

Online Model Identification for State of Charge Estimation for Lithium-ion Batteries with Missing Data

Hao Jin, Ling Mao*, Keqing Qu, Jinbin Zhao, Fen Li

School of Electric Engineering, Shanghai University of Electric Power, Shanghai 200090, China

*E-mail: maoling2290@shiep.edu.cn

Received: 11 October 2022/ Accepted: 14 November 2022/ Published: 27 December 2022

Online model identification is critical for determining the accurate state of charge (SOC) for a battery based on a model, which depends on the use of complete and reliable measurement data. To ensure precision for parameter identification and reliability for SOC estimation in practical applications, the occurrence of data loss and noise interference must be considered. In this paper, a first-order resistor–capacitor equivalent circuit model is developed to simulate the “black box” system of a battery. A recursive least square method based on a variable interval auxiliary model is proposed to compensate for the missing data in an unreliable actual environment. Meanwhile, a forgetting factor is introduced to prevent the influence of historical data in parameter identification. To further reduce the noise effects on SOC estimation, the extended Kalman filter (EKF) is applied to the algorithm. The proposed method is verified using CALCE experimental data. The experimental results show that the proposed method can be used to realize accurate model parameter identification and reliable online SOC estimation under conditions of data loss and noise interference.

Keywords: lithium-ion battery; model parameter identification; data loss; state of charge; auxiliary model

1. INTRODUCTION

Battery energy storage systems are widely used in electric vehicles (EVs), smart grids and many other fields in response to the need to reduce carbon emissions and improve energy efficiency. Among them, lithium-ion batteries (LIBs) stand out due to their high energy density, long cycle life and low self-discharge rate [1]. However, with the widespread application of battery energy storage equipment, the safety and reliability problems of LIBs are becoming increasingly prominent [2]. To ensure the safety and reliability of a battery, an advanced battery management system (BMS) is required to monitor and manage the various battery states [3], such as state of charge (SOC), state of health (SOH) and state of power (SOP). This can ensure the safety of the battery, prevent the occurrence of overcharge and discharge and make full use of the best battery performance. The SOC

represents the remaining available battery power as a percentage of the total capacity, which is one of the most important states in a BMS, and provides an important reference for battery safety management, charge and discharge control, energy management and other functions of a battery system.

The LIB SOC estimation methods mainly include the ampere hour integration method [5], model-based method [6] and data-driven method [7]. Among them, the model-based SOC estimation method is widely considered because of its high accuracy and robustness [8]. Model-based SOC estimation methods can be divided into equivalent circuit models (ECMs), electrochemical models and neural network models [9-11]. The ECM is widely studied due to its high accuracy and low calculation cost [12,13].

The accuracy of SOC estimation based on a ECM depends on the result of circuit parameter identification. At present, circuit model parameter identification methods mainly include RLS [14], PLSLS [15], FRLS [16], RTLS [17], genetic algorithms [18], etc. The RLS algorithm shows good performance for online parameter identification because of its advantages in terms of operation speed. In addition, to prevent the influence of historical data, these parameter identification methods introduce the forgetting factor, which also correspondingly increases the volatility in the parameter identification results.

In brief, the above online parameter identification methods can all show good performance in a laboratory environment, but there are still many challenges for practical applications. Due to defects in the sensor, noise and electromagnetic interference, the actual voltage and current measurement data contain a large amount of noise data [22], and when the amount of data reaches a certain scale, abnormal data will be produced [23] and data loss will occur in SOC [24] and SOH estimation [25]. In addition, to enhance the LIB system management, the collected battery data are transmitted to the cloud data management platform through a network [26-30], and the problems in the network environment can also lead to data loss [31,32]. For example, the data in the battery management unit (BMU) of EVs is transmitted through the in-vehicle network because of its advantages of reduced wiring, faster communication speed and easier troubleshooting [33]. However, due to physical constraints and cost considerations, the communication resources allocated to each electric vehicle are usually limited, which can lead to network congestion or even data loss when the BMU sends increasingly more data [34]. These disturbances greatly affect the accuracy of model parameter identification and further reduce the performance for SOC estimation. However, the above methods do not consider the influence of these disturbances on the parameter identification results. Therefore, how to obtain the ideal SOC estimation result in the case of a noisy environment and missing data is an urgent problem to be solved, which is the motivation of this study.

To address the noise problem, many scholars have carried out in-depth research. Hu et al. [35] proposed a dual fractional extended Kalman filter to estimate the SOC, which has a certain accuracy and robustness. Jiang et al. [36] proposed a new adaptive square root extended Kalman filter algorithm, and the Sage-Husa adaptive filter was used to update the noise variables, which also leads to good accuracy and robustness for SOC estimation. However, the above two methods ignore the importance of parameter identification in the circuit model. Wei et al. [37] estimated the noise variance through the cooperation of least squares (LS) and a variable projection algorithm (VPA) to improve the

noise immunity performance for parameter identification, but its effectiveness in a LIB has not yet been verified. After that, Wei et al. [22] combined instrumental variable estimation and the parametric method of the bilinear principle to compensate for the deviation caused by noise in the process of model identification, which shows good performance for SOC estimation accuracy and noise tolerance.

Meanwhile, the integrity of the data is also a critical condition for parameter identification [38]. Several simple methods can be used for the case of missing data in a BMS, such as the interpolation method, cluster analysis algorithm and regression analysis algorithm. The interpolation method is only applicable to the case where the missing values are scattered and the error is large [39]. In the clustering analysis algorithm, the filling accuracy of the K-means algorithm is high, but when the attribute of the missing data deviates far from the other attributes, this method cannot be used to find a similar attribute for replacement and ensure stability [40]. Due to the limitations of the method, the above methods considering the data loss are only applicable offline and cannot be used during real-time online SOC estimation. Recently, Chen et al. [23] proposed a data-unavailability-resistant nonlinear recursive filtering algorithm to estimate the online SOC in an unreliable industrial environment, but the performance for parameter identification was not mentioned.

It can be seen that although the online parameter identification method for immune noise can be used to obtain a more accurate SOC estimation, most of the reported methods do not take into account the occurrence of data loss. In addition, the above methods for battery data loss cannot ensure accuracy and cannot be used to carry out online parameter identification and SOC estimation. Therefore, to solve the abovementioned problems, an online SOC estimation method is proposed in this paper. The contributions of this paper are mainly summarized as follows: 1) The random data loss for a LIB, which is characterized by a discrete-time deterministic system with unavailable output, is considered for SOC estimation. 2) A recursive algorithm based on an auxiliary model is proposed to identify the reliable parameter for a ECM by compensating for the missing data due to sensor defects and network fluctuations. 3) The EKF is applied to eliminate the SOC estimation error caused by environmental noise and ensure robustness for online SOC estimation. 4) The proposed method is validated by comparison with the RLS method under two typical dynamic driving cycles.

The rest of this paper is organized as follows: in Section 2, the modelling of an ECM is introduced, and the SOC-open circuit voltage (OCV) curve is fitted. Section 3 presents the principle and workflow for the recursive least square method based on the variable interval auxiliary model and EKF algorithm. In Section 4, the performance of the proposed method and traditional method for different working conditions is compared and discussed. The conclusion is presented in Section 5.

2. EQUIVALENT CIRCUIT MODELING

At present, ECMs are widely used for model-based battery SOC estimation due to the advantages of simple calculation and clear physical meaning. Therefore, an accurate ECM is the key to obtaining the dynamic parameters for a battery and provides the basis for parameter identification. An ECM can be divided into internal resistance models, n-order RC models and so on. In light of the

complexity and computational cost of the model [41], this paper selects the first-order RC circuit model shown in Fig. 1 for analysis. According to Kirchhoff's law, the first-order RC circuit model can be expressed as follows:

$$\begin{cases} I_0 = \frac{U_p}{R_p} + C_p \frac{dU_p}{dt} \\ U = U_{OC} + I_0 R_0 + U_p \end{cases} \quad (1)$$

where R_0 is the internal ohmic resistance; R_p and C_p are the polarization resistance and polarization capacitance, respectively, which are used to simulate the dynamic characteristics of the battery terminal voltage; C is the battery capacity; U , U_{OC} and U_p are the terminal voltage, OCV and polarization voltage of the battery, respectively; and I_0 denotes the loading current.

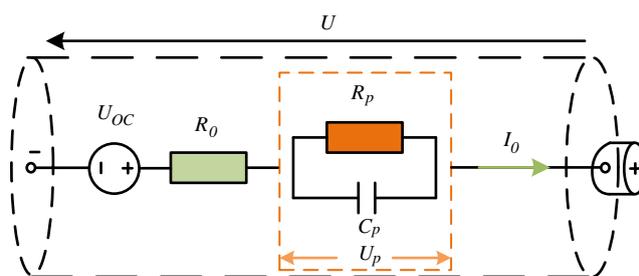


Figure 1. The first-order RC circuit model

By defining $E = U - U_{OC}$ and Laplace transforming Equation (1), the transfer function of the system can be expressed as follows:

$$G(s) = \frac{U(s)}{I_0(s)} = R_0 + \frac{R_p}{R_p C_p s} \quad (2)$$

This paper uses the following bilinear transformation method to obtain the transfer function for a discrete system:

$$s = \frac{2}{T_S} \frac{1 - z^{-1}}{1 + z^{-1}} \quad (3)$$

where T_S is the sample time.

Therefore, the discrete transfer function can be obtained as follows:

$$G(z) = \frac{U(z^{-1})}{I_0(z^{-1})} = \frac{a_2 + a_3}{1 - a_1 z^{-1}} \quad (4)$$

where

$$\begin{cases} a_1 = -\frac{T_S - 2R_p C_p}{T_S + 2R_p C_p} \\ a_2 = \frac{R_0 T_S + R_p T_S + 2R_0 R_p C_p}{T_S + 2R_p C_p} \\ a_3 = \frac{R_0 T_S + R_p T_S - 2R_0 R_p C_p}{T_S + 2R_p C_p} \end{cases} \quad (5)$$

The corresponding difference equation for Equation (3) is given by

$$U(k) = a_1U(k - 1) + a_2I_0(k) + a_3I_0(k - 1) \tag{6}$$

Which can be transferred to the discrete-time system:

$$y(k) = \varphi^T(k)\theta(k) + v(k) \tag{7}$$

where the information vector $\varphi(k) = [U(k - 1) \quad I_0(k) \quad I_0(k - 1)]^T$, the parameter vector to be identified $\theta(k) = [a_1 \quad a_2 \quad a_3]^T$, and $v(k)$ is zero-mean white noise.

Finally, R_0 , R_p , and C_p can be expressed as follows:

$$\begin{cases} R_0 = T_s \frac{a_2 - a_3}{T_s + a_1} \\ R_p = 2 \frac{a_1 a_2 + a_3}{a_1^2 - 1} \\ C_p = \frac{T_s(a_1 + 1)^2}{-4a_1 a_2 a_3} \end{cases} \tag{8}$$

The SOC-OCV curve required for model parameter identification can be obtained by a SOC-OCV test. In this paper, the incremental current test for an INR18650-20R battery in the University of Maryland test is used to analyse the SOC-OCV relationship. Specifications for the tested battery cell are shown in Table 1. The environmental temperature of the incremental current test is 25 °C, and the frequency is set to 1 Hz.

Table 1. Specifications for the tested battery cell

| Parameters | Value |
|-----------------------|-----------------|
| Type | INR18650-20R |
| Rated Capacity | 2000 mAh |
| Cell Chemistry | LNMC/Graphite |
| Upper Cut-off Voltage | 4.2 V |
| Lower Cut-off Voltage | 2.5 V |
| Weight | 45.0 g |
| Length | 64.85 ± 0.15 mm |

To obtain a better fitting effect, this paper compares the SSE , R^2 , $AdjustR^2$ and $RMSE$ for different orders of polynomial fitting. The formula of the index is expressed as follows:

$$\begin{cases} SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ AdjustR^2 = 1 - (1 - R^2) \frac{n - 1}{n - k} \\ RMSE = \sqrt{SSE/n} \end{cases} \tag{9}$$

Compared with low-order polynomials, the curve formed by fitting high-order polynomials can show more details, but, at the same time, the calculation complexity and calculation cost are greatly increased. Therefore, the 8th-order polynomial is finally adopted to fit the SOC-OCV curve. The fitting result is shown in Fig. 2.

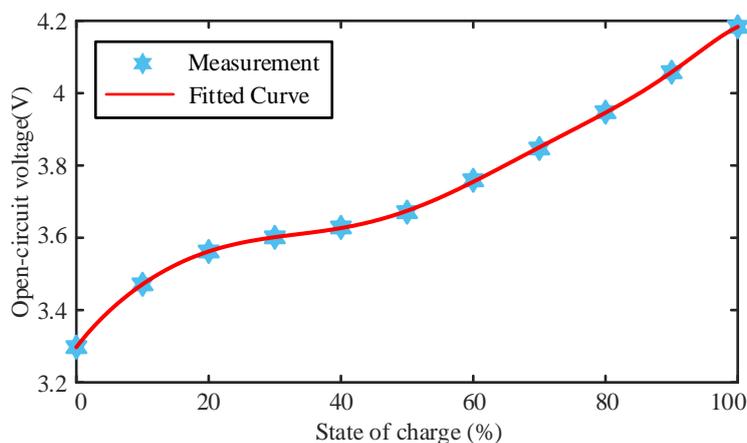


Figure 2. SOC-OCV curve

3. JOINT ALGORITHMS FOR VI-AM-RLS AND EXTENDED KALMAN FILTER WITH MISSING DATA

3.1 Parameter identification under different situations of missing data

The input and output data are not completely available in the data loss system. Due to various practical reasons, the data loss system can be divided into three cases [42]:

- (1) Loss of output data: the input data are complete, but part of the output data are lost;
- (2) Loss of input data: the output data are complete, but part of the input data are lost;
- (3) Loss of input and output data: only some input and output data are available.

The most common situation is case 1 because the input data are basically known and the original data are complete and available. Therefore, this case is worthy of further study. If case 2 does arise, it can be transformed into the problem of case 1 by inverting the system under some standard assumptions, such as stability and minimum phase. Case 3 is a completely different and complex situation that does not need to be discussed. Therefore, in this paper, we mainly focus on the loss of output data, which is the partial loss of voltage data for a LIB.

3.2 VI-AM-RLS parameter identification method with missing data and noise

For a system with missing data, the conventional least square algorithm cannot be directly applied to complete parameter identification. Therefore, to solve this problem, the least square method based on the variable interval auxiliary model (VI-AM-RLS) is proposed to identify the parameters that build an auxiliary model based on the measurable variables for the original system to compensate for the missing data.

In Section 1, considering the occurrence of data loss, the actual output model can be expressed as follows:

$$\begin{cases} y(k) = x(k) + v(k) \\ x(k) = \varphi^T(k)\theta(k) \end{cases} \quad (10)$$

where the intermediate variable $x(k)$ is the output of the system without any interference, which cannot be directly measured due to the noise interference in the actual system; $y(k)$ is the actual measured system output, which is disturbed by the noise variable $v(k)$.

To deal with the missing data, time series are defined as $\{k_s, s = 0, 1, \dots\}$, which meets $0 = k_0 < k_1 < k_2 < k_3 < \dots < k_{s-1} < k_s < \dots$, and $k^* = k_s - k_{s-1} \geq 1$, so that when $k = k_s$, measurement data $y(k)$ and $\varphi(k)$ are available and the series $\{y(k_s), \varphi(k_s): s = 0, 1, \dots\}$ contains all observable output data.

By replacing k with k_s in Equation 10, we can obtain the identification model:

$$\begin{cases} y(k_s) = \varphi^T(k_s)\theta(k_s) + v(k_s) \\ x(k_s) = \varphi^T(k_s)\theta(k_s) \end{cases} \quad (11)$$

The information vector $\varphi(k_s)$ contains $U(k_s - 1)$, $I_0(k_s)$ and $I_0(k_s - 1)$, in which $I_0(k_s)$ and $I_0(k_s - 1)$ are available and $U(k_s - 1)$ contains $u(k_s - 1)$ with missing data. Therefore, the unknown vector $u(k_s - 1)$ can be replaced with the output of the auxiliary model [42].

The auxiliary model can be constructed using the system input $I_0(k_s)$, as shown in Fig. 3:

$$u_a(k_s) = \varphi_a^T(k_s)\theta_a(k) \quad (12)$$

where the information vector $\varphi_a^T(k_s)$ is the polynomial of the same order as $\varphi(k_s)$ and $\theta_a(k)$ is the parameter vector of the auxiliary model.

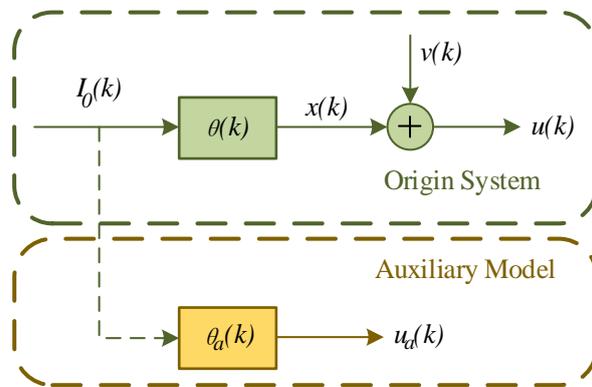


Figure 3. The missing-data battery system with the auxiliary model

Use $\hat{\theta}(k_s)$ and $\hat{\varphi}(k_s)$ as the parameter vector θ_a and information vector $\varphi_a(k_s)$ of the auxiliary model, which represent estimates of θ and $\varphi(k_s)$, respectively. It is obvious that

$$\begin{cases} \theta_a(k) = \hat{\theta}(k_s) \\ \varphi_a(k_s) = \hat{\varphi}(k_s) \end{cases} \quad (13)$$

The output of the auxiliary model can be expressed as

$$u_a(k_s - i) = \hat{\varphi}^T(k_s - i)\hat{\theta}(k_s) \quad (14)$$

By using $u_a(k_s - i)$ to replace the unmeasurable output $u(k)$, the problem of unknown variables in the information vector can be solved so that the noise interference can be effectively reduced. VI-AM-RLS can be used to determine the result for the case of when data loss occurs by replacing the missing data $u(k_s - i)$ with $u_a(k_s - i)$:

$$\begin{cases} e(k_s) = y(k_s) - \hat{\varphi}(k_s)\hat{\theta}(k_{s-1}) \\ \hat{\theta}(k_s) = \hat{\theta}(k_{s-1}) + P(k_s)\hat{\varphi}(k_s)[y(k_s) - e(k_s)] \\ P^{-1}(k_s) = P^{-1}(k_{s-1}) + \hat{\varphi}(k_s)\hat{\varphi}^T(k_s) \end{cases} \quad (15)$$

In the missing data section, the estimate for the parameter vector remains unchanged:

$$\hat{\theta}(k) = \hat{\theta}(k_s), k_s \leq k \leq k_{s+1} - 1 \quad (16)$$

We define the gain matrix $L(k_s)$ as follows:

$$L(k_s) = P(k_s)\varphi(k_s) = P(k_{s-1})\hat{\varphi}(k_s)[1 + \hat{\varphi}^T(k_s)P(k_{s-1})\hat{\varphi}(k_s)]^{-1} \quad (17)$$

Let λ ($0 < \lambda \leq 1$) be the forgetting factor. According to identification Model (11), we define

$$J(\theta) = \sum_{i=1}^s \lambda^{s-i} [y(k_i) - \varphi^T(k_i)\theta]^2 \quad (18)$$

By minimizing $J(\theta)$, the covariance matrix $P(k_s)$ can be updated as follows:

$$P(k_s) = \frac{1}{\lambda} [I_n - L(k)\hat{\varphi}^T(k_s)]P(k_{s-1}) \quad (19)$$

The initial value of $\hat{\theta}(k_0)$ is generally a small real vector, and $P(k_0)$ is a large positive-definite matrix. In this paper, $\hat{\theta}(k_0) = 1_n/p_0$, $P(k_0) = p_0I_n$, $p_0 = 10^6$. To a certain degree, the VI-AM-RLS dealing with missing data uses the output of the auxiliary model for parameter estimation. The detailed process for the VI-AM-RLS algorithm for model parameter identification is listed in Table 2.

Table 2. Model parameter identification based on VI-AM-RLS

| |
|---|
| (1) Initialization |
| $\varphi(k), \theta(k), y(k), u_a(k_s + i), P(k_0), p_0, \lambda$ |
| (2) Information and Parameter Vector |
| $\begin{cases} \varphi(k) = [U(k-1) & I_0(k) & I_0(k-1)]^T \\ \theta(k) = [a_1 & a_2 & a_3]^T \end{cases}$ |
| (3) Auxiliary Model |
| $u_a(k_s + i) = \hat{\varphi}^T(k_s + i)\hat{\theta}(k_s)$ |
| (4) Gain and Covariance Matrix |
| $\begin{cases} L(k_s) = P(k_{s-1})\hat{\varphi}(k_s)[1 + \hat{\varphi}^T(k_s)P(k_{s-1})\hat{\varphi}(k_s)]^{-1} \\ P(k_s) = \frac{1}{\lambda} [I_n - L(k)\hat{\varphi}^T(k_s)]P(k_{s-1}) \end{cases}$ |
| (5) Model Parameter Update |
| $\begin{cases} e(k_s) = y(k_s) - \hat{\varphi}^T(k_s)\hat{\theta}(k_{s-1}) \\ \hat{\theta}(k_s) = \hat{\theta}(k_{s-1}) + P(k_s)\hat{\varphi}(k_s)e(k_s) \end{cases}$ |

The flowchart for the battery SOC estimation based on the VI-AM-RLS parameter identification method is illustrated in Fig. 4. The measured current and voltage data from the sensor and network are sent to the proposed algorithm using an auxiliary model to obtain the parameter vector $\theta(k) = [a_1 \ a_2 \ a_3]^T$. After that, the value of R_0 , R_p , and C_p in the first-order RC model can be obtained from Equation 8, which can be used to calculate the model voltage. The SOC result can be finally estimated through the use of the extended Kalman filter algorithm.

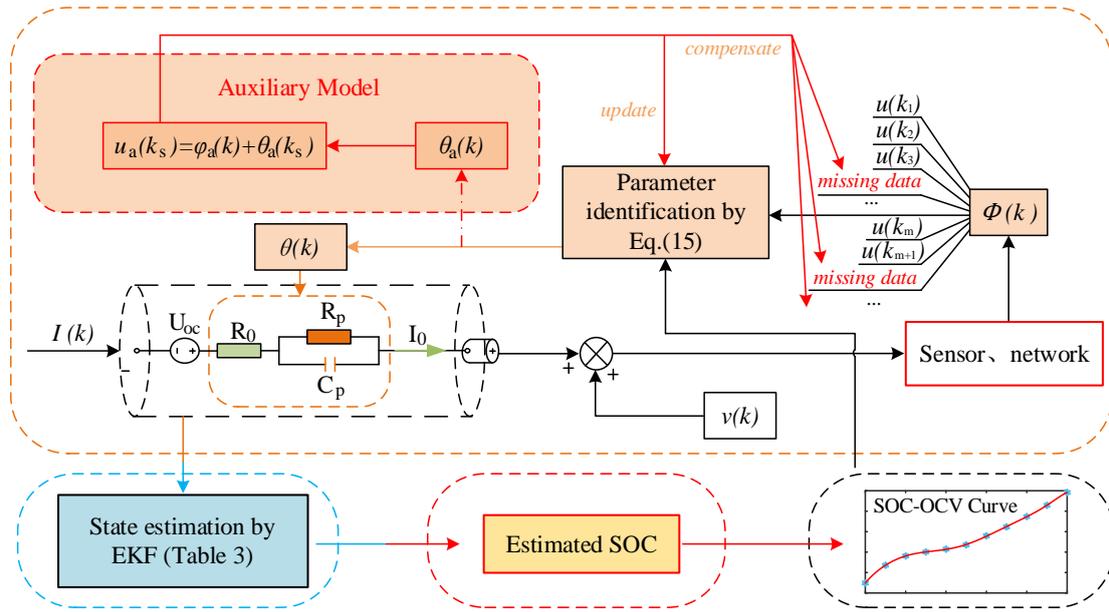


Figure 4. Flowchart for the proposed method for battery SOC estimation

3.3 SOC estimation with an extended Kalman filter

The joint RLS-based method and extended Kalman filter (EKF) are widely used in parameter and state coestimation [43]. The core idea of the estimation is to correct the dynamic states of the system and bias states by using the estimated values for the bias state. Since the EKF can only be applied to linear systems and ECM as a typical nonlinear model cannot be applied directly, the principle of the EKF algorithm used in this paper is to expand the nonlinear model to obtain an approximate linearized model, and then the estimation result can be obtained by the EKF.

The state equation of the nonlinear system can be expressed as follows:

$$\begin{cases} x_{k+1} = f(x_k, u_k) + w_k \\ y_k = g(x_k, u_k) + v_k \end{cases} \quad (20)$$

where f and g are the nonlinear functions of the system, w_k is the process noise, v_k is the measurement noise, and both are independent, zero mean, Gaussian noise.

The linearized model of the system can be obtained by linearizing the nonlinear function in the system by using the first-order Taylor series:

$$\begin{cases} x_{k+1} = A_k x_k + [f(\hat{x}_k, u_k) - A_k \hat{x}_k] + w_k \\ y_k = C_k x_k + [g(\hat{x}_k, u_k) - C_k \hat{x}_k] + v_k \end{cases} \quad (21)$$

where A_k and C_k are the state transition and observation matrices, respectively:

$$\begin{cases} A_k = \frac{\partial f(x_k, u_k)}{\partial x_k} = \begin{bmatrix} 1 - T_s/R_p C_p & 0 \\ 0 & 1 \end{bmatrix} \\ C_k = \frac{\partial g(x_k, u_k)}{\partial x_k} = \begin{bmatrix} 1 & \frac{\partial U_{OC}}{\partial SOC} \end{bmatrix} \end{cases} \quad (22)$$

After discretizing the first-order RC model shown in Equation (1) and assuming the state vector $x_k = [U_p \quad SOC]^T$, the state equation can be obtained as follows:

$$\begin{cases} \begin{bmatrix} U_{p,k+1} \\ SOC_{k+1} \end{bmatrix} = \begin{bmatrix} 1 - T_s/R_p C_p & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} U_{p,k} \\ SOC_k \end{bmatrix} + \begin{bmatrix} T_s/C_p \\ T_s/C \end{bmatrix} I_k \\ U_k = [1 \quad 0] \begin{bmatrix} U_{p,k} \\ SOC_k \end{bmatrix} + R_0 I_k + U_{OC,k} \end{cases} \quad (23)$$

so that the state can be predicted and updated according to the recursive process of the EKF algorithm shown in Table 3.

Table 3. SOC estimation based on EKF

| |
|--|
| (1) Initialization |
| $\begin{cases} x_0^+ = E(x_0) \\ P_0^+ = var(x_0) \end{cases}$ |
| (2) Prediction |
| $\begin{cases} \hat{x}_k^- = f(\hat{x}_{k-1}^-, u_{k-1}) \\ P_{x,k}^- = A_{k-1} P_{k-1}^+ A_{k-1}^T + Q_{k-1} \end{cases}$ |
| (3) Update |
| $\begin{cases} K_k = P_{x,k}^- C_k^T (C_k P_{x,k}^- C_k^T + R_k)^{-1} \\ \hat{x}_k^+ = \hat{x}_k^- + K_k [y_k - g(\hat{x}_k^-, u_k)] \\ P_{x,k}^+ = (I - K_k C_k) P_{x,k}^- \end{cases}$ |

where \hat{x}_k^- is the predicted value for the estimated state, \hat{x}_k^+ is the updated estimated state, K_k is the Kalman gain matrix, $P_{x,k}^-$ is the predicted value for the covariance matrix, $P_{x,k}^+$ is the updated covariance matrix, and Q_k and R_k is the variance in the process noise and measurement noise, respectively.

After estimation of the SOC by the EKF, U_{OC} is calculated by a fit to the SOC-OCV curve to obtain the predicted model voltage, as shown in Fig. 2, which can be circularly used in VI-AM-RLS.

4. EXPERIMENT

In this section, to analyse the complex battery loading conditions, test subjects were exposed to two dynamic testing profiles, including a dynamic stress test (DST) and a federal urban driving schedule (FUDS). Figure 5 plots the current and voltage profiles for the DST and FUDS profiles. All the experimental data for the LIB used in this paper were obtained from the CALCE battery research group of the University of Maryland, which is an open battery test database. [dataset] The environmental temperature was set to 25 °C, and the initial SOC was 80%. As shown in Figs. 6(a) and (b), the parameter identification model can be used to accurately predict the voltage response of the test battery under two different working conditions. The error between the measured voltage and predicted voltage is within 3%, as shown in Figs. 6(c) and (d). The model parameters are identified online and offline. In this paper, the experimental parameter identification reference values are obtained by offline identification under the noise-free case. The DST condition for the INR18650-20R battery at 25°C tested by the University of Maryland is used for offline parameter identification, where

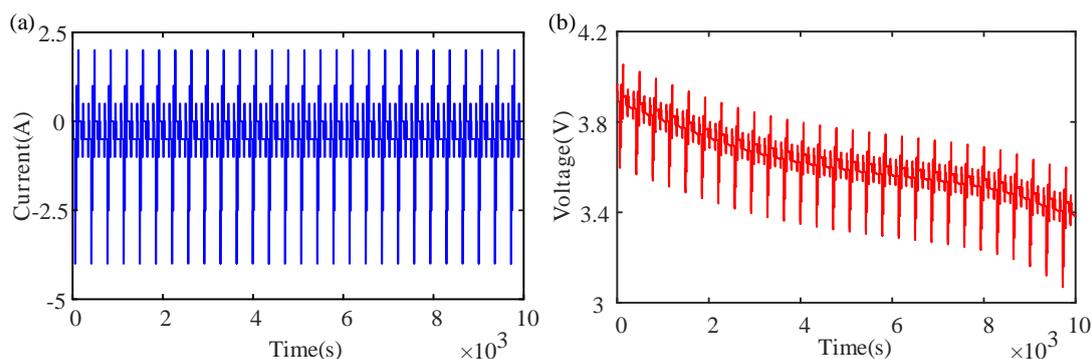
a general method is adopted. The process mainly includes two steps: (1) Because the internal resistance R_0 is proportional to the instantaneous voltage drop under the known discharge pulse, which can be achieved by injecting a current I_0 and measuring the attenuation of the relevant voltage response. (2) The polarization resistance R_p and polarization capacitance C_p are estimated by an offline least square estimation algorithm. As shown in Figs. 6(e)-(g), the black triangle represents the offline parameter identification result, which is provided for reference, and the red line represents the online parameter identification result.

The performance of the framework proposed in this paper is verified by comparison with that obtained for current mainstream algorithms, such as RLS, SVM, FOM, LSVPA and RLS-EKF which is the most common traditional method. The average errors in the SOC estimation result are shown in Table 4. It can be observed that the average error using the VI-AM-RLS proposed in this work is the lowest. The average MAE, MAPE and RMSE for the proposed method are 0.64%, 0.71% and 0.85%, respectively.

In addition, in an actual battery management system, we face many complex and changeable situations. Different situations lead to different forms of missing data. For example, under electromagnetic interference or due to the limitation of sensor equipment, there are occasional random losses of individual data; due to the limitation of the data transmission system or computer memory, paragraph data or data packets can be lost.

Table 4. Error analysis for the model proposed in this work compared to that for other models

| Model | Average MAE% | Average MAPE% | Average RMSE% |
|-----------------------|--------------|---------------|---------------|
| RLS [44] | 4.24 | 4.33 | 4.54 |
| SVM [45] | 2.58 | 2.69 | 2.87 |
| FOM [46] | 2.16 | 2.24 | 2.35 |
| LSVPA [47] | 1.52 | 1.67 | 1.74 |
| RLS-EKF [48] | 1.84 | 1.95 | 2.17 |
| VI-AM-RLS [this work] | 0.64 | 0.71 | 0.85 |



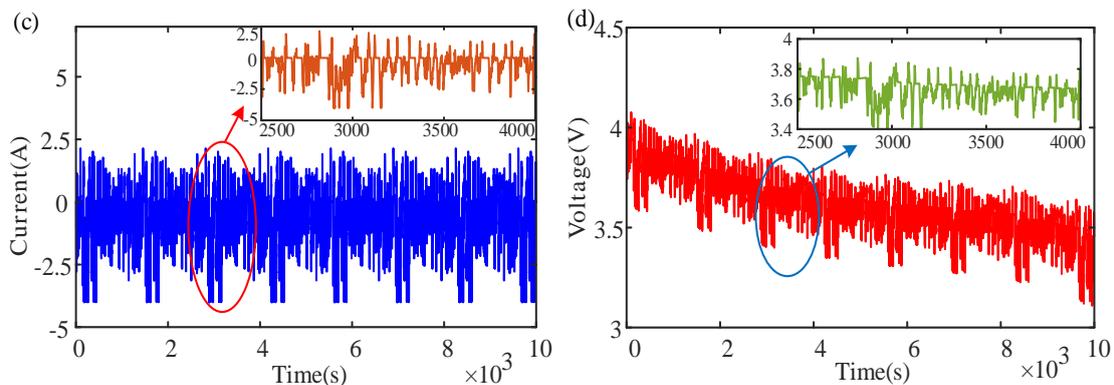
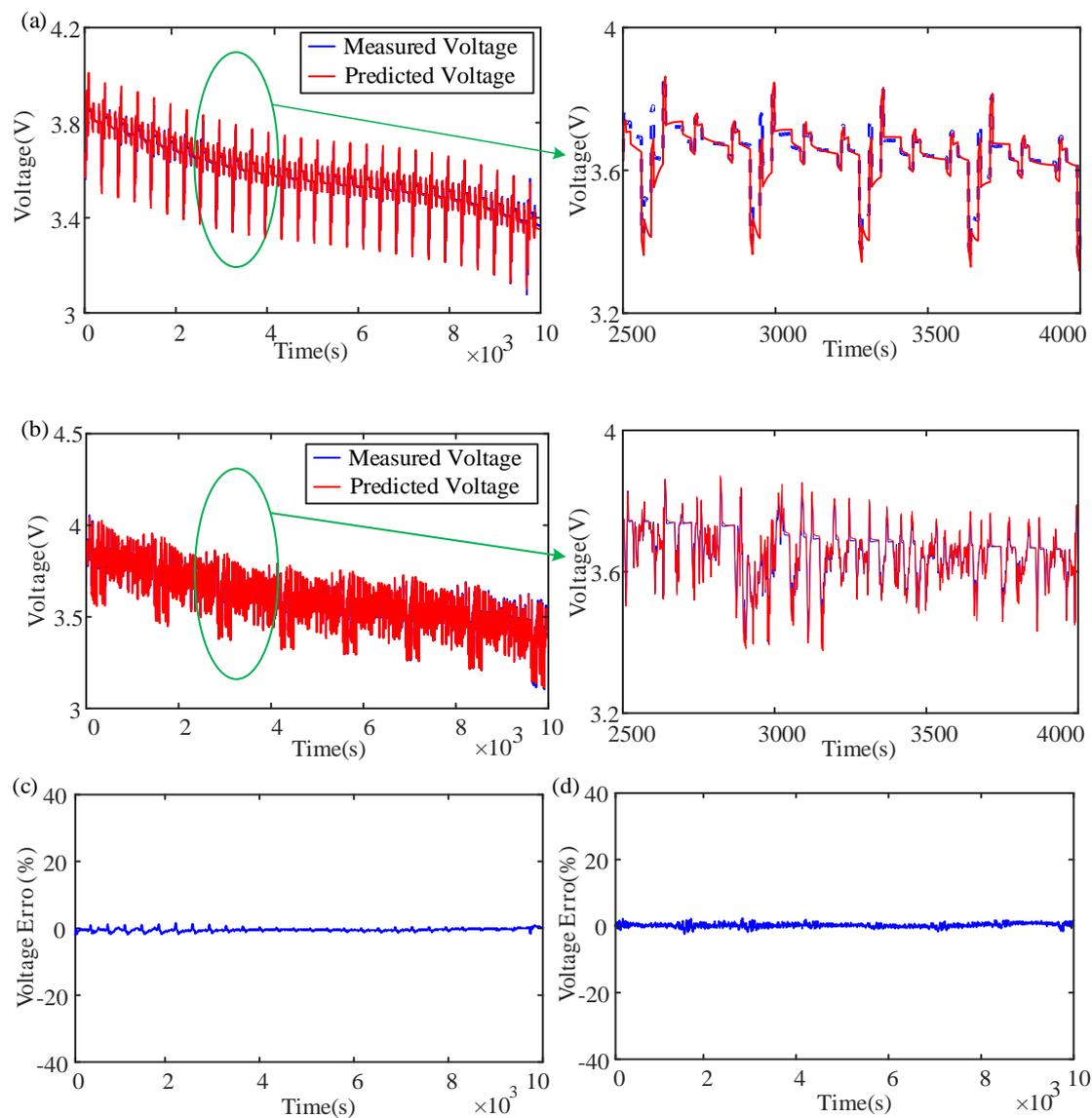


Figure 5. (a) Current profiles for the DST; (b) voltage profiles for the DST; (c) current profiles for the FUDS; (d) voltage profiles for the FUDS.



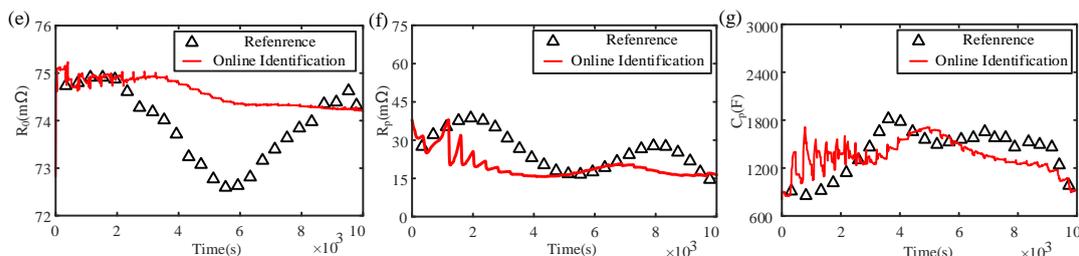


Figure 6. (a) Evolution for the predicted and measured voltages in the DST; (b) Evolution of the predicted and measured voltages in the FUDS; (c) Error between the predicted and measured voltages in the DST; (d) Error between the predicted and measured voltages in the DST; (e) The internal resistance R_0 ; (f) The polarization resistance R_p ; (g) The polarization capacitor C_p .

4.1 Parameter identification and SOC estimation with missing data in the DST

In the DST condition, the voltage signals are subjected to random single data losses, and the measured voltage is compensated by the method proposed above.

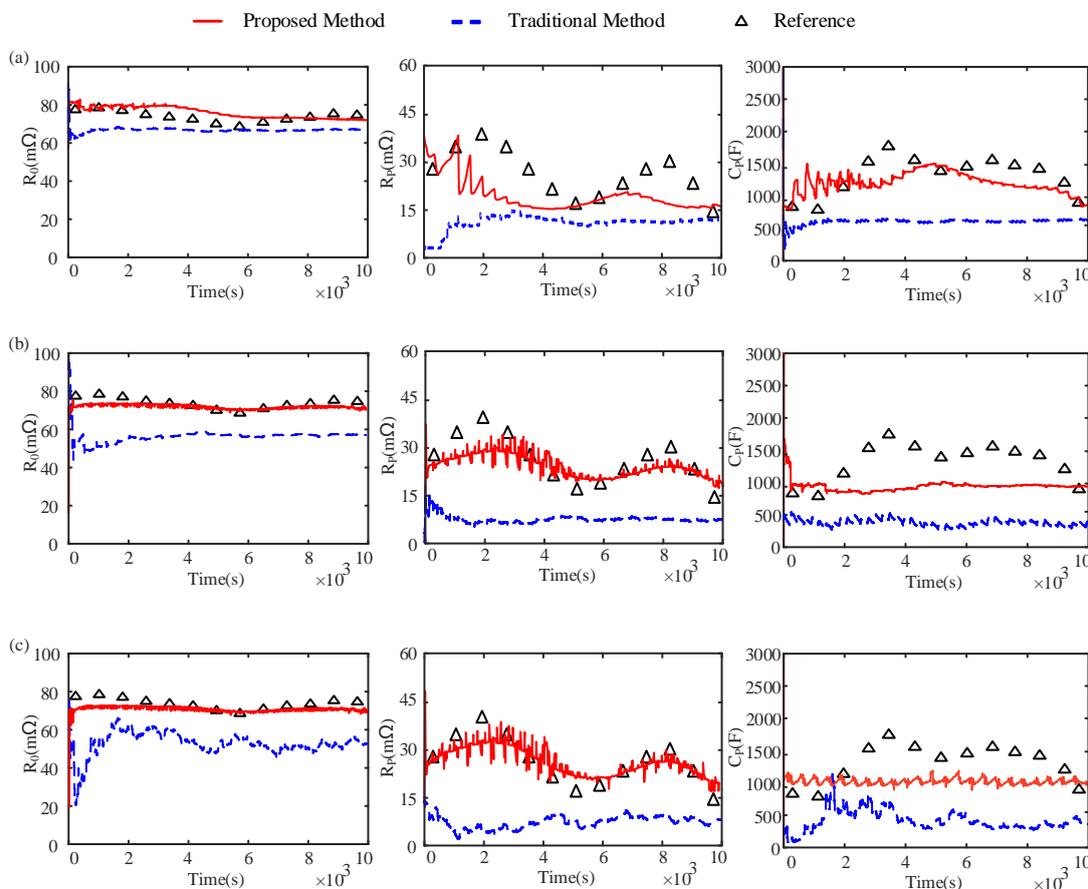


Figure 7. (a) Parameter identification with complete data; (b) parameter identification with a data loss rate of $\Pr\{\alpha_1(k) = 0\} = 0.1$; (c) parameter identification with a data loss rate of $\Pr\{\alpha_1(k) = 0\} = 0.2$.

The battery random single data loss rates are set as $\Pr\{\alpha_1(k) = 0\} = 0.1$ and $\Pr\{\alpha_2(k) = 0\} = 0.2$. The traditional method involves the combination of RLS and EKF for parameter identification. In addition, to simulate the noise environment, the variance in the current and voltage noise are added as $\sigma_v^2 = 20mV^2$ and $\sigma_i^2 = 20mA^2$, respectively.

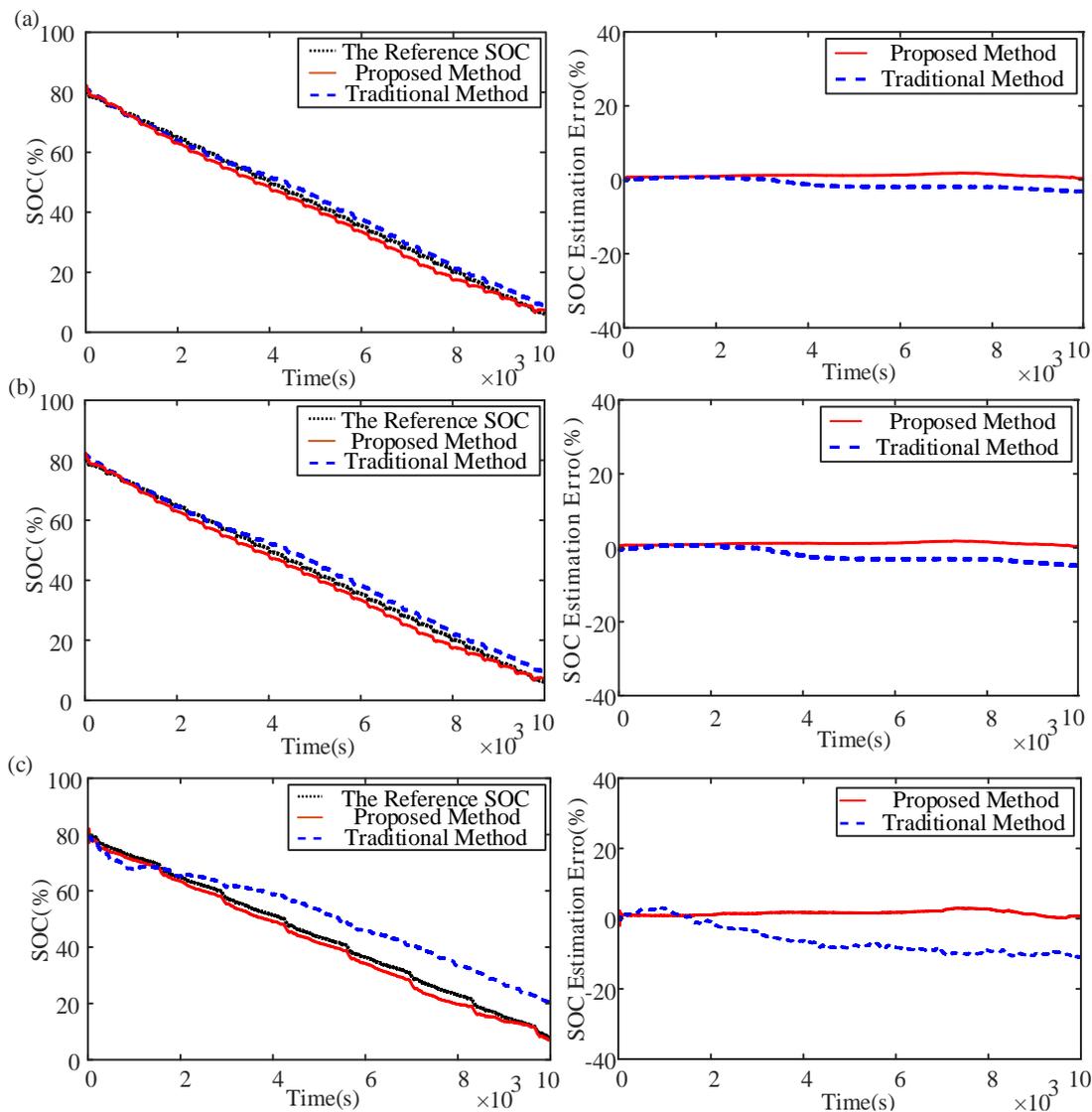


Figure 8. (a) SOC estimation result with complete data; (b) SOC estimation result with a data loss rate of $\Pr\{\alpha_2(k) = 0\} = 0.1$; (b) SOC estimation result with a data loss rate of $\Pr\{\alpha_2(k) = 0\} = 0.2$.

The parameter identification results are shown in Fig. 7, and the reference value is the parameter identification result obtained under offline conditions. The result of SOC estimation is shown in Fig. 8, and the reference SOC obtained using the ampere hour integration method under a noiseless environment and with complete input and output data is shown for comparison. When the data loss rate is 10%, the traditional SOC estimation method and the method proposed in this paper can quickly converge to the reference value in the initial stage. At the same time, the proposed method maintains high robustness and converges near the reference value. With an increase in the missing data

rate to 20%, the traditional method gradually deviates from the reference value and converges again after a long period of iteration. The fluctuation in the traditional method is more intense and difficult to converge, and a final error of 11.32% is obtained. In contrast, the proposed method still converges near the reference value in these cases and maintains high robustness. The MAE and RMSE for the two cases is shown in Table 4.

Table 5. Comparison of MAE and RMSE with single data losses

| Loss rate | 10% | | 20% | |
|------------|-------|-------|--------|--------|
| | MAE | RMSE | MAE | RMSE |
| VI -AM-RLS | 0.92% | 0.96% | 2.56% | 2.63% |
| RLS | 3.34% | 3.57% | 11.32% | 11.51% |

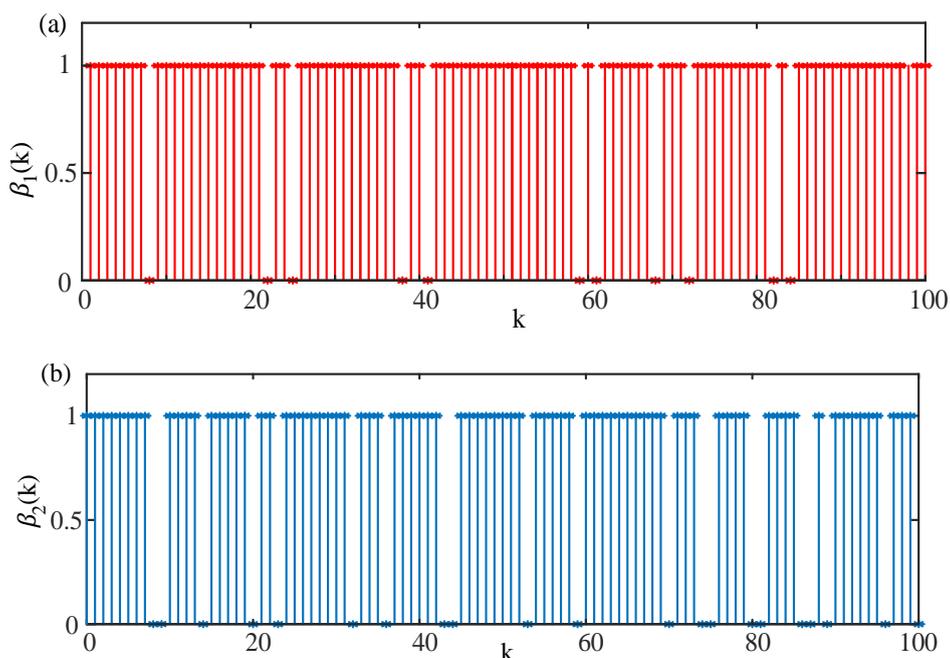


Figure 9. (a) The state of data packet losses with $\Pr\{\beta_1(k) = 0\} = 0.1$; (b) The state of data packet losses with $\Pr\{\beta_2(k) = 0\} = 0.2$.

For the case of data packet losses, the data loss rates are set to $\Pr\{\beta_1(k) = 0\} = 0.1$ and $\Pr\{\beta_2(k) = 0\} = 0.2$. The state of data packet losses is shown in Figs. 9 (a) and (b), where ‘0’ and ‘1’ indicate the missing and available data packets, respectively, and the length of every data loss packet is 1%. Similar to the single data loss, the variance in the current and voltage noise is added as $\sigma_v^2 = 20mV^2$ and $\sigma_i^2 = 20mA^2$, respectively.

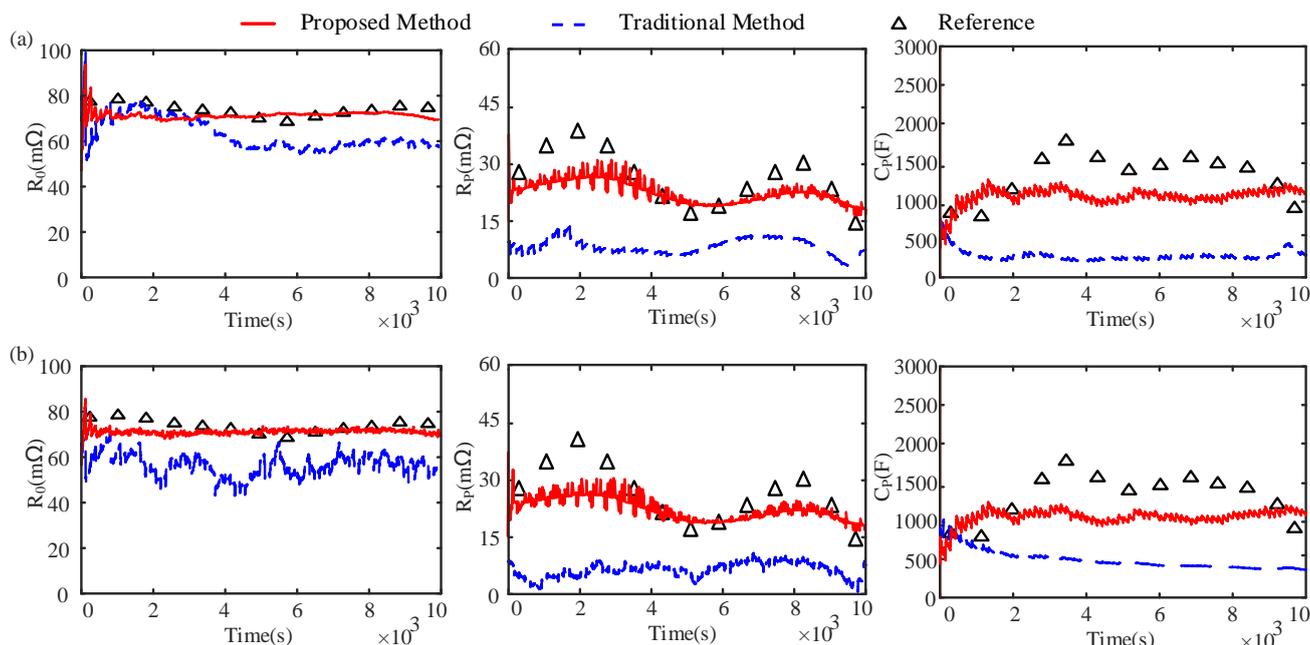


Figure 10. (a) Parameter identification with a data packet loss rate of $\Pr\{\beta_1(k) = 0\} = 0.1$; (b) parameter identification with a data packet loss rate of $\Pr\{\beta_1(k) = 0\} = 0.2$.

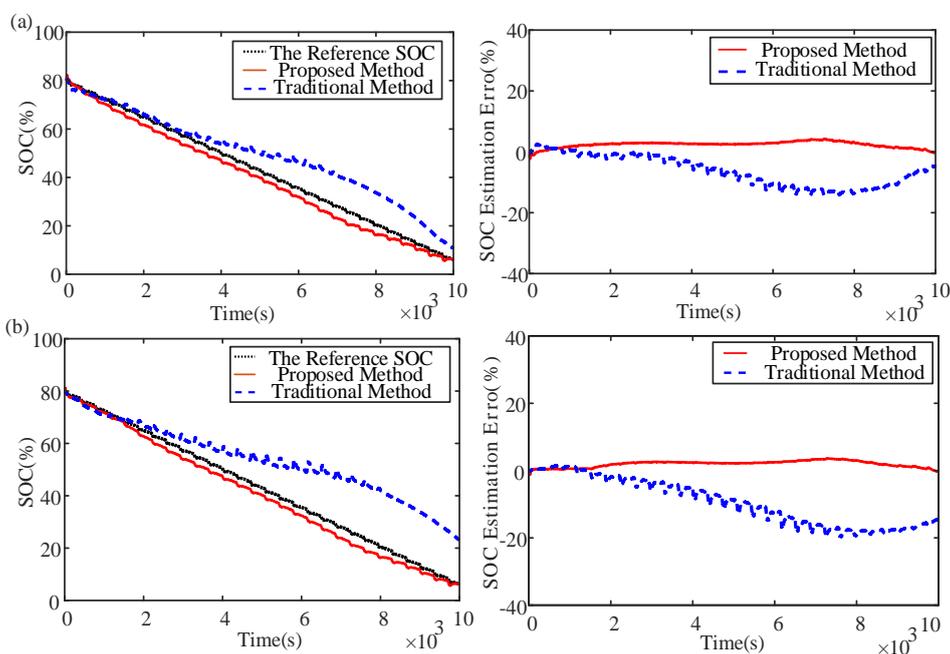


Figure 11. (a) SOC estimation result with a data packet loss rate of $\Pr\{\beta_1(k) = 0\} = 0.1$; (b) SOC estimation result with a data packet loss rate of $\Pr\{\beta_1(k) = 0\} = 0.2$.

The parameter identification results are shown in Fig. 10. It can be observed that for the case of data packet loss, the estimated SOC results shown in Fig. 11(a) for the proposed method also show a higher convergence, less fluctuation and less error than the result obtained using the traditional method shown in Fig. 11(b). In addition, the parameter identification result for RLS shows obvious fluctuations, which illustrates that although the traditional method has a degree of accuracy in SOC

estimation, it cannot accurately reflect the identification results obtained for circuit parameters in the ECM. The MAE and RMSE are shown in Table 5.

Table 6. Comparison of the MAE and RMSE with data packet losses

| Loss rate | 10% | | 20% | |
|-----------|--------|--------|--------|--------|
| | MAE | RMSE | MAE | RMSE |
| VI-AM-RLS | 2.82% | 2.91% | 3.11% | 3.19% |
| RLS | 13.23% | 13.65% | 15.44% | 15.92% |

4.2 Parameter identification and SOC estimation with missing data packets in FUDS

To further demonstrate the performance of the proposed method under different working conditions, experimental verification under the condition of data loss in a FUDS test was carried out in this paper. The voltage signals were subjected to a combination of random single data losses and data packet losses. The battery data loss rates were set as $\Pr\{\gamma_1(k) = 0\} = 0.2$ and $\Pr\{\gamma_2(k) = 0\} = 0.3$, where $\gamma_1(k) = \alpha_1(k) + \beta_1(k)$; $\gamma_2(k) = \alpha_2(k) + \beta_2(k)$. To simulate the noise environment, the variance in the current and voltage noise was added as $\sigma_v^2 = 20mV^2$ and $\sigma_i^2 = 20mA^2$, respectively.

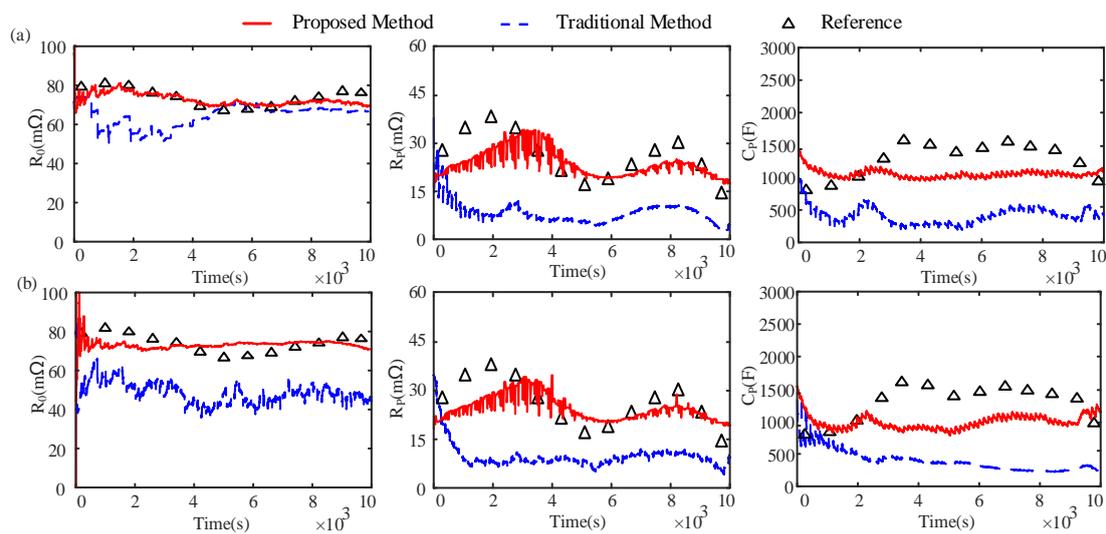


Figure 12. (a) Parameter identification with a data packet loss rate of $\Pr\{\gamma_1(k) = 0\} = 0.2$; (b) parameter identification with a data packet loss rate of $\Pr\{\gamma_1(k) = 0\} = 0.3$.

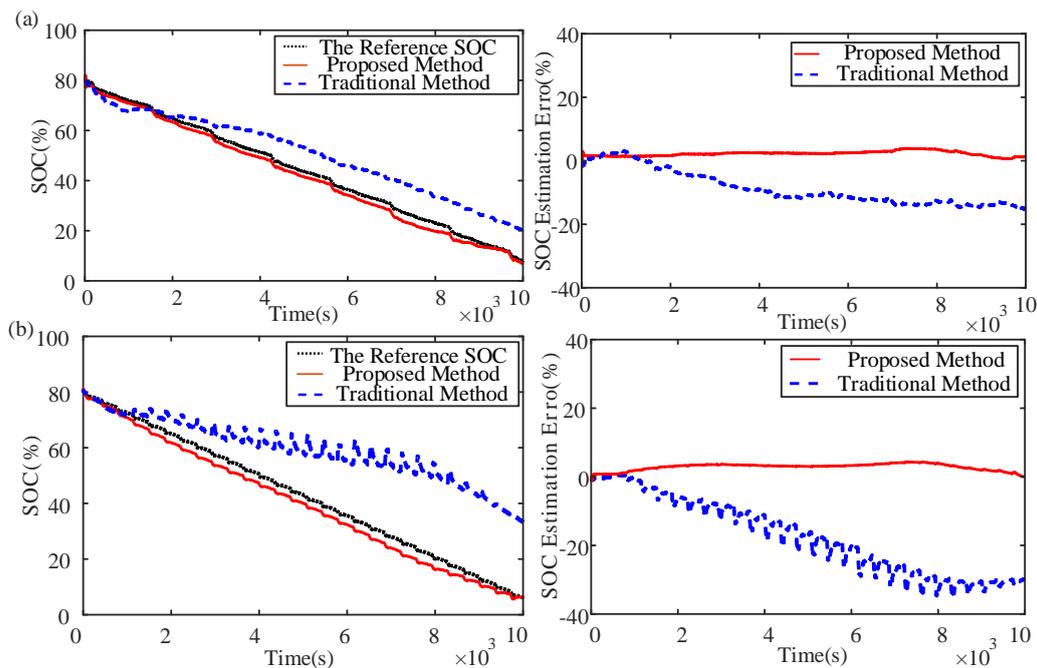


Figure 13. (a) SOC estimation result with a data packet loss rate of $\Pr\{\gamma_1(k) = 0\} = 0.2$; (b) SOC estimation result with a data packet loss rate of $\Pr\{\gamma_1(k) = 0\} = 0.3$.

When the combined data loss rate is set as $\Pr\{\gamma_1(k) = 0\} = 0.2$, the parameter identification results are shown in Fig. 12(a). The traditional method shows a slow convergence speed and drastic fluctuation. The method proposed in this paper shows small fluctuations and high accuracy and can quickly converge to the vicinity of the reference value. Both the traditional and proposed methods can finally converge to the reference value. The error for SOC estimation is shown in Fig. 13(a). The final error for the proposed method is less than 3%, while the error for the traditional method is more than 15%.

When the combined data loss rate is increased to $\Pr\{\gamma_2(k) = 0\} = 0.3$, as shown in Fig. 12 (b), the parameter identification fluctuation for the proposed method is drastic in the initial stage and converges quickly, while the traditional method cannot converge to the reference with increasing data loss. The final estimated SOC error for the traditional method is more than 30%, and the final error for the method proposed in this paper is less than 4%. When the rate of missing data reaches 30, the traditional method cannot accurately reflect the identification results obtained for the circuit parameters in the equivalent circuit model and the SOC estimation. The MAE and RMSE for the estimation results are shown in Table 6.

Table 7. Comparison of the MAE and RMSE obtained for combined data losses

| Loss rate | 20% | | 30% | |
|------------|--------|--------|--------|--------|
| | MAE | RMSE | MAE | RMSE |
| VI -AM-RLS | 2.88% | 2.94% | 3.43% | 3.51% |
| RLS | 15.44% | 15.87% | 32.86% | 34.47% |

Therefore, the experiment based on the FUDS test further proves the adaptability and robustness of the method proposed for the situation of combined data loss.

5. CONCLUSION

In this paper, a recursive least square method based on a variable interval auxiliary model is proposed for parameter identification and SOC estimation for LIBs when experiencing data loss and noise interference. This conclusion follows from the facts listed below:

1). The random missing data were compensated for by an auxiliary model so that VI-AM-RLS was applied to identify the model parameter. In addition, the EKF was combined to eliminate the SOC estimation error caused by environmental noise.

2). Simulation results further demonstrate the feasibility and efficiency of the proposed method. Even if the data loss rate is set to 30%, the accuracy and convergence for the estimation results also show good performance.

3). Under two different working conditions and various numerical simulations, the maximum SOC estimation error for the proposed method is 3.19% and 3.51%, respectively, which is far less than the 15.92% and 34.47% obtained for the RLS.

However, the proposed method is only validated in DST and FUDS at 25°C. There are many other factors that are not considered, such as the influence of temperature and battery ageing. Both of these factors can lead to an error in the SOC-OCV curve and further affect the parameter identification, which will be included in future work.

References

1. A. El-Kharbachi, O. Zavorotynska, M. Latroche, F. Cuevas, V. Yartys and M. Fichtner, *Exploits*, 8 (2020) 817.
2. L.G. Lu, X. Han, J. Li, J. Hua and M. Ouyang, *J. Power Sources*, 46 (2013) 226.
3. K. Liu, K. Li, Q. Peng and C. Zhang, *Front. Mech. Eng.*, 33 (2019) 14.
4. S. Tamilselvi, Z. Quan and H. He, *Sustainability*, 6 (2021) 13.
5. X. Zhang and D. Sun, *Electrochim Acta*, 34 (2015) 226.
6. J. Meng and R. Chen, *IEEE Trans. Ind. Electron.*, 10 (2019) 7717.
7. K. Liu, Y. Shang, Q. Ouyang and W.D. Widanage, *IEEE Trans. Ind. Electron.*, 68 (2021). 3170.
8. X. Bian, Z. Quan and H. He, *IEEE Trans. Transp. Electr.*, 7 (2021) 399.
9. Y. Wang, J. Tian, Z. Sun, L. Wang, R. Xu, M. Li and Z. Chen, *Renewable Sustainable Energy Rev.*, 12 (2020) 131.
10. J. Meng, D. Stroe, M. Ricco, G. Luo and R. Teodorescu, *Appl Sci-Basel*, 8 (2018) 37.
11. H. Chaoui and C.C. Ibe-Ekeocha, *IEEE Trans. Veh. Technol.*, 66 (2017) 8773.
12. 1 A. Fotouhi, D. Auger, K. Propp, S. Longo and M. Wild, *Renewable Sustainable Energy Rev.*, 56 (2016) 1008.
13. Q. Wang, J. Wang, P. Zhao, J. Kang, F. Yan and C. Du, *Electrochim Acta*, 228 (2017) 146.
14. H. He, X. Zhang, R. Xiong, Y. Xu and H. Guo, *Energy*, 39 (2012) 310.
15. Z. Zeng, J. Tian, D. Li and Y. Tian, *Energies*, 11 (2018) 468.

16. Z. Wei, C. Zou, F. Leng, B. Soong and K. Tseng, *IEEE Trans. Ind. Electron.*, 65 (2018) 1336.
17. X. K.Chen and D. Sun, *Adv. Manuf.*, 33 (2015) 202.
18. G. Luo, J. Meng, X. Ji, X. Cai and F. Gao, *J. Energy Storage*, 18 (2017) 2451.
19. F. Sun, X Hu, Y. Zou and L. Siguang, *Energy*, 36 (2011) 3531.
20. M. Gholizadeh and F.R. Salmasi, *IEEE Trans. Ind. Electron.*, 61 (2014) 1335.
21. W. Li, M. Rentemeister, J. Badeda, D. Jöst, D. Schulte and D. Sauer, *J. Energy Storage*, 30 (2020) 361.
22. Z. Wei, G. Dong, X. Zhang, J. Pou, Z. Quan and H. He, *IEEE Trans. Ind. Electron.*, 68 (2021) 312.
23. J. Hong, Z. Wang and Y. Yao, *Appl. Energy*, 18 (2019) 251.
24. H. Chen, E. Tian, and L. Wang, *IEEE Trans. Ind. Electron.*, 69 (2022) 5175.
25. A. Yang and G. Dong, *IEEE Trans. Veh. Technol.*, 72 (2017) 985.
26. S. Li, H. He, C. Su and P. Zhao, *Appl. Energy*, 72 (2020) 275.
27. W. Li, M. Rentemeister, J. Badeda, D. Jöst, D. Schulte and D. Sauer, *J. Energy Storage*, 30 (2020) 4524.
28. A. Adhikaree and L. Zhang, *IEEE Trans. Veh. Technol.*, 28 (2018) 113
29. T. Tanizawa, T. Suzumiya and K. Ikeda, *Fujitsu Sci. Tech. J.*, 51 (2015) 27.
30. S. Yang, Y. Cao, S. Zhou, Y. Hua, X. Zhou and X. Liu, *IEEE Access*, 18 (2020) 156232.
31. D. Du, R. Chen, M. Fei and K. Li, *IEEE Trans. Signal Inf. Process. Networks*, 34 (2017) 744.
32. D. A. Popescu and A.W. Moore, *IEEE Trans. Netw. Serv. Manage.*, 15 (2021) 3753.
33. X. Zhu , H. Zhang, J. Wang and Z. Fang, *IEEE Trans. Veh. Technol.*, 64 (2015) 4985.
34. Z. Shuai, H. Zhang, J. Wang, J. Li and M. Ouyang, *Control Eng. Pract.*, 24 (2014) 55.
35. X. Hu, H. Yuan, C. Zou, Z. Li and L. Zhang, *IEEE Trans. Veh. Technol.*, 67 (2018) 10319.
36. C. Jiang, C. Jiang, S. Wang, B. Wu, C. Fernandez, X. Xiong and J. Coffie-Ken, *Energy*, 52 (2021) 219.
37. Z. Wei, D. Zhao, H. He, W. Cao and G. Dong, *Appl. Energy*, 73 (2020) 268.
38. D. Wang, S. Zhang, M. Gan and J. Qiu, *IEEE Trans. Ind. Inf.*, 16 (2020) 2500.
39. J. Li and A.D. Heap, *Environ. Modell. Software*, 53 (2014) 173.
40. Q. Yang and Z. Liu, *IEEE Access*, 11 (2012) 232
41. P. Shen, M. Ouyang, L. Lu, J. Li and X. Feng, *IEEE Trans. Veh. Technol.*, 67 (2018) 92.
42. F. Ding and J. Ding, *Int. J. Adapt. Control Signal Process.*, 57 (2009) 180.
43. Y. Li, J. Chen and F. Lan, *J. Power Sources*, 17 (2020) 456.
44. J. Zhang, Y. Wei and H. Qi, *Appl. Math. Modell.*, 11 (2016) 6040
45. F. Xie, S. Wang, Y. Xie, C. Fernandezb, X. Li and C. Zou, *Int. J. Electrochem. Sci.*, 15 (2020) 7935
46. R. Xiong, J. Tian, W. Shen and F. Sun, *IEEE Trans. Veh. Technol.*, 68 (2019) 4130
47. Z. Wei, D. Zhao, H. He, W. Cao and G. Dong, *Appl. Energy*, 268 (2020) 114932
48. S. Zhang, H. Tao, K. Bi, W. Yan and H. Ni, *J. Phys.*, 2216 (2022) 12002