

Defining Level of Service Criteria of Urban Streets Using Neural Gas Clustering

A. K. Das^{1*}, A. K. Patnaik², A. N. Dehury³, P. K. Bhuyan⁴, U. Chattaraj⁵,
M. Panda⁶

¹ M. Tech Student, Department of Civil Engineering, National Institute of Technology Rourkela

² M. Tech Student, Department of Civil Engineering, National Institute of Technology Rourkela

³ M. Tech Student, Department of Civil Engineering, National Institute of Technology Rourkela

⁴ Assistant Professor, Department of Civil Engineering, National Institute of Technology Rourkela

⁵ Assistant Professor, Department of Civil Engineering, National Institute of Technology Rourkela

⁶ Professor, Department of Civil Engineering, National Institute of Technology Rourkela

Abstract: The traffic flow condition in India is highly heterogeneous. This study aims at defining the Level of Service (LOS) categories for highly heterogeneous traffic condition for urban streets in Indian context. Global Positioning System (GPS) is used as a powerful tool for collecting speed data at every one second interval. For this study Free Flow Speed (FFS) data and average travel speed data are collected on five corridors of Mumbai city. These speed data are analyzed using neural gas clustering method. The optimal number of clusters is obtained using six validation parameters. Neural gas is used two times in this study process. First neural gas is applied on FFS data to obtain the speed ranges of different urban street classes. Second time neural gas is used on the average speed data to obtain the speed ranges of different LOS categories. The above study confirms that the FFS ranges for different urban street classes and speed ranges for different LOS categories are found to be lower than that suggested by HCM 2000.

Keywords: Clustering, Free Flow Speed, Global Positioning System, Level of Service, Neural gas

I. INTRODUCTION

Due to rapid urbanization the urban streets in India suffers from substantially poor operating conditions. This study aims at assessing the operating condition. Earlier techniques involved the use of probe vehicle for collecting the travel time data. But the accuracy with this method varied from technician to technician. With emerging technologies and development of computer systems the use of GPS has come to effect. GPS can be used to record the travel time, travel speed and location as latitude and longitude. Using GPS speed data can be collected in every one second interval. GPS becomes very effective in collecting large number of speed data and handling of such large number of speed data also becomes very easy. The occurrence of error is also minimized with the use of GPS.

Defining LOS is basically a classification problem and cluster analysis is the most suitable technique that can be used for such kind of problem. In this study neural gas algorithm has been used for analysis purpose. Neural gas algorithm is used on the FFS data to classify the urban street segments into various urban street classes and to obtain the FFS speed ranges of different urban street classes. Neural gas algorithm is again used for the second time on the average travel speed data obtained during the peak and off-peak hours to determine the speed ranges of different LOS categories. Six validation measures have been used in this study to determine the optimal number of clusters. The speed ranges obtained in this study are found to be lower than HCM 2000.

II. LITERATURE REVIEW

Highway Capacity Manual (HCM, 2000) defined LOS as, “a qualitative 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 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 [1]. The studies of 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 [2]. Kita and Fujiwara (1995) mentioned that quantitative measures of the level of service are not the level of service itself, but merely characteristics of traffic conditions which have rather a strong relationship to the level of service of the traffic, and not necessarily shows the quality of service perceived by the drivers [3]. Spring (1999) stated that the use of distinct boundaries or thresholds

limits the subjective and continuous characteristics of the quality of service [4]. Kittelson and Roes (2001) have noted down that the current HCM methodologies have not been based upon user perception surveys [5]. Flannery et al. (2005) while relating quantitative to qualitative service measuring methods for urban streets found that level of service calculated by HCM 2000 methodology, predicted 35% of the variance in mean driver rating. The authors have suggested that LOS does not completely represent drivers' assessments of performance of urban streets because drivers perceive the quality of urban street segments in several dimensions, including travel efficiency, sense of safety, and aesthetics [6]. 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 [7]. Kaufman and Rousseeuw (1990) presented a *k*-medoid algorithm which they call PAM (Partition Around Medoids) [8]. Chakroborthy and Kikuchi (1990) have discussed the application of fuzzy set theory to the analysis of highway capacity and level of service. The authors have shown the limitations of the current procedure to determine highway capacity and service level. In this study, fuzzy numbers were used to represent the values of input variables and output variables which were involved in calculating capacity and service level [9]. Lingra (1995) compared grouping of traffic pattern using the Hierarchical Agglomerative Clustering and the Kohonen Neural Network methods in classifying traffic patterns [10]. Brilon, and Estel, (2010) have offered standardized methods that permit a differentiated evaluation of saturated flow (LOS F) conditions beyond static consideration of traffic conditions in German Highway Capacity Manual [11]. Basu, et al. (2006) modeled passenger car equivalency for urban mid-block using stream speed as measure of equivalence. In this research a neural network approach was used to illustrate the effects of traffic volume and its composition level on the stream speed [12].

III. CLUSTER ANALYSIS

Neural gas algorithm was proposed by Martinez and Schulzen in 1991. This algorithm allows to find a generalization for a set of data represented as a group with similar characteristics.

3.1 Neural gas algorithm

The algorithm for Neural gas is as follows:

Step1: Initialize the set A to contain N units C_i .

$$\text{Where } A = \{C_1, C_2, \dots, C_n\}$$

With reference vectors $W_{c_i} \in R^n$ chosen randomly according to expected distribution of 'X' samples.

Also initialize the time parameter t as $t=0$

Step 2: A random input 'X' is provided.

Step 3: Order all elements of A according to their distance to X, i.e., find the sequence of indices (i_0, i_1, \dots, i_{N-1}) such that W_{i_0} is the reference vector closest to X, W_{i_1} is the reference vector second closest to X and W_{i_k} , $k=0, \dots, N-1$ is the reference vector such that k vectors W_j exist with $\|X - W_j\| < \|X - W_k\|$. We denote $K_i(X, A)$ the number k associated with W_{i_i} .

Step 4: Adapt reference vectors according to

$$\Delta W_i = e(t) \cdot h_\lambda(k_i(X, A)) \cdot (X - W_i)$$

With the following time dependencies.

$$\lambda(t) = \lambda_i \left(\frac{\lambda_f}{\lambda_i} \right)^{\frac{t}{t_{\max}}} \tag{1}$$

$$e(t) = e_i \left(\frac{e_f}{e_i} \right)^{\frac{t}{t_{\max}}} \tag{2}$$

$$h_\lambda(k) = e^{-k/\lambda(t)} \tag{3}$$

Step 5: Increase the time parameter t :

$$t = t + 1$$

Step 6: If $t < t_{\max}$, go to step 2.

3.2 Validation Measures

Validation measures are used to obtain the optimal number of clusters. Optimal number of clusters depicts the number of groups into which a data set has to be divided. A single validation parameter is not sufficient to define the optimal number of clusters. Hence, six validation parameters have been used in this study.

A) Duda Index

$$duda = \frac{J_e(2)}{J_e(1)} = \frac{W_k + W_e}{W_m} \tag{4}$$

Where, $J_e(2)$ is the sum of squared errors within cluster when the data are partitioned into two clusters and $J_e(1)$ gives the squared errors when only one cluster is present.

W_k, W_l, W_m are defined as W_q

$$W_q = \sum_{k=1}^q \sum_{i \in c_k} (x_i - c_k)(x_i - c_k)^T \tag{5}$$

is the within group dispersion matrix for data clustered into q clusters.

It is assumed that clusters c_k and c_l are merged to form c_m .

$B_{kl} = W_m - W_k - W_l$, if $c_m = c_k \cup c_l$

n_i = number of observations in cluster c_i , $i = k, l, m$

The optimal number of clusters is the smallest q such that

$$duda = 1 - \frac{2}{\pi p} - z \sqrt{\frac{2 \left(1 - \left(\frac{8}{\pi^2 p} \right) \right)}{n_m p}} = \text{Critical Value} \tag{6}$$

Where, p = number of variables in the data set

Z = Standard normal score

B) McClain Index

This index consists of the ratio of two terms. The first term is the average within cluster distance divided by the number of within cluster distances. The denominator value was the average between cluster distance divided by the number of cluster distances. It is computed as follows:

$$mclain = \frac{mcan(\sum_{k=1}^q \sum_{i=1}^{n_k} \sum_{j=i+1}^{n_k} d_{ij})}{mcan(\sum_{k=1}^q \sum_{i \in c_k} \sum_{l=k+1}^q \sum_{j \in c_l} d_{ij})} \tag{7}$$

q is the number of clusters.

n_k is the number of objects in the k^{th} cluster, $k \in [1, \dots, q]$ k in $[1, \dots, q]$

d_{ij} = distance between i^{th} and j^{th} objects.

The minimum value of the index is used to indicate the optimal number of cluster.

C) PtBiserial Index

$$ptbiserial = \frac{(\overline{d_b} - \overline{d_w})((f_w f_b)/(n^2 d))^{\frac{1}{2}}}{S_d} \tag{8}$$

Where,

d_w = sum of within cluster distances

d_b = sum of between cluster distances

$\overline{d_w}, \overline{d_b}$ = respective means

S_d = standard deviation of all distances

n_d = total number of distances

f_w = number of within cluster distances

f_b = number of between cluster distances

Its value varies between -1 to +1 and the maximum values indicate the optimal number of clusters.

D) Gplus Index

$$G(+)= \frac{2S(-)}{n_d(n_d-1)} \tag{9}$$

Where,

$S(-)$ is the number of discordant comparisons i.e the number of times where two points which were in the same cluster had a larger distance than two points not clustered together.

n_d = total number of distances (which is the same as the total number of observations or objects under study).

Minimum values of the index are used to determine the optimal number of clusters in the data.

E) Tau Index

Tau index is computed as follows:

$$\text{Tau} = \frac{S(+)-S(-)}{\left[\left(\frac{n_d(n_d-1)}{2-t}\right) \times \left(\frac{n_d(n_d-1)}{2}\right)\right]^{0.5}} \tag{10}$$

Where,

S(+) is the number of concordant comparisons.

S(-) is the number of discordant comparisons.

n_d is the total number of distances (which is the same as the total number of observations or objects under study.)

t is the number of comparisons of two pairs of points where both pairs represent within cluster comparisons or both pairs are between cluster comparisons.

The maximum value in the hierarchy sequence was taken as indicating the correct number of clusters.

F) Ratkowsky Index

The index is based on the formula:

$$\frac{\bar{s}}{q^2} \tag{11}$$

The value of \bar{s} is equal to the average of the ratios of B/T where B stands for the sum of squares between the clusters for each variable and T for the total sum of squares for each variable. The optimal number of clusters is that value of q for which $\bar{s}/q^{(1/2)}$ has its minimum value.

IV. STUDY AREA AND DATA COLLECTION

4.1 Study Area

Mumbai, the business capital of India was chosen as the study area. In total five corridors were chosen for this study purpose, including four North-South corridor and one East-West corridor. During morning hours commuters move from North to South for work trip and return back towards North during evening hours. Four North-South corridors include Eastern express highway extending upto South (Corridor-1), LBS Road extending upto South via Ambedkar Road (Corridor-2), Western express highway extending upto marine drive (Corridor-3), SV Road extending upto 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 1.

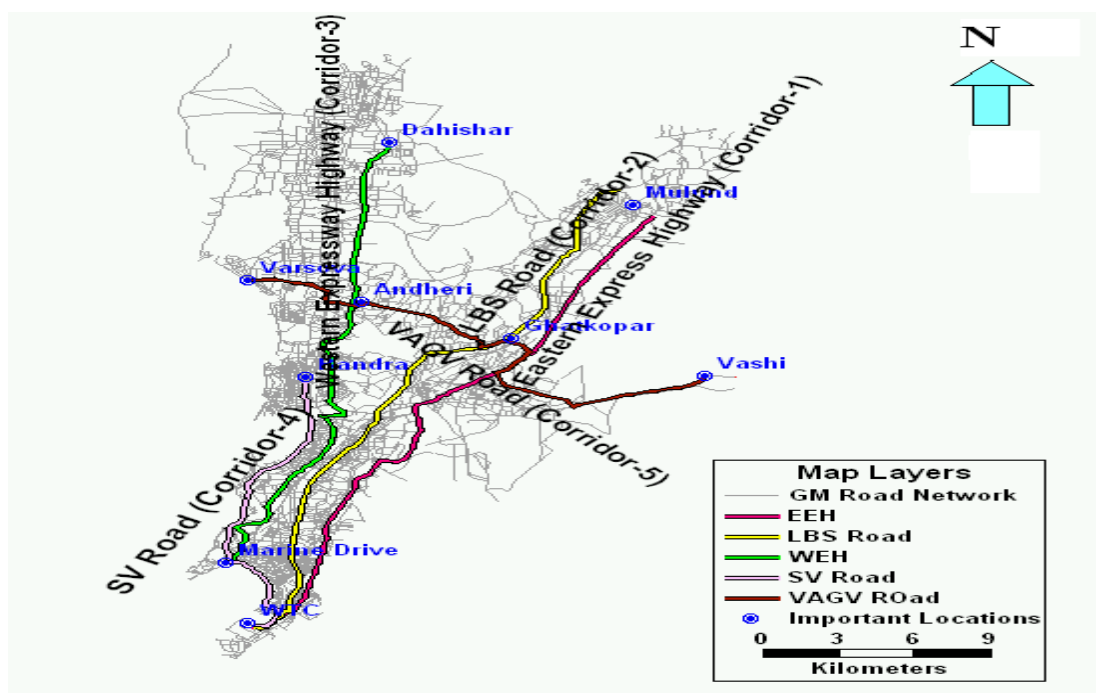


Fig.1 Map showing selected corridors of Greater Mumbai

4.2 Data Collection

GPS has been used as a powerful tool for the collection of data for this study purpose. Mid-sized vehicle (car) was fitted with GPS receiver so as to collect the travel speed data and travel time data.

First type survey was carried to obtain the free flow speed data. For the collection of free flow speed data, sufficient knowledge is required about the time duration when the traffic volume is less than or equal to 200 vehicles per lane per hour. 24-hour traffic volume count survey was conducted prior to the collection of free flow speed. The traffic volume data were collected on 45 stations on seven screen lines. In order to accomplish this study traffic volume per lane per hour was calculated for roads coming under this study area. It was established that free flow condition (less than 200 vehicles/l/h) is approaching at 12 midnight and all roads sections are having free flow traffic condition from 1AM to 5 AM. Hence free flow speed for all these corridors were collected during these hours.

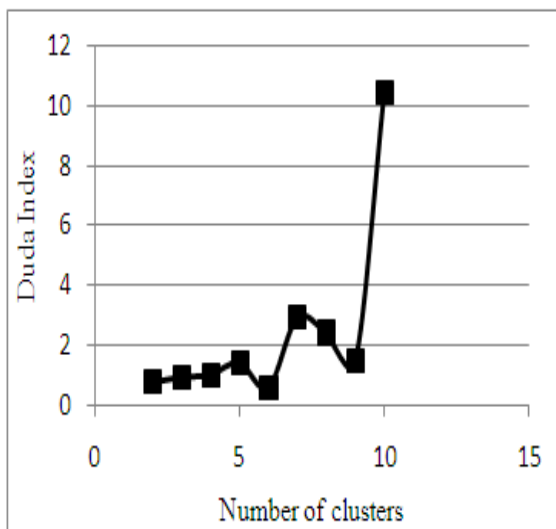
The second type of data collected was the congested travel speed. congested travel speed survey was conducted both during peak and off-peak hours on both directions of all corridors. Minimum number of trips covered for each directions of travel and for the study hours is three and can extend up to six. After the data is collected it is transferred back to computer using Pathfinder office version 3.0.

Third kind of data included roadway inventory details. For completing this study details on segments like segment number, number of lanes, pedestrian activity, access density, speed limit, number of flyovers, roadside development, construction activity, separate right turn lane, date and day of data collection and segment length were collected. Proper segmentation technique was applied during the data collection which is the directional stretch of road section immediately after signalized intersection to the location point immediately after the next signal.

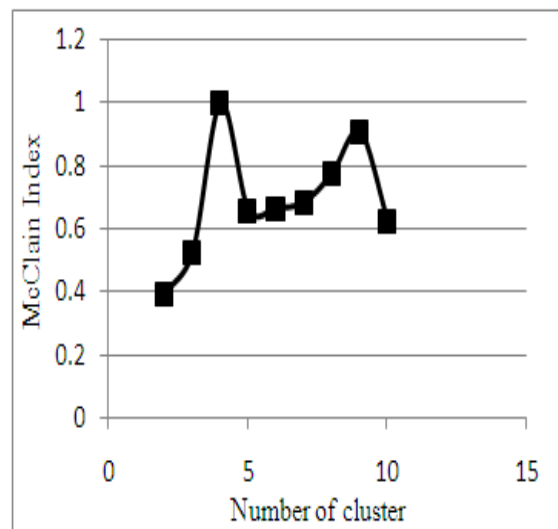
V. RESULTS AND ANALYSIS

For carrying out this study purpose six validation parameters has been used to determine the optimal number of clusters. FFS data is used as an input to classify the urban street segments. The values of the optimal number of clusters varied from four to seven numbers of clusters. The values of the validation parameters obtained are plotted in Fig. 2(A) to Fig. 2(F).

Literature survey indicates that the minimum value of Duda Index gives the optimal number of clusters for a set of data. The index is minimum for six numbers of clusters as shown in Fig. 2(A). For McClain Index the maximum value yields the optimal number of clusters. The index is maximum for four numbers of clusters as shown in Fig. 2(B). Study of literature reveals that the maximum value of PtBiserial Index gives the optimal number of clusters. The index value is maximum for four numbers of clusters as shown in Fig. 2(C). For Gplus Index the minimum value gives the optimal number of clusters. The index value is minimum for four number of clusters as shown in Fig. 2(D). Literature survey shows that the maximum value of Tau Index gives the optimal number of clusters. The index value is maximum for four numbers of clusters as shown in Fig. 2(E). The maximum value of Ratkowsky Index gives the optimal number of clusters. The index returns the maximum value for seven numbers of clusters as shown in Fig. 2(F). Out of the six validation parameters four validation parameters returned the value as four. Hence the urban street segments were classified into four classes using Neural gas algorithm.



A: Duda Index vs Number of Clusters



B: McClain Index vs Number of Clusters

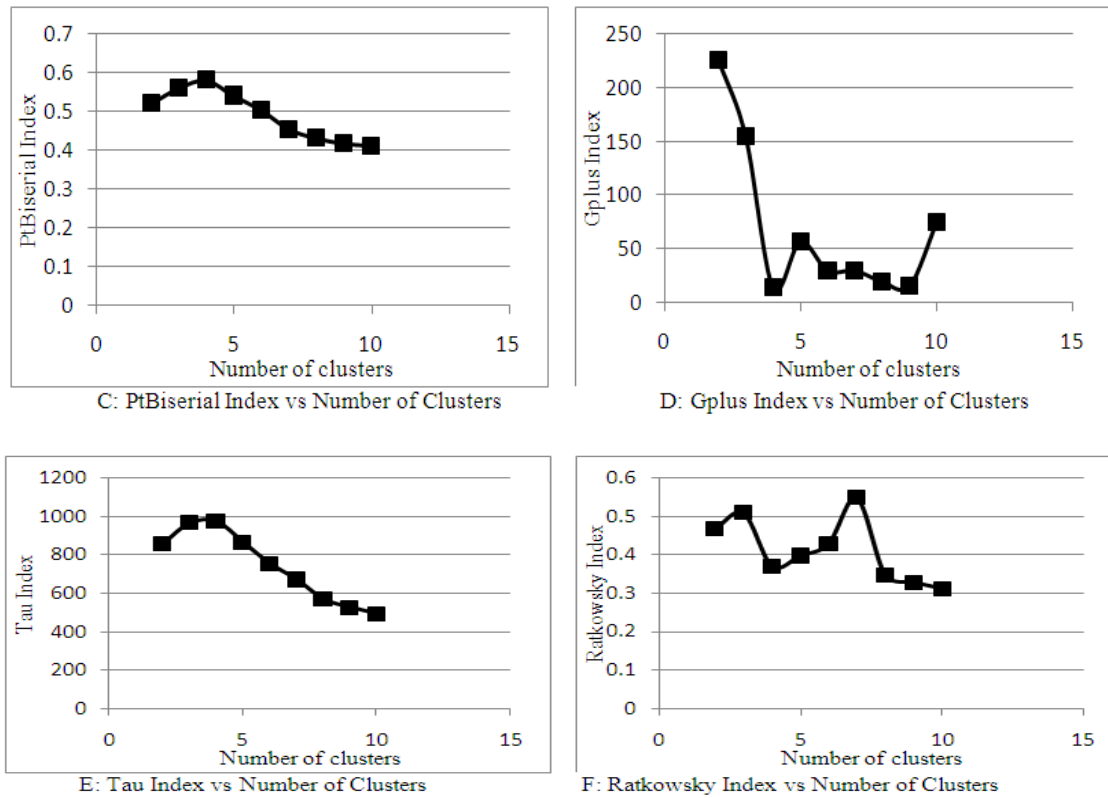


Fig. 2 Validation measures for optimal number of clusters using Neural gas clustering

The free flow speed data collected from the hundred street segments of Mumbai city were clustered using neural gas algorithm. These free flow speed data were used for the classification of the street segments into four number of classes. The result obtained is represented in Fig. 3. The speed ranges can be depicted from the legends indicated in the figure.

The average travel speed obtained during the peak and off-peak hours were clustered using neural gas algorithm in six numbers of clusters. These six clusters give the speed ranges for different urban street classes. The result of the clustering is shown in Fig.4 (A) to Fig.4 (D). The legend in Fig.4 (A) to Fig.4 (D) gives the speed ranges for six LOS categories of urban street classes obtained by using neural gas clustering. The speed ranges for LOS categories found using neural gas clustering is shown in Table 1. It was observed that the values of free flow speed range of urban street classes and speed ranges of LOS categories obtained from the present study are different from that mentioned in HCM-2000.

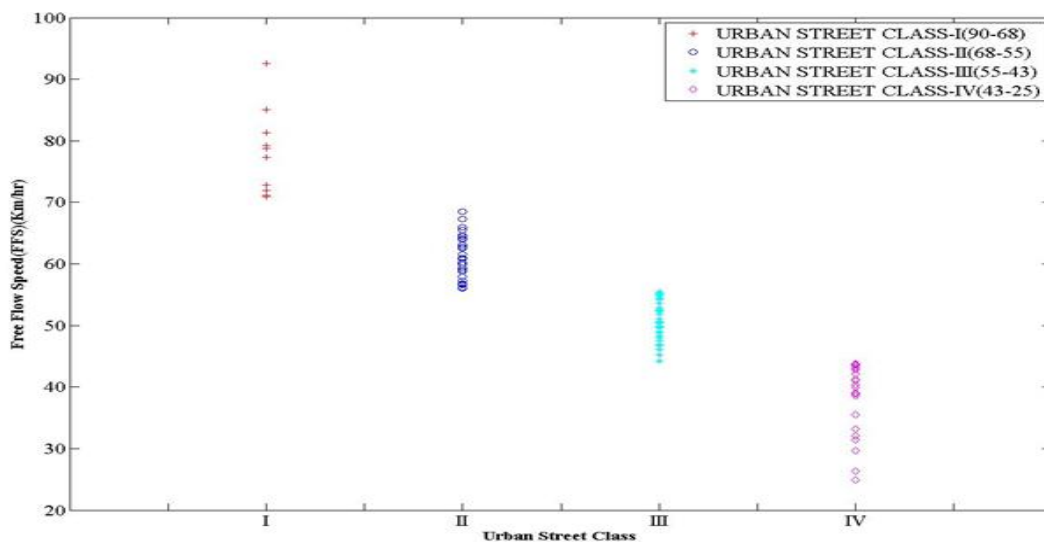
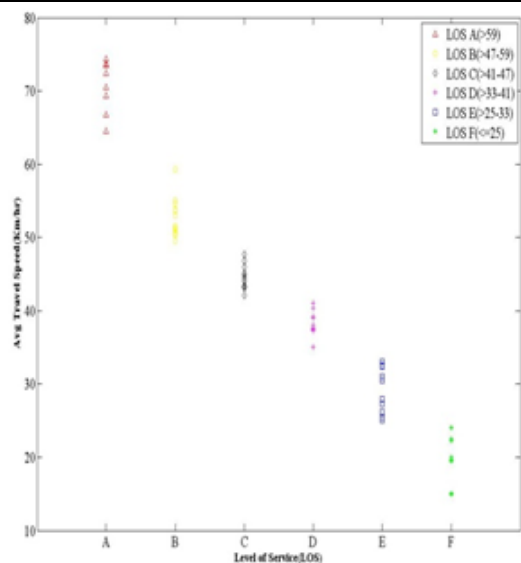
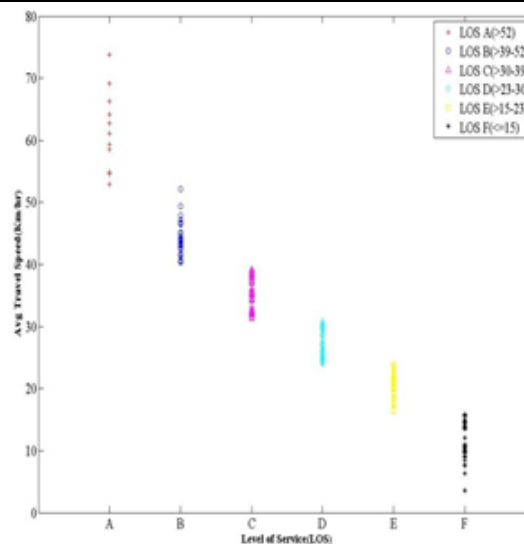


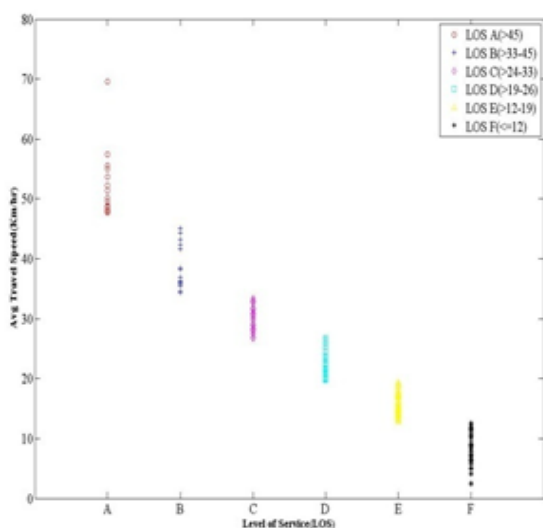
Fig. 3 Neural gas Clustering of FFS for Urban Street Classification



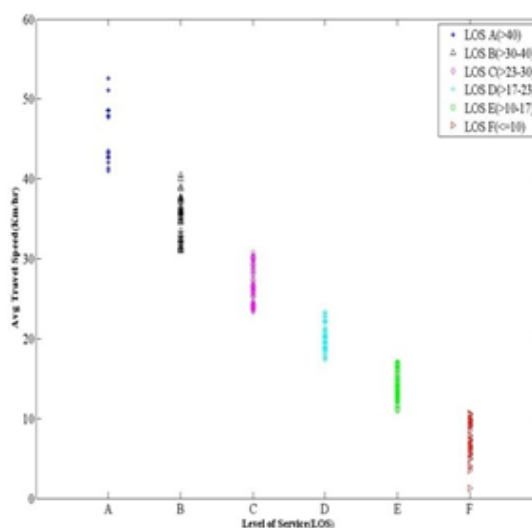
A: LOS of Urban Street Class I



B: LOS of Urban Street Class II



C: LOS of Urban Street Class III



D: LOS of Urban Street Class IV

Fig. 4 Level of service of urban street classes (I-IV) using neural gas clustering on average travel speeds

Table 1: Urban Street Speed Ranges for different LOS Proposed in Indian Conditions by neural gas Clustering method

Urban Street Class	I	II	III	IV
Range of free-flow speed (FFS)	90 to 68 km/h	68 to 55 km/h	55 to 43 km/h	43 to 25 km/h
Typical FFS	75km/h	60km/h	47km/h	35 km/h
LOS	Average Travel Speed (Km/h)			
A	>59	>52	>45	>40
B	>47-59	>39-52	>33-45	>30-40
C	>41-47	>30-39	>24-33	>23-30
D	>33-41	>23-30	>19-26	>17-23
E	>25-33	>15-23	>12-19	>10-17
F	≤25	≤15	≤12	≤10

The speed ranges of different LOS categories and urban street classes depend on the physical features of road segments under consideration. The speed values obtained after clustering are checked for coherence with the physical characteristics of road segments under consideration obtained during inventory survey. Road segments with good geometric condition and surrounding environmental condition demonstrate good LOS i.e. LOS “A”, “B”, “C”. Moreover road segment with poor traffic flow, poor geometric condition and surrounding environmental conditions demonstrate poor LOS i.e. LOS “D”, “E”, “F”.

VI. SUMMARY AND CONCLUSION

GPS has been used as a powerful tool for the collection of speed data. In the present study an attempt has been made to define the LOS criteria for developing countries with heterogeneous traffic condition. Six validation parameters have been used in the study purpose which helps us to classify the urban street segments into four numbers of classes (I-IV). Neural gas algorithm is used two times in the study. First it is used on the free flow speed data to classify the urban street segment into four numbers of classes. This helps us to obtain the speed ranges of the urban street classes. Neural gas is used for the second time on the average travel speed to determine the speed ranges of different LOS categories. The above study confirms that the speed ranges obtained in Indian context are different than that mentioned in HCM 2000. This is due to the side friction developed due to pedestrian movements across and along the stretch of road, road side vending and road side parking activities.

REFERENCES

- [1]. HCM (2000). “Highway Capacity Manual.” Transportation Research Board, Washington, D.C.
- [2]. Baumgartner, W.E. (1996). “Level of service: Getting ready for the 21st century.” ITE Journal, January, 36-39.
- [3]. Kita, H., and Fujiwara, E. (1995). “Reconsideration on the level of service and a proposed measure.” Proceedings of 15th Annual Meeting of JSTE, Japanese, 25–28.
- [4]. 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., 63–72.
- [5]. Kittelson, W.K., and Roes, R.P. (2001). “Highway capacity analysis after the highway capacity manual 2000.” Transportation Research Record, 1776, Transportation Research Board, Washington, D.C., 10–16.
- [6]. Flannery, A., Wochinger, K., and Martin, A. (2005). “Driver assessment of service quality on urban streets.” Transportation Research Record, 1920, Transportation Research Board, Washington, D.C., 25–31.
- [7]. Taylor, M.A.P., Woolley, J. E., and Zito, R. (2000). “Integration of the global positioning system and geographical systems for traffic congestion studies.” Transportation Research Part-C, 8, 257–285.
- [8]. Kaufman, L., and Rousseeuw, P.J. (1990). Finding groups in data: An introduction to cluster analysis, Wiley, New York.
- [9]. Chakroborthy, P. and Kikuchi, S. (1990). “Application of fuzzy set theory to the analysis of capacity and level of service of highways.” In the proceedings of First International Symposium on Uncertainty Modeling and Analysis, IEEE Computer Press, College Park, Maryland, 146-150.
- [10]. Lingras, P. (1995). “Classifying highways: Hierarchical grouping versus kohonen neural networks.” Journal of Transportation Engineering, ASCE, 121(4), 364–368.
- [11]. Brilon, W. and Estel, A. (2010) Differentiated Analysis of Level of Service F within the German Highway Capacity Manual. Journal of the Transportation Research Board, 2173, pp.36-44.
- [12]. Basu, D. et al. (2006) Modeling passenger car equivalency at an urban mid-block using stream speed as a measure of equivalency. International Journal of Transport Economics, Engineering and Law, 34, pp.75-87.