Data-Driven Battery State of Health Estimation Based on Random Partial Charging Data

Zhongwei Deng, *Member, IEEE*, Xiaosong Hu, *Senior Member, IEEE*, Penghua Li, Xianke Lin, *Member, IEEE*, and Xiaolei Bian

***Abstract*—The rapid development of battery technology has promoted the deployment of electric vehicles (EVs). To ensure the healthy and sustainable development of EVs, it is urgent to solve the problems of battery safety monitoring, residual value assessment, and predictive maintenance, which heavily depends on the accurate state-of-health (SOH) estimation of batteries. However, many published methods are unsuitable for actual vehicle conditions. To this end, a data-driven method based on the random partial charging process and sparse Gaussian process regression (GPR) is proposed in this article. First, the random capacity increment sequences ($\triangle Q$) at different voltage segments are extracted from the partial charging process. The average value and standard deviation of $\triangle Q$ are used as features to indicate battery health. Second, correlation analysis is conducted for three types of batteries, and high correlations between the features and battery SOH are verified at different temperatures and discharging current rates. Third, by using the proposed features as inputs, sparse GPR models are constructed to estimate the SOH. Compared with other data-driven methods, the sparse GPR has the highest estimation accuracy, and its average maximum absolute errors are only 2.88%, 2.52%, and 1.51% for three different types of batteries, respectively.**

***Index Terms*—Capacity increment, feature extraction, lithium-ion battery, random charging segment, sparse Gaussian process, state-of-health.**

I. Introduction

THE RAPID development of battery technology has accelerated transportation electrification, resulting in a large-scale deployment of electric vehicles (EVs). The Chinese and European governments have implemented a series of policies to promote the development of EVs and they plan to have 80 and 30 million EVs on the road by 2030 [1], [2], respectively. However, many technical problems relating to EVs are still unsolved, leading to frequent safety incidents and consumer anxiety. For reliable and efficient battery operation, a battery management system (BMS) is indispensable to online monitor battery states and control charge or discharge behaviors. Wang *et al.* [3], [4] systematically summarized the battery modeling and state estimation techniques for advanced BMS. To enhance the penetration of EVs, it is urgent to solve the problems, e.g., battery safety monitoring, residual value assessment, and predictive maintenance, which heavily depends on accurate evaluation of battery health status [5], [6]. There are two fundamental approaches addressing battery health evaluation problem. The first one is to estimate the current health status, corresponding to battery state-of-health (SOH) estimation problem, while the second one is to predict the remaining time or number of cycles, corresponding to battery remaining useful life (RUL) prognosis [7]. As more attention is paid to the current state of battery health in automotive applications, this article focuses on battery SOH estimation problem.

Battery SOH is originally introduced to quantify the degree of aging of a battery. Current studies in the literature show that battery aging is mainly caused by irreversible reactions inside the battery, which usually result in an increase in internal resistance and a decrease in capacity [8], [9]. Many methods have been proposed to calculate battery internal resistance, such as electrochemical impedance spectroscopy (EIS) [10], hybrid pulse power characterization (HPPC) [11], and parameter identification based on equivalent circuit models (ECMs) [12]. By comparison, the battery capacity cannot be directly measured due to the incomplete charging and discharging process in actual applications, and it is also significantly affected by temperature, current rate, aging, and historical aging path [13]. The above reality makes accurate estimation of battery capacity more challenging and attracts many researchers to devote themselves to it [14]–[16]. To address this challenge, this article investigates the capacity-based SOH estimation for batteries. In this scenario, SOH is defined as the ratio of actual capacity to rated capacity, so SOH is 100% when the battery is new. For vehicle applications, the battery is considered to reach its end-of-life (EOL) when its SOH reaches 80% [17]. After the EOL in the vehicle, the battery is expected to be used in second-life applications, which also requires evaluation of the SOH [18].

Manuscript received September 5, 2021; revised November 13, 2021; accepted December 9, 2021. Date of publication December 13, 2021; date of current version January 19, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 52102420, in part by China Postdoctoral Science Foundation under Grant 2021M693725, in part by the Chongqing Natural Science Foundation for Postdoctor under Grant cstc2020jcyjbsh0040, and in part by Distinguished Young Scholars under Grant cstc2021jcyj-jqX0001. Recommended for publication by Associate Editor W. Cao. *(Corresponding authors: Zhongwei Deng; Xiaosong Hu; Penghua Li.)*

Zhongwei Deng and Xiaosong Hu are with the College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China (e-mail: dengzhongw@cqu.edu.cn; xiaosonghu@ieee.org).

Penghua Li is with the College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: lipenghua88@163.com).

Xianke Lin is with the Department of Mechanical Engineering, Ontario Tech University, Oshawa, ON L1G 0C5, Canada (e-mail: xianke.lin@uoit.ca).

Xiaolei Bian is with the Department of Chemical Engineering, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden (e-mail: xiaoleib@kth.se).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TPEL.2021.3134701.

Digital Object Identifier 10.1109/TPEL.2021.3134701

Generally, there are three categories of methods for battery SOH estimation in the literature. The first one is the direct calculation method, which records the full charge and discharge process and calculates actual battery capacity according to its definition. Another improved method is to calculate the accumulated charge (Q) in a large state-of-charge (SOC) interval, and the actual capacity can be derived from $Q/\triangle$SOC [19]. However, this method requires high-precision SOC estimation with voltage correction. Besides, through various orthogonal experiments, a semiempirical model can be established to directly calculate the capacity degradation under different temperatures, current rates and depth of discharge (DOD) [20]. The second one is the model-based method, which mainly depends on ECMs [21] or electrochemical models [22]. In an ECM, battery capacity can be regarded as a parameter or state of the model. Therefore, regressive least-squares methods [23], Kalman filters [24], [25], particle filters [26], moving horizon estimators [27], and their variants can be used to estimate it in real time. In a pseudo-two-dimensional (P2D) model, which has been regarded as the first principle model of lithium-ion batteries [28], battery capacity can be derived based on relevant electrochemical parameters [29]. However, complicated operating conditions and coupling degradation mechanisms make it difficult to establish an accurate battery model throughout the battery life cycle. The third one is the data-driven method, which regards the SOH estimation as a typical regression problem. Many advanced machine learning methods have been used to map the nonlinear relationship between the battery SOH and input data, e.g., the support vector machine (SVM) [12], relevance vector machine (RVM) [30], long short-term memory network (LSTM) [31], convolutional neural network (CNN) [32], and Gaussian process regression (GPR) [16], [33]. To capture the cross-correlate capacity degradation trends between different battery cells, Chehade *et al.* [34], [35] proposed a collaborative GPR and a multioutput convolved GPR, and both achieve accurate battery capacity prediction based on the number of cycles.

In order to construct accurate data-driven models for battery SOH estimation, it is often necessary to extract features that are closely correlated with battery capacity in advance. Incremental capacity (IC) and differential voltage (DV) analysis can not only identify the aging mechanisms of batteries [8], but also provide features for SOH estimation [16], [36]. The major drawback of this method is that it requires the battery operating at a low current rate. In our previous study [37], a general method is proposed to extract features from different discharging conditions, including constant current (CC), pulse current, and dynamic current conditions. In a vehicle application, the most commonly used charging protocol is the constant current and constant voltage (CC–CV) charging protocol, which is relatively stable during the whole life cycle of battery pack. Taking the advantage of this, various features have been extracted from charging process to represent battery health, typically including CC or CV charging time [33], voltage curve slope [38], time change during equal voltage interval [39], and voltage change during equal time interval [40]. However, a complete or specific charging process is usually required to extract these features. Since the battery is usually recharged before it is depleted, the charging process usually starts at a random SOC and could end at any point until before being fully charged. Therefore, it is crucial to construct the battery SOH estimation model based on a random partial charging process.

To bridge this research gap, a battery SOH estimation method based on the random partial charging process and sparse GPR is proposed in this article, and its schematic diagram is shown in Fig. 1. First, random capacity increment sequences at different voltage segments are derived from partial charging processes, and their statistical characteristics are regarded as features to indicate battery health. Then, correlation analysis is conducted for three different types of batteries to verify the effectiveness of the extracted features. The sparse GPR with better computational efficiency and accuracy is employed to establish the SOH estimation model. Finally, datasets of three different types of batteries under different temperatures and discharging current rates are used to validate the proposed method. Note that in the training process, all random charging segments are used as the inputs of the sparse GPR model; while in the online application, the battery SOH can be estimated by any random charging segment. The main contributions of this article include the following.

1) Battery SOH estimation based on random partial charging data. Any CC charging segment can be used to estimate SOH in the online application. Unlike the traditional methods where different models are built for different segments, our proposed method requires only one model. To the best of our knowledge, this is the first study to achieve battery health estimation based on a random charging process.
2) An SOH estimation model with higher accuracy and lower computational cost is constructed. Because random charging segments are used to extract features, there are thousands of training samples for each battery cell, resulting in a long training time for regular GPR and possible overfitting. The sparse GPR with inducing points optimization can effectively solve the above problems.
3) The proposed method is verified by three different types of batteries at different temperatures and discharging current rates, and the results are compared with other typical machine learning methods, i.e., the linear multiple linear regression (MLR), SVM, RVM, and CNN.

The rest of this article is organized as follows. Section II introduces the battery datasets; the feature extraction is described in Section III; the principle of regular and sparse GPR is illustrated in Section IV; the results of the SOH estimation are presented in Section V. Finally, Section VI concludes this article.

II. Battery Datasets

Datasets used in this article include battery cycle aging data for three different types of batteries, namely, lithium nickel cobalt aluminum oxide (NCA), lithium nickel manganese cobalt oxide (NMC), and lithium iron phosphate (LFP). They are all obtained from Sandia National Lab [41], which include data produced at different temperatures and discharging conditions. The battery information and experimental setting are listed in Table I, and more details can be found in the website [42].

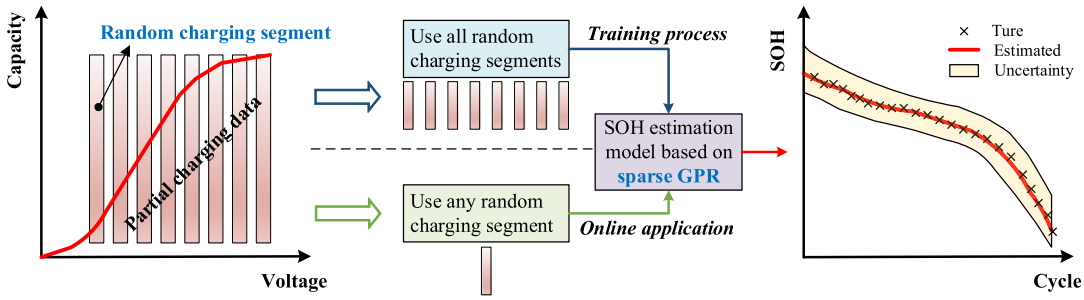

Fig. 1. Schematic diagram of battery SOH estimation based on random partial charging data.

TABLE I
BATTERY DATASETS FOR HEALTH ESTIMATION

Battery types	Manufacturers	Nominal capacity (Ah)	Temperature (°C)	Charge and discharge policies
NCA	Panasonic	3.2	15/25/35	0.5C CC-CV & 0.5C/1C/2C CC
NMC	LG Chem	3	15/25/35	0.5C CC-CV & 0.5C/1C/2C/3C CC
LFP	A123	1.1	15/25/35	0.5C CC-CV & 0.5C/1C/2C/3C CC

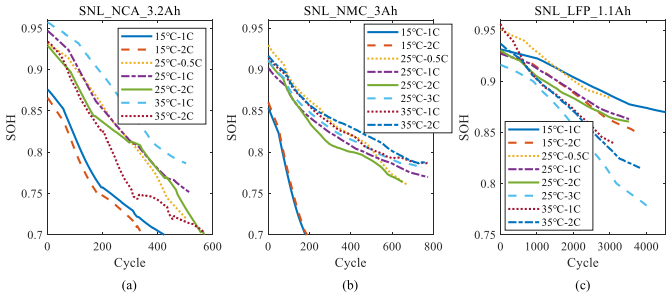

Fig. 2. Battery SOH versus cycle number for three types of batteries. (a) NCA cells. (b) NMC cells. (c) LFP cells.

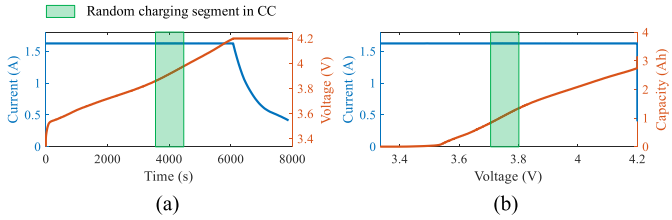

Fig. 3. Battery typical charging process. (a) Current and voltage curves with respect to time. (b) Current and charge capacity curves with respect to voltage.

The NCA battery is from Panasonic with a nominal capacity of 3.2Ah; the NMC comes from LG Chem with a nominal capacity of 3Ah; the LFP is from the A123 system and has a nominal capacity of 1.1Ah. All battery cells are cycled under the same charging profile (0.5C CC–CV), but different CC discharging profiles with different current rates (0.5C, 1C, 2C, and 3C), different DODs (100%, 60%, and 20%) and different temperatures (15, 25, and 35 °C). To evaluate battery capacity degradation, only 100% DOD data are used in this article.

For these three types of batteries, the SOH declining curves of representative cells are plotted in Fig. 2. It can be observed that battery SOH degradation shows different trends for different types of batteries and is significantly affected by temperature and current rate. For LFP, the discharge current rate has more impact on battery degradation than temperature, while for NCA and NMC, the discharge current rate has less impact on battery degradation than temperature. Due to the complicated internal aging mechanism of different batteries, it is difficult to propose a general physics-based method to achieve an accurate estimation of battery SOH. In contrast, data-driven methods could obtain high-precision SOH estimation for different batteries by extracting effective features as inputs and applying model training to optimize parameters.

III. FEATURES EXTRACTED FROM THE RANDOM CHARGING PROCESS

In the vehicle application, the battery discharging process is heavily dynamic, while the charging process is relatively stable, and a constant charging strategy is usually carried out during the entire life cycle of the battery pack. Therefore, battery charging data is preferred for SOH estimation. However, since the battery is usually recharged before it is depleted, the charging process can start from any random SOC and end at any point until full SOC. Accordingly, it is vital to extract features from the random partial charging process.

A. Feature Extraction

For typical charging processes, such as CC–CV charging policy, the battery current and voltage profiles are illustrated in Fig. 3(a). In the vehicle application, the available battery data are sampling time (t), battery current (I), battery voltage (V), and battery temperature (T). Because not every battery has a temperature sensor, the temperature is usually not an ideal input. Battery charge capacity (Q) can be obtained by integrating

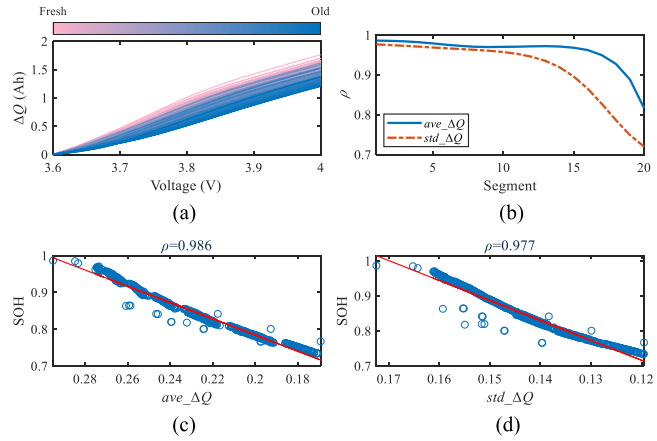

Fig. 4. Battery features and their correlations to battery SOH. (a) Evolution of $\triangle Q$ over aging cycles. (b) Correlation results of ave_$\triangle Q$ and std_$\triangle Q$ for different segments. (c) Correlation of ave_$\triangle Q$ for the first segment. (d) Correlation of std_$\triangle Q$ for the first segment.

current over time. For data-driven SOH estimation, the output is battery discharge capacity, while the inputs can be the direct measurement data, usually including I, V, and Q [32], [43], or features extracted from the raw data [37], [44]. Since the charging process in actual applications is random and partial, to estimate battery SOH based on any random segment, the training inputs of the model need to contain all the random segments. Therefore, for each cycle, there will be several training samples that have the same capacity but different inputs.

In this article, the charge capacity curve with respect to battery voltage, as shown in Fig. 3(b), is used to extract features. It can be observed that only charging data at the CC process is effective in this case. To extract features, a reasonable voltage range needs to be determined. Given starting voltage (V_{start}), end voltage (V_{end}), and voltage interval ($\triangle V$), the data of ith cycle can be divided into L_i intervals

$$L_i = \frac{V_{\text{end}} - V_{\text{start}}}{\Delta V}. \tag{1}$$

By setting a fixed length (n) for random charging segment, the number of segments of each cycle can be obtained as follows:

$$M_i = \text{floor}\left(\frac{L_i - n}{c}\right) + 1 \tag{2}$$

where c is the stride size, and the function floor(.) gets the largest integer no more than the input value. An obvious limitation of the random charging segment is that the actual charge capacity is unknown. To address this issue, the capacity sequence, $Q = [Q_1, Q_2, \ldots, Q_n]$, is replaced by the capacity increment sequence, $\triangle Q = Q - Q_1$.

The evolution of $\triangle Q$ over aging cycles is illustrated in Fig. 4(a), which has a corresponding voltage range of [3.6 V, 4.0 V]. It can be observed that the $\triangle Q$ curve moves downwards as the degree of aging increases, which indicates that there is an implicit relationship between $\triangle Q$ and battery health. Therefore, $\triangle Q$ can be used to estimate battery SOH, and several previous studies have used $\triangle Q$ as the input of the data-driven model directly [32], [39], [45]. To further improve the estimation accuracy and reduce uncertainty, the statistical characteristics of $\triangle Q$ are used as features for SOH estimation in this article.

TABLE II
PARAMETERS FOR FEATURES EXTRACTION

Battery types	V_{start} (V)	V_{end} (V)	$\triangle V$ (V)	n	c
NCA/NMC	3.6	4.19	0.01	40	1
LFP	3.0	3.59	0.01	40	1

The parameters setting for features extraction is listed in Table II. The selected voltage ranges cover almost the entire CC charging processes of different batteries. The V_{start} is set to a voltage point corresponding to about 5% SOC of battery, because 5% battery SOC is usually reserved to avoid battery over-discharge in practical applications, while, the V_{end} is set to a value slightly deviating from the charging cut-off voltage to avoid the influence of data fluctuations when switching to CV charging. According to the setting, each cycle can produce 20 charging segments. Due to the same voltage platform, the parameters settings are the same for NCA and NMC cells. In contrast, LFP cells have a lower voltage platform, so the voltage range for features extraction is changed to [3.0 V, 3.59 V], while other parameters remain unchanged. The effect of segment length (n) on battery SOH estimation will be analyzed in the following section.

B. Correlation Analysis

The correlations between the battery SOH and the statistical characteristics of $\triangle Q$, including average value (ave_$\triangle Q$) and standard deviation (std_$\triangle Q$), are analyzed in this section. Generally, the higher the correlation between the input features and the output of a data-driven model, the better the accuracy of the model. Pearson correlation coefficient (ρ) has been widely used to quantify the degree of linear correlation between two variables, and its formula is as follows [46]:

$$\rho_{x_j} = \frac{\sum(\boldsymbol{x}_j - \bar{x}_j)(\boldsymbol{y} - \bar{y})}{\sqrt{\sum(\boldsymbol{x}_j - \bar{x}_j)^2 \sum(\boldsymbol{y} - \bar{y})^2}} \tag{3}$$

where $\boldsymbol{x}_j$ is the feature sequence, $\boldsymbol{y}$ is the battery SOH sequence, and $\bar{x}_j$ and $\bar{y}$ are their average values.

First, the correlation analysis is conducted on an NCA cell, which is cycled under 25°C and 1C discharging conditions. For a given charging segment, which is determined by battery voltage, the corresponding $\triangle Q$ at different aging cycles can be extracted, then the correlations between battery SOH and the statistical characteristics of $\triangle Q$ can be calculated. The correlation results of ave_$\triangle Q$ and std_$\triangle Q$ for different segments are shown in Fig. 4(b), and fitting curves between battery SOH and the features of the first segment are presented in Fig. 4(c) and (d). The voltage sequence of the first segment is [3.6 V: 0.01 V: 4.0 V], and the larger the segment number, the higher voltage it corresponds. It can be observed that both ave_$\triangle Q$ and std_$\triangle Q$ have high correlations to battery SOH for almost all segments. For the segment with a number over 10, as the increase of the number, which means the voltage goes up, the

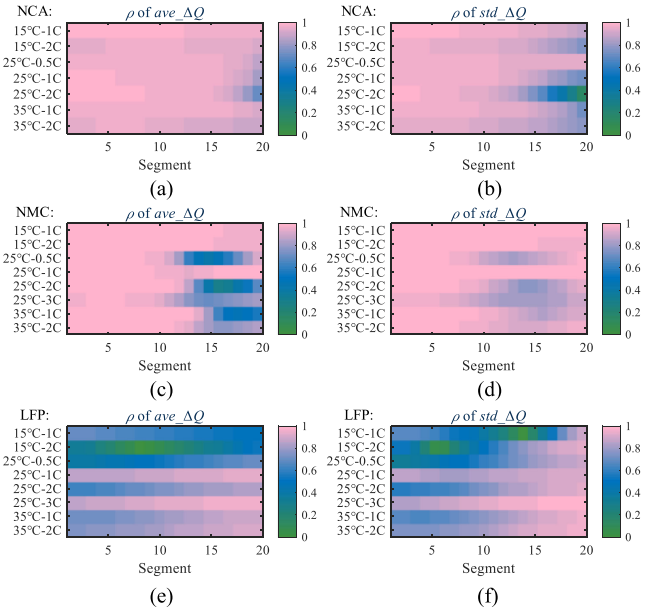

Fig. 5. Correlation analysis results of two features for three different types of batteries. (a) ave_$\triangle Q$ for NCA cells. (b) std_$\triangle Q$ for NCA cells. (c) ave_$\triangle Q$ for NMC cells. (d) std_$\triangle Q$ for NMC cells. (e) ave_$\triangle Q$ for LFP cells. (f) std_$\triangle Q$ for LFP cells.

correlation of ave_$\triangle Q$ decreases slowly while the correlation of *std*_$\triangle Q$ decreases relatively quickly.

Then, the correlation analysis is extended to all battery cells under different temperatures and discharging current rates for the three types of batteries. The correlation results of the two features for different batteries are shown in Fig. 5, in which the correlations under different conditions are compared for each type of battery. As illustrated in Fig. 5, most of the correlations of the two features are greater than 0.9 for NCA and NMC cells, and 0.8 for LFP cells. All features for three types of batteries under 25°C and 1C discharging conditions have a very high correlation to battery SOH. Temperature and current rate have an obvious influence on the correlation, but no specific rules can be observed. For NCA and NMC cells, high correlations can be obtained at low temperatures. In contrast, the low temperature has a negative effect on the correlations of LFP cells, and a higher current rate exacerbates the effect.

IV. Battery Health Estimation Based on Gaussian Process Regression

The GPR has been successfully applied in many regression applications due to its advantages in nonparametric modeling and probabilistic prediction. Accurate battery SOH estimation can be obtained by using effective features as inputs [37]. However, with the increase of data scale, the computational complexity of regular GPR increases dramatically, which hinders its online use in vehicles. To overcome this problem, this article proposes an SOH estimation method based on sparse GPR.

A. Regular Gaussian Process Regression

The Gaussian process (GP) can be defined as a collection of finite random variables which follow a Gaussian distribution [47]. Formally, for any input set $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]$, if its probability distribution over function $\boldsymbol{F}(\boldsymbol{x}) = [f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \ldots, f(\boldsymbol{x}_n)]$ follows a joint Gaussian distribution, $\boldsymbol{F}(\boldsymbol{x})$ is a GP, and can be described as $f(\boldsymbol{x}) \sim GP(m(\boldsymbol{x}), k(\boldsymbol{x}_i, \boldsymbol{x}_j))$, where the mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ are defined as follows:

$$m(\boldsymbol{x}) = E(f(\boldsymbol{x})) \tag{4}$$

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = E\left[\left(f(\boldsymbol{x}_i) - m(\boldsymbol{x}_i)\right)\left(f(\boldsymbol{x}_j) - m(\boldsymbol{x}_j)\right)\right]. \tag{5}$$

$m(\boldsymbol{x})$ is usually set to zero when the prior information is unknown. $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is also called the kernel function, which has a large effect on the estimation accuracy of GP. For a specific application, a combined kernel function can be constructed based on prior knowledge to achieve better performance [48]. Among various kernel functions [49], a squared exponential kernel function is the most widely used, and is expressed as follows:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma_f^2 \exp\left(\frac{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2l^2}\right) \tag{6}$$

where σ_f and l are hyperparameters, which determine the amplitude of the kernel function and the importance of each input feature, respectively.

For battery SOH estimation, which is a typical regression problem, observations usually contain Gaussian white noises, and can be modeled as follows:

$$y = f(\boldsymbol{x}) + N\left(0, \sigma_n^2\right) \tag{7}$$

where y is the observation, $\boldsymbol{x}$ is the input features, and $f(\boldsymbol{x})$ is the implicit function between the features and battery SOH, and is approximated by GP. By assuming $m(\boldsymbol{x})$ of GP equal to zero, the prior distribution of the observation can be described as follows:

$$\boldsymbol{y} \sim N\left(0, \boldsymbol{K}_{x,x} + \sigma_n^2\right) \tag{8}$$

where $\boldsymbol{K}_{x,\,x}$ is an n-dimensional covariance matrix made up of $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. To improve the estimation accuracy, hyperparameter set $\Theta = [\sigma_f, l, \sigma_n]$ needs to be optimized according to the maximum likelihood method [47], which can be expressed as follows:

$$\begin{aligned}\log p(\boldsymbol{y}|\boldsymbol{X}, \Theta) = & -\frac{1}{2}\log\left(\det\left(\boldsymbol{K}_{x,x} + \sigma_n^2 \boldsymbol{I}_n\right)\right) \\ & -\frac{1}{2}\boldsymbol{y}^T\left[\boldsymbol{K}_{x,x} + \sigma_n^2 \boldsymbol{I}_n\right]^{-1}\boldsymbol{y} - \frac{n}{2}\log 2\pi\end{aligned} \tag{9}$$

where n is the number of training samples and $\boldsymbol{I}_n$ is a unit matrix with n dimensions. Then, a gradient-based algorithm can be used to obtain the optimal hyperparameters. For a newly arrived dataset $\boldsymbol{x}^*, f^*$ is the estimated output. The joint prior distribution of $\boldsymbol{f}$ and f^* can be written as follows:

$$p(\boldsymbol{f}, f^*) \sim N\left(\begin{bmatrix}\boldsymbol{0} \\ 0\end{bmatrix}, \begin{bmatrix}\boldsymbol{K}_{x,x} & \boldsymbol{K}_{x,x^*} \\ \boldsymbol{K}_{x^*,x} & \boldsymbol{K}_{x^*,x^*}\end{bmatrix}\right). \tag{10}$$

According to Bayes' theorem, the estimated posteriori distribution can be derived as follows:

$$p\left(f^*|\boldsymbol{X},\boldsymbol{y},\boldsymbol{x}^*\right)=N\left(\mu^*,\Sigma^*\right) \tag{11}$$

where

$$\begin{aligned}\mu^*&=\boldsymbol{K}_{x^*,x}\left[\boldsymbol{K}_{x,x}+\sigma_n^2\boldsymbol{I}_n\right]^{-1}\boldsymbol{y}\\ \Sigma^*&=\boldsymbol{K}_{x^*,x^*}-\boldsymbol{K}_{x^*,x}\left[\boldsymbol{K}_{x,x}+\sigma_n^2\boldsymbol{I}_n\right]^{-1}\boldsymbol{K}_{x,x^*}.\end{aligned} \tag{12}$$

The mean value μ^* gives the best estimation of f^*, the estimation uncertainty is quantified by its variance value Σ^*, and a 95% confidence interval can be calculated by $[\mu^*-1.96\sqrt{\Sigma^*},\mu^*+1.96\sqrt{\Sigma^*}]$.

B. Sparse Gaussian Process Regression

For battery SOH estimation based on random charging segments, each charging process can generate dozens of training samples (usually more than 10), thus thousands of training samples can be produced for the entire aging process of a battery. Due to the inversion of the matrix in (9), the regular GPR takes $O(n^3)$ time to optimize the hyperparameters for the training set with n samples. According to (12), once the inversion is known, the prediction time of regular GPR is $O(n^2)$ for each test sample.

To reduce the computing burden, a general method is to modify the joint prior distribution $p(\boldsymbol{f},f^*)$by introducing inducing variables $\boldsymbol{u}=[u_1,u_1,\ldots,u_m]$ [50]. The corresponding input set of $\boldsymbol{u}$ is called inducing points $\boldsymbol{X}_u$, which can be chosen from the original training set. Given the inducing points, $p(\boldsymbol{f},f^*)$can be rewritten as follows:

$$p\left(\boldsymbol{f},f^*\right)=\int p\left(\boldsymbol{f},f^*\,|\boldsymbol{u}\right)p\left(\boldsymbol{u}\right)d\boldsymbol{u} \tag{13}$$

where $p(\boldsymbol{u})=N(\boldsymbol{0},\boldsymbol{K}_{u,u})$. Furthermore, given $\boldsymbol{u}$, it can be assumed that $\boldsymbol{f}$ and f^* are conditionally independent, thus (13) can be modified as follows:

$$p\left(\boldsymbol{f},f^*\right)\simeq q\left(\boldsymbol{f},f^*\right)=\int q\left(\boldsymbol{f}\,|\boldsymbol{u}\right)q\left(f^*\,|\boldsymbol{u}\right)d\boldsymbol{u}. \tag{14}$$

Different methods have been proposed to approximate inducing conditionals $q(\boldsymbol{f}|\boldsymbol{u})$ and $q(f^*|\boldsymbol{u})$, typically, including subset of regression (SOR) approximation, deterministic training conditional (DTC) approximation, partially independent training conditional (PITC) approximation, and fully independent training conditional (FITC) approximation [50]. In this article, the FITC approximation is used to obtain a better result. Due to the fully independence assumption, training conditional $q(\boldsymbol{f}|\boldsymbol{u})$ can be calculated by

$$\begin{aligned}q\left(\boldsymbol{f}\,|\boldsymbol{u}\right)&=\prod_{i=1}^{n}p\left(f_i\,|\boldsymbol{u}\right)\\&=N\left(\boldsymbol{K}_{x,u}\boldsymbol{K}_{u,u}^{-1},\operatorname{diag}\left[\boldsymbol{K}_{x,x}-\boldsymbol{Q}_{x,x}\right]\right)\end{aligned} \tag{15}$$

where $\boldsymbol{Q}_{a,b}=\boldsymbol{K}_{a,u}\boldsymbol{K}_{u,u}^{-1}\boldsymbol{K}_{u,b}$, and the test conditional keeps exact as follows:

$$q\left(f^*|\boldsymbol{u}\right)=p\left(f^*|\boldsymbol{u}\right). \tag{16}$$

By inserting the inducing conditionals into (14), and integrating over $\boldsymbol{u}$, the joint prior distribution can be derived as follows:

$$\begin{aligned}&q\left(\boldsymbol{f},f^*\right)\\&\sim N\left(\begin{bmatrix}\boldsymbol{0}\\0\end{bmatrix},\begin{bmatrix}\boldsymbol{Q}_{x,x}-\operatorname{diag}\left[\boldsymbol{Q}_{x,x}-\boldsymbol{K}_{x,x}\right] & \boldsymbol{Q}_{x,x^*}\\ \boldsymbol{Q}_{x^*,x} & \boldsymbol{K}_{x^*,x^*}\end{bmatrix}\right).\end{aligned} \tag{17}$$

According to Bayes' theorem, the estimated posteriori distribution can be obtained as follows:

$$q\left(f^*|\boldsymbol{X},\boldsymbol{y},\boldsymbol{x}^*\right)=N\left(\tilde{\mu}^*,\tilde{\Sigma}^*\right) \tag{18}$$

where

$$\begin{aligned}\tilde{\mu}^*&=\boldsymbol{K}_{x^*,u}\Omega\boldsymbol{K}_{u,x}\Lambda^{-1}\boldsymbol{y}\\ \tilde{\Sigma}^*&=\sigma_n^2+\boldsymbol{K}_{x^*,x^*}-\boldsymbol{Q}_{x^*,x^*}+\boldsymbol{K}_{x^*,u}\Omega\boldsymbol{K}_{u,x^*}\end{aligned} \tag{19}$$

where

$$\begin{aligned}\Omega&=\left(\boldsymbol{K}_{u,u}+\boldsymbol{K}_{u,x}\Lambda^{-1}\boldsymbol{K}_{x,u}\right)^{-1}\\ \Lambda&=\operatorname{diag}\left[\boldsymbol{K}_{x,x}-\boldsymbol{Q}_{x,x}+\sigma_n^2\boldsymbol{I}_n\right].\end{aligned}$$

Again, the maximum likelihood method can be used to optimize the hyperparameters for this sparse GPR [50]. The computing burden of the training process becomes $O(m^2n)$, which is linear with n, while the prediction time is only $O(m^2)$ for each test sample [51]. Compared with the regular one, the sparse GPR can effectively reduce the computing burden. Generally, a larger m leads to better accuracy for sparse GPR, but it also sacrifices computational efficiency at the same time.

V. Results and Discussions

In this section, the battery SOH estimation results based on the regular and sparse GPR models are presented, and their performance is compared with other advanced machine learning methods, including MLR, SVM, RVM, and CNN. It should be noted that in addition to the two features used as input, the mean voltage of the segment is also used as input to distinguish different charging segments. To evaluate the estimation accuracy, mean absolute error (MAE) and root-mean-square error (RMSE) are chosen to quantify the estimation error, and their formulas are given by

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{20}$$

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

where y_i is the real value and n is the number of samples. MATLAB is used to conduct the methods in a personal computer (Intel Core i7-8565U CPU 1.8 GHz and 7.85-GB RAM available). To verify the improvement in computational efficiency, the running times of the GPR models are recorded and compared. The battery cycling life is usually lower than one thousand cycles for NCA and NMC cells, and over three thousands for LFP cells. To avoid the malfunction of MATLAB due to out of memory, when

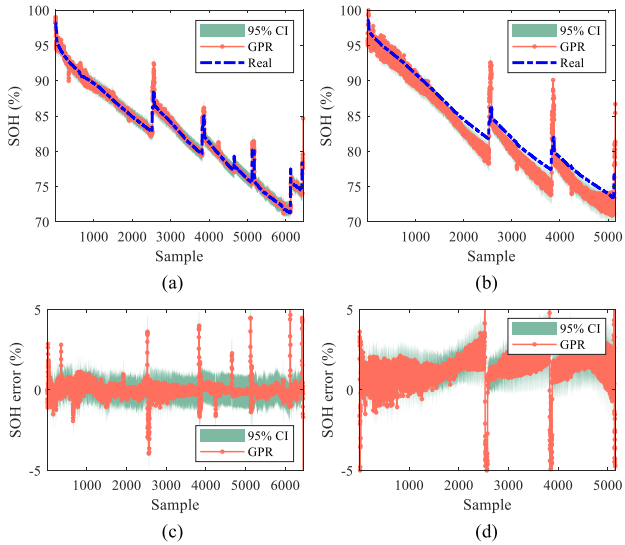

Fig. 6. Battery SOH estimation results based on the regular GPR for NCA cells. (a) Training results. (b) Test results. (c) Training errors. (d) Test errors.

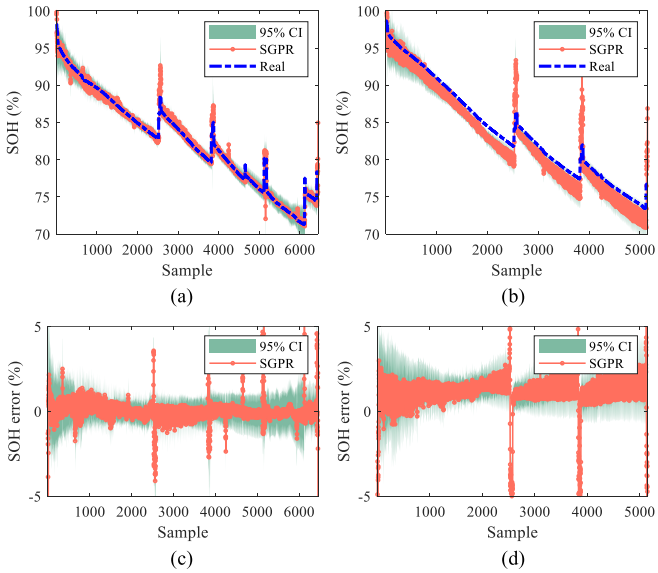

Fig. 7. Battery SOH estimation results based on the sparse GPR for NCA cells. (a) Training results. (b) Test results. (c) Training errors. (d) Test errors.

conducting GPR-based SOH estimation, the stride size c is set to 2 for NCA and NMC cells, and 4 for LFP cells.

A. *Estimation Results for the Same Condition*

First, the GPR models are trained on a cell under 25°C and 1C discharging conditions, and are tested on another cell with the same battery type and cycling condition. The SOH estimation results of NCA cell based on the regular and sparse GPR models are shown in Figs. 6 and 7, respectively, in which the 95% confidence interval (CI) is also provided and the sparse GPR is denoted by SGPR. The corresponding statistical errors and computational times are listed in Table III. For the sparse GPR, the number of inducing points is set to 500. The indexes of the inducing points are regarded as hyperparameters and are optimized together with the parameters of the kernel function during the training process.

TABLE III
STATISTICAL ERRORS AND COMPUTATIONAL TIMES OF THE BATTERY SOH ESTIMATION

Criterion	Regular GPR		Sparse GPR	
	Training	Test	Training	Test
MAE	0.26%	1.61%	0.29%	1.41%
RMSE	0.51%	1.83%	0.64%	1.62%
Samples	6450	5150	500[1]	5150
Time (s)	1182s	76.79s	9.44s	0.36s

[1] Number of inducing points.

It can be observed that by using the proposed features as inputs, the GPR models can be fully trained with small errors. For the regular GPR, MAE, and RMSE are only 0.26% and 0.51%, respectively, while for the sparse GPR, the errors just increase slightly. Note that the sparse GPR only uses 500 points for training, which leads to a large computational time reduction, more than 93% of the running time of the regular GPR. In the test process, MAE and RMSE of the regular GPR increase to 1.61% and 1.83%, respectively, which are still very good for actual applications. Compared with the regular GPR, the errors of the sparse one are further reduced by 0.20% and 0.21%, respectively, and its time is reduced by more than 96%. The higher accuracy of the sparse GPR on test samples is due to the optimization of inducing points, which reduces the overfitting of the training process. The above results indicate that for large-scale data samples, the sparse GPR has advantages over the regular one in terms of accuracy and computational efficiency. Besides, in both training and test processes, larger errors all occur at the points where SOH shows a step change due to battery resting and capacity recovery phenomenon.

The SOH estimation results of NMC and LFP cells based on the regular and sparse GPR models are shown in Figs. S1–S4 in the supplementary. Similar conclusions can be drawn for NMC and LFP cells as for NCA cells, and the superiority of the sparse GPR has been verified again.

To evaluate the performance of GPR models, other commonly used machine learning methods, including MLR, SVM, RVM, and CNN are also used to estimate battery SOH.

The MLR is a basic method to model the linear relationship between input features and output target. The SVM can transform a low-dimensional nonlinear problem into a high-dimensional linear problem. Although the RVM has an identical functional form to SVM, it is derived based on Bayesian framework, and can provide uncertainty evaluation of prediction results. The CNN has been widely used in image recognition, and it does not require an extra feature extraction process, it can conduct feature extraction by itself. Note that all methods, except CNN, use the same features as input; while CNN directly uses the voltage and capacity increment sequences as inputs. CNN does not require an extra feature extraction process, and it can conduct feature extraction by itself. All the mentioned methods are trained on a cell under 25°C and 1C discharging conditions and are tested on another cell with the same battery

TABLE IV
SOURCES, STRUCTURES, AND INITIAL PARAMETERS OF THE USED METHODS

Methods	Sources	Model structures or initial parameters
MLR	By using the *regress* function in MATLAB	--
SVM [12]	By using the codes in https://www.csie.ntu.edu.tw/~cjlin/libsvm/	'-s 4 -t 2 -c 2.2 -g 2.8 -p 0.01'
RVM [30]	By using the codes in https://www.relevancevector.com	's',5e-1,'b',5e-1
CNN [32]	By using the *Deep learning* toolbox in MATLAB	The constructed network mainly includes two 2D convolutional layers, one maximum pooling layer and one fully-connected layer
GPR [33]	By using the codes in http://www.gaussianprocess.org/gpml/code/matlab/doc/	covfunc = {@covSEiso} hyp.cov =log([5 10]) likfunc = {@likGaus} hyp.lik = log(1e-1) meanfunc=[]

TABLE V
SOH ESTIMATION ERRORS OF NCA BATTERY FOR DIFFERENT METHODS

	Training		Test	
Method	MAE	RMSE	MAE	RMSE
MLR	1.92%	2.45%	2.12%	2.79%
SVM	1.26%	1.76%	1.63%	2.28%
RVM	1.53%	2.04%	1.86%	2.34%
CNN	1.34%	1.99%	1.70%	2.32%
Regular GPR	0.26%	0.51%	1.61%	1.83%
Sparse GPR	0.29%	0.64%	1.41%	1.62%

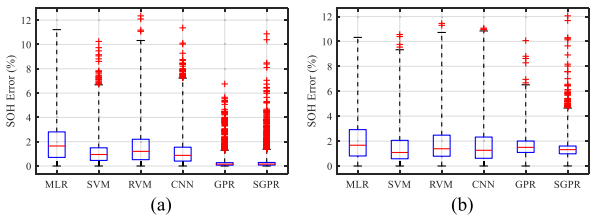

Fig. 8. Error distributions of SOH estimation for different methods. (a) Training process. (b) Test process. On each box, the central mark indicates the median, and the bottom and top edges of the box represent the 25th and 75th percentiles, respectively. Excluding the outliers which are drawn separately using the '+' symbol, the whiskers extend at both ends to the most extreme data points.

type and cycling condition. The sources of the used methods and the model structures or initial parameters are listed in Table IV.

For the NCA cell, the statistical errors of different methods are listed in Table V, and the corresponding error distributions are illustrated in Fig. 8. For all the methods, MAE and RMSE are lower than 2.2% and 2.8%, respectively, which indicates the effectiveness of the extracted features. The worst errors occur in MLR, because it cannot capture the nonlinearity between battery SOH and the features. The sparse GPR has the smallest errors in the test process, and its maximum errors are much lower than that of other methods with the removal of outliers.

For NMC and LFP cells, the statistical errors of different methods are listed in Tables S1 and S2 in the supplementary. Among all the methods, the best accuracy is achieved in the sparse GPR for these two types of batteries. It is worth noting that severe overfitting happens on the regular GPR in the SOH

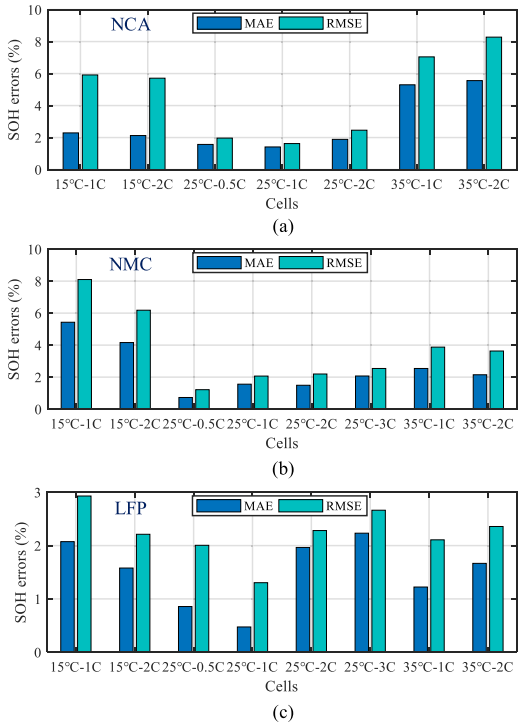

Fig. 9. Battery SOH estimation errors under different temperatures and discharging current rates. (a) NCA cells. (b) NMC cells. (c) LFP cells.

estimation of LFP cell, which leads to erroneous estimated values in the test process, while the sparse GPR can effectively avoid the above problem and maintains high accuracy.

B. Estimation Results for Different Conditions

For battery SOH estimation, it is valuable to construct a model that can be trained on a specific condition and maintain high estimation accuracy under different conditions. This section evaluates the generalization ability of the proposed GPR-based model for different temperatures and discharging current rates. For the three types of batteries, the GPR models are all trained at 25°C and 1C discharging conditions, and are tested on different conditions. The number of inducing points is also set to 500 for NMC cell, but 1000 for LFP cell due to its longer cycling life. The estimation errors of the sparse GPR models are plotted in Fig. 9. For all three types of batteries, the smallest errors occur at the same temperature (25°C) in the training set. An increase or decrease in temperature results in larger estimation errors. At 15°C, increasing the discharging current rate could reduce the estimation errors. The possible reason is that the higher rate leads to more heat production, making the internal temperature and degradation process of the battery cell closer to the training cell. In contrast, at 35°C, the higher rate makes the battery degradation more severe, further deviating from the degradation characteristics of the training cell, and finally resulting in larger errors. Overall, the sparse GPR model has an excellent SOH estimation accuracy for different batteries, and its average MAEs are 2.88%, 2.52%, and 1.51% for NCA, NMC, and LFP cells, respectively, which indicates its good generalization to different cycling conditions.

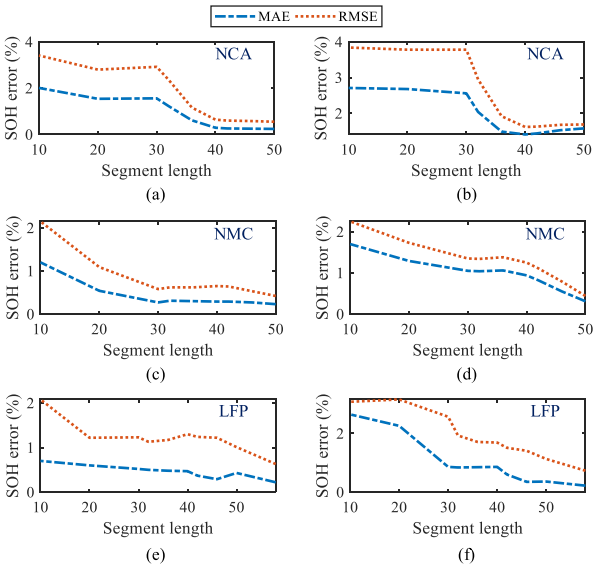

Fig. 10. Influence of segment length on battery SOH estimation. (a), (c), (e) Training process of NCA cell, NMC cell, and LFP cell, respectively. (b), (d), (f) Test process of NCA cell, NMC cell, and LFP cell, respectively.

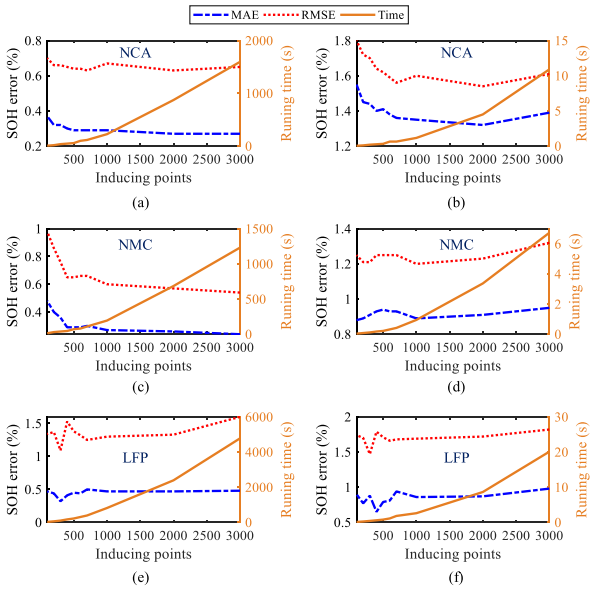

Fig. 11. Influence of inducing points number on battery SOH estimation. (a), (c), (e) Training process of NCA cell, NMC cell, and LFP cell, respectively. (b), (d), (f) Test process of NCA cell, NMC cell, and LFP cell, respectively.

C. Sensitivity Analysis of Segment Length

Since the features for SOH estimation are the statistical characteristics of $\triangle Q$, it is necessary to analyze the effect of the sequence length of $\triangle Q$ on the estimation accuracy. Since the sparse GPR has been verified to have better performance, only the effect on sparse GPR is analyzed in this section. The MAE and RMSE of SOH estimation as a function of the segment length for three types of batteries are shown in Fig. 10. In the training process, the two errors of all battery cells generally show a downward trend as the sequence length increases, and gradually become stable after the sequence length exceeds 40 for NCA cell and 30 for NMC cell. In the test process, the estimation errors of different cells also present a downward trend as the sequence length increases, except for NCA cell, which increase gradually when the sequence length exceeds 40. Note that the segment length is the length of the random charging sequence $\triangle Q$, and the shorter it is, the more likely it is to be extracted from random charging process. Given the estimation accuracy and the probability of the segments, it is recommended to choose the sequence length within 30–40 range for battery SOH estimation.

D. Sensitivity Analysis of Inducing Points Number

The regular GPR uses all samples to train model, while the sparse GPR only uses inducing points to train model, and a better performance has been obtained by the latter one. The above results indicate that more sample number is not always better so that it is important to analyze the effect of the number of inducing points on the estimation accuracy and computational time. The influence of inducing points number on battery SOH estimation is shown in Fig. 11, in which the left y-axis represents estimation errors and the right y-axis represents running time. It can be seen that as the number of inducing points increases, the MAE and RMSE of different cells just fluctuate slightly in the training process, except for NMC cell with a quick decrease in the beginning. In the test process, the estimation errors of different cells keep relatively stable as the increase of inducing points, except for NCA cell, which has a large decrease in the beginning, and a slight increase trend is observed for all cells after the inducing points over 2000. Besides, as the number of inducing points increases, the running time increases quadratically for the training and test processes, which has been described mathematically in Section IV-B. Considering both the computational time and estimation accuracy, it is suggested to choose the number of inducing points within 500–1500 range for battery SOH estimation.

VI. Conclusion

In EVs, it is important to accurately monitor the battery health status. To enable battery SOH estimation under actual vehicle operating condition, a data-driven method based on a random partial charging process and the sparse GPR is proposed in this article.

In the battery CC charging process, it has been observed that there is an implicit relationship between the capacity increment sequence $\triangle Q$ and battery aging status. The average value (ave_$\triangle Q$) and standard deviation (std_$\triangle Q$) of $\triangle Q$ within different voltage segments are extracted as features to indicate battery health. Pearson correlation analysis is conducted on three types of batteries, and the proposed two features are verified to have high correlations to battery SOH under different temperatures and discharging current rates. Most of the correlations are greater than 0.9 for NCA and NMC cells, and 0.8 for LFP cells.

By using the proposed two features and the mean value of voltage segment as the inputs, the regular and sparse GPR models are constructed to estimate battery SOH, and their performance is compared with other typical data-driven methods, including

MLR, SVM, RVM, and CNN. Among them, the sparse GPR has been verified with the highest estimation accuracy, and its average MAEs under different conditions are 2.88%, 2.52%, and 1.51% for NCA, NMC, and LFP cells, respectively. Compared with the regular GPR, the sparse one not only reduces the calculation time, but also has higher estimation accuracy because the overfitting problem of the regular one can be addressed by the optimization of the inducing points. Besides, the sensitivity analysis indicates that it is better to set the sequence length within 30–40 range and the number of inducing points within 500–1500 range.

In our future work, we will try to combine the advanced collaborative GPR and multioutput convolved GPR with the sparse mechanism to achieve more accurate battery SOH estimation.

REFERENCES

[1] F. Zhao, F. Liu, Z. Liu, and H. Hao, "The correlated impacts of fuel consumption improvements and vehicle electrification on vehicle greenhouse gas emissions in China," *J. Cleaner Prod.,* vol. 207, pp. 702–716, 2019.

[2] N. Manthey, "EU to target 30 million electric vehicles by 2030," 2020. [Online]. Available: https://www.electrive.com/2020/12/06/eu-to-target-30-million-electric-vehicles-by-2030

[3] Y. Wang *et al.*, "A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems," *Renewable Sustain. Energy Rev.*, vol. 131, 2020, Art. no. 110015.

[4] Y. Wang, L. Wang, M. Li, and Z. Chen, "A review of key issues for control and management in battery and ultra-capacitor hybrid energy storage systems," *eTransportation,* vol. 4, 2020, Art. no. 100064.

[5] D. Roman, S. Saxena, V. Robu, M. Pecht, and D. Flynn, "Machine learning pipeline for battery state-of-health estimation," *Nat. Mach. Intell.,* vol. 3, no. 5, pp. 447–456, 2021.

[6] A. E. Mejdoubi, H. Chaoui, H. Gualous, P. V. D. Bossche, N. Omar, and J. V. Mierlo, "Lithium-ion batteries health prognosis considering aging conditions," *IEEE Trans. Power Electron.*, vol. 34, no. 7, pp. 6834–6844, Jul. 2019.

[7] X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics," *Joule*, vol. 4, no. 2, pp. 310–346, 2020.

[8] X. Han, M. Ouyang, L. Lu, J. Li, Y. Zheng, and Z. Li, "A comparative study of commercial lithium ion battery cycle life in electrical vehicle: Aging mechanism identification," *J. Power Sources*, vol. 251, pp. 38–54, 2014.

[9] A. Barré, B. Deguilhem, S. Grolleau, M. Gérard, F. Suard, and D. Riu, "A review on lithium-ion battery ageing mechanisms and estimations for automotive applications," *J. Power Sources*, vol. 241, no. 11, pp. 680–689, 2013.

[10] D. Andre, M. Meiler, K. Steiner, C. Wimmer, T. Soczka-Guth, and D. U. Sauer, "Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy—I: Experimental investigation," *J. Power Sources,* vol. 196, no. 12, pp. 5334–5341, 2011.

[11] U. Energy, "Battery test manual for plug in hybrid electric vehicles," *V. TP Energy Efficiency Renewable Energy*, editor, Idaho Operations Office, 2014.

[12] Z. Deng, L. Yang, Y. Cai, H. Deng, and L. Sun, "Online available capacity prediction and state of charge estimation based on advanced data-driven algorithms for lithium iron phosphate battery," *Energy,* vol. 112, pp. 469–480, 2016.

[13] 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, 2014.

[14] Y. Li *et al.*, "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review," *Renewable Sustain. Energy Rev.,* vol. 113, 2019, Art. no. 109254.

[15] X. Bian, Z. G. Wei, W. Li, J. Pou, D. U. Sauer, and L. Liu, "State-of-health estimation of Lithium-ion batteries by fusing an open-circuit-voltage model and incremental capacity analysis," *IEEE Trans. Power Electron.*, vol. 37, no. 2, pp. 2226–2236, Feb. 2022.

[16] J. Wu, X. Cui, H. Zhang, and M. Lin, "Health prognosis with optimized feature selection for lithium-ion battery in electric vehicle applications," *IEEE Trans. Power Electron.,* vol. 36, no. 11, pp. 12646–12655, Nov. 2021.

[17] K. A. Severson *et al.*, "Data-driven prediction of battery cycle life before capacity degradation," *Nat. Energy*, vol. 4, no. 5, pp. 383–391, 2019.

[18] X. Lai *et al.*, "Sorting, regrouping, and echelon utilization of the large-scale retired lithium batteries: A critical review," *Renewable Sustain. Energy Rev.*, vol. 146, 2021, Art. no. 111162.

[19] A. Farmann, W. Waag, A. Marongiu, and D. U. Sauer, "Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles," *J. Power Sources*, vol. 281, pp. 114–130, 2015.

[20] M. Schimpe, M. E. von Kuepach, M. Naumann, H. C. Hesse, K. Smith, and A. Jossen, "Comprehensive modeling of temperature-dependent degradation mechanisms in lithium iron phosphate batteries," *J. Electrochem. Soc.,* vol. 165, no. 2, pp. A181–A193, 2018.

[21] D. Xiao *et al.*, "Reduced-coupling coestimation of SOC and SOH for lithium-ion batteries based on convex optimization," *IEEE Trans. Power Electron.,* vol. 35, no. 11, pp. 12332–12346, Nov. 2020.

[22] Z. Lyu, R. Gao, and L. Chen, "Li-ion battery state of health estimation and remaining useful life prediction through a model-data-fusion method," *IEEE Trans. Power Electron.,* vol. 36, no. 6, pp. 6228–6240, Jun. 2021.

[23] R. Zhu, B. Duan, J. Zhang, Q. Zhang, and C. Zhang, "Co-estimation of model parameters and state-of-charge for lithium-ion batteries with recursive restricted total least squares and unscented Kalman filter," *Appl. Energy,* vol. 277, 2020, Art. no. 115494.

[24] X. Lai *et al.*, "Capacity estimation of lithium-ion cells by combining model-based and data-driven methods based on a sequential extended Kalman filter," *Energy*, vol. 216, 2020, Art. no. 119233.

[25] Z. Song, X. Wu, X. Li, J. Sun, H. F. Hofmann, and J. Hou, "Current profile optimization for combined state of charge and state of health estimation of lithium ion battery based on Cramer–Rao bound analysis," *IEEE Trans. Power Electron.*, vol. 34, no. 7, pp. 7067–7078, Jul. 2019.

[26] S. Schwunk, N. Armbruster, S. Straub, J. Kehl, and M. Vetter, "Particle filter for state of charge and state of health estimation for lithium–iron phosphate batteries," *J. Power Sources,* vol. 239, pp. 705–710, 2013.

[27] Z. Wei, J. Hu, H. He, Y. Li, and B. Xiong, "Load current and state-of-charge coestimation for current sensor-free lithium-ion battery," *IEEE Trans. Power Electron.,* vol. 36, no. 10, pp. 10970–10975, Oct. 2021.

[28] Z. Deng, H. Deng, L. Yang, Y. Cai, and X. Zhao, "Implementation of reduced-order physics-based model and multi-parameters identification strategy for lithium-ion battery," *Energy,* vol. 138, pp. 509–519, 2017.

[29] Z. Deng, X. Hu, X. Lin, Y. Kim, and J. Li, "Sensitivity analysis and joint estimation of parameters and states for all-solid-state batteries," *IEEE Trans. Transp. Electrif.*, vol. 7, no. 3, pp. 1314–1323, Sep. 2021.

[30] H. Li, D. Pan, and C. L. P. Chen, "Intelligent prognostics for battery health monitoring using the mean entropy and relevance vector machine," *IEEE Trans. Syst., Man, Cybern., Syst.,* vol. 44, no. 7, pp. 851–862, Jul. 2014.

[31] P. Li *et al.*, "State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network," *J. Power Sources*, vol. 459, 2020, Art. no. 228069.

[32] S. Shen, M. Sadoughi, M. Li, Z. Wang, and C. Hu, "Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries," *Appl. Energy*, vol. 260, 2020, Art. no. 114296.

[33] 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, 2018.

[34] A. A. Chehade and A. A. Hussein, "A multioutput convolved Gaussian process for capacity forecasting of li-ion battery cells," *IEEE Trans. Power Electron.,* vol. 37, no. 1, pp. 896–909, Jan. 2022.

[35] A. A. Chehade and A. A. Hussein, "A collaborative gaussian process regression model for transfer learning of capacity trends between li-ion battery cells," *IEEE Trans. Veh. Technol.,* vol. 69, no. 9, pp. 9542–9552, Sep. 2020.

[36] J. Tian, R. Xiong, and W. Shen, "State-of-health estimation based on differential temperature for lithium ion batteries," *IEEE Trans. Power Electron.,* vol. 35, no. 10, pp. 10363–10373, Oct. 2020.

[37] Z. Deng, X. Hu, X. Lin, L. Xu, Y. Che, and L. Hu, "General discharge voltage information enabled health evaluation for lithium-ion batteries," *IEEE/ASME Trans. Mechatron.*, vol. 26, no. 3, pp. 1295–1306, Jun. 2021.

[38] X. Hu, Y. Che, X. Lin, and Z. Deng, "Health prognosis for electric vehicle battery packs: A data-driven approach," *IEEE/ASME Trans. Mechatron.,* vol. 25, no. 6, pp. 2622–2632, Dec. 2020.

[39] J. Tian, R. Xiong, W. Shen, J. Lu, and X.-G. Yang, "Deep neural network battery charging curve prediction using 30 points collected in 10 min," *Joule*, vol. 5, no. 6, pp. 1521–1534, 2021.
[40] D. Liu, Y. Song, L. Li, H. Liao, and Y. Peng, "On-line life cycle health assessment for lithium-ion battery in electric vehicles," *J. Cleaner Prod.*, vol. 199, pp. 1050–1065, 2018.
[41] Y. Preger *et al.*, "Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions," *J. Electrochem. Soc.*, vol. 167, no. 12, 2020, Art. no. 120532.
[42] B. Archive, "Homepage of Battery Archive," 2021. [Online]. Available: www.batteryarchive.org
[43] S. Shen, M. Sadoughi, X. Chen, M. Hong, and C. Hu, "A deep learning method for online capacity estimation of lithium-ion batteries," *J. Energy Storage,* vol. 25, 2019, Art. no. 100817.
[44] Y. Che, Z. Deng, X. Lin, L. Hu, and X. Hu, "Predictive battery health management with transfer learning and online model correction," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1269–1277, Feb. 2021.
[45] C. Qian *et al.*, "Convolutional neural network based capacity estimation using random segments of the charging curves for lithium-ion batteries," *Energy*, vol. 227, 2021, Art. no. 120333.
[46] Y. Zhou, M. Huang, Y. Chen, and Y. Tao, "A novel health indicator for on-line lithium-ion batteries remaining useful life prediction," *J. Power Sources,* vol. 321, pp. 1–10, 2016.
[47] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning (no. 3).* Cambridge, MA, USA: MIT Press, 2006.
[48] K. Liu, X. Hu, Z. Wei, Y. Li, and Y. Jiang, "Modified gaussian process regression models for cyclic capacity prediction of lithium-ion batteries," *IEEE Trans. Transp. Electrif.,* vol. 5, no. 4, pp. 1225–1236, Dec. 2019.
[49] M. Liu, G. Chowdhary, B. C. da Silva, S.-Y. Liu, and J. P. How, "Gaussian processes for learning and control: A tutorial with examples," vol. 38, no. 5, pp. 53–86, 2018.
[50] J. Q. Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *J. Mach. Learn. Res.,* vol. 6, pp. 1939–1959, 2005.
[51] G. Ozcan, M. Pajovic, Z. Sahinoglu, Y. Wang, P. V. Orlik, and T. Wada, "Online battery state-of-charge estimation based on sparse gaussian process regression," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2016, pp. 1–5.

Zhongwei Deng (Member, IEEE) received the B.S. degree from Jilin University, Changchun, China, in 2014, and the Ph.D. degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2019.

He is currently an Assistant/Postdoctoral Researcher with the College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, China. His research interests include data-driven and electrochemical mechanism modeling, parameter identification, states estimation, health diagnosis, and second-life utilization of lithium-ion batteries.

Xiaosong Hu (Senior Member, IEEE) received the Ph.D. degree in automotive engineering from the Beijing Institute of Technology, Beijing, China, in 2012.

He did scientific research and completed the Ph.D. dissertation from the Automotive Research Center University of Michigan, Ann Arbor, MI, USA, from 2010 to 2012. He is currently a Professor with the State Key Laboratory of Mechanical Transmissions, College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, China. He was a Postdoctoral Researcher with the University of California (Berkeley), Chalmers University of Technology, and Swiss Federal Institute of Technology. His research interests include modeling and control of alternative powertrains and energy storage systems.

Dr. Hu was the recipient of several prestigious awards/honors, including the IEEE ITSS Young Researcher/Engineer Award, SAE Ralph Teetor Educational Award, Energies Young Investigator Award, Emerging Sustainability Leaders Award in 2016, EU Marie Currie Fellowship in 2015, ASME DSCD Energy Systems Best Paper Award in 2015, and Beijing Best Ph.D. Dissertation Award in 2013.

Penghua Li was born in 1984. He received the B.S. degree in electronic information science and technology and the Ph.D. degree in control theory and control engineering from Chongqing University, Chongqing, China, in 2008 and 2012, respectively.

He is currently a Professor with the Chongqing University of Posts and Telecommunications (CQUPT), Chongqing, China. He is the Deputy Director with the Department of Measurement and Control, Automation College, CQUPT. He was also a Senior Visiting Scholar with the Vienna University of Technology.

Dr. Li is the Standing Committee Member of the Intelligent Transportation Professional Committee, CQAIS and the Director of the Chongqing Artificial Intelligence Society (CQAIS). He chaired the 27th and 30th China Conference on control and decision-making neural networks. He was the recipient of many talent titles, e.g., the Outstanding Youth of Chongqing and the Young Scientific and Technological Talent of Chongqing. He was also the recipient of the First Prize of Chongqing Science and Technology Progress Award twice in 2018 and 2019.

Xianke Lin (Member, IEEE) received the B.S. degree from Zhejiang University, Hangzhou, China, in 2009, and the Ph.D. degree in mechanical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2014.

He has extensive industrial experience with Fiat Chrysler Automobiles, Mercedes-Benz R&D North America, and General Motor of Canada. He is currently an Assistant Professor with the Department of Automotive and Mechatronics Engineering, Ontario Tech University, Oshawa, ON, Canada. His research interests include hybrid powertrain design and control strategy optimization, multiscale/multiphysics modeling, and optimization of energy storage systems.

Xiaolei Bian received the B.Eng. and M.Sc. degrees in mechanical engineering from Jilin University, Changchun, China, in 2014 and 2017, and the Ph.D. degree in chemical engineering from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2021.

He is currently a Postdoctoral Researcher with the Department of Chemical Engineering, KTH Royal Institute of Technology. His research interests include modeling and state estimation of lithium-ion batteries.