

An Iterative Markov Chain Approach for Generating Vehicle Driving Cycles

2011-01-0880

Published
04/12/2011Qiuming Gong, Shawn Midlam-Mohler, Vincenzo Marano and Giorgio Rizzoni
Ohio State Univ.

Copyright © 2011 SAE International

doi:10.4271/2011-01-0880

ABSTRACT

For simulation and analysis of vehicles there is a need to have a means of generating drive cycles which have properties similar to real world driving. A method is presented which uses measured vehicle speed from a number of vehicles to generate a Markov chain model. This Markov chain model is capable of generating drive cycles which match the statistics of the original data set. This Markov model is then used in an iterative fashion to generate drive cycles which match constraints imposed by the user. These constraints could include factors such number of stops, total distance, average speed, or maximum speed.

In this paper, systematic analysis was done for a PHEV fleet which consists of 9 PHEVs that were instrumented using data loggers for a period of approximately two years. Statistical analysis using principal component analysis and a clustering approach was carried out for the real world velocity profiles. After dividing the real velocity profiles into segments, they were clustered into several different clusters based on statistical data. This data set is also used to generate the Markov chain model technique described above which is the central development of this work. The work of this paper is a part of a larger project in which a mass simulation of a neighborhood of PHEVs will be conducted based on statistical representations of key factors such as vehicle usage patterns, vehicle characteristics, and market penetration of PHEVs. This approach can be used as critical input of the large scale simulation model to generate random velocity profiles of various driving patterns.

INTRODUCTION

The PHEV fleet study project which developed the vehicle database used in this work is a part of a broader research

program underway at The Ohio State University's Center for Automotive Research called SMART@CAR. The main goals of this project are to create and maintain a database containing all charging and duty cycle data collected from a growing PHEV fleet and to use this data to make predictions of the impact of these vehicles on the grid in the near future. The detailed information about the database can be found in [1]. The database includes data for driving and charging sessions of different aspect of the vehicle variables.

The first step of the research is to analyze the usage patterns of a PHEV fleet based on collected real driving data. The next step is then to build a PHEV energy analysis model which can make use of the usage patterns of the PHEV customers and vehicle characteristics to estimate the energy consumption of the PHEV fleet. The final goal is to study the impact of PHEVs on the electric grid on a neighborhood scale through large scale simulation of vehicles. With stochastic inputs, like vehicle weight, generated velocities, initial SOC, *etc.*, the approach is designed to estimate the energy consumption and fuel economy results of individual vehicles as well as the overall fleet. The model also includes driving pattern recognition (DPR) to identify different types of driving cycles and estimate performance. The schematic diagram of the energy analysis model is shown in [Figure 1](#), where p stands for probability, and v is velocity.

To generate velocity profiles for different driving pattern, the driving patterns should be defined first. The importance of driving patterns to the control of vehicles and energy management of PHEVs were pointed out in many research works. Driving patterns, such as road type and traffic condition, trend and style, and vehicle operation modes have various degrees of impacts on vehicle fuel consumptions [2, 3, 4, 5, 6, 7]. The knowledge about online driving pattern was

incorporated into control strategies of intelligent vehicle control in [8, 9, 10, 11, 12, 13, 14].

Some of these works tried to define the driving patterns in different ways. A detailed description about the driving pattern factors was discussed and their influence on fuel use and exhaust emission were analyzed in [6]. To find independent measures to describe the dimensions of urban driving patterns and to investigate which properties have main effect on emissions and fuel-use, 62 driving pattern parameters were calculated for each of 19,230 driving patterns collected in real traffic. The author used factorial analysis approach to reduce the initial 62 parameters to 16 independent driving pattern factors. Driving pattern recognition was used in [8, 9, 10]. In [8], Jeon *et al.* selected 6 RDPs (Representative Driving Pattern), including urban, suburban and expressway. On each RDP, optimized parameters for control were found by off-board calculation. Artificial Neural Network (ANN) based driving pattern recognition was conducted for real-time control. Once the current driving pattern was distinguished, optimized parameters for the selected RDP were used in the real time sub-optimization. While in [9], Lin *et al.* used similar method as that in [8]. Instead of using ANN, a simple rule-based control strategy was used to recognize the RDP. In [10], 5 RDP was first generated by rules; then a rule-based control algorithm was extracted from the result of DP on each generated RDP, finally, a multi-mode driving control was realized by switching the control parameters in each RDP. Bo *et al.* proposed a method that used statistical pattern recognition and results in a very simple algorithm that was easily implemented in real time in [11]. In [12-13], an intelligent energy management agent (IEMA) was used for parallel hybrid vehicles. They used only 40 of the 62 parameters as in [8] and then added seven new parameters for the driving pattern recognition. Driving information was incorporated into the vehicle power management in [14]. An intelligent system was developed to predict the current traffic conditions and Neural learning for predicting the driving environment, such as road type and traffic congestion, was developed.

For generating velocity profiles, some previous research work tried to use the information from external systems like ITS (Intelligent Transportation Systems) or GIS (Geographic Information Systems) [15]. The model approach for the urban data in that paper can't represent the real driving pattern precisely by using some simple models and the complicated modeling approach for highway part is too difficult to calibrate and thus not easy to use for the real time implementation.

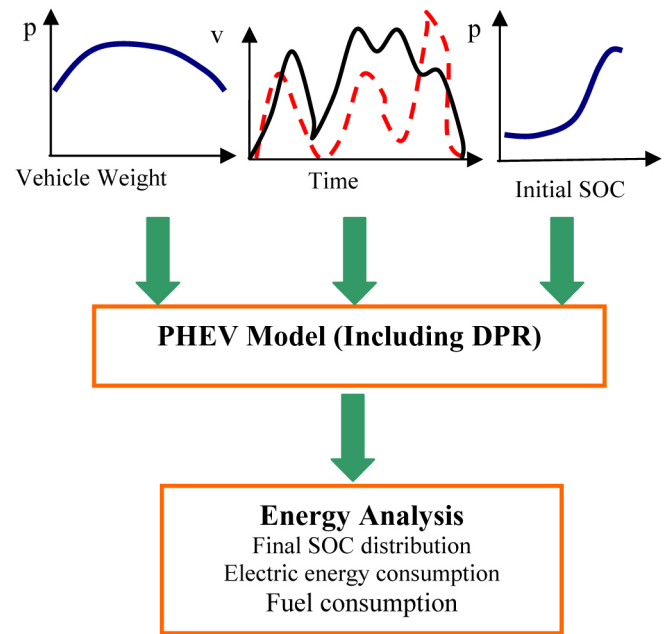


Figure 1. Monte Carlo simulation of PHEV energy analysis

Some research tried to use traffic information in vehicle control, or power management of HEV, PHEV [16-17]. Similar to [11], statistical analysis and clustering approach were used for driving cycles in [18]. The authors divided the driving cycles into segments called kinematic sequences and then clustering approach was used to separate those kinematic sequences into groups based on statistical information. It also proposed an approach to generate the driving cycle by choosing the kinematic sequences from the existing database randomly following distributions obtained from the kinetic sequences. For long and regular driving cycles, the approach may be effective and appropriate. However, to generate short and more precise driving cycles, the approach is not adequate. Markov chain models are an effective way to generate a representative driving pattern in a statistical way. Authors in [19] tried to divide the velocity profiles into segments, and clustered them into different model event bins, then, Markov train process was applied to these model events for constructing the velocity profiles. In [20], a Markov chain model was used for the velocity profiles generation based on observations from some standard driving cycles.

In this paper, a Markov chain model will be combined with some traffic information to generate more precise velocity profiles for different driving patterns. With mass real driving data available for a PHEV fleet, a systematical statistical approach was carried for the real driving data, then a Markov chain model was used based on the clustering results. Furthermore, traffic information was used as the constraints for the stochastic velocity generation model. With this information, more precise velocity profiles can be generated

Table 1. Charging frequency compare of 9 PHEVs

PHEV No.	1	2	3	4	5	6	7	8	9
Charging Frequency Ratio	1:3.5	1:2.06	1:2.2	1:8.16	1:2.69	1:4.67	1:5.95	1:3.95	1:4.05

for different driving pattern under different traffic density conditions.

FLEET DATA COLLECTION AND INITIAL COMPARISONS

In paper [21], real world data of 6 PHEVs of about 1 year driving were collected at that point. While in this paper, 9 PHEVs' real word data of approximately two years driving were collected from a growing database. This database currently contains more than 70,000 miles of vehicle data. Similarly the data collection approach and general analysis approach were adopted in this paper also. From the available datasets, the important variables for charging and driving were selected respectively. Real-world data of the selected variables for 9 PHEVs were compared and analyzed to gain insight of the charging and driving patterns of various customers and PHEV models. Figure 2 shows some distributions of several variables based on the real driving data; the distributions of those variables show the usage patterns of the PHEVs. The obtained statistic results will also be used for the stochastic input model development, such as the driving distance, initial SOC distribution, etc.

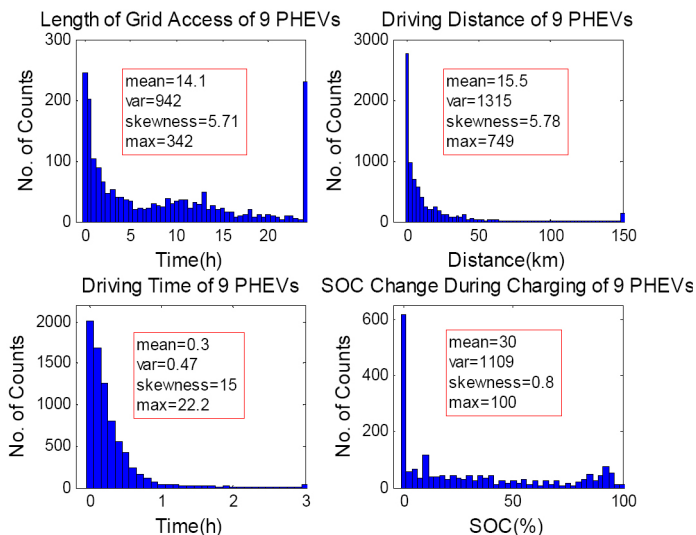


Figure 2. Data analysis for 9PHEVs

To see the similarities and differences of these 9 PHEVs, the mean, variance, skewness of the variables chosen for the study were collected and compared. In Figure 3, the mean of some key variables are compared. The data sets indicate the similarities in some of the driving and charging patterns in some PHEVs while also show difference in some PHEVs. We define the ratio of charging activity as: Number of

charging sessions/Number of driving sessions. This value shows the charging frequency of the PHEV. For the 9 PHEVs, the values of this ratio are collected in Table 1. The results show that PHEV No.4 and No.7 are less frequently charged compared to other PHEVs on a per trip basis.

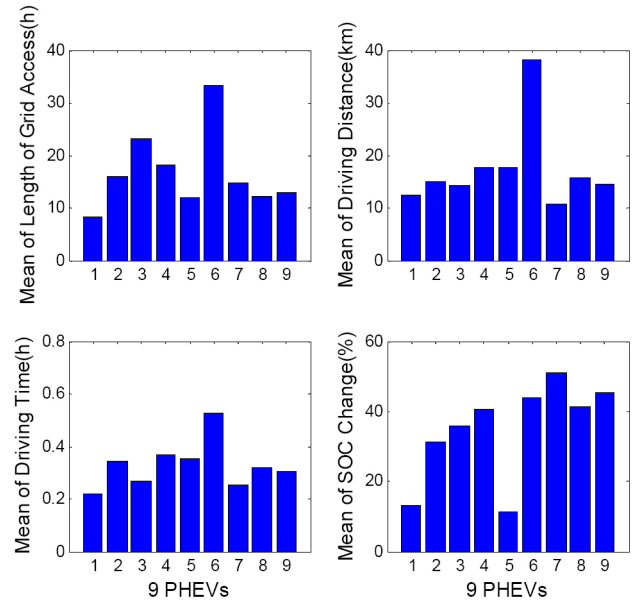


Figure 3. Mean comparison of some data

STATISTICAL ANALYSIS OF REAL VELOCITY PROFILES

In this section, statistical analysis of real world velocity profiles is discussed to define different driving patterns. The driving data from a single vehicle from May 2009 to Apr 2010 was collected and analyzed. For this vehicle, there are 673 driving sessions during this period. Here, a driving session is defined as the duration from key on to key off in driving. After observation of the data, very short sessions usually are meaningless sessions, such as all zero data. To get much cleaner and complete driving sessions, those bad sessions were deleted. After deleting sessions with length less than 250 seconds and sessions with very short distance, there are 530 sessions left. The velocity profiles of these 530 driving sessions were used for the statistical analysis as described in this section.

Statistical Metrics

For the real driving velocities, a systematic statistical analysis was carried out. The velocity profiles are divided into smaller segments such that each segment is formed by velocity

profile between two consecutive velocity stops. The statistical analysis and clustering is based on these velocity segments. A set of statistic characteristics is used to describe each velocity segment, as shown in [Table 2](#). This statistical metric is adapted from [11] [21-22]. In this paper, the segments were used for study, two of the statistical variables were not used: PKE, per_idle_t, since these two variables are not very suitable to describe segment based analysis. Instead, as pointed out in [18], driving distance (d_cyc) and driving time (t_cyc) were used. In this way each velocity segment is represented by a 18 element vector, Driving Cycle Statistical Metrics Vector (DCSMV), $DCSMV_{org,1 \times 18}$.

Table 2. Statistical metric of driving cycles

Statistical Metric	
V_ms_mean	Mean Velocity
V_run_ms_mean	Mean Run Velocity
V_peak_ms	Peak Velocity
a_peak	Peak Acceleration
d_peak	Peak Deceleration
a_mean	Mean Acceleration
d_mean	Mean Deceleration
a_RMS	Root-Mean-Square of Acceleration
PKE	Positive Acceleration Kinetic Energy per Unit Distance (PKE)
per_dec_t	Percent of Time Decelerating
per_acc_t	Percent of Time Accelerating
per_cruise_t	Percent of Time Cruising
per_idle_t	Percent of Time Idling
per_acc_dist	Percent of Distance Accelerating
per_dec_dist	Percent of Distance Decelerating
Per_cruise_dist	Percent of Distance Cruising
V_std	Standard Deviation of Velocity
a_std	Standard Deviation of Acceleration

Principal Component Analysis

From the statistical metrics, 18 metrics form the state vector, which indicates the characteristics of a certain driving cycle. However, some metrics may be strongly related. Such repetition increases the complexity of calculation and principal component analysis (PCA) is used to reduce the size of state vector.

PCA involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. These principal components form a group of orthogonal basis for space of the data. Standardizing the data is often preferable when the variables are in different units. In this case, data are standardized by dividing each component of the vector by its corresponding standard deviation among all the velocity segments. The output of principle component analysis is a

coefficient matrix that converts large dimensional vector to a lower dimensional vector.

First, all 18 metrics among all the collected driving cycles are standardized, as shown in [Figure 4](#). Sometimes it makes sense to compute principal components for raw data. However, this is only appropriate when all the variables are in the same units. Standardizing the data is often preferable when the variables are in different units. In this case, data are standardized by dividing each component of the vector by its corresponding standard deviation among all the driving cycles [11]. From the standardized data, segment time and distance vary much among all the segments compared to other variables.

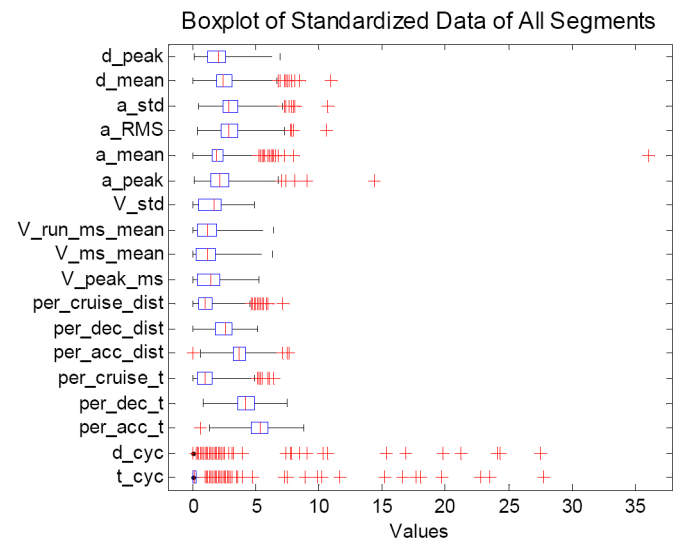


Figure 4. Boxplot of the 18 standardized statistical variables of all segments

The standardized cycle state vector can be used in principal component analysis. In [Figure 5](#), the curve of the accumulated variance shows that the first 5 principal components determined from the PCA represent more than 90% of the variance of the original cycle state vector variables. Thus, the 1 by 18 cycle vector can be converted into a 1 by 5 vector, for the purpose of data reduction. [Equation \(1\)](#) shows how original Driving Cycle Statistical Metrics Vector (DCSMV) is converted into a reduced size vector.

$$DCSMV_{PCA,1 \times 5} = DCSMV_{org,1 \times 18} \times Coef_{PCA,18 \times 5} \quad (1)$$