Received April 26, 2020, accepted May 13, 2020, date of publication May 22, 2020, date of current version June 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2996659

Estimating Travel Times of Mid-Volume Buses Considering Right-of-Way Variation

JUNLAN CHEN1, ZHENJUN ZHU2, XIUCHENG GUO1, YI LI1, AND JUMEI ZHU1

1Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Nanjing 211189, China
2College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing 210037, China

Corresponding author: Xiucheng Guo (seuguo@163.com)

This work was supported by the Scientific Research Foundation for Advanced Talents of Nanjing Forestry University under Grant 163106041.

ABSTRACT Transit agencies often provide estimates of bus travel times to downstream stops. This study aims to improve the perceived reliability of bus transit systems and enhance their competitiveness. This study considers the characteristics of low headway and high demand for mid-volume bus lanes. Considering the variation in right-of-way with respect to both time and space, a stop-based bus route is built to divide the road into sections. Available real-time data from a schedule-based mid-volume bus route are used, including bus global positioning system (GPS) data, road condition information, and weather. Based on the accelerated failure time (AFT) model, a dynamic travel time model considering right-of-way variation is established to estimate bus travel times between adjacent stops and explore the specific impact of bus right-of-way variation on the travel time. The AFT model is chosen because it can reveal the significance of different variables to estimate travel times, and simultaneously estimate expected travel times as well as travel time uncertainty. The experimental results indicate that bus right-of-way variation significantly affects travel times. In contrast to the linear model, the parameter estimated by the AFT model conforms better to expectations, especially for long-distance travel.

INDEX TERMS Travel time, mid-volume bus, accelerated failure time model, GPS data, right-of-way variation.

I. INTRODUCTION

With rapid urbanization and expanding motorization, the increase in household income has led to a rise in private car ownership in China. Residents are increasingly dependent on commuting by private car, which exacerbates traffic congestion and emissions in a metropolis [1], [2]. For instance, in Shanghai, a megacity in China, the urban area has reached 3,640 km², with the population density increasing to 3,814 people per square kilometer, and private car ownership increasing to 616 vehicles per square kilometer in 2017. The number of passengers using public transportation daily has reached 17.96 million, which accounts for 49% of the total travel modes [3]. Therefore, methods need to be developed to relieve traffic congestion, protect the travel rights of low-income people, and implement traffic demand management fairly and productively within the limits of urban public resources [4].

Effective right-of-way allocation policies can improve the efficiency of public transportation and guarantee the reliability of public transport travel times, which can attract more people to travel by public transport [5]. In the public transportation system, mid-volume buses have a higher travel time reliability when the right-of-way is not fully guaranteed, which is in contrast to the subway and regular small-volume buses. Mid-volume buses play an important role in improving the functional hierarchy of the public transportation system, servicing passengers in short and medium distances. The transport capacity and speed of the mid-volume buses is 5–20,000 people per hour and 20–35 km per hour, respectively, with departure intervals reducing to 2.5 min in peak hours. Mid-volume buses are a popular mode of public transport considering passenger corridors, lack of subway coverage, and road conditions. Many mid-volume bus routes have achieved operational benefits in China. For example, the number of passengers using the mid-volume No. 71 bus route in Shanghai, China, has reached 51,000 people daily.

Compared with the subway, the mid-volume bus is more susceptible to traffic factors, weather, and road conditions,
especially the right-of-way variation. Therefore, the reliability of mid-volume bus travel time is a crucial aspect of bus service quality, in which the right-of-way even plays a decisive role. Furthermore, in the process of the promotion of mid-volume buses, some roads cannot guarantee the right-of-way, which is limited by red lines and restrictions (e.g., the change in the number of lanes and the availability of bus lanes). In addition, the right-of-way use time for many bus lanes is limited in the morning and evening rush hours, such as during the hours of 7:00–9:00 and 17:00–19:30. During the non-use time of the bus lane, other vehicles can also pass. It is of considerable significance to estimate the travel time while considering the right-of-way from perspectives of time and space, which can improve the punctuality rate of a bus, attract people to take public transportation, and alleviate traffic congestion.

Various methods have been used to estimate bus travel time. These methods can be classified into four categories: models based on historical data, statistical models (time series, classic regression analysis), machine learning models (artificial neural network, support vector machines), and Kalman filtering models [6].

Models based on historical data assume that the current and future travel patterns depend on historical data, i.e., average travel time and average speed [7]. Existing studies have combined real-time location data from global positioning system (GPS) receivers with historical average travel speed, calculated the weighted mean value from the bus and current real-time speed, and predicted the travel time through the ratio of distance to speed [8]–[10]. Cortés et al. presented a method based on GPS data that systematically analyzed average bus speed [11]. However, the models developed with historical data cannot respond to conditions.

Statistical models such as time series and classic regression models are commonly used to estimate travel times. Time series models assume that the historical traffic patterns will remain the same in the future, and their precision is highly dependent on the correspondence between real-time and historical traffic patterns [12]. Regression models build transparent relationships between travel times and a set of independent variables that can affect travel times [13]–[15]. Patnaik et al. developed a set of regression models to estimate bus travel times using data collected by automatic passenger counters installed in buses [16]. In general, the applicability of the regression models is limited because variables in transportation systems are highly inter-correlated [17].

Machine learning algorithms are a powerful tool for modeling complicated problems in which the relationship between input and output variables is not clear [18], [19]. The artificial neural network (ANN) is a popular approach to solving complex non-linear problems. The ANN models have shown good performance in predicting bus arrival times [20], [21]. The support vector machine is a specific type of statistical learning algorithm, which has been shown to outperform the historical average time-series methods [22], [23]. Julio et al. implemented and compared different machine learning methods to estimate bus travel speeds using real-time information about traffic conditions [24]. Although machine-learning models perform better than linear regression models, they are difficult to interpret and need a large amount of data to train the system [25].

Kalman filtering is an efficient recursive procedure that estimates the future states of dependent variables. It originated from the state-space representations in modern control theory [26]. Chien et al. [12] as well as Cathey and Dailey [27] introduced Kalman filtering methods to estimate bus travel time by using real-time and historical data [12], [27]. Bai et al. applied the Kalman filtering-based dynamic algorithm to adjust bus travel times with the latest bus operation information and estimated baseline travel times [28]. A model based on the Kalman filtering algorithm needs continuous real-time feeds, and data fluctuations might cause difficulties in solving the time lag problem [14].

Besides the estimation methods, the factors affecting the distribution of travel times have been divided into the road, traffic environment, vehicle, and sudden factors [16]. Geometric conditions, route length, number of intermediate stops, and intersections are usually included in road factors. Traffic environment factors are generally expressed by weather and temperature, while the vehicle and sudden factors include speed, arrival time, departure time, travel time, incidents, and accidents. For a bus travel time estimation system, the primary data sources are tracking devices, such as a GPS. GPS-equipped buses provide a trajectory of the buses along their respective routes throughout the day by updating the location information every 5–10 s [17]. The bus transport agency uses the advanced traveler information systems (ATIS). This is one of the main intelligent transportation systems (ITS) that tracks their vehicles and collects bus operation information. By using ATIS, bus transit agencies have increasingly adopted the automatic vehicle location (AVL) and automatic passenger counter (APC) systems [29]. Moreover, information on travel time, arrival time, departure time, and speed along a route are available for these buses. However, no previous study has incorporated the right-of-way constraints into the estimation of bus travel times, such as the change of lane numbers and the availability of bus lanes.

To fill these knowledge gaps, this study used the regression model, considering the right-of-way and other influential factors, to estimate the mid-volume bus travel times. The most significant advantage of the regression model is that it reveals the significance of variables for estimating travel times. In addition, it estimates the expected travel times and travel time uncertainty. Through the analysis of the right-of-way factors, this study provides some recommendations to promote the mid-volume bus.

The remainder of this study is organized as follows. Section 2 describes the data sources and variables used to develop the travel time estimation model and the data processing procedure. Section 3 describes the methodological
modeling frameworks, including the survival model and linear regression model. Section 4 provides the overall estimation results and model evaluation. Finally, Sections 5 and 6 provide discussion and conclusions.

II. DATA

A. DATA SOURCES

Data used for this study were collected from the 17 km mid-volume No. 71 bus route—known as the “ground subway”—that connects Hongqiao Hub, the largest transportation hub, with Bun, the most prosperous business circle in Shanghai, China. Particularly, only the data of the section containing stops 1–5 were used in this study. These stops were monitored from August 26 to September 8, 2018. The selected section covered complex bus right-of-way variations between adjacent stops. For stops 1 to 3, the road has no bus lane, and the number of lanes changes. For stops 3 to 4, only part of the road has a bus lane. For stops 4 to 5, the entire road has a bus lane. The route and selected sections are shown in Fig. 1.

The average daily passenger flow of this line has reached 51,000 people, which ranks first among bus lines in Shanghai. Various factors may affect the uncertainty of travel times, which in turn causes unstable operations. The high mid-volume bus arrival frequency could result in a problem in which the real travel time between adjacent bus stops of a bus in front is unavailable. This is because the headway is often less than the travel time between adjacent bus stops, which means that many buses are running between adjacent stops simultaneously. In this case, it is impossible to use the travel time of the bus in the front. This partially reflects the road condition to estimate the travel time of the bus behind with the missing data. Therefore, the influence of the road conditions should be considered, especially the bus right-of-way variation, such as the bus lane, and the number of lanes and intersections. The methods developed in this study will likely provide more precise estimates when applied to routes with lower demand patterns, lower departure frequencies, and more consistent right-of-way conditions.

The No. 71 bus operates using a schedule-based scheme in which buses are dispatched regularly by schedule. Forty buses were generally in operation during the study periods, with scheduled headways ranging from 2–3 min during high peak departures and 4 min during non-peak departures. All buses in the No. 71 fleet are equipped with AVL systems. In these systems, GPS receivers are usually interfaced with the global system for mobile communications (GSM) modems and placed in the buses. The systems basically record point locations in latitude-longitude pairs, date, time, speed, and direction of the buses. The acquisition frequency of the GPS data is 6–12 times/min, which means that the GPS data is updated every 5–10 s.

B. DATA PROCESSING

The goal of this study is to estimate bus travel times between adjacent stops primarily using real-time GPS data. A total of 2,262,911 observations were available for modeling.
These data were examined and cleaned to eliminate erroneous or potentially inaccurate information. The data processing process is shown in Fig. 2.

1) DATA TRANSFORMING

A stop-based approach considering right-of-way is proposed in this study. There are two major technical challenges that we addressed in our approach. The first challenge is to accurately divide a bus route into sections based on bus stops. Existing data do not have information on the number of times buses go “in” and “out” of bus stops. The second challenge is to reflect the impact of right-of-way on mid-volume travel times. The central data transformation is described below:

**Arrival, departure, and travel time**

Arrival and departure time information from the GPS data were considered for the development of the estimation model. First, by calculating the distance between a bus and bus stop \(i\), observation distances of less than 50 m. The zero velocity observations were then sorted by time. The first zero velocity observation distances of less than 50 m. The zero velocity estimations are no longer reliable, which has a significant impact on operational continuity. The data were screened to reflect the impact of right-of-way on mid-volume travel times. The GPS data correspond to the arrival time, and the last one corresponds to the departure time. The travel time of the bus corresponds to the arrival time, and the last one corresponds to the departure time. The travel time of the bus can be expressed as equation (1). Relevant variable descriptions can be seen in Table 1.

\[
T_i^{\text{trav}} = T_i^{\text{arr}} - T_i^{\text{dep}} \tag{1}
\]

**Right-of-way**

Right-of-way is separated into the two main aspects of time and space, respectively. Between them, the spatial right-of-way of the bus is composed of the right-of-way on the road sections and at the intersections. The following will describe the data transformation of the two aspects.

Above all, to express the impact of the temporal right-of-way of the bus, 0-1 variable \(t\) is defined.

\[
t = \begin{cases} 
1, & \text{bus lane usage time} \\
0, & \text{otherwise} 
\end{cases} \tag{2}
\]

When it comes to the spatial right-of-way of the bus, three key factors should be considered, namely bus lanes, the number of lanes, and intersections. Roads are segmented based on the right-of-way to ensure that a section has the same number of lanes and bus lane setting. In this study, it is assumed that the bus lane usage time is the same and synchronized.

\[
p_{i}^{\text{buslane}} = \sum_{i}^{a} \frac{L_{i,j}}{L_{i}} \ast t \,
\begin{cases} 
\sum_{i}^{a} \frac{L_{i,j}}{L_{i}}, & t = 1 \\
0, & t = 0
\end{cases} \tag{3}
\]

\[
N_{i}^{\text{lane}} = \sum_{i} \frac{L_{i,j}}{L_{i}} \ast [(M - 1) \ast t + 1] + \sum_{i} \frac{a+b}{L_{i}} \ast j,
\begin{cases} 
\sum_{i} \frac{a}{L_{i}} \ast M + \sum_{i} \frac{a+b}{L_{i}} \ast j, & t = 1 \\
\sum_{i} \frac{a}{L_{i}} + \sum_{i} \frac{a+b}{L_{i}} \ast j, & t = 0
\end{cases} \tag{4}
\]

\[
N_{i}^{\text{pointert}} = N_{i}^{\text{inter}} - N_{i}^{\text{priorityinter}} \tag{5}
\]

where \(i\) is the identifier for a bus stop. Between adjacent stops \(i\) and \(i+1\), \(a\) is the number of sections with a bus lane; \(b\) is the number of sections without the bus lane; and \(j\) is the number of lanes other than bus lanes in one section. In addition, \(M\) represents the number of bus lanes in an equivalent non-bus lane when \(t = 1\). This study gives an assumption that \(M\) is large enough for buses to run without disturbance of other vehicles. Specifically, the value of \(M\) is set as 5. All variable descriptions are shown in Table 1.

2) DATA SCREENING

First, some observations have shown incomplete GPS track information, or severe deviation from bus lines. Traffic accidents or GSM system errors may cause these phenomena. Operational interruptions may result from various reasons, including driver rests, driver shifts, mechanical breakdowns, and emergencies. When an interruption occurs, travel time estimations are no longer reliable, which has a significant impact on operational continuity. The data were screened to...
TABLE 1. Variable descriptions (N = 10990).

| Variable                        | Symbol | Description                                                                 |
|---------------------------------|--------|-----------------------------------------------------------------------------|
| Weekday                         |        | Day from Monday to Friday                                                   |
| Peak time                       |        | 10:00 am and 5:00 pm to 7:00 pm                                            |
| Bus ID                          | k      | An identifier for the bus k                                                |
| Arrival time                    | \( T_{i,k}^{\text{dep}} \) | The time that bus \( k \) arrives at the stop \( i \) (s)                  |
| Departure time                  | \( T_{i,k}^{\text{arr}} \) | The time that bus \( k \) leaves from the stop \( i \) (s)                 |
| Travel time                     | \( T_{i,k}^{\text{trav}} \) | Amount of time it takes the bus \( k \) to travel from stop \( i \) to \( i + 1 \) (s) |
| Stop ID                         | \( i \) | An identifier for the bus stop                                              |
| Trip ID                         |        | An identifier for the trip                                                  |
| Section ID                      |        | (Distinguish between two different directions)                              |
| Temperature                     |        | Average temperature during the hour (°C)                                  |
| Humidity                        |        | Average relative humidity during the hour                                  |
| Weather                         |        | Weather during the hour                                                    |
| Percentage of bus lanes         | \( p_i^{\text{bus lane}} \) | Percentage of bus lanes between the stop \( i \) to stop \( i + 1 \)         |
| Length of adjacent stops        | \( L_i \) | Length between the two adjacent stops \( i \) to stop \( i + 1 \) (m)        |
| Number of lanes                 | \( j \) | Number of lanes other than the bus lanes in one direction for a section     |
| Length of section               | \( L_{i,j} \) | Length of the section under the same right-of-way with \( j \) lanes, between the two adjacent bus stops \( i \) to stop \( i + 1 \) (m) |
| Weighted number of lanes        | \( N_{i}^{\text{lane}} \) | The weighted number of lanes between adjacent stops \( i \) to stop \( i + 1 \) |
| Number of intersections         | \( N_{i}^{\text{inter}} \) | The number of total intersections between adjacent stops \( i \) and \( i + 1 \) |
| Number of bus non-priority      | \( N_{i}^{\text{no inter}} \) | Number of intersections where buses have no priority to pass between adjacent stops \( i \) and \( i + 1 \) |
| Number of bus priority          | \( N_{i}^{\text{priority inter}} \) | The number of intersections where a bus has the priority to pass between adjacent stops \( i \) and \( i + 1 \) |

identify any travel times between consecutive bus stops that exceeded 15 min. Second, because bus operation time is from 6:30 am to 11:30 pm, observations of the data generated in other periods were deleted. Finally, the unreasonable short travel times are also deleted based on the maximum speed of the bus, which was 60 km/h, and the distance between adjacent stops.

3) DATA DESCRIPTION

Table 2 shows descriptions of the variables used in this study. Variables can be divided into two categories: weather and peak time. The standard deviation of the travel time is larger than that of other variables, which indicates that many factors influence the bus travel time.

Because the westward and eastward bus stops are set slightly differently, the selected section was numbered based on trips, as shown in Fig. 3, from trip 0 to trip 5. Clustering travel times by period is vital because travel times between stops usually vary over the time of day, resulting in different travel patterns. The travel time of the six trips (trip 0 to trip 5 are drawn in sequence) every day over two weeks is shown in Fig. 4. Fig. 4 shows that the travel time of trip 0, the first trip of every day, is higher (approximately 1000 s) than that of the other four trips (200 s to 300 s on average) as a result of the longer distance between the bus stop 1 and bus stop 3. The travel time distribution over different days of the week seem to be nearly the same. Compared with a weekday, the travel times during a weekend will be relatively flat.

III. METHODS

This section describes two types of models. By using a stop-based model, linear regression models were considered for comparison purposes because this modeling method has been extensively in the literature to estimate bus travel times. The AFT survival models were explicitly proposed for application to the problem at hand.
because they can simultaneously estimate expected travel times and travel time uncertainty [18].

Fig. 5 shows an overview of the proposed approach. First, the inputs required for using the proposed model are real-time GPS data [8], weather data, and road conditions containing bus right-of-way information. Second, through the bus GPS data, including longitude, latitude, and speed information, and the actual stop location information, the bus route is segmented based on the location of the stop and time period segmentation is obtained. Then the dynamic travel time estimation model of the stop-based bus considering right-of-way is established based on the AFT survival model.

The travel time of the bus $k$ between the two adjacent stops $i$ and $i + 1$ includes the running time, delay on the road, and delay at the intersections [13], [30]. There are many reasons for the delay, such as distance, change of lane number,
signal control, and interference from pedestrians and other vehicles. The stop-based model is given as follows:

\[ T_{i,k}^{\text{trav}} = T_{i,k}^{\text{ariv}} - T_{i,k}^{\text{depa}} = T_{i,k}^{\text{running}} + T_{i,k}^{\text{roaddelay}} + T_{i,k}^{\text{interdelay}} \] (6)

To estimate the unknown parameters in (6) considering the right-of-way, the following data transformation, which was mentioned in Section 3, can be used. The length of adjacent stops represents the running time. The percentage of bus lanes and the weighted number of lanes correspond to an additive change in the expected running time. An AFT survival model assumes that an independent variable \( T \) is the time before the occurrence of a specific event. The event is the arrival time at a bus stop after it has traveled to time \( t \). The survival function represents the probability that the event happens after a certain time [32].

Another important function in the survival model is the hazard function \( h(t) \). It is the probability that a bus will arrive at the stop before it has traveled to time \( t \) during a very short interval \( (t, t + \Delta t) \). The hazard function can be expressed as follows:

\[ h(t) = \lim_{\Delta t \to 0} \frac{\text{Prob}(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{1 - F(t)} \] (11)

where \( f(t) \) and \( F(t) \) denote the probability density function and the cumulative distribution function of \( t \), respectively [33].

IV. REGRESSION RESULTS AND MODEL EVALUATION

Eighty percent (80%) of the data were randomly selected for modeling, and 20% of data were used to test the model. Both the linear regression models and survival models were developed using the analysis dataset described in Section 2. The mean absolute error (MAE) and root mean squared error (RMSE) were used as the measures of model performance, which represent the average percentage difference between the observed value (in this case the observed arrival times at a bus stop) and estimated value (in this case the estimated arrival times at a bus stop). In addition, the Akaike information criterion (AIC) is essentially an estimated measure of the quality for each of the models as they relate to one another for a certain set of data, thus making it an ideal method for model selection. More specifically, based on the probabilities for the assumed model, the AIC works to balance the trade-offs between the complexity of a given model and its goodness of fit. For this model, \( k \) represents the number of parameters. In addition, the log-likelihood is a measure of model fit. The higher the number, the better the fit.

\[ MAE = \frac{1}{n} \sum |\hat{y} - \hat{y}^\ast| \] (13)

\[ RMSE = \sqrt{\frac{1}{n} \sum (y - \hat{y})^2} \] (14)

\[ AIC = 2k - 2(\log \text{likelihood}) \] (15)
TABLE 3. Estimated results of The AFT model vs. Linear model.

| Variable                  | AFT model                          | Linear regression model |
|---------------------------|------------------------------------|-------------------------|
|                           | $\chi^2$ | Sig. | AIC   | $\rho^2$ | RMSE(s) | AIC   | $R^2$ | RMSE (s) |
| Percentage of bus lanes   | 4756    | 0.000 | 160.267 | 0.746 | 28 | 238.776 | 0.547 | 52 |
| Length                    | 0.110   | 0.030 | 0.000 | 13.808 | 0.015 | 0.005 | 0.03 |
| Number of weighted lanes  | -3.728  | 0.128 | 0.011 | 6.529 | -0.052 | 0.022 | 0.016 |
| Number of non-priority intersections | -4.516 | 0.351 | 0.000 | 165.734 | 3.738 | 0.340 | 0.000 |

A. REGRESSION RESULTS

The regression results of the AFT and the linear models are shown in Table 3. A smaller MAE for the survival models indicates that the magnitudes of the error terms for the survival models are generally smaller than the error terms for the linear model. The AIC value for the AFT survival model indicates that the AFT survival model fits the data better than the linear model from a probabilistic perspective. The close values of the RMSE for all the models indicate that there are a small number of observations with larger error terms in the survival models than for the linear model. From these observations, it can be concluded that the survival models provide slightly better overall predictions than the linear model and that the performance of the various forms of the survival models are approximately equivalent. Furthermore, the $R^2$ value for the linear model is 0.547, which means that approximately 54.7% of the observed variation can be explained by the model’s inputs. This result shows that the fitting effect of the linear model is not good.

According to the AFT model regression, considering the right-of-way variation, all three variables (percentage of bus lanes, weighted number of lanes, number of non-priority intersections) are significant in influencing the bus travel times. The most significant factor is the percentage of bus lanes followed by the number of non-priority intersections, and the number of weighted lanes. Meteorology, temperature, and humidity have a negative influence on travel time, whereas weekdays and peak times have a positive association with travel time. In contrast to the AFT model, some of the linear model parameter estimates were not consistent with expectations, such as peak time, percentage of bus lanes, the number of weighted lanes, and the number of non-priority intersections.

One of the most important results of the AFT modeling technique is survival probability and hazard function. Fig. 6 illustrates the change in the survival probability and hazard function by increasing the travel time. The survival probability of the AFT model decreases as the travel time grows. The hazard function of the AFT model increases as the travel time increases. Fig. 6 (a) and (b) show that when the travel time is more than 300 s, the survival probability is low, and the hazard function has a sharp increase when the travel time is more than 1,000 s. It is a remarkable fact that the larger the percentage of the bus lanes is, the faster the survival probability decrease and the hazard function increases, as shown in Fig. 6 (c) and (d).

B. MODEL EVALUATION

After the model estimation, the 20% validation dataset (2,198 samples) was used to compare with the estimative capabilities of the linear and survival models. Because the travel time of trip 0 is obviously greater than that of the other four trips, the validation dataset was divided into two groups: trip 0 and trip 1 to 5. Table 4 shows the evaluation of regression models. This study found that the AFT model accurately estimates the travel time due to its low RMSE, MAE, and AIC values. Apart from this, the AFT model performs better than the linear model in regression estimation of the travel time for longer distances. Besides, when the AFT
model accepts missing data, the precisions of estimation will be enhanced a step forward. Therefore, the AFT model could calculate the accumulated ending probability of travel times.

V. DISCUSSION
The survival model was specifically considered because it could estimate the distribution of travel times. Variables such as the percentage of bus lanes, length, number of weighted lanes, number of non-priority intersections, and weekday and peak times were found to be statistically significant at a confidence level of 0.95 in the models, and their parameter estimates were all consistent with expectations. The categories of variable weather, clear, cloudy, and rain, were not found to be significant at a confidence level of 0.60. The possible reason is that, although rainfall and snowfall may have a great impact on bus travel times, the data collection was in the summer, and from August 26 to September 8, 2018, it only rained for a few hours, and there was no snow. In comparison with the survival model, linear regression models were estimated, but they were not efficient. In particular, the parameter estimation of some variables did not conform to expectations.

Variables about right-of-way were more important than other variables in the model. On the one hand, when the headway of the mid-volume bus is very small, and many buses are traveling between two adjacent stops at the same time, the travel time of the bus in front cannot be used as a reference to reflect the road conditions. It is necessary to consider various factors of complicated road conditions separately. On the other hand, bus travel times change with right-of-way variations. Take the top three parameters about right-of-way as an example. The most significant parameter, the percentage of bus lanes, is negatively correlated with travel time and has the highest Wald statistic, suggesting that increasing the proportion of bus lanes can significantly reduce bus travel times. It can be seen from the survival probability in Fig. 6 that, as the percentage of bus lanes increases, the reliability of bus travel times is improved, and the probability of arriving at the bus stop in less time has increased. Similar to the first parameter, the number of weighted lanes is also negative. This indicates that under the circumstance of no bus lane, or the more bus lanes on the road, the bus travel times will be shorter. Finally, there is a positive correlation between bus travel times and the number of non-priority intersections, the third typical parameter, which implies that the more non-priority intersections there are, the longer the travel time is. The uncertainty of bus intersection delays can be reduced by green wave controls at the intersections.

For mid-volume buses with high travel time reliability requirements, it would be helpful if bus temporal and spatial right-of-way can be ensured by establishing bus lanes and increasing dedicated time for bus lanes. However, due to the restrictions of road red line and the limitation of the number of lanes, it is generally impossible to expand the capacity of the road by adding lanes or changing one lane into the bus lane. Therefore, it is suggested that high occupancy vehicle lanes be built for buses and vehicles with multiple passengers. In addition, mid-volume buses are usually on schedule-based routes, and the actual headway is usually less than the scheduled headway, which may contribute to the condition that two buses arrive at the same stops within a very short time. It is suggested that the signal controls at the intersections can be used to adjust the actual headway of buses.

The estimation of bus travel times between two adjacent bus stops uses real-time data and GPS data. By combining the longitude and latitude information of the bus with that of the stops, the arrival and departure time data of the bus can be obtained. Subsequently, the AFT model is used to process the obtained data and provide travel time information at the next stops with minimal error. Apart from this, the static data, right-of-way information, and variables about time, peak time, and weekday reflect the traffic condition of the road. The model will improve the reliability of the public transport system and assist passengers in planning their trips to minimize waiting time, thus attracting more travelers to bus trips and helping relieve traffic congestion.

VI. CONCLUSIONS
This study investigated the problem of estimating mid-volume bus travel times between two adjacent stops with real-time traffic information. The AFT survival model was applied to estimate the bus travel times to the next stop as a function of bus GPS data, right-of-way information, and weather data that would be available in real-time. A stop-based bus route was built to divide the road into sections considering two aspects of right-of-way variation: time and space. The spatial right-of-way of the bus, involving both road sections and intersections, is influenced by three primary factors, namely, the percentage of the bus lane, weighted number of lanes, and prioritized intersections. The travel times between adjacent stops were then divided into running time and road delay. The AFT model was chosen because it reveals the significance of variables in estimating travel times as well as simultaneously estimating expected travel times and travel time uncertainty. Compared with the linear model, the parameter estimation by the AFT model conformed well to expectations, and the AFT model performs better in the regression estimation of travel times for longer distances. The methods developed in this study will likely

### TABLE 4. Evaluation for regression models.

| Form          | Trip 0 | Trip 1 to 5 |
|---------------|--------|-------------|
|               | AIC    | RMSE (s)    | MAE (s)    | AIC    | RMSE (s)    | MAE (s)    |
| Linear model  | 123,476| 95          | 69         | 238,780| 83          | 66         |
| AFT model     | 50,179 | 49          | 28         | 156,145| 54          | 39         |
provide more precise estimations when applied to routes with low demand patterns, low departure frequencies, and more consistent road conditions. The findings of this study could improve the accuracy of estimations of bus travel times and could be used in various ITS applications, such as passenger information systems.

One of the limitations of this study is that the data were collected in August and September when there was no snow and little rain in Shanghai, and the estimation model cannot be extended to other seasons. In the future, the travel time in other seasons can be modeled more accurately by considering the effects of rain and snow. Another limitation is that the bus lane for No. 71 mid-volume bus is used 24 h, so the impact of the bus lane used at different periods on travel time can be further explored. In addition, this study only estimated the mid-volume bus travel times between two adjacent stops. A long-period study might increase the inaccuracy of the results because of fluctuations in traffic volume and accumulation of deviations between the actual and scheduled times. In the future, the study can focus on verifying the uncertainty of the model prediction, the factors affecting the accuracy, and the sensitivity of the prediction.

REFERENCES

[1] Z. Zhu, Z. Li, Y. Liu, H. Chen, and J. Zeng, “The impact of urban characteristics and residents’ income on commuting in China,” Transp. Res. D, Transp. Environ., vol. 57, pp. 474–483, Dec. 2017, doi: 10.1016/j.trd.2017.09.015.

[2] K. Shi, B. Di, K. Zhang, C. Feng, and L. Srivichet, “Detrended cross-correlation analysis of urban traffic congestion and NO2 concentrations in chengdu,” Transp. Res. D, Transp. Environ., vol. 61, pp. 165–173, Jun. 2018, doi: 10.1016/j.trd.2016.12.012.

[3] H. Suzuki and R. Cervero, Transforming Cities With Transit: Transit and Land-Use Integration for Sustainable Urban Development, Washington, DC: World Bank, 2013, doi: 10.1596/978-0-8213-9745-9.

[4] D. Sun, H. Luo, L. Fu, W. Liu, X. Liao, and M. Zhao, “Predicting bus arrival time on the basis of global positioning system data,” Transp. Res. Rec., J. Transp. Res. Board, vol. 2034, no. 1, pp. 62–72, Jun. 2007, doi: 10.3141/2034-08.

[5] J. Strauss and L. F. Miranda-Moreno, “Speed, travel time and delay for intersections and road segments in the montreal network using cyclist smartphone GPS data,” Transp. Res. D, Transp. Environ., vol. 57, pp. 155–171, Dec. 2017, doi: 10.1016/j.trd.2017.09.001.

[6] C. E. Cortés, J. Gibson, A. Gschwendner, M. Munizaga, and M. Zúñiga, “Commercial bus speed diagnosis based on GPS-monitored data,” Transp. Res. C, Emerg. Technol., vol. 19, no. 4, pp. 695–707, Aug. 2011, doi: 10.1016/j.trc.2010.12.008.

[7] S. I.-J. Chien, Y. Ding, and C. Wei, “Dynamic bus arrival time prediction with artificial neural networks,” J. Transp. Eng., vol. 128, no. 5, pp. 429–438, 2002, doi: 10.1061/(ASCE)0733-947X(2002)128:5(429).

[8] Y. Z. Farid, E. Christofa, and L. Paget-Seekins, “Estimation of short-term bus travel time by using low-resolution automated vehicle location data,” Transp. Res. Rec., J. Transp. Res. Board, vol. 2539, no. 1, pp. 113–118, Jan. 2016, doi: 10.3141/2539-13.

[9] B. A. Kumar, L. Vanajakshi, and S. C. Subramanian, “Pattern-based bus travel time prediction under heterogeneous traffic conditions,” Transp. Res. Board, Nat. Res. Council, Washington, DC, USA, Tech. Rep. 14-2513, Jan. 2014, doi: 10.13140/RE.1.2.1338.5448.

[10] X. Hua, W. Wang, Y. Wang, and M. Ren, “Bus arrival time prediction using machine learning in multi-route and multi-data,” Transport, vol. 33, no. 2, pp. 543–554, May 2017, doi: 10.1016/j.trc.2017.12.055.

[11] J. Patnaik, S. Chien, and A. Bladikas, “Estimation of bus arrival times using APC data,” J. Public Transp., vol. 7, no. 1, pp. 1–20, Mar. 2004, doi: 10.5038/2375-0901.7.1.1.

[12] H. Xu and J. Ying, “Bus arrival time prediction with real-time and historic data,” Cluster Comput., vol. 20, no. 4, pp. 3099–3106, Dec. 2017, doi: 10.1007/s10586-017-1066-1.

[13] Z. Yu, J. S. Wood, and V. V. Gayah, “Using survival models to estimate bus travel times and associated uncertainties,” Transp. Res. C, Emerg. Technol., vol. 74, pp. 366–382, Jan. 2017, doi: 10.1016/j.trc.2016.11.013.

[14] S. Khetarpaul, S. K. Gupta, S. Malhotra, and L. V. Subramaniam, “Bus arrival time prediction using a modified amalgamation of fuzzy clustering and neural network on spatio-temporal data,” in Proc. Australas. Database Conf. Cham, Switzerland: Springer, Jun. 2015, pp. 142–154, doi: 10.1007/978-3-319-21613-1_12.

[15] J. Amita, S. S. Jain, and P. K. Garg, “Prediction of bus travel time using ANN: A case study in delhi,” Transp. Res. Procedia, vol. 17, pp. 263–272, 2016, doi: 10.1016/j.trpro.2016.10.091.

[16] J. Ma, J. Chan, G. Ristanoski, S. Rajasegarar, and C. Leckie, “Bus travel time prediction with real-time traffic information,” Transp. Res. C, Emerg. Technol., vol. 105, pp. 536–549, Aug. 2019, doi: 10.1016/j.trc.2019.06.015.

[17] M. Yang, C. Chen, L. Wang, X. Yan, and L. Zhou, “Bus arrival time prediction using support vector machine with genetic algorithm,” Neural Netw. World, vol. 26, no. 3, pp. 205–217, 2016, doi: 10.14311/NNW.2016.26.011.

[18] B. Yu, Z. Z. Yang, and J. Wang, “Bus travel-time prediction based on bus speed,” Proc. Inst. Civil Eng. Transp., vol. 163, no. 1, pp. 3–7, Mar. 2010, doi: 10.1680/tcet.2010.163.1.3.

[19] N. Julio, R. Giesen, and P. Lizzana, “Real-time prediction of bus travel speeds using traffic shockwaves and machine learning algorithms,” Res. Transp. Econ., vol. 59, pp. 250–257, Nov. 2016, doi: 10.1016/j.retrec.2016.07.019.

[20] Z. K. Gurmuru and W. Fan, “Artificial neural network traffic travel time prediction models for buses using only GPS data,” J. Public Transp., vol. 17, no. 4, pp. 45–55, Jun. 2014, doi: 10.5038/2375-0901.17.4.1066.

[21] B. Yu, W. H. K. Lam, and M. L. Tam, “Bus arrival time prediction at bus stop with multiple routes,” Transp. Res. C, Emerg. Technol., vol. 19, no. 6, pp. 1157–1170, Dec. 2011, doi: 10.1016/j.trc.2011.01.003.

[22] F. W. Cathey and D. J. Dailey, “A prescription for transit arrival/departure prediction using automatic vehicle location data,” Transp. Res. C, Emerg. Technol., vol. 11, no. 3–4, pp. 241–264, Jun. 2003, doi: 10.1016/S0924-7963(03)00022-8.

[23] C. Bai, Z.-R. Peng, Q.-C. Lu, and J. Sun, “Dynamic bus travel time prediction models on road with multiple bus routes,” Comput. Intell. Neurosci., vol. 2015, Aug 2015, Art. no. 432389.

[24] H. Chang, D. Park, S. Lee, H. Lee, and S. Baek, “Dynamic multi-interval bus travel time prediction using bus transit data,” Transportmetrica, vol. 6, no. 1, pp. 19–38, Jan. 2010, doi: 10.1080/18128600903292591.

[25] W. Chen, X. Liu, D. Chen, and X. Pan, “Setting headways on a bus route under uncertain conditions,” Sustainability, vol. 11, no. 10, p. 2823, May 2019, doi: 10.3390/su11102823.
JUNLAN CHEN was born in Lu’an, Anhui, China. She received the B.S. degree in traffic engineering from Southeast University, Nanjing, in 2018, where she is currently pursuing the M.S. degree in traffic engineering.

She has authored one article. Her research interests include public transit, transport modeling, and planning. From 2017 to 2018, as the Group Leader, her team’s SRTP project on the abortion experiment of more than 300 people was rated as the National Undergraduate Innovation and Entrepreneurship Project.

ZHENJUN ZHU received the Ph.D. degree in transportation engineering from Southeast University, Nanjing, China, in 2019. From 2017 to 2018, he was a Visiting Ph.D. Student with the Department of City and Regional Planning, University of California at Berkeley, Berkeley. He is currently a Lecturer with the College of Automobile and Traffic Engineering, Nanjing Forestry University. As the Project Leader or a Key Researcher, he has been involved in a number of transportation planning projects in China over the past five years. His research interests include transport modeling and transportation planning. He has authored or coauthored over 20 articles in refereed journals and proceedings.

XIUCHENG GUO was born in Shenzhou, Zhejiang, China, in 1964. He received the B.E. degree in highway and city road specialty and the M.E. and Ph.D. degrees in transportation planning and management from Southeast University, Nanjing, in 1986 and 2002, respectively.

He is currently a Professor with the School of Transportation, Southeast University. He is the author of 30 books and more than 300 articles in refereed journals and proceedings. His research interests include urban transport planning and design, road safety technology, integrated regional transportation planning and management, urban planning and transit operation, traffic information, and traffic control. He is a member of the Urban Transport Planning Academic Committee of China Urban Planning Society and the Deputy Director of the Jiangsu Province Professional Committee of Comprehensive Transportation of the Urban Planning Society.

Yi Li received the B.S. degree in transportation engineering from Southeast University, Nanjing, China, in 2013, where she is currently pursuing the M.S. degree in transportation planning and management. From 2017 to 2019, as the Project Leader or a Key Researcher, she has been involved in a number of transportation planning projects in China. Her research interests include transport modeling and transportation planning.

Ms. Li, as a Key Researcher, she and her team had obtained the Second Prize in the National Competition of Transport Science and Technology for students in China, from 2014 to 2015.

Jumei Zhu is currently pursuing the bachelor’s degree in traffic engineering with Southeast University, Nanjing, China. From 2018 to 2019, as a Key Researcher, her team’s SRTP project was rated as a provincial university student innovation and entrepreneurship project.