ARTICLE IN PRESS

International Journal of Transportation Science and Technology xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

journal homepage: www.elsevier.com/locate/ijtst

Concept of heterogeneity index for urban mixed traffic

Ashutosh Pandey [a], Manjul Sharma [b], Subhadip Biswas [c,*]

[a] Post Graduate Scholar, Department of Civil Engineering, National Institute of Technology, Hamirpur, Hamirpur-177005, Himachal Pradesh, India
[b] Ph.D. Scholar, Department of Civil Engineering, Motilal Nehru National Institute of Technology, Allahabad, Prayagraj-211004, Uttar Pradesh, India
[c] Assistant Professor, Department of Civil Engineering, National Institute of Technology, Hamirpur, Hamirpur-177005, Himachal Pradesh, India

ARTICLE INFO

Article history:
Received 9 November 2021
Received in revised form 4 January 2022
Accepted 19 February 2022
Available online xxxx

Keywords:
Heterogeneous traffic
Urban roads
Gaussian process regression
PCU
Clustering

ABSTRACT

Heterogeneity is one of those characteristics which differentiate traffic conditions of a developing country from other developed nations. The heterogeneity which represents the diversity among vehicle categories is suspected to have adverse influences on lane discipline, congestion potential, and road users' safety. However, the influence of heterogeneity on the above-mentioned parameters has been only measured indirectly by considering traffic composition as an indicator of the prevailed heterogeneity-level. No direct relationship between the heterogeneity and other parameters has been established as there is no methodology available yet for quantifying the heterogeneity of mixed traffic. The present study addresses this problem and conceptualizes the 'Heterogeneity Index' (HI) to quantify the heterogeneity present in a mixed traffic stream. HI is conceived as a measure of the dispersion of Passenger Car Units (PCU) for different vehicle categories from its central value. A higher value of HI signifies more diverse vehicle categories present in the traffic stream. PCU of a vehicle category was estimated using the speed-based method and the individual speeds were predicted based on classified volumes using the Gaussian Process Regression model developed in this study. This paper also recommends several categorical levels for easy perception about the intensity of heterogeneity. Further, the sensitivity analysis explored the dynamic aspects of HI. Results showcased how the HI of a traffic stream may vary subject to the combined or the individual change in traffic volume, traffic composition and classified speeds. The outcomes of the study will be useful to estimate the intensity of heterogeneity that prevailed within a mixed traffic stream with varying traffic conditions.

1. Introduction

The exponential growth in urban population is resulting in a massive surge in vehicles operating on urban roadway facilities. By 2030, the world's total vehicle growth will exceed 2 billion in non-OECD (Organization for Economic Cooperation and Development) countries accounting for more than half of the global vehicles (Dargay et al., 2007). While the picture of automobile growth is similar in both developed and developing countries but other traffic characteristics are significantly unalike. While the traffic stream of developed countries consists of majorly cars and heavy vehicles, a variety of vehicles is

Peer review under responsibility of Tongji University and Tongji University Press.
* Corresponding author.
E-mail address: sbiswas@nith.ac.in (S. Biswas).

https://doi.org/10.1016/j.ijtst.2022.02.008
2046-0430/

Please cite this article as: A. Pandey, M. Sharma and S. Biswas, Concept of heterogeneity index for urban mixed traffic, International Journal of Transportation Science and Technology, https://doi.org/10.1016/j.ijtst.2022.02.008

generally observed on urban roads of developing countries (Dhamaniya and Chandra, 2013; Pandey and Biswas, 2021). Indian Highway Capacity Manual (CSIR- Central Road Research Institute, 2017) identified nine categories of motorized vehicles and two non-motorized categories while analyzing the capacity of urban roads in India. Not only are the physical dimensions of these vehicles diverse, but a wide diversity also exists in their speed profiles and maneuverability (Manghat et al., 2017). This heterogeneity of traffic creates additional challenges in analyzing the aggregated traffic movements and eventually designing the urban roadway facilities in developing countries. In this regard, Passenger Car Unit (PCU) or Passenger Car Equivalent (PCE) is an established concept that is generally adopted while converting the heterogeneous traffic volume into its homogeneous equivalent. HCM 2000 (Transportation Research Board, 2000) defined PCE as "*the number of passenger cars displaced by a single heavy vehicle of a particular type under specified roadway, traffic and control conditions*". Researchers (Arasan and Krishnamurthy, 2008; Biswas et al., 2020; Chandra and Kumar, 2003; Seguin et al., 1982; Transport and Road Research Laboratory, 1965) adopted this concept and extended the analysis towards finding PCU values of other vehicle categories along with heavy vehicles specifically in a heterogeneous traffic situation. Six popular PCU estimation methods namely, headway method (Craus et al., 1980; Krammes and Crowley, 1986; Rongviriyapanich and Suppattrakul, 2005; Subotić et al., 2016), homogeneous coefficient method, Walker's method (Werner and Morrall, 1976), multiple linear regression method (Aerde and Yagar, 1984), Huber's method (Huber, 1982), and speed-based method (Basu et al., 2006; Biswas et al., 2020, 2017a, 2017b; Chandra and Kumar, 2003; Rahman and Nakamura, 2005) exist in literature. Among these, the speed-based method (Chandra and Kumar, 2003) has great compatibility with heterogeneous traffic conditions and has become the most popular method in the last two decades for estimating PCU factors of the heterogeneous traffic stream. Characteristics of PCU factors have also evolved with time. While different studies and capacity manuals (Adnan, 2014; Ben-Edigbe and Ferguson, 2005; Fan, 1990; Yeung et al., 2015) recommended the static values of PCU, many studies (Arkatkar and Arasan 2010; Biswas et al., 2017c; Dhamaniya and Chandra 2016; Mishra et al. 2017) in recent times showed its dynamic nature. It was observed that the PCU of a vehicle changes with the change in surrounding traffic and other conditions (Sharma and Biswas, 2020).

Urban traffic streams are generally heterogeneous everywhere but the 'degree of heterogeneity' may not be the same as it varies largely with time and location (Suvin et al., 2020; Yang et al., 2017). It is quite conspicuous from the above discussion that there are many studies reported in the literature that recommended PCU factors to covert a heterogeneous traffic volume into an equivalent homogeneous. In that way, PCU is only useful to deal with the heterogeneous traffic volume but unable to quantify the degree of heterogeneity that prevailed within the traffic stream. For example, two traffic streams having the same set of PCU values do not guarantee the same degree of heterogeneity present in those traffic streams. Hence, in this regard, no study has been conducted yet to quantify the 'degree of heterogeneity' of a mixed traffic stream. A summary regarding the research gap and the necessity of proposing a new measure to quantify the heterogeneity is given below.

- The traffic streams of the majority of developing countries are heterogeneous unlike the homogeneous traffic of developed nations. Due to this and a few other dissimilarities, the norms given in US-HCM and other guidelines cannot be implemented directly in a developing country's context that has been thoroughly reported in the literature (Luca and Acqua, 2014; Raj et al., 2019). However, the consideration of this heterogeneity has always been made at a categorical level; either 'heterogeneous' or 'homogeneous'. No previous attempt has been made to know 'if it is heterogeneous, then how much heterogeneous?'. The heterogeneity level is not the same for all roads rather it is anticipated to have significant spatial and temporal variations. In this context, an effective measurer is required that can quantify the degree of heterogeneity for varying mixed traffic conditions.
- Today, the degree of heterogeneity of a mixed traffic stream is perceived based on the observed traffic composition. But, it can only give a vague idea about the heterogeneity and cannot measure it quantitatively. For example, two mixed traffic streams feature the following traffic compositions; a) cars: two-wheelers: heavy vehicles → 50:30:20, b) cars: two-wheelers: heavy vehicles → 40:20:40. Based on the traffic composition, it can only be said that both streams are heterogeneous but it is not possible to identify which is more heterogeneous and to what extent.
- It is anticipated that the increase in heterogeneity of a traffic stream degrades the lane discipline and the safety of road users and increases the chances of congestion in a typical urban road scenario of a developing country (Matcha et al., 2020; Saini and Biswas, 2021; Shen and Yang, 2020; Wen et al., 2018). However, these influences of heterogeneity have not been evaluated yet due to the complete absence of a method in literature to quantify the heterogeneity itself.
- Treatments like 'provisions of dedicated lanes for a particular vehicle class' can be made based on a heterogeneity-based criterion. More specifically, when the heterogeneity of a mixed traffic stream exceeds a predefined critical value, a dedicated lane would be provided for the vehicle class which has the maximum impact on the heterogeneity. After segregating the vehicle class from the mainstream, the improvement made in the heterogeneity-level can also be assessed. All of these aforementioned works can be implemented in the future if a methodology for quantifying the heterogeneity-level is developed.

On this background, the present study is conceived to quantify the degree of heterogeneity of a mixed traffic stream and to examine its variation with the change in traffic conditions. This quantification of heterogeneity will guide the transportation planners to assess the amount of heterogeneity present on a road within a given period of time and also act as a surrogate measure to predict the chances of crashing, congestion, and overall lane- indiscipline.

2. Study design

The present study conceptualizes the 'Heterogeneity Index' to quantify the heterogeneity present in a mixed traffic stream. The heterogeneity of a mixed traffic stream can be quantified based on a parameter that can measure the dispersion of different PCU values from their overall centroid. In this attempt, the study thoroughly reviewed several dispersion parameters and concluded that either standard deviation/variance or coefficient of variation (COV) would be a suitable dispersion parameter for this purpose. Although the selection of dispersion parameter depends upon the dataset and its scale, in general, former studies (Weber et al., 2004) have opted for COV due to its more interpretability as compared to the standard deviation. Hence, the present study adopted a proportion-based variant of COV for effective quantification of the Heterogeneity Index. Further, for estimation of PCU values, a speed-based method proposed by Chandra and Kumar (2003) was adopted due to its dynamic nature and simplicity. Classified speeds used in the PCU formula are governed by several influencing factors. Hence, there is a need to develop a classified speed prediction model before quantifying the Heterogeneity Index. In this regard, this study employed the Gaussian Process Regression (GPR) technique for predicting the classified speed of an urban mixed traffic stream based on a given circumstance. Therefore, the methodology of quantifying heterogeneity is divided into two parts; i) development of classified speed model using Gaussian Process Regression-based approach, ii) estimation of PCU and Heterogeneity Index.

2.1. Methodology for development of classified speed model using Gaussian Process regression

Classified speeds which are taken as inputs in this PCU estimation, may vary with the change in traffic conditions. Therefore, the development of a classified speed model is a prerequisite before arriving at the estimation of PCU or Heterogeneity Index. Over the last century, the development of the volume-based traffic speed model had received perpetual emphasis as the in-field collection of speed data turns out to be infeasible more often than not. In the absence of considerable side frictional elements (curb parking, bus-stops, pedestrian movements, traffic movements through access points, etc.), the average speed of a vehicle category is completely governed by the overall traffic volume and its compositional characteristics in the context of a mid-block heterogeneous traffic stream. Therefore, the classified traffic volumes which represent both traffic volume and its composition can be considered as the design variables to predict the classified speeds. In recent times, the usefulness of the volume-based classified speed model in the context of heterogeneous traffic has been realized and few studies have come up with different speed models. Dhamaniya and Chandra (2016) modified Greenshields (Greenshields et al., 1935) model to take it towards the individual vehicle category level. Eq. (1) suggested by Dhamaniya and Chandra (2016) is useful to predict the average speed of individual vehicles for a given set of classified traffic volumes.

$$V_i = a_0 - \sum_{i=1}^{n} a_i \left(\frac{Q_i}{V_i}\right) \tag{1}$$

V_i = average speed of a vehicle category 'i'.
Q_i = volume of a vehicle category 'i'.
a_0 = intercept.
a_i = regression coefficients.
n = number of vehicle categories present in a traffic stream.

It is to be mentioned here that an iterative approach is essentially required to use this model (Eq. (1)) as the speed (V_i) appears at both sides of the equation. Therefore, it eventually loses the suitability of the model in-field application. Moreover, the model is developed based on the assumption that the overall speed-density relationship is following a particular trend (linear). This can also be considered as a drawback of this model. In this regard, the present study adopted Gaussian Process Regression-based approach to develop a classified speed model which also overcomes the shortcomings associated with the existing model.

Gaussian Process Regression (GPR) is a non-parametric approach (Schulz et al., 2018; Wan and Sapsis, 2017). Instead of assuming a particular trend, GPR considers all admissible functions that match the data set. The basic model of GPR can be divided into two parts; a) regression part and b) stochastic part. The regression part performs approximation on the global scale while the stochastic part deals with the approximation on the local scale. Suppose, a training dataset S is considered in the following manner; $S = \{(x_i, y_i); i = 1, 2, \cdots n\}$. y_i is a dependent or response variable $\{y_i \in R\}$ and $\times$ is a multi-dimensional vector containing 'd' number of independent variables $\left\{x_i \in R^d\right\}$. The regression model with Gaussian noise can be described using Eq. (2).

$$y = x^T \beta + \varepsilon \tag{2}$$

Here, ε is the Gaussian Noise which is added to the model assuming that the actual observed values of response variable y will differ from function values $x^T\beta$. ε follows a Gaussian distribution with a mean 0 and an error variance of σ^2 as given in Eq. (3). For modeling purposes, the values of β (coefficient) and σ^2 (error variance) can be determined based on the dataset kept for the training purpose.

$$\varepsilon\, N(0, \sigma^2) \tag{3}$$

Further, a GPR technique proposes latent variables $f(x_i)$, i = 1, 2, ..., n, (from a Gaussian Process) and explicit basis functions h(x) to model the response variable. A Gaussian process can be defined as a set of random variables which follow a joint (multivariate) Gaussian distribution for any number of random variables in the dataset. If $f(x), \{x \epsilon R^d\}$ is a Gaussian process, then a joint multi-Gaussian distribution of random variables with n observations $\{f(x_1), f(x_2), \cdots f(x_n)\}$ will also be a Gaussian Process with the mean 0 and the covariance $k(x_i, x_j)$ as given in Eq. (4). $k(x_i, x_j)$ which is known as kernel function, estimates the covariance between two observation points. The covariance is directly governed by the hyper-parameters or kernel parameters (θ) (Wan and Sapsis, 2017).

$$f(x)\ GP(0, k(x_i, x_j)) \tag{4}$$

Hence, Eq. (2) is transformed as.

$$y = \underbrace{\overbrace{h(x)^T \beta}^{Regression\ \ Part}}_{Approximation\ \ on\ \ global\ \ scale} + \underbrace{\overbrace{f(x)}^{Stochastic\ \ Part}}_{Approximation\ \ on\ \ local\ \ scale} \tag{5}$$

Here, h(x) is a set of explicit basis functions which set the primary trend of the model to be constructed.

The present study primarily considered six popular kernel functions i.e. i) Squared Exponential (SE), ii) Matern 3/2, iii) Matern 5/2, iv) Automatic Relevance Determination (ARD) Squared Exponential (ASE), v) ARD Matern 3/2 (AM32) and vi) ARD Matern 5/2 (AM52) and attempted to find the best suitable function based on error analysis. Also, two hyper-parameters are primarily involved in different kernel functions; i) the characteristic length scale (σ_l) and ii) the signal standard deviation (σ_f). These hyper-parameters can be determined by the process of optimization. The present study adopted Maximum Likelihood Estimation (MLE) technique to maximize the log-marginal likelihood of the training data. The values of hyper-parameters corresponding to the maximum likelihood were taken as the optimal hyper-parameter for each kernel function. The marginal likelihood function can be optimized through any convenient mathematical tool, provided that the kernel function is differentiable with respect to its hyper-parameter. In this study, a gradient-based optimizer was used. Details of error analysis for finding the desirable kernel function are discussed in *Section 4*.

The present study adopts the individual volumes of different vehicle categories as the input variables in the GPR method to predict the classified speeds. For this purpose, the conventional method i.e. multiple linear regression can also be used as an alternative to the GPR-based approach. However, the weakness of multiple linear regression models would lie in assuming the trend of speed-volume relationship before model development (Dhamaniya and Chandra, 2016). On the other hand, GPR based approach is flexible as it can select the most suitable kernel function depending upon the nature of data. In addition, GPR performs approximation at two different levels and is expected to predict the outcome more accurately as compared to the conventional method. Due to these, GPR is more efficient than the conventional regression technique in the case when the input variables are multiple and the size of the dataset is limited.

2.2. Methodology for estimation of PCU and heterogeneity index

Estimation of PCU is a prerequisite before arriving at quantifying the Heterogeneity Index. Although several methods for PCU estimation are available in the literature (Sharma and Biswas, 2020), the speed-based method is the most suitable one for heterogeneous traffic streams with poor lane discipline. This is due to its simplicity and the ability to capture the dynamic nature of PCU (Sharma and Biswas, 2020). Therefore, the present study adopted the speed-based method and estimated PCU of a vehicle category using the formula proposed by Chandra and Kumar (2003) as given in Eq. (6).

$$PCU_{i(5-\text{min})} = \frac{V_c/V_i}{A_c/A_i} \tag{6}$$

$PCU_{i\ (5\text{-min})}$ = PCU of a vehicle category 'i' corresponding to a 5-minutes time interval.
V_c = Average speed of car (km/h).
V_i = Average speed of vehicle type 'i' (km/h).
A_c = Projected rectangular area of car (m^2).
A_i = Projected rectangular area of vehicle type 'i' (m^2).

The classified speeds predicted by the GPR model developed at the previous phase were used as inputs in Eq. (6) to estimate PCU factors of different vehicle categories. However, the projected rectangular areas for different vehicle categories were measured in the field.

The present study conceptualizes the 'Heterogeneity Index' to quantify heterogeneity in a mixed traffic stream. Heterogeneity Index (HI) is defined as the adapted coefficient of variation of PCU factors estimated for different categories of vehicles with respect to its overall centroid value within a given time. The diversity in static and dynamic characteristics among different vehicles is reflected by the variation in PCU factors and HI represents this variation. This concept is further demonstrated using Fig. 1 which exhibits the ranges of PCU factors estimated for different vehicle categories viz. motorized two-wheeler (2WH), three-wheeler (3WH), car, heavy vehicle (HV), and bicycle. Here, it is needless to mention that the PCU of a car is 1 (constant) and hence, its variation cannot be seen. However, PCU factors of other vehicle categories can be observed as diverse as their ranges are not even overlapping with each other. For a particular time interval, the standard deviation of

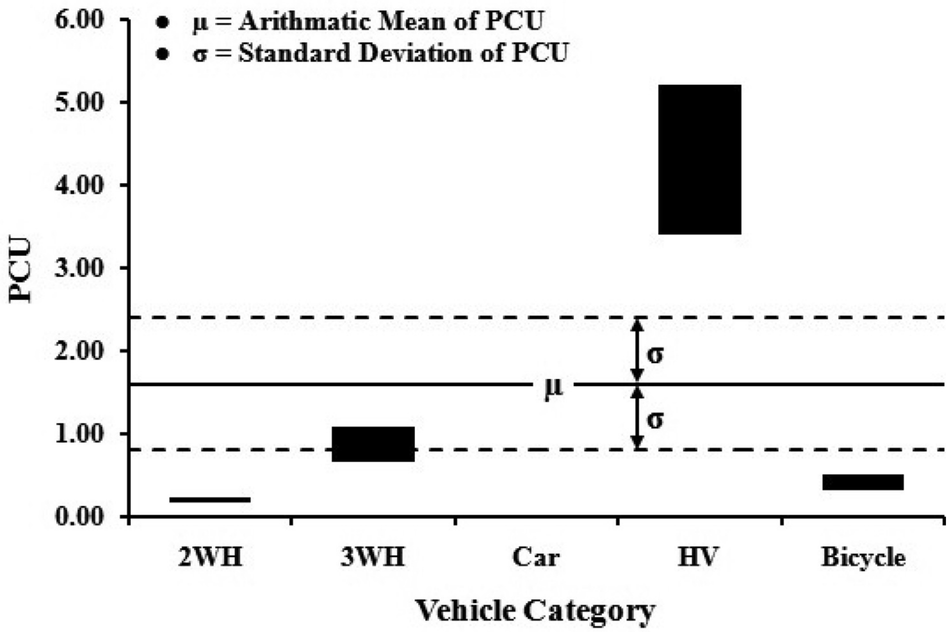


Fig. 1. Concept of Heterogeneity Index.

different PCU factors from its arithmetic mean replicates the diversity of vehicle categories present in the traffic stream. However, this study adopted a proportion-based variant of COV to include the influence of individual proportions of different vehicle categories in determining HI. Hence, HI in this study is proposed as the ratio of the proportion-based standard deviation (σ) and the arithmetic mean (μ) of PCU factors for a particular time interval and can be determined using Eq. (7).

$$HeterogeneityIndex(\%)_{(5-min)} = \frac{\sqrt{\sum_i \left(P_i \times PCU_i^2\right) - \left(\sum_i P_i \times PCU_i\right)^2}}{\sum_i (P_i \times PCU_i)} \times 100 \quad (7)$$

HI $(\%)_{(5\text{-min})}$ = Heterogeneity Index of a traffic stream corresponding to 5-minutes time interval.

P_i = Proportion of a vehicle category 'i' within a traffic stream.

In Fig. 1, as the diversity among PCU values increases, the standard deviation (the numerator of Eq. (7)) will increase. It further leads to an increase in HI as per Eq. (7). Thus, the proposed HI model can capture the diversity among PCU values and estimate the level of heterogeneity present within a traffic stream. The minimum possible value of HI is 0 which would indicate a homogeneous traffic stream composed of a single category of vehicles. However, there is no upper limit of HI. It can be even higher than 125% ($\sigma > 1.25\mu$) for highly heterogeneous traffic. PCU values for different vehicle categories estimated in the previous phase were used in Eq. (7) and accordingly, HI was estimated corresponding to each 5-min time interval.

PCU factors are dynamic in nature. Therefore, HI which captures the diversity among PCU factors, is also expected to vary with the change in prevailing traffic conditions. Hence, there is a need to classify the range of HI into several categorical levels to provide an easy perception of the intensity of heterogeneity present in a traffic stream. In this regard, the K-medoid clustering technique was adopted to classify the observed range of HI into several thresholds. Later, the thresholds obtained were verified using two validation measures; Davies-Bouldin Index and Silhouette Index. Further, to examine the influences of traffic volume, traffic composition, and classified speeds on HI, a sensitivity analysis was carried out. In the first phase, traffic composition was kept constant at its overall observed values and traffic volume was varied gradually within its observed range. Conversely in the second phase, proportions of different vehicles were simultaneously varied keeping the traffic volume constant at a predefined value. In the third phase, the speed of a particular vehicle category was varied keeping other speeds constant at the same predefined values. On all occasions, the change in HI was monitored.

A flow chart showing different phases of the research methodology adopted in this study is exhibited in Fig. 2.

3. Collection of traffic data

Traffic data were collected at a road section on Palam Road in New Delhi, the capital and a major metropolitan of India. The section has the following characteristics.

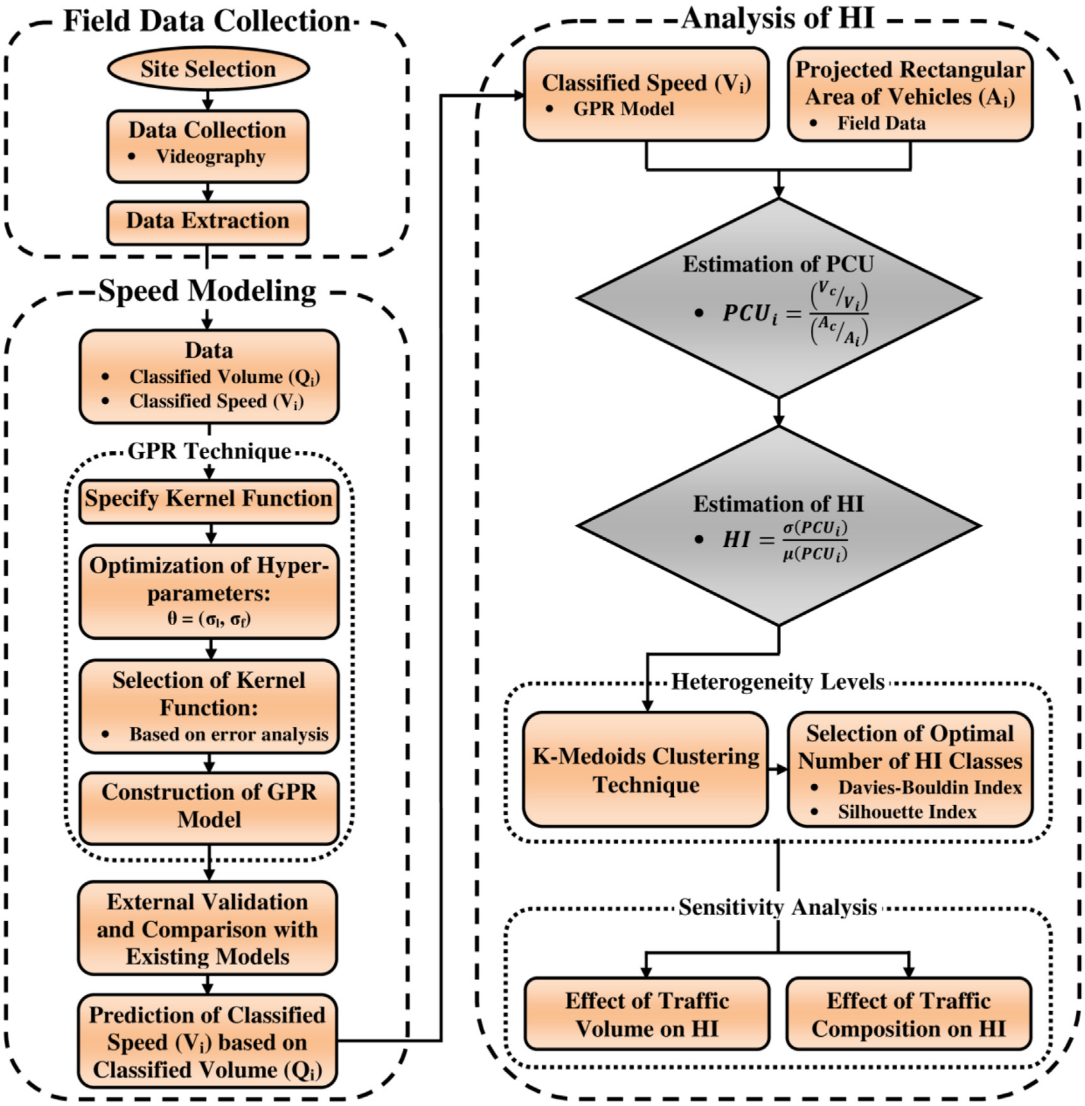


Fig. 2. Methodology followed in the research.

i. The section is straight and located in plain and leveled terrain.
ii. The section is located on a two-lane undivided urban road.
iii. The section is free from any side friction (curb parking, bus-stop, pedestrian movement, traffic movement through access points, etc.).
iv. The section is located at a distance of more than 500 m on either side from the nearby intersection.
v. Surface condition of the road is good without any pavement distress.

The entire traffic data collection was based on the videography survey followed by manual extraction of the data from the video files. A 30 m longitudinal stretch with a carriageway width of 9 m was chosen on this section for speed data collection as shown in Fig. 3a. Traffic data were collected using the videography technique by mounting a camera at the top of a 4.57 m tall camera-stand. The set-up was placed roadside in such a way that the traffic movements on the selected stretch could be clearly captured as shown in Fig. 3b. Videography was carried out at normal weather conditions uninterruptedly for 7 hours which include peak as well as off-peak hours. Initially, the 30 m section was marked with on-site temporary lines to collect the video files. After the videography, the video files were taken to the laboratory and the data required were extracted manually. For this, the video files were edited by adding virtual lines which define the boundaries of the stretch as shown in Fig. 3.

Based on manual observation while playing the video files, vehicles were classified into five different categories viz. cars, two-wheelers (2WH), heavy vehicles (HV), three-wheelers (3WH), and bicycles. Classified traffic volumes i.e. the volumes

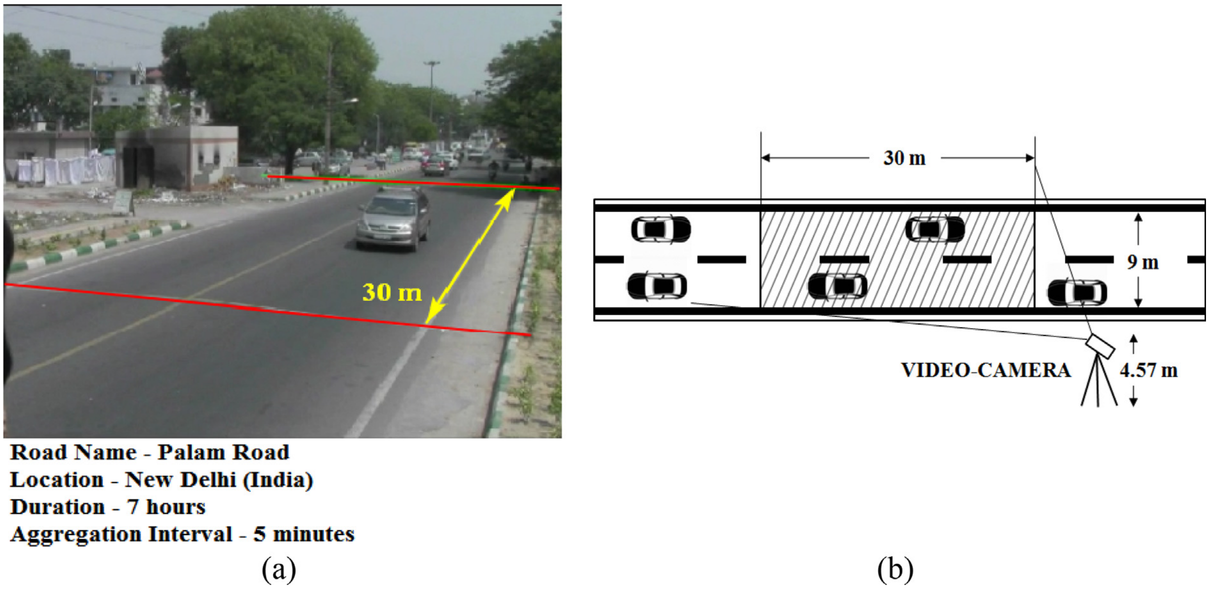


Fig. 3. a) Selected road section at Palam Road (New Delhi) and b) Videography setup.

under individual vehicle category were measured in both directions by counting the number of vehicles crossing the road section within each 5-minute time interval. Sample size under each vehicle category extracted from the video is given in Table 1. On the other hand, the speed of a vehicle was measured based on the time taken by the vehicle to cross the 30 m long stretch. As the videos were recorded at 25 frames per second, each frame accounted for 0.04 seconds and thus the entry and exit time of a vehicle on the stretch was measured with an accuracy of 0.04 seconds. Therefore, the classified speeds i.e. space mean speed under each vehicle category was estimated for every 5-minute time interval. The range of average speed data for each vehicle category is given in Table 1. In this way, 7 hours of traffic data yielded 84 data-points of classified traffic volumes and speeds. A good variation (1512 veh/h to 3636 veh/h) in hourly traffic volume was observed in the extracted traffic data. However, the directional split (the ratio of one-directional traffic volume to the total traffic volume) was found to vary within a narrow range of 40:60 to 50:50. The proportion of different vehicle categories in the traffic stream also varied with time as given in Table 1.

4. Analysis and results

Traffic data collected from the field were analyzed to estimate HI adopting the methodology discussed earlier. The entire analysis part is divided into five phases; i) Development of classified speed model using GPR-based method, ii) External validation and comparative performance assessment of the proposed model and the existing model, iii) Estimation of HI, iv) Development of heterogeneity levels using clustering technique and v) Sensitivity analysis to examine the effects of traffic

Table 1
Proportions and average projected rectangular areas of different vehicle categories measured in the field.

		Vehicle Category				
		Car	**Two-Wheelers**	**Heavy Vehicles**	**Three-Wheelers**	**Bicycles**
Vehicles included		All passenger cars	Motorized two-wheelers	Buses, LCVs* and trucks	Auto-rickshaws	Non-motorized two-wheelers
Sample Size		7862	6347	495	748	1378
Traffic Composition (%)	**Minimum**	27	23	0	1	3
	Mean	47	38	3	5	9
	Maximum	64	48	20	10	15
Average Projected Rectangular Area (m^2)		6.73	1.2	24.54	4.48	0.86
Average Speed (km/h)	**Minimum**	37.39	36.84	29.02	29.07	13.50
	Mean	47.64	42.91	37.43	37.26	15.59
	Maximum	56.71	53.85	51.66	46.70	17.38

* LCVs = Light Commercial Vehicles.

volume, traffic composition and classified speed on HI of the traffic stream. Results obtained in each of these phases are reported below.

4.1. Development of classified speed model

For training purposes, 85% of these data points (71 data) were selected using simple random sampling where each of 84 data-points had equal probability to be selected. The rest 13 data-points were kept aside for testing the model and selecting the suitable Kernel function. Six popular kernel functions viz. Squared Exponential (SE), Matern 3/2 (M32), Matern 5/2 (M52), ARD Squared Exponential (ASE), ARD Matern 3/2 (AM32), and ARD Matern 5/2 (AM52), were primarily considered. GPR-based speed models were developed separately considering each of the kernel functions and optimizing their hyper-parameters using a gradient-based optimizer (quasi-newton). Formulations of different kernel functions and the optimized values of hyper-parameters are given in Table A.1 & Table A.2 of *Appendix* respectively. Therefore, a total of 30 GPR-based speed models were developed in a combination of five vehicle categories and six kernel functions. To assess the performance of each model, the classified traffic volumes of the testing data were taken as inputs in the developed models and the classified speeds predicted by these models were recorded. Subsequently, these predicted speeds were compared with the classified speeds observed in the testing data. Although a good agreement between these two speeds was observed for the majority of the models, the error analysis was carried out to identify the most suitable kernel function for the present analysis. For this purpose, an error parameter, 'Mean Absolute Percentage Error' (MAPE) was employed to examine the accuracy of each model in predicting the classified speeds. MAPE can be calculated using Eq. (8).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{V^{P}{}_{i} - V^{O}{}_{i}}{V^{O}{}_{i}}\right| \tag{8}$$

V_i^P = Predicted speed for data point 'i'.
V_i^O = Observed speed for data point 'i'.
n = number data points in testing data.

Fig. 4 shows the values of MAPE obtained for different speed models developed using various kernel functions. Each spoke of the radar chart (Fig. 4) shows the obtained MAPE value corresponding to a particular vehicle category. It was observed that ARD Matern 3/2 (AM32) kernel function had the least average MAPE value (2.64%) as compared to other kernel functions. Also by visual inspection on Fig. 4, the red-shaded area for AM32 enclosed by MAPE values looks the smallest among all. This indicates that ARD Matern 3/2 (AM32) is the most suitable kernel function for the present job and the same was adopted for further analysis. Finally, the developed GPR-based speed model could predict the average speed of a vehicle category for a given set of classified traffic volumes. However, before using this model in quantifying heterogeneity, the accuracy of the model should be checked based on a new set of data.

4.2. External validation and performance assessment of the model

To validate the performance of the model, another study location having similar road geometry was chosen on Vigyan Path, Chandigarh, and traffic data of 1 hour and 30 minutes (i.e. 18 data-points) were collected similarly as discussed before. Observed speeds on Vigyan Path were then compared with the speeds predicted by the GPR-based model based on the same classified volumes observed on the road. Satisfactory neighborliness was observed between the observed and the predicted speeds for each vehicle category as exhibited in Fig. 5. Table 2 shows that the error (MAPE) involved in predicting the average speed of individual categories was consistently found below 7.5%.

To assess the performance of the proposed model in comparison to the existing model, a new set of classified speed models was developed adopting the alternative methodology proposed by Dhamaniya and Chandra (2016). The accuracy of these models was also tested based on the same classified traffic data collected at Vigyan Path. Results of accuracy-check given in Table 2 evidence that the proposed GPR-based model outperformed the other model in predicting the classified speeds when applied upon a new set of data. Therefore, the classified speeds predicted by this model were further used in the estimation of PCU factors for different vehicles and subsequently, the estimation of HI.

4.3. Estimation of heterogeneity index

Heterogeneity Index (HI) is determined based on the variation observed in PCU factors for different vehicle categories and PCU is estimated using the method proposed by Chandra and Kumar (2003). The average speeds in each time interval and the projected rectangular areas (given in Table 1) for different vehicle categories were taken as inputs in Eq. (6) to estimate PCU factors. Owing to the change in average speeds with time, PCU factors were observed varying within the following range as given in Table 3.

Thereafter, HI was determined separately for each time interval using Eq. (7). A wide variation (52.43 to 125.73%) of HI was observed on Palam Road for different time intervals. Further, when the values of HI estimated based on the observed classified speeds were compared with those predicted by the proposed GPR-based framework, a good proximity (MAPE = 2.82%) was observed. Hence, the proposed GPR-based model has the potential to predict the HI of a traffic stream

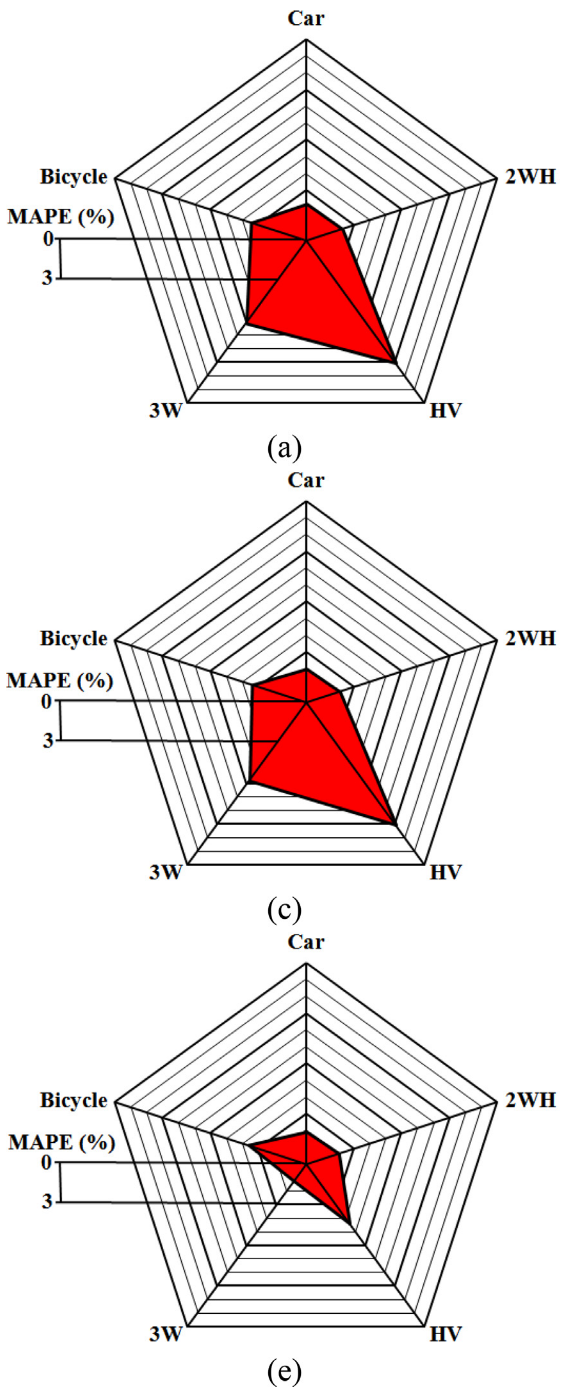


Fig. 4. Mean Absolute Percentage Error (MAPE) obtained in predicting the average speed of individual vehicle category using (a) SE, (b) M32, (c) M52, (d) ASE, (e) AM32, and (f) AM52 as the kernel function.

with reasonable accuracy. However, when the information regarding the classified speeds is available, it is advised to only use the HI formula (Eq. (7)) proposed in this study for quantifying heterogeneity. On the other hand, when the in-field collection of speed data is infeasible, GPR-based speed models coupled with the HI formula can be utilized.

4.4. Heterogeneity levels

The present study further aims to classify the range of HI into a few categorical levels for easy perception about the intensity of heterogeneity. K-medoids clustering technique was adopted to determine the thresholds of different levels of

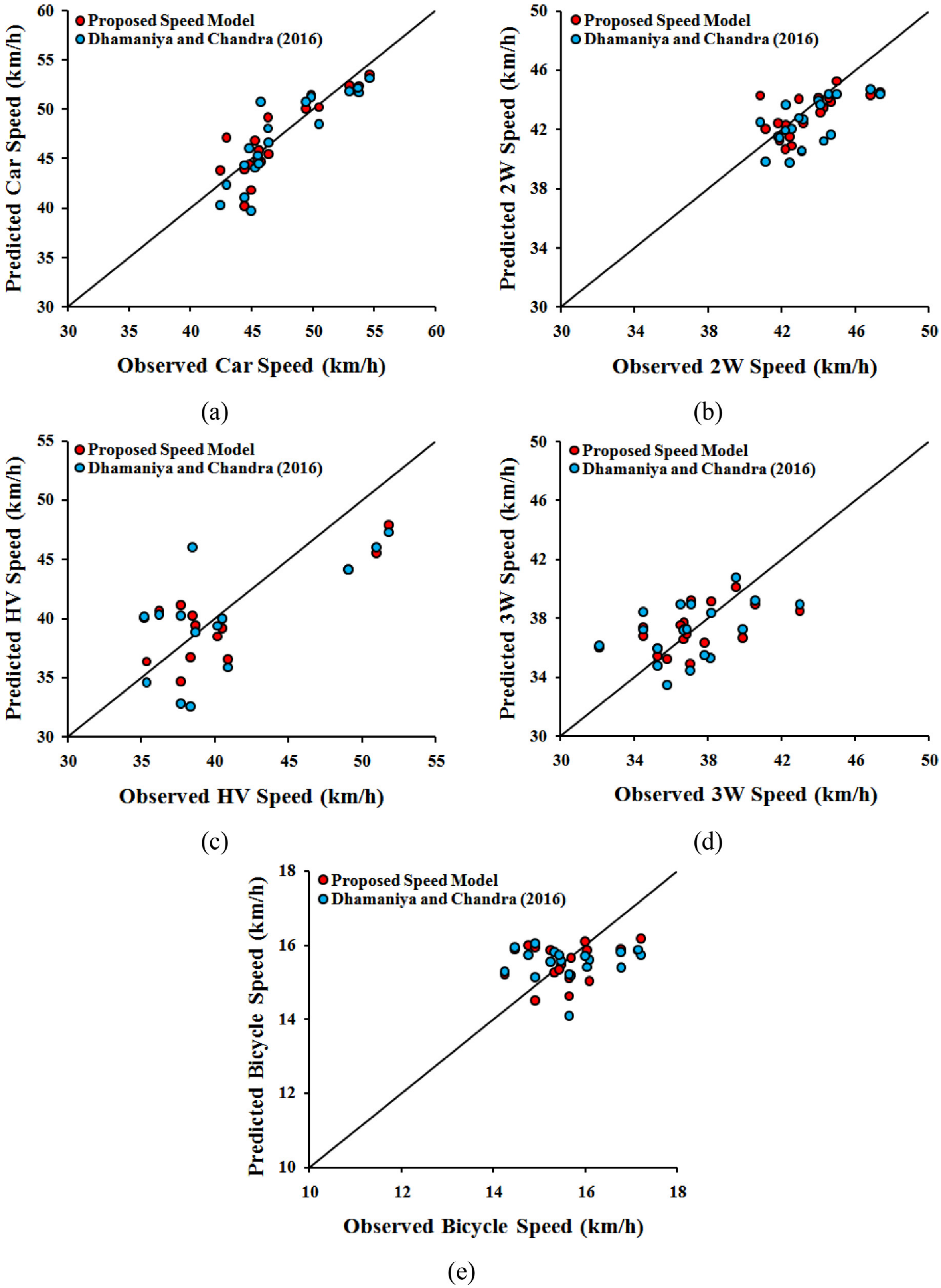


Fig. 5. Validation and comparison of classified speed models for (a) Car, (b) Two-Wheeler, (c) Heavy Vehicle, (d) Three-Wheeler and (e) Bicycle.

Table 2
Comparison of accuracy in predicting classified speeds by the proposed model and the existing model.

Speed model	MAPE (%) involved in predicting classified speeds for different vehicle categories					
	Car	2WH	HV	3WH	Bicycle	Average
Proposed GPR-based model	3.20	2.73	7.34	4.58	4.31	**4.43**
Dhamaniya and Chandra (2016)	3.65	2.75	8.98	5.35	5.03	**5.15**

Table 3
Observed values of PCU and Heterogeneity Index (%).

Vehicle Categories	PCU			Heterogeneity Index (%)		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
Two-wheelers	0.18	0.22	0.20	52.43	125.73	90.14
Heavy vehicles	3.41	5.20	4.15			
Three-wheelers	0.66	1.08	0.85			
Bicycles	0.32	0.50	0.39			

HI. K-medoids clustering technique which is also known as 'Partitioning Around Medoids', performs clustering around a medoid and distributes the data-points among the pre-defined 'k' number of clusters. A medoid is a data-point that has the least dissimilarity (i.e. least sum of distances to other data-points) in the whole dataset. In the first step, the algorithm randomly selects 'k' number of data-points as medoids and calculates the distance from each data-point to different medoids. Based on the closest medoid, each data-point is moved to the corresponding cluster. Then, the total sum of distances for the developed clusters is calculated. Again, medoids are updated by substituting them with other data-points of the dataset. Repetitions of previous steps are continued until the maximum number of iterations is reached or there is no further need for rearranging data-points. Steps followed in K-medoids clustering are also exhibited in the flow-chart as shown in Fig. 6.

K-medoids clustering was performed on the observed data of HI separately considering the number clusters (k) as 3, 4, 5, and 6, and thereby three distinct sets of clusters were developed. It took nearly 150 iterations to reach saturation on each occasion. k was varied to identify the optimal number of clusters based upon the strength of the formation. The strength of each set of clusters was examined using two validation measures; Davies-Bouldin Index (DBI) and Silhouette Index (SI). DBI is defined as the ratio of 'the average within-cluster distances of data-points to their centres' to 'the between-cluster distances of centres of adjacent clusters' (Davies and Bouldin, 1979). Silhouette value (Rousseeuw, 1987) considers the average distance from a data-point to other points in the same cluster (a_i) and the minimum average distance from a data-point to other members in a different cluster (b_i). Silhouette value can be determined using the following formula.

$$SilhouetteValue = \frac{(b_i - a_i)}{\max(a_i, b_i)} \tag{9}$$

The average silhouette value of data-points within the same cluster is known as Silhouette width and the average of Silhouette widths of all clusters is known as Silhouette Index (SI). DBI and SI may vary within the range of '0 to 1′ and '-1 to 1′ respectively. Lower DBI and higher SI indicate the stronger form of clusters. Fig. 7(a) shows the variation in strength in terms of DBI and SI for varying k. The minimum DBI (0.50) and the maximum SI (0.62) which indicate the maximum strength, were found corresponding to three clusters. The silhouette value of each data-point within a cluster is also presented in Fig. 7(b) for three clusters.

Hence, the observed range of HI was classified into three categorical levels; Mild, Moderate and Severe. The upper and the lower boundaries of each cluster were considered as the thresholds of that particular heterogeneity level as given in Table 4. These heterogeneity levels will help the users to build up an easy perception about the intensity of the heterogeneity present within the mixed traffic stream for a given time.

4.5. Sensitivity analysis

Sensitivity analysis was performed to study the influences of traffic volume and its composition on HI. Classified traffic volumes (Q_i) i.e. the input variables considered in the GPR model were varied in such a manner so that it can capture the influence of traffic volume and traffic composition simultaneously on the average speed of individual vehicle categories. These individual speeds were then taken to Eq. (6) & (7) successively to monitor the influences on HI. Details of the analysis examining the influences of traffic volume, traffic composition, and classified speed on HI are discussed below in three phases.

4.5.1. Effect of traffic composition on heterogeneity index

In the first phase, the classified traffic volumes were varied in such a way so that the overall traffic volume could be kept constant at a pre-defined value and the proportions of different vehicle categories could be varied systematically within the

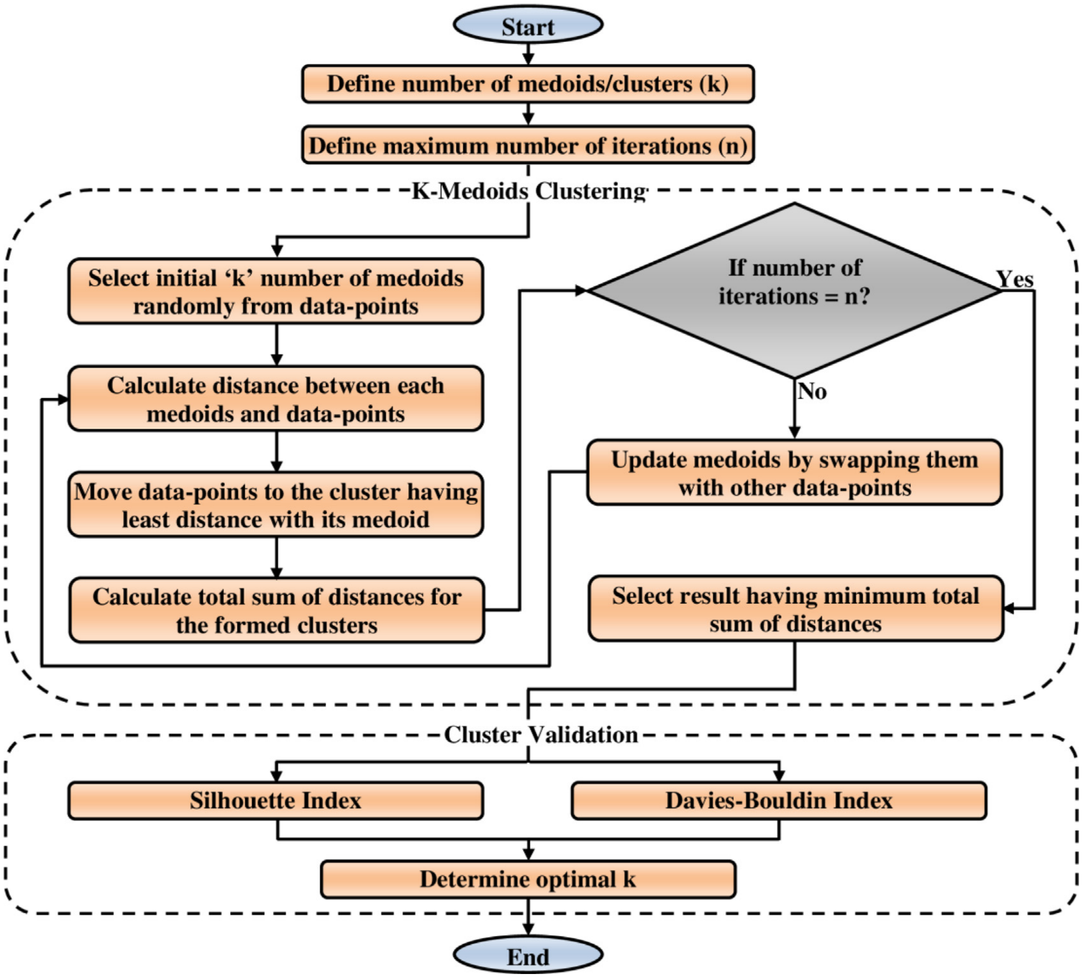


Fig. 6. Steps followed in K-medoids Clustering.

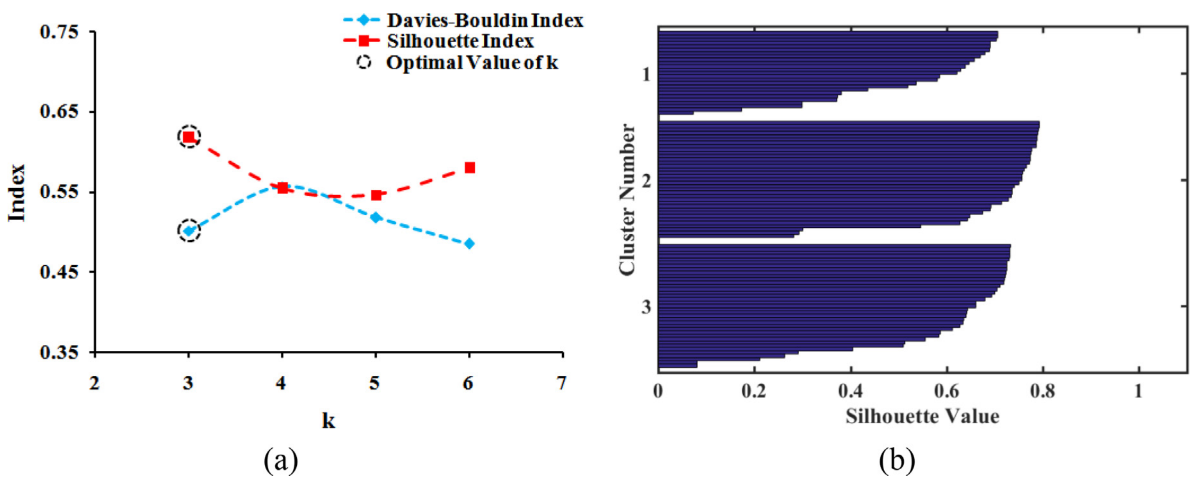


Fig. 7. Results of K-medoids clustering showing (a) obtained SI and DBI values for three, four, five, and six clusters and (b) Silhouette plot for three clusters.

observed ranges given in Table 1. Due to the limitation in graphical representation, the proportions of all vehicle categories could not be varied at a time. Firstly, the proportion of cars and two-wheelers which possess the major share in the traffic stream was varied simultaneously within the range of 35–45% and 25–35% respectively keeping the combined proportion

Table 4
Recommended heterogeneity levels and corresponding thresholds of HI.

Heterogeneity Levels	HI thresholds (%)
Mild	< 80
Moderate	80 – 100
Severe	> 100

constant at 70%. Later, the proportions of other vehicles were also varied. For each combination of classified volumes, HI was estimated using the GPR model and Eq. (6) & (7). Fig. 8 exhibits the variation of HI subject to the change in traffic composition.

It was observed that HI decreases with the increase in car proportion. Since the 'car' was considered as the standard vehicle in PCU estimation, an increase in its proportion replacing other vehicles reduces the heterogeneity prevalent within a traffic stream. If the traffic volume and other conditions remain unaltered, a 1% increase in the car proportion replacing the same proportion of two-wheeler leads to an approximate 2.5% decrease in HI. On the other hand, an increase in the proportion of heavy vehicles (trucks and buses) which possess higher PCU factors (Fig. 1) as compared to other vehicle categories, results in more heterogeneity in the traffic stream.

4.5.2. Effect of traffic volume on heterogeneity Index

In the second phase of sensitivity analysis, the classified traffic volumes in the GPR model were altered in such a way so that the traffic volume could be varied within its observed range while keeping the traffic composition constant primarily. Similar to the previous phase, HI was estimated for each combination of classified volumes and plotted against the overall traffic volume as shown in Fig. 9.

As may be seen, HI increases with the increase in traffic volume in a mixed traffic stream having a lesser proportion of cars. The impact of traffic volume on HI was observed to vary significantly depending upon the prevalent traffic composition. However, a maximum of 1.5% increase in HI was observed with every 100 veh/h increase in traffic volume. This trend can be justified in the following manner. For the same traffic composition, the increase in traffic volume reduces the average speed irrespective of vehicle category. However, the rate of this reduction is not similar for all vehicles. Generally, large-sized vehicles find it more difficult to manoeuvre at this increased traffic volume condition resulting in a high reduction rate of the average speed. On the other hand, small-sized vehicles generally possess high manoeuvrability and hence, become less affected when traffic volume increases. PCU which considers the ratio of the average speeds for standard cars and the subject

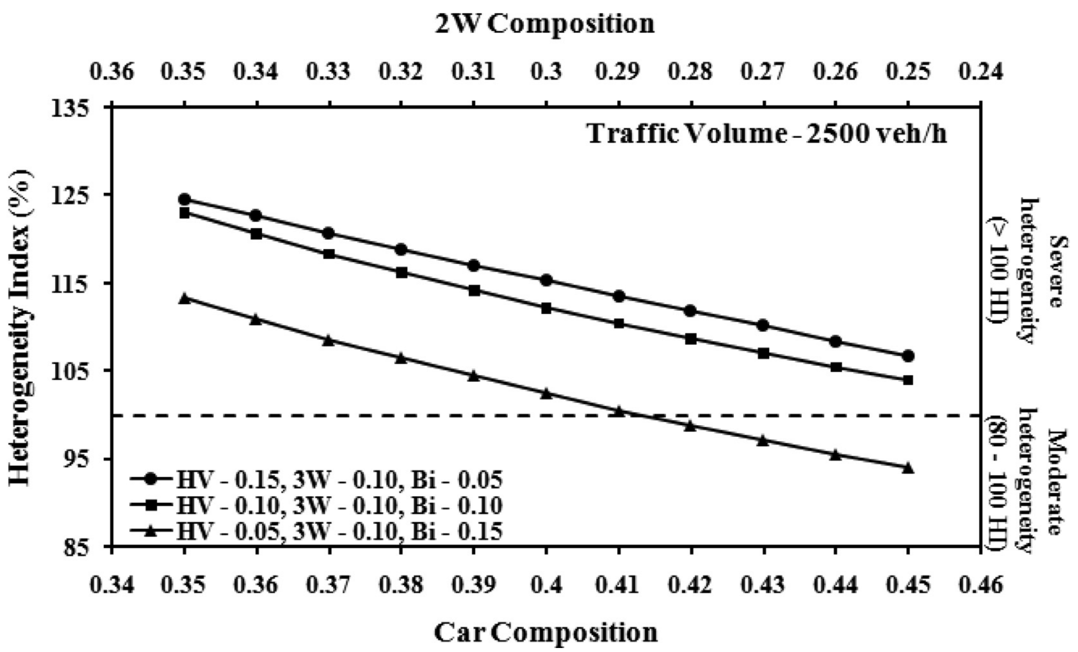


Fig. 8. Effect of traffic composition on Heterogeneity Index (%).

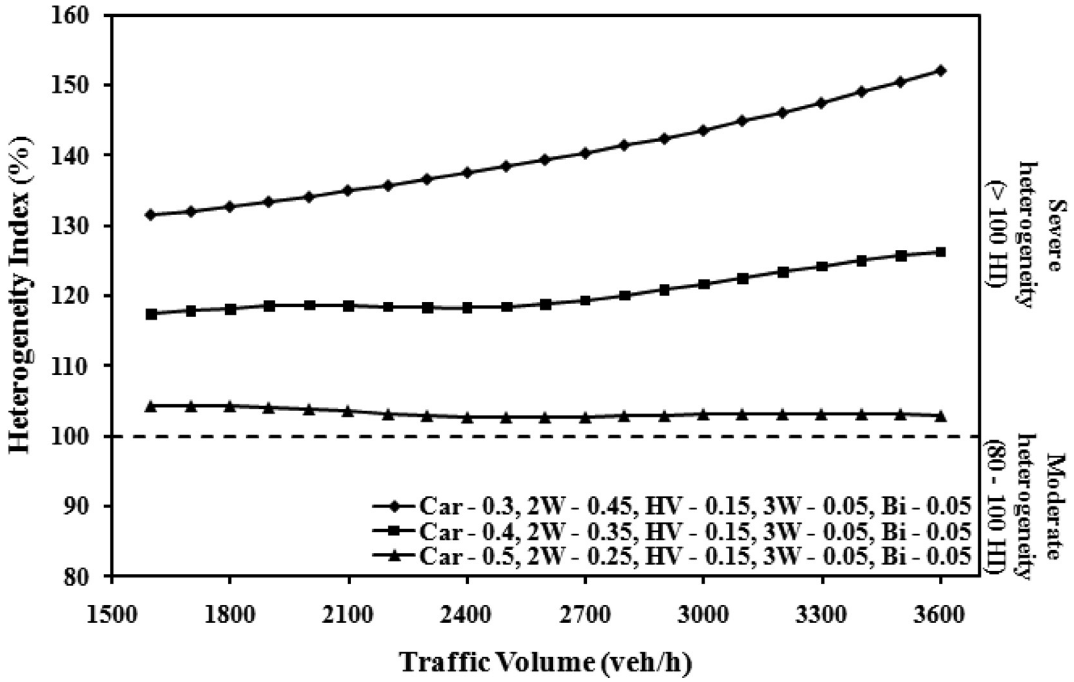


Fig. 9. Effect of traffic volume (veh/h) on Heterogeneity Index (%).

vehicle, changes accordingly. It was observed that due to the increase in traffic volume, PCU increases for large-sized vehicles but decreases for small vehicles. Therefore, in a traffic stream composed of different vehicle categories, the diversity among PCU factors increases with the increase in traffic volume. Subsequently, HI which captures this diversity also increases.

The above discussion is enough to expound on the significance of both traffic volume and its composition in determining the HI of a traffic stream. In fact, the change in traffic composition can take the traffic condition to a new heterogeneity level single-handedly as shown in Fig. 8. Therefore, the graphs developed in sensitivity analysis are useful to predict HI for a given combination of traffic volume and its composition.

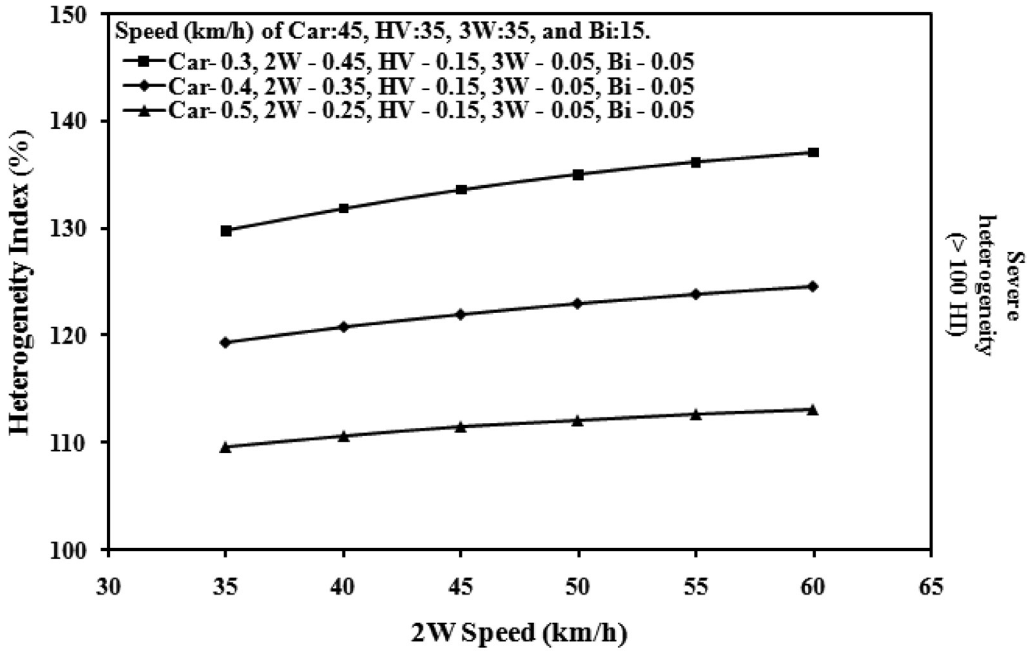
4.5.3. Effect of classified speed on heterogeneity index

After analyzing the effects of traffic composition and volume, the present study examined the effect of classified speeds on HI. For this analysis, the speed of all vehicle classes was kept constant at its observed means (rounded off to nearby suitable value) except the subject vehicle. The speed of the subject vehicle was varied from 35 to 60 km/h with a 5 km/h interval. If the classified speeds are known, there is no need to use the developed GPR-based speed models. The analysis was carried out at three different traffic compositions similar to the ones used in section 4.5.2. In the first phase, the speed of a small-sized vehicle (motorized two-wheelers) was increased gradually. With reference to Eq. (6), it is comprehended that PCU of 2 W decreases with the increase of its own speed. Since PCU of 2 W is already on the lower side as compared to other vehicles in the traffic stream, a further decrease in PCU_{2W} results in an increase in diversity among PCU values. Therefore, as may be seen in Fig. 10(a), the increase in speed of 2 W eventually leads to the increase in HI.

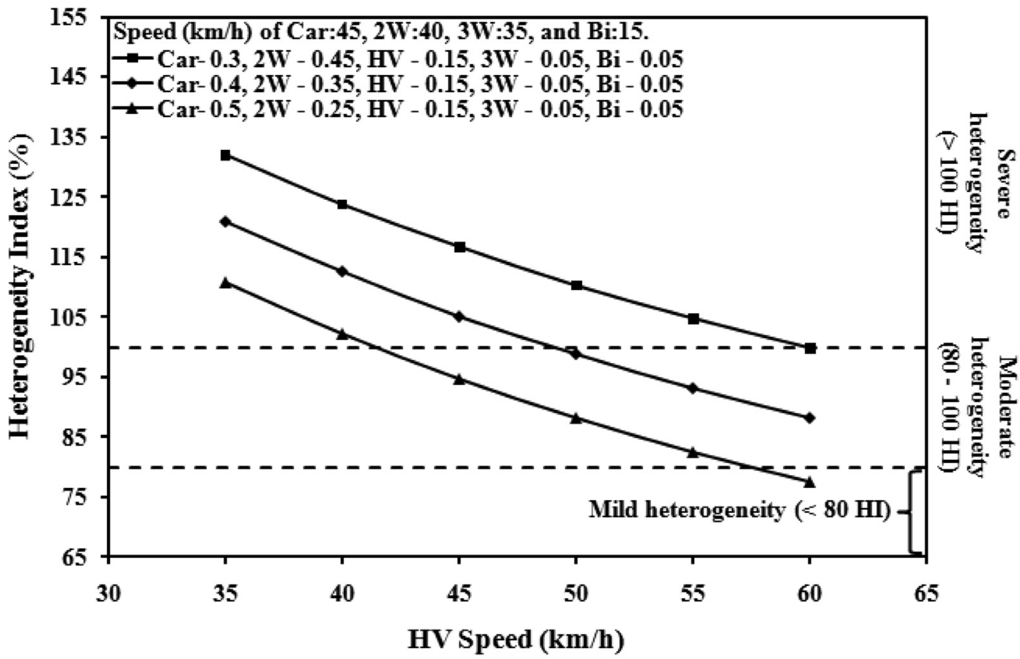
The exact opposite trend was observed when the speed of a large-sized vehicle increases in the traffic stream. In the second phase, the speed of the heavy vehicle (HV) was increased systematically from 35 to 60 km/h keeping other parameters unchanged. An increase in the speed of HV results in a decrease in PCU_{HV} as per Eq. (6). Since PCU_{HV} is on the higher side as compared to other vehicles in the traffic stream, a reduction in PCU_{HV} decreases the diversity among PCUs. Hence in Fig. 10 (b), HI which captures this diversity, decreases as a result of increasing the speed of HV. With each 5 km/h increase in the average speed of 2 W, HI was observed increasing by 2% if other prevailing conditions remain unchanged. In contrast, a 5 km/h increase in the average speed of HV causes an 8% decrease in HI. It is to be acknowledged here that these graphs (Fig. 10) are only useful when the classified speeds are known. As the collection of classified volume data is reasonably easier as compared to the classified speeds, the graphs (Fig. 8&9) exhibiting the impact of traffic volume and its composition on HI will be more useful from the user's perspective.

5. Conclusions

This paper presents a novel methodological framework to quantify the heterogeneity that prevailed within a heterogeneous traffic stream. The study conceptualizes a parameter called 'Heterogeneity Index' (HI) and forwards a Gaussian Process



(a)



(b)

Fig. 10. Effect of (a) two-wheeler speed and (b) heavy vehicle speed on Heterogeneity Index (%).

Regression-based model to estimate it. The model developed addresses the limitation associated with the existing model and also performs satisfactorily in predicting the Heterogeneity Index. Therefore, the outcomes of the study (GPR model, HI formula, and sensitivity graphs) will be useful to quantify the heterogeneity of a mixed traffic stream with varying traffic conditions. Other findings of this study are summarized below.

- The performance of GPR-based classified speed models is highly governed by the kernel function adopted in the development of the model. In this study, six different sets of speed models were developed using six popular kernel functions viz. Squared Exponential, Matern 3/2, Matern 5/2, ARD Squared Exponential, ARD Matern 3/2 and ARD Matern 5/2. Among these, ARD Matern 3/2 was found to be the most efficient kernel function in predicting the average speed of a vehicle category based on a given set of classified traffic volumes.
- To validate the performance of the model, a new set of traffic data collected at a different road section was considered. The model performed satisfactorily as it predicted the classified speeds with an average MAPE of 4.43% convincingly lower than the other existing model (5.15% MAPE).
- Ranges of PCU factors estimated for different vehicle categories are distinct and non-overlapping (0.18–0.22, 0.66–1.08, 3.41–5.20, and 0.32–0.50 respectively for two-wheeler, three-wheeler, heavy vehicle, and bicycle). This variation in PCU factors is accountable for the dynamic nature of HI which also varies with prevailing traffic conditions. HI was observed varying within the range of 52.43 to 125.73%.
- For the easy perception of the intensity of heterogeneity, the range of HI was divided into a few categorical levels. For this purpose, the K-medoids clustering technique was employed and the strength of developed clusters was assessed by Davies-Bouldin Index and Silhouette Index. Based on the results obtained in cluster-validation tests, this paper recommends three heterogeneity levels; a) Mild (<80% HI), b) Moderate (80–100% HI), and c) Severe (>100% HI).
- The impact of traffic composition on HI is so significant that the change in the factor can take the traffic condition to a new heterogeneity level single-handedly. HI increases with the increase in heavy vehicle's proportion or the decrease in car's proportion. As traffic volume increases, the PCU factor increases for large-sized vehicles but decreases for small-sized vehicles. Therefore, HI which represents the dispersion of PCU factors from its central value also increases with the increase in traffic volume.
- The influence of speed on HI is not similar for all vehicle categories. The increase in the speed of smaller sized vehicles leads to an additional increase in HI. However, the exact opposite trend was observed when the speed of a larger sized vehicle increases in the traffic stream.
- The present study proposes a methodology to estimate the HI of a traffic stream, particularly at mid-block sections. The work can be extended towards estimating the same at intersections and network-level. As mentioned earlier, there might be influences of heterogeneity of a traffic stream on lane discipline, congestion potential, and road users' safety. Therefore, the impacts of HI on these factors can be examined in future studies. Outcomes will be helpful for transportation planners in choosing the right policy to improve the lane discipline and to reduce the congestion and the crash potential of a transportation network.

Notations

The following symbols are used in this paper:

$\mathbf{A_c}$ = Projected rectangular area of car (m^2);
$\mathbf{A_i}$ = Projected rectangular area of vehicle type 'i' (m^2);
$\mathbf{a_i}$ = Average distance from a data-point to other points in the same cluster;
$\mathbf{b_i}$ = Minimum average distance from a data-point to other members in a different cluster;
d = total number of predicting variables;
f (x) = Latent variables;
$\mathbf{GP\ (0,\ k(x_i,\ x_j))}$ = Gaussian Process with the mean 0 and the covariance $k(x_i, x_j)$;
h (x) = Explicit basis function;
k = Number of clusters;
$\mathbf{k\ (x_i,\ x_j)}$ = kernel function showing covariance between inputs x_i and x_j where i and j are observation points and $i \neq j$, i, j =1,2,. . ..n;
$\mathbf{N\ (0,\ \sigma^2)}$ = Gaussian or normal distribution with a mean 0 and an error variance of σ^2;
$\mathbf{PCU_i}$ = Passenger Car Unit of a vehicle category 'i';
$\mathbf{Q_i}$ = Volume of a vehicle category 'i';
r = Euclidean distance between x_i and x_j;
$\mathbf{V_c}$ = Average speed of car (km/h);
$\mathbf{V_i}$ = Average speed of a vehicle category 'i';
$\mathbf{V_i^O}$ = Observed speed for data point 'i';
$\mathbf{V_i^P}$ = Predicted speed for data point 'i';
x = Multi-dimensional vector containing 'd' number of independent variables;
y = Response variable;
$\boldsymbol{\alpha_i}$ = Regression coefficients;
$\boldsymbol{\alpha_0}$ = Intercept;
$\boldsymbol{\beta}$ = Coefficient;

(*continued*)

$\mathbf{A_c}$ = Projected rectangular area of car (m^2);
$\boldsymbol{\varepsilon}$ = Gaussian noise;
$\boldsymbol{\theta}$ = Hyper-parameters or kernel parameters;
μ = Arithmetic mean;
$\boldsymbol{\sigma}$ = Standard deviation;
$\boldsymbol{\sigma_f}$ = Signal standard deviation;
$\boldsymbol{\sigma_l}$ = Characteristic length scale; and
$\boldsymbol{\sigma_m}$ = Separate length scale for each m^{th} predicting variable where m=1,2,3….d

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Not Applicable.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix

Table A1
Formulas of various kernel functions.

Kernel function	Abbreviation used	Formula	Eqn. No.
Squared Exponential	SE	$k(x_i, x_j\|\theta) = \sigma_f^2 e^{\left[-\frac{1}{2}\frac{(x_i - x_j)^T (x_i - x_j)}{\sigma_l^2}\right]}$	(10)
Matern 3/2	M32	$k(x_i, x_j\|\theta) = \sigma_f^2 \left(1 + \frac{\sqrt{3}r_1}{\sigma_l}\right) e^{\left[-\frac{\sqrt{3}r_1}{\sigma_l}\right]}$ Where,	(11)
		$\boldsymbol{r}_1 = \sqrt{(x_i - x_j)^T (x_i - x_j)}$	(12)
Matern 5/2	M52	$k(x_i, x_j\|\theta) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r_1}{\sigma_l} + \frac{5r_1^2}{3\sigma_l^2}\right) e^{\left[-\frac{\sqrt{5}r_1}{\sigma_l}\right]}$	(13)
ARD Squared Exponential	ASE	$k(x_i, x_j\|\theta) = \sigma_f^2 e^{\left[-\frac{1}{2}\sum_{m=1}^{d} \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}\right]}$	(14)
ARD Matern 3/2	AM32	$k(x_i, x_j\|\theta) = \sigma_f^2 \left(1 + \sqrt{3}r_2\right) e^{\left[-\sqrt{3}r_2\right]}$ Where,	(15)
		$\boldsymbol{r}_2 = \sqrt{\sum_{m=1}^{d} \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}$	(16)
ARD Matern 5/2	AM52	$k(x_i, x_j\|\theta) = \sigma_f^2 \left(1 + \sqrt{5}r_2 + \frac{5}{3}r_2^2\right) e^{\left[-\sqrt{5}r_2\right]}$	(17)

Table A2
The values of kernel parameters optimized using Quasi-Newton process.

Vehicle Categories	Kernel Parameters	Kernel Functions					
		SE	M32	M52	ASE	AM32	AM52
Car	σ_{L1}	165.1	171.9	175.0	173.5	176.2	185.2
	σ_{L2}				48016.3	43632.0	43091.6
	σ_{L3}				254.4	300.4	293.6
	σ_{L4}				88.5	92.0	93.6
	σ_{L5}				153.5	144.6	155.8
	σ_{F}	2.3	2.4	2.3	2.2	2.3	2.3
2 W	σ_{L1}	148.3	161.5	159.0	135.7	132.0	145.4
	σ_{L2}				152037.2	38828183.7	610820.0
	σ_{L3}				218.8	236.4	240.2
	σ_{L4}				82.2	98.8	93.8
	σ_{L5}				101865.6	1167646.7	92454.0
	σ_{F}	1.6	1.7	1.6	1.6	1.7	1.6
HV	σ_{L1}	38.5	33.6	35.3	223.3	383687.4	270658.7
	σ_{L2}				33.1	23.6	20.0
	σ_{L3}				285.7	1549381.3	410129.2
	σ_{L4}				39.6	10.6	10.1
	σ_{L5}				194296.5	1136004.7	1558615.8
	σ_{F}	4.8	4.8	4.8	3.4	3.7	3.7
3 W	σ_{L1}	109.5	36.3	87.8	261.7	256.4	265.0
	σ_{L2}				181979.6	40410783.8	101177.1
	σ_{L3}				122.5	116.9	121.0
	σ_{L4}				232111.3	417092.5	5422661.6
	σ_{L5}				16.8	15.4	16.0
	σ_{F}	1.3	3.3	1.5	2.8	3.1	3.0
Bicycle	σ_{L1}	121.1	130.2	130.7	282.0	299.6	297.3
	σ_{L2}				40328.3	216959.2	55388.7
	σ_{L3}				152.8	148.6	150.9
	σ_{L4}				58.4	56.0	58.9
	σ_{L5}				710242.2	296.2	297.5
	σ_{F}	0.6	0.6	0.6	0.5	0.6	0.6

References

Adnan, M., 2014. Passenger car equivalent factors in heterogenous traffic environment-are we using the right numbers? Procedia Eng. 77, 106–113. https://doi.org/10.1016/j.proeng.2014.07.004.

Aerde, M.V., Yagar, S., 1984. Capacity, Speed, and Platooning Vehicle Equivalents for Two-Lane Rural Highways. Transp. Res. Rec. J. Transp. Res. Board 971, 58–67.

Arasan, V.T., Krishnamurthy, K., 2008. Effect of traffic volume on PCU of vehicles under heterogeneous traffic conditions. Road Transp. Res. 17, 32–49.

Arkatkar, S.S., Arasan, V.T., 2010. Effect of gradient and its length on performance of vehicles under heterogeneous traffic conditions. J. Transp. Eng. 136, 1120–1136. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000177.

Basu, D., Maitra, S.R., Maitra, B., 2006. Modelling passenger car equivalency at an urban mid-block using stream speed as a measure of equivalence. Eur. Transp. Trasp. Eur. 34, 75–87.

Ben-Edigbe, J., Ferguson, N., 2005. Extent of capacity loss resulting from pavement distress. Proc. Inst. Civ. Eng. - Transp. 158, 27–32.

Biswas, S., Chakraborty, S., Chandra, S., Ghosh, I., 2017a. Kriging-Based Approach for Estimation of Vehicular Speed and Passenger Car Units on an Urban Arterial. J. Transp. Eng. Part A Syst. 143, 04016013. https://doi.org/10.1061/JTEPBS.0000031.

Biswas, S., Chandra, S., Ghosh, I., 2020. An advanced approach for estimation of PCU values on undivided urban roads under heterogeneous traffic conditions. Transp. Lett. 12, 172–181. https://doi.org/10.1080/19427867.2018.1563268.

Biswas, S., Chandra, S., Ghosh, I., 2017b. Estimation of Vehicular Speed and Passenger Car Equivalent Under Mixed Traffic Condition Using Artificial Neural Network. Arab. J. Sci. Eng. 42, 4099–4110. https://doi.org/10.1007/s13369-017-2597-9.

Biswas, S., Chandra, S., Ghosh, I., 2017c. Lambert W Function Based Semi-Analytical Framework for Estimation of Speed and Passenger Car Unit. Journal of the Eastern Asia Society for Transportation Studies 12, 1588–1603. https://doi.org/10.11175/easts.12.1588.

Chandra, S., Kumar, U., 2003. Effect of Lane Width on Capacity under Mixed Traffic Conditions in India. J. Transp. Eng. 129, 155–160. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:2(155).

Craus, J., Polus, A., Grinberg, I., 1980. A revised method for the determination of passenger car equivalencies. Transp. Res. Part A Gen. 14, 241–246. https://doi.org/10.1016/0191-2607(80)90040-0.

CSIR- Central Road Research Institute, 2017. Indian Highway Capacity Manual (Indo-HCM). New Delhi, India.

Dargay, J., Gately, D., Sommer, M., 2007. Vehicle Ownership and Income Growth, Worldwide: 1960–2030. Energy J. 28, 143–170. https://doi.org/10.2307/41323125.

Davies, D.L., Bouldin, D.W., 1979. A Cluster Separation Measure. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1, 224–227. https://doi.org/10.1109/TPAMI.1979.4766909.

Dhamaniya, A., Chandra, S., 2016. Conceptual Approach for Estimating Dynamic Passenger Car Units on Urban Arterial Roads by Using Simultaneous Equations. Transp. Res. Rec. J. Transp. Res. Board 2553, 108–116. https://doi.org/10.3141/2553-12.

Dhamaniya, A., Chandra, S., 2013. Concept of stream equivalency factor for heterogeneous traffic on urban arterial roads. J. Transp. Eng. ASCE 139, 1117–1123. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000581.

Fan, H.S.L., 1990. Passenger car equivalents for vehicles on Singapore expressways. Transp. Res. Part A Gen. 24, 391–396.

Greenshields, B.D., Bibbins, J.R., Channing, W.S., Miller, H.H., 1935. A Study of Traffic Capacity. Highw. Res. Board Proc. 14, 448–477.

Huber, M.J., 1982. Estimation of Passenger-Car Equivalents of Trucks in Traffic Stream. Transp. Res. Rec. J. Transp. Res. Board 869, 60–70.

Krammes, R.A., Crowley, K.W., 1986. Passenger Car Equivalents for Trucks on Level Freeway Segments. Transp. Res. Rec. 1091, 10–17.

Luca, M.D., Acqua, G.D., 2014. Calibrating the passenger car equivalent on italian two line highways: a case study. TRANSPORT 29, 449–456. https://doi.org/10.3846/16484142.2013.845854.

Manghat, D., Aswathy, K.V., Karuppanagounder, K., 2017. Automobile level of service criteria for two-lane undivided heterogeneous urban corridors. Eur. Transp./Trasp. Eur. 63, 1–14.

Matcha, B.N., Namasivayam, S.N., Fouladi, M.H., Ng, K.C., Sivanesan, S., Noum, S.Y.E., 2020. Simulation Strategies for Mixed Traffic Conditions: A Review of Car-Following Models and Simulation Frameworks. J. Eng. 1–22. https://doi.org/10.1155/2020/8231930.

Mishra, R., Kumar, P., Arkatkar, S.S., Sarkar, A.K., Joshi, G.J., 2017. Novel Area Occupancy-Based Method for Passenger Car Unit Estimation on Multilane Urban Roads Under Heterogeneous Traffic Scenario. Transp. Res. Rec. J. Transp. Res. Board 2615, 82–94. https://doi.org/10.3141/2615-10.

Pandey, A., Biswas, S., 2021. Development of 'speed ratio' based level of service criteria on undivided urban streets in mixed traffic context. Can. J. Civ. Eng. 48, 1169–1180. https://doi.org/10.1139/cjce-2020-0104.

Rahman, M., Nakamura, F., 2005. Measuring passenger car equivalents for non-motorized vehicle (rickshaws) at mid-block sections. J. East. Asia Soc. Transp. Stud. 6, 119–126. https://doi.org/10.11175/easts.6.119.

Raj, P., Sivagnanasundaram, K., Asaithambi, G., Ravi Shankar, A.U., 2019. Review of Methods for Estimation of Passenger Car Unit Values of Vehicles. J. Transp. Eng. Part A Syst. 145, 04019019. https://doi.org/10.1061/jtepbs.0000234.

Rongviriyapanich, T., Suppattrakul, C., 2005. Effects of motorcycles on traffic operations on arterial streets. J. East. Asia Soc. Transp. Stud. 6, 137–146. https://doi.org/10.11175/easts.6.137.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Saini, H.K., Biswas, S., 2021. Estimating lateral placement and lane indiscipline of urban mixed traffic of a developing country: An ANN-assisted approach. Can. J. Civ. Eng. 48, 1571–1581. https://doi.org/10.1139/cjce-2020-0250.

Schulz, E., Speekenbrink, M., Krause, A., 2018. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. J. Math. Psychol. 85, 1–16. https://doi.org/10.1016/j.jmp.2018.03.001.

Seguin, E.L., Crowley, K.W., Zweig, W.D., 1982. Passenger Car Equivalents on Urban Freeways: Report No. - FHWA/RD-81/156. Washington D.C.

Sharma, M., Biswas, S., 2020. Estimation of Passenger Car Unit on urban roads: A literature review. Int. J. Transp. Sci. Technol. 10.1016/j.ijtst.2020.07.002.

Shen, J., Yang, G., 2020. Crash Risk Assessment for Heterogeneity Traffic and Different Vehicle-Following Patterns Using Microscopic Traffic Flow Data. Sustain. 12, 1–18. https://doi.org/10.3390/su12239888.

Subotić, M., Tubić, V., Marić, B., 2016. PCE in analysis models of the number of following vehicles on a two-lane road. Int. J. Traffic Transp. Eng. 6, 25–37. https://doi.org/10.7708/ijtte.2016.6(1).03.

Suvin, P.V., Karmakar, A.R., Islam, S.Y., Mallikarjuna, C., 2020. Criteria for Temporal Aggregation of the Traffic Data from a Heterogeneous Traffic Stream. Transp. Res. Procedia 48, 3401–3412. https://doi.org/10.1016/j.trpro.2020.08.114.

Transport and Road Research Laboratory, 1965. Research on Road Traffic. London.

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

Wan, Z.Y., Sapsis, T.P., 2017. Reduced-space Gaussian Process Regression for Data-Driven Probabilistic Forecast of Chaotic Dynamical Systems. Phys. D Nonlinear Phenom. 345, 40–55. https://doi.org/10.1016/j.physd.2016.12.005.

Weber, E.U., Shafir, S., Blais, A.R., 2004. Predicting Risk Sensitivity in Humans and Lower Animals: Risk as Variance or Coefficient of Variation. Psychol. Rev. 111, 430–445. https://doi.org/10.1037/0033-295X.111.2.430.

Wen, H., Sun, J., Zeng, Q., Zhang, X., Yuan, Q., 2018. The Effects of Traffic Composition on Freeway Crash Frequency by Injury Severity: A Bayesian Multivariate Spatial Modeling Approach. J. Adv. Transp. 1–7. https://doi.org/10.1155/2018/6964828.

Werner, A., Morrall, J.F., 1976. Passenger Car Equivalencies of Trucks, Buses, and Recreational Vehicles for Two-Lane Rural Highways. Transp. Res. Rec. J. Transp. Res. Board 615, 10–17.

Yang, S., Wu, J., Qi, G., Tian, K., 2017. Analysis of traffic state variation patterns for urban road network based on spectral clustering. Adv. Mech. Eng. 9, 1–11. https://doi.org/10.1177/1687814017723790.

Yeung, J.S., Wong, Y.D., Secadiningrat, J.R., 2015. Lane-harmonised passenger car equivalents for heterogeneous expressway traffic. Transp. Res. Part A 78, 361–370. https://doi.org/10.1016/j.tra.2015.06.001.