Vehicle-type Specific Headway Distribution in Freeway Work Zone: A Nonparametric Approach

Shangjia Dong
Ph.D Student
School of Civil and Construction Engineering
Oregon State University
Corvallis, OR, 97331
Phone: +1 541-737-XXXX
Fax: +1 541-737-3052
Email: shanga.dong@gmail.com

Haizhong Wang (Corresponding Author)
Assistant Professor, Ph.D.
School of Civil & Construction Engineering
Oregon State University
220 Owen Hall
Corvallis, OR 97331
Phone: +1 541-737-8874
Fax: +1 541-737-3052
Email: haizhong.wang@oregonstate.edu

David Hurwitz
Assistant Professor, Ph.D.
School of Civil & Construction Engineering
Oregon State University
220 Owen Hall
Corvallis, OR 97331
Phone: +1 541-737-9242
Fax: +1 541-737-3052
Email: david.hurwitz@oregonstate.edu

Submitted to TRB 93th Annual Meeting at Washington D. C. Jan. 2014
Number of words:
Number of figures and tables: n × 250
Total:

Abstract

Headway is significant to the traffic flow control, and many researches are conducted on this topic. Previous work mostly focused on the parametric models, which based on certain assumptions, thus its reliability remain discussing. This paper employs the nonparametric distribution model with Gaussian kernel functions to investigate the data work zone. Without any assumptions, Gaussian kernel model is capable to catch the intrinsic features from empirical headway data for depicting the headway distribution. The nonparametric model would be more applicable and desirable in various scenarios. Also, we aim on the vehicle type-specified model: car-car, car-van, car-truck, van-car, van-van, van-truck, truck-car, truck-van, and truck-truck. The K-S test confirmed the good performance of the nonparametric model, all K-S statistics and hypothesis test indicate that nonparametric model with Gaussian kernel-based model is better than parametric model with lognormal distribution. Experiments were further conducted on the nine types of headway to provide a visual evidence. The Gaussian kernel model shows very good capability in describing the probability density function and cumulative density function, the relative error is also small and limit to 0. The lognormal distribution indicate a good fit in approximate headway distribution, we use lognormal distribution to compare with Gaussian kernel model, results shows Gaussian kernel model performs better in approximating and the relative error is steady and small while the lognormal distribution has a big fluctuation in the beginning. All the results find that nonparametric distribution model with Gaussian kernel functions has a better goodness-of-fit in type-specified work zone scenario.

1 Introduction

1.1 Background

Vehicle time headway, defined as the elapsed time in seconds between the arrival of the leading and following vehicle at an observation point, is a measure of space between two successive vehicles. The modelling of vehicle time headway distribution is essential to many

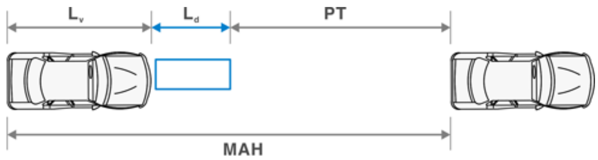

FIGURE 1 Definition of Vehicle Time Headway

aspects of traffic flow fundamental analysis, for example, capacity estimation, microscopic simulation, and safety analysis (i.e., time-to-collision) [1]. The fundamentality and significance to microscopic traffic flow modelling and simulation has led to many investigations on this topic.

For traffic simulation models, a key component of determining the simulation model performance is vehicle inter-arrival times. Researchers, therefore, devote considerable efforts on the accurate headway distribution models [10]. In addition, time headway could be considered as the reciprocal of flow rate [7]. Under certain circumstances, vehicle time headway can be used to estimate the road capacity. An accurate headway distribution would help engineer to maximize the road capacity and minimize the vehicle delays [7]. Additionally, as it is related to vehicle merging and lane-changing behaviour, it is essential to estimate road capacity or adjust the signal control parameters at signalized intersections. Furthermore, it is also related to traffic safety, driver behaviour, and traffic flow theory. Since vehicle time headway modelling has important ramifications for applications [6], ranging from traffic control to the safety issue. Hence, an inspection on headway distributions is essential and important.

Much studies have been conducted on the headway distribution model. Zhang [12] proposed a nonparametric model with Gaussian kernel method. He investigated the freeway scenario without specifying the vehicle type. The results shows very well and better than the studied parametric model. Weng [11] studied vehicle type-specific headway distribution in work zone. He showed the most fitted distribution in every vehicle type. Our paper innovatively proposed a study on the work zone distribution using nonparametric method in [12]. Better than Zhang's model, we did a specification on vehicle type which showed very necessary in headway study, and we investigated the work zone scenario. Different from the Weng's study, we adopt the nonparametric model and conducted a new specification on the vehicle type. Parametric methods share a common feature, that provide some methods first and use empirical data to confirm which is better. The crux is our limited recognition on the distribution. And parametric methods often established on certain assumptions. Nonparametric model inherent a flexible form and less assumption, demonstrate a better performance than the parametric model [12] make us believe it would be the best method, and it's better in extracting the statistical features inherent in the data, which is the reason we choose the nonparametric model.

1.2 Objective of Paper

The objective of this paper is to present a vehicle-type specific headway distribution study through a non-parametric approach.

1.3 Paper organization

The remainder of this paper is organized as follows. Theory and methodology were introduced in the section 3. Section 4 described study area, data and experiment design. Then the statistical test and analysis were conducted in section 5 and visual performance and results analysis were presented in section 2. A discussion on the motivation of methods and the results were exhibited in section 7. At last, the conclusion of the paper was drawn in section 8, and acknowledgement was made in section **??**.

2 Literature Review

Many headway distribution models have been derived and calibrated using empirical traffic data. In general, these models could be categorized into two groups [7]: single statistical distribution models and mixed models.

2.1 Single Distribution Models

Representatives of the single statistical distribution family include normal distribution, log-normal distribution [15], Weibull distribution [14], the Erlang distribution, exponential distribution, log-logistic distribution, Cowan's M3 and M4 distribution [13], inverse gaussian distribution and Gamma distribution etc. For instance, Sun and Benekoal [14] used Weibull distribution model to describe the vehicle headways in work zone. Jang et al. [10] [17] examined that Johnson SU distribution, together with Johnson SB distribution and lognormal distribution are transformations of a normal distribution, which can be employed to depict most naturally occurring uni-modal sets of data. Jin et al. [9] studied the departure headways and indicated that the distribution headways in a queue approximate a certain log-normal distribution. Al-Ghamdi [19] recommended four headway distribution models at different flow rates, such as negative exponential distribution for the low flow rate, shifted exponential and gamma distribution for the middle flow rate, and Erlang distribution for the high flow rate. Riccardo's et al. analyzed case study on rural two-lane two-way roads [3] suggested that inverse Weibull distribution best fits the headways observed for the most of situations, better performed than the log logistic, person 5 and person 6, regardless of flow rate range. Yin's et al. [4] studied the dependence of headway distribution on traffic status and showed that log-normal distribution is adequate to fit headway when that traffic is in free flow state, and log-logistical distribution is suitable in congestion state. Serge and Hein [5] presented and Branston's generalized queueing model for headway distribution and a new estimation method is proposed.

2.2 Mixed Distribution Models

Many of the stationary distributions can fit the empirical data of free flow but not the congested flow, and their performances are not satisfactory. Then mixed headway distribution models were introduced to better capture the headway distribution characteristics. The representatives of mixed models include double displaced negative exponential distribution (DDNED) [16], normal distribution + shifted negative exponential distribution [8], negative exponential distribution + shifted negative exponential distribution [8] Generalized Queuing Model (GQM) [7] and Semi-Poisson distribution etc. For instance, Zhang et al. [7] found that double displaced negative exponential distribution (DDNED) and lognormal distribution best fit the high occupancy vehicle (HOV) lane and regular lanes. In a vehicle type-specific but car dominant case, Ye et al. [8] proved that normal distribution + shifted negative exponential distribution could not fit the data well, while negative exponential distribution + shifted negative exponential distribution fitted very well [8]. Some mixed distribution were developed based on the assumption that a headway H consists of two components, $H = T + U$, where T is the "tracking or following" component and U is the "free" component [7], according to this, many important models are derived such as Cowan M1-M4, the Generalized Queuing Model [18], and Semi-Poisson model. Among these, Cowan's M3 model are widely investigated and applied for its simplicity and easy approximation in describing longer headways [12]. Because the explicit expression for the Laplace transform of the following-vehicle headway distribution is required, the use of Semi-Poisson model has been limited [7]. In [7], Cowan M3, Cowan M4 and GQM, DDNED etc. have been analysed.

2.3 Vehicle-type Specific Headway Distribution

Considering the empirical traffic compositions, researchers started exploring the impact of vehicle-types on the headways. Ye and Zhang [8] categorized headways into four types according to different combination of vehicle types (leader-follower pairs). They adopt three distribution models for the four headway types: the shifted negative exponential distribution for truck-car and truck-truck types, the Erlang distribution for the car-truck type, and a composite model for the car-car type [11]. Weng et al. [11] conducted the test and concluded that headways are strongly related to the types of the leading and following vehicles. The results show that the investigation of the headway by four types is reasonable [11]. In the examination, they found that lognormal best fit the Car-Car headway type, as well as the Car-Truck headway, and inverse Gaussian distribution is best for the Truck-Car and Truck-Truck headway [11]. They also concluded that four factors: traffic flow rate, percentage of trucks, lane position and intensity of work zone activity, may influence the location and scale of a headway distribution model [11].

Although single distribution models are simple and easy to apply, they are typically inadequate when approximating shorter vehicle time headways. However, mixed distribution are more flexible to describe headways than single distribution, but the calibration process in general is challenging, and the parameter estimation is difficult as well due to the complicated structures of the probability density functions [7].

TABLE 1 A summary of the vehicle headway distribution study

Headway Study	Mixed	Vehicle-type Specific	Parametric	Non-parametric	Scenario
Zwahlen (1999)	√		√		Freeway
Yin (2009) [4]	√		√		Urban roadways
Ye (2009) [8]	√	√	√		Freeway
Jang (2011) [10]	√		√		Suburban Arterial
Riccardo (2012) [3]			√		Rural two-lane two-way road
Zhang (2013) [12]	√			√	Freeway
Weng (2013) [11]	√	√	√		Work Zone

2.4 Parametric vs. Nonparametric

Despite many distribution models have been investigated and applied, all these models are parametric and assume that the headway follows a known distribution or a composite distribution. The existing parametric approaches, no matter what functional form of the distribution is, share a common nature. It starts with a bold assumption the headway follow some types of distributions and then check with empirical data. Whichever fits the empirical distribution the best is favored. Therefore, the result sometimes is inaccurate because of our limited capacity to guess the known distribution with the inevitable mis-perception. Furthermore, parametric headway model require strict prior knowledge and certain conditions which are difficult to meet. Although these parametric models are simple and intuitive to understand, the goodness-of-fit for these model varies with the location and traffic flow level [12]. Essentially, deterministic models could not properly account for the stochastic nature of the variables or the transient nature of the traffic [2]. Mixed models typically fit real situation better, but as the cost of complex conformation and calibration. While nonparametric model could work better due to their flexibility in forms and ability of extracting statistical features of observed headways without referring to assumed distribution models with specific parameters [12]. As [12] stated, the nonparametric Gaussian kernel headway models outperform the traditional parametric model because of the flexible data modeling ability, requires few stringent hypotheses and can sufficiently handle subtle and complicated interactions among vehicles, and do not rely on the assumption that the data drawn from a particular distribution. So its applicability and compatibility is much wider than the traditional parametric methods. Also the test showed the model is independent to specific sample data and could be generalized to suit different sample data under similar traffic scenario.

3 Theory/Methodology

Most tests were conducted based on the data that are collected from freeway or HOV, compared with uninterrupted traffic, work zone traffic has unique characteristics [11]. We would study the nonparametric headway distribution based on work zone vehicle date. In headway distribution modeling, two goodness-of-fit tests are used to judge how well a distribution fits the sample data: the Chi-square test and Kolmogorov-Smirnov (K-S) test [11]. And we would adopt K-S tests to determine the goodness-of-fit in the work zone traffic scenario. In most case, parametric methods were adopted to measure headway distribution, and lognormal distribution and some certain distribution always demonstrates a better performance. In addition, with many research conducted on the headway distribution study, little study focus on the nonparametric model. Compare to the parametric method, nonparametric methods exhibit a good capability in capture the intrinsic character. A Gaussian Kernel-Based approach in modeling headway distribution based on the freeway headway data [12] shows a good result. We could employ this method to have an examination on the work zone data. If possible, the influence of the car type to the headway could be considered, and the headway mode could be classified into specific groups as the [11] [8] conducted.

3.1 Nonparametric Model

The estimated PDF of the Gaussian kernel model is calculated as follows[12]:

$$f\left(x\right)=\frac{1}{nh}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-X_i}{h}\right)^2} \tag{1}$$

where X_i is an individual headway measured. The h was computed by

$$h=1.06\delta n^{-1/5} \tag{2}$$

where δ is the standard deviation of the data set and n is the size of the data set.

This equation indicates that a high value of the PDF could be get when more sample points closely aggregate at a certain point. And as a linear composition of Gaussian kernels, the PDF have differentiable and continuous characteristics derived from the kernels [12] could strengthen the smoothness of the density curve.

3.2 Parametric Models

In this paper, we use the better performed lognormal distribution to compare with the Gaussian kernel function. The model are shown as follows:

$$f\left(x\right)=\frac{e^{-\frac{1}{2}\left(\frac{lnx-\mu'}{\delta'}\right)^2}}{x\sqrt{2\pi}\delta'} \tag{3}$$

with

$$\mu'=ln\frac{\mu^2}{\sqrt{\delta^2+\mu^2}} \tag{4}$$

$$\delta'=\sqrt{ln[1+\left(\frac{\delta}{\mu}\right)^2]} \tag{5}$$

where μ is the mean of the data set, and δ^2 is the variance. The lognormal distribution perform good in many investigations on the headway distribution model.

4 Experimental section

4.1 Study Area Description

The work zone site is on I-91 in Massachusetts, the data was collected in 2005. This section is easy to take care of. :

4.2 Data Description

The work zone traffic flow data was collected over a week long time horizon.

4.3 Experiment Design

The data was collected in 2005 on I-91 in Massachusetts. Based on the FHWA Vehicle Classification scheme, we classify the vehicles into three types: Car, Van, and Truck. Then the traffic flow consists of nine types: Car-Car, Car-Van, Car-Truck, Van-Car, Van-Van, Van-Truck, Truck- Car, Truck-Van, and Truck-Truck. According to the time-line of the day, separate the day into four periods: morning peak, off-noon, afternoon peak and evening. Without losing generality, experiments based on the specific flow types in given time periods are illustrated as follows. In order to examine the goodness-of-fit, we employ the K-S test to provide the statistical evidence, and compare Gauss kernel model with lognormal distribution which tested as the best distribution in Car-Car and Car-Truck type in work zone[11]. Also, We selected some of them to make a visualized performance comparison.

5 Statistical Analysis: Probability Metric

In order to give an numerical expression instead of only visual performance, we conducted a statistical test to investigated the goodness-of-fit of the nonparametric method. There are two often-used measuring criteria to examine the goodness-of-fit, Chi-square test and Kolmogorov-Smirnov (K-S) test. The Chi-square test, however, is too strict that a model would be thrown off with only a few "bad" fits[12]. In this study, we adopt K-S test to measure goodness-of-fit of the selected nonparametric method. The K-S test is a form of minimum distance estimation used as a nonparametric test of equality of one-dimensional probability distribution used to compare a sample with a reference probability distribution (one-sample K-S test), or to compare two samples (two-sample K-S test). The Kolmogorov-Smirnov statistic quantifies the distance between empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples.

5.1 One-sample K-S Test

The one-sample K-S test is defined as

$$D_n = \sup_x |F_n(x) - F(x)| \tag{6}$$

where $F_n(x)$ denotes the empirical distribution function, and $F(x)$ denotes the proposed cumulative distribution function. The term D_n is the maximum vertical distance between $F_n(x)$ and $F(x)$, and n is the sample size. Two sample K-S test evaluates the difference between the CDF of the distribution of two sample data.

5.2 Two-sample K-S Test

The Kolmogorov-Smirnov test (KS test) is usually used to obtain a probability of similarity between two distributions to determine whether two datasets differ significantly. The KS-test is non-parametric and distribution free meaning that it has the advantage of making no assumption about the distribution of data. The mechanism behind this test is to obtain the cumulative distribution function of the two distributions that needs to be compared. The Kolmogorov-Smirnov distance (KS distance) is a simple measure which is defined as the maximum value of the absolute difference between two cumulative distribution functions. Kolmogorov-Smirnov distance measures the largest absolute difference between two distribution functions $F(t)$ and $G(t)$ for varying t. In the similar setting, the Kolmogorov-Smirnov distance is defined by

$$\rho_K(X,Y) := ||F-G||_\infty = \sup_{t\subset\mathbb{R}} |P(X \leq t) - P(Y \leq t)| = \sup_t |F(t) - G(t)| \tag{7}$$

The supremum is the least upper bound of a set. Given a sample of observations $x = (x_1, \ldots, x_n)$, the empirical distribution function F_n is given by the following expression

$$F_n(t) = \frac{1}{n}\#\{x_i|x_i \leq t\} \tag{8}$$

Where $\#\{\ldots\}$ denotes the number of elements contained in the set $\{\ldots\}$ and F_n defines a discrete probability distribution function on the real line and for large values of n the empirical distribution converges to the theoretical one.

In our study, we adopt the two-sample K-S test to measure the goodness-of-fit. A smaller K-S statistic value indicates a better goodness-of-fit, and in two-sample K-S test, the decision to reject the null hypothesis is based on comparing the $p - value$ with significance level α. The comparison of K-S statistics and hypothesis test are illustrated in Table 1, the null hypothesis is that two sample of data are generated from the same distribution and the significant confidence is 95%.

After both nonparametric model and parametric model were employed for each data set, their overall performance was examined and evaluated. The comparisons were illustrated in Table 2. The hypothesis test is conducted at 95% significant confidence. From the table, we noted that nonparametric model performs better than the parametric model in most of scenarios. This indicate that nonparametric method is better in describing headway model.

TABLE 2 Statistical test results comparison between Nonparametric model and Parametric model

		Nonparametric model with Gaussian Kernel function		Parametric model Lognormal distribution	
	Period	K-S statistic	Hypothesis test	K-S statistic	Hypothesis test
Car-Car	Morning	0.0714	not reject	0.1327	not reject
	Off-noon	0.0166	not reject	0.2500	reject
	Afternoon	0.0400	not reject	0.2533	reject
	Evening	0.0667	not reject	0.1704	reject
Car-Van	Morning	0.1695	not reject	0.2881	reject
	Off-noon	0.1029	not reject	0.2647	reject
	Afternoon	0.0571	not reject	0.1857	not reject
	Evening	0.0370	not reject	0.1296	not reject
Car-Truck	Morning	0.0872	not reject	0.2081	reject
	Off-noon	0.1412	not reject	0.3058	reject
	Afternoon	0.0667	not reject	0.1867	not reject
	Evening	0.2916	reject	0.2311	reject
Van-Car	Morning	0.0684	not reject	0.1453	not reject
	Off-noon	0.0400	not reject	0.2400	reject
	Afternoon	0.0634	not reject	0.1111	not reject
	Evening	0.0640	not reject	0.1520	reject
Van-Van	Morning	0.1557	not reject	0.2131	reject
	Off-noon	0.1363	not reject	0.2727	reject
	Afternoon	0.0746	not reject	0.1493	not reject
	Evening	0.0593	not reject	0.0847	not reject
Van-Truck	Morning	0.0658	not reject	0.1579	not reject
	Off-noon	0.0769	not reject	0.2615	reject
	Afternoon	0.0400	not reject	0.1200	not reject
	Evening	0.0580	not reject	0.1449	not reject
Truck-Car	Morning	0.1240	not reject	0.2479	reject
	Off-noon	0.0340	not reject	0.2203	not reject
	Afternoon	0.0469	not reject	0.1406	not reject
	Evening	0.0579	not reject	0.2479	reject
Truck-Van	Morning	0.0602	not reject	0.1325	not reject
	Off-noon	0.0833	not reject	0.2167	not reject
	Afternoon	0.1231	not reject	0.1538	not reject
	Evening	0.0661	not reject	0.0826	not reject
Truck-Truck	Morning	0.0588	not reject	0.1029	not reject
	Off-noon	0.0656	not reject	0.2459	reject
	Afternoon	0.0465	not reject	0.1163	not reject
	Evening	0.1212	not reject	0.1818	reject

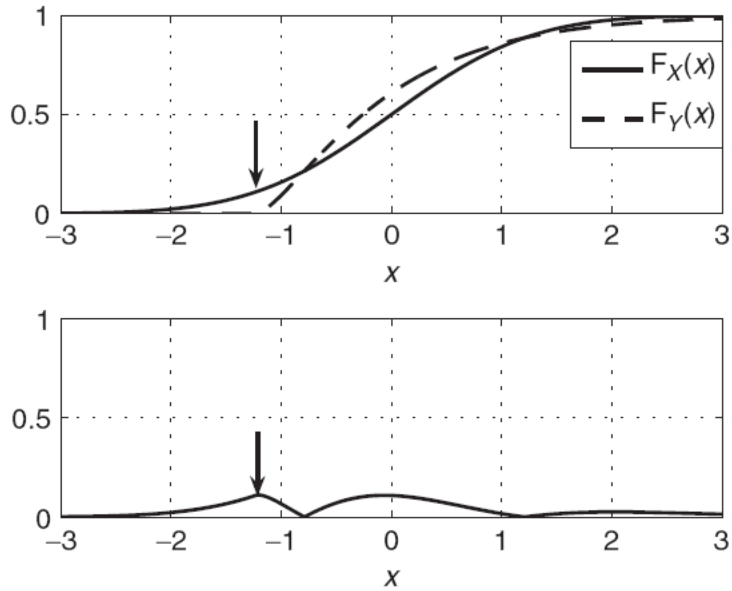

FIGURE 2 Kolmogorov-Smirnov Distance

For example, for Car-Car type, during the morning period, the K-S test statistic of Gaussian kernel-based model is 0.0714, and the corresponding value of lognormal distribution model is 0.1327. Although both model do not reject the null hypothesis test, that headway data follow the proposed model, but the K-S test statistic value of Gaussian kernel model is smaller than lognormal distribution model. Under some circumstance, nonparametric method with Gaussian kernel-based model performs much better than the parametric model, as in Van-Car type, off-noon period, the hypothesis that headway data follow Gaussian kernel model was not rejected while the hypothesis that headway data follow lognormal model was rejected. However, both model cannot provide satisfactory goodness-of-fit for the headway of Car-Truck type in the evening period, both reject the null hypothesis at $\alpha = 0.05$. Also noticed that most of traffic flow types in off-noon period reject the null hypothesis under parametric model condition except Truck-Car and Truck-Van type, while nonparametric model didn't reject. From the whole test results, we can confidently conclude that nonparametric model with Gaussian kernel based method is better than the studied parametric model.

6 Results Analysis

The fundamental statistical characteristics of headway data collected for this study are shown in Table 3.

Some rules could be concluded from the table below. For example, in the morning period, the means of headways are all 7 around and standard deviation are 9 and 10 around. The means of headways are 5 around and standard deviation are 6 around in off-noon and afternoon period. The means and standard deviation of handways are all above 10. The distinctions of the data demonstrate the specification of vehicle types are necessary and

TABLE 3 Fundamental statistical analysis of collected workzone headway data

	Period	Sample size	Means of headways (second)	Standard deviation (second)	Minimum value (second)	Maximum value (second)
Car-Car	Morning	9285	7.19	10.19	196	0
	Off-noon	23849	5.27	6.10	60	0
	Afternoon	14044	5.02	6.15	75	0
	Evening	1212	14.58	18.86	264	0
Car-Van	Morning	3477	7.41	9.46	118	0
	Off-noon	8976	5.58	6.30	68	0
	Afternoon	4567	5.24	6.08	70	0
	Evening	1497	11.34	13.24	108	0
Car-Truck	Morning	2459	7.76	9.65	149	0
	Off-noon	6469	6.24	6.53	85	0
	Afternoon	3023	6.31	7.20	75	0
	Evening	1212	14.58	18.86	264	0
Van-Car	Morning	3503	7.61	9.79	117	0
	Off-noon	8867	5.49	6.31	75	0
	Afternoon	4602	5,31	6.43	63	0
	Evening	1484	12.23	14.85	125	0
Van-Van	Morning	1708	7.15	9.41	122	0
	Off-noon	4494	5.32	5.99	66	0
	Afternoon	2061	5.17	6.08	67	0
	Evening	623	11.81	15.13	118	0
Van-Truck	Morning	1128	7.41	8.90	76	0
	Off-noon	3121	6.00	6.26	65	0
	Afternoon	1191	6.08	6.53	50	0
	Evening	433	13.85	13.98	69	0
Truck-Car	Morning	2406	7.51	9.89	121	0
	Off-noon	6505	5.62	6.34	59	0
	Afternoon	2958	5.51	6.46	64	0
	Evening	1200	13.88	16.16	121	0
Truck-Van	Morning	1113	7.35	9.15	83	0
	Off-noon	2933	5.62	5.94	60	0
	Afternoon	1201	5.25	5.83	65	0
	Evening	416	14.71	17.74	121	0
Truck-Truck	Morning	866	7.49	8.74	68	0
	Off-noon	2532	5.99	6.35	61	0
	Afternoon	951	5.75	6.44	43	0
	Evening	590	13.43	17.08	165	0

correct.

We further present the visual comparisons of the two methods to provide a direct reflection on the two methods. The relative error in the following figures are calculated by

$$relative\ error = \frac{model\ generated\ data - observed\ data}{observed\ data} \tag{9}$$

6.1 Car-Car Headway Distribution

Figure 3 show visualized performance comparisons between the nonparametric method and parametric method using the headway data collected from work zone during evening period. Figure 3 presents the relative error of this two methods. The curves of both cumulative and probability density function directly reflect that nonparametric method is better than parametric method in approximating the observed headway data. From the K-S test nonparametric model did not reject null hypothesis, while parametric model reject null hypothesis. Visual performance of the curves confirms the conclusion drawn form statistical test.

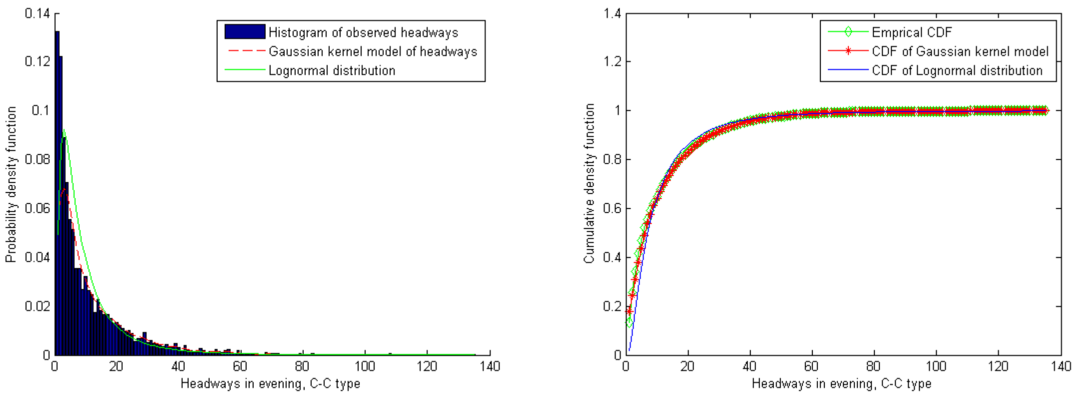

FIGURE 3 Probability Density Functions and Cumulative Density Functions of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Car-Car type in evening period

6.2 Car-Van Headway Distribution

Figure 5 show visualized performance comparisons between the nonparametric method and parametric method using the headway data collected from work zone during morning peak period. Figure 6 presents the relative error of this two methods. The curve also proved that nonparametric method is better. The relative error of nonparametric method in the beginning is small and steady while parametric method is large and fluctuate a little.

6.3 Car-Truck Headway Distribution

Figure 7 show visualized performance comparisons and figure 9 presents the relative error. As showed in the figure 7, in the PDF approximation, nonparametric method fit the data

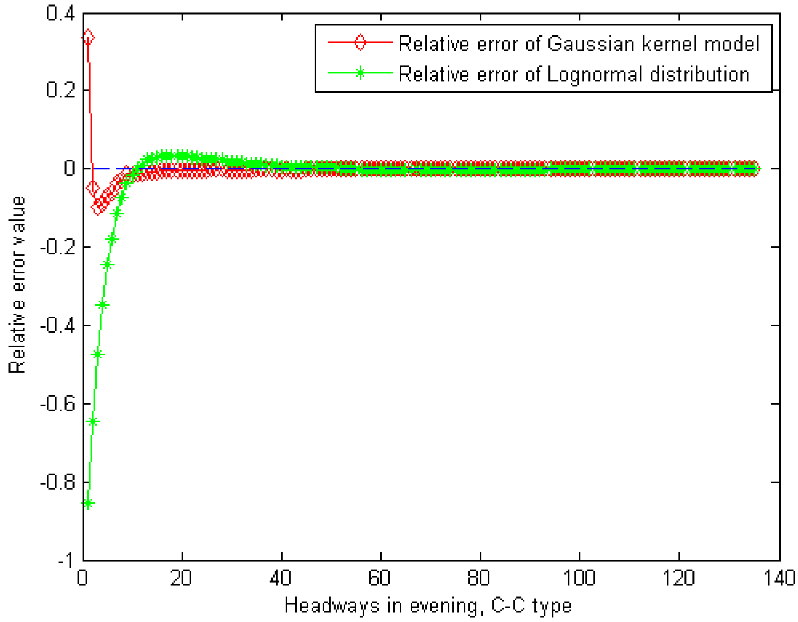

FIGURE 4 Relative Error of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Car-Car type in evening period

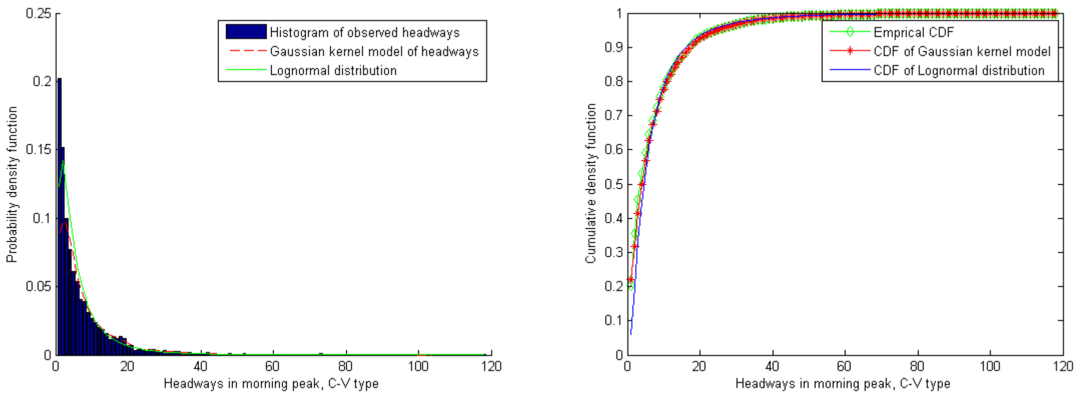

FIGURE 5 Probability Density Functions and Cumulative Density Functions of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Car-Van type in morning peak period

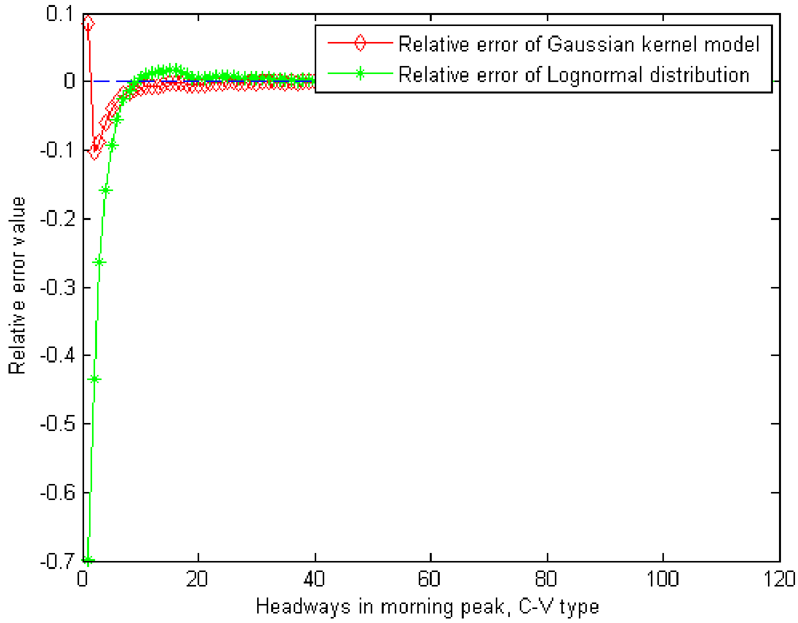

FIGURE 6 Relative Error of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Car-Van type in morning peak period

well but cannot get to a high point, lognormal method depicts the data trend well but cannot fit the data accurately. Figure 9 shows the relative error of both methods would fluctuate in the beginning, but tend to steady and small afterwards.

6.4 Van-Van Headway Distribution

Additionally, we further investigate the Van-Van type flow in the afternoon peak to verify the transferability of the nonparametric method with gaussian kernel function. Figure 9 showed the experimental results. The visual comparisons support the conclusion that overall goodness-of-fit for the nonparametric model with gaussian kernel function is acceptable for different headway samples. And the relative error provide the numerical evidence that gaussian kernel model is better than parametric model. This implies that proposed kernel model can be conducted to model headway on different time periods and flow types.

6.5 Van-Truck Headway Distribution

Following are experimental figures of two more flow types in different time periods. All figures shows a consist support to the nonparametric model.

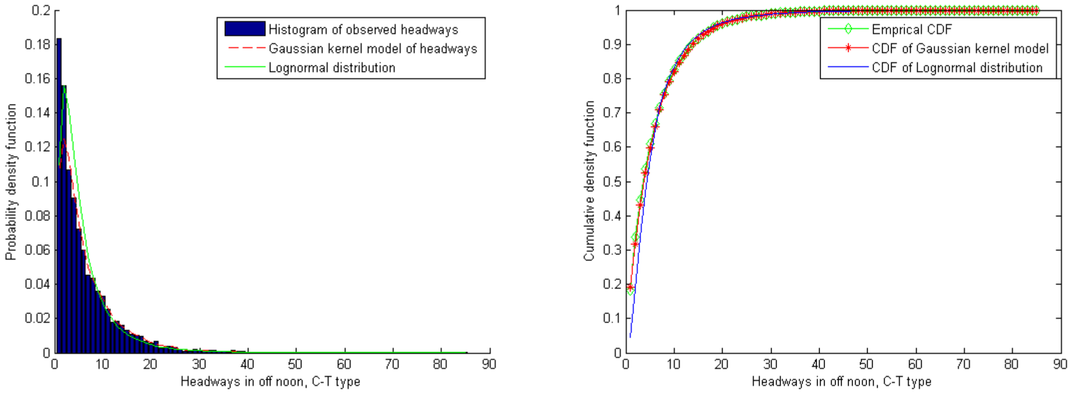

FIGURE 7 Probability Density Functions and Cumulative Density Functions of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Car-Truck type in off noon period

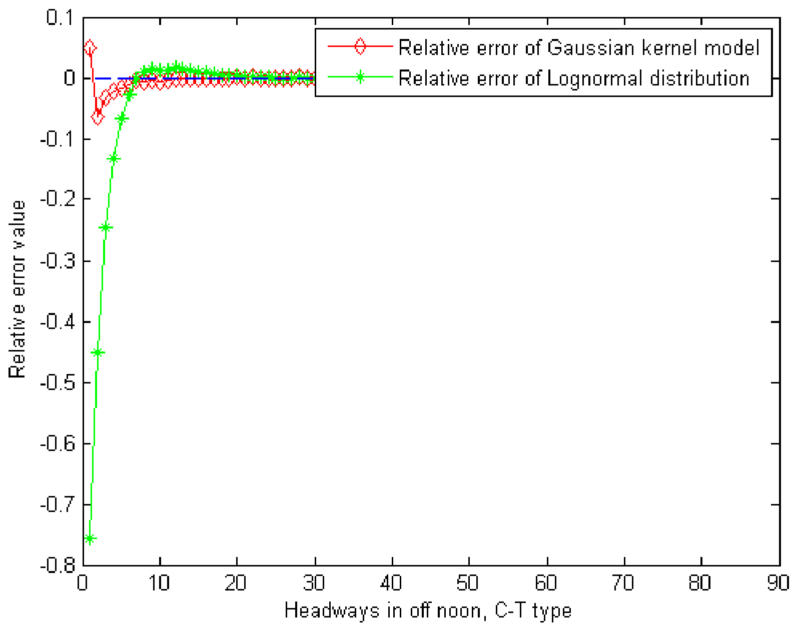

FIGURE 8 Relative Error of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Car-Truck type in off noon period

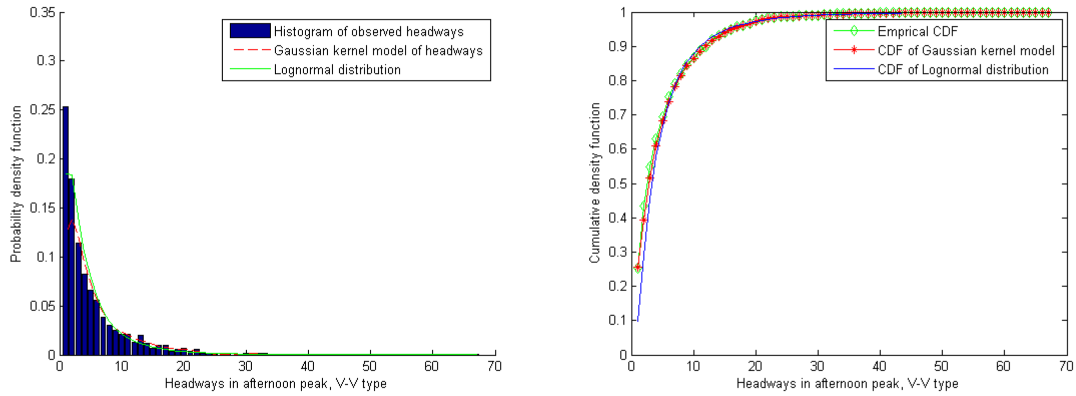

FIGURE 9 Probability Density Functions and Cumulative Density Functions of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Van-Van type in afternoon peak period

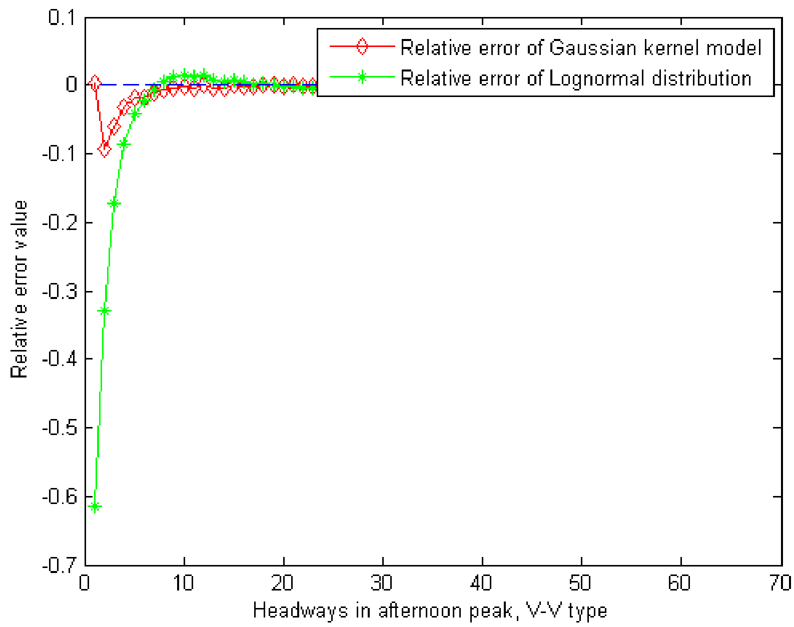

FIGURE 10 Relative Error of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Van-Van type in afternoon peak period

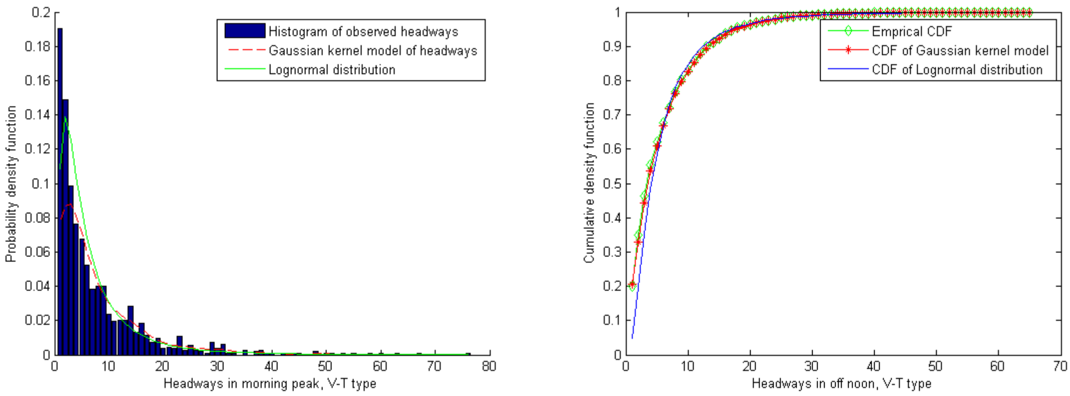

FIGURE 11 Probability Density Functions and Cumulative Density Functions of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Van-Truck type in morning peak period

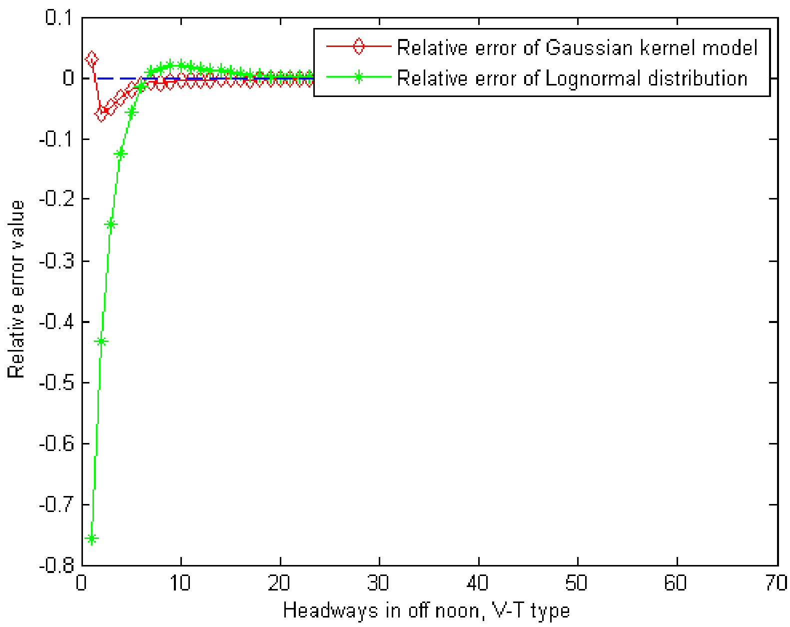

FIGURE 12 Relative Error of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Van-Truck type in morning peak period

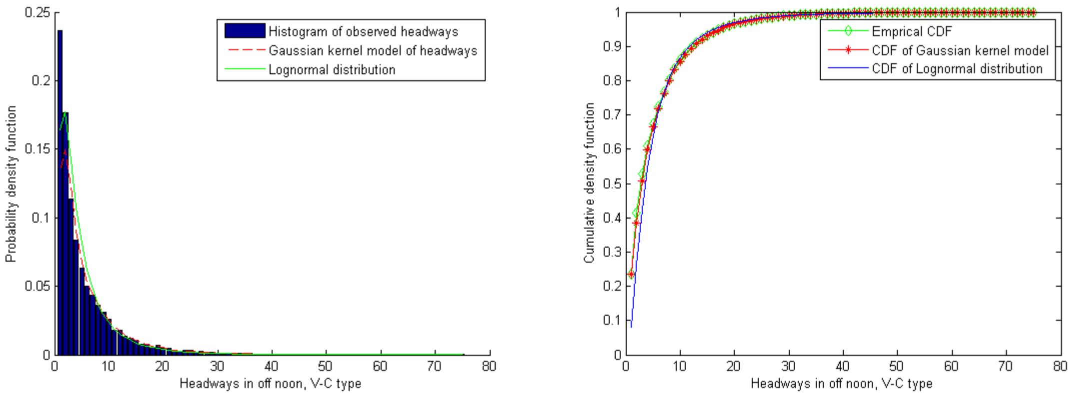

FIGURE 13 Probability Density Functions and Cumulative Density Functions of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Van-Car type in off noon period

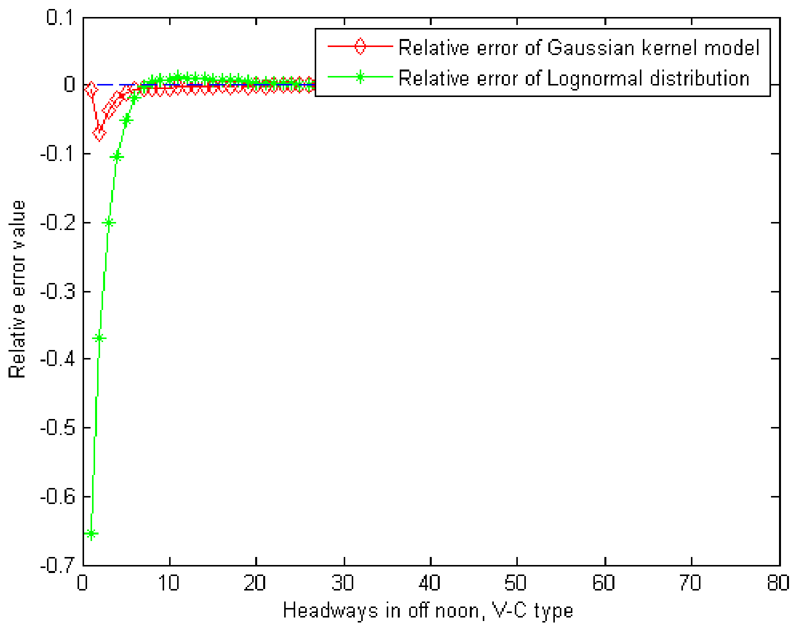

FIGURE 14 Relative Error of the Lognormal Distribution and the Nonparametric Model with Gaussian Kernels Fitted to headway Data of Van-Car type in off noon period

6.6 Van-Car Headway Distribution

Noted from the above figures, same with the conclusion, studied nonparametric model performs better than compared parametric model. as it shows in the relative error figure, big error lies in the lognormal distribution's CDF curve's beginning, which maybe the reason why most of null hypothesis were rejected in off-noon period under parametric model.

7 Discussion

In this paper, we employ nonparametric model with Gaussian kernel-based method to model the vehicle type-specific headway distribution in the work zone. Work zone is important in our daily life since it would significantly effect the traffic flow, research on this topic would accelerate the improvement of traffic control. As to the vehicle type-specific headway research, much research were conducted on this but only four types: Car-Car, Car-Truck, Truck-Car, Truck-Truck. We adopt the FHWA Vehicle Classification scheme, add Van into the vehicle type. From psychological perspective, drivers prefer to keep a large distance if there are a large vehicle ahead. Since the length of the Truck and Car contains a large difference, add Van into the vehicle type would make the model more reasonable. And the statistical and visual test shows nonparametric model with Gaussian kernel-based method is better than parametric model with lognormal distribution which performs best for the Car-Car and Car-Truck type[11], which support our measures on the vehicle type specification.

8 Conclusion

Headway is a fundamental concept in traffic flow theory and simulation research. Most researches were conducted based on the parametric method. But the parametric model contains so many underlying hypothesis that cannot fit to every scenario. Hence, the traditional methods could not provide a good fit to the headway. This paper employ the nonparametric model with Gaussian kernel function, which requires few hypothesis and could handle complicated interaction among vehicles, represent headway characteristics without any artificial assumptions.

A case study on I-91 in Massachusetts was conducted in this paper. Based on the collected data in the work zone, we conducted a K-S test to examine the accuracy of the method in approximating the empirical data. The K-S test indicates that nonparametric model with Gaussian kernel performs better than the parametric model with lognormal distribution. With only one case the nonparametric model reject the null hypothesis, the parametric model reject the null hypothesis often. Later the visualized performances were presented. All the PDF and CDF figures shows nonparametric model is better in fitting the data than the parametric model do, and relative error curve finds the reason why parametric model often reject null hypothesis is that the relative error is large in the beginning. Our research provide help information in modeling the headway.

Although the tests and figures indicate a good capability in modeling headway distribution, but the duration of the data is not long, a long-term headway investigation would be benefit to the traffic flow research. And this paper utilized the nonparametric model with Gaussian kernel, other forms of kernel function could be investigated to examine if they would have a better performance.

References

[1] R. Pueboobpaphan, D. Park, Y. Kim, and S. Choo. Time headway distribution of proble vehicles on single and multiple lane highways. *KSCE Journal of Civil Engineering*, 17(4): 824-836, 2013.

[2] Zwahlen H T, Oner E, Suravaram K R. Approximated headway distributions of free-flowing traffic on Ohio freeways for work zone traffic simulations[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2007, 1999(1): 131-140.

[3] Riccardo R, Massimiliano G. An empirical analysis of vehicle time headways on rural two-lane two-way roads[J]. *Procedia-Social and Behavioral Sciences*, 2012, 54: 865-874.

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

[5] Serge P. Hoogendoorn, Hein Botma. Modeling and Estimation of Headway Distributions. *transportation research record 1591*, Paper No. 970476.

[6] Chen X, Li L, Zhang Y. A Markov model for headway/spacing distribution of road traffic[J]. *Intelligent Transportation Systems, IEEE Transactions*, 2010, 11(4): 773-785.

[7] Zhang G, Wang Y, Wei H, et al. Examining headway distribution models with urban freeway loop event data[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2007, 1999(1): 141-149.

[8] Ye F, Zhang Y. Vehicle type-specific headway analysis using freeway traffic data[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2009, 2124(1): 222-230.

[9] Jin X, Zhang Y, Wang F, et al. Departure headways at signalized intersections: A log-normal distribution model approach[J]. *Transportation research part C: emerging technologies*, 2009, 17(3): 318-327.

[10] Jang J, Park C, Kim B, et al. Modeling of time headway distribution on suburban arterial: Case study from South Korea[J]. *Procedia-Social and Behavioral Sciences*, 2011, 16: 240-247.

[11] Weng J, Meng Q, Fang Fwa T. Vehicle headway distribution in work zones[J]. *Transportmetrica A: Transport Science*, 2013 (ahead-of-print): 1-19.

[12] Zhang Guohui, Wang Yinhai. A Gaussian Kernel-Based Approach for Modeling Vehicle Headway Distributions[J]. *transportation science*, 2013, articles in adcance.

[13] Cowan R J. Useful headway models[J]. *Transportation Research*, 1975, 9(6): 371-375.

[14] Sun D, Benekohal R F. Analysis of work zone gaps and rear-end collision probability[J]. *Journal of Transportation and Statistics*, 2005, 8(2): 71.

[15] Greenberg I. The log normal distribution of headways[J]. *Australian Road Research*, 1966, 2(7).

[16] Griffiths J D, Hunt J G. Vehicle headways in urban areas[J]. *Traffic Engineering and Control*, 1991, 32(10).

[17] Johnson N L. Systems of frequency curves generated by methods of translation[J]. *Biometrika*, 1949, 36(1/2): 149-176.

[18] Branston D. Models of single lane time headway distributions[J]. *Transportation Science*, 1976, 10(2): 125-148.

[19] Al-Ghamdi A S. Analysis of time headways on urban roads: Case study from Riyadh[J]. *Journal of transportation engineering*, 2001, 127(4): 289-294.