Transportation Research Part C 70 (2016) 53–68

Contents lists available at ScienceDirect

Transportation Research Part C

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

A copula-based approach to accommodate the dependence among microscopic traffic variables

Yajie Zou [a,1], Yunlong Zhang [b,*]

[a] Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 201804, China
[b] Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, United States

ARTICLE INFO

Article history:
Received 13 May 2014
Received in revised form 3 September 2015
Accepted 11 November 2015
Available online 7 December 2015

Keywords:
Microscopic traffic variables
Speed
Headway
Copula
Multivariate distribution

ABSTRACT

Developing microscopic traffic simulation models requires the knowledge of probability distributions of microscopic traffic variables. Although previous studies have proposed extensive mathematical distributions for describing traffic variables (e.g., speed, headway, vehicle length, etc.), these studies usually consider microscopic traffic observations to be independent variables and distributions for these variables are investigated separately. As a result, some traditional approaches consider microscopic traffic variables as independent inputs to the traffic simulation process and these methods may ignore the possible dependence among different traffic variables.

The objectives of this paper are to investigate the dependence structure among microscopic traffic variables and to examine the applicability of the copula approach to the joint modeling of these variables. Copulas are functions that relate multivariate distribution functions of random variables to their one-dimensional marginal distribution functions. The concept of copulas has been well recognized in the statistics field and recently has been introduced in transportation studies. The proposed copula approach is applied to the 24-h traffic data collected on IH-35 in Austin, Texas. The preliminary data analysis indicates that there exists dependence among microscopic traffic variables. Moreover, the modeling and simulation results suggest that copula models can adequately accommodate and accurately reproduce the dependence structure revealed by the traffic observations. Overall, the findings in this paper provide a framework for generating multiple microscopic traffic variables simultaneously by considering their dependence.

1. Introduction

Developing microscopic traffic simulation models requires the knowledge of probability distributions of microscopic traffic variables (i.e., speed, time and distance headway, vehicle type, etc.). It is important to find appropriate mathematical distributions to describe the measured microscopic traffic variables, because the input parameters for some car-following models and microscopic simulation models are generated based on some form of mathematical models. For example, CORSIM provides three types of vehicle generation distributions: uniform distribution, normal distribution, and Erlang distribution. In VISSIM and SimTraffic, vehicles are usually generated on the basis of an exponential distribution (Tian et al., 2002). Schultz and Rilett (2004) used lognormal and normal distributions to generate input parameters for car-following sensitivity

* Corresponding author. Tel.: +1 979 845 9902; fax: +1 979 845 6481.
E-mail addresses: yajiezou@hotmail.com (Y. Zou), yzhang@civil.tamu.edu (Y. Zhang).
[1] Tel.: +86 13681865023.

http://dx.doi.org/10.1016/j.trc.2015.11.003

factors (i.e., headway) in Pitt car-following model. Traditionally, distributions for microscopic traffic variables are investigated separately and independent of each other. For speed distributions, normal (Leong, 1968; Mclean, 1979), log-normal (Haight and Mosher, 1962; Gerlough and Huber, 1976), gamma (Haight, 1963), normal mixture models (Ko and Guensler, 2005; Park et al., 2010) and skew-t mixture models (Zou and Zhang, 2011) have been considered. For headway distributions, exponential (Cowan, 1975), normal, gamma, lognormal, log–logistic (Yin et al., 2009), Cowan M3 model (Luttinen, 1999), M4 model (Hoogendoorn and Bovy, 1998), the generalized queuing model and the semi-Poisson model (Wasielewski, 1979) have been used. Compared with speed and headway, few researchers examined the distribution of vehicle lengths. Previously, Wang and Nihan (2004) and Zhang et al. (2008) used normal distributions to fit vehicle length data for short and long vehicles.

Despite a large number of previous works have examined the distributions of microscopic traffic variables, few studies have considered the possible dependence among microscopic traffic variables. For speed and headway, Luttinen (1992) found out that speed limit and road category have a considerable effect on the statistical properties of vehicle headways. Taieb-Maimon and Shinar (2001) conducted a field study to evaluate drivers' actual headways in car-following situation and their results showed that drivers adjusted the distance headways in relation to speed. Brackstone et al. (2009) revealed that there is a limited dependence of following headway on speed and the most successful relationship fit of headway and speed is an inverse relationship. Yin et al. (2009) studied the dependence of headway distributions on the traffic condition (speed pattern) and concluded that different headway models should be used for distinct traffic conditions (speed patterns). Zou et al. (2014) analyzed 24-h freeway speed and headway data and found that there was a weak correlation between speed and headway. In addition, vehicle type is known as an important factor in the car following situation. Previous studies (Ye and Zhang, 2009; Sarvi, 2013) showed that passenger cars usually travel further behind long vehicles than when following short vehicles and long vehicles also take longer time headways when following other vehicles due to their less agile operating characteristics. Ravishankar and Mathew (2011) also modified the Gipp's car-following model to incorporate vehicle-type dependent parameters. Wang et al. (2015) considered vehicle type and developed a game theoretic approach for predictive lane-changing and car-following control.

Some traditional microscopic simulation models consider vehicle speeds, arrival times and vehicle types as independent inputs to the simulation process. As a result, the same headway distribution may be assumed for different speed levels or vehicle types and this assumption may neglect the possible variability of headway distribution across speed or vehicle type. To overcome this potential problem, it is necessary to construct multivariate distributions to accommodate the dependence among microscopic traffic variables. One common way to jointly model several variables is the classical families of multivariate distribution (i.e., multivariate normal, lognormal and gamma distributions). However, the main limitation of this approach is that the individual behavior of the variables must then be characterized by the same parametric family of univariate distributions (Genest and Favre, 2007). Copula models, which can avoid this restriction, have been introduced in transportation studies (see, e.g., Bhat and Eluru (2009), Bhat and Sener (2009), Spissu et al. (2009), Eluru et al. (2010), Zou et al. (2014), Kim and Mahmassani (2014)). The objectives of this study are to (1) investigate the dependence structure between microscopic traffic variables (i.e., speed, headway and vehicle length); (2) examine the applicability of the copula approach to the joint modeling of these variables.

The remainder of the paper is organized as follows. In Section 2, we introduce the concept of copula and describe various types of copulas. In Section 3, we provide parameter estimation method for copula models and discuss different methods for generating samples from copulas. Section 4 presents the datasets used in this study and discusses the application of copulas in modeling microscopic traffic variables. Section 5 provides findings and conclusions.

2. Overview of the copula approach

2.1. Concept of copulas

The concept of copula was first proposed by Sklar (1959) and the interests in copulas and their application in the statistics field have grown over the last decades (see Genest and MacKay (1986), Genest and Rivest (1993), Nelsen (2006)). Recently, the copula method has received much attention from the finance, hydrological modeling, econometrics and transportation fields (see, Embrechts et al. (2002), Cherubini et al. (2004), Zhang and Singh (2006), Bhat and Eluru (2009)).

What are copulas? Copulas are functions that join or "couple" multivariate distribution functions to their one-dimensional marginal distribution functions (Nelsen, 2006). For continuous random variables X and Y, the Sklar's theorem (1959) stated that let $H(x, y)$ be a joint cumulative distribution function (cdf) with continuous marginal distributions $F(x)$ and $G(y)$, then there exists a bivariate copula C:

$$H(x, y) = C(F(x), G(y)) \tag{1}$$

where $C : [0, 1]^2 \rightarrow [0, 1]$ = copula.

A valid model for (X, Y) can be obtained from Eq. (1) if $F(x)$ and $G(y)$ are selected from parametric families of distributions. Moreover, a rich set of copula types C are available for generating the joint cdf $H(x, y)$. These copula types include the Gaussian copula, the Farlie–Gumbel–Morgenstern copula, and various Archimedean copulas. One advantage of the copula approach is that the selection of a model for representing X and Y can proceed independently from the choice of the marginal distributions (Genest and Favre, 2007).

For continuous distribution functions $F(x)$ and $G(y)$, the generalized inverse functions are defined by $F^{-}(t) = \inf\{x|F(x) \geqslant t\}$ and $G^{-}(t) = \inf\{y|G(x) \geqslant t\}$, respectively. Let $U = F(X)$ and $V = G(Y)$, then based on the probability integral transform, U and V are uniformly distributed random variables with support $[0, 1]$. We can obtain $F(x) = \Pr(X < x) = \Pr(F^{-1}(U) < x) = \Pr(U < F(x))$ and $G(y) = \Pr(Y < y) = \Pr(G^{-1}(V) < y) = \Pr(V < G(y))$.

Let $H(x, y)$ be a distribution function with continuous marginal distributions $F(x)$ and $G(y)$, then for any $u, v \in [0, 1]$, the copula function can be defined as (Nelsen, 2006):

$$C(u, v) = H(F^{-}(u), G^{-}(v)) \tag{2}$$

2.2. Measuring dependence

There are different ways to measure dependence. Some measures are scale-invariant (i.e., these measures remain unchanged under strictly increasing transformations of the random variables). Two widely known scale-invariant measures of association are Kendall's tau and Spearman's rho. Besides the Kendall's tau and Spearman's rho, one traditional correlation coefficient needs to be mentioned is the Pearson's product-moment correlation coefficient, which measures the linear dependence between random variables. Compared with the rank-based correlation, the linear correlation has the deficiency that it is not invariant under nonlinear strictly increasing transformations (Embrechts et al., 2002). Embrechts et al. (2002) also pointed out that for multivariate distributions which possess a simple closed-form copula, the moment-based correlations (i.e. Pearson's correlation coefficient) may be difficult to calculate and the determination of rank-based correlation (i.e., Kendall's tau and Spearman's rho) may be easier. Therefore, considering the advantages of rank-based correlation, the Kendall's tau and Spearman's rho are used to characterize the dependence structure for different types of copulas described in the following section.

2.3. Family of bivariate copulas

2.3.1. Bivariate Gaussian copulas

The Gaussian copula can be obtained using the inversion method. The 2-dimensional Gaussian copula with linear correlation matrix Σ is given by:

$$C_\Sigma(u, v) = \Phi_\Sigma(\Phi^{-1}(u), \Phi^{-1}(v)) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(-\frac{s^2 - 2\theta st + t^2}{2(1-\theta^2)}\right) dsdt \tag{3}$$

where $\Sigma = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$ is the correlation matrix, with parameter $\theta \in (-1, 1)$, Φ_Σ is a standard bivariate normal distribution and Φ is a standard normal distribution. If $\theta = 0$, the Gaussian copula becomes to the independent copula. Dependence parameter θ and Kendall's tau have the relationship, that is, $\tau = (2/\pi)\sin^{-1}(\theta)$. The 2-dimensional Gaussian copula density function is given by:

$$c_\Sigma(u, v) = \frac{1}{\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\omega^T\left(\Sigma^{-1} - I_2\right)\omega\right) \tag{4}$$

where $\omega^T = \left(\Phi^{-1}(u), \Phi^{-1}(v)\right), I_2$ is the 2×2 identity matrix.

2.3.2. The Farlie–Gumbel–Morgenstern copula

The joint cdf of a bivariate distribution constructed by the FGM copula can be described as follows:

$$C_\theta(u, v) = uv[1 + \theta(1-u)(1-v)] \tag{5}$$

where θ is a parameter of the copula function and for absolutely continuous marginal distributions, we need $|\theta| \leqslant 1$ (Schucany et al., 1978).

And the density of the FGM copula is provided by:

$$c_\theta(u, v) = [1 + \theta(2u-1)(2v-1)] \tag{6}$$

The FGM copula has the limitation that only if the correlation of two variables is weak, the FGM can provide an effective way for constructing a bivariate distribution. The correlation structure of FGM copula has been investigated for various continuous marginal distributions such as uniform, normal, exponential, gamma and Laplace distributions. For the rank-based dependence measures, Schucany et al. (1978) showed that, regardless of the forms of marginal distributions, θ and concordance-based correlation (Kendall's tau $\tau_{X,Y}$ and Spearman's rho $\rho_{X,Y}$) satisfy the following equations:

$$\tau_{X,Y} = \frac{2}{9}\theta \tag{7}$$

$$\rho_{X,Y} = \frac{\theta}{3} \tag{8}$$

Since θ is in $[-1, 1]$, the FGM copula can allow weak positive and negative dependence and $\tau_{X,Y}$ and $\rho_{X,Y}$ are bounded on $[-\frac{2}{9}, \frac{2}{9}]$ and $[-\frac{1}{3}, \frac{1}{3}]$, respectively.

2.3.3. Bivariate Archimedean copulas

Archimedean copulas are important class of copulas and these copulas are widely applied for a few reasons: (1) Archimedean copulas have a simple and explicit form expression; (2) they are characterized by a single parameter function φ that meets certain requirements; (3) a variety of families of copulas which belong to this class. Archimedean copulas were introduced by Genest and MacKay (1986). One parameter Archimedean copulas are briefly introduced in the following paragraph, further details can be found in Nelsen (2006).

As defined in Nelsen (2006), let φ be a continuous, strictly decreasing function from $[0, 1]$ to $[0, \infty]$ such that $\varphi(1) = 0$. The pseudo-inverse of φ is the function $\varphi^{[-1]} : [0, \infty] \to [0, 1]$ such that $\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & 0 \leqslant t \leqslant \varphi(0) \\ 0 & \varphi(0) \leqslant t \leqslant \infty \end{cases}$. If we assume $\varphi(0) = \infty$, then $\varphi^{[-1]} = \varphi^{-1}$, and we have $\varphi(\varphi^{[-1]}(t)) = t$. Using functions φ and φ^{-1}, the definition of one parameter Archimedean copulas is given as:

$$C_\theta(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \tag{9}$$

The function φ is called a generator of the copula. When $\varphi(0) = \infty$, φ is said to be a strict generator and $C_\theta(u, v)$ in Eq. (9) is a strict Archimedean copula. In the following paragraphs, several well-known one-parameter families of Archimedean copulas, along with their generators are described.

2.3.3.1. Ali–Mikhail–Haq copula. The Ali–Mikhail–Haq copula, proposed by Ali et al. (1978), can allow for weak positive and negative dependence. The generator function is $\varphi(t) = \ln \frac{1-\theta(1-t)}{t}$, with $\theta \in [-1, 1)$, and the corresponding Ali–Mikhail–Haq copula function is as follows:

$$C_\theta(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)} \tag{10}$$

Kendall's tau is related to θ by $\tau = \frac{3\theta-2}{3\theta} - \frac{2(1-\theta)^2}{3\theta^2} \ln(1-\theta)$, so that $-0.182 < \tau < 0.333$. The density function of Ali–Mikhail–Haq copula is given by (Hofert et al., 2012):

$$c_\theta(u, v) = \frac{(1-\theta)^3}{\theta^2} \frac{h_\theta^A(u, v)}{u^2 v^2} Li_{-2}\left\{h_\theta^A(u, v)\right\} \tag{11}$$

where $h_\theta^A(u, v) = \theta \frac{u}{1-\theta(1-u)} \frac{v}{1-\theta(1-v)}$ and $Li_s(z) = \sum_{k=1}^{\infty} z^k / k^s$.

2.3.3.2. The Clayton copula. If the generator function is selected as $\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$, with $\theta \in (0, \infty)$, the Archimedean copula is called the Clayton copula. It is given by:

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \tag{12}$$

The Clayton copula was first proposed by Clayton (1978) and allows only positive dependence. Kendall's tau is related to θ by $\tau = \frac{\theta}{\theta+2}$, so that $0 < \tau < 1$. If θ tends to 0, the Clayton copula becomes independent copula. The density function of Clayton copula is given by (Hofert et al., 2012):

$$c_\theta(u, v) = (1+\theta) u^{-(\theta+1)} v^{-(\theta+1)} (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta - 2} \tag{13}$$

2.3.3.3. The Frank copula. If we choose $\varphi(t) = -\ln \frac{e^{-\theta t}-1}{e^{-\theta}-1}$, with $\theta \in (-\infty, \infty) \setminus \{0\}$, the Archimedean copula is called the Frank copula. It is given by:

$$C_\theta(u, v) = -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right) \tag{14}$$

with $\tau = 1 - \frac{4}{\theta}[1 - D_1(\theta)]$, where $D_1(\theta)$ is the first order Debye function $D_k(\theta)$ which is defined as $D_k(\theta) = \frac{k}{\theta^k} \int_0^\theta \frac{t^k}{e^t - 1} dt$.

The Frank copula was proposed by Frank (1979). It can allow for both positive and negative dependence. The range of τ is $(-1, 1)$ and if θ tends to 0, the Frank copula becomes independent copula. The density function of Frank copula is given by (Hofert et al., 2012):

$$c_\theta(u,v) = \left(\frac{\theta}{1-e^{-\theta}}\right) Li_{-1}\left\{h_\theta^F(u,v)\right\} \frac{\exp(-\theta(u+v))}{h_\theta^F(u,v)} \tag{15}$$

where $h_\theta^F(u,v) = (1-e^{-\theta})^{-1}(1-\exp(-\theta u))(1-\exp(-\theta v))$.

2.3.3.4. The Gumbel copula. The Gumbel copula, also known as the Gumbel–Hougaard copula, was first introduced by Gumbel (1960). The generator function for this copula is $\varphi(t) = (-\ln t)^\theta$, and the corresponding copula function is

$$C_\theta(u,v) = \exp\left(-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\right) \tag{16}$$

The Gumbel copula only accommodates positive dependence and Kendall's tau is related to θ by $\tau = 1-\theta^{-1}$, so that $0 < \tau < 1$. If $\theta = 1$, the Gumbel copula becomes independent copula. The density function of Gumbel copula is given by (Hofert et al., 2012):

$$c_\theta(u,v) = \theta^2 \exp\left\{-t_\theta(u,v)^\alpha\right\} \frac{(-\ln u)^{\theta-1}(-\ln v)^{\theta-1}}{t_\theta(u,v)^2 uv} P_{2,\alpha}^G(t_\theta(u,v)^\alpha) \tag{17}$$

where $\alpha = 1/\theta, P_{2,\alpha}^G(x) = \sum_{k=1}^2 \varepsilon_{2k}^G(\alpha)x^k$, and $\varepsilon_{2k}^G(\alpha) = \frac{2}{k!}\sum_{j=1}^k \binom{k}{j}\binom{\alpha j}{2}(-1)^{2-j}$.

2.3.3.5. The Joe copula. The Joe copula, discussed by Joe (1993, 1997), has a generator function $\varphi(t) = -\ln[1-(1-t)^\theta]$. The Joe copula is defined as:

$$C_\theta(u,v) = 1 - \left[(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\right]^{1/\theta} \tag{18}$$

with $\tau = 1 + \frac{4}{\theta}D_J(\theta)$, where $D_J(\theta) = \int_{t=0}^1 \frac{[\ln(1-t^\theta)](1-t^\theta)}{t^{\theta-1}}dt$.

Like the Clayton and Gumbel copulas, the Joe copula cannot account for negative dependence. The range of τ is $(0,1)$. If θ tends to 0, the Joe copula becomes independent copula. The density function of Joe copula is given by (Hofert et al., 2012):

$$c_\theta(u,v) = \theta \frac{(1-u)^{\theta-1}(1-v)^{\theta-1}}{\left\{1-h_\theta^J(u,v)\right\}^{1-\alpha}} P_{2,\alpha}^J\left\{\frac{h_\theta^J(u,v)}{1-h_\theta^J(u,v)}\right\} \tag{19}$$

where $\alpha = 1/\theta, h_\theta^J(u,v) = \left\{1-(1-u)^\theta\right\}\left\{1-(1-v)^\theta\right\}, P_{2,\alpha}^J(x) = \sum_{k=0}^1 \varepsilon_{2k}^J(\alpha)x^k, \varepsilon_{2k}^J(\alpha) = S(2,k+1)\frac{\Gamma(k+1-\alpha)}{\Gamma(1-\alpha)}$ and $S(j,k)$ is the Stirling numbers of the second kind.

2.4. Multivariate Gaussian copulas

The copula of the n-variate normal distribution with $n \times n$ correlation matrix P is

$$C_P(\mathbf{u}) = \Phi_P(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)) \tag{20}$$

where Φ_P represents the joint distribution function of the n-variate standard normal distribution function with correlation matrix P, and Φ^{-1} is the inverse of the distribution function of the univariate standard normal distribution. For the multivariate Gaussian copula, correlation matrix P and Kendall's tau have the relationship, that is $\tau_{X_i,X_j} = \frac{2}{\pi}\sin^{-1}(\rho_{ij})$ (Embrechts et al., 2003; Demarta and McNeil, 2005).

In the trivariate case the copula expression can be written as

$$C_P(u_1,u_2,u_3) = \int_{-\infty}^{\Phi^{-1}(u_1)}\int_{-\infty}^{\Phi^{-1}(u_2)}\int_{-\infty}^{\Phi^{-1}(u_3)} \frac{1}{(2\pi)^{3/2}|P|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{w}^T P^{-1}\mathbf{w}\right)d\mathbf{w} \tag{21}$$

where $P = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$ is the symmetrical correlation matrix with $-1 \leqslant \rho_{ij} \leqslant 1$ $(i,j = 1,2,3)$; $\mathbf{w} = (w_1,w_2,w_3)^T$ represents the corresponding integral variables.

Besides the multivariate Gaussian copula, multivariate Archimedean copulas are also widely used for modeling multivariate distribution of multiple random variables. The multivariate Archimedean copulas include the symmetric Archimedean copula and the asymmetric Archimedean copula (which is also called nested Archimedean copula). Note that the symmetric Archimedean copula is a special case of the asymmetric Archimedean copula. The symmetric Archimedean copula suffers from a very limited dependence structure since all k-margins are identical; they are distribution functions of n exchangeable $U(0,1)$ random variables (Embrechts et al., 2003). As a consequence of this exchangeability property, all mutual dependences among variables are modeled by only one Archimedean 2-copula (Grimaldi and Serinaldi (2006). On the other hand, the asymmetric Archimedean copula allows for nonexchangeability and a part of all possible mutual dependences can be modeled in a different way. For more details about multivariate Archimedean copulas, interested readers can see Grimaldi and Serinaldi (2006). Compared with the multivariate Gaussian copulas which are able to model all range of dependence, the multivariate Archimedean copula families ($n \geqslant 3$) can model only positive dependence. Thus, considering the possible inverse relationship between speed and headway, only multivariate Gaussian copulas are considered.

3. Estimation of θ and random variate generation

Given a parametric family (C_θ) of copulas and a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from continuous random variables (X, Y), the first step is to select appropriate marginal distributions for each variable. Then the data can be transformed onto the copula scale using the probability integral transform. The next step is to estimate θ. Genest and Favre (2007) reviewed various nonparametric methods for estimating θ and they recommend using ranked-based estimators since the ranks of the observations are the best summary of the joint behavior of the random pairs. Two straightforward estimators are based on Kendall's Tau and Spearman's Rho. These two rank-based estimators are explained in the following example.

If the dependence structure of a random pair (X, Y) can be appropriately modeled by the FGM copula described in Eq. (5). Thus, as discussed above, there exist relations between the parameter θ and Kendall's Tau and Spearman's Rho, which are

$$\tau_{X,Y} = \frac{2}{9}\theta \tag{22}$$

$$\rho_{X,Y} = \frac{\theta}{3} \tag{23}$$

Since $\tau_{X,Y}$ and $\rho_{X,Y}$ can be computed from the sample pairs, a simple and intuitive approach to estimating θ would be

$$\widehat{\theta} = \frac{9}{2}\tau_{X,Y} \tag{24}$$

$$\widehat{\theta} = 3\rho_{X,Y} \tag{25}$$

$\tau_{X,Y}$ and $\rho_{X,Y}$ are rank-based, and this estimation strategy may be seen as a nonparametric adaptation of the method of moments (Genest and Favre, 2007).

Another method for estimating θ is called the method of maximum pseudolikelihood, which requires that C_θ be absolutely continuous with density c_θ. Compared with Kendall's Tau and Spearman's Rho, the maximum pseudolikelihood estimator has the advantage that it does not require the dependence parameter θ to be real. However, this method also involves a lot of numerical work and requires the existence of a density c_θ. Thus, for simplicity, the Kendall's Tau based estimator is adopted in this research. For detailed procedure of using the maximum pseudolikelihood estimator, see Genest et al. (1995). Note that Joe (1997, Chap. 10) also introduced a parametric two-step procedure referred to the inference from margins (IFM) method for estimating θ. Kim et al. (2007) pointed out that the IFM estimator depends on the choice of margins, and may run the risk of being unduly affected if selection of the margins turn out to be inappropriate.

One of the primary applications of copulas is in simulation and Monte Carlo studies (Nelsen, 2006). Based on Sklar's theorem, the copula can be used as a tool for generating observations (x, y) of a pair of random variables (X, Y) from copula function C_θ with marginal distributions $F(x)$ and $G(y)$. Specifically, we need to generate uniform random variates (u, v) from the desired copula C_θ, and then use the inverse distribution function method to transform the data, $(x, y) = (F^{-1}(u), G^{-1}(v))$. Three algorithms described in the previous studies for copula simulation are conditional distribution method, sampling algorithm for Gaussian copulas and sampling algorithm for Archimedean copulas (for the complete description of the three algorithms, see (Nelsen, 2006)).

4. Application

4.1. Dependence among microscopic traffic variables

The dataset used in this study was collected at a location on IH-35 in Austin, Texas. IH-35 has four lanes in the southbound direction and the free flow speed is 60 mile/h (or 96.56 km/h) for all types of vehicles. Due to the heavy traffic demand and a large volume of heavy vehicles, the data collection site is typically congested during the morning and afternoon peak hours. The data were collected using a Peek ADR 6000 detector, which includes two inductive loops embedded in the pavement. The detector generates accurate outputs of arrival time, presence time, speed, length, and classification for each individual vehicle (Ye et al., 2006). This dataset was extensively analyzed in the previous studies (Ye and Zhang, 2009; Park et al., 2010; Zou and Zhang, 2011). The data have 27,920 vehicles with recorded speed values, arrival times and vehicle lengths in a 24-h period (from 00:00 to 24:00, December 11, 2004), including 24,011 (86%) passenger vehicles and 3909 (14%) heavy vehicles. For this dataset, the headway value between two consecutive vehicles is the elapsed time between the arrivals of a pair of vehicles. The arrival times were recorded in second (s); the observed speeds were recorded in m/s; and the vehicle

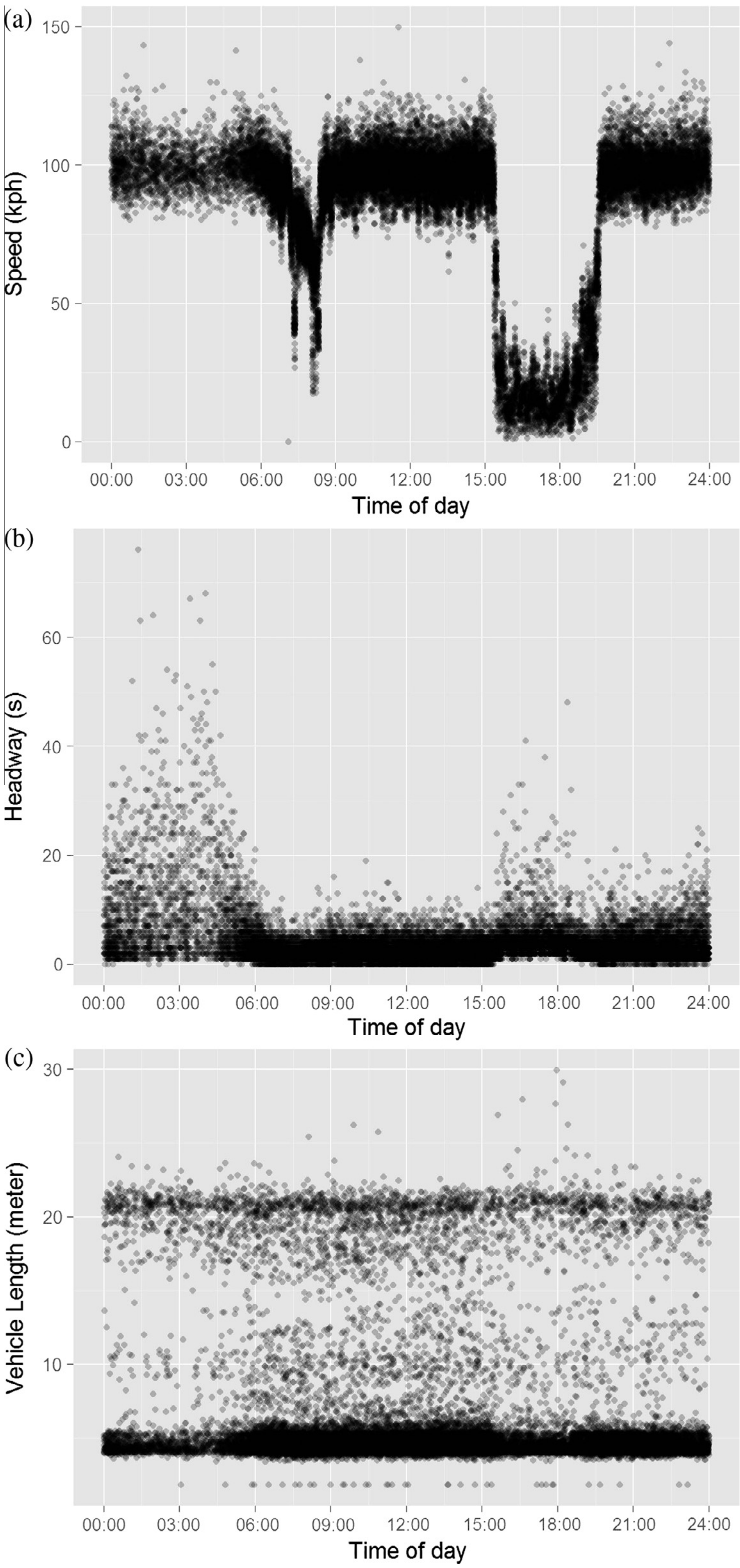

Fig. 1. (a) Speed scatter plot by time of day; (b) headway scatter plot by time of day; (c) vehicle length scatter plot by time of day.

lengths were recorded in meter (m). Fig. 1(a–c) display the scatter plots of speed, headway and vehicle length by time of day. Because of large samples in the dataset, semi-transparent points are used to alleviate some of the over-plotting in Fig. 1. Fig. 1(c) indicates that the observed vehicles seem to consist of two sub-populations: one at about 5 m, representing passenger vehicles, and the other at about 22 m, representing trucks and buses.

We first examine their dependence structure among three traffic variables. Since the 24-h traffic data collected on IH-35 consist of distinct traffic flow conditions, it is possible that the dependence structure between traffic variables may vary depending on the traffic condition. Thus, we first evaluate the hourly dependence among speed, headway and vehicle length for the 24-h period. For each hour, Kendall's tau τ, and Spearman's rho ρ_S are used to measure the dependence. The computed values of Kendall's tau τ, and Spearman's rho ρ_S for each of the 24-h are given in Table 1. As shown in Table 1 below, the dependence structure among three traffic variables exhibits different characteristics. First, for speed and headway, the dependence structure is stable under the same traffic condition, but change significantly between different traffic conditions. Generally speaking, for the off-peak period, when the flow rate is below 1000 vehicles/h (i.e., 00:00–06:00 and 23:00–24:00), Kendall's tau values indicate that speed and headway have negligible effect on each other; when the flow rate is above 1000 vehicles/h (i.e., 06:00–07:00, 09:00–15:00 and 20:00–23:00), τ ranges between 0.08 and 0.15 and speed and headway have a very weak positive correlation. On the other hand, for the peak period, when the flow rate is below 1000 vehicles/h (i.e., 16:00–19:00), speed and headway have a weak negative dependence. Note that compared to the afternoon peak period (most speed values are below 40 kph from 16:00 to 19:00), the morning peak period (a large portion of speed values are above 50 kph 7:00–8:00) has a different correlation relationship. Second, Kendall's tau and Spearman's rho values indicate that there exists a very limited relationship between speed and vehicle length. Third, headway and vehicle length have the strongest dependence during the afternoon peak period (i.e., 16:00–19:00). For the copula modeling approach, parameter θ is related to Kendall's tau and it is assumed to be fixed. Thus, this modeling approach cannot capture the varying characteristics of dependence structure between speed and headway. However, under the same traffic condition, the dependence structure among speed, headway and vehicle length is quite stable. In the following section, the traffic data observed under the congested traffic condition (from 16:00 to 19:00) are considered to demonstrate the usefulness of copula methods for constructing bivariate models. This is because the relationship between speed and headway and the influence of vehicle length on headway is more obvious in the car following situation. Fig. 2(a–d) show the scatter plots of speed, headway and vehicle length for the time period from 16:00 to 19:00. Note that there were 2360 vehicles observed between 16:00 and 19:00.

We further examine the dependence among speed, headway and vehicle length using the chi-plot which was proposed by Fisher and Switzer (2001). The chi-plot depends on the data through the values of their ranks and it is defined as follows:

$$H_i = \frac{1}{n-1}\#\{j \neq i : X_j \leqslant X_i, Y_j \leqslant Y_i\} \tag{26}$$

$$F_i = \frac{1}{n-1}\#\{j \neq i : X_j \leqslant X_i\} \tag{27}$$

Table 1
Hourly dependence among speed, headway and vehicle length for the 24-h period.

Time period	Count (vehicles)	Speed and headway		Speed and vehicle length		Headway and vehicle length	
		τ	ρ_S	τ	ρ_S	τ	ρ_S
0–1	457	−0.02	−0.03	−0.02	−0.03	−0.02	−0.02
1–2	354	−0.01	−0.02	−0.08	−0.11	−0.02	−0.03
2–3	301	−0.01	−0.01	−0.12	−0.18	−0.03	−0.04
3–4	277	−0.05	−0.08	−0.06	−0.08	−0.03	−0.05
4–5	346	−0.02	−0.02	−0.01	−0.01	−0.04	−0.06
5–6	709	0.05	0.07	−0.11	-0.16	0.03	0.04
6–7	1594	0.15	0.21	−0.02	−0.03	0.08	0.11
7–8	2039	0.04	0.05	0.02	0.03	0.10	0.13
8–9	1851	0.05	0.07	0.02	0.02	0.09	0.12
9–10	1701	0.11	0.15	−0.04	−0.05	0.03	0.04
10–11	1653	0.13	0.17	−0.06	−0.10	0.06	0.08
11–12	1707	0.10	0.13	−0.03	−0.05	0.09	0.12
12–13	1748	0.11	0.15	−0.06	−0.08	0.08	0.11
13–14	1739	0.11	0.15	−0.02	−0.03	0.04	0.05
14–15	1722	0.12	0.16	−0.01	−0.01	0.11	0.14
15–16	1295	−0.35	−0.46	0.00	0.00	0.07	0.09
16–17	755	−0.34	−0.45	0.01	0.01	0.14	0.19
17–18	676	−0.33	−0.45	−0.01	−0.01	0.14	0.19
18–19	929	−0.36	−0.49	0.02	0.04	0.13	0.17
19–20	1446	−0.11	−0.13	0.01	0.01	0.12	0.16
20–21	1241	0.11	0.15	−0.03	−0.04	0.04	0.06
21–22	1267	0.08	0.11	−0.01	−0.01	0.05	0.07
22–23	1185	0.09	0.12	−0.01	−0.02	0.05	0.07
23–24	927	0.03	0.04	−0.03	−0.05	0.01	0.02

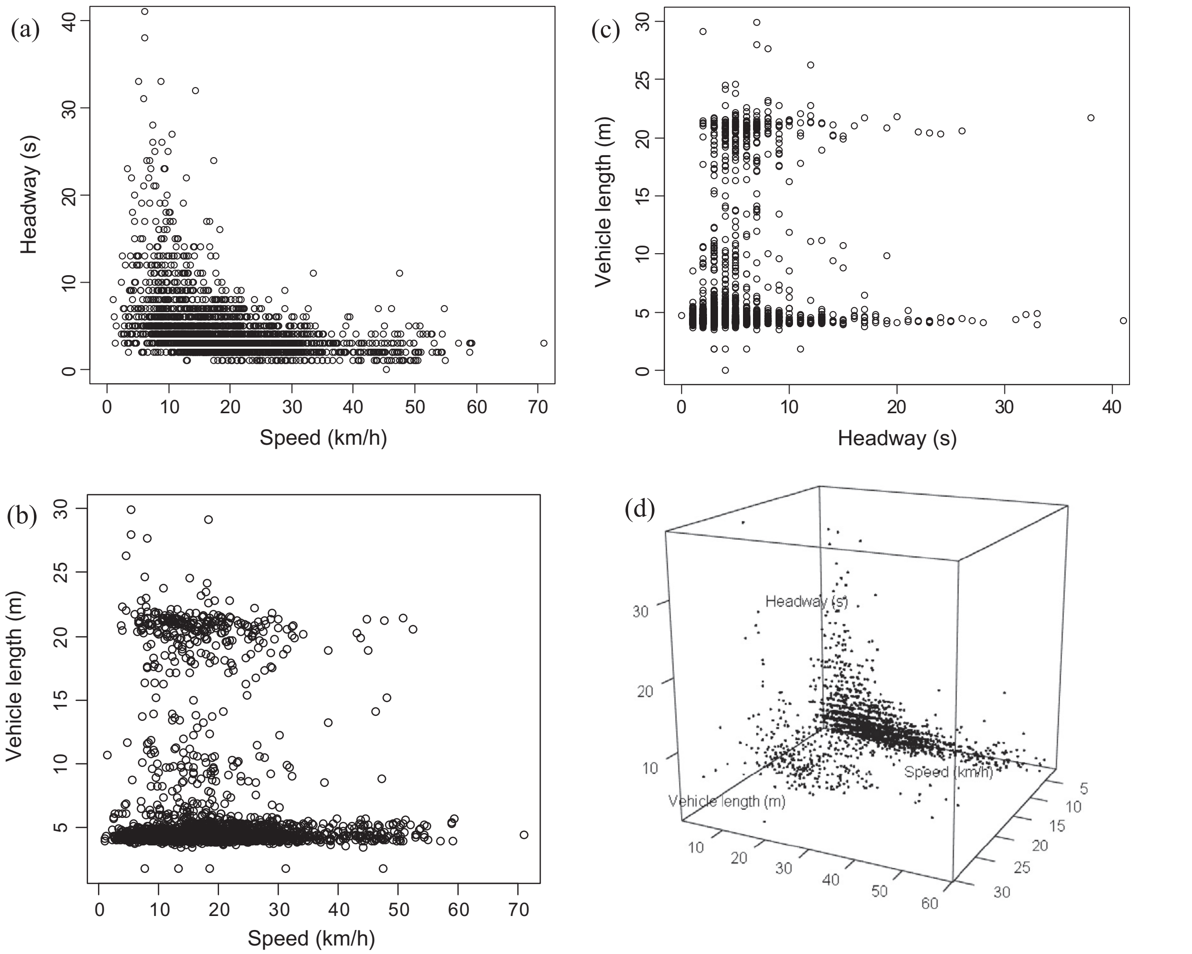

Fig. 2. Scatter plot of (a) speed and headway; (b) speed and vehicle length; (c) headway and vehicle length; (d) speed, headway and vehicle length for time period from 16:00 to 19:00.

and

$$G_i = \frac{1}{n-1}\#\{j \neq i : Y_j \leqslant Y_i\} \tag{28}$$

The above quantities depend exclusively on the ranks of the observations. A chi-plot is a scatter plot of the pairs (λ_i, χ_i), where

$$\chi_i = \frac{H_i - F_i G_i}{\sqrt{F_i(1-F_i)G_i(1-G_i)}}$$

and

$$\lambda_i = 4\text{sign}\{(F_i - 1/2)(G_i - 1/2)\} \max\left\{(F_i - 1/2)^2, (G_i - 1/2)^2\right\}.$$

To avoid outliers, Fisher and Switzer (2001) recommend that $|\lambda_i| \leqslant 4\left(\frac{1}{n-1} - \frac{1}{2}\right)^2$. Fig. 3(a–c) show the chi-plots for the traffic data observed from 16:00 to 19:00. Dashed blue lines are the 95% confidence band and values of χ_i measure the degree of departures from the hypothesis that speed and headway are independent. As shown in Fig. 3(a), almost all points lie below the 95% probability region and this confirms the presence of negative association between speed and headway. For speed and vehicle length, Fig. 3(b) shows that many data points are within the dashed blue lines and the remaining points are either above or below the 95% probability region. Since the area inside the confidence interval means independent, the finding from Fig. 3(b) is consistent with the results reported in Table 1 that the evidence in support of the dependence between speed and vehicle length is generally lacking. Fig. 3(c) demonstrates that most points are lying above the 95% probability region and while some of the points fall inside the confidence band. This pattern corroborates the presence of positive association between headway and vehicle length.

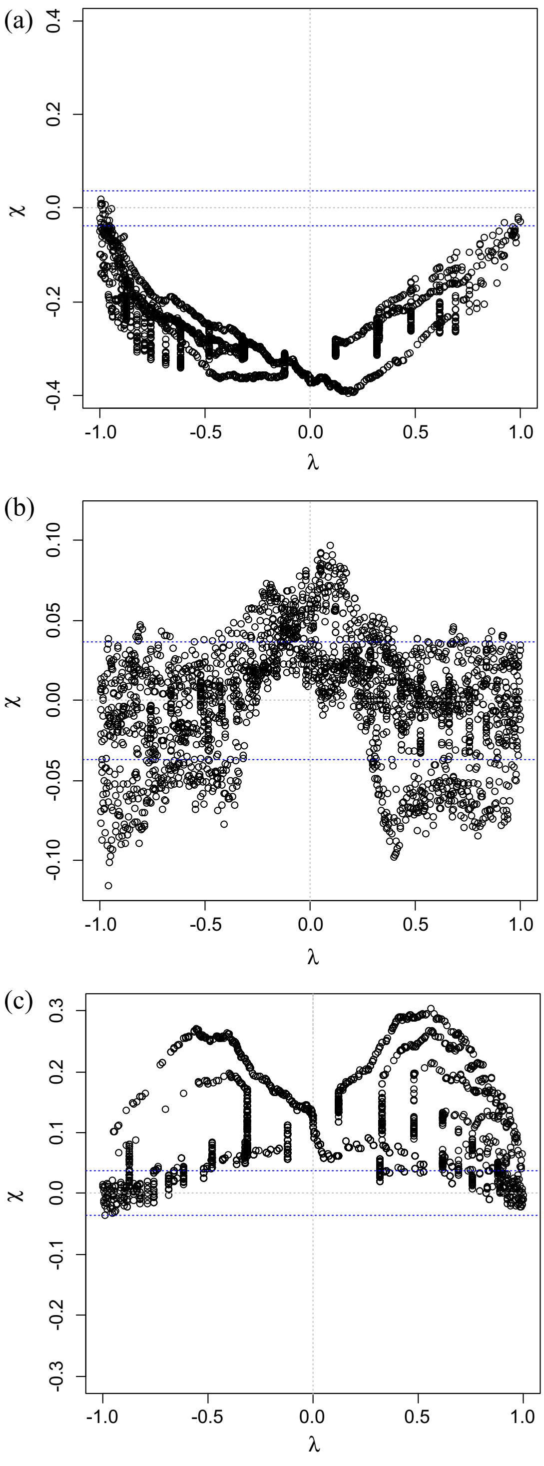

Fig. 3. Chi-plot for (a) speed and headway; (b) speed and vehicle length; (c) headway and vehicle length.

4.2. Marginal distribution

In selecting the marginal distributions, we model speed using normal, log-normal, skew-normal and skew-t distributions. Normal and lognormal distributions are widely used to describe observed speed. A recent study (Zou and Zhang, 2011) examined the fitting performance of skew-normal and skew-t distributions and their findings suggest that compared to normal distribution, skew-normal or skew-t distribution can be a better alternative for describing the speed data. For headway data, three commonly used single distribution models are investigated: gamma, lognormal and log–logistic distributions. Considering the excess skewness, kurtosis and bimodality present in vehicle length distribution, three mixture models are selected, which are 2-component normal mixture distribution, 2-component skew-normal mixture distribution and 2-component skew-t mixture distribution. The parameters were estimated by the maximum likelihood method. The best fitted distributions for speed, headway and vehicle length were selected using log-likelihood, the Akaike information criterion (AIC) and root mean square error (RMSE) values. Table 2 reports the log-likelihood, AIC and RMSE values of different speed, headway and vehicle length models. Larger log-likelihood and smaller AIC and RMSE values indicate a better overall fit. For the speed data, the skew-t model is better than other models in term of goodness of fit index and normal model provide the least fitting result. In the meantime, the headway data were examined using gamma, lognormal and log–logistic models. The performance of headway models is not consistent. Based on the results, the log–logistic model has the highest log-likelihood and lowest AIC and RMSE values and the gamma model provides the least satisfactory fitting performance. As discussed above, the bimodality of the vehicle length distribution indicates the presence of 2 different clusters. Thus, 2-component mixture distributions were used. The fitting results illustrate that the 2-component skew-t distribution can provide a more accurate description of the bimodal vehicle length distribution than the other two mixture models. Thus, the skew-t, log–logistic and 2-component skew-t distributions are selected as the marginal distributions for describing speed, headway and vehicle length, respectively.

4.3. Optimal copula model selection

In this section, we modeled the dependence between speed and headway, and headway and vehicle length using different families of copulas. Note that speed and vehicle length are assumed to be independent due to lack of evidence to support the association between each other. The possible explanation is that cars and trucks have the same speed limit on IH-35. The

Table 2
Log-likelihood, AIC and RMSE values of different fitted probability distributions for each traffic variable.

Traffic variable	Fitted marginal distributions	Log-likelihood	AIC	RMSE
Speed	Normal	−8751.15	17506.30	14.75
	Log-normal	−8521.78	17047.56	11.30
	Skew-normal	−8495.50	16997.00	9.69
	Skew-t	−8476.43	16960.86	8.01
Headway	Log-normal	−5250.48	10504.95	23.06
	Gamma	−5488.46	10980.93	48.51
	Log–logistic	−5193.94	10391.89	16.74
Vehicle length	2-component normal	−3492.39	6996.78	28.08
	2-component skew-normal	−3262.86	6541.72	29.10
	2-component skew-t	−3035.33	6090.67	28.30

Table 3
The estimation of Kendall's tau τ and parameter θ of different copulas.

	τ	Gaussian	FGM	Gumbel	Clayton	Ali–Mikhail–Haq	Frank	Joe
Speed and headway	−0.37	−0.55	NA[a]	1.59[b]	1.19[b]	NA[a]	−3.80	2.08[b]
Headway and length	0.13	0.21	0.59	1.15	0.30	0.51	1.21	1.27

[a] NA means that the parameter θ for that copula is not applicable. This is because for the FGM copula, it can model the correlated random variables with $-2/9 \leqslant \tau \leqslant 2/9$; for the Ali–Mikhail–Haq copula, it can model the correlated random variables with $-0.182 < \tau < 0.333$.
[b] For Gumbel, Claytion and Joe copulas, since they can only model positive correlated random variables (i.e., Kendall's tau $\tau > 0$), we used a new variable $\text{speed}_{new} = -\text{speed}$. The speed_{new} and headway show a positive correlation, Kendall's tau = 0.37.

Table 4
The log-likelihood, AIC and RMSE values of different copulas.

	Goodness-of-fit statistics	Gaussian	FGM	Gumbel	Clayton	Ali–Mikhail–Hq	Frank	Joe
Speed and headway	LL[a]	−13305.72	NA	−13407.06	−13470.51	NA	−13315.74	−13613.04
	AIC	26625.44	NA	26828.12	26955.02	NA	26645.48	27240.08
	RMSE	1.08	NA	1.24	1.44	NA	1.12	1.92
Headway and length	LL	−8179.75	−8180.72	−8182.55	−8206.01	−8188.79	−8179.55	−8198.61
	AIC	16385.50	16387.44	16391.11	16438.03	16403.59	16385.09	16423.21
	RMSE	4.20	4.27	4.23	4.46	4.32	4.26	4.45

[a] LL denotes log-likelihood.

traffic data observed in the congested traffic condition (16:00–19:00) were used. Different copulas were used and the most appropriate copulas were identified. The calculated Kendall's tau and estimated values of parameter θ of each copula are provided in Table 3. Note that Kendall's tau for speed and headway is -0.37. Thus, some copulas cannot be used given that the degrees of dependence they span were insufficient to account for the association observed between speed and headway. For original speed and headway data, only Gaussian and Frank copulas are applicable. For the Clayton, Gumbel and Joe copulas,

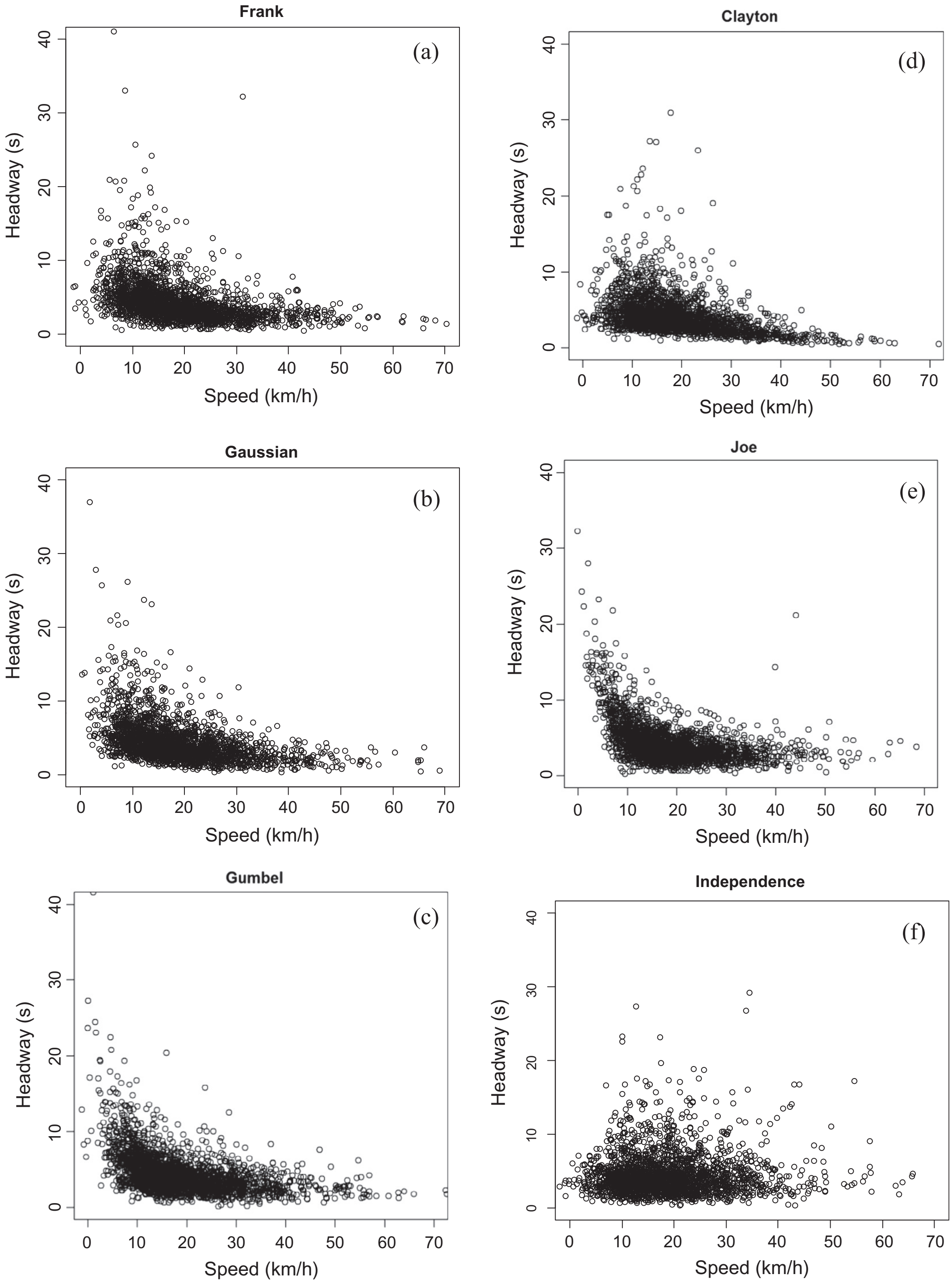

Fig. 4. Transformed samples for (a) the Frank copula with parameter $\theta = -3.80$; (b) the Gaussian copula with parameter $\theta = -0.55$; (c) the Gumbel copula with parameter $\theta = 1.59$; (d) the Clayton copula with parameter $\theta = 1.19$; (e) the Joe copula with parameter $\theta = 2.08$; (f) the independent copula.

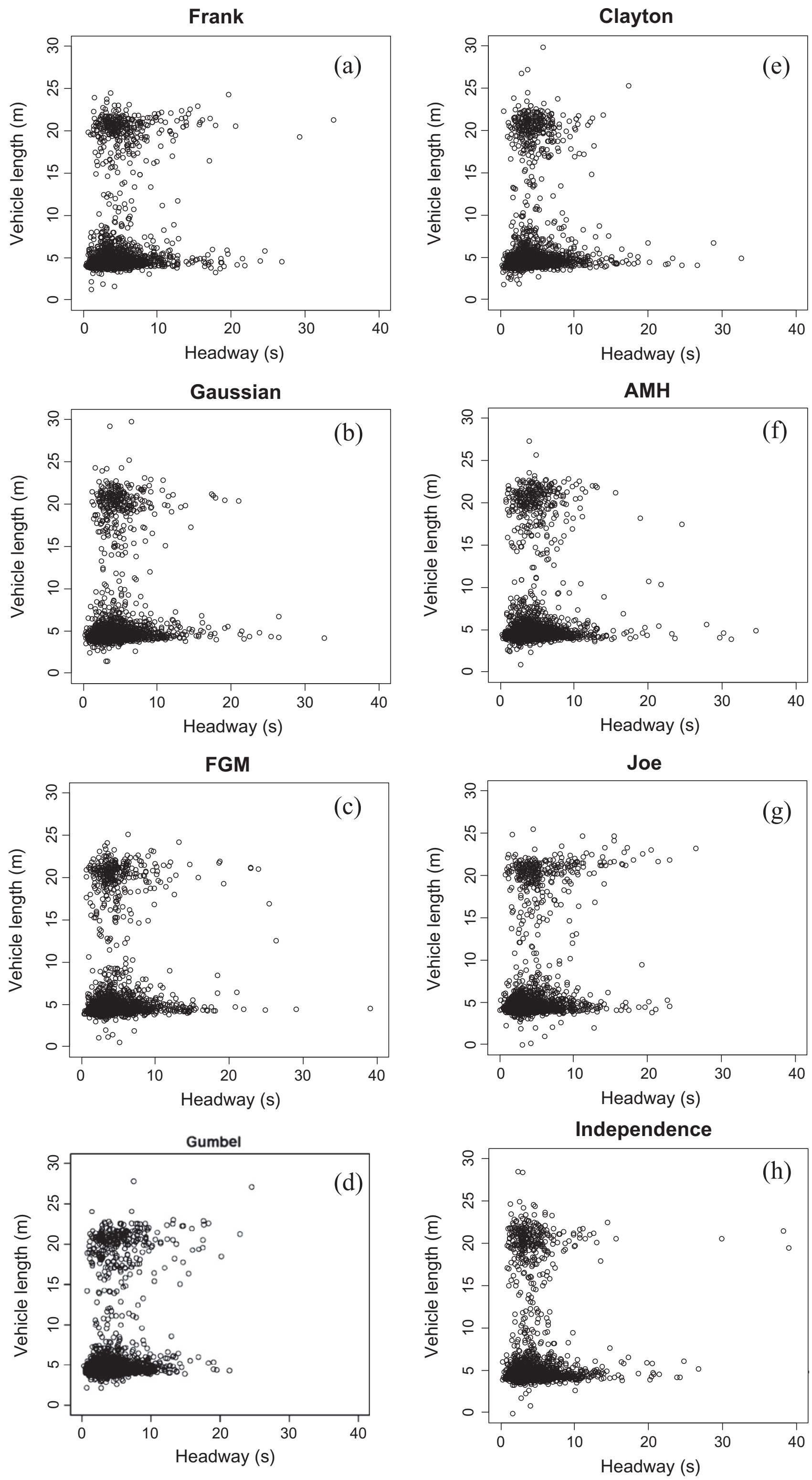

Fig. 5. Transformed samples for (a) the Frank copula with parameter $\theta = 1.21$; (b) the Gaussian copula with parameter $\theta = 0.21$; (c) the FGM copula with parameter $\theta = 0.59$; (d) the Gumbel copula with parameter $\theta = 1.15$; (e) the Clayton copula with parameter $\theta = 0.3$; (f) the AMH copula with parameter $\theta = 0.51$; (g) the Joe copula with parameter $\theta = 1.27$; (h) the independent copula.

we convert the observed speed to a new variable by taking the negative speed values, $Speed_{new} = -$observed speed. $Speed_{new}$ and headway show a positive correlation (Kendall's tau = 0.37). Then, the Clayton, Gumbel and Joe copulas are used to fit the new speed and headway data. For FGM and AMH copulas, the calculated Kendall's tau value exceeds the range of copula parameter. Thus, five copulas (Gaussian, Frank, Clayton, Gumbel and Joe) are used to construct bivariate distribution of speed and headway. The best copula model is selected based on log-likelihood, AIC and RMSE values. For speed and headway data, the Gaussian copula can give larger log-likelihood and smaller AIC and RMSE values than other considered copulas. For headway and vehicle length data, all copulas are viable and the goodness-of-fit statistics for each copula model are provided in Table 4. Overall, the Gaussian copula is found as the best fitted copula for headway and vehicle length data.

One natural way to check the adequacy of copula models is to compare the scatter plot of observations with an artificial dataset of the same size generated from fitted copulas. Using the random variate generation algorithm previously introduced, 2360 pairs (U_i, V_i) were simulated from the Gaussian, Frank, Clayton, Gumbel and Joe copulas with specified θ values. Then, the 2360 pairs (U_i, V_i) from each copula model were transformed back into the original units using the marginal distribution identified in the marginal distribution section for speed and headway. Fig. 4 displays the simulated speed and headway samples. Assuming $\theta = 0$ for the FGM copula, the independent speed and headway samples were also generated for the purpose of comparison. The actual observations are provided in Fig. 2(a). As shown in Fig. 4, the simulated samples from copula models can accurately reproduce the dependence structure revealed by the speed and headway observations. Moreover, the inappropriateness of the independent model is apparent, as it is hard to observe the inverse relationship

Table 5
Parameters and fitting evaluation of trivariate Gaussian copula.

Parameter			LL	AIC	RMSE
$\rho_{speed\&headway}$	$\rho_{speed\&vehiclelength}$	$\rho_{headway\&vehiclelength}$			
−0.55	0.01	0.21	−16306.47	33148.44	0.20

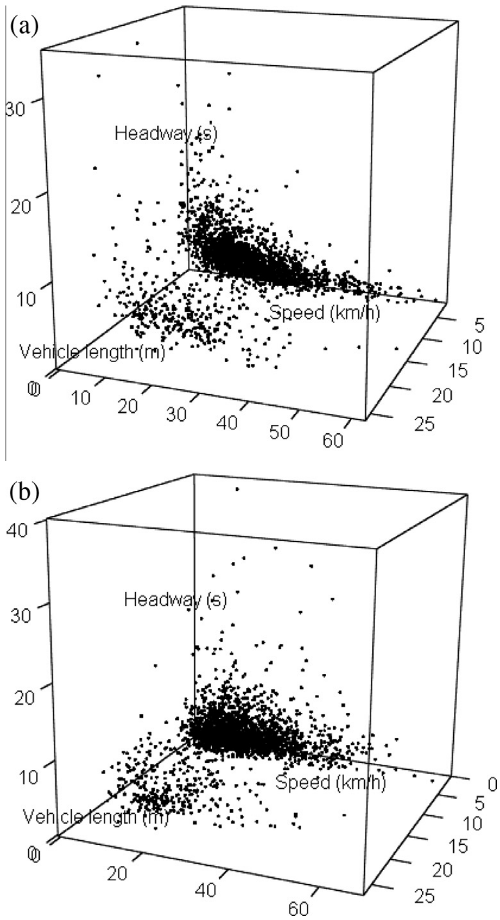

Fig. 6. Transformed samples for (a) the trivariate Gaussian copula; (b) the independent copula.

between speed and headway from Fig. 4(f). The same procedure was repeated for the headway and vehicle length data using various copulas with specified θ values. Fig. 5 exhibits the simulated headway and vehicle length samples. Due to the very weak dependence between headway and vehicle length, it is hard to tell from Fig. 5 whether the actual observations can be more accurately reproduced by considering the dependence structure.

The parameters of trivariate Gaussian copula are estimated and provided in Table 5. The log-likelihood, AIC and RMSE values are employed to measure the fitting performance. Using the random variate generation algorithm for the Gaussian copulas, 2360 vectors $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)$ were simulated with the specified correlation matrix P. Then, the 2360 vectors $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)$ were transformed back into the original units using the marginal distribution selected for speed, headway and vehicle length. Fig. 6 displays the simulated samples. Assuming $\mathrm{P} = I_3$, where I_3 is the 3-dimentional identity matrix, the independent speed, headway and vehicle length samples were also generated for the purpose of comparison. The actual observations are provided in Fig. 2(d). Since there is an inverse relationship between speed and headway for both passenger cars and trucks, the simulated samples from the trivariate Gaussian copula can accurately reproduce this dependence structure. However, it is difficult to observe the inverse relationship between speed and headway from Fig. 6(b).

5. Summary

This study documented the application of copula models for constructing the distribution of traffic variables (speed, headway and vehicle length) using recorded data collected on IH-35. Before constructing multivariate distributions, we first evaluated the hourly dependence among speed, headway and vehicle length for the 24-h period. For each hour, Kendall's tau τ, and Spearman's rho ρ_S are used to measure the dependence. Based on the analysis results, the important conclusions can be summarized as follows:

(1) The relationship between speed and headway and the influence of vehicle length on headway are most obvious for the time period from 16:00 to 19:00, which is the busiest time of the day on IH-35.
(2) There exists a weak positive dependence between headway and vehicle length under both congested and uncongested traffic conditions. And vehicle length does influence following headway as trucks and buses usually keep larger following time headways than cars at the same speed level.

After evaluating the dependence among speed, headway and vehicle length, copula models were used to construct bivariate and trivariate distributions and goodness-of-fit statistics showed that the proposed copula models can adequately describe the multivariate distributions of microscopic traffic variables. Moreover, the simulated samples from some families of copulas can accurately reproduce the actual relationship between traffic variables. Since speed and headway usually have a negative correlation under the congested traffic condition, the degrees of dependence some copulas span are insufficient to account for the association. In this study, Gaussian, Frank, Clayton, Gumbel and Joe copulas are applicable to the speed and headway data. Overall, the findings in this paper provide a framework for generating multiple microscopic traffic variables simultaneously by considering their dependence. For future works, first, for the considered copula models, it is interesting to use the hour (or traffic condition) as the explanatory variable to describe copula parameter. And the goodness of fit results from copula models with fixed copula parameter and variable copula parameter should be compared. Second, the proposed copula modeling approaches can be used to analyze the microscopic traffic data collected from multiple freeway locations.

Acknowledgment

This research is financially supported by Grant No. 51138003 from the National Natural Science Foundation of China.

References

Ali, M.M., Mikhail, N., Haq, M.S., 1978. A class of bivariate distributions including the bivariate logistic. J. Multivariate Anal. 8, 405–412.
Bhat, C.R., Eluru, N., 2009. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Transport. Res. Part B – Methodol. 43, 749–765.
Bhat, C.R., Sener, I.N., 2009. A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. J. Geograph. Syst. 11, 243–272.
Brackstone, M., Waterson, B., McDonald, M., 2009. Determinants of following headway in congested traffic. Transport. Res. Part F – Traffic Psychol. Behav. 12, 131–142.
Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula Methods in Finance. Wiley.com.
Clayton, D.G., 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65, 141–151.
Cowan, R.J., 1975. Useful headway models. Transport. Res. 9, 371–375.
Demarta, S., McNeil, A.J., 2005. The t copula and related copulas. Int. Stat. Rev. 73, 111–129.
Eluru, N., Bhat, C.R., Pendyala, R.M., Konduri, K.C., 2010. A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. Transportation 37, 603–626.
Embrechts, P., Lindskog, F., McNeil, A., 2003. Modelling dependence with copulas and applications to risk management. Handbook Heavy Tailed Distrib. Financ. 8, 329–384.
Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls. Risk Manage.: Value Risk Beyond, 176–223.
Fisher, N., Switzer, P., 2001. Graphical assessment of dependence. Is a picture worth 100 tests? Am. Stat. 55, 233–239.

Frank, M.J., 1979. On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. Aequat. Math. 19, 194–226.
Genest, C., Favre, A.C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. J. Hydrol. Eng. 12, 347–368.
Genest, C., Ghoudi, K., Rivest, L.-P., 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82, 543–552.
Genest, C., Mackay, R.J., 1986. Copules archimédiennes et families de lois bidimensionnelles dont les marges sont données. Can. J. Stat. 14, 145–159.
Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate Archimedean copulas. J. Am. Stat. Assoc. 88, 1034–1043.
Gerlough, D.L., Huber, M.J., 1976. Traffic Flow Theory.
Grimaldi, S., Serinaldi, F., 2006. Asymmetric copula in multivariate flood frequency analysis. Adv. Water Resour. 29, 1155–1167.
Gumbel, E.J., 1960. Bivariate exponential distributions. J. Am. Stat. Assoc. 55, 698–707.
Haight, F.A., 1963. Mathematical Theories of Traffic Flow. Academic Press.
Haight, F.A., Mosher, W.W., 1962. A practical method for improving the accuracy of vehicular speed distribution measurements. In: Highway Research Board Bulletin 341. Highway Research Board, Washington, D.C., pp. 92–116.
Hofert, M., Mächler, M., McNeil, A.J., 2012. Likelihood inference for Archimedean copulas in high dimensions under known margins. J. Multivariate Anal. 110, 133–150.
Hoogendoorn, S.P., Bovy, P.H., 1998. New estimation technique for vehicle-type-specific headway distributions. Transport. Res. Rec.: J. Transport. Res. Board 1646, 18–28.
Joe, H., 1993. Parametric families of multivariate distributions with given margins. J. Multivariate Anal. 46, 262–282.
Joe, H., 1997. Multivariate Models and Dependence Concepts. CRC Press.
Kim, G., Silvapulle, M.J., Silvapulle, P., 2007. Comparison of semiparametric and parametric methods for estimating copulas. Comput. Stat. Data Anal. 51, 2836–2850.
Kim, J., Mahmassani, H.S., 2014. A finite mixture model of vehicle-to-vehicle and day-to-day variability of traffic network travel times. Transport. Res. Part C: Emerg. Technol. 46, 83–97.
Ko, J., Guensler, R.L., 2005. Characterization of congestion based on speed distribution: a statistical approach using Gaussian mixture model. In: Transportation Research Board Annual Meeting, 2005. Citeseer.
Leong, H., 1968. The distribution and trend of free speeds on two lane two way rural highways in New South Wales. In: Australian Road Research Board (ARRB) Conference, 4th, 1968, Melbourne.
Luttinen, R., 1992. Statistical properties of vehicle time headways. Transport. Res. Rec., 92–98
Luttinen, R.T., 1999. Properties of Cowan's M3 headway distribution. Transport. Res. Rec.: J. Transport. Res. Board 1678, 189–196.
Mclean, J., 1979. Observed speed distributions and rural road traffic operations. In: Australian Road Research Board Conference Proc.
Nelsen, R.B., 2006. An Introduction to Copulas. Springer.
Park, B.J., Zhang, Y.L., Lord, D., 2010. Bayesian mixture modeling approach to account for heterogeneity in speed data. Transport. Res. Part B – Methodol. 44, 662–673.
Ravishankar, K., Mathew, T.V., 2011. Vehicle-type dependent car-following model for heterogeneous traffic conditions. J. Transport. Eng. 137, 775–781.
Sarvi, M., 2013. Heavy commercial vehicles-following behavior and interactions with different vehicle classes. J. Adv. Transport. 47 (6), 572–580.
Schucany, W.R., Parr, W.C., Boyer, J.E., 1978. Correlation structure in Farlie–Gumbel–Morgenstern distributions. Biometrika 65, 650–653.
Schultz, G.G., Rilett, L., 2004. Analysis of distribution and calibration of car-following sensitivity parameters in microscopic traffic simulation models. Transport. Res. Rec.: J. Transport. Res. Board 1876, 41–51.
Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges, Université Paris 8.
Spissu, E., Pinjari, A.R., Pendyala, R.M., Bhat, C.R., 2009. A copula-based joint multinomial discrete–continuous model of vehicle type choice and miles of travel. Transportation 36, 403–422.
Taieb-Maimon, M., Shinar, D., 2001. Minimum and comfortable driving headways: reality versus perception. Hum. Factors: J. Hum. Factors Ergon. Soc. 43, 159–172.
Tian, Z.Z., Urbanik, T., Engelbrecht, R., Balke, K., 2002. Variations in capacity and delay estimates from microscopic traffic simulation models. Transport. Res. Rec.: J. Transport. Res. Board 1802, 23–31.
Wang, M., Hoogendoorn, S.P., Daamen, W., van Arem, B., Happee, R., 2015. Game theoretic approach for predictive lane-changing and car-following control. Transport. Res. Part C: Emerg. Technol. 58, 73–92.
Wang, Y., Nihan, N.L., 2004. Dynamic estimation of freeway large-truck volumes based on single-loop measurements. Intelligent Transportation Systems. Taylor & Francis, pp. 133–141.
Wasielewski, P., 1979. Car-following headways on freeways interpreted by the semi-Poisson headway distribution model. Transport. Sci. 13, 36–55.
Ye, F., Zhang, Y., 2009. Vehicle type-specific headway analysis using freeway traffic data. Transport. Res. Rec.: J. Transport. Res. Board 2124, 222–230.
Ye, Z., Zhang, Y., Middleton, D.R., 2006. Unscented Kalman filter method for speed estimation using single loop detector data. Transport. Res. Rec.: J. Transport. Res. Board 1968, 117–125.
Yin, S.C., Li, Z.H., Zhang, Y., Yao, D., Su, Y.L., Li, L., 2009. Headway distribution modeling with regard to traffic status. In: 2009 IEEE Intelligent Vehicles Symposium, vols. 1 and 2, pp. 1057–1062.
Zhang, L., Singh, V.P., 2006. Bivariate flood frequency analysis using the copula method. J. Hydrol. Eng. 11, 150–164.
Zhang, Y., Xie, Y., Ye, Z., 2008. Estimation of large truck volume using single loop detector data. In: Transportation Research Board 87th Annual Meeting.
Zou, Y., Zhang, Y., 2011. Use of skew-normal and skew-t distributions for mixture modeling of freeway speed data. Transport. Res. Rec.: J. Transport. Res. Board 2260, 67–75.
Zou, Y., Zhang, Y., Zhu, X., 2014. Constructing a bivariate distribution for freeway speed and headway data. Transportmet. A: Transport Sci. 10 (3), 255–272.