ISSN: 2993-6594

Research Article

Open Access Journal of Disease and Global Health

A Joint SOC Estimation Method Based on FFRLS-AEKF-LSTM

Min Wei[1], Xian Tao Yu[2] and Jiang Cheng Ding[3*]

[1]*School of Automotive Engineering, Wuhan University of Technology, China*

[2]*Wuhan University of Technology, China*

[3]*Wuhan University of Technology, China*

***Corresponding Author**
Jiang Cheng Ding, School of Automotive Engineering, Wuhan University of Technology, China.

Submitted: 2025, Jun 18; **Accepted:** 2025, Jul 14; **Published:** 2025, Aug 06

Citation: Wei, M., Yu, X. T., Ding, J. C. (2025). A Joint SOC Estimation Method Based on FFRLS-AEKF-LSTM. *Ope Acce Jou Dis Glo Heal, 3*(2), 01-09.

Abstract

The state of charge (SOC) of a battery is a core component of battery management systems (BMS) and has become a key research focus because of its significant role in the development of clean energy. To address the limitations of traditional SOC estimation methods, this paper proposes a joint SOC estimation algorithm based on FFRLS-AEKF-LSTM. First, a second-order RC equivalent circuit model of the battery is established, using data from hybrid pulse power characterization (HPPC) tests as input. The model parameters are initially identified the forgetting factor recursive least squares (FFRLS) method. Then, the adaptive extended Kalman filter (AEKF) algorithm is employed to update and estimate the SOC iteratively. Finally, the parameters obtained from the AEKF, together with voltage and current data under HPPC conditions, are used to train a long short-term memory (LSTM) neural network to predict the SOC. Experimental results show that the proposed joint algorithm achieves a root mean square error (RMSE) and mean absolute error (MAE) of less than 1%, demonstrating excellent performance.

Keywords: Lithium-Ion Battery, State of Charge, Long Short-Term Memory Network

1. Introduction

In recent years, industrial technologies have developed rapidly, and scientific and technological progress has advanced at an unprecedented pace. However, this progress has also led to environmental pollution and energy shortages. In response to the national strategies of "carbon peaking" and "carbon neutrality," China is actively promoting a cleaner energy consumption structure and striving to achieve both goals [1]. As a clean and pollution-free green energy source, lithium-ion batteries have been widely applied because of their high energy density, long lifespan, low self-discharge rate, and strong stability—particularly within the new energy vehicle industry chain. Therefore, the battery management system (BMS), which plays a key role in real-time monitoring of electric vehicle status, is of central importance in the development of new energy vehicles. Accurately estimating the state of charge (SOC) of individual lithium-ion battery cells in a system is considered the lifeline of a BMS and is critical to energy management and safety protection in electric vehicles [2].

However, owing to the complexity of internal lithium battery structures, their electrochemical reactions are affected by various factors, making them nonlinear and time-varying. This makes direct measurement difficult, and accurate SOC estimation has thus become a research hotspot in recent years. Currently, four main methods exist for SOC estimation in lithium batteries [3]: (1) The open-circuit voltage method requires the battery to rest for a long time until its internal electrochemical reactions stabilize; then, the SOC is determined on the basis of voltage lookup tables. Owing to its long resting time, this method is impractical in real-world applications. (2) The ampere-hour integration method is simple and computationally efficient, but it relies heavily on sensor accuracy and is prone to accumulated measurement errors, leading to poor precision in practice. (3) The Kalman filtering method is typically based on equivalent battery models, but it has limitations when dealing with nonlinear systems and requires optimization. Reference proposed a second-order

extended Kalman filter (EKF) algorithm to improve upon the weak nonlinearity handling and low estimation accuracy of the first-order version [4]. (4) Data-driven methods typically apply machine learning techniques to estimate the SOC. These methods require large datasets and extensive training but offer high accuracy in modeling nonlinear systems. Reference proposed an improved gated recurrent unit (GRU) neural network algorithm optimized via a Hunter–Prey Optimization (HPO) algorithm, resulting in SOC estimation errors within 5% [5].

To address the limitations of existing approaches, this paper proposes a joint SOC estimation algorithm based on FFRLS-AEKF-LSTM. First, a second-order RC equivalent battery model is constructed. The parameters are subsequently identified via the forgetting factor recursive least squares (FFRLS) method. The adaptive extended Kalman filter (AEKF) algorithm is applied to estimate preliminary SOC values and relevant battery parameters. Finally, a long short-term memory (LSTM) neural network is built, using the AEKF-derived features as input, to accurately predict the SOC.

2. Materials and Methods

2.1. Modeling of the Battery

The identification of battery parameters relies heavily on the establishment of an accurate battery model, as the model directly affects the precision of SOC estimation. Currently, the most commonly used battery models include equivalent circuit models and electrochemical models, among which the former is preferred for parameter identification because of its relatively high accuracy and relatively simple structure [6]. In this study, a second-order Thevenin equivalent circuit model is adopted. Compared with the traditional Thevenin model, this configuration offers improved accuracy while maintaining moderate complexity, thus avoiding issues such as nonconvergence. As illustrated in the figure below, R_1 and R_2 represent the polarization resistances, C_1 and C_2 are the polarization capacitors, R_0 is the ohmic internal resistance of the battery, U_{oc} denotes the open-circuit voltage, and U_t is the terminal voltage.

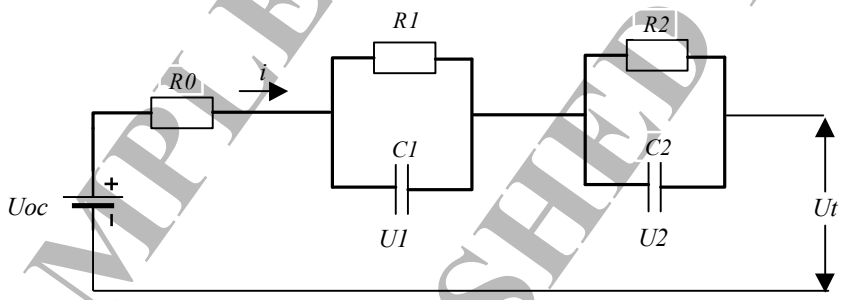

Figure 1: Model of a Second-Order RC Battery

On the basis of Kirchhoff's laws, the following set of equations is established:

$$\begin{cases} U_{oc} = U_t + U_1 + U_2 + iR_0 \\ \dfrac{\mathrm{d}U_1}{\mathrm{d}t} = \dfrac{i}{C_1} - \dfrac{U_1}{R_1C_1} \\ \dfrac{\mathrm{d}U_2}{\mathrm{d}t} = \dfrac{i}{C_2} - \dfrac{U_2}{R_2C_2} \end{cases} \quad (1)$$

The preliminary calculation formula for the SOC based on the ampere-hour integration method is as follows:

$$\mathrm{SOC}(t) = \mathrm{SOC}(0) - \frac{\eta \int_0^t i(t)\mathrm{d}t}{C_0} \quad (2)$$

In this formula, $SOC(t)$ represents the battery's state of charge at time, $SOC(0)$ is the initial state of charge, denotes the current, η is the Coulombic efficiency, and C_0 is the battery capacity.

By combining Equations (1) and (2) and selecting, SOC, U_1 and U_2 as state variables, with U_t as the output variable, the discrete state-space equation can be derived as follows:

$$\begin{cases} \begin{bmatrix} SOC(k) \\ U_1(k) \\ U_2(k) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \exp\left(-\frac{t_s}{R_1C_1}\right) & 0 \\ 0 & 0 & \exp\left(-\frac{t_s}{R_2C_2}\right) \end{bmatrix} \begin{bmatrix} SOC(k-1) \\ U_1(k-1) \\ U_2(k-1) \end{bmatrix} + \begin{bmatrix} -\frac{t_s}{Q} \\ R_1\left(1-\exp\left(-\frac{t_s}{R_1C_1}\right)\right) \\ R_2\left(1-\exp\left(-\frac{t_s}{R_2C_2}\right)\right) \end{bmatrix} [i(k-1)] \\ U_t(k) = U_{oc}(SOC(k)) - R_0 i(k) - U_1(k) - U_2(k) \end{cases} \quad (3)$$

In this equation, t_s denotes the sampling time, and Q represents the battery capacity.

2.2. Parameter Identification

The subject of this experiment is an 18650 lithium-ion battery. The nominal capacity of the battery is 1.5 Ah, and the ambient temperature is maintained at 25 °C. The dataset was obtained through the hybrid pulse power characterization (HPPC) discharge test, which provides critical operating data for battery modeling. The HPPC test applies standard discharge pulses at various states of charge (SOC) to capture the voltage response, laying a foundation for subsequent parameter identification. The detailed test procedure is as follows [7]:

- Fully charge the battery via constant current–constant voltage (CC–CV) charging and then let it rest.
- Divide the SOC into equal intervals (typically 10%) and discharge at 0.2C current until the SOC reaches the target interval.
- The battery rests to stabilize its electrochemical state, and the open-circuit voltage (OCV) is recorded; then, the battery continues discharging for 10 s at a current of 1C.
- After a 1-minute rest, the battery was charged at 0.75C for 10 s, followed by another 2-minute rest.
- Repeat steps 2–4 until the battery reaches the minimum SOC.

Through HPPC testing, we not only obtain accurate voltage data corresponding to each SOC interval but also calculate the battery's ohmic and polarization resistances on the basis of the observed voltage changes, which serve as the basis for subsequent parameter identification.

2.3. Soc–Ocv Curves

The open-circuit voltage (OCV) of a battery is closely related to its state of charge (SOC). Before performing parameter identification, it is necessary to establish the relationship between the two—namely, the SOC–OCV curve. By fitting the curve, the OCV–SOC relationship can be obtained as shown below.

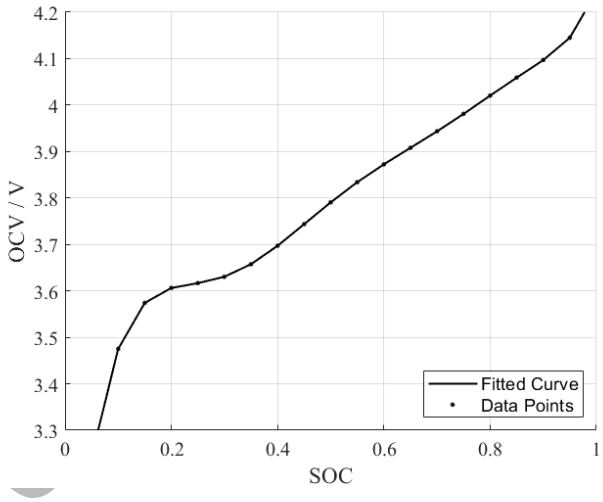

Figure 2: SOC–OCV fitting curve

2.4. Parameter Identification via FFRLS

To perform parameter identification via the forgetting factor recursive least squares (FFRLS) method, it is necessary to transform the equivalent circuit model. By applying the Laplace transform to Equation (1), we can obtain:

$$U_{oc}(s)=i(s)\left(R_0+\frac{R_1}{1+R_1C_1s}+\frac{R_2}{1+R_2C_2s}\right)+U_0(s) \tag{4}$$

The transfer function of the second-order Thevenin model is as follows:

$$\begin{aligned}G(s)&=\frac{U_{oc}(s)-U_0(s)}{i(s)}=R_0+\frac{R_1}{1+R_1C_1s}+\frac{R_2}{1+R_2C_2s}\\&=\frac{R_0\tau_1\tau_2s^2+(R_0\tau_1+R_0\tau_2+R_1\tau_3+R_2\tau_1)s+R_0+R_1+R_2}{\tau_1\tau_2s^2+(\tau_1+\tau_2)s+1}\end{aligned} \tag{5}$$

$\tau_2=R_2C_2$

In this equation, $\tau_1=R_1C_1$ and $\tau_2=R_2C_2$; then, applying the bilinear transformation yields:

$$s=\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}} \tag{6}$$

$$G(z^{-1})=\frac{c_3+c_4z^{-1}+c_5z^{-2}}{1-c_1z^{-1}-c_2z^{-2}} \tag{7}$$

The discretized form of Equation (6) in difference equation form can be written as:

$$U_{(k)}=U_{oc(k)}-U_{0(k)}=c_1U_{(k-1)}+c_2U_{(k-2)}+c_3i_{(k)}+c_4i_{(k-1)}+c_5i_{(k-2)} \tag{8}$$

Among them:

$$\begin{cases}R_0=(c_3-c_4+c_5)/(1+c_1-c_2)\\ \tau_1\tau_2=T^2(1+c_1-c_2)/[4(1-c_1-c_2)]\\ \tau_1+\tau_2=T(1+c_2)/(1-c_1-c_2)\\ R_0+R_1+R_2=(c_3+c_4+c_5)/(1-c_1-c_2)\\ R_0\tau_1+R_0\tau_2+R_2\tau_1+R_1\tau_2=4(c_3-c_5)/[T(1-c_1-c_2)]\end{cases} \tag{9}$$

Equation (7) is transformed into the basic form of the FFRLS algorithm:

$$\begin{cases}y_{(k)}=\boldsymbol{\theta}_{(k)}\phi_{(k)}\\ \boldsymbol{\theta}_{(k)}=\begin{bmatrix}c_1 & c_2 & c_3 & c_4 & c_5\end{bmatrix}\\ \phi_{(k)}=\begin{bmatrix}U_{(k-1)} & U_{(k-2)} & i_{(k)} & i_{(k-1)} & i_{(k-2)}\end{bmatrix}^{\mathrm{T}}\end{cases} \tag{10}$$

According to the FFRLS algorithm:

$$\begin{cases}\boldsymbol{\theta}_{(k)}=\boldsymbol{\theta}_{(k-1)}+\mathbf{K}_{(k)}\left[\mathbf{y}_{(k)}-\phi^{\mathrm{T}}{}_{(k)}\boldsymbol{\theta}_{(k-1)}\right]\\ \mathbf{K}_{(k)}=\dfrac{\mathbf{P}_{(k-1)}\phi_{(k)}}{\lambda+\phi^{\mathrm{T}}{}_{(k)}\mathbf{P}_{(k-1)}\phi_{(k)}}\\ \mathbf{P}_{(k)}=\dfrac{1}{\lambda}\left[\mathbf{I}-\mathbf{K}_{(k)}\phi^{\mathrm{T}}{}_{(k)}\right]\mathbf{P}_{(k-1)}\end{cases} \tag{11}$$

In this formula, θ_k is the parameter vector ϕ_k to be estimated, is the known input vector, K_k is the gain vector, _P_k is the error vector, λ and is the forgetting factor. The system parameters are estimated step by step through recursive calculations. The battery parameters can be obtained by substituting the estimated system parameters θ into equation (8).

2.5. Extended Kalman Filter Algorithm

The extended Kalman filter (EKF) is an improved version of the traditional Kalman filter (KF). Since the Kalman filter is applicable only to linear systems and many real-world engineering systems are nonlinear, the EKF was developed to address this limitation. Its core idea is to approximate the nonlinear functions through local Taylor expansion, thereby converting them into linear problems that can be handled with standard Kalman filtering techniques. For a nonlinear system, the state–space equations can be expressed as follows:

$$\begin{cases} x_k = A_{k-1}x_{k-1} + B_{k-1}u_{k-1} + \omega_k \\ y_k = C_k x_k + D_k u_k + v_k \end{cases} \tag{12}$$

In this equation, ω_k and v_k represent the unknown process noise and measurement noise, respectively, and x_k is the system state vector at time step k, corresponding to $[SOC(k)\ U_1(k)\ U_2(k)]^T$ in Equation (3). u_k is the input vector at time k, corresponding to the current $i(k)$ in Equation (3), and y_k is the output variable at time k, corresponding to the terminal voltage $U_t(k)$ in Equation (3). The matrices B, C, and D represent the system parameter matrices, specifically:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & exp^{\frac{-t_s}{R_1C_1}} & 0 \\ 0 & 0 & exp^{\frac{-t_s}{R_2C_2}} \end{bmatrix}, B = \begin{bmatrix} \frac{-t_s}{Q} \\ R_1(1-exp\frac{-t_s}{R_1C_1}) \\ R_2(1-exp\frac{-t_s}{R_2C_2}) \end{bmatrix}, C = \left[\frac{\partial U_{oc}(k)}{\partial SOC} - 1 - 1\right], D = [-R_0]$$

2.6. Adaptive Extended Kalman Filter Algorithm

Since the traditional EKF algorithm assumes a constant process and measurement noise, which limits its applicability and accuracy, this paper adopts the adaptive extended Kalman filter (AEKF) algorithm. The core idea of the AEKF lies in dynamically adjusting the noise parameters to significantly enhance estimation accuracy. Specifically, the algorithm introduces the following:

$$\begin{cases} e_k = y_k - \hat{y}_{k|k-1} \\ Q_k = (1-m)Q_{k-1} + mK_k e_k e_k^{\cdot} K_k^{\cdot} \\ R_k = (1-m)R_{k-1} + m\left(e_k e_k^{\cdot} - C_k P_{k|k} C_k^{\cdot}\right) \end{cases} \tag{13}$$

In this equation, e_k represents the measurement error, whereas Q_k and R_k denote the process noise covariance matrix and the observation noise covariance matrix, respectively. In this paper, both Q and R are set to 0.01. The symbol m represents the weight forgetting factor, which is chosen as 0.98 in this paper. The main steps of the AEKF algorithm are as follows:

(1) Set the initial values of the matrices x_0, P_0, Q_0, and R_0.

(2) State prediction and covariance prediction: $\begin{cases} x_k^- = A_{k-1}x_{k-1}^+ + B_{k-1}u_{k-1} \\ P_k^- = A_{k-1}P_{k-1}^+ A_{k-1}^{\cdot} + Q_{k-1} \end{cases}$

(3) State parameter iteration: $\begin{cases} K_k = P_k^- C_k^{\cdot}\left(C_k P_k^- C_k^{\cdot} + R_k\right)^{-1} \\ x_k^+ = x_k^- + K_k\left(y_k - C_k x_k^-\right) \\ P_k^+ = \left(I - K_k C_k\right)P_k^- \end{cases}$

(4) Adaptive update: $\begin{cases} Q_k = (1-m)Q_{k-1} + mK_k e_k e_k^{\cdot} K_k^{\cdot} \\ R_k = (1-m)R_{k-1} + m\left(e_k e_k^{\cdot} - C_k P_k^- C_k^{\cdot}\right) \end{cases}$

2.7. LSTM Neural Network

Long short-term memory (LSTM) is a special type of recurrent neural network (RNN) that replaces ordinary hidden nodes with memory cells. It addresses the problems of gradient vanishing and explosion that traditional RNNs often encounter during backpropagation,

enabling it to effectively learn long-term dependencies. LSTM performs well in time series prediction tasks and is therefore used in this study to estimate the state of charge of lithium-ion batteries [8].

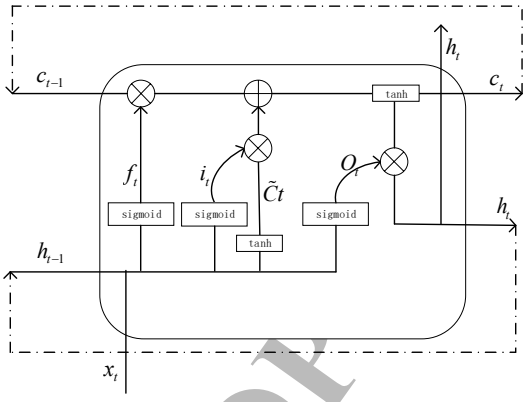

Figure 3: LSTM Network Structure Diagram

LSTM introduces a gating mechanism and memory cell to control the flow of information. Its general structure is shown in the figure, with core components including the forget gate, input gate, control unit (memory cell), and output gate. First, the input gate processes the input vector; then, the forget gate and input gate are calculated. On the basis of their results, the long-term memory is updated. Finally, the output gate is computed to determine the final output. The specific formulas are as follows:

In the above formulas, W represents the weight matrix, b represents the bias vector, x_t represents the current input vector, h_t represents the current hidden state, C_t represents the current memory cell state, o_t represents the current output, σ represents the sigmoid activation function, and *tanh* represents the hyperbolic tangent function.

3. Results

3.1. Joint Algorithm Design of FFRLS-AEKF-LSTM

The LSTM neural network can effectively handle long-term dependency issues, including trends and periodic changes. Therefore, it can be trained offline for regression prediction. This paper proposes a joint FFRLS-AEKF-LSTM algorithm. The parameters obtained through the FFRLS-EKF algorithm not only expand the LSTM network dataset but also exhibit a high correlation with the target SOC values, making them more suitable as input parameters to improve prediction accuracy.

First, operational condition data are input. The FFRLS algorithm is used to identify battery parameters, and the initial SOC value is obtained through the OCV-SOC relationship as the input for the AEKF algorithm. The AEKF algorithm then performs iterative filtering and outputs the Kalman gain K_k, covariance matrix P_k, and estimation error *eSOC*. These, along with the battery terminal voltage and current, are used as inputs for the LSTM network, thus expanding the network's dataset. The actual SOC value is used as the network output to ensure feature relevance. Finally, the neural network is constructed and trained. Eighty percent of the total dataset is used as the training set, and the remaining 20% is used as the test set to validate the network's accuracy. The flowchart of the FFRLS-AEKF-LSTM algorithm is shown below.

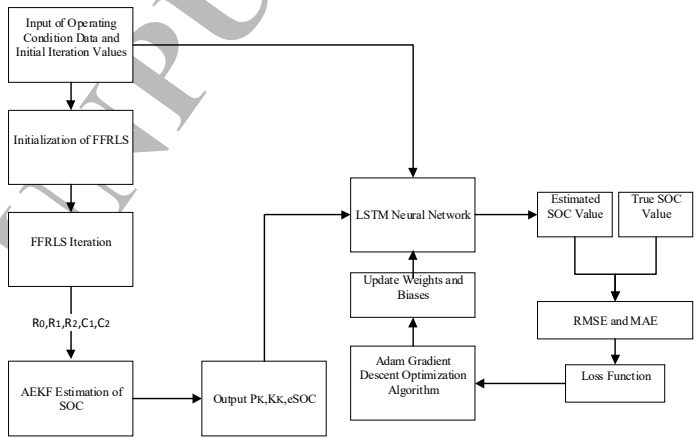

Figure 4: Flow of the FFRLS-AEKF-LSTM algorithm

3.2. Experimental Results

In this experiment, battery discharge data under HPPC conditions were used as the research subject. After completing the theoretical research, a simulation model of the combined algorithm was built via MATLAB software. The LSTM neural network module was implemented on a hardware platform equipped with an Intel i5-13650HX CPU and an NVIDIA 4060 laptop GPU with 8 GB memory.

The LSTM neural network was configured with an input layer of dimension 5 and a fully connected output layer of dimension 1. The hidden layers consisted of two stacked LSTM layers, followed by a dropout layer with a rate of 0.2 to enhance generalization and prevent overfitting. The initial learning rate was set to 0.005, and the optimizer used was Adam, with its learning rate adjusted to 0.01 to facilitate faster convergence of the network.

As shown in the figure below, by the 150th epoch and after approximately 500 iterations, the SOC prediction error became stable, with the RMSE remaining below 0.1.

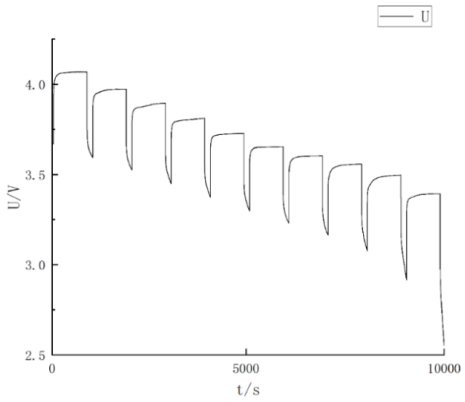

Figure 5: Battery voltage data under HPPC Conditions

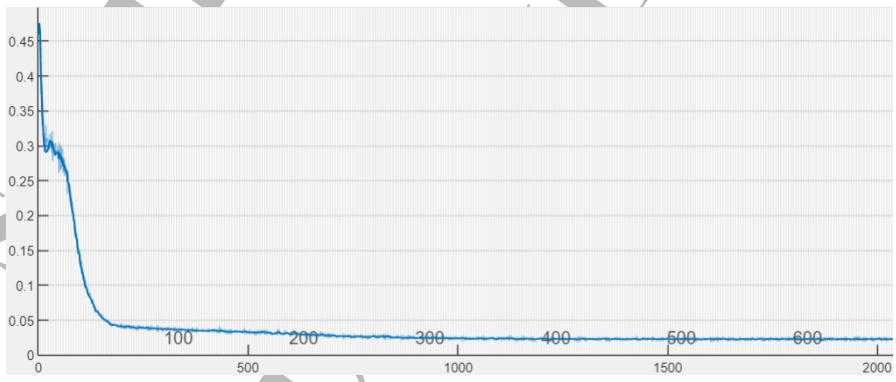

Figure 6: RMSE iteration curve

In this experiment, the mean absolute error (MAE) and root mean square error (RMSE) are used as the primary evaluation metrics. Both are common methods for measuring the difference between the predicted and true values. The MAE calculates the average of the absolute differences between the predictions and actual values, reflecting the overall average deviation of the model. A smaller MAE value indicates better prediction accuracy, and it is less sensitive to outliers. In contrast, the RMSE is the square root of the mean of the squared prediction errors, which penalizes larger errors more heavily and better captures the fluctuations in prediction accuracy. The formulas for both metrics are shown below:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad (15)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

In the above formulas, y_i represents the true value, and $\hat{y}_i$ represents the predicted value.

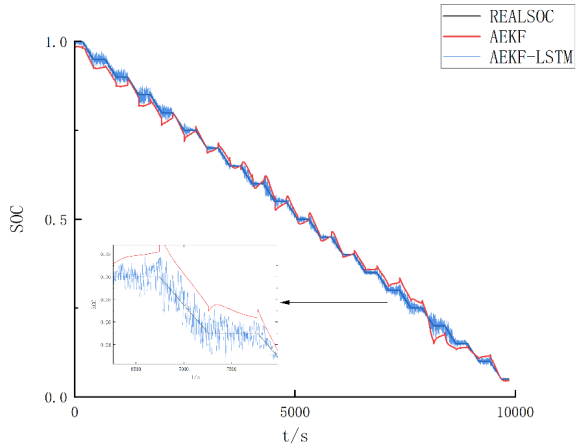

Figure 7: SOC Prediction Results for the Test Set

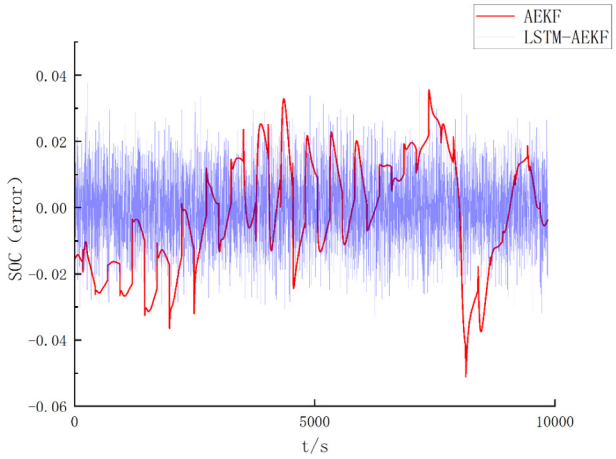

Figure 8: Comparison of SOC Estimation Errors under Different Algorithms

4. Conclusion

The SOC prediction results of the FFRLS-AEKF algorithm and the FFRLS-AEKF-LSTM combined algorithm were compared based on 10,000 recorded data points, as shown in Figure 7. The prediction errors of both algorithms relative to the actual SOC are illustrated in Figure 8. It can be observed that the FFRLS-AEKF-LSTM algorithm exhibits better convergence in SOC prediction compared to the FFRLS-AEKF algorithm, with a maximum error of less than 0.04, while the maximum error of the FFRLS-AEKF algorithm reaches 0.05.

Specifically, the RMSE and MAE of the FFRLS-AEKF method are 1.47% and 1.72%, respectively, whereas the RMSE and MAE of the proposed FFRLS-AEKF-LSTM method are reduced to 0.83% and 1.04%, respectively. This demonstrates the superior accuracy and robustness of the proposed method. Additionally, the model exhibits good convergence during training, with the prediction error stabilizing after approximately 150 epochs. These results further validate the feasibility and practical value of the combined algorithm in modeling complex battery systems, providing a solid theoretical and engineering foundation for high-precision SOC estimation.

Discussion

This paper focuses on the estimation of the lithium-ion battery state of charge and proposes a hybrid estimation algorithm based on FFRLS-AEKF-LSTM by integrating parameter identification, filtering algorithms, and deep learning models. The proposed method fully utilizes the dynamic parameter identification capability of the FFRLS algorithm, the nonlinear state optimization strength of the AEKF algorithm, and the temporal modeling advantages of the LSTM neural network, thereby achieving more accurate SOC prediction.

The experimental results demonstrate that the proposed hybrid algorithm outperforms traditional methods in both accuracy and stability, exhibiting good convergence and robustness. The RMSE and MAE of the predicted results remain below 1%, meeting the high-precision

SOC estimation requirements of battery management systems in new energy vehicles and energy storage systems. This research provides an effective algorithmic foundation for the realization of intelligent battery management and contributes technical insights for the safe and efficient operation of the new energy industry under China's "dual carbon" strategy.

References

1. Yu, D. F., Liu, X., Du, T., et al. (2025). Analysis of the Current Situation and Frontier Trends of Carbon Neutrality in China's Automotive Industry Based on Knowledge Graphs [J]. *Times Auto,* (11): 8–11.
2. Miao, Y. (2024). *Research on state of charge estimation of electric vehicle battery systems*.
3. Cui, X., Huang, Y., & Wu, X. (2024). *A review of methods and applications for lithium battery state-of-charge estimation*. Electronic Measurement Technology, 47(20), 41–59.
4. Zhao, J. (2018). *Modeling and simulation of lithium-ion battery SOC estimation based on second-order* EKF [Master's thesis, Xi'an University of Science and Technology].
5. 贺伟, 马鸿雁, 张英达, 李晟延, & 王帅. (2023). 基于改进门控循环单元神经网络的锂电池组荷电状态预测. 科学技术与工程, 23(12), 5102-5109.
6. Li, Q., Hong, Z., Lin, H., et al. (2024). *Research on power battery parameter modeling and SOC prediction based on RC equivalent circuit models of different orders*. Auto Magazine, (04), 83–85.
7. Li, C., Zhang, Y., Zhao, Y., et al. (2024). *Online parameter identification of lithium battery equivalent model based on HPPC-FFRLS. Journal of Henan Polytechnic University* (Natural Science Edition), 43(05), 127–132.
8. Gao, J., Hao, S., & Dang, Y. (2025). R*emaining useful life prediction of lithium-ion batteries based on LSTM-Transformer model parameter optimization* [J/OL]. Automotive Engineer, 1–7.