Transportation Research Interdisciplinary Perspectives 24 (2024) 101053

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

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

Reliability Assessment of Vehicle-to-Vehicle Communication Networks through Headway Distribution and Information Propagation Delay

Marzieh Khakifirooz [a,*], Michel Fathi [b], Lili Du [c]

[a] Department of Industrial Engineering, Tecnologico de Monterrey, Mexico
[b] Department of Information Technology & Decision Sciences, G. Brint Ryan College of Business, University of North Texas, USA
[c] Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL, USA

ARTICLE INFO

Keywords:
Headway distribution
Information propagation delay
Reliability
Kullback Leibler divergence
Vehicle-to-vehicle communication network

ABSTRACT

Several developed vehicle spaces and time headway distribution models in traffic flow theory have been widely used in the literature, reflecting the primary uncertainty in drivers' car-following movements and explaining the traffic flow stochastic features. Moreover, effective vehicle-to-vehicle (V2V) communication is a key to decentralizing traffic information systems. Accurate vehicle headway distribution estimation will ensure reliable communication and benefit passengers' safety and comfort. Consider several proposals for headway distribution in the literature; this paper studies the effect of space-headway distribution on information propagation delay in assessing the reliability of the V2V communication networks. We utilize the properties of reliability measurement for headway data by introducing the quasi-maximum likelihood estimator (QMLE) to measure the effects that different headway models have on estimating the parameters of headway distribution in a V2V communication network. The statistical analysis is then applied to the real Next-Generation Simulation (NGSIM) data. It validates the proposed methodology and formulations by measuring the effects of model selection on headway data and information propagation delay on the reliability of the V2V network. The results show that the effect of headway distribution when the vehicle's transmission range is smaller than the road segmentation is not negligible, especially when cars are very distant. Based on our results, we recommend new metrics based on Kullback–Leibler divergence for model selection of headway data, thereby enhancing the reliability of the V2V network.

1. Introduction

Vehicle headway is a crucial metric in transportation and traffic flow analysis. A vehicle space-headway is the distance between two consecutive vehicles passing the same point in the same lane, which explains how vehicles spread on the road (Biswas et al., 2021). The concept of vehicle time-headway is similar. Modeling uncertainties in traffic flows and drivers' car-following movements with vehicle headway distribution models are common in traffic flow analysis. Furthermore, many academics are interested in headway distribution research for various applications, including evaluating road capacity, traffic signal scheduling (Qian et al., 2021), and driver acceleration behavior (Singh and Kathuria, 2021).

With the advancement of mobility solutions, transportation systems constantly strive for sustainable development through smart technology to maintain a safe and quick response to the demands of society 5.0 (Fathi et al., 2019). In this context, a vehicular ad hoc network (VANET) facilitates data exchange among vehicles and from roadside infrastructure to vehicles. The VANET's communication resource management relies heavily on headway distribution (Du et al., 2016). Therefore, the correct headway distribution pattern must be determined to analyze traffic, connectivity, and infrastructure-related phenomena in VANET. It is necessary to know how much information may be transferred on the road between vehicles and how many nodes (connecting vehicles) are required to establish a stable connection to have a reliable VANET (Ukkusuri and Du, 2008).

The likelihood of information propagation delay in a VANET with either congestion or free traffic flow conditions is not negligible. On the other hand, a variety of other sources of uncertainty, such as inaccurate data, simulator randomness, model discrepancy, and statistical variability in estimating and calibrating input parameters, can significantly widen the gap between real-world and simulated data, affecting the accuracy of the headway distribution and, as a result, the connectivity in VANET (Weber et al., 2021). Consequently, there will be significant

* Corresponding author.
E-mail addresses: mkahkifirooz@tec.mx (M. Khakifirooz), mfathi@unt.edu (M. Fathi), lili.du@essie.ufl.edu (L. Du).

https://doi.org/10.1016/j.trip.2024.101053
Received 14 September 2023; Received in revised form 27 January 2024; Accepted 25 February 2024
Available online 29 February 2024

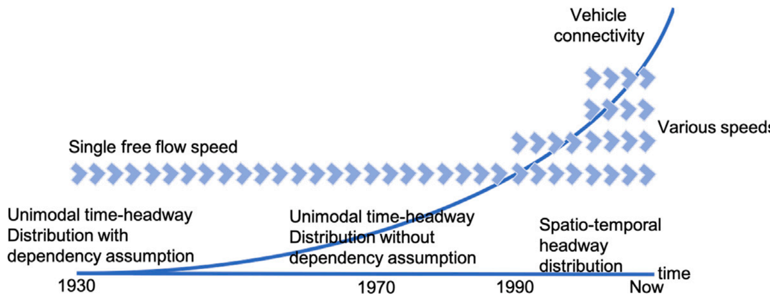

Fig. 1. Summary of vehicle headway distribution models and development during the time (Li and Chen, 2017).

errors in decision-making. Any inaccurate estimation of related parameters in VANET will cause an inaccurate estimation of information propagation delay and, therefore, an unreliable network. Therefore, selecting an inaccurate distribution for headway data will significantly affect traffic flow rate calculations and connectivity reliability.

Despite these facts, this work aims to investigate the impact of vehicle headway distribution on VANET reliability and information propagation delay. For this purpose, we consider situations where having a known distribution of space headway is not a reasonable assumption and, therefore, the true model is unknown. In this case, we force the use of an estimation model, thereby running the risk of misspecifying the model. To investigate this, we start by defining a measure of the discrepancy between different models, called the Kullback–Leibler divergence, and investigate the effect of such a risk on the stability and reliability of decisions on information propagation delay.

The remainder of this paper is structured as follows: Section 2 discusses the literature review of the previously published studies in this area. Section 3 describes the problem of reliability failure in VANET due to model misspecification, followed by Section 4, which describes the problem and formulates the time delay in VANET. Section 5 proposes the reliability measurement of VANET based on headway distribution and the theoretical computation of this assessment for information propagation delay. Finally, the numerical experiments in Section 6 validate the proposed approaches. We conclude this research in Section 7.

2. Literature review

In the literature, attention to the reliability assessment of VANET comes mainly from hardware functionality, road infrastructures, data management, and broadcasting systems. To the best of our knowledge, only a couple of papers see this problem from the theoretical point of view in real-world simulation. These few works cover the effect of vehicle transmission range, vehicle density on the connectivity probability (Hussain et al., 2019), and string stability of a one-vehicle look-ahead platoon under different network loss scenarios (Alsuhaim et al., 2021). Therefore, our approach to considering the efficiency of selecting headway distribution to assess the reliability of VANET is novel. Therefore, in this section, we focus on the challenges in selecting proper headway distribution addressed in the literature.

There are several headway distribution models in the literature. A comprehensive literature review on modeling headway data and its inferences in macroscopic/microscopic traffic flow analysis is discussed in Li and Chen (2017). They classified the vehicle headway distribution models into 1930–1970s, 1970–1990s, and 1990s-present, and their findings are summarized in Fig. 1.

According to Fig. 1, it can be seen that several different distributions have been used in the literature for modeling the headway data. Therefore, practitioners and researchers have critical questions about which distribution is better. Under what conditions would you use this distribution? What will happen if the selected distribution is not the best fit for the data? To answer these questions, Table 1 summarizes some of the most important headway distribution models in the literature, along with their applications and characteristics. In most early literature, as shown in Fig. 1, the headway distribution is denoted as a time-headway distribution based on vehicle speed instead of a space-headway distribution.

Regarding several approaches to selecting the best distribution for both time and space headway data, it is necessary to understand traffic phase transition and behavior to assist with planning successful vehicle-to-vehicle (V2V) communication in a VANET. Considering this need, we can categorize the existing literature into three different groups based on their contributions, as follows:

I : Discussing the applications of headway distribution in VANET,
II : Comparing different distributions for specific measurements,
III : Showing the bias and uncertainties in selecting an improper headway distribution.

2.1. Group I literature

Back to the historical review of literature, the order of implementation of headway distribution in various studies can be categorized into four main applications: (1) modeling the traffic flow for the use of pedestrians, allocating traffic lights, and pedestrian crossings; (2) estimating the traffic noise for pollution reduction (air and sound) and urban resilience management; (3) identifying abnormal road conditions and enhancing the public transportation service level (Toledo et al., 2010; Jamili and Aghaee, 2015; Morales et al., 2020), and (4) managing the inter-vehicle communications in VANET due to the rapid development of intelligent transportation systems (ITS) and AVs.

Yang and Recker (2005) utilized vehicle distribution in a very turbulent and stable traffic stream to simulate traffic information propagation based on V2V communication; their simulation model calculated the probability of successful communication and traffic information propagation distance for different roadways under incident and incident-free conditions. Saito (2006) evaluated the performance of combined vehicular communication in ITS and measured the delay of indirect information delivery using mean space headway, vehicle speed, and device connectivity penetration ratio. Ukkusuri and Du (2008) studied the geometric connectivity of VANET with disturbance to capture drivers' uncertain behavior in traffic flow. Also, they analytically derived the lower bound of reachable neighbor vehicles for each vehicle to maintain high connectivity.

Wang et al. (2010) studied the robust information propagation process in a traffic stream with a general identically and independently distributed (i.i.d) concept for V2V spacing and calculated the expected value, variance of information propagation distance, and connectivity between two vehicles. Yin et al. (2013) derived the probability distribution of information propagation distance and its expectation and

Table 1
Some representative headway distribution models.

Distribution	Characteristics	Application
Gaussian	1: Each event is independent of other events; 2: Equal intervals of time include an equal number of events	Movement of vehicles is considered a random series.
Negative exponential	The traffic stream is normally distributed	When the occurrence of the interval between two consecutive vehicles is greater than a random variable
Shifted exponential (Cowan's M2)	Has a location parameter that shows the minimum headway between vehicles	A $M/D/1$ queueing system with service time equal to the location parameter
Shifted Lognormal	Consider a minimum space between vehicles which present as shift parameter	Traffic stream at high flows when individual headway distribution is Lognormal
Hyperlang	Linear combination of exponential and Erlang distributions	Model single-lane traffic road flow in two-lane two-way road
Hyperexponential	The distance among vehicles is large enough that no vehicle is bound on one side by other vehicles.	When the flow rate rises above a few dozen vehicles per hour and line
Semi-Poisson model	1: Data is semi-random, 2: There is a threshold that separates data in random and uniform partitions multiplied by weighted parameter	Applies for negatively autocorrelated headway at very high flow roads
Cowan's M3	Special case of shifted exponential distribution when location parameter has a mixed distribution	There is a mixture of vehicles: a group that follows a specific space headway and another group that drives freely with larger space headway than the first group
Mixed-vehicle-type (Hoogendoorn and Bovy, 1998)	Rely on Branston's generalized queueing model	Emphasize vehicle type and period of the day
Type I Extreme Value (Hung et al., 2003)	The length of data has a quadratic relationship with parameters	Use when modeling vehicle discharge headway at signal-controlled intersections
Double Displaced Negative Exponential (Zhang et al., 2007)	The mixture of two components of headway: (1) tracking or following, and (2) free component	Modeling headway of high-occupancy vehicle lanes
Double exponential (Jin et al., 2009)	Consisting of two exponential distributions glued together on each side of a threshold	Modeling vehicle discharge headway at signal-controlled intersections
Farlie–Gumbel–Morgenstern (Zou et al., 2014)	Maximize correlation coefficient between marginal distributions	Describe the characteristics of speed and headway simultaneously
Inverse Gaussian (Kong and Guo, 2016)	Describes the distribution of the time a Brownian motion with positive drift takes to reach a fixed positive level	Consider multi-lane freeway with car–truck interaction
Pearson 5 and 6 (Roy and Saha, 2018)	Special cases of inverse-gamma and beta-prime distributions, respectively, to control the skewness and kurtosis.	Modeling heavy traffic flow and providing a decent fit under off-peak flow
Gaussian mixture (Zhou and Zhu, 2020)	Sum of weighted Gaussian distributions with positive parameters	Utilized when the distribution of speed data exhibited bimodality and skewness
Mixture model (Li et al., 2021a)	The model can be applied for most of the common headway distribution, considering classification results based on the correlation coefficient between macroscopic and microscopic traffic flow	Modeling the headway data by combining weighted random free flow, steady car-following, steady free flow, and block car-following models.
Uniform (Li et al., 2021b)	When an arbitrary outcome lies between certain bounds	For measuring the homogeneous degree of headway distribution

variance for connectivity of V2V on two parallel roadways using headway distribution. Wang et al. (2015) analyzed the vehicle connectivity in VANET with two parallel roads; when the distance between roads is smaller than $\frac{\sqrt{3}}{2}$ of transition rate, then the probability distribution of the information propagation distance and its expectation and variance are derived.

Wang et al. (2018a) developed a distributed optimization algorithm for V2V communication in a multi-agent system to minimize the communication time delay over a directed graph. Wang et al. (2018b) proposed an extended car-following model at un-signalized intersections to examine the impacts of V2V communication on micro-driving behavior. Tan et al. (2020) developed a car-following model under V2V communication to improve traffic safety in the sand-dust environment. They showed that V2V communication is active in reducing headway fluctuations, acceleration, and velocity when there is a disturbance in the sand-dust area's traffic flow. Li et al. (2020) analyzed active and inactive V2V communication for car-following behavior with the effect of sudden acceleration changes using headway distribution. In summary, it can be seen that the application of headway distribution in VANET has the following pattern:

(1) Simulate the traffic flow and density for different time windows and road types based on headway distribution	→	(2) Estimate the communication power by the number of connected vehicles using traffic flow and headway data	→	(3) Measure the probability of communication failure and information propagation delay according to the communication power and headway information.

Therefore, if the headway distribution is not selected properly, there would be a dramatic bias in simulating traffic flow and communication power and estimating the probability of communication failure and information propagation delay in VANET.

2.2. Group II literature

Most studies only gave a rough description of the fitness of a headway distribution for a particular traffic scenario without rigorous statistical tests. Statistical goodness of fit tests is the most notable reference to prove the advantages of selected distributions. For instance, Weng et al. (2014) disaggregated the vehicle headway for car-truck interaction. Their results showed that the headway distribution of car/truck interaction is significantly different. They used Mann–Whitney and Kruskal–Wallis tests to show that the car/truck interaction's headway distribution is significantly different. Their methodology determined the best headway distribution by maximum-likelihood estimation and the Kolmogorov–Smirnov test as a function of the traffic flow rate, work intensity, percentage of trucks, and lane position, impacting vehicle headway in work zones. In another study, Kong and Guo (2016) studied vehicle headway distribution characteristics on the multi-lane freeway, considering car-truck interaction and different vehicle types, and fitted the best headway distribution among possible choices using statistical inferences.

2.3. Group III literature

There will likely be numerous models that match the data well and appear to be viable models for analyzing any particular dataset. Because various models provide different inferences, statistical goodness of fit is a required metric but is not a sufficient reason for distribution selection. Therefore, researchers need to consider that the validity of traditional statistical results rests on the assumption that the underlying model is correct, which is a much stronger criterion than just confirming that the model gives a good fit to the data (Copas and Eguchi, 2020). Nevertheless, investigating the effect of not selecting the best/better distribution can provide an additional level of certainty for accepting a distribution as a good fit.

To the best of our knowledge, only a few studies have this perspective and discuss the side-effects of selecting an improper distribution for headway data. For instance, Abuelenin and Abul-Magd (2015) studied the moment analysis (mean, variance, skewness, and kurtosis) of freeway-traffic clearance distribution. They computed V2V spacing in a multi-lane freeway and calculated the correlation of vehicles' spacing in different lanes. They found that the calculated kurtosis differs from the skewness for different lanes during the different time intervals. The skewness and kurtosis as functions of the mean V2V spacing produce sharp peaks at critical densities expected for transitions between different traffic phases. Besides, Blumenfeld and Weiss (1975) showed that for a fixed traffic density of moderate or heavy traffic, increasing the distribution order decreases the variance of traffic noise and emphasizes higher frequencies in the spectral density.

The effect of selecting an improper distribution caused by mathematical modeling or the interpolation of observations of the error term can be a disruptive factor in the simulation and management of transportation systems. Considering this fact, some indexes, such as mean square error, which is very common in measuring the efficiency of estimators and models, do not show the exact variability and bias of the estimator (even from nearly unbiased estimators, such as the maximum likelihood estimator). In this situation, asymptotic metrics (when the limits of sample size tend to be infinite), such as asymptotic efficiency and bias, are more practical to show the power of estimators and models. In summary, in terms of contribution and implementation, our research can be compared with a few literature in the field, as summarized in Table 2. Therefore, considering the gap in the literature, the main contributions of this paper are:

1. to introduce a new metric based on Kullback Leibler divergence as a reliability measurement index for estimating the delay of information propagation in the VANET system;
2. to bring the attention of practitioners and researchers in the field of transportation technology to the impact of improper distribution selection;
3. to derive the asymptotic distribution and properties of time delay on tail quantiles (when vehicles are very far from each other) of headway distribution.

3. Problem explanation: Reliability under headway model uncertainty

The importance of statistical challenges in transportation study and practice will grow as new mobility technologies and systems develop. For instance, to effectively predict the dynamics of autonomous vehicles (AVs) in the transportation network, V2V, vehicle-to-infrastructure, and infrastructure-to-vehicle technologies, because AVs are not yet realistic, driving simulations are increasingly being used to identify such dynamics or surrogates. Therefore, addressing statistical-related issues directly impacts the successful implementation of new transportation systems.

In traffic simulation, little attention has been paid to uncertainty considerations in simulation model inputs. Instead, the focus is mainly on analyzing the stochastic outputs of various performance measures resulting from different models. Bayarri et al. (2004) assessed uncertainties in traffic simulation by model calibration and validation. The calibration and validation process of traffic simulation models for application to a transportation network is inherently tricky, and it is typically handled informally using a variety of ad hoc approaches. Field data, which are often restricted and expensive to get but necessary for defining inputs to the simulation model and assessing the model's reliability, are explicit parts of the calibration and validation process. The quantification and systematization of the calibration/validation process reveal the statistical challenges of using such data to determine a model's validity. The validation is usually phrased as "does the model accurately represent reality?" However, the answer to this question is straightforward: *models are not perfect*.

Researchers and practitioners can choose the most effective modeling approach and find acceptable procedures for evaluating and validating statistical results by better understanding the "model-building" process. The computation of sensitivity factors such as marginal effects and statistical performance metrics, including goodness-of-fit and prediction accuracy, are common examples of these techniques. However, structural discrepancies between various state-of-the-art statistical and econometric methodologies may limit the applicability or practicality of some measures. In this context, best practices in transportation data analysis may not follow a uniform standard for the pre- and post-implementation phases of statistical modeling. Because of the disparities in model implementation and evaluation and the data-intensive nature of transportation analyses, the accuracy of some models may be questionable.

Therefore, we can (and have to) quantify the accuracy of a model by:

$$\Pr[|\text{reality} - \text{simulation prediction}| < \delta] > \alpha \tag{1}$$

where δ is the tolerable difference (how close) and α is the level of assurance (how certain), with statements of what reality means, what is needed to make sense, and how to compute the probabilities involved. Reality refers to an operationally possible measure of a network's actual performance (e.g., it could be the time between two consecutive vehicles to pass a certain location).

Accomplishing the task in (1) is to minimize δ while making sense of (Pr), yet this is a challenging task. Therefore, a closer examination of the model's inputs based on simulators is required. Geometric inputs, such as lane widths and bus stops, can be easily obtained by exact measurements or documented sources. However, some parameters can be discovered by calibration or the use of field data, including:

- Type A : Parameters that can be directly estimated from field data (such as vehicle arrival rates, turning ratio, and traffic volume), but estimation includes some percentage of uncertainties.
- Type B : Parameters that are not directly measurable (such as the degree of driver aggressiveness) or depend on the choice, such as discharge headway distribution.

Table 2
Comparison of this study with a similar study on the reliability of VANET.

Reference	Objective	Methodology	Key Findings	Contribution
Hussain et al. (2019)	Assessing VANET reliability considering vehicle transmission range and density	Real-world simulation	Effect of vehicle transmission range and density on connectivity probability	Theoretical perspective on VANET reliability
Alsuhaim et al. (2021)	Studying string stability in one-vehicle look-ahead platoon under network loss	Network loss scenarios analysis	Impact of network loss on platoon stability	Insights into string stability in VANETs under network loss
Li and Chen (2017)	Comprehensive review of headway distribution models	Literature review	Classification and evolution of headway distribution models over time	Historical perspective on headway distribution models
Current study	Assess the reliability of VANET focusing on headway distribution efficiency	Analyzing headway distribution in VANETs; real-world simulation	Novel insights into the efficiency of different headway distributions for VANET reliability	Introducing a new approach in VANET reliability assessment by focusing on headway distribution efficiency

Type C : Tuning parameters (e.g., free-flow speed, lost time) that are not real but are required by the simulation model.

Some of the parameters mentioned above can be replaced in the calibration process with a sensitivity analysis. The remaining parameters are often adjusted until the model output fits the field data of the real traffic system. However, the accuracy of how the model is “well fitted” is subjective to the other existing models and is informal with the opinion of experts. Therefore, there are two major issues with this type of calibration approach. First, there could be a “misspecification” problem, which allows the model to be tuned in various ways while creating potentially huge uncertainty in the estimated parameters, resulting in inaccurate model predictions. Second, “over-fitting” may occur, in which fitting obscures model flaws and results in potentially very inaccurate models outside the range of observed field data.

In this regard, if we have more than one candidate model for fitting data, one interesting question would be, “How much can we compensate/repudiate the bias of estimators by changing from one model to another and keeping the connectivity in the VANET network reliable?” To answer this question, we need to find the difference between the effect of each simulated model on real data, measure how significant the changes of models are, and determine if it is necessary and cost-effective to look for a better model.

Several phenomena can make the accuracy of headway distribution questionable. For instance, the long time/space headway is associated with the high intensity in the driving zone (Weng et al., 2014). The omission of drivers’ familiarity with driving zones and forcing a fixed value on the driving intensity’s coefficient can lead to low accuracy of the headway model (Weng et al., 2019). Therefore, knowledge of vehicle headway distribution is essential for estimating connectivity in VANET. With the rise of VANETs, connectivity became a fundamental requirement for applying such networks. Connectivity analyses require very accurate distribution models.

4. Problem formulation of time delay in VANET

To the best of our knowledge, there are only a few models in the literature for estimating the time delay of information propagation. In many other studies, the tolerance of delay is considered to be known. Zarei and Rahmani (2016) considered the vehicles’ physical movement as a catch-up process and ended with multihop transmission through connected vehicles as a forwarding process. Therefore, the information propagation process cyclically renews, and the information propagation speed is related to the number of renewal cycles for delivering information. They employ the distribution of connectivity distance (headway distribution) in VANET, the probability that at least one pair of vehicles communicate, and the number of connected vehicles for modeling information propagation delay. Similarly, for one-way or two-way road segments with many lanes, Du and Dao (2014) developed analytical formulations to estimate information propagation time delay over a V2V communication network. Zarei and Rahmani (2016) added the average vehicle speed carrying information, transmission rate, and transmission time to Du and Dao (2014) models. As a more comprehensive model, we employ the proposed model by Du and Dao (2014) for estimating the time delay of information propagation in this study to assess the influence of model misspecification.

Table 3
List of mathematical notations and expressions in this study.

Notations	Definitions
i	Vehicle’s index. $i = 0, \ldots, n$
k	Number of hob nodes when vehicles are well connected as an instantaneous transmission happens.
t_1	Time delay in an intermittent transmission flow.
t_2	Time delay in a ferry transmission flow.
S	The distance (space headway) between two consecutive vehicles.
r	The transmission range.
τ	The wireless transmission time delay.
Y	The information propagation distance in a ferry transmission flow.
X	The information propagation distance in an intermittent transmission flow.
T	The total time delay.
L	The length of the road segment.
v_i	The average speed of the vehicle i carrying information.
$f(s)$	The space headway distribution between two vehicles in a lane or across different lanes.
B_i	ith vehicle in a sequence of vehicles within a traffic flow or network
P	Probability of successfully making instantaneous transmission hob with the next vehicle.
u_i	ith percentile of a given probability distribution

4.1. Modeling time delay of information propagation

Consider the main mathematical notations in this study, as summarized in Table 3. In a dynamic traffic flow, the wireless connection is occasionally connected (resulting in instantaneous transmission) and broken (leading to ferry transmission) due to relative movement between vehicles. In a mildly congested traffic flow, intermittent transmission is common. Intermittent transmission is exemplified by the ferry and instantaneous communications. They are most common in sparse and heavily congested traffic flows, where wireless communication between two cars occurs only occasionally or regularly. An intermittent transmission continues a pattern of instantaneous and ferry transmissions until the information reaches the end of the road section. Several vehicles are well connected as an instantaneous transmission happens, and information is smoothly conveyed from node i to $i + k$ (see Fig. 2).

When each vehicle in the VANET system can successfully communicate with at least one other vehicle as a communication hop; thus, according to Du and Dao (2014), intermittent transmission follows each hop of wireless transmission and will succeed if S is less than r (i.e., $S < r$). If the space headway between a transmitter and a receiver

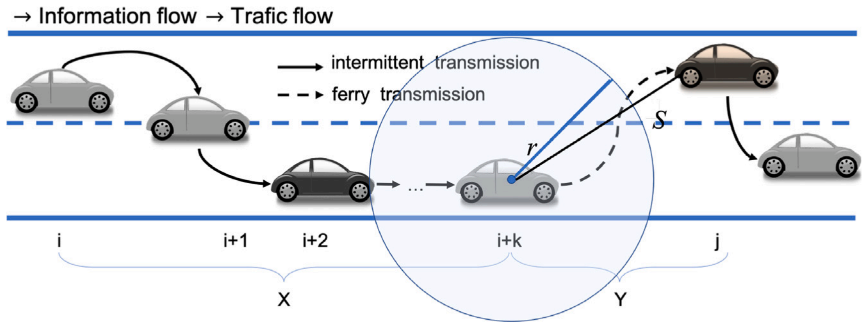

Fig. 2. Schematic diagram of intermittent and ferry transmission in a one-way road.

(two consecutive vehicles) is $S > r$, the information flows by ferry transmission.

The equivalent time delay in an instantaneous transmission flow is computed by $t_1 = k\tau$. In a VANET, a piece of information may travel multiple hops in an instantaneous transmission along with the well-connected vehicle network on a road segment until the communication connection fails; at this point, the information propagation becomes ferry transmission. Ferry transmission is common due to the lack of traffic. In particular, a vehicle will transport information until it meets another vehicle, such as at node $i+k$, then transported by node $i+k$ until it reaches the next node (see Fig. 2). As a result, in a ferry transmission flow, $t_2 = Y/v_i$ is used to determine the corresponding time delay.

The expected time delay for one intermittent transmission (i.e., an instantaneous transmission followed by a ferry transmission and vice versa) is multiplied by the expected times when the intermittent transmission occurs to estimate the time delay of a piece of information traversing a road segment. This concept is expressed mathematically by Du and Dao (2014) as follows:

$$E(T) = \left(E(t_1) + E(t_2)\right) \frac{L}{E(X) + E(Y)} = \left(E(k)\tau + \frac{E(Y)}{E(v_i)}\right) \frac{L}{E(X) + E(Y)}, \tag{2}$$

where

$$E(X) = E(k)E(S|S \le r), \tag{3}$$

and

$$E(Y) = \frac{E(S|S > r) - r}{E(v_i - v_j)}. \tag{4}$$

where $E(v_i - v_j)$ represents the average relative speed of two vehicles i and j.

Considering the dynamic feature of propagation distance under instantaneous and ferry transmission and the number of hops. The conditional random variable $S|S \le r$ represents the space headway in instantaneous transmission, given that information always propagates to its nearest neighbor, and can be expressed as follows:

$$E(S|S \le r) = \frac{\int_0^r s f(s)ds}{\int_0^r f(s)ds}, \tag{5}$$

given

$$f(S|S \le r) = \frac{dF(S|S \le r)}{ds} = \frac{f(s)}{\int_0^r f(s)ds},$$

and

$$F(S|S \le r) = \Pr(0 < S < s | 0 < S \le r) = \frac{\Pr(0 < S < s, 0 < S \le r)}{\Pr(0 < S \le r)} = \frac{\int_0^s f(s)ds}{\int_0^r f(s)ds}$$

Similarly, $E(S|S > r)$ can be calculated using $S|S > r$ as a conditional random variable:

$$E(S|S > r) = \frac{\int_r^{+\infty} s f(s)ds}{\int_r^{+\infty} f(s)ds}, \tag{6}$$

given

$$f(S|S > r) = \frac{dF(S|S > r)}{ds} = \frac{f(s)}{\int_r^{+\infty} f(s)ds},$$

and

$$F(S|S > r) = \Pr(0 < S < s | r < S < +\infty) = \frac{\Pr(r < S < s)}{\Pr(r < S < +\infty)} = \frac{\int_r^s f(s)ds}{\int_r^{+\infty} f(s)ds}.$$

The mathematical background of $E(k)$ is presented in the following sub-section.

4.2. Estimating $E(k)$: expected hops in an instantaneous transmission

A piece of information may propagate multiple hops in an instantaneous transmission along with a road segment's well-connected vehicle network until the communication link breaks and the information propagation becomes a ferry transmission. As a result, k, where the number of hops equals the number of cars that can communicate effectively within an instantaneous transmission.

To calculate $E(k)$, we consider $n+1$ vehicles on the road segment and mark them with numbers ranging from 0 to n from left to right. Then, if an instantaneous transmission starts at any individual vehicle, $B_i (i = 0, \dots, n)$, evenly with probability $\Pr(B_i) = \frac{1}{n+1}$, and P_k represents the probability of k hops of successive transmission; therefore, the probability of an instantaneous transmission starting at vehicle i and only successively propagating k hops is $\Pr(P_k \cap B_i) = \Pr(P_k|B_i)\Pr(B_i)$. As a result, we treat k as a random variable and investigate its probability concerning B_i using the below strategy:

hops = 0	For all vehicles, there is a probability of $1 - P$ failing to make an instantaneous transmission hop with the next vehicle except vehicle B_n (the last vehicle), with the probability that one cannot make a transmission.	$\Pr(P_0|B_{i=1,\dots,n-1}) = 1 - P$ & $\Pr(P_0|B_n) = 1$
hops = 1	For all vehicles, there is a probability of making one successful instantaneous transmission (P_1) with the next vehicle and then failing ($1 - P$) except vehicle B_{n-1}, with the probability P_1 that can make the transmission with vehicle n without failure.	$\Pr(P_1|B_{i=1,\dots,n-2}) = P_1(1 - P)$ & $\Pr(P_1|B_{n-1}) = P_1$ & $\Pr(P_1|B_n) = 0$

hops = 2	For all vehicles, there is a probability of making two successful instantaneous transmissions (P_2) with the next vehicle and then failing $(1-P)$, except for vehicle B_{n-2} with the probability P_2 that can make the transmission with vehicles $n-1$ and n without failure.	$\Pr(P_2\|B_{i=1,\dots,n-3}) = P_2(1-P)$ & $\Pr(P_2\|B_{n-2}) = P_2$ & $\Pr(P_2\|B_{n-1,n}) = 0$
⋮		
hops = k	For all vehicles, there is a probability to make k successful instantaneous transmission (P_k) with the next vehicle and then fail $(1-P)$, except vehicle B_{n-k} with probability P_k can make the transmission with vehicle $n-k+1,\dots,n$ without failure.	$\Pr(P_k\|B_{i=1,\dots,n-k-1}) = P_k(1-P)$ & $\Pr(P_k\|B_{n-k}) = P_k$ & $\Pr(P_k\|B_{n-k+1,\dots,n}) = 0$
⋮		
hops = n	Only vehicle B_0 with probability P_n can make n successful instantaneous transmissions with vehicles $1,\dots,n$ without failure.	$\Pr(P_n\|B_0) = P_n$ & $\Pr(P_n\|B_{1,\dots,n}) = 0$

Consider $g(k)$ as a probability of an instantaneous transmission with only k hops, or $g(k) = \bigcup_{i=0}^{n} \Pr(P_k \cap B_i)$a general formulation of $g(k)$ can be derived as follows:

$$g(k) = \frac{(n-k)P_k(1-P) + P_k}{n+1} \tag{7}$$

Consequently, the $E(k)$ under free-flow traffic conditions can be easily calculated by $E(k) = \sum_k kg(k)$. However, the space between two vehicles is relatively small in congested traffic conditions. The movement of the following vehicle needs to consider the movement of the leading vehicle in front to keep it safe; therefore, the spacing between any two consecutive vehicles is dependent, which causes difficulty in accurately calculating P_k. To address this challenge, Du and Dao (2014) provides a lower and upper bound for $E(k)$. In this study, we employ the same approach as (Du and Dao, 2014), where the lower and upper bounds of $E(k)$ are calculated as follows, respectively:

$$E(k) \geq \sum_{k=1}^{n} k \frac{(n-k)(kP-(k-1))(1-P) + kP - (k-1)}{n+1} \tag{8}$$

$$E(k) \leq \sum_{k=1}^{n} k \frac{(n-k)(2P-P^2)(1-P) + 2P - P^2}{n+1} \tag{9}$$

5. Quantile estimation of information propagation delay and reliability assessment of VANET

In the literature on traffic flow modeling, considering the type of traffic flow from free flow to heavily congested traffic, Exponential, Lognormal, and Normal distributions are among the most commonly used distributions in fitting headway data (Yin et al., 2009); however, these may not be the best choice for modeling headway distribution. The more precise the headway distribution, the more unified the communication system between vehicles will be; thus, various applications contributing to passenger and vehicle safety will be available. In addition, in most of the literature, a uni-modal structure for traffic flow is considered. In reality, the traffic can change from very congested to a very sparse model during the day. Therefore, (even in a very congested traffic flow), giving attention to tail information (when vehicles are very far from each other) in the connectivity measurement of VANET is necessary.

In this section, we show derivations of quantile function based on entropy and its relation with the expected value of a random variable to assess the reliability of the VANET network based on information propagation delay for two cases of (1) single distribution and (2) competitiveness of two distributions.

5.1. Single distribution

To perform the communication among vehicles, consider n sample headway data. Let S_i be the ith observation, with $f(s)$ and $F_S(s_i)$ as the probability and cumulative distribution functions, respectively, where $1 - F_S(s_i)$ can be considered as a survival function. Therefore, the measure of uncertainty is defined by:

$$\xi(S) = -\int_0^\infty \log(f(s))f(s)dS = -E[\log f(s)], \tag{10}$$

as the expected uncertainty contained in the distribution of the predictability of an outcome of S, known as the Shannon entropy measure. Based on this idea, the residual Shannon entropy of S at point r is:

$$\begin{aligned}\xi(S|S>r) &= -\int_r^\infty \frac{f(s)}{F(S|S>r)} \log\left(\frac{f(s)}{F(S|S>r)}\right) dS = \\ &\log F(S|S>r) - \frac{1}{F(S|S>r)} \int_r^\infty \log(f(s))f(s)dS = \\ &1 - \frac{1}{F(S|S>r)} \int_r^\infty \log \frac{f(s)}{F(S|S>r)} f(s)dS\end{aligned} \tag{11}$$

A similar function can be obtained when $S < r$.

In addition, the quantile function of $log(S_i)$ can be defined as

$$Q_{S_i}(u_i) = F_S^{-1}(u_i) = \inf\{s_i|F_S(s_i) \geq u_i\}, \quad 0 \leq u_i \leq 1 \tag{12}$$

where $F_S(Q(u)) = u$. Similar to nations for (5) and (6) we can define the quantile density function by $q(u) = \frac{Q(u)}{du}$, where $q(u)f_S(Q(u)) = 1$. Therefore, the $\xi(S)$ in (10) can be expressed as:

$$\xi(S) = \int_0^1 \log q(u)du \tag{13}$$

Clearly, by knowing either $q(u)$ or $Q(u)$, the expression for $\xi(S)$ is quite straightforward. In addition, an equivalent definition for the entropy of quantile function is given by:

$$\xi(Q(u)) = \log(1-u) + \frac{1}{1-u} \int_u^1 \log q(u)du \tag{14}$$

$\xi(Q(u))$ measures the expected uncertainty contained in the conditional density about the predictability of an outcome of S until the $100(1-u)\%$ point of distribution. Furthermore, differentiating (14) with respect to u results in:

$$\frac{d\xi(Q(u))}{du} = -\frac{1}{1-u} + \frac{1}{(1-u)^2} \int_u^1 \log q(p)dp - \frac{1}{1-u} \log q(u)$$

or

$$(1-u)\frac{d\xi(Q(u))}{du} = -1 + \xi(Q(u)) - \log(1-u) - \log q(u),$$

then

$$q(u) = \exp\left(\xi(Q(u)) - (1-u)\frac{d\xi(Q(u))}{du} - \log(1-u) - 1\right). \tag{15}$$

Therefore, when $E(X) = \int_0^1 Q_X(u)du$, we can use the notations in (15) for updating (2) based on quantile function and entropy. In a numerical analysis in the next section, we will update the upper limit of the integral of $E(X)$ based on $Q(u)$ to focus on only tail quantiles and assess the reliability of the VANET network based on information propagation delay.

5.2. Two competitive distributions

In practice, to minimize the effects of information propagation delay, we want to select a distribution that, while minimizing δ in (1), does not overestimate or underestimate the data. In this regard, there might be more than one candidate distribution for modeling data. Therefore, let us consider the problem of predicting the unknown parameter θ as the property of space headway distribution. Given a collection of predictors, we select the predictor $\hat{\theta}$ that minimizes the discrepancy with θ. The common approach is to select $\hat{\theta}$ that minimizes an unbiased estimate of the expected squared loss, or $\mathrm{E}\left|\theta-\hat{\theta}\right|^2$. In this context, we investigate the interest of going beyond squared losses by estimating a loss function grounded on an information-based criterion, namely, the Kullback–Leibler divergence.

The Kullback–Leibler divergence can measure the information loss when an alternative distribution such as $f'_S(s|\theta')$ is used to approximate the underlying distribution given by

$$I(\theta,\theta')=\int \mathrm{d}f_S(s|\theta)\log\frac{\mathrm{d}f_S(s|\theta)}{\mathrm{d}f'_S(s|\theta')} \tag{16}$$

Unlike square loss, the Kullback–Leibler divergence does not measure the discrepancy between an unknown parameter and its estimate but between the (unknown) distribution f of S and its estimate f'. Therefore, it is invariant with one-to-one reparametrization of parameters and transformation of S because the transforms do not affect the quantity of information carried by S. While the squared loss is defined irrespective of the noise distribution, the Kullback–Leibler divergence could adjust its penalty for the scale and shape of the deviation, and it accounts for heteroscedasticity.

The Kullback–Leibler divergence is a non-symmetric discrepancy measure. The model in (16) could give interpretation to the question "How well $f'_S(s|\theta')$ explains an independent copy of S?", and the following equation could provide an answer to "how well $f'_S(s|\theta')$ generates an independent copy of X":

$$I(\theta',\theta)=\int \mathrm{d}f'_S(s|\theta')\log\frac{\mathrm{d}f'_S(s|\theta')}{\mathrm{d}f_S(s|\theta)} \tag{17}$$

Therefore, to know which distribution is better and what will happen if the selected distribution is not the best fit for the data, we extend the result in the last section to two competitive distributions. To do this, let f and f' be two continuous distributions with quantile functions $Q(u)$ and $Q'(u)$, respectively. Assume that $Q''(u)=Q'^{-1}(Q(u))$ denotes the quantile function of $F_S(F_S'^{-1})$, where $0\le u\le 1$. Then, the quantile-based Kullback–Leibler divergence is given by:

$$\xi(S_\theta,S_{\theta'})=\int_0^1\left(\log\frac{d}{dp}Q''(p)\right)dp \tag{18}$$

For more information regarding the derivation details in (18), one can refer to Sankaran et al. (2016).

Assuming that $E(X)$ of both distributions is finite, we can use the properties of the cumulative quantile function and derive the equivalent quantile-based Kullback–Leibler divergence in (18) as a function of the expected value of both distributions and the cumulative quantile function as follows:

$$\begin{aligned}\xi^*(S_\theta,S_{\theta'})&=\int_0^1(1-p)\left(\log\frac{1-p}{1-Q'^{-1}(Q(p))}\right)dQ(p)\\&\quad-\left[\int_0^1(1-p)dQ(p)-\int_0^1(1-p)dQ'(p)\right]\\&=\int_0^1(1-p)\left(\log\frac{1-p}{1-Q''(p)}\right)q(p)dp+E(S|\theta)-E(S|\theta')\end{aligned} \tag{19}$$

In the numerical example of this study, using (19), we measure the distance between the two pairs of candidate distributions for modeling headway data based on their quantile functions and show how significant the effect of advancing one distribution to another one is on the reliability of VANET and information propagation delay.

6. Numerical example

To illustrate the effect of uncertainty in model selection of headway data on the reliability of VANET based on information propagation delay, we use the Next-Generation Simulation (NGSIM) data to evaluate our proposed methodology and formulations. We selected data that includes detailed vehicle trajectory data on southbound US 101 (as an example for mixed flow) and Lankershim Blvd (as an example for congestion flow) in Los Angeles, CA, and Peachtree Street in Atlanta, Georgia (as an example for free flow). The data was collected through a network of synchronized digital video cameras.

The data set provides attributes that include vehicle length, width, class, acceleration, velocity, ID, frame ID, total frames, global time, local X and Y, global X and Y (based on the state plane), lane ID, preceding and following vehicles' ID, and time and space headway (both in the same lane). For this empirical data, the transmission range and road segment's length are set to $r=300$ and $L=500$ m, respectively. For data preparation, we also removed all observations with a time headway equal to 9999.99 (where vehicles move at zero speed). The NGSIM dataset provides the space headway for the proceeding vehicle in the same line; however, a vehicle can stay in an instantaneous transmission if another vehicle is in the other lanes with a distance shorter than r. Therefore, we utilize the sorted data based on the global time (regardless of the lane ID) and then calculate the space headway.

We monitor the space headway data until the condition (S_i+vehicle length $\le r$) fails and consider this a ferry transmission. Fig. 3 illustrates the histogram plot of S for three selected roads in this study with summary statistics of the ferry and instantaneous transmissions. In addition, since for each vehicle ID, there are multiple frames captured by sensors/cameras, we merged observations based on lane ID, vehicle ID, preceding, and following vehicles and considered the average time and space of merged observation for space headway and time headway, respectively, for each unique observation.

We expect to see the effect of uncertainty of the headway model in the data set with higher variation more significantly. Therefore, according to the histogram plot in Fig. 3 and the corresponding variance of each road equal to $(Var(\text{US101}),Var(\text{Lankershim}),Var(\text{Peachtree}))=(5054.475,3485.12,6147.207)$, we are expecting to see a higher level of entropy for US101, and Peachtree street in compare with Lankershim Blvd, validating this finding can result in a conclusion that congestion traffic flow (because it causes less variation in space headway) is less sensitive by the effects of model selection, and eventually mixed and free traffic flow (because they cause more variation in space headway) have a more sensitive behavior in this paradigm. In addition, both fatty tails (right and left) of Peachtree Street and Lankershim Blv histograms tell us that the effects of model selection in both tails quantiles will be high, while these effects are more obvious only for the lower tail quantiles of US101.

Following the data pre-processing step, calculating the theoretical entropy of $E(T)$ became straightforward. To do that, we separated data into two groups of $S\le 300$ for calculating $E(X)$ and $S>300$ for calculating $E(Y)$ in (3) and (4), respectively. However, the alternative equations of quantile-based entropy in (15) are utilized. We calculated $E(X)$ and $E(Y)$ for empirical data under different distribution settings (some of these distributions may not be a proper choice for our empirical data, but to show how serious the effect of models on the reliability of VANET we intentionally included them into our analysis).

Eventually, $E(X)$, $E(Y)$, $E(k)$, τ, and v_is are utilized to estimate $E(T)$ in (2). We considered P as a Bernoulli experiment that results in successful communication of one vehicle within the network as $\Pr(0\le S\le r)=\int_0^r f(s)ds$. For calculating wireless transmission time, τ, since the instantaneous communication usually happens in a highly congested traffic flow where the wireless connection between two vehicles happens constantly and is very small (microseconds) compared to ferry communication, we set τ as a very small and constant value (122 μs). Finally, v_i is calculated based on the headway space

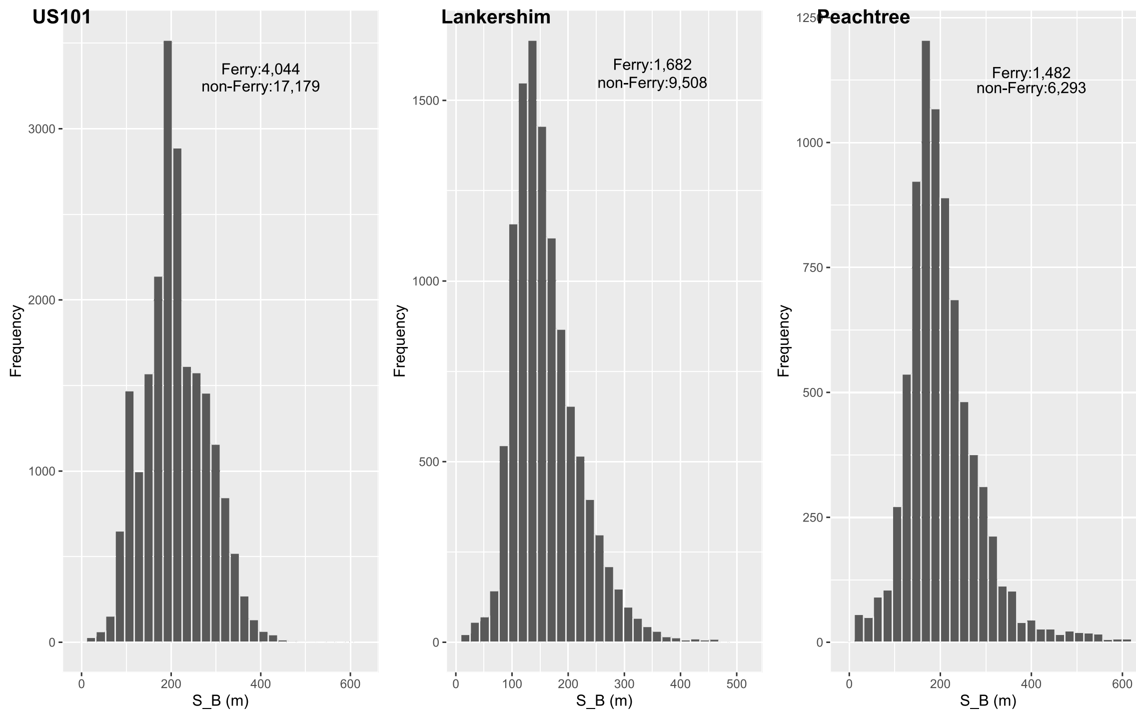

Fig. 3. Histogram plot of space headway data (S) for selected roads (legends show the variation of space headway for each traffic flow).

and time ratio. To estimate $E(X)$ and $E(Y)$ and show the effect of model uncertainty on the reliability of VANET, we considered four distributions, including Weibull, Normal, Lognormal, and Gamma distributions. The purpose behind selecting these distributions was only for illustration and did not relate to how well or poorly these distributions fit the headway data. Thereafter, several quantiles including $u = \{0.5, 0.75, 0.8, 0.9, 0.95, 0.975, 0.99, 0.999\}$ are considered and quantile-based entropy is only calculated for $\%u$ of the smallest value of the data.

The most common values of u for decision makers are 0.5 (known as median quantile), 0.25, 0.75 (known as lower and upper quartiles, respectively), and 0.95 and higher values (known as upper tail quantiles). The median quantile is pivotal for understanding the central tendency of the data. It provides insight into the typical or median delay within the traffic system, which is essential for general operational planning and performance assessment. The lower quartile is useful for identifying the lower bounds of headway or delays where the majority of the data points lie. This can be particularly relevant for setting minimum service standards or understanding the best-case operational scenarios. The upper quartile helps recognize the upper limits of the more typical range before entering the tail extremities. It aids in capacity planning and understanding regular congestion points and can be a marker for the onset of less frequent, more severe traffic conditions. Tail quantiles are crucial in assessing the risk of rare but impactful events, such as significant traffic delays. They are essential for preparing for potential worst-case scenarios, ensuring the resilience of traffic models against outliers, and complying with safety and reliability standards. Decision-makers utilize these quantiles to develop strategies for robust traffic management and to plan for emergencies.

We expect that when data are for lower quantiles (we only have instantaneous transmission), we see low entropy or receive high information gain by fitting selected distributions to headway data. The main reason for this phenomenon is the high frequency of data due to high congestion traffic flow. However, when we include the higher quantile information when traffic turns to be more sparse, and we have ferry transmission, due to the high volume of sparsity, the entropy increases, and therefore, no matter how the selected distribution is a good choice to fit the headway data, there would be a low-high information gain from the model for upper tail quantiles.

Table 4 summarizes the quantile-based entropy for each road and distribution and the estimated value of $E(T)$ under each scenario. As motioned earlier in Section 5, based on the relationship between the expected value and cumulative quantile function, we employed quantile-based entropy to estimate the $E(X)$ and $E(Y)$. Note that, in Table 4, one of the reasons that the value of $\xi(Q(u)|S > r)$ is very high compared with $\xi(Q(u)|S < r)$, is because we did not have enough observations that fit in the condition $S > r$. This includes only %12.5 of data from US101, %2.5 of Lankershim, and %10 of Peachtree roads.

The result in Table 4 validates our earlier assumptions in which the upper tail quantiles with ferry transmission in sparse traffic flow are more sensitive to model selection, and regardless of how distribution is well fitted with the majority of data (mainly in instantaneous transmission status), there is always a high level of uncertainty (high entropy) for ferry transmission data.

From Fig. 4, we can discern that, across all cases, the Lognormal distribution more accurately models the headway distribution and subsequent propagation delay calculation. The alignment of the Lognormal distribution with the empirical data is particularly notable, suggesting that it effectively captures the nuances of real-world traffic dynamics. Lower quantiles correspond to scenarios where vehicles are in close proximity (instantaneous transmission status), a common characteristic in right-skewed distributions. Fig. 3 shows that the empirical distribution for all the selected roads is indeed right-skewed, indicating a greater frequency of lower quantiles for short headway distances. Consequently, Fig. 4 shows that the expected time delay, $E(T)$, decreases as vehicles become closer (from lower quantiles, getting closer to the central quantiles in this figure). This trend intensifies towards the upper tail quantiles, where the headways are much longer (ferry transmission status). Such a pattern is expected and aligns with typical traffic behavior, where time delays are shorter for closely packed vehicles and increase as the distance between vehicles expands.

Table 4
The quantile-based entropy for each road and distribution under different quantile values, where (L, U) denotes lower and upper bounds of $E(k)$. Note: entropy is always between 0 and 1, and closer to zero means high information gain, and closer to 1 indicates low information gain.

Road $E(k)(L, U)$	Distribution	Terms	Quantiles							
			0.5	0.75	0.8	0.9	0.95	0.975	0.99	0.999
US101	Weibull	$\xi(Q(u)\|S<r)$	0.11339	0.12797	0.11339	0.09692	0.09092	0.08907	0.08815	0.08738
4.24		$\xi(Q(u)\|S>r)$	0.74844	0.77182	0.74844	0.71912	0.70769	0.7053	0.70332	0.7017
(2.107, 6.374)	Normal	$\xi(Q(u)\|S<r)$	0.1088	0.13648	0.1219	0.10543	0.09942	0.09758	0.09666	0.09589
		$\xi(Q(u)\|S>r)$	0.46157	0.51426	0.49088	0.46157	0.45013	0.44774	0.44576	0.44415
	Lognormal	$\xi(Q(u)\|S<r)$	0.22261	0.25028	0.23571	0.21924	0.21323	0.21139	0.21046	0.20969
		$\xi(Q(u)\|S>r)$	0.35286	0.40555	0.38217	0.38217	0.35286	0.34142	0.33705	0.33544
	Gamma	$\xi(Q(u)\|S<r)$	0.17682	0.1914	0.17682	0.16035	0.15434	0.1525	0.15158	0.15081
		$\xi(Q(u)\|S>r)$	0.41541	0.43879	0.41541	0.3861	0.37466	0.37227	0.37029	0.36868
Lankershim	Weibull	$\xi(Q(u)\|S<r)$	0.09932	0.11965	0.09932	0.08124	0.07428	0.07256	0.07125	0.07017
5.677		$\xi(Q(u)\|S>r)$	0.62	0.65659	0.62	0.6049	0.59779	0.59278	0.59278	0.59278
(3.893, 7.461)	Normal	$\xi(Q(u)\|S<r)$	0.08306	0.11921	0.09888	0.0808	0.07387	0.07212	0.0708	0.06973
		$\xi(Q(u)\|S>r)$	0.46719	0.51888	0.48229	0.46719	0.46008	0.45507	0.45507	0.45507
	Lognormal	$\xi(Q(u)\|S<r)$	0.09804	0.13419	0.11386	0.09579	0.08882	0.0871	0.08579	0.08471
		$\xi(Q(u)\|S>r)$	0.39036	0.44205	0.40546	0.39036	0.38326	0.37825	0.37825	0.37825
	Gamma	$\xi(Q(u)\|S<r)$	0.08256	0.10289	0.08256	0.06448	0.05752	0.0558	0.05449	0.05341
		$\xi(Q(u)\|S>r)$	0.42931	0.46589	0.42931	0.4142	0.4071	0.40209	0.40209	0.40209
Peachtree	Weibull	$\xi(Q(u)\|S<r)$	0.102	0.11534	0.102	0.07665	0.06996	0.068	0.06657	0.06657
1.303		$\xi(Q(u)\|S>r)$	0.7271	0.7271	0.7271	0.68835	0.67169	0.66657	0.66657	0.66243
(0,5.549)	Normal	$\xi(Q(u)\|S<r)$	0.06961	0.10574	0.09239	0.06704	0.06036	0.05839	0.05697	0.05697
		$\xi(Q(u)\|S>r)$	0.6046	0.64335	0.64335	0.6046	0.58795	0.58283	0.58283	0.57868
	Lognormal	$\xi(Q(u)\|S<r)$	0.27688	0.31295	0.29961	0.27426	0.26758	0.26561	0.26419	0.26419
		$\xi(Q(u)\|S>r)$	0.48662	0.52536	0.52536	0.48662	0.46996	0.46484	0.46484	0.4607
	Gamma	$\xi(Q(u)\|S<r)$	0.19009	0.20344	0.19009	0.16475	0.15806	0.1561	0.15467	0.15467
		$\xi(Q(u)\|S>r)$	0.56087	0.56087	0.56087	0.52212	0.50546	0.50034	0.50034	0.4962

To show the effect of model selection on $E(T)$ and the reliability of VANET, instead of directly using distributions, we utilized the quantile functions to estimate $E(S|S<r)$ and $E(S|S>r)$ for the calculation of $E(T)$. While estimating $E(T)$ based on quantile functions, we considered the same quantile information for calculating the average and relative speeds and $E(k)$. Our primary analysis found that regardless of the speed, $E(X)$ and $E(Y)$ always have the same patterns. In addition, according to our findings, although speed and $E(k)$ are values that can be changed under different quantiles, even if they remain fixed for all quantiles and equal to estimated values for complete information, still $E(T)$ remains sensitive to the model.

Considering the results provided in Fig. 3, our assessments at the beginning of this section are validated by the result in Fig. 1, which means traffic flows with higher variations are more sensitive to the effects of model selection, and these effects in tail quantiles, when there is a very large (ferry transmission) or very short distance (during congestion traffic flow) among vehicles, are more significant. In conclusion, the reliability of VANET is questionable due to information propagation delays in very crowded or sparse traffic. Besides, Fig. 1 shows that having more hops (during congested traffic flow) on the road does not guarantee a low information propagation delay (see lower quantiles in Fig. 2).

The results of our calculation for $E(T)$, can be effectively contrasted with the findings from Du and Dao (2014) to highlight the significance of reliability assessment in VANET connectivity. Du and Dao (2014) conducted a detailed analysis of similar datasets, focusing on various vehicle scenarios, road types (one-way and two-way), and traffic conditions (congested and free-flow). They provided average time delay calculations for each road type and traffic condition, offering a comprehensive perspective on delay variations across different traffic scenarios. Their result is as follows:

- $E(T)$ for one-way segment with congested flow: 42.07 s
- $E(T)$ for one-way segment with free flow: 46.4 s
- $E(T)$ for two-way segment with congested flow: 25.4 s
- $E(T)$ for two-way segment with free flow: 17.8 s

Our approach and findings in this study are alleged closely within the context of one-way road segmentation results presented by Du and Dao (2014). In our analysis, the concepts of free flow and congested flow are analogous to the tail and central quantiles, respectively. It is notable that for different road scenarios, the central quantiles' $E(T)$ in our study closely matches the 42 s average delay reported by Du and Dao (2014) (see Fig. 4). However, for free-flow conditions, which correspond to tail quantiles in our research, our estimated $E(T)$ is closer to 80 s. This discrepancy highlights the importance of incorporating reliability aspects into analysis, as our approach reveals a higher delay. This indicates a significant bias in earlier estimations, underscoring the effectiveness of our methodology in capturing the more realistic traffic dynamics, especially in decision-making contexts where higher delays are critical.

According to the simulation result in Fig. 1, the Lognormal distribution always performs better than Normal, Gamma, and Weibull distributions, specifically not under the very end tails, which are the most common traffic flows in all selected roads. Therefore, the proper selection of models is the most critical decision in the reliability assessment of VANET. To show how significant the effect of misspecifying the distribution would be on information propagation delay, we consider the scenario where the better distribution is Lognormal but fitted by the Weibull distribution. We considered the quantile-based function for two competitive distributions in (19) and defined two metrics as relative bias (RB) and relative variability (RV) for measuring the effects of misspecification on information propagation delay and reliability of VANET as follows:

$$RB = \frac{E(T|\hat{\theta}) - E(T|\hat{\theta}')}{E(T|\hat{\theta})} = \frac{E(T|\xi^*(S_\theta, S_{\theta'}))}{E(T|\xi(S_\theta))} \tag{20}$$

$$RV = \frac{E(T|\xi^*(S_\theta, S_{\theta'}))}{E(T|\xi^*(S_{\theta'}, S_\theta))}. \tag{21}$$

The results of RB and RV for different quantile functions in estimating $E(T)$ are depicted in Figs. 5 and 6, respectively. From the figures, theoretically speaking, if the better distribution is Lognormal and the fitted distribution is Weibull, by comparing the results, there could be about 3.5%, 3%, and 1.5% of bias in roads with mixed congestion and free traffic flows, respectively, when vehicles have low distance. These effects are about 2.5%, 2.5%, and 1% for similar traffic flows when vehicles have higher distances. Although these numbers may not look very critical and strongly depend on road and traffic conditions,

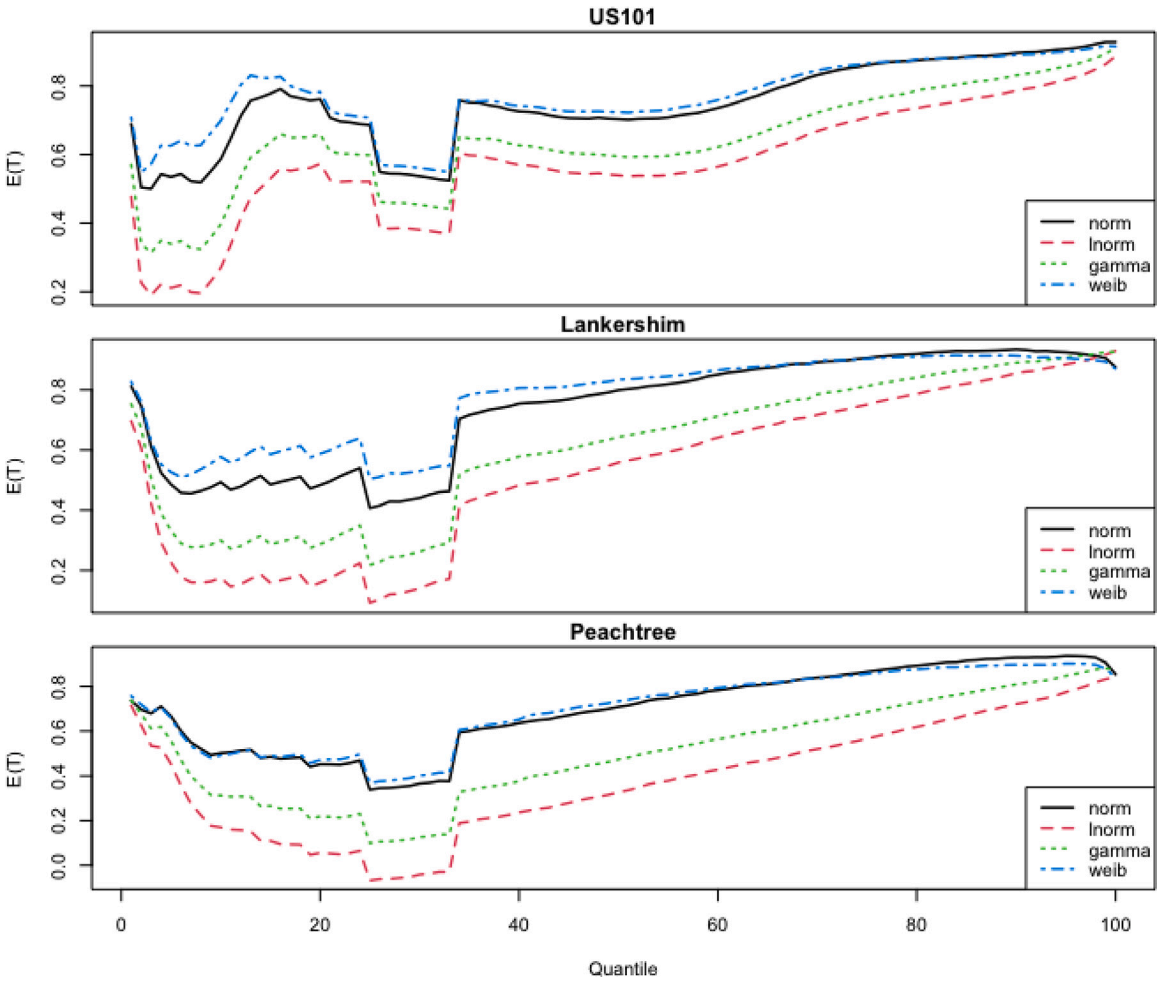

Fig. 4. Estimated $E(T)$ based on quantile-based entropy for each percentile under different distribution and road settings. Note: $E(T)$ is in 10 s.

imagine a low-speed/low-distance traffic flow that automatically leads to a high information propagation delay; if drivers drive at speeds as low as 20 km/h, there would be about 42 s of bias in estimating the $E(T)$, if instead of a better distribution like Lognormal, a weaker distribution like Weibull fit the X and Y data. During 42 s at a speed of only 20 km/h, the vehicle can pass 2.5 m, while the reaction and braking distance are only 6 m at such a speed (see Fig. 5). [1]

Fig. 6 depicts the result for the RV of each road. The result confirms our findings from analysis RB for tail quantiles; in addition, these plots tell us that for US101 with a mixed traffic flow, Weibull is always underestimating the property of $E(T)$ (RV is always smaller than 1), while for the case of Lankershim Blvd., Weibull is usually overestimating the properties of $E(T)$. A similar conclusion with US101 can be withdrawn for Peachtree Street. Generally, we can conclude that information propagation delay on roads with mixed and free traffic flow can be underestimated and overestimated in highly congested traffic flow. Therefore, it can be concluded that if the space headway distribution is not properly selected, the underestimation of information propagation delay will cause accidents in the mix and even free traffic flow roads.

7. Conclusion

Headway modeling has attracted many researchers during the past decades. In this paper, we developed a novel methodology for studying the effect of model uncertainty on information propagation delays in the reliability of V2V communication networks, which may be beneficial for vehicle safety, traffic control estimation, and prediction. We evaluate the effects of model selection of space headway data when different distributions fit the data and calculate the quantile-based entropy to estimate the information propagation delay in a V2V communication network. Under different models, the behavior of $E(T)$ as the expected time delay value is calculated for different quantile values.

The proposed model for considering the effect of model selection on information propagation delay in the reliability assessment of IVC networks can be implemented to quantify propagation and coverage of time delay information for local transportation networks. Also, CAV technologies open new issues for exploring digital data-sharing opportunities. One of the future challenges will be gathering different sources of information about vehicle headway distribution. Moreover, this study examines the time delay caused by instantaneous transmission, which is small in a single road segment but significant as information spreads across a large network. Extending to the network level would be a potential research area. The proposed analytical formulations, which capture the time delay of information propagating over a road segment, serve as a solid foundation for further research into these complex areas under the effects of proper model selection and network reliability assessment.

In addition, it is worth mentioning that when the size of the headway data is large enough because estimators are approximately unbiased, bias is not an issue for practitioners. However, to reduce the bias and inefficiency of model selection when the fitted model turns

[1] https://mobilityblog.tuv.com/en/calculating-stopping-distance-braking-is-not-a-matter-of-luck/.

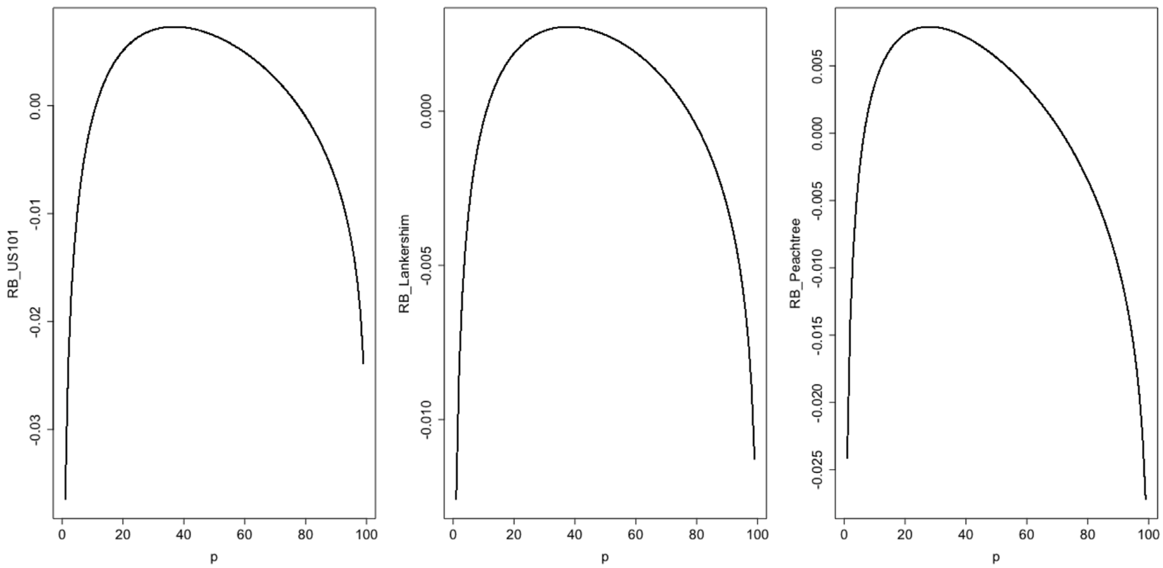

Fig. 5. RB of estimating $E(T)$ based on quantile-based entropy between two distributions, when the true distributions are Lognormal but mistreated as Weibull.

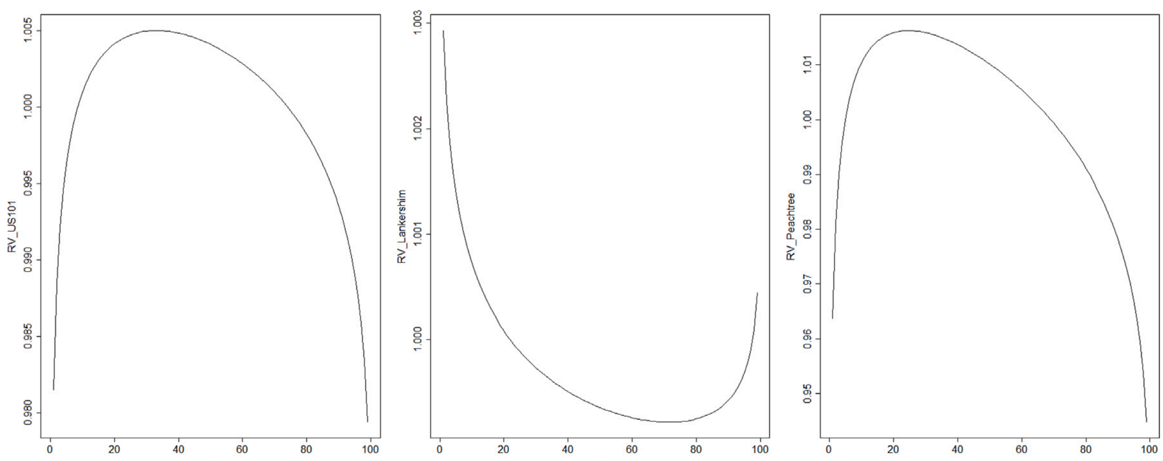

Fig. 6. RV of estimating $E(T)$ based on quantile-based entropy between two distributions, when the true distributions are Lognormal but mistreated as Weibull.

out to be incorrect, defining a robust test plan, such as our proposed entropy-based goodness of fit test, is necessary. In practical situations, there are cases where mixture distribution is the true distribution, but an unimodal distribution incorrectly fits the data; for instance, separated distributions for ferry and instantaneous transmissions are recommended. Therefore, defining a correct mixture model or stochastic process that can describe the dynamic behavior or data is worth investigating.

Finally, using the proposed model to detect, forecast, and manage freeway congestion based on a better estimate of headway distribution and data availability would improve the efficiency and reliability of AI-based traffic congestion detection, forecasting (short, medium, and long term), and freeway management.

CRediT authorship contribution statement

Marzieh Khakifirooz: Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Michel Fathi:** Conceptualization, Validation, Writing – original draft. **Lili Du:** Conceptualization, Methodology, Supervision.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

Marzieh Khakifirooz acknowledges funding from Tecnologico de Monterrey's OPEX program. Michel FATHI's research was supported by the University of North Texas College of Business Summer Research Grant. The research of Lili Du was supported by NSF grants CMMI 2213459 and 1817346.

References

Abuelenin, S.M., Abul-Magd, A.Y., 2015. Moment analysis of highway-traffic clearance distribution. IEEE Trans. Intell. Transp. Syst. 16 (5), 2543–2550.

Alsuhaim, A., Rayamajhi, A., Westall, J., Martin, J., 2021. Adapting time headway in cooperative adaptive cruise control to network reliability. IEEE Trans. Veh. Technol. 70 (12), 12691–12702.

Bayarri, M., Berger, J.O., Molina, G., Rouphail, N.M., Sacks, J., 2004. Assessing uncertainties in traffic simulation: A key component in model calibration and validation. Transp. Res. Rec. 1876 (1), 32–40.

Biswas, R.K., Friswell, R., Olivier, J., Williamson, A., Senserrick, T., 2021. A systematic review of definitions of motor vehicle headways in driver behaviour and performance studies. Transp. Res. F 77, 38–54.

Blumenfeld, D., Weiss, G., 1975. Effects of headway distributions on second order properties of traffic noise. J. Sound Vib. 41 (1), 93–102.

Copas, J., Eguchi, S., 2020. Strong model dependence in statistical analysis: goodness of fit is not enough for model choice. Ann. Inst. Statist. Math. 72 (2), 329–352.

Du, L., Dao, H., 2014. Information dissemination delay in vehicle-to-vehicle communication networks in a traffic stream. IEEE Trans. Intell. Transp. Syst. 16 (1), 66–80.

Du, L., Gong, S., Wang, L., Li, X.-Y., 2016. Information-traffic coupled cell transmission model for information spreading dynamics over vehicular ad hoc network on road segments. Transp. Res. C 73, 30–48.

Fathi, M., Khakifirooz, M., Pardalos, P.M., 2019. Optimization in Large Scale Problems: Industry 4.0 and Society 5.0 Applications, Volume 152.

Hoogendoorn, S.P., Bovy, P.H., 1998. New estimation technique for vehicle-type-specific headway distributions. Transp. Res. Rec. 1646 (1), 18–28.

Hung, W., Tian, F., Tong, H., 2003. Discharge headway at signalized intersections in Hong Kong. J. Adv. Transp. 37 (1), 105–117.

Hussain, S., Wu, D., Wang, X., Memon, S., Bux, N.K., Saleem, A., 2019. Reliability and connectivity analysis of vehicular ad hoc networks for a highway tunnel. Int. J. Adv. Comput. Sci. Appl. 10 (4).

Jamili, A., Aghaee, M.P., 2015. Robust stop-skipping patterns in urban railway operations under traffic alteration situation. Transp. Res. C 61, 63–74.

Jin, X., Zhang, Y., Wang, F., Li, L., Yao, D., Su, Y., Wei, Z., 2009. Departure headways at signalized intersections: A log-normal distribution model approach. Transp. Res. C 17 (3), 318–327.

Kong, D., Guo, X., 2016. Analysis of vehicle headway distribution on multi-lane freeway considering car–truck interaction. Adv. Mech. Eng. 8 (4), 1687814016646673.

Li, L., Chen, X.M., 2017. Vehicle headway modeling and its inferences in macroscopic/microscopic traffic flow theory: A survey. Transp. Res. C 76, 170–188.

Li, T., Hui, F., Liu, C., Zhao, X., Khattak, A.J., 2020. Analysis of V2V messages for car-following behavior with the traffic jerk effect. J. Adv. Transp. 2020.

Li, X., Zhang, B., Liu, Y., 2021b. A little bit flexibility on headway distribution is enough: Data-driven optimization of subway regenerative energy. Inform. Sci. 554, 276–296.

Li, H., Zhou, Y., Xing, Y., Zhang, C., Zhang, X., 2021a. An improved mixed distribution model of time headway for urban roads based on a new traffic state classification method. IEEE Access 9, 12635–12647.

Morales, D., Muñoz, J.C., Gazmuri, P., 2020. A stochastic model for bus injection in an unscheduled public transport service. Transp. Res. C 113, 277–292.

Qian, G., Guo, M., Zhang, L., Wang, Y., Hu, S., Wang, D., 2021. Traffic scheduling and control in fully connected and automated networks. Transp. Res. C 126, 103011.

Roy, R., Saha, P., 2018. Headway distribution models of two-lane roads under mixed traffic conditions: a case study from india. Eur. Transp. Res. Rev. 10 (1), 1–12.

Saito, H., 2006. Performance analysis of combined vehicular communication. IEICE Trans. Commun. 89 (5), 1486–1494.

Sankaran, P., Sunoj, S., Nair, N.U., 2016. Kullback–Leibler divergence: A quantile approach. Statist. Probab. Lett. 111, 72–79.

Singh, H., Kathuria, A., 2021. Analyzing driver behavior under naturalistic driving conditions: A review. Accid. Anal. Prev. 150, 105908.

Tan, J., Qin, X., Gong, L., 2020. Using vehicle-to-vehicle communication to improve traffic safety in sand-dust environment. Int. J. Environ. Res. Public Health 17 (4), 1165.

Toledo, T., Cats, O., Burghout, W., Koutsopoulos, H.N., 2010. Mesoscopic simulation for transit operations. Transp. Res. C 18 (6), 896–908.

Ukkusuri, S., Du, L., 2008. Geometric connectivity of vehicular ad hoc networks: Analytical characterization. Transp. Res. C 16 (5), 615–634.

Wang, B.X., Adams, T.M., Jin, W., Meng, Q., 2010. The process of information propagation in a traffic stream with a general vehicle headway: A revisit. Transp. Res. C 18 (3), 367–375.

Wang, D., Wang, Z., Chen, M., Wang, W., 2018a. Distributed optimization for multi-agent systems with constraints set and communication time-delay over a directed graph. Inform. Sci. 438, 1–14.

Wang, X.B., Yin, K., Yan, X., 2015. Vehicle-to-vehicle connectivity on parallel roadways with large road separation. Transp. Res. C 52, 93–101.

Wang, T., Zhao, J., Li, P., 2018b. An extended car-following model at un-signalized intersections under V2V communication environment. PLoS One 13 (2), e0192787.

Weber, J.S., Neves, M., Ferreto, T., 2021. VANET simulators: an updated review. J. Braz. Comput. Soc. 27 (1), 1–31.

Weng, J., Gan, X., Du, G., 2019. Random coefficient models for work zone headway distribution. J. Transp. Eng. A 145 (10), 04019042.

Weng, J., Meng, Q., Fwa, T.F., 2014. Vehicle headway distribution in work zones. Transp. A 10 (4), 285–303.

Yang, X., Recker, W., 2005. Simulation studies of information propagation in a self-organizing distributed traffic information system. Transp. Res. C 13 (5–6), 370–390.

Yin, S., Li, Z., Zhang, Y., Yao, D., Su, Y., Li, L., 2009. Headway distribution modeling with regard to traffic status. In: 2009 IEEE Intelligent Vehicles Symposium. IEEE, pp. 1057–1062.

Yin, K., Wang, X.B., Zhang, Y., 2013. Vehicle-to-vehicle connectivity on two parallel roadways with a general headway distribution. Transp. Res. C 29, 84–96.

Zarei, M., Rahmani, A.M., 2016. Renewal process of information propagation in delay tolerant VANETs. Wirel. Pers. Commun. 89 (4), 1045–1063.

Zhang, G., Wang, Y., Wei, H., Chen, Y., 2007. Examining headway distribution models with urban freeway loop event data. Transp. Res. Rec. 1999 (1), 141–149.

Zhou, J., Zhu, F., 2020. Modeling the fundamental diagram of mixed human-driven and connected automated vehicles. Transp. Res. C 115, 102614.

Zou, Y., Zhang, Y., Zhu, X., 2014. Constructing a bivariate distribution for freeway speed and headway data. Transp. A 10 (3), 255–272.