State of Charge Estimation of Lithium-ion Batteries using Hybrid Machine Learning Technique

Manjot S. Sidhu, *Student Member, IEEE*, Deepak Ronanki, *Student Member, IEEE* and Sheldon Williamson, *Senior Member, IEEE*
Advanced Storage Systems and Electric Transportation (ASSET) Laboratory
Smart Transportation Electrification and Energy Research (STEER) Group
Department of Electrical, Computer, and Software Engineering
Faculty of Engineering and Applied Science
University of Ontario-Institute of Technology
2000 Simcoe Street North, Oshawa, ON L1G 0C5, Canada
Email:sheldon.williamson@uoit.ca

***Abstract*—The pivotal features of low self-discharge, high energy density and long calendar life lead the Lithium-ion (Li-ion) batteries as being a mainstream energy storage source in electric vehicles (EVs). A meticulous estimation of the state of charge (SOC) is indispensable for ensuring safe and reliable operations in battery powered EVs. However, SOC estimation of Li-ion battery with high accuracy have become a major challenge in the automotive industry. To fulfill reliable operation in EVs, researchers have proposed numerous SOC estimators through model based or machine learning techniques. This paper presents an improved SOC estimation of Li-ion battery using random forest (RF) regression, which is robust and effective for controlling dynamic systems. To ensure good resilience and accuracy, a Gaussian filter is adopted at the final stage to minimize the variations in the SOC estimation. The proposed SOC estimator is verified on the experimental data of the Li-ion battery under Federal test driving schedules and different operating temperatures. Results show that the proposed SOC estimator displays sufficient accuracy and outperforms the traditional artificial intelligence based approaches.**

***Index Terms*—Artificial intelligence, battery management systems, Gaussian processes, lithium- ion batteries, machine learning, random forest regression.**

I. Introduction

The rapid depletion of petroleum-based energy resources, global warming and environmental pollution caused by fossil fuels have lead to the triumph of electric vehicle (EV) technologies [1]. These systems employ a battery as the primary energy source, which plays a crucial role in the range and lifetime of EVs. To meet the current requirements and automotive standards, enhancing the battery performance and its life have become an active area of research. Over the past years, EV technologies have been equipped with different energy storage systems including lead-acid, Nickel-Cadmium (Ni-Cd), Nickel-metal hydride (Ni-MH) and li-ion batteries [2]. Among them, Li-ion batteries have become promising energy storage systems for EVs due to high energy density, low self-discharge, high thermal voltage, long lifespan and environmental friendliness. However, managing these batteries from over-charging or over-discharging is an important aspect in order to avoid explosions or fire catastrophes [3].

Estimation of SOC of the battery is the main index to evaluate the range of the EV as it indicates the amount of energy for electrical assist operation. Moreover, the life of the battery is improved if it is operated within a particular SOC range [4], [5]. Thus, a meticulous estimation of SOC is essential to ensure the safe operation of the batteries as well as effective energy management. However, it is a challenging task to achieve due to complex chemical reactions and nonlinear characteristics of the battery. Furthermore, the SOC cannot be measured directly, and must be estimated from other battery quantities. Thus, SOC estimation of a Li-ion battery is a complicated process and depends on various parameters such as its internal resistance, capacitance and operating temperature [6].

Numerous SOC estimation algorithms have been established well in the literature [7]–[22]. Among them, Coulomb counting (ampere-hour method) has gained huge popularity due to its simple implementation with low-power consumption. However, it is an open-loop method and its estimation process tends to drift from the original values due to cumulative effect of the measurement and calculation errors by the integration function. Furthermore, it fails to determine initial value of the SOC as well as introduces a bias in the estimation process [10]. Open circuit voltage (OCV) is an off-line method which achieves SOC estimation with high accuracy. Nevertheless, OCV requires long duration for it to be in stable condition. Other methods including electrochemical models [11] and electrochemical impedance spectroscopy [12] are utilized to estimate the SOC. Nonetheless, these methods are sensitive to the temperature and aging effects.

Adaptive methods used for SOC estimation based on Kalman filter [13]–[15] are also presented in the literature. Popular extensions of Kalman filter used for SOC estimation include extended Kalman filter [16], unscented Kalman filter [17] and adaptive unscented Kalman filter [18]. Particle filter [19], [20] and H-infinity filter [21] have also been made use of to estimate SOC. Other approaches for SOC estimation based of machine learning algorithms include support vector regression (SVR) [23] and neural networks (NN) [24] are also

978-1-7281-4878-6/19/$31.00 ©2019 IEEE

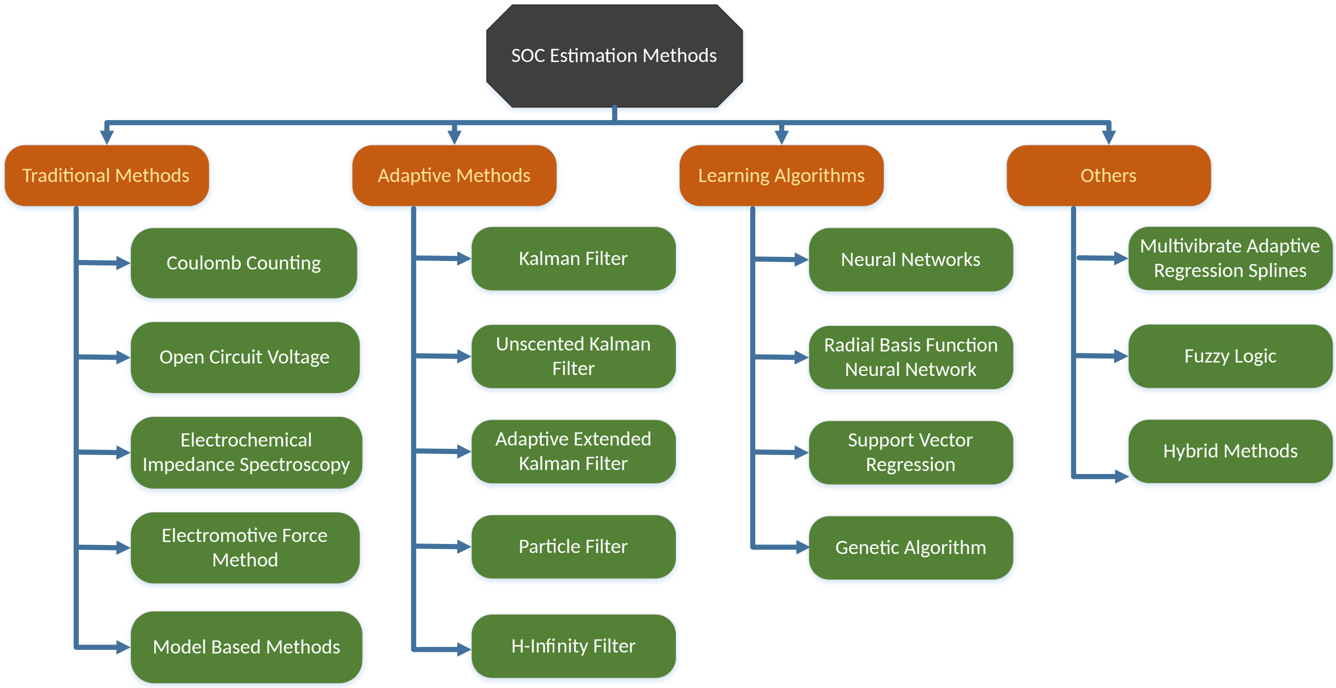

Fig. 1: Classification of popular SOC estimation algorithms.

proposed. SVR is used to predict SOC of lithium-iron battery taking current, voltage and cell temperatures as inputs [25]. This method successfully kept maximum errors below 6% and maintaining accuracy. NN is used to estimate SOC by employing measured current and voltage as inputs and error in results are minimized by using an unscented Kalman filter [26]. Kang *et. al* [27] proposed radial basis function neural network (RBFNN) to estimate SOC using current, voltage and capacity with an absolute error to less than 2%. The most popular SOC estimation methods are classified in Fig. 1.

However, the aforementioned estimation methods involve complex processes and increases computational load for real time implementation. Moreover, no proper studies are performed under dynamic conditions such as Federal test driving schedules. In this paper, an SOC estimation method based on RF regression and Gaussian filter is presented. At first, SOC is predicted using RF regression which takes 4 inputs namely current, voltage, voltage difference from last second and SOC estimated at previous second. Gaussian filter is introduced in next stage to remove variations from the results generated by RF regression. Data used in this study is explained in Section II. Section III explains the method used in this experiment and it also explains the performance evaluation metrics used in the experiment. Results are explained in the Section IV and Section V contains the conclusions drawn from the analysis.

II. Experimental Data

To evaluate the effectiveness of proposed SOC estimation technique, data used in this study was taken from two drive cycles namely dynamic stress test (DST) and US06 highway drive cycle (US06). The duration of DST is 360 s in which current goes from 2 to -4 A. The DST shown in figure 2 is designed by united states advanced battery consortium (USABC) [28] and it covers current with varying intensity and also accounts for regenerative charging. Therefore it was used as training data for the proposed model.

TABLE I: Specifications of test battery

Type	Nominal Capacity (Ah)	Nominal Voltage (V)	Cut-off Voltage (V)	Maximum Current (A)
18650 LNMC	2.0	3.60	2.4 - 4.2	2.2

The model's accuracy and generalization capability was tested by using the data from US06 test profile. The test profile US06 shown in figure 3, is 600 s in duration and it simulates highway driving. US06 test profile simulates real life driving conditions of electric vehicles and are considerably complex than the DST test profile. The above mentioned test profiles were run on a Lithium Nickel Manganese Cobalt Oxide (LNMC) battery whose are specifications are mentioned in Table I.

A. *Data Normalization*

The data used for features of this experiment is scaled between the range of [0,1]. The features were normalized by:

$$x = 2\frac{x - x_{min}}{x_{max} - xmin} \tag{1}$$

where x_{min} and x_{max} are minimum and maximum values of input vector x. Both training and testing data were normalized.

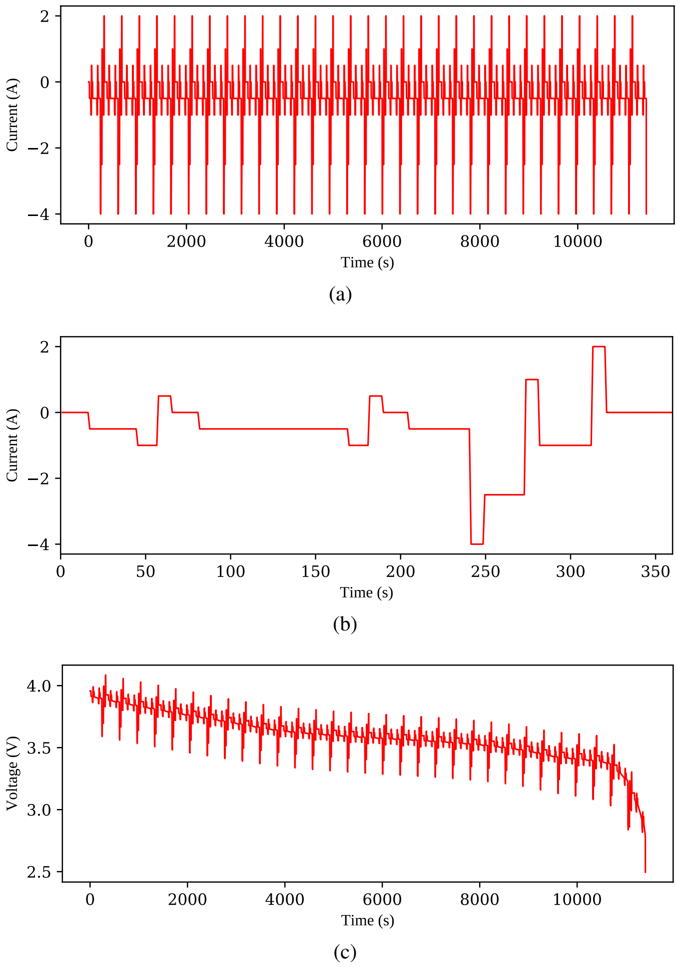

Fig. 2: DST Profile a) Full current cycle b) Zoomed version of current Cycle c) Full voltage cycle during complete test profile.

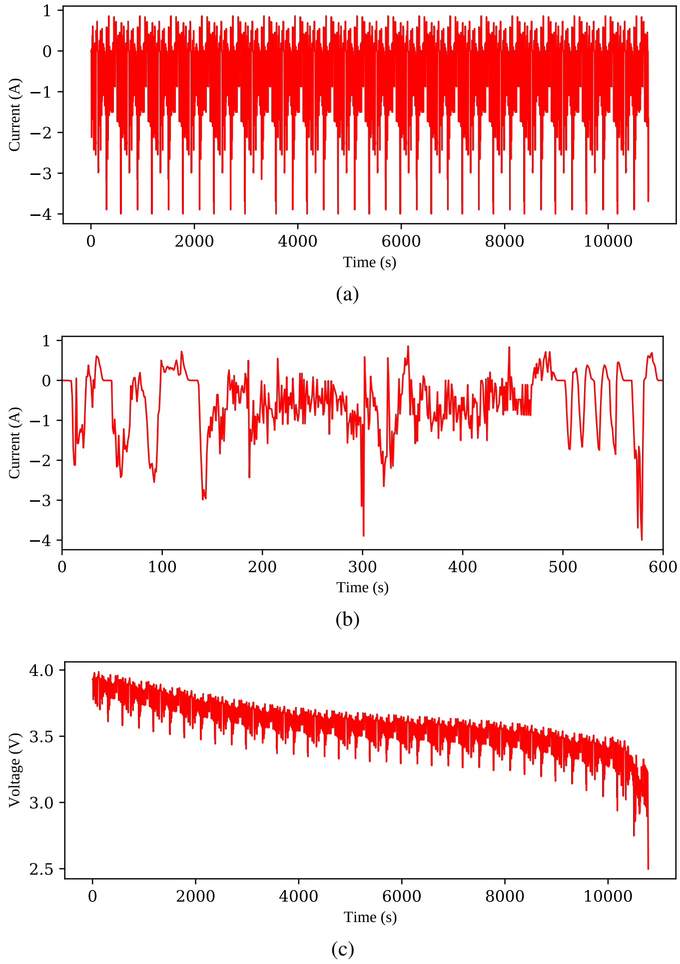

Fig. 3: US06 Profile a) Full current cycle b) Zoomed version of US06 current cycle c) Full voltage cycle during the US06 test profile.

B. Input Features and Output of Model

The inputs used for training the model and predicting SOC are current (A), voltage (V), Δ V (difference between voltage from last second) and SOC_{t-1} (SOC at the end of last second). Because SOC_{t-1} is used in predicting SOC, the SOC from the previous step was divided by 100. For example if the SOC predicted at last step is 70 then SOC_{t-1} for current step will be 70/100 i.e. 0.70. The inputs and range of inputs are shown in Table II. For training the model, SOC was calculated using the equation given below:

$$SOC(t+1) = SOC(t) + \frac{1}{C_n}\int_{t}^{t_0} n.I(t).dt \qquad (2)$$

where C_n represents the capacity of battery, t and t_0 represents the time interval, n is coulombic efficiency during charge/discharge and $I(t)$ is real time discharge current.

III. SOC Estimation Method

The proposed approach involves a two-stage process, which consists of RF regression and Gaussian filter. In the first stage, RF regression method is employed to estimate the SOC based on voltage, current and previous SOC sample data. Gaussian filter is introduced in the final stage so that variability produced from the RF regression can be reduced in order to obtain results with high accuracy rate. The overall block diagram of the proposed SOC technique is shown in Fig. 4.

TABLE II: Input variables used in the study.

Input Variables	Range
Current (A)	(-4) - 2
Voltage (V)	2.4 - 4.2
Δ V	0.0 - 0.1
SOC_{t-1}	0.0 - 0.8

A. Random Forest Regression

RF regression [29] generates many decision trees for regression and its output is calculated by averaging the output of all decision trees. The working principle of RF regression is shown in Fig. 5. The decision tree [30] is a model that does not have any prior tree structure. The structure of the tree depends on the complexity of the training data during

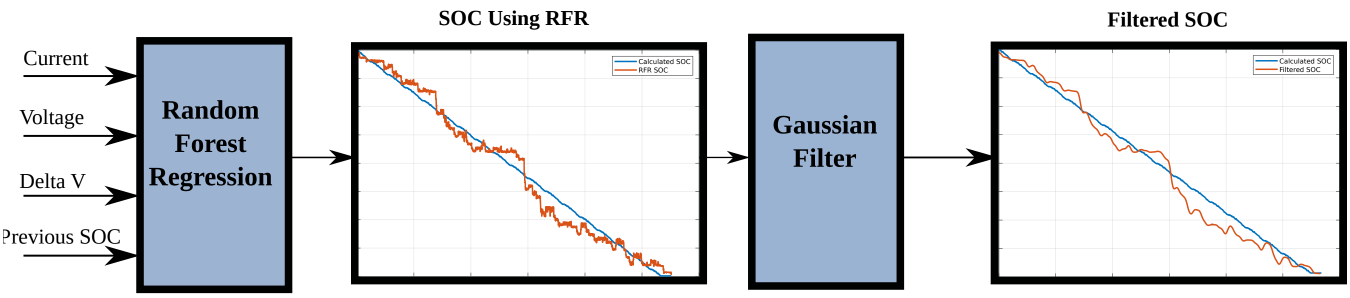

Fig. 4: Block diagram of the proposed SOC estimation method

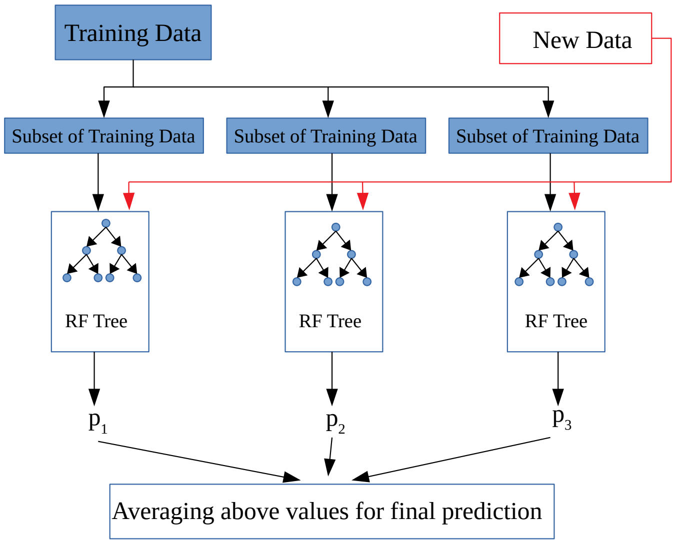

Fig. 5: Working principle of RF regression.

the learning stage. The Decision tree comprises of two nodes: decision node and leaf node. Every sample of training data is evaluated by the decision nodes and passed onto different nodes depending on the value of the features of the sample.

RF regression generates regression trees using the training data X, which is given by, $X = x_1, x_2, x_3 \dots\dots\dots x_n$ which produces forest. This method produces k outputs $T_1(x), T_2(x), \dots\dots, T_k(x)$ corresponding to each tree. To calculate the final result, all of tree predictions are averaged with the equation given below:

$$RF\,(X) = \frac{1}{k}\sum_{k=1}^{k}\hat{T}_k\,(x)\,. \tag{3}$$

Steps included in RF are shown below:

1) First of all, inputs for RF regression model are identified which in our case are Current, Voltage, ΔV and SOC_{t-1}.
2) Grow each tree in the forest, while optimal number of trees (n = 200) selected, making use of new training set which is generated from original data with replacement. Finding optimal trees depends on evaluation of RF done by using Mean squared error (MSE).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - Y_i) \tag{4}$$

where n represents number of samples, P_i represents predicted values and Y_i represents true values.
3) Predict SOC values for new data by averaging the predictions of n trees.

The output produced from RF regression contained small variations. To remove the variations, Gaussian filter [31]–[33] is introduced in second stage. The equation for Gaussian filter is as follows:

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}}exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \tag{5}$$

where μ is mean value, σ is standard deviation of Gaussian distribution.

IV. Results and Discussion

A. Performance Evaluation

Statistics used for evaluating the model are mean absolute error (MAE) and coefficient of determination (COD). Both of these can be calculated by using predicted and true values.

1) MAE: It is calculated by using predicted values and real values of the SOC. MAE is the unweighted average of absolute differences between predicted and true values. MAE is given by:

$$MAE = \frac{\sum_{i=1}^{n}|p_i - y_i|}{n} \tag{6}$$

where p_i is the predicted value of target variable and y_i is the true value of target variable. Smaller values of MAE means better performance of the model.

2) COD: It indicates variance in predicted variable that is predicted from the features set. COD can be calculated using sum of squares. It's value always remains between 0 and 1, 1 is indicative of the model being a perfect fit. It can be calculated from following equation.

$$COD = 1 - \frac{SS_{res}}{SS_{tot}} \tag{7}$$

where, total sum of squares (SS_{tot}) is given by:

$$SS_{tot} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{8}$$

Mean ($\bar{y}$) of samples is calculated by:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} (y_i) \tag{9}$$

Residual sum of squares (SS_{res}) is given by:

$$SS_{res} = \sum_{i=0}^{n} e_i^2 \tag{10}$$

where e is the difference between true values and the values predicted by model.

B. Discussion

TABLE III: MAE for US06 test profile at different temperatures.

Temperature (^oC)	RFR	SVR	NN
0	2.59	3.89	4.87
25	2.88	5.70	6.90
45	2.88	3.08	3.75

The study was conducted using Python version 3.6 and MATLAB 2018a. The comparison between filtered RFR and filtered NN, filtered SVR is shown in Figs. 6 (a) - (c). Fig. 6 (a) shows the results at 0°C which shows RFR tracks the true SOC line better than other methods. Fig. 6 (b) shows the results of RFR at 25°C and Fig. 6 (c) shows the results from 45°C. Comparison of performance of all methods is shown in Table III. RFR shows the minimum value of MAE of 2.59 at 0°C while at the same temperature MAE for SVR and NN is 3.89 and 4.87 respectively. At 25°C, the value of MAE for RFR is 2.88 and at 45°C RFR has same MAE value. SVR shows worst results at 25°C where the MAE jumps to 5.70. SVR shows the MAE of 3.08 at 45°C. NN shows the worst performance at 25°C with MAE of 6.90 and it shows best performance at 45°C with MAE of 3.75. At all three temperatures, filtered RFR shows better performance than SVR and NN. Table IV shows the coefficient of determination for all the models tested in this experiment. At 25°C SVR and NN show worst performance with COD value of 0.92 and 0.88 respectively. The COD of RF regression is 0.97 through all operating temperatures tested, which means 97% variability

TABLE IV: COD for US06 test profile at different operating temperatures.

Temperature (^oC)	RFR	SVR	NN
0	0.97	0.95	0.93
25	0.97	0.92	0.88
45	0.97	0.97	0.96

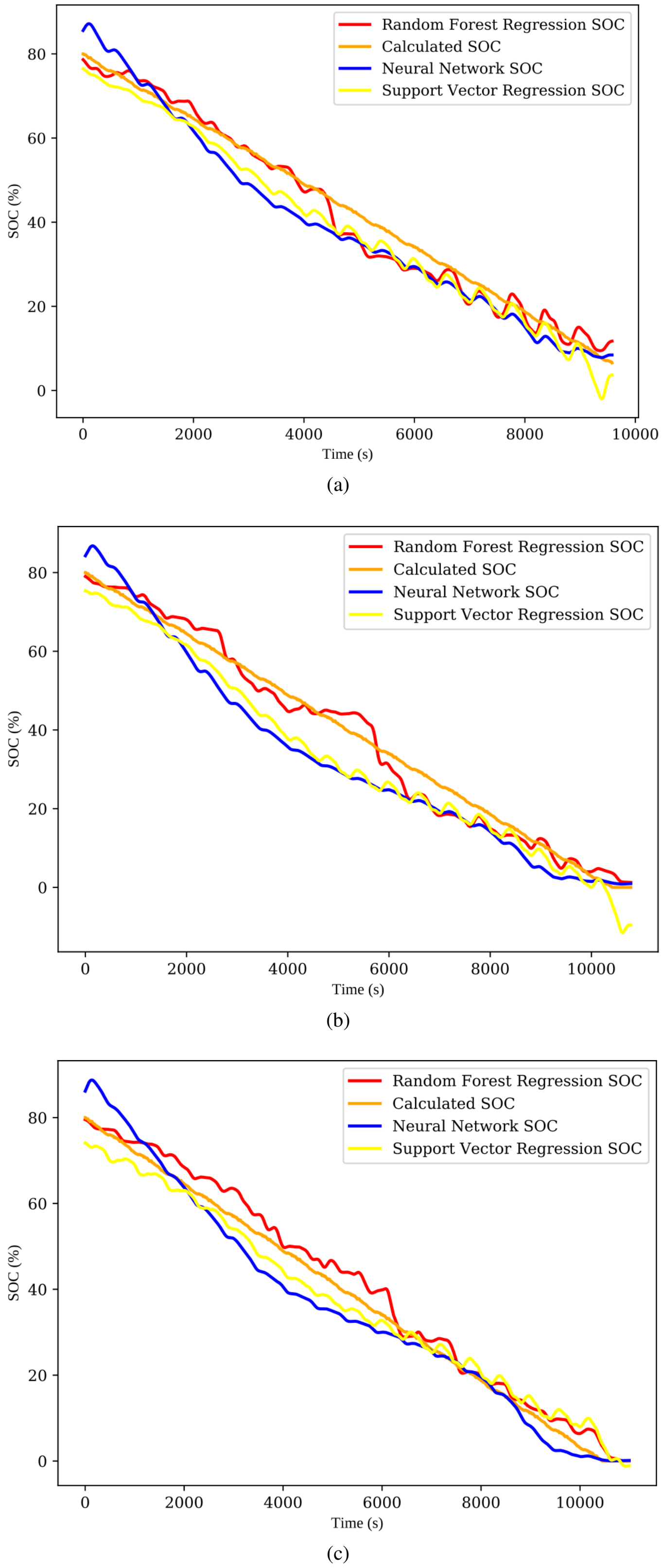

Fig. 6: SOC estimation results at different operating temperatures a) 0°C b) 25°C c) 45°C.

in predictions can be explained by the features used. On the other hand, SVR shows COD value of 0.95 at $0°$C and 0.97 at $45°$C. While NN performance was poor with COD values of 0.93 at $0°$C and 0.96 at $45°$C.

V. Conclusion

A meticulous SOC estimation is a major challenge in battery powered electrified vehicles due to non-linear characteristics and complex chemistry of the batteries. In this paper, hybrid SOC estimation method combining the RF regression and Gaussian filter is presented. It is observed that the proposed method reduces MAE and COD by 2.88 and 0.97, respectively for US06 cycle compared to SVR and NN. The proposed SOC estimation method eliminates prior knowledge of battery model and implementation of complex mathematical equations. Comprehensive test results confirm that adaptability of the proposed SOC estimation approach for real-time applications. Furthermore, the proposed method effectively works with a greater accuracy under dynamic conditions and outperforms conventional methods such as SVR and NN methods. The proposed approach could be designed for modularized battery design and applied to different battery chemistries for electric vehicles as well as sustainable energy applications.

References

[1] D. Ronanki, A. Hemasundar, and P. Parthiban, "A small 4-wheeler ev propulsion system using dtc controlled induction motor," in *Proceedings of the World Congress on Engineering*, vol. 2, 2013.

[2] Y. Jiang, L. Kang, and Y. Liu, "A unified model to optimize configuration of battery energy storage systems with multiple types of batteries," *Energy*, vol. 176, pp. 552–560, 2019.

[3] J. Garche, A. Jossen, and H. Döring, "The influence of different operating conditions, especially over-discharge, on the lifetime and performance of lead/acid batteries for photovoltaic systems," *Journal of Power Sources*, vol. 67, no. 1-2, pp. 201–212, 1997.

[4] S. S. Williamson, A. K. Rathore, and F. Musavi, "Industrial electronics for electric transportation: Current state-of-the-art and future challenges," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 5, pp. 3021–3032, 2015.

[5] D. Ronanki and S. S. Williamson, "Modular multilevel converters for transportation electrification: challenges and opportunities," *IEEE Transactions on Transportation Electrification*, vol. 4, no. 2, pp. 399–407, 2018.

[6] R. Zhang, B. Xia, B. Li, L. Cao, Y. Lai, W. Zheng, H. Wang, and W. Wang, "State of the art of lithium-ion battery soc estimation for electrical vehicles," *Energies*, vol. 11, no. 7, p. 1820, 2018.

[7] P. Shen, M. Ouyang, L. Lu, J. Li, and X. Feng, "The co-estimation of state of charge, state of health, and state of function for lithium-ion batteries in electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 1, pp. 92–103, 2018.

[8] Y. Xing, W. He, M. Pecht, and K. L. Tsui, "State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures," *Applied Energy*, vol. 113, pp. 106–115, 2014.

[9] D. Li, J. Ouyang, H. Li, and J. Wan, "State of charge estimation for limn2o4 power battery based on strong tracking sigma point kalman filter," *Journal of Power Sources*, vol. 279, pp. 439–449, 2015.

[10] T. Hansen and C.-J. Wang, "Support vector based battery state of charge estimator," *Journal of Power Sources*, vol. 141, no. 2, pp. 351–358, 2005.

[11] M. Corno, N. Bhatt, S. M. Savaresi, and M. Verhaegen, "Electrochemical model-based state of charge estimation for li-ion cells," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 1, pp. 117–127, 2014.

[12] M. Li, "Li-ion dynamics and state of charge estimation," *Renewable Energy*, vol. 100, pp. 44–52, 2017.

[13] P. Spagnol, S. Rossi, and S. M. Savaresi, "Kalman filter soc estimation for li-ion batteries," in *2011 IEEE International Conference on Control Applications (CCA)*. IEEE, 2011, pp. 587–592.

[14] I. Baccouche, S. Jemmali, B. Manai, N. Omar, and N. Amara, "Improved ocv model of a li-ion nmc battery for online soc estimation using the extended kalman filter," *Energies*, vol. 10, no. 6, p. 764, 2017.

[15] S. Sepasi, R. Ghorbani, and B. Y. Liaw, "A novel on-board state-of-charge estimation method for aged li-ion batteries based on model adaptive extended kalman filter," *Journal of Power Sources*, vol. 245, pp. 337–344, 2014.

[16] J. Lee, O. Nam, and B. Cho, "Li-ion battery soc estimation method based on the reduced order extended kalman filtering," *Journal of power sources*, vol. 174, no. 1, pp. 9–15, 2007.

[17] W. Wang, X. Wang, C. Xiang, C. Wei, and Y. Zhao, "Unscented kalman filter-based battery soc estimation and peak power prediction method for power distribution of hybrid electric vehicles," *IEEE Access*, vol. 6, pp. 35 957–35 965, 2018.

[18] Y. Li, C. Wang, and J. Gong, "A multi-model probability soc fusion estimation approach using an improved adaptive unscented kalman filter technique," *Energy*, vol. 141, pp. 1402–1415, 2017.

[19] B. Xia, Z. Sun, R. Zhang, D. Cui, Z. Lao, W. Wang, W. Sun, Y. Lai, and M. Wang, "A comparative study of three improved algorithms based on particle filter algorithms in soc estimation of lithium ion batteries," *Energies*, vol. 10, no. 8, p. 1149, 2017.

[20] Y. Wang, C. Zhang, and Z. Chen, "A method for state-of-charge estimation of lifepo4 batteries at dynamic currents and temperatures using particle filter," *Journal of power sources*, vol. 279, pp. 306–311, 2015.

[21] R. Xiong, Q. Yu, L. Y. Wang, and C. Lin, "A novel method to obtain the open circuit voltage for the state of charge of lithium ion batteries in electric vehicles by using h infinity filter," *Applied energy*, vol. 207, pp. 346–353, 2017.

[22] M. S. Sidhu, D. Ronanki, and S. Williamson, "Hybrid state of charge estimation approach for lithium-ion batteries using k-nearest neighbour and gaussian filter-based error cancellation," in *2019 IEEE 28th International Symposium on Industrial Electronics (ISIE)*, June 2019, pp. 1506–1511.

[23] J. Meng, G. Luo, and F. Gao, "Lithium polymer battery state-of-charge estimation based on adaptive unscented kalman filter and support vector machine," *IEEE Transactions on Power Electronics*, vol. 31, no. 3, pp. 2226–2238, 2015.

[24] M. A. Hannan, M. S. H. Lipu, A. Hussain, M. H. Saad, and A. Ayob, "Neural network approach for estimating state of charge of lithium-ion battery using backtracking search algorithm," *Ieee Access*, vol. 6, pp. 10 069–10 079, 2018.

[25] J. C. A. Anton, P. J. G. Nieto, C. B. Viejo, and J. A. V. Vilán, "Support vector machines used to estimate the battery state of charge," *IEEE Transactions on power electronics*, vol. 28, no. 12, pp. 5919–5926, 2013.

[26] W. He, N. Williard, C. Chen, and M. Pecht, "State of charge estimation for li-ion batteries using neural network modeling and unscented kalman filter-based error cancellation," *International Journal of Electrical Power & Energy Systems*, vol. 62, pp. 783–791, 2014.

[27] L. Kang, X. Zhao, and J. Ma, "A new neural network model for the state-of-charge estimation in the battery degradation process," *Applied Energy*, vol. 121, pp. 20–27, 2014.

[28] T. Duong, "Usabc and pngv test procedures [us advanced battery consortium (usabc); partnership for a new generation of vehicles (pngv)]," 2000. [Online]. Available: https://avt.inl.gov/sites/default/files/pdf/battery/usabc_manual_rev2.pdf

[29] Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J. C.-W. Chan, P. van den Bossche, J. Van Mierlo, and N. Omar, "Random forest regression for online capacity estimation of lithium-ion batteries," *Applied energy*, vol. 232, pp. 197–210, 2018.

[30] L. Breiman, *Classification and regression trees*. Routledge, 2017.

[31] Y. Li, M. Abdel-Monem, R. Gopalakrishnan, M. Berecibar, E. Nanini-Maury, N. Omar, P. van den Bossche, and J. Van Mierlo, "A quick on-line state of health estimation method for li-ion battery with incremental capacity curves processed by gaussian filter," *Journal of Power Sources*, vol. 373, pp. 40–53, 2018.

[32] K. Ito, "Gaussian filter for nonlinear filtering problems," in *Decision and Control, 2000. Proceedings of the 39th IEEE Conference on*, vol. 2. IEEE, 2000, pp. 1218–1223.

[33] H. H. Afshari, S. A. Gadsden, and S. Habibi, "Gaussian filters for parameter and state estimation: A general review of theory and recent trends," *Signal Processing*, vol. 135, pp. 218–238, 2017.