

Development of vehicle driving cycles based on the real traffic dataset

Maksim Potashnikov^{1*}, Victoria Shishkina¹, Alexandr Muravev¹, and Alexandr Kartashov¹

¹Bauman Moscow State Technical University, 105005 Moscow, Russia

Abstract. The paper describes methods for generating (modeling) representative driving cycles of the vehicle used to solve engineering problems in the design of electric vehicles, such as resource calculations, determination of the required capacity of traction batteries and evaluation of power reserve, etc. The general approaches used in the processing of real traffic data and algorithms for modeling driving cycles using deterministic and probabilistic approaches are described. This paper presents driving cycles that ensure convergence with parameters corresponding to the real conditions of vehicle movement. The developed driving cycles can be used in the design of electric vehicle transmission components and allow for the analysis of operational properties.

1 Introduction

Driving cycles are quite popular in the vehicles design, as they are useful for the analysis of operational properties and can be used as a source of primary information for generating loads in the calculations of vehicle components. One of the advantages of this approach is that driving cycles reflect certain operating conditions of vehicles and allow for comparative analysis both for different traffic conditions, for example, urban and suburban traffic cycles, and for individual regions within the country or in different countries.

Among the approaches for modeling driving cycles, there are four main groups [1]:

- micro-trip based cycle construction;
- segment based cycle construction;
- modal cycle construction;
- cycle construction based on machine learning.

To assess the convergence of the obtained driving cycles with real traffic data, they are compared with each other according to the following criteria [2]:

- mean speed (km/h);
- maximum speed (km/h);
- average positive road slope (%);
- average negative road slope (%);
- mean acceleration (m/s^2);

* Corresponding authors: potashnikov@bmstu.ru

- mean deceleration (m/s^2);
- percentage of time spent in acceleration mode (%);
- percentage of time spent in deceleration mode (%);

It is worth noting that there are many ready-made driving cycles used to assess the operational properties of vehicles in various driving conditions, but there are no driving cycles that take into account the specifics of the operating conditions of different regions of the Russian Federation. In this regard, the main purpose of this work is to obtain a driving cycle that takes into account the peculiarities of movement in the specified regions of operation. The data obtained as a result of long-term operation of mainline tractors in various regions of the Russian Federation were taken as the basis for the development and an analysis of various methods used to generate driving cycles was carried out. Based on the results of the analysis of various methods for the development of driving cycles, two options were chosen, the micro-trips method and the Markov chains method, with some modification of them.

The synthesis of driving cycles was carried out to determine the load conditions of electromechanical transmissions within the framework of the project "Creation of high-tech production of mechatronic transmissions of promising KAMAZ trucks and buses with electric energy storage and hydrogen fuel cells", carried out by Bauman Moscow State Technical University jointly with PJSC "KAMAZ", carried out under the Agreement on the provision of Subsidies dated April 7, 2022 No. 02-17560/2021.

2 Micro-trip based cycle construction using k-means clustering

One of the most widespread methods of generating driving cycles of a vehicle is an approach using micro-trips - short sections of the real vehicle movement measurements. This method is based on obtaining a set of micro-trips with certain parameters, such as average speed, average acceleration, average deceleration, etc., and the subsequent formation of a representative cycle of movement from them. There are various ways to select micro-trips from the data. Among the most common there are six methods [3]:

- division into sections by stops;
- division into fixed time intervals;
- division into sections with a fixed length;
- division into sections between two intersections;
- division into sections with a certain speed interval;
- division into sections by acceleration values.

In this paper, the choice was made on dividing the traffic data by vehicle stops, since it provides good convergence, and also simplifies the procedure for combining micro-trips for both the speed signal and the road slope values.

The method is based on the formation of a driving cycle from a set of micro-trips from the dataset obtained earlier at the division stage. There are several ways to choose micro-trips for a cycle, such as random, quasi-random [4], modal [5], etc. In this paper, the method using clustering for preliminary grouping of micro-trips was chosen [6], since it shows better convergence with the real traffic measurements, and also has greater reproducibility.

Data clustering belongs to the unsupervised learning group of machine learning tasks. One of the most common algorithms is k-means. This method is based on the search for specific clusters of data, as well as the search for their centers, the so-called cluster centers. To solve the clustering problem, a set of criteria is needed on the basis of which the grouping will be carried out. The following criteria were used:

- mean speed (km/h);
- mean acceleration (m/s^2);

- mean deceleration (m/s^2);
- trip time (sec);
- trip distance (m);
- number of road slopes over 5%;

The initial choice of a set of parameters was based on the recommendations [2] and included a larger number of criteria, but during the analysis we selected the parameters that made the most significant contribution to the clustering algorithm and the rest were excluded from consideration.

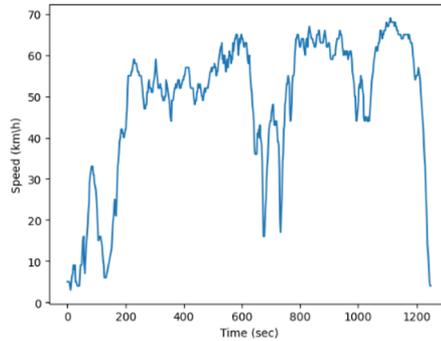


Fig. 1. Dependence of speed on time in a single micro-trip

Additionally, to use the k-means algorithm, it is necessary to determine the required number of clusters into which the data will be divided. For this task, there are methods for selecting the optimal number of clusters based on calculating the average distance from the elements to the center of the corresponding clusters. To calculate the optimal number of clusters, two of the most common methods were used – the "elbow" method and the CH-index method. Based on the analysis using both methods, it was revealed that the optimal number for the problem is 6 clusters.

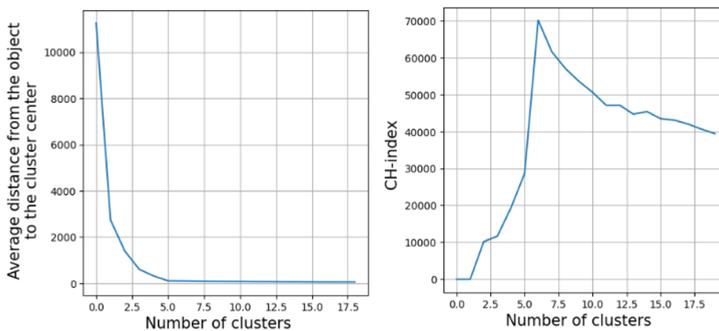


Fig. 2. Illustration for the analysis of the optimal number of clusters using the "elbow" and CH-index methods

The figure (Fig. 3) shows the results of clustering, where each point corresponds to one micro-trip, and the colors of the dots symbolize belonging to the corresponding cluster. To be able to analyze the results, the 6-dimensional space was divided into several two-dimensional ones. Each two-dimensional figure has two criterion presented along the abscissa and ordinate axes. This data performance is very convenient, for example, to identify "useless" parameters that distort the result of the algorithm.

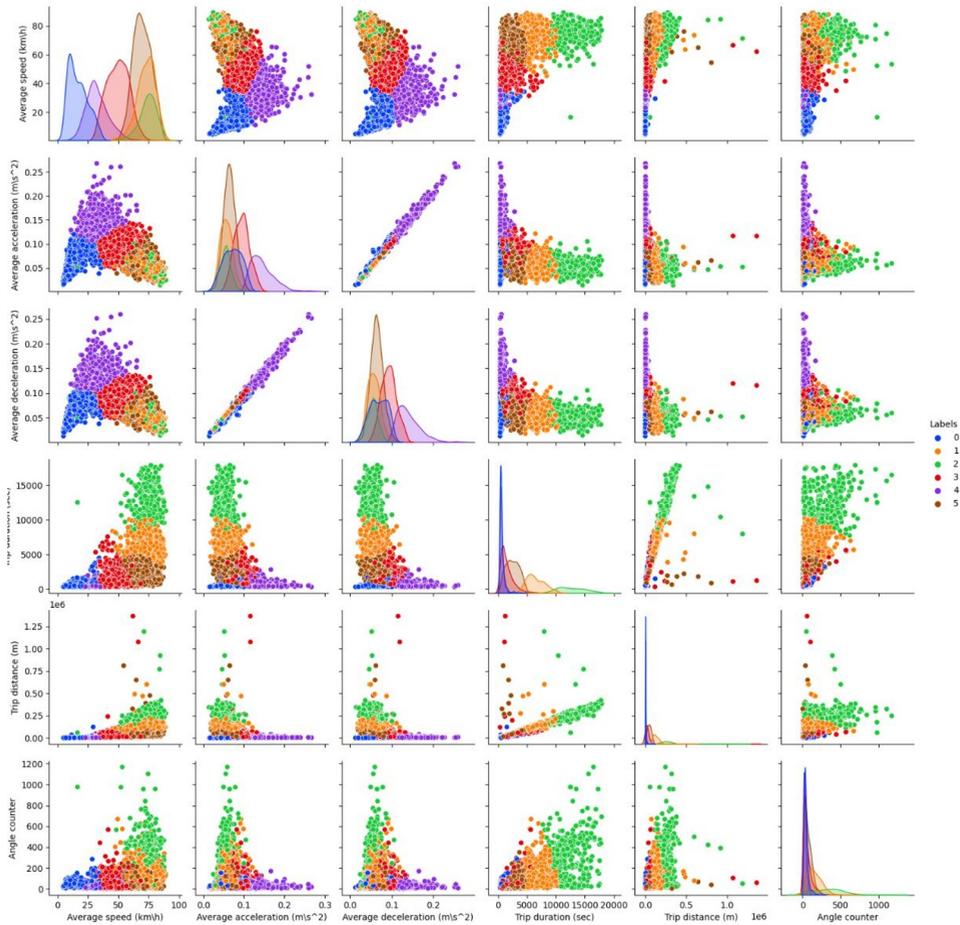


Fig. 3. Illustration of clustering results

To obtain the driving cycle of a vehicle, it is necessary to get a final selection of micro-trips from the clusters. For a more reliable reflection of statistical parameters in the driving cycle, equal shares of micro-trips were selected from each cluster. To exclude from the driving cycle a large amount of data located on the boundaries of clusters, the algorithm for selecting micro-trips from the cluster was refined, in comparison with random. So the probability of choosing a micro-trip from the cluster was inversely proportional to the ratio of the Euclidean measure of the distance to this data sample from the center of the corresponding cluster to the average Euclidean measure of the distance from all elements to the cluster center. At the same time, micro-trips, the distance from which to the center of the cluster is greater than the average value, were completely excluded from consideration.

As a result, an algorithm was developed that generates driving cycles of arbitrary duration. An example of the generated cycle is shown on the graph (Fig. 4).

The main criterion for assessing the quality of the generated driving cycle is its convergence with a complete information of the movement of the car according to a number of criteria discussed in paragraph 1. It is also necessary to find a compromise between the length of the driving cycle and the degree of convergence with the dataset. For this reason, a number of cycles of different lengths were generated to analyze the average deviation by criteria from the dataset (Fig. 5).

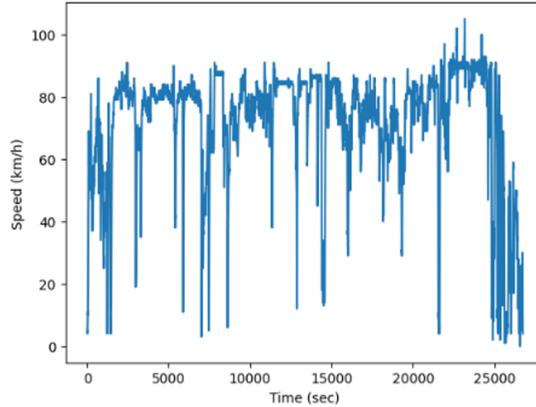


Fig. 4. Example of the constructed driving cycle

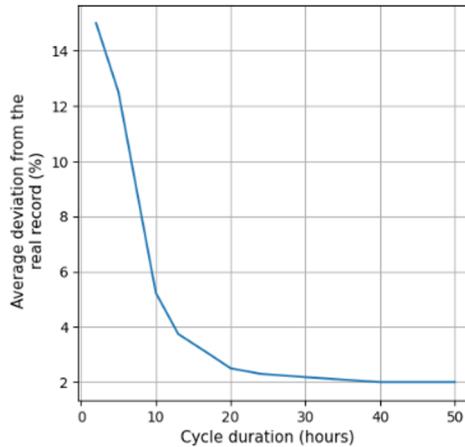


Fig. 5. Convergence comparison of cycles with different duration

As a result of the comparison, it was found that with a driving cycle duration of more than twenty hours, convergence improvement occurs slightly (by less than 0.2%), which means that for the processed data and the method used, the optimal cycle duration is 20 hours.

3 Cycle construction using the Markov chains method

The Markov chains method is used to construct a driving cycle and is a discrete sequence of states, where each is taken from a discrete state space satisfying the Markov property.

The Markov property is presented in formula (1), which shows that the conditional probability distribution of future states of the process depends only on the current state, and not on the sequence of events that preceded it.

$$P(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_{n+1} = x | X_n = x_n) \quad (1)$$

where X_{n+1} – future state, X_n – current state, X_1, \dots, X_{n-1} – previous states.

Since the speed of the vehicle at the next moment of time depends only on the speed of the vehicle at the current moment of time, then modeling its change using Markov chains is possible.

The Markov chains cycle simulation algorithm contains the following steps:

- real vehicle movement data processing;
- formation of vehicle speed states;
- formation of the transition matrix and calculation of transition probabilities;
- cycle construction using Monte-Carlo method.

For further simulation of the driving cycle, states were introduced by vehicle speed and road angle, according to which a transition matrix was compiled, where each element determines the probability of transition from one state to another. The transition probabilities were obtained from the data of a vehicle movement measurements using the formula (2).

$$P_{ij} = P(X_{n+1} = x_j | X_n = x_i) = \frac{Q_{ij}}{\sum_{j=1}^{j=n} Q_{ij}} \quad (2)$$

where Q_{ij} – the number of transitions from state X_n to state X_{n+1} .

After constructing the transition matrix, the Monte Carlo method is used to select simulated vehicle speed data and road slopes from real vehicle movement data.

To form the final cycle of movement, two variations of the Markov chains method were considered [7-8]: a method in which a probabilistic value or characteristic (in our case, the speed of the vehicle) is obtained using a random variable generator, on the basis of which a further selection of a random state is carried out, and a simplified Markov chains method based on the generation of a random state.

The algorithm using a random variable generator consists of four steps [7-8]:

- 1) setting i as the current state;
- 2) randomly selecting the number k in $[0; 1]$;
- 3) making the decision on taking into account under the condition in formula 3:

$$\sum_{j=0}^{j=n-1} P_{ij} < k < \sum_{j=0}^{j=n} P_{ij} \quad (3)$$

- if the outcome is positive, the vehicle speed data is written randomly from the data sample for state j to the end of the driving cycle;
 - if the outcome is negative, the algorithm returns to step 2.
- 4) When the required cycle time is reached, the algorithm operation is terminated, otherwise the algorithm returns to step 1, and state j is considering as the current state.

The algorithm according to the simplified method consists of four steps:

- 1) setting i as the current state;
- 2) randomly selecting a new assumed state j and probability of transition P_{ij} ;
- 3) making the decision on taking into account with probability P_{ij} :
 - if the outcome is positive, the vehicle speed data is written randomly from the data sample for state j to the end of the driving cycle;
 - if the outcome is negative, the algorithm returns to step 2.
- 5) when the required cycle time is reached, the algorithm operation is terminated, otherwise the algorithm returns to step 1, and state j is considering as the current state.

As a result, an algorithm was developed for constructing cycles using Markov chains method. The graph of the speed obtained as a result of the cycle construction is shown on Figure 6.

To check the convergence of the real movement data and constructed cycle, depending on the duration of the cycle, there were constructed cycles with a duration of 1.5, 3, 5, 10 and 20 hours (Fig. 7).

The ordinate axis shows the difference between the average values of the parameters of the simulated cycle from the vehicle movement measurements in percentages. As can be seen from the graph, convergence, at which the difference between the simulated parameters and the real ones is less than 5%, is achieved on the ten-hour duration of the constructed cycle.

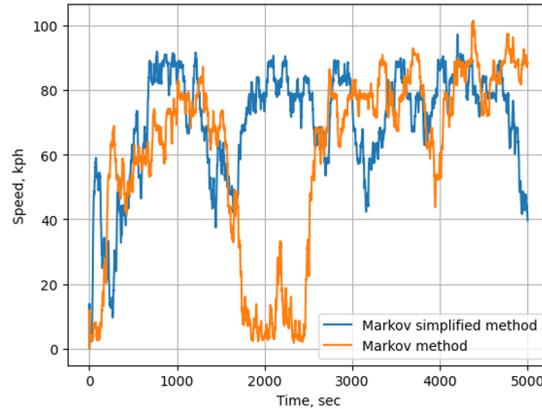


Fig. 6. Velocity graphs obtained as a result of Markov chains simulation

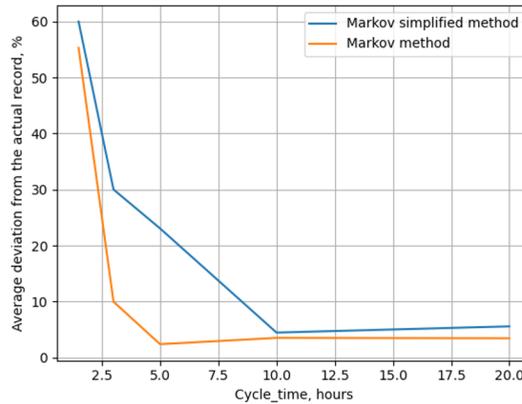


Fig. 7. Convergence of a cycles constructed by the Markov chains method

A variation of the Markov chains method using a random variable generator, unlike the simplified method, has a better convergence of the parameters of the simulated cycle to the real parameters and a higher simulation speed.

4 Comparison of methods for constructing driving cycles

So, the paper considered two of the most popular methods of generating driving cycles of a vehicle. Both methods proved to be effective, as they have good convergence with the dataset with a relatively short duration of the constructed cycles.

Table 1 presents the statistical parameters of the driving cycles of the vehicle in comparison with the dataset on which the construction was based.

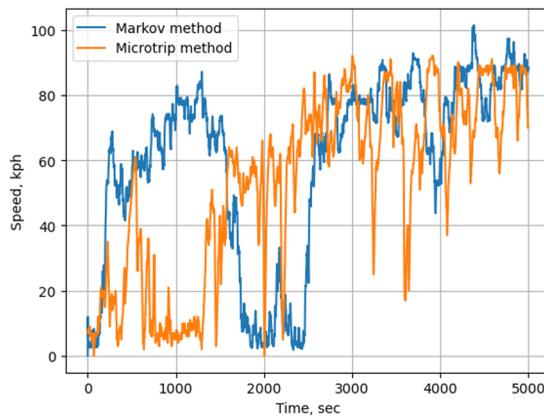
According to Table 1, as well as the graphs Fig. 7 and Fig. 5, it is possible to judge the advantages and disadvantages of both methods.

The method based on micro-trips has greater convergence with the dataset due to the modernization of the classical method using clustering and the method of selecting elements of each cluster. However, the cycles generated by this method begin to have sufficient convergence only when the route time is 15-20 hours, which can be attributed to the disadvantages of the method.

Table 1. Comparison of statistical parameters of the driving cycles and the original vehicle movement measurements

	Real traffic data	Markov Chains	Micro-trip using k-means clustering
Mean speed, km/h	66.1342	65.1646	67.2311
Mean positive slope, degrees	0.933	0.9763	0.9412
Mean negative slope, degrees	-0.9483	-0.9915	-0.9317
Mean acceleration, m/s^2	0.1988	0.1711	0.1963
Mean deceleration, m/s^2	-0.2302	-0.193	-0.2364
Percentage of time spent in acceleration mode	31.83	30.29	30.27
Percentage of time spent in deceleration mode	27.44	30.07	24.82
Maximum speed, km/h	120	113.265	111

Markov chains method has less convergence with dataset, this is most evident in the difference in the average acceleration and deceleration in comparison with the vehicle movement measurements. However, the relative deviation according to the comparison criteria remains acceptable. The advantage of the Markov chains method is that the cycles begin to converge already at a duration of about 10 hours, which significantly differs for the better from the cycles obtained by the micro-trips method.

**Fig. 8.** Comparison of sections of driving cycles obtained using different methods

Analyzing the graphs in Fig. 8, it can be seen that the cycle obtained by the "micro-trips" method has a greater number of sharp speed drops in comparison with the cycle obtained by

the Markov chains method. Since the cycle of micro-trips includes sections from the actual measurements of vehicle movement, it can be concluded that the method of "micro-trips" better reflects the real dynamics of the movement of the target vehicle.

5 Conclusion

As a result of the analysis of the two most common methods of generating driving cycles of a vehicle based on real traffic data, two driving cycles were obtained that can be used in the design of transmission elements and the evaluation of the operational properties of the vehicle. Modeling by the micro-trips method using the k-means clustering algorithm turned out to be preferable in comparison with the approach using the Markov chains method, since the micro-trips method provides greater convergence with the dataset: the deviation of the average parameter values from the real cycle on a twenty-hour cycle for "micro-trips" is less than 3%, and for the Markov chains method – 5%. At the same time, the cycles generated by the "micro-trips" method begin to have sufficient convergence only at a length of 15-20 hours, unlike the Markov chains method, which begins to have sufficient convergence at a length of about 10 hours. Therefore, if a large number of iterations are needed in calculations using driving cycles, preference should be given to the Markov chains method.

References

1. L. Czege, A. Vamosi, I. Kocsis, *IJEMS* **5(2)** (2020) <https://doi.org/10.21791/IJEMS.2020.2.31>
2. R.M. Desineedi, S. Mahesh, G. Ramadurai, *Developing driving cycles using k-means clustering and determining their optimal duration*, WCTR 2019 Mumbai 26-31 May 2019 <https://doi.org/10.1016/j.trpro.2020.08.268>
3. P. Nouri, C. Morency, *Journal of the Transportation Research Board* **2627**, 86–92 (2017) <https://doi.org/10.3141/2627-10>
4. Jie Lin, Debbie A. Neimeier, *Atmospheric Environment* **36**, 5759–5770 (2002) [https://doi.org/10.1016/S1352-2310\(02\)00695-7](https://doi.org/10.1016/S1352-2310(02)00695-7)
5. Zhen Dai, Deb Niemeier, *Douglas Eisinger, Task Order* **66** (2008)
6. Subramanya Nagaraj, *Algorithm to generate Synthetic Driving Cycle using Real driving data*, Chalmers University of Technology (Gothenburg, Sweden 2020)
7. Xiaodong Liu, Jian Ma, Xuan Zhao, Juan Du, Yanfeng Xiong, *Journal of Advanced Transportation* **21** (2020) <https://doi.org/10.1155/2020/3871703>
8. Zhenpo Wang, Jin Zhanga, Peng Liua, Changhui Qua, Xiaoyu Lia, *ICAE2018* **6** (2018) <https://doi.org/10.1016/j.egypro.2019.01.389>