Study on the driving cycle construction for city hybrid bus

Jiankun Peng[1,a], Deng Pan[1,b] and Hongwen He[1,c *]

[1] National Engineering Laboratory for Electric Vehicles, Beijing Institute of Technology, Beijing 100081, China

[a]pengjiankun870719@163.com, [b]hdpandeng@163.com, [c]hwhebit@bit.edu.cn

Keywords: driving cycle, kinematic sequences, principal component, cluster analysis

Abstract. The driving date of hybrid bus which can be represent the city driving cycle is fully collected, the original data is divided into several kinematic sequences according to the velocity information, and the city hybrid bus driving cycle is built by using principal component analysis and cluster analysis. The results show that, the correlation coefficient of constructed driving cycle and original dada reaches 97%, the actual operation situation of hybrid bus can be fully reflected, this study has good practical value.

1 Introduction

The emission and fuel economy of city hybrid bus are greatly impacted by city driving cycle [1], the control strategy development of hybrid vehicle is often based on driving cycle as the design basis [2]. The prevailing driving cycles are developed for traditional fuel vehicle, such as the European NEDC cycle, the America New York Bus cycle, the Japan 10-15 cycle and China typical urban cycle [3], but these cycles can not adequately represent the operation characteristics of the hybrid vehicle. Therefore, the development of hybrid city vehicle driving cycle is necessary.

In this paper. The development process and data processing method of driving cycle is described in detail, and the hybrid bus driving cycle of Zhengzhou city can be combined by applying statistical theory of probability distributions, the design basis for developing and researching hybrid city bus which is in line with the characteristics of Zhengzhou city bus can be provided.

2 The original traffic data acquisition

The original vehicle traffic data can reflect urban road traffic condition and vehicle driving characteristics, which is the data source for driving cycle construction. Therefore, the traffic data as the basis for driving cycle, the original traffic data collected more accurate and detailed, then through the analysis of these original data can get the vehicle driving characteristics which are more real and objective, the resulting driving cycle are more representative and pertinence. The two hybrid bus routes of Zhengzhou city bus are chosen for the study objective of traffic acquisition experiment, the experiment last two weeks and collected more than 600 thousands traffic data which is provided the data assurance for the late study.

Table 1 Bus route profile of data acquisition

Route	1	2
One-way distance	15Km	21Km
Stops	27	36
Single run time	2.5~3h	3.5~4h
Acquisition time	7:00~9:00 ; 10:00~12:00 ; 16:00~18:00	7:00~11:00；14:00~18:00

The required measured parameters of experiment include: vehicle velocity, vehicle instantaneous acceleration, road slope. The experiment equipment is OXTS Inertial+. The experiment acquisition system is shown in Fig. 1.

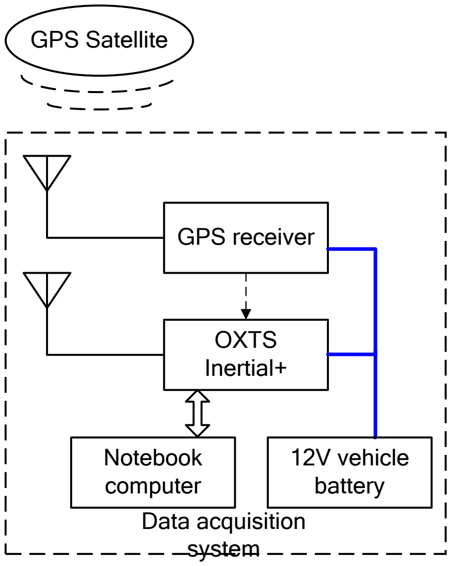


Figure 1 GPS and inertial navigation combined system

OXTS Inertial + measurement accuracy of the parameters are listed in Tab. 2:

Table 2 Measurement accuracy of the parameters

Postion accuracy	Velocity accuracy	Acceleration accuracy			
		Deviation	Linearity	Scale factor	Range
2cm 1σ	0.05 km/hRMS	10 mm/s^2 1σ	0.01%	0.1% 1σ	100 m/s^2

This paper adopts OXTS Inertial+ system to measure vehicle velocity and acceleration. Compare with the traditional method of reading ABS velocity signals, this method can avoid tire pressure and slip on the impact of velocity measurement, improve the measurement accuracy. Therefore, the data sampling frequency is properly increased to retain more details in the velocity sequence. 5Hz sampling frequency was selected in the experiment.

3 Driving cycle construction

The so-called kinematic sequence refers from an idling state to the next idling state, a complete kinematic sequence typically contains idling, acceleration, deceleration and constant speed these four running states.

The steps of kinematic sequences which based on the analysis method of multi-characteristic parameters are as follows: Firstly all kinematic sequences are extracted from the original data and its characteristic parameters are calculated, secondly the characteristic parameters of kinematic sequences are dimension reduced by using principal component analysis, lastly the kinematic sequences which have similar characteristic parameters are clustered into a category through combining with the cluster analysis, and the final driving cycle can be consisted of the several kinematic sequences from the same category with a random selection [4,5].

3.1 Analysis method of kinematic sequences

3.1.1 The kinematic sequences dimidiation and its characteristic parameters selection

According to the definition of kinematic sequences, 2055 sequences are got from original data. Due to the research objective is hybrid bus, the velocity threshold value of idle sequence is taken 0.7km/h, and the threshold value of acceleration is taken 0.15m/s2.

13 parameters are selected for principal component analysis and cluster analysis of kinematics sequences. The parameters are shown in the Tab. 3.

3.1.2 Characteristic parameters matrix of kinematic sequences and its standardization

After the statistical computation for characteristic parameters of each kinematic sequence, fill in the table as follows, and constitute the 2055×13 characteristic parameters matrix of kinematic sequences.

Table 3 The characteristic parameters of kinematic sequences

Parameters	Meaning	Unit
N	The total number of sampling points	
N_a	The sampling points of acceleration sequences	
N_d	The sampling points of deceleration sequences	
N_c	The sampling points of uniform sequences	
N_i	The sampling points of stop sequences	
V_{max}	The maximum velocity	Km/h
V_m	The average velocity	Km/h
V_{sd}	The standard deviation of velocity	Km/h
α_{max}	The maximum acceleration	m/s^2
α_{min}	The minimum acceleration	m/s^2
α_{sd}	The standard deviation of acceleration	m/s^2
α_a	The average acceleration of acceleration sequences	m/s^2
α_d	The average acceleration of deceleration sequences	m/s^2

Table 4 Characteristic parameter matrix

No.	N	N_a	⋯	α_a	α_d
1	185	53	⋯	0.503	-0.885
2	242	60	⋯	0.797	-0.555
3	268	146	⋯	0.524	-0.907
4	29	2	⋯	0.745	-0.189
⋯	⋯	⋯	⋯	⋯	⋯
2054	571	273	⋯	0.530	-0.759
2055	158	69	⋯	0.707	-0.711

The characteristic parameters have the difference of dimension due to the various units, it will result in the dispersion degree of the parameters is high, and the impact on the total variance of an indicator which has a greater variance is larger than any other indicators. Therefore, the characteristic parameters matrix of kinematic sequences should be standardized, namely each column mean is 0, and each column variance is 1.

If set the i-th row, j-th column element of the kinematic sequences parameters matrix is x_{ij}, the element of the corresponding position after normalization is y_{ij}, the normalization process can be expressed as:

$$y_{ij}=\frac{x_{ij}-\overline{x_j}}{S_j} \qquad (1)$$

where, $\overline{x_j}=\frac{1}{n}\sum_{i=1}^{n}x_{ij}$, $S_j=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij}-\overline{x_j}\right)^2}$

Standardized results are shown in Tab. 5.

Table 5 Normalization for characteristic parameter matrix

No.	N	N_a	⋯	α_a	α_d
1	-0.326	-0.314	⋯	0.073	-1.014
2	-0.058	-0.229	⋯	1.158	-0.051
3	0.064	0.809	⋯	0.150	-1.076
4	-1.060	-0.930	⋯	0.966	1.017
⋯	⋯	⋯	⋯	⋯	⋯
2054	1.490	2.343	⋯	0.173	-0.645
2055	-0.453	-0.121	⋯	0.825	-0.505

3.1.3 Principal component analysis

Principal component analysis can transform the original many corresponding indexes into the few independent comprehensive indexes. Principal component analysis can meet in the premise of as much as possible to maintain the original data information, the original variable linear combination transform to a new variable, and new variables independent to each other. Principal component analysis plays a major role of dimensionality reduction and simplification of data structure.

Several principal components can be obtained based on principal component analysis, a new sampling score matrix of principal component which can represent more than 90% of the original sampling information are compressed from the original overall sampling matrix. The new matrix will be used for cluster analysis to classify the overall sampling of kinematic sequences.

Set Σ is the covariance matrix of $X=\left(X_1,\ X_2,\cdots,\ X_p\right)^T$, the characteristic parameter of Σ and corresponding orthogonal unit are $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$ and $e_1,\ e_2,\cdots,\ e_p$, thus the i-th principal component of X is:

$$Y_i = e_i^T X = e_{i1}X_1 + e_{i2}X_2 + \cdots e_{ip}X_p, \left(i=1,2,\cdots,p\right) \quad (2)$$

According to the above equation, the 13 components contribution rates of kinematic sequences characteristic parameter matrix after standardization are obtained.

Seen from the Tab. 6, the cumulative contribution rate of first four main components has reached 94.33%, the first four principal components can be used to compress the original overall sample matrix to a new sample which maintains more than 90% of the original sample information according to Equation 2, then principal component score matrix is obtained. This matrix will be used for cluster analysis.

Table 6 Contribution of principal component

No.	Variance	Contribution	Cumulative contribution
1	8.6229	66.33%	66.33%
2	1.8227	14.02%	80.35%
3	1.1413	8.78%	89.13%
4	0.6753	5.19%	94.33%
5	0.2721	2.09%	96.42%
6	0.2049	1.58%	97.99%
7	0.0866	0.67%	98.66%
8	0.0604	0.46%	99.13%
9	0.0559	0.43%	99.56%
10	0.0318	0.24%	99.80%
11	0.0189	0.15%	99.95%
12	0.0069	0.05%	100.00%
13	-1.804e-16	0.00%	100.00%

3.1.4 The cluster analysis of kinematic sequences

Cluster analysis the method that research individual classification according to the thing itself characteristic. The individual in the same class have greater similarity and difference in the different classes is the principal of cluster analysis. Cluster analysis includes a wide range contents, it can have a variety of methods to do data classification. Dynamic cluster analysis is adopted in this paper.

Dynamic cluster analysis also known as average value cluster method. This method can automatically determine the center location of k group. Calculating the distance of each record to the center location, then adding these records to a group according to the principal of nearest distance. Recalculating the center location of the new group, and repeating the above steps until it reached a certain standard. This method is fast and suitable for large amounts of data classification.

About selecting accumulation point, this paper carried out according to the principle of minimum and maximum. The n samples were divided into k categories, firstly select two farthest distance samples among all samples as the two initial accumulation points, so

$$d\left(x_{11}, x_{12}\right) = d_{11,12} = \max\left\{d_{ij}\right\} \quad (3)$$

Then, select the third accumulation point x13, so as to satisfy the equation 4:

$$\min\left\{d\left(x_{13}, x_{1n}\right), n=1,2\right\} = \max\left\{\min\left[d\left(x_j, x_{1n}\right), n=1,2\right]\right\} \quad (4)$$

And so on, according to the same principal to select until find k initial accumulation pints $L^{(1)}=\{x_{11},x_{12},\cdots,x_{1k}\}$. Calculate the distance among the selected point and other points, and all the sampling points with the initial accumulation point as the center are divided into k disjoint sets based on nearest principal, denoted as $G^{(1)}=\{G_1^{(1)},G_2^{(1)},\cdots,G_k^{(1)}\}$. Then, on the basis of $G^{(1)}$, calculating a new cluster point $L^{(2)}=\{x_{21},x_{22},\cdots,x_{2k}\}$:

$$x_{2i}=\frac{1}{n_i}\sum_{x_l\in G_i^1}x_l, i=1,2,\cdots,k \tag{5}$$

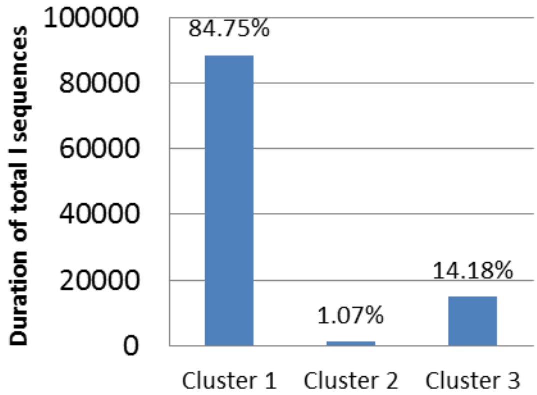
where xi is all sampling points in the current cluster set.

Thus the second round of clustering points can be calculated, a new accumulation point can be regarded as the gravity or the midpoint of its cluster set. Then from L(2), re-devising the sampling points into k disjoint sets, denoted $G^{(2)}=\{G_1^{(2)},G_2^{(2)},\cdots,G_k^{(2)}\}$. Such repeated calculation continues until $L^{(m)}=L^{(m+1)}$.

3.1.4 Driving cycle combination

The classification of cluster sets is set to 3, the kinematic sequences duration and the time proportion of these three cluster sets are shown in Fig. 2.

Figure 2 Duration of total kinematic sequences



The duration of designed driving cycle is defined to 1200s, the duration of the 3 cluster classes are calculated according to the time proportion, then the kinematic sequences are randomly selected from each cluster sets until the corresponding cluster duration meet the requirements. The ultimate fusion driving cycle are shown in Fig. 3.

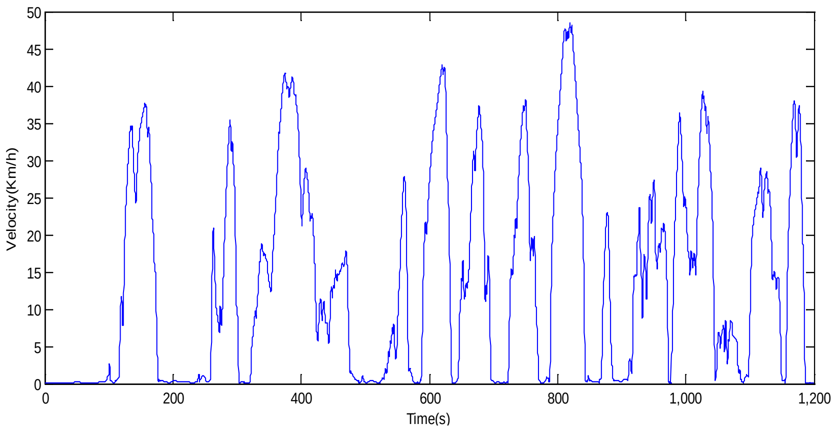


Figure 3 The ultimate fusion driving cycle

To verify the effectiveness of the constructed driving cycle, the characteristic parameters of the driving cycle are compared with the original data, the comparison results are shown in Tab. 7.

Table 7 Comparison of statistical value

Parameters	Original data	Driving cycle
Proportion of stop time	24.09%	32.51%
Proportion of uniform time	19.91%	16.80%
Proportion of acceleration time	30.31%	27.30%
Proportion of deceleration time	25.69%	23.39%
Standard deviation of velocity	13.67	13.742
Maximum velocity	53.762	48.489
average velocity	14.8962	13.688
Minimum acceleration	-5.399	-2.777
Maximum acceleration	4.507	2.459
Standard deviation of acceleration	0.5944	0.542

The correlation coefficient is an intensity metric that describes direct linear relationship between two variables. The calculation expression is expressed as:

$$\rho=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \qquad (6)$$

where $Cov(X,Y)$ is the covariance of variable X and Y, $Var(X)$, $Var(Y)$ are variance of variable X and Y, respectively.

According to Equation 6, the calculated correlation coefficient is 0.97, the constructed road driving cycling with the original experimental data are highly correlated, it indicated that the build cycle is objective and effective.

4 Conclusions

(1) In this paper, several kinematic sequences are obtained by dividing Zhengzhou city hybrid bus driving data, and the driving cycle of hybrid bus for Zhengzhou city are constructed by using principal component and cluster analysis.

(2) The constructed driving cycle compared with the original velocity acquisition data, the proportion of characteristic parameters are the basically same, the correlation coefficient reaches 97%, indicating that the construction method for driving cycle in this paper is correct, the constructed driving cycle of hybrid bus for Zhengzhou City, is true and reliable.

Acknowledgments

This work was supported by the National High Technology Research and Development Program of China (2013BAG05B00) in part, the Program for New Century Excellent Talents in University (NCET-11-0785) and Beijing Institute of Technology Post Graduate Students Innovation Foundation in part. The author would also like to thank the reviewers for their corrections and helpful suggestions.

References

[1] Fontaras G, Pistikopoulos P, Samaras Z. Experimental evaluation of hybrid vehicle fuel economy and pollutant emissions over real-world simulation driving cycles [J]. Atmospheric environment, 2008, 42(18): 4023-4035.

[2] Amjad S, Neelakrishnan S, Rudramoorthy R. Review of design considerations and technological challenges for successful development and deployment of plug-in hybrid electric vehicles [J]. Renewable and Sustainable Energy Reviews, 2010, 14(3): 1104-1110.

[3] Silva C, Ross M, Farias T. Evaluation of energy consumption, emissions and cost of plug-in hybrid vehicles [J]. Energy Conversion and Management, 2009, 50(7): 1635-1643.

[4] Shi Qin, Zheng Yubo, Jiang Ping. A Research on Driving Cycle of City Roads Based on Microtrips [J]. Automotive Engineering, 2011, 33(3): 256-261. (in Chinese)
[5] Hung W T, Tong H Y, Lee C P, et al. Development of a practical driving cycle construction methodology: A case study in Hong Kong[J]. Transportation Research Part D: Transport and Environment, 2007, 12(2): 115-128.