Extraction of typical driving cycles in plateau based on improved short-stroke method

Feng Shi¹, Shichao Cui²±⁴, Han Bu¹, Bingshuo Chen³ and Yulong He³

¹ China Automotive Engineering Research Institute Co., Ltd., 401122 Chongqing, Chongqing, China;
² Chang’an University, 710064 Xi’an, Shaanxi, China;
³ Beijing University of Technology, 100124 Beijing, Beijing, China

Email: 2495829728@qq.com

Abstract. Vehicle driving cycles is a speed time curve used to describe the characteristics of vehicle driving in a specific traffic environment. It can represent the law of speed change in the tested area. The commonly used extraction methods of driving cycles are short stroke method and Markov chain method. However, in the mountainous areas of the plateau, the poor road environment makes the driving cycles very complex. The driving cycles of automobile in plateau area has the characteristics of small proportion of idling condition, and the result of using conventional short stroke method to construct driving cycles is not ideal. In view of this, this paper proposes an improved short stroke method, which uses three parameters of speed, torque and time to divide the kinematic segments. The collected test data is divided into one kinematic segment every 60s. If the value of the 60s torque in the kinematic segment is large (greater than 85% of the total test data), it is divided into the minimum value in the 55-65s interval. Compare the actual driving cycles with the extracted driving cycles, and it is found that the driving cycles constructed by the improved short distance method are consistent with the driving characteristics of vehicles in the plateau area.

1. Introduction

Vehicle driving cycles is a speed time curve used to describe the characteristics of vehicle driving in a specific traffic environment, which can represent the law of speed change in the tested area [1]. The constructed driving cycles are mainly used for the research of vehicle dynamic performance, fuel consumption, pollutant emission control and other aspects.

The relevant committee members of many countries in the world have obtained the vehicle driving cycles curve through a large number of tests, including the vehicle test condition curve and engine condition curve of passenger cars, light commercial vehicles and heavy commercial vehicles, which is of great significance for vehicle energy conservation and emission reduction. At present, the three typical driving cycles widely used in the world are NEDC, FTP75 and Japan10-15[2-5].

In addition, many scholars are committed to the study of urban driving cycles[6, 7]. However, there are few studies on the driving cycles of vehicles in plateau areas. The mountainous environment of plateau makes the driving cycles of vehicles very bad. Compared with the smooth road section, the amount of oxygen pumped into the engine will be significantly reduced when the vehicle is driving on the climbing condition, which makes the mixture in the cylinder relatively thick, and the mixture is not fully burned in the engine combustion chamber, resulting in increased fuel consumption and serious emission pollution.
For the climbing section, the existing driving cycle analysis method can not reflect the actual driving situation. The extraction method proposed in this paper is combined with the actual road slope information, which can better reflect the actual conditions.

2. Data acquisition and preprocessing

2.1. Data acquisition

In order to make the driving cycles of vehicles in the plateau area have typical characteristics, the test road is selected from the south section of G318 Sichuan Tibet highway (Deda Township Haizishan section) in the Qinghai Tibet Plateau. The lowest altitude of the test section is 3610.4m, and the highest altitude is 4591.6m.

In this test, drivers who participated in road test for many times are selected, and the test vehicle is sedan. Drive the test vehicle on the test road through the autonomous driving method. The test vehicle is equipped with an OBD (On Board Diagnostics) instrument, which collects and records the vehicle speed, mileage, time, engine output torque, geographical location and other information in the process of vehicle driving at the frequency of 1Hz.

2.2. Data preprocessing

In the process of data acquisition, there may be some problems such as poor interface contact, insensitivity of sensors to speed perception, etc., which lead to data acquisition or transmission errors. Therefore, the collected data will be preprocessed first, and the filtered data will be constructed. Combined with the principle of driving cycle extraction, the data screening conditions are as follows[8, 9]:

(1) If the vehicle speed is 0km/h for a long time (more than 200s), it is considered that the data acquisition device is separated from the test vehicle and the acquisition data is invalid;

(2) If the acceleration is more than 5m/s², it is considered that the equipment is faulty (the acceleration time of 100 km/h is about 7s, and the vehicle dynamic performance in plateau area will be reduced).

The filtered partial condition data segment is shown in Figure 1.

![Figure 1. Collect data fragments.](image)

3. Extraction of vehicle driving cycles

The common methods of constructing driving cycles of automobile are short stroke method[10], Markov chain method[11] and fixed length intercept method. Among them, the short stroke method divides the kinematic segments according to the idle speed: the kinematic segment analysis method is also called the short stroke method. Because it is easy to understand, this method is widely used at home and abroad to build driving cycles in large cities. Kinematic segment analysis is defined as the...
relationship between speed and time from the beginning of one idle speed to the beginning of the next idle speed, or from the end of one idle speed to the end of the next idle speed. It is characterized by simple principle and convenient operation. For the climbing section, the idle condition of the vehicle is less, so there is a great disadvantage in the extraction of the driving cycle based on the short stroke method - less division of kinematic segments, resulting in the final result is not ideal.

In the process of vehicle climbing, the output torque of engine can well reflect the road slope information. Therefore, this paper selects three parameters of speed, torque and time to divide the kinematic segments.

3.1. Eigenvalue selection
First, preprocess the data collected from the experiment to eliminate the abnormal data. Then, the time and length are divided, and the torque is taken as the reference factor to set a certain amount of relaxation. Take the speed as the analysis index to construct the driving cycle. Because the driving cycles of the vehicle is relatively stable during the climbing process, a longer time is used to divide the kinematic segments. The collected test data is divided into one kinematic segment every 60s. If the torque value in the 60s of the kinematic segment is large (greater than 85% of the total test data), it is divided into the minimum value in the 55th-65s.

In the process of driving in mountainous areas of the plateau, there are mainly acceleration, uniform and deceleration conditions. According to experience, the following settings are made:
- Acceleration condition: when vehicle acceleration \( a > 0.15 \text{m/s}^2 \) and vehicle speed \( v \neq 0 \);
- Deceleration condition: when the vehicle acceleration \( a < -0.15 \text{m/s}^2 \), and the vehicle speed \( v \neq 0 \);
- Uniform speed condition: the condition when vehicle acceleration \(-0.15 \text{m/s}^2 \leq a \leq 0.15 \text{m/s}^2\) and vehicle speed \( v \neq 0 \).

11 parameters, including acceleration time (\( T_a \)), constant speed time (\( T_c \)), deceleration time (\( T_d \)), average speed (\( V_m \)), maximum speed (\( V_{\text{max}} \)), average acceleration (\( A_m \)), maximum acceleration (\( A_{\text{max}} \)), maximum deceleration (\( A_{\text{min}} \)), driving distance (\( S \)), standard deviation of speed (\( V_{sd} \)) and standard deviation of acceleration (\( A_{sd} \)), are selected as characteristic variables.

The eigenvalues of each kinematic segment can be calculated. The eigenvalues of these samples can describe the driving cycles of the vehicle when climbing Plateau area, At the same time, the above formula can be used to calculate the overall characteristic parameters of the test data, and provide effective data for further research on the extraction of vehicle climbing conditions. The eigenvalue matrix of each kinematic segment is shown in Table 1:

3.2. Principal component analysis
The eigenvalues of each kinematic segment are the basis of constructing vehicle driving cycle[12]. The eigenvalues of each kinematic segment can be taken as variables, and each variable can provide certain information of vehicle road driving state. Some of these variables will contain some overlapping information, that is, there is a certain correlation between variables, and they are not independent of each other. For example, acceleration time, uniform speed time and deceleration time have the relationship between these three eigenvalues.
If 11 eigenvalue variables are used to classify the kinematic segments, although they can reflect the driving cycles information in detail, the calculation process will become more complex and it is difficult to analyze the results due to the large number of variables. If one or two eigenvalues are used to evaluate the driving cycles of the vehicle, a lot of effective information will be lost, which means that the constructed condition cannot represent the actual condition. How to determine the number of variables, not only to make the variables reflect more information of vehicle driving cycles, but also not to make the calculation process too complex, which requires the use of principal component analysis.

The eigenvalue matrix which has been divided into 127 kinematic segments is analyzed by principal component analysis.

The contribution rate and cumulative contribution rate of each principal component are shown in Table 2:

| Serial number | characteristic value | Contribution rate (%) | Cumulative contribution rate (%) |
|---------------|----------------------|------------------------|----------------------------------|
| M1            | 6.059                | 55.082                 | 55.082                           |
| M2            | 2.252                | 20.474                 | 75.556                           |
| M3            | 1.259                | 11.444                 | 87.000                           |
| M4            | 0.741                | 6.733                  | 93.732                           |
| M5            | 0.384                | 3.491                  | 97.223                           |
| M6            | 0.167                | 1.515                  | 98.739                           |
| M7            | 0.061                | 0.557                  | 99.295                           |
| M8            | 0.036                | 0.325                  | 99.620                           |
| M9            | 0.033                | 0.300                  | 99.920                           |
| M10           | 0.009                | 0.078                  | 99.998                           |
| M11           | 0.000                | 0.002                  | 100.000                          |

According to the contribution rate and cumulative contribution rate of each principal component in Table 2, three principal components M1, M2, M3 are extracted, and the cumulative contribution rate of the first three principal components has reached 87%, and the authenticity of 13% is discarded. The problem is simplified into three variables, which shows that the principal component analysis method is very successful. The correlation coefficients between the first three principal components and the 11 eigenvalues are shown in Table 3:

|                   | M1     | M2     | M3     |
|-------------------|--------|--------|--------|
| Ta                | 0.883  | -0.051 | 0.359  |
| Tc                | -0.878 | 0.37   | -0.125 |
| Td                | 0.615  | -0.5   | -0.142 |
| vm                | 0.581  | 0.762  | -0.181 |
| vmax              | 0.854  | 0.437  | -0.214 |
| am                | 0.288  | 0.313  | 0.895  |
| amax              | 0.786  | -0.447 | 0.147  |
| amin              | -0.831 | 0.241  | 0.402  |
| S                 | 0.505  | 0.776  | -0.154 |
| vsd               | -0.755 | -0.242 | -0.063 |
| asd               | -0.92  | 0.274  | 0.064  |
The first principal component mainly reflects acceleration time, uniform speed time, deceleration time, maximum speed, maximum acceleration, maximum deceleration, standard deviation of speed and standard deviation of acceleration;

The second principal component mainly reflects the average speed and driving distance;

The third principal component mainly reflects the average acceleration.

The first three principal components represent the information of 11 feature parameters of each kinematic segment, indicating that the first three principal components can represent the overall feature value of the kinematic segment.

The eigenvalues of each kinematic segment are calculated, the main factors are selected by principal component analysis according to the cumulative contribution rate, the K-means clustering algorithm is used to cluster, and the kinematic segments are constructed according to the corresponding proportion.

3.3. K-means clustering
K-means clustering is a simple and efficient clustering algorithm[13]. By calculating the distance between different samples to determine the close relationship between them, the data objects with the same characteristics are classified into the same category, and are distinguished from the data with less correlation. In this paper, based on the distance of different kinds of clustering centers, segments are clustered.

The specific steps of K-means clustering algorithm are as follows:
(1) K sample points are selected from the score matrix of PCA, and K sample points are regarded as K initial clustering centers;
(2) The Euclidean distance from all data points in each sample data set to each initial clustering center point is calculated in turn.
(3) Iterative convergence
Calculate the average value of each cluster, and as a new center point, repeat the above process until the k center points no longer change (convergence), or perform enough iterations to output the partition results.

Figure 2 is the result of clustering. 127 samples are clustered into 3 clusters.

There are 49 fragments in the first cluster, 38 fragments in the second cluster and 40 fragments in the third cluster. According to the distance from each segment to the cluster center, the segment used to construct the driving cycle is selected. The extraction period is 800s and 14 sample segments are selected. The first type of selection fragments 63, 36, 50, 68, 127, 25, the second type of selection fragments 77, 102, 108, 97, the third type of selection fragments 74, 31, 76, 41. According to the selected segments, the driving cycles are constructed in chronological order.
In the process of driving cycle extraction, the points with larger speed may be connected with the points with smaller speed, resulting in abnormal acceleration value. The median value average filtering method is adopted to filter these points: take the abnormal acceleration as the center, select 14 adjacent points before and after, remove the maximum and minimum acceleration of these points, and replace the abnormal acceleration points with the arithmetic average of the remaining points. The extraction conditions are shown in Figure 3.

Figure 3. Improved short stroke method to extract driving cycles I.

3.4. fixed length intercept method

Fixed length intercept method is a commonly used extraction method of driving cycles[12]: from the test data, a certain step is used to cut the data segment in turn as the alternative driving cycles, and the characteristic parameters, such as average speed, are used as the evaluation criteria. The characteristic parameters of each candidate driving cycles are calculated and compared with the experimental data, the driving cycles with the largest correlation coefficient is the representative driving cycles.

According to experience, the construction cycles are set as 800s, and the one with the minimum error between the extraction driving cycles and the actual cycles characteristic value will be selected according to the fixed step intercept method, as shown in the Figure 4.

Figure 4. fixed length intercept method to extract driving cycles II.

3.5. Analysis

Through the case analysis of vehicle driving cycle extraction in plateau area on the test road, the method of combining torque and time length is used to divide the kinematic segment, and the typical
driving cycle is constructed by principal component analysis and K-means clustering algorithm. Compare the actual driving cycles with the extracted driving cycles, as shown in Table 4:

|                  | P<40 | P40~50 | P50~60 | P60~70 | P>70 | Pd       | Pc       | Pa       |
|------------------|------|--------|--------|--------|------|----------|----------|----------|
| Actual driving   | 2.55%| 11.05% | 76.35% | 5.23%  | 4.82%| 37.91%   | 26.97%   | 35.11%   |
| cycles           |      |        |        |        |      |          |          |          |
| Extracted       | 0.33%| 14.52% | 81.85% | 3.30%  | 0.00%| 37.42%   | 27.15%   | 35.43%   |
| driving cycles I|      |        |        |        |      |          |          |          |
| Extracted       | 4.63%| 21.63% | 39.38% | 18.88% | 15.00%| 36.84%   | 31.28%   | 31.88%   |
| driving cycles II|      |        |        |        |      |          |          |          |
| Error value I   | 2.22%| -3.47% | -5.50% | 1.93%  | 4.82%| 0.50%    | -0.18%   | -0.32%   |
| Error value II  |      |        |        |        |      | 36.97%   | -13.65%  | -10.18%  |
|                  |      |        |        |        |      |          |          |          |

Through the comparison, it is found that the driving cycles constructed by the improved short stroke method are consistent with the driving characteristics of vehicles in the plateau area.

4. Conclusions
In this paper, the extraction method for the plateau mountainous road is proposed, and the extraction conditions and actual conditions obtained by principal component analysis and K-means clustering are compared. The ratio of speed and driving cycles of each section can better fit the actual conditions, which shows that the extraction conditions can be better used as the typical driving cycles of this road.

(1) In the process of kinematic segment division, the engine output torque and time length are used as the division basis. When the engine output torque at the division point exceeds a certain threshold (greater than the average value of the overall driving cycle), the minimum value of torque is selected as the actual division point of kinematic segment in the adjacent cell. Using this method;

(2) First, the eigenvalues are determined, then principal component analysis is carried out to reduce the number of variables;

(3) Through the K-means clustering of the principal component results, the kinematic segments are grouped into three categories according to the experience, and the appropriate number of segments is selected to construct the driving cycle.

References
[1] Liu B, Shi Q, Qiu D, et al. 2017 Driving cycle construction based on improved ant colony optimization algorithm and precision analysis J. Journal of Hefei University of Technology (Natural Science Edition) 40(10) 1297-1302
[2] Li L, You S, Yang C, et al. 2016 Driving-behavior-aware stochastic model predictive control for plug-in hybrid electric buses J. Applied Energy 162 868-879
[3] Sciarretta A, Guzzella L 2007 Control of hybrid electric vehicles J. IEEE Control Systems Magazine 27(2) 60-70
[4] Bayindir K. C, Gozukucuk M. A, Teke A. 2011 A comprehensive overview of hybrid electric vehicle: Powertrain configurations, powertrain control techniques and electronic control units J. Energy Conversion and Management 52(2) 1305-1313
[5] He W, Xiang C, Zhang D, et al. 2015 J. Modelling and control of a two-mode power-split hybrid powertrain. International Journal of Electric and Hybrid Vehicles 7(2) 139-158
[6] Liu Y, Xia H, Yao Y, et al. 2018 J. A method of vehicle driving cycle development based on combined principal component analysis and fuzzy C means clustering. Highway Transportation Technology
[7] Li Y, Ren T, Shao P, et al. 2019 Development of driving cycle of bus in Xi'an city based on
Markov chain. *J. China Sciencepaper* 14(02) 121-128

[8] Li N. 2016 *D.* The construction and study of vehicle driving cycle based on urban road. *Hebei Agricultural University.*

[9] Wang N 2012 *D.* Construction of the urban road driving cycle and research of the fuel consumption. *Journal of Hefei University of Technology*

[10] Ren T 2019 *D.* Research on development of typical driving cycles based on markov chain and application research of driving. *Chang’an University*

[11] Hu Z 2016 *D.* Recognition and Sorting of Workpiece Based on Pattern Recognition and Machine Vision. *Kunming University of science and technology*

[12] Cao D 2016 *D.* Study on the algorithm of city driving cycle construction. *Beijing University of Technology*

[13] Liu Z, Zhu P, Liu X, et al. 2019 *J* Research on improved k-means and driving cycle construction. *Automotive Technology* 11 57-62