

Article

A Battery Health Monitoring Method Using Machine Learning: A Data-Driven Approach

Shehzar Shahzad Sheikh ¹, Mahnoor Anjum ², Muhammad Abdullah Khan ²,
Syed Ali Hassan ², Hassan Abdullah Khalid ¹, Adel Gastli ^{3,*} and Lazhar Ben-Brahim ³

¹ U.S.-Pakistan Center for Advanced Studies in Energy, (USPCAS-E), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; shehzar@ieee.org (S.S.S.); hakhalid@uspcase.nust.edu.pk (H.A.K.)

² School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; manjum.bee15seecs@seecs.edu.pk (M.A.); mabdullah.bee15seecs@seecs.edu.pk (M.A.K.); ali.hassan@seecs.edu.pk (S.A.H.)

³ Department of Electrical Engineering, Qatar University (QU), Doha 2713, Qatar; brahim@qu.edu.qa

* Correspondence: adel.gastli@qu.edu.qa; Tel.: +974-55653362

Received: 20 May 2020; Accepted: 29 June 2020; Published: 15 July 2020



Abstract: Batteries are combinations of electrochemical cells that generate electricity to power electrical devices. Batteries are continuously converting chemical energy to electrical energy, and require appropriate maintenance to provide maximum efficiency. Management systems having specialized monitoring features; such as charge controlling mechanisms and temperature regulation are used to prevent health, safety, and property hazards that complement the use of batteries. These systems utilize measures of merit to regulate battery performances. Figures such as the state-of-health (SOH) and state-of-charge (SOC) are used to estimate the performance and state of the battery. In this paper, we propose an intelligent method to investigate the aforementioned parameters using a data-driven approach. We use a machine learning algorithm that extracts significant features from the discharge curves to estimate these parameters. Extensive simulations have been carried out to evaluate the performance of the proposed method under different currents and temperatures.

Keywords: battery health monitoring; feature extraction; knee-point calculation; machine learning; state of health

1. Introduction

In recent decades, global climate has significantly degraded due to the rampant increase in earth's temperature by carbon-emitting fossil fueled vehicles and conventional power generation resources. This has driven the research and development of energy storage systems. In the evolution of energy storage devices, lithium-ion (Li-ion) has emerged as one of the best-suited and widely-implemented solutions with extensive usage in low-power portable electronic devices, electric vehicles' power sources, and grid-scale energy storage [1]. High energy density, high power density, safety, more life-cycles, and low self-discharge are the main attributes of the Li-ion battery; making it superior to other types of batteries such as lead-acid, sodium sulfur (NaS) and nickel metal hydride (NiMH) [2].

The electrical characteristics of Li-ion batteries (i.e., energy-density, power-density, and C-rate) have been improved significantly, however, the non-linear charging capacity and the accelerated rate of aging still remain a problem [3]. To overcome such inherent pitfalls of Li-ion batteries, a regulatory battery management system is crucial. It must protect the overall assembly and monitor the optimal energy usage of the battery [4]. Such a system needs to monitor and ensure the proper state of functions

of the battery including its state-of-health (SOH) and state-of-charge (SOC), which, consequently optimize the battery performance [5–7]. As an example; in electrical vehicles, the SOH is analogous to an odometer and SOC mimics the fuel gauge used in fueled vehicles. If SOH and SOC are estimated accurately, they can prevent overcharging and overheating of the battery, and hence extend its life [8]. However, they can not be measured directly by electrical equipment. They are estimated from battery variables, such as its load current and terminal voltage. Many techniques have been developed to estimate SOC and SOH and each has its own advantages and shortfalls [9].

The SOC of a battery shows the currently available charge, which changes over the utility of load current drawn from the battery [10]. Present literature classifies the estimation of SOC into three broad categories; data-driven approach, adaptive methods, and hybrid techniques [11]. The data-driven approach is further classified into two methods; the methods that directly measure the SOC and the methods that are model-based [12]. The direct measurement methods include open-circuit voltage method (OCV) [13], Coulomb counting method (CCM) [14] and electrochemical impedance spectroscopy (EIS) [15]. The OCV provides accurate SOC but requires relaxation time intervals between two cyclic measurements and is prone to more errors when OCV-SOC has a flat curve. The CCM measures the rate of discharging current and integrates the current and time to find SOC [16]. However, this method depends on the accuracy of the current sensor and due to the open-loop nature of CCM, it accumulates error [17]. The last method of the data-driven approach is model-based and it is an online method that takes waveforms as inputs for models that output SOC. Two prominent model-based methods have been widely used; the electrical equivalent model and the electrochemical models. The biggest shortfall of the model-based methods is that they can only be applied on fresh cells for model parameters and are invalid for aged cells [18].

Methods like Kalman filter (KF), fuzzy logic (FL), support vector machines (SVM) and machine learning techniques are categorized in adaptive methods for SOC estimation [19–22]. Amongst these, Kalman filter is the most employed method; which verifies SOC through real-time state estimation with high accuracy. Lastly, hybrid models avail the benefits of both data-driven and adaptive methods such as extended/uncentered Kalman filters.

The SOH quantifies the condition of the battery as compared to the ideal battery. The internal resistance of the battery can be used as a proxy for SOH as it changes over time. However, it is difficult to calculate the internal resistance in real-time. The SOH can be estimated by two methods; the data-driven approach and adaptive methods [23]. In the data-driven approach, the discharging voltage and load current cycles can be used to estimate the SOH [24]. However, this technique requires depth of knowledge and understanding of the correlations between different operations and age of the battery. The accuracy of this method depends on the size of the dataset; a large authentic dataset will yield the best results [25]. Adaptive methods employ only those parameters that contribute in aging. These parameters should be examined carefully as they will be used throughout the operation. There are many techniques to estimate the SOH, including Kalman filter, extended Kalman filter (EKF), fuzzy logic, intelligent algorithms (e.g., genetic algorithm (GA) and artificial neural networks (ANN)), least-squared polynomial regression (LSPR), and particle filters. For prognosis of battery, EKF is commonly used and is easy to implement. Online estimation for the SOH and SOC can be done using recursive least-squares with electromotive force. Battery parameters, diffusion capacitance and the SOH can be estimated using GA [26,27]. The discharging waveform are mapped by using least squared polynomial regression (LSPR), ANNs are used for mapping features and particle filters are used for prediction. Such techniques provide good efficiency but are computationally expensive in terms of their pre-processing stages. The compounded uncertainties can raise beyond control in estimation, if not carefully managed [28].

In [28], the authors have estimated the end of discharge (EOD) by mapping discharge curves over load current using LSPR and ANN to predict the SOH for different loading profiles. However, many assumptions have been made, i.e., predetermination of the slope values, and lengthened mapping processes for prediction of EOD.

This paper aims to overcome these assumptions by introducing an improved method for SOH estimation. The main contributions can be summarized as follows:

- Accurate definition of knee points in the discharge curves i.e., initial voltage drop and threshold value.
- Implementation of a simplified machine learning model for SOH estimation. The proposed trained models have the ability to accurately predict and estimate results for different current
- Estimating the SOC using the discharge curve of the battery.

2. Background on Battery Characteristics

A battery is a converter or transducer of chemical energy into electrical energy, and vice versa. It consists of an anode and a cathode (i.e., electrodes) immersed in an electrolyte solution. An electrical potential is created in the battery due to the difference in electric charge between the two electrodes in a cell, which is determined by the product of the reaction and the reactants (i.e., standard Gibbs Free energies) [29]. The data has batteries whose OCV reading is observed at 297.15 K and 277.15 K, and at 1 atmospheric pressure. This OCV is not available at the terminal when the battery is in use, due to the voltage drop by different processes, e.g., resistive forces inside the battery (electrolyte), terminal leads, and potential drop [30].

It should be noted that these parameters of the battery are current dependable and can only be observed as the current is drawn from the battery. This curve changes over time for a given current level. Figure 1 shows an example for a current of 2.0 A.

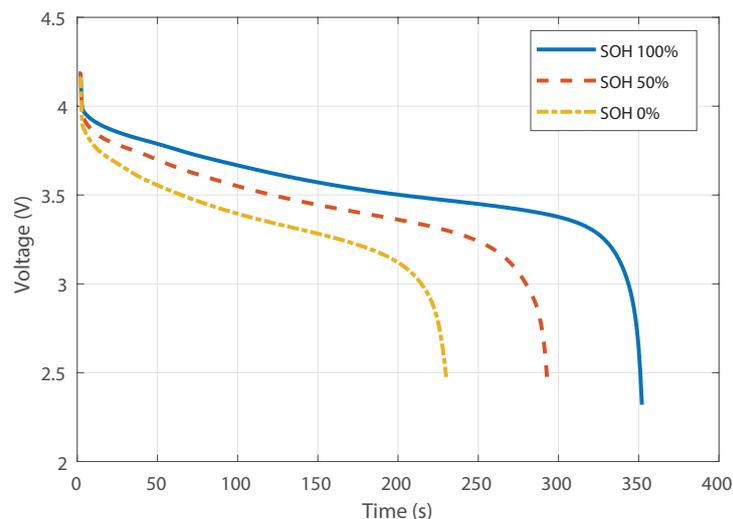


Figure 1. Battery discharge curves at constant load current of 2.0 A.

Since the current drawn provides such an important understanding about the losses inside the batteries, it helps in quantifying the performance of the battery. In literature, the current drawn from the battery is related to the commonly used term C-rate. The C-rate (C/r) is defined as the amount of current drawn in a single discharge cycle of the battery for a given number of hours with respect to its nominal capacity [31]. If a battery has a C-rate of $C/10$ (0.1 A) for a battery of 1.0 Ah, then this battery will be able to provide charge for 10 h. The amount of energy delivered by a battery and area under the discharge curve also rely on the C-Rate. Figure 2 shows different current, i.e., 1.0 A and 2.0 A, and for different temperatures, i.e., 297.15 K and 277.15 K.

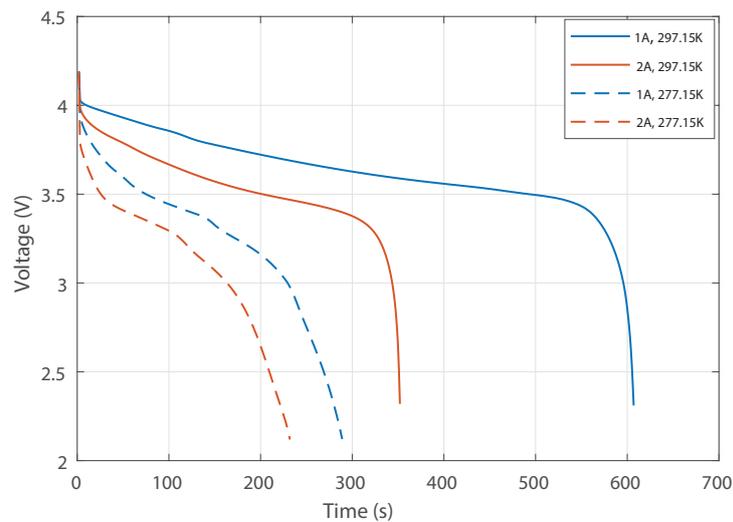


Figure 2. Discharge curves for different current, (i.e., 1.0 A and 2.0 A) and for different temperatures (297.15 K and 277.15 K).

There are other parameters which represent the physical properties of batteries, and based on such parameters, different batteries can be used for different applications e.g., grid-scale applications and portable electronic devices. Such parameters include specific energy, specific power, and life cycle. Specific energy is defined as energy per unit weight. Energy stored per unit volume is called energy density. This is an important parameter for the application of portable electronic devices where long-lasting charge is required. Specific power or power per volume is defined as the power density, and is important when sudden charge/discharge is required such as in EVs brakes [32]. Lastly, battery life cycle defines how many charging/discharging cycles a battery can provide before it depletes beyond acceptable limits.

3. Battery Modeling Approach

There are several models that have been used to represent internal characteristics of Li-ion batteries, e.g., black, gray and white box models [2]. The first principal method is categorized as a white-box model; developed with internal battery dynamics such as intercalation kinetics, ion diffusions, and electric potentials. However, the internal electrochemical reactions can be very dull and computationally intractable due to partial-differential equations [33]. Another approach is to use algorithms which are merged with low-order modeling and rely on the slight sacrifice of precision and physical representation [34]. An alternative method is data-driven; it is a model-free approach where machine learning techniques are used. This is a gray/black box model also called the electrical equivalent circuit model [35]. In this model, internal characteristics of a battery are modeled in terms of resistors, capacitors, and inductors. We use the lumped model with approximation of IR drop i.e., the voltage drop caused by electrolyte resistance represented as R_o . The activation polarization is denoted as a charge transfer resistance R_p and C_p is the dual layer capacitance in parallel with R_p .

Mathematical Model of a Battery

The battery characteristics can be identified using the Randles models [36]. The generalized Randles models have also been used in bio-technologies where a combination of RC and voltage potential is modeled for electrode-tissues interface (ETI) in electroencephalography [37]. In [36], authors have explored the identifiability of Randles models; consisting of [38,39]:

- R_o representing an ohmic resistance due to the conduction of charge carriers through electrolyte and metallic conduction.

- a series of parallel resistor and capacitor connections (i.e., activation polarization) representing the charge transfer resistance and double layer capacitance respectively.

The number of RC parallel circuits defines the order of the circuit model [40].

The electrical behavior of a dual polarization (DP) circuit is represented by following equations [17]:

$$\frac{dU_{p1}}{dt} = \frac{I_t}{C_{p1}} - \frac{U_{p1}}{R_{p1}C_{p1}} \quad (1)$$

$$\frac{dU_{p2}}{dt} = \frac{I_t}{C_{p2}} - \frac{U_{p2}}{R_{p2}C_{p2}} \quad (2)$$

The output equation in the case of dual polarization (DP) model is given by [17]:

$$y = U_{oc}(SOH) - U_{p1} - U_{p2} - R_o \times I_t, \quad (3)$$

where $U_{oc}(SOC)$ is the internal voltage of the battery, and U_{p1} and U_{p2} are the voltages across the two RC parallel branches, respectively.

4. Machine Learning for SOH and SOC

Machine learning algorithms are employed here for robust estimation of the battery SOH and SOC. The goal is to calculate the SOH and SOC of a particular battery using a trained regression model, given the battery's discharge curve and the current voltage-of-operation. The SOH is determined using a machine learning algorithm trained on the discharge curves provided by the NASA Ames prognostic dataset [41], which is shown in Table 1 for different current levels (i.e., 1.0 A, 2.0 A) and for different temperature levels, (i.e., 297.15 K and 277.15 K).

Table 1. Batteries with their current and temperature levels.

S.No	Battery	Current	Temperature
1	B0005	2.0 A	297.15 K
2	B0006	2.0 A	297.15 K
3	B0007	2.0 A	297.15 K
4	B0045	1.0 A	277.15 K
5	B0046	1.0 A	277.15 K
6	B0047	1.0 A	277.15 K

This model looks at a discharge curve of the battery and determines its SOH percentage relative to a reference battery state. This healthy state of the battery is assumed to be the state at the first discharge in the dataset. The SOC at a particular voltage-of-operation (V_o) is determined by utilizing the nature of the discharge curve. The battery has been discharged from a maximum (V_{max}) to a minimum (V_{min}) voltage. The SOC is estimated by calculating the area under the discharge curve from V_o to V_{min} and dividing this area by the total area under the curve. This provides a quantitative percentage measure of how much charge is left in the battery at a particular voltage-of-operation.

This SOC is then plotted against SOH to get a one-to-one mapping. The end-result is the estimation of SOC given SOH at a specific voltage-of-operation. The following subsection provides more details on the evaluation of SOH and SOC (as shown in block diagram for SOH and SOC Estimation in Figure 3).

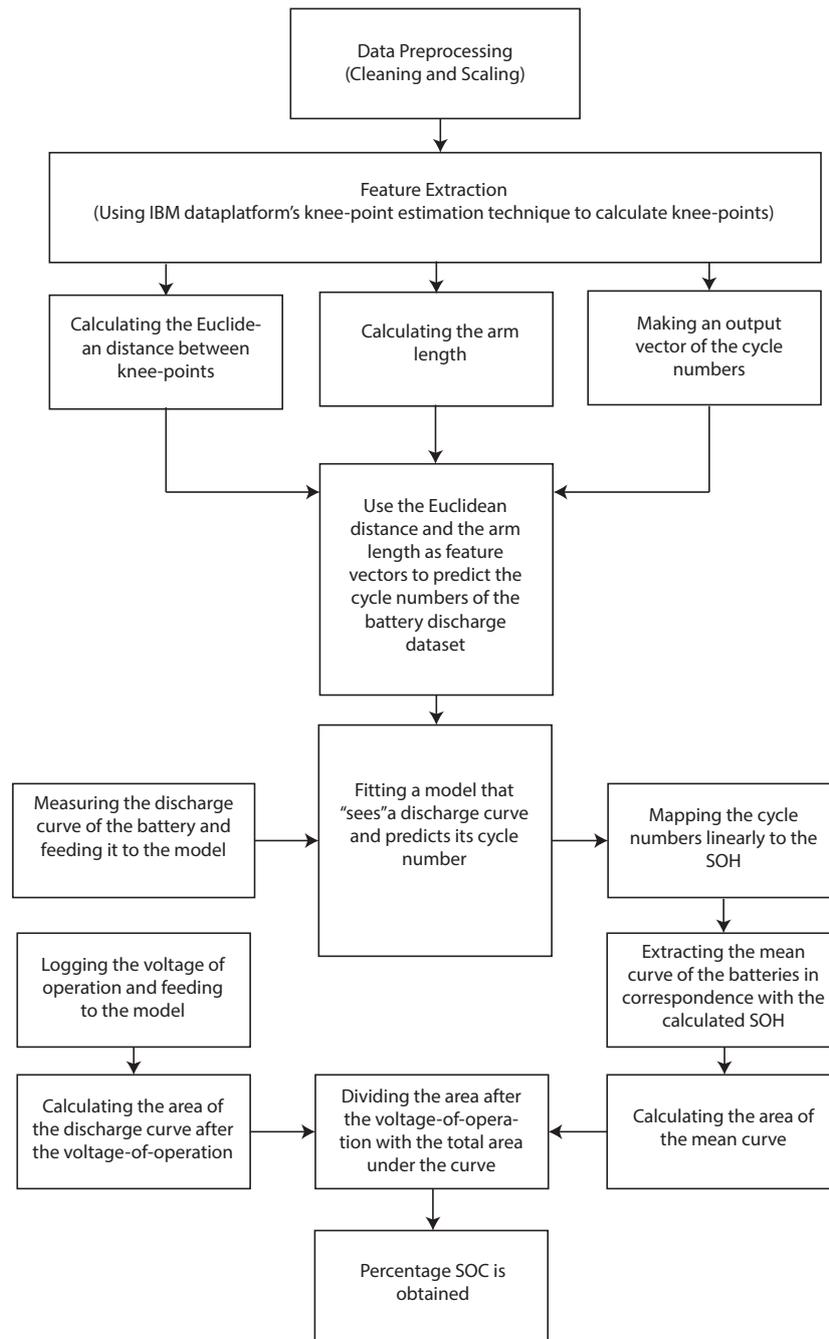


Figure 3. Block diagram of SOH and SOC estimation with machine learning.

4.1. Knee-Point Calculation

We first convert the discharge voltage vectors into measurable and logical features for SOH estimation. An exemplary discharge curve is shown in Figure 4, where points A and B show the maximum voltage (V_{max}) and minimum voltage (V_{min}) of the discharge curve, respectively. Points C and D show the two knee-points of the curve. The relative health of the battery decreases with a decrease in the distance between the two knee-points. Hence, the first input parameter becomes the pseudo linear region between the two knee-points of the discharge curve.

An extended version of the algorithm proposed in the IBM data platform [42] is utilized here for the knee point estimation. Specifically, the calculation of the knee points involves two steps; first the equation of the line from the highest voltage V_{max} to the lowest voltage V_{min} of the curve is determined i.e., the straight line passing through points A and B. Then, the largest perpendicular distance from

the line AB to the discharge curve is identified. The longest perpendicular distance will represent one of the knee-points of the curve denoted as D . The second knee point C can be determined using a similar approach by applying the algorithm in a region between the start of the curve and the point of intersection between the graph and line AB .

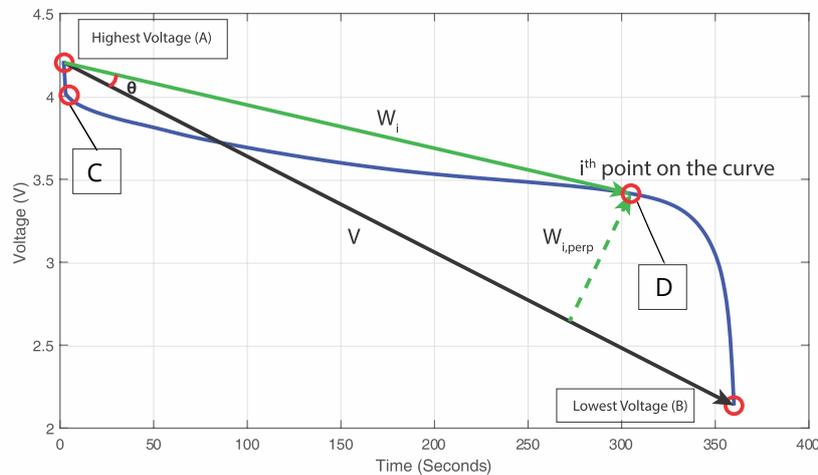


Figure 4. Knee points calculation.

To calculate the perpendicular distance from the line to the graph, the dot product is used. Let $AB = V$ and let there be any N points sampled from the discharge curve. Note that the data set already contains N points for each curve of battery discharge. Let $\mathbb{N} = \{n_1, \dots, n_N\}$ be the set of all points on the curve such that $|\mathbb{N}| = N$ where $|\cdot|$ denotes the cardinality of the set \mathbb{N} . For each $n_i, i \in \{1, 2, 3, \dots, N\}$, a vector W_i is drawn from point A such that the set of vectors is denoted by $\mathbb{W} = \{W_1, \dots, W_i, \dots, W_N\}$. For each W_i , a dot product between V and W_i is taken such that the angle between them is given as:

$$\theta_i = \cos^{-1} \left(\frac{V \cdot W_i}{|V| |W_i|} \right), \quad (4)$$

where $|\cdot|$ denotes the magnitude operator. The value of the perpendicular distance in terms of the single V, W_i pair is calculated by,

$$W_{i,perp} = W_i \sin \theta_i \quad (5)$$

The knee-point D , is then found by taking the maxima, i.e.,

$$D = \max_i (W_{1,perp}, W_{2,perp}, \dots, W_{N,perp}) \quad (6)$$

For calculating the other knee-point C , the range of the search has to be reduced. This is done by finding the intersection of line AB and the discharge curve, and setting the upper limit of the search equal to the intersection point. Thus, N reduces to a lower value.

4.2. Feature Engineering

An important step in machine learning is the preparation of the dataset to fit our needs. After logging the discharge voltage and time for points A, B, C and D of the battery dataset, the features that would be valuable input parameters in the machine learning model, are extracted. In order to extract these features for the training of the machine learning model, the Euclidean distances d_{AC} , d_{CD} and d_{DB} are calculated. Note that the relative health of the battery decreases with a decrease in the distance between the two knee-points. Hence, the first input parameter becomes the Euclidean distance between the two knee-points of the discharge curve (i.e., d_{CD}). The line between points C and D is termed the pseudo-linear region due to its linear nature. Now, after determining the first

input parameter in the form of the pseudo-linear region, the lengths AC and DB are analyzed as other input parameters to the ML model. The input parameters to the ML algorithm are chosen to be linearly independent to cater for redundancy. The Euclidean distance d_{CD} is identified as a significant feature, therefore other features must be checked if they are linearly independent from the first feature or not. Figure 5 represents the relationship between d_{CD} and the other two candidate features such as d_{AC} and d_{DB} for one of the batteries. As d_{DB} provides a more non-linear relationship to d_{CD} than d_{AC} , therefore, the second input feature is selected as d_{DB} . The length d_{CD} is called as *arm length*, as it is the length of the offshoot discharge arm of the battery voltage curve.

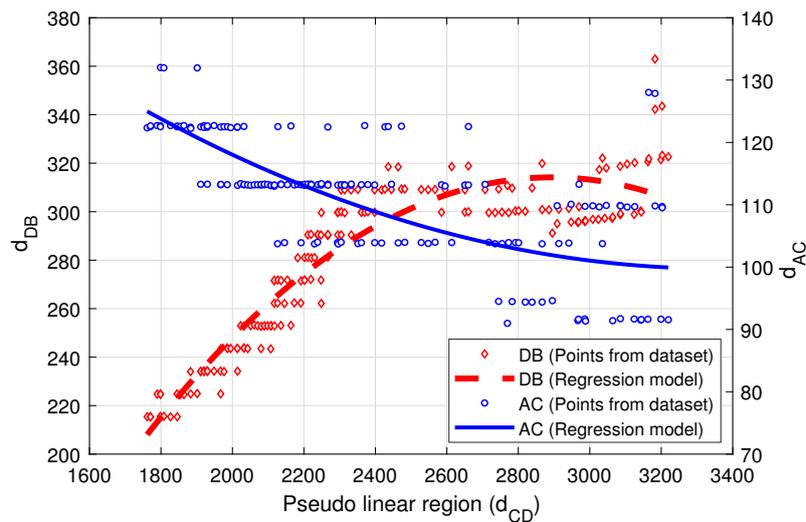


Figure 5. Relationship between d_{CD} and the other two features such as d_{AC} and d_{DB} for one of the batteries.

4.3. Data Preprocessing

The features extracted from each of the discharge curves is used to create the input data matrix, as:

$$\mathbf{X} = [x_1, x_2] \quad (7)$$

where x_1 and x_2 are the $(D \times 1)$ vectors representing the extracted features from the discharge curves if D is the number of discharge curves of a battery. Our dataset contains the discharge curves of a battery in a sequential order. The first discharge curve is assumed to correspond to a healthy battery. After one cycle, the next curve is obtained by charging and discharging this battery again. Hence the n th discharge curve is taken on a battery that has undergone n charges and $(n - 1)$ discharges; consequently losing some of its health. We use this characteristic of the dataset to create the output vector of the ML model. The final form of the dataset is:

$$\Delta = [\mathbf{X}, \mathbf{y}], \quad (8)$$

where each element of \mathbf{y} denotes the cycle number corresponding to x_1, x_2 features.

4.4. Machine Learning

This dataset is fed into the ML model to determine the best fit. The root mean squared error (RMSE) of each model is calculated to analyze the authenticity and feasibility of the model. It is concluded that a polynomial regression model of 3rd degree provides the best fit with the least RMSE. The hypothesis equation utilized for regression is given below:

$$h_{\theta}(x_1, x_2) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1 x_2 + \theta_5 x_1^3 + \theta_6 x_1^2 x_2 \quad (9)$$

The values of the parameters are determined by using the gradient descent algorithm. This algorithm quantitatively compares the predicted value of cycle number from the model to the ground truth y of the dataset. The following equation calculates the RMSE of the ML model;

$$J(\theta) = J(\theta_0, \dots, \theta_6) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x_1^{(i)}, x_2^{(i)}) - y_m)^2, \quad (10)$$

where J is the cost or error function quantifying the RMSE of the model, m is the size of the dataset and θ is the vector containing the model parameters $\theta_0, \dots, \theta_6$. To minimize this error, we must tweak the model parameters and find the optimum values recursively. To get the new values of the model parameters θ_{n+1} , we take the partial derivatives of the error term with respect to each of the parameters in vector θ_n and subtract this derivative term from θ_n .

$$\theta_{n+1} := \theta_n - \alpha \frac{\partial}{\partial \theta_n} J(\theta_0, \dots, \theta_6) \quad (11)$$

This process is then repeated with the new parameter values to minimize the error function and push it closer to zero. An ideal cost function will have a value of 0.

The first discharge cycle will correspond to a battery health of 100% and the last discharge cycle will provide the discharge curve of the least “healthy” battery.

4.5. SOC Estimation

The ML model uses the entire discharge curve to estimate the SOH of the battery. Now the current voltage-of-operation is provided as another parameter to the algorithm. The goal is to provide an estimated amount of charge, that is left in the battery at a certain voltage-of-operation (V_a). To do so, the curve from the maximum V_o to the minimum V_{min} is used. As provided in the NASA Ames prognostic repository, the battery has been discharged till +2.7 V. Hence, only part of the curve up to that point will be used.

After extracting the voltage curve from V_{max} to +2.7 V, the area A_T under the curve is calculated. Now, the given voltage V_o , which is the current voltage-of-operation of the battery, is used. The total charge remaining in a battery of a specific SOH, at a particular V_o is calculated by dividing the area under the curve after V_o to +2.7 V, by the total area of the curve. This gives an estimate of the percentage of charge left in the battery at a particular voltage level. The formula for SOC% calculation is given by:

$$SOC\% = \frac{\int_{V_o}^{+2.7V} V(t) dt}{A_T}, \quad (12)$$

where $V(t)$ is the discharge curve function with respect to time and the nominator term represents the area from V_o to +2.7 V.

5. Results and Discussions

In this section, the method is implemented on six batteries with different current and temperature levels. The first three batteries are modeled with a current level of 1.0 A and a temperature of 297.15 K. Similarly, the other three batteries are modeled with a current level of 2.0 A and a temperature of 277.15 K. Note that each battery has different cycles, therefore, each battery represents a different trend when plotted against the pseudo-linear region (d_{CD}) and arm length (d_{DB}).

The discharge curve labeling has been implemented by knee point calculation. This step accurately specifies the threshold voltages for feature estimation. d_{CD} and d_{DB} are extracted as features since d_{CD} represents the capacity of the battery and d_{DB} provides remaining 10% of charge with EOD point. This not only helps in finding the SOH, but also the SOC at any instance of time. Moreover, this algorithm could have also estimated SOH of batteries at higher temperature and load current levels, given a larger dataset. Nevertheless, implementing the proposed algorithm using 277.15 K

and 297.15 K temperature and 1.0 A and 2.0 A current levels, respectively are sufficient verification of this method.

The datasets of batteries are divided into two major portions; training and testing. Such classification of data is important to verify the performance of the developed models. In doing so, our classification is purely done randomly i.e., out of all the discharge curves, 70% of them are randomly chosen for training and remaining 30% of the discharge curves are dedicated to the testing of the model.

First, the results showing the relationship between the two extracted features of a battery curve versus the cycle number or SOH are presented. Note that the dataset contains the cycle number of each discharge curve, whereas the battery health is generally taken in terms of SOH. Therefore, a mapping between cycle number and SOH is performed, such that the greater the cycle number the lesser the SOH.

Figure 6–8 represent the contour plots of the models used for the estimation of the SOH of the batteries at 297.15 K. The training dataset used in each case is shown by points on the graph. It can be inferred from the graph that with the increase in the d_{CD} and d_{DB} the SOH also increases. It should be noted that the SOH is given in the form of a percentage metric but the contour lines go above 100% on the right side of the contour plots. The reason behind this apparent anomaly is that the ML model predicts the SOH without any regard to the nature of the SOH. Thus, no input features are ever provided that take the SOH above 100%. For this model, SOH is just a numeric value though there are no “real” data points in the above 100% region. Similarly, Figures 9–11 show contour plots of the models estimating the SOH at 277.15 K. Note that, in the contour plot of each of the battery, the blue circles show the cycles, whereas the lines correspond to SOH. Similarly, if these points are connected together a 3rd polynomial trend can be obtained as represented by Equation (9).

The increase in the SOH of the battery causes the overall area of the voltage curve to increase. This increased area is in the denominator of the term that calculates the percentage SOC and henceforth, decreases the SOC. This percentage decreases much more rapidly as compared to a healthy battery because of its reduced capacity. Thus, it can be concluded, for a particular value of the voltage, the SOC of the battery increases with the decrease in health, i.e., the battery loses charge faster when it has a low SOH. The overall curve is uneven and has higher peaks at the ends. The main reason for such abnormality is due to the sparsity of the dataset. The inclusion of such result may depict rough estimation. However, it is very important to draw a relationship between SOH and SOC together for a particular voltage, which not only broadens the understanding of these two parameters but also introduces a correlation between them.

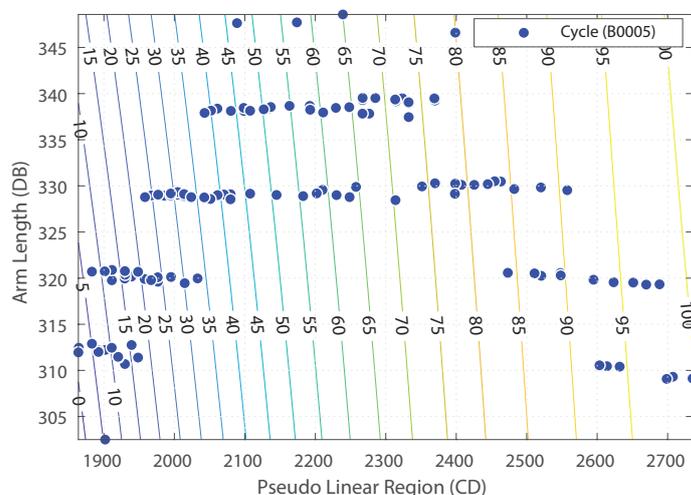


Figure 6. Contour plot showing the effect of cycle number on Arm Length and Pseudo Linear Region at 297.15 K.

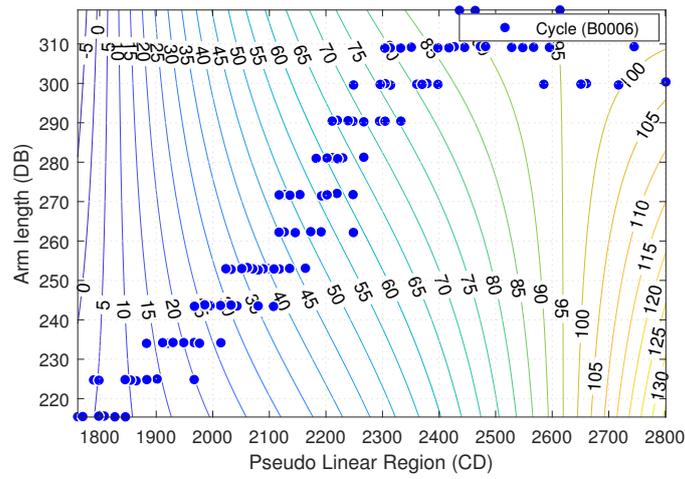


Figure 7. Contour plot showing the effect of cycle number on Arm Length and Pseudo Linear Region at 297.15 K.

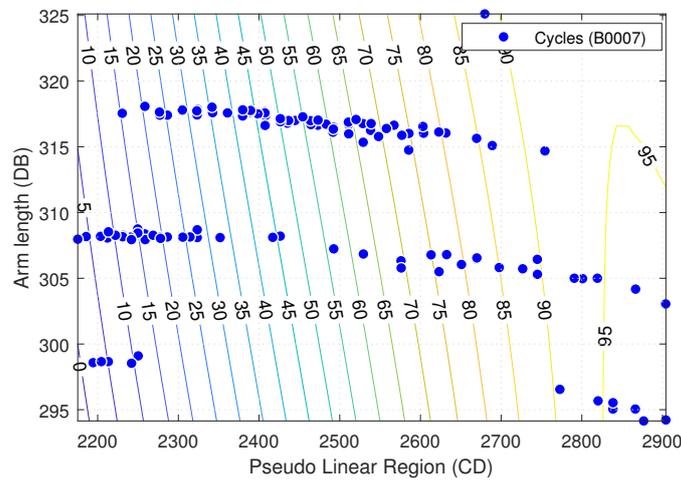


Figure 8. Contour plot showing the effect of cycle number on Arm Length and Pseudo Linear Region at 297.15 K.

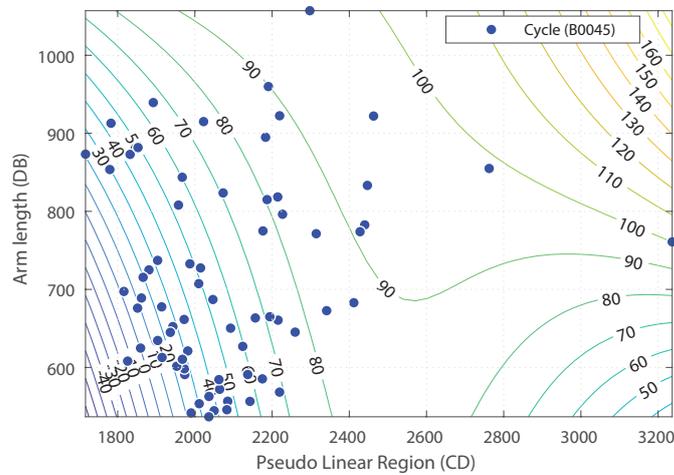


Figure 9. Contour plot showing the effect of cycle number on Arm Length and Pseudo Linear Region at 277.15 K.

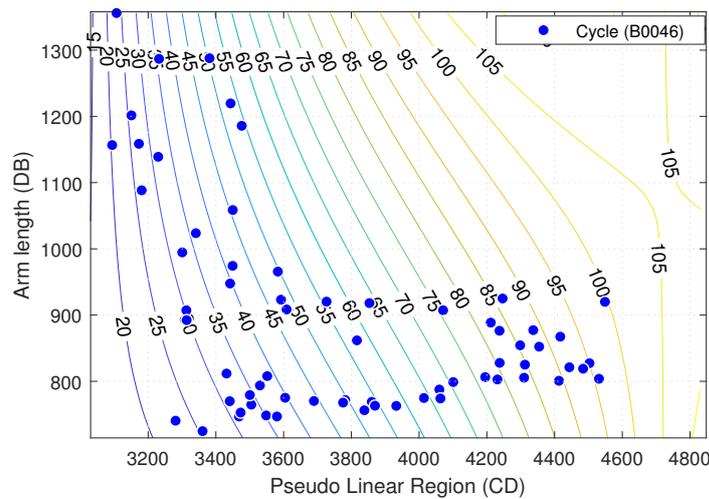


Figure 10. Contour plot showing the effect of cycle number on Arm Length and Pseudo Linear Region at 277.15 K.

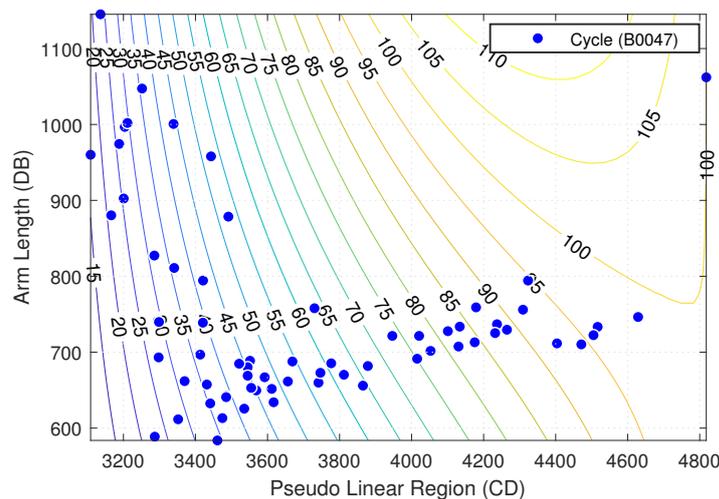


Figure 11. Contour plot showing the effect of cycle number on Arm Length and Pseudo Linear Region at 277.15 K.

Finally, it is worthwhile noting that the accuracy of the three batteries (B0045-B0047) with a temperature of 277.15 K and a current of 2.0 A, has lower values (87.86% to 77.87%) in comparison to that of 297.15 K and 1.0 A batteries (93.08% to 98.00%). This is mainly due to the small size of datasets and larger gaps between adjacent discharge curves, which increase the error between training versus testing results. Such an error can be minimized with larger and more precise datasets. The accuracy shown in Figure 12 is calculated by:

$$Accuracy = 100 - \frac{1}{N} \left[\sum_{i=1}^{N_{test}} \frac{SOH_i - SOH_i^{(t)}}{SOH_i^{(t)}} \right] \times 100 \tag{13}$$

where, N_{test} is the number of discharge cycles of the test dataset, SOH_i is the i_{th} term of estimated SOH, and $SOH_i^{(t)}$ is the i_{th} true value of discharge curve driven by the training dataset.

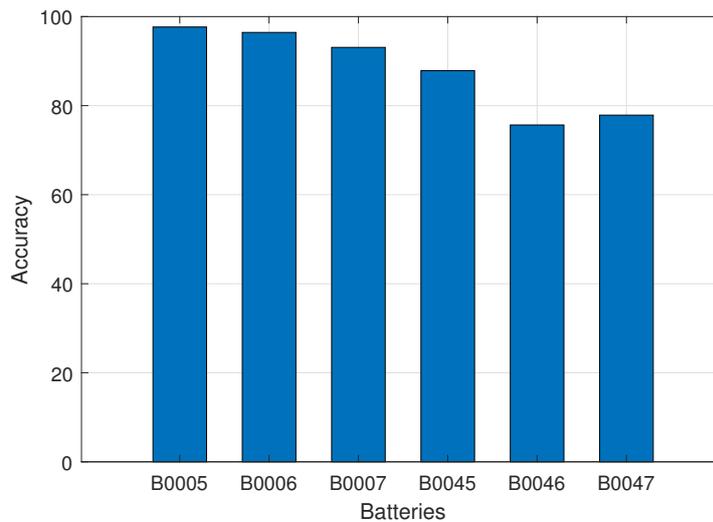


Figure 12. Accuracy results of training versus testing dataset of batteries.

6. Conclusions

This work summarized some of the important contributions for battery management systems (BMS). In an initial stage of this research, important findings are made in the estimation of SOH and SOC using the ML approach. As the behavior of the battery is non-linear and cannot be predicted directly, therefore, an indirect estimation and prediction method is needed. In the estimation of battery SOH and SOC, several features were extracted from discharge curves of different batteries with current and temperature levels, i.e., B0006 is a battery with 2.0 A and 297.15 K, whereas, B0047 is a battery with 1.0 A and 277.15 K. During the extraction of features, a knee-point calculation technique was implemented, which identified knee-points numerically. The calculated features are fed to the ML model, which provide SOH over all discharge cycles of the battery with respect to the pseudo-linear region and arm length. In addition, the SOC% was estimated by considering the area under the curve of the measured voltage. This technique has not only estimated the correct SOC% for a discharge cycle, but has also provided a decreasing trend with respect to all training cycles. In other words, if we are given a battery at any instant of time, we can estimate its SOC by considering its voltage as a reference point and using reverse engineering to calculate the SOH.

However, due to a limited dataset, the obtained error was significant, which can be minimized in with the use of larger datasets. Despite the fact that the Polynomial regression models stand out over other ML techniques as they provide the best RMSE, this might not be possible with large datasets. Therefore, better methods such as ANN and support vector machines can be implemented in the future. Finally, since the proposed model needs the entire discharge curve to calculate the SOH and SOC of the battery, and with a larger dataset, recurrent neural networks might be a better choice for online estimation of SOH.

Author Contributions: Conceptualization, S.S.S.; methodology, S.S.S., M.A., M.A.K. and S.A.H.; formal analysis, S.S.S. and M.A.; writing—original draft preparation, S.S.S., M.A., M.A.K. and S.A.H.; writing—review and editing, S.A.H., H.A.K., A.G. and L.B.-B.; visualization, H.A.K. and S.A.H.; supervision, H.A.K. and S.A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: Authors thank NASA Ames prognostics for providing open source battery data repository for advancing this domain knowledge and also thank George Alberto Salazar, Human Computer Interface Technical Discipline Lead at Johnson Space Center NASA for his support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dunn, B.; Kamath, H.; Tarascon, J. Electrical Energy Storage for the Grid: A Battery of Choices. *Science* **2011**, *334*, 928–935. [[CrossRef](#)] [[PubMed](#)]
2. Hu, X.; Zou, C.; Zhang, C.; Li, Y. Technological Developments in Batteries: A Survey of Principal Roles, Types, and Management Needs. *IEEE Power Energy Mag.* **2017**, *15*, 20–31. [[CrossRef](#)]
3. Kim, T.; Song, W.; Son, D.; Ono, L.; Qi, Y. Lithium-ion batteries: Outlook on present, future, and hybridized technologies. *J. Mater. Chem.* **2019**, *7*, 2942–2964. [[CrossRef](#)]
4. Stewart, P.; Bingham, C. Electrical Power and Energy Systems for Transportation Applications. *Energies* **2016**, *9*, 545. [[CrossRef](#)]
5. Kim, Y.; Hwang, C.; Kim, E.; Cho, C. State of Charge-Based Active Power Sharing Method in a Standalone Microgrid with High Penetration Level of Renewable Energy Sources. *Energies* **2016**, *9*, 480. [[CrossRef](#)]
6. Ren, H.; Zhao, Y.; Chen, S.; Wang, T. Design and implementation of a battery management system with active charge balance based on the SOC and SOH online estimation. *Energy* **2019**, *166*, 908–917. [[CrossRef](#)]
7. Tang, X.; Wang, Y.; Zou, C.; Yao, K.; Xia, Y.; Gao, F. A novel framework for Lithium-ion battery modeling considering uncertainties of temperature and aging. *Energy Convers. Manag.* **2019**, *180*, 162–170. [[CrossRef](#)]
8. Wei, Z.; Leng, F.; He, Z.; Zhang, W.; Li, K. Online State of Charge and State of Health Estimation for a Lithium-Ion Battery Based on a Data–Model Fusion Method. *Energies* **2018**, *11*, 1810. [[CrossRef](#)]
9. Zhang, M.; Fan, X. Review on the State of Charge Estimation Methods for Electric Vehicle Battery. *World Electr. Veh. J.* **2020**, *11*, 23. [[CrossRef](#)]
10. He, H.; Xiong, R.; Fan, J. Evaluation of Lithium-Ion Battery Equivalent Circuit Models for State of Charge Estimation by an Experimental Approach. *Energies* **2011**, *4*, 582–598. [[CrossRef](#)]
11. Watrin, N.; Blunier, B.; Miraoui, A. Review of adaptive systems for lithium batteries State-of-Charge and State-of-Health estimation. In Proceedings of the 2012 IEEE Transportation Electrification Conference and Expo (ITEC), Dearborn, MI, USA, 18–20 June 2012.
12. Li, C.; Xiao, F.; Fan, Y. An Approach to State of Charge Estimation of Lithium-Ion Batteries Based on Recurrent Neural Networks with Gated Recurrent Unit. *Energies* **2019**, *5*, 1592. [[CrossRef](#)]
13. Xing, Y.; He, W.; Pecht, M.; Tsui, K.L. State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures. *Appl. Energy* **2014**, *113*, 106–115. [[CrossRef](#)]
14. Ng, K.S.; Moo, C.S.; Chen, Y.P.; Hsieh, Y.C. Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries. *Appl. Energy* **2009**, *86*, 1506–1511. [[CrossRef](#)]
15. Andre, D.; Meiler, M.; Steiner, K.; Walz, H.; Soczka-Guth, T.; Sauer, D.U. Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. II: Modelling. *J. Power Sources* **2011**, *196*, 5349–5356. [[CrossRef](#)]
16. Roscher, M.A.; Sauer, D.U. Dynamic electric behavior and open-circuit-voltage modeling of LiFePO₄-based lithium ion secondary batteries. *J. Power Sources* **2011**, *196*, 331–336. [[CrossRef](#)]
17. Khalid, M.; Sheikh, S.S.; Janjua, A.K.; Khalid, H.A. Performance validation of electric vehicle's battery management system under state of charge estimation for lithium-ion battery. In Proceedings of the 2018 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), Quetta, Pakistan, 12–13 November 2018; pp. 1–5. [[CrossRef](#)]
18. Xiong, R.; Sun, F.C.; He, H.W. Data-driven State-of-Charge estimator for electric vehicles battery using robust extended Kalman filter. *Int. J. Automot. Technol.* **2014**, *15*, 89–96. [[CrossRef](#)]
19. Lee, J.; Nam, O.; Cho, B.H. Li-ion battery SOC estimation method based on the reduced order extended Kalman filtering. *J. Power Sources* **2007**, *174*, 9–15. [[CrossRef](#)]
20. Talha, M.; Asghar, F.; Kim, S.H. A Neural Network-Based Robust Online SOC and SOH Estimation for Sealed Lead–Acid Batteries in Renewable Systems. *Arab. J. Sci. Eng.* **2018**, *44*, 1869–1881. [[CrossRef](#)]
21. Anton, J.C.A.; Nieto, P.J.G.; Viejo, C.B.; Vilán, J.A.V. Support Vector Machines Used to Estimate the Battery State of Charge. *IEEE Trans. Power Electron.* **2013**, *28*, 5919–5926. [[CrossRef](#)]
22. Huang, S.C.; Tseng, K.H.; Liang, J.W.; Chang, C.L.; Pecht, M.G. An Online SOC and SOH Estimation Model for Lithium-Ion Batteries. *Energies* **2017**, *10*, 512. [[CrossRef](#)]
23. Vignarooban, K.; Chu, X.; Chimatapu, K.; Ganeshram, P.; Pollat, S.; Johnson, N.G.; Kannan, A.M. State of health determination of sealed lead acid batteries under various operating conditions. *Sustain. Energy Technol. Assess.* **2016**, *18*, 134–139. [[CrossRef](#)]

24. Ma, Z.; Yang, R.; Wang, Z. A novel data-model fusion state-of-health estimation approach for lithium-ion batteries. *Appl. Energy* **2019**, *237*, 836–847. [CrossRef]
25. Fang, Q.; Wei, X.; Lu, T.; Dai, H.; Zhu, J. A State of Health Estimation Method for Lithium-Ion Batteries Based on Voltage Relaxation Model. *Energies* **2019**, *12*, 1349. [CrossRef]
26. Chen, Z.; Mi, C.C.; Fu, Y.; Xu, J.; Gong, X. Online battery state of health estimation based on Genetic Algorithm for electric and hybrid vehicle applications. *J. Power Sources* **2013**, *240*, 184–192. [CrossRef]
27. Coleman, M.; Lee, C.K.; Zhu, C.; Hurley, W.G. State-of-Charge Determination From EMF Voltage Estimation: Using Impedance, Terminal Voltage, and Current for Lead-Acid and Lithium-Ion Batteries. *IEEE Trans. Ind. Electron.* **2007**, *54*, 2550–2557. [CrossRef]
28. Saxena, A.; Celaya, J.R.; Roychoudhury, I.; Saha, S.; Saha, B.; Goebel, K. Designing data-driven battery prognostic approaches for variable loading profiles: Some lessons learned. In Proceedings of the First European Conference of Prognostics and Health Management Society, PHM 2012, Dresden, Germany, 3–5 July 2012; pp. 72–732.
29. Zhang, Y.; Shang, Y.; Cui, N.; Zhang, C. Parameters Identification and Sensitive Characteristics Analysis for Lithium-Ion Batteries of Electric Vehicles. *Energies* **2017**, *11*, 19. [CrossRef]
30. McKissock, B.; Loyselle, P.; Vogel, E. Guidelines on Lithium-ion Battery Use in Space Applications. NASA, National Aeronautics and Space Administration. 2009. Available online: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20090023862.pdf> (accessed on 12 May 2020).
31. Lee, K.T.; Dai, M.J.; Chuang, C.C. Temperature-Compensated Model for Lithium-Ion Polymer Batteries with Extended Kalman Filter State-of-Charge Estimation for an Implantable Charger. *IEEE Trans. Ind. Electron.* **2018**, *65*, 589–596. [CrossRef]
32. Rucker, F.; Bremer, I.; Linden, S.; Badedo, J.; Sauer, D.U. Development and Evaluation of a Battery Lifetime Extending Charging Algorithm for an Electric Vehicle Fleet. *Energy Procedia* **2016**, *99*, 285–291. [CrossRef]
33. Gao, L.; Liu, S.; Dougal, R.A. Dynamic lithium-ion battery model for system simulation. *IEEE Trans. Components Packag. Technol.* **2002**, *25*, 495–505. [CrossRef]
34. Fleischer, C.; Waag, W.; Heyn, H.M.; Sauer, D.U. On-line adaptive battery impedance parameter and state estimation considering physical principles in reduced order equivalent circuit battery models part 2. Parameter and state estimation. *J. Power Sources* **2014**, *262*, 457–482. [CrossRef]
35. Zhang, X.; Zhang, W.; Lei, G.A. A Review of Li-ion Battery Equivalent Circuit Models. *Trans. Electr. Electron. Mater.* **2016**, *17*, 311–316. [CrossRef]
36. Alavi, S.M.; Mahdi, A.; Jacob, P.E.; Payne, S.J.; Howey, D.A. Structural identifiability analysis of fractional order models with applications in battery systems. *arXiv* **2015**, arXiv:1511.01402.
37. Mihajlovic, V.; Grundlehner, B.; Vullers, R.; Penders, J. Wearable, Wireless EEG Solutions in Daily Life Applications: What are we Missing? *IEEE J. Biomed. Health Inform.* **2015**, *19*, 6–21. [CrossRef] [PubMed]
38. Macdonald, J.R.; Johnson, W.B. Chapter 1—Fundamentals of impedance spectroscopy. In *Impedance Spectroscopy: Theory, Experiment, and Applications*; John Wiley and Sons, Inc.: New Jersey, NJ, USA, 2018; pp. 1–20. [CrossRef]
39. Buller, S.; Thele, M.; De Doncker, R.W.; Karden, E. Impedance-Based Simulation Models of Supercapacitors and Li-Ion Batteries for Power Electronic Applications. *IEEE Trans. Ind. Appl.* **2005**, *41*, 742–747. [CrossRef]
40. Westerhoff, U.; Kroker, T.; Kurbach, K.; Kurrat, M. Electrochemical impedance spectroscopy based estimation of the state of charge of lithium-ion batteries. *J. Energy Storage* **2016**, *8*, 244–256. [CrossRef]
41. NASA-Battery Data-Set, National Aeronautics and Space Administration. Available online: <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/battery> (accessed on 12 May 2020).
42. Curve Knee Estimation, IBM-Watson. Available online: <https://dataplatform.cloud.ibm.com/analytics/notebooks/5\4d79c2a-f155-40ec-93ec-ed05b58afa39/> (accessed on 12 May 2020).

