Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Bicycle Comfort Level Rating (BCLR) Model for Urban Street Segments in Developing Countries

Sambit Kumar Beura[a], Haritha Chellapilla[a], Prasanta Kumar Bhuyan[a,*]

[a]*Department of Civil Engineering, National Institute of Technilogy Rourkela, Odisha 769008, India*

Abstract

An in-depth understanding of on-street bicyclists' perceived sense of comfort under heterogeneous traffic flow conditions is very much complex and important as well. The present study explores these concerns and proposes a "Bicycle Comfort Level Rating" (BCLR) model for the assessment of urban street segments carrying heterogeneous traffic in mid-sized cities. Variables on roadway geometrics, built-environments and traffic flow conditions were collected from a large number of 60 road segments located in three Indian cities. Socio-demographic information, and perceived comfort ratings of around 150 bicyclists were also collected from each segment. Nine quantitative variables having significant ($p < 0.001$) influences on bicyclists' perceived comfort levels were identified and included in a step-wise regression analysis. The resulting model is highly reliable in the present context and has reported a high coefficient of determination (R^2) value of 0.87 with averaged observations. A sensitivity analysis carried out on the BCLR model parameters has reported that, motorized and non-motorized traffic volumes, pavement conditions and roadway width are by far the most influencing parameters and are contributing 20.21%, 17.37%, 11.91% and 11.49% respectively to the prediction of comfort ratings. Genetic Programming (GP) clustering was used to define the ranges of comfort levels A-F (excellent-worst). Results showed that, around 97% of all studied segments are offering average ('C') or inferior comfort levels at the present scenario. The BCLR model along with other findings of this study may assist the transportation planners and engineers in taking judicious decisions to establish bicycle-friendly road infrastructures.

Keywords: Urban road segment; Heterogeneous traffic; Bicycle level of service; Genetic programming clustering

1. Introduction

In many countries around the world, the developmental stages of road infrastructures have been primarily focusing on the safe management of motorized traffic. Conversely, the perceived comfort levels of bicycle users are being

* Corresponding author. Tel.: +91-661-246 2349; fax: +91-661-246 2349.
E-mail address: pkbtrans@gmail.com

highly neglected. Facilities favourable for bicycle use such as separate bicycle lanes, shared-use paths and wide shoulders are not frequently available on the street segments. As a result, bicyclists are inevitably using the lanes available on the main carriageway to fulfil their mobility requirements. However, these lanes are predominately used by motorists and thus bicyclists are cogently deprived from availing the minimum space desired for a smooth riding. Under such conditions, bicyclists are encountering a very complex interaction with various categories of vehicles widely varying in their sizes and operational characteristics. An in-depth understanding of bicycle operations under such an environment is very much complex and important as well. An innovative approach for the investigation of operational characteristics of bicyclists measured through comfort ratings is presented in this paper. Researchers in developed countries have made significant effort to explore these concerns under homogeneous traffic flow conditions only. However, in case of developing countries, the traffic flow on the main-carriageway is highly heterogeneous where vehicle users do not follow the lane discipline. Thus, the findings of previous studies cannot fulfill the mobility requirements under heterogeneous traffic conditions. So far, the modelling of bicyclists' perceived comfort levels under such traffic flow conditions has not been given a systemic approach. These models in fact play vital roles while making any plan of actions for the enhancement of service qualities offered by existing transportation facilities. These models also assist while designing new bicycle-friendly road networks.

The present study has primarily focused on the development of a "Bicycle Comfort Level Rating" (BCLR) model thorough an in-depth investigation of bicycling behavior under prevailing road conditions in Indian mid-sized cities (population size within 0.5-1.0 million). For analyses purposes, a large quantity of data sets (roadway geometric details, built environmental characteristics, traffic flow parameters and socio-demographic information of users) were collected from a large number of 60 road segments. These segments are located in different parts of India and well represent the variability and complexities persisting in the bicycling environment. Each studied segment was rated by 150 bicycle users from varying socio-demographic backgrounds based on its ability to satisfy on-street bicyclists. The ratings were collected using a 6-point Likert scale (1 = 'excellent' and 6 = 'worst') and named as "Bicycle Comfort Level Ratings" (BCLRs). A thorough investigation was carried out to identify which attributes have significant influences on the perceived BCLRs in the present context.

Preliminary data analysis showed that nine quantitative variables (road attributes) are contributing to the perceived comfort level of bicyclists in the mid-sized cities. Though there were some tendencies, socio-demographic attributes such as gender, age, bicycling experiences and trip purpose had insignificant contribution in the present context. Thus, quantitative variables road attributes were only inputted in a step-wise regression analysis to build the BCLR model. The resulting model is highly reliable and has a high coefficient of determination (R^2) of 0.86 with averaged observations. Nash–Sutcliffe model efficiency coefficient (E) (Nash and Sutcliffe, 1970), error measuring parameters and several other statistical parameters were also applied to assess the goodness-of-fit and prediction precision of the model. A sensitivity analysis was carried out to identify the relative importance of input parameters in predicting BCLRs of urban street segments. These results will help the transportation planners and engineers in deciding which parameters should be primarily prioritized in order to enhance the comfort levels of bicyclists. Obtained BCLRs were classified into six comfort classes ranging from 'A' (excellent) to 'F' (worst). For this purpose, one of the most recent and reliable cluster techniques namely, Genetic Programming (GP) clustering was used in this study. This defined scale in combination with the BCLR model would help to determine what level of bicycle comfort a desired segment is offering at its present scenario. A detailed discussion of each of above aspects are given in succeeding sections.

2. Review of Literature

Existing bicycle service models are primarily based on homogeneous traffic conditions and are not transferable to heterogeneous traffic conditions in developing countries. However, a detailed review of those models provided a strong basis for carrying out the present research work. Crucial findings of these studies are summarized in the succeeding sub-section. This is followed by a brief introduction on the Genetic Programming (GP) cluster technique used in this study to define the ranges of bicycle comfort levels (A-F).

2.1. Existing bicycle models

Several studies have been conducted in the recent years to relate roadway geometrics and homogeneous traffic flow conditions with the operational characteristics of bicyclists. Bicycle Safety Index Rating (BSIR) model (Davis, 1987) is the initial model developed to assess the bicycle service criteria of street segments. This model is comprised of two sub-models namely, Roadway Segment Index (RSI) and Intersection Evaluation Index (IEI) model. RSI model is a function of traffic volume, number of lanes, speed limit, width of outside traffic lane, pavement conditions and location factors. This model neglects the influences of percentage of heavy vehicles and on-street parking turn-over, etc. Modified Roadway Condition Index (RCI) model (Epperson, 1994) is a modified version of the RSI model. In this model, the location and pavement factors were modified and the lane width term was multiplied by speed limit to place greater weightage on narrow roads with high vehicular speeds. Bicycle Suitability Rating (BSR) model (Davis, 1995) is also a modified version of RSI model which signifies the important roles of traffic volume and traffic speed in bicycle service criteria.

Developers of another bicycle service model namely, Interaction Hazard Score (IHS) model (Landis, 1994) identified the important roles of two more influencing variables namely roadside land use intensity and curb cut (on-street parking) frequency. Further, the importance of curb-lane was reflected in Bicycle Stress Level (BSL) model (Sorton and Walsh, 1994). This model exclusively considers three different parameters of the curb-lane to define the bicycle service criteria such as curb-lane width, curb-lane traffic volume and average traffic speed in the curb-lane. BSL model was further improvised and BCI model (Harkey et al., 1998) was developed. BCI model introduced some new influencing parameters such as bicycle lane parameters and right turning vehicles parameters. Highway Capacity Manual (HCM, 2010) considered a broad range of factors such as effective width of the outside through lane, mid-segment demand flow rate, number of through lanes, motorized vehicle running speed, percentage of heavy vehicles and pavement condition for defining the bicycle service criteria.

Some of the other studies have revealed that, the interference from pedestrians and non-motorized vehicles considerably degrade the bicycle service quality (Mozer, 1994) while, well maintained pavement surfaces and the provision of bicycle lanes positively influence the same (Chellapilla et al., 2016; FDOT, 2009; Harkey et al., 1998; Jensen, 2007; Landis, 1997). Frequency of driveways also have considerable negative influence on bicyclists' perceived sense of comfort (Davis, 1987; Davis, 1995; Epperson, 1994; Harkey et al., 1998; Landis, 1994). Traffic volume largely influences bicyclists' perceived comfort levels under heterogeneous traffic flow conditions (Beura et al., 2016); and with the provision of separate bicycle lanes, bicyclists gain better confidence to ride further from the edge of roadways (Hallett, 2006; Hunter 2005).

From the above discussions it could be summarized that, several roadway geometric and traffic flow variables have considerable influences on bicyclists' perceived sense of comfort. The role of individual variables changes with the change in roadway environmental conditions. In this regard, an extensive investigation was carried out in this study to identify the influencing variables under heterogeneous traffic flow conditions. In addition to the variables considered in previous studies, the role of few more variables (for an example roadside stoppages of intermittent public transits) were also investigated. Subsequently, a new bicycle service prediction model was developed which is basically a new decision support system for the transportation planners and designers.

2.2. Cluster technique

Traditionally, researchers have been defining the ranges of bicycle service categories (A-F) by taking the overall perceived satisfaction score as a boundary between service classes 'C' and 'D' and further deriving the ranges of all classes at equal difference. This method distinguishes the service classes at equal intervals and does not assess the similarity or dis-similarity that does exist among a group of data points (service scores). In contrast, the clustering techniques help to recognize several distinct modalities persisting within the data set. They assemble a set of data in such a way that, data in the same group (or cluster) are more similar to each other than to those in other groups. Objects inside a cluster are closer to the 'centre' of the same cluster than to the centre of other clusters. A good clustering method produces clusters with the property that their intra-cluster distance is smaller and inter-cluster distance is larger. Cluster analysis has been proved as a powerful method for classifying traffic engineering data on which deterministic modelling and regression analysis have been previously applied (Beura et al., 2017; Prassas, 1996).

The present study has defined the ranges of comfort levels (A-F) in an advanced and meaningful way by using the Genetic Programming (GP) clustering technique. GP is an evolutionary algorithm and an extension of the genetic algorithm (GA) invented by Holland (1975). The idea of evolving programs was proved, promoted and developed into a practical tool by Koza (1992). The existing literatures about application of GP as a tool for classification has been surveyed and an idea about the different ways in which it can help in the construction of accurate and reliable classifiers has been reported by Cristobal et al. (2013). Previously, Patnaik and Bhuyan (2016) have applied the GP clustering tool on speed data to define the service classes of urban streets.

3. Research Methodology

In the present study, the Step-wise regression analysis is used to develop the BCLR model and the Genetic Programming (GP) cluster tool is used to define the ranges of bicycle comfort levels (A-F). In practice, the regression-based models are more intuitive and very much easy to be followed by the engineers, planners and administrators. Non-modelers also well understand the sensitivity of regression models (percentage change in the model output with percentage change in the model inputs). On the other hand, the GP clustering tool is one of the recently introduced and most reliable cluster techniques available for classification purposes. Following sub-sections briefly describe the basic principles of step-wise regression analysis, statistical parameters used to assess the prediction precision of the resulting model, and the basic principles of GP cluster tool.

3.1. Step-wise regression analysis

The step-wise regression analysis was carried out in following four major steps:

- Variables having significant ($p < 0.001$) influence on bicyclists' perceived comfort levels were identified through the application of Spearman correlation analysis
- Several combined and/or transformed forms (e.g. power, inverse, square root, exponential and logarithmic) of the identified variables were developed and included in the regression analysis
- Significance of the coefficients obtained for original variables as well as their combined and/or transformed forms were tested using *t*-test
- The regression model one which well satisfied the required significance criteria was selected and reported

The mathematical expression showing the relationship among BCLRs of a road segment and its important attributes is presented in Eq. (1).

$$BCLR = a_1 f(X_1) + a_2 f(X_2) + ... + a_n f(X_n) + c \qquad (1)$$

Where, each of $X_1, X_2, \ldots, X_n$ represent an individual input variable or combination of two or more variables, $a_1, a_2, \ldots, a_n$ represent the estimated coefficients, and c represent the constant term.

Backward elimination process was chosen during the regression analysis because of the advantage that, in this process computer system first places all variables in the expected model and then examines the contribution of each predictor. If *t*-statistics of any predictor is not significant, then it removes it from the analysis and re-estimates the model. Then, contributions of remaining predictors are reassessed and the process continues till all significant predictors are identified. Hence, the chance of missing a significant variable in the modelling process is negligible.

3.2. Prediction performance assessing parameters

Several statistical parameters discussed below are used in this study to assess the prediction performance of BCLR model.

(a) Nash–Sutcliffe efficiency coefficient (E): Though coefficient of determination (R^2) is commonly used in various fields of research to assess the efficiency of predictive models, the same parameter has been criticized by several researchers as being a biased estimate. Hence, another statistical parameter, E (Nash and Sutcliffe, 1970), is also used in this study to test the performance of BCLR model. *E*, defined below, compares actual and predicted values of the output and evaluates how good the model is able to explain total variance in the data.

$$E = (E_1 - E_2)/E_1 \tag{2a}$$

where

$$E_1 = \sum_{i=1}^{s} (BCLR_{PC} - \overline{BCLR_{PD}})^2 \tag{2b}$$

$$E_2 = \sum_{i=1}^{s} (BCLR_{PC} - BCLR_{PD})^2 \tag{2c}$$

and $BCLR_{PC}$, $\overline{BCLR_{PC}}$, $BCLR_{PD}$ are the perceived, average of perceived and predicted values of BCLRs respectively, and *s* is the total number of observations (i.e., 45 and 15 in training and validation stages respectively).

(b) Error measuring parameters: Average absolute error (AAE), maximum absolute error (MAE) and root mean square error (RMSE) of $BCLR_{PC}$ and $BCLR_{PD}$ were used in this study to assess the errors in model predictions. The less is the error better is the prediction precision of a model.

(c) Arithmetic calculations of $BCLR_{PD}/BCLR_{PC}$ (i.e., the ratio of $BCLR_{PD}$ and $BCLR_{PC}$) values: Under ideal conditions, i.e. when $BCLR_{PD}$ values are equals to the corresponding $BCLR_{PC}$ values, the mean (μ) of $BCLR_{PD}/BCLR_{PC}$ values becomes 'one' and standard deviation (σ) of the same becomes 'zero'. Thus, the developed BCLR model is expected to produce a μ value very close to 'one' and a σ value close to 'zero'. Also, μ value above 'one' indicates over predictions and below 'one' indicates under predictions of the model.

(d) 50% and 90% cumulative probabilities (P_{50} and P_{90} respectively) of $BCLR_{PD}/BCLR_{PC}$ values: The value of P_{50} below 'one' implies under predictions and above 'one' implies over predictions of a model. Similarly, P_{90} parameter reflects the variation in the values of $BCLR_{PD}/BCLR_{PC}$ for the total observations. Thus, P_{50} and P_{90} values close to 'one', indicate a high efficiency of the predictive model. In order to estimate these parameters, the $BCLR_{PD}/BCLR_{PC}$ values of all investigated segments were calculated and arranged in an ascending order. Further, their cumulative probability (*P*) was calculated by using the following equation:

$$P = i/(s+1) \tag{3}$$

where *i* is the index or serial number of the segment. By using the obtained results, P_{50} and P_{90} values were estimated.

(e) Overfitting Ratio (OR): Quantitatively, the generalization ability of any empirical model is represented as the 'OR' value calculated as the ratio of RMSE in validation stage to the same in training stage. 'OR' value equal to 'one' indicates perfect generalization of an empirical model.

3.3. Genetic programming (GP) clustering

In GP, genetic algorithm (GA) are the search algorithm that mimic the process of natural evolution where each individual is a candidate solution. It explores the algorithm search space and evolve computer programs to perform a defined task. Computer programs are developed which are usually represented in the memory as tree structures. Trees can easily be evaluated in a recursive manner. For easier development and evaluation of the mathematical expressions, every terminal node possesses an operand and every tree node possesses an operator function. Thus, GP usually favors the use of programming languages that naturally embody tree structures. The GP clustering technique was executed in the following five steps to classify BCLRs into six comfort levels (A-F):

- Initial population of the candidate solutions or individuals (i.e., BCLRs) were taken for clustering. The tree node representing class and terminal node representing end class were used to construct a legal tree. Each individual was formed by the clustering trees and each of them is associated with one of the cluster which is to be formed.
- The object centroid distance was calculated (by Euclidean distance formula) which gave the fitness value for which the individuals had higher chance to survive. Then each BCLR was assigned to a certain group (A, B,…, or F) according to minimum distance.

- The iteration process was adopted to reach the goal matrix or maximum generation. In each step of the iteration processes, new centroids of each group as well as new object centroid distances were calculated based on new members of BCLRs those got accumulated in that group.
- New distance matrices were formed with respect to the minimum distance obtained. This process was iterated till a new distance matrix was achieved with same values as obtained in the previously produced matrix. This indicated that the object would not change its group any more and the computation of the cluster analysis has been reached its stability. Thus, the iteration process was stopped at this point.
- Each BCLR was assigned to the desired group (A, B,…, or F) according to the new minimum distance. In this way the final classification of BCLRs were achieved by the application of GP clustering technique.

4. Site Selection and Data Collection

The basic principle adopted in selecting suitable sites for the data collection was “the variabilities and complexities persisting in the urban bicycling environments of midsized cities should be reflected in the sample”. In order to accomplish this, about 60 road segments were identified in this study which varied from each other with respect to following four major criteria.

(1) **Location inside the city:** Average speed, volume and composition of the vehicular traffic on any road network is significantly influenced by its location inside a city. For an example, the flow of heavy vehicles is often restricted on the road networks joining the city centers. Conversely, roads at outskirts carry heavy vehicles along with significant percentage of four-wheelers at a very high operating speed. The proportion of independent modes of transportation are generally dominant on commercial vehicles on roadways situated at residential areas. Conversely, road networks in commercial areas generally attract a significant number of commercial vehicles from different parts of the city. The traffic volume on these roadways is also normally higher than those in residential areas. The location of the roadways also has significant influence on several other properties such as roadside developments, on-street parking activities and pedestrian volume, etc. In this regard, road segments from different parts of the cities were considered in this study for data collection.

(2) **Geometric details:** The performance of a roadway is hugely determined by its geometrical elements such as umber of lanes, width of the carriageway, presence and width of bicycle lane, paved shoulder and median, etc. Thus, the selected sites were made represent wide variations persisting in the geometric design standards of the existing road networks in mid-sized cities of a developing country like India.

(3) **Traffic volume and its composition:** In mid-sized cities of developing countries, separate bicycle lane facilities are seldom available due to which bicyclists inevitably use the lanes available on the main-carriageway. Under such situations, bicyclists encounter a very complex interaction with various categories of vehicles. Hence, road segments carrying very low as well as very high volume of traffic were included in the study area. The traffic volume as well as its composition normally changes with differing transportation facilities in different parts of a nation. Thus, roadway segments from three different Indian states were included in the study corridors. Sites selected in this study carry around 22 types of vehicles at their present scenario.

(4) **Roadside developmental pattern:** The roadside developmental pattern also has a significant influence on quality of bicycling. This factor hugely influences travel demand, on-street parking activities, and several other factors. In this regard, road networks form residential areas, rural fields, office or institutional areas, commercial areas as well as industrial areas were considered in the present study.

Sixty different segments selected in this study for data collection purposes are basically the parts of road networks of three Indian mid-sized cities namely, Bhubaneswar (Capital of Odisha state), Rourkela (Steel city of Odisha state) and Kottayam (Kerala state). These cities are geographically located in different parts of the country as shown in Fig. 1. Bhubaneswar, the capital city of Odisha state, is primarily an administrative city. It is the largest city of the state and is commonly known for its economic and religious importance. In the recent past, it has emerged as an information technology (IT) and education hub. Rourkela city is another large city of Odisha state, which is well-known for the first ever steel plant integrated into the public sector in India. On the other hand, Kottayam is a municipal town in the Indian state of Kerala and is the administrative capital city of Kottayam district.

Fig. 1. Locations of Indians cities covering the study area.

From the selected 60 segments, 29 segments belong to Bhubaneswar city, 19 segments belong to Rourkela city and remaining 12 segments belong to Kottayam city. These segments are basically the parts of 2, 4, 6 and 8-laned urban road corridors. Roughly 50% of road segments have the provision of separate sidewalk facilities. The traffic flow on these roads is highly heterogeneous and thus the average travel speed varies in a lower range of 25-45 km/h. The vehicular traffic volume normally varies from 350 PCUs/h up to 5800 PCUs/h during the peak hours. Heavy vehicles contribute up to 7% of the total traffic. The vehicular movement on few segments is restricted to one-way and most of the segments allow two-way movement. The roadside land-use patterns are very much diversified and include residential, commercial, office and industrial developments. City roads are often characterized by legal or illegal on-street parking activities. The vehicular ingress-egress to the on-street parking area often varies up to 4000 veh/h/km during the peak hours. The overall data sets collected in this study for the development of BCLR model include both quantitative and qualitative data which are discussed in the following sub-sections. Approximately 50 working days' data sets were collected in between September 2015 to March 2016.

4.1. Quantitative data

Quantitative data sets basically include geometric, traffic and other built-environmental attributes. During the inventory survey, geometric attributes such as: width of carriageway, shoulder, parking lane, sidewalk, median kerb and gutter were measured using measuring tapes. The pavement surface conditions were rated using a 5-point scale where, 5 = excellent and 1 = worst. Roadside land-use pattern was rated using a 3-point scale where, 1 = highly commercial, 0.5 = moderately commercial, and 0 = minimal or non-commercial. The operating speed of motor vehicles on Indian roads under mixed traffic conditions is not as high as in developed countries; and a large variation exists among speeds of vehicles. Hence the speed measures such as: spot speed or space mean speed that are normally calculated for homogeneous traffic, should not be considered for mixed traffic situations. In this regard, the mid-segment traffic flow on each segment was videotaped during the peak hours (i.e., either morning 8:30-11:00 AM or evening 4:00-6:30 PM) over a longitudinal trap of 30 meter. The average time taken by motor-vehicles to cross this trap was extracted with an accuracy of 0.1 second. Subsequently, the average traffic speed on each segment was calculated by dividing the crossing distance (30 meter) by the average crossing time of motor-vehicles. Other variables collected during inventory study and visual inspections are: total number of driveways connecting the segment; approximate vehicular ingress volume (veh/h) to each driveway during the peak hours; and interruptions caused from authorized/un-authorized stoppages of intermittent public transits (1 = high, 0.5 = medium, 0 = minimal).

Recorded video clips were also used to extract several other parameters such as traffic volume, pedestrian volume (ped/h), percentage of heavy vehicles (%), average vehicular ingress-egress to the on-street parking area (veh/h/km) and average headway of on-street vehicular encounters (min). In this study, the running average method was used to determine the peak hour traffic volume on each segment. In order to bring all categories of motor-vehicles into a single measuring unit, Passenger Car Unit per hour (PCUs/h), volumes of different categories of motor-vehicles were multiplied by corresponding PCU values recommended by Indian Road Congress code of practice 106 (IRC, 1992). However, non-motorized traffic volume (*NMV*) was expressed in bicycles/h by using Eq. (2). In this equation, the number of non-motorized vehicles other than bicycles has been multiplied by a weightage factor of 'four' because, the PCU values of these vehicles as recommended by the IRC-106 are approximately four times of that of a bicycle.

$$NMV = \text{No. of bicycles} + (4 \times \text{No. of non-motorized vehicles other than bicycles}) \quad (4)$$

4.2. Qualitative data

Qualitative data sets include, (1) perceived comfort levels of bicyclists with the studied segments, and (2) socio-demographic information of bicyclists whose perceived comfort levels are assessed. In order to obtain these data sets, an innovative questionnaire was designed and an extensive roadside interview of bicycle users was carried out. Both regular and occasional bicycle users were selected from the field itself and included in the perception survey. Although this survey approach is a very time consuming and costly method while gathering a large quantity of data, it reflects realistic responses of the participants. The participants of this survey are possibly well experienced with the prevailing road conditions and are assumed as the best examiners of bicycling comfort levels. Bicyclists those who were involved in this survey were made sure to have a minimum one year of bicycling experience on the urban roadways.

On each segment, at least 150 effective participants from different socio-demographic backgrounds were interviewed. The survey was normally conducted at the end point of each segment, and bicyclists those who recently covered the full length of the desired segment were interviewed. Each participant was provided with the questionnaire sheet mentioned earlier. Socio-demographic information and driver characteristics were collected by ensuring the participants that all information would be used for academic purpose only. These qualitative attributes include gender, age, household size, income, educational background, bicycling experiences (years), daily average bicycling distance, and bicycle trip purpose. Scales used to collect these attributes are shown in Table 1. The percentage distributions of survey participants shown in this table ensure that, this study has considered a significant variation in socio-demographic attributes as well as driver characteristics.

After obtaining the said information form a participant, he/she was asked to rate the desired segment based on a simple question: "How would you rate the overall comfort level that you perceived while using bicycle on the road segment?" The responses (BCLRs) were collected using a 6-point Likert scale where, 1 = 'excellent' and 6 = 'worst' level of perceived comfort. This survey resulted in a total of 9,000 (150 participants × 60 segments) effective BCLRs for all studied segments. Obtained BCLRs had a mean of 3.47 and standard deviation of 0.66. For checking the data sufficiency, Cochran's sample size formula (Cochran, 1997) was used and the allowed error in estimation of the mean perceived BCLRs was calculated. The error in using this amount of data set was limited within 1% as estimated at 95% confidence level. This minimal error showed the sufficiency in data sets collected in this study.

5. Analysis and Results

The discussions included in this section are exclusively focused on selection of variables, development and significance tests of BCLR model, scopes for its applications around the world, and defining the ranges of bicycle comfort levels (A-F) through the application of Genetic Programming (GP) clustering tool.

5.1. Variables selection

In this study, Spearman's correlation analysis was carried out to identify which variables are significantly affecting the perceived comfort levels of on-street bicyclists. A wide array of both quantitative variables (field observed road attributes) and qualitative attributes (socio-demographics and driver characteristics) were used as independent variables and perceived BCLRs were used as the dependent variable. Results showed that, nine road attributes are significantly ($p < 0.001$) affecting the perceived comfort levels of on-street bicyclists in the present context. These variables along with their notations and descriptive statistics are shown in Table 2.

Table 1. Descriptions and scales of qualitative attributes.

Attribute	Scale and Distribution	Percentage
Gender	1 = Female	42.0%
	2 = Male	58.0%
Age	1 = ⩽ 20	4.6%
	2 = 21–30	38.0%
	3 = 31–40	26.0%
	4 = 41–50	17.4%
	5 = 51–60	10.0%
	6 = ⩾ 60	4.0%
Educational background	1 = Matric or less	18.7%
	2 = Intermediate	32.0%
	3 = Graduate	41.3%
	4 = PG or above	8.0%
Bicycling experience (years)	1 = < 5	34.7%
	2 = 5–10	42.0%
	3 = > 10	23.3%
Daily average bicycling distance (km)	1 = < 5	28.7%
	2 = 5–10	55.3%
	3 = 11–20	14.0%
	4 = > 20	2.0%
Bicycle trip purpose	1 = Recreation	27.3%
	2 = Exercise	22.7%
	3 = Personal errands	18.0%
	4 = Education	16.0%
	5 = Commuting to/from work	8.0%
	6 = Visit a friend (or relative)	8.0%
User type	1 = Regular bicycle user	42.6%
	2 = Occasional bicycle user	57.4%

Table 2. Descriptions and scales of qualitative attributes.

Sl. No.	Variable (Notation)	Unit or Scale	Minimum	Maximum	Mean	Standard Deviation
1	Roadway width (*RW*)	m	3	14	7.45	2.95
2	Pavement condition index (*PCI*)	5-Point scale	2.5	4.5	3.81	0.39
3	Peak hour motorized traffic volume (*PHMV*)	PCUs/h	286	4912.6	2085.2	1330.5
4	Peak hour non-motorized traffic volume (*NMV*)	Veh/h	30	1277	210.1	215.6
5	Average traffic speed (*S*)	Km/h	24	50	35.92	6.85
6	Percentage of heavy vehicles (*%HV*)	%	0	6.97	1.56	1.72
7	Vehicular ingress-egress to on-street parking area (*P*)	Veh/h/km	0	6000	745.78	1442.45
8	Interruptions from roadside stoppages of intermittent public transits (*IIPT*)	3-point scale	0	1	0.41	0.40
9	Roadside commercial activities (*CA*)	3-point scale	0	1	0.46	0.44

Note: The 5-point scale varies from 5 (excellent) to 1 (worst) in an ordered manner, and the 3-point scale varies as 1=high, 0.5=medium, 0=minimal.

The correlations among all important variables with the output variable (perceived BCLR) are shown in Table 3. As observed, all independent variables are strongly correlated with perceived BCLRs with higher correlation coefficients. On the other hand, the inter-collinearities among independent variables are not very substantial. Thus, multi-collinearity does not exist among these selected variables and all are contributing in the model building process independently. In the present context, male bicyclists seemed to be more satisfied than females, and elderly bicyclists seemed to be more dissatisfied than the youth. However, these tendencies were not significant at the 99% significance level ($p < 0.01$). Other socio-economic attributes and driver characteristics (e.g., gender, bicycling experience level, household size, income, working class, etc.) were also observed to be dominated by quantitative road attributes listed in Table 2. Therefore, only quantitative attributes were used in the BCLR model development process.

Table 3. Correlations among input and output variables.

	BCLR	*RW*	*PCI*	*PHMV*	*NMV*	*S*	*%HV*	*P*	*IIPT*	*CA*
BCLR	1.000	-	-	-	-	-	-	-	-	-
RW	0.406**	1.000	-	-	-	-	-	-	-	-
PCI	-0.560**	0.212	1.000	-	-	-	-	-	-	-
PHMV	0.495**	0.219	0.175	1.000	-	-	-	-	-	-
NMV	0.486**	0.211	0.206	0.181	1.000	-	-	-	-	-
S	-0.289**	0.221	0.201	0.190	0.206	1.000	-	-	-	-
%HV	0.443**	-0.010	-0.097	0.044	-0.166	-0.199	1.000	-	-	-
P	0.407**	0.199	-0.033	0.193	0.158	0.105	-0.191	1.000	-	-
IIPT	0.424**	0.204	-0.029	0.189	0.159	0.204	-0.178	0.191	1.000	
CA	0.651**	-0.173	-0.203	0.201	-0.141	-0.123	0.047	0.186	0.180	1.000

Note: **Correlation is significant at the 0.01 level (2-tailed).

5.2. BCLR model development, significance tests and cluster analysis

Data of randomly selected 45 segments (75% of total) were utilized for the development of BCLR model and the remaining data were used for the model validation. A step-wise multivariable regression analysis was carried out by considering all primary variables (listed in Table 2) as independent variables. The averages of all perceived BCLRs (i.e., overall perceived BCLRs) obtained for individual segments were used as the dependent variable. Variables in their original forms, or different combined and/or transformed forms were gone through numerous trials to find out the best configuration in the regression model. Table 4 shows the coefficients of modelled parameters and their statistics as obtained in the most efficient model.

Table 4. Model coefficients and statistics.

Model Term	Coefficient	Standard Error	*t*-statistic	Significance (*p*-value)
Constant	2.412	0.642	3.758	0.001
ln(*PHMV*/*RW*)	0.502	0.080	6.237	< 0.001
NMV/100	0.162	0.034	4.811	< 0.001
PCI	-0.664	0.143	-4.653	< 0.001
S×(1+*%HV*)	0.003	0.001	3.663	0.001
(1+*IIPT*)×(*P*/100)	0.006	0.002	3.260	0.002
CA	0.425	0.133	3.207	0.003

As observed in the above table, all coefficients are associated with negligible standard errors, and their *t*-statistics are highly significant at the 95% confidence level ($p < 0.05$). Thus, it is well convinced that the model estimations are stable and unbiased. Incorporating these estimated coefficients, the BCLR model proposed for street segments of mid-sized cities under heterogeneous traffic conditions could be expressed as follows:

$$\begin{aligned} \text{BCLR} = {} & 2.412 + 0.502 \times \ln\left(PHMV/RW\right) + 0.162 \times \left(NMV/100\right) - 0.664 \times PCI \\ & + 0.003 \times S \times \left(1 + \%HV\right) + 0.006 \times \left(1 + IIPT\right) \times \left(P/100\right) + 0.425 \times CA \end{aligned} \tag{5}$$

The BCLR model has reported a high coefficient of determination (R^2 = 87) between predicted and overall perceived BCLRs during its training stage. Its performance during the validation stage (with data of 15 segments) is shown in Fig. 2. As observed, the coefficient of determination between predicted and overall perceived BCLRs is as high as 0.85. Further, the trend-line has made an angle of 44.4^0 with the horizontal which is very much close 45^0. Thus, the model has well satisfied the required criteria of model validation.

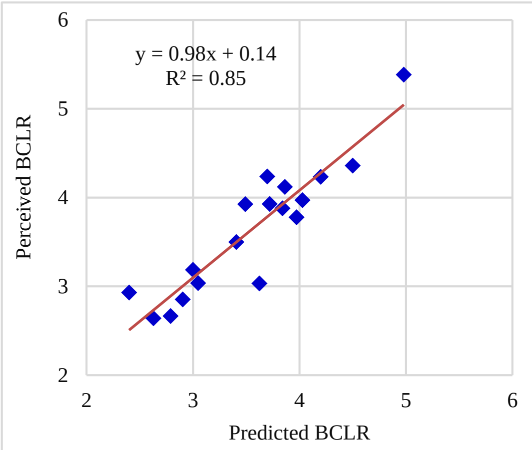

Fig. 2. Plot of average perceived and model predicted BCLRs.

Table 5 summarizes the prediction performance of BCLR model while assessed through the application of R^2 and several other statistical parameters. These statistical parameters include *E*, AAE, RMSE, MAE, mean (μ) and standard deviation (σ) of $BCLR_{PD}/BCLR_{PC}$ values, 50% and 90% cumulative probabilities (P_{50} and P_{90} respectively) of $BCLR_{PD}/BCLR_{PC}$ values, and the overfitting ratio (OR). As observed, BCLR model has a very good prediction performance in both training and validation stages. It has produced reasonably higher values of R^2 and *E* in both the stages. All error measuring parameters also have got very smaller numerical values. Moreover, the μ-values (1.01, 1.03) of $BCLR_{PD}/BCLR_{PC}$ are observably very much close to ‘one’ and σ-values (0.06, 0.08) are very much close to ‘zero’. Thus, the model has very good prediction performance in terms of all these statistical parameters. Values of P_{50} and P_{90} parameters reported in this table are estimated from the cumulative probability plot of $BCLR_{PD}/BCLR_{PC}$ values shown in Fig. 3. This figure depicts that the developed model has fair prediction performance at any cumulative percentage value. P_{50} (1.01, 1.01) and P_{90} (1.08, 1.15) values conclude that, though the model has marginal over prediction at the 90% cumulative probability in its validation stage, it has an excellent performance at other cumulative percentages. The value of ‘OR’ obtained for the model is 1.42 which is very much close to ‘one’. This concludes that the proposed model has a good generalization ability in the present context.

Table 5. Performance of BCLR model in the present context.

Data Set	R^2	*E*	AAE	MAE	RMSE	μ	σ	P_{50}	P_{90}	OR
Training	0.87	0.87	0.18	0.64	0.21	1.01	0.06	1.01	1.08	1.42
Validation	0.85	0.81	0.21	0.59	0.30	1.03	0.08	1.01	1.15	

The empirical bicycle model reported in this paper predicts the BCLR value for a street segment by using the required field observations. In order to determine the bicycle comfort level (A-F) from this predicted BCLR value, a scale has been defined using the GP clustering technique. Here, the number of comfort classes were chosen as six in order to remain consistent with several other relevant studies including HCM (2010). Fig. 4 shows the ranges of BCLRs in each comfort class as obtained through the GP cluster analysis. This figure depicts that, a BCLR value of below 1.75 corresponds to the ‘excellent’ comfort level ‘A’. On the other end of this scale, a BCLR value of above 5.20 corresponds to the ‘worst’ comfort level ‘F’. The BCLRs ranges of other intermediate comfort levels are as shown in the figure.

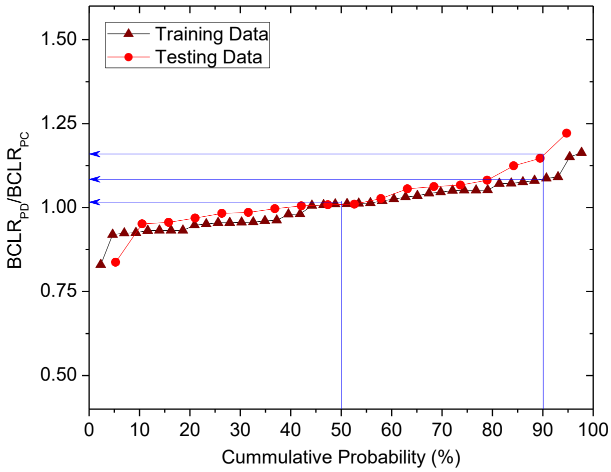

Fig. 3. Cumulative probability plot of BCLRPD/BCLRPC values.

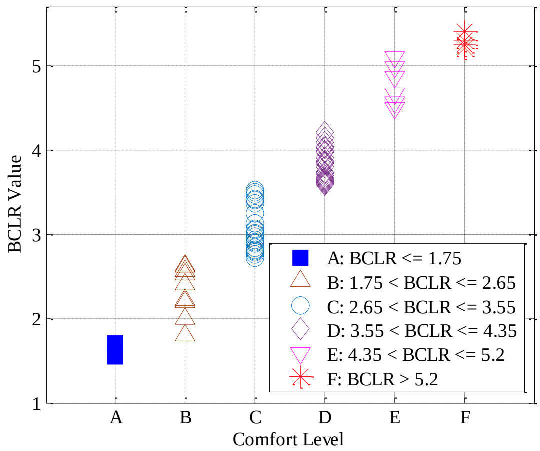

Fig. 4. BCLR ranges of comfort levels (A-F) defined using GP clustering.

Field applications of the GP scale has shown that, none of the studied segments is offering excellent levels of comfort at its present scenario. From the remaining around 3.3%, 46.7%, 44.0%, 5.0% and 1% of the total studied segments are offering bicycle comfort levels of B, C, D, E and F respectively. This means, around 90.7% of all segments are offering average or below average levels of comfort, (C and D). Thus, the service levels of Indian roadways should be largely enhanced for the betterment of bicyclists. The crucial findings of this study will largely help the transportation planners and engineers in this regard. In the present study, the inputs of BCLR model are collected form a wide range (excellent-worst) of road conditions. These variations as shown in Table 2 expectedly well explain the variabilities and complexities in road conditions of mid-sized cities. Thus, the developed model possibly has got tremendous potential for its reliable applications in mid-sized cities of India and other developing countries around the world. However, in case of metropolitan cities and similar other locations where the field observations are outside the ranges of variables reported in this paper, the study effort can easily be duplicated to meet the varied requirements.

5.3. Sensitivity analysis

A sensitivity analysis was carried out to distinguish the relative importance of independent variables accommodated in the BCLR model. Percentage contribution of each input in the prediction of model output was calculated by using Eqs. (6a-6b) (Gandomi et al., 2013) and results are summarized in Table 6.

$$N_i = f_{\max}(x_i) - f_{\min}(x_i) \tag{6a}$$

$$S_i = N_i \Big/ \sum_{i=1}^{n} N_i \times 100 \tag{6b}$$

Where, N_i is the difference between the maximum and minimum value of the predicted outcome over the i^{th} input variable expressed as $f_{max}(x_i)$ and $f_{min}(x_i)$ respectively, which were calculated by putting the maximum and/or minimum (as applicable) values of i^{th} input variable and mean values of remaining inputs in the BCLR model; S_i is the sensitivity in percentage; n is the number of input variables.

Table 6. Relative importance of model inputs.

Variable	*RW*	*PCI*	*CA*	*SPD*	*PHMV*	*%HV*	*NMV*	*IIPT*	*P*
S_i (%)	11.49	11.91	6.32	3.96	20.21	11.11	17.37	2.38	9.24
Rank	4	3	7	8	1	5	2	9	6

In the above table, a higher S_i value designates higher order of importance of an input variable. Among all inputs, *PHMV* parameter has shown the highest contribution of 20.21% in the prediction of BCLRs and acquired the rank of 'one'. This observation concludes that, the motorized traffic extremely hampers bicyclists' perceived sense of comfort under heterogeneous traffic flow environment. Thus, the minimization of bicycle-vehicle interaction is primarily expected in order to enhance the perceived comfort levels of bicyclists. This could be achieved through the provision of separate bicycle lane, wide outer lane or paved shoulder facilities. *NMV* parameter has observably acquired the second rank in the present context. Thus, the effective management of non-motorized traffic is also well desired to enhance the bicyclists' perceived comfort levels. The ranks of remaining variables show that, *PCI* and *RW* are the third and fourth most important variables in the present context. Thus, these parameters should also be largely prioritized in order to enhance the quality of bicycling under heterogeneous traffic flow conditions. *%HV*, *P*, *CA*, *SPD*, and *IIPT* parameters also significantly influences bicyclists' perceived comfort levels in a decreasing order of importance. Thus, these parameters also require attentions for the better operation of bicycle mode on urban road segments.

5.4. Comparison of BCLR model with existing models

Existing bicycle models are developed for homogeneous traffic conditions and are not well transferable to heterogeneous traffic flow environment. These existing models have considered either 'per lane traffic volume' or 'outside lane volume' as one major traffic flow parameter. However, roadways in developing countries carry several categories of vehicles which share the entire roadway width without any segregations or lane disciplines. Moreover, several roadways, particularly in mid-sized cities, are unstripped (no lane marking). Hence, instead of 'per lane traffic volume' parameter, 'traffic volume per unit width of roadway' (*PHMV*/*RW*) has been considered in the BCLR model development process. Moreover, motorized and non-motorized vehicular volumes are considered separately as these parameters have different effect on bicyclists' perceived comfort level. Roadside stoppages of intermittent public transits (authorized/unauthorized) cause considerable discomforts to bicyclists in developing countries. Hence, a new parameter *IIPT* has also been introduced in the BCLR model developmental process. These are some of the key factors which differentiate the BCLR model from existing bicycle models.

6. Conclusions

Several crucial conclusions are documented from this research which is primarily focused on the modelling of bicycle comfort levels offered by road segments in mid-sized cities operating under heterogeneous traffic flow conditions. The data analysis has shown that; nine quantitative variables are primarily influencing the perceived comfort levels of on-street bicyclists in the present context. These parameters include roadway width, pavement condition index, peak hour motorized and non-motorized traffic volumes, average traffic speed, percentage of heavy vehicles, on-street parking activities, interruptions from roadside stoppages of intermittent public transits, and roadside commercial activities. Socio-demographic and other characteristics of the bicyclists (e.g., gender, age, bicycling experience level, household size, income, working class, etc.) were observed to be less important in the mid-sized cities. A step-wise regression-based Bicycle Comfort Level Rating (BCLR) model is developed by considering the nine primary variables as independent variables and overall perceived comfort scores as the dependent variable. The resulting model is highly reliable and has reported high coefficient of determination (R^2) values of 0.87 and 0.85 in its training and validation stages respectively. Several other statistical parameters including Nash–Sutcliffe model efficiency coefficient (E) and error measuring parameters are also applied to assess the prediction performance of this model. These extensive investigations have justified the well reliability of the model in the present context.

A sensitivity analysis carried out on the modelled variables has reported that the peak hour motorized and non-motorized vehicular volume, pavement condition index and roadway width are by far the most important variables in the present context. These variables are contributing by 20.21%, 17.37%, 11.91% and 11.49% respectively in the prediction of BCLRs of road segments. Thus, it is arguably to say that, the minimization of bicycle-vehicle interactions and effective management of bicycle traffic through the provision of separate bicycle lanes (or wide outer lane) is the most important criteria to enhance bicyclists' comfort levels. The provision of separate bicycle lanes also provides a dedicated riding space to the bicycle users. Subsequently, the well maintenance of pavement surfaces and widening of roadways are two other key factors to enhance bicyclist's perceived comfort levels. Few other important observations noted in this regard are: minimization of on-street parking activities, imposing restrictions on the movement of heavy vehicles and vending activities on the city roads particularly during the peak hours of traffic flow.

The stratification of BCLRs into six comfort levels (A-F) through the application of Genetic Programming (GP) clustering has reported that, around 97% of all studied segments are offering average ('C') or inferior comfort levels to the bicyclists at the present scenario. Under do nothing scenario, performance of existing facilities will degrade further over time period due to several reasons such as gradual increase in traffic volume and all type of activities within the urban areas. Hence the quality of existing road surfaces and road side aesthetics should be timely augmented for the betterment of bicyclists. In this decision support system, transportation planners and engineers in the developing countries can take help of BCLR model and other crucial findings of this study. This BCLR model is highly reliable, statistically calibrated and greatly suitable for its applications in mid-sized cities under heterogeneous traffic conditions. This model will help in long-term transportation planning and development of bicycle friendly road infrastructures. Issues related to roadways having the provision of separate bicycle lane(s) are not investigated in this study as the same facilities were not found in the study areas. However, the same can be investigated in similar other studies. The variations in living standards and trip purposes of users in high income and metropolitan cities are noticeably higher than the same in mid-sized cities. Thus, the role of qualitative variables needs further investigation in these big cities. However, the BCLR model has been trained and validated with input attributes collected form a widely diversified (excellent to worst quality) bicycling environments persisting in mid-sized cities. Thus, the model has got tremendous potential for its reliable applications developing countries around the world.

References

Beura, S.K., Chellapilla, H., Jena, S., Bhuyan, P.K., 2017. Service Quality Assessment of Shared Use Road Segments: A Pedestrian Perspective, in: Deiva Sundari, P., Dash, S., Das, S., Panigrahi, B. (Eds.), Proceedings of 2nd International Conference on Intelligent Computing and Applications, Advances in Intelligent Systems and Computing 467, Springer, Singapore.

Beura, S.K., Kumar, N.K., Bhuyan, P.K., 2016. Level of Service for bicycle through movement at signalized intersections under heterogeneous traffic flow conditions, Proceedings of 12th Transportation Planning and Implementation Methodologies for Developing Countries (TPMDC), IIT Bombay, Mumbai, India.

Chellapilla, H., Beura, S.K., Bhuyan, P.K., 2016. Modeling bicycle activity on multi-lane urban road segments in Indian context and prioritizing bicycle lane to enhance the operational efficiency, Proceedings of 12th Transportation Planning and Implementation Methodologies for Developing Countries (TPMDC), IIT Bombay, Mumbai, India.

Cochran, W.G., 1997. Sampling techniques, third ed. John Wiley & Sons, New York.

Cristobal, R., Espejo, P.G., Amelia, Z., Raul, R.J., 2013. Web Usage Mining for predicting final marks of students that use Moodle courses, Computer Applications in Engineering Education 21.1, 135–146.

Davis J., 1987. Bicycle safety evaluation, Auburn University, City of Chattanooga, and Chattanooga-Hamilton County Regional Planning Commission, Chattanooga, TN.

Davis, J., 1995. Bicycle test route evaluation for urban road conditions. Transportation Congress: Civil Engineers-Key to the World Infrastructure, Transportation Congress, Volumes 1 and 2: Civil Engineers—Key to the World's Infrastructure, American Society of Civil Engineers (ASCE), San Diego, CA, pp. 1063-1076.

Epperson, B., 1994. Evaluating suitability of roadways for bicycle use: Toward a cycling level-of-service standard, Transportation Research Record 1438, 9–16.

FDOT., 2009. Quality/Level of Service Handbook, Florida Department of Transportation (FDOT), Tallahassee, FL.

Gandomi, A.H., Yun, G.J., Alavi, A.H., 2013. An evolutionary approach for modeling of shear strength of RC deep beams, Materials and Structures 46, 2109–2119.

Hallett, I., Luskin, D., Machemehl, R., 2006. Evaluation of on-street bicycle facilities added to existing roadways, Center for Transportation Research, University of Texas at Austin, Austin.

Harkey, D.L., Reinfurt, D.W., Knuiman, M., Stewart, J.R., Sorton, A., 1998. Development of the bicycle compatibility index: a level of service concept, Transportation Research Record 1636, 13-20.

Highway Capacity Manual (HCM)., 2010. Highway Capacity Manual, Transportation Research Board, Washington, D.C., 1650 p.

Holland, J. H., 1975. Adaptation in Natural and Artificial Systems, University of Michigan, 183 p.

Hunter, W.W., Feaganes, J.R., Srinivasan, R., 2005. Wide curb lane conversions: The effect on bicycle and motor vehicle interaction, Transportation Research Record 1939, 37–44.

Indian Road Congress (IRC)., 1992. Guidelines for capacity of urban roads in plain areas, Indian Road Congress, 106, New Delhi.

Jensen, S.U., 2007. Pedestrian and bicycle level of service on roadway segments, Transportation Research Record 2031, 43–51.

Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press.

Landis, B.W., 1994. Bicycle interaction hazard score: A theoretical model. Transportation Research Record 1438, 3–8.

Landis, B.W., Vattikuti, V.R., Brannick, M.T., 1997. Real-Time human perceptions: Toward a bicycle level of service, Transportation Research Record 1578, 119–126.

Mozer, D., 1994. Calculating multi-mode levels-of-service, International Bicycle Fund, Seattle, WA.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I-A discussion of principles, Journal of Hydrology 10.3, 282–290.

Patnaik, A.K., Bhuyan, P.K., 2016. Application of genetic programming clustering in defining LOS criteria of urban street in Indian context, Travel Behaviour and Society 3, 38–50.

Prassas, E.S., Roess, R.P., Mcshane, W.R., 1996. Cluster analysis as tool in traffic engineering, Transportation Research Record 1551, 39–48.

Sorton, A., Walsh, T., 1994. Bicycle stress level as a tool to evaluate urban and suburban bicycle compatibility. Transportation Research Record 1438, 17–24.