



Article

Method of Predicting SOH and RUL of Lithium-Ion Battery Based on the Combination of LSTM and GPR

Jiahui Zhao, Yong Zhu *, Bin Zhang, Mingyi Liu, Jianxing Wang, Chenghao Liu and Yuanyuan Zhang

China Huaneng Group Clean Energy Research Institute (CERI), Beijing 102209, China

* Correspondence: y_zhu@qny.chng.com.cn

Abstract: The state of health and remaining useful life of lithium-ion batteries are important indicators to ensure the reliable operation of these batteries. However, because they cannot be directly measured and are affected by many factors, they are difficult to predict. This paper presents method of jointly predicting state of health and RUL based on the long short-term memory neural network and Gaussian process regression. This method extracts the batteries' health factors from the charging curve, selects health factors with more relevance than the setting standard as the characteristic of capacity by the maximum information coefficient method, and establishes the battery aging and remaining useful life prediction models with Gaussian process regression. On this basis, the long short-term memory neural network is used to predict the trend of the change in health factors with the increase in cycles, and the results are input into a Gaussian process regression aging model to predict the state of health. Taking the health factors and state of health as the characteristics of remaining useful battery life, a battery remaining useful life model based on Gaussian process regression is established, and the change trend in the remaining useful life can be obtained by inputting the predicted health factors and state of health. In this study, four battery data sets with different depths of charge were used to verify the accuracy and adaptability of the algorithm. The results show that the proposed algorithm has high accuracy and reliability.



Citation: Zhao, J.; Zhu, Y.; Zhang, B.; Liu, M.; Wang, J.; Liu, C.; Zhang, Y. Method of Predicting SOH and RUL of Lithium-Ion Battery Based on the Combination of LSTM and GPR. *Sustainability* **2022**, *14*, 11865. <https://doi.org/10.3390/su141911865>

Academic Editor: Jifeng Song

Received: 9 August 2022

Accepted: 13 September 2022

Published: 21 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 have many advantages, such as high energy density, low self-discharge, and long service life, and have been applied in many fields [1,2]. However, with the continuous charging and discharging process of lithium-ion batteries, the performance of these batteries deteriorates, e.g., through a reduction in capacity and an increase in impedance, which may create internal short circuit, thermal runaway, or other problems, leading to equipment and system failures, and even catastrophic accidents [3]. Therefore, it is very important to accurately estimate and predict the battery state to improve the reliability of the battery. State of health (SOH) and remaining useful life (RUL) are important indicators used to characterize the degree of battery aging [4].

At present, SOH and RUL estimation methods are mainly divided into model-based methods and data-driven methods. Model-based methods include the equivalent circuit model and the electrochemical model, according to different modeling mechanisms. According to the physical and chemical reactions inside the battery, the electrochemical model deduces the degradation mechanism of battery performance in relation to the mechanism, such as lithium-ion loss, active material loss, and conductivity loss.

The pseudo two-dimensional (P2D) [5] model describes the internal dynamic mechanism of lithium-ion batteries through a series of partial differential equations (PDE). It can accurately estimate the battery state. This model is suitable for batteries with different materials and can be developed and extended to more complex multifield coupling

models. However, for complex models such as batteries, find analytical solutions with PDE is difficult [6]. Generally, the single particle model (SPM) can be adopted, or PDE can be reduced to ordinary differential equations to reduce the amount of calculation and the complexity of the model. Reference [7] established mathematical equations based on porous electrode theory and concentrated solution theory to describe the internal physical and chemical behaviors of a power battery in the process of charging and discharging, and meshed and discretized the PDE using the finite analysis method to reduce their order to that of an ordinary differential equation. Reference [8] established an electrochemical thermal coupling model, reduced the order of the electrochemical model, and estimated the battery SOH based on the model parameters. However, simplifying the model also reduces its accuracy, so it is worth considering how to balance the amount of calculation and the accuracy of the model [9–12].

The equivalent circuit model ignores the complex physical and chemical reactions inside the battery, and simulates the output characteristics of the battery through a series parallel circuit of resistance and capacitance. Reference [13] compared and analyzed the complexity, accuracy, and robustness of 12 equivalent circuit models in SOH estimation. Based on the second-order equivalent circuit model, Reference [14] proposed a new SOC-SOH relation function, which can be used for capacity update in SOC estimation. However, the equivalent circuit model usually has poor accuracy and generalizability, and is only suitable for limited operating conditions.

Some scholars have used empirical models to describe the capacity deterioration curve, and use the Kalman filter (KF), particle filter (PF) [15], and its improved forms, such as the unscented Kalman filter (UKF), adaptive unscented Kalman filter (AUKF) [16], interactive multi model particle filter, or other extrapolated models [17], to estimate SOH and predict RUL. However, empirical models cannot reflect the impact of operating conditions on battery life. The semiempirical model, established according to the aging mechanism, can take into account the impact of different service conditions on battery aging performance. It is mainly established based on external factors, including the operating environment and working condition of the battery, such as temperature, charge and discharge ratio, depth of discharge (DOD), charge and discharge cut-off voltage [18], etc. Generally, the Arrhenius, inverse power law, and double exponential formulas are used as the basis of the semiempirical model. Reference [19] studied the aging characteristics of LiFePO₄ batteries, and established a generalized battery life model based on ampere hour throughput, charge and discharge rates, and temperature with the Arrhenius formula. Reference [20] established a three-parameter capacity decline model based on the double exponential formula, and the result showed that the model had a better prediction effect than the double exponential formula. In addition to temperature and magnification, Reference [21] considered the influence of cut-off voltage on LFP battery capacity and internal resistance, explored the coupling effect of two influencing factors on battery aging, and established a multi-influence factor coupling the life prediction model combined with the Arrhenius formula and the inverse power law formula. However, semiempirical models cannot reflect the complexity of the internal electrochemical reaction and degradation mechanism of the battery, which leads to the low prediction accuracy of the semiempirical model in the late stage of battery decline.

Data-driven methods can be used to directly mine the deterioration information and the evolution law of the health state of lithium-ion batteries through historical data, without establishing a clear model formula. Existing methods include neural networks and regression methods, such as the artificial neural network (ANN), support vector machine (SVM), autoregressive model (AR model), etc. Reference [22] predicted the battery RUL online with the AR model improved by the particle swarm optimization algorithm. The calculation of the AR model is simple, but the prediction result has no expression of uncertainty. Reference [23] proposed the RUL prediction method based on the autoregressive integrated moving average (ARIMA) model. However, the ARIMA model requires the stability of time series data and has high requirements for battery

operating conditions. Reference [24] considered a dynamically driven recurrent network (DDRN) to reduce the complexity of the network structure and improve the robustness of the algorithm. However, DDRN has the problem of gradient disappearance. Reference [25] established a regression model with an SVM algorithm to estimate battery RUL by training battery terminal voltage and voltage derivative during charging. However, due to its own characteristics, the SVM algorithm easily falls into local minima.

Gaussian process regression (GPR) is a probability prediction method based on a Bayesian framework. It has the characteristics of a nonparametric expression and can be used to solve regression problems with large dimensions, a small amount of data, and nonlinearity [26,27]. GPR is simpler than neural network algorithms, and can greatly reduce the amount of calculation required for RUL prediction.

SOH and RUL predictions based on data-driven methods can be divided into direct and indirect methods. Based on the trend in the capacity decline, the former takes the curve as a feature to directly predict the change in battery residual capacity. Due to the different trends in batteries in different recession periods, this method cannot accurately predict the recession trend of the whole life cycle of batteries. Indirect data-driven methods can achieve better a prediction effect by extracting health factors (HFs) that can characterize the degree of battery aging and taking them as the features to predict battery RUL. Reference [28] extracted the constant current (CC) charging time and constant voltage (CV) charging time as the HFs. Reference [29] selected the time required for the voltage change in the same time interval during the discharge process and the same temperature change during the discharge process to predict battery RUL. Reference [30] extracted the time when the voltage dropped from 3.9 to 3.5 V during the discharge process and the time required for the temperature to rise to the maximum. Due to the large uncertainty in the discharge conditions during actual operation, the charging curve is relatively stable, so extracting features from it is easy.

Because simple one-to-one mapping is not possible between the remaining available capacity, or SOH, and the number of battery cycles due to the capacity regeneration phenomenon, or RUL, traditional data-driven methods are not applicable. Moreover, SOH and RUL are strongly correlated; thus, separately estimating them may reduce the prediction accuracy. To fully reflect the health status of a battery, it is necessary to comprehensively diagnose the current SOH and RUL of the battery to solve the problem of no one-to-one mapping for both SOH and RUL and the problem of the lack of consideration of the correlation between SOH and RUL. However, commonly used methods often estimate only one of them, and separately estimating the two increases the amount of computation and the algorithm complexity.

Moreover, to provide instruction for maintenance and fault alarms, it is necessary to predict the trend of the change in batteries' SOH and RUL in the future. SOH and RUL cannot be sufficiently predicted with direct data-driven methods; however, they can be predicted by predicting the change trend of HFs in the future. Based on the above reasons, in this study, we developed a data-driven prediction method combining GPR and long short-term memory (LSTM) to utilize both advantages of the regression and indirect data-driven methods. Specifically, we extracted HFs from the charging curve of the battery, analyzed the correlation between HFs and capacity with the maximum information coefficient (MIC), and selected the HFs with high correlation. Through LSTM, the change trend of HFs in the aging process was predicted, and the predicted HFs were used as the input and capacity as the output of Gaussian process regression model, to accurately estimate RUL.

The remainder of this paper is structured as follows: we first introduce the extraction and selection of HFs, then define the overall procedure of SOH and RUL estimation. We then describe the verification results of the method, and finally provide our conclusions.

2. Health Factor Selection and Optimization

2.1. Extracting Health Factors

We selected 4 battery data sets from the lithium-ion battery test data published by NASA to verify the method due to the limited experiment conditions [31]. The battery parameters and operating conditions are shown in Table 1.

Table 1. Battery parameters and operating conditions.

Battery Index	CC Charge Current (A)	Rated Voltage (V)	Charging Cut-Off Current (mA)	Discharge Current (A)	Discharging Cut-Off Current (V)	Rated Capacity (Ah)
B0005	1.5	4.2	20	2	2.7	2
B0006	1.5	4.2	20	2	2.5	2
B0007	1.5	4.2	20	2	2.2	2
B0018	1.5	4.2	20	2	2.5	1.86

We only extracted HFs in the process of battery charging. When charging the battery, CC charging is first carried out, and CV charging is started after the voltage rises to the charging cut-off voltage. Then, the battery keeps CV charging until the current drops to the charging cut-off current. We selected six health characteristics from the charging current and voltage curves, including:

1. HF1: the time of voltage rising from 3.9 to 4.2 V during CC charging;
2. HF2: voltage increase within 600 s after reaching 3.9 V in CC charging process;
3. HF3: current reduction within 900 s after entering CV charging process.

Because a battery can be represented with an 1st-order RC equivalent circuit and the response of the equivalent circuit compies with exponential expression, after the battery starts CV charging, the fitting curve expression of the current is follows:

$$I = HF_4 + HF_5 e^{-HF_6/t} \quad (1)$$

where I denotes the current; HF_4 – HF_6 denote the selected HFs, which are related to the battery time constant.

Figures 1 and 2 show the current and voltage curves of a B0005 battery at the 40th, 80th, 120th, and 160th cycles. The capacity curve of the battery is shown in Figure 3. HF_1 – HF_6 after standardization are shown in Figure 4. During the cycle of battery charging and discharging, due to the side reaction between the electrode and electrolyte, lithium ions are continuously consumed, and the capacity shows a downward trend. However, during repeated processes of charging and discharging, the side reaction products may be counteracted to some extent. Therefore, compared with the previous cycle, the battery performance of the next cycle is better, and the capacity relatively increases. This phenomenon is called capacity regeneration, which causes the local dynamic fluctuations [32,33] in capacity.

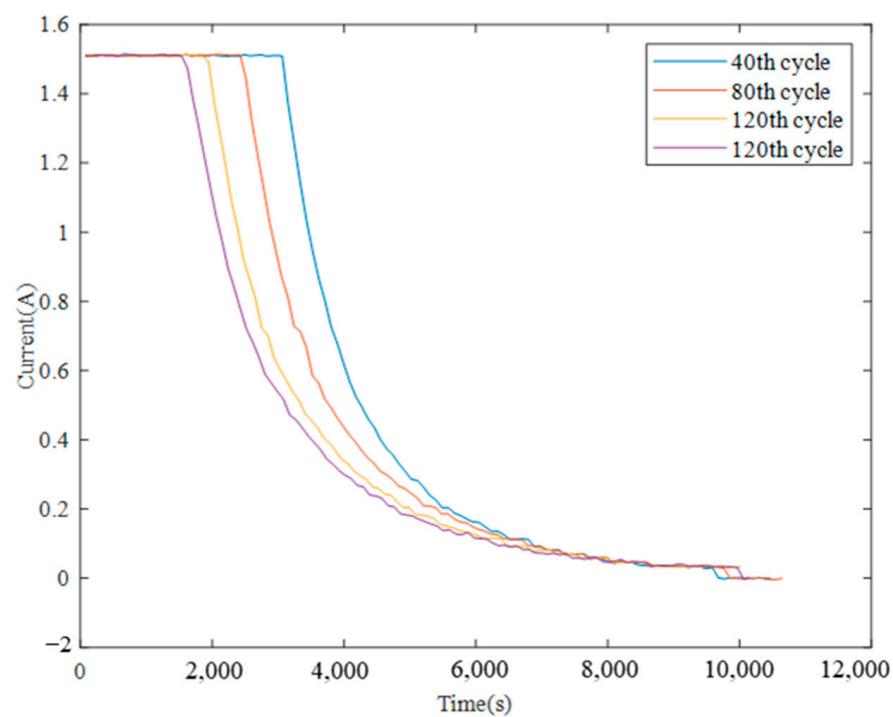


Figure 1. Battery current curve of charging process.

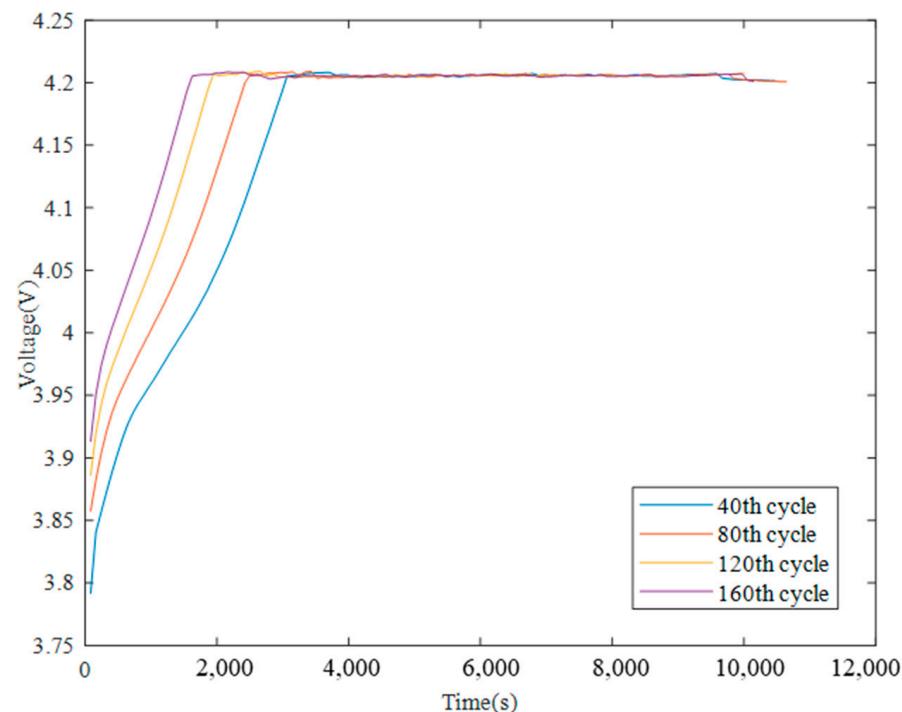


Figure 2. Battery voltage curve of charging process.

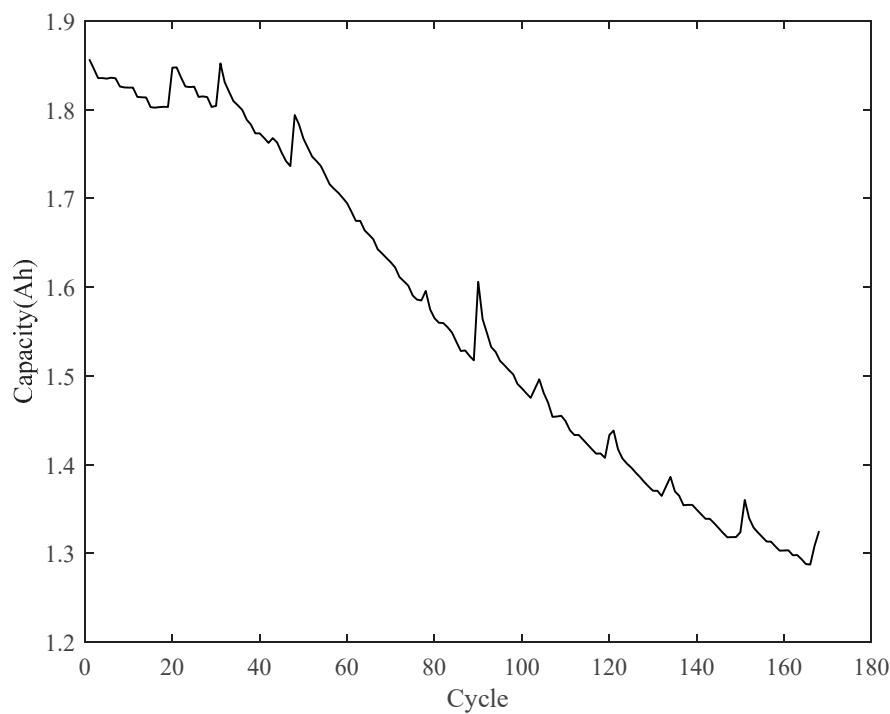


Figure 3. Battery capacity attenuation curve.

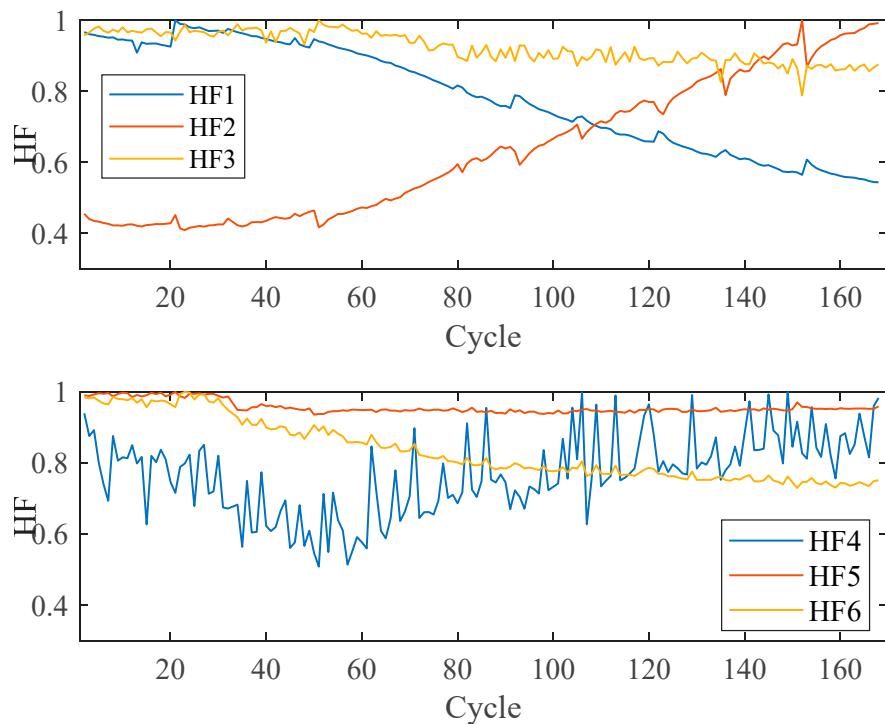


Figure 4. Curves of HF_1 – HF_6 during charging process.

2.2. Selecting Health Factors

As shown in Figure 4, the change trends of each HF are different, so it is difficult to directly determine the correlation between HFs and the capacity. Therefore, we used the maximum information coefficient (MIC) to measure the correlation between HFs and the capacity.

The basic principle of the MIC is that when calculating the correlation between two variables, the two variables are meshed on the scatter diagram formed in two-dimensional

space; then, the joint approximate probability density of the two variables is obtained according to the mesh, to obtain the mutual information value between the two variables for normalization. There are corresponding normalized mutual information values at different grid numbers and positions, and the maximum value is the MIC. The advantage of the MIC lies in its generality and equality. The generality means that when there are enough samples, it can identify a wide range of correlations, not limited to specific types, such as linear correlations, periodic functional correlations, parabolic correlations, and no-functional correlations. Equality means that for different correlation types, when adding the same degree of noise, the MICs between variables are similar. The MIC is evaluated as follows [34]:

$$mic(x, y) = \max_{a \times b < B} \frac{I(x; y)}{\log \min(a, b)} \quad (2)$$

where a and b are the number of meshing in the x and y directions, respectively; I is the mutual information value of variables x and y ; and B is set to the amount of data to the power of 0.6. For discrete variables X, Y , the scatter diagram composed of X, Y is meshed in i columns and j rows with given i, j , and the maximum mutual information value is calculated [35]:

$$MIC[x; y] = \max_{|X||Y| < B} \frac{I[X; Y]}{\log(\min(|X|, |Y|))} \quad (3)$$

Figure 5 shows the correlation analysis results between HF_1 – HF_6 and capacity. It can be seen that the correlations between the selected HF and capacity are different. Because variables are commonly regarded as highly correlated when correlation coefficients are larger than 0.8, HF with an MIC value greater than 0.8 were selected as the input of GPR model, that is, HF_1 , HF_2 , HF_3 , and HF_6 .

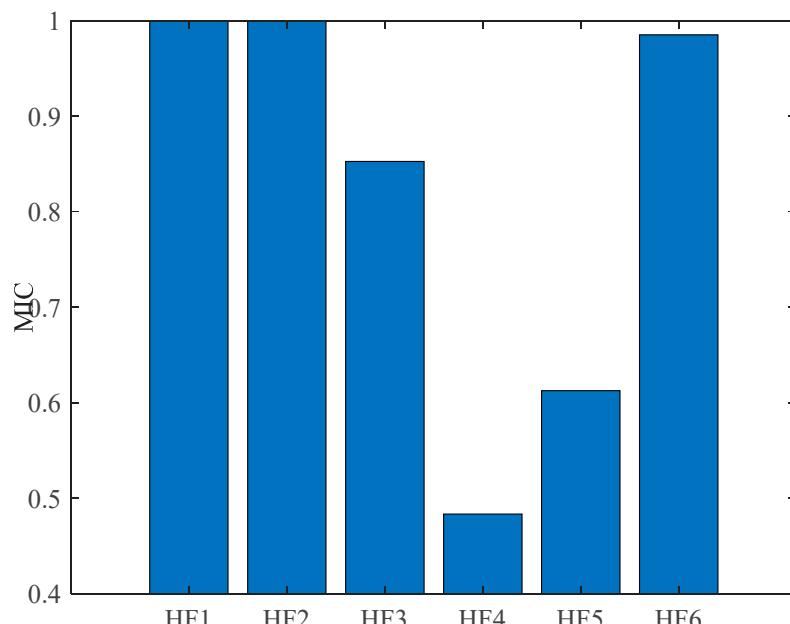


Figure 5. MICs of HF_1 – HF_6 .

3. SOH and RUL Prediction

3.1. GPR Model

The probability density function of univariate Gaussian distribution is expressed as follows [35]:

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4)$$

where σ^2 is the variance of random variables; μ is the mean value of random variables. Assuming that each dimension is independent of each other, the multivariate Gaussian distribution can be expressed as:

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i) \quad (5)$$

Therefore, the Gaussian regression process can be expressed as [27]:

$$f(x) \sim N(\mu(x), k(x, x)) \quad (6)$$

where $\mu(x)$ represents the mean function; $k(x, x')$ represents the covariance function, which is also known as the kernel function, and is expressed as [27]:

$$k(x, x') = \theta_0 e^{-\frac{(x-x')^2}{2\theta_1^2}} + \sigma^2 \delta_{ij} \quad (7)$$

where δ_{ij} is the Dirac function. When $i = j$, $\delta_{ij} = 1$; otherwise, it is 0.

The prior model of the Gaussian process is $f(x) \sim N(\mu(x), k(x, x))$. Supposing that this model is $f(x) \sim N(\mu_f, K_{ff})$, and the observed data (the given discrete data) are (\hat{x}, \hat{y}) and \hat{y} , and $f(x)$ satisfies the joint Gaussian distribution; then, the joint probability density formula can be expressed as [27]:

$$\begin{bmatrix} f(x) \\ \hat{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_f \\ \mu_y \end{bmatrix}, \begin{bmatrix} K_{ff} & K_{fy} \\ K_{fy}^T & K_{yy} \end{bmatrix}\right) \quad (8)$$

$$\begin{aligned} K_{ff} &= k(x, x) \\ K_{fy} &= k(x, \hat{x}) \\ K_{yy} &= k(\hat{x}, \hat{x}) \end{aligned} \quad (9)$$

where x is the independent variable that needs to be predicted; \hat{x} is the observed independent variable, which is already known. The following can be determined with Bayesian probability expression:

$$f \sim N(K_{fy}^T + \mu_f, K_{yy} - K_{fy}^T K_{ff}^{-1} K_{fy}) \quad (10)$$

Then, the predicted mean value is calculated as follows:

$$y_{mean} = K_{fy}^T K_{ff}^{-1} \hat{y} \quad (11)$$

The predicted error is calculated as follows:

$$y_\sigma = K_{yy} - K_{fy}^T K_{ff}^{-1} K_{fy} \quad (12)$$

If the super parameters θ_0 and θ_1 in the covariance function are determined, then the predicted value can be obtained.

The hyperparameters can be optimized by the maximum likelihood expression. The expression of maximum likelihood is as follows:

$$\begin{aligned} \log p(y|\theta_0, \theta_1) &= \log N(0, K_{yy}(\theta_0, \theta_1)) \\ &= -\frac{1}{2} y^T K_{yy}^{-1} y - \frac{1}{2} \log |K_{yy}| - \frac{N}{2} \log(2\pi) \end{aligned} \quad (13)$$

The partial derivative of the likelihood function can be solved by the conjugate gradient method:

$$\begin{cases} \frac{\partial}{\partial \theta_i} \log p(y|x, \theta) = \frac{1}{2} \left[\begin{bmatrix} \alpha \alpha^T - (K_{ff}(x, x))^{-1} \cdot \\ \frac{\partial (K_{ff}(x, x))}{\partial \theta_i} \end{bmatrix} \right] \\ \alpha = (K_{ff}(x, x))^{-1} y \end{cases} \quad (14)$$

3.2. LSTM

The LSTM cell is composed of an input gate, a forgetting gate, an output gate, and cell state.

1. Input gate determines how much of the input data of the network needs to be saved to the unit state at the current moment;
2. Forgetting gate determines how much unit state at the previous moment needs to be retained at the current moment;
3. Output gate controls how much current unit state needs to be output at the current moment.

The LSTM cell replaces the recurrent neural network (RNN) cell in the hidden layer with LSTM cells, so that it has long-term memory ability. After continuous evolution, the cell structure of the most widely used LSTM model is shown in Figure 6, and its forward calculation method can be expressed as:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (15)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (16)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (17)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (18)$$

$$h_t = o_t \tanh c_t \quad (19)$$

where i , f , c , and o are input gate, forgetting gate, cell state, and output gate, respectively; W and b are the corresponding weight coefficient matrix and offset, respectively; σ and \tanh are the sigmoid function and hyperbolic tangent activation function, respectively.

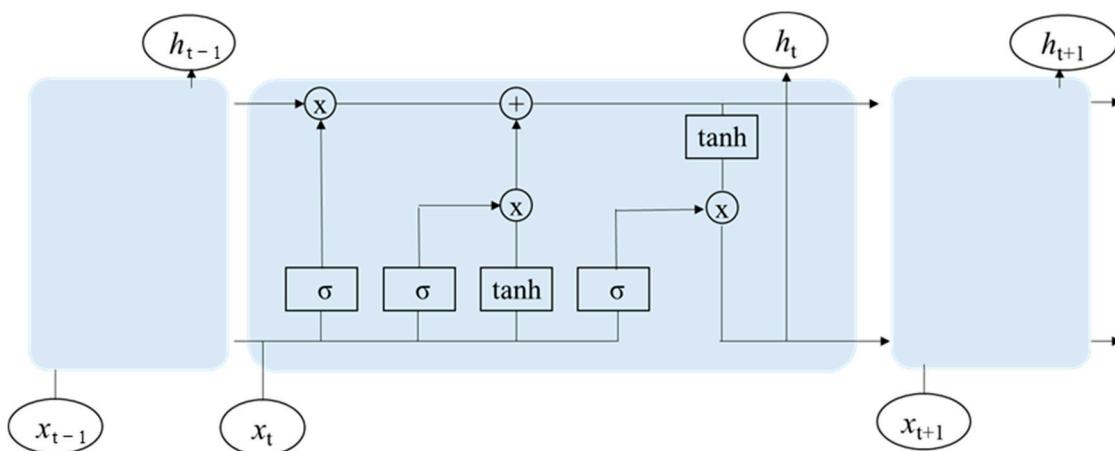


Figure 6. Principle of LSTM.

For the LSTM training process, the backpropagation through time (BPTT) algorithm is adopted, which is similar to the back propagation (BP) algorithm. There are roughly 4 steps:

1. Calculate the output value of LSTM cells according to the forward calculation method Formulas (15–19);
2. Calculate the error term of each LSTM cell backwards, with directions of time and network level;
3. Calculate the gradient of each weight value according to the corresponding error term;
4. The weight value is updated by optimization algorithm for gradient.

We used adaptive moment estimation (Adam) to optimize the gradient. Adam is a first-order optimization algorithm that can replace the traditional random gradient descent

process. It can iteratively update the weights of neural networks based on training data. The algorithm combines the first-order momentum of traditional SGDM (SGD with momentum) and the second-order momentum of AdaDelta. It not only has the advantage of adaptive learning, but also shows good convergence speed and effect.

3.3. Structure of SOH and RUL Prediction

After obtaining the measured HF series data of the first k cycles, the trend in the HFs with the number of cycles can be predicted with LSTM. Each predicted HF is input into the three GPR HF-LSTM models, and the corresponding SOH prediction value can be obtained through the GPR SOH model. Offline training is carried out by setting three standardized HFs and capacity as the input and output, respectively. The RUL prediction process of battery based on HFs is shown in Figure 7.

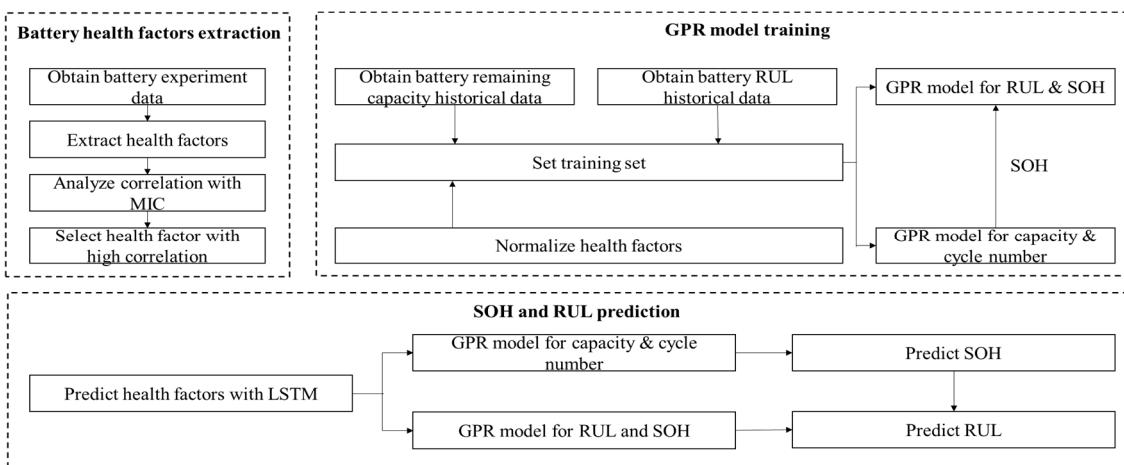


Figure 7. Structure of SOH and RUL prediction.

4. Results

In order to verify the prediction accuracy and reliability of the proposed algorithm for different batteries, we used the data of four batteries provided in Table 1. Mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE) were used as the indicators of algorithm performance:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (20)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (21)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (22)$$

where n is the length of training set; y_i is the experiment data of SOH or RUL; \hat{y}_i is prediction data of SOH or RUL.

In each cycle, the selected four HFs were extracted from the charging current, voltage curve. In order to verify the accuracy and reliability of the established battery aging model, the first half of the four batteries' data was used as the training set, and the second half was used as the test set. To improve the accuracy of prediction, when predicting with LSTM, the sliding window method was adopted; that is, certain cycle data were predicted based on the training set, and the obtained prediction results were added to the end of the training set to predict the next cycle data. Figure 8 shows the prediction results of the HFs based on LSTM, in which the black line is the real HF data, and the red line is the prediction data based on LSTM. The prediction results based on the LSTM sliding window prediction

method better followed the trend in HF, but there was still a certain deviation from the experimental values in the late stage of battery recession.

Using the four HFs predicted by LSTM as the input of the GPR model to predict the SOH of batteries, Figure 9 shows the prediction results and relative errors of batteries B0005, B0006, B0007, and B0018 from left to right. The black line is the experimental data, and the red line is the predicted data, which follows the experimental data well. Table 2 shows the predicted SOH of the four batteries. Although there was a certain deviation in the prediction of the HFs in the late period of battery recession, the GPR model could still maintain a high-precision prediction of SOH, including the local capacity regeneration part, and the prediction MAE was less than 0.01, which verified the effectiveness of the method.

Table 2. SOH prediction results of 4 batteries based on GPR.

Battery Index	MSE	RMSE	MAE
B0005	1.55×10^{-4}	0.012	0.007
B0006	1.62×10^{-4}	0.013	0.008
B0007	7.66×10^{-5}	0.009	0.005
B0018	2.11×10^{-4}	0.015	0.010

In order to explore the influence of different training set sizes on the effectiveness of the model, 50%, 60%, 70%, and 80% training set sizes were used to predict the SOH of B0005.

The prediction results are shown in Figure 10. From left to right are the prediction results with training set sizes of 60%, 70%, and 80% of the original data set length for battery B0005. It can be seen that the estimated values of the four batteries were quite close to the true values, and the corresponding SOH estimation error was smaller. When the training set was 80% of the original data set, the relative error of all points was within 1%.

RUL estimation is an estimation of the number of cycles left for the battery capacity to decay to EOL. In order to ensure the accurate prediction of the change in HFs and reduce the computational complexity as much as possible, we took the predicted HFs and SOH as the input of the GPR model and the remaining cycle number (RUL) as the output. The prediction results are shown in Figure 11. The black line is the real RUL and the red line is the estimated value.

It can be seen from the results that the change trend in the RUL prediction result basically conformed to the real RUL change trend. The relative prediction error was basically distributed within $\pm 0.5\%$ for B0005, $\pm 0.4\%$ for B0006, $\pm 1\%$ for B0007, and $\pm 1\%$ for B0018. Table 3 shows the calculation results of MSE, RMSE, and MAE. The MAE of each battery was less than one.

Table 3. RUL prediction results of 4 batteries based on GPR.

Battery Index	MSE	RMSE	MAE
B0005	9.4526×10^3	97.2244	0.2746
B0006	7.1663×10^3	84.6543	0.2550
B0007	7.4450×10^3	86.2844	0.3326
B0018	30.6684	5.5379	0.5789

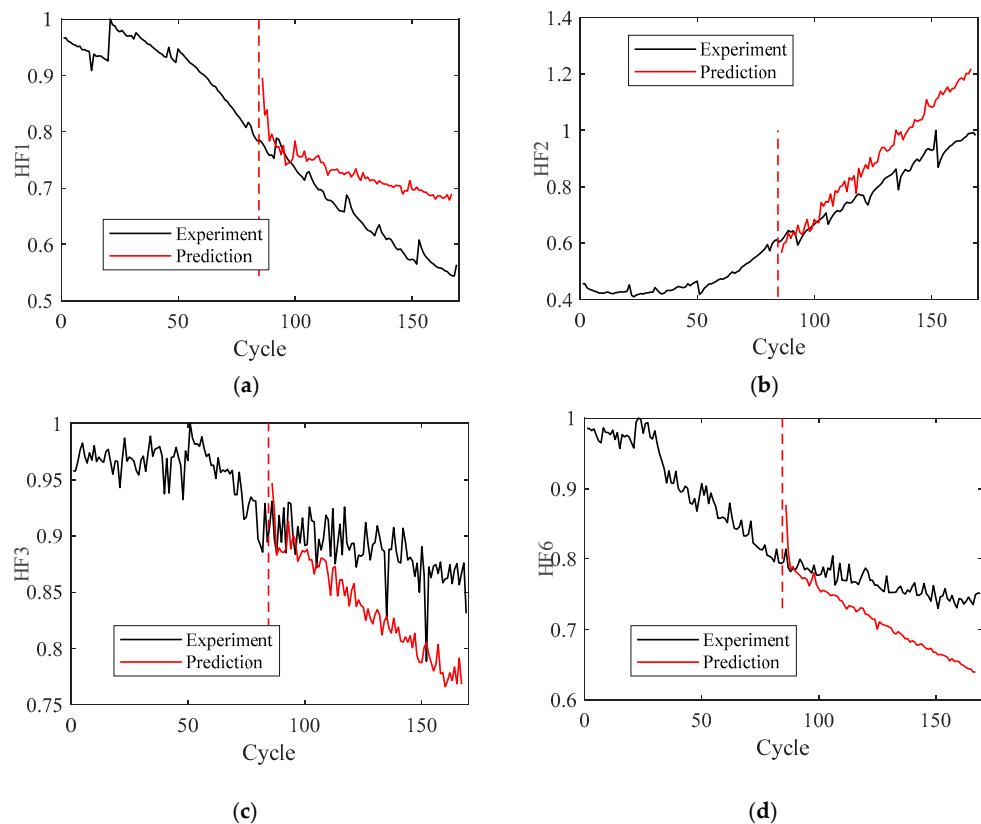


Figure 8. Prediction result of HFs based on LSTM: (a) prediction of HF_1 ; (b) prediction of HF_2 ; (c) prediction of HF_3 ; and (d) prediction of HF_6 .

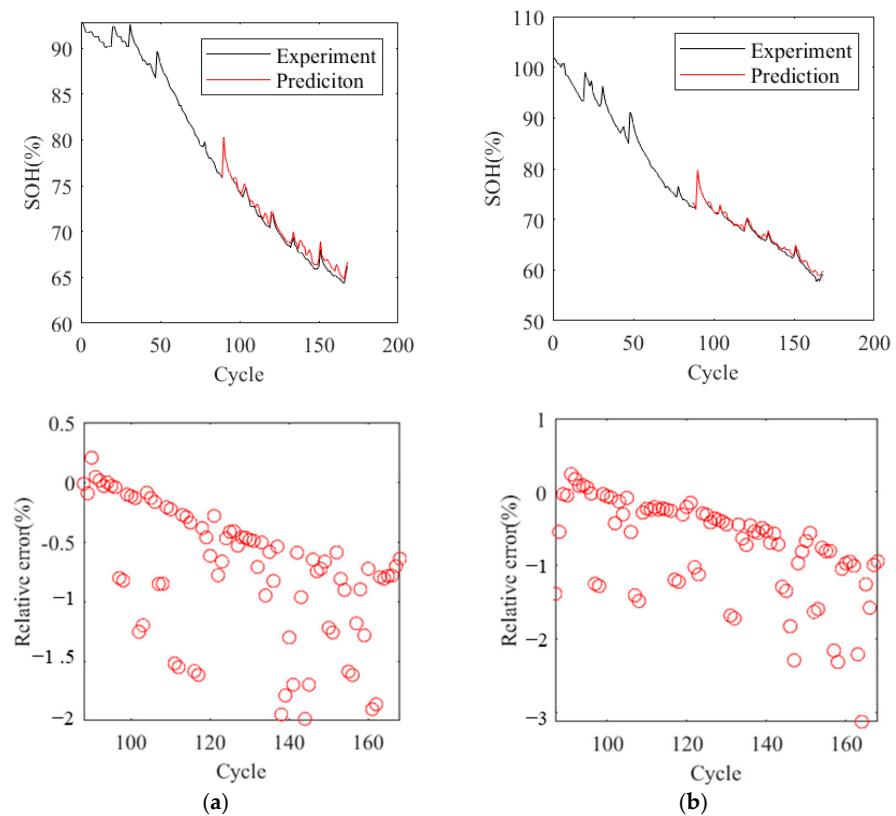


Figure 9. Cont.

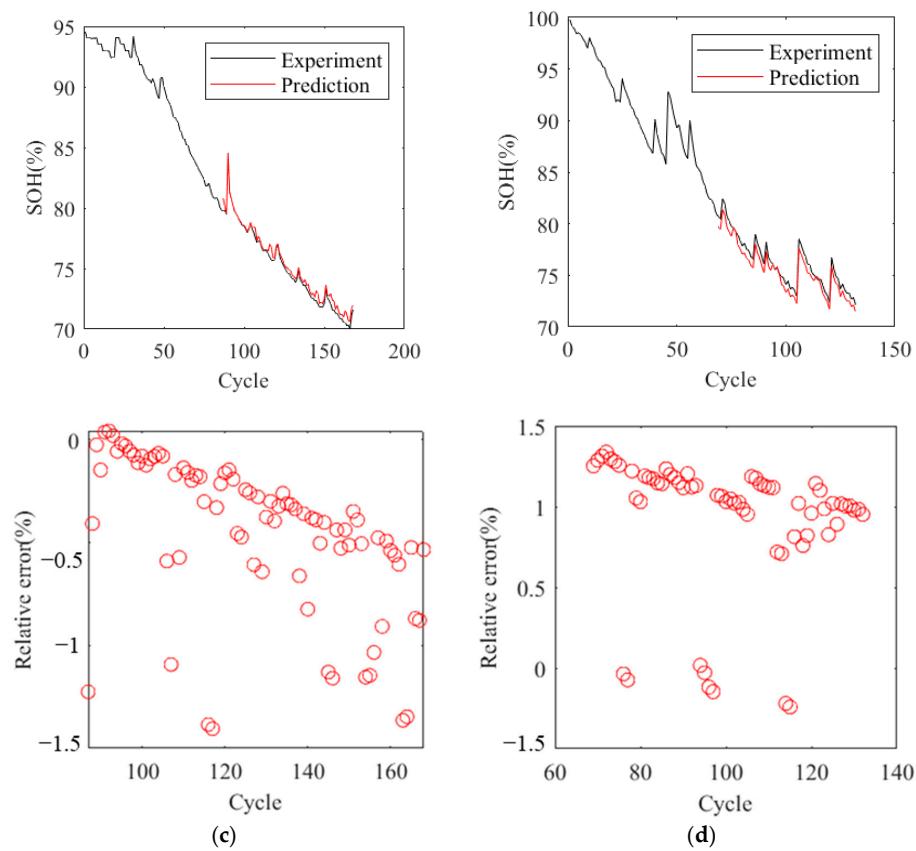


Figure 9. SOH prediction result of capacity for 4 batteries based on GPR: SOH prediction result of (a) B0005 and its relative error; (b) B0006 and its relative error; (c) B0007 and its relative error; (d) B0018 and its relative error.

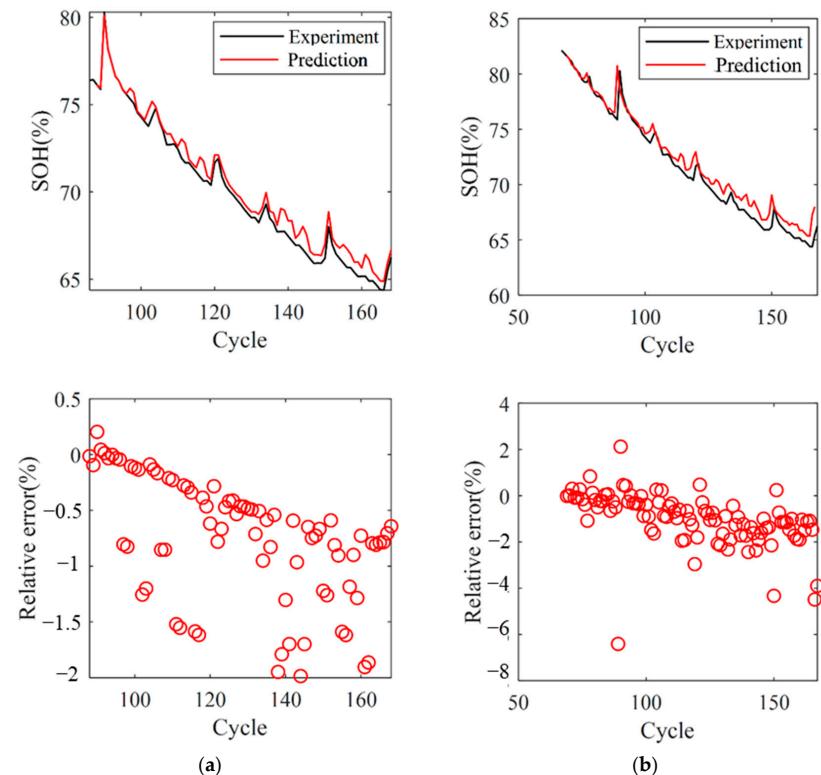


Figure 10. Cont.

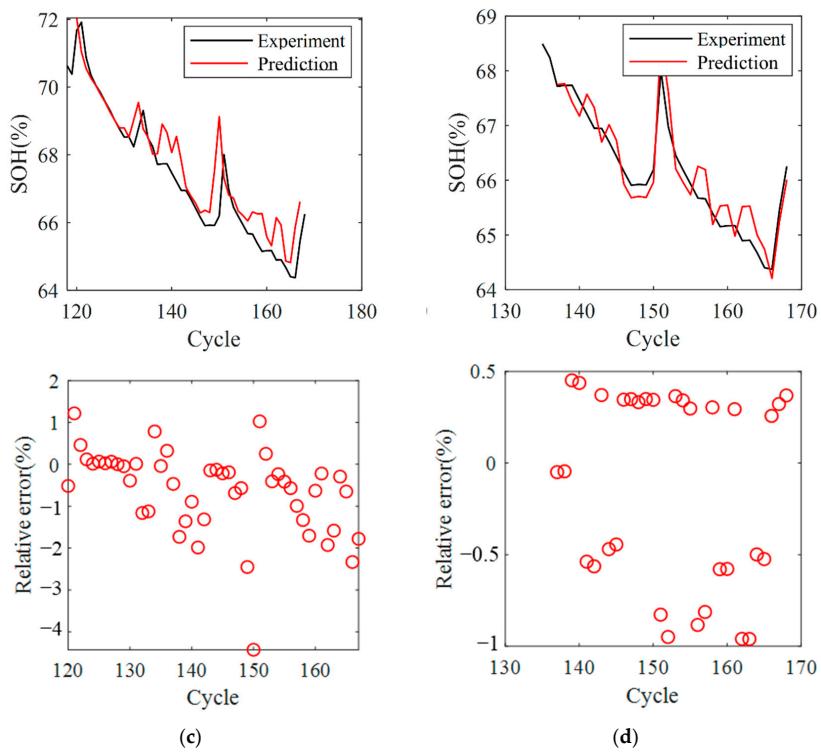


Figure 10. Prediction result of SOH for B0005 based on GPR with different-sized training sets: (a) 50% data set and its relative error; (b) 60% data set and its relative error; (c) 70% data set and its relative error; (d) 80% data set and its relative error.

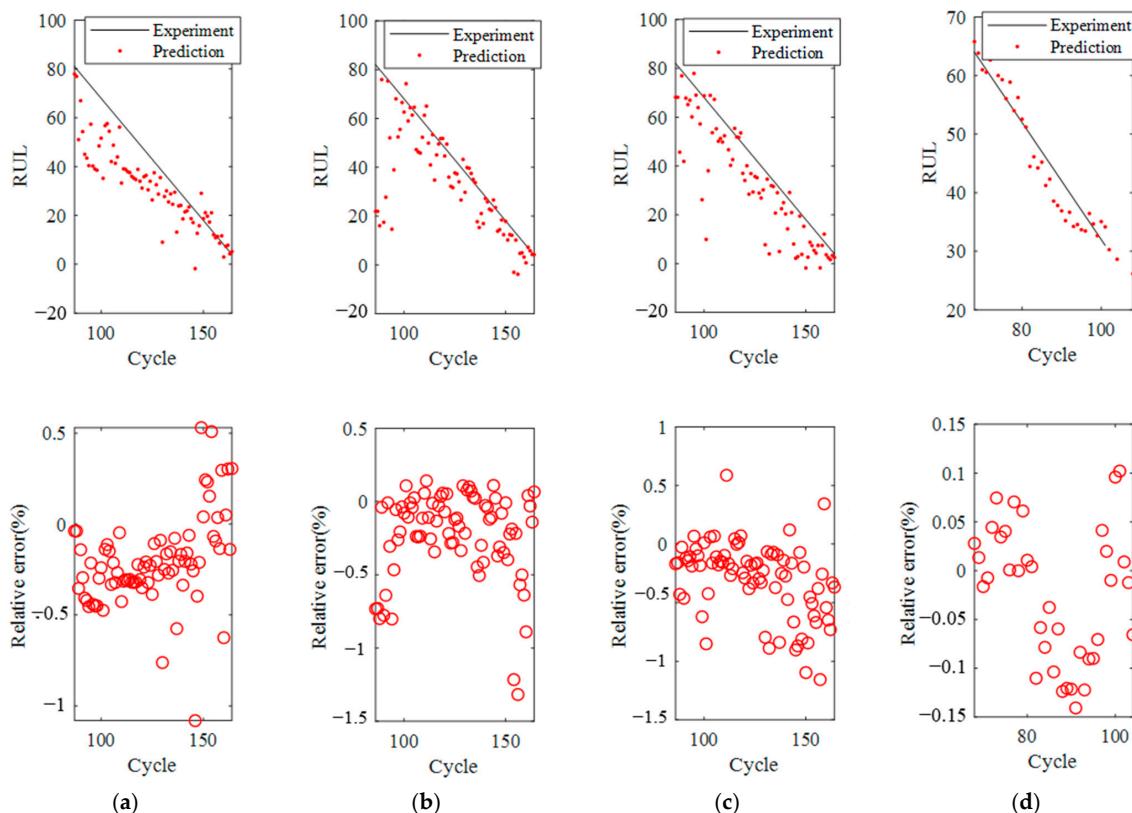


Figure 11. RUL prediction result of 4 batteries based on predicted SOH: (a) B0005; (b) B0006; (c) B0007; (d) B0018.

5. Conclusions

In this study, a joint estimation method of SOH and RUL based on GPR and LSTM was designed. By extracting six HFs from the charging current and voltage curves and analyzing the correlations between HFs and capacity with MIC, four HFs with the highest correlation were selected as the characteristic inputs of SOH in the GPR model. Then, four HFs were predicted based on LSTM, and the predicted HFs were input into the GPR model for SOH to obtain the predicted SOH. Then, the predicted SOH was input into the GPR model for RUL to predict RUL. In order to verify the reliability and accuracy of the algorithm, we used a four-battery data set with different discharge depths from NASA data. The results showed that the proposed joint estimation method of SOH and RUL based on GPR and LSTM has high accuracy and reliability.

Author Contributions: Methodology, J.Z., Y.Z. (Yong Zhu) and J.W.; software, Y.Z. (Yong Zhu), Y.Z. (Yuanyuan Zhang) and J.W.; supervision, M.L.; validation, J.Z., Y.Z. (Yuanyuan Zhang) and J.W.; visualization, M.L., B.Z. and J.W.; writing—original draft, J.Z. and J.W.; writing—review and editing, J.Z. and C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Huaneng Group Headquarters Science and Technology Project of the Key Technology Research and System Development of Group Level Intelligence Operations Platform Construction (HNKJ21-H52) and the Fault Diagnosis Technology Research and System Development of Lithium-ion Battery Energy Storage Station Based on Mass Data (HNKJ21-H52-004).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used during the study are available from the NASA's official website.

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

References

1. Sarmah, S.B.; Kalita, P.; Garg, A.; Niu, X.-D.; Zhang, X.-W.; Peng, X.; Bhattacharjee, D. A Review of State of Health Estimation of Energy Storage Systems: Challenges and Possible Solutions for Futuristic Applications of Li-Ion Battery Packs in Electric Vehicles. *J. Electrochem. Energy Convers. Storage* **2019**, *16*, 040801.
2. Kim, I.-S. A Technique for Estimating the State of Health of Lithium Batteries Through a Dual-Sliding-Mode Observer. *IEEE Trans. Power Electron.* **2010**, *25*, 1013–1022. [[CrossRef](#)]
3. Saha, B.; Goebel, K.; Christoffersen, J. Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Trans. Inst. Meas. Control.* **2009**, *31*, 293–308. [[CrossRef](#)]
4. Liao, L.; Köttig, F. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Trans. Rel.* **2014**, *63*, 191–207. [[CrossRef](#)]
5. Doyle, M.; Fuller, T.F.; Newman, J. Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. *J. Electrochem. Soc.* **1993**, *140*, 1526–1533. [[CrossRef](#)]
6. Abada, S.; Marlair, G.; Lecocq, A.; Petit, M.; Sauvant-Moynot, V.; Huet, F. Safety focused modeling of lithium-ion batteries: A review. *J. Power Sources* **2016**, *306*, 178–192. [[CrossRef](#)]
7. Li, Z. Research on Electrochemical Modeling and Aging State Identification of Lithium Ion Power Battery. Master's Thesis, Beijing University of Technology, Beijing, China, 2018. [[CrossRef](#)]
8. Huang, H. Aging State Estimation of Lithium Ion Battery Based on Electrochemical Thermal Coupling Model. Master's Thesis, Harbin Institute of Technology, Harbin, China, 2021. [[CrossRef](#)]
9. Ahmed, R.; El Sayed, M.; Arasaratnam, I.; Tjong, J.; Habibi, S. Reduced-Order Electrochemical Model Parameters Identification and State of Charge Estimation for Healthy and Aged Li-Ion Batteries—Part II: Aged Battery Model and State of Charge Estimation. *IEEE J. Emerg. Sel. Top. Power Electron.* **2014**, *2*, 678–690. [[CrossRef](#)]
10. Zou, C.; Zhang, L.; Hu, X.; Wang, Z.; Wik, T.; Pecht, M. A review of fractional-order techniques applied to lithium-ion batteries, lead-acid batteries, and supercapacitors. *J. Power Sources* **2018**, *390*, 286–296. [[CrossRef](#)]
11. Hussein, A.A. Capacity fade estimation in electric vehicles Li-ion batteries using artificial neural networks. 2013 IEEE Energy Conversion Congress and Exposition. *IEEE Trans. Ind. Appl.* **2014**, *51*, 2321–2330. [[CrossRef](#)]
12. Moura, S.J.; Argomedo, F.B.; Klein, R.; Mirtabatabaei, A.; Krstic, M. Battery State Estimation for a Single Particle Model With Electrolyte Dynamics. *IEEE Trans. Control. Syst. Technol.* **2017**, *25*, 453–468. [[CrossRef](#)]

13. Hu, X.; Li, S.; Peng, H. A comparative study of equivalent circuit models for Li-ion batteries. *J. Power Sources* **2015**, *198*, 359–367. [[CrossRef](#)]
14. Tong, S.; Klein, M.P.; Park, J.W. On-line optimization of battery open circuit voltage for improved state-of-charge and state-of-health estimation. *J. Power Sources* **2015**, *293*, 416–428. [[CrossRef](#)]
15. Luo, Y. Research on Residual Life Prediction Method of Lithium Ion Battery Based on Particle Filter. Master’s Thesis, Harbin Institute of Technology, Harbin, China, 2012.
16. Wei, K.; Chen, Q. Lithium ion power battery state estimation based on adaptive unscented Kalman filter algorithm. *J. Chin. Electr. Eng. Sci.* **2014**, *34*, 445–452. [[CrossRef](#)]
17. Su, X.; Wang, S.; Pecht, M.; Zhao, L.; Ye, Z. Interacting multiple model particle filter for prognostics of lithium-ion batteries. *Microelectron. Reliab.* **2017**, *70*, 59–69. [[CrossRef](#)]
18. Zhu, Q.; Yu, X.; Wu, Q.; Xu, Y.; Chen, F.; Huang, R. Semi empirical aging model of high energy density lithium ion batteries. *Energy Storage Sci. Technol.* **2022**, *11*, 2324. [[CrossRef](#)]
19. Wang, J.; Liu, P.; Hicks-Garner, J.; Sherman, E.; Soukiazian, S.; Verbrugge, M.; Tataria, H.; Musser, J.; Finamore, P. Cycle-life model for graphite-LiFePO₄ cells. *J. Power Sources* **2011**, *196*, 3942–3948. [[CrossRef](#)]
20. Wang, D.; Miao, Q.; Pecht, M. Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model. *J. Power Sources* **2013**, *239*, 253–264. [[CrossRef](#)]
21. Li, Z.; Lu, L.; Ouyang, M.; Xiao, Y. Modeling the capacity degradation of LiFePO₄/graphite batteries based on stress coupling analysis. *J. Power Sources* **2011**, *196*, 9757–9766. [[CrossRef](#)]
22. Long, B.; Xian, W.; Jiang, L.; Liu, Z. An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries. *Microelectron. Reliab.* **2013**, *53*, 821–831. [[CrossRef](#)]
23. Zhou, Y.; Huang, M. Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model. *Microelectron. Reliab.* **2016**, *65*, 265–273. [[CrossRef](#)]
24. Chaoui, H.; Ibe-Ekeocha, C.C. State of Charge and State of Health Estimation for Lithium Batteries Using Recurrent Neural Networks. *IEEE Trans. Veh. Technol.* **2017**, *66*, 8773–8783. [[CrossRef](#)]
25. Li, X.; Shu, X.; Shen, J.; Xiao, R.; Yan, W.; Chen, Z. An On-Board Remaining Useful Life Estimation Algorithm for Lithium-Ion Batteries of Electric Vehicles. *Energies* **2017**, *10*, 691. [[CrossRef](#)]
26. Hu, J.; Wang, J. Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression. *Energy* **2015**, *93*, 1456–1466. [[CrossRef](#)]
27. Sahinoglu, G.O.; Pajovic, M.; Sahinoglu, Z.; Wang, Y.; Orlik, P.V.; Wada, T. Battery State-of-Charge Estimation Based on Regular/Recurrent Gaussian Process Regression. *IEEE Trans. Ind. Electron.* **2017**, *65*, 4311–4321. [[CrossRef](#)]
28. Williard, N.; He, W.; Osterman, M.; Pecht, M. Comparative Analysis of Features for Determining State of Health in Lithium-Ion Batteries. *Int. J. Progn. Health Manag.* **2013**, *4*, 1–7. [[CrossRef](#)]
29. Ma, Q.; Wang, Y.; Yang, W.; Tao, B.; Zheng, Y. A Novel Health Index for Battery RUL Degradation Modeling and Prognostics. In Proceedings of the 2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS), Dali, China, 24–27 May 2019; pp. 1077–1081. [[CrossRef](#)]
30. Jia, J.; Liang, J.; Shi, Y.; Wen, J.; Pang, X.; Zeng, J. SOH and RUL Prediction of Lithium-Ion Batteries Based on Gaussian Process Regression with Indirect Health Indicators. *Energies* **2020**, *13*, 375. [[CrossRef](#)]
31. Available online: <https://catalog.data.gov/dataset> (accessed on 2 July 2022).
32. Orchard, M.E.; Tang, L.; Saha, B.; Goebel, K.; Vachtsevanos, G.J. Risk-Sensitive Particle-Filtering-based Prognosis Framework for Estimation of Remaining Useful Life in Energy Storage Devices. *Stud. Inform. Control.* **2010**, *19*, 209–218. [[CrossRef](#)]
33. Qin, T.; Zeng, S.; Guo, J.; Skaf, Z. A Rest Time-Based Prognostic Framework for State of Health Estimation of Lithium-Ion Batteries with Regeneration Phenomena. *Energies* **2016**, *9*, 896. [[CrossRef](#)]
34. Zhang, Y.; Shang, P. KM-MIC: An improved maximum information coefficient based on K-Medoids clustering. *Commun. Nonlinear Sci. Numer. Simul.* **2022**, *111*, 106418. [[CrossRef](#)]
35. Zhao, F.; Li, Y.; Feng, W. Recent Advances in Applying Vulcanization/Inverse Vulcanization Methods to Achieve High-Performance Sulfur-Containing Polymer Cathode Materials for Li-S Batteries. *Small Methods* **2018**, *2*, 1800156. [[CrossRef](#)]