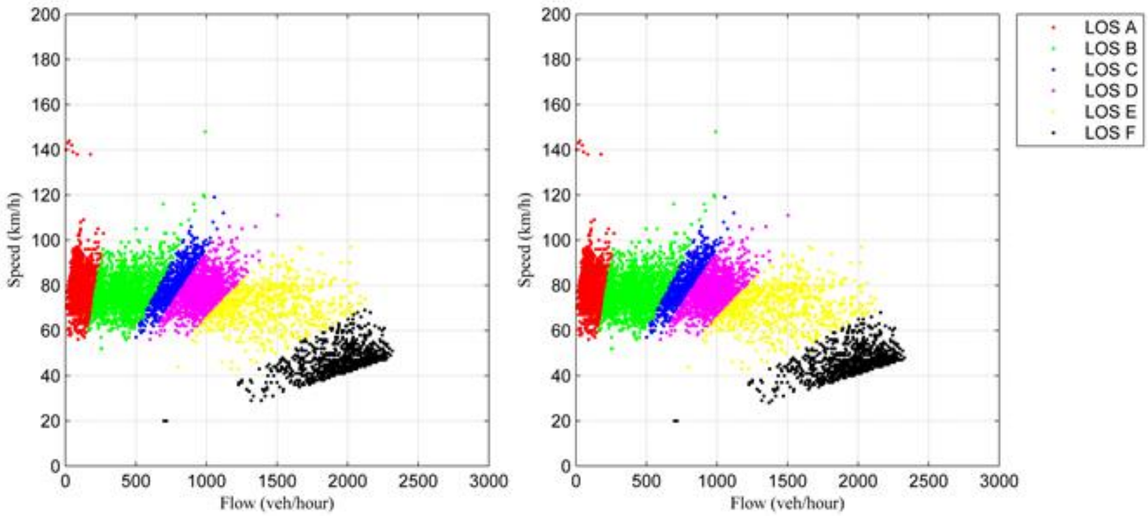

Figure 4.20 : City Block Method result for 7920 and 8040.

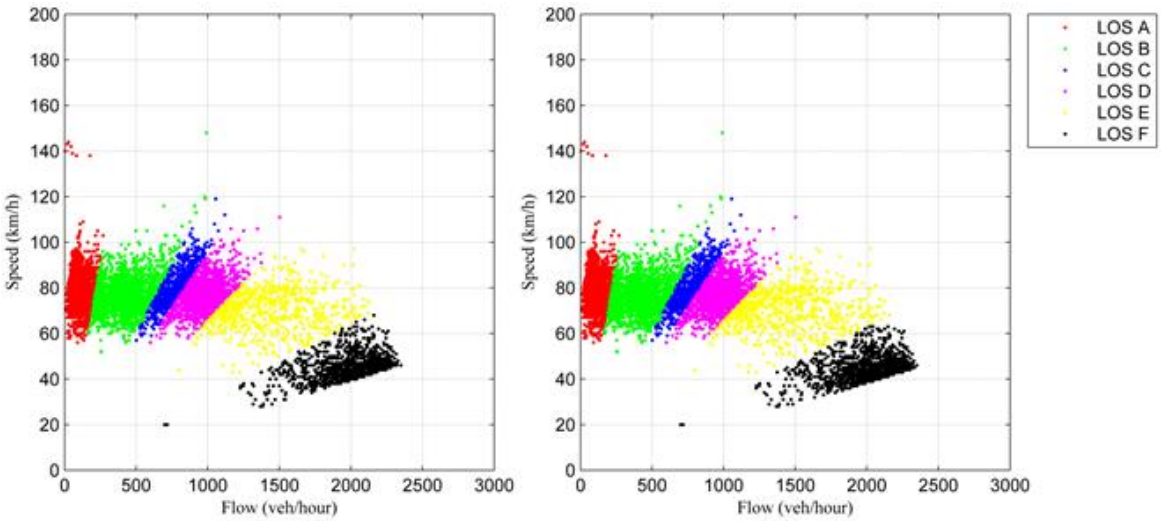

Figure 4.21 : City Block Method result for 8160 and 8280.

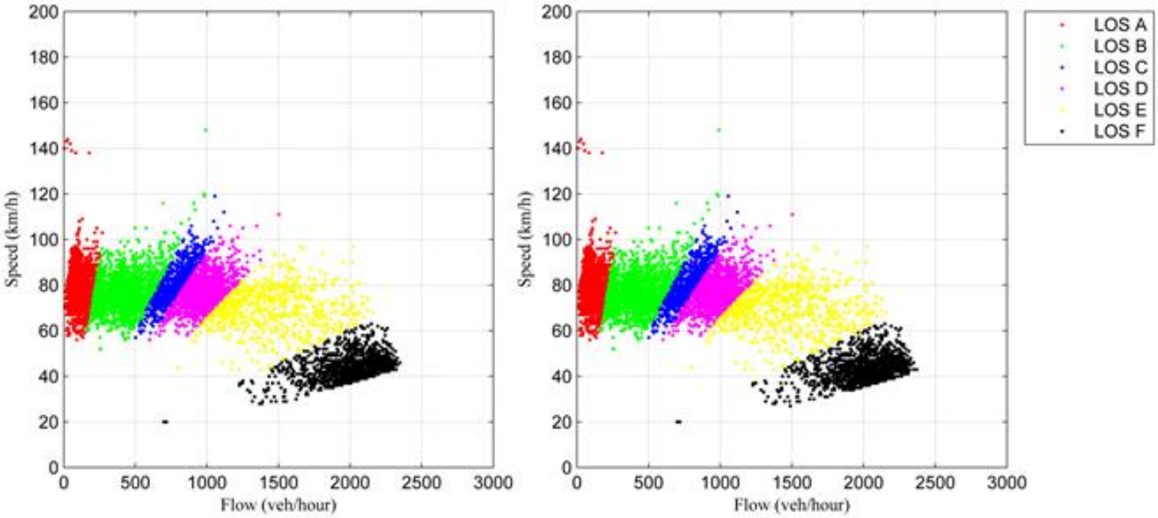

Figure 4.22 : City Block Method result for 8400 and 8520.

4.4.3 Clustering by fuzzy c-means

Figure 4.23 is provided to present the process of fuzzy c-means clustering. It should be noted in Step 4, fuzziness parameter, termination criteria and norm- induce matrix are relatively selected as 2, 0.00001 and A=I (Pal and Bezdek, 1995).

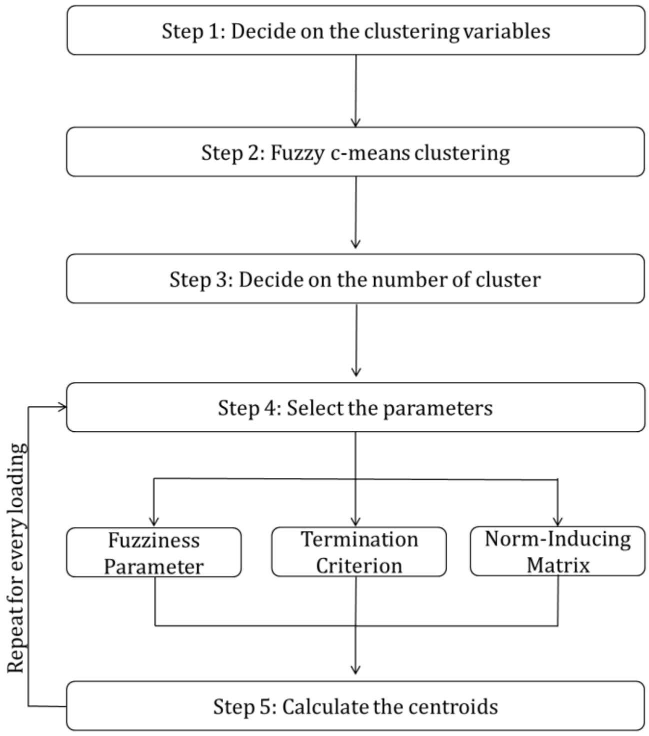

Figure 4.23 : Flow chart of fuzzy c-means clustering.

Figure 4.24, Figure 4.25, Figure 4.26, Figure 4.27 and Figure 4.28 show each 4 hours time interval that starts from 7440, and ends in 8520.

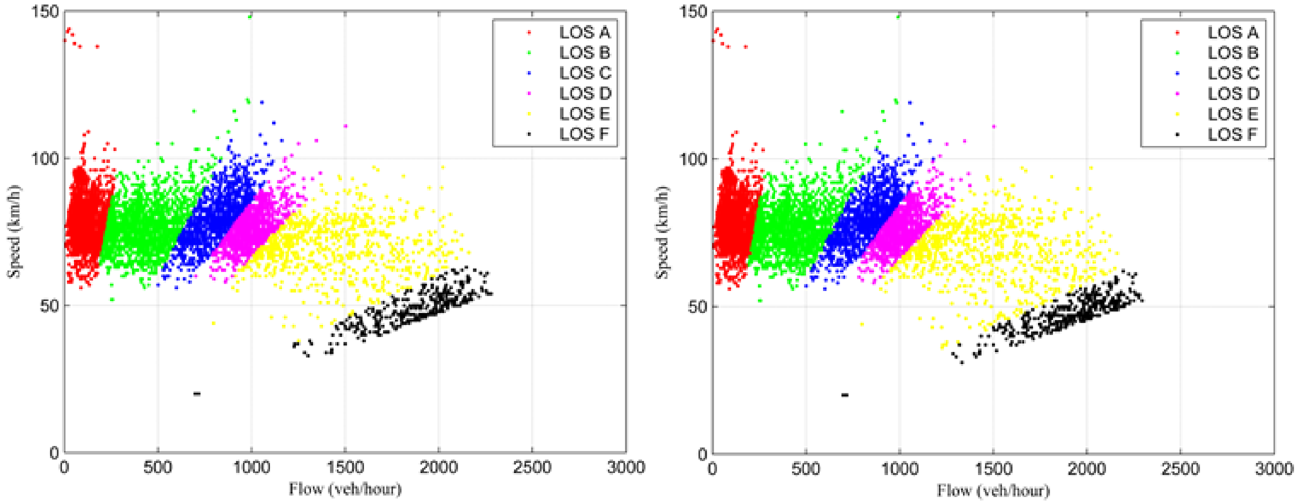

Figure 4.24 : Fuzzy C-means Method result for 7440 and 7560.

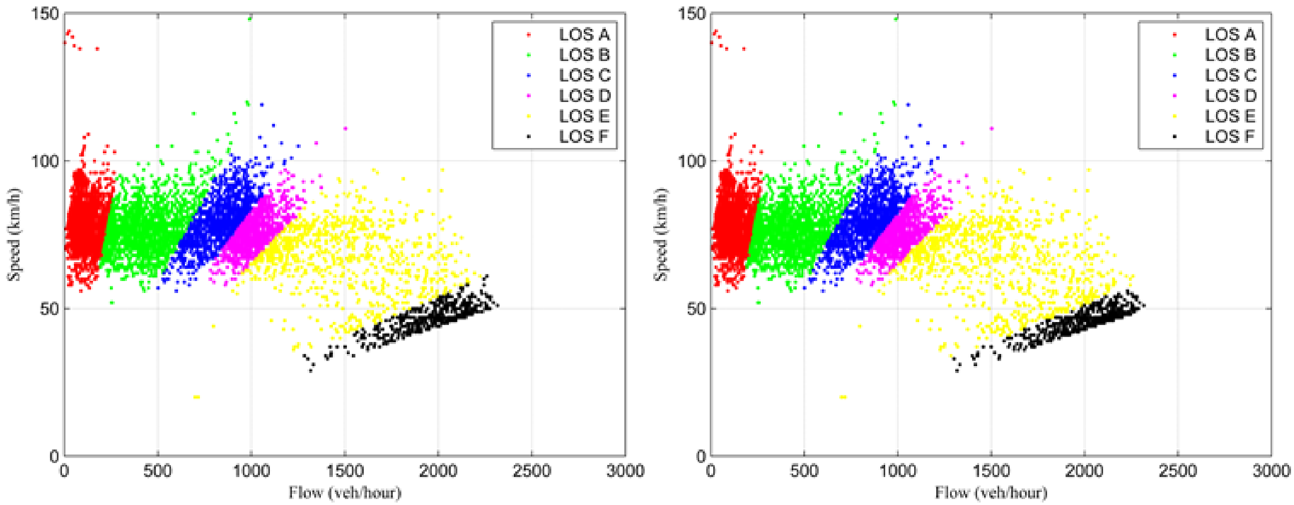

Figure 4.25 : Fuzzy C-means Method result for 7680 and 7800.

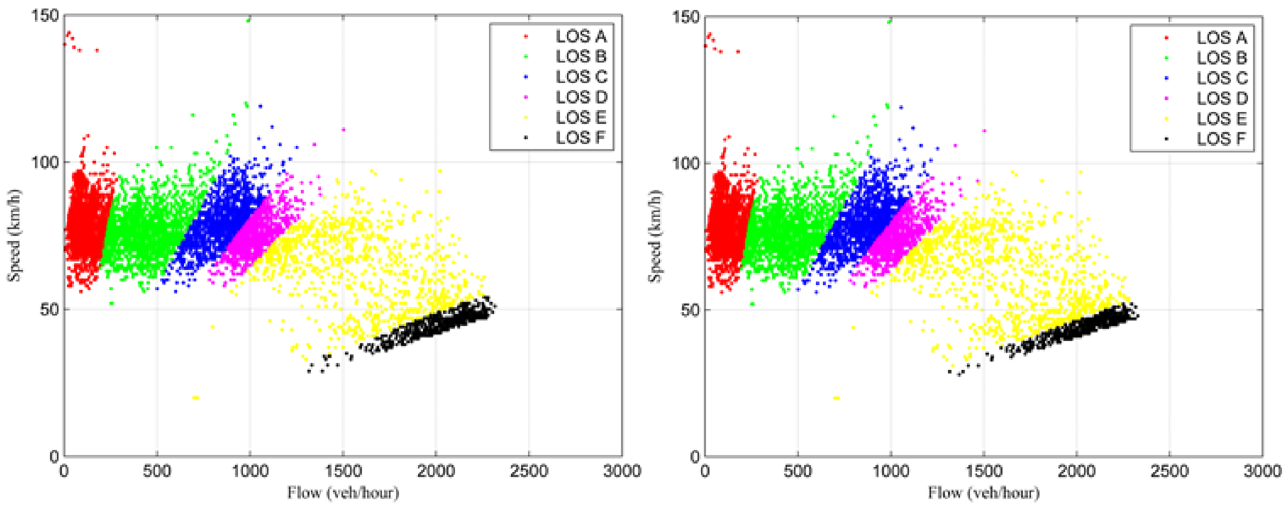

Figure 4.26 : Fuzzy C-means Method result for 7920 and 8040.

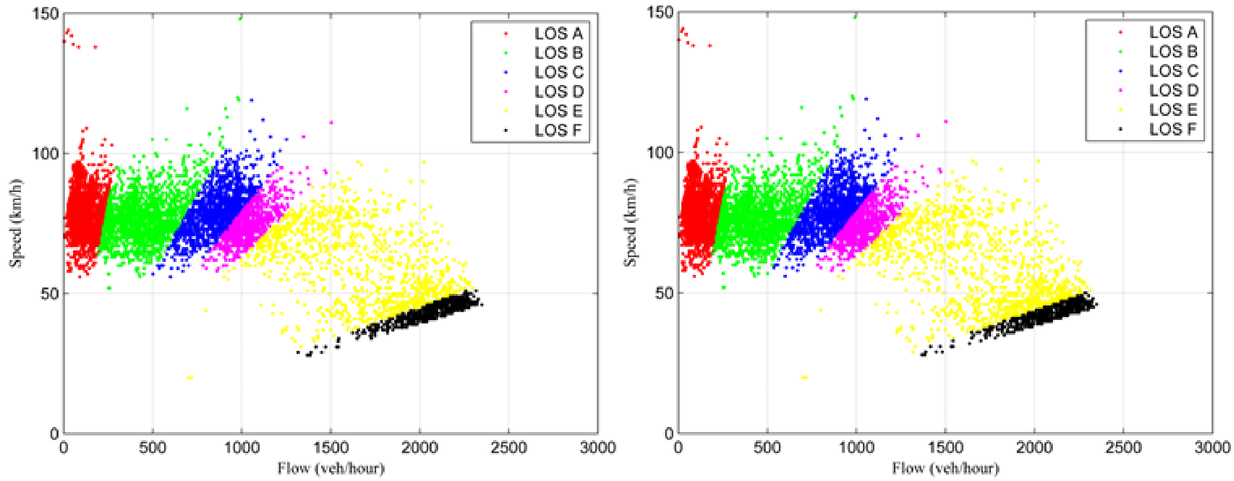

Figure 4.27 : Fuzzy C-means Method result for 8160 and 8280.

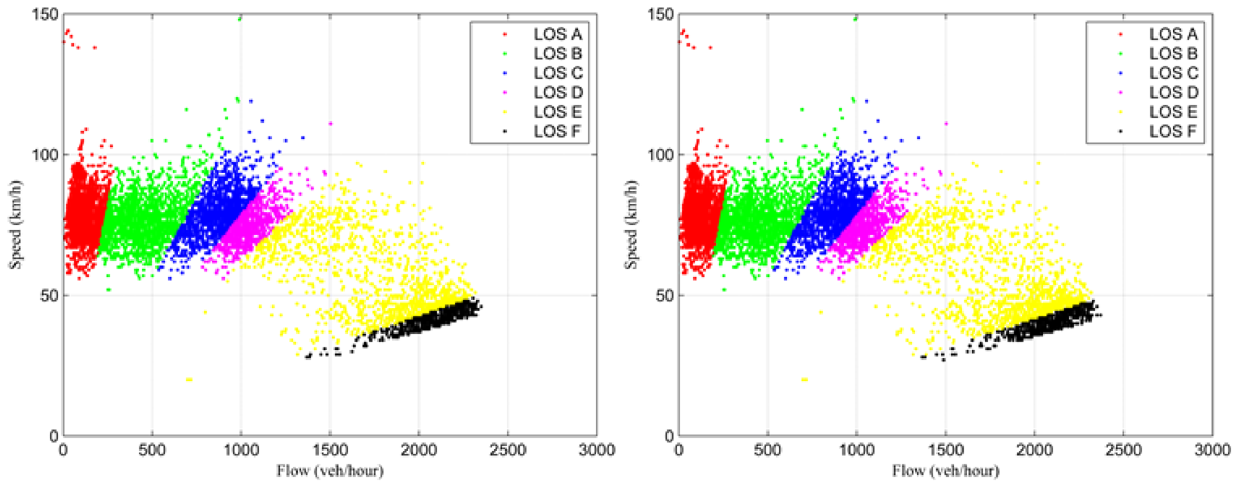

Figure 4.28 : Fuzzy C-means Method result for 8400 and 8520.

4.5 Comparative Evaluations

In order to comparatively evaluate the clustering performances of multivariate and fuzzy methods on lane-based densities relative to deterministic clustering, a number of statistical criteria, including the root mean squared error (RMSE), the mean square error (MSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the coefficient of determination (R^2).

The performances of K- means clustering and fuzzy c- means clustering, using both the Square Euclidean and City Block Distance measures, are evaluated in two cases, with two assumptions.First is the case that static centroids for deterministic HCM method are assumed, and called SHCM where;

1) Each level of service is accepted as a cluster.

2) Centroids are calculated by subtracting the boundary values of the LOS in interest, as given in Equation **(4.4)**,
3) And the result is divided into two.

Second case assumes dynamic centroids for deterministic HCM method and is called DHCM where;

1) Each level of service is accepted as a cluster,
2) Mean for each level of service is calculated,
3) And the previous step is repeated for each loading.

In the light of the information given on computing centroids for level of service , the boundaries are defined in Highway Capacity Manual (2010), Figure 4.29 shows how the boundaries are computed for K-means and Fuzzy c- means.

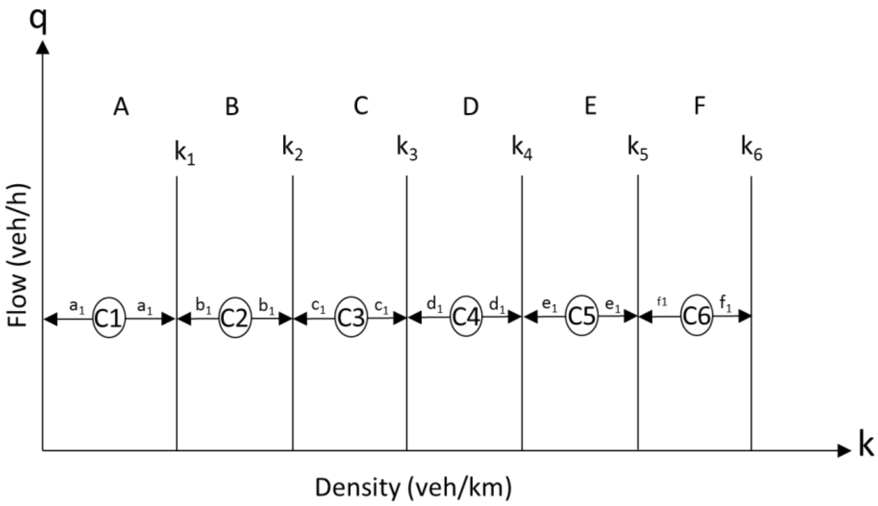

Figure 4.29 : Computing sketch of centroids.

The coefficient of determination is defined as the proportion of variance 'explained' by the regression model makes it useful as a measure of success of predicting the dependent variable from the independent variables (Nagelkerke, 1991). It is computed as Equation **(4.22)**

$$R^2 = \left(\frac{n\sum P_\mathrm{i}R_i - (\sum P_i)(\sum R_i)}{\left(n\left(\sum {P_i}^2\right) - \left(\sum P_i\right)^2\right)\left(n\left(\sum {R_i}^2\right) - \left(\sum R_i\right)^2\right)}\right)^2 \quad \textbf{(4.22)}$$

where P_i is filtered and modeled values, R_i is value of clustering result and n is number of level of services. Table 4.3 and Table 4.4 represent respectively the coefficients of determinant for K-means methods, fuzzy c-means to DHCM and SHCM.

Table 4.3 : Coefficient of Determination for DHCM.

Clustering Method	Size of Data Set									
	7440	7560	7680	7800	7920	8040	8160	8280	8400	8520
Fuzzy c-means	0,98	0,98	0,97	0,97	0,97	0,96	0,96	0,96	0,95	0,95
Square Euclidiean	0,98	0,98	0,98	0,97	0,97	0,97	0,96	0,96	0,96	0,96
City Block	0,94	0,94	0,97	0,98	0,98	0,98	0,98	0,98	0,98	0,98

Table 4.4 : Coefficient of Determination for SHCM.

Clustering Method	Size of Data Set									
	7440	7560	7680	7800	7920	8040	8160	8280	8400	8520
Fuzzy c-means	0,96	0,97	0,96	0,96	0,96	0,95	0,95	0,95	0,95	0,95
Square Euclidiean	0,96	0,97	0,96	0,97	0,96	0,96	0,96	0,96	0,96	0,96
City Block	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99

The RMSE is a frequently used measure of the difference between values predicted by a model and the values actually resulted from the clustering that is being modelled. The RMSE is computed as given by Equation **(4.23)**:

$$\sqrt{\frac{\sum_{i=1}^{n}(P_i - R_i)^2}{n}} \qquad \textbf{(4.23)}$$

Figure 4.30 and Figure 4.31 show respectively the variations of RMSE for DHMC and SHMC to fuzzy c-means and K-means clustering.

Figure 4.30 : Variation of RMSE for DHCM to Fuzzy c- means and K-means clustering.

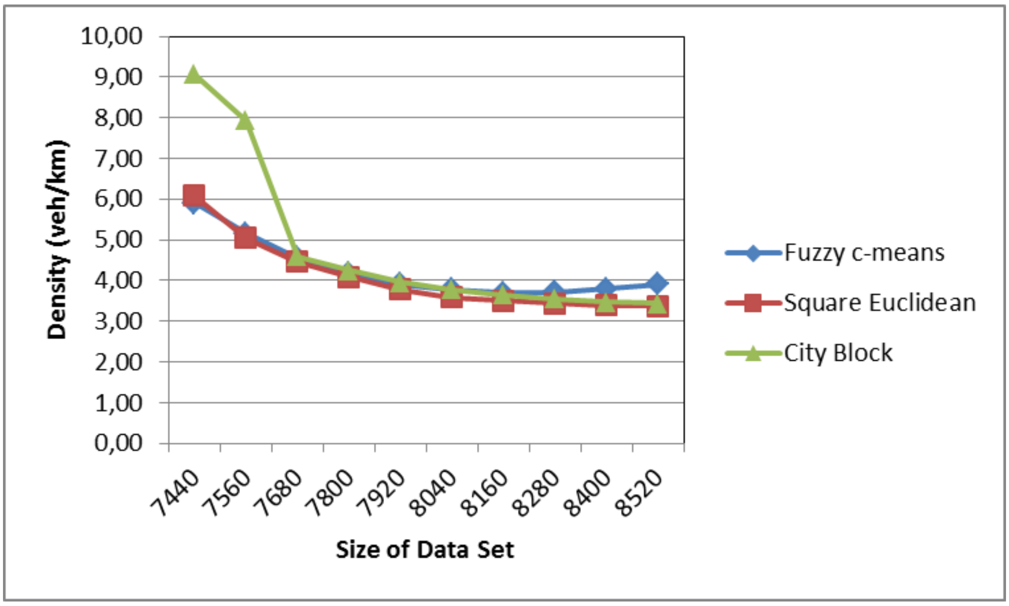

Figure 4.31 : Variation of RMSE for SHCM to Fuzzy c- means and K-means clustering.

The MAE is a quantity used to measure how close predictions are to the eventual outcomes. MAE can be calculated as given in Equation **(4.24)**;

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - R_i| \tag{4.24}$$

Figure 4.32 and Figure 4.33 show respectively the variations of MAE for DHMC and SHMC to fuzzy c-means and K-means clustering.

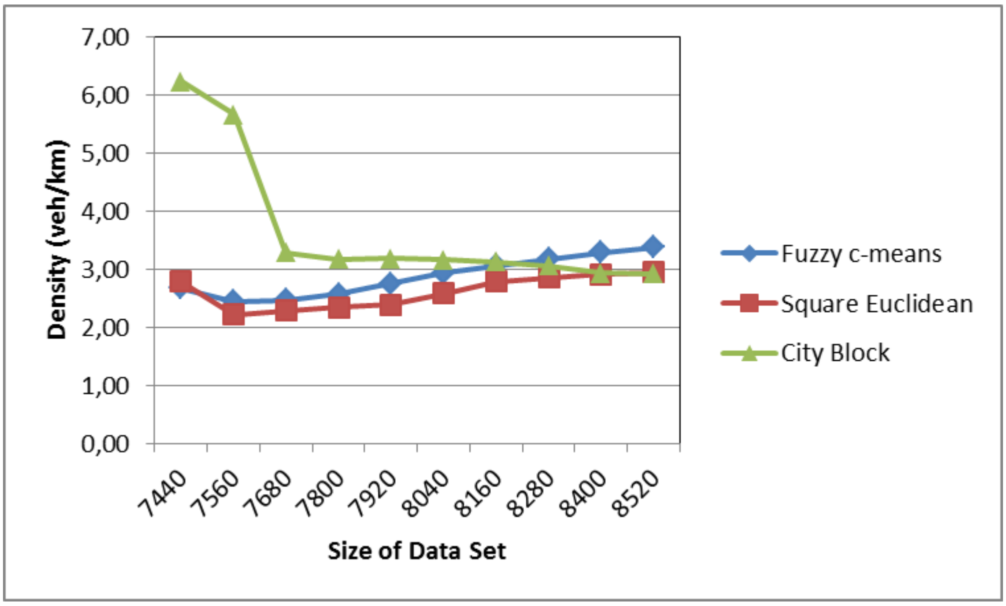

Figure 4.32 : Variation of MAE for DHCM to Fuzzy c- means and K-means clustering.

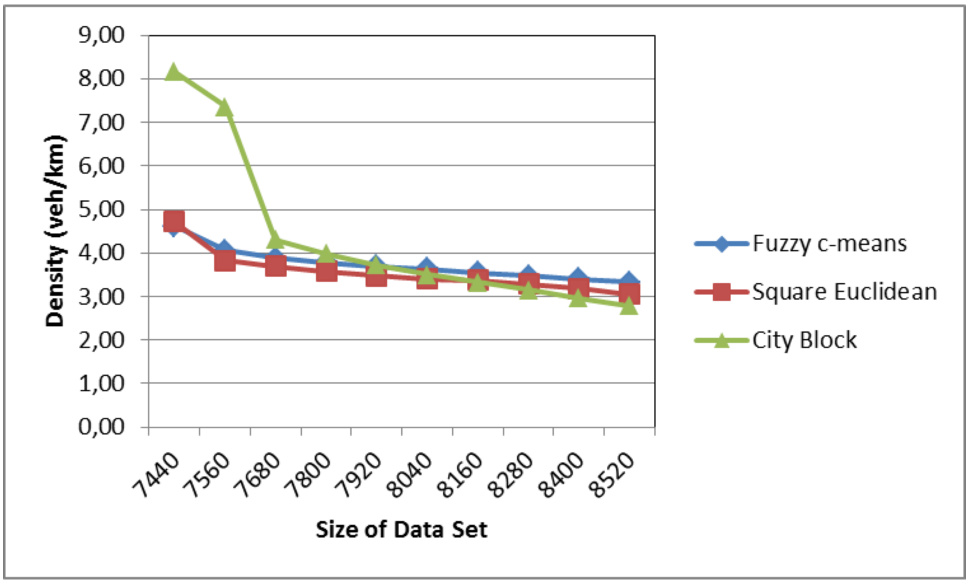

Figure 4.33 : Variation of MAE for SHCM to Fuzzy c- means and K-means clustering.

The mean square error of an estimator measures the average of the squares of the "errors", that is, the difference between the estimator and what is estimated. MSE can be computed as given by Equation **(4.25)**;

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - R_i)^2 \tag{4.25}$$

Figure 4.34 and Figure 4.35 show respectively the variations of MSE for DHMC and SHMC to fuzzy c-means and K-means clustering.

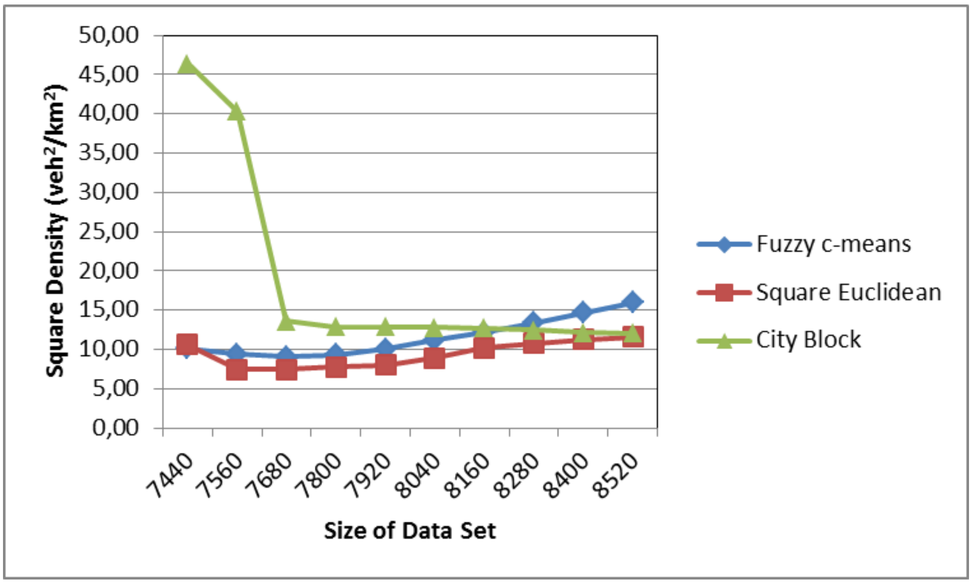

Figure 4.34 : Variation of MSE for DHCM to Fuzzy c- means and K-means clustering.

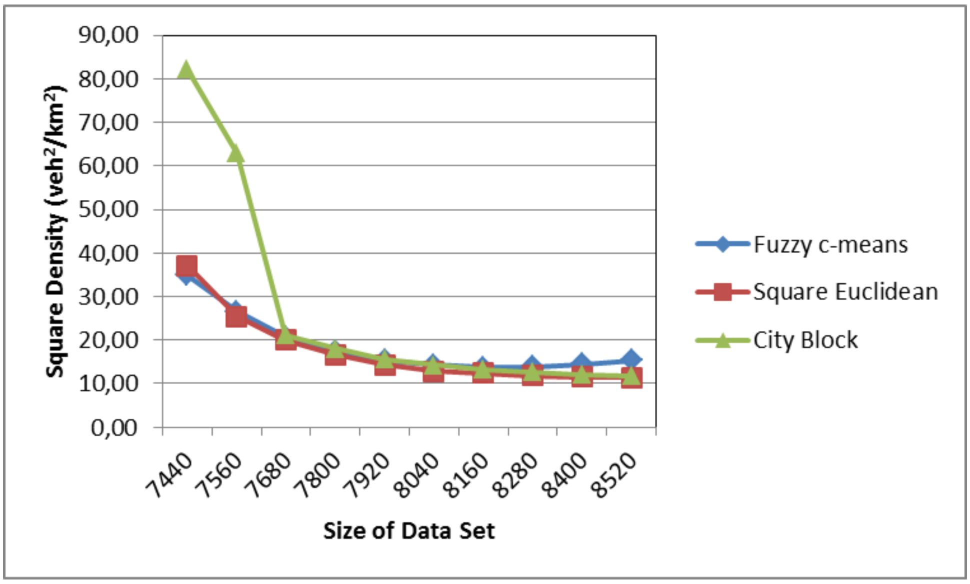

Figure 4.35 : Variation of MSE for SHCM to Fuzzy c- means and K- means clustering.

The mean absolute percentage error (MAPE) is a measure of accuracy of a method for constructing fitted time series values in statistics, specifically in trend estimation. It is defined by the formula;

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_i - R_i}{P_i}\right| \quad \textbf{(4.26)}$$

Figure 4.36and Figure 4.37 show respectively the variations of MAPE for DHMC and SHMC to fuzzy c-means and K-means clustering.

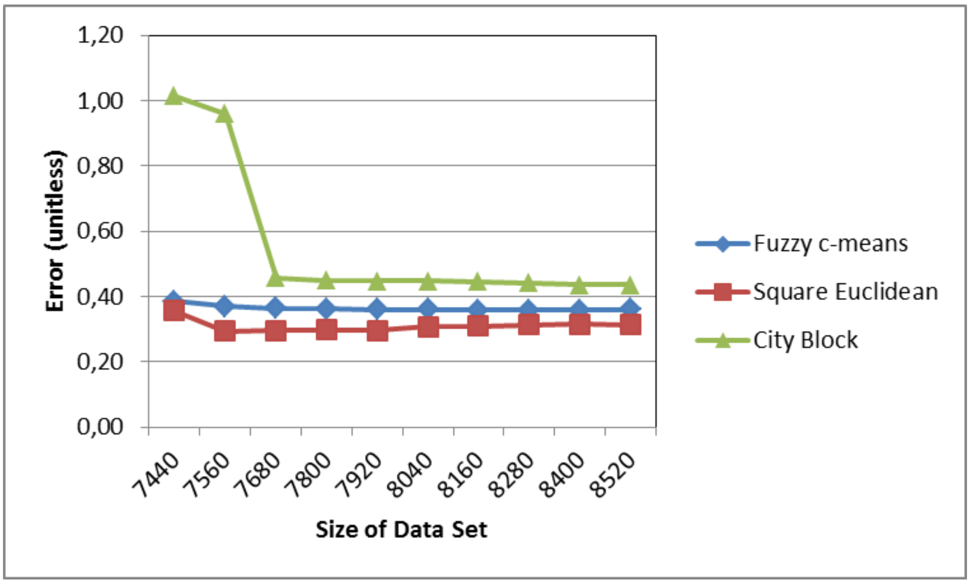

Figure 4.36 : Variation of MAPE for DHCM to Fuzzy c- means and K- means clustering.

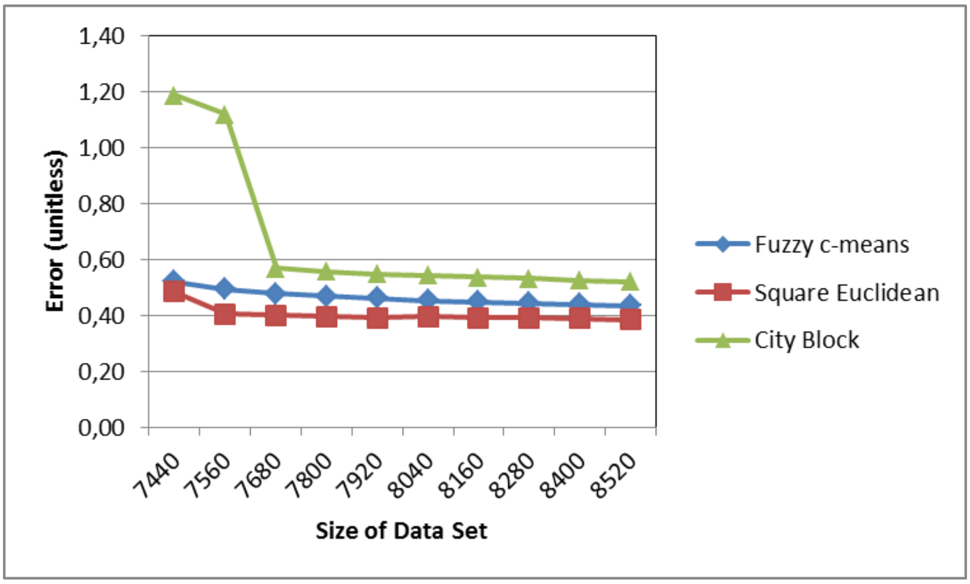

Figure 4.37 : Variation of MAPE for SHCM to Fuzzy c- means and K-means clustering.

It is shown that for all the figures which represent variations of error respectively RMSE, MAE, MSE and MAPE , K- means clustering by City Block Distance descended sharply between 7440 and 7680 before plunging back down to its normal level. In contrast to City Block Distance method, both Square Euclidean and fuzzy c-means methods show either a slight decrease or increase together.

5. CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

The aim of the current study has been to employ and comparatively evaluate the performance of clustering methods in helping to capture flow state variations over the fundamental digram of traffic flow. On this purpose the dynamic classification performance of multivariate and fuzzy clustering methods are investigated using outputs of a traffic simulation model run with the point measurement data over a freeway segment.

Three clustering methods, all of which are dynamic due to their processing nature, are separately applied to cluster flow conditions that are simulated by a macroscopic traffic flow model. The comparative evaluation is presented considering the static level of service classification approach in Highway Capacity Manual.

The perfomance analyses are focused explicitly on the application of clustering for partitioning states of traffic flow, using three clustering algorithms: K-means clustering by Square Euclidean Distance, K-means clustering by City Block Distance and fuzzy C-means. Clustering algorithms have the flexibility to specify the number of clusters. Instead of determining the number of clusters, it has been used user-defined number of clusters as six.

Categorization has been based on similarities and dissimilarities of traffic flow variables without specifying arbitrary values to bound states. Clustering methods are processed in a time-varying fashion to partition the fundamental diagrams at selected temporal resolutions. As it is seen from the above section, system is loaded for each 4 hours. Clustering measures at selected times are used to statistically evaluate each method’s performance. It is straight to conclude from the comprasions presented that the K-means clustering by Square Euclidean Distance and fuzzy c-means methods perform better and appear to yield to classifications consistent with two types of LOS, which are calculated by using HCM-defined LOS.

5.2 Future Research Directions

As a future extension of the present study, the author suggests setup an overall dynamic clustering process that begins by collecting data and ends in the interpretation of clustering results, so that flow state variations among successive time intervals can be sequentially captured and help to specify discontinuities in traffic flow, such as breakdowns and capacity drops. The usefulness of such an overall process in modeling and detection, and hence, in management of incidents is out of question. Aside, separating the level of service E from the level of service F within the clustering process also deserves further investigation that may affect the performance of unstable flow state classification.

REFERENCES

Anderberg, M.J. (1973) Cluster Analysis for applications, Academic Press, 1973.

Antoniou, C., Koutsopoulos, H. N., Yannis, G. (2013) Dynamic data-driven local traffic state estimation and prediction, *Transportation Research Part C*, 34.

Azimi, M., Zhang, Y. (2010) Categorizing Freeway Flow Conditions by Using Clustering Methods, *Transportation Research Board*, 2173.

Banks, J. H. (1999) Investigations of Some Characteristics of Congested Flow. *In Transportation Research Record: Journal of the Transportation Research Board*, pp. 128-134, No. 1678, TRB, National Research Council, Washington, D.C.

Bezdek, J. C. (1981) Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

Bin, Z., Chen, J., Xia, S., Jin, Y. (2008) Data Analysis of Vessel Traffic Flow Using Clustering Algorithms, *International Conference on Intelligent Computation Technology and Automation*.

Celikoglu, H. B. (2007) A dynamic network loading model for traffic dynamics modeling, *IEEE Transactions on Intelligent Transportation Systems*, 8(4), 575-83.

Celikoglu, H.B. and Dell'Orco, M. (2007). Mesoscopic simulation of a dynamic link loading process. Transportation Research Part C: Emerging Technologies, vol. 15, no. 5, pp. 329–344.

Celikoglu, H.B. (2013) An Approach to Dynamic Classification of Traffic Flow Patterns. *Computer-Aided Civil and Infrastructure Engineering*, vol. 28, no. 4, pp. 273-288.

Çelikoğlu, H.B., Gedizlioğlu, E. (2012) Lecture Notes in Trafik Teknigi ve Etudleri, ITU.

Çelikoglu, H.B., Gürcanli, G.E., and Ciğizoğlu, H.K. (2010) Hybrid Models of Neural Network Method for Road Safety Regulations: Safety Index Calibration and Intelligent Transportation Systems Based Safety Control (in Turkish). Final report of national research project with no. 108M299, the Scientific and Technological Research Council of Turkey (TÜBİTAK)

D.R. Drew (1965) Deterministic Aspects of Freeway Operations and Control. *Highway Research Record*, 99: pp. 48-58.

Daganzo, C.F., (1994). The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research Research Part B: Methodological*, vol. 28, no.4, pp. 269-287.

Drake, J., J. Schofer, and A. D. May (1967) A Statistical Analysis of Speed Density Hypotheses. Proc., *3rd International Symposium on the Theory of Traffic Flow*, Elsevier North Holland, Inc., New York.

Dunn, J. (1974) C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, Vol. 3, No. 3, pp. 32-57.

Edie, L.C., Discussion on traffic stream measurements and definitions. *In Proceedings of the 2nd Int. Symp. Theory Traffic Flow*, Paris, France, page 139154, 1963

Gartner N., Messer , C.J., Rathi, A. K. (2001)**,** Revised Monograph of Traffic Flow Theory: A State-of-the-Art Report. *Transportation Research Board*.

Gerlough, D.L., Huber, M. J. (1975)**,** Traffic Flow Theory A Monograph, *Transportation Research Board*, Special Report 165, TRB, Washington D.C.

Greenberg, H. (1959), An Analysis of Traffic Flow. Operations Research, Vol. 7, pp. 79-85.

Greenshields, B. D. (1935), A Study of Traffic Capacity. *Highway Research Board Proceedings*, Vol. 14, pp. 448-477.

Haight, F.A. (1971)**,** Vehicles as Particles. (Book Reviews: Kinetic Theory of Vehicular Traffic by Prigogine and Herman). *Science*, 173(3996):513.

Hall, F. L., V. F. Hurdle, and J. H. Banks (1992) Synthesis of Recent Work on the Nature of Speed-Flow and Flow-Occupancy (or Density) Relationships on Freeways. *In Transportation Research Record 1365, TRB, National Research Council*, Washington, D.C., pp. 12-18.

Helbing, D., A. Hennecke, and M. Treiber (1999). Phase Diagram of Traffic States in the Presence of Inhomogeneities. *Physical Review Letters*, Vol. 82, No. 21, pp. 4360-4363.

Herrera, J. C. & Bayen, A. M. (2010), Incorporation of Lagrangian measurements in freeway traffic pattern estimation, *Transportation Research Part B: Methodological*, 44(4), 269-87.

Highway Capacity Manual (1985), Transportation Research Board of The National Academies.

Highway Capacity Manual (1994), Transportation Research Board of The National Academies.

Highway Capacity Manual (1997), Transportation Research Board of The National Academies.

Highway Capacity Manual (2000), Transportation Research Board of The National Academies.

Highway Capacity Manual (2010), Transportation Research Board of The National Academies.

Illner, R., Bohun, C.S., McCollum, S. and Roode, T.V. (2005) Mathematical Modelling: A Case Study Approach. American Mathematical Society.

Jain, A.K., Dubes, R.C. (1988) Algorithms for Clustering Data, Prentice Hall, NJ.

Jain, A.K., Murty, M.N., Flynn, P.J. (1999) **,** Data Clustering: A Review, *ACM Computing Surveys*, Vol. 31, No. 3.

Kerner, B. & Rehborn, H. (1996)**,** Experimental properties of complexity in traffic flow, *Physical Review E*, 53(5), R4275-R78.

Kerner, B. S. (2004), The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory, Springer, Heidelberg.

Lighthill, M.J. and Whitham, G. B., (1955). On Kinematic Waves II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society of London Part A*, vol. 229, no. 1178, pp. 317-345.

MacQueen, J. (1967) Some Methods for Classification and Analysis of Multivariate Observations. Proc., *5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1 (L. M. Le Cam and J. Neuman, eds.), University of California Press, Berkeley, pp. 281-297.

May, A. (1990), Traffic Flow Fundamentals. Prentice Hall, Englewood Cliffs, N.J.

Mooi, E. and Sarstedt, M., (2011). A Concise Guide to Market Research, Springer-Verlag , Berlin.

Munjal, P.K. and Pipes, L.A. (1971),. Propogation of On-Ramp Density Perturbations on UniDirectional and Two and Three Lane Freeways. *Transportation Research Part B*, 5(4):241-255.

Munoz, L., Sun, X., Horowitz, R. & Alvarez, L. (2003), Traffic density estimation with the Cell Transmission Model, *in Proceedings of the 2003 American Control Conference*, Denver, Colorado June 4-6, Vol. 5, 3750-55.

Murat, Y.S. ve Sekerler, A. (2009). "Trafik kaza Verilerinin Kumelenme Analizi Yontemi ile Modellenmesi*". iMO, Teknik Dergi*. S. 4759-4777.

Nagelkerke, N.J.D. (1991) A Note on a General Definition of the Coefficient of Determination, *Biometrika*, Vol. 78, No. 3.

Ni, Daiheng (2013). Lecture Notes in Traffic Flow Theory. University of Massachusetts Amherst.

Oh, C., Tok, A., Ritchie, S.G. (2005) Real-Time Freeway Level of Service Using Inductive-Signature-Based Vehicle Reidentification System. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 6, No. 2, pp. 138-146.

Pal, N.R. and Bezdek, J.C. (1995), On cluster validity for the fuzzy c-means model. *IEEE T. Fuzzy Systems* Vol 3, No:3.

Park, B. (2002) Hybrid Neuro-Fuzzy Application in Short-Term Freeway Traffic Volume Forecasting. *In Transportation Research Record: Journal of the Transportation Research Board, No. 1802, Transportation Research Board of the National Academies*, Washington, D.C., pp. 190-196.

Richards, P.I., (1956). Shock Waves on The Highway. *Operations Research*, vol. 4, no. 1, pp. 42-51.

Sumalee, A., Zhong, X., Pan, L. & Szeto, Y. (2011), Stochastic cell transmission model (SCTM): a stochastic dynamic traffic model for traffic pattern surveillance and assignment, *Transportation Research Part B: Methodological*, 45(3), 507-33.

Sun, L., and J. Zhou (2005). Development of Multiregime Speed-Density Relationships Using Cluster Analysis. *In Transportation Research Record: Journal of the Transportation Research Board, No. 1934, Transportation Research Board of the National Academies*, Washington, D.C., pp. 64-71.

Sun, X., Muñoz, L., and Horowitz, R., (2003). Highway Traffic Pattern Estimation Using Improved Mixture Kalman Filters for Effective Ramp Metering Control. *Proceedings of the 42nd IEEE Conference on Decision and Control*, Vol. 6, pp. 6333-6338.

Treiber, M. & Helbing, D. (2002), Reconstructing the spatio-temporal traffic dynamics from stationary detector data, *Cooperative Transportation Dynamics*, 1(3), 3.1-3.24.

Treiber, M., Kesting, A. & Helbing, D. (2010), Three-phase traffic theory and two-phase models with a fundamental diagram in the light of empirical stylized facts, *Transportation Research Part B: Methodological*, 44(8-9), pp. 983-1000.

Underwood, R.T. (1961)**,** . Speed, Volume and Density Relationships. Quality and Theory of Traffic Flow, *Yale University Report*, New Haven, Connecticut.

US DOT (Department of Transportation) FHWA (Federal Highway Administration) (2006), Traffic Detector Handbook, 3rd edn, Vol. I, Publication No. FHWA-HRT-06-108.

Vlahogianni, E I., Karlaftis, M. G., Golias, J.C. (2014), Short-term Traffic Forecasting: Where We Are and Where We're Going. *Transportation Research Part C: Emerging Technologies*, 43(1), 3-19, 2014.

Wahba, G., (1990). Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, *Society for Industrial and Applied Mathematics (SIAM)*, Philadelphia.

Wang, Y. and Papageorgiou, M., (2005). Real-Time Freeway Traffic Pattern Estimation Based on Extended Kalman Filter: A General Approach. *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 141-167.

Wang, Y., Papageorgiou, M., and Messmer, A., (2007). Real-Time Freeway Traffic Pattern Estimation Based on Extended Kalman Filter. *Transportation Science*, vol. 41, no. 2, pp. 167-181.

Wei, L. and Li, M. (2012), A fuzzy fusion algorithm to identify the traffic states by floating car data, *9th International Conference on Fuzzy Systems and Knowledge Discovery*.

Xia, J., and Chen, M. (2007), A Nested Clustering Technique for Freeway Operating Condition Classification. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 22, No. 6, pp. 430-437.

Xia, J., and Chen, M. (2007), Defining Traffic Flow Phases Using Intelligent Transportation Systems-Generated Data. *Journal of Intelligent Transportation Systems*, Vol. 11, No. 1, pp. 15-24.

Xu, R. and Wunsch D.C. (2008), Clustering, *Wiley-IEEE Press*, New Jersey.

Yang, H., and Qiao, F. (1998),. Neural Network Approach to Classification of Traffic Flow States, *Journal of Transportation Engineering*, Vol. 124, No. 6, pp. 521-525.

CURRICULUM VITAE

Name Surname: Mehmet Ali Silgu

Place of Birth: Istanbul

Date of Birth: 26/ 01/ 1988

E-mail: msilgu@itu.edu.tr

EDUCATION:

B.Sc.: 2010, Karadeniz Technical University, Faculty of Engineering, Geological Engineering

2011, Karadeniz Technical University, Faculty of Engineering, Civil Engineering

PUBLICATIONS ON THE THESIS:

Silgu, M.A., Celikoglu, H.B.,(2014) K-means clustering method to classify freeway traffic flow patterns, *Pamukkale University Journal of Engineering Sciences*, Vol. 20, No. 6, s. 232-239, ISSN: 2147-5881, (doi: 10.5505/pajes.2014.36449).

Silgu, M.A., Celikoglu, H.B., (2014) Non-hierarchical clustering methods to classify freeway traffic flow states, *11th International Congress on Advances in Civil Engineering*, Istanbul, Turkey, October 21 – October 25, 2014.