IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 31, 2020, accepted January 12, 2021, date of publication January 19, 2021, date of current version January 27, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3052820*

# Development of a Typical Urban Driving Cycle for Battery Electric Vehicles Based on Kernel Principal Component Analysis and Random Forest

**LU WANG, JIAN MA, XUAN ZHAO, AND XUEBO LI**
School of Automobile, Chang'an University, Xi'an 710064, China

Corresponding authors: Jian Ma (majian@chd.edu.cn) and Xuan Zhao (zhaoxuan@chd.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1600604; in part by the Fok Yingdong Young Teachers Fund Project under Grant 171103; in part by the Key Research and Development Program of Shaanxi under Grant 2019ZDLGY15-01, Grant 2019ZDLGY15-02, and Grant 2020ZDLGY16-02; and in part by the Natural Science Basic Research Program of Shaanxi under Grant 2020JQ-381.

**ABSTRACT** Great concerns have been raised on the driving cycle due to its critical importance in vehicle design, energy management strategy, and energy consumption forecast of new energy vehicles. Taking Xi'an city as a case, a novel method of driving cycle development for battery electric vehicles is proposed in this paper. First, the chase car method and on-board measurement method are combined to collect sufficient real driving data, which are randomly divided into two parts for developing and validating the target cycle. Then the nonlinear dimension reduction of characteristic parameters with respect to the micro-trips is achieved by employing kernel principal component analysis, and an improved clustering method is developed for constructing candidate cycles, in which the K-means clustering algorithm is applied in the training of random forest. The target cycle is selected from the candidate cycles by determining the assessment criteria with consideration of the characteristic parameters and the speed-acceleration distribution probability. Finally, a comparative study of different methods is implemented to illustrate the effectiveness of the proposed method. The typicality of the target cycle is revealed by analyzing the discrepancies between the target cycle and other legislative cycles.

**INDEX TERMS** Urban driving cycle, battery electric vehicles, random forest, kernel principal component analysis.

## I. INTRODUCTION

Owing to the advantages of environmental friendliness, simple structure, and high energy conversion efficiency, new energy vehicles have great development prospects with strong policy support [1], [2]. A driving cycle is a second-by-second vehicle speed-time profile in a specific region, which is usually selected to describe the driving characteristics [3], [4]. As the input of vehicle design, power matching optimization, and certification standard of energy consumption, the driving cycle is generic technology in the automobile industry [5].

Driving cycles are grouped into legislative and non-legislative cycles according to whether it is adopted by the national or regional government as the standard cycle. The widely recognized legislative cycles in the world include JC08 in Japan, UDDS, FTP75, SFTP, HWFET in the United States, and NEDC in Europe [6]–[8]. The World Regulatory Harmonization Forum (WP.29) proposed the worldwide harmonized light vehicles test procedure (WLTP) for developing a universal cycle, which was implemented from 2017 to evaluate vehicle emissions and energy consumption [9]. Initially, China has been adopting NEDC as the standard cycle for vehicle design and test, but NEDC cannot represent the real driving behavior of the vehicle in China since the test procedure of NEDC is based on the actual road operation in Europe. To fill the technical gap, China automotive test cycle project group has collected the real driving data of 5050 vehicles with a total of 55 million kilometers in 41 typical cities since 2015 and issued two national standards in 2019, *China automotive test cycle-part 1: Light-duty vehicles* and *China automotive test cycle-part 2: Heavy-duty commercial vehicles* [10], [11]. The construction of China automotive test cycles has

The associate editor coordinating the review of this manuscript and approving it for publication was Chandan Kumar.

important reference value for guiding vehicle development and improving vehicle dynamic and economic performance.

Ample research has been conducted on developing non-legislative cycles in various regions for different aims. For instance, cycles of Dublin [4], Istanbul [12], Tianjin [13], Florence [5], and Hefei [14] were established to study energy consumption and driving range. Cycles of Mexico [15], Hanoi [16], Aleppo [17], and Taipei [18] were mainly used to research emissions. The research objects of the above cycles include internal combustion engine vehicles (ICEVs), hybrid electric vehicles, battery electric vehicles (BEVs), and motorcycles. By reviewing previous studies, it can be inferred that the driving cycles varies from region to region and the driving characteristics of different vehicle type are also unlike [19]. For BEVs, the driving range is one of the primary factors considered by users, reflected in an inaccurate nominal driving range that easily causes users to be anxious [20], [21]. Therefore, there is an urgent need to estimate the energy consumption and driving range more accurately. However, China light-duty vehicle test cycle-passenger car (CLTC-P) represents the average characteristics of light-duty passenger vehicles, which cannot provide targeted input for the test of BEVs. Thus, developing a tailored driving cycle for BEVs is the research focus at the current stage.

Existing methods of developing driving cycles can be divided into the micro-trip-based method and Markov analysis-based method. The micro-trip-based method divides the driving data into many micro-trips, then the specified duration or mileage driving cycles are obtained by connecting several micro-trips according to different criteria. Differently, Markov analysis-based method divides the driving data into kinematic fragments and develops the cycles based on Markov analysis, which is increasingly used in combination with Monte-Carlo simulation [15], [22]. In consideration of the complexity and non-repeatability of the combination of Markov analysis and Monte-Carlo method, more attention has been devoted to the micro-trip-based method, in which principal component analysis (PCA) and K-means clustering algorithm are extensively used in relevant literature [5], [23]. Although the PCA and K-means algorithm is effective, only one candidate cycle can be obtained from the same original driving data. Besides, PCA can only extract linear features of characteristics, and K-means is too sensitive to the initial cluster center to achieve global optimization [24].

Intending to solve the above problems, we propose a novel method of developing a driving cycle. The main contributions of our work as follows.

1) we collect the driving data of BEVs in different periods within Xi'an. Parts of the preprocessed original data are randomly selected to construct the candidate driving cycles, whereas the remaining data are used for validation. The main characteristics of micro-trips are extracted based on Kernel principal component analysis (KPCA), and the combination of K-means and random forest (RF) algorithm is adopted to cluster micro-trips.

2) We determine the assessment criterion of the candidate cycles and quantitatively described the differences between candidate cycles and driving data by the performance value. The candidate cycle with the minimum performance value is chosen as the target cycle. The comparisons from diverse aspects reveal the effectiveness of the proposed method and the typicality of the target cycle.

The structure of the paper is as follows. Data acquisition and preprocessing are provided in Section II. In Section III, the construction of candidate cycles is explained. Section IV introduces the development of the target cycle. Validation and comparison of the target cycle are carried out in Section V. Lastly, the conclusion is summarized in Section VI.

## II. DATA ACQUISITION AND PREPROCESSING

The driving cycle development involves data collection, generation of micro-trips, selection of characteristic parameters and assessment criteria, and construction of target cycles.

### A. ROUTE PLANNING AND DATA ACQUISITION

The representativeness of a driving cycle to the traffic pattern of a region firstly depends on the rationality of the selected experimental route [25], [26]. We planned the route according to the urban road topology, traffic flow statistics of various roads, and resident travel surveys. Traffic flow monitoring statistics on expressways, main roads, secondary roads, and branch roads were conducted at 11 predetermined monitoring points in Xi'an. Analytic hierarchy process (AHP) is a method to transform a multi-objective decision problem into a quantitative calculation problem, effectively solving the weight calculation and decision analysis. AHP was applied for the determination of the experimental route by the combination of the traffic flow on the different roads and the weights of fast travel and convenient travel, which were obtained by the survey of residents' travel psychology [1], [27]. Fig. 1 shows the experimental route, and its total length is 38.46km with 38 traffic lights, of which the proportion of the expressways, main roads, secondary roads, and branch roads are 29.96%, 24.80%, 26.85%, and 18.39%, respectively.

The methods of data collection include the chase car method, on-board measurement method, and a combination of both [22], [23], [28], [29]. The chase car method means that the driver randomly follows a target vehicle on the predetermined route. If the target vehicle drives out of the experimental area or the driving behavior is abnormal, the experimental vehicle will search for another target to follow [30], [31]. Given the low demand for resources and convenient implementation, this method has been widely adopted, but still exist two problems. One is that the target is easy to confuse, especially when entering a new road section or in heavy traffic [22]. The other is that drivers may likely change driving styles when followed. The on-board measurement method refers to that Global Positioning System (GPS) and On-Board Diagnosis (OBD) are installed on the experimental vehicle

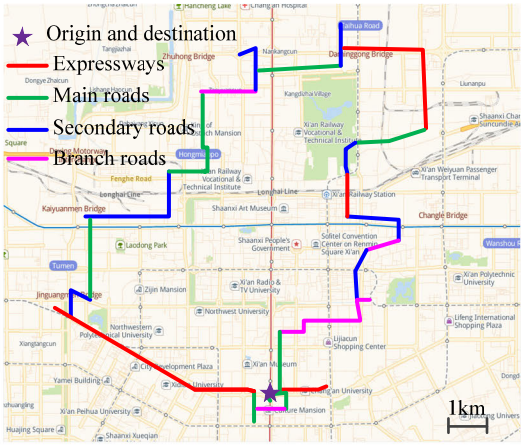


**FIGURE 1.** Experimental route.

to record its driving information along the predetermined route, which can obtain reliable data with high cost [3], [26]. In general, the chase car method and on-board measurement method have their pros and cons. We combined the two methods to collect real driving data on the circulation route for a week to eliminate the influence of weekdays and weekends on the representativeness of data. The daily experiment time consisted of 7:30-9:30 in the morning peak, 12:00-13:30 in the afternoon off-peak, 17:30-19:30 in the evening peak, and 19:30-21:00 in the evening off-peak period. Further, to avoid the fatigue caused by the repeated driving on the same route and eliminate the impact of different driving styles on driving data, 28 drivers were selected to drive the experimental vehicles, who are familiar with road conditions, have rich driving experience, and a stable driving style. BYD E6, battery electric passenger cars with a high penetration rate in Xi'an, was adopted in the experiment. Fig. 2 shows the experimental vehicle and main equipment, including GPS, OBD, and VBOX, with a frequency of 1Hz.


**FIGURE 2.** Experimental vehicle and main equipment.

### B. DATA PREPROCESSING

Considering the quality and continuity of data, we cannot directly use the original driving data to develop the target cycle [32]. For example, the speed of some points is greater than 100km/h, and some discontinuous speed leads to abnormal acceleration. As shown in Fig. 3, the maximum and minimum acceleration of this part are $8\text{m/s}^2$ and $-10\text{m/s}^2$, respectively, which are beyond the range that can be achieved by normal driving of vehicles. Due to the sensor noise, the speed may not be zero when the BEVs are stationary, and the speed less than 0.5km/h was set as 0. In light of the speed limit of the experimental route by regulations, the speed exceeded 70km/h was eliminated. Besides, the acceleration threshold values were set as $-2.3\text{m/s}^2$ and $2.1\text{m/s}^2$, respectively. For the points where the acceleration exceeded the threshold, the mean speed at the previous second and the next second was taken as the current speed, and the acceleration was recalculated. After 150 iterations, the speed and acceleration of all time points were in a reasonable range. Fig. 3 shows the comparison of partial speed and acceleration before and after preprocessing.

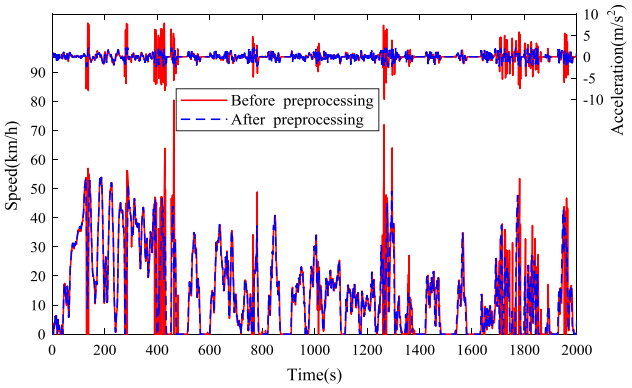


**FIGURE 3.** Comparison of speed and acceleration.

For ICEVs, the driving data are generally divided into four different kinematic fragments, including idling, accelerating, cruising, and decelerating. Somewhat differently, BEVs have stopping fragments instead of idling fragments. The segmentation of the kinematic fragments requires a comprehensive consideration of speed and acceleration. Given the corresponding threshold values, the specific division criteria as

$$\begin{cases} v = 0 \ \& \ \text{abs}(a) \le 0.15, & \text{stopping} \\ v > 0 \ \& \ a > 0.15, & \text{accelerating} \\ v > 0 \ \& \ a < -0.15, & \text{decelerating} \\ v > 0 \ \& \ \text{abs}(a) \le 0.15. & \text{cruising} \end{cases} \tag{1}$$

where $v$ (km/h) and $a$ ($\text{m/s}^2$) are the speed and acceleration of the vehicle, respectively.

## III. CONSTRUCTION OF CANDIDATE CYCLES

Based on our previous research [27], [33], we proposed a novel and effective method for constructing the candidate

cycles. First, the micro-trips are generated and their main characteristic parameters are calculated. Second, KPCA is used to reduce the dimension of characteristic parameters for micro-trips, then the reduced principal components score is clustered into specific categories by the combination of K-means clustering and RF. Finally, the candidate cycle is constructed by splicing several micro-trips. Furthermore, it is worth noting that the preprocessed driving data are divided into two parts, randomly selecting 80% to construct the target cycle, and the remaining 20% are used for validation.

### A. THE GENERATION OF MICRO-TRIPS

As the elementary blocks for driving cycles, the micro-trip of ICEVs refers to the speed-time profile between two consecutive idling fragments, including idling, accelerating, cruising, and decelerating phases [31]. Nevertheless, the micro-trip of BEVs was defined as the trip between two consecutive stopping fragments in this study, starting and ending with the speed of 0. We obtained 1414 valid micro-trips by excluding those with running time less than 10s. For the generated micro-trips, appropriate characteristic parameters need to be extracted to reflect the driving feature fully. Different parameters have been reported in the literature [4], [32], [34], and 14 parameters related to time, speed, and acceleration were extracted in this paper to describe micro-trips are shown in Table 1. The values of characteristic parameters are given in Table 2.

**TABLE 1. Characteristic parameters for describing micro-trips.**

| Description | Symbol | Definition | Unit |
|---|---|---|---|
| Time-related | $T$ | Driving time | s |
| | $T_a$ | Acceleration time | s |
| | $T_d$ | Deceleration time | s |
| | $T_c$ | Cruising time | s |
| | $T_s$ | Stopping time | s |
| Speed-related | $V_{\max}$ | Maximum speed | km/h |
| | $V_m$ | Mean speed | km/h |
| | $V_{mr}$ | Mean speed (stopping excluded) | km/h |
| | $V_{std}$ | Standard deviation of speed | km/h |
| Acceleration-related | $A_{\max}$ | Maximum acceleration | m/s$^2$ |
| | $A_m$ | Mean acceleration | m/s$^2$ |
| | $D_{\min}$ | Minimum deceleration | m/s$^2$ |
| | $D_m$ | Mean deceleration | m/s$^2$ |
| | $A_{std}$ | Standard deviation of acceleration and deceleration | m/s$^2$ |

### B. CHARACTERISTIC PARAMETERS DIMENSIONALITY REDUCTION

We depicted and calculated 14 parameters of micro-trips, and these parameters had a significant correlation with each other. Therefore, it is necessary to search for new low-dimensional features independent of each other and contain most of the vital information of the original data to avoid the dimensionality curse and feature redundancy. These low-dimensional features are not limited to a subset of the original features. In many cases, they are a linear or nonlinear combination of the original features.

PCA has been extensively applied as a linear dimensionality reduction method due to well reconfigurability and separability with fewer calculations [35]. PCA is calculated based on the covariance matrix of the observed data, and the widely existing nonlinear problems cannot be handled adequately. KPCA is an improved algorithm of PCA by employing nonlinear methods to extract principal components, among which the kernel function is the link between linear and nonlinear [36]. In recent years, KPCA has been widely used in nonlinear problems such as feature extraction and regression [37]–[39]. Assuming that $X$ is the sample set in the input space, the samples in $X$ are mapped to the high-dimensional Hilbert space by the nonlinear mapping. Then PCA is implicitly performed in this high-dimensional feature space. The kernel method is based on the inner product transformation of vectors as

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \tag{2}$$

where $K$ is the kernel function; $x_i$ and $x_j$ are samples belonging to $X$, and $\Phi$ is the mapping function from the input space to the high-dimensional feature space.

The kernel function is a continuous kernel of a positive integral operator satisfying the Mercer theorem [40]. Several representative kernel functions can be seen in Table 3.

In this paper, the detailed procedures of using KPCA to reduce the dimensionality of micro-trips characteristic parameters are given as follows.

1) For eliminating the influence of the order of magnitude on the results, the characteristic parameters matrix with the size of 1414 × 14 was standardized;

2) In KPCA, selecting a kernel function and determining the related parameters are the most important because the degree of capturing nonlinear characteristic of the data is dependent on them. This paper was carried out with a Gaussian RBF kernel, in which $\sigma$ represents the width of the Gaussian function. A relatively suitable parameter $\sigma$ as 38.73 was obtained through parameter optimization, and the kernel matrix $K_m$ was calculated;

3) The centralizer of $K_m$ was calculated by (3), and its eigenvalues and eigenvectors were calculated;

$$K_c = K_m - A \times K_m - K_m \times A + A \times K_m \times A \tag{3}$$

where $K_c$ is the centralizer of $K_m$; $A$ is the matrix with the size of $m \times m$, in which all elements equal to $1/m$, and $m$ is the number of samples in the input space. The value of $m$ is 1414 in this paper.

4) Results of KPCA are shown in Table 4 (only the first five principal components are given). The first four principal components were selected according to the criteria of variance larger than 1 and the cumulative probability of variance larger than 85%;

5) In the end, the principal component score matrix was calculated as the variable of output space.

For comparison, PCA was also employed to reduce the dimensionality of characteristic parameters. The results showed that the number of principal components was 3,

**TABLE 2.** Characteristic parameters of micro-trips.

| Number | $T$ | $T_a$ | $T_d$ | $T_c$ | $T_s$ | $V_{\max}$ | $V_m$ | $V_{mr}$ | $V_{std}$ | $A_{\max}$ | $A_m$ | $D_{\min}$ | $D_m$ | $A_{std}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 132 | 16 | 12 | 14 | 90 | 16.45 | 2.57 | 8.08 | 4.61 | 1.25 | 0.58 | -2.13 | -0.86 | 0.40 |
| 2 | 77 | 36 | 27 | 7 | 7 | 68.62 | 33.74 | 37.11 | 22.04 | 1.72 | 0.67 | -2.30 | -0.97 | 0.91 |
| 3 | 63 | 13 | 15 | 4 | 31 | 26.96 | 7.36 | 14.50 | 9.21 | 1.35 | 0.68 | -1.32 | -0.66 | 0.51 |
| 4 | 57 | 15 | 15 | 1 | 26 | 26.57 | 8.50 | 15.63 | 10.04 | 1.26 | 0.59 | -1.70 | -0.65 | 0.53 |
| 5 | 50 | 18 | 23 | 8 | 1 | 33.36 | 21.38 | 21.82 | 8.58 | 2.05 | 0.80 | -1.87 | -0.70 | 0.83 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| 1414 | 55 | 14 | 16 | 15 | 10 | 10.22 | 3.37 | 4.12 | 3.18 | 0.59 | 0.39 | -0.87 | -0.33 | 0.30 |

**TABLE 3.** Representative kernel functions.

| | |
|---|---|
| Linear kernel | $K(x_i,x_j) = x_i \cdot x_j$ |
| Polynomial kernel | $K(x_i,x_j) = [(x_i \cdot x_j)+1]^p$ |
| Gaussian radial basis function (RBF) kernel | $K(x_i,x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})$ |
| Multilayer perceptron kernel | $K(x_i,x_j) = \tanh(v(x_i \cdot x_j)+c)$ |

**TABLE 4.** Results of KPCA-total variance explained.

| Component | Variance | %of Variance | Cumulative (%) |
|---|---|---|---|
| 1 | 8.1637 | 52.23 | 52.23 |
| 2 | 3.8064 | 24.35 | 76.58 |
| 3 | 1.2587 | 8.05 | 84.63 |
| 4 | 1.1455 | 7.33 | 91.96 |
| 5 | 0.4410 | 2.82 | 94.78 |

and the cumulative variance contribution rate was 85.49%. However, KPCA selected 4 principal components, and the cumulative variance contribution rate was 91.96%. It can be concluded that KPCA dimensionality reduction can reflect the nonlinear characteristics of the data while containing more original data information.

### C. CLUSTERING OF MICRO-TRIPS BASED ON THE COMBINATION OF K-MEANS AND RANDOM FOREST

K-means clustering takes the local minimum of the sum of squares of errors as the objective function [41]. Although it is widely used because of its simplicity, easy implementation, and good interpretability, K-means clustering is hard to achieve global optimization. Therefore, we introduced the RF algorithm to optimize and modify the K-means clustering results. RF is a classification algorithm proposed by Breiman in 2001, which integrates multiple decision trees by ensemble learning to improve the generalization ability of the classification method [42]. Compared with other classification algorithms, RF can run effectively on large data sets due to parallel training implementation. Besides, RF can handle the data with high feature dimensions. On account of these advantages, RF has been studied extensively in classification, regression, and data mining [43], [44]. The procedures of RF classification are as follows.

1) If the number of samples in the training data set is $k$, $k$ samples are extracted from the training data set as the samples at the root node of each decision tree by bootstrap aggregating sampling. Each sample set has the same size with different contents. In each round of random sampling, about 36.8% of the training data are not selected, that is, data of out-of-bag;

2) If the number of variables is $m$ in the sample set, $n$ ($n \leqslant m$) is randomly selected as the number of variables when the nodes of the decision tree are split, then the information contained in $n$ variables is calculated, selecting the variables with the most classification ability as the node split features;

3) Each tree grows as much as possible without pruning;

4) A forest is generated by building many decision trees. For classification problems, each decision tree is a classifier, and the classification is determined by voting based on the classification results of each tree. As a consequence, the classification effect of RF is better than that of any single classifier due to the combination of multiple classifiers.

There are three steps in the clustering of micro-trips by K-means and RF. In the first step, we determined three categories of low-speed, medium-speed, and high-speed by analyzing the actual driving behavior. In the second step, the K-means clustering algorithm was applied in the training of RF. K-means clustering was used to cluster the micro-trips into three categories, in which 30 micro-trips closest to the center of the category were selected as the training samples. In the third step, the classification prediction of the remaining micro-trips was completed by learning the input micro-trips in the RF.

The classification effect of RF mainly depends on two factors. One is the correlation between any two trees in the forest, and a greater correlation will worsen the classification effect. The other is the classification effect of each tree in the forest, and the better the classification effect of each tree, the better the classification effect of the whole forest. Both factors are related to the number $n$ of the selected variables, and $n$ was taken as 3 in this paper. The classification results of K-means clustering combined with RF (KR) and K-means clustering are shown in Fig. 4 and Table 5. Fig. 4 shows the scatter plots of the $V_m$, $A_m$, and $D_m$ of all micro-trips belonging to different categories. It can be seen that the results of clustering using only K-means are different from the KR. Apparently, the micro-trips that initially belonged to the second category are assigned to the first category after KR clustering, which may be due to the more significant correlation between these micro-trips and the second category.

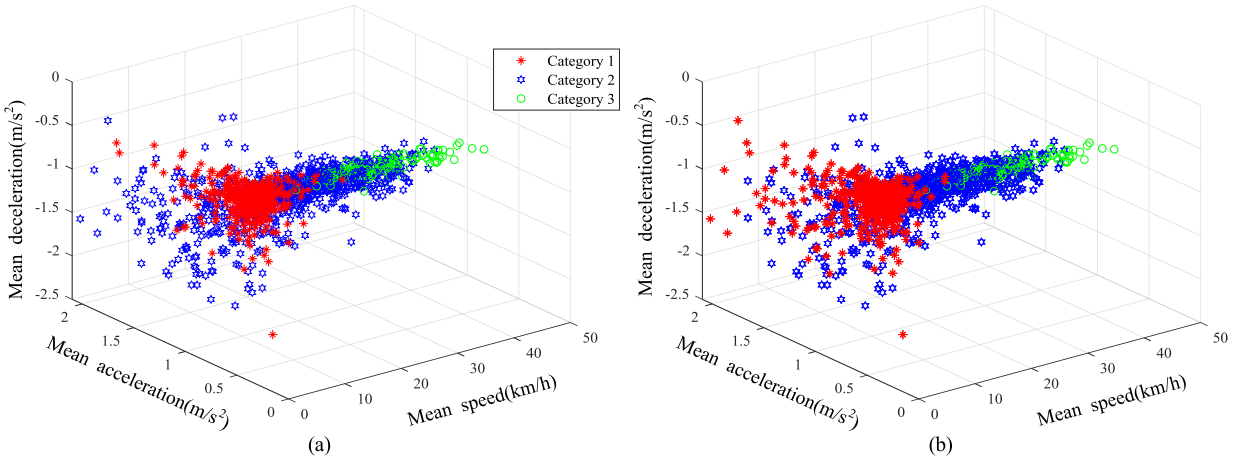


**FIGURE 4.** The results of clustering. (a) K-means; (b) KR.

**TABLE 5.** K-means and KR classification results.

| Parameter | | $V_m$ | $A_m$ | $D_m$ |
|---|---|---|---|---|
| K-means | Category 1 | 2.98 | 0.53 | -0.54 |
| | Category 2 | 16.13 | 0.90 | -0.92 |
| | Category 3 | 27.35 | 0.68 | -0.70 |
| KR | Category 1 | 2.81 | 0.61 | -0.61 |
| | Category 2 | 17.05 | 0.86 | -0.89 |
| | Category 3 | 27.44 | 0.68 | -0.70 |

As listed in Table 5, the parameter values characterize the average characteristics of each category. The parameters of the same category obtained by the K-means and KR are slightly different. The results of KR show that the $V_m$ of the first category is very low, with the value of 2.81km/h, the $A_m$ and $D_m$ are 0.61 m/s$^2$ and $-0.61$m/s$^2$, respectively. The second category has a $V_m$ of 17.05km/h, and it has the largest $A_m$ and $D_m$. The third category has the highest $V_m$, and the $A_m$ and $D_m$ are in the middle. It can be concluded that three categories respectively corresponded to the actual congestion, normal and smooth driving modes, representing the low-speed, medium-speed, and high-speed conditions.

We discussed the validity metrics from Compactness (*CP*), Separation (*SP*), Davies-Bouldin Index (*DB*), and Dunn Validity Index (*DVI*) to quantitatively evaluate the clustering performance of different methods [45].

(1) As an internal validation index of clusters, *CP* of each cluster is calculated as the average 1-norm distance between all samples and its cluster center as

$$\overline{CP}_i = \frac{1}{|\Omega_i|} \sum_{x_i \in \Omega_i} \|x_i - w_i\| \tag{4}$$

where $\overline{CP}_i$ is the compactness of cluster $i$; $\Omega_i$ represents the cluster $i$; $|\Omega_i|$ is the number of samples contained in $\Omega_i$; $x_i$ is the sample in $\Omega_i$; $w_i$ is the center of $\Omega_i$. The *CP* of global variables is expressed as

$$CP = \frac{1}{K} \sum_{i=1}^{K} \overline{CP}_i \tag{5}$$

where $K$ denotes the number of clusters. The lower value of *CP*, the smaller the distance within the clusters.

(2) *SP* is defined as the average Euclidean distance between the centers of every two clusters as (6). The higher value of *SP*, the greater the distance between clusters.

$$SP = \frac{2}{K^2 - K} \sum_{i=1}^{K} \sum_{j=i+1}^{K} \|w_i - w_j\|_2 \tag{6}$$

(3) *DB* measures distance both between and within the clusters. Divide the sum of the average distances between any two clusters by the Euclidean distance between two cluster centers, and then find the maximum value and divide by the number of clusters to get the *DB* as

$$DB = \frac{1}{K} \sum_{i=1}^{K} \max_{j \neq i} \left( \frac{\overline{C_i} + \overline{C_j}}{\|w_i - w_j\|_2} \right) \tag{7}$$

where $\overline{C_i}$ is the average distance between samples in $\Omega_i$. The lower value of *DB*, the higher similarity within the clusters, and the higher dissimilarity between the clusters.

(4) *DVI* is defined as the minimum distance between samples (between clusters) divided by the maximum distance between samples (within clusters). The higher value of *DVI*, the higher similarity within the clusters, and the higher dissimilarity between the clusters. *DVI* is calculated as

$$DVI = \frac{\min\limits_{0<m\neq n\leq K} \left\{ \min\limits_{\substack{\forall x_i \in \Omega_m \\ \forall x_j \in \Omega_n}} \{\|x_i - x_j\|\} \right\}}{\max\limits_{0<m\leq K} \max\limits_{\forall x_i, x_j \in \Omega_m} \{\|x_i - x_j\|\}} \tag{8}$$

**TABLE 6.** The clustering validity metrics.

| Metrics | *CP* | *SP* | *DB* | *DVI* |
|---|---|---|---|---|
| K-means | 2.385 | 6.087 | 1.403 | 0.440 |
| KR | 2.093 | 6.254 | 1.322 | 0.579 |

Table 6 shows the classification validity metrics of K-means and KR methods. Compared with K-means clustering, the *CP* and *DB* of KR clustering are smaller, whereas *SP* and *DVI* values are larger. It is proved that the proposed hybrid clustering algorithm can effectively improve the similarity within the same cluster and reduce the similarity between the different clusters.

### D. DETERMINE THE DURATION OF EACH CATEGORY

The determination of cycle duration is an essential part of the construction driving cycles. Durations of cycles are different, for instance, the total duration of NEDC, WLTC, and CLTC-P are 1180s, 1800s, and 1800s, respectively, and the non-legislative cycles are mostly between 900-1800s [14], [23], [26]. The duration was determined to be 1200s for Xi'an BEVs urban cycle (XBUC). XBUC consisted of three categories, and the duration of each category was obtained by multiplying the total duration by the time proportion of each corresponding category as

$$T_i = \frac{T_{\Omega_i}}{\sum_{i=1}^{K} T_{\Omega_i}} \times T_{total} \tag{9}$$

where $T_{\Omega i}$ represents the duration of $\Omega_i$; $T_{total}$ denotes the total duration of the cycle.

Furthermore, micro-trips were selected based on its Pearson correlation coefficient with its category. The larger the correlation coefficient is, the greater the micro-trip can represent the overall characteristics of its category. The calculation of the correlation coefficient as

$$r(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \tag{10}$$

where $Cov(X, Y)$ is the covariance of $X$ and $Y$; $\sigma_X$ and $\sigma_Y$ are the standard deviations of $X$ and $Y$, respectively.

Based on the Pearson correlation coefficient and the run time of each micro-trip, a few appropriate micro-trips were selected in each category to represent the characteristics of the belonging category. We constructed 10 candidate driving cycles by splicing the selected micro-trips.

## IV. DEVELOPMENT OF XI'AN BEVS URBAN CYCLE

By evaluating and comparing the performance value of the candidate cycles, a target cycle is selected from them to represent the original driving characteristic best. The determination of assessment criteria in this study as follows.

### A. DETERMINATION OF ASSESSMENT CRITERIA

It is of great significance to determine a set of rational and representative assessment criteria for the development of driving cycles when selecting the best cycle from candidate cycles according to the statistical representativeness [22]. Based on the maximum similarity principle, we took the constructed 10 candidate cycles as input and adopted the following steps to output the best cycle.

The first step is to give the indication parameters, which are different from the micro-trips characteristic parameters and can be used to characterize the cycle characteristics. The indication parameters including $V_m$, $V_{mr}$, $A_m$, $D_m$, acceleration proportion ($P_a$), deceleration proportion ($P_d$), cruising proportion ($P_c$), and stopping proportion ($P_s$) (defined as the ratio of the time of each kinematic fragment to the total driving time, for instance, $P_a = T_a/T \times 100\%$), and the relative positive acceleration (RPA) as well as positive kinetic energy (PKE) related to dynamics. RPA is an important parameter to describe the power demand of the driving cycle profile as [9]

$$\mathrm{RPA} = \frac{1}{dist} \int_0^T v_t \cdot a_t^+ dt \tag{11}$$

where *dist* (m) is the distance of the whole trip; $T$ is the driving time of the trip; $v_t$ is the instantaneous speed of the moment $t$; $a_t^+$ is the instantaneous acceleration of the moment $t$, only the positive acceleration is considered.

PKE is a parameter describing the positive kinetic energy of vehicles, and its calculation as [15]

$$\mathrm{PKE} = \frac{1}{dist} \sum_{t=2}^{T} v_t^2 - v_{t-1}^2 \tag{12}$$

where only the moment $v_t > v_{t-1}$ is considered.

The absolute relative error (ARE) of 10 indication parameters were calculated between the candidate cycles and the driving data in the second step. The candidate cycles were then selected when the ARE of each indicator parameter less than 10% [34], and the mean absolute relative error (MARE) of each candidate cycle was calculated. Speed-acceleration probability distribution (SAPD) is a parameter to describe the speed and acceleration states [4], and the $\mathrm{SAPD_{diff}}$ is defined as the percentage error of the frequency between the driving data and other cycles in different speed-acceleration bins as

$$\mathrm{SAPD_{diff}} = \frac{\sum i(\mathrm{SAPD_{cycle}}(i)\text{-}\mathrm{SAPD_{data}}(i))^2}{\mathrm{SAPD_{data}}(i)^2} \tag{13}$$

where $i$ is the bin of SAPD; $\mathrm{SAPD_{cycle}}$ is the SAPD of candidate cycle; $\mathrm{SAPD_{data}}$ is the SAPD of driving data.

Finally, the performance value (PV) of candidate cycles was defined as the mean value of MARE and $\mathrm{SAPD_{diff}}$. The candidate cycle with the minimum PV was selected as the target cycle, that is, the XBUC.

### B. MAIN CHARACTERISTICS OF XI'AN BEVS URBAN CYCLE

The total duration of XBUC is 1200s, including the low-speed category of 160s, medium-speed category of 603s, and high-speed category of 437s. The speed and acceleration profiles of XBUC are shown in Fig. 5. The main characteristics of XBUC and its three categories are shown in Table 7.

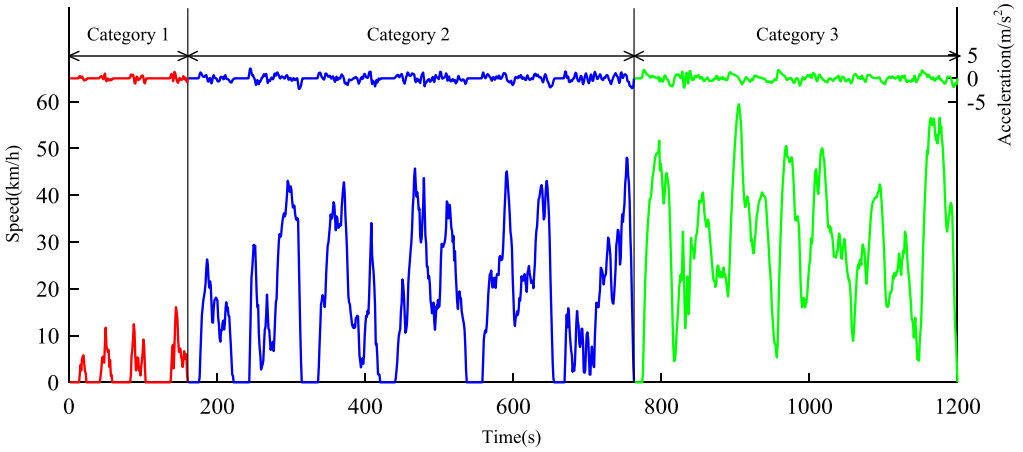


**FIGURE 5.** Speed and acceleration profiles of XBUC.

**TABLE 7.** Main characteristics of XBUC.

| Characteristics | XBUC | Category 1 | Category 2 | Category 3 |
|---|---|---|---|---|
| $T$ | 1200 | 160 | 603 | 437 |
| $Dist$[a] | 6719 | 107 | 2902 | 3710 |
| $V_{max}$ | 59.452 | 16.071 | 48.034 | 59.452 |
| $V_m$ | 20.159 | 2.413 | 17.328 | 30.563 |
| $V_{mr}$ | 24.313 | 5.218 | 21.194 | 31.352 |
| $A_m$ | 0.620 | 0.520 | 0.645 | 0.607 |
| $D_m$ | -0.659 | -0.452 | -0.709 | -0.645 |
| $P_a$ (%) | 34.58 | 18.13 | 35.16 | 39.82 |
| $P_d$ (%) | 32.59 | 20.62 | 31.84 | 37.52 |
| $P_c$ (%) | 15.75 | 7.50 | 14.76 | 20.14 |
| $P_s$ (%) | 17.08 | 53.75 | 18.24 | 2.52 |

[a]*Dist* (m) is the distance of the trip.

It can be seen from Table 7 that the $V_m$ of the first category is the smallest, and the $P_s$ is the largest, up to 53.75%, which can be inferred that the corresponding driving condition is congested. Compared with the first category, the second category has a larger $V_m$ and $P_c$, and its $P_s$ is smaller, which indicates that the driving condition is relatively smooth. The third category has the highest $V_m$ and $P_c$, which means that the corresponding condition is smooth.

To further illustrate the speed-acceleration distribution, this paper presents SAPDs of the driving data and XBUC, as shown in Fig. 6. It can be included that XBUC and driving data have very similar SAPD in that the probability is high when the acceleration is in the range of [−0.5, 0.5] and the speed is very low, whereas the rest of the bins are close to 0. The $SAPD_{diff}$ between XBUC and overall driving data, driving data for development and validation are 3.03%, 3.05%, and 3.52%, respectively, which further indicates the applicability of the developed cycle.

## V. VALIDATION AND COMPARISON OF DRIVING CYCLES

To verify the representativeness and typicality of the developed driving cycle, we compared PCA and K-means clustering with the proposed method and studied the discrepancies between the XBUC and several legislative cycles.

### A. COMPARATIVE ANALYSIS OF DIFFERENT METHODS

Based on the same driving data for development and validation, the target cycle was also developed by the conventional PCA and K-means clustering (driving cycle-PK), as shown in Fig. 7. The duration of the driving cycle-PK is the same as XBUC and also consists of three categories. The characteristic parameters of the target cycles based on the two methods are shown in Table 8. PVs between XBUC and the driving data used for development and validation are 2.72% and 2.95%, respectively, whereas the PVs between driving cycle-PK and driving data are 3.64% and 3.81%, respectively.

**TABLE 8.** Comparison of characteristic parameters.

| Parameters | D-data[a] | V-data[b] | XBUC | Driving cycle-PK |
|---|---|---|---|---|
| $V_m$ | 20.062 | 19.293 | 20.159 | 18.518 |
| $V_{mr}$ | 24.377 | 23.712 | 24.313 | 22.515 |
| $A_m$ | 0.633 | 0.622 | 0.620 | 0.588 |
| $D_m$ | -0.646 | -0.648 | -0.659 | -0.594 |
| $P_a$ (%) | 33.04 | 32.87 | 34.58 | 32.42 |
| $P_d$ (%) | 32.33 | 31.48 | 32.59 | 32.16 |
| $P_c$ (%) | 16.93 | 17.02 | 15.75 | 17.67 |
| $P_s$ (%) | 17.70 | 18.63 | 17.08 | 17.75 |
| PKE | 0.595 | 0.582 | 0.576 | 0.547 |
| RPA | 0.256 | 0.250 | 0.256 | 0.238 |
| PV(&D-data) (%) | N/A | N/A | 2.72 | 3.64 |
| PV(&V-data) (%) | N/A | N/A | 2.95 | 3.81 |

[a]D-data is the driving data for developing the target cycle; [b]V-data is the remaining driving data for validating the target cycle.

Meanwhile, Fig. 8 shows the proportion of acceleration, deceleration, cruising, and stopping visually. The distribution of acceleration, deceleration, cruising, and stopping of the two target cycles are both very similar to that of the driving data. Relatively speaking, the proportion of acceleration and deceleration of XBUC are slightly larger.

The results show that the target cycles based on the two methods can both reflect the driving characteristics. Still, the proposed method is more in line with the real vehicle driving data, which proves its availability and reliability.

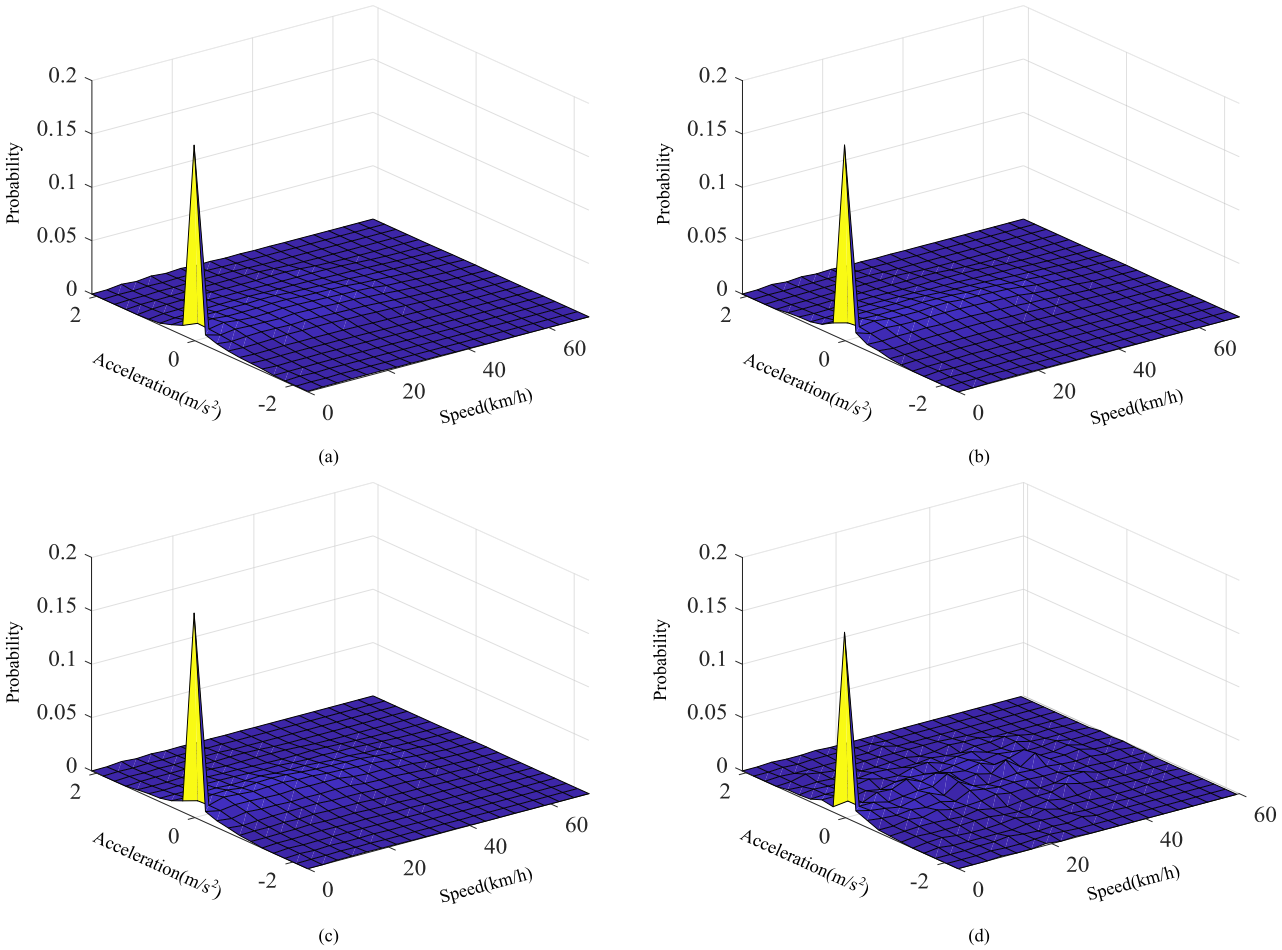


**FIGURE 6.** Comparison of SAPDs. (a) Overall driving data; (b) Driving data for development; (c) Driving data for validation; (d) XBUC.

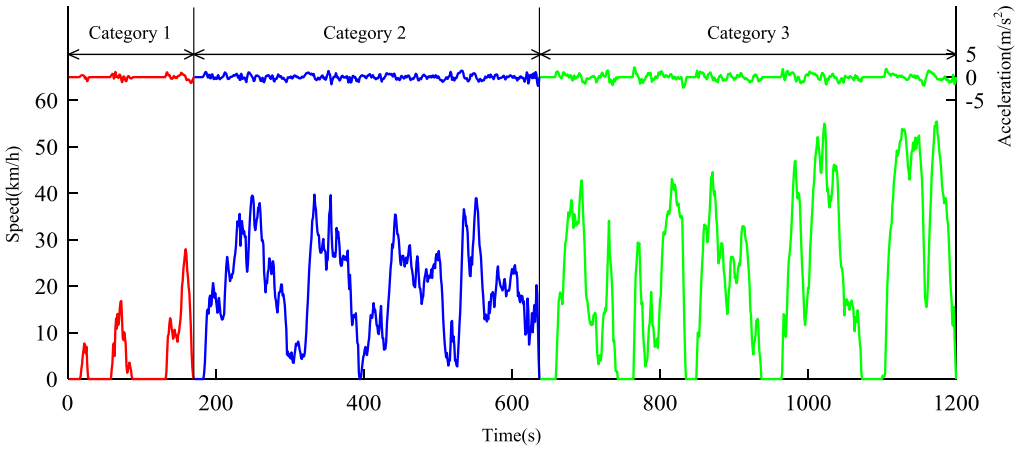


**FIGURE 7.** Speed and acceleration profiles of driving cycle-PK.

### *B. COMPARATIVE ANALYSIS BETWEEN XI'AN BEVS URBAN CYCLE AND LEGISLATIVE CYCLES*

Three legislative cycles were selected to compare with XBUC in this section. Table 9 shows the characteristic parameters of different cycles.

As shown in (1), the same kinematic fragments division principle is used to calculate the characteristic parameters of the three legislative cycles. According to Table 9, apparent differences exist between the XBUC and three legislative cycles. Apparently, the research objects and duration of

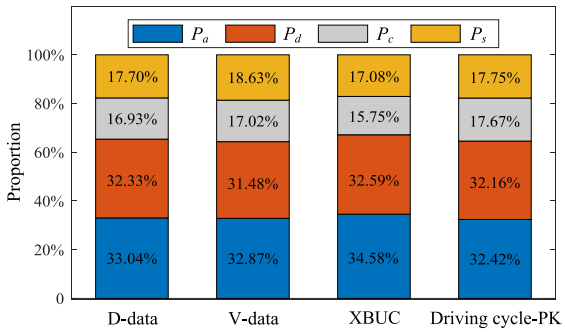


**FIGURE 8.** $P_a$, $P_d$, $P_c$, and $P_s$ of the target cycles and driving data.

**TABLE 9.** Comparison of XBUC and legislative cycles.

| Parameters | NEDC | WLTC | CLTC-P | XBUC |
|---|---|---|---|---|
| Vehicle type | ICEVs | LDVs | LDVs | BEVs |
| $T$ | 1180 | 1800 | 1800 | 1200 |
| $V_m$ | 33.642 | 46.533 | 28.956 | 20.159 |
| $V_{mr}$ | 43.484 | 53.282 | 37.176 | 24.313 |
| $A_m$ | 0.533 | 0.542 | 0.450 | 0.620 |
| $D_m$ | -0.748 | -0.577 | -0.494 | -0.659 |
| $P_a$ (%) | 23.22 | 30.61 | 28.61 | 34.58 |
| $P_d$ (%) | 16.61 | 28.94 | 26.45 | 32.59 |
| $P_c$ (%) | 37.54 | 27.78 | 22.83 | 15.75 |
| $P_s$ (%) | 22.63 | 12.67 | 22.11 | 17.08 |
| PKE | 0.222 | 0.308 | 0.347 | 0.576 |
| RPA | 0.111 | 0.152 | 0.166 | 0.256 |

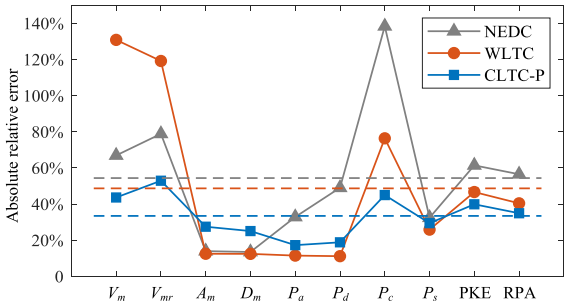


**FIGURE 9.** The absolute relative error of characteristic parameters.

XBUC and NEDC, WLTC, CLTC-P are different. To further analyze the differences between XBUC and three legislative cycles, ARE of characteristic parameters are shown in Fig. 9. The dotted lines in Fig. 9 show the MARE between XBUC and NEDC, WLTC, CLTC-P are 54.38%, 48.69%, and 33.44%, respectively. It can be inferred that XBUC and CLTC-P have the most similar driving characteristics, followed by WLTC and NEDC. The parameters of $V_m$, $V_{mr}$, and $P_c$ show the most significant difference between XBUC and three legislative cycles. ARE of $A_m$ and $D_m$ for three legislative cycles are the smallest, but the proportion of kinematic fragments is significantly different from that of XBUC. Quantitative analyses for parameters show that operating characteristics of XBUC are lower speed, more violent acceleration and deceleration, and higher PKE and RPA, which indicates that the urban traffic condition is even worse. To further explain the speed and acceleration distributions of XBUC and three legislative cycles, Fig. 10 shows box plots of the speed and acceleration, and the speed-acceleration probability distribution is shown in Fig. 11.

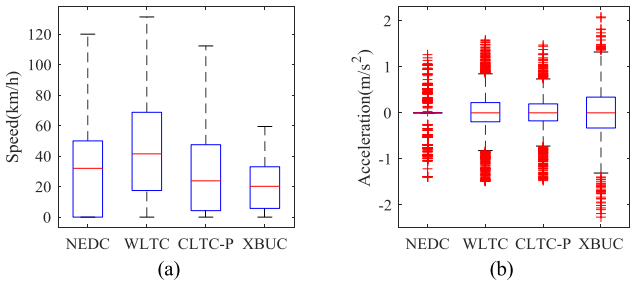


**FIGURE 10.** Box plots of speed and acceleration distribution. (a) Speed; (b) Acceleration.

From Fig. 10(a), the minimum values of speed are equal to 0 for all cycles. The 25th percentile range between 0 and 17.7km/h. The median values equal 32.0km/h for NEDC, 41.5km/h for WLTC, 23.8km/h for CLTC-P, and 19.5km/h for XBUC. The 75th percentile range between 32.6km/h and 68.8km/h. The maximum values range between 59.5km/h and 131.3km/h. It can be inferred from the interquartile range that the speed distribution of WLTC is the most dispersed, whereas that of XBUC is the most concentrated. It can be observed from Fig. 10(b) that the median values of acceleration are equal to 0 for all cycles. The acceleration distribution of XBUC is the most dispersed, whereas the acceleration of NEDC is the most concentrated. According to the comprehensive analysis of the above results, BEVs have unique driving characteristics and have greater acceleration than ICEVs. It can be concluded from Fig. 11 that NEDC, WLTC, CLTC-P, and XBUC have one thing in common, that is, the distribution probability in the range of speed [0, 5] and acceleration [−0.2, 0.2] is the highest, whereas the probability of other bins is small. The main difference is that NEDC and WLTC have 6 and 3 distinct peaks, respectively, whereas CLTC-P and XBUC have only 1 distinct peak. Although the SAPD of CLTC-P and XBUC are similar, their speed and acceleration ranges are different. Specifically, the speed range of XBUC is smaller, but the acceleration range is larger, reflecting the more congested traffic conditions in the city. Compared with XBUC, three legislative cycles belong to composite cycles, which integrate the driving characteristics of urban and suburban areas, so the speed is relatively high. The XBUC developed in this paper is mainly aimed at the operation condition of the urban environment, belonging to the urban driving cycles. The qualitative and quantitative discrepancies between XBUC and the three legislative cycles indicate the uniqueness of the XBUC and the necessity to develop the cycle.

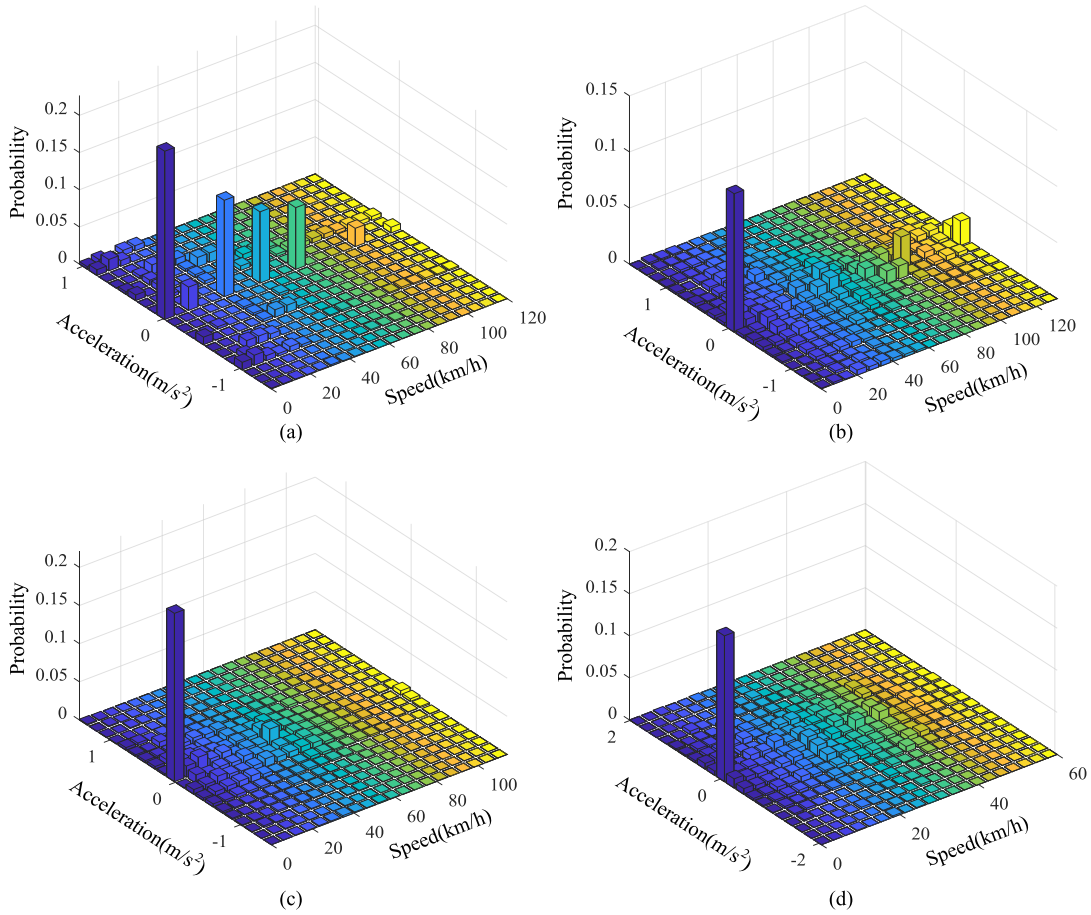


**FIGURE 11.** SAPDs of four cycles. (a) NEDC; (b) WLTC; (c) CLTC-P; (d) XBUC.

## VI. CONCLUSION

Based on kernel principal component analysis and random forest algorithm, we proposed a novel method to develop an urban driving cycle for battery electric vehicles, which provides an accurate input for estimating energy consumption and driving range. The experimental route was determined according to the qualitative and quantitative analysis rather than the subjective experience of the author. The real driving data in Xi'an city were collected by the combination of the chase car method and on-board measurement method for one week. The collected driving data were then randomly divided into two parts, 80% used to develop the target cycle, and the remaining data adopted to validate the target cycle.

Considering the deficiency of the conventional PCA and K-means clustering method, the dimension of micro-trips was reduced by the kernel principal component analysis, and an improved clustering method combined K-means with the random forest was adopted to cluster the micro-trips and construct the candidate cycles. The performance value was then defined and calculated, and the candidate cycle with the minimum performance value was selected as the target cycle, which can best reflect the urban traffic conditions and residents' travel characteristics.

At last, the comparative analysis of the different methods showed that the effectiveness of the proposed method. The significant discrepancies were found by comparing the target cycle to other legislative cycles, especially CLTC-P, and it can be inferred that the current legislative cycles may not apply to new energy vehicles. However, it is worth further thinking about whether the discrepancies are mainly attributed to the vehicle's power system characteristics or urban road properties and traffic congestion levels, or both. In future studies, it is imperative to develop the driving cycles for subdivided vehicle types in specific regions.

## REFERENCES

[1] X. Zhao, X. Zhao, Q. Yu, Y. Ye, and M. Yu, "Development of a representative urban driving cycle construction methodology for electric vehicles: A case study in Xi'an," *Transp. Res. D, Transp. Environ.*, vol. 81, Apr. 2020, Art. no. 102279.

[2] X. Liu, J. Ma, X. Zhao, J. Du, and Y. Xiong, "Study on driving cycle synthesis method for city buses considering random passenger load," *J. Adv. Transp.*, vol. 2020, pp. 1–21, Feb. 2020.

[3] W. T. Hung, H. Y. Tong, C. P. Lee, K. Ha, and L. Y. Pao, "Development of a practical driving cycle construction methodology: A case study in Hong Kong," *Transp. Res. D, Transp. Environ.*, vol. 12, no. 2, pp. 115–128, Mar. 2007.

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

[5] L. Berzi, M. Delogu, and M. Pierini, "Development of driving cycles for electric vehicles in the context of the city of Florence," *Transp. Res. D, Transp. Environ.*, vol. 47, pp. 299–322, Aug. 2016.
[6] H. Y. Tong, W. T. Hung, and C. S. Cheung, "Development of a driving cycle for Hong Kong," *Atmos. Environ.*, vol. 33, no. 15, pp. 2323–2335, Jul. 1999.
[7] J. Huertas, L. Quirama, M. Giraldo, and J. Díaz, "Comparison of three methods for constructing real driving cycles," *Energies*, vol. 12, no. 4, p. 665, Feb. 2019.
[8] N. T. Jeong, S. M. Yang, K. S. Kim, M. S. Wang, H. S. Kim, and M. W. Suh, "Urban driving cycle for performance evaluation of electric vehicles," *Int. J. Automot. Technol.*, vol. 17, no. 1, pp. 145–151, Feb. 2016.
[9] M. Tutuianu, P. Bonnel, B. Ciuffo, T. Haniu, N. Ichikawa, A. Marotta, J. Pavlovic, and H. Steven, "Development of the world-wide harmonized light duty test cycle (WLTC) and a possible pathway for its introduction in the European legislation," *Transp. Res. D, Transp. Environ.*, vol. 40, pp. 61–75, Oct. 2015.
[10] *China Automotive Test Cycle—Part 1: Light-Duty Vehicles*, Standard GB/T 38146.1-2019, 2019.
[11] *China Automotive Test Cycle—Part 2: Heavy-Duty Commercial Vehicles*, Standard GB/T 38146.2-2019, 2019.
[12] H. Kaymaz, H. Korkmaz, and H. Erdal, "Development of a driving cycle for Istanbul bus rapid transit based on real-world data using stratified sampling method," *Transport. Res. D, Transport. Environ.*, vol. 75, pp. 123–135, Oct. 2019.
[13] Z. Jing, G. Wang, S. Zhang, and C. Qiu, "Building Tianjin driving cycle based on linear discriminant analysis," *Transp. Res. D, Transp. Environ.*, vol. 53, pp. 78–87, Jun. 2017.
[14] Q. Shi, B. Liu, Q. Guan, L. He, and D. Qiu, "A genetic ant colony algorithm-based driving cycle generation approach for testing driving range of battery electric vehicle," *Adv. Mech. Eng.*, vol. 12, no. 1, Jan. 2020, Art. no. 168781401990105.
[15] J. I. Huertas, M. Giraldo, L. F. Quirama, and J. Diaz, "Driving cycles based on fuel consumption," *Energies*, vol. 11, no. 11, pp. 1–13, Nov. 2018.
[16] Y.-L.-T. Nguyen, T.-D. Nghiem, A.-T. Le, and N.-D. Bui, "Development of the typical driving cycle for buses in Hanoi, Vietnam," *J. Air Waste Manage. Assoc.*, vol. 69, no. 4, pp. 423–437, Apr. 2019.
[17] H. Achour and A. G. Olabi, "Driving cycle developments and their impacts on energy consumption of transportation," *J. Cleaner Prod.*, vol. 112, pp. 1778–1788, Jan. 2016.
[18] G.-H. Tzeng and J.-J. Chen, "Developing a taipei motorcycle driving cycle for emissions and fuel economy," *Transp. Res. D, Transp. Environ.*, vol. 3, no. 1, pp. 19–27, Jan. 1998.
[19] Q. Wang, H. Huo, K. He, Z. Yao, and Q. Zhang, "Characterization of vehicle driving patterns and development of driving cycles in Chinese cities," *Transp. Res. D, Transp. Environ.*, vol. 13, no. 5, pp. 289–297, Jul. 2008.
[20] C.-H. Lee and C.-H. Wu, "A novel big data modeling method for improving driving range estimation of EVs," *IEEE Access*, vol. 3, pp. 1980–1993, 2015.
[21] J. Bi, Y. Wang, Q. Sai, and C. Ding, "Estimating remaining driving range of battery electric vehicles based on real-world data: A case study of Beijing, China," *Energy*, vol. 169, pp. 833–843, Feb. 2019.
[22] H. Y. Tong and W. T. Hung, "A framework for developing driving cycles with on-road driving data," *Transp. Rev.*, vol. 30, no. 5, pp. 589–615, Sep. 2010.
[23] Y. Yang, T. Li, T. Zhang, and Q. Yu, "Time dimension analysis: Comparison of Nanjing local driving cycles in 2009 and 2017," *Sustain. Cities Soc.*, vol. 53, Feb. 2020, Art. no. 101949.
[24] W. Yan, M.-J. Li, Y.-C. Zhong, C.-Y. Qu, and G.-X. Li, "A novel k-MPSO clustering algorithm for the construction of typical driving cycles," *IEEE Access*, vol. 8, pp. 64028–64036, 2020.
[25] A. Ashtari, E. Bibeau, and S. Shahidinejad, "Using large driving record samples and a stochastic approach for real-world driving cycle construction: Winnipeg driving cycle," *Transp. Sci.*, vol. 48, no. 2, pp. 170–183, May 2014.
[26] N. H. Arun, S. Mahesh, G. Ramadurai, and S. M. S. Nagendra, "Development of driving cycles for passenger cars and motorcycles in Chennai, India," *Sustain. Cities Soc.*, vol. 32, pp. 508–512, Jul. 2017.
[27] X. Zhao, Q. Yu, J. Ma, Y. Wu, M. Yu, and Y. Ye, "Development of a representative EV urban driving cycle based on a k-means and SVM hybrid clustering algorithm," *J. Adv. Transp.*, vol. 2018, pp. 1–18, Nov. 2018.
[28] 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," *Atmos. Environ.*, vol. 97, pp. 353–362, Nov. 2014.
[29] Z. Xiao, Z. Dui-Jia, and S. Jun-Min, "A synthesis of methodologies and practices for developing driving cycles," *Energy Procedia*, vol. 16, pp. 1868–1873, Jan. 2012.
[30] Z. Yu, S. Shi, S. Wang, Y. Mu, W. Li, C. Xu, and M. Zhang, "Statistical inference-based research on sampling time of vehicle driving cycle experiments," *Transp. Res. D, Transp. Environ.*, vol. 54, pp. 114–141, Jul. 2017.
[31] S. H. Kamble, T. V. Mathew, and G. K. Sharma, "Development of real-world driving cycle: Case study of pune, india," *Transp. Res. D, Transp. Environ.*, vol. 14, no. 2, pp. 132–140, Mar. 2009.
[32] H. Gong, Y. Zou, Q. Yang, J. Fan, F. Sun, and D. Goehlich, "Generation of a driving cycle for battery electric vehicles: A case study of Beijing," *Energy*, vol. 150, pp. 901–912, May 2018.
[33] X. Zhao, J. Ma, S. Wang, Y. Ye, Y. Wu, and M. Yu, "Developing an electric vehicle urban driving cycle to study differences in energy consumption," *Environ. Sci. Pollut. Res.*, vol. 26, no. 14, pp. 13839–13853, May 2019.
[34] X. Zhao, Y. Ye, J. Ma, P. Shi, and H. Chen, "Construction of electric vehicle driving cycle for studying electric vehicle energy consumption and equivalent emissions," *Environ. Sci. Pollut. Res.*, vol. 27, no. 30, pp. 37395–37409, May 2020.
[35] I. T. Jolliffe, "Principal component analysis and factor analysis," in *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer, 2002, ch. 7, pp. 150–166, doi: 10.1007/0-387-22440-8_7.
[36] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
[37] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.
[38] R. Rosipal, M. Girolami, L. J. Trejo, and A. Cichocki, "Kernel PCA for feature extraction and de-noising in nonlinear regression," *Neural Comput. Appl.*, vol. 10, no. 3, pp. 231–243, Dec. 2001.
[39] Y. Zhang and C. Ma, "Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS," *Chem. Eng. Sci.*, vol. 66, no. 1, pp. 64–72, Jan. 2011.
[40] J. Mercer, "Functions of positive and negative type, and their connection with the theory of integral equations," in *Proc. Roy. Soc. London*, 1909, pp. 415–446.
[41] H. Ismkhan, "Ik-means−+: An iterative clustering algorithm based on an enhanced version of the k-means," *Pattern Recognit.*, vol. 79, pp. 402–413, Jul. 2018.
[42] L. Breiman, "Random forests," *Mach. Learn. Arch.*, vol. 45, no. 1, pp. 5–32, 2001.
[43] M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," *BMC Med. Informat. Decis. Making*, vol. 11, no. 1, p. 51, Jul. 2011.
[44] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern Recognit.*, vol. 44, no. 2, pp. 330–349, Feb. 2011.
[45] A. Fahad, N. Alshatri, Z. Tari, A. Alamri, I. Khalil, A. Y. Zomaya, S. Foufou, and A. Bouras, "A survey of clustering algorithms for big data: Taxonomy and empirical analysis," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 267–279, Sep. 2014.

**LU WANG** received the B.S. degree from Chang'an University, Xi'an, China, in 2016, where she is currently pursuing the Ph.D. degree in automobile engineering. Her research interests include vehicle driving cycle construction and power battery fault diagnosis.

**JIAN MA** received the Ph.D. degree in transportation engineering from Chang'an University, China, in 2001.

He is currently a Professor with the School of Automobile, Chang'an University. He has undertaken more than 30 government sponsored research projects, such as 863 projects and key transportation projects in Western China. He published more than 90 academic articles at home and abroad. He has authored four books. His main research interests include vehicle dynamics, electric vehicle and clean energy vehicle technology, and the automobile detection technology and theory.

**XUAN ZHAO** received the B.S., M.S., and Ph.D. degrees in vehicle engineering from Chang'an University, China, in 2007, 2009, and 2012, respectively. He is currently a Professor with the School of Automobile, Chang'an University. He has undertaken over nine government sponsored works, including the National Key Research and Development Program of China and the China Postdoctoral Science Foundation. His main research interests include electric vehicle control strategy and electric vehicle lightweight design.

**XUEBO LI** received the B.S. degree from Chang'an University, Xi'an, China, in 2016, where he is currently pursuing the Ph.D. degree in automobile engineering. His research interests include parameter state estimation and control of regenerative braking for electric vehicles.