IEEE TRANSACTIONS ON ENERGY CONVERSION

Data-Driven State of Health Estimation Method of Lithium-ion Batteries for Partial Charging Curves

Jinrui Tang, *Member, IEEE*, Yang Li, *Senior Member, IEEE*, Shaojin Wang, Binyu Xiong, *Member, IEEE*, Xiangjun Li, Jinxuan Pan, Qihong Chen, *Member, IEEE*, and Peng Wang, *IEEE Fellow*

***Abstract*—State of health (SOH) is one of the most important performance indicators of lithium-ion batteries (LIBs). Accurate estimation of SOH is a prerequisite for the safe and reliable operation of LIBs. Traditional SOH estimation methods predominantly rely on complete charging cycle data acquired through laboratory testing. However, in practical application, the charging behaviors of electric vehicle users are random and unpredictable, making the partial charging curves difficult to utilize the traditional methods. This work introduces a novel data-driven approach to estimating a battery's SOH for partial charging cases. Firstly, a curve fitting method is proposed to extract health indicators (HIs) from partial charging voltage data, where novel HIs based on the energy-voltage curve are extracted. A composite Gaussian process regression-based data-driven method is proposed to achieve highly accurate SOH estimation. The method's adaptability to real-world partial charging habits is evaluated through three representative scenarios derived from extensive charging behavior reports of EV users. The impact of partial charging on HI extraction is analyzed based on the three identified scenarios. The proposed method is verified using a combination of our laboratory testing data and the Oxford open dataset. The results show that the proposed framework demonstrates the ability to estimate SOH accurately and strong robustness to various partial charging behaviors.**

***Index Terms*—Lithium-ion battery, health indicator, data-driven method, partial charging, state of health estimation.**

I. INTRODUCTION

LITHIUM-ION batteries (LIBs) have widespread adoption across a diverse range of applications, including electric vehicles (EVs), grid-scale energy storage, and mobile electronic devices, attributed to their superiority in energy density, energy conversion efficiency, and working voltage[1]. However, LIB battery cells inevitably experience gradual aging, leading to degraded performance and potential pack-level failures. The state of health (SOH) is a vital metric for signifying this degradation, typically defined as the ratio of the present to the rated battery capacity [2]. Since SOH is not directly measurable, it is imperative to estimate the SOH accurately based on available measurements such as terminal voltage, applied current, and surface temperature.

Typically, SOH estimation algorithms can be categorized into model-based and data-driven methods. Model-based techniques use mechanism knowledge of LIBs to depict the aging trajectory and potential failure modes [3-6]. The models can be classified into electrochemical models [3], empirical models [4], and equivalent circuit models [5]. Despite their wide usage, model-based methods encounter challenges in precisely identifying model parameters, as they may change significantly under large variations of the operating conditions [6].

Data-driven methods have provided another pathway for SOH estimation. Extensive research efforts have been made in recent years, and massive algorithms have been developed [7-13]. Machine learning is one of the most important data-driven methods for modeling complex and nonlinear degradation behavior of LIBs based on historical data. Many machine learning algorithms have been applied for SOH estimation, including artificial neural network (ANN) and its variants [7], [8], support vector machine (SVM) [9], correlation vector machine [10], Gaussian process regression (GPR) [11], and model migration and ensemble learning [12]. In these algorithms, some characteristic parameters are extracted as model inputs from testing and operation data of charging/discharging current and temperature profiles. These parameters can reflect battery aging and are referred to as health indicators (HIs) [13]. The accuracy of the estimation algorithms can be affected by the selection of HIs on top of the nature of the machine learning itself, such as the number of input/output layers, hyperparameters, loss function, training time, learning rate schedule, and computer accuracy. In [14], [15], temperature-variation-based HIs were used for SOH estimation. However, the drawback is that temperature data are significantly affected by environmental conditions and noises. Other widely used HIs were extracted from the charging current curves, most based on the constant-current constant-voltage (CC-CV) strategy [16], [17], [18]. Compared to the discharging curves, the charging curves are much more regular and easier for HIs extraction [19]. Well-investigated HIs include those based on CC-CV curves, such as time interval, voltage interval [16], incremental capacity analysis (ICA) [17], and differential voltage analysis [18].

Most of the studies mentioned above rely on the datasets of

This work was supported by the National Natural Science Foundation of China under Grant #52177221. (*Corresponding author: Binyu Xiong*)

Jinrui Tang and Binyu Xiong are with the State Key Laboratory of Maritime Technology and Safety, Hubei Province New Energy Power Battery Engineering Research Center. (e-mail: tangjinrui@whut.edu.cn; bxiong2@whut.edu.cn).

Yang Li is with the Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden. (e-mail: yangli@ieee.org).

Shaojin Wang and Qihong Chen are with the School of Automation, Wuhan University of Technology, Wuhan 430070, China. (e-mail: shaojinwang@whut.edu.cn; chenqh@whut.edu.cn).

Xiangju Li is with the Energy Storage System Integration and Configuration Technology Research Lab. at the China Electric Power Research Institute. (e-mail: li_xiangjun@126.com).

Jinxuan Pan is with the State Grid Yichun Power Supply Company. (e-mail: 370200010@qq.com)

Peng Wang was with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore (e-mail: epwang@ntu.edu.sg).

complete CC-CV curves obtained from the battery testing phase. However, in the operation phase, an EV user usually begins to charge the batteries before their depletion [20]. The charge/discharge processes become random and unpredictable depending on the driving conditions and the user's driving behaviors, which makes the charge/discharge profiles usually partial and irregular. A partial charge/discharge dataset does not include many HIs used in the literature, leading to a failure for HI extraction and poor SOH estimation results. For instance, there can be some limitations to using traditional HIs based on fixed voltage or time intervals since extracted due to the incomplete charging curves. Therefore, selecting a feasible combination of HIs from the partial charging curves poses a challenge for SOH estimation in real-world scenarios.

A straightforward and simple solution to the above problem is to divide the partial charging cases into several categories and establish different HIs combinations for each category. Some initial attempts have been made for such investigations [21-23]. For example, Wei *et al*. [21] considered two partial charging cases based on the initial charging terminal voltage. These cases include two main situations: When there is moderate partial charging with an initial terminal voltage below 4.0 V and when there is heavy partial charging with an initial terminal voltage between 4.0 V and 4.1 V. Different sets of HIs were extracted from partial constant current charging data with different initial charging voltages, including morphological incremental capacity features and voltage entropy information. This method was verified when the initial charging voltage was as low as 3.7 V during partial charging. Similarly, He *et al.* [23] proposed a method to estimate the SOH using incremental capacity (IC) characteristics of partial CC charging data. However, in the above works, although schemes show strong regularity, their results were biased towards the laboratory discharge case,

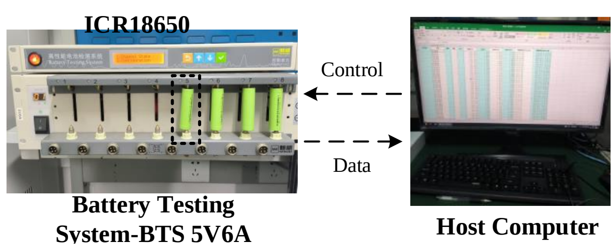

Fig.1. Battery testing platform.

TABLE I
OPERATING CONDITIONS AND PARAMETERS OF LCO CELLS

Battery Number	Charging Mode	Charging Current (A)	Discharging Mode	Discharging Current (A)
Cell 1	CC-CV	1.5 (1C)	CC	1.5
Cell 2	CC-CV	3.0 (2C)	CC	3.0
Cell 3	CC-CV	1.5 (1C)	WLTC	-
Cell 4	CC-CV	3.0 (2C)	WLTC	-

possibly because the discharge patterns during the driving of EVs were not employed or simulated by unrealistic random walks.

In real EV operation, due to the variability of driving roads and driving habits, the discharge curve will be in a disordered situation rather than a regular distribution. Although the above

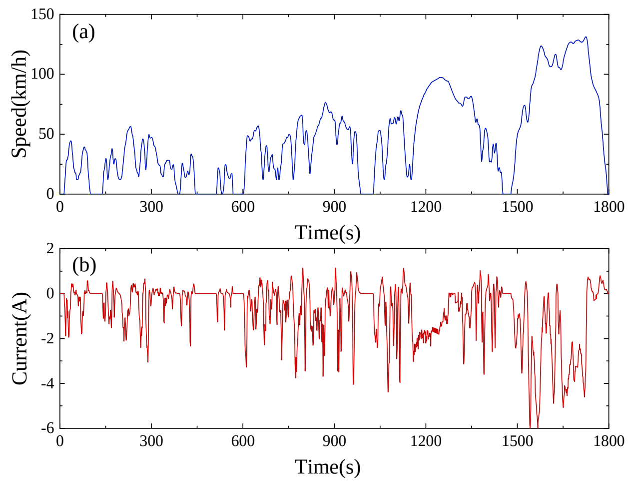

Fig. 2. WLTC speed-current profile. (a) WLTC driving cycle for Class 3b vehicles. (b) Battery current profile corresponding to the driving cycle.

works start from partial charging curves and can extract effective characteristic parameters from a wide range of partial charging curves, no battery data close to real vehicle operation were used to validate the model. Despite the attempts made in [22] [23], the discharge model is still biased towards the laboratory type of discharge model and different from the real operating situation. Bian *et al*. [22], He *et al*. [23], and Stroe *et al*. [24] also studied the feasibility of their proposed SOH estimation methods under partial charging. However, these studies did not consider user charging behavior habits when classifying partial charging cases. For instance, EV drivers tend to use state-of-charge (SOC) rather than terminal voltage as the main criterion to make their charging decisions [25]. The relationship between SOC and terminal voltage varies as the battery ages [21]. Therefore, the SOC and user charging behavior should be considered when classifying partial charging cases, but unfortunately, the relevant investigation is lacking in the literature.

In view of the above, a data-driven method is proposed to estimate battery SOH for partial charging cases. Moreover, the proposed method is verified by our laboratory testing data and the Oxford open dataset. The main contributions of the paper are as follows:

1) A curve fitting method is proposed to extract HIs from partial charging voltage data. More than 20 HIs are extracted from charging voltage data, including novel HIs based on the energy-voltage curve.

2) A composite Gaussian process regression (CGPR) based data-driven method is proposed to estimate battery SOH. CGPR can be regarded as two GPR models. In the model training phase, the estimated residual in the model training phase of one GPR model (global term) is modeled by another GPR model (local term). More flexible and accurate SOH estimation is achieved by well-designed composite kernel functions.

3) Three representative partial charging cases are selected for case studies based on charging behavior reports of EV users. The proposed framework is verified in these cases.

The rest of this paper is structured as follows: In Section II, the two battery aging datasets used in this study are introduced. In Section III, the proposed SOH estimation framework for partial charging cases is introduced. In Section IV, three representative partial charging cases are defined for the case study, and the proposed method is verified by our experimental data and the Oxford open dataset. Finally, the conclusion and further discussion are given in Section V.

II. BATTERY AGING DATASET

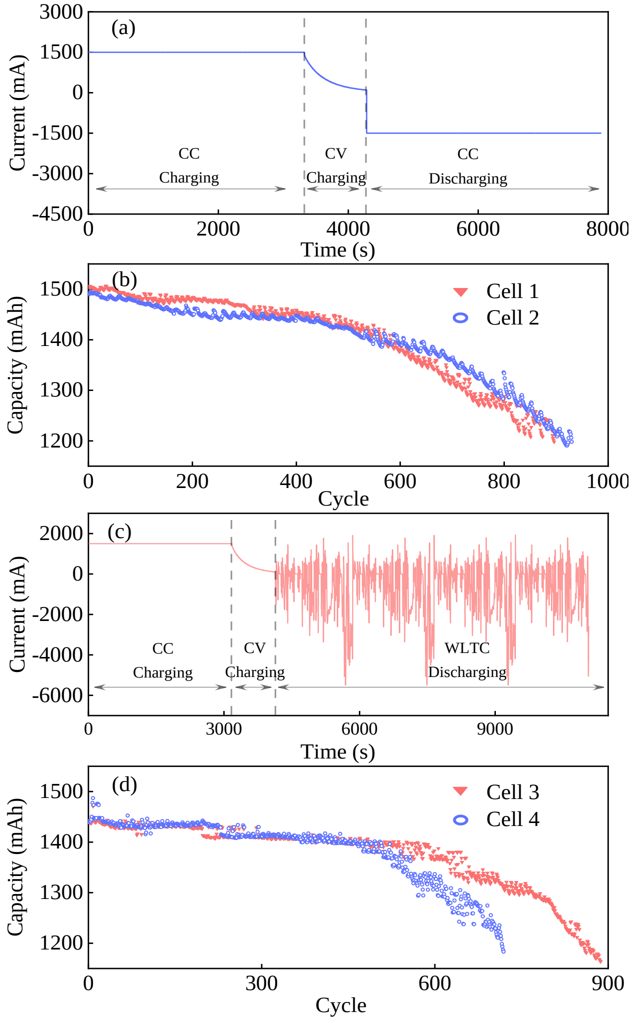

Fig. 3. Battery testing condition and remaining capacity. (a) Current profile of Cell 1 and Cell 2 in a CC-CV cycle. (b) Capacity fading curves of Cell 1 and Cell 2. (c) Current profile of Cell 3 and Cell 4 in a WLTC cycle. (d) Capacity fading curves of Cell 3 and Cell 4.

Two different battery aging datasets are used in this study, i.e., our multi-condition battery aging dataset and the Oxford open dataset. The common characteristic of these batteries is the complex discharge conditions, which are more practical compared to traditional constant current discharge conditions.

A. Multi-Condition Aging Tests of LCO cells

In this study, four 18650 lithium cobalt oxide (LCO) cells with a rated capacity of 1.5 Ah are used for the test. The battery aging experimental platform is shown in Fig. 1. Different operating conditions are adopted for the four cells as given in Table I: Cell 1 and Cell 3 experience a 1-h CC charging with 1.5 A (1C), and Cell 2 and Cell 4 are charged 0.5 h using 3 A (2C) CC. At the end of each charge or discharge, the cells were rested for 10 min and cycled until the remaining capacity degraded to 80% of the rated value.

Specifically, the CC-CV charging mode is adopted for all cells. These cells are further divided into two groups, and

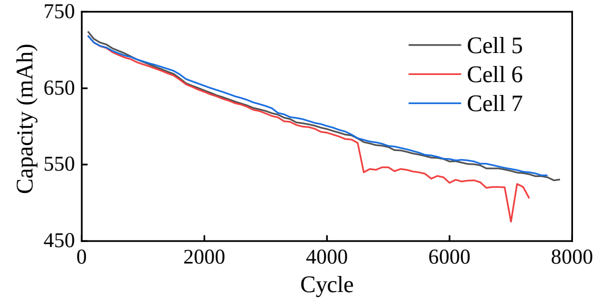

Fig. 4. Capacity fading curves of Cell 5 to Cell 7.

different discharge strategies are adopted. CC discharge condition is adopted for the first group of cells, including Cells 1 and 2. For the second group of cells, Cells 3 and 4, a new test cycle known as the worldwide harmonized light vehicles test cycle (WLTC) is adopted to mimic the discharge condition, which includes four driving conditions: Urban area (low speed), suburban area (medium speed), rural area (high speed), and expressway (super high speed), with a total duration of 1800 s. The simulated condition is obtained from the WLTC class 3b driving cycle, shown in Fig. 2(a). Based on vehicle dynamics [26], the speed-time curve of the vehicle is converted into the current-time curve of the power battery, as shown in Fig. 2(b). The speed v is converted into the current I according to,

$$P=(mgf_r+\frac{1}{2}\rho_a C_D A_f v^2+m\delta a+mgi)v \tag{1}$$

$$I=-\left[\frac{1}{\eta_w}\frac{1+\mathrm{sgn}(P)}{2}+\eta_r\frac{1-\mathrm{sgn}(P)}{2}\right]\frac{P}{\alpha\cdot N\cdot V_b} \tag{2}$$

where P is the required traction power of the EV, and α is the acceleration of the EV. More details about the conversion of the current curve, including the meanings of other parameters, can be found in [26].

The battery testing condition and remaining capacity of the battery are shown in Fig. 3. The current curve of the first group of cells in a single cycle is shown in Fig. 3(a). Fig. 3(b) shows the capacity fading curve of Cells 1 and 2. The current curve of the second group of cells in a single cycle is shown in Fig. 3(c). Furthermore, Fig. 3(d) shows the capacity fading curve of Cells 3 and 4. The results show that the capacity of LIBs decays nonlinearly, and there is a periodic capacity regeneration phenomenon. This is because the predominant mechanisms for LIB capacity degradation include the loss of lithium-ion inventory (LLI) and loss of active materials (LAM) caused by various side reactions[27]. However, the products of these side reactions are sometimes unstable and may decompose and release lithium ions as the battery operates, which leads to a periodic increase of capacity. In addition, the aging paths vary with LIBs due to inevitable cell-to-cell inconsistency.

B. Oxford Open Dataset

The Oxford battery degradation data set includes the degradation data of 8 small Li-ion pouch cells produced by the KOKAM [21], [28]. These cells with a rated capacity of 740 mAh are charged at a constant current of 1C and discharged

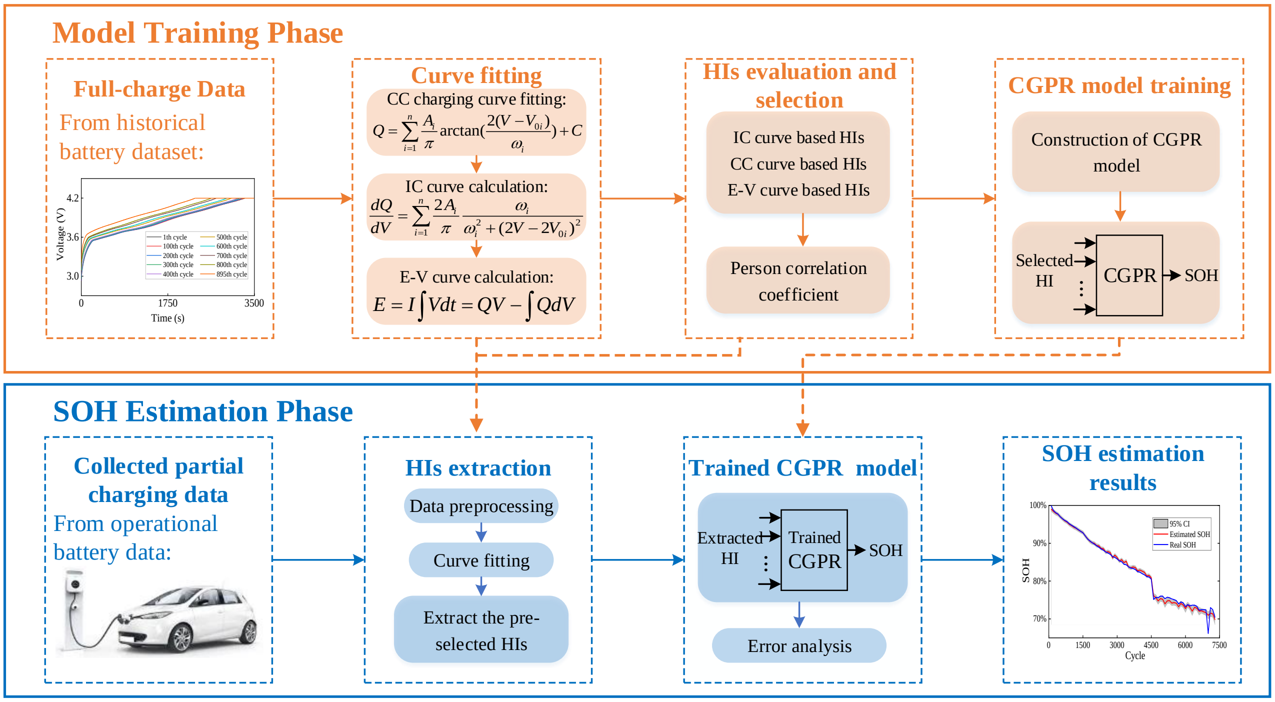

Fig.5. SOH estimation framework for partial charging cases.

under the urban Artemis profile. The SOH is tested every 100 cycles. In this study, we select the first three cells in the dataset for framework verification. The three cells are denoted as Cells 5 to 7. The capacity fading curves of Cell 5 to Cell 7 are shown in Fig. 4.

III. SOH ESTIMATION FRAMEWORK

A SOH estimation framework for partial charging cases is proposed in this section. As shown in Fig. 5, the framework consists of model training and SOH estimation phases. In the framework, a new curve-fitting-based method is proposed to extract HI from charging voltage data. In addition, composite Gaussian process regression (CGPR) is proposed for fusing extracted HI to estimate battery SOH. In this section, the HI extraction method and the CGPR model will be introduced based on the measured data of Cell 1.

A. HI Extraction Based on Curve Fitting

Although all batteries are charged under the CC-CV protocol in the laboratory, we assume only the CC data are available for HI extraction, considering the CV mode is less adopted in practice. A large amount of aging information is contained in the charging voltage curve of LIB, but there are some indirect HIs that are difficult to extract. In partial charging cases, further loss of aging information increases the difficulty of HI extraction. The ICA method is based on the slope of the charging voltage curve, which is not affected by the partial voltage segment. Therefore, the ICA method is expected to extract indirect HI even in partial charging cases.

ICA is an in-situ health diagnosis technology that can convert the voltage plateau in the voltage curve into the peak value on the incremental capacity (IC) curve. IC is defined as:

$$IC = \frac{dQ}{dV} = I\frac{dt}{dV} \tag{3}$$

where IC represents the original IC curve, V is the charging voltage, I is the charging current, and t is the charging time. Q is the charging capacity, which equals the integral of the charging current over time.

The IC method can be divided into differential filtering and curve fitting. The differential filtering method approximates the derivative using the ratio of two differences. Specifically, in (3), dt and dV are replaced by $t_k - t_{k\text{-}1}$ and $V_k - V_{k\text{-}1}$, respectively, where k represents the step of the discrete point. The original IC curve is obtained using this approximation, and the resulting noises are then removed by applying certain filtering algorithms [29].

Variation in the experimental IC curve with battery aging of Cell 1 is shown in Fig. 6. As the battery ages, the peak of the IC curves will weaken and deviate regularly. This phenomenon is related to the fading of capacity and becomes more significant in the presence of the polarization effect. It can be observed that the first peak gradually becomes unable to be identified at a high current rate of 1C after 700 cycles. This indicates that using IC curves at high current rates as an in situ diagnostic technique can have some limitations in analyzing the aging mechanism of batteries.

Considering the feasibility of further extracting other HIs, the curve-fitting method is adopted in this study. The commonly used peak fitting functions include the Gaussian and Lorentz functions [29]. The Lorentz function is used in this study to express the IC curve due to its simplicity. Then, the IC curve can be fitted by:

$$\frac{dQ}{dV} = \sum_{i=1}^{n} \frac{2A_i}{\pi} \frac{\omega_i}{\omega_i^2 + (2V - 2V_{0i})^2} \tag{4}$$

where n is the number of peaks, A_i is the area of the ith peak, ω_i is the width of the ith peak at half height, and V_{0i} is the center position of the ith peak. n is determined by the phase transition characteristics of the electrode material. n is set to 4 for LCO cells used in this study. It should be noted that one can always

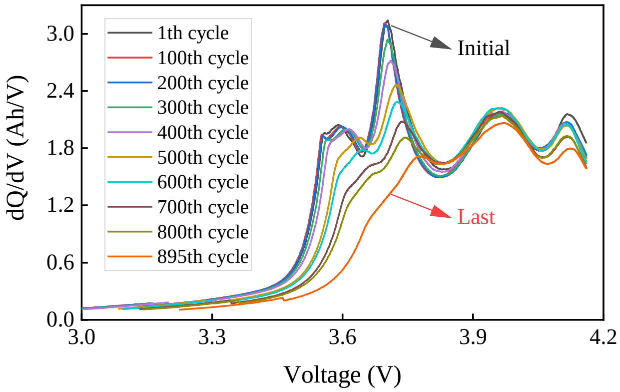

Fig. 6. Variation in IC curve with battery aging of Cell 1.

obtain the first peak through fitting IC curves by (3), although for some cases, the first peak is difficult to identify by the naked eye, as seen in Fig. 6.

To determine the unknown parameters such as A_i, ω_i, and V_{0i}, (4) is integrated to describe the relationship between charging capacity, Q, charging voltage, and V, which can be expressed as:

$$\begin{cases} E=\int Pdt = I\int Vdt=\int VdQ = QV-\int QdV \\ \int QdV = \sum_{i=1}^{n}\left[\frac{A_i(V-V_{0i})}{\pi}\arctan\left(\frac{2(V-V_{0i})}{\omega_i}\right)-\frac{A_i\omega_i}{4\pi}\ln\left(1+\frac{4(V-V_{0i})^2}{\omega_i^2}\right)\right]-C_1V-C_2 \end{cases} \tag{5}$$

$$Q=\sum_{i=1}^{n}\frac{A_i}{\pi}\arctan\left(2\frac{V-V_{0i}}{\omega_i}\right)+C_1 \tag{6}$$

All parameters can be obtained by fitting the raw Q-V curve in (5). Q-V curve is another equal expression form of the CC charging voltage curve when the current is constant. The Q-V curve is used instead of the CC charging voltage curve in the following context.

Moreover, the charging energy E is the integral of power over time and can be used to reflect battery aging.

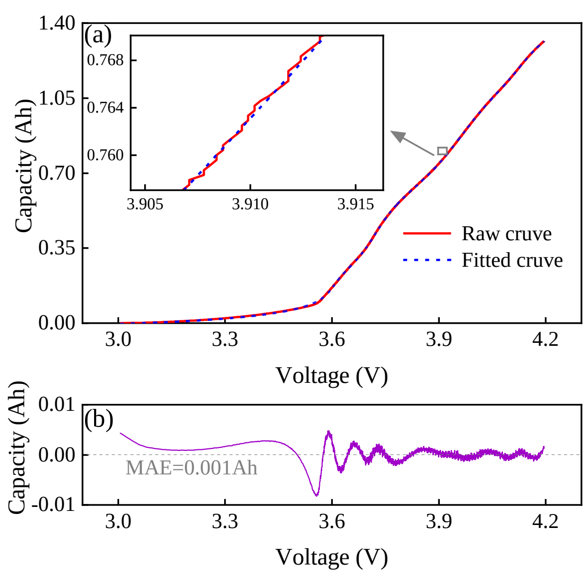

Fig.7. CC charging curve fitting. (a) Curve fitting result. (b) Fitting error.

In the above equations, $3n+1$ parameters need to be identified. In this study, a nonlinear least-squares method is used for curve fitting. The following constraints are set for the parameters based on the properties of the electrode material:

$$\begin{cases} A_i \in (0,1) \ , i=(1,2,3,4) \\ V_{0i}\in(3.5,4.2) \ , i=(1,2,3,4) \\ \omega_i \in (0,0.3) \ , i=(1,2,3,4) \end{cases} \tag{7}$$

Using the full charging cycle data of Cell 1 as an illustration, Fig. 7(a) displays the fitting results of the CC charging curve, while Fig. 7(b) presents the associated errors. This indicates that the proposed function can precisely fit the raw CC charging curve, and the mean absolute error (MAE) is only 0.001 Ah. In Fig. 7(a), the enlarged figure shows that the raw CC charging curve is fluctuating. In contrast, the fitted curve is smooth and preserves the aging information in the raw curve. The fitted IC curve, CC charging curve, E-V curve, and the HIs are shown in Fig. 8. Various HIs can be extracted and plotted in the subfigures.

In Fig. 8(a), the peak height H_1-H_4, peak position V_1-V_4,

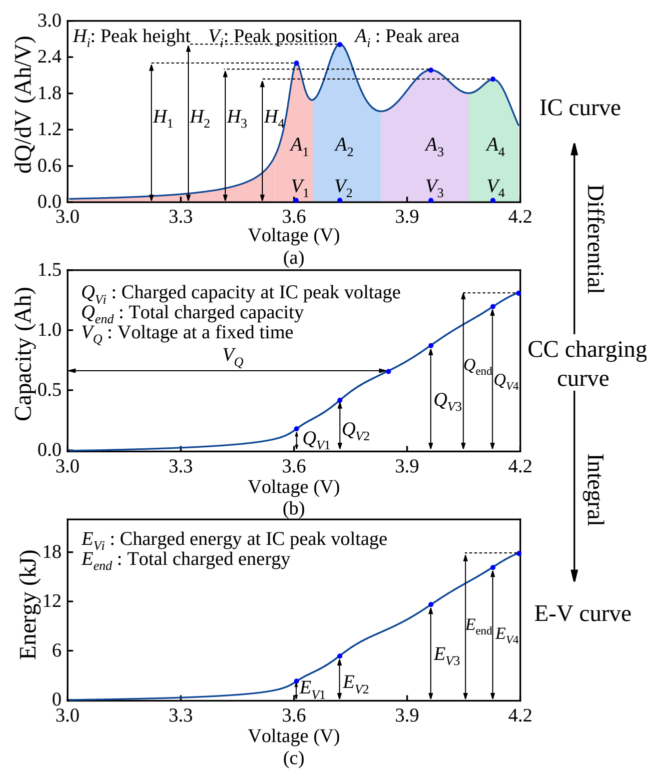

Fig.8. Curve fitting and HI extraction based on (a) IC curve, (b) CC charging curve, and (c) E-V curve.

and peak area A_1-A_4 are plotted as the key HIs. In fact, the peak characteristics on the battery IC curve are influenced by various factors including the type of cell and charging and discharging conditions. In this study, we use an LCO cell as a

$$\begin{pmatrix} y(\mathbf{x}) \\ \mathbf{y} \end{pmatrix} \sim N_{1+n}\left[\begin{pmatrix} \mu \\ \mu\mathbf{1} \end{pmatrix}, \begin{pmatrix} \tau^2+\sigma^2 v(\mathbf{x}) & (\tau^2\mathbf{g}(\mathbf{x})+\sigma^2 v^{1/2}(\mathbf{x})\sum^{1/2}\mathbf{l}(\mathbf{x}))^{\mathrm{T}} \\ \tau^2\mathbf{g}(\mathbf{x})+\sigma^2 v^{1/2}(\mathbf{x})\sum^{1/2}\mathbf{l}(\mathbf{x}) & \tau^2\mathbf{G}+\sigma^2\sum^{1/2}\mathbf{L}\sum^{1/2} \end{pmatrix}\right] \tag{8}$$

reference to plot the IC curve presented in Fig. 6, identifying four prominent peaks at approximately 3.6 V, 3.7 V, 4.0 V, and 4.1 V. On the other hand, Zhang *et al*. [30] utilized NCM cells for their IC curve analysis, observing two peaks around 3.6 V and 3.83 V. This highlights how different types of batteries can exhibit varying numbers of peaks and peak locations on the IC curve. Similarly, Wei *et al*. [21] used the same LCO cell as our study, but their IC curves showed peaks at 3.8 V, 3.9 V, and 4.0 V. However, due to differences in battery-rated capacity, discharge conditions, discharge cutoff voltage, and other parameters, even cells of the same type can display variations in their IC curve peaks. We notice that most IC curves extracted typically exhibit three or even two peaks. In contrast, the LCO cell employed in our paper can reveal up to four effective peaks. It is important to note that even if a certain cell's IC curve only shows two or three peaks, all these are included in the HIs of our study, ensuring that no information regarding battery degradation is overlooked.

In addition, the peak positions, $V_1 - V_4$, are regarded as important nodes to identify battery aging since they represent the electrode phase transition voltages. They are also served to further extract the HIs of $Q_{V1} - Q_{V4}$ and $E_{V1} - E_{V4}$ in the following steps [21], [31].

In Fig. 8(b), the charging capacity $Q_{V1} - Q_{V4}$, the charging capacity, Q_{end}, and the voltage, V_Q, when the charging capacity is equal to half of Q_{end} are extracted as key HIs. The charging capacity, $Q_{V1} - Q_{V4}$, corresponds to the peak position, the charging capacity Qend of the whole charging process, and the voltage VQ when the charging capacity is equal to half of Q_{end}.

In Fig. 8(c), the charging energy, $E_{V1} - E_{V4}$, the charging energy E_{end}, and the average charging power P are the key HIs extracted from the E-V curve. The charging energy, $E_{V1} - E_{V4}$, corresponds to the peak positions, $V_1 - V_4$. The charging energy E_{end}, and the average charging power P can be expressed as follows:

$$P = \frac{\Delta E}{\Delta t} = I\frac{\Delta E}{\Delta Q} \tag{9}$$

where ΔE is the charging energy, Δt is the charging time interval, I is the charging current, and ΔQ is the charging capacity.

All these HIs can be calculated directly by (4), (5), and (7) without additional processing of the raw data. Therefore, the curve fitting method proposed in this study effectively utilized the aging information in the charging voltage data and extracted 24 HIs. The influence of partial charging on HI extraction and further selection of HIs will be introduced in Section IV.

B. SOH Estimation Based on CGPR

The Gaussian process regression (GPR) is a nonparametric machine learning model. Compared with ANN, SVM, and other popular machine learning models, the main advantage of GPR is that its output has confidence intervals and higher estimation accuracy [32]. A traditional GPR model can be described as:

$$\begin{cases} m(\mathbf{x}) = 0 \\ k(\mathbf{x},\mathbf{x}') = \sigma_f^2 \exp\left(\dfrac{-(\mathbf{x}-\mathbf{x}')^2}{2l^2}\right) \end{cases} \tag{10}$$

where σ_f and l are hyperparameters of square index (SE) function, which control output scale and distance scale. **x** and **x′** represent any two independent feature vectors.

According to (7), the obtained function distribution is defined as a Gaussian process:

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}')) \tag{11}$$

However, the basic GPR only adopts the distance-based SE kernel. Here, "distance" refers to the Mahalanobis distance between two HI vectors. The original multi-dimensional sample data $X_{n\times m}$ (n rows and m columns):

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \tag{12}$$

where each row represents a test sample and n = 2 since only two HI vectors are considered. $X_i = (x_{1i}, x_{2i})$ denotes the ith dimension of the sample (m in total). The multi-dimensional data sample is denoted as $X = (X_1, X_2, \ldots, X_m)$ [33]. The overall mean of a sample is,

$$\mu_X = (\mu_{X_1}, \mu_{X_2} \ldots \mu_{X_m}) = (\frac{x_{11}+x_{21}}{2}, \frac{x_{12}+x_{22}}{2}, \ldots, \frac{x_{1m}+x_{2m}}{2}) \tag{13}$$

The covariance matrix is:

$$\Sigma_X = \mathrm{E}\{(X-\mu_X)^T(X-\mu_X)\} = \frac{1}{2}(X-\mu_X)^T(X-\mu_X) \tag{14}$$

The Mahalanobis distance between the two HI vectors is [34]:

$$d^2 = (X-\mu_X)\Sigma_X^{-1}(X-\mu_X)^T \tag{15}$$

It is abbreviated here as |**x**-**x′**|. In this condition, training samples with a long distance from the testing samples will hardly affect the testing results. Therefore, the estimation performance of the basic GPR model declines significantly in the area where the training samples are sparse [35].

GPR models with a single kernel function may perform poorly in areas with fewer training samples. In order to improve the flexibility and accuracy of SOH estimation, the composite Gaussian process regression (CGPR) model [36] is adopted for SOH estimation in our study. The CGPR introduces a composite structure of global and local terms. The global term is first used to model the overall mapping relationship between HIs and SOH. Then, the local term is added to model the estimated residual of the global term. Specifically, the CGPR model can be expressed as:

$$f(\mathbf{x}) = Z_{global}(\mathbf{x}) + v(\mathbf{x})Z_{local}(\mathbf{x}) \tag{16}$$

where $Z_{global}(\mathbf{x})$ is the global term, and $Z_{local}(\mathbf{x})$ is the local term. The wave function $v(\mathbf{x})$ is the standardized volatility function which fluctuates around the unit value, and it is a measure of the local volatility. The value of $v(\mathbf{x})$ is related to the global term

$Z_{global}(\mathbf{x})$ and its estimated residual [36].

In this study, the nonstationary process is not considered, and $v(\mathbf{x})$ is set to constant "1" to simplify the calculation. Equation (11) is thus simplified to:

$$f(\mathbf{x}) = Z_{global}(\mathbf{x}) + Z_{local}(\mathbf{x}) \tag{17}$$

Here, $Z_{global}(\mathbf{x})$ and $Z_{local}(\mathbf{x})$ are two independent Gaussian processes:

$$\begin{cases} Z_{global}(\mathbf{x}) \sim GP(\mu, \tau^2 g(\mathbf{x}, \mathbf{x}')) \\ Z_{local}(\mathbf{x}) \sim GP(0, \sigma^2 l(\mathbf{x}, \mathbf{x}')) \end{cases} \tag{18}$$

where μ and τ^2 are the mean function and variance of the global Gaussian process $Z_{global}(\mathbf{x})$, respectively. $g(\mathbf{x},\mathbf{x}')$ is the kernel function of $Z_{global}(\mathbf{x})$, which describes the global trend of data. σ^2 is the variance of the local Gaussian process $Z_{local}(\mathbf{x})$. $l(\mathbf{x},\mathbf{x}')$ is the kernel function of $Z_{local}(\mathbf{x})$, which describes the local variation of data. It is worth noting that this model can be equivalent to modeling the mean function of the local Gaussian process $Z_{local}(\mathbf{x})$ with the global Gaussian process $Z_{global}(\mathbf{x})$.

Combining (16) and (17) yields the expression of CGPR, i.e.,

$$f(\mathbf{x}) \sim GP(\mu, \tau^2 g(\mathbf{x}, \mathbf{x}') + \sigma^2 l(\mathbf{x}, \mathbf{x}')) \tag{19}$$

The specific process of how CGPR works is explained as follows. The training and testing processes of the CGPR model are similar to that of the basic GPR. Firstly, in building a CGPR modeling framework, the optimal hyperparameter is solved in the training phase by minimizing the negative log marginal likelihood based on the maximum likelihood estimation method. Secondly, solve for the joint distribution of the training and test set outputs [36], [37].

Finally, the mean $\hat{y}(\mathbf{x})$ and variance $\mathrm{cov}(\hat{y})$ of predicted values can be calculated as:

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \left(\mathbf{g}(\mathbf{x}, \hat{\mathbf{x}}) + \lambda \mathbf{l}(\mathbf{x}, \hat{\mathbf{x}})\right)^T (\mathbf{G} + \lambda \mathbf{L})^{-1} (\mathbf{y} - \hat{\mu} \mathbf{I}_n) \tag{20}$$

$$\begin{aligned} \mathrm{cov}(\hat{y}) = {} & \mathbf{g}(\hat{\mathbf{x}}, \hat{\mathbf{x}}) + \lambda \mathbf{l}(\hat{\mathbf{x}}, \hat{\mathbf{x}}) - \\ & \left(\mathbf{g}(\mathbf{x}, \hat{\mathbf{x}}) + \lambda \mathbf{l}(\mathbf{x}, \hat{\mathbf{x}})\right)^T (\mathbf{G} + \lambda \mathbf{L})^{-1} \left(\mathbf{g}(\mathbf{x}, \hat{\mathbf{x}}) + \lambda \mathbf{l}(\mathbf{x}, \hat{\mathbf{x}})\right) \end{aligned} \tag{21}$$

with

$$\hat{\mu} = \left(\mathbf{I}_n{}^T (\mathbf{G} + \lambda \mathbf{L})^{-1} \mathbf{I}_n\right)^{-1} \mathbf{I}_n{}^T (\mathbf{G} + \lambda \mathbf{L})^{-1} \mathbf{y} \tag{22}$$

where $\mathbf{x}$ represents the training sample input, $\hat{\mathbf{x}}$ represents the testing sample input, $\lambda=\sigma^2/\tau^2$ is the variance ratio, and the value is between 0 and 1, for the global Gaussian process usually accounts for a larger proportion in the prediction. When $\lambda = 0$, CGPR is reduced to the basic GPR model. $\mathbf{G}$ and $\mathbf{L}$ are the covariance matrices of the training set for the global and local Gaussian processes, respectively. $\mathbf{y}$ is the output matrix of the training set, and $\mathbf{I}_n$ is the n-dimensional identity matrix.

In addition, the 95% confidence interval is calculated by (23), which quantifies the uncertainty in the prediction process.

$$CI_{95\%} = \left[\hat{y}(\mathbf{x}) - 1.96\sqrt{\mathrm{cov}(\hat{y})} \ , \ \hat{y}(\mathbf{x}) + 1.96\sqrt{\mathrm{cov}(\hat{y})}\right] \tag{23}$$

In this study, the mean function of the global Gaussian process is set to zero. We compare common kernel functions and find that linear kernel functions yield the most effective results and that a strong linear relationship exists between HIs and SOH. Consequently, we decided to employ linear functions as the kernel functions throughout our paper. The kernel function of the local Gaussian process is set as the Matern covariance function, which is used to describe the local variation of the relationship between HIs and SOH. The kernel function of the constructed CGPR model is represented by (24). The CGPR model is used to fuse selected HI and estimate battery SOH.

$$\begin{cases} g(\mathbf{x}, \mathbf{x}') = a\mathbf{x}' + b \\ l(\mathbf{x}, \mathbf{x}') = \sigma_{f2}^2 \left(1 + \dfrac{\sqrt{3}|\mathbf{x} - \mathbf{x}'|}{\lambda_2}\right) \exp\left(-\dfrac{\sqrt{3}|\mathbf{x} - \mathbf{x}'|}{\lambda_3}\right) \end{cases} \tag{24}$$

IV. RESULTS AND DISCUSSION

In this section, the aging data of Cells 1-7 are used to verify the SOH estimation framework. To verify the proposed framework in partial charging cases, three representative partial charging cases are defined for the case study based on charging behavior reports of EV users.

Many studies on partial charging, including [21-23], utilize a classification approach. The goal of categorizing different instances of partial charging is to reduce the amount of required samples in the training set. This arises from the numerous variations in user charging behavior, whether in terms of SOC or voltage intervals. For data-driven methods to function effectively, the input parameters for both training and testing must carry the same meaning. In scenarios involving partial charging, it is equivalent to requiring an identical SOC range. In real-world applications, it is impractical to pre-establish a corresponding training set for every individual charging session of a user. To address this difficulty, the method proposed in this study involves defining several commonly observed SOC charging intervals. All charging behaviors that encompass these intervals are then grouped into one case. In addition, it is important to note that the algorithm proposed in this study does not rely on the classification of partial charging cases. For example, parameters such as SOC or voltage intervals can serve as model inputs, enabling the algorithm to classify the case automatically as needed. The influence of partial charging on HI extraction is analyzed. Finally, the proposed framework is verified in the three representative partial charging cases.

A. Definition of Representative Partial Charging Cases Based on User Behavior Reports

The initial SOC and final SOC are the key states describing the shapes of each partial charging curve [25]. In practice, the initial and final SOCs are random and uncertain since the decisions of EV users to start or end their charge actions are based on various factors. These factors include remaining SOC, time limits, infrastructure locations, driver anxiety ratios, peak hours, charging types, driving experience, electricity tariff, etc. [38]. Among these factors, EV users usually start or terminate the charging process based on two main factors, i.e., the range anxiety level and the charge time adequacy [39]. The range anxiety level refers to an EV user's fear that the battery does not have sufficient charge to reach the destination, which is related to the initial SOC. Chaudhari *et al*. [38] pointed out that the user's initial SOC of charging is affected by range anxiety. Higher range anxiety usually leads to a higher initial charging SOC. The charge time adequacy indicates whether an EV user has adequate time to fully charge the battery, which is related to the final SOC. This depends on the user's availability and charging type. For instance, an EV user may decide to fully charge the battery when he/she has adequate time at the

TABLE II
CLASSIFICATION OF NINE PARTIAL CHARGING CASES

Range Anxiety Level (Initial SOC)	**Charge Time Adequacy** (Final SOC)		
	Short	Medium	Long
Low	Case I	Case II	Case III
Medium	**Case IV**	Case V	**Case VI**
High	Case VII	Case VIII	**Case IX**

destination (home or workplace) or to fast charge halfway, subject to limited waiting time.

In this paper, the classification of nine partial charging cases is shown in Table II. Range anxiety level, which determines the initial SOC, is divided into three levels, namely low, medium, and high. Charge time adequacy, which determines the final SOC, is divided into short, medium, and long levels, respectively. For example, Case I indicates that the users may experience a low-range anxiety level and a short charge time. Here, three representative partial charging cases, Case IV, Case VI, and Case IX, are selected according to the statistical analysis of vehicle charging data in [40-42] to validate the feasibility of this method.

Case IV: EV users have medium-range anxiety but short charge time. This group of EV users corresponds to that in [40], and they can endure battery SOC discharged in a lower SOC but do not have sufficient time for a full charge. We select the initial SOC lower than 40% and the final SOC higher than 60% for this case, which means that the SOC data for SOH estimation is between 40% to 60%.

Case VI: EV users suffer from medium-range anxiety and long charge time. This group of EV users can endure battery SOC discharged in a lower SOC and have sufficient time for a full charge, corresponding to the users in [41]. Therefore, we select the initial SOC lower than 40% and the final SOC higher than 85% as the unique SOC range for this case.

Case IX: EV users suffer from high-level range anxiety and long charging time. This group of EV users cannot tolerate low battery SOC, which corresponds to the users in [42]. We select a high initial SOC of 70% to represent the charging behavior of people with the highest range anxiety.

Case 0 (Base Case): The standard full charge case with SOC ranging from 10% to 85% is considered. This case serves as the control group or benchmark.

The definition of the three representative partial charging cases is summarized in Table III. All cases are listed below with the range anxiety, charge time adequacy, and reference voltage segment.

TABLE III
THREE REPRESENTATIVE PARTIAL CHARGING CASES

Category	Initial SOC	Range anxiety	Final SOC	Charge Time	Ref. voltage segment
Case IV	⩽40%	Medium	≥60%	Short	3.8V~3.95V
Case VI	⩽40%	Medium	≥85%	Long	3.8V~4.2V
Case IX	⩽70%	High	≥85%	Long	4.0V~4.2V
Case 0	⩽10%	Low	≥85%	Long	3.2V~4.2V

B. Influence of Partial Charging on HI Extraction

In partial charging cases, lack of specific data could affect the extraction of HIs and even lead to the missing of HIs. This section assesses the feasibility of the estimation model through three more common cases. In the process of extracting HIs, we observed that the three cases with different partial charging curves enable the extraction of various and effective features. For instance, the V_2 feature extracted from the IC curve is present in Cases VI and IV but absent from Case IX. This discrepancy stems from the incomplete, partial charging curve in Case IX, highlighting the limitation associated with partial charging curves. Our study intends to develop an applicable SOH estimation model based on partial charging curves. Therefore, even if there are two or even more features from IC curves, CC and E-V curves with high correlation with each other are used in a certain case. The purpose is to ensure that at least one feature can be identified in other cases. This strategy is crucial to prevent scenarios where no valid features are extractable, enhancing the overall generalizability of the proposed estimation methods.

Specifically, when the initial SOC is high, and the initial voltage is beyond a certain peak voltage value, the fitted peak

TABLE IV
RESULTS OF CORRELATION ANALYSIS

	IC Curve Based HIs											
HI type	Peak height				Peak position				Peak area			
HI name	H_1	H_2	H_3	H_4	V_1	V_2	V_3	V_4	A_1	A_2	A_3	A_4
Case 0	**0.9771**	0.9731	-0.1951	0.8628	-0.9872	**-0.9886**	0.6131	0.4534	0.9659	0.9931	**0.9982**	0.4728
Case VI	~	0.8371	0.1858	0.8076	~	**-0.9662**	0.0274	-0.0374	~	0.8216	**0.9037**	0.0976
Case IX	~	~	~	0.5888	~	~	~	-0.7341	~	~	~	**0.9438**
Case IV	~	0.8751	-0.1331	~	~	**0.9629**	-0.7367	~	~	0.0573	**-0.9724**	~

	CC Curve Based HIs						**E-V Curve Based HIs**					
HI type	Charged capacity at IC peak voltage				Total charging capacity	Voltage at a fixed time	Charged energy at IC peak voltage				Total charging energy	Average charging power
HI name	Q_{V_1}	Q_{V_2}	Q_{V_3}	Q_{V_4}	$Q_{V_{end}}$	V_Q	E_{V_1}	E_{V_2}	E_{V_3}	E_{V_4}	E_{end}	P
Case 0	0.9374	0.9868	0.9959	**0.9981**	0.9969	**-0.9904**	0.9446	0.9923	0.9971	**0.9985**	0.9973	**-0.9857**
Case VI	~	0.2387	0.9907	**0.9971**	0.9943	**-0.9938**	~	0.2829	0.9908	**0.9973**	0.9939	**-0.9899**
Case IX	~	~	~	**0.9941**	0.9874	**-0.9846**	~	~	~	**0.9935**	0.9867	**-0.9816**
Case IV	~	-0.0477	0.9901	~	**0.9935**	**-0.9873**	~	0.0011	0.9903	~	**0.9977**	**-0.9873**

Bolded fonts: The final selected HIs. ~ : Not applicable.

point in the IC curve will diminish and finally disappear until the initial voltage reaches the next peak position. For example, in Case VI, the initial SOC is 40%. The initial voltage is about 3.8 V, which exceeds the position of the second peak, V_2 = 3.7 V. As a result, the first peak point will disappear, missing the first peak HIs of H_1, Q_{V1}, and E_{V1}, and also affect the value of the second peak HIs of H_2, Q_{V2}, and E_{V2}. Therefore, the total number of peaks, n, needs to be adjusted in partial charging curves. A proper number of peaks, n, can be set for a partial charging case by referring to the full IC curve. For example, the initial voltage of Case VI is within the segment of the second peak. Consequently, the number of peaks n in Case VI is set to 3. Similarly, the numbers of peaks n in Cases IX and IV are set to 1 and 2, respectively.

Furthermore, various HIs in partial charging cases are analyzed by the Pearson correlation method, and the combination of HIs in each case is selected. The Pearson correlation coefficient is calculated by

$$r = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(SOH_i - \bar{SOH})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{N}(SOH_i - \bar{SOH})^2}} \tag{25}$$

where r is the Pearson correlation coefficient, and the value is within [-1, 1]. A coefficient with an absolute value closer to 1 indicates a stronger correlation between corresponding HI and SOH. i is the number of cycles, N is the total number of charging cycles, and X is the HIs to be evaluated.

The Pearson correlation coefficients for each partial charging case are given in Table IV. As a benchmark, Case 0, or the Full-Charge Case, is analyzed in Table IV. In Case 0, IC-based HIs show a high correlation with SOH. Although there are still some high correlation HIs, it is difficult to guarantee the accuracy and robustness of SOH estimation using only these HIs. Fortunately, the HIs based on the CC charging curve and *E-V* curve maintain a high correlation in partial charging cases. In particular, the capacity or the energy HIs based on the peak position tend to show the highest correlation, although some HIs based on the peak position will also disappear in partial charging cases. In addition, in different partial charging cases, HIs with the same name may have various physical meanings. For example, though the peak position remains unchanged in different cases, the charging capacity represented by Q_{V1} varies due to a change in the initial SOC.

During the selection process, HIs with a correlation coefficient below 0.9 are excluded first, and more types of HIs are tested from the remaining HIs since HIs of the same type may contain redundant information. The selected HIs for each case are highlighted in bold font in Table IV.

C. Multi-Case SOH Estimation Results Under Constant Current Condition

Based on the data of Cell 1 and Cell 2, the feasibility of SOH estimation in different partial charging cases is verified in this section. A widely used verification method is to divide the data into the training set and the testing set based on the number of cycles. The advantage of this method is that both the training and testing sets are obtained from the same battery. This can eliminate the interference caused by cell inconsistency in

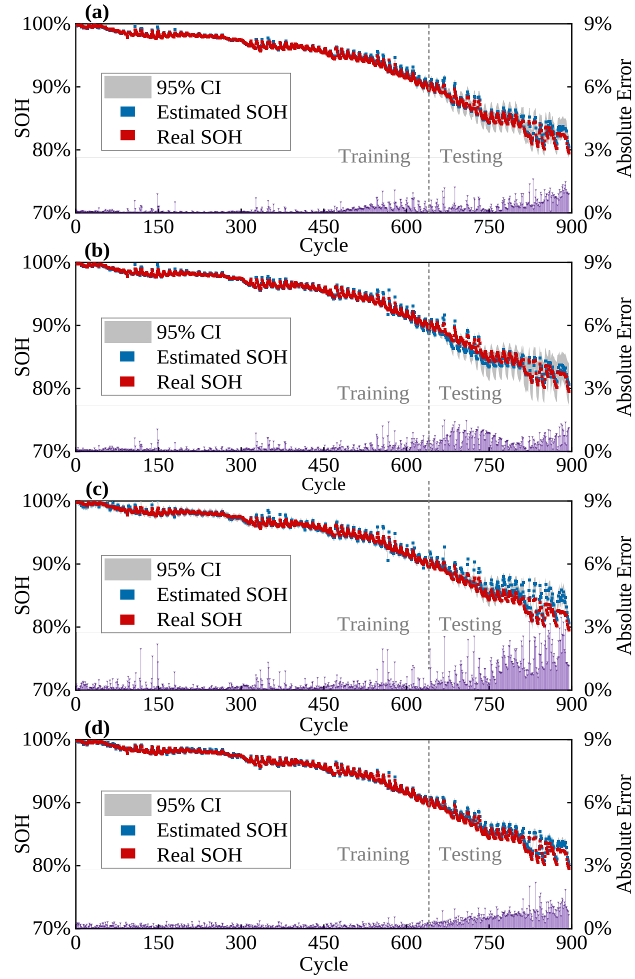

Fig. 9. SOH estimation results of Cell 1 in different Case. (a) Case 0, (b) Case VI, (c) Case IX, (d) Case IV.

TABLE V
MAES OF THE TESTING ERROR OF MULTI-CASE SOH ESTIMATION FOR CELL 1 AND CELL 2

Cell No.	Case 0	Case VI	Case IX	Case IV
Cell 1	0.44%	0.58%	0.91%	0.65%
Cell 2	0.67%	0.96%	0.85%	1.08%

manufacturing processes. Thus, the feasibility of HI selection and estimation methods can be directly verified. Accordingly, model training is based on the cycles when SOH is between 90% and 100%, and model testing is based on the cycles when SOH is between 80% and 90%.

The estimation results of Cells 1 and 2 are shown in Fig. 9. Figs. 9(a)-(d) on the left are the results of Cell 1 for different charging cases, and others are the results of Cell 2. The estimation results of Case 0 are shown in Figs. 9(a) as a comparison. The MAE of the testing error is presented in Table V.

Overall, the estimated SOH is close to the real SOH in all cases, and the MAE is within 1%. SOH Estimating results of all cycles, including training cycles and testing cycles, are shown in the figure. Among them, the error is minor in the training cycles and almost completely fitted, indicating that the selected HIs can adequately reflect the battery aging. Although the error of the test cycles rises, most errors are below 3%. This result shows that the proposed model has good generalization ability. Even in the cycles without training samples, the model can still describe the local variation of SOH.

In addition, the confidence interval of the testing phase is significantly wider than that of the training phase, which shows that the model can express the uncertainty well.

Comparing the SOH estimation results in different charging cases, the error in Case 0 is the smallest, and the error in Case IX is the most significant. This is because the available data in Case II is the fewest and only contains the voltage segment of Peak 4. Table V shows that the correlation between the HIs extracted from Peak 4 and SOH is lower than those from other peaks. The main peak of the IC curve can precisely reflect battery aging best [43]. For LCO cells in this study, the main peak is Peak 2, according to Fig. 8(a). The MAE in Cases VI and III are close since the data of both cases contain the voltage segment with the main peak. It is noted that the models of Cases IV and IV can also be used in Case VI since the HIs required by Cases IX and IV can be extracted from that in Case VI. However, the SOH estimation error will increase due to the further lack of data. Especially in Case IX, the error has almost doubled compared to Case VI.

D. Verification Under Different Operating Conditions

Based on the data of Cell 1 to Cell 4, the multi-case SOH estimation effect under different operating conditions is analyzed in this section.

In practical applications, the model training needs to be carried out based on the historical data of other batteries since the exact SOH data of the battery itself is unavailable. Historical data of the same type of battery are commonly used for model training to ensure the similarity of electrochemical characteristics. However, the consistency between the discharge conditions of the training battery and the testing battery is not ensured since the actual discharge condition of the battery in EVs is uncertain. Therefore, it is necessary to study the robustness of the SOH estimation method under different operating conditions. Accordingly, the CGPR model trained based on all the data of Cells 1 and 2 is used to directly estimate the SOH of Cells 3 and 4, respectively. Cells 1 and 2 for training are under the constant current discharge condition, and Cells 3 and 4 for testing are under the WLTC simulated discharge condition.

The estimation results of Cells 3 and 4 are shown in Fig. 10. Figs. 10(a), (b), (c), and (d) on the left are the results of Cell 3 for different charging cases, and the other subfigures show the results of Cell 4. The MAE is integrated into Table VI. It can be seen that the estimated SOH is close to the real value in all charging cases. The estimation error is less than 3% in most cycles, and the MAE is within 1.5%. These errors mainly come from the differences in manufacturing processes and discharge conditions. It shows that the proposed method is strongly robust to various discharge conditions. More importantly, the comparison of estimation results in different partial charging cases shows the same pattern as in Section IV-C.

The results show that the SOH estimation method proposed is robust to different operating conditions. Under the same charging conditions, the difference in discharge conditions has

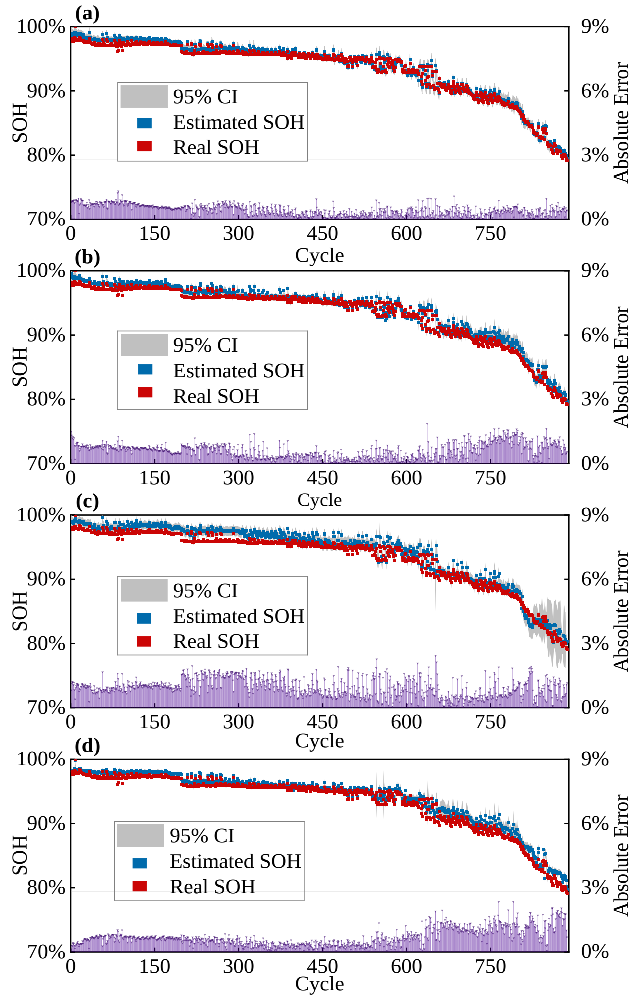

Fig. 10. SOH estimation results of Cell 3 in different Case. (a) Case 0, (b) Case VI, (c) Case IX, (d) Case IV.

TABLE VI
MAES OF THE TESTING ERROR OF MULTI-CASE SOH ESTIMATION FOR CELL 3 AND CELL 4

Cell Number	Case 0	Case VI	Case IX	Case IV
Cell 3	0.63%	0.89%	1.42%	1.02%
Cell 4	0.84%	0.95%	1.25%	0.96%

a limited impact on the proposed SOH estimation method. Therefore, the proposed method is useful for real-world applications. Based on the data of batteries under the CC discharging condition in the laboratory, the accurate estimation of SOH of EV batteries under complex discharge conditions can be realized.

E. Verification on the Oxford Open Dataset

Based on the data of Cell 5 to Cell 7, the feasibility of multi-case SOH estimation for different battery types is verified in this subsection.

Specifically, the model is trained based on Cell 5, and Cells 6 and 7 data are used for testing. Since the complete IC curve of Cell 5 to Cell 7 has only three peaks, the number of peaks n in the HI extraction process needs to be adjusted. According to

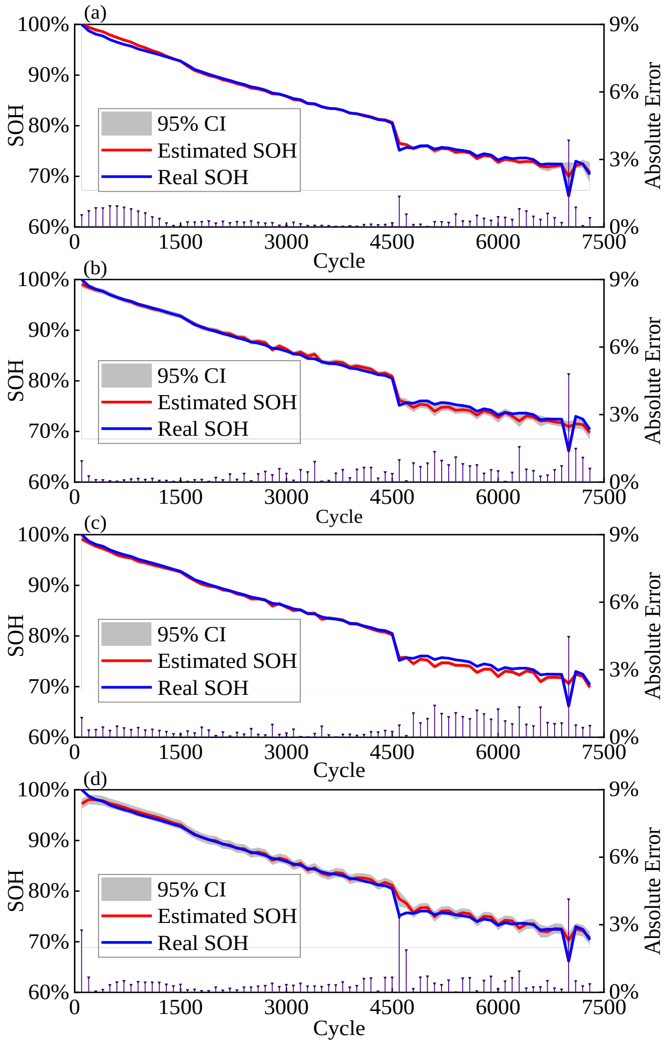

Fig. 11. SOH estimation results of Cell 6 in (a) Case 0, (b) Case VI, (c) Case IX, and (d) Case IV.

TABLE VII
MAES OF THE TESTING ERROR OF MULTI-CASE SOH ESTIMATION FOR CELL 6 AND CELL 7

Cell Number	Case 0	Case VI	Case IX	Case IV
Cell 6	0.57%	0.75%	0.82%	0.75%
Cell 7	0.48%	0.51%	0.71%	0.63%

the complete IC curve, the number of peaks n is set to 2, 1, and 1 for the three partial charging cases.

The multi-case SOH estimation results of Cells 6 and 7 are shown in Fig. 11, and the MAE of testing data is integrated into Table VII. Overall, the cells in the Oxford battery degradation dataset have good consistency, so the model trained from Cell 5 can accurately estimate the SOH of Cells 6 and 7, and the average error is within 0.6%. Note that the SOH estimation error under Case 0 is still the smallest, and the error in Case VIIII is the largest.

In the three representative partial charging cases, the proposed method can accurately estimate the SOH of these cells. It is shown that the proposed method can be easily applied to different types of batteries by only properly adjusting the number of peaks n.

F. Limitations and Future Work

In our present study, we have successfully extracted many valid HIs from the IC curve, CC charging curve, and E-V curve. The estimation performance of the model across different scenarios, including scenarios with high initial SOC charging (i.e., Case IX) and other conditions, are verified by utilizing three cases that span a wide range of charging segments. These cases affirm the validity of our approach in the majority of cases. Moreover, we choose SOC as a more relevant visualization parameter for driver users, providing a clear advantage over the majority of existing studies in this research field. These factors highlight the strengths of the proposed method compared to the aforementioned papers and other existing literature.

However, it should be pointed out that the proposed method has limitations. One limitation is that it cannot be applicable in mixed scenarios, such as car rental scenarios with different user behaviors. This paper assumes that the car is owned by a private family/driver with certain fixed driving patterns during its life span. Another limitation is that due to the limited literature on driver behavior statistics, only three sub-models are selected for SOH estimation based on three representative cases, which can cover the whole scenario. These are shortcoming that needs to be addressed in our future work.

V. CONCLUSION

Accurate estimation of SOH is a prerequisite for the safe and reliable operation of LIBs. Previous SOH methods mainly rely on the full charging cycle data for SOH estimation.

Considering the lack of aging information in partial charging data, this study proposes a curve-fitting method to extract HIs. More than 20 HIs are extracted from charging voltage data, including novel HIs based on the energy-voltage curve. Furthermore, the composite Gaussian process regression-based data-driven method is proposed to ensure accurate SOH estimation. To verify the proposed framework in partial charging cases, three representative partial charging cases are defined for the case study based on charging behavior reports of EV users. Finally, the proposed method is verified under WLTC operation conditions and different battery types by our laboratory testing data and Oxford open dataset. The results show that the proposed framework demonstrates the ability to estimate SOH accurately and strong robustness to representative partial charging cases. In our future research, a unified SOH estimation framework for random partial charging cases will be investigated.

REFERENCES

[1] Z. Wei, J. Zhao, H. He, G. Ding, H. Cui, and L. Liu, "Future smart battery and management: Advanced sensing from external to embedded multi-dimensional measurement," *J. Power Sources*, vol. 489, p. 229462, Mar. 2021, doi: 10.1016/j.jpowsour.2021.229462.

[2] S. Khaleghi Rahimian and Y. Tang, "A practical data driven battery state of health estimation for electric vehicles," *IEEE Trans. Ind. Electron.*, vol. 70, no.1, pp. 1973–1982, Feb. 2023, doi: 10.1109/TIE.2022.3165295.

[3] Y. Li, Z. Wei, B. Xiong, and D. M. Vilathgamuwa, "Adaptive ensemble-based electrochemical–thermal degradation state estimation

of lithium-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 69, no. 7, pp. 6984–6996, Jul. 2022, doi: 10.1109/TIE.2021.3095815.

[4] X. Cong, C. Zhang, J. Jiang, W. Zhang, and Y. Jiang, "A hybrid method for the prediction of the remaining useful life of lithium-ion batteries with accelerated capacity degradation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12775–12785, Nov. 2020, doi: 10.1109/TVT.2020.3024019.

[5] M. Cacciato, G. Nobile, G. Scarcella, and G. Scelba, "Real-time model-based estimation of SOC and SOH for energy storage systems," *IEEE Trans. Power Electron.*, vol. 32, no. 1, pp. 794–803, Jan. 2017, doi: 10.1109/TPEL.2016.2535321.

[6] W. Vermeer, G. R. Chandra Mouli, and P. Bauer, "A comprehensive review on the characteristics and modeling of lithium-ion battery aging," *IEEE Trans. Transp. Electrif.*, vol. 8, no. 2, pp. 2205–2232, Jun. 2022, doi: 10.1109/TTE.2021.3138357.

[7] X. Shu, J. Shen, G. Li, Y. Zhang, Z. Chen, and Y. Liu, "A flexible state-of-health prediction scheme for lithium-ion battery packs with long short-term memory network and transfer learning," *IEEE Trans. Transp. Electrif.*, vol. 7, no. 4, pp. 2238–2248, Dec. 2021, doi: 10.1109/TTE.2021.3074638.

[8] G. Dong, W. Han, and Y. Wang, "Dynamic Bayesian network-based lithium-ion battery health prognosis for electric vehicles," *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, pp. 10949–10958, Nov. 2021, doi: 10.1109/TIE.2020.3034855.

[9] L. Cai, J. Meng, D.-I. Stroe, J. Peng, G. Luo, and R. Teodorescu, "Multiobjective optimization of data-driven model for lithium-ion battery SOH estimation with short-term feature," *IEEE Trans. Power Electron.*, vol. 35, no. 11, pp. 11855–11864, Nov. 2020, doi: 10.1109/TPEL.2020.2987383.

[10] 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, doi: 10.1109/TSMC.2013.2296276.

[11] K. Liu, Y. Li, X. Hu, M. Lucu, and W. D. Widanage, "Gaussian process regression with automatic relevance determination kernel for calendar aging prediction of lithium-ion batteries," *IEEE Trans. Ind. Inform.*, vol. 16, no. 6, pp. 3767–3777, Jun. 2020, doi: 10.1109/TII.2019.2941747.

[12] C. She, Y. Li, C. Zou, T. Wik, Z. Wang, and F. Sun, "Offline and online blended machine learning for lithium-ion battery health state estimation," *IEEE Trans. Transp. Electrif.*, vol. 8, no. 2, pp. 1604–1618, Jun. 2022, doi: 10.1109/TTE.2021.3129479.

[13] S. Jiang and Z. Song, "A review on the state of health estimation methods of lead-acid batteries," *J. Power Sources*, vol. 517, p. 230710, Jan. 2022, doi: 10.1016/j.jpowsour.2021.230710.

[14] 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, doi: 10.1109/TPEL.2020.2978493.

[15] A. El Mejdoubi, A. Oukaour, H. Chaoui, H. Gualous, J. Sabor, and Y. Slamani, "State-of-charge and state-of-health lithium-ion batteries' diagnosis according to surface temperature variation," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2391–2402, Apr. 2016, doi: 10.1109/TIE.2015.2509916.

[16] 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, doi: 10.1016/j.jpowsour.2018.03.015.

[17] J. He, Z. Wei, X. Bian, and F. Yan, "State-of-health estimation of lithium-ion batteries using incremental capacity analysis based on voltage–capacity model," *IEEE Trans. Transp. Electrif.*, vol. 6, no. 2, pp. 417–426, Jun. 2020, doi: 10.1109/TTE.2020.2994543.

[18] S. Zhang, X. Guo, X. Dou, and X. Zhang, "A rapid online calculation method for state of health of lithium-ion battery based on Coulomb counting method and differential voltage analysis," *J. Power Sources*, vol. 479, p. 228740, Dec. 2020, doi: 10.1016/j.jpowsour.2020.228740.

[19] Z. Wang, S. Zeng, J. Guo, and T. Qin, "State of health estimation of lithium-ion batteries based on the constant voltage charging curve," *Energy*, vol. 167, pp. 661–669, Jan. 2019, doi: 10.1016/j.energy.2018.11.008.

[20] A. Alsabbagh and C. Ma, "Distributed charging management of electric vehicles considering different customer behaviors," *IEEE Trans. Ind. Inform.*, vol. 16, no. 8, pp. 5119–5127, Aug. 2020, doi: 10.1109/TII.2019.2952254.

[21] Z. Wei, H. Ruan, Y. Li, J. Li, C. Zhang, and H. He, "Multistage state of health estimation of lithium-ion battery with high tolerance to heavily partial charging," *IEEE Trans. Power Electron.*, vol. 37, no. 6, p. 11, 2022.

[22] 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, doi: 10.1109/TPEL.2021.3104723.

[23] J. He, S. Meng, X. Li, and F. Yan, "Partial charging-based health feature extraction and state of health estimation of lithium-ion batteries," *IEEE J. Emerg. Sel. Topic Power Electron.*, vol. 11, no. 1, pp. 166–174, Feb. 2023, doi: 10.1109/JESTPE.2022.3143831.

[24] D. I. Stroe, V. Knap, and E. Schaltz, "State-of-health estimation of lithium-ion batteries based on partial charging voltage profiles," *ECS Trans.*, vol. 85, no. 13, pp. 379–386, Jun. 2018, doi: 10.1149/08513.0379ecst.

[25] Y. Wang, E. Yao, and L. Pan, "Electric vehicle drivers' charging behavior analysis considering heterogeneity and satisfaction," *J. Clean. Prod.*, vol. 286, p. 124982, Mar. 2021, doi: 10.1016/j.jclepro.2020.124982.

[26] V.-H. Duong, H. A. Bastawrous, K. Lim, K. W. See, P. Zhang, and S. X. Dou, "Online state of charge and model parameters estimation of the LiFePO4 battery in electric vehicles using multiple adaptive forgetting factors recursive least-squares," *J. Power Sources*, vol. 296, pp. 215–224, Nov. 2015, doi: 10.1016/j.jpowsour.2015.07.041.

[27] X. Jia, C. Zhang, Y. Li, C. Zou, L. Y. Wang, and X. Cai, "Knee-point-conscious battery aging trajectory prediction based on physics-guided machine learning," *IEEE Trans. Transp. Electrif.*, vol. 10, no. 1, pp. 1056–1069, Mar. 2024, doi: 10.1109/TTE.2023.3266386.

[28] C. Birkl, "Diagnosis and prognosis of degradation in lithium-ion batteries," http://purl.org/dc/dcmitype/Text, University of Oxford, 2017. Accessed: Oct. 28, 2023. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:7d8ccb9c-1469-4209-9995-5871fc908b54

[29] X. Li, J. Jiang, L. Y. Wang, D. Chen, Y. Zhang, and C. Zhang, "A capacity model based on charging process for state of health estimation of lithium ion batteries," *Appl. Energy*, vol. 177, pp. 537–543, Sep. 2016, doi: 10.1016/j.apenergy.2016.05.109.

[30] C. Zhang *et al.*, "An adaptive battery capacity estimation method suitable for random charging voltage range in electric vehicles," *IEEE Trans. Ind. Electron.*, vol. 69, no. 9, pp. 9121–9132, Sep. 2022, doi: 10.1109/TIE.2021.3111585.

[31] Y. Yang, S. Chen, T. Chen, and L. Huang, "State of health assessment of lithium-ion batteries based on deep Gaussian process regression considering heterogeneous features," *J. Energy Storage*, vol. 61, p. 106797, May 2023, doi: 10.1016/j.est.2023.106797.

[32] X. Hu, Y. Che, X. Lin, and S. Onori, "Battery health prediction using fusion-based feature selection and machine learning," *IEEE Trans. Transp. Electrification*, vol. 7, no. 2, pp. 382–398, Jun. 2021, doi: 10.1109/TTE.2020.3017090.

[33] Q. Wang, Z. Wang, L. Zhang, P. Liu, and Z. Zhang, "A novel consistency evaluation method for series-connected battery systems based on real-world operation data," *IEEE Trans. Transp. Electrif.*, vol. 7, no. 2, pp. 437–451, Jun. 2021, doi: 10.1109/TTE.2020.3018143.

[34] W. Diao, I. H. Naqvi, and M. Pecht, "Early detection of anomalous degradation behavior in lithium-ion batteries," *J. Energy Storage*, vol. 32, p. 101710, Dec. 2020, doi: 10.1016/j.est.2020.101710.

[35] J. N. Fuhg, M. Marino, and N. Bouklas, "Local approximate Gaussian process regression for data-driven constitutive models: development and comparison with neural networks," *Comput. Methods Appl. Mech. Eng.*, vol. 388, p. 114217, Jan. 2022, doi: 10.1016/j.cma.2021.114217.

[36] S. Ba and V. R. Joseph, "Composite Gaussian process models for emulating expensive functions," *Ann. Appl. Stat.*, vol. 6, no. 4, Dec. 2012, doi: 10.1214/12-AOAS570.

[37] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. in Adaptive computation and machine learning. Cambridge, Mass: MIT Press, 2006.

[38] K. Chaudhari, N. K. Kandasamy, A. Krishnan, A. Ukil, and H. B. Gooi, "Agent-based aggregated behavior modeling for electric vehicle charging load," *IEEE Trans. Ind. Inform.*, vol. 15, no. 2, pp. 856–868, Feb. 2019, doi: 10.1109/TII.2018.2823321.

[39] J. Yang *et al.*, "Charging demand analysis framework for electric vehicles considering the bounded rationality behavior of users," *Int. J.*

Electr. Power Energy Syst., vol. 119, p. 105952, Jul. 2020, doi: 10.1016/j.ijepes.2020.105952.

[40] P. Fieltsch, H. Flämig, and K. Rosenberger, "Analysis of charging behavior when using battery electric vehicles in commercial transport," *Transp. Res. Procedia*, vol. 46, pp. 181–188, 2020, doi: 10.1016/j.trpro.2020.03.179.

[41] C. Siddique, F. Afifah, Z. Guo, and Y. Zhou, "Data mining of plug-in electric vehicles charging behavior using supply-side data," *Energy Policy*, vol. 161, p. 112710, Feb. 2022, doi: 10.1016/j.enpol.2021.112710.

[42] Q. Xing, Z. Chen, Z. Zhang, R. Wang, and T. Zhang, "Modelling driving and charging behaviours of electric vehicles using a data-driven approach combined with behavioural economics theory," *J. Clean. Prod.*, vol. 324, p. 129243, Nov. 2021, doi: 10.1016/j.jclepro.2021.129243.

[43] B. Ospina Agudelo, W. Zamboni, and E. Monmasson, "Application domain extension of incremental capacity-based battery SoH indicators," *Energy*, vol. 234, p. 121224, Nov. 2021, doi: 10.1016/j.energy.2021.121224.

Jinrui Tang (Member, IEEE) received the B.S. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2009, and the Ph.D. degree in electrical engineering from the School of Electrical and Electronics Engineering, Huazhong University of Science and Technology, Wuhan, in 2014. He is currently an Associate Professor with the School of Automation, Wuhan University of Technology, Wuhan. His research interests include power system protection, distribution systems, and energy storage systems.

Yang Li (Senior Member, IEEE) received the B.E. degree in electrical engineering from Wuhan University, Wuhan, China, in 2007 and the Ph.D. degree in power engineering from Nanyang Technological University (NTU), Singapore, in 2015. From 2015 to 2018, he was a Research Fellow with the Energy Research Institute, NTU, and the School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Australia, respectively. He is currently a Researcher at the Chalmers University of Technology, Gothenburg, Sweden. His research interests include modeling and control of renewable generation systems, energy storage systems, and power electronic converters.

Dr. Li serves as an Associate Editor for several journals, such as IEEE Transactions on Industrial Electronics, IEEE Transactions on Transportation Electrification, and IEEE Transactions on Energy Conversion.

Binyu Xiong (Member, IEEE) received the M.Sc. and Ph.D. degrees in power engineering from Nanyang Technological University, Singapore, in 2011 and 2016, respectively. He is currently a professor at the School of Automation, Wuhan University of Technology, Wuhan, China. His research interests include electrical and thermal modeling of batteries, battery system control, large-scale energy storage systems, and renewable energy generation.

Shaojin Wang received the M.S. degree in electrical engineering from Hubei University of Technology in 2023. He is currently pursuing the Ph.D. degree in the School of Automation, Wuhan University of Technology, Wuhan, China. His research interests include battery modeling.

Jinxuan Pan received the M.S. degree in electrical engineering from Wuhan University of Technology, Wuhan, China, in 2023. He is now working at State Grid Yichun Power Supply Company. His research interests include state estimation and fault diagnosis of lithium-ion batteries in power systems and electrical vehicles.

Xiangjun Li (Senior Member, IEEE) received the Ph.D. degree in electrical and electronic engineering from the Kitami Institute of Technology (KIT), Kitami, Japan, in March 2006. He is currently the Director of the Energy Storage System Integration and Configuration Technology Research Lab. Energy Storage and Electrotechnics Dept. at the China Electric Power Research Institute since Jun. 2019. He is a Fellow of IET and Fellow of CAA, chartered engineer of the UK, and is one of the senior members of CSEE, CAS, and CES, etc.

Qihong Chen (Member, IEEE) received the Ph.D. degree in control science and engineering from Southeast University, Nanjing, China, in 2003. He is currently a Professor with the Wuhan University of Technology, Wuhan, China. His current research interests include grid-tied inverters and predictive control.

Wang Peng (Fellow, IEEE) received the B.Sc. degree in electronic engineering from Xian Jiaotong University, Xi'an, China, in 1978, the M.Sc. degree from Taiyuan University of Technology, Taiyuan, China, in 1987, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Saskatchewan, Saskatoon, SK, Canada, in 1995 and 1998, respectively. Currently, he is a full Professor with the School of Electrical and Electronic Engineering at Nanyang Technological University, Singapore. He is an Associate Editor or Guest Editor-in-Chief of IEEE Transactions on Smart Grid, IEEE Transactions on Power Delivery, Journal of Modern Power Systems and Clean Energy, CSEE Journal of Power and Energy Systems, and so on. His current research interests include power system planning and operation, renewable energy planning, solar/electricity conversion systems and power system reliability analysis.