State of Charge Estimation for Lithium-Ion Batteries at Various Temperatures by Extreme Gradient Boosting and Adaptive Cubature Kalman Filter

Weilu Hou, Qin Shi, Yiwen Liu, Liquan Guo, Xiaonan Zhang, and Ji Wu, *Member, IEEE*

***Abstract*— State of charge (SOC) plays a crucial role in battery management systems (BMSs) as it significantly impacts battery lifespan and energy efficiency. However, accurately estimating SOC is challenging due to the highly nonlinear electrochemical characteristics of batteries. Traditional machine learning algorithms struggle with substantial SOC estimation errors and limited convergence capability, particularly when faced with significant current fluctuations. Recursive algorithms heavily depend on battery models, which can introduce uncertainties during computation, potentially leading to system instability or divergence. Therefore, achieving accurate SOC monitoring has become a significant technical obstacle. To address these challenges, this study presents a novel SOC estimation algorithm named XGBoost-ACKF, which combines the strengths of extreme gradient boosting (XGBoost) and adaptive cubature Kalman filter (ACKF). XGBoost establishes a nonlinear mapping model between input and output characteristics, while ACKF filters and estimates the SOC approximation obtained from XGBoost. This integration yields highly accurate SOC estimation. Experimental results demonstrate that the proposed algorithm surpasses existing approaches in terms of estimation accuracy, generalization ability, and error convergence. Across various temperatures and testing conditions, the algorithm achieves a mean absolute error (MAE) and root-mean-square error (RMSE) of less than 1.06% and 1.25%, respectively. Furthermore, both RMSE and MAE are reduced by over 20% compared to the extended Kalman filter (EKF) and gradient-boosted decision tree (GBDT) methods. These findings establish the proposed algorithm as a promising solution for accurate SOC estimation in BMS applications.**

***Index Terms*— Extreme gradient boosting (XGBoost)-adaptive cubature Kalman filter (ACKF) algorithm, lithium-ion batteries, state of charge (SOC) estimation, various temperatures.**

Manuscript received 28 June 2023; revised 23 November 2023; accepted 2 December 2023. Date of publication 25 December 2023; date of current version 4 January 2024. This work was supported in part by the University Synergy Innovation Program of Anhui Province under Grant GXXT-2020-076, in part by the Innovation Project of New Energy Vehicle and Intelligent Connected Vehicle of Anhui Province, and in part by the Fundamental Research Funds for the Central Universities of China under Grant JZ2023HGTB0287. The Associate Editor coordinating the review process was Dr. Yang Song. *(Corresponding author: Ji Wu.)*

The authors are with the School of Automotive and Transportation Engineering and the Engineering Research Center for Intelligent Transportation and Cooperative Vehicle-Infrastructure of Anhui Province, Hefei University of Technology, Hefei 230009, China (e-mail: houweilu@mail.hfut.edu.cn; shiqin@hfut.edu.cn; 472363578@qq.com; guoliquan@hfut.edu.cn; zhangxiaonan@mail.hfut.edu.cn; wu.ji@hfut.edu.cn).

Digital Object Identifier 10.1109/TIM.2023.3346509

I. Introduction

THE sustainable transport report by the International Energy Agency highlights that the transport sector contributes to 24% of global carbon emissions, with road transport accounting for more than 85% of total transport carbon emissions. With the depletion of fossil fuels and the urgent need to combat global warming, replacing conventional fuel vehicles with electric vehicles has become a necessity [1]. Lithium-ion batteries, which have high energy density, long cycle life, high discharge power, and are environmentally friendly, are widely used in electric vehicles [2], [3]. To ensure battery performance and extend battery life while improving energy efficiency, monitoring the state of charge (SOC) is crucial for battery management systems (BMSs) [4], [5].

However, the batteries' highly nonlinear electrochemical characteristics make it difficult to directly estimate the SOC using traditional methods [6]. Therefore, researchers worldwide have conducted extensive research to develop new methods. One approach for handling nonlinear systems is to treat the system as a black-box model using machine learning algorithms. To deal with the time-varying nature of the problem, the system is continuously iterated, and the time is regularly updated. Currently, there are three commonly used methods for SOC estimation, including the ampere-hour integral method [7], [8], [9], recursive algorithms (e.g., extended Kalman filter (EKF), particle filter, and unscented Kalman filter) [10], [11], [12], and machine learning algorithms (e.g., support vector machines, decision trees, and neural networks) [13], [14].

The ampere-hour integral method is the most widely used method for SOC estimation due to its simple calculation process that does not require knowledge of the internal structure and external circuit characteristics of the battery. Song et al. [15] introduced a high-pass filter to estimate battery parameters by injecting high-frequency and medium-frequency currents for SOC estimation. Zhang et al. [16] proposed a new weighted recursive least squares method to estimate the fast and slow dynamic parameters of the battery separately. However, the computation time of the ampere-hour integral method is much longer than other methods, and it has a large cumulative error because it is sensitive to the noise of the

current sensor and the measured current does not match the actual current [17].

As time has passed, the simple ampere-hour integral method is no longer accurate enough for SOC estimation. Recursive algorithms, which model an equivalent circuit of electronic components to derive state space equations and combine them with an estimation algorithm, can significantly improve the effects of noise disturbances from electric vehicle loads on battery currents and terminal voltages. Jiang et al. [18] and He et al. [19] improved the traditional Kalman filter and obtained the extended Kalman particle filter and the central difference Kalman filter, respectively. These filters have better stability and robustness and greatly improve the estimation accuracy compared to traditional Kalman filters. Zhong et al. [20] constructed a sliding mode observer based on a fractional-order model and utilized Lyapunov's stability theory to analyze its convergence. The observer can overcome parameter uncertainty, modeling error, and measurement error. Recursive algorithms have better error convergence ability because they correct the SOC estimation continuously, but they may accumulate uncertainty during the computation process, resulting in an unstable or divergent system and much longer computation time than other methods [21].

Recently, machine learning algorithms have become the mainstream method for SOC estimation. These algorithms build a black-box model using a large amount of battery experimental data to directly output the estimated SOC. Yang et al. [22] developed a gated recurrent neural network that considers the previous SOC information and measurement information to achieve better estimation accuracy. Ni and Yang [23] proposed a physics-constrained neural network that minimizes both data mapping loss and physical constraint loss simultaneously. This is achieved by enforcing physical constraints between two consecutive time steps to ensure that the estimates adhere to the model equations, effectively reducing the issue of spurious errors. Machine learning algorithms have better efficiency, convenience, and ease of implementation than other methods. However, the estimation results may be biased due to noise, and the machine learning algorithm may be difficult to train because of its long computation time and the need for a large dataset [24].

Recursive algorithms and machine learning algorithms have exhibited significant potential in various aspects of vehicle engineering, particularly in the field of SOC estimation. Researchers have explored the combination of these two algorithmic approaches to develop novel estimation methods. In one study, Tian et al. [25] proposed a hybrid approach by integrating the long–short-term memory (LSTM) network with the cubature Kalman filter (CKF) for SOC estimation in electric vehicles. The results demonstrated notable enhancements in estimation accuracy and convergence capability. However, the literature lacks a comprehensive performance comparison among different algorithms, highlighting the need for further research. In addition, the practical implementation of LSTM poses challenges due to its long training time and complicated parameter settings, which necessitate careful consideration in real-world applications. To address these issues, Tian et al. [26] introduced a hybrid model that combines a deep neural network with Kalman filtering for SOC estimation. This approach not only improves robustness against random noise and error peaks but also incorporates migration learning to enhance adaptive ability in dynamic vehicular environments.

Ambient temperature is another factor that needs to be considered. Currently, existing estimation algorithms cannot directly be applied to SOC estimation under different temperatures, and their estimation accuracy is poor at low temperatures [27], [28]. Xing et al. [29] proposed a temperature-based model and an estimation method that combines open circuit voltage (OCV), SOC, and temperature. The model parameters are adjusted at each sampling step to deal with the uncertainty caused by variations in the operating environment. Liu et al. [30] proposed a temperature compensation model based on a dual-particle filter to overcome the influence of temperature perturbations on model parameters. These methods mentioned above are complex in operation and suffer from issues such as individual differences and battery aging.

Aiming to address the drawbacks of low estimation accuracy and large cumulative error of the ampere-hour integral method, limited convergence capability and estimation accuracy of recursive algorithms, and long training time and complex parameter setting of machine learning algorithms, a lithium-ion battery SOC estimation method using extreme gradient boosting (XGBoost)-adaptive CKF (ACKF) is proposed. First, the XGBoost algorithm is utilized to establish a SOC estimation model for initial SOC estimation. Then, after constructing the state-space equations, the ACKF algorithm is introduced to filter and secondary estimate the approximate SOC output of the XGBoost algorithm, achieving accurate SOC estimation. Finally, the estimation accuracy of the XGBoost-ACKF algorithm is validated by comparing the SOC estimation errors of some common algorithms under various temperature conditions, specifically the Economic Commission for Europe (ECE) and Urban Dynamometer Driving Schedule (UDDS) conditions.

The main contributions of this article are as follows.

1) First, it employs ACKF to construct the state space equations for the battery pack. It uses current as input data and applies the ampere-hour integration method to derive the SOC of the battery pack, which is used as the observed SOC. Then, it harnesses XGBoost to estimate the battery pack's SOC and uses it as the measured SOC. Finally, it utilizes ACKF to mitigate the noise errors in the observed SOC. ACKF possesses the capability to dynamically adjust parameters, making it better suited to handle uncertainties and noise, thereby enhancing the accuracy of battery SOC estimation.
2) The proposed XGBoost-ACKF method demonstrates the robustness and the ability to effectively learn battery states across various temperatures. It provides accurate SOC estimation while accommodating various operating conditions.
3) The effectiveness of the XGBoost-ACKF approach is validated using actual BMS measurement data under ECE and UDDS conditions at temperatures of 0 °C, 25 °C, and 40 °C. The results showcase the

generalization capability of the method under different operating conditions and temperatures. Overall, the XGBoost-ACKF method offers improved training speed, higher prediction accuracy, and robust SOC estimation for lithium-ion batteries across diverse temperatures and operating conditions.

The rest of this article is organized as follows. Section II describes the architecture of the XGBoost-ACKF. Section III focuses on the construction of the XGBoost algorithm, including the training process using the battery dataset. In addition, it introduces the state space equations and presents the integration of the ACKF algorithm into the overall framework. Section IV discusses the experimental results, shedding light on the performance and effectiveness of the XGBoost-ACKF method. Finally, Section V concludes the entire article.

II. Estimation Method

A. Definition of SOC

The SOC is typically defined as the ratio of the current remaining capacity to the nominal capacity, and the detailed expression is as follows:

$$\mathrm{SOC}_t = \frac{C_{\mathrm{remain},t}}{C_{\mathrm{nominal}}} = \mathrm{SOC}_0 - \frac{\eta \int_0^t I_t dt}{Q_n} \tag{1}$$

where SOC_t is the SOC at the moment t, SOC_0 is the initial SOC, I_t is the load current (positive for discharge and negative for a charge), $C_{\mathrm{remain},t}$ is the battery remaining capacity at the moment t, C_{nominal} is the battery nominal capacity, η is the charge-discharge efficiency, and Q_n is the battery maximum available capacity.

B. Estimation Method Architecture

Aiming to address the drawbacks of low estimation accuracy and large cumulative error of the ampere-hour integral method, limited convergence capability and estimation accuracy of recursive algorithms, and long training time and complex parameter setting of machine learning algorithms, a lithium-ion battery SOC estimation method using XGBoost-ACKF is proposed. The architecture of the XGBoost-ACKF is shown in Fig. 1, where four steps are designed to estimate the SOC.

Step 1: Voltage (U), current (I), temperature (T), and SOC are collected from the lithium-ion battery pack and subjected to preprocessing. Then, the XGBoost algorithm is employed to establish a nonlinear mapping model between input characteristics and the output characteristic with the goal of obtaining the SOC_X.

Step 2: By utilizing the current data from the battery and the SOC_X obtained in Step 1, the state space equations are constructed to derive SOC_A and Z_k.

Step 3: The output SOC approximation Z_k obtained in Step 2 is filtered and subjected to secondary estimation using the ACKF algorithm to obtain SOC_E.

Step 4: Combine the SOC_E obtained in Step 3 and the BMS control strategy to control the charging and discharging of the lithium-ion battery pack.

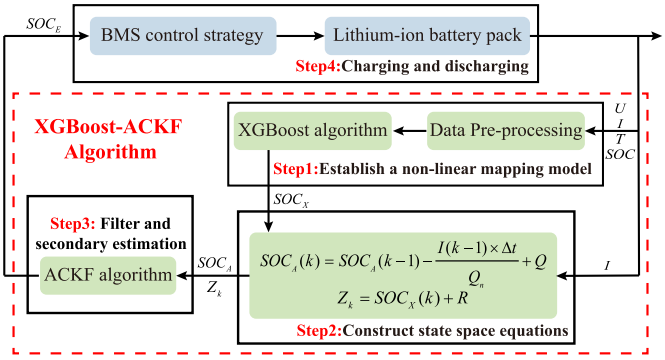

Fig. 1. Architecture of SOC estimation for lithium-ion batteries using XGBoost-ACKF.

III. XGBoost-ACKF

A. XGBoost Algorithm

The XGBoost algorithm is a powerful machine learning algorithm that combines both linear and tree models, making it well-suited for SOC estimation [31]. It adapts to the task by adjusting the complexity of the model using a second-order Taylor expansion of the loss function based on a gradient-boosting decision tree (GBDT) and incorporating a regularization. As a result, the XGBoost algorithm achieves faster training speed and higher estimation accuracy by optimizing the objective function [32].

The XGBoost algorithm approximates the actual SOC by incorporating a new tree model into the prediction results of the previous generation after each iteration, and the detailed expression is as follows:

$$\hat{y}_j^{(m)} = \sum_{l=1}^{m} f_l(x_j) = \hat{y}_j^{(m-1)} + f_m(x_j) \tag{2}$$

where m is the number of iterations; j is the number of samples; $f_l(\cdot) \in Z$, $f_l(\cdot)$ is a subfunction of L, L is the set of all trees with the following equation:

$$L = \{f_l(x) = w_{q(x)}\}(w \in R^M, q : R^d \to M) \tag{3}$$

where $w_{q(x)}$ is the weight vector of the leaf nodes, q is the mapping relation for the leaf nodes of the sample, M is the total number of leaf nodes, d is the number of characteristics.

The objective function of the XGBoost algorithm comprises both a loss function and a regularization function, and the detailed expression is as follows:

$$F_{\mathrm{obj}} = \sum_{j=1}^{J} l(y_j, \hat{y}_j) + \sum_{j=1}^{m} \Omega(f_j) \tag{4}$$

where J is the total of samples, $l(\cdot)$ is the loss function, y_j is the actual SOC, $\hat{y}_j$ is the estimated SOC, $\Omega(\cdot)$ is the regularization function, which is used to suppress model complexity and prevent over-fitting with the following equation:

$$\Omega(f_m) = \gamma M + \frac{1}{2}\lambda \sum_{j=1}^{M} w_j^2 \tag{5}$$

where γ and λ penalty coefficients for the regularization function, respectively. The approximate objective function is obtained by neglecting the known values and constant terms after performing a second-order Taylor expansion of the loss function to at $f_m(x_j) = 0$ and aggregating the tree leaf nodes, and the detailed expression is as follows:

$$F_{\text{obj}}^{(m)} = \sum_{j=1}^{M}\left[\left(\sum_{j\in\psi_j} g_j\right)w_j + \frac{1}{2}\left(\sum_{j\in\psi_j} h_j + \lambda\right)w_j^2\right] + \gamma M \tag{6}$$

where $\psi_j = \{l|q(x_l) = j\}$ is the set of samples falling into the leaf node j; g_j and h_j are the first-order partial derivative and second-order partial derivative of the loss function with respect to $\hat{y}_j{}^{(m-1)}$, respectively. The objective subfunctions of the leaf node are independent. Therefore, the objective function is minimized when the subfunctions of each leaf node are minimized, and the weight of the leaf node j with the following equation:

$$w_j^* = -\frac{\sum_{j\in\psi_j} g_j}{\sum_{j\in\psi_j} h_j + \lambda}. \tag{7}$$

The final objective function of the XGBoost algorithm is as follows:

$$F_{\text{obj}}^* = -\frac{1}{2}\sum_{j=1}^{M}\frac{\left(\sum_{j\in\psi_j} g_j\right)^2}{\sum_{j\in\psi_j} h_j + \lambda} + \gamma M. \tag{8}$$

The XGBoost algorithm initiates with a tree of depth 0 and systematically explores all potential splitting schemes for the input characteristics. The optimal splitting position is determined by calculating the improvement in the objective function before and after the split. Following the training of one tree, the next round of training is conducted, and the optimal model is obtained through iterative refinement.

In this article, weak learners are selected for integrated learning within the regression tree model. To ensure the model's generalization ability, the training data is collected from dynamic stress testing (DST), while the test data is gathered from ECE and UDDS.

During the training process, data preprocessing is carried out to ensure estimation accuracy. In this article, the data was measured accurately through the experimental test bench, data preprocessing includes duplicate value removal, outlier screening, and missing value imputation. First, duplicate data in the dataset is eliminated. Second, discrete values in the dataset are identified and designated as missing values, and then filled with mean value imputation. Finally, data normalization is applied to the optimized dataset.

In addition, the grid search method is employed to fine-tune the parameters of the XGBoost algorithm, aiming to obtain the optimal model and reduce errors [33]. The steps for using the grid search method to fine-tune penalty coefficients in the XGBoost algorithm are as follows. First, define the potential range of penalty coefficient values. Second, create an XGBoost regression model and conduct cross-validation with the penalty coefficients. Employ a grid search library to carry out grid search and record the performance metrics for each combination. Finally, identify the parameter combination with the best performance metrics and apply it to the XGBoost model.

Furthermore, an early stop mechanism is implemented to mitigate overfitting risks and decrease training time. The early stopping mechanism refers to utilizing the loss function as a metric to monitor model performance. After each iteration, an evaluation is conducted. If this metric fails to show improvement within a defined number of consecutive monitoring windows, it is inferred that the model training has reached an optimal stage, prompting the early termination of the training process.

B. ACKF Algorithm

The SOC estimated by the XGBoost algorithm contains some level of noise. To mitigate these noise fluctuations, the SOC is filtered and subjected to secondary estimation using the ACKF algorithm. Before applying the ACKF algorithm for filtering and secondary estimation, the state space equations are established with the following equation:

$$\begin{cases} \text{SOC}_A(k) = \text{SOC}_A(k-1) - \dfrac{I(k-1)\times\Delta t}{Q_n} + Q_k \\ Z_k = \text{SOC}_X(k) + R_k \end{cases} \tag{9}$$

where $\text{SOC}_A(k)$ is the SOC estimation by ampere-hour integral method at the moment k, Δt is the sampling time, $\text{SOC}_X(k)$ is the SOC estimation by XGBoost algorithm at the moment k, Z_k is the SOC measurement at the moment k, Q_k is the process noise covariance at the moment k, and R_k is the measurement noise covariance at the moment k.

The CKF algorithm enhances the traditional Kalman filter by introducing a third-order spherical criterion and utilizing a set of cubature points to approximate the state mean and covariance of a nonlinear system [34]. Compared to other filtering algorithms, CKF offers higher accuracy, and stability, and is independent of the system dimension. It serves as a powerful tool for solving nonlinear state estimation problems. In this article, considering that the process noise covariance and measurement noise covariance may vary and cannot be assumed as stable constants throughout the experiment, which could result in significant estimation errors, the ACKF algorithm can solve this problem effectively [35]. The ACKF algorithm aims to improve noise estimation accuracy by employing an adaptive method based on the CKF algorithm. The ACKF algorithm comprises the following four main processes.

Step 1: Parameter initialization, define the initial state variable $\hat{x}_0$ and the initial error covariance P_0.

Step 2: Time update.

Step 2.1: Calculate the cubature points

$$S_k = \text{chol}(P_k) \tag{10}$$

$$x_{i,k} = S_k\xi_i + \hat{x}_k, \quad i = 1, 2, \ldots, 2n \tag{11}$$

where n is the number of state variables, $\text{chol}(\cdot)$ is the Cholesky decomposition and returns a lower triangular Cholesky factor, and ξ_i is the ith set of cubature points.

Step 2.2: Propagate the cubature points and calculate the estimated state variable

$$x_{i,k+1|k} = f\left(x_{i,k}, u_k\right) \tag{12}$$

$$\bar{x}_{k+1} = \frac{1}{2n}\sum_{i=1}^{2n} x_{i,k+1|\mathrm{k}} \tag{13}$$

where $f(\cdot)$ is a state function in a nonlinear system.

Step 2.3: Calculate the estimated error covariance

$$P_{k+1|k} = \frac{1}{2n}\sum_{i=1}^{2n} x_{i,k+1|\mathrm{k}}\left(x_{i,k+1|\mathrm{k}}\right)^T - \bar{x}_{k+1}\left(\bar{x}_{k+1}\right)^T + Q_k. \tag{14}$$

Step 3: Measurement update.

Step 3.1: Calculate the cubature points

$$S_{k+1|k} = \mathrm{chol}\left(P_{k+1|k}\right) \tag{15}$$

$$x_{i,k+1} = S_{k+1|k}\xi_i + \bar{x}_{k+1}, \quad i = 1, 2, \ldots, 2n. \tag{16}$$

Step 3.2: Propagate the cubature points and calculate the estimated state variable

$$Z_{i,k+1} = g\left(x_{i,k+1}, u_{k+1}\right) \tag{17}$$

$$\hat{Z}_{k+1} = \frac{1}{2n}\sum_{i=1}^{2n} Z_{i,k+1}. \tag{18}$$

Step 3.3: Calculate the estimated error covariance and estimated reciprocal covariance

$$P^z_{k+1} = \frac{1}{2n}\sum_{i=1}^{2n} Z_{i,k+1}\left(Z_{i,k+1}\right)^T - \hat{Z}_{k+1}\left(\hat{Z}_{k+1}\right)^T + R_k \tag{19}$$

$$P^{xz}_{k+1} = \frac{1}{2n}\sum_{i=1}^{2n} x_{i,k+1}\left(Z_{i,k+1}\right)^T - \bar{x}_{k+1}\left(\hat{Z}_{k+1}\right)^T. \tag{20}$$

Step 3.4: Calculate the Kalman gain

$$K_{k+1} = P^{xz}_{k+1}/P^z_{k+1}. \tag{21}$$

Step 3.5: Update the state variable and error covariance

$$\hat{x}_{k+1} = \bar{x}_{k+1} + K_{k+1}\left(Z_{k+1} - \hat{Z}_{k+1}\right) \tag{22}$$

$$P_{k+1} = P_{k+1|k} - K_{k+1}P^z_{k+1}{K_{k+1}}^T. \tag{23}$$

Step 4: Adaptive update.

Step 4.1: Calculate SOC residual covariance

$$F_k = \sum_{i=k-L_w+1}^{k} e_i {e_i}^T \tag{24}$$

where e_i is the SOC residual covariance and L_w is the window size for covariance matching.

Step 4.2: Update process noise covariance and measurement noise covariance

$$Q_k = K_k F_k {K_k}^T \tag{25}$$

$$R_k = F_k + \frac{1}{2n}\sum_{j=1}^{2n}\left(Z_{i,k} - Z_k\right)\left(Z_{i,k} - Z_k\right)^T. \tag{26}$$

Step 4.3: There are three kinds of covariance matrices in the proposed method, including initial error covariance P_0, process noise covariance Q_0, and measurement noise covariance R_0. The initial values of the covariance matrices will affect the SOC estimation accuracy. It is worth mentioning that the initial values of P_0, Q_0, and R_0 in the filter are set to the optimal value by several cycles before the experiment [36]. In this article, $P_0 = \mathrm{diag}(10^{-1}, 10^{-6}, 10^{-6})$, $Q_0 = \mathrm{diag}(10^{-6}, 10^{-6}, 10^{-6})$, and $R_0 = 10^{-1}$.

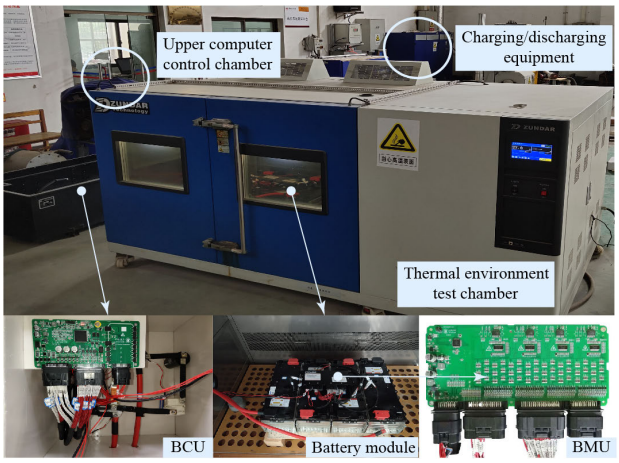

Fig. 2. Lithium-ion battery test bench.

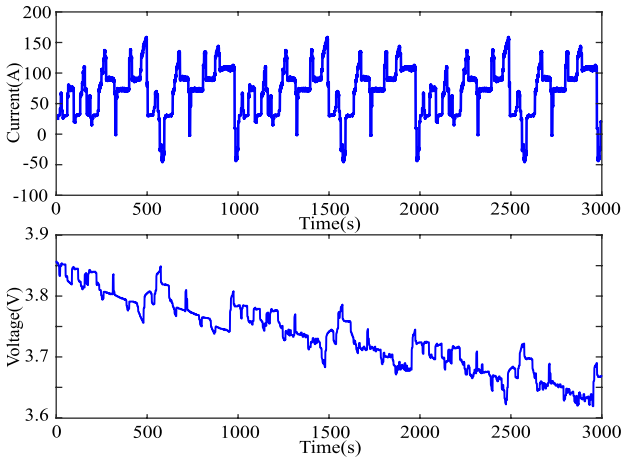

Fig. 3. Current and voltage variation curves under ECE condition.

IV. Experimental Results and Analysis

In this article, driving conditions are simulated using a battery test bench, and data such as voltage, current, and temperature are collected for training the XGBoost-ACKF algorithm. The battery test bench setup includes charging/discharging equipment, a thermal environment test chamber, an upper computer control chamber, a battery module, and a BMS comprising a battery control unit (BCU) and a battery management unit (BMU), as shown in Fig. 2.

ECE and UDDS provide current and voltage data for SOC estimation, but they exhibit differences due to variations in driving conditions and driving modes, as shown in Figs. 3 and 4. Specifically, ECE replicates a mix of urban and highway driving scenarios, encompassing diverse speed profiles and various road conditions, which may result in significant fluctuations in current and voltage between urban

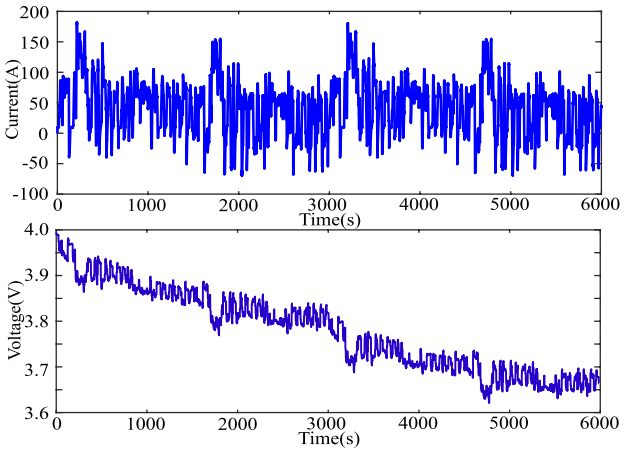

Fig. 4. Current and voltage variation curves under UDDS condition.

and highway segments. On the contrary, UDDS predominantly emulates urban driving conditions, marked by a significant amount of low-speed travel, frequent stops, and rapid accelerations, resulting in frequent and abrupt changes in current over short time intervals.

Considering the significant influence of temperature on the maximum available battery capacity, it is important to note that using the nominal battery capacity instead of the maximum available battery capacity can result in substantial errors in SOC estimation. Therefore, based on the battery test bench, the maximum available battery capacities at 0 °C, 25 °C, and 40 °C were determined as 83.9, 92.7, and 96.5 Ah, respectively. To evaluate the generalization ability of the proposed algorithm under different operating and temperature conditions, the algorithm was tested using data from the ECE and UDDS operating conditions at temperatures of 0 °C, 25 °C, and 40 °C. This validation process aimed to assess the algorithm's performance across various scenarios and temperature variations.

A. Experiments Under ECE Condition

ECE is a standardized test method for measuring vehicle energy consumption, widely employed in electric vehicle research within Europe. To assess the robustness of the algorithms, the initial SOC value of 0.5 is set for the experiments. The SOC estimation results, corresponding errors, and the error probability distribution plots of various algorithms at three different temperatures are shown in Figs. 5–7.

Overall, all three hybrid algorithms demonstrate the ability to estimate the decreasing trend of SOC. However, there is a noticeable difference in estimation accuracy among the algorithms. As shown in Figs. 5–7, the XGBoost-ACKF algorithm exhibits superior estimation accuracy, with SOC estimation results closer to the reference values. Analyzing the error probability distribution plots, it becomes evident that the mean error of XGBoost-ACKF consistently remains closer to 0 than the other two algorithms (GBDT-ACKF and XGBoost-EKF). Specifically, also under the ECE condition, the GBDT-ACKF algorithm and the XGBoost-EKF algorithm

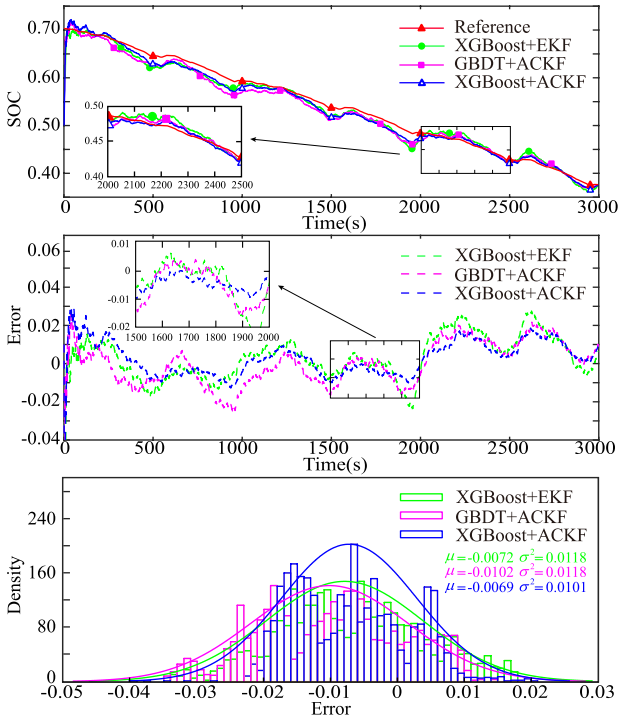

Fig. 5. SOC estimation results under ECE condition at 0 °C.

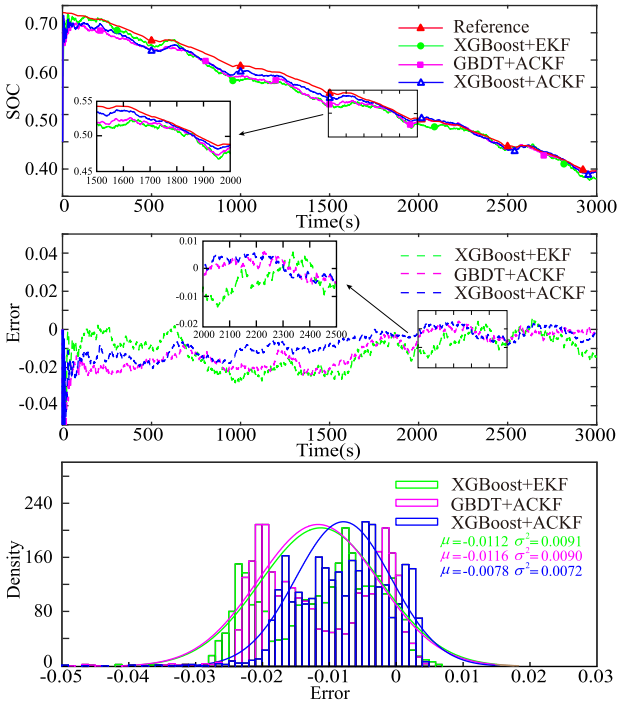

Fig. 6. SOC estimation results under ECE condition at 25 °C.

can reflect the trend of the SOC changes, but they have a wider error distribution than the XGBoost-ACKF algorithm. As shown in Figs. 5–7, the error probability distribution of XGBoost-ACKF algorithm consistently exhibits mean and variance values closest to 0 compare to other two algorithms (GBDT-ACKF and XGBoost-EKF) across three different temperatures. The mean error probability distribution of XGBoost-ACKF remains below 0.0078, with a variance not

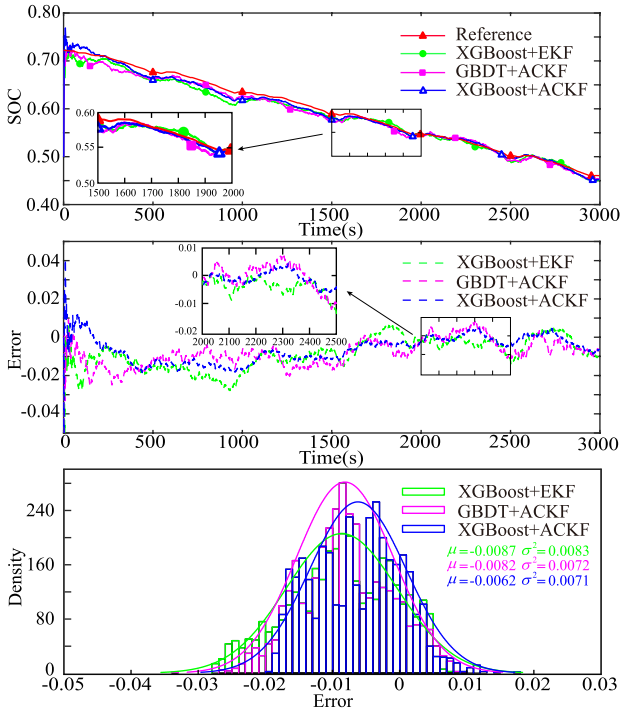

Fig. 7. SOC estimation results under ECE condition at 40 °C.

TABLE I

MAE OF VARIOUS ALGORITHMS UNDER ECE CONDITION AT VARIOUS TEMPERATURES

Algorithm	Temperature (°C)		
	0	25	40
XGBoost	0.0237	0.0163	0.0161
XGBoost-EKF	0.0117	0.0124	0.0105
GBDT-ACKF	0.0124	0.0118	0.0093
XGBoost-ACKF	0.0106	0.0094	0.0085

exceeding 0.0101. Compared to XGBoost-EKF, it reduces by 30.36% and 33.90%, respectively, and compared to GBDT-ACKF, it decreases by 32.76% and 33.90%, respectively. Moreover, as the temperature gradually rises, the error distribution becomes more concentrated. This implies that at higher temperatures, generating significant deviations becomes increasingly challenging, leading to more dependable results. Moreover, the SOC estimation error is smoother and more stable compared to the other two algorithms (GBDT-ACKF and XGBoost-EKF). The superior performance of XGBoost-ACKF can be attributed to two main factors. First, compared to GBDT, the XGBoost algorithm expands the Taylor objective function to the second order and incorporates a regularization function for the leaf weights. This allows the model to capture more information about the objective function and achieve lower variance. Second, the adaptive update of the process noise covariance and measurement noise covariance in the ACKF algorithm significantly reduces SOC estimation errors caused by the uncertainty and complexity of the noise.

Tables I and II present error comparisons of the three hybrid algorithms (GBDT-ACKF, XGBoost-EKF, and XGBoost-ACKF) under the ECE condition. It is evident that the combination of recursive algorithms and machine

TABLE II

RMSE OF VARIOUS ALGORITHMS UNDER ECE CONDITION AT VARIOUS TEMPERATURES

Algorithm	Temperature (°C)		
	0	25	40
XGBoost	0.0249	0.0178	0.0173
XGBoost-EKF	0.0147	0.0154	0.0130
GBDT-ACKF	0.0156	0.0147	0.0119
XGBoost-ACKF	0.0125	0.0119	0.0105

learning algorithms effectively reduces the mean absolute error (MAE) and root-mean-square error (RMSE), with reductions exceeding 100% in the ECE condition. Comparing XGBoost-ACKF and XGBoost, both algorithms demonstrate significantly reduced MAE and RMSE, controlled within 1.60%. These improvements can be attributed to the ACKF and EKF algorithms, which reduce SOC peak fluctuations during drastic changes in current and voltage, leading to a substantial reduction in SOC estimation errors. Moreover, compared to XGBoost-EKF, XGBoost-ACKF achieves a 24.19% reduction in MAE and a 24.67% reduction in RMSE. Compared to GBDT-ACKF, XGBoost-ACKF achieves a 20.34% reduction in MAE and a 21.09% reduction in RMSE. These results further validate the necessity and effectiveness of the adaptive update of the process noise covariance and measurement noise covariance.

B. Experiments Under UDDS Condition

UDDS is more complicated than ECE and is commonly used to verify various system and vehicle algorithm verification conditions. Similar to the ECE test, the initial SOC value is set to 0.5 to evaluate the robustness of the algorithms. The SOC estimation results, corresponding errors, and the error probability distribution plots of various algorithms at three different temperatures are shown in Figs. 8–10.

As shown in Figs. 8–10, all three hybrid algorithms are capable of converging to the final reference value. As shown in Figs. 8–10, our findings reveal that while GBDT-ACKF algorithm boasts a mean error closer to 0 compared to XGBoost-ACKF algorithm, GBDT-ACKF algorithm consistently remains the most scattered error distribution. The error probability distribution of the XGBoost-ACKF algorithm consistently exhibits variance values closest to 0 compared to the other two algorithms (GBDT-ACKF and XGBoost-EKF) across three different temperatures. On the other hand, the XGBoost-ACKF algorithm demonstrates a more concentrated error distribution while also maintaining an average error closer to 0 when compared to the XGBoost-EKF algorithm. However, in comparison to the other two algorithms, XGBoost-ACKF demonstrates superior estimation accuracy and adaptability across different temperatures. Moreover, XGBoost-ACKF exhibits smoother SOC estimation errors with lower peaks and faster convergence time. The improved performance of XGBoost-ACKF can be attributed to the adaptive update of the process noise covariance and measurement noise covariance in the ACKF algorithm. Unlike the EKF algorithm, which may have a mismatch between the initial noise covariance

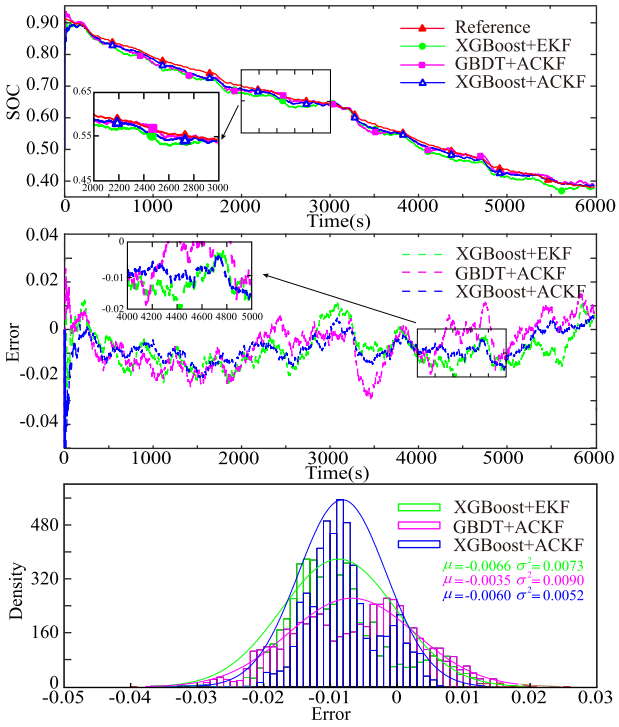

Fig. 8. SOC estimation results under UDDS condition at 0 °C.

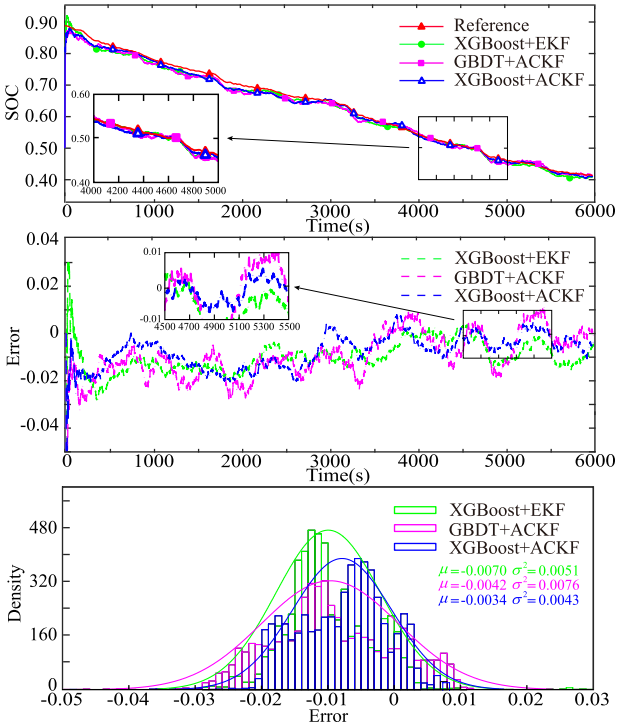

Fig. 9. SOC estimation results under UDDS condition at 25 °C.

TABLE III
MAE OF VARIOUS ALGORITHMS UNDER UDDS CONDITION AT VARIOUS TEMPERATURES

Algorithm	Temperature (°C)		
	0	25	40
XGBoost	0.0167	0.0171	0.0148
XGBoost-EKF	0.0113	0.0115	0.0102
GBDT-ACKF	0.0100	0.0123	0.0094
XGBoost-ACKF	0.0096	0.0093	0.0082

TABLE IV
RMSE OF VARIOUS ALGORITHMS UNDER UDDS CONDITION AT VARIOUS TEMPERATURES

Algorithm	Temperature (°C)		
	0	25	40
XGBoost	0.0183	0.0183	0.0151
XGBoost-EKF	0.0132	0.0137	0.0127
GBDT-ACKF	0.0124	0.0152	0.0122
XGBoost-ACKF	0.0114	0.0119	0.0113

and the actual noise, ACKF corrects this mismatch by dynamically updating the noise covariates based on the SOC residual covariance. This adaptive update mechanism enables ACKF to quickly adjust and reduce estimation errors caused by noise discrepancies. Overall, XGBoost-ACKF proves to be a more accurate and robust algorithm, providing better estimation results and faster convergence, especially when faced with the challenges presented by the UDDS test and varying temperature conditions.

The error comparisons of the three hybrid algorithms under UDDS condition are shown in Tables III and IV. The results demonstrate notable improvements in the MAE and RMSE achieved by XGBoost-ACKF. Specifically, both the MAE and RMSE of XGBoost-ACKF are significantly reduced and controlled to be within 1.20%. When compared to XGBoost-EKF, XGBoost-ACKF shows a reduction of 19.13% in MAE and 13.64% in RMSE. Similarly, when compared to GBDT-ACKF, XGBoost-ACKF exhibits a reduction of 24.39% in MAE and 21.71% in RMSE.

C. Experiments Under Various Temperature

From Figs. 5–10, it is evident that all three hybrid algorithms are capable of estimating the decreasing trend of SOC under different temperatures. However, the SOC estimation results at low temperatures (0 °C) exhibit noticeable fluctuations compared to those at room temperature (25 °C and 40 °C) under both ECE and UDDS conditions. Similarly, when examining Tables I–IV, it is observed that the SOC estimation results and errors at room temperature are superior to those at low temperature for both ECE and UDDS conditions. Specifically, the MAE and RMSE of the three hybrid algorithms remain below 1.71% and 1.83%, respectively, under ECE and UDDS conditions at room temperature. On the other hand, under low-temperature conditions, the MAE and RMSE of the three hybrid algorithms increase to 2.37% and 2.49%, respectively, for both ECE and UDDS conditions. This experimental outcome indicates that low temperature has a more significant impact on the internal characteristics of batteries, resulting in decreased SOC estimation accuracy compared to room temperature conditions. At low temperatures, the battery's internal reactions become less active, leading to reduced migration efficiency of lithium ions and decay in the maximum available capacity. Conversely, at room temperature, the battery's internal resistance decreases, improving the discharge capacity and increasing the maximum available capacity.

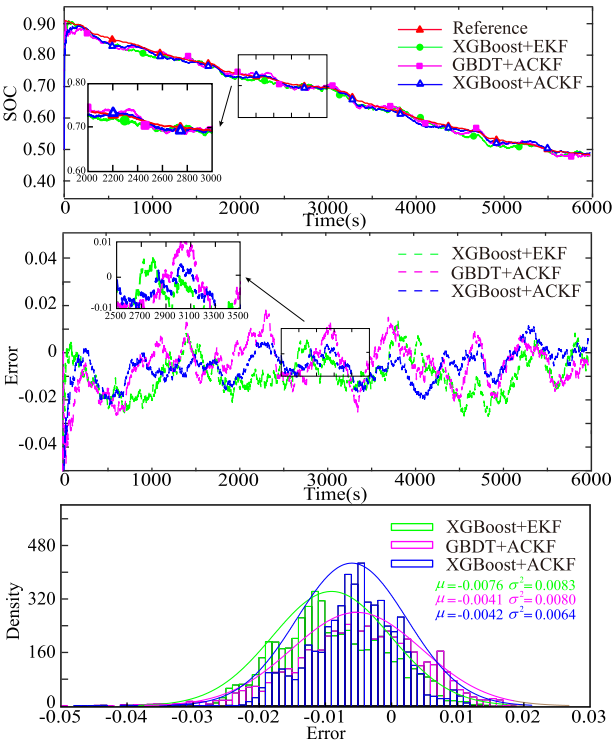

Fig. 10. SOC estimation results under UDDS condition at 40 °C.

Therefore, it is recommended to preheat the battery before charging and discharging at low temperatures to enhance estimation performance. However, even at low temperatures, XGBoost-ACKF exhibits high estimation accuracy. Compared to XGBoost, XGBoost-ACKF achieves a remarkable reduction in MAE from 2.37% to 1.06% and RMSE from 2.49% to 1.25%. These results demonstrate that XGBoost-ACKF effectively learns the battery's state under various temperature conditions and provides satisfactory estimation accuracy with robustness.

D. Runtime Comparison

These programs are run on a laptop equipped with an Intel Core i7-8565U CPU 1.8 GHz and 7.85 GB RAM. The XGBoost-ACKF algorithm model is trained using the data extracted from the ECE and UDDS conditions, and subsequently, the execution is repeated ten times under identical conditions. A comparison of execution times in ten times under ECE and UDDS conditions is shown in Table V. From Table V, it is evident that XGBoost-ACKF significantly reduces the runtime compared to GBDT-ACKF. This improvement can be attributed to the characteristics of the XGBoost algorithm, which incorporates a regularization term into the GBDT. By adjusting the complexity of the model and implementing an early stop mechanism during training, XGBoost achieves faster training speed and higher prediction accuracy. While the runtime of XGBoost-ACKF may be slightly increased compared to XGBoost and XGBoost-EKF, the benefits in terms of estimation accuracy and robustness are effectively guaranteed. The modest increase in runtime is a worthwhile tradeoff considering the improved performance of XGBoost-ACKF.

TABLE V
RUNTIMES OF VARIOUS ALGORITHMS UNDER ECE AND UDDS CONDITIONS

Algorithm	Runtimes(s)	
	ECE	UDDS
XGBoost	0.524	0.798
XGBoost-EKF	0.948	1.203
GBDT-ACKF	1.673	1.967
XGBoost-ACKF	1.117	1.354

V. CONCLUSION

In this article, a SOC estimation method based on the XGBoost-ACKF algorithm is proposed for SOC estimation at various temperatures. First, the XGBoost is used to establish a nonlinear mapping model between the input characteristics (voltage, current, temperature, and SOC at the previous moment) and the output characteristic (SOC at the present moment). This model captures the complex relationships between the input variables and SOC. Then, the ACKF is used to filter and secondary estimate the output SOC approximation obtained from XGBoost. This step helps to eliminate noise fluctuations and improve the accuracy of SOC estimation. Through these steps, the accurate estimated SOC at various temperatures is obtained.

To validate the effectiveness of the XGBoost-ACKF algorithm, we conduct experiments using BMS-measured data under ECE and UDDS conditions at temperatures of 0 °C, 25 °C, and 40 °C. The experimental results demonstrate the superiority of the proposed algorithm. Compared with XGBoost-EKF and GBDT-ACKF, the XGBoost-ACKF algorithm achieves a significant reduction in MAE, with reductions of 24.19% and 24.39%, respectively. The RMSE of XGBoost-ACKF is reduced by 24.67% and 21.71% compared to XGBoost-EKF and GBDT-ACKF, respectively. Moreover, both MAE and RMSE of XGBoost-ACKF remain below 1.06% and 1.25% at different temperatures for both ECE and UDDS conditions. Overall, the proposed algorithm demonstrates higher estimation accuracy, stronger generalization ability, and effective error convergence.

REFERENCES

[1] K. Moustakas, M. Loizidou, M. Rehan, and A. S. Nizami, "A review of recent developments in renewable and sustainable energy systems: Key challenges and future perspective," *Renew. Sustain. Energy Rev.*, vol. 119, Mar. 2020, Art. no. 109418.

[2] K. Liu, K. Li, Q. Peng, and C. Zhang, "A brief review on key technologies in the battery management system of electric vehicles," *Frontiers Mech. Eng.*, vol. 14, no. 1, pp. 47–64, Mar. 2019.

[3] Y. Wang and Z. Chen, "A framework for state-of-charge and remaining discharge time prediction using unscented particle filter," *Appl. Energy*, vol. 260, Feb. 2020, Art. no. 114324.

[4] M. Wu, L. Qin, G. Wu, Y. Huang, and C. Shi, "State of charge estimation of power lithium-ion battery based on a variable forgetting factor adaptive Kalman filter," *J. Energy Storage*, vol. 41, Sep. 2021, Art. no. 102841.

[5] K. Yu, H. Wang, L. Mao, Q. He, and Q. Wu, "IC curve-based lithium-ion battery SOC estimation at high rate charging current," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.

[6] L. Ling, D. Sun, X. Yu, and R. Huang, "State of charge estimation of lithium-ion batteries based on the probabilistic fusion of two kinds of cubature Kalman filters," *J. Energy Storage*, vol. 43, Nov. 2021, Art. no. 103070.

[7] R. Xiong, L. Li, Q. Yu, Q. Jin, and R. Yang, "A set membership theory based parameter and state of charge co-estimation method for all-climate batteries," *J. Cleaner Prod.*, vol. 249, Mar. 2020, Art. no. 119380.

[8] S. M. Qaisar, "Event-driven approach for an efficient Coulomb counting based li-ion battery state of charge estimation," *Proc. Comput. Sci.*, vol. 168, pp. 202–209, Jan. 2020.

[9] S. Wang, Y. Yang, and K. Guo, "An improved recursive total least squares estimation of capacity for electric vehicle lithium-iron phosphate batteries," *Math. Problems Eng.*, vol. 2020, pp. 1–12, Jun. 2020.

[10] X. Qiu, W. Wu, and S. Wang, "Remaining useful life prediction of lithium-ion battery based on improved cuckoo search particle filter and a novel state of charge estimation method," *J. Power Sources*, vol. 450, Feb. 2020, Art. no. 227700.

[11] S. Wang et al., "A novel safety assurance method based on the compound equivalent modeling and iterate reduce particle-adaptive Kalman filtering for the unmanned aerial vehicle lithium ion batteries," *Energy Sci. Eng.*, vol. 8, no. 5, pp. 1484–1500, May 2020.

[12] Q. Shi, Z. Jiang, Z. Wang, X. Shao, and L. He, "State of charge estimation by joint approach with model-based and data-driven algorithm for lithium-ion battery," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.

[13] J. Chen, Y. Zhang, J. Wu, W. Cheng, and Q. Zhu, "SOC estimation for lithium-ion battery using the LSTM-RNN with extended input and constrained output," *Energy*, vol. 262, Jan. 2023, Art. no. 125375.

[14] X. Tang, Y. Zhou, F. Gao, and X. Lai, "Joint estimation of state-of-charge and state-of-health for all cells in the battery pack using 'leader-follower' strategy," *eTransportation*, vol. 15, Jan. 2023, Art. no. 100213.

[15] Z. Song et al., "The sequential algorithm for combined state of charge and state of health estimation of lithium-ion battery based on active current injection," *Energy*, vol. 193, Feb. 2020, Art. no. 116732.

[16] C. Zhang, W. Allafi, Q. Dinh, P. Ascencio, and J. Marco, "Online estimation of battery equivalent circuit model parameters and state of charge using decoupled least squares technique," *Energy*, vol. 142, pp. 678–688, Jan. 2018.

[17] W. Cao, S. Wang, C. Fernandez, C. Zou, C. Yu, and X. Li, "A novel adaptive state of charge estimation method of full life cycling lithium-ion batteries based on the multiple parameter optimization," *Energy Sci. Eng.*, vol. 7, no. 5, pp. 1544–1556, Oct. 2019.

[18] J. Zhengxin, S. Qin, W. Yujiang, W. Hanlin, G. Bingzhao, and H. Lin, "An immune genetic extended Kalman particle filter approach on state of charge estimation for lithium-ion battery," *Energy*, vol. 230, Sep. 2021, Art. no. 120805.

[19] L. He, Y. Wang, Y. Wei, M. Wang, X. Hu, and Q. Shi, "An adaptive central difference Kalman filter approach for state of charge estimation by fractional order model of lithium-ion battery," *Energy*, vol. 244, Apr. 2022, Art. no. 122627.

[20] Q. Zhong, F. Zhong, J. Cheng, H. Li, and S. Zhong, "State of charge estimation of lithium-ion batteries using fractional order sliding mode observer," *ISA Trans.*, vol. 66, pp. 448–459, Jan. 2017.

[21] S. Sun, Z. Gao, and K. Jia, "State of charge estimation of lithium-ion battery based on improved Hausdorff gradient using wavelet neural networks," *J. Energy Storage*, vol. 64, Aug. 2023, Art. no. 107184.

[22] F. Yang, W. Li, C. Li, and Q. Miao, "State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network," *Energy*, vol. 175, pp. 66–75, May 2019.

[23] Z. Ni and Y. Yang, "A combined data-model method for state-of-charge estimation of lithium-ion batteries," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[24] X. Liu, Q. Li, L. Wang, M. Lin, and J. Wu, "Data-driven state of charge estimation for power battery with improved extended Kalman filter," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.

[25] Y. Tian, R. Lai, X. Li, L. Xiang, and J. Tian, "A combined method for state-of-charge estimation for lithium-ion batteries using a long short-term memory network and an adaptive cubature Kalman filter," *Appl. Energy*, vol. 265, May 2020, Art. no. 114789.

[26] J. Tian, R. Xiong, W. Shen, and J. Lu, "State-of-charge estimation of $LiFePO_4$ batteries in electric vehicles: A deep-learning enabled approach," *Appl. Energy*, vol. 291, Jun. 2021, Art. no. 116812.

[27] F. Yang, S. Zhang, W. Li, and Q. Miao, "State-of-charge estimation of lithium-ion batteries using LSTM and UKF," *Energy*, vol. 201, Jun. 2020, Art. no. 117664.

[28] Z. Zhou, Y. Li, Q.-G. Wang, and J. Yu, "Health indicators identification of lithium-ion battery from electrochemical impedance spectroscopy using geometric analysis," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–9, 2023.

[29] Y. Xing, W. He, M. Pecht, and K. L. Tsui, "State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures," *Appl. Energy*, vol. 113, pp. 106–115, Jan. 2014.

[30] X. Liu, Z. Chen, C. Zhang, and J. Wu, "A novel temperature-compensated model for power li-ion batteries with dual-particle-filter state of charge estimation," *Appl. Energy*, vol. 123, pp. 263–272, Jun. 2014.

[31] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[32] W. Cao, Y. Liu, H. Mei, H. Shang, and Y. Yu, "Short-term district power load self-prediction based on improved XGBoost model," *Eng. Appl. Artif. Intell.*, vol. 126, Nov. 2023, Art. no. 106826.

[33] P. Liashchynskyi and P. Liashchynskyi, "Grid search, random search, genetic algorithm: A big comparison for NAS," 2019, *arXiv:1912.06059*.

[34] Z. Ning, Z. Deng, J. Li, H. Liu, and W. Guo, "Co-estimation of state of charge and state of health for 48 V battery system based on cubature Kalman filter and H-infinity," *J. Energy Storage*, vol. 56, Dec. 2022, Art. no. 106052.

[35] J. Zhang, T. Bi, and H. Liu, "Dynamic state estimation of a grid-connected converter of a renewable generation system using adaptive cubature Kalman filtering," *Int. J. Electr. Power Energy Syst.*, vol. 143, Dec. 2022, Art. no. 108470.

[36] K. Li, F. Zhou, X. Chen, W. Yang, J. Shen, and Z. Song, "State-of-charge estimation combination algorithm for lithium-ion batteries with Frobenius-norm-based QR decomposition modified adaptive cubature Kalman filter and H-infinity filter based on electro-thermal model," *Energy*, vol. 263, Jan. 2023, Art. no. 125763.

Weilu Hou received the master's degree in transportation engineering from the School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei, China, in 2023, where he is currently pursuing the Ph.D. degree with the School of Automotive and Transportation Engineering.

His research interests primarily focus on electric vehicle big data modeling and analysis.

Qin Shi received the master's degree from Jilin University, Changchun, China, in 1989, and the Ph.D. degree from the Hefei University of Technology, Hefei, China, in 2006.

Since 2005, she has been working as a Professor with the School of Automotive and Transportation Engineering, Hefei University of Technology. Currently, she holds the position of Director at the Engineering Research Center for Intelligent Transportation and Cooperative Vehicle-Infrastructure of Anhui Province. Throughout her career, she has successfully completed two Surface Projects of the National Natural Science Foundation of China and one International Exchange and Cooperation Project of the National Natural Science Foundation of China. Her research interests primarily revolve around intelligent connected vehicles, vehicle-road collaboration, intelligent transportation, and optimization of logistics systems.

Yiwen Liu received the master's degree in vehicle engineering from the School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei, China, in 2023.

His research interests primarily focus on electric vehicle battery management and control.

Liquan Guo received the Ph.D. degree from Dalian Maritime University, Dalian, China, in 2019.

Since 2019, he has been serving as an Assistant Professor at the School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei, China. His research interests primarily lie in transport modeling, transport big data, machine learning, operations research, and transport demand analysis.

Xiaonan Zhang received the B.E. degree in transportation from the Zhengzhou University of Aeronautics, Zhengzhou, China, in 2021. She is currently pursuing the master's degree with the School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei, China.

Her research interests primarily focus on electric vehicle big data modeling and analysis.

Ji Wu (Member, IEEE) received the B.E. degree in automation from the Hefei University of Technology (HFUT), Hefei, China, in 2011, and the Ph.D. degree in control science and technology from the University of Science and Technology of China, Hefei, in 2018.

From 2016 to 2017, he was a Guest Ph.D. Student with the Department of Energy Technology, Aalborg University, Aalborg, Denmark. He has been a Lecturer with the School of Automotive and Transportation Engineering, HFUT, since 2018, where he leads the Cybernetics Energy Transportation Laboratory (CETL). His areas of research are the modeling, control, and optimization of complex systems, including battery energy storage systems, electric vehicles, and microgrids.