Genetic Algorithm Fuzzy Clustering using GPS data for Defining Level of Service Criteria of Urban Streets

Smruti Sourava Mohapatra[1], P.K Bhuyan[2*], K.V.Krishna Rao[3]

[1] *M. Tech. Student, Department of Civil Engineering, National Institute of Technology Rourkela, India 769008*

[2]*Assistant Professor, Department of Civil Engineering, National Institute of Technology Rourkela, India 769008*

[3]*Professor, Civil Engineering Department, IIT Bombay, India 400076*

Abstract

Developing countries like India need to have proper Level of Service (LOS) criteria for various traffic facilities as this helps in planning, design of transportation projects and also allocating resources to the competing projects. The LOS analysis for urban street followed in India is an adaptation of HCM-2000 methodology but the methodology is relevant for developed countries having homogenous traffic flow. In this research an attempt has been made to establish a framework to define LOS criteria of urban street in Indian context keeping in mind the geometric and surrounding environmental characteristics. Defining LOS criteria is basically a classification problem for which cluster analysis is a suitable technique can be applied. In this research a hybrid algorithm comprising of Genetic Algorithm (GA) and Fuzzy C-mean is utilized. As input to the clustering algorithm GA-Fuzzy a lot of speed data is required. From literature review GPS is found to be a suitable tool for collecting second by second speed data and GIS is suitable in handling large amount of speed data. The clustering algorithm is used twice in this study. First the GA-Fuzzy algorithm was used to classify Free Flow Speed (FFS) data into number of classes in order to get the FFS ranges of different urban street classes. To determine the optimal number of cluster using FFS data five cluster validation parameters are considered. After getting the FFS ranges for different urban street classes the same GA-Fuzzy algorithm is used on average travel speed data collected during both peak and off-peak hours to determine the speed ranges of different LOS categories. From this analysis the free flow speed ranges for different urban street classes and the speed ranges for different LOS categories are defined and the values are found to be lower than that suggested by HCM-2000. The coherence of the clustering result in classification of urban streets into four classes and speed values into six LOS categories is agreed with the physical and surrounding environmental characteristics of road segments under the study area. From this analysis it is also found that good LOS can't be expected from urban street segment for which physical and surrounding environmental characteristics are not good.

Keywords: Urban Streets, Level of Service, GPS, Genetic Algorithm (GA), Fuzzy C-mean clustering, Cluster Validation

---

1. Introduction

* Corresponding author: P. K. Bhuyan (pkbtrans@gmail.com)

In developing country like India urban infrastructure are not spared from the ill effect of urbanization because of burgeoning economic advancement. As a result of which urban streets are highly congested and commuter are exposed to very poor operating condition. To improve the quality of traffic flow in this prevailing condition first of all the operating condition should be evaluated.

As this will affects the planning, design, operational aspects of transportation projects and also it influences the allocation of limited financial resources among competing transportation projects. The free flow speed ranges for urban street categories and speed ranges of level of service categories that are specified in HCM (2000) have been followed in India for level of service analysis of urban streets. In fact speed ranges mentioned in HCM (2000) are suitable for developed countries having homogenous traffic flow. In developing countries like India traffic on roads are highly heterogeneous comprising vehicles with wide ranging static and dynamic characteristics such as vehicle size, engine power, acceleration/deceleration, manoeuvring capabilities, etc. Typically traffic in developed countries follows lane discipline and physical dimensions of vehicles do not vary much. Whereas, twenty two types of vehicles having wide variation in physical size travels on Indian roads. As a result of which vehicular travel speed is comparatively less under heterogeneous flow condition.

As Level of Service (LOS) is not well defined for highly heterogeneous traffic flow on urban streets in India, an attempt has been made to define LOS criteria in this study. Average travel speed of through-vehicles for segments under the each street corridor is the basis of defining LOS of urban streets. Traditionally, for collecting travel time data probe vehicle is used but this method is quite susceptible to human error. Recent research has demonstrated the feasibility of using Global Positioning System (GPS) receiver in recording location as latitude-longitude, travel time and travel speed. GPS receiver records location and speed automatically at regular sampling periods. This method of data collection eliminates human error and also quite efficient in handling large amount of speed data. In this study the advantage of using GPS in speed data collection was utilized.

Defining LOS is basically a classification problem. From literature review it was found that cluster analysis is the most suitable technique for the classification of the large amount of speed data acquired thorough GPS receiver. In this study Genetic Algorithm Fuzzy (GA-Fuzzy) set theory is used for the clustering purpose. The clustering algorithm was used twice in this research. First, GA-Fuzzy clustering was used on Free Flow Speed (FFS) data to get speed ranges of urban street classes. After defining the speed ranges GA-Fuzzy algorithm was used for the second time on average travel speed data to get the speed ranges of different LOS categories. To get the optimal number of cluster using FFS data five different Validation parameters were used. The coherence of the clustering result for the classification of urban streets and LOS categories were verified with geometric and surrounding environmental characteristics of street segments. The overall framework of this study is shown in the Figure 1.

2. Literature Review

The concept of "practical capacity" presented in the 1950 HCM (TRB, 1950) gave rise to the concept of LOS. In the dynamicity of time the concept has been modified and made more practical in the 1965 HCM (TRB, 1965) and 1985 HCM (TRB, 1985). The current definition of LOS being followed is that defined in 2000 HCM (TRB, 2000). Level of service in the Highway Capacity Manual (HCM 2000) defined as ''a quality measure describing operational conditions within a traffic stream, generally in terms of service measures such as speed and travel time, freedom to maneuver, traffic interruptions, comfort and convenience.''. The current HCM methodologies is resulted from a combination of consulting studies, research and the discussions of Highway Capacity and Quality of Service committee (Pecheux et al. 2000). The HCM 2000 also designates six levels of service for each type of facility, from "A" to "F," with LOS "A" representing the best operating conditions and LOS "F" the worst.

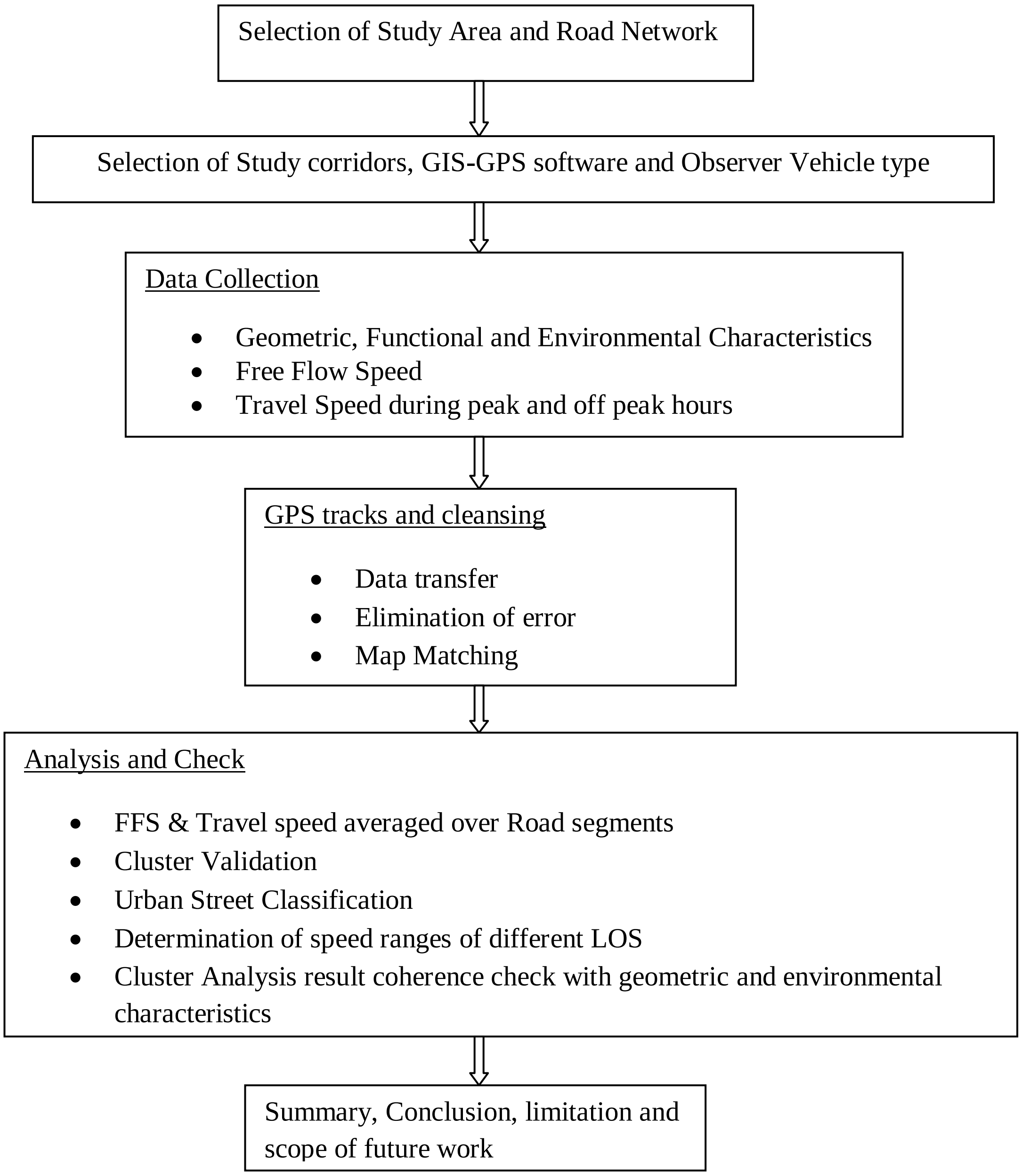


Figure-1: Overall framework of the study

In recent years, several alternatives have been suggested for improving the scaling and determination of LOS. For example, the study of Baumgaertner (1996) 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. Thus, travel conditions that would have been viewed as intolerable in the 1960s are considered normal by today's motorists, especially commuters. Kita and Fujiwara (1995) stated LOS not just as a traffic operating condition and tried to find the relationship of LOS with driver's perception. 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. Marwah and Singh (2000) have attempted in providing classification of level of service categories for heterogeneous traffic condition in Indian cities. Based on their simulation results of benchmark roads and traffic composition the level of service were classified into four groups (LOS I-IV). In an another study for heterogeneous traffic flow condition on urban roads Maitra et al. (1999) considered congestion as measure of effectiveness for LOS and redefined LOS boundaries. Based on quantified congestion, LOS boundaries have been categorized into nine groups "A" to "I" in the stable zone and one LOS "J" for the unstable operation.

Bhuyan and Rao (2011) defined the free flow speed ranges of urban street classes and speed limits of LOS categories using Hierarchical Agglomerative Clustering (HAC) and data collected by GPS handheld receiver in Indian context. Maitra, B. et. al. (2004) have shown the effect of different types of vehicles on congestion trough congestion model for developing countries. The established model can be used as a tool for formulating traffic management measures for urban roads. Basu, et. al. (2006) modeled passenger car equivalency for urban mid-block using stream speed as measure of equivalence. In this study a neural network approach was explored to capture the effects of traffic volume and its composition level on the stream speed. Clark (2008) from his study upon New Zealand traffic raised question mark about the LOS "F". The author suggested for a new LOS to be termed as F+ or G 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. (2003) 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. Ivana 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. Kikuchi and Chakroborty (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.

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 and number of passengers records the elapsed time information at predefined check points. This method is very much susceptible to human error (Turner, et. al. 1998). With improvement computers Distance Measuring Instrument (DMI) came as the solution of Floating car method for measurement of the speed distance using pulses from a sensor attached to the test vehicle’s transmission (Quiroga and Bullock, 1998). Turner et. al. (1998) found limitation in this method relating to installation of the DMI unit and data storage problems. 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 and become one of the alternatives to moving car observer method for field data collection. So in this study GPS receiver was used extensively for the purpose of data collection. Basically three types of data were collected in this; they are free flow speed, average travel speed during peak and off peak hours and data related to geometric characteristics of the road.

It was convinced from literature review that LOS criteria that are described in HCM 2000 have many pit falls upon which various researcher suggested different improvement measures. In Indian context no methodology has been developed in regards to define LOS criteria. Urban street LOS is a function of travel speed and to define LOS a large amount of speed data are required. Advancement of information technology made GIS-GPS as an efficient method to acquire and handle large amount of speed data. Defining LOS is basically a classification problem for which cluster analysis was found to be a suitable technique. The advantage of using GA-Fuzzy clustering for this classification problem is that the algorithm utilizes the optimization process of Genetic Algorithm in Fuzzy c-mean clustering.

3. Cluster Analysis

To get a close optimal solution in this research a hybrid algorithm based on fuzzy c-means in association with genetic algorithm (GA) is used. This hybrid algorithm has global search of GA and local search capability of FCM so it can solve the clustering problem in a more efficient way.

3.1 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 ssatisfying the following conditions

$$F = [\mu_{ki}], \mu_{ki} \in [0,1], k = \overline{1,M}, i = \overline{1,C} \quad (1)$$

$$\sum_{i=\overline{1,C}} \mu_{ki} = 1, k = \overline{1,M} \quad (2)$$

$$0 < \sum_{k=\overline{1,M}} \mu_{ki} < N, i = \overline{1,C} \quad (3)$$

Step 3. Calculate the centers of clusters: $V_i = \dfrac{\sum_{k=\overline{1,N}} (\mu_{ki})^m \cdot X_k}{\sum_{k=\overline{1,N}} (\mu_{ki})^m}, i = \overline{1,c}$ (4)

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

$$D_{ki} = \sqrt{\left\| X_k - V_i \right\|^2}, k = \overline{1,M}, i = \overline{1,c} \quad (5)$$

Here X is the observation matrix

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

$$\text{If } D_{ki} > 0: \quad \mu_{ki} = \frac{1}{\left(D_{ik}{}^2 \cdot \sum_{j=1,c} \frac{1}{D_{jk}{}^2}\right)^{1/(m-1)}} \quad (6)$$

$$\text{If } D_{ki} = 0: \quad \mu_{kj} = \begin{cases} 1, j = i \\ 0, j \neq i, j = \overline{1,c} \end{cases} \quad (7)$$

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.

3.2 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 $x_j = (x_{j1}, x_{j2}, ..., x_{jn}) \frac{\pi}{3}$ and the center of cluster $v_i = (v_{k1}, v_{k2}, ..., v_{kn}), m \in (1, \infty)$ 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.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

3.3 Validation measure

Quality of clustering results can be evaluated using clustering algorithm. It has been mainly used to evaluate and compare whole partitions, resulting from different algorithms or resulting from the same algorithms under different parameters. Validation measure is used to determine the optimal number of cluster the data set is to be clustered into (Bensaid et al., 1996). Different validity measures have been proposed in the literature, none of them is perfect by oneself, and therefore several indices are used in this study, such as: C-index, Weighted inter-intra index, Hartigan index, R-squared Index, and Krzanowski-Lai Index.

A) C-index

The C-index is defined as (Hubert and Schultz 1976):

$$\frac{s - s_{\min}}{s_{\max} - s_{\min}} \tag{8}$$

Where S is the sum of distances over all pairs of objects form the same cluster, n is the number of those pairs and S_{min} is the sum of the n smallest distances if all pairs of objects are considered. Likewise S_{max} is the sum of the n largest distances out of all pairs. Hence a small value of C indicates a good clustering.

B) Weighted inter-intra index

This index tries to find the optimal number of cluster by high overall quality of cluster $\phi^{(Q)}$ and a small number of cluster *k*. The quality of cluster is determined as follow (Strehl and Ghosh, 2002):

$$\phi^{(Q)}(X,\lambda)=1-\frac{\sum_{i=1}^{n_i}\frac{n_i}{n-n_i}\sum_{j\in\{1,\dots i-1,i+1\dots k\}} n_j\times \text{inter}(X,\lambda,i,j)}{\sum_{i=1}^{k} n_i\times \text{intra}(X,\lambda,i)} \tag{9}$$

Where

$$\text{inter}(X,\lambda,i,j)=\frac{1}{n_i \text{X} n_j}\sum_{\lambda_a=i,\lambda_b=j} S(X_s,X_b) \tag{10}$$

And

$$\text{intra}(X,\lambda,j)=\frac{2}{(n_i-1)\text{X}n_i}\sum_{\lambda_a=\lambda_b=i,b>a} S(X_a,X_b) \tag{11}$$

Here *I and j* are cluster indices. X_a *and* X_b are two vertices.

C) Hargitan index

For a clustering result with *k* clusters, the overall fitness for the clustering can be expressed as the square of the error for all samples :

$$err(k)=\sum_{i=1}^{k}\sum_{j=1,j\in C_i}^{N} d^2(X_{j,}X_{c_i}) \tag{12}$$

Where N is the number of data points, where X_i is a M-component vector, representing M features for sample I . d is the distance between data sample X_j and the centre X_{C_i} .
Then, *Hartigan* index *H(k),* for *k* partitioning, is expressed as follow:

$$H(k)=(n-k-1)\frac{err(k)-err(k+1)}{err(k+1)} \tag{13}$$

D) R-squared Index

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

$$\text{RS}=\frac{SS_t-SS_w}{SS_t} \tag{14}$$

Where

$$S_t = \sum_{j=1}^{d} \sum_{k=1}^{n_j} (x_k - \overline{x_j})^2 \quad \text{and} \quad SS_w = \sum_{\substack{i=1...n_c \\ j=1...d}} \sum_{k=1}^{n_{ij}} (x_k - \overline{x_j}) \tag{15}$$

Here d is number of clusters,n_c is optimum number of cluster n_j is Number of element in j^{th} dimension in the whole data set, $\bar{x}_j$ is expected value in j^{th} dimension, n_{ij} is the Number of elements in i^{th} cluster j^{th} dimension.

E) Krzanowski-Lai (KL) Index

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

$$KL(k) = \left| \frac{DIFF(k)}{DIFF(k+1)} \right| \tag{16}$$

Where

$$DIFF(k) = (k-1)^{2/p} W(k-1) - k^{2/p} W(k) \tag{17}$$

and p denotes the number of features in the data set, k denotes the number of clusters, *W (k)* denotes the within cluster sum of square of the partition.

4. Study Corridor and Data Collection Techniques

4.1 Study Corridor

For this study the commercial capital of India Mumbai City was considered. The city has linear pattern of transport network having predominant North-South commuter movements. South Mumbai houses various work place so during morning time people move towards South for work and during evening hours they returns towards north to their homes in the Suburbs of Mumbai. So In this study five major corridors were chosen out of which four are north-south corridors and one is east-west corridor. The North-South corridors are 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 the only East-West corridor is Versova- Andheri- Ghatkopar- Vashi (VAGV) (Corridor-5). These five corridors are overlapped on the GIS base map of Greater Mumbai are shown in Fig.2.

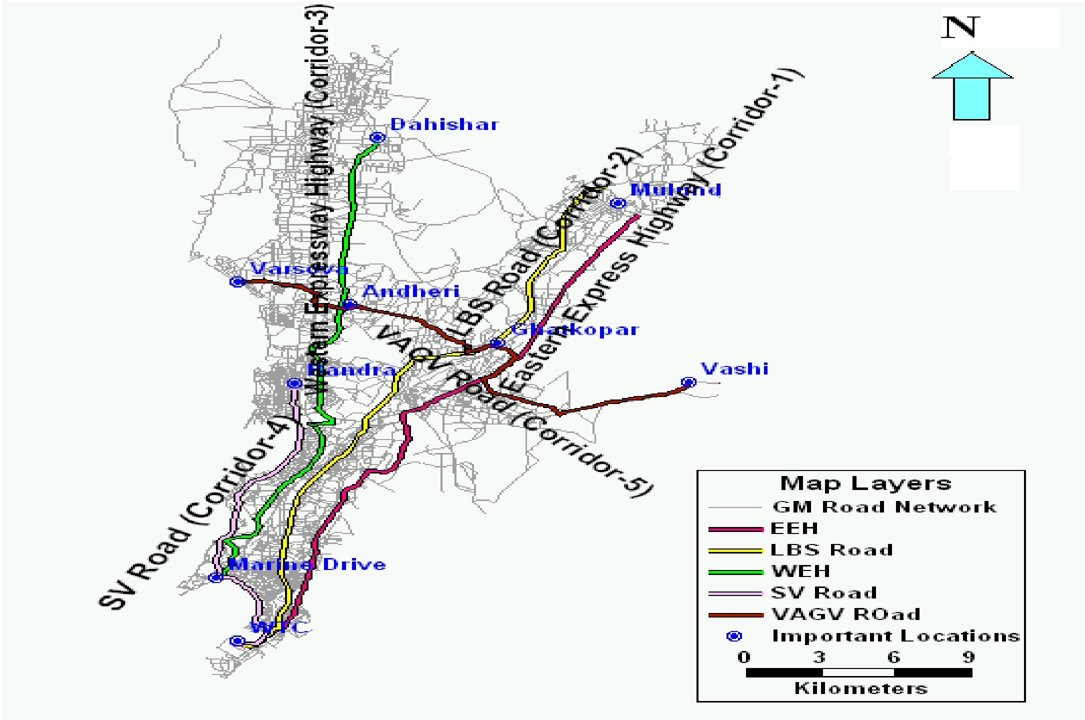


Fig.2 Map Showing Selected Corridors of Greater Mumbai

4.2 Data Collection

The probe vehicle used in this research work is mid-sized cars. This vehicle was fitted with Trimble Geo-XT GPS receiver, so that it was adjusted to log 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 can be derived, including travel time, stopped time, travel speeds (instantaneous and average), and various congestion indices. In order to get unbiased data sets we used three mid-sized cars and took the help of three drivers on different days of the survey work.

The first type is roadway inventory details. In this survey Details on segments like segment number, number of lanes, median types, pedestrian activity, road side development, access density, construction activity, speed limit, separate right turn lane, number of flyovers, date and day of data collection and segment length were collected. During the collection of inventory details proper segmentation technique was applied, which is the directional stretch of road section immediately after signalized intersections to the location point immediately after the next signal.

The second type of survey conducted was to find free flow speed. Before going for the free flow speed data collection, the duration during which the traffic volume is less than or equal to 200 vehicles per lane per hour should be known. For that a detailed 24 hour traffic volume count survey was conducted prior to collection of FFS .The traffic volume data were collected on 45 stations on seven screen lines. From survey data traffic volume per lane per hour were calculated for roads comings under the study area. It was found that free flow traffic condition (less than 200veh/ln/hr) is approaching at 12 mid-nights and all road sections are having free flow traffic conditions from 1 AM to 5 AM. Hence free flow speed for all these roads were collected using GPS receiver fitted on a probe vehicle during these hours. 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.

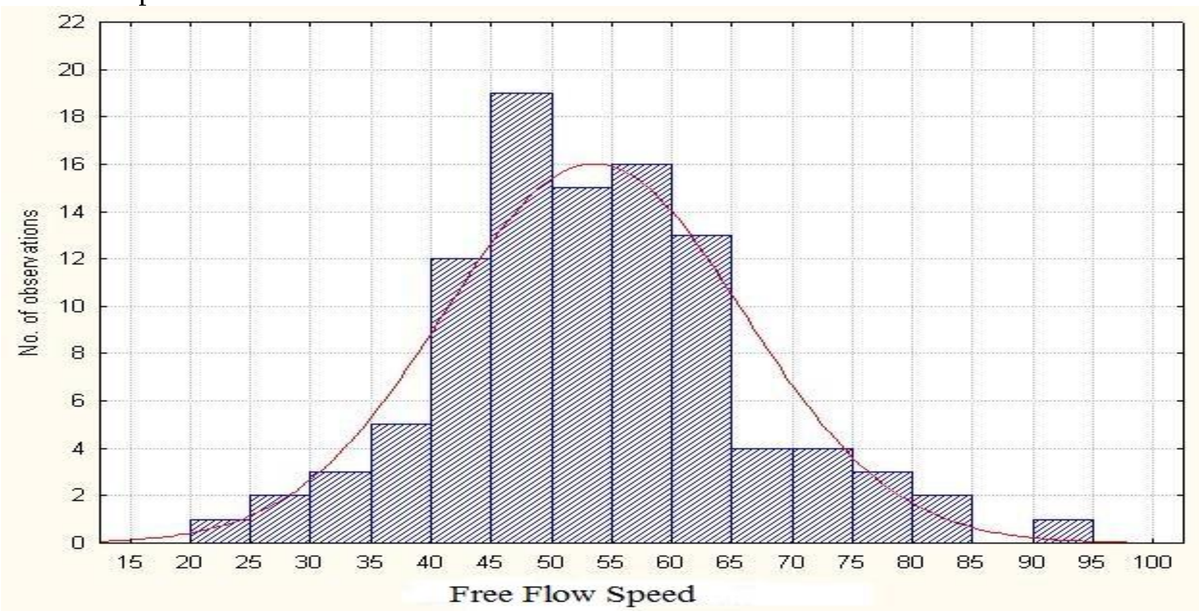
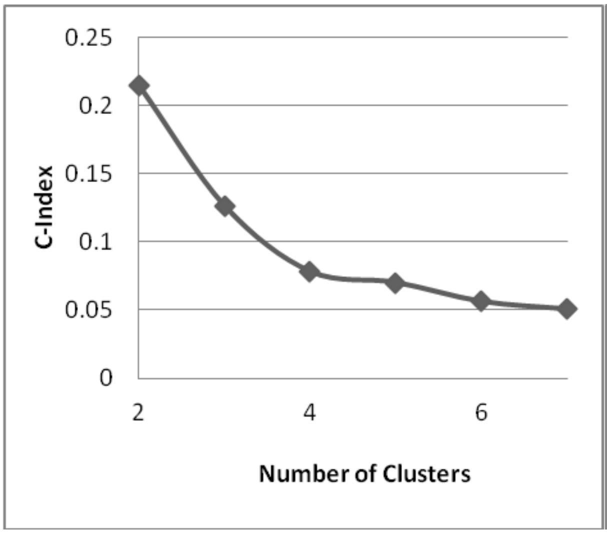


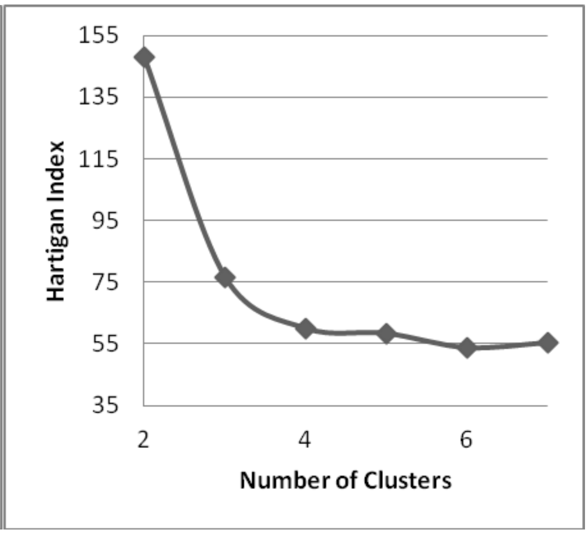
Fig.3: Probability distribution plot of collected FFS data using GPS receiver

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 all corridors. Number of trips covered for each direction of travel and for the study hours (peak, off-peak and free-flow) is at least 3 and sometimes it is up to six trips. After data has been collected in the field, it has been transferred back to the office computer by using Pathfinder version 3.00. The accuracy of field data was improved significantly using *differential correction.*

5. Results and Analysis

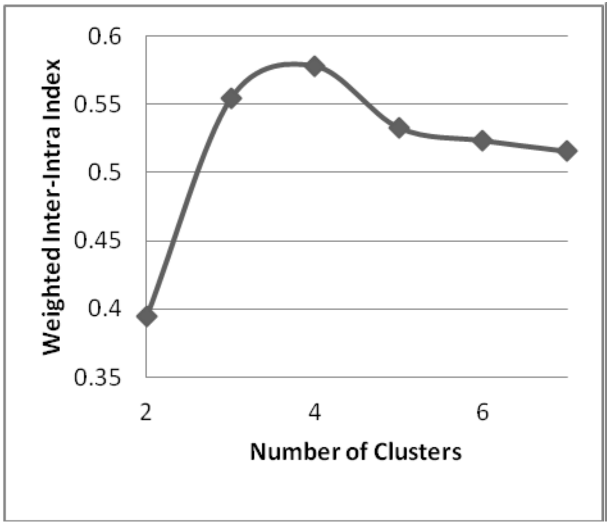
Both input data (free flow speed) to GA-Fuzzy clustering and its output (cluster centres) found from the cluster analysis are used in computing the values of cluster validation parameters. The values of five validation measures obtained for 2 to 7 number of clusters are plotted in Fig.4 (A) to Fig.4 (E).

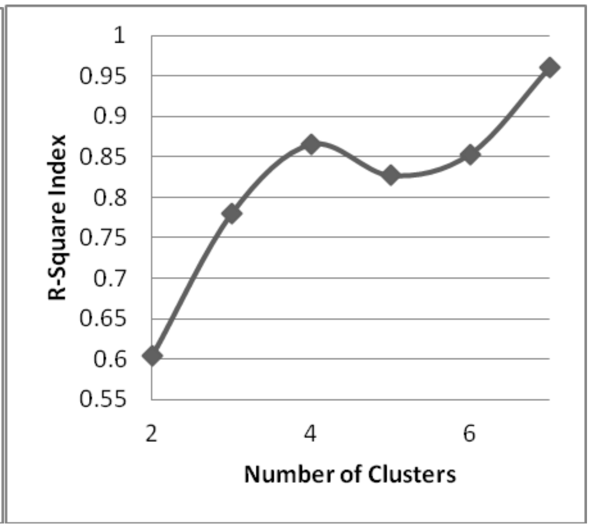




A: C-Index vs. Number of cluster

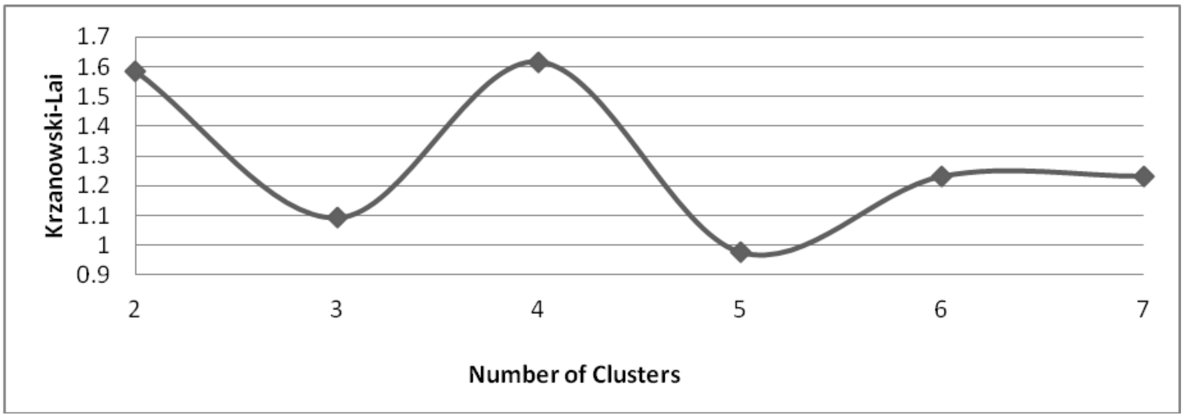
B: Hartigan Index vs. Number of cluster





C: Weighted Inter-Intra Index vs. Number of Cluster

D: R-Squared Index vs. Number of cluster



E: Krzanowski-Lai Index vs. Number of cluster

Fig. 4: 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 Fig.4 (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 Fig. 4(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 Fig. 4(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. Fig.4(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 Fig.4(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 Fig.5. 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.

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 Fig.6(A) to Fig.6(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.2. 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.

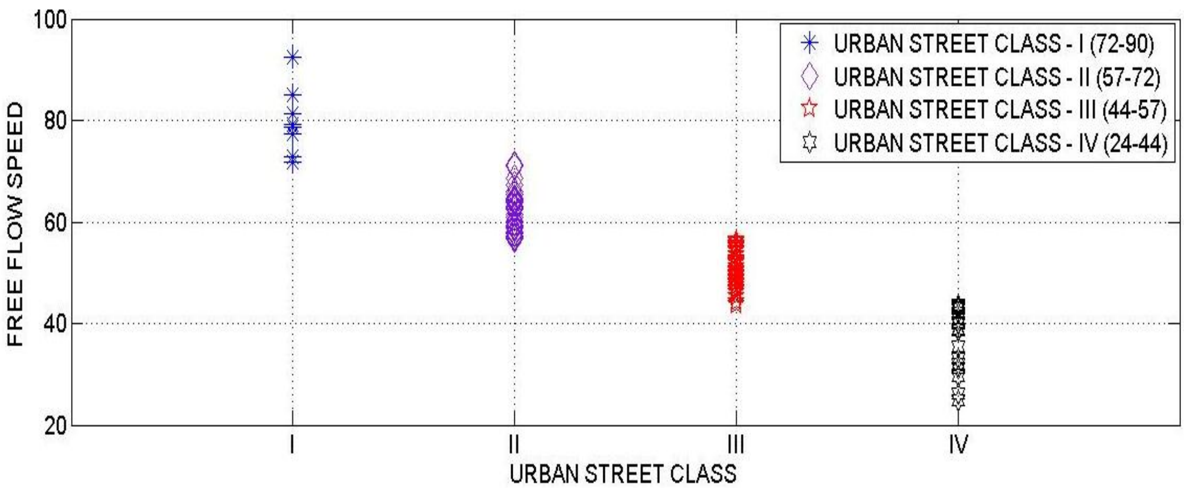
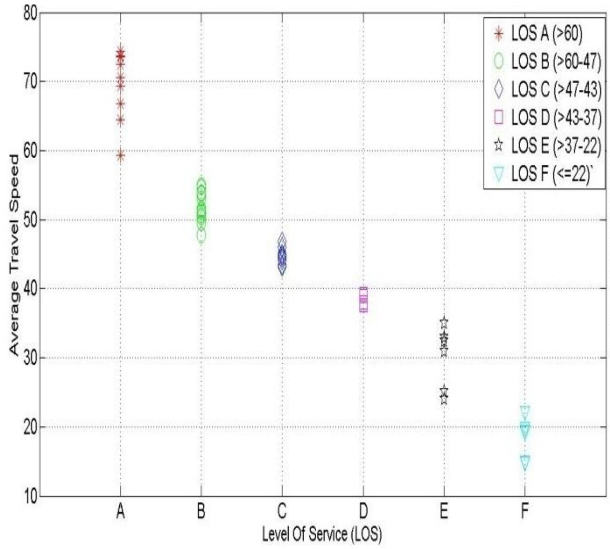
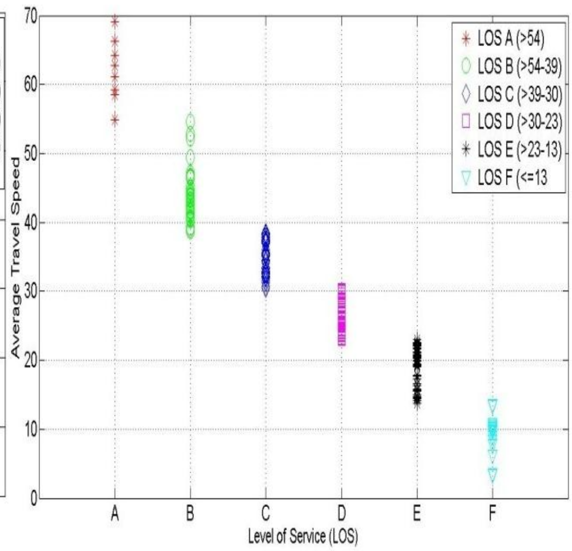


Figure 5: GA-Fuzzy Clustering of FFS for Urban Street Classification.

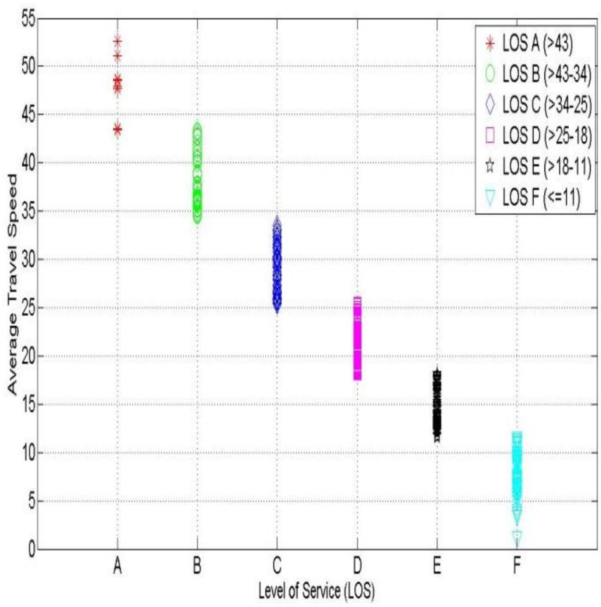
Average travel speed of different Level of Service categories in terms of percentage of free-flow speeds was found. The approximately percentage value for LOS ‘A’ to LOS “F” are 85, 70, 56, 43, 28 and 16-28 respectively in the present study. Whereas, in HCM (2000) it has been mentioned these values are 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.



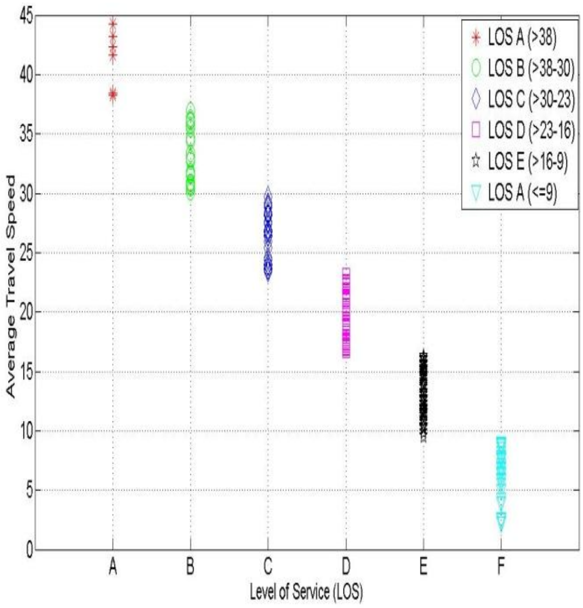
A: LOS of Urban Street Class I



B: LOS of Urban Street Class II



C: LOS of Urban Street Class III



D: LOS of Urban Street Class IV

Figure.6: Level of service of urban street classes (I-IV) using GA-Fuzzy clustering on average travel speeds

Table 2: 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

The coherence of the clustering result in the classification of urban streets into four classes and speed values into six LOS categories was verified with geometric and surrounding environmental characteristics of street segments and one urban street corridor out of the five urban street corridors is shown in the Appendix-1 for clarification. Table.3 contains various geometric and surrounding environmental characteristics of 15 street segments of Urban Street Corridor-3 that were acquired during inventory survey and also the different Levels of Services experienced on these street segments. Geometric and surrounding environmental characteristics were checked with the LOS category for a particular road segment. From this analysis it was inferred that good level of service “A”, “B”, “C”, can be attended in good traffic flow, geometric and surrounding environmental conditions. When traffic flow condition is poor and surrounding environmental conditions are not so good it is difficult to achieve good level of service rather in these conditions vehicles have to travel in poor level of service like “D”, “E”, and “F”. So travel speed the parameter used to determine LOS is quite capable in defining LOS categories for a particular urban street segment.

5. Summary and Conclusion

In this research an attempt is made to define the LOS criteria of urban street of developing courtiers having heterogeneous traffic flow condition. From literature review GPS was found as an efficient tool for collecting large amount of speed data. Five validation parameters were used to get the optimal number of cluster using FFS data for the classification of urban streets into number of classes. GA-Fuzzy clustering algorithm was used on the collected speed data for two times for the classification of the data set into number of groups. Finally, in this research FFS speed ranges for different urban street classes and speed ranges for different LOS categories were defined.

Considering five cluster validation parameters and the geometric and surrounding environmental characteristics it is well convinced to classify various urban street segments into four classes in Indian context. Hence, agreed with the classification of street segments into four classes as mentioned in HCM 2000. From this study it was found that the free flow speed ranges of Urban Street Class-IV is significantly lower than that mentioned in HCM 2000 because heterogeneous traffic flow and roads having varying geometric and surrounding characteristics. The speed ranges of levels of service categories under all the four urban street classes are found to be lower compared to the values mentioned in HCM 2000 because of substantial percentage of vehicles travelling on roads are slow moving. Side friction developed because of unwanted movement of pedestrians along and across the road sections compelled travelers to reduce vehicular speed. Road side vendors and on-street parking occupy substantial portion of road sections resulting in reduced space is available to the road users for through movement of vehicles. With the application of GPS a large amount of speed data were able to collected at a very short time period which are of course very accurate also. Hence this tool can be applied to collect speed data and cluster analysis can be applied to define the speed ranges for various countries to define the levels of service of their own rather than following some values which are not very much appropriate for them.

References

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

Basu, D., Maitra Roy, S., Maitra, B. (2006) “Modeling passenger car equivalency at an urban mid-block using stream speed as measure of equivalency.” *European Transport /Trasporti Europei*, 34, pp. 75-87.

Baumgartner, W.E. (1996) “Level of service: Getting ready for the 21st century.” *ITE Journal*, January, pp. 36-39.

Bhuyan, P.K, Krishna Rao K.V. (2011) “Defining level of service criteria of urban streets in Indian context”, *European Transport /Trasporti Europei*, 49, pp. 38-52.

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

Dandan,T., Wei,W., Jian,L., Yang, B. (2007) “Research on Methods of Assessing Pedestrian Level of Service for Sidewalk.*” J Transpn Sys Eng & IT* , 7(5), pp.74−79.

Fang,F.C., Elefteriadou,L., Pecheux,K.K., Pietrucha,M.T. (2003) “Using Fuzzy Clustering of User Perception to Define Levels of Service at Signalized Intersections.” *Journal of Transportation Engineering,* ASCE, 129 (6), pp.657-663.

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

Flannery, A. ,Rouphail, N., Reinke, D. (2008) “Analysis and Modeling of Automobile Users’ Perceptions of Quality of Service on Urban Streets.” *Transportation Research Record,* 2071, Transportation Research Board, Washington, D.C., pp. 26-34.

HCM (2000) “Highway Capacity Manual” *Transportation Research Board,* Washington, D.C.

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

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

Kikuchi,S., Chakroborty,P. (2007) “Frameworks to Represent the Uncertainty when Determining the Level of Service.” *Transportation Research Record*, 1968, pp. 53-62.

Kita, H., Fujiwara, E. (1995). “Reconsideration on the level of service and a proposed measure.” Proceedings of *15th Annual Meeting of JSTE,* Japanese, pp.25–28.

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

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

Maitra,B., Sikdar, P. K., Dhingra, S. L. (2004) "Modeling of Congestion: A Tool for Urban Traffic Management in Developing Countries", *European Transport / Trasporti Europei*, 27, pp. 45-56

Marwah, B.R., Singh, B. (2000) “Level of service classification for urban heterogeneous traffic: A case study of Kanpur metropolis.” Paper presented at the *Fourth International Symposium on Highway Capacity*, Hawaii, June-July, pp.271-286.

Ndoh, N.N., Ashford, N.J. (1994) “Evaluation of transportation level of service using fuzzy sets.” *73rdAnnual Meeting of TRB*, Transportation Research Board, Washington, D.C.

Pattnaik, S.B., Ramesh Kumar, K. (1996). “Level of service of urban roads based on users' perception.” Civil Engineering Systems, 14, pp.87-110.

Pecheux, K.K., Pietrucha, M.T., Jovanis, P.P. (2000) “User perception of level of service at signalized intersections: Methodological issues.” *79th Annual Meeting of TRB*, Transportation Research Board, Washington, D.C., pp.322-335.

Quiroga, C., Bullock, D. (1998). “Travel time studies with global positioning and geographic information systems: An integrated methodology.” *Transportation Research Part C*, 6(1–2), pp.101–127.

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

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

Shouhua,C., Zhenzhou, Y., Chiqing, Z., Li,Z. (2009) “LOS Classification for Urban Rail Transit Passages Based on Passenger Perceptions.” *J Transpn Sys Eng & IT*, 9(2), pp.99-104.

Spring, G.S. (1999) ‘‘Integration of safety and the highway capacity manual.’’ *Transportation Research Circular E-C018: 4th International Symposium on Highway Capacity*, Transportation Research Board, Washington, D.C., pp. 63–72.

Strehl, A., Ghosh, J. (2003) “Relationship-based clustering and visualization for high-dimensional data mining.” *INFORMS Journal on Computing*, 15, pp. 208-230.

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

TRB (1965). Special report 87: Highway capacity manual, *Transportation Research Board*, Washington, D.C.

TRB (1985). Special report 209: Highway capacity manual, Transportation Research Board, Washington, D.C.

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

Appendix-1: Levels of Service of Urban Street Classes in different Physical and Traffic Characteristics on Corridor -3

Location of Street Segments		*Free Flow Speed (Km/hr)*	*Geometric Characteristics*				*Surrounding Environmental Characteristics*			*Average Travel Speeds and Level of Service (LOS)*							
Seg ment . No.	*Urban Street Class*		*Segment Length (Km)*	*Lanes*	*Median Type*	*Access Density*	*Road Side Development*	*Parking*	*Pedestrian Activity*	*M-N-S*	*LOS*	*M-S-N*	*LOS*	*E-N-S*	*LOS*	*E-S-N*	*LOS*
1	I	71.1	2.2	6-LD	Barrier	Low	Medium density	Significant	Some	28.6	E	28.6	E	30.5	E	37.3	D
2	II	61.5	1.8	6-LD	Barrier	Low	Medium density	Some	Some	19.2	E	30.1	D	24.5	D	27.3	D
3	II	65.9	6.4	6-LD	Barrier	Low	Low density	No	Little	40.4	B	45.2	A	44.9	B	40.4	B
4	III	43.2	4.3	6-LD	Fence	Medium	Medium density	Some	Some	38.2	B	47.1	A	41.7	B	37.3	B
5	I	77.3	4.6	6-LD	Barrier	Low	Low density	No	Some	37.0	E	56.1	B	50.4	B	44.6	C
6	II	57.9	0.8	6-LD	Plantation	Medium	Low density	No	Some	16.9	E	27.1	D	17.4	E	38.3	C
7	II	58.8	0.99	6-LD	Plantation	Medium	Low density	No	Some	28.7	D	32.1	D	20.7	E	24.7	D
8	III	44.2	0.38	4-LD	Mountable	Medium	High density	Significant	Usually	20.2	D	27.0	C	25.5	C	48.0	A
9	II	54.2	8.15	4-LD	Mountable	Medium	Low density	Some	Little	27.4	D	33.2	C	27.9	D	20.2	E
10	II	54.4	0.9	4-LD	Plantation	Medium	Low density	Some	Little	15.5	E	30.1	C	12.9	F	15.9	E
11	II	56.7	1.4	3-LU	No	Medium	Low density	Some	Little	21.8	E	19.2	E	44.4	B	16.6	E
12	III	43.0	.79	2-LU	No	High	High density	Significant	Usually	22.5	D	26.0	C	26.4	C	22.2	D
13	IV	24.9	.37	2-LU	No	High	High density	Significant	Usually	25.9	C	22.8	D	11.3	E	8.9	F
14	IV	31.4	.90	2-LU	No	High	High density	Significant	Usually	13.2	E	31.5	B	31.5	B	23.7	C
15	IV	29.6	0.95	2-LU	No	High	High density	Significant	Usually	17.7	D	15.3	E	24.4	C	8.3	F

Note: LD-Lane Divided; LU- Lane Undivided; MNS-Morning-North-South; MSN-Morning-South-North; ENS-Evening-North-South;ESN-Evening -South-North