Alexandria Engineering Journal (2016) **55**, 407–418

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

ORIGINAL ARTICLE

Divisive Analysis (DIANA) of hierarchical clustering and GPS data for level of service criteria of urban streets

Ashish Kumar Patnaik [a,*], **Prasanta Kumar Bhuyan** [a,1], **K.V. Krishna Rao** [b,2]

[a] *Department of Civil Engineering, National Institute of Technology, Rourkela, 769008, India*
[b] *Department of Civil Engineering, IIT Bombay, Powai, Mumbai, Maharastra 400076, India*

Received 26 February 2014; revised 5 April 2015; accepted 8 November 2015
Available online 29 December 2015

KEYWORDS

Urban street;
Level of Service (LOS);
Free flow speed (FFS);
GPS;
DIANA

Abstract Level of Service (LOS) for heterogeneous traffic flow on urban streets is not well defined in Indian context. Hence in this study an attempt is taken to classify urban road networks into number of street classes and average travel speeds on street segments into LOS categories. Divisive Analysis (DIANA) Clustering is used for such classification of large amount of speed data collected using GPS receiver. DIANA algorithm and silhouette validation parameter are used to classify Free Flow Speeds (FFS) into optimal number of classes and the same algorithm is applied on speed data to determine ranges of different LOS categories. Speed ranges for LOS categories (A–F) expressed in percentage of FFS are found to be 90, 70, 50, 40, 25 and 20–25 respectively in the present study. On the other hand, in HCM (2000) it has been mentioned these values are 85 and above, 67–85, 50–67, 40–50, 30–40 and 30 and less percent respectively.

1. Introduction

The traffic and transportation facility of any country significantly defines its development. The developing countries like India must have a well defined LOS analysis procedure to develop a good road network, because, it is very essential for the planning, design of transportation system and allocation of limited resources to the competing projects. The FFS ranges for urban street classes and speed ranges of LOS categories that are specified in HCM [1] have been followed in India for LOS analysis of urban streets. In fact speed ranges mentioned in HCM [1] are suitable for developed countries having homogenous traffic flow. In developing countries like India traffic on roads is highly heterogeneous, and twenty two types of vehicles having wide variation in physical size travel on roads, as a result of which vehicular travel speed is comparatively less under heterogeneous flow condition.

In transportation engineering, according to HCM [2] "Level of Service (LOS) is a quantitative stratification of a performance measure or measures that represent quality of

---

[*] Corresponding author.

E-mail addresses: ashishkumar.patnaik@yahoo.in (A.K. Patnaik), pkbtrans@gmail.com (P.K. Bhuyan), kvkrao@iitb.ac.in (K.V. Krishna Rao).

[1] Tel.: +91 661 246 2313.

[2] Tel.: +91 22 2576 7305.

Peer review under responsibility of Faculty of Engineering, Alexandria University.

http://dx.doi.org/10.1016/j.aej.2015.11.003
1110-0168 © 2015 Faculty of Engineering, Alexandria University. Production and hosting by Elsevier B.V.

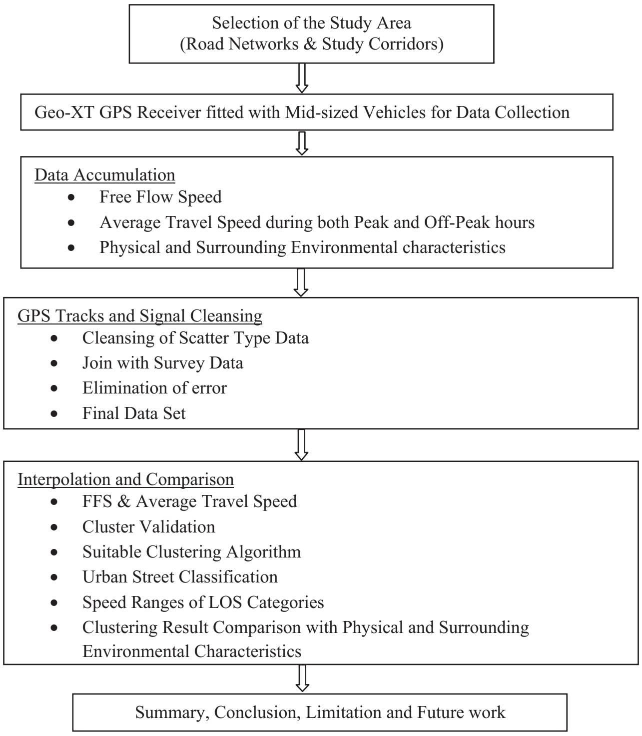

Figure 1 Overall framework of the study.

service. The measures used to determine LOS for transportation system elements are called "*service measures*." The HCM designates the Urban street facilities into six different types of LOS ranging from "A" to "F". LOS A represents the best quality of road and hence best serviceability and LOS F represent the worst facility. The Urban street classification or LOS is based on average through-vehicle travel speed for that segment or for the entire street under consideration. As 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 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. The best way of collecting data is by moving-observer method. However accuracy with this technique varies from technician to technician. The location can be pointed according to their longitude-latitude and GPS gives the accurate measure of their position and hence travel-speed data can be recorded at different points. GPS receivers can record location and speed automatically at regular sampling periods; hence a considerable data can be collected in terms of travel speed and travel time.

Defining LOS of urban street classes is basically an approach to classify average travel speeds on road segments of the entire road network into number of groups. 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 through GPS receiver. In this study Divisive Analysis (DIANA) Clustering is used for the classification of data set. The data set used in this study was obtained from 10 to 12 travel runs taken on five major urban corridors in the city of Mumbai, India. The total length of these corridors is about 140 km. These corridors, on the whole, were divided into 100 street segments. Comprehensive data sets of free-flow speed, travel speeds during both peak and off- peak hours, inventory details and classified traffic volume data were used. The clustering algorithm was used twice in this research. First, DIANA clustering was used on Free Flow Speed (FFS) data to get FFS ranges of urban street classes. After defining the speed ranges DIANA 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 silhouette width was used as Validation parameters. The physical characteristics of urban street corridor-5 with 19 street segments are illustrated in Appendix A in which the coherence of the clustering result for the classification of urban streets and LOS categories was verified with geometric and surrounding environmental characteristics of street segments. Analyzing the data it is found that urban street segments can be classified into four classes in Indian context. The speed limits of LOS categories are proportionately lower than that values mentioned in HCM [1] and speed values expressed in terms of percentage of FFS are marginally different from those mentioned in HCM [2]. The overall framework of the study is as shown in Fig. 1.

2. Literature review

The current definition of LOS being followed is that defined in 2000 HCM. Level of service in the Highway Capacity Manual (HCM [1]) is 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." Several studies have been performed on LOS analysis of urban street classes. Patel and Joshi [3] investigated the behavior of mixed traffic stream speed and flow rate on an access controlled urban arterials in Surat city in Gujarat state of India. The study established the thresholds of level of service based on volume to capacity ratio by using cluster analysis approach. Arasan and Vedagiri [4] developed micro simulation technique to study the effect of provision of reserved bus lanes on the flow of non-lane based heterogeneous traffic on urban roads in India. The study indicated that exclusive bus lanes were intended to enhance the LOS of bus transport in terms of speed, reliability, safety, etc. Polus and Cohen [5] presented flow characteristic measure for two lane rural highways, the flow, average platoon length, traffic intensity, the percent time spent following and the freedom of flow. The LOS is estimated from the freedom of flow parameters. Bhuyan and Nayak [6] presented a classification and analysis of the results achieved using various tools for the estimation of level of service of urban streets.

Dandan et al. [7] 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. [8] found that the LOS criteria of walkways proposed by HCM [1] 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 which is lower than that suggested by HCM [1]. In this study it has been found that body size, culture, gender and age influence the LOS classification. Fang and Pechuex [9] studied LOS of signalized intersections 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. Lee and Jam [10] proposed six LOS design standards in Hong Kong mass transit railway stations. Light/clear visibility on the Hong Kong MTR station stairways was found to be major factor of concern, with interest in the environment being of low priority. Azimi and Zhang [11] have applied three pattern recognition methods (K-means, fuzzy C-means, and CLARA (clustering large applications)) to classify freeway traffic flow conditions on the basis of flow characteristics. The classification results from the three clustering methods were compared with the HCM [1] LOS. Lee et al. [12] proposed LOS standards for signalized crosswalks in Hong Kong commercial/shopping areas which explicitly take the bidirectional pedestrian flow effects into account. The study defined explicitly the LOS boundaries for different levels of bidirectional flow regarding area occupancy, pedestrian flow and walking speed. Roess et al. [13] have authored "Level of Service 2010 and beyond" and attempted to address on the history of the LOS concept and its use in the planning, design, and analysis of traffic facilities. Ko et al. [14] have conducted an extensive survey to know the performance measures that significantly affect truck drivers' perceptions of LOS on various roadway types. It has been identified various performance measures through the analysis of survey data and the results lay the groundwork for future research that can focus on the actual development of quantitative LOS methods for truck mode.

Chen et al. [15] have developed a methodology using Fuzzy Neural Networks to access the LOS perceived by road users at signalized intersections. In this study, a neural network containing fuzzy reasoning experiences was employed to combine the perceived attributes in order to determine LOS. Cao et al. [16] have conducted a stated preference survey to study the factors influencing the LOS perceived by passengers at Platforms of Beijing Urban Rail Transit. In this research, it has been observed that congestion level of the platform was the most important factor influencing the LOS of the platform, followed by passenger order, air quality, information signs, and waiting time.

Zegeer et al. [17] have developed default values to represent input parameters to the approach methodology used in the analysis of Capacity and level of service of roads when they are difficult to measure or estimate. It has been observed that out of several default parameters, nineteen parameters have shown high degree of sensitivity in influencing service measure results in the appropriate methodology. Dowling et al. [18] have developed a methodology for the assessment of the quality of service provided by urban streets for the flow of traffic by various modes on the road network at national level. In this research the authors have categorized urban travels into four types (motorized vehicle, transit mode, bicycle rider, and walk mode) and hence developed separate LOS models for each mode of travel. Chakroborty and Kikuchi [19] 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. 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&B) and splitting existing LOS F into two categories. Choocharukul et al. [20] examined road user perceptions of freeway LOS by presenting study participants with a series of video clips of various traffic conditions and asking them their perception of LOS. A random effect ordered probability model is then used to statistically link participant- recorded perceptions of LOS with measurable traffic conditions and participant characteristics. Semeida [21] established the relationship between road geometric characteristics and heavy vehicles by applying statistical modeling. Then the LOS and the capacity of two ways modeling by artificial neural network (ANN). Becher [22] described the development of a procedure for an overall evaluation of traffic light controlled intersections, taking into consideration the effects of mean delays on the quality evaluation from the point of view of the road user. Ameri et al. [23] determined the capacity of two lane suburban roads and its effect on LOS. The results show that level of service based on average travel speed is better than the LOS based on percent time spent following. These are some of researches carried out at different locations under different traffic conditions which give a strong background for further research carried out in this study in defining LOS criteria in Indian context.

3. Divisive Analysis Clustering (DIANA)

3.1. DIANA algorithm

DIANA is a hierarchical clustering technique which constructs the hierarchy in the inverse order. It approaches the reversal algorithm of Agglomerative Hierarchical Clustering. There is one large cluster consisting of all n objects. At each subsequent step, the largest available cluster is split into two clusters until finally all clusters, comprise of single objects. Thus, the hierarchy is built in $n - 1$ steps. In the first step of an agglomerative method, all possible fusions of two objects are considered leading to $n\ (n - 1)/2$ combinations. In the divisive method based on the same principle, there are $2^{n-1} - 1$ possibilities to split the data into two clusters. This number is considerably larger than that in the case of an agglomerative method. To avoid such large calculations the following steps have been followed:

Step-1: The DIANA clustering is followed by Agglomerative Hierarchical Clustering up to the cluster contains all the objects. Then the Divisive Analysis Clustering (DIANA) follows the top-down approach assuming it single cluster having level $L\ (0) = n$ and sequence number $m = 0$.

Step-2: The most dissimilar pair of clusters in the current cluster is found out; that is (r), (s) in which $d\ [(r), (s)] = \min d\ [(i), (j)]$, where min is the complete pairs of cluster in the current cluster.

Step-3: The sequence number is incremented in the manner $m = m + 1$. The cluster is broken into clusters (r) and (s) to form next cluster to make the level of clustering: $L\ (m1) = d\ [(r)]$ and $L\ (m2) = d\ [(s)]$.

Step-4: The distance matrix (D) is updated by adding the rows and columns corresponding to clusters (r) and (s).

The similarity between the new cluster, denoted by (*r*, *s*) and old cluster (*k*) is defined in this way:

$$D[(k),(r,s)] = \min d[(k),(r)], d[(k),(s)]$$

If all objects are distinct clusters, then stop; otherwise proceed to step-2.

For example, consider a sample data set *x*, made up of six objects, say average free flow speed (ffs) values in kmph on six street segments. The data set can be defined as a matrix.

$$x = [\mathrm{ffs1}; \mathrm{ffs2}; \mathrm{ffs3}; \mathrm{ffs4}; \mathrm{ffs5}; \mathrm{ffs6}]$$
$$= [85.00; 92.56; 72.85; 50.66; 39.89; 38.58]$$

The object centroid distance is calculated by Euclidean distance. If $p = (p_1, p_2)$ and $q = (q_1, q_2)$, then the Euclidean distance is given by $d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$, and then the distance between cluster centroid to each object having distance matrix – D_0 is calculated using single linkage distance function. Similarly for all other distance matrices (D_1–D_9) same single linkage distance function is used. The distance matrix (D_1–D_5) is followed by Agglomerative Hierarchical Clustering. After finding the cluster contains all the objects, Divisive Analysis Clustering (DIANA) is started assuming it single cluster having level *L* (0) = *n* and sequence number *m* = 0 until to get single-single object in all clusters: (*A*, *B*, *C*, *D*, *E*, *F*). Here the distance matrix (D_6–D_{10}) is calculated by Divisive Analysis Clustering (DIANA).

Distance matrix – D_0

Dist.	*A*	*B*	*C*	*D*	*E*	*F*
A	0	7.56	12.15	34.34	45.11	46.42
B	7.56	0	19.71	41.9	52.67	53.98
C	12.15	19.71	0	22.19	32.96	34.27
D	34.34	41.9	22.19	0	10.77	20.08
E	45.11	52.67	32.96	10.77	0	1.31
F	46.42	53.98	34.27	12.08	1.31	0

Minimum distance between cluster *E* and *F* is 1.31; hence clustered together.

Distance matrix – D_1

Dist.	*A*	*B*	*C*	*D*	*E, F*
A	0	7.56	12.15	34.34	?
B	7.56	0	19.71	41.9	?
C	12.15	19.71	0	22.19	?
D	34.34	41.9	22.19	0	?
E, F	?	?	?	?	0

? marks under a column or row means two elements are clubbed together at this point of calculation but distance from this new point to all other points are not known.

Minimum distance between cluster *A* and *B* is 7.56.

Distance matrix – D_2

The cluster *A* and *B* is grouped into single cluster name (*A*, *B*).

Dist.	*A, B*	*C*	*D*	*E, F*
A, B	0	?	?	?
C	?	0	22.19	32.96
D	?	22.19	0	10.77
E, F	?	32.96	10.77	0

? marks under a column or row means two elements are clubbed together at this point of calculation but distance from this new point to all other points are not known.

Using single linkage, the minimum distance between original of the two clusters is specified using the input matrix. From the distance matrix it is found that the closest distance between clusters happens between cluster *D* and (*E*, *F*) at distance 10.77. Again to cross check the closest distance between the cluster, distance between cluster (*E*, *F*) with cluster *A* and *B* is computed separately is as shown below;

Distance between cluster (*E*, *F*) and *A* is

$$d(E,F) \rightarrow A \quad = \min(d\ EA, d\ FA)$$
$$= \min(45.11, 46.42) = 45.11$$

Distance between cluster (*E*, *F*) and *B* is

$$d(E,F) \rightarrow B \quad = \min(d\ EB, d\ FB)$$
$$= \min(52.67, 53.98) = 52.67$$

Distance matrix – D_3

It can be seen that the closest distance between clusters happens between cluster *D* and (*E*, *F*) at distance 10.77. Thus, result in clustering these together into cluster ((*E*, *F*), *D*).

Dist.	*A, B*	*C*	*D*	*E, F*
A, B	0	12.15	34.34	34.34
C	12.15	0	22.19	32.96
D	34.34	22.19	0	10.77
E, F	45.11	32.96	10.77	0

Distance matrix – D_4

Dist.	*A, B*	*C*	*(E, F), D*
A, B	0	12.15	34.34
C	12.15	0	22.19
(E, F), D	34.34	22.19	0

The minimum distance appears between cluster (*A*, *B*) and *C* at distance 12.15. Thus, clustered them together ((*A*, *B*), *C*).Again to cross check the closest distance between the clusters, Distance between cluster (*A*, *B*) and ((*E*, *F*), *D*) is calculated as

$$d(A,B) \rightarrow ((E,F),D) \; = \min(d\,AE, d\,AF, d\,AD, d\,BE, d\,BF, d\,BD)$$
$$= \min(45.11, 46.42, 34.34, 52.67, 53.98, 41.9) = 34.34$$

Distance matrix – D_5

Dist.	((*A*, *B*), C)	((*E*, *F*), *D*)
((*A*, *B*), *C*)	0	22.19
((*E*, *F*), *D*)	22.19	0

$$\begin{aligned} D(A,B,C) \rightarrow ((E,F),D) &= d\ AE, d\ AF, d\ AD, d\ BE, d\ BF, d\ BD, d\ CE, d\ CF, d\ CD \\ &= \min(45.11, 46.42, 34.34, 52.67, 53.98, 41.9, 32.96, 34.27, 22.14) \\ &= 22.19 \end{aligned}$$

From the distance matrix, it has been found that ((*E*, *F*), *D*) and ((*A*, *B*), *C*) are merged into cluster {((*E*, *F*), *D*), ((*A*, *B*), *C*)}; Here, the cluster contains all the objects, and thus terminates the Agglomerative Hierarchical Clustering computation. Since it approaches the reversal algorithm of Agglomerative Hierarchical Clustering, the cluster {((*E*, *F*), *D*), ((*A*, *B*), *C*)} is assumed to be single cluster. In this study, Divisive Analysis Clustering (DIANA) is computed from distance matrix (D_6–D_{10}) to get last clusters contain single object that is (*A*, *B*, *C*, *D*, *E* and *F*).

Distance matrix – D_6

The single cluster (((*E*, *F*), *D*), ((*A*, *B*), *C*)) is split into cluster ((*E*, *F*), *D*) and ((*A*, *B*), *C*) at distance 22.19.

Dist.	((*A*, *B*), *C*)	((*E*, *F*), *D*)
((*A*, *B*), *C*)	0	22.19
((*E*, *F*), *D*)	22.19	0

$$d((E,F),D) \rightarrow ((A,B),C)$$
$$= \min(d\,EA, d\,EB, d\,EC, d\,FA, d\,FB, d\,FC, d\,DA, d\,DB, d\,DC)$$

Distance matrix – D_7

$$d((E,F),D) \rightarrow ((A,B),C)$$
$$= \min(45.11, 52.67, 32.96, 46.42, 53.98, 34.27, 34.34, 41.9, 22.19)$$
$$= 22.19$$

The cluster ((*A*, *B*), *C*) are split into cluster (*A*, *B*) and *C* at a distance 12.15.

Dist.	*A*, *B*	*C*	((*E*, *F*), *D*)
A, *B*	0	12.15	34.34
C	12.15	0	22.19
((*E*, *F*), *D*)	34.34	22.19	0

Distance matrix – D_8

The cluster ((*E*, *F*), *D*) are split into cluster *D* and (*E*, *F*) at distance 10.77.

Dist.	*A*, *B*	*C*	*D*	*E*, *F*
A, *B*	0	12.15	34.34	34.34
C	12.15	0	22.19	32.96
D	34.34	22.19	0	10.77
E, *F*	45.11	32.96	10.77	0

Distance matrix – D_9

The cluster (*A*, *B*) are split into cluster *A* and *B* at distance 7.56.

Dist.	*A*	*B*	*C*	*D*	*E*, *F*
A	0	7.56	12.15	34.34	45.11
B	7.56	0	19.71	41.9	52.67
C	12.15	19.71	0	22.19	32.96
D	34.34	41.9	22.19	0	10.77
E, *F*	45.11	52.67	32.96	10.77	0

Distance matrix – D_{10}

The cluster (*E*, *F*) are split into cluster *E* and *F* at distance 1.31 and it also created the single–single object, in all clusters (*A*, *B*, *C*, *D*, *E* and *F*). Thus terminate the Divisive Analysis Clustering (DIANA) computation.

Dist.	*A*	*B*	*C*	*D*	*E*	*F*
A	0	7.56	12.15	34.34	45.11	46.42
B	7.56	0	19.71	41.9	52.67	53.98
C	12.15	19.71	0	22.19	32.96	34.27
D	34.34	41.9	22.19	0	10.77	20.08
E	45.11	52.67	32.96	10.77	0	1.33
F	46.42	53.98	34.27	12.08	1.31	0

Results (DIANA)

1. In the beginning, the distance matrix (D_1–D_5) is computed by Agglomerative Hierarchical Clustering to get the cluster contains all the objects that is (((*E*, *F*), *D*), ((*A*, *B*), *C*)).
2. DIANA is reverse of Agglomerative Hierarchical Clustering followed by top-down approach applied in distance matrix (D_6–D_{10}). Cluster (((*E*, *F*), *D*), ((*A*, *B*), *C*)) are split into cluster ((*E*, *F*), *D*) and ((*A*, *B*), *C*) at distance 22.19.
3. The cluster ((*A*, *B*), *C*) are split into cluster (*A*, *B*) and *C* at a distance 12.15.
4. The cluster ((*E*, *F*), *D*) are split into cluster *D* and (*E*, *F*) at distance 10.77.
5. The cluster (*A*, *B*) are split into cluster *A* and *B* at distance 7.56.
6. The cluster (*E*, *F*) are split into cluster *E* and *F* at distance 1.31.
7. In the end, the single–single object is created, in all clusters (*A*, *B*, *C*, *D*, *E* and *F*).
8. The last clusters contain single object, thus, terminated.

3.2. Silhouettes

Silhouette value S is expressed for each object as follows.

$$S = (b - a)/\max(a, b)$$

Here, a particular object i is in cluster A and a is equal to the average dissimilarity of i to all other objects in A. For every other cluster not equal to A, cluster B has the smallest average dissimilarity between its objects and i which is equal to b. The cluster B is the nearest neighbor of objects i. A wide silhouette indicates large silhouette values and hence a pronounced cluster. The other dimension of a silhouette is its height, which simply equals to the number of objects in a group. In order to obtain an overview, the silhouettes of the different clusters are printed below each other. In this way the entire clustering can be displayed by means of a single plot, which enables us to distinguish clear-cut clusters from weak ones. The average of the silhouettes for all objects in a cluster is called the average silhouette width of that cluster. The average of the silhouettes for the entire data set is called the average silhouette width for the entire data set. The choice of optimal number of clusters is one of the most difficult problems of cluster analysis, for which no unique solution exists. This average silhouette width for the entire data set should be as high as possible and is used for the selection of optimal number of clusters. For application this maximum value of average silhouette width for the entire data set is called the silhouette coefficient. The silhouette coefficient is a dimensionless quantity which is at most equal to 1.

4. Study corridor and data collection techniques

4.1. Study corridors

The commercial capital of India, Mumbai Metropolitan is considered for this study. The metropolitan has linear pattern of transport network having predominant North–South commuter movements. South Mumbai houses various workplaces so during morning time people move toward South for work and during evening hours they returns toward 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 shown in Fig. 2.

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 provide 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 such as 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 h 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 was calculated for roads coming under the study area. It was found that free flow traffic condition (less than 200 veh/ln/h) 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 was collected using GPS receiver fitted on a probe vehicle during these hours. It is observed that the probe vehicle had maintained free flow speed between 40 km/h and 65 km/h on significant number of observed street segments and few segments had maintained free flow speed below 40 km/h or above 65 km/h. 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 have 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. Result and analysis

Data collected using the GPS handheld were transferred to office computer. The data were compiled following several steps: the GPS data were brought to GIS platform by using TransCAD software. The data set was viewed and observed before transferred to comma delaminated CSV file, from which it was copied to MS Excel. In MS Excel, the travel speed data under each segment were averaged over for each travel run. Average travel speeds over street segments were further averaged separately for peak and off-peak hours. Similarly the average free flow speeds of segments were calculated following the same principle. These average travel speed data were used as input to the DIANA program separately.

In the first step, free flow speed data were used in DIANA, and then input data (free flow speed) and output data (cluster centers) found from the analysis are used in computing validation parameters. The values of the Silhouette width obtained for 2–10 numbers of clusters are plotted in Fig. 3.

The value of Silhouette width is interpreted to obtain the optimum number of clusters in deciding the classification of street segments into number of Urban street classes. Lesser

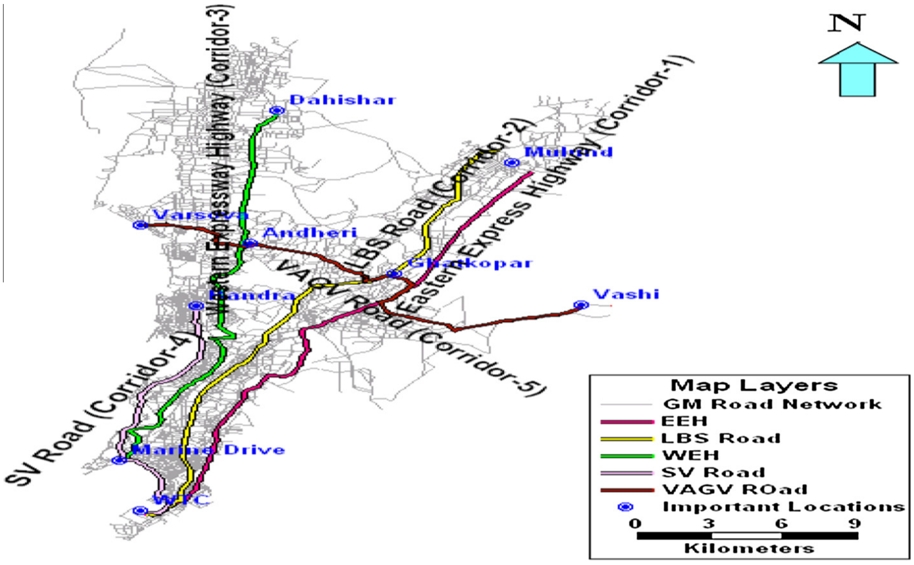

Figure 2 Map showing selected corridors of greater Mumbai.

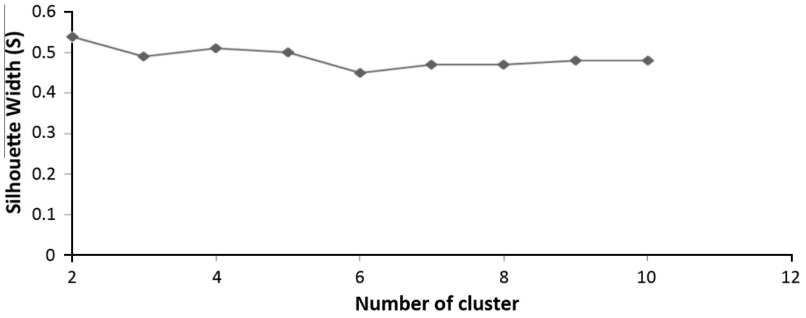

Figure 3 Validation measures for optimal number of clusters using DIANA clustering.

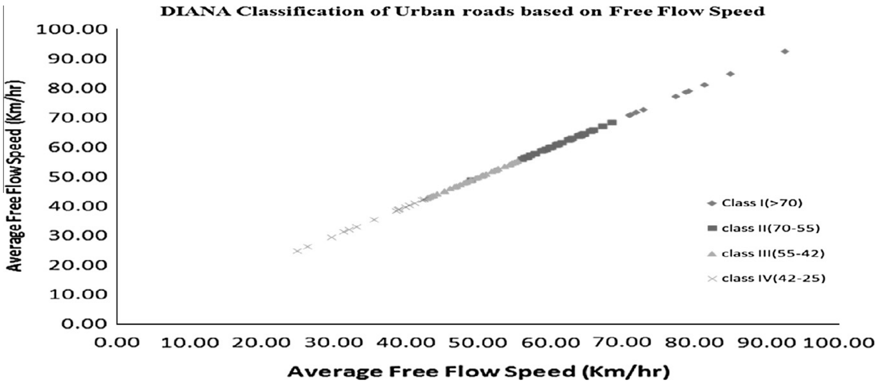

Figure 4 Classification of Indian urban roads based on free flow speed using DIANA.

number of clusters is preferably adopted by considering the difference in silhouette width prior to that number of cluster; hence four numbers of clusters are preferred over five in this case. Hence, considering Silhouette width for different numbers of clusters, the optimal number of clusters was found to be 4. So it has been decided to categorize the urban streets into four classes based on free flow speeds on DIANA clustering. Thus free-flow speed ranges of urban street classes are defined in Indian context as shown in Fig. 4. 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. From Fig. 4 it has been observed that FFS ranges of Urban Street IV are lesser compared to those values mentioned in HCM [1]. Large road network with a significant number of road segments is having varying road geometric characteristics and unplanned road side development compel vehicles to reduce the FFS to as low as 25 km/h. FFS on few road segments are significantly high; wide road with less road side development may help in attaining these speed limits. While substantial percentage of road segments are having FFS speed limits between 42 and 70 km/h, which shows most of road segments are of average type and users get average quality of service. This signifies that road network needs substantial improvement to provide better quality of service to the users.

Direction-wise average travel speeds calculated on street segments are used on DIANA clustering to define speed ranges for LOS categories for each class. In Fig. 5(A–D) the speed values are shown by different symbols depending onto which LOS category they belong. The legend in Fig. 5(A–D) gives

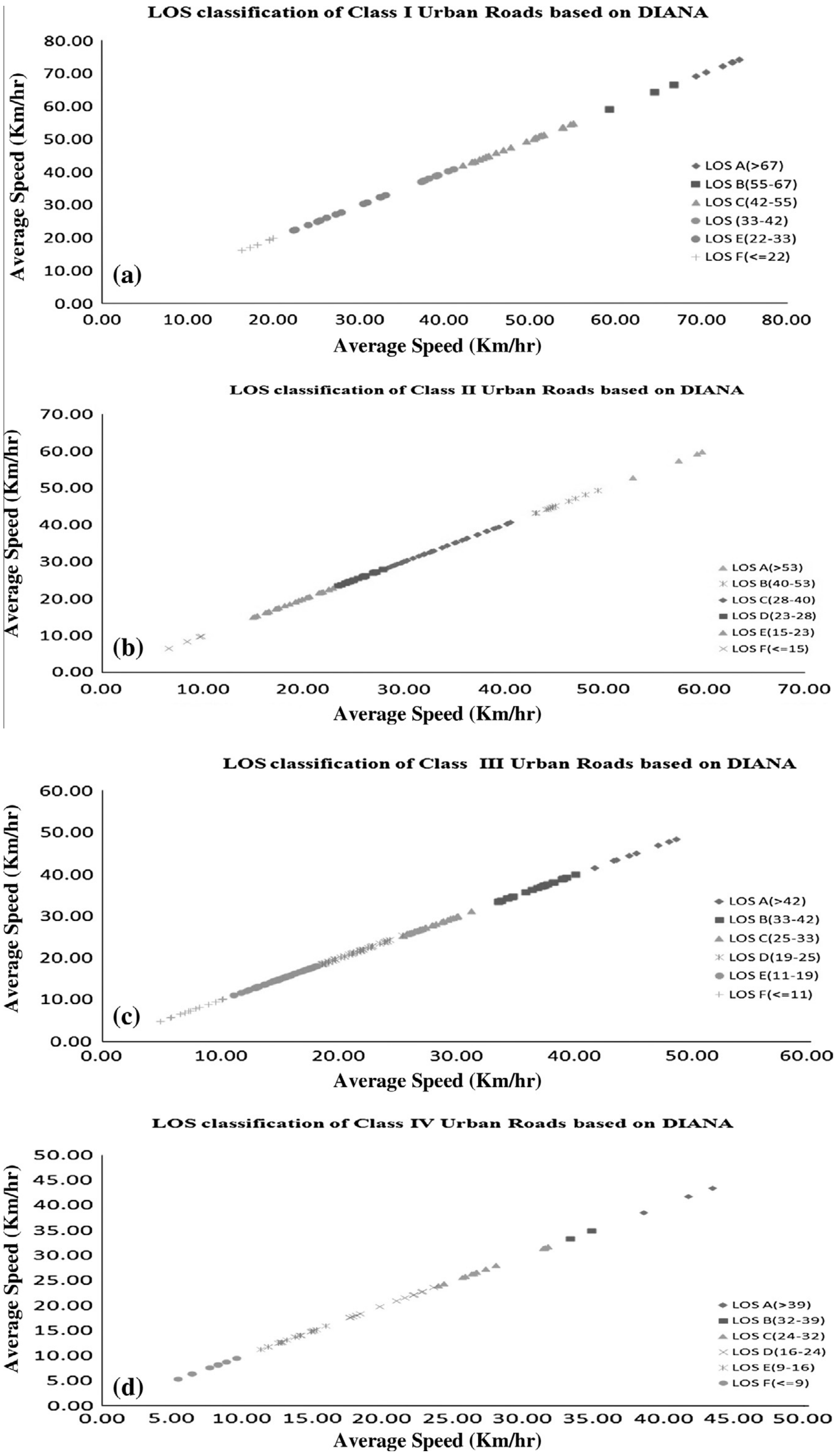

Figure 5 Level of service of urban street classes (I–IV) using DIANA clustering on average travel speed.

Table 1 Urban Street Speed ranges of LOS Categories using Divisive Analysis Clustering (DIANA).

Urban Street Class	I	II	III	IV
Range of free-flow speed (km/h)	93–70	70–55	55–42	42–25
Typical FFS (km/h)	75	60	50	40
LOS	Average travel speed (km/h)			
A	> 67	> 53	> 42	> 39
B	> 55–67	> 40–53	> 33–42	> 32–39
C	> 42–55	> 28–40	> 25–33	> 24–32
D	> 33–42.	> 23–28	> 19–25	> 16–24
E	> 22–33	> 15–23	> 11–19	> 9–16
F	⩽22	⩽15	⩽11	⩽9

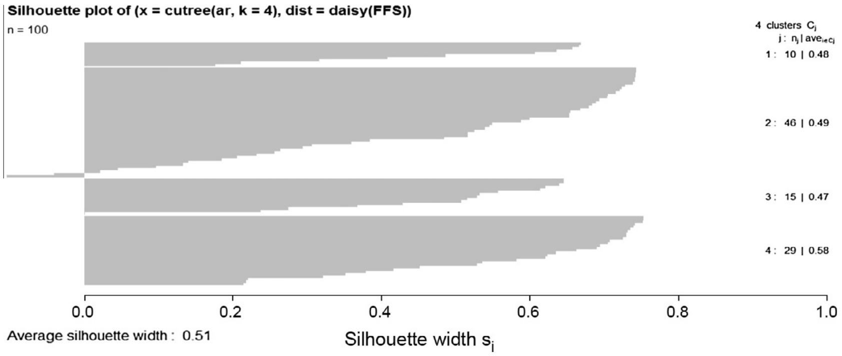

Figure 6 Silhouette plot of urban street class using DIANA clustering.

the speed ranges for six LOS categories of urban street classes obtained by using DIANA clustering and speed ranges are shown in Table 1. From Table 1 it can be stated that free-flow speed ranges of urban street classes and speed ranges of LOS categories based on DIANA clustering in Indian context are significantly different from the suggested values applicable to western countries with homogeneous traffic flow condition. Highly heterogeneous traffic flows with substantial percentage of vehicles are having slow moving with large dimension compel normal traffic to reduce the speed and hence result in lower speed ranges for LOS categories of urban streets. Also side friction developed due to commercial activity by roadside vendors, unwanted on-street parking, haphazard pedestrian activity coupled with random movements of street dogs etc. causes this reduction in speed limit of LOS categories.

Fig. 6 shows Silhouettes plot for urban street classes using free flow speeds. From this figure it is found that thickness of silhouettes of Urban Street Class II (shown as 2) and IV (shown as 4) is comparatively high which means large numbers of street segments are falling under Urban Street Classes II and IV. Also it is found that Silhouette width of data points under Urban Street Classes I–IV lies between 0.47 and 0.58. This indicates that free-flow speed data points form reasonably bonded within each urban street class. Also it is found that speed data points are well bonded within each level of service categories using silhouette validation parameter. A comparison between percent FFS values suggested in HCM [1,2] and result obtained in this research using DIANA clustering method has been illustrated in Table 2. It has been observed that speed ranges for level of service categories (A–F) expressed in percentage of free-flow speeds were found to be 90, 70, 50, 40, 25 and 20–25 respectively by applying Divisive Analysis Clustering (DIANA) method in the present study. On the other hand, in HCM [2] it has been mentioned that these values are 85 and above, 67–85, 50–67, 40–50, 30–40 and 30 and less percent respectively. It is observed that average travel speed expressed in terms of percentage of FFS by applying DIANA clustering method for LOS categories "E" and "F" is significantly lower than the mentioned in HCM [1] and for LOS categories from "A" to "D" remains same.

Table 2 Comparisons of percent FFS values for each LOS category.

Level of service	Typical % FFS (HCM [1])	% FFS (HCM [2])	Typical % FFS (DIANA Clustering method)
A	90	> 85	90
B	70	67–85	70
C	50	50–67	50
D	40	40–50	40
E	33	30–40	25
F	25–33	< 30	20–25

6. Summary and conclusion

In this research an attempt is made to define the LOS criteria of urban street of developing countries like India having heterogeneous traffic flow condition. From literature review it is known that GPS is an efficient tool for collecting large amount of speed data. Silhouette width was used as the validation parameters for the optimal number of cluster using FFS data for the classification of urban streets into number of classes. DIANA clustering algorithm was used twice on the collected speed data for the classification of the data set into number of groups. Finally, the classification of Urban Street was done and LOS was defined.

Considering the geometric and surrounding environmental characteristics it is well convinced to classify various urban street segments into four classes in Indian context. Hence, it agreed with the classification of street segments into four classes as mentioned in HCM [1]. From this study it was found that the free flow speed ranges of Urban Street Class-I to Urban Street Class-III are within the range defined by HCM [1] but Class-IV is significantly lower than that mentioned in HCM [1] because of heterogeneous traffic flow and roads having varying geometric and surrounding characteristics. The speed ranges of Levels of Service categories under the entire four urban street classes are found to be varying compared to the values mentioned in HCM [1] because substantial percentage of vehicles traveling on roads are slow moving. The HCM [2] describes that LOS is still a function of through vehicle speed. Determination of arterial segment running time is a function of environmental factors that influence free flow speed such as speed limit, access point density, area type, functional class, median type, curb presence and the presence of on-street parking. For LOS A speed range was found to be 90% of FFS which is 85% according to HCM [1], because some road segments are located in the suburban or outskirt to the city on which traffic flow is proportionately low which perhaps is due to lesser number of commuters use private owned vehicles rather prefer to use public transit for long distance travel. From the DIANA clustering of FFS data it can be seen that less number of roads in Mumbai is of high speed design (Street Class-I) or highly congested (Street Class-IV). More number of road segments are 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 burden on urban road infrastructure because of ever growing vehicular traffic volume. The other issues that can be noted are that high value of v/c ratio makes the roads less available for high speed movement of vehicles. Unwanted movement of pedestrians along and across the road sections creates a side friction and compelled travelers to reduce vehicular speed. A large number of Road side vendors and on-street parking occupy substantial portion of road sections result in high v/c ratio which reduces the service quality of road segments.

It has been observed that DIANA is a very successful clustering tool that be applied for all kinds of urban roads have varying traffic flow. 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. In the developing countries like India, there is heterogeneous traffic flow and the LOS definition makes significant role in improving traffic facilities. Also this study result has got strong application in transportation planning, operation and design. Rather using mid-sized vehicles for data collection purpose, this study can be extended to collect data using different types of vehicles and further study can be carried out.

Appendix A. (Level of service of urban street classes in different physical and traffic characteristics on Corridor-5)

Location of street segments		Free flow speed (km/h)	Geometric characteristics			Surrounding environmental characteristics				Average travel speeds (km/h) and Level of Service (LOS)							
Segment no.	Urban street class		Segment length (km)	No. of lanes	Median type	Access density	Road Side development	Parking	Ped. activity	MEW	LOS	MWE	LOS	EEW	LOS	EWE	LOS
1	67.28	5.441	6-LD	Concrete barrier	Very low	Low density	No	Very little	59.3	C	73.8	A	57.4	C	69.1	B	
2	III	50.39	2.449	6-LD	Concrete mountable	Moderate	Medium density	Significant	Some	43.3	B	35.8	B	33.8	C	14.4	E
3	II	57.86	1.568	6-LD	Concrete mountable	very low	Low density	Significant	Some	47.2	B	44.3	B	48.2	B	43.2	B
4	II	64.64	2.077	6-LD	Fence	very low	Low density	No	Very little	66.2	A	46.5	B	59.8	A	52.9	A
5	III	45.26	1.323	4-LD	Plantation	Moderate	Medium density	No	Little	17.5	E	43.5	B	29.2	C	33.5	C
6	IV	33.14	0.574	4-LD	Concrete mountable	High	High density	Some	Some	12.8	E	15.1	E	12.7	E	15	E
7	IV	32.12	3.383	3-LD	Concrete mountable	High	High density	Significant	Usually	14.2	E	15.1	E	11.9	E	11.8	E
8	III	47.52	1.032	4-LD	Metallic	High	Medium density	Significant	Some	17.4	E	17.6	E	12.3	E	16.3	E

9	III	47.53	0.277	4-LD	Concrete mountable	High	Medium density	Significant	Some	14.9	E	5.9	F	13.5	E	15	E
10	III	48.92	1.85	4-LD	Concrete mountable	High	Medium density	Significant	Some	15.6	E	14.7	E	20.9	D	19.1	D
11	IV	42.17	0.473	4-LD	Concrete mountable	High	Medium density	Some	Some	16	E	5.4	F	17.9	D	15.1	E
12	III	48.26	0.483	4-LD	Concrete mountable	High	Medium density	Significant	Some	33.4	C	13.8	E	6.4	F	18.1	D
13	IV	43.79	0.686	4-LD	Concrete mountable	High	Medium density	Significant	Some	9.6	E	15.4	E	6.6	F	23.3	D
14	IV	43.56	0.445	4-LD	Metallic	Moderate	Medium density	Significant	Some	14.2	E	22.8	D	20.2	D	10.1	E
15	IV	35.53	1.172	3-LD	Concrete barrier	Moderate	Medium density	Significant	Some	11.2	E	12.1	E	19.7	D	20.4	D
16	III	52.49	0.243	3-LD	Concrete mountable	Moderate	Medium density	Significant	Some	13.8	E	14.9	E	12.6	E	14.1	E
17	III	48.96	0.693	4-LD	Concrete mountable	Moderate	Medium density	Significant	Some	9.6	F	20.5	D	13.2	E	15	E
18	III	48.86	0.572	4-LD	Concrete mountable	Moderate	Medium density	Significant	Some	8.2	F	13.1	E	7	F	14.9	E
19	I	67.28	5.441	6-LD	Concrete barrier	Very low	Low density	No	Very little	17.1	F	16.2	F	22.1	F	12.2	F

Note: LD – Lane Divided, MEW – Morning-East–West, MWE – Morning-West-East, EEW – Evening-East–West, EWE – Evening-West-East.

References

[1] HCM, Highway Capacity Manual, Transportation Research Board, Washington DC, 2000.

[2] HCM, Highway Capacity Manual, Transportation Research Board, Washington DC, 2010.

[3] C.R. Patel, G.J. Joshi, Capacity and LOS for urban arterial road in Indian mixed traffic condition, Proc. Soc. Behav. Sci. 48 (2012) 527–534.

[4] V.T. Arasan, P. Vedagiri, Simulating heterogeneous traffic flow on roads with and without bus lanes, J. Infrastruct. Syst. 15 (4) (2009) 305–312.

[5] A. Polus, M. Cohen, Theoretical and empirical relationships for the quality of flow and for a new level of service on two lane highways, J. Transp. Eng. 135 (6) (2009) 380–385.

[6] P.K. Bhuyan, M.S. Nayak, A review on level of service analysis of urban streets, Transp. Rev. Trans. Transdiscipl. J. 33 (2) (2013) 219–238.

[7] T. Dandan, W. Wei, L. Jian, B. Yang, Research on methods of assessing pedestrian level of service for sidewalk, J. Transp. Syst. Eng. IT 7 (5) (2007) 74–79.

[8] C. Shouhua, Y. Zhenzhou, Z. Chiqing, Z. Li, LOS classification for urban rail transit passages based on passenger perceptions, J. Transp. Syst. Eng. IT 9 (2) (2009) 99–104.

[9] F.C. Fang, K.K. Pecheux, Fuzzy data mining approach for quantifying signalized intersection level of services based on user perceptions, J. Transp. Eng., ASCE 135 (6) (2009) 349–358.

[10] J. Lee, W. Lam, Level of service for stairway in hongkong underground stations, J. Transp. Eng. 129 (2) (2003) 196–202.

[11] M. Azimi, Y. Zhang, Categorizing freeway flow conditions by using clustering methods, J. Transp. Res. Board, Transp. Res. Board Nat. Acad. 2173 (2010) 105–114.

[12] J. Lee, P. Goh, W. Lam, New level of service standard for signalized crosswalks with bi-directional pedestrian flows, J. Transp. Eng. 131 (2005) 957–960.

[13] R.P. Roess, M.A. Vandehey, W. Kittelson, Level of service 2010 and beyond, J. Transp. Res. Board, Transp. Res. Board Nat. Acad. 2173 (2010) 20–27.

[14] B. Ko, S. Washburn, D.S. McLeod, Performance measures for truck level of service an exploratory survey analysis, J. Transp. Res. Board, Transp. Res. Board Nat. Acad. 2130 (2009) 120–128.

[15] X. Chen, D. Li, N. Ma, C. Shao, Prediction of user perceptions of signalized intersection level of service based on fuzzy neural networks, J. Transp. Res. Board, Transp. Res. Board Nat. Acad. 2130 (2009) 7–15.

[16] S. Cao, Z. Yuan, Y. Li, P. Hu, L. Johnson, Study on service-level classification of platforms in beijing urban rail transit, J. Transp. Res. Board, Transp. Res. Board Nat. Acad. 2112 (2009) 127–135.

[17] J.D. Zegeer, M. Blogg, K. Nguyen, M. Vandehey, Default values for highway capacity and level-of-service analyses, J. Transp. Res. Board, Transp. Res. Board Nat. Acad. 2071 (2008) 35–43.

[18] R. Dowling, A. Flannery, B. Landis, T. Petritsch, N. Rouphail, P. Ryus, Multimodal level of service for urban streets, J. Transp. Res. Board, Transp. Res. Board Nat. Acad. 2071 (2008) 1–7.

[19] S. Kikuchi, P. Chakroborty, Frameworks to represent the uncertainty when determining the level of service, Transp. Res. Rec. 1968 (2007) 53–62.

[20] K. Choocharukul, K.C. Sinha, F.L. Mannering, User perceptions and engineering definitions of highway level of service. An explanatory statistical comparison, Transp. Res. Part A 38 (2004) 677–689.

[21] A.M. Semeida, New models to evaluate the level of service and capacity for rural multi-lane highways in Egypt, Alex. Eng. J. 52 (2013) 455–466.

[22] T. Becher, A new procedure to determine a user oriented level of service of traffic light controlled crossroads, Proc. Soc. Behav. Sci. 16 (2011) 515–525.

[23] M. Ameri, R. Moayedfar, F. Jafari, Determine the capacity of two lane suburban roads with neural networks and effect of speed on level of service, Eur. Transp. Res. Rev. 5 (4) (2013) 179–184.