Int. J. Electrochem. Sci., 17 **(2022)** Article Number: 220832, doi: 10.20964/2022.08.34

International Journal of
ELECTROCHEMICAL
SCIENCE
www.electrochemsci.org

State of Health Estimation for Li-ion Batteries using Improved Gaussian Process Regression and Multiple Health Indicators

Hao Dong[1], *Ling Mao*[1,*], *Keqing Qu*[1], *Jinbin Zhao*[1], *Fen Li*[1], *Lei Jiang*[2]

[1] School of Electrical Engineering, Shanghai Electric Power University, Shanghai 200090, China
[2] Shanghai University of Engineering Science and Technology, Shanghai 201620, China
[*]E-mail: maoling2290@shiep.edu.cn

Received: 7 May 2022 / *Accepted:* 4 June 2022 / *Published:* 4 July 2022

Accurately estimating the state of health (SOH) of Li-ion batteries (LIBs) is critical to ensure safe and stable battery operation. However, during the online operation of LIBs, it is difficult to directly measure the SOH. To overcome this challenge, this paper proposes an online SOH estimation method for LIBs based on multiple health factors (HFs) and improved Gaussian process regression. First, by analysing the dQ/dV curve and dV/dT curve of the LIB, this study finds the data interval with the highest correlation with the battery SOH to extract multidimensional health features. Then, the dimensionality reduction process is performed by the principal component analysis (PCA) method to reduce the computational complexity. Using this approach, a Gaussian process regression (GPR) model with combined kernel functions was proposed to establish the mapping relationship between the predicted values of HFs and SOH. Finally, test experiments are performed on two public lithium battery ageing datasets, and the obtained results show that the estimation error of the proposed method is kept within 1.5%, with high accuracy and reliability.

Keywords: Li-ion battery; state of health; health factors; Gaussian process regression; combined kernel functions

1. INTRODUCTION

Li-ion batteries (LIBs) are widely used in multiple applications due to their high energy density, low cost and long life [1-2]. However, with repeated charge and discharge cycles and long-term use, LIBs experience irreversible chemistry inside the battery. This increases the size of LIBs, which affects the normal use of LIBs and even causes safety accidents [3]. Therefore, real-time monitoring of the SOH of LIBs is critical to the safe and reliable operation of LIBs over their entire lifespan [4-6]. Currently, the SOH is usually defined as the ratio between the actual available capacity and the initial capacity of the LIBs, reflecting the current capacity of LIBs to store and supply energy at the beginning of their life.

When the SOH drops to 70-80%, it is considered the end-of-life (EOL) of LIBs. When the endpoint is reached, the ageing battery should be replaced.

To date, researchers have proposed a variety of approaches to SOH estimation. These methods are mainly divided into three aspects: direct measurement methods, model-based methods and data-driven methods.

The direct measurement method directly reflects the current SOH of the LIBs by measuring the internal resistance or actual available capacity of the LIBs. As the battery ages, its internal resistance will increase; thus, the measurement of internal resistance can indirectly reflect the SOH. Regarding the internal resistance of LIBs, electrochemical impedance spectroscopy (EIS) and AC pressure drop internal resistance measurement methods are usually adopted [7-8]. For capacity measurement, the open circuit voltage method is usually used to estimate the current capacity of LIBs [9-10]. Although the computational complexity of the direct measurement method is not high, it is difficult to apply it in online operation conditions because of the strict hardware requirements.

Model-based methods are divided into equivalent circuit models and electrochemical models according to the mechanism of the model. The equivalent circuit model simulates the complex physical and chemical reactions in the battery through basic electronic components. In [11], the SOH of LIBs was estimated by establishing a fractional-order model and EIS, but the resistance value was not sensitive enough in the ageing process of LIBs. Cheng et al. [12] proposed a new decoupling weighted recursive least squares (DWRLS) method to solve the problem of low accuracy in parameter estimation under different dynamic mechanisms of LIBs. This method estimates the battery dynamic parameters separately. Other researchers use the physical and chemical reactions inside the battery to establish an electrochemical model to derive the battery performance degradation mechanism related to the material properties to estimate the SOH[13]. In [14], a quasi-two-dimensional porous electrode model (pseudo two-dimensional model, P2D model) was built based on porous electrode theory, and then a series of partial differential equations (PDEs) were used to describe the internal dynamic mechanism of LIBs. The state can be estimated more accurately, and the parameters can be adjusted to adapt to LIBs with different electrode materials, but PDE has a difficult problem to solve. Although the computational complexity can be reduced by simplifying the model, the accuracy and application range of the model will also be reduced.

Therefore, the model-based method can solve the problem of simultaneous estimation of online parameters and states, but it is highly dependent on the accuracy of the model. Because model parameters are often determined under fixed operating conditions, it is only applicable to limited operating conditions and is not applicable under other operating conditions. Thus, it cannot be extended to online operation.

The data-driven method analyses and mines the implicit performance evolution law in the battery performance degradation data. It does not need to consider the complex physicochemical reactions inside the cell and directly establishes the mapping relationship between the SOH and the health indicators that can reflect the ageing state of the LIBs. Its flexibility and the advantages of not requiring modelling have attracted considerable attention in academia and industry [15]. The main data-driven methods can be divided into nonprobabilistic methods and probabilistic methods, such as nonprobabilistic methods, including support vector machine (SVM) [16-17], artificial neural network (ANN) [18], and Box–Cox

transformation (BCT) [19]. Nonprobabilistic methods cannot establish a potential probability model and cannot give an expression of uncertainty in the result. Probabilistic methods are important for SOH estimation, which not only predict the value of the SOH but also give the degree of uncertainty of the predicted value [20]. In [21], the historical capacity data decomposed the main trends and the fluctuation trends. The main trend of historical capacity is estimated through LSTM, and GPR is used to estimate the fluctuation part caused by capacity regeneration. It not only has a high prediction accuracy but also quantifies the uncertainty of the estimation results.

To achieve online SOH prediction, in addition to choosing a good modelling method, the selection of HF is also very important for characterizing the ageing state of LIBs. The charge/discharge curve of the LIBs can be more easily obtained than the capacity for the battery management system (BMS). In [22], four HFs are extracted as the input of GPR to estimate SOH. However, two of the HFs require a complete battery charging phase, and the other two HFs are also obtained at specific voltage points. As a result, the expandability of these health indicators is limited. In [23], the average value of voltage fluctuations over a period of time was extracted from the discharge curve as an HF for SOH estimation. However, in practice, the discharge curve of LIBs changes with the load. Therefore, it becomes difficult to extract health indicators from an unstable discharge curve. Although the abovementioned data-driven approaches have compensated for some of the shortcomings of the model-based and direct measurement approaches, they still face the following challenges.

(1) The abovementioned methods often use all charging data directly as input or extract features as input to estimate the SOH. However, it is difficult to determine the initial state of charge (SOC) at the beginning of battery charging and the SOC at the end of charging due to the uncertainty of the actual working conditions. The actual charging data are often fragmented data; thus, the SOH estimation model should be designed based on the fragmented data.

(2) Previous studies usually only consider the voltage or current factors of lithium batteries in the analysis of the ageing process of batteries and do not consider the temperature change factors of batteries. In fact, there is a close relationship between the temperature of the battery and the SOH thus far.

Additionally, long-term resting of LIBs will cause a short-term rebound in battery capacity. Due to its severe nonlinearity, it is difficult for existing models to predict it accurately. To overcome these difficulties, this paper proposes a state-of-health estimation method for Li-ion batteries based on Improved Gaussian Process Regression (IGPR) and Multiple Health Factors. First, the ageing data of the battery are collected in the offline stage, and the multidimensional HFs are extracted from the constant current charging stage by qualitatively analysing the dQ/dV curve and dT/dV curve of the battery. Then, PCA is used to reduce its dimensionality to obtain indirect HF, which is input into the GPR model of the combined kernel function. The obtained results show that by combining the characteristics of different kernel functions, the method proposed in this paper can simultaneously track the main trend and fluctuation part of battery ageing.

2. MATHEMATICAL ANALYSIS

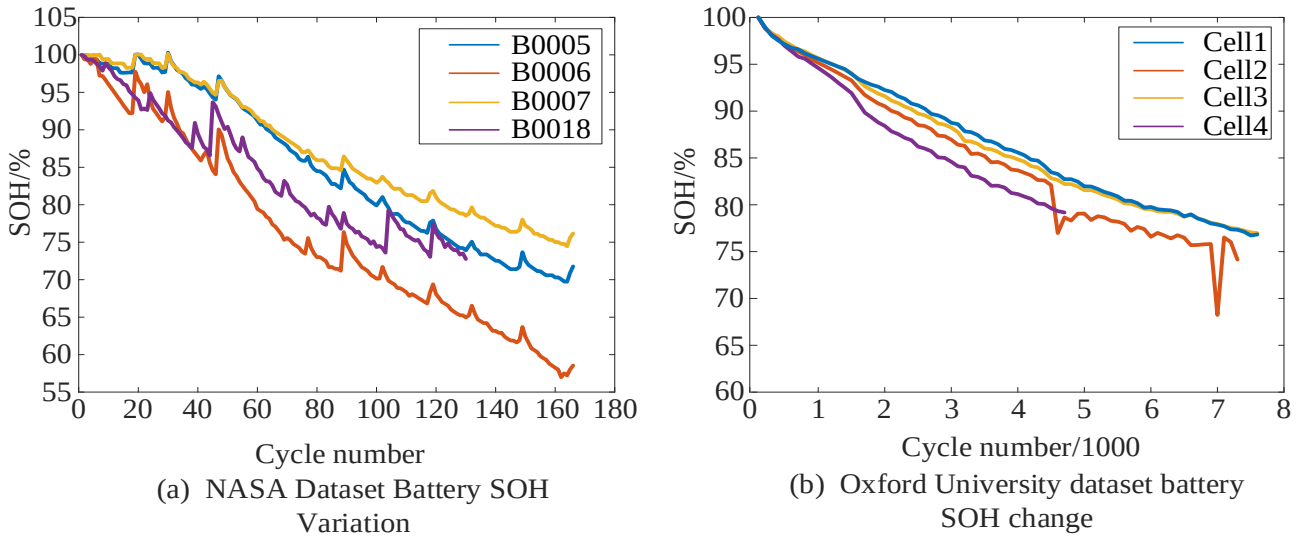
During the operation of the battery, its interior is a highly dynamic process. Currently, the SOH is generally expressed as the ratio between the current maximum available capacity of the battery and the initial capacity, as follows:

$$SOH = \frac{C_{cur}}{C_{nom}} \times 100\% \quad (1)$$

C_{cur} is the current maximum capacity of the battery, and C_{nom} is the nominal capacity of the battery. When the battery ages to a certain extent, it needs to be timely replaced for safety reasons. Here, the end-of-life (EOL) of the battery is defined as 70% of the nominal capacity of the battery.

2.1 Lithium battery aging dataset analysis

Public battery data provided by NASA and Oxford are mainly used as simulation experiment data in this study [24-25]. The LG Chem 18650 cylindrical battery used in the NASA dataset has a rated capacity of 2.1 Ah. In the ageing experiment, the lithium battery was charged with a constant current of 1.5 A until the voltage increased to 4.2 V. Then, in the constant voltage stage, the current is gradually reduced to 20 mA, and the discharge uses a constant current of 2 A to discharge to the cut-off voltage. The Oxford University battery ageing dataset uses four Kokam lithium cobalt oxide pouch cells with a rated capacity of 740 mA h. In the ageing test, the lithium battery was repeatedly charged at a constant current of 1.48 A, and its discharge process simulated the urban driving conditions of Artemis. During the experiment, the capacity was measured every 100 cycles, and the SOH change curves of the lithium batteries in the two datasets are shown in Figure 1.



(a) NASA Dataset Battery SOH Variation

(b) Oxford University dataset battery SOH change

Figure 1. Variation curve of battery SOH with the number of cycles

2.2 Health factor extraction

The correct selection of HF is the key to achieving accurate SOH estimation. In the daily use of lithium batteries, the physical quantities that the battery management system can provide are the battery's voltage, current and temperature. Therefore, HF can be constructed from these directly available battery operating data to characterize battery ageing. Because the discharge process of an LIB largely depends on the load, the discharge current will vary with the change in the load, and extracting HIs from the discharge process is not suitable for actual working conditions. In contrast, the charging stage can be controlled by the BMS system to obtain stable charging data. Figure 2 shows the voltage and temperature curves of the battery with increasing ageing. As the SOH of the battery decreases, the time for the charging voltage to reach the cut-off voltage gradually shortens, and the temperature also gradually increases. Therefore, health characteristics that characterize battery ageing can be extracted from the charging voltage curve and temperature curve.

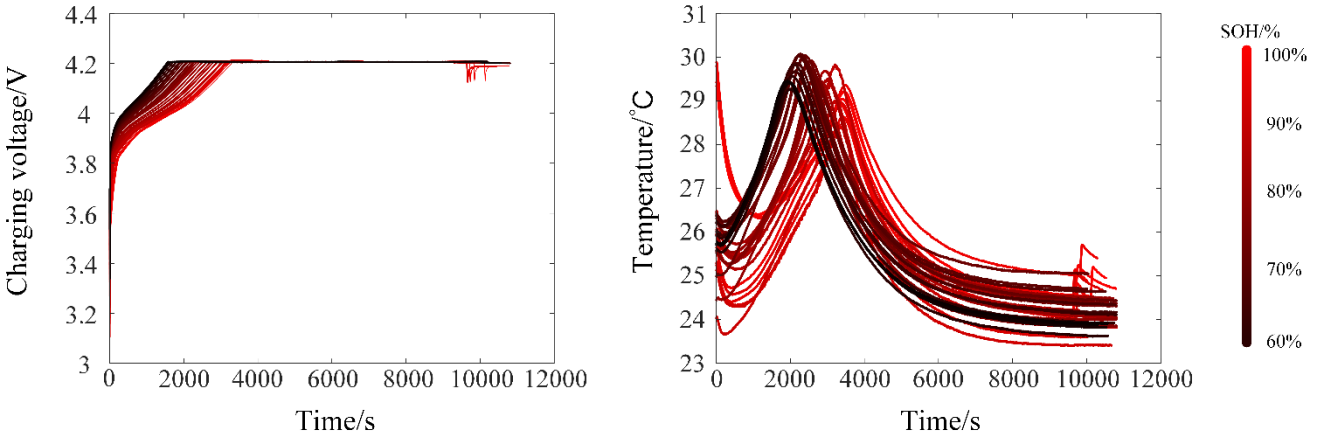


Figure 2. The voltage and temperature of the battery change curve with ageing

Currently, with the rapid advancement of cloud computing technology, a large amount of battery data is easier to collect, which will provide a good foundation for data-driven methods. However, it will become difficult to extract high-quality HF from a large amount of data; thus, secondary analysis of the features becomes more important. In addition, because the charging initiation state of charge in daily battery operation cannot be determined, the obtained charging data are often fragmented. Therefore, by analysing the dQ/dV curve and dV/dT curve of the battery, this study selects the small segment data with the highest correlation with the battery SOH from a large number of charging data to extract HF. The calculation formulas of the dQ/dV curve and dV/dT curve are as follows:

$$\frac{dQ}{dV} \approx \frac{Q(k)-Q(k-N)}{V(k)-V(k-N)} \tag{2}$$

$$\frac{dT}{dV} \approx \frac{T(k)-T(k-N)}{V(k)-V(k-N)} \tag{3}$$

Due to the limited resolution of the sensor, the method of directly using the numerical derivative will be sensitive to noise. To solve this problem, dQ/dV and dT/dV are approximated by finite

differences of *N* sample points. *Q*(*k*), *T*(*k*) and *V*(*k*) in the abovementioned formula are the capacity, temperature and voltage at time k, respectively. The larger is the *N* value, the lower is the sensitivity to noise and the better is the SOH estimation accuracy. However, if the *N* value is too large, the SOH estimation accuracy will be reduced due to the change in the peak value of the submerged curve. Therefore, the size of *N* must be reasonably selected. In this study, the value of *N* is 10. Smoothing the curve is also important because the acquired data often contain considerable noise. Therefore, the Savitzky–Golay filter is selected in this study to smooth the dQ/dV and dT/dV curves, and the obtained results are shown in Figure 3.

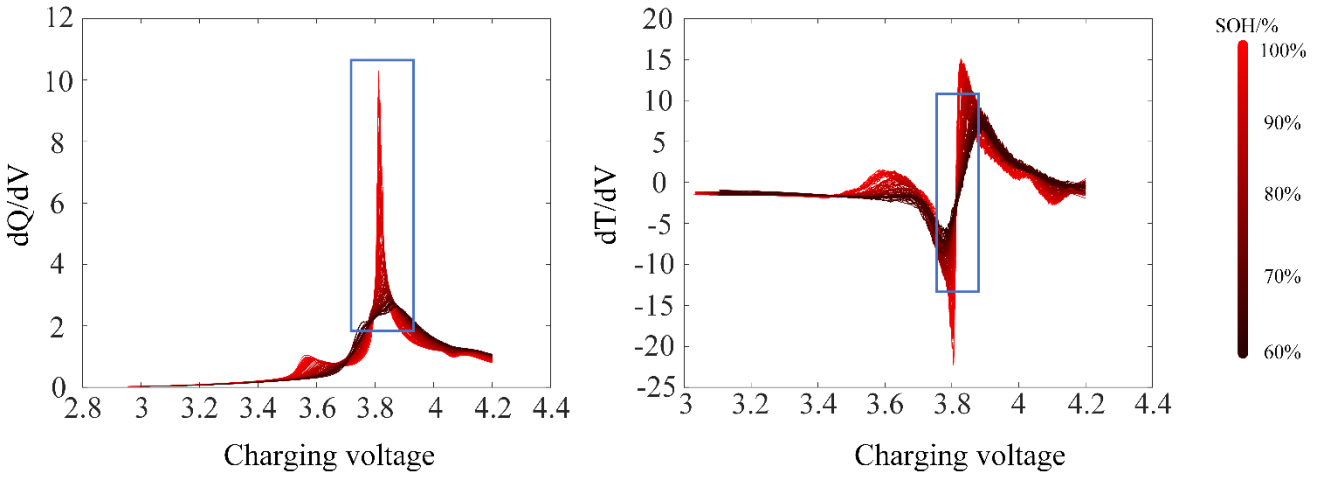


Figure 3. dQ/dV and dT/dV curves of Cell 1

First, from the dQ/dV curve, it can be seen that there is a high correlation between voltage and capacity between 3.75 V and 3.85 V; thus, these data need to be considered. Second, the relationship between the capacity and temperature of the battery is analysed, which can be expressed by dQ/dT. The calculation formula is as follows:

$$\frac{dQ}{dT} = \frac{dQ}{dV} \times \frac{dV}{dT} \tag{4}$$

The abovementioned formula shows that the correlation between capacity and temperature is proportional to dQ/dV and inversely proportional to dV/dT. From 3.8 V to 3.9 V, the rate of change of temperature and voltage is the smallest, based on the dV/dT curve. Finally, through a local traversal search, it is determined that the voltage, temperature and capacity are most correlated within the range of 3.8-3.85 V. The interval with the highest correlation between voltage, temperature, and capacity in the NASA battery ageing dataset is 3.95-4.0 V. Therefore, this data interval is selected to extract the health factor of lithium batteries. The integral of voltage versus time is set as HF_1, the integral of temperature versus time is set as HF_2, and the current number of cycles is set as HF_3. The calculation process is as follows:

$$HF_1(i) = \int_{t_1}^{t_2} Vcc(i)\, dt \tag{5}$$

$$HF_2(i) = \int_{t_1}^{t_2} Tcc(i)\ dt \quad (6)$$

$$HF_3 = Cycle(i) \quad (7)$$

In the abovementioned formula, t_1 to t_2 represent the time required for the constant current charging voltage V_{cc} to change from V_1 to V_2; T_{cc} represents the temperature of the battery during constant current charging, and i represents the ith charge–discharge cycle.

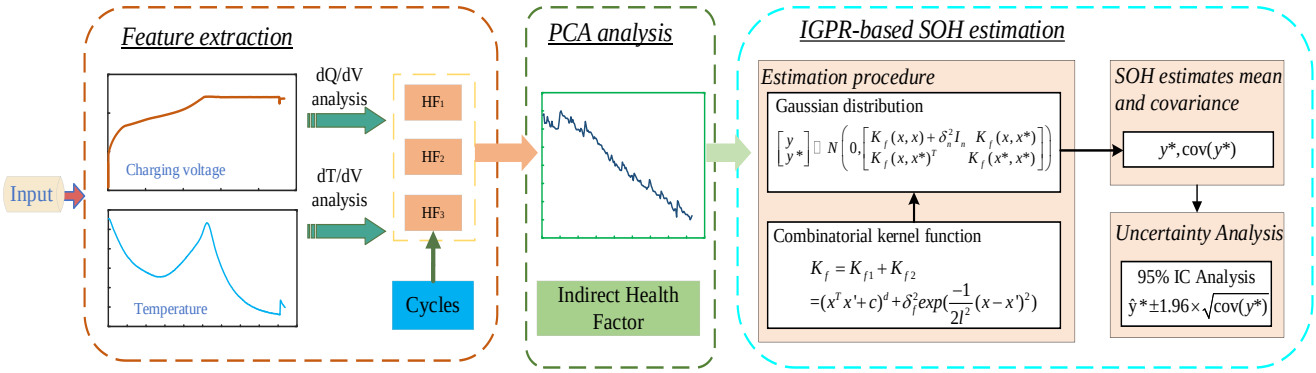


Figure 4. IGPR-based SOH estimation framework

The basic framework of IGPR-based SOH estimation is shown in Figure 4, which is mainly divided into three parts: HF extraction, principal component analysis and SOH estimation. In the first part, the dQ/dV and dT/dV curves are smoothed by the SG filtering algorithm, and the voltage and temperature ranges with strong correlation with SOH are found through analysis to extract HF. In the second part, while preserving the original information as much as possible, PCA is performed on the extracted multidimensional features to obtain IHF. The third part establishes the nonlinear mapping relationship between IHF and battery SOH through the IGPR model, outputs the estimated value of SOH and gives a 95% confidence interval.

2.3 Correlation analysis

To quantitatively evaluate the correlation between HF and SOH selected in this study, this study adopts the Pearson and Spearman coefficients, and the specific calculation is as follows:

$$Pearson = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad (8)$$

$$Spearman = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (9)$$

In the abovementioned formula, X and Y are the sample populations, x_i and y_i are the sample individuals, and the value of the correlation coefficient is between -1 and +1. The closer the absolute

value is to 1, the higher is the degree of correlation between the two. When equal to 0, there is no linear relationship between the two.

2.3 Principles of principal component analysis

To comprehensively analyse the factors of battery ageing, the introduced HF can reflect the SOH of the battery to varying degrees. However, there will be overlapping information between these HFs. To reduce the computational complexity, principal component analysis (PCA) was chosen to optimize HF. While retaining the original information, the original three HFs were recombined into a new vector and recorded as the Indirect Health Factor (IHF). The specific steps are as follows.

First, set $_X$=[HF_1, HF_2, HF_3], which is a matrix of *n***m* order, *n* is the number of samples, and *m* is the number of vectors. The formula for calculating the covariance matrix *S* is as follows:

$$S = \frac{1}{m-1}(X^{*T}X^{*}) \tag{10}$$

X^* in the abovementioned formula is the standardized *X*, and the eigenvector u_i and eigenvalue λ_i of X^*, i=1,2,…,*k*, can be calculated by the following formula:

$$Su_i = \lambda_i u_i \tag{11}$$

Let U=[u_1,u_2,…,u_k]; the matrix *Z* after dimension reduction can be obtained by the following formula:

$$Z = X^{*} \times U \tag{12}$$

2.4 Gaussian Process Regression Principle

Battery online health monitoring results always involve different types of uncertainties. The presentation of uncertainty is much more advantageous for the estimated results than just an estimated result. GPR is a kernel function-based machine learning approach that can be combined with prior knowledge based on a Bayesian framework to achieve predictions. The associated uncertainty is then described by providing confidence intervals for the predicted values. This paper describes GPR from the perspective of function space. The GPR model can be completely determined by the mean function $m(x)$ and the covariance function $k(x, x')$:

$$\begin{cases} m(x) = E(f(x)) \\ k(x,x') = E[(f(x)+m(x))(f(x')-m(x'))^T] \end{cases} \tag{13}$$

In the abovementioned formula, δ_f^2 is the signal variance, and l is the characteristic length scale. In reality, the observed value of the model will be disturbed by various noises. Therefore, independently distributed Gaussian noise is added to the model, which is expressed as follows:

$$k(x,x') = \delta_f^2 exp(\frac{-1}{2l^2}(x-x')^2) \tag{14}$$

In the abovementioned formula, $\varepsilon \sim N(0, \delta_n^2)$, and y is the output observation value after adding noise. According to the Bayesian principle, the prior distribution of the model is expressed as follows:

$$y = f(x) + \varepsilon \tag{15}$$

In the abovementioned formula, $\varepsilon \sim N(0, \delta_n^2)$, and y is the output observation value after adding noise. According to the Bayesian principle, the prior distribution of the model is expressed as follows:

$$y \sim N(0, K_f(x,x) + \delta_n^2 I_n) \tag{16}$$

$$\begin{cases} K_f(x,x) = (k_{ij})_{n\times n} \\ k_{ij} = \delta_f^2 \exp(\dfrac{-(x_i - x_j)^2}{2l^2}) \end{cases} \tag{17}$$

In the abovementioned formula, $K_f(x, x)$ is an $n \times n$ symmetric positive definite matrix, $\delta_n^2 I_n$ is the noise covariance matrix, and I_n is the n-dimensional unit matrix. k_{ij} is used to describe the similarity between x_i and x_j. The higher is the similarity, the greater is the k_{ij}. In the random process of the Gaussian process, the finite-dimensional subset of any random variable obeys the joint normal distribution; thus, the joint prior distribution of the training set (x, y) and the test set (x^*, y^*) is shown in the following formula.

$$\begin{bmatrix} y \\ y* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K_f(x,x) + \delta_n^2 I_n & K_f(x,x^*) \\ K_f(x,x^*)^T & K_f(x^*,x^*) \end{bmatrix}\right) \tag{18}$$

Then, the output is predicted by calculating the conditional distribution:

$$p(y^* | x, y, x^*) \sim N(y^* | m^*, \mathrm{cov}(y^*)) \tag{19}$$

In Equation 19, $p(y^*|x, y, x^*)$ also follows the Gaussian distribution, m^* can be regarded as the predicted value y^*, and $cov(y^*)$ can be regarded as the agreement reflecting its uncertainty. Variance matrix. The specific expressions of m^* and $cov(y^*)$ are as follows:

$$\begin{cases} m^* = K_f(x,x^*)^T [K_f(x,x) + \delta_n^2 I_n]^{-1} y \\ \mathrm{cov}(y^*) = K_f(x^*,x^*) - K_f(x,x^*)^T [K_f(x,x) + \delta_n^2 I_n]^{-1} K_f(x,x^*) \end{cases} \tag{20}$$

The hyperparameter $[\delta_f, l, \delta_n]$ in the abovementioned formula is solved by minimizing the negative log marginal likelihood (NLML), and the specific expression is as follows:

$$NLML = -\log p(y | x, \theta) \tag{21}$$

By taking the partial derivative of the likelihood function, the conjugate gradient method is used to solve it. The specific expression is as follows:

$$\begin{cases} \dfrac{\partial}{\partial \theta_i} \log p(y | x, \theta) = \dfrac{1}{2}\left\{ \alpha\alpha^T - [K_f(x,x) + \delta_n^2 I_n]^{-1} * \dfrac{\partial (K_f(x,x) + \delta_n^2 I_n)}{\partial \theta_i} \right\} \\ \alpha = [K_f(x,x) + \delta_n^2 I_n]^{-1} y \end{cases} \tag{22}$$

When the hyperparameters are solved, the GPR model can be determined. Then, the test data are estimated according to the mean function and variance, and the 95% confidence interval of the prediction result can be expressed as follows:

$$[m^* - 1.96 \times \sqrt{\mathrm{cov}(y^*)}\ ,\ m^* + 1.96 \times \sqrt{\mathrm{cov}(y^*)}] \tag{23}$$

2.4 Improved Gaussian Process Regression Model

In this study, by analysing the ageing process of lithium batteries, it is found that in addition to the main capacity decay trend, there are also local fluctuations due to battery capacity regeneration, as shown in Figure 1. Of note, the choice of GPR kernel function has a great influence on the accuracy of

the model. Previous models usually use a single kernel function for estimation, which makes it difficult to capture two different ageing trends at the same time. Therefore, this paper proposes a GPR estimation method based on combined kernel functions, which linearly couples two kernel functions with different functions to improve the accuracy of the model. The combined kernel function K_f is shown in the following formula, where K_{f1} is the polynomial kernel function used to describe the main trend part of battery ageing, and K_{f2} is the SE kernel function used to describe the fluctuation part. The calculation formula is as follows:

$$K_f = K_{f1} + K_{f2} = (x^T x' + c)^d + \delta_f^2 exp(\frac{-1}{2l^2}(x - x')^2) \quad (24)$$

3. EXPERIMENTAL ANALYSIS

To verify the accuracy and reliability of the method proposed in this study, the mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) are used to evaluate the evaluation indicators of model performance. These indicators are defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}| y_i - y_i | \quad (25)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}| \frac{y_i - y_i}{y} | \times 100\% \quad (26)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_i)^2}{n}} \quad (27)$$

In the formula, y_i represents the predicted value, $\hat{y}_i$ represents the true value, and n represents the numbers of data.

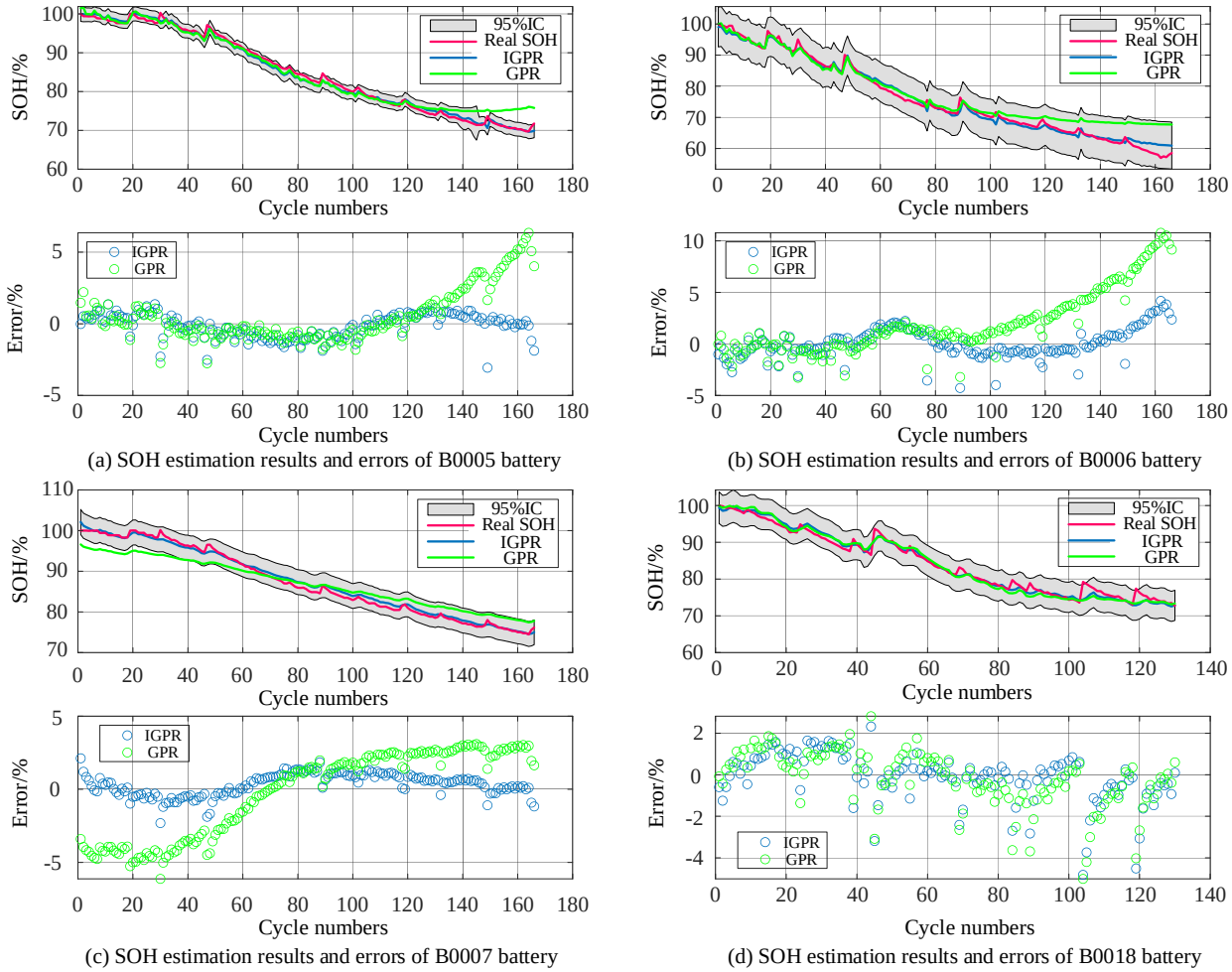
3.1 Correlation analysis results between IHF and battery SOH

To quantitatively measure the strength of the correlation between the IHF constructed in this paper and the SOH of the battery, the Pearson coefficient and the Spearman coefficient are selected for evaluation. The calculation results are shown in Table 1. It can be seen from the obtained results that the health factor constructed in this study is greater than 0.9 in both datasets. The data length required for feature extraction is only 50 mV, which can be easily obtained in practical applications.

Table 1. Calculation results of the correlation coefficient between the IHF and lithium battery SOH

	B0005	B0006	B0007	B0018	Cell 1	Cell 2	Cell 3	Cell 4
Pearson	0.9952	0.9932	0.9943	0.9905	0.9942	0.9845	0.9962	0.9978
Spearman	0.9956	0.9967	0.9966	0.9804	0.9977	0.9939	0.9991	0.9993

3.2 SOH estimation results



(a) SOH estimation results and errors of B0005 battery

(b) SOH estimation results and errors of B0006 battery

(c) SOH estimation results and errors of B0007 battery

(d) SOH estimation results and errors of B0018 battery

Figure 5. SOH estimation results and errors of 5, 6, 7 and 18 batteries

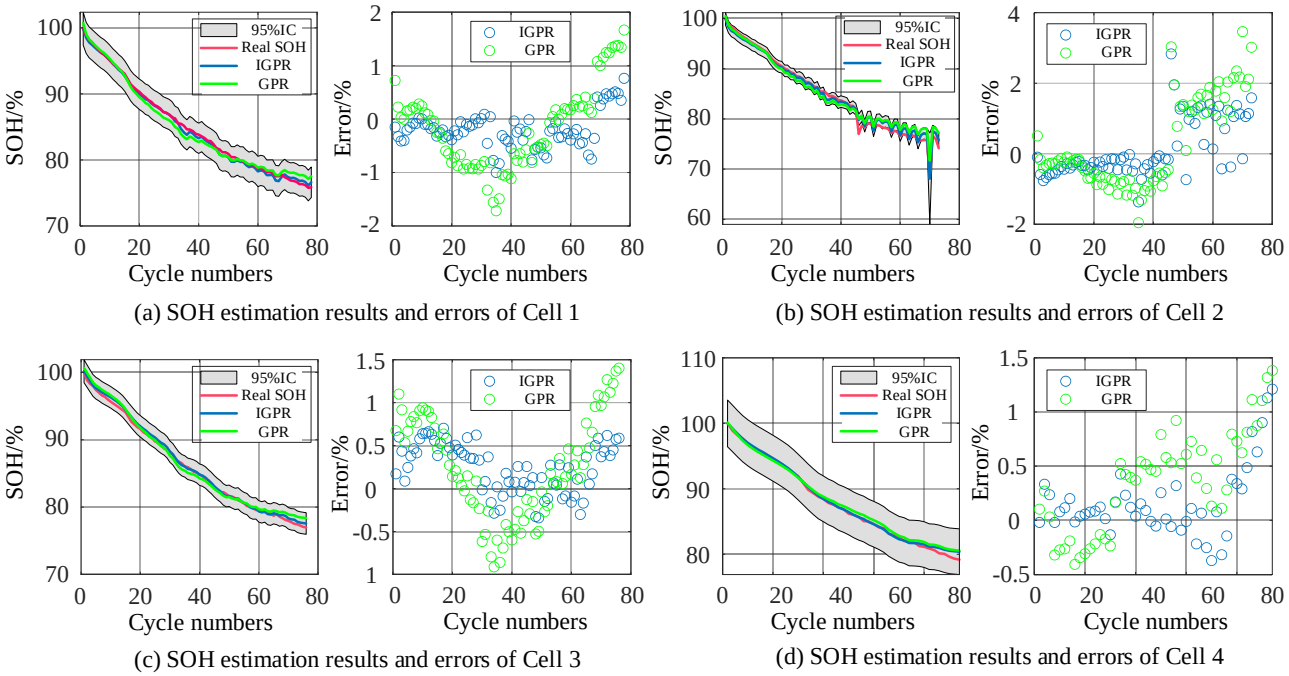
Table 2. Calculation results of the SOH estimation error for batteries 5, 6, 7 and 18

Model	Test set	Training set	MAE/%	MAPE/%	RMSE/%
GPR	B0005	B0006, B0007, B0018	1.31	1.69	1.88
	B0006	B0005, B0007, B0018	2.41	2.63	2.53
	B0007	B0005, B0006, B0018	2.62	2.96	2.92
	B0018	B0005, B0006, B0007	1.05	1.26	1.38
IGPR	B0005	B0006, B0007, B0018	0.69	0.81	0.85
	B0006	B0005, B0007, B0018	1.10	1.43	1.37
	B0007	B0005, B0006, B0018	0.68	0.79	0.81
	B0018	B0005, B0006, B0007	0.86	1.01	1.17

To verify the accuracy of the method proposed in this paper, one battery in the NASA or Oxford University battery ageing dataset is selected as the test set each time, and the remaining battery data are used as the training set. Figures 5(a)-(d) show the SOH prediction results and error percentages of the NASA dataset. The blue curve in the figure represents the estimated value of IGPR, the green curve

represents the estimated value of GPR, and the red curve represents the true value of SOH. The grey shaded area represents the 95% confidence interval, which is used to describe the uncertainty of the estimation result. The smaller is the confidence interval, the better is the estimation effect.

Compared with the single-core GPR model, the IGPR model proposed in this paper can not only accurately track the ageing trend of the battery but also accurately track the local capacity fluctuation part. From the error percentage results, it can be seen that the prediction error in the NASA dataset is approximately 4% except for a few points, and most of them are distributed within 2%. When the SOH of the battery is lower than 70%, the estimated effect of the model worsens. This occurs because with the ageing of the battery, the nonlinear degree of the relationship between the health characteristics and the SOH increases, causing the error to increase. In fact, when the battery capacity decays to 70-80% of the rated capacity, the performance of the battery will decline exponentially, and the battery should be timely replaced. This threshold is generally called the end-of-life threshold, and data below this threshold are rarely obtained in actual operation; thus, a large estimation error can be accepted. Table 2 shows the calculation results of the error indicators MAE, MAPE and RMSE. It can be seen that the error indicators of the SOH estimation on the NASA dataset by the method proposed in this paper are all within 2%.



(a) SOH estimation results and errors of Cell 1

(b) SOH estimation results and errors of Cell 2

(c) SOH estimation results and errors of Cell 3

(d) SOH estimation results and errors of Cell 4

Figure 6. SOH estimation results and errors of Cell 1-Cell 4

To fully verify the applicability of our method, Fig. 6 shows the SOH estimation results and errors of our method in the Oxford battery ageing dataset Cell1-Cell4. It can be seen that the estimation accuracy of the GPR model using the combined kernel function is higher, and most of the errors are distributed within 1%. The calculation of the SOH estimation error index based on GPR and IGPR is

shown in Table 3. On the basis of Cell1-Cell4 batteries, the RMSE of the IGPR model was reduced by 0.4%, 0.39%, 0.27% and 0.25%, respectively, compared with the GPR model.

Table 3. Calculation results of the SOH estimation error of Cell 1-Cell 4

Model	Test set	Training set	MAE/%	MAPE/%	RMSE/%
GPR	Cell 1	Cell 2, Cell 3, Cell 4	0.63	0.75	0.78
	Cell 2	Cell 1, Cell 3, Cell 3	1.04	1.25	1.21
	Cell 3	Cell 1, Cell 2, Cell 4	0.53	0.62	0.64
	Cell 4	Cell 1, Cell 2, Cell 3	0.49	0.58	0.59
IGPR	Cell 1	Cell 2, Cell 3, Cell 4	0.32	0.39	0.38
	Cell 2	Cell 1, Cell 3, Cell 4	0.64	0.78	0.82
	Cell 3	Cell 1, Cell 2, Cell 4	0.32	0.35	0.37
	Cell 4	Cell 1, Cell 2, Cell 3	0.24	0.29	0.34

3.3 Model Comparison Analysis

To verify the superiority of the method proposed in this paper, current mainstream algorithms such as SVM [26], ANN [27], LSTM [28] and ELM [29] are selected for comparison. The error results are shown in Table 4. When the IGPR model proposed in this paper is used, the estimated error index is the lowest, with the average MAE, MAPE and RMSE values of 0.61%, 0.62% and 0.73%, respectively. Of note, compared with other mainstream algorithms, this model can also provide the degree of uncertainty of the prediction results, which is very important for engineering practice.

Table 4. Error analysis of different models

Model	Average MAE/%	Average MAPE/%	Average RMSE/%
SVM	1.74	1.76	1.93
ANN	2.13	2.21	2.17
LSTM	1.86	1.84	1.91
ELM	1.26	1.47	1.49
IGPR	0.60	0.61	0.76

4. CONCLUSION

In this paper, an online estimation method of lithium battery SOH based on multiple health factors and improved Gaussian process regression is proposed. Experiments are performed on two public datasets, and the obtained results show that the method proposed in this paper can not only accurately track the main ageing trend and local fluctuation part of the battery but also give the uncertainty of the estimated results. The main contributions are divided into the following two aspects.

(1) To solve the problem of incomplete data obtained during the online SOH estimation of lithium batteries. Through the voltage and temperature data provided by the battery management system during the charging process, the health factor extracted using the dQ/dV curve and dT/dV curve analysis only uses the data within the 50 mV voltage interval. These fragmented data are more readily available during actual battery operation than full charge data.

(2) A Gaussian process regression model combining the polynomial kernel function and SE kernel function is used to estimate the main degradation trend and fluctuation part of the ageing process of lithium batteries. The experimental results show that the estimation accuracy of the proposed method is higher than that of the GPR model with a single kernel function.

CONFLICTS OF INTEREST

The authors declare there is no potential conflict of interest involved in this writing process.

ACKNOWLEDGMENTS

Project Supported by National Natural Science Foundation of China 52177184

References

1. A. S. Mogodal, and K. M. Zohdy, *Int J Electrochem Sci.*, 15 (2020) 1128.
2. X. Zhang, Y. Qin, C. Yuen, L. Jayasinghe and X. Liu, *IEEE Trans Industr Inform*, 17 (2021) 6820.
3. X. Xin, S. L. Wang, C. M. Yu, J. Cong, and J. Coffie-Ken, *Int J Electrochem Sci*., 15 (2020) 2216
4. C. She, Z. Wang, F. Sun, P. Liu and L. Zhang, *IEEE Trans Industr Inform*, 16 (2020) 3345.
5. J. H. Huang, S. L. Wang, W. H. Xu, C. Fernandez, Y. C. Fan and X. P. Chen, *Int J Electrochem Sci*., 16 (2021) 21075.
6. G. Dong, F. Yang, Z. Wei, J. Wei and K. Tsui, *IEEE Trans Industr Inform*, 16 (2020) 4736.
7. W. Waag, S. Kbitz, and D. U. Sauer, *Applied Energy*, 102 (2013) 885.
8. S. F. Schuster, T. Bach, E. Fleder, J. Müller, M. Brand, G. Sextl, and A. Jossen, *J Energy Storage*, 1 (2015) 44.
9. L. Chen, W. Lin, J. Li, B. Tian, and H. Pan, *Energy*, 106 (2016) 662.
10. C. R. Birkl, E. McTurk, M. R. Roberts, P. G. Bruce, and D. A. Howey, *J Electrochem Soc*., 162 (2015) A2271.
11. R. Xiong, J. Tian, H. Mu, and C. Wang, *Appl. Energy*, 207 (2017) 372.
12. C. Zhang, W. Allafi, Q. Dinh, P. Ascencio, and J. Marco, *Energy*, 142 (2018) 678.
13. C. F. Pan, Y. Chen, L. M. Wang and Z. G. He, *Int J Electrochem Sci*., 14 (2019) 9537.
14. S. Abada, G. Marlair, A. Lecocq, M. Petit, V. Sauvant-Moynot, and F. Huet, *J Power Sources*, 306 (2016) 178.
15. K. Liu., *Renew. Sust. Energ. Rev*., 113 (2019) 109254.
16. J. Meng, L. Cai, G. Luo, D. I. Stroe, and R. Teodorescu, *Microelectron Reliab*, 88 (2018) 1216.
17. J. Wei, G. Dong, and Z. Chen, *IEEE Trans. Ind. Electron*, 65 (2018) 5634.
18. L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao and M. J. Deen, *IEEE Trans Industr Inform*, 17 (2021) 3478.
19. Y. Zhang, R. Xiong, H. He, and M. G. Pecht, *IEEE Trans. Ind. Electron*, 66 (2019) 1585.
20. K. Liu, Y. Shang, Q. Ouyang and W. D. Widanage, *IEEE Trans. Ind. Electron*, 68 (2021) 3170.
21. L. Feng, W. F. Xiaohui, and G. Yong, *Appl. Sci*., 6 (2016) 166.
22. D. Y ang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, *J. Power Sources*, 384 (2018) 387.

23. Y. P. Zhou, M. H. Huang, Y. P. Chen, and Y. Tao, *J. Power Sources*, 321 (2016) 1-10.
24. D. T. Liu, J. B. Zhou, and H. T. Liao, *IEEE T SYST MAN CY-S.*, 45 (2015) 915-928.
25. C. Birkl, Diagnosis and Prognosis of Degradation in Lithium-Ion Batteries ," Univ. Oxford, Oxford, U. K., 2017.
26. C. Lei, J. Meng, D. Stroe, J. Peng and R. Teodorescu, *IEEE T POWER ELECTR*, 35 (2020) 99.
27. H. Dai, G. Zhao, M. Lin, J. Wu, and G. Zheng, *IEEE Trans. Ind. Electron*, 66 (2018) 7706–7716.
28. Y. Tan and G. Zhao, *IEEE Trans. Ind. Electron*, 67 (2020) 8723–8731.
29. H. Pan, Z. Lv, H. Wang, H. Wei, and L. Chen, *Energy*, 60 (2018) 466–477.