Analysis of Driving Behavior's Impact on Battery Discharge Rate for Electric Vehicles

Yew Ming Yeap
Department of Infrastructure, Energy and Environment
Institute for Infocomm Research
Singapore
yeap_yew_ming@i2r.a-star.edu.sg

Duong Tran
Department of Infrastructure, Energy and Environment
Institute for Infocomm Research
Singapore
tran-d@i2r.a-star.edu.sg

***Abstract*—A robust battery management system (BMS) is important for electric vehicle (EV) to preserve the battery health. As EV is gaining popularity as a mode of public transport, the fleet operator should be concerned of optimizing the usage of a fully charged battery, this boils down to the driving behavior. This paper investigates the relation of driving parameters to the battery discharge rate. The driving behavior is associated with vehicle speed, motor speed, throttle position and brake pressure. The data is collected from experiment where the EV is operated through urban roads in Singapore. Principal Component Analysis (PCA) is used to reduce dimensionality of data. We identified an interesting feature in the data distribution, which can be quantified by centroid. Using this feature, we conducted the regression analysis and found out that the centroid has a linear relationship with the duration of battery retaining its state of charge (SOC), resulting in the Pearson correlation coefficient of 0.92.**

***Index Terms*—Driving behavior, driving parameters, data analysis, battery management system, electric vehicles**

I. Introduction

Transportation is a bedrock of a country's economy; it provides accessibility of people to goods, jobs, services and social activities. However, transportation reportedly accounts for 24% of the global carbon dioxide emission, of which road transport makes up about 74% [1] due to dependence on fossil fuel. Driven by push for more decarbonized economy, some countries are aggressively formulating policies (carbon tax, subsidy) and building infrastructures (charging station) to incentive uptake of battery operated electric vehicles (EV). To meet the low emission goal, transport operators, which have taxis and buses on the road for long hours, are replacing their current fleet with EV. It is always in the interest of fleet management to ensure that the fully charged battery gives the best mileage possible. One of the factors that determine the discharge rate is the EV driving pattern.

The analysis of driving style requires a robust data acquisition platform to provide useful data. Controller Area Network (CAN) is a communication system in vehicle that allows sensors and electronic control units (ECU) exchanging data with each other. The parameters of interest are, for example, speed, steering wheel, throttle position, brake pressure, battery state-of-charge (SOC), battery output voltage and current, etc. The survey paper [2] summarized the driving style recognition methods proposed in literature and the parameters that have been used in their study.

Machine learning is one of the most popular approaches for driving style recognition. It can handle large amount of data and adaptively improve the classification when given more data. There are two types of machine learnings: unsupervised and supervised. For unsupervised learning, *k*-means clustering has been used to determine the clusters of driving style/behavior [3]. Constantinescu *et al.* [4] analyzed the driving data using Hierarchical Cluster Analysis (HCA) first, then Principal Component Analysis (PCA). The driving style is characterized into six groups of "aggressiveness". Campo *et al.* [5] modeled the driving style classifier using extreme learning machine (ELM). The clusters established by the HCA determined the number of outputs for the ELM. For supervised learning, support vector machine (SVM) was proven to have achieved reasonable accuracy [6]. The implementation of neural network (NN) was explored by Xu *et al.* successfully classifying driver's behavior based on throttle position, brake pressure and speed [7]. Semisupervised learning has been applied for the same purpose [8–10]. As the name suggests, it combines the advantages of unsupervised learning, which works well on unlabeled data, and supervised learning, which tends to generate clean classification.

The knowledge of driving style and external condition is important to understand its relationship with energy consumption. Braun and Rid [11] presented that EV has the best performance in urban area compared to internal combustion vehicle. When driving on country roads, the vehicles tend to be on steady-speed motion, in this case combustion engine has better efficiency. It was demonstrated that aggressive driving (repeated acceleration/deceleration) drains the energy from battery faster than gentle driving [12, 13]. Moreover, the energy consumption is independent of the initial SOC of battery. Climate is a determinative factor that affecting battery discharge rate in two conflicting ways. On one hand, the battery efficiency in delivering and receiving current is marginally higher in high ambient temperature [14]; on the other hand, hot weather compels drivers to turn up the use of air-conditioning resulting in high energy demand [15]. Most studies in literature thus far only focus on the effect of single driving parameter on the energy consumption. As battery is

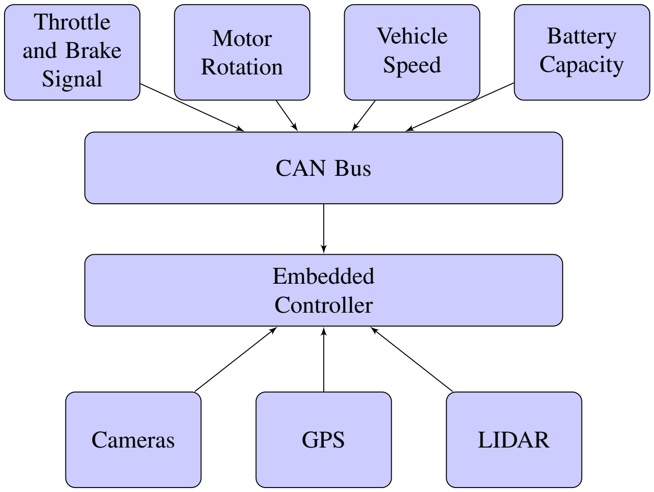

Fig. 1. Data acquisition in experimental EV.

an integral part of EV, there are studies on how to prolong its mileage. Martínez *et al.* [16] modeled the EV using fuzzy logic and used the model to optimize the energy demand. An average saving of 3-4% was reported. Jeong *et al.* [8] proposed mixed usage of battery and supercapacitor so that charge and discharge cycle of battery can be reduced, hence longer shelf life.

In this paper, we are looking into the relationship between the battery capacity and different driving parameters. The real data is sourced from the experimental EV driving in the urban road of Singapore. The CAN bus captures several important variables such as battery SOC, vehicle speed, motor speed, throttle position, brake pressure, etc. It will be analyzed how these four parameters in combination affects the discharge rate of battery. We perform PCA to reduce the variables to just two components, allowing us to find the unique feature that represents the driving behavior. Then, we use regression analysis to estimate the time it takes for the SOC to drop by 1% based on the given feature.

The organization of this paper is structured as follows. Section II explains how the driving data was collected from the experiment and the driving parameters that are of interest for analysis. Section III presents our approach of analyzing the driving data and interesting insights on the relationship between driving parameters and battery discharge rate. The discussion on the results of the analysis and the ways to improve the experiment are given in Section IV. Lastly, the paper is concluded in Section V.

II. Driving Data

A. Experiments and Data Collection

The driving data was collected from the sensor system installed in the EV. The detection of obstacles in 360-degree vision was done by the LIDAR sensor, mounted on the roof. It is important to monitor traffic condition, weather and pedestrian crossing the road, therefore we installed three in-car cameras to capture the video when the EV was operated.

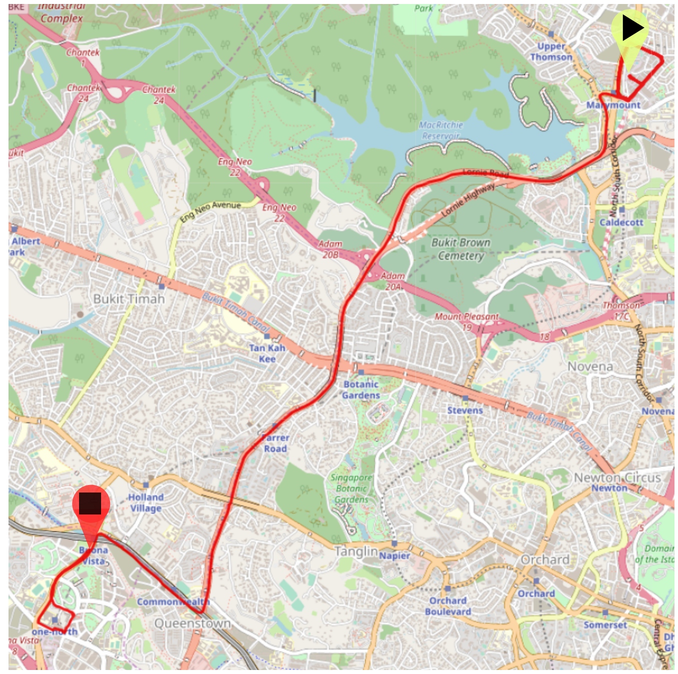

Fig. 2. One of the driving routes in Singapore.

We also had GPS to track and review the route, which can help us understand the road type (major road, expressway) and condition, and how it affects the speed profile. LIDAR data, GPS locations, video data and CAN bus data were fed to an in-vehicle embedded controller to be organized in pre-defined structure and timestamped. The flow in the data acquisition is summarized in Fig. 1.

The experimental EV was deployed in the open driving environment. There was no fixed driving route. The GPS sensor was installed in the EV to record its movement. The results reported in this paper were collected from the experiments over the span of 11 days. One of the driving routes is shown in Fig. 2, with the path marked in red, and starting and ending point indicated. In this route, the EV was driven through semi-expressway and major urban road.

B. Driving Parameters

When collected from the CAN bus, the raw data, still in the hexadecimal format, has to be decoded into readable value. About 7 signals can be extracted from the raw data, but the following parameters are of the interest for data analysis:

1) Battery SOC (0-100%)
2) Vehicle speed (km/h)
3) Throttle position (0-100%)
4) Brake pressure (0-100%)
5) Motor speed (RPM)

The combination of 2-5 parameters is sufficient to capture the driving behavior. It is obvious that the vehicle speed is directly related to the driver's behaviour. For example, aggressive drivers tend to drive fast in urban road despite strict speed regulation. Although the information of acceleration/deceleration is not available, they can be indirectly

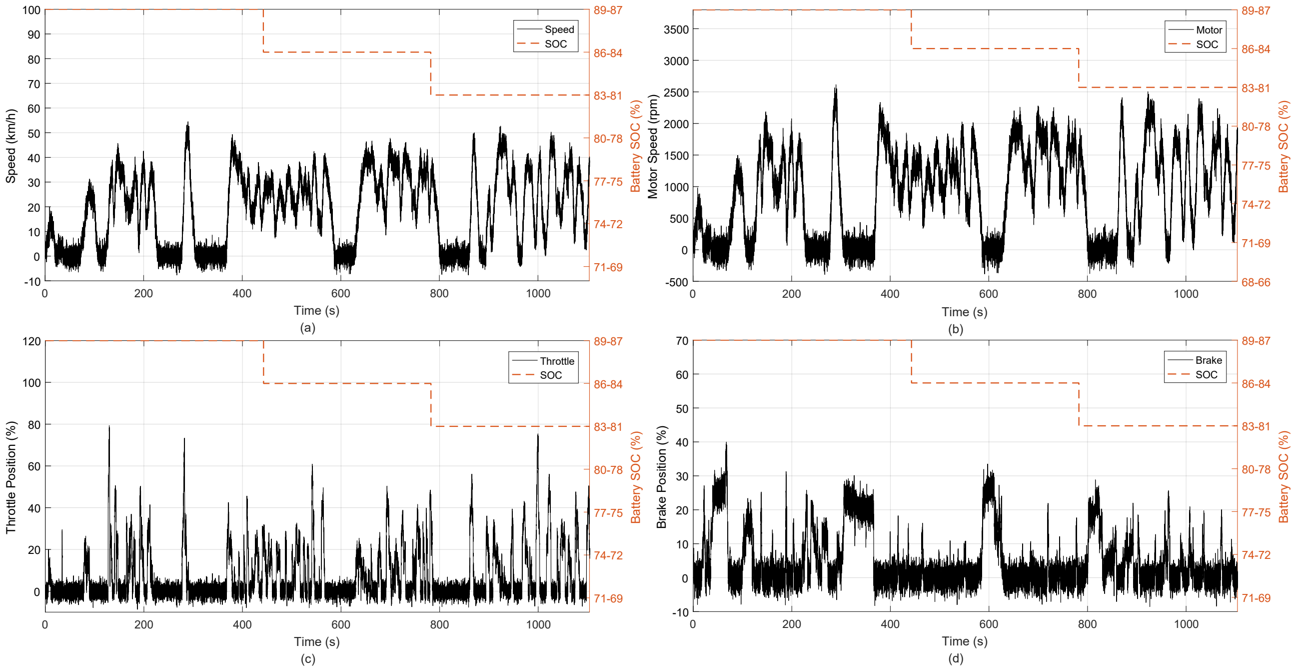

Fig. 3. Driving pattern when battery SOC drops from 89% to 81%. (a) Vehicle speed, (b) Motor speed, (c) Throttle position, (d) Brake pressure.

represented by throttle and brake. Lastly, motor speed reflects the efficiency of translating the motor rotational work into linear movement. In most cases, the motor speed should correlate with the vehicle speed.

Fig. 3 illustrates the trend of battery consumption for the corresponding driving pattern. It is clear that the motor speed follows the pattern of vehicle speed quite closely. We can interpret the figure by looking at three time segments, each representing the battery SOC range. When the battery is 86-84%, the driver appeared to press the throttle more frequently compared to the previous time segment. Hence, the battery discharged at faster rate. Because of the low precision level of measured SOC, we are not able to see the effect of regeneration when the vehicle decelerates.

III. Data Analysis

This paper employs exploratory data analysis to find interesting characteristics of time-series driving dataset. To that end, principal component analysis (PCA) is a suitable tool because we are interested in different driving parameters forming multivariable data.

PCA essentially re-expresses a set of n observations of k variables so that it results in smaller sets of components of the original variables. With the reduction in dimensions, the data analysis will be especially easy to handle. Consider that the dataset is described in the form of $k \times n$ matrix, $\mathbf{X}$, where the n columns are the observations and the k rows are the variables. To transform matrix $\mathbf{X}$ into a new matrix $\mathbf{Y}$ ($k \times n$), $\mathbf{X}$ should be projected onto $k \times k$ matrix, $\mathbf{P}$, as described in (1).

$$\mathbf{Y} = \mathbf{PX}. \tag{1}$$

Here, $\mathbf{P}$ is also known as basis. Each row of $\mathbf{P}$, $\{\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}, \cdots, \mathbf{p_k}\}$ is the vector that describes the direction in which the corresponding row of $\mathbf{X}$ is projected. PCA dictates that $\mathbf{P}$ should be equal to the eigenvector of $\mathbf{XX}^T$. Then, we organize the rows of $\mathbf{P}$ by inserting the eigenvectors in the order of their corresponding eigenvalues. This is equivalent to saying that PC1 (highest eigenvalue) is placed on the first row, and so on, as illustrated in (2). For more detailed derivation of PCA, one can refer to Shlens' paper [17].

$$\mathbf{P} = \begin{pmatrix} \mathbf{e_1} \\ \mathbf{e_2} \\ \vdots \\ \mathbf{e_k} \end{pmatrix} \ni d_1 > d_2 > ... > d_k. \tag{2}$$

where $\mathbf{e}$ is the eigenvector of $\mathbf{XX}^T$ and the corresponding eigenvalue is represented by d. Note that the new basis, $\{\mathbf{e_1}, \mathbf{e_2}, \cdots, \mathbf{e_k}\}$ are orthogonal.

In this study, we are investigating the impacts of vehicle speed, motor speed, throttle position and brake pressure on battery SOC. Here, PCA is applied to reduce the four variables of interest into two components (PC1 and PC2). The data analysis is run using the Pandas and Scikit-learn Python modules.

From Fig. 3, it is clear that the four variables are measured on different scales. For example, speed is in the unit of km/h, while throttle position is quantified by percentage. Applying PCA on these data directly will result in inaccurate interpretation of relationship between two variables, as the data in each variable is spread out in different manner. To resolve this issue, it is a good practice to modify the data by standardizing to zero mean and unit variance.

TABLE I
EIGENVECTOR AND PC SCORE OF EACH PC AXIS.

	PC1		PC2		PC3	
Variable	**Eigenvector**	**Score**	**Eigenvector**	**Score**	**Eigenvector**	**Score**
Vehicle Speed	0.637	0.940	-0.162	-0.155	0.294	0.273
Motor Speed	0.650	0.960	-0.194	-0.186	0.174	0.161
Throttle Position	0.254	0.375	0.966	0.927	0.038	0.035
Brake Pressure	-0.329	-0.486	0.048	0.045	0.939	0.872
Eigenvalue	2.182		0.919		0.863	
Explained Variance	54.6		23		21.6	

We examine the effectiveness of PCA based on several indicators: eigenvalue, eigenvector, PC score and explained variance. Firstly, eigenvalue is the measure of variability of transformed data. Generally agreed criterion has it that the PC should have eigenvalue greater than 1 to be considered meaningful. PC score is the composition of the variable as a result of transformation. Lastly, explained variance reveals the percentage of entire data covered by the PC. Table I shows the indicators as outputted from the PCA of our dataset. PC1 has an eigenvalue of 2.182 and covers 54.6% of total variation. The eigenvector and PC score suggest that vehicle and motor speeds make up the major components in PC1. For PC2, it represents 23% of total variation, and throttle position is the most significant variable. The variables that are known to require the energy from battery are reflected in PC1 and PC2. Using these two axes, the PCA is able to capture 77.6% of variation in the dataset.

Fig. 4 shows the scatterplot of PC1 and PC2 for two of the analyzed datasets, each is associated with the duration for SOC to drop by 1%. The durations of T2, marked by circle, and T4 are 339.35 s and 874.35 s, respectively. In T2, the EV was on active driving mode, infrequent stopping, which causes the battery to discharge faster compared to T4. It is also worth pointing out that there seems to be a pattern how the duration is related to the data distribution. The data in T4 is sparsely spread out in the first quadrant (positive PC1 and PC2) whereas the data in T2 is more concentrated near origin. We are only interested in the first quadrant, because it is where vehicle speed, motor speed and throttle position are positive and above mean. The brake pressure does not influence the discharge rate of battery the same way like the aforementioned variables, rather it results in regeneration, which could be an interesting focus in another study.

One of the ways to characterize the data distribution is to calculate the centroid. Consider $\mathbf{X_{Ti}} = (\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_j})$, where Ti is the segment representing battery retaining its SOC and j is the number of data points, is the observation in the first quadrant of PCA plot. The centroid is essentially the mean point of a cluster, or described as follows:

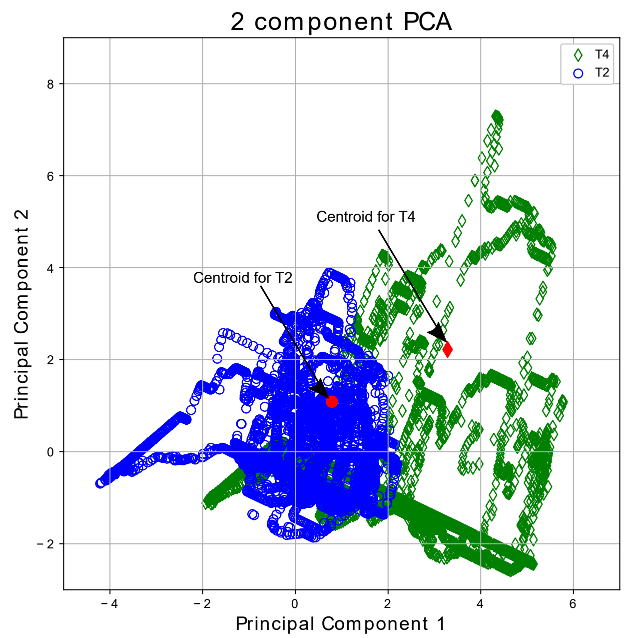

Fig. 4. PCA plot of PC1 and PC2 for driving data of different durations. T2 and T4 are 339.35 s and 874.35 s, respectively.

$$\mathbf{c_{Ti}} = \frac{1}{j}\sum_{i=1}^{j} \mathbf{x_i}, \mathbf{x_i} \in \mathbf{X_{Ti}}. \tag{3}$$

The process of arriving at this point of data analysis can be summarized in Algorithm 1. The fact that the data for each driving parameter is transmitted at different baud rate on CAN bus requires post-processing to resample the data so that they are all on the same timestep, as spelled out in Step 4. The resampling also equalizes the length of each variable. Recalling the matrix $\mathbf{X}$ in (1), it has n observations and k variables, making up the full $k \times n$ matrix. If one of the variables collected has deficient number of observations, we use resampling technique to expand its length to k.

Algorithm 1 PCA of driving dataset

Load data of battery SOC, vehicle speed (**v**), motor speed (**m**), throttle position (**t**) and brake position (**b**)
Segment each variable based on SOC duration Ti ($\mathbf{v_{Ti}}$, $\mathbf{m_{Ti}}$, $\mathbf{t_{Ti}}$, $\mathbf{b_{Ti}}$)
for number of segment, $Ti = T1, T2, \cdots, Tk$ **do**
    Remove noise, align all variables on same timestep ($\mathbf{v'_{Ti}}$, $\mathbf{m'_{Ti}}$, $\mathbf{t'_{Ti}}$, $\mathbf{b'_{Ti}}$)
    Normalize $\mathbf{v'_{Ti}}$, $\mathbf{m'_{Ti}}$, $\mathbf{t'_{Ti}}$, $\mathbf{b'_{Ti}}$
Concatenate all variables in dataframe
Perform PCA
for number of segment, $Ti = T1, T2, \cdots, Tk$ **do**
    $\mathbf{X_{Ti}}$ = observations in the first quadrant
    Calculate centroid for $\mathbf{X_{Ti}}$ using (3)

IV. Discussion

After finding the centroid for each segment Ti, we are interested to study its relationship with the duration of battery retaining its SOC before it drops by 1%. The Euclidean distance of centroid from the origin is calculated. Then, we conduct the linear regression to generate a model that can predict the SOC duration for the given driving parameters.

The result of the linear regression is illustrated in Fig. 5, following a trend that when centroid distance increases, the SOC duration increases. In other words, the SOC duration can be predicted linearly. This is corroborated by the Pearson correlation coefficient (0.92) obtained in this analysis, which indicates a strong, positive linear relationship between SOC duration and centroid distance. It can be explained as follows. Constantly high vehicle and motor speeds and repeated acceleration are associated with aggressive driving. When these parameters are expressed in PC axes, the data appears to be compactly distributed in the first quadrant of PCA plot, resulting in a centroid closer to origin. The opposite can be said for the gentle driving with frequent deceleration, whereby large centroid distance is related to long SOC duration.

There are some performance indices we can use to evaluate the accuracy of the regression analysis, which are summarized as follows:

- MAE: 77.636
- MSE: 8953.683
- R^2: 0.8243

The mean absolute error (MAE) indicates that the model predicts the result with average error of 77.636 s. The model has the mean squared error (MSE) of 8953.683, this is largely due to the prevalent presence of outlier in the centroid distance between 1 to 7. The R^2 is a measure of how close the data are to the fitted regression line, ranging between 0 to 1. In our case, an R^2 of 0.8243 suggests that the model is reasonably good at fitting the given data.

The following comments are cited based on the results presented in this paper:

1) Understanding the influence of driving parameters is necessary to develop the battery management system (BMS). The results have shown that the battery SOC can be affected by the way the EV is operated. Hence, the BMS can be incorporated with the SOC prediction function by utilizing the readily available CAN bus data. This information will be useful for the driver to make conscious effort changing their driving behavior in order to prolong the battery life.
2) It is commonly known that temperature is a critical factor which causes battery degradation. However, the data of battery temperature was not available at the time the experiment was conducted. The regression analysis can be further improved by taking into account the temperature in future.
3) The capability of regenerating energy back to battery is what differentiates EV from conventional vehicle. Due to the poor measurement precision of battery SOC in this study, we are not able to clearly see the effect of regeneration. It is definitely in our interest to improve the data collection approach by using higher precision, which will provide us microscopic view of charging and discharging dynamics of battery.
4) Overall, our regression model does a good job at fitting all the provided data. However, we observe a tendency that most of the data fall in the centroid distance between 1 to 7, as shown in Fig. 5. The model can be significantly improved if we have more data outside this region ($>$ 10). To collect these data, we will have to manipulate the way the EV is driven. It will be interesting to see if the data are still following the regression line.

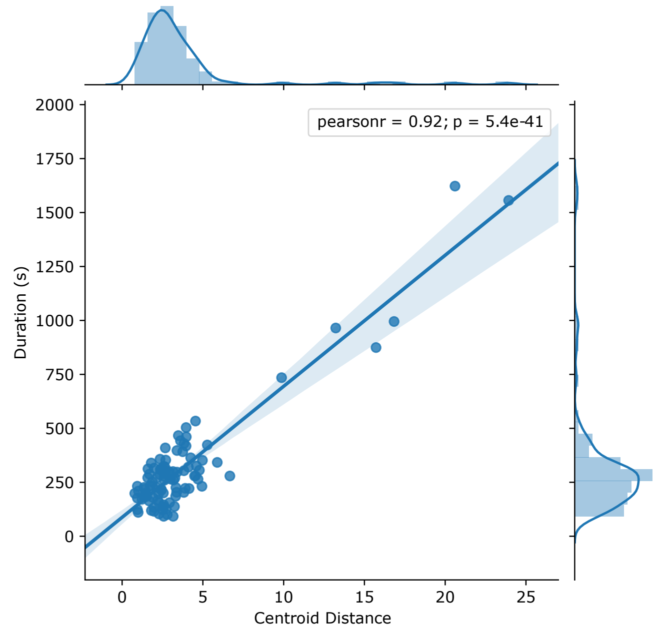

Fig. 5. Influence of PC components on SOC duration. Each point represents battery discharging segment (Ti).

V. Conclusion

Battery is an integral part of electric vehicle (EV). As more transport operators are deploying EV in their fleet, a robust battery management system (BMS) will become essential to prolong the mileage of the battery. To that end, it is important to understand how the driving behavior can influence the battery health.

In this paper, we have looked into the relationship between different driving parameters and battery state of charge (SOC). The data was collected from experiment of driving electric vehicle (EV) in uncontrolled condition in Singapore. The data of interest are vehicles speed, motor speed, throttle position, brake pressure and battery SOC. Due to multidimensionality of data, we intended to reduce it using Principal Component Analysis (PCA). We observed that the distribution of the data in the first quadrant of PCA plot exhibits a pattern in relation to the duration of battery retaining its SOC. The characteristic of distribution can be represented by its center point, or called centroid. Then, we performed regression analysis to generate a model that can predict the SOC duration for the given centroid. The regression fitted all data points reasonably well with R^2 of 0.824. We suggested a few approaches in hope to improve the experiment in future.

References

[1] International Energy Agency, *CO2 Emissions from Fuel Combustion 2018*, 2018.

[2] C. Marina Martinez, M. Heucke, F. Wang, B. Gao, and D. Cao, "Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 666–676, March 2018.

[3] U. Fugiglando, E. Massaro, P. Santi, S. Milardo, K. Abida, R. Stahlmann, F. Netter, and C. Ratti, "Driving behavior analysis through can bus data in an uncontrolled environment," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 737–748, Feb 2019.

[4] Z. Constantinescu, C. Marinoiu, and M. Vladoiu, "Driving style analysis using data mining techniques," *International Journal of Computers, Communications & Control (IJCCC)*, vol. V, no. 5, pp. 654–663, December 2010.

[5] I. del Campo, E. Asua, V. Martínez, O. Mata-Carballeira, and J. Echanobe, "Driving style recognition based on ride comfort using a hybrid machine learning algorithm*," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov 2018, pp. 3251–3258.

[6] H. Woo, Y. Ji, Y. Tamura, Y. Kuroda, T. Sugano, Y. Yamamoto, A. Yamashita, and H. Asama, "Driver classification in vehicle following behavior by using dynamic potential field method," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Oct 2017, pp. 1–6.

[7] L. Xu, J. Hu, H. Jiang, and W. Meng, "Establishing style-oriented driver models by imitating human driving behaviors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2522–2530, Oct 2015.

[8] J. Jeong, J. Choi, H. Seo, Y. Park, and S. W. Cha, "Analysis of fuel economy and battery life depending on the types of hev using dynamic programming," in *2013 World Electric Vehicle Symposium and Exhibition (EVS27)*, Nov 2013, pp. 1–5.

[9] S. Li, W. Wang, Z. Mo, and D. Zhao, "Cluster naturalistic driving encounters using deep unsupervised learning," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, June 2018, pp. 1354–1359.

[10] W. Wang, J. Xi, A. Chong, and L. Li, "Driving style classification using a semisupervised support vector machine," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 5, pp. 650–660, Oct 2017.

[11] A. Braun and W. Rid, "Energy consumption of an electric and an internal combustion passenger car. a comparative case study from real world data on the erfurt circuit in germany," *Transportation Research Procedia*, vol. 27, pp. 468 – 475, 2017, 20th EURO Working Group on Transportation Meeting, EWGT 2017, 4-6 September 2017, Budapest, Hungary.

[12] C. Bingham, C. Walsh, and S. Carroll, "Impact of driving characteristics on electric vehicle energy consumption and range," *IET Intelligent Transport Systems*, vol. 6, no. 1, pp. 29–35, March 2012.

[13] S. Rangaraju, L. D. Vroey, M. Messagie, J. Mertens, and J. V. Mierlo, "Impacts of electricity mix, charging profile, and driving behavior on the emissions performance of battery electric vehicles: A belgian case study," *Applied Energy*, vol. 148, pp. 496 – 505, 2015.

[14] Z. Younes, L. Boudet, F. Suard, M. Gérard, and R. Rioux, "Analysis of the main factors influencing the energy consumption of electric vehicles," in *2013 International Electric Machines Drives Conference*, May 2013, pp. 247–253.

[15] W. Li, P. Stanula, P. Egede, S. Kara, and C. Herrmann, "Determining the main factors influencing the energy consumption of electric vehicles in the usage phase," *Procedia CIRP*, vol. 48, pp. 352 – 357, 2016, the 23rd CIRP Conference on Life Cycle Engineering.

[16] D. A. Martínez, J. D. Poveda, and D. Montenegro, "Li-ion battery management system based in fuzzy logic for improving electric vehicle autonomy," in *2017 IEEE Workshop on Power Electronics and Power Quality Applications (PEPQA)*, May 2017, pp. 1–6.

[17] J. Shlens, "A tutorial on principal component analysis," *arXiv preprint arXiv:1404.1100*, 2014.