Int. J. Electrochem. Sci., 17 **(2022)** Article Number: 221135, doi: 10.20964/2022.11.25

International Journal of ELECTROCHEMICAL SCIENCE
www.electrochemsci.org

State of Health Estimation and Remaining Useful Life Estimation for Li-ion Batteries Based on a Hybrid Kernel Function Relevance Vector Machine

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

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

Received: 12 August 2022 / *Accepted:* 16 September 2022 / *Published:* 10 October 2022

Accurate estimations of the state of health (SOH) and the remaining useful life (RUL) of lithium batteries are important indicators that ensure the safe and stable operation of a battery system. However, these two health indicators are difficult to estimate during online operation. This paper proposes a joint estimation method for SOH and RUL based on the hybrid-kernel RVM (H-RVM) method. The method extracts the segment data features of the charging voltage, current and temperature, online ,by analysing the battery incremental capacity (IC) curve and obtains the indirect health factor (IHF) by reducing the dimension through a principal component analysis (PCA). Then, an ageing model of a lithium battery is established by the RVM algorithm. On this basis, another RVM is used to perform a multistep prediction of the IHF, combining the prediction results with the battery ageing model and comparing the failure thresholds to attain a RUL estimation. Finally, three groups of battery data, under different ageing conditions, are used for verification. The results show that the method proposed in this paper has high accuracy and stability.

Keywords: Li-ion battery; state of health; health factors; remaining useful life; relevance vector machine

1. INTRODUCTION

As a clean energy source, lithium-ion batteries are widely used in electric vehicles, electronic products and various mobile energy storage devices [1-3]. Compared with other energy storage batteries, they have the advantages of high energy density, low self-discharge rate and long cycle life [4]. However, these advantages will decrease with the decay of the battery life, which can easily lead to battery system failures and even safety accidents [5]. Therefore, the SOH and RUL of the battery are estimated in real time to characterize the ageing degree of the battery and to provide a risk warning to remind people to maintain and replace the battery. The RUL is generally defined as the number of cycles required to decay

from the current state to the end of life (EOL), and the EOL is usually set as the moment when the SOH of the battery drops to 70%-80%, based on the actual situation.

Battery aging is the result of the coupling of multiple factors and is highly nonlinear, which complicates a battery's SOH and RUL estimations [6]. In recent years, researchers have conducted substantial amounts of research on this topic, mainly divided into research on model-based methods and data-driven methods. Model-based methods generally simulate the complex changes inside the battery through an equivalent circuit model. The internal resistance and capacity can be identified online. A series of filtering algorithms, such as the Kalman filter, particle filter and their extensions, are used to achieve an SOH estimation [7-8]. However, such methods are often only applicable to fixed operating conditions, and the accuracy of the model is required to be high when estimating a battery's SOH online. The online identification of internal resistance and capacity is easily affected by the accuracy of the sensor, making the estimation accuracy of the model low. For RUL estimations, the model-based methods can only provide point predictions and cannot track the long-term ageing process of the batteries, which is not effective for a long-term prediction.

With the rapid development of artificial intelligence and machine learning methods, data-driven methods have attracted increasing attention from researchers [9]. Because there is no need to consider the complex physical and chemical changes inside the battery, the state estimation of the battery is obtained by establishing a nonlinear mapping relationship between the measurement data of the external voltage, current and temperature of the battery and the battery capacity. Due to their advantages of flexibility and applicability, the data driven methods have gradually become the mainstream battery SOH and RUL estimation methods. Data-driven methods can be roughly divided into two categories, namely, the nonprobabilistic methods and probabilistic methods [10]. The Nonprobabilistic methods obtain point estimations by establishing a relationship between the input and output, as is done in support vector machines (SVMs), artificial neural networks (ANNs) and extreme learning machines (ELMs) [11-13]. Although this kind of method has a good estimation effect, the output of its estimation result cannot provide an expression of uncertainty. For long-term prediction problems such as RUL, probabilistic estimation results are more beneficial to the actual battery operation. Therefore, an RVM based on the SVM model combined with Bayesian theory of probability learning is widely used in regression and prediction.

With the rapid development of big data technology, a large amount of practical, multidimensional real-time battery data is uploaded to the cloud, which offers great potential for the data-driven approaches. For the joint estimation of the SOH and RUL of lithium batteries, in addition to choosing an appropriate data-driven model, screening out the healthy features that can correctly characterize battery ageing from a large amount of data is also crucial to establishing a correct battery ageing model. [14] achieved one-step and multistep predictions of battery capacities by using historical capacity data for EMD decomposition as the input to the GPR model. Although this method has good short-term and long-term prediction effects, it is difficult to directly obtain historical capacity data during online operations. [15] extracted the mean value of a part of the voltage fluctuation from the discharge voltage curve of the battery as a healthy HF to attain RUL estimations. However, the discharge process experienced in the actual use of the battery is often a random process, and its voltage data are very unstable, so it is difficult to track online. [16] extracted the time difference of an equal voltage interval

from the constant current charging stage of the battery as the input of the ELM model to obtain an online estimation of the SOH. The charging curve of the battery is more stable than the discharging curve, which makes it easy to perform feature extractions. However, for both short-term and long-term predictions of battery health, the temperature changes cannot be ignored.

The literature on state estimations of lithium batteries is usually based on separate estimations of SOH and RUL, and there are few studies on the joint estimation method of the two [17]. In fact, both are important indicators for battery health assessments, and both are indispensable in the actual operation of the battery. In addition, the results of the mutual coupling between the SOH and RUL cause the estimation of the SOH to affect the future ageing trend of the battery, which is bound to also affect the RUL prediction of the battery. Therefore, to operationalize the health management of a battery's entire life cycle and to reduce the resources wasted on separate calculations, it is necessary to jointly estimate the SOH and RUL of the battery.

Based on the above reasons, this paper conducts feature extraction from a battery's charging current, voltage and temperature curves by analysing the IC curve of the battery. To reduce the computational complexity, PCA is used for dimensionality reduction to obtain the indirect health feature, IHF. Pearson and Spearman correlation coefficients were used to evaluate the correlation between IHF and SOH. In addition, a method for estimating the SOH and RUL of lithium batteries based on the mixed kernel function, RVM, is proposed. The IHF is used as the input to indicate the SOH to the RVM, and the RVM establishes the battery's ageing model. On this basis, a long-term estimation model of IHF with ageing is established through another RVM model, and a multistep prediction of the SOH is implemented through an integrations with the battery ageing model, and an accurate estimation of RUL is obtained.

2. MATHEMATICAL ANALYSIS

The available capacity of the battery is the factor that most directly reflects the battery's health status. Therefore, the SOH of a lithium-ion battery is defined as the ratio of the current maximum capacity to the initial capacity, and the specific calculation is as follows:

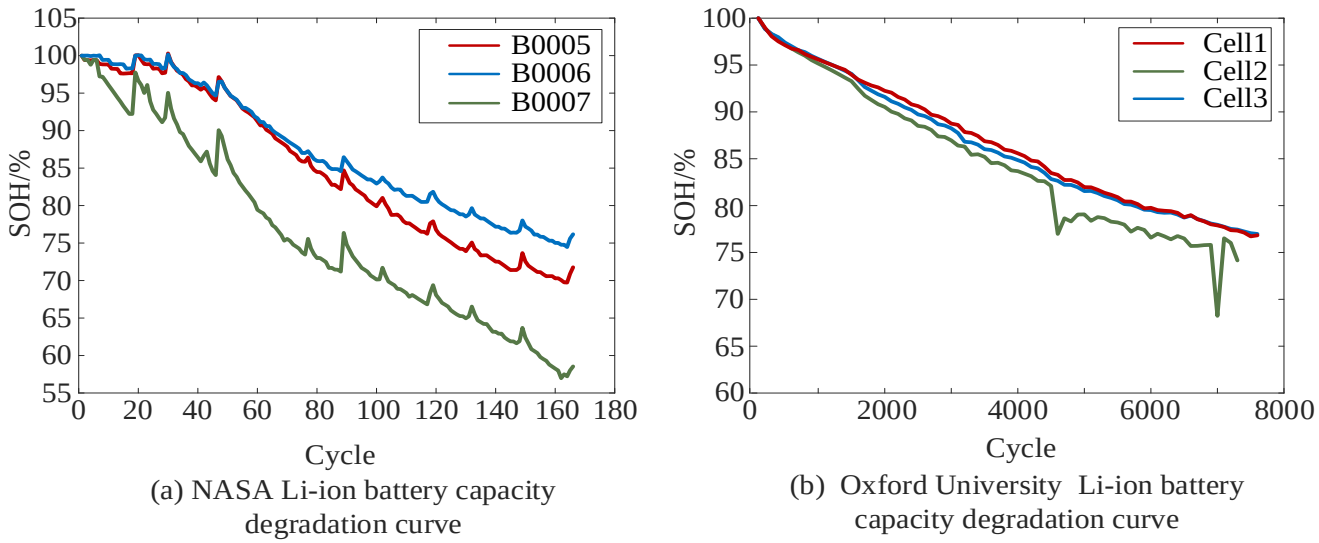
$$SOH = \frac{C_{cur}}{C_{nom}} \times 100\% \quad (1)$$

C_{cur} is the current maximum capacity of the battery, and C_{nom} is the initial capacity of the battery. As the number of times the battery is used increases, its SOH will continue to decrease. The End-of-life (EOL) of the lithium-ion battery is considered to be reached when the SOH decreases to 70%-80%.

2.1 Lithium battery ageing dataset analysis

This paper chooses to use two public datasets to experimentally verify the method proposed in this paper. The Oxford University battery dataset contains 3 Kokam lithium cobalt oxide pouch cells, labelled Cell1-Cell3. Its rated capacity is 740 mAh [18]. The ageing test was carried out at room temperature (40 °C), the battery was charged with a current of 2C, and a dynamic discharge was

performed by simulating the actual operating conditions of a car in an urban area. After 100 cycles, a constant current charge and discharge were performed with a current of 1 C, and the voltage, current and surface temperature data of the battery were recorded every second. The NASA dataset comes from the NASA Ames Prognostics Centre of Excellence (PCoE) [19]. The battery has a rated capacity of 2 Ah and is marked as B0005, B0006 and B0007. Under room temperature conditions of 24 degrees Celsius, the battery is repeatedly charged at a 1.5 A-4.2 V constant current and constant voltage mode, and the cut-off current is 20 mA. Then, a discharge current of 2 A is used to discharge to the cut-off voltages of 2.7 V, 2.5 V and 2.5 V, and the Arbin tester is used to record the voltage, current and temperature data of the battery. Figure 1 shows the change curve of the battery SOH with the ageing process



(a) NASA Li-ion battery capacity degradation curve

(b) Oxford University Li-ion battery capacity degradation curve

Figure 1. Li-ion battery capacity degradation curve

2.2 Health factor extraction

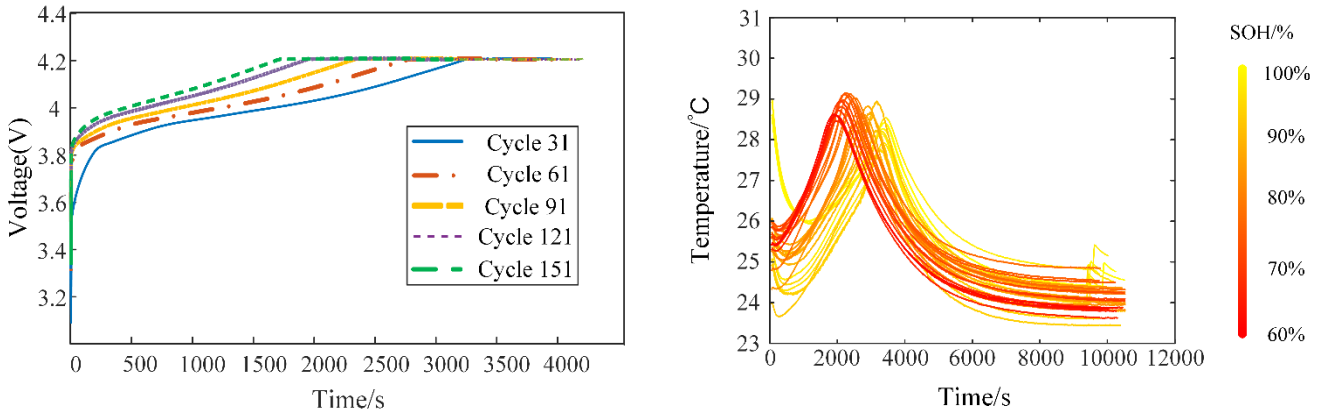


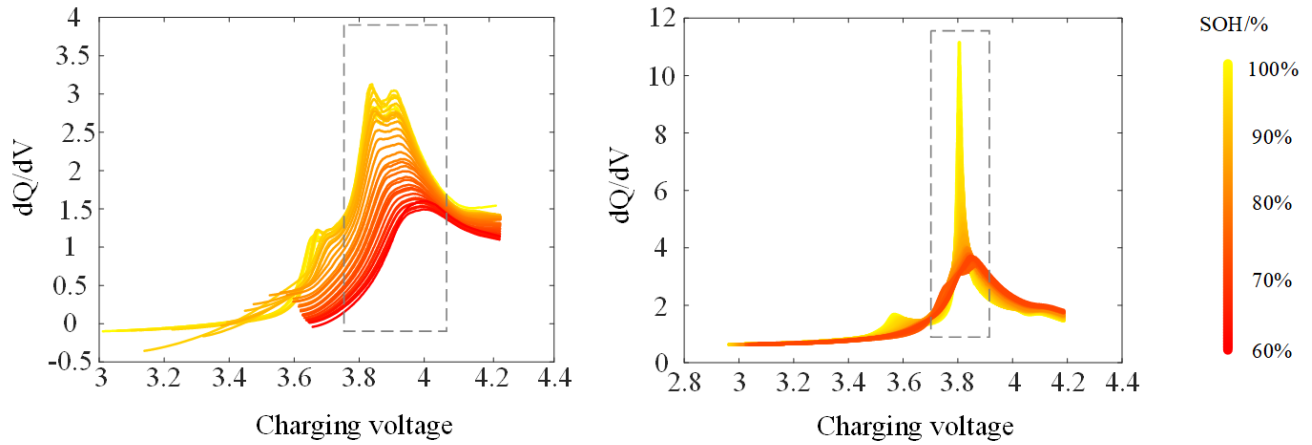
Figure 2. Battery charging voltage curve and temperature curve

For the health feature measurements, in addition to considering the strength of the feature correlation, it is also necessary to pay attention to the convenience, rationality and practicability of extraction during the actual use of the battery. As the number of battery charge–discharge cycles increases, the ageing state of the battery continues to increase, and its charging curve will also change accordingly. Figure 2 shows that the time to reach the cut-off voltage in the constant current charging stage of the battery is continuously shortened, and the temperature during the charging process gradually deepens with ageing. The charge voltage curve and temperature curve are in good agreement with the decay of the SOH. In contrast, since the actual discharge condition is not as smooth as the charging condition, it is not suitable for feature extraction. Therefore, the segment data of suitable charging voltages and temperatures can be selected for feature extraction.

With the rapid development of big data technology, the state estimation model can obtain a large amount of the real-time operating data of the battery. To reduce the length of the input data, obtaining high-quality data segments becomes particularly important. To better reflect the influence of capacity decline on the voltage curve, this paper adopts the IC analysis method. The formula for calculating the IC curve is as follows:

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

Q(k) and V(k) in the above formula represent the capacity and voltage data at the kth time, respectively. Since the sensor will be affected by noise in the measurement process, the finite difference of the N sampling points is adopted for approximation. Although a larger value of N will reduce the effect of noise, too large a value will mask the change in the peak of the IC curve. After weighing the pros and cons, the N value is set to 10. The calculation result of the IC curve is shown in Figure 3.



(a) IC curve of B0005 battery (b) IC curve of Cell 2 battery

Figure 3. IC curves of the B0005 battery and Cell 2 battery

Fig. 3(a) and Fig. 3(b) show that there is a strong correlation between the capacity decay and the voltage near the peak of the IC curve. For subsequent quantitative analyses, the voltage interval of the

peak interval is recorded as [V1, V2]. As shown in Figure 4, the integrals of voltage and current in this interval, over time, are used as HF1 and HF2. At the same time, to consider the influencing factors of the battery charging temperature, the integral of the temperature over time in this interval is recorded as HF3. The specific calculation process is as follows:

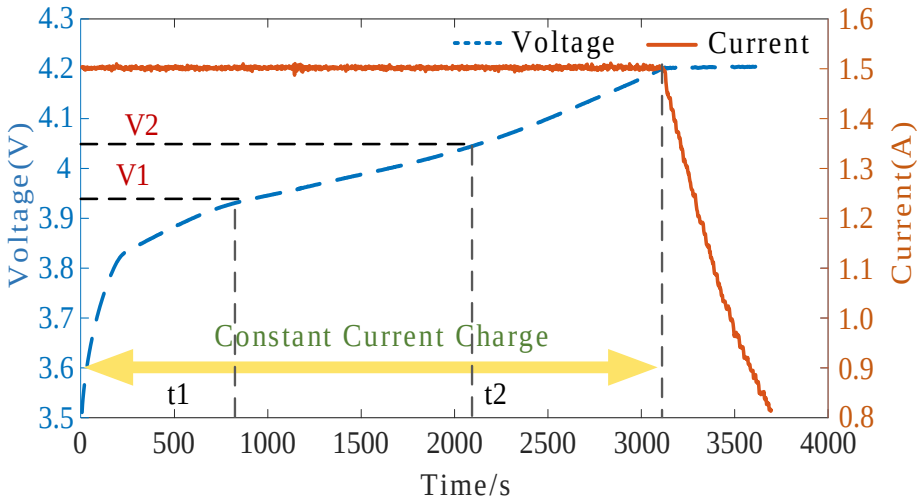


Figure 4. Constant current charging voltage and current curve of the B0005 battery

$$HF_1(i) = \int_{t_1}^{t_2} Vcc(i)\ dt \quad (3)$$

$$HF_2(i) = \int_{t_1}^{t_2} Icc(i)\ dt \quad (4)$$

$$HF_3 = \int_{t_1}^{t_2} Tcc(i)\ dt \quad (5)$$

In the above formula, t_1 and t_2 represent the time when the voltage is at V_1 and V_2, respectively, V_{cc} represents the charging voltage curve, I_{cc} represents the charging current curve, T_{cc} represents the charging temperature curve, and i represents the ith cycle of the battery.

2.3 SOH and RUL Estimation Framework

The SOH and RUL characterize the ageing status of the battery on different time scales, considering the connection and difference between the two. In this paper, a hybrid kernel function, RVM, is proposed to fit the SOH and RUL estimation models for lithium batteries. The overall framework is shown in Figure 5. The IHF is obtained by extracting three HFs from the charging current, voltage and temperature data and then performing PCA dimensionality reduction processing on them. Using IHF as the input of C-RVM, a battery ageing model is established to attain an SOH estimation. Then, the IHF is used as the input of the I-RVM to predict the overall trend, and the multistep prediction results are brought into the C-RVM model to attain the RUL estimations.

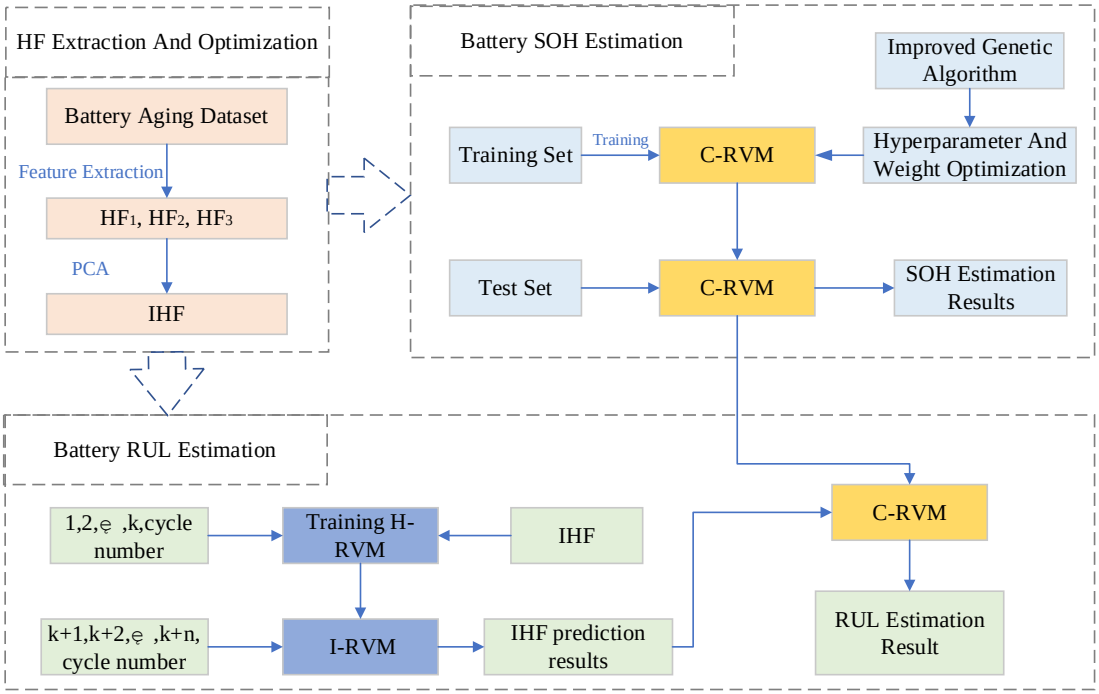


Figure 5. SOH and RUL Estimation Framework

2.4 Correlation analysis

To further analyse the correlation between the IHF and SOH extracted in this paper under the different voltage ranges, this paper uses the Pearson and Spearman correlation coefficients. It is calculated as follows:

$$Pearson = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad (6)$$

$$Spearman = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (7)$$

The Pearson and Spearman correlation coefficients can be used to evaluate the correlation between two variables. The larger the absolute value is, the higher the correlation. Among them, X and Y represent the IHF and SOH sample populations, respectively, and x_i and y_i represent the sample individuals [20].

2.5 PCA

Three health features were extracted from the charging voltage, current and temperature data in the previous section. They contain different degrees of overlapping ageing information. To reduce the redundancy of the input data of the model and increase the calculation speed, this paper selects PCA to

perform dimensionality reduction processing on the three HFs to obtain the IHF. The specific PCA calculation process is as follows.

First, define Y as a n*m-order matrix containing three HFs, and then calculate the covariance matrix Q of the matrix Y.

$$Q = \frac{1}{m-1}(Y^{*T}Y^{*}) \tag{8}$$

Standardize the matrix Y to obtain Y^{*}, and then calculate the eigenvector p_i and eigenvalue b_i by the following formula:

$$Qp_i = b_i p_i \tag{9}$$

In the above formula, Q is a n*n matrix, p_i is a n*1 vector, and b_i is a scalar. From this, the matrix T , after a dimensionality reduction, can be obtained:

$$T = Y^{*} \times P \tag{10}$$

where P=[p_1,p_2,…,p_k], i=1, 2, …, k, and the component with the largest contribution rate of each principal component is obtained as IHF by the following formula.

$$r_i = b_i / \sum_{i=1}^{k} b_i \times 100\% \tag{11}$$

Finally, the order of the input matrix of the model is changed from n*m to n*1, which greatly reduces the computational complexity of the model.

2.6 Relevance Vector Machine

RVM is a regression method based on the Bayesian sparse kernel, which overcomes the main limitations of SVM, such as requiring the kernel function to be a positive definite kernel [21]. Compared to SVM, RVM usually produces a sparser model, which in turn makes it run faster on the test set. The specific estimation process is as follows:

Suppose the sample dataset used for training is $\{x_n, t_n | n = 1,2, \dots, N\}$, x_n represents the sample input value used for training, and t_n represents the output of the independent distribution and establishes a functional relationship between x_n and t_n:

$$t_n = \phi(x_n)w_n + \varepsilon_n \tag{12}$$

In Formula 10, $\phi(x_n) = K(x, x_n)$ represents the kernel function, and ε_n represents the additional Gaussian noise of $\varepsilon_n \sim N(0, \sigma^2)$. Since t_n are independent of each other, the likelihood function of the training set can be expressed as follows:

$$p(t \mid w, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\{-\frac{1}{2\sigma^2} \| t - \phi w \|^2\} \tag{13}$$

In the above formula, $w = [w_0, w_1, \dots, w_n]^T$, ϕ is an $N \times (N+1)$ matrix, and w_i satisfies the Gaussian distribution with the prior distribution mean 0 and variance α_i^{-1}, expressed as follows:

$$P(w \mid a) = \prod_{n=0}^{N} N(w_n \mid 0, a_n^{-1}) \tag{14}$$

In the above formula, the hyperparameter $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_N]^T$, and each α_i has a w_i corresponding to it. According to the Bayesian principle, the expression of the posterior distribution of the weight w_i is as follows:

$$P(t \mid a, \sigma^2) = \begin{cases} = \int P(t \mid w, \sigma^2) P(t \mid a) dw \\ = (2\pi)^{-\frac{N}{2}} \mid C \mid^{-\frac{1}{2}} \exp\{-\frac{1}{2} t^T C^{-1} t\} \end{cases} \quad (15)$$

In Equation 15, C is the covariance, $C = \sigma^2 I + \phi A^{-1} \phi^T$. Taking the partial derivatives of the hyperparameters α and σ^2 and then making them equal to zero, we can obtain their iterative expressions:

$$a_i^{new} = \frac{r_i}{\mu_i^2} \quad (16)$$

$$(\sigma^2)^{new} = \frac{\| t - \phi u \|^2}{N - \sum_1^N r_i} \quad (17)$$

In the above formula, r_i is the ith element on the main diagonal. In the above iterative solution process, it is found that most of α_i is infinite, which means that the parameter control variable $p(w_i) = N(0, \alpha_i^{-1})$ obeys the standard normal distribution. Therefore, most of the weights are distributed near zero, which realizes the sparse representation of the variables. Variables that are not zero are called the "relevance vector".

2.7 Hybrid kernel function relevance vector machine based on improved GA optimization

By analysing the ageing curve of the battery, this paper finds that the capacity fading trend can be decomposed into a main degradation part and a local fluctuation part. The RVM models constructed by different kernel functions will produce different prediction results, and the influence of the kernel function will further increase in the long-term prediction process. Considering that the traditional RVM with a single kernel function has difficulty tracking the main linear degradation trend and the local nonlinear degradation part at the same time, this paper proposes a mixed kernel function RVM estimation method. The linear kernel function K_{f1} is used to describe the main linear ageing trend for the capacity, and the radial basis kernel function K_{f2} is used for the local fluctuation part, as shown in the following formula:

$$K_{f1}(x, z) = x^T z + c \quad (18)$$

$$K_{f2}(x, z) = \exp(-\frac{\| x - z \|^2}{2\sigma^2}) \quad (19)$$

The linear combination of multiple kernel functions can synthesize the expression ability of each kernel function of the model. The mathematical expression of the linear combination of the two kernel functions is as follows:

$$K_f = w_1 K_{f1}(x, z) + w_2 K_2(x, z) \quad (20)$$

In the formula, w_1 and w_2 are the weight coefficients of the linear combination of the kernel functions. The correct choice of the weight coefficients, w_1 and w_2, and the hyperparameters, k_1 and k_2, of the two kernel functions determines the accuracy of the model. Therefore, this paper will use the mean absolute error (MAE) as the objective function and use the improved genetic algorithm to perform global optimization on [w_1, w_2, k_1, k_2]. The "random selection by probability" method used by the genetic algorithm naturally avoids the local optimum trap that other optimization algorithms often encounter. However, when seeking a solution with higher accuracy, its local optimization ability is insufficient,

resulting in an excessively long calculation time. Therefore, the hill-climbing algorithm, with its excellent local selection ability, is selected to combine with the genetic algorithm, and the genetic hill-climbing (GA-HC) search algorithm is proposed, which can quickly find the optimal exact solution.

3. EXPERIMENTAL ANALYSIS

To evaluate the accuracy of the framework proposed in this paper, this paper selects three current mainstream error indicators, namely, the mean absolute error mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}| y_i - y_i | \tag{21}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}| \frac{y_i - y_i}{y} |\times 100\% \tag{22}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_i)^2}{n}} \tag{23}$$

where y_i is the estimated value of the model and $\hat{y}_i$ is the actual observation data [22].

3.1 Calculation results of the correlation coefficient between IHF and SOH

The Pearson coefficient and Spearman coefficient were used to quantitatively evaluate the strength of the IHF correlation proposed in this paper. After determining the approximate interval through the IC curve and then gradually calculating the field interval, the optimal feature extraction interval of B0005 and B0007 is 3.98 V-4.0 V, and the optimal extraction interval of Cell1 and Cell2 is 3.8 V-3.85. Table 1 gives the calculated results of the Pearson and Spearman coefficients of the IHF and SOH. The results show that the two correlation coefficients of IHF and SOH selected in this paper are both above 0.9, and the required fragment data are only 20 mV and 50 mV.

Table 1. Calculation results of the Pearson and Spearman coefficients of IHF and SOH

	B0005	B0007	Cell 1	Cell 2
Pearson	0.9949	0.9938	0.9934	0.9841
Spearman	0.9957	0.9957	0.9966	0.9934

3.2 Single battery experimental results

To verify the estimation ability of the proposed model under small data samples, the first 30% of the data of the 5, 6 and 7 cells in the NASA battery ageing dataset are selected as training data, and the last 70% of the data are selected as test data. Figure 6(a)-(c) shows the SOH estimation results for 3 batteries, where the red curve is the estimated SOH value, the blue curve is the actual value of the SOH

and the grey area represents the 95% confidence interval. Through the analysis, it can be found that the RVM model based on the mixed kernel function can accurately track the ageing state of the battery, and the estimated results are distributed within the 95% confidence interval. From the error box plot in Figure 6(d), it can be seen that the maximum estimation errors are kept within 2%. The calculation results of the MAE, MAPE and RMSE of the single-core RVM and H-RVM are shown in Table 2. The results show that the H-RVM still has a high estimation accuracy even when the sample data of the single-cell battery are small, and the average error is less than 1%.

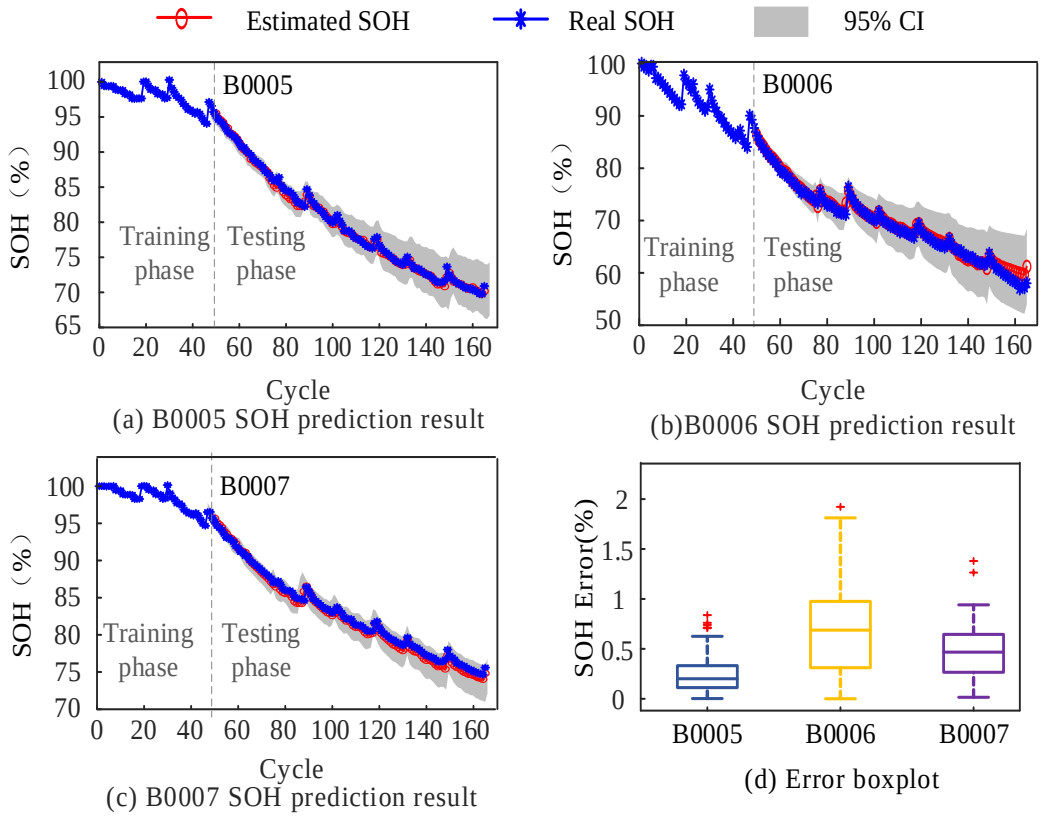


Figure 6. SOH prediction results and errors of the B0005, B0006 and B0007 lithium-ion batteries

Table 2. Error calculations of the estimated results for the B0005, B0006 and B0007 batteries

Model	Battery number	MAE/%	MAPE/%	RMSE/%
RVM	B0005	1.63	1.72	1.73
	B0006	2.14	2.35	2.27
	B0007	1.93	2.04	2.07
H-RVM	B0005	0.31	0.32	0.42
	B0006	0.71	0.95	0.81
	B0007	0.40	0.61	0.57

3.3 Multibattery experimental results

In this section, multibattery experiments are designed to verify the rationality and accuracy of the proposed prediction method. Each battery in the Oxford dataset is used as test data in turn, and the remaining batteries are used as training data. Figure 7 shows the results and error boxplots of 3 sets of multicell experiments. Table 3 lists the MAE, MAPE and RMSE of the multicell experimental results based on the single-core RVM and H-RVM.

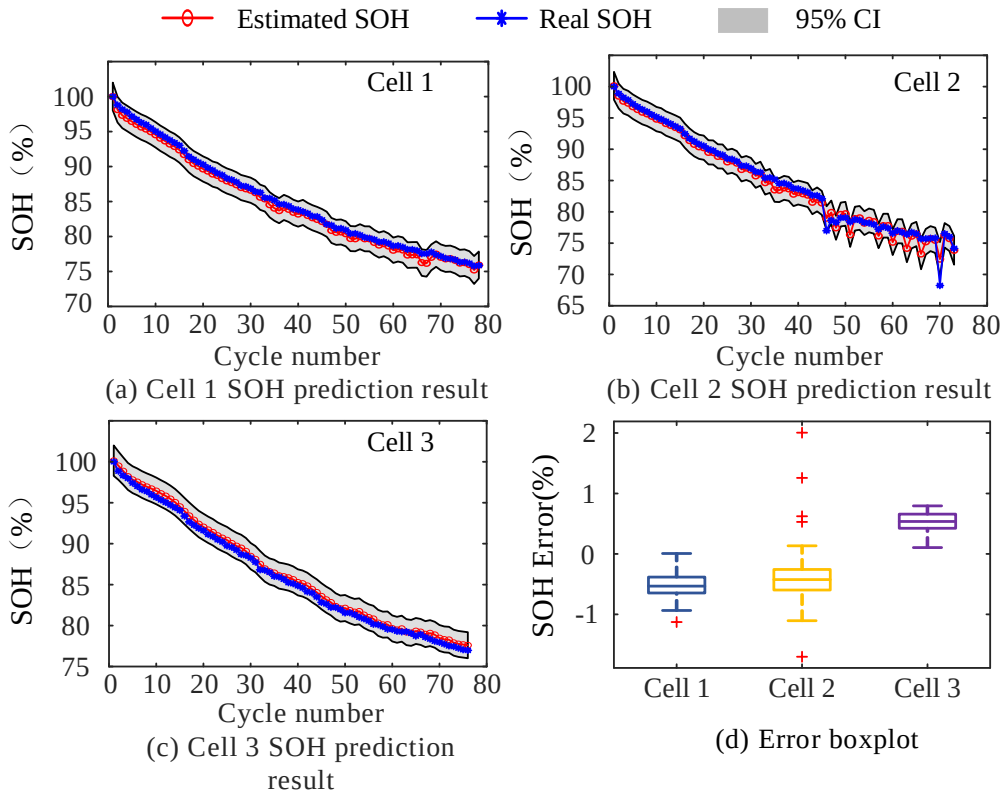


Figure 7. SOH estimation results and errors of Cell 1, Cell 2 and Cell 3

Table 3. Error calculations of the estimated results for the Cell 1, Cell 2 and Cell 3 batteries

Model	Training data	Test set	MAE/%	MAPE/%	RMSE/%
RVM	Cell 2, Cell 3	Cell 1	0.64	0.81	0.84
	Cell 1, Cell 3	Cell 2	1.17	1.34	1.36
	Cell 1, Cell 2	Cell 3	0.69	0.77	0.73
H-RVM	Cell 2, Cell 3	Cell 1	0.56	0.49	0.57
	Cell 1, Cell 3	Cell 2	0.62	0.70	0.82
	Cell 1, Cell 2	Cell 3	0.57	0.49	0.53

The training set and test set in the multibattery experiment are the data of different batteries, which are closer to the actual application situation. Figure 7 shows that the relative errors of the SOH estimation results based on the H-RVM are within 2%, except for individual points, which still have high accuracy. Table 3 quantitatively evaluates the prediction error of the multibattery experiment. The results show that, compared with the RVM with a single kernel function, the three error index values of the proposed framework are all within 1%. The experimental results verify that the estimation method based on the H-RVM can adapt to different training sets and has strong practicability and robustness.

3.3 RUL estimation result

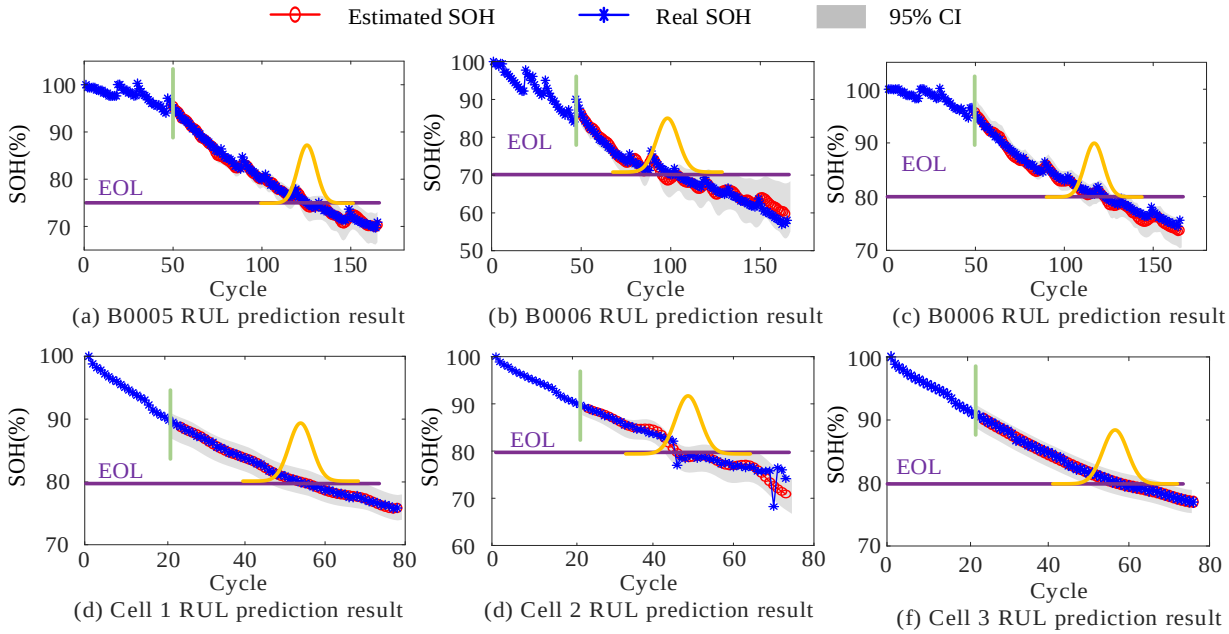


Figure 8. RUL estimation result

Table 4. Error in RUL estimation result

Battery No.	Starting Point of Prediction	Actual RUL	Predicted RUL	RUL uncertainty	Prediction Error
B0005	50	75	73	(70,85)	2
B0006	50	22	21	(17,33)	1
B0007	50	73	72	(63,79)	1
Cell 1	20	34	33	(26,44)	1
Cell 2	20	26	28	(24,36)	2
Cell 3	20	38	38	(32,46)	0

To estimate the RUL, it is necessary to predict the future IHF and use all the predicted values as the input of the H-RVM. The estimated RUL value is obtained by calculating the number of cycles between the current time and the EOL. In this paper, I-RVM is used to establish the mapping relationship between IHF and the cycle times, and the multistep predicted value of IHF is combined with the battery

ageing model to perform a multistep estimation of the SOH. Figure 8 shows the RUL estimation results. The starting estimation point for the NASA battery dataset is 50, and the starting estimation point for the Oxford battery dataset is 20. According to the different ageing conditions of the batteries, the EOL of the B0005 battery is set to 0.75, the EOL of the B0006 battery is set to 0.7 and the EOL of the remaining batteries is set to 0.8. Table 4 shows the estimation error of the RUL and gives the estimated interval of the RUL by the upper and lower bounds of the confidence interval. The estimation errors of RUL are all within 3 times, and it can be seen that the model proposed in this paper still maintains a high accuracy in long-term prediction.

3.4 Model Comparison Analysis

In previous work, this research team designed an IGPR based on the combined kernel function to obtain an online SOH estimation for the lithium-ion batteries and adopted the gradient optimization algorithm to optimize the parameters of the model [22]. However, this model only achieves a one-step prediction of the battery state of health and considers multistep predictions. On this basis, this paper designs an H-RVM model optimized by the GA-HC algorithm. Not only can the current SOH of the battery be predicted, but the RUL of the battery can also be estimated by predicting the SOH in multiple steps. By using the SOH and RUL joint estimation model, the accuracy is also improved on the basis of saving computing resources. The ageing of batteries can be identified on different time scales, which has important implications for engineering practices.

To verify the performance of the framework proposed in this paper, this paper compares it with the current mainstream algorithms, such as ANN, LSTM, SVM and GPR. The average errors of the estimation results of SOH and RUL are shown in Table 5. It can be seen that the average error using the H-RVM model proposed in this work is the lowest. The average MAE, MAPE and RMSE of the proposed method are 0.53%, 0.61% and 0.64%, respectively. It is worth noting that, compared with other algorithms, the H-RVM proposed in this paper not only estimates the RUL single point but also gives the uncertainty interval of the estimation result. The interval prediction of RUL has important guiding significance for practical work.

Table 5. Error analysis between the model proposed in this work and other models

Model	Average MAE/%	Average MAPE/%	Average RMSE/%	AE
ANN[23]	2.31	2.43	2.67	6
LSTM[24]	2.07	2.16	2.31	5
SVM[25]	1.86	1.77	1.81	4
GPR[26]	1.11	1.04	1.24	3
H-RVM[this work]	0.53	0.60	0.64	1

4. CONCLUSION

In this paper, a method for joint online estimation of the SOH and RUL of lithium batteries based on a hybrid kernel function is proposed. The model is experimentally verified by 6 sets of batteries in two public datasets. The results show that the method proposed in this paper maintains a high prediction accuracy in both the short-term and long-term prediction processes. This paper's main research contributions are reflected in three areas:

(1) By analysing the IC curve, the charging current, voltage and temperature of the battery are all used as battery ageing factors for feature extractions. To be more suitable for online operations, the data in the voltage intervals of 20 mV and 50 mV are used as the basis for feature extractions through correlation analyses.

(2) The RVM model with a linear kernel function and a Gaussian kernel function coupled with each other is adopted to track the linear and nonlinear parts of battery ageing, and the weight coefficients and hyperparameters are optimized using an improved genetic algorithm.

(3) By using the multistep prediction of IHF as the input of the battery ageing model, an online joint estimation of SOH and RUL is attained so that the battery health can be more accurately evaluated.

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. X. Liu, J. Li, Z. Yao, Z. Wang, R. Si, and Y. Diao, *Energy Rep*, 8 (2022) 217-223.
2. Z. Yang, D. Patil, and B. Fahimi, *IEEE Trans. Transp. Electrific*, 4 (2018) 147–156.
3. T. Shibagaki, Y. Merla, and G. J. Offer, *J. Power Sources*, 374 (2018) 188–195.
4. R. Li, W. Li, and H. Zhang, Int. *J. Electrochem. Sci*, 17 (2022) 220212.
5. Z. Chen, X. Shu, R. Xiao, W. Yan, Y. Liu, and J. Shen, *Int. J. Energy Res*, 43 (2019) 4344–4358.
6. L. Zhang, Z. Mu, and C. Sun, *IEEE Access*, 6 (2018) 17729–17740.
7. K. Qian, and X. Liu, *J Energy Storage*, 44 (2021) 103319.
8. X. Qiu, W. Wu, and S. Wang, *J. Power Sources*, 450 (2020) 227700.
9. A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, *J. Power Sources*, 239 (2013) 680-688.
10. P. Wang, L. Fan, and Z. Cheng, *CSEE Jpes*, 4 (2022) 1523-1534.
11. W. Xiong, Y. Mo, and C. Yan, *IEEE Access*, 9 (2021) 1870-1881.
12. Y. Tan, and G. Zhao, *IEEE T Ind Electron*, 67 (2020) 8723-8731.
13. L. Chen, H. Wang, B. Liu, Y. Wang, Y. Ding, and H Pan, *Energy*, 215 (2021) 119078.
14. K. Liu, Y. Shang, Q. Ouyang and W. D. Widanage, *IEEE Trans. Ind. Electron*, 68 (2021) 3170.
15. M. Ali, A. Zafar, S. H Nengroo, S. Hussain, and G. Park, *Energies*, 12 (2019) 4366.
16. B. Gou, Y. Xu, and X. Feng, *IEEE T Veh Technol*, 69 (2020) 10854-10867.
17. P. Li, Z. Zhang, R. Grosu, Z. Deng, J. Hou, Y. Rong, and R. Wu, *Renew Sust Energ Rev*, 156 (2020) 111843.

18. C. Birkl, Diagnosis and Prognosis of Degradation in Lithium-Ion Batteries, Univ. Oxford, Oxford, U. K, 2017.
19. D. T. Liu, J. B. Zhou, and H. T. Liao, *IEEE T Syst Man Cy-s*., 45 (2015) 915-928.
20. Q. Li, D. Li, K. Zhao, L. Wang, and K. Wang, *J Energy Storage*, 50 (2022) 104215.
21. Z. Chen, N. Shi, Y. Ji, M. Niu, and Y. Wang, *Energy*, 234 (2021) 121269.
22. H. Dong, L. Mao, K. Qu, J. Zhao, F. Li, and L. Jiang, *Int. J. Electrochem. Sci*, 17 (2022) 220832.
23. J. Wu, C. Zhang, and Z. Chen, *Appl Energ*, 173 (2016) 134-140
24. Y. Toughzaoui, S. BamatiToosi, H. Chaoui, H. Louahlia, R. Petrone, S. LeMasson, and H. Gualous, *J Energy Storage*, 51 (2020) 104520
25. J. Meng, L. Cai, G. Luo, D. Stroe, and R. Teodorescu, *Microelectron Reliab*, 88 (2018) 1216-1220.
26. Z. Lyu, R. Gao, and X. Li, *J. Power Sources*, 483 (2021) 229131.