

A novel autoregressive rainflow-integrated moving average modeling method for the accurate state of health prediction of lithium-ion batteries.


HUANG, J., WANG, S., XU, W., SHI, W. and FERNANDEZ, C.

2021

© 2021 by the authors. Licensee MDPI, Basel, Switzerland.

Article

A Novel Autoregressive Rainflow—Integrated Moving Average Modeling Method for the Accurate State of Health Prediction of Lithium-Ion Batteries

Junhan Huang ¹, Shunli Wang ^{1,*} , Wenhua Xu ¹, Weihao Shi ¹ and Carlos Fernandez ^{2,*}

- ¹ School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China; huangjunhan@mails.swust.edu.cn (J.H.); xuwenhua@mails.swust.edu.cn (W.X.); shiweihao@mails.swust.edu.cn (W.S.)
- ² School of Pharmacy and Life Sciences, Robert Gordon University, Aberdeen AB10-7GJ, UK
- * Correspondence: wangshunli@swust.edu.cn (S.W.); c.fernandez@rgu.ac.uk (C.F.); Tel.: +86-15884655563 (S.W.)

Abstract: The accurate estimation and prediction of lithium-ion battery state of health are one of the important core technologies of the battery management system, and are also the key to extending battery life. However, it is difficult to track state of health in real-time to predict and improve accuracy. This article selects the ternary lithium-ion battery as the research object. Based on the cycle method and data-driven idea, the improved rain flow counting algorithm is combined with the autoregressive integrated moving average model prediction model to propose a new prediction for the battery state of health method. Experiments are carried out with dynamic stress test and cycle conditions, and a confidence interval method is proposed to fit the error range. Compared with the actual value, the method proposed in this paper has a maximum error of 5.3160% under dynamic stress test conditions, a maximum error of 5.4517% when the state of charge of the cyclic conditions is used as a sample, and a maximum error of 0.7949% when the state of health under cyclic conditions is used as a sample.

Keywords: lithium-ion battery; state of health; rainflow; autoregressive integrated moving average model



Citation: Huang, J.; Wang, S.; Xu, W.; Shi, W.; Fernandez, C. A Novel Autoregressive Rainflow—Integrated Moving Average Modeling Method for the Accurate State of Health Prediction of Lithium-Ion Batteries. *Processes* **2021**, *9*, 795. <https://doi.org/10.3390/pr9050795>

Academic Editor: Mingxia Gao

Received: 1 April 2021
Accepted: 29 April 2021
Published: 30 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Lithium-ion batteries are widely used in agriculture, communications, industry, and other fields because of their unique advantages, such as their small size, low cost, and long life. From electronic products to electric vehicles, from airplanes and large equipment to energy storage equipment, they are all used as power and control equipment. Because of the widespread use of lithium-ion batteries, there are higher requirements for the accuracy of battery predictions. The battery management system (BMS) monitors and manages the entire process of power battery work, including estimating the state of charge (SOC), state of health (SOH), thermal management, fault diagnosis, and balance management, as well as other aspects. Accurate detection of the SOH is a reference for other aspects of the battery management system. After a long period of use, the capacity of the battery will decay. At this time, charging and discharging according to the factory capacity will cause excessive charging and discharging, which will seriously affect the life of the battery itself and the related equipment. Therefore, it is particularly important to accurately predict the health of the battery.

The estimation methods of battery SOH include empirical-based methods [1–4], model-based methods [5,6], and data-driven methods [7–10]. Empirical-based methods include the cycle number method, ampere-hour method, weighted ampere-hour method, and event-oriented aging accumulation method. Model-based methods use algorithms to estimate based on equivalent models. Data-driven methods include support vector machine (SVM) [11,12], autoregressive moving average (ARMA), particle filtering (PF) [13–15], and neural network [16], etc. The empirical-based method uses some empirical knowledge in

the process of battery use to give a rough estimate of battery life based on certain statistical laws, and can only be used for life prediction in specific situations when the empirical knowledge of the battery use is sufficient. The model in the model-based approach requires fine parameters and a high degree of complexity. Moreover, the tests for aging factors [17–19] are complicated, and there are difficulties in establishing a perfect aging mechanism model. Data-driven [20–23] prediction based on the object system does not require mechanistic knowledge of the object system, but is based on the data collected through a variety of data analysis learning methods used to mine the implicit information from the prediction, thus avoiding the complexity of model acquisition, and is a more practical prediction method. For example, Dong et al. [24] realized lithium-ion battery SOH monitoring and remaining useful life prediction [25–28] based on a support vector regression-particle filter. Chen et al. [29] proposed a new method to estimate SOH based on the second-order central difference particle filter. This method can solve the particle degeneracy phenomenon of the particle filter by optimizing the importance of the probability density function. Chen et al. [30] estimated the state of health of lithium-ion batteries based on the fusion of the autoregressive moving average model and the Elman neural network [31–33].

Considering that the ternary lithium-ion battery has strong nonlinear characteristics, but the content volume attenuation during the life cycle [34] is difficult to change suddenly, based on the cycle method and data-driven ideas, the improved Rainflow algorithm is used to calculate the SOH and establish the autoregressive integrated moving average (ARIMA) prediction model, and uses the augmented Dickey–Fuller (ADF), KPSS, and Akaike information criterion (AIC) multiple inspection methods to determine the optimal state of the model, and compares the predicted value with the true value in order to get the prediction effect.

2. Mathematical Analysis

2.1. Forecasting Process

To identify as few model parameters as possible and to improve the prediction accuracy and efficiency, this research is based on the improved Rainflow counting method and the optimized test ARIMA model so as to predict the lithium-ion battery SOH. The prediction flow chart is shown in Figure 1.

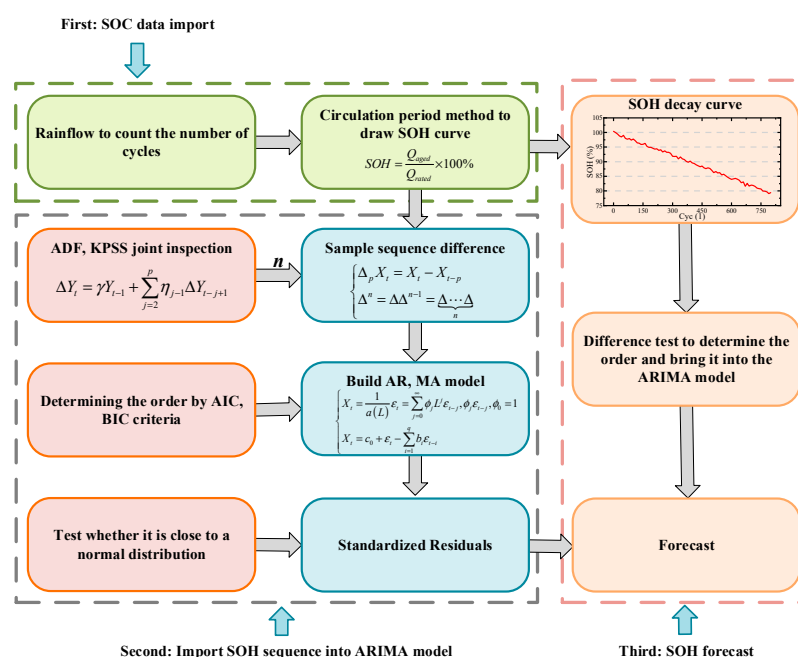


Figure 1. State of health (SOH) prediction process.

In Figure 1, the SOH prediction can be divided into three parts. First, import the data, and derive the SOH attenuation curve through the improved rain flow counting method. Then, import the obtained SOH sequence into the ARIMA model, and perform differential stationery, sequence determination, and residual test processing on it. Finally, the original data and the data processed by the model are integrated and predicted.

2.2. Improved Rainflow Algorithm

For the Rainflow algorithm, the main function of this counting method is to simplify the actual measured load history into several load cycles, and each cycle is a damage accumulation. At the same time, to reduce the number of half-cycles, it is necessary to reconstruct the time history of the data before counting and move the absolute value of the peak or trough to the starting point of the process, as shown in Figure 1. In the estimation of the lithium-ion battery life, the load is SOC, both a set of SOC–time curves are obtained, and then the entire coordinate system is rotated 90 degrees clockwise, and the time coordinate axis is vertically downward. The roof goes down as shown in Figure 2.

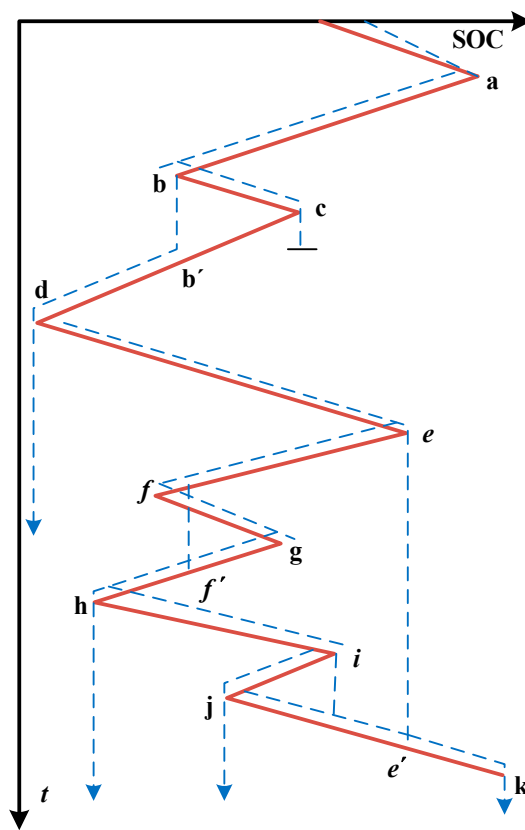


Figure 2. SOC-time process analysis.

The main principle of the Rainflow algorithm is to imitate the process of raindrops falling along the eaves. When raindrops fall to the next eaves, a loop will be formed. The SOC time record shown in Figure 2 includes three complete cycles ($b-c-b'$, $f-g-f'$, and $i-j-e'$) and three half-cycles ($a-b-d$, $d-e-e'$, and $e-f-h$).

The flow chart of the improved Rainflow algorithm is shown in Figure 3.

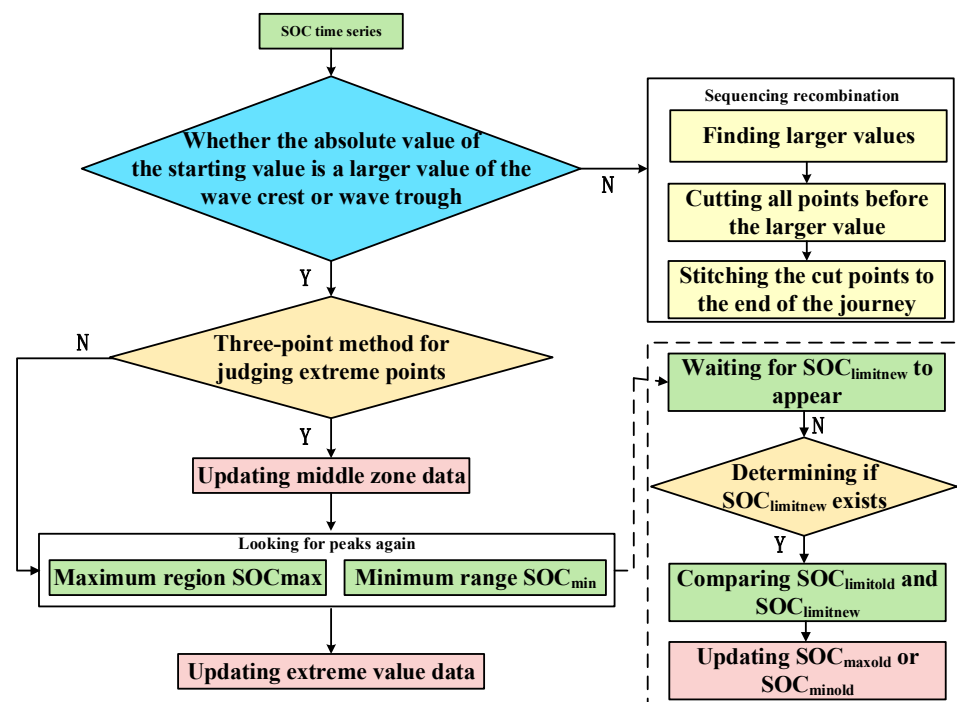


Figure 3. Improvement of the Rainflow process.

First, judge whether the starting value is the peak value. If not, the data before the peak value are cut to the end of the data segment, and the peak value is taken as the starting value. The maximum and minimum values are determined by the three-point method, and the data are updated at any time. The middle zone is used to store extreme points. According to the counting principle of the Rainflow algorithm mentioned above, lithium-ion batteries may experience some small cycles during a deep cycle. These small cycles may be a small range of charge and discharge, or they may be the oscillation caused by the electrochemical properties of the lithium-ion batteries. Because of the advantages of lithium-ion batteries, the impact of these small cycles on the life of a lithium-ion battery can be ignored. In addition, the traditional Rainflow algorithm is sensitive to the size of the cycle. Therefore, the small cycles in these processes can be filtered out by the improved Rainflow algorithm.

The traditional Rainflow algorithm estimates the load level by calculating the average stress. This method can also be used as a reference for the lithium-ion battery load level, but the traditional calculation formula is not suitable for the lithium-ion battery load calculation. The formula modification is shown in Equation (1).

$$D_m = \frac{1 - DOD_{\min} + DOD_{\max}}{2} \quad (1)$$

DOD_{\max} and DOD_{\min} in Equation (1) are the maximum and minimum discharge depths (DOD), respectively, of a group of charge–discharge cycles. Input the SOC sequence to the algorithm, and then the data are obtained by counting the Rainflow of the SOC. The calculated results are in the range of 0–100. The larger the calculated value, the heavier the load of the lithium-ion battery, and vice versa. Therefore, the average DOD of D_m is used to evaluate the battery load.

The main idea of using the rain flow counting method is to calculate the battery SOH using the cycle period method, and the main calculation is shown in Equation (2).

$$SOH = \frac{Q_{aged}}{Q_{rated}} \times 100\% \quad (2)$$

Q_{aged} represents the rated capacity of the battery when it leaves the factory. Q_{rated} represents the available capacity of the battery after it is put into use. The advantage of the cyclic period method is that it avoids cumbersome parameter identification and is simple and fast.

2.3. Second-Order Stationarity Test

Second-order stationarity requires that the input sample time series undergoes curve fitting, and the current trend will continue for some time in the future and hold inertia, and the first-order and second-order moments of the second-order stationary time series will not change with time. To make the SOC time series meet the requirements of second-order stability, it is necessary to perform a different operation on it. The operation equation is shown in Equation (3).

$$\begin{cases} \Delta_p X_t = X_t - X_{t-p} \\ \Delta^n = \Delta \Delta^{n-1} = \underbrace{\Delta \cdots \Delta}_n \end{cases} \quad (3)$$

In Equation (3), Δ is the difference operator and Δ_p represents the p -step difference. The time series $\{X_t\}$ obtained by the first difference of the time series is the first-order difference. If the difference is performed multiple times, the time series $\{\Delta^n\}$ is the n -th order difference. Differential data can improve the signal accuracy and remove the common error interference.

Although the difference is simple and effective, if the trend is stable, the difference of the sequence will cause excessive difference, resulting in the loss of effective signals. If the difference order is too low, the data will not meet the requirements of stationarity and the prediction results will diverge. Therefore, this paper uses the combined test of augmented Dickey–Fuller (ADF) and KPSS to determine the reasonable difference order.

In a real situation, most time series have a high-order autocorrelation, so the ADF method is used to investigate high-order autoregression in the model. The time-series data Y_t can be represented by the superposition of the weight of the historical data and the random disturbance. The expression is shown in Equation (4).

$$\begin{cases} Y_t = \sum_{j=1}^p \alpha_j Y_{t-j} + \varepsilon_t \\ \Delta Y_t = \gamma Y_{t-1} + \sum_{j=2}^p \eta_{j-1} \Delta Y_{t-j+1} \\ \gamma = \sum_{i=1}^p \alpha_i - 1 \\ \eta_i = - \sum_{i=1}^{p-1} \alpha_{i+1} \end{cases} \quad (4)$$

In Equation (4), α_j is the autoregressive coefficient and ε_t is the random disturbance term. Subtract Y_{t-1} from both ends of the equal sign of the first sub-expression, and then write it down as the second sub-expression after differential transformation. The first sub-formula is a linear differential equation. When $\gamma = 0$, the corresponding characteristic equation has at least one unit root. At this time, the stationarity of the sequence $\{Y_t\}$ is in a critical state, that is, it is a non-stationary sequence. Therefore, it is necessary to continue to differentiate the sequence until $\gamma < 0$ for the new sequence, that is, the sequence is stable. Therefore, based on Equation (4), the original hypothesis and alternative hypothesis of ADF are $H_0: \gamma = 0$ and $H_1: \gamma < 0$. The original hypothesis is that the original sequence is not stationary, and the alternative hypothesis is that the sequence is stationary.

Although ADF has a good adaptability in high-level inspections, it is a single-ended inspection, which inevitably leads to an excessively high inspection order, resulting in an excessive difference in the series, which in turn leads to a loss of valid data. Therefore, this research introduces the KPSS test and uses KPSS and ADF to test the sequences together. Only when two tests pass at the same time can the difference order be determined. KPSS

test is to remove the intercept term and trend term from the sequence to be tested to construct the statistic LM. The inspection process is shown in Equation (5).

$$\begin{cases} y_t = x_t\delta + \varepsilon_t, y_0 = 0, \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2) \\ \hat{u}_t = y_t - x_t\hat{\delta} \\ LM = \sum_t S(t)^2 / (T^2 f_0), S(t) = \sum_{i=1}^t \hat{u}_i \end{cases} \quad (5)$$

In Formula (5), x_t is a vector sequence of exogenous variables, including the intercept term, or intercept term, and the trend term of the tested sequence is y_t . The second sub-formula is used to estimate the residual sequence by using the least-squares method to regress and judge whether the original sequence has a unit root by checking whether the residual has a unit root. The third sub-formula is the LM statistic, and f_0 is the residual error under the condition of zero frequency. The null hypothesis of the KPSS test is $H_0: \gamma = 0$, and the alternative hypothesis is $H_1: \gamma < 0$. The assumptions of KPSS and ADF are opposite; the original hypothesis is that the sequence is stationary, and the alternative hypothesis is that the sequence is not stationary. When the LM statistic is less than 3 critical values, the null hypothesis is rejected, that is, the sequence has a unit root.

2.4. ARIMA Model Establishment

The autoregressive model (AR) uses itself as the process of regression variables and uses the linear combination of random variables at a certain time in the previous period to describe the linear regression process of random variables at a certain time in the future. Compared with other linear regressions, autoregressive does not use x to predict y , but uses x to predict x itself. Set the time series $\{X_t\}$ as Equation (6).

$$X_t = \varepsilon_t + a_0 + \sum_{i=1}^p a_i X_{t-i} \quad (6)$$

where $\{\varepsilon_t\}$ is the white noise sequence, a_0, a_1, \dots , and a_p is $p + 1$ real number; this model is called the p -order autoregressive model, which is recorded as the $AR(p)$ model, and $\{X_t\}$ suitable for this model is $AR(p)$ sequence. The autoregressive coefficient polynomial defined in $AR(p)$ is the autoregressive coefficient polynomial of the $AR(p)$ model, as shown in Equation (7).

$$\begin{cases} a(u) = 1 - a_0 - \sum_{i=1}^p a_i u^i \\ \partial(L) = 1 - \sum_{i=1}^p a_i L^i, \partial(L)X_t = \varepsilon_t \\ X_t = \frac{1}{a(L)}\varepsilon_t = \sum_{j=0}^{\infty} \phi_j L^j \varepsilon_{t-j}, \phi_j \varepsilon_{t-j}, \phi_0 = 1 \end{cases} \quad (7)$$

The first sub-formula in Equation (7) is the autoregressive coefficient polynomial. Put the lag operator Δ in Equation (3) into the first sub-form to get the second sub-form; L represents the lag operator, and at this time, the operator expression of the $AR(p)$ model can be expressed by ε_t . The rule can be found by bringing in multiple different p values, and finally, the solution of $AR(p)$ is shown in the third sub-formula, where Φ_j is the sum of the coefficients for all items in the $\{X_t\}$ sequence.

The function of the moving average (MA) model is to add constraints to the AR model, and a finite number of parameters b is proposed to limit parameter a in the AR model and to avoid the divergence of the AR model. The model of MA is shown in Equation (8).

$$\begin{cases} X_t = c_0 + 1 - \varepsilon_t \sum_{i=1}^p b_i L^i, q > 0 \\ X_t = a_0 + \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t - \sum_{j=1}^q b_j \varepsilon_{t-j} \end{cases} \quad (8)$$

In Equation (8), c_0 is a constant and $\{\varepsilon_t\}$ is the white noise sequence. To make the $\{X_t\}$ sequence stationary, the absolute value of b_1 must be less than 1, otherwise, the $\{b_i\}$ sequence will diverge. Therefore, the combination of the AR model and the MA model becomes the ARMA model, as shown in the second sub-formula of Equation (8).

Combining the second-order smoothing process of the data with the ARMA model is the ARIMA model. There are three main parameters in the ARIMA model—the difference order d , the order p of the AR model, and the order q of the MA model. From Equation (3) to Equation (5), the value of d can be determined. The values of p and q will be determined using the Akaike information criterion (AIC).

$$\begin{cases} AIC = 2k - 2\ln(L) \\ AIC(p_0, q_0) = \min_{1 \leq p, q \leq M(N)} AIC(p, q) \end{cases} \quad (9)$$

In Equation (9), k is the number of parameters, and the number depends on p and q in Equation (8). When the accuracy is guaranteed, the smaller the value of k , the better. L is the maximum likelihood value of the model. Because the maximum likelihood requires a high data volume and is difficult to solve, the residual variance obtained by the least square estimation is often used as an approximate substitute in practice. AIC conducts a comprehensive analysis of k and $\ln(L)$. When the order of the model increases, $\ln(L)$ usually decreases. When the number of observation data is given, the value of k increases as the order of the model increases. When the model order is gradually increased and the data are fitted, the value of AIC shows a downward trend. At this time, the residual variance of the model decreases faster, and $\ln(L)$ will play a decisive role. When the order rises to a certain point, the AIC value reaches a minimum. Subsequently, no matter how the model order increases, the residual variance is almost unchanged. At this time, the influence of $\ln(L)$ is weakened, and k will play a key role. Given the highest order $M(N)$ in advance, take m_0 and n_0 as the best order of the model.

2.5. Residual Test

To ensure that the order of p and q in ARMA is appropriate, a residual test is also needed. The residual is the residual signal after subtracting the signal fitted by the model from the original signal. If the residuals are randomly distributed normally and are white noise sequences, it means that all of the effective signals have been extracted into the ARMA model. Therefore, the significance test of the model is the white noise test of the residual sequence. The original hypothesis, alternative hypothesis, and Ljung–Box (LB) statistics are shown in Equation (10).

$$\begin{cases} H_0 : \varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_m = 0, \forall m \geq 1 \\ H_1 : \text{at least exists } \varepsilon_k \neq 0, \forall m \geq 1, k \leq m \\ LB = n(n+2) \sum_{k=1}^m \left(\frac{\hat{\varepsilon}_k^2}{n-k} \right) \sim \chi^2(m), \forall m > 0 \end{cases} \quad (10)$$

In Equation (10), n is the number of samples and $\hat{\varepsilon}_k^2$ is the correlation coefficient of the k -order lag of the sample. This statistic obeys the χ^2 distribution with m degrees of freedom. Given the significance level α , the rejection domain is $LB > X_{1-\alpha, m}^2$. If the null hypothesis is rejected, it means that there is still relevant information in the residual sequence, and the fitting model is not significant. If the null hypothesis cannot be rejected, the fitted model is considered to be significantly effective.

3. Experimental Analysis

3.1. Construction of the Experimental Platform

This research has a reliable software and hardware foundation and stable battery research laboratory conditions. It has complete experimental equipment, including an electronic load, high-precision digital oscilloscope, UTP3313TFL DC power supply (Haixu

Instrument Co., Ltd., Shenzhen, China), low-noise current amplifier (NT59-179, ST570) (Wanlong Co., Ltd., Suzhou, China), low-noise voltage amplifier (SR560)(stanford research systems, Inc., Sunnyvale, CA, USA), three-layer independent temperature control high- and low-temperature test box (BTT-331C)(Dongguan bell Experimental Equipment Co., Ltd., Guangdong, China), power battery module test system BTS750-200-100-4(Yakeyuan Electric Co., Ltd., Shenzhen, China), simulated high-altitude and low-voltage test box (BE-8104), power cell high-rate charge and discharge tester (CT-4016-5V100A-NTFA)(Yakeyuan Electric Co., Ltd., Shenzhen, China), and power battery drop test bench (BF-F-315ST)(Dongguan bell Experimental Equipment Co., Ltd., Guangdong, China), as well as and supporting experimental equipment, some of which are shown in Figure 4.

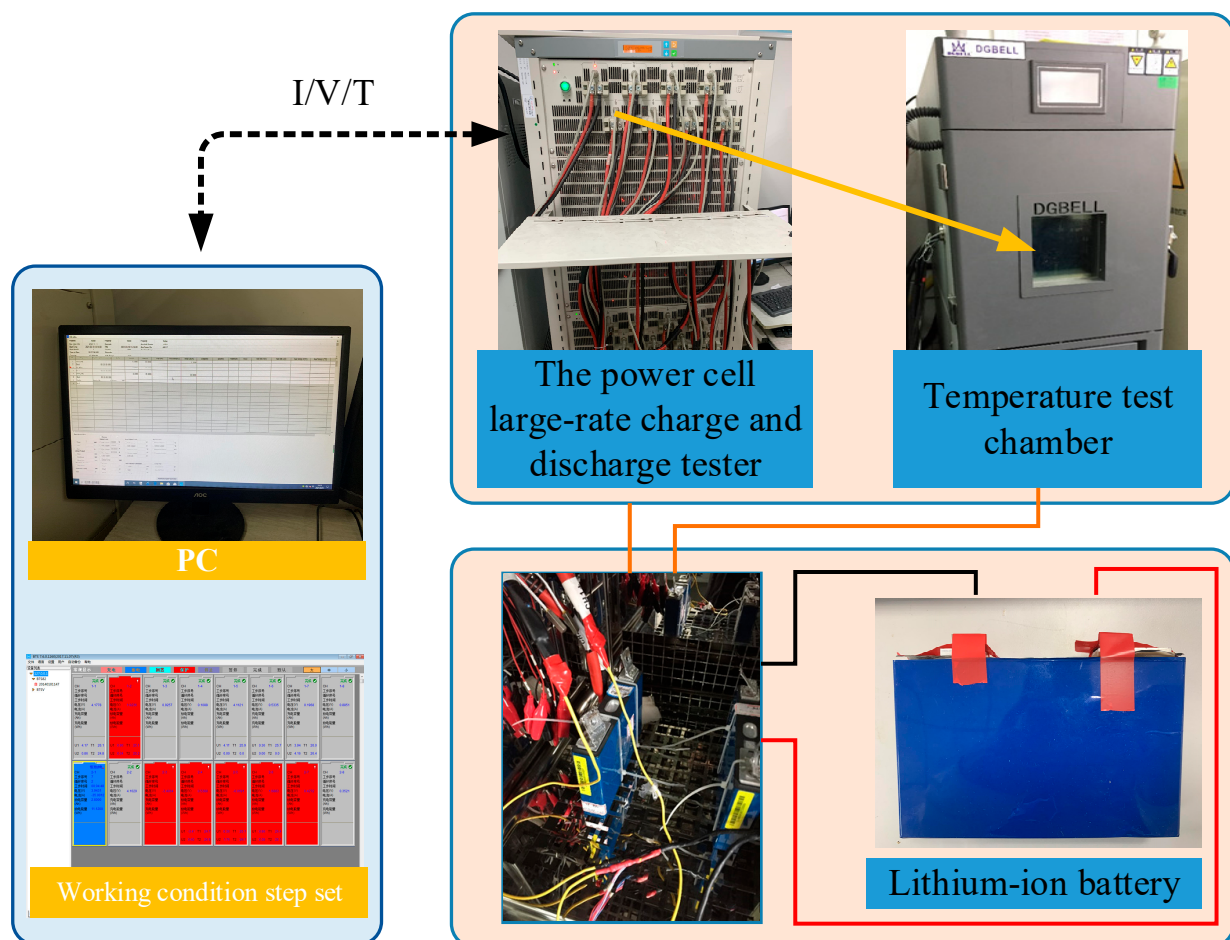


Figure 4. Experiment platform.

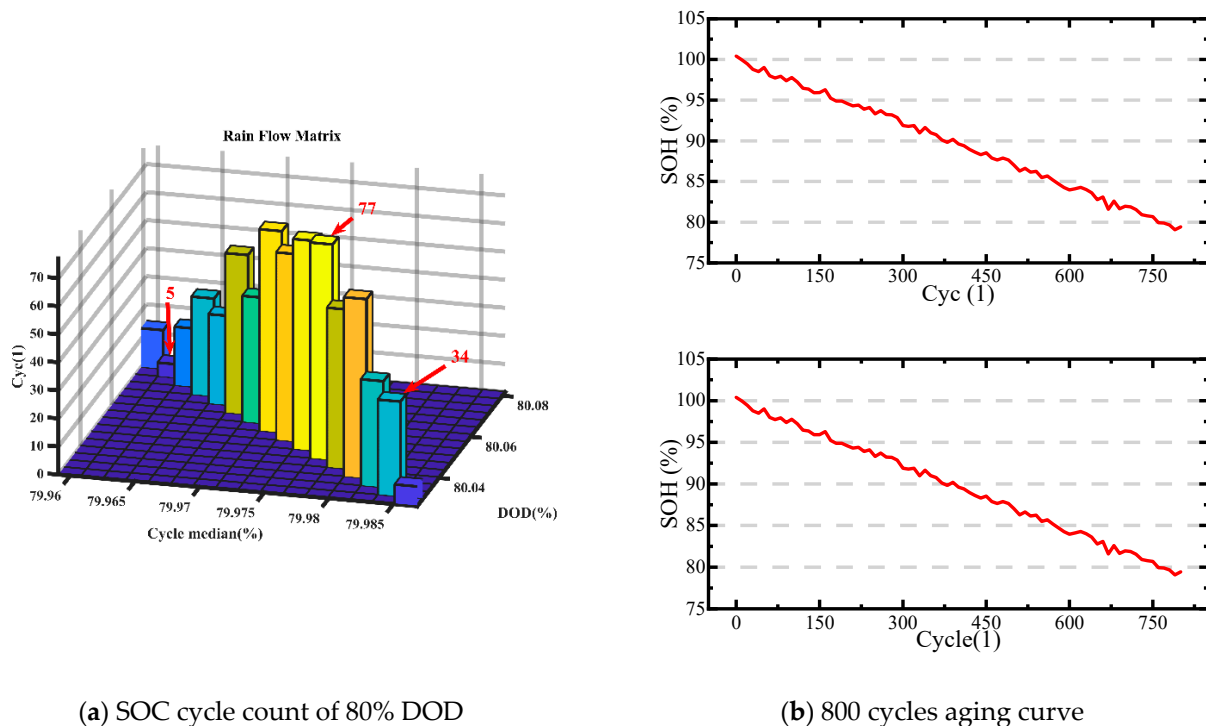
As shown in Figure 4, the ternary lithium-ion battery is placed in a thermostat and is connected to the power experiment device through a professional line. We set the experimental conditions on the PC side. The experimental platform is stable and reliable and has an automatic protection mechanism for power failure. Detailed experimental data can be obtained, including the current, voltage, capacity, time, temperature, energy, etc.

3.2. Rainflow Algorithm to Calculate SOH

To explore the effect of the ARIMA prediction model, a lithium-ion battery cycle life experiment was carried out. This study uses a ternary lithium-ion battery as the experimental object. According to the parameters given by the battery manufacturer, the recommended SOC window of the battery is 10%~90%, and the theoretical cycle life under deep discharge (80% DOD) is about 800 times. In real life, in most cases, the battery will

not be fully discharged before charging, so 80% DOD is a more balanced choice. Therefore, in this experiment, 800 times 80% depth of discharge (DOD) tests were performed on lithium-ion batteries at a temperature of 25 °C, and a capacity measurement was performed every 10 cycles. According to the principle of the cycle method, the SOH attenuation curve of the lithium-ion battery is drawn based on the improved Rainflow counting method.

The cyclic median value in Figure 5a represents the average stress, and the calculation formula is shown in Equation (1). The end of life of a lithium-ion battery is defined as damage when the current full capacitance drops to 80% of the initial capacity. In fact, when the battery capacity is less than 80%, it can still be used, but the performance will decline. The meaning of the average stress here is that when the damage condition of the lithium-ion battery exceeds the defined value, the stress level should be reduced when in use, and one should avoid further deep discharge of battery, i.e., reduce battery load. Figure 5a shows the corresponding DOD of a lithium-ion battery under different average stress. Cyc represents the number of cycles, corresponding to the median value of each cycle of the battery, and the total number of cycles is 800 when all of the numbers are added. CYC is the number of cycles of the battery in each cycle median case, and the total number of cycles is 800. Figure 5b is based on the number of cycles obtained in Figure 5a using the cycle period method to draw the SOH attenuation curve.



(a) SOC cycle count of 80% DOD

(b) 800 cycles aging curve

Figure 5. Life decay curve for 800 cycles at 80% discharge depths (DOD).

3.3. ADF and KPSS Jointly Verify the Differential Sequence

The SOH of the lithium-ion battery decays slowly and smoothly, so the SOH sample of the cycle condition is stable and inertial. The result of the joint test of ADF and KPSS is 0th order, that is, the sample sequence itself is stable and no difference is required. The sample sequence is shown in Figure 5b.

To better verify the accuracy of the prediction model, the study used the SOH sequence of an 80% DOD cycle test, the SOC sequence of a 10% DOD cycle test, and the SOC data of a dynamic stress test (DST) as the prediction samples.

Figure 6a,c are the SOC curves under DST operating conditions and 10% depth-of-discharge cycle operating conditions, respectively. The joint test of ADF and KPSS in

Figure 6b shows that the order of difference for the sequence to reach stationary is second. To verify the different test results, Figure 6b makes the first-, second-, and third-order differences on the sequence of Figure 6b $\delta 1$, $\delta 2$, and $\delta 3$ represent the first, second, and third order differences, respectively. The mean value of $\delta 1$ is unstable, $\delta 3$ is over-differentiated and there are many burrs, and only the mean value of the $\delta 2$ sequence is stable without unnecessary burrs. Figure 6d is the same, the second-order difference satisfies the second-order stationary condition.

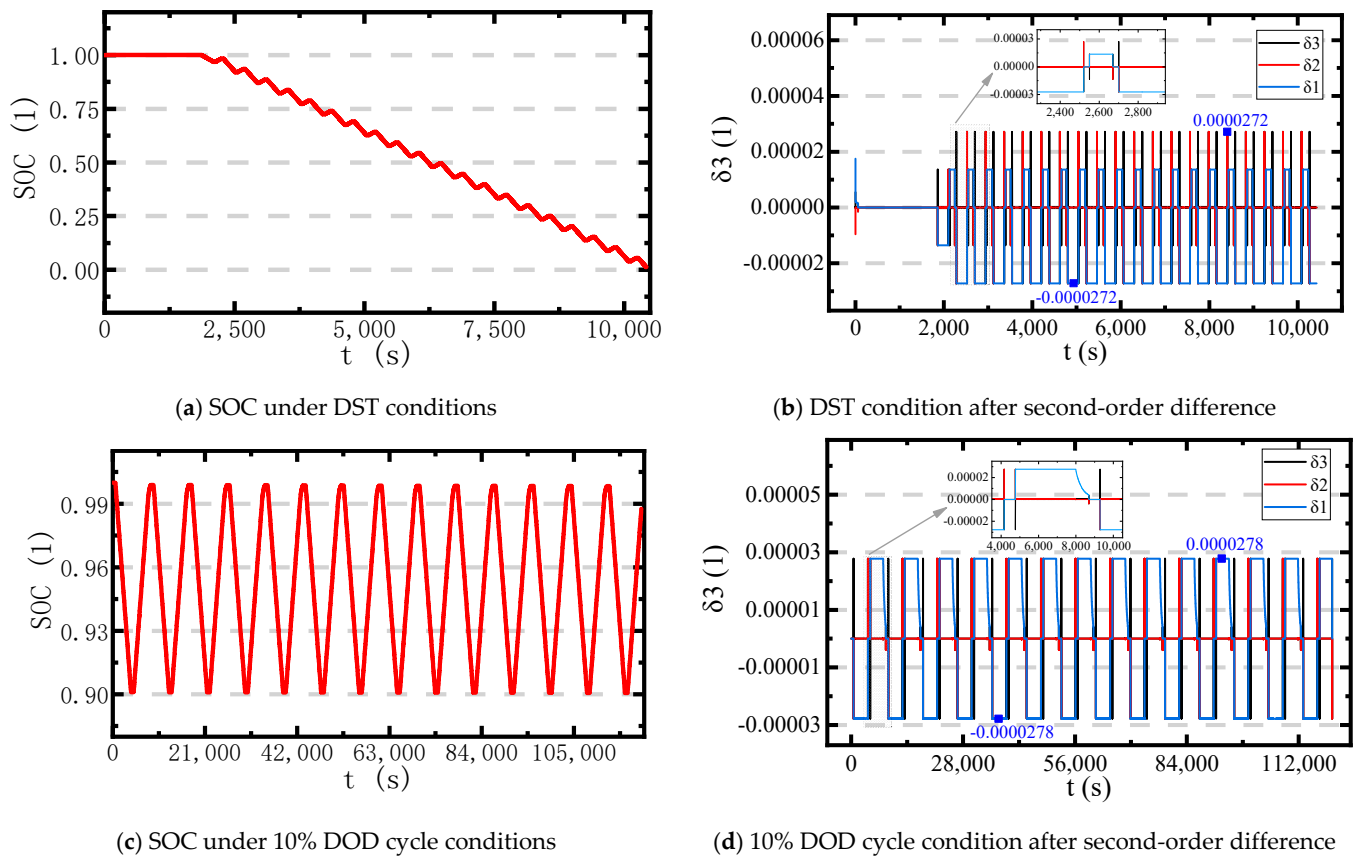
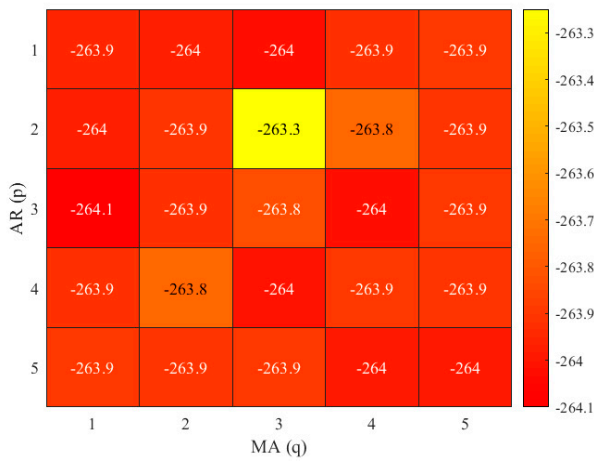


Figure 6. Second-order smoothing of different working condition data.

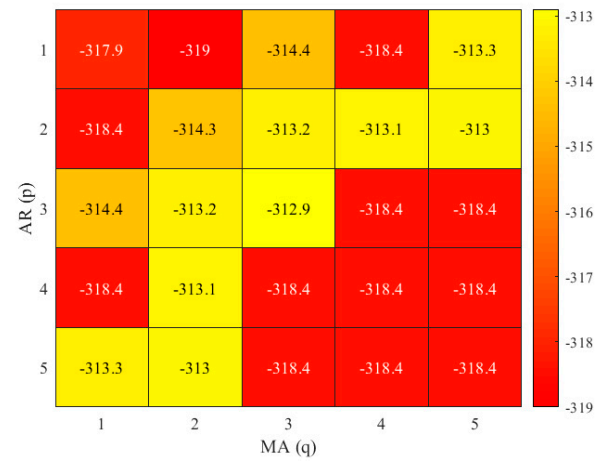
3.4. Complex Condition Analysis

It can be seen from Equation (9) that in the AIC criterion, the model parameter quantity k and the maximum likelihood value L are in dynamic equilibrium, and the optimal solution can only be found from many solutions. When making predictions, it is hoped that the fewer model parameters, the better, so the smaller the value of k , the better. The larger L , the more accurate the result. Therefore, the smaller the value of AIC, the better the corresponding p and q values.

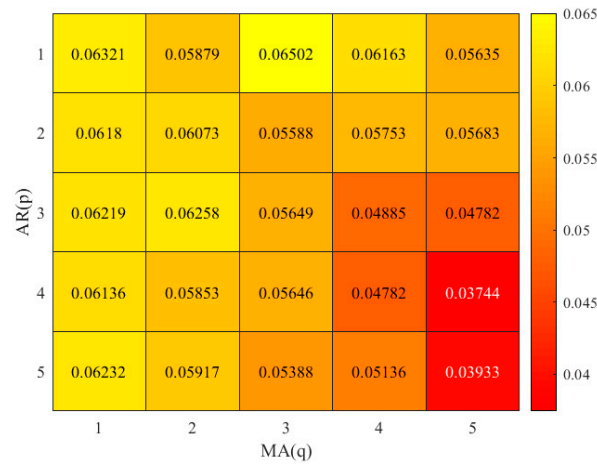
The optimal solution can be found intuitively in the form of a heat map. Finding the optimal solution through AIC allows for finding the minimum value. The minimum value in the heat map has the darkest color. The parameters corresponding to p and q in Figure 7a are 3 and 1, respectively. The parameters corresponding to p and q in Figure 7b are 1 and 2, respectively. The parameters corresponding to p and q in Figure 7c are 4 and 5, respectively. If the p or q order is selected too high, then k and L in Equation (9) will have an absolute effect on one side, which may cause the error of the ARIMA model to increase. Therefore, when the AIC criterion is used in this study, the maximum order is set to 5.



(a) p and q parameter setting order under DST conditions



(b) 10% DOD cycle condition SOC sequence p and q parameter setting order



(c) 80% DOD cycle conditions SOH sequence p and q parameter setting order

Figure 7. Akaike information criterion (AIC) criteria for parameter selection.

3.5. Residual Error Test

To facilitate the test, this article mainly uses quantile–quantile (Q–Q) graphs to test whether the data conform to the normal distribution. The function of the Q–Q chart is to compare whether the quantiles of the two columns of data are distributed on the straight line of $y = x$. The Q–Q chart examines the distribution of data by comparing the quantiles of a column of sample data with the quantiles of a column of data of known distribution. The residual error test of the DST condition and cycle condition is shown in Figure 8.

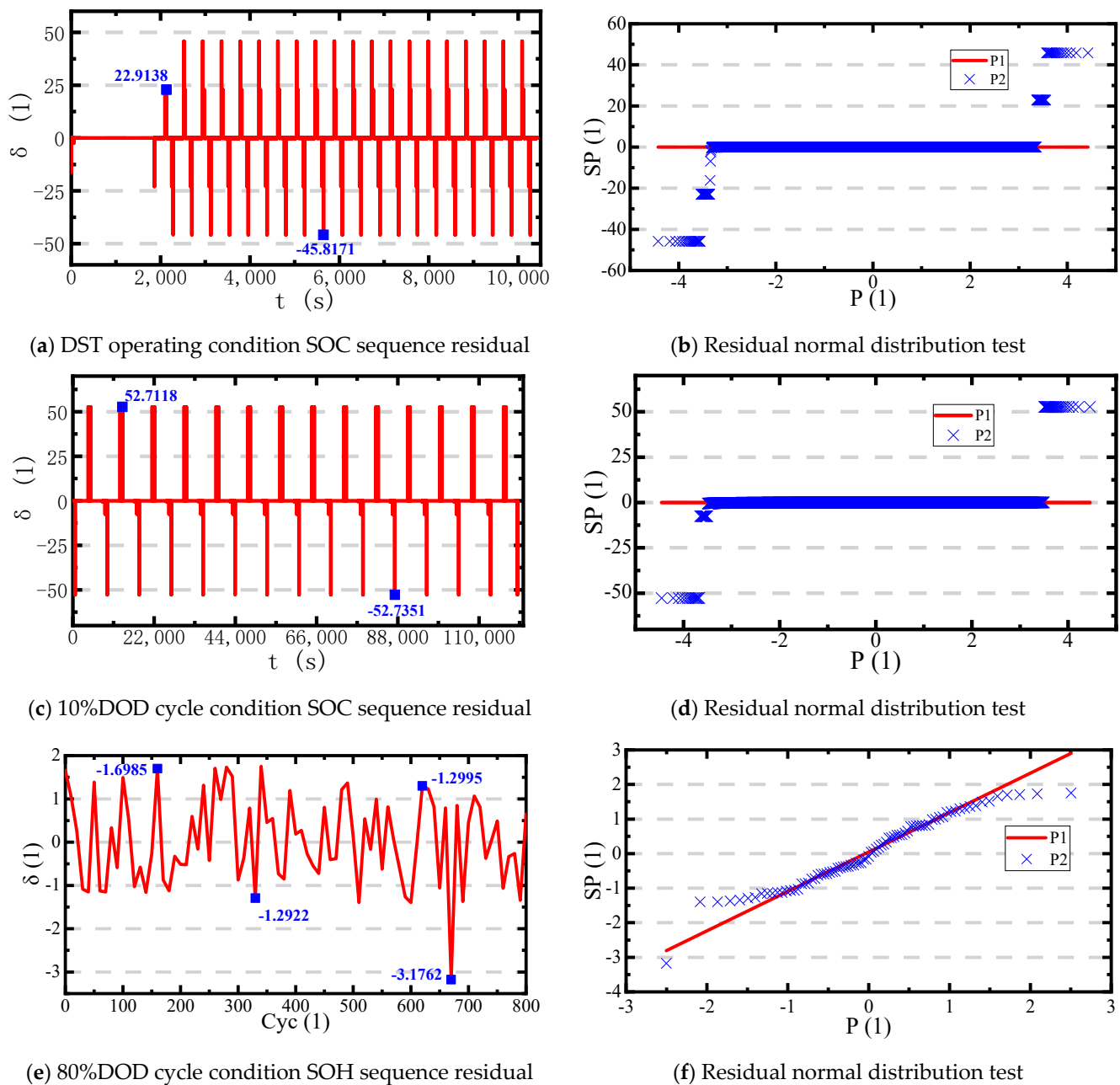


Figure 8. Q-Q chart test whether the residuals conform to the normal distribution.

Figure 8a,c,e (are the residual sequence of samples under different working conditions, and Figure 8b,d,f are the Q–Q diagrams corresponding to the residuals of each working condition. SP represents the quantile of the input sample, and P represents the standard normal quantile. Compare the SP and P series to test whether the residuals are close to a normal distribution. Curve P1 represents the standard normal distribution, and curve P2 represents the distribution of the residuals. The higher the degree of overlap between curve P2 and curve P1, the closer the residual is to a normal distribution. It can be seen from Figure 8 that the residuals are close to the normal distribution under the three working conditions.

3.6. Predictive Verification

All of the parameters can be imported into the ARIMA model to predict the future. For the aperiodic sequence forecasting length, 10% of the sample length is usually sampled.

For a cyclical sequence forecasting time, the forecasting time can be extended. The study uses the method of comparing the predicted value with the true value and the method of using the confidence interval as the error range to verify the prediction. The predictions of the three operating conditions are shown in Figure 8.

In Figure 9a, S1 represents the true value of the first 80% of the entire SOC sequence. The data of S1 are used to establish the ARIMA model, S2 represents the predicted value of 20% after the SOC sequence, and S3 is the true value of 20% after the SOC sequence, which is used for comparison with the predicted value of S2. Figure 9b shows the error between the predicted value and the true value. The prediction error is stable within 5.3160% under DST conditions. Figure 9c,d are similar, except that the ratio of the data used for modeling to the predicted data in Figure 9c is 7:3. The prediction error under 10% DOD cycle conditions is within 5.4517%. One unit of step in Figure 9e corresponds to 10 cycles of the battery; S1 represents the true value of SOH, S2 represents the predicted value, and S3 represents the 95% confidence interval. SOH is predicted when the lithium-ion battery is cycled 500 times and 800 times, respectively, and the predicted number of steps is 10 steps, that is, the prediction of 100 cycles in the future. As it is difficult for lithium-ion batteries to have sudden changes in SOH under normal use, and because of the non-linear characteristics of lithium-ion batteries, it is difficult to track SOH in real-time, so this paper draws the confidence interval to lock the error range of SOH, as shown in Figure 9f. E1 represents the error between the upper half confidence interval and the predicted value, and E2 represents the error between the lower half confidence interval and the predicted value. The results show that the prediction error for the next 10 steps is within 1%, that is, the SOH will fluctuate based on the predicted value and the amplitude will not exceed 1%. The method proposed in this study has a maximum prediction error of 5.4517% under these three conditions, which is 3.3% less than that in the reference [27] (the maximum error is 8.8%).

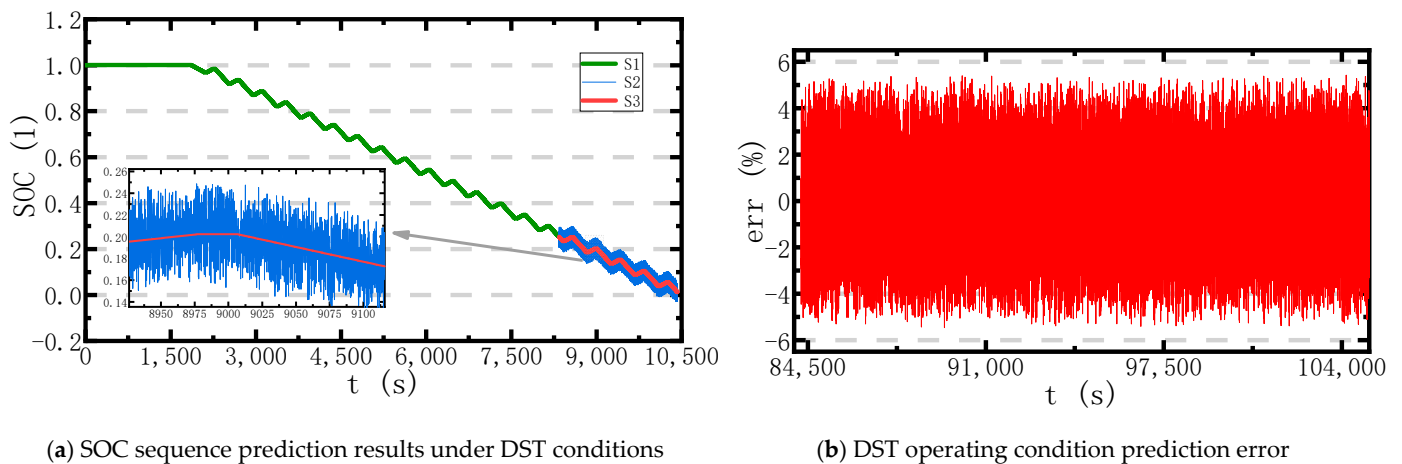


Figure 9. Cont.

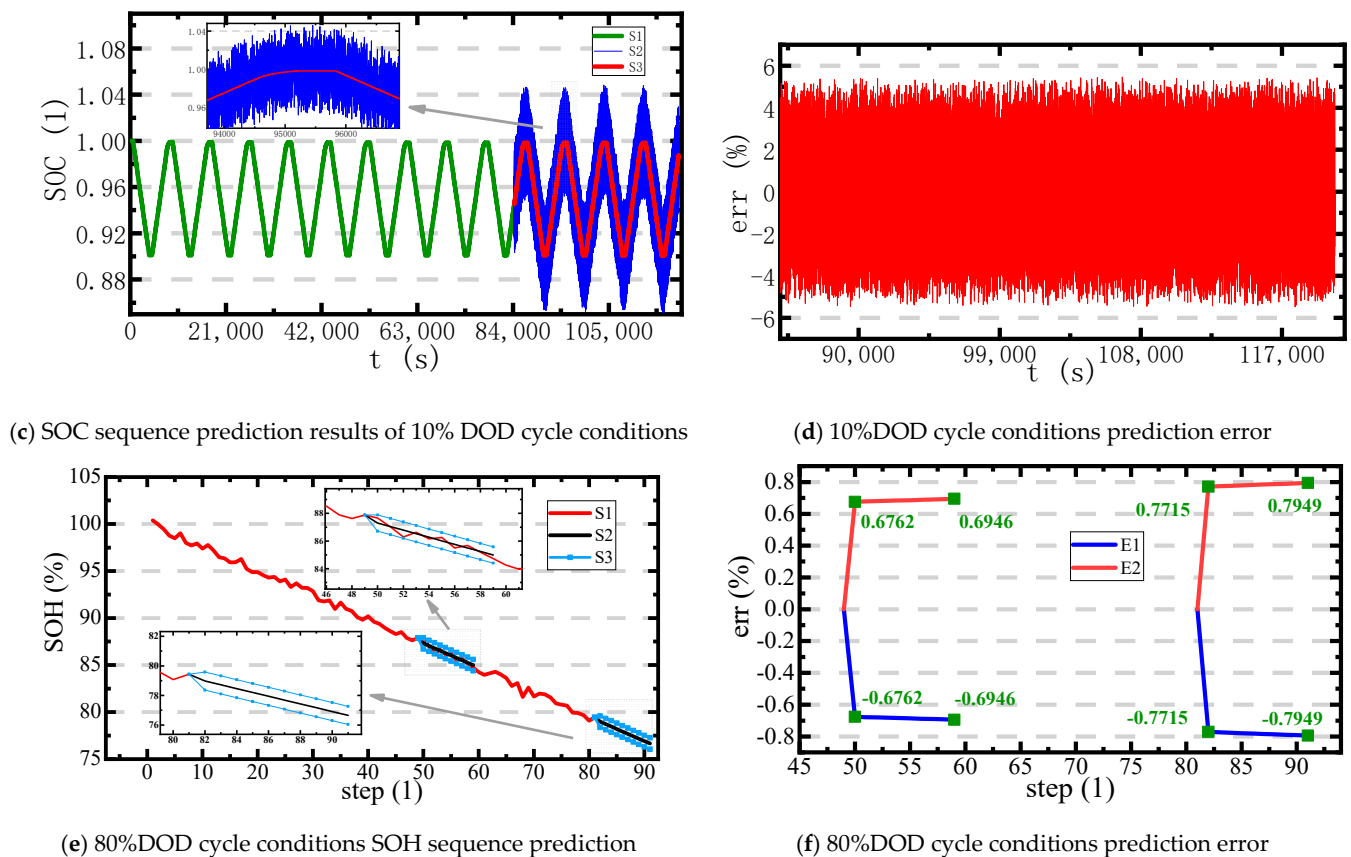


Figure 9. Model prediction effect.

4. Conclusions

The accurate estimation and prediction of lithium-ion battery SOH is the key and difficult point of lithium-ion battery condition monitoring. This article is based on the idea of the cycle method, and calculates the battery SOH with an improved rain flow counting method. Then, we built an ARIMA model based on data-driven thinking, using a more accurate inspection method, further streamlining the parameter identification steps, and finally predicting the future SOH trend of the battery in combination with a variety of actual operating conditions. The experimental results show that the prediction error of the complex working conditions is controlled within 5.4517%, and the prediction error of simple working conditions is controlled within 0.7949%. The model has high calculation efficiency and high accuracy for SOH prediction.

Author Contributions: Conceptualization, J.H.; Data curation, J.H. and W.S.; Investigation, W.X.; Methodology, J.H.; Project administration, S.W. and C.F.; Resources, S.W. and C.F. Writing—original draft, J.H.; Writing—review & editing, S.W. and C.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 61801407.

Data Availability Statement: Publicly archived datasets: <https://www.researchgate.net/project/DD-SOH-Test>.

Acknowledgments: The work was supported by the National Natural Science Foundation of China (no. 61801407), the Sichuan Science and Technology Program (no. 2019YFG0427), the China Scholarship Council (no. 201908515099), and the Fund of Robot Technology Used for Special Environment Key Laboratory of Sichuan Province (no. 18kftk03).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jiang, Y.; Zhang, J.; Xia, L.; Liu, Y. State of Health Estimation for Lithium-Ion Battery Using Empirical Degradation and Error Compensation Models. *IEEE Access* **2020**, *8*, 123858–123868. [[CrossRef](#)]
2. Lin, C.-P.; Cabrera, J.; Yang, F.; Ling, M.-H.; Tsui, K.-L.; Bae, S.-J. Battery state of health modeling and remaining useful life prediction through time series model. *Appl. Energy* **2020**, *275*, 115338. [[CrossRef](#)]
3. Li, P.; Zhang, Z.; Xiong, Q.; Ding, B.; Hou, J.; Luo, D.; Rong, Y.; Li, S. State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network. *J. Power Sources* **2020**, *459*, 228069. [[CrossRef](#)]
4. Zhou, Y.; Huang, M.; Pecht, M. Remaining useful life estimation of lithium-ion cells based on k-nearest neighbor regression with differential evolution optimization. *J. Clean. Prod.* **2020**, *249*, 119409. [[CrossRef](#)]
5. Tan, X.; Tan, Y.; Zhan, D.; Yu, Z.; Fan, Y.; Qiu, J.; Li, J. Real-Time State-of-Health Estimation of Lithium-Ion Batteries Based on the Equivalent Internal Resistance. *IEEE Access* **2020**, *8*, 56811–56822. [[CrossRef](#)]
6. Tian, J.; Xiong, R.; Yu, Q. Fractional-Order Model-Based Incremental Capacity Analysis for Degradation State Recognition of Lithium-Ion Batteries. *IEEE Trans. Ind. Electron.* **2019**, *66*, 1576–1584. [[CrossRef](#)]
7. Gou, B.; Xu, Y.; Feng, X. State-of-Health Estimation and Remaining-Useful-Life Prediction for Lithium-Ion Battery Using a Hybrid Data-Driven Method. *IEEE Trans. Veh. Technol.* **2020**, *69*, 10854–10867. [[CrossRef](#)]
8. Wu, Y.; Xue, Q.; Shen, J.; Lei, Z.; Chen, Z.; Liu, Y. State of Health Estimation for Lithium-Ion Batteries Based on Healthy Features and Long Short-Term Memory. *IEEE Access* **2020**, *8*, 28533–28547. [[CrossRef](#)]
9. Zhou, D.; Li, Z.; Zhu, J.; Zhang, H.; Hou, L. State of Health Monitoring and Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Temporal Convolutional Network. *IEEE Access* **2020**, *8*, 53307–53320. [[CrossRef](#)]
10. Tian, H.X.; Qin, P.L. State of health prediction for lithium-ion batteries with a novel online sequential extreme learning machine method. *Int. J. Energy Res.* **2021**, *45*, 2383–2397. [[CrossRef](#)]
11. Tan, X.J.; Zhan, D.; Lyu, P.; Rao, J.; Fan, Y. Online state-of-health estimation of lithium-ion battery based on dynamic parameter identification at multi timescale and support vector regression. *J. Power Sources* **2021**, *484*, 229233. [[CrossRef](#)]
12. Xiao, B.; Liu, L. Rapid measurement method for lithium-ion battery state of health estimation based on least squares support vector regression. *Int. J. Energy Res.* **2021**, *45*, 5695–5709. [[CrossRef](#)]
13. Shi, E.; Xia, F.; Peng, D.; Li, L.; Wang, X.; Yu, B. State-of-health estimation for lithium battery in electric vehicles based on improved unscented particle filter. *J. Renew. Sustain. Energy* **2019**, *11*, 024101. [[CrossRef](#)]
14. Wang, X. A Novel Gaussian Particle Swarms optimized Particle Filter Algorithm for the State of Charge Estimation of Lithium-ion Batteries. *Int. J. Electrochem. Sci.* **2020**, *15*, 10632–10648. [[CrossRef](#)]
15. Bi, Y.; Yin, Y.; Choe, S.-Y. Online state of health and aging parameter estimation using a physics-based life model with a particle filter. *J. Power Sources* **2020**, *476*, 228655. [[CrossRef](#)]
16. Zhang, S.; Zhai, B.; Guo, X.; Wang, K.; Peng, N.; Zhang, X. Synchronous estimation of state of health and remaining useful lifetime for lithium-ion battery using the incremental capacity and artificial neural networks. *J. Energy Storage* **2019**, *26*, 100951. [[CrossRef](#)]
17. Tang, X.; Wang, Y.; Zou, C.; Yao, K.; Xia, Y.; Gao, F. A novel framework for Lithium-ion battery modeling considering uncertainties of temperature and aging. *Energy Convers. Manag.* **2019**, *180*, 162–170. [[CrossRef](#)]
18. Zhao, M.; Zhong, S.; Fu, X.; Tang, B.; Dong, S.; Pecht, M. Deep Residual Networks With Adaptively Parametric Rectifier Linear Units for Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2021**, *68*, 2587–2597. [[CrossRef](#)]
19. Juarez-Robles, D.; Vyas, A.A.; Fear, C.; Jeevarajan, J.A.; Mukherjee, P.P. Overcharge and Aging Analytics of Li-Ion Cells. *J. Electrochem. Soc.* **2020**, *167*, 090547. [[CrossRef](#)]
20. Liu, K.; Shang, Y.; Ouyang, Q.; Widanage, W.D. A Data-Driven Approach with Uncertainty Quantification for Predicting Future Capacities and Remaining Useful Life of Lithium-ion Battery. *IEEE Trans. Ind. Electron.* **2021**, *68*, 3170–3180. [[CrossRef](#)]
21. Eleftheroglou, N.; Mansouri, S.S.; Loutas, T.; Karvelis, P.; Georgoulas, G.; Nikolakopoulos, G.; Zarouchas, D. Intelligent data-driven prognostic methodologies for the real-time remaining useful life until the end-of-discharge estimation of the Lithium-Polymer batteries of unmanned aerial vehicles with uncertainty quantification. *Appl. Energy* **2019**, *254*, 113677. [[CrossRef](#)]
22. Cai, L.; Meng, J.; Stroe, D.-I.; Peng, J.; Luo, G.; Teodorescu, R. Multiobjective Optimization of Data-Driven Model for Lithium-Ion Battery SOH Estimation With Short-Term Feature. *IEEE Trans. Power Electron.* **2020**, *35*, 11855–11864. [[CrossRef](#)]
23. Pan, W.; Chen, Q.; Zhu, M.; Tang, J.; Wang, J. A data-driven fuzzy information granulation approach for battery state of health forecasting. *J. Power Sources* **2020**, *475*, 228716. [[CrossRef](#)]
24. Mawonou, K.S.; Eddahech, A.; Dumur, D.; Beauvois, D.; Godoy, E. State-of-health estimators coupled to a random forest approach for lithium-ion battery aging factor ranking. *J. Power Sources* **2021**, *484*, 229154. [[CrossRef](#)]
25. Liu, K.; Ashwin, T.; Hu, X.; Lucu, M.; Widanage, W.D. An evaluation study of different modelling techniques for calendar ageing prediction of lithium-ion batteries. *Renew. Sustain. Energy Rev.* **2020**, *131*, 110017. [[CrossRef](#)]
26. Ma, Y.; Wu, L.; Guan, Y.; Peng, Z. The capacity estimation and cycle life prediction of lithium-ion batteries using a new broad extreme learning machine approach. *J. Power Sources* **2020**, *476*, 228581. [[CrossRef](#)]
27. 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. [[CrossRef](#)]

28. Qiu, X.; Wu, W.; Wang, S. Remaining useful life prediction of lithium-ion battery based on improved cuckoo search particle filter and a novel state of charge estimation method. *J. Power Sources* **2020**, *450*, 227700. [[CrossRef](#)]
29. Chen, Y.; He, Y.; Li, Z.; Chen, L.; Zhang, C. Remaining Useful Life Prediction and State of Health Diagnosis of Lithium-Ion Battery Based on Second-Order Central Difference Particle Filter. *IEEE Access* **2020**, *8*, 37305–37313. [[CrossRef](#)]
30. Chen, Z.; Xue, Q.; Xiao, R.; Liu, Y.; Shen, J. State of Health Estimation for Lithium-ion Batteries Based on Fusion of Autoregressive Moving Average Model and Elman Neural Network. *IEEE Access* **2019**, *7*, 102662–102678. [[CrossRef](#)]
31. Zhao, X.; Xuan, D.; Zhao, K.; Li, Z. Elman neural network using ant colony optimization algorithm for estimating of state of charge of lithium-ion battery. *J. Energy Storage* **2020**, *32*, 101789. [[CrossRef](#)]
32. Li, W.; Jiao, Z.; Du, L.; Fan, W.; Zhu, Y. An indirect RUL prognosis for lithium-ion battery under vibration stress using Elman neural network. *Int. J. Hydrog. Energy* **2019**, *44*, 12270–12276. [[CrossRef](#)]
33. Li, X.; Zhang, L.; Wang, Z.; Dong, P. Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and Elman neural networks. *J. Energy Storage* **2019**, *21*, 510–518. [[CrossRef](#)]
34. Liu, C.; Wang, Y.; Chen, Z. Degradation model and cycle life prediction for lithium-ion battery used in hybrid energy storage system. *Energy* **2019**, *166*, 796–806. [[CrossRef](#)]