Case Study

Nonparametric Modeling of Vehicle-Type-Specific Headway Distribution in Freeway Work Zones

Shangjia Dong[1]; Haizhong Wang[2]; David Hurwitz[3]; Guohui Zhang[4]; and Jianjun Shi[5]

Abstract: This paper presents a vehicle-type-specific headway distribution analysis in a freeway work zone. The goal of this paper is to provide a vehicle-type-specific model with different time periods using empirical work-zone data from highway I-91 in Greenfield, MA. A nonparametric approach with a Gaussian kernel is used to describe the vehicle-type-specific headway distribution in a freeway work zone. No assumption is required on how the headways should be distributed for nonparametric methods. The vehicles are classified into Car, Van, and Truck based on a FHWA vehicle-type classification scheme. Statistical tests indicate nonparametric distribution with Gaussian kernel outperforms the lognormal distribution in statistical sense according to the χ^2 values. Further, another work zone dataset from Jacksonville, FL is utilized to examine the mixed headway scenario without specifying vehicle types. The $K-S$ and Chi-square test results suggest the necessity of considering the vehicle types separately. Following the results, a discussion regarding why nonparametric model is better and the future research directions are presented. **DOI: 10.1061/(ASCE)TE.1943-5436.0000788.** *© 2015 American Society of Civil Engineers.*

Author keywords: Headway distribution; Freeway work zone; Nonparametric approach.

Introduction

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 as can be seen from Fig. 1.

The modelling of vehicle time headway distribution is essential to many aspects of fundamental traffic flow analysis, for example, capacity estimation, microscopic simulation, and safety analysis (i.e., time-to-collision) (Pueboobpaphan et al. 2013).

For traffic simulation models, a key component of determining the simulation model performance is vehicle interarrival times. Researchers, therefore, devote considerable efforts to headway distribution models (Jang et al. 2011). Since time headway can be considered as the reciprocal of flow rate (Zhang et al. 2007), under certain circumstances, vehicle time headway can be used to estimate the road capacity. An accurate headway distribution would help engineers to maximize road capacity and minimize vehicle delays (Zhang et al. 2007). Additionally, as headway is related to vehicle merging and lane-changing behavior, it is essential for estimating road capacity or signal timing parameters at signalized intersections. Furthermore, vehicle interarrival times are also related to traffic safety, driver behavior, and traffic flow theory (Chen et al. 2010). Hence, an inspection of headway distributions is essential and important.

A spectrum of studies have been conducted on vehicle headway distribution models. Recently, Dey and Chandra (2009) proposed gamma and lognormal distribution for desired time gap and time headway in a steady car-following state on two-lane roads under mixed traffic conditions. It was observed that a two-wheeler's desired time gap is the minimum of all other categories of vehicles, while the desired time gap for tractors is the maximum of all vehicles. Zhang and Wang (2013) proposed a nonparametric model with a Gaussian kernel model investigating a freeway scenario but without specifying vehicle types. The results showed the superiority of nonparametric models over parametric families. Ai et al. (2010) retrieved reliable vehicle trajectory data from observation-based video data. Through an examination, it showed significant differences in vehicle-type-specific headways under average traffic conditions, uncongested flow, and congested flow. Simultaneously, Weng et al. (2013) conducted a vehicle-type-specific headway distribution in work zones using empirical data from Singapore through a parametric approach. The results showed a decent approximation of the parametric model for every vehicle type. Jiang et al. (2011) classified vehicles into two categories: Car and heavy goods vehicles (HGV). The results indicated the average time headways of HGV–Car and HGV–HGV patterns were about 0.5 s more than that of Car–Car and Car–HGV. However, parametric methods share a common flaw: the empirical data follow behavioral assumptions. That is, the use of assumptions introduces uncertainty in the modeling technique, which make it difficult to judge which is the true model for headway distribution. Therefore, parametric methods are often established on behavioral assumptions while the nonparametric model is flexible without dependending on prior assumptions. This has been verified by the recent study of Zhang and Wang (2013). Further, nonparametric models can extract the statistical features inherent in the empirical data. Considering these points, this paper proposes a study on

[1]Graduate Student, School of Civil and Construction Engineering, Oregon State Univ., 101 Kearney Hall, Corvallis, OR 97331. E-mail: dongs@onid.oregonstate.edu

[2]Assistant Professor, School of Civil and Construction Engineering, Oregon State Univ., 101 Kearney Hall, Corvallis, OR 97331 (corresponding author). E-mail: Haizhong.Wang@oregonstate.edu

[3]Assistant Professor, School of Civil and Construction Engineering, Oregon State Univ., Corvallis, OR 97331. E-mail: david.hurwitz@oregonstate.edu

[4]Assistant Professor, Dept. of Civil Engineering, MSC01 1070, 1, Univ. of New Mexico, Albuquerque, NM 87131. E-mail: guohui@unm.edu

[5]Associate Professor, Beijing Key Laboratory of Transportation Engineering, Beijing Univ. of Technology, Beijing 100124, China. E-mail: jjshi@bjut.edu.cn

Note. This manuscript was submitted on January 9, 2014; approved on April 10, 2015; published online on July 9, 2015. Discussion period open until December 9, 2015; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering***, © ASCE, ISSN 0733-947X/05015004(13)/$25.00.

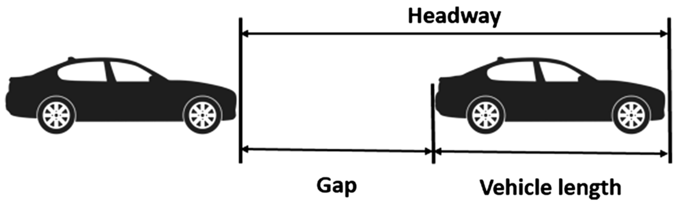

Fig. 1. Definition of vehicle time headway

vehicle headway distribution using a nonparametric method in freeway work zones. In addition, a vehicle-type-specific headway distribution analysis is conducted in a freeway work zone, which is different from the approach of Zhang and Wang (2013).

Objective of the Paper

The purpose of this paper is to present a vehicle-type-specific headway distribution model through a nonparametric approach for a freeway work-zone scenario. This problem is challenging because the availability of large work-zone vehicle-type-specific traffic flow data is limited. A nonparametric Gaussian kernel model is used to evaluate the vehicle-type-specific headway distribution. Then, the nonparametric model is compared against parametric models such as a lognormal distribution. The suggested approach is illustrated with case studies and experimental results.

Paper Organization

The remainder of this paper is organized as follows. A thorough review is presented in literature review section. Theory and methodology are introduced in "Methodology" section. "Experimental Section" describes study area, data and experimental design. Then the statistical tests methodology are provided in "Statistical Analysis: Probability Metric" section, and the descriptive statistics are exhibited in "Descriptive Statistical Analysis" section. In "Results Analysis," the goodness-of-t test and visual performance of the results are presented. At last, "Summary, Conclusion, and Future Work" concluded the paper with future remarks.

Literature Review

Many headway distribution models have been derived and calibrated using empirical traffic data. In general, these models can be categorized into two groups: single statistical distribution models and mixed models (Zhang et al. 2007). The authors summarized the existing relevant headway distribution studies in Fig. 2.

Single Distribution Models

Representatives of the single statistical distribution family include normal distribution, log-normal distribution (Greenberg 1966), Weibull distribution (Sun and Benekohal 2005), the Erlang distribution (Al-Ghamdi 2001; Zhang and Wang 2013), exponential distribution, log-logistic distribution, inverse Gaussian distribution, and Gamma distribution (Al-Ghamdi 2001; Yin et al. 2009). For instance, Sun and Benekohal (2005) used a Weibull distribution model to describe the vehicle headways in work zones. Jang et al. (2011) examined a Johnson SU distribution together with a Johnson SB distribution (Johnson 1949). A lognormal distribution is transformation of a normal distribution, which can be employed to describe naturally occurring unimodal sets of data. Jin et al. (2009) studied departure headways and indicated that the headway distribution in a queue resembles a lognormal distribution. Al-Ghamdi (2001) recommended four headway distribution models at different flow rates, such as a negative exponential distribution for low flow rates; shifted exponential and gamma distributions for the middle flow rate; and an Erlang distribution for high flow rates. Riccardo and Massimiliano (2012) analyzed a few case studies on rural two-lane two-way roads and suggested that the inverse Weibull distribution fits the empirical headway data better. Yin et al. (2009) studied the dependence of headway distribution on traffic status and showed that the lognormal distribution is adequate to fit headway when traffic is in a free-flow state, and the log-logistical distribution is suitable in congested state.

Mixed Distribution Models

Many of the single distributions can fit empirical headway in free-flow condition but not congested flow. The poor predictive capability makes their performance unsatisfying. Mixed headway distribution models, therefore, were pursued to better capture headway dynamics. Examples of mixed models used to predict headway distributions include double displaced negative exponential distribution (DDNED) (Griffiths and Hunt 1991), combined normal distribution and shifted negative exponential distribution (Ye and Zhang 2009), combined negative exponential distribution and shifted negative exponential distribution (Ye and Zhang 2009), generalized queuing model (GQM) (Zhang et al. 2007) and semi-poisson distribution etc. For instance, Zhang et al. (2007) found that double displaced negative exponential distribution (DDNED) and lognormal distribution better described high-occupancy vehicle (HOV) lanes and regular lanes. In a vehicle-type-specific and car-dominant case, Ye and Zhang (2009) proved that combined negative exponential distribution and shifted negative exponential distribution is better than combining a combined normal distribution and shifted negative exponential distribution (a mixed distribution) in describing empirical headway. Some mixed distributions 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 (Zhang et al. 2007). According to this construct, many important models have been derived such as the Cowan M1–M4 (Cowan 1975), the generalized queuing model (Branston 1976), and the semi-poisson model. Among these, Cowan's M3 model is widely investigated and applied for its simplicity and easy approximation of describing longer headways (Zhang and Wang 2013). Vasconcelos et al. (2012) proposed a simultaneous numerical estimation (SNE) to estimate the parameters of Cowan's M3 headway distribution. The requirement of explicit expression for the Laplace transform of the following-vehicle headway distribution makes the use of the semi-poisson model limited (Zhang et al. 2007). Further, Cowan M3, Cowan M4, GQM, and DDNED were extensively analyzed by Zhang et al. (2007).

Vehicle-Type-Specific Headway Distribution

Considering empirical traffic compositions, researchers started exploring the impact of vehicle types on headway distribution. Ye and Zhang (2009) categorized headways into four types according to different combinations of vehicle types (leader–follower). They adopted three distribution models for the four headway types: the shifted negative exponential distribution for Truck–Car and Truck–Truck types, an Erlang distribution for the Car–Truck type, and a composite model for the Car–Car type (Ye and Zhang 2009). Weng et al. (2013) conducted an experiment and concluded that headways are strongly related to the types of the leading and following vehicles. In the examination, they found lognormal distribution best fits the Car–Car headway type, as well as the Car–Truck

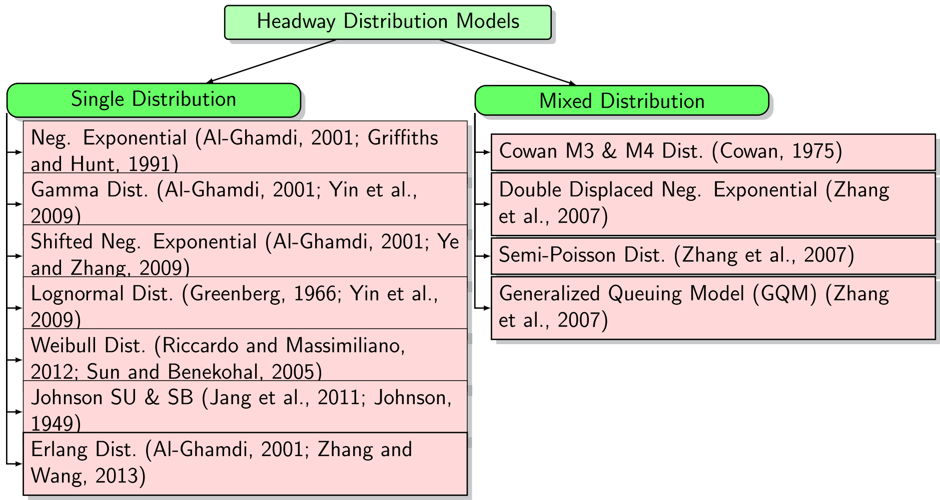

Fig. 2. Summary of vehicle headway distribution models in literature

headway, and that an inverse Gaussian distribution is appropriate for Truck–Car and Truck–Truck headways. From the study, it was concluded that the location and scale of a headway distribution model may be influenced by four factors: traffic flow rate, percentage of trucks, lane position, and intensity of work-zone activity.

Although single distribution models are simple and easy to apply, they are typically inadequate with approximating shorter (i.e., under 3 s) vehicle time headways. Meanwhile, mixed distribution models are more flexible to describe headways 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 (Zhang et al. 2007). Table 1 shows the scope of recent studies.

Parametric versus Nonparametric Approaches

Despite the fact that many distribution models have been investigated and applied, the existing parametric approaches share a common feature: the headway are assumed to follow a particular distribution and then that assumption is checked against empirical data. The distribution that fits the empirical headway better is selected as the preferable alternative. Therefore, the result is sometimes inaccurate because of our limited capacity to correctly choose the preferable distribution. Furthermore, parametric headway models require strict prior knowledge that is often not available. Although these parametric models are simple and intuitive to understand, such as Weibull and Poisson distribution, the goodness-of-fit varies with the location and level of traffic flow (Zhang and Wang 2013). In addition, one of the major limitations of parametric models is that you have to make an assumption about the shape parameter of the headway distribution. There are ways to limit the possibility of making an incorrect choice, but it's difficult to tell whether the correct parametric model has been chosen. Thus, if the assumption can be justified, parametric method is preferred, otherwise, a nonparametric technique should be pursued. Essentially, deterministic models do not properly account for the stochastic nature of variables or the transient nature of traffic (Zwahlen et al. 2007). Mixed models typically depict real situations better. But at the cost of complex conformation and calibration, nonparametric models work better due to their flexibility and ability of extracting the statistical features of observed headways without referring to assumed distribution models with specific parameters (Zhang and Wang 2013). As Zhang and Wang (2013) stated, the nonparametric Gaussian kernel headway model outperforms the traditional parametric model because its flexible data modeling ability, which requires few stringent hypotheses, can sufficiently handle subtle and complicated interactions among vehicles and does not rely on the assumption that the data are drawn from a particular distribution. The applicability and compatibility is greater than traditional parametric methods. Also, the transferability test showed the model is independent to specific sample data and could be generalized to suit different sample data under a similar traffic scenario.

Methodology

Most tests were conducted based on the data collected from freeways or HOV lanes, compared with uninterrupted traffic; however, work-zone traffic has unique characteristics (Weng et al. 2013). This paper considers a nonparametric headway distribution based on work-zone vehicle data. In headway distribution modeling, goodness-of-fit tests are used to judge how a distribution fits the sample data (Weng et al. 2013). In this study, the K-S test was adopted to determine the goodness-of-fit in the work-zone traffic

Table 1. Summary of Recent Vehicle Headway Distribution Studies

Headway study	Mixed	Vehicle-type specific	Parametric	Nonparametric	Scenario
Zwahlen et al. (2007)	X	—	X	—	Freeway
Yin et al. (2009)	X	—	X	—	Urban roadways
Ye and Zhang (2009)	X	X	X	—	Freeway
Jang et al. (2011)	X	—	X	—	Suburban arterial
Riccardo and Massimiliano (2012)	—	—	X	—	Rural two-lane two-way road
Zhang and Wang (2013)	X	—	—	X	Freeway
Weng et al. (2013)	X	X	X	—	Work zone
Dong et al. (2015) this study	—	X	—	X	Work zone

scenario. A Chi-square test was utilized to further validate the results. In Jin et al. (2009) and Yin et al. (2009), parametric methods were adopted to measure headway distribution, in which lognormal distribution demonstrate a distinguishing performance with better goodness-of-fit test results. Therefore, lognormal distribution was selected to make the comparison in this study.

Nonparametric Model

The estimated probability density function (PDF) of the Gaussian kernel model is defined as follows (Zhang and Wang 2013):

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

where X_i = individual headway measurement. The h was computed by

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

where δ = standard deviation of the data set; and n = sample size of the data set. As a linear composition of Gaussian kernels, the PDF has differentiable and continuous characteristics derived from the kernels (Zhang and Wang 2013) that could strengthen the smoothness of the density curve. Kernel density estimators belong to a class of estimators called nonparametric density estimators. In comparison to parametric estimators where the estimator has a fixed functional form (structure) and the parameters of this function are the only information to store, nonparametric estimators have no fixed structure and depend upon all the data points to reach an estimate.

Parametric Models

In this paper, the lognormal distribution is compared to the Gaussian kernel function. The models are shown as follows:

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

with

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

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

where μ = mean of the data set; and δ^2 = variance.

Experimental Section

Study Area and Data Description

The work-zone site is on highway I-91 in Greenfield, Massachusetts. The data was collected in 2005 (Heaslip 2007) and the detailed information is presented in Table 2. The length of the work zone is approximately 0.92 km (0.57 mi) with one lane closed for a bridge rehabilitation project northbound and southbound over the B&M railroad. The proportion of heavy vehicles was 1.67. The work-zone traffic-flow data were collected over a week-long time horizon using pneumatic sensors. The data was collected at the following five specific locations, shown in Fig. 3:

Table 2. I-91 Work-Zone Site and Traffic Description (Data from Heaslip 2007)

Type	Data
Work zone	—
Length	0.19 km (0.12 mi)
Number of lanes open	1/direction
Duration (LC)	Permanent
Timing (LC)	24 h
Posted WZ speed	104.6 km/h (65 mph)
Availability of alternative route	No
R.type	Bridge
R.strategy	3 phase
Roadway	Interstate
Limited access	No
Senior center	3.22 km (2 mi)
Pavement condition	Good
Nearby attractor	Retail
Driveways	0
Curb cuts	0
Geometric	—
Number of lanes	2
Lane width	3.66 m (12 ft)
Shoulder width	Varies
Maximum upgrade %	2.84%
Maximum downgrade %	−1.01
Superelevation	Yes
V. curve type	Crest
V. curve length	120 m
Design speed	104.6 km/h (65 mph)
Sight distance	No
Bridges	1
Traffic	—
ADT	43,500
DHV	3,433
PHV	2,871
AM/PM LOS	B/B
Truck	1.67
Posted limit	104.6 km (65 mph)
Speed	115.9 km (72 mph)
Mean speed	107.8 km (67 mph)
Crash rate	44/year
Crash type	Single vehicle
Injury crashes	10%
PDO crashes	90%
% of old drivers	14%

- Location 1: 4.82 km (3 mi) north of the work area;
- Location 2: 3.22 km (2 mi) from the work zone and after the location of variable message sign (VMS);
- Location 3: 2.41 km (1.5 mi) form the work zone directly under the bridge;
- Location 4: inside the taper; and
- Location 5: inside the work area.

The highest volumes were observed on Sunday afternoons. During the week, the a.m. peaks were higher than the p.m. peaks except for Friday, when the evening peak was higher than the morning peak and spread out over a 5-hour period. In the evening, the data show that the speeds at Location 5 were higher than the speeds in the taper, which can be attributed to its location at the end of the work area. Evening hours also show higher speeds, which may be attributed to a lack of congestion and work activities within the work zone. During the Sunday afternoon congestion, the speed at Location 5 was higher than at Location 4. This is partially due to the slow speeds that were caused by congestion in advance of the work zone. At Location 5, the vehicles speed up leaving the work zone whereas vehicles at Location 4 were in a stop-and-go traffic flow.

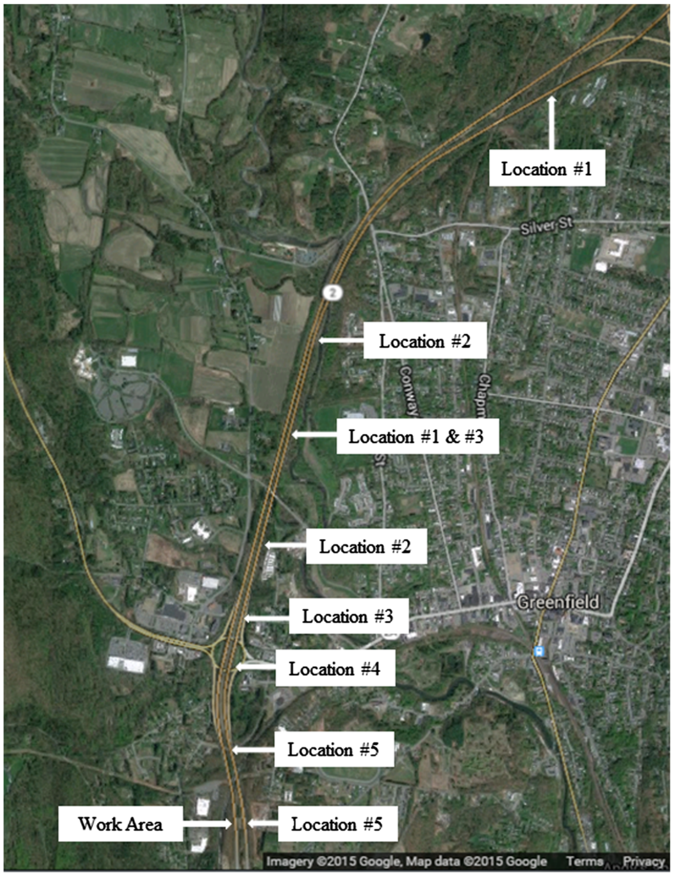

Fig. 3. Study site of work zone on I91 in Greenfield, Massachusetts (imagery ©2015 Google, map data ©2015 Google)

The shoulder width of this work zone is 3.66 m (12 ft). The free flow speed is 112 km/h (70 mph) with 2 lanes in each direction. The annual average daily traffic (ADT) is about 72,000 vehicles.

Experiment Design

Based on the FHWA (2013) Vehicle Classification scheme, the authors classify the vehicles into three types: Car (1–2), Van (3), and Truck (4–14). Then the traffic composition consists of nine types: Car–Car, Car–Van, Car–Truck, Van–Car, Van–Van, Van–Truck, Truck–Car, Truck–Van, and Truck–Truck. Use the *for* loop algorithm, the data were separated into nine categories correponding to these nine types. The time of the day was separated into four periods: morning peak, off-noon, afternoon peak, and evening. For example, find the traffic flow where the leading vehicle is a Car and following vehicle is a Truck. Then within this group, selected the Car–Truck type of headway that happens in the morning. This yields the Car–Truck morning dataset. Repeating this procedure derives 36 subsets of data. Without losing generality, experiments based on the specific flow types in given time periods are illustrated as follows:

- Morning period (6:00–9:30 a.m.);
- Off-noon period (9:30–16:30 p.m.);
- Afternoon period (16:30–19:30 p.m.); and
- Evening period (19:30–22:00 p.m.)

Statistical Analysis: Probability Metric

In order to give a numerical expression instead of exclusively considering visual inspection on the results, two statistical tests were conducted to investigated the goodness-of-fit of both methods: Chi-square test and K-S test. The K-S test is a form of minimum-distance estimation used as a nonparametric test 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 K-S statistic quantifies the distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution or between the empirical distribution functions of two samples. A Chi-square distribution is a nonparametric test, which is a statistical test applied to set of categorical data to evaluate how likely it is that any observed differences between the sets arose by chance.

Two-Sample K-S Test

The K-S test is commonly used to obtain a probability of similarity between two distributions to determine whether two datasets differ significantly. The K-S test is nonparametric and assumption-free, meaning that it has the advantage of making no assumption about the distribution of data. The purpose of this test is to obtain the cumulative distribution function of the two distributions that need to be compared. The K-S distance is a measure defined as the maximum value of the absolute difference between two cumulative distribution functions; it measures the largest absolute difference between two distribution functions $F(t)$ and $G(t)$ for varying t. In a similar setting, the K-S distance is defined by

$$\rho_K(X,Y) := \|F-G\|_\infty = \sup_{t\subset\mathbb{R}}|P(X\leq t)-P(Y\leq t)| = \sup_t|F(t)-G(t)| \quad (6)$$

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:

$$F_n(t) = \frac{1}{n}\#\{x_i|x_i\leq t\} \quad (7)$$

where $\#\{\ldots\}$ = number of elements contained in the set $\{\ldots\}$; and F_n = discrete probability distribution function on the real line. For large values of n, the empirical distribution converges to the theoretical one.

Chi-Square Test

The Chi-square test is utilized to examine if the sample data came from a population with certain distribution. The value of the test-statistic is

$$\chi^2 = \sum_{i=1}^{n}\frac{(O_i-E_i)^2}{E_i} \quad (8)$$

where χ^2 = Pearson's cumulative test statistic, which asymptotically approaches a χ^2 distribution; O_i = observed frequency for bin i; E_i = expected (theoretical) frequency for bin i as asserted by the null hypothesis; and n = number of categories, groupings, or possible outcomes.

A distinguishing feature of the Chi-square goodness-of-fit test is that it could be applied to any univariate distribution for which you can calculate the cumulative distribution function (CDF). The Chi-square goodness-of-fit test is applied to binned data, but this is not a restriction, because you can simply calculating a histogram or frequency table.

Table 3. Fundamental Statistical Analysis of Collected Work Zone Headway Data

Type	Period	Sample size	Means of headways (s)	Standard deviation (s)	Maximum value (s)	Minimum value (s)
Car–Car	Morning	9,285	7.19	10.19	196	0
	Off-peak	23,849	5.27	6.10	60	0
	Afternoon	14,044	5.02	6.15	75	0
	Evening	1,212	14.58	18.86	264	0
Car–Van	Morning	3,477	7.41	9.46	118	0
	Off-peak	8,976	5.58	6.30	68	0
	Afternoon	4,567	5.24	6.08	70	0
	Evening	1,497	11.34	13.24	108	0
Car–Truck	Morning	2,459	7.76	9.65	149	0
	Off-peak	6,469	6.24	6.53	85	0
	Afternoon	3,023	6.31	7.20	75	0
	Evening	1,212	14.58	18.86	264	0
Van–Car	Morning	3,503	7.61	9.79	117	0
	Off-peak	8,867	5.49	6.31	75	0
	Afternoon	4,602	5,31	6.43	63	0
	Evening	1,484	12.23	14.85	125	0
Van–Van	Morning	1,708	7.15	9.41	122	0
	Off-peak	4,494	5.32	5.99	66	0
	Afternoon	2,061	5.17	6.08	67	0
	Evening	623	11.81	15.13	118	0
Van–Truck	Morning	1,128	7.41	8.90	76	0
	Off-peak	3,121	6.00	6.26	65	0
	Afternoon	1,191	6.08	6.53	50	0
	Evening	433	13.85	13.98	69	0
Truck–Car	Morning	2,406	7.51	9.89	121	0
	Off-peak	6,505	5.62	6.34	59	0
	Afternoon	2,958	5.51	6.46	64	0
	Evening	1,200	13.88	16.16	121	0
Truck–Van	Morning	1,113	7.35	9.15	83	0
	Off-peak	2,933	5.62	5.94	60	0
	Afternoon	1,201	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-peak	2,532	5.99	6.35	61	0
	Afternoon	951	5.75	6.44	43	0
	Evening	590	13.43	17.08	165	0

Descriptive Statistical Analysis

The fundamental statistical characteristics of headways are shown in Table 3 to give a perception on the dataset. For example, in the morning period, the mean of the headway was approximately 7 s and the standard deviation was approximately 9–10 s. The means of headway was approximately 5 s, and standard deviation was approximately 6 s for both off-noon and afternoon periods. The means and standard deviations of headway were both above 10 s in evening period.

Results Analysis

In order to examine the goodness-of-fit for both models, two-sample K-S test and Chi-square were employed to provide statistical evidence. A smaller K-S statistic value indicates a better goodness-of-fit, and the decision to reject the null hypothesis is made by comparing the p-value with the significance level α. The comparison of K-S statistics and hypothesis tests are illustrated in Table 4. The null hypothesis is that the two samples of data are generated from the same distribution and the significant confidence level is 95%.

It can be concluded from Table 4 that the nonparametric model performs better than the parametric model in most scenarios. For example, for Car–Car type headways, during the morning period, the K-S test statistic of the Gaussian kernel-based model is 0.0714, and the corresponding value of the lognormal distribution model is 0.1327. Although both models do not reject the null hypothesis test, the K-S test statistic value of the Gaussian kernel model is smaller than the lognormal distribution model. Under some circumstances such as Van–Car type headways, the nonparametric method with a Gaussian kernel-based model performs much better than the parametric model. During the off-noon period, the hypothesis that the headway data follows Gaussian kernel model was not rejected while the hypothesis that headway data follow the lognormal model was rejected. However, both models cannot provide satisfactory goodness-of-fit for the Car–Truck type headways in the evening period: both reject the null hypothesis at $\alpha = 0.05$. Also most traffic flow types during the off-noon period rejected the null hypothesis under parametric model conditions except Truck–Car and Truck–Van type while the nonparametric model did not reject. These findings indicate that the nonparametric model with a Gaussian kernel displayed a better performance than the studied parametric model of lognormal distribution. In Table 5, Chi-square test are conducted for further examination.

Observed from Table 5, neither distribution rejects the null hypothesis, that is, the empirical data follows the distribution. However, most χ^2 values of the nonparametric distribution are lower than in parametric distribution column. Since the lower the χ^2 value, the better the empirical data fits the distribution, the nonparametric distribution displays its strength in depicting

Table 4. Comparison of K-S Test Results between Nonparametric and Parametric Models, I-91

Type	Period	Nonparametric model with Gaussian kernel function		Parametric model lognormal distribution	
		K-S statistic	Hypothesis test	K-S statistic	Hypothesis test
Car–Car	Morning	0.0714	Not reject	0.1327	Not reject
	Off-peak	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-peak	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-peak	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-peak	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-peak	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-peak	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-peak	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-peak	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-peak	0.0656	Not reject	0.2459	Reject
	Afternoon	0.0465	Not reject	0.1163	Not reject
	Evening	0.1212	Not reject	0.1818	Reject

the work-zone headway distribution. The relative error is presented to provide visual comparisons of the goodness-of-fit index. It is defined as

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

Car–Car and Van–Car Headway Distribution

Figs. 4(a–d) present a visual performance comparisons of Car–Car and Van–Car headway distributions through a probability density function and cumulative distribution function in the evening and over a day's off-noon period between the nonparametric method and parametric method using the headway data collected from the I-91 work zone. Figs. 4(e and f) present the relative error of these two methods. The curves of both CDF and PDF directly reflect that the nonparametric method is better than the parametric method in approximating the observed headway data. In the K-S test, the nonparametric model did not reject the null hypothesis while the parametric model did. Relative error plots at the bottom of Fig. 4 confirms the conclusion drawn from statistical test.

Car–Van and Car–Truck Headway Distribution

Figs. 5(a–d) exhibit a visual performance of Car–Van and Car–Truck headway distributions between the nonparametric and parametric method using the headway data collected from the work zone during the morning peak and off-noon periods. Figs. 5 (e and f) present the relative error of these two methods. The relative error of the nonparametric method in the beginning is small and steady while the relative error of the parametric method is large and fluctuates at the beginning. This observation is consistent with the earlier findings in goodness-of-fit test, namely that the nonparametric model outperforms the parametric model.

Van–Van and Van–Truck Headway Distribution

This section provides further analysis of Van–Van and Van–Truck headway distributions during both the afternoon peak and off-noon periods. Figs. 6(a–d) show the experimental results, which verified the transferability of the nonparametric method with a Gaussian kernel function. The visual comparison supports the argument that the overall goodness-of-fit for the nonparametric model with Gaussian kernel function is favorable for different headway samples. In addition, the relative error plots in Figs. 6(e and f) provide supporting numerical evidence that the Gaussian kernel model is

J. Transp. Eng., 2015, 141(11): 05015004

Table 5. Comparison of χ^2 Test Results between Nonparametric and Parametric Models, I-91

Type	Period	Nonparametric model with Gaussian kernel function		Parametric model lognormal distribution	
		χ^2 stats	Hypothesis test	χ^2 stats	Hypothesis test
Car–Car	Morning	3.8×10^{-4}	Not reject	2.1×10^{-3}	Not reject
	Off-peak	1.9×10^{-4}	Not reject	5.5×10^{-4}	Not reject
	Afternoon	2.2×10^{-4}	Not reject	4.6×10^{-3}	Not reject
	Evening	9.9×10^{-4}	Not reject	2.2×10^{-3}	Not reject
Car–Van	Morning	8.6×10^{-4}	Not reject	3.2×10^{-3}	Not reject
	Off-peak	3.4×10^{-4}	Not reject	5.6×10^{-3}	Not reject
	Afternoon	4.3×10^{-4}	Not reject	5.7×10^{-3}	Not reject
	Evening	2.3×10^{-3}	Not reject	2.9×10^{-3}	Not reject
Car–Truck	Morning	5.6×10^{-4}	Not reject	3.6×10^{-3}	Not reject
	Off-peak	2.2×10^{-4}	Not reject	4.9×10^{-3}	Not reject
	Afternoon	6.5×10^{-4}	Not reject	5.3×10^{-3}	Not reject
	Evening	1.8×10^{-3}	Not reject	1.6×10^{-3}	Not reject
Van–Car	Morning	9.5×10^{-4}	Not reject	3.3×10^{-3}	Not reject
	Off-peak	2.9×10^{-4}	Not reject	5.1×10^{-4}	Not reject
	Afternoon	6.5×10^{-4}	Not reject	6.1×10^{-3}	Not reject
	Evening	2.4×10^{-3}	Not reject	2.3×10^{-3}	Not reject
Van–Van	Morning	1.4×10^{-3}	Not reject	3.1×10^{-3}	Not reject
	Off-peak	4.4×10^{-4}	Not reject	5.7×10^{-3}	Not reject
	Afternoon	7.9×10^{-4}	Not reject	6.0×10^{-4}	Not reject
	Evening	5.4×10^{-3}	Not reject	2.5×10^{-3}	Not reject
Van–Truck	Morning	2.0×10^{-3}	Not reject	4.9×10^{-3}	Not reject
	Off-peak	4.5×10^{-4}	Not reject	6.0×10^{-3}	Not reject
	Afternoon	1.3×10^{-3}	Not reject	8.2×10^{-3}	Not reject
	Evening	4.8×10^{-3}	Not reject	2.7×10^{-3}	Not reject
Truck–Car	Morning	1.1×10^{-3}	Not reject	3.3×10^{-3}	Not reject
	Off-peak	3.6×10^{-4}	Not reject	6.2×10^{-3}	Not reject
	Afternoon	5.7×10^{-4}	Not reject	6.6×10^{-3}	Not reject
	Evening	3.2×10^{-3}	Not reject	1.5×10^{-3}	Not reject
Truck–Van	Morning	2.1×10^{-3}	Not reject	5.3×10^{-3}	Not reject
	Off-peak	4.1×10^{-4}	Not reject	7.1×10^{-3}	Not reject
	Afternoon	7.4×10^{-4}	Not reject	6.8×10^{-3}	Not reject
	Evening	7.8×10^{-3}	Not reject	2.7×10^{-3}	Not reject
Truck–Truck	Morning	1.9×10^{-3}	Not reject	5.7×10^{-3}	Not reject
	Off-peak	4.3×10^{-4}	Not reject	6.8×10^{-3}	Not reject
	Afternoon	1.4×10^{-3}	Not reject	1.09×10^{-2}	Not reject
	Evening	3.8×10^{-3}	Not reject	3.2×10^{-3}	Not reject

superior to the parametric model in approximating empirical headways. This implies that the proposed Gaussian kernel model can be used to model the different headways' time periods with varying traffic flow conditions in a freeway work-zone setting.

Model Goodness-of-Fit

According to the goodness-of-fit tests, the nonparametric distribution demonstrates its strength in depicting headways with specific vehicle types in various time periods. Still, the necessity of specifying the vehicle types and time periods needs further investigation. An experiment that does not separate the vehicle types and time periods was developed as shown Tables 6 and 7. The null hypothesis is that the sample data follow the proposed distribution.

As can be concluded, when all vehicle types are treated as a whole, the K-S test rejected the null hypothesis while the Chi-square test did not. In the vehicle-type-specific with four time periods approach, K-S test did not reject most null hypotheses, and its goodness-of-fit also demonstrated a really good result. Considering these findings, it is necessary to model the headway in different traffic types with certain time periods, which validates the significance of our research. Figs. 7(a and b) still favor the conclusion that the nonparametric model performs better than the parametric model.

Transferability Test

As Koppelman and Wilmot (1982) stated, "First, we define transfer as the application of a model, information, or theory about behavior developed in one context to describe the corresponding behavior in another context. We further define transferability as the usefulness of the transferred model, information, or theory in the new context." There are two categories of transferability tests: temporal transferability tests and spatial transferability tests. The key distinction between them is that temporal transferability focuses on application of a model developed using data collected in different periods of time, while spatial transferability focuses on application of a model developed using data collected in different spatial area. In this paper, both temporal and spatial are involved (Fox et al. 2014).

A common used statistical test in the literature is the transferability test statistic (TTS), which assesses the transferability of the base-model parameters b in the transfer context t, under the hypothesis that the two sets of parameters are equal (Fox et al. 2014)

$$\mathrm{TTS}_t(\beta_b) = -2[LL_t(\beta_b) - LL_t(\beta_t)] \tag{10}$$

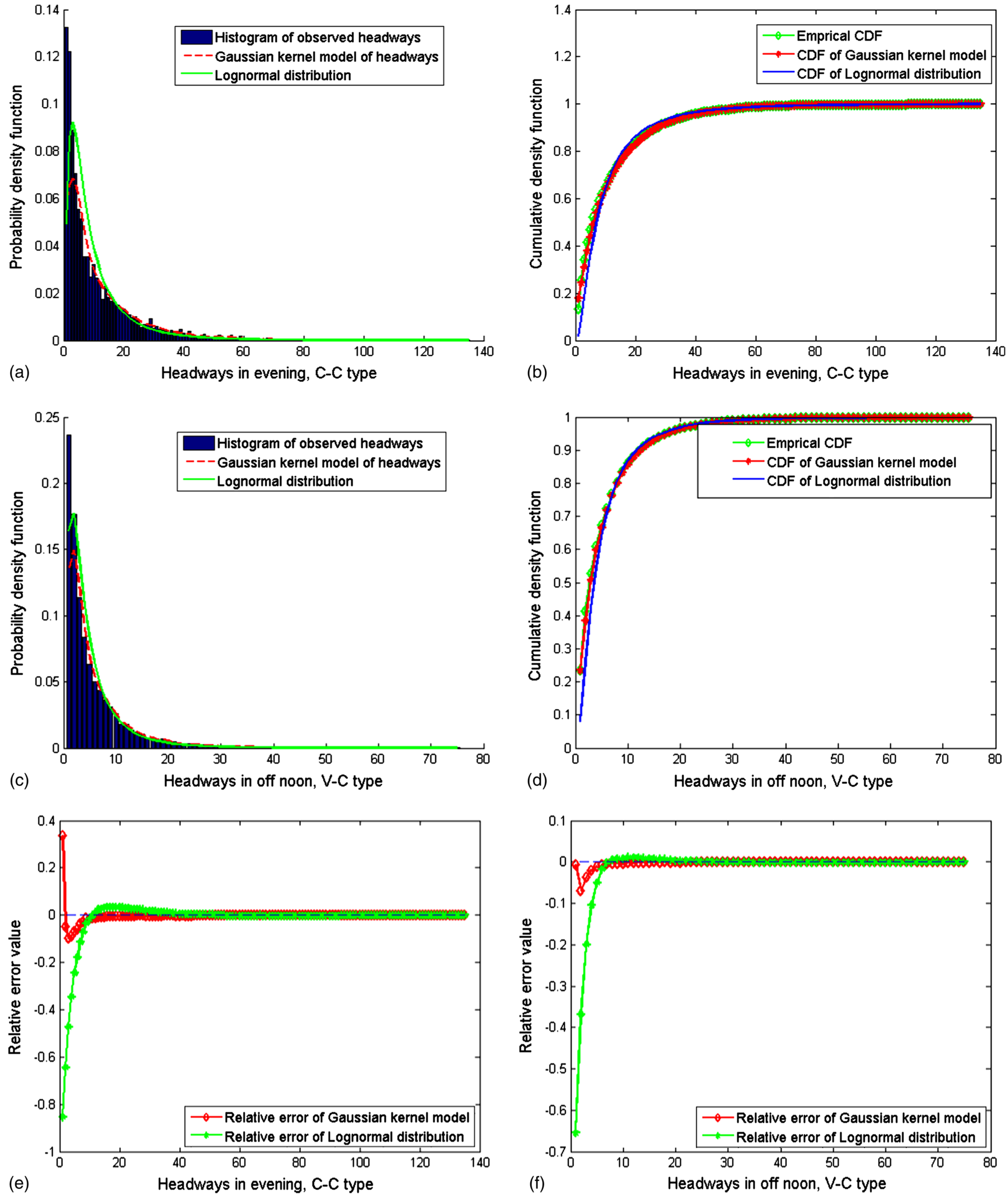

Fig. 4. Probability density functions and cumulative density functions of the lognormal distribution and the nonparametric model with Gaussian kernels fitted to headway data of Car–Car type, and the relative errors in evening and off-noon periods: (a) Car–Car, PDF, evening; (b) Car–Car, CDF, evening; (c) Van–Car, PDF, off-noon; (d) Van–Car, CDF, off-noon; (e) Car–Car, evening; (f) Van–Car, evening

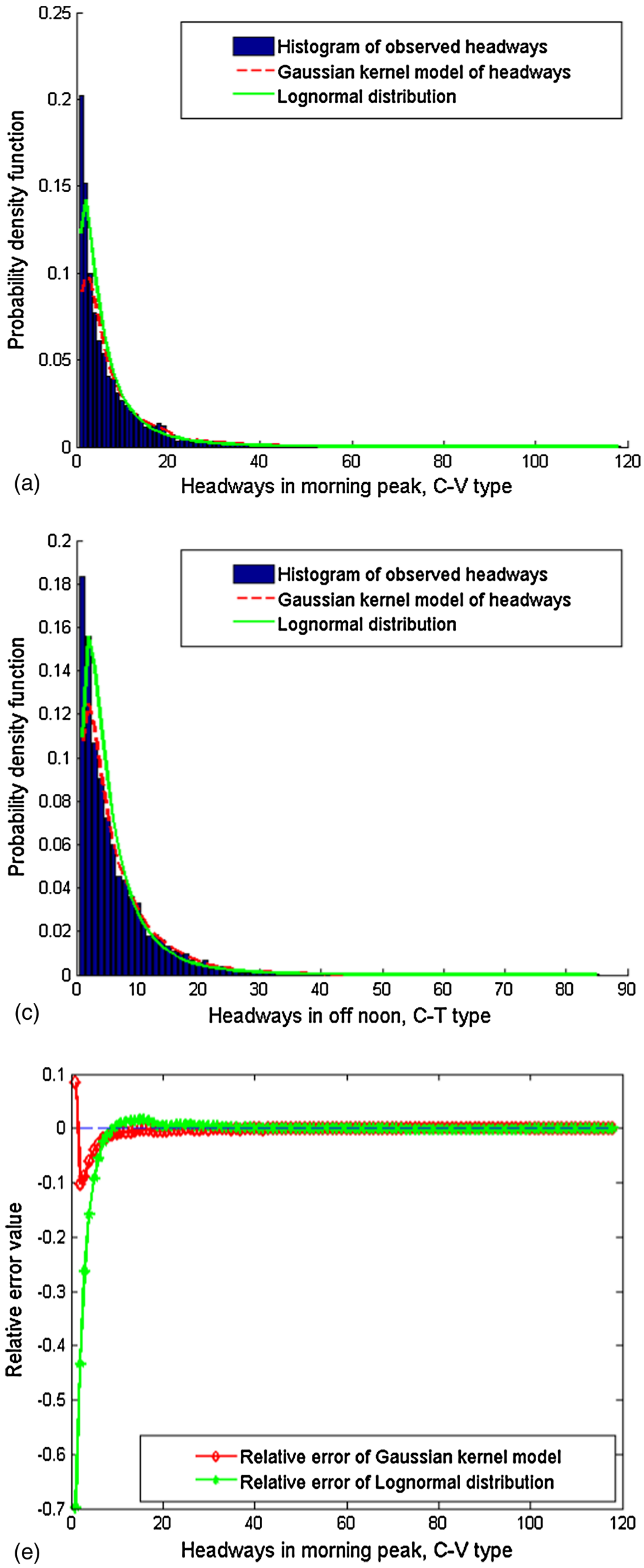

Fig. 5. Probability density functions and cumulative density functions of the lognormal distribution and the nonparametric model with Gaussian kernels fitted to headway data of Car–Van and Car–Truck types, and the relative errors in morning peak and off-noon periods: (a) Car–Van, PDF, morning peak; (b) Car–Van, CDF, morning peak; (c) Car–Truck, PDF, off-noon; (d) Car–Truck, CDF, off-noon; (e) Car–Van, evening peak; (f) Car–Truck, off-noon

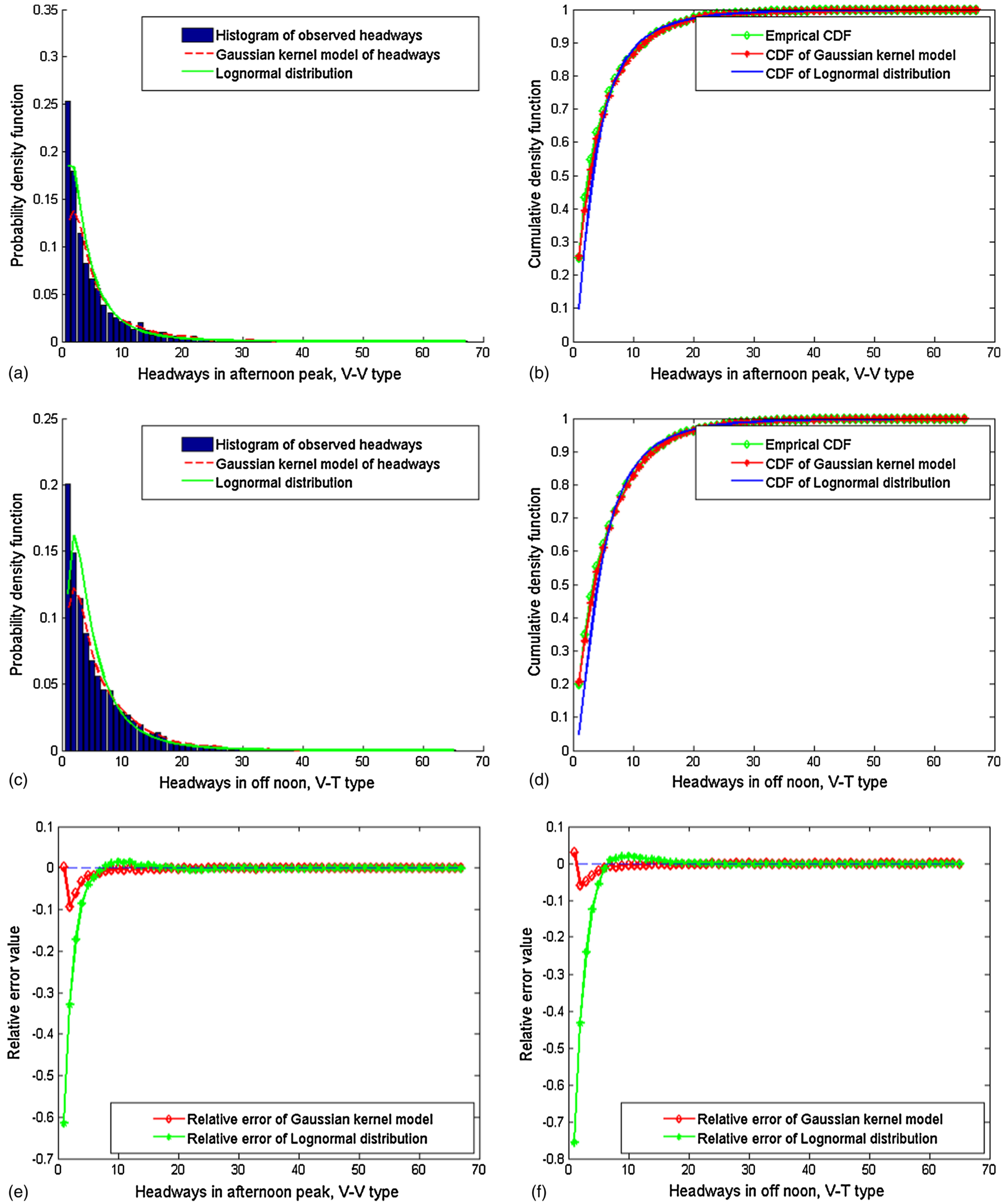

Fig. 6. Probability density functions and cumulative density functions of the lognormal distribution and the nonparametric model with Gaussian kernels fitted to headway data of Van–Van and Van–Truck types, and the relative errors in afternoon peak and off-noon periods: (a) Van–Van, PDF, afternoon peak; (b) Van–Van, CDF, afternoon peak; (c) Van–Truck, PDF, off-noon; (d) Van-Truck, CDF, off-noon; (e) Van–Van, afternoon peak; (f) Van–Truck, off-noon

Table 6. Comparison of K-S Test Results between Nonparametric and Parametric Models, Jacksonville

Method	Measurement	Result
Nonparametric model	K-S stats	0.8
with Gaussian kernel function	Hypothesis test	Reject
Parametric model	K-S stats	0.65
lognormal distribution	Hypothesis test	Reject

Table 7. Comparison of χ^2 Test Results between Nonparametric and Parametric Models, Jacksonville

Method	Measurement	Result
Nonparametric model	χ^2 stats	0.01425
with Gaussian kernel function	Hypothesis test	Not reject
Parametric model	χ^2 stats	0.05305
lognormal distribution	Hypothesis test	Not reject

where $LL_t(\beta_b)$ = log-likelihood for the model applied to the transfer data; $LL_t(\beta_t)$ = log-likelihood for the model estimated on the transfer data; and $\text{TTS} = \beta^2$ distributed with degree of freedom equal to the number of model parameters.

In this paper, two different datasets collected from different times and locations are utilized to examine the transferability of the nonparametric model. These two datasets were extracted from different traffic flows, Car–Car and Car–Van, which would further enhance the validity of the transferability test.

After calculation of the log-likelihood, model based on the Car–Van dataset reaches $LL_t(\beta_b = -469.8964)$ and the model based on Car–Car dataset achieves $LL_t(\beta_t) = -395.7796$. This yields $\text{TTS} = 148.2338$. The χ^2 statistic with 60 degree of freedom under a 99% confidence level is 91.952. Compared to the derived $\text{TTS} = 148.2338$, the result of the log-likelihood test provides statistically significant evidence, at the 99% confidence level, that the model could be employed in different datasets.

Summary, Conclusion, and Future Work

In this paper, a nonparametric model with a Gaussian kernel-based method is used to model the vehicle-type-specific headway distribution in a freeway work zone. This work adopts the FHWA Vehicle Classification scheme, classifying vehicle into three types: Car, Van, Truck. Due to the requirement of underlying behavior assumptions, a nonparametric model is pursued in this paper. The distribution with Gaussian kernel function is used because of its capability in handling complex interactions among vehicles in a freeway work zone setting as well as the fact that it requires no prior assumption about the data. A work zone on I-91 in Greenfield, Massachusetts is selected as the study site for this paper. A comparison between the Gaussian kernel-based model and lognormal model was presented by conducting a K-S test and Chi-square test.

The K-S test results indicate that the nonparametric model with Gaussian kernel performs better than the parametric model with lognormal distribution. In only one case (Car–Truck type headway in the evening period) does the nonparametric model reject the null hypothesis. This is partially due to the fact that there is a lower traffic volume in the observation period. In Chi-square test, the lower χ^2 value of the nonparametric distribution indicates its superiority in depicting headway over the parametric distribution. Visual comparisons by plotting empirical headway distribution are also presented. Both the PDF and CDF plots show the nonparametric model performs better than the parametric model in fitting the empirical headways. This observation is supported by the relative error curve. It is found the reason why the null hypothesis is rejected often by parametric models is mainly because the fluctuation of the relative error is large at the starting period. One possible explanation is that parametric distribution's assumption on the shape parameter lead to the inaccuracies in fitting the data. In order to further validate the results, a different work zone dataset from Jacksonville without specifying the vehicle type and time period was utilized to test both distributions. The K-S test rejected the null hypothesis while it did not reject it in vehicle-type-specific case, which suggests it is necessary to consider the vehicle type separately with different time periods. The following transferability test demonstrated the powerful ability of the nonparametric approach to be employed in different datasets.

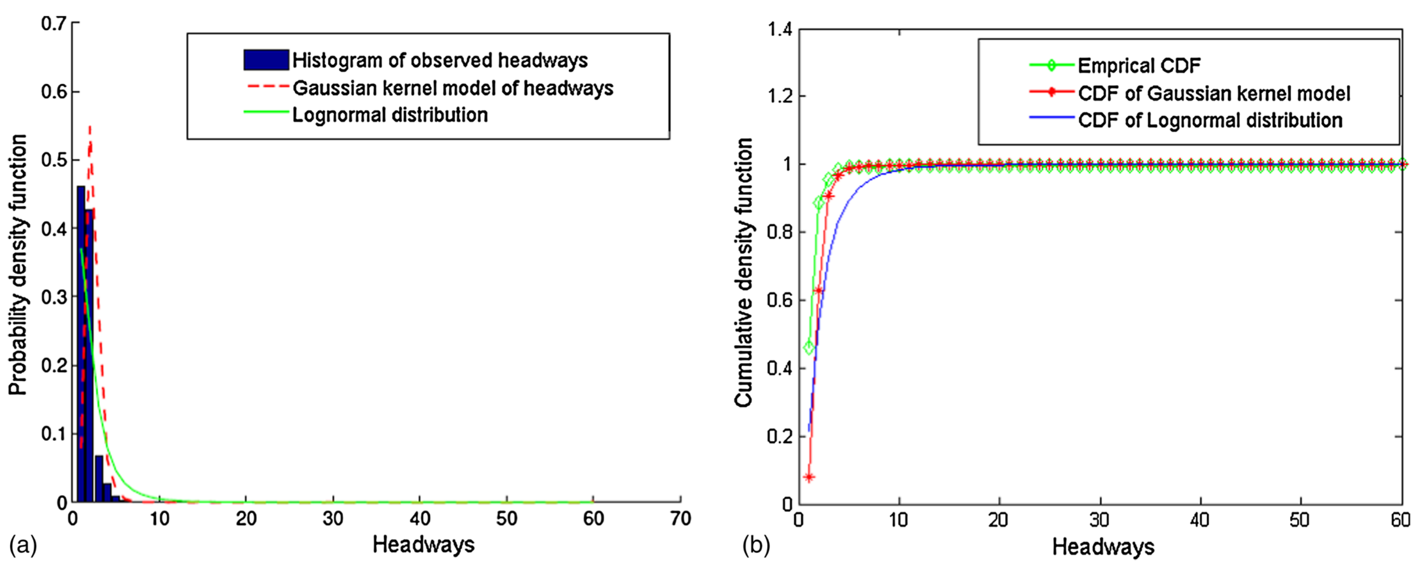

Fig. 7. Probability density functions and cumulative density functions of the lognormal distribution and the nonparametric model with Gaussian kernels fitted to headway data, and the relative errors of both methods

Although the statistical tests and numerical comparisons indicate the robustness of the nonparametric method in modeling work-zone vehicle-type-specific headway distributions, the duration of the data is not long enough to conclude that the nonparametric method is the natural candidate in this modeling endeavor. A more comprehensive work-zone traffic dataset is desired with different work zone scenarios from both interstate highways and arterials, such as pavement resurfacing, bridge rehabilitation, lane widening, or utility work. Other than the adopted Gaussian kernel function, other forms of kernel functions are worth investigating to evaluate the bias of statistical methods. In summary, this research adds strength to the literature in the area of modeling work-zone vehicle-type-specific headway distributions in terms of the merits of parametric and nonparametric methods.

References

Ai, Q., Yao, Z., Wei, H., and Li, Z. (2010). "Identifying characteristics of freeway traffic headway by vehicle types using video trajectory data." *ICCTP 2010: Integrated Transportation Systems: Green, Intelligent, Reliable*, 1984–1992.

Al-Ghamdi, A. S. (2001). "Analysis of time headways on urban roads: Case study from Riyadh." *J. Transp. Eng.*, 10.1061/(ASCE)0733-947X(2001)127:4(289), 289–294.

Branston, D. (1976). "Models of single lane time headway distributions." *Transp. Sci.*, 10(2), 125–148.

Chen, X., Li, L., and Zhang, Y. (2010). "A Markov model for headway/spacing distribution of road traffic." *IEEE Trans. Intell. Transp. Syst.*, 11(4), 773–785.

Cowan, R. J. (1975). "Useful headway models." *Transp. Res.*, 9(6), 371–375.

Dey, P. P., and Chandra, S. (2009). "Desired time gap and time headway in steady-state car-following on two-lane roads." *J. Transp. Eng.*, 10.1061/(ASCE)0733-947X(2009)135:10(687), 687–693.

Dong, S., Wang, H., Hurtwiz, D., Zhang, G., and Shi, J. (2015). "Nonparametric modeling of vehicle-type-specific headway distribution in freeway work zones." *J. Transp. Eng.*, in press.

FHWA. (2013). "Federal highway administration (fhwa) vehicle classification scheme." ⟨http://www.fhwa.dot.gov/⟩ (May 21, 2015).

Fox, J., Daly, A., Hess, S., and Miller, E. (2014). "Temporal transferability of models of mode-destination choice for the greater toronto and hamilton area." *J. Transp. Land Use*, 7(2), 41–62.

Greenberg, I. (1966). "The log normal distribution of headways." *Aust. Road Res.*, 2(7), 14–18.

Griffiths, J. D., and Hunt, J. G. (1991). "Vehicle headways in urban areas." *Traffic Eng. Control*, 32(10), 458–462.

Heaslip, K. P. (2007). "Modeling driver behavior in work zones: An evaluation of traffic flow impacts in freeway work zones with full lane closures." Ph.D. thesis, Univ. of Massachusetts Amherst, Amherst, MA.

Jang, J., Park, C., Kim, B., and Choi, N. (2011). "Modeling of time headway distribution on suburban arterial: Case study from south korea." *Procedia Soc. Behav. Sci.*, 16, 240–247.

Jiang, J., Lu, J., and Lv, Z. (2011). "Car-following behavior of different vehicles on ordinary highways." *Proc., Int. Conf. of Chinese Transportation Professionals*.

Jin, X., et al. (2009). "Departure headways at signalized intersections: A log-normal distribution model approach." *Transp. Res. Part C Emerg. Technol.*, 17(3), 318–327.

Johnson, N. L. (1949). "Systems of frequency curves generated by methods of translation." *Biometrika*, 36(1–2), 149–176.

Koppelman, F. S., and Wilmot, C. G. (1982). "Transferability analysis of disaggregate choice models." *HS-035 371*.

Pueboobpaphan, R., Park, D., Kim, Y., and Choo, S. (2013). "Time headway distribution of proble vehicles on single and multiple lane highways." *J. Civ. Eng.*, 17(4), 824–836.

Riccardo, R., and Massimiliano, G. (2012). "An empirical analysis of vehicle time headways on rural two-lane two-way roads." *Procedia Soc. Behav. Sci.*, 54(4), 865–874.

Sun, D., and Benekohal, R. F. (2005). "Analysis of work zone gaps and rear-end collision probability." *J. Transp. Stat.*, 8(2), 71.

Vasconcelos, L., Silva, A. B., Seco, l., and Silva, J. P. (2012). "Estimating the parameters of cowans m3 headway distribution for roundabout capacity analyses." *Baltic J. Road Bridge Eng.*, 7(4), 261–268.

Weng, J., Meng, Q., and Fwa, T. F. (2013). "Vehicle headway distribution in work zones." *Transportmetrica A Transp. Sci.*, 10(4), 1–19.

Ye, F., and Zhang, Y. (2009). "Vehicle type-specific headway analysis using freeway traffic data." *Transportation Research Record 2124*, Transportation Research Board, Washington, DC.

Yin, S., Li, Z., Zhang, Y., Yao, D., Su, Y., and Li, L. (2009). "Headway distribution modeling with regard to traffic status." *Intelligent Vehicles Symp.*, IEEE, 1057–1062.

Zhang, G., and Wang, Y. (2013). "A Gaussian kernel-based approach for modeling vehicle headway distributions." *Transp. Sci.*, 48(2), 206–216.

Zhang, G., Wang, Y., Wei, H., and Chen, Y. (2007). "Examining headway distribution models with urban freeway loop event data." *Transportation Research Record 1999*, Transportation Research Board, Washington, DC.

Zwahlen, H. T., Oner, E., and Suravaram, K. R. (2007). "Approximated headway distributions of free-flowing traffic on Ohio freeways for work zone traffic simulations." *Transportation Research Record 1999*, Transportation Research Board, Washington, DC.