

# A novel Bayesian multivariate linear regression model for online state-of-health estimation of Lithium-ion battery using multiple health indicators

Zhiqiang Lyu <sup>a,1</sup>, Geng Wang <sup>a,1</sup>, Cao Tan <sup>b,\*</sup>

<sup>a</sup> State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116000, China

<sup>b</sup> School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, China

## ARTICLE INFO

### Keywords:

Lithium-ion battery  
State-of-health  
Health indicator  
Bayesian multivariate linear regression

## ABSTRACT

An accurate state of health is critically significant to evaluate the battery aging level and to ensure electric vehicle security and reliability. This paper presents a novel Bayesian-based method to quantify battery capacity degradation using multiple health indicators extracted from the battery Thevenin model at different temperatures. In this method, the linearity and monotonicity between the extracted health indicators and capacity degradation are analyzed by exploiting the correlation analysis. To estimate the battery's state of health (SOH) accurately, the Bayesian multivariate linear regression is introduced to develop an online state of health estimator. Experimental tests are conducted on two batteries with the same specifications to verify the efficiency of the proposed method. Additionally, the interval estimation of numerical results contains reasonable uncertainty expression, which validates the robustness and reliability of the proposed method.

## 1. Introduction

To tackle global energy shortages and environmental deterioration, electric vehicles (EVs) have received wide attention [1]. For EV applications, Lithium-ion batteries have been received widespread attention due to their high energy and power density, long service life, and eco-friendliness [2]. However, there remain some major problems that need to be settled for performance enhancement of Lithium-ion batteries in EVs, such as state of health (SOH) estimation, life prediction, fault diagnosis, and so on. Among these problems, the battery SOH estimation is one of the most crucial and challenging issues for the battery management system (BMS) [3]. Accurate battery SOH estimation contributes to determining the degree of battery degradation, predicting the remaining useful life, and providing instruction for rational battery replacement [4].

As a kind of intricate and nonlinear electrochemical system, it is difficult to estimate the battery SOH accurately due to the internal multifaceted electrochemical reactions and external performance changes during the battery aging process. Various approaches have been proposed for online SOH estimation in the past few years [5,6]. These methods can mainly be classified into three categories: (1) direct

assessment approach, (2) adaptive approach, and (3) data-driven approach. In the direct assessment approach, the main technical methods include the coulomb counting method and the impedance spectroscopy method. For the coulomb counting method, the batteries are fully charged and discharged to obtain the static battery capacity [7]. As for the impedance spectroscopy method, the spectroscopies in the different frequencies sweep the battery at different states of charge (SOC) to calculate the specific model parameters and estimate the battery SOH. However, this method is difficult to implement online because of its large computation [8]. The main concept of the adaptive approach is to connect the battery signals with battery SOH through the battery electrochemical model or equivalent circuit model using adaptive techniques. Bartlett et al. [9] proposed a reduced-order electrochemical model for a composite electrode battery, while a dual-observer is applied to estimate SOC and capacity. Xiong et al. [10] used the finite analysis method to simplify the electrochemical model. Then, five characteristic parameters were applied to estimate the battery SOH based on the relationship between the battery degradation and model parameters of the pseudo-two-dimensional (P2D) model. The electrochemical model can explicitly characterize the spatiotemporal dynamics of electrochemical reactions inside batteries from the viewpoint of

\* Corresponding author.

E-mail address: [njusttancao@yeah.net](mailto:njusttancao@yeah.net) (C. Tan).

<sup>1</sup> These authors contributed equally to this work and should be considered co-first authors.



Fig. 1. Architecture of battery test system.

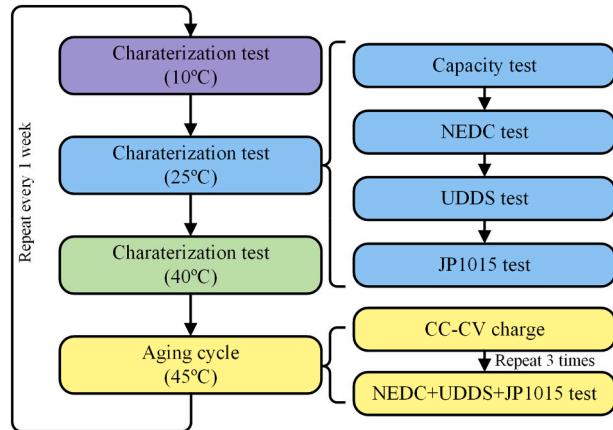


Fig. 2. Experimental schedule.

electrochemistry. Nevertheless, solving the elusive partial differential equations efficiently of the electrochemical model obstructs the embedded application [11]. An equivalent circuit model (ECM), i.e., a purely mathematical model based on common electrical components, is adopted as a popular model type of batteries [12]. To avoid the lack of capacity estimation accuracy in commonly used joint/dual extended Kalman filter, Hu et al. [13] proposed a multiscale framework using an extended Kalman filter (EKF) for SOC and capacity estimation with higher accuracy and efficiency. Andre et al. [14] estimated the SOC and SOH by use of a dual Kalman filter in combination with a support vector machine and verified the estimation results using different cycle profiles. Schwunk et al. [15] presented a particle filter (PF) based approach for the SOC and SOH estimation to overcome the shortcoming that Kalman-Filter-based methods represent a state by the mean and the variance of a Gaussian probability density function. Bi et al. [16] achieved a dynamic battery SOH estimation using the genetic resampling particle filter to eliminate the effects of multi-source noise. Using these techniques, such as Kalman filter, PF, and so on, the closed-loop SOH estimator was developed. However, one drawback of these techniques is that the battery aging level must be determined prior, which is difficult and expensive for a complex battery aging process [17]. Besides, the effectiveness of the adaptive approaches relies on the robustness and credibility of the battery models [7]. Data-driven approaches have also gained great attention for battery SOH estimation owing to their

flexibility and model-free characteristics. In general, data-driven approaches do not have to obtain prior knowledge of battery aging, and it is considered to be a “black box” model with non-physical parameters. You et al. [18] improved the recurrent neural network by introducing a long/short-term memory effect for battery SOH estimation using real-time BMS data (e.g., terminal voltage, load current, and temperature). To estimate SOH in more practical environments, Guo et al. [19] proposed a rational analysis and principal component analysis to extract and optimize health features, then, the relevance vector machine (RVM) was applied to fitting the relationship between the health features and capacity. Based on discharging voltage sequences, Hu et al. [7] achieved battery health prognosis using sparse Bayesian predictive modelling and sample entropy. Richardson et al. [20] proposed a Gaussian process regression (GPR) for in-situ capacity estimation using small portions of voltage-time data under constant current operation.

Almost all of the above methods used some battery characteristic parameters as the health indicators (HIs) to describe the battery aging indirectly and battery degradation modelling. With the extraction of HIs, researchers accumulated plentiful research achievements. Liu et al. [21,22] found that the time interval for the equal discharging voltage difference or the discharging voltage difference within an equal time interval during each discharge cycle could be used as HIs to quantify the capacity degradation. Xiong et al. [10] determined some characteristic parameters from the electrochemical model to describe the battery SOH. Bi et al. [16] considered the ohmic resistance of a second-order ECM for the battery as the evaluation index of battery SOH. Hu et al. [7] and Yang et al. [17] expounded the validity of sample entropy and charging parameters as HIs, respectively. Wang et al. [23] and Torai et al. [24] utilized differential voltage and differential capacity to estimate battery SOH, respectively. In Ref. [25], we have demonstrated the effectiveness of the ohmic internal resistance and polarized internal resistance extracted from the battery Thevenin model as the HIs. The HIs can be obtained easily using the recursive least-squares (RLS) method and utilized to quantify the capacity degradation effectively.

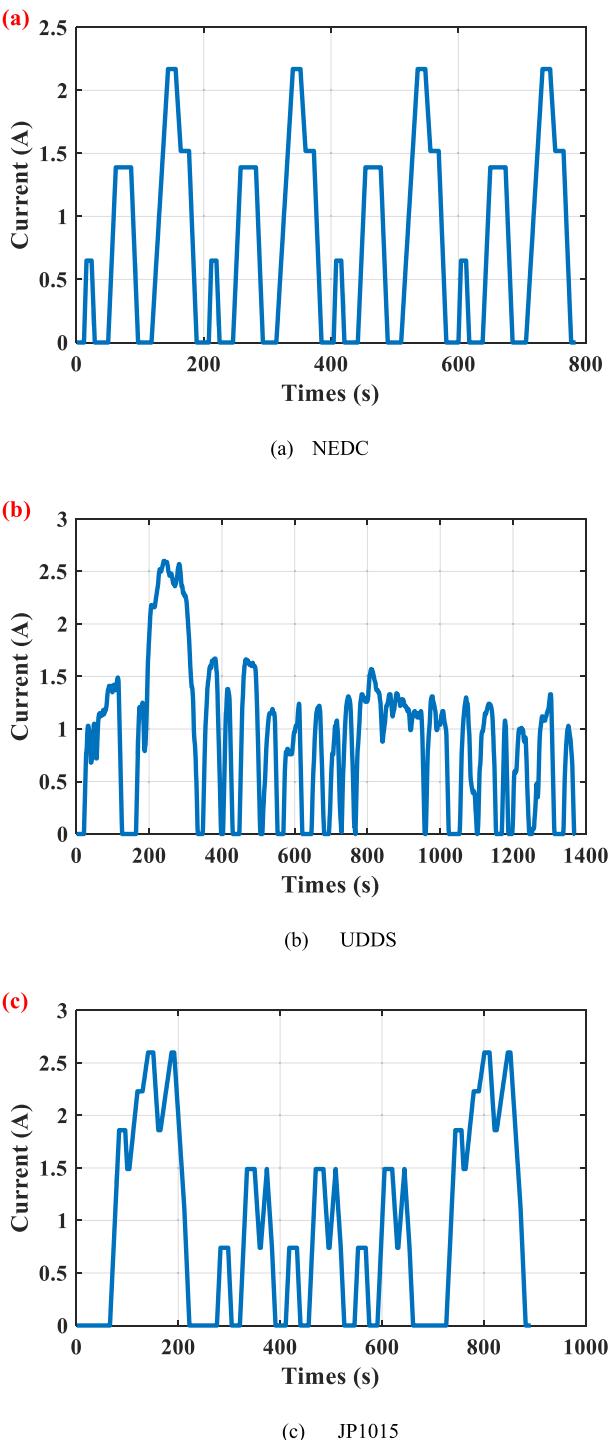
In summary, combining the effective HIs within the data-driven approaches contributes to realizing accurate and robust SOH estimation. In statistics, Bayesian multivariate linear regression (BMLR) is a kind of method for regression analysis in which the statistical analysis is implemented based on the Bayesian framework. As one of the data-driven approaches, BMLR has the advantages of simplicity and robustness, and it has been widely used in economics, statistics, and biochemistry, etc. The results of BMLR can be explicated in the probability-based form, which exhibits the reliability interpretation. Therefore, the BMLR shows great potential for online SOH estimation when taking the intricate battery aging process into account.

In this paper, integrating the excellences of the adaptive approach and data-driven approach, a joint battery model and BMLR model has been proposed for online battery SOH estimation. First, according to the characteristics of the Thevenin model, the increments of ohmic resistance and polarization resistance are defined as battery HIs and served as inputs of the BMLR, with which a novel SOH estimation framework is developed. Second, battery aging data at different temperatures are used to verify the performance of the proposed BMLR.

The remainder of the paper is organized as follows: In Section 2, the Battery testing system, experimental schedule, and battery HIs are exhibited. In Section 3, the BMLR is introduced, and the SOH estimation framework is developed. In Section 4, the results and discussions are presented. Finally, in Section 5, the conclusions are summarized.

## 2. Lithium-ion battery test and HIs

Battery SOH, as a key indicator of battery aging, is usually defined by capacity or ohmic resistance. Herein, according to our previous work [25], the SOH is defined as the ratio of the remaining battery capacity ( $C_{RBC}$ ) to the initial battery capacity ( $C_{IBC}$ ) at a certain temperature:



**Fig. 3.** Loading profiles.

$$SoH = \frac{C_{RBC}}{C_{IBC}} \times 100\% \quad (1)$$

## 2.1. Battery testing system and schedule

To simulate the practical environments of the Lithium-ion battery, the battery testing system was established as shown in Fig. 1.

The test bench consists of two chargers (ITECH IT6952A), two electronic loads (ITECH IT8511A+), two LiNMC 18650 batteries from Samsung (NO.A and NO.B), an environmental chamber, and a control computer. More device parameters are detailed in Ref. [25].

The two batteries were placed in the same environmental chamber using the specific clamps and tested by the experimental schedule independently, as shown in Fig. 2.

Each experimental period is composed of three characterization tests at 10 °C, 25 °C, and 40 °C, and the accelerated aging test (aging cycle) at 45 °C. Each characterization test includes the capacity test at 0.2C, a New European Driving Cycle (NEDC) test [see Fig. 3 (a)], a UDDS (Urban Dynamometer Driving Schedule) test [see Fig. 3 (b)], and a JP1015 (Japanese 10.15 Mode Driving Schedule) test [see Fig. 3 (c)] consecutively. The three representative dynamic loading profiles are scaled down to avoid exceeding the charging-discharging limits of batteries. Limited by the device limitation, the regenerative charging during the dynamic loading profiles is not able to be implemented. In the characterization test, the capacity test is employed for capacity calibration while the dynamic loadings are utilized for online SOH estimation. In the aging cycle, a joint driving schedule that combines the three dynamic loading profiles is conducted and repeated three times to accelerate the battery aging process. The experimental schedule is implemented until the measured battery capacity at 25 °C is reduced to 80% of the rated capacity. The battery aging data at different temperatures also contribute to verifying the performance and universality of the proposed method for SOH estimation.

## 2.2. HI extraction and evaluations

Thevenin model [see Fig. 4 (a)] is identified as the best choice considering the model complexity, accuracy, and robustness [26].

Here, OCV is the open-circuit voltage;  $I_L$  is the load current;  $R_0$  is the ohmic resistance;  $R_1$  and  $C_1$  is the polarization resistance and polarization capacitance, respectively;  $U_t$  is the terminal voltage. For many identification algorithms for battery model parameters, the RLS method shows the potential to determine the real-time model parameters.

Neglecting the complex electrochemical reactions inside the batteries, it can be recognized as a single-input single-output system. The only input of this model is the load current  $I_L$ , and the only output is the terminal voltage  $U_t$ . Therefore, the same NEDC is input to the battery and model, respectively, the errors between the model output and voltage measurement can reflect the model accuracy. The results of the model identification through one NEDC are shown in Fig. 4 (b) to (f). As shown in Fig. 4 (b), the curve of the estimated voltage shows good agreement with voltage measurement, and the errors are less than 10 mV as shown in Fig. 4 (c). The result suggests that the Thevenin model is of higher precision in simulating the dynamic process of battery discharging. Furthermore, the model parameters as shown in Fig. 4 (d) to (f) are accurate. Meanwhile, the model parameters remain unchanged when the SOC is greater than or equal to 30%. Thus, the mean values of  $R_0$  and  $R_1$  for 80–30% SOC can be calculated as the aging parameters at certain aging levels:

$$\begin{cases} R_0^i = \text{mean}(R_0(SOC = 30\%) + \dots + R_0(SOC = 80\%)) \\ R_1^i = \text{mean}(R_1(SOC = 30\%) + \dots + R_1(SOC = 80\%)) \end{cases} \quad (2)$$

where  $R_0^i$  and  $R_1^i$  are the mean values of  $R_0$  and  $R_1$ , The superscript  $i$  donates the  $i$ th measurement result. mean is used to calculate the mean value. The SOC is calculated using the Coulomb-counting method according to the definition of the method:

$$SOC(t) = SOC(t_0) - \int_{t_0}^t \frac{\eta I_L(\tau)}{C_a} d\tau \quad (3)$$

where  $(\cdot)$  is time for SOC calculation.  $\eta$  is the Coulombic efficiency,  $C_a$  is the present maximum available capacity.

To avoid complex offline calculations and using the electrochemical impedance spectrometry (EIS) data from certain specialized devices, the battery HIs need to be robust and easily accessible. According to the characteristics between model parameters and SOC [25], herein, we

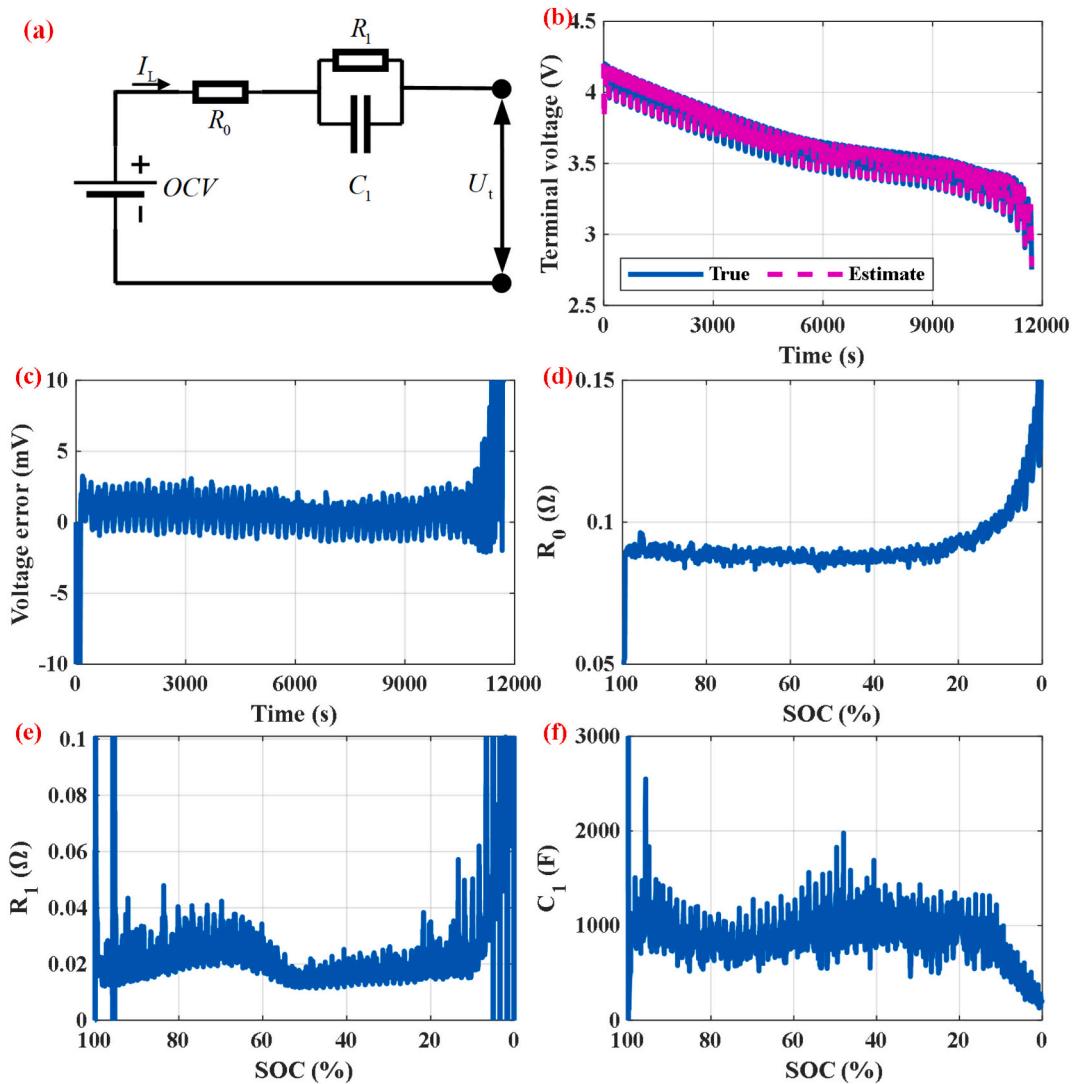


Fig. 4. Thevenin model and identification.

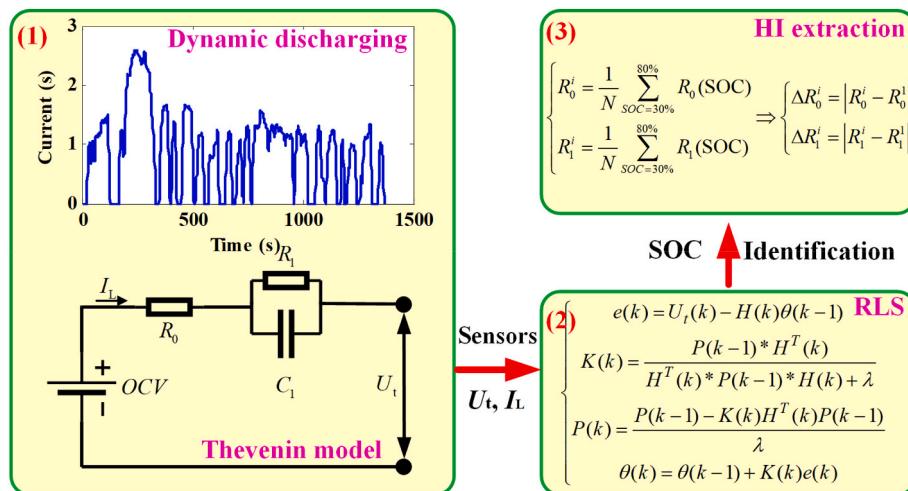


Fig. 5. HI extractor.

extracted the increment of ohmic resistance ( $\Delta R_0$ ) and polarization resistance ( $\Delta R_1$ ) for 80–30% SOC as the HIs to describe the battery capacity degradation ( $\Delta C$ ) as shown in Eq. (4):

**Table 1**  
Correlation analysis.

Temperature	Type	Method	NO.A	NO.B
10 °C	$\Delta R_0$	$r_p$	0.9630	0.9663
		$r_s$	0.9997	0.9997
	$\Delta R_1$	$r_p$	0.9660	0.9445
		$r_s$	0.9667	0.9615
25 °C	$\Delta R_0$	$r_p$	0.9836	0.9892
		$r_s$	0.9997	0.9997
	$\Delta R_1$	$r_p$	0.9891	0.9890
		$r_s$	0.9951	0.9960
40 °C	$\Delta R_0$	$r_p$	0.9922	0.9904
		$r_s$	0.9997	0.9997
	$\Delta R_1$	$r_p$	0.9866	0.9911
		$r_s$	0.9954	0.9945

**Table 2**  
Gibbs sampling.

**Step1:** Initialization  $x^{(0)} \sim q(x)$   
**Step2:** Iteration sampling  
For  $i = 1, 2, \dots, N$   
 $x_1^{(i)} \sim p(X_1 = x_1 | X_2 = x_2^{i-1}, X_3 = x_3^{i-1}, \dots, X_N = x_N^{i-1})$   
 $x_2^{(i)} \sim p(X_2 = x_2 | X_1 = x_1^{i-1}, X_3 = x_3^{i-1}, \dots, X_N = x_N^{i-1})$   
 $\vdots$   
 $x_N^{(i)} \sim p(X_N = x_N | X_1 = x_1^{i-1}, X_2 = x_2^{i-1}, \dots, X_{N-1} = x_{N-1}^{i-1})$   
End  
**Step3:** Removing the “burn-in” sample point  
**Step4:** Obtaining the  $X_1, X_2, X_3, \dots, X_N$

$$\begin{aligned}\Delta C^i &= |C_{IBC} - C^i| \\ \Delta R_0^i &= |R_0^i - R_0^1| \\ \Delta R_1^i &= |R_1^i - R_1^1|\end{aligned}\quad (4)$$

where the  $R_0^1$  and  $R_1^1$  are the initial values of ohmic resistance and polarization resistance, respectively.

To better describe the process of HIs extraction, a flowchart is given as shown in Fig. 5. The process consists of three steps:

- (1) The battery is discharged with dynamic loading profiles, and the Thevenin model is utilized to depict the dynamic and static characteristics of the battery.
- (2) Based on the real-time battery measurement information, the RLS method is employed to identify the model parameters; Meanwhile, SOC is estimated by the combination of the OCV- and Coulomb-counting-based methods.

- (3) The ohmic resistance ( $R_0$ ) and polarization resistance ( $R_1$ ) for 80–30% SOC are calculated by summation and average calculation. Then, the HIs are extracted with subtraction.

To quantitatively evaluate the relevance at the different temperatures between the extracted HIs and the capacity degradation, the Pearson correlation analysis and Spearman rank correlations are utilized. Pearson correlation analysis is the most extensively used quantitative method to determine if there is a kind of linear correlation between two variables, and Spearman rank correlation analysis is a non-parametric test used to measure the strength of association between two variables. What's more, if both the Pearson correlation coefficient ( $r_p$ ) and Spearman's rank correlation coefficient ( $r_s$ ) are approximately equal to 1, it suggests that the two variables are monotonic and linear. The correlation analysis of the extracted HIs and capacity degradation at different temperatures is in Table 1.

In Table 1, all the correlation coefficients are greater than 0.96 at various temperatures. The results manifest the effectiveness and robustness of the extracted HIs. In addition, due to the monotonicity of the battery aging process, Spearman's rank correlation coefficients are relatively closer to 1 regardless of different temperatures. For Pearson correlation coefficients, the linearity of HIs at high temperature is stronger than that at low temperature. In conclusion, it suggests that the extracted HIs are effective and reasonable to describe battery capacity degeneration owing to the robust linearity and monotonicity at various temperatures.

### 3. Method for SOH estimation

#### 3.1. Bayesian multivariate linear regression

Multiple linear regression is one of the statistical tools used to clarify the variability between the response variables ( $y$ ) and explanatory variables ( $x$ ) [27], and multiple linear regression is expressed as:

$$y_t = \beta x_t + \varepsilon_t, t = 1, 2, 3, \dots \quad (5)$$

where  $\beta$  is the regression (slope) coefficients corresponding to the variables  $x_t$ ;  $\varepsilon_t$  is the regression disturbance having a mean of 0 and a variance of  $\sigma^2$ .

The Bayesian approach treats  $\beta$  and  $\sigma^2$  of multiple linear regression as random variables [28]. In short, the probability distributions of the parameters are updated with the help of parameters information composed of a likelihood function and prior distributions. The likelihood function is the parameters information provided by samples, and the likelihood of multiple linear regression with a random sample is

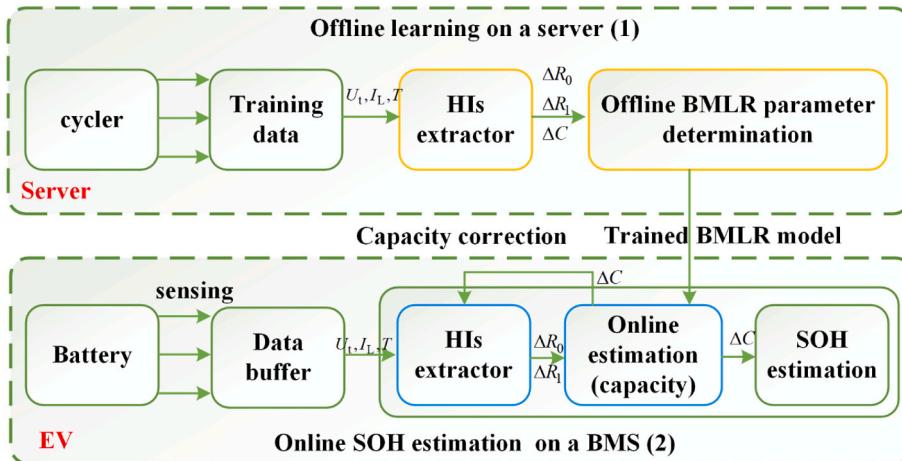


Fig. 6. The framework of the battery SOH estimator.

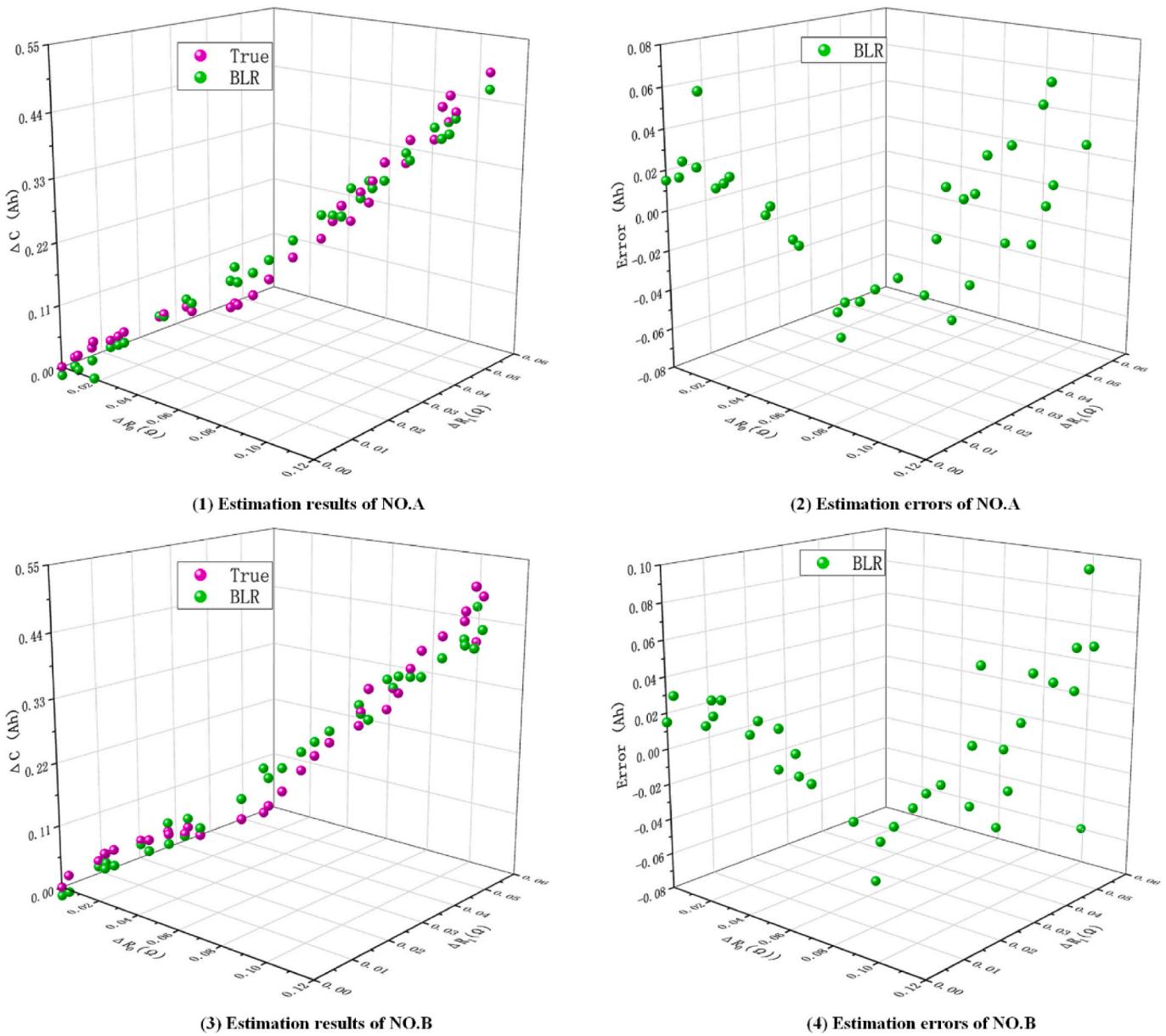


Fig. 7. Capacity degradation estimation using BMLR.

expressed as:

$$\ell(\beta, \sigma^2 | y, x) = \prod_{t=1}^T P(y_t | x_t, \beta, \sigma^2) \quad (6)$$

where  $P(y_t | x_t, \beta, \sigma^2)$  is the conditional probability density function of  $y_t$  given by the parameters and induced by the conditional distribution of  $\varepsilon_t$ . Typically,  $x_t$  is considered as a fixed quantity. If the disturbances are independent, homoscedastic, and Gaussian, then

$$\ell(\beta, \sigma^2 | y, x) = \prod_{t=1}^T \phi(y_t; x_t \beta, \sigma^2) \quad (7)$$

where  $\phi(y_t; x_t \beta, \sigma^2)$  is the Gaussian probability density with mean  $x_t \beta$  and variance  $\sigma^2$ , evaluated at  $y_t$ .

Prior distributions are the distribution of the parameters believed

before observing the data. The confidence of the knowledge about the parameter can be adjusted by adjusting the prior variance. In practice, the prior rather than the real distribution of parameters is preferred as a matter of convenience. For multiple linear regression, prior distributions are typically denoted as  $\pi(\beta)$  and  $\pi(\sigma^2)$ . The joint posterior distribution of  $\beta$  and  $\sigma^2$  is obtained utilizing Bayes' Rule [28]:

$$\pi(\beta, \sigma^2 | y, x) = \frac{\pi(\beta) \pi(\sigma^2) \ell(\beta, \sigma^2 | y, x)}{\int_{\beta, \sigma^2} \pi(\beta) \pi(\sigma^2) \ell(\beta, \sigma^2 | y, x) d\beta d\sigma^2} \propto \pi(\beta) \pi(\sigma^2) \ell(\beta, \sigma^2 | y, x) \quad (8)$$

If  $\beta$  depends on  $\sigma^2$ , then the prior should be replaced with  $\pi(\beta | \sigma^2)$ . The denominator is the distribution of the response given the predictors and is a constant after observing  $y$ . Hence, the posterior is often as being proportional to the numerator.

A posterior is like other joint probability distribution of random

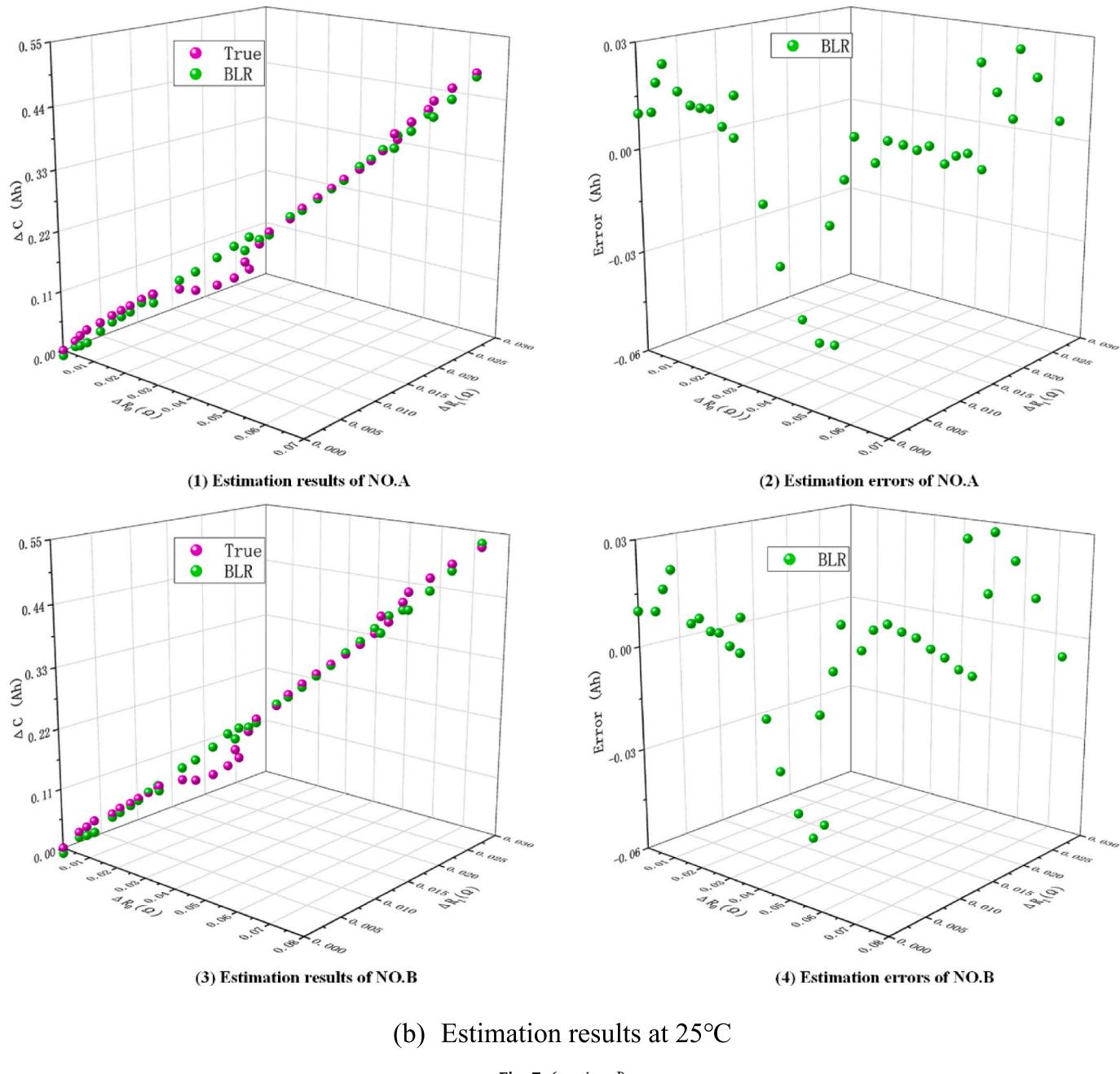


Fig. 7. (continued).

variables, and it has also involved the known information about parameters. Posterior estimation involves integrating functions of parameters concerning the posterior. The expected value of  $\beta$  given the data is

$$\hat{\beta} = E(\beta|y, x) = \int_{\beta, \sigma^2} \beta \pi(\beta, \sigma^2|y, x) d\beta d\sigma^2 \quad (9)$$

This estimated value offers an interpretation, and it also provides the minimum mean squared error owing to a minimum of  $E[(\hat{\beta} - \beta)^2|y, x]$ . Therefore, other losses, the median, mode, or a quantile can be a Bayes estimator.

The maximum a priori estimate the value of the parameter that maximizes the posterior distribution.

Given the data, the predicted response  $\hat{y}$  of the explanatory variable  $\hat{x}$  is a random variable with posterior predictive distribution [29]:

$$\pi(\hat{y}|y, x, \hat{x}) = \int_{\beta, \sigma^2} f(\hat{y}|\beta, \sigma^2, \hat{x}) \pi(\beta, \sigma^2|y, x) d\beta d\sigma^2 \quad (10)$$

The outcome can be considered as the conditional expected value of the probability distribution of  $y$  concerning the posterior distribution of the parameters.

### 3.2. Gibbs sampling

Integration methods in Eq. (10) rely on the functional form of the product  $\pi(\beta)\pi(\sigma^2)\ell(\beta, \sigma^2|y, x)$  in Eq. (8). However, the posterior cannot be tractable analytically due to the unknown probability distribution of the intricate battery aging process. Under certain conditions, numerical integration can be implemented using Monte Carlo or Markov Chain Monte Carlo (MCMC) sampling [30]. The statistic of a posterior distribution can be obtained with  $N$  simulated samples from the distribution:

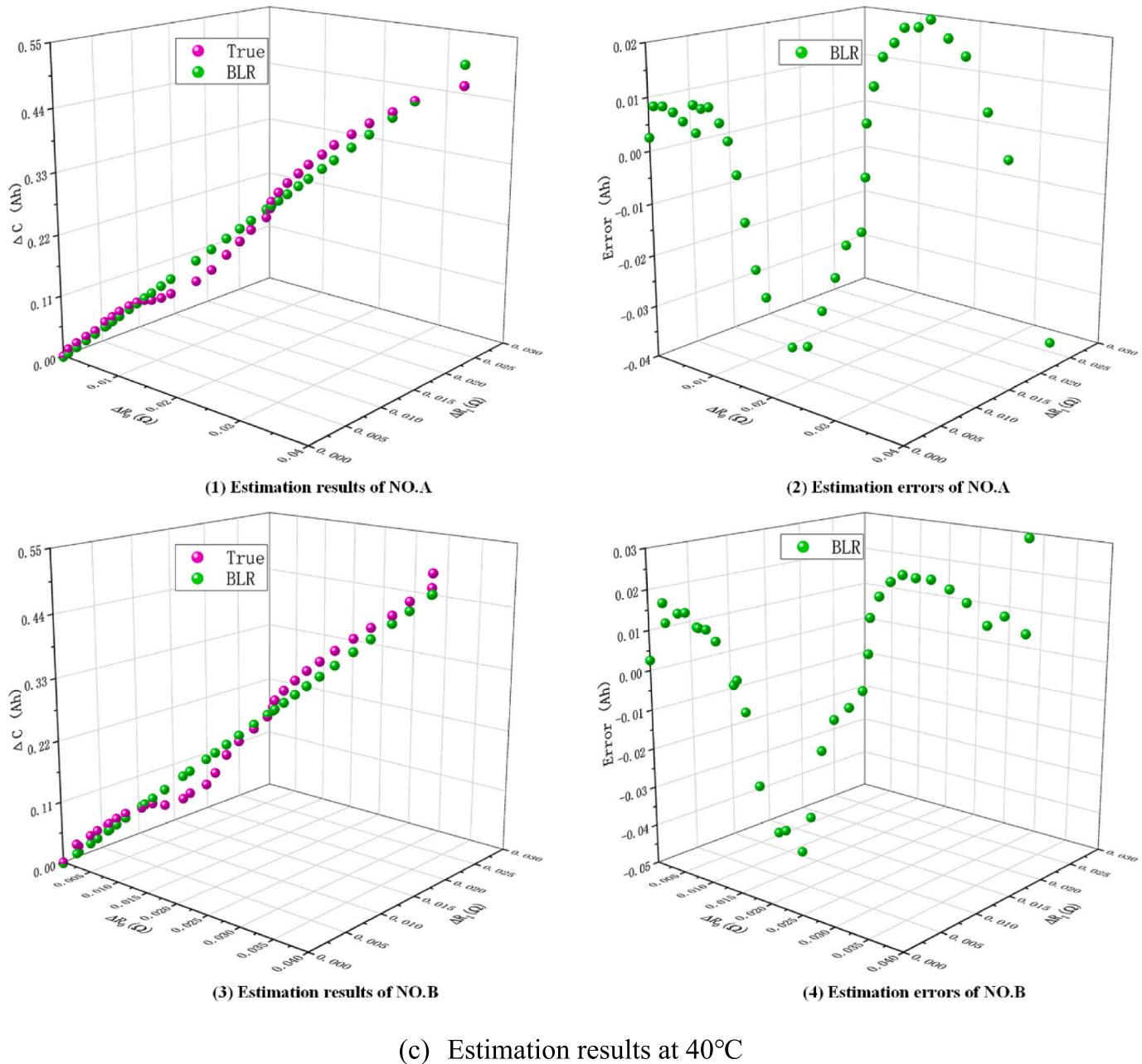


Fig. 7. (continued).

**Table 3**  
Comparison results of capacity degradation estimation.

Temperatures	NO.A				NO.B			
	Max	MAE	RMSE	STD	Max	MAE	RMSE	STD
10 °C	0.0641	0.0268	0.0323	0.0328	0.0995	0.0310	0.0378	0.0384
25 °C	0.0562	0.0158	0.0215	0.0217	0.0572	0.0154	0.0214	0.0217
40 °C	0.0357	0.0153	0.0181	0.0184	0.0439	0.0178	0.0209	0.0207

$$E[f(s)]_p \approx \frac{1}{N} \sum_{i=1}^N f(s^i) \quad (11)$$

where  $p$  is the posterior distribution of interest,  $f(s)$  is the desired expectation.

The Gibbs sampling was proposed in the early 1990s, as one of the popular MCMC techniques, is introduced in this paper. The main

concept of Gibbs sampling is to settle the problem of sampling from the high-dimensional joint distribution into a series of samples from low-dimensional conditional distributions. The Markov chain generated by Gibbs sampling is assumed to be ergodic, thus, if the sampling number is large enough, the samples would reach stationary distribution regardless of the initial sample values. According to the sampling theorem, these sample points are also subject to the posterior edge probability

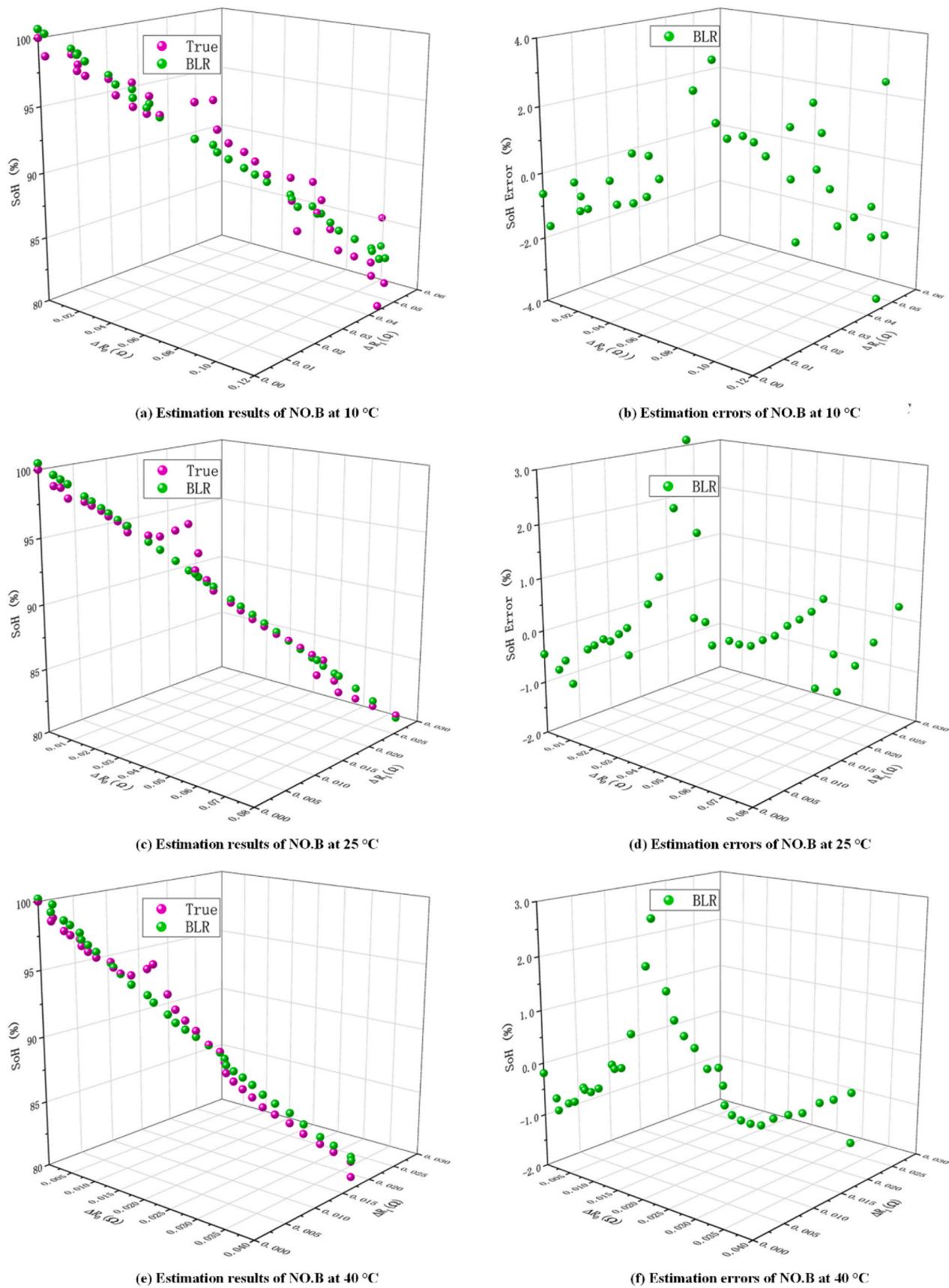


Fig. 8. SOH estimation results of NO-B using BMLR.

**Table 4**  
Comparison results of SOH estimation.

Temperatures	NO.B			
	Max/%	MAE	RMSE	STD
10 °C	3.74	0.123	0.0150	0.0153
25 °C	3.50	0.062	0.0094	0.0095
40 °C	2.78	0.070	0.0089	0.0089

density function of each set of parameters. The steps of the Gibbs sampling are shown in Table 2 [31].

Therefore, the posterior of BMLR in Eq. (10) can be used to get the numerical solution by resorting to Gibbs sampling.

### 3.3. Battery SOH estimator

To realize online SOH estimation, a developed framework based on the proposed BMLR and extracted HIs is shown in Fig. 6.

The estimation framework is divided into two parts:

(1) Offline learning on a server:

(I) Through charging and discharging a reference battery circularly, the cycler data is collected as the training date.

(II) Based on the real-time measured battery information, including the terminal voltage ( $U_t$ ), load current ( $I_L$ ), and temperatures ( $T$ ), the HIs extractor, as shown in Fig. 5, is employed to extract the battery HIs ( $\Delta R_0$  and  $\Delta R_1$ ), meanwhile, the battery capacity fading ( $\Delta C$ ) is extracted through Eq. (4).

(III) Through the training and cross-validation, the parameters of BMLR are determined. Then trained BMLR model is input to the BMS in EV for the next SOH estimation.

(2) Online SOH estimation on a BMS in EV:

(I) When a verification battery is undergoing dynamic discharging, the real-time battery information ( $U_b$ ,  $I_L$ , and  $T$ ) is sensed by a series of sensors, and the battery data is stored in a data buffer.

(II) The battery data in the data buffer is processed by the HIs extractor (see Fig. 5) to extract the battery HIs ( $\Delta R_0$  and  $\Delta R_1$ ).

(III) The extracted HIs are the input of the trained BMLR model for online capacity estimation. On the one hand, the output ( $\Delta C$ ) of the BMLR model is employed to calculate the SOH. On the other hand, the  $\Delta C$  is utilized to update the battery SOC estimation as shown in Eq. (3) in the HI extractor.

So far, a BMLR-based framework is developed to estimated battery SOH online through the offline training data from the reference battery and online measured battery from the verification battery.

## 4. Results and discussion

In this section, the estimation results based on BMLR are exhibited and discussed. It should be noted that the reference battery is NO.A and validation battery is NO.B, respectively. The battery capacity degradation is estimated firstly, then, the remaining battery capacity ( $\hat{C}_{RBC}$ ) is computed. Finally, the battery SOH is computed based on the computed battery capacity according to Eq. (12).

$$\hat{S}OH_i = \frac{\hat{C}_{RBC}}{C_{IBC}} \times 100\% = \frac{C_{IBC} - \Delta \hat{C}^i}{C_{IBC}} \times 100\% \quad (12)$$

where the  $\hat{C}_{RBC}^i$ ,  $\hat{S}OH_i$  and  $\Delta \hat{C}^i$  is the  $i$ th estimates of the remaining battery capacity, SOH, and capacity degradation, respectively.

### 4.1. Capacity degradation estimation

It is significant to explore the effectiveness and robustness of the proposed BMLR for battery SOH estimation. The HIs ( $\Delta R_0$  and  $\Delta R_1$ ) at various temperatures are extracted primarily based on the aging data of

the reference battery, and the developed estimator can then be applied to estimate the capacity degradation ( $\Delta C$ ) with the extracted HIs of the validation battery. Herein, the actual aging data used are processed ahead of time by low-pass filtering, then, the estimation results obtained using BMLR at different temperatures are shown in Fig. 7.

As seen in Fig. 7, as the increment of ohmic resistance ( $\Delta R_0$ ) and polarization resistance ( $\Delta R_1$ ), the battery capacity degradation ( $\Delta C$ ) also increases with an approximately linear trend, and the linear trend becomes increasingly visible with a rise of temperature. Additionally, it can be seen that some individual points deviate from the mainstream degenerate trajectory [see Fig. 7, sub-graphs (1)], such cases are more likely caused by the extremely complex electrochemical reactions and side reactions. However, the trajectories of estimated battery capacity degradation are approximate linear distributing along with the extracted HIs. The estimated capacity degradation ( $\hat{\Delta C}$ ) of the reference battery has a good performance on tracing actual capacity degradation at various temperatures due to the approximately linear relationship between HIs and capacity degradation. Furthermore, the tracking effects at 25 °C and 40 °C are better than the tracking effect at 10 °C owing to the higher linearity at high temperatures [see Fig. 7, sub-graphs (2)]. Nonetheless, the overfitting has not occurred during determining BMLR parameters. Therefore, there is a good performance when the BMLR with determined parameters is exploited to estimate the battery capacity degradation of the validation battery [see Fig. 7, sub-graphs (3)]. The estimation errors are also kept within a reasonable range [see Fig. 7, sub-graphs (4)]. Like the reference battery, the estimation results are better at higher temperatures.

To display the performance on battery capacity degradation estimation, some performance indices, including the maximum error (Max), mean absolute error (MAE), root mean square error (RMSE), and error standard deviation (STD) are introduced. The performance indices which are from statistics are often used to evaluate the differences between an estimator and the observed actual values. In addition, the results are also given to compare with the proposed BMLR. The detailed comparison is shown in Table 3.

As shown in Table 3, It is concluded that: (1) At different temperatures, the BMLR has a good performance on the reference battery and verification battery, respectively; For instance, at 10 °C, the error standard deviations for the two batteries based on BMLR are 0.00323 and 0.0384, respectively. (2) With the increase of temperature, Table 3 also reveals that there has been a gradual increase in estimation accuracy with the temperature. Using the same BMLR model, the maximum errors of NO.B are 0.0995, 0.0572, and 0.0439, respectively, at the three temperatures. This phenomenon might be related to the effect of temperature on electrochemical reaction kinetics.

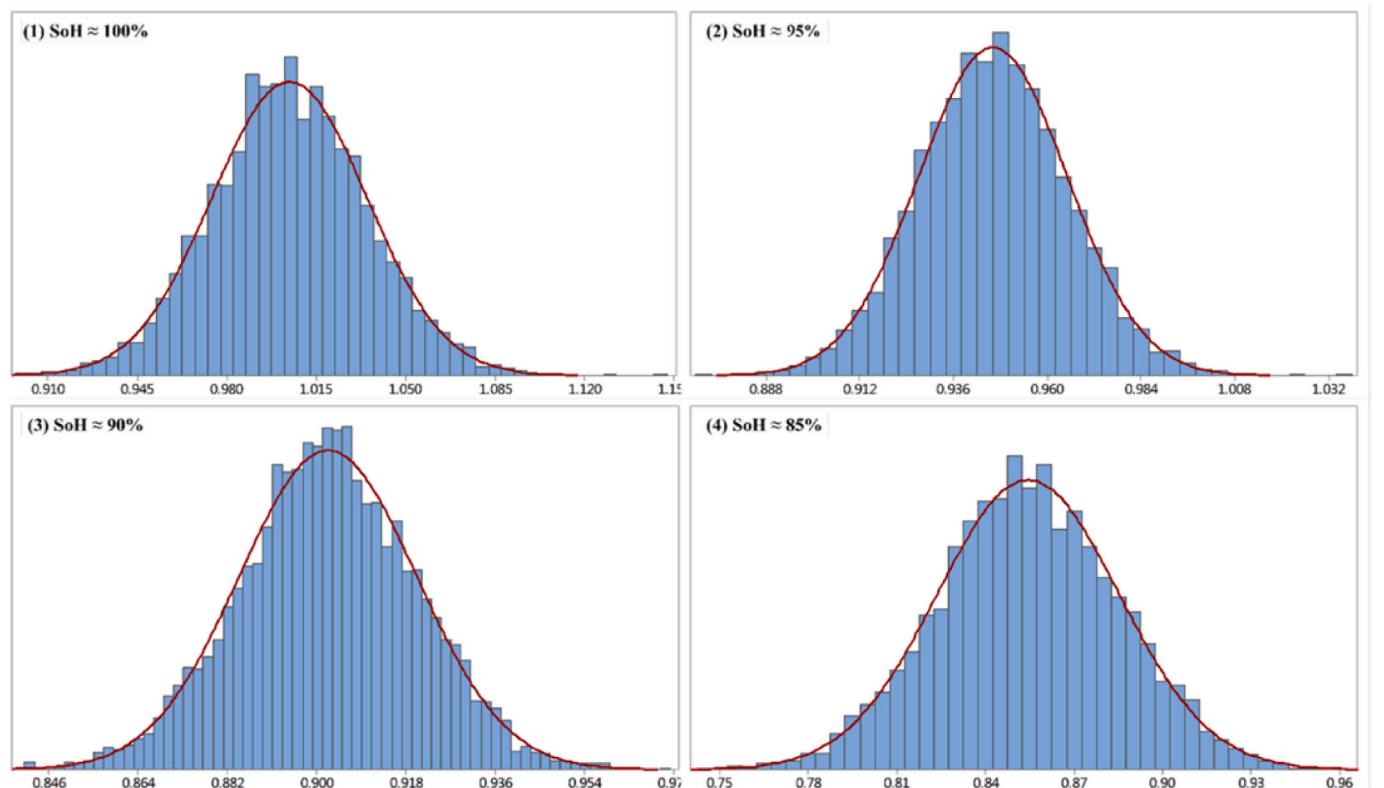
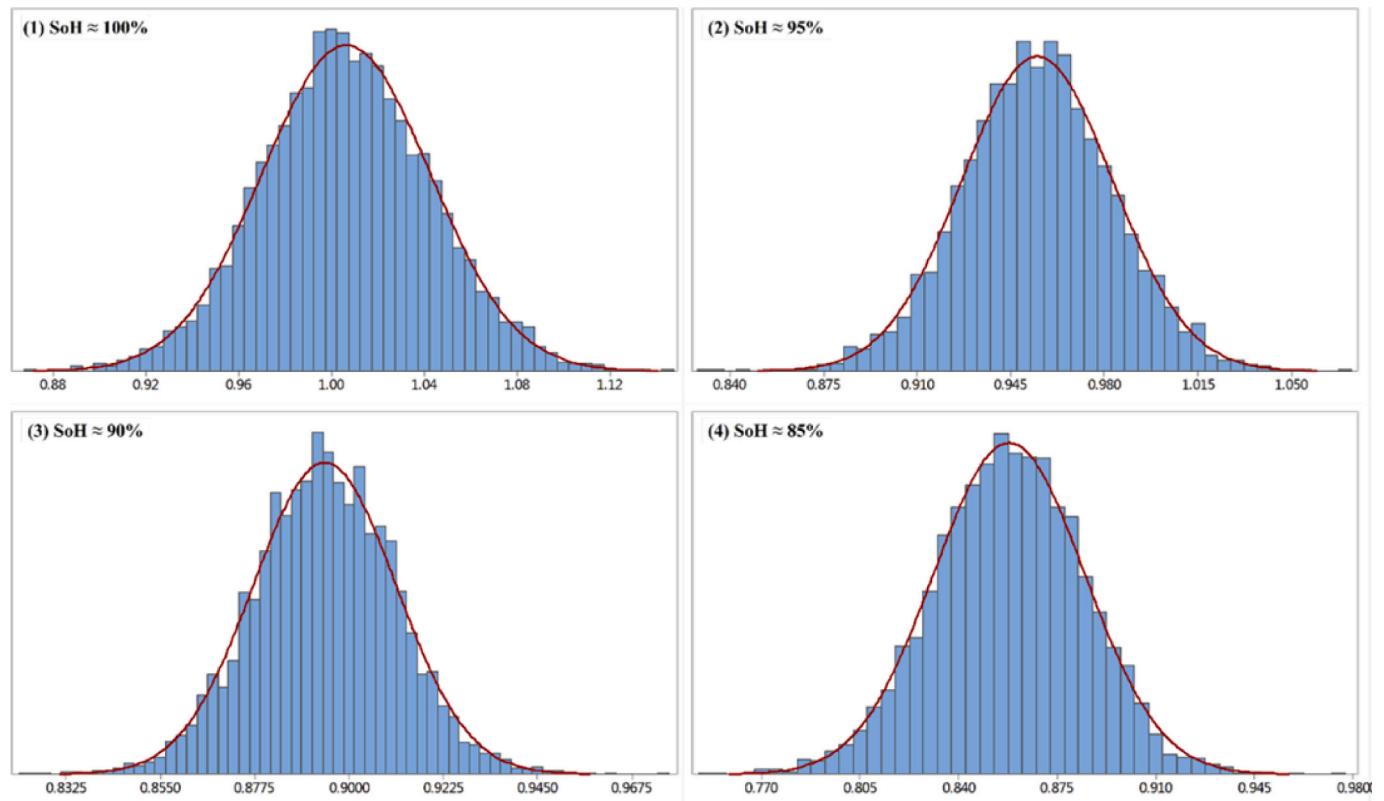
Given the above, the validities of the extracted HIs and the developed SOH estimation framework are verified. The extracted HIs can trace the trajectory of battery capacity degradation effectively, and the developed estimation framework has strong robustness for various temperatures.

### 4.2. SOH estimation

Based on the calculated battery capacity degradation the SOH can be estimated according to Eq. (12). The SOH estimation results of battery NO.B using BMLR are shown in Fig. 8. There is a good agreement between the estimated SOH and the true SOH values at different temperatures [see Fig. 8 (a, c, e)], nevertheless, the performance at 25 °C and 40 °C is relatively better than those at low temperature. The margins of SOH errors using BMLR are in the region of [-4.0%, 4.0%] [see Fig. 8 (b)], [-2.0%, 4.0%] [see Fig. 8 (d)], and [-2.0%, 3.0%] [see Fig. 8 (f)], respectively, at low and high temperatures.

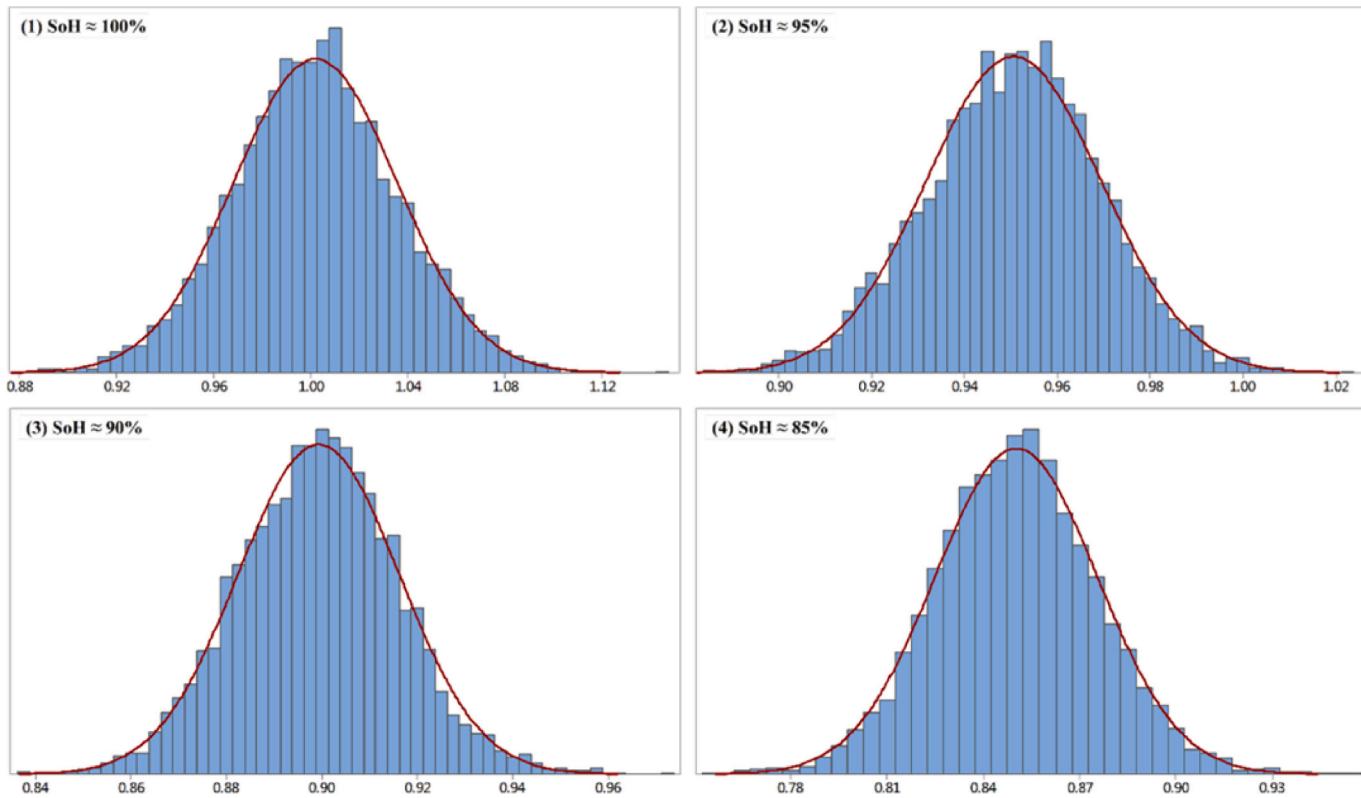
In addition, the evaluation indices are introduced again to evaluate the performance of the linear regression techniques, the results are summarized in Table 4.

As shown in Table 4, the maximum SOH errors are less than 4.0%,



(b) SOH Probability density distribution of NO.B at 25°C

Fig. 9. SOH Probability density distribution of BMLR.



(c) SOH Probability density distribution of NO.B at 40°C

Fig. 9. (continued).

**Table 5**  
Comparison results of SOH estimation.

SOH	Temperatures	95% CI
100%	10 °C	[0.9352, 1.0775]
	25 °C	[0.9450, 1.0640]
	40 °C	[0.9358, 1.0681]
95%	10 °C	[0.9122, 0.9933]
	25 °C	[0.9087, 0.9834]
	40 °C	[0.9136, 0.9876]
90%	10 °C	[0.8676, 0.9333]
	25 °C	[0.8671, 0.9372]
	40 °C	[0.8668, 0.9324]
85%	10 °C	[0.7980, 0.8985]
	25 °C	[0.8028, 0.9033]
	40 °C	[0.8007, 0.8997]
80%	10 °C	[0.7646, 0.9115]
	25 °C	[0.7002, 0.8964]
	40 °C	[0.7369, 0.8907]

and it is even lower when the temperature is higher, such as a maximum of 3.74 at 10 °C while 2.78 at 40 °C. The MAE and RMSE, as the parameters to illustrate the dispersion degree and convergence performance, show that the BMLR has good accuracy and robustness on SOH estimation at various temperatures. In addition, the estimation error of BMLR also fluctuates less with changes from the point of STD with the values of 0.0153, 0.0095, and 0.0089, respectively, at the three different temperatures.

As mentioned above, the BMLR can provide the results with probability-based form. The estimation results at different aging levels (about 100%, 95%, 90%, and 85% of SOH) and temperatures (10 °C, 25 °C, and 40 °C) are given in Fig. 9.

As can be seen from Fig. 9, the estimation results of BMLR at certain SOH are a normal distribution from the fitted normal curves.

Additionally, the normal test and the Probability Density Distributions (PDFs) of the estimation results are also given. Through the graphical description, The PDFs provides a statistical measure to evaluate the SOH estimation. In the PDFs, the area under the red lines can represent the interval in which the values of estimated SOH fall. The bars provide the probability of a discrete random SOH estimate occurring. The peaks also indicate the mean SOH estimation. Thus, these normal distribution parameters can well illustrate the normality of SOH estimates. Meanwhile, the 95% confidence intervals (CIs) calculated by Eq. (13) are listed in Table 5. The interval estimation that provides a range of SOH estimation that may contain the unknown parameter can express the uncertainty of SOH estimation effectively. To sum up, the PDFs and the CIs transform the point estimate into an interval estimate, the transformations can effectively ensure the accuracy and robustness of SOH estimation.

$$\text{bounds of 95\%confidence range} = \overline{\text{SOH}}_{\text{estimate}} \pm 1.96 \times \text{cov}(\text{SOH}_{\text{estimate}}) \quad (13)$$

#### 4.3. Discussion

The good performances of the battery SOH estimation results suggest the effectiveness of the extracted HIs and the SOH estimation framework. As for the HIs, they can be extracted easily through the Thevenin model and recursive least-squares method to estimate battery capacity degradation. Furthermore, the extracted HIs are closely related to capacity degradation. The correlation analysis results show that there are good linearity and monotonicity between the extracted health indicator and capacity degradation at various temperatures. The phenomena also imply that they are most likely attributed to the same aging mechanism, such as loss of Li-ion inventory and loss of active electrode material [32]. The solid electrolyte interphase growth and decomposition, lithium plating, and electrolyte decomposition may consume the free li-ions

[33]. The active mass of the electrode would also reduce for the insertion and release of lithium due to the electrical contact and other reasons. The loss of Li-ion inventory is one of the prime contributors to capacity degradation. Meanwhile, the loss of active electrode material not only leads to capacity degradation but also increase the battery impedance. Therefore, the same mechanistic origins determine that the resistance-based HIs can be used to quantify capacity degradation effectively.

In terms of methods, the proposed BMLR has better performance on battery capacity degradation. From a mathematical point of view, the regression (slope) coefficients in the BMLR are considered as random variables instead of unknown variables. The data utilization BMLR can be virtually 100%, only using the training sample can effectively and accurately determine the complexity of the model. In addition, the overfitting in maximum likelihood estimation is nonexistent for BMLR. The probability-based results also elaborate on the uncertainty and robustness of the proposed BMLR [28]. Thus, it is seen that there are fluctuations in the first half of degradation trajectories, and the maximum errors occurred in the same place (Fig. 7 and Fig. 8). The primary cause is capacity regeneration, the battery capacity regeneration phenomenon represents a sudden increment in the available capacity that happens at the next cycles, and the phenomenon is in connection with various physicochemical aspects, such as loading profiles and working temperatures. Especially, the capacity regeneration phenomenon is particularly significant at low temperature because of the temperature on the kinetics electrochemical reactions.

Therefore, the advantages of the extracted HIs and the proposed BMLR are beneficial for the online implementation of the proposed framework for SOH estimation for EVs.

## 5. Conclusion

Combine the advantage of the adaptive approach and data-driven

approach, a joint model-based and BMLR-oriented framework was developed for online SOH estimation. And a series of battery aging tests to simulate the dynamic working condition of electric vehicles are implemented covering different temperature ranges. To quantify the battery capacity degeneration effectively, the multiple extracted HIs based on the battery Thevenin model were verified through Pearson correlation analysis and Spearman correlation analysis. To ensure accurate and robust SOH estimation, a BMLR is introduced to develop a two-section framework for SOH estimation. The experimental results show that both the maximum estimation errors are less than 4.0% and the mean absolute errors are approximately equal to 1% at various temperatures. In the future, the capacity regeneration phenomenon and lower temperatures will be paid more attention to.

## CRediT authorship contribution statement

**Zhiqiang Lyu:** Conceptualization, Methodology, Software, Writing - original draft; **Geng Wang:** Writing - review & editing, Visualization; **Cao Tan:** Supervision, Project administration, 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.

## Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 51905319).

## Appendix A. Appendix of RLS

The electrical behavior of the Thevenin model can also be written as follows:

$$\begin{cases} OCV = U_t + I_L R_0 + U_1 \\ I_L = \frac{U_1}{R_1} + C_1 \frac{dU_1}{dt} \end{cases} \quad (A-1)$$

where  $OCV$  denotes the OCV and  $I_L$  represents the load current.  $R_1$  and  $C_1$  are, respectively, the polarization resistance and polarization capacitance of the RC network,  $R_0$  is the ohmic internal resistance,  $U_t$  is the terminal voltage and  $U_1$  is the voltage across the parallel RC network.

Further, the Eq. (A-1) can be rewritten as follows in the frequency domain:

$$U_t(s) = OCV(s) - I_L(s) \left( R_0 + \frac{R_1}{1 + R_1 C_1 s} \right) \quad (A-2)$$

Define  $E_t = U_t - OCV$ , the transfer function  $G(s)$  of Eq. (A-2) can be written as follows:

$$G(s) = \frac{E_t(s)}{I_L(s)} = -R_0 - \frac{R_1}{1 + R_1 C_1 s} = -\frac{R_0 + R_1 + R_0 R_1 C_1 s}{1 + R_1 C_1 s} \quad (A-3)$$

A bilinear transformation method is employed for the discretization calculation of Eq. (A-4) and the result is shown in Eq. (A-4).

$$\begin{cases} G(z^{-1}) = -\frac{\frac{R_0 T + R_1 T + 2R_0 R_1 C_1}{T + 2R_1 C_1} + \frac{R_0 T + R_1 T - 2R_0 R_1 C_1}{T + 2R_1 C_1} z^{-1}}{1 + \frac{T - 2R_1 C_1}{T + 2R_1 C_1} z^{-1}} \\ s = \frac{2}{T} \frac{1 - z^{-1}}{1 + z^{-1}} \end{cases} \quad (A-4)$$

$$\text{Define } \begin{cases} a_1 = -\frac{T - 2R_1 C_1}{T + 2R_1 C_1} \\ a_2 = -\frac{R_0 T + R_1 T + 2R_0 R_1 C_1}{T + 2R_1 C_1} \\ a_3 = -\frac{R_0 T + R_1 T - 2R_0 R_1 C_1}{T + 2R_1 C_1} \end{cases}$$

$$\begin{cases} R_0 = \frac{(a_3 - a_2)}{1 + a_1} \\ R_1 = \frac{2(a_3 - a_1 a_2)}{a_1^2 - 1} \\ C_1 = \frac{-(a_1 T + T)(a_1 + 1)}{4(a_3 + a_1 a_2)} \end{cases} .$$

Eq. (A-2) is rewritten as Eq. (A-5) after discretization, where  $k = 1, 2, 3\cdots$ .

$$E(k) = a_1 E(k-1) + a_2 I_L(k) + a_3 I_L(k-1) \quad (\text{A-5})$$

And then Eq. (A-5) is rewritten as Eq. (A-6).

$$\begin{aligned} U_t(k) &= OCV(k) - a_1 OCV(k-1) + a_1 U_t(k-1) + a_2 I_L(k) + a_3 I_L(k-1) \\ &= (1 - a_1) OCV(k) + a_1 U_t(k-1) + a_2 I_L(k) + a_3 I_L(k-1) \end{aligned} \quad (\text{A-6})$$

Define  $\begin{cases} H(k) = [1 \quad U_t(k-1) \quad I_L(k) \quad I_L(k-1)] \\ \theta(k) = [(1 - a_1) OCV(k) \quad a_1(k) \quad a_2(k) \quad a_3(k)]^T \end{cases}$ , then.

$$Y(k) = U_t(k) = H(k)\theta(k) \quad (\text{A-7})$$

In case of online application, the  $U_t(k)$  and  $I_L(k)$  are sampled at constant period, the vector  $\theta$  can be identified by a recursive least square algorithm according to Eq. (15), and then the model parameters can be solved by the expressions of  $a_1, a_2, a_3$ .

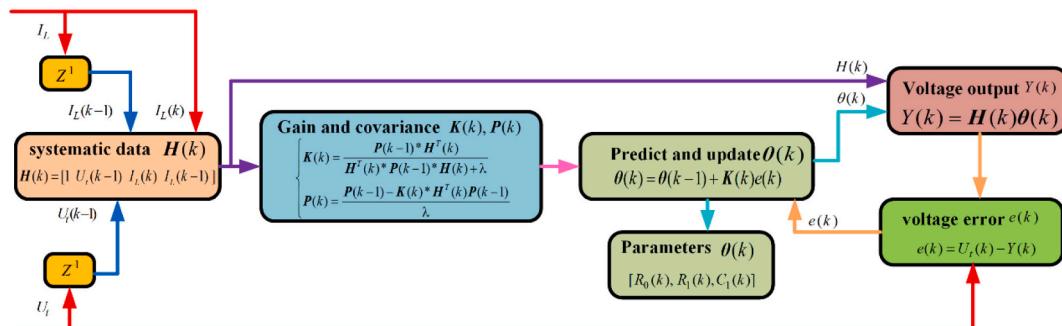
$$Y(k) = H(k)\theta(k) + e(k) \quad (\text{A-8})$$

With some modifications and improvements, the frequently-used RLS method can be used to optimize the system model. The problem of a standard RLS is that it usually results in the saturation phenomenon. This disadvantage is owing to the exponential growth of the covariance matrix. Therefore, based on the standard RLS method, an optimal forgetting factor is employed to give less weight to old data and more weight to recent data, the parameter estimates are updated at every sample interval on the basis of current and terminal voltage measurement at time  $k$  and previous estimates at time ( $k-1$ ). The system identification is realized as follows:

$$\left\{ \begin{array}{l} e(k) = U_t(k) - H(k)\theta(k-1) \\ \mathbf{K}(k) = \frac{\mathbf{P}(k-1)^* \mathbf{H}^T(k)}{\mathbf{H}^T(k)^* \mathbf{P}(k-1)^* \mathbf{H}(k) + \lambda} \\ \mathbf{P}(k) = \frac{\mathbf{P}(k-1) - \mathbf{K}(k) \mathbf{H}^T(k) \mathbf{P}(k-1)}{\lambda} \\ \theta(k) = \theta(k-1) + \mathbf{K}(k) e(k) \end{array} \right. \quad (\text{A-9})$$

where,  $e(k)$  is the prediction error of the battery terminal voltage.  $H(k)$  is the matrix of systematic data and  $\theta(k)$  is the matrix of systematic parameters.  $H(k)$  and  $\theta(k)$  are defined and preset [30].  $K(k)$  is the algorithm gain and  $P(k)$  is the covariance matrix at the moment  $k$ ,  $\lambda$  is the forgetting factor, typically  $\lambda \in [0.95, 1]$ .

The schematic diagram for the online identification of the Thevenin model parameters is shown in Fig. 8. The initial values of the parameter vector  $\theta(1)$  and its error covariance matrix  $P(1)$  are preset. Based on load current  $I_L(k)$ , and terminal voltage  $U_t(k)$  which are measured, the values of  $\theta(k)$  at each moment can be adjusted adaptively.



A-Fig. 1: Scheme of the identification of the Thevenin model parameters.

## References

- [1] M. Maures, A. Capitaine, J.Y. Delétage, J.M. Vinassa, O. Briat, Microelectron. Reliab. 114 (2020), 113798.
- [2] K. Laadjal, A.J.M. Cardoso, Electronics 10 (2021) 1588.
- [3] X. Qu, Y. Song, D. Liu, X. Cui, Y. Peng, Microelectron. Reliab. 114 (2020), 113857.
- [4] M. Ge, Y. Liu, X. Jiang, J. Liu, Measurement 174 (2021), 109057.
- [5] M.S.H. Lipu, M.A. Hannan, A. Hussain, M.M. Hoque, P.J. Ker, M.H.M. Saad, A. Ayob, J. Clean. Prod. 205 (2018) 115.
- [6] X. Hu, F. Feng, B. Liu, K. Liu, L. Zhang, J. Xie, Renew. Sust. Energ. Rev. 114 (2019), 10934.
- [7] X. Hu, J. Jiang, D. Cao, B. Egardt, IEEE Trans. Ind. Electron. 63 (2016) 2645.
- [8] H. Tian, P. Qin, K. Li, Z. Zhao, J. Clean. Prod. 261 (2020), 120813.
- [9] A. Bartlett, J. Marcicki, S. Onori, G. Rizzoni, X.G. Yang, T. Miller, IEEE Trans. Control Syst. Technol. 24 (2016) 384.
- [10] R. Xiong, L. Li, Z. Li, Q. Yu, H. Mu, Appl. Energ. (2018) 219.
- [11] R. Xiong, L. Li, J. Tian, J. Power Sources 405 (2018) 18.
- [12] W. Waag, C. Fleischer, D.U. Sauer, J. Power Sources 258 (2014) 321.
- [13] Y. Zou, X. Hu, H. Ma, S.E. Li, J. Power Sources 273 (2015) 793.
- [14] D. Andre, C. Appel, T. Soczka-Guth, D.U. Sauer, J. Power Sources 224 (2013) 20.
- [15] S. Schwunk, N. Armbruster, S. Straub, J. Kehl, M. Vetter, J. Power Sources 239 (2013) 705.
- [16] J. Bi, T. Zhang, H. Yu, Y. Kang, J. Yan, Appl. Energ. 182 (2016) 558.
- [17] D. Yang, X. Zhang, R. Pan, Y. Wang, Z. Chen, J. Power Sources 384 (2018) 387.
- [18] G.W. You, S. Park, D. Oh, IEEE Trans. Ind. Electron. 64 (2017) 4885.
- [19] P. Guo, Z. Cheng, L. Yang, J. Power Sources 412 (2019) 442.
- [20] R.R. Richardson, C.R. Birk, M.A. Osborne, D.A. Howey, Ieee T. IndInform 15 (2019) 127.
- [21] D. Liu, J. Zhou, H. Liao, Y. Peng, X. Peng, IEEE Trans. Syst.Man Cybern. Syst. 45 (2015) 915.
- [22] D. Liu, H. Wang, Y. Peng, W. Xie, H. Liao, Energies 6 (2013) 3654.
- [23] L. Wang, C. Pan, L. Liu, Y. Cheng, X. Zhao, Appl. Energ. 168 (2016) 465.
- [24] S. Torai, M. Nakagomi, S. Yoshitake, S. Yamaguchi, N. Oyama, J. Power Sources 306 (2016) 62.

- [25] H. Pan, Z. Lü, H. Wang, H. Wei, L. Chen, Energy 160 (2018) 466.
- [26] R. Xiong, W. Shen, Advanced Battery Management Technologies for Electric Vehicles, Wiley, Singapore, 2018.
- [27] S.T. Rachev, J.S.J. Hsu, B.S. Bagasheva, F.J. Fabozzi, Bayesian Linear Regression Model, John Wiley & Sons Inc, 2015 chapter 4.
- [28] R.C. Marin J M, Bayesian Core: A Practical Approach to Computational Bayesian Statistics, Springer Science & Business Media, 2007.
- [29] T. Chen, E. Martin, Anal. Chim. Acta 631 (2009) 13.
- [30] Y.J. Jung, J.P. Hobert, Stat. Probab. Lett. 95 (2014) 92.
- [31] L. Martino, V. Elvira, G. Camps-Valls, Digit. Signal Process. 74 (2018) 1.
- [32] C.R. Birk, M.R. Roberts, E. Mcturk, P.G. Bruce, D.A. Howey, J. Power Sources 341 (2017) 373.
- [33] A. Barré, B. Deguilhem, S. Grolleau, M. Gérard, F. Suard, D. Riu, J. Power Sources 241 (2013) 680.