Research Article

Development of a Driving Cycle for Fuzhou Using *K*-Means and AMPSO

Minrui Zhao ⓘ**, Hongni Gao, Qi Han** ⓘ**, Jiaang Ge, Wei Wang** ⓘ**, and Jue Qu**

Air and Missile Defense College, Air Force Engineering University, Xi'an 710038, China

Correspondence should be addressed to Wei Wang; b6135958qincina@163.com

The driving cycle is a speed-to-time curve, a fundamental technique in the automotive industry, and also a basis to set standards for fuel consumption and emissions of vehicles. A driving cycle is developed based on firsthand driving data collected from fieldwork. First, bad data in the original dataset are preprocessed, the time-series standard smoothing algorithm is used to smoothen the data, and Lagrange's interpolation is used to realize data interpolation. Next, the rules for kinematic fragment extraction are set to divide the data into kinematic fragments. Last, an evaluation system of kinematic fragment feature parameters is built. On that basis, the *K*-means clustering method is used to cluster the dimensionally reduced data, and the adaptive mutation particle swarm optimization (AMPSO) algorithm is employed to select the optimal fragments from candidate fragments to develop a driving cycle. The experiment result shows that the developed driving cycle can represent the kinematic features of the experiment car and provides a basis for the development of a driving cycle for Fuzhou.

1. Introduction

Energy consumption and exhaust emissions are two major issues the automotive industry needs to address [1]. Exhaust emissions account for a major cause for air pollution and have drawn worldwide attention [1–3]. A car driving cycle is a curve representing the speed of a vehicle versus time; it is a common fundamental technique that the automotive industry employs to simulate actual traffic conditions [4], assess energy economy [5], and exhaust emissions [6]. Currently, typical driving cycles in the world include the Worldwide Light-duty Test Cycle (WLTC) [7], the Federal Test Procedure (FTP), the New European Driving Cycle (NEDC), and the Japanese JC08 Cycle.

NEDC is currently used in China to assess energy economy and emission levels, but it is a modal cycle constructed based on ideal conditions that are unachievable in reality [9], consisting of fragments of constant-speed modes, constant acceleration modes, and constant deceleration modes [10]. Figure 1 shows the NEDC. The emission levels and fuel consumption data obtained by using this ideal cycle will be far removed from reality. André et al. [11] found that

to use one single cycle in different driving conditions would lead to inaccurate assessment results. Research by Lin and Niemeier [12] revealed that regional driving differences would result in significant driving cycle differences. Therefore, using one single driving cycle for emission estimation would be defective. Therefore, in order to construct a driving cycle that accords with a city's actual conditions, it is necessary to base the research on the actual driving data of that specific city. Many existing studies on driving cycles are based on local driving conditions. Ho et al. [13] developed a Singapore driving cycle (SDC) for roads and traffic conditions in central business districts of Singapore. Fotouhi and Montazeri-Gh [14] developed a driving cycle for Tehran using *K*-means clustering; other studies have developed driving cycles for different regions such as Toronto waterfront area [15], Chennai in India [16], Hamburg [17], Tianjin [18], Edinburgh, and Abu Dhabi [19].

Two major methods are currently used for the development of driving cycles. The first is the clustering method [20], which divides the original data into kinematic fragments (microtrips), clusters these fragments, selects fragments with the least deviation from the general condition,
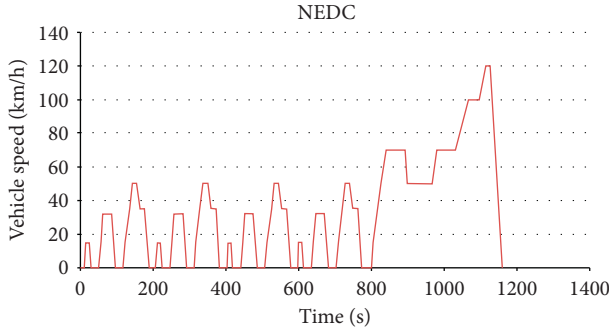
FIGURE 1: New European driving cycle [8].

and combines these fragments into a driving cycle [21, 22]. Typical clustering methods for driving cycle development include *K*-means clustering [23], hierarchical clustering [24], and fuzzy clustering [25]. The central step in clustering is the selection of proper microtrip fragments. Clustering requires less computation than random selection methods, but the determination of the value of *k* (the number of clusters) relies on previous experience. The other driving cycle development method is the Markov chain-based vehicle speed prediction method [26]. This method considers the vehicle speed changes as a random process and builds a vehicle state transfer matrix to predict the vehicle's speed. However, this method entails large amounts of data, the accuracy of prediction relies on the number of iterations of the Markov algorithm [27, 28], and the kinematic fragments lack continuality [29].

To develop a car driving cycle for Fuzhou, we analyzed the local driving features using the clustering method. We then constructed a microtrip fragment selection method that would identify fragments with the least deviation from the average feature parameters and selected the kinematic fragments using the adaptive mutation particle swarm optimization (AMPSO). We developed a typical car driving cycle based on driving data collected in Fuzhou. The statistical features of the driving cycle were also analyzed and compared to other cycles. The algorithm used in this study can provide a reference for driving cycle development, and the research results will provide a basis for the design of vehicle estimation standards.

2. Methods

The sample data used in this study are obtained from field surveys on roads in Fuzhou by a university to research on the driving cycle of lightweight cars. Thus, the sample data can reflect the actual road and traffic conditions in Fuzhou. The dataset was used in the 2019 China Postgraduate Mathematical Modeling Competition, the top competition in mathematical modeling in China. Therefore, we believe that this dataset is reliable and can reflect the real situation of the road. The sample consists of three data files with data collected by the same car at different time periods. The sampling frequency is 1 Hz, and the three files contain 185,726, 145,826, and 164,915 pieces of data, respectively. The collected data include the time, GPS vehicle speed, longitude

and latitude, engine speed, engine torque percentage, and instantaneous fuel consumption. In order to develop a vehicle driving cycle, only the GPS speed is analyzed in this study. Other information including the acceleration and driving distance is obtained by differential operation and integral operation. Because of minor problems of the sampling device and the vehicle, there are errors in the collected data. When the sample data are processed, the following five types of data will be removed: first, discontinuous data caused by the loss of vehicle speed data; second, data of abnormal acceleration and deceleration; third, data collected when the vehicle is under a long-time halt state; fourth, abnormal data collected when the vehicle moves at a low speed for a long time; and last, data collected during an idling event longer than 180 s. Figure 2 shows the steps of the driving cycle's construction.

2.1. Hypotheses

(1) We assume all the sampled data, except the aforementioned five types of abnormal or invalid data, are valid

(2) If the vehicle moves at a speed below 10 km/h in an on-and-off state for over 30 s, it is considered that there is a traffic congestion, and this situation is regarded as an idling event

(3) Data on a speed above 180 km/h are considered as abnormal data (i.e., the speed should not be 150% higher than the set maximum)

(4) If a data segment remains incomplete after data interpolation, the segment of data till the next starting time point of an idling event will be removed and not included in the kinematic fragments (microtrips)

2.2. Data Preprocessing.
GPS speed data are used as an indicator for driving cycle development in this study. The initial value of each data file at the starting time is set as 1. Figure 3 shows part of the speed-time chart obtained by the preprocessed data.

Some bad data still exist in the preprocessed data files, such as discontinuous data, abnormal acceleration/deceleration, long-time halt, long-time traffic congestion, and long-time idling events.

These bad data can be classified into two groups: abnormal data and missing data.

(1) Abnormal data: this group includes data of long-time idling events and data of abnormal acceleration/ deceleration

(2) Missing data: this group includes short-time missing data and long-time missing data

With regard to data of abnormal acceleration and deceleration in the first group, the range of acceleration/deceleration is set as $[-8 \, m/s^2, \, 4 \, m/s^2]$ according to the Specifications of Acceleration and Deceleration for Coach
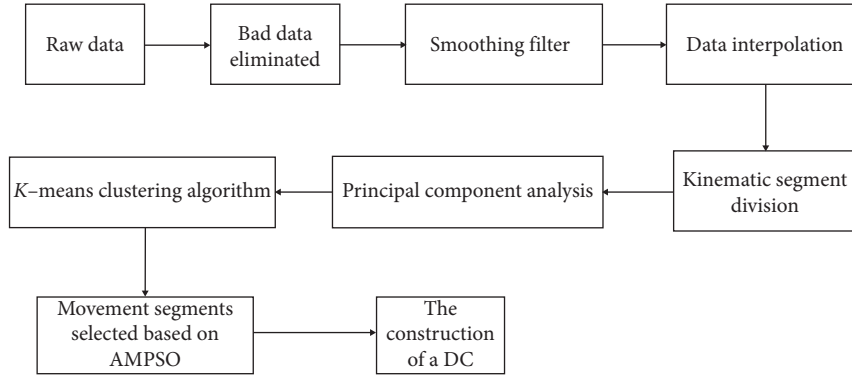
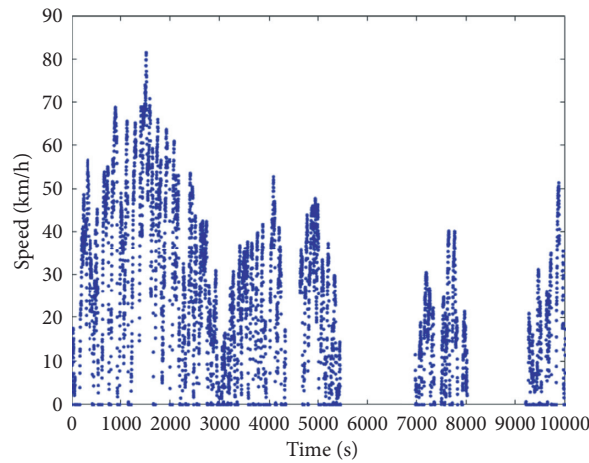FIGURE 2: Driving cycle construction flowchart.

FIGURE 3: Part of the speed-time chart.

[30]. Data on idling events that last longer than 180 s are deleted.

With regard to short-time missing data that are few in the data files, interpolation is performed to fill the gap between data that have an interval less than 3 s and nonzero values on both sides of the interval (given idling situations). The data set with lots of continuous missing data are deleted.

Data of idling events are divided into the following two groups: long-time traffic congestion is considered as an idling event (speed set as 0), and data of long-time idling events are deleted. According to the definition of long-time congestion, data of driving at a speed lower than 10 km/h for no less than 30 s are searched and reset as 0.

The three data files after data preprocessing have 178,753, 143,535 and 159,777 pieces of data.

2.3. Kinematic Fragment Selection Methods.
A kinematic fragment (also called "microtrip") is an excursion between two successive time points at which the vehicle is stopped, i.e., a trip between two idling events. A complete kinematic fragment includes the four fragments: an idle fragment, an acceleration fragment, a cruise fragment, and a deceleration fragment. According to the standards of the WLTC, the following fragment extraction rules are set:

(1) One fragment should last more than 10 s

(2) The idle time should be less than 180 s

(3) For acceleration, $a \leq 4\,m/s^2$; and for deceleration, $a \geq -8\,m/s^2$

(4) The maximum speed should be above 10 km/h

In accordance with the four rules, the preprocessed driving data are segmented into kinematic fragments, as shown in Figure 4. The numbers of fragments for the three sample data files are 1476, 1020, and 1224, respectively.

2.4. Extraction and Computation of Feature Parameters.
The major objective of this study is to develop a vehicle driving cycle with a 1200–1300 s speed time series. The vehicle driving features reflected by the cycle should be representative of corresponding features of the source data. As shown in Table 1, 16 feature parameters are chosen in this study. We choose these parameters based on the common practice of DC constructing.

In Table 1, parameters of speed and time can be collected directly, and the collection frequency is 1 Hz. For any given time point i, $t_{i+1} - t_i = 1$, so the value of acceleration can be obtained by using the following equation:

$$a_{i,i+1} = \frac{v_{i+1} - v_i}{t_{i+1} - t_i} \times \frac{1000}{3600} = \frac{v_{i+1} - v_i}{3.6}, \quad i = 1, 2, \ldots, n, \quad (1)$$
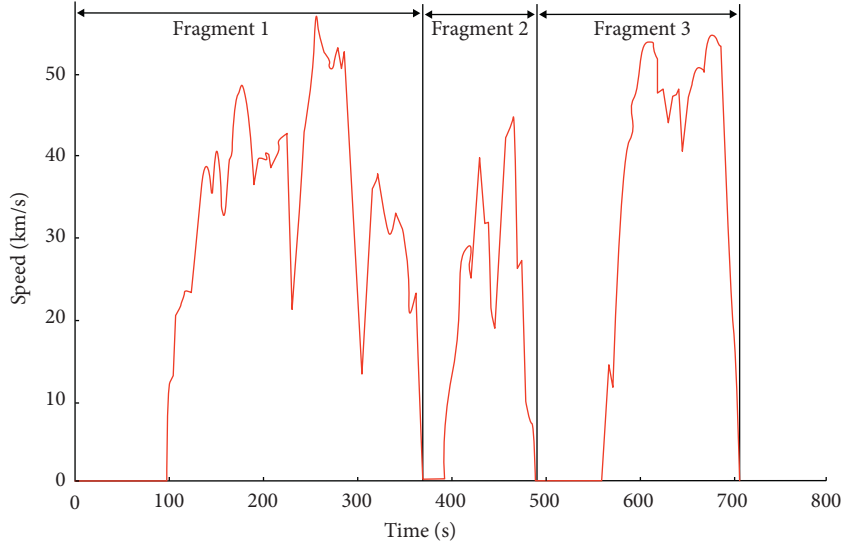
FIGURE 4: Segmentation of kinematic fragments.

TABLE 1: Feature parameters for the division of kinematic fragments.

No.	Feature parameter (unit)	Definition
1	$T\,(s)$	Duration of the fragment
2	$P_a\,(\%)$	Acceleration time percentage
3	$P_d\,(\%)$	Deceleration time percentage
4	$P_c\,(\%)$	Cruise time percentage
5	$P_i\,(\%)$	Idle time percentage
6	$S\,(km)$	Driving distance
7	$v_{max}\,(km/h)$	Maximum speed
8	$a_{max}\,(m/s^2)$	Maximum acceleration
9	$a_{min}\,(m/s^2)$	Maximum deceleration
10	$V_a\,(km/h)$	Average speed
11	$V_x\,(km/h)$	Average moving speed
12	$a_a\,(m/s^2)$	Average acceleration
13	$a_d\,(m/s^2)$	Average deceleration
14	v_{me}	Average speed of cruise time
15	V_{std}	Speed standard deviation
16	a_{std}	Acceleration standard deviation

where $a_{i,i+1}$ is the acceleration from the ith second to the $(i+1)^{th}$ second (m/s^2), v_i is the speed at the ith second (km/h), and t_i is the time point of the ith second (s).

Parameters including v_{max}, v_m, v_{me}, and v_{std} can be calculated by using the following equation:

$$v_{max} = \max\{v_1, v_2, \ldots, v_i\},$$

$$v_m = \text{mean}\{v_1, v_2, \ldots, v_k\} = \frac{S}{T},$$

$$v_{me} = \text{mean}\{v_j, v_{j+1}, \ldots, v_k\} = \frac{S}{(T - T_i)}, \quad (2)$$

$$v_{std} = \text{std}\{v_1, v_2, \ldots, v_k\} = \sqrt{\frac{1}{k-1} \sum_{i=1}^{k} (v_i - v_m)^2}.$$

Other parameters including a_{max}, a_{min}, a_a, and a_d can be calculated by using the following equation:

$$a_{max} = \max\{a_1, a_2, \ldots, a_{k-1}\},$$

$$a_{min} = \min\{a_1, a_2, \ldots, a_{k-1}\},$$

$$a_a = \text{mean}\{a_i | a_i \geq 0.15\} = \sum \frac{\{a_i | a_i \geq 0.15\}}{T_a}, \quad (3)$$

$$a_d = \text{mean}\{a_i | a_i \leq -0.15\} = \sum \frac{\{a_i | a_i \leq -0.15\}}{T_d}.$$

For each fragment, the starting speed and the ending speed are 0, so the average acceleration is 0. Thus, the acceleration standard deviation is formulated as the following equation:

$$a_{std} = \text{std}\{a_1, a_2, \ldots, a_{k-1}\} = \sqrt{\frac{1}{k-1} \sum_{i=1}^{k} a_i^2}. \quad (4)$$

T is the time length of a complete fragment, Ta is the time duration when $a \geq 0.1\,m/s^2$, T_d is the duration when $a \leq -0.1\,m/s^2$, and $T_e = T - T_a - T_d - T_i$. T_i is the idling time.

Table 2 presents the feature parameters for the validity assessment of the kinematic fragments.

Parameters including P_i, P_a, P_b, and P_c can be calculated by equations (5)–(8):

$$P_c = \frac{T_c}{T}, \quad (5)$$

$$P_a = \frac{T_a}{T}, \quad (6)$$

$$P_d = \frac{T_d}{T}, \quad (7)$$

TABLE 2: Feature parameters for the validity assessment of the kinematic fragments.

No.	Feature parameters (unit)	Definition
1	T_i (s)	Idle time
2	T_a (s)	Acceleration time
3	T_d (s)	Deceleration time
4	T_e (s)	Cruise time
5	P_{0-10} (%)	Time percentage at speed 0~10 km/h
6	P_{10-20} (%)	Time percentage at speed 10~20 km/h
7		Time percentage at speed 20~30 km/h
8	P_{30-40} (%)	Time percentage at speed 30~40 km/h
9	P_{40-50} (%)	Time percentage at speed 40~50 km/h
10	P_{50-60} (%)	Time percentage at speed 50~60 km/h
11	P_{60-70} (%)	Time percentage at speed 60~70 km/h
12	P_{70-80} (%)	Time percentage at speed 70~80 km/h
13	P_{80+} (%)	Time percentage at speed over 80 km/h

$$P_c = \frac{T_c}{T}. \tag{8}$$

Other parameters including P_{1-10}, P_{10-20}...P_{80+} can be obtained by using the following equations:

$$
\begin{aligned}
P_{0-10} &= \frac{\text{count}\{v_i | v_i \le 10\}}{T}, \\
P_{10-20} &= \frac{\text{count}\{v_i | 10 < v_i \le 20\}}{T}, \\
&\cdots \\
P_{80+} &= \frac{\text{count}\{v_i | v_i > 80\}}{T}.
\end{aligned} \tag{9}
$$

After data cleansing, fragmentation, and parameter extraction, a total of 3720 pieces of 28-dimension data were obtained at last. The clustering method is used to cluster the microtrips (kinematic fragments), but the high dimension of data reduces the computation efficiency and undermines the clustering effect. Thus, the PCA method is employed to reduce the dimension of data.

2.5. Data Dimension Reduction Using PCA and Factor Analysis (FA)

(1) Principal component analysis:

Principal component analysis is a method that transforms a set of complex variables into a few principal components by adding several variables to reduce the number of variables while minimizing the loss of information. This method simplifies data and obtains results with more effective information [31]. The eigendecomposition method is used in this study.

(2) FA is an extension of PCA [32]. FA is a method that describes relationships between individual variables of a dataset. When FA is employed for decomposition, the common factors $f = (f_1, f_2, \ldots, f_n)'$ and the special factors $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)'$ are obtained, and then the original variables are modeled as linear combinations of common factors.

2.6. Fragment Selection Based on the Minimum Deviation from the Average Feature Parameters

(1) After the kinematic fragments are divided by using the clustering algorithm, all fragments are uploaded into a library of k-category fragments. The number of feature parameters of a microtrip is i, and c_i is the number of fragments in the category of $i = (1, 2, \ldots, k)$. First, standardized processing is performed on the feature parameters of fragments, as shown in the following equation:

$$
d_{pq,n} = \frac{\left(x_{pq,n} - \min x_{q,n}\right)}{\left(\max x_{q,n} - \min x_{q,n}\right)},
$$
$$
p = 1, 2, \ldots, c_n, q = 1, 2, \ldots, i, n = 1, 2, \ldots, k, \tag{10}
$$

where $x_{pq,n}$ is the qth parameter value in the nth fragment in the n category, $\max x_{q,n}$ and $\min x_{q,n}$ refer to the maximum and minimum values of the qth parameter in the n category, respectively, and $d_{pq,n}$ is the parameter value after standardized processing.

(2) The sum of all nondimensionalized feature parameters in the library of the i ($i = 1, 2, \ldots, k$) category is calculated, as shown in the following equation:

$$
d_{p,n} = \sum_{q=1}^{i} d_{pq,n}, \quad p = 1, 2, \ldots, c_i. \tag{11}
$$

(3) The average of all nondimensionalized parameters in the n ($n = 1, 2, \ldots, k$) category is calculated and summed up, as shown in the following equation:

$$
Z_n = \sum_{q=1}^{i} \frac{\left(\overline{x}_{q,n} - \min x_{q,n}\right)}{\left(\max x_{q,n} - \min x_{q,n}\right)}, \tag{12}
$$

where $\overline{x}_{q,n}$ is the average of the qth parameter of all fragments in the n category and Z_n is the sum of all nondimensionalized feature parameters of all fragments in the n category.

(4) The time percentage of each fragment in the cycle is calculated, as shown in equations (13) and (14):

$$P_{n1} = \frac{\sum_{p=1}^{m_k} T_{p,n}}{\sum_{n=1}^{k} \sum_{p=1}^{c_n} T_{ip,n}} \times T_{e1}, \qquad (13)$$

$$P_{n2} = \frac{\sum_{p=1}^{m_k} T_{p,n}}{\sum_{n=1}^{k} \sum_{p=1}^{c_n} T_{ip,n}} \times T_{e2}, \qquad (14)$$

where P_{n1} and P_{n2} are the time percentage of the n-category fragments in the driving cycle to be obtained and $T_{p,n}$ is the time duration of the pth fragment in the n category. $T_{e1} = 1200$ and $T_{e2} = 1300$ are the time range of the final driving cycle.

(5) The difference between the sum of all feature parameters in the i category and the sum of all average parameters in the i category $d_{p,n}$ is calculated, and its absolute value is adopted. The candidate driving cycle fragments are sequenced in the order of the size of the interpolated values, and the fragments are selected one after another in this order till the time duration of the selected fragment exceeds the value of P_n in equations (13) and (14).

(6) The fragments selected from each category are combined to construct a representative driving cycle.

2.7. Fragment Selection Using AMPSO.
In the process of preexperiment, we found that the traditional PSO algorithm would lead to premature algorithm or fall into local minimum. After comparing the traditional PSO, AMPSO, and hybrid differential evolution particle swarm optimization (DEPSO), it was found that AMPSO has the best global search ability. Therefore, AMPSO algorithm was selected. AMPSO follows the fundamental steps of PSO [33]: first, the swarm is initialized, a particle is randomly chosen in the feasible solution space, and the state of the selected particle is represented by three indicators—position, speed, and fitness. After that, the optimal solution of the selected particle and the optimal solution of the whole swarm are tracked to update through continuous iteration. When all particles stop searching in a dimension, the adaptive mutation strategy is triggered, the dimensional activity factor is introduced, and the quality particles are maintained for subsequent iterations to avoid precocity and local minimum problem. In Section 2.6, we proposed a selection criterion, and in this section, we proposed a specific implementation algorithm for selecting.

The optimization function of the algorithm is shown in the following equation:

$$Q = \varepsilon_1 \sum_k \max\left(Z_{q_k k} - Y_k\right) + \varepsilon_2 \sum_k \frac{\min\left(\left|\sum_{q_k} a_k k - T_{k_1}\right|, \left|\sum_{q_k} a_k k - T_{k_2}\right|\right)}{T_{k_1} - T_{k_2}}, \qquad (15)$$

where q_k is the sequence number of selected fragments in the k category, $Z_{q_k k}$ is the average feature parameter of the q_k^{th} fragment in the k category, and ε_1 and ε_2 are constant coefficients (both are set as 0.5).

3. Results and Discussion

3.1. Analysis of Feature Parameters Using FA and PCA.
The values of the 15 feature parameters for the 3720 fragments obtained in Section 2.3 are calculated. PCA is performed using SPSS (SPSS Inc., 16.0) on these 15 feature parameters. Typically, correlation test should be performed before PCA analysis.

3.1.1. Correlation Test.
The FA method can find common factors and latent representative factors from the group of variables. Correlation among variables is a premise for factor analysis, so correlation tests must be made before factor analysis. Table 3 shows the correlation between feature parameters. Only six feature parameters are listed here.

Table 4 shows the results of the KMO and Bartlett's test.

The KMO value in Table 4 is 0.961 (>0.7), and Sig. is 0, which indicates strong correlation between the feature parameters, and thus, factor analysis is feasible.

3.1.2. Extraction of Principal Components and Common Factors.
Table 5 shows the result of PCA.

In the data used to develop the driving cycle, the components numbered 1, 2, 3, and 4 have an eigenvalue above 1 and can explain 31.651%, 20.236%, 13.220%, and 9.786% variance, respectively, and the percentage of cumulative variance explained reaches 74.893%, which means much information is explained. Thus, the components 1, 2, 3, and 4 are extracted as the principal components, and other components are dispensed with because they contain little information.

Figure 5 shows the scree test result.

Then, the component matrix is extracted, as shown in Table 6.

The common factors are stored for subsequent operations, and regression is performed to obtain the component score coefficient matrix. The method of varimax with Kaiser normalization is used for factor rotation. New variables generated by using SPSS are placed at the end of the sample data as values for the common factors. The output factor coefficients are shown in Table 7.

Table 7 shows the component matrix after fragment selection. Factor rotation minimizes the number of variables of high loadings in each factor and better explains the common factors.

TABLE 3: Correlation among feature parameters.

	Fragment duration	Acceleration time percentage	Deceleration time percentage	Cruise time percentage	Idle time percentage	...	Acceleration standard deviation
Fragment duration	1.000	0.083	−0.075	0.081	−0.052	...	−0.242
Acceleration time percentage	0.083	1.000	0.140	−0.081	−0.528	...	0.241
Deceleration time percentage	−0.075	0.140	1.000	0.064	−0.650	...	0.357
Cruise time percentage	0.081	−0.081	0.064	1.000	−0.616	...	−0.100
Idle time percentage	−0.052	−0.528	−0.650	−0.616	1.000	...	−0.252
⋮	⋮	⋮	⋮	⋮	⋮	...	...
Acceleration standard deviation	−0.242	0.241	0.357	−0.100	−0.252	...	1.000

TABLE 4: KMO and Bartlett's test.

Kaiser–Meyer–Olkin measure of sampling adequacy		0.961
Bartlett's test of sphericity	Approx. chi-square	328,726.009
	df. (degree of freedom)	91
	Sig. (significance)	0.000

As shown in Tables 6 and 7, the first principal component represents the cruise time, idle time, driving distance, maximum speed, average speed, and average driving speed. The second principal component reflects the maximum deceleration, average deceleration, and acceleration standard deviation. The third reflects the fragment duration, and the fourth reflects the acceleration time. The four principal components cover all the 15 constructed feature parameters and hence can fully reflect the features of the fragments. By multiplying the normalized driving speed data matrix by the principal component matrix, we obtain the principal component score matrix. Equation (16) shows how the principal components are calculated. Scores of the four principal components are taken as the research object for clustering.

$$
\begin{aligned}
\text{FAC\_1} =\ & 0.476T + 0.368P_a + 0.296P_d + 0.590P_e - 0.712P_i + 0.745S + 0.911v_{\max} \\
& - 0.170a_{\max} - 0.192a_{\min} \\
& + 0.927v_a + 0.895v_x - 0.390a_a + 0.140a_d + 0.397v_{\text{std}} - 0.044a_{\text{std}}, \\
\text{FAC\_2} =\ & -0.040T + 0.282P_a + 0.227P_d \\
& - 0.137P_e - 0.178P_i - 0.037S + 0.116v_{\max} \\
& + 0.594a_{\max} - 0.778a_{\min} - 0.035v_a - 0.049v_x + 0.612a_a - 0.746a_d \\
& + 0.338v_{\text{std}} + 0.911a_{\text{std}}, \\
\text{FAC\_3} =\ & 0.061T + 0.624P_a + 0.368P_d - 0.553P_e \\
& - 0.168P_i - 0.082S - 0.078v_{\max} + 0.012a_{\max} \\
& + 0.038a_{\min} - 0.220v_a - 0.222v_x - 0.407a_a + 0.367a_d \\
& + 0.438v_{\text{std}} + 0.024a_{\text{std}}, \\
\text{FAC\_4} =\ & 0.061T + 0.624P_a + 0.368P_d \\
& - 0.553P_e - 0.168P_i - 0.082S - 0.078v_{\max} \\
& + 0.012a_{\max} + 0.038a_{\min} - 0.220v_a - 0.222v_x - 0.407a_a \\
& + 0.367a_d + 0.348v_{\text{std}} + 0.024a_{\text{std}}.
\end{aligned}
\tag{16}
$$

3.2. Clustering Result Analysis.

The scores of four principal components are clustered, and the number of condensation points is set as 3. The K-means clustering algorithm is realized by software to divide the 3720 kinematic fragments

TABLE 5: Total variance explained.

Component	Initial eigenvalues			Sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	4.748	31.651	31.651	4.748	31.651	31.651
2	3.035	20.236	51.887	3.035	20.236	51.887
3	1.983	13.220	65.108	1.983	13.220	65.108
4	1.468	9.786	74.893	1.468	9.786	74.893
5	0.999	6.660	81.553			
6	0.806	5.372	86.926			
7	0.674	4.495	91.420			
8	0.433	2.889	94.309			
9	0.381	2.542	96.851			
10	0.211	1.408	98.260			
11	0.147	0.977	99.237			
12	0.080	0.535	99.772			
13	0.027	0.182	99.954			
14	0.007	0.046	100.000			
15	−8.3E-16	−5.555E-15	100.000			
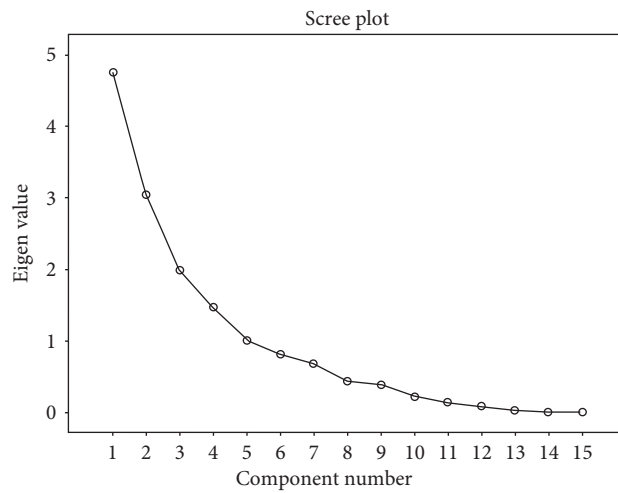
FIGURE 5: PCA scree test.

TABLE 6: Component matrix.

	Component			
	1	2	3	4
Fragment duration	0.476	−0.040	0.672	0.061
Acceleration time percentage	0.368	0.282	−0.065	0.624
Deceleration time percentage	0.296	0.227	−0.593	0.368
Cruise time percentage	0.590	−0.137	−0.347	−0.553
Idle time percentage	−0.712	−0.178	0.569	−0.168
Driving distance	0.745	−0.037	0.429	−0.082
Maximum speed	0.911	0.116	0.157	−0.078
Maximum acceleration	−0.170	0.594	0.482	0.012
Maximum deceleration	−0.192	−0.778	−0.039	0.038
Average speed	0.927	−0.035	−0.114	−0.220
Average driving speed	0.895	−0.049	0.078	−0.222
Average acceleration	−0.390	0.612	0.142	−0.407
Average deceleration	0.140	−0.746	0.105	0.367
Speed standard deviation	0.397	0.338	0.417	0.438
Acceleration standard deviation	−0.044	0.911	−0.258	0.024

TABLE 7: Rotational component matrix.

|	Component			
	1	2	3	4
Fragment duration	0.397	−0.114	−0.278	0.660
Acceleration time percentage	0.021	0.007	0.610	0.486
Deceleration time percentage	0.078	0.050	0.782	−0.078
Cruise time percentage	0.784	−0.015	0.068	−0.415
Idle time percentage	−0.541	−0.021	−0.771	0.056
Driving distance	0.700	−0.096	−0.107	0.486
Maximum speed	0.836	0.023	0.176	0.379
Maximum acceleration	−0.186	0.572	−0.225	0.448
Maximum deceleration	−0.147	−0.716	−0.216	−0.254
Average speed	0.923	−0.072	0.243	0.071
Average driving speed	0.898	−0.080	0.097	0.195
Average acceleration	−0.187	0.757	−0.309	−0.098
Average deceleration	−0.008	−0.837	−0.002	0.146
Speed standard deviation	0.132	0.121	0.193	0.754
Acceleration standard deviation	−0.098	0.847	0.414	0.033
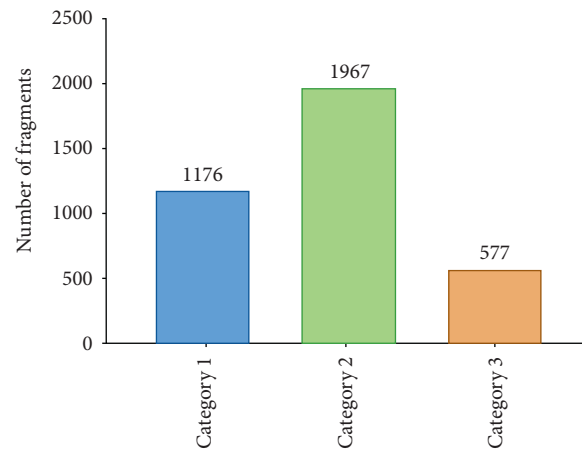
FIGURE 6: Number of fragments in three categories.

into three categories. The respective number of fragments (microtrips) of these three categories is 1176, 1967, and 577, as shown in Figure 6.

The driving features reflected in the microtrips in each category are further analyzed, and the parameters and equations in Section 2.4 are used to obtain the composite feature parameters of these categories, as shown in Table 8.

On the basis of Table 8, Figures 7–9 are obtained.

As shown in Figure 8, there are patterns in the distribution of composite feature parameters. According to these patterns, category 1, 2, and 3 are considered as a low-speed, medium-speed, and high-speed category of fragments, respectively. The following patterns can be concluded:

(1) With regard to the number of kinematic fragments, the number of medium-speed fragments is the highest, reaching 1957, followed by 1176 low-speed fragments and 577 high-speed fragments.

(2) With regard to the driving time percentage, the low-speed fragments have the highest idle time percentage at 58.98%, and the high-speed fragments have the lowest idle time percentage at 4.39%. The constant-speed time percentage reaches the highest at 35.16% in the high-speed fragments while hitting the lowest at 9.28% in the low-speed fragments.

(3) In terms of speed distribution percentage, the percentage of speed at 0–10 km/h reaches the highest at 72.08% in the category of low-speed fragments, and the percentage of speed over 80 km/h reaches the highest at 17.09% in the category of high-speed fragments.

TABLE 8: Composite feature parameters of fragments in three categories.

	Category 1	Category 2	Category 3
Number of fragments	1176	1967	577
Driving time (s)	108,429	227,210	132,055
Acceleration time (s)	18,898	79,114	44,485
Deceleration (s)	15,512	63,795	35,336
Constant-speed time (s)	10,066	45,784	46,431
Idle time (s)	63,953	38,517	5803
Driving distance (km)	198.537	1555.13	1964.93
Maximum speed (km/h)	58	122.1	173.8
Maximum acceleration (m/s^2)	3.91	3.912	3.94
Maximum deceleration (m/s^2)	−5.92	−7.29	−7.58
Average speed (km/h)	6.59	24.64	53.57
Average driving speed (km/h)	16.07	29.67	56.03
Average acceleration in the acceleration mode (m/s^2)	0.53	0.46	0.35
Average deceleration in the deceleration mode (m/s^2)	−0.54	−0.58	−0.44
Acceleration time percentage	0.174289	0.348198	0.336867
Deceleration time percentage	0.143061	0.280775	0.267585
Constant-speed time percentage	0.092835	0.201505	0.351603
Idle time percentage	0.589815	0.169522	0.043944
Percentage of speed at 0–10 (km/h)	0.72084	0.248818	0.062565
Percentage of speed at 10–20 (km/h)	0.150725	0.148162	0.03881
Percentage of speed at 20–30 (km/h)	0.085245	0.19857	0.065223
Percentage of speed at 30–40 (km/h)	0.035037	0.19706	0.10978
Percentage of speed at 40–50 (km/h)	0.007682	0.141213	0.166658
Percentage of speed at 50–60 (km/h)	0.00047	0.051151	0.179153
Percentage of speed at 60–70 (km/h)	0	0.011514	0.122161
Percentage of speed at 70–80 (km/h)	0	0.00331	0.084775
Percentage of speed > 80 (km/h)	0	0.000202	0.170876

3.3. Development of the Vehicle Driving Cycle and Analysis.
A proper vehicle driving cycle is usually between 1200 and 1300 s. The time duration of fragments that should be selected in each category is obtained based on the clustering result, and the AMPSO is employed to find fragments with the minimum deviation from the composite feature parameters. As the optimal fragments obtained by using the AMPSO differ every time the algorithm is performed, multiple experiments are conducted, and fragments with the minimum deviation from the composite feature parameters are selected so that driving cycle 1 is obtained, as shown in Figure 10. Moreover, the same number of experiments is conducted using the random selection method, which leads to driving cycle 2, as shown in Figure 11. These two cycles are compared. Their time lengths are 1216 s and 1261 s, respectively.

3.4. Constructed Driving Cycles vs. Existing Driving Cycles.
To better display the comparison result, Figures 12–14 are produced based on Tables 9 and 10.

As shown in Figures 12–14, driving cycle 1 has smaller errors than driving cycle 2 in terms of four parameters—average speed, average driving speed, average acceleration, and average deceleration. Driving cycle 1 is closer to the result of pre-processed experiment data than driving cycle 2 in terms of the acceleration percentage, deceleration percentage, constant-speed percentage, and idle percentage, as well as the driving speed frequency.

Figure15 shows the final driving cycle. Comparison among the NEDC, FTP75, WLTC, experiment data, and the final driving cycle developed in this study is performed, as shown in Table 11.

As the result shows, the method proposed in this study improves the effect of the random experiment method and solves the problem of low reliability. Compared to typical driving cycles abroad, the driving cycle developed herein better suits the actual road and traffic conditions of Fuzhou urban district in terms of the average speed, average acceleration, and percentages of driving modes. The vehicle driving features of the developed driving cycle are closer to the data sources and are thus representative.
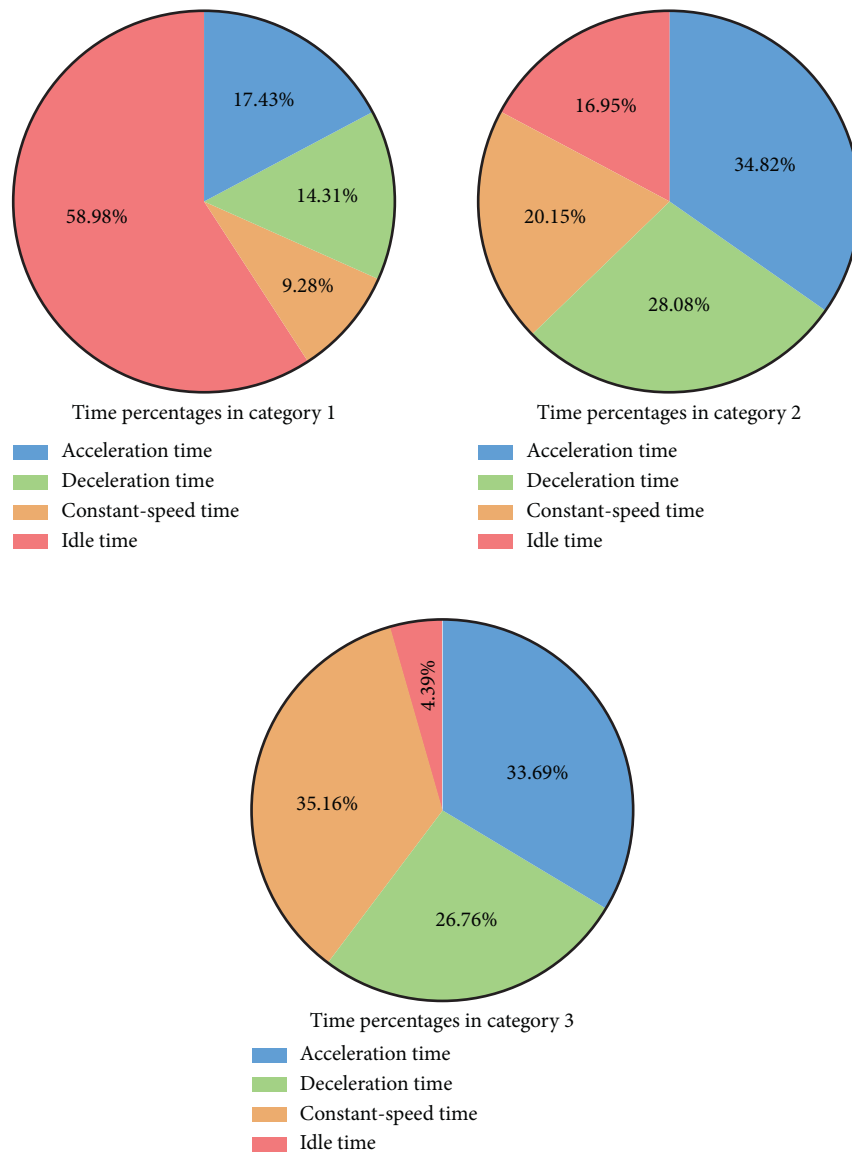
Figure 7: Time percentage of fragments in (a) category 1, (b) category 2, and (c) category 3.
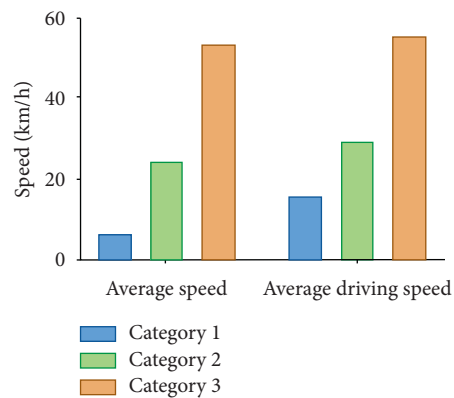

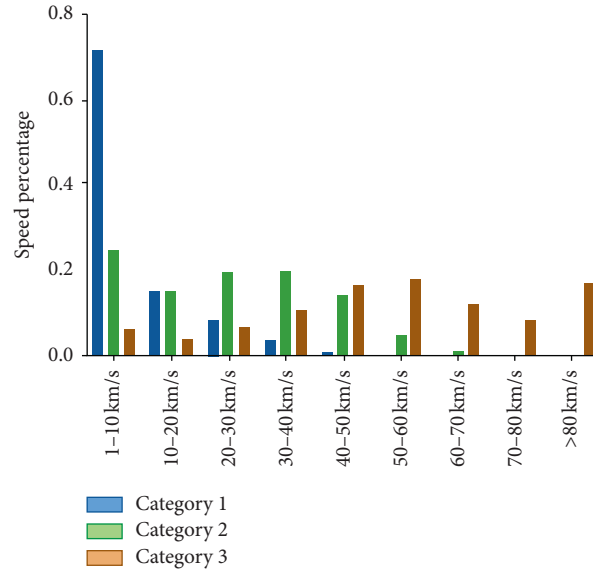Figure 8: Average speed and average driving speed in each category.

Figure 9: Distribution of speed in each category.

Table 9: Comparison between constructed driving cycles and existing driving cycles.

	Experiment data	Driving cycle 1	Driving cycle 2
Average speed (km/h)	28.62	27.58	29.24
Average driving speed (km/h)	37.25	36.86	39.91
Average acceleration in the acceleration mode (m/s^2)	0.44	0.44	0.44
Average deceleration in the deceleration mode (m/s^2)	−0.54	−0.63	−0.64
Acceleration percentage (%)	30.53	32.08	31.11
Deceleration percentage (%)	24.87	24.42	25.93
Constant-speed percentage (%)	21.45	19.90	20.67
Idle percentage (%)	23.15	23.60	22.28

Table 10: Comparison of validity analysis feature parameters.

	Experiment data	Driving cycle 1	Driving cycle 2
Percentage of speed at 1–10 km/h (%)	30.57	30.18	31.96
Percentage of speed at 10–20 km/h (%)	11.79	12.20	11.94
Percentage of speed at 20–30 km/h (%)	13.46	11.76	12.10
Percentage of speed at 30–40 km/h (%)	13.49	16.12	15.26
Percentage of speed at 40–50 km/h (%)	11.74	19.16	18.64
Percentage of speed at 50–60 km/h (%)	7.55	6.69	7.00
Percentage of speed at 60–70 km/h (%)	4.01	3.89	3.10
Percentage of speed at 70–80 km/h (%)	2.55	0	0
Percentage of speed over 80 km/h (%)	4.83	0	0

Table 11: Comparison with three typical driving cycles.

|	Constant-speed percentage (%)	Acceleration percentage (%)	Idle percentage (%)	Deceleration percentage (%)	Average acceleration in the acceleration mode (m/s^2)	Average deceleration in the deceleration mode (m/s^2)	Average speed (km/h)
NEDC	36.8	23.2	23.7	17.3	0.48	−0.67	33.2
FTP75	25.6	19.09	26.45	28.85	0.66	−0.75	34.1
WLTC	25.3	31.9	12.6	30.2	1.65	−1.7	46.5
Experiment data	21.44	30.53	23.15	24.87	0.4369	−0.5436	28.6
Developed driving cycle	19.90	32.08	23.60	24.42	0.4375	−0.6327	27.6
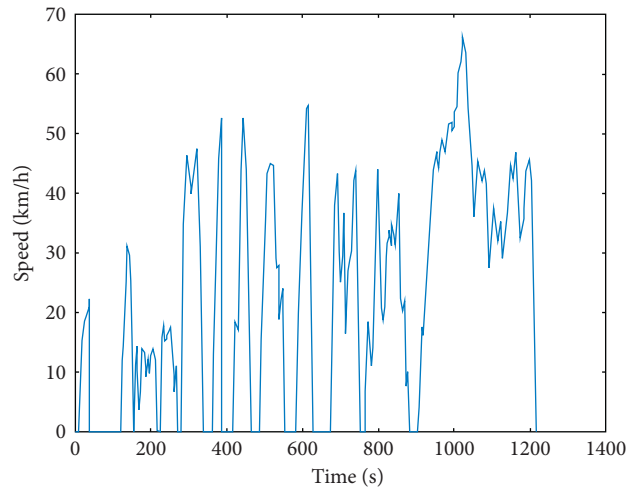
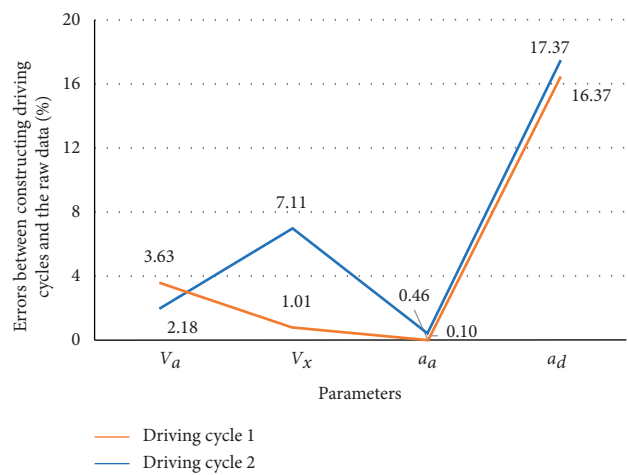FIGURE 10: Driving cycle 1.
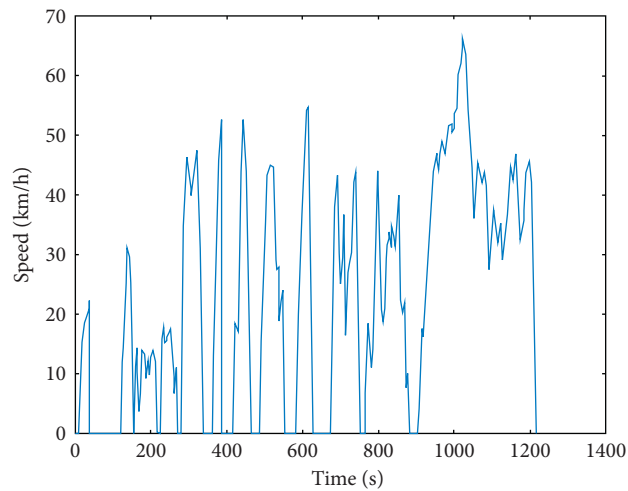
FIGURE 11: Driving cycle 2.

FIGURE 12: Errors between the two cycles and the experiment data (parameters on speed and acceleration).
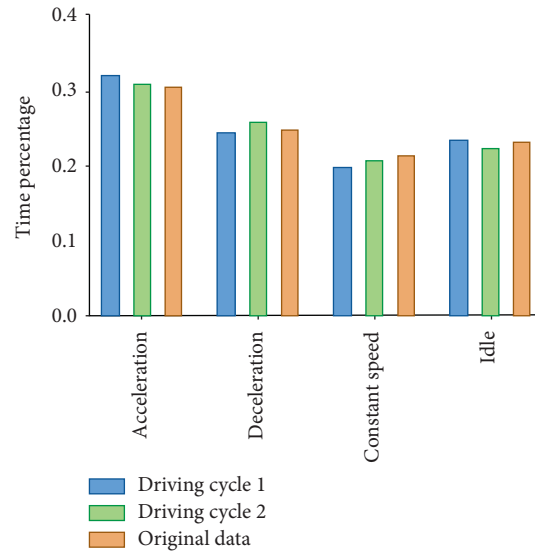
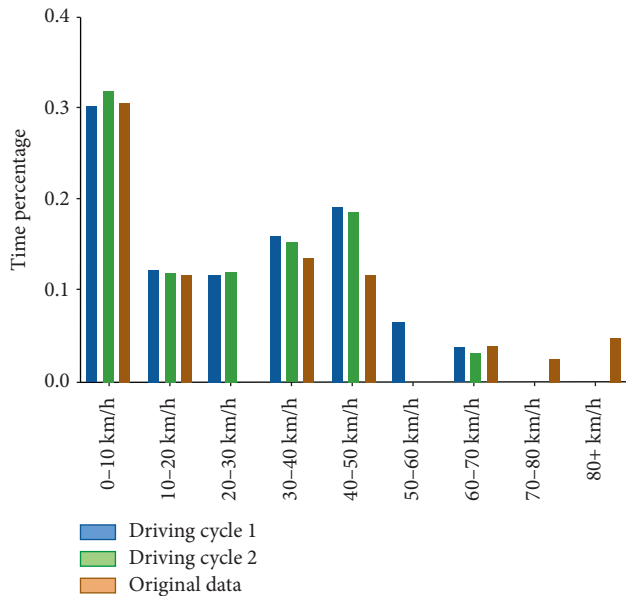FIGURE 13: Distribution of driving modes under the two driving cycles.

FIGURE 14: Speed distribution under the two driving cycles.

4. Conclusion

The driving cycle is an important benchmark for vehicle design and evaluation. A driving cycle is built in this study based on actual driving data in Fuzhou. Analysis of vehicle usage and driving features in Fuzhou reveals that the actual driving cycle in Fuzhou differs much from standard driving cycles. Therefore, it is necessary to develop and use specific driving cycles for the evaluation of car emissions.

With driving data collected by a minibus, this study preprocessed the raw data, performed K-means clustering on the fragments, screened the fragments using AMPSO, and developed a driving cycle in Fuzhou. The developed driving cycle is compared to existing driving cycles. The developed driving cycle can reflect the actual driving

conditions. The acceleration percentage under the four driving modes shows the largest deviation of 1.5%. The developed driving cycle can also be used for the evaluation of fuel consumption and emissions of vehicles in Fuzhou.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] A. Jenn, I. M. L. Azevedo, and J. J. Michalek, "Alternative fuel vehicle adoption increases fleet gasoline consumption and greenhouse gas emissions under United States corporate average fuel economy policy and greenhouse gas emissions standards," *Environmental Science & Technology*, vol. 50, no. 5, pp. 2165–2174, 2016.

[2] C. K. Gately, L. R. Hutyra, S. Peterson, and I. Sue Wing, "Urban emissions hotspots: quantifying vehicle congestion and air pollution using mobile phone GPS data," *Environmental Pollution*, vol. 229, pp. 496–504, 2017.

[3] R. Smit, M. Cope, and L. Knibbs, "Building an Australian national motor vehicle inventory for air pollution population exposure assessment," *Air Quality and Climate Change*, vol. 53, no. 2, p. 7, 2019.

[4] P. Nyberg, E. Frisk, and L. Nielsen, "Using real-world driving databases to generate driving cycles with equivalence properties," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 4095–4105, 2015.

[5] J. Brady and M. O'Mahony, "Development of a driving cycle to evaluate the energy economy of electric vehicles in urban areas," *Applied Energy*, vol. 177, pp. 165–178, 2016.

[6] K. Kivekäs, A. Lajunen, J. Vepsäläinen, and K. Tammi, "City bus powertrain comparison: driving cycle variation and passenger load sensitivity analysis," *Energies*, vol. 11, no. 7, p. 1755, 2018.

[7] M. Tutuianu, P. Bonnel, B. Ciuffo et al., "Development of the World-wide harmonized Light duty Test Cycle (WLTC) and a possible pathway for its introduction in the European legislation," *Transportation Research Part D: Transport and Environment*, vol. 40, pp. 61–75, 2015.

[8] N. Romain, "The different driving cycles," March 2020, http://www.car-engineer.com/the-different-driving-cycles.

[9] J. Pavlovic, B. Ciuffo, G. Fontaras, V. Valverde, and A. Marotta, "How much difference in type-approval CO2 emissions from passenger cars in Europe can be expected from changing to the new test procedure (NEDC vs. WLTP)?" *Transportation Research Part A: Policy and Practice*, vol. 111, pp. 136–147, 2018.

[10] P. Brabec and X. Zhou, "Simulation of the ride for passenger car in NEDC and wltp driving cycles," in *Proceedings of the 24th International Conference Engineering Mechanics*, Svratka, Czech Republic, 2018, May.

[11] M. André, R. Joumard, R. Vidon, P. Tassel, and P. Perret, "Real-world European driving cycles, for measuring pollutant emissions from high- and low-powered cars," *Atmospheric Environment*, vol. 40, no. 31, pp. 5944–5953, 2006.

[12] J. Lin and D. A. Niemeier, "Regional driving characteristics, regional driving cycles," *Transportation Research Part D: Transport and Environment*, vol. 8, no. 5, pp. 361–381, 2003.

[13] S.-H. Ho, Y.-D. Wong, and V. W.-C. Chang, "Developing Singapore driving cycle for passenger cars to estimate fuel consumption and vehicular emissions," *Atmospheric Environment*, vol. 97, pp. 353–362, 2014.

[14] A. Fotouhi and M. J. S. I. Montazeri-Gh, "Tehran driving cycle development using the k-means clustering method," *Scientia Iranica*, vol. 20, no. 2, pp. 286–293, 2013.

[15] G. Amirjamshidi and M. J. Roorda, "Development of simulated driving cycles for light, medium, and heavy duty trucks: case of the Toronto Waterfront Area," *Transportation Research Part D: Transport and Environment*, vol. 34, pp. 255–266, 2015.

[16] K. S. Nesamani and K. P. Subramanian, "Development of a driving cycle for intra-city buses in Chennai, India," *Atmospheric Environment*, vol. 45, no. 31, pp. 5469–5476, 2011.

[17] R. Günther, T. Wenzel, M. Wegner, and R. Rettig, "Big data driven dynamic driving cycle development for busses in urban public transportation," *Transportation Research Part D: Transport and Environment*, vol. 51, pp. 276–289, 2017.

[18] Z. Jing, G. Wang, S. Zhang, and C. Qiu, "Building Tianjin driving cycle based on linear discriminant analysis," *Transportation Research Part D: Transport and Environment*, vol. 53, pp. 78–87, 2017.

[19] Al Zaidi and Ahmed, *Investigation of Driving Cycles as Tools to Assess Travel Demand Management in Edinburgh and Abu Dhabi*, Edinburgh Napier University, Edinburgh, Scotland, 2013.

[20] Z. Chen, L. Li, B. Yan, C. Yang, C. Marina Martinez, and D. Cao, "Multimode energy management for plug-in hybrid electric buses based on driving cycles prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2811–2821, 2016.

[21] S. R. Kancharla and G. Ramadurai, "Incorporating driving cycle based fuel consumption estimation in green vehicle routing problems," *Sustainable Cities and Society*, vol. 40, pp. 214–221, 2018.

[22] S. Li, M. Hu, C. Gong, S. Zhan, and D. Qin, "Energy management strategy for hybrid electric vehicle based on driving condition identification using KGA-means," *Energies*, vol. 11, no. 6, p. 1531, 2018.

[23] I. N. Anida and A. R. Salisa, "Driving cycle development for Kuala Terengganu city using k-means method," *International Journal of Electrical & Computer Engineering*, vol. 9, no. 3, pp. 2088–8708, 2019.

[24] S. B. Xie, T. Liu, H. Chen, and L. Wei, "Spatial driving cycle for city bus considering energy management for electric vehicles," *China Journal of Highway and Transport*, vol. 31, no. 1, p. 18, 2018.

[25] M. Ahlborn and M. Wortmann, "The core–periphery pattern of European business cycles: a fuzzy clustering approach," *Journal of Macroeconomics*, vol. 55, pp. 12–27, 2018.

[26] Y. Li, J. Peng, H. He, and S. Xie, "The study on multi-scale prediction of future driving cycle based on Markov chain," *Energy Procedia*, vol. 105, pp. 3219–3224, 2017.

[27] M. Fries, A. Baum, M. Wittmann, and M. Lienkamp, "Derivation of a real-life driving cycle from fleet testing data with the markov-chain-monte-carlo method," in *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2550–2555, IEEE, Maui, HI, USA, 2018, November.

[28] S. Tewiele, P. Driesch, T. Weber, and D. Schramm, "Clustering of real BEV driving data with subsequent driving cycle construction using Markov chains," in *Proceedings of the AmE 2018-Automotive Meets Electronics; 9th GMM-Symposium*, pp. 1–6, VDE, Dortmund, Germany, 2018, March.

[29] J. Françoise, A. Roby-Brami, N. Riboud, and F. Bevilacqua, "Movement sequence analysis using hidden Markov models: a case study in Tai Chi performance," in *Proceedings of the 2nd International Workshop on Movement and Computing*, pp. 29–36, ACM, Vancouver, Canada, 2015, August.

[30] Y. Peng, Z. Yuan, and H. Yang, "Development of a representative driving cycle for urban buses based on the K-means cluster method," *Cluster Computing*, vol. 22, pp. 6871–6880, 2018.

[31] J. Lee and Y. Choe, "Robust PCA Based on Incoherence With Geometrical Interpretation," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1939–1950, 2018.

[32] W. Sun and J. Sun, "Daily PM 2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm," *Journal of Environmental Management*, vol. 188, pp. 144–152, 2017.

[33] H. T. Liang and F. H. Kang, "Adaptive mutation particle swarm algorithm with dynamic nonlinear changed inertia weight," *Optik—International Journal for Light and Electron Optics*, vol. 127, no. 19, 2016.