

Received January 14, 2020, accepted January 28, 2020, date of publication January 30, 2020, date of current version February 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2970530

Copula-Based Travel Time Distribution Estimation Considering Channelization Section Spillover

YI YU¹, MENGWEI CHEN², HONGSHENG QI¹, (Member, IEEE), AND DIANHAI WANG¹

¹Department of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China

²College of Civil Engineering and Architecture, Zhejiang University of Technology, Hangzhou 310014, China

Corresponding author: Dianhai Wang (eveyu@zju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61773338.

ABSTRACT Travel time is inherently uncertain in urban networks due to volatile traffic flows, signal controls, bus stops and disturbances from pedestrians. An effective way to characterize such uncertainty is by estimating Travel Time Distribution (TTD). However, conventional TTD models are lack of considering the interactions between turning movements within the link. Whereas the phenomenon of the Channelization Section Spillover (CSS) is very common, leading to strong interactions between turning movements. In this study, by incorporating the correlation of turning movements in the TTD model, that is, considering CSS, copula-based link-level and path-level TTD models are built. First, based on the empirical data, the correlations of turning movements are analyzed, and then the applicability of the various copula models are examined. The marginal distribution of each turning movement is described using parametric and non-parametric regression analysis, respectively. Then, the best-fitting copula is determined based on correlation parameters and the goodness-of-fit tests. As a case study, the chosen model is applied in estimating link-level and path-level TTD in an arterial road in Hangzhou, China, and compared with the model that did not consider the CSS before. Both results indicate that the copula-based approach can precisely capture the positively correlated relationship between turning movements during peak hours. Furthermore, higher TTD estimation accuracy demonstrates the significance to consider CSS, particularly in peak hours.

INDEX TERMS Travel time distribution, turning movements correlation, channelization section spillover, copula.

I. INTRODUCTION

Travel Time Distribution (TTD) is the fundamental component of the reliability and variability of travel time estimation, which is essential for route guidance [1] and advanced traveler information system in urban traffic [2]. TTD is also useful to do travel time prediction on highways [3], evaluate traffic control policies [4] and to provide detailed model support for simulation software. What's more, it is TTD that enables detecting congestions and incidents and accessing the level of intersection service, which is crucial to traffic operators and planners [5].

TTD estimation is challenging because of the complexity and interrupted nature of urban traffic flows [6], usually caused by signal control, bus stops, roadside parking, disturbances from pedestrians and vehicle interactions. Under such complicated interactions, arterial travel times often exhibit distinct shapes of distributions. Thus, most studies

have focused on the estimation of the average travel time. No study of the TTD model has been found to consider the correlation of turning movements within the link. Moreover, rarely work has been done to quantify the variability of either link-level or path-level TTD and explore the interdependence between link TTDs along arterials. Some related work about link and path TTD is provided below.

Link-level TTD can capture the nature of interrupted flow under traffic signal control and has wide applications in practice. Initially, based on Automatic Number Plate Recognition (ANPR) data, link TTD can be directly obtained and estimated by assuming unimodal distributions, like Weibull, exponential, lognormal [7], or normal distributions. For instance, Hunter proposed a probabilistic model and employed an EM algorithm to simultaneously learn the likely paths and the parameters of lognormal link travel time distributions using the travel time observations [8]. Later, multimodal using GPS data were presented. Ji and Zhang used high-resolution bus probe data to estimate link travel times and revealed bimodal TTD at the segment level.

The associate editor coordinating the review of this manuscript and approving it for publication was Dalin Zhang.

Then, to evaluate such bimodal TTD, they took the hierarchical Bayesian mixture model into consideration [5]. Ma *et al.* modeled the link-level bus travel time reliability of different types of roads and identified the causes of service unreliability [9]. They further used the Gaussian Mixture Model (GMM) to fit the link-level TTD and proved the superiority of GMM in supporting service reliability analysis [10]. Using the sources of floating car data, Yun and Qin developed a framework for determining the minimum sample size of floating cars to estimate TTD in situations when link travel times may follow multistate distributions [11]. Hans and Chiabaut proposed an aggregated diagram that provided a direct assessment of vehicle travel times with respect to their departure and traffic flow [12]. Another aspect is focusing on delay distribution at signalized intersections. Zheng and Van investigated urban link-level TTD and gained delay distribution to quantify the delay uncertainty [13].

As for path-level TTD analysis, one common approach is using statistical models to fit real travel time observations. Increasing researchers tend to use multimodal or mixture models to characterize path-level TTD [14]. However, in most studies, the assumption of the independence of individual link-level TTD is made [15], which appears not appropriate any more. Geroliminis and Skabardonis began the path-level travel time variance calculation under the assumption of linear correlations between consecutive link travel times [16]. To extend the work, they used Markov chains to estimate path-level TTD through the integration of correlated link TTDs. They got the sequence of pair-wise Markov matrixes which was calculated by assuming the transitions between different link pairs are conditionally independent [17]. Chen *et al.* introduced a copula-based approach to model arterial TTD by considering spatial link correlations [18]. Considering the correlation of both time and space, Ma *et al.* further developed a heuristic clustering method based on the link TTD and traffic conditions to obtain trip TTDs [19]. Recently, Zhang *et al.* developed a deep learning method to estimate trip TTDs in an urban road network. And by modeling the joint distribution of travel times of link pairs with the consideration of spatiotemporal correlations, a DCWD algorithm was proposed which could successfully characterize the traffic state transitions. The DCWD algorithm improved the trip TTD model accuracy [20]. In addition to the TTD models, other traffic variables such as travel demand in the simulation of the transportation system are also proved to be dependable. Ng *et al.* proposed a correlation-based approach that was able to represent a wider range of dependence structures than the multivariate normal distribution [21]. Zeng *et al.* elaborated a reliable pathfinding method considering the correlation between links and verified it in a real-world network in Tokyo, Japan. The results proved that it can characterize the link and path TTD validly [22].

In summary, many efforts have been made in characterizing link and path TTD. However, less research has paid attention to the interaction between turning movements within the link, which is nonnegligible. The most common phenomenon

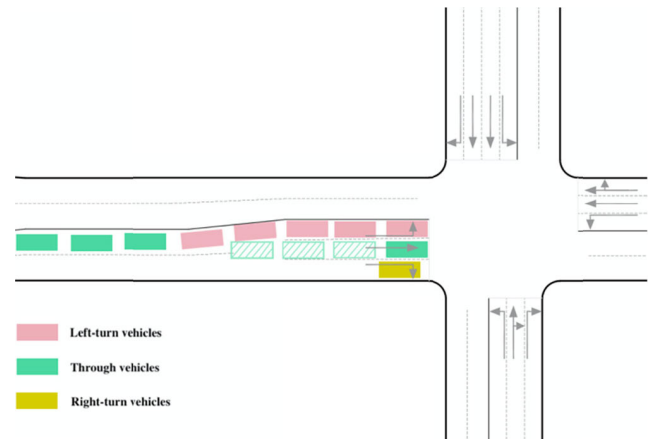


FIGURE 1. Typical Spillover of the left-turn vehicles in channelization section.

is Channelization Section Spillover (CSS), as illustrated in Fig. 1. Urban arterials customarily incorporate channelization design, where three turning movements are mixed in the upstream section and separated in the channelization section. It is evident that three turning movements have a strong nonlinear dependence. As illustrated in Fig. 1, during peak hours, the left-turn vehicles overflow the channelization section and hinder the path of through vehicles. It is necessary to take this interplay into account in travel time modeling.

Note that, vehicles' turning movements data could be obtained by roadside detectors, it is capable of studying the correlation among turning movements within the link. Even so, the dependence structure has not been explicitly addressed so far. To overcome it, we introduce the copula model to model link-level and path-level TTD. Compared to previous path-level TTD study, one distinct improvement of our work is firstly considering both the CSS within the link and the correlation among links. Meanwhile, as a feature of the copula model, the dependence structure is unaffected by the types of marginal distributions, which helps us get greater flexibility in correlating individual movement TTDs. It is proved that these efforts help provide more accurate model of link-level and path-level TTD estimation on urban arterials. The inputs to the proposed copula model are travel times of different turning movements, and the output is TTD over the study periods.

This paper is organized as follow: First, the concept and comparison of various types of copulas are present. Then, the method of modeling TTD is elaborated, including estimating the marginal distribution of travel times and identifying the best-fitting copula. In the following section four, we conduct two case studies at an arterial road in Hangzhou, China, to illustrate the application of the proposed method. In Case one, we statistically examine the correlations between left-turn and through travel time and then estimate link-level TTD by copula models. In Case two, we model the path-level TTD by taking into account the CSS in the downstream and the correlation among upstream and downstream. Next, the results are compared with empirical

bimodal distributions and previous research that has not considered CSS [18]. In the last section, we draw conclusions and provide recommendations for future work.

II. BASIC PROPERTIES OF COPULA

To incorporate the impact of CSS in estimating TTD, we introduced the copula approach, which is widely used in modeling multivariate distributions among multiple random variables with preprocessed marginal distribution [23]. The copula is a multivariate joint distribution function, which provides a flexible way of describing the correlations among several random variables in isolation from their marginal distributions. Besides, different types of marginal distributions are allowed. In specific, by selecting the appropriate marginal distribution from the parametric families of univariate distributions and picking the best-fit copula, we can accurately characterize TTD in a copula-based model with the CSS and the correlations of links considered. Variables in the copula-based model stand for the travel times of different turning movements, meanwhile, parameters in the model can reflect the correlations of links and influence of CSS. The fundamental theorem on copula is given below.

A. CONCEPT OF COPULA

Sklar firstly proposed the term copula, representing to connect or to join [23]. The primary purpose of copulas is to describe the integration of several random variables. In the last decades, copulas have been applied widely in vast fields, such as quantitative finance, economics, hydrology, etc. Recently, the function was introduced into the traffic field. Bhat and Eluru initially proposed a copula-based approach in modeling the built environment effects on travel behavior [24]. Wan and Kornhauser modeled the dependence structure of a sequence of link travel time distribution by a lagged Gaussian copula to get an approximation of path travel time distribution [25]. Lately, Chen applied a copula-based approach to characterize the travel time reliability of urban arterial [26].

The copula is a multivariate joint distribution function where each variable has a uniform marginal probability distribution. Sklar's theorem states that any n -variate joint distribution $F(x_1, x_2, \dots, x_n)$ can be written in terms of univariate marginal distribution functions $F_1(x_1), F_2(x_2), \dots, F_n(x_n)$ and a certain copula function C . The relationship between the variables are described as

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)).$$

Explicitly, Sklar's theorem can be expressed as the corresponding joint probability density function (PDF) $f(\cdot)$ as:

$$f(x_1, x_2, \dots, x_n) = c(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i) \quad (1)$$

where $F_i(x_i)$ is the cumulative density function (CDF) of marginal distribution; $f_i(x_i)$ is the marginal distribution function; $c(\cdot)$ is the copula density function, which can be

represented as:

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

where $C(\cdot)$ is the copula function, if marginal distributions are all continuous, $C(\cdot)$ is unique.

B. FAMILY OF COPULAS

1) ELLIPTICAL COPULA

a: GAUSSIAN COPULA

Gaussian copula is the most frequently applied copula function, which is an elliptical normal distribution copula with Pearson's correlation parameter [26]. The 2-dimensional Gaussian copula can be represented as:

$$C_\theta(u_1, u_2) = \Phi_\theta(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \quad (3)$$

where $u_1 = \Phi(x_1)$ is the marginal distribution of X_1 , $u_2 = \Phi(x_2)$ is the marginal distribution of X_2 ; u_1 and u_2 can be any arbitrary marginal CDF, either parametric or non-parametric, which is the most distinct difference from other joint normal CDF; and $u_1, u_2 \in [0, 1]$; Φ^{-1} denotes the inverse CDF of a standard normal; $\Phi_\theta(\cdot, \cdot)$ is the standard bivariate CDF with a parameter number θ ($\theta \in [-1, 1]$).

The copula density can be denoted by:

$$c_\theta(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C_\theta(u_1, u_2) = \frac{\exp\left(\frac{2\theta\Phi^{-1}(u_1)\Phi^{-1}(u_2) - \theta^2(\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)^2)}{2(1-\theta^2)}\right)}{\sqrt{1-\theta^2}} \quad (4)$$

In terms of its copula density, the bivariate Gaussian copula can be expressed as:

$$C_\theta(u_1, u_2) = \int_0^{u_1} \int_0^{u_2} c_\theta(s, t) ds dt \quad (5)$$

b: T COPULA

To characterize fat tails of variables, the copula of the bivariate t-distribution is universally used in the financial field, called t-copula. The joint CDF of t-copula is:

$$C_{v,\theta}^t(u_1, u_2) = t_{v,\theta}\left[t_v^{-1}(u_1), t_v^{-1}(u_2)\right] \quad (6)$$

The density function of t-copula is therefore as:

$$c_{v,\theta}^t(u_1, u_2) = \frac{f_{v,\theta}\left[t_v^{-1}(u_1), t_v^{-1}(u_2)\right]}{f_v\left[t_v^{-1}(u_1), t_v^{-1}(u_2)\right]} \quad (7)$$

where $u_1, u_2 \in [0, 1]$; t_v^{-1} denotes the inverse CDF of the standard t-distribution with the degree of freedom v ; $t_{v,\theta}$ is the bivariate t-distribution CDF with correlation parameter θ and degree of freedom v ; correlation parameter θ and degree of freedom v are usually estimated based on actual data by using the maximum likelihood method; f_v is the standard

t-distribution PDF; $f_{v,\theta}$ is the bivariate t-distribution PDF with correlation parameter θ .

Note: When $v \rightarrow \infty$, t-copula turns into Gaussian copula.

2) ARCHIMEDEAN COPULA

Archimedean family of copulas is also widely used, and is constructed in a completely different way from elliptical copula by using a generator function φ . And the Archimedean copula is uniquely determined by the generator function φ . The universal form of the bivariate Archimedean copula is described as following [27]:

$$C(u_1, u_2) = \varphi^{-1}[\varphi(u_1) + \varphi(u_2)]. \quad (8)$$

where $u_1, u_2 \in [0, 1]$; the function φ is a generator of the Archimedean copula, which is a continuous and strictly decreasing function. For any $t \in [0, 1]$, $\varphi(0) = \infty$, $\varphi(1) = 0$, $\varphi'(t) < 0$, $\varphi''(t) > 0$. The function φ^{-1} is the pseudo-inverse of φ , also continuous and strictly decreasing.

a: CLAYTON COPULA

The generator function of Clayton copula is $\varphi(t) = t^{-\theta} - 1$, with $\theta \in ((-1, 0) \cup (0, \infty))$. Clayton copula is represented as:

$$C_{Cl}^\theta(u_1, u_2) = \max\{(u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}, 0\} \quad (9)$$

The density function of Clayton copula is calculated as:

$$c_{Cl}^\theta = (1+\theta)(u_1 u_2)^{-\theta-1} (u_1^{-\theta} + u_2^{-\theta} - 1)^{-2-\frac{1}{\theta}} \quad (10)$$

where $u_1, u_2 \in [0, 1]$; when parameter $\theta \rightarrow \infty$, variables u_1, u_2 tend to be independent.

b: GUMBEL COPULA

The generator function of Gumbel copula is $\varphi(t) = (-\ln t)^\theta$, with $\theta \in [1, \infty]$. The function of Gumbel copula is:

$$C_{Gu}^\theta(u_1, u_2) = \exp\left\{-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{\frac{1}{\theta}}\right\} \quad (11)$$

The density function of Gumbel copula is denoted as:

$$c_{Gu}^\theta = \frac{C_{Gu}^\theta(u_1, u_2)(\ln u_1 \ln u_2)^{\theta-1}}{u_1 u_2 [(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{2-\frac{2}{\theta}}} \times \left\{1 + (\theta - 1)[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{\frac{1}{\theta}}\right\} \quad (12)$$

where $u_1, u_2 \in [0, 1]$. When parameter $\theta \rightarrow \infty$, variables u_1, u_2 tend to fully dependent.

c: FRANK COPULA

The generator function of Frank copula is $\varphi(t) = -\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$, with $\theta \neq 0$. Frank copula is defined as:

$$C_\theta(u_1, u_2) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right) \quad (13)$$

TABLE 1. Characteristics of different copula functions.

Family	Name	Dependence Structure	Dependence Parameter θ Domain	Strength	Weakness
Elliptical Copula	Gaussian	Positive or Negative	$[-1, 1]$	Axisymmetry	Tail Dependence, Leptokurtosis, And Heavy Tails
	T	Positive or Negative	$[-1, 1]$	Tail Dependence	Only Suit Symmetrical Variables
Archimedean Copula	Clayton	Only Positive	$(0, \infty)$	Lower Tail Dependence	Upper Tail Dependence
	Gumbel	Only Positive	$[1, \infty)$	Upper Tail Dependence	Lower Tail Dependence
	Frank	Positive or Negative; Weak	$(-\infty, 0) \cup (0, \infty)$	Symmetrical Dependence	Tail Dependence

The density function of Frank copula is calculated by [28]:

$$c_\theta(u_1, u_2) = \left(\frac{\theta}{1 - e^{-\theta}}\right) Li_{-1}\left[h_\theta^F(u_1, u_2)\right] \frac{\exp(-\theta(u_1 + u_2))}{h_\theta^F(u_1, u_2)} \quad (14)$$

where $u_1, u_2 \in [0, 1]$, $h_\theta^F(u_1, u_2) = (1 - e^{-\theta})^{-1} (1 - \exp(-\theta u_1) (1 - \exp(-\theta u_2)))$.

To choose the best copula function, we compare the characteristics of different copulas, as shown in Table 1.

Considering the relationship between travel time of turning movements within the link and combining goodness-of-fit parameters, the best-fitted copula can be selected. Due to space limitations, only the binary form of copulas is illustrated here, in which Gaussian and t copula can be deduced to multivariate copulas.

III. MODELING LINK AND PATH TTD

Taking CSS into consideration, the PDF of travel time is not merely a sum of turning movements travel time, but the joint PDF of them which can be deduced based on previously proposed copula method.

A. ESTIMATION OF THE MARGINAL DISTRIBUTION

Due to the heterogeneous traffic condition in the urban network, travel time may follow the multimodal distribution. Therefore, the Gaussian Mixture Model (GMM) is proposed to fit turning movements' TTD as a popular parametric estimator.

The finite GMM with k components can be described as:

$$p(y | u_1, u_2, \dots, u_k; s_1, s_2, \dots, s_k; \pi_1, \pi_2, \dots, \pi_k) = \sum_{j=1}^k \pi_j N(u_j, s_j^{-1}) \quad (15)$$

where u_j are the means of component j ; s_j are the inverse variances of component j ; π_j are the mixing proportion; N is a Gaussian distribution with specified mean and variance.

Kernel smoothing estimator is as well applied to describe TTD of turning movements. The density at x using kernel smoothing estimator is provided by:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (16)$$

where K is the kernel function; h is the kernel smoothing estimator's bandwidth; X_i is the observed data. A trade-off exists for the selection of h considering the bias of the estimation and its variance. The bandwidth h is chosen according to the optimality rules [29].

To select the marginal distribution, we assume that turning movements and upstream travel time conforms to a specific distribution. Two statistic tests, Kolmogorov-Smirnov (K-S) test and Cramér-von Mises (CvM) test, and Hellinger Distance (HD) are applied to access the goodness-of-fit of GMM and kernel smoothing estimator.

Here the K-S test is employed to access whether the converted distribution conforms to the uniform distribution. K-S test provides a p_{ks} value. A larger p_{ks} means a better performance in the probability integral transform. At the same time, the CvM is used to measure the similarity between the estimated CDF and the empirical CDF by providing a p_{CvM} value. A larger p_{CvM} means a better estimation of empirical CDF. To confirm the fitting performance of the non-parametric approach, the Hellinger Distance (HD) is applied to evaluate the similarity between the estimated distribution and the empirical one [30]. And a smaller HD implies a close distribution.

B. MEASURING VARIABLE DEPENDENCE

Plenty of approaches can be used to measure the dependence between variables. For two-dimension variables, a general way is Pearson's linear correlation:

$$\rho_P(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\sigma^2(X) \sigma^2(Y)}} \quad (17)$$

where $\text{cov}(X, Y)$ is the covariance of variables X and Y ; the standard deviation of X and Y are $\sigma^2(X)$ and $\sigma^2(Y)$, respectively.

Another commonly used dependence measurement is Kendall's tau, which is defined as the probability of concordance minus the probability of discordance, calculated by:

$$\tau(X, Y) = P\left[(X - \tilde{X})(Y - \tilde{Y}) > 0\right] - P\left[(X - \tilde{X})(Y - \tilde{Y}) < 0\right] \quad (18)$$

where (\tilde{X}, \tilde{Y}) is an independent copy of (X, Y) . A positive τ value means the simultaneous movement of both variable ranks, whereas a negative value indicates when one variable's rank went down, the other went up.

Besides the Pearson's linear correlation and Kendall's tau, another traditional correlation coefficient is the Spearman's rho, which is also rank-based like Kendall's tau.

In the comparison of these three dependence measurements, Kendall believed the confidence intervals of Spearman's rho are unreliable compared to Kendall's tau, and the absolute value of Pearson's linear correlation coefficient is generally overestimated [31]. Holding a similar view, Embrechts pointed out that compared with the rank-based correlation, the linear correlation has the deficiency that it is not invariant under nonlinear strictly increasing transformations. And as for multivariate distributions that possess a simple closed-form copula, the moment-based correlations (Pearson's correlation coefficient) may be difficult to calculate and the determination of rank-based correlation (Kendall's tau and Spearman's rho) may be easier [30]. In the following section, all measurements are applied to characterize the dependence of turning movements, but Kendall's tau is used as the main indicator in the estimation of parameters in copula.

C. ESTIMATION OF θ IN COPULA

After estimating the marginal distributions, the observed data is converted onto the copula scale by the probability integral transform. The next step is to estimate θ in copula. Spearman's rho and Kendall's tau can be used to estimating θ . Here, we take Gaussian copula and Kendall's tau as an example.

If the dependence structure of a random pair (X, Y) was appropriately modeled by the Gaussian copula, the estimated parameter θ can be expressed as $\tau = (2/\pi) \sin^{-1}(\theta)$, where $\tau \in [-1, 1]$. According to the equation, once the dependence estimators are known, the correlation parameter of copula θ can be estimated.

D. IDENTIFICATION OF THE BEST-FITTING COPULA

CvM test, K-S test, Log-likelihood, RMSE and Akaike Information Criterion (AIC) values are considered as the goodness-of-fit measures for comparing alternative copulas. The CvM statistic and K-S test are based on the empirical process comparing the empirical copula with a parametric estimate of the copula derived under the null hypothesis. CvM test has been proved to be more potent than the K-S test to a little deviation from the hypothesis distribution [32]. The models with higher p_{ks} and p_{CvM} values are better in terms of not being rejected. Setting the confidence interval in 0.05, larger Log-likelihood with lower AIC and RMSE values indicate a better-fitting copula.

E. MODELING LINK TTD

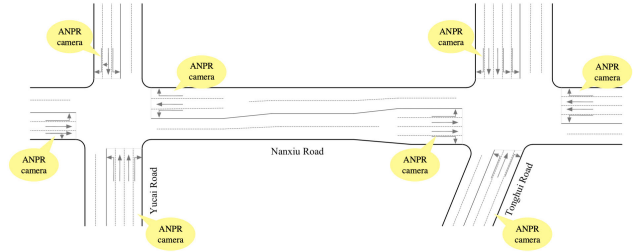
After estimating the copula model, the link travel time distribution can be expressed as:

$$F(\lambda t_x^0) = p(T_u \leq t_x^0 \cup T_v \leq t_x^0) \quad (19)$$

where $F(\lambda t_x^0)$ represents the probability of link travel time below λt_x^0 ; $p(\cdot)$ is the density of the estimated copula where $T_u \leq t_x^0$ union $T_v \leq t_x^0$.



(a) An Overview Map and an Image of the Channelization Section



(b) The Layout of Link 1 in Nanxiu Road

FIGURE 2. Study area of Nanxiu road selected in Hangzhou, China.

F. MODELING PATH TTD

The path travel time distribution can be expressed as:

$$F_z(\lambda t_x^0) = p\left(\left(T_{u1} + T_v \leq \lambda t_x^0\right) \cup \left(T_{u2} + T_v \leq \lambda t_x^0\right) \cup \left(T_{u3} + T_v \leq \lambda t_x^0\right) \dots \cup \left(T_{un} + T_v \leq \lambda t_x^0\right)\right) \quad (20)$$

where $F_z(\lambda t_x^0)$ represents the probability of path travel time below λt_x^0 ; T_{u1} is the travel time of left-turn vehicles; T_{u2} is the travel time of through vehicles; T_{u3} is the travel time of right-turn vehicles; T_v is the travel time of upstream vehicles; $p(T_{u1} + T_v \leq \lambda t_x^0)$ is the cumulative density of the estimated copula where $T_{u1} + T_v \leq \lambda t_x^0$.

IV. CASE STUDY

The proposed methodology was evaluated on an arterial in Hangzhou, China. Following presented is the study area description and data processing. Based on the investigation, the copula was chosen, and the related model was constructed for TTD estimation. Then, the link-level TTD model was compared with the empirical distribution fitting approach and the result reflected that the copula model performed better than the GMM. Next, the path-level TTD model considering CSS was compared with the copula model ignoring CSS, and the former performed better.

A. STUDY AREA DESCRIPTION

The study site is located in Nanxiu Road, Hangzhou, Zhejiang Province. As shown in Fig. 2(a), it is a 1,270 m long urban arterial consisting of two links. Within link 1, there are left-turn, through and right-turn lane one each in the channelization section, while two lanes at upstream section with total length of 540 m. In link 2, there are one left-turn lane, two through lanes, one right-turn lane in the channelization section, and two lanes of 710 m at upstream section. There are some entries and exits for primary school, college and residential districts along the links.

As a main commuter arterial of the city, there is extensive traffic demand through the day in Nanxiu Road. The signal system's cycle time length in the intersection of link 1 ranges from 60 s to over 120 s; link 2 ranges from 100 s to over 170 s. The release of right-turn vehicles in link 1 is not limited.

The travel time data were collected by Automatic Number Plate Recognition (ANPR) systems. The cameras are located at the entrance of each roadway in the intersections, as illustrated in Fig. 2(b). Based on image processing, ANPR cameras can catch the license plates of passing vehicles, then records the vehicles' license plates, time stamps, turning movements and vehicle types. According to the data collected by upstream and downstream ANPR cameras, actual travel times through the links and intersections can be obtained by matching identified vehicle license plates and comparing their time stamps.

Taking the features of urban roads into consideration, we can observe many outliers of travel time samples, which make data noisy. A major part of that noise was attributed to vehicles that, after passing the upstream camera, stop before continuing their trip (loading-unloading, parking, shopping, etc.). In such occasion, the vehicles were detected by the downstream ANPR cameras significantly later, and this phenomenon caused travel times to be notably higher than usual. In the same context, buses stopping as part of their regular routes produce another source of the noise. According to the field investigation, we proposed a simple filtering method for the data, which confirmed that during peak hours, most of the delayed vehicles had maximum queuing time less than three red phases. Thus, 400 s, approximately the time of three cycle lengths and 90th percentile travel time of all the vehicles, was set as the upper bound value for outlier elimination in link 1. Ten seconds was set as the lower bound value to avoid repeated or error detection. After a robust filtering method, the data of weekdays in Dec 2018, altogether the ground truth travel times of 4954 vehicles, were obtained and would be utilized for the copula modeling. The collected dataset is divided into the training set (70%) and the testing set (30%).

To demonstrate the need of considering CSS and the superiority of copula modeling, two trials was conducted within the study area. One tested the copula modeling performance of link-level TTD by considering the CSS within link 1. Since the release of right-turn vehicles are not limited, only left-turn and through vehicles were taken into consideration in case one. The other case modeled the path-level TTD, considering both the CSS within link 2 and correlation between link 1 and link 2.

B. APPLICATION OF THE COPULA MODEL IN LINK-LEVEL TTD

1) DEPENDENCE BETWEEN TURNING MOVEMENTS

Considering that time of day, road congestions and occurrence of any event may change the travel time significantly; it is possible that the dependence structure among turning movements travel time varies relying on

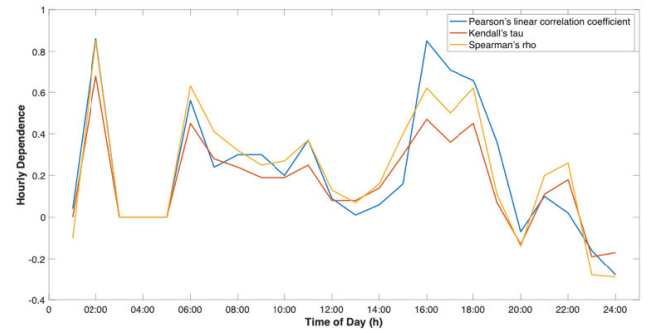
TABLE 2. Hourly Dependence between left-turning and through travel times for a 24h period.

Time Period	Count (vehicle number)	Correlation Coefficient		
		ρ_P	τ	ρ_S
00:00-01:00	5	0.04	0.00	-0.10
01:00-02:00	7	0.86	0.68	0.85
02:00-03:00	2	0.00	0.00	0.00
03:00-04:00	1	0.00	0.00	0.00
04:00-05:00	2	0.00	0.00	0.00
05:00-06:00	5	0.56	0.45	0.63
06:00-07:00	64	0.24	0.28	0.41
07:00-08:00	112	0.30	0.24	0.32
08:00-09:00	108	0.30	0.19	0.25
09:00-10:00	102	0.20	0.19	0.27
10:00-11:00	117	0.37	0.25	0.37
11:00-12:00	102	0.09	0.08	0.13
12:00-13:00	118	0.01	0.08	0.07
13:00-14:00	107	0.06	0.14	0.16
14:00-15:00	106	0.16	0.30	0.40
15:00-16:00	115	0.85	0.47	0.62
16:00-17:00	176	0.71	0.36	0.50
17:00-18:00	157	0.66	0.45	0.62
18:00-19:00	81	0.36	0.07	0.11
19:00-20:00	43	-0.07	-0.13	-0.14
20:00-21:00	55	0.10	0.11	0.20
21:00-22:00	72	0.02	0.18	0.26
22:00-23:00	29	-0.16	-0.19	-0.28
23:00-24:00	13	-0.28	-0.17	-0.29

the traffic conditions. Thus, we firstly evaluated the hourly dependence between left-turning and through travel times in link 1, which was based on ANPR data at a vehicle-to-vehicle level on Dec. 4, 2018. For every hour, Pearson's linear correlation coefficient ρ_P , Kendall's tau τ , and Spearman's rho ρ_S were calculated. The results are presented in Table 2.

As shown in Table 2 and Fig. 3, the tendency of three correlation coefficients is similar. The values of Spearman's rho ρ_S are mostly higher than those of Kendall's tau τ , which verifies that Pearson's linear correlation coefficient ρ_P in absolute value is generally overestimated, as mentioned in Section III. Fig. 3 also illustrates that the dependence structure varies significantly during different times of the day. Generally, a strong correlation exists in peak hours, specifically in afternoon peak hours.

The line 'count' in Table 2 shows the vehicle counts through the left-turn lane in Nanxiu Road, collected at 60-min intervals. Notably, the number of samples during midnight or early morning is not enough, leading correlation coefficients to be abnormally high or equal to zero.

**FIGURE 3.** Hourly dependence between left-turning and through travel times for a 24h period.

Furthermore, the hourly dependence curve depicted in Fig. 3 shows that the afternoon peak starts from 16:00 to 20:00. Afternoon peaks are not clear in time compared with morning peaks. The possible reason for this phenomenon is that commercial activities prolong peak periods from early afternoon to late evening.

In essence, the level of correlation between travel time in turning movements may get influenced by numerous factors, such as traffic volume and signal settings. To accurately test the goodness of the copula model, the same period in days was chosen, which reflected a stable dependence structure. Particularly, in morning peak hours (07:00-09:00), correlation coefficients are a bit higher than the off-peak hour, which are suitable to be characterized. In the research period, the signal cycle is 120 s.

Given the numbers of samples and features of dependence structure, the data of weekdays' morning peak period were chosen to construct the copula model. The dependence estimation results are: Kendall's tau τ is 0.21, Pearson's rho ρ_P is 0.35, Spearman's rho ρ_S is 0.28. Adequately 1,109 samples altogether. In the following sections, the parameters of alternative copulas can be determined by dependence parameters.

2) INVESTIGATION OF MARGINAL TTD

The estimation of each turning movement's marginal distribution served as a foundation for the copula approach. Based on kernel smoothing estimation, assisted with the interpretation of density curves depicted in Fig. 5, the travel time of turning movements appears to follow a multimodal distribution. In order to choose a rational number of components in GMM, AIC and BIC values for different components were calculated. As shown in Fig. 4, GMM with 3 components has the minimum AIC value and a relatively small BIC value. Therefore, we chose GMM3 to fit both left-turn and through travel time. Red curves in Fig. 5 are the fitting results.

3) OPTIMAL COPULA MODEL SELECTION

As aforementioned, Kendall's tau τ for two turning movement's travel times is 0.21, indicating that the dependence between two turning movements is both positive and weak. According to the characteristics of copulas in Table 1, all functions can be applied in this case. The parameters θ of

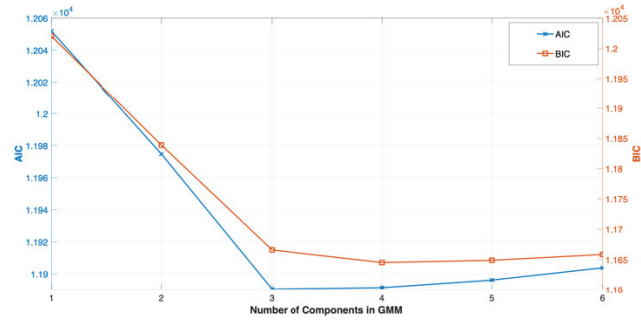


FIGURE 4. AIC and BIC values for a different number of GMM components.

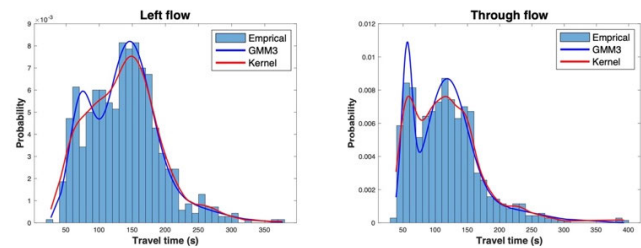


FIGURE 5. Fitting curves of GMM3 and kernel smooth for left-turn flow and through flow.

TABLE 3. The parameters θ of different copulas.

Time Periods	Gaussian	T	Clayton	Frank	Gumbel
θ peak hours	0.21	0.33	0.53	1.94	1.27
off-peak hours	0.22	0.24	0.25	1.43	1.16

different copulas was calculated based on Kendall’s tau τ , as shown in Table 3. The statistics in Table 3 can reflect the interactions between left-turning and through movements. By comparing the θ value in Table 3, it can be found that the parameters of the copula model can mathematically capture correlations of travel times within the link. During peak hours, θ values are generally higher, especially in the Frank copula and Clayton copula. In the meantime, the correlation between left-turning and through travel times are simultaneously high as illustrated in Table 2. Particularly, the degree of freedom ν in T copula is estimated using the maximum likelihood method, in this case $\nu = 2.47$ in peak hours, $\nu = 8.96$ in off-peak hours.

Afterward, to determine the best-fitting copula, the goodness-of-fit statistics of different copulas is calculated and presented in Table 4. The favorable copula is supposed to meet requirements: the biggest p_{ks} value, the biggest p_{cvm} value in the case of smallest CvM statistics; a larger Log-likelihood value and smaller AIC value. Overall, Frank copula appeared as the best.

After that, we drew the surface of estimated copula density of Frank copula and frequency histogram of empirical data, as presented in Fig. 6. Note the existence of extremely high values on the upper side of the frequency histogram near the node (1, 1). Accordingly, the estimated Frank copula density

TABLE 4. Goodness-of-fit tests of different copulas during peak hours.

Goodness-of-Fit Statistics	Gaussian	T	Clayton	Frank	Gumbel
Log Likelihood	-2146.5	-2097.4	-2169.8	-2145.3	-2112.9
AIC	4294.9	4198.8	4341.6	4292.6	4227.7
RMSE	0.0156	0.0137	0.0214	0.0137	0.0144
p_{ks}	0.6543	0.2795	0.0675	0.9839	0.7651
p_{cvm}	0.7516	0.5951	0.1100	0.9899	0.8997
cvm	0.0700	0.0980	0.3324	0.0247	0.0461

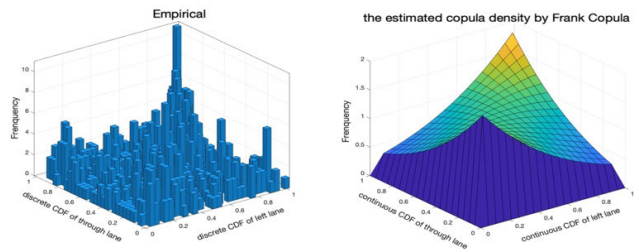


FIGURE 6. The empirical frequency histogram and the estimated copula density of the optimal frank copula in 07:00-09:00 a.m.



FIGURE 7. An image of CSS in link 1 during peak hours.

confirms the presence of upper tail dependence, implying that the slow-speed flow occurs simultaneously and frequently on left-turn and through lanes during peak hours, which needs additional attention, implying the possibility of CSS. In the field investigation, we did find the spillover of the through vehicles in the channelization section during peak hours, which hindered the path of left-turn vehicles, as shown in the image in Fig. 7.

As mentioned in Table 1, Frank copula has symmetrical characteristics, indicating that after the probability integral transform, the dependence structure of travel times in left-turn and through lanes tend to be symmetrical. The left and right tail dependences near (0, 1) and (1, 0) in frequency histogram are remarkable, but are not accurately captured in the estimated copula density. This phenomenon indicates that Frank copula is not ideal for identifying the right tail and left

TABLE 5. The goodness-of-fit tests of frank copula, GMM2, GMM3 and unimodal distributions.

Goodness-of-Fit Statistics	Frank Copula	GMM2	GMM3	Normal	Lognormal	Gamma	Weibull
p_{ks}	0.84	0.23	0.7	0.36	0.59	0.46	0.59
p_{CvM}	0.62	0.38	0.6	0.2	0.52	0.64	0.45
Cvm	0.087	0.15	0.1	0.24	0.11	0.089	0.13
HD	0.26	0.43	0.27	0.46	0.38	0.36	0.36

tail dependences. More types of copulas are supposed to be tested in future work.

4) ESTIMATION OF LINK TTD

Unimodal distributions, e.g., Normal, Lognormal, Gamma and Weibull distributions, were first used to characterize TTD. Compared with unimodal distributions, GMM has also become of interest. In order to figure out the performance of these distributions, the goodness-of-fit values were calculated, and the results were listed in Table 5. The numerical values show that GMM3's fitting results are generally better than unimodal distributions and the GMM2. Thus GMM3 was chosen as a representative of multimodal distributions to compare with the copula model.

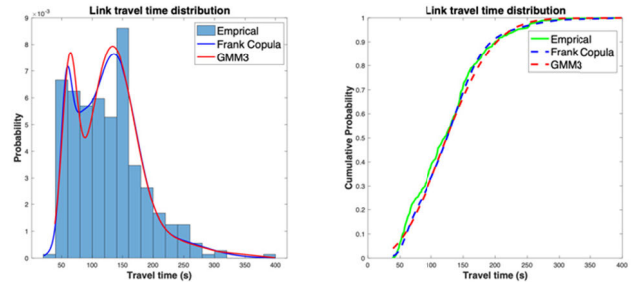
Fig. 8 represents the fitting results. The goodness-of-fit measurements of Frank copula and GMM3 are also presented in Table 5. The statistics of higher p_{ks} and p_{CvM} and lower Cvm and HD indicate that Frank copula performs superior on the GMM3. Besides, the comparison of GMM2 and GMM3 above proves that the copula model is more flexible than GMM, due to the no need of any assumptions for input variables. And the copula model avoids the instability of the GMM which is caused by random initialization of the parameters.

C. APPLICATION OF THE COPULA MODEL IN PATH-LEVEL TTD

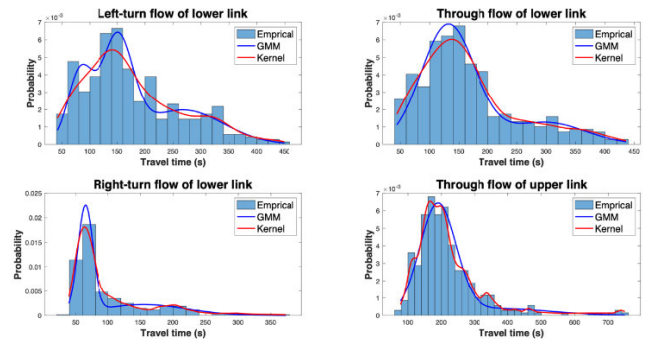
The CSS appearing within the link also has a great influence on the path-level TTD model. By considering both the CSS and the correlation between links, we constructed a path-level TTD model and compared with the existing research [26], which only considered the correlation of link travel time. Based on the above, the study period was set at the early peak and afternoon off-peak in weekdays.

1) DEPENDENCE BETWEEN LEFT-TURN, THROUGH, RIGHT-TURN MOVEMENTS, AND LINKS

For peak hours and off-peak hours, Pearson's linear correlation coefficient ρ_P , Kendall's tau τ , and Spearman's rho ρ_S were calculated. The results are presented in Table 6. The travel time correlation within the link is relatively high, particularly in peak hour, whereas the correlation between the upstream and downstream is weak. The strong correlation within the link in peak hours indicates the possibility of CSS in downstream. The weak correlation between upstream and

**FIGURE 8.** Link TTD estimated by frank copula and GMM3.**TABLE 6.** Dependence between turning movements in peak and off-peak hours.

Time Period	Combination Options	Correlation Coefficient		
		ρ_P	τ	ρ_S
07:00-09:00	Left & Through	0.73	0.49	0.63
	Through & Right	0.68	0.41	0.57
	Left & Right	0.65	0.44	0.61
	Upstream & Downstream	0.02	0.04	0.05
	Left-Turn & Through	-0.04	-0.06	-0.09
10:00-15:00	Through & Right-Turn	0.07	0.20	0.12
	Left-Turn & Right-Turn	0.07	0.05	0.07
	Upstream & Downstream	-0.09	-0.04	-0.06

**FIGURE 9.** Fitting curves of GMM and kernel smooth for each turning movement's flow and upstream flow.

downstream means that the spillover only existed in the channelization section, vehicles didn't overflow the link or affect the traffic in the upstream.

2) INVESTIGATION OF MARGINAL TTD

Kernel smoothing estimation and GMM were used for marginal distribution estimation, as presented in Fig. 9. Since the right-turn vehicles are not restricted by the signal, the right-turn travel time presents a unimodal distribution. Additionally, it can be observed that there are limited visual discrepancies between the GMM and kernel estimator, which means statistics results are required to decide which way to choose.

Larger p_{ks} , p_{CvM} values, and smaller HD values always indicate a better overall fit. As reported in Table 7, although

TABLE 7. Goodness-of-fit tests of GMM and kernel.

	Goodness-of-Fit Statistics	GMM	Kernel
Left	p_{ks}	0.4563	0.7687
	p_{CvM}	0.9888	0.9426
	HD	0.2486	0.2402
Through	p_{ks}	0.8141	0.9139
	p_{CvM}	0.9899	0.9899
	HD	0.2455	0.2365
Right	p_{ks}	0.5746	0.4843
	p_{CvM}	0.4977	0.9899
	HD	0.2486	0.2310
Upstream	p_{ks}	0.8019	0.8842
	p_{CvM}	0.8124	0.8124
	HD	0.2496	0.2301

TABLE 8. Goodness-of-fit tests of copulas.

Goodness-of-Fit Statistics	Gaussian	T
Log Likelihood	-1104.09	-1092.68
AIC	2210.18	2189.36
BIC	2214.03	2197.05
RMSE	0.04	0.04

GMM has an advantage in characterizing TTD similarly, the kernel performs better on probability integral transform, which is indicated by larger p_{ks} values. Overall, kernel smoothing estimation provided a comparable performance with GMM in this case.

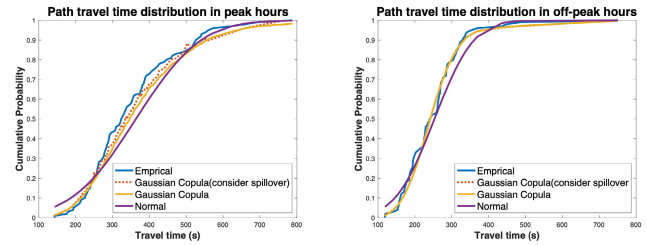
3) OPTIMAL COPULA MODEL SELECTION

For multivariable distribution, the candidate copula is Gaussian copula and t copula. The comparison of these two copula models is shown in Table 8. The goodness-of-fit statistics of Gaussian copula and t copula are close to each other. Overall, Gaussian copula appeared as the best-fitting copula function.

4) ESTIMATION OF PATH TTD

Since the integral of multivariate TTD model is difficult, the Monte-Carlo simulation were used to get the results of the CDF of path travel time. Compared with the path TTD model that only considers the correlation between links, the path TTD model also considering the influence of CSS within downstream performs better, particularly in peak hour, as shown in Fig. 10.

For a more comprehensive comparison, statistics values are presented in Table 9. The numerical value proved it's necessary to consider the mutual influence of CSS within downstream. The values of the p_{ks} and p_{CvM} are much higher

**FIGURE 10. Link TTD estimated by Gaussian copula (with and without considering CSS).****TABLE 9. The Goodness-of-fit tests of gaussian copula (with and without consider CSS).**

Time Period	Goodness-of-Fit Statistics	Gaussian Copula (considering CSS)	Gaussian Copula
Peak Hours	p_{ks}	0.74	0.30
	p_{CvM}	0.65	0.28
	cvm	0.09	0.20
	HD	0.20	0.27
	p_{ks}	0.72	0.75
Off-peak Hours	p_{CvM}	0.79	0.79
	cvm	0.06	0.06
	HD	0.21	0.21

than the model ignoring the CSS. Due to the existence of the CSS during the peak periods, the correlation of turning movements within the link is enhanced, which makes the superiority of the copula model more prominent. The copula-based model can mathematically capture correlation and appropriately characterize the path-level TTD during peak hours. However, during off-peak hours, two models have similar behavior.

Path-level TTD is usually complicated and difficult to calibrate because of the complexity and interrupted nature of urban traffic flows, caused by signal control, bus stops, roadside parking, disturbances from pedestrians and vehicle interactions. But the copula-based function simplifies the relationship, turn it into a mathematical model. Using the parameters in the copula-based model, it is able to describe the traffic flows' correlation within the link and between links. The reasonable copula-based model is proved to be accurate and may have wide applications in practice. However, due to the limited data amount and poor correlation between links, the TTD model of the path-level is less reliable than that of the link-level. In future research, more data on various paths should be collected and modeled.

V. CONCLUSIONS AND DISCUSSIONS

This paper introduced a copula-based method to estimate link-level and path-level TTD by taking both the CSS and the correlation of links into consideration. The method, without any assumptions for input variables, was applied in modeling TTD on an arterial road in Hangzhou, China. We compared

link-level TTD estimated by copula model to the Gaussian mixture model. Moreover, path-level TTD model considering CSS was compared to the previous researcher's copula model, which only noting the correlation between links [20]. All results prove the necessity of considering CSS. Also, the superiority of the copula model in the travel time modeling is illustrated. The main findings are summarized as below:

(1) Movement-level travel time represents significantly different dependence structures in different time of day. Such changeable correlation could be regarded as the overall effect of the CSS, which should not be ignored.

(2) For link-level TTD estimation, by taking CSS into account, the copula model with marginal distributions from movement-level travel time performs reasonably better than empirical Gaussian mixture model. The statistical superiority of the copula model is proved.

(3) When modeling path-level TTD, in off-peak periods, the copula model considering CSS has no apparent superiority over the copula model ignoring CSS, due to weak correlation. In peak hours, with the increasing correlation among movements and links, taking CSS into account makes the TTD model performs much more accurate. The necessity to consider CSS in path-level TTD has been proven.

The proposed TTD model can make travel time estimation more accurate. On applications, if the underlying determinants of the travel time are to be analyzed, a more precise TTD model could be applied as a research basis. What's more, the underlying determinants could be introduced as variables in the copula-based TTD model, making the model more accurate and universal. With additional factors, the improved statistic TTD fitting could support effective travel time reliability analysis and average travel time prediction. Besides, the reliable TTD model is expected to have wide applications at a more macro level, such as grasping the network-wide traffic states index and targeting specific strategies aimed at improving traffic states.

It is worth noting that in this study, we only compare the copula model with unimodal distributions and GMM when characterizing link-level TTD. The results and conclusions derived above may have some limitations. To extend the initial findings, we are supposed to test other types of distribution models using more data. What's more, in the path-level TTD model where Gaussian copula and T copula are adopted, they may have limitations. More multi-dimensional copulas would be conducted in future work. Also, to have practical value, additional inputs are to be introduced to the copula-based model including link length, lane count, signal parameters, day of the week, time period of the day and so on. With adequate factors, a general and complete spatial-temporal TTD model can be used to predict the average travel time and its reliability. Besides, as this paper only focuses on link-level and short path-level TTD model, future research will extend the TTD model to paths with more than two links, route-level, and network-level. The algorithm will be optimized to be faster for the application.

REFERENCES

- [1] G. Kim, "Travel time estimation in vehicle routing problem," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Dec. 2017, p. 5.
- [2] M. Biswas, A. R. Pinjari, and S. K. Dubey, "Travel time variability and route choice: An integrated modelling framework," in *Proc. 11th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2019, pp. 737–742.
- [3] J. Rupnik, J. Davies, B. Fortuna, A. Duke, and S. S. Clarke, "Travel time prediction on highways," in *Proc. IEEE Int. Conf. Comput. Inf. Technol.; Ubiquitous Comput. Commun.; Dependable, Autonomic Secure Comput.; Pervasive Intell. Comput.*, Oct. 2015, pp. 1435–1442.
- [4] S. Shiravi, L. Fu, and M. Muresan, "Continual retiming of traffic signals using big travel time data," in *Proc. 4th Int. Conf. Transp. Inf. Saf. (ICTIS)*, Aug. 2017, pp. 784–793.
- [5] Y. Ji and H. M. Zhang, "Travel time distributions on urban streets: Estimation with hierarchical Bayesian mixture model and application to traffic analysis with high-resolution bus probe data," presented at the 92nd Annu. Meeting Transp. Res. Board, 2013.
- [6] J. Tang, F. Liu, Y. Zou, W. Zhang, and Y. Wang, "An improved fuzzy neural network for traffic speed prediction considering periodic characteristic," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2340–2350, Sep. 2017.
- [7] N. Uno, F. Kurauchi, H. Tamura, and Y. Iida, "Using bus probe data for analysis of travel time variability," *J. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 2–15, Feb. 2009.
- [8] T. Hunter, P. Abbeel, R. Herring, and A. Bayen, "Path and travel time inference from GPS probe vehicle data," in *Proc. Neural Inf. Process. Syst. Found. (NIPS)*, Vancouver, BC, Canada, 2009, p. 8.
- [9] Z.-L. Ma, L. Ferreira, M. Mesbah, and A. T. Hojati, "Modeling bus travel time reliability with supply and demand data from automatic vehicle location and smart card systems," *Transp. Res. Rec.*, vol. 2533, no. 1, pp. 17–27, Jan. 2015, doi: [10.3141/2533-03](https://doi.org/10.3141/2533-03).
- [10] Z. Ma, L. Ferreira, M. Mesbah, and S. Zhu, "Modeling distributions of travel time variability for bus operations: Bus travel time distribution model," *J. Adv. Transp.*, vol. 50, no. 1, pp. 6–24, Jan. 2016, doi: [10.1002/atr.1314](https://doi.org/10.1002/atr.1314).
- [11] M. Yun and W. Qin, "Minimum sampling size of floating cars for urban link travel time distribution estimation," *Transp. Res. Rec.*, vol. 2673, no. 3, pp. 24–43, Mar. 2019.
- [12] E. Hans, N. Chiabaut, and L. Leclercq, "Clustering approach for assessing the travel time variability of arterials," *Transp. Res. Rec.*, vol. 2422, no. 1, pp. 42–49, Jan. 2014.
- [13] F. Zheng and H. Van Zuylen, "Uncertainty and predictability of urban link travel time: Delay distribution-based analysis," *Transp. Res. Rec.*, vol. 2192, no. 1, pp. 136–146, Jan. 2010.
- [14] E. Kazagli and H. N. Koutsopoulos, "Estimation of arterial travel time from automatic number plate recognition data," *Transp. Res. Rec.*, vol. 2391, no. 1, pp. 22–31, Jan. 2013.
- [15] Y. Iida, "Basic concepts and future directions of road network reliability analysis," *J. Adv. Transp.*, vol. 33, no. 2, pp. 125–134, 1999.
- [16] N. Geroliminis and A. Skabardonis, "Real time vehicle reidentification and performance measures on signalized arterials," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Sep. 2006, pp. 188–193.
- [17] M. Ramezani and N. Geroliminis, "On the estimation of arterial route travel time distribution with Markov chains," *Transp. Res. B, Methodol.*, vol. 46, no. 10, pp. 1576–1590, Dec. 2012.
- [18] P. Chen, W. Zeng, M. Chen, G. Yu, and Y. Wang, "Modeling arterial travel time distribution by accounting for link correlations: A copula-based approach," *J. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 28–40, Jan. 2019.
- [19] Z. Ma, H. N. Koutsopoulos, L. Ferreira, and M. Mesbah, "Estimation of trip travel time distribution using a generalized Markov chain approach," *Transp. Res. C, Emerg. Technol.*, vol. 74, pp. 1–21, Jan. 2017, doi: [10.1016/j.trc.2016.11.008](https://doi.org/10.1016/j.trc.2016.11.008).
- [20] K. Zhang, N. Jia, L. Zheng, and Z. Liu, "A novel generative adversarial network for estimation of trip travel time distribution with trajectory data," *Transp. Res. C, Emerg. Technol.*, vol. 108, pp. 223–244, Nov. 2019, doi: [10.1016/j.trc.2019.09.019](https://doi.org/10.1016/j.trc.2019.09.019).
- [21] M. Ng, K. Kockelman, and S. Waller, "Relaxing the multivariate normality assumption in the simulation of transportation system dependencies: An old technique in a new domain," *Transp. Lett.*, vol. 2, no. 2, pp. 63–74, Apr. 2010, doi: [10.3328/tl.2010.02.02.63-74](https://doi.org/10.3328/tl.2010.02.02.63-74).
- [22] W. Zeng, T. Miwa, Y. Wakita, and T. Morikawa, "Application of Lagrangian relaxation approach to α -reliable path finding in stochastic networks with correlated link travel times," *Transp. Res. C, Emerg. Technol.*, vol. 56, pp. 309–334, Jul. 2015, doi: [10.1016/j.trc.2015.04.018](https://doi.org/10.1016/j.trc.2015.04.018).

- [23] A. Sklar, "Random variables, joint distribution functions, and copulas," *Kybernetika*, vol. 9, no. 6, pp. 449–460, 1973.
- [24] C. R. Bhat and N. Eluru, "A copula-based approach to accommodate residential self-selection effects in travel behavior modeling," *Transp. Res. B, Methodol.*, vol. 43, no. 7, pp. 749–765, Aug. 2009.
- [25] K. Wan and A. L. Kornhauser, "Link-data-based approximation of path travel time distribution with Gaussian copula estimated through lasso," presented at the 89th Annu. Meeting Transp. Res. Board, 2010.
- [26] M. Chen, G. Yu, P. Chen, and Y. Wang, "A copula-based approach for estimating the travel time reliability of urban arterial," *Transp. Res. C, Emerg. Technol.*, vol. 82, pp. 1–23, Sep. 2017.
- [27] C. Genest and L.-P. Rivest, "Statistical inference procedures for bivariate archimedean copulas," *J. Amer. Stat. Assoc.*, vol. 88, no. 423, pp. 1034–1043, Sep. 1993.
- [28] M. Hofert, M. Mächler, and A. J. Mcneil, "Likelihood inference for Archimedean copulas in high dimensions under known margins," *J. Multivariate Anal.*, vol. 110, pp. 133–150, Sep. 2012.
- [29] A. W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford, U.K.: Clarendon, 2004.
- [30] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde, "Bayesian measures of model complexity and fit," *J. Roy. Stat. Soc. B (Stat. Methodol.)*, vol. 64, no. 4, pp. 583–639, Oct. 2002, doi: [10.1111/1467-9868.00353](https://doi.org/10.1111/1467-9868.00353).
- [31] M. G. Kendall, *Rank Correlation Methods*. Oxford, U.K.: Griffin, 1948.
- [32] S. Das and A. K. Maurya, "Multivariate analysis of microscopic traffic variables using copulas in staggered car-following conditions," *Transportmetrica A: Transp. Sci.*, vol. 14, no. 10, pp. 829–854, Nov. 2018.



YI YU was born in 1995. She received the B.S. degree in civil engineering from Zhejiang University, Zhejiang, China, in 2017, where she is currently pursuing the Ph.D. degree. Her research interests include traffic flow theories and traffic state identification.



MENGWEI CHEN received the Ph.D. degree in road and transportation engineering from Zhejiang University, Hangzhou, China, in 2019. She is currently a Research Assistant with the Zhejiang University of Technology. Her research interests are transportation planning and travel behavior.



HONGSHENG QI (Member, IEEE) received the Ph.D. degree in transportation engineering from Jilin University, in 2011. He is currently an Assistant Professor of traffic engineering with Zhejiang University, Hangzhou, China. His research interests are in the area of traffic control, traffic flow theories, and big data in traffic research.



DIANHAI WANG received the Ph.D. degree in transportation engineering from Beijing Jiaotong University, Beijing, China, in 1995.

He is currently a Full Professor with the Institute of Transportation Engineering, College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China. His research interests include traffic control, traffic flow theory, and traffic information. On these topics, he has published more than 200 articles in related journals, such as

Transportation Research part A/C and the *Journal of Urban Planning and Development*. He is a member of the China Higher Education Steering Committee, Traffic Engineering Subject and the China City Planning Academic Committee. He has received the National Science and Technology Progress Award, in 2004, and the Science and Technology Progress Award of Jilin province, in 2009.

...