future transportation

Article

The Construction of Xi'an Urban Bus Driving Cycle: A Case Study

Yaohua Li *, Dengwang Zhai, Hong Ding and Rajibul Islam

School of Automobile, Chang'an University, Xi'an 710064, China
* Correspondence: yaohua.li@chd.edu.cn

Abstract: At present, there are many methods to construct a vehicle driving cycle, such as the micro-trip-based method and the Markov chain method. Different methods have different advantages and disadvantages. To compare these methods, this paper uses the micro-trip-based method, the Markov chain method, and the method combining micro-trips and Markov chains to construct the representative driving cycle of a Xi'an urban bus based on the driving data of the Xi'an No.2 bus. Firstly, the driving data is collected and preprocessed. Then, representative driving cycles are constructed based on different methods. Finally, different driving cycles based on different methods are compared. By calculating and comparing characteristic parameters, velocity distributions, acceleration distributions, and vehicle-specific power distributions between different representative driving cycles, this paper shows the advantages and disadvantages of different construction methods.

Keywords: driving cycle; micro-trip-based method; Markov chain method; micro-trip-based and Markov chain method

check for updates

Citation: Li, Y.; Zhai, D.; Ding, H.; Islam, R. The Construction of Xi'an Urban Bus Driving Cycle: A Case Study. *Future Transp.* **2023**, *3*, 92–107. https://doi.org/10.3390/futuretransp3010006

Received: 28 October 2022
Revised: 8 January 2023
Accepted: 9 January 2023
Published: 11 January 2023

1. Introduction

The driving cycle is a speed–time profile that represents vehicle driving characteristics [1–8], which can be used to study vehicle parameter matching, control strategy optimization, fuel economy, and other aspects. In the development and testing of vehicles, engineers often adopt driving cycles. Because of differences in city size, geographical characteristics, road types, road topology, vehicle ownership, and other factors, the characteristics of the driving cycle in different cities and regions are different [9]. Therefore, when purchasing the urban bus, the relevant departments can select the most suitable bus for the city by referring to the fuel economy and other evaluation indicators of the vehicle tested under the representative driving cycle of the city [10]. Hence, it is important to select the appropriate method to accurately construct the driving cycle with different regional driving characteristics [11]. At present, many researchers have used different methods to construct the driving cycles of many regions. Some select micro-trip-based methods [12,13], some select the Markov chain method [9,14], and some combine several methods to construct driving cycles [15,16]. They all verified the feasibility of the methods they selected, but their conclusions are based on their data. The advantages of different methods are not comparable. Therefore, based on the same driving data, this paper adopts different methods to construct the driving conditions of the same route, and finds out the advantages and disadvantages of different methods through comparison, to provide references for the selection of methods to construct the working conditions.

The route intensity method is used to select representative routes. The bus route is made up of bus stations. The more frequently the bus station appears across all bus routes, the higher utilization the bus station has, and the more representative it is. The route intensity is shown in (1), where μ_i is the route intensity of the ith bus route, λ_j is the occurrence number of the jth bus station of the ith bus route among all bus routes, and n is the total number of bus stations on the ith bus route.

$$\mu_i = \frac{\sum_{j=1}^{n} \lambda_j}{n} \tag{1}$$

After analysis of the route intensity of all Xi'an bus lines and consultation with the Xi'an bus company, Bus Route No. 2 was selected among more than 200 bus routes. There are 33 bus stops, the origin station is in the southwest corner of the Second Ring Road in Xi'an, and the terminal station is in the northeast corner of the Second Ring Road. It generally runs from east to west, covers the First Ring Road and the Second Ring Road, and passes through the urban main road. The selected route can reflect the traffic conditions in urban areas in Xi'an.

In an actual situation, when the bus is in an idling condition, the speed collected by the GPS and CAN bus equipment is not necessarily zero, which will reduce the authenticity of the collection data and thus reduce the accuracy of the driving cycle. Therefore, idling calibration is required. Referring to relevant papers and considering the actual situation of the vehicle, we defined the data with a speed less than 1.5 km/h as idle speed data and assigned it to zero.

Due to various factors, such as bad weather and building obstruction, there will be a lot of noise in data collected by the GPS. Moreover, when the equipment is disturbed, the collected data may have an abnormal value, which will cause the vehicle acceleration to be greater than the true value. By consulting the bus driver and combining this with the actual situation, we know that the maximum acceleration of the bus will not be greater than 3 m/s^2 and the maximum deceleration will not be greater than -4 m/s^2. The data exceeding the limit point is regarded as cusp data, and linear interpolation is carried out several times until the cusp data is within the limit point.

The data collected by GPS and CAN bus equipment is continuous. However, there are some specific situations, such as waiting for departure and driving between the bus terminal and the bus company. These data are invalid for driving cycle construction, which will affect the accuracy of driving cycle construction. By investigating the record of the bus company dispatching station, the actual following car timing, and the monitoring platform track verification, these data are stripped from the continuous data to ensure the validity of the collected data.

2. Micro-Trip-Based Method

The flow diagram for constructing the Xi'an No.2 bus driving cycle by the micro-trip-based method is shown in Figure 1.

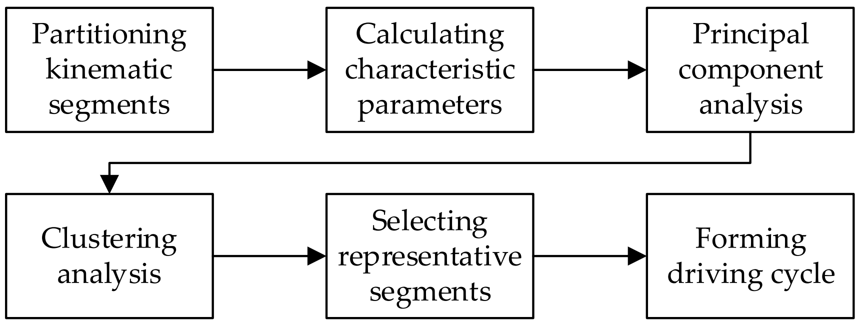


Figure 1. Flow diagram of driving cycle construction based on the micro-trip-based method.

2.1. Partitioning Micro-Trips

A kinematic segment is a driving process between two adjacent starting points (or stopping points), usually consisting of an idle part and a driving part [17]. This paper stipulates that the time length of a kinematic segment should not be less than 20 s, so micro-trips less than 20 s were combined with the following micro-trip until the time length requirement was met, so as to not miss some micro-trips with frequent starts and stops.

The amount of vehicle driving data has a huge impact on the subsequent construction of the driving cycle. Generally speaking, the larger the amount of data, the closer the result will be to the theoretical value. When the data volume reaches a certain level and continues to increase, the accuracy of the results will not be greatly improved, but it will increase the time of early data collection and the difficulty of late driving cycle construction. Therefore, it is important to determine the appropriate amount of data. In this paper, eight characteristic parameters, including proportion of acceleration, proportion of deceleration, proportion of uniform speed, proportion of idling, average speed, average running speed, average acceleration and average deceleration were selected to define acceleration proportional stability, deceleration proportional stability, uniform speed proportional stability, idling stability, average speed stability, average running speed stability, average acceleration stability, and average deceleration stability (K_1, $K_2 \cdots K_8$). The average value is defined as comprehensive stability (K_a), which is used as the evaluation index to determine the saturation of sampled data. The calculation method is shown in (2) and (3).

$$K_m(n) = \frac{\overline{N}_m(n) - \overline{N}_m(n-1)}{\overline{N}_m(n)} \tag{2}$$

$$Ka = \frac{1}{8}\sum_{m=1}^{8} K_m(n) \tag{3}$$

where $K_m(n)$ is the stability of the m-th characteristic parameter of n sets of sampled data, and $\overline{N}_m(n)$ is the m-th cumulative characteristic parameter of sampled data from the first set to the n-th set.

According to (2) and (3), the comprehensive stability Ka was calculated, and it was found that the Ka value gradually converges to 0 as the number of sets of sampled data increases, as shown in Figure 2. It is defined so that when the comprehensive stability for five consecutive times satisfies $|K_a| < 0.002$, the sampled data volume reaches saturation. It can be seen from Figure 2 that the data saturation determination condition is satisfied from the 46th set of sampled data. Therefore, the first 50 sets of sampled data were selected as the entire dataset to construct the bus driving cycle, and 50 sets of preprocessed data were partitioned into 2565 micro-trips [9,18].

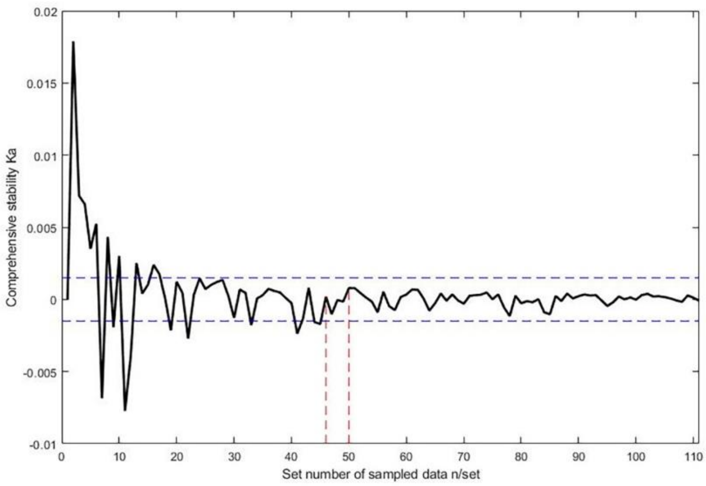


Figure 2. Comprehensive stability K_a calculation results.

2.2. Calculating Characteristic Parameters

In this paper, 19 characteristic parameters that have an important influence on the driving cycle construction are selected [18–20] as shown in Table 1. It is stipulated that the bus is in an acceleration state when its acceleration is greater than or equal to 0.15 m/s^2, and the bus is in a deceleration state when its acceleration is less than or equal to -0.15 m/s^2.

Table 1. All characteristic parameters.

Number	Characteristic Parameters	Abbreviation	Unit
1	Running time	T	s
2	Acceleration time	Ta	s
3	Deceleration time	Td	s
4	Uniform speed time	Tc	s
5	Idling time	Ti	s
6	Running distance	L	m
7	Maximum speed	Vmax	km/h
8	Average speed	Vm	km/h
9	Average running speed	Vmr	km/h
10	Standard deviation of speed	Vsd	km/h
11	Maximum acceleration	Amax	m/s^2
12	Average acceleration	Am	m/s^2
13	Maximum deceleration	Dmax	m/s^2
14	Average deceleration	Dm	m/s^2
15	Standard deviation of acceleration	Asd	m/s^2
16	Proportion of acceleration	Pa	%
17	Proportion of deceleration	Pd	%
18	Proportion of uniform speed	Pu	%
19	Proportion of idling	Pi	%

2.3. Principal Component Analysis

Firstly, the values of the characteristic parameters of each kinematic segment were calculated. The first 15 characteristic parameters were selected for principal component analysis. The dispersion of the characteristic parameters' value will increase due to the different dimensions of the characteristic parameters of micro-trip. Therefore, before principal component analysis, z-score standardization was carried out for the values of the characteristic parameters.

Principal component analysis (PCA) is a dimension-reduction algorithm, which can reduce the number of characteristic parameters. It was conducted for the standardized parameters by SPSS software. The principal components variance, variance contribution rate, and cumulative variance contribution rate are shown in Table 2.

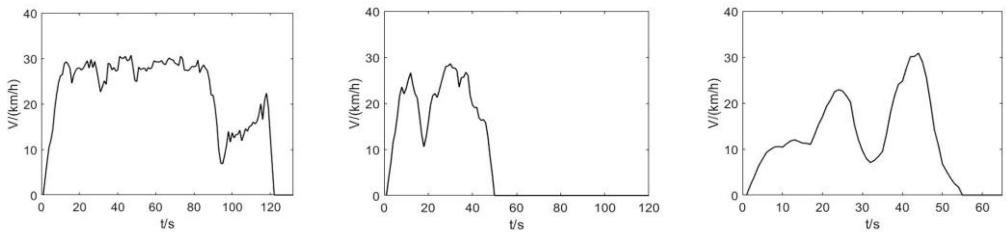
According to Table 2, the cumulative variance contribution rate for the first four principal components is 87.55%. This means only four components need to be analyzed instead of the original 15 characteristic parameters.

Table 2. Principal component analysis results.

Principal Component	Principal Component Variance	Variance Contribution Rate / %	Cumulative Variance Contribution Rate / %
M_1	7.171	47.807	47.807
M_2	3.537	23.582	71.389
M_3	1.416	9.437	80.826
M_4	1.009	6.728	87.554
…	…	…	…

2.4. Clustering Analysis

The above four principal components were analyzed by K-means clustering in SPSS software [21], and they can be grouped into three types of micro-trip with different characteristics [10,19]. The number of micro-trips in cluster 1 is 613, the number of micro-trips in cluster 2 is 222, and the number of micro-trips in cluster 3 is 1730. The average characteristic parameter values of each cluster are shown in Table 3. The most representative micro-trips of the first, second, and third clusters are shown in Figure 3.



(a) (b) (c)

Figure 3. (**a**) Representative kinematic segment of Cluster 1; (**b**) Representative kinematic segment of Cluster 2; (**c**) Representative kinematic segment of Cluster 3.

Table 3. Characteristics of each cluster.

Characteristic Parameters	Cluster 1	Cluster 2	Cluster 3
Vm	20.352	8.309	14.057
Vmr	22.775	19.612	18.522
Am	0.537	0.557	0.574
Dm	−0.570	−0.567	−0.627
Pa	33.4	15.9	31.5
Pd	30.4	14.6	26.5
Pu	25.6	11.8	17.9
Pi	10.6	57.6	24.1
Vsd	11.456	11.943	12.175
Asd	0.542	0.379	0.557

It can be seen from Figure 2 and Table 3 that the average speed and average running speed of the Cluster 1 kinematic segments are the largest, respectively 20.352 km/h and 22.775 km/h, with a minimum idling proportion of 10.6% and a maximum cruising proportion of 25.6%, representing a driving cycle with smooth traffic. The average speed of the Cluster 2 kinematic segments is 8.309 km/h, with a maximum idling proportion of 57.6% and a minimum cruising ratio of 11.8%, representing a driving cycle with heavy traffic. The average speed and average running speed of the Cluster 3 kinematic segments are lower than for the Cluster 1 kinematic segments, which are 14.057 km/h and 18.522 km/h respectively. The Cluster 3 kinematic segments represent the driving cycle with normal traffic.

2.5. Forming Driving Cycle

According to the principle of minimum distance from the cluster center, the representative micro-trips were selected to form a driving cycle. The results of clustering analysis in SPSS software can show the clusters of micro-trips and the distance from the cluster center. The distances from micro-trips to the cluster center were ranked from least to most by the adopting bubble method, as shown in Table 4.

According to the clustering results, the number ratio for the three clusters of micro-trips is 613:222:1730. The total time lengths of the three clusters of micro-trips were 81,514 s, 29,477 s, and 99,183 s respectively, and the ratio was 2.77:1:3.36. For the driving cycle, cycle duration should be not too short or too long. If it is too short, it cannot reflect the real conditions. If it is too long, it will cost too much time and money in future testing. Thus, the time length of the driving cycle was set to about 1300 s, and the time lengths for the three clusters of micro-trips were about 504.2 s, 182.3 s, and 613.5 s respectively. The corresponding number ratio of the three clusters of micro-trips is 4:1:11. Therefore, the first four micro-trips were selected from the first cluster, the first kinematic segment was selected from the second cluster, and the first eleven micro-trips were selected from the third cluster. Segments 1573, 2005, 311, 353 (Cluster 1), 2021 (Cluster 2), 568, 1652, 2024, 1744, 1083, 183, 1271, 1802, 1227, 329 and 506 (Cluster 3) were selected to form the representative driving cycle for the Xi'an No.2 bus, as shown in Figure 4. The running time is 1281 s, the running distance is 5.13 km, the maximum speed is 38.15 km/h, the maximum acceleration is 1.76 m/s^2, and the maximum deceleration is -2.17 m/s^2.

Table 4. Distances and the segment number from micro-trips to the cluster center.

Cluster 1	Distance	0.217	0.229	0.304	0.343	0.377	...
	Segment number	1573	2005	311	353	211	...
Cluster 2	Distance	0.729	0.760	0.800	0.892	1.064	...
	Segment number	2021	1382	987	689	2327	...
Cluster 3	Distance	0.293	0.293	0.294	0.340	0.343	...
	Segment number	568	1652	2024	1744	1083	...

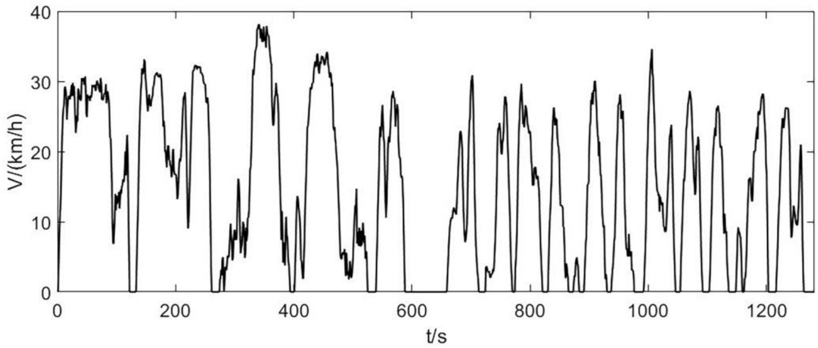


Figure 4. Clustering representative driving cycle.

3. Markov Chain Method

The flow diagram for constructing the Xi'an No.2 bus driving cycle by the Markov chain method is shown in Figure 5.

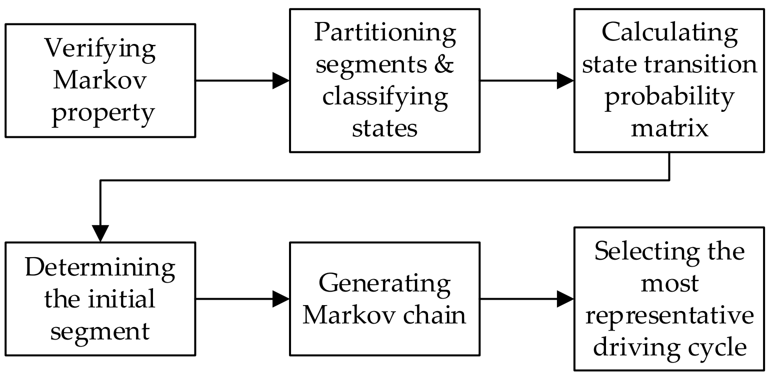


Figure 5. Flow diagram of driving cycle construction based on the Markov chain method.

3.1. *Verifying the Markov Property*

The state of a random process at time T is only related to the state at time T-1 and has nothing to do with the previous state. In a word, it has no after-effect, and the random process is called the Markov process. Correlation analysis was conducted on the entire dataset to calculate the correlation coefficients at different time intervals of 1 s, 5 s, 10 s, and

30 s respectively according to (4), where X is the velocity vector and Y is the velocity vector after a certain time interval adjacent to X, and the results are shown in Table 5.

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}} \tag{4}$$

Table 5. Correlation coefficients at different time intervals.

Time Interval	Correlation Coefficient
1 s	0.9865
5 s	0.8299
10 s	0.5863
30 s	0.1808

The correlation of vehicle speed data decreases with an increase in time interval, as shown in Table 5. When the time interval is 1 s, the correlation coefficient of vehicle speed reaches 0.9865; that is, the running state of the vehicle in the next period is only related to the current running state, but has nothing to do with the previous running state, which conforms to the definition of a Markov chain. Thus, it is proved that the data collected at the sampling frequency of 1 Hz has Markov characteristics, and the Markov chain method can be used to construct the driving cycle of the No.2 bus.

3.2. Partitioning Segments and Classifying States

The entire dataset was partitioned into four types of unit segments: idle segments, acceleration segments, deceleration segments, and uniform speed segments. The flow diagram of the segment partitions is shown in Figure 6.

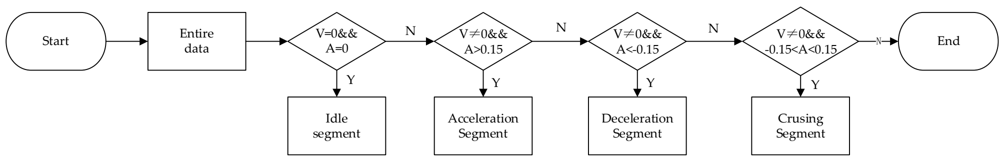


Figure 6. Flow diagram of segment partitions.

After partitioning segments, we calculated the average speed of unit segments and classified unit segments into different states according to the average speed. According to the statistics of the entire dataset, the speed of the bus is generally not more than 40 km/h. Therefore, unit segments were partitioned into 10 states according to their average speed (State 1: 0~4 km/h, State 2: 4~8 km/h... State 10: ≥36 km/h), and each unit segment was numbered and its number and state information was recorded.

3.3. Calculating the State Transition Probability Matrix

The state change process of the bus from time T to time T+1 is called the state transition process. The one-step transition probability between all states constitutes the state transition probability matrix P. The number of unit segments in different states is counted to calculate the transition probability between different states according to (5), where N_{ij} is the number of segments transferred from state i to state j, and P_{ij} is the transition probability from state i to state j [9].

$$p_{ij} = \frac{N_{ij}}{\sum_j N_{ij}} \tag{5}$$

The calculation result for the state transition probability matrix P is shown in (6), and the corresponding state transition probability distribution diagram is shown in Figure 7. It can be known from common sense that the probability that the bus maintains its current state while running is the largest, which is reflected in the transition probability matrix in that the value of the main diagonal element is the maximum among the corresponding row elements. It can be seen from Figure 7 that the state transition probability matrix conforms to this feature.

$$P = \begin{pmatrix} 0.5926 & 0.2165 & 0.0862 & 0.0603 & 0.0387 & 0.0056 & 0.0001 & 0 & 0 & 0 \\ 0.2173 & 0.4862 & 0.2040 & 0.0675 & 0.0187 & 0.0062 & 0.0002 & 0 & 0 & 0 \\ 0.0929 & 0.1762 & 0.4333 & 0.1645 & 0.0839 & 0.0423 & 0.0067 & 0.0002 & 0 & 0 \\ 0.0919 & 0.0510 & 0.1536 & 0.4059 & 0.1583 & 0.0724 & 0.0486 & 0.0187 & 0.0013 & 0 \\ 0.0653 & 0.0151 & 0.0660 & 0.1525 & 0.3966 & 0.1788 & 0.0663 & 0.0392 & 0.0183 & 0.0019 \\ 0.0126 & 0.0052 & 0.0335 & 0.0581 & 0.1491 & 0.4629 & 0.2096 & 0.0514 & 0.0134 & 0.0048 \\ 0.0007 & 0.0008 & 0.0060 & 0.0341 & 0.0389 & 0.1262 & 0.5837 & 0.1884 & 0.0188 & 0.0024 \\ 0.0001 & 0.0007 & 0.0001 & 0.0107 & 0.0218 & 0.0203 & 0.1294 & 0.7150 & 0.0984 & 0.0040 \\ 0 & 0 & 0 & 0.0011 & 0.0135 & 0.0138 & 0.0194 & 0.1642 & 0.6996 & 0.0884 \\ 0 & 0 & 0 & 0.0003 & 0.0058 & 0.0105 & 0.0071 & 0.0149 & 0.2587 & 0.7025 \end{pmatrix} \tag{6}$$

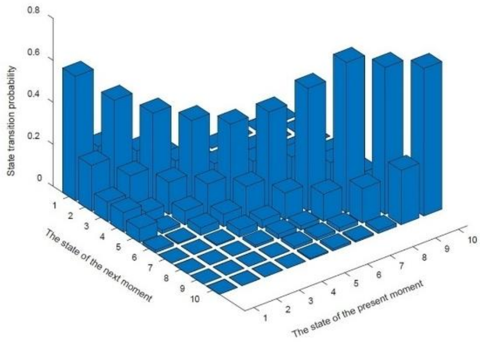


Figure 7. State transition probability distribution.

3.4. Determining the Initial Segment

In the construction of a bus driving cycle based on the Markov chain method, the state of the next segment depends on the state of the current segment and the corresponding state transition probability, so it is necessary to determine the initial state of the bus for the construction of subsequent driving cycle [22]. A too short or too long initial segment is not suitable. If it is too short, it cannot reflect the buses starting condition, and, if it is too long, it will affect the quality of the driving cycle. According to analysis of the data on the bus starting, 50s fixed durations starting from the stationary state were selected from the entire dataset as candidate initial segments in this paper. The V-A matrices of the candidate initial segments and the entire dataset were calculated. The V-A matrix of the entire dataset is shown in Table 6.

Then two matrices were converted into two one-dimensional arrays in the same way, and the similarity Ω of the two arrays was calculated according to (7).

$$\Omega = \frac{\sum_{i=1}^{n} AiBi}{\sqrt{\sum_{i=1}^{n} A_i{}^2}\sqrt{\sum_{i=1}^{n} B_i{}^2}} \tag{7}$$

where Ai is the i-th value in the one-dimensional array corresponding to the candidate initial segment, Bi is the i-th value in the one-dimensional array corresponding to the entire dataset, and n is the length of the one-dimensional array. The candidate initial segment with the maximum similarity to the entire dataset V-A matrix was selected as the initial segment, as shown in Figure 8.

Table 6. The V-A matrix of the overall data Unit %.

V(km/h) a(m/s^2)	[0,5)	[5,10)	[10,15)	[15,20)	[20,25)	[25,30)	[30,35)	[35,40)	[40,45)	[45,50]
(2.5,3]	0	0	0.002	0	0	0	0	0	0	0
(2,2.5]	0.003	0.004	0.003	0.003	0	0	0	0	0	0
(1.5,2]	0.007	0.016	0.012	0.008	0.005	0.001	0	0	0	0
(1,1.5]	0.041	0.078	0.064	0.038	0.019	0.011	0.001	0	0	0
(0.5,1]	0.220	0.242	0.224	0.170	0.081	0.029	0.009	0	0	0
(0,0.5]	0.575	0.570	0.555	0.540	0.459	0.225	0.076	0.012	0	0
(−0.5,0]	1.655	1.141	1.195	1.301	1.450	1.391	0.670	0.123	0.002	0
(−1,−0.5]	25.107	2.257	2.152	2.144	2.796	5.692	5.002	1.561	0.012	0
(−1.5,−1]	2.120	2.585	2.311	2.344	2.919	5.664	5.682	2.025	0.023	0
(−2,−1.5]	1.292	1.803	1.705	1.610	1.758	1.731	1.051	0.362	0.011	0
(−2.5,−2]	0.168	0.331	0.527	0.569	0.459	0.285	0.159	0.057	0.011	0
(−3,−2.5]	0	0.024	0.060	0.094	0.090	0.055	0.037	0.010	0.003	0
(−3.5,−2]	0	0	0.011	0.017	0.023	0.022	0.007	0.003	0	0
(−4,−3.5]	0	0	0.004	0.005	0.003	0.002	0.003	0	0	0

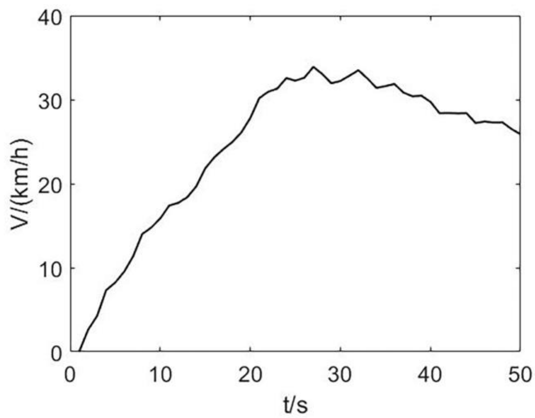


Figure 8. The initial segment.

3.5. Generating a Markov Chain

After determining the initial segment, a random array conforming to the state transition probability matrix was generated with the Monte Carlo simulation method to determine the next segment. Assuming that the current segment state is State 6, the transition probabilities corresponding to the next segment states are 0.0126, 0.0052, 0.0335, 0.0581, 0.1491, 0.4629, 0.2096, 0.0514, 0.0134, and 0.0048. The (0, 1) interval is divided into 10 intervals according to the probability value. The program generates a random number r that satisfies the uniform distribution in (0, 1) interval. The next segment is state k when the random number r meets (8).

$$\sum_{j=1}^{\mathrm{k}} p_{ij} \leq r \leq \sum_{j=1}^{k+1} p_{ij} \tag{8}$$

where i is the current segment state, j is the next segment state, and P_{ij} is the transition probability from state i to state j. After determining the state of the next segment, the segment with the minimum difference between its initial speed and the current segment's final speed is selected as the optimal segment in the segment set corresponding to the state, then it is spliced with the current segment, and each segment is selected only once. The above process is repeated until the Markov chain satisfies the time requirement (about 1300 s) and the final speed is zero. The process flow diagram is shown in Figure 9.

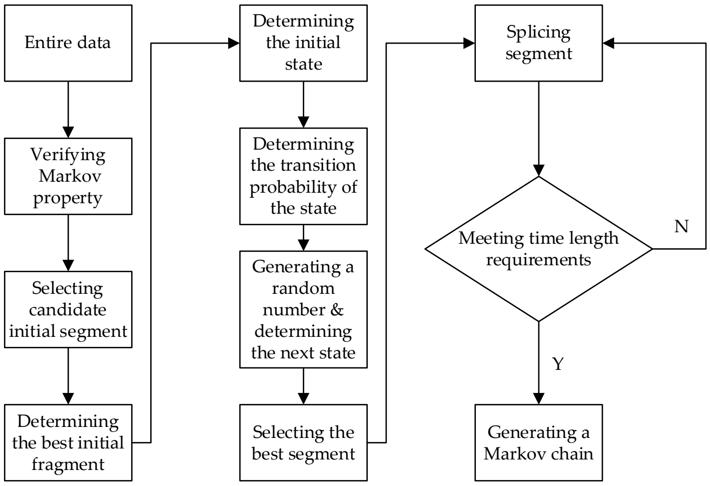


Figure 9. Flow diagram of generating a Markov chain.

3.6. Selecting the Most Representative Driving Cycle

One Markov chain represents one driving cycle. Every driving cycle constructed by the Markov chain method is different, so it is necessary to generate several Markov chains and select a candidate driving cycle with the minimum deviation from the entire dataset as the representative driving cycle. We constructed 50 candidate driving cycles. Ten characteristic parameters, including average speed, average running speed, the standard deviation of speed, average acceleration, average deceleration, the standard deviation of acceleration, proportion of acceleration, proportion of deceleration, proportion of uniform speed, and proportion of idling were selected for the deviation evaluation index. The average deviation was calculated to describe the deviation between the candidate driving cycles and the overall data according to (9) and (10).

$$\delta_i = \frac{|A_i - A|}{A} \times 100\% \tag{9}$$

$$\overline{\delta} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \tag{10}$$

where A_i is the i-th class characteristic parameter of the candidate driving cycles, A is the i-th class characteristic parameter of the overall data, δ_i is the deviation of the i-th characteristic parameter, and $\overline{\delta}$ is the average deviation. The average deviations between 50 candidate driving cycles and the overall data were calculated, and the candidate driving cycle with the minimum average deviation was selected as the representative driving cycle [9], as shown in Figure 10. The running time is 1277 s, the running distance is 5.73 km, the maximum speed is 39.42 km/h, the maximum acceleration is 2.43 m/s^2, and the maximum deceleration is -2.44 m/s^2.

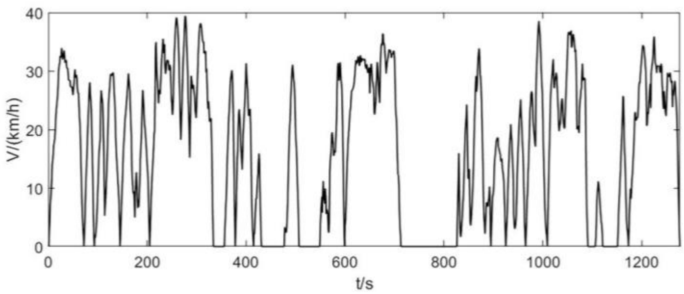


Figure 10. Markov chain representative driving cycle.

4. Combined Micro-Trip and Markov Chain Method

The flow diagram for constructing the Xi'an No.2 bus driving cycle by the combined micro-trip and Markov chain method is shown in Figure 11.

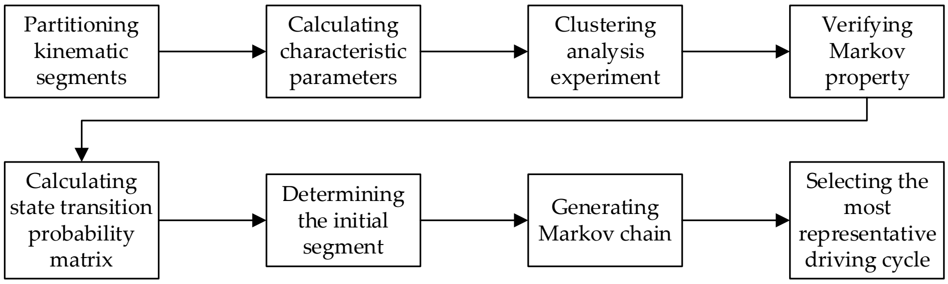


Figure 11. Flow diagram for driving cycle construction based on the combined micro-trip and Markov chain method.

4.1. Clustering Analysis Experiment and Calculating State Transition Probability Matrix

The partitioning of micro-trips, the definitions of characteristic parameters, and the standardization of characteristic parameters are the same as in the clustering analysis method above.

The combined micro-trip and Markov chain method takes different clusters from clustering analysis as different states and then generates a Markov chain based on the Markov chain method. If the number of micro-trips belonging to a certain cluster is too small, the probability of transition to this cluster will be extremely small, which in turn leads to a small probability of selecting a micro-trip belonging to this cluster during the driving cycle construction. This is equivalent to reducing the amount of data. Therefore, the characteristic parameters should be reasonably selected for clustering analysis to avoid the above phenomenon.

To determine appropriate characteristic parameters, 30 different combinations of characteristic parameters were selected. SPSS software was used to conduct a k-means clustering analysis experiment on 2565 micro-trips, and the number of clusters was 3. After analyzing the clustering results, the four characteristic parameters of maximum speed, the standard deviation of speed, maximum acceleration, and maximum deceleration were selected as the best combination of characteristic parameters for clustering analysis, which is the most in line with the actual driving cycle of the bus.

The Markov property verification is the same as in the Markov chain method above, and the calculation results for the state transition probability matrix are shown in (11).

$$P = \begin{Bmatrix} 0.735786 & 0.169743 & 0.094471 \\ 0.088969 & 0.755284 & 0.155747 \\ 0.079954 & 0.138903 & 0.781143 \end{Bmatrix} \quad (11)$$

4.2. Generating Markov Chains and Constructing Driving Cycles

The process of generating a Markov chain is like the Markov chain method above. We determine the state of the initial segment based on the cluster it belongs to. The state of the next unit segment is then determined according to the state transition matrix, each unit segment in the corresponding state segment set is selected for long-segment splicing, and the similarity between the V-A matrix of the current long segment and the V-A matrix for the overall data is calculated respectively. We select the unit segment with the maximum similarity in the corresponding state segment set, and each unit segment can only be selected once.

This step is repeated until the time length of the current long segment is appropriate. After each step, the similarity between the V-A matrix of the current long segment and the V-A matrix of the overall data is calculated according to (4). In this way, a candidate driving cycle is constructed. The above steps are repeated to construct a total of 50 candidate driving cycles. The candidate driving cycle with the minimum average deviation is selected as the final representative driving cycle according to (6) and (7) [23]. The representative driving cycle is shown in Figure 12. The running time is 1292 s, the running distance is 5.64 km, the maximum speed is 38.48 km/h, the maximum acceleration is 2.44 m/s^2, and the maximum deceleration is -2.78 m/s^2.

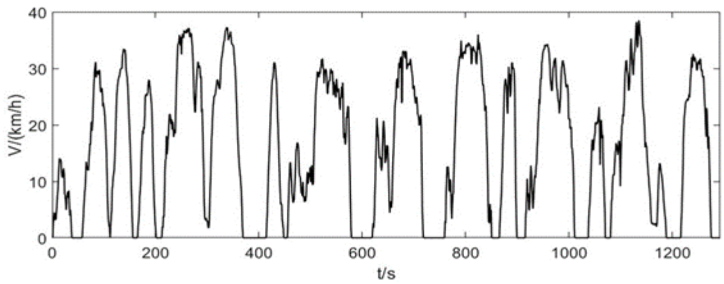


Figure 12. Combined micro-trip and Markov chain representative driving cycle.

5. Comparison of Driving Cycles

5.1. Comparison of Characteristic Parameters

To evaluate the driving cycles of the Xi'an No.2 bus constructed based on the above three methods, average speed, average running speed, the standard deviation of speed, average acceleration, average deceleration, standard deviation of acceleration, the proportion of acceleration, proportion of deceleration, proportion of uniform speed, and proportion of idling were selected as deviation evaluation indicators. The deviations between the driving cycles and the overall data are described by the average deviations, to reflect the accuracy of the driving cycles. After calculation, Deviation of characteristic parameters between the driving cycles constructed based on the above three methods and the overall data can be obtained respectively, as shown in Table 7.

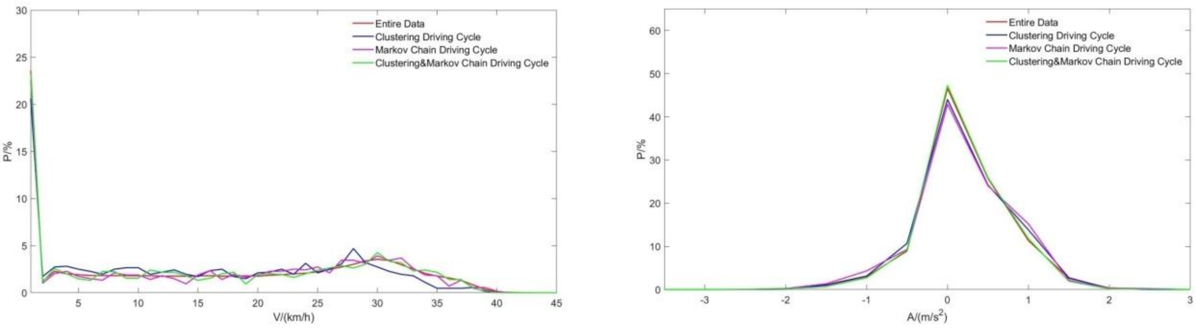
Table 7. Comparison results of characteristic parameters.

Characteristic Parameters	**Vm**	**Vmr**	**Vsd**	**Am**	**Dm**	**Add**
Overall data	15.687	20.527	12.577	0.557	−0.597	0.530
Micro-trip driving cycle	14.405	18.144	11.426	0.572	−0.604	0.533
Deviation	8.2%	11.6%	9.1%	2.8%	1.2%	0.6%
Markov chain driving cycle	16.163	20.870	12.431	0.563	−0.632	0.573
Deviation	3.0%	1.7%	1.2%	1.0%	5.9%	8.1%
Micro-trip anf Markov chain driving cycle	15.707	20.456	12.515	0.535	−0.575	0.499
Deviation	0.1%	0.3%	0.5%	4.0%	3.6%	5.9%
Characteristic parameters	**Pa**	**Pd**	**Pu**	**Pi**	**Average deviation**	
Overall data	30.1	26.3	20.0	23.6	0	
Micro-trip driving cycle	30.9	27.3	21.2	20.6		
Deviation	2.8%	3.7%	5.7%	12.6%	5.8%	
Markov chain driving cycle	34.9	29.6	12.9	22.6		
Deviation	15.9%	12.4%	35.5%	4.4%	8.9%	
Micro-trip and Markov chain driving cycle	29.6	26.2	21.0	23.2		
Deviation	1.4%	0.7%	4.8%	1.5%	2.3%	

As can be seen from Table 7, the deviations of the three driving cycles are all within a reasonable range. The average deviation of the driving cycle based on the combined micro-trip and Markov chain method is the least, 2.3%. The deviations of average speed, average running speed, and standard deviation of speed are the least, 0.1%, 0.3%, and 0.5%, respectively, and the deviations of the proportion of acceleration, proportion of deceleration, proportion of uniform speed, and proportion of idling are small, but the deviations of average acceleration, average deceleration and standard deviation of acceleration are large, 4.0%, 3.6%, and 5.9%, respectively. The average deviation of the driving cycle based on the micro-trip method is moderate, 5.8%. The deviations of average acceleration, average deceleration, and standard deviation of acceleration are the least, 2.8%, 1.2%, and 0.6%, respectively, and the deviations of the proportion of acceleration, proportion of deceleration, proportion of uniform speed, and proportion of idling are small, but the deviations of average speed, average running speed, and standard deviation of speed are larger, 8.2%, 11.6%, and 9.1%, respectively, to make the average deviation larger. The average deviation of the driving cycle based on the Markov chain method is the largest, 8.9%. The deviations of average speed, average running speed, and standard deviation of speed are lower, 3.0%, 1.7%, and 1.2%, respectively; the deviations of average acceleration, average deceleration, and standard deviation of acceleration are larger; and the deviations of the proportion of acceleration, proportion of deceleration, proportion of uniform speed, and proportion of idling are the largest, 16.2%, 12.4%, 35.5%, and 4.4%, respectively, which leads to the largest average deviation.

5.2. Comparison of Speed Distribution and Acceleration Distribution

The probability distribution diagrams of speed and acceleration for the overall data and three driving cycles are shown in Figure 13.



(**a**) (**b**)

Figure 13. Speed distribution and acceleration distribution. (**a**). Speed distribution. (**b**). acceleration distribution.

It can be seen from Figure 12 that both the speed distribution probability curve of the Markov chain driving cycle and the speed distribution probability curve of the combined micro-trip and Markov chain driving cycle fit well with the speed distribution probability curve of the entire dataset. While the speed distribution probability curve of the micro-trip driving cycle fluctuates greatly, the deviation between it and the curve for the overall data is larger.

It can be seen from Figure 13 that the acceleration distribution probability curve of the combined micro-trip and Markov chain driving cycle fits best with the acceleration distribution probability curve of the overall data, so that the deviations of the proportion of acceleration, proportion of deceleration, proportion of uniform speed, proportion of idling between the driving cycle and the overall data are all the least.

5.3. Comparison of Vehicle Specific Power Distribution

Vehicle Specific Power (VSP) is defined as the instantaneous power demand of the vehicle divided by its mass, and the commonly used unit is kW/t [23]. For the same type of vehicle, the greater the VSP, the better the power of the vehicle. VSP is an important parameter that can reflect the driving cycle and the emissions of the bus. Therefore, the accuracy of different driving cycles can be compared by comparing the VSP distribution of different driving cycles and the VSP distribution of the entire dataset. For the No.2 bus, the calculation formula is shown in (12), where v is vehicle speed (m/s), a is vehicle acceleration (m/s^2), g is the gravity acceleration (9.8 m/s^2), s is road gradient (taken 0), C_R is the rolling resistance coefficient (taken 0.00938), ρ_a is the air density (1.207 kg/m^3 at 20 °C), C_D is the wind drag coefficient (taken 0.6), A is the frontal cross-section (8.0 m^2), m is the gross mass (16,975 kg), and N is a total number of VSP bins [17].

$$\begin{gathered}\text{VSP} = v(a + gs + gC_R) + \tfrac{1}{2}\rho_a \tfrac{C_D A}{m} v^3 = v(a + 0.09192) + 0.000171v^3 \\ \forall : N - 0.5 \leq \text{VSP} \leq N + 0.5 \\ \text{VSP Bin} = N, N \in [-20, 20], N \in Z\end{gathered} \tag{12}$$

VSPs of the overall data and three driving cycles were calculated respectively, then VSPs were divided into different bins according to a certain interval. In this paper, 1 kW/t is adopted as the incremental interval, the value range of VSP Bin is [−20, 20], and the VSP distribution is obtained.

Root-mean-square error (RMSE) was employed to assess the similarity between the three driving cycles and the overall data [24]. The RMSE is calculated according to (13). The lower the value of RMSE is, the more similar the driving cycles are to the overall data. [25].

$$\text{RMSE} = \sqrt{\frac{\sum_{i=-20}^{20} (\text{Bin}_{e,i} - \text{Bin}_{k,i})^2}{L - 1}} \tag{13}$$

where i is the ID number of the VSP bin, $\text{Bin}_{e,i}$ is the time proportion of the VSP bin i of the entire dataset, and $\text{Bin}_{k,i}$ is the time proportion of the VSP bin i of the *k-th* driving cycle (k = 1,2,3). L is the total number of VSP bins (L = 41). The calculation results are shown in Table 8.

Table 8. The RMSE between three driving cycles and the overall data.

Driving Cycle	RMSE
Micro-trip driving cycle	0.0034
Markov chain driving cycle	0.0120
Micro-trip and Markov chain driving cycle	0.0031

It can be seen from Table 8 that the RMSEs of the three driving cycles are all in a reasonable range. The RMSE of the VSP distribution between the combined micro-trip and Markov chain driving cycle and the overall data is the lowest (0.0031), so the method of the combined micro-trip and Markov chain driving cycle has the maximum similarity to the overall data. The RMSE of the VSP distribution between the micro-trip driving cycle and the overall data is also very low (0.0034), so the similarity between the micro-trip driving cycle and the overall data is also high. The RMSE of the VSP distribution between the Markov chain driving cycle and the overall data is the highest (0.0120), so the Markov chain driving cycle has the minimum similarity to the overall data.

6. Conclusions

Based on the driving data of the Xi'an No.2 bus, this paper uses the micro-trip-based method, Markov chain method, and combined micro-trip and Markov chain method to construct the representative driving cycle of a Xi'an urban bus. Different driving cycles constructed based on different methods have different deviations from the entire data, which reflects the advantages and disadvantages of the different methods.

(1) The micro-trip-based method can effectively classify micro-trips, and the selected micro-trips also have obvious driving characteristics. The constructed representative driving cycle has a small deviation from the overall data. However, the calculation process for the micro-trip-based method is relatively complex. Standardization and principal component analysis are required because the number of defined characteris-

tic parameters is large. Meanwhile, micro-trips in each cluster need to be sorted and spliced in accordance with the time ratio, which leads to a large amount of calculation.

(2) The core idea of the Markov chain method is to regard the process of vehicle speed changing with time as a Markov process and use this property to construct a driving cycle. This method does not need to partition micro-trips to ensure the continuity of driving data. However, its segment partition method is relatively rough, and the Markov chain has randomness, which leads to large uncertainty in constructing driving cycles; therefore, the constructed representative driving cycle has a larger deviation from the entire data.

(3) The combined micro-trip and Markov chain method combines the advantages of the micro-trip-based method and the Markov chain method. The micro-trip-based method can determine the best combination of characteristic parameters, and the Markov chain method can reflect the random characteristics of driving, closer to the actual driving characteristics of the bus. The constructed representative driving cycle has the minimum deviation from the overall data, and this method is suitable for a situation where the accuracy of the driving cycle is very high. However, this method needs standardization, calculation of the clustering effect under different combinations of characteristic parameters, and calculation of the similarity between the current long segment and the entire data when splicing long segments. Therefore, the computational complexity is high, and the construction process is particularly complicated.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L.; software, D.Z.; validation, D.Z. and H.D.; formal analysis, D.Z. and H.D.; investigation, D.Z. and H.D.; resources, D.Z. and H.D.; data curation, D.Z. and H.D.; writing—original draft preparation, D.Z. and H.D.; writing—review and editing, R.I.; visualization, R.I.; supervision, Y.L.; project administration, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Xi'an Beilin District Science and Technology Project (Grant NO. GX2252).

Data Availability Statement: Due to privacy, the data is unavailable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Brady, J.; O'mahony, M. Development of a driving cycle to evaluate the energy economy of electric vehicles in urban areas. *Appl. Energy* **2016**, *177*, 165–178. [CrossRef]
2. Galgamuwa, U.; Perera, L.; Bandara, S. Developing a General Methodology for Driving Cycle Construction: Comparison of Various Established Driving Cycles in the World to Propose a General Approach. *J. Transp. Technol.* **2015**, *5*, 191–203. [CrossRef]
3. Huertas, J.; Quirama, L.; Giraldo, M.; Díaz, J. Comparison of Three Methods for Constructing Real Driving Cycles. *Energies* **2019**, *12*, 665. [CrossRef]
4. Liu, X.; Ma, J.; Zhao, X.; Du, J.; Xiong, Y. Study on Driving Cycle Synthesis Method for City Buses considering Random Passenger Load. *J. Adv. Transp.* **2020**, 3871703. [CrossRef]
5. Mongkonlerdmanee, S.; Koetniyom, S. Development of a Realistic Driving Cycle Using Time Series Clustering Technique for Buses: Thailand Case Study. *Eng. J.* **2019**, *23*, 49–65. [CrossRef]
6. Quirama, L.F.; Giraldo, M.; Huertas, J.I.; Jaller, M. Driving cycles that reproduce driving patterns, energy consumptions and tailpipe emissions. *Transp. Res. Part D Transp. Environ.* **2020**, *82*, 102294. [CrossRef]
7. Tong, H.Y. Development of a driving cycle for a supercapacitor electric bus route in Hong Kong. *Sustain. Cities Soc.* **2019**, *48*, 101588. [CrossRef]
8. Tong, H.Y.; Ng, K.W. A bottom-up clustering approach to identify bus driving patterns and to develop bus driving cycles for Hong Kong. *Environ. Sci. Pollut. Res.* **2021**, *28*, 14343–14357. [CrossRef]
9. Li, Y.H.; Ren, T.Y.; Shao, P.D.; Song, W.P.; Li, Z.Y.; Gou, Q.Z. Development of driving cycle of bus in Xi'an city based on Markov chain. *China Sci.* **2019**, *14*, 121–128.
10. Li, Z.Y. Driving Cycle Construction of Xian and Parameters Matching of Hybrid Electric Vehicle. Ph.D. Thesis, Chang'an University, Xi'an, China, 2018.

11. Ou, S.Q.; Zhou, Y.F.; Lian, J.; Jia, P.; Tian, B. Development of Hybrid City Bus's Driving Cycle. In Proceedings of the 2011 International Conference on Electric Information and Control Engineering, Wuhan, China, 15–17 April 2011; IEEE: Piscataway, NJ, USA; pp. 2112–2116.
12. Peng, Y.H.; Zhuang, Y.; Yang, H.B. Development of a representative driving cycle for urban buses based on the K-means cluster method. *Clust. Comput. J. Netw. Softw. Tools Appl.* **2019**, *22*, 6871–6880.
13. Qin, D.T.; Zhan, S.; Qi, Z.G.; Chen, S.J. Driving cycle construction using K-means clustering method. *J. Jilin Univ. Eng. Technol. Ed.* **2016**, *46*, 383–389.
14. Gong, H.M.; Zou, Y.; Yang, Q.K.; Fan, J.; Sun, F.; Goehlich, D. Generation of a driving cycle for battery electric vehicles: A case study of Beijing. *Energy* **2018**, *150*, 901–912. [CrossRef]
15. Mayakuntla, S.K.; Verma, A. A novel methodology for construction of driving cycles for Indian cities. *Transp. Res. Part D Transp. Environ.* **2018**, *65*, 725–735. [CrossRef]
16. Cao, Q.; Li, J.; Qu, D.W. Construction of Typical Driving Cycle for Passenger Cars in the City of Dalian. *J. Shanghai Jiaotong Univ.* **2018**, *52*, 1537–1542.
17. Shi, S.M.; Zou, G.L.; Liu, L.; Kui, H.; Wu, D.; Huang, C.; Liu, M. Study on the Fuzzy Clustering Method of the Microtrips for Passenger Car Driving Cycle in Changchun. In Proceedings of the 5th IEEE Vehicle Power and Propulsion Conference (VPPC 09), Dearborn, MI, USA, 7–10 September 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1338–1343.
18. Pouresmaeili, M.A.; Aghayan, I.; Taghizadeh, S.A. Development of Mashhad driving cycle for passenger car to model vehicle exhaust emissions calibrated using on-board measurements. *Sustain. Cities Soc.* **2018**, *36*, 12–20. [CrossRef]
19. Liu, P. Study on Xi'an XX Road City Bus Driving Cycle and Parameters Matching of Hybrid Electric City Bus Dynamic Systems. Ph.D. Thesis, Chang'an University, Xi'an, China, 2017.
20. Galgamuwa, U.; Perera, L.; Bandara, S. Development of a driving cycle for Colombo, Sri Lanka: An economical approach for developing countries. *J. Adv. Transp.* **2016**, *50*, 1520–1530. [CrossRef]
21. Berzi, L.; Delogu, M.; Pierini, M. Development of driving cycles for electric vehicles in the context of the city of Florence. *Transp. Res. Part D Transp. Environ.* **2016**, *47*, 299–322. [CrossRef]
22. Liu, L.; Huang, C.S.; Lu, B.W.; Shi, S.; Zhang, Y.; Cheng, J. Study on the Design Method of Time-Variant Driving Cycles for EV Based on Markov Process. In Proceedings of the IEEE Vehicle Power and Propulsion Conference (VPPC), Seoul, Republic of Korea, 9–12 October 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1277–1281.
23. Degraeuwe, B.; Weiss, M. Does the New European Driving Cycle (NEDC) really fail to capture the NOX emissions of diesel cars in Europe? *Environ. Pollut.* **2017**, *222*, 234–241. [CrossRef] [PubMed]
24. Lai, J.X.; Yu, L.; Song, G.H.; Guo, P.; Chen, X. Development of City-Specific Driving Cycles for Transit Buses Based on VSP Distributions: Case of Beijing. *J. Transp. Eng.* **2013**, *139*, 749–757. [CrossRef]
25. Lai, J.X.; Yu, L.; Song, G.H.; Guo, P.; Chen, X. Establishing Method of Driving Cycles for Transit Buses Based on VSP Distributions. *J. Transp. Syst. Eng. Inf. Technol.* **2014**, *14*, 209–214+228.