IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 30, 2019, accepted October 9, 2019, date of publication October 14, 2019, date of current version October 28, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2947294

State-of-Health Prediction For Lithium-Ion Batteries With Multiple Gaussian Process Regression Model

XUEYING ZHENG AND XIAOGANG DENG, (Member, IEEE)
College of Control Science and Engineering, China University of Petroleum (East China), Qingdao 266580, China

Corresponding author: Xiaogang Deng (dengxiaogang@upc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61403418 and Grant 21606256, in part by the Shandong Provincial Key Program of Research and Development under Grant 2018GGX101025, in part by the Fundamental Research Funds for the Central Universities under Grant 17CX02054, and in part by the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China, under Grant ICT1900306.

ABSTRACT State-of-health (SOH) prediction for lithium-ion batteries is a challenging and important topic in the modern industry. With the advent of cloud-connected devices, there are huge amounts of the battery degradation trend data available. How to make full use of these existing degradation data for the SOH prediction is a valuable problem deserving deep research. Aiming at this problem, a multiple Gaussian process regression (MGPR) method is proposed for the SOH prediction of lithium-ion batteries. In this work, the health indicators (HIs) are firstly extracted from the charging process curves of the batteries, and the mutual information analysis is used to select the important HIs which are strongly correlated to the SOH. These selected HIs are applied as the regression model input for describing the aging procedure of the battery effectively. Then, Gaussian process regression modeling is performed on the different batteries to bring multiple GPR models. Lastly, a weighting strategy based on the prediction uncertainty is designed to integrate the predictions from the multiple GPR models. The method validations are executed on the battery datasets from NASA, and the results show that the proposed MGPR method has higher prediction accuracy than the basic GPR method.

INDEX TERMS Lithium-ion batteries, Gaussian process regression, mutual information, state-of-health.

I. INTRODUCTION

Because of high single-cell voltage, large energy density and long cycle life, lithium-ion batteries are widely used in many fields including electric vehicles, communication stations, and thermometers, etc. [1]–[3]. As the core component of various systems, batteries are often related to the system performance degradation or the complete failure. Therefore, the configuration of intelligent and efficient battery management system (BMS) is especially important for monitoring the safety, reliability and availability of batteries [4], [5]. The state-of-health (SOH) is used to quantitatively measure the ability of the lithium-ion battery system to store and deliver the energy while tracking the degradation trend of the battery [6]. Capacity and internal resistance are the typical health indicators for lithium-ion batteries. However, they are difficult to measure online due to many factors such as the series and parallel connection of the cells [7], [8]. Therefore, it is important to design the lithium-ion battery SOH prediction method by mining the intrinsic status information from the available online measured parameters.

The present SOH prediction technologies are mainly divided into two categories including model-based methods and data-driven methods. Model-based methods (e.g., equivalent circuit models, electrochemical models, etc.) require a large amount of prior knowledge to describe the degradation process and the failure mode. Jong and Cho [9] established a battery equivalent circuit model and used the extended Kalman filter (EKF) method to predict the SOH of lithium-ion batteries. Based on the existing empirical model [10], Yu *et al*. [11] developed a particle filter based SOH prediction method by considering the reduction of battery capacity and the self-charging behavior. Guha and Patra [12] combined the capacity reduction model and the internal resistance growth

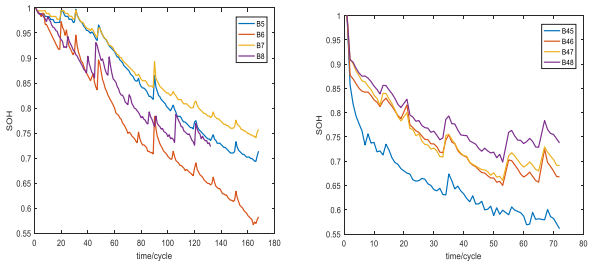
The associate editor coordinating the review of this manuscript and approving it for publication was Xiao-Sheng Si.

model for the SOH prediction of batteries. Wang *et al*. [13] developed a battery prediction method that relies on a state space model of discharge rate.

Data-driven methods extract the hidden information and evolution rule from the battery data to realize the SOH prediction of lithium-ion batteries. Different to the model-based methods, data-driven methods overcome the problem of requiring the accurate physical models. Wang *et al*. [14] presented a SOH prediction method based on multi-state non-linear Brownian motion. Principal component analysis (PCA) is one well-known data analysis tool [15]. Guo *et al*. [16] utilized PCA to optimize the health features extracted from the charging curve, and then used relevance vector machine to estimate the battery capacity. Li *et al*. [17] extracted the characteristics of the charge-discharge curve and used the ensemble learning neural network to improve the accuracy of SOH prediction performance. Yang *et al*. [18] built the SOH prediction model by training a three-layer back propagation (BP) neural network. These present data-driven methods usually give the deterministic results and do not consider the uncertainty of the predictions [19], [20].

As one emerging method, Gaussian process regression (GPR) is designed based on the Bayesian framework, and has the ability of explaining the uncertainty of the prediction results in probabilistic form. Because of its advantage over the other data-driven methods, GPR has been one of the important data-driven SOH prediction methods [21]. Li and Xu [22] applied GPR to identify the model parameters and then carried out the particle filter to predict the remaining useful life. Wang *et al*. [23] proposed the incremental capacity analysis to extract the peak and the position as the GPR input to predict the SOH. Yang *et al*. [24] extracted four features from the charging curve as the input of GPR, and applied the gray correlation method to analyze the correlation between features and SOH. Yu [25] carried out the empirical mode decomposition (EMD) to decompose the battery capacity time series and then integrated multi-scale logistic regression (LR) and GPR for the SOH prediction.

Although these aforementioned GPR methods have demonstrated the successful applications in the SOH prediction field, there are some limitations deserving further studies. One the one hand, data feature extraction and deep analysis are not enough. Although the present methods [23], [24] have designed some features for SOH prediction, the extracted features are still not sufficient and the importance on the SOH prediction is not analyzed comprehensively. One the other, the present methods perform the SOH prediction based on the historical data of the tested battery and omit the integrated utilization of the other available batteries [22]–[25]. In fact, with the development of cloud-connected devices, we can collect a lot of degradation data from the batteries installed on the different units. These existing different lithium-ion batteries can provide the important information on SOH prediction. As shown in Fig. 1 [26], [27], some lithium-ion batteries under the same conditions, collected from NASA [28], [29], have the similar attenuation



(a) B5, 6, 7 and 18 (b) B45, 46, 47 and 48

FIGURE 1. SOH curves of different battery under the same conditions.

curves during the decay process. Therefore, how to take full advantage of the existing other batteries to predict the given battery SOH is an important problem.

Based on the above analysis, this paper is to propose a multiple Gaussian process regression (MGPR) method for predicting the SOH of the lithium-ion battery. The main contributions include two aspects. (1) One contribution is about the construction and selection of the health indicators (HIs). We extract eight HIs by inspecting the trajectories of the charging processes, and the mutual information analysis is applied to select the key HIs according to the correlation between the HIs and the SOH. The key HIs are regarded as the prediction model input so that the aging of the battery can be reflected effectively. (2) Another contribution lies in the development of the multiple GPR (MGPR) model. Multiple existing batteries are well utilized as the model training sets and the single GPR model is built for each training battery. Then all the GPR models are integrated for a holistic MGPR model by the weighting strategy based on the prediction uncertainty.

The structure of this paper is listed as follows: Section II introduces the proposed model, including the HI selection rules using the mutual information analysis, and the MGPR model based on prediction variance. The experimental results including the training data under the same conditions and different conditions are further analyzed in Section III. Finally, Section IV draws some conclusions of the article.

II. METHODOLOGY

A. THE METHOD FRAMEWORK

Considering the sufficient utilization of the lithium-ion batteries data, this paper proposes a MGPR method for the SOH prediction of lithium-ion batteries. As shown in Fig. 2, the flowchart of the proposed method involves three parts: feature extraction, offline modeling, and online testing. Feature extraction is the basis of offline modeling and online testing. A series of health indications (HIs) are extracted by investigating the current, voltage and temperature curves during battery charging. The mutual information analysis is used to analyze the correlation between HIs and SOH, and some key HIs are obtained. In the offline modeling stage, several Gaussian process regression models are established using the lithium-ion battery data. The model input is the HIs obtained in the feature extraction stage, while the model output is

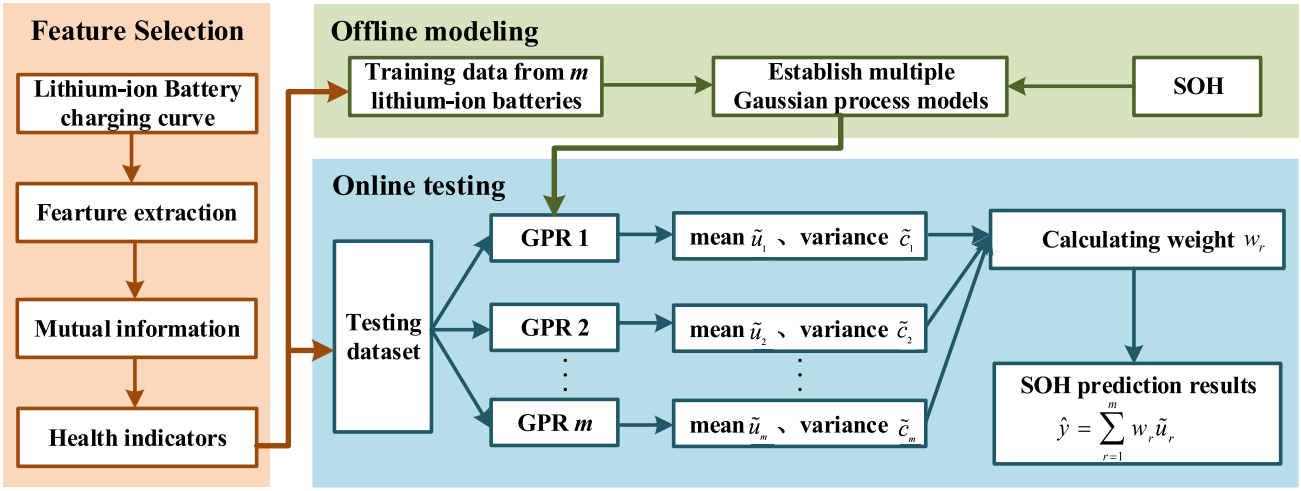


FIGURE 2. The SOH prediction flowchart based on the MGPR method.

the SOH. In the online testing phase, the multiple GPR models from the offline modeling stage are separately tested to obtain the corresponding prediction results and prediction variances. For each GPR model, its weight is computed based on the prediction uncertainty. Finally, the SOH prediction result of a lithium-ion battery is given by the weighted outputs of multiple GPR models.

B. FEATURE SELECTION USING MUTUAL INFORMATION

1) LITHIUM-ION BATTERY DATA

In this work, the cycling data of lithium-ion batteries are obtained from the Prognostics Center of Excellence (PCoE) at Ames Research Center, NASA [28], [29]. The charging process consists of two phases: (Constant current) CC phase and (Constant voltage) CV phase. In the CC phase, lithium-ion battery is charged at a constant current of 1.5A until the battery voltage reaches 4.2V. Then in the CV phase, the lithium battery is under the constant voltage until the charging current drops to 20mA. The discharge process is performed under a constant current. Table 1 shows the operating conditions and parameters for several sets of batteries. From this table, it can be seen that these batteries can be classified into several groups. E.g., B5, B6, B7, and B18 can be categorized together because of the same discharge current and the same ambient temperature.

SOH describes the health status of the battery and reflects the current rated capacity of the battery, expressed as:

$$SOH = \frac{C_t}{C_0}, \tag{1}$$

where C_t is the capacity of the t-th cycle and C_0 is the initial capacity. In addition to the rated capacity of the battery, the parameters such as internal resistance of the battery and number of cycles can also be used as characterization parameters for SOH.

2) FEATURE EXTRACTION

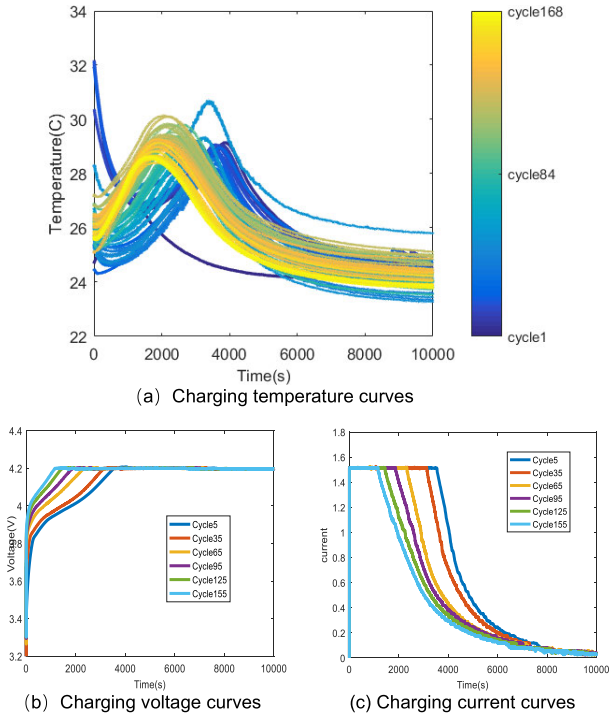
The voltage, current and temperature curves of the charging process of the battery B6 at different cycle times are shown in the Fig. 3. As the lithium-ion battery degrades, the voltage, current and temperature variations during charging process are regular. Therefore, extracting the HIs of the lithium-ion battery from the charging curve helps to reflect the deterioration trend of the battery. According to the characteristics of lithium-ion battery charging curve with different cycle times, eight HIs are extracted as follows.

TABLE 1. Batteries with their operating parameters.

Battery ID	Discharge current(A)	End voltage(V)	Ambient temperature
B5	2A	2.7	24℃
B6	2A	2.5	
B7	2A	2.2	
B18	2A	2.5	
B29	4A	2.0	43℃
B30	4A	2.2	
B31	4A	2.5	
B32	4A	2.7	
B33	4A	2.0	4℃
B34	4A	2.2	
B36	2A	2.7	
B45	1A	2.0	4℃
B46	1A	2.2	
B47	1A	2.5	
B48	1A	2.7	

a: HI1: THE TIME CORRESPONDING TO THE HIGHEST TEMPERATURE IN THE TIME INTERVAL FROM 1000s TO END

By most curves shown in Fig. 3 (a), we can see that the time to reach the maximum temperature $T_{\max}$ is gradually reduced as the cycle number increases. For example, the first



(a) Charging temperature curves

(b) Charging voltage curves

(c) Charging current curves

FIGURE 3. Charge profiles of B6 with different cycle times.

cycle arrives the highest temperature at the 3908s, while the 50th and 100th cycle arrive the highest temperature at the 3361s and 2335s, respectively. For avoiding the influence of the initial temperature, the first 1000s are excluded. The HI1 is defined as the time $T_{\max}$ corresponding to the highest temperature in the interval from 1000s to end.

$$\boldsymbol{HI1} = t_{T_{\max}}. \tag{2}$$

b: HI2: THE MAXIMUM TEMPERATURE IN THE INTERVAL FROM 1000s TO END

In the Fig. 3 (a), each maximum temperature in the temperature curve may be related to the battery attenuation. Although the maximum temperatures in the curve have no obvious trend in the whole cycles, the regular changes can be observed for some continuous cycles. For example, from the 122th to 130th cycle, the maximum temperature performs a decreasing trend. For a comprehensive consideration, the maximum temperature $T_{\max}$ in the time interval from 1000s to end is defined as the second HI:

$$\boldsymbol{HI2} = T_{\max}. \tag{3}$$

c: HI3: THE END TIME OF THE CC MODE

As shown in Fig. 3 (b) and (c), the end time t_{cc} of the CC mode decreases as the number of cycle increases. This time indicates the charge capacity in the CC mode. With this time, the battery polarization phenomenon can be exhibited. As the cycle number increases, the polarization becomes more serious and the end time t_{cc} decreases gradually. For example, the t_{cc} values for the 1st, 100th, and 160th cycle are 3609s, 1689s and 1066s, respectively. So, we define the CC mode end time t_{cc} as the third HI:

$$\boldsymbol{HI3} = t_{cc}. \tag{4}$$

d: HI4: THE TIME DURATION FROM 3.9V TO 4.2V UNDER THE CC MODE

In the CC mode, as the number of cycle increases, the time for the same voltage change are very different. Especially, at the end of CC mode, the slopes of different voltage curves have the clear distinction. Therefore, we define the time difference for the voltage change from 3.9V (v_1) to 4.2V (v_2) as the fourth HI:

$$\boldsymbol{HI4} = t_{v_2} - t_{v_1}, \tag{5}$$

where t_{v_1}, t_{v_2} are the time instants corresponding to 3.9V and 4.2V, respectively.

e: HI5: THE VOLTAGE INCREASEMENT IN 500s UNDER THE CC MODE

In the CC mode, different cycles have the different the voltage growth rates, which can be used as an indicator of the battery SOH. By defining the 3.9V as the starting point v_1, the voltage is increased to v_2 after 500s. The voltage increasement in 500s is defined as the fifth HI:

$$\boldsymbol{HI5} = v_2 - v_1. \tag{6}$$

This index is similar to the HI4. The difference is that HI4 is about the time change in the fixed voltage zone, while HI5 is the voltage change in the fixed time zone.

f: HI6: THE TIME DURATION FROM 1.2A TO 0.5A IN THE CV MODE

The CV mode plays an important role in weakening the polarization caused by CC charging. In the CV mode, the change in the current at different cycles is related to the battery degradation. The time duration required from 1.2A (i_1) to 0.5A (i_2) is defined as the sixth HI:

$$\boldsymbol{HI6} = t_{i_2} - t_{i_1}, \tag{7}$$

where t_{i_1} is the time instant at the 1.2A, while t_{i_2} is the time instant at the 0.5A. The longer the HI6 is, the more serious the battery aging is. For the first cycle, HI6 value is 982s. By contrast, the HI6 values of the 50th and 100th cycle are 1068s and 1266s, respectively.

g: HI7: THE CURRENT CHANGE IN 1000s UNDER THE CV MODE

In the CV mode, the voltage growth rate changes for different cycle times. By setting the starting current i_3 as 1.5A, the current after 1000s is denoted as i_4. The seventh HI is given by

$$\boldsymbol{HI7} = i_3 - i_4. \tag{8}$$

With the increasing cycle, the HI7 has a decreasing trend generally. The first cycle has the HI7 as 0.95A, while the 160th cycle's HI7 is 0.69A.

TABLE 2. The results of mutual information analysis.

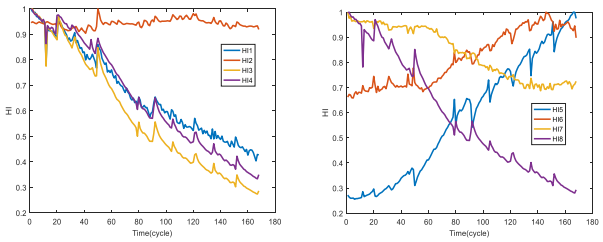
Battery ID	HI1	HI2	HI3	HI4	HI5	HI6	HI7	HI8
B5	0.7598	0.5143	0.7984	0.7608	0.7489	0.6614	0.6147	0.7891
B6	0.7496	0.5034	0.7748	0.7712	0.7409	0.6444	0.6593	0.7577
B29	0.8081	0.7322	0.8145	0.8215	0.8483	0.7579	0.7912	0.8500
B30	0.8377	0.7867	0.8620	0.8255	0.8470	0.7848	0.7337	0.8435
B33	0.5128	0.3301	0.5767	0.5957	0.5203	0.3918	0.2647	0.5767
B36	0.5701	0.2794	0.4825	0.4433	0.4088	0.2672	0.2638	0.4789
B45	0.5677	0.5712	0.6519	0.6584	0.5232	0.6821	0.6055	0.6584
B46	0.5984	0.6087	0.7059	0.6591	0.6882	0.6632	0.6357	0.6980

h: HI8: THE AREA UNDER THE CC MODE

The voltage area in the CC mode decreases as the battery decays. Denote the end of the CC mode as t_{cc}, and the integration of the voltage $v(t)$ in the CC mode is expressed by

$$\boldsymbol{HI8} = \int_0^{t_{cc}} v(t)dt. \tag{9}$$

Take the battery B6 as one example, and eight HIs are extracted and normalized as shown in Fig. 4. By this figure, we can see that all these indices change regularly regarding to the cycle number. The next task is to select some key indices for the SOH prediction modeling.



(a) HI1-4 (b) HI5-8

FIGURE 4. Eight HIs of battery B6.

3) MUTUAL INFORMATION

In order to select the key HIs for GPR modeling, some quantitative analyses are critical. Mutual information is a measure of the correlation degree between the two variables. Assuming two random variables $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$, the mutual information between them can represent a measure of how much the $\boldsymbol{g}_1$ can provide information for $\boldsymbol{g}_2$. The mutual information value determines the importance of the variable $\boldsymbol{g}_1$ to the variable $\boldsymbol{g}_2$, which facilitates a more rational selection of the input variable [30]. Therefore, mutual information can reveal the correlation between nonlinear variables, which is consistent with the nonlinear characteristics of lithium-ion battery sample data.

For two given random variables $\boldsymbol{g}_1 = [g_{11}, g_{12}, \cdots, g_{1n}]$ and $\boldsymbol{g}_2 = [g_{21}, g_{22}, \cdots g_{2n}]$ with n samples, the mutual information is denoted by

$$im(\boldsymbol{g}_1, \boldsymbol{g}_2) = \sum_{i=1}^{n} p(g_{1i}, g_{2i}) \log\left(\frac{p(g_{1i}, g_{2i})}{p(g_{1i})p(g_{2i})}\right), \tag{10}$$

where $p(g_{1i}, g_{2i})$ is the joint probability function, while $p(g_{1i})$ and $p(g_{2i})$ are the marginal probability distribution functions. The mutual information can be computed by the histogram method [30], [31].

In the Table 2, we list the mutual information values between the 8 HIs and the SOH for the 8 applied batteries. Furthermore, for each battery, the mutual information values are sorted, as shown in Table 3. Then the order sum of each HI is calculated and listed in the last row of the Table 3. It can be seen that HI3, HI4, HI8 are corresponding to the three smallest MI order sum values. So, we select these three HIs as the input to the model.

C. MULTIPLE GAUSSIAN PROCESS REGRESSION MODELING

Gaussian Process (GP) is a collection of finite number of random variables following Gaussian distribution [32]–[34]. The MGPR method is to establish the multiple basic GPR models and combine their outputs to provide the holistic result. For the r-th basic GPR model, it is used to approximate the target output $f_r(\boldsymbol{x}_r)$, following the probability distribution as

$$f_r(\boldsymbol{x}_r) \sim GP(m(\boldsymbol{x}_r), k(\boldsymbol{x}_r, \boldsymbol{x'}_r)), \tag{11}$$

where $\boldsymbol{x}_r$ is d-dimensional input vectors, $m(\boldsymbol{x}_r)$, $k(\boldsymbol{x}_r, \boldsymbol{x'}_r)$ are mean and covariance function, respectively, expressed by

$$m(\boldsymbol{x}_r) = E[f_r(\boldsymbol{x}_r)], \tag{12}$$

$$k(\boldsymbol{x}_r, \boldsymbol{x'}_r) = E[(f_r(\boldsymbol{x}_r) - m_r(\boldsymbol{x}_r))(f_r(\boldsymbol{x'}_r) - m_r(\boldsymbol{x'}_r))]. \tag{13}$$

Usually, $m(\boldsymbol{x}_r)$ can be set to be zero. $k(\boldsymbol{x}_r, \boldsymbol{x'}_r)$ is selected as the squared exponential covariance (SE) function, denoted by

$$k(\boldsymbol{x}_r, \boldsymbol{x'}_r) = \sigma_{fr}^2 \exp\left(-\frac{(\boldsymbol{x}_r - \boldsymbol{x'}_r)^2}{2l_r^2}\right), \tag{14}$$

where the signal variance σ_{fr}^2 represents the output scale, the l_r is the characteristic length scales. In many cases,

TABLE 3. The orders of the mutual information values.

Battery ID	HI1	HI2	HI3	HI4	HI5	HI6	HI7	HI8
B5	4	8	1	3	5	6	7	2
B6	4	8	1	2	5	7	6	3
B29	5	8	4	3	2	7	6	1
B30	4	6	1	5	2	7	8	3
B33	5	7	3	1	4	6	8	2
B36	1	6	2	4	5	7	8	3
B45	7	6	4	3	8	1	5	2
B46	8	1	2	5	3	7	4	6
Sum	38	50	18	26	34	48	52	22

we assume that the observed data can be represented as an implicit function with Gaussian noise as

$$\boldsymbol{y}_r = f_r(\boldsymbol{X}_r) + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(0, \sigma_{nr}^2), \tag{15}$$

where $\boldsymbol{X}_r$ is the data matrix composed of the n observations of the input vector $\boldsymbol{x}_r$, $\boldsymbol{y}_r$ is the observation corresponding to $\boldsymbol{X}_r$, $\boldsymbol{\varepsilon}$ is white noise whose mean is 0 and variance is $\boldsymbol{\sigma}_{nr}^2$. Hence the prior distribution of $\boldsymbol{y}_r$ is denoted as

$$\boldsymbol{y}_r \sim N(0, K(\boldsymbol{X}_r, \boldsymbol{X}_r) + \sigma_{nr}^2 \boldsymbol{I}_n), \tag{16}$$

where $K(\boldsymbol{X}_r, \boldsymbol{X}_r)$ is the n-dimensional symmetric positive definite matrix and its element $K_{ij} = k(\boldsymbol{X}_r(i,:), \boldsymbol{X}_r(j,:))$ describes the correlation between the i-th sample $\boldsymbol{X}_r(i,:)$ and the j-th sample $\boldsymbol{X}_r(j,:)$, $\boldsymbol{I}_n$ is the n-dimensional identity matrix.

The hyper-parameter set $\Theta_r = [\sigma_{fr}, l_r, \sigma_{nr}]$ can be determined by maximizing the log-likelihood, denoted by

$$L_r = \log p(\boldsymbol{y}_r|\boldsymbol{X}_r, \Theta_r) = -\frac{1}{2}\boldsymbol{y}_r^T [K(\boldsymbol{X}_r, \boldsymbol{X}_r) + \sigma_{nr}^2 \boldsymbol{I}_n]^{-1} \boldsymbol{y}_r - \frac{1}{2}\log(\det(K(\boldsymbol{X}_r, \boldsymbol{X}_r) + \sigma_{nr}^2 \boldsymbol{I}_n) - \frac{n}{2}\log 2\pi. \tag{17}$$

The conjugate gradient method is used to find the optimal hyper-parameter, and the maximum value of the objective function is obtained by the derivative of the log-likelihood function. The optimization process is solved based on

$$\begin{cases} \dfrac{\partial}{\partial \Theta_r}\log(\boldsymbol{y}_r|\boldsymbol{X}_r, \boldsymbol{\Theta}_r) = \dfrac{1}{2} tr\left\{\left[\boldsymbol{\alpha}\boldsymbol{\alpha}^T - (K(\boldsymbol{X}_r, \boldsymbol{X}_r) + \sigma_{nr}^2 \boldsymbol{I}_n)^{-1} \dfrac{\partial(K(\boldsymbol{X}_r, \boldsymbol{X}_r) + \sigma_{nr}^2 \boldsymbol{I}_n)}{\partial \boldsymbol{\Theta}_r}\right]\right\} \\ \boldsymbol{\alpha} = \left[K(\boldsymbol{X}_r, \boldsymbol{X}_r) + \sigma_{nr}^2 \boldsymbol{I}_n\right]^{-1} \boldsymbol{y}_r. \end{cases} \tag{18}$$

For the test input $\boldsymbol{x}^*$, the joint distribution of the observed value $\boldsymbol{y}_r$ and the predicted output y^* are expressed as

$$\begin{bmatrix} \boldsymbol{y}_r \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(\boldsymbol{X}_r, \boldsymbol{X}_r) + \sigma_{nr}^2 \boldsymbol{I}_n & K(\boldsymbol{X}_r, \boldsymbol{x}^*) \\ K(\boldsymbol{x}^*, \boldsymbol{X}_r) & K(\boldsymbol{x}^*, \boldsymbol{x}^*) \end{bmatrix}\right), \tag{19}$$

where $K(\boldsymbol{X}_r, \boldsymbol{x}^*) = K(\boldsymbol{x}^*, \boldsymbol{X}_r)^T$ is the covariance matrix between $\boldsymbol{X}_r$ and $\boldsymbol{x}^*$, and $K(\boldsymbol{x}^*, \boldsymbol{x}^*)$ is the covariance of $\boldsymbol{x}^*$.

The posterior distribution of the predicted value y^* is expressed by

$$p(y^*|\boldsymbol{X}_r, \boldsymbol{y}_r, \boldsymbol{x}^*) = N(\tilde{u}_r, \tilde{c}_r), \tag{20}$$

where the prediction mean $\tilde{u}_r$ and the prediction variance $\tilde{c}_r$ are given as

$$\tilde{u}_r = K(\boldsymbol{x}^*, \boldsymbol{X}_r)[K_r(\boldsymbol{X}_r, \boldsymbol{X}_r) + \sigma_{nr}^2 \boldsymbol{I}]^{-1} \boldsymbol{y}_r, \tag{21}$$

$$\tilde{c}_r = K(\boldsymbol{x}^*, \boldsymbol{x}^*) - K(\boldsymbol{x}^*, \boldsymbol{X}_r)^T \times [K(\boldsymbol{X}_r, \boldsymbol{X}_r) + \sigma_{nr}^2 \boldsymbol{I}]^{-1} K(\boldsymbol{X}_r, \boldsymbol{x}^*). \tag{22}$$

Then, for the test input $\boldsymbol{x}^*$, its prediction mean $\tilde{u}_r$ and variance $\tilde{c}_r$ from the r-th GPR model can be obtained. The prediction mean is used as the prediction result, and the prediction variance $\tilde{c}_r$ gives a confidence interval to represent the uncertainty corresponding to each measurement result. The 95% confidence interval is given by

$$\left[\tilde{u}_r - 1.96\sqrt{\tilde{c}_r}, \tilde{u}_r + 1.96\sqrt{\tilde{c}_r}\right]. \tag{23}$$

In the MGPR model, the uncertainty of the predicted value $\tilde{u}_r$ obtained by the r-model is used to determine the weight w_r of the model to determine the accuracy of each GPR prediction. The uncertainty is related to $\tilde{c}_r$, and the weight of each basic GPR model can be obtained by prediction variance $\tilde{c}_r$ as

$$w_r = \frac{1/\tilde{c}_r}{\sum_{j=1}^{m} 1/\tilde{c}_j}. \tag{24}$$

Then the final output of the MGPR is

$$\hat{y} = \sum_{r=1}^{m} w_r \tilde{u}_r. \tag{25}$$

One problem which should be noted is about the prediction uncertainty. The GPR based methods has the advantage of explaining the prediction uncertainty. For the single GPR model, the prediction uncertainty can be obtained by

TABLE 4. The prediction results for the battery B6.

	1-GPR	2-MGPR(i)	2-MGPR(ii)	3-MGPR	ZGPR
MAPE(%)	7.0850	1.7198	1.6125	1.6535	2.2079
RMSE	0.0624	0.0173	0.0173	0.0167	0.0230

Eq. (23). However, in the proposed method, the prediction uncertainty of each single GPR model is utilized to design the model weight for a weighted MGPR model. Therefore, the prediction uncertainty of MGPR model is missing. The determination of the MPGR prediction uncertainty is one problem deserving the future studies.

III. SIMULATIONS

In order to verify the proposed MGPR method, we use the NASA battery dataset to perform simulation experiments. The applied batteries are the commercially available Li-ion 18650 sized rechargeable batteries, which are tested under three different operational profiles (charge, discharge, and impendence) [26], [29]. The accuracy of the SOH estimation is evaluated in terms of root mean square error (RMSE) and absolute error percentage (MAPE) [35]. The RMSE calculation formula is given by

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i^* - \hat{y}_i)^2}{n}}, \tag{26}$$

while the MAPE calculation formula is defined as

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^* - \hat{y}_i}{y_i^*}\right| \times 100\%, \tag{27}$$

where y_i^* is the true value, $\hat{y}_i$ is the estimated value, and n is the number of samples.

A. BATTERIES IN THE SAME CONDITIONS

In the real industrial scenarios, there are often some batteries which are with the similar running conditions. The similar batteries have the similar degradation mechanisms, and it is possible to achieve better prediction results by integrating these existing similar batteries for the SOH prediction modeling. To verify this, two groups of similar batteries are illustrated. One group involves the batteries B5, B6, B7 and B18, while another group includes the batteries B45, B46, B47 and B48.

1) CASE ONE

In this case, we select the batteries B5, B7, and B18 as training set, and apply the battery B6 as a testing set. All these involved batteries have the same conditions, which can be seen by Table 1. To give a deep discussion, five models are developed as follows.

a) 1-GPR model with individual training data sets: the method is to establish a basic GPR model in which the battery B18 is used to train the model and the battery B6 is applied for testing.

b) 2-MGPR(i) model with two training data set: the method is to establish the MGPR model, in which B5 and B18 are used for training, and B6 is tested.

c) 2-MGPR(ii) model with two training data sets: the method is to establish the MGPR model, in which B7 and B18 are used for training, and B6 is tested.

d) 3-MGPR model with three training data sets: the method is to establish the MGPR model, in which B5, B7 and B18 are used for training, and B6 is tested.

e) ZGPR model with three training data sets: The method is to establish a basic GPR model in which B5, B7, B18 are trained and B6 are tested.

The test results of all models are plotted in Fig. 5 and the performance indices are listed in Table 4. The SOH curves used for training and testing are listed in Fig. 5 (a)-(d) for the five applied models, respectively. The prediction results by 1-GPR model is seen in Fig. 5 (e) with a large prediction bias, which has the RMSE of 0.0624 and the MAPE of 7.085%. When two batteries are used for model training, that means the application of 2-MGPR(i) and 2-MGPR(ii) models, the results are seen in the Fig. 5 (f) and (g), respectively. It is clear that their performances outperform the basic 1-GPR method. Compared to the 1-GPR model, 2-MGPR(i) and 2-MGPR(ii) methods reduce the RMSE from 0.0624 to 0.0173, which improve the prediction accuracy by 72.3%. However, due to the use of different training sets, 2-MGPR(i) and 2-MGPR(ii) have the different MAPE values, which are both lower than 1-GPR MAPE. For the 3-MGPR model shown in Fig. 5(h), the RMSE is further reduced to 0.0167 because of three training datasets are involved. Therefore, increasing the number of data sets can enhance the diversity of the model and improve the accuracy of the prediction. However, it should be noted that to perform the basic GPR modeling on the multiple datasets simply may not result in the good SOH prediction performance. To validate this, the ZGPR model, which is a basic GPR model trained by the three battery datasets, is carried out and the SOH prediction result on the battery B6 is given the Fig. 5 (i). The RMSE and MAPE of ZGPR are 0.023, and 2.2079%, respectively, which are larger than the MGPR's. To compare the results of 3-MGPR and ZGPR, we can observe that the MGPR method can provide better SOH prediction performance when the same training datasets are used. The absolute error curves of all the methods are given in Fig. 5 (j), while the error analysis is demonstrated in Fig. 5 (k). In this case, the ZGPR model has lower prediction error than the 1-GPR method because

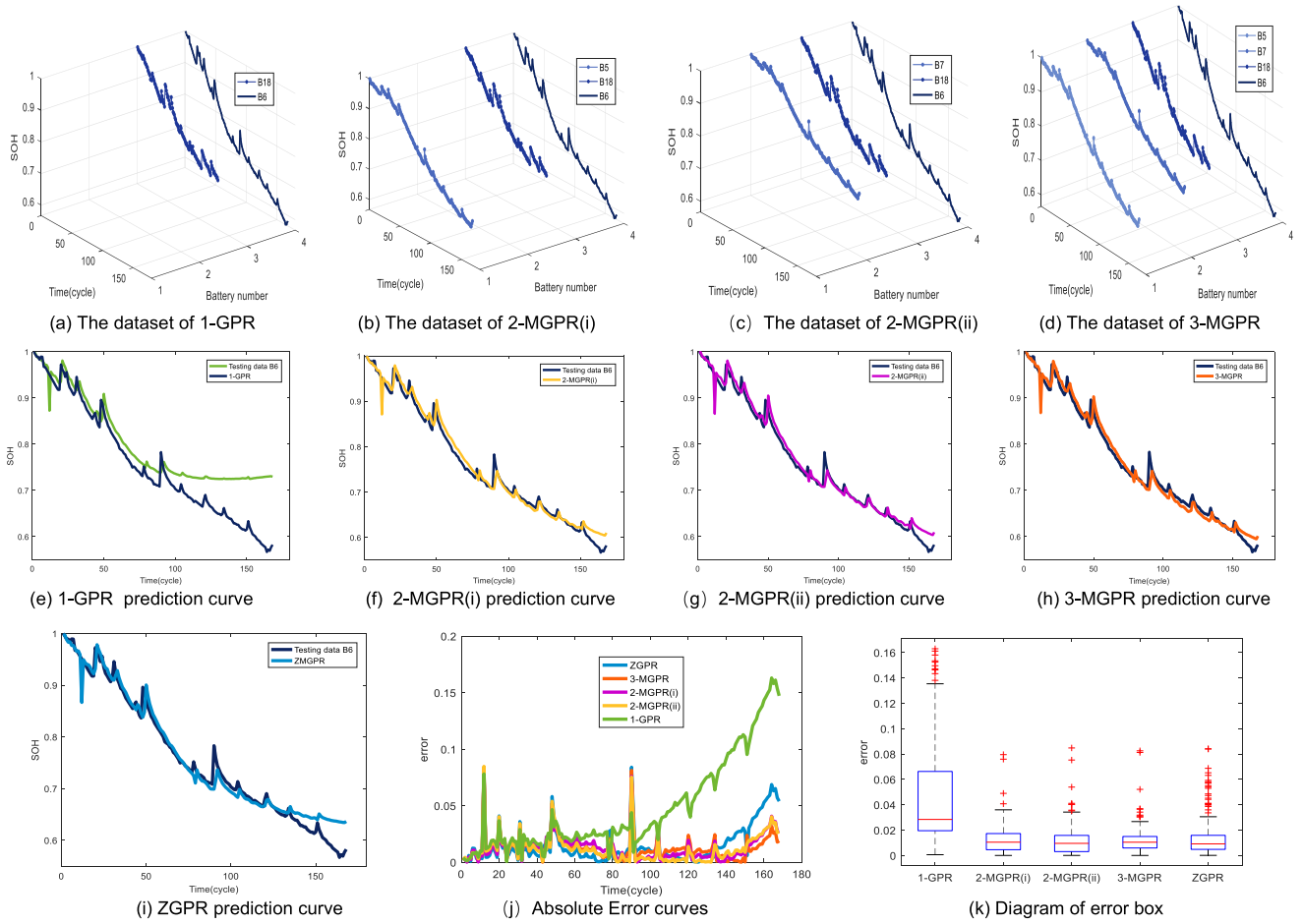


FIGURE 5. The prediction results of the battery B6. (a)-(d) are for more clearly indicating the relationship between the training set and the test set, where the y-axis 1-3 are corresponding to the training set B5, B7, B18, respectively, and the y-axis 4 is the test set B6. (e)-(i) represent the comparison of the predicted results of each model with the test values. (j) shows the error plots for the five methods. (k) is a wire box diagram of the error.

of the introduction of more training datasets. Furthermore, the proposed MGPR method outperforms the ZGPR method because of the multiply modeling strategy based on the model weighting.

2) CASE TWO

This case selects B45, B46, and B48 as the training sets, and B47 as a testing set. The five models of the experiment are as follows:

a) 1-GPR model with individual training data set: the method is to establish a basic GPR model in which B48 is trained and B47 battery is tested.

b) 2-MGPR(i) model with two training data sets: the method is to establish the MGPR model, in which B45 and B48 are used for training, and B47 is tested.

c) 2-MGPR(ii) model with two training data sets: the method is to establish the MGPR model, in which B46 and B48 are used for training, and B47 is tested.

d) 3-MGPR model with three training data sets: the method is to establish the MGPR model, in which B45, B46 and B48 are used for training, and B47 is tested.

e) ZGPR model with three training data sets: The method is to establish a basic GPR model, in which B45, B46, B48 batteries are trained and B47 is tested.

The test results are shown in Fig. 6 and Table 5. Fig. 6 (a)-(d) plot the SOH curves of the training and testing batteries for different methods. The SOH prediction using the 1-GPR method, shown in the Fig. 6 (e), is not satisfactory with the large RMSE 0.0761 and the large MAPE 9.6808%. The results of 2-MGPR(i) and 2-MGPR(ii) are shown in Fig. 6 (f) and (g), respectively. The two models have higher prediction accuracy, which reduce the RMSE from 0.0761 to 0.0415 and 0.0384, respectively. The prediction RMSE is improved by 45.5% and 49.5%, respectively. Also MAPE is reduced to 4.559% and 3.8893%, respectively. With more training datasets including B45, B46 and B48, the 3-MGPR model achieves the better SOH prediction result, which reduces the RMSE by 22.4% and 16.1% compared to 2-MGPR(i) and 2-MGPR(ii), respectively. It should be noted that ZGPR can not achieve a better prediction than 1-GPR in this case. The RMES and MAPE of ZGPR are 0.2606 and 23.1902%, respectively, which are clearly larger than the

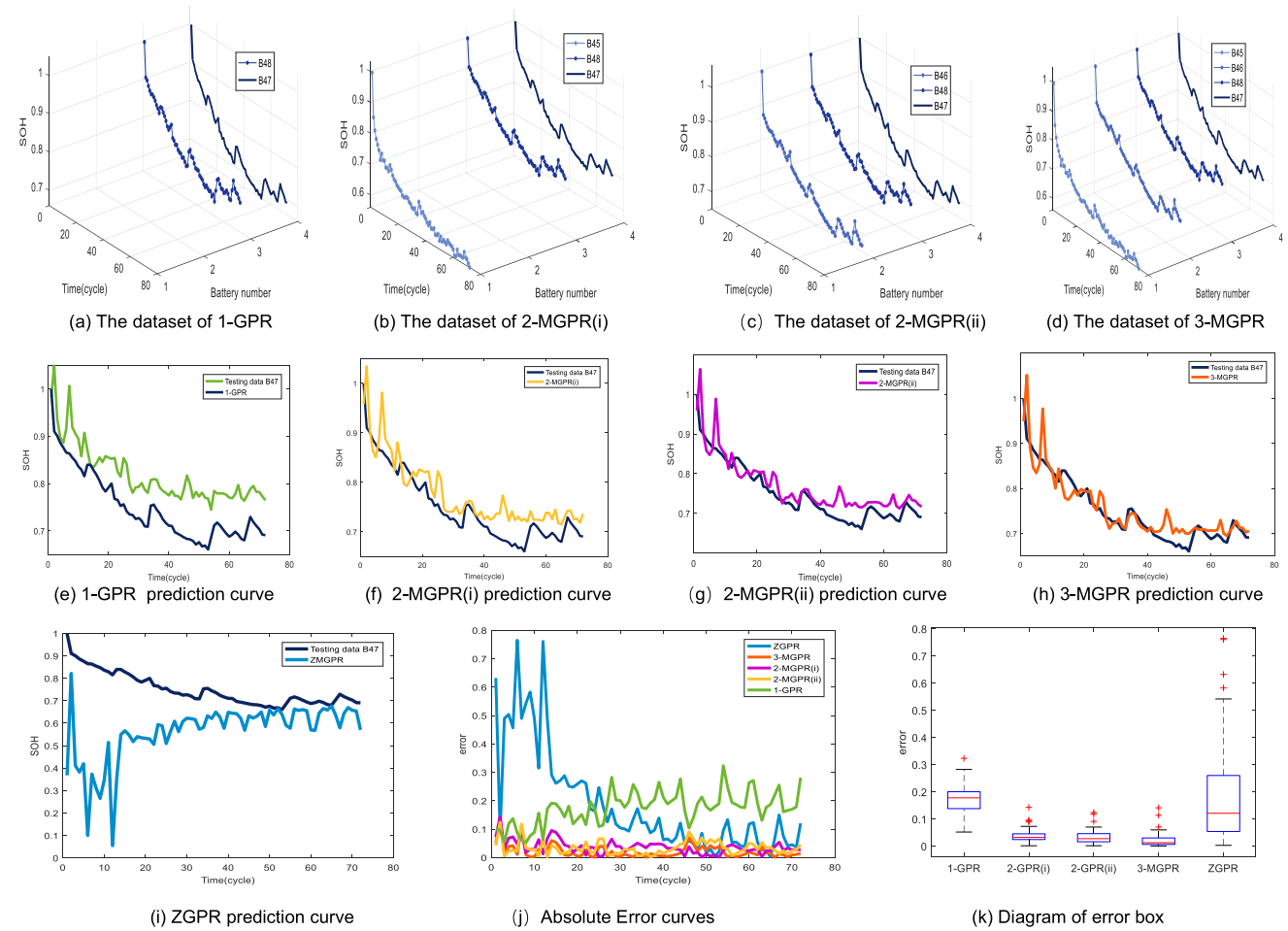


FIGURE 6. The prediction results of the battery B47. (a)-(d) are for more clearly indicating the relationship between the training set and the test set, where the y-axis 1-3 are corresponding to the training set B45, B46, B48, respectively, and the y-axis 4 is the test set B47. (e)-(i) represent the comparison of the predicted results of each model with the test values. (j) shows the error plots for the five methods. (k) is a wire box diagram of the error.

TABLE 5. The prediction results for the battery B47.

	1-GPR	2-MGPR（i）	2-MGPR（ii）	3-MGPR	ZGPR
MAPE(%)	9.6808	4.5590	3.8893	2.8020	23.1902
RMSE	0.0761	0.0415	0.0384	0.0322	0.2606

other methods. This further demonstrates that the simple use of more training datasets without the multiple modeling strategy can not ensure the performance improvement. The training datasets with the different characteristic may bring the increasing of prediction error. By contrast, the MGPR method can make the rational use of the multiple training datasets and distinguish the influence of different training datasets by the weighting strategy. The absolute error curves and the error block diagrams of the five methods are given in Fig. 6 (j) and (k). We can observe that 3-MGPR method performs best.

B. BATTERIES IN DIFFERENT CONDITIONS

In order to further verify the effectiveness of the algorithm, several sets of experimental data under different experimental conditions are also selected for method testing. In this case, the SOH prediction model is trained using B5, B29, B34, and is tested on B6. It should be pointed out that B5 and B6 are from the same conditions, while the batteries B29 and B34 involve different experimental conditions. The experimental results are shown in Fig. 7 and Table 6. The 1-GPR models trained with B5, B29 and B34 obtain the RMSE values of 0.0176, 0.1654 and 0.0776, respectively. The results are reasonable because the batteries under different conditions lead to the high prediction error, while the battery from the same condition can provide a good prediction model. With all the three batteries for GPR modeling, the ZGPR method has the RMSE of 0.0262 and the MAPE of 2.9605%. This is better than the GPR models based on the single battery B29 and B34. When the proposed

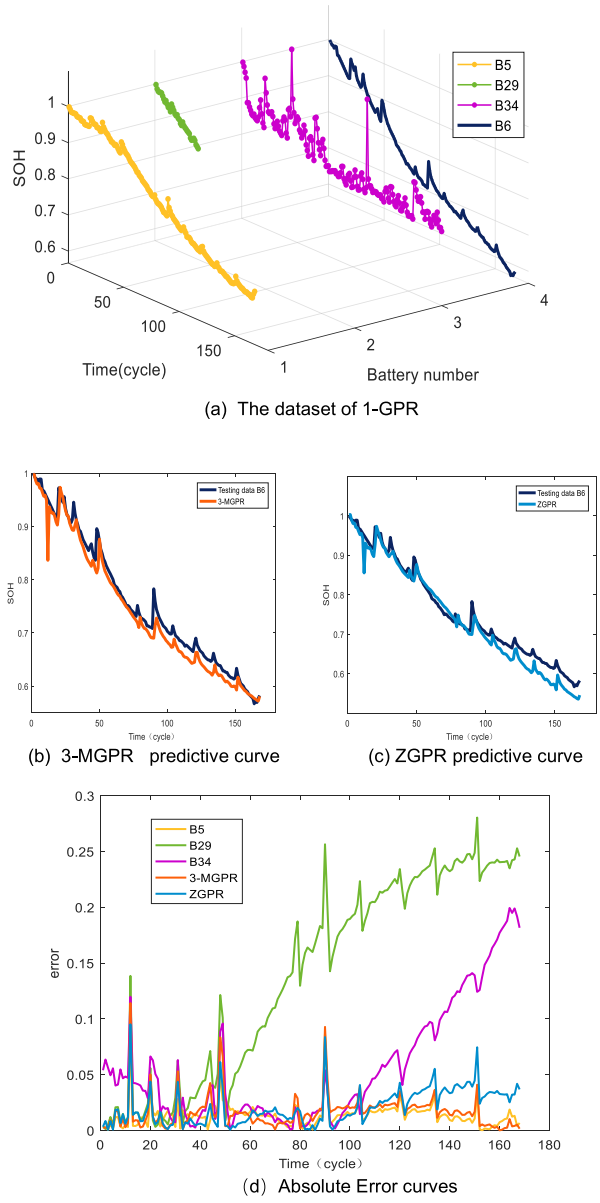


FIGURE 7. The prediction results of B6 battery. (a)-(d) are for more clearly indicating the relationship between the training set and the test set, where the y-axis 1-3 are corresponding to the training set B5, B29, B34, and the y-axis 4 is the test set B6. (b)-(c) represent the comparison of the predicted results of 3-GPR and ZGPR with the test values. (d) shows the error plots for the five methods.

TABLE 6. The prediction results for the battery B6 based on the training batteries with different conditions.

	1-GPR (B5)	1-GPR (B29)	1-GPR (B34)	3-MGPR	ZGPR
MAPE (%)	1.8439	20.5495	8.4404	2.1912	2.9605
RMSE	0.0176	0.1654	0.0776	0.0224	0.0262

MGPR method is applied, the MGPR method still predicts the trend of SOH degradation relatively well with the RMSE of 0.0224 although the batteries used for model training include the different types of batteries. Looking back the results in Table 4, the SOH prediction RMSE by the 3-MGPR is 0.0167. That means MGPR considering the batteries under different conditions may lead to a slight performance degradation, but the results from MGPR are still satisfactory.

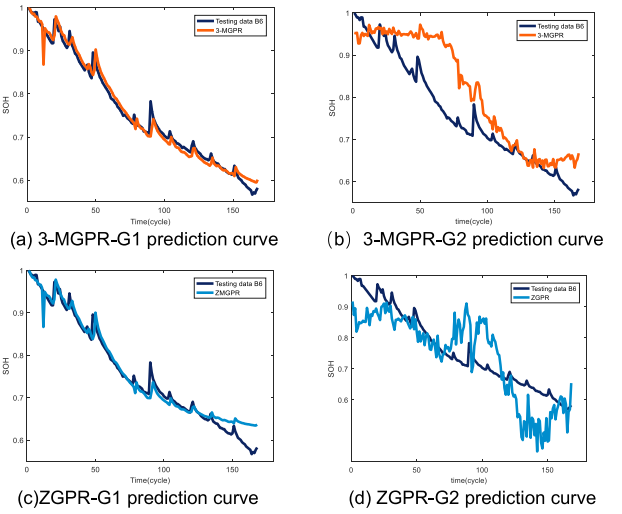


FIGURE 8. The prediction results of B6 Battery by the 3-MGPR and ZGPR with different index groups.

TABLE 7. The prediction results of B6 battery with different index groups.

	3-MGPR-G1	ZGPR-G1	3-MGPR-G2	ZGPR-G2
MAPE(%)	1.6535	2.2079	8.2487	9.6071
RMSE	0.0167	0.0230	0.0797	0.0869

C. FURTHER DISCUSSIONS

In this paper, we apply the mutual information analysis to select the key HIs. To demonstrate the effectiveness of the index selection strategy based on the mutual information analysis, we give some further discussions. For the case one in the section III-A, we develop the SOH prediction models via two groups of indices, respectively. Group 1 involves the indices HI3, HI4 and HI8, which have been identified as the key HIs in the previous discussions. For comparison, Group 2 includes the indices HI2, HI6 and HI7. Based on these two groups of HIs, the 3-MGPR and ZGPR models are developed and the corresponding models are denoted by 3-MGPR-G1, 3-MGPR-G2, ZGPR-G1 and ZGPR-G2, respectively. The SOH predictions on the battery B6 are plotted in the Fig. 8, and the corresponding statistical results are given in Table 7. For the 3-MGPR model, it can be seen that group 1 in the Fig. 8 (a) gives better SOH prediction than group 2. The RMSE obtained by the 3-MGPR model with group 1 is the 0.0167, while the RMSE from the 3-MGPR model with group 2 is 0.0797. The MAPE values for group 1 and group 2 are 1.6535% and 8.2487%, respectively. It is obvious that the 3-MGPR model with the index group 2 brings the higher prediction error. Similar results

can be observed for the ZGPR method. When the group 2 is applied, the model prediction RMSE and MAPE are 0.0869, and 9.6071%, respectively, which are larger than the ones from group 1. So, the group 1, selected based on the mutual information analysis, can predict the change of the SOH more precisely. These results demonstrate the effectiveness of the proposed HI selection strategy.

TABLE 8. The SOH prediction results for all the batteries.

Case No.	Tested battery	RMSE		MAPE(%)	
		3-MGPR	ZGPR	3-MGPR	ZGPR
1	B 5	0.0096	0.0109	0.8047	0.9448
	B6	0.0167	0.0230	1.6535	2.2079
	B7	0.0129	0.0120	1.2099	1.0530
	B18	0.0228	0.0579	2.2627	6.3816
2	B45	0.1279	0.1238	19.6709	19.7455
	B46	0.0333	0.0609	4.1978	7.0702
	B47	0.0322	0.2606	2.8021	23.1902
	B48	0.0857	0.0947	10.6760	11.8313
mean		0.0426	0.0805	5.4097	9.0531

In the above discussions, only one battery is applied for testing. For the deep analysis, we test all the batteries by the leave-one-out strategy. That means, among the four batteries of every case, each battery is tested while the other three batteries are used as the training set. The SOH prediction results for all the batteries are tabulated in Table 8. By this table, we observe that the proposed 3-MGPR method achieves the smaller RMSE and MAPE than ZGPR for the batteries B5, B6, B18, B46, B47 and B48. For the battery B7 and B45, the 3-MGPR has a little higher RMSE than ZGPR. The mean RMSE of 3-MGPR is 0.0426, which is smaller than the ZGPR's mean RMSE 0.0805, while the mean MAPE of 3-MGPR is 5.4097%, clearly smaller than the ZGPR's MAPE 9.0531%. To sum up, the application by the leave-one-out strategy demonstrate the proposed multiple GPR method can give better prediction than the basic ZGPR on the whole.

IV. CONCLUSION

This paper designs one method for the SOH prediction of lithium-ion batteries based on multiple Gaussian process regression model. In this proposed method, the mutual information analysis is used to select the key HIs, and the multiple GPR models from the existing batteries are combined by applying the weighting strategy based on the prediction uncertainty. The proposed model is validated by several sets of lithium-ion battery data involving the same conditions and the different conditions. Based on the applications results, some conclusions can be drawn. Firstly, compared to the basic GPR method, the MGPR method can improve the SOH prediction performance by applying the multiple GPR modeling strategy. Secondly, considering more training datasets from the same conditions benefits the SOH prediction for the MGPR method. Thirdly, even if the training datasets include some batteries from the different conditions, the MGPR method still does well.

REFERENCES

[1] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, "Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods," *J. Power Sources*, vol. 239, pp. 680–688, Oct. 2013.

[2] S. M. Rezvanizaniani, Z. Liu, Y. Chen, and J. Lee, "Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility," *J. Power Sources*, vol. 256, pp. 110–124, Jun. 2014.

[3] M. S. H. Lipu, M. A. Hannan, A. Hussai, M. M. Hoque, J. J. Ker, M. H. M. Saad, and A. Ayob, "A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: Challenges and recommendations," *J. Cleaner Prod.*, vol. 205, pp. 115–133, Dec. 2018.

[4] J. Wei, G. Dong, and Z. Chen, "Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5634–5643, Jul. 2018.

[5] Y. Song, D. Liu, C. Yang, and Y. Peng, "Data-driven hybrid remaining useful life estimation approach for spacecraft lithium-ion battery," *Microelectron. Rel.*, vol. 75, pp. 142–153, Aug. 2017.

[6] H. Zhang, Q. Miao, X. Zhang, and Z. Liu, "An improved unscented particle filter approach for lithium-ion battery remaining useful life prediction," *Microelectron. Rel.*, vol. 81, pp. 288–298, Feb. 2018.

[7] C. Wang, N. Lu, S. Wang, Y. Cheng, and B. Jiang, "Dynamic long short-term memory neural-network-based indirect remaining-useful-life prognosis for satellite lithium-ion battery," *Appl. Sci.*, vol. 8, no. 11, pp. 2078–2089, Oct. 2018.

[8] D. Liu, J. Zhou, H. Liao, Y. Peng, and X. Peng, "A health indicator extraction and optimization framework for lithium-ion battery degradation modeling and prognostics," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 45, no. 6, pp. 915–928, Jun. 2015.

[9] J. Kim and B.-H. Cho, "State-of-charge estimation and state-of-health prediction of a li-ion degraded battery based on an EKF combined with a per-unit system," *IEEE Trans. Veh. Technol.*, vol. 60, no. 9, pp. 4249–4260, Nov. 2011.

[10] B. Saha, K. Goebel, S. Poll, and J. Christophersen, "Prognostics methods for battery health monitoring using a Bayesian framework," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 2, pp. 291–296, Feb. 2009.

[11] J. Yu, B. Mo, D. Tang, H. Liu, and J. Wan, "Remaining useful life prediction for lithium-ion batteries using a quantum particle swarm optimization-based particle filter," *Qual. Eng.*, vol. 29, no. 3, pp. 536–546, 2017.

[12] A. Guha and A. Patra, "State of health estimation of lithium-ion batteries using capacity fade and internal resistance growth models," *IEEE Trans. Transport. Electrific.*, vol. 4, no. 1, pp. 135–146, Mar. 2018.

[13] D. Wang, F. Yang, Y. Zhao, and K.-L. Tsui, "Battery remaining useful life prediction at different discharge rates," *Microelectron. Rel.*, vol. 78, pp. 212–219, Nov. 2017.

[14] D. Wang, Y. Zhao, F. F. Yang, and K.-L. Tsui, "Nonlinear-drifted Brownian motion with multiple hidden states for remaining useful life prediction of rechargeable batteries," *Mech. Syst. Signal Process.*, vol. 93, no. 1, pp. 531–544, Sep. 2017.

[15] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Deep principal component analysis based on layerwise feature extraction and its application to nonlinear process monitoring," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 6, pp. 2526–2540, Nov. 2019.

[16] P. Y. Guo, Z. Cheng, and L. Yang, "A data-driven remaining capacity estimation approach for lithium-ion batteries based on charging health feature extraction," *J. Power Sources*, vol. 412, pp. 442–450, Feb. 2019.

[17] Y. Li, S. Zhong, Q. Zhong, and K. Shi, "Lithium-ion battery state of health monitoring based on ensemble learning," *IEEE Access*, vol. 7, pp. 8754–8762, Jan. 2019.

[18] D. Yang, Y. Wang, R. Pan, R. Chen, and Z. Chen, "A neural network based state-of-health estimation of lithium-ion battery in electric vehicles," *Energy Procedia*, vol. 105, pp. 2064–20159, May 2017.

[19] Z. Ge, "Process data analytics via probabilistic latent variable models: A tutorial review," *Ind. Eng. Chem. Res.*, vol. 57, pp. 12646–12661, Aug. 2018.

[20] E. Schulz, M. Speekenbrik, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *J. Math. Psychol.*, vol. 85, pp. 1–16, Aug. 2018.

[21] D. Liu, J. Pang, J. Zhou, Y. Peng, and M. Pecht, "Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression," *Microelectron. Rel.*, vol. 53, no. 6, pp. 823–839, Jun. 2013.
[22] F. Li and J. Xu, "A new prognostics method for state of health estimation of lithium-ion batteries based on a mixture of Gaussian process models and particle filter," *Microelectron. Rel.*, vol. 55, no. 7, pp. 1035–1045, Jun. 2015.
[23] Z. Wang, J. Ma, and L. Zhang, "State-of-health estimation for lithium-ion batteries based on the multi-Island genetic algorithm and the Gaussian process regression," *IEEE Access*, vol. 5, pp. 21286–21295, 2017.
[24] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, "A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve," *J. Power Sources*, vol. 384, pp. 387–395, Apr. 2018.
[25] J. Yu, "State of health prediction of lithium-ion batteries: Multiscale logic regression and Gaussian process regression ensemble," *Rel. Eng. Syst. Saf.*, vol. 174, pp. 82–95, Jun. 2018.
[26] Y.-J. He, J.-N. Shen, J.-F. Shen, and Z.-F. Ma, "State of health estimation of lithium-ion batteries: A multiscale Gaussian process regression modeling approach," *AIChE J.*, vol. 61, no. 5, pp. 1589–1600, Mar. 2015.
[27] L. Li, A. A. F. Saldivar, Y. Bai, and Y. Li, "Battery remaining useful life prediction with inheritance particle filtering," *Energies*, vol. 12, no. 14, p. 2784, Jul. 2019.
[28] B. Saha, K. Goebel, and J. Christophersen, "Comparison of prognostic algorithms for estimating remaining useful life of batteries," *Trans. Inst. Meas. Control*, vol. 31, nos. 3–4, pp. 293–308, 2009.
[29] B. Saha and K. Goebel, "Battery data set," NASA Ames Prognostics Data Repository, NASA Ames, Moffett Field, CA, USA, Tech. Rep., 2007. [Online]. Available: https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/
[30] X. Deng and J. Deng, "Incipient fault detection for chemical processes using two-dimensional weighted SLKPCA," *Ind. Eng. Chem. Res.*, vol. 58, no. 6, pp. 2280–2295, Jan. 2019.
[31] W. Qian and W. Shu, "Mutual information criterion for feature selection from incomplete data," *Neurocomputing*, vol. 168, pp. 210–220, Nov. 2015.
[32] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2005.
[33] Z. Chen and B. Wang, "How priors of initial hyperparameters affect Gaussian process regression models," *Neurocomputing*, vol. 275, pp. 1702–1710, Jan. 2018.
[34] Z. Ge, "Distributed predictive modeling framework for prediction and diagnosis of key performance index in plant-wide processes," *J. Process Control*, vol. 65, pp. 107–117, May 2018.
[35] O. Onur and B. Yön, "Elimination of a measurement problem: A robust prediction model for missing eigenvector value to assess earthquake induced out-of-plane failure of infill wall," *Measurement*, vol. 144, pp. 88–104, Oct. 2019.

XUEYING ZHENG received the bachelor's degree in engineering from Liaoning Shihua University, in 2016. She is currently pursuing the master's degree with the College of Control Science and Engineering, China University of Petroleum, Qingdao. Her research interests include machine learning and industrial data mining.

XIAOGANG DENG received the B.Eng. and Ph.D. degrees from the China University of Petroleum, Dongying, China, in 2002 and 2008, respectively, where he is currently an Associate Professor with the College of Information and Control Engineering. From October 2015 to October 2016, he was a Visiting Scholar with the Department of Electronics and Computer Sciences, University of Southampton, Southampton, U.K. His research interests include industrial process modeling and simulation, data-driven fault detection and diagnosis, and control performance monitoring.