2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Distance Until Charge Prediction and Fuel Economy Impact for Plug-in Hybrid Vehicles

Payam Naghshtabrizi, Johannes Kristinsson, Hai Yu, and Ryan McGee

Abstract— **For a Plug-in Hybrid-Electric Vehicle (PHEV) a close to linear battery depletion profile is near optimal in terms of fuel economy, given that the driven Distance Until Charge (DUC) is known. However the intended driven distance until charge is not always known. We propose a pattern recognition method to predict the DUC based on the key-on time and day observations and the on-board stored data in vehicle. We review the energy management strategy and we show a fuel economy improvement despite prediction errors.**

I. INTRODUCTION

A Plug-in Hybrid-Electric Vehicle (PHEV) is a mix between a regular Hybrid Electric Vehicle (HEV) and a Battery Electric Vehicle (BEV). Contrary to the HEV which aims at keeping a nominal fixed State of Charge (SOC) for its battery, the PHEV has a bigger battery with extra capacity to store energy that will be consumed during driving and then recharged from the electric grid when the vehicle reaches a charging station (e.g. overnight). The base strategy of a PHEV has two modes of operations shown in Fig. 1: Charge Depletion (CD) mode and Charge Sustaining (CS) mode. During the CD operating mode, the vehicle propulsion is relying primarily on the electric drivetrain using the energy stored in the battery. In this mode, the PHEV runs like a BEV. When the battery SOC reaches a pre-defined level (e.g., 30%), the CS mode starts. In this mode, the PHEV runs like a HEV, both the battery and the engine provide the required propulsion energy while the SOC is kept constant. The solid red line in Fig. 1 shows the base energy management strategy of a PHEV that depicts SOC versus the driven distance and the two operating modes. In the figure, the vehicle switches from CD to CS mode at around 30% SOC level and keeps the battery energy within a few percentage points of this level.

However, this strategy is not optimal in terms of fuel consumption and if the distance that will be traveled until the next charge is known in advance, the battery depletion profile (rate of usage) can be adapted in order to improve Fuel Economy (FE). It was shown that a close to linear battery depletion profile for a PHEV such as the dotted blue line in Fig. 1, is near optimal in terms of FE, given that the driven distance is known [1], [2]. In this case, the engine works in a more efficient region and a better fuel economy can be achieved. Gong et al. [3] found a global optimal SOC profile for a given speed profile by solving an optimization based on a dynamic programming. However, it is hard to implement such an algorithm on a real world drive cycle since the future speed profile is unknown. Gong et al. [4] used the GPS information and a neural network to model the trip, based on the past driving history. They used a model to construct a near optimal solution through dynamic programming; however, this method requires significant computation. Tulpule et al. [5] proposed a simple method where the SOC is depleted linearly with the traveled distance given that the Distance Until (the next) Charge (DUC) is known. They called this strategy the blended mode for which the depletion rate is slower. The authors showed that the blended strategy has better net energy consumption compared with the CD-CS strategy. Karbowski et al. [6] defined the total energy loss of the engine and battery to obtain a global optimal strategy for different speed profiles and driving cycles and then they proposed a real-time rule-based controller. They also observed that the DUC is an important factor to determine the SOC depletion strategy such that the energy losses are minimized and the engine is forced to operate in the most efficient operating points.

While it is impossible to know the intended DUC beforehand, a combination of statistical estimation based on pattern recognition and optional driver input will allow the vehicle to estimate the DUC. Although there are several papers for route prediction and travel time prediction [7], [8], [9], there is no publication regarding prediction of DUC to our best knowledge. In this paper, we propose a method to estimate the DUC based on multiple estimators and historical data consisting of Time Of Day (TOD) and Day Of Week (DOW) of the first key-on after each charge. Each one of the estimators use a different subset of historical data, called filtered training sets, to provide a DUC prediction. All estimators use the Recursive Least Square (RLS) estimation method [10] which is easily implementable using existing computational power in a vehicle. The difference between the estimators is that they use a different subset of the historical data to predict the DUC value. Our initial study using Weka [11], showed that there is only a minor difference between the RLS method results and other methods such as support vector machine, k-nearest neighbor, neural network, and rule based pattern recognition methods for this particular problem. Therefore, we choose the RLS for its simplicity and ease of implementation.

We propose a method to predict the final DUC by using the predictions from the estimators. This final prediction will be used by the energy management of the PHEV to obtain a depletion profile such as the dotted blue line in Fig. 1. We associate a metric called Quality of Prediction (QoP) to each of

The authors are with the Ford Motor Company, Vehicle and Battery Controls Department. The corresponding author can be reached at pnaghsht@ford.com.

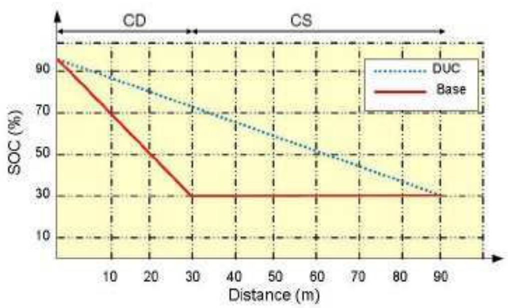


Fig. 1. The solid red graph shows the SOC for the base energy management strategy with CD and CS modes. The dotted blue graph shows the SOC for the proposed method called DUC.

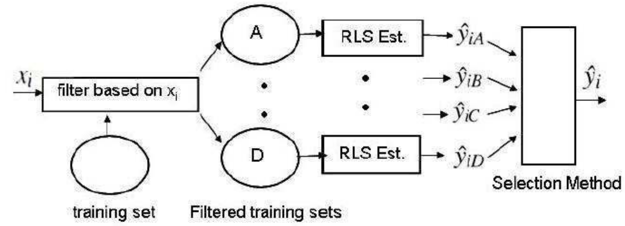


Fig. 2. This figure shows the structure of the overall estimator that predicts DUC based on DOW and TOD observation and the training set.

the predictions provided by the estimators. The final DUC is the average of predictions with a QoP higher than a computed level called normal QoP. Under some conditions, the overall estimator would not give any estimation due to a low QoP of most of the estimators. We will see in Section IV that the proposed method is quite effective in estimating DUC based on TOD and DOW observations. For route prediction, from the trip observation and destination intent, Simmons et al. [8] used hidden Markov chain that allowed them to incorporate TOD, DOW, and speed into their estimator. Their results showed that only speed is a significant help in boosting their prediction accuracy. However, our result for distance until charge prediction (that may consist of several trips) shows that DOW and TOD in conjunction with multiple predictors can predict DUC accurate enough for our purpose.

This paper is organized as follows. In Section II, we briefly present the overall system. In Section III, we review the data collection plan used for this project and in Section IV we focus on the proposed prediction method. In section V we briefly review the energy management strategy and we investigate the fuel economy impact of DUC strategy on a PHEV by comparing the base case with the ideal case that the driver enters DUC manually through HMI and the case that DUC is predicted and necessarily a prediction error is present.

II. System Overview

The Prediction System is part of the whole DUC system and it is responsible for providing an estimation or prediction of an intended DUC. One of the sub-systems is The Energy Management Controls, that controls the engine, the electrical motor, the battery, and other powertrain components so that they work seamlessly together to propel the vehicle. These control systems decide when to turn the electrical motor on or off depending on the present power demand and the battery SOC. One of the control systems is the Trip-based Energy Management control (TEMC). This system is responsible for controlling the actual battery depletion to obtain a target depletion profile, e.g., dotted line in Fig. 1 which is based on the DUC value.

III. Data Collection

We used data loggers from Davis called CarChip (http://www.carchip.com/) to gather data. The device powers itself from the OBD II port and starts logging the vehicle speed as soon as the vehicle is used. Fig. 3 shows the scatter plot of the DUC for 4 users where the value of recorded DUCs are bined. The range between 0 to 100 miles is divided into 10 bins and any DUC value more than 100 miles is also represented by the last bin. In the figures, day 1 to 7 represent Monday to Sunday respectively but for better visualization, a random number drawn from a uniform distribution in the interval $[-0.25, 0.25]$ is added to the number assigned to days. For example any DOW between 0.75 and 1.25 represents Monday in Fig. 3. Note that discretizing DUC is just for the visualization and not for the prediction since our initial studies with Weka showed that the discretization error lead to bad DUC predictions even if we increase the number of bins (for this reason we did not use a discrete prediction method such as support vector machine).

IV. DUC Estimation

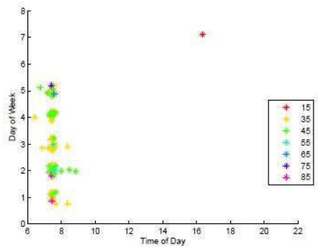
We divide the collected data for each driver into two sets: a training set and a test set. For each driver, the test set contains 16 samples randomly selected from the collected data and the rest form the training set. The training set is used for prediction and the test set is used to determine how good the prediction is. Alternatively, we could use other methods such as folding to provide test and training sets.

Fig. 2 shows the structure of the overall estimator. The objective is to estimate y_i, $i \in \{1, \cdots, 16\}$, the i-th DUC in the test set for each driver, where the corresponding estimate is denoted by $\hat{y}_i$. The estimate $\hat{y}_i$ is based on the observation x_i, $i \in \{1, \cdots, 16\}$, its DOW and TOD, and the training set data (x_j, y_j), $j \in \{1, \cdots, n\}$ where n represents the number of samples in the training set for each driver. This objective mimics the real world scenario where past TOD, DOW, DUC are saved in an on-board memory and the DUC is predicted at the first key-on after each charge by observing TOD and DOW.

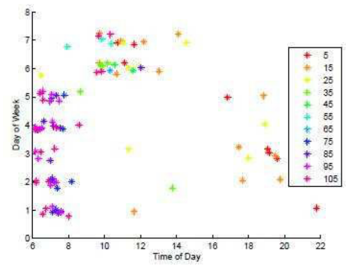
A. Unfiltered and Filtered training sets

We use multiple data sets based of the observation x_i to provide a prediction:

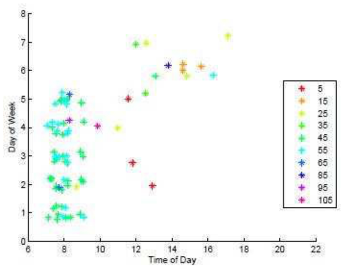
- Unfiltered training set includes all the samples in the training set.



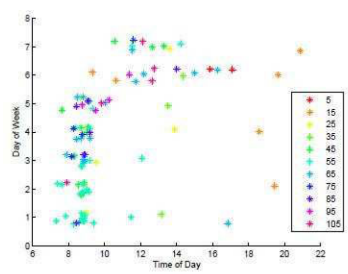
(a) User 1



(b) User 2



(c) User 3



(d) User 4

Fig. 3. Figures show the scatter plot of the recorded DUCs for user 1 to 4 where the value of DUCs are bined. The range between 0 to 100 miles is divided into 10 bins and any DUC value more than 100 miles is also represented by the last bin, and the bins are presented by their mid-points.

- Filtered training set A includes the samples in the training set whose DOW type, weekend or weekday, matches with x_i's DOW type.
- Filtered training set B includes the samples in the training set whose DOW, Monday to Sunday, matches with x_i's DOW.
- Filtered training set C includes the samples in the training set whose TOD are within 1.5 hours of x_i's TOD.
- Filtered training set D includes the samples that are in both sets B and C.

TABLE I

THIS TABLE SHOWS THE RMSE OF ESTIMATED DUC (IN MILES) BASED ON THE UNFILTERED AND FILTERED TRAINING SETS. THE FINAL SHOWS THE RMSE FOR THE FINAL DUC ESTIMATION BASED ON (1) AND THE LAST COLUMN SHOWS THE NUMBER OF TEST SET SAMPLES FOR WHICH THE FINAL PREDICTION IS PROVIDED.

User	Unf.	A	B	C	D	Final	No.
U1	4.2	4.1	4.6	4.1	4.6	4.0	16
U2	23.9	20.5	22.0	18.7	23.4	15.0	10
U3	21.9	20.5	16.0	22.1	26.6	6.1	11
U4	14.3	14.2	19.2	19.3	17.9	8.2	11

For example for x_i=(Monday,10), set A includes all historical data that their DOW is Monday through Friday, set B includes all samples whose DOW is Monday, its set C includes all samples whose TOD is between 8:30 and 11:30, and set D includes all samples that their DOW is Monday and their TOD is between 8:30 and 11:30.

After finding filtered training sets A to D for an observation x_i, we use the RLS method to provide a prediction for DUC showed by $\hat{y}_{ik}$, that represents the estimation for the i-th sample in the test set based on the k-th filtered training set where $k \in \{A, B, C, D\}$. More details of RLS method can be found in [10]. It is important to note that TOD is a numerical variable but DOW is a symbol and 1 to 7 represent the assigned values to symbols. Usually the RLS is used just for numerical variables and for the mixed case, RLS is used for each symbol. But in our case we choose to treat the numerical values representing symbols as numbers because this approach boosts out prediction accuracy. The reason is that there is a strong relationship between the outcome of symbols, e.g., DUC value for Saturday and Sunday.

B. Selection Method

By using the filtered training sets described in Section IV-A and the RLS method, five different predictions for a test sample are computed. Now the question is, how to choose the final estimation of DUC from multiple predictions provided by filters?

The following observations are crucial for setting the selection rules:

- Variance of a filtered training set used for prediction is an important measure for QoP. If a filtered training set has a high variance it means that there is a high variability and randomness in the data used for prediction and one should not expect a good prediction in this case. It is possible that one or two filters have low QoP and others have high QoP. For example, consider the following case: for a given driver, if TOD is 6:30 to 8:30 she drives close to 60 miles because her job is 30 miles away from her house. However, if TOD is 10:30 she drives 20 miles because she does not work that day. Also we assume that she has a flexible work schedule and sometimes she works on Wednesdays and sometimes she does not. With these assumptions, given the input observations TOD=7:00 and DOW=Wednesday, filtered training set B can have high variance and one should use the prediction provided by other filters.

- The number of samples in a filtered training set is an important factor as well, because if there is only one sample in a filtered set, the variance is zero but the QoP is factitiously high. Consequently we consider a prediction corresponding to less that 3 samples in the filtered training set as low QoP even if their variance is low, because the prediction is not trustworthy.
- We can choose a fixed value as a threshold for acceptable QoP so that if a filtered set variance is above that threshold, the corresponding estimation is discarded. Alternatively, we can choose the variance of unfiltered training set as the normal variance of the data and discard a filter estimation if the variance of the corresponding filtered training set is above this level. We choose the latter case since some drivers have a more random driving pattern. For example, for the driver 1 the normal variance is 5 and for the driver 2 the normal variance is 25. For this case we do not have to find a fixed level for the threshold and it is determined by calculation based on the drivers driving history.

Based on the above observations, the selection rules and DUC estimation are as follows:

1) For a given input observation x_i, $i \in \{1, \cdots, 16\}$ form the filtered training sets A, B, C, D and find predictions by the RLS method.
2) For all $k \in \{A, B, C, D\}$, compute

$$\eta_k = \begin{cases} 0 & \eta_k \leq 3, \\ 1 & \text{otherwise}, \end{cases}$$

where n_k is the number of samples in the filtered set k and η_k is a corresponding flag of the filter k. When the flag η_k is zero, the corresponding prediction is discarded and will not be used to compute the final DUC in (1).

3) For all $k \in \{A, B, C, D\}$, compute

$$h_k = \begin{cases} 0 & \sigma_k \geq \sigma_u \\ 1 & \text{otherwise}, \end{cases}$$

where σ_k, σ_u represent standard deviation of filtered training set k and the unfiltered one. When the flag h_k is zero, the corresponding prediction is discarded and will not be used to compute the final DUC in (1).

4) Compute the DUC estimation for the input observation x_i according to

$$\hat{y}_i = \begin{cases} \frac{\sum_k \eta_k h_k \hat{y}_{ik}}{\sum_k \eta_k h_k} & \sum_k \eta_k h_k > 2, \\ \text{no estimation} & \text{otherwise}, \end{cases} \qquad (1)$$

The final DUC is the average of predictions which both of the corresponding flags η_k, h_k are 1. If the condition $\sum_k \eta_k h_k > 2$ does not hold, the overall estimator would not give any estimation due to a low QoP of most of the estimators.

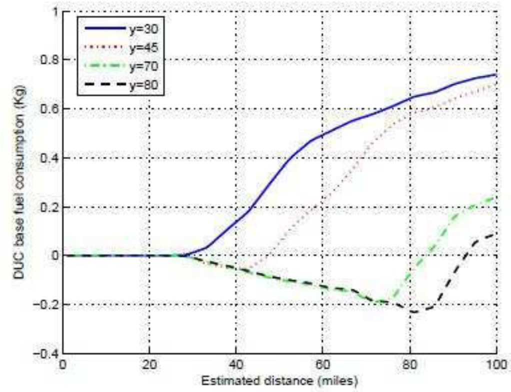


Fig. 4. Cost vs. the estimated DUC ($\hat{y}$) for $30, 45, 70, 80$ miles for intended DUC (y).

Table I shows the Root Mean Square Error (RMSE) of the estimated DUC (in miles) for the samples in the unfiltered and filtered test sets. The last column shows the number of test set samples for which the final prediction is provided. Note that the total number of samples for other cases is 16 (except for the filter D which are $16, 14, 15, 15$) which is the number of samples in the test sets for users. The RMSE has improved significantly for two reasons. First, the proposed algorithm for final prediction is successful in refusing to provide a prediction that otherwise would be a poor prediction of the DUC. The second reason for the improvement is due to the fact that the final method provides the estimate for fewer samples, which may reduce the RMSE. Since there is no penalty for not providing an estimate, the comparison between final method and other methods becomes somewhat difficult. However, the FE results presented in the next section clearly and unambiguously show the superiority of the final method over the other ones (see Table II) since no prediction can affect FE negatively.

V. BENEFIT OF DUC ESTIMATION BASED ON FE

A closed-loop control strategy for PHEV energy management was developed at the Ford Motor Company to obtain a SOC profile such as the dotted blue line in Fig. 1. Then the strategy was simulated and implemented on a real PHEV to compare the FE with the base strategy and a significant FE improvement was achieved. The underlying assumption was that the intended DUC is known; However, in our work the DUC is not provided and we predict the DUC that will be used by energy management. The goal of this section is to show the impact of the prediction on FE using the simulation.

We use the simulation model to find the FE for several driving cycles for a PHEV vehicle for two cases: an exact intended DUC (entered by the driver) and the estimated DUC based on the algorithm discussed in Section IV. The results for US06 driving cycle which is the focus of this paper, are saved in a look-up table to form the function duckFuelCost$(y, \hat{y})$ that represents the fuel consumption of the PHEV when the vehicle is driven y miles while the estimated DUC used by TEMC is set to $\hat{y}$ miles.

TABLE II

THIS TABLE SHOWS THE RMSE OF ESTIMATED DUC (IN MILES) BASED ON THE UNFILTERED AND FILTERED TRAINING SETS. THE FINAL SHOWS THE RMSE FOR THE FINAL DUC ESTIMATION BASED ON (1) AND THE LAST COLUMN SHOWS THE NUMBER OF TEST SET SAMPLES FOR WHICH THE FINAL PREDICTION IS PROVIDED.

User	All	*A*	*B*	*C*	*D*	Final 1
U1	-0.15	-0.14	-0.27	-0.14	-0.27	-0.16
U2	0.39	0.74	0.43	1.00	0.76	1.22
U3	0.11	0.07	-0.81	0.12	-1.75	0.22
U4	0.45	0.43	-0.07	0.40	0.13	0.93

TABLE III

FE IMPACT OF FINAL METHOD 1 AND 2 OF PREDICTION AND THE BEST CASE FOR WHICH THE EXACT INTENDED DUC IS KNOWN IN TERMS OF KILOGRAMS OF GASOLINE FOR 16 SAMPLES IN THE TEST SET FOR USER 1 TO 4. THE % COLUMNS SHOWS THE PERCENTAGE OF FUEL SAVED COMPARED WITH THE BASE STRATEGY.

User	Final 1	%	Final 2	%	best	%
U1	-0.16	-0.38	0	0	0.06	0.15
U2	1.22	1.59	1.21	1.59	2.14	2.79
U3	0.22	0.35	0.19	0.30	1.26	1.97
U4	0.55	0.72	0.58	0.80	1.28	1.82
Ave.	0.55	0.72	0.58	0.80	1.28	1.82

We define *cost* as the difference between the fuel consumption of the base strategy (the red curve in Fig. 1) and the DUC strategy (a depletion curve such as the blue curve in Fig. 1). So a negative cost represents fuel saving achieved by the DUC strategy and the positive cost means that more fuel is used by the DUC strategy compared with the fuel consumption of the base strategy. The latter case represents the risk of using the DUC strategy in terms of fuel consumption. Fig. 4 shows the cost vs. the estimated DUC ($\hat{y}$) for $30,45,70,80$ miles for intended DUC (y). Below we discuss the important factors affecting the benefit or risk of using DUC strategy with estimation in terms of fuel consumption concluded from Fig. 4

- The base case changes mode from CD to CS around 30 miles. If both y and $\hat{y}$ are shorter than 30 miles, the base and DUC strategy are the same and the cost is zero.
- When no final DUC estimation is provided, it is the same as the case that $\hat{y}$ = 0 and the base strategy will be followed and the cost is zero.
- Perfect DUC estimation is the case that $y = \hat{y}$ where the estimated DUC exactly matches with the intended DUC. The blue line in Fig. 1 shows the strategy for this case when $y = \hat{y} = 90$ miles. For this case, fuel consumption is always better than the base case if $y > 30$ miles (and otherwise the consumptions are equal). The minimum of curves in Fig. 4 for $y = 45,70,80$ represent this case that occur at $\hat{y} = 45,70,80$ miles.
- For a given driven distance for example $y = 70$, any value of DUC estimation in the interval $30 < \hat{y} < 83$ provides some fuel saving benefit compared with the base strategy. This interval decreases as the real DUC (y) decreases, so there is less room for estimation error to obtain some fuel saving benefit. Since there is always some estimation errors, if the estimated DUC ($\hat{y}$) is smaller than for example 40 miles, it is better not to use the DUC estimate and follow the base strategy.
- There is a high risk in terms of fuel consumption in over-estimation, as indicated by a sharp increase of cost functions for $\hat{y}$ larger than y. This means that in case of over-estimation, the fuel consumption can be much worse than the base strategy. However, there is more room to tolerate under-estimation.

We use the cost function to find whether using DUC prediction improves the fuel consumption compared with the base strategy. Table II summarizes the FE impact for 16 test samples of 4 users if the DUC estimation based on the unfiltered and filtered training sets and the final estimation based on (1) are used by TEMC. The FE impact represents the difference of gasoline consumption between the base strategy and the DUC strategy in terms of kilograms of fuel. Consequently more positive numbers are more desirable since they represent fuel saving achieved by using the DUC strategy. Note that when no final DUC estimation is provided, then $\hat{y} = 0$ and the FE impact is zero.

It is important to note when the real DUC (the perfect estimation) is less than the point that the CD mode changes to the CS mode, roughly 30 to 37 miles, the FE impact is very close to zero and the difference can be considered as numerical noise produced by simulation. So we propose final method 2 that is the same as the final method 1 with the difference that we use the base strategy instead of DUC when the predicted DUC is smaller that 37 to reduce the effect of estimation error and noise. Table III shows the FE impact of method 1 and 2 and the percentage of fuel saving compared with the base strategy. To investigate the effect of prediction error on the FE, we also present FE impact of the base case for which we assume the exact intended DUC is known and used by TEMC ($y_i = \hat{y}_i$). The results clearly show the benefit of DUC strategy with the prefect and the predicted DUC for up to 1.82% and 0.8% FE improvement on average for 4 drivers (note that without user 1, the average improvements are 2.38% and 1.07%).

Remark 1: It is important to notice that *good* DUC estimation does not translate to *good* fuel saving of DUC strategy compared with the base case. From the Table I, it is clear that the DUC estimations for user 1 have much better QoP than the results for user 2. However, Table II or Table III show that fuel saving for user 2 is much higher than user 1. As mentioned above the reason is that real DUC values for user 1 are shorter than 37 and consequently the base strategy and the DUC strategy are both in the CD mode.

VI. CONCLUSION AND FUTURE WORK

Our results confirmed that if the intended DUC is exactly known, this knowledge can be used by energy management algorithm of a PHEV to improve fuel economy. However the intended DUC is not always known. We proposed a method to predict the DUC based on the TOD, DOW key-on observations and the stored data for 4 different users. We briefly reviewed the energy management strategy and we showed the fuel economy improvement for the users when the predictive DUC strategy is used.

There are a few ideas to continue research in this area. Instead of predicting the DUC by minimizing a training set summation of square errors of the DUC and then checking the FE improvement, we can minimize a cost function based on the FE directly. In future we study the case that DUC estimation can be updated on-board based on the online GPS information and the trip chain observations, charging locations, and speed traces. By updating the DUC estimate on-line it is possible to improve the DUC prediction and the FE significantly, especially for under-estimated DUC values for Users 3 and 4, since after driving more than the initial estimated DUC it becomes obvious that under-estimation has happened and the initial estimation needs to be revised.

REFERENCES

[1] R. C. DeVault, "Just-in-tim battery charge depletion control for PHEVs and E-REVs for maximum battery life," in *2009 SAE World Congress, No. 2009-01-1384*, April 2009.

[2] P. Tulpule, S. Stockar, and G. Rizzoni, "Optimality assessment of equivalent consumption minimization strategy for PHEV applications," in *ASME 2009 Dynamic Systems and Contr. Conf. , No. DSCC2009-2748*, April 2009, pp. 265 – 272.

[3] Q. Gong, Y. Li, and Z. Peng, "Trip based optimal power management of plug-in hybrid electric vehicle with advanced traffic modeling," in *2008 SAE World Congress, No. 2008-01-1316*, 2008.

[4] ——, "Power management of plug-in hybrid electric vehicles using neural network based trip modeling," in *Proc. of the 2009 Amer. Contr. Conf.*, 2009, pp. 4601 – 4606.

[5] P. Tulpule, V. Marano, and G. Rizzoni, "Effects of different PHEV control strategies on vehicle performance," in *Proc. of the 2009 Amer. Contr. Conf.*, 2009, pp. 3950 – 3955.

[6] D. Karbowski, A. Rousseau, S. Pagerit, and P. Sharer, "Plug-in vehicle control strategy: From global optimization to real time application," in *22th Int. Electric Vehicle Symp.*, Oct. 2006.

[7] J. Froehlich and J. Krumm, "Route prediction from trip observations," in *2008 SAE World Congress, No. 2008-01-0201*, April 2008.

[8] R. Simmons, B. Browning, Y. Zhang, and V. Sadekar, "Learning to predict driver route and destination intent," in *IEEE Intell. Transportation Syst. Conf.*, 2006, pp. 127 – 132.

[9] D. Patterson, L. Liao, D. Fox, and H. Kautz, "Inferring high-level behavior from low-level sensors," in *UbiComp 2003: Ubiquitous Computing, Springer: Seattle, Washington USA*, 2003, pp. 73 – 89.

[10] S. Haykin, *Adaptive Filtering Theory*, 4th ed. Prentice Hall, 2002.

[11] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Elsevier, 2005.