Journal of Highway and Transportation Research and Development
2024,18(2):56-66
ISSN 2095-6215
DOI: 10.26599/HTRD.2024.9480014

# Using License Plate Recognition Data to Gain Insight into Urban Travel Time Distributions

Xiaoqin Luo*

(Wuhan Planning and Design Institute, Wuhan, Hubei 430014, China)

**Abstract**: Travel time distribution is an essential tool for measuring urban traffic performance, a subject that has been studied for decades. This paper conducts a comprehensive investigation of two types of travel time distributions using extensive license plate recognition data from Automatic Number Plate Recognition techniques on four signalized arterials in Guiyang, China. The travel time plane distributions presented in the overlay charts of observed travel times usually exhibit significant stratified data strips. When considering signal schemes, we observe that the cycle times of the first upstream and last downstream intersections are the determining factors for the data patterns of travel time plane distributions. We also investigate the characteristics of single or multiple peaks within various departure time windows. The results indicate that travel time statistical distributions are more likely to exhibit multiple states under short time windows. As for the shapes of travel time statistical distributions, skewness and kurtosis are used as descriptive statistics. The results show that the majority of statistical distributions are positively skewed and leptokurtic, and the skewness is highly correlated with the kurtosis. A stable skewness and kurtosis at a relatively lower level may be caused by lower travel time reliabilities.

**Key words**: traffic engineering; travel time distribution; urban roads; signalized intersections; license plate recognition data; ANPR cameras

## 1 Introduction

The distribution travel time is a crucial factor estimatiing and predicting travel time reliability on urban roads. Research on fitting various kinds of continuous distributions to observed travel time data began decades ago. The choice of distribution forms often depends on the observed properties of the field data. Initally, the normal distribution was considered the most suitable form. However, increasing studies have found that skewed distributions, such as the lognormal distribution, Weibull distribution, Gamma distribution, better fit the observed travel times[1-2]. For example, SUSILAWATI demonstrated the Burr distribution can better fit the positive-skewed travel times, and road lengths and saturation degrees can influence the shapes of distributions[3-5]. DELHOME' research suggests that the Halphen distribution also performs well when fitting the real data with long tails and positive skewness[6]. CAO and SRINIVASAN adopted truncated and the shifted forms of distributions to match field data considering of non-negative travel times, demonstrating the superiority of these modified forms over the original forms[7-8].

Later, mixed models were proposed when multistate travel times were discovered. The common method to observe the multimodality of a distribution is to visualize observed travel times by plotting frequency distribution histograms. Besides, a mathematical technique called the Hartigan dip test is used to test for multimodality[5,9-10]. Some studies point out that the two-component mixture normal distribution is useful in modeling the observed data[5,11-13]. Moreover, normal mixture distributions with more than two components have also been used[14-16]. GUO suggests that multimodality and skewness can be simultaneously captured using with mixed lognormal distributions, concluding that the main factors causing skewness are be uneven distributions between different traffic states and within the same traffic state[14]. Chen adopts mixed forms of normal distribution, lognormal distribution, and other types of distributions to fit observed travel times[17].

Some studies try different kinds of models to find the

Received 8 August 2023; Revised 19 December 2023; Accepted 2 March 2024

* *E-mail address*: 810061761@qq.com

most appropriate one for the travel time data in different locations and periods[9, 18-21]. CHAI and MAZLOUMI demonstrate that travel times follow Weibull distribution or normal distribution during peak hours, and a lognormal distribution during off-peak hours[19-20]. MAZLOUMI and LI find that travel times follow lognormal distribution at large time windows (e. g. , an hour or a day), but follow normal distribution at small time windows (e. g. , 5 min)[18-19]. Using travel time data over a month, GUO demonstrates that travel time distributions forms remain stable across different aggregation time intervals (e. g. , 5, 10, 30, 60 min)[22].

To date, different forms of distributions have been appropriate for different studies. The fitting degree depends on the consistenciy between the properties of the distributions and the observed travel times, such as skewness or the number of peaks. To the best of our knowledge, existing research focuses more on distribution forms rather than the properties of observed travel times. However, many distribution forms may be suitable for the same data. For example, both Weibull distribution and Gamma distribution can fit skewed travel times. Therefore, this paper focus on the properties of observed travel times.

For a given urban road, the geometric structure and driving behaviors are assumed to remain unchanged or similar, leaving traffic demand and signal controls as the primary determinants of travel times[12]. Hence, the relationship between travel time properties and signals should be carefully considered. In this paper, we use the data recorded by Automatic Number Plate Recognition cameras and signal machines to empirically investigate urban travel time distributions. The data collected on four arterials in Guiyang, China is first introduced, followed by a simple description to explain the main processes. The next section characterizes some properties of urban travel time distributions and explores the relationships with some potential factors, such as signal schemes and time windows. The conclusions are summarized in the final section of the paper.

## 2 Data

The license plate recognition (LPR) data for the study were supplied from Automatic Number Plate Recognition (ANPR) cameras installed at the stop lines of several main arterials in Guiyang, China. The ANPR cameras record vehicles' license plate numbers and passing timestamps when the vehicles pass through the stop lines. The LPR data of four arterials were collected on Saturday, February 21st, 2016. Table 1 list the number of lanes and the all-day detected volumes in each stop line, with the intersection points, P1 to P23, depicted in Figure 1. It should be noted that not all lanes were installed with ANPR cameras. From Table 1, the amount of detected LPR data is considerable in spite of some lower detected volumes in some intersections, which provides a basis to obtain sufficient travel time data. According to the signal schemes recorded by signal machines, we collected phase-to-phase travel times ($T_t$) between any two intersections within an arterial by re-recognizing vehicles' passing timestamps. Here, phase-to-phase travel times correspond to the vehicles passing stop lines in the same phases in both the first and last stop lines. At the same time, the travel time data was divided into different datasets according to the pre-timed periods of each end of the intersections. Therefore, the travel time samples in the same dataset were collected under the same signals, which allows us to distinguish signals' influence.

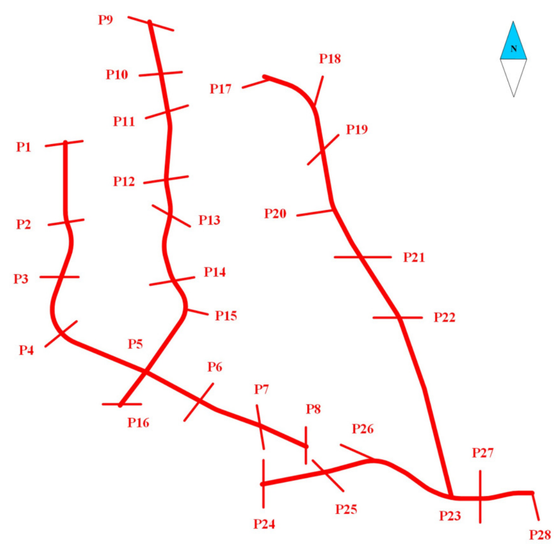



**Fig. 1 Study site in Guiyang, China**

From the ANPR cameras, we get 863 datasets of travel times. Since some datasets appear to be small samples due to the low traffic volumes at midnight, the short period of pre-timed signals, or the low coverage of ANPR cameras in some locations, we take the sample size of

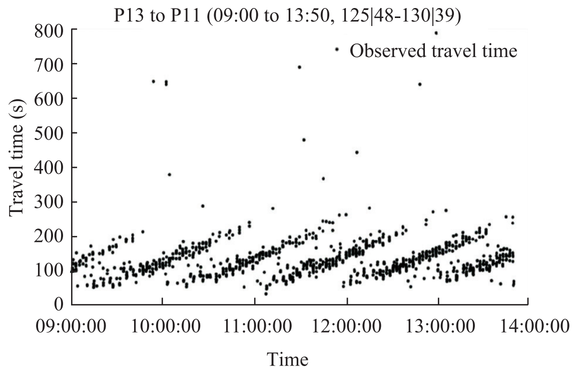
300 as a threshold to select 398 datasets to do further analyses. Among these datasets, 163 of them are from two nearby intersections; 116, 75, and 41 datasets are from three, four, five adjacent intersections respectively. For any dataset, four types of data are used: period, signal schemes (including cycle lengths, green times), $T_t$ statistical distributions, and overlay charts of observed travel times. As shown in Figure 2(a), the overlay chart visualizes observed travel times on a rectangular coordinate plane, with timestamps on the $x$-axis and travel times on the $y$-axis. Since the over charts describe travel times distributions on a rectangular plane, we rename it as the $T_t$ plane distribution to echo the $T_t$ statistical distribution. Figure 2 shows an example of these two distributions from P13 to P11 between 9:00 and 13:50, with the period and the signal schemes listed.

**Table 1 Details of the four arterials**

| Arterial number | | | Details of each arterials | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | From north to south | Intersection point | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | |
| | | Number of lanes | 5 | 5 | 6 | 4 | 6 | 4 | 7 | 3 | |
| | | Detected volumes (veh) | 24 651 | — | 6 065 | 10 341 | 17 309 | 10 676 | 12 655 | 8 846 | |
| | From south to north | Intersection point | P8 | P7 | P6 | P5 | P4 | P3 | P2 | P1 | |
| | | Number of lanes | 3 | 5 | 4 | 5 | 4 | 6 | 5 | 5 | |
| | | Detected volumes (veh) | 9 109 | 14 285 | 10 675 | 12 723 | — | 21 267 | 33 743 | 33 984 | |
| 2 | From north to south | Intersection point | P9 | P10 | P11 | P12 | P13 | P14 | P15 | P5 | P16 |
| | | Number of lanes | 3 | 4 | 5 | 5 | 6 | 5 | 5 | 5 | 6 |
| | | Detected volumes (veh) | 7 020 | 28 983 | 24 937 | 41 515 | 25 818 | 26 337 | 14 624 | 19 891 | 8 561 |
| | From south to north | Intersection point | P16 | P5 | P15 | P14 | P13 | P12 | P11 | P10 | P9 |
| | | Number of lanes | 6 | 5 | 5 | 5 | 6 | 5 | 5 | 5 | 4 |
| | | Detected volumes(veh) | 13 003 | 16 113 | 18 793 | 20 511 | 99 77 | 33 027 | 24 872 | 26 335 | 21 932 |
| 3 | From north to south | Intersection point | P17 | P18 | P19 | P20 | P21 | P22 | P23 | | |
| | | Number of lanes | 7 | 5 | 5 | 7 | 9 | 5 | 5 | | |
| | | Detected volumes(veh) | 22 083 | 31 235 | 39 112 | 22 424 | 34 200 | 24 813 | 8 162 | | |
| | From south to north | Intersection point | P23 | P22 | P21 | P20 | P19 | P18 | P17 | | |
| | | Number of lanes | 6 | 5 | 8 | 6 | 6 | 4 | 7 | | |
| | | Detected volumes(veh) | 8 114 | 20 002 | 22 234 | 33 744 | 40 538 | 38 001 | 53 908 | | |
| 4 | From north to south | Intersection point | P24 | P25 | P26 | P23 | P27 | P28 | | | |
| | | Number of lanes | 6 | 6 | 5 | 6 | 4 | 3 | | | |
| | | Detected volumes(veh) | 20 208 | 22 244 | 12 620 | 8 114 | 14 583 | 11 530 | | | |
| | From south to north | Intersection point | P28 | P27 | P23 | P26 | P25 | P24 | | | |
| | | Number of lanes | 3 | 3 | 4 | 5 | 5 | 6 | | | |
| | | Detected volumes(veh) | 8 196 | 13 053 | 8 880 | 12 630 | 20 184 | 15 117 | | | |

## 3 Methodology

The study methodology aims to explore travel time distributions in urban roads by focusing on two dimensions of distributions. In $T_t$ plane distributions, the data pattern in overlay charts is investigated by analyzing the parameters of signals; see more explanations in the next sub-section. The premise of showing significant data strips [see Figure 2(a)] is first examined, followed by qualitative statistics of data strip patterns with different signal schemes and route constitutions (e.g., two intersections or three intersections). Then, the multimodality of $T_t$ statistical distributions is tested by the Hartigan dip test, followed by correlation analyses between the descriptive statistics (including skewness $\beta_1$ and kurtosis $\beta_2$) and some potential relative factors. The Hartigan dip




(a) Plane distribution (b) Statistical distribution

* P13 to P11 (09:00 to 13:50, 125|48-130|39) refers to data obtained from P13 to P15 between 9:00 and 13:50, with upstream cycle 125 s, upstream green 48 s, downstream cycle 130 s, and downstream green 39 s.

**Fig. 2 An example**

test is a method to test the multimodality by calculating the dip statistic as the maximum difference between the empirical distribution function of all samples and the unimodal distribution function that minimizes the maximum difference[5,9]. It produces a probability that the observed data could come from a unimodal distribution and can be used to test the null hypothesis that the data are unimodal. Here, the $p$-value of the Hartigan dip test is also considered as a descriptive statistic of the shape of statistical distributions, just as skewness $\beta_1$ and kurtosis $\beta_2$.

## 4 Travel time distribution analysis

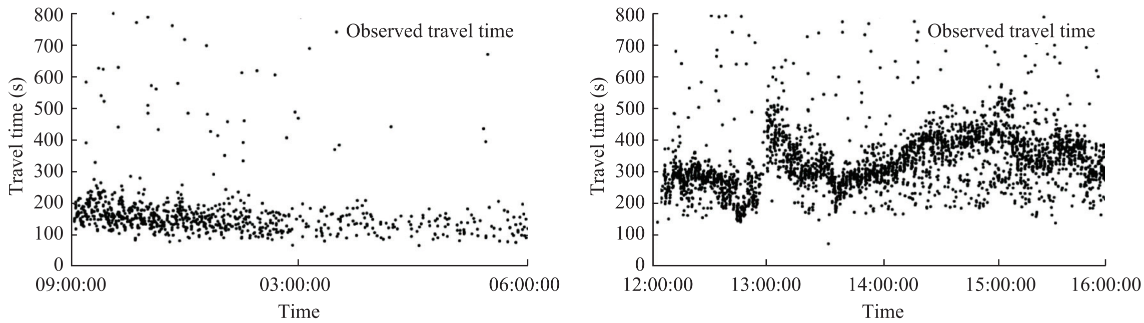
### 4.1 Travel time plane distributions

The $T_t$ plane distribution has been used as a visual tool to study the travel time data[16, 23-24]. Young explores the patterns of $T_t$ plane distributions by overlaying the travel times obtained from Bluetooth on different days (e. g. , 30 days). In this paper, the sample size of travel times obtained from ANPR cameras is relatively high. Then, the data collected at the same time can be directly used for analyses, thus avoiding the influence of different traffic conditions on different days. This section aims to explore the data patterns of $T_t$ plane distributions at various kinds of signal schemes and route constitutions. The former is designed to understand whether the relationship between upstream and downstream signals will influence data strip patterns. The latter is to understand whether different route constitutions will affect travel time patterns.

**Table 2 Frequency of datasets with and without significant data strips**

| Number of datasets divided by periods | | | | |
|---|---|---|---|---|
| Periods | 0:00 to 6:00 | | 6:00 to 24:00 | |
| With significant data strips | 2 | | 197 | |
| Without significant data strips | 72 | | 127 | |
| Number of datasets divided by route constitutions | | | | |
| Number of intersections | 2 | 3 | 4 | 5 |
| With significant data strips | 79 | 62 | 35 | 23 |
| Without significant data strips | 84 | 54 | 40 | 21 |

As shown in Table 2, the observed travel times are more likely to form significant data strips from 6:00 to 24:00, which are more stochastic from 0:00 to 6:00. Here, the $T_t$ plane distributions are divided into distributions with and without significant data strips according to the apparent degree of data strips. Distributions with significant data strips are strictly defined when significant data strips appear during the whole study period, while distributions with the existence of stochastic travel times are divided into another category. We list two typical examples of $T_t$ plane distributions without significant data strips in Figure 3. In the first case [see Figure 3(a)], the traffic flows are stochastic and keep stable at a low flow rate on the four-intersection route from P17 to P20, especially between 3:00 and 6:00. Another case mainly occurs when green ratios are relatively long, e. g. , as much as three quarters from P20 to P17 between 12:00 and 16:00 [see Figure 3(b)]. Therefore, a preliminary conclusion is that significant data strips in $T_t$ plane distributions are more likely to appear with higher traffic flows




(a) P17 to P20 (00:00 to 06:00, 65|20-54|15) (b) P20 to P17 (12:00 to 16:00, 120|85-120|80)

**Fig. 3 Examples of $T_t$ plane distributions without significant data strips**

and lower green ratios. Besides, the number of intersections seems to have no direct relationship with the significance of data strips (see Table 2). We might think a route with more intersections is less likely to show significant data strips because the travel time samples become smaller due to the detection failure or the existence of branch flows. But the data strips remain apparent in the routes with more than five intersections in the data used. It may be the result of the higher arterial traffic flows and the good quality of the ANPR data. Besides, we can deduce the existence of data strips over long arterials in spite of the possible insufficient samples which make the data strips hard to observe.

Then, the 199 datasets with significant data strips are divided according to the route constitutions and the cycles of the first upstream and last downstream intersections, $c_u$ and $c_d$ (see Table 3). Here we use the slopes of data strips to represent different types of $T_t$ plane distributions, where Z, P, N, PN represent zero, positive, negative, positive and negative, respectively. It can be observed that the majority of zero-slope data strips mainly occur in the case of equal upstream and downstream cycle times. As for the positive (or negative) slopes, more datasets are concentrated in the case of shorter (or longer) upstream cycle times. Such concentration of distributions is more distinct on the routes with more intersection. For $T_t$ plane distributions with both positive and negative slopes, they seem to appear equally with shorter and longer upstream cycle times.

**Table 3 Classifications of $T_t$ Plane Distributions With Significant Data Strips**

| Number of intersections | 2 | | | | 3 | | | | 4 | | | | 5 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Slope of data strip | Z | P | N | PN | Z | P | N | PN | Z | P | N | PN | Z | P | N | PN |
| $C^u>C^d$ | 3 | 1 | 26 | 6 | — | — | 20 | 5 | — | — | 12 | 3 | — | — | 11 | — |
| $C^u=C^d$ | 9 | 2 | — | — | 10 | — | — | — | 6 | — | — | — | 2 | — | — | — |
| $C^u<C^d$ | 3 | 22 | 3 | 4 | 1 | 19 | 2 | 5 | — | 9 | — | 5 | 1 | 9 | — | — |

To further understand the slopes of data strips, we imitate the analyses of intersection delay patterns in [25] to show the travel time patterns in a single lane between two fixed-time intersections. As shown in Figure 4(a), vehicles enter the link saturatedly during upstream green intervals, travel with the free-flow speed, then queue up before the downstream stop line, and finally depart during downstream green intervals. Vehicles' travel times on the link (from 50 m to 150 m) are marked as blue lines. It can be observed that travel times vary from cycle to cycle due to the changes in offsets. The latter cycle always has the offset $C^u-C^d$ seconds longer than the previous cycle, thus resulting in the travel time $C^u-C^d$ seconds shorter. If we connect the travel times of different cycles, a data strip $I$ marked as red quadrilateral can be observed, the same as the data strip in $T_t$ plane distributions. The mismatch between upstream and downstream cycles leads to $C^d-C^u$ seconds of travel time displacement between adjacent cycles. This is also the reason that the slope signs of significant data strips are consistent with the signs of $C^d-C^u$ in most cases.

However, according to the results in Table 3, the change in offsets is not the only factor that affects data strips. For example, equal cycle times of upstream and downstream intersections also form positive data strips. Here, a case with equal upstream and downstream cycles is designed in Figure 4(b) to exclude the influence of offset changes and explore other influencing factors. The travel time jump caused by the oversaturation and the starts of downstream reds can be observed in each cycle. Besides, two types of data strips, II and III, can be found, which is related to the variation of queues between cycles. More and more oversaturated vehicles are accumulated into the next cycle, so vehicles need more time to depart. Besides, compared with vehicles in type II, vehicles in type III need to wait for one more downstream red light and form a shorter strip which is hard to observe. We guess that when we observe one type of significant data strips in a $T_t$ plan distribution, there may be another less significant data strip with a different slope. In such a case, we still classify the types of $T_t$ plane distributions according to the significant degree of data strips, since the empirical analysis in this paper is more like a description of visual impressions. To conclude, changes in queues also affect the slopes of data strips.

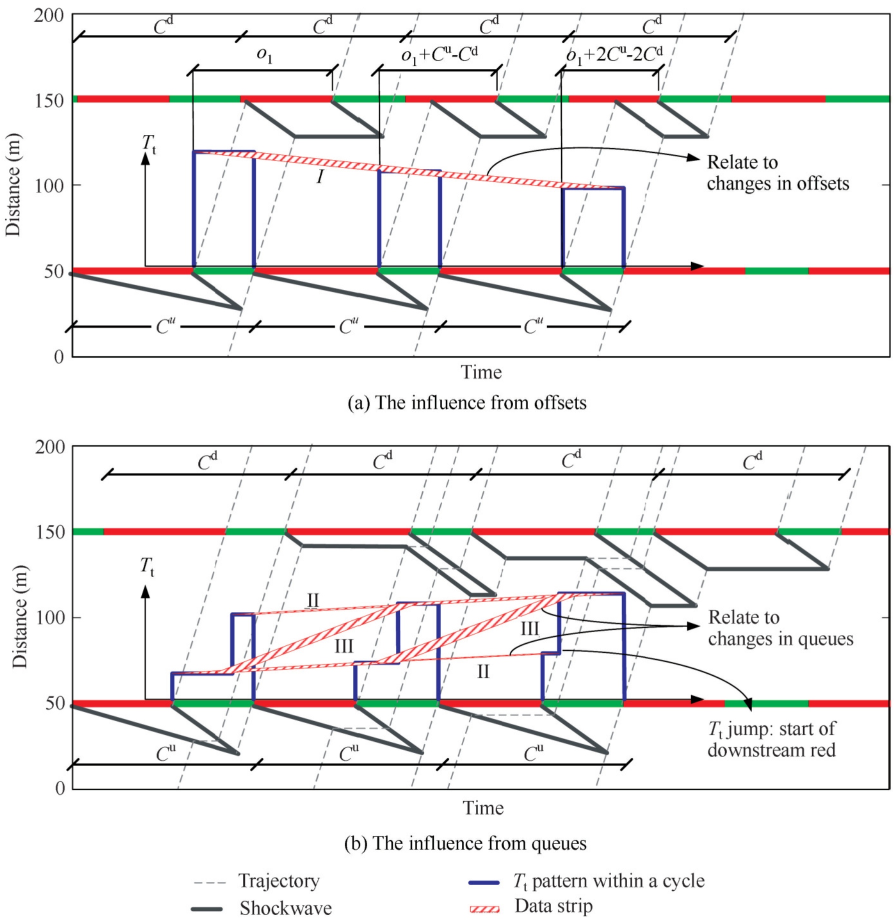
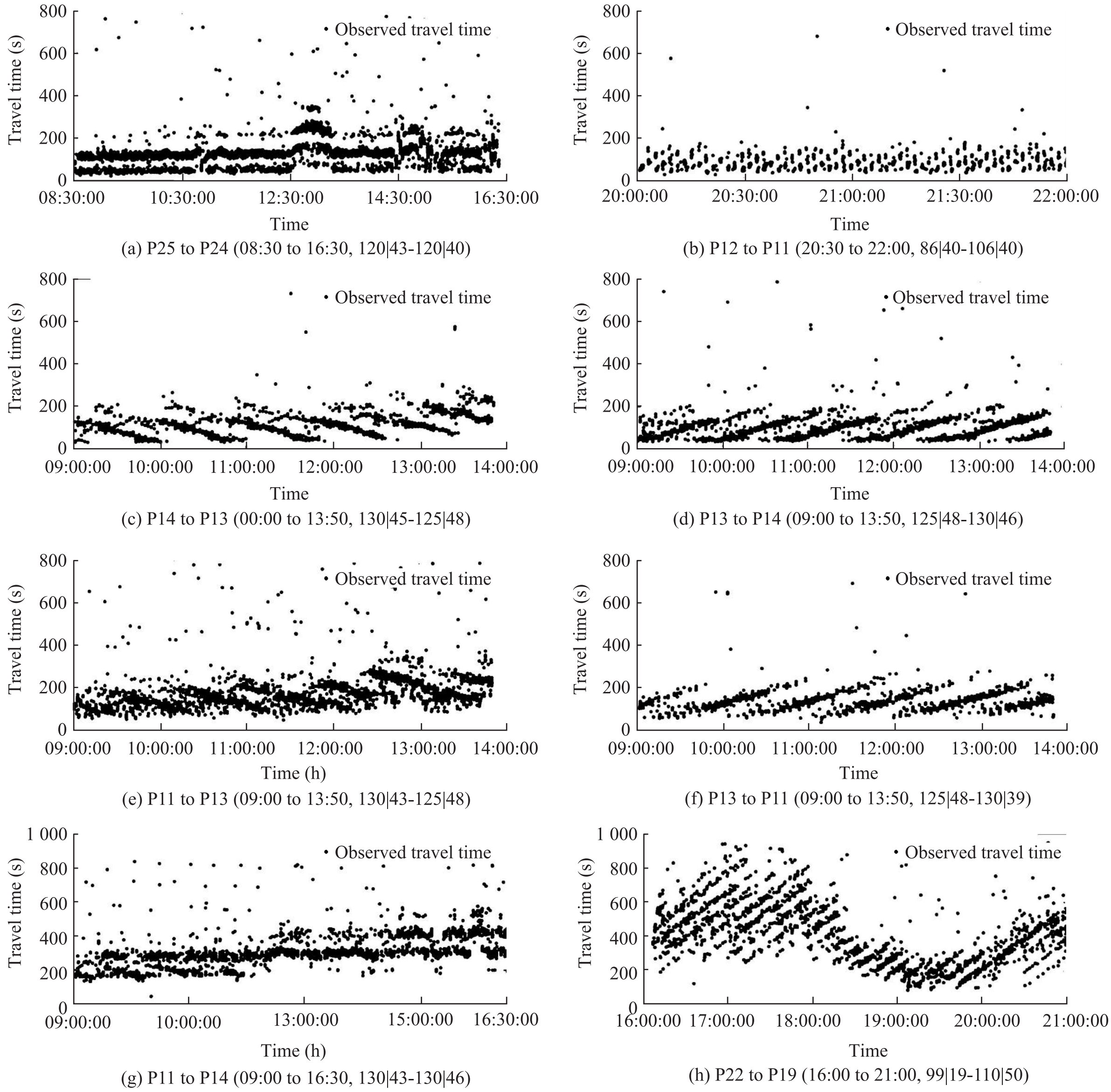



**Fig. 4 Travel time patterns between two stop lines**

We also show some examples of the $T_t$ plane distributions with significant data strips in Figure 5. The first four subplots are from routes with two neighboring intersections, while the following subplots are from roads with three and four intersections. It seems that the number of intersections does not weaken the significance of stratified data strips. The slopes of the eight $T_t$ plane distributions are Z, PN, N, P, N, P, Z, and P, respectively. A two-way $T_t$ plane distributions of the same intersection pairs, P13 and P14, as well as P11 and P13, are depicted from Figures 5(c) to (f). It seems that the slopes of the data strips in the two directions have the opposite arithmetic signs. Such a phenomenon also demonstrates that the types of data strips are directly affected by the relationships between the upstream and downstream cycle times. Besides, we also depict the $T_t$ plane distributions of a route with four intersections and its sub route or link [see Figures 5(g), (e) and (d)]. The $T_t$ plane distri-




(a) P25 to P24 (08:30 to 16:30, 120|43-120|40)

(b) P12 to P11 (20:30 to 22:00, 86|40-106|40)

(c) P14 to P13 (00:00 to 13:50, 130|45-125|48)

(d) P13 to P14 (09:00 to 13:50, 125|48-130|46)

(e) P11 to P13 (09:00 to 13:50, 130|43-125|48)

(f) P13 to P11 (09:00 to 13:50, 125|48-130|39)

(g) P11 to P14 (09:00 to 16:30, 130|43-130|46)

(h) P22 to P19 (16:00 to 21:00, 99|19-110|50)

**Fig. 5 Examples of $T_t$ plane distributions with significant data strips**

butions of the sub-route from P11 to P13 and the sub-link from P13 to P14 appear to be totally different with the slopes approaching opposite numbers. On the route from P11 to P14, the two different travel time tendencies from P11 to P13 and from P13 to P14 in the subs are just offset, thus making the travel times over the route be stable in each strip. Such zero-slope data strips over the route can also be deduced from the equal cycle times in the first and last intersections, P11 and P14. Therefore, we just need to focus on the first and last intersections if we only care about the plane distributions of the whole route rather than the middle links. One thing is noting that more kinds of travel time modes may appear at the same time over the arterials with more complicated route constitutions. For example, six types of travel times are observed at 17:00 from P22 to P19 [see Figure 5(h)], which also suggests the travel times over long arterials are less reliable.

## 4.2 Travel time statistical distributions

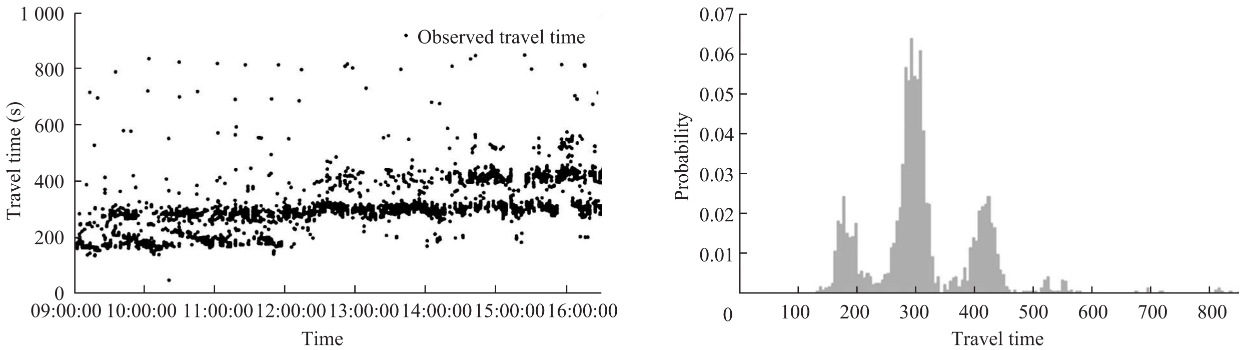
This section aims to explore the nature of $T_t$ statistical distributions, including unimodal or multimodal, skewness $\beta_1$, and $\beta_2$ kurtosis. The 398 datasets are first divided into more sub-datasets with different time windows, including 10, 15, 30, 60, 120, 180, 240, 300, and 360 min. To avoid the influence of small sample sizes, the sub-datasets with sample size less than 100 are eliminated. For each sub-dataset, three shape descriptive statistics of the corresponding distribution are calculated, with the ratios of distributions in different kinds of shapes listed in Table 4.

**Table 4 Ratios of distributions with different shape descriptive statistics**

| Time window (min) | | 10 | 15 | 30 | 60 | 120 | 180 | 240 | 300 | 360 |
|---|---|---|---|---|---|---|---|---|---|---|
| Numbers of distributions | | 880 | 1378 | 1784 | 1481 | 830 | 612 | 280 | 157 | 109 |
| Ratios of distributions (%) | Dip test, $p \leqslant 0.05$ | 82 | 65 | 38 | 18 | 14 | 15 | 16 | 24 | 11 |
| | Dip test, $p > 0.05$ | 18 | 35 | 63 | 82 | 86 | 85 | 84 | 76 | 89 |
| | $\beta_1 \leqslant 0$ | 6 | 4 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| | $\beta_1 > 0$ | 94 | 96 | 98 | 99 | 100 | 100 | 100 | 100 | 100 |
| | $\beta_2 \leqslant 3$ | 13 | 11 | 8 | 6 | 4 | 8 | 3 | 6 | 1 |
| | $\beta_2 3$ | 87 | 89 | 92 | 94 | 96 | 92 | 97 | 94 | 99 |

* The ratios of distributions where $\beta_1$ equals to 0 or $\beta_2$ equals to 3 are zero under each $T_w$.

From the view of peak numbers, the distributions with longer time windows are more likely to have single peaks ($p > 0.05$), while the distributions with shorter time widows are more likely to be multimodal ($p \leqslant 0.05$). For example, the ratios of distributions with more than one peak decrease from 82% to 11% with the increase of time windows from 10 min to 360 minutes, which means a long time window makes the short-term multimodal properties hidden. Such a phenomenon may be the result of non-zero slopes of $T_t$ plane distributions since long time windows allow more uniform appearances of multistate travel times. Under all the time windows, the routes with equal upstream and downstream cycle times (which correspond to zero slopes) are more likely to have multistate statistical distributions. Besides, travel times with zero slopes in $T_t$ plane distributions are more likely to form separated sub statistical distributions, therefore making the overlaps between the sub distributions smaller (see Figure 6).




(a) P11 to P14 (09:00 to 16:30, 130|43-130|46)

(b) P11 to P14 (09:00 to 16:30, 130|43-130|46)

**Fig. 6 An example of distribution with equal upstream and downstream cycle times**

As to skewness and kurtosis, there are two indexes to assess how far a distribution deviates from the symmetrical normal distribution with the same median and standard deviation. Skewness($\beta_2$) is a measure of the asymmetry of a random variable about its mean. For a unimodal distribution, a negative skew indicates that a longer tail is on the left side of the distribution, while a positive skew indicates that the longer tail is on the right. Kurtosis ($\beta_2$) is a measure of the "tailedness" of the distribution of a random variable with higher kurtosis in the case of infrequent extreme deviations (or outliers) and lower kurtosis in the case of frequent modestly sized deviations. It is common to compare the kurtosis of a distribution to the kurtosis of any univariate normal distribution (where the kurtosis is 3). The results in Table 4 demonstrate that the statistical distributions of most sub-datasets are positively skewed ($\beta_1 > 0$) and leptokurtic ($\beta_2 > 3$) under both short and long time windows. Such high ratios suggest that non-normal models may better fit observed travel times in most cases, such as lognormal distribution and Weibull distribution. A scatter plot of skewness and kurtosis under different time windows are depicted in Figure 7. It seems that time windows almost do not correlate with skewness and kurtosis since the values are overlapped. An interesting observation is that the two descriptive statistics are highly correlated, with the Pearson´ correlation coefficient 0.92. Such a high correlation suggests that the two descriptive statistics may be influenced by the same factors, so the high (or the low) skewness and kurtosis coincide.

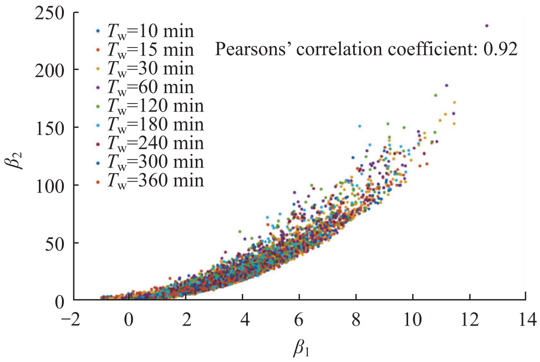



**Fig. 7 Scatter plot of kurtosis and skewness**

To understand the nature of skewness and kurtosis, we depict their cumulative frequency curves under different cases in Figure 8. The factors we considered are route constitutions, measures of central tendency (means, medians, and modes of travel times), and measures of dispersion (interquartile ranges and variances of travel times). Since the skewness highly correlates to the kurtosis, we take the skewness as an example to show the results. As shown in Figure 8(a), the maximum skewness over four or five intersections is less than the maximum ones over two or three intersections. It means that the skewness is more stable in the case of long routes with more intersections. Such stable values may be the result of the lower travel time reliabilities over the long routes, which makes the influence of extreme deviations smaller. As to the measures of means, medians, and modes, we divided these values into six levels [see Figure 8(b), (c) and (d)]. The cumulative frequency curves under each level are very similar in the three subplots, and the ranges of skewness get smaller with the increase of $T_t$ concentration measures. For urban arterials, higher central tendencies of travel times usually occur on the routes with more intersections or highly loaded traffic volumes, which also correspond to low travel time reliabilities. For the quartile deviation and variance, a higher travel time dispersion means a lower travel time reliability, which also results in more stable skewness [see Figure 8(e) and (f)]. Therefore, we can conclude that a lower travel time reliability allows a stable skewness at a relatively lower level, while the fluctuation ranges of the skewness are much wider in the case of higher travel time reliabilities. For kurtosis, the same conclusion can be obtained [see Figure 8(g) to (l)]. Such a phenomenon that both skewness and kurtosis decrease with the increase of travel time variabilities can be interpreted as the limited impacts of the infrequent extreme travel times in the case of low reliabilities.

## 5 Conclusions

The analyses in this study are based on data from Automatic Number Plate Recognition cameras collected from four signalized arterials. We use phase-to-phase travel times between intersections on an arterial roads to explore travel time patterns and statistical distributions.

First, travel time distributions for planes with higher traffic flows and lower green ratios are more likely to exhibit significant data strips. These patterns are influenced by the cycle times of the first upstream and last downstream intersections, and changes in queues.

We speculate that two types of data strips may coexist in the same travel time distribution, though sometimes only one significant type is observable. In most cases, the slopes of significant data strips have the same signs as the differences between downstream and upstream cycles time. Equal downstream and upstream cycles typically result in zero-slope data strips. A longer upstream cycle is more likely to form positive-slope data strips, while a shorter upstream cycle is more likely to create negative-slope data strips. Essenctially, data patterns in travel time plane distributions are closely related to the traffic signals.

Second, the study examines the nature of travel time statistical distributions us the Hartigan dip test along with skewness and kurtosis. We find that departure time windows significantly influence the number of peaks in travel time distributions. A long time window is more likely to result in a single-peak distribution, while a short time window is more likely to lead to a distribution with multiple peaks.

Furthermore, travel time distributions corresponding to zero-slope data strips tend to exhibit separated sub-distributions and demonstrate multistate properties across all time windows. Regarding skewness and kurtosis, the results indicate that most travel time distributions are positively skewed and leptokurtic, with these two values highly correlated. We infer that skewness and kurtosis degrees are related to the travel time reliabilities. In ca-

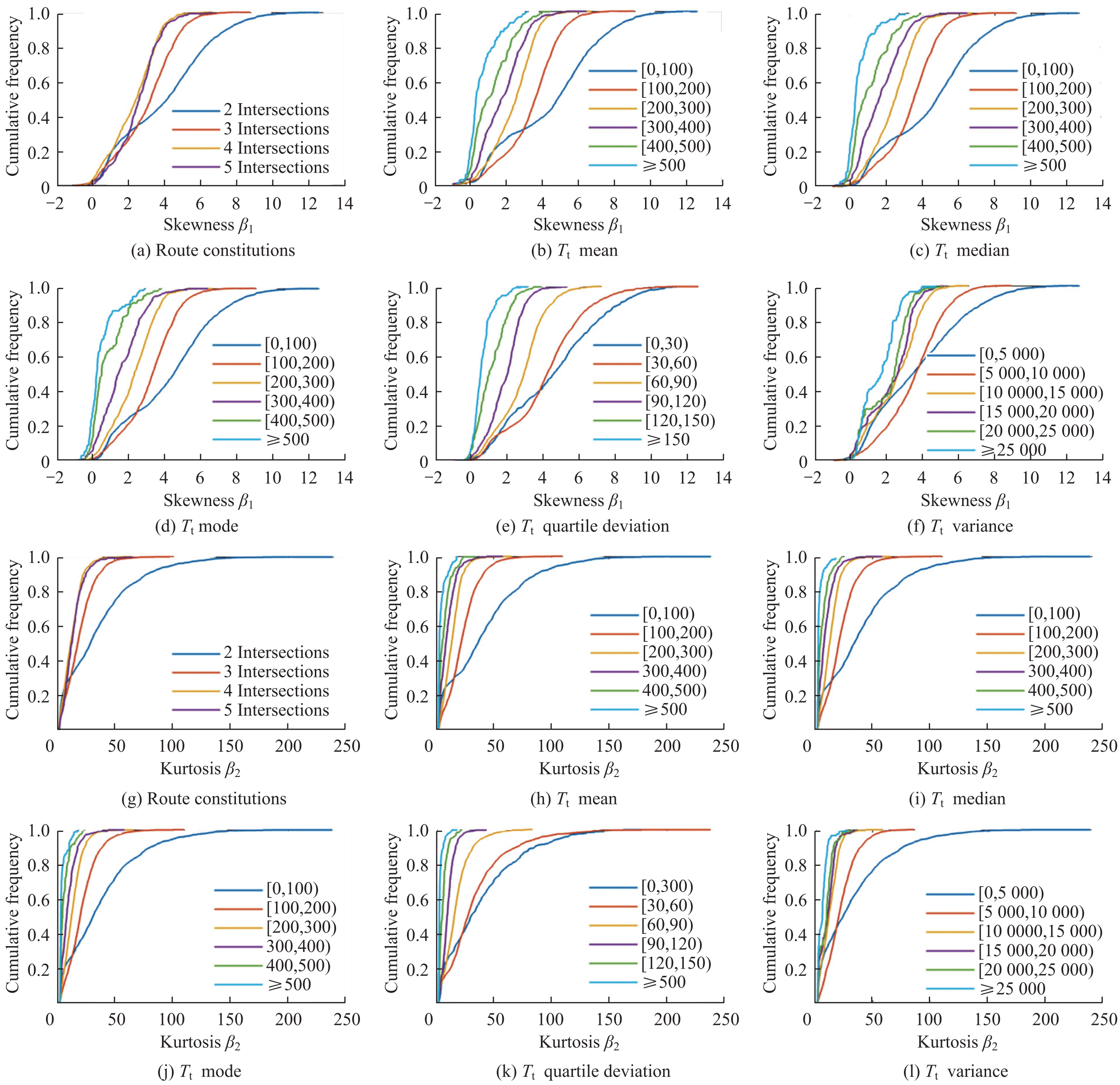



**Fig. 8 Cumulative frequency curves of skewness and kurtosis**

ses of lower reliabilities, infrequent extreme travel times have limited impacts on travel time distributions, keeping skewness and kurtosis relatively stable at lower level. Although this analysis helps us understand the properties of travel time distributions, it remains at a qualitative level. Therefore, our ongoing work is focusing on quantitative analysis from an analytical perspective. For example, we aim to derive an analytical expression for the slope of the data strip while considering offsets and queues simultaneously.

## Acknowledgement

This work was supported by Wuhan Science and Technology Program (2023010188060006, Collaborative Governance between Urban Transport and Urban Spatial Based on Traceability Features of flows on the Composite Corridor). The author would like to thank Wuhan Planning and Design Institute and IITS of Zhejiang University for their support.

## Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

### References

[1] WARDROP J G. Some Theoretical Aspects of Road Traffic Research [J]. Proceedings of the Institution of Civil Engineers, 1952, 1 (3): 325-362.

[2] HERMAN R, LAM T. Trip Time Characteristics of Journeys to and from Work [J]. Transportation and Traffic Theory, 1974, 6: 57-86.

[3] SUSILAWATI, TAYLOR M A P, SOMENAHALLI S. Modelling Urban Travel Time Variability with The Burr Regression Technique [C] // Australasian Transport Research Forum 2011. Adelaide, Australia: ATRF, 2011: 34.

[4] SUSILAWATI, TAYLOR M A P, SOMENAHALLI S. Urban Arterial Road Travel Time Variability Modelling Using Burr Regression [C] // 93rd Transportation Research Board Annual Meeting. Washington DC: TRB, 2014.

[5] SUSILAWATI, TAYLOR M A P, SOMENAHALLI S V C. Distributions of Travel Time Variability on Urban Roads [J]. Journal of Advanced Transportation, 2011, 47 (8): 720-736.

[6] DELHOME R, BILLOT R, EL FAOUZI N E. Travel Time Statistical Modeling with the Halphen Distribution Family [J]. Journal of Intelligent Transportation Systems, 2017, 21 (6): 452-464.

[7] CAO P, MIWA T, MORIKAWA T. Modeling Distribution of Travel Time in Signalized Road Section Using Truncated Distribution [J]. Procedia-Social and Behavioral Sciences, 2014, 138: 137-147.

[8] SRINIVASAN K K, PRAKASH A A, SESHADRI R. Finding Most Reliable Paths on Networks with Correlated and Shifted Log - Normal Travel Times [J]. Transportation Research Part B: Methodological, 2014, 66: 110-128.

[9] YANG F, YUN M P, YANG X G. Travel Time Distribution under Interrupted Flow and Application to Travel Time Reliability [J]. Journal of Transportation Research Record, 2014, 2466 (1): 114-124.

[10] HARTIGAN A J, HARTIGAN P M. The Dip Test of Unimodality [J]. The Annals of Statistics, 1985, 13 (1): 70-84.

[11] TAYLOR M A P, SOMENAHALLI S. Travel Time Reliability and the Bimodal Travel Time Distribution for an Arterial Road [J]. Road and Transport Research, 2010, 19 (4): 37-50.

[12] FENG Y, HOURDOS J, DAVIS G A. Probe Vehicle Based Real-Time Traffic Monitoring on Urban Roadways [J]. Transportation Research Part C: Emerging Technologies, 2014, 40: 160-178.

[13] FU F J, MA D F, WANG D H, et al. An Optimization Method of Time Window Based on Travel Time and Reliability [J]. Mathematical Problems in Engineering, 2015, 2015: 921480.

[14] GUO F, LI Q, RAKHA H. Multistate Travel Time Reliability Models with Skewed Component Distributions [J]. Transportation Research Record, 2012, 2315 (1): 47-53.

[15] YANG Q C, WU G Y, BORIBOONSOMSIN K, et al. Arterial Roadway Travel Time Distribution Estimation and Vehicle Movement Classification Using a Modified Gaussian Mixture Model [C] // 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). The Hague, Netherlands: IEEE, 2013.

[16] LI R M, QIAN X D, WU H B. Gaussian Mixture Model for Urban Link Travel Time Analysis [J]. Journal of Transportation Systems Engineering and Information Technology, 2016, 16 (4): 178-184.

[17] CHEN P, YIN K, SUN J. Application of Finite Mixture of Regression Model with Varying Mixing Probabilities to Estimation of Urban Arterial Travel Times [J]. Transportation Research Record, 2014, 2442 (1): 96-105.

[18] LI R M, ROSE G, SARVI M. Using Automatic Vehicle Identification Data to Gain Insight into Travel Time Variability and its Causes [J]. Transportation Research Record, 2006, 1945 (1): 24-32.

[19] MAZLOUMI E, CURRIE G, ROSE G. Using GPS Data to Gain Insight into Public Transport Travel Time Variability [J]. Journal of Transportation Engineering, 2010, 136 (7): 623-631.

[20] CHAI H J, LI R M, GUO M. Travel Time Distribution and Estimation of Urban Traffic Using Vehicle Identification Data [J]. Journal of Transportation Systems Engineering and Information Technology, 2012, 12 (6): 41-47.

[21] SHI C Y, CHEN B Y, LI Q Q. Estimation of Travel Time Distributions in Urban Road Networks Using Low-frequency Floating Car Data [J]. ISPRS International Journal of Geo-Information, 2017, 6 (8): 253.

[22] GUO J H, LI C G, QIN X, et al. Analyzing Distributions for Travel Time Data Collected Using Radio Frequency Identification Technique in Urban Road Networks [J]. Science China Technological Sciences, 2019, 62 (1): 106-120.

[23] YOUNG S E, SHARIFI E, DAY CHRISTOPHER M, et al. Visualizations of Travel Time Performance Based on Vehicle Reidentification Data [J]. Transportation Research Record, 2017, 2646 (1): 84-92.

[24] WANG D H, FU F J, LUO X Q, et al. Travel Time Estimation Method for Urban Road Based on Traffic Stream Directions [J]. Transportmetrica A-Transport Science, 2016, 12 (6): 479-503.

[25] BAN X G, HERRING R, HAO P, et al. Delay Pattern Estimation for Signalized Intersections Using Sampled Travel Times [J]. Transportation Research Record, 2009, 2130 (1): 109-119.