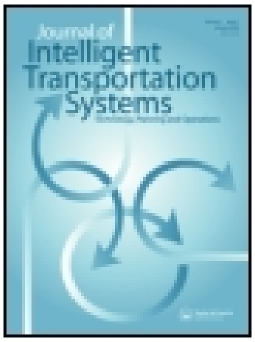

Journal of Intelligent Transportation Systems
Technology, Planning, and Operations

ISSN: 1547-2450 (Print) 1547-2442 (Online) Journal homepage: http://www.tandfonline.com/loi/gits20

Modeling arterial travel time distribution by accounting for link correlations: a copula-based approach

Peng Chen, Weiliang Zeng, Min Chen, Guizhen Yu & Yunpeng Wang

To cite this article: Peng Chen, Weiliang Zeng, Min Chen, Guizhen Yu & Yunpeng Wang (2018): Modeling arterial travel time distribution by accounting for link correlations: a copula-based approach, Journal of Intelligent Transportation Systems, DOI: 10.1080/15472450.2018.1484738

To link to this article: https://doi.org/10.1080/15472450.2018.1484738

Published online: 17 Oct 2018.

JOURNAL OF INTELLIGENT TRANSPORTATION SYSTEMS
https://doi.org/10.1080/15472450.2018.1484738

Taylor & Francis
Taylor & Francis Group

Check for updates

Modeling arterial travel time distribution by accounting for link correlations: a copula-based approach

Peng Chen[a,b], Weiliang Zeng[c], Min Chen[a], Guizhen Yu[a,b], and Yunpeng Wang[a,b]

[a]Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety Control, School of Transportation Science and Engineering, Beihang University, Beijing, China; [b]Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing, China; [c]School of Automation, Guangdong University of Technology, Guangdong, China

ABSTRACT

The estimation of urban arterial travel time distribution (TTD) is critical to help implement Intelligent Transportation Systems (ITS) and provide travelers with timely and reliable route guidance. The state-of-practice procedure for arterial TTD estimation commonly assumes that the path travel time follows a certain distribution without considering link correlations. However, this approach appears inappropriate since travel times on successive links are essentially dependent along signalized arterials. In this study, a copula-based approach is proposed to model arterial TTD by accounting for spatial link correlations. First, TTDs on consecutive links along one arterial in Hangzhou, China are investigated. Link TTDs are estimated through the nonparametric kernel smoothing method. Link correlations are analyzed in both unfavorable and favorable coordination cases. Then, Gaussian copula models are introduced to model the dependent structure between link TTDs. The parameters of Gaussian copula are obtained by Maximum-Likelihood Estimation (MLE). Next, path TTDs covering consecutive links are estimated based on the estimated copula models. The results demonstrate the advantage of the proposed copula-based approach, compared with the convolution without capturing link correlations and the empirical distribution fitting methods in both unfavorable and favorable coordination cases.

ARTICLE HISTORY
Received 10 November 2016
Revised 18 May 2018
Accepted 19 May 2018

KEYWORDS
Copula; link correlations; signal coordination; travel time distribution; urban arterial

Introduction

Travel time distribution (TTD) and travel time reliability (TTR) serve as significant measures for evaluating the operational efficiency of transportation facilities (Etienne, Nicolas, & Ludovic, 2015; Ji et al., 2018; Zhang, Wang, Chen, He, & Yu, 2017), comparing different traffic control and management strategies (Chen, Sun, & Qi, 2017a; Ramezani & Geroliminis, 2015) and delivering travelers about timely and reliable route information (Zeng, Miwa, Wakita, & Morikawa, 2015). The national SHRP2 Reliability research (SHRP2 Report S2-L02-RR-2, 2014) defined TTR as consistency of travel time over time. It has been well noted that the analysis of TTR is equally important as the commonly adopted average travel time (Bates, Polak, Jones, & Cook, 2001).

The interrupted flow on arterials makes TTD estimation more challenging than uninterrupted flow on freeway and expressway facilities (Chen, Tong, Lu, & Wang, 2018; Tang, Liu, Zou, Zhang, & Wang, 2017). Under complicated interactions of traffic regimes and signal control, arterial travel times exhibit distinct shapes of distributions under different traffic characteristics. The state-of-practice approaches for arterial TTD estimation can be generally classified into two categories based on available data sources (Rakha, El-Shawarby, Arafeh, & Dion, 2006). Based on path-level data sources, e.g., using automatic license plate recognition and automatic vehicle identification (AVI), path TTD can be directly obtained and estimated by assuming certain unimodal distributions, e.g., Normal and Lognormal (Emam & Ai-Deek, 2006; Susilawati, Taylor, & Somenahalli, 2013; Uno, Kurauchi, Tamura, & Iida, 2009), or multimodal distributions, e.g., Gaussian Mixture distribution (Chen, Yin, & Sun, 2014; Feng, Hourdos, & Davis, 2012; Guo, Rakha, & Park, 2010; Kazagli & Koutsopoulos, 2013). However, due to the high labor and operation costs, such data sources are scarce and usually suffer from undesirable accuracy (Feng, Sun, & Chen, 2015). In general, only limited arterials can be detected, which may sometimes fail to reflect the path TTD of overall traffic in

CONTACT Weiliang Zeng weiliangzeng@gdut.edu.cn School of Automation, Guangdong University of Technology, Guangdong, China.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/gits.

a network-wide scale. The alternative approach is to derive path TTD based on link-level data sources, e.g., probe vehicles with GPS devices, which enables to cover a wider area than stationary equipment. The link TTDs are first estimated and then assembled to a joint distribution (e.g., Normal) by assuming the independence between individual link TTDs (Dailey, Wall, Maclean, & Cathey, 2000). However, such approach appears inappropriate since travel times on successive links are essentially dependent along signalized arterials. For example, when one link becomes congested, the adjacent links also get affected by this congestion. As pointed by Iida (1999), mutual relationships or dependencies among links should not be overlooked in reliability analysis.

To address this problem, this study adopts the copula model in econometrics (Trivedi & Zimmer, 2006), to estimate path TTD given a link set by accounting for spatial correlation between link travel times. Compared to multivariate distributions, one distinct feature of the copula model is that the dependence structure is unaffected by the types of marginal distributions, which offers greater flexibility in correlating individual link TTDs. Note that the temporal correlation between link travel time and link arrival time is not studied here but was presented in an earlier study (Chen, Liu, Qi, & Wang, 2013). The effect of signal coordination on the form of link and path TTD is also scrutinized in this study. It is hoped that these efforts may help provide practical insights into how link-level data sources can be fully utilized for path TTD estimation on signalized arterials.

The reminder of this article is organized as follows. A literature review of path travel time estimation is first presented, followed by a brief description of the copula theory and the estimation procedure for copula models. Then, a case study is conducted at one arterial in Hangzhou City, China to illustrate the application of the proposed methodology. The spatial correlations between link travel times are statistically examined and path TTDs are estimated by copula methods under both unfavorable and favorable signal coordination cases. The results are compared with the estimates without considering link correlations as well as empirical bimodal or multimodal distributions. The last section draws conclusions and provides recommendations for future work.

Literature review

Travel time along signalized arterials has been extensively studied in the literature. The vast majority of the related work has focused on the estimation of the average travel time. The representative studies can be referred to Highway Capacity Manual (2010), Skabardonis and Geroliminis (2005), Skabardonis and Geroliminis (2008), and Liu and Ma (2009). However, less 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. A thorough review is provided below.

Link-level TTD can capture the nature of interrupted flow under traffic signal control and has wide applications in practice. Ji and Michael Zhang (2013) used bus probe data to estimate travel times on urban street and identified bimodal TTD at link-level. The two modes relates to the travels with and without delays, respectively. Then, a mixture model was developed to characterize such bimodal TTD. Zheng and van Zuylen (2010) analyzed the delay distribution as the main component of link TTD, and built a probabilistic model to estimate delay distribution by accounting for stochastic arrivals and departures at signalized intersections. Through simulating the cyclic evolution of delay distribution, the temporal correlation between vehicle arrival time and link travel time was illustrated and demonstrated.

When planning trips, travelers care more about path-level TTD than link-level. A common approach to analyze path-level TTD is using statistical models to fit real travel time observations. Early studies resorted to unimodal distributions (e.g., Normal, Lognormal, Gamma, Weibull, and exponential distribution, etc.). One can refer to the studies by Emam and Ai-Deek (2006), Uno et al. (2009), and Susilawati et al. (2013). On the other hand, it has been argued that unimodal distribution may not well represent the variation of path travel times. For instance, travel times under free-flow and congested conditions can be significantly different. Increasing researchers tend to use multimodal or mixture models to characterize path TTD. Guo et al. (2010) and Feng et al. (2012) used a mixture of Gaussian distributions to estimate path TTD, while Kazagli and Koutsopoulos (2013) used a mixture of lognormal distributions. Chen et al. (2014) further used a finite mixture of regression model with varying mixing probabilities to explore path TTD under the impact of signal control. Such approach helps establish a connection between TTDs and underlying traffic states, through which the detailed analysis of the probabilities of each state and an overall better fitting can be achieved. However, as the basic input of the method, the path-level travel time data are not often available in urban network, which restrains the applicability of the distribution fitting approach. In real, a path usually includes multiple links and there exist a series of paths between one Origin-Destination (OD) pair. Considering that the link-level travel time data are commonly available, it will be more feasible to estimate path TTD based on individual link TTDs (Chen, Yu, Chen, & Wang, 2017b).

In most studies on path travel time estimation, the assumption of the independence of individual link travel times is made (Iida, 1999). Such assumption may be generally accepted when only the average path travel time is interested. However, when one is interested in TTD estimation, the assumption of link independence appears not appropriate any more. He, Liu, Kornhauser, and Ran (2002) stated 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. Dailey et al. (2000) and Pattanamekar, Park, Rilett, Lee, and Lee (2003) suggested using joint probability density function to model the correlated link TTD. Geroliminis and Skabardonis (2006) calculated path travel time variance under the assumption of linear correlations between consecutive link travel times. As an extended work, Ramezani and Geroliminis (2012) used Markov chains to estimate path TTD through the integration of correlated link TTDs. The product of the sequence of pair-wise Markov matrixes was calculated by assuming the transitions between different link pairs are conditionally independent. However, the proposed approach was data intensive and the effect of signal timing and coordination has not been analyzed, which is supposed to have direct impact on link-level TTD along arterials.

In summary, the characteristics of link TTD under different signal control strategies needs further investigation. The correlation between successive link TTDs on arterials has not been explicitly analyzed. Furthermore, based on individual link TTDs, how to characterize path TTD by accounting for their spatial correlation remains a critical task. Thus, copula models, which have been recognized and employed in the econometrics field (Trivedi & Zimmer, 2006) and recently are popular in transportation research (Bhat & Eluru, 2009; Chen et al., 2017b; Zou & Zhang, 2016), are introduced in this study to interpret the dependent structure of link TTDs and predict empirical TTD for a given path. In the next section, we provide a brief description of the copula theory, the Gaussian Copula model adopted in this study and its estimation method.

Methodology

As stated in Ramezani and Geroliminis (2012), given individual link TTDs, a simple model for path TTD estimation is to aggregate them independently. Assume a path consisted of K links, the path TTD can be computed as follows:

$$\mathrm{TTD}_K = \mathrm{TTD}_1 \times \mathrm{TTD}_2 \times \ldots \times \mathrm{TTD}_K \quad (1)$$

$$\left(\mathrm{TTD}_i \times \mathrm{TTD}_j\right)(t) \triangleq \int_{-\infty}^{\infty} \mathrm{TTD}_i(\tau)\mathrm{TTD}_j(t-\tau)\mathrm{d}\tau \quad (2)$$

where the (*) mathematical operator represents convolution and the left term in Eq.(2) is the probability density over time t for two links (i, $j=1, 2, \ldots, k$) given both TTDs *a priori*. The convolution method considers the independence of link TTDs and ignores the spatiotemporal correlation that exists between successive links.

To incorporate the correlation of link TTDs into path TTD estimation, a copula approach, which has been used for modeling multivariate distributions among random variables with pre-specified marginal distributions (Sklar, 1973), is adopted in this study. The copula approach enables a flexible way to characterize nonlinear dependence among consecutive links travel times regardless of their marginal distributions, i.e., the dependence structure is not influenced by the assumed marginal distributions. It will provide flexibility in correlating individual link travel times since they may not even have the same marginal distributions under different traffic flow and signal control strategies. The basic theorem on multivariate copulas is provided below.

Copulas

According to Sklar's theorem (Sklar, 1973), for an n-variate distribution $F(x_1, x_2, \ldots, x_n)$ with marginal distributions $F_1(x_1)$, $F_2(x_2), \ldots,$ $F_n(x_n)$, there exists an n-dimensional copula C such that for all $x_1, x_2, \ldots, x_n$:

$$F(x_1, x_2, \ldots, x_n) = C\left(F_1(x_1),\ F_2(x_2), \ldots, F_n(x_n)\right) \quad (3)$$

If marginal distributions are all continuous, then C is unique. Equation (3) essentially 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. As a general case, the bivariate distribution of two random variables x_1 and x_2 can be given by:

$$F(x_1, x_2) = C\left(F_1(x_1),\ F_2(x_2); \theta\right) = C(u_1, u_2; \theta) \tag{4}$$

where $F(x_1, x_2)$ is the joint distribution of x_1 and x_2; $u_1 = F_1(x_1)$ and $u_2 = F_2(x_2)$ are the corresponding marginal distributions of x_1 and x_2, respectively; $C(u_1, u_2; \theta)$ is the copula function; θ is the copula parameter describing the dependency between x_1 and x_2, which can be determined by Pearson's linear correlation coefficient, or rank correlation coefficient, e.g., the Spearman and Kendall correlation coefficients, or tail dependence, which relates to the amount of dependence at the upper-quadrant tail or lower-quadrant tail of a bivariate distribution.

A set of copula types can be found in Trivedi and Zimmer (2006) to describe the dependence between random variables, including the Gaussian copula, the FGM copula, and the Archimedean class of copulas. In this study, as an initial effort to employ copula methods, only Gaussian copula is introduced and utilized to model the dependent structure between individual link TTDs.

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 is essentially equivalent to a bivariate normal distribution if and only if the two marginal distributions are Gaussian and the dependent structure between them is a Gaussian copula function. The two-dimensional Gaussian copula can be defined as

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

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

$$\frac{\partial}{\partial u_1} C(u_1, u_2; \theta) = \Phi\left(\frac{\Phi^{-1}(\mathrm{u}_2) - \theta\Phi^{-1}(\mathrm{u}_1)}{\sqrt{1-\theta^2}}\right) \tag{6}$$

$$\frac{\partial}{\partial u_2} C(u_1, u_2; \theta) = \Phi\left(\frac{\Phi^{-1}(\mathrm{u}_1) - \theta\Phi^{-1}(\mathrm{u}_2)}{\sqrt{1-\theta^2}}\right) \tag{7}$$

The copula density is given by

$$c(u_1, u_2; \theta) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2; \theta) = \frac{1}{\sqrt{1-\theta^2}} \times \exp\left(\frac{2\theta\Phi^{-1}(\mathrm{u}_1)\Phi^{-1}(\mathrm{u}_2) - \theta^2\left(\Phi^{-1}(\mathrm{u}_1)^2 + \Phi^{-1}(\mathrm{u}_2)^2\right)}{2(1-\theta^2)}\right) \tag{8}$$

In terms of its copula density, the 2D Gaussian copula can be written as

$$C(u_1, u_2; \theta) = \int_0^{\mathrm{u}_1} \int_0^{\mathrm{u}_2} c(s, t; \theta) \mathrm{d}s\mathrm{d}t \tag{9}$$

Besides, one typical property of the Gaussian copula about its tail dependence is asymptotic independence. That is, regardless of the level of correlation assumed, upper-quadrant tail or lower-quadrant tail events appear to be independent in each margin. Thus, Gaussian copula is essentially unable to capture the dependence in the tails as compared to the other copulas, e.g., Clayton is radially symmetric with strong left tail dependence and weak right tail dependence and its right tail dependence goes to zero at right extreme. To explore other types of copula functions with tail dependence will be left as a future work.

Two-stage estimation for copula models

For copula models, because the dependence modeled through the copula is separated from the marginal distributions, Maximum Likelihood Estimation (MLE) method is utilized to estimate the marginal distributions and the copula model in two stages. It is important from a practical point of view because numerical optimization depends on a good starting point if the dimension of the parameter vector is large.

Estimation of marginal distributions

Both parametric and nonparametric estimators can be used to characterize the marginal distribution of link travel times. Parametric estimators include Normal, Lognormal, Gamma, and Weibull, etc. Nonparametric estimators include a kernel smoothing estimator with a Gaussian kernel. The density at x by the kernel smoothing method is given by

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{10}$$

where K is the kernel function, h is a smoothing parameter for the bandwidth and X_i represents the observed data. The selection of h needs to consider the bias of the estimation and its variance. In this study, Gaussian kernel is selected and the bandwidth h chosen in light of the optimality rules in Bowman

Figure 1. Study site: arterial schematic in Hangzhou City, China.

and Azzalini (1997), which are also implemented in Matlab.

Estimation of the copula

After estimation of marginal distributions, $\tilde{n}_j$ via a univariate maximum likelihood for $j = 1, 2, \ldots, n$ as shown in Eq. (3) can be obtained. Then, the copula parameter θ describing the dependency between the travel times of different links is estimated by maximizing the likelihood

$$L_C = \sum_{i=1}^{n} \log c\left(F_1(x_1; \tilde{n}_1),\ F_2(x_2; \tilde{n}_2), \ldots, F_n(x_n; \tilde{n}_n); \theta\right) \tag{11}$$

Note that in the process of estimation, each marginal link TTDs are transformed to values between 0 and 1. Based on uniformly distributed link TTDs, the study of dependence is therefore not subject to any difference in marginal link TTDs, which helps to achieve stable estimation of the dependent structure. Then, path TTD can be generated based on the estimated marginal link TTDs and copula model.

Case study

The proposed methodology was evaluated for the through movement of one arterial in Hangzhou, China. Following presented is the study site description and data preparation process; based on investigation of link TTDs, the Gaussian copula model was constructed for path TTD estimation under both unfavorable and favorable coordination cases; last, the comparison was made for path TTD estimates by the copula model, the convolution, and the empirical distribution fitting approach.

Study site description and data preparation

The study site is 1.7 km long stretch of a major urban arterial in Hangzhou, China. The study section consists of four signalized intersections with three links, as illustrated in Figure 1. There are two through lanes on each direction of the north-south arterial, with additional turning lanes at intersection approaches, and the speed limit is 40 km/h. Traffic signals are all four-phase operating with cycle lengths ranging from 170 to 180 s.

In addition to the geometric data available, a field study was conducted from 7 am to 10 am on 1 June 2007, to obtain comprehensive through movement counts, turning movement counts, and signal timing plans. The study period covers time-varying traffic flow patterns, e.g., a short low flow off-peak pattern, a longer high flow peak pattern, and a short mid-flow post-peak pattern. The field observation showed unfavorable signal progression along the arterial, especially for link 1. Due to platoon dispersion under long link spacing and the impact of mid-link traffic (from Zhaohui Road, as shown in Figure 1), travel times on link 1 showed significant fluctuation and less correlation with nearby links. Thus, based on the signal control plan at Intersection 2 downstream of link 1, signal timings were adjusted for Intersections 3 and 4 to enable a better coordination favoring the southbound direction. The traffic flow data (through and turning movement every 15 min) and signal timing data (both unfavorable signal coordination and the improved one) were coded into the VISSIM microscopic simulation model (PTV, 2008).

To ensure that the VISSIM model will correctly represent traffic performance at the site, calibration was conducted and VISSIM model parameters were adjusted to reproduce local driver behavior and traffic conditions. Considering that capacity has a significant effect on the system performance, i.e., delay and queues, an initial calibration was performed to

Table 1. The default and calibrated values of the identified capacity parameters in VISSIM.

Parameter	Average standstill distance	Additive part of safety distance	Multiplicative part of safety distance	Waiting time before diffusion
Default value	2 m	2 m	2 m	60 s
Calibrated value	2.5 m	1.5 m	3 m	60 s

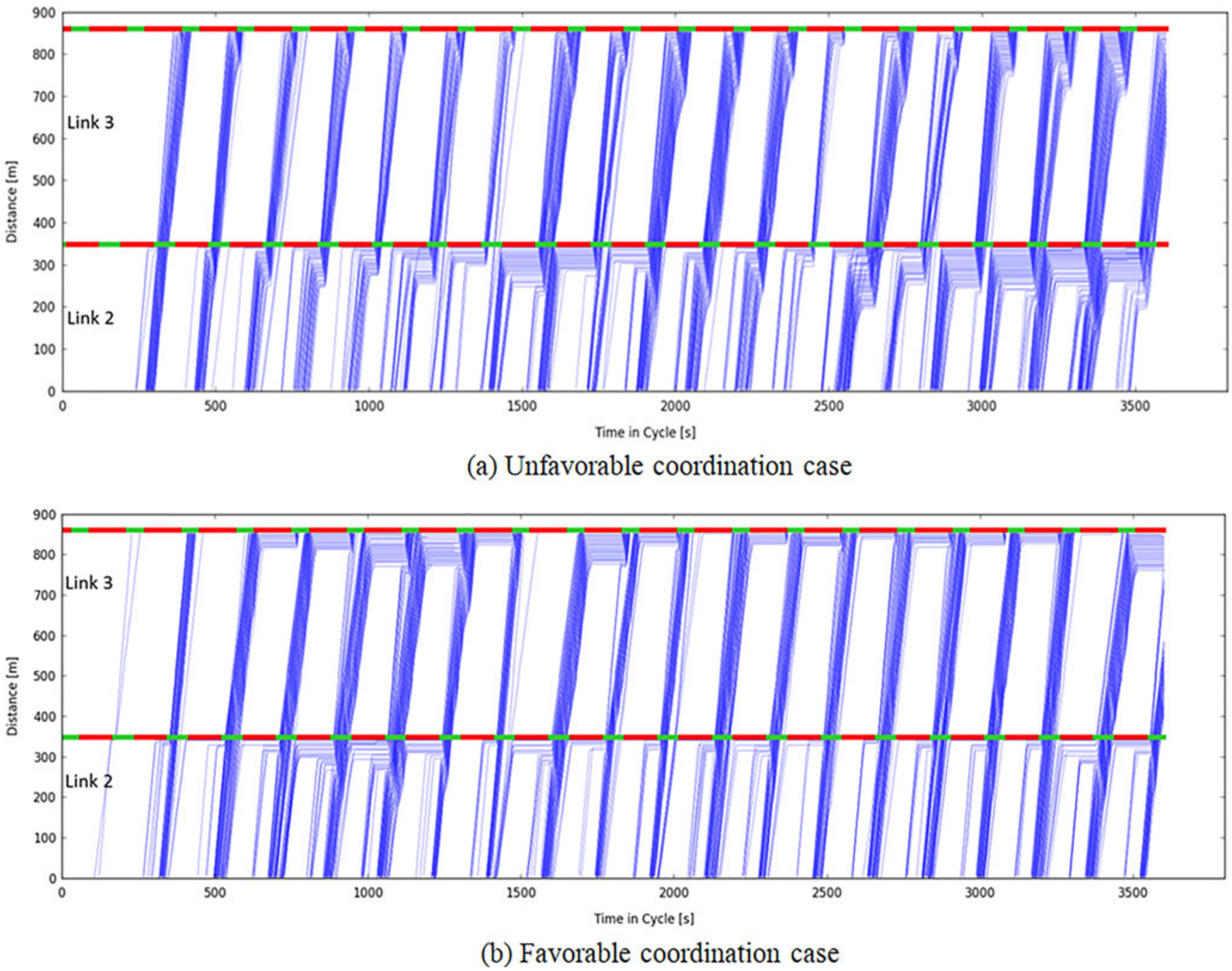

Figure 2. Time-space diagrams with vehicle trajectories during 7 am–8 am of the analysis period. (a) Unfavorable coordination case. (b) Favorable coordination case.

identify the values for the capacity adjustment parameters that cause the model to best reproduce observed traffic capacities in the field. First, based on field measurements and model estimates of capacity, a global calibration was performed to identify the appropriate value of the capacity parameters. Then, link-specific capacity parameters were selected and adjusted to fine-tune the model so that it more precisely matches the field-measured capacities.

According to field investigation, the key capacity parameters in VISSIM were identified as: (1) in Wiedemann 74 car following model: Average standstill distance, Additive part of safety distance, Multiplicative part of safety distance; and (2) in lane changing model: Waiting time before diffusion. By referring to the potential range of the identified capacity parameters in practice, Latin Orthogonal Design was employed to fine-tune the optimal combination of calibration parameters for minimizing the squared error between the field observations and the simulation model. Last, the overall traffic performance predicted by the simulation model was compared to the field measurements of travel time, queue lengths, and the duration of queuing. The results of calibrated parameters are shown in Table 1.

Based on the calibrated simulation model, time-space diagrams with vehicle trajectories were made for both unfavorable and favorable coordination cases, as shown in Figure 2. The diagrams help illustrate the site congestion. It can be found that after signal adjustment between intersections the traffic flow on the arterial become smoother with most of the vehicles traveling within the green band. Of particular interest for this study is the difference of link TTDs correlation under unfavorable and favorable signal coordination, and whether the proposed copula approach can capture such difference in path TTD estimation.

Before estimating arterial path travel time, the definition of an individual link needs to be made clear. Generally, a link should be defined such that consecutive links are contiguous over a section of the road and the link travel time includes the intersection delay (Bhaskar, Chung, & Dumont, 2009). Only in this way can travel time estimates from such links be used to obtain path travel time estimates, where a path consists of multiple links. Thus, a link in this study is defined from upstream stop-bar to downstream stop-bar, as illustrated in Figure 3. Accordingly, the lengths of three links are 764, 433, and 507 m, respectively, as in Figure 1. Link travel time is defined as the travel time from the moment the vehicle enters the upstream of the link to the moment it leaves the downstream intersection. The link travel time is then calculated as

$$TT_i = t_i' - t_i \tag{12}$$

Based on the definition of link travel time above, the simulation model was applied to estimate the TTDs of links 2 and 3 in southbound direction with a 3-hour simulation.

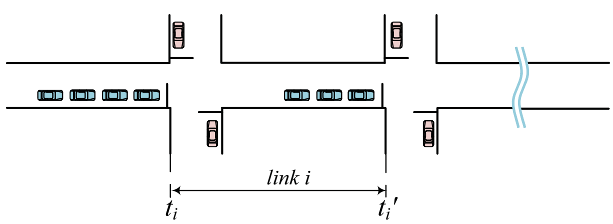

Figure 3. Definition of link travel time.

Investigation of link TTD

First, link TTDs were investigated by analyzing the travel times of all vehicles crossing successive links 2 and 3 during the evaluation period. Figure 4 gives a scatter diagram of joint TTD, with each dot representing the travel time of one vehicle in each of the two links. The shape of individual link TTDs was estimated by the kernel smoothing method, as in Eq. (10). Assumed that the feature of link travel time correlation have significant difference due to signal coordination, we analyzed the performance of the proposed model for unfavorable and favorable signal coordination cases, respectively.

In unfavorable coordination case, as shown in Figure 4(a), the values of individual travel times for each link scatter in a wide range of the joint distribution space. It is hard to identify any correlation or dependence pattern between scattered travel times along successive links. While in favorable coordination case, as shown in Figure 4(b), four travel time states

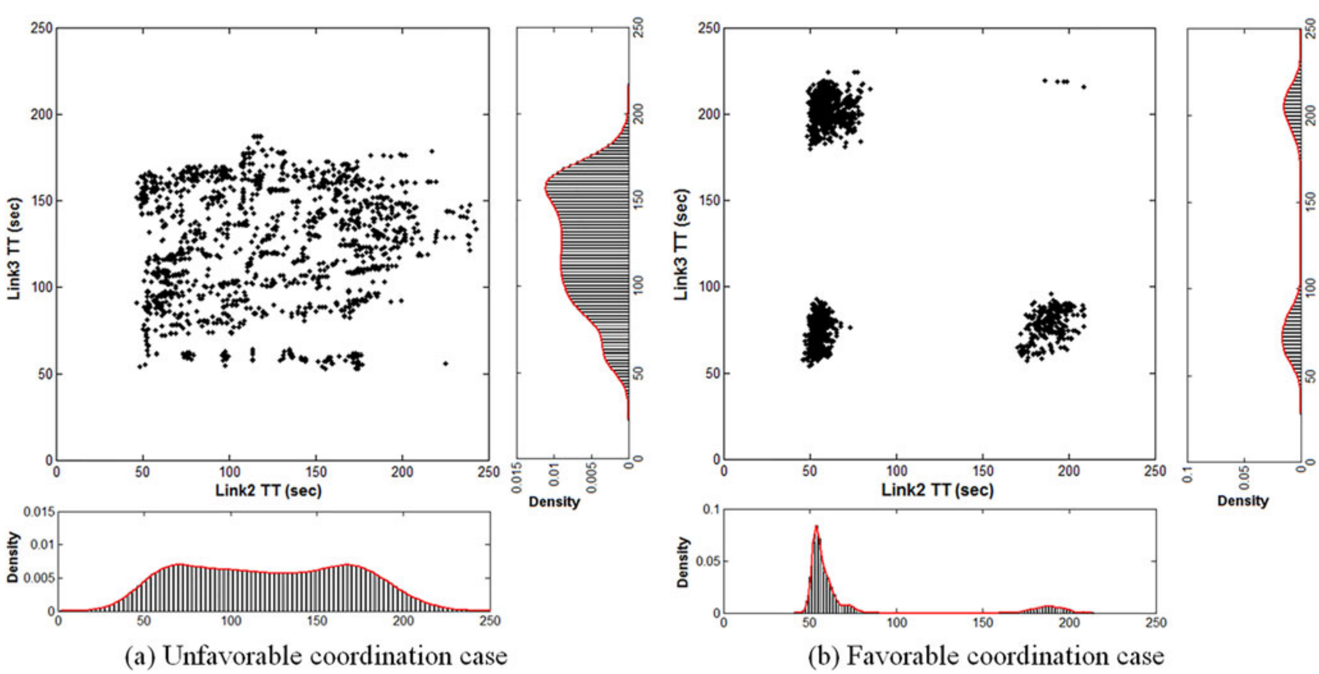

Figure 4. TTDs of successive links 2 and 3. (a) Unfavorable coordination case. (b) Favorable coordination case.

Table 2. Link travel time analysis.

	Link No.	Time period	No. of samples	Mean	Standard deviation	Coefficient of variation
Unfavorable coordination case	2	7 am–8 am	597	110.41	47.00	0.43
		8 am–9 am	705	139.56	49.75	0.36
		9 am–10 am	685	116.84	38.46	0.33
		Overall	1987	122.97	47.01	0.38
	3	7 am–8 am	597	95.41	11.66	0.12
		8 am–9 am	705	132.45	12.67	0.10
		9 am–10 am	685	142.12	41.77	0.29
		Overall	1987	124.65	32.91	0.26
Favorable coordination case	2	7 am–8 am	598	67.50	37.98	0.56
		8 am–9 am	787	74.33	45.30	0.61
		9 am–10 am	648	86.41	52.88	0.61
		Overall	2033	76.17	46.55	0.61
	3	7 am–8 am	598	122.86	64.93	0.53
		8 am–9 am	787	141.53	66.45	0.47
		9 am–10 am	648	127.79	64.26	0.50
		Overall	2033	131.66	65.81	0.50

Table 3. Correlation between travel times of links 2 and 3.

	Time period	No. of samples	Mean	Covariance	Correlation coefficient
Unfavorable coordination case	7 am–8 am	597	(110.41, 95.41)	179.053	0.13
	8 am–9 am	705	(139.56, 132.45)	218.90	0.15
	9 am–10 am	685	(116.84, 142.12)	−216.96	−0.14
Favorable coordination case	7 am–8 am	598	(67.50, 122.86)	−219.97	−0.09
	8 am–9 am	787	(74.33, 141.53)	−920.83	−0.31
	9 am–10 am	648	(86.41, 127.79)	−1192.92	−0.35

and corresponding correlation patterns can be identified for successive link travel times. The four travel time states include fast in both links, fast in link 2 but slow in link 3, slow in link 2 but fast in link 3, and slow in both links, which are also found by Ramezani and Geroliminis (2012). The state boundaries are quite clear compared to unfavorable coordination case and the link TTDs show apparent bimodal distributions implying two significantly different states. On the other hand, it shows the clear positive correlation (points along 45 degree diagonal line) and negative correlation (points along −45 degree diagonal line) in the favorable coordination case. It can be explained by the green-wave control such that some of the vehicles go through the successive links without stop (points in lower left) while some have to stop once (points in upper left and lower right). Notice that a desirable green-wave control should avoid the link travel times locating in the upper right side of Figure 4(b).

Detailed analyses of individual link travel times are given in Table 2. As shown, almost both link travel times first increase during the morning peak period, i.e., 7 am–9 am, then decrease during the mid-flow post-peak periods, i.e., 9 am–10 am, except for link 3 in unfavorable coordination case. It is worth noticing that the overall average travel time on link 2 decreases significantly (from 122.97 s to 76.17 s) only at a relatively small cost of travel time increment (from 124.65 s to 131.66s) on link 3 if we improve the signal timing strategy from unfavorable coordination to favorable. In addition, a larger variability of link travel times, indicated by coefficient of variation, is found for favorable coordination case. Note that signal coordination favors the mean value of link or path travel times, while the variability accordingly increases to some extent with the link TTDs being divided into clear-cut bimodal distributions, as illustrated in Figure 4(b).

Furthermore, analyses of link travel time correlation are given in Table 3. The Pearson correlation coefficients were computed to indicate the correlation for the pair of the successive links based on the link travel time of each vehicle. The results in Table 3 show the correlation between link travel times in both unfavorable and favorable coordination cases. Interestingly, it is found that the link travel time correlation is positive in most of the time periods (except for the period 9 am–10 am) for the unfavorable coordination case, while it is negative in all the time periods for the favorable case. This phenomenon is likely resulted from the integrated impact of traffic demand and signal control. For the unfavorable coordination case, the change of traffic demand may have a greater impact on the link correlation, while the signal control has a relatively smaller impact. Thus, the link travel time correlation naturally shows a positive feature. For example, as the traffic flow increases, the travel time for link 2 and link 3 increases simultaneously. However, this is not true in the favorable coordination case, because the signal

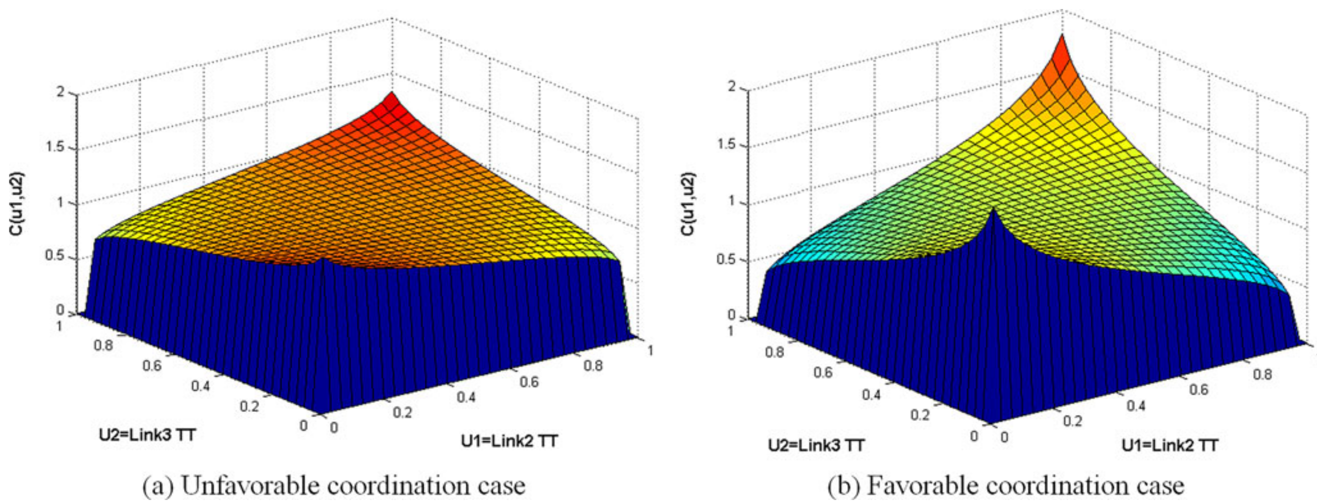

Figure 5. Estimated copula density. (a) Unfavorable coordination case. (b) Favorable coordination case.

control, rather than traffic flow, may have a leading impact on link correlation, which is shown in Figure 4.

Construction of Gaussian copula model

Using the two-stage estimator introduced in the last section, marginal link TTDs were first estimated by the kernel smoothing method, as illustrated in Figure 4, and then Gaussian copula models were constructed for unfavorable and favorable coordination cases, respectively. The Gaussian copula parameters θ are estimated to be 0.103 and 0.211 for both cases. The rank correlation coefficients, i.e., the Spearman and Kendall correlation coefficients, are estimated to be 0.066 and 0.099 for unfavorable case, and 0.135 and 0.201 for favorable case. Those indices imply the difference of dependence between travel times of successive links under the impact of different signal control strategies.

Based on the estimated copula parameters, copula densities for the two cases are shown in Figure 5. It is evident in Figure 5(a) that less dependence can be identified for successive link travel times in the unfavorable coordination case, whereas, the copula density in Figure 5(b) shows an increasing dependence in the favorable case. Note that a certain lower tail dependence and upper tail dependence can be identified, i.e., near (0,0) and (1,1), implying that the dependence between low travel time observations and between high travel time observations of successive links needs additional attention, especially in favorable coordination case. In reality, travelers are more interested in the co-occurrence of high or low travel time on the focused consecutive links. In copula theory (Trivedi & Zimmer, 2006), different copula models yield different upper or lower tail dependence measures, which may describe different likelihoods of experiencing high or low travel time on the unknown link given the same experience on the known link. However, given the relative ease of construction and estimation, the Gaussian copula method is adopted in this study as an initial effort but can hardly capture such tail dependence.

Estimation of path TTD

Based on the constructed Gaussian copula models, path TTDs were estimated for both unfavorable and favorable coordination cases. Using the experienced travel times on upstream link 2 as an input, the models predict the travel times on downstream link 3 by considering the dependence between them. The PDF and CDF of real path TTD, the results by the copula method and the convoluted estimation by Eq. (2) are shown in Figure 6. For unfavorable coordination case, path TTD on successive links 2–3 shows a unimodal distribution as in Figure 6(a). The estimation of TTD by the copula method is slightly better than the convoluted one. A possible reason is that the correlation between link TTDs is not significant in unfavorable coordination case. In such situation, the copula method considering the correlation between link TTDs has no apparent superiority over the convolution method without capturing the link correlation. While for favorable coordination case, path TTD shows apparently bimodal distributions. The TTD by the copula method is close to the real one and more accurate compared to the convoluted estimation. It implies with the increasing correlation between link TTDs under signal coordination, the superiority of the copula method over convolution becomes

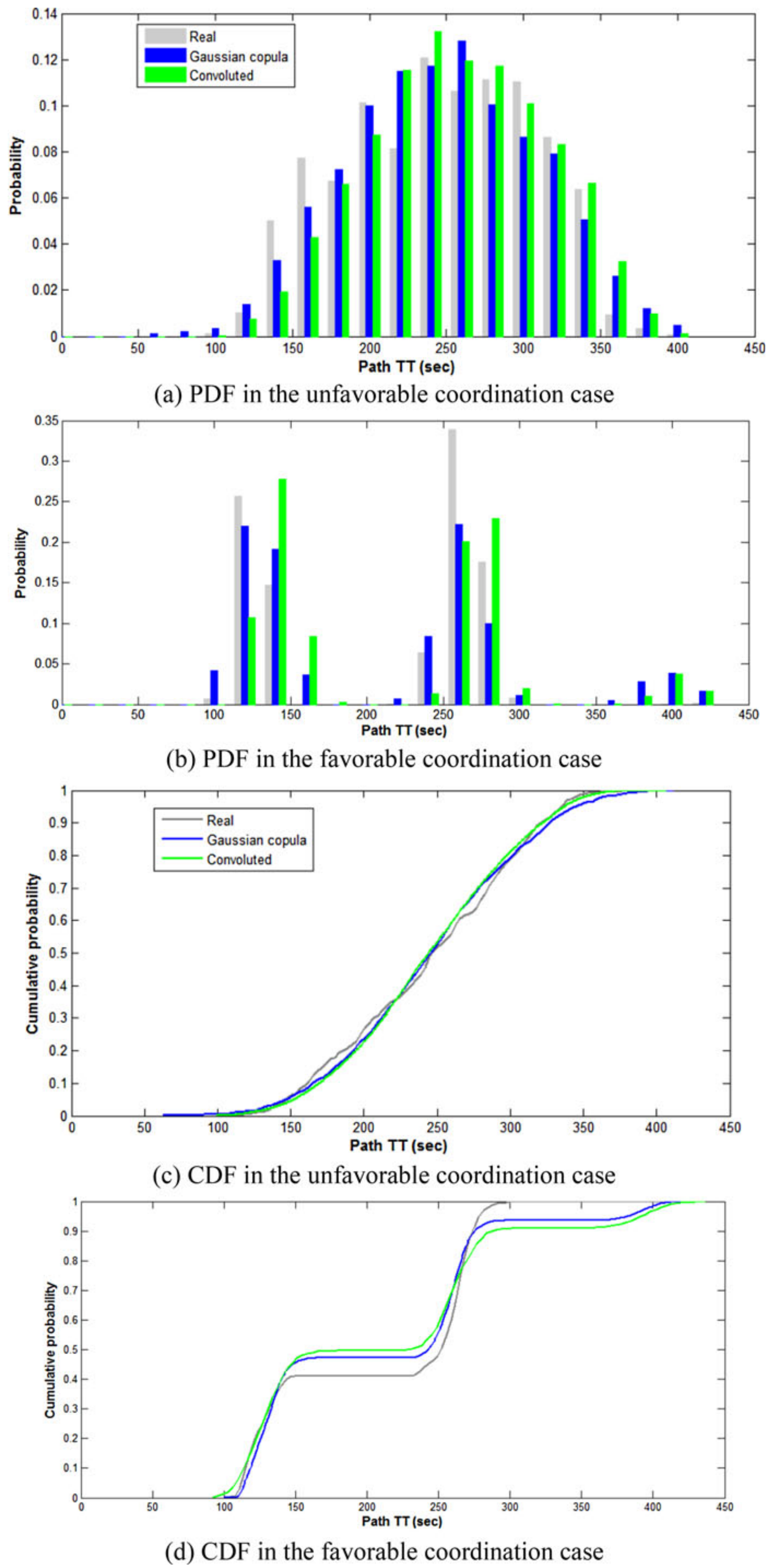

Figure 6. Path TTD estimation. (a) PDF in the unfavorable coordination case. (b) PDF in the favorable coordination case. (c) CDF in the unfavorable coordination case. (d) CDF in the favorable coordination case.

Table 4. Kolmogorov–Smirnov statistics of real TTD with estimated TTDs.

Type of distribution	Unfavorable coordination case	Favorable coordination case
Gaussian Copula	0.050	0.081
Convolution	0.066	0.117
Normal	0.068	0.254
Lognormal	0.086	0.282
Gumbel	0.082	0.271
Gaussian Mixture (with 2 components)	0.024	0.095

significant. Note that five samples with extreme path travel times, as illustrated in Figure 4(b) corresponding to slow traffic states in both links, lead to an overestimation of TTD in such boundaries by both convolution and copula methods. The potential reasons are provided below. For convolution method, it is prone to overfitting problem given scarce data points in certain intervals of sample distributions. For the copula method, this study adopts the Gaussian Copula one, which assumes asymptotic independence in both lower tail and upper tail, i.e., when travel times on two links which depend on each other through Gaussian Copula both have extreme low or high values, these extreme values tend to distribute independently. However, the five samples with extreme high path travel times, corresponding to the upper tail, are apparently not independent. Thus, the adopted Gaussian copula method fails to consider the joint distribution in the tails. Other types of copulas, e.g., Gumbel copula and BB1 copula capturing the tail dependence (Trivedi & Zimmer, 2006), are supposed to be tested for path TTD estimation in future work.

For more comprehensive comparison, the Kolmogorov–Smirnov statistic is introduced to quantify the maximum distance between CDFs of real TTD with the estimated ones by Gaussian copula, convolution, unimodal distributions (including Normal, Lognormal, Gumbel) and bimodal Gaussian Mixture distribution (with two components). The results are shown in Table 4. It can be seen that the advantage of the Gaussian copula method over the convolution and the distribution fitting methods are promising in both unfavorable and favorable coordination cases, though the Gaussian Mixture method performs slightly better than the Gaussian copula method in the unfavorable coordination case, with a K-S statistics of 0.024 versus 0.050). In the favorable coordination case, the Gaussian Mixture method has a slightly larger K-S statistics, i.e., 0.095, over Gaussian Copula, i.e., 0.081. Figure 6(b) indicates that there may exist the additional component (>2) of the path TTD, which cannot be adequately explained by the assumed bimodal Gaussian mixture model. Worth noticing is that the Gaussian copula method utilizes the link-level TTDs to estimate the unknown path-level TTD, while the distribution fitting methods attempt to approximate the known path-level TTD (unimodal or bimodal as in both cases). The copula approach demonstrates to have statistical superiority and practical value in real application.

Conclusions and future work

This article introduces a copula-based approach to characterize the dependent structure between links along arterials and then aggregate the individual link TTDs to estimate path TTD by accounting for spatial link correlation. To validate the copula-based model in an ideal tested scenario, VISSIM simulation with calibration is utilized to generate travel time data on one arterial in Hangzhou, China. Link-level TTD and their correlations are analyzed in both unfavorable and favorable coordination cases. Path TTD estimates by the copula model are compared to those by the convolution and the empirical distribution fitting approach. The main findings are summarized below:

1. Link-level TTDs can hardly be represented by unimodal distributions in both unfavorable and favorable coordination cases. Instead, link TTDs show apparent bimodal distributions implying two significantly different states. Especially in favorable coordination case, link TTDs are divided into clear-cut bimodal distributions, implying the significant impact of signal control upon link travel times.
2. Link travel times are found spatially dependent. Small correlation or dependence pattern is identified between scattered travel times along successive links in unfavorable coordination case. While in favorable case, TTDs of successive links under signal coordination show the clear positive correlation (points along 45 degree diagonal line) and negative correlation (points along −45 degree diagonal line). It implies that path TTD estimation needs to carefully consider such spatial link correlation, for which the copula method can be an appropriate approach.
3. For path TTD estimation, the copula-based approach has no apparent superiority over the convolution method in unfavorable coordination case, due to weak correlation between links. While for favorable coordination case, with the increasing correlation between links, path TTD estimates by the copula method are close to the

real ones and more accurate compared to the convolution and the empirical distribution fitting approach.

It is noteworthy that in this study we only analyzed the correlation between two neighboring links TTD under signal control. The results and conclusions draw above may be limited in scope. To extend initial findings, it will be necessary to analyze and model the path TTD consisting of increasing number of links. Other types of copula functions with different properties are supposed to be tested and compared for goodness-of-fit statistics (Trivedi & Zimmer, 2006). For instance, a Clayton copula is able to capture lower tail dependence; a Gumbel copula is able to capture upper tail dependency; a BB1 copula may capture both lower and upper tail dependence in the joint distribution of successive link travel times, of which traveler may be more concerned. Next, it remains considerable work to quantify the spatial link correlation under different signal control strategies and traffic conditions. The signal timing, offset setting, different levels of mismatch at upstream and down signals need to be inspected in detail. Besides, Zou, Yang, Zhang, Tang, and Zhang (2017) implied that when the traffic is characterized by heterogeneity, i.e., the observed link travel time data are generated from different sub-populations, the correlation value between consecutive links will vary depending on the traffic conditions. One potential approach will be to explore the applicability of the finite mixtures of multivariate distributions to characterize the heterogeneity existing in link travel time data. Last but not least, to improve the applicability of the proposed method in field implementation, experiments with real data, e.g., GPS data by probe vehicles (which need to be decomposed from path-level to link-level), will help examine the practical estimation of link travel times. These are directions for our future research.

Acknowledgments

The authors appreciate the National Natural Science Foundation of China (No. U1564212 and 51508014) and the Fundamental Research Funds for the Central Universities for kind support for this research.

Funding

The study was financially supported by the National Natural Science Foundation of China (No. U1564212 and 51508014) and the Fundamental Research Funds for the Central Universities.

Notes on contributors

Peng Chen received the Ph.D. degree from Nagoya University, Japan, in 2012. He is an assistant professor with the School of Transportation Science and Engineering, Beihang University, China. His research interest include urban traffic control and operation, travel time reliability, traffic flow modeling and simulation, intersection safety assessment.

Weiliang Zeng received the Ph.D. degree in civil engineering in Nagoya University, Japan, in 2016. He has been appointed by Guangdong University of Technology as an associate professor in department of automation. His current research interest include autonomous vehicle routing, route guidance system, traffic simulation and pedestrian behavior modeling.

Min Chen is currently working toward the master degree in the School of Transportation Science and Engineering, Beihang University, Beijing, China. Her research interests include urban traffic operation and control.

Guizhen Yu received the Ph.D. degree from Jilin University, China, in 2003. He is a full Professor with the School of Transportation Science and Engineering, Beihang University, China. His research interests include intelligent vehicle and urban traffic operations.

Yunpeng Wang is a professor in the School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety Control, Beihang University, Beijing, China. His research interests include intelligent transportation systems and connected vehicle.

References

Bates, J., Polak, J. W., Jones, P., & Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E*, *37*(2-3), 191–229.

Bhaskar, A., Chung, E., & Dumont, A. G. (2009). Integrating Cumulative Plots and Probe Vehicle for Travel Time Estimation on Signalized Urban Network. 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. *Transportation Research Part B: Methodological*, *43*(7), 749–765.

Bowman, A., & Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The Kernel approach with s-plus illustrations*. US: Oxford University Press.

Chen, P., Liu, H., Qi, H. S., & Wang, F. J. (2013). Analysis of delay variability at isolated signalized intersections. *Journal of Zhejiang University Science A*, *14*(10), 691–704.

Chen, P., Sun, J., & Qi, H. (2017). Estimation of delay variability at signalized intersections for urban arterial performance evaluation. *Journal of Intelligent Transportation Systems*, *21*(2), 94–110.

Chen, P., Tong, R., Lu, G., & Wang, Y. (2018). The α-reliable path problem in stochastic road networks with link correlations: a moment-matching-based path finding algorithm. *Expert Systems with Applications*, *110*, 20–32.

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. *Transportation Research Record: Journal of the Transportation Research Board*, *2442*(1), 96–105.

Chen, M., Yu, G., Chen, P., & Wang, Y. (2017). A copula-based approach for estimating the travel time reliability of urban arterial. *Transportation Research Part C: Emerging Technologies*, *82*, 1–23.

Dailey, D., Wall, Z., Maclean, S., & Cathey, F. (2000). *An algorithm and implementation to predict the arrival of transit vehicles*. Dearborn, MI: IEEE Intelligent Transportation Systems Conference (ITSC).

Etienne, H., Nicolas, C., & Ludovic, L. (2015). Applying variational theory to travel time estimation on urban arterials. *Transportation Research Part B: Methodological*, *78*, 169–181.

Emam, E., & Ai-Deek, H. (2006). Using real-life dual-loop detector data to develop new methodology for estimating freeway travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board*, *1959*, 140–150.

Feng, Y., Hourdos, J., & Davis, G. A. (2012). Bayesian model for constructing arterial travel time distributions using GPS probe vehicles. TRB 91st Annual Meeting, Washington, DC.

Feng, Y., Sun, J., & Chen, P. (2015). Vehicle trajectory reconstruction using automatic vehicle identification and traffic count data. *Journal of Advanced Transportation*, *49*(2), 174–194.

Geroliminis, N., & Skabardonis, A. (2006). *Real time vehicle reidentification and performance measures on signalized arterials*. Toronto, Canada: IEEE Intelligent Transportation Systems Conference (ITSC).

Guo, F., Rakha, H., & Park, S. (2010). Multistate model for travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board*, *2188*(1), 46–54.

He, R. R., Liu, H. X., Kornhauser, A. L., & Ran, B. (2002). *Temporal and spatial variability of travel time*. Irvine, CA: Institute of Transportation Studies, UC.

Highway Capacity Manual. (2010). Transportation Research Board, National Research Council, Washington, DC.

Iida, Y. (1999). Basic concepts and future directions of road network reliability analysis. *Journal of Advanced Transportation*, *33*(2), 125–134.

Ji, X., Ban, X., Zhang, J., & Ran, B. (2018). Moment-based travel time reliability assessment with Lasserre's relaxation. *Transportmetrica B: Transport Dynamics*. doi: 10.1080/21680566.2018.1434018

Ji, Y., & Michael Zhang, 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. TRB 92th Annual Meeting, Washington, DC.

Kazagli, E., & Koutsopoulos, H. N. (2013). Estimation of arterial travel time from automatic number plate recognition data. *Transportation Research Record: Journal of the Transportation Research Board*, *2391*(1), 22–31.

Liu, H., & Ma, W. T. (2009). A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials. *Transportation Research Part C: Emerging Technologies*, *17*(1), 11–26.

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. *Transportation Research Part C: Emerging Technologies*, *11*(5), 331–354.

PTV, VISSIM Version 5.1 User Manual. (2008).

Rakha, H., El-Shawarby, I., Arafeh, M., & Dion, F. (2006). *Estimating path travel-time reliability*. Toronto, Canada: IEEE Intelligent Transportation Systems Conference (ITSC).

Ramezani, M., & Geroliminis, N. (2012). On the estimation of arterial route travel time distribution with Markov chains. *Transportation Research Part B: Methodological*, *46*(10), 1576–1590.

Ramezani, M., & Geroliminis, N. (2015). Queue profile estimation in congested urban networks with probe data. *Computer-Aided Civil and Infrastructure Engineering*, *30*(6), 414–432.

Skabardonis, A., & Geroliminis, N. (2005). Real-time estimation of travel times on signalized arterials. Proceedings of the 16th International Symposium on Transportation and Traffic Theory, 387–406

Skabardonis, A., & Geroliminis, N. (2008). Real-time monitoring and control on signalized arterials. *Journal of Intelligent Transportation Systems*, *12*(2), 64–74.

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, *9*, 449–460.

Strategic Highway Research Program (SHRP 2) Report S2-L02-RR-2. (2014). Guide to Establishing Monitoring Programs for Travel Time Reliability.

Susilawati, S., Taylor, M. A., & Somenahalli, S. V. (2013). Distribution of travel time variability. *Journal of Advanced Transportation*, *47*(8), 720–736.

Tang, J., Liu, F., Zou, Y., Zhang, W., & Wang, Y. (2017). An improved fuzzy neural network for traffic speed prediction considering periodic characteristic. *IEEE Transactions on Intelligent Transportation Systems*, *18*(9), 2340–2350.

Trivedi, P. K., & Zimmer, D. M. (2006). Copula modeling: An introduction for practitioners. *Foundations and Trends® in Econometrics* , *1*(1), 1.

Uno, N., Kurauchi, F., Tamura, H., & Iida, Y. (2009). Using bus probe data for analysis of travel time variability. *Journal of Intelligent Transportation Systems*, *13*(1), 2–15.

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. *Transportation Research Part C: Emerging Technologies*, *56*, 309–334.

Zhang, Z., Wang, Y., Chen, P., He, Z., & Yu, G. (2017). Probe-data driven travel time forecasting for urban expressways by matching similar spatiotemporal traffic patterns. *Transportation Research Part C: Emerging Technologies*, *85*, 476–493.

Zheng, F., & van Zuylen, H. J. (2010). Uncertainty and predictability of urban link travel time: a delay distribution based analysis. Transportation research record. *Journal of the Transportation Research Board*, *2259*, 80–95.

Zou, Y., Yang, H., Zhang, Y., Tang, J., & Zhang, W. (2017). Mixture modeling of freeway speed and headway data using multivariate skew-t distributions. *Transportmetrica A: Transport Science*, *13*(7), 657–678.

Zou, Y., & Zhang, Y. (2016). A copula-based approach to accommodate the dependence among microscopic traffic variables. *Transportation Research Part C: Emerging Technologies*, *70*, 53–68.