Markov Property Analysis of the Real-World Driving Data and Its Application

Nguyen Thi Yen Lien[1,2], Nghiem Trung Dung[1,*]
[1]Hanoi University of Science and Technology
[2]University of Transport and Communication
Received: March 16, 2017; Accepted: May 25, 2018

Abstract

There are a number of methods that can be used for the development of a driving cycle. Among them, the Markov chain is a promising approach which has been being widely studied in recent years in many developed countries but still is scarce in Vietnam. In this study, the real–world driving data of bus system in Hanoi are proved to be a stationary time-series and have the Markov property, based on the analysis of random number series which are the values of instantaneous speed of the bus recorded by the GPS in a whole bus route. The typical driving cycle of the bus route No.9 was then developed based on this finding. The developed cycle was assessed in the comparison with the on – road driving data. A good conformity of this driving cycle with the real –world driving data was observed.

Keywords: *Markov chain, stationary series, driving cycle, Hanoi bus, emission factor.*

1. Introduction

A driving cycle is a time series of vehicle speeds recorded at successive time points; it is developed from the data collected by driving the testing vehicle on the real road network. The driving cycle provides the basis data for vehicle design and the important index for emission measuring. In recent years, there were several studies on real driving cycle building in Vietnam such as Tong et al (2011) [1] and Tuan Anh et al (2012) [2].

The most widely used method for developing the driving cycle is based on the clustering and combination algorithm of microtrips. One of the most important limitations of this method is that it does not fully reflect the vehicle's operating conditions, typically referred to as the intensity and duration of modal events, within the driving cycle [3, 4]. Jie Lin (2002) was the first to use this method to construct the driving cycle in her doctoral dissertation. In recent years, many scholars have also used Markov chain for designing the driving cycle such as Brady et al (2013) [5], Ashtari et al (2014) [6] and so on.

A Markov chain is a random process with the property that the conditional probability of the process value at any future time depends only on its value at the current time (Gubner, 2006). The description of the present state fully captures all the information that could influence the future evolution of the process. A discrete-time Markov chain is a sequence of random variables X_1, X_2, X_3, ... with the Markov property, namely that, given the present state, the future and past states are independent [7]. The formula is as follows:

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

The set of random variables X_n is called the state space of the chain. The conditional probabilities P_{ij}: = $P(X_{n+1} = j|X_n = i)$ are called transition probabilities. Note that the sum of all probabilities leaving a state must be unity.

$$\sum_j P_{ij} = \sum_j P(X_{n+1}=j \mid X_n=i)=1 \quad (2)$$

In the Jie Lin's dissertation, driving patterns such as acceleration, deceleration, cruising, idling are defined as the states in Markov process. Other scholars in the more recent studies proposed to use vehicle velocity and acceleration as states for the Markov chain to capture features of naturalistic driving [7]. So the information of instantaneous velocity is used in this study to analyze the Markov property of real-world driving data. The analytical results are very important as they provide with a good basis for the application of the Markov chain theory to develop the driving cycle.

2. Methodology

2.1. Vehicle Dynamics Analysis

The most important thing in the application of Markov chain is the selection of the number of states. These states can be determined based on vehicle

* Corresponding author: Tel: +84.912.238.386
Email: dung.nghiemtrung@hust.edu.vn

dynamics equation. The dynamics equation for the vehicle motion is [7]:

$$F_{net} = F_{prop} - F_{RR} - F_{WR} - F_{GR} = m_e.a_{veh} = m_e.\dot{v}_{veh} \quad (3)$$

Where F_{net} is the net force applied to the vehicle, F_{prop} is the propulsion force from the power train, F_{RR} is the rolling resistance force, F_{WR} is the wind resistance force, F_{GR} is the grade resistance force and all other external forces applied to the vehicle, m_e is the equivalent vehicle mass, v_{veh} is the vehicle velocity and a_{veh} is the vehicle acceleration.

In the equation (3), vehicle dynamics can be represented fully by using vehicle velocity and acceleration. Hence, vehicle velocity and acceleration are selected to analyze the Markov property of the real-world driving data and the typical driving cycle that was developed on the basis of the real-world driving data.

2.2. Route selection

The routes should normally be selected in the light of the purposes and objectives of the development of the driving cycles. In this study, to determine the Markov property of the real–world driving data, the bus route No.9 in Hanoi was selected. The summary information of this bus route is presented in Table 1.

Table 1. The information of the bus route No.9 in Hanoi

Route	Type of route	Starting point	Finishing point	Length of the route (km)
09	Closed route	Hoan Kiem Lake	Hoan Kiem Lake	19.7
	Average duration: 4449 seconds Average velocity: 15.9 km/h			

2.3. Data collection

There are three major approaches for collecting real-world driving data that are the chase car method, the on-board measurement, and using Global Positioning System (GPS) technology. Among them, GPS data acquisition devices have been proven to be an useful tool for gathering real-world driving data because the data of vehicle instantaneous speed and position can be captured continuously by this technology [8, 9]. A GPS, Garmin etrex vista HCx, was used in this study to collect the travel data of a bus in the bus route No.9. Data were recorded continuosly, from the starting point at around 6 am to the finishing point at around 8 pm. Data were recorded with the time step of one second to avoid losing information. The MapSource software was used to convert data from GPS equipment to Excel file.

2.4. Data analysis

2.4.1. Test of stationarity

In this study, with the way of data collection as mentioned above, the continuous data of velocity are converted into discrete data to make a sequence of discrete random variables with the time step of one second. So we have a time series (X_t) in which t = 1, 2,…n,… and the value of X_t is an observed instantaneous speed value at time t. This time series must be tested on stationarity before it is used in other analysis. In this study, the method of unit root testing using the Augmented Dickey–Fuller (ADF) tests was used to verify the stationarity for one real-world driving data series of the bus route No.9. The EVIEWS 8.0 software was used to do so. The results show that the real-world driving data series is a stationary time series because the absolute value of the test statistics ($|\tau| = 17.2$) exceeds the test critical values so the time series is stationary [10].

2.4.2. Analysis of Markov property of real-world driving data

The time series (X_t) with 7590 observed points above is used to analyze the Markov property. Supposing that the value of X_t at the current time is V_i, at the future time (next state) is V_{i+1}, and at the past time is V_{i-5}, V_{i-60},... Then (X_t) will have the Markov property if we could determine the fact that the next state velocity (V_{i+1}) depends only on the current state (V_i) and not on the sequence of events that preceded (V_{i-j} with $j \neq 0, \forall j < i$). The analysis of the correlation between the speeds at the future time (V_{i+1}) and the present (V_i) and the past time as (V_{i-5} or V_{i-60}),... were applied to determine this dependence. If the results indicate that there is a good correlation between V_i and V_{i+1}, and that the correlation get worse or that there is almost no correlationship between V_{i+1} and V_{i-5} or V_{i-60} then the real-world driving data are proved to have the Markov property.

Besides, a Matlab code was also developed to discretize the sequence of abovecollected speed data into time series with time step of 5 seconds. So we have a time series (V_t) with 1518 data points. This time series is used to analyze the correlation between V_t and V_{t+1} to determine the suitable time step for the conservation of the Markov property of the sequence of random variables which are instantaneous velocity values.

The collected data are splitted into trips, in which each trip is a instantaneous velocity values series between the starting point and the finishing point. In this study, we chose randomly two trips that describe moving of a vehicle from the starting point to the finishing point and coming back the starting point.

2.5. Driving cycle development

There are a number of methods that can be used for the development of a driving cycle. In this study, the methodology which is based on the Markov chain theory was used. In order to do so, firstly, the velocity-acceleration (VA) probability must be calculated. Therefore, the two-dimension distribution map of velocity and acceleration is divided into grids to define the states. Each grid stands for one kind of state. All one-step state probabilities can be arranged in a matrix which is called the transition probability matrix (TPM), where each element contains the probabilities for every other state to be the next in the chain.

The size of the matrix is determined by the maximum velocity and the absolute maximum acceleration, combined with the resolutions for velocity and acceleration. The number of rows (n_r) and columns (n_c) are calculated as follows [11]:

$$n_r = 2.\frac{|a|_{\max}}{a_{res}} + 1 \quad (4)$$

$$n_c = 2.\frac{v_{\max}}{v_{res}} + 1 \quad (5)$$

When a TPM has been created, it is possible to start generating a driving cycle. The process generating a driving cycle is described on Figure 1. This process is started in the idle state (zero velocity and acceleration).

2.6. Driving cycle evaluation

The assessment of the conformity of developed driving cycle with the real-world driving data is done based on the analysis of the difference between the Speed Acceleration Frequency Distribution (SAFD) of the developed driving cycle and the real-world driving data. In order to calculate SAFD, the speed and acceleration fields are divided into equal cells (being called bins) and the probability for each cell is determined. $SAFD_{diff}$ is the percentage difference between the SAFD of all bins in the real-world driving data and the developed driving cycle as defined by Equation 6. The smaller $SAFD_{diff}$ is, the higher the commonality between the two cycles is [5, 6].

$$SAFD_{diff} = \frac{\sum_i (SAFD_{cycle}(i) - SAFD_{data}(i))^2}{\sum_i (SAFD_{data}(i))^2} \quad (6)$$

Where, i is the i^{th} bin in the SAFD, $SAFD_{cycle}$ is the SAFD of the developed driving cycle, and $SAFD_{data}$ is the SAFD the collected data.

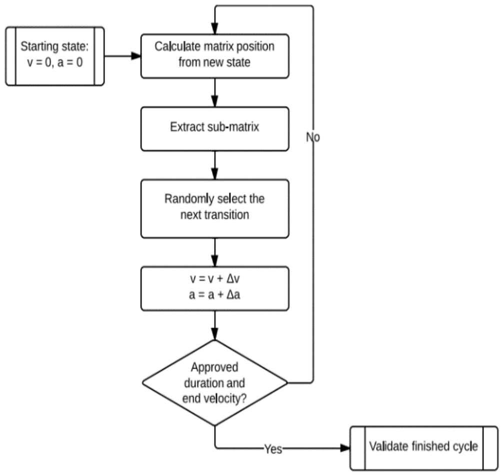

Figure 1. The process generates a driving cycle [11]

3. Results and discussions

3.1. Markov property of the real-world driving data of Hanoi bus

The correlation of speeds between V_i and V_{i+1} is shown on Figure 2. It can be seen that the coefficient of determination, $R^2 = 0.9538$, means that 95.38% of the variability of velocity in the next state (V_{i+1}) is explained by velocity at the current state (V_i). In other words, the correlation of V_i and V_{i+1} is very good, meaning that the velocity of the next state depends strongly on the velocity of the present state. However, the correlation between V_i and V_{i-5} gets worse (Figure 3). And lastly, there is almost no correlation of speeds at current time with the time farther back in the past (for example V_i and V_{i-60}) (Figure 4). In other words, the velocity value at the future state does not depend on the velocity values at past states.

So, the farther values of speed in sequence of random variables are, the weaker correlation between them are. In other words, the velocity of the next time has a very good linear relation with the velocity of the present time but has almost no correlation with the velocity of the past time. It means that the sequence of speeds that were discretized with the time step of one second has Markov property. This finding contributes to verify the results of vehicle dynamics analysis reported in the study of ShumingShi et al.[4]. And the real-world driving data of Hanoi bus only has the Markov property when they are discretized with small time step (one second) because, when they are discretized with time step of 5 seconds, the correlation of V_i and V_{i+1} become weaker (see in Figure 5).

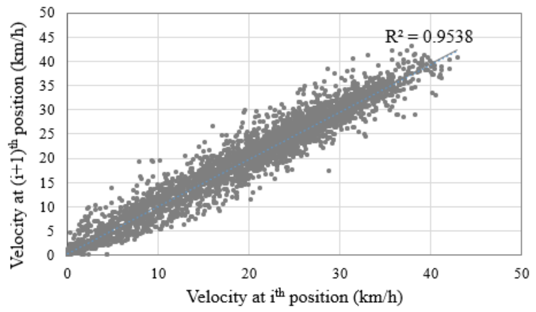

Figure 2. Correlation of V_i and V_{i+1}

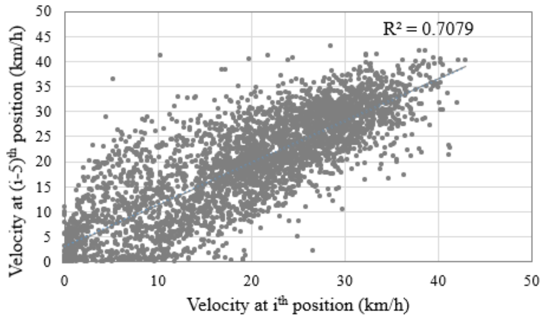

Figure 3. Correlation of V_i and V_{i-5}

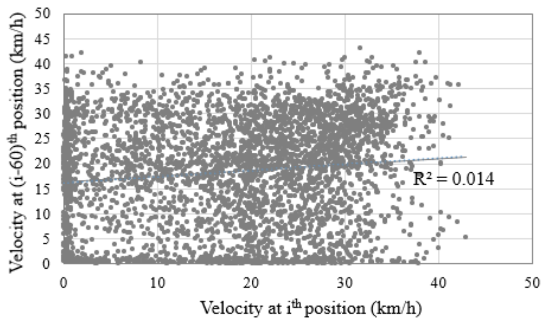

Figure 4. Correlation of V_i and V_{i-60}

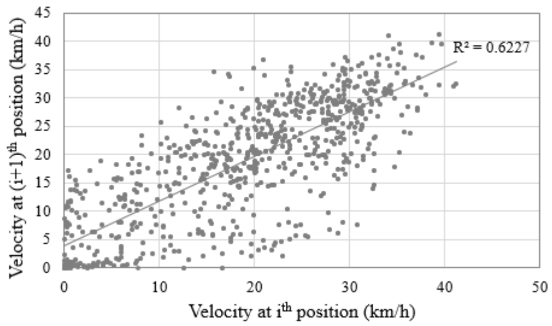

Figure 5. Correlation of V_i and V_{i+1} with time step 5 second

3.2. Typical driving cycle of the bus route No.9

The input data consist of 19 driving cycles where each driving cycle is the data of real-world vehicle instantaneous speeds which were recorded from the starting point to finishing point of the bus route No.9. With v_{res} = 1km/h and a_{res} = 0.2m/s^2, the size of TPM which was determined based on the input data is 51 rows and 53 columns.

The proposed driving cycle of the bus route No.9 was developed based on the Markov theory and is shown on Figure 6. The driving cycle characteristics is shown on Table 2.

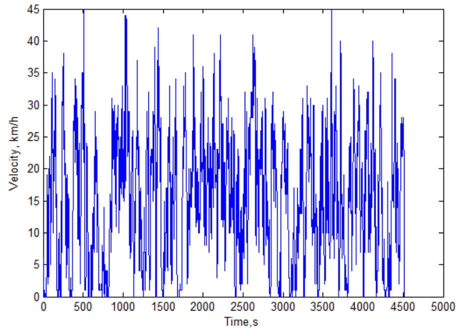

Figure 6. Speed-time curves of representative driving cycle

Table 2. The characteristic parameters of proposed driving cycle for the bus route No.9

<table>
<tr><th>Parameters</th><th>Values</th><th colspan="2">Parameters</th><th>Values</th></tr>
<tr><td>Length (km)</td><td>19.65</td><td rowspan="5">Time proportion of driving modes (%)</td><td>Idling</td><td>6.01</td></tr>
<tr><td>Duration (s)</td><td>4174</td><td>Acceleration</td><td>34.23</td></tr>
<tr><td rowspan="3">Average velocity (km/h)</td><td rowspan="3">16.9</td><td>Deceleration</td><td>31.43</td></tr>
<tr><td>Cruising</td><td>16.60</td></tr>
<tr><td>Creeping</td><td>11.73</td></tr>
</table>

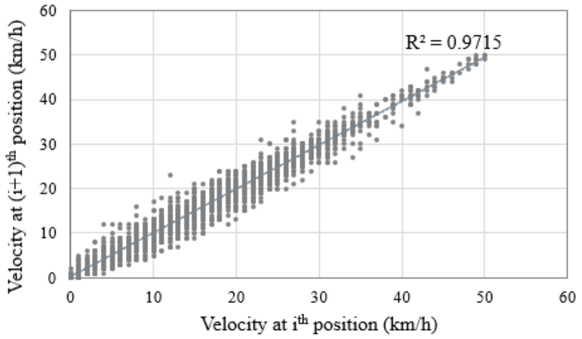

Figure 7. Correlation of speeds between one second in the developed driving cycle

The velocity data of this driving cycle are then used to analyze the correlation of speed between one second (V_i and V_{i+1}) (Figure 7).

It can be seen from Figure 7 that the velocity of the present time and next time has a good linear relation (the correlation coefficient is 0.9856). This result is very similar to the one which was based on the FTP test cycle and used in the study of Shuming Shi et al.[4].

3.3. Conformity estimation

Conventionally, the conformity of the developed cycle and the real-world driving data is estimated based on the $SAFD_{diff}$ in Equation 6.

For the recorded driving data of the bus route No.9 and the developed driving cycle, the $SAFD_{diff}$ is 11.3%. This result is smaller (meaning better) than other studies which used other methods (microtrip or trip snippets) for designing the driving cycle. For example, the test cycle FTP72 has $SAFD_{diff}$ = 17.7%, the test cycle FTP75 has $SAFD_{diff}$ = 30.4% [6]. With the smaller $SAFD_{diff}$, it can be concluded that the driving cycle developed in this study is fitted in the real-world driving data better than other studies. Therefore, the driving cycle, which is based on the Markov chain theory, can be more representative for the real-world driving data because the real-world driving data has Markov property as mentioned in Section 3.1.

4. Conclusions

The relationship of speed versus time at any moment in the driving process is uncertain, so an actual driving cycle is needed to be considered as a random process. In this study, the continuous series of real-world vehicle speeds were discretized with step of one second to avoid losing information, therefore we have a time series (speed vs. time). This time series is a stationary time series, in which, its stationarity was tested based on the EVIEWS software. This time series also has a very good correlation of speed values with the time step of one second (V_i and V_{i+1}), and this correlation gets worse when using longer time steps.

For the driving data that are discretized with time step of one second, the velocity value at the next time depends only on the velocity value at present time and does not depend on the velocity value in the past.

Consequently, it can be concluded that the collected real-world driving data of Hanoi bus has the Markov property when they are discretized with time step of one second. Therefore, when developing the driving cycle for Hanoi bus, it is recommended to use the methods that follow these essential characteristics. A test of the development of the driving cycle which is based on the Markov theory was performed with the real-world driving data of the bus route No.9 in Hanoi. The driving cycle developed is fitted very well in the real-world driving data ($SAFD_{diff}$ = 11.3%). In this approach, the nature of randomness of the driving data can be reserved. Therefore, the development of the driving cycle based on the Markov chain theory is a proper and promising approach.

References

[1]. Tong H. Y., H. D. Tung, W. T. Hung,H. V. Nguyen, Development of driving cycles for motorcycles and light-duty vehicles in Vietnam, Atmospheric Environment, 45 (2011) 5191-5199.

[2]. Le T. A., T. M. Pham, T. T. Nguyen,V. D. Nguyen, Measurements of Emission factors and Fuel Consumption for Motocycles on a Chassis Dynamometer based on a localized driving cycle, ASEAN Engineering Journal, part C, 1 (2012) 74-86.

[3]. Lin J.,D. A. Niemeier, Estimating Regional Air Quality Vehicle Emission Inventories: Constructing Robust Driving Cycles, Transport Science, 37 (2003) 330-346.

[4]. Shi S., N. Lin, Y. Zhan, J. Cheng, C. Huang, L. Liu,B. Lu, Research on Markov property analysis of driving cycles and its application, Transportation Research Part D, 47 (2016) pp. 171-181.

[5]. Brady J.,M. O.Mahony, The development of a driving cycle for the greater Dublin area using a large database of driving data with a stochastic and statistical methodology, Proceedings of the ITRN2013, (2013)

[6]. Ashtari A., E. Bibeau,S. Shahidinejad, Using Large Driving Record Samples and a Stochastic Approach for Real-World Driving Cycle Construction: Winnipeg Driving Cycle Transportation Science, 48 (2014) pp.170 - 183.

[7]. Lee T.-K.,Z. S. Filipi, Synthesis of real-world driving cycles using stochastic process and statistical methodology, International Journal of Vehicle Design, 57 (2011) p.17-36.

[8]. Tong H. Y.,W.T.Hung, A Framework for Developing Driving Cycles with On-Road Driving Data, Transport Reviews: A Transnational Transdisciplinary Journal, 30 (2010) 589-615.

[9]. Niemeier D. A., T. Limanond,J. E. Morey, Data collection for driving cycle development: evaluation of data collection protocols, Department of Civil and Environmental Engineering, Institute of Transportation Studies, University of California, Davis (1999).

[10]. Gujarati D. N., Basic Econometrics, Fourth Edition, McGraw−Hill (2004).

[11]. Önnegren E. T. o. P., Driving cycle generation using statistical analysis and Markov chain, Department of Electrical Engineering, Likopings University, Sweden (2013).