ORIGINAL ARTICLE

Multivariate Analysis on Dynamic Car-Following Data of Non-lane-Based Traffic Environments

Sanhita Das[1] · **Anuj Budhkar**[2] · **Akhilesh Kumar Maurya**[1] · **Avijit Maji**[3]

Abstract

The difficulties of the microscopic models in accurate representation of the real traffic phenomena stem from its complexities in the collection and processing of reliable time-series car-following data of non-lane-based traffic environments. Proper estimation of car-following data can suitably ameliorate the realism of traffic sub-models and is still a demanding task. This study describes an image-based in-vehicle trajectory data collection system for the estimation of reliable dynamic time-series data, using camera calibration and in-vehicle GPS information. A copula-based methodological framework is also investigated in this study for evaluating safety in the car-following processes, by accommodating the dependence structure of longitudinal gap, centerline separation and vehicle speeds. Results of the study demonstrated the importance of centerline separation in apprehending the car-following processes. In particular, the probability of maintaining lower gaps increases with the decrease in speed and increase in centerline separation. A 15–20% reduction in the longitudinal gaps is observed for speeds greater than 60 kmph. As importantly, the study recommends the applicability of tri-variate Gaussian copula in assessing the safety or 'safe distance-keeping' criteria of drivers in the car-following processes, which can indeed augment the accurate representation of drivers' behavior and development of the car-following models, advanced driver assistance systems and for safety evaluation in the car-following process of non-lane-based traffic environments.

Keywords Instrumented vehicle · Heterogeneous · Mixed traffic · Safety · Centerline separation

Introduction

The utilization of microscopic simulation modeling and intelligent transport systems (ITS) to evaluate traffic system performance has proved the inadequacy of the existing models, including car-following ones, to accurately represent the behavioral phenomena in complex real-world contexts. Despite the recent advancements of new technologies (such as ITS) for road system applications, achieving a detailed understanding of the dynamic behavioral response of drivers in 'car-following processes' has been a major safety concern since decades [1].

Although many studies have supported the car-following theories and development of its subsequent sub-models, calibration and validation of the models as well as an empirical verification of the underlying assumptions have led to serious difficulties with both collection and processing of accurate, unbiased time-series data in a common space–time reference system [2, 3]. As improved accuracy in the experimental data collection can substantially ameliorate the behavioral phenomena from a microscopic perspective and the realism of traffic sub-models, proper estimation of time-series data has still proven to be challenging.

Anecdotal evidence, however, demonstrates several techniques for gathering time-series data which include, static laboratory simulators [4], instrumented vehicles equipped

✉ Sanhita Das
  sanhi.das25@gmail.com

  Anuj Budhkar
  anujbudhkar@gmail.com

  Akhilesh Kumar Maurya
  maurya@iitg.ac.in

  Avijit Maji
  avimaji@civil.iitb.ac.in

1 Department of Civil Engineering, Indian Institute of Technology, Guwahati, Assam 781039, India

2 Department of Civil Engineering, Indian Institute of Engineering Science and Technology, Shibpur, West Bengal 711103, India

3 Department of Civil Engineering, Indian Institute of Technology, Bombay, Maharashtra 400076, India

with radar sensors [1, 5, 6] or onboard global positioning system (GPS) receivers. Although driving simulator-based experiments lack the flexibility of representing real traffic phenomena; howbeit, in the domain of instrumented vehicles, the dynamics of the following vehicle can be thoroughly analyzed for long stretches along a route. On one hand, instrumented vehicles can adequately capture the relative spacing and speeds using sensors; while on the other hand, an accurate positional data of the equipped vehicles can be obtained with the satellite-based GPS technology [7].

The advancements in real-time GPS technology have expedited a new horizon in the field of traffic engineering. In the context of car-following theories, Hatipkarasulu et al. [8] and Jiang and Li [9] have underlined the flexibility of vehicle-mounted GPS receivers to collect the position and speed data with a positional accuracy of 1–5 m and speed accuracy of 0.16 kmph. Additionally, by utilizing time-series data for a platoon of ten vehicles in a probing field, Gurusinghe et al. [7] demonstrated the superiority of the GPS technology in accurately estimating data than other conventional methods. Punzo and Simonelli [2] used kinematic differential GPS instruments to investigate the methodological issues of car-following model calibration and validation by accurately monitoring the trajectories of four vehicles in a platoon under real traffic conditions. The peer-reviewed literature has also elucidated the efficacy of real-time differential GPS receivers in suitably modeling the following behavior of cars [10–12].

Despite the high expected accuracy of kinematic GPS in vehicle tracking, the estimation of time-series trajectories to apprehend the car-following process of weak-lane discipline traffic is indeed a demanding task. Recent literature has underlined the importance of lateral descriptor (centerline separation) of traffic in modeling the staggered-following behavior [13–16], where the following vehicles often tend to maintain some lateral separation with its immediate leader, either to perceive the forward visual field with more confidence or to anticipate the behavioral response of the front vehicles and its associated proximity risks in the car-following processes. However, the requirements for staggered-following trajectory data are indeed stringent. For instance, in the field of satellite-based GPS technology, inter-vehicle longitudinal spacing can be obtained from the recorded vehicle positions with an accuracy of 1 m; on the contrary, the estimation of lateral separation from the GPS receivers may produce unreliable results. Understanding that a proper evaluation of car-following behavior in non-lane-based traffic environments requires an accurate characterization of the microscopic traffic variables and reliable experimental data, a suitable integration of the satellite-based GPS technology with an image-based data collection scheme can substantially improve the collection and processing of continuous time-series data for the staggered-following scenario.

Vehicle detection and tracking in mixed traffic conditions remain a challenge. Intrusive and non-intrusive techniques have been tried and tested; however, it is observed that only the image-based technique is able to trace vehicles with better accuracy [17]. Further, several efforts have been made to extract vehicle trajectories from video images. Some efforts were initially made by Metkari et al. [18], Jin et al. [19] and others by marking road sections by strips. In some works, the road sections were physically marked with strips of known width. On the other hand, efforts were made to calibrate the field data-points from their corresponding points visible in the image. Fukui [20], Courtney et al. [21], Bas and Chrisman [22] and Fung et al. [23] are the leading efforts in this regard. Errors were observed with changing camera orientations, and modifications necessary to avoid errors due to camera orientation angles are best described in Fung et al. [23]. Camera calibration techniques and image detection techniques are simultaneously used in several vehicle trajectory data extraction softwares, and the most popular one for mixed traffic is traffic analyzer and enumerator (TRAZER). However, these softwares need camera at fixed position and are not useful if dynamic data need to be obtained from within the stream. There is, thus, a need for semi-manual data collection method, which can manually classify and trace vehicle trajectories of neighboring vehicles. Softwares based on this concept were used to obtain vehicle trajectories [24, 25] or to study inter-vehicular gaps [26, 27]. However, in their study, the authors recorded a video of traffic stream from a static camera placed at a high vantage point. The entire vehicular interaction cannot be recorded by this technique. If the entire interaction between vehicles needs to be captured, one has to either compromise on the video resolution to mark the trajectory (thereby increasing error) or deploy complicated methods like video-stitching by recording from several simultaneous videos which were not attempted earlier. It is, therefore, necessary to use an instrumented vehicle for capturing video of the entire car-following interactions.

In this context, efforts are being directed to the development of a suitable data collection technique for estimation of reliable dynamic time-series car-following data for non-lane-based traffic. The study reported in this paper attempts to establish an image-based in-vehicle trajectory data collection system to process the microscopic variables (such as longitudinal gap, centerline separation, vehicle speeds, accelerations, etc.) using camera calibration and in-vehicle GPS information on straight roads, and to provide a copula-based methodological approach for the safety evaluation of vehicles in car-following process.

Firstly, a brief overview of the car-following processes is described, followed by a detailed description of the data collection methodology which forms the core of this research and the resulting data estimation process is provided. A

discussion of the safety evaluation in the car-following process and methodological approach examining the safety criteria follows. Finally, analysis and conclusions providing an overview of the suitability of data collection exercise complete the paper.

Fig. 1 A typical car-following 'spiral' showing the variation of relative speed with longitudinal spacing

Car-Following Behavior

The car-following behavior is a control process in which each driver attempts to maintain a desired following distance behind the lead vehicle by accelerating or decelerating in response to the actions of the vehicle in-front. This continuous adjustment in vehicle speeds is governed by the drivers' perceptions of adjudging the leading vehicle speed as well as their own speeds, proximity to the desired following headway (or, longitudinal gap, LG) and response time delay which may overcompensate for small deviations from a target point [28, 29], resulting in the following 'spirals' which can be expressed in the relative speed-relative spacing plane (Fig. 1).

Such spirals are indicative of the car-following processes of the driver across all speed ranges in staggered-following scenario and his/her perceptions to maintain a safe desired headway. The following behavior in weak-lane discipline traffic, however, requires the consideration of an indicator of lateral interaction [30], namely centerline separation (CS) for a precise representation of staggered-following behavior.
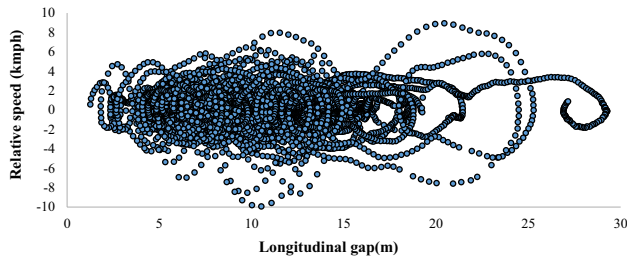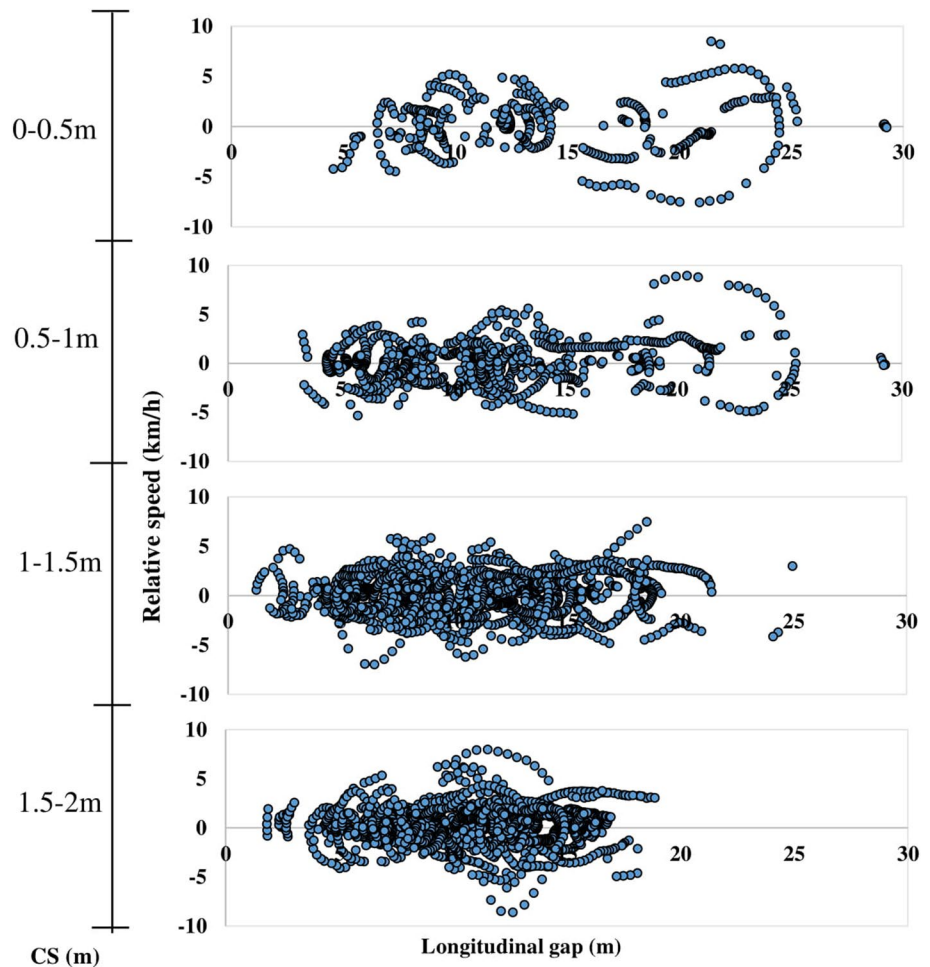
Fig. 2 Typical following 'spirals' for different CS levels

In essence, centerline separation describes the off-centeredness between the leading and following vehicles in the car-following process. Figure 2 illustrates the car-following 'spirals' for different CS levels of staggered-following scenario.

The 'spirals' in the figure envisage that the continuous adjustment of the driver to the behavior of the vehicle in-front occurs at all levels of CS and the inter-vehicle spacing gradually decreases with the increase in staggered positions of vehicles. This is an indicative of a typical staggered-following behavior of non-lane-based traffic environments where the following vehicles often tend to circumvent the encumbrance of the lead vehicle to have a clear forward view and anticipate the criticality of the car-following process. As a result, they tend to maintain large centerline separation, thereby resulting in reduced spacing.

Such behavior illustrates the consideration of CS in the car-following theories, an accurate data estimation and processing of which will enrich the realism of model development and representation of drivers behavior in simulation modeling.

Experimental Data

Details of the data used in this study and a comprehensive discussion about the experimental setup, calibration process and the issues regarding data collection are provided in this section.

Instrumented Vehicle Setup and Data Collection

The data used in this study were collected from a series of experiments conducted in March 2018 along straight sections of the NH-27, a four-lane divided carriageway in Guwahati, India, under real traffic conditions. Experiments were conducted by driving two vehicles in the car-following state along rural highway covering a total stretch length of 37.33 km during the afternoon hours, 2:00–4:30 p.m. The vehicles were equipped with GPS receivers [Racelogic video V-box] that recorded the position of each vehicle at 0.1-s interval, and a video system attached to the windshield of the cars, allowing a complete visual record of the experiment. From the positional data of GPS receivers, vehicle speeds, longitudinal/lateral acceleration are calculated through successive derivations of the space traveled; while inter-vehicle longitudinal and lateral gaps are evaluated from the synchronized recorded video by a semi-automated trajectory extracting system [detail discussion in the "Calibration process"].

In the process of data collection, careful attention was devoted to the experimental setup. The drivers of both the vehicles were familiar with the traveled path, but were not aware of the purpose of experimental design. The driver of the leading vehicle was instructed to maintain different speeds in the range of 30–90 kmph at an interval of 10 kmph each (that is, 30 kmph, 40 kmph, 50 kmph, 60 kmph, 70 kmph, 80 kmph and 90 kmph) based on prevailing traffic conditions of the considered road stretch. On the other hand, the driver of the following vehicle was asked to follow the leading vehicle maintaining a safe distance based on his perception. In response to the actions of the leading vehicle, the following process of the subject vehicle was then captured based on the proposed data collection technique. As the aim of the experiment was to collect car-following data, the lane-changing maneuver and intrusions of other vehicles were avoided. Data for any unimpeded intrusions in unavoidable circumstances were discarded. In particular, the car-following data were processed for only such scenarios when the subject vehicle was under the influence of the leading instrumented vehicle. Moreover, the study was conducted for only straight sections where effect of external factors such as curves and gradients affecting the traffic flow was not considered. An example of a car-following scenario and a staggered-following scenario recorded from the camera attached to the following vehicle is presented in Fig. 3.

(a)

(b)
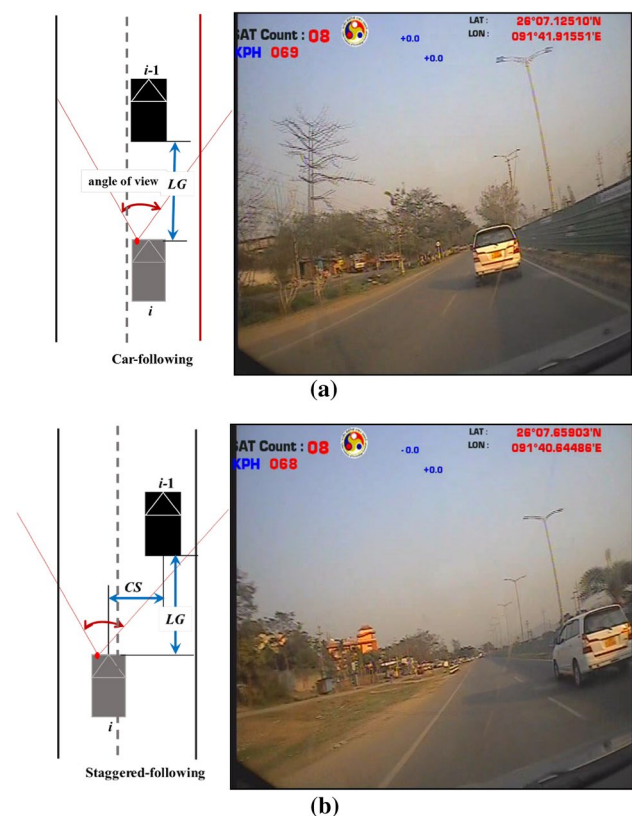
Note: LG-Longitudinal gap; CS- Centerline separation; Red dot indicates the position of the camera attached to the following vehicle, *i*

Fig. 3 A typical **a** car-following and **b** staggered-following scenario recorded from the camera attached to the following vehicle. *LG* longitudinal gap, *CS* centerline separation; Red dot indicates the position of the camera attached to the following vehicle, i (color figure online)

Calibration Process

In an attempt to extract and analyze the inter-vehicle spacing (both longitudinal and lateral) from the video footage, a semi-automated trajectory system is employed in this research where each vehicle's location on the screen is recognized and tracked at different time stamps using frame-by-frame analysis. The screen coordinates of the vehicle's position recorded at each tracking are then transformed to real-world coordinates by employing the camera calibration equations devised by Fung et al. [23]. For brevity the detailed process is not discussed in this paper. The same calibration technique was also used to study the car-following behavior and lateral gap maintaining behavior of vehicles in non-lane-based traffic scenario [17, 26, 27].

The four reference points in the video sequence and their respective coordinates in the real world need to be captured and based on the calculated camera parameters (pan angle, tilt angle, swing angle, focal length and distance of the plane from the camera lens) and screen coordinate of each point, the real-world coordinate of the respective point can be estimated. As described in Fig. 4, four endpoints of the road edge markings are selected in this process to form a calibration pattern of rectangle *ABCD*.

With reference to Fig. 4, four endpoints representing an exact rectangle in the world coordinates of known dimensions (AB = CD and AC = BD) are selected and using the semi-automated trajectory extractor, the corresponding end-point is traced to obtain its respective screen coordinates. Each end-point is, however, clicked several times on the screen, and for further data estimation and processing, the average pixel value for each point is considered to reduce errors in manual mouse-clicks. For the estimation of

inter-vehicle spacing, it is considered that the edge of the test vehicle (EG or FH) is parallel to the road edge (AB or CD), and the longitudinal and lateral distance of the end-point (either B or D) from one front corner of the test vehicle (either E or F) is known. For any point P representing the mid-point of the rear bumper of the leading vehicle, the transformed real-world coordinates are estimated similar to the methodology described above. With all known measured distances and transformed real-world coordinates, the inter-vehicle spacing can be accurately estimated according to the expression described below:
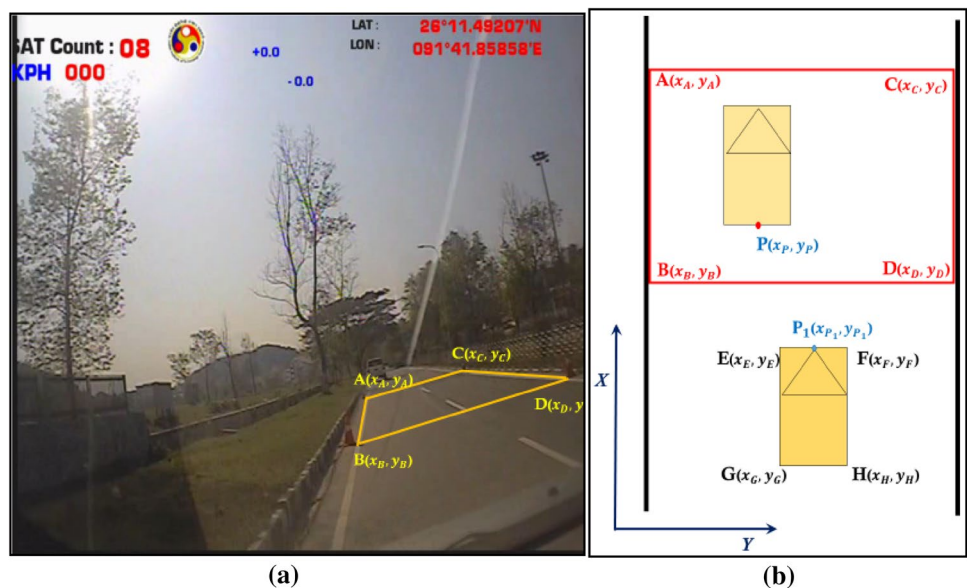
$$\text{Longitudinal gap} = (x_P - x_B) + (x_B - x_E),$$

$$\text{Centerline separation} = \left| y_P - y_B - (y_E - y_B) + (y_{P1} - y_E) \right|,$$

where the distances $(x_B - x_E)$, $(y_E - y_B)$ and $(y_{P1} - y_E)$ are accurately measured in the field during the calibration exercise (that is, before the start of experiment when the vehicle is stationary).

The real time of sampling, the global coordinates and speed of each vehicle involved in the experiment are directly obtained from the GPS receivers at 10-Hz frequency; while the positional data of the vehicle in-front are extracted from the video recorders at each 5-Hz frequency level. After suitable transformation, the GPS data and the extracted video data are then synchronized to obtain speed of both the vehicles, global time, longitudinal gap and centerline separation at every 0.2-s intervals. While extracting data from the video, an extractor may make an erroneous mouse click with a fixed standard deviation (in pixels) on the screen, and an average extractor makes mouse-clicking with an accuracy of six pixels [26]. These six pixels correspond to different

Fig. 4 Camera calibration technique considered in the study **a** rectangle ABCD viewed from the camera **b** top view (or, bird eye view) of the calibration pattern along with the lead vehicle

(a)   **(b)**

distances in the field, and error in calculating longitudinal gap will increase as field distance from point B (refer Fig. 4) increases. The maximum error (normally distributed, 3σ) increases from 1.2% for a point exactly on point B, to 11.7% for a point 30 m ahead of B for the experiment conducted on Guwahati roads (when point B was 8 m ahead of test vehicle's front edge). Moreover, the accuracy in the speed measurement is 0.16 kmph [31].

The underlying assumptions of the experimental setup process are (a) the road under surveillance should be reasonably straight such that both the vehicles rest on the same plane and (b) careful attention should be paid to the fixture of the camera on the vehicle since any change in the camera orientation may result in improper data estimation. Such considerations can indeed produce significant reliable and accurate experimental data for understanding the car-following process of non-lane-based traffic streams.

Results

Preliminary Analysis

The synchronized GPS and extracted video data resulted in 7002 cases of car-following events (data being obtained at each 0.2-s intervals) which would essentially help in understanding the following behavior as well as its associated proximity risks. A summary of the descriptive statistics of the inter-vehicle spacing (both longitudinal and lateral) and vehicle speeds for the extracted data is presented in Table 1.

Statistics of the traffic variables show that the car-following data cover a wide range of vehicle speeds lying in the range of 9–78 kmph. The extent of off-centeredness (CS) of the subject vehicle with the front leader is observed to range from a minimum of zero to a maximum of 2.50 m. The observed range of longitudinal gap further justifies that the dataset covers the entire spectrum of all car–car interactions because the peer-reviewed literature suggests a maximum longitudinal gap of 30 m for a close-following behavior [15, 32].

The skewness and kurtosis values of longitudinal gap further indicate that the distributions have sharper peaks and heavier tails than that of a normal distribution; while kurtosis value of CS indicates that the data are normally distributed. Most of the speed data are characterized by skewness close to zero and negative skewness values indicating that the speed data are more symmetrically distributed than that of longitudinal gap, the tails of the distribution are lighter and have a flatter peak than a normal distribution; a comparison of the mean and median values further corroborates this result.

Dependence Structure Between the Traffic Variables

With an aim to comprehend the behavioral characteristics of longitudinal gap (LG), vehicle speed and CS and the potential dependence relationship between them, a summary of mean and median values of inter-vehicle spacing (LG, CS) for different ranges of vehicle speeds is presented in Table 2.

The table shows that the mean and median values of longitudinal gap increase with the increase in following vehicles speeds and decrease with the increase in centerline separation. Further, vehicle speeds are observed to follow a decreasing trend with centerline separation. This observed trend is indicative of the fact that longitudinal

Table 1 Descriptive statistics of the microscopic traffic variables

Statistical parameters	Longitudinal gap (m)	Centerline separation (m)	Speed (kmph)
Mean	12.32	1.39	49.96
Median	11.89	1.44	49.47
Standard deviation	4.31	0.48	10.11
Minimum	2.94	0.00	9.09
Maximum	29.88	2.50	78.38
Skewness	0.68	− 0.46	0.03
Kurtosis	0.61	0.03	− 0.10

Table 2 Statistics of the variables for different speed ranges

Speed Range (kmph)	Longitudinal gap (m)		Centerline separation (m)		Speed (kmph)		Sample
	Mean	Median	Mean	Median	Mean	Median	
< 20	4.97	3.81	1.76	1.81	15.21	16.67	25
20–30	7.45	7.19	1.48	1.48	25.62	26.79	92
30–40	10.47	10.74	1.55	1.54	36.75	37.01	803
40–50	11.92	11.92	1.48	1.53	43.96	43.78	2659
50–60	12.79	11.81	1.27	1.28	55.13	55.06	2213
60–70	13.86	13.56	1.28	1.34	63.76	63.18	1044
70–80	15.44	13.46	1.25	1.27	72.78	72.29	167

gap and vehicle speeds are dependent on the lateral positioning of vehicles in car-following processes of non-lane-based traffic streams. Because following vehicles at large CS encounters a wide forward field of view and tends to avoid the encumbrance of lead vehicles, they tend to follow the leaders closely maintaining lower longitudinal gap and speed. The correlation coefficients of LG and CS ($\tau = -0.257, \rho_s = -0.376, r = -0.439$), speed and CS ($\tau = -0.198, \rho_s = -0.293, r = -0.264$), and LG and speed ($\tau = -0.172, \rho_s = 0.254, r = 0.293$) further corroborate the dependence relationship of the microscopic traffic variables.

To attain additional insights into the statistical differences of inter-vehicle spacing according to the following vehicle's speeds, a one-way analysis of variance test is conducted; the results of which indicated significant statistical difference in longitudinal gaps [$F(6,6996) = 110.46; p < 0.001$] and centerline separations [$F(6,6996) = 70.51; p < 0.001$] across all speed ranges. This finding justifies that the longitudinal and lateral separations of the subject vehicle with the vehicle in-front vary significantly according to the speeds of the following vehicle in the car-following processes.

Phenomenology of Safe 'Distance-Keeping' in the Car-Following Scenario

Achieving a comprehensive understanding of how drivers control their vehicles while following another vehicle, how substantial the safety problem is, how much percentage of drivers exhibit such unsafe critical events and how the safety requirements can be modeled, is still a matter of debate and requires more attention. The homeostasis theory proposed by Wilde [33] emphasizes that a driver always tend to maintain a constant level of risk exposure by adjusting his behavior, that is, by controlling speed and headway. Similarly, Summala [34] proposed the 'zero-risk theory' which states that a driver acts to control the risk level when the risk exceeds the safety margins. Although quantifying the 'safety' is difficult, many pioneering efforts were devoted to recommend the threshold values for headways which could essentially distinguish between relatively safe and dangerous encounters in the car-following processes of lane-based traffic [35–40]. Anecdotal literature on the car-following models, however, suggest the consideration of different parameters namely, space headway, speed, relative speed, relative acceleration, desired speed, maximum acceleration, etc. to predict the decision of the following vehicle in response to the actions of the leader. Amongst all, space headway (summation of longitudinal gap and vehicle length) and speed form the basis of most of the car-following models. The drivers in real-world scenario can actually perceive available distance more precisely than time measurement and considering further the utilization of safe longitudinal gap information in the development of car-following models (safety distance

models or Gipp's model), evaluation of safe longitudinal gap is assessed in this study rather than time-headway. Particularly, with regard to non-lane-based traffic scenario, there is still a paucity of research concerning the applicability and evaluation of safety criteria, where in addition to speed and headway, centerline separation has proved to be an essential indicator in describing the vehicle following processes.

Understanding that the utilization of safety indicator in weak-lane discipline traffic requires an integration of the longitudinal (longitudinal gap and vehicle speed) as well as the lateral descriptor (centerline separation), the headway thresholds recommended in the literature may not represent the actual scenario in non-lane-based traffic environments as these minimum thresholds are envisaged to vary across different lateral positions of vehicles. A methodological approach using copulas is thus employed in this research to accommodate the dependence structure of the time-series data of longitudinal gap, vehicle speeds and centerline separation. A concoction of the peer-reviewed literature reveals that the average minimum safe and comfortable time-headways in car-following scenario lie in the range of 0.64–1.78 s [35, 36, 38–40] for vehicle speeds of 45–150 kmph.

With an aim to quantify the 'safety' in car-following processes, the lower 5% values of longitudinal gaps at each speed and centerline separation are selected in this study. Firstly, the CS data were segregated into different groups (0–0.5 m, 0.5–1 m, 1–1.5 m, 1.5–2 m and 2–2.5 m) and for each CS group, the corresponding dataset of longitudinal gap and speeds were separated as well. Secondly, a variety of speeds belonging to each CS group were then further separated into approximate speed values (± 3 kmph) of 30 kmph, 40 kmph, 50 kmph, 60 kmph, 70 kmph, 80 kmph and 90 kmph and the corresponding LG data for each speed were also separated. Finally, for each demarcated CS group and approximate speed value, the minimum 5% data of LG were selected for evaluating the safe distance requirements in the car-following process. This lower 5% LG values obtained for each speed and CS ranges are then compared with the recommended thresholds described in the existing literature and are accordingly, anticipated as the minimum safe distance-keeping requirements for the driver in the following scenario. Figure 5 depicts variation of the average safe longitudinal gap (considering lower 5% values) with vehicle speeds and centerline separation.

The plots depict a pragmatic increasing relationship of the longitudinal gap with speed for each level of CS and also a decreasing trend of safe longitudinal gap and speed with CS is observed for speeds of 30 kmph, 40 kmph, 50 kmph, 60 kmph, 70 kmph and 80 kmph. Considering lower CS values as a replication of actual car-following events of lane-based traffic, a direct comparison of the safe longitudinal gap (Fig. 5) with the recommended thresholds can be acquired for CS less than 0.5 m. In our study, the average

Fig. 5 Variation of the safe longitudinal gap with speed and CS
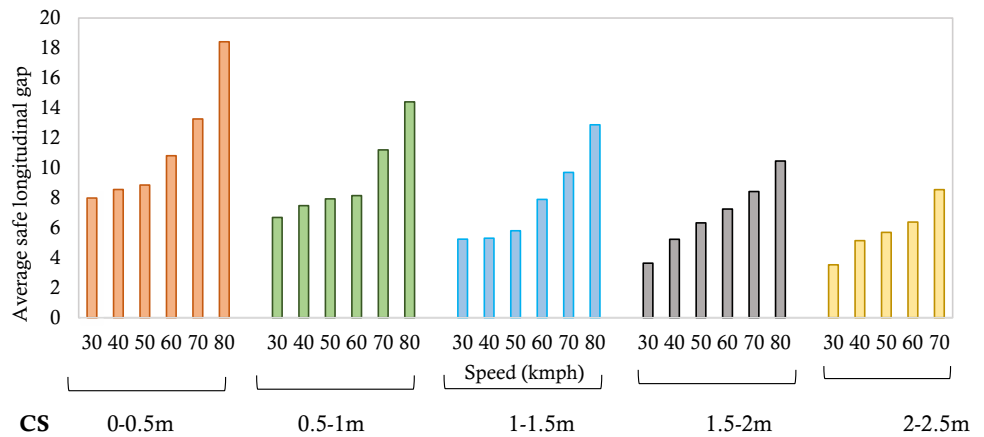
safe longitudinal gaps for speed of 40 kmph, 50 kmph and 80 kmph are obtained as 8.55 m, 8.85 m and 18.41 m, respectively, which is similar to the safe LG value indicated by Duan et al. [40] where they found the safe longitudinal gap range as 8.70–23.10 m for speeds of 45 kmph and 90 kmph. Similarly, in a study by Taieb-Maimon and Shimar [38], the range of safe time headway obtained was 0.64–0.69 s for vehicle speeds of 50 kmph, 60 kmph, 70 kmph, 80 kmph, 90 kmph and 100 kmph. Direct comparison of the recommended safe longitudinal gap value (8.88–19.17 m) with our data (8.85–18.41 m for speeds in the range of 50–80 kmph) indicated that the safe longitudinal gaps considered in our study for different vehicle speeds are similar to the values obtained by Taieb-Maimon and Shinar's [38] work. This indeed justifies that the average safe longitudinal gaps considered in our study (corresponding to lower 5% values) can be used to represent the safe following conditions in car-following and staggered-following cases of non-lane-based traffic environments. The lower 5% values of longitudinal gaps corresponding to each speed and CS level are, therefore, utilized for the development of the copula model.

Application of Copulas in Safety Evaluation

As identified previously, the risk level of drivers defined for lane-based homogeneous traffic may not imply the same level of risk in non-lane-based traffic cases. Considering the dependence relationship among safe LG, speed and CS, a copula-based methodological framework can address the relationship of micro-level parameters used in the car-following models with the lateral descriptor which will, in turn, ascertain the propensity of crash risks at any lateral positioning of vehicle in non-lane-based traffic streams.

Concept of Copulas

A copula is a joint cumulative distribution function that links a stochastic multivariate relationship to its univariate

marginal distributions of any dimension, such that each margin is uniformly distributed in [0, 1]. For uniformly distributed continuous random variables $(X_1, X_2, \ldots, X_n)$ with marginal cumulative distribution functions $(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))$, a joint n-dimensional cumulative distribution function (CDF) $F(x_1, x_2, \ldots, x_n)$ can be generated as follows:

$$
\begin{aligned}
F(x_1, x_2, \ldots, x_n) &= \mathbb{P}(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n) \\
&= \mathbb{P}(F_1(X_1) < F_1(x_1), F_2(X_2) \\
&\quad < F_2(x_2), \ldots, F_n(X_n) < F_n(x_n)) \\
&= C_\theta(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))
\end{aligned}
\tag{1}
$$

Equation (1) shows that the joint CDF F can be described by the margins $F_1, F_2, \ldots, F_n$ and the bivariate copula C, which captures the dependency structure among $X_1, X_2, \ldots, X_n$.

Selection of Univariate Marginal Distributions

The first step in the estimation of a suitable copula model requires proper selection of optimal marginal probability distributions. Several probability distributions were employed in this study to select the best-fitted marginal probability distribution function for safe LG data, speed and CS data. The best-fitted distribution was selected when the Akaike-Information Criteria (AIC) and three statistical goodness-of-fit tests namely Kolmogorov–Smirnov (K–S) test, Anderson–Darling (A–D) and Cramer–von Mises (C–vM) test statistics were minimal, rendering the null hypothesis unable to be rejected at $\alpha = 0.05$ [16, 41]. Table 3 presents a summary of the results of the goodness-of-fit test.

Based on the goodness-of-fit index, AIC and log-likelihood criterion, logistic distribution was selected as the best fitted for safe longitudinal gap [location = 12.514, scale = 2.893] and speed data [location = 47.221,

Table 3 Results of goodness-of-fit tests

Variables	Marginal distributions	Log-likelihood	AIC	Goodness-of-fit measures		
				K–S	CvM	A–D
Longitudinal gap	Burr	− 2916.96	5837.91	0.063	0.591	3.496
	Logistic	**− 2915.74**	**5837.48**	**0.054**	**0.492**	**2.562**
	Weibull	− 2921.91	5847.82	0.064	0.743	3.873
	Gamma	− 2949.63	5903.26	0.068	0.779	5.199
	Lognormal	− 2943.52	5891.05	0.088	1.339	8.339
CS	Lognormal	− 1022.52	2049.03	0.169	9.429	51.464
	Logistic	**− 774.27**	**1552.54**	**0.080**	**1.816**	**10.733**
	Weibull	− 754.76	1513.51	0.050	0.493	3.413
	Gamma	− 867.36	1738.72	0.125	5.332	28.747
	Normal	**− 748.65**	**1511.05**	**0.057**	**0.758**	**4.347**
Speed	Skew-t	− 4452.11	8908.21	0.482	8.241	40.117
	Lognormal	− 3773.22	7550.44	0.207	5.607	31.137
	Logistic	**− 3656.51**	**7317.02**	**0.148**	**5.024**	**26.179**
	Gamma	− 3714.38	7442.76	0.189	4.973	26.667
	Normal	− 3662.02	7328.04	0.152	4.824	25.133

Bold values indicate the best-fitted univariate model

scale = 5.995]; whereas, normal distribution [mean = 1.284, standard deviation = 0.525] provided the best fit for CS.

Estimation of Suitable Copula Model

Constructing a joint distribution model for 3D variables (LG, CS, speed) requires good assessment of suitable tri-variate copula functions in comparison with their observations. The selection of suitable copula model was performed intuitively based on the designated dependence domains and the nature of association of the data [42].

Considering the flexibility of multivariate Gaussian copula in modeling all ranges of dependence structure (− 1, 1) and the widespread popularity of Archimedean copulas, five copula functions namely, Gaussian, Gumbel, Frank, Clayton and Joe are employed in this study. However, since the multivariate Archimedean copulas (Frank, Gumbel, Clayton and Joe) lack the flexibility of modeling negative dependence structure for higher dimensions ($n \geq 3$), the original CS data are transformed to a new variable CS', where $CS' = \frac{1}{CS}$, and subsequently the tri-variate Frank, Gumbel, Clayton and Joe copulas are used to fit the LG, speed and the transformed centerline separation data. Moreover, the performances of tri-variate Gaussian, Frank, Clayton and Joe copulas with/

without the transformed variables are assessed based on log-likelihood and AIC criterion, and the parameters of each copula are estimated using maximum pseudo-likelihood method. The dependence parameter value (θ) of each copula function along with the log-likelihood (LL) and Akaike's information criterion (AIC) values are listed in Table 4.

The results indicate that tri-variate Gaussian copula showed the highest log-likelihood and the lowest AIC values for LG, CS and speed data, followed by Frank, Clayton, Gumbel and Joe copulas, respectively. On the basis of performance measures of the copula functions presented in Table 4, it is, thus, adjudged that tri-variate Gaussian copula, with logistic distribution for LG and speed, and normal distribution for CS, could be further investigated to assess the level of safety in the following processes of non-lane-based traffic environments.

Non-Exceedance Conditional Probability Distributions

For a particular non-exceedance conditional probability, several combinations of the microscopic variables exceeding a certain threshold can be obtained from the developed tri-variate Gaussian copula model. Figure 6 shows the non-exceedance conditional probabilities of safe longitudinal gap

Table 4 Parameter estimation and its associated LL values for the transformed tri-variate Archimedean copulas and Gaussian copula

	Gaussian [LG, CS, Speed]	Archimedean copulas [LG, CS', Speed]			
		Gumbel	Clayton	Frank	Joe
θ	− 0.51, 0.45, − 0.35	1.346	0.459	2.55	1.494
LL	− 7084.631	− 8430.171	− 8393.996	− 8361.705	− 8492.915
AIC	14175.262	16862.342	16789.992	16725.41	16987.83

Fig. 6 Conditional probabilities $\mathbb{P}(LG \leq lg | VS \geq vs, CS \geq cs)$ with centreline separation being equal to **a** 0.10 m and **b** 1.50 m
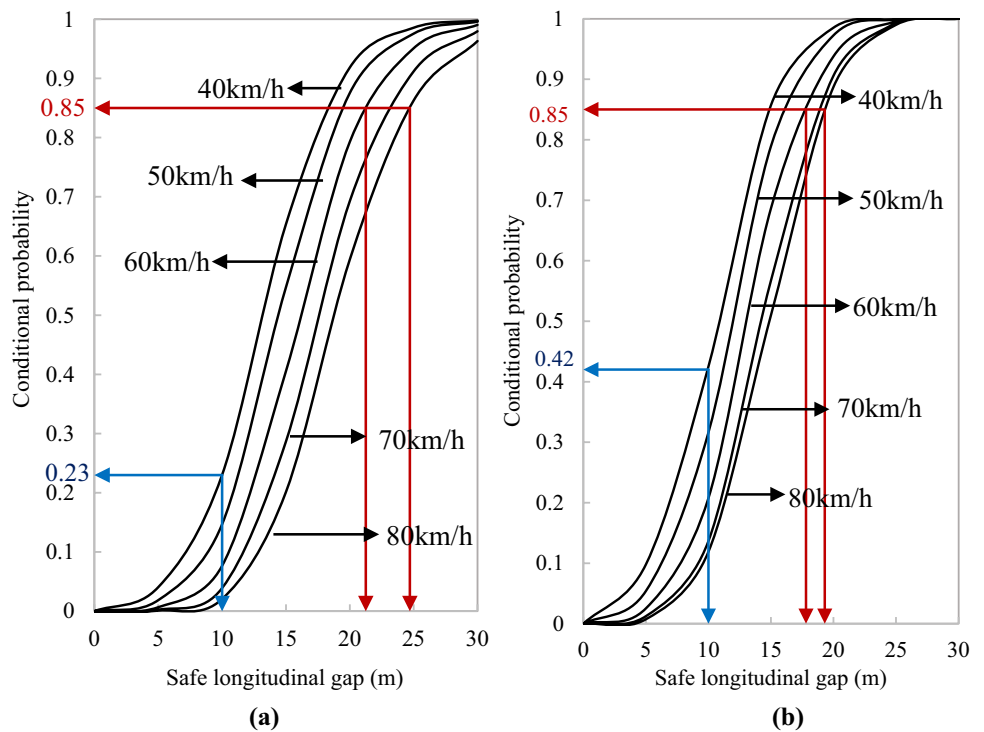
given both vehicle speeds (VS) and centerline separation exceeding certain thresholds.

From the figure, it can be observed that the conditional distribution of safe longitudinal gap gradually increases with the corresponding conditional factor (vehicle speed) decreasing and also it increases with the increase in centerline separation. For instance, the non-exceedance conditional probability for safe longitudinal gap $\mathbb{P}(LG \leq 10m)$ with speeds exceeding 40 kmph in car-following state (CS = 0.1 m) is 0.23; whereas for large centerline separation (CS = 1.5 m), the probability is obtained as 0.42. This increase in probability is due to the staggered car-following in which the drivers tend to follow the leaders more closely maintaining larger centerline separation in order to have a clear field of view of the forward scenario. Similarly, for a conditional probability of 0.85, the safe longitudinal gaps corresponding to vehicle speeds exceeding 60 kmph and 80 kmph in car-following state are obtained as 21.3 m and 24.7 m; whereas longitudinal gaps of 17.8 m and 19.3 m are observed for the staggered-following scenario, respectively (about 15–20% decrease in longitudinal gaps for staggered following). This clearly indicates that the decreasing relationship of LG with CS holds true for each value of vehicle speed.

The joint probability density distributions of the longitudinal gap and centerline for vehicle speeds exceeding thresholds of 40 kmph, 50 kmph, 60 kmph and 70 kmph are represented in Fig. 7.

The bivariate LG–CS plots of Fig. 7 clearly depict a reciprocal dependent relationship between the 2D variables

for each speed level. It can be observed that there is a gradual rightward shift in the LG-axis as speed increases, which is quite expected. For example, a majority of the drivers in car-following state (say CS = 0) are observed to maintain safe longitudinal gaps of 17.5 m, 18.75 m, 22.5 m, 23.75 m for speeds exceeding 40 kmph, 50 kmph, 60 kmph and 70 kmph, respectively; while for any arbitrary CS value (say CS = 1 m), the corresponding longitudinal gaps are observed as 14.75 m, 16 m, 17.5 m and 20 m, respectively, which justifies the negative dependent relationship of LG and speed with CS and positive degree of association between LG and speed. The results of the study, therefore, signify that car-drivers in staggered-following cases perceive the scenario ahead with more confidence and can anticipate the lead vehicle's behavior with better predictability; therefore, they usually tend to follow the leading vehicles closely maintaining lower longitudinal gaps and speeds at higher CS levels.

Such information can provide useful insights in the car-following models, by which the safe 'distance-keeping' requirements at different speeds and CS values can be evaluated, which indeed can help in a better replication and representation of actual drivers' behavior in the following processes.
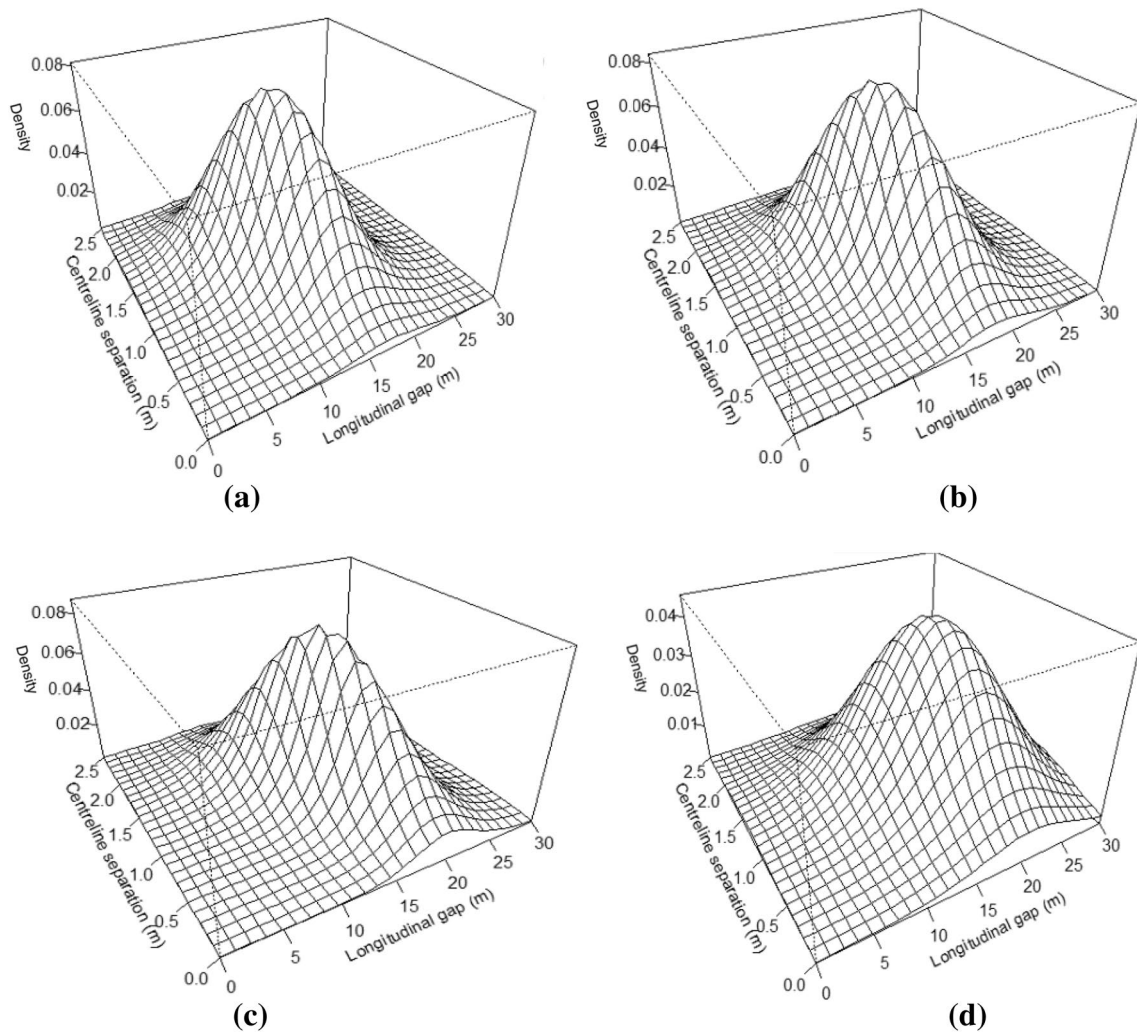
Fig. 7 Joint probability density plots for safe LG and CS with **a** speed ≥ 40 kmph and **b** speed ≥ 50 kmph, **c** speed ≥ 60 kmph and **d** speed ≥ 70 kmph

Possible Applications of the Current Study

The tri-variate Gaussian copula can provide useful insights in the development of car-following models, by which the safe 'distance-keeping' requirements at different speeds and CS values can be evaluated. The widely used safety distance car-following models hypothesize that drivers usually try to maintain a sufficient distance with the leading vehicles at different speeds so as to avoid a collision if the leading vehicle suddenly applies brakes. In such models, the information on the safe longitudinal gap thresholds obtained from the copulas at different speeds and CSs can be directly utilized in the model development which will help in a better replication and representation of actual drivers' behavior in the microsimulation models.

In addition, the results of the study can also find possible applications in providing advanced collision warning system adaptation databases for the evaluation of conflict severities in the car-following process of non-lane-based traffic streams. With the advent of driving-assistance systems, vehicles are now equipped with adaptive cruise control (ACC)/collision avoidance systems (CAS) in which the system maintains a safe headway to the vehicle in-front according to the settings predefined by the users, and also warns the drivers of upcoming potential threats and imminent collisions. In particular, for a given probability value, if the safe longitudinal gap exceeds the available gap at specified CS and speed levels, the warning and intervention systems used for detecting safety hazards in the car-following processes may be triggered, so that any risks of imminent collisions can be avoided. The work undertaken in this study can, therefore, be useful in the development of car-following models, advanced driver assistance systems and for safety

evaluation in the car-following process of non-lane-based traffic streams.

Conclusions and Future Scope

The major contribution of this study lies in the focus on car-following data collection and estimation of non-lane-based traffic environments. Efforts are being directed to the development of an image-based in-vehicle trajectory data collection system for an estimation of reliable dynamic time-series car-following data using camera calibration and in-vehicle GPS information. This study not only describes the stringent requirements of the data but also provides a copula-based methodological framework for the safety evaluation of vehicles in the following processes.

First, the car-following 'spirals' indicated that the drivers' adjustments of speeds in the following process occur at all levels of centerline separations and the inter-vehicle spacing follows a decreasing trend with the lateral separation between the vehicles. Preliminary analysis on longitudinal gap (LG), speed and centerline separation (CS) also corroborated a reciprocal dependent relationship between LG and CS, speed and CS, and a positive dependent relationship between LG and speed, complementing earlier researches in mixed traffic. The drivers in car-following state (say CS = 0) are observed to maintain safe longitudinal gaps varying from 17.5 to 23.75 m for speeds exceeding 40–70 kmph, while the gaps for 1 m CS are observed in the range of 14.75–20 m for the same speeds. Understanding that an estimation of the safe longitudinal spacing in the following process can inherently enrich the realism of car-following model development and representation of realistic drivers' behavior, a tri-variate copula model was developed considering lower 5% longitudinal gap data for each speed and CS level. Based on the performance measures of the copula functions, a tri-variate Gaussian model with logistic distribution for LG and speed, and normal distribution for CS, was found to assess the level of safety in the following processes of non-lane-based traffic environments. The conditional and joint probability distributions further demonstrated the importance of CS in modeling the car-following behavior of non-lane-based traffic streams.

The results obtained in this study can find suitable applications in the car-following model development, advanced driver assistance systems and in safety evaluation of non-lane-based traffic environments as discussed in the "Possible applications of the current study". For future scope, the results of the study can be extended for other leader–follower vehicle pairs that are more prevalent in mixed traffic streams. Various driver behaviors such as drifting, shying away and closing in reactions, changing lateral positions and the overtaking behavior can be studied using camera calibration methodology. The copula models ensure robust

representation of obtained datasets thereby informing the effect of all explanatory terms in gap-keeping. Also, further studies may need to be conducted in diverse cities to capture the driving behavior variability.

References

1. Brackstone M, McDonald M (2007) Driver headway: how close is too close on a motorway? Ergonomics 50(8):1183–1195
2. Punzo V, Simonelli F (2005) Analysis and comparison of microscopic traffic flow models with real traffic microscopic data. Transp Res Rec 1934:53–63
3. Punzo V, Formisano D, Torrieri V (2005) Part 1: traffic flow theory and car following: nonstationary kalman filter for estimation of accurate and consistent car-following data. Transp Res Rec 1934:1–12
4. Winsum W, Heino A (1996) Choice of time headway in car following and the role of time to collision information in braking. Ergonomics 39:579–592
5. Brackstone M, Sultan B, McDonald M (2002) Motorway driver behaviour: studies on car following. Transp Res Part F Traffic Psychol Behav 5(1):31–46
6. Brackstone M, Waterson B, McDonald M (2009) Determinants of following headway in congested traffic. Transp Res Part F Traffic Psychol Behav 12(2):131–142
7. Gurusinghe GS, Nakatsuji T, Azuta Y, Ranjitkar P, Tanaboriboon Y (2002) Multiple car-following data using real-time kinematic global positioning system. Transp Res Rec 1802:166–180
8. Hatipkarasulu Y, Wolshon B, Quiroga CA (2000) Global positioning system approach for the analysis of car-following behavior. In: Transp Res Board 79th Annual Meeting of the Transportation Research Board, Washington, DC
9. Jiang Y, Li S (2001) Measuring and analyzing vehicle position and speed data at work zones using global positioning system. In: Transp Res Board 80th Annual Meeting of the Transportation Research Board, Washington, DC
10. Shekleton S, Australia M (2002) A GPS study of car following theory. In: Conference of Australian Institutes of Transport Research (CAITR)
11. Mathew TV, Ravishankar KVR (2012) Neural network based vehicle-following model for mixed traffic conditions. Eur Transp 52:1–15
12. Colombaroni C, Fusco G (2014) Artificial neural network models for car following: experimental analysis and calibration issues. J Intel Transp Syst 18(1):5–16
13. Gunay B (2007) Car following theory with lateral discomfort. Transp Res Part B Methodol 41(7):722–735
14. Jin S, Wang D, Tao P, Li P (2010) Non-lane-based full velocity difference car following model. Phys A 389(21):4654–4662
15. Zheng L, Zhong S, Jin PJ, Ma S (2012) Influence of lateral discomfort on the stability of traffic flow based on visual angle car-following model. Phys A 391(23):5948–5959
16. Das S, Maurya AK (2018) Multivariate analysis of microscopic traffic variables using copulas in staggered car-following conditions. Transp A Transp Sci 14(10):1–28

17. Budhkar AK, Maurya AK (2017) Multiple-leader vehicle-following behavior in heterogeneous weak lane discipline traffic. Transp Dev Econ 3(2):20

18. Metkari M, Budhkar A, Maurya AK (2013) Development of simulation model for heterogeneous traffic with no lane discipline. Procedia Soc Behav Sci 104:360–369

19. Jin S, Huang Z, Tao P, Wang D (2011) Car-following theory of steady-state traffic flow using time-to-collision. J Zhejiang Univ Sci A 12(8):645–654

20. Fukui I (1981) TV image processing to determine the position of a robot vehicle. Pattern Recognit 14(1–6):101–109

21. Courtney JW, Magee MJ, Aggarwal JK (1984) Robot guidance using computer vision. Pattern Recognit 17(6):585–592

22. Bas EK, Crisman JD (1997) An easy to install camera calibration for traffic monitoring. In: Proceedings of IEEE Conference on Intel Transp Syst, pp 362–366

23. Fung GS, Yung NH, Pang GK (2003) Camera calibration from road lane markings. Opt Eng 42(10):2967–2977. https://doi.org/10.1117/1.1606458

24. Munigety CR, Vicraman V, Mathew TV (2014) Semiautomated tool for extraction of microlevel traffic data from videographic survey. J Transp Res Rec 2443:88–95

25. Kanagaraj V, Asaithambi G, Toledo T, Lee T (2015) Trajectory data and flow characteristics of mixed traffic. J Transp Res Rec. https://doi.org/10.3141/2491-01

26. Budhkar AK, Maurya AK (2016) Modeling gap-maintenance in heterogeneous no lane discipline traffic. Presented in 1st International Conference of Transportation Infrastructure and Materials (ICTIM-2016), Chang'an University, Xi'an, China. ISBN: 978-1-60595-367-0

27. Budhkar AK, Maurya AK (2017) Characteristics of lateral vehicular interactions in heterogeneous traffic with weak lane discipline. J Mod Transp. https://doi.org/10.1007/s40534-017-0130-1

28. Schmidt R (1993) Unintended acceleration: human performance considerations. In: Peacock B, Karwowski W (eds) Automotive ergonomics. Taylor and Francis, London

29. Braun M, Coleman CS, Drew DA (1983) Modules in applied mathematics: volume 1 differential equation models. Springer, New York. https://doi.org/10.1007/978-1-4612-5427-0. ISBN 978-1-4612-5429-4

30. Kanagaraj V, Treiber M (2018) Self-driven particle model for mixed traffic and other disordered flows. Phys A 509:1–11

31. Racelogic (2015) RACELOGIC support centre. https://racelogic.support/. Accessed 24 Nov 2017

32. Asaithambi G, Joseph J (2018) Modeling duration of lateral shifts in mixed traffic conditions. J Transp Eng Part A Syst 144(9):04018055

33. Wilde GJS (1982) The theory of risk homeostasis: implications for safety and health. Risk Anal 2:209–225

34. Summala H (1988) Risk control is not risk adjustment: the zero-risk theory of driver behavior and its implications. Ergonomics 31(4):491–506

35. Siebert FW, Oehl M, Pfister HR (2014) The influence of time headway on subjective driver states in adaptive cruise control. Transp Res Part F Traffic Psychol Behav 25:65–73

36. Purucker C, Rüger F, Schneider N, Neukum A, Färber B (2014) Comparing the perception of critical longitudinal distances between dynamic driving simulation, test track and vehicle in the loop. Advances in human aspects of Transp. In: Proceedings of the 5th AHFE Conference. Krakau, Polen, vol 19, pp 421–430

37. Ota H (1994) Distance headway behavior between vehicles from the viewpoint of proxemics. IATSS Res 18(2):5–14

38. Taieb-Maimon M, Shinar D (2001) Minimum and comfortable driving headways: reality versus perception. Hum Factors 43(1):159–172

39. Ayres TJ, Li L, Schleuning D, Young D (2001) Preferred time-headway of highway drivers. In: Proceedings of Intel Transpn Syst, IEEE, pp 826–829

40. Duan J, Li Z, Salvendy G (2013) Risk illusions in car following: is a smaller headway always perceived as more dangerous? Saf Sci 53:25–33

41. Das S, Maurya AK (2018) Bivariate modeling of time headways in mixed traffic streams: a copula approach. Transp Lett. https://doi.org/10.1080/19427867.2018.1537209

42. Chowdhary H, Escobar LA, Singh VP (2011) Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. Hydrol Res 42(2–3):193–216