A Novel Remaining Useful Life Prediction Method for Capacity Diving Lithium-Ion Batteries

Kaidi Gao, Jingyun Xu, Zuxin Li, Zhiduan Cai, Dongming Jiang, and Aigang Zeng

ABSTRACT: To be prepared for the capacity diving phenomena in future capacity deterioration, a hybrid method for predicting the remaining useful life (RUL) of lithium-ion batteries (LIBs) is proposed. First, a novel empirical degradation model is proposed in this paper to improve the generalization applicability and accuracy of the algorithm. A particle filter (PF) algorithm is then implemented to generate the original error series using prognostic results. Next, a discrete wavelet transform (DWT) algorithm is designed to decompose and reconstruct the original error series to improve the data validity by reducing the local noise distribution information. A relatively less approximate component is selected as the reconstructed error series, which preserves the primary evolutionary information. Finally, to make full use of the information contained in the PF algorithm’s prognosis results, the support vector regression (SVR) algorithm is utilized to correct the PF prognosis results. The results indicate that long–short-term deterioration progress and RUL prediction tasks can both benefit from significant performance improvements.

1. INTRODUCTION

At present, lithium-ion batteries (LIBs) are widely used in various fields, and constitute an essential green energy source. However, the complex chemistry material and operating environment of LIBs lead to an increase in the internal resistance and a drop in the capacity and power of the cell during system operation.1,2 Because the electrical performance, safety, and stability of LIBs vary with remaining useful life (RUL), one of the most critical issues is to determine the RUL.3,4 RUL can be defined as the length of time from the current to the end of the useful life.5 Therefore, accurate and reliable RUL prediction techniques can improve battery management techniques to extend battery life and avoid the loss caused by accidental battery failure, which is no doubt of great importance.6 Extensive studies in this field have indicated that RUL prediction methods can be classified into approximately three categories: model-based methods, data-driven methods, and hybrid methods.

With respect to the model-based methods, these include establishing electrochemical models,7,8 equivalent circuit models,9,10 and empirical degradation models11,12 combined with filtering techniques such as extended Kalman filter (EKF),13 unscented Kalman filter (UKF), particle filter (PF),14–17 and unscented particle filter (UPF)18 to analyze the complex degradation processes and failure mechanisms inside the LIBs. In particular, the electrochemical and equivalent circuit models are computationally complex and difficult to measure, and the modeling process is quite complicated. However, the empirical degradation model relies on the historical degradation data to construct an accurate dynamic degradation model that, combined with the filtering technique, can accurately reflect the trend of LIBs state parameters over time and the recurrence relationship between two adjacent observed variables. The empirical degradation model has the advantages of many cycle-life models, less computational effort, fewer parameters, and higher fitting accuracy than the other two models.

Relative to the data-driven methods, the mining of information contained in degraded data can be flexibly implemented through numerous data analysis methods without obtaining any information about the internal mechanism of the LIBs. This method includes signal processing (empirical modal decomposition (EMD),19 and wavelet transform (WT)20), machine learning (support vector machines (SVM),21–23 related vector machines (RVM),24 artificial neural networks (ANN),25 long–short-term memory network (LSTM)26), time series (autoregression (AR)27 and autoregressive integrated moving average28), and statistical analysis methods (Gauss process...
regression)\textsuperscript{29} etc. Nuhic et al.\textsuperscript{1} used support vector machine (SVM) embedding to diagnose and predict RUL. Wang et al.\textsuperscript{23} proposed an RUL method for predicting LIBs using artificial bee colony (ABC) optimized support vector regression (SVR) parameters. Although data-driven methods have been widely used in battery RUL prediction, these methods require a large amount of recorded data to predict the trend of battery degradation and computational cost, and there are still some research gaps to be addressed.\textsuperscript{30} However, SVR can overcome the drawbacks of small samples, large computational size, and excellent generalization capability.

The model-based method obtains good performance only for short-term RUL predictions. If the prediction cycle becomes longer, the uncertainty worsens due to the diving capacity of LIBs. In contrast, the data-driven method can improve this problem by fully exploiting the information in the original error series and bringing into play the various advantages of both ways. Therefore, hybrid methods for the prediction of RUL of LIBs are proposed. Li et al.\textsuperscript{31} proposed a creative integrated model with the residual correction to maximize the use of residual information: a labeling algorithm for randomly sampling the UKF prediction error set to obtain some error subsets and construct an integrated model. Chang et al.\textsuperscript{32} proposed a new hybrid method based on error correction to predict the RUL of LIBs, which incorporates the algorithms of UKF, CEEMD, and RVM. Cong et al.\textsuperscript{33} proposed an improved empirical degradation model that reduces the parameter computation of the state space model and then used the CEEMD reconstructed UPF to predict the error series, and finally GPR to expect the future error series.

Concerning our specific research question, the current capacity is 80% or 90% of the initial capacitance, leading to faster capacity degradation. A prerequisite for RUL prediction of LIBs based on the hybrid methods should be that the empirical degradation model is guaranteed in terms of accuracy, explainability, and applicability. Saha et al.\textsuperscript{34} found that PF has superior accuracy than EKF, RVM, and SVM and generated a time-varying probability distribution that best encapsulates the uncertainty inherent in the system model. Xing et al.\textsuperscript{35} validated and analyzed the performance of two different models, showing the polynomial model’s weakness and the exponential model’s better predictive performance.

Besides, since less information relating to capacity dives is included in the early cycles of monitoring, the model-based and data-driven parts must be specifically selected and integrated. There is no doubt that it is unreasonable to train early data using only data-driven methods. Considering the specific issues and challenging aspects mentioned above, the data-driven part of the hybrid method can overcome these problems. The SVR algorithm,\textsuperscript{22} a typical data-driven method in the estimation of SOH and RUL, has the advantage of handling nonlinear systems with small samples. It uses inner product kernel functions for nonlinear mapping of high-dimensional spaces, superior generalization, and computational speed. A large number of studies have recognized the effective performance and applicability of SVR in hybrid methods.\textsuperscript{25,26,32,33} PF is a probabilistic statistical algorithm that estimates the discriminant parameters by calculating the sample mean of a collection of particles. It does not require knowledge of the noise model of the system to estimate data disturbed by task from noise; it can also handle linear or nonlinear systems. The RUL prediction process involves solving problems based on nonlinear, non-Gaussian systems, which has led to the widespread use of particle filtering.\textsuperscript{14–17} PF is used to estimate and adjust the model parameters to track the battery degradation process with nonlinear and non-Gaussian properties. In addition, the predictions of RUL are in a narrow probability distribution. PF has the ability to express the uncertainty of each cycle,\textsuperscript{38–40} which in combination with SVR corrects for future errors, so that the hybrid approach plays a crucial role. However, the data preprocessing technique DWT is essential to make the SVR converge; without preprocessing, a large number of training attempts cannot produce a converged SVR. DWT is able to retain the original features of the original signal while extracting different refined features for each resolution; that is, the nonstationary signal can be approximated and detailed decomposition can occur by scale function and wavelet function, respectively. DWT not only overcomes the shortcomings of the Fourier transform but also has good localization characteristics in the time domain and frequency domain, which is one of the most widely used in signal processing and becoming a powerful tool for analyzing nonstationary signals.\textsuperscript{41,42}

After the model-based part identified the model parameters, the most appropriate empirical model was explored using least-squares fitting performances. The SVR algorithm based on VC dimensional theory uses the particle swarm optimization algorithm to optimize the model parameters. The hyperparameters of a properly optimized SVR for the particle swarm algorithm can compress thousands of training data to a decisive number of support vectors. PF based on sequential Monte Carlo methods and Bayesian inference can effectively describe the uncertainties in the prediction results.

To sum up, the work in this paper focuses on two aspects. First, an empirical capacity degradation model with segmentation is proposed with the widely popular traditional exponential model. The performance evaluation of this model is verified to be better than the exponential model and another improved model for the later accelerated degradation stages. Besides, an
adaptive inflection point determination method is proposed to help the apparent degradation model predict the diving trend.

Second, a hybrid method based on the RUL prediction error correction for LIBs is proposed, which combines the PF algorithm, SVR algorithm, and DWT algorithm, and is implemented in three phases in the following steps. The simplified framework for RUL prediction proposed in this paper is shown in Figure 1.

In the first phase, the LIBs historical data are preprocessed and analyzed, and different empirical degradation models are experimentally studied. Second, an adaptive inflection point determination method is proposed, and finally, the PF algorithm is used to obtain the prediction results, and the original error series is accepted as a byproduct.

In the next phase, the DWT algorithm is used to decompose and reconstruct the error series, and through statistical analysis, the approximate signal with high coefficient as the reconstructed series, that is, provide training data for the next stage of the prediction error.

The SVR algorithm predicts the prognostic error in the last phase, correcting the PF-based predictive result and obtaining the final RUL prediction and probability density distribution.

The remainder of the paper is organized as follows: Section 2 presents the experimental validation data set and proposes an analytical verified performance expression for the empirical capacity degradation model with a capacity degradation phenomenon. Sections 3 to 5 illustrate the detailed algorithm for the proposed three-phase hybrid method. Experimental analysis is performed in section 6, and conclusions are discussed in section 7.

2. CAPACITY DEGRADATION MODEL

2.1. LIBs Capacity Degradation Data. In this paper, the experimental validation data set was derived from a Stanford University study that used APR18650M1A batteries manufactured by A123 Systems, which are widely used in RUL prediction studies. In particular, the data set was divided into three batches (2017–05–12; 2017–06–30; 2018–04–12), each with approximately 48 cells. Four batteries (lithium-ion phosphate (LFP)/graphite, nominal capacity of 1.1 Ah, nominal voltage of 3.3 V) were selected and labeled as A01–A04. All batteries were charged using a one-step or two-step fast charging protocol (C1(Q1)–C2, where C1 and C2 are the first and second constant current steps, respectively, and Q1 is the state of charge (SOC) at the switching current when 80% SOC when charging with 1C–CV mode, cutoff voltage 3.6 V). All batteries are discharged with a 4C constant current, and a cutoff voltage of 2 V as detailed in Table 1. Four lithium batteries are tested at the same temperature (30 °C).

 Apparently, the LIBs’ capacity decreases slowly in the early stages, and after about 400–500 cycles, the power starts to dive, as shown in Figure 2a. After some cycles, the LIBs’ capacity does not merely drop one way but shows a rebound. Figure 2b shows the smoothing results of the Savitzky-Golay filter for the four LIBs’ capacities.

Table 1. Four A123 LIBs Charging and Discharging Protocol Details

| battery | charge policy readable | barcode          | channel | cycles | ambient temp |
|---------|------------------------|------------------|---------|--------|--------------|
| A01     | 6C(60%)-3C             | EL150800460640   | 29      | 731    | 30 °C        |
| A02     | 6C(60%)-3C             | EL150800460436   | 30      | 757    | 30 °C        |
| A03     | 7C(40%)-3C             | EL150800460601   | 38      | 648    | 30 °C        |
| A04     | 7C(30%)-3.6C           | EL150800460622   | 40      | 703    | 30 °C        |

2.2. Autoregressive Integrated Moving Average Model. The autoregressive integrated moving average (ARIMA) model can overcome the problem of weak resistance of the empirical degradation model to the external environment and improve the real-time performance of the model. Next, the paper will introduce the empirical degradation model of segmented ideas and the adaptive method of inflection point establishment.

The system state equations and observation equations need to be constructed to complete PF for RUL prediction of LIBs first. In the paper, two different model transitions are designed to describe the system well and are easy to build to meet the practice’s real-time and flexibility requirements. In the referenced literature, the double exponential model is widely used in the RUL prediction of LIBs due to its fluency, smoothness, and simplicity. It is expressed as

\[ Q(k) = ae^{-b}k + ce^{-d}k \]  

where \( k \) is the number of charge/discharge cycles, \( Q(k) \) is the battery capacity at the \( k \)th time, and \( a, b, c, \) and \( d \) are constants that vary with time (\( a \) and \( b \) are related to the internal impedance of the battery, \( c \) and \( d \) are aging parameters).

In this regard, the exponential combined polynomial models and single exponential models were excluded due to the difficulty of achieving a balance between the number of parameters and the observation error. The literature mentions that an improved double exponential model significantly improves the late sharp degradation by subtracting nonpositive expressions, denoted as

\[ Q(k) = e^{-a+b}k - e^{-c+d}k \]  

where \( e^{-a+b}k \) and \( -e^{-c+d}k \) subscales represent the initial capacity degradation and accelerated capacity degradation processes. The latter degradation is updated iteratively by the nonpositive expression \( -e^{-c+d}k \). It is of obvious advantage that the capacity of the late cycle decreases by subtracting an exponential function.

The ARIMA model can observe the current moment based on the historical moment. The ARIMA model is suitable for real-time state modeling and can be used as the measured model of the PF model. The ARIMA model consists of three components: (1) an autoregressive (AR) process; (2) establishing the stationary time series by differential order; (3) a moving average (MA) process. The following is the general form of ARIMA \((p, d, q)\)

\[ Q(k) = \sum_{j=1}^{p} \theta_j Q(k-j) + \sum_{i=1}^{q} \phi_i Q(k-i) + \epsilon_k \]  

where the \( p \) measured backward value \( Q(k-j) \) and the \( \epsilon_k \) modeling error were used to get the present measured value \( Q(k) \); \( r_j \) and \( \theta_i \) denote the parameters of the observed values at \( j \) moments and the parameters of the error term at \( i \) moments, respectively; \( q \) represents the number of moving error terms.

The autocorrelation function (ACF) and partial autocorrelation function (PACF) check for stationary series and whether
the AR and MA components should be included. If not smooth, then they are checked differentially until stationary. This method determines the $p$, $d$, and $q$ by competent observation. To avoid subjectivity, Akaike Information Criterion (AIC)\textsuperscript{27,38} would be used.
and Bayesian Information Criterion (BIC)\textsuperscript{27,40} are introduced to determine the model order. The AIC and BIC values are calculated after an ARIMA model with different orders is constructed for a given time series in this paper. The most appropriate model for the time series is the smallest AIC and BIC values.

\[
AIC = 2m - 2 \ln(L) \tag{4}
\]

\[
BIC = \ln(n)^* m - 2 \ln(L) \tag{5}
\]

where \(m\), \(n\), and \(L\) denote the number of model parameters, the number of samples, and the likelihood function, respectively.

The results of the models (1) and (2) performance fits are shown in Table 2, and it is clear that model (2) outperforms model (1) in both SSE and RMSE throughout the degradation, as well as \(R^2\) is closer to 1. However, model (1) performs even better when the field of view is placed in the predegradation period. Model (1) and model (3) proposed in this paper are feasible as early and late model schemes, respectively. Determination of the true capacity dive points for cells A01, A02, A03, and A04 at points 518 cycles, 521 cycles, 459 cycle, and 491 cycles, respectively, is made by paralleling the linkage line manual set with the initial test capacity and the failure threshold capacity points shown in the A01 battery in Figure 2c. As in Figure 2d, there is a particular error in determining the inflection point based on the \(\sigma\) standard interval of the capacity difference curve. Taking cell A01 as an example, to prevent overfitting from occurring, results were qualitatively determined by the values of 2 and 4 for AIC and BIC, respectively, in Table 3.

In the paper, the proposed adaptive approach to determine the inflection point is to identify the \(\sigma\) interval criterion using the differential capacity curve.\textsuperscript{44} Moreover, the current capacity is less than 90% of the initial capacity introduced as one of the conditions. However, it is worth noting that the inflection points found by the adaptation are not the same as the true capacity dive points shown in Figure 2c,d. Then, the inflection point determined in the above way can complete the switching of the model. In the case of the segmented model discussed in the paper, the inflection point is often found before the true capacity dive point, which makes it advantageous to use a small amount of information to correct the ARIMA model parameters before the capacity dive appears.

In review, it can be proven that the segmentation idea scheme proposed in this thesis accurately describes the different degradation stages and improves the empirical degradation model with insufficient real-time,\textsuperscript{41,42} reduces the estimation of unnecessary parameters, and dramatically reduces the computational effort.

3. PF-BASED PROGNOSTIC RESULTS

Accurate prediction of RUL requires an accurate degeneracy model and depends on the tuning of the particle state parameters. The two different degradation models are used for LIBs that show diving phenomena to reduce the transformation requirements of parameters at the inflection point.

The PF algorithm is a Bayesian filtering algorithm based on the Monte Carlo method, which can handle any nonlinear, non-Gaussian problem.

\[
\begin{aligned}
\mathbf{x}_k &= f(\mathbf{x}_{k-1}) + u_k \\
\mathbf{z}_k &= h(\mathbf{x}_k) + v_k
\end{aligned} \tag{6}
\]

where \(x_k\) represents the system state at the moment \(k\), \(u_{k-1}\) represents the process noise at the moment \(k - 1\), and \(f(\cdot)\) is a linear or nonlinear function, establishing the relationship between the current state and the state at the last moment, \(z_k\) represents the observed value at the time \(k\), \(v_k\) is the measurement noise, and \(h(\cdot)\) is a linear or nonlinear function, which establishes the relationship between the state value and the measured value at the same time. \(p(x_k | x_{k-1})\) is the prior probability of the state equation and \(p(z_k | x_k)\) represents the observation distribution as a likelihood function. The proposed two different empirical capacity degradation models eqs 1 and 3 can be used as measurement in eq 6. The process of particle filtering algorithm involves the following steps:

- Step 0: Initialization. \(N\) particles \(x_0^i\) are generated by sampling in the prior probability \(p(x_0), w_0^i = 1/N, i = 1, ..., N\).
- Step 1: Importance sampling. \(i = 1, ..., N\) sampling \(x_k^i \sim q(x_k | x_{k-1}, z_k)\).

\[
w_k^i = \frac{p(z_k^i | x_k^i)p(x_k^i | x_{k-1}^i)}{\sum_{i=1}^{N} w_k^i} \tag{7}
\]

- Step 2: Weight calculation.

\[
w_k^i = \frac{w_k^i}{\sum_{i=1}^{N} w_k^i} \tag{8}
\]

- Step 3: Resampling. Calculate the number of valid samples.

\[
\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N} (w_k^i)^2} \tag{9}
\]

If \(\hat{N}_{eff} < N_{eff}\), \(\{x_k^i, w_k^i\}\) generated based on importance weights \(w_k^i\) by residual resampling method.
Figure 3. (a) Complete iterative process of the model (1). (b) Iterative process parts of the model (1). (c) Iterative process parts of the model (3). (d) Iterative process parts of the model (2).
Step 4: State estimation.

\[ \hat{x}_k \approx \sum_{i=1}^{N} w_i x_{i(k)} \]  

(10)

The following implementation of the PF method using A01 battery is an example:

1. The tangent to the A01 cell curve by a linkage line set manually parallel to the initial capacity of the cell and the end of life (EOL) point to 80% is being identified as the true capacity dive point, i.e., the 518th cycle.

2. An adaptive approach to determining the inflection point is proposed. It is the 470th point determined by the A01 cell using the \( \mu \pm 3\sigma \) interval criterion of the differential capacity decline curve, and the current capacity is less than 90% of the initial capacity. The segmented empirical degradation model early \( \alpha e^{-k} + \beta e^{k} \) indicates a slowly decreasing trend and later \( \sum_{j=1}^{p} r_j Q(k-j) + \epsilon_k \) indicates a diving phase.

3) The improved PF with 499 cycles after the termination of the iterative process as training data confirms that the improved PF can accurately identify the dive before it appears, thus effectively identifying the root-mean-square error in predicting subsequent EOL points of about 1%. The required training data for cells A02, A03, and A04 were determined similarly for 465, 472, and 466 cycles, respectively.

However, since the iterative process of the single model in the later parameters implies that subsequent fluctuations significantly slow down the parameter variation, it is inevitable that the final estimated parameters have hysteresis or overshoot in the single model. The segmentation model proposed in this paper can improve the problem of slow parameter fluctuations and changes in the single index model at different stages and also has excellent predictive trends in the late diving stage shown in Figure 3.

Figure 3. Besides, the training data set needs to be fully utilized for this and the error series can predict the possible range of subsequent data, that is, the real data minus the prognosis results. For this reason, the future evolution of the errors is predicted by using a data-driven approach. The related details are described in the following sections.

4. IMPROVE TRAINING DATA USING DWT

The WT method is able to analyze and multilayer decompose the signal at different scales and extract it into approximate components through a finite decomposition and reconstruction process, resulting in a smoothed residual series as the training data for prediction.

Here the DWT denoising process, that is, the Mallat algorithm is briefly explained.\textsuperscript{52} The original signal is passed through two complementary filters in the first decomposition layer to produce the approximate and detailed components. The second decomposition layer decomposes the approximate and detailed components based on the approximate composition of the first decomposition layer, and so on. The more decomposition layers there are, the less the approximate

![Low-frequency and high-frequency signals of the original capacity error series by the DWT.](image)

**Figure 4.**

### Table 4. DWT Results of A01 Battery

| signal | low-frequency | high-frequency |
|--------|---------------|----------------|
|        | Var (\(\cdot e^{-x}\)) | \(\rho\) | Var (\(\cdot e^{-x}\)) | \(\rho\) |
| 1      | 8.3069        | 0.9824        | 3.0275          | 0.2004  |
| 2      | 8.1364        | 0.9608        | 3.1227          | 0.1622  |
| 3      | 7.8828        | 0.9601        | 1.4097          | 0.1202  |
| 4      | 7.7476        | 0.9467        | 1.2592          | 0.1702  |
| 5      | 7.6430        | 0.9409        | 0.8188          | 0.1192  |
| 6      | 7.3933        | 0.9276        | 2.1889          | 0.2866  |
| 7      | 6.8431        | 0.9129        | 2.1889          | 0.2866  |
| original series | 8.6663        |                |                 |         |
Figure 5. SVR prediction error results of the four battery.
component obtained, and the smaller the variance of the reconstructed signal.

To reduce the computational drawbacks that appear in the WT, the DWT method is similar to the fast Fourier transform in Fourier variation to reduce information redundancy by constructing an orthogonal wavelet method to obtain more time-frequency information.

The signal correlation is assessed using the eq 11 Pearson’s correlation coefficient $\rho$ (value between $-1$ and $1$) and the variance. A low correlation is stated as a value near 0 when evaluating the correlation between the decomposed signal and the original signal.

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \quad (11)$$

where $\sigma_x$ and $\sigma_y$ are denoted as the standard deviation of the signals. The covariance $\text{cov}(X, Y)$ represents the error of the overall two signals.

Since the high-frequency components are considered to be insignificant for tendencies of error series evolution, the reconstructed error series could be made up of low components with the dominant information on error evolution shown in Figure 4. Therefore, the low-frequency component reconstructed signals in the sixth layer (the original signal minus the high-frequency component signals in the first six layers) are selected as the long-term trend of error evolution. The correlations of the low-frequency and high-frequency signals are shown in Table 4.

The data quality is enhanced by deleting some low-correlation high-frequency signals from the residual components. It can be seen that the trend features of the global fluctuations of the residual components and reconstructed low-frequency signals are contained in a smoother, less noisy error series. Thus, it was used in the data-driven algorithm, which laid a solid foundation for the successful predictions of support vector regression that followed.

5. Predicting Future Errors Using SVR

As a nonlinear regression for VC dimensional theory and minimizing structural risk, SVR implements regression with statistical learning theory.\textsuperscript{16,23} The SVR method uses a nonlinear relation to map the sample set to a high-dimensional space and train the input and output relationships of the sample set, which can be expressed as

$$E(x) = \omega \phi(x) + b \quad (12)$$

where $E(x)$ denotes the capacity of error, which is the corresponding output, and $x$, $\omega$, $b$ represent the input data (the feature vector of the sample), weight, and intercept, respectively. $\phi(x)$ is the high dimensional feature space. The slack variables $\{\xi_k\}_{k=1}^N$ and the penalty factor $C$ are introduced to solve eq 13 by translation to optimization problems as follows:

$$\min R(\omega, \xi_k) = \frac{1}{2} \omega^T \omega + C \sum_{k=1}^N (\xi_k + \bar{\xi}) \quad (13)$$

Figure 6. A detailed flowchart of the proposed hybrid method.
Figure 7. Algorithm prediction results of (a) A01, (b) A02, (c) A03, and (d) A04 batteries.
\[ E(x_k) - \omega \phi(x_k) - b \leq \epsilon + \xi_k \]
\[
\text{s.t.} - E(x_k) + \omega \phi(x_k) + b \leq \epsilon + \xi_k
\]
\[
\xi_k \geq 0; \quad \xi_k \geq 0 \quad k = 1, 2, \ldots, N
\]

where \( \epsilon (\epsilon > 0) \) is the insensitivity loss coefficient. The Lagrange multiplier approach and the Karush-Kuhn-Tucker condition are introduced and translated to the dual form to address the problem of constrained optimization:

\[
\max R(\alpha^*_k, \alpha_k) = \frac{1}{2} \sum_{k=1}^{N} (\alpha^*_k - \alpha_k)(\alpha^*_k - \alpha_k)\phi(x_k)\phi(x_k)
\]
\[
- \epsilon \sum_{k=1}^{N} (\alpha_k + \alpha^*_k) + E(x_k) \sum_{j=1}^{N} (\alpha^*_j - \alpha_j)
\]

\[
\text{s.t.} \sum_{k=1}^{N} (\alpha_k - \alpha^*_k) = 0
\]
\[
0 \leq \alpha_k, \quad \alpha^*_k \leq C, \quad k = 1, 2, \ldots, N
\]

where \( \alpha_k \) and \( \alpha^*_k \) are Lagrange multipliers, and finally solved by a convex optimization problem to obtain \( \omega \) and \( b \):

\[
\omega = \sum_{k=1}^{N} (\alpha^*_k - \alpha_k)\phi(x_k)
\]

\[
b^* = \frac{1}{N_n} \sum_{i,j} \left[ E(x_i) - \sum_{i=1}^{N} (\alpha^*_k - \alpha_k)K(x_i, x_j) - c \right] + \frac{1}{N_n} \sum_{i,j} \left[ E(x_i) - \sum_{j=1}^{N} (\alpha^*_j - \alpha_j)K(x_i, x_j) + c \right]
\]

where \( N_n \) indicates the number of support vectors and \( K(x_i, x_j) = \phi(x_i)\phi(x_j) \) is the kernel function. Finally, the hyperplane can be expressed as

\[ E(x) = \omega \phi(x) + b^* \]

One of the most popular kernel functions in machine learning is the radial basis kernel function, which can be indicated as

\[ K_{\text{RBF}}(x, x_i) = \exp(-\gamma ||x - x_i||^2) \gamma > 0 \]

where \( \gamma \) is the hyperparameter to get the best balance between training models and generalization ability. The other hyperparameter \( C \) is a balance between the complexity of the support vector and the misclassification rate. The two hyperparameters are determined by the particle swarm algorithm that uses a set of particle swarms moving in the search space. The optimal particles are obtained by information interaction between individuals. The reconstructed error series is used as the training data set to build the SVR prediction model, and the prediction curves are shown in Figure 5.

Finally, the SVR algorithm implements the data-driven part of the hybrid method to generate the predicted error series. The effective and comprehensive utilization of training information is realized, including cycle capacity analysis, the establishment of a segmented model of capacity diving, model-based prognostic errors, and data-driven correction of prognostic results.

### 6. RESULTS AND DISCUSSION

The whole RUL prediction will be finalized by solving eqs 1 and 2. Once the inflection point, or the true capacity diving point, is reached, the model is updated to improve prediction accuracy, as shown in Figure 6. This study utilizes 80% of LIB capacity as EOL data for experiments and validation. The projected values were compared to genuine experimental data to assess the method’s performance, as shown in Figure 7.

The rapid rate of a terminal recession leads to slight differences between the number of cycles in the RUL prediction results. It is difficult to use the number of cycles to describe the error clearly. Therefore, the error and other effective regression indicators in the predicted capacity difference are rationally discussed to evaluate the prediction performance.

Reviewing section 2, the verification analysis of three models has been completed. It is obvious that model (1) is superior to model (2) in the early degradation stage of LIBs. In comparison to prior hybrid approaches, the hybrid method suggested in this study significantly reduces computation time while preserving prediction and RUL accuracy, as shown in Table 5. The prediction performance of the hybrid algorithm being better than that of the other algorithm is reflected quantitatively by the evaluation metrics of the four cells. These would be strong evidence validating the ability of the method to predict the accelerated capacity decay of the battery under these operating conditions.

In this case, the prediction starting point is 219 cycles. The RUL prediction results of four batteries is shown in Figure 7. The RUL predicted by A01, A02, A03, and A04 is 515, 532, 430, and 479, respectively. In addition, assuming that the prediction results conform to Gaussian distribution, the variances are 18.5, 11.4, 10.2, and 10.4, respectively. Through the statistical regression analysis of the prognosis results, the effectiveness of the hybrid method proposed in this paper including root-mean-square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and Max abs error is proven. However, it is worth noting that, after using the hybrid prognostic technique, the specific capacity decline curves with error correction become one of the critical areas of interest.

### Table 5. Prediction Methods Performance Evaluation Results of the Four Batteries

| Battery | Method                  | RMSE (\(\epsilon^2\)) | MAE  | MAPE | Max abs error (\(\epsilon^2\)) | Time (h) | Variance (\(\epsilon^2\)) | EOL Data | MRUL  |
|---------|-------------------------|------------------------|------|------|-------------------------------|----------|---------------------------|----------|-------|
| A01     | PF+CEEMD+GPR            | 0.85                   | 0.61 | 0.62 | 1.93                          | 25.60    | 19.2                      | 518      | 509   |
|         | PF+DWT+SVR              | 0.61                   | 0.41 | 0.43 | 1.48                          | 2.9      | 18.5                      | 515      | 509   |
| A02     | PF+CEEMD+GPR            | 0.28                   | 0.22 | 0.21 | 1.18                          | 21.17    | 11.9                      | 528      | 534   |
|         | PF+DWT+SVR              | 0.35                   | 0.22 | 0.22 | 0.92                          | 3.59     | 11.4                      | 532      | 534   |
| A03     | PF+CEEMD+GPR            | 0.95                   | 0.78 | 0.79 | 1.90                          | 36.66    | 12.5                      | 429      | 430   |
|         | PF+DWT+SVR              | 1.02                   | 0.84 | 0.85 | 1.96                          | 4.81     | 10.2                      | 430      | 430   |
| A04     | PF+CEEMD+GPR            | 0.31                   | 0.23 | 0.23 | 0.79                          | 45.48    | 10.8                      | 482      | 483   |
|         | PF+DWT+SVR              | 0.26                   | 0.21 | 0.21 | 0.78                          | 5.40     | 10.4                      | 479      | 483   |
Figure 8. Predicted results of A01, A02, A03, and A04 batteries by three models.
Table 6. Different Models Predict Performance Evaluation Results of the Four Batteries

| data | model | RMSE($\epsilon$) | MAE | MAPE | Max abs error($\epsilon$) | variance | PRUL | MRUL |
|------|-------|------------------|-----|------|---------------------------|----------|------|------|
| A01  | (1)   | 0.95             | 0.20| 0.21 | 1.61                      | 20.9     | 234  | 229  |
|      | (2)   | 0.23             | 0.20| 0.21 | 0.67                      | 12.1     | 227  | 229  |
|      | (3)   | 0.13             | 0.13| 0.13 | 0.31                      | 2.2      | 231  | 229  |
| A02  | (1)   | 0.36             | 0.21| 0.22 | 0.76                      | 11.2     | 286  | 288  |
|      | (2)   | 0.25             | 0.21| 0.22 | 0.58                      | 7.7      | 283  | 288  |
|      | (3)   | 0.15             | 0.13| 0.14 | 0.28                      | 1.4      | 290  | 288  |
| A03  | (1)   | 0.46             | 0.15| 0.15 | 1.55                      | 9.8      | 165  | 173  |
|      | (2)   | 0.17             | 0.15| 0.15 | 0.48                      | 6.1      | 170  | 173  |
|      | (3)   | 0.13             | 0.11| 0.12 | 0.22                      | 1.1      | 174  | 173  |
| A04  | (1)   | 0.41             | 0.22| 0.23 | 0.75                      | 11.4     | 227  | 232  |
|      | (2)   | 0.26             | 0.22| 0.23 | 0.64                      | 10.9     | 228  | 232  |
|      | (3)   | 0.25             | 0.20| 0.20 | 0.43                      | 1.5      | 232  | 232  |

The hybrid method reduces computational costs, and since only a single run of the technique is required to capture the overall capacity decline trend accurately, there is no need to repeat the model-based portion of the run for subsequent cycles.

The training data set has been explained respectively in section 2 and section 3. The predictions are based on A01 battery cycles of 499 to 731, A02 battery cycles of 465 to 757, A03 battery cycles of 472 to 648, and A04 battery cycles of 466 to 703, respectively. Then, in this case, the performance of the three different models in the late diving stage is shown in Figure 8. Model (3) predicted RULs of A01, A02, A03, and A04 batteries are 231, 290, 174, and 232. The distribution of the expected EOL is assumed to be Gaussian, with variances of 2.2, 1.4, 1.1, and 1.5 for the four cells, respectively. The prognostic results of the metrics assessed by statistical conventional models are shown in Table 6. All the reduced evaluation metrics can prove the good predictive performance of the ARIMA model. A slight variance means that the uncertainty expression has high precision and a small range of predictive possibilities.

7. CONCLUSION

A model-data hybrid method for RUL prediction of LIBs based on error correction is proposed. Based on the adaptive segmented empirical degradation model, a hybrid method for lithium-ion batteries RUL prediction with the capacity diving phenomenon is presented, using the PF algorithm as the model part. Then the DWT error series is decomposed into the data-driven part based on the SVR algorithm to predict the error series laid successfully. On the one hand, to overcome the problem of the poor real-time performance of the empirical degradation model, a segmented model was proposed to ensure better performance in both the early and late degradation stages. On the other hand, a data-driven component was introduced to correct the predictive error further. The hybrid method enables a comprehensive and effective utilization of training information to be achieved, including cycling capacity degradation data, building segmented empirical capacity degradation models with capacity diving, and model-based PF prognostic errors. The RUL prediction framework proposed by this hybrid method can guarantee accurate RUL prediction results.

AUTHOR INFORMATION

Corresponding Author

Jingyun Xu — School of Engineering, Huzhou University, Huzhou City 316007, China; orcid.org/0000-0002-2779-

6059; Phone: +0572)13511231020; Email: xjy@zjhu.edu.cn

Authors

Kaidi Gao — School of Engineering, Huzhou University, Huzhou City 316007, China
Zuxin Li — Institute of Technology, Huzhou College, Huzhou City 313000, China; orcid.org/0000-0002-5816-9728
Zhiduan Cai — Institute of Technology, Huzhou College, Huzhou City 313000, China
Dongming Jiang — School of Engineering, Huzhou University, Huzhou City 316007, China
Aigang Zeng — Zhejiang Tianheng New Materials Co, Huzhou City 313103, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c03043

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by the Basic Public Welfare Research Program of Zhejiang Province, China, under Grant (LGG22F030023, LGG21F010002), Huzhou public welfare applied research project (2021GZ11), and in part by Huzhou University Graduate Student Research Innovation Project Subjects (2022KYCX55).

REFERENCES

(1) Nuhic, A.; Terzimehic, T.; Soczka-Guth, T.; Buchholz, M.; Dietmayer, K. Health diagnosis and remaining useful life prognostics of LIBs using data-driven methods. J. Power Sources 2013, 239, 680–688.
(2) Wu, J.; Zhang, C.; Chen, Z. An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks. Appl. Energy 2016, 173, 134−140.
(3) Barré, A.; Deguilhem, B.; Grolleau, S.; Gérard, M.; Suard, F.; Riu, D. A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. J. Power Sources 2013, 241, 680−689.
(4) Dong, H.; Jin, X.; Lou, Y.; Wang, C. Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter. J. Power Sources 2014, 271, 114−123.
(5) Tian, H.; Qin, P.; Li, K.; Zhao, Z. A review of the state of health for lithium-ion batteries: Research status and suggestions. J. Cleaner Prod. 2020, 261, 120813−120843.
(6) Wei, J.; Dong, G.; Chen, Z. Remaining Usefullife Prediction and State of Health Diagnosis for Lithium-Ion Batteries Using Particle Filter and Support Vector Regression. IEEE Trans. Ind. Electron. 2018, 65, S634−S643.
(7) Xiong, R.; Li, L.; Li, Z.; Yu, Q.; Mu, H. An electrochemical model based degradation state identification method of Lithium-ion battery for all-climate electric vehicles application. Appl. Energy 2018, 219, 254–275.

(8) Guha, A.; Patra, A. Online Estimation of the Electrochemical Impedance Spectrum and Remaining Useful Life of Lithium-Ion Batteries. IEEE Trans. Instrum. Meas. 2018, 67, 1836–1849.

(9) Hu, X.; Li, S.; Peng, H. A comparative study of equivalent circuit models for Li-ion batteries. J. Power Sources 2012, 198, 359–367.

(10) Jiang, L.; Li, Q.; Chen, W.; Yu, J.; Liu, J. Study on parameter identification of third-order RQ equivalent circuit of PEMFC based on Nelder-Mead optimization. J. Power Supply 2019, 17, 12.

(11) Singh, P.; Chen, C.; Tan, C. M.; Huang, S. C. Semi-Empirical Capacity Fading Model for SOH Estimation of Li-Ion Batteries. Appl. Sci. 2019, 9, 3012–3027.

(12) Wang, J.; Liu, P.; Hicks-Garnier, J.; Sherman, E.; Soukiazian, S.; Verbrugge, M.; Tatara, H.; Musser, J.; Finamore, P. Cycle-life model for graphite-LiFePO 4 cells. J. Power Sources 2011, 196, 3942–3948.

(13) Charkhgard, M.; Farrokh M. State-of-Charge Estimation for Lithium-Ion Batteries Using Neural Networks and EKF. IEEE Trans. Ind. Electron. 2010, 57, 4178–4187.

(14) Ma, Y.; Chen, Y.; Zhou, X.; Chen, H. Remaining Useful Life Prediction of Lithium-Ion Battery Based on Gaussian–Hermitte Particle Filter. IEEE Trans. Control Syst. Technol. 2019, 27, 1788–1795.

(15) Dong, G.; Chen, Z.; Wei, J.; Ling, Q. Battery Health Prognosis Using Brownian Motion Modeling and Particle Filtering. IEEE Trans. Ind. Electron. 2018, 65, 8646–8655.

(16) Dong, H.; Jin, X.; Lou, Y.; Wang, C. Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter. J. Power Sources 2014, 271, 114–123.

(17) Li, D. Z.; Wang, W.; Ismail, F. A Mutated Particle Filter Technique for System State Estimation and Battery Life Prediction. IEEE Trans. Instrum. Meas. 2014, 63, 2034–2043.

(18) Chen, L.; Chen, J.; Wang, H.; Wang, Y.; An, J.; Yang, R.; Pan, H. Remaining Useful Life Prediction of Battery Using a Novel Indicator and Framework With Fractional Grey Model and Unscented Particle Filter. IEEE Trans. Power Electron. 2020, 35, 5850–5859.

(19) Qiao, J.; Liu, X.; Chen, Z. Prediction of the Remaining Useful Life of Lithium-Ion Batteries Based on Empirical Mode Decomposition and Deep Neural Networks. IEEE Access 2020, 8, 42760–42767.

(20) Ibrahim, M.; Steiner, N. J.; Jemei, S.; Hissel, D. Wavelet-Based Approach for Online Fuel Cell Remaining Useful Lifetime Prediction. IEEE Trans. Ind. Electron. 2016, 63, 5057–5068.

(21) Xu, J.; Zhen, A.; Cai, Z.; Wang, P.; Gao, K.; Jiang, D. State of Health Diagnosis and Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Multi-Feature Data and Mechanism Fusion. IEEE Access 2021, 9, 85431–85441.

(22) Gao, D.; Huang, M. Prediction of Remaining Useful Life of Lithium-ion Battery based on Multi-kernel Support Vector Machine with Particle Swarm Optimization. J. Power Electron. 2017, 17, 1288–1297.

(23) Wang, Y.; Ni, Y.; Lu, S.; Wang, J.; Zhang, X. Remaining Useful Life Prediction of Lithium-Ion Batteries Using Support Vector Regression Optimized by Artificial Bee Colony. IEEE Trans. Veh. Technol. 2019, 68, 9543–9553.

(24) Zhang, C.; He, Y.; Yuan, L.; Xiang, S. Capacity Prognostics of Lithium-Ion Batteries using EMD Denoising and Multiple Kernel RVM. IEEE Access 2017, 5, 12061–12070.

(25) Pan, H.; Lu, Z.; Wang, H.; Wei, H.; Chen, L. Novel battery state-of-health online estimation method using multiple health indicators and an extreme learning machine. Energy 2018, 160, 466–477.

(26) Qu, J.; Liu, F.; Ma, Y.; Fan, J. A Neutral-Network-Based Method for RUL Prediction and SOH Monitoring of Lithium-Ion Battery. IEEE Access 2019, 7, 87178–87191.

(27) Long, B.; Xian, W.; Jiang, L.; Liu, Z. An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries. Microelectron. Reliab. 2013, 53, 821–831.

(28) Zhou, Y.; Huang, M. Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model. Microelectron. Reliab. 2016, 65, 265–273.

(29) Peng, Y.; Hou, Y.; Song, Y.; Pang, J.; Liu, D. Lithium-Ion Battery Prognostics with Hybrid Gaussian Process Function Regression. Energies 2018, 11, 1420–1440.

(30) Severson, K. A.; Attia, P. M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M. H.; Aykol, M.; Herring, P. K.; Fraggedakis, D. Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy 2019, 4, 2084–2089.

(31) Li, Z.; Fang, H.; Xiao, Z. A novel hybrid model based on ensemble strategy for lithium-ion battery residual life prediction. 2018 Chinese Automation Congress 2018, 2084–2089, DOI: 10.1109/ CAC.2018.8623088.

(32) Chang, Y.; Fang, H.; Zhang, Y. A new hybrid method for the prediction of the remaining useful life of a lithium-ion battery. Appl. Energy 2017, 206, 1564–1578.

(33) Cong, X.; Zhang, C.; Jiang, J.; Zhang, W.; Jiang, Y. A Hybrid Method for the Prediction of the Remaining Useful Life of Lithium-Ion Batteries With Accelerated Capacity Degradation. IEEE Trans. Veh. Technol. 2020, 69, 12775–12785.

(34) Saha, B.; Goebel, K.; Christophersen, J. Comparison of prognostic algorithms for estimating remaining useful life of batteries. Trans. Inst. Meas. Control 2009, 31, 293–308.

(35) Xing, Y.; Ma, E. W. M.; Tsui, K. L.; Pecht, M. A case study on battery life prediction using particle filtering. Proceedings of the IEEE 2012 Prognostics and System Health Management Conference 2012, 1–6, DOI: 10.1109/PHM.2012.6228847.

(36) Patil, M. A.; Tagade, P.; Harirhan, K. S.; Kolake, S. M.; Song, T.; Yeo, T.; Doo, S. A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation. Appl. Energy 2015, 159, 285–297.

(37) Hu, X.; Che, Y.; Lin, X.; Onori, S. Battery Health Prediction Using Fusion-Based Feature Selection and Machine Learning. IEEE Trans. Transp. Electrification 2021, 7, 382–398.

(38) Xing, Y.; Ma, E.; Tsui, K. L.; Pecht, M. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. Microelectron. Reliab. 2013, 53, 811–820.

(39) Zhang, L.; Ma, Z.; Sun, C. Remaining Useful Life Prediction for Lithium-Ion Batteries Based on Exponential Model and Particle Filter. IEEE Access 2018, 6, 17729–17740.

(40) Cong, X.; Zhang, C.; Jiang, J.; Zhang, W.; Jiang, Y.; Jia, X. An Improved Unscented Particle Filter Method for Remaining Useful Life Prognostic of Lithium-ion Batteries with Li(NiMnCo)O2 Cathode with Capacity Diving. IEEE Access 2020, 8, 58717–58729.

(41) Wang, Y.; Pan, R.; Yang, D.; Tang, X.; Chen, Z. Remaining Useful Life Prediction of Lithium-ion Battery Based on Discrete Wavelet Transform. Energy Procedia 2017, 105, 2053–2058.

(42) Lee, S.; Kim, J. Discrete wavelet transform-based denoising technique for advanced state-of-charge estimator of a lithium-ion battery in electric vehicles. Energy 2015, 83, 462–473.