

## Article

# State of Health Estimation and Remaining Useful Life Prediction of Lithium-Ion Batteries by Charging Feature Extraction and Ridge Regression

Minghu Wu <sup>1,2</sup>, Chengpeng Yue <sup>2</sup>, Fan Zhang <sup>1,2,\*</sup>, Rui Sun <sup>2</sup>, Jing Tang <sup>1,2</sup>, Sheng Hu <sup>1,2</sup>, Nan Zhao <sup>1,2</sup> and Juan Wang <sup>1,2</sup>

- <sup>1</sup> Hubei Key Laboratory for High-Efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan 430068, China; wuxx1005@mail.hbut.edu.cn (M.W.); jingtang@hbut.edu.cn (J.T.); 20181008@hbut.edu.cn (S.H.); nzhaoh@mail.hbut.edu.cn (N.Z.); happywj@hbut.edu.cn (J.W.)
- <sup>2</sup> School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China; 102110269@hbut.edu.cn (C.Y.); 102210269@hbut.edu.cn (R.S.)
- \* Correspondence: joyce\_zhang@hbut.edu.cn

**Abstract:** The state of health (SOH) and remaining useful life (RUL) of lithium-ion batteries are critical indicators for assessing battery reliability and safety management. However, these two indicators are difficult to measure directly, posing a challenge to ensure safe and stable battery operation. This paper proposes a method for estimating SOH and predicting RUL of lithium-ion batteries by charging feature extraction and ridge regression. First, three sets of health feature parameters are extracted from the charging voltage curve. The relationship between these health features and maximum battery capacity is quantitatively evaluated using the correlation analysis method. Then, the ridge regression method is employed to establish the battery aging model and estimate SOH. Meanwhile, a multiscale prediction model is developed to predict changes in health features as the number of charge-discharge cycles increases, combining with the battery aging model to perform multistep SOH estimation for predicting RUL. Finally, the accuracy and adaptability of the proposed method are confirmed by two battery datasets obtained from varying operating conditions. Experimental results demonstrate that the prediction curves can approximate the real values closely, the mean absolute error (MAE) and root mean square error (RMSE) calculations of SOH remain below 0.02, and the maximum absolute error (AE) of RUL is no more than two cycles.

**Keywords:** lithium-ion batteries; state of health; remaining useful life; charging feature extraction; ridge regression; maximum battery capacity; aging model; multiscale prediction model



**Citation:** Wu, M.; Yue, C.; Zhang, F.; Sun, R.; Tang, J.; Hu, S.; Zhao, N.; Wang, J. State of Health Estimation and Remaining Useful Life Prediction of Lithium-Ion Batteries by Charging Feature Extraction and Ridge Regression. *Appl. Sci.* **2024**, *14*, 3153. <https://doi.org/10.3390/app14083153>

Academic Editor: Elza Bontempi

Received: 4 March 2024

Revised: 28 March 2024

Accepted: 4 April 2024

Published: 9 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Lithium-ion batteries are commonly employed in electronics, new energy vehicles, and power source systems owing to their high energy density, long life cycle, and low self-discharge rate [1,2]. However, as the number of charge-discharge cycles increases, performance degradation will occur for the battery, such as a decrease in the maximum available capacity and an increase in internal resistance [3]. If the battery is not replaced promptly, great safety risks will be posed to the equipment, seriously affecting the stability and reliability of the power systems [4,5]. Accurately estimating battery status is essential for effective battery management, and the state of health (SOH) and remaining useful life (RUL) of lithium-ion batteries are critical indicators for evaluating battery performance, which can effectively quantify the degree of aging [6].

In the field of battery health assessment, SOH is a widely used macro indicator to measure the overall health of a battery. In addition to SOH, the micro-health parameters of the battery are also important aspects in assessing the health status, including the

volume fraction of negative electrode active materials, solid-phase diffusion coefficient, electrolyte concentration, and so on. Micro-health parameters are for the performance of the active material and electrolyte inside the battery, and changes in micro-health parameters can present the internal health state of the battery [7]. Although micro-health parameters can provide more detailed information about the internal state of the battery, their measurement and evaluation processes are relatively complex and require advanced models and algorithms. In contrast, SOH is more easily assessed by battery capacity, internal resistance, and other parameters. Therefore, in this study, SOH is chosen as the primary indicator for assessing battery health and the mainstream definition of SOH describes it as the ratio of maximum usable capacity to rated capacity [8]. As the number of charge-discharge cycles increases, the SOH will continuously decrease. When SOH decreases to 70%~80%, the battery is typically recognized as the end of life (EOL), and the number of cycles to degradation from the current state to EOL is called the RUL of the battery [9,10].

In addition, there is a correlation between estimating SOH and predicting RUL, as both can be obtained by estimating battery capacity. Model-based and data-driven methods are included in the main research methods [11]. Model-based methods can be classified as equivalent circuit models and electrochemical models, each based on distinct modeling mechanisms. In the equivalent circuit model, intricate physical and chemical processes within the battery are simplified and basic electronic components are used to simulate battery output [12]. Zhang et al. [13] introduced a recursive least square weighted decoupling method to enhance parameter estimation accuracy across various dynamic regimes. Misyris et al. [14] used recursive least squares with variable forgetting factors to identify parameters of the equivalent circuit model, achieving internal impedance identification and capacity estimation. Wang et al. [15] developed an equivalent circuit model using the constant voltage charging current curve to predict battery SOH by determining relevant circuit model features. However, establishing a precise model of battery aging is challenging owing to the intricate internal reaction mechanism [16]. Electrochemical models model the dynamics of battery performance degradation by simulating physical and chemical processes within the battery [17]. Khodadadi Sadabadi et al. [18] simulated the battery charging and discharging process by establishing an electrochemical model to analyze the affected factors of battery capacity degradation, thereby predicting RUL. Hong et al. [19] designed an enhanced single-particle model capable of predicting battery aging states using aging research data from LMO-NMC-cathode graphite-anode batteries. Nevertheless, it is more difficult to develop a precise battery degradation model because the electrochemical model is dynamic and nonlinear. The above model-based methods provide an accurate representation of the external dynamic characteristics of the battery. However, a substantial amount of prior knowledge is required to use the method, and frequent adjustments to the model according to the specific type of lithium-ion battery and varying operating conditions are needed, making it unsuitable for real-time forecasting [20].

In contrast, data-driven methods do not require intensive examination of the battery's aging mechanism, but battery performance parameters are directly extracted, such as voltage, current, resistance, temperature, and other data as health features. The health feature parameters are analyzed and modeled with the trend of capacity degradation to estimate the internal state and performance of the batteries. Data-driven methods offer the advantages of efficiency, simplicity, and ease of use, and have attracted extensive attention in various fields [21]. Zhang et al. [22] used the rate of temperature change associated with capacity degradation as input to time convolution networks and SOH as output, achieving good prediction accuracy. Wu et al. [23] implemented RUL prediction by analyzing the change rule of the terminal voltage curve during battery charging and extracting degradation features as inputs to the feed-forward neural network. Khumprorn et al. [24] introduced an application of deep neural networks to extract eight features from charge-discharge curves to predict SOH and RUL, with better results. Chen et al. [25] used capacity and equal voltage drop discharge time as health features, preprocessed health

features using ensemble empirical mode decomposition (EEMD), introduced phase space reconstruction to optimize input sequences, and finally combined it with support vector regression (SVR) to complete the prediction of RUL. Zhang et al. [26] combined partial incremental capacity with an artificial neural network for estimating battery status under constant current discharge conditions. Deng et al. [27] used constant current charging time and capacity, as well as constant voltage charging time, as features for training the SVR model to predict SOH. Chen et al. [28] advanced a hybrid model combining a modified adaptive noise algorithm for performing EEMD decomposition of capacity data, predicting each decomposition component separately using LSSVR. Li et al. [29] investigated changes in battery charging status under various health conditions. The particle swarm algorithm was employed to optimize the kernel function of the support vector machine for the joint estimation of SOC and SOH. Experimental results show that the method is remarkably adaptable and feasible. Tian et al. [30] used support vector machines to predict trends in capacity degradation by extracting health features from the temperature differential curve during constant current battery charging. Wu et al. [31] extracted features from charging capacity and incremental capacity data to model battery aging through ridge regression. Experimental results showed that compared to SVR and Gaussian process regression, more reliable estimates with a simpler structure and lower computational cost were provided by ridge regression using selected features. The data-driven methods mentioned above are simple to implement and achieve high prediction accuracy when reliable training sets are used. However, due to dynamic changes in electrochemical reactions within the battery. This includes factors such as fine-tuning the electrode material structure, electrolyte redistribution, and repair of the solid electrolyte interface layer, which can lead to a phenomenon of capacity regeneration in the battery within a short time [32]. The existence of this capacity regeneration phenomenon can lead to a situation where the battery has the same SOH but different RUL in two different cycles. This makes it difficult to assess the degree of battery aging comprehensively and accurately by relying only on the SOH. Therefore, it is necessary to consider the SOH and RUL of the current battery to make a comprehensive diagnosis of the degree of aging throughout its life cycle. However, current common evaluation methods usually estimate only one of them, while estimating both separately will inevitably increase computational and algorithmic complexity.

Based on the previous analysis, a method is proposed for estimating SOH and predicting RUL of lithium-ion batteries by charging feature extraction and ridge regression. First, three sets of health feature parameters are extracted from the charging voltage curve, and the Pearson correlation coefficient is applied to analyze these health features for maximum battery capacity. Then, the ridge regression method is applied to develop the battery aging model and estimate the SOH. On this basis, a multiscale prediction model is developed to predict trends in health features as charge-discharge cycles increase. The results of the multiscale prediction model are integrated with the battery aging model to estimate SOH through multiple steps, enabling RUL prediction. Finally, the accuracy and adaptability of the proposed method are confirmed by two battery datasets procured under varied operating conditions. Experimental results show that SOH and RUL are accurately predicted for the proposed method with high accuracy and reliability.

The remainder of this paper is structured as follows: Section 2 presents battery degradation datasets and explains the process of extracting health features. Section 3 describes the rationale for the selected models and algorithms, as well as the overall framework for estimating SOH and predicting RUL. Section 4 provides experimental validation of the method, followed by discussion and analysis of the results. Section 5 summarizes the work carried out in this paper.

## 2. Health Feature Parameters

### 2.1. Battery Degradation Datasets

This paper used two datasets of lithium-ion batteries under various operating conditions to assess the prediction performance of the proposed method. The first dataset was

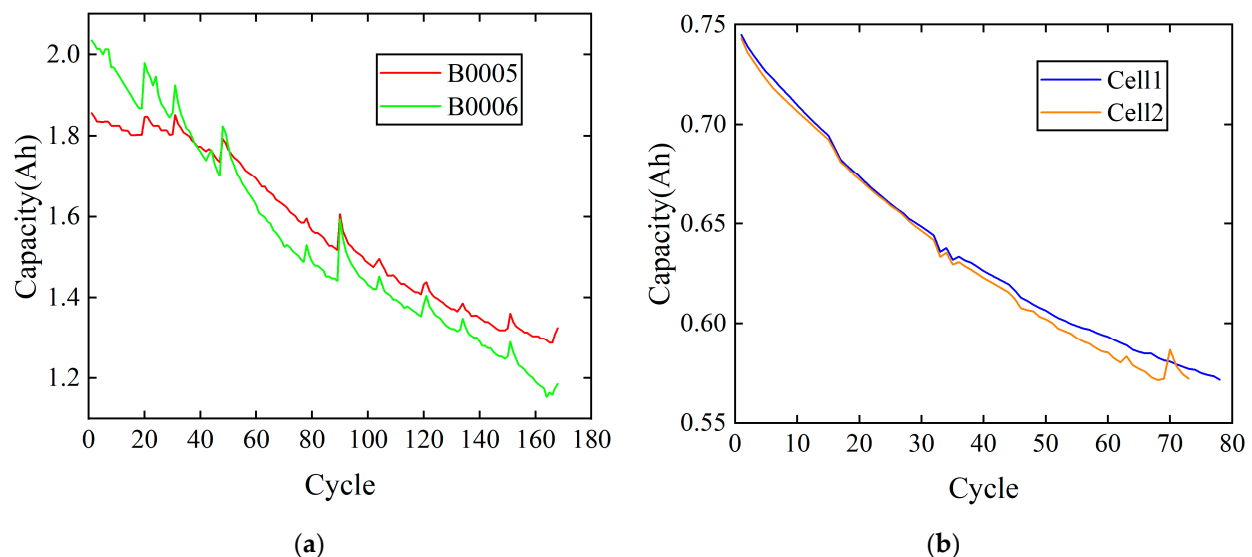
from NASA lithium-ion battery dataset [33], and B0005 and B0006 batteries were selected as research objects. Both batteries had a rated capacity of 2.0 Ah, and both underwent 168 charge-discharge cycles, setting the battery failure threshold at 70% (1.4 Ah) of its rated capacity. The experimental data was generated through charge-discharge experiments at 24 °C, and the battery was initially charged at 1.5 A in constant current mode until the voltage reached 4.2 V. Then, the mode was changed to constant voltage charging until the charging current decreased to 20 mA. The discharge process employed a constant current of 2 A until the B0005 and B0006 batteries reached voltages of 2.7 V and 2.5 V.

The second dataset was obtained from Oxford University's lithium-ion battery aging dataset [34], where Cell1 and Cell2 batteries were designated as research objects. The rated capacity of both batteries was 0.74 Ah and underwent 78 and 73 charge-discharge cycles, respectively. The failure threshold for lithium-ion batteries was set at 80% (0.59 Ah) of their rated capacity. The aging experiments were conducted at 40 °C. The batteries were charged initially at 1.48 A in constant current mode and then discharged to simulate Artemis urban driving conditions. After every 100 charge-discharge cycles, a capacity calibration experiment was conducted until the end of battery life, and the aging test was completed.

These two datasets provide capacity degradation data and related information for the entire battery life cycle, which is suitable for state estimation experiments. The above specific battery parameters are detailed in Table 1, and the battery capacity variation with charge-discharge cycles is illustrated in Figure 1.

**Table 1.** The parameters and operating conditions of the selected batteries.

Battery	Charging Current (A)	Rated Capacity (Ah)	Temperature (°C)	Number of Cycles
B0005	1.5	2	24	168
B0006	1.5	2	24	168
Cell1	1.48	0.74	40	78
Cell2	1.48	0.74	40	73



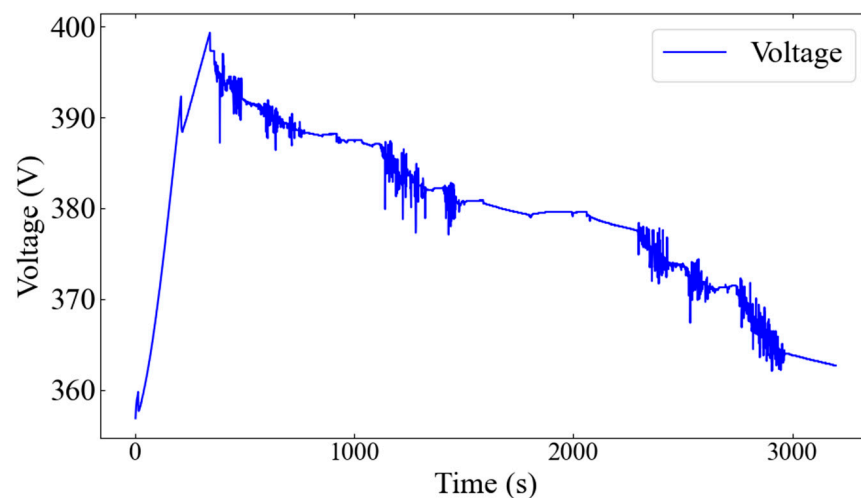
**Figure 1.** Battery capacity decline curve: (a) Battery B0005 and B0006 in the NASA dataset; (b) Battery Cell1 and Cell2 in the Oxford dataset.

## 2.2. Health Features Extraction

The data-driven estimation method establishes a mapping relationship with maximum battery capacity using multidimensional health feature data. Therefore, how to select health features is crucial for data-driven methods to accurately describe trends in battery capacity

degradation. The operating phases of lithium-ion batteries mainly involve the charging and discharging phases. During the charging phase, a fixed constant current and constant voltage mode is commonly implemented. Compared to the irregularity of the discharging phase under actual operating conditions, it is easier to extract stable and effective health feature data under the charging phase [35]. Characterization of the charging phase is relatively easy to analyze and extract because the charging process is usually carried out in a controlled constant voltage and constant current mode. Because, in practice, the discharge phase of Li-ion batteries may be affected by a variety of factors, resulting in irregularities in discharge data. For example, factors such as environmental conditions and usage behavior can lead to more complexity in extracting regular data during the discharge phase. In real life, when operating an electric vehicle artificially, the driver's behaviors such as acceleration, deceleration, and braking will lead to fluctuations in the battery discharge current and voltage; the vehicle load (e.g., the number of passengers and the weight of cargo) will also affect the battery discharge rate and voltage output. The combined effect of these external factors leads to the fact that the discharge data of electric vehicles under actual operating conditions tend to show a high degree of irregularity. Conversely, constant voltage and constant current charge states make it easier to extract stable and effective health characteristics data.

To demonstrate the irregularity of the discharge data more intuitively, as shown in Figure 2, which shows the variation curve of the battery pack voltage of an electric vehicle under actual operating conditions. The discharge voltages show a rather chaotic trend in actual operation. These irregularities can create difficulties for our modeling and prediction. Therefore, the feature data of the charging voltage curve are analyzed in this paper.

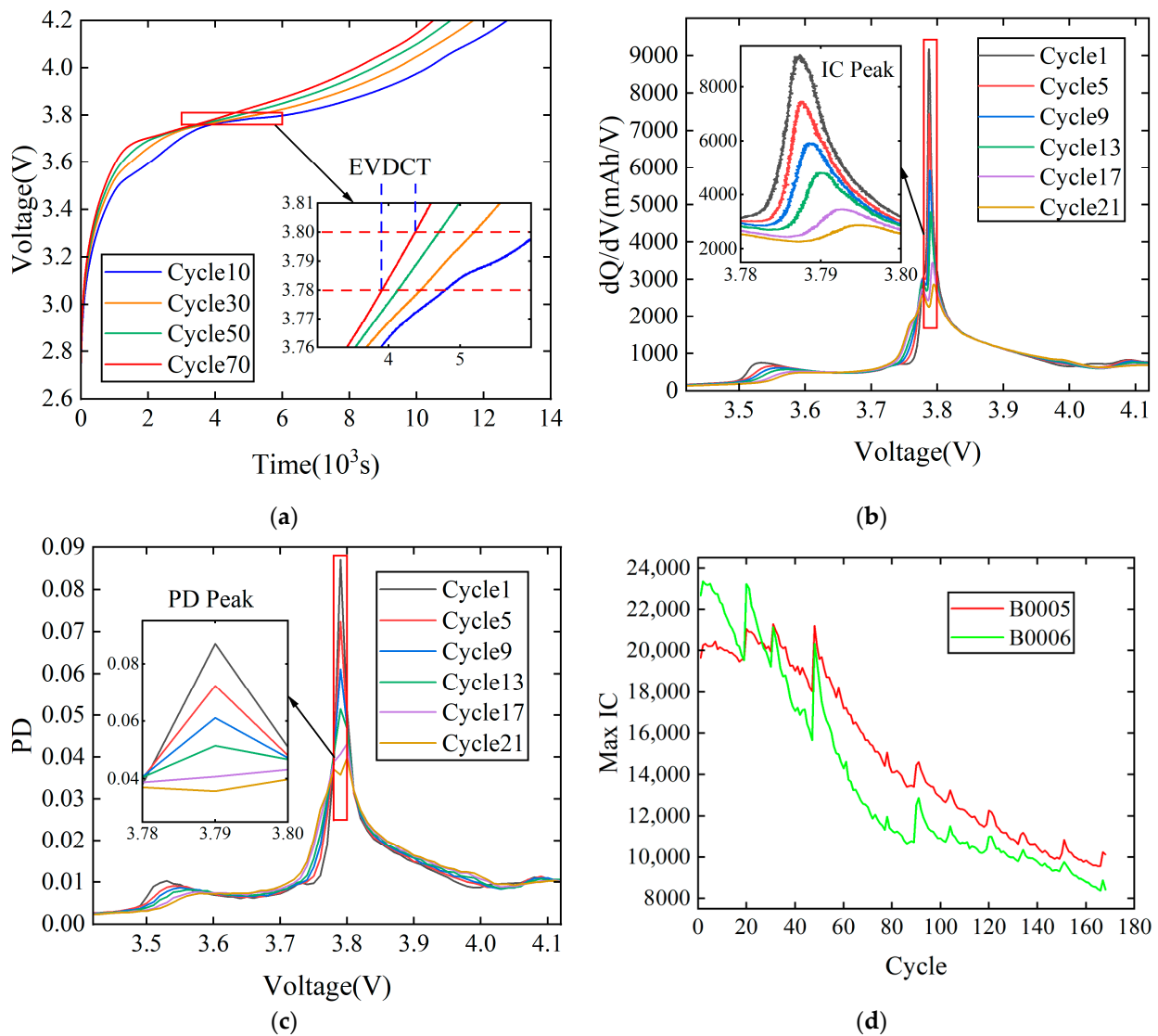


**Figure 2.** Discharge voltage curve under actual vehicle operation.

For example, the charging voltage curve of the Cell1 battery changes with battery degradation as shown in Figure 3a. In the initial and final stages, the curve is characterized by steepness. Nevertheless, the curve is relatively flat and is known as the plateau voltage in the intermediate state. The plateau voltage duration gradually shortens as the charging and discharging cycles increase, which aligns with the decreasing maximum capacity of the battery. Consequently, equal voltage difference charging time (EVDCT) can be used as a key feature parameter for assessing battery status. The incremental capacity analysis (ICA) method elucidates how the decrease in maximum battery capacity is related to changes in charging voltage. The ICA method calculates the difference  $dQ/dV$  between capacity  $Q$  and voltage  $V$  to accentuate incremental capacity related to distinct voltage intervals during charging. Figure 3b displays the change in the capacity increment curve, with the highest peak decreasing gradually with an increasing number of cycles, reflecting the trend of maximum battery capacity. The probability density (PD) analysis method correlates voltage distribution with capacity decline by calculating probability density for each voltage value.



Figure 3c shows the correlation between peak PD curve and capacity decline. By observing Figure 3a–c, it is evident that the voltage data needed for feature extraction in Cell1 battery mainly ranges from 3.78 V to 3.80 V. This indicates that the prediction process does not necessitate complete charge voltage data and only a smaller amount of data needs to be collected, resulting in reduced data computation and improved processing efficiency.



**Figure 3.** The process of extracting health features: (a) The charging curves of battery Cell1; (b) The capacity increment curves of battery Cell1; (c) The voltage probability density curves of battery Cell1; (d) The Max IC curves of B0005 and B0006 batteries.

To show the above conclusions more clearly, trends in IC peaks over the entire cycle process of B0005 and B0006 batteries are summarized. Changes and curves over the cycles are plotted as shown in Figure 3d. The max IC trend gradually decreases with increasing cycles, while showing some volatility, manifested as a global decline and a local increase. The trend closely resembles the downward trend in the actual maximum battery capacity, which verifies the validity of the selected health feature parameters in this paper.

### 2.3. Correlation Analysis

This paper utilizes Pearson correlation analysis to quantify the relationship between health features and battery capacity. The calculation formula is:

$$\rho_{X,Y} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (1)$$

where  $X_i$  and  $\bar{X}$  represent health features and their mean values,  $Y_i$  and  $\bar{Y}$  are battery capacity and mean value, and  $N$  represents data length.

The Pearson correlation coefficient serves as a statistical metric used to evaluate the magnitude of the association between two variables. A stronger linear relationship between the two variables is indicated when a correlation coefficient is close to 1 in absolute value. In the field of statistics, it is widely recognized that when the absolute value of the correlation coefficient exceeds 0.8, a robust correlation between the two variables can be considered to exist. Table 2 shows the correlation coefficients for each health feature with battery capacity. The results reveal significant linear correlations between the three health feature parameters extracted in this paper and battery capacity. Accordingly, EVDCT, IC peak, and PD peak can be effectively employed as health features to characterize battery capacity degradation.

**Table 2.** Pearson correlation coefficient of health features.

Health Features	B0005	B0006	Cell1	Cell2
EVDCT	0.9915	0.9941	0.8666	0.9930
IC Peak	0.9953	0.9802	0.8829	0.9016
PD Peak	0.8287	0.9085	0.9247	0.9212

### 3. Methods

A feasible battery degradation prediction model needs to be constructed after extracting health features from battery charging data. Through the analysis of the health features, it is shown that the correlation coefficients between the selected inputs and outputs are above 0.82, and most of them are above 0.90, indicating a very high correlation between the selected health features and maximum battery capacity. Therefore, linear regression can be used for fitting models because it is computationally inexpensive and can be trained offline.

From the previous section, feature variables are extracted and transformed from the charging voltage curve. Therefore, there is bound to be overlapping information between them, that is, the problem of multicollinearity exists. The presence of a high correlation between data features can contribute to overfitting the linear regression model, ultimately affecting the accuracy of predictions. In addition, a simple linear regression model is extremely susceptible to noise from input data, and even small errors can lead to significant changes in output variables.

#### 3.1. Ridge Regression

Ridge regression is an improved bias estimation method based on least square estimation [36]. The method optimizes the model by introducing the term L2 regularization to obtain more realistic and reliable regression coefficients, effectively solving the problems of multicollinearity among variables and improving the overall fitting effect of the model. The ridge regression formula is defined as follows:

$$\hat{\beta}_{ridge} = \underset{\beta}{\operatorname{argmin}} \left[ \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right] \quad (2)$$

where  $\hat{\beta}_{ridge}$  is the vector of coefficients obtained from ridge regression,  $\beta_j$  is the coefficients of the  $j$ th feature,  $y_i$  is the target value of the  $i$ th sample,  $x_{ij}$  is the  $j$ th feature value of the

$i$ th sample,  $n$  is the number of samples, and  $p$  is the number of features. Where  $\lambda$  is the coefficient of the penalty term,  $\lambda \geq 0$ . The greater the coefficient value, the more drastic the contraction of the coefficient vector. When  $\lambda = 0$ , the above formula represents the least squares estimation (LSE), so LSE can be considered a special case of ridge regression.

The matrix form of ridge regression is:

$$\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T Y \quad (3)$$

where  $I$  is the identity matrix, when the eigenvector has a collinearity problem, the matrix  $X^T X$  is irreversible and cannot be solved for  $\hat{\beta}_{ridge}$ . The penalty term added in ridge regression makes the matrix  $X^T X + \lambda I$  full rank and invertible, which makes  $\hat{\beta}_{ridge}$  solvable, and it can be concluded that ridge regression functions to reduce the singularity of the eigenmatrix, and at the same time compresses the size of the coefficient vectors, thus mitigating the risk of overfitting.

### 3.2. Multiscale Prediction Model

#### 3.2.1. Ensemble Empirical Mode Decomposition

EEMD is an improved EMD method based on noise-assisted analysis [37]. The principle is to decompose the signal  $n$  times by EMD and add white noise with fixed variance and zero means to the initial signal in each decomposition process. By using the property that white noise has zero means to offset the noise effect and taking the integrated mean result as the final result, the modal mixing problem of the EMD algorithm has been effectively solved, ensuring the automatic distribution of the signal in the appropriate time scale.

The EEMD algorithm decomposes complex non-stationary signals into intrinsic mode function (IMF) components at various time scales. Compared to the initial signal, the fluctuation amplitude of each IMF component is reduced, and more precise forecasting results can be obtained by analyzing the IMFs separately. This method has significant advantages in extracting the fluctuation patterns of complex sequences and amplifying the precision of the prediction model. The EEMD algorithm consists of the following specific steps:

- (1) Add white noise sequence  $w_i(t)$  to the initial signal  $x(t)$  to obtain new signal  $X_i(t)$ :

$$X_i(t) = x(t) + w_i(t) \quad (4)$$

where  $i$  is the number of white noise additions.

- (2) The EMD decomposition of  $X_i(t)$  is performed to obtain the sum of the IMF components of each order and the residual component  $r_{es}$  after decomposition:

$$X_i(t) = \sum_{j=1}^n c_{ij}(t) + r_{es} \quad (5)$$

where  $c_{ij}(t)$  is the  $j$ th IMF component derived from the decomposition after adding white noise for the  $i$ th time, and the value of  $j$  is in the range of  $1 \sim n$ .

- (3) Repeat the above two steps  $M$  times and add the IMF components obtained each time, then calculate the mean value as the result:

$$c_j(t) = \frac{1}{M} \sum_{i=1}^M c_{ij}(t) \quad (6)$$

where  $c_j(t)$  is the  $j$ th IMF component average obtained after EEMD decomposition of the initial signal.



### 3.2.2. Multiple Linear Regression

When there are two or more independent variables in regression analysis, and the independent variable and the dependent variable have a linear correlation, the resulting analysis is called multiple linear regression (MLR). Compared to conventional neural networks, MLR offers high prediction accuracy, easy parameter adjustment, and fast operation speed when dealing with low frequency components that are highly periodic and have a flat trend [38]. The expression and expansion of the MLR matrix are:

$$Y = X \times \beta + \mu \quad (7)$$

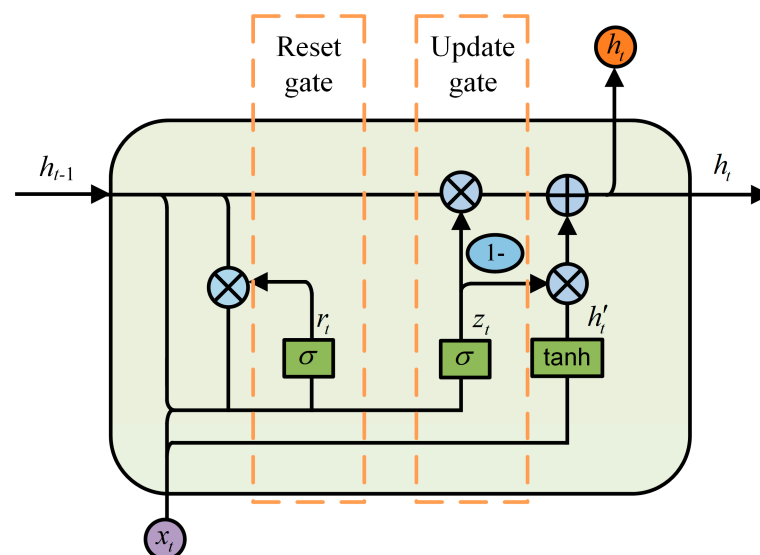
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ 1 & x_{21} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nn} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \quad (8)$$

where  $Y$  is the dependent variable,  $X$  is the independent variable,  $\beta$  is the regression coefficient, and  $\mu$  is the random variable. The regression function can be obtained by solving the parameters using the least squares method, and its formula is:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (9)$$

### 3.2.3. Gated Recurrent Unit

The gated recurrent unit (GRU) optimizes the long short-term memory by combining the forgetting gate and the input gate into an update gate. This optimizes network parameters, enhances convergence speed, and effectively reduces the risk of data overfitting. GRU has excellent performance in handling complex, fluctuating non-smooth, and nonlinear data with excellent performance. Due to the design of the gating mechanism, the GRU model can extract and retain important “key information” in high-frequency components with better capture of long-term dependence in sequence data. The GRU network structure is illustrated in Figure 4.



**Figure 4.** Internal structure diagram of GRU network.

The formulas for its status and output are as follows:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (10)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (11)$$

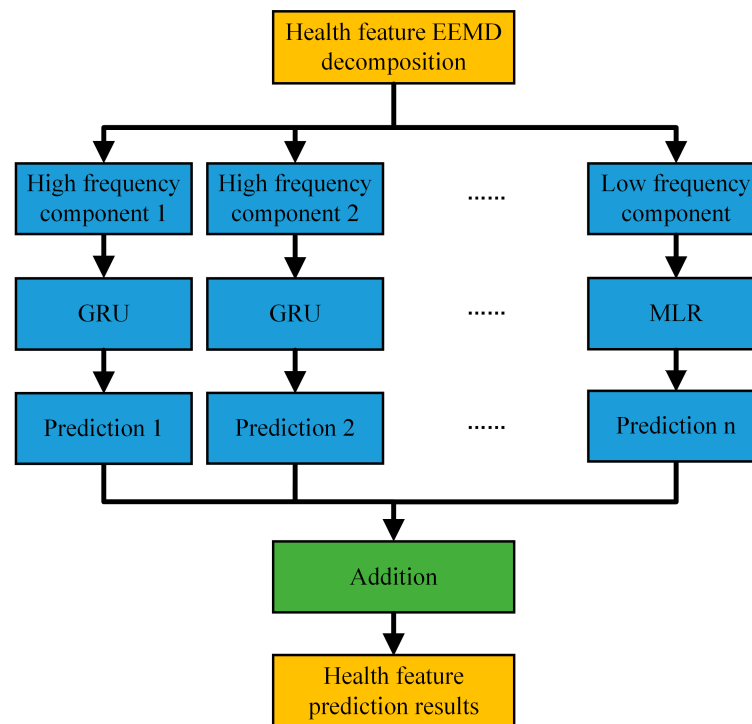
$$h'_t = \tanh(W_{h'} \cdot [r_t * h_{t-1}, x_t]) \quad (12)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * h'_t \quad (13)$$

where  $x_t$  is the input value at the current moment;  $z_t$  is the value of the update gate at the current moment;  $r_t$  is the value of the reset gate at the current moment;  $h_t$  and  $h_{t-1}$  are the state of the implicit layer at the current moment and the previous moment, respectively;  $h'_t$  is the activation state of the implicit layer at the current moment;  $W_r$ ,  $W_z$  and  $W_{h'}$  are the weight matrices;  $\sigma$  and  $\tanh$  are the activation functions.

### 3.2.4. Multiscale Prediction Modeling

From Figure 3d, as the number of cycles increases, health features show a tendency to decline like capacity, posing a challenge to the model training process. This paper proposes a multiscale prediction model to solve the problem. First, EEMD is utilized to break down the health feature sequence into its high and low frequency components, which have simpler fluctuation patterns and significant frequency characteristics. Then, high frequency and low frequency components are predicted using GRU and MLR models, respectively. Finally, the prediction results from both models are merged to generate a new dataset of health features. The flowchart is shown in Figure 5.



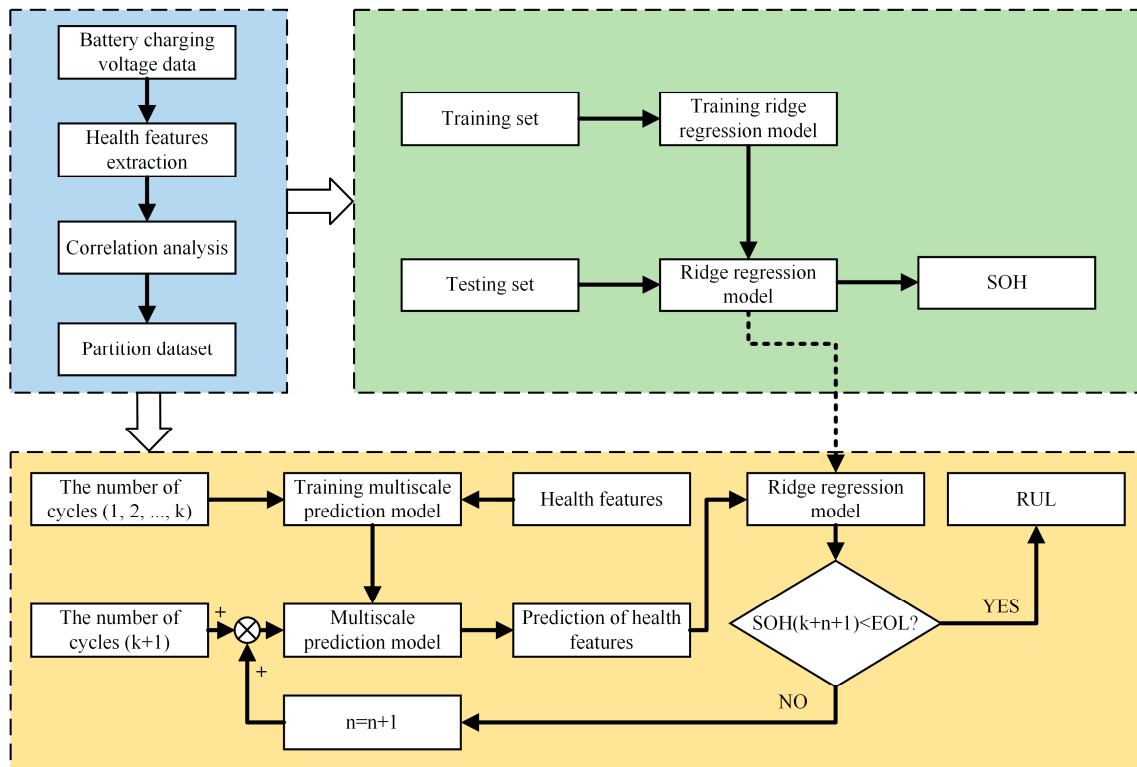
**Figure 5.** Flowchart of multiscale prediction model.

### 3.3. Overall Prediction Framework

SOH and RUL serve as crucial indicators for assessing the degree of battery aging, but they are based on different principles and application scopes. Considering the relationship and difference between SOH and RUL, this paper suggests a coupling framework for evaluating and predicting both SOH and RUL in an integrated manner.

The framework of the estimation method is illustrated in Figure 6. Among them, the blue area represents the data processing process, the green area represents the SOH estimation process, and the yellow area indicates the RUL prediction process. Specific steps are as follows: Initially, charging voltage data for lithium-ion batteries is collected, and EVDCT, IC Peak, and PD Peak are extracted from them as health features. These features are then used as inputs and SOH as outputs to develop a battery aging model using ridge regression. Meanwhile, the multiscale prediction model is employed to forecast the health

features of future cycles. Finally, the results are combined with the ridge regression model to make a multistep SOH prediction and achieve RUL prediction.



**Figure 6.** Flowchart of SOH and RUL estimation.

## 4. Results and Discussion

### 4.1. Evaluation Metrics

Prediction results are evaluated using mean absolute error (MAE), root mean square error (RMSE), and absolute error (AE). The following formulas are presented:

$$MAE = \frac{1}{n} \sum_{i=1}^n |x(i) - \hat{x}(i)| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x(i) - \hat{x}(i))^2} \quad (15)$$

$$AE = |T_{RUL} - \hat{T}_{RUL}| \quad (16)$$

where  $n$  is the number of predicted cycles,  $x(i)$  is the real value,  $\hat{x}(i)$  is the predicted value,  $T_{RUL}$  is the number of cycles at the end of battery life in the actual case, and  $\hat{T}_{RUL}$  is the predicted value of RUL.

### 4.2. SOH Estimation Results

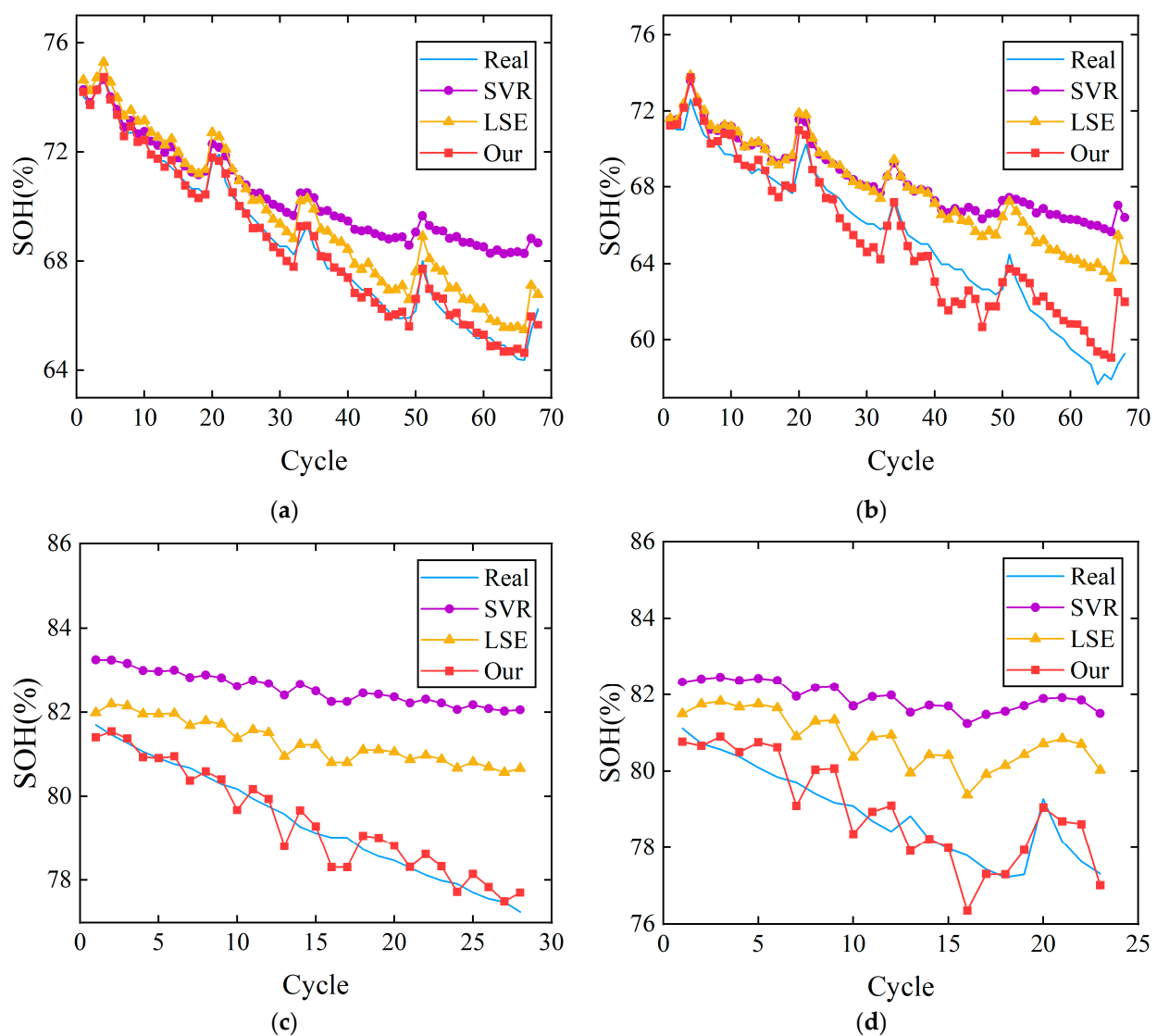
Divide experimental data into training and test sets, and the ridge regression model is constructed by the training set, while its performance is assessed by the test set. Specifically, the first 100 cycle data from B0005 and B0006 of NASA battery datasets are used as training sets, and subsequent cycle data are used as test sets. Similarly, the first 50 cycles of Cell1 and Cell2 in the Oxford battery dataset are chosen as training sets, followed using the remaining cycle data as test sets. Throughout each prediction process, three selected health features are selected from the charging voltage curve and utilized as inputs for the ridge regression model, with SOH estimates as outputs.

To confirm the superiority of the ridge regression model, a comparative experiment was conducted with SVR and LSE. Specific descriptions of the three forecasting methods are illustrated in Table 3.

**Table 3.** Methods statement.

Prediction Methods	Describe	Advantages	Disadvantages
SVR	Support Vector Regression	Strong generalization ability	Complex parameter tuning
LSE	Least Squares Estimation	Simplicity and Efficiency	Sensitive to outliers
Our method	Ridge Regression	Handles multicollinearity well	Regularization selection

Figure 7 displays the SOH prediction results for three methods using various batteries, and a comparison of their estimation errors is shown in Table 4.



**Figure 7.** Estimation results of three different methods on the battery SOH: (a) B0005; (b) B0006; (c) Cell1; (d) Cell2.

**Table 4.** MAE and RMSE of comparison for different model predictions.

Battery	Methods	MAE	RMSE
B0005	SVR	0.0339	0.0411
	LSE	0.0177	0.0186
	Our Method	0.0042	0.0051
B0006	SVR	0.0628	0.0770
	LSE	0.0521	0.0597
	Our Method	0.0188	0.0193
Cell1	SVR	0.0236	0.0246
	LSE	0.0142	0.0155
	Our Method	0.0022	0.0027
Cell2	SVR	0.0226	0.0235
	LSE	0.0143	0.0153
	Our Method	0.0036	0.0045

Our method shows superior performance in predicting battery SOH, as shown in Figure 7. Its prediction curve closely tracks the real SOH degradation curve, accurately capturing the trend of battery capacity decline. Moreover, Our method effectively adapts to the deterioration process of the battery and its capacity regeneration phenomenon. Compared with the SVR and LSE methods, Our method exhibits a notable enhancement in predictive precision. Most of its predicted SOH values closely match the real values, making it well suited for predicting the real SOH trend. However, a careful comparison can be found in Figure 7c,d, where the simulated results are slightly different from the actual results and appear to jump up and down. This may be due to errors in data collection or processing resulting in bias in the input data used for model training. Therefore, the datasets used for training and testing need to be further reviewed to minimize noise disturbances and outliers and to ensure data accuracy.

Table 4 shows the MAE and RMSE calculation results, and Our method prediction error results on the four battery datasets show extremely low values, which remain below 0.02. Moreover, under the same conditions, Our method exhibits the smallest prediction error. Especially for Cell1 and Cell2, the MAE and RMSE are maintained within 0.004 and 0.005. This shows that the ridge regression model proposed in this paper not only adapts effectively to different datasets but also exhibits superior prediction accuracy.

Among them, the estimation errors of both Cell1 and Cell2 are relatively small, because there is no obvious fluctuation during battery decline and the overall degradation trend is relatively smooth. Therefore, all three prediction methods effectively capture the real SOH curve trends for Cell1 and Cell2 batteries. However, for battery B0006, the degradation trend is more drastic, and fluctuations up and down are more obvious. Within 60 cycles, the SOH of this battery decreased by about 15% and significant capacity regeneration occurred, resulting in a large error in the prediction effect for battery B0006. However, Our method outperforms the other two methods in predicting the real degradation trend with the smallest error, and this also reflects the applicability of Our method in dealing with battery data with drastic and fluctuating degradation trends.

#### 4.3. RUL Prediction Results

RUL prediction is based on the predicted starting point set in the SOH estimation. The left side of Figure 8 shows the real value and estimated value. The real value of RUL is obtained by calculating the cycles needed to reach EOL from the current cycle count. The red circle on the right side of Figure 8 shows the absolute error between the real and estimated values. From Figure 8, the estimated and real values of all four batteries closely match, with a maximum AE of only 2. This observation shows the high accuracy and adaptability of the proposed RUL prediction method.

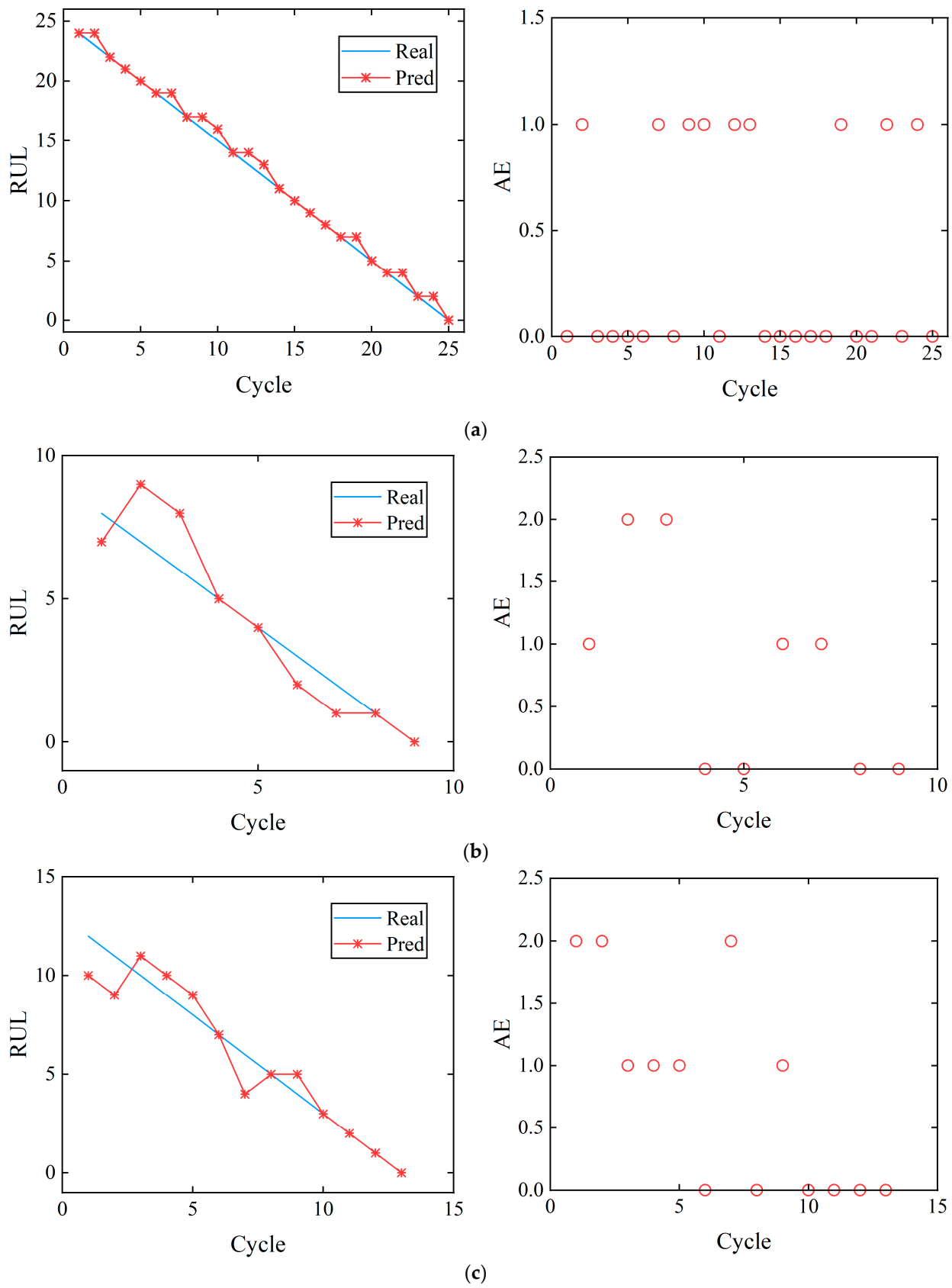
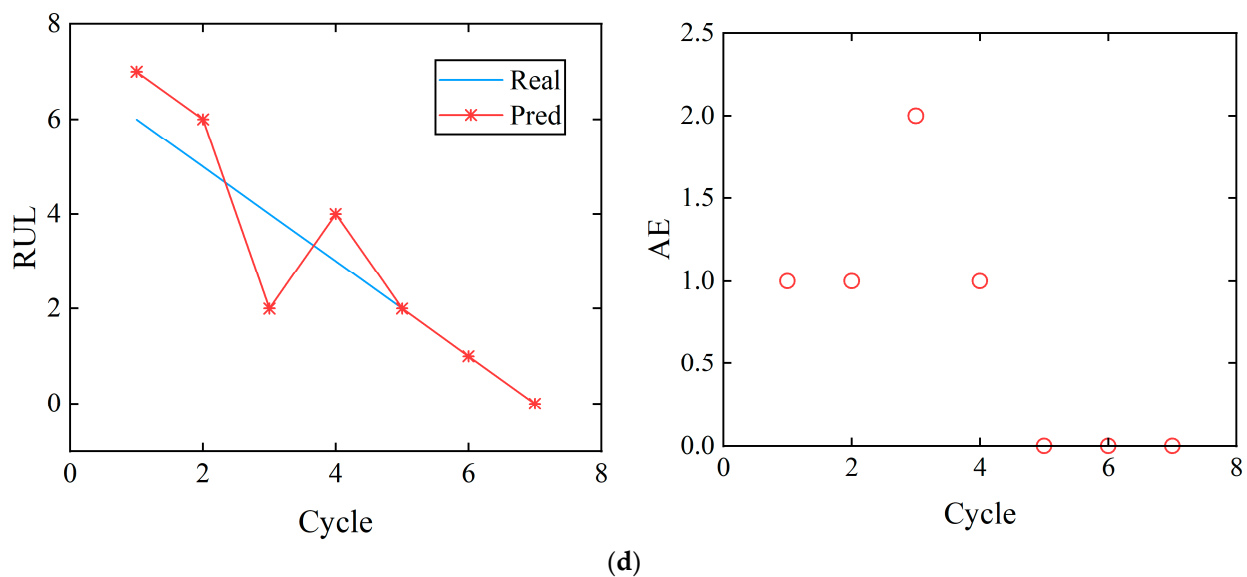


Figure 8. Cont.





**Figure 8.** Results of the RUL prediction: (a) B0005; (b) B0006; (c) Cell1; (d) Cell2.

Notably, the prediction accuracy of health features improves with more cycles, resulting in a decrease in the corresponding RUL estimation error. The overall errors are small in this paper, mainly because the multiscale prediction model can decompose time series signals into components at different frequencies and use the appropriate prediction model for targeted prediction. This method can effectively capture the feature details of the data and can well predict the trend of health features, thus greatly improving the precision and robustness of the RUL prediction method.

Table 5 shows the MAE and RMSE calculation results. Wherein the battery B0005 starts from cycle 101 until it reaches EOL, and the multistep SOH prediction is performed for 25 cycles to obtain the RUL prediction results, with a prediction error of 0.3600 for MAE and 0.6000 for RMSE. Compared to the B0005 battery, the error metrics of the other three batteries are slightly larger, which is mainly due to the poor results caused by the smaller amount of data. However, the MAE of all batteries is below 0.8 and the highest RMSE value is 1.1094, suggesting that the RUL prediction method proposed in this paper can achieve reliable estimation despite the limited data available.

**Table 5.** Results of RUL prediction error.

Evaluation Metrics	B0005	B0006	Cell1	Cell2
MAE	0.3600	0.7778	0.7692	0.7143
RMSE	0.6000	1.1055	1.1094	1.0000

#### 4.4. Results of the Joint SOH and RUL Evaluation

Both SOH and RUL can be used to evaluate the aging state of a battery. However, SOH is evaluated based on the current battery capacity state, while RUL is evaluated based on the number of cycles remaining in the battery. During battery use, the presence of capacity regeneration may cause the battery to have the same SOH in different cycling cycles. However, this does not mean that the battery has recovered to the previous degree of aging at the corresponding SOH, and therefore further judgment is required. Given the more obvious phenomenon of capacity regeneration in B0005, two points with the same SOH but different RUL in B0005 were selected for comparative analysis.

As shown in Table 6, the remaining usable capacity of the B0005 battery is comparable between cycles 111 and 121, but the RUL of the two differ by 10 cycles. This indicates that although SOH can reflect the current capacity state of the battery, the degree of battery aging cannot be comprehensively and accurately assessed by SOH alone. The joint estimation

method proposed in this paper can simultaneously predict the SOH and RUL of the battery at the current moment, and the prediction results are both well-fitted to the real values with high prediction accuracy and reliability. Combining SOH and RUL estimation results helps to assess battery health more comprehensively and accurately.

**Table 6.** Results of the joint SOH and RUL evaluation.

Current Number of Cycles	Real Capacity (Ah)	Estimated Capacity (Ah)	Estimation Error (Ah)	Real RUL	Prediction of RUL	Prediction Error
111	1.4387	1.4382	0.0005	14	14	0
121	1.4383	1.4336	0.0047	4	4	0

## 5. Conclusions

This paper proposes a method that utilizes charging feature extraction and ridge regression to estimate SOH and predict RUL of lithium-ion batteries. By extracting three sets of health features that can reflect battery degradation from the charging voltage curve. These health features are used as input and SOH as output, leading to the development of a battery aging model based on ridge regression. On this basis, a multiscale prediction model is designed to predict the changing trend of health features. The results of the multiscale prediction model are integrated with the battery aging model to estimate SOH through multiple steps, thus achieving an accurate prediction of RUL. The accuracy and reliability of the proposed method are confirmed using two battery datasets under varied operating conditions to validate its performance. Experimental results show that the prediction curves can approximate reality closely, the MAE and RMSE calculations of SOH remain below 0.02, and the maximum AE of RUL is no more than two cycles. By combining the SOH and RUL estimation results, a comprehensive and accurate prediction of battery aging can be obtained. Therefore, the proposed SOH and RUL prediction methods offer high accuracy and adaptability, serving as valuable references for SOH and RUL prediction of existing lithium-ion batteries.

However, the method has not yet been validated in other more complex working conditions, such as the impact of temperature factors on battery aging. Battery temperature changes can reduce the reliability of battery SOC estimation and can even pose a risk of thermal runaway; accurate battery SOC estimation in temperature-varying environments can avoid battery thermal runaway [39]. This may result in the inability to provide a more comprehensive reflection of the battery state. Therefore, in future work, the method needs to be further optimized to consider the temperature factor and perform joint estimation to adapt to more complex operating environments.

**Author Contributions:** Conceptualization, M.W. and C.Y.; methodology, C.Y. and F.Z.; software, C.Y.; validation, C.Y. and R.S.; formal analysis, J.T. and S.H.; investigation, N.Z. and J.W.; resources, M.W.; data curation, J.T. and S.H.; writing—original draft preparation, C.Y.; writing—review and editing, F.Z. and R.S.; project administration, N.Z. and J.W.; funding acquisition, M.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Science and Technology Project of Hubei Province (No. 2022BEC017), Natural Science Foundation of Hubei Province (No. 2022CFA007), and Science and Technology Project of Hubei Province (No. 2022BEB016).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available at: <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoc/pcoc-data-set-repository> (accessed on 2 February 2024) and <https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac> (accessed on 2 February 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Wang, G.; Lyu, Z.; Li, X. An Optimized Random Forest Regression Model for Li-Ion Battery Prognostics and Health Management. *Batteries* **2023**, *9*, 332. [\[CrossRef\]](#)
- Zhang, M.; Chen, W.; Yin, J.; Feng, T. Health Factor Extraction of Lithium-Ion Batteries Based on Discrete Wavelet Transform and SOH Prediction Based on CatBoost. *Energies* **2022**, *15*, 5331. [\[CrossRef\]](#)
- Liu, H.; Li, Y.; Luo, L.; Zhang, C. A Lithium-Ion Battery Capacity and RUL Prediction Fusion Method Based on Decomposition Strategy and GRU. *Batteries* **2023**, *9*, 323. [\[CrossRef\]](#)
- Deng, L.; Shen, W.; Wang, H.; Wang, S. A rest-time-based prognostic model for remaining useful life prediction of lithium-ion battery. *Neural Comput. Appl.* **2020**, *33*, 2035–2046. [\[CrossRef\]](#)
- Tian, H.; Qin, P.; Li, K.; Zhao, Z. A review of the state of health for lithium-ion batteries: Research status and suggestions. *J. Clean. Prod.* **2020**, *261*, 120813. [\[CrossRef\]](#)
- Lipu, M.S.H.; Hannan, M.A.; Hussain, A.; Hoque, M.M.; Ker, P.J.; Saad, M.H.M.; Ayob, A. A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations. *J. Clean. Prod.* **2018**, *205*, 115–133. [\[CrossRef\]](#)
- Xu, J.; Sun, C.; Ni, Y.; Lyu, C.; Wu, C.; Zhang, H.; Yang, Q.; Feng, F. Fast Identification of Micro-Health Parameters for Retired Batteries Based on a Simplified P2D Model by Using Padé Approximation. *Batteries* **2023**, *9*, 64. [\[CrossRef\]](#)
- Fan, Y.; Xiao, F.; Li, C.; Yang, G.; Tang, X. A novel deep learning framework for state of health estimation of lithium-ion battery. *J. Energy Storage* **2020**, *32*, 101741. [\[CrossRef\]](#)
- Amogne, Z.E.; Wang, F.-K.; Chou, J.-H. Transfer Learning Based on Transferability Measures for State of Health Prediction of Lithium-Ion Batteries. *Batteries* **2023**, *9*, 280. [\[CrossRef\]](#)
- Lin, C.-P.; Cabrera, J.; Yang, F.; Ling, M.-H.; Tsui, K.-L.; Bae, S.-J. Battery state of health modeling and remaining useful life prediction through time series model. *Appl. Energy* **2020**, *275*, 115338. [\[CrossRef\]](#)
- Roman, D.; Saxena, S.; Robu, V.; Pecht, M.; Flynn, D. Machine learning pipeline for battery state-of-health estimation. *Nat. Mach. Intell.* **2021**, *3*, 447–456. [\[CrossRef\]](#)
- Chun, H.; Kim, J.; Kim, M.; Lee, J.; Lee, T.; Han, S. Capacity Estimation of Lithium-Ion Batteries for Various Aging States Through Knowledge Transfer. *IEEE Trans. Transp. Electr.* **2022**, *8*, 1758–1768. [\[CrossRef\]](#)
- Zhang, C.; Allafi, W.; Dinh, Q.; Ascencio, P.; Marco, J. Online estimation of battery equivalent circuit model parameters and state of charge using decoupled least squares technique. *Energy* **2018**, *142*, 678–688. [\[CrossRef\]](#)
- Misyris, G.S.; Tegnér, T.; Marinopoulos, A.G.; Doukas, D.I.; Labridis, D.P. Battery energy storage systems modeling for online applications. In Proceedings of the 2017 IEEE Manchester PowerTech, Manchester, UK, 18–22 June 2017; pp. 1–6.
- Wang, Z.; Zeng, S.; Guo, J.; Qin, T. State of health estimation of lithium-ion batteries based on the constant voltage charging curve. *Energy* **2019**, *167*, 661–669. [\[CrossRef\]](#)
- Qiu, X.; Wu, W.; Wang, S. Remaining useful life prediction of lithium-ion battery based on improved cuckoo search particle filter and a novel state of charge estimation method. *J. Power Sources* **2020**, *450*, 227700. [\[CrossRef\]](#)
- Bole, B.; Kulkarni, C.S.; Daigle, M. Adaptation of an Electrochemistry-based Li-Ion Battery Model to Account for Deterioration Observed Under Randomized Use. In Proceedings of the Annual Conference of the PHM Society, Fort Worth, TX, USA, 29 September–2 October 2014; Volume 6.
- Khodadadi Sadabadi, K.; Jin, X.; Rizzoni, G. Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health. *J. Power Sources* **2021**, *481*, 228861. [\[CrossRef\]](#)
- Hong, J.; Lee, D.; Jeong, E.-R.; Yi, Y. Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning. *Appl. Energy* **2020**, *278*, 115646. [\[CrossRef\]](#)
- Zhang, X.; Sun, J.; Shang, Y.; Ren, S.; Liu, Y.; Wang, D. A novel state-of-health prediction method based on long short-term memory network with attention mechanism for lithium-ion battery. *Front. Energy Res.* **2022**, *10*, 972486. [\[CrossRef\]](#)
- Li, Y.; Liu, K.; Foley, A.M.; Zülke, A.; Bercibar, M.; Nanini-Maury, E.; Van Mierlo, J.; Hoster, H.E. Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109254. [\[CrossRef\]](#)
- Zhang, D.; Zhao, W.; Wang, L.; Chang, X.; Li, X.; Wu, P. Evaluation of the State of Health of Lithium-Ion Battery Based on the Temporal Convolution Network. *Front. Energy Res.* **2022**, *10*, 929235. [\[CrossRef\]](#)
- Wu, J.; Zhang, C.; Chen, Z. An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks. *Appl. Energy* **2016**, *173*, 134–140. [\[CrossRef\]](#)
- Khumprum, P.; Yodo, N. A Data-Driven Predictive Prognostic Model for Lithium-Ion Batteries based on a Deep Learning Algorithm. *Energies* **2019**, *12*, 660. [\[CrossRef\]](#)
- Chen, L.; Zhang, Y.; Zheng, Y.; Li, X.; Zheng, X. Remaining useful life prediction of lithium-ion battery with optimal input sequence selection and error compensation. *Neurocomputing* **2020**, *414*, 245–254. [\[CrossRef\]](#)
- Zhang, S.; Zhai, B.; Guo, X.; Wang, K.; Peng, N.; Zhang, X. Synchronous estimation of state of health and remaining useful lifetime for lithium-ion battery using the incremental capacity and artificial neural networks. *J. Energy Storage* **2019**, *26*, 100951. [\[CrossRef\]](#)
- Deng, Y.; Ying, H.; Jiaqiang, E.; Zhu, H.; Wei, K.; Chen, J.; Zhang, F.; Liao, G. Feature parameter extraction and intelligent estimation of the State-of-Health of lithium-ion batteries. *Energy* **2019**, *176*, 91–102. [\[CrossRef\]](#)
- Chen, Y.; Duan, W.; Ding, Z.; Li, Y. Battery Life Prediction Based on a Hybrid Support Vector Regression Model. *Front. Energy Res.* **2022**, *10*, 899804. [\[CrossRef\]](#)

29. Li, R.; Li, W.; Zhang, H.; Zhou, Y.; Tian, W. On-Line Estimation Method of Lithium-Ion Battery Health Status Based on PSO-SVM. *Front. Energy Res.* **2021**, *9*, 693249. [\[CrossRef\]](#)
30. Tian, J.; Xiong, R.; Shen, W. State-of-Health Estimation Based on Differential Temperature for Lithium Ion Batteries. *IEEE Trans. Power Electron.* **2020**, *35*, 10363–10373. [\[CrossRef\]](#)
31. Wu, J.; Cui, X.; Zhang, H.; Lin, M. Health Prognosis With Optimized Feature Selection for Lithium-Ion Battery in Electric Vehicle Applications. *IEEE Trans. Power Electron.* **2021**, *36*, 12646–12655. [\[CrossRef\]](#)
32. Zhao, L.; Wang, Y.; Cheng, J. A hybrid method for remaining useful life estimation of lithium-ion battery with regeneration phenomena. *Appl. Sci.* **2019**, *9*, 1890. [\[CrossRef\]](#)
33. Zhang, L.; Ji, T.; Yu, S.; Liu, G. Accurate Prediction Approach of SOH for Lithium-Ion Batteries Based on LSTM Method. *Batteries* **2023**, *9*, 177. [\[CrossRef\]](#)
34. Birkl, C. *Diagnosis and Prognosis of Degradation in Lithium-Ion Batteries*; University of Oxford: Oxford, UK, 2017.
35. Guo, P.; Cheng, Z.; Yang, L. A data-driven remaining capacity estimation approach for lithium-ion batteries based on charging health feature extraction. *J. Power Sources* **2019**, *412*, 442–450. [\[CrossRef\]](#)
36. Wang, X.; Wang, X.; Ma, B.; Li, Q.; Shi, Y.-Q. High Precision Error Prediction Algorithm Based on Ridge Regression Predictor for Reversible Data Hiding. *IEEE Signal Process. Lett.* **2021**, *28*, 1125–1129. [\[CrossRef\]](#)
37. Mao, L.; Xu, J.; Chen, J.; Zhao, J.; Wu, Y.; Yao, F. A LSTM-STW and GS-LM Fusion Method for Lithium-Ion Battery RUL Prediction Based on EEMD. *Energies* **2020**, *13*, 2380. [\[CrossRef\]](#)
38. Chen, Z.; Li, L.; Cui, W.; Yang, S.; Wang, Y.; Wang, D. Remaining useful life prognostics of lithium-ion batteries based on a coordinate reconfiguration of degradation trajectory and multiple linear regression. *Front. Energy Res.* **2023**, *10*, 1033039. [\[CrossRef\]](#)
39. Zhang, R.; Li, X.; Sun, C.; Yang, S.; Tian, Y.; Tian, J. State of Charge and Temperature Joint Estimation Based on Ultrasonic Reflection Waves for Lithium-Ion Battery Applications. *Batteries* **2023**, *9*, 335. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.