

Research papers

State of health estimation method for lithium-ion batteries using incremental capacity and long short-term memory network



Zhaopu Zhang^a, Haitao Min^a, Hangang Guo^a, Yuanbin Yu^a, Weiyi Sun^{a,*}, Junyu Jiang^a, Hang Zhao^b

^a State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130022, China

^b Vehicle Components Department, Changchun Automobile Test Center, Changchun 130011, China

ARTICLE INFO

Keywords:

Batteries
State of health
Incremental capacity
Health feature variables
Long short-term memory network

ABSTRACT

Knowing the battery state of health (SOH) is essential to ensure safety and reliability in the operation of electric vehicles (EVs). However, it is still difficult to accurately estimate the SOH values of batteries owing to the complexity of their aging mechanisms. In this context, an SOH estimation method using incremental capacity (IC) and long short-term memory (LSTM) network was investigated in this study. We proposed an improved IC curve acquisition method based on reference voltage, which can retain the important features of the IC curve and reduce computational efforts. Based on the correlation between the incremental capacity and SOH of the battery, health feature variables were extracted from the IC curves. Moreover, considering the time-series characteristics and long-term dependency of battery degradation, we adopted the LSTM network to develop the SOH estimation model. The accuracy and reliability of the proposed model were verified. The results showed that the proposed IC curve acquisition method reduced the calculation time by 11.49 % compared to that using the Gaussian filter. Compared with the estimation results obtained using the support vector machine (SVM) and artificial neural network (ANN), the estimation results of the proposed IC curve acquisition method combined with the LSTM network had the minimum mean absolute percentage error (MAPE). The MAPE of the estimation results was <2 % for all the different battery samples.

$$SOH = \frac{Q_{actual}}{Q_{initial}} \quad (1)$$

1. Introduction

In recent years, environmental pollution and fossil fuel depletion have attracted significant attention. Electric vehicles (EVs), which can alleviate these increasingly serious problems due to their high efficiency and environmental friendliness, have become one of the key research fields in the automotive industry [1,2]. Characterized by their high energy density, high specific power and long cycle life, lithium-ion batteries (LIBs) have been widely used in EVs as energy storage systems [3–5]. Practically, LIBs age with continued operation, leading to a capacity loss and an increase of internal resistance [6]. The aging level is characterized by the battery state of health (SOH). Regarding EVs, people are particularly concerned about the driving range, which is related to battery capacity. Hence, the SOH is often represented as the ratio of the actual battery capacity to the initial battery capacity, as shown in Eq. (1).

An accurate SOH is essential for estimating the state of charge (SOC) of a battery and ensuring safety during its operation [7,8]. Estimating the SOH of LIBs is difficult because of their complex degradation process. The techniques proposed for SOH estimation can generally be divided into three categories: direct measurements, model-based methods and data-driven approaches [9,10]. Direct measurements of the SOH are mainly performed using experiments, such as complete discharge-charge experiments [11], electrochemical impedance spectroscopy (EIS) [12], hybrid pulse power characterization (HPPC) [13], which are easy to implement and are widely used. However, they are time-consuming and require specific test conditions, making them difficult to apply in EVs. Model-based methods mainly use equivalent circuit models or electrochemical/physical models for SOH estimation. For example, Chen et al. [14] proposed an online SOH estimation method based on an equivalent circuit model. The internal resistance of the model was estimated using a Genetic Algorithm, and the SOH of the

* Corresponding author.

E-mail address: swy_18@jlu.edu.cn (W. Sun).

Nomenclature

SOH	State of health
EV	Electric vehicle
IC	Incremental capacity
LSTM	Long short-term memory
SVM	Support vector machine
ANN	Artificial neural network
LIB	Lithium-ion battery
SOC	State of charge
EIS	Electrochemical impedance spectroscopy
HPPC	Hybrid pulse power characterization
SEI	Solid electrolyte interface
PDE	Partial differential equation
CC-CV	Constant current-constant voltage
ICA	Incremental capacity analysis
DVA	Differential voltage analysis
BMS	Battery management system
MA	Moving average
GPR	Gaussian process regression
RNN	Recurrent neural network
MSE	Mean squared error

battery was then calibrated. However, the estimation performance was sensitive to model accuracy, which was difficult to guarantee in real time. Electrochemical/physical models reveal the mechanisms of battery capacity loss, such as solid electrolyte interface (SEI) growth, lithium plating, and loss of electrode active material [15,16]. These models comprise many partial differential equations (PDEs), leading to a high computational cost. Additionally, obtaining the parameters of these models is difficult, potentially resulting in significant errors in SOH estimation. To solve the above problems, data-driven approaches have become increasingly important in SOH estimation owing to their model-free and flexible characteristics [17]. In these approaches, the extraction of high-quality feature variables related to battery degradation and the selection of appropriate algorithms are two key issues.

To extract feature variables, analyzing data on constant current-constant voltage (CC-CV) charging is a common method. Regional CV charging current and CV duration are strongly correlated with battery aging [18]. Since the CV charging phase is typically time-consuming, obtaining complete data in practical applications is difficult. Recently, incremental capacity analysis (ICA) and differential voltage analysis (DVA) have become increasingly important tools for extracting feature variables. The IC and DV curves obtained through CC charging process can be used to reveal battery degradation mechanisms, and the peak values and position of the curves can be used for SOH estimation [19,20]. This method can be easily implemented in a battery management system (BMS) and is suitable for different batteries and operating conditions [21]. Since drivers tend to recharge their EVs before the LIBs are completely discharged, maintaining the integrity of CC charging is practically difficult. Techniques for feature variables extraction using partial CC-CV charging have gradually attracted attention. Wei et al. [22] proposed a multistage SOH estimation method with a high tolerance to heavy partial charging, obtaining a high SOH estimation accuracy.

To realize SOH estimation, it is necessary to apply machine learning or intelligent algorithms to combine feature variables. Yang et al. [23] extracted feature variables from CC-CV charging data as model inputs, and proposed a Gaussian process regression (GPR) model for SOH estimation. Klass et al. [24] presented an SOH estimation method based on support vector machines (SVM), which was suitable for on-board applications with respect to processing power and memory restrictions. Kashkooli et al. [25] used calendar-aged battery data to train an

Table 1
Cell parameters.

Parameter	Component/Value
Nominal voltage	3.7 V
Cathode material	LiNi _{0.8} Co _{0.15} Al _{0.05} O ₂
Anode material	Graphite
Nominal capacity	2 Ah
Size	18,650

artificial neural network (ANN) for SOH estimation. These aforementioned algorithms are unsuitable for time series-related systems. Battery aging is a time-series process, and historical data are also a considerable feature of information for SOH estimation [26]. Thus, algorithms such as recurrent neural networks (RNN) and long short-term memory (LSTM) networks that can handle time-series data have been gradually investigated for SOH estimation, and their superior performances have been proved [27].

In summary, data-driven approaches take the feature variables under different levels of battery aging as the input dataset. After training using algorithms, data-driven models that can perform SOH estimation according to these feature variables are obtained. Thus, the quality of feature variables determines the accuracy of the estimation results. Most commonly, the IC and DV curves are used to extract the feature variables. However, owing to derivative operations, these curves are subject to noise, resulting in peak disappearances [28]. Different methods have been proposed to smooth IC curves. Li et al. [29] compared two common techniques: moving average (MA) and Gaussian filter. The smoothing effect of the MA is highly dependent on the number of selected smoothing points. As the number of smoothing points increases, the smoothing effect improves, but the important features of the curves tend to deviate. Compared to the MA, the Gaussian filter can clearly identify the peak values. However, it is more complicated and requires increased computational effort. Currently, the BMS needs to handle an increasing number of functions besides just SOH estimation, and its computational resources are limited. Although external cloud devices can be used as a solution, they require effective and reliable Internet communication, which is not always easy to obtain [30]. Thus, a simple IC curve acquisition method that can retain the important features of the curve and reduce the required computational effort is beneficial for practical SOH estimation in EVs.

To overcome the challenges of SOH estimation based on data-driven approaches, this paper investigates the battery aging data. An improved IC curve calculation method based on reference voltage is proposed and analyzed to retain the feature of the curve without significantly increasing the computational effort. Based on the smoothed IC curves, the battery degradation mechanism is analyzed, and the battery health feature variables are extracted. Considering the time-series characteristics and long-term dependency of battery aging, long short-term memory (LSTM) network is employed for SOH estimation based on health feature variables. And the effectiveness and robustness of the proposed SOH method are verified using data from the NASA data repository.

The remainder of this paper is organized as follows: In Section 2, an improved IC curve calculation method is proposed and compared with the conventional calculation and smoothing methods. Section 3 describes the mechanisms of battery aging revealed by the IC curves and the method for extracting health feature variables. In Section 4, the SOH estimation method based on LSTM network is proposed and the feasibility of the method is verified. The main conclusions are summarized in Section 5.

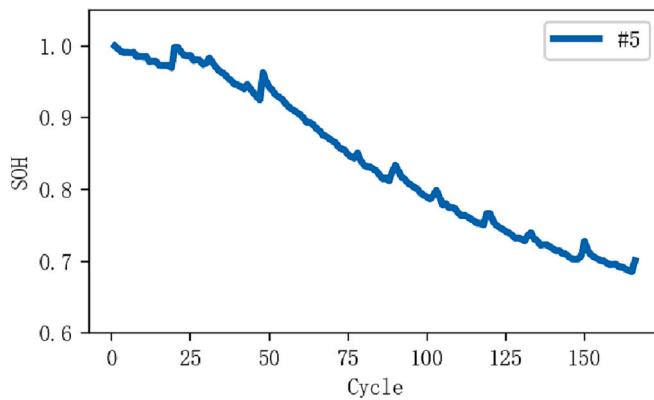


Fig. 1. Battery degradation curve.

2. Battery aging data investigation and IC curve acquisition

2.1. Battery aging data investigation

Battery aging data were collected from the NASA Ames Prognostics Center of Excellence [31]. We selected cells #5, #6, #7 and #18, which are widely used in SOH estimation investigations [32], as the research objects in this study. The basic parameters of these cells are listed in Table 1. All the batteries were cycled at room temperature (24 °C). The CC-CV charge mode was adopted for charging process. Specifically, the charging current of the CC stage was 1.5 A. When the voltage reached 4.2 V, the battery was charged under the CV mode until the charging current was <20 mA. The battery was discharged under the CC mode, where the discharge current was 2 A. And the cut-off voltages of the batteries numbered #5, #6, #7, and #18 were 2.7, 2.5, 2.2, and 2.5 V, respectively.

Taking cell #5 as an example, the change in the SOH with the number of cycles is shown in Fig. 1. It can be seen that battery SOH decreases with the increase of cycle number. However, there are some fluctuations in SOH decline due to capacity regeneration during the cycle process. To reduce the impact of these fluctuations on SOH estimation, the time-series characteristics of battery degradation should be fully considered. Additionally, it is necessary to further analyze the charging and discharging curves under different values of SOH to extract the feature variables which can be used to reveal the battery degradation mechanism and construct an accurate battery SOH estimation method.

2.2. Conventional IC curve acquisition method

During the operation of an LIB, material-level phase transitions occur at the electrodes. The phase transitions and changes during the aging of a battery affect its voltage characteristics. The ICA method is an effective tool for extracting the implicit information of the battery voltage. Based on the variation of the IC curve, the battery degradation mechanisms can be investigated. Moreover, the IC curve can also be used for SOH estimation [33]. Specifically, the IC curve expresses the change rate of capacity over the voltage range. The battery capacity and voltage during CC charging can be calculated to obtain the IC curve, as shown in Eqs. (2) and (3).

$$Q = It \quad (2)$$

$$V = f(Q) \quad (3)$$

where I is the charging current, t is the charging time, and V is the battery voltage. The conventional calculation method of the IC curve can be expressed as Eq. (4).

$$(f^{-1})' = \frac{dQ}{dV} = I \frac{dt}{dV} = g(V) \quad (4)$$

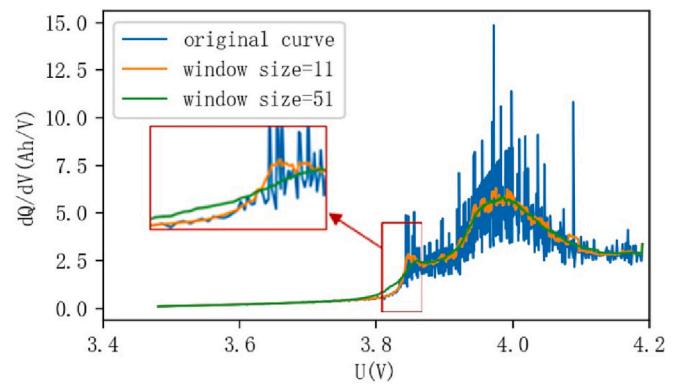


Fig. 2. Comparison of original and smoothed IC curves.

The discrete form of Eq. (4) can be represented as

$$\frac{dQ}{dV} = \frac{I(t_{k+1} - t_k)}{V_{k+1} - V_k} \quad (5)$$

Practically, the battery voltage and current are sampled by the BMS at certain time intervals, and the sampling time determines the quality of the IC curve. A longer sampling time may result in the loss of information, whereas a shorter one may produce large fluctuations in the curve. The obtained IC curves are also affected by the measurement noise. Therefore, filtering algorithms are applied to smoothen the curves. The MA is a common method to solve the problem, which can be expressed as:

$$y(i) = \frac{1}{M} \sum_{j=0}^{M-1} x(i+j) \quad (6)$$

where M is the fixed window size of the series, $x(\bullet)$ and $y(\bullet)$ are the input and the output signals, respectively. The original and the smoothed IC curves with different window sizes are shown in Fig. 2. It can be seen that the original IC curve is very noisy and it is difficult to extract useful health features. Even if the curve can be smoothed by MA, the filtering result is highly dependent on the window size. When the window size is small, the filtered curve still fluctuates greatly, making it difficult to identify the health features. When the window size is large, peak features such as position and intensity become inaccurate, potentially resulting in significant errors in SOH estimation.

2.3. Proposal and analysis of an improved IC curve calculation method

2.3.1. Improved IC curve calculation method

In order to reduce the noise in the IC curve and extract accurate health features, an improved IC curve calculation method based on reference voltage is proposed in this study. The proposed method can be generalized as follows.

Step 1: Record the time series (T_s) and the corresponding voltage series (V_s) and current series (I_s) during CC charging, as shown in Eqs. (7)–(9). The capacity series (dQ_s) corresponding to T_s can be expressed by Eqs. (10) and (11).

$$T_s = [T_1, T_2, T_3, \dots, T_l] \quad (7)$$

$$V_s = [V_1, V_2, V_3, \dots, V_l] \quad (8)$$

$$I_s = [I_1, I_2, I_3, \dots, I_l] \quad (9)$$

$$dQ_s = [dQ_1, dQ_2, dQ_3, \dots, dQ_l] \quad (10)$$

$$dQ_i = \begin{cases} 0, & i = 1 \\ I_i(T_i - T_{i-1}), & i \neq 1 \end{cases} \quad (11)$$

Step 2: Select the appropriate voltage start point (V_{start}), voltage

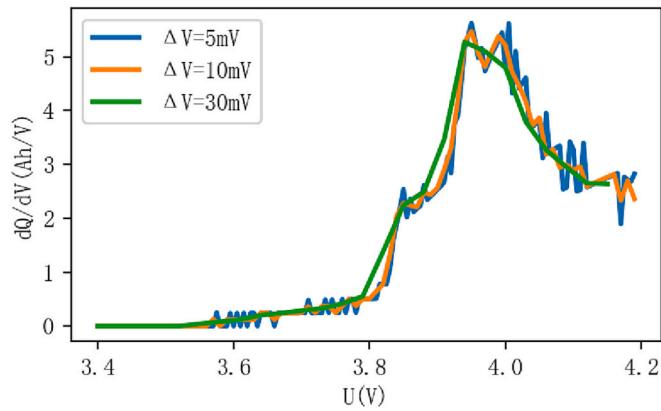


Fig. 3. Comparison of IC curves with different voltage intervals.

endpoint (V_{end}) and voltage interval (ΔV) based on the voltage range of the charging process. The reference voltage series (V_{re}) can be obtained by Eqs. (12) and (13).

$$V_{re} = [V_{re,1}, V_{re,2}, V_{re,3}, \dots, V_{re,m}] \quad (12)$$

$$V_{re,j} = V_{start} + (j - 1)\Delta V \quad (13)$$

Step 3: Create a reference incremental capacity series (IC) corresponding to V_{re} , as shown in Eq. (14), where the initial value of each element is set to 0.

$$IC = [IC_1, IC_2, IC_3, \dots, IC_m] \quad (14)$$

Step 4: Update the values of IC by processing every element in V_s and dQ_s . Specifically, input V_i into the comparison array $[|V_i - V_{re,1}|, |V_i - V_{re,2}|, |V_i - V_{re,3}|, \dots, |V_i - V_{re,m}|]$. Find the $V_{re,j}$ corresponding to the minimum value in the array. Then update the value of IC_j corresponding to $V_{re,j}$ using Eq. (15).

$$IC_j = IC_j + dQ_i \quad (15)$$

Step 5: Divide each element in the updated IC by ΔV , and the IC curve can be obtained based on the reference voltage and reference incremental capacity series.

2.3.2. Quality analysis of IC curve

The calculation platform used to obtain IC curves was a personal computer. The processor of the platform was AMD Ryzen 7-6800H (3.20 GHz) with 16GB main memory, and the programming language was Python 3. The IC curves obtained by the improved method are shown in Fig. 3. Compared with the original IC curve obtained by the conventional method in Fig. 2, the improved method can significantly reduce the curve noise. This is because the sampling time in BMS is usually short in the conventional method, making the voltage difference between any two adjacent sampling times too small, resulting in larger noise. Under the improved method, the reference voltage interval is fixed, and the capacity increment at one reference voltage consists of several capacity increments at different collected voltages. This can not only avoid the noise caused by small voltage difference, but also reduce the impact of the measurement noise during sampling.

As shown in Fig. 3, when the voltage interval is set to 5 mV, more features on the IC curve can be retained. However, the curve is subject to more noise. When the voltage interval is set to 30 mV, the IC curve is smoother but some important features are submerged. It can be concluded that the quality of the IC curve is determined by reference voltage interval. When the voltage interval is set to 5, 10, and 30 mV, the computational time is 29.01, 14.89, and 4.84 s, respectively. It can be seen that the smaller the selected reference voltage interval, the greater the computational effort. Therefore, in order to achieve a compromise between the IC curve quality and computational effort, the voltage

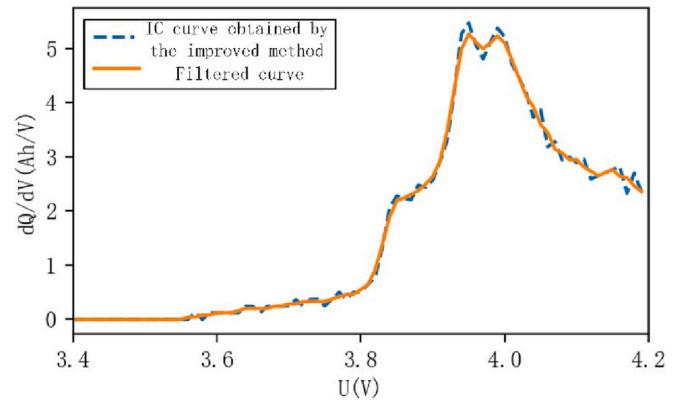


Fig. 4. IC curve obtained by the improved method and its filtered curve.

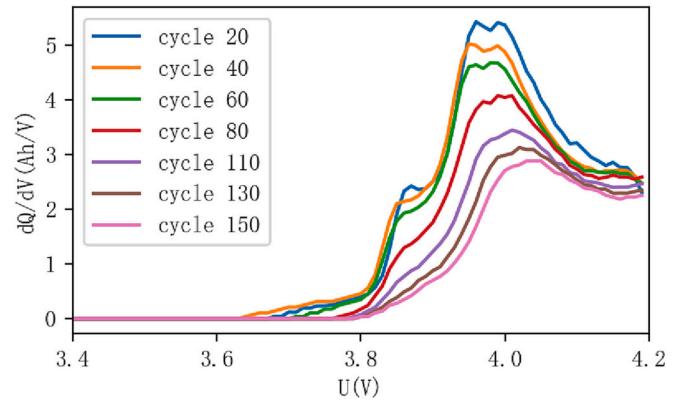


Fig. 5. IC curves under different cycle numbers.

interval is set to 10 mV for further study of SOH estimation.

After determining the reference voltage interval, the MA method was used to smoothen the IC curve obtained by the improved method. The window size was set to 3. The IC curve obtained by the improved method and the filtered curve are shown in Fig. 4. It can be seen that the filtered curve is smooth and can accurately indicate the features. The computational time involved in acquiring the IC curve using the proposed method is 11.49 % less than that required in obtaining it using a Gaussian filter.

Overall, the improved IC calculation method has several advantages. First, the IC curve obtained by the improved method can accurately indicate the features without excessive noise. In addition, this method does not adopt complex filtering algorithms for decreasing the computational efforts. Finally, the improved method calculates the IC curve based on the preset reference voltage series, ensuring the unity of the input dimensions in the data-driven model.

3. Analysis of battery aging mechanism and extraction of health feature variables

3.1. Analysis of battery aging mechanism

The peaks in the IC curve can reveal phase transitions in the battery. The decrease in the peak intensity reflects the loss of electrode active materials and lithium inventory. And the shift in the peak position reflects the change in the battery internal resistance [34]. The IC curves of battery #5 under different cycle numbers are shown in Fig. 5. It can be seen that most of the capacity is charged in the voltage range of 3.8 V to 4.2 V. For the fresh battery, there are four peaks near the voltage values of 3.86, 3.95, 4, and 4.16 V, respectively. With the increase of the cycle number, all the IC peaks shrink and the positions of the peaks shift to the

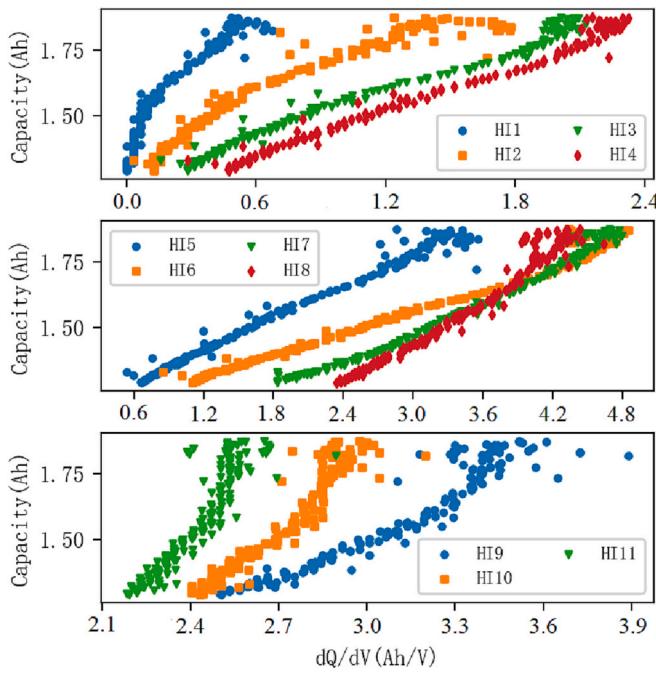


Fig. 6. Relationship between feature variables and battery capacity.

right. After 80 cycles, the first peak gradually fades away, the second peak and the third peak are gradually transformed into a large peak. It can be concluded that with battery degradation, the battery suffers from a significant loss of electrode active materials and lithium ions, resulting in the reduction of the battery capacity. In addition, the battery internal resistance increases gradually.

3.2. Extraction of health feature variables

In previous studies, the peak intensity, position, or slope of the IC curve are usually extracted as the health feature variables. However, as shown in Fig. 5, some of the peaks fade away with battery aging, making it difficult to estimate the battery SOH at the later stage of the battery life span. Aiming to solve this problem, the incremental capacity values in the voltage range of 3.8 V to 4.1 V were selected as the feature variables. Specifically, the voltage range was divided equally at a voltage interval of 30 mV, and a voltage array ($V_h = [3.8, 3.83, 3.86, \dots, 4.1] \text{ V}$) can be obtained. Subsequently, incremental capacity values corresponding to the elements in the voltage array were extracted to construct the health feature variables ($HI = [HI_1, HI_2, HI_3, \dots, HI_{11}]$).

Fig. 6 shows the relationship between the health feature variables and the actual capacity of battery #5 under different battery health states. It can be seen that each health feature variable is close to a linear

relationship with the actual battery capacity. The Pearson correlation analysis method was applied to quantitatively analyze the strength of the correlation, as shown in Eq. (16).

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (16)$$

where r_{xy} is the correlation coefficient, which ranges from -1 to 1 , x_i is the feature variable with index i , y_i is the actual battery capacity corresponding to x_i , \bar{x} and \bar{y} are the average values of x and y , respectively.

The correlation analysis results for cells #5, #6, #7, and #18 are shown in Fig. 7. A correlation coefficient close to $1/-1$ indicates a strong positive/negative correlation. As shown in Fig. 7, most feature variables have a strong positive correlation with the actual battery capacity. However, for cell #7, the correlation coefficients of HI_{10} and HI_{11} are <0.8 . Similarly, for cell #18, HI_1 and HI_2 are also weakly correlated with battery health. In order to ensure a high quality for the extracted health feature variables and make them applicable to different batteries, the variables HI_3 to HI_9 are adopted to develop the SOH estimation model.

4. SOH estimation based on the LSTM network

4.1. Description of the LSTM network

The LSTM network proposed by Hochreiter and Schmidhuber is a kind of improved recurrent neural network (RNN) [35]. RNN can remember previous information, making it quite suitable for solving time-series problems [36]. However, exploding gradients may occur when the time series increases. Compared with the basic RNN, the LSTM can decide to update or discard information through a special network structure called a “gate”, thus overcoming such shortcomings. The LSTM unit structure is shown in Fig. 8.

As shown in Fig. 8, the input of the LSTM unit includes the input layer at the current moment (x_t), hidden layer at the previous moment (h_{t-1}) and cell state at the previous moment (C_{t-1}). The output includes the hidden layer and cell state at the current moment (h_t , C_t). There are three gates to determine the information that should be memorized or forgotten: the input gate, forget gate, and output gate. The gates selectively transform the information through a sigmoid layer and dot products. The representations for the different gates of the LSTM are as follows [37].

Forget gate (f_t): The forget gate can be described as Eq. (17). It can be seen that the forget gate squashes x_t and h_{t-1} into 0 to 1. When $f_t = 1$, information should be completely retained; otherwise, this information should be completely discarded. When $0 < f_t < 1$, part of information is reserved.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (17)$$

Input gate (i_t): Through the input gate, the information which needs

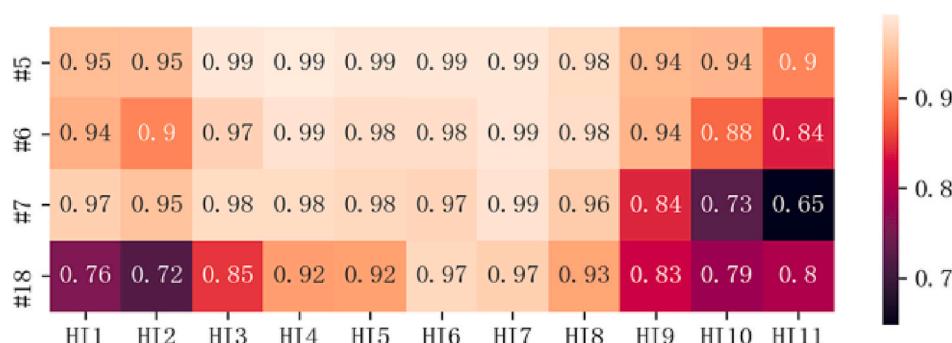


Fig. 7. Heatmap of the Pearson correlation coefficient.

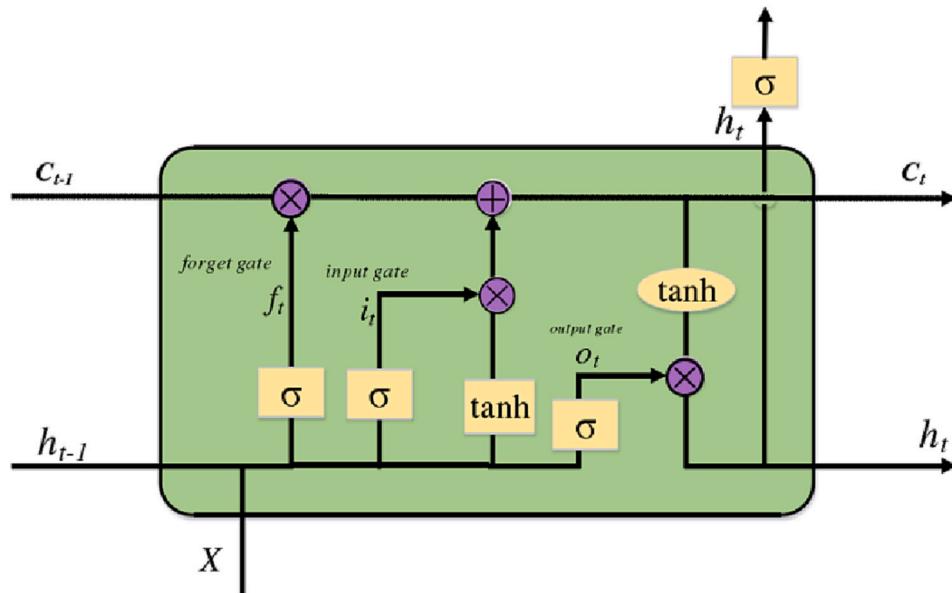


Fig. 8. LSTM network structure.

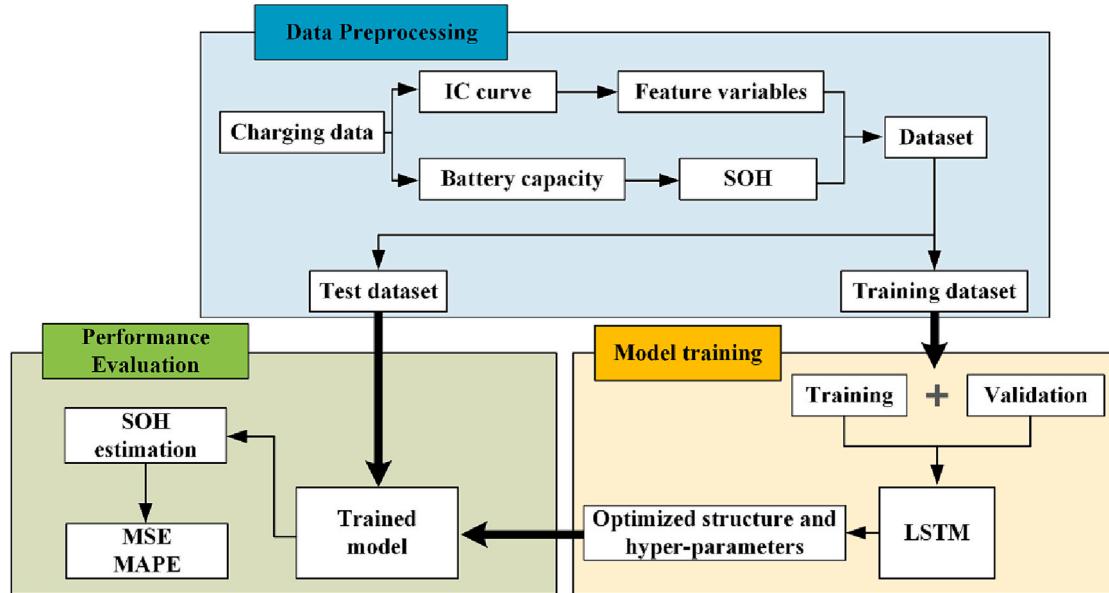


Fig. 9. Overview of SOH estimation process.

to be stored in the cell state is determined. This is done in two steps. Firstly, a sigmoid function is used to decide the values to be updated, as shown in Eq. (18).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (18)$$

Then a vector of new candidate values (\tilde{C}_t) is created using a tanh function and added to the state. The calculation of \tilde{C}_t is expressed as Eq. (19).

$$\tilde{C}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (19)$$

The current cell state can be determined using Eq. (20).

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (20)$$

Output gate (o_t): The output gate determines the information that should be outputted, as shown in Eq. (21). An additional tanh function is used to squash the output values into -1 to 1 , as shown in Eq. (22).

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (21)$$

$$h_t = o_t \tanh(C_t) \quad (22)$$

In Eqs. (17)–(22), σ is the sigmoid activation function, \tanh is the tan hyperbolic function, b is the bias, W_x and W_h are the weights for the input and last output, respectively.

4.2. SOH estimation based on LSTM

Considering that battery aging is a time-series process and the LSTM network is suitable for solving such problems, an SOH estimation method based on this network is proposed. The process of the SOH estimation method is illustrated in Fig. 9. This method consists of three stages: data preprocessing, model training, and performance evaluation. In the data preprocessing stage, the health feature variables extracted

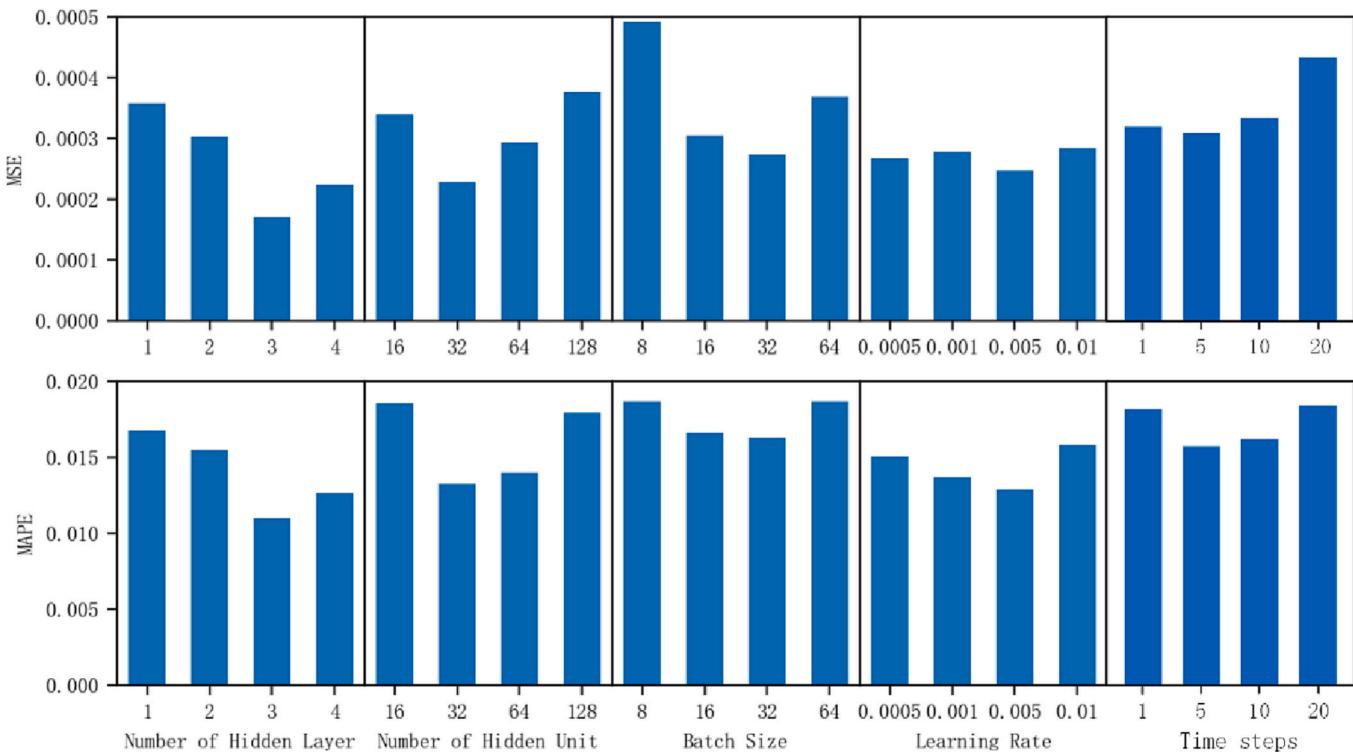


Fig. 10. Estimation performance with different hyper-parameters.

from the IC curves and the SOH of the battery during aging are processed into a dataset which can meet the requirements of the LSTM network. Then the dataset is divided into a training dataset and a test dataset. In the model training stage, the training dataset is split into training and validation datasets. The training data are used to train the LSTM network, and the validation data are used to optimize the model structure and hyper-parameters. In the performance evaluation stage, the trained model is used to estimate battery SOH based on the test dataset. The performance of the SOH estimation method is evaluated using the mean squared error (MSE) and the mean absolute percentage error (MAPE).

4.2.1. Data preprocessing

The obtained data should be processed to make them suitable for the LSTM network. The input of the LSTM network needs to contain variables under different times to express the relativity of the predicted objects in a time series. According to Section 3, the health feature variables under the i -th charging cycle (HI_i) can be expressed by Eq. (23). In order to meet the requirements of the LSTM network, a matrix (IP_i) containing HI_i along with the health feature variables of the $s-1$ cycles before the i -th charging cycle was established, as shown in Eq. (24). The input (IP) of the LSTM network can be constructed by Eq. (25), where n is the total number of cycles.

$$HI_i = [HI_{3,i}, HI_{4,i}, HI_{5,i}, \dots, HI_{9,i}] \quad (23)$$

$$IP_i = \begin{bmatrix} HI_{3,i-s+1} & HI_{4,i-s+1} & \dots & HI_{9,i-s+1} \\ \vdots & \vdots & & \vdots \\ HI_{3,i-2} & HI_{4,i-2} & \dots & HI_{9,i-2} \\ HI_{3,i-1} & HI_{4,i-1} & \dots & HI_{9,i-1} \\ HI_{3,i} & HI_{4,i} & \dots & HI_{9,i} \end{bmatrix} \quad (24)$$

$$IP = [IP_1, IP_2, IP_3, \dots, IP_i, \dots, IP_n] \quad (25)$$

The output (OP) of the network is the SOH under different cycles, which can be described by Eq. (26).

$$OP = [SOH_1, SOH_2, SOH_3, \dots, SOH_i, \dots, SOH_n] \quad (26)$$

After determining the input and output of the LSTM network, the dataset was divided into a training dataset and a test dataset. In this study, the first 70 % of the data were used for training and the last 30 % of the data were used to evaluate the SOH estimation performance.

4.2.2. Model training and optimization

In the model training stage, the training dataset was further split into training and validation datasets. The validation data were used to evaluate the estimation performance with different model structures and hyper-parameters. Based on the validation results, the model can be optimized.

The K-fold cross validation approach was used to select the model structure and hyper-parameters. In this approach, the training dataset is randomly divided into K groups. Then K times of training and validation are performed. In each training and validation process, one group is selected as the validation data, and the remaining groups are merged for training [38]. The MSE and the MAPE were used to quantitatively evaluate the estimation performance, which are expressed as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (27)$$

$$MAPE = 100\% \times \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (28)$$

where \hat{y}_i is the estimated values, y_i is the measured values, N is the number of samples.

The number of hidden layers, number of hidden units, batch size, learning rate, and time steps are essential hyper-parameters of the LSTM network. These parameters affect the accuracy, convergence speed, and computational effort of the model. The performance of SOH estimation with different hyper-parameters was investigated through the K-fold cross validation approach, as shown in Fig. 10. Considering the estimation performance and computational effort comprehensively, the

Table 2
Optimized hyper-parameters.

Hidden layer	Hidden unit	Batch size	Learning rate	Time steps
3	32	32	0.005	5

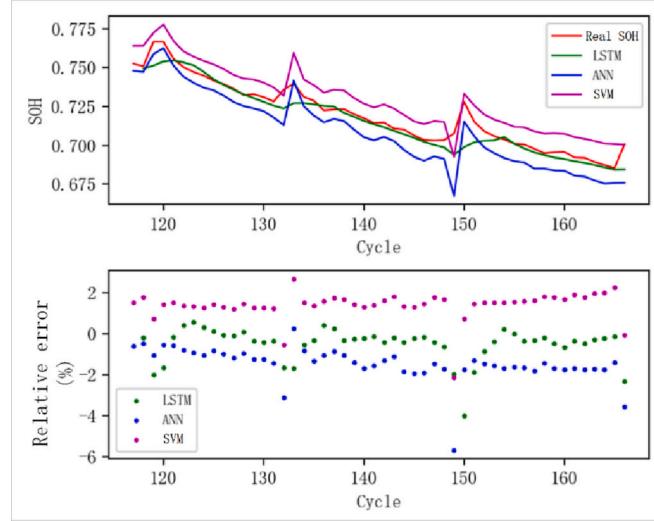


Fig. 11. Comparison of SOH estimation performance using different methods.

optimized hyper-parameters were selected, as shown in Table 2.

4.2.3. Estimation results and discussion

The data of battery #5 were used for training and testing the SOH estimation model. After optimization, the estimation performance of the model was evaluated. In order to intuitively show the advantages of the SOH estimation method based on LSTM network, the SOH estimation results using ANN and SVM are compared with the results of the proposed method, as shown in Fig. 11. Overall, all three methods have good SOH estimation results. Among them, the estimation results based on the LSTM network are closest to the real SOH, and the maximum relative error is 4 %. Since the LSTM network can completely reflect the time-series dependence of the SOH changes during the battery aging process, the estimation results are more stable and more in line with practical battery degradation.

To verify the reliability of the proposed model, data from batteries #7 and #18 are used to evaluate the SOH estimation performance based on different methods. As shown in Fig. 12(a), in the early stages of battery aging, the estimation results of the three methods have large deviations due to the differences in the battery and cycle conditions. During the cycling process, the cut-off voltage of battery #7 is lower, resulting in a more severe loss of electrode active material, which is mainly reflected in the incremental capacity in the low voltage range. However, the health feature variables are extracted from the higher voltage range, leading to an error in the SOH estimation. After 50 cycles, the estimation results under the three methods gradually approach the real SOH. Fig. 12(b) shows the estimation results for battery #18. It can be seen that compared with that of battery #5, the estimation performance of battery #18 becomes worse. The maximum relative error of estimation using the LSTM model reaches 6.24 %. It can be concluded that the inconsistency of the battery has a certain impact on the estimation performance using data-driven approaches.

The estimation performance of the different methods for different batteries can be quantitatively evaluated from the MSE and MAPE values in the estimation results, which are summarized in Table 3. It can be concluded that, although the estimation performance deteriorates when the methods are migrated to the other batteries, the MAPE of the

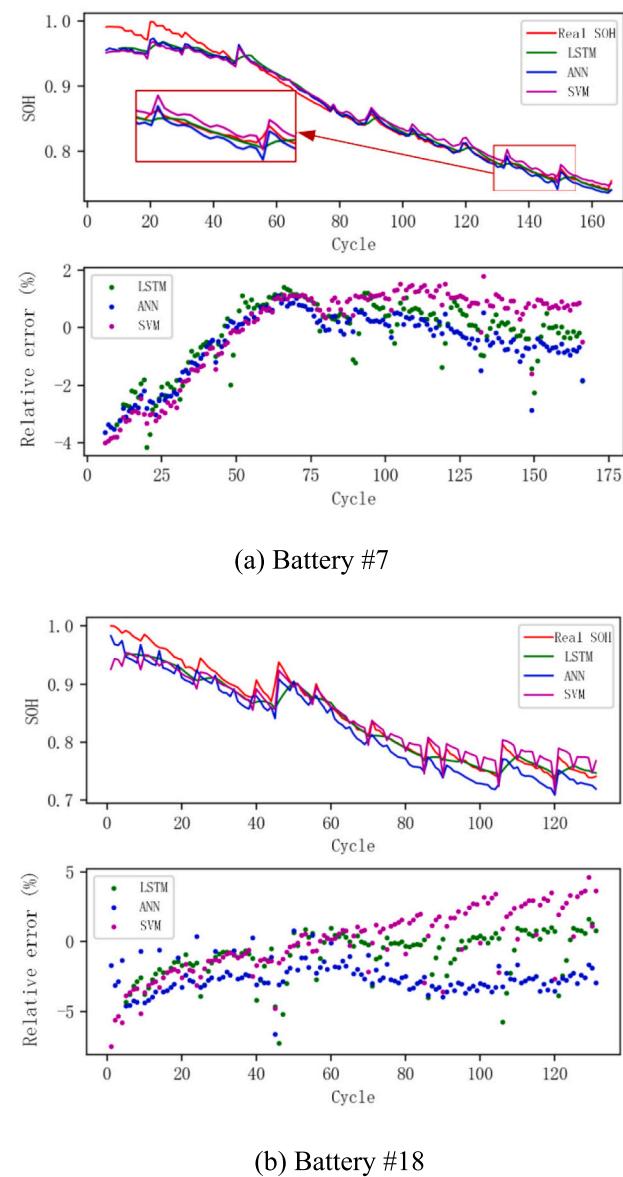


Fig. 12. SOH estimation for different batteries.

Table 3
Summary of SOH estimation performance results.

Battery number	Algorithm	MSE	MAPE
#5	LSTM	0.0000503	0.0061
	ANN	0.0001443	0.0147
	SVM	0.0001246	0.0150
#7	LSTM	0.0001695	0.0098
	ANN	0.0002144	0.0101
	SVM	0.0002889	0.0138
#18	LSTM	0.0002898	0.0133
	ANN	0.0005612	0.0259
	SVM	0.0004296	0.0193

estimation results based on the LSTM network is still <2 %, showing that the estimation performance is satisfactory. This means that the proposed model can reliably estimate the SOH of different battery samples. In addition, compared with the different methods, SOH estimation under the proposed method provided better results. This is because the proposed method can completely reflect the time-series dependence of the battery aging process, thereby reducing the influence of noise and

abnormal values on the estimation results.

5. Conclusions

In this study, an SOH estimation method for LIBs using the incremental capacity and LSTM network was investigated. Firstly, the battery aging data was studied, and an improved IC curve acquisition method based on reference voltage was proposed. In addition, the health feature variables were extracted from the IC curves. The correlation between the variables and battery health states was validated. Moreover, a SOH estimation model based on LSTM was developed, and its estimation performance was also studied. The main conclusions of the study are summarized as follows:

- (1) The original IC curves obtained by the improved method can accurately retain the health features without excessive noise. It is unnecessary to process the curves using complex filtering algorithms that can effectively reduce computational efforts.
- (2) The incremental capacity values in the voltage of 3.8 V to 4.1 V are selected to extract health feature variables. The correlation coefficients of the selected health feature variables are >0.8 , indicating that the variables have a strong correlation with the battery SOH.
- (3) Compared with the SOH estimation methods based on ANN and SVM, the estimation results of the proposed method using LSTM have the minimum MSE and MAPE. And the proposed estimation method can reduce the influence of noise and abnormal values on the estimation results.
- (4) Although the estimation performance deteriorates when the SOH estimation method using LSTM is migrated to different batteries, the MAPE of the estimation results is still $<2\%$, indicating that the proposed model has good reliability.

The insights provided in this study are useful for estimating the SOH of batteries. In future study, it is essential to further investigate the online SOH estimation method of the battery packs equipped on EVs. Accurate state estimation methods for batteries can contribute to the application in EVs.

CRediT authorship contribution statement

Zhaopu Zhang: Methodology, Investigation, Writing- Original draft.
 Haitao Min: Validation, Formal analysis, Funding acquisition.
 Hangang Guo: Validation, Formal analysis.
 Yuanbin Yu: Validation, Data Curation, Writing-Review & Editing.
 Weiyi Sun: Validation, Formal analysis, Software, Funding acquisition.
 Junyu Jiang: Validation, Formal analysis, Writing-Revision.
 Hang Zhao: Validation, Data Curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data was collected from the NASA Ames Prognostics Center of Excellence

Acknowledgements

This research is supported by National Natural Science Foundation of China (Grant No. 52102458), Jilin Province Science and Technology Development Program (Grant No. 20210301023GX), and data support

from the NASA Ames Prognostics Center of Excellence.

References

- [1] T.R. Hawkins, O.M. Gausen, A.H. Strømman, Environmental impacts of hybrid and electric vehicles—a review, *Int. J. Life Cycle Assess.* 17 (2012) 997–1014, <https://doi.org/10.1007/s11367-012-0440-9>.
- [2] F. Sun, R. Xiong, H. He, A systematic state-of-charge estimation framework for multi-cell battery pack in electric vehicles using bias correction technique, *Appl. Energy* 162 (2016) 1399–1409, <https://doi.org/10.1016/j.apenergy.2014.12.021>.
- [3] D. Ansean, M. Gonzalez, V.M. Garcia, J.C. Viera, J.C. Anton, C. Blanco, Evaluation of LiFePO₄ batteries for electric vehicle applications, *IEEE Trans. Ind. Appl.* 51 (2015) 1855–1863, <https://doi.org/10.1109/SmartMILE.2013.6708211>.
- [4] H. Min, B. Wang, W. Sun, Z. Zhang, Y. Yu, Y. Zhang, Research on the combined control strategy of low temperature charging and heating of lithium-ion power battery based on adaptive fuzzy control, *Energies* 13 (2020) 1584, <https://doi.org/10.3390/en13071584>.
- [5] X. Hu, C. Zou, C. Zhang, Y. Li, Technological developments in batteries: a survey of principal roles, types, and management needs, *IEEE Power Energ. Mag.* 15 (2017) 20–31, <https://doi.org/10.1109/MPPE.2017.2708812>.
- [6] M. Galeotti, C. Giannamico, L. Cinà, S. Cordiner, A.D. Carlo, Synthetic methods for the evaluation of the state of health (SOH) of nickel-metal hydride (NiMH) batteries, *Energy Convers. Manag.* 92 (2015) 1–9, <https://doi.org/10.1016/j.enconman.2014.12.040>.
- [7] C. Zou, C. Manzie, D. Nešić, A.G. Kallapur, Multi-time-scale observer design for state-of-charge and state-of-health of a lithium-ion battery, *J. Power Sources* 335 (2016) 121–130, <https://doi.org/10.1016/j.jpowsour.2016.10.040>.
- [8] G. Dong, Z. Chen, J. Wei, Q. Ling, Battery health prognosis using brownian motion modeling and particle filtering, *IEEE Trans. Ind. Electron.* 65 (2018) 8646–8655, <https://doi.org/10.1109/TIE.2018.2813964>.
- [9] Z. Ma, R. Yang, Z. Wang, A novel data-model fusion state-of-health estimation approach for lithium-ion batteries, *Appl. Energy* 237 (2019) 836–847, <https://doi.org/10.1016/j.apenergy.2018.12.071>.
- [10] J. He, Z. Wei, State-of-health estimation of lithium-ion batteries using incremental capacity analysis based on voltage-capacity model, *IEEE Trans. Transp. Electrific.* 6 (2020) 417–426, <https://doi.org/10.1109/TTE.2020.2994543>.
- [11] R. Xiong, L. Li, J. Tian, Towards a smarter battery management system: a critical review on battery state of health monitoring methods, *J. Power Sources* 405 (2018) 18–29, <https://doi.org/10.1016/j.jpowsour.2018.10.019>.
- [12] M. Galeotti, L. Cinà, C. Giannamico, S. Cordiner, A. Di Carlo, Performance analysis and SOH (state of health) evaluation of lithium polymer batteries through electrochemical impedance spectroscopy, *Energy* 89 (2015) 678–686, <https://doi.org/10.1016/j.energy.2015.05.148>.
- [13] N. Omar, J. Van Mierlo, B. Verbrugge, P. Van den Bossche, Power and life enhancement of battery-electrical double layer capacitor for hybrid electric and charge-depleting plug-in vehicle applications, *Electrochim. Acta* 55 (2010) 7524–7531, <https://doi.org/10.1016/j.electacta.2010.03.039>.
- [14] Z. Chen, C.C. Mi, Y. Fu, J. Xu, X. Gong, Online battery state of health estimation based on genetic algorithm for electric and hybrid vehicle applications, *J. Power Sources* 240 (2013) 184–192, <https://doi.org/10.1016/j.jpowsour.2013.03.158>.
- [15] S. Yang, Y. Hua, D. Qiao, Y. Lian, Y. Pan, Y. He, A coupled electrochemical-thermal-mechanical degradation modelling approach for lifetime assessment of lithium-ion batteries, *Electrochim. Acta* 326 (2019), 134928, <https://doi.org/10.1016/j.electacta.2019.134928>.
- [16] J. Li, K. Adewuyi, N. Lotfi, R.G. Landers, J. Park, A single particle model with chemical/mechanical degradation physics for lithium ion battery state of health (SOH) estimation, *Appl. Energy* 212 (2018) 1178–1190, <https://doi.org/10.1016/j.apenergy.2018.01.011>.
- [17] Y. Zhen, M. Ouyang, L. Lu, J. Li, Understanding aging mechanisms in lithium-ion battery packs: from cell capacity loss to pack capacity evolution, *J. Power Sources* 278 (2015) 287–295, <https://doi.org/10.1016/j.jpowsour.2014.12.105>.
- [18] A. Eddahetch, O. Bariat, J.M. Vinassa, Determination of lithium-ion battery state-of-health based on constant-voltage charge phase, *J. Power Sources* 258 (2014) 218–227, <https://doi.org/10.1016/j.jpowsour.2014.02.020>.
- [19] M. Berecibar, M. Garmendia, I. Gandiaga, J. Crego, I. Villarreal, State of health estimation algorithm of LiFePO₄ battery packs based on differential voltage curves for battery management system application, *Energy* 103 (2016) 784–796, <https://doi.org/10.1016/j.energy.2016.02.163>.
- [20] Z. Wang, J. Ma, L. Zhang, State-of-health estimation for lithium-ion batteries based on the multi-island genetic algorithm and the Gaussian process regression, *IEEE Access* 5 (2017) 21286–21295, <https://doi.org/10.1109/ACCESS.2017.2759094>.
- [21] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, P. Van den Bossche, Critical review of state of health estimation methods of Li-ion batteries for real applications, *Renew. Sust. Energ. Rev.* 56 (2016) 572–587, <https://doi.org/10.1016/j.rser.2015.11.042>.
- [22] Z. Wei, H. Ruan, Y. Li, J. Li, C. Zhang, H. He, Multistage state of health estimation of lithium-ion battery with high tolerance to heavily partial charging, *IEEE Trans. Power Electron.* 37 (2022) 7432–7442, <https://doi.org/10.1109/TPEL.2022.3144504>.
- [23] D. Yang, X. Zhang, R. Pan, Y. Wang, Z. Chen, A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve, *J. Power Sources* 384 (2018) 387–395, <https://doi.org/10.1016/j.jpowsour.2018.03.015>.
- [24] V. Klapp, M. Behm, G. Lindbergh, A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation,

- J. Power Sources 270 (2014) 262–272, <https://doi.org/10.1016/j.jpowsour.2014.07.116>.
- [25] A.G. Kashkooli, H. Fathiannasab, Z. Mao, Z. Chen, Application of artificial intelligence to state-of-charge and state-of-health estimation of calendar-aged lithium-ion pouch cells, J. Electrochem. Soc. 166 (2019) A605–A615, <https://doi.org/10.1149/2.0411904jes>.
- [26] L. Yao, J. Wen, S. Xu, J. Zheng, J. Hou, Z. Fang, Y. Xiao, State of health estimation based on the long short-term memory network using incremental capacity and transfer learning, Sensors 22 (2022) 7835, <https://doi.org/10.3390/s22207835>.
- [27] Y. Toughzaoui, S.B. Toosi, H. Chaoui, H. Louahlia, R. Petrone, S.L. Masson, H. Gualous, State of health estimation and remaining useful life assessment of lithium-ion batteries: a comparative study, J. Energy Storage 51 (2022), 104520, <https://doi.org/10.1016/j.est.2022.104520>.
- [28] X. Li, Z. Wang, J. Yan, Prognostic health condition for lithium battery using the partial incremental capacity and Gaussian process regression, J. Power Sources 421 (2019) 56–67, <https://doi.org/10.1016/j.jpowsour.2019.03.008>.
- [29] Y. Li, M. Abdel-Moneim, R. Gopalakrishnan, M. Berecibar, E. Nanini-Maury, N. Omar, P. Van den Bossche, J. Van Mierlo, A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by gaussian filter, J. Power Sources 373 (2018) 40–53, <https://doi.org/10.1016/j.jpowsour.2017.10.092>.
- [30] S. Yang, Z. Zhang, R. Cao, M. Wang, H. Cheng, L. Zhang, Y. Jiang, Y. Li, B. Chen, H. Ling, Y. Lian, B. Wu, X. Liu, Implementation for a cloud battery management system based on the CHAIN framework, Energy AI 5 (2021), 100088, <https://doi.org/10.1016/j.egyai.2021.100088>.
- [31] B. Saha, K. Goebel, Battery Data Set, NASA Ames Prognostics Data Repository, 2007, <http://ti.arc.nasa.gov/project/prognostic-data-repository>.
- [32] T. Pham, T. Le, D. Dang, H. Bui, H. Pham, L. Truong, M. Nguyen, H. Vo, Q.T. Tho, ARNS: a data-driven approach for SoH estimation of lithium-ion battery using nested sequence models with considering relaxation effect, IEEE Access 10 (2022) 117067–117083, <https://doi.org/10.1109/ACCESS.2022.3217478>.
- [33] E. Riviere, P. Venet, A. Sari, F. Meniere, Y. Bultel, LiFePO₄ battery state of health online estimation using electric vehicle embedded incremental capacity analysis, IEEE Veh.Power Propuls.Conf. 2015 (2015) 1–6, <https://doi.org/10.1109/VPPC.2015.7352972>.
- [34] Y. Gao, J. Jiang, C. Zhang, W. Zhang, Z. Ma, Y. Jiang, Lithium-ion battery aging mechanisms and life model under different charging stresses, J. Power Sources 356 (2017) 103–114, <https://doi.org/10.1016/j.jpowsour.2017.04.084>.
- [35] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [36] S. Li, C. Ju, J. Li, R. Fang, Z. Tao, B. Li, T. Zhang, State-of-charge estimation of lithium-ion batteries in the battery degradation process based on recurrent neural network, Energies 14 (2021) 306, <https://doi.org/10.3390/en14020306>.
- [37] Y. Liu, X. Shu, H. Yu, J. Shen, Y. Zhang, Y. Liu, Z. Chen, State of charge prediction framework for lithium-ion batteries incorporating long short-term memory network and transfer learning, J. Energy Storage 37 (2021), 102494, <https://doi.org/10.1016/j.est.2021.102494>.
- [38] T. Wong, P. Yeh, Reliable accuracy estimates from k-fold cross validation, IEEE Trans. Knowl. Data Eng. 32 (2020) 1586–1594, <https://doi.org/10.1109/TKDE.2019.2912815>.