SN Computer Science (2023) 4:287
https://doi.org/10.1007/s42979-023-01756-7

ORIGINAL RESEARCH

Microscopic Analysis of Time Headway of Vehicles on Two-Lane Highways with Mixed Traffic

Rupali Roy[1] · Pritam Saha[1]

Received: 4 January 2022 / Accepted: 21 February 2023 / Published online: 27 March 2023

Abstract
The efficacy of single distributions to model time headway data for mixed traffic with faster and slower vehicles is marginal. Effects of such traffic exaggerate on two-lane roads where both directions of traffic use a single carriageway. This paper aims to provide insights on modelling headways in this context based on field data. Shifted Erlang, and Exponential distributions can model headways under low flow when platoons are infrequent, and the proportion of shorter headways is insignificant. Moderate and heavy flow witness frequent interactions, platooning and significant slowing of vehicles. The use of mixture distribution exhibits aptness in describing the headways of the following and free vehicles; however, it entails a complex mathematical approach for parameter calibration, aggravating further if traffic is mixed. The paper introduces the concept of combined distribution as an alternative. It illustrates the effectiveness of the combined distribution function with shorter and longer headways modelled separately while describing headway data, respectively, at moderate and heavy flow. Lognormal and Gamma distribution models aptly described the shorter headways. Lognormal, Gamma and negative exponential models were found appropriate for longer ones. The combined models confirmed that the probability of shorter headways at moderate and heavy flow is relatively high compared to longer ones.

Keywords Highways · Mixed traffic · Time headway · Distributions · Goodness-of-fit test · Probability plot

Introduction

Time headway between successive vehicles is a critical microscopic traffic-flow parameter. It helps to perform a wide range of analyses for road traffic, including road safety studies, car following-lane changing behaviour, capacity and level-of-services [1]. Thus, researching the characteristics of vehicle headways and then developing appropriate distribution models have been an essential look-out to the traffic analysts while increasing roadway capacity and reducing vehicle delays. Besides, suitable headway distribution models must be applied in microscopic simulation models to generate possible vehicle events. While using the negative exponential function to describe the empirical distribution of headway data is quite common, several researchers reported alternatives that can clearly illustrate such distribution, especially when the prevalent traffic is mixed in character and exhibits frequent vehicle-following interactions.

Vehicle-following interaction is reasonably frequent on two-lane roads, which aggravates further if the prevailing traffic is mixed and composed of various vehicle types in terms of their static and dynamic characteristics. Since both directions of traffic use a single carriageway on such roads, a range of vehicles, including non-motorized ones, use the same road space and also, the flow of traffic in one direction affects flow in the other direction. Faster vehicles overtake slower ones using the adjacent opposing lane through the gap acceptance process. Hence, under such traffic operations, headway events range from shorter during platoon movements to longer when vehicles move freely, and their speed is not restricted. Headways are sometimes even close to zero values when drivers shift on the adjacent lane during overtaking operation [2]. Further, there are instances when a driver finds it tricky to evaluate the acceptable gap under such traffic and hesitates to perform overtaking operations even if the overtaking vehicle

✉ Pritam Saha
pritam@civil.iiests.ac.in

Rupali Roy
rupaliroy.nit@gmail.com

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

is on the adjacent lane. As a result, an oscillating movement between initial and adjacent lanes occurs, resulting in zero values of headways.

Statistically, the exponential model can describe the headway data if the co-efficient of variation is one [2, 3]. However, based on empirical investigations, many researchers reported that the heterogeneity effect of traffic stream results in a deviation of co-efficient of variation; this makes the exponential model inappropriate in describing headways. The hypo-exponential model is suitable when the coefficient of variation is less than one; whereas, a hyper-exponential model is considered statistically valid when it exceeds unity. Erlang distribution function exhibits similar applicability like hypo-exponential distribution in a stochastic process. Though a single distribution model is simple and easy to apply at a low flow level, the performance of such model forms is unsatisfactory, particularly at moderate and heavy flow because of their limited capabilities in approximating smaller headways [4]. The study describes the concepts of using shifted distributions while improving the accuracy of single distribution models when the coefficient of variation is less than one. The study also proposes the combined distribution function as an alternative of mixture distribution for describing headways, respectively, at moderate and heavy flow. Combined distribution contemplates modeling of shorter and longer headways separately based on a simplistic approach. For grouping the data into shorter and longer headways, the current study used partitional clustering and found K-means algorithm with Euclidean distance measure to exhibit aptness in this regard. The appropriate models for headway data were selected using a method based on the K-S test and the examination of probability plots.

In India, interurban highways are, in general, two/ multi-lanes. Moreover, most of the country's national and state highways and a significant portion of major district roads are of two-lane status. Thus, there is a pressing need to develop capacity standards for such highways, especially when the prevailing traffic is mixed and has a wide variety of vehicles in its composition. Since vehicle-following behaviour varies considerably across vehicle types, most distribution models proposed so far have been reported to exhibit incompatibility in such a mixed-mode movement of traffic, especially on two-lane roads where interaction occurs in both directions. Therefore, the present study aimed to develop appropriate headway distribution models based on microscopic analysis of the time headway of vehicles considering such heterogeneity in the traffic mix. In doing so, it intended to investigate identifying following and free vehicles in the mixed traffic stream and, thereby, developing an appropriate approach to modelling time headways considering following and free vehicles. Eventually, the study developed a comprehensive understanding of distributions of headways which is fundamental in traffic flow study and its' simulation issues.

Research Background

Many studies reported investigations on vehicle time headways and proposed several distribution models to describe headways. A variety of factors, including traffic flow [5], composition [2], and roadway condition affect distribution of headways and put restrictions on the use of those models. In an attempt to identify the boundaries in the traffic flows where the distribution of headways changes, a study found that at low flow, when interactions among the vehicles are insignificant negative exponential distribution fits the observed headways aptly. However, at medium and heavy flow with a frequent car-following interaction, Shifted Exponential/Gamma and Erlang distribution provided an excellent fit to the empirical distributions [3].

There have been several studies that reported a range of compatible distributions like generalized extreme value [6], hyperlang [7] and negative exponential [2] for urban roads. Interestingly a study found the distribution of headways for urban roads different even across lanes. It indicated Lognormal and Gamma distributions with a shift of 0.24 s and 0.69 s appropriate in passing and middle lanes, respectively [8]. Notably, the applications of Lognormal and Gamma distributions for headways in urban roads are widespread. For instance, modeling preferred time headway and drivers' time headway [9] and driver's desired time gap and time headway in a steady car-following state [10].

A handful of studies indicated mixed distributions like gamma-generalized queuing model (gamma-GQM) and double displaced negative exponential distribution (DDNED) for modeling headways. Gamma-GQM exhibits its aptness for a wide range of time headways [11], considering traffic flow and composition as endogenous and weather and geometric features as exogenous factors [12]. A study aimed at assessing the performance of various headway models using freeways data observed DDNED to provide a decent fit to data points. However, the study found shifted lognormal distribution to fit such data well for other types of roads [4]. Some recent studies reported Log-Pearson 3, Lognormal, and Log-logistic, Pearson 5 and Pearson 6 distributions suitable for a range of flow levels by examining headway data of two-lane roads with mixed traffic conditions [13, 14].

A spectrum of studies focused on analyzing vehicle-type-specific headways for a variety of roads and found that more often than not, headways are largest for truck-truck combinations and smallest in the case of car-car. Hence, the distribution of headways varies with the headway type. For instance, Erlang, lognormal, and inverse Gaussian models are appropriate for car-truck headways, composite and lognormal for car-car headways, shifted negative exponential, inverse Gaussian and lognormal models for truck-car and truck-truck headways [1, 15, 16]. Based on traffic data of two-lane highways, a study indicated Pearson-III to be suitable for a realistic replication of headways over a range of vehicle combinations and time of day [17]. However, empirical distributions of vehicle-type-specific headways for such highways vary considerably in the event of mixed traffic due to the risk-taking behaviour of the driver population. A couple of studies in this context highlighted this fact and proposed predictive models for understanding car-following behaviour [13, 18]. Another recent study in context to such traffic investigated the applicability of the copula approach for different vehicle-pair combinations and found Frank copula suitable for almost all leader-follower vehicle pairs [19].

While several studies have addressed the distribution of headways, a handful attempted to describe their characteristics under mixed traffic situations. Recent experiences with such traffic revealed that headways of lead-lag vehicles largely depend on the length of the lead vehicle [20], and the interactions cease beyond six seconds headway [21] at vehicle-following situations. Since the desired headway also varies across drivers [22], the current study aimed to systematically investigate time headways of vehicles under mixed traffic and propose a representative distribution of such headways.

Study Motivations

Mixed traffic with a range of vehicle types that include two/three-wheelers and non-motorized modes like paddle tricycle, etc. is symptomatically different from the traffic that exhibits more or less homogeneous vehicle types. More often than not, mixed traffic, as considered in the current study, exhibits vehicles that are considerably different in their statics and dynamics. They do not follow lane discipline while sharing the same road space and occupy any available lateral position. At times, smaller vehicles occupy the gaps between larger vehicles in the traffic stream, resulting in a significant increase in shorter headways even at moderate and low flow. Thus, headway consists of 'following' and 'free' components, and a composite distribution can model them for such traffic. For years, several studies derived quite a few important composite models according to this concept; however, calibration and estimation of parameters were not easy for those distributions due to the inherent complexities in analyzing such traffic and the complicated structure of those model forms. Moreover, vehicles' arrival patterns change noticeably with traffic flow, resulting in different distributions to fit better at different flow levels. Thus, it is critical to discern the precise distribution of time headways for a road with a given flow condition.

Data and Methods

Field Study

Design of field study aimed at capturing time headways, covering a range of flow levels. Eight study sites (road segments) were selected for collecting traffic data on a stretch of a national highway (NH-8; two-lane highway) in India. The road stretch provides access to a city at its western end, and the study section extends about twenty kilometres towards the east of the city limit. Therefore, the paper describes traffic heading towards the urban area as westbound and opposing traffic as eastbound. Since the effect of intersections, curvatures, and ribbon developments were outside the scope of this paper; the current study made efforts to select road segments so that they are free from these effects. All the chosen segments have a reasonably long straight stretch with good and uniform pavement conditions (see Fig. 1). Field investigations also included a pilot study for validation and selected two highway segments (SH-13 and NH-11) that exhibit more or less similar roadway and traffic characteristics.

The current study used video cameras to capture traffic data to achieve adequate precision and accuracy in the dataset. The site layout plan for data collection includes a reference line across the pavement and two camera installation points on either side of it to capture both directional movements. The study was conducted during daylight hours and continued for several days covering peak and off-peak hours (each for about two hours) to capture an extensive range of traffic flow levels. For video data extraction, the study utilized computers with a large display screen. It noted the necessary readings, i.e., vehicle type, axle configurations and time when a vehicle crosses the reference line. Then the study computed the time headway of vehicles using the extracted traffic data. Since, it is necessary to determine the class interval for the observed data to obtain a smooth histogram, the study applied Struge's rule [23, 24] for such assessment.

Figure 2a–b shows the composition of traffic consisting of two/ three-wheelers, bus, truck, car, and non-motorized ones, including paddle tri-cycle. A look into the figure reveals that the share of two-wheeler and car is significant, ranging from about 20–35 and 20–40 percent, respectively. Also, city-bound traffic consists of a substantial proportion of non-motorized vehicles; in a way that they share around 20 percent of total traffic. The study attempted to be acquainted with the reason behind such paradox through an informal opinion poll. Responses from about 50 road users indicate that lack of adequate public transport facilities compel commuters to use para-transit modes like paddle tri-cycle or personalized transportation like cars, motorized two-wheelers, etc.

Fig. 1 A view of the study section (image by author)

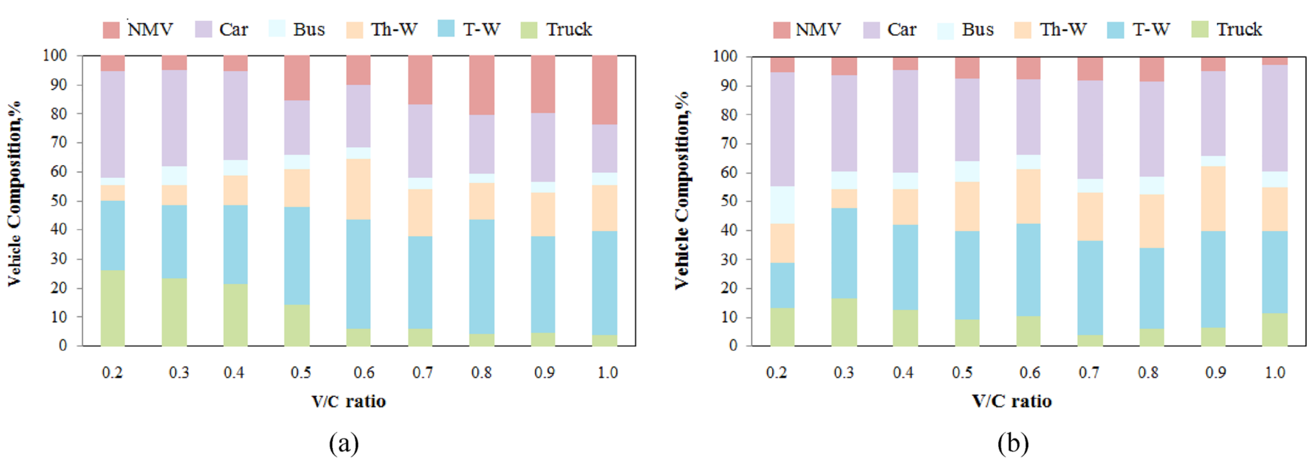

Fig. 2 Vehicle composition at the study sites: **a** westbound and **b** eastbound traffic [*NMV* non-motorized vehicle, *Th-W* three-wheeler, *T-W* two-wheeler, *v/c* ratio–volume to capacity ratio]

Further, traffic flow levels were grouped into a low, moderate and heavy flow based on traffic platooning and expressed in v/c ratio: derived from a capacity value of 2300 pc/h [25]. As moderate flow relates to stable flow, it would be rational to assume that the average number of headways inside and outside platoons under such flow is approximately equal. The current study found that the flow levels corresponding to a v/c ratio of 0.5–0.7 represent the moderate flow. Accordingly, flow levels at v/c ratio of 0.4 or less and 0.8 or more were, respectively, considered as low and heavy flow (see Table 1). Table 1 provides the sample size and also, details of headway frequencies at different flow levels.

Modeling Techniques

A suitable headway distribution model development calls for an approach to examine the statistical models with the observed headways (see Fig. 3). The use of negative exponential distribution to describe headways is quite widespread. However, several researchers have proposed using alternate models to illustrate the headway distribution pattern explicitly. Notably, these models are Erlang and Gamma distributions that provide a decent fit for an extensive range of flows and lognormal distribution, which has a theoretical connection to the car-following models [3]. The flexibility and compatibility of Gamma distribution make it suitable for

Table 1 Details of traffic condition and observed headway data

Traffic condition	V/C ratio	Sample size (number of headways)	Frequency of headways (h) less than 't' sec											
			1.5	4.5	7.5	10.5	13.5	16.5	19.5	22.5	25.5	28.5	31.5	34.5
Low flow: interaction between vehicles is very less	0.2	160	30	35	25	20	13	11	8	7	4	3	4	–
	0.3	875	259	127	116	85	69	44	42	40	26	30	20	17
	0.4	1082	390	197	149	82	76	50	36	40	23	22	8	9
Moderate flow: avg. number of headways inside and outside platoons is equal	0.5	1388	632	246	157	92	66	60	41	28	28	10	18	10
	0.6	760	374	146	89	54	30	23	23	9	7	5	–	–
	0.7	1001	561	173	107	53	39	35	16	10	5	2	–	–
Heavy flow: congested traffic involving slowing and stopping	0.8	970	580	159	90	58	35	15	22	5	4	2	–	–
	0.9	634	408	105	54	34	9	8	6	7	2	1	–	–
	1.0	464	309	83	37	22	6	4	2	0	1	0	–	–

extensive use in modelling headway data [4]. By the same token, Erlang distribution also has wide applicability owing to its relation to the exponential and Gamma distributions.

The exponential distribution function is considered statistically valid if the headway data's co-efficient variation equals one [2, 3]. The hypo-exponential model is considered suitable for distributions of such data with a coefficient of variation of less than one. A hyper-exponential model is statistically valid for those with a coefficient of variation of more than one [26]. Typically, Erlang distribution represents a hypo-exponential distribution function [26], and hyper-exponential distribution can be well represented by a mixed distribution function [27, 28]. As the study schemed to develop headway distribution models, data corresponding to an extensive range of traffic flow were fitted to different distributions that meet the required traits (see Fig. 4). The following sections provide a detailed insight into it.

An Application of Shifted Headway Distribution Model

Hypo-exponential distribution appears to be appropriate if the observed data's coefficient of variation is less than one. Erlang distribution is found to have similar applicability to hypo-exponential distribution in the stochastic process [26]. Applying single distribution models is straightforward; however, they have limited capabilities in approximating shorter headways. Therefore, the concept of shifted single distribution models is applied to improve the accuracy of single distribution models [4]. Equations 1 and 2 demonstrate the cumulative density functions of shifted negative exponential and shifted Erlang distributions, indicating a τ seconds shift to the right. Accordingly, investigations were made with these functions with a range of shifts (0 to 3 s), keeping a step of 0.015 s to determine a shift against which the function gives the best result.

Shifted negative exponential distribution function is given by,

$$F(h) = 1 - \exp(-\lambda(h - \tau)) \quad (1)$$

Shifted Erlang distribution function is given by,

$$F(h) = \left(\frac{\Gamma(h-\tau)/_{\beta}(m)}{\Gamma(m)}\right) \quad (2)$$

where: h = headways, λ = continuous inverse scale parameter ($\lambda > 0$), m = shape parameter (positive integer), β = continuous scale parameter (Erlang distribution) ($\beta > 0$).

An Approach of Combined Headway Distribution Model

Overtaking opportunities and proportion of vehicles spaced widely start declining on two-lane roads when flow rate upturns. The situation further aggravates in the event of

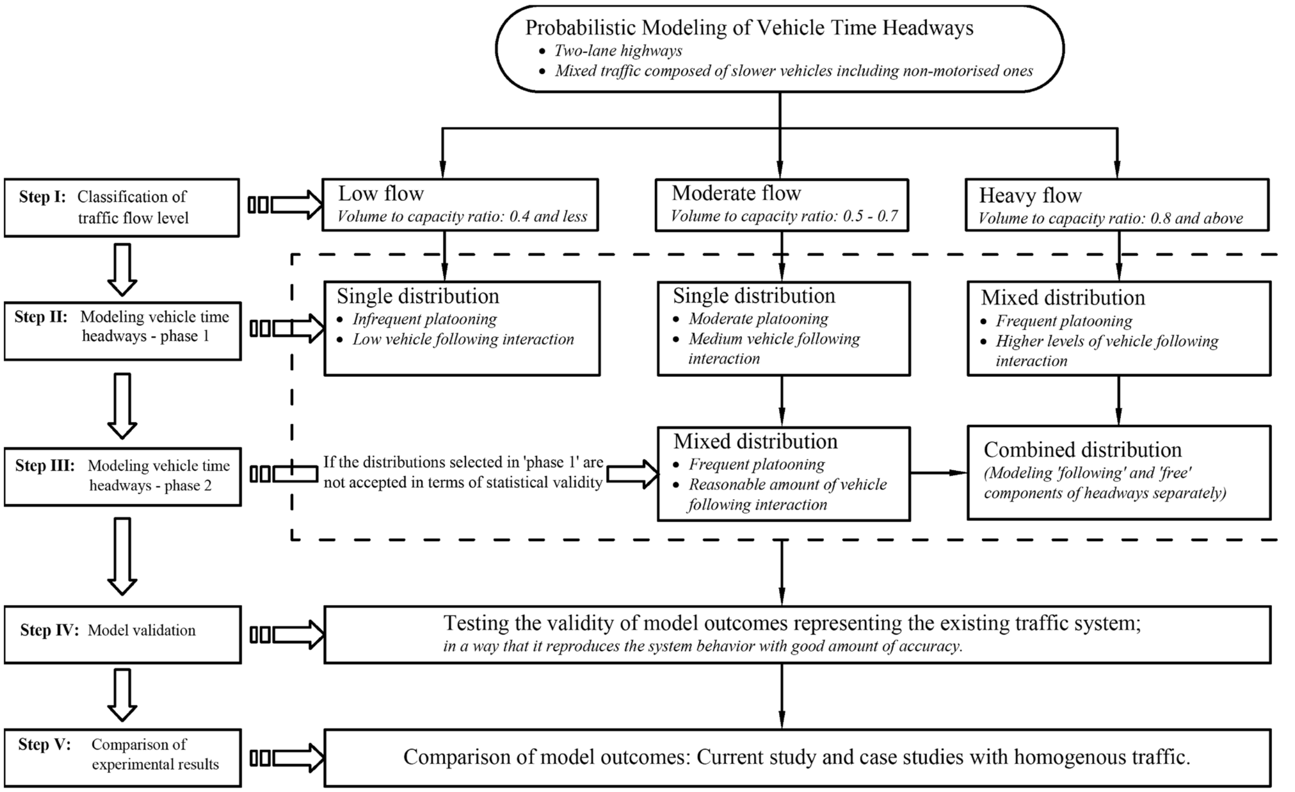

Fig. 3 A schematic representation of the proposed framework for probabilistic modelling of time headway of vehicles with varying sizes and speeds on two-lane highways

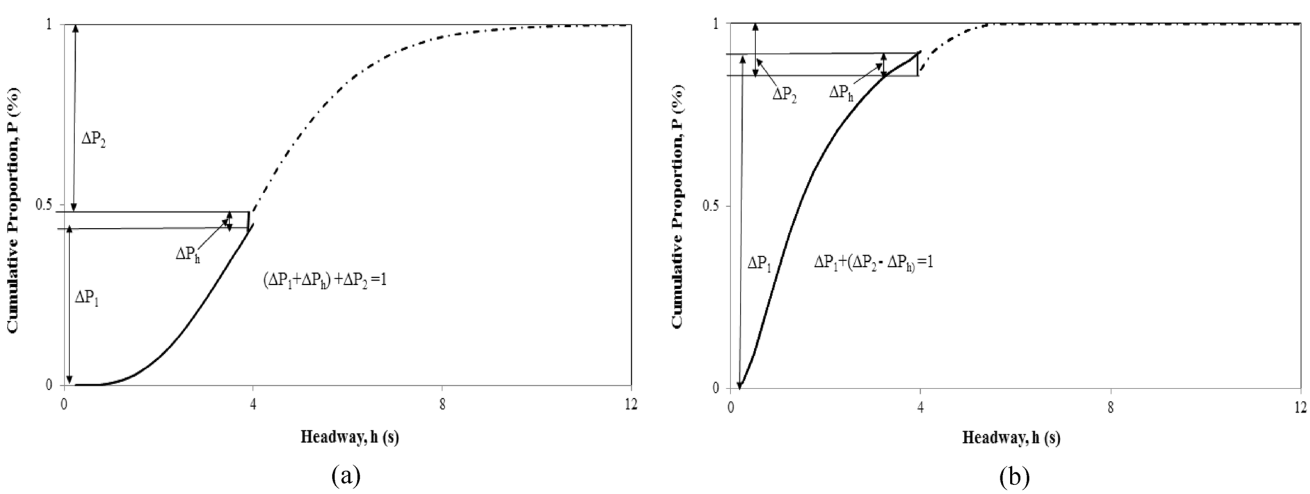

Fig. 4 A schematic representation of the combining headway distributions by adjusting the predicted probabilities at the estimated headway threshold ‘h’: **a** Positive adjustment and **b** Negative adjustment

heterogeneity and a sizable proportion of slower vehicles in the traffic mix, resulting in a significant amount of shorter headways, particularly at moderate and heavy flow. There have been several researchers who proposed the concept of mixed distribution models for describing headways in such situations: these are ‘double displaced negative exponential distribution’ [29], ‘generalized queuing model’[30], ‘Cowan M3 and M4 model’ [4]. Equations 3–5 demonstrate these mixed distribution functions:

Double Displaced Negative Exponential Distribution (DDNED) function is given by,

$$f(h) = \begin{cases} \phi * \Upsilon_1 * \exp\left(-\Upsilon_1 * (h-d)\right) + (1-\Phi) * \Upsilon_2 * \exp\left(-\Upsilon_2 * (h-d)\right) & h \geq d \\ 0 & h < d \end{cases} \quad (3)$$

Generalized Queuing Model function is given by,

$$f(h) = \theta * g(h) + (1-\theta) * \Upsilon * \exp(-\Upsilon * h) * \int_0^h g(x) * \exp(\Upsilon * x)dx \quad (4)$$

Cowan M3 distribution function is given by,

$$f(h) = (1-\Phi)\delta(h-\tau) + \Phi\lambda e^{-\lambda(h-\tau)}u(h-\tau) \quad (5)$$

where: *f (h)* is the probability density function; ϕ is a weighting factor constrained by $0 < \phi \leq 1$; Υ_1 and Υ_2 are constants associated with the flow status, and *d* is a displaced parameter; θ = proportion of vehicles tracking at the minimum headway; $= \delta(h-\tau)$unit impulse (Dirac delta) function; x = proportion of the following headway; γ = exponential decay function.

The concept's premise is that headway comprises 'following' and 'free' components. Accordingly, mixed headway distribution models are supposed to capture headway dynamics better. However, calibration and parameter estimation of such distribution models are tricky because of their complicated structure [4]. This is particularly true for mixed traffic situations where prevailing vehicles exhibit a wide variety of static and dynamic characteristics.

Accordingly, the combined distribution function was considered an alternative of mixture distribution while describing moderate and heavy flow headways to overcome this difficulty. The simplistic analytical approach considers modelling shorter (lower fraction) and longer (upper fraction) headways separately using appropriate distribution functions. Figure 4 demonstrates the approach of combining headway distributions by adjusting the predicted probabilities at the estimated headway threshold 'h'. An adjustment ΔP_h was made on the proportional contributions of the lower and upper fractions, i.e., ΔP_1 and ΔP_2 to ensure that the total probability is one. Partitional clustering technique exhibits its aptness in identifying the headway threshold. Lognormal and Gamma distributions seem appropriate at car-following situations because of their flexibility and compatibility [4], and the negative exponential model describes the widely spaced data well. Therefore, Lognormal and Gamma distributions were applied on headway data for describing shorter headways, and they all were applied for longer headways. Thus, the combined distribution represents 'following' and 'free' components of headways separately based on the estimated headway threshold 'h' as displayed in Eq. 6–9.

Combined lognormal and Gamma distribution function is given by,

$$F(h) = \int_o^h \Phi\left(\frac{\ln h - \mu}{\sigma}\right)dh + \int_h^\alpha \frac{\Gamma h/_\beta(\alpha)}{\Gamma(\alpha)}dh\Delta P_h \quad (6)$$

ΔP_h= Difference in probability between lognormal and Gamma distribution.

Combined Gamma and lognormal distribution function is given by,

$$F(h) = \int_0^h \frac{\Gamma h/_\beta(\alpha)}{\Gamma(\alpha)}dh + \int_h^\infty \phi\left(\frac{\ln h - \mu}{\sigma}\right)dh \pm \Delta P_h \quad (7)$$

ΔP_h= Difference in probability between Gamma and lognormal distribution.

Combined Gamma and exponential distribution function is given by,

$$F(h) = \int_0^h \frac{\Gamma h/_\beta(\alpha)}{\Gamma(\alpha)}dh + \int_h^\alpha (1 - exp(-\lambda h))dh \pm \Delta P_h \quad (8)$$

ΔP_h= Difference in probability between Gamma and exponential distribution.

Combined lognormal and exponential distribution function is given by,

$$F(h) = \int_0^h \phi\left(\frac{\ln h - \mu}{\sigma}\right)dh + \int_h^\alpha (1 - exp(-\lambda h))dh \pm \Delta P_h \quad (9)$$

ΔP_h= Difference in probability between lognormal and exponential distribution. Where: μ = continuous location parameter, σ = continuous scale parameter (Lognormal distribution), α = continuous shape parameter $(\alpha > 0)$.

Evaluation of Goodness-of-Fit

Selection of appropriate models that exhibit compatibility to the observed data entails an appropriate method based on goodness-of-fit (GOF) tests. The use of Chi-square and Kolmogorov–Smirnov (K–S) tests for extracting GOF test statistics are quite common in traffic analysis. However, the K–S test provides specific benefits over the chi-square test, like (a) it can use data with a continuous distribution and (b) there is no minimum frequency per test interval [1]. Accordingly, the paper applied the K–S test on the way to identify compatible models. The GOF test statistic, 'D' is the discrepancy between the cumulative percentage of the measured and expected frequency: the largest discrepancy over the entire measurement indicates the value [21, 31]. Also, quantile–quantile plots help visualize the goodness-of-fit graphically by comparing their observed and estimated quantiles.

Analysis and Results

Descriptive Statistics of Headway Data

The possible capacity of the highway section was observed to be around 2300 pc/h [25]. Headway data were collected considering directional segments separately. A wide range of traffic volume (expressed in terms of pc/h; calculated based on passenger car equivalents as suggested by IRC 64: 1990 [32]) corresponding to volume to capacity ratio of 0.2–1.0 was covered. Table 2 displays the descriptive statistics of the headway data. A careful examination of Table 2 indicates that the median is less than the mean and the standard deviation decreases with the flow rate at all flow scopes. It signifies the concentration of shorter headways and is attributed to the high risk-ability of driver populations, which eventually lessens traffic safety. Besides, the mean and standard deviation should result in a 45° plot, i.e., coefficient of variation should be one in case of a negative exponential distribution. However, an empirical investigation indicates a deviation, exhibiting the negative exponential model's inappropriateness in describing headways. Moreover, the coefficient of variation was less than one upto a flow level that corresponds to a volume to capacity ratio of 0.3 for both the directions and exceeds the value at low to moderate, moderate and heavy flow (volume to the capacity ratio: 0.4–1.0).

Headway Distribution at Low Flow Level

Erlang and negative exponential distributions with shifts ranging from 0 to 3 s (with a step of 0.015 s) were used in modeling headways at low flow. A comparison of the K–S test statistic obtained with 5 percent level-of-significance for the selected models, was made to examine the extent of fit to the observed data. The test statistic decreased with the increase in shifts, which started increasing beyond an optimal value. Accordingly, the trend discerns an optimal shift for the fitted models; Erlang distribution model with shifts of 0.675, 1.41 and 2.76 s and exponential distribution with 0.495 s shift were found to have acceptable statistical validity in terms of K–S test statistic. Table 3 provides the details of goodness-of-fit and estimated parameters.

Figure 5a displays the cumulative density function of the selected models. It indicates that the proportion of shorter headways at a volume to capacity ratio of 0.2 is less for city-bound (westbound) traffic, increasing with the flow. Eastbound traffic points to a diametric characteristic wherein

Table 2 Descriptive statistics of headways at different flow levels

V/C ratio	Westbound traffic			Eastbound traffic		
	Mean	Median	Standard deviation	Mean	Median	Standard deviation
0.2	14.87	10.50	12.65	10.03	4.50	9.22
0.3	11.57	7.50	10.64	10.10	7.50	9.18
0.4	8.12	4.50	8.52	8.75	4.50	9.02
0.5	6.70	4.50	6.81	7.27	4.50	8.22
0.6	5.70	4.50	6.11	6.00	1.50	7.05
0.7	4.86	1.50	5.54	4.69	1.50	5.05
0.8	4.49	1.50	4.92	4.50	1.50	5.17
0.9	3.97	1.50	4.21	3.60	1.50	5.12
1.0	3.66	1.50	4.10	3.55	1.50	4.78

Table 3 Goodness-of-fit Details and Estimated Parameters of the Fitted Distributions

V/C ratio	K-S test (D-value)		Shift value (Second)		Fitted distribution	Estimated parameters
	Exponential	Erlang	Exponential	Erlang		
0.2*	**0.0942**	0.1907	**0.4950**	2.820	Shifted Exponential	$\lambda = 0.0673$
0.3*	0.1328	**0.1312**	2.595	**1.410**	Shifted Erlang	$m = 1; \beta = 9.8681$
0.2$^{@}$	0.1579	**0.1507**	2.955	**0.675**	Shifted Erlang	$m = 1; \beta = 8.5335$
0.3$^{@}$	0.1735	**0.1624**	2.790	**2.760**	Shifted Erlang	$m = 1; \beta = 9.5023$

*West bound traffic; $^{@}$East bound traffic; λ: continuous inverse scale parameter; m: shape parameter (positive integer); β: continuous scale parameter [Bold values indicate smallest D-value]

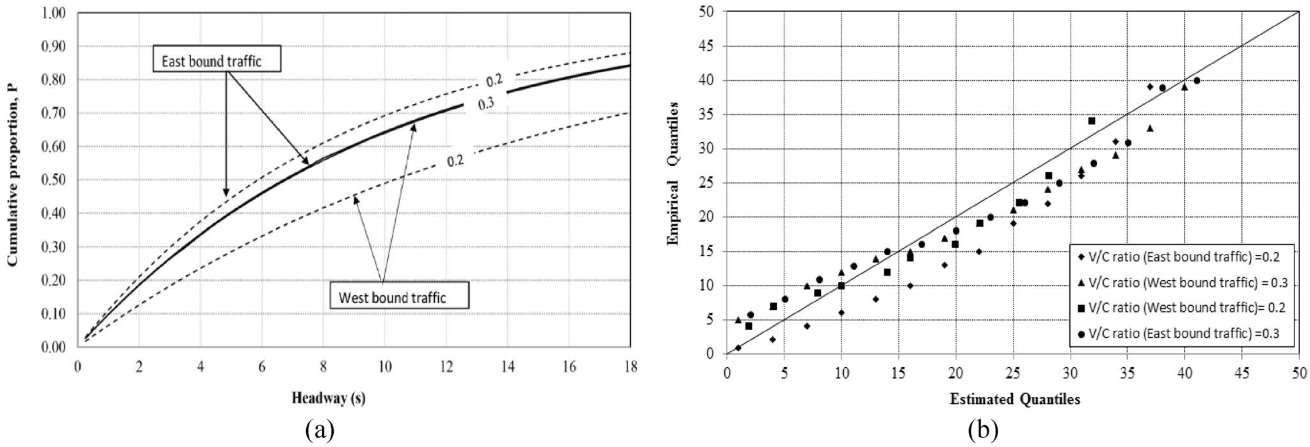

Fig. 5 Shifted headway distributions at low flow: **a** Cumulative density functions **b** Q–Q plot

shorter headways were significant at such flow level. This is because city-bound traffic was quite less during the morning, and it starts increasing as the day progresses. In contrast, trips to the city outskirts were sizable during the morning, and, thenceforth, it decreases. Figure 5b indicates that the data points are clustered in about a 45° plot signifying satisfactory agreement between the theoretical and observed data distribution.

Headway Distribution at Moderate and Heavy Flow

The proportion of shorter headways was observed to be considerably high at the moderate and high flow levels. It was, therefore, essential to describe shorter and longer headways separately. Accordingly, the field data, collected at different flow scopes, was grouped into clusters or subsets such that the data in each subset are similar to each other [33]. The data set similarity is defined by a distance measure; this plays a vital role in obtaining correct clusters [33]. However, the selection of distance measures depends on the data type. The present study's observed data set is of interval type as it contains a range of continuous values [34]. Therefore, the Euclidean distance measure was adopted in the present study as it corresponds to the interval data type. The number of clusters was selected, given shorter and longer headways, and then the partitioning algorithm was used for identifying the boundaries. The mean of each cluster was used to determine its centroid. Accordingly, the k-means algorithm was applied, adopting Euclidean distance measure as an effective heuristic method of partitional clustering [35]. The algorithm appeared efficacious since it provides a faster computational platform and helps interpret findings through easy visualization of clusters. The study performed 100 iterations with a random set of initial clusters, calculating the distance between data points and the cluster centre while arriving at the optimized cluster. The limiting value of shorter headways was found to be 4 s; a similar value was also obtained by a study conducted in India [36].

The study considered lognormal and gamma distribution models to describe the shorter headways and lognormal, Gamma and negative exponential distribution models for the larger headways. Table 4 elucidates the goodness-of-fit details of the selected models and the calibrated model parameters. Since the limiting value of shorter headway was 4 s, shorter headways were modelled considering headway data up to 4 s, and the remaining headways were modelled as longer headways separately. The selection of best-fitted models used a method based on K–S test results (see Table 4).

Selected best-fitted distributions for shorter (following component) and longer (free component) headways were accordingly combined (see Eq. 6–9) and applied for describing the data set. Figure 6 illustrates the cumulative density function of combined models and indicates that the shorter headway's probability is relatively higher than longer headways. This attributes to the fact that platoon formation is frequent and car following interaction is quite high at moderate and high flow. Figure 7 displays the quantile plots of the combined models and indicates that the data points are close to the straight line signifying satisfactory agreement between the theoretical model and the distribution of the observed data.

Validation of the Proposed Model

The current study further attempted to test the validity of the model outcomes to ensure that they represent the existing traffic system with a good amount of accuracy. Two different highway sections that exhibit similar roadway and traffic characteristics were selected for the pilot study, and traffic data were collected, respectively, at the low and moderate flow levels. A comparison between the outcomes of the shifted Erlang/ shifted negative exponential models and empirical headways obtained at low flow indicated the

Table 4 Goodness-of-fit details and estimated parameters of the selected distributions

	V/C ratio	Shorter headways		Longer headways			Estimated parameters	
		Lognormal	Gamma	Lognormal	Gamma	Exponential	Shorter headways	Longer headways
West bound traffic	0.4	0.19345	**0.17950**	0.1428	**0.1383**	0.4540	$\alpha=2.6448$; $\beta=0.8611$	$\alpha=7.4232$; $\beta=0.3789$
	0.5	**0.23205**	0.24084	**0.1172**	0.1187	0.4556	$\mu=0.3695$; $\sigma=0.8114$	$\mu=0.9602$; $\sigma=0.4194$
	0.6	**0.20605**	0.20675	**0.1427**	0.1432	0.4577	$\mu=0.4549$; $\sigma=0.7817$	$\mu=0.5423$; $\sigma=0.8168$
	0.7	0.26508	**0.25685**	0.1632	**0.1622**	0.4667	$\alpha=4.8001$; $\beta=0$ .9544	$\alpha=6.026$; $\beta=0.7138$
	0.8	0.27781	**0.26696**	0.1327	**0.1289**	0.4666	$\alpha=1.7022$; $\beta=0.9918$	$\alpha=6.832$; $\beta=0.3789$
	0.9	0.26713	**0.26529**	**0.1443**	0.1475	0.4749	$\alpha=4.8351$; $\beta=0.7539$	$\mu=1.0985$; $\sigma=0.8996$
	1.0	0.27597	**0.25811**	**0.1730**	0.1751	0.4803	$\alpha=1.6505$; $\beta=1.0135$	$\mu=1.1174$;$\sigma=0.1985$
East bound traffic	0.4	**0.21253**	0.21348	0.1422	**0.1407**	0.4462	$\mu=0.4026$; $\sigma=0.7822$	$\alpha=6..531$; β=0.4603
	0.5	**0.23211**	0.23748	**0.1481**	0.1482	0.4721	$\mu=0.3553$; $\sigma=0.8130$	$\mu=0.9055$; $\sigma=0.4837$
	0.6	**0.20523**	0.22213	0.2497	**0.2256**	0.5080	$\mu=0.5018$; $\sigma=0.8401$	$\alpha=7.0215$; $\beta=0.4232$
	0.7	0.42970	**0.40975**	**0.1388**	0.1411	0.4746	$\alpha=4.9778$; $\beta=0.8171$	$\mu=1.2035$; $\sigma=0.8761$
	0.8	0.27278	**0.25856**	**0.1340**	0.1366	0.4550	$\alpha=0.2585$; $\beta=0.7893$	$\mu=0.8653$; $\sigma=0.3621$
	0.9	0.28101	**0.26112**	0.2217	**0.2192**	0.4597	$\alpha=1.8156$; $\beta=0.8442$	$\alpha=8.6817$; $\beta=0.2419$
	1.0	0.26135	**0.23247**	0.1956	**0.1472**	0.4632	$\alpha=1.2463$; $\beta=0.6542$	$\alpha=6.6846$; $\beta=0.3683$

[Bold values indicate smallest D-value]

variability of predictions. The study expressed them in terms of the standard error of the estimate (SEE) and found the values ranging between 0.04 and 0.06. (see Fig. 8a). The empirical probability of headways measured at moderate flow was compared with the outcomes of combined distribution models, and the values were found to be ranging between 0.02 and 0.06 (see Fig. 8b). These indicate a good agreement between expected and empirical probabilities, and the slight deviation is possibly due to variations in traffic composition and driver's behaviour across study sites.

Discussions

The current study investigates headway distribution models when the prevailing traffic is mixed. Based on field measurements, the study attempted to characterize vehicle headways using distribution functions statistically. The investigations indicated shifted Erlang and shifted negative exponential distributions appropriate at low flow and 'combined distribution' at moderate and heavy traffic flow. Most drivers move with widely spaced headways at low flow when they have enough freedom to manoeuvre. Frequent interactions at moderate and heavy flow result in the formation of platoons. Vehicles entrapped inside platoons start moving in following with shorter headways and overtake through the gap acceptance process. Table 5 demonstrates a steady increase in 'probability of headway less than 't' sec' at moderate flow (upto a v/c ratio of 0.4–0.6), which, however, starts decreasing at the higher flow level. A further investigation clarifies this fact and indicates that most faster vehicles initiate overtaking at a moderate flow level and move with shorter headways in such operations. However, limited overtaking opportunities at heavy flow restrict them from taking such initiatives and accordingly, they start moving in following keeping adequate spacing.

The current study compared the outcomes of combined distribution models with those obtained from single distribution models [21]. It was found that a single distribution model results in a lower probability at moderate flow (v/c ratio: 0.4–0.6) than the combined one. It is, however, quite the opposite in the case of heavy flow when a single distribution model results in a higher probability (see Table 5). The current study, thus, makes it evident that a combined distribution model exhibits its aptness in approximating headway data at moderate and heavy flow under mixed traffic.

Conclusion, Limitation and Scope for Further Study

The paper presents an investigation on time headways of two-lane highways under mixed traffic situations. The study encompassed extensive traffic flow levels ranging from low to heavy for analyzing the driver's behaviour in choosing headways. Observations indicate that a part of the driver population follows impeding vehicles keeping unsafe headways at the car-following situation, i.e., less than 2 s [37]. Statistical parameters were computed, and the mean was observed to be larger than the median at all flow levels, signifying the concentration of shorter headways; this attributes to the high risk-taking behaviour of the driver population, eventually resulting in safety reduction [38].

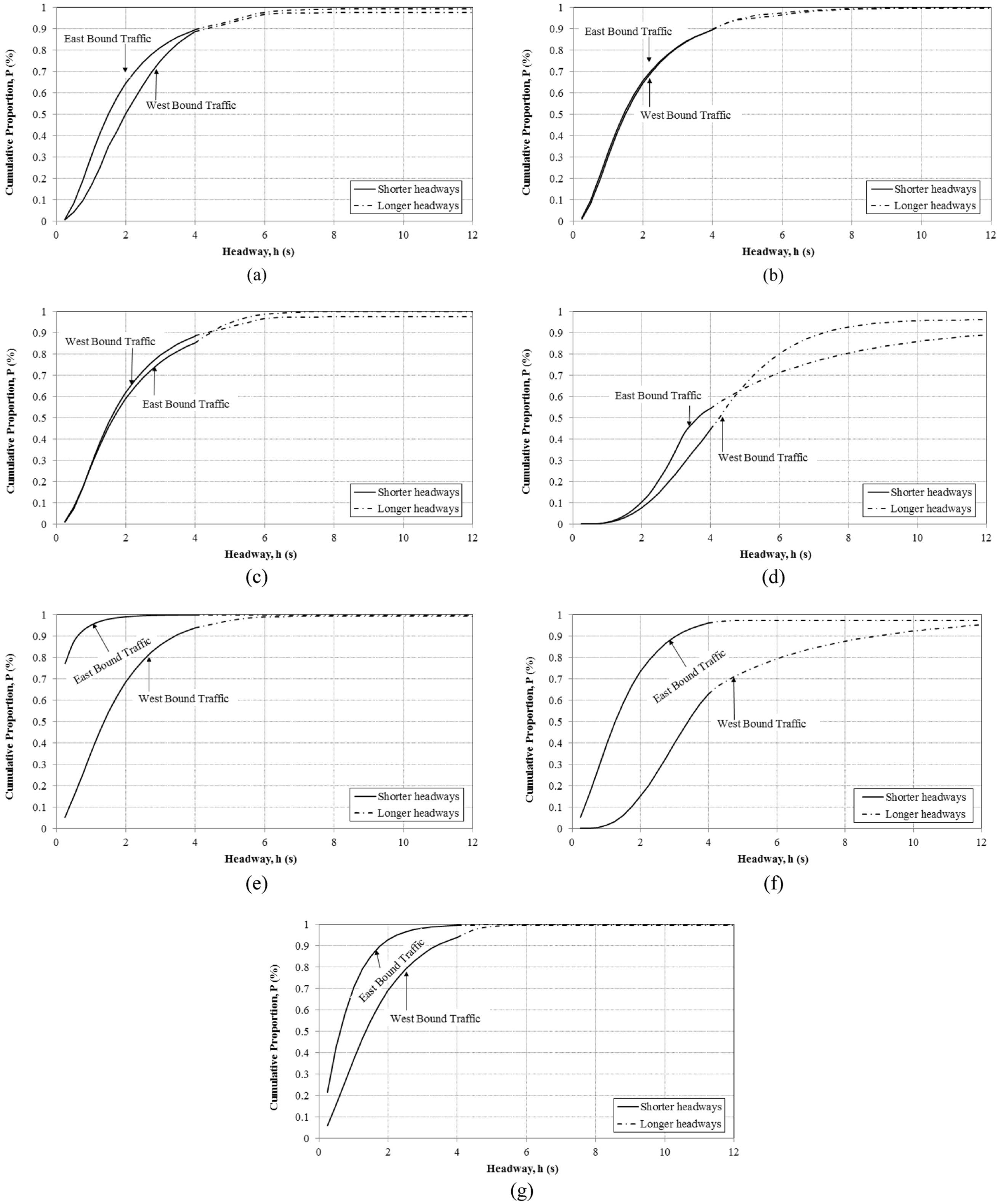

Fig. 6 Cumulative distributions of combined models at different volume-to-capacity ratio: **a** 0.4; **b** 0.5; **c** 0.6; **d** 0.7; **e** 0.8; **f** 0.9; **g** 1.0

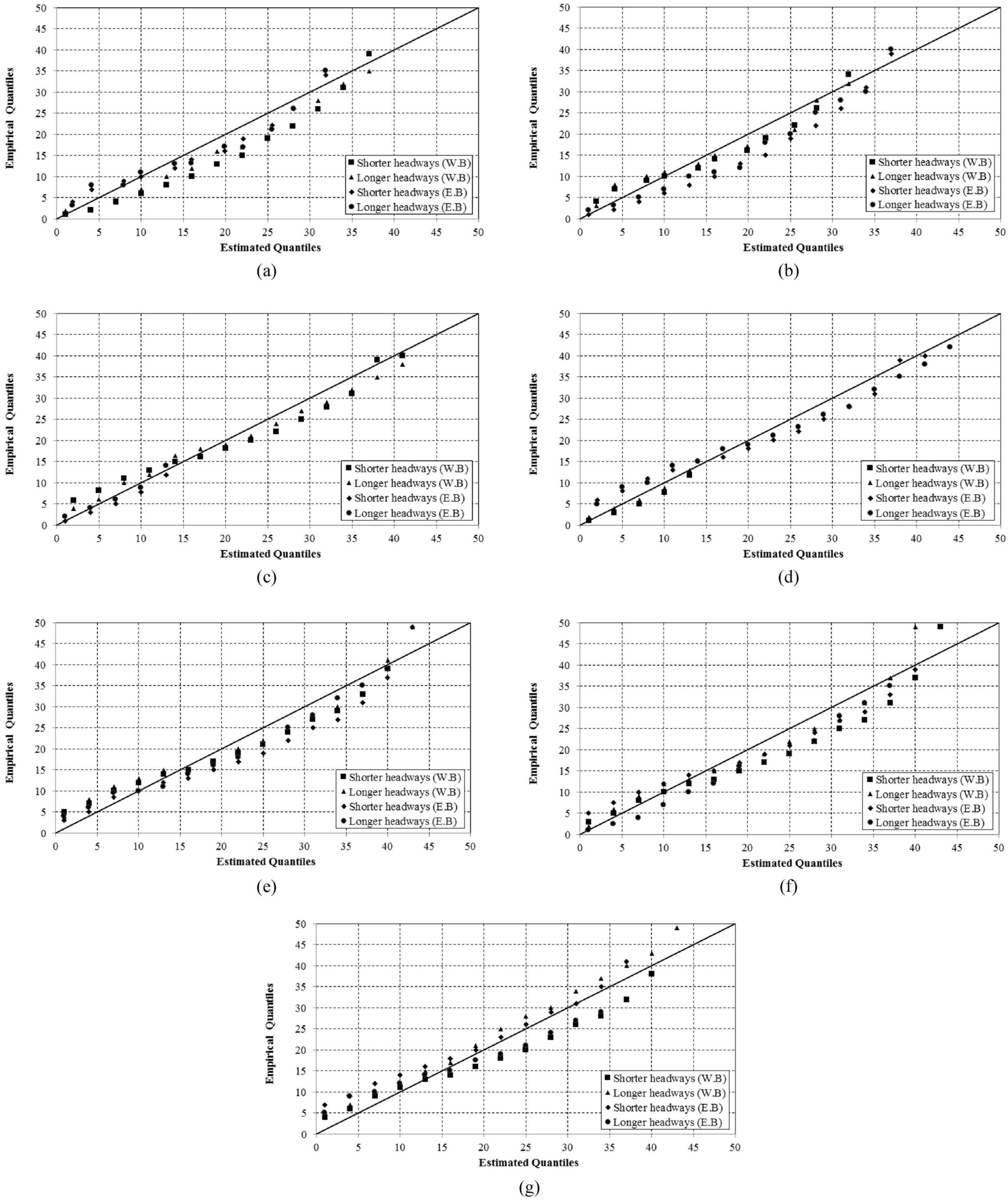

Fig. 7 Quantile–Quantile plots of combined models at different volume-to-capacity ratio: **a** 0.4; **b** 0.5; **c** 0.6; **d** 0.7; **e** 0.8; **f** 0.9; **g** 1.0

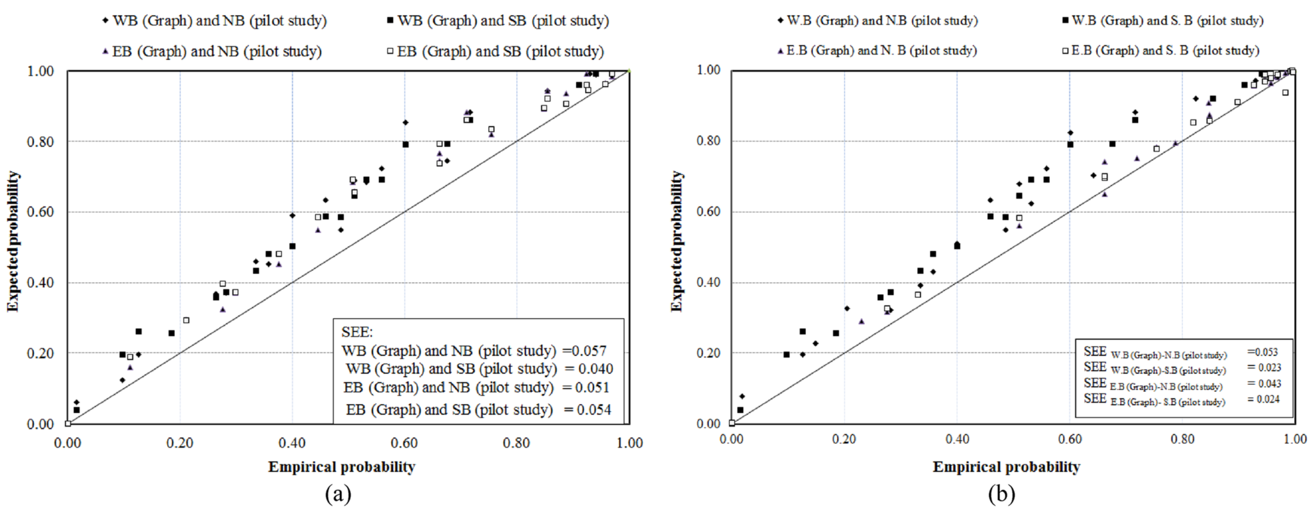

Fig. 8 Agreement between the probabilities of headways obtained from the graph and pilot field studies: **a** Shifted and **b** Combined model

Table 5 Comparison of headway probabilities: combined (current study) and single [21] distribution models

	Headway (Sec)	V/C ratio						
		0.4	0.5	0.6	0.7	0.8	0.9	1.0
Combined distribution: current study	< 3.0	0.749	0.826	0.834	0.312	0.852	0.552	0.862
	< 4.5	0.906	0.934	0.942	0.612	0.959	0.724	0.901
	< 7.5	0.974	0.981	0.985	0.852	0.992	0.925	0.924
	< 10.5	0.975	0.996	0.999	0.925	0.993	0.952	0.992
	< 13.5	0.982	0.998	0.999	0.936	0.995	0.995	0.996
	< 16.5	0.991	0.999	0.999	0.945	0.996	0.996	0.997
Single distribution [21]	< 3.0	0.312	0.352	0.412	0.523	0.567	0.621	0.668
	< 4.5	0.451	0.512	0.572	0.672	0.714	0.762	0.801
	< 7.5	0.568	0.625	0.712	0.772	0.788	0.862	0.911
	< 10.5	0.712	0.778	0.822	0.862	0.888	0.935	0.951
	< 13.5	0.758	0.832	0.884	0.912	0.925	0.952	0.963
	< 16.5	0.868	0.887	0.916	0.935	0.952	0.962	0.975

Further, empirical investigation reveals that the heterogeneity effect results in a deviation of the coefficient of variation and makes the exponential model inappropriate for describing headways. The study found that the hypo-exponential model is suitable when the coefficient of variation is less than one. In contrast, the hyper-exponential model is statistically valid when it exceeds unity. The following sections provide detailed insights into the research findings indicating limitations and scope for future work.

Conclusion

The present study used shifted-Erlang/negative exponential distributions as hypo-exponential and combined distribution as hyper-exponential. It found Erlang distribution with shifts of 0.675, 1.41 and 2.76 s and exponential distribution with a 0.495-s shift to exhibit acceptable statistical validity. The combined distribution function appeared as a suitable alternative to mixture distribution while describing headway data at moderate and heavy flow. A simplistic approach modelled shorter and longer headways separately using (a) Lognormal and Gamma distribution models in describing the shorter headways and (b) Lognormal, Gamma and negative exponential models for the longer ones. Lognormal distribution can describe shorter and longer headways up to moderate flow levels. However, Gamma distribution often best fits the data obtained in high and congested flow due to its flexibility and compatibility.

Headway models, established for describing low-flow states, indicate an insignificant proportion of shorter headways. In contrast, combined models confirm that the probability of shorter headways at moderate and heavy flow is relatively high compared to longer ones. The study observed that frequent vehicle following interaction at such flow levels results in platooning and entrapment of faster vehicles inside those platoons, compelling them to move with shorter headways. Further, empirical examinations substantiate that arrival pattern changes considerably with the change in flow and traffic composition resulting in inter-arrival time or time headways following different distributions at different flow levels.

Future Scope and Limitations

The models have been developed based on field data collected from a straight stretch of roads that pass through plain terrain. The study did not experiment with the impacts of varying geometric conditions, such as vertical and horizontal curvature and pavement conditions. Further, the scope of the paper considers endogenous conditions, such as (a) traffic flow rate and composition, (b) effects of non-motorized vehicles and (c) lead-lag vehicle pair combinations. It does not explain the impacts of exogenous conditions like weather and night visibility on headways. Thus, the study demonstrates the need for introducing a robust time headway modelling technique based on comprehensive field data.

Acknowledgements The part of the analysis presented in the paper has used the data collected in the CSIR-CRRI, New Delhi sponsored project "Development of Indian Highway Capacity Manual (INDO-HCM)". The authors sincerely acknowledge CSIR-CRRI. The authors thank the anonymous referees for their suggestions to improve the paper.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Ye F, Zhang Y. Vehicle type-Specific headway analysis using freeway traffic data. Transp Res Rec. 2009;2124(1):222–30. https://doi.org/10.3141/2124-22.
2. Thamizh Arasan V, Koshy RZ. Headway distribution of heterogeneous traffic on urban arterials. J Inst Eng (India). 2003;84:210–5.
3. Al-Ghamdi AS. Analysis of time headways on urban roads: case study from Riyadh. J Transp Eng. 2001;127(4):289–94. https://doi.org/10.1061/(ASCE)073947X(2001)127:4(289).
4. Zhang G, Wang Y, Wei H, Chen Y. Examining headway distribution models with urban freeway loop event data. Transp Res Rec. 2007;1999(1):141–9. https://doi.org/10.3141/1999-15.
5. Li L, Chen XM. Vehicle headway modeling and its inferences in macroscopic/microscopic traffic flow theory: a survey. Transp Res C Emer. 2017;76:170–88. https://doi.org/10.1016/j.trc.2017.01.007.
6. Panichpapiboon S. Time-headway distributions on an expressway: case of Bangkok. J Transp Eng. 2015;141(1):05014007. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000731.
7. Chandra S, Kumar R. Headway modelling under mixed traffic on urban roads. Road Transp Res. 2001;10(1):61–71.
8. Abtahi SM, Tamannaei M, Haghshenash H. Analysis and modeling time headway distributions under heavy traffic flow conditions in the urban highways: case of Isfahan. Transport. 2011;26(4):375–82. https://doi.org/10.3846/16484142.2011.635694.
9. Bham GH, Ancha SR. 2006 Statistical models for preferred time headway and time headway of drivers in steady state car-following. In Applications of Advanced Technology in Transportation pp 344–349
10. Dey PP, Chandra S. Desired time gap and time headway in steady-state car-following on two-lane roads. J Transp Eng. 2009;135(10):687–93. https://doi.org/10.1061/(ASCE)0733-947X(2009)135:10(687).
11. Ha DH, Aron M, Cohen S. Time headway variable and probabilistic modeling. Transp Res C-Emer. 2012;25:181–201. https://doi.org/10.1016/j.trc.2012.06.002.
12. Rossi R, Gastaldi M, Pascucci F. Gamma-GQM time headway model: endogenous effects in rural two-lane two-way roads. Procedia Soc Behav Sci. 2014;111:859–68. https://doi.org/10.1016/j.sbspro.2014.01.120.
13. Das S, Maurya AK. Time headway analysis for four-lane and two-lane roads. Transp Dev Econ. 2017;3:9. https://doi.org/10.1007/s40890-017-0039-8.
14. Roy R, Saha P. Headway distribution models of two-lane roads under mixed traffic conditions: a case study from India. Eur Transp Res Rev. 2018;10(3):1–12. https://doi.org/10.1007/s12544-017-0276-2.
15. Weng J, Meng Q, Fwa TF. Vehicle headway distribution in work zones. Transportmetrica A. 2014;10(4):285–303. https://doi.org/10.1080/23249935.2012.762564.
16. Kong D, Guo X. Analysis of vehicle headway distribution on multi-lane freeway considering car–truck interaction. Adv Mech Eng. 2016;8(4):1–12. https://doi.org/10.1177/1687814016646673.
17. Hoogendoorn SP, Bovy PH. New estimation technique for vehicle-type-specific headway distributions. Transp Res Rec. 1998;1646(1):18–28. https://doi.org/10.3141/1646-03.
18. Roy R, Saha P. Analysis of vehicle-type-specific headways on two-lane roads with mixed traffic. Transport. 2020;35(6):588–604. https://doi.org/10.3846/transport.2020.14136.
19. Das S, Maurya AK. Bivariate modeling of time headways in mixed traffic streams: a copula approach. Transp Lett. 2020;12(2):138–48. https://doi.org/10.1080/19427867.2018.1537209.
20. Penmetsa P, Ghosh I, Chandra S. Evaluation of performance measures for two-lane intercity highways under mixed traffic conditions. J Transp Eng. 2015;141(10):04015021. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000787.
21. Saha P, Roy R, Sarkar AK, Pal M. Preferred time headway of drivers on two-lane highways with heterogeneous traffic. Transp Lett. 2019;11(4):200–7. https://doi.org/10.1080/19427867.2017.1312859.
22. Risto M, Martens MH. Driver headway choice: a comparison between driving simulator and real-road driving. Transp Res F-Traf. 2014;25:1–9. https://doi.org/10.1016/j.trf.2014.05.001.
23. Sturges HA. The choice of a class interval. J Am Stat Assoc. 1926;21(153):65–6. https://doi.org/10.1080/01621459.1926.10502161.

24. Scott DW. Sturges' rule. WIREs Comp Stat. 2009;1(3):303–6. https://doi.org/10.1002/wics.35.
25. Saha P, Sarkar AK, Pal M. Evaluation of speed–flow characteristics on two-lane highways with mixed traffic. Transport. 2017;32(4):331–9. https://doi.org/10.3846/16484142.2015.1004369.
26. Bolch G, Greiner S, De Meer H, Trivedi KS. Queueing networks and Markov chains: modeling and performance evaluation with computer science applications. John Wiley & Sons; 2006.
27. Luttinen RT. (1996). Statistical analysis of vehicle time headways. Publication 87, Helsinki University of Technology. Transp Eng, Otaniemi
28. Singh LN, Dattatreya GR. Estimation of the hyperexponential density with applications in sensor networks. Int J Distrib Sens N. 2007;3(3):311–30. https://doi.org/10.1080/15501320701259925.
29. Griffiths JD, Hunt JG. Vehicle headways in urban areas. Traffic Eng Control. 1991;32(10):458–62.
30. Branston D. Models of single lane time headway distributions. Transp Sci. 1976;10(2):125–48. https://doi.org/10.1287/trsc.10.2.125.
31. Lilliefors HW. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. J Am Stat Assoc. 1967;62(318):399–402. https://doi.org/10.1080/01621459.1967.10482916.
32. Indian Roads Congress. 1990. Guidelines for capacity of roads in rural areas. IRC 64: 1990, New Delhi
33. Vimal A, Valluri SR, Karlapalem K. 2008. An Experiment with Distance Measures for Clustering. In COMAD: 241–244
34. Peng W, Li T. (2006). Interval data clustering with applications. In 18th IEEE International Conference on Tools with Artificial Intelligence: 355–362. https://doi.org/10.1109/ICTAI.2006.71
35. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans Pattern Anal Mach Intell. 2002;24(7):881–92. https://doi.org/10.1109/TPAMI.2002.1017616.
36. Dey PP, Chandra S, Gangopadhaya S. Speed distribution curves under mixed traffic conditions. J Transp Eng. 2006;132(6):475–81. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:6(475).
37. Luttinen RT. Statistical properties of vehicle time headways. Transp Res Rec. 1992;1365:92–8.
38. Saha P, Roy R. A descriptive study of vehicle classwise headways using mixed traffic data. J Inst Eng India Ser A. 2022. https://doi.org/10.1007/s40030-022-00666-w.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.