Remaining Useful Lifetime Prediction of Lithium-Ion Batteries Based on Fragment Data and Trend Identification

Yiqing Lu, *Student Member, IEEE*, Ye Shi, *Member, IEEE*, Yu Liu, *Senior Member, IEEE*, and Haoyu Wang, *Senior Member, IEEE*

***Abstract*—Existing methods for predicting lithium-ion battery remaining useful lifetime (RUL) rely on complete capacity degradation data or extensive historical profiles. However, such sufficient conditions are usually unavailable in practical battery usage. To cope with this issue, a framework for RUL estimation with fragment data is proposed. The framework utilizes a small amount of prior knowledge as benchmark data to create an empirical model-based predictive method for estimating RUL by fragment historical data during nonlinear degradation stage. A more specified parameter initialization is obtained by trend identification of the fragment. Particle filter (PF) algorithm is utilized for model parameter update with proposed improved resampling strategy. RUL predictions using two different datasets demonstrate the effectiveness of the proposed method. An error margin of less than ten cycles in RUL predictions is consistently achieved in CS2 dataset when employing fragments ranging from 50 to 60 cycles. And an error margin of around 20 cycles is achieved in CX2 dataset by fragments ranging from 60 to 80 cycles. The proposed method renders a more precise and stable predictive result with high confident level.**

***Index Terms*—Lithium-ion battery, prediction, remaining useful life, trend identification.**

I. Introduction

A. Research Background and Problem Definition

LITHIUM-ION battery excels among different energy storage technologies due to its low self-discharge rate, long charge and discharge cycles, and stable electrochemical characteristics [1], [2], [3]. The capacity of lithium-ion batteries degrades over time. The degradation is caused by extrinsic factors, such as temperature and C-rate, as well as intrinsic factors, such as parasitic reactions [4], [5]. As the increasing cycles of repeated charging and discharging process, the battery experiences capacity fade, leading to reduced runtime and performance inefficiencies. To effectively address the challenge of capacity degradation in lithium-ion batteries, accurate prediction of capacity trends and estimation of the remaining useful lifetime (RUL) is essential.

The RUL of a battery refers to the number of operational cycles it can undergo before its capacity deteriorates to the failure threshold [6]. Methods for capacity prediction and RUL estimation typically adhere to a structured procedure. A predictive model is formulated to determine the health status of the battery based on battery performance characteristics. Then, selected performance characteristics of the target battery are gathered from its historical operational cycles. By integrating the obtained features with the predictive model, the number of cycles before the battery reaches its failure threshold is calculated, which will be considered as RUL.

Such model always require a dataset comprising multiple batteries of a similar type and operating under consistent usage patterns, thereby enabling a comprehensive reflection from battery characteristic to health status. However, greater variability in working condition and usage scenarios exist in the real-world battery, making it challenging to obtain a comprehensive and consistent dataset. Furthermore, a number of portable or mobile devices are not equipped with advanced battery management systems. Such devices often lack the capability to collect multiple performance characteristics and archive extensive historical battery data. Therefore, an RUL prediction method based on fragment data and limited background knowledge is proposed to handle the aforementioned scenario.

B. Related Work

Various approaches have been explored to predict battery capacity degradation and estimate RUL. The method of modeling battery degradation behavior can be categorized into physics-based models, empirical models, and data-driven models.

Physics models directly incorporate the chemical reactions occurring inside the battery, providing an accurate description of capacity degradation mechanisms [7]. However, due to the complexity of chemical reactions, physics models often involve more intricate parameters [8], [9]. Although reduced-order models [10] and equivalent circuit models [4], [11] are proposed for

Received 4 November 2024; accepted 29 December 2024. Date of publication 31 January 2025; date of current version 21 April 2025. This work was supported by the National Natural Science Foundation of China under Grant 52077140. Paper no. TII-24-5839. *(Corresponding author: Haoyu Wang.)*

The authors are with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China, and also with the Shanghai Engineering Research Center of Energy Efficient and Custom AI IC, Shanghai 201210, China (e-mail: luyq1@shanghaitech.edu.cn; shiye@shanghaitech.edu.cn; liuyu@shanghaitech.edu.cn; wanghy@shanghaitech.edu.cn).

Digital Object Identifier 10.1109/TII.2025.3528583

simplification, most of the parameters can still only be sampled and tracked by dedicated equipment. Consequently, they are unsuitable for online estimation applications.

In recent years, artificial intelligence has gained tremendous popularity [12], with a particular focus on data-driven models. These approaches emphasize the utilization of data for learning, without relying on explicit knowledge of underlying degradation mechanisms. Various data-driven algorithms, such as support vector machines [13], Gaussian process regression (GPR) [14], [15], and various artificial neural networks [16], [17], have been employed to construct predictive models based on input–output relationships. In a recent study [17], a BiGRU-TSAM network was established based on three or more complete degradation curves as training data to extract battery features. Another study [18] requires an even larger set of features from repeated charge and discharge cycles to ensure the accuracy of the predictive model. Although it is claimed in [15] that a model-free GPR approach can extract statistical characteristics from historical battery running data, it is important to note that the accuracy of prediction still heavily relies on the length of available historical data before making predictions. In [19] and [20], advanced strategies such as transfer learning and virtual data generation are proposed to address the challenge of limited historical data in model derivation and prediction. These methods involve synthesizing measured data with a pretrained model to enhance the training and prediction process. However, a well-prepared pretrain model still requires adequate dataset of battery operation data to yield accurate capacity predictions.

Therefore, for most data-driven methods, the availability of sufficient training data is paramount to establishing a robust and generalized model. The model's accuracy is significantly influenced by the quality and representativeness of the training data and can be adversely affected when dealing with limited or noisy data in practical usage.

On the contrary, empirical models are more readily integrated into probabilistic frameworks for RUL prediction. These models employ various functions to characterize the degradation process [21], [22]. Consequently, they can effectively capture the overall degradation trend of the battery with relatively limited data acquisition. However, the inherent lack of physical interpretation in empirical models means that they primarily capture the shape of capacity decay trends and may lack adaptability. As a result, the parameters obtained from an empirical model should be updated and adjusted, considering the actual measurements of the target battery. State–space methods, such as the Kalman filter (KF) [23] and particle filter (PF) [24], [25], are widely utilized for parameter update, which is accomplished by integrating information from both the model and actual measurements.

As demonstrated in [26], PF exhibits superior accuracy across a spectrum of empirical and physics-based models compared to a group of filtering algorithms, due to its inherent flexibility and capability to manage nonlinear, non-Gaussian states. Furthermore, in [27] and [28], adaptations of the PF have been introduced to tackle the challenge of particle degeneracy and to augment the effectiveness of the predictions. However, what is not often thoroughly clarified is the critical dependence of the PF algorithm's accuracy on the proper configuration of hyperparameters and initial values.

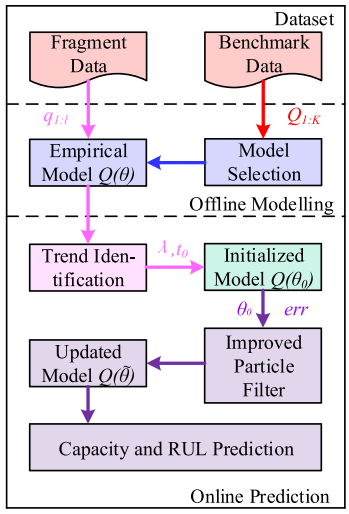


Fig. 1. Proposed framework for RUL prediction with fragment historical data.

The hyperparameters usually consist of the noise distribution of the state and observation equation, which reflects the modeling and measurement error in the algorithm. Modeling error is significantly influenced by the initial guess of the model parameters at the start of the PF. Generally, a smaller modeling error is conducive to achieving greater stability and precision in predictions. The majority of existing model-based methodologies operate under the assumption that complete historical data starting from the very beginning of battery degradation is accessible. Therefore, the initial guess can be set equal to the parameters obtained from empirical model to achieve a small modeling error. In [21], to consider the diversity between the battery data for modeling and the target battery, the model parameters were initialized via Dempster–Shafer theory. The method is proved effective for integrating diverse battery data and is applicable in parametric initialization. Nonetheless, when dealing with fragment data with unknown starting cycle still complicates the initialization process. Since the inherent differences in the working environment and usage patterns between different target batteries, an optimal initial value in one set of test may not be directly transferable to another target battery. The existing approaches face challenges when applied to RUL estimation scenarios involving fragmented data.

C. Motivations and Contributions of This Work

In this sense, a capacity and RUL estimation method using fragment data is proposed, as shown in Fig. 1. The method requires a prior degradation curve as a benchmark to represent one possible degradation process and establish an empirical model. Battery degradation usually consists of a mild linear phase and a accelerated nonlinear phase [29]. During the linear degradation phase, the trend observed from a fragment does not necessarily presage the subsequent nonlinear, rapid degradation phase. Therefore, predictions made during the linear phase may

not accurately account for the accelerated degradation that occurs in the nonlinear phase. In this work, fragments are first classified according to its degradation rate compared to the nominal capacity.

For the fragment capturing a sufficient duration on nonlinear phase, its trend is identified by comparing the degradation pattern from the empirical model. This trend identification is used for a more fitted initialization of PF. The model is further updated with PF algorithm based on the measured capacity of the fragment, which provides a better description of the fragment dynamics. By updating the model, the capacity and RUL can be predicted. The contributions and innovations of the proposed method can be summarized as follows.

1) The proposed method only requires a single complete capacity degradation curve of a specific type of battery under certain operational conditions as the benchmark. This benchmark serves as prior knowledge and models the general trend.
2) A trend identification approach is employed to determine the initial guess for the model parameters describing the fragment data. This provides a more appropriate initialization for the parameter update algorithm, narrowing the parameter searching range and increasing algorithm stability. The parameter update algorithm is also improved based on the PF, resulting in more accurate predictions and a more concentrated distribution of prediction results.
3) A capacity and RUL estimation framework is proposed, allowing flexibility in choosing the modeling and parameters. Satisfactory RUL estimation can be achieved even with a basic modeling and prediction method.

II. Methodology

A. Selection of Empirical Model

To effectively model battery degradation trends, empirical models are utilized due to their simplicity and intuitive form. Various expressions of empirical models have been developed to describe the degradation process of different types of batteries, such as the double exponential model, polynomial model, and ensemble model [22]. These models are represented by (1), (2), and (3) as shown in the following:

$$Q = a\exp(b\cdot k) + c\exp(d\cdot k) \tag{1}$$

$$Q = ak^2 + bk + c \tag{2}$$

$$Q = a\exp(b\cdot k) + ck^2 + d \tag{3}$$

where $k = 1, 2, 3, \cdots K$ is the cycles of repeated charging and discharging of the battery cell and the maximum length of model is denoted as K.

The three different models are applied to fit the degradation curve from two typical degradation curve in the MATLAB environment using the Levenberg–Marquardt algorithm to minimize the root-mean-square error (RMSE). The fitting performance of each model on the benchmark curve of the two datasets is depicted in Fig. 2.

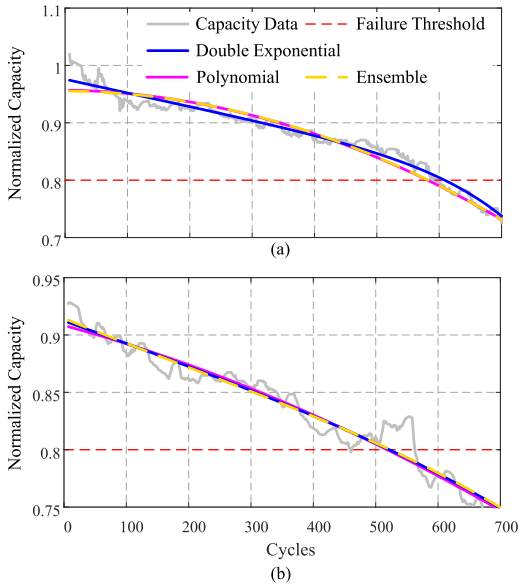


Fig. 2. Fitting performance among different models applied to the typical degradation curve of two datasets. (a) CS2-37. (b) CX2-37.

Among the three models considered, both double exponential model and ensemble model demonstrate superior fitting performance. However, the inclusion of a quadratic term in the ensemble model introduces potential concerns regarding model stability. Hence, the double exponential model is selected as the empirical model, with the parameters of a, b, c, and d determined based on the benchmark curve. The derived empirical model from benchmark curve is denoted as $Q_{1:K}$.

B. Fragment Trend Identification

In the conventional PF method, the initial states and the variance of noise for particles are often determined empirically. Typically, increasing the selection of modeling error can expand the search space for particles, potentially aiding in the discovery of optimal model parameters. Given the nonlinear and exponential expression of degradation, even minor variations in individual parameters can accumulate to significantly impact the future trend predictions. This can adversely affect the convergence and stability of the PF algorithm. It is crucial to constrain the variance of each parameter within a narrow range to maintain robustness. Therefore, initializing model parameters with a small modeling error is essential for the effectiveness of the filtering algorithm.

For a typical capacity fragment observed in practical battery operation, the initial cycle is often unknown. The fragment is noted as $q_{1:l_{\text{frag}}}$, where l_{frag} is the fragment length. Since the fragment share a similar degradation trend with the benchmark model Q, the relationship between the fragment and benchmark can be expressed as

$$q(t) = Q(\lambda t + t_0) \tag{4}$$

where $t = 1, 2, \cdots l_{frag}$. By examining the normalized capacities at both the starting and ending point of fragment, it is feasible to identify a corresponding segment on the benchmark model that exhibit an equivalent degree of degradation. Recognizing

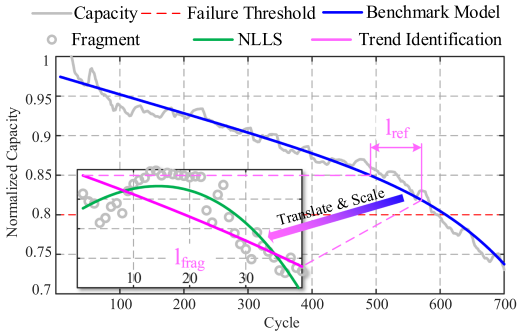


Fig. 3. Examples of trend identification approach applied to sampled fragment data.

that fragment data of target battery may be influenced by measurement errors and variability stemming from local dynamic characteristics, the starting and ending capacity value to fit for the fragment is defined as

$$\hat{q}(1) = \frac{1}{2}(q(1) + \max(q(t))$$
$$\hat{q}(l_{\text{frag}}) = \frac{1}{2}(q(l_{\text{frag}}) + \min(q(t)). \tag{5}$$

The corresponding segment in the benchmark model is characterized by its initiation at the k_{ref} cycle and a duration of l_{ref} cycles. Therefore

$$\hat{q}(1) = Q(\lambda + t_0) = Q(k_{\text{ref}})$$
$$\hat{q}(l_{\text{frag}}) = Q(\lambda l_{\text{frag}} + t_0) = Q(k_{\text{ref}} + l_{\text{ref}}). \tag{6}$$

The target fragment can be considered as a translate and scale transformation of the benchmark model, following:

$$\lambda = l_{\text{ref}}/(l_{\text{frag}} - 1)$$
$$t_0 = k_{\text{ref}} - \lambda. \tag{7}$$

The initial guess of the model parameters based on trend identification can be expressed as

$$a_0 = a\exp(bt_0)$$
$$b_0 = \lambda b$$
$$c_0 = c\exp(dt_0)$$
$$d_0 = \lambda d \tag{8}$$

where $[a_0; b_0; c_0; d_0]$ are the parameters of the benchmark model. As an example shown in Fig. 3, the pink curve is the initial guess of the parameters based on identifying the trend of fragment. The green curve is the initial guess directly based on fitting the fragment with double exponential model using nonlinear least squares (NLLS).

Although the initial guess derived by NLLS aims to minimize the discrepancy between the model and the data fragment, this approach can render the model more sensitive to the influence of outliers and inherent fluctuations within the fragment. Given that there exist capacity recovery within the fragment, the NLLS fitting process could be adversely affected by this phenomenon. The resultant estimate may exhibit an initial rise followed by a decline, potentially leading to a misalignment between the model parameters and the actual trajectory of battery capacity degradation.

By utilizing the trend identification of the fragment, the empirical model with refined initial parameters $\theta_0 = [a_0; b_0; c_0; d_0]$ can effectively capture the trend of fragments and subsequently be utilized for capacity and RUL prediction.

C. Parameter Update for Capacity Prediction

To improve the prediction of the future trend, the model parameters are further updated based on the specific measured fragment data. Given the strong nonlinearity of battery capacity degradation and the potential presence of complex noise distribution in the sampled data, an improved PF-based parameter update method is employed.

To model the uncertainty of the parameters, the noise of parameters a, b, c, and d, as well as the error of model output is assumed to follow a normal distribution. The state equation of the battery degradation process can be expressed as follows:

$$a_t = a_{t-1} + v_a$$
$$b_t = b_{t-1} + v_b$$
$$c_t = c_{t-1} + v_c$$
$$d_t = d_{t-1} + v_d \tag{9}$$
$$q_t = a_t \exp(b_t \cdot t) + c_t \exp(d_t \cdot t) + v_r \tag{10}$$

where $[a_t; b_t; c_t; d_t]$ are the parameters estimated at cycle t, and q_t is the measured capacity from fragment at cycle t. $v_a \sim N(0, \sigma_a^2)$, $v_b \sim N(0, \sigma_b^2)$, $v_c \sim N(0, \sigma_c^2)$, $v_d \sim N(0, \sigma_d^2)$ are the noise variation of state transition and $v_r \sim N(0, \sigma_r^2)$ is the noise variation of output.

The parameters are initialized considering the trend and predicted position of the fragment, as described in (7). The updated parameters at cycle t are denoted as $X_t = [a_t, b_t, c_t, d_t]$. Given a series of measurements from the fragment data $q_{1:l_{frag}}$, the goal is to estimate the proper probability distribution of the parameter $P(X_t|q_{1:t})$. Within the Bayesian framework, the posterior distribution $P(X_t|q_{1:t})$ can be recursively derived from the distribution $P(X_{t-1}|q_{1:t-1})$ obtained from the previous cycle.

According to the Chapman–Kolmogorov equation, the prior probability distribution of X_t considering posterior distribution $P(X_{t-1}|q_{1:t-1})$ obtained from cycle $k-1$ can be expressed as

$$P(X_t|q_{1:t-1}) = \int P(X_t|X_{t-1})P(X_{t-1}|q_{1:t-1})\mathrm{d}X_{t-1}. \tag{11}$$

At cycle t, new observation q_t is obtained and used to update the prior distribution via Bayes' rule

$$P(X_k|q_{1:t}) = \frac{P(X_t|q_{1:t-1})P(q_t|X_t)}{P(q_t|q_{1:t-1})} \tag{12}$$

where $P(q_t|q_{1:t-1})$ is a normalized constant. However, analytically evaluating these distributions is challenging due to the complex integrals. Therefore, Monte Carlo simulation is utilized to approximate the probability density function with a set of

particles and their associated weights

$$P(x_t|q_{1:t}) = \sum_i w_t^i \delta(X_t - X_t^i) \tag{13}$$

where $X_t^i (i = 1, 2, 3, \ldots, M)$ represents a set of particles drawn from a distribution, and w_t^i signifies the Bayesian importance weight associated with the particles. Since the precise distribution of $P(X_t|q_{1:t})$ is unknown, the particles are sampled using a manually designed importance sampling function. As the number of sampled particles increases, the simulation of the distribution becomes more representative of the true underlying process. The distribution of particles weights is updated in a recursive formula expressed as

$$w_t^i = w_{t-1}^i P(q_t|X_t^i). \tag{14}$$

Typically, the likelihood $P(q_t|X_t^i)$ is decided regarding the residual between particles and measurements at cycle t, which is expressed as

$$z_t^i = q_t^i - q_t$$

$$P(q_t|X_t^i) = \frac{1}{\sqrt{2\pi}\sigma_r} \exp(-(z_t^i)^2/2\sigma_r^2). \tag{15}$$

In conventional PF algorithm, to mitigate the issue of particle degeneracy, number of effective particles (N_{eff}) is calculated after each iteration. Resampling is triggered if N_{eff} drops under a predetermined threshold. However, since a more precise initialization of model parameters, a rather small value of variance in state equation is obtained, leading to a reduced disparity in particle weights. By conventional random or stratified resampling, diminished fluctuations in particle movement potentially lead to a wide-spread of particles range. Therefore, a manual intervention is applied to enhance the convergence. Specifically, the particles with lower weights are intentionally converged toward the particles $X_{\max}$ with the highest weight using the following equation:

$$X_k^{i\prime} = X_k^i + \beta(X_{\max} - X_k^i) \tag{16}$$

where β is a random step with manually controlled upper bound to decide the tendency of particle convergence. As a result, a more concentrated particle distribution and more stable results can be obtained through iteration. The updated parameters for each iteration can be expressed as

$$\tilde{X}_k = \sum_{j=1}^{M} w_k^i X_k^i. \tag{17}$$

The weighted average of the particles after the final iteration serves as an expectation of the future degradation trend of the fragment. In addition, the independent models represented by each particle contribute to the predicted probability density (PDF) distribution.

III. Experiment Conduction

A. Dataset Description

The experiment utilizes two types of batteries from the Center for Advanced Life Cycle Engineering (CALCE) in University of Maryland [22]: CS2 and CX2 cells, both with $LiCoO_2$ cathodes but differing in configuration and size. CS2 cells have a capacity of 1.1 Ah, while CX2 cells have 1.35 Ah. Each type undergoes repeated charge-discharge cycles without strict thermal control. CS2 cells are charged at 1 C constant current until 4.2 V, then voltage hold until current falls below 0.05 A, and discharged at 1C constant current until 2.7 V. CX2 cells follow the same profile but the current rate is at 0.5 C.

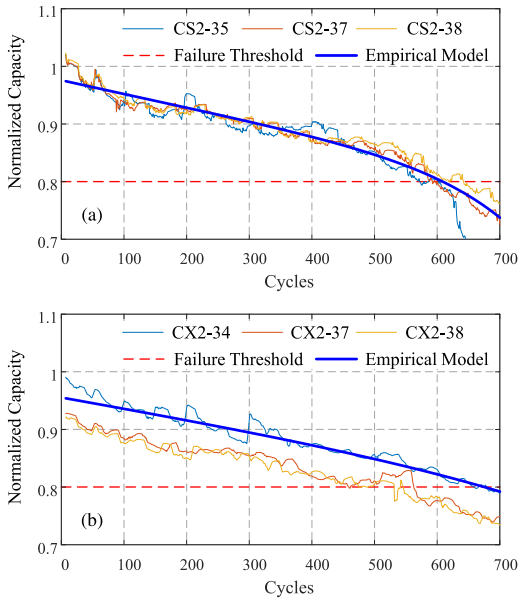


Fig. 4. Dataset utilized in experiment. (a) CS2 cells. (b) CX2 cells.

Three cells are selected for each type. One serves as the benchmark for establishing an empirical model based on its degradation trend, while the other two are practical batteries for validating the proposed method using fragment data to predict capacity and RUL. For CS2 cells, CS2-37 is the benchmark, and CS2-35, CS2-38 are for validation. For CX2 cells, CX2-34 is the benchmark, and CX2-37, CX2-38 is for validation. Both types is considered reach EOL when capacity degrades to 80% of rated capacity. Fig. 4 shows the normalized capacity degradation curves, revealing different trends between types and variations within the same type due to intrinsic parameters.

B. Experiment Design for RUL Prediction

The experiment follows the proposed framework shown in Fig. 4. CS2 and CX2 are independent datasets and are tested separately. We take the capacity and RUL prediction experiment on CS2 cells as an example. The experiment procedure can be summarized into the following steps.

1) Data processing: The benchmark data (CS2-37) undergoes data smoothing and outliers removal.
2) Empirical model Q establishment: Double exponential model is selected to represent the degradation trend based on curve fitting of the benchmark data.
3) Fragment generation: Fragments are generated from the validation data using a starting cycle k_0 and a fragment length l_{frag}.

4) Trend identification: A trend identification approach is applied to identify the segment that best represents the fragment trend in Q. The model parameters are initialized according to (7) for subsequent parameter updates.
5) Improved PF algorithm: The improved PF algorithm is applied to tune the model parameters, considering the dynamics of the fragment. The updated model, denoted as Q_{pred}, is obtained as the expectation of the updated model parameters from (17).
6) RUL prediction: The updated model Q_{pred} is extended to determine the cycle when Q_{pred} reaches the failure threshold, which serves as the predicted RUL. The absolute error (AE) between the real RUL and the predicted RUL is calculated to evaluate the prediction accuracy. In addition, the width of PDF is computed as an indicator of model uncertainty and stability.
7) Repeat different fragments or apply on other target batteries.
8) For the validation of the universality of the proposed framework, CS2-38 and CX2-37 are also considered as a new benchmark data and repeat the testing process.

C. Parameter Configuration for Prediction Algorithm

Since it is more necessary for the capacity degradation trend and RUL prediction when battery enters the nonlinear degradation stage, the fragment data selected for prediction should include adequate length at the nonlinear stage. As the degradation mechanism is not the primary focus of this work, the onset of the nonlinear degradation stage for battery cells is identified when the cell capacity drops to 85% of its nominal capacity.

To compare the proposed method with conventional methods, two different filtering algorithms, KF and PF, are employed for prediction purposes. Both algorithms are initialized using NLLS and trend identification methods. Given that NLLS and trend identification provide a precise initial estimate for fitting the fragment data, the noise variance of the state equation is set to 10% of the initial parameter values. The measurement noise variance is set at 1%, a uniform value across all fragment input and filter algorithms.

For the KF, implementing adaptive noise variance adjustment can reduce the algorithm's sensitivity to the initial estimate. Thus, the adaptive Kalman filter (AKF), initialized with an estimate from NLLS, is chosen as a representative method. Regarding the PF, when the fragment includes linear stage data, the proposed method aligns with the conventional PF, serving as a tracker and gradually update model parameters to fit the degradation trend. An enhanced resampling strategy is implemented as the algorithm progresses into the nonlinear stage. The manual intervention parameter, β, which governs the step of particle convergence, is set to a random value not exceeding 0.5. This setting ensures a controlled aggregation of particles without being overly aggressive, maintaining a balance between convergence and diversity in the PF.

To assess the performance of the methods, comparative analysis is conducted utilizing two principal metrics, AE and the width of PDF (W_{PDF}). The specific calculation formula is as follows:

$$\begin{aligned} \text{AE} &= |\text{RUL}_p - \text{RUL}_t| \\ W_{\text{PDF}} &= \max(\text{RUL}_{p,i}) - \min(\text{RUL}_{p,i}) \end{aligned} \tag{18}$$

AE quantifies the discrepancy between the predicted RUL and the true RUL, which serves as a direct reflection of the prediction accuracy. PDF is constructed from the aggregated distribution of particles within the PF algorithm. The width of the PDF represents the span between the maximum and minimum predictions derived from the particle ensemble. The PDF width listed in the table indicates the change using the proposed improved resampling method compared to the random resampling method. A more constricted PDF signifies a tighter clustering of particles, thereby augmenting the robustness and reliability of the predictive results.

IV. Results and Discussions

Figs. 5–10 show fragments of varying lengths from the nonlinear degradation stage of different batteries, represented by dashed blue lines as historical data inputs for prediction models. The figures also display prediction results from various models using curves of different colors: green for AKF, yellow for conventional PF, and purple for PF with improved resampling strategy. The shaded areas of corresponding colors indicate the PDF ranges for the PF algorithm, showing prediction uncertainty. Predicted RUL, AE, and PDF width are detailed in Tables I and II for analysis and comparison.

A. Comparison of Prediction Algorithm

The results indicate that predictions derived from the AKF are susceptible to the influence of outlier anomalies. AKF primarily operates by extending future estimates based on the local trajectory of historical data. Consequently, the stability of predictions is not constant and can be significantly impacted by measurement errors or capacity regeneration. In the case of the CS-38 dataset, particularly when the fragment length is set to 35, the fragment ended at a region where capacity regeneration occurs. The AKF model, in response to this increase, tends to follow the trend, thereby projecting a ascending future trend. Similar results can also be found in CX2-38 battery.

Compared with AKF, PF demonstrates a modest enhancement in its resilience to fluctuations, primarily attributed to its nonparametric nature. This characteristic allows PF to adapt more flexibly to the underlying dynamics of the system, thereby providing a improvement in the robustness of predictions against variability. Despite this, the predictive accuracy still encounters challenges in precisely capturing the actual degradation trends of system using NLLS initialization.

By applying the initial guess with trend identification, the predictive accuracy is significantly enhanced. For CS2-35, AE is decreased to one cycle, eight cycles, and one cycle at the fragment length of 40, 50, and 60 cycles, respectively. This is a marked improvement over the results with NLLS initialization, which yielded errors of 42 cycles, 21 cycles, and 18 cycles for the same fragment lengths. For CS2-38, the application of trend

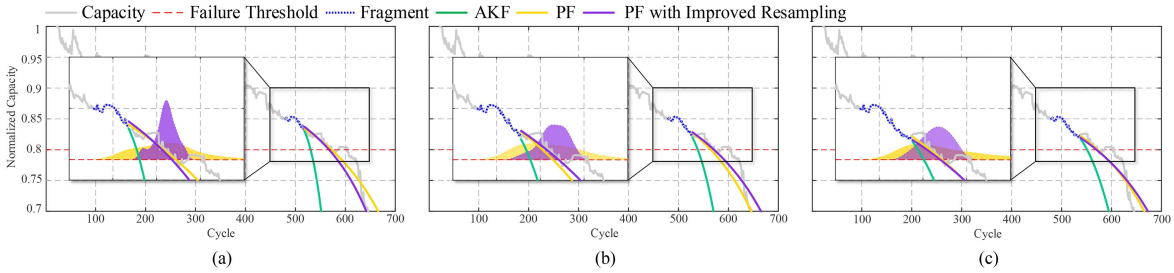


Fig. 5. CS2-35 capacity prediction and RUL prediction based on the benchmark data of CS2-37. (a) $k_0 = 478$, $l_{frag} = 40$. (b) $k_0 = 478$, $l_{frag} = 50$. (c) $k_0 = 478$, $l_{frag} = 60$.

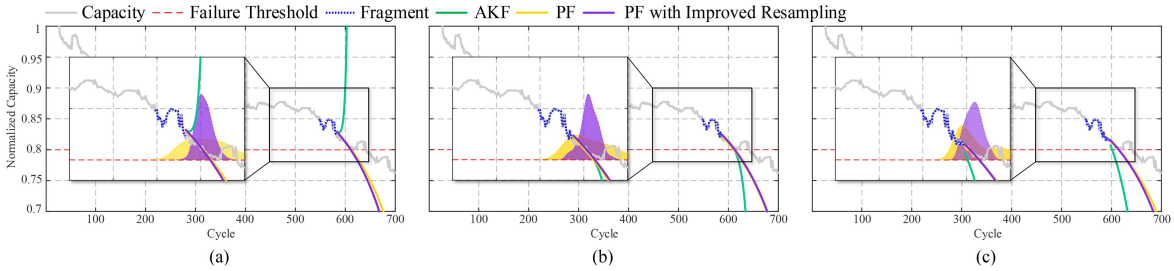


Fig. 6. CS2-38 capacity prediction and RUL prediction based on the benchmark data of CS2-37. (a) $k_0 = 548$, $l_{frag} = 35$. (b) $k_0 = 548$, $l_{frag} = 40$. (c) $k_0 = 548$, $l_{frag} = 50$.

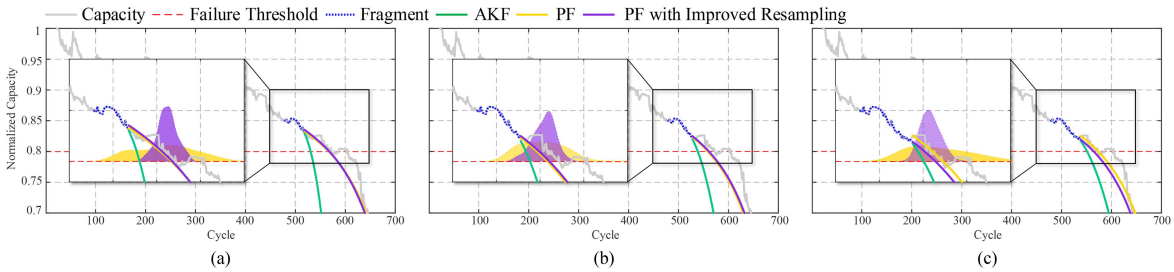


Fig. 7. CS2-35 capacity prediction and RUL prediction based on the benchmark data of CS2-38. (a) $k_0 = 478$, $l_{frag} = 40$. (b) $k_0 = 478$, $l_{frag} = 50$. (c) $k_0 = 478$, $l_{frag} = 60$.

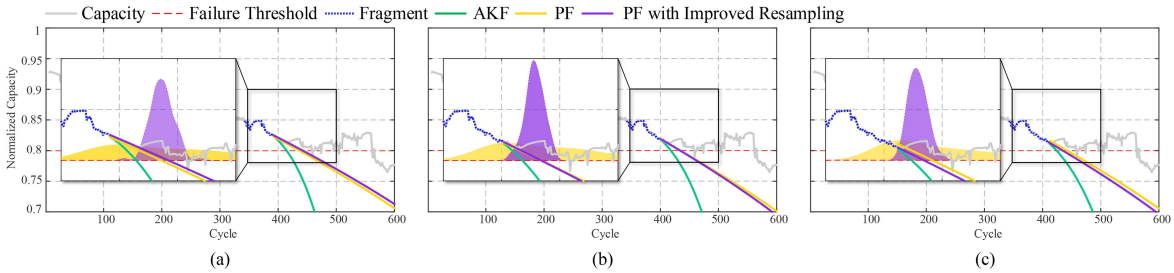


Fig. 8. CX2-37 capacity prediction and RUL prediction based on the benchmark data of CX2-34. (a) $k_0 = 342$, $l_{frag} = 50$. (b) $k_0 = 342$, $l_{frag} = 60$. (c) $k_0 = 342$, $l_{frag} = 70$.

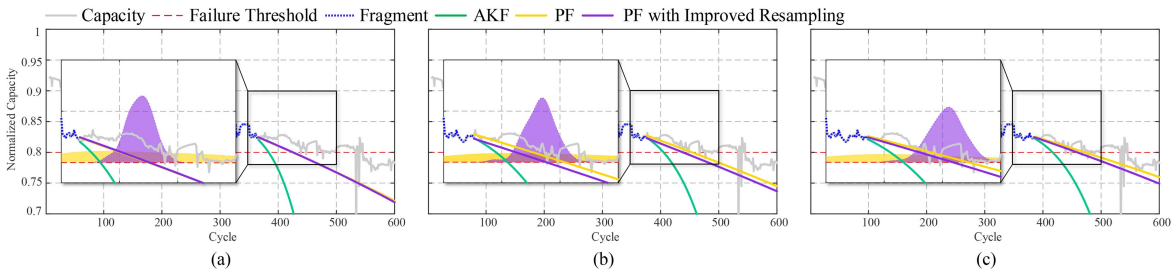


Fig. 9. CX2-38 capacity prediction and RUL prediction based on the benchmark data of CX2-34. (a) $k_0 = 301$, $l_{frag} = 65$. (b) $k_0 = 301$, $l_{frag} = 75$. (c) $k_0 = 301$, $l_{frag} = 85$.

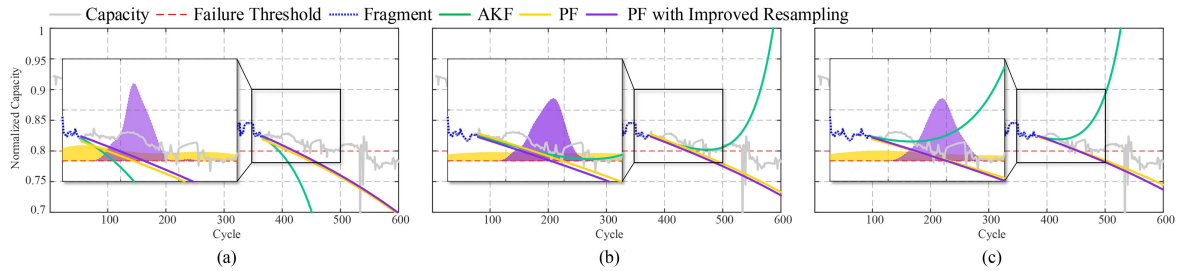


Fig. 10. CX2-38 capacity prediction and RUL prediction based on the benchmark data of CX2-37. (a) $k_0 = 301$, $l_{frag} = 65$. (b) $k_0 = 301$, $l_{frag} = 75$. (c) $k_0 = 301$, $l_{frag} = 85$.

TABLE I
RUL PREDICTION RESULTS ON CS2 CELLS

Cell	Bench-mark	Fragment			AKF			PF			PF			+ Improved Resampling		
		k_0	l_{frag}	RUL_t	Initial	RUL_p	AE	Initial	RUL_p	AE	Initial	RUL_p	AE	RUL_p	AE	PDF width
CS2-35		478	40	56		14	42		14	42		55	1	51	5	210 → 52
(EOL =	CS2-37	478	50	46	NLLS	12	34	NLLS	25	21	Trend	38	8	45	1	130 → 69
573)		478	60	36		15	21		18	18		35	1	33	3	190 → 62
CS2-35		478	40	56		14	42		14	42		50	6	51	5	120 → 52
(EOL =	CS2-38	478	50	46	NLLS	12	34	NLLS	25	21	Trend	32	14	34	12	120 → 63
573)		478	60	36		15	21		18	18		37	1	28	8	160 → 61
CS2-38		548	35	37		/	/		4	33		31	6	29	8	91 → 47
(EOL =	CS2-37	548	40	32	NLLS	25	7	NLLS	3	29	Trend	24	8	26	6	129 → 51
619)		548	50	22		5	17		13	9		19	3	18	4	114 → 38

TABLE II
RUL PREDICTION RESULTS ON CX2 CELLS

Cell	Bench-mark	Fragment			AKF			PF			PF			+ Improved Resampling		
		k_0	l_{frag}	RUL_t	Initial	RUL_p	AE	Initial	RUL_p	AE	Initial	RUL_p	AE	RUL_p	AE	PDF width
CX2-37		342	50	68		14	54		12	56		46	22	52	16	217 → 66
(EOL =	CX2-34	342	60	58	NLLS	12	46	NLLS	7	51	Trend	33	25	36	22	163 → 42
459)		342	70	48		11	37		9	39		34	14	25	23	178 → 38
CX2-38		301	65	74		5	69		/	/		60	14	61	13	221 → 53
(EOL =	CX2-34	301	75	64	NLLS	20	44	NLLS	31	33	Trend	80	16	64	0	354 → 85
439)		301	85	54		32	22		42	12		88	34	74	20	371 → 84
CX2-38		301	65	74		5	69		6	68		47	27	56	18	268 → 73
(EOL =	CX2-37	301	75	64	NLLS	/	/	NLLS	31	33	Trend	75	11	64	0	303 → 62
439)		301	85	54		/	/		42	12		67	13	65	11	406 → 62

identification for initialization also results in a significant reduction in AE. Specifically, for fragment lengths ranging from 35 to 50 cycles, the AE is consistently below ten cycles, indicating a higher level of precision with PF.

Moreover, the implementation of the improved resampling method has been shown to significantly narrow the PDF width by over 50%, while resulting in a marginal difference in AE when compared to the conventional PF approach. The error fluctuations observed with different fragment lengths are reduced. This suggests that the improved resampling method contributes to a more robust and reliable predictive framework capable of handling variations in target battery.

B. Effect of Fragment Length and Position

Fragments at the nonlinear degradation stage are deemed to have a more significant impact on the prediction of capacity trends and RUL. Therefore, the primary focus of this work is on fragments that commence at the onset of the nonlinear degradation stage. However, if the fragment length is excessively long, it may approach the failure threshold, thereby limiting the effectiveness of early predictions. To balance this, fragment selection is restricted to no more than half of the total cycles between the 85% and 80% capacity thresholds, which translates to approximately 30–60 cycles for CS2 batteries.

Comparing the prediction effect of the proposed PF with trend identification, a longer fragment length can generally encapsulate more detailed degradation information, leading to more reliable predictions. Nevertheless, the position of the fragment is also a critical variable, considering dynamics, such as capacity regeneration. In CS2-35, the observed increase in AE at the 50-cycle fragment length utilizing PF can be attributed to the occurrence of a capacity regeneration event immediately following the fragment. Given the absence of prior knowledge regarding this capacity increase, anticipating such a deviation in the degradation pattern is inherently challenging for filter algorithm. Conversely, when the fragment length is extended to 60 cycles, the predictive model is able to incorporate the newly observed capacity increase. This updated information allows for a more accurate trend identification of fragment, thereby enhancing the precision of future trend prediction.

For CS2-38 where capacity regeneration is not that obvious in the prediction range, the predictive accuracy of PF is enhanced as the fragment approaches to the 80% threshold.

C. Generalization Ability

The generalization ability of the proposed method is further validated by incorporating diverse benchmark data and on different battery type. A new benchmark empirical model with different parameters is established based on CS2-38 in place of CS2-37, which is shown in Fig. 7. The final prediction result exhibits a fluctuation of ± 5 cycles relative to the outcomes obtained using the CS2-37 benchmark. This observation demonstrates its capability to accommodate any reliable benchmark data to obtain effective prediction.

The prediction method is also applied to CX2 battery dataset. Due to the more noisy data and pronounced fluctuations, the capacity curve is prone to multiple crossings of the 80% nominal capacity failure threshold. This variability can introduce bias in the computation of the actual RUL. In this work, the first instance when the capacity curve degrades to the failure threshold is selected as the EOL. This selection may result in an overly conservative estimation of the true RUL, potentially overestimating the error. Consequently, as the AE data presented in the Table II, the prediction error associated with the CX2 dataset is observed to be greater.

However, as can be seen from Figs. 8 to 10, the PF prediction model remains highly informative for forecasting future capacity trends. For the CX2-38 dataset, the predictive error is constrained to approximately ten cycles, utilizing fragments with a length of 75–85 cycles. This level of accuracy is consistently achieved across two distinct benchmark models, demonstrating the robustness of the approach. In addition, the implementation of an improved resampling strategy has effectively gathered the distribution of particles, thereby narrowing the width of the PDF width by over 60% and enhancing the precision of the predictions.

Conversely, the CX2-37 dataset presents a capacity regeneration phenomenon occurring too much near the EOL. This regeneration inevitably leads to a more substantial error, since none of the selected fragment includes the potential trend at such a final stage. Nonetheless, in the majority of cases, the proposed methodology is capable of delivering estimates that are both reasonably accurate and stable.

V. Conclusion

In this article, a framework for estimating the RUL is proposed to address the challenge of fragmented data in RUL prediction. The framework utilizes a small amount of prior knowledge to create a general benchmark model that captures the degradation trend. It then extracts the trend from the fragment data to identify its degradation trend from the benchmark. The parameters are further updated based on the measured fragment data utilizing PF algorithm.

To validate the proposed framework, RUL predictions are conducted using two different datasets CS2 and CX2 with different battery types. In CS2 dataset, an error margin of less than ten cycles in RUL predictions is consistently achieved, particularly when employing fragments ranging from 40 to 50 cycles at the nonlinear degradation phase. In CX2 dataset, although an error margin of around 20 cycles exist in RUL predictions due to the inherent volatility of the battery data, the predictive model's stability and accuracy have nonetheless been enhanced relative to the AKF. It is plausible that the precision of predictions could be further elevated with the acquisition of more refined and densely sampled data.

The results demonstrate the effectiveness of the proposed method for RUL prediction based on the improved PF algorithm and trend identification approach. By leveraging the trend identification of fragments as the initial estimate for the filtering algorithm, the resultant predictions are rendered more precise and stable. Moreover, the refinement of the resampling phase within the PF algorithm serves to constrict the PDF width, thereby enhancing confidence in the predictive result.

Furthermore, the proposed framework provides a foundation and guidance for dealing with fragmented data in RUL estimation for practical battery usage scenarios. Future work can focus on applying more advanced modeling techniques, extracting additional features and information from the fragment data, and developing more accurate parameter identification methods. These advancements have the potential to further improve the accuracy of RUL estimation and enhance decision-making processes in various industries.

References

[1] X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics," *Joule*, vol. 4, no. 2, pp. 310–346, Feb. 2020.
[2] J. Xie, Y. Weng, and H. D. Nguyen, "Health-informed lifespan-oriented circular economic operation of li-ion batteries," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 2749–2760, Mar. 2023.
[3] Z. Jiao et al., "LightGBM-based framework for lithium-ion battery remaining useful life prediction under driving conditions," *IEEE Trans. Ind. Informat.*, vol. 19, no. 11, pp. 11353–11362, Nov. 2023.
[4] A. Barré, B. Deguilhem, S. Grolleau, M. Gérard, F. Suard, and D. Riu, "A review on lithium-ion battery ageing mechanisms and estimations for automotive applications," *J. Power Sources*, vol. 241, pp. 680–689, May 2013.
[5] W. Vermeer, G. R. C. Mouli, and P. Bauer, "A comprehensive review on the characteristics and modeling of lithium-ion battery aging," *IEEE Trans. Transport. Electrific.*, vol. 8, no. 2, pp. 2205–2232, Jun. 2022.
[6] K. Song, D. Hu, Y. Tong, and X. Yue, "Remaining life prediction of lithium-ion batteries based on health management: A review," *J. Energy Storage*, vol. 57, Nov. 2023, Art. no. 106193.
[7] X. Wang et al., "Research progress of battery life prediction methods based on physical model," *Energies*, vol. 16, no. 9, Apr. 2023, Art. no. 3858.
[8] J. Christensen and J. Newman, "A mathematical model for the lithium-ion negative electrode solid electrolyte interphase," *J. Electrochem. Soc.*, vol. 151, no. 11, Oct. 2004, Art. no. A1977.
[9] M. Jafari, K. Khan, and L. Gauchia, "Deterministic models of Li-ion battery aging: It is a matter of scale," *J. Energy Storage*, vol. 20, pp. 67–77, Sep. 2018.
[10] A. V. Randall, R. D. Perkins, X. Zhang, and G. L. Plett, "Controls oriented reduced order modeling of solid-electrolyte interphase layer growth," *J. Power Sources*, vol. 209, pp. 282–288, Feb. 2012.
[11] S. Nejad, D. Gladwin, and D. Stone, "A systematic review of lumped-parameter equivalent circuit models for real-time estimation of lithium-ion battery states," *J. Power Sources*, vol. 316, pp. 183–196, Mar. 2016.
[12] Y. Xiang, H. S.-H. Chung, and H. Lin, "Light implementation scheme of ANN-based explicit model-predictive control for DC-DC power converters," *IEEE Trans. Ind. Informat.*, vol. 20, no. 3, pp. 4065–4078, Mar. 2024.

[13] S. Li, H. Fang, and B. Shi, "Remaining useful life estimation of lithium-ion battery based on interacting multiple model particle filter and support vector regression," *Rel. Eng. Syst. Saf.*, vol. 210, Feb. 2021, Art. no. 107542.
[14] K. Liu, X. Hu, Z. Wei, Y. Li, and Y. Jiang, "Modified Gaussian process regression models for cyclic capacity prediction of lithium-ion batteries," *IEEE Trans. Transport. Electrific.*, vol. 5, no. 4, pp. 1225–1236, Dec. 2019.
[15] J. Meng, M. Yue, and D. Diallo, "A degradation empirical-model-free battery end-of-life prediction framework based on Gaussian process regression and Kalman filter," *IEEE Trans. Transport. Electrific.*, vol. 9, no. 4, pp. 4898–4908, Dec. 2023.
[16] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen, "A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3478–3487, May 2021.
[17] J. Zhang et al., "A data-model interactive remaining useful life prediction approach of lithium-ion batteries based on PF-BiGRU-TSAM," *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 1144–1154, Feb. 2024.
[18] K. A. Severson et al., "Data-driven prediction of battery cycle life before capacity degradation," *Nat. Energy*, vol. 4, no. 5, pp. 383–391, Mar. 2019.
[19] Z. Zhou, Y. Liu, M. You, R. Xiong, and X. Zhou, "Two-stage aging trajectory prediction of LFP lithium-ion battery based on transfer learning with the cycle life prediction," *Green Energy Intell. Transp.*, vol. 1, Apr. 2022, Art. no. 100008.
[20] Y. Tian, Q. Dong, J. Tian, X. Li, G. Li, and K. Mehran, "Capacity estimation of lithium-ion batteries based on optimized charging voltage section and virtual sample generation," *Appl. Energy*, vol. 332, Dec. 2023, Art. no. 120516.
[21] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method," *J. Power Sources*, vol. 196, no. 23, pp. 10314–10321, Aug. 2011.
[22] Y. Xing, E. W. Ma, K.-L. Tsui, and M. Pecht, "An ensemble model for predicting the remaining useful performance of lithium-ion batteries," *Microelectron. Rel.*, vol. 53, no. 6, pp. 811–820, Jan. 2013.
[23] N. Zhang, A. Xu, K. Wang, X. Han, W. Hong, and S. H. Hong, "Remaining useful life prediction of lithium batteries based on extended Kalman particle filter," *IEEJ Trans. Electr. Electron. Eng.*, vol. 16, no. 2, pp. 206–214, Jan. 2021.
[24] J. Zhang, Y. Jiang, X. Li, M. Huo, H. Luo, and S. Yin, "An adaptive remaining useful life prediction approach for single battery with unlabeled small sample data and parameter uncertainty," *Rel. Eng. Syst. Saf.*, vol. 222, Jan. 2022, Art. no. 108357.
[25] L.-H. Ye, S.-J. Chen, Y.-F. Shi, D.-H. Peng, and A.-P. Shi, "Remaining useful life prediction of lithium-ion battery based on chaotic particle swarm optimization and particle filter," *Int. J. Electrochem. Sci.*, vol. 18, Mar. 2023, Art. no. 100122.
[26] E. Walker, S. Rayman, and R. E. White, "Comparison of a particle filter and other state estimation methods for prognostics of lithium-ion batteries," *J. Power Sources*, vol. 287, pp. 1–12, Apr. 2015.
[27] Y. Wang, Y. Peng, and T. W. Chow, "Adaptive particle filter-based approach for RUL prediction under uncertain varying stresses with application to HDD," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6272–6281, Sep. 2021.
[28] M. Ahwiadi and W. Wang, "An enhanced particle filter technology for battery system state estimation and RUL prediction," *Measurement*, vol. 191, Jan. 2022, Art. no. 110817.
[29] R. Fang, P. Dong, H. Ge, J. Fu, Z. Li, and J. Zhang, "Capacity plunge of lithium-ion batteries induced by electrolyte drying-out: Experimental and modeling study," *J. Energy Storage*, vol. 42, Jul. 2021, Art. no. 103013.

Yiqing Lu (Student Member, IEEE) received the B.S. degree in information engineering from Zhejiang University, Hangzhou, China, in 2020. He is currently working toward the Ph.D. degree in electrical engineering with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China.

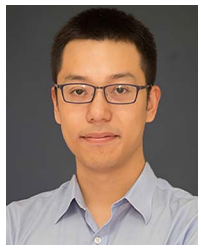
His research interests include AI-enabled battery modeling and management.

Ye Shi (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Technology Sydney, Broadway, NSW, Australia, in 2018.

From 2017 to 2019, he was a Research Assistant with the University of New South Wales, Sydney, NSW, Australia, and from 2019 to 2020, a Postdoctoral Fellow with the University of Technology Sydney. Since January 2021, he has been an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China. His research interests include optimization algorithms for artificial intelligence, machine learning, and smart grid.

Yu Liu (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical power engineering from Shanghai Jiao Tong University, Shanghai, China, in 2011 and 2013, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2017.

He is currently a tenured Associate Professor with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China. He has authored or coauthored two book chapters and more than 110 peer-reviewed journal and conference papers. His research interests include modeling, protection, fault location, and state/parameter estimation of power systems and power electronic systems.



Haoyu Wang (Senior Member, IEEE) received the B.S. degree with distinguished honors from Zhejiang University, Hangzhou, China, in 2009, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2014, both in electrical engineering.

In 2023, he was a Visiting Academic Fellow with the University of Cambridge, Cambridge, U.K. In 2014, he joined the School of Information Science and Technology, ShanghaiTech University, Shanghai, China, where he is currently a Full Professor. His research interests include power electronics, plug-in electric and hybrid electric vehicles, data-driven battery modeling, renewable energy harvesting, and power management integrated circuits.

Dr. Wang is an Associate Editor for IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE TRANSACTIONS ON TRANSPORTATION ELECTRIFICATION, and *CPSS Transactions on Power Electronics and Applications*. He is also a Guest Associate Editor for IEEE TRANSACTIONS ON POWER ELECTRONICS.