# Statistical Analysis of PHEV Fleet Data

Qiuming Gong, Shawn Midlam-Mohler, Vincenzo Marano, Giorgio Rizzoni, Yann Guezennec
Center for Automotive Research
The Ohio State University
Columbus, USA

***Abstract*—The added load that a PHEV (Plug-in Hybrid Electric Vehicle) fleet imposes on the existing electrical grid is of great concern to the electric utility industry. In this paper, analysis was done for a PHEV fleet which consists of 6 PHEVs that were instrumented using data loggers for a period of approximately one year. Systematic analysis using a clustering approach was carried out for the real world velocity profiles. A driving pattern recognition algorithm was developed based on the clustering of the results and Markov-chain model was used for the stochastic velocity generation for different driving patterns. 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.**

***Keywords- PHEV, grid interaction, fleet study***

## I. INTRODUCTION

This paper is based on the work of a plug-in hybrid electric vehicles (PHEV) fleet study. The project is 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. The detailed information about the database can be found in [1]. The database of PHEV data for use by the consortium members and university researchers has been developed to provide insights on PHEV usage patterns, impacts on vehicle performance and to the electrical grid.

The objective of this research is first to analyze the usage patterns of a PHEV fleet based on collected real driving data. The next phase is 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. The schematic diagram of the energy analysis model is shown in Fig.1.

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.

Driving patterns have great impact on fuel economy of vehicle or power split control strategies of PHEVs. Various research works have suggested that road type and traffic condition, trend and style, and vehicle operation modes have various degrees of impacts on vehicle fuel consumptions [2-7]. Some researchers in intelligent vehicle power control have begun to explore the ways to incorporate the knowledge about online driving pattern into control strategies [8-14].

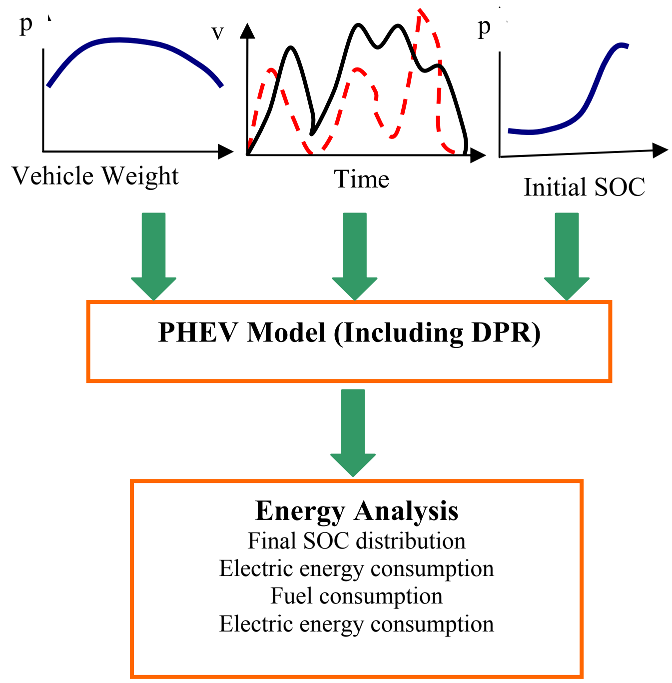


Figure 1. Monte Carlo Simulation of PHEV Energy Analysis

In [6], a detailed description about the driving pattern factors was discussed and their influence on fuel use and exhaust emission were analyzed. The paper aimed 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 19230 driving patterns collected in real traffic. By using factorial analysis, the initial 62 parameters were reduced to 16 independent driving pattern factors.

In [8], Jeon *et al.* selected 6 RDPs (Representative Driving Pattern), including urban, suburban and expressway. On each RDP, optimized parameters for control are found by off-board calculation. For real-time control, driving pattern recognition is conducted by using Artificial Neural Network (ANN). Once the current driving pattern is distinguished, optimized parameters for the selected RDP are used in the real time sub-optimization. In [9], Lin *et al.* used similar method as that in [8]. Instead of using ANN, a simple rule-based control strategy is used to recognize the RDP. In [10], Jeon *et al.* first generated

5 RDP by rules; then a rule-based control algorithm is extracted from the result of DP on each generated RDP; at last, a multi-mode driving control was realized by switching the control parameters in each RDP. In [11], Bo *et al.* proposed a method that uses statistical pattern recognition and results in a very simple algorithm that is easily implemented in real time. In [12-13], the author proposed an intelligent energy management agent (IEMA) 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. In [14], the author incorporated the driving information into the vehicle power management. An intelligent system has to be developed to predict the current traffic conditions. Neural learning for predicting the driving environment, such as road type and traffic congestion, was developed. In this paper, the approach described in [11] will be used for the driving pattern study and recognition.

For prediction of the driving patterns, some work was based on external systems like traffic information system [15]. Modeling approaches were also used for the driving cycle generations in those papers. However, the complicated modeling approach gave much difficulty in calibration and real time implementation. In [16], statistical analysis and clustering approach were used for driving cycles. It divided the driving cycles into segments called kinematic sequences. 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. For long and regular driving cycles like in that paper, 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.

The diagram in Fig. 1 represents the overall objective of this research, while this paper is focused on the statistical analysis of the real world driving data and the stochastic velocity generation model based on Markov train model. In section II, fleet data of PHEVs will be introduced and some initial comparisons will be discussed. Statistical analysis of real velocity profiles will be studied in section III. In section IV, a Markov chain model-based velocity generation approach will be discussed following by some simulation results in section V. Finally a conclusion will be made in section VI.

## II. Fleet Data Collection and Initial Comparisons

### A. *Fleet Data Collection*

In this paper, 6 PHEVs' real word data were collected from a growing database. This data base currently contains more than 50,000 miles of vehicle data. From the available datasets, the important variables for charging and driving are selected respectively. Real-world data of the selected variables for 6 PHEVs are 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*.

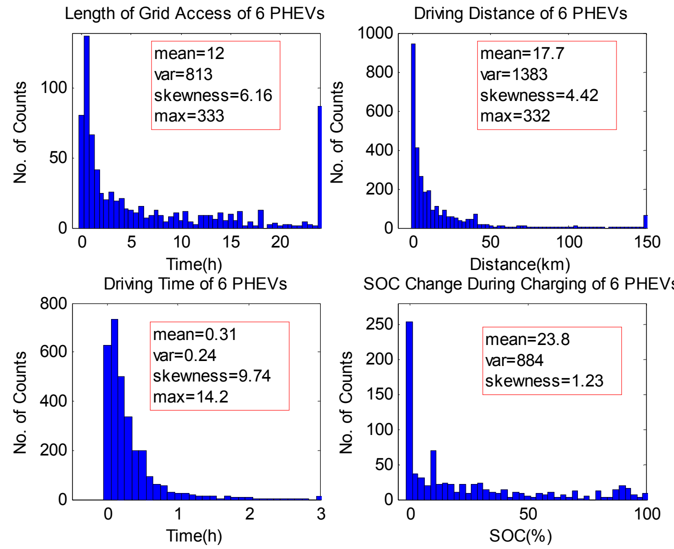


Figure 2. Data Analysis of 6 PHEVs (a)

### B. *Initial Comparison of Fleet Data*

To see the similarities and differences of these 6 PHEVs, the mean, variance, skewness of the variables chosen for the study were collected and compared. In Fig.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.

The cumulative grid access time of the 6 PHEVs are compared in Fig.4. Five of them have very similar trend, while the 6th one is quite different. The normalized figure is obtained by dividing the each data set by the summation of that data set. The figures show the charging pattern of the 6 PHEVs.

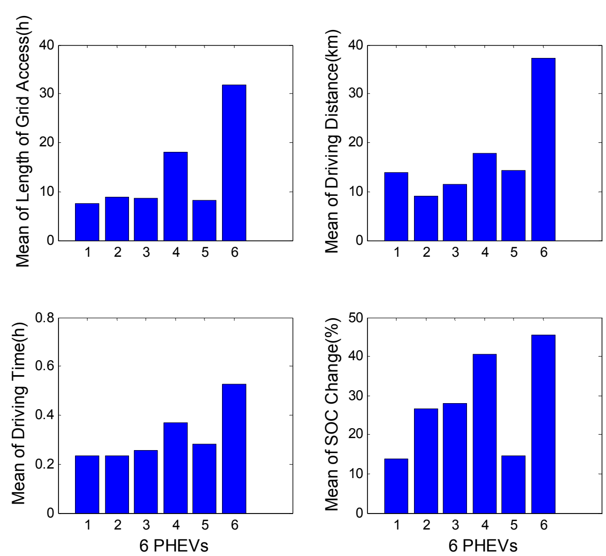


Figure 3. Mean Comparison of Some Data

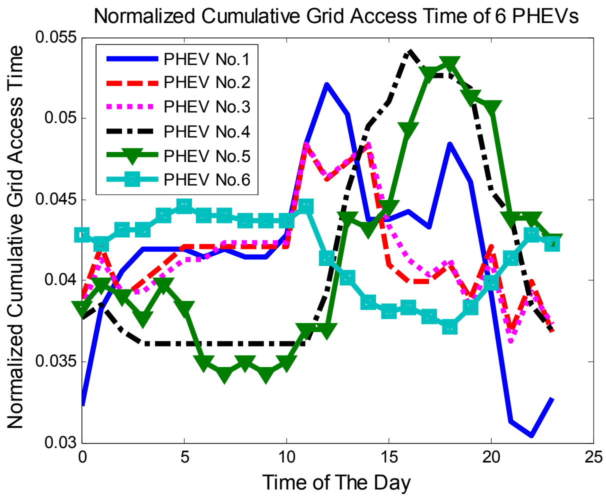

Figure 4. Normalized Cumulative Grid Access Time of 6 PHEVs

### C. Charging Frequency Comparison

We define the ratio of charging activity as: Number of charging sessions/Number of driving sessions. This value will show the charging frequency of the PHEV. For the 6 PHEVs, the values of this ratio are 1:3.5, 1:2.06, 1:2.2, 1:8.16, 1:2.69, 1:4.67 respectively. The results show that PHEV No.4 and No.6 are less frequently charged compared to other PHEVs on a per trip basis.

## III. STATISTICAL ANALYSIS OF REAL VELOCITY PROFILES

In this section, the driving data from a single vehicle from May 2009 to Apr 2010 is 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. So after deleting sessions with length less than 250 s 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.

### A. Statistical Metrics

In [17], 21 statistical metrics are used to represent the characteristic of a driving cycle. The same approach was also used in [11]. However, some statistical metrics are not necessarily needed in the classification of driving cycles, such as the “Duration”, “Distance” and “Number of Stops per Unit Distance”. Thus, the remaining 18 statistical metrics were used in [11] which were also used in this paper as listed in Table 1.

TABLE 1. 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_aec_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 |

### B. Principal Component Analysis

From the statistical metrics, 18 metrics form a state vector, which indicates the characteristics of a certain driving cycle. However, some metrics may be strongly related, for example the “Percent of Distance Accelerating” and “Percent of Time Accelerating”. Such repetition increases the complexity of calculation. The idea of principal component analysis is used to reduce the size of state vector.

First, all 18 metrics among all the collected driving cycles are standardized, as shown in Fig.5. 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.

Then the standardized cycle state vector can be used in principal component analysis. In Fig.6, 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.

$$DCSMV_{PCA,1\times5} = DCSMV_{org,1\times18} \times Coef_{PCA4,18\times5} \qquad (1)$$

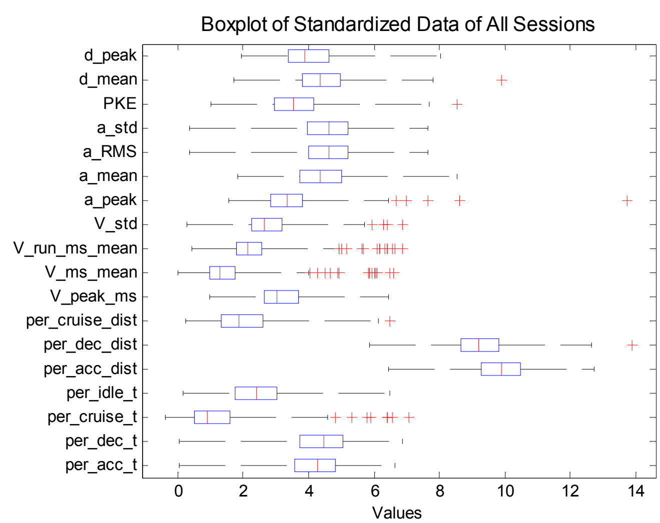

Figure 5. Boxplot of the 18 Standardized Statistical Variables of All Driving Sessions

### *C. Clustering*

A clustering approach is then used to group the velocity profiles into classes. Through a trial and error approach, a two group clustering is found to be better than 3 or more by looking at the silhouette plots of the trials. The silhouette values of each component in each class are shown in Fig.7. The silhouette plot displays a measure of how close each point in one cluster is to points in the neighboring clusters. This measure ranges from +1, indicating points that are very distant from neighboring clusters, through 0, indicating points that are not distinctly in one cluster or another, to -1, indicating points that are probably assigned to the wrong cluster [18]. The average velocity and average absolute value of acceleration for each session are collected for both clusters, and relation between this value and driving time are compared in Fig.8.

The differences between the clusters in Fig.8 indicate the effectiveness of the clustering. In the clustering results, cluster 1 has typically urban driving. In cluster 2, it includes all the typical highway driving, but it also contains some urban driving sessions which may have close distance to typical highway driving in statistical sense.

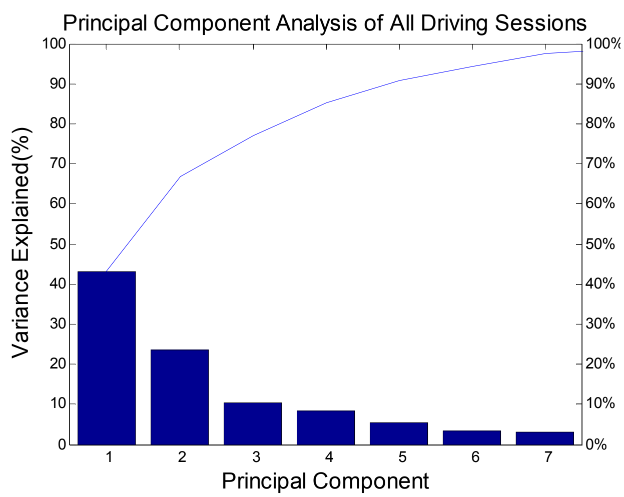


Figure 6. Principal Component Analysis of All Driving Sessions

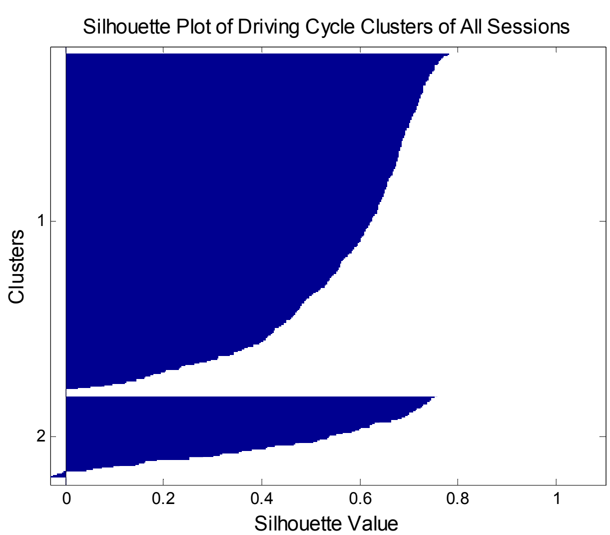


Figure 7. Clustering Results of All Driving Sessions

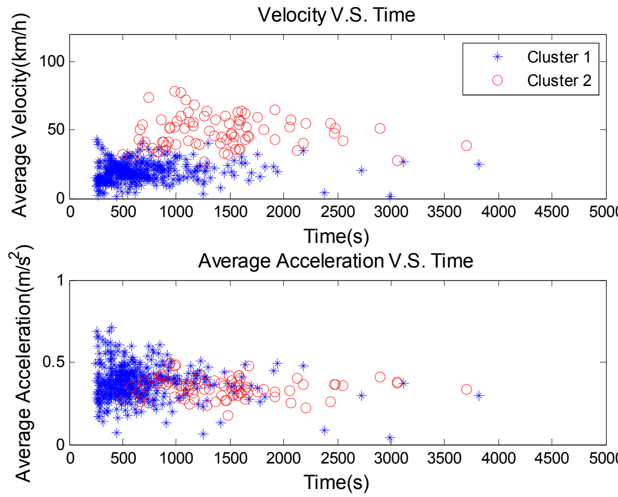


Figure 8. Comparison of Average Velocity and Average Acceleration of the Two Clusters

The resulting statistical data from this step will be used for the later driving pattern recognition and the grouped velocity profiles will be used for the stochastic velocity generation using Markov-chain model.

## IV. Markov Chain Model Based Velocity Generation

In [19-20], stochastic DP was used for the power management of conventional HEV and PHEV. A stationary Markov chain was used to generate the power demand from the driver. Here, we will use the same approach to generate a velocity profiles instead of the power demand. The Markov model is described in (2), which shows that the next state of the system is just dependent on the state of current step.

$$Prob\{X(t+1) = j | X(t) = i\} = f(i, j) \tag{2}$$

For our case, the state vector in the Markov chain is defined as $X_k = (V_k, a_k)$ where *a* is acceleration and *V* is velocity, and the probability distribution for a combination of *a* and *V* at the next step is given by the transition probabilities.

$$P(A_{k+1} = a_{k+1}, V_{k+1} = v_{k+1} | a_k, v_k) = p_{k,k+1} \tag{3}$$

To obtain the transition probability model for the Markov chain, several steps needs to be done: 1) Collect the drive cycle data, *V*; 2) Obtain the acceleration data *a*; 3) Using nearest-neighbor quantization, the sequence of observations (*v, a*) was mapped into a sequence of quantized states ($v_i$, $a_j$).

For each acceleration state, the transition probabilities for velocity are determined from the real world velocity data by counting the occurrence of each transition and visiting times of the state. In (4), $m_{i,j}$ is the number of occurrences of the transition from $v_i$ to $v_j$ for a certain acceleration rate, and

$m_i = \sum_{j=1}^{n} m_{i,j}$ is the total number of times that $v_i$ has occurred at the acceleration rate.

$$p_{i,j} = \frac{m_{i,j}}{m_i} \tag{4}$$

## V. Results of Stochastic Velocity Generation

So far, almost one year driving data of a PHEV has been considered for this study. These 530 cycles are clustered into two groups using K-means clustering approach. Cluster 1 and cluster 2 are quite typical sessions of highway and urban, and the Markov chain model was developed for these two driving patterns. This clustering procedure is necessary as the statistical behavior of vehicles during highway driving and urban driving have been found to be distinctly different. By separating them, it is possible to generate driving cycles which more faithfully represent these two different types of driving.

Based on the observation of the real world velocity data sets, the velocity and acceleration ranges of urban real velocity data are 0 - 124 km/h, -7 - 5.3 m/s$^2$, while for highway cycles they are 0 - 138 km/h, -6 - 5 m/s$^2$. For numerical calculation, velocity and acceleration data are mapped into a quantized set; the resolution of quantization of velocity is chosen as 1 km/h. For acceleration data, the histograms show that the density for absolute values less than 2 is much higher than the density out of this range. Based on the histogram of acceleration for urban driving and highway driving case, the resolution of quantization of acceleration within [-2, 2] m/s$^2$ is chosen as 0.1 m/s$^2$, while out of this range the resolution is chosen as 0.5 m/s$^2$. The probability transition matrix for $a_k=0$, and urban driving pattern case is shown in Fig.9. The transition matrix shows that the velocity states have higher correlation to the states close to them.

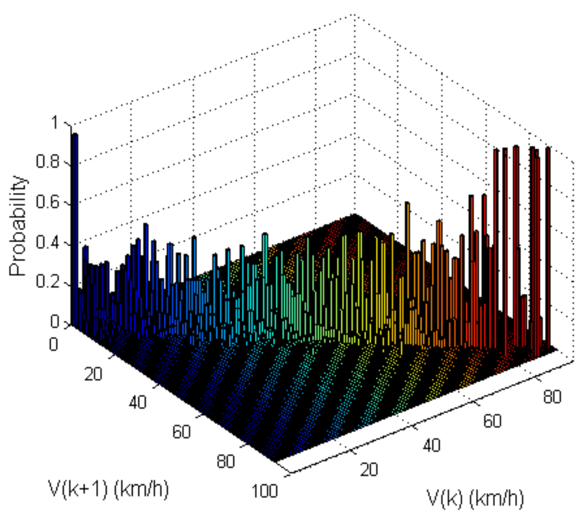


Figure 9. Probability Transition Matrix for a=0 of Urban Case

A case of generated velocity for urban driving is shown in Fig.10; to determine how similar these velocities are to real data, the phase plot of the generated cycle is compared to a typical urban cycle selected from the database (Fig.11). Results show that the generated velocity generally has very similar phase to the selected urban cycle.

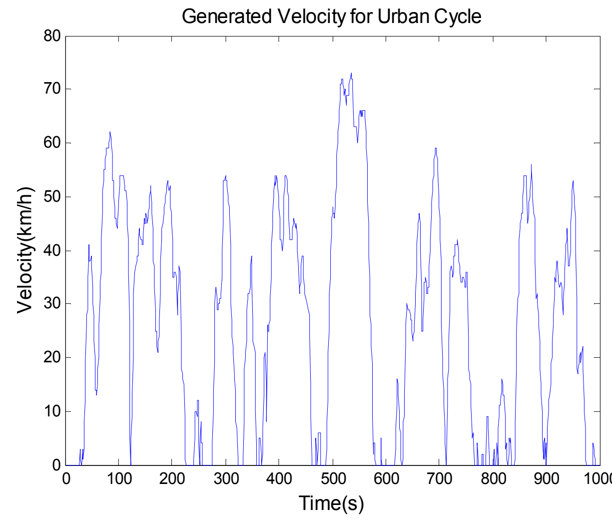


Figure 10. A Generated Velocity Profile for Urban Driving

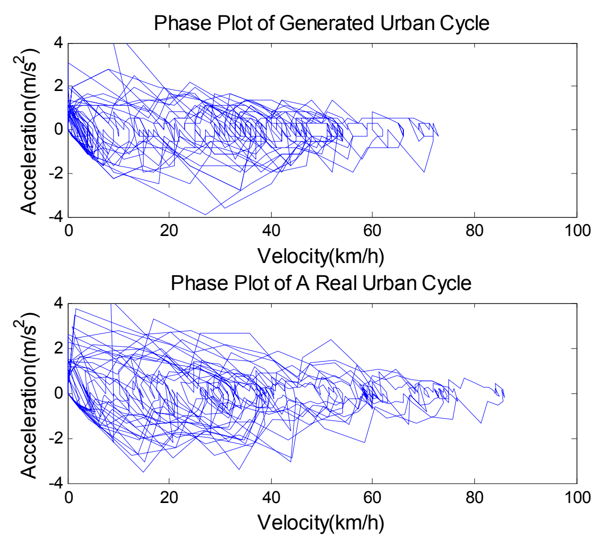


Figure 11. Phase Plot Comparison of Generated Cycle to A Real Urban Cycle

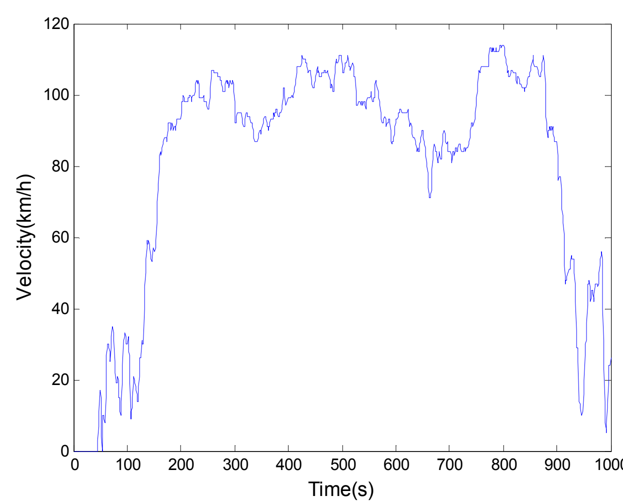


Figure 12. A Generated Velocity Profile for Highway Driving

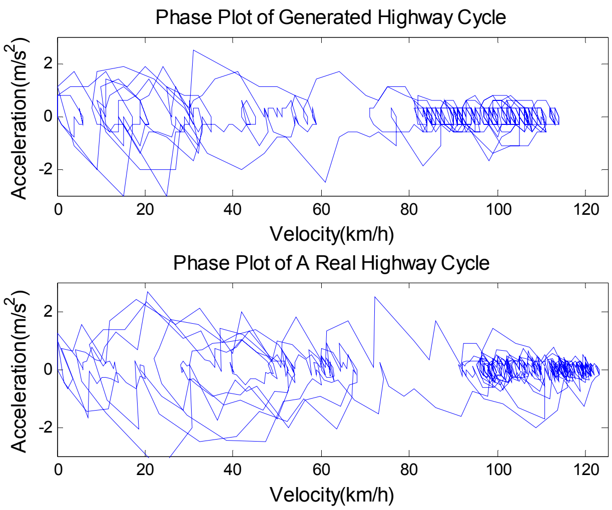


Figure 13. Phase Plot Comparison of Generated Cycle to A Real Highway Cycle

Similarly, a case of generated velocity for urban driving is shown in Fig.12, and the corresponding comparison of phase plot to a selected typical highway cycle is shown in Fig.13. The obtained results show that the generated velocity generally also has very similar phase to the selected highway cycle.

In the future, the developed velocity generation model will be integrated in to the mass simulation model with other stochastic input models.

## VI. CONCLUSIONS

In this paper, some general and basic analysis was done for a PHEV fleet consisting of 6 PHEVs. Systematic analysis using clustering approach was carried for the real world velocity profiles. Velocity profiles were grouped into classes. Driving pattern recognition has been developed based on the clustering results, and Markov-chain model was used for the stochastic velocity generation for different driving pattern. One year's real velocity profiles were collected and grouped into 2 classes based on the clustering approach. The stochastic velocity generation model can generate velocity profiles which represent the specific driving pattern well based on the comparison of the phase plot to the typical real driving cycles.

The work of this paper is a part of a mass simulation PHEV energy analysis model with uncertain inputs of vehicle characteristics, initial conditions and usage patterns. The PHEV energy analysis model will be developed to predict energy consumption in a statistic way and further be used for the PHEV-Grid analysis and smart charging study in the future work.

## ACKNOWLEDGMENT

This work is supported by the SMART@CAR Consortium members at the Ohio State University Center for Automotive Research.

## REFERENCES

[1] Midlam-Mohler, Shawn; Ewing, Sean; Marano, Vincenzo; Guezennec, Yann; Rizzoni, Giorgio, "PHEV fleet data collection and analysis," *5th IEEE Vehicle Power and Propulsion Conference*, VPPC '09, p 1205-1210, 2009

[2] Bata, R., Yacoub, Y., Wang, W., Lyons, D., Gambino, M., Rideout, G. "Heavy duty testing cycles: survey and comparison" *SAE Paper 942 263* (1994)

[3] Carlson, T.R., Austin, R.C. "Development of speed correction cycles". *Sierra Research, Inc.*, Sacramento, CA, Report SR97-04-01 (1997)

[4] Cloke, J., Harris, G., Latham, S., Quimby, A., Smith, E., Baughan, C. "Reducing the environmental impact of driving: a review of training and in-vehicle technologies". *Transport Res. Lab*., Crowthorne, U.K., Report 384 (1999)

[5] Ericsson, E. "Variability in urban driving patterns". *Transportation Res. Part D* 5, 337– 354 (2000)

[6] Ericsson, E. "Independent driving pattern factors and their influence on fuel-use and exhaust emission factors". T*ransportation Res. Part D* 6, 325–341 (2001)

[7] De Vlieger, I., De Keukeleere, D., Kretzschmar, J. "Environmental effects of driving behaviors and congestion related to passenger cars." *Atmosph. Environ.* 34, 4649–4655 (2000)

[8] Jeon S., Jo S., Park Y., Lee J.M., "Multi-Mode Driving Control of a Parallel Hybrid Electric Vehicle Using Driving Pattern Recognition", *Journal of Dynamic Systems, Measurement, and Control, Transactions of the ASME*, Vol.124, pp.141-148, 2002

[9] Lin C.C., Peng H., Jeon S., Lee J.M., "Control of a Hybrid Electric Truck Based on Driving Pattern Recognition", *Proceedings of the Advanced Vehicle Control Conference*, Hiroshima, Japan, 2002

[10] Jeon S., Lee J.M., Park Y.I., "Adaptive multi-mode control strategy for a parallel hybrid electric vehicle based on driving pattern recognition", Proceedings of IMECE, 2003 ASME international Mechanical Engineering Congress, Washington, D.C., November 15-21, 2003

[11] Bo, Gu; Rizzoni G.., "An adaptive algorithm for hybrid electric vehicle energy management based on driving pattern recognition", *Proceedings of 2006 ASME International Mechanical Engineering Congress and Exposition, Dynamic Systems and Control Division*, 2005.

[12] Langari, R., Won, J.-S. "Intelligent energy management agent for a parallel hybrid vehicle-part I: system architecture and design of the driving situation identification process". *IEEE Transactions on Vehicular Technology* 54(3), 925–934 (2005)

[13] Won, J.-S., Langari, R. "Intelligent energy management agent for a parallel hybrid vehicle-part II: torque distribution, charge sustenance strategies, and performance results". *IEEE Transactions on Vehicular Technology* 54(3), 935–953 (2005)

[14] Murphey, Yi L.; Chen, ZhiHang; Kiliaris, Leo; Park, Jungme; Kuang, Ming; Masrur, Abul; Phillips, Anthony, "Neural learning of driving environment prediction for vehicle power management," *Proceedings of the International Joint Conference on Neural Networks*, p 3755-3761, 2008

[15] Gong, Q., Li, Y. and Peng, Z.-R., "Trip Based Optimal Power Management of Plug-in Hybrid Electric Vehicle with Advanced Traffic Modeling," *SAE Transaction,* 2008

[16] Dembski N., Rizzoni G.., Soliman A., "Development of Refuse Vehicle Driving and Duty Cycles", *SAE Int'l Congress and Exposition* Paper No. 2002-01-1165, Detroit, MI, 2005.

[17] Dembski N., Guezennec Y., Soliman A., "Analysis and Experimental Refinement of Real-World Driving Cycles", *SAE Int'l Congress and Exposition Paper* No. 2002-01-0069, Detroit, MI, March 2002.

[18] Matlab help notes, "K-Means Clustering".

[19] Lin, Chan-Chiao (Dept. of Mechanical Engineering, University of Michigan, MI 48109-2125); Peng, Huei; Grizzle, J.W. "A Stochastic Control Strategy for Hybrid Electric Vehicles," *Proceedings of the American Control Conference*, v 5, p 4710-4715, 2004

[20] S. J. Moura, H. K. Fathy, D. S. Callaway, and J. L. Stein, "A stochastic optimal control approach for power management in plug-in hybrid electric vehicles," *Proceedings of the 2008 ASME Dynamic Systems and Control Conference*, 2008