Driving Cycle Prediction Model Based on Bus Route Features

Denggao Huang, Hui Xie*, Hongjie Ma and Qiang Sun
State key laboratory of engines, Tianjin University, Tianjin 300072, PR China

Abstract—Bus fuel economy is deeply influenced by the driving cycles, which vary for different route conditions. Buses optimized for a standard driving cycle are not necessarily suitable for actual driving conditions, and, therefore, it is critical to predict the driving cycles based on the route conditions. To conveniently predict representative driving cycles of special bus routes, this paper proposed a prediction model based on bus route features, which supports bus optimization. The relations between 27 inter-station characteristics and bus fuel economy were analyzed. According to the analysis, five inter-station route characteristics were abstracted to represent the bus route features, and four inter-station driving characteristics were abstracted to represent the driving cycle features between bus stations. Inter-station driving characteristic equations were established based on the multiple linear regression, reflecting the linear relationships between the five inter-station route characteristics and the four inter-station driving characteristics. Using kinematic segment classification, a basic driving cycle database was established, including 4704 different transmission matrices. Based on the inter-station driving characteristic equations and the basic driving cycle database, the driving cycle prediction model was developed, generating drive cycles by the iterative Markov chain for the assigned bus lines. The model was finally validated by more than 2 years of acquired data. The experimental results show that the predicted driving cycle is consistent with the historical average velocity profile, and the prediction similarity is 78.69%. The proposed model can be an effective way for the driving cycle prediction of bus routes.

Keywords: Driving cycle prediction, route features, city bus, Markov chain, driving characteristics

1 INTRODUCTION

Buses are environmentally friendly in terms of per capita energy consumption and emissions. They are fairly inexpensive and bring convenience to urban residents. The fuel cost accounts for 60% ~77% of a bus system’s operating costs, depending on the driving conditions (Rohani, 2012). To improve the fuel economy and reduce emissions, the vehicle powertrain design and control strategies are usually optimized by standard driving cycles, such as NEDC (New European Driving Cycle), FTP (Federal Test Procedure), and UDDS (Urban Dynamometer Driving Schedule). The China city bus test cycle is widely applied in bus design, control strategy optimization and regulation tests (Xing and Han et al., 2014). However, the actual driving cycles vary for different route environments, driving habits, traffic conditions and bus types. Many researchers’ results show that there are large differences between the actual driving cycles and standard driving cycles (Hung and Tong et al., 2007; Wang and Huo et al., 2008; Fotouhi and Montazeri-Gh, 2013; Ho and Wong et al., 2014). The fuel economy can be
significantly improved when the design and control of buses are optimized based on the actual driving cycles (Ivanič, 2007; Lee and Adornato et al., 2011). The driving cycle synthesis is important for bus factory calibration before actual operation. The synthesis of bus route driving cycles needs large datasets and is not suitable for extensive promotion. To conveniently predict representative driving cycles of bus routes, some driving cycle prediction methods were proposed based on route geographical information, traffic flow information, route traffic laws and environmental information. Route geographical information is commonly applied to predict driving cycle types, easily extracted from GIS (Geographic Information Systems). André (André and Villanova, 2004) showed that the route geographical typology can account for the driving cycle types of bus lines. Jiménez (Jiménez and Serradilla et al., 2014) proposed a neural network to recognize the driving cycle types of bus lines by macroscopic route features, without requiring an operation driving dataset. The route characteristics, travel time, commercial speed, annual statistics, irregularity of travel, and “urban context” were also used to describe the driving cycle features. Multiple linear regression was utilized to establish the relations between the average speed, 15th percentile speed, 85th percentile speed and route characteristics, and the predicted driving cycle characteristics can also be utilized in the identification of driving cycles (Rui and Lukic, 2011).

The prediction of velocity profiles of bus routes has been attracted attention in recent years. The bus velocity profile along the travel distance can be predicted by the off-line historic model and the on-line AR model (Pan and Song, 2012). The on-line AR model was applied to update the velocity values from the previous vehicle state. The off-line historic model provided the average speed as a function of the travel distance, but many driving experiments were required. To reduce the dependency on the experimental dataset, Minett (Minett and Salomons et al., 2011) proposed a ‘synthetic’ speed profile model based on the route speed limit and the bus acceleration limits, indicating that representative speed profiles can be predicted by the route regulations and vehicle kinematic limits. Valera (Valera and Heriz et al., 2013) proposed a driving cycle prediction model based on an artificial neural network, which processed vehicle velocity measurements and information from modern navigation systems, but a precise driver model and the actual driving scenery were needed. Johannesson (Johannesson and Asbogard et al., 2005) established a driving speed prediction model by Markov chain based only on the route elevation, but the predicted driving speed was not that of a representative driving cycle.

This study aims to establish a bus route driving cycle prediction model to predict the representative driving cycle of bus routes. Because of the bus route features, trip schedule, and residents’ travel habits, the city bus driving cycles are very special. Particularly, the following are four typical features of driving cycles:

a) City buses run on fixed routes and need to stop at each bus station;

b) The running periods, total time and departure timetable are relatively fixed;

c) The drivers are familiar with the traffic infrastructure, and their reaction behaviors tend to stability;

d) The average distance of bus stations is approximately 1 km, and the number of route elements is limited.

Considering these typical features, the relations between the driving cycle characteristics and route features were analyzed by inter-station data segments. Inter-station driving characteristic equations were established by a characteristic analysis between the inter-station driving characteristics and route features. Using kinematic segment classification, a basic driving cycle database was established based on driving data collection. Based on the inter-station driving characteristic equations and the basic driving cycle database, the driving cycle prediction model was finally built, generating drive cycles by the iterative Markov chain for the assigned bus lines. The rest of this paper is organized as follows. In the 2nd section, the research platform and driving data collection are described. In the 3rd section, the whole process of the driving cycle prediction model for bus routes is presented. The evaluation results are presented in the 4th section and discussed in the 5th section. Finally, the paper ends with conclusions in the 6th section.
2 Experimental driving data collection

Two typical bus lines (Line 503 and Line 516) were monitored, as shown in Fig.1, in Tianjin, China. Three road types (urban, suburban, and rapid way) can be found in Line 503. The data measuring experiment was begun in May 2012 and lasted for two years. The operating data were acquired by a bus on-board diagnostic port and a GPS (global position system) module. The acquisition data of Line 503 include 1460 cycles, covering the criteria of seasons, trip time, driving habits, region types and weather. The acquisition data of Line 516 include more than 840 cycles. The total travel distances of Line 503 and Line 516 are 71.4 km and 21 km, respectively. The acquisition data include six buses and 19 drivers, with the fuel consumption, speed, pedal operation, gear, GPS, and vehicle mass, recorded at a frequency of 10 Hz (Ma and Xie et al., 2015). The data of Line 503 were utilized to establish the driving cycle prediction model, and the acquisition data of Line 516 were utilized to validate the model.

The acquisition data were divided by the GPS data of bus station. The dataset architecture of the acquisition data is illustrated in Fig.2. The velocity profiles change with the travel distance from the departing bus stations. Eleven inter-station variables are measured with the travel distance. Each inter-station segment is related to the driver, bus ID, station ID, weather, and trip time. For the bus mass estimation, the raw data filtering criteria and calculation process are illustrated in the reference (Wenjuan and Jing et al., 2015).

Fig.1 Bus routes of Line 503 (red) and Line 516 (yellow)

Fig.2 Acquisition data of each inter-station segment
The inter-station distance, inter-station traffic light numbers, inter-station curve numbers and road types of Line 503 and Line 516 were also collected, as shown in Fig.3. The x axes of subfigures are bus station ID. The road type values 0, 1, and 2 represent the rapid way, suburban, and urban areas, respectively. In Line 503, the 33rd bus station is the demarcation station between the suburban area and rapid way area, and the 54th bus station is the demarcation station between the urban area and rapid way areas. The demarcation stations are determined by the city lay-out and traffic flow surveys. The traffic light numbers and inter-station curve numbers in urban areas are larger than those in the rapid way areas and suburban areas. The average inter-station distance of Line 503 is 1.09 km and that of Line 516 is 0.73 km. The ranges of inter-station distance, inter-station traffic light numbers, and inter-station curve numbers in Line 503 are larger than those of Line 516. Line 503 is thus more complex and provides more information correlating route features and driving characteristics than Line 516.

3 Modeling methodology

3.1 Framework of driving cycle prediction model

The structure chart of driving cycle prediction model is shown in Fig.4. The main parts of the driving cycle prediction model are the inter-station driving characteristic equations, basic driving cycle database and iterative Markov chain. The model inputs are the inter-station route characteristics, and the model output is the driving cycle profile. The inter-station driving characteristic equations are the statistical relations between the inter-station route characteristics and inter-station driving characteristics. The basic driving cycle database contains a group of Markov chain transmission matrices, reflecting the relations between the inter-station driving characteristics and the transmission matrices. The iterative Markov chain is a method for driving cycle profile generation based on the inter-station driving characteristics and the transmission matrices.

The driving cycle prediction process for a bus route is shown as the following steps:

Step 1, divide the bus route into several inter-station segments, and abstract the inter-station route characteristics (route features);
Step 2, predict the inter-station driving characteristics based on the inter-station route characteristics, according to the inter-station driving characteristic equations;

Step 3, search the Markov chain transmission matrices from the basic driving cycle database, indexed by the predicted inter-station driving characteristics;

Step 4, generate inter-station time-velocity profiles based on the transmission matrices for each inter-station segment, constrained by the predicted inter-station driving characteristics;

Step 5, link the inter-station time-velocity profiles to construct a bus route representative driving cycle.

To establish the driving cycle prediction model shown in Fig.4, the construction principles of the inter-station driving characteristic equations, the basic driving cycle database and the iterative Markov chain are illustrated in the following subsections.

3.2 Inter-station driving characteristic equations

3.2.1 Relations between inter-station driving characteristics and route characteristics

The driving cycle characteristics were usually defined from the features of driving habits, velocity profile, traffic conditions, route features, and weather (Ericsson, 2001; Brundell and Ericsson, 2005). Based on the acquisition data of each inter-station segment, 27 inter-station characteristics were defined, as shown in Table 1.

The 27 inter-station characteristics include inter-station driving characteristics, driver habits characteristics, environmental characteristics and inter-station route characteristics. To adequately describe the driving cycle between two bus stations, all inter-station characteristics related to fuel economy should be fully searched in advance. Then, the abstracted inter-station driving characteristics from the inter-station characteristics can be relied.

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Meaning(Unit)</th>
<th>No.</th>
<th>Variable</th>
<th>Meaning(Unit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(v_{s,max})</td>
<td>Maximum velocity (km/h)</td>
<td>15</td>
<td>(AccPed_{m,p})</td>
<td>Mean of acceleration pedal opening rate (%/s)</td>
</tr>
<tr>
<td>2</td>
<td>(v_{s,m})</td>
<td>Mean velocity (km/h)</td>
<td>16</td>
<td>(AccPed_{sd})</td>
<td>Standard deviation of acceleration pedal opening rate (%)</td>
</tr>
<tr>
<td>3</td>
<td>(Pa_a)</td>
<td>Acceleration time ratio(%)</td>
<td>17</td>
<td>(RPA_a)</td>
<td>Positive acceleration factor(m/s²)</td>
</tr>
<tr>
<td>4</td>
<td>(Pd_s)</td>
<td>Deceleration time ratio(%)</td>
<td>18</td>
<td>(SN)</td>
<td>Stop times per kilometer (1/km)</td>
</tr>
<tr>
<td>5</td>
<td>(Pc_r)</td>
<td>Cruise time ratio (%)</td>
<td>19</td>
<td>(AirCnd)</td>
<td>Air conditioning state</td>
</tr>
<tr>
<td>6</td>
<td>(Pi_i)</td>
<td>Idling time ratio (%)</td>
<td>20</td>
<td>(T_r)</td>
<td>Environmental temperature (℃)</td>
</tr>
<tr>
<td>7</td>
<td>(a_{m,p})</td>
<td>Mean positive acceleration (m/s²)</td>
<td>21</td>
<td>(IsWorkday)</td>
<td>Weekend-1 or weekday-0</td>
</tr>
<tr>
<td>8</td>
<td>(a_{m,n})</td>
<td>Mean negative acceleration (m/s²)</td>
<td>22</td>
<td>(Weather)</td>
<td>Weather (sun, cloud, rain, snow)</td>
</tr>
<tr>
<td>9</td>
<td>(G2Per_r)</td>
<td>2nd gear time ratio (%)</td>
<td>23</td>
<td>(\rho_{station})</td>
<td>Density of bus stations (1/km)</td>
</tr>
<tr>
<td>10</td>
<td>(G3Per_r)</td>
<td>3rd gear time ratio (%)</td>
<td>24</td>
<td>(\rho_{RedLight})</td>
<td>Density of traffic lights (1/km)</td>
</tr>
<tr>
<td>11</td>
<td>(G4Per_r)</td>
<td>4th gear time ratio (%)</td>
<td>25</td>
<td>(\rho_{Curve})</td>
<td>Density of curves (1/km)</td>
</tr>
<tr>
<td>12</td>
<td>(G5Per_r)</td>
<td>5th gear time ratio (%)</td>
<td>26</td>
<td>(Ele)</td>
<td>Average grade between bus stations</td>
</tr>
<tr>
<td>13</td>
<td>(M_{Bus})</td>
<td>Bus mass (kg)</td>
<td>27</td>
<td>(RoadType)</td>
<td>Road type between bus stations</td>
</tr>
<tr>
<td>14</td>
<td>(AccPed_{m})</td>
<td>Mean acceleration pedal opening (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The No.1~17 inter-station characteristics were calculated, corresponding to the velocity profile and driver operations in each inter-station segment. The calculation methods of the No.1~17 inter-station characteristics were adapted from previous references (Fukutomoi, 2004; Shiqi and Yafu et al., 2011). \( SN \) was calculated by Formula (1), where \( n_{stop} \) is the number of stop times between two consecutive bus stations, and \( S_{station} \) is the distance between two consecutive bus stations. The air conditioning state \( \text{AirCnd} \) was defined as Formula (2). Three environmental characteristics were defined as \( T_\omega, \text{IsWorkday} \) and \( \text{Weather} \).

\[
SN_j = \frac{n_{stop,j}}{S_{station,j}}
\]

\[
\text{AirCnd} = \{\text{on} - 1, \text{off} - 0\}
\]

The route features have a significant effect on the driving cycles and fuel consumption. The inter-station route features are shown in Fig.5. The bus station density, traffic light number, curve number, elevation and road type are the main features of a bus route. Five inter-station route characteristics were defined for the route features, including \( \rho_{\text{Station}}, \rho_{\text{TrafficLight}}, \rho_{\text{Curve}}, \text{Elev}_{\omega} \) and \( \text{RoadType}_\omega \). The calculation equations of the five inter-station route characteristics are shown in Formula (3)–Formula (6). Where \( j \) is the inter-station serial number, \( N_{\text{TrafficLight}} \) is the number of traffic lights between two bus stations, \( N_{\text{Curve}} \) is the total number of curves between bus stations, and \( \DeltaElev \) is the altitude deviation between two bus stations.

\[
\rho_{\text{Station},j} = \frac{1}{S_{station,j} + S_{station,j-1}}
\]

\[
\rho_{\text{TrafficLight},j} = \frac{N_{\text{TrafficLight},j}}{S_{station,j}}
\]

\[
\rho_{\text{Curve},j} = \frac{N_{\text{Curve},j}}{S_{station,j}}
\]

\[
\text{Elev}_{\omega,j} = \frac{\max(\DeltaElev_j)}{S_{station,j}}
\]

The relations between the inter-station characteristics and fuel economy are complex. To represent the relations, a very complex model involving many parameters is required. However, this research focuses on obtaining a more general understanding to support the abstraction of the inter-station driving characteristics. A multivariate regression between the inter-station characteristics and fuel economy was processed, without interaction terms. Iterative stepwise regression was applied to reduce the multi-collinearity of the regression process. The multivariate regression coefficients were estimated, as shown in Table 2. The regression coefficients were normalized to reflect the sensitivity degree of the inter-station characteristics on fuel economy. \( t \) is the studentized residual of the independent variable, and Sig. is the validation test of the independent variable. When the Sig. value is less than 0.05, the corresponding independent variable is available for the regression analysis. The variance inflation factor (VIF) measures by how much the variance of an estimated regression coefficient is increased because of multi-collinearity (Vittinghoff and Glidden et al., 2011). The regression \( R^2 \) is 0.707, and the regression significance is 0.00. The variance inflation factors of the inter-station characteristics are all less than 5.
The regression equation and dependent variable selection are correct. The standard coefficient of $AirCnd$ is 0.32, which means that the fuel economy of the city bus is deeply influenced by the state of the air conditioning. $RPA_i$, $v_{s,m}$ and $SN$ are the inter-station driving characteristics, and their standard coefficients are greater than 0.1. $AccPed_{sid}$, $G3Per_i$ and $G2Per_i$ represent the driver habits, and their coefficients are lower than the coefficients of the inter-station driving characteristics. $\rho_{\text{Station}}$ and $\rho_{\text{RedLight}}$ represent the route features, but the coefficients are lower than 0.1. $RPA_i$, $v_{s,m}$ and $SN$ are the sorted inter-station driving characteristics, which are closely related to the bus fuel economy. The linear correlation between $v_{s,m}$ and $v_{s,max}$ is 0.834. Considering the fuel economy and driving cycle description, the final four inter-station driving characteristics were extracted to represent driving cycle features between bus stations, which are $RPA_i$, $v_{s,m}$, $SN$ and $v_{s,max}$.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Standardized Coefficients</th>
<th>$t$</th>
<th>Sig.</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$AirCnd$</td>
<td>0.32</td>
<td>70.64</td>
<td>0.00</td>
<td>1.63</td>
</tr>
<tr>
<td>$RPA_i$</td>
<td>0.30</td>
<td>51.38</td>
<td>0.00</td>
<td>2.63</td>
</tr>
<tr>
<td>$v_{s,m}$</td>
<td>-0.29</td>
<td>-44.15</td>
<td>0.00</td>
<td>3.46</td>
</tr>
<tr>
<td>$SN$</td>
<td>0.19</td>
<td>34.39</td>
<td>0.00</td>
<td>2.52</td>
</tr>
<tr>
<td>$AccPed_{sid}$</td>
<td>0.15</td>
<td>32.44</td>
<td>0.00</td>
<td>1.81</td>
</tr>
<tr>
<td>$\rho_{\text{Station}}$</td>
<td>0.05</td>
<td>9.55</td>
<td>0.00</td>
<td>1.82</td>
</tr>
<tr>
<td>$G3Per_i$</td>
<td>0.09</td>
<td>22.15</td>
<td>0.00</td>
<td>1.42</td>
</tr>
<tr>
<td>$Pa_i$</td>
<td>0.08</td>
<td>15.73</td>
<td>0.00</td>
<td>2.08</td>
</tr>
<tr>
<td>$G2Per_i$</td>
<td>0.07</td>
<td>18.02</td>
<td>0.00</td>
<td>1.26</td>
</tr>
<tr>
<td>$M_{bus}$</td>
<td>0.05</td>
<td>11.03</td>
<td>0.00</td>
<td>1.72</td>
</tr>
<tr>
<td>$\rho_{\text{RedLight}}$</td>
<td>0.03</td>
<td>7.73</td>
<td>0.00</td>
<td>1.16</td>
</tr>
</tbody>
</table>

Table 2 Multivariate regression coefficients of bus fuel economy

Both the route features and driver habits have an influence on the inter-station driving characteristics (Lu and Huang et al., 2016). The effect contrast between the driver habits and the route features on the four inter-station driving characteristics is shown in Fig. 6. The x axis is the bus station number in the subplots of Fig. 6, and the lengths of the columns represent the changing range according to the driver habits. The min/max values in each subplot represent the changing range according to the route features. Considering the distributions of $v_{s,m}$, $v_{s,max}$, $RPA_i$ and $SN$ in Fig. 6, the changing ranges according to the station differences are larger than those affected by the driver differences.

The standard deviations of $v_{s,m}$, $v_{s,max}$, $RPA_i$ and $SN$ affected by the driver habits and the route features are compared in Table 3. In contrast with the driver habit differences, the four inter-station driving characteristics are more deeply affected by the route differences. The relations between the driving characteristics and route characteristics should be given priority to be analyzed in the inter-station driving characteristic equations.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Route stations</th>
<th>Driver habits</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{s,m}$</td>
<td>10.99 km/h</td>
<td>4.7 km/h</td>
</tr>
<tr>
<td>$v_{s,max}$</td>
<td>7.30 km/h</td>
<td>5.01 km/h</td>
</tr>
<tr>
<td>$SN$</td>
<td>3.05 1/km</td>
<td>2.75 1/km</td>
</tr>
<tr>
<td>$RPA_i$</td>
<td>1.25 m/s²</td>
<td>1.23 m/s²</td>
</tr>
</tbody>
</table>

Table 3 Standard deviations affected by route features and driver habits
To establish inter-station driving characteristic equations, the Spearman correlations between the inter-station driving characteristics and the route characteristics and the environmental characteristics were analyzed. The Spearman correlation assesses monotonic relationships. If there are no repeated data values, a perfect Spearman correlation of $+1$ or $-1$ occurs when each of the variables is a perfect monotonic function of the other (Vittinghoff and Glidden et al., 2011). The correlation results are shown in Table 4. There are significant correlations between $\rho_{\text{daylight}}$, $\rho_{\text{station}}$, $\text{RoadType}$, $\rho_{\text{curve}}$, $Elev_{T}$ and the four inter-station driving characteristics.

**represents that the correlation significance is greater than 0.01.

3.2.2 Model setup of inter-station driving characteristic equations

Based on the relations between the driving characteristics and route characteristics, linear models, shallow networks and the deep architecture of NNs (neural networks) can be used to construct the relation equations. The
training sample size of the linear model is relatively small, and its analysis results are intuitive. Generalized linear regression analysis was finally used. The architecture of the inter-station driving characteristic equations was described as Formula (7). The statistical relations between the inter-station driving cycle characteristics and the inter-station route characteristics are revealed by the regression coefficients, where $a_{i1} \sim a_{45}$ are parameters.

$$
\begin{bmatrix}
  v_{s,m} \\
  v_{s,\text{max}} \\
  \text{SN} \\
  \text{RPA}_1
\end{bmatrix} =
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\
  a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\
  a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
  a_{41} & a_{42} & a_{43} & a_{44} & a_{45}
\end{bmatrix} \times
\begin{bmatrix}
  \rho_{\text{Station}} \\
  \rho_{\text{RoadLight}} \\
  \rho_{\text{Curve}} \\
  \text{Elev}_{s,m} \\
  \text{RoadType}_s
\end{bmatrix} +
\begin{bmatrix}
  a_{01} \\
  a_{02} \\
  a_{03} \\
  a_{04} \\
  a_{05}
\end{bmatrix}
$$

(7)

The coefficients of Formula (7) can be estimated by the acquired data of Line 503, shown in Table 5. The coefficients ($a_{11}, a_{12}, a_{13}, a_{15}$) are negative, which means that $v_{s,m}$ will decrease when $\rho_{\text{Station}}, \rho_{\text{RoadLight}}, \rho_{\text{Curve}},$ and $\text{RoadType}_s$ increase. The tendency of the maximum velocity $v_{s,\text{max}}$ is almost the same as $v_{s,m}$. The coefficients of $\rho_{\text{Station}}, \rho_{\text{RoadLight}},$ and $\text{RoadType}_s$ for $\text{SN}$ is positive, which means that the buses will stop more frequently in short station sections with high densities of traffic intersections and bus stations. The $\text{RPA}_1$ is positively related to $\rho_{\text{Station}}, \rho_{\text{RoadLight}},$ and $\text{RoadType}_s$, which means that the driving cycle will be more aggressive when the driving conditions are more complicated. The $R^2$ of $v_{s,m}, v_{s,\text{max}}, \text{SN}$ and $\text{RPA}_1$ regression equations are 0.47, 0.55, 0.41 and 0.49, respectively, which means that the equations have moderate correlations. The regression residual distributions of the inter-station driving characteristic equations are shown in Fig.7. The medians of the four residuals are close to zero and located in the center of the boxplots. Based on the validation theory of multiple linear regression, the regression equations can be adapted to predict the inter-station driving cycle characteristics based on the inter-station route characteristics.

**Table 5** Standardized coefficients of the inter-station driving characteristic equations

<table>
<thead>
<tr>
<th>$a_{i1} \sim a_{45}$</th>
<th>$v_{s,m}$</th>
<th>$v_{s,\text{max}}$</th>
<th>$\text{SN}$</th>
<th>$\text{RPA}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{\text{Station}}$</td>
<td>-1.44</td>
<td>-3.66</td>
<td>0.82</td>
<td>0.34</td>
</tr>
<tr>
<td>$\rho_{\text{RoadLight}}$</td>
<td>-1.61</td>
<td>-1.32</td>
<td>0.59</td>
<td>0.15</td>
</tr>
<tr>
<td>$\rho_{\text{Curve}}$</td>
<td>-0.03</td>
<td>-0.50</td>
<td>-0.24</td>
<td>0.00</td>
</tr>
<tr>
<td>$\text{Elev}_{s,m}$</td>
<td>0.18</td>
<td>-0.05</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$\text{RoadType}_s$</td>
<td>-6.88</td>
<td>-2.52</td>
<td>0.79</td>
<td>0.59</td>
</tr>
<tr>
<td>Constance</td>
<td>39.24</td>
<td>58.66</td>
<td>0.81</td>
<td>3.13</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.47</td>
<td>0.55</td>
<td>0.41</td>
<td>0.49</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>6.82</td>
<td>4.54</td>
<td>1.77</td>
<td>0.65</td>
</tr>
<tr>
<td>Sig.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Fig.7 Residual error distribution of statistical regression estimate
3.3 Basic driving cycle database construction

In the previous section, \( v_{s,m}, v_{s,max}, SN \) and \( RPA \), are the main inter-station driving characteristics profiled in the inter-station driving cycle. Similarly, the driving cycle types can also be characterized by the four inter-station driving characteristics. The intervals of \( v_{s,m}, v_{s,max}, SN \) and \( RPA \), can be used for driving cycle categorization, but they are needed to compensate the prediction accuracy of the inter-station driving characteristic equations.

According to Table 3, the standard deviations of \( v_{s,m}, v_{s,max}, SN \) and \( RPA \), were 6.82 km/h, 4.54 km/h, 1.77 1/km and 0.65 m/s², respectively, and the intervals were divided as follows:

a) The range of \( v_{s,m} \) is \([0,60) \) km/h, and its interval are 5 km/h;

b) The range of \( v_{s,max} \) is \([0,70) \) km/h, and its intervals are 10 km/h;

c) The range of \( SN \) is \([0,15) \) 1/km, and its intervals are 2 1/km;

d) The range of \( RPA \) is \([0,8) \) m/s², and its intervals are 1 m/s².

Although the intervals of \( v_{s,m} \) were larger than its standard deviation, the accuracy of the driving cycle classification can be modified by the cooperation of \( v_{s,max}, SN \) and \( RPA \). With the combination of the four driving characteristic intervals, 4704 types of inter-station basic driving cycle can be categorized. The driving cycles can be represented by speed-acceleration transmission matrices (Shuming and Nan et al., 2013), and generated by calculating the transmission matrix with Monte Carlo simulations, which is the core concept of the Markov chain (Gong and Midlam-Mohler et al., 2011; Tae-Kyung and Filipi, 2011). The speed-acceleration transmission matrices of the basic types can be obtained by the kinematic slips matching their discrete driving cycle characteristics, shown in Formula (8).

\[
T_{trans,l} = p\{v_{l+1,j+1} \mid (v_{l,j}, a_{l,j})\} \tag{8}
\]

\( T_{trans,l} \) is the speed-acceleration transmission matrix of the Markov chain for the basic driving cycles. \( v_{l+1,j} \) is the \( j+1 \)th velocity interval of the \( l \)th basic driving cycle, and \( a_{l,j+1} \) is the \( j+1 \)th acceleration interval of the \( l \)th basic driving cycle. 4704 basic driving cycle transmission matrices and their corresponding characteristics were grouped into a basic driving cycle database, shown in Fig. 8. The process of establishing the basic driving cycle database can be divided into three steps. The first step is to sort the basic driving cycle types with the interval combinations of the four driving characteristics; the second step is to extract the bus kinematic slips matching the driving characteristic intervals; and the third step is to calculate the speed-acceleration transmission matrix for each bus kinematic slips.

![Fig.8 Reflection architecture of the basic driving cycle database](image-url)
The basic driving cycle types can be identified using Formula (9). If the predicted inter-station driving characteristics are calculated by the inter-station driving characteristic equations, the driving cycle types can be indexed by the four driving characteristic intervals.

\[ T_{\text{trans}, k} = \{ T_{\text{trans}} | \hat{v}_{s,m,i} \in S_{r,s}, \hat{v}_{r,\text{max},i} \in S_{r_{\text{max},s}}, \hat{S}_{N,s} \in S_{N,s}, R_{\text{PA},i} \in S_{R_{\text{PA},s}} \} \]  

(9)

\( T_{\text{trans}, k} \) represents the speed-acceleration transmission matrix of the \( k \)th basic driving cycle type. \( T_{\text{trans}} \) represents the basic driving cycle database including 4704 basic driving cycle types, which is established by historical data of the same type of city buses. \( \hat{v}_{s,m,i}, \hat{v}_{r,\text{max},i}, \hat{S}_{N,s}, \) and \( R_{\text{PA},i} \) represent the predicted inter-station driving characteristics of the \( i \)th bus station, and \( S_{r,s}, S_{r_{\text{max},s}}, S_{N,s}, \) and \( S_{R_{\text{PA},s}} \) represent the \( k \)th combination of the fourth driving characteristics intervals.

### 3.4 Iterative Markov chain for velocity profile generation

The Iterative Markov chain was used to generate velocity profiles between bus stations based on the speed-acceleration transmission matrix and the kinematic constraints. The transmission matrix can be easily obtained by the inter-station driving characteristic equations and the basic driving cycle database. The state flow of the velocity profile chain generation by the iterative Markov chain is shown in Fig.9. \( \hat{v}_{s,m,i}, \hat{v}_{r,\text{max},i}, \hat{S}_{N,s}, \) and \( R_{\text{PA},i} \) represent the four predicted inter-station driving characteristics and the speed-acceleration transmission matrix, and \( S_{\text{station},i} \) represents the \( i \)th bus station distance. The bus route driving cycle prediction is to generate and combine all of the inter-station velocity profiles. Because of the four typical features of city buses mentioned in the introduction section, the kinematic constraints for the iterative Markov chain are as follows:

1) The predicted velocity profiles between bus stations start and end with 0 m/s;
2) The average and maximum deviations of the predicted velocity profiles are less than \( \lambda \) times of \( \hat{v}_{s,m,i} \) and \( \hat{v}_{r,\text{max},i} \);
3) The initial distance of the predicted velocity profile is appropriate to \( \lambda \) times the station distance. The distance increases by 10 meters when a proper driving cycle is not obtained within 1000 iterations. This operation is intended to reduce the number of iterations.

Based on the iterative Markov chain, the inter-station velocity profile can be automatically generated and represent the features of the speed-acceleration transmission matrices, especially the driving characteristics of average velocity, maximum velocity, and driving distance.

![Fig.9 State flow of velocity profile generation by the iterative Markov chain](image-url)
4 Application of driving cycle prediction model

In summary, the calibration parameters of the driving cycle prediction model based on bus route features are \( a_{0i} \sim a_{4i} \) and \( T_{\text{max},i} \). Based on the acquired data of Line 503, it is easy to estimate the calibration parameters of the driving cycle prediction model. For the extensive promotion validation of the driving cycle prediction model, the route characteristics of Line 516 were used as the inputs, and the acquired data of Line 516 were used as the experimental references. According to the route characteristics of Line 516 in Fig.3, the predicted inter-station driving characteristics are shown in Fig.10. The solid lines are the predicted values, and the dash lines are the experimental values calculated by the acquired data. The predicted values are consistent with the experimental values for Line 516. The average prediction error of \( v_{s,n} \) is 5.40 km/h, that of \( v_{s,\text{max}} \) is 2.82 km/h, that of \( SN \) is 1.33 1/km and that of \( RPA \) is 0.63 m/s². The variation tendencies of \( v_{s,n} \), \( v_{s,\text{max}} \), \( SN \) and \( RPA \) are in compliance with the experimental data. The prediction errors increased at the 25~29th station. The reason is that these stations are in the approach to the suburban area; therefore, their traffic conditions may be mixed.

![Fig.10 Inter-station driving characteristics of predicted and experimental results](image1)

![Fig.11 Predicted driving cycle of Line 516](image2)
In this case, the kinematic constraints $\lambda_\alpha$ and $\lambda_\beta$ are 0.2 and 1, respectively, and the predicted driving cycle of Line 516 is shown in Fig.11. The average velocity of the predicted driving cycle is 19.03 km/h, the maximum velocity is 42.75 km/h, the total duration is 4129 s, and the total distance is 21.89 km. The inter-station distance of the predicted driving cycle and the experimentally measured value is shown in Fig.12. The average absolute deviation of the inter-station distance is 17.01%, and the total distance deviation of Line 516 is only 4.23%.

![Fig.12 Inter-station distance of the predicted driving cycle and actual experimental values](image)

The average velocity profile was calculated by each travel position from 810 cycles of driving data in Line 516. In order to intuitively compare the average velocity profile with the predicted driving cycle, the time-axis (x-axis) of the predicted driving cycle was transferred to travel distance by the integral of velocity. The average velocity profile and the predicted velocity profile were compared, shown in Fig.13. There are some very similar kinematic slips, shown as Arrow (a) ~ Arrow (i). To some degree, the predicted driving cycle profile can represent the historical average velocity profile. Some kinematic slips are different. This may be caused by two reasons. On the one hand, the predicted velocity profile contains random features of the speed-acceleration transmission matrices, which is generated from the Markov chain. On the other hand, the historical average velocity profile is not a real driving profile, but rather a statistical result.

![Fig.13 Comparison between the historical average velocity profile and the predicted driving cycle profiles](image)
The velocity and acceleration distributions are widely used for the evaluation of the driving cycle synthesis and is closely related to the fuel economy of vehicles (Tutuianu and Marotta et al., 2015). When the distributions of the proposed driving cycle are coincide with acquired database, the proposed driving cycle are usually treated as the representative driving cycle. To validate the performance of the predicted driving cycle, the velocity and acceleration distribution comparison between the predicted driving cycle and the experimentally acquired data must still be conducted. The velocity and acceleration distributions were compared, as shown in Fig.14. The velocity and acceleration distributions of the experimental driving cycle were calculated based on 854 cycles of velocity profiles. The velocity and acceleration distributions of the predicted driving cycle basically coincide with those of the experimental data.

![Velocity distribution contrast](image1.png) ![Acceleration distribution contrast](image2.png)

Fig.14 Velocity and acceleration distribution contrasts between the predicted and experimental results.

The integral areas of the velocity and acceleration distribution profiles over the x axis are 1. When the range correlations of the predicted and experimental distribution profiles are the same, the predicted and experimental distribution profiles properly overlap. Therefore, the prediction precision was quantified by the Spearman correlation of the velocity and acceleration distributions. The results of Spearman correlation calculation are shown in Table 6. The Spearman correlation coefficients are 0.81 and 0.95 for the distributions of velocity and acceleration, respectively. $r_{DC}$ is the production of the Spearman correlation coefficients, which is the comprehensive similarity of the velocity and acceleration distributions, quantifying similarity level of the velocity and acceleration distributions. The chi-square test is true, and the $r_{DC}$ of the predicted driving cycle is 78.69%. The distributions of the predicted driving cycle profile are similar to that of acquired data. It can be inferred that the predicted driving cycle profile can be treated as a representative driving cycle of Line 516.

### Table 6 Spearman correlation between predicted driving cycle and experimental results

<table>
<thead>
<tr>
<th></th>
<th>Spearman coefficient</th>
<th>$\chi^2$ (H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity distribution</td>
<td>0.81</td>
<td>-</td>
</tr>
<tr>
<td>Acceleration distribution</td>
<td>0.95</td>
<td>-</td>
</tr>
<tr>
<td>$r_{DC}$</td>
<td>78.69%</td>
<td>0</td>
</tr>
</tbody>
</table>

### 5 Discussion

For Line 503, Line 516 and the China city bus test cycle (*China bus*), the velocity and acceleration distributions are shown in Fig.15. The velocity and acceleration distributions of Line 503 and Liner 516 were calculated based on the acquired dataset. The velocity distribution density of Line 503 is greater than that of Line 515 in the velocity region of [50, 60] km/h. Because of sample size limitations, the velocity distribution density...
profile of the China city bus test cycle is rougher than that of the other two profiles. The velocity distributions of Line 503, Line 516 and Chinabus are different. The acceleration distribution density of Chinabus is more concentrated around the acceleration 0 m/s², but those of Line 503 and Line 516 are relatively diverging.

![Velocity distribution contrast](image1)

![Acceleration distribution contrast](image2)

(a) Velocity distribution contrast of acquired data  
(b) Acceleration distribution contrast of acquired data

Fig.15 Velocity and acceleration distribution contrast between Line 516, Line 503 and China city bus test cycle

The driving cycle similarities $r_{DC,c}$ between Line 516, Line 503 and the China city bus test cycle were calculated, as shown in Fig.16. The driving cycle similarity between Line 516 and the China city bus test cycle is 58.70% and that between Line 503 and the China city bus test cycle is 41%. The driving cycles of bus lines are less similar to the China city bus test driving cycle. The similarity of the predicted driving cycle of Line 516 is 78.69%, and higher than the driving cycle similarity 68% between Line 503 and Line 516. Considered the similarity degree, the driving cycle prediction model is an effective way for the driving cycle prediction of bus lines, based on the bus route features. When the city bus control strategies are optimized by a driving cycle that is similar to actual driving conditions, more fuel savings can be achieved for the city buses (Rui and Lukic, 2011). If the bus manufacturers can obtain the bus route features in advance, the bus control strategy can be optimized based on the proposed prediction model in the factory settings. Furthermore, city buses can be designed with better fuel economy in advance. It’s the benefit of the driving cycle prediction model based on bus route features.

![Driving cycle similarity](image3)

Fig.16 Similarity between Line 516, Line 503 and the China city bus test cycle

The proposed research provides an opportunity to optimize the bus design and control based on the driving cycle prediction model. Some drawbacks of the driving cycle prediction model can be solved with more data
samples. The driving cycle prediction model was established by the acquired data of Line 503. The coefficients between the inter-station route characteristics and inter-station driving characteristics may be changed in different cities, and multivalent linear equations may be insufficient to summarize the relations. However, the selection of the inter-station route characteristics and inter-station driving characteristics can be utilized in other non-linear modelling methods, for example, BP, deep learning and machine learning. When the volume of the acquired bus data increases, the driving cycle prediction model can be made more precise using training data selection. Hence, a novel driving cycle prediction method based only on the bus route features was introduced.

6 Conclusions

A novel approach for driving cycle prediction based on bus route features was proposed. Distinct from the methods mentioned in the introduction section, the proposed model predicted the bus route representative driving cycle based only on the route features. The construction principles of driving cycle prediction model were illustrated, including the inter-station driving characteristic equations, the basic driving cycle database and the iterative Markov chain. Based on the regression analysis, four inter-station driving characteristics and five inter-station route characteristics were abstracted to represent the features of driving cycles and routes. The relations between the inter-station driving characteristics and the inter-station route characteristics were briefly summarized as the inter-station driving characteristic equations. The basic driving cycle database was set up, including 4704 types of velocity-acceleration transmission matrices. The iterative Markov chain was improved to generate velocity profiles between bus stations, and the driving cycle of the bus route was constructed by combinations of the generated velocity profiles.

The performance of the driving cycle prediction model was evaluated using experimental data. Compared with the inter-station distance and inter-station velocity profile, the predicted driving cycle profile can represent the historical average velocity profile to some degree. Through the similarity of the velocity and acceleration distributions, the prediction precision of the predicted driving cycle is 78.69%, better than the similarities between Line 516 and Line 503. The proposed model can be an effective method to predict representative driving cycles of bus routes, supporting bus design, control strategy optimization and perform testing.

Future work will have two parts, model updating and vehicle optimization. For model updating, the authors will continue the extensive data collection. As the measured data accumulated, the statistical parameters and velocity-acceleration transmission matrices of the proposed model will be further improved in different cities. A self-learning strategy is being designed to make the model update process more efficient. For vehicle optimization, the driving cycle prediction model will be applied to a plug-in hybrid electric bus, and the fuel saving results will be evaluated in actual driving conditions.

7 Acknowledgements

This work was supported by the national international cooperation in science and technology special project (grant number 2013DFG62890) and the National Natural Science Foundation of China (grant number 51307119).

References:


Highlights
1. 27 inter-station characteristics are reduced to four inter-station driving characteristics and five inter-station route features, and their relations are summarized by linear equations.
2. A novel model for driving cycle prediction based on bus route features is proposed, including inter-station driving characteristics equations, a basic driving cycle database and an iterative Markov Chain.
3. The predicted driving cycle velocity profile is similar to the historical average velocity profile, and its distributions of velocity and acceleration are coincide with experimental driving data.