Energy 262 (2023) 125514

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

State of health estimation for lithium-ion batteries using geometric impedance spectrum features and recurrent Gaussian process regression

Yong Zhou [a], Guangzhong Dong [b], Qianqian Tan [a], Xueyuan Han [a], Chunlin Chen [a], Jingwen Wei [a,*]

[a] School of Management and Engineering, Nanjing University, Nanjing 210093, China
[b] School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, Guangdong 518055, China

ARTICLE INFO

Keywords:
Lithium-ion battery
State of health
Gaussian process regression
Electric vehicles
Electrochemical impedance spectroscopy

ABSTRACT

Due to lithium-ion batteries' complex behaviors, accurate estimation of state-of-health is still a critical challenge in battery systems' prognosis and health management. Most existing efforts in battery health prognosis focus on feature engineering using low-frequency sampled time-domain response. These efforts may not completely reflect the battery health status in automotive applications due to information missing in the high or medium frequency range. This paper proposes a data-driven state-of-health estimation method using high and medium frequency range impedance spectroscopy data. First, battery health indicators are extracted from electrochemical impedance spectroscopy data. It is found that the Nyquist diagram shows semicircle characteristics at high and medium frequency ranges. The center and radius of this circle show high dependence on battery health. Then, a recurrent Gaussian process regression with a one-step delay feedback loop is designed to provide a smooth and accurate battery state-of-health estimate. Finally, the proposed health indicators and state-of-health estimators are validated using experimental data on different cells. The results demonstrate the high accuracy and robustness of the proposed health indicators and state-of-health estimator, suggesting a 1.12% estimation error. This study shows the prospect of health prognosis using robust geometric impedance spectrum indicators in energy storage systems.

1. Introduction

1.1. Motivation and challenges

Nowadays, the rapidly developing electric vehicles have become one of the best candidates for the transformation of transportation sectors, to face the challenges of environmental pollution and carbon emissions caused by excessive consumption of fossil fuels [1,2]. State-of-the-art LIBs (e.g., lithium iron phosphate, $LiFePO_4$) offer the best trade-off between power/energy density and costs for energy storage in electric vehicles. They have been attracting special attention due to low self-discharge and long lifetime [3]. However, the performance of LIBs will degrade due to repeated charging/discharging cycles. When the quality characteristics deteriorate to some unacceptable level, the battery system will be deemed failed and it will even lead to catastrophic loss. Therefore, accurately estimating battery SOH becomes a critical task in PHM of energy storage systems [4].

The battery SOH is defined as the ratio between its current total and the rated capacity [5]. Estimating battery SOH is a challenging task since battery charging/discharge processes involve complex and coupling electrochemical side reactions inside the battery [6]. In addition, the battery degradation characteristics are affected by multiple stress factors including depth-of-discharge, temperature, loading profiles, and SOC ranges, etc. These stress factors show nonlinear and coupling effects on battery degradation behaviors, and it is thus difficult to build accurate aging models to capture battery degradation dynamics. Meanwhile, these challenges constitute a major incentive to take advantage of advanced diagnosis and prognosis methodologies. Among them, EIS is one of the most promising methods to characterize the degradation dynamics of lithium-ion batteries [7]. In this paper, we aim to diversify a data-driven SOH estimation approach by taking advantage of geometric features of the EIS test data and model-free properties of GPR. It is worth noting that EIS testing on batteries usually requires a standalone battery EIS tester. However, the onboard application is not a necessary requirement for the battery SOH estimation functions, which can be performed when the vehicle is parked, inspection, and maintained. Thus, the proposed algorithm is the potential to be an effective supplement to battery health diagnosis.

* Corresponding author.
E-mail address: jwwei@nju.edu.cn (J. Wei).

https://doi.org/10.1016/j.energy.2022.125514
Received 5 June 2022; Received in revised form 14 September 2022; Accepted 19 September 2022
Available online 23 September 2022

Nomenclature

Abbreviations

AC	Alternating Current
ANN	Artificial neural networks
DC	Direct Current
ECM	Equivalent circuit model
EIS	Electrochemical impedance spectroscopy
GPR	Gaussian process regression
HI	Health indicator
LIB	Lithium-ion battery
MAE	Mean Absolute Error
PBM	Physics-based model
PCC	Pearson Correlation Coefficient
PHM	Prognosis and Health Management
R-GPR	Recurrent Gaussian process regression
RMSE	Root Mean Square Error
SCC	Spearman Correlation Coefficient
SEI	Solid electrolyte interphase
SOC	State-of-charge
SOH	State of health
SVM	Support vector machine

Notation

(X, Y)	Center position of the semicircle in Nyquist diagram
$\mathbb{E}$	Expectation operator
$\mathbf{0}$	Zero vector
$\mathbf{I}$	Unit matrix
ρ_{pcc}	Pearson Correlation Coefficient
ρ_{scc}	Spearman Correlation Coefficient
j	Imaginary unit
R	Radius of the semicircle in Nyquist diagram
Z'	Real part of impedance
Z''	Imaginary part of impedance
$Z(j\omega)$	Impedance at frequency ω

1.2. Literature review

A plethora of relevant technologies and methodologies have been developed and reported in the field of battery SOH estimation [8]. These studies can be roughly divided into two categories, namely, model-based and data-driven methods.

1.2.1. Model-based approaches

The model-based methods first establish a mathematical model to capture battery dynamic behaviors. Then, parameter identification and state observation algorithms are designed to online monitor SOH-related parameters, such as capacity and internal resistance. Among them, ECMs are likely the most popular ones due to their simplicity and good accuracy [9]. For instance, Qiu et al. [10] proposed a backward smoothed square root cubature Kalman filter and a multi-scale hybrid Kalman filter for joint SOC and SOH estimation based on a second-order ECM. Wei et al. [11] proposed a multi-time scale framework for joint SOC and SOH estimation, in which the dual extended Kalman filter was designed to estimate SOH estimation. Dong et al. [12] proposed a two-layer hierarchical approach for joint SOC and lifetime prediction, where the SOH was estimated based on the measured current and robust SOC estimates within one discharging cycle. Compared to ECMs, PBMs can improve model prediction accuracy and provide insights into the interior of the battery. Thus, some of the internal SOH-related parameters are based on model-based observer approaches. For instance, Bartlet et al. [13] proposed an online SOH estimator by estimating the loss of active materials in composite electrodes through a reduced-order PBM, similar work can also be found in [14]. The advantages of these model-based approaches include their simplicity and adaptability. They are also closed-loop and easy to implement. However, the drawbacks of these methods lie in that the SOH-related parameters may be unidentifiable and battery models may be unobservable. In addition, the convergency and stability of these observers may depend on the type of loading profiles [15].

1.2.2. Data-driven approaches

Due to the flexibility and model-free characteristics, data-driven approaches are gaining increasing attention in battery SOH estimation. Roman et al. [16] proposed a machine learning pipeline for battery SOH estimation, which includes three steps, HI extraction, offline training machine learning using experimental data, and online estimation using the trained model. According to the types of HI selection, these works can be categorized into three subcategories, namely, convenient-variables-based, statistical-metrics-based, and signal-based [17]. The convenient-variables-based HIs directly employ measurable variables, such as constant-current charging time, the instantaneous voltage drop at the start of discharge, the open-circuit voltage [18], and the Euclidean distance between important inflection points in the discharge curve [19]. The statistical-metrics-based HIs are extracted as different statistical metrics of battery operation data, and then, the relationship between these metrics and the battery SOH is established, for example, statistical properties in [20] and sample entropy in [21]. The signal-based HIs regard battery measurements as sequential signals, then signal processing methods can be employed to extract HI, for example, incremental capacity analysis in [22] and differential voltage analysis in [23]. Ng et al. [24] summarized the machine learning models as four categories: ANN [25], SVM [26], random forest regression [27], and GPR [28]. The advantages of these data-driven techniques are that they are more flexible than model-based approaches due to their model-free characteristics. In addition, they also show advantages in chemistry-agnostic modeling capability and the ability to analyze a wide range of degradation mechanisms and operating conditions, including rare loading events that are often overlooked by simplified models or physics-based simulations [16]. However, all of the above data-driven models that take time-domain variables (voltage, current, and temperature) as HIs may lose accuracy, since the time-domain variables may fail to provide reliable degradation information for certain scenarios. It can be found in [16] that some time-domain HIs show weak dependence on battery aging. Thus, identifying and extracting reliable HIs become the main bottleneck of the adoption of data-driven approaches.

To overcome the above drawbacks, recent studies have revealed that battery EIS is strongly dependent on the SOH since EIS measurements over a wide range of frequencies provide rich information about the dynamic characteristics of the battery and pave the way for precise estimation of the battery status [29]. For instance, Xiong et al. [30] employed ECMs to fit the EIS test data, then the SEI resistance was selected as the health indicator and was further used to predict battery capacity. Mingant et al. [31] first employed free voltage and current signal analysis method to generate battery "quasi" EIS, then the EIS-based HI was extracted as the parameters of an ECM. Fu et al. [32] employed fast Fourier transform to generate EIS using time-domain variables, then impedance at certain characteristic frequency points was employed as HI. Bo Jiang et al. [33] summarized fixed-frequency impedance feature can realize outstanding performance in battery SOH estimation. And Xiaojia Su et al. [34] extracted HIs related to the diffusion coefficient of lithium-ion through online recognition of low-frequency electrochemical impedance spectroscopy (LEIS) by step waves and realized the rapid estimation of lithium-ion battery capacity combined with GPR. The

advantage of frequency-domain characteristics is that they can provide richer insight information than time-domain measurements. Current techniques employ either the whole spectrum or data at characteristic frequency points. However, the EIS test at the lower frequency range is quite time-consuming, and the HIs extracted from certain characteristic frequencies may lack robustness since the EIS test may be affected by environmental conditions, equipment performances, and measurement methods. On the other hand, GPR is likely the most popular machine learning tool due to its strong nonlinear mapping ability, simple parameters, and its advantages of quantifying uncertainty. However, GPR establishes battery degradation models by mapping the extracted HIs and SOH, which ignores the battery degradation dynamics, i.e., the relationship between adjacent cycles. In fact, the battery degradation path is sequential data. To solve this problem, an alternative solution is to use R-GPR [35].

1.3. Key contributions

Considering the above research gaps, SOH estimation methods are proposed using R-GPR and HIs extracted from geometric EIS features. To be specific, the Nyquist diagram of battery EIS characteristics at the high and middle-frequency range is close to a semicircle, whose center and size are directly related to battery SOH. Thus, three HIs, including the position of this semicircle (represented as the center coordinates in the Nyquist diagram) and radius are extracted. Then, to improve the SOH estimation accuracy, the regular GPR is upgraded by adding a feedback loop to the input. The main contributions can be summarized as follows.

1. The HIs are all extracted in the middle and high-frequency range of EIS data, avoiding the time-consuming low-frequency test. In addition, the proposed geometric HIs are more robust than HIs extracted from certain characteristic frequency points.
2. Diverse HIs extracted from EIS data are compared in this study using Spearman correlation and Pearson coefficients, and the results indicate that the proposed HIs show higher dependence on battery SOH.
3. On the basis of conventional GPR, we consider the characteristics of battery capacity attenuation and feedback the capacity estimation value of the previous sample to the input for estimation, so that the prediction is smoother and the estimation accuracy is further improved.

1.4. Paper organization

The remainder of this paper is organized as follows. Experimental data and EIS test principles are presented in Section 2. Section 3 gives out the methodology regarding HI extraction and R-GPR, and Section 4 presents the evaluation of the proposed methodologies. Lastly, conclusions and remarks are summarized in Section 5.

2. EIS principles and experimental data

2.1. Principle and test method of EIS

EIS is a real-time, non-invasive, and informative measurement that has not been fully utilized in battery diagnosis so far. It is calculated by applying a small amplitude AC signal of different frequencies to the electrochemical system and measuring the variation of the ratio of AC signal voltage $u = U\sin(\omega t + \phi_u)$ to current $i = I\sin(\omega t + \phi_i)$ with the frequency of the sine wave ω [rad/s]. Fig. 1(a) shows the measurement principle of the battery impedance, the impedance can be calculated as follows:

$$Z(j\omega) \triangleq Z' + jZ'' = \frac{U\sin(\omega t + \phi_u)}{I\sin(\omega t + \phi_i)} \tag{1}$$

Taking the Z' as the abscissa and $-Z''$ as the ordinate, a Nyquist plot, i.e., EIS, is obtained to demonstrate the battery impedance at different frequencies.

2.2. Experimental data

The proposed methodology is validated using EIS data obtained in [36]. The experiments were conducted on 12 commercially available 45-mAh Eunicell LR2032 Li-ion coin cells using continuous charge and discharge cycles. The cell electrolyte material is $LiCoO_2$/graphite. Experiments were conducted by placing the cells at three temperatures of 25 °C (25C01-25C08), 35 °C (35C01 and 35C02) and 45 °C (45C01 and 45C02) for cycling, each cycle consists of a 1C-rate (45 mA) constant current-constant voltage charge up to 4.2 V and a 2C-rate (90 mA) constant current discharge down to 3 V. C-rate is the measurement of the charge and discharges current with respect to its nominal capacity. EIS measurements in the frequency ranging from 0.02 Hz to 20 kHz were performed under various operating conditions at nine different stages (denoted as State I, ⋯ IX) according to the SOC with and without Direct DC and relaxation. The battery capacity is then measured after an odd number of cycles. EIS and capacity data according to cycles, were used in this study. Taking cell 25C01 as an example, the relationship between EIS and cycle number is illustrated in Fig. 1(b). It can be seen that the size and position of the semicircle at the middle and high-frequency range are changing along with cycle numbers, i.e., battery SOH.

3. Methodologies

PCC in [29] and SCC in [37] are used to test and compare the performance of all the HIs, which can be calculated as

$$\rho_{pcc} = \frac{\mathbb{E}(\mathbf{x},\mathbf{y}) - \mathbb{E}(\mathbf{x})\mathbb{E}(\mathbf{y})}{\sqrt{\mathbb{E}(\mathbf{x}^2) - \mathbb{E}(\mathbf{x})^2}\sqrt{\mathbb{E}(\mathbf{y}^2) - \mathbb{E}(\mathbf{y})^2}} \tag{2}$$

$$\rho_{scc} = 1 - 6\sum_{k=1}^{n}\frac{d_k^2}{n(n^2-1)} \tag{3}$$

where $\mathbf{x}$ and $\mathbf{y}$ are two related vectors. d_k denotes the difference in ranks of the "kth" element between two variables. n denotes the number of data points of the two variables. In statistics, PCC and SCC reflect the direction and extent of the changing trend between the two variables, and their values range from -1.0 to $+1.0$ indicating that the two variables are independent. A higher absolute value of the correlation coefficient indicates a stronger relationship between variables.

3.1. Impedance-based HI extraction

3.1.1. Benchmark HIs

Although EIS has not been fully utilized in battery diagnosis, many researchers have explored the SOH of batteries through EIS feature extraction. In order to verify the proposed HIs extracted from geometric features, two categories of benchmarks in existing literature are used. The first category is the HIs extracted from parameters of fractional-order ECM fitted using EIS test data. The second category is EIS data at certain characteristic frequency points.

Benchmark 1 (ECM-based HIs) A convenient way to model battery frequency-domain properties is to employ fractional-order ECMs. It is found that some parameters of the ECMs can be employed as HIs due to their dependence on battery SOH. By using parameter fit tools, such as `zfit` in [38], the most commonly used HIs extracted through fitting fractional-order ECMs are listed in Table 1.

Benchmark 2 (HIs Extracted from Fixed Characteristic Frequencies): Another way to extract HIs is to directly select EIS data from some individual characteristic points. As shown in Fig. 2 [32], six health features (HF), denoted as HFi, $i = 1,\ldots,6$, are extracted at three characteristic points, which are also listed in Table 1.

Table 1
Correlation evaluation of the HIs.

HIs	R_s [39]	R_1 [39]	R_2 [39]	R_1 [40]	R_2 [40]	R_i [30]	R_{ct} [30]	R_{sei} [30]
PCC	−0.868	0.822	−0.896	−0.873	−0.882	−0.906	−0.767	0.784
SCC	−0.987	0.905	−0.984	−0.988	−0.981	−0.989	−0.960	0.960
HIs	R_0 [31]	Q_1 [31]	R_0/Q_1 [31]	17.8 Hz [36]	HF1 [32]	HF2 [32]	HF3 [32]	HF4 [32]
PCC	−0.998	0.984	−0.998	0.911	−0.850	−0.388	−0.716	−0.945
SCC	−0.998	0.984	−0.998	0.990	−0.989	−0.356	−0.873	−0.968
HIs	HF5 [32]	HF6 [32]	X	Y	R			
PCC	−0.940	0.860	0.954	0.963	0.962			
SCC	−0.934	0.767	0.995	0.991	0.993			

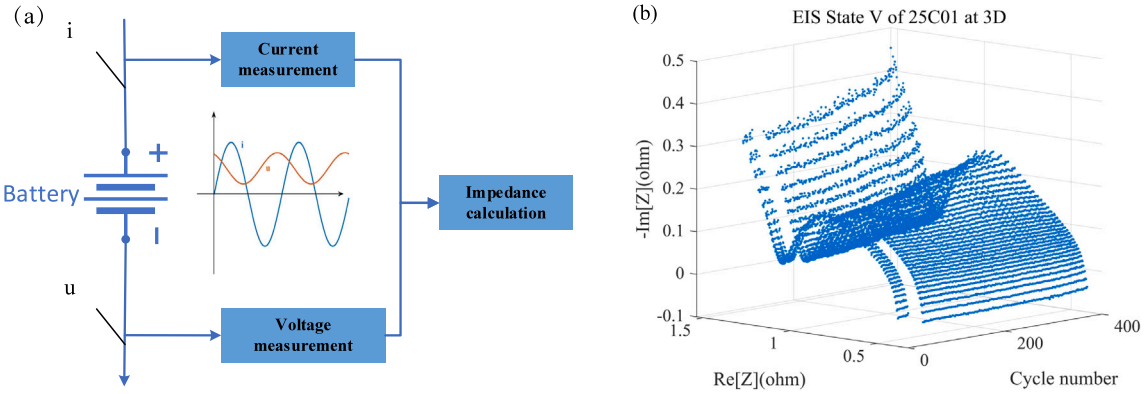


Fig. 1. Measurement principle of EIS test and 3D impedance spectrum of cell 25C01.

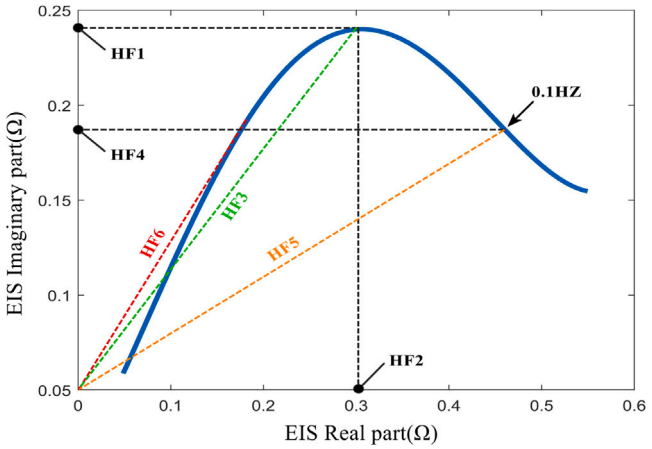


Fig. 2. Schematic diagram of Benchmark 2 HIs(slashes denote the slope).

3.1.2. The proposed geometric-based HIs

For the ECM-based HIs, though, the fractional-order models can provide a certain physical explanation for battery characteristics, the parameterization process requires the EIS data in a whole frequency range. In addition, the selection of the ECMs is still an open problem in the field of battery modeling. For the HIs extracted from certain characteristic frequencies, the advantage is that it is easy to implement and no complex fitting algorithm is required. The major drawback is that the EIS test data may be affected by diverse factors. Thus, the HIs may lose robustness against uncertain measurement conditions and wrong measurement results. Another important aspect is that both benchmarks require EIS data at a low-frequency range. However, the acquisition of low-frequency impedance is very time-consuming and low-frequency impedance is unstable during charging and discharging.

Current researchers believe that the solid electrolyte interphase (SEI) film growth is a major mechanism of battery degradation, which is an irreversible side reaction. As the battery charge and discharge cycles increase, the SEI film will continue to grow. This film will not only directly consume the active lithium, but will also adhere to the electrode surface, thereby increasing the resistance to lithium-ion diffusion, which manifests itself as an increase in cell resistance, resulting in capacity loss and power degradation [41]. Additionally, the first semicircle in the EIS is typically associated with the SEI [42]. With battery aging, the first semicircle not only gradually moves downward in position but also increases in curvature. Therefore, we expect to propose some features to characterize this change in order to determine the decay of lithium-ion batteries.

To overcome the above problems, this paper proposes an HI extraction method from geometric impedance features. As shown in Fig. 3, the Nyquist diagram of battery EIS at high and medium frequency ranges is close to a semicircle. Thus, the EIS characteristics at high and medium frequency ranges can be described as

$$(Z' - X)^2 + (Z'' - Y)^2 = R^2 \tag{4}$$

For experimental data $(Z'_j, Z''_j), j = 1, 2, \ldots, N_s$, X, Y and R can be obtained through solving following optimization problem

$$\arg\min_{\theta} \sum_{j=1}^{N_s} (Z'_j)^2 + (Z''_j)^2 + \theta_1 Z'_j + \theta_2 Z''_j + \theta_3, \tag{5}$$

where $\theta = [\theta_1, \theta_2, \theta_3]$ denotes the unknown parameters. This problem can be solved by the least squares method, we then can get

$$X = -\frac{\theta_1}{2}, Y = -\frac{\theta_2}{2},$$
$$R = \frac{\sqrt{\theta_1^2 + \theta_2^2 - 4\theta_3}}{2}.$$

In this paper, the curve segment from 57.4 Hz to 20 kHz (20 points) in EIS data is employed to obtain (X, Y) and R. The results are demonstrated in Fig. 3. It can be seen the center position and radius are highly related to battery degradation. It is worth noting that only states V and IX are considered in this study since they are collected when the battery reaches equilibrium status.

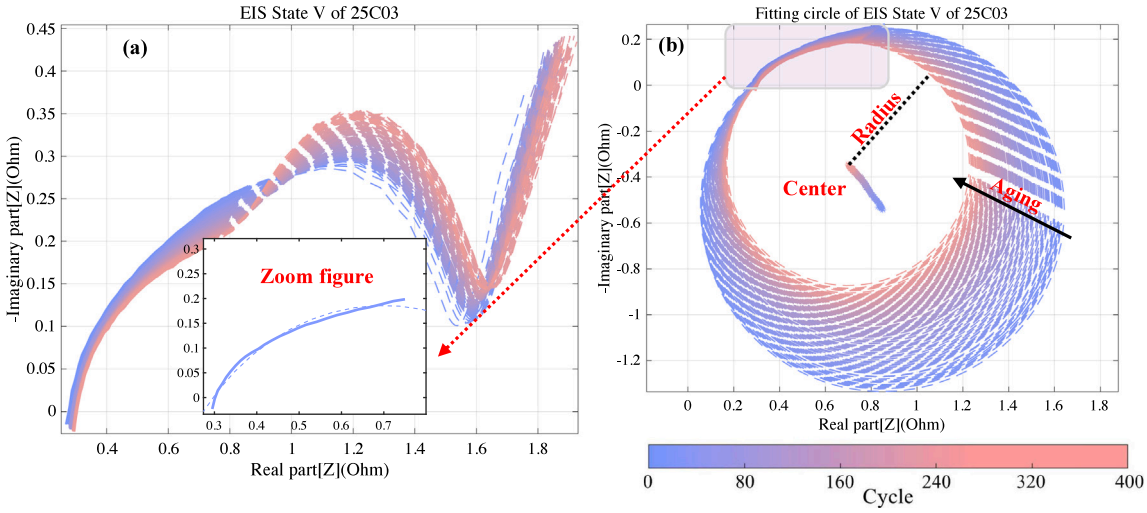
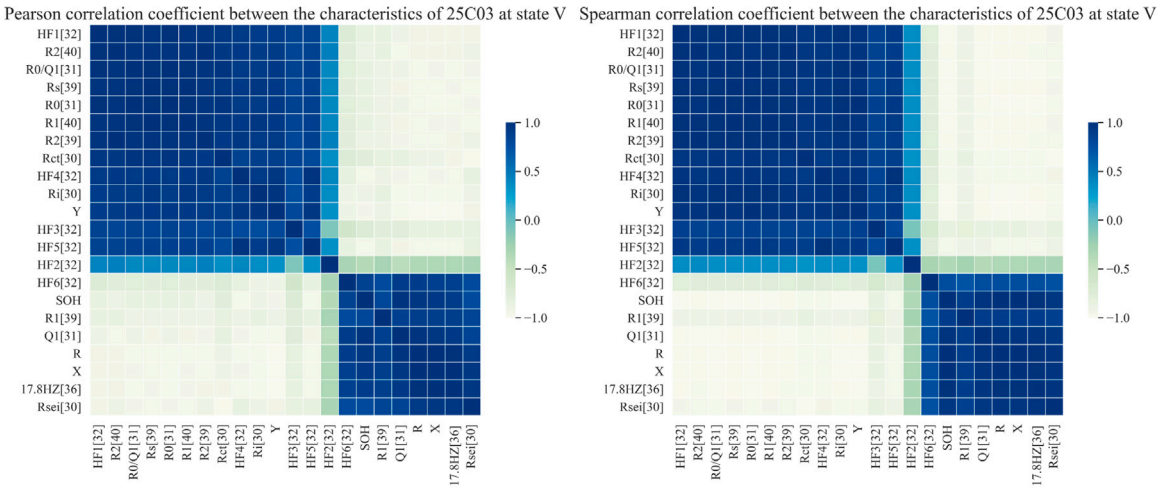
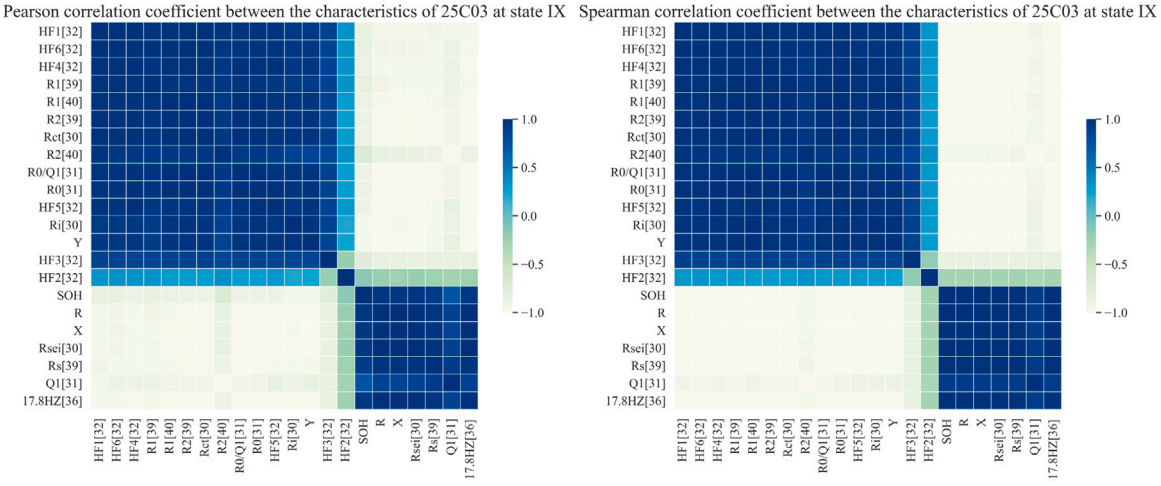


Fig. 3. Schematic diagram of geometric HIs (solid lines denote the selected data, dash lines denote the fitted circle).

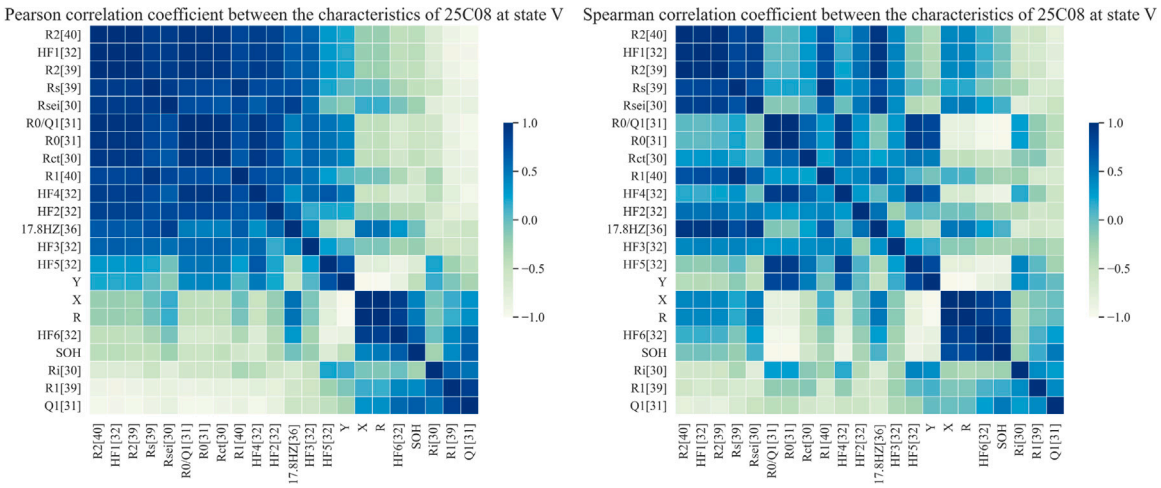


(a) Correlation between features of cell 25C03 at state V.

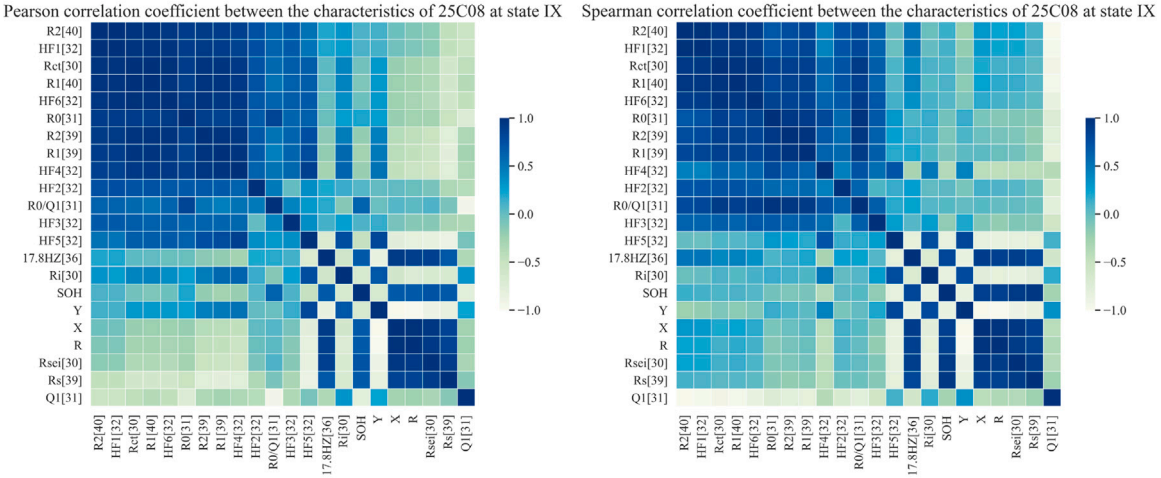


(b) Correlation between features of cell 25C03 at state IX

Fig. 4. Heatmap for feature sensitivity analysis of EIS spectrum of cell 25C03.



(a) Correlation between features of cell 25C08 at state V.



(b) Correlation between features of cell 25C08 at state IX.

Fig. 5. Heatmap for feature sensitivity analysis of EIS spectrum of cell 25C08.

To quantify the correlation between HIs and SOH and compare the performance of different HIs, the PCC and SCC between different HIs and SOH are calculated in Table 1, and the results are based on data collected from cell 25C03 at state V. It is clear that the proposed HIs, including center position and radius, show a high correlation to battery degradation.

To further validate the proposed HIs, the EIS test data of cells 25C03 and 25C08 at states V and IX are employed to generate the correlation matrix, which indicates the dependency of two or more variables on each other. The results are shown in Figs. 4 and 5. It can be seen in both figures that R and X are positively related to battery SOH, while Y is negatively related to battery SOH. It also demonstrates that the proposed HIs are more robust than other benchmarks for different cells and states. All these results indicate that the proposed HIs are suitable for accurate SOH estimation.

3.2. SOH estimation based on recurrent GPR

After obtaining suitable HIs, we turn to another key component of the machine learning learner for a data-driven SOH estimator. The machine learning learner employs a general fitting function with optimization hyperparameters. These parameters are tuned by using training data to deliver the desired behaviors. Then, the trained learner can be used to predict SOH for other battery cells. Prediction accuracy and uncertainty quantification are two of the most commonly used desired metrics. To this end, the GPR is likely the most popular learner in the field of battery SOH estimation due to its good nonlinear mapping and uncertainty quantification capability.

3.2.1. Regular GPR

Let $\mathfrak{D} = (\mathcal{X}, \mathcal{Y})$ denote the training dataset, comprising N data pairs, $\mathcal{X} = \{\mathbf{x}_j\}_{j=1}^N, \mathcal{Y} = \{\mathbf{y}_j\}_{j=1}^N$, where $\mathbf{x} \in \mathbb{R}^D$ and output $\mathbf{y} \in \mathbb{R}$. The GPR assumes that there is a latent function $\mathbf{y}_j = f(\mathbf{x}_j) + \epsilon_j$, which can map $\mathbf{x}_j$ to its corresponding output $\mathbf{y}_j$, $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$ denotes zero-mean Gaussian noise. The key of GPR is to assume any set of function values $\mathbf{f} = [f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_N)]$ as a multivariate Gaussian distribution, i.e.,

$$f(\mathbf{x}) \sim \mathcal{GP}(0, \mathbf{K}(\mathbf{x}, \mathbf{x}')), \tag{6}$$

where $\mathbf{K}$ is a matrix with element

$$\mathbf{K}_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j), \tag{7}$$

where $\kappa(\cdot, \cdot)$ denote a kernel function. In this study, the square exponential covariance function is employed, i.e.,

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2l^2}\right), \tag{8}$$

where σ_f and l are two hyperparameters. The goal of GPR is to predict a set of observations $\mathbf{f}^\star = [f(\mathbf{x}_1^\star), f(\mathbf{x}_2^\star), \cdot, f(\mathbf{x}_N^\star)]^\mathrm{T}$ at test inputs $\mathcal{X}^\star$ for the

training dataset $\mathfrak{D}$. Considering (6), the joint probability distribution of $\mathcal{Y}$ with $\mathbf{f}^\star$ can be calculated by

$$\begin{pmatrix} \mathcal{Y} \\ \mathbf{f}^\star \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{K}+\sigma^2\mathbf{I} & \mathbf{K}_\star \\ \mathbf{K}_\star^{\mathrm{T}} & \mathbf{K}_{\star\star} \end{pmatrix}\right), \tag{9}$$

where $(\mathbf{K}_\star)_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j^\star)$, and $(\mathbf{K}_{\star\star})_{ij} = \kappa(\mathbf{x}_i^\star, \mathbf{x}_j^\star)$.

Thus, according to (9), the conditional probability density $p(f_\star|\mathcal{X}_\star, \mathcal{X}, \mathcal{Y})$ can be calculated by

$$p(f_\star|\mathcal{X}_\star, \mathcal{X}, \mathcal{Y}) = \mathcal{N}(f_\star|\mu_\star, \Sigma_\star) \tag{10}$$

where the mean $\mu_\star$ and variance $\Sigma_\star$ can be formulated as

$$\mu_\star = \mathbf{K}_\star^{\mathrm{T}}(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\mathcal{Y},$$
$$\Sigma_\star = \mathbf{K}_{\star\star} - \mathbf{K}_\star^{\mathrm{T}}(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\mathbf{K}_\star.$$

Thus, we can predict battery SOH $f_\star$ for any given inputs $\mathbf{x}_\star$.

3.2.2. R-GPR for SOH estimation

To improve the SOH estimation performance regarding the accuracy, a recurrent structure is applied to the regular GPR. As shown in Fig. 6(a), we implement a feedback loop with a one-tap delay. The previously estimated SOH value at cycle $k-1$ is fed back to the input layer so that the previous SOH together with the extracted HIs is employed to predict battery SOH at cycle k. In fact, battery SOH will not change abruptly within one or two adjacent cycles, and this can be implemented by a closed-loop feedback mechanism. Thus, we implement one-tap delays in the regular GPR to improve estimation accuracy. The proposed algorithm is listed in Algorithm 1.

Algorithm 1 SoH estimation using R-GPR.

1: **Training part:**
2: **Step 1:** Obtain a training dataset, $\mathfrak{D} = (\mathcal{X}, \mathcal{Y})$, where $\mathcal{X}$ includes the radius length of the EIS fitting circle, the abscissa of the circle center, and the ordinate of the circle center under the current cycle of the battery, together with the previous SOH value of the last cycle, and $\mathcal{Y}$ are the corresponding current cycle battery SOH value.
3: **Step 2:** Initialize hyperparameters,Θ,
4: **Step 3:** Apply the conjugate gradient method to find the optimal values of the hyperparameters that minimize the negative marginal log-likelihood function (equivalently maximize the marginal log-likelihood function).
5: **Estimation part:**
6: **Step 4:** Obtain the mean and variance of the predictive distribution given optimal hyperparameters, training dataset, $\mathfrak{D}$, and test input $\mathbf{x}^\star$ as follows:

$$\mu_\star = \mathbf{K}_\star^{\mathrm{T}}(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\mathcal{Y}$$
$$\Sigma_\star = \mathbf{K}_{\star\star} - \mathbf{K}_\star^{\mathrm{T}}(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\mathbf{K}_\star$$

where $\mu_\star$ is the SOH estimate.

3.3. Implementation details of the proposed methodology

Before the formal training of the model, we need to do some auxiliary work. Firstly, six cells at room temperature (25 °C) were selected from the aging data of the cell to divide the training set and the test set. Among them, cells 25C02, 25C04, 25C06, and 25C07 were used as the training set, and cells 25C03 and 25C08 were used for testing. Second, the input vectors in the training process are normalized to balance the weights of each HI.

To characterize the estimation performance for any variable ψ, the RMSE and the MAE with the definition of

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N}(\psi_j - \hat{\psi}_j)^2}{N}} \tag{11}$$

$$MAE = \frac{1}{N}\sum_{j=1}^{N}|\psi_j - \hat{\psi}_j| \tag{12}$$

are used, where $\hat{\psi}_j$ and ψ_j are the estimate and measurement. N is the data length.

4. Results and discussion

In this section, experimental results of different cells under different conditions are employed to validate the proposed HIs and R-GPR-based SOH estimator.

4.1. Evaluation of different HIs

To validate the performance of different HIs, we first extracted all the HIs listed in Table 1. The extraction methods are presented in Section 3. These HIs are categorized into seven groups. Four of them are extracted by fitting the four ECMs described in Table 1. The fifth group is extracted using certain characteristic frequencies, i.e., six HFs in [32]. The sixth group is the EIS data at 17.8 Hz extracted by [36]. The last group is the proposed geometric-based HIs. The regular GPR is employed to evaluate their performance in SOH estimation. The EIS data of cell 25C03-V is employed. The results are shown in Fig. 6(b), while the numeric results are listed in Table 2. It can be seen that the geometric-based HIs proposed in this paper show the highest accuracy in predicting battery SOH, with RMSE and MAE within 1.38% and 0.93% for SOH and within 0.51 mAh and 0.34 mAh for capacity, respectively. These results indicate the good performance of the proposed geometric-based HIs. In addition, it can also be seen in Fig. 7 that the regular GPR-based SOH estimation shows frequent fluctuations. The reason is that the GPR-based approach directly maps HIs to SOH. The fluctuations of HIs will directly reflect SOH. This motivates us to propose the R-GPR to improve estimation accuracy.

4.2. Evaluation of different learners

To further validate the performance of the proposed R-GPR-based SOH estimator, eight datasets from two individual cells 25C03 and 25C08, and four SOC statuses, states IV, V, VI and IX are employed to test the trained R-GPR. Regular GPR, SVM, and ANN are employed to validate its superiority. The results are shown in Fig. 7, where the yellow region denotes the 95% confidence intervals. It should be noted that the SVM uses a Gaussian kernel, while the ANN contains two hidden layers each containing ten neurons. The numeric results are listed in Tables 3 and 4. From Fig. 7, it can be seen that all the methods can follow the reference SOH value. This indicates that the proposed HIs have a strong correlation to battery degradation. In addition, the R-GPR shows the highest SOH estimation accuracy for all four datasets, and also provides a more smooth SOH estimation curve, while the results of the other three algorithms show high fluctuations. The SOH estimation errors in Fig. 8 show that the maximum absolute SOH errors are mostly less than 2% for all eight test datasets. From Table 3, it can be seen that the RMSEs of R-GPR are less than 1.50% for the four datasets. And the R-square coefficients of determination for R-GPR are mostly more than 0.9 for eight test datasets, which are referenced in Table 4, R-square can be calculated as

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2} \tag{13}$$

where y and $\hat{y}$ denote SOH true value and estimate value, $\bar{y}$ denote the mean of y.

For the results of different cells, it can be seen that the SOH estimation results of cell 25C03 are more accurate than those of cell 25C08. For the three benchmark learners, the RMSE and MAE of cell 25C03 are less than 2.68% for SOH and less than 0.99 mAh for capacity, while those of cell 25C08 are less than 6.38% for SOH and less than

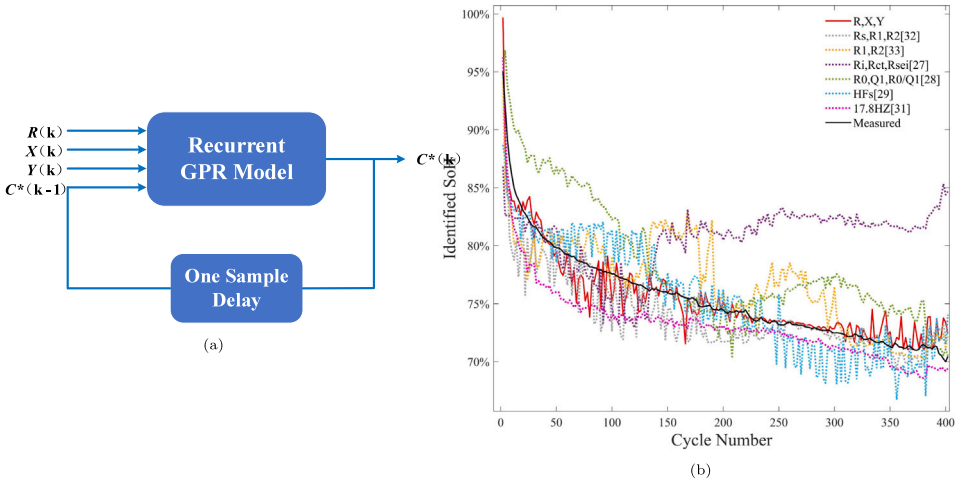


Fig. 6. (a) SOH estimation of Recurrent GPR, (b) SOH estimation for different HIs for cell 25C03-V.

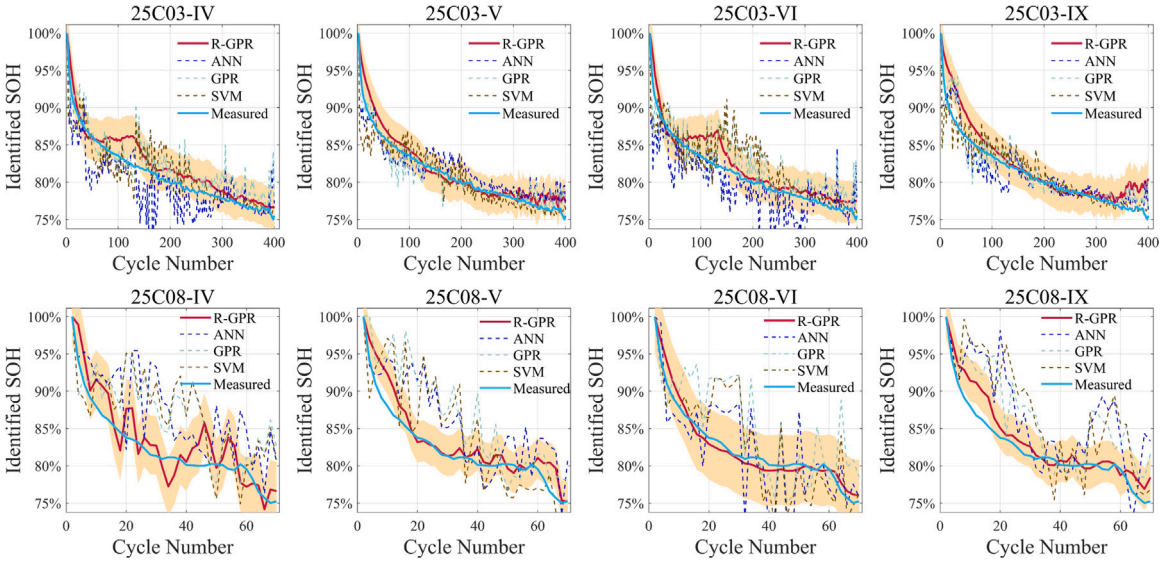


Fig. 7. Comparison of different data-driven learners.

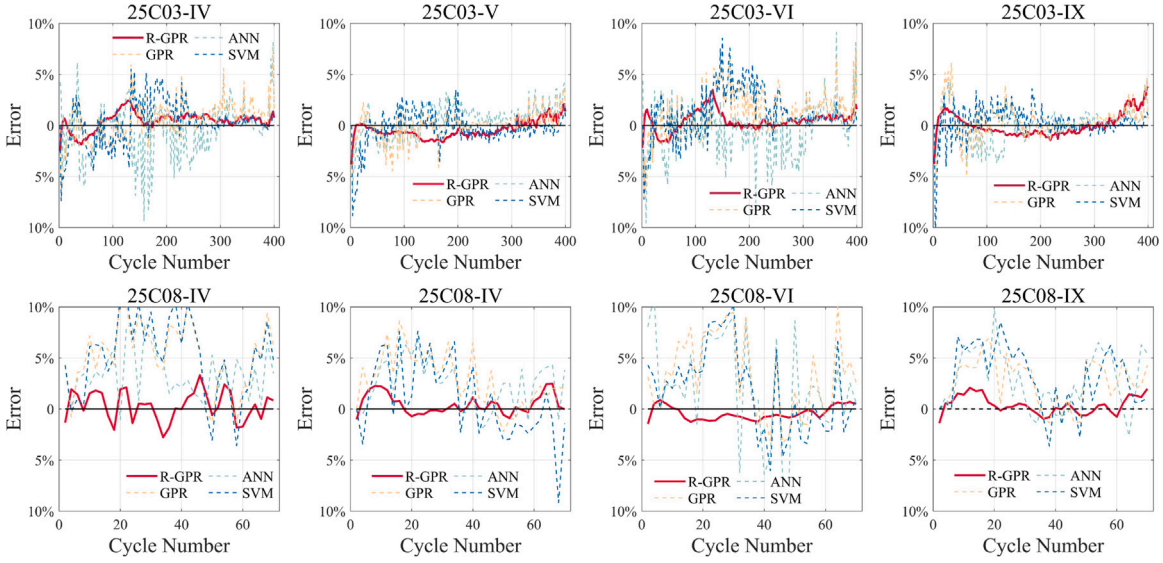


Fig. 8. SOH estimation errors of different data-driven learners.

Table 2
Numeric results of SOH and Capacity estimation using different HIs.

HIs	X, Y, R	ECM1 [39]	ECM2 [40]	ECM3 [30]	ECM4 [31]	17.8 Hz [36]	HFs in [32]
MAE-SOH [%]	0.93	1.70	2.20	6.58	3.36	2.28	1.92
RMSE-SOH [%]	1.38	2.30	2.98	7.68	3.99	2.54	2.33
MAE-Capacity [mAh]	0.34	0.63	0.81	2.42	1.24	0.84	0.71
RMSE-Capacity [mAh]	0.51	0.84	1.10	2.83	1.479	0.94	0.86

Table 3
Numeric results of SOH estimation using different learners.

Methods		R-GPR	GPR	SVM	ANN
25C03-IV	MAE-SOH [%]	0.80	1.56	1.51	1.91
	RMSE-SOH [%]	0.96	1.97	2.07	2.65
25C03-V	MAE-SOH [%]	0.79	0.89	1.07	1.43
	RMSE-SOH [%]	0.94	1.31	1.71	1.70
25C03-VI	MAE-SOH [%]	0.78	1.99	2.00	1.89
	RMSE-SOH [%]	1.04	2.54	2.68	2.56
25C03-IX	MAE-SOH [%]	0.87	1.16	0.96	1.19
	RMSE-SOH [%]	1.12	1.70	1.54	1.65
25C08-IV	MAE-SOH [%]	1.29	5.40	5.51	3.38
	RMSE-SOH [%]	1.50	6.12	6.38	4.17
25C08-V	MAE-SOH [%]	0.77	3.15	3.07	2.76
	RMSE-SOH [%]	1.06	3.87	3.88	3.30
25C08-VI	MAE-SOH [%]	0.71	4.09	3.74	4.42
	RMSE-SOH [%]	0.79	5.12	4.57	5.41
25C08-IX	MAE-SOH [%]	0.76	3.01	3.58	3.21
	RMSE-SOH [%]	0.96	3.53	4.31	4.05

Table 4
R-Squared coefficient of determination for R-GPR.

Datasets	25C03-IV	25C03-V	25C03-VI	25C03-IX
R-Square	0.94	0.94	0.93	0.91
Datasets	25C08-IV	25C08-V	25C08-VI	25C08-IX
R-Square	0.81	0.92	0.95	0.92

2.17 mAh for capacity. The reason is that the data length of cell 25C08 is much shorter than cell 25C03. This result shows that the estimation accuracy of the other three benchmarks may decrease in the case of a small degradation path, while the proposed R-GPR still works well. This indicates that the proposed R-GPR is more robust and more applicable than other benchmark algorithms.

5. Conclusion

In this paper, the SOH estimation of lithium-ion batteries is performed based on geometric impedance HIs and R-GPR. Firstly, considering that collecting low-frequency EIS data is a time-consuming process, this paper proposes three HIs based on the geometric shape of the EIS Nyquist diagram. The EIS data at the high and medium frequency range is employed to fit a semicircle. The position and radium of this cycle are found highly correlated to battery SOH and thus used as battery HIs. After proposing the reliable HIs, a recurrent Gaussian process regression is trained, and the predictions are performed by the trained model over different datasets. Performance indices such as Pearson correlation coefficient, Spearman correlation coefficient, RMSE, and MAE are used to evaluate the robustness of the accuracy of the proposed HIs and SOH estimators. The results indicate that the proposed methods can accurately estimate battery SOH using reliable health indicators. The geometric-based HIs are found to have a strong correlation to battery SOH. The RMSE and MAE of the proposed R-GPR are less than 1.50% in SOH and less than 0.51 mAh in capacity for four different datasets and show strong robustness against the size of the test datasets. Therefore, the proposed method shows good potential for practically embedding in the BMS to accurately monitor battery SOH.

Our future work will focus on SOH estimation at an early stage, which is challenging due to the weak correlation between HIs and SOH at battery early life. In addition, we will also consider various temperature conditions and the scenario where some of the cycle data of the battery is missing.

CRediT authorship contribution statement

Yong Zhou: Writing – original draft, Conceptualization, Methodology, Investigation. **Guangzhong Dong:** Writing – review & editing. **Qianqian Tan:** Formal analysis, Review & editing. **Xueyuan Han:** Formal analysis. **Chunlin Chen:** Writing – review & editing. **Jingwen Wei:** Formal analysis, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported in part by the Fundamental Research Funds for the Central Universities, China under Grant number 0221-14380010, the Natural Science Foundation of Jiangsu Province, China under Grant number BK20200333, and the National Natural Science Fund of China, China under Grant number 62203209

References

[1] Xie J, Lu Y-C. A retrospective on lithium-ion batteries. Nature Commun 2020;11(1):1–4. http://dx.doi.org/10.1038/s41467-020-16259-9.
[2] Chen J, Zhang Y, Li X, Sun B, Liao Q, Tao Y, Wang Z. Strategic integration of vehicle-to-home system with home distributed photovoltaic power generation in Shanghai. Appl Energy 2020;263:114603. http://dx.doi.org/10.1016/j.apenergy.2020.114603.
[3] Dong G, Wei J, Zhang C, Chen Z. Online state of charge estimation and open circuit voltage hysteresis modeling of LiFePO4 battery using invariant imbedding method. Appl Energy 2016;162:163–71. http://dx.doi.org/10.1016/j.apenergy.2015.10.092.
[4] Hu X, Xu L, Lin X, Pecht M. Battery lifetime prognostics. Joule 2020;4(2):310–46. http://dx.doi.org/10.1016/j.joule.2019.11.018.
[5] Huang S, Liu C, Sun H, Liao Q. State of health estimation of lithium-ion batteries based on the regional frequency. J Power Sources 2022;518:230773. http://dx.doi.org/10.1016/j.jpowsour.2021.230773.
[6] Dong G, Wei J. A physics-based aging model for lithium-ion battery with coupled chemical/mechanical degradation mechanisms. Electrochim Acta 2021;395:139133. http://dx.doi.org/10.1016/j.electacta.2021.139133.
[7] Tröltzsch U, Kanoun O, Tränkler H-R. Characterizing aging effects of lithium ion batteries by impedance spectroscopy. Electrochim Acta 2006;51(8–9):1664–72. http://dx.doi.org/10.1016/j.electacta.2005.02.148.
[8] Shah FA, Shahzad Sheikh S, Mir UI, Owais Athar S. Battery health monitoring for commercialized electric vehicle batteries: Lithium-ion. In: 2019 International conference on power generation systems and renewable energy technologies (PGSRET). 2019, p. 1–6. http://dx.doi.org/10.1109/PGSRET.2019.8882735.
[9] Khalid M, Sheikh SS, Janjua AK, Khalid HA. Performance validation of electric vehicle's battery management system under state of charge estimation for lithium-ion battery. In: 2018 International conference on computing, electronic and electrical engineering (ICE Cube). 2018, p. 1–5. http://dx.doi.org/10.1109/ICECUBE.2018.8610969.

[10] Qiu X, Wu W, Wang S. Remaining useful life prediction of lithium-ion battery based on improved cuckoo search particle filter and a novel state of charge estimation method. J Power Sources 2020;450:227700. http://dx.doi.org/10.1016/j.jpowsour.2020.227700.
[11] Wei J, Chen C. A multi-timescale framework for state monitoring and lifetime prognosis of lithium-ion batteries. Energy 2021;229:120684. http://dx.doi.org/10.1016/j.energy.2021.120684.
[12] Dong G, Xu Y, Wei Z. A hierarchical approach for finite-time H-∞ state-of-charge observer and probabilistic lifetime prediction of lithium-ion batteries. IEEE Trans Energy Convers 2022;37(1):718–28. http://dx.doi.org/10.1109/TEC.2021.3109896.
[13] Bartlett A, Marcicki J, Rhodes K, Rizzoni G. State of health estimation in composite electrode lithium-ion cells. J Power Sources 2015;284:642–9. http://dx.doi.org/10.1016/j.jpowsour.2015.03.080.
[14] Sadabadi KK, Jin X, Rizzoni G. Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health. J Power Sources 2021;481:228861. http://dx.doi.org/10.1016/j.jpowsour.2020.228861.
[15] Ouyang Q, Chen J, Zheng J. State-of-charge observer design for batteries with online model parameter identification: A robust approach. IEEE Trans Power Electron 2019;35(6):5820–31. http://dx.doi.org/10.1109/TPEL.2019.2948253.
[16] Roman D, Saxena S, Robu V, Pecht M, Flynn D. Machine learning pipeline for battery state-of-health estimation. Nat Mach Intell 2021;3(5):447–56. http://dx.doi.org/10.1038/s42256-021-00312-3.
[17] Dong G, Han W, Wang Y. Dynamic Bayesian network-based lithium-ion battery health prognosis for electric vehicles. IEEE Trans Ind Electron 2020;68(11):10949–58. http://dx.doi.org/10.1109/TIE.2020.3034855.
[18] Lin HT, Liang TJ, Chen SM. Estimation of battery state of health using probabilistic neural network. IEEE Trans Industr Inform 2012;9(2):679–85. http://dx.doi.org/10.1109/TII.2012.2222650.
[19] Sheikh SS, Anjum M, Khan MA, Hassan SA, Khalid HA, Gastli A, Ben-Brahim L. A battery health monitoring method using machine learning: A data-driven approach. Energies 2020;13(14). http://dx.doi.org/10.3390/en13143658.
[20] Li F, Xu J. A new prognostics method for state of health estimation of lithium-ion batteries based on a mixture of Gaussian process models and particle filter. Microelectron Reliab 2015;55(7):1035–45. http://dx.doi.org/10.1016/j.microrel.2015.02.025.
[21] Hu X, Jiang J, Cao D, Egardt B. Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling. IEEE Trans Ind Electron 2015;63(4):2645–56. http://dx.doi.org/10.1109/TIE.2015.2461523.
[22] Ma Z, Jiang J, Shi W, Zhang W, Mi CC. Investigation of path dependence in commercial lithium-ion cells for pure electric bus applications: Aging mechanism identification. J Power Sources 2015;274:29–40. http://dx.doi.org/10.1016/j.jpowsour.2014.10.006.
[23] Chen X, Hu Y, Li S, Wang Y, Li D, Luo C, Xue X, Xu F, Zhang Z, Gong Z, Li Y, Yang Y. State of health (SoH) estimation and degradation modes analysis of pouch NMC532/graphite li-ion battery. J Power Sources 2021;498:229884. http://dx.doi.org/10.1016/j.jpowsour.2021.229884.
[24] Ng M-F, Zhao J, Yan Q, Conduit GJ, Seh ZW. Predicting the state of charge and health of batteries using data-driven machine learning. Nat Mach Intell 2020;2(3):161–70. http://dx.doi.org/10.1038/s42256-020-0156-7.
[25] Fan Y, Xiao F, Li C, Yang G, Tang X. A novel deep learning framework for state of health estimation of lithium-ion battery. J Energy Storage 2020;32:101741. http://dx.doi.org/10.1016/j.est.2020.101741.
[26] Klass V, Behm M, Lindbergh G. A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation. J Power Sources 2014;270:262–72. http://dx.doi.org/10.1016/j.jpowsour.2014.07.116.
[27] Li Y, Zou C, Berecibar M, Nanini-Maury E, Chan JCW, van den Bossche P, Van Mierlo J, Omar N. Random forest regression for online capacity estimation of lithium-ion batteries. Appl Energy 2018;232:197–210. http://dx.doi.org/10.1016/j.apenergy.2018.09.182.
[28] Yang D, Zhang X, Pan R, Wang Y, Chen Z. A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. J Power Sources 2018;384:387–95. http://dx.doi.org/10.1016/j.jpowsour.2018.03.015.
[29] Babaeiyazdi I, Rezaei-Zare A, Shokrzadeh S. State of charge prediction of EV Li-ion batteries using EIS: A machine learning approach. Energy 2021;223:120116. http://dx.doi.org/10.1016/j.energy.2021.120116.
[30] Xiong R, Tian J, Mu H, Wang C. A systematic model-based degradation behavior recognition and health monitoring method for lithium-ion batteries. Appl Energy 2017;207:372–83. http://dx.doi.org/10.1016/j.apenergy.2017.05.124.
[31] Mingant R, Bernard J, Sauvant-Moynot V. Novel state-of-health diagnostic method for Li-ion battery in service. Appl Energy 2016;183:390–8. http://dx.doi.org/10.1016/j.apenergy.2016.08.118.
[32] Fu Y, Xu J, Shi M, Mei X. A fast impedance calculation based battery state-of-health estimation method. IEEE Trans Ind Electron 2022;69(7):7019–28. http://dx.doi.org/10.1109/TIE.2021.3097668.
[33] Jiang B, Zhu J, Wang X, Wei X, Shang W, Dai H. A comparative study of different features extracted from electrochemical impedance spectroscopy in state of health estimation for lithium-ion batteries. Appl Energy 2022;322:119502. http://dx.doi.org/10.1016/j.apenergy.2022.119502.
[34] Su X, Sun B, Wang J, Zhang W, Ma S, He X, Ruan H. Fast capacity estimation for lithium-ion battery based on online identification of low-frequency electrochemical impedance spectroscopy and Gaussian process regression. Appl Energy 2022;322:119516. http://dx.doi.org/10.1016/j.apenergy.2022.119516.
[35] Sahinoglu GO, Pajovic M, Sahinoglu Z, Wang Y, Orlik PV, Wada T. Battery state-of-charge estimation based on regular/recurrent Gaussian process regression. IEEE Trans Ind Electron 2017;65(5):4311–21. http://dx.doi.org/10.1109/TIE.2017.2764869.
[36] Zhang Y, Tang Q, Zhang Y, Wang J, Stimming U, Lee AA. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. Nature Commun 2020;11(1):1706. http://dx.doi.org/10.1038/s41467-020-15235-7.
[37] Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018;18(3):91–3. http://dx.doi.org/10.1016/j.tjem.2018.08.001.
[38] Dellis J-L. Zit.software. 2020, https://www.mathworks.com/matlabcentral/fileexchan-ge/19460-zfit.
[39] Stroe DI, Swierczynski M, Stan AI, Knap V, Teodorescu R, Andreasen SJ. Diagnosis of lithium-ion batteries state-of-health based on electrochemical impedance spectroscopy technique. In: 2014 IEEE Energy conversion congress and exposition (ECCE). 2014, p. 4576–82. http://dx.doi.org/10.1109/ECCE.2014.6954027.
[40] Eddahech A, Briat O, Bertrand N, Delétage J-Y, Vinassa J-M. Behavior and state-of-health monitoring of Li-ion batteries using impedance spectroscopy and recurrent neural networks. Int J Electr Power Energy Syst 2012;42(1):487–94. http://dx.doi.org/10.1016/j.ijepes.2012.04.050.
[41] Zhou F, Bao C. Analysis of the lithium-ion battery capacity degradation behavior with a comprehensive mathematical model. J Power Sources 2021;515:230630. http://dx.doi.org/10.1016/j.jpowsour.2021.230630.
[42] Andre D, Meiler M, Steiner K, Wimmer C, Soczka-Guth T, Sauer D. Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. I. Experimental investigation. J Power Sources 2011;196(12):5334–41. http://dx.doi.org/10.1016/j.jpowsour.2010.12.102, Selected papers presented at the 12th Ulm ElectroChemical Talks (UECT):2015 Technologies on Batteries and Fuel Cells.