LEVEL OF SERVICE CRITERIA OF URBAN STREETS IN INDIAN CONTEXT USING ADVANCED CLASSIFICATION TOOLS

SMRUTI SOURAVA MOHAPATRA

DEPARTMENT OF CIVIL ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY

ROURKELA-769008.

2012

LEVEL OF SERVICE CRITERIA OF URBAN STREETS IN INDIAN CONTEXT USING ADVANCED CLASSIFICATION TOOLS

Thesis

Submitted in partial fulfillment of the requirements
For the degree of

Master of Technology
in
Transportation Engineering

By
Smruti Sourava Mohapatra
Roll No. **210CE3279**

Under the guidance of

Prof. U. Chattaraj

Prof. P.K. Bhuyan

DEPARTMENT OF CIVIL ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA-769008.
2012

NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA- 769008

CERTIFICATE

This is to certify that project entitled, **"Level of Service Criteria of Urban Streets in Indian Context using Advanced Classification Tools"** submitted by **SMRUTI SOURAVA MOHAPATRA** in partial fulfillment of the requirements for the award of **Master of Technology** Degree in **Civil Engineering** with specialization in **Transportation Engineering** at National Institute of Technology, Rourkela is an authentic work carried out by him under my supervision and guidance. To the best of my knowledge, the matter embodied in this Project review report has not been submitted to any other university/ institute for award of any Degree or Diploma.

Prof. U. Chattaraj
Department of Civil Engineering
National Institute of Technology,
Rourkela- 769008

Prof. P.K. Bhuyan
Department of Civil Engineering
National Institute of Technology,
Rourkela-769008

ACKNOWLEDGEMENT

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study. First and foremost, my utmost gratitude to Dr**. U. Chattaraj** and **Dr. P.K. Bhuyan** whose sincerity and encouragement I will never forget. **Dr. Chattaraj** and **Dr. Bhuyan** have been my inspiration as I hurdle all the obstacles in the completion this research work.

My sincere thanks to **Prof. M. Panda** former HOD of Civil Engineering Department, NIT Rourkela for providing valuable co-operation and needed advice generously all along my M.Tech study.

I would like to extend my gratefulness to **Dr. S. Sarangi**, Director, NIT Rourkela and **Dr. N. Roy**, HOD, Civil Engineering Department, NIT Rourkela for providing necessary facilities for my research work.

I also want to convey sincere thanks to all my friends at NIT, Rourkela for making my stay in the campus a pleasant one. The co-operation shown by **B.B. Sutradhar**, **S. Purohit** and **D. Kar** my batch mates of Transportation Engineering Specialization cannot be ignored.

Last but not the least, my parents and the one above all of us, the omnipresent God, for answering my prayers for giving me the strength and courage always.

Smruti Sourava Mohapatra

Roll No-210CE-3279

CONTENTS

Items	Page No.
Certificate	**i**
Acknowledgement	**ii**
Contents	**iii**
Abstract	**vi**
List of Figures	**viii**
List of Tables	**x**
Symbols	**xi**
1. Introduction	**1-9**
1.1 General	1
1.2 Statement of Problem	6
1.3 Objective and Scope	7
1.4 Organization of Report	9
2. Urban Streets and Level of Service	**10-20**
2.1 Introduction	10
2.2 Urban Streets	
2.2.1 Introduction	10
2.2.2 Functional and design categories of urban streets	11
2.2.3 Determining urban street class	13
2.2.4 Segmenting the urban streets	13
2.3 Level of Service	14
2.3.1 Introduction	14
2.3.2 Factors affecting Level of Service	16
2.4 Methodology	19
2.5 Summary	20

3. Review of Literature 21-33

3.1 Introduction 21
3.2 Use of GIS-GPS tool in collecting Highway Inventory data 25
3.3 Methods of Cluster Analysis 27
3.3.1 Introduction 27
3.3.2 Self Organizing Map (SOM) Clustering 28
3.3.3 Affinity Propagation (AP) Clustering 29
3.3.4 GA-Fuzzy Clustering 31
3.3.5 Partition Around Medoid (PAM) Clustering 32
3.4 Summary 33

4. Study Area and Data Collection 34-41

4.1 Introduction 34
4.2 Study Corridor and Data Collection 34
4.2.1 Base Map Preparation 34
4.2.2 Study Corridors 34
4.2.3 Data Collection 36
4.3 Aggregating speed on Each Segment 40
4.4 Summary 41

5. Cluster Analysis 42-62

5.1 Artificial Neural Network
5.1.1 Introduction 42
5.1.2 Types of Neural Network 43
5.1.3 SOM Algorithm 45
5.2 Affinity Propagation (AP)
5.2.1 Introduction 46
5.2.2 AP Algorithm 48

5.3 GA-Fuzzy
5.3.1 Introduction 49
5.3.1 GA-Fuzzy Algorithm 51
5.4 Partition Around Medoid (PAM)
5.4.1 Introduction 53
5.4.2 PAM Algorithm 53
5.5 Cluster Validation 55
5.6 Summary 62

6. Result Analysis for Defining LOS 63-91
6.1 Introduction 63
6.2 Application of Cluster Analysis Methods in Defining LOS Criteria of Urban Streets 63
6.2.1 SOM Clustering 64
6.2.2 AP Clustering 70
6.2.3 GA Fuzzy Clustering 77
6.2.4 PAM Clustering 82
6.3 Selecting the Best Clustering Method in Defining LOS Criteria of Urban Streets 87
6.4 Summary 90

7. Summary, Conclusions and Future Scope 90-96
7.1 Summary 92
7.2 Conclusion 93
7.3 Limitation and Future Scope 96

References 97-104
Appendix 105
List of Publications 106

Abstract

In India Level of Service (LOS) is not well defined for urban streets. The analysis procedure followed in India is that developed by HCM 2000. Speed ranges of LOS categories for various urban street classes defined by HCM are appropriate for developed countries having homogenous type of traffic flow. India being a developing country its traffic is very much heterogeneous having vehicles of different operational characteristics. So LOS criteria in Indian context should be defined correctly considering the traffic and geometric characteristics of urban streets. Defining LOS is very important as this is the first step of LOS analysis. LOS analysis is necessitated as this affects planning, design, operational aspects of transportation projects and also allocating resources to the competing projects. Defining LOS is basically a classification problem and application of cluster analysis is found to be a suitable technique to solve the problem. Advanced clustering techniques like Artificial Neural Network (ANN), Affinity Propagation (AP), Genetic Algorithm-Fuzzy (GA-Fuzzy), Partition around Medoid (PAM) can be used to solve the classification problems. Before applying these algorithms for clustering purpose various cluster validation parameters are used to determine the optimal number of cluster for the input data set. This cluster validation parameter holds significance as this decides the numbers of urban street class the free flow speed (FFS) data should be clustered into. After deciding the number of urban street class, the above four algorithms are used in this research in two steps. First, the free flow speed data were clustered to determine the urban street class of each segment. After determination of the urban street class of each segment average travel speed in peak and off peak hour is used in clustering methods to determine the speed ranges of LOS categories. For this study lot of speed data is required for which GPS is found to be the most suitable method of data collection and hence extensively used. Free flow speed and average travel speed during peak and of peak hours and inventory details of road segments are used in this study. All these data are collected from secondary source for this research work. The FFS speed ranges of different urban street class and travel speed ranges of different LOS categories found from different algorithms are found to be different. So to decide the best clustering algorithm for this study four cluster quality evaluation parameters are used. The best clustering algorithm for this study is decided and the FFS and travel speed ranges resulted from this

algorithm is suggested for Indian context. These FFS and travel speed ranges are found to be significantly lower than that mentioned in HCM-2000. Heterogeneous traffic flow and roads having varying geometric and surrounding environmental characteristics are the major reasons for these lower values in FFSs. Presence pedestrians, slow moving vehicles, roadside vendors and on-street parking creates side friction that reduces the travel speed of a commuter. From the clustering analysis of FFS data it can be seen less number of roads in Mumbai are of high speed design (street class-I) or highly congested (street class-IV). More number of road segments is of suburban (street class II) or intermediate (street class III) type. It can be suggested that Greater Mumbai region needs substantial geometric improvements to mitigate the demand of exponentially growing traffic.

Key words: Urban Street, Level of Service, Travel Speed, GPS, Affinity Propagation Clustering, ANN, SOM clustering, PAM clustering, GA-Fuzzy clustering, Cluster Validation

List of Figures

Figure 1.2 Overall framework of the study 8

Figure 2.1 Types of Urban Street Segments 14

Figure 2.2 Urban Streets Methodology for LOS by HCM-2000 19

Figure 5.1 Map showing selected corridors of greater Mumbai 36

Figure 5.2 Probability distribution plot of collected FFS data using GPS receiver 38

Figure 5.1 Flowchart of AP Clustering 48

Figure 6.1 Validation measures for optimal number of clusters using SOM clustering 65

Figure 6.2 SOM weight Positions plot 66

Figure 6.3 SOM sample hit Plot 67

Figure 6.4 SOM Clustering of FFS for Urban Street Classification 68

Figure 6.5 Level of service of urban street classes (I-IV) using SOM clustering on average travel speeds 69

Figure 6.6 Validation measures for optimal number of clusters using AP clustering 71

Figure 6.7 FFS data distribution plot after AP clustering 74

Figure 6.8 AP Clustering of FFS for Urban Street Classification 75

Figure 6.9 Level of service of urban street classes (I-IV) using AP clustering on average travel speed 76

Figure 6.10 Validation measures for optimal number of clusters using GA-Fuzzy clustering 78

Figure 6.11 GA-Fuzzy Clustering of FFS for Urban Street Classification 80

Figure 6.12 Level of service of urban street classes (I-IV) using GA-Fuzzy clustering on average travel speeds 81

Figure 6.13 Validation measures for optimal number of clusters using PAM clustering 84

Figure 6.14 FFS data distribution plot after PAM clustering 85

Figure 6.15 PAM Clustering of FFS for Urban Street Classification 85

Figure 6.16 Level of service of urban street classes (I-IV) using PAM clustering on average travel speeds 86

Figure 6.17 Silhouette width vs. Number of cluster Plot for different clustering algorithms 88

Figure 6.18 Calinski-Harabasz Index vs. Number of cluster Plot for different clustering algorithms 89

Figure 6.19 Dunn Index vs. Number of cluster Plot for different clustering algorithms 89

Figure 6.20 R-square Index vs. Number of cluster Plot for different clustering algorithms 90

List of Tables

Table 2.1 Urban Street LOS by HCM (HCM, 2000) 15

Table 5.1: Parameters of GA-Fuzzy Algorithm 52

Table 6.1 Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by SOM Method 70

Table 6.2 Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by AP Method 77

Table 6.3 Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by GA-Fuzzy Method 82

Table 6.4 Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by PAM Method 87

Table AI. FFS and Average travel speed during peak and off peak hours of Corridor-3 106

Symbols

$a\ (i,k)$	Availabilities Matrix
$a(i)$	average distance of a data point i to other data point in the same cluster
$b(i)$	the average distance of the that particular data point to all the data points belonging to the nearest cluster
C	Cluster centre
c	Number of cluster
d	Distance between data points
d_{ce}	Inter-cluster distance
$d(c_i,c_j)$	Dissimilarity function between two clusters c_i and c_j
$d(i,j)$	Distance between object i and j
D	Sum of the distances between all the pairs of data points of a particular cluster
D_{min}	Minimum value of D possible from the data set
D_{max}	Maximum value of D possible from the data set
D_c	Intra-cluster distance
$i,,j,k$	Matrix indices
$h_{b(x)}$	Neighborhood Function
m	Dimension of the data set
$m_c(t)$	Best Matching Unit (BMU)
$m_i(t)$	Input weight vector
n	Number of data points
n_{ij}	Number of data points of j^{th} dimension that belongs to cluster i
n_j	Number of data points of j^{th} dimension
n_k	Number of data point in cluster k
p	Number of pairs that can be formed taking for data points of a particular cluster
P	Number of pairs can be formed from the data set

r (i,k)	Responsibility Matrix
s(i,k)	Similarity Matrix
x(t)	Any arbitrary input data point
X	Input Matrix
$X_{a,} X_b$	Any arbitrary chosen data point
W(c)	Sum of square of partition of a cluster
$\alpha(t)$	Learning rate of width function
$\phi^{(Q)}$	Quality of cluster
$\phi^{(T)}$	A function to get best quality of cluster with least numbers of clusters
μ_{ki}	Membership function

Chapter 1

Introduction

1.1 General

The rapid growth of urbanization and the rise of megacities with many millions of inhabitants cause tremendous challenges to the developing countries. Keeping with the global trend economic growth has ushered India to go under tremendous urbanization in last century which become more significant after the independence. As for the magnitude, in 1901, only 25 million people constituting 10.84 per cent of population lived in urban areas of India. In the 100 years since then, the urban population has grown 12 times and it is now around 285 million people constituting 28 per cent of the total population. In the following 20 years (2001-21), the urban population will nearly double itself to reach about 550 million. According to the World Urbanization Prospects (the 1996 Revision), the urban population in the year 2025 will rise to 42.5 per cent which is 566 million (UN Department of Economic and Social Affairs,1996). Urban areas in India, which includes wide ranges of mega cities, cities, towns are not all that lucky in terms of intra & inter city transportation. Transport in this context has been a victim of ignorance, neglect and confusion all these at once. In 2002, 58.8 million vehicles were plying on Indian roads. According to statistics provided by the Ministry of Road Transport & Highways, Government of India, the annual rate of growth of motor vehicle population in India has been about 10 percent during the last decade. The basic problem is not the number of vehicles in the country but their concentration in a few selected cities, particularly in metropolitan cities (million plus). It is alarming to note that 32 percent of these vehicles are plying in metropolitan cities alone, which constitute about 11 percent of the total population (MORTH,2003).

India being a developing country our traffic especially in urban street is very much heterogeneous consisting various kind of vehicles having different operational characteristics. In current scenario no suitable methodology is available for determination of Level of Service (LOS) of heterogeneous traffic conditions. Determining level of service for urban street is very much important as this is the first step of level of service analysis procedure. This affects the planning, design, and operational aspects of transportation projects as well as the allocation of limited financial resources among competing transportation projects. So there is a bare need of determining the level of service criteria in Indian context.

Urban street level of service is primarily a function of travel speed along segments, and is calculated from field data (HCM). Floating car method is the most common technique to acquire speed data. In this method as a driver drives the vehicle a passenger records the elapsed time information at predefined check points. This recording of elapsed time can be done by pen and paper, audio recorder or with a small data recording device. This method is advantageous as require very low skilled technician and very minimal cost equipments. It has a drawback as it is a labor intensive method. Human errors that often result from this labor intensiveness include both recording errors in the field and transcription errors as the data is put into an electronic format (Turner, *et. al.,* 1998). With improvement in portable computers Distance Measuring Instrument

(DMI) came as the solution for floating car method. DMI measures the speed distance using pulses from a sensor attached to the test vehicle's transmission (Quiroga and Bullock, 1998). This method also has some limitations like very complicated wiring is required to install a DMI unit to a vehicle. Frequent calibration and verification factors unrelated to the unit are necessary to store making the data file large and which leads to data storage problem. (Turner, *et. al.*, 1998; Benz and Ogden, (1996). The development of Information Technology and advancement of Global Positioning System (GPS) has largely overcome the data quality and quantity shortcomings of the manual and DMI methods of collecting travel time data .become one of the alternatives to moving car observer method for field data collection. In this method a GPS receiver mounted on a vehicle automatically records location of urban corridor and speed at regular sampling interval. A single technician can perform this task quite easily with accuracy. The field data collected through GPS receiver should refer to a geographical position and additional tool is required for this purpose. Geographical Information System (GIS) comes as the solution for this problem with an advantage of assigning the parameters received through GPS to existing geographical data base. This automated procedure provides convenience, consistency, finer precession and accuracy than the conventional procedure. This automated procedure helps to collect large amount of travel time and speed data.

Level of Service (LOS) concept germinated from the concept of "practical capacity" presented in the 1950 HCM. In the 1965 Highway Capacity Manual (HCM) LOS was stated as "qualitative measure of the effect of numerous factors, which include speed and travel time, traffic interruptions, freedom to maneuver, safety, driving comfort and convenience, and operating cost." In the 1985 HCM the statement of 1965 HCM was clarified by incorporating two significant factors i.e. “Qualitative major of operational factors” and “Perception of motorist and passengers” however “Operation Cost” was dropped. In 1965 and 1985 HCM the LOS was described by the six classes from “A” to “F” defined, based on the combination of travel time and the ratio of traffic flow rate to the capacity, because travel time was recognized as a dominant factor of the service quality. Highway capacity as well as LOS is dynamic, ever-changing and evolving phenomenon. The number of vehicles on the road, the amount of congestion, vehicle performance characteristics and geometric standards has significantly changed the environment in which a driver has to drive (Kittelson, 2000). In course of time the definition of LOS went in an evolution process and what we are following now is the LOS defined in 2000 HCM (TRB, 2000).Highway Capacity Manual (HCM, 2000) defined LOS as “a quality measure describing operational conditions within a traffic stream, generally in terms of such service measures as speed and travel time, freedom to maneuver, traffic interruptions, and comfort and convenience.”. The HCM designates six levels of service, A-F; describe operations from best to worst for each type of facility. Figure 1.1 illustrates Levels of service as a step-function, having a distinct speed value as boundary and each representing a range of operating conditions.

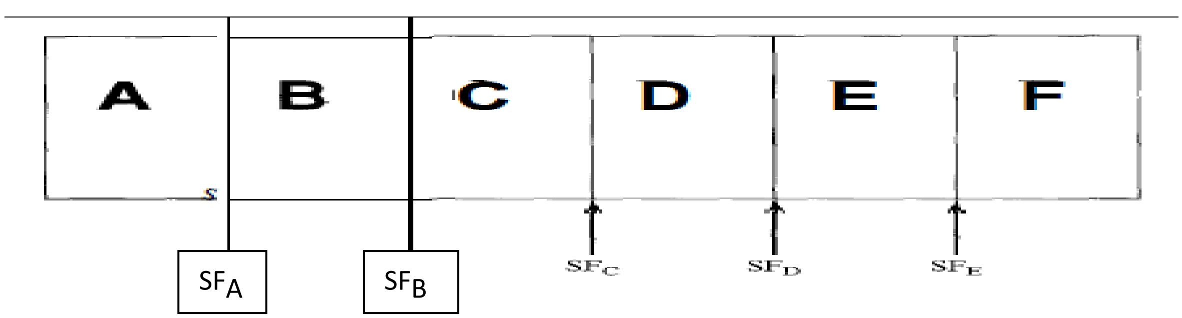


Figure 1.1. Level of Service as a step function

The above figure also shows service flow rate for each type of LOS. This service flow rate is nothing but maximum capacity for a particular type of designated LOS. Service flow rate can be defined for LOS A-E but for LOS F which corresponds to very unstable flow having poor quality of service, Service flow rate cannot be defined (Roess et al. 1994). To derive LOS criteria various studies have been performed. Cameron (1996) and Baumgaertner (1996) tried to justify an extended criteria from "A" to "J" and "A" to "I", respectively. Their common concern however was that longer delay occurs due to increasing congestion. However, criteria representing traffic conditions beyond LOS "F" proposed by adding extra categories appears to be somewhat arbitrary. Heterogeneity of traffic, speed regulations, frequency of intersections, presence of bus stops, on-street parking, roadside commercial activities, pedestrian volumes etc are taken as factor affecting the LOS of urban streets (IRC, 1990).

In simple language defining LOS can be called as a classification problem which can be solved using the various clustering tool available. Prassas et al. (1996) applied the cluster analysis tools to a set of traffic engineering data on which deterministic modeling and regression analysis have been applied before. From this study the authors have concluded that cluster analysis is a powerful exploratory technique and helps in identifying several distinct modalities within the traffic data. Lingra (1995) compared grouping of traffic pattern using the Hierarchical Agglomerative Clustering and the Kohonen Neural Network methods for classifying traffic patterns. Such an approach is useful in using hour-to-hour and day-to-day traffic variations in addition to the monthly traffic-volume variation in classifying highway sections. Researchers have been applying other classification techniques in addition to the clustering techniques, in this regard. Wei et al. (1996) explored the feasibility of vehicle classification in real world situations using an integrated model that they have developed from back propagation Artificial Neural Network (ANN) model and image processing. Yang and Qiao (1998) have developed a classification method for traffic flow states for Chinese highways using neural network approach. The method sampled actual flow data for each possible case, recognized their different characteristics, and then sorted them into various clusters using the neural network pattern recognition techniques.

1.2 Statement of the Problem

Like every coin have two sides the urbanization also showing the consequences by taking toll on the urban infrastructure. Urban streets are not spared from the ill effect of urbanization. As a result of which the operating condition of road available to the commuter decreasing day by day. The urban road networks in recent times are badly suffering from the problems like decreasing speeds, increased congestion, increased travel time, and decreased level of service and increase in accident rates. The ever-increasing demand for urban networks and variety of problems that occur daily has outstripped the ability of traditional traffic management practices. Building new and bigger roads is not the answer to solve the present transportation problems because of presence of hindrance in financial and environmental issues. So available urban transport network need to be utilized more efficiently which can achieved by proper management of traffic flow in urban networks. This requires a huge amount of travel time, travel speed and delay data.

Floating car method was used traditionally for collection of travel speed data. Although this method is very simple but it has some lacuna like accuracy variation from technician to technician and the possibility of missing and inaccurate marking of some check points.

Recent research has demonstrated the feasibility of using Global Positioning System (GPS) and Geographic Information System (GIS) technologies for automating the travel time data collection, reduction, and reporting when using a probe vehicle. The new automated procedures provide consistency, automation, finer levels of resolution, and better accuracy in measuring travel time, delay and speed rather than traditional techniques. As a result, large amounts of reliable travel time, delay and speed data can be collected and processed.

Determining LOS for urban street is very much important as this is the first step of LOS analysis procedure. This affects the planning, design, and operational aspects of transportation projects as well as the allocation of limited financial resources among competing transportation projects. India being a developing country the traffic especially in urban street is very much heterogeneous consisting various kind of vehicles having different operational characteristics. In current scenario no suitable methodology is available for the determination of LOS of heterogeneous traffic conditions. So there is a bare need of determining the LOS criteria in Indian context.

1.3 Objectives and Scope

Based on the above problem statement, the objectives of the study are:

- To classify the urban street segments into various classes using free flow speed data acquired through GPS and data clustering technique.
- To define free flow speed ranges of urban street classes and speed ranges of LOS categories using advanced clustering algorithms.

- To find the most suitable cluster analysis algorithm in defining FFS ranges of urban street classes and speed ranges of LOS categories.

The overall framework of this study is illustrated in Figure 1.2.

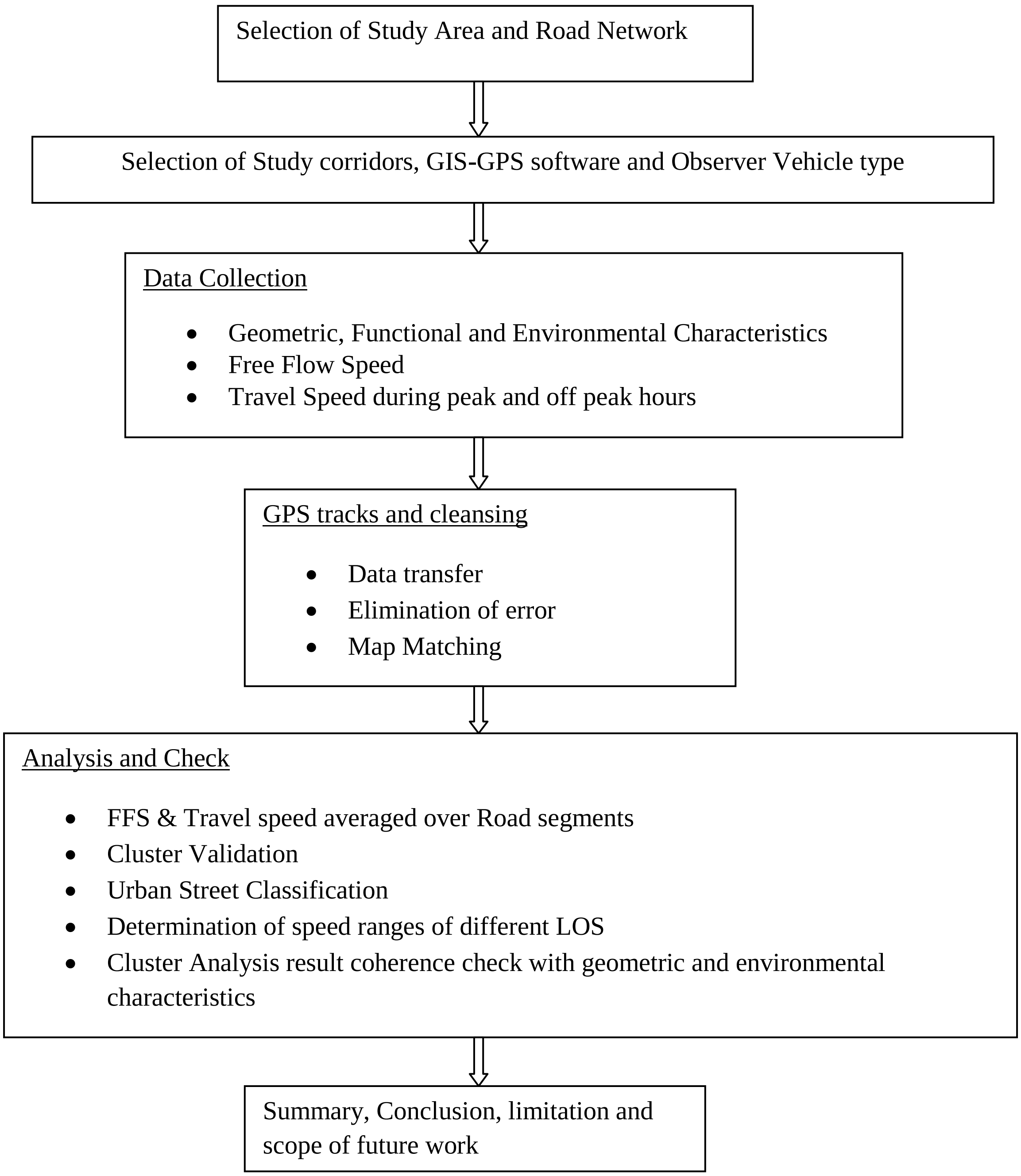


Figure 1.2 Overall framework of the study

1.4 Organization of Report

This report comprises of seven chapters. The first chapter provides an introduction to this research and also describes the objective and scope of this study. Second chapter gives an insight into the two major components of this research work i.e. urban street and level of service. In third chapter a thorough discussion on various literatures related to level of service, use of GIS-GPS in traffic data collection and clustering techniques are presented. Fourth chapter provides idea about the study area of this work and methodology of data collection. The fifth chapter a detailed description on the cluster techniques and cluster validation measure are presented for clear understanding of this research. Result and analysis of the findings are illustrated in sixth chapter. The seventh chapter gives a summary of this work and conclusion of the research. Limitations in the current study and scope for future work are highlighted.

References and Appendix are given at the end of this report.

Chapter 2

Urban Streets and Level of Service Concepts

2.1 Introduction

A foresight has been given in this chapter to the two major component of this report. The term “urban street”, as used in this work, refers to urban arterials and collectors, including those in downtown areas. Urban street quality of service requires quantitative measures to characterize operational conditions within a traffic stream. Level of service is a quality measure describing operational conditions within a traffic stream, generally in terms of such service measure as speed and travel time, freedom to maneuver, traffic interruptions, and comfort and convenience (HCM, 2000).

2.2 Urban Streets

2.2.1 Introduction

In the hierarchy of street transportation facilities, urban streets (including arterial and collectors) are ranked between local streets and multilane suburban and rural highways. The difference is determined principally by street function, control conditions, and the character and intensity of roadside development. Arterial streets are roads that primarily serve longer through trips. However, providing access to abutting commercial and residential land uses is also an important function of arterials. Collector streets provide both land access and traffic circulation within residential,

commercial, and industrial areas. Their access function is more important than that of arterials, and unlike arterials their operation is not always dominated by traffic signals. Downtown streets are signalized facilities that often resemble arterials. They not only move through traffic but also provide access to local businesses for passenger cars, transit buses and trucks. Turning movements at downtown intersections are often greater than 20 percent of total traffic volume because downtown flow typically involves a substantial amount of circulatory traffic. Pedestrian conflicts and lane obstructions created by stopping or standing taxis, buses, trucks and parking vehicles that cause turbulence in traffic flow are typical of downtown streets. Downtown street function may change with the time of the day; some downtown streets are converted to arterial-type operation during peak hour operation. Multilane suburban and rural highways differ from urban streets in the following ways; road side development is not as intense, density of traffic access points is not as intense, density of traffic access points is not as high, as signalized intersections are more than 3.0 km apart.
These conditions result in a smaller number of traffic conflicts, smoother flow, and dissipation of the platoon structure associated with traffic flow on an arterial or collector with traffic signals. The speed of vehicles on urban streets is influenced by three main factors: street environment, interaction among vehicles, and traffic control. As a result, these factors also affect quality of service.

2.2.2 Functional and design categories of urban streets

Estimating Speed, delay and LOS for an urban street or an intersection requires geometric data and demand data. Taking field measurements for use as inputs to an analysis is the most reliable way to generate parameter values. Default values should be considered only when this is not feasible. The classification step is used herein to determine the appropriate design category. The design category depends on the posted speed limit, signal density, driveway/access-point density, and other design features. Four urban street classes are defined in HCM. The classes are designated by number (i.e., I, II, III, and IV) and reflect unique combinations of street function and design. The functional component is separated into two categories: principal arterial and minor arterial. The design component is separated into four categories: high-speed, suburban, intermediate, and urban.

High-speed Design

High-speed design represents an urban street with a very low driveway/access-point density, separate left-turn lanes, and no parking. It may be multilane divided or undivided or a two lane facility with shoulders. Signals are infrequent and spaced at long distances. Roadside development is low density, and the speed limit is typically 75 to 90 km/hr. This design category includes many urban streets in suburban settings.

Suburban Design

Suburban design represents a street with a low driveway/access-point density, separate left turn lanes and no parking. It may be multilane divided or undivided or a two-lane facility with shoulders. Signals are spaced for good progressive movement (up to three signals per kilometer). Roadside development is low to medium density, and speed limits are usually 65 to 75 km/hr.

Intermediate Design

Intermediate design represents an urban street with a moderate driveway/access-point density. It may be a multilane divided, an undivided one-way, or a two-lane facility. It may have some separate or continuous left-turn lanes and some portions where parking is permitted. It has a higher density of roadside development than the typical suburban design and usually has two to six signals per kilometer. Speed limits are typically 50 to 65 km/hr.

Urban Design

Urban design represents an urban street with a high driveway/access-point density. It frequently is an undivided one-way or two-way facility with two or more lanes. Parking is usually permitted. Generally, there are few separate left-turn lanes, and some pedestrian interference is present. It commonly has four to eight signals per kilometer. Roadside development is dense with commercial uses. Speed limits range from 40 to 55 km/hr.

2.2.3 Determining urban street class

In defining LOS, determination of urban street's class of each urban street segment is the first step. This can be based on direct field measurement of the FFS or on an assessment of the subject street's functional and design categories. If the FFS measurements are not available, the street's functional and design categories must be used to identify its class. The functional category is identified first, followed by the design category. In practice, there are sometimes ambiguities in determining the proper categories. The measurement or estimation of the free-flow speed is a great aid in this determination, because each urban street class has a characteristic range of free-flow speeds, HCM (2000).

2.2.4 Segmenting the urban streets

Segment is the basic unit of a facility on which capacity and level of service analysis is performed. This basically a one-directional stretch between two consecutive signalized intersections. At the start of the analysis, the location and length of the urban street to be considered must be defined. All relevant physical, signal, and traffic data should be identified. Consideration should be given to the extent of the urban street; generally at least 1.5 km is necessary in downtown areas and 3.0 km in other areas. Figure2.1 illustrates the segment concept on one- and two-way streets for Indian conditions. In this study, the starting point of segment is identified during inventory survey. During inventory survey the exact location where the vehicles achieved unconstrained speed after leaving the stopped line is identified.

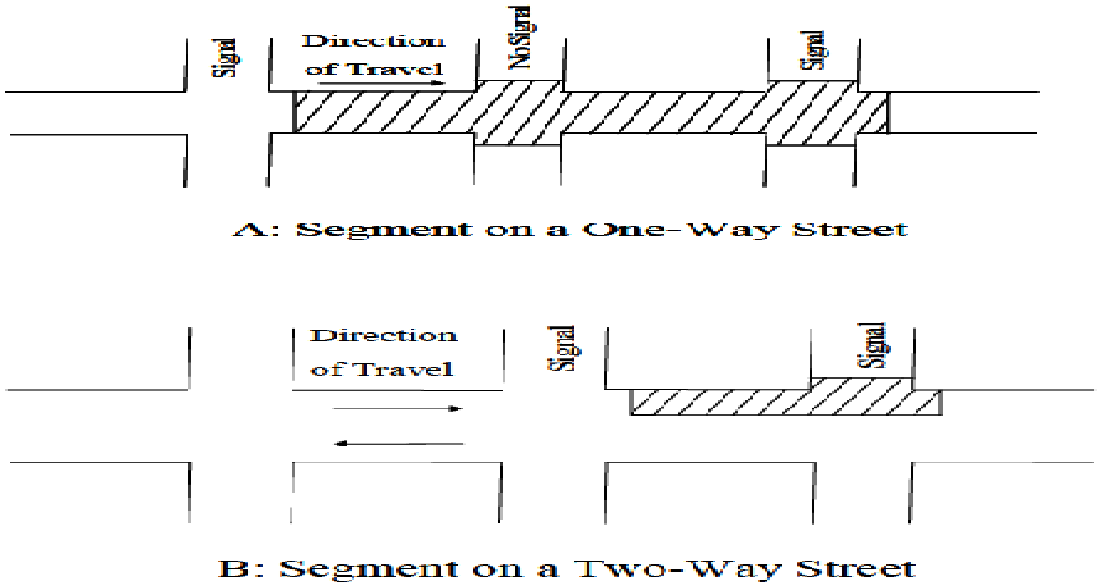


Figure 2.1 Types of Urban Street Segments

2.3 Level of Service

2.3.1 Introduction

Urban Street Level of Service (LOS) is based on average through-vehicle travel speed for the segment or for the entire street under consideration. Travel speed is the basic service measure for urban streets. The average travel speed is computed from the running times on the urban street and the control delay of through movements at signalized intersections. The LOS for urban streets is influenced both by the number of signals per kilometer and by the intersection control delay. Inappropriate signal timing, poor progression, and increasing traffic flow can degrade the LOS substantially. Streets with medium-to-high signal densities (i.e., more than one signal per kilometer) are more susceptible to these factors, and poor LOS might be observed even before significant problems occur. On the other hand, longer urban street segments comprising heavily loaded intersections can provide reasonably good LOS, although an individual signalized intersection might be operating at a lower level. The term through vehicle refers to all vehicles passing directly through a street segment and not turning. Table 2.1 lists urban street LOS criteria based on average travel speed and urban street class according to HCM-2000. It is being noted that if demand volume exceeds capacity at any point on the facility, the average travel speed might not be a good measure of the LOS.

Table 2.1 Urban Street LOS by HCM (HCM, 2000)

Urban Street Class	I	II	III	IV
Range of free-flow speed (FFS)	90 to 70 km/hr	70 to 55 km/hr	55 to 50 km/hr	55 to 40 km/hr
Typical FFS	80km/hr	65km/hr	55km/hr	45km/hr
LOS	Average Travel Speed (km/hr)			
A	>72	>59	>50	>41
B	>56-72	>46-59	>39-50	>32-41
C	>40-56	>33-46	>28-39	>23-32
D	>32-40	>26-33	>22-28	>18-23
E	>26-32	>21-26	>17-22	>14-18
F	≤ 26	≤ 21	≤ 17	≤ 14

LOS "A" describes primarily free-flow operations at average travel speeds, usually about 90 percent of the Free-Flow Speed (FFS) for the given street class. Vehicles are completely unimpeded in their ability to maneuver within the traffic stream. Control delay at signalized intersections is minimal.

LOS "B" describes reasonably unimpeded operations at average travel speeds, usually about 70 percent of the FFS for the street class. The ability to maneuver within the traffic stream is only slightly restricted, and control delays at signalized intersections are not significant.

LOS "C" describes stable operations; however, ability to maneuver and change lanes in midblock locations may be more restricted than at LOS B, and longer queues, adverse signal coordination, or both may contribute to lower average travel speeds of about 50 percent of the FFS for the street class.

LOS "D" borders on a range in which small increases in flow may cause substantial increases in delay and decreases in travel speed. LOS "D" may be due to adverse signal progression, inappropriate signal timing, high volumes, or a combination of these factors Average travel speeds are about 40 percent of FFS.

LOS "E" is characterized by significant delays and average travel speeds of 33 percent or less of the free-flow speed. Such operations are caused by some combination of adverse progression, high signal density, high volumes, extensive delays at critical intersections, and inappropriate signal timing.

LOS "F" is characterized by urban street flow at extremely low speeds, typically one third to one-fourth of the FFS. Intersection congestion is likely at critical signalized locations, with high delays, high volumes, and extensive queuing.

2.3.2 Factors affecting level of service

Many of the procedures provide a formula or simple tabular or graphic presentations for a set of specified standard conditions, which must be adjusted to account for prevailing conditions that do not match. The standard conditions so defined are termed base conditions. Base conditions assume good weather, good pavement conditions, and users familiar with the facility, and no impediments to traffic flow.

Base conditions for uninterrupted-flow facilities include the following:

1. Lane widths of 3.6 m,

2. Clearance of 1.8 m between the edge of the travel lanes and the nearest obstructions or objects at the roadside and in the median,

3. Only passenger cars in the traffic stream (no heavy vehicles),

4. Level terrain,

5. No no-passing zones on two-lane highways, and

6. No impediments to through traffic due to traffic control or turning vehicles. In most facility type, prevailing conditions differ from the base conditions, and level of service must include adjustments. Prevailing conditions are generally categorized as roadway, traffic and control.

a) Roadway conditions

Roadway conditions include geometric and other elements. They can affect a performance measure such as speed, but not the capacity or maximum flow rate of the facility.

Roadway factors include the following:

1. Number of lanes,

2. The type of facility and its development environment,

3. Lane widths,

4. Shoulder widths and lateral clearances,

5. Design speed,

6. Horizontal and vertical alignments, and

7. Availability of exclusive turn lanes at intersections.

b) Traffic conditions

Traffic conditions that influence capacities and service levels include vehicle type and lane or directional distribution.

- **Vehicle type**

The entry of heavy vehicles-that is, vehicles other than passenger cars into the traffic stream affects the number of vehicles that can be served. Trucks, buses, and recreational vehicles are the three groups of heavy vehicles. Heavy vehicles adversely affect traffic in two ways:

1. They are larger than passenger cars and occupy more roadway space.

2. They have poorer operating capabilities than passenger cars.

- **Directional and lane distribution**

In addition to the distribution of vehicle types, two other traffic characteristics affect level of service: directional distribution and lane distribution. Nevertheless, each direction of the facility usually is designed to accommodate the peak flow rate in the peak direction. Typically, morning peak traffic occurs in one direction and evening peak traffic occurs in the opposite direction. Lane distribution also is a factor on multilane facilities. Typically, the shoulder lane carries less traffic than other lanes.

c) Control conditions

For interrupted-flow facilities, the control of the time for movement of specific traffic flows is critical to level of service. The most critical type of control is the traffic signal. The type of control in use, signal phasing, allocation of green time, cycle length, and the relationship with adjacent control measure affect operations.

2.4 Methodology

This methodology provides the framework for the evaluation of urban streets. The field data on travel time and travel speed are available; this framework can be used to determine the street's level of service. Also, the direct measurement of the travel speed along an urban street provides an accurate estimate of LOS without using the computations. Figure 2.6illustrates the basic method for determining LOS on an urban street by HCM (2000).

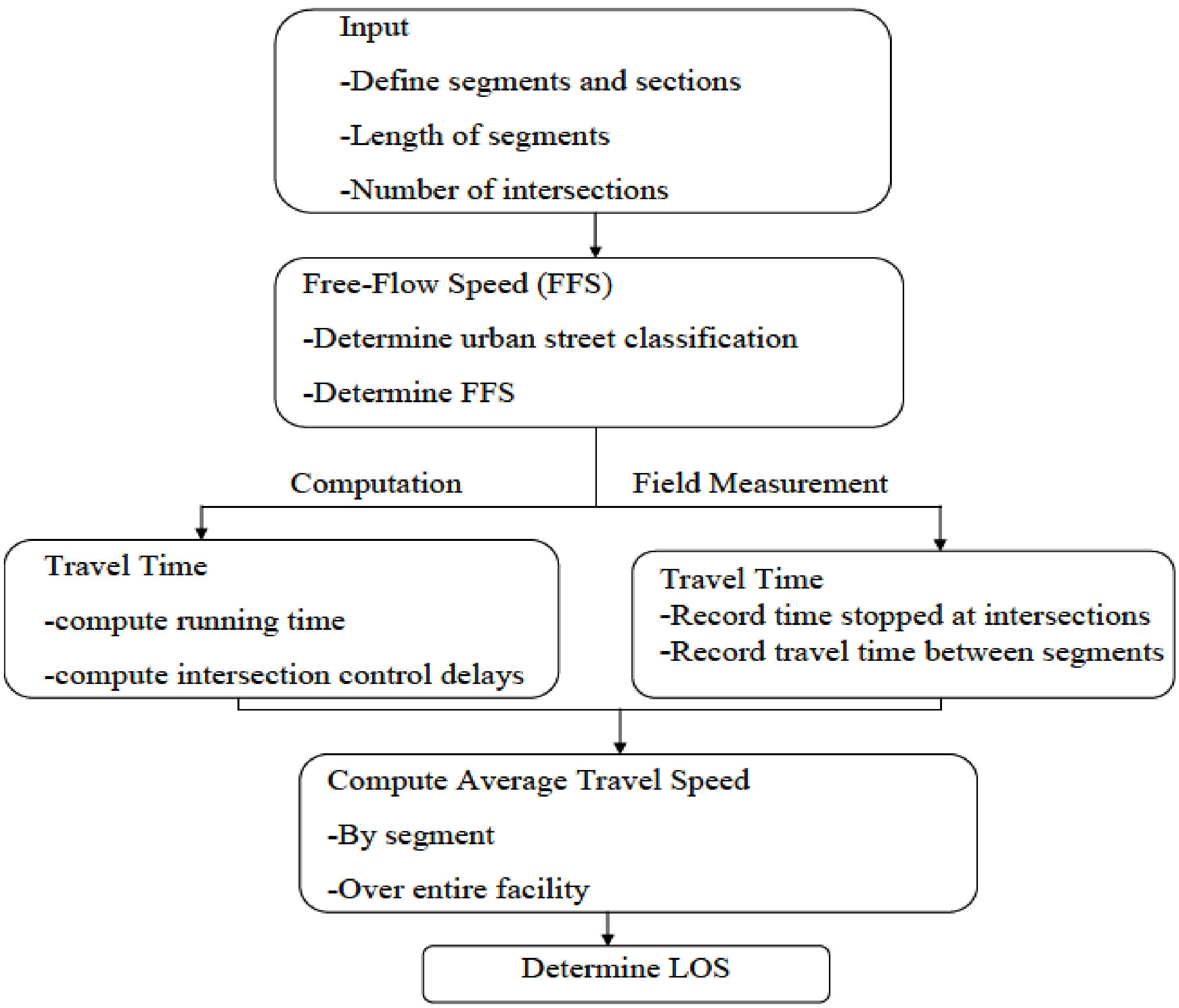


Figure 2.2 Urban Streets Methodology for LOS by HCM-2000

2.5 Summary

This chapter gives a brief understanding on urban streets and Level of Service analysis. Various nomenclature used in this study are described in lucid manner. The LOS analysis procedure given in HCM-2000 is described. Various influencing factors that affect LOS analysis are described. Speed ranges of LOS categories of different urban street classes are also shown. The next chapter describes various literatures that gave a platform for this study.

Chapter 3

Review of Literature

3.1 Introduction

Capacity research started in way back 1920 but there was no consensus on the particular methodology that was adopted for a particular highway capacity. For the first time the level-of-service concept was introduced in the 1965 HCM as a convenient way to describe the general quality of operations on a facility with defined traffic, roadway, and control conditions. Using a letter scale from A to F, a terminology for operational quality was created that has become an important tool in communicating complex issues to decision-makers and the general public. In 1965 HCM the LOS was not defined on quantitative manner. This concept was redefined in relation to several traffic conditions in the HCM of version 1985 (TRB, 1985). The measures of level of service adopted in the HCM of version 1985, which describe the characteristics of traffic conditions under operation, include travel speed, traffic flow rate, and traffic density, for each types of roads. These measures are useful to evaluate and compare the level of service of road sections of the same type. HCM 2000 defines level of service as follows: “Level of service (LOS) is a quality measure describing operational conditions within a traffic stream, generally in terms of such service measures as speed and travel time, freedom to maneuver, traffic interruptions, and comfort and convenience.”The six defined levels of service, A-F, describe operations from best to worst for each type of facility. When originally defined, models did not exist for the prediction of precise quality measures for many types of facilities. This is no longer true. Every facility type has levels of service defined in terms of a specific measure of effectiveness.

Several researchers tried to find out various alternatives and came up with various solutions for defining and determining the LOS. Baumgaertner (1996), Cameron (1996) and Brilon (2000) all provided some insight into the limitations of the current LOS measure. Baumgaertner pointed out that the continuous growth of urban populations, vehicle ownership, average trip length, and number of trips has resulted in a significant increase in traffic volumes. Today motorist became more adapted to urban congestion so the traffic condition which was viewed as intolerable 1960 now considered normal. Kita and Fujiwara (1995) stated LOS not just as a traffic operating condition but tried to find the relationship of LOS with driver's perception. Different type of road section has different type of measures of effectiveness and making it impossible to evaluate and compare the LOS of different road segments with varying characteristics. Spring (1999) raised question mark over LOS being a step function. He found service quality being a continuous and subjective matter so it is not wise to use a distinct boundary or threshold value for determining a particular level of service. Kittelson and Roess (2001) pointed out that the current methodology of determining LOS is not based upon user perception. Clark (2008) raised question mark about the LOS "F" stating it to be very broad. He suggested for a new LOS to be termed as F+ or G. His study specially refers to the type of traffic condition prevailing in New Zealand.

Shao and Sun (2010) proposed a new concept on LOS. The author categorized LOS into two parts: Level of facility supply and Level of traffic operation. Travel speed to free flow speed ratio was considered as evaluation index of traffic operation. Fuzzy set was used by authors to categorize traffic operation into different groups. Flannery et.al (2008) incorporated user perception to estimate LOS of Urban street facilities using a set of explanatory variables that describe the geometry and operational effectiveness. Fang et. al. (2011) determined speed-flow curves of different segments of an interchange by developing a simulation model using VISSIM software. By taking density as classification index the author determined the LOS ranges from the speed-flow curves. Arasan and Vedagiri (2010) through computer simulation studied the effect of a dedicated bus lane on the LOS of heterogeneous traffic condition prevailing in India. The author also estimated the probable modal shift by commuter when a dedicated bus lane is introduced. Cavara et.al. (2011) found traditional algorithms are not very much suitable for analysis of large amount of speed data. So the author developed a state-of-the-art hybrid algorithm for this purpose and classified urban roads based on vehicle track and infrastructural data collected through GPS. Chung (2003) tried to determine the travel pattern along a particular route of Tokyo metropolitan area. Small to large ration (SLR) clustering algorithm was used by the author to cluster historical travel time data for this purpose. It was found from the research day time travel pattern can be classified into three categories i.e weekdays, Saturday and Sunday (including holidays) but night time travel pattern didn't have any such group to classify. Chakroborthy and Kikuchi (2007) utilized Fuzzy set in order to find the uncertainty associated with the LOS categories. Six frameworks were proposed by the authors in order to determine the uncertainty associated under each LOS category.

Dandan et. al. (2007) did not consider traffic flow as the only parameter to access the LOS of various traffic facilities. Not going with traditional research the author analyzed the pedestrian LOS with user perception along with physical facilities and traffic flow operation. In this research the authors have elaborated that primary factors for classification of LOS can be determined by utilizing mass survey data and statistical software SPSS. Shouhua et.al. (2009) found the LOS criteria of walkways proposed by HCM 2000 are not suitable for China. The authors have taken user perception into consideration for classification of LOS at urban rail transit passages and found the limit for LOS standards suitable for China is lower than that suggested by HCM 2000. From this research body size, culture, gender and age found to influence the LOS classification. Fang and Pechuex (2009) studied about the LOS of a signalized intersection taking user perception into account. Unsupervised data clustering technique such as fuzzy c-means clustering was used to get distinct cluster of user perceived delay and service rating. Clustering result was analyzed according to approach membership, delay membership and rating membership. The author found that it is appropriate to differentiate LOS into six categories as described in HCM but proposed a new six levels of service by merging existing LOS A and B and splitting existing LOS F into two categories. Ndoh and Ashford (1994) developed a model to evaluate airport passenger services quality using fuzzy set theory techniques. The authors tried to incorporate user perception in evaluation of service quality instead of just considering traffic parameters for this purpose. Pattnaik and Ramesh Kumar (1996) developed methodology to define level of service of urban roads taking into account users' perceptions.

To get the threshold limit of different LOS we have to first classify different urban street segments into four different category of road according to the free flow speed (FFS). For this classifying clustering technique can be used as a tool. Cluster analysis groups objects based on the information found in the data describing their relationships. The goal in cluster analysis is that the objects in a group will be similar to one another and different from the objects in other groups. After classifying the road segment each segment of a particular class of urban street, the average travel speed is used to find the threshold average travel speed value for determining the LOS. For this clustering large amount of speed data is required.

3.2. Use of GIS and GPS tools in collection of highway inventory data

Floating car method is the most common technique to acquire speed data. In this method as a driver drives the vehicle a passenger records the elapsed time information at predefined check points. This recording of elapsed time can be done by pen and paper, audio recorder or with a small data recording device. This method is advantageous as require very low skilled technician and very minimal cost equipments. It has a drawback as it is a labor intensive method. Human errors that often result from this labor intensiveness include both recording errors in the field and transcription errors as the data is put into an electronic format (Turner, *et. al.*, 1998). With improvement in portable computers Distance Measuring Instrument (DMI) came as the solution for floating car method. DMI measures the speed distance using pulses from a sensor attached to the test vehicle's transmission (Quiroga and Bullock, 1998). This method also has some limitations like very complicated wiring is required to install a DMI unit to a vehicle. Frequent calibration and verification factors unrelated to the unit are necessary to store making the data file large and which leads to data storage problem. Recent research has demonstrated the feasibility of using GPS receiver in recording location as longitude-latitude, travel time and travel speed. However, additional tools are required to provide a linear reference to these point locations. Fortunately, Geographical Information System (GIS) can be used for this purpose, with the added advantage that the resulting parameters can be entered directly into the existing geographic database (Kennedy, 2002).

Taylor et al. (2000) have developed an integrated Global Positioning System and Geographical Information System for the collection of on-road traffic data from probe vehicle. The system was further integrated with the engine management system of a vehicle to provide time-tagged data on GPS position and speed, distance traveled, acceleration, fuel consumption, engine performance, and air pollutant emissions on a second-by-second basis. Owusu et. al. (2006) tried for management of vehicular traffic in urban road by the data acquired through GPS. They found GIS-GPS environment quite efficient in handling large amount of traffic data and also they tried to give idea to a planner to know the speed which is not acceptable to the driver in congested traffic situation using GPS data.

Cesar and Bullock (1998b) described a new methodology for performing travel time studies using Global Positioning System and Geographic Information System technologies. In this study, the authors have documented the data collection, data reduction, and data reporting procedures using GPS and GIS technologies. For data collection the authors have used GPS receivers to automatically collect coordinates of logged points, travel time, and travel speed at regular sampling periods, for example every one second. Global Positioning Systems (GPS) are one of several available technologies which allow individual vehicle trajectories to be recorded and analyzed. McNally et. al. (2002) designed a flexible GPS-based data collection which incorporates GPS, data logging capabilities, two-way wireless communications, and a user interface in an embedded system which eliminates driver interaction. They also found it suitable to have idea about the individual vehicle trajectory data which is required to get the idea of travel demand at network level. Atikom et. al. (2011) proposed a technique to identify road traffic congestion levels from velocity of mobile sensors with high accuracy and consistent with motorists' judgments. Human perceptions were used to rate the traffic congestion levels into three levels: light, heavy, and jam.

After getting the speed data using GIS-GPS tools the next objective is to classify these acquired speed data into various groups. The following section elaborates about the literature of clustering techniques used in this study. The applicability of these algorithms for clustering of speed data was found suitable and relevant to our study from the following works.

3.3 Methods of Cluster Analysis

3.3.1 Introduction

After getting the speed data various clustering method can be applied for grouping the data according to their similarity. Clustering in simple is a process in which groups can be formed from a large data set. Various cluster analysis approaches have been applied on different kind of classification problems. From previous research work successful application of Self Organizing Map (SOM), Affinity Propagation (AP), GA-Fuzzy and Partition Around Medoid (PAM) clustering algorithms in solving clustering problem was established. These four algorithms were used in this study for classification of urban streets and LOS categories. Brief descriptions on some previous research related to these algorithms are given in following section.

3.3.2 Self Organizing Map (SOM) Clustering

Self Organizing Map (SOM) is one of the Artificial Neural Network (ANN) having the inherent capability to learn the pattern of input. The application this particular problem to define the LOS of urban street artificial neural network (ANN) is used for clustering of speed data. Levy et. al (1994) compared the ability of supervised and unsupervised learning method for classification and clustering. Lingra (1995) compared grouping of traffic pattern using the Hierarchical Agglomerative Clustering and the Kohonen Neural Network methods in classifying traffic patterns. It is advantageous to use hierarchical grouping on a small subset of typical traffic patterns to determine the appropriate number of groups and change its parameters to reflect the changing traffic patterns. Such an approach is useful in using hour-to-hour and day-to-day traffic variations in addition to the monthly traffic-volume variation in classifying highway sections. Garni and Abdennour (2008) developed a technique using the ANN neural network to detect and count the vehicles plying on road from the video graph data.

Sanchez et. al. (2002) proposed a back propagation neural network to discriminate zones of high mineral potential in the Rodalquilar gold field, south-east Spain, using remote sensing and mineral exploration data stored in a GIS database. A neural network model with three hidden units was selected by means of the k-fold cross-validation method. The trained network estimated a gold potential map efficiently, indicating that both previously known and unknown potentially mineralized areas can be detected. Yang and Qiao (1998) used neural network to classify traffic flow state. Author applied a self-organizing neural network pattern recognition method to classify highway traffic states into some distinctive cluster centers. Jian-ming (2010) developed a combined ANN and Genetic Algorithm method for the prediction of traffic volume in Sanghai Metropolitan Area. The accuracy of prediction of traffic volume of future traffic improved significantly with this combined algorithm. Cetiner et. al. (2010) developed a back propagation Neural Network traffic flow model for prediction of traffic volume of Istanbul City. The model uses the historical data at major junctions of the city for prediction of future traffic volume. Florio and Mussone (1995) have taken the advantage of application of ANN in classification problem to develop the flow-density relationship of a motorway. The author defined the stability and instability of spacing of vehicle in traffic stream. Murat and Basken (2006) used ANN for determination of non uniform delay which is part of total vehicular delay at signalized intersections. Sharma et. al. (1994) studied and compared the learning ability of both supervised and unsupervised type of learning method for clustering. Self Organizing Map (SOM) is a type of ANN which can learn to detect regularities and correlations in their input and responses in future accordingly.

3.3.3 Affinity Propagation (AP) Clustering

Affinity propagation is a theoretic clustering method recently developed by Frey and Dueck (2007). This algorithm simultaneously considers all of data points as possible exemplar (center point) where each message is sent to reflect the latest interest which is owned by each data point to be able to select another data points as their exemplar. In recent past researchers have used this efficient and accurate algorithm in solving various clustering problems.

Frey and Deuck (2007) used AP algorithm to cluster images of faces and genes in microarray data. The authors found AP to cluster data with much lower error than other methods, and it did the clustering in less than one-hundredth the amount of time. Conroy and Xi (2009) developed a semi-supervised AP algorithm for face-image clustering and functional Magnetic Resonance Imaging (fMRI) volumetric pixel clustering. Xia et.al. (2008) presents two variants of AP for grouping large scale data with a dense similarity matrix. The local approach was Partition Affinity Propagation (PAP) and the global method was landmark affinity propagation (LAP). Refianti et.al. (2012) compared accuracy and effectiveness of AP and K-Means algorithm. The authors have found that AP to be more effective than K-Means by implementing these algorithms on the relationship between two variables i.e Grade Point Average (GPA) and duration of Bachelor-Thesis completion at Gunadarma University. Zhang and Zhuang (2008) presented a modified AP algorithm called voting partition affinity propagation (voting-PAP) which is a method of clustering using evidence accumulation. Yang and Bruzzone (2010) used AP for classifying large amount of remote sensing images data quite accurately. Authors found the algorithm to be very much efficient for clustering of data for which training data is not available. Yang et.al. (2010) used this newly developed AP clustering algorithm in traffic engineering. The authors have proposed a model-based temporal association scheme and novel pre-processing and post-processing operations which together with affinity propagation make a quite successful method for vehicle detection and on road traffic surveillance. Zhang et.al. (2012) proposed an instant traffic clustering algorithm using AP to find points on road having similar traffic pattern. Authors found the algorithm to be suitable in predicting the traffic pattern and for finding the influence of traffic pattern at one point to that at another point. These are some of researches carried out at different locations under different traffic conditions which give a strong back ground for further research carried out in this study in defining LOS criteria in Indian Context.

3.3.4 GA-Fuzzy Clustering

Many researchers have utilized Genetic Algorithm in optimizing the clustering problems. Zhou et. al (2010) developed a Genetic Fuzzy Clustering Algorithm combining FCM clustering and Genetic Algorithm (GA). Fuzzy clustering is a process in which a data point is not assigned to a single cluster rather each data point posses a membership function. This membership function indicates the strength of the data point. Alata et. al. (2008) optimized the FCM clustering algorithm using GA. The researcher used the subtractive clustering algorithm to provide the optimal number of clusters needed by FCM algorithm by optimizing the parameters of the subtractive clustering algorithm by an iterative search approach and then to find an optimal weighting exponent (m) for the FCM algorithm. Wei et. al.(2010) found FCM clustering has inherent problem being very time consuming and having poor clustering result. The authors exploited the search capability of GA by enhancing the global search process. GA is used in traffic engineering field by various researchers to solve traffic engineering problems. Lingras et. al (2004) utilized Genetic Algorithm to estimate the missing traffic count. 50% of permanent traffic counts have missing data. The researchers designed a genetically designed regression model having very high precision. Kidwai et. al (2005) presented a optimized model for bus scheduling. The bus scheduling problem was solved in two levels. First level minimum bus in a route with a guarantee of load feasibility was determined. In second level fleet size in first step was taken as upper bound and fleet size was again minimized using GA.

3.3.5 Partition Around Medoid (PAM) Clustering

Another clustering which is used in this current study is Partition around Medoid (PAM). PAM was presented by Kaufman and Rousseeuw (1990).This algorithm attempts to minimize the total distance between objects within each cluster. The algorithm proceeds through two phases. In the first phase, an initial clustering is obtained by the successive selection of reprentative objects until *k* representative objects have been found. The first representative object is the one for which the sum of the dissimilarities to all objects is as small as possible. This reprentative object is the most centrally located in the set of objects. Subsequently, at each step another object is selected. This object is the one which decreases the objective function as much as possible. In the second phase of the algorithm, it attempts to improve the set of representative objects and therefore also to improve clustering yielded by this set. Swami and Jain (2006) integrated Association Rule Mining and Classification to produce Associative Classification techniques which showed better classification accuracy. The idea to use this model for cluster analysis was introduced by Vinod (1969). In a study, Spath (1985) uses random starting cluster configurations for *k*-medoid algorithm. In *k*-medoid cluster analysis, centers are selected from the data points itself. Otherwise, *k*-medoid is just like *k*-means algorithm. Typical aspects of these algorithms are that they provide a large number of statistics by which a thorough investigation of the clustering results is made possible, particularly by means of validation parameters.

3.4 Summary

A thorough literature review is discussed in this chapter related to LOS, speed data collection methods and the four clustering algorithm used in this study. From literature review it was found that there are many limitations in the current LOS methodology given in HCM-2000. GPS was found to be an efficient and accurate technique for collection of speed and road inventory data. Cluster analysis was found to be suitable for grouping of FFS and travel speed data in order to define FFS speed ranges of urban street classes and travel speed ranges of LOS categories respectively.

Chapter-4 discusses about the Study area and Data collection technique for this study.

Chapter 4

Study Area and Data Collection

4.1 Introduction

Previous chapter described the literatures that gave the foundation for this research. In this chapter, details of study area, data collection and the database preparation are described. The study area for this study is taken as Greater Mumbai and Kolkata City. The required GIS layers of the study area for present study obtained from a secondary source. The following section describes the details of road networks and route systems on which the probe vehicle fitted with a Geo-XT GPS receiver was made to run several times. Type and timing of data collection, data smoothing and data compilation are also discussed in detail

4.2 Study Corridors and Data Collection

4.2.1 Base map preparation

The GIS layers of the study area for present analysis are obtained from a secondary source. A detailed roadway inventory survey was also carried out for preparing the digitized GIS base map of the road network.

4.2.2 Study corridors

Five important road corridors of the city of Mumbai in Maharastra state, India are taken up for the present study. Grater Mumbai (GM) is an Island city with a linear pattern of transport network having predominant North-South commuter movements. Passengers move towards south for work trip in the morning hours and return back towards the north in the evening hours. Hence, four north-south corridors and one east-west corridor have been chosen for this study. Major roads like Eastern express highway extending up to south (Corridor-1), LBS Road extending up to south via Ambedkar road (Corridor-2), Western express highway extending up to marine drive (Corridor-3), SV road extending up to south via Veer Savarkar road (Corridor-4) and Versova- Andheri- Ghatkopar- Vashi (VAGV) (Corridor-5) are included. These five corridors overlapped on the GIS base map of Greater Mumbai are shown in Figure 5.1. These five road corridors as a whole cover 100 street segments with 101 signalized intersections. These corridors were selected from such a large road network, as these have varying road geometric characteristics. Total length of road included in this study is approximately 140 kilometers. Roadway width varies significantly from location to location and in this study it includes two lane undivided roads to eight lane divided roads. Traffic movements are two-way on almost all segments except very few on which traffic movement is restricted to a single direction only. Direction-wise roadway capacity expressed in terms of passenger car unit per hour (PCU/hr) on these road corridors varies from 1500 to 3000. Similarly, free flow speeds (FFS) have a large variation on the road corridors i.e. on some segments FFS is 90 km/hr (Kilometer per hour) and on some segments FFS is mere 25 km/hr. The travel speed also varies widely, which ranges from 5km/hr on some road corridors with highly congested traffic during peak hours to 75 km/hr on corridors during off-peak hours.

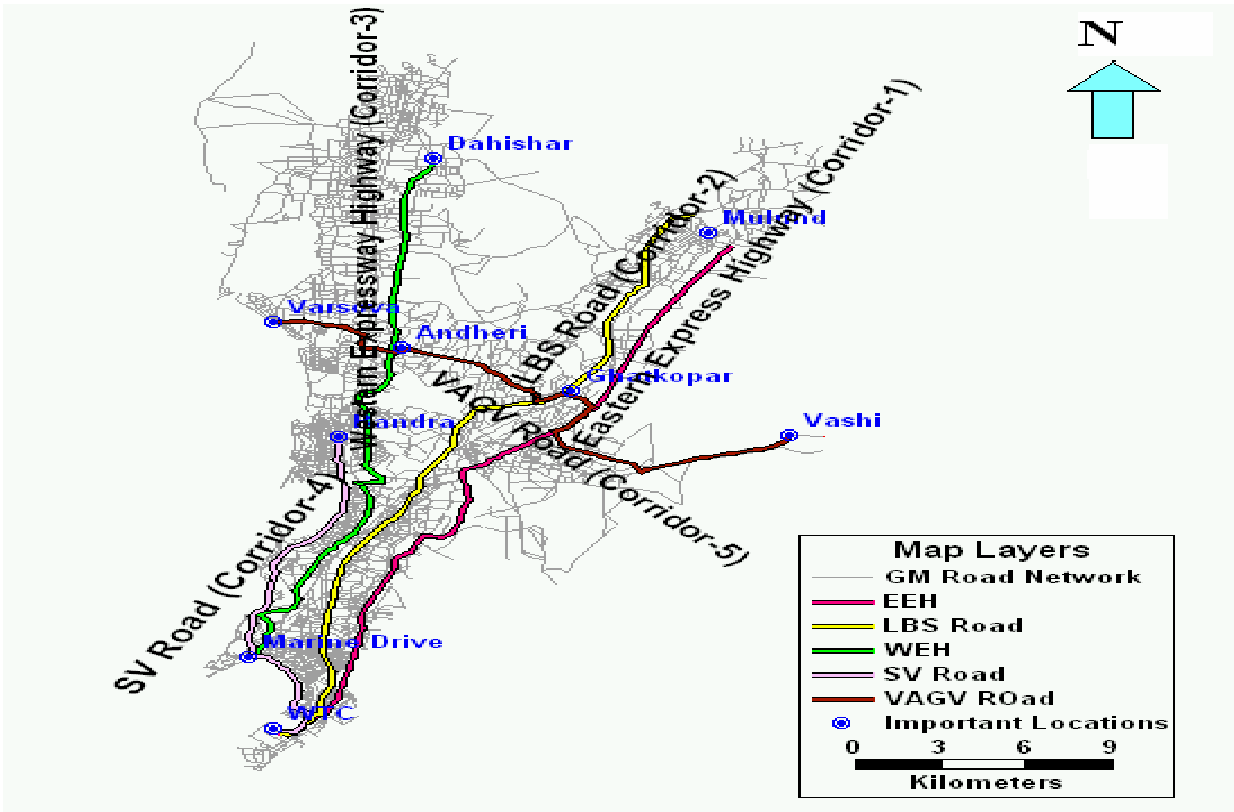


Figure 5.1 Map showing selected corridors of greater Mumbai

4.2.3 Data collection

From extensive literature survey it was found that midsized vehicle fitted with GPS is most suitable for data collection in this study. Midsized vehicle reflect the average of the wide range in vehicle sizes for the heterogeneous traffic flow condition on urban streets in Indian context. It was convenient to employ mid-sized vehicle for this kind of study in which large data samples were collected. Time and resources were the major constraints in the use of different vehicle types in this study. Hence, probe vehicles used in this research work were mid-sized vehicles. These vehicles were fitted with Trimble Geo-XT GPS receiver, capable of logging speed data continuously (at time intervals of one second). The GPS data provides both spatial and time/distance based data from which various traffic parameters were derived, including travel time, stopped time, travel speeds (instantaneous and average). In order to get unbiased data sets, three mid-sized vehicles and the help of three drivers was taken on different days of the survey in this study.

Basically three types of data sets were collected such as roadway inventory details, free flow speed and travel speed during peak and off peak hours. A GPS receiver logs information in the form of features and attributes. A feature is a physical object or an event in the real world for which we want to collect position and descriptive information. A feature is of point, line or area type. We can define a set of attributes for each feature type. An attribute is a piece of descriptive information about the feature. Attributes are of menu, numeric, text, date and time types in data dictionary. Segment number, number of lanes on roadways, median type, parking conditions, pedestrian activity, road side development, access density, commercial activity and speed limits etc. were collected during inventory survey.

The second type of survey conducted was to find the free flow speeds on all these corridors. Before going for the free flow speed data collection, we need to know the time period during which traffic volume would be less than or equal to 200 vehicles per lane per hour. Detailed 24 hour traffic volume count survey was conducted in the month of April; 2005.The traffic volume data were collected on 45 counting stations on seven screen lines covering the whole of Greater Mumbai region. From this survey data, traffic volume per lane per hour was calculated for those roads covered on these five corridors. From this data, it was found that free flow traffic condition (less than 200veh/ln/hr) is approaching at 12 mid-night and all road sections were having free flow traffic conditions from 1 AM to 5 AM. Hence, free flow speed on all these corridors were collected using GPS receiver fitted in probe vehicles during these hours. The third type of data collected was congested travel speed. Congested travel speed survey was conducted during both peak and off-peak hours on both directions of travel on all the corridors. 10-12 travel runs were made on all these five corridors consisting of 100 street segments.

The probability distribution of collected FFS data is shown in Fig.3. From this figure it is observed that the probe vehicle had maintained free flow speed between 40 km/hr to 65 km/hr on significant number of observed street segments and few segments had maintained free flow speed below 40km/hr or above 65 km/hr.

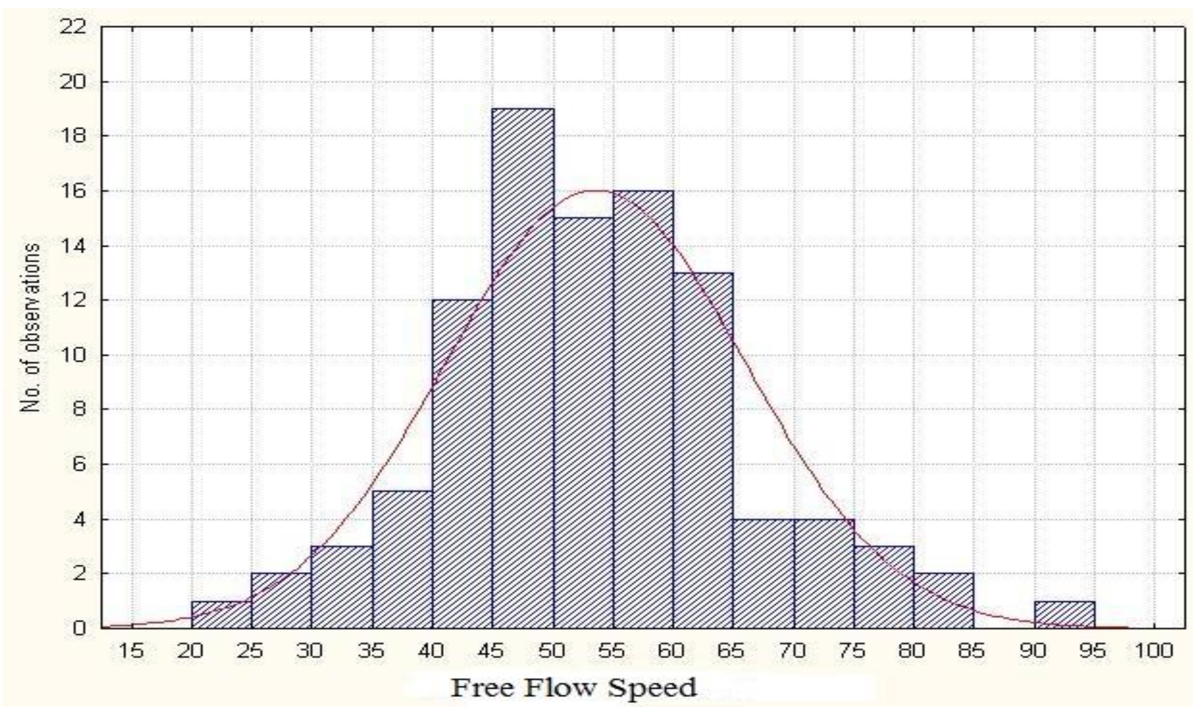


Figure 5.2 Probability distribution plot of collected FFS data using GPS receiver

After the field survey, collected data were transferred to the office computer by using Pathfinder office version 3.00. Collected data files were in the *.ssf extension format. Differential correction was applied on all these collected rover files by merging with base station files. Base station files were in *.dat, or RINEX (Receiver Independent EXchange) format and Pathfinder office automatically converts *.dat and RINEX files into *.ssf extension format during differential correction. Latitude/Longitude as the coordinate system and WGS 1984 as datum were used as the default settings in the end product software built in GIS environment called TransCAD. The accuracy of field data was significantly improved through differential correction process.

Depending on the database used, the differentially corrected data files were exported in a format those were used in the end product software TransCAD. The following steps were followed while exporting a file

1. While clicking the export tool, the most recently used data files were selected by default as input files.

2. Looking at output folder, this folder was where all export files were created.

3. Data files were transferred into ESRI shape file format.

4. The coordinate system latitude/ longitude were used for the exported data.

5. Starting the export process, and

6. Closing the export process.

After files were exported into TransCAD, the SHP file (exported data in shape file format), SHX file (index files), and DBF file (attribute data associated with the shape files) were found in the output folder.

The process and idea behind applying differential correction is explained in more detail in the section below. The receiver at base station measures the timing errors and then provides correction information to the other receivers that are roving around. That way virtually all errors can be eliminated from the system, even the pesky selective availability error that the USA

Department of Defense puts in on purpose. The idea is simple. Put the reference receiver on a point that has been very accurately surveyed. This reference station receives the same GPS signals as the roving receiver but instead of working like a normal GPS receiver it attacks the equations backwards. Instead of using timing signals to calculate its position, it uses its known position to calculate timing. It figures out what the travel time of the GPS signals should be, and compares it with what they actually are. The difference is an "error correction" factor. The receiver then transmits this error information to the roving receiver so that it can use it to correct its measurements. In this study, however, the differential correction is applied offline, as there is no real time communication between the base station and the rover.

4.3 Aggregating Speed on each Segment

Many segmentation schemes are possible in a highway network. We can define a detailed network composed of relatively short segments, say 0.15 km in length, or a “coarse" network composed of relatively long segments, say 3 km in length. Shorter segment lengths imply a larger number of segments but at the same time, segment speeds, which are closer to the original GPS, point speeds. Conversely, longer segments imply a fewer number of segments but, at the same time, segment speeds which may be farther from the original GPS point speeds. One way to apply required segmentation technique is to borrow concepts from HCM 2000. A segment is the directional length of a road stretch from downstream end of one signalized intersection to next signalized intersection.

4.4 Summary

This chapter provided the details of the study area, data collection and database preparation. The details of corridors on which GPS data was collected were discussed. Pathfinder was used to prepare data dictionary and the inventory details were collected using the prepared data dictionary. It was also discussed how the timing of free-flow speed data collection was fixed based upon the traffic volume data.

The next chapter gives idea about the cluster analysis algorithms used in this study and also about the various cluster validation parameters used in the research work in order to determine the optimal number of cluster and to select the best clustering algorithm.

Chapter 5

Cluster Analysis

5.1 Artificial Neural Network (ANN)

5.1.1 Introduction

Artificial Neural Networks are mathematical formulations that mimic nervous system operations of brain. This is a soft computing technique which successfully applied in a various application. ANN is used to learn pattern and relationship in data. The ANN mimics the human ability to adapt to changing circumstances and the current environment. They learn from the events that have happened from past apply this learning to future environment. ANN consists of many nodes i.e. processing unit analogous to neuron in the brain. Each node has a node function and also some local parameters. Modification of local parameter changes the node function. Neural network may be of single or multiple layers. Single layer consists of input neurons and output neurons. Multi layer artificial neural network consists of input layer, output layer and hidden layer.

5.1.2Types of Artificial Neural Network

A) Feed forward neural network

A feed forward neural network is an artificial neural network where connections between the units do not form a directed cycle. This is different from recurrent neural networks. The feed forward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.

B) Radial basis function network

Radial basis functions are powerful techniques for interpolation in multidimensional space. A RBF is a function which has built into a distance criterion with respect to a center. RBF networks have two layers of processing: In the first, input is mapped onto each RBF in the 'hidden' layer. In regression problems the output layer is then a linear combination of hidden layer values representing mean predicted output. The interpretation of this output layer value is the same as a regression model in statistics. In classification problems the output layer is typically a sigmoid function of a linear combination of hidden layer values.

C) Kohonen self-organizing network

A self-organizing map (SOM) or self-organizing feature map (SOFM) is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space. This makes SOMs useful for visualizing low-dimensional views of high-dimensional data, akin to multidimensional scaling. The model was first described as an artificial neural network by the Finnish professor Teuvo Kohonen, and is sometimes called a Kohonen map. Like most artificial neural networks, SOMs operate in two modes: training and mapping. Training builds the map using input examples. It is a competitive process, also called vector quantization. Mapping automatically classifies a new input vector. A self-organizing map consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a regular spacing in a hexagonal or rectangular grid. The self-organizing map describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to first find the node with the closest weight vector to the vector taken from data space. Once the closest node is located it is assigned the values from the vector taken from the data space.

D) Recurrent neural network

A recurrent neural network (RNN) is a class of neural network where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior. Unlike feed forward neural networks, RNNs can use their internal memory to process arbitrary sequences of inputs. This makes them applicable to tasks such as handwriting recognition, where they have achieved the best known results.

5.1.3 SOM Algorithm

Among various types of ANN algorithms, in this study Self Organizing Map is used for clustering of speed data because of its inherent capability to learn the pattern of input. SOM is trained iteratively being inspired by neural networks in the brain. Self Organization Map (SOM) uses a competition and cooperation mechanism to achieve unsupervised learning. In SOM, a set of nodes is arranged in a geometric pattern which is typically a 2-dimensional lattice. This arrangement of neuron may be grid, hexagonal or random topology. In this research Hexagonal topology is used. Each node is associated with a weight vector with the same dimension as the input space. The purpose of the SOM is to find a good mapping. During training, each node is presented to the map so also the input data associated with it. An input weight vector of same dimension as that of input data dimension was given to the ANN. The clustering using SOM algorithm was done in two steps.

STEP: 1

The input data is compared with all the input weight vectors $m_i(t)$ and the *Best Matching Unit* (*BMU*) on the map is identified. The *BMU* is the node having the lowest Euclidean distance with respect to the input pattern $x(t)$. The final topological organization of the map is heavily influenced by this distance. *BMU* $m_c(t)$ is identified by:

For all $i, \| x(t) - m_c(t) \| \leq \| x(t) - m_i(t)$ (1)

STEP: 2

Weight vectors of BMU are updated as

$$m_i(t+1) = m_i(t) + \alpha h_{b(x)i}(x(t) - m_i(t)) \quad (2)$$

Here $h_{b(x)}$ is the neighborhood function, which is

$$h_{b(x)} = \alpha(t) e^{\left(-\frac{\left\| r_i - r_{b(x)} \right\|^2}{2\sigma^2(t)}\right)} \tag{3}$$

Where $0 < \alpha(t) < 1$ is the learning rate factor which decreases with each iteration. r_i And $r_{b(x)}$ are the locations of neuron in the input lattice. $\alpha(t)$ defines the width of the neighborhood function. The above two steps were repeated iteratively till the pattern in input was processed.

5.2 Affinity Propagation (AP)

5.2.1 Introduction

Affinity propagation is a low error, high speed, flexible, and remarkably simple clustering algorithm that may be used in forming teams of participants for business simulations and experiential exercises, and in organizing participants' preferences for the parameters of simulations. The four-equation algorithm is easy to encode into a computer program. Affinity propagation is a graph theoretic clustering method recently developed by Frey and Dueck (2007), who have tested it against *k*-centers clustering, an iterative partitioning method similar to the popular *k*-means procedure that is available on SPSS 15, differing in that *k*-means clusters items around a computed central values whereas *k*-centers clusters them around exemplars, each one being the most central item of its cluster. When applied to a large database of human faces and a large database of mouse DNA segments, Frey and Dueck found that affinity propagation gave rise to smaller errors and arrived at its solution at least two orders of magnitude faster, an important consideration because clustering data is inherently a computationally intensive problem. Moreover, unlike *k*-centers and *k*-means, affinity propagation is more flexible in two ways. First, it does not require the user to specify the number of clusters in advance. Rather, the user selects initial "self-similarity" values from a set derived from the data itself, such that lower self-similarity values give rise to a smaller number of clusters.

Mézard (2007) points out that affinity propagation is known in computer science as a message-passing algorithm, and suggests that the algorithm can be understood by taking an anthropomorphic viewpoint. Thus, imagine that each item being clustered sends messages to all other items informing its targets of each target's relative attractiveness to the sender. Each target then responds to all senders with a reply informing each sender of its availability to associate with the sender, given the attractiveness messages that it has received from all other senders. Senders absorb the information, and reply to the targets with messages informing each target of the target's revised relative attractiveness to the sender, given the availability messages it has received from all targets. The message-passing procedure proceeds until a consensus is reached on the best associate for each item, considering relative attractiveness and availability. The best associate for each item is that item's exemplar, and all items sharing the same exemplar are in the same cluster. Essentially, the algorithm simulates conversation in a gathering of people, where each in conversation with all others seeks to identify his or her best representative for some function.

5.2.2Affinity Propagation Algorithm

Figure 5.1 shows the steps those are performed during AP clustering.

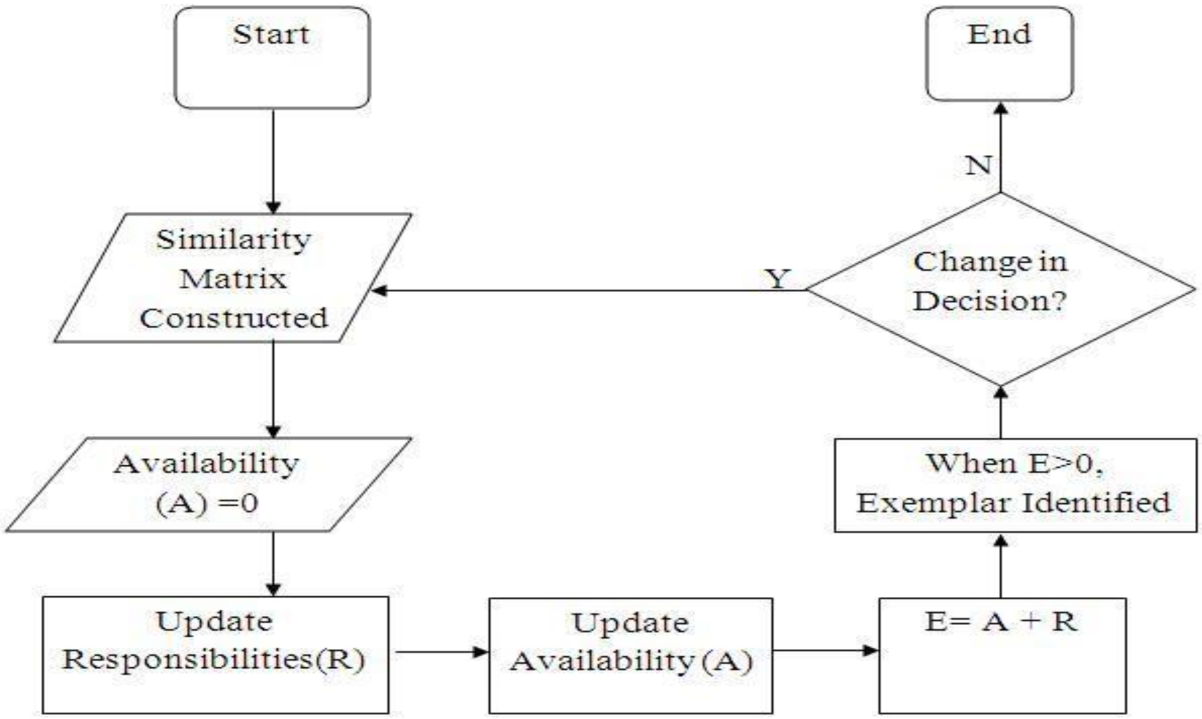


Figure 5.1 Flowchart of AP Clustering

Procedurally, the algorithm operates on three matrices: a similarity (s) matrix, a responsibility (r) matrix, and availability (a) matrix. Results are contained in a criterion (c) matrix. These matrices are iteratively updated by four equations, where *i* and *k* refer, respectively, to the rows and columns of the associated matrix.

The algorithm is processed in the following seven steps to find out the exemplar which is the cluster centre.

Steps:

1. Input similarity matrix *s(i,k)*: the similarity of point *i* to point *k*.

2. Initialize the availabilities $a(i, k)$ to zero: $a(i, k)=0$.

3. Updating all responsibilities $r\ (i,k)$:

$$r(i,k) \leftarrow s(i,k) - \max_{k' \neq k}\{a(i,k') + s(i,k')\} \tag{4}$$

4. Updating all availabilities *a (i,k):*

$$a(i,k) \leftarrow \min\{0, r(k,k) + \sum_{i': i' \notin \{i,k\}} \max\{0, r(i',k)\}\} \text{ for } k \neq i \tag{5}$$

5. Availabilities and responsibilities matrix were added to monitor the exemplar decisions. For a particular data point *i ;* $a(i,k) + r(i,k) > 0$ for identification exemplars.

6. If decisions made in step 3 did not change for a certain times of iteration or a fixed number of iteration reaches, go to step 5. Otherwise, go to step 1.

7. Assign other data points to the exemplars using the nearest assign rule that is to assign each data point to an exemplar which it is most similar to.

5.3 GA-Fuzzy Algorithms (GA)

5.3.1 Introduction

Genetic Algorithms are search algorithms that are based on concepts of natural selection and natural genetics. Genetic algorithm was developed to simulate some of the processes observed in natural evolution, a process that operates on chromosomes (organic devices for encoding the structure of living being). The genetic algorithm differs from other search methods in that it searches among a population of points, and works with a coding of parameter set, rather than the parameter values themselves. It also uses objective function information without any gradient information. The transition scheme of the genetic algorithm is probabilistic, whereas traditional methods use gradient information because of these features of genetic algorithm; they are used as general purpose optimization algorithm. They also provide means to search irregular space and hence are applied to a variety of function optimization, parameter estimation and machine learning applications.

The evolutionary process of a GA is a highly simplified and stylized simulation of the biological version. It starts from a population of individuals randomly generated according to some probability distribution, usually uniform and updates this population in steps called generations. Each generation, multiple individuals are randomly selected from the current population based upon some application of fitness, bred using crossover and modified through mutation to form a new population.

• **Crossover** – exchange of genetic material (substrings) denoting rules, structural components, features of a machine learning, search, or optimization problem

• **Selection** – the application of the fitness criterion to choose which individuals from a population will go on to reproduce

• **Replication** – the propagation of individuals from one generation to the next

• **Mutation** – the modification of chromosomes for single individuals

5.3.2 GA Fuzzy Clustering Algorithm

- **FCM clustering**

Step 1. Set Algorithm Parameters: c - the number of clusters; m - exponential weight; ε- Stop setting algorithm,

Step 2. Randomly generate a fuzzy partition matrix F satisfying the following conditions

$$F = [\mu_{ki}], \mu_{ki} \in [0,1], k = \overline{1, M}, i = \overline{1, c} \tag{6}$$

$$\sum_{i=\overline{1,C}} \mu_{ki} = 1, k = \overline{1, M} \tag{7}$$

$$0 < \sum_{k=\overline{1,M}} \mu_{ki} < N, i = \overline{1, c} \tag{8}$$

Here μ_{ki} - membership function

Step 3. Calculate the centers of clusters: $V_i = \dfrac{\sum_{k=\overline{1,N}} (\mu_{ki})^m \cdot X_k}{\sum_{k=\overline{1,N}} (\mu_{ki})^m}, i = \overline{1, c}$ (9)

Step 4. Calculate the distance between the objects of the X and the centers of clusters:

$$D_{ki} = \sqrt{\left\| X_k - V_i \right\|^2}, k = \overline{1, M}, i = \overline{1, c} \tag{10}$$

Here X is the observation matrix

Step 5. Calculate the elements of a fuzzy partition $(i = \overline{1, c}, k = \overline{1, M})$:

$$\text{If } D_{ki} > 0: \quad \mu_{ki} = \frac{1}{\left(D_{ik}{}^2 \cdot \sum_{j=\overline{1,c}} \frac{1}{D_{jk}{}^2}\right)^{1/(m-1)}} \tag{11}$$

$$\text{If } D_{ki} = 0: \quad \mu_{kj} = \begin{cases} 1, j = i \\ 0, j \neq i, j = \overline{1, c} \end{cases} \tag{12}$$

Step 6. Check the condition $\left\| F - F^* \right\|^2 < \varepsilon$ Where F^* is the matrix of fuzzy partition on the previous iteration of the algorithm. If "yes", then go to step 7, otherwise - to Step 3.

Step 7. End.

➢ Genetic algorithm

The quality of cluster result is determined by the sum of distances from objects to the centers of clusters with the corresponding membership values: $J = \sum_{k=1}^{m} \sum_{i=1}^{c} (\mu_{ki})^m d(v_k, x_i)$ where $d(v_i, x_j)$ is the Euclidean distances between the object $x_j = (x_{j1}, x_{j2}, \ldots, x_{jn}) \frac{\pi}{3}$ and the center of cluster $v_i = (v_{k1}, v_{k2}, \ldots, v_{kn}), m \in (1, \infty)$ is the exponential weight determining the fuzziness of clusters.

The local minimum obtained with the fuzzy *c*-means algorithm often differs from the global minimum. Due to large volume of calculation realizing the search of global minimum of function *J* is difficult. GA which uses the survival of fittest gives good results for optimization problem. GA doesn't guarantee if the global solution will be ever found but they are efficient in finding a "Sufficiently good" solution within a "sufficient short" time.

The parameters used for the algorithm are elaborated in Table.1.

Table 5.1: Parameters of GA-Fuzzy Algorithm

Algorithm Type	*Parameter*	*Value*
Fuzzy C-Means	Stop Parameter	0.1
Fuzzy C-Means	Exponential Weight	2.0
Genetic Algorithm	Number of Individual	10
Genetic Algorithm	Elite	2
Genetic Algorithm	Generations	200
Genetic Algorithm	Frequency of Mutation	0.5
Genetic Algorithm	Crossover	1.0

5.4 Partition around Medoid (PAM)

5.4.1 Introduction

In k-medoid algorithm unlike in k-mean clustering where the mean of the items in each cluster is calculated, a representative item, or medoid, is chosen for each cluster in every iteration. Medoids for each cluster are calculated by finding object *i* within the cluster that minimizes

$$\sum_{j \in c_i} d(i, j) \tag{13}$$

Where C_i is the cluster containing object *i* and *d(i,j)* is the distance between objects *i* and *j*.

Partitioning about medoids (PAM) also clusters objects about *k* medoids, where *k* is specified in advance. However, the algorithm takes the form of a steepest ascent hill climber, using a simple swap neighborhood operation. In each iteration medoid object *i* and non-medoid object *j* are selected that produce the best clustering when their roles are switched. The objective function used is the sum of the distances from each object to the closest medoid. As this search phase of the algorithm is slow, the initial set of medoids is constructed in a greedy build phase. Starting with an empty set of medoids, objects are added one at a time until *k* medoids have been selected. At each step, the new medoid is selected so as to minimize the objective function.

5.4.2 Partition around Medoid (PAM) Algorithm

The algorithm is intended to find a sequence of objects called medoid that are centrally located in clusters. Objects that are tentatively defined as medoids are placed into a set S of selected objects. If O is the set of objects that the set U = O − S is the set of unselected objects. The goal of the algorithm is to minimize the average dissimilarity of objects to their closest selected object. Equivalently, we can minimize the sum of the dissimilarities between object and their closest selected object.

The algorithm has two phases:

(i) In the first phase, BUILD, a collection of k objects is selected for an initial set S.

(ii) In the second phase, SWAP, one tries to improve the quality of the clustering by exchanging selected objects with unselected objects.

The goal of the algorithm is to minimize the average dissimilarity of objects to their closest selected object. Equivalently, we can minimize the sum of the dissimilarities between object and their closest selected object.

For each object p we maintain two numbers:

• D_p, the dissimilarity between p and the closest object in S, and

• E_p, the dissimilarity between p and the second closest object in S.

- **The BUILD phase entails the following steps:**

1. Initialize S by adding to it an object for which the sum of the distances to all other objects is minimal.

2. Consider an object $i \in U$ as a candidate for inclusion into the set of selected objects.

3. For an object $j \in U - \{i\}$ compute D_j, the dissimilarity between j and the closest object in S.

4. If $D_j > d(i, j)$ object j will contribute to the decision to select object i

5. Compute the total gain obtained by adding i to S as $g_i = \sum_{j \in U} C_{ij}$

6. Choose that object that maximizes g_i; let $S = S \cup \{i\}$ and $U = U - \{i\}$

The second phase, SWAP, attempts to improve the set of selected objects and, therefore, to improve the quality of the clustering. This is done by considering all pairs $(i,h) \in S \times U$ and consists of computing the effect T_{ih} on the sum of dissimilarities between objects and the closest selected object caused by swapping i and h, that is, by transferring i from S to U and transferring h to from U to S.

5.5 Cluster Validation measures

Quality of clustering result obtained from a clustering algorithm can be checked by various cluster validation measures. These Validation parameters been mainly used to evaluate and compare whole partitions, resulting from different algorithms or resulting from the same algorithms under different parameters. Most common application of cluster validation is to determine the optimal number of cluster for a particular data set (Bensaid et al., 1996). Different validity measures have been proposed in the literature, none of them is perfect by oneself, and therefore several parameters are used in study.

A) C-index (CI)

C-index is an index which can be used to get optimal number of cluster from a particular set of data. C-Index (CI) is as follows (Hubert and Schultz 1976):

$$CI = \frac{D - D_{\min}}{D_{\max} - D_{\min}} \tag{14}$$

Here D, D_{min}, D_{max} are calculated as follows. Let p be the number of pairs can be formed taking data points belonging to a single cluster. D is the sum of the distances between all these pairs of data points in a particular cluster. P is the number of possible pairs can be formed from the data set. These P numbers of pairs can be ordered according to their distances and p pairs with largest and smallest distances between samples can be selected. D_{min} is the sum of the smallest distance of p and D_{max} is the sum of the largest distance of p. When pairs of samples with small distances are in same cluster the index will be small so it corresponds to optimal number of cluster.

B) Weighted inter-intra index (WI)

This index tries to find the optimal number of cluster with high overall quality of cluster $\phi^{(Q)}$ for a small number of clusters c. This index can be utilized to get right number of cluster for a particular data set. The objective here is to maximize intra-cluster similarity and minimize inter-cluster similarity.

The inter-cluster and intra-cluster similarity are given bellow (Strehl and Ghosh,2002):

$$\mathrm{inter}(X,\lambda,i,j) = \frac{1}{n_i \times n_j} \sum_{\lambda_a=i,\lambda_b=j} S(X_a, X_b) \tag{15}$$

And

$$\mathrm{intra}(X,\lambda,j) = \frac{2}{(n_i - 1) \times n_i} \sum_{\lambda_a=\lambda_b=i,b>a} S(X_a, X_b) \tag{16}$$

S is the similarity matrix. The inter cluster similarity is undefined for singleton cluster. The quality measure $\phi^{(Q)}$ is found out by the ratio of weighted average inter-cluster to weighted average intra-cluster similarity.

$$\phi^{(Q)}(X,\lambda) = 1 - \frac{\sum_{i=1}^{c} \frac{n_i}{n-n_i} \sum_{j\in\{1,\dots i-1,i+1\dots c\}} n_j \times \mathrm{inter}(X,\lambda,i,j)}{\sum_{i=1}^{n} n_i \times \mathrm{intra}(X,\lambda,i)} \tag{17}$$

Here *i and j* are cluster indices. *n is* the number of data points in the input matrix. *X* is the input matrix. X_a and X_b are two vertices. λ is a vector containing the output. $\phi^{(Q)}$ can have highest value 1 and lowest value 0. The target function to achieve high quality $\phi^{(Q)}$ with lowest *c* is $\phi^{(T)} \in [0,1]$ which is:

$$\phi^{(T)}(c) = 1 - \frac{2c}{n}\phi^{(Q)}(c) \tag{18}$$

Strehl and Ghosh (2002) stated that the highest value of Weighted Inter-Intra Index (WI) gives the optimal number of cluster for a given data set**.**

C) Hartigan index (HI)

Hartigan index (HI) is applied to get the optimal number of cluster for a set of data. The index is calculated as follow (Hartigan,1975):

$$HI = (n - c - 1)\frac{err(c) - err(c+1)}{err(c+1)}$$

Here $err(c)$ describes the fitness of clustering result.

$$err(c) = \sum_{i=1}^{c}\sum_{a=1, a\in i}^{n} d^2(X_{a,} C_i) \tag{19}$$

In this formulation *c is* number of clusters and $c \geq 1$; *n* is the number of data points. *d* is the distance between data sample X_a and the centre C_i. *i*, *j* are indices. The highest lowest value indicates the optimal number of cluster for the input data. Bolshakova et al. 2003 and Dudoit et al. 2002 found that the lowest value of C-Index (CI) and Hartigan Index (HI) signifies the optimal number of cluster for a particular set of data.

D) R-squared Index (RSI)

R-squared index (RSI) is defined as follow (Sharma,1996) :

$$\mathrm{RSI} = \frac{SS_t - SS_w}{SS_t} \tag{20}$$

Where

$$SS_t = \sum_{j=1}^{m}\sum_{k=1}^{n_j}(X_A - \overline{x_j})^2 \quad \text{and} \quad SS_w = \sum_{\substack{i=1\ldots c \\ j=1\ldots m}}\sum_{k=1}^{n_{ij}}(X_A - \overline{x_j})$$

Here c is number of clusters, m is the number of dimensions of data, n_j is the number of data values of j dimension, n_{ij} corresponds to number of data values of j dimension that belongs to cluster i. $\overline{x_j}$ is mean data value of j dimension. X_A chosen any arbitrary data point.

Halkidi et al. (2001) affirmed that from R-squared Index (RI) the optimal number of cluster can be found from the Index vs. Number of cluster graph. The author proposed that the optimal number of cluster is that point for which the parametric value on the graph goes downward.

E) Krzanowski-Lai Index (KLI)

The index of Krzanowski and Lai (KLI) is defined as (Krzanowski and Lai,1985) :

$$KLI(c) = \left|\frac{DIFF(c)}{DIFF(c+1)}\right|$$

Where

$$DIFF(c) = (c-1)^{\frac{2}{p}} W(c-1) - c^{\frac{2}{p}} W(c) \tag{21}$$

And p denotes the number of features in the data set, c denotes the number of clusters and $c \geq 2$, $W(c)$ denotes the within cluster sum of square of the partition. Dudoit et al. (2002) suggested that the highest value of Krzanowski-Lai index (KLI) gives the optimal number of cluster for a given data set.

F) Davies-Bouldin Index (DBI)

This index is a function of the ratio of the sum of within-cluster scatter to between-cluster separation. This DB index is defined by (Davies and Bouldin,1979):

$$DBI = \frac{1}{c} \sum_{i,j=1}^{c} \max_{i \neq j} \left\{ \frac{D_c(i) + D_c(j)}{d_{ce}(i,j)} \right\}$$

Here $D_c = \frac{\sum_i \left\| X_a - C_i \right\|}{N_i}$ and $d_{ce} = \left\| C_i - C_j \right\|$ (22)

Here D_c and d_{ce} are intra-cluster and inter-cluster distances respectively. The intra cluster distances are calculated by average of pair wise distances from points in the cluster to the cluster centroid. The inter -cluster distance between two cluster is computed as the distance between their centroids. X_a any chosen arbitrary data that belongs to cluster i. N_i is number of data that belong to cluster i. C_i and C_j cluster centre of i and j cluster. So, when the cluster is compact and far from each other the ratio is small. Consequently Davies- Bouldin index is small for good cluster. Dimitriadou et al. (2002) found that optimal numbers of cluster for a data set is the number of cluster for which the Davies-Bouldin index (DBI) value is the lowest.

G) Silhouette Width Index

This index was proposed by Rousseeuw (1987) to evaluate clustering results. Silhouette width is a composite index which reflects the compactness and separation of the clusters. For each data point i the Silhouette width is calculated as follows:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{23}$$

Where *a(i)* is the average distance of a data point *i* to other data point in the same cluster, *b(i)* is the average distance of the that particular data point to all the data points belonging to the nearest cluster. The average *s(i)* of all data points reflects the quality of clustering result. Larger silhouette value signifies good cluster.

H) Calinski-Harabasz index

Calinski-Harabasz (1974) suggested the index for cluster validation purpose. This index uses the quotient between the intra-cluster average squared distance and inter-cluster average squared distance.

$$F = \frac{\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 - \sum_{k}^{c}\sum_{i=1}^{n_k}(X_i - \overline{X_k})^2\right] \Big/ c-1}{[\sum_{k}^{c}\sum_{i=1}^{n_k}(X_i - \overline{X_k})^2] \Big/ n-c} \tag{24}$$

Here n is total number of input data points. c is number of cluster. n_k is number of data points in cluster k (k=1,2,…c), and X_i and $\overline{X_k}$ are observation vectors for input data i and the centroid for group k, respectively. This parameter was found to be the best global statistic criterion in cluster evaluation by Milligan and Cooper (1985).

I) Dunn Index

The index was formulated by Dunn (1974) in order to check the quality of cluster resulted from a clustering algorithm. The equation is defined by:

$$D_{nc} = \min_{i=1,\ldots,nc}\left\{\min_{j=i+1,\ldots,nc}\left(\frac{d(c_i,c_j)}{\max\limits_{k=1,\ldots,nc}\left(diam(c_k)\right)}\right)\right\} \tag{25}$$

Here $d(c_i,c_j)$ is the dissimilarity function between two clusters c_i and c_j defined as $d(c_i,c_j) = \min\limits_{x\in C_i, y\in C_j} d(x,y)$ and $diam(c)$ is the diameter of a cluster which is the measure of dispersion of the clusters. The $diam(c)$ of the cluster can be defined as follows $diam(C) = \max\limits_{x,y\in C} d(x,y)$.

When the dataset contains compact and well-separated clusters, the distance between clusters will be large and diameter of the cluster will be small. So when the index value is large the clusters are compact and well-separated.

5.6 Summary

In this chapter an insight is given into the cluster analysis and various algorithms associated with it. Details of the four clustering algorithm i.e. SOM, AP, GA-Fuzzy and PAM used in this study is elaborated. Various cluster validation parameter and their significance in finding the optimum number of cluster for the input data set is described. These validation parameters are used to determine Number of Urban street class suitable in Indian context. Using validation parameters the most appropriate clustering algorithm for clustering of speed data in order to define free flow speed ranges of different urban street classes and speed ranges of different LOS categories is determined.

Chapter 6

Result Analysis for Defining LOS

6.1 Introduction

Result of cluster analysis is discussed in this chapter. In clustering the SOM, GA, AP and PAM algorithms were used. A basic idea about the algorithm used is described. By clustering using the above four algorithm the urban street segments were classified into four classes from the free flow speed data. After defining the segment into a particular class of urban street speed range for six LOS was found out.

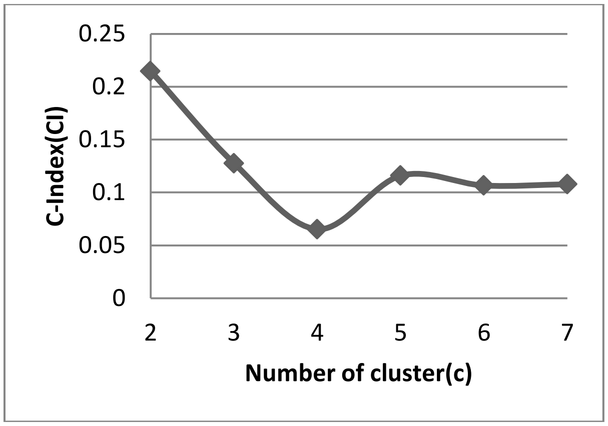
6.2 Application of Cluster Analysis Methods in Defining LOS Criteria of Urban Streets.

Average travel speeds were calculated direction wise on each segment. Four advanced cluster analysis techniques (SOM, AP, GA Fuzzy and PAM) were applied in two stages. Firstly, clustering methods were applied on average free flow speeds of all segments and free-flow speeds were classified into four groups. Each range of free flow speed found out indicates to an urban street class of I to IV. Secondly, clustering methods were applied on average travel speeds that were collected during peak and off peak hours on street segments for each of the urban street classes. In the second case, speeds were classified into six groups (A to F) for six categories of levels of service; thus speed ranges for level of service categories were defined in Indian context.

6.2.1 SOM Clustering

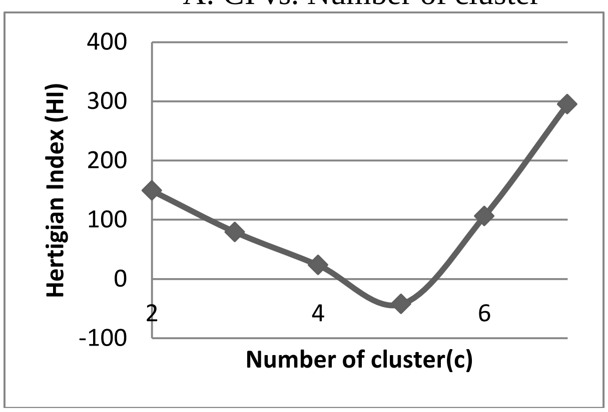
The free flow speed data acquired through GPS receiver was clustered using the SOM algorithm of ANN. For determination of the parametric value of validation measures, free flow speed data and cluster centre found from ANN analysis was used. In this research six validation parameters were used. Value of validation parameters were obtained for 2 to 7 number of cluster and were plotted in Figure 6.1 (A) to Figure 6.1 (F).

These six number of validation parameters were used to know the optimum number of cluster for this particular data set of free flow speed. By knowing the optimum number of cluster we can classify the urban street segments into that number of Urban street classes. It is always considered that lesser number of clusters is better if variation in validation parameters is minimal. Literature says that the lowest value of C-Index (CI) and Hertigian Index (HI) signifies the optimal number of cluster for a particular set of data. Figure 6.1 (A) and Figure 6.1 (C) show that the index are lowest for 4 number of clusters. Also, available literature says that the highest value of Weighted Inter-Intra Index (WI) gives the optimal number of cluster for a given data set; which is 4 as shown in Figure 6.1 (B). For R-squared Index (RSI), the optimal number of cluster is that point from where the Index vs. Number of cluster graph goes downward. Figure 2(D) shows the R-squared Index goes downward beyond four clusters; this goes in hand with CI, HI and WI giving the optimal number of cluster as 4. The highest value of Krzanowski-Lai index (KLI) gives the optimal number of cluster. From Figure 2 (E) it can be seen that KLI is highest for 3 numbers of cluster. Davies-Bouldin indicex (DBI) is shown in Figure 2(F) for which the optimal numbers of cluster is the point at which the index value is the lowest one and found to be 6. Out of six validation parameters considered in this study four parameters give the optimal cluster value as 4 which is also same as suggested by HCM-2000. That is the reason for which in this research the urban street segments were classified into four Classes by using the SOM of ANN.

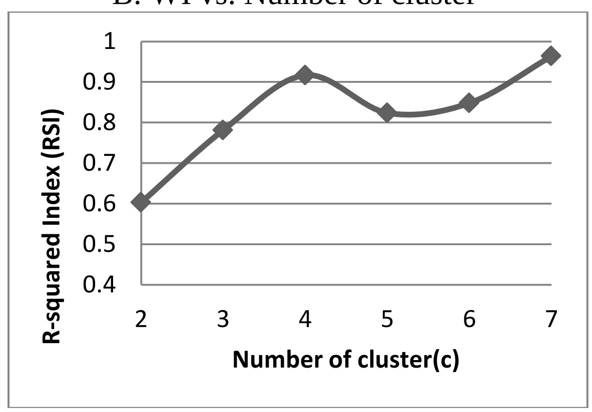


A: CI vs. Number of cluster

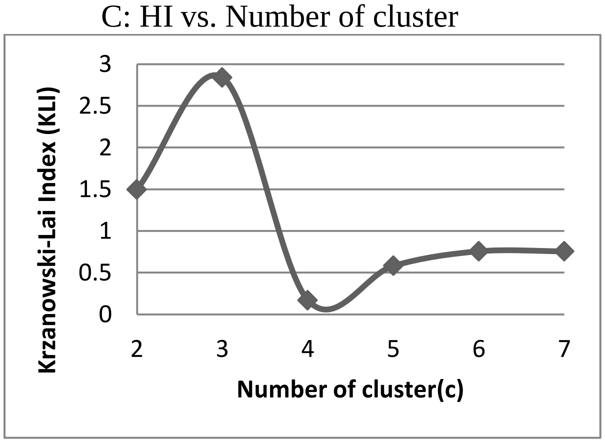
B: WI vs. Number of cluster



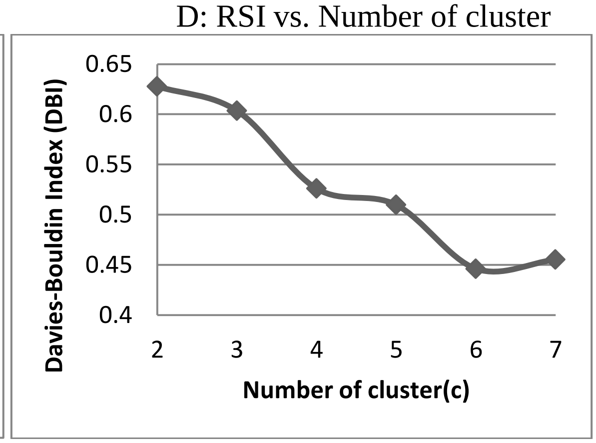
C: HI vs. Number of cluster



D: RSI vs. Number of cluster



E: KLI vs. Number of cluster



F: DBI vs. Number of cluster

Figure 6.1 Validation measures for optimal number of clusters using SOM clustering

Figure 6.2 shows the SOM weight positions for the input vector. This SOM weight positions plot shows how the SOM classified the input vector with the weight vector associated with it.

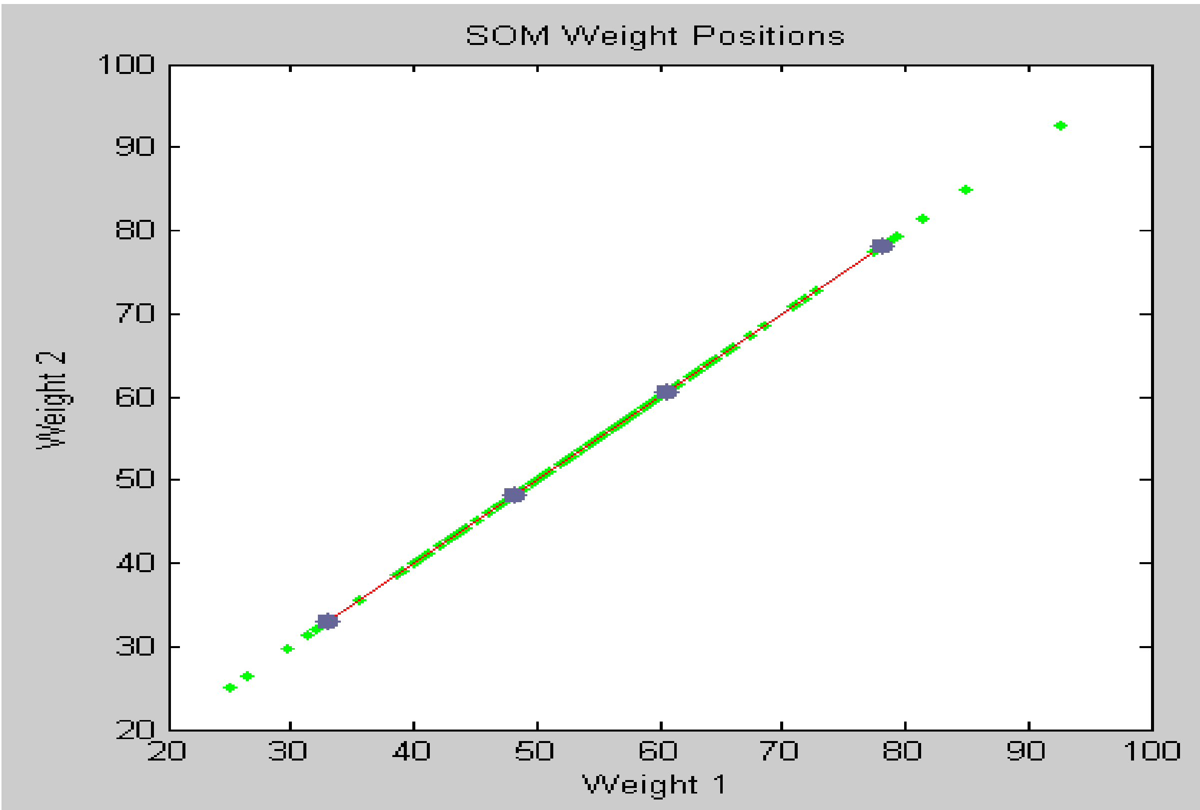


Figure 6.2 SOM weight Positions plot

In this study 100 urban street segments of five urban street corridors were analyzed. So to get the FFS ranges of different urban street class FFS of these 100 urban street segments were clustered using ANN. Self Organizing Map (SOM) which is known to be the best neural network for clustering data was utilized for this purpose. To cluster these FFS data four neurons were used in the SOM which acted as the centroid of a particular cluster. The SOM sample hit plot is shown in Figure 6.3. This figure shows clustering of FFS data around each neuron which acted as a centroid. As we have considered a hexagonal topology the neurons are shaped as hexagon. The number inside the hexagon depicts the amount of data associated with a particular cluster. From Figure 6.3 we can see two clusters have more numbers of data than the other two clusters.

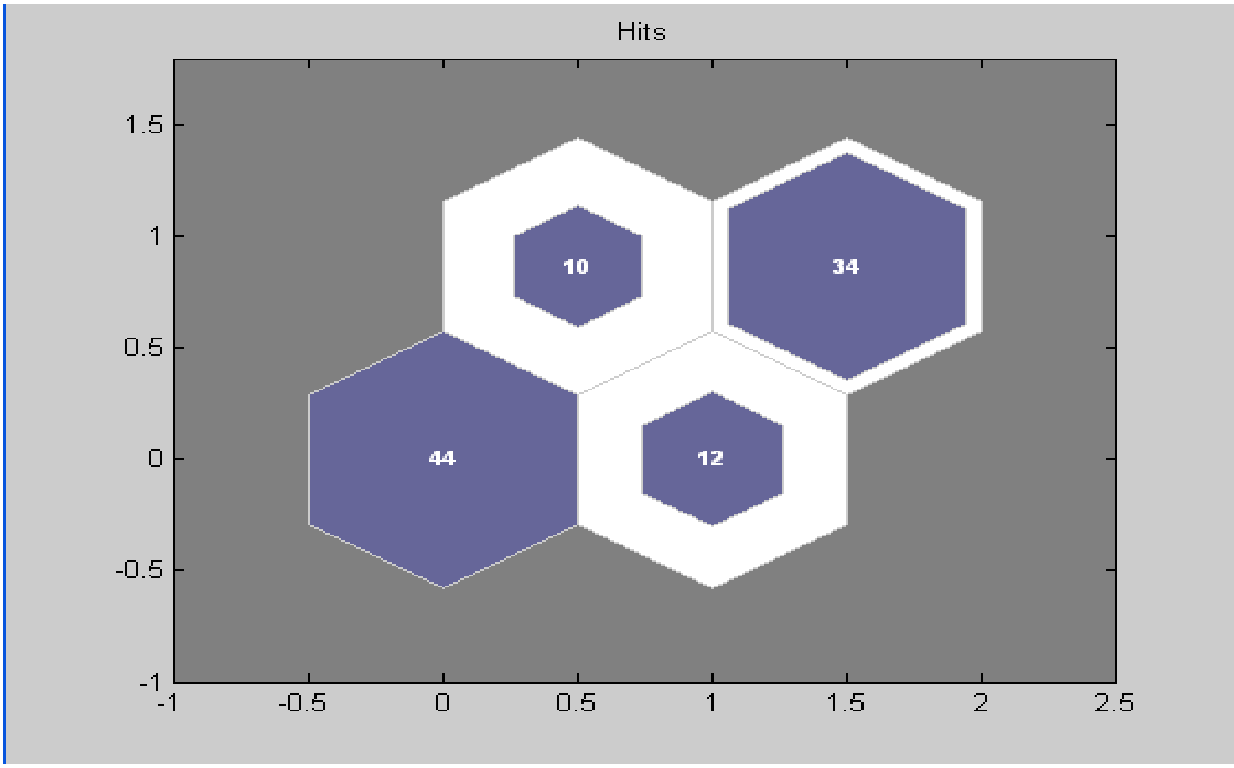


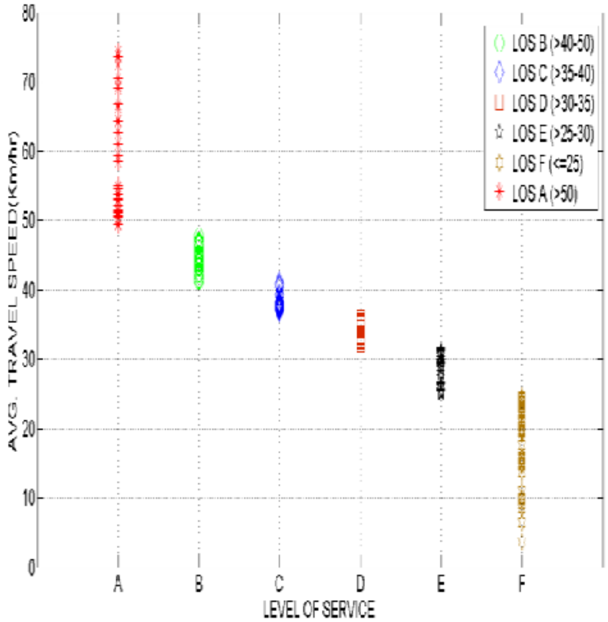
Figure 6.3 SOM sample hit Plot

Figure 6.4 shows the speed ranges for different urban street classes. Different symbol in the plot used for different urban street class. Observing these two figures it can be inferred most of the street segments belong to Urban Street class-I and Urban street class-II. In SOM sample hit plot hexagon with larger value corresponds to Urban street class-I and Urban Street class-II. It is observed from the collected data set that when a street segment falls under particular urban street class is agreed with the geometric and surrounding environmental condition of the road segments as well.

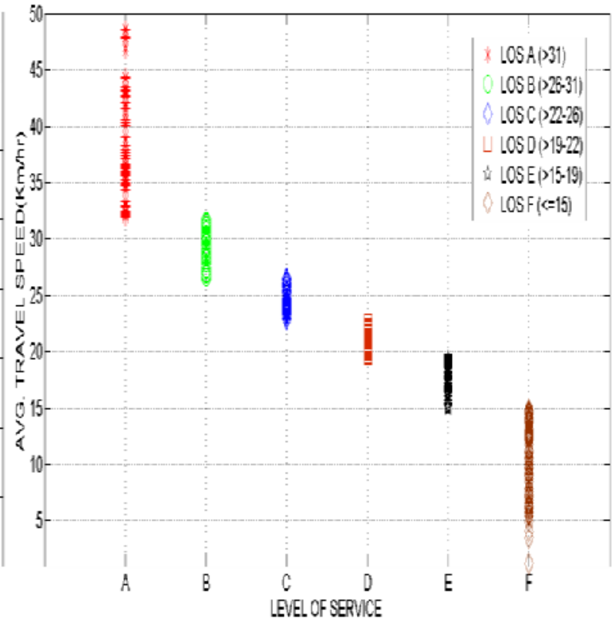
It has been found that there is very good correlation between free flow speed and geometric and environmental characteristics of streets under considerations.

Figure 6.4 SOM Clustering of FFS for Urban Street Classification

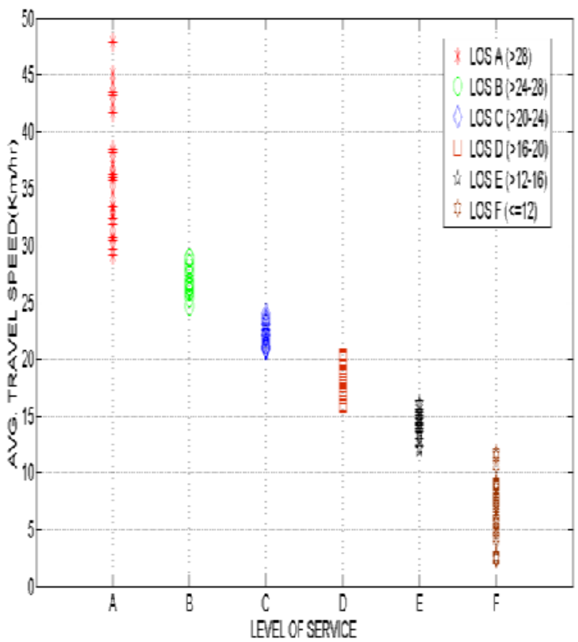
After classification of urban streets into number of classes, direction wise average travel speed on street segments during both peak and off peak hours were clustered using SOM to find the speed range of level of service categories. In fig. 5 the speed values are shown by different symbols depending on to which LOS category they belong. The legends in fig. 5 (A-D) gives the speed ranges for the six LOS categories obtained by using SOM clustering. The speed ranges for LOS categories found using SOM clustering is also shown in Table 6.1.



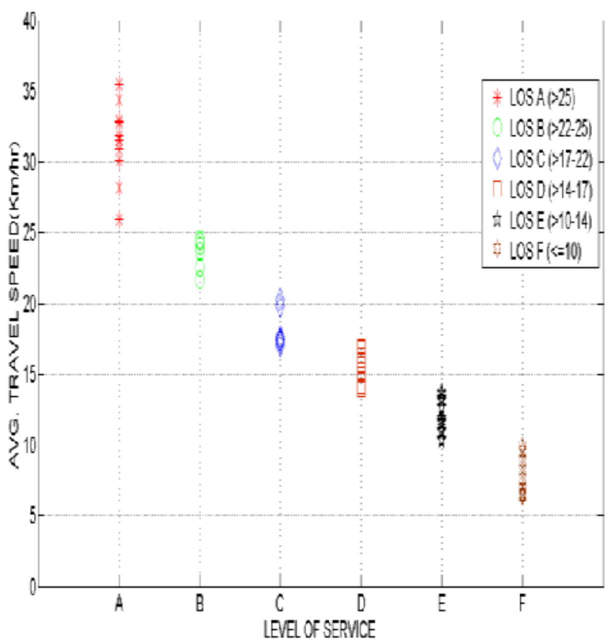
A: LOS of Urban Street Class I



B: LOS of Urban Street Class II



C: LOS of Urban Street Class III



D: LOS of Urban Street Class IV

Figure 6.5 Level of service of urban street classes (I-IV) using SOM clustering on average travel speeds

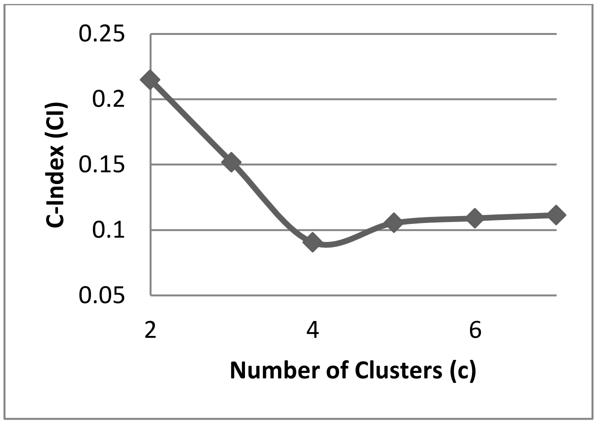
Table 6.1 Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by SOM Method

Urban Street Class	**I**	**II**	**III**	**IV**
Range of Free Flow Speed (FFS)	63 to 85 km/h	46 to 63 km/h	35 to 46 km/h	25 to 35 km/h
Typical FFS	68km/h	54km/h	42km/h	32 km/h
LOS	Average Travel Speed (Km/h)			
A	>50	>31	>28	>25
B	>40-50	>26-31	>24-28	>22-25
C	>35-40	>22-26	>20-24	>17-22
D	>30-35	>19-22	>16-20	>14-17
E	>25-30	>15-19	>12-16	>10-14
F	≤25	≤15	≤12	≤10

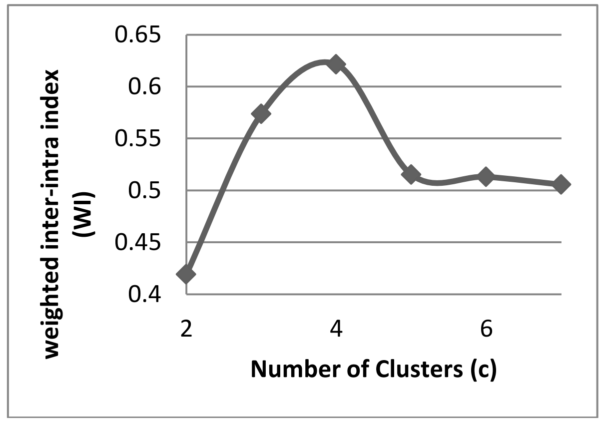
Speed ranges for level of service categories (A-F) expressed in percentage of free-flow speeds were found to be approximately 80, 65, 50, 45, 35 and 19-35 respectively in the present study. Whereas, in HCM (2000) it has been mentioned these values are 90, 70, 50, 40, 33 and 25-33 percent respectively.

6.2.2 Affinity Propagation (AP) clustering

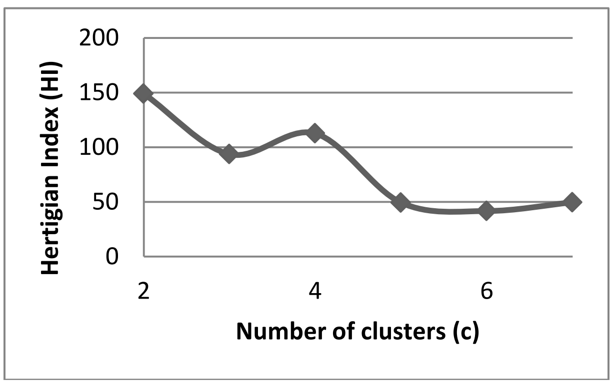
AP which is a very much new clustering tool developed in recent past is used to get the speed range for different urban street classes and travel speed ranges of LOS categories. The same six validation parameters that are used for ANN clustering are also used for AP clustering. For determination of the parametric value of validation measures, FFSs data acquired through GPS receivers which are averaged over street segment and their cluster centres found from AP cluster analysis is used as input to the validation parameters. Values of validation parameters are obtained for 2 to 7 number of cluster are plotted in Figure 6.6 (A) to Figure 6.6 (F).



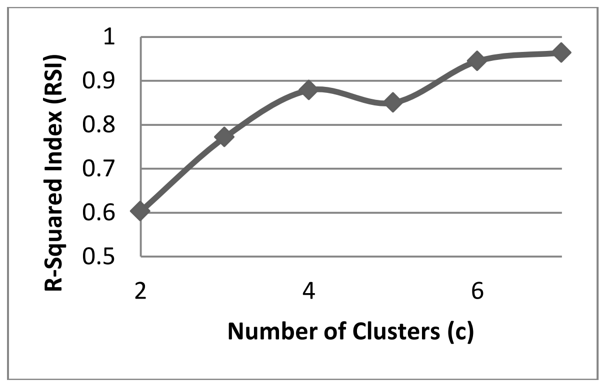
A: CI vs. Number of cluster



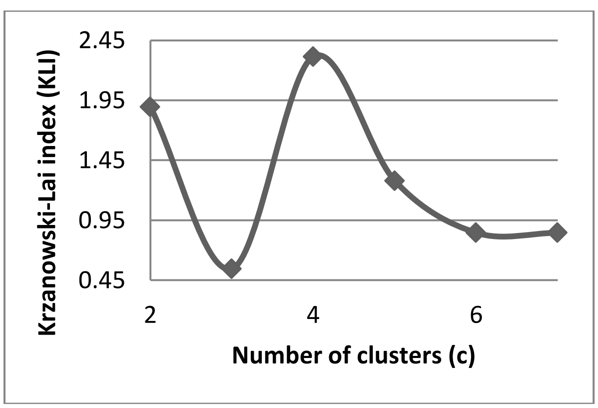
B: WI vs. Number of cluster



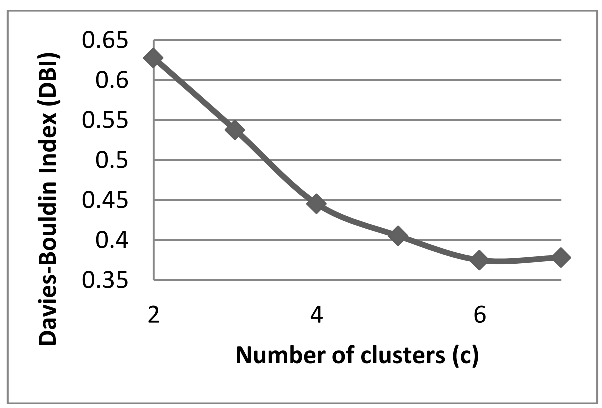
C: HI vs. Number of cluster



D: RSI vs. Number of cluster



E: KLI vs. Number of cluster



F: DBI vs. Number of cluster

Figure 6.6 Validation measures for optimal number of clusters using AP clustering

These six validation parameters are used to know the optimum number of cluster for FFS data set. Optimal number of cluster is basically concern with the quality of cluster obtained by applying a particular clustering algorithm to a particular data set. Every algorithm has its natural way of classification of the data set into number of groups. When a data set is clustered into its optimal number of cluster the quality of the cluster is best as the variation between the data points belonging to a particular cluster is minimal.

By knowing the optimum number of cluster the urban street segments can be classified into that number of Urban street classes. It is always considered that lesser number of clusters is preferred as optimal cluster if variation in parametric values between two consecutive numbers of clusters is minimal. Figure 6.6 (A) shows that the index is lowest for 4 numbers of clusters and beyond 4 number of cluster the variation in parametric values is very minimal so 4 is taken as the optimal number of cluster for CI. Further it can be clarified that when the FFS data are classified into four groups the pairs of FFS data with small distances (difference in speed values) falls under each group and hence sum of these values for four groups is minimal. Weighted inter-intra index vs. Number of cluster graph is plotted in Figure 6.6 (B). The index is having highest value for 4 number of cluster. So 4 is chosen as the optimal number of cluster going according to the literature. For four number of cluster the similarity in FFS values within a group is maximized and the similarity among four groups is minimized. Figure 6.6 (C) shows the index value is lowest for 5 numbers of clusters. 5 can be chosen as optimal number of cluster as literature states that number of cluster giving the lowest HI value is taken as optimal number of cluster. This index calculates the total square error of all the clusters and determines the index value for 2 to 7 number of clusters. When the FFS data is clustered into 5 numbers of clusters total square errors is found to be minimal. Figure 6.6 (D) shows the RSI goes downward beyond four clusters; this goes in hand with CI, HI and WI giving the optimal number of cluster as 4. In this case when FFS data clustered into 4 numbers of clusters sum of the square of FFS values of a particular cluster is maximum. From Figure 6.6 (E) it can be observed that KLI is highest for 4 numbers of clusters so 4 is chosen as the optimal number of cluster. When FFS values are clustered into 4 numbers of clusters the difference between FFS values of all clusters to their respective cluster centre is minimal. DBI is shown in Figure 6.6 (F) for which the optimal number of cluster is found to be 6 because according to literature number of cluster for which index value is lowest is chosen as optimal number of cluster. So if the FFS data set will be clustered into 6 numbers of groups the cluster centroid of each cluster will be far from each other and all the FFS values coming under a particular cluster will have the least distance from the cluster centroid. In simple words these parameter determines the quality of cluster obtained by classifying the FFS data into number of groups. It is found that when the input data set is clustered into optimal number of cluster the quality of cluster is best because street segments clustered into a particular group has the least difference in FFS and also street segments clustered into different cluster posses the maximum difference in FFS. Out of six validation parameters considered in this study four parameters give the optimal cluster value as 4 which is also same as suggested by HCM-2000. That is the reason for which in this research the urban street segments were classified into four Classes by using the AP algorithm.

The FFS data obtained by GPS receiver was clustered using this algorithm which is based on passing message between data point. After each road segment been assigned to a particular class of road the average travel speed data acquired during morning and evening peak hour was clustered into 6 group using AP. From this clustering the speed range for each particular type of LOS was determined. Figure 6.7 shows how the FFS data been clustered in the 2D space. Each color represents a particular cluster which corresponds to a particular class of urban street.

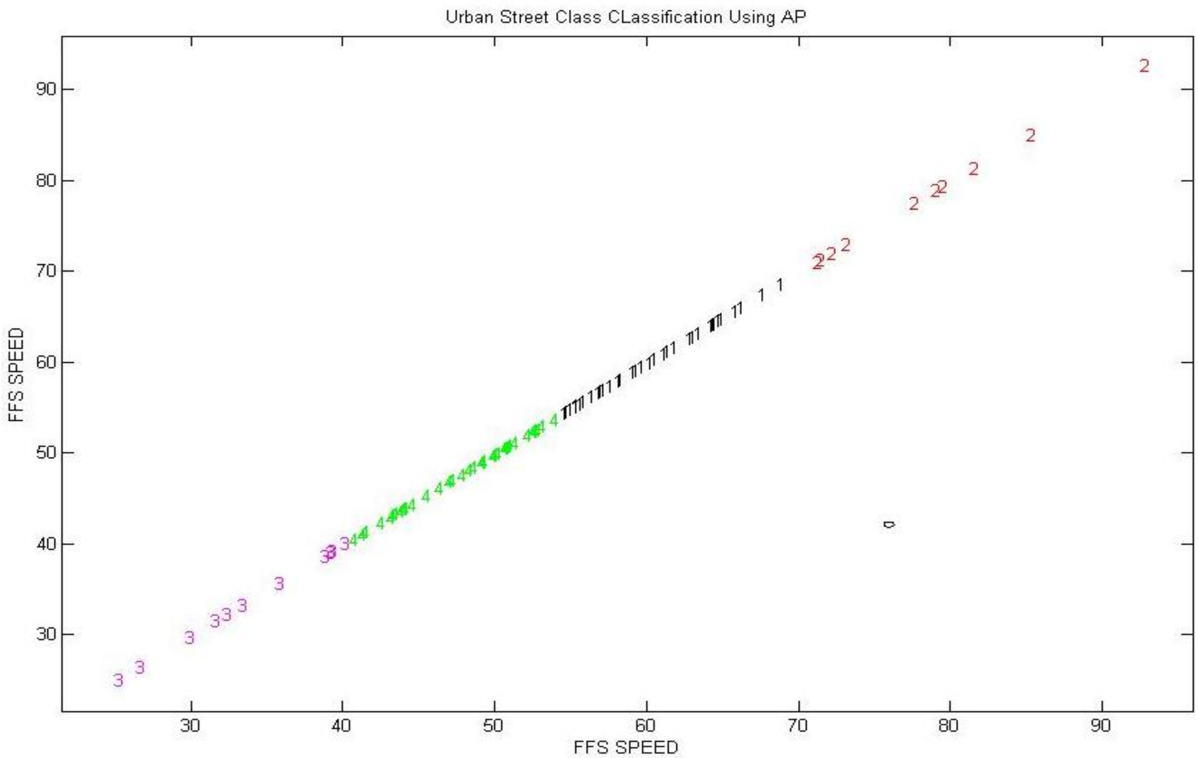


Figure 6.7 FFS data distribution plot after AP clustering

Hundred urban street segments of five urban street corridors were analyzed in this research. So to get the FFS ranges of different urban street class FFS of these 100 urban street segments were clustered using AP. AP which is known to a very flexible and accurate clustering technique is utilized for this purpose. Speed data are given to the algorithm in form of a similarity matrix. As from validation parameter analysis it was found the optimal number of cluster to be 4, number of exemplar chosen is four for clustering o FFS. The distribution of data after AP clustering is shown in Figure 6.8. Different symbols represent the class of urban street. Four different colours used for four different urban street classes. From Figure 6.8 It can be seen most of the street segments belong to Urban street class-I and Urban street class-II. Very few street segments belong to Urban street class-I.

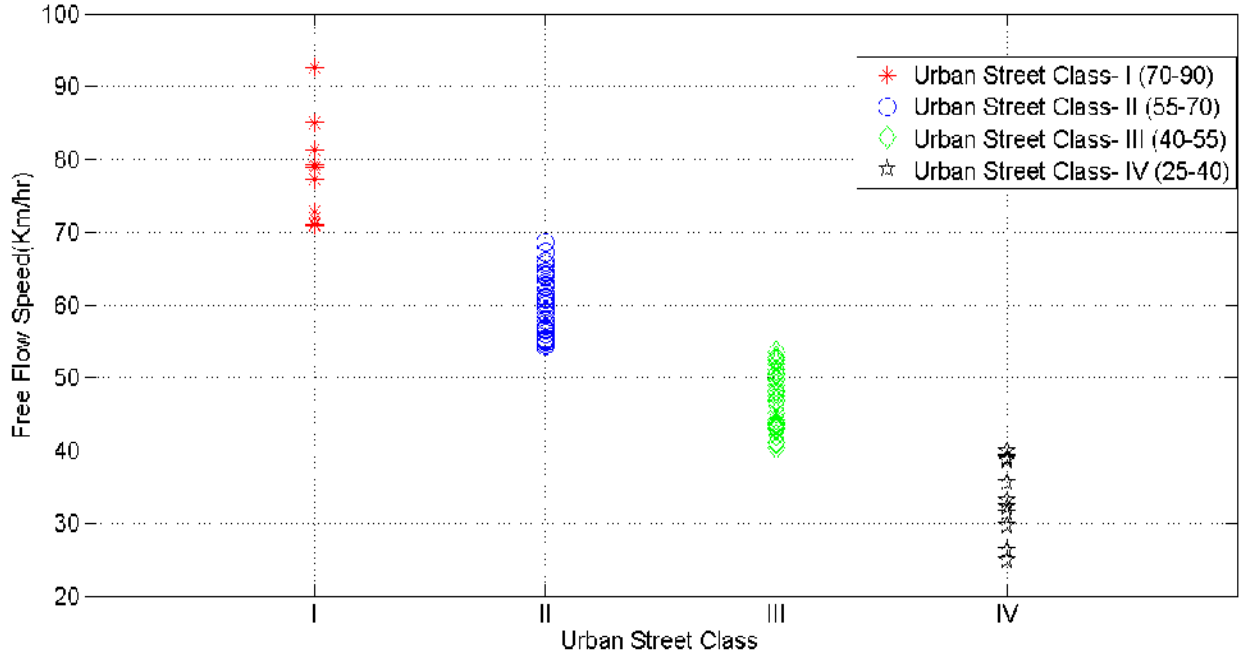
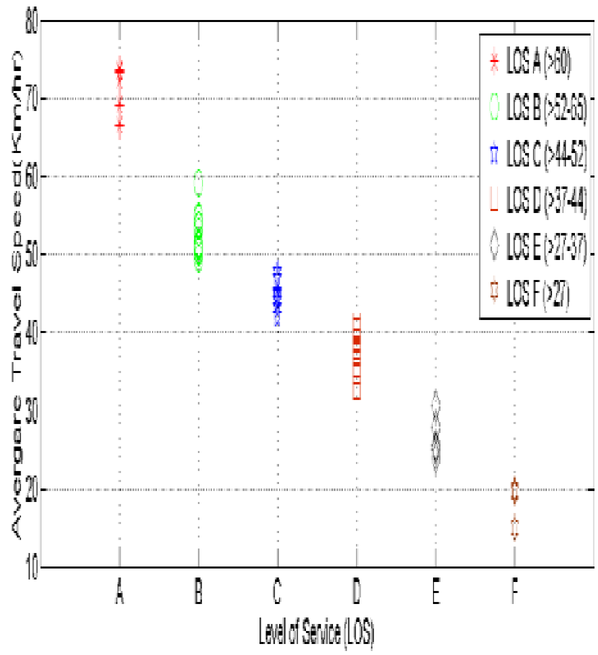
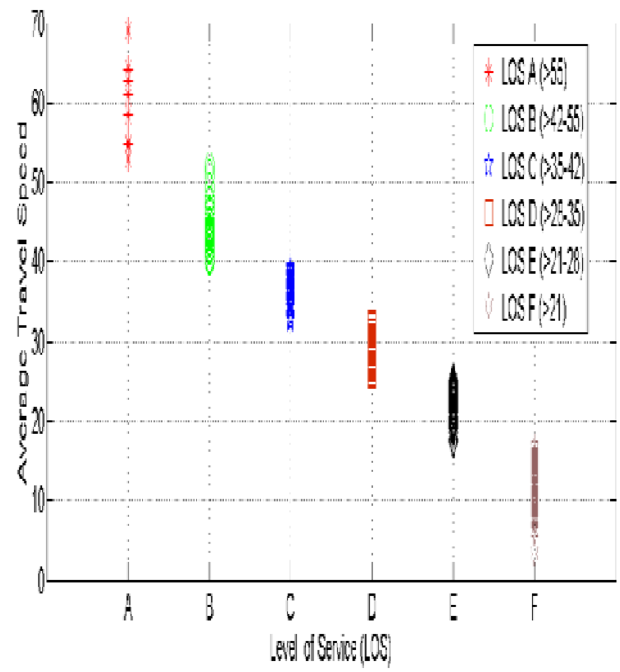


Figure 6.8 AP Clustering of FFS for Urban Street Classification

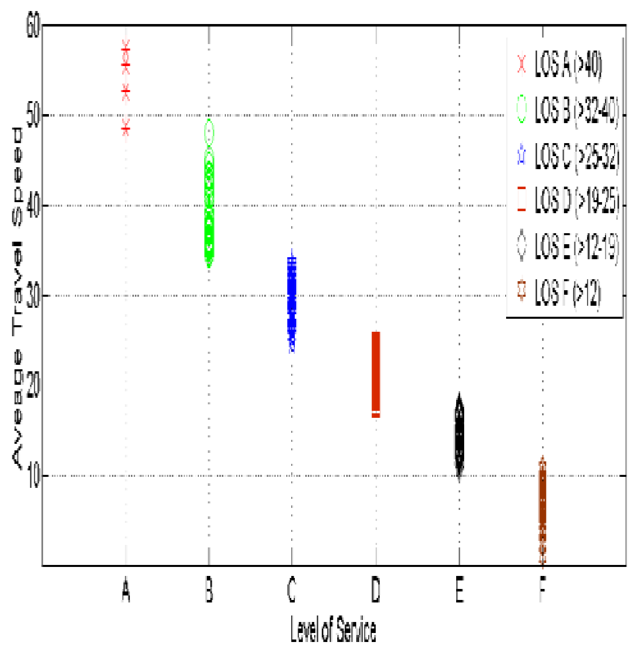
AP clustering is used for second time after classification of urban street segments into four different urban street classes. For four classes of urban street classes LOS speed ranges were defined using AP clustering. In this clustering direction wise average travel speed data of both peak and off peak hours were given as input to the clustering algorithm. In Figure 6.9 the clustering result are complied. The legends in Figure 6.9 (A-D) gives the speed ranges for the six LOS categories obtained by using AP clustering. The speed speed ranges for level of service categories (A-F) expressed in percentage of free-flow speeds were found to be approximately 80, 62, 48, 35, 23 and 18-23 respectively in the present study. Whereas, in HCM (2000) it has been mentioned these values are 90, 70, 50, 40, 33 and 25-33 percent respectively.



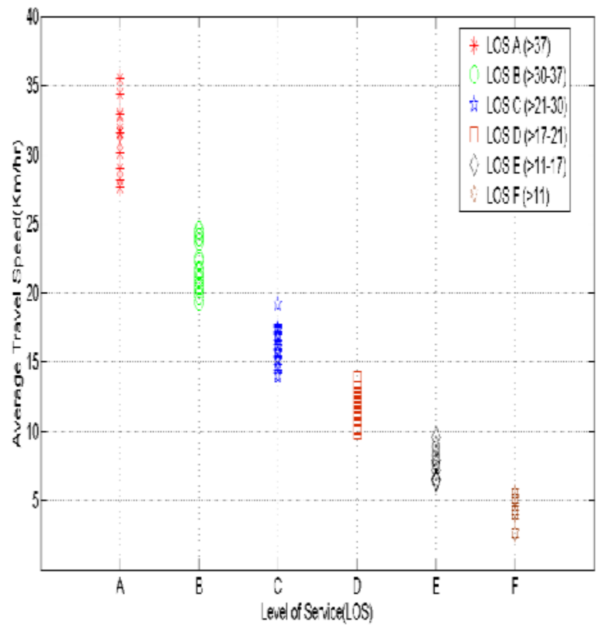
A: LOS of Urban Street Class I



B: LOS of Urban Street Class II



C: LOS of Urban Street Class III



D: LOS of Urban Street Class IV

Figure 6.9 Level of service of urban street classes (I-IV) using AP clustering on average travel speed

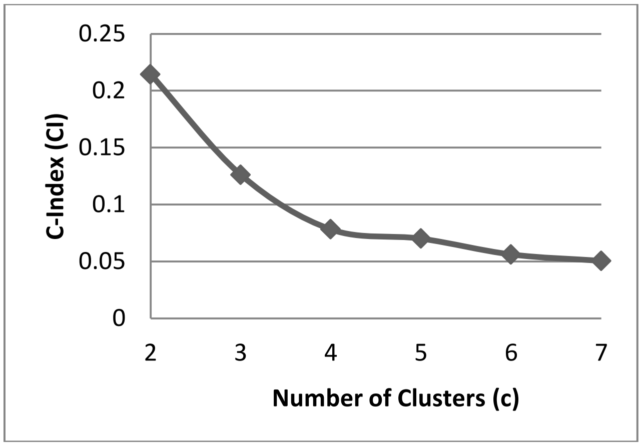
Table 6.2 Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by AP Method

Urban Street Class	**I**	**II**	**III**	**IV**
Range of free-flow speed (FFS)	90 to 70 km/h	70 to 55 km/h	55 to 40 km/h	40 to 25 km/h
Typical FFS	75km/h	60km/h	47km/h	35 km/h
LOS	**Average Travel Speed (Km/h)**			
A	>60	>52	>45	>28
B	>50-60	>40-52	>35-45	>20-28
C	>42-50	>32-40	>25-35	>14-20
D	>33-42	>25-32	>17-25	>10-14
E	>22-33	>17-25	>12-17	>6-10
F	>22	>17	>12	>6

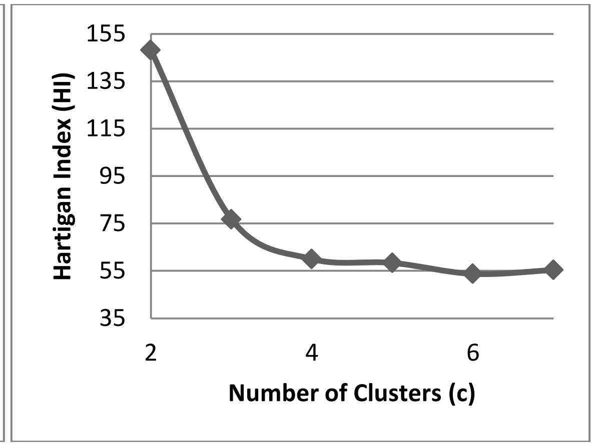
6.2.3 GA-Fuzzy clustering

The most popular method in fuzzy classification called Fuzzy *C*-Means (FCM) clustering is used for clustering free flow speed of urban street segments. From this clustering the type of urban class particular segment belongs to is determined than the average travel speed of each segment is used for clustering process to know the speed range for a particular LOS. GA is used for optimization of search, to get the local minima which differ from global minima. The search for the global minimum can't be realized due to a large volume of calculations, but GA is used to get a sufficiently good solution.

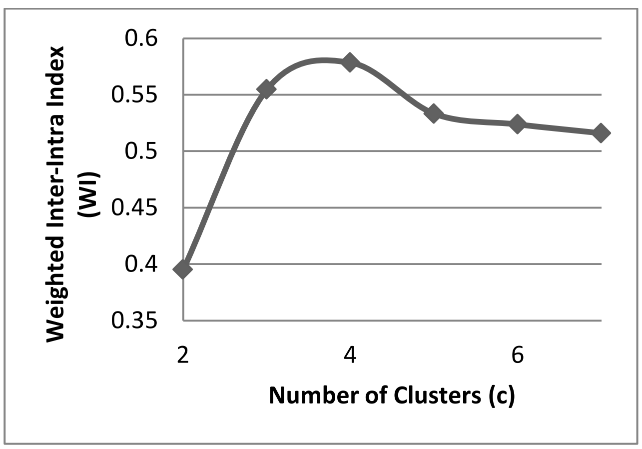
Like the previous two algorithm both input data (free flow speed) to GA-Fuzzy clustering and its output (cluster centres) found from the cluster analysis are used in computing the values of cluster validation parameters. The values of five validation measures obtained for 2 to 7 number of clusters are plotted in Figure 6.10 (A) to Figure 6.10 (E).



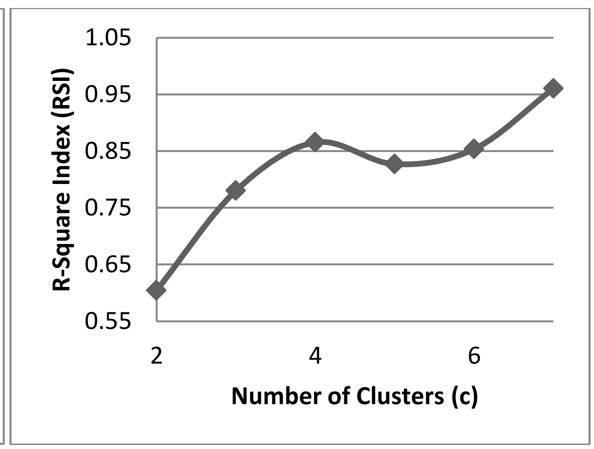
A: CI vs. Number of cluster



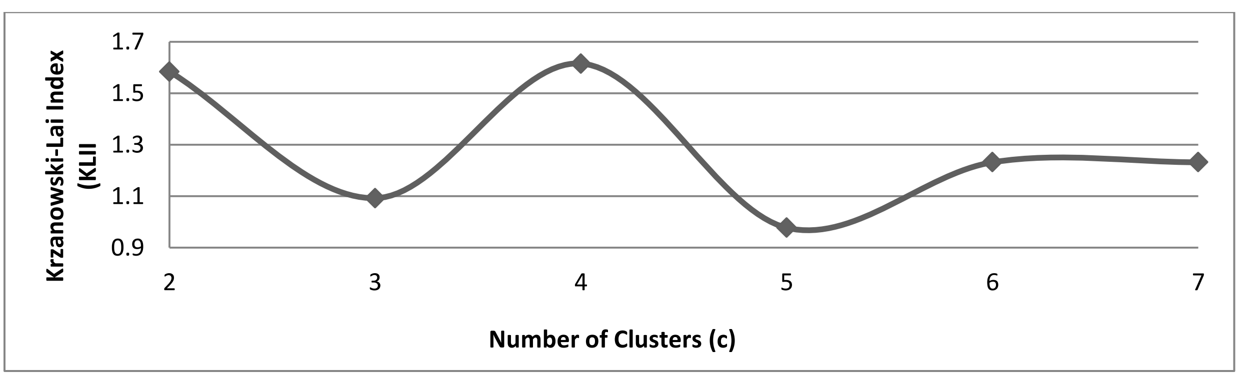
B: HI vs. Number of cluster



C: WI vs. Number of Cluster



D: RSI vs. Number of Cluster



E: KLI vs. Number of cluster

Figure 6.10 Validation measures for optimal number of clusters using GA-Fuzzy clustering

Five validation parameters are interpreted to obtain the optimum number of clusters for deciding the classification of street segments into different urban street classes. If variation in parametric values from one cluster to the next cluster is not significant it is always considered to go for lesser number of clusters. From Literature review it was believed that the lowest value of C-Index (CI) signifies the optimal number of cluster for a particular set of data. The index value is lowest for four numbers of clusters which is shown in Figure 6.10 (A). Beyond 4 numbers of clusters the index value all most remains same and the variation is minimal. So 4 is taken as the optimal number of clusters for Hartigan Index (HI) as shown in Figure 6.10 (B). Also from literature review it was found that the highest value of Weighted Inter-Intra Index (WI) gives the optimal number of cluster for a given data set; which is 4 as shown in Figure 6.10 (C). For R-squared Index (RI), the optimal number of cluster is that point from where the Index vs. Number of cluster graph goes downward. Figure 6.10 (D) shows that R-squared Index goes downward beyond four clusters. The highest value of Krzanowski-Lai Index (KI) gives the optimal number of cluster and from Figure 6.10(E) it can be seen that KI is highest for 4 numbers of cluster. Thus this goes in hand with CI, HI, WI and RI. Hence all the five validation parameters considered in this study give the optimal cluster value as 4 which is also same as suggested by HCM-2000. That is the reason for which in this research the urban street segments were classified into four Classes by using GA-Fuzzy clustering technique.

In this study data collected from five major urban corridors of Mumbai city comprising of 100 street segments were analyzed. Second wise free flow speed data collected during the night hours using GPS receiver are averaged over each segment. Average of these averaged values taken for each travel run on street segments are used by a hybrid GA-Fuzzy algorithm for the classification of street segments into number of classesThe result obtained using the GA-Fuzzy algorithm for clustering purpose has been illustrated in Figure 6.11. Different types of symbols are used for each type of urban street class. From the figure it can be inferred that more number of street segments (free flow speed data points) belong to Urban Street Class-II and Urban Street Class-III than Urban Street Class-I and Urban Street Class-IV.

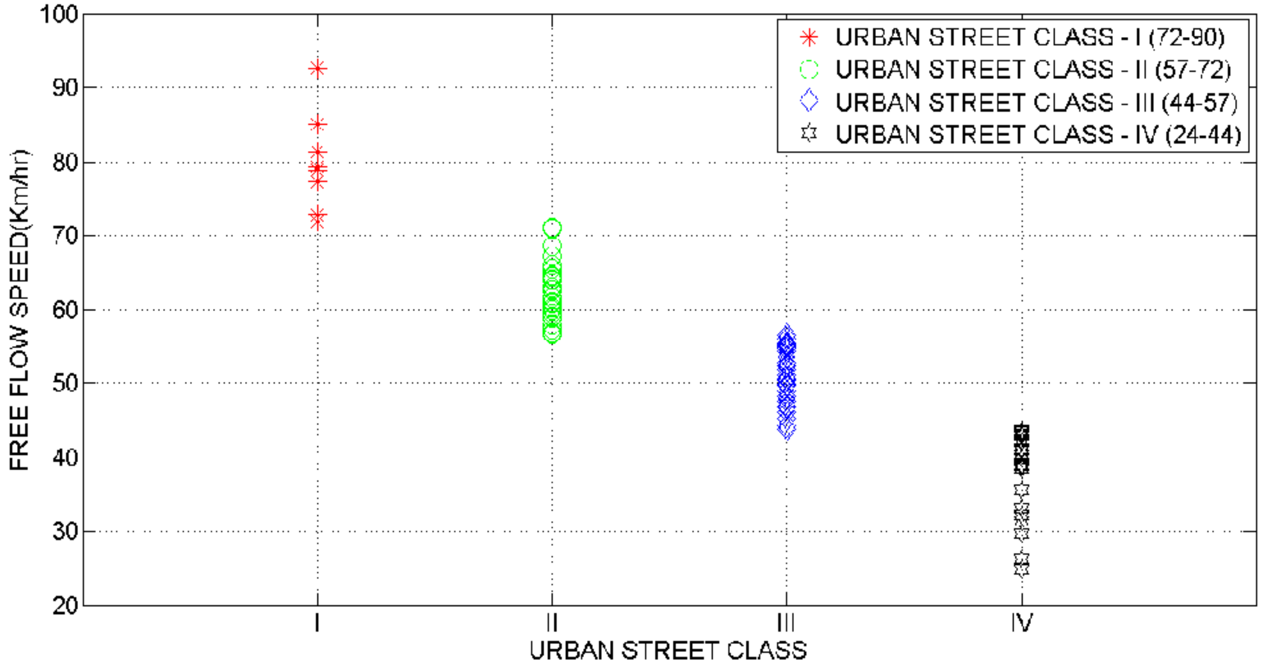
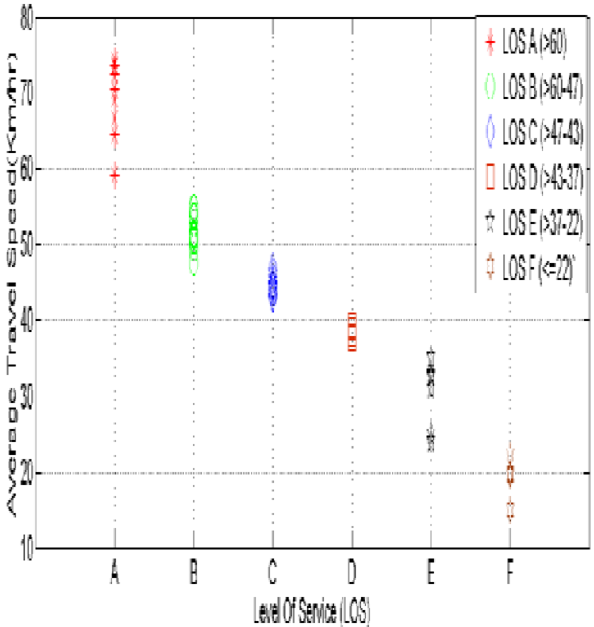
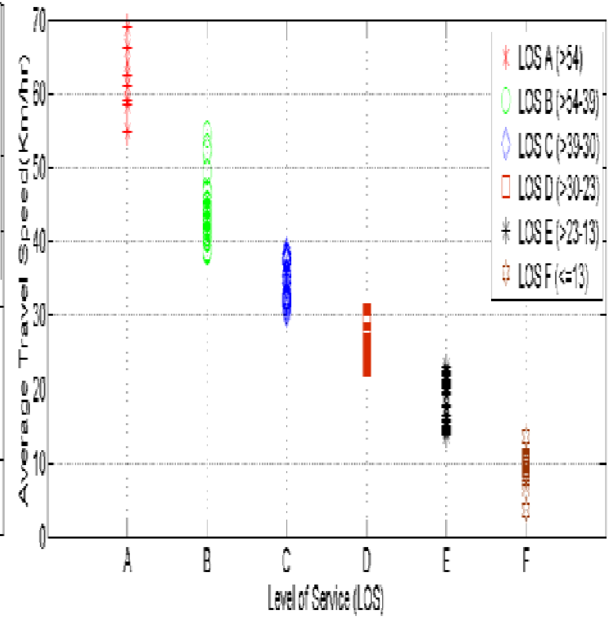


Figure 6.11 GA-Fuzzy Clustering of FFS for Urban Street Classification

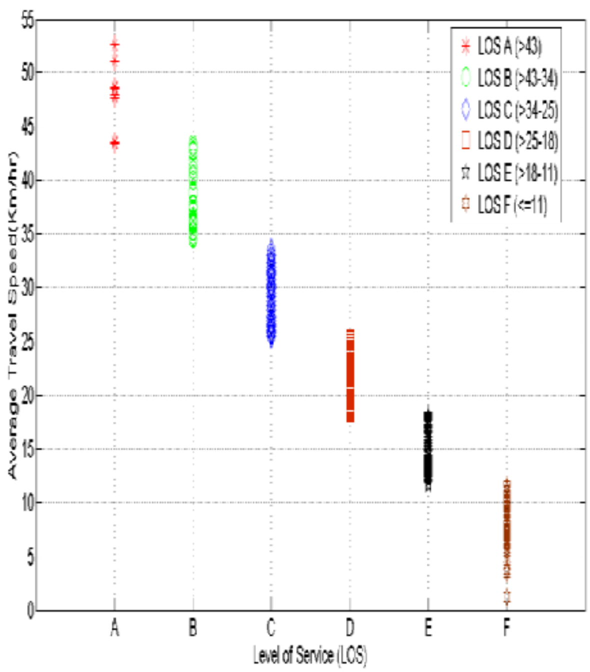
The same state of the art algorithm was used for second time to the average travel speed data acquired during peak and off peak hour of the above stated five urban street corridors. The GA-Fuzzy algorithm clustered the average speed data into six clusters to give the speed ranges of different urban street classes. The result of the clustering is shown in Figure 6.12 (A) to Figure 6.12(D). Each Level of Service (LOS) for a particular urban street class illustrated in the figure with a unique symbol. The speed ranges for each individual Level of Service categories elaborated in Table 6.3. It can be outlined that the free flow speed range of urban street classes and speed ranges of level of service categories that was resulted from this cluster analysis are significantly lower than that stated in HCM-2000.



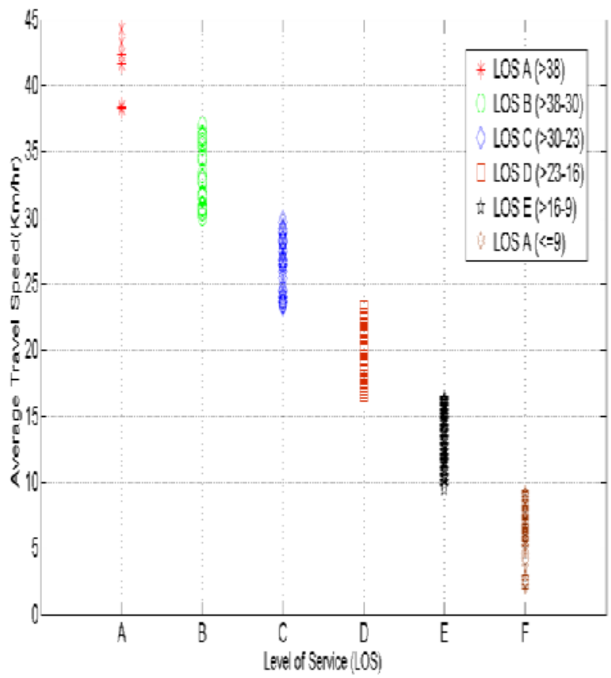
A: LOS of Urban Street Class I



B: LOS of Urban Street Class II



C: LOS of Urban Street Class III



D: LOS of Urban Street Class IV

Figure 6.12 Level of service of urban street classes (I-IV) using GA-Fuzzy clustering on average travel speeds

Average travel speed of different Level of Service categories in terms of percentage of free-flow speeds was found. The approximately percentage value for LOS 'A' to LOS "F" are 85, 70, 56, 43, 28 and 16-28 respectively in the present study. Whereas, in HCM (2000) it has been mentioned these values are 90, 70, 50, 40, 33 and 25-33 percent respectively. The lower values of percentage of free-flow speed for different LOS categories point at the poor quality of road and heterogeneity of traffic condition prevailing in Indian urban street corridors.

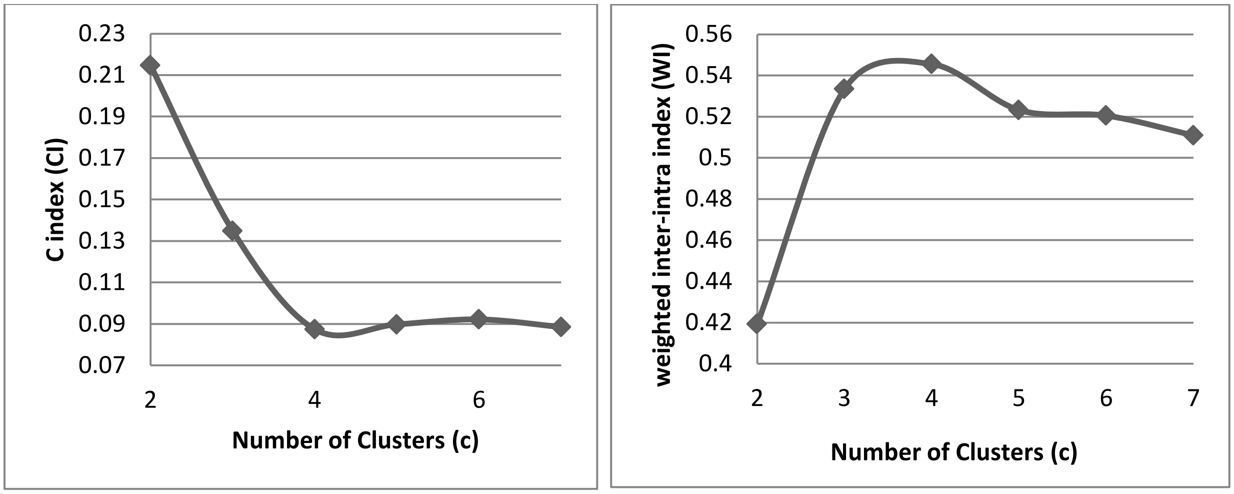
Table 6.3 Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by GA-Fuzzy Method

Urban Street Class	I	II	III	IV
Range of Free Flow Speed (FFS)	72 to 90 km/h	57 to 72 km/h	44 to 57 km/h	24 to 44 km/h
Typical FFS	77km/h	63km/h	50km/h	35 km/h
LOS	Average Travel Speed (Km/h)			
A	>60	>54	>43	>38
B	>60-47	>54-39	>43-34	>38-30
C	>47-43	>39-30	>34-25	>30-23
D	>43-37	>30-23	>25-18	>23-16
E	>37-22	>23-13	>18-11	>16-9
F	≤22	≤13	≤11	≤9

6.2.4 Partition Around Medoid (PAM) clustering

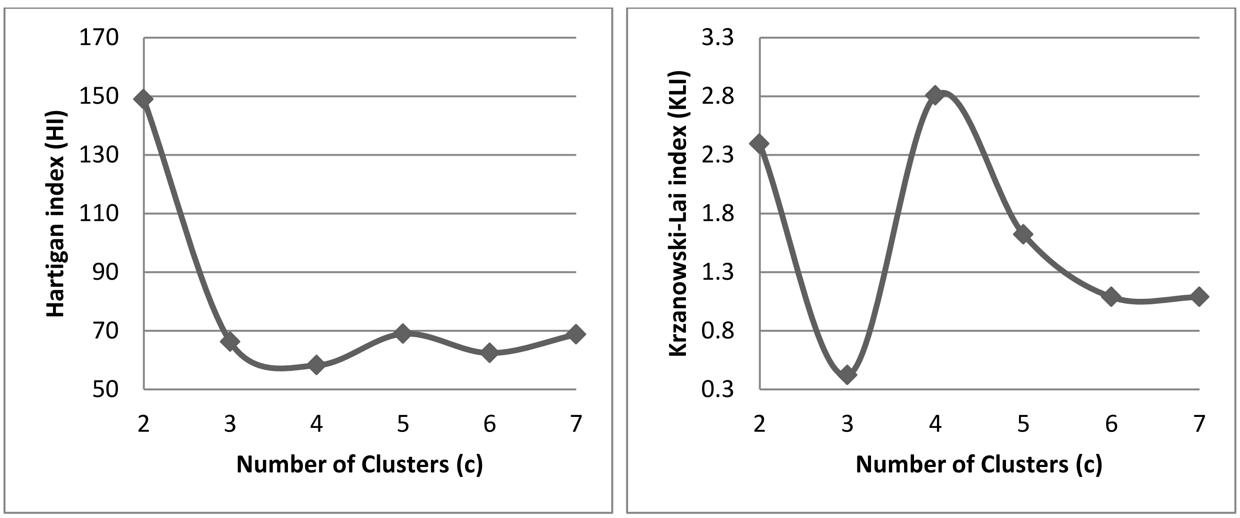
This algorithm attempts to minimize the total distance between objects within each cluster. The algorithm proceeds through two phases. In the first phase, an initial clustering is obtained by the successive selection of representative objects until *k* representative objects have been found. The first representative object is the one for which the sum of the dissimilarities to all objects is as small as possible. This representative object is the most centrally located in the set of objects. Subsequently, at each step another object is selected. This object is the one which decreases the objective function as much as possible. In the second phase of the algorithm, it attempts to improve the set of representative objects and therefore also to improve clustering yielded by this set.

To know the optimal number of cluster for this algorithm four validation parameters are used. For determination of the parametric value of validation measures, FFSs data acquired through GPS receivers which are averaged over street segment and their cluster centres found from PAM cluster analysis is used as input to the validation parameters. Values of validation parameters are obtained for 2 to 7 number of cluster are plotted in Figure 6.13 (A) to Figure 6.13 (D). These four validation parameters are used to know the optimum number of cluster for FFS data set. Figure 6.13 (A) illustrates C Index. The index value is lowest for 4 numbers of clusters and beyond 4 the variation is minimal. So going with literature 4 is taken as optimal number of cluster. Weighted inter-intra Index vs. Number of clusters graph is shown in Figure 6.13 (B). As discussed previously the highest value of this index is for 4 number of cluster so 4 can be chosen as the optimal number of cluster. HI and KLI are shown in Figure 6.13 (C) and Figure 6.13 (D) respectively. HI value is lowest for 4 number of cluster and KLI value is highest for 4 numbers of clusters. As discussed previously lowest value of HI and highest value of KLI signifies optimal number of clusters. For this algorithm all the parameter considered gave 4 as the optimal number of cluster for the data set. So the FFS data can be clustered into four numbers of groups. So FFS will be clustered into four numbers of urban street classes. HCM-2000 also opines the same to classify urban street segments into four numbers of classes.



A: CI vs. Number of cluster

B: WI vs. Number of cluster



C: HI vs. Number of Clusters

D: KLI vs. Number of Clusters

Figure 6.13 Validation measures for optimal number of clusters using PAM clustering

After demining the number of groups the FFS data to be clustered into, using PAM algorithm FFS data are clustered into four numbers of groups to know the FFS ranges of each urban street class. The clustering result of PAM is shown in a 2D space in Figure 6.14. In the figure each color corresponds to a particular urban street class.

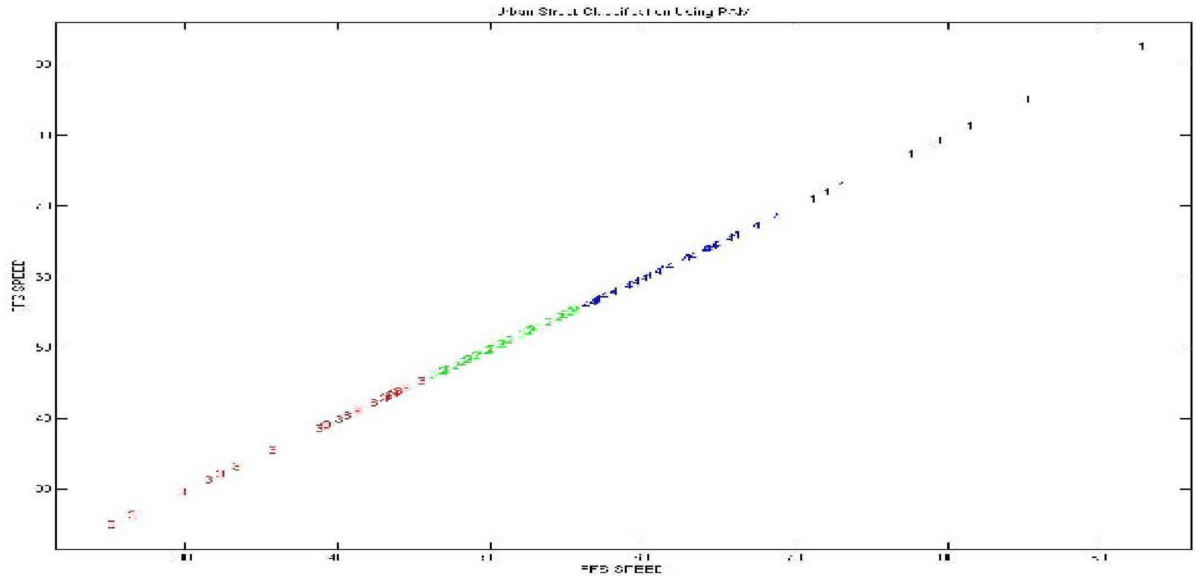


Figure 6.14 FFS data distribution plot after PAM clustering

In Figure 6.15 each FFS data is shown. Different colors are used for the FFS data that belongs to different urban street classes. From the figure the FFS ranges of each urban street class can also be determined. From the figure it can be inferred that very few number of street segments belong to Urban street class-I.

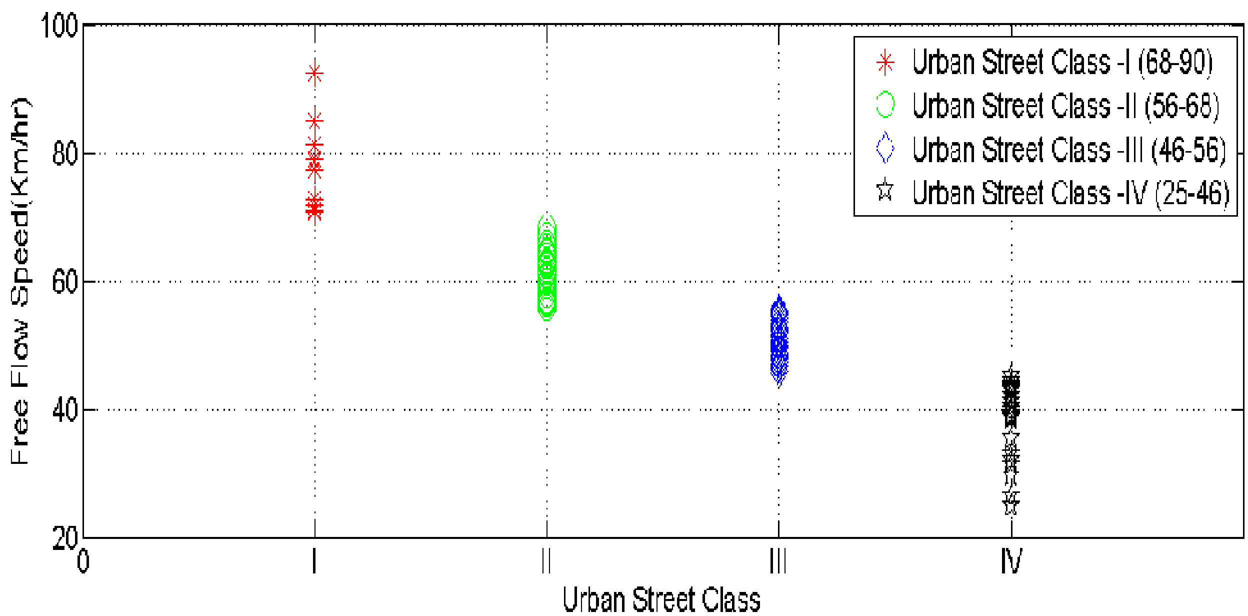
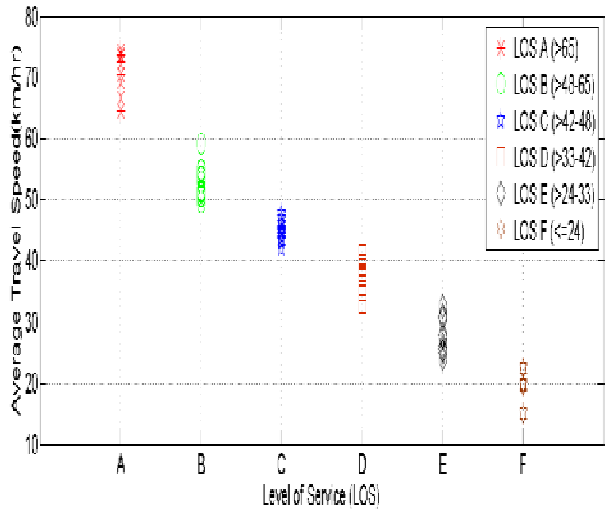
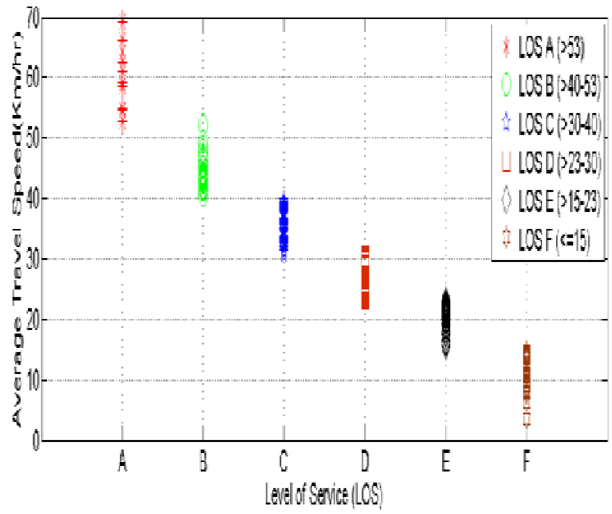


Figure 6.15 PAM Clustering of FFS for Urban Street Classification

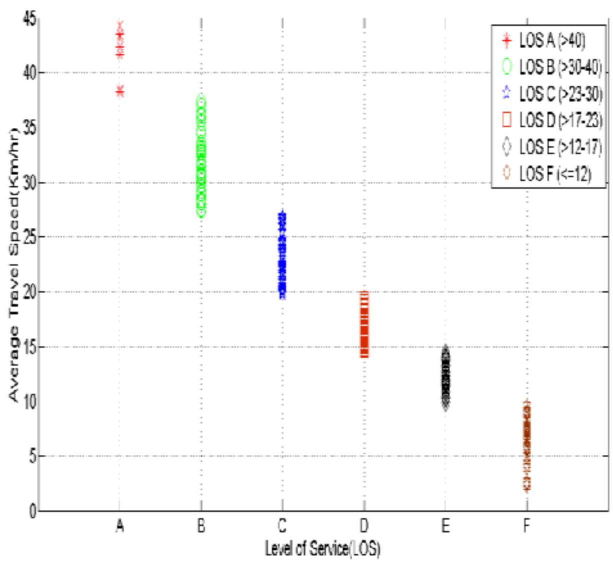
The variation in FFS ranges of different urban street class from this algorithm and previously discussed algorithm is not significant but the speed ranges of different LOS categories that is achieved after clustering of travel speed data by the same PAM algorithm is remarkable. The clustering result is elaborated in Figure 6.16 (A-D). Each LOS category is illustrated with different symbol and color.



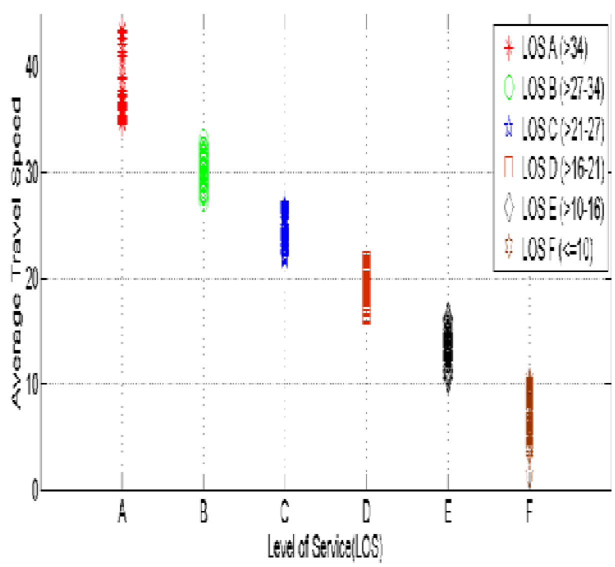
A: LOS of Urban Street Class I



B: LOS of Urban Street Class II



C: LOS of Urban Street Class III



D: LOS of Urban Street Class IV

Figure 6.16 Level of service of urban street classes (I-IV) using PAM clustering on average travel speeds

The speed ranges for level of service categories (A-F) expressed in percentage of free-flow speeds were found to be approximately 85, 75, 60, 45, 33 and 20-30 respectively in the present study. The speed ranges of urban street classes and LOS categories are given in Table 6.4.

Table 6.4 Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by PAM Method

Urban Street Class	I	II	III	IV
Range of free-flow speed (FFS)	90 to 68 km/h	68 to 56 km/h	56 to 46 km/h	46 to 25 km/h
Typical FFS	75km/h	60km/h	50km/h	40 km/h
LOS	Average Travel Speed (Km/h)			
A	>65	>53	>40	>34
B	>48-65	>40-53	>30-40	>27-34
C	>42-48	>30-40	>23-30	>21-27
D	>33-42	>23-30	>17-23	>16-21
E	>24-33	>15-23	>12-17	>10-16
F	≤24	≤15	≤12	≤10

6.3 Selecting the Best Clustering Method in Defining LOS Criteria of Urban Streets

In this study four advance clustering tools are used in to get the FFS ranges of different urban street classes and to define travel speed ranges of LOS categories but to get the most suitable algorithm for the input data four parameters are used in this study. In order to get optimal number of cluster for a particular algorithm indices value for 2-7 numbers of clusters were calculated. To determine the best clustering algorithm four cluster quality evaluation parameters are used. Index value of each parameter is calculated for four algorithms considering 2-7 numbers of clusters. Different colors are used for each algorithm to illustrate the result clearly.

In Figure 6.17 Silhouette width vs. Number of cluster is shown. Previously it was found 4 is the optimal number of cluster. So for 4 number of cluster the algorithm having highest Silhouette width index is the best clustering algorithm in this context. From figure it can be seen that AP is having highest index value.

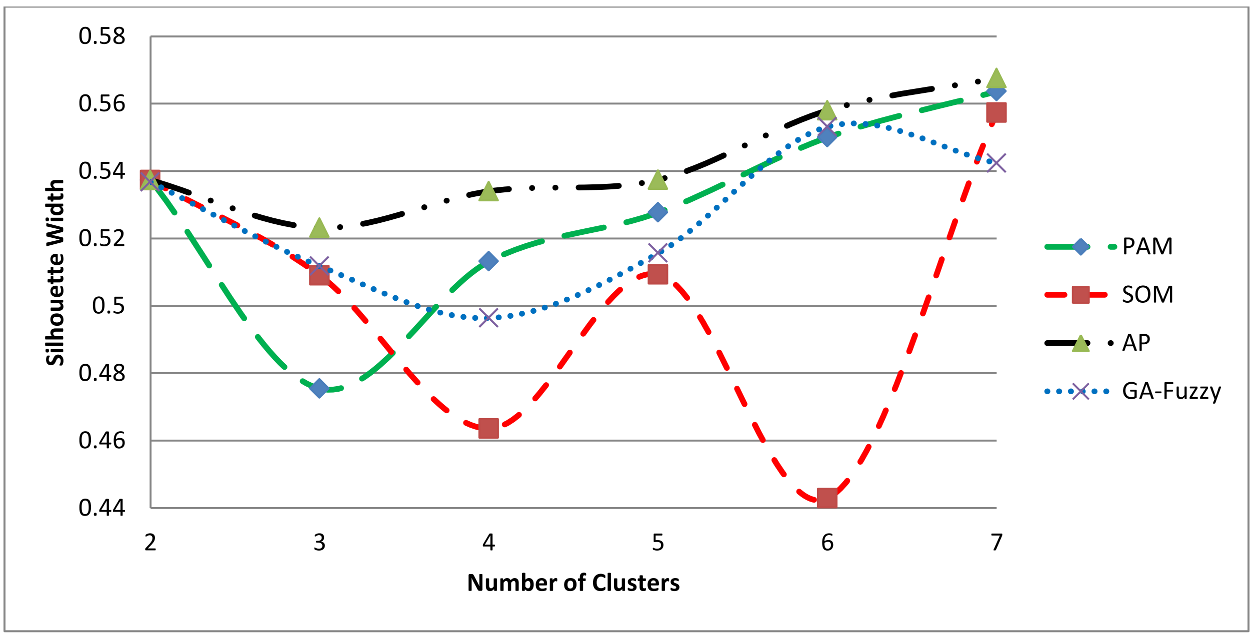


Figure 6.17 Silhouette width vs. Number of cluster Plot for different clustering algorithms

Calinski-Harabasz Index vs Number of cluster plot is given in Figure 6.18. Like the previous parameter for 4 number of cluster AP algorithm is having the highest index values. So going with the literature this index also validates the result obtained from Silhouette width.

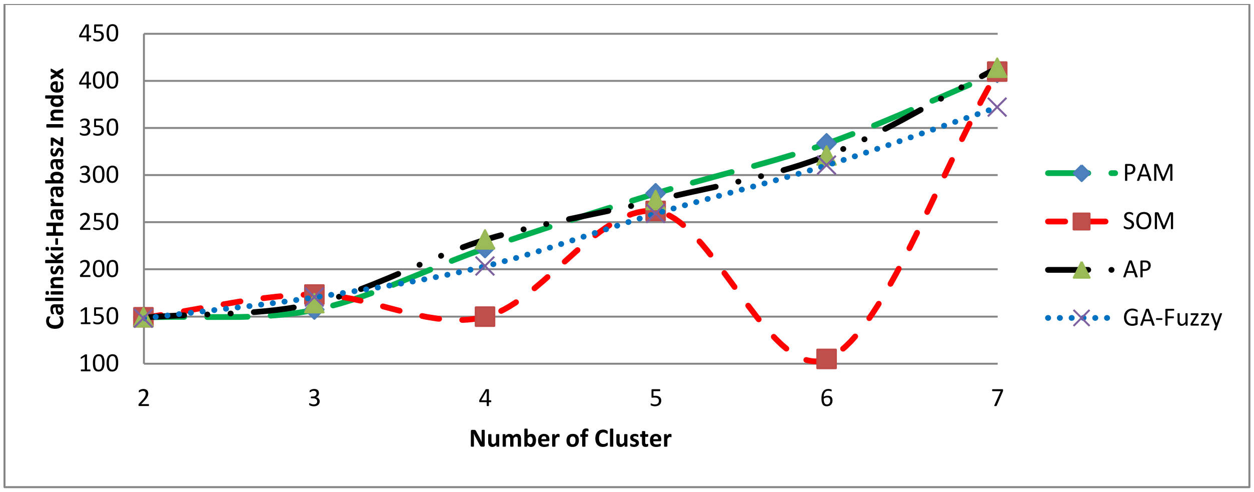


Figure 6.18 Calinski-Harabasz Index vs. Number of cluster Plot for different clustering algorithms

Dunn Index vs. Number of cluster for four different algorithms is illustrated in Figure 6.19. The last two indices showed highest value for AP algorithm for any number of clusters but in this case AP algorithm is having highest value for 2 to 5 numbers of clusters. In this study 4 was found as optimal number of cluster and AP is having highest value at that point so going with previous parameters Dunn Index also selected AP as the best algorithm for this study.

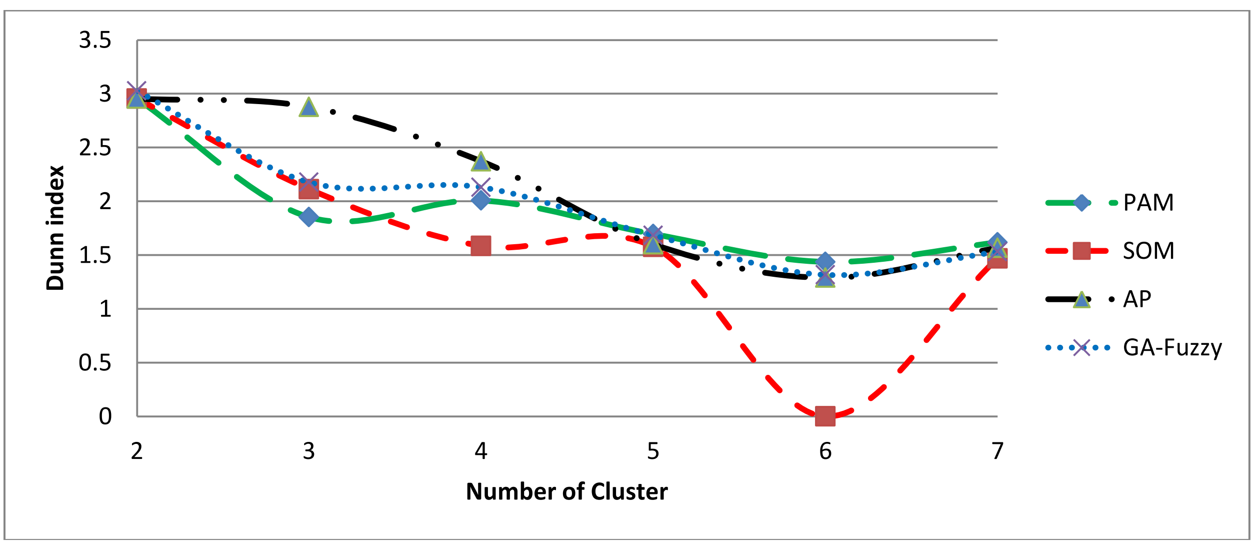


Figure 6.19 Dunn Index vs. Number of cluster Plot for different clustering algorithms

The same R-square Index (RSI) that was used previously for determining the optimal number of cluster for the input data set is again used to decide the best clustering algorithm. The index value of RSI is calculated for 4-7 numbers of clusters for all the four algorithms used in this study. The index values for PAM, AP and GA-Fuzzy algorithm are found to be all most same but AP having slightly higher value. So going in hand with three previous parameters RSI also selected AP as the most appropriate clustering algorithm for clustering of speed data for this study.

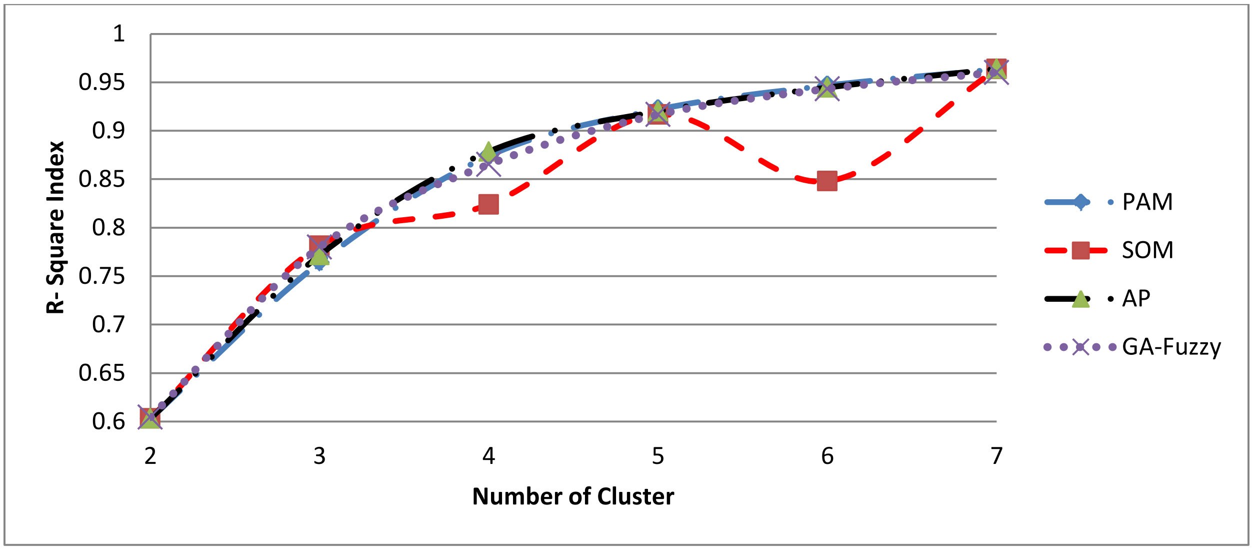


Figure 6.20 R-square Index vs. Number of cluster Plot for different clustering algorithms

6.4 Summary

The speed data collected using GPS receiver were clustered using four clustering algorithms namely SOM, AP, GA-Fuzzy and AP algorithms. Various cluster validation measures were applied to get the optimal number of cluster for all the four algorithms applied. Cluster validation which holds significance in deciding the number of urban street classes the FFS data should be clustered into. These four algorithms were applied twice in each case. First FFS data were clustered to get the FFS ranges of different urban street classes and secondly the clustering algorithm is used for travel speed data to get the speed ranges of different LOS categories. After getting the cluster result the speed ranges were given in tabular form for all the four clustering algorithm. Four cluster quality evaluation parameters i.e. Silhouette width, Calinski-Harabasz Index, Dunn Index and R-square Index were used to determine the best clustering algorithm for this study.

The next chapter elaborates on summary, conclusion, limitation and future work possible.

Chapter 7

Summary, Conclusions and Future Scope

7.1 Summary

In this study, various limitations in current HCM-2000 methodology for defining LOS criteria was found out and an attempt has been made to develop methodologies to define LOS criteria for urban streets in Indian context. GPS was used to collect the speed and inventory data and GIS was used handle these data. Applications of GIS and GPS for traffic data collection were reviewed from literature. The concept of urban street classification based on free-flow speeds, function and geometric characteristics of street segments are presented. Also important influencing factors that affect level of service classifications of urban streets are enumerated.

From literature it was found that cluster analysis is the suitable technique that can be applied for the classification of urban streets and level of service categories. SOM, AP, GA Fuzzy, PAM were used as a tool to cluster the speed data to classify the road segments into various classes and also to define the speed ranges of the LOS. By using above four algorithms FFS speed were clustered into four different groups corresponding to different urban street classes. Secondly, clustering methods were applied on average travel speeds on street segments of each class of urban street during peak and off peak hours. In the latter case, speeds were classified into six categories for six levels of service; thus speed ranges for level of service categories were defined for Indian conditions. Number of cluster into which the data set should be clustered into is given as priori to the clustering algorithm. To determine the optimum number of cluster various cluster validation parameters were used as not a single parameter is self sufficient. In this study as a whole 9 validation parameters were used i.e. C-Index, Weighted inter-intra index (WI), Hartigan index (HI), R-squared Index (RSI) , Krzanowski-Lai Index (KLI), Davies-Bouldin Index (DBI), Silhouette width, Calinski-Harabasz Index and Dunn Index. Out of these 9 parameters first 6 parameters were used to get optimal number of cluster for the input data set and last 3 parameters were used to determine the best clustering algorithm for this study. R-square Index was used for both the purpose.

7.2 Conclusion

Following conclusions are derived from the present study in defining the Level of Service (LOS) in Indian context:

- Various cluster validation measures, based on their applicability is used to find the optimal number of clusters for SOM, AP, GA-Fuzzy and PAM algorithm. After thorough analysis it was decided to classify urban street into four classes (I-IV) in Indian context. Free flow speed ranges for different urban street classes were found out and for each algorithm the range was found to be different. The speed ranges were lower than that mentioned in HCM-2000. Heterogeneous traffic flow and roads having varying geometric and surrounding environmental characteristics are the major reasons for these lower values in FFSs.

- After determining the FFS ranges of different urban street classes speed ranges of LOS categories were also found using the four different clustering algorithms. These speed ranges resulted from different clustering algorithm were found to be significantly different from each other. In order to get the most suitable clustering algorithm in defining LOS criteria a thorough study of four clustering quality evaluation parameters was carried out. All the four parameters showed AP to be most suitable clustering algorithm for this study so AP was selected as the clustering method in defining LOS criteria in Indian context. Speed ranges obtained from AP clustering is decided to be most relevant in Indian context.

The following LOS criteria for urban streets in Indian context are suggested:

Urban Street Class	**I**	**II**	**III**	**IV**
Range of free-flow speed (FFS)	90 to 70 km/h	70 to 55 km/h	55 to 40 km/h	40 to 25 km/h
Typical FFS	75km/h	60km/h	47km/h	35 km/h
LOS	**Average Travel Speed (Km/h)**			
A	>60	>52	>45	>28
B	>50-60	>40-52	>35-45	>20-28
C	>42-50	>32-40	>25-35	>14-20
D	>33-42	>25-32	>17-25	>10-14
E	>22-33	>17-25	>12-17	>6-10
F	>22	>17	>12	>6

- These speed ranges of different LOS categories found to be proportionately lower than those suggested in HCM-2000. Unwanted movement of pedestrians along and across the road sections produces undesirable side friction that constrains travelers to reduce vehicular speed. Road side vendors and on-street parking occupy substantial portion of road sections which compels the commuter to move through a narrower space can be inferred as the reasons behind these lower values.
- The speed ranges for level of service categories (A-F) expressed in percentage of free-flow speeds were found to be approximately 80, 62, 48, 35, 23 and 18-23 respectively in the present study. Whereas, in HCM (2000) it has been mentioned these values are 90, 70, 50, 40, 33 and 25-33 percent respectively.
- From clustering results it can be seen less number of roads in Mumbai are of high speed design (street class-I) or highly congested (street class-IV). More number of road segments is of suburban (street class II) or intermediate (street class III) type so it can be suggested that Greater Mumbai region needs substantial geometric improvements to mitigate the burden on urban road infrastructure because of ever growing vehicular traffic volume.
- . From this study the applicability of GPS in collection of speed data with high precision in short time is established. So this tool can be exploited by other developing and developed countries to collect speed data and cluster analysis can be applied to define the speed ranges of LOS categories of their own rather than following some values which are not completely appropriate for the local condition.

7.3 Limitations and Future Scope

There are some limitations in this study and opportunity lies in future studies to eliminate these limitations.

- This study is done for the city of Mumbai. Similar study can be carried out in other cities of India, as India having significant diversities among its people and their driving characteristics.
- For this research only midsized vehicle is used for data collection purpose. All though midsized vehicle has significance presence in urban roads of India and data collection using these vehicles is convenient and easy but to get complete picture of heterogeneous traffic flow further study can be done using more numbers of modes.
- In defining LOS criteria user perception should be given due consideration. Along with quantitative analysis qualitative analysis from stated preference survey needed to be given due consideration. The relation between the qualitative and quantitative study need to be established.

References

Arasan, V.T., Vedagiri,P. Study of the impact of exclusive bus lane under highly heterogeneous traffic condition. *Public Transport* ,2010, Vol. 2(1),p. 135-15.

Baumgartner, W.E. Level of service: Getting ready for the 21st century. *ITE Journal (Institute of Transportation Engineers),* 1996, Vol. 66 (1), p. 36-39

Bensaid, A.M., Hall, L.O., Bezdek, J.C., Clarke, L.P., Silbiger, M.L., Arrington, J.A., Murtagh, R.F. Validity guided (re) clustering with applications to image segmentation. *IEEE Transactions on Fuzzy Systems*, 1996, Vol. 4(2), p. 112-123.

Bezdek, J. C. *Pattern recognition with fuzzy objective function algorithm*, Plenum, New York, 1981.

Bezdek, J.C., Pal, N.R. Some new indexes of cluster validity. *IEEE Transactions on System, Man, and Cybernetics-Part B: Cybernetics*, 1998, Vol. 28(3), p.301-315.

Buyan, P.K. Defining Level of Service criteria for urban streets in Indian context. Published Ph.D. Dissertation, IIT Bombay, Mumbai, India, 2009.

Bhuyan, P.K., Rao K.V.K. FCM Clustering Using GPS Data for Defining Level of Service Criteria of Urban Streets. *Transport Problems: an International Scientific Journal,* Poland, 2010, Vol. 5(4), p. 105-113.

Bolshakova, N., Azuaje, F. Estimating the number of clusters in DNA microarray data. *Methods of Information in Medicine*, 2006, Vol. 45 (2), p. 153-157

Cao, S.H., Yuan, Z.Z., Zhang, C.Q., Zhao, L. LOS Classification for Urban Rail Transit Passages Based on Passenger Perceptions. *Journal of Transportation Systems Engineering and Information Technology, 2009*, Vol. 9 (2), p. 99-104.

Cesar, A., Bullock, D. Travel time studies with global positioning and geographic information systems: An integrated methodology. *Transportation Research Part C*, 1998,
Vol. 32 (6), p.101-127.

Cetiner, B,G., Sari,M., Borat, O. A neural network based traffic-flow model, *Mathematical and Computational Applications*, 2010 Vol. 15, No. 2, p. 269-278.

Clark, I. Level of Service F: Is it really bad as it gets? In Proc.,*IPENZ Transportation Group Conference*, New Plymouth, November, 2008.

Dimitriadou, E., Dolnicar, S., Weingessel, A., An examination of indexes for determining the Number of Cluster in binary data sets. *Psychometrika*, 2002, Vol. 67(1), p.137-160.

Dudoit, S., Fridlyand, J. A prediction-based resampling method for estimating the number of clusters in a dataset.*Genome biology*,2002, Vol. 3 (7), p. 36.

Fang, F.C., Pecheux, K.K. Fuzzy data mining approach for quantifying signalized intersection level of services based on user perceptions. Journal of Transportation Engineering, 2009, Vol. 135 (6), 349-358.

Flannery, A., Rouphail, N., Reinke, D. Analysis and Modeling of Automobile Users' Perceptions of Quality of Service on Urban Streets. *Transportation Research Record*, 2071, Transportation Research Board, Washington, D.C., 2008, ASCE ,p. 26-34.

Flannery, A., Wochinger, K., Martin, A. Driver assessment of service quality on urban streets. *Transportation Research Record*, 1920 Transportation Research Board, 2005, Washington, D.C., p. 25–31.

Frey, B.J., Deuck, D. Clustering by Passing Messages between Data Points. *Science*, 2007, Vol. 315(5814), p. 972-997.

Gallagher, J. Travel time data collection using GPS. In Proc., *Proceedings of National Traffic Data Acquisition Conference*, New Mexico, 1996, p.148-161.

Al-garni, S., Abdennour, A. A Neural Network Based Traffic Flow Evaluation System for Highways, *Journal of King Saud University of Engineering Science*s, 2008 Vol. 20 (1), p. 37-46.

Highway Capacity Manual. *Transportation Research Board*, 1950, Washington, D.C.

Highway Capacity Manual. *Transportation Research Board*, 1965, Washington, D.C.

Highway Capacity Manual. *Transportation Research Board*, 1985, Washington, D.C.

Highway Capacity Manual. *Transportation Research Board*, 2000, Washington, D.C.

Halkidi, M., Batistakis, Y., Vazirgiannis, M. Cluster validity methods: Part I and II. *SIGMOD Record*, 2002, Vol. 31 (2) , p. 40-45

Hubert, L., Schultz, J. Quadratic assignment as a general data-analysis strategy. *British Journal of Mathematical and Statistical Psychologie*, 1976, p.190-241.

IRC. *Guidelines for capacity of urban roads in plain areas*, IRC: 106, 1990 New Delhi.

Ivana,C., Zvonko, K., Marjana, P. Hybrid approach for urban roads classification based on GPS tracks and road subsegments data. *Promet-Traffic & Transportation*,2011, Vol. 23(4), p. 289-296.

Kasturi, J., Acharya, R., Ramanathan, M. An information theoretic approach for analyzing temporal patterns of gene expression. *Bioinformatics*, 2003, Vol. 19 (4), p. 449-458.

Kaufman, L., Rousseeuw, P.J. *Finding groups in data: An introduction to cluster analysis*, Wiley, New York, 1990.

Kidwai, F.A., Marwah, B.R., Deb, K., Karim, M.R., A genetic algorithm based bus scheduling model for transit network. In Proc., *Eastern Asia Society for Transportation Studies*, 2005, Vol. 5, p. 477 – 489.

Kita, H., Fujiwara, E. Reconsideration on the level of service and a proposed measure. In Proc., *15th Annual Meeting of JSTE*, Japanese, 1995, p. 25–28.

Kittelson, W.K., Roess, R.P. Highway capacity analysis after the highway capacity manual 2000. *Transportation Research Record*, 1776, Transportation Research Board, Washington, *D.C.* 2001, p. 10–16.

Krzanowski W, Lai Y. A criterion for determining the number of groups in a dataset using sum of squares clustering. *Biometrics***,** 1985, Vol. 44, p.23-34.

Maitra, B., Sikdar, P.K., Dhingra, S.L. Modelling congestion on urban roads and assessing level of service. *Journal of Transportation Engineering*, 1999, Vol. 125 (6), ASCE, p. 508-514.

Marwah, B.R., Singh, B. Level of service classification for urban heterogeneous traffic: A case study of Kanpur metropolis. In Proc., *Fourth International Symposium on Highway Capacity*, Hawaii, June-July, 2000, p. 271-286.

Murat S. Y., Baskan O. Modeling vehicle delay at signalized junction: Artificial Neural Network approach, *Journal of Scientific and Industrial Research,* 2006, Vol. 65, p.558-564.

Ndoh, N.N., Ashford, N.J. Evaluation of transportation level of service using fuzzy sets. In Proc., *73rdAnnual Meeting of TRB*, Transportation Research Board, Washington, D.C., 2004.

Shao, M., Sun, L. United evaluation model of traffic operation level for different types of urban road. Journal *of Tongji University*, 2010, Vol. 38 (11) ,p. 1593-1598.

Spring, G. S. Integration of safety and the highway capacity manual. In *Proc., 4th International Symposium on Highway Capacity, Transportation Research Board, Washington, D.C*., 1999, p. 63–72.

Refianti, R., Mutira, A.B., Juarna, A., Ikhsan, S.N. Analysis and implementation of algorithm clustering Affinity Propagation and K-means at data student based on GPA and duration of bachelor-Thesis Completion. *Journal of Theoretical and Applied Information Technology,* 2012, Vol. 35(1), p.69-76.

Shao, M., Sun, L. United evaluation model of traffic operation level for different types of urban road. Journal *of Tongji University,* 2010, Vol. 38 (11) ,p. 1593-1598.

Sharma, S.C. *Applied Multivariate Techniques.* John Wiley and Sons. 1996.

Spring, G. S. Integration of safety and the highway capacity manual. In *Proc., 4th International Symposium on Highway Capacity, Transportation Research Board, Washington, D.C.*, 1999, p. 63–72.

Strehl, A., Ghosh, J. Relationship-based clustering and visualization for high-dimensional data mining. *INFORMS Journal on Computing*, 2003, Vol. 15, p. 208-230.

Taylor, M.A.P., Woolley, J. E., and Zito, R. Integration of the global positioning system and geographical systems for traffic congestion studies. *Transportation Research Part-C*, 2000, Vol. 8, 257–285.

Tan, D., Wang, W., Lu, J., Bian, Y. Research on Methods of Assessing Pedestrian Level of Service for Sidewalk. *Journal of Transportation Systems Engineering and Information Technology*, 2007, Vol. 7 (5),p. 74-79.

Turner,S.M., Eisele,W.L., Benz, R.J., Holdener, D. J. *Travel time data collection handbook*, Texas Transportation Institute, The Texas A&M Univ. System, College Station, Texas, 1998..

Wang, K., Wang, B., Peng, L. CVAP: Validation for cluster analyses. *Data Science Journal*, 2009, Vol. 8 , p. 88-93

Wei, C.Y.H., Chang, C.C., Wang, S.S Vehicle classification using advanced technologies. *Transportation Research Record*, 1551, TRB, Washington, D.C., 1996, p. 45–50.

Wei, C., Tingjin, L., Jizheng, W., Yanqing, Z., An Improved Genetic FCM Clustering Algorithm. In Proc., *2nd International Conference on Future Computer Communication (ICFCC)*, Wuhan, China, 2010, Vol. 1, p. 45-48

Xia, D., Wu, F., Zhang, X., Zhuang, Y. Local and global approaches of affinity propagation clustering for large scale data. *Journal of Zhejiang University*, 2008, Vol. 9(10), p. 1373-1381.

Yang, C., Bruzzone, L. A Fuzzy-Statistics-based Affinity Propagation technique for clustering in multispectral images. *IEEE transactions on geosciences and remote sensing*, 2010, Vol. 48(6), p. 2647-2659.

Yang, H., Giao, F. Neural network approach to classification of traffic flow states, *Journal of Transportation Engineering*,1998, Vol. 124(6), p.521–525.

Zhang, X., Wu, F., Zhuang, Y. Clustering by evidence accumulation on affinity propagation in Proc., *IEEE International Conference on Pattern Recognition*, 2008, p. 1 – 4.

Zhang, B., Xing, K., Cheng, X., Huang, L., Bie, R. Traffic clustering and online traffic prediction in vehicle networks: A Social Influence Perspective in Proc., *IEEE INFOCOM*, Orlando, Florida USA, March 25-30, 2012.

Appendix-I

The table illustrates FFS and Average travel speed data of 15 segments belonging Corridor-3. Both these data collection procedure is elaborated in Chapter-4.

Table AI. FFS and Average travel speed during peak and off peak hours of Corridor-3

Corridor-3					
Segment No.	Average Free-Flow Speed (km/hr)	Duration and Direction of Travel			
		M-N-S	M-S-N	E-N-S	E-S-N
		Average Travel Speed (km/hr)			
1	71.14	28.64	28.63	30.5	37.33
2	61.56	19.24	30.08	24.51	27.28
3	65.91	40.41	45.19	44.94	40.43
4	43.19	38.17	47.06	41.68	37.29
5	77.36	37.04	56.11	50.39	44.61
6	57.92	16.88	27.14	17.35	38.27
7	58.83	28.74	32.13	20.68	24.65
8	44.27	20.17	27.01	25.47	47.95
9	54.27	27.39	33.21	27.93	20.15
10	54.39	15.50	30.10	12.87	15.86
11	56.68	21.80	19.22	44.39	16.63
12	43.00	22.50	25.99	26.39	22.21
13	24.94	25.91	22.84	11.31	8.85
14	31.38	13.21	31.49	31.50	23.66
15	29.62	17.67	15.27	24.40	8.25

List of Publications

Journals:

Submitted

1. Mohapatra, S.S., Bhuyan, P.K. (2012) “Self Organizing Map (SOM) of Artificial Neural Network (ANN) for defining Level of Service criteria of urban streets.” *International Journal of Traffic and Transport Engineering (IJTTE).*

2. Mohapatra, S.S., Bhuyan, P.K. (2012) “Affinity Propagation Clustering In defining Level of Service criteria of urban Streets.” *Transport.*

3. Mohapatra, S.S., Bhuyan, P.K., Rao, K.V.K. (2012) “Genetic Algorithm Fuzzy Clustering using GPS data for defining Level of Service criteria of urban Streets.” *Journal of European Transport.*