Transportation Research Part C 82 (2017) 1–23

Contents lists available at ScienceDirect

Transportation Research Part C

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

A copula-based approach for estimating the travel time reliability of urban arterial

Min Chen, Guizhen Yu, Peng Chen *, Yunpeng Wang

School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Infrastructure System and Safety Control, Beihang University, Beijing 100191, China

ARTICLE INFO

Article history:
Received 24 November 2016
Received in revised form 13 June 2017
Accepted 13 June 2017

Keywords:
Urban arterial
Travel time distribution
Travel time reliability
Segment correlation
Copula

ABSTRACT

Estimating the travel time reliability (TTR) of urban arterial is critical for real-time and reliable route guidance and provides theoretical bases and technical support for sophisticated traffic management and control. The state-of-art procedures for arterial TTR estimation usually assume that path travel time follows a certain distribution, with less consideration about segment correlations. However, the conventional approach is usually unrealistic because an important feature of urban arterial is the dependent structure of travel times on continuous segments. In this study, a copula-based approach that incorporates the stochastic characteristics of segments travel time is proposed to model arterial travel time distribution (TTD), which serves as a basis for TTR quantification. First, segments correlation is empirically analyzed and different types of copula models are examined. Then, fitting marginal distributions for segment TTD is conducted by parametric and non-parametric regression analysis, respectively. Based on the estimated parameters of the models, the best-fitting copula is determined in terms of the goodness-of-fit tests. Last, the model is examined at two study sites with AVI data and NGSIM trajectory data, respectively. The results of path TTD estimation demonstrate the advantage of the proposed copula-based approach, compared with the convolution model without capturing segments correlation and the empirical distribution fitting methods. Furthermore, when considering the segments correlation effect, it was found that the estimated path TTR is more accurate than that by the convolution model.

1. Introduction

Travel time reliability (TTR), defined as consistency of travel time over time by SHRP 2 Reliability research (SHRP 2 Report S2-L02-RR-2, 2014), has been increasingly recognized as an important measure for assessing the operation efficiency of road facilities, evaluating alternative traffic management strategies (SHRP 2 Report S2-L07-RR-2, 2014) and providing travelers with timely and reliable route guidance (SHRP 2 Report S2-L14-RW-1, 2014). Bates et al. (2001) suggested that the analysis of TTR is as important as, if not more important than, the traditional analysis of average travel time. To fully characterize TTR, travel time distribution (TTD) needs to be determined as a prior, which helps measure the risk for on-time arrival probability and find a reliable path for risk-averse travelers (Zeng et al., 2015). Ma et al. (2016) also stated that better understanding of TTD is a prerequisite for exploring the causes of unreliability.

* Corresponding author.
E-mail addresses: chenmin_cn@163.com (M. Chen), yugz@buaa.edu.cn (G. Yu), cpeng@buaa.edu.cn (P. Chen), ypwang@buaa.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.trc.2017.06.007
0968-090X/

Compared to uninterrupted freeway facilities, the interrupted nature of travel flows on arterials make TTD estimation much more challenging. Due to complex interactions between volatile traffic regimes and signal control strategies, as well as the correlation between neighboring segments, the resulting travel times on arterials often present various distributions under different levels of congestion (Chen et al., 2013). The state-of-practice approaches for arterial TTD estimation have focused on developing models for estimating the average travel time. Detailed introduction and discussion of these average travel time models have been provided by Highway Capacity Manual (2010), Skabardonis and Geroliminis (2005, 2008), and Liu and Ma (2009). However, less effort has been made to quantify the variability of either segment-level or path-level TTD and analyze the interdependence between segment TTDs along signalized arterials. A review of the related research is provided below.

Segment-level TTD on arterials can capture the interrupted nature of traffic flow under signal control and has various applications in practice, including evaluation on the level of service of intersections, congestion and incident detection, and real time dynamic control of traffic systems. Ji and Zhang (2013) utilized high-resolution bus probe data to estimate travel times on urban streets for short segments, and revealed predominantly bimodal TTD at the segment level, with one mode corresponding to travels without delay and the other travels with delay. Then, a hierarchical Bayesian mixture model was used to characterize such bimodal TTD, analyze traffic operation and identify congestion. Zheng and van Zuylen (2010) investigated urban segment-level TTD by mainly focusing on the delay distribution at signalized intersections. A probabilistic delay distribution model was developed considering both stochastic arrivals and departures at signals. The key influencing factors, e.g., arrival process, degree of saturation, saturation flow rate, were examined in detail. The results indicated a temporal correlation between arrival time and segment travel time, and demonstrated the evolutions of delay distribution on a cycle-by-cycle basis. Mohammad et al. (2015) proposed a probabilistic travel time model for a single segment, which was compatible with the algorithms developed to compute the probability density function (PDF) of a path.

In terms of route guidance, path-level TTD is of more concern to travelers than segment-level TTD. The state-of-art studies on path-level TTD put significant effort into identifying the best statistical model for fitting travel time observations. Unimodal distributions, e.g., Normal, Lognormal, Gamma, Weibull, Exponential and Burr distributions, were often used to characterize TTD for a given path, as can be found in the literature (Eman and Al-Deek, 2006; Uno et al., 2009; Susilawati et al., 2013). However, increasing researchers (Guo et al., 2010; Feng et al., 2012; Kazagli and Koutsopoulos, 2012; Chen et al., 2014) pointed out a unimodal distribution may not sufficiently represent the variation of path TTD. For example, travel times under free-flow conditions and under congested conditions can differ substantially. Guo et al. (2010) proposed a multi-state model to fit a mixture of Gaussian distributions into travel time observations along one arterial at San Antonio, Texas. Feng et al. (2012) adopted a mixture of normal distributions to estimate TTD for arterial routes using GPS probe vehicles. Kazagli and Koutsopoulos (2012) used a mixture of two lognormal distributions to model TTD on urban routes in Sweden based on AVI data. Chen et al. (2014) further presented a finite mixture of regression model with varying mixing probabilities (weights) to gain a better understanding of urban path TTD through consideration of signal timings. The essence of the above multi-state mixture models is to establish a connection between TTDs and the underlying travel time states, which allows for the quantitative evaluation of the probability of each state and results in better fitting. However, such distribution fitting methods are largely dependent on the path-level travel time observations, which may not be available in large-scale urban network. Considering that various paths may exist between one Origin-Destination (OD) pair and one path may include multiple segments, it is more realistic in practice to estimate path-level TTD based on individual segment TTDs given the commonly available data sources at segment-level. The route choice is then made by choosing the best path from all feasible paths according to certain reliability measures.

To date, a fundamental assumption imbedded in most reliability research upon segment-level or path-level TTD in a road network is the independence of individual segment travel times (Iida, 1999). If one is only interested in a deterministic measure of travel times, e.g., average segment travel time or path travel time, this assumption is generally acceptable. However, when focused on TTD, the assumption of segment independence is evidently unrealistic. He et al. (2002) mentioned that the assumptions made for long-term (such as peak hours, non-peak hours, daily and seasonal) TTD estimation, i.e., (1) travel times on all separate route sections are independent and (2) trip times per unit distance on all sections are identically distributed, may not be valid for the short-term estimation of route TTD. They investigated the temporal and spatial variations and correlations of travel times in short term based on vehicle tracking data from Paramics simulation. Strong evidence of significant correlation was found between segment travel times. The results suggested that the estimation of path TTD needs to account for the correlation between individual segments and it is necessary to include both temporal and spatial variability of travel times when developing advanced route guidance systems.

To model the dependency and correlation of segment TTDs, Pattanamekar et al. (2003) suggested that in order to estimate the conditional mean and variance of one segment travel time given the observation of the other, the joint PDF is needed. However, instead of estimating the function in detail, the authors only used a three-point polynomial approximation to estimate average travel time. Rakha et al. (2006) pointed out that the assumption of segment travel time independence ignores any covariance across segments and they tried to estimate the variance of freeway path travel time by modeling dependence between segment travel time variances. This method makes the results tractable; however, the limitation is that the variances cannot fully characterize the travel time characteristics. Geroliminis and Skabardonis (2006) estimated the variance of one urban route travel time by assuming linear correlations between successive segment travel times. In fact, the dependence between successive segment travel times is complex. Herring et al. (2010) proposed a Coupled Hidden Markov Model to model the dynamics of segment travel times in urban network based on probe vehicle data. It is assumed that each seg-

ment has one state evolving over time based on a time-invariant state transition matrix. Given the state of one segment, its TTD (Gaussian) is independent from all other traffic variables and state transition is correlated to the states of the spatial neighbors of the segment. Ramezani and Geroliminis (2012) used Markov chains to predict path TTD through the integration of travel time correlation of path's successive segments. It was assumed that the traffic state of the current segment and that of the nearby segment forms a Markov chain. A two-dimensional (2D) diagram was established for travel times on every two consecutive segments. A heuristic grid clustering method was developed to cluster every 2D diagram to rectangular sub spaces (states) with regard to travel time homogeneity. Finally, path TTD was estimated using the product of the sequence of these pair-wise Markov matrices assuming the transitions between different segment pairs are conditionally independent. Recently, Westgate et al. (2016) presented a regression approach to model TTDs for ambulances at the trip level, which naturally incorporates dependence between segment travel times. However, ambulances traveling at lights-and-sirens speeds are less affected by traffic than other vehicles and the ambulances trips may not appropriately reflect the features of urban road travel time features.

To sum up, the correlation between consecutive segment TTDs on urban arterials has not been explicitly addressed so far. Additionally, after characterizing individual segment TTDs, how to assemble them by considering their spatial correlation and further characterize the TTD for a path remains a challenge. To this end, this study introduces the copula model, previously used in econometrics (Trivedi and Zimmer, 2007), to the estimation of path TTD given a segment set by considering the spatial correlation between segment travel times. The inputs to the proposed copula model are segment travel times, and the output is path TTD over the study time periods. Besides, a complete process of path TTR estimation based on the copula model is presented.

The remainder of this paper is organized as follows. A brief introduction to the concept of copula and a description of various types of copulas are first presented. Next, we provide path TTD estimation with two models, and the parameter estimation method and the goodness-of-fit test for copula were implemented after that. Then, a case study is conducted at one arterial in Shanghai, China to illustrate the application of the proposed methodology. The spatial correlations between segment travel times are statistically examined and path TTDs are estimated by copula methods. The results are compared with the estimates without considering segment correlations as well as empirical bimodal distributions. In addition, the NGSIM data were used to examine whether the proposed copula framework is applicable and transferable to different case studies. The last section draws conclusions and provides recommendations for future work.

2. Copula theory

To incorporate the impact of correlation of segment TTDs in the estimation of path TTD, a copula approach, which has been widely used for modeling multivariate distributions among random variables with pre-specified marginal distributions (Sklar, 1973), is proposed in this study. The copula approach provides a flexible way of describing nonlinear dependence among consecutive segment travel times in isolation from their marginal distributions. In other words, the dependence structure is unaffected by the marginal distributions assumed, which provides flexibility in correlating individual segment travel times, considering they may not even have the same marginal distributions under different traffic states. The basic theorem on multivariate copulas is given below.

2.1. Concepts of copula

The copula theory was first proposed by Sklar (1959) and their applications in the statistics field have grown over the last decades. After that it was increasingly used to analyze problems in the field of finance, economic and hydrology. Recently, the copula method has been applied in transportation studies. Bhat and Eluru (2009) applied a copula-based approach to sample selection in the context of built environment effects on travel behavior. Zou et al. (2014) and Zou and Zhang (2016) modeled the relationship between microscopic traffic flow parameters with the copula method.

Copula is a function that relates multivariate distribution functions of random variables to their one-dimensional marginal distribution functions. According to Sklar's theorem (1973), for an n-variate distribution function $F(x_1, x_2, \ldots, x_n)$ with marginal distribution functions $F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$, there exists a certain copula function C which meets the relationship $F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))$. If marginal distributions are all continuous, C is unique. It states that the multivariate distribution F of random variables $x_1, x_2, \ldots, x_n$ can be expressed in terms of a copula function C and their marginal distributions.

Specifically, Sklar's theorem can be represented as

$$f(x_1, x_2, \ldots, x_n; \theta) = c(F_1(x_1, \theta_1), F_2(x_2, \theta_2), \ldots, F_n(x_n, \theta_n); \theta_c) \prod_{i=1}^{n} f_i(x_i; \theta_i) \tag{1}$$

$$c(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) = \frac{\partial C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n))}{\partial F_1(x_1) \partial F_2(x_2), \ldots, \partial F_n(x_n)} \tag{2}$$

where $f(x_1, x_2, \ldots, x_n; \theta)$ is the joint PDF with the parameter θ; c denotes the density of the copula function; $f_i(x_i; \theta_i)$ is the marginal distribution function with parameter $\theta_i (i = 1, 2, 3, \ldots)$; and $F_i(x_i), (i = 1, 2, 3, \ldots)$ denotes the cumulative density function (CDF) of the marginal distribution.

Thus we can use the two-stage procedure to construct a copula model. First, to determine the marginal distributions of segment travel times. Second, to choose a favorable copula which can characterize the dependence between segment travel times.

2.2. Measuring dependence

In fact, there are different ways to measure dependence. For the random variables X, Y, Pearson's linear correlation coefficient was the most commonly used. And it can be calculated by

$$\rho_p(X, Y) = \frac{cov(X, Y)}{\sqrt{\sigma^2(X)\sigma^2(Y)}} \tag{3}$$

where $cov(X, Y)$ is the covariance of X and Y; $cov(X, Y) = \mu(XY) - \mu(X)\mu(Y)$; $\mu(XY)$ denotes the mean of variables X and Y; $\mu(X), \sigma(X)$ and $\mu(Y), \sigma(Y)$ are the mean and standard deviation of X and Y, respectively.

The Pearson's linear correlation coefficient has the deficiency that it is not invariant under nonlinear strictly increasing transformations (Embrechts et al., 2002). On the contrary, two widely known scale-invariant measures of association (Kendall's tau and Spearman's rho) are more flexible in characterizing the dependence. Embrechts et al. (2002) ever pointed out that for multivariate distributions which possess a simple closed-form copula, the linear correlation coefficient may be difficult to describe and the determination of rank-based correlation may be easier. As Kendall and Gibbons (1990) argued, the confidence intervals for Spearman's rho are less reliable and interpretable than confidence intervals for Kendall's tau. Kendall's tau measure of dependence between two random variables (X, Y) is defined as the probability of concordance minus the probability of discordance. This can be represented as

$$\tau(X, Y) = \mathrm{P}[(X - \widetilde{X})(Y - \widetilde{Y}) > 0] - \mathrm{P}[(X - \widetilde{X})(Y - \widetilde{Y}) < 0] \tag{4}$$

where $(\widetilde{X}, \widetilde{Y})$ is an independent copy of the vector (X, Y). The first expression on the right side is the probability of concordance of (X, Y) and $(\widetilde{X}, \widetilde{Y})$, and the second expression on the right side is the probability of discordance of the same two vectors. In common with other measures of correlation, Kendall's tau will produce values ranging from -1 to 1, with a positive correlation indicating that the ranks of both variables increase together while a negative correlation indicates that as the rank of one variable increases, that of the other one decreases.

In summary, considering the advantages of rank-based correlation, Kendall's tau, which is taken as the main measure indicator, and Spearman's rho are used to characterize the dependence between segments travel time in the following section.

2.3. Family of bivariate copulas

2.3.1. Bivariate Gaussian copula

As the most frequently used one, the Gaussian copula is an elliptical and symmetric copula, since it is simply the copula of the elliptical bivariate normal distribution. Using a Gaussian copula with normal margins is essentially equivalent to a bivariate normal distribution. The 2-dimensional Gaussian copula can be defined as

$$C(u_1, u_2; \theta) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta) \tag{5}$$

where Φ^{-1} denotes the inverse CDF of a standard normal, $\Phi_2(.,.;\theta)$ is the bivariate CDF with Pearson's correlation parameter $\theta (-1 \leqslant \theta \leqslant 1)$. Note u_1 and u_2 can be any arbitrary marginal CDF, either parametric or non-parametric, which evidently distinguishes the Gaussian copula from the joint normal CDF.

The copula density is given by

$$\begin{aligned} c(u_1, u_2; \theta) &= \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2; \theta) \\ &= \frac{1}{\sqrt{1-\theta^2}} \exp\left(\frac{2\theta\Phi^{-1}(u_1)\Phi^{-1}(u_2) - \theta^2(\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)^2)}{2(1-\theta^2)}\right) \end{aligned} \tag{6}$$

In terms of its copula density, the 2-dimensional Gaussian copula can be represented as

$$C(u_1, u_2; \theta) = \int_0^{u_1} \int_0^{u_2} c(s, t; \theta) ds dt \tag{7}$$

2.3.2. The Farlie-Gumbel-Morgenstern copula

The joint CDF of the bivariate Farlie-Gumbel-Morgenstern (FGM) Copula is provided by:

$$C_\theta(u_1,u_2)=u_1u_2[1+\theta(1-u_1)(1-u_2)] \tag{8}$$

where θ is a parameter of the copula function and it ranges in $|\theta|\leqslant 1$.

The density of the FGM Copula can be described as follows:

$$c_\theta(u_1,u_2)=[1+\theta(2u_1-1)(2u_2-1)] \tag{9}$$

There is a limitation when using the FGM Copula that only if the correlation of two variables is weak, the FGM Copula can provide an effective way for constructing a bivariate distribution. The parameter θ, Kendall's tau τ and Spearman's rho ρ satisfy the following equations:

$$\tau=\frac{2}{9}\theta \tag{10}$$

$$\rho=\frac{1}{3}\theta \tag{11}$$

Since parameter θ meets the formula $|\theta|\leqslant 1$, the FGM Copula can allow both positive and negative dependence.

2.3.3. Bivariate Archimedean copula

Another widely used family of copula functions is the Archimedean family, which is constructed in a completely different way from the normal copula and the FGM copula by using a generator function φ. N-dimension Archimedean copulas are defined as follows:

$$C(u_1,u_2,\ldots,u_n)=\varphi^{[-1]}\{\varphi(u_1)+\varphi(u_2)+\cdots+\varphi(u_n)\} \tag{12}$$

$$\varphi^{[-1]}(t)=\begin{cases}\varphi^{-1}(t), & 0\leqslant t\leqslant\varphi(0)\\ 0, & \varphi(0)\leqslant t\leqslant\infty\end{cases} \tag{13}$$

where $C(u_1,u_2,\ldots,u_n)$ is a certain copula, which denotes the dependence structure of random variables $X_1,X_2,\ldots,X_n$; $u_i=F(x_i)$ is the marginal distribution of X_i; the function φ, called a generator of the Archimedean copula, is a continuous and strictly decreasing function which meets $\varphi(0)=\infty,\varphi(\infty)=0$; the pseudo-inverse of φ is the function $\varphi^{[-1]}$, which also is a continuous and strictly decreasing function meeting $\varphi^{[-1]}(0)=1,\varphi^{[-1]}(\infty)=0$.

According to Eq. (12), it can be learnt that if the generator function is definite, its corresponding copula function is also explicit. Thus, different generator functions result in different types of Archimedean copulas, namely, Ali-Mikhai-Haq (AMH) copula, Clayton copula, Frank copula, Gumbel copula and Joe copula. Different types of Archimedean copulas have their own requirements on dependence. The Gumbel copula, Clayton copula and Joe copula allow only positive dependence. The Frank copula allows for both positive and negative dependence, while the AMH copula and FGM copula characterize only weak dependence. For the weak and negative dependence between segments travel time in this paper, here only AMH copula, Frank copula and their corresponding parameters estimation are introduced as follows:

2.3.3.1. Ali-Mikhail-Hap copula. If the generator function $\varphi(t)=ln\frac{1-\theta(1-t)}{t}$, with $-1\leqslant\theta<1$, the copula is Ali-Mikhail-Hap (AMH) copula, proposed by Ali et al.(1978). The AMH copula is

$$C_\theta(u_1,u_2)=\frac{u_1u_2}{1-\theta(1-u_1)(1-u_2)} \tag{14}$$

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

$$c_\theta(u_1,u_2)=\frac{(1-\theta)^3}{\theta^2}\frac{h_\theta^A(u_1,u_2)}{u_1^2u_2^2}Li_{-2}\{h_\theta^A(u_1,u_2)\} \tag{15}$$

where $h_\theta^A(u_1,u_2)=\theta\frac{u_1}{1-\theta(1-u_1)}\frac{u_2}{1-\theta(1-u_2)}$ and $Li_s(z)=\sum_{k=1}^{\infty}z^k/k^s$.

2.3.3.2. Frank copula. If the generator function $\varphi(t)=-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$, with $\theta\neq0$, the copula is called Frank copula, proposed by Frank (1979). It is expressed as

$$C_\theta(u_1,u_2)=-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right) \tag{16}$$

The Kendall's tau is related to θ by $\tau=1-\frac{4}{\theta}[1-D_1(\theta)]$, where $D_1(\theta)$ is the first order Debye function $D_k(\theta)$ defined as $D_k(\theta)=\frac{k}{\theta^k}\int_0^\theta\frac{t^k}{e^t-1}dt$, so that $-1<\tau<1$. The density function of Frank copula is given by Hofert et al. (2012):

$$c_\theta(u_1,u_2)=\left(\frac{\theta}{1-e^{-\theta}}\right)Li_{-1}\{h_\theta^F(u_1,u_2)\}\frac{\exp(-\theta(u_1+u_2))}{h_\theta^F(u_1,u_2)} \tag{17}$$

where $h_\theta^F(u_1,u_2)=(1-e^{-\theta})^{-1}(1-\exp(-\theta u_1)(1-\exp(-\theta u_2))$.

2.4. Trivariate Gaussian copula

Similar to the bivariate Gaussian copula, in the trivariate case the copula expression can be represented as

$$\begin{aligned}C_P(u)&=\Phi_P(\Phi^{-1}(u_1),\Phi^{-1}(u_2),\Phi^{-1}(u_3)\\&=\int_{-\infty}^{\Phi^{-1}(u_1)}\int_{-\infty}^{\Phi^{-1}(u_2)}\int_{-\infty}^{\Phi^{-1}(u_3)}\frac{1}{(2\pi)^{3/2}|P|^{1/2}}\exp\left(-\frac{1}{2}w^TP^{-1}w\right)dw\end{aligned} \tag{18}$$

where $P=\begin{bmatrix}1&\rho_{12}&\rho_{13}\\\rho_{12}&1&\rho_{23}\\\rho_{13}&\rho_{23}&1\end{bmatrix}$ is the symmetric correlation matrix with $-1\leqslant\rho_{ij}\leqslant 1$ $(i,j=1,2,3)$;

$w=(w_1,w_2,w_3)^T$ represents the corresponding integral variables.

Besides the trivariate Gaussian copula, trivariate Archimedean copulas are also widely used for modeling multivariate distribution of multiple random variables. However, compared with trivariate Gaussian copula which is applicable to all ranges of dependence, trivariate Archimedean copulas can only model positive dependence. Thus, considering the possible relationship between segment travel times, only trivariate Gaussian copula are considered here.

3. Modeling travel time reliability

Before modeling path TTR, the definition of an individual segment is explained. Generally, a segment is defined such that consecutive segments are contiguous over a section of the road (Bhaskar et al., 2009). Only in this way travel time estimates from such segments can be used to obtain path travel time estimates, where a path consists of multiple segments.

Besides, it is necessary to make clear the definition of TTR and the measure for TTR. TTR is defined as the level of consistency in travel conditions over time, and is measured by describing the distribution of travel times that occur over a substantial period of time (SHRP 2 Report S2-L02-RR-2, 2014). For more detail, it is presented in Fig. 1.

The definition of path TTR can be expressed as

$$R_x=P(T_x<\lambda t_x^0) \tag{19}$$

where R_x denotes TTR of the path x; T_x is travel time for path x; t_x^0 is travel time for path x in free traffic flow; λ denotes a constant representing the standard for expected travel time.

Van Lint et al. (2008) pointed out that the skewness and width of TTD should also be calculated when estimating path TTR. Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right. The larger the absolute value of the skewness is, the higher the probability of extreme travel times will be. Here, the third moment of the distribution s, i.e., the more common definition of the skewness in statistics, is used to quantify the skewness of the TTD. s can be represented as

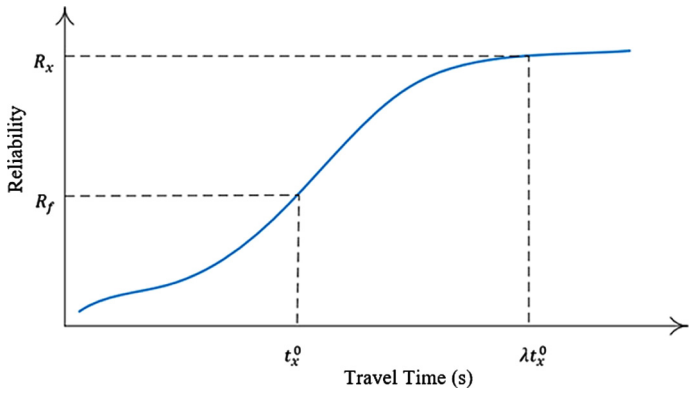


Fig. 1. Definition of travel time reliability.

$$s = \frac{E(x-\mu)^3}{\sigma^3} \tag{20}$$

where μ is the mean of variable x, σ denotes the standard deviation of x, and $E(t)$ represents the expected value of the quantity t.

The width of TTD λ^{var} is the distance between the 90th and 50th percentile relative to the median, which can be expressed as

$$\lambda^{var} = \frac{T_{90} - T_{10}}{T_{50}} \tag{21}$$

where $T_{10} < T_{50} < T_{90}$, T_{XX} denotes the XX percentile value. The wider the distribution λ^{var} is, the lower the reliability will be.

3.1. Convolution model

Given all individual segment TTDs, the simplest model for path TTD estimation is to aggregate those independently by convoluting segment TTDs (Ramezani and Geroliminis, 2012). One path consists of n segments, the path TTD is then computed according to:

$$TTD_N = TTD_1 * TTD_2 * \ldots * TTD_n \tag{22}$$

$$(TTD_i * TTD_j)(t) \triangleq = \int_{-\infty}^{\infty} TTD_i(\tau) TTD_j(t-\tau) d\tau \tag{23}$$

where the $(*)$ mathematical operator expresses convolution (Killmann and Collani, 2001) and the left term in Eq. (23) is the probability density over time t for two segments (i, j = 1, 2,..., n) given both TTDs a priori. It is clear that the convolution method considers independence and thus neglects the spatiotemporal correlation between segment TTDs.

3.2. Copula model

When taking the correlation into consideration, the probability density function (PDF) of path travel time is not the convolution of PDFs of each segment travel time, but the joint PDF of path travel time which was deduced based on the proposed copula method. Then we can estimate path TTR according to its definition. As mentioned above, the TTD of each segment needs to be determined first, and then choose the best fitting copula function to characterize the dependence between segments.

3.2.1. Estimation of marginal distribution

As aforementioned, segment travel times may follow multimodal distributions due to heterogeneous traffic states. Thus, Gaussian Mixture Model (GMM), as a parametric estimator and a representative of multimodal distribution, is preferred in practice to fit segment TTD.

The finite GMM with k components (Rasmussen, 1999) may be represented as

$$\mathrm{p}(\mathrm{y}|u_1, u_2, \ldots, u_k; s_1, s_2, \ldots, s_k; \pi_1, \pi_2, \ldots, \pi_k) = \sum_{j=1}^{k} \pi_j N(u_j, s_j^{-1}) \tag{24}$$

where u_j are the means, s_j are the precisions (inverse variances), π_j are the mixing proportions (which must be positive and sum to one) and N is a (normalized) Gaussian with specified mean and variance.

Besides, kernel smoothing estimator is superimposed to assist with the interpretation of density curves. The density at x by the kernel smoothing method is given by

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{25}$$

where K is the kernel function, h is a smoothing parameter called the bandwidth and X_i is the observed data. Trade-off exists for the selection of h considering the bias of the estimation and its variance. In this study, Gaussian kernel is selected and the bandwidth h chosen according to the optimality rules in Bowman and Azzalini (1997).

In selecting the marginal distribution, we assume that segment travel times follow a certain distribution. The best fitted distribution for segment travel times is selected in terms of Log-likelihood, the Akaike information criterion (AIC) and Kolmogorov-Smirnov (K-S) test. A larger Log-likelihood indicates a better fitting model. AIC values can be calculated by

$$AIC = -2 \times \ln(\textit{maximized likelihood for the model}) + 2 \times \textit{numbers of fitted parameters} \tag{26}$$

A smaller AIC value indicates a better fit. The K-S test (Hollander and Liu, 2008) is a standard non-parametric test to state whether samples are distributed according to a hypothetical distribution (in opposition to other tests like the T-test or the chi-squared test that tests the means and the normality of the data, respectively). The test is based on the K-S statistic which

is computed as the maximum difference between the empirical and the hypothetical cumulative distributions. The test provides a *p-value* which informs the goodness of the fit. A low *p-value* indicates that the data does not follow the hypothetical distribution.

3.2.2. Estimation of θ in copula

After estimating the marginal distributions, the data can be transformed onto the copula scale using the probability integral transform. The next step is to estimate copula parameter of θ. Since the ranks of the observations are the best summary of the joint behavior of the random pairs, it was recommended using the ranked-based estimators. Kendall's tau and Spearman's rho are the basis of the two straightforward estimators; the former was used to characterize the dependence between segment travel times. Here, we take the FGM copula with the parameter of Kendall's tau τ as an example.

If the dependence structure of a random pair (X, Y) can be appropriately modeled by the FGM copula, the estimated parameter $\hat{\theta}$ can be expressed as

$$\hat{\theta} = \frac{9}{2}\tau \tag{27}$$

The rank-based estimation strategy may be seen as a nonparametric adaption of the moments (Genest and Favre, 2007). Once the ranked-based estimators are known, the correlation parameter of the copula θ can be estimated, which is implemented in R packages (R Development Core Team, 2008).

The other method for estimating θ is called maximum likelihood, compared with Kendall's tau and Spearman's rho, which has the advantage that it does not require the dependence parameter θ to be real. However, this method involves lots of numerical work. Besides, the inference functions for margins (IFM) estimators depend on the choice of marginal distributions, and may run the risk of being unduly affected if selections of the marginal distributions turn out to be inappropriate. This study adopts the Kendall's tau based estimator.

3.2.3. Identification of the best-fitting copula

Cramér-von Mises (CvM) statistics, Log-likelihood and AIC values are considered as the goodness-of-fit measures for comparing alternative copulas. Compared with K-S statistic, CvM statistics yield the best blanket goodness-of-fit test procedures for copula models (Breymann et al., 2003; Genest et al., 2006, 2009). The CvM statistic is based on the empirical process comparing the empirical copula with a parametric estimate of the copula derived under the null hypothesis. The CvM function represents the type of distance between the true and observed copulas. Approximate *p*-values for the test statistic can be obtained by means of a parametric bootstrapping approach. The p_c -value represents the level at which the copula is not rejected, meaning that the models with higher p_c-values are better in terms of them not being rejected. Larger Log-likelihood and lower AIC values also indicate a better-fitting copula.

3.2.4. Modeling path travel time reliability

Based on the dependence between segments travel time, we can determine the candidate copulas for path TTD description. Among them, there exists a most appropriate copula for describing such dependence in terms of the goodness-of-fit tests. Here, as an example, we present the process of building path TTD based on the FGM copula model. According to Eqs. (1) and (9), the PDF of path travel time can be expressed as

$$\begin{aligned} f_{T_1T_2,\ldots,T_n}(t_1,t_2,\ldots,t_n) &= c(F_{T_1}(t_1),F_{T_2}(t_2),\ldots,F_{T_n}(t_n))\prod_{i=1}^{n} f_{T_i}(t_i) \\ &= [1+\theta(2F_T(\mathrm{t})-1)]\prod_{i=1}^{n} f_{T_i}(t_i) \end{aligned} \tag{28}$$

where $F_T(\mathrm{t}) = (F_{T_1}(t), F_{T_2}(t), \ldots, F_{T_n}(t))$; $F_{T_i}(t_i)$ is the CDF of segment i; $f_{T_i}(t_i)$ is the PDF of segment i.

After selecting the segment TTD, combing with Eq. (19), we can calculate the path TTR. The details are provided as follows:

$$R_x = p(T_x < \lambda t_x^0) = F_z(\lambda t_x^0) \tag{29}$$

$$\begin{aligned} \mathrm{F}_z(\lambda t_x^0) &= p(T_1 + T_2 + \cdots + T_n \leqslant \lambda t_x^0) \\ &= \iint_{Z<\lambda t_x^0} \cdots \int f(t_1 t_2,\ldots,t_n)dt_1dt_2,\ldots,dt_n \end{aligned} \tag{30}$$

where the relationship between Z and T_i can be represented by $Z = T_1 + T_2 + \cdots + T_n$, and it is the sum of several segment travel times.

Last, path TTR can be expressed by

$$R_x = \iint_{Z<\lambda t_x^0} [1+\theta(2F_T(t)-1)]\prod_{i=1}^{n} f_{T_i}(t_i)dt_1dt_2,\ldots,dt_n \tag{31}$$

4. Case study

The proposed methodology was evaluated for the through movement of two arterials in Shanghai, China and Los Angeles, California, respectively. Following presented is the study site description and data preparation process; based on investigation of segment TTDs, the copula model was constructed for path TTD estimation; then, the comparison was made for path TTDs estimated by the copula model, the convolution and the empirical distribution fitting approach; last, based on estimated TTDs, the path TTR was evaluated for different approaches.

4.1. Study site description

4.1.1. Changshou Road

The study site is a 1.8 km long stretch of one major urban arterial named Changshou Road in Shanghai, China. There are four through lanes in each direction of the arterial, and the speed limit is 40 km/h. The study section consists of seven signalized intersections operated by the Sydney Coordinated Adaptive Traffic System (SCATS), as illustrated in Fig. 2(a). A three-phase signal timing plan was applied at all the intersections. Due to a wide range of traffic conditions through the day, the system cycle lengths ranged from 80 to over 200 s dependent on the time-varying traffic demand.

Fig. 2. Study site: (a) Changshou Road, (b) Lankershim Street schematic.

The travel time data (ever used in Chen et al., 2014) were collected by using automatic license plate recognition and automatic vehicle identification (AVI). The AVI cameras were installed at the downstream ends of the segments, i.e., Intersection No. 1: Chang'an intersection, Intersection No. 4: Shanxi North intersection, and Intersection No. 7: Jiaozhou intersection, as illustrated in Fig. 2(a). Accordingly, two continuous segments are defined between three AVI locations. That is, Segment 1 is from Chang'an intersection to Shanxi North intersection, which is 991 m long. Segment 2 is from Shanxi North intersection to Jiaozhou intersection, which has a length of 813 m.

AVI cameras registered the license plates of passing vehicles and the corresponding time stamps. Actual travel times through the segments can be obtained through matching identified vehicle license plates by both upstream and downstream AVI cameras and comparing their time stamps. Considering that Changshou Road was surrounded by commercial buildings, there are outliers of travel time samples between two AVI cameras, e.g., picking up a passenger and parking along the street. Field investigation showed that during peak periods most of the delayed vehicles have maximum queuing time less than two red phases. Thus, 300 s, i.e., approximately the time of two cycle lengths and the 90th percentile travel time of all the vehicles, was set as the upper bound value for outlier elimination in this study. From 0:00 on August 25th (Monday) to 24:00 on August 29th (Friday), 2008, altogether the ground truth travel times of 4257 vehicles on two segments were obtained and will be utilized for the copula modeling as proposed in this study.

4.1.2. Lankershim Street

The second selected study site is Lankershim Street, an arterial in Los Angeles, California. It is approximately 490 m in length with four intersections, three links and three to four through lanes in each direction. The speed limit on this north-south arterial is 35mph and we only study the southbound direction of traffic. The Lankershim Street dataset is part of the NGSIM program (NGSIM, 2005), which was managed by Federal Highway Administration and intended to provide a dataset of arterial vehicle trajectories for traffic behavioral analyses. Detailed traffic data (vehicle trajectories) were collected using video cameras between 8:30 a.m. and 8:45 a.m. at a resolution of 10 frames per second on June 16, 2005. The dataset contained comprehensive individual vehicle trajectories with time, link and direction stamp, from which the link travel times of vehicles are calculated. Fig. 3b shows the study area schematic. Intersection numbering is incremented from the north-most location. Link1 is from Intersection No. 4 to Intersection No. 3, Link2 is from Intersection No. 3 to Intersection No. 2, and Link3 is from Intersection No. 2 to Intersection No. 1.

4.2. Application of copula models in Changshou Road

4.2.1. Dependence between segments

Considering time of day, congestion of the road and occurrence of any event that may change the travel time through the path significantly, it is possible that the dependence structure between segments travel time may vary relying on the traffic conditions. Thus, we first evaluate the hourly dependence among segments travel time, which is based on AVI data at the vehicle-to-vehicle level. For each hour, Pearson's linear correlation coefficient ρ_P, Kendall's tau τ, and Spearman's rho ρ_S are used to measure the segment dependence. The results are presented in Table 1.

As shown in Table 1, the values of Spearman's rho ρ_S are all higher than those of Kendall's tau τ, and it tested and verified the view of Kendall's and Gibbons (1990) that the confidence intervals for Spearman's rho are less reliable and interpretable than those for Kendall's tau, while Pearson's linear correlation coefficient in absolute value was generally overestimated. Thus, we use Kendall's tau that owns high reliability to analyze the dependence between segments travel time.

To make a direct comparison of the dependence changes of segments travel time among different times of day presented in Table 1, the curve of Kendall's tau over the 24 h period is shown in Fig. 3. The results illustrate that the dependence structure between segments travel time shows weak correlation in general along Changshou Road. There exist several alternative

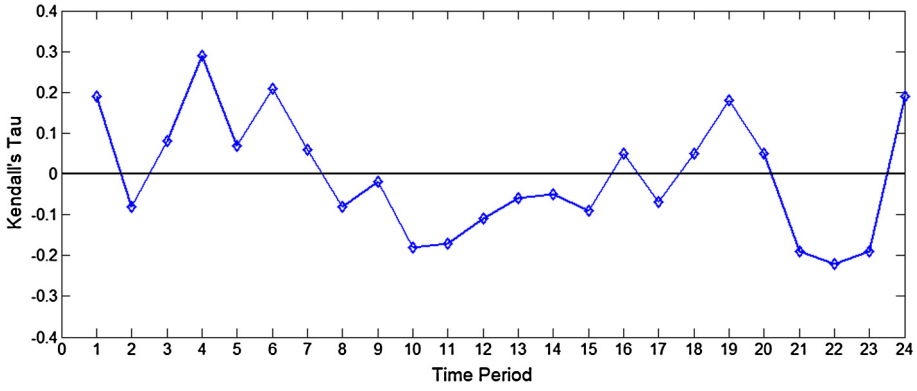


Fig. 3. Dependence between Segment 1 and Segment 2 by Kendall's tau.

Table 1
Hourly dependence between Segment 1 and Segment 2 for the 24 h period.

Number	Time period	Count (vehicles)	Segment 1 and Segment 2		
			ρ_P	τ	ρ_S
1	0–1	29	0.27	0.19	0.30
2	1–2	25	−0.23	−0.08	−0.13
3	2–3	19	−0.33	0.08	0.09
4	3–4	8	0.12	0.29	0.38
5	4–5	6	−0.06	0.07	0.09
6	5–6	72	0.33	0.21	0.31
7	6–7	230	0.09	0.06	0.04
8	7–8	467	−0.14	−0.08	−0.13
9	8–9	610	−0.03	−0.02	−0.03
10	**9–10**	**353**	**−0.24**	**−0.18**	**−0.27**
11	**10–11**	**311**	**−0.21**	**−0.17**	**−0.26**
12	11–12	301	−0.20	−0.11	−0.17
13	12–13	295	−0.11	−0.06	−0.10
14	13–14	178	−0.03	−0.05	−0.06
15	14–15	150	−0.17	−0.09	−0.14
16	15–16	137	0.07	0.05	0.09
17	16–17	163	−0.14	−0.07	−0.11
18	17–18	374	0.02	0.05	0.09
19	18–19	105	0.24	0.18	0.27
20	19–20	130	0.05	0.05	0.09
21	20–21	108	−0.32	−0.19	−0.27
22	21–22	97	−0.36	−0.22	−0.32
23	22–23	48	−0.22	−0.19	−0.30
24	23–24	41	0.17	0.19	0.30

copula functions to characterize such weak correlation. It is worth noting that the numbers of samples during the time periods of 0:00–6:00 and 19:00–24:00 are not enough. The reason can be attributed to the fact that less people travel during midnight or early in the morning; while with respect to 19:00–21:00, it is mainly because that one segment in a path can be part of many other paths, and the number of available data samples for a given segment is far greater than that for any path that goes through the segment. According to the statistics, due to the branches existed and the randomness of people's traveling OD, the number of data samples both in Segment 1 and Segment 2 accounts for only 20% of the traffic on individual segments. On the other hand, the vehicle-to-vehicle travel time samples serve as the most accurate data to represent segment-level and path-level travel time.

For the identified peak periods (i.e., 7:00–9:00, 11:00–13:00, 16:00–18:00), $|\tau| < 0.1$ or $|\tau| \approx 0.1$, Kendall's tau indicates that the adjacent segments are hardly dependent with each other during these peak periods. In essence, the level of correlation between segments travel time may get influenced by numerous parameters and variables, e.g., segment length, traffic volume, signal settings. Fig. 4a presents the through traffic counts by SCATS detectors at 5-min intervals on August 25, 2008 at Intersections No. 4 and No. 7, as illustrated in Fig. 2a. The traffic flows at the two intersections were close and had similar time-varying characteristics. Furthermore, afternoon peaks were not clear compared with morning peaks. Commercial activ-

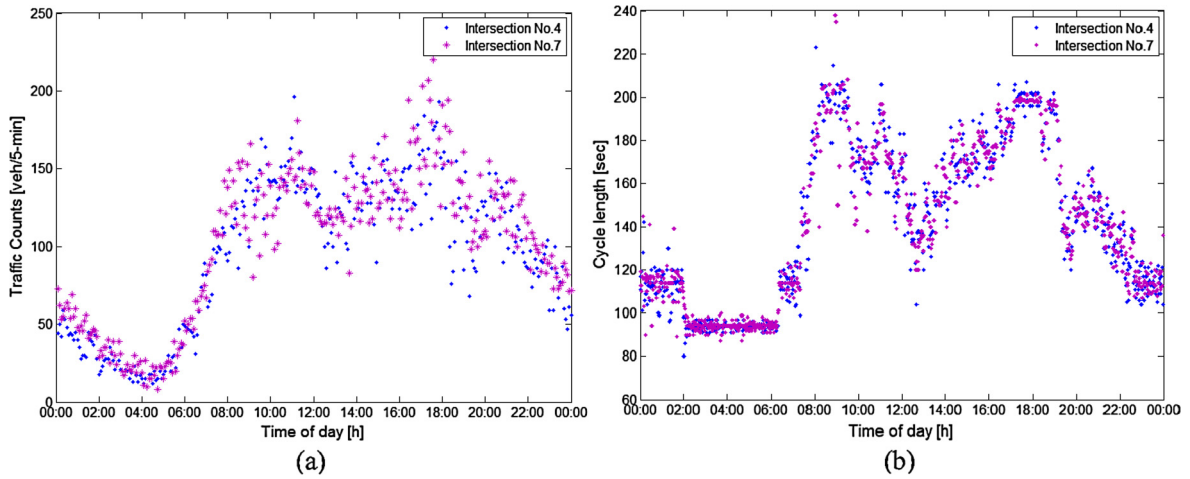


Fig. 4. Intersections No. 4 and No. 7 on August 25, 2008: (a) traffic counts; (b) cycle lengths.

ities extended the peak periods from morning to late evening. Fig. 4a shows that only after 9 p.m. does traffic demand start decreasing. Fig. 4b shows the cycle lengths at Intersections No. 4 and No. 7. The two plots follow similar patterns for time-varying traffic volume.

Besides, weak dependency could be explained by the fact that traffic flow at the site was temporally out of capacity, thus the impact of signal control performed weakly. Conversely, for off-peak periods (i.e., 9:00–11:00), $-0.20 \leqslant \tau \leqslant -0.15$, Kendall's tau suggests that there exists certain negative dependence. One potential reason is that the road capacity meets traffic demand and the favorable signal control between segments domains. Note that for the case of unfavorable coordination, the change of traffic demand may have a greater impact on segment correlation, while the signal control has a relatively smaller impact. Thus, the segments travel time correlation naturally shows a positive feature. For example, as the traffic flow increases, the travel times for two segments increase simultaneously. However, this is not true in the favorable coordination case, because the signal control rather than traffic flow may have a leading impact on segment correlation. On the other hand, such week correlation might be the case that segments consist of several links and intersections. Last but not least, the vehicle-to-vehicle travel times are affected by heterogeneous driver characteristics. Taken together, the correlation can be regarded as the comprehensive effect of temporal-spatial aggregation and driver characteristics.

Given the numbers of samples and dependence structure, the following model construction and results of analysis are based on the data of 9:00–11:00 a.m. Table 2 shows the dependence estimation result. Note that there were 664 vehicles observed during this period. Kendall's tau τ is -0.164. In the following sections, the parameters of alternative copulas are determined by τ.

4.2.2. Investigation of segment TTD

For the copula approach, the marginal distributions of segment travel times need to be specified separately. Based on the Kernel smoothing estimation results superimposed to assist with the interpretation of density curves in Fig. 5, segment travel times appear to follow a multimodal distribution with three peaks. Thus, GMM3 was chose to fit both Segment 1 and Segment 2 travel times. The fitting results are presented by red curves in Fig. 5. Besides, Table 3 presents the GMM3 fitting parameters. The first component of GMM3 implies the situation when vehicles travel at desired speeds in the "green wave" of signal coordination along the segment. The second component is for partially delayed vehicles, which were caught in red phase at certain intersections. The third component corresponds to congested situations, while the mixture weights of 0.10 for segment 1 and 0.09 for segment 2 imply that the path experienced only a short duration of congestions. The parameters of GMM3 represent the characteristics of traffic conditions in a quantitative manner and help provide comprehensive insights into the heterogeneity of segments TTD.

4.2.3. Optimal copula model selection

As aforementioned, Kendall's tau for Segment 1 and Segment 2 is -0.164. It indicates the dependence between Segment 1 and Segment 2 travel times is weak and negative. Thus, the candidate copulas are AMH, Frank, FGM and Gaussian Copula. Here we present the results of estimated parameters θ of different copulas in Table 4.

After estimating the parameters of different copulas, the best-fitting copula model needs to be determined. Table 5 represents Log-likelihood, AIC values and CvM statistical tests of different copulas. The favorable copula is supposed to meet the requirements, namely (1) the biggest *p-value* in the case of smaller CvM statistic; (2) a larger Log likelihood value and smaller AIC value. Overall, FGM copula is found as the best-fitting copula model in this case.

The scatter plots of empirical observations and an artificial dataset of the sample size generated from the fitted FGM copula are presented in Fig. 6. It indicates that the copula model can reasonably reconstruct the 2D segment travel times with certain correlation.

4.2.4. Estimation of path TTD

Path TTD estimation is the preliminary preparation for TTR analysis. In general, there are two approaches for estimating path TTD. One is to identify the best statistical model for fitting travel time observations; the other is to model travel time observations for path TTD prediction.

Unimodal distributions, e.g., Normal, Lognormal, Gamma and Weibull distributions were first used to characterize TTD for the given path. Compared with unimodal distributions, GMM has also become of interest (Chen et al., 2016). Based on empirical observations of path TTD, GMM2 was selected as a representative of multimodal distributions. The fitting results of unimodal distributions and GMM2 are shown in Fig. 7. For a more comprehensive comparison, Log-likelihood, AIC values and the skewness and width of TTD for empirical observations are presented in Table 6. It indicates that Gamma distribution and

Table 2
Hourly dependence between Segments 1 and 2 for the period of 9:00–11:00 a.m.

Time period	Count (vehicles)	Segments 1 and 2		
		ρ_P	τ	ρ_S
9:00–11:00 a.m.	664	−0.21	**−0.164**	−0.25

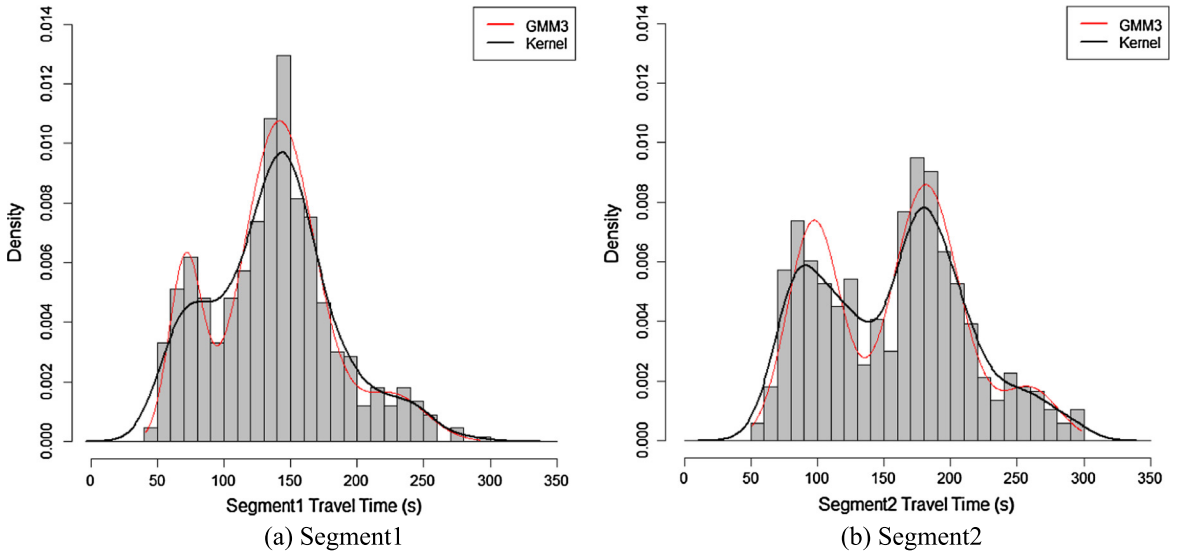


Fig. 5. Fitting Curves of GMM and Kernel Smooth for Changshou Road.

Table 3
The parameters of GMM3 for Changshou Road.

Segment	Distribution	Mean (μ)	Sigma^2 (σ^2)	Weight (π)
Segment 1	GMM3	(71.20, 141.65, 226.30)	(151.68, 704.02, 628.85)	(0.18, 0.72, 0.10)
Segment 2	GMM3	(97.43, 181.33, 259.45)	(402.30, 619.14, 452.39)	(0.37, 0.54, 0.09)

Table 4
Estimated parameters θ of different copulas based on Kendall's tau τ for Changshou Road.

τ	AMH	FGM	Frank	Gaussian
−0.164	−0.89	−0.74	−1.51	−0.25

Table 5
Goodness-of-fit tests of different copulas for Changshou Road.

	Goodness-of-fit statistics	AMH	FGM	Frank	Gaussian
Segment1 and Segment2	CvM	0.0377	**0.0209**	0.0251	0.0211
	p-value	0.0175	**0.2872**	0.1094	0.2722
	Log-likelihood	−6969.009	**−6964.12**	−6963.441	−6973.73
	AIC	13958.02	**13948.24**	13946.88	13967.46

GMM2 perform relatively better. When further referring to Log-likelihood and AIC values, GMM2 is identified as the best fitting model.

From the aforementioned analysis, it is clear that the segment-level TTDs are more complex than the path-level TTDs. For segment travel times, there exists explicit multimodal distribution phenomenon, since travel times are more sensitive to intersections and stop delays in a shorter distance. While for the entire path with a longer distance, travel times show approximately a unimodal distribution. One possible reason is that with the increasing number of intersections, TTD would become more dispersed and the impact of traffic congestion developed at a certain part on urban road upon path TTD is not explicit compared with a shorter segment.

The path TTDs estimated by FGM Copula and the convolution model are shown in Fig. 8. It is evident that the estimation by the copula model is closer to the empirical PDF and CDF than the convolution model. To confirm that, the K-S tests for the estimated TTDs by FGM copula, the convolution model and GMM2 were conducted, as presented in Table 7. The results confirm the advantage of FGM copula over the convolution and the distribution fitting method, and demonstrate the statistical superiority of the proposed copula approach.

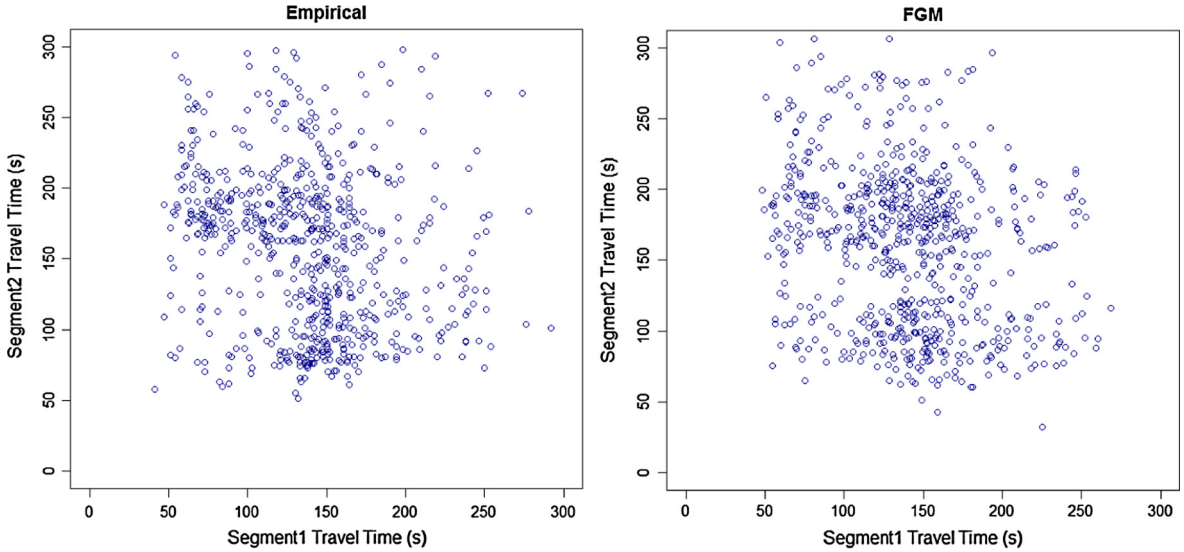


Fig. 6. Empirical and simulated scatter diagrams.

4.2.5. Path TTR evaluation

TTR has been increasingly recognized as a critical factor contributing to the efficiency and service quality of transportation system. The copula model provides a flexible and superior TTD estimation than its alternatives, which helps provide the maximum information for reliability analysis. It thus enables a comprehensive assessment of any path TTR of interest.

In this section, the path TTR for different models will be quantified and compared. To be specific, path TTR can be estimated by Eq. (19) when ignoring segments correlation impact, and by Eq. (31) when taking segments correlation impact into consideration.

Based on the maximum speed limit along the path (i.e., 40 km/h), the free travel times were defined as 73 s and 89 s for Segment 1 and Segment 2, respectively. Thus, the free travel time for the entire path is 162 s. Besides, different reliability parameters λ were enumerated to compare different models with the empirical reliability quantification. The values of reliability parameter λ report the standards for the expected travel time. A smaller λ value indicates that travelers' expectation for travel time in a certain path is higher, and travelers want to reach the destination in faster speeds. Conversely, if λ value is larger, it suggests that people allow for longer travel times along the path. Table 8 displays the path TTR for different models.

The results show that path TTR constantly increases with the increase of parameter λ. When ignoring the segments correlation impact, path TTR will be underestimated. Thus, it is necessary to incorporate the segment correlation into consideration when estimating path TTD. Specifically, when $\lambda < 1.7$, path TTR was overestimated by the convolution model, while the copula model performs desirably compared with the observation. For travelers, they neither want to reach the destination extremely early, nor late for the date. The accurate path TTR information assists them in responding to travel time uncertainty in a reasonable way. Overall, the copula model is a promising approach to calculate path TTR.

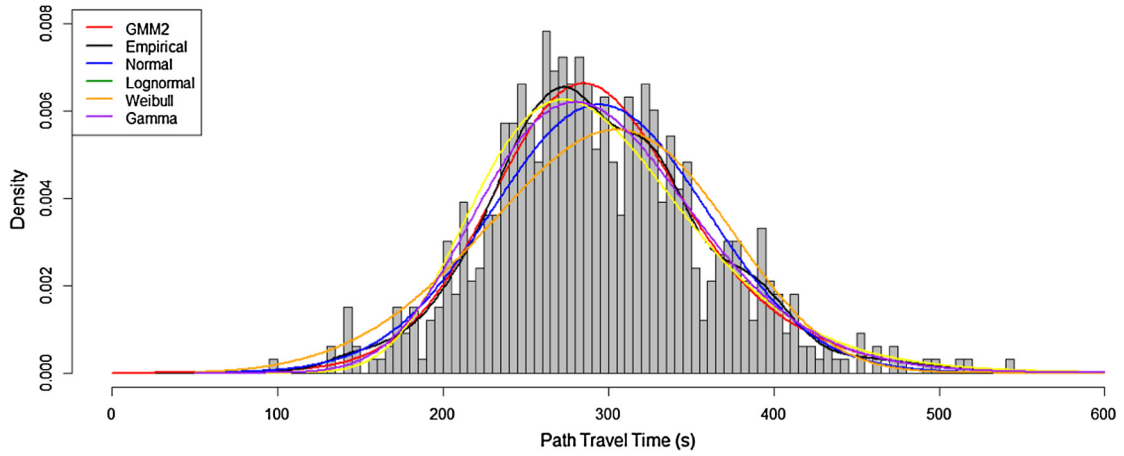
After calculation of path TTR, the skewness and width of TTDs are of interest. As presented in Table 9, the results indicate that the skewness and width estimated by the copula model are closer to the empirical observations. The convolution model, which fails to consider the correlation between segments travel time, cannot characterize the skewness of path TTD accurately. The results help emphasize the importance of accounting for the correlation between segment travel times when estimating path TTD and reliability related measures.

4.3. Application of copula models in Lankershim Street

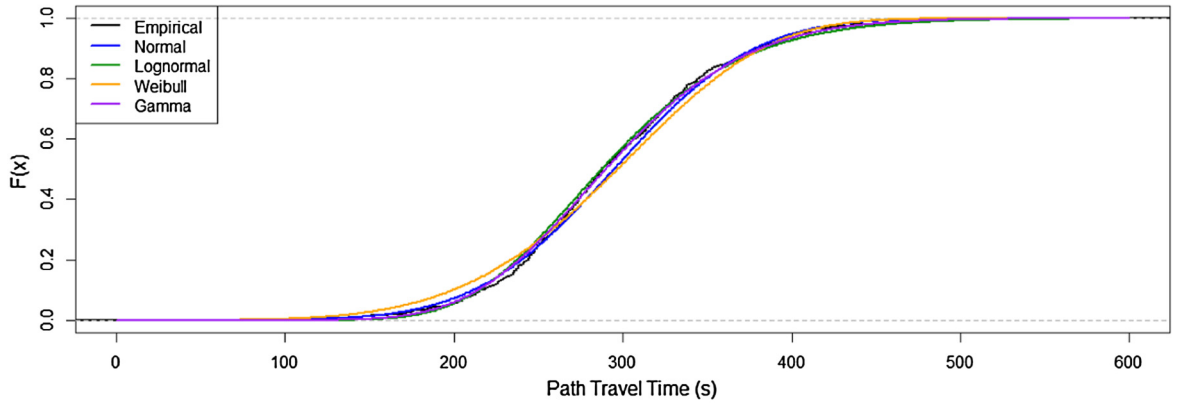
The case study of Changshou Road with only 2 segments is of limitation. In order to examine whether the proposed copula framework is applicable and transferable to different case studies, the NGSIM data were applied as another case study. Here the NGSIM data in Lankershim Street were used to construct copula models with marginal distributions of both segment-level and link-level travel times.

4.3.1. Dependence between links

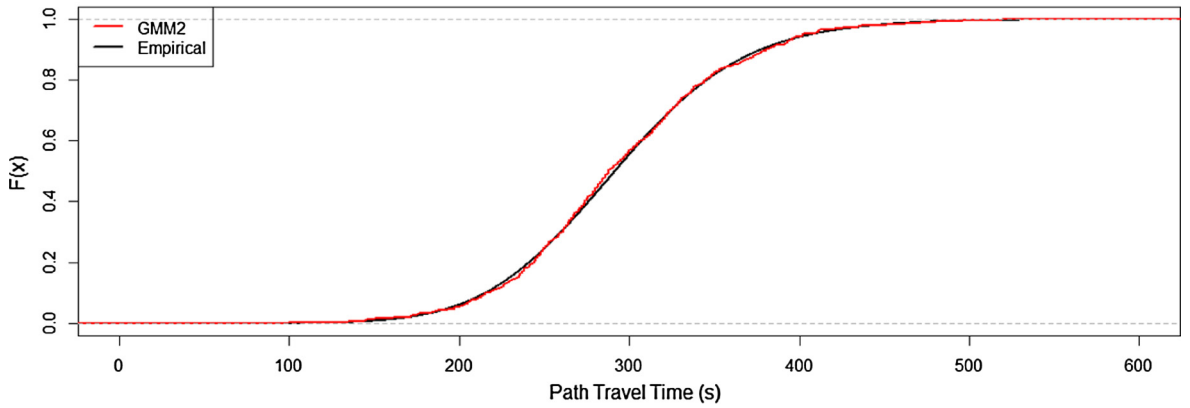
The correlations of both segment-level and link-level travel times estimated by Kendall's tau are presented in Table 10. Similar to the case study based on AVI data, the results illustrate that the dependence structure between segments travel time shows weak correlation in general. Link travel times show both negative and positive correlations, while the depen-



(a) PDF of path travel time



(b) CDF of path travel time (Unimodal)



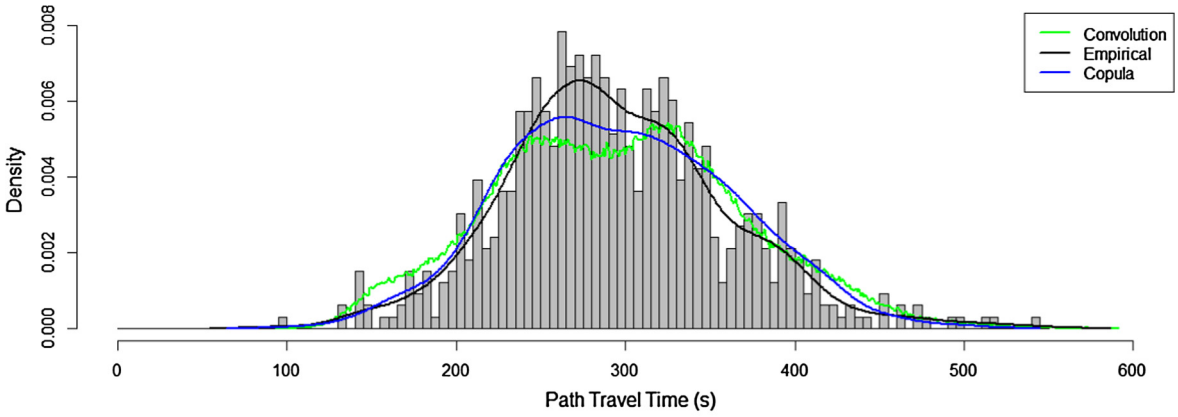
(c) CDF of path travel time (GMM2)

Fig. 7. Path TTD fitting for Changshou Road.

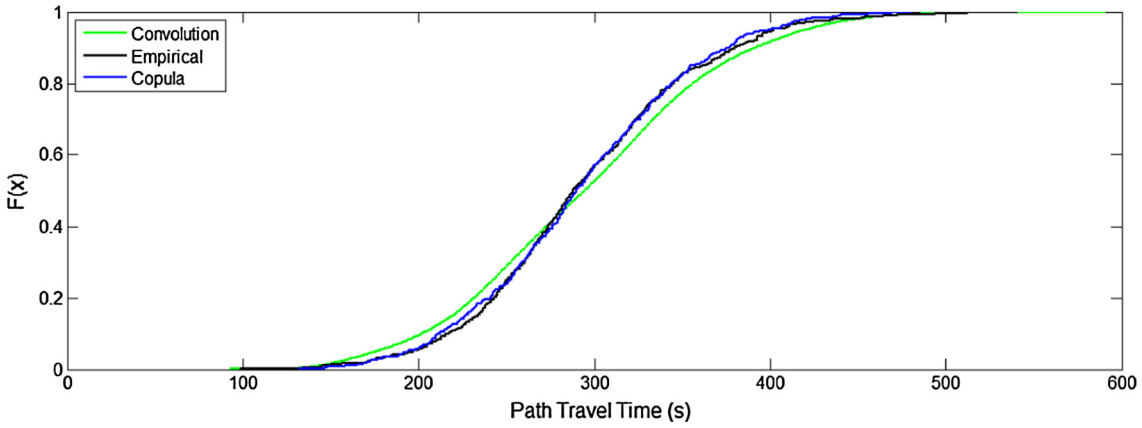
dences between segment travel times are all negative. This is the effect of spatial aggregation. For segment-level travel times, the correlation between Link1-2 and Link3 is so weak that it is difficult to characterize. Thus, Link1 and Link2-3 travel times are used to construct the segment-level copula model.

Table 6
Fitting results of different distributions for path TTD of Changshou Road.

Type of distribution	Log-likelihood	AIC	s	λ^{var}
Weibull	−3734.31	7472.629	−0.23	0.62
Normal	−3713.13	7430.259	0	0.57
Log normal	−3717.83	7439.654	0.96	0.59
Gamma	−3709.53	7423.067	0.50	0.58
GMM2	−3705.09	7414.180	0.49	0.58
Empirical	\	\	0.37	0.57



(a) PDF of path travel time



(b) CDF of path travel time

Fig. 8. Path TTD estimation for Changshou Road.

Table 7
K-S statistics of real TTD versus estimated TTDs for Changshou Road.

Model	K-S Test		
	h	*p-value*	k-s stat
FGM (GMM3)	**0**	**0.8141**	**0.0346**
Convolution	1	0.0079	0.0643
GMM2	0	0.4236	0.0482

4.3.2. Investigation of link TTD

The marginal distributions of both link travel times and segment travel times need to be specified separately. Based on the results of kernel smoothing estimation, link travel times indicate a multimodal distribution under the temporal aggre-

Table 8
Path TTR estimation by different models for Changshou Road.

λ	Path TTR		
	Empirical	Convolution	FGM Copula
1.5	0.2003	0.2544	0.2364
1.6	0.2967	0.3339	0.3223
1.7	0.4111	0.4114	0.4111
1.8	0.5169	0.4859	0.5256
1.9	0.6024	0.5619	0.6175
2.0	0.7018	0.6506	0.7063

Table 9
The skewness and width of TTD for Changshou Road.

TTR	Empirical	Convolution	Copula
S	0.37	0.57	**0.36**
λ^{var}	0.57	0.65	**0.58**

Table 10
Dependence between different segments for Lankershim Street.

Coefficients	Link1 & Link2	Link2 & Link3	Link1 & Link3	Link1 & Link2-3	Link1-2 & Link3
τ	**−0.14**	**0.14**	**−0.16**	**−0.23**	−0.05

gation of 15 min. Thus, GMM is used to fit individual link travel times and Link2-3 travel times as illustrated in Fig. 9. For further comparison, the K-S test, Log-likelihood and AIC values are used to examine the goodness of fit for 2-component and 3-component GMMs, as shown in Table 11. h = 0 indicates the fact that path travel time is from the specific model can be rejected at the 95% confidence interval. Larger Log-likelihood and smaller AIC values also indicate a better fitting result. Table 11 shows that GMM3 is a good fitting model for Link1, Link2 and Link2-3, while GMM2 is more appropriate for Link3. The GMM parameters of different link TTDs are presented in Table 12.

4.3.3. Optimal copula model selection

After the estimation of link and segment TTDs, an optimal copula model should be determined. Based on the dependence identified in Table 10, bivariate copulas and trivariate copulas appear appropriate to represent dependent link and segment travel times.

For bivariate copula, we considered Link1 as a segment with one link and Link2-3 as a segment with 2 links for modeling. Compared with Kendall's tau (−0.05) for Link1-2 and Link3 travel times, Link1 and Link2-3 with Kendall's tau (−0.23) are more appropriate to construct copula models. Thus, Link1 and Link2-3 travel times were chosen as marginal distributions for modeling. Given the dependence between Link1 and Link2-3, the candidate copulas are Frank and Gaussian. Table 13 presents the results of estimated parameters θ for the two copulas.

After estimating the parameters of different copulas, the best-fitting copula model needs to be determined. Goodness-of-fit tests of different copulas are presented in Table 14. The optimal copula model is supposed to meet the requirements, namely (1) the biggest *p-value* in the case of smaller CvM statistic; (2) larger Log-likelihood; (3) smaller AIC values. Overall, Gaussian copula is determined as the best-fitting copula model.

For trivariate copula, considering that the correlation between link travel times in Lankershim Street is not only negative but also positive, Gaussian copula was adopted to model path TTD. The corresponding parameters and the estimation results are presented in Table 15.

4.3.4. Estimation of path TTD

Before estimating path TTD, the Kernel smoothing was performed to analyze path TTD characteristics, as illustrated in black curves in Fig. 10. It was found that path TTD shows a multimodal distribution. Thus, we first chose GMM with different components to fit path TTD. The fitting results are shown in Fig. 10. For a more comprehensive comparison, Log-likelihood, AIC values and K-S test results are provided in Table 16. GMM3, with larger Log-likelihoods, smaller AIC values and h = 0, indicates the best fitting model.

The path TTDs estimated by the copula models, the convolution models and GMM3 are shown in Fig. 11. Considering that the path consists of three links, 3-dimensional (3D) models enable TTD estimation from individual link-level, while 2-dimensional (2D) models enable the estimation performed at segment-level, i.e., one segment may consist of more than

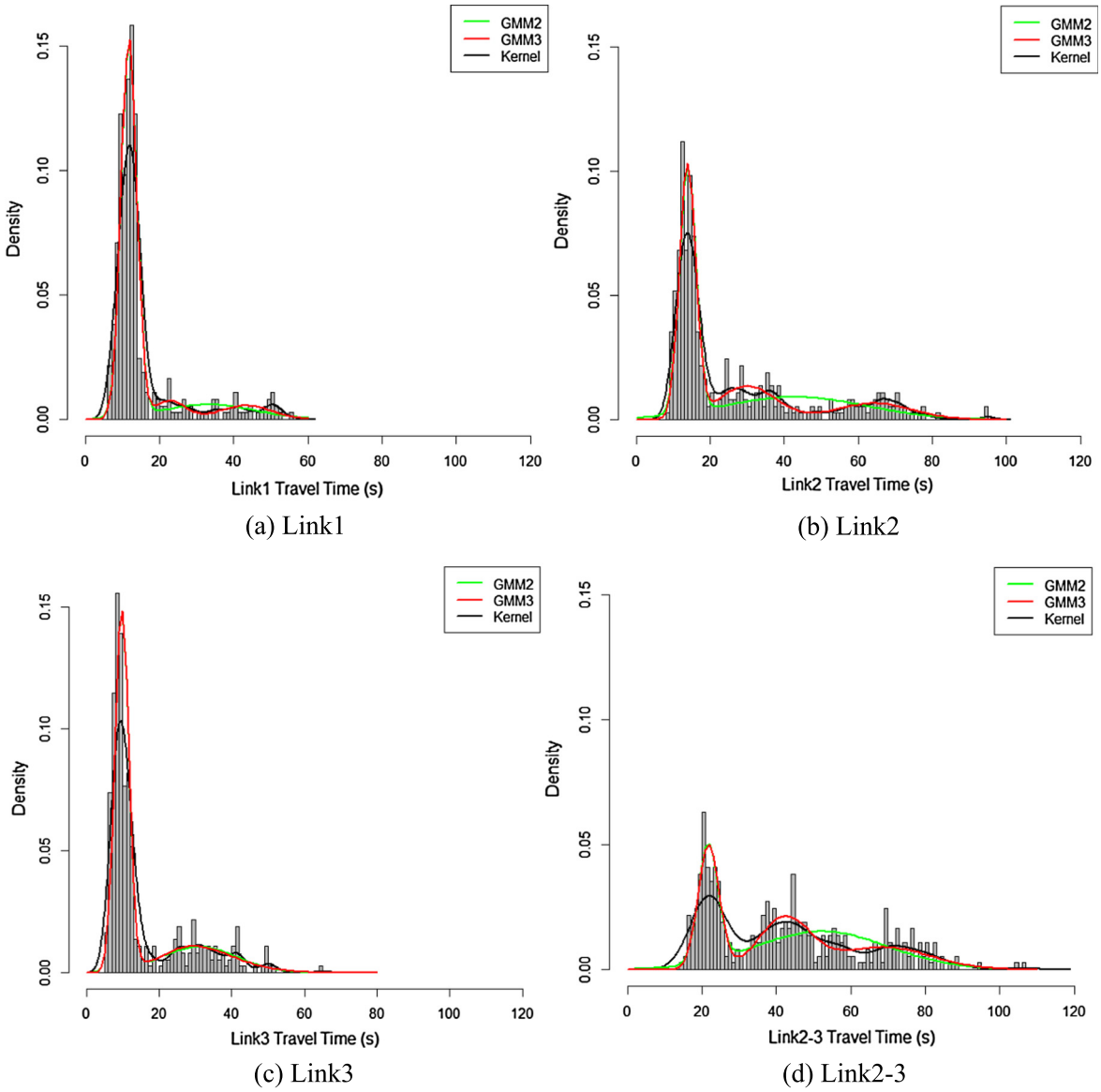


Fig. 9. Link and segment TTD fitting for Lankershim Street.

one link. In Fig. 11, it is evident that the 3D convolution model performs worse than the 3D copula model. Compared to the 2D convolution model, the path TTD estimated by the 2D copula model is closer to the empirical observations.

Table 11
Log-likelihood, AIC values and K-S tests for different link and segment GMMs in Lankershim Street.

Segment	Distribution	Log-likelihood	AIC	K-S test		
				h	*p-value*	k-s stat
Link1	**GMM3**	**−1060.144**	**2138.287**	**0**	**0.8683**	**0.0437**
	GMM2	−1068.232	2148.463	0	0.9727	0.0355
Link2	**GMM3**	**−1342.650**	**2703.299**	**0**	**0.5106**	**0.0601**
	GMM2	−1351.734	2751.467	0	0.4530	0.0628
Link3	GMM3	−1125.557	2269.113	0	0.9128	0.0410
	GMM2	**−1126.003**	**2264.006**	**0**	**0.9999**	**0.0246**
Link2-3	**GMM3**	**−1511.618**	**3043.236**	**0**	**0.9960**	**0.0301**
	GMM2	−1532.206	3078.411	0	0.0964	0.0902

Table 12
The parameters of different segment GMMs for Lankershim Street.

Segment	Distribution	Mean (μ)	Sigma^2 (σ^2)	Weight (π)
Link1	GMM3	(11.72, 22.69, 43.21)	(4.52, 17.85, 52.41)	(0.82, 0.08, 0.10)
Link2	GMM3	(13.85, 29.89, 64.27)	(4.85, 61.10, 109.01)	(0.56, 0.27, 0.17)
Link3	GMM2	(9.71, 30.91)	(3.69, 113.92)	(0.71, 0.29)
Link2-3	GMM3	(21.79, 42.14, 68.59)	(8.18, 46.97, 167.33)	(0.36, 0.35, 0.29)

Table 13
The estimation of parameters θ of different copulas based on Kendall's tau τ(NGSIM data: segment-level).

τ	Gaussian	Frank	AMH	FGM
−0.23	−0.35	−2.13	NA	NA

Table 14
Goodness-of-fit test results of different copulas for Lankershim Street.

	Goodness-of-fit statistics	**Gaussian**	Frank
Link1 and Link2-3	CvM	**0.0248**	0.0287
	p-value	**0.1454**	0.0784
	Log-likelihood	**−2551.31**	−2555.30
	AIC	**5140.62**	5148.61

Table 15
Trivariate Gaussian parameters (NGSIM data: link-level).

Parameter			Trivariate Gaussian Copula		
$\tau_{link1\&2}$	$\tau_{link2\&3}$	$\tau_{link1\&3}$	θ	Log-Likelihood	AIC
−0.14	0.15	−0.17	(−0.22, 0.23, −0.26)	−3577.67	7180.34

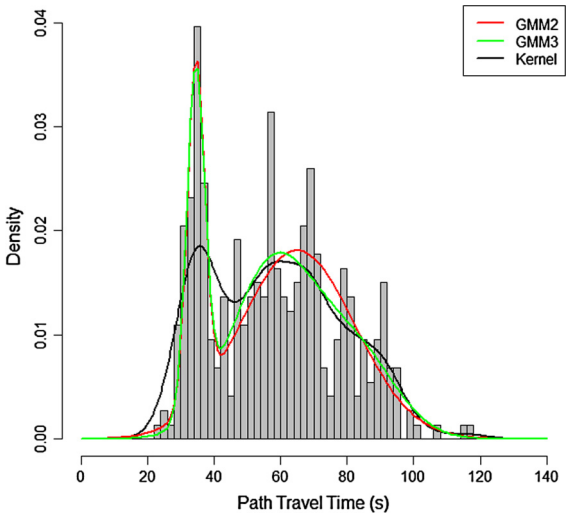
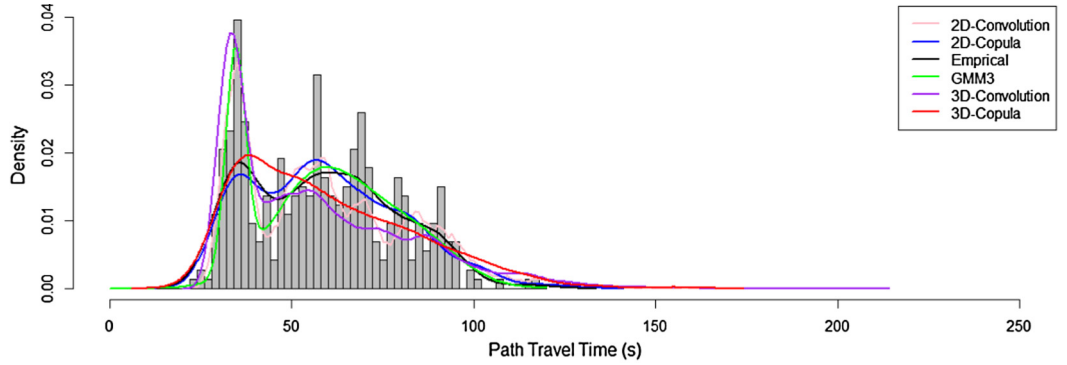


Fig. 10. Path TTD fitting with GMM for Lankershim Street.

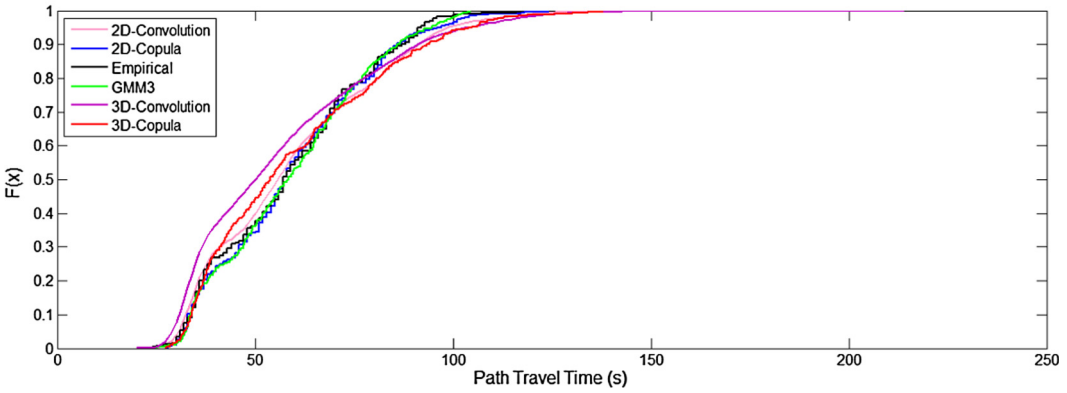
For further comparison, the K-S statistics for different models are presented in Table 17. Worth noticing is that the copula model utilizes the link-level or segment-level TTDs to estimate the unknown path-level TTD, while the distribution fitting methods, e.g., GMM3, attempt to approximate the known path-level TTD. The results in Table 18 help confirm the statistical superiority of the proposed copula approach. Besides, due to not considering link correlation, the convolution model performs worse with the increasing dimension, i.e., the number of links.

Table 16
Goodness-of-fit test results of Path TTD with GMM for Lankershim Street.

Segment	Distribution	Log-likelihood	AIC	K-S Test		
				h	*p-value*	k-s stat
Path	GMM2	−1558.97	3129.932	0	0.5106	0.0601
	GMM3	**−1556.83**	**3131.669**	0	0.8159	0.0464



(a) PDF of Path travel time



(b) CDF of Path travel time

Fig. 11. Path TTD estimation for Lankershim Street.

Table 17
K-S statistics of real TTD versus estimated TTDs for Lankershim Street.

Model	K-S Test		
	h	*p-value*	k-s stat
3D-Copula	0	0.3990	0.0656
3D-Convolution	1	0	0.1451
GMM3	0	0.5106	0.0601
2D-Copula	**0**	**0.9128**	**0.0410**
2D-Convolution	0	0.0866	0.0650

4.3.5. Path TTR evaluation

After the TTD estimation, path TTR was calculated by referring to Eq. (19). Based on the maximum speed limit along the path (i.e., 35 mph), the free flow travel time for the entire path is estimated to be 31 s. The path TTR estimation results are presented in Table 18. A smaller λ indicates a higher standard for expected travel time by the travelers, vice versa. It was found that when $\lambda > 2.3$, path TTR is up to 75%. Overall, the convolution model overestimated path TTR, while the estimation results by the copula model agree well with the empirical observations.

Table 18
Path TTR for different models of Lankershim Street.

λ	Path TTR		
	Empirical	2D-Convolution	2D-Copula
1.5	0.3169	0.3533	0.306
1.6	0.3607	0.3985	0.3415
1.7	0.4044	0.4518	0.388
1.8	0.4399	0.5049	0.4563
1.9	0.5246	0.5617	0.5328
2.0	0.5847	0.608	0.5956
2.1	0.623	0.6541	0.6393
2.2	0.6803	0.6915	0.6885
2.3	0.7486	0.7272	0.735
2.4	0.7814	0.752	0.7678
2.5	0.8005	0.7725	0.7869

Table 19
The skewness and width of TTD by different models for Lankershim Street.

TTR	Empirical	2D-Convolution	2D-Copula
s	0.32	0.47	**0.34**
λ^{var}	0.93	1.07	**0.93**

In addition, the skewness and width of TTDs by different models are of interest. As presented in Table 19, the skewness and width estimated by the copula model are closer to the empirical observations, which are consistent with the case study of Changshou Road based on AVI data. Note that the skewness of TTDs in both case studies are all positive which indicate that the path TTDs spread more to the right side toward larger values. With respect to the width of TTD λ^{var}, a larger value indicates a relatively volatile traffic states and unreliable travel time. Compared with Table 9, travel times in Changshou Road appears more reliable than Lankershim Street.

5. Conclusions and future work

This paper introduced a copula-based approach to characterize the dependent structure between segments for a given path on arterials and then aggregated the individual segment TTDs to estimate path TTD by accounting for spatial segment correlation. The method does not make any assumptions for the model input variables. Path TTD estimated by the copula model were compared to those by the convolution and the empirical distribution fitting approach in both AVI data case and NGSIM data case. Last, path TTR was evaluated by three indicators for different approaches. The main findings are summarized below:

(1) Segment-level TTDs can hardly be represented by unimodal distributions. Instead, segment TTDs showed apparent multimodal distributions and implied significantly different states as a result of spatial and temporal aggregation.
(2) Segment travel times were found spatially and weakly dependent at the study sites. Such week correlation could be regarded as the comprehensive effect of temporal-spatial aggregation and heterogeneous driver characteristics.
(3) For path TTD estimation, the copula model with marginal distributions from segment-level travel times performs better than from link-level. The statistical superiority of the copula method over the convolution and the distribution fitting methods was also demonstrated.
(4) For path TTR evaluation, when ignoring segments correlation impact, it was found that path TTR was overestimated by the convolution model within certain ranges. By comparison, the path TTR and the skewness and width estimated by the copula model were found closer to the empirical observations in both case studies.

It is noteworthy that in this study we only tested Gaussian copula when modeling three continuous links travel time using NGSIM data. The results and conclusions derived above may be limited in scope. To extend initial findings, it will be necessary to test other types of copulas for an increasing number of links. Furthermore, to improve the applicability of the proposed methodology in field implementation, experiments with real day-to-day data (e.g. GPS data by probe vehicles) would be conducted. The inputs to the proposed copula model are link travel times and since they are not directly reported by probe vehicles, the estimation of link travel times from GPS data is of interest (Hellinga et al., 2008; Ramezani et al., 2015). GPS data by probe vehicles (which need to be decomposed from path-level to segment-level), will help examine the practical estimation of path TTDs. These are directions for our future research.

Acknowledgment

The authors appreciate the National Natural Science Foundation of China (51508014, U1564212 and 51408538) and the Open Project of Beijing Key Laboratory of Urban Road Intelligent Traffic Control for kind support for this research.

References

Ali, M.M., Mikhail, N.N., Haq, M.S., 1978. A class of bivariate distributions including the bivariate logistic. J. Multivar. Anal. 8 (3), 405–412.
Bates, J., Polak, J.W., Jones, P., Cook, A., 2001. The valuation of reliability for personal travel. Transp. Res. Part E 37 (2–3), 191–229.
Breymann, W., Dias, A., Embrechts, P., 2003. Dependence structures for multivariate high-frequency data in finance. Quant. Finance 3 (1), 1–14.
Bhaskar, A., Chung, E., Dumont, A., 2009. Integrating cumulative plots and probe vehicle for travel time estimation on signalized urban network. In: The 9th Swiss Transport Research Conference, Monte Varita..
Bhat, C.R., Eluru, N., 2009. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Transp. Res. Part B 43 (7), 749–765.
Bowman, A., Azzalini, A., 1997. Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations. Oxford University Press, US.
Chen, P., Liu, H., Qi, H., Wang, F., 2013. Analysis of delay variability at isolated signalized intersections. J. Zhejiang Univ. Sci. A 14 (10), 691–704.
Chen, P., Yin, K., Sun, J., 2014. Application of finite mixture of regression model with varying mixing probabilities to estimation of urban arterial travel times. Transp. Res. Rec.: J. Transp. Res. Board 2442, 96–105.
Chen, P., Sun, J., Qi, H., 2017. Estimation of delay variability at signalized intersections for urban arterial performance evaluation. J. Intell. Transp. Syst. 21 (2), 94–110.
Eman, E.B., Al-Deek, H.M., 2006. Using real-life dual-loop detector data to develop new methodology for estimating freeway travel time reliability. Transp. Res. Rec.: J. Transp. Res. Board 1959, 140–150.
Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls. Risk Manage. Value Risk Beyond, 176–223.
Frank, M.J., 1979. On the simultaneous associativity of F(x, y) and x+y-F(x, y). Aequationes Math. 19 (1), 194–226.
Feng, Y., Hourdos, J., Davis, G.A., 2012. Bayesian model for constructing arterial travel time distributions using GPS probe vehicles. In: The Transportation Research Board 91st Annual Meeting, Washington D.C..
Genest, C., Quessy, J.F., Rémillard, B., 2006. Goodness-of-fit procedures for copula models based on the integral probability transformation. Scand. J. Stat. 33 (2), 337–366.
Genest, C., Favre, A.C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. J. Hydrol. Eng. 12 (4), 347–368.
Genest, C., Rémillard, B., Beaudoin, D., 2009. Goodness-of-fit tests for copulas: a review and a power study. Insur.: Math. Econ. 44 (2), 199–213.
Geroliminis, N., Skabardonis, A., 2006. Real time vehicle reidentification and performance measures on signalized arterials. In: IEEE Intelligent Transportation Systems Conference, Toronto, Canada.
Guo, F., Rakha, H., Park, S., 2010. Multistate model for travel time reliability. Transp. Res. Rec.: J. Transp. Res. Board 2128, 46–54.
He, R.R., Liu, H.X., Kornhauser, A.L., Ran, B., 2002. Temporal and Spatial Variability of Travel Time. Institute of Transportation Studies, UC Irvine.
Hellinga, B., Izadpanah, P., Takada, H., Fu, L., 2008. Decomposing travel time measured by probe-based traffic monitoring systems to individual road segments. Transp. Res. Part C 16 (6), 768–782.
Herring, R., Hofleitner, A., Abbeel, P., Bayen, A., 2010. Estimating arterial traffic conditions using sparse probe data. In: IEEE Intelligent Transportation Systems Conference, Madeira, Portugal.
Highway Capacity Manual, 2010. Transportation Research Board. National Research Council, Washington, D.C..
Hofert, M., Mächler, M., Mcneil, A.J., 2012. Likelihood inference for Archimedean copulas in high dimensions under known margins. J. Multivar. Anal. 110 (5), 133–150.
Hollander, Y., Liu, R., 2008. Estimation of distribution of travel times by repeated simulation. Transp. Res. Part C: Emerg. Technol. 16 (2), 212–231.
Iida, Y., 1999. Basic concepts and future directions of road network reliability analysis. J. Adv. Transp. 33 (2), 125–134.
Ji, Y., Zhang, M.H., 2013. Travel time distributions on urban streets: their estimation with a hierarchical bayesian mixture model and applications to traffic analysis using high-resolution bus probe data. In: The Transportation Research Board 92nd Annual Meeting, Washington D.C..
Kazagli, E., Koutsopoulos, H.N., 2012. Estimation of arterial travel time from automatic number plate recognition data. Transp. Res. Rec.: J. Transp. Res. Board 2391, 22–31.
Kendall, M.G., Gibbons, J.D., 1990. Rank Correlation Methods. Griffin, London.
Killmann, F., von Collani, E., 2001. A note on the convolution of the uniform and related distributions and their use in quality control. Econ. Qual. Control 16 (1), 17–41.
Liu, H., Ma, W.T., 2009. A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials. Transp. Res. Part C: Emerg. Technol. 17 (1), 11–26.
Ma, Z., Ferreira, L., Mesbah, M., Zhu, S., 2016. Modeling distributions of travel time variability for bus operations. J. Adv. Transp. 50 (1), 6–25.
Mohammad, A., Tobias, E., Ugur, D., Cyrus, S., 2015. Probabilistic estimation of link travel times in dynamic road networks. In: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems.
NGSIM, 2005. Next Generation Simulation. <http://ngsim.fhwa.dot.gov>.
Pattanamekar, P., Park, D., Rilett, L., Lee, J., Lee, C., 2003. Dynamic and stochastic shortest path in transportation networks with two components of travel time uncertainty. Transp. Res. Part C: Emerg. Technol. 11 (5), 331–354.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 <http://www.R-project.org>.
Rakha, H.A., EI-Shawarby, I., Arafeh, M., Dion, F., 2006. Estimating path travel-time reliability. In: IEEE Intelligent Transportation Systems Conference, ITSC'06. IEEE, pp. 236–241.
Ramezani, M., Geroliminis, N., 2012. On the estimation of arterial route travel time distribution with Markov chains. Transp. Res. Part B: Meth. 46, 1576–1590.
Ramezani, M., Geroliminis, N., 2015. Queue profile estimation in congested urban networks with probe data. Comput.-Aided Civil Infrastruct. Eng. 30 (6), 414–432.
Rasmussen, C.E., 1999. The infinite Gaussian mixture model. Adv. Neural. Inf. Process. Syst. 159 (16), 554–560.
Skabardonis, A., Geroliminis, N., 2005. Real-time estimation of travel times on signalized arterials. In: Proceedings of the 16th International Symposium on Transportation and Traffic Theory, pp. 387–406.
Skabardonis, A., Geroliminis, N., 2008. Real-time monitoring and control on signalized arterials. J. Intell. Transp. Syst. 12 (2), 64–74.
Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris 8, 229–231.
Sklar, A., 1973. Random variables, joint distribution functions, and copulas. Kybernetika 9 (6), 449–460.
Strategic Highway Research Program (SHRP 2) Report S2–L02-RR-2, 2014. Guide to Establishing Monitoring Programs for Travel Time Reliability.
Strategic Highway Research Program (SHRP 2) Report S2–L07-RR-2, 2014. Design Guide for Addressing Nonrecurrent Congestion.
Strategic Highway Research Program (SHRP 2) Report S2–L14-RW-1, 2014. Effectiveness of Different Approaches to Disseminating Traveler Information on Travel Time Reliability.

Susilawati, S., Taylor, M.A., Somenahalli, S.V., 2013. Distribution of travel time variability. J. Adv. Transp. 47 (8), 720–736.
Trivedi, P.K., Zimmer, D.M., 2007. Copula modeling: an introduction for practitioners. Foundations and Trends in Econometrics 1(1), Now Publishers, 2007.
Uno, N., Kurauchi, F., Tamura, H., Iida, Y., 2009. Using bus probe data for analysis of travel time variability. J. Intell. Transp. Syst. 13 (1), 2–15.
Van Lint, J.W.C., Van Zuylen, H.J., Tu, H., 2008. Travel time unreliability on freeways: why measures based on variance tell only half the story. Transp. Res. Part A: Policy Pract. 42 (1), 258–277.
Westgate, B., Woodard, D., Matteson, D., Henderson, S., 2016. Large-network travel time distribution estimation for ambulances. Eur. J. Oper. Res. 252 (1), 322–333.
Zeng, W., Miwa, T., Wakita, Y., Morikawa, T., 2015. Application of lagrangian relaxation approach to α-reliable path finding in stochastic networks with correlated link travel times. Transp. Res. Part C: Emerg. Technol. 56, 309–334.
Zheng, F., van Zuylen, H.J., 2010. Uncertainty and predictability of urban link travel time: a delay distribution based analysis. Transp. Res. Rec.: J. Transp. Res. Board 2259, 80–95.
Zou, Y., Zhang, Y., 2016. A copula-based approach to accommodate the dependence among microscopic traffic variables. Transp. Res. Part C: Emerg. Technogl. 70, 53–68.
Zou, Y., Zhang, Y., Zhu, X., 2014. Constructing a bivariate distribution for freeway speed and headway data. Transportmetrica A: Transp. Sci. 10 (3), 255–272.