batteries

*Article*

# State of Charge Estimation for Sodium-Ion Batteries Based on LSTM Network and Unscented Kalman Filter

**Xiangang Zuo [1], Xiaoheng Fu [1], Xu Han [1], Meng Sun [2] and Yuqian Fan [2,3,*]**

[1] School of Information Engineering, Henan Institute of Science and Technology, Xinxiang 453003, China; zuoxg2002@hist.edu.cn (X.Z.); 21202402237@stu.hist.edu.cn (X.F.); 942s@stu.hist.edu.cn (X.H.)
[2] School of Computer Science and Technology, Henan Institute of Science and Technology, Xinxiang 453003, China; 21202402213@stu.hist.edu.cn
[3] Dongguan Institute, Sun Yat-sen University, Dongguan 523000, China
* Correspondence: fanyq3@mail3.sysu.edu.cn

**Abstract**

With the increasing application of sodium-ion batteries in energy storage systems, accurate state of charge (SOC) estimation plays a vital role in ensuring both available battery capacity and operational safety. Traditional Kalman-filter-based methods often suffer from limited model expressiveness or oversimplified physical assumptions, making it difficult to balance accuracy and robustness under complex operating conditions, which may lead to unreliable estimation results. To address these challenges, this paper proposes a hybrid framework that combines an unscented Kalman filter (UKF) with a long short-term memory (LSTM) neural network for SOC estimation. Under various driving conditions, the UKF—based on a second-order equivalent circuit model with online parameter identification—provides physically interpretable estimates, while LSTM effectively captures complex temporal dependencies. Experimental results under CLTC, NEDC, and WLTC cycles demonstrate that the proposed LSTM-UKF approach reduces the mean absolute error (MAE) by an average of 2% and the root mean square error (RMSE) by an average of 3% compared to standalone methods. The proposed framework exhibits excellent adaptability across different scenarios, offering a precise, stable, and robust solution for SOC estimation in sodium-ion batteries.

**Keywords:** SOC estimation; long short-term memory network; unscented Kalman filter; equivalent circuit model



Academic Editor: King Jet Tseng

Received: 23 June 2025
Revised: 13 July 2025
Accepted: 17 July 2025
Published: 18 July 2025

**Citation:** Zuo, X.; Fu, X.; Han, X.; Sun, M.; Fan, Y. State of Charge Estimation for Sodium-Ion Batteries Based on LSTM Network and Unscented Kalman Filter. *Batteries* **2025**, *11*, 274. https://doi.org/10.3390/batteries11070274



## 1. Introduction

Sodium-ion batteries (SIBs), as an emerging energy storage technology, have gained considerable attention due to their advantages, such as low cost, environmental friendliness, and abundance of raw materials. These characteristics make them highly suitable for various application scenarios including stationary energy storage, electric vehicles, low-speed electric vehicles, and renewable energy systems [1]. With ongoing technological advancements and cost reductions, SIBs are expected to achieve even broader adoption in the near future [2].

State of charge (SOC) is a critical indicator for battery energy management and operational control. Accurate SOC estimation not only helps prevent overcharging or over-discharging of the battery but also extends the battery's lifespan and ensures the safe operation of the battery system [3–5]. As a core function of the battery management system (BMS), SOC estimation is inherently challenging due to its sensitivity to factors such as

aging, temperature, and operational dynamics. Nevertheless, precise SOC estimation is crucial as it prevents unexpected system failures and ensures safety for both the battery and its users [6]. Furthermore, by monitoring the rate of SOC change during charge and discharge cycles, users can intuitively evaluate the battery's health status and remaining useful life [7–10].

Current research on SOC estimation methods can be broadly categorized into three groups: traditional estimation methods, model-based methods, and learning-based methods [11].

Traditional approaches mainly include the ampere-hour integration method and the open-circuit voltage (OCV) method. The ampere-hour integration method, also known as Coulomb counting, estimates SOC by integrating the charge/discharge current over time [12]. The OCV method predicts SOC based on the known relationship between the battery's OCV and its SOC [13]. However, the ampere-hour integration method is an open-loop technique and is highly sensitive to initial SOC errors and current measurement noise. Meanwhile, the OCV method requires long rest periods to obtain accurate readings, which is impractical for real-time applications [14].

Model-based estimation methods are primarily represented by the Kalman filter family. These algorithms have demonstrated significant advantages in battery SOC prediction due to their ability to fuse model dynamics with real-time measurements through recursive estimation [15]. The extended Kalman filter (EKF), for example, estimates SOC by minimizing the covariance between measured and predicted values using a battery model. However, as battery systems are inherently nonlinear, the linearization process in the EKF often leads to estimation accuracy loss [16]. The unscented Kalman filter (UKF) addresses this limitation by using the nonlinear transformation of sigma points, thereby improving estimation accuracy without requiring linearization. Essentially, these methods construct an equivalent circuit model (ECM) to translate battery dynamics into a mathematically tractable form, providing physically interpretable estimations—an important distinction from purely data-driven methods. For instance, Machin A et al. [17] propose a UKF–based SOC estimation method with favorable performance in terms of computational efficiency and estimation accuracy. Takyi-Aninakwa P et al. [18] introduce a Kalman-filter-based ECM tailored for electric vehicle applications, though lacking experimental validation. Zhang Y et al. [19] introduce an adaptive singular value decomposition-based UKF that integrates all unknown variables while maintaining a vector-form covariance matrix for enhanced fusion estimation.

Learning-based methods, also referred to as data-driven methods, leverage machine learning algorithms for SOC estimation. These methods are known for their strong nonlinear fitting capabilities and high estimation accuracy, though they often require large amounts of high-quality offline training data. Their performance is significantly influenced by dataset quality and the training strategy. Typical implementations involve constructing nonlinear mappings between input features and SOC using neural networks, with approaches such as end-to-end estimation, hybrid frameworks, and feature enhancement techniques. For example, Prasanthi A et al. [7] present a bidirectional long short-term memory-based hybrid prediction method, achieving high estimation accuracy but facing increasing computational demands with larger data volumes. Wang J et al. [20] propose an LSTM–based method that reduces SOC estimation errors under aging conditions, although its feedback mechanism depends heavily on accurate voltage prediction and is sensitive to noise, requiring frequent parameter updates. Despite their strong nonlinearity handling, these methods often suffer from issues such as high data dependency, poor online adaptability, and limited interpretability. To address these challenges, some studies have integrated learning-based methods with Kalman filtering. Xu H et al. [21] propose a hybrid method

combining genetic algorithms, LSTM, and improved adaptive Kalman filtering, showing robust performance but requiring retraining over time. Zhang Z et al. [22] introduce an EKF combined with particle swarm optimization and LSTM for accurate SOC estimation in power batteries, effectively mitigating estimation errors under noisy conditions. Xie Y et al. [23] compare filter-based and deep learning-based SOC estimation methods, including recursive least squares (RLS), EKF, and gated RNNs, showing improved accuracy and reduced computational time, although there is a lack of neural network structure optimization. Chen J et al. [24] introduce an enhanced method using a multi-layer LSTM and suboptimal fading EKF capable of fast convergence and strong tracking ability for dynamic data, demonstrating high estimation accuracy and robustness.

In recent years, data-driven methods have emerged as some of the most promising solutions for SOC estimation, aiming to overcome the limitations of traditional and model-based approaches. Neural networks, with their strong adaptability, self-learning capability, and powerful modeling performance, have been widely used to estimate the SOC of SIBs [25]. These model-free approaches depend heavily on the quality and quantity of input data, as well as on proper training and hyperparameter selection. In such methods, the battery is treated as a "black box" rather than a physical model, reducing computational complexity and the number of required variables. They directly model complex nonlinear relationships, such as electrochemical reactions, aging effects, and self-discharge, using extracted features without prior knowledge of the system [26]. These methods aim to estimate SOC by learning from nonlinear battery data, considering factors such as ohmic resistance, polarization effects, and self-discharge, thus achieving high accuracy. This is in contrast to model-based approaches, which require these factors to be explicitly incorporated into the observer design [27].

In SOC estimation, single-method strategies face notable technical bottlenecks. Data-driven methods based on LSTM, for example, require high-quality, labeled datasets that cover a wide range of operating conditions, which are costly to acquire and offer limited generalization to unseen scenarios. Model-based methods such as the UKF depend on the accuracy of ECMs, where the complexity of parameter identification increases exponentially with model order—typically beyond third-order models—and where multi-state estimation under time-varying conditions imposes significant computational burdens on embedded systems. These challenges highlight the difficulty of achieving a balance between robustness, adaptability, and real-time performance using a single approach.

Motivated by these limitations, a hybrid LSTM-UKF method is proposed in this study. The method combines the long-term temporal modeling capabilities of LSTM with the accurate state estimation features of the UKF in nonlinear systems, aiming to leverage observational battery data for improved SOC estimation. By integrating LSTM and UKF with a forgetting-factor recursive least squares (FFRLS) algorithm for the online identification of RC model parameters, the proposed approach effectively overcomes the drawbacks of standalone methods and enhances estimation stability and accuracy. Experimental results show that the fused model reduces prediction error by 2% compared to single-method approaches. Furthermore, the effects of training and testing under different operating conditions on the accuracy of the LSTM-UKF framework are also investigated.

The remainder of this paper is structured as follows: Section 2 introduces the battery model, LSTM network, and UKF algorithm; Section 3 presents the algorithm design; Section 4 discusses the performance of the proposed LSTM-UKF method under various operating conditions; and Section 5 provides the conclusions.

## 2. Model Development

### *2.1. Battery Modeling*

An accurate battery model is essential for identifying the external characteristics of the battery and achieving high SOC estimation accuracy. The battery model typically consists of resistors, capacitors, and voltage sources arranged as an equivalent circuit to characterize the internal behavior of SIBs. Moreover, the accuracy and reliability of the model improve with its order. Although a first-order RC model can be used to identify the parameters of SIBs, it fails to capture the system dynamics with sufficient precision—particularly under varying temperature and terminal voltage conditions.

To address this limitation, a second-order RC model was adopted. This model not only provides a more comprehensive representation of the battery's external characteristics but also enables more accurate identification of internal parameters. Therefore, the dynamic behavior of the sodium-ion battery was represented using a second-order RC ECM as illustrated in Figure 1.

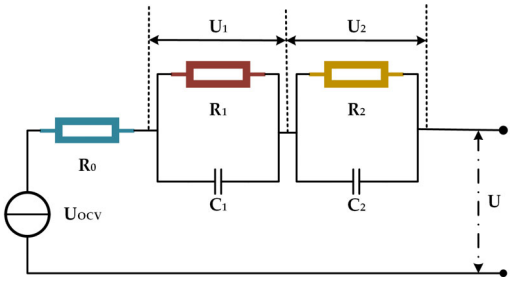


**Figure 1.** Second-order RC equivalent circuit model.

The model consists of an ideal voltage source $U_{ocv}$, input current $I_t$, ohmic internal resistance $R_0$, polarization resistances $R_1$ and $R_2$, equivalent polarization capacitors $C_1$, $C_2$, and battery terminal voltage $U$. According to Kirchhoff's voltage law, the model equation is expressed as Equation (1):

$$\begin{cases} U = U_{ocv}(SOC(t)) - U_1 - U_2 - I(t)R_0 \\ C_1 \frac{dU_1}{dt} = I(t) - \frac{U_1}{R_1} \\ C_2 \frac{dU_2}{dt} = I(t) - \frac{U_2}{R_2} \end{cases}, \quad (1)$$

In the equation, $U_{ocv}(SOC(t))$ denotes the OCV as a function of the SOC, where SOC itself is a function of time $t$.

### *2.2. UKF*

In sodium-ion battery SOC prediction, the UKF is commonly combined with an ECM to form a hybrid framework, integrating physical modeling and state estimation. A second-order ECM is established, with parameters identified through experimental data, and the state-space equations are constructed accordingly. The UKF is employed for state estimation: after initialization, a set of sigma points is generated via the unscented transform, and the prior state is predicted through nonlinear state equations; then, the posterior state is updated by incorporating the measured terminal voltage via the observation equation. This framework achieves closed-loop correction by dynamically adjusting the noise covariance, accurately handling the nonlinear characteristics between SOC and OCV, and maintaining the SOC estimation error within $\pm 3\%$ while preserving the physical interpretability of the model. The main steps are as follows:

Sigma Point Calculation

$$A = \alpha^2(n + \kappa) - n, \quad (2)$$

$$W_0^m = \frac{A}{n+\lambda}, W_0^c = \frac{A}{n+\lambda} + \left(1-\alpha^2+\beta\right), \tag{3}$$

$$W_i^m = W_i^c = \frac{1}{2(n+\lambda)}, i = 1,2,\ldots,2n, \tag{4}$$

Initialization

$$x_0 = E(x_0), P_0 = E\left[(x_0-\bar{x}_0)(x_0-\bar{x}_0)^T\right], \tag{5}$$

Sigma Point Sampling

$$x_{k-1}^0 = x_{k-1}, \tag{6}$$

$$x_{k-1}^i = x_{k-1} + \sqrt{(n+\lambda)P_{k-1}}, x_{k-1}^{i+n} = x_{k-1} - \sqrt{(n+\lambda)P_{k-1}}, i = 1,\ldots,n, \tag{7}$$

Time Update

$$x_{k|k-1} = \sum_{i=0}^{2n} W_i^m f\left(x_{k-1}^i\right), \tag{8}$$

$$P_{k|k-1} = \sum_{i=0}^{2n} W_i^c \left[f\left(x_{k-1}^i\right) - x_{k|k-1}\right]\left[f\left(x_{k-1}^i\right) - x_{k|k-1}\right]^T + Q_k, \tag{9}$$

Measurement Update

$$y_k^i = g\left(x_k^i\right), \tag{10}$$

$$\hat{y}_k = \sum_{i=0}^{2n} W_i^m y_k^i, \tag{11}$$

$$P_{y_k} = \sum_{i=0}^{2n} W_i^c \left(y_k^i - \hat{y}_k\right)\left(y_k^i - \hat{y}_k\right)^T + R_k, \tag{12}$$

$$P_{xy_k} = \sum_{i=0}^{2n} W_i^c \left(x_k^i - x_{k|k-1}\right)\left(y_k^i - \hat{y}_k\right)^T, \tag{13}$$

Kalman Gain and State Update

$$K_k = P_{xy_k} P_{y_k}^{-1}, \tag{14}$$

$$x_k = x_{k|k-1} + K_k(y_k - \hat{y}_k), \tag{15}$$

$$P_k = P_{k|k-1} - K_k P_{y_k} K_k^T, \tag{16}$$

This approach effectively suppresses the influences caused by sensor errors, model parameter uncertainties, and environmental variations, enabling the real-time and accurate estimation of the battery's internal states. It can efficiently capture higher-order statistical characteristics under nonlinear transformations, thereby significantly improving estimation accuracy and numerical stability.

### *2.3. LSTM*

In the task of SOC prediction for SIBs, the LSTM model, as a purely data-driven approach, can directly learn the nonlinear dynamic characteristics embedded in the battery's charging and discharging processes from multidimensional historical input data, such as current, voltage, and temperature, thus avoiding the complexity of physical modeling. Traditional recurrent neural networks often struggle to capture long-term dependencies due to issues such as vanishing or exploding gradients when processing long sequences. As illustrated in Figure 2, LSTM overcomes these limitations by introducing memory cells and gating mechanisms—including input, forget, and output gates—that selectively retain or discard information. This enables the network to effectively model key information over long time sequences, significantly enhancing its ability to extract temporal features.

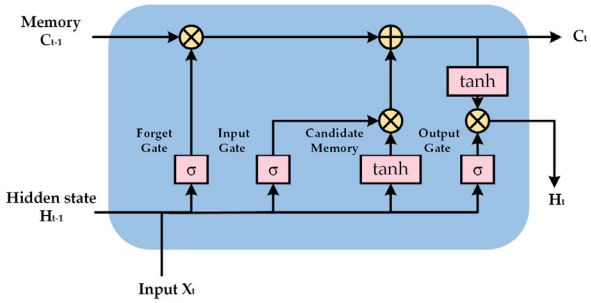


**Figure 2.** LSTM.

This method possesses strong generalization capability and high prediction accuracy, making it particularly suitable for modeling dynamic battery behaviors under complex operating conditions. With the advancement of big data and computational power, the application of LSTM in intelligent BMSs has become increasingly widespread. It provides robust support for achieving high-precision, real-time SOC estimation and represents a promising direction for future data-driven battery modeling and management.

## 3. Algorithm Design

To enhance robustness and accuracy in nonlinear SOC estimation, a hybrid LSTM-UKF strategy was introduced. While UKF provides real-time error correction and physical interpretability, LSTM captures nonlinear time-dependent behaviors. Their outputs were dynamically fused using weights updated via FFRLS, forming an adaptive SOC estimation architecture.

### *3.1. Algorithm Procedure*

By analyzing the current mainstream SOC estimation algorithms, a hybrid SOC estimation method is proposed that integrates LSTM and UKF. The aim of this fusion algorithm is to combine the theoretical advantages of physical modeling with the temporal learning capability of data-driven methods to achieve high-precision SOC prediction under dynamically complex operating conditions. The overall implementation process of the algorithm is illustrated in Figure 3 and mainly consists of three parts: a parameter identification and UKF module, an LSTM module, and a fusion mechanism.

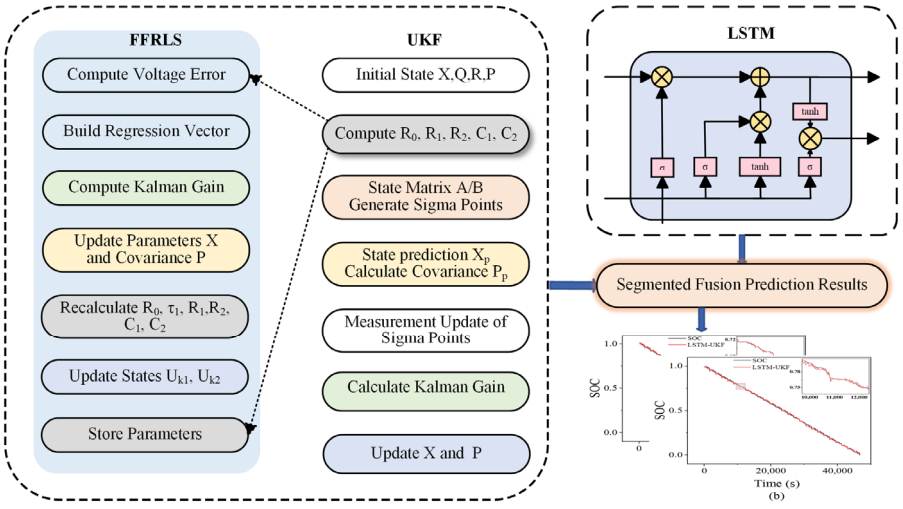


**Figure 3.** Algorithm process.

Firstly, the parameters of the second-order ECM were calculated using the FFRLS algorithm. Then, the identified parameters were employed to initialize the battery system

state variables in the UKF module, including the state vector and covariance matrix. The state transition matrix was derived based on the constructed ECM of the battery. Subsequently, the state was predicted and updated using observation data; the Kalman gain was computed; and the current SOC of the battery was recursively estimated. This module effectively copes with noise interference and parameter uncertainties, demonstrating strong robustness.

The LSTM module extracts features from historical operational data, which are normalized and preprocessed before constructing the LSTM prediction model. This module deeply mines the nonlinear dynamic characteristics of the battery during different charge/discharge phases, thereby enhancing prediction accuracy.

During the fusion stage, to ensure the accuracy of the output results, the outputs of the UKF and LSTM were combined with different weighting ratios.

Although the LSTM-UKF fusion structure is generally applicable across battery chemistries, all components in this study were implemented specifically for SIBs. The equivalent circuit model parameters were identified using SIB–specific dynamic current-voltage data, and the OCV curve was fitted using experimental SIB OCV-SOC measurements. The FFRLS algorithm was also initialized using parameters tuned to the polarization behavior and time constants of SIBs. As such, no assumptions, model structures, or data from lithium-ion batteries were reused. This ensures that the modeling framework is intrinsically tailored to SIB characteristics.

### *3.2. Parameter Identification*

In this study, a battery parameter identification model was established, aimed at achieving accurate SOC estimation. During the filtering process of parameter identification, an accurate battery model is essential. The internal SOC of the battery is closely related to external voltage, load current, temperature, and other relevant variables. The accuracy of the model directly affects the estimation precision of the Kalman filter and thus determines the reliability of the results. Among various models, the ECM is widely adopted due to its convenience for parameter identification and its minimal interference with battery characteristics.

In system analysis, system identification refers to the process of selecting a model equivalent to the measured system from a predefined model class by collecting the system's input and output data. Simply put, system identification involves using optimization algorithms to iteratively adjust model parameters by minimizing the error between the model output and actual output until the optimal model is obtained. In the parameter identification domain, the least squares method serves as a fundamental estimation technique applicable to both static and dynamic systems, as well as linear and nonlinear systems. It is favored for its simplicity and ease of implementation. The core concept of least squares identification is to compute the optimal solution within a given model structure framework to minimize the sum of squared errors.

#### 3.2.1. RLS

The RLS algorithm belongs to the category of adaptive filtering algorithms, primarily used for the online estimation of parameters in dynamic systems. This algorithm obtains the optimal weight vector by minimizing the sum of squared errors and possesses the capability of recursively updating the estimates without the need to store historical data. Widely applied in system parameter identification, the RLS algorithm is typically derived based on the discrete-time least squares criterion to achieve parameter estimation. Its core

principle can be expressed as $U(t) = \varphi^T(t)$, where $\varphi(t)$ represents the observation data matrix and $\theta$ denotes the parameter vector as shown in Equation (17):

$$\begin{cases} \varphi(t) = \begin{bmatrix} U(t-1) & U(t-2) & I(t) & I(t-1) & I(t-2) \end{bmatrix}^T \\ \theta = \begin{bmatrix} k_1 & k_2 & k_3 & k_4 & k_5 \end{bmatrix} \end{cases} \tag{17}$$

The parameter identification of the RLS algorithm is expressed as Equation (18):

$$\begin{cases} \mathrm{K}(k+1) = \frac{P(k)\varphi(k+1)}{1+\varphi^T(k+1)P(k)\varphi(k+1)} \\ e(k+1) = y(k+1) - \varphi^T(k+1)\theta(k) \\ \theta(k+1) = \theta(k) + K(k+1)e(k+1) \\ P(k+1) = I - K(k+1)\varphi^T(k+1)P(k) \end{cases}, \tag{18}$$

where $P(k)$ is the covariance matrix; $\mathrm{K}(k)$ is the gain matrix; $e(k)$ is the error; $y$ is the output variable; and $I$ is the identity matrix.

3.2.2. FFRLS

The Forgetting Factor FFRLS algorithm is based on the RLS algorithm, incorporating a forgetting factor λ as a weight in the observation data matrix and system output vector. When new observation data arrive, the algorithm adjusts the weighting ratio of old and new data through exponential weighting, performing a weighted average to update the identified parameters. Thus, as the input variables change and observation data accumulate, the FFRLS algorithm can respond more quickly and obtain more optimal identification parameters.

When λ equals 1, the FFRLS algorithm degenerates to the traditional RLS algorithm. Since λ is a constant, if the parameter error is small during online identification, introducing λ may increase the parameter estimation error. However, when the parameter error is large, an appropriate choice of λ can optimize the algorithm by accelerating the convergence speed of the online identification process, thereby reducing the error. Therefore, selecting a suitable λ not only improves convergence speed but also effectively reduces the error.

To perform parameter identification of the established second-order RC ECM of the battery, the battery model needs to be transformed into a mathematical form applicable for least squares methods. Based on the ECM shown in Figure 1, the following can be derived:

$$U_{OCV}(s) = \left(R_0 + \frac{R_1}{R_1C_1s} + \frac{R_2}{R_2C_2s}\right)I(s) + U(s), \tag{19}$$

Let $\tau_1 = R_1C_1$ and $\tau_2 = R_2C_2$. By discretizing Equation (19) and introducing parameters $a_1$, $a_2$, $a_3$, $a_4$, and $a_5$, the following expression can be obtained:

$$\begin{cases} a_1 = \frac{-(\tau_1+\tau_2)T-2\tau_1\tau_2}{T^2+(\tau_1+\tau_2)T+\tau_1\tau_2} \\ a_2 = \frac{\tau_1\tau_2}{T^2+(\tau_1+\tau_2)T+\tau_1\tau_2} \\ a_3 = \frac{(R_0+R_1+R_2)T^2}{T^2+(\tau_1+\tau_2)T+\tau_1\tau_2} + \\ \quad \frac{[R_1\tau_2+R_2\tau_1+R_0(\tau_1+\tau_2)]T+\tau_1\tau_2R_0}{T^2+(\tau_1+\tau_2)T+\tau_1\tau_2} \\ a_4 = \frac{-[R_1\tau_2+R_2\tau_1+R_0(\tau_1+\tau_2)]T-2\tau_1\tau_2R_0}{T^2+(\tau_1+\tau_2)T+\tau_1\tau_2} \\ a_5 = \frac{\tau_1\tau_2R_0}{T^2+(\tau_1+\tau_2)T+\tau_1\tau_2} \end{cases}, \tag{20}$$

Equation (19) can thus be rewritten as follows:

$$\begin{gathered} U_{OCV}(k) - U(k) = a_1[U_{OCV}(k-1) - U(k-1)] + \\ a_2[U_{OCV}(k-2) - U(k-2)] + a_3 I(k) + a_4 I(k-1) +, \\ a_5 I(k-2) \end{gathered} \tag{21}$$

The above equation can be directly applied to the RLS identification method with a forgetting factor, where $\theta = [a1,\ a2,\ a3,\ a4,\ a5]^T$ is used as the parameter vector for direct identification. The battery model parameters can then be derived from the identification results of these parameters.

In the FFRLS algorithm, the choice of the forgetting factor λ is crucial. To achieve optimal identification performance, λ is usually determined experimentally. During experiments, different values of λ can be tested to observe their effects on the identification results. By comparing the identification errors and convergence speeds under various λ values, the most suitable λ for the current system can be selected. Once the optimal λ is determined, it can be applied in the FFRLS algorithm to precisely identify the battery model.

During the identification process, new observational data are continuously fed into the algorithm, and the weighting of data is adjusted according to the forgetting factor λ. As data accumulate, the algorithm gradually converges to the final identified parameters.

These identified parameters can be used to describe the internal characteristics and dynamic behavior of the battery, providing essential support for battery management and control. The FFRLS–based identification of the battery model allows for a more accurate prediction of the battery's remaining capacity and lifespan, thereby improving battery utilization efficiency and safety.

## 4. Experimental Analysis

### *4.1. Platform Setup*

In this study, voltage and current data of the battery under actual operating conditions were collected using a self-developed experimental platform. This setup aims to simulate the real working state of the battery and provide a solid foundation for validating the superiority of the proposed estimation method. During the fusion stage, to ensure the stability of the predictions, a weighted fusion of the outputs from UKF and LSTM with equal weights was initially applied. In the final one-eighth segment of the prediction horizon, due to increasing deviation of the LSTM module's error, the fusion weights were adjusted to 0.8 for UKF and 0.2 for LSTM, making the prediction rely more on the stability and disturbance rejection capability of the UKF model. The self-developed experimental platform is shown in Figure 4.

To prevent look-ahead bias and ensure reliable time-series learning, all datasets were partitioned strictly in chronological order. For each battery cell, the first 70% of time-series data was used for training, and the remaining 30% for testing. This approach simulates realistic online estimation scenarios and ensures that the LSTM model does not gain access to future information during training. No overlap or mixing occurred between the training and testing periods.

Furthermore, to evaluate the generalization ability of the model across different battery samples, cross-cell validation was conducted. Specifically, the model was trained on the entire dataset from Cell 1 and tested on Cell 2, and vice versa. This procedure helps assess whether the model captures transferable temporal representations rather than overfitting to the characteristics of a specific cell. Within each cell, data partitioning remained strictly chronological, ensuring full temporal separation between training and testing sets.

Battery testing was conducted using a Neware BTS-5V12A testing system (Neware BTS-5V12A, manufactured by Neware Technology Limited, Shenzhen, China), with a voltage and current accuracy of 0.05%. Two batches of sodium-ion batteries were selected: six 18650E-1300 cells and several NaCR26700-ME30 cells. The 18650E-1300 sodium-ion batteries have a nominal capacity of 1.3 Ah and a nominal voltage of 3.1 V, while the NaCR26700-ME30 sodium-ion batteries have a nominal capacity of 3 Ah and a nominal voltage of 3 V. All experiments were carried out in a temperature-controlled chamber (Sanwood SMG-150-CC) at a constant temperature of 25 °C, with a temperature deviation of less than or equal to ±1 °C.

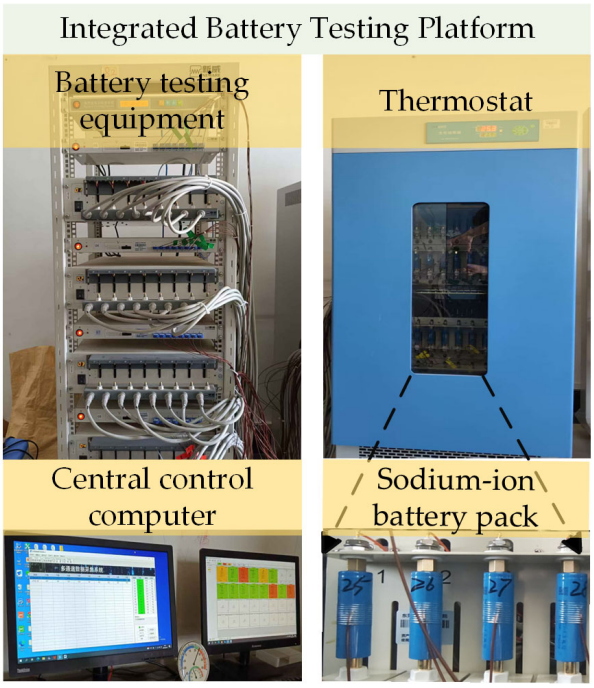


**Figure 4.** Experimental platform.

### *4.2. Comparison Under Different Operating Conditions*

In practical operation, a single working condition cannot accurately reflect prediction accuracy. Therefore, datasets collected under different operating conditions but within the same ambient temperature were used as training data for the model, enabling the training data to better represent the performance of the proposed method.

As shown in Figure 5, the LSTM-UKF fusion model demonstrates excellent SOC prediction performance under various operating conditions. Its core advantage lies in the ability to jointly model dynamic disturbances, nonlinear characteristics, and noisy environments. Notably, under scenarios with significant condition variations, the LSTM-UKF fusion model can quickly adapt its prediction strategy and maintain high prediction accuracy, whereas single models tend to exhibit prediction deviations and fail to capture the real-time variation of battery SOC accurately. These results validate the superior performance of the LSTM-UKF fusion model in handling complex operating conditions, offering a novel approach and methodology for the accurate SOC estimation of SIBs.

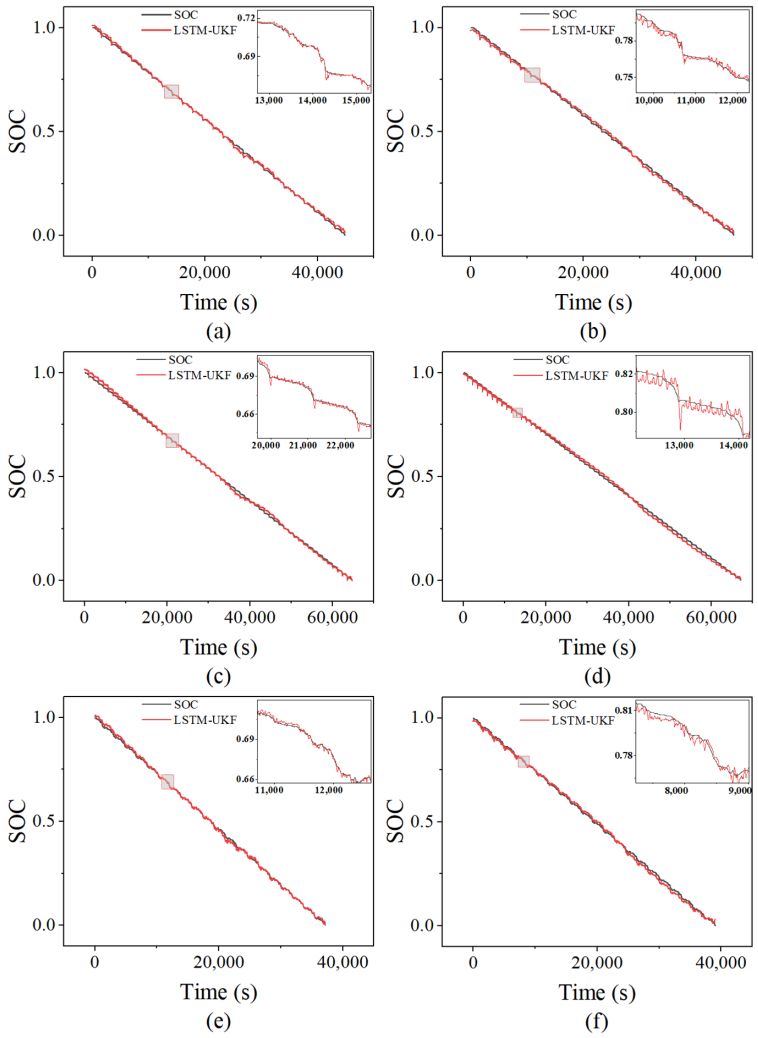


**Figure 5.** Prediction results. (**a**) Cell 1 under CLTC; (**b**) Cell 2 under CLTC; (**c**) Cell 1 under NEDC; (**d**) Cell 2 under NEDC; (**e**) Cell 1 under WLTC; and (**f**) Cell 2 under WLTC.

### *4.3. Comparison Under Different Methods*

To evaluate the effectiveness of the proposed LSTM-UKF fusion model compared to conventional approaches, we conducted a comprehensive comparison using three baseline methods: standalone UKF, standalone LSTM, and the proposed LSTM-UKF fusion. All methods were tested under identical experimental conditions across CLTC, NEDC, and WLTC driving cycles.

The standalone UKF model relies entirely on the equivalent circuit model and is effective in handling sensor noise through recursive estimation. However, it struggles with nonlinearities, parameter drift, and aging effects, particularly under dynamic driving conditions. Conversely, the LSTM model, being data-driven, captures nonlinear temporal dependencies effectively but lacks real-time correction and physical constraints, which may lead to accumulated prediction errors over long horizons.

The proposed hybrid method integrates the temporal modeling capability of LSTM with the state correction ability of UKF. Fusion is guided by a confidence weighting mechanism dynamically adjusted using FFRLS, which allocates greater influence to the model with better short-term accuracy. This synergy results in improved estimation stability, especially under variable load conditions. To evaluate the effectiveness of these models, detailed experiments were conducted, and the results are illustrated in Figure 6.

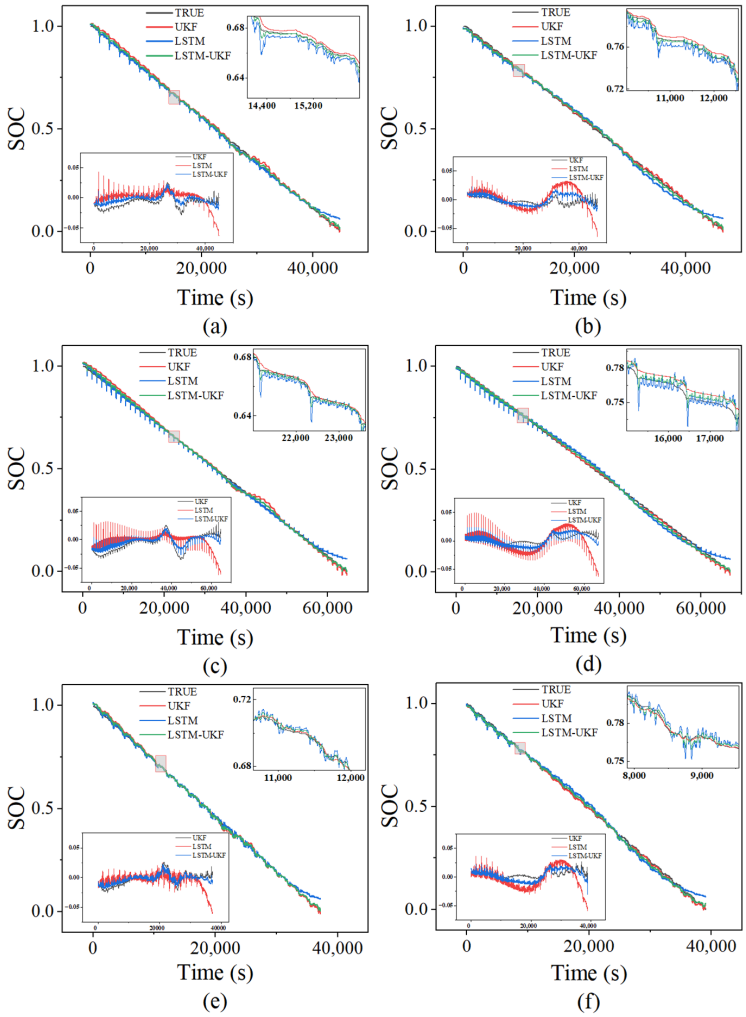


**Figure 6.** Prediction results and errors under different models. (**a**) Cell 1 under CLTC; (**b**) Cell 2 under CLTC; (**c**) Cell 1 under NEDC; (**d**) Cell 2 under NEDC; (**e**) Cell 1 under WLTC; and (**f**) Cell 2 under WLTC.

Tables 1 and 2 present the RMSE and nRMSE values for each method across the three driving conditions. While all methods perform within the 3% RMSE range, the hybrid method consistently shows lower nRMSE values. To further validate the significance of these improvements, Wilcoxon signed-rank tests were conducted between the fusion model and each baseline. The results ($p < 0.05$) confirm that the performance enhancements are statistically significant.

**Table 1.** MAE values of the three models.

| MAE | NEDC1 | WLTC1 | CLTC1 | NEDC2 | WLTC2 | CLTC2 |
|---|---|---|---|---|---|---|
| LSTM | 0.70 | 0.76 | 0.77 | 1.59 | 1.54 | 1.56 |
| UKF | 1.23 | 0.79 | 1.00 | 0.56 | 0.57 | 0.52 |
| LSTM-UKF | 0.68 | 0.55 | 0.57 | 0.91 | 0.89 | 0.80 |

**Table 2.** RMSE values of the three models.

| RMSE | NEDC1 | WLTC1 | CLTC1 | NEDC2 | WLTC2 | CLTC2 |
|---|---|---|---|---|---|---|
| LSTM | 0.70 | 0.76 | 0.77 | 1.59 | 1.54 | 1.56 |
| UKF | 1.23 | 0.79 | 1.00 | 0.56 | 0.57 | 0.52 |
| LSTM-UKF | 0.68 | 0.55 | 0.57 | 0.91 | 0.89 | 0.80 |

As Table 3 The Wilcoxon signed-rank test was applied to compare the proposed LSTM-UKF fusion model with standalone LSTM and UKF across six driving segments. The results show that the fusion model significantly outperforms the standalone LSTM in both MAE and RMSE metrics ($p$ = 0.03125). However, no statistically significant difference was observed compared to the UKF ($p$ = 0.84375), likely due to UKF's strong baseline under certain conditions. These results validate the effectiveness of the hybrid model in reducing learning-related error accumulation, especially over LSTM.

**Table 3.** Wilcoxon test results for LSTM-UKF vs. baseline models.

| Comparison | Metric | W | *p*-Value | Significant ($p < 0.05$) |
|---|---|---|---|---|
| LSTM-UKF vs. LSTM | MAE | 0.0 | 0.03125 | Yes |
| LSTM-UKF vs. UKF | MAE | 9.0 | 0.84375 | No |
| LSTM-UKF vs. LSTM | RMSE | 0.0 | 0.03125 | Yes |
| LSTM-UKF vs. UKF | RMSE | 9.0 | 0.84375 | No |

### *4.4. Comparison Under Noise Disturbance*

To evaluate the robustness and stability of the proposed LSTM-UKF fusion algorithm under complex environments, Gaussian white noise was artificially introduced in the experiments to simulate common non-ideal factors, such as sensor inaccuracies, environmental disturbances, and model uncertainties. Specifically, voltage noise with a standard deviation of σ = 10 mV and current noise with σ = 50 mA were added to the measurement signals. Multiple interference levels with signal-to-noise ratios (SNRs) of 20 dB, 30 dB, and 40 dB were set to cover the disturbance range commonly encountered in industrial scenarios.

As shown in Figure 7 and Table 4, even under strong noise interference, the RMSE of the LSTM-UKF algorithm remained stably controlled within 1.04%, with only a slight increase compared to the noise-free condition, demonstrating significantly better performance than the standalone UKF or LSTM models. This result fully validates the engineering applicability and generalization capability of the algorithm in complex real-world environments.

**Table 4.** Evaluation metrics of the LSTM-UKF model with added noise.

| | NEDC1 | WLTC1 | CLTC1 | NEDC2 | WLTC2 | CLTC2 |
|---|---|---|---|---|---|---|
| MAE | 0.57 | 0.55 | 0.68 | 0.80 | 0.89 | 0.91 |
| RMSE | 0.73 | 0.73 | 0.89 | 0.89 | 1.01 | 1.04 |

Further analysis reveals that the robustness of the fusion algorithm mainly benefits from the complementary advantages of LSTM and UKF in prediction mechanisms: LSTM captures the system's nonlinear dynamic characteristics by learning from historical data, possessing strong short-term memory capability; meanwhile, the UKF utilizes state-space modeling and Kalman gain to dynamically correct observation errors, effectively suppressing random disturbances caused by sensor noise. Especially when noise intensity increases, the filtering advantage of the UKF is fully leveraged to provide real-time correction of biases in LSTM outputs. Simultaneously, the memory units of LSTM help alleviate the UKF's sensitivity to model mismatch to some extent. The synergistic interaction of these two modules within a unified framework achieves effective noise suppression and error propagation control.

In summary, the proposed LSTM-UKF fusion architecture not only provides high-precision SOC estimation under ideal conditions but also maintains stable and reliable

performance in harsh environments, demonstrating strong anti-interference capability and practical deployment potential.

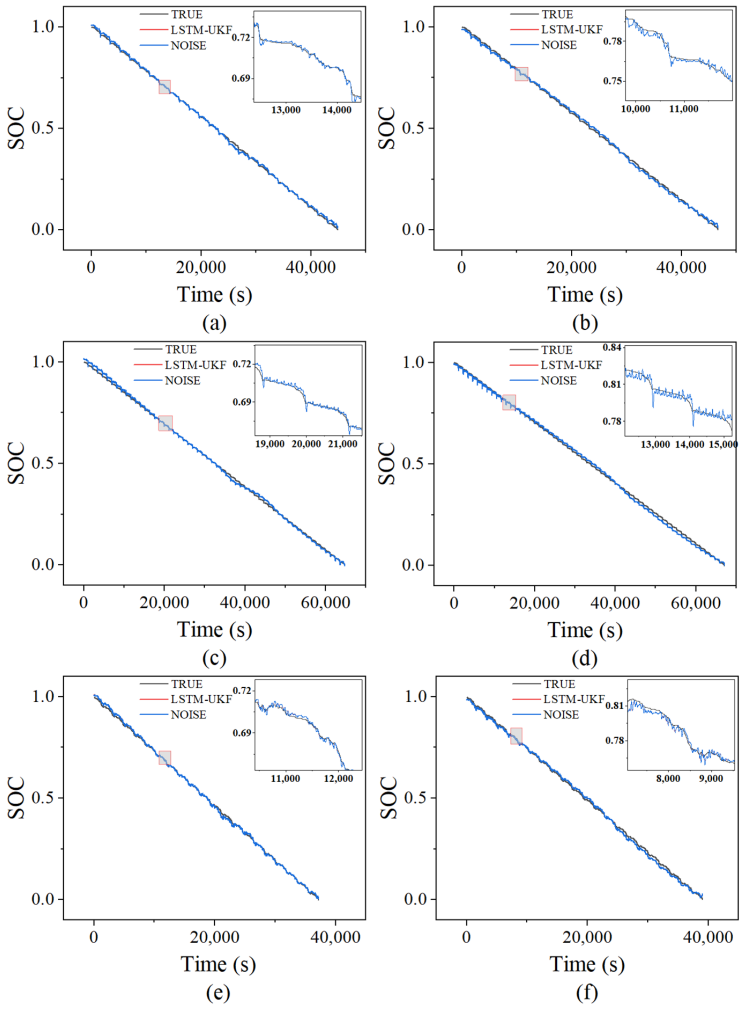


**Figure 7.** Prediction results under added noise. (**a**) Cell 1 under CLTC; (**b**) Cell 2 under CLTC; (**c**) Cell 1 under NEDC; (**d**) Cell 2 under NEDC; (**e**) Cell 1 under WLTC; and (**f**) Cell 2 under WLTC.

## 5. Conclusions

In this paper, a hybrid framework integrating LSTM and a UKF is proposed to improve the SOC estimation accuracy of SIBs. Through comprehensive experimental analysis and comparisons, the superior performance of the proposed method under various operating conditions, different algorithms, and noise disturbances is validated. The framework combines LSTM's capability of learning dynamic temporal features with the UKF's physical constraints based on an ECM. By employing weighted fusion and online parameter identification via the FFRLS algorithm, the proposed approach reduces the mean absolute error (MAE) and root mean square error (RMSE) by approximately 30% on average compared to individual methods under CLTC, NEDC, and WLTC driving cycles. Moreover, the method maintains an RMSE less than or equal to 1.04% even under noise interference, significantly enhancing estimation robustness.

This approach not only improves prediction accuracy and robustness but also further refines SOC estimation precision by leveraging the FFRLS algorithm with a forgetting factor for battery model parameter identification. It provides a novel strategy and method for the accurate SOC estimation of SIBs, holding significant implications for battery management.

Future research can focus on optimizing computational efficiency and extending the framework to other battery systems, thereby offering effective solutions for intelligent BMSs.

**Author Contributions:** X.Z.: methodology, validation, writing—review and editing, resources, and funding acquisition. X.F.: validation, data curation, software, and writing—original draft. X.H.: conceptualization, validation, and writing—review and editing. M.S.: conceptualization and validation. Y.F.: investigation, visualization, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Key Scientific and Technological Project of Henan Province (252102211021, 252102211019, 252102241042).

**Data Availability Statement:** The experiments of this study are still in progress. If you need our data for relevant research, please contact the corresponding author.

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that may appear to influence the work reported in this paper.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| SIBs | Sodium-ion Batteries |
| UKF | Unscented Kalman Filter |
| LSTM | Long Short-Term Memory |
| SOC | State of Charge |
| BMS | Battery Management System |
| OCV | Open Circuit Voltage |
| EKF | Extended Kalman Filter |
| RLS | Recursive Least Squares |
| FFRLS | Forgetting Factor Recursive Least Squares |
| ECM | Equivalent Circuit Model |

## References

1. Chen, J.; Zhang, Y.; Li, W.; Cheng, W.; Zhu, Q. State of charge estimation for lithium-ion batteries using gated recurrent unit recurrent neural network and adaptive Kalman filter. *J. Energy Storage* **2022**, *55*, 105396. [CrossRef]
2. Zhao, H.; Liao, C.; Zhang, C.; Wang, L.; Wang, L. State-of-charge estimation of lithium-ion battery: Joint long short-term memory network and adaptive extended Kalman filter online estimation algorithm. *J. Power Sources* **2024**, *604*, 234451. [CrossRef]
3. Yun, S.T.; Kong, S.H. Data-driven in-orbit current and voltage prediction using Bi-LSTM for LEO satellite lithium-ion battery SOC estimation. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, *58*, 5292–5306. [CrossRef]
4. Lyu, L.; Jiang, B.; Zhu, J.; Wei, X.; Dai, H. An Adaptive Combined Method for Lithium-Ion Battery State of Charge Estimation Using Long Short-Term Memory Network and Unscented Kalman Filter Considering Battery Aging. *Batter. Supercaps* **2024**, *7*, e202400441. [CrossRef]
5. Zhu, T.; Wang, S.; Fan, Y.; Zhou, H.; Zhou, Y.; Fernandez, C. Improved forgetting factor recursive least square and adaptive square root unscented Kalman filtering methods for online model parameter identification and joint estimation of state of charge and state of energy of lithium-ion batteries. *Ionics* **2023**, *29*, 5295–5314. [CrossRef]
6. Zhao, J.; Qian, X.; Jiang, B. Lithium battery state of charge estimation based on improved variable forgetting factor recursive least squares method and adaptive Kalman filter joint algorithm. *J. Energy Storage* **2024**, *100*, 113392. [CrossRef]
7. Prasanthi, A.; Shareef, H.; Abd Khalid, S.; Selvaraj, J. Optimized forgetting factor recursive least square method for equivalent circuit model parameter extraction of battery and ultracapacitor. *J. Energy Storage* **2025**, *119*, 116298. [CrossRef]
8. Pai, H.Y.; Liu, Y.H.; Ye, S.P. Online estimation of lithium-ion battery equivalent circuit model parameters and state of charge using time-domain assisted decoupled recursive least squares technique. *J. Energy Storage* **2023**, *62*, 106901. [CrossRef]
9. Hossain, M.; Haque, M.E.; Arif, M.T. Kalman filtering techniques for the online model parameters and state of charge estimation of the Li-ion batteries: A comparative analysis. *J. Energy Storage* **2022**, *51*, 104174. [CrossRef]
10. Liang, Y.; Wang, S.; Fan, Y.; Takyi-Aninakwa, P.; Xie, Y.; Fernandez, C. An error covariance correction-adaptive extended Kalman filter based on piecewise forgetting factor recursive least squares method for the state-of-charge estimation of lithium-ion batteries. *J. Energy Storage* **2023**, *68*, 107629. [CrossRef]

11. Long, T.; Wang, S.; Cao, W.; Zhou, H.; Fernandez, C. An improved variable forgetting factor recursive least square-double extend Kalman filtering based on global mean particle swarm optimization algorithm for collaborative state of energy and state of health estimation of lithium-ion batteries. *Electrochim. Acta* **2023**, *450*, 142270. [CrossRef]
12. Madani, S.S.; Shabeer, Y.; Allard, F.; Fowler, M.; Ziebert, C.; Wang, Z.; Panchal, S.; Chaoui, H.; Mekhilef, S.; Dou, S.X.; et al. A comprehensive review on lithium-ion battery lifetime prediction and aging mechanism analysis. *Batteries* **2025**, *11*, 127. [CrossRef]
13. Dini, P.; Colicelli, A.; Saponara, S. Review on modeling and soc/soh estimation of batteries for automotive applications. *Batteries* **2024**, *10*, 34. [CrossRef]
14. Houache, M.S.; Yim, C.H.; Karkar, Z.; Abu-Lebdeh, Y. On the current and future outlook of battery chemistries for electric vehicles—Mini review. *Batteries* **2022**, *8*, 70. [CrossRef]
15. Machín, A.; Morant, C.; Márquez, F. Advancements and challenges in solid-state battery technology: An in-depth review of solid electrolytes and anode innovations. *Batteries* **2024**, *10*, 29. [CrossRef]
16. Fan, T.E.; Liu, S.M.; Tang, X.; Qu, B. Simultaneously estimating two battery states by combining a long short-term memory network with an adaptive unscented Kalman filter. *J. Energy Storage* **2022**, *50*, 104553. [CrossRef]
17. Murawwat, S.; Gulzar, M.M.; Alzahrani, A.; Hafeez, G.; Khan, F.A.; Abed, A.M. State of charge estimation and error analysis of lithium-ion batteries for electric vehicles using Kalman filter and deep neural network. *J. Energy Storage* **2023**, *72*, 108039. [CrossRef]
18. Takyi-Aninakwa, P.; Wang, S.; Liu, G.; Bage, A.N.; Bobobee, E.D.; Appiah, E.; Huang, Q. Enhanced extended-input LSTM with an adaptive singular value decomposition UKF for LIB SOC estimation using full-cycle current rate and temperature data. *Appl. Energy* **2024**, *363*, 123056. [CrossRef]
19. Zhang, Y.; Tu, L.; Xue, Z.; Li, S.; Tian, L.; Zheng, X. Weight optimized unscented Kalman filter for degradation trend prediction of lithium-ion battery with error compensation strategy. *Energy* **2022**, *251*, 123890. [CrossRef]
20. Wang, J.; Zuo, Z.; Wei, Y.; Jia, Y.; Chen, B.; Li, Y.; Yang, N. State of charge estimation of lithium-ion battery based on GA-LSTM and improved IAKF. *Appl. Energy* **2024**, *368*, 123508. [CrossRef]
21. Xu, H.; Xu, Q.; Duanmu, F.; Shen, J.; Jin, L.; Gou, B.; Wu, F.; Zhang, W. State of Charge Estimation of Lithium-ion Batteries Based on EKF Integrated with PSO-LSTM for Electric Vehicles. *IEEE Trans. Transp. Electrif.* **2024**, *11*, 2311–2321. [CrossRef]
22. Zhang, Z.; Fan, Y.; Tian, J.; Kuang, H.; Li, M.; Pan, T.; Wang, Y.; Liu, X. Online Estimation of Model Parameters and State of Charge for Lithium-Ion Battery Using Multitimescale Recurrent Neural Networks. *IEEE Trans. Ind. Electron.* **2025**, *72*, 8119–8129. [CrossRef]
23. Xie, Y.; Wang, S.; Zhang, G.; Fan, Y.; Fernandez, C.; Blaabjerg, F. Optimized multi-hidden layer long short-term memory modeling and suboptimal fading extended Kalman filtering strategies for the synthetic state of charge estimation of lithium-ion batteries. *Appl. Energy* **2023**, *336*, 120866. [CrossRef]
24. Chen, J.; Zhang, Y.; Wu, J.; Cheng, W.; Zhu, Q. SOC estimation for lithium-ion battery using the LSTM-RNN with extended input and constrained output. *Energy* **2023**, *262*, 125375. [CrossRef]
25. Zhang, Z.; Chen, S.; Lu, L.; Han, X.; Li, Y.; Chen, S.; Wang, H.; Lian, Y.; Ouyang, M. High-Precision and Robust SOC Estimation of LiFePO4 Blade Batteries Based on the BPNN-EKF Algorithm. *Batteries* **2023**, *9*, 333. [CrossRef]
26. Zhang, Z.; Shao, J.; Li, J.; Wang, Y.; Wang, Z. SOC estimation methods for lithium-ion batteries without current monitoring. *Batteries* **2023**, *9*, 442. [CrossRef]
27. Pattnaik, S.; Kumar, M.R.; Mishra, S.K.; Gautam, S.P.; Appasani, B.; Ustun, T.S. DC bus voltage stabilization and SOC management using optimal tuning of controllers for supercapacitor based PV hybrid energy storage system. *Batteries* **2022**, *8*, 186. [CrossRef]