Int. J. Electrochem. Sci., 17 **(2022)** Article ID: 220555, doi: 10.20964/2022.05.61

International Journal of
ELECTROCHEMICAL
SCIENCE
www.electrochemsci.org

A Novel Seasonal Autoregressive Integrated Moving Average Method for the Accurate Lithium-ion Battery Residual Life Prediction

Yifen Hu[1], *Shunli Wang*[2,*], *Junhan Huang*[2], *Paul Takyi-Aninakwa*[2], *Xianpei Chen*[2]

[1] School of Automotive and Mechanical-Electronic Engineering, Xinyang Vocational and Technical College, Xinyang 464000, China
[2] School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China;
[*]E-mail: 497420789@qq.com

Received: 11 August 2021 / *Accepted:* 21 March 2022 / *Published:* 5 April 2022

Lithium-ion batteries are widely used in electric vehicles (EVs), unmanned aerial vehicles (UAVs), and smart devices because of their high specific energy, long service life, and environmental-friendly. The remaining useful life (RUL) prediction is extremely important for the evaluation of the state of health (SOH) of the battery. Also, it is an important indicator to improve its safety in a variety of applications. In this study, a novel seasonal autoregressive integrated moving average (SARIMA) prediction model is proposed. The proposed model adds periodic parameter optimization to fit the nonlinear characteristics of the battery, including maximum likelihood estimation (MLE) and Akaike information criterion (AIC) to filter the parameters more accurately. So, to effectively solve the shortcomings of the traditional prediction methods, such as complex parameter acquisition, low prediction accuracy, and a large amount of sample data. The method proposed in this paper simplifies the remaining useful life prediction process, ensures high accuracy, and improves the safe operation and reliability of lithium-ion batteries. The model can give the prediction confidence bounds, and the maximum prediction error under complex working conditions is 4.62%.

Keywords: lithium-ion battery; remaining useful life prediction; seasonal autoregressive integrated moving average;

1. INTRODUCTION

With the continuous advancement of global science and technology, the energy and environmental crisis is increasingly worsening. As a kind of clean energy with zero gas emission, the lithium-ion battery has the advantages of high specific energy, low self-discharge rate, long life, and

environmental-friendly [1, 2]. It is widely used in electronic products, new electric vehicles, energy storage systems, aerospace, military communications, and other fields. The performance of the lithium-ion battery continually degrades with an increasing cycle of charge and discharge. If it is not replaced in time, it may cause the performance of the electrical equipment to decrease or even malfunction and shut down. The battery management system (BMS) evaluates the battery states and gives indications when there is a need to perform maintenance or replacement duty, which ensures the long-term stability and safe operation of the battery [3-6]. Both the state of health (SOH) [7-9] and remaining useful life (RUL) [10-13] are important indicators that is characterized by the degree of battery aging [14, 15]. SOH is generally expressed as the ratio between the current maximum practical capacity and the initial capacity of the battery. RUL is defined as the number of cycles remaining for the battery to decay from the current state to the end of life (EOL) [16, 17]. EOL is selected based on real-time usefulness of the battery. It is usually set as the time when the battery's SOH drops to 80% [18].

In recent years, researchers have conducted a lot of research on SOH estimation and RUL prediction [19, 20]. The main methods for SOH estimation and RUL prediction include the model-based and data-driven methods [21, 22]. The model-based methods can be divided into equivalent circuit models and electrochemical models using different modeling mechanisms. The equivalent circuit model ignores the complex physical and chemical processes to simulate the output effect of the battery using basic electrical circuit elements. Electrochemical model is a mechanism model established after fully considering the internal structure of the battery and its electrochemical reaction in the process of charge and discharge. It is an extremely complex model with many parameters. Qu et al. [23] proposed a neural network method that combines long and short-term memory (LSTM) networks with particle swarm optimization and method research for RUL prediction and SOH monitoring of lithium-ion batteries. Zhou et al. [24] proposed a lithium-ion battery SOH monitoring model framework based on a time convolution network (TCN). Model prediction and magnification are used in the model to improve the ability to capture local regeneration, thereby improving the overall accurate prediction of the model. Zhang et al. [25] proposed a deterioration modeling and considered the recovery phenomenon of lithium-ion battery affecting RUL prediction method, and established a deterioration model of lithium-ion battery considering recovery effect. The method based on the equivalent circuit model often has a small amount of calculation, but the model parameters are often determined under fixed conditions, which makes it only suitable for limited operating conditions. Also, compensation during the prediction increase the complexity of the model.

The data-driven method does not need to analyze the internal mechanism of the battery. It directly establishes the connection between the input characteristics and the battery capacity through historical data. In recent years, the data-driven method is widely used in battery state estimation and is divided into probabilistic and non-probabilistic methods. The probabilistic method such as artificial neural network (ANN), support vector machine (SVM) [26, 27], and autoregressive model (AR) [28-33], etc, is point estimation, and it establishes a relationship between the input and output values. However, these methods cannot establish a potential probability model and express the uncertainty in the estimated value. The importance of RUL estimation is not only to predict the RUL value but also to show the degree of uncertainty in the prediction.

The existing data-driven prediction methods have the disadvantage that they require a large amount of training sample data. Also, they are difficult to simulate the nonlinear characteristics of real capacity attenuation, and the prediction results do not give a range of prediction which leads to the results that are inaccurate. In this paper, the seasonal autoregressive integrated moving average model is used, which has attracted extensive attention because of its nonparametric and uncertain expression. Also, the proposed model offers an efficient machine learning method that makes predictions based on a very small amount of data, and the adaptability of the prediction results is more obvious than that of the ARIMA and closer to the real-time operation. The SARIMA is an extension of the traditional autoregressive integrated moving average (ARIMA) model. So, there is the need to remove the periodicity first. The method of removal is to conduct an ARIMA test at the periodic interval. Meanwhile, a non-stationary and non-periodic time series is obtained, and then used based on the analysis conducted by the ARIMA model.

2. MATHEMATICAL ANALYSIS

2.1. Forecasting process

To obtain the least number of parameter identification of the battery model and improve the prediction accuracy and efficiency. This paper adds the concept of sequence periodicity based on the ARIMA model and uses the SARIMA model to predict the RUL of lithium-ion batteries. The forecast flowchart is shown in Fig. 1.

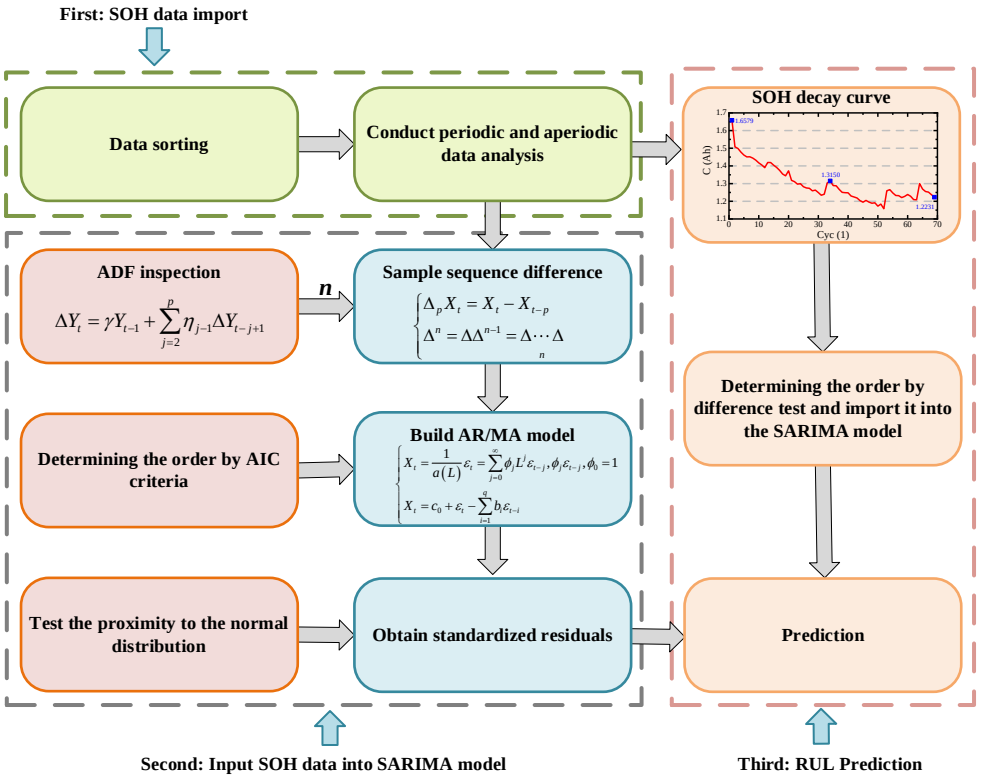


Figure 1. Forecasting flowchart

In Fig. 1, the RUL prediction is divided into three parts. Firstly, import the data, make a preliminary analysis of the periodicity and aperiodicity of the data. Secondly, the obtained SOH sequence is imported into the SARIMA model, the differential sequence stability is determined, and the residual test processing is performed. Finally, integrate the unprocessed data and the data processed by the model, and make predictions.

2.2. Covariance stationary

Covariance stationary is the stationarity defined based on the second moment of the time series. The first and second moments of weakly stationary time series do not change with time. If the second moment of the time series $\{y_t\}$ is finite, it satisfies Eq. (1).

$$\begin{cases} E(y_t) = E(y_{t-j}) = \mu \\ Var(y_t) = Var(y_{t-j}) = \sigma^2 \\ Cov(y_t, y_{t-s}) = Cov(y_{t-j}, y_{t-j-s}) = \gamma_s \end{cases} \tag{1}$$

For t, j, and s, where σ^2, μ, and γ_s are constants, the time series $\{y_t\}$ is said to be weakly stationary.

In the weakly stationary definition in Eq. (1), the first and second sub-forms indicate that its time series has a finite constant mean and covariance. The third sub-form indicates that the autocovariance of the weakly stationary time series is only related to the time lag s, and it has nothing to do with the starting position t of time. In summary, the first and second moments of a weakly stationary time series are constants with time.

Since the autocovariance of a weakly stationary time series is only related to the time lag s and not t. The autocovariance $\text{cov}(y_t,y_{t-s})$ in the form of the binary function in the third sub-form in Eq. (1) is written as the unary function y_s, which is related to the time lag s. In addition, the autocorrelation coefficient of a stationary time series is only related to the time lag. So, the autocorrelation coefficient of a stationary time series is abbreviated as the univariate function ρ_s which is also related to the time lag, as shown in Eq. (2).

$$\rho_{s=} \frac{\gamma_s}{\gamma_0} = \frac{\gamma_s}{\sigma^2} \tag{2}$$

Since the autocorrelation coefficient of a stationary time series is a function of the time lag s, the ρ_s is usually called the autocorrelation function or the autocovariance.

The stationarity of time series is the basic premise for many statistical operations in time series analysis. Therefore, the non-stationary time series data often need to be converted into stationary data. The stationarity process of the time series is random, and its unconditional joint probability distribution is not time-dependent, so the parameters such as the mean and variance of the stationary time series are also independent of time.

To transform the non-stationary process into a stationary process, the difference method is adopted, as shown in Eq. (3). The difference between the consecutive values of the time series, and the new time series data is obtained, which is tested to get the new correlations.

$$\begin{cases} \Delta_p X_t = X_t - X_{t-p} \\ \Delta^n = \Delta\Delta^{n-1} = \Delta\cdots\Delta \\ \qquad\qquad\quad n \\ \Delta_s X_t = X_t - X_{t-s} = \left(1 - L^s\right) X_t \end{cases} \tag{3}$$

In Eq. (3), Δ is the difference operator, Δp is the p-order difference, and Δs is the *s*-step difference. If there are multiple differences, the time series $\{\Delta n\}$ is expressed as a *n*-th order difference. Differential data can improve the accuracy of the signal and eliminate common error interference. The typical use of *s*-step difference is for periodic series to make the prediction result as close to the real-time operation as possible. For the time series, in addition to the *p*-order and the *s*-step difference alone, sometimes it is necessary to perform two difference operations jointly. The first-order four-step difference ($\Delta\Delta_4$) or the first-order twelve-step ($\Delta\Delta_{12}$) difference is commonly used in the time series modeling process.

2.3. Single root test

The difference of any sequence requires a suitable order. The excessive difference will result in the loss of effective information. Also, insufficient difference order will be inaccurate, and its data cannot meet the stationarity requirements. In this paper, the Augmented Dickey-Fuller (ADF) test is used to determine a reasonable difference order, as shown in Eq. (4).

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

In Eq. (4), α_j is the autoregressive coefficient, and ε_t is the random interference term. Both ends of the equal sign of the first sub-expression are subtracted from $Y_{t\text{-}1}$, and the second sub-expression is obtained after differential transformation. The first sub-expression is a linear differential equation. When γ=0, the corresponding characteristic equation has at least one unit root. Currently, the stationarity of the sequence $\{Y_t\}$ is a critical state, which is a non-stationary sequence. Therefore, for a new sequence, it is necessary to continue to differentiate until γ<0, which will make the sequence stationary. According to Eq. (4), when the original assumption of ADF is *H*0: γ=0, the original sequence is non-stationary. Therefore, the selected alternative hypothesis is *H*1: γ<0 to make the sequence stationary.

2.4. SARIMA model

There are several important parameters to be solved in the SARIMA model, which are the acyclic parameters *p, d,* and *q,* the cyclic parameters *P, D, and Q,* and the cycle period *s* of the time series.

2.4.1. Parameter order determination based on the AIC

The traditional method of parameter order identification uses autocorrelation and partial autocorrelation functions. This method is more challenging, and the recognition of complex data is not high, and the order identification is not inaccurate. This paper uses the AIC to directly identify the order to make up for the shortcomings of the traditional method, as shown in Eq. (5).

$$\begin{cases} AIC(p,q) = -2\ln[L] + 2r \approx N\ln\left(\hat{\sigma}_a^2\right) + 2r/N \\ AIC(p,q) = \min\limits_{1\le p,q\le M(N)} AIC(p,q) \end{cases} \tag{5}$$

In Eq. (5), $r=p+q$ is the number of independent parameters of the model; $\hat{\sigma}_a^2$ is the maximum likelihood estimation of residual variance. Due to the difficulty in solving the maximum likelihood estimation, the moment or least square estimation is commonly used in practical applications. The resulting residual variance approximates instead. The AIC function consists of two parts: the first part reflects the quality of the model simulation; the second part indicates the number of model parameters. The AIC function appropriately synthesizes the model fitting accuracy and the number of parameters in the model. When the model order increases, the first term in the AIC function gradually decreases. But for a given observation data N, the second term in the function increases with the order of the model. If the order of the model is increased to fit the data, the AIC value will show a downward trend. At this time, the first part plays a key role which results in the rapid decrease of the residual variance of the model. When a certain order is reached, the AIC value gets to a minimum. Then, as the order of the model continues to increase, the residual variance is slightly improved, so the second part plays a key role, and the value of AIC increases with the order of the model.

2.4.2. Residual test

In order to ensure the accuracy of the order, residual test is needed. The signal obtained by subtracting the fitting signal from the original signal is the residual signal.. If the residuals satisfy a random distribution and are not autocorrelated, it is a white noise signal. This shows that the useful signal has been extracted into the AR/MA model.. This paper used autocorrelation and partial autocorrelation functions to test the autocorrelation and partial autocorrelation of the residuals.

The autocorrelation function (ACF) is used to measure the correlation between observations of every k time unit (y_t and y_{t-k}) in a time series, as shown in Eq. (6).

$$\begin{cases} \rho_0 = \alpha_1\rho_1 + \sigma_\varepsilon^2/\gamma_0 - \beta_1(\alpha_1-\beta_1)\sigma_\varepsilon^2/\gamma_0 - \beta_2\left[\alpha_1(\alpha_1-\beta_1)+\alpha_2-\beta_2\right]\sigma_\varepsilon^2/\gamma_0 \\ \rho_1 = \alpha_1 - \beta_1\sigma_\varepsilon^2/\gamma_0 - \beta_2(\alpha_1-\beta_1)\sigma_\varepsilon^2/\gamma_0 \\ \rho_2 = \alpha_1\rho_1 - \beta_2\sigma_\varepsilon^2/\gamma_0 \\ \rho_3 = \alpha_1\rho_2 \\ \quad\vdots \\ \rho_s = \alpha_1\rho_{s-1} \end{cases} \tag{6}$$

In Eq. (6), according to the first two equations, the values of ρ_1 and γ_0 can be obtained. Using the obtained values of ρ_1 and γ_0, the third equation, ρ_2 is also be obtained. And all subsequent ρ_s ($s\geq3$) is calculated according to the recursive formula $\rho_s=\alpha_1\rho_{s-1}$.

The partial autocorrelation function (PACF) refers to the relevant measure of the impact of the lagged variable on the current variable under the condition of given random variables, or after eliminating the interference of the intermediate random variable, as shown in Eq. (7).

$$\begin{cases} \varphi_{ss}=\dfrac{D_s}{D} \\ D=\begin{vmatrix} \rho_0 & \rho_1 & \cdots & \rho_{s-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{s-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{s-1} & \rho_{s-2} & \cdots & \rho_0 \end{vmatrix}, D_s=\begin{vmatrix} \rho_0 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & \rho_0 & \cdots & \rho_2 \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{s-1} & \rho_{s-2} & \cdots & \rho_s \end{vmatrix} \end{cases} \tag{7}$$

Eq. (7) is a relational equation solved according to Cramer's law. Ψ_{ss} is the PACF with the lag number s, and ρ_i in the matrix is the autocorrelation coefficient with the lag i.

Finally, the Q-Q chart is used to test whether the residuals are close to the normal distribution, as shown in Eq. (8).

$$\begin{cases} y=f\left(\left(x-m\right)/std\right) \\ x=n\times std+m \end{cases} \tag{8}$$

In Eq. (8), $f(n)$ is the probability density function, m is the sample mean, and std is the sample standard deviation. In this paper, the Q-Q chart is used to analyze whether different data sets are of the same distribution. It is also used to verify whether the data set conforms to the normal distribution.

3. EXPERIMENTAL ANALYSIS

3.1. Construction of the experimental platform

In terms of experimental conditions, it has a reliable software testing platform and a stable hardware testing basis. The experiment is equipped with complete equipment which are the electronic load, DC power supply, three-layer independent temperature control, high and low temperature test bench, power battery module test system, simulated high and low voltage test bench, power battery high-rate test bench, charge and discharge tester, power battery drop test bench, and other supporting experimental equipment, as shown in Fig. 2.

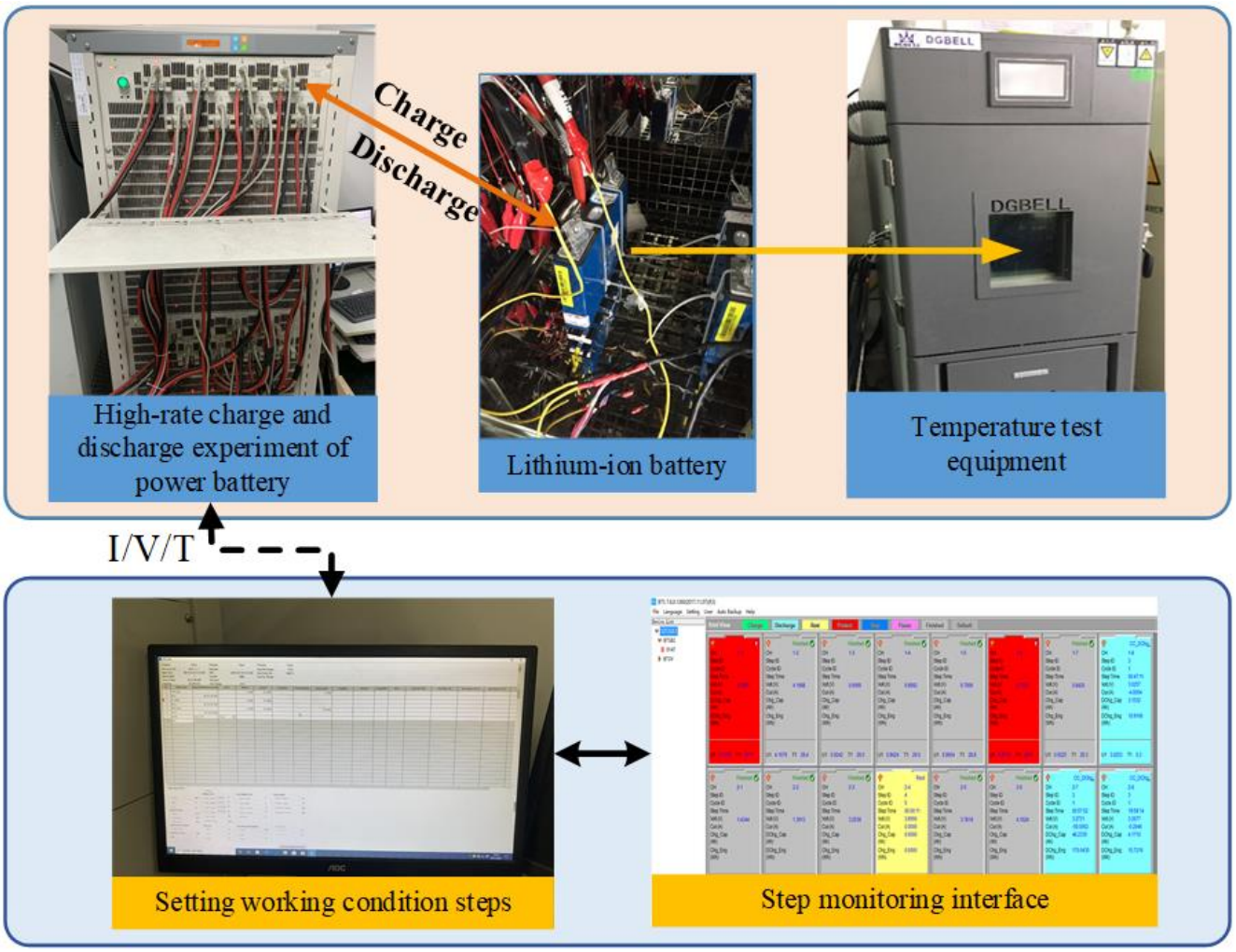


Figure 2. Experimental platform

As shown in Fig. 2, the lithium-ion battery is placed in a thermostat and connected to the power device through a high-current connection line. The experimental data of different working conditions are obtained by setting the experimental steps on the operations monitor. The experimental platform has automatic power cut-off protection measures to make it stable and reliable. It can obtain detailed experimental data over time, including current, voltage, capacity, temperature, energy, etc.

In this research, the algorithm is compared and verified in four different working conditions using four different batteries (B1, B2, B3, and B4). The experimental test procedure for each battery is as follows:

(1) Measure the lithium-ion battery (B1) with a capacity of 2.03Ah, discharge at a constant current of 2A until the battery voltage drops to 2.5V and 168 cycles.

(2) Measure the lithium-ion battery (*B*2) with a capacity of 1.66Ah, charge it in a 1.5A constant current mode. Then continue to charge in a constant voltage mode until the charging current drops to 20mA, discharge at 4A until the battery voltage drops to 2.0V, and 197 cycles.

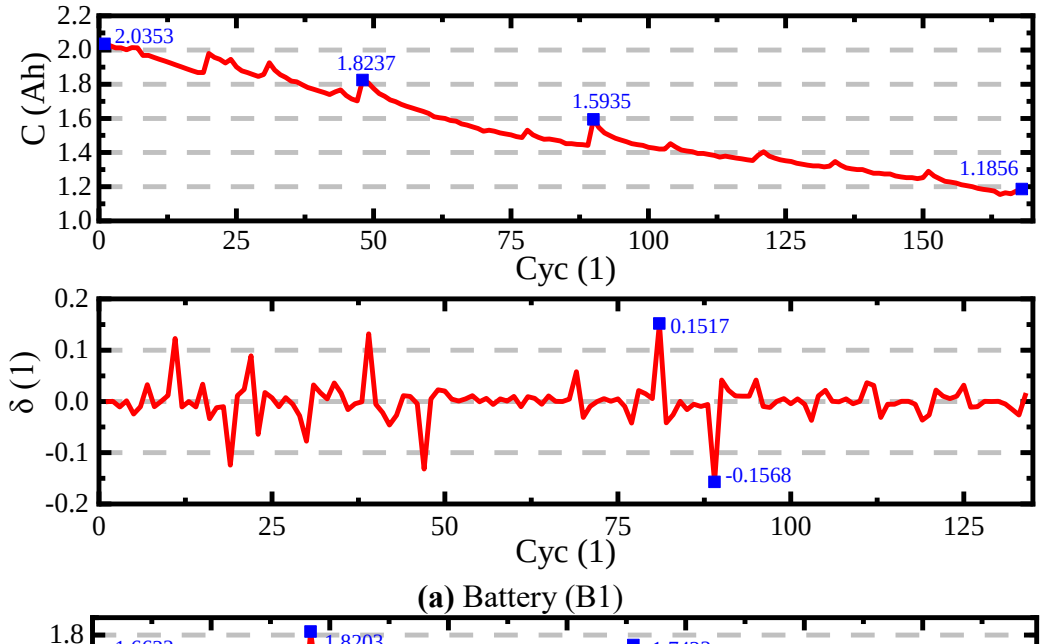
(3) Measure the lithium-ion battery (*B*3) with a capacity of 1.65Ah, charge it in a constant current mode of 1.5A until the battery voltage reaches 4.2V. Then, continue to charge in a constant

voltage mode until the charging current drops to 20mA, and use a fixed load current of 1A to discharge the voltage to 2.7V and 69 cycles.

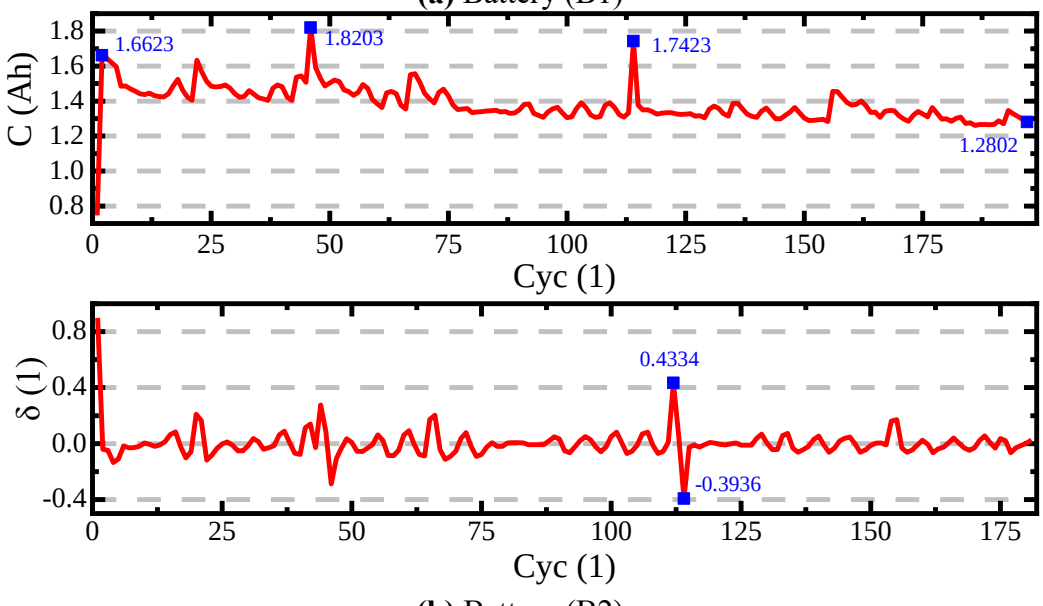
(4) Measure the lithium-ion battery (*B*4) with a capacity of 1.16Ah, charge it in a constant current mode of 1.5A until the battery voltage reaches 4.2V. Then, continue to charge in a constant voltage mode until the charging current drops to 20mA, and use a fixed load current of 2A is discharged to 2.2V and 97 cycles.

3.2. Stationary test

Before establishing the model, it is necessary to test the stability of the sequence. In the way of Eq. (1) to (3), the sequence is first differentiated and then tested by ADF.



(a) Battery (B1)



(b) Battery (B2)

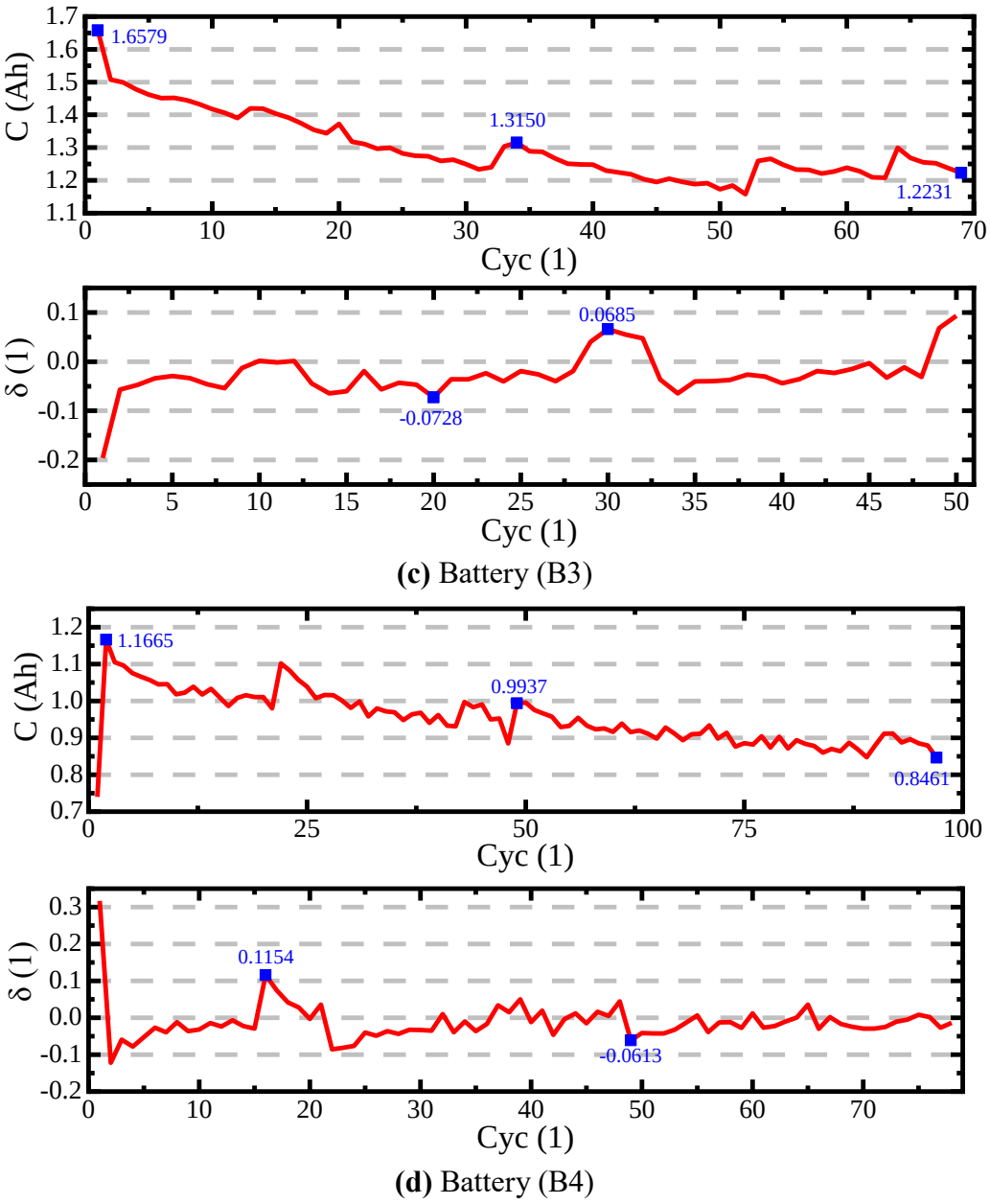


Figure 3. Capacity attenuation curve and differential sequence: **(a)** Battery (B1), **(b)** Battery (B2), **(c)** Battery (B3), **(d)** Battery (B4)

For the aperiodic part, it is usually from the order 0 to the order 3 differences. If the ADF test passes before order 3, the difference stops. If the ADF fails to pass the test after order 3, it is necessary to select accuracy. The difference beyond order 3 causes the loss of effective information of the sample sequence. Usually, for a periodic partial difference, the fixed value of order 1 is selected. The capacity attenuation curve and difference results are shown in Fig. 3.

Fig. 3 shows the different results for the four working conditions after passing the ADF test. B1 battery non-periodic difference order 1, periodic difference order 1; B2 battery non-periodic difference order 0, periodic difference order 1; B3 battery non-periodic difference order 0, periodic difference order 1; B4 battery non-periodic difference order 0, periodic difference order 1. It can be observed from Fig.

3 that the difference sequence is more stable than the original sequence and meets the modeling requirements.

3.3. Parameter order determination based on AIC

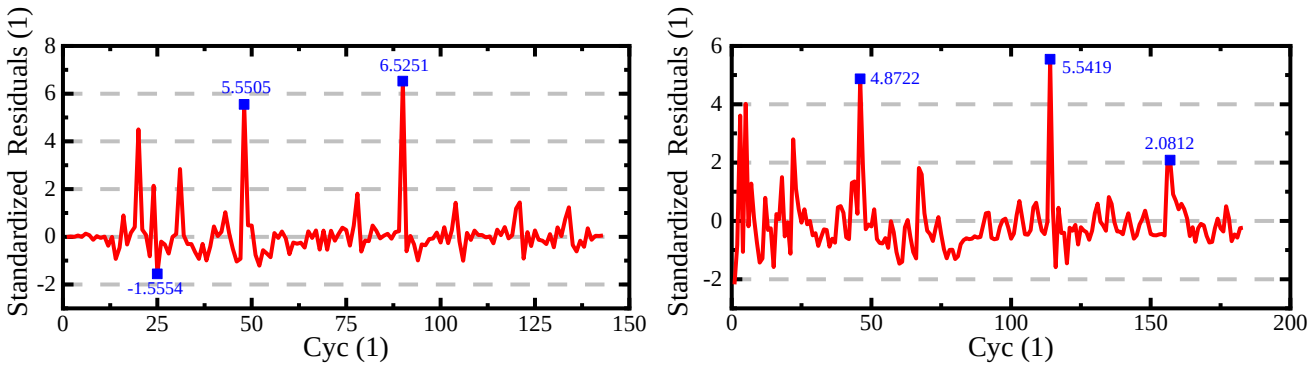
The AIC is used to find a model that best explains the data but contains the least free parameters. AIC considers the accuracy of data fitting and whether it is over-fitting. Therefore, the preferred model should be the one with the minimum AIC value. Calculate the AIC value of n models at a time and choose among them. The model corresponding to the smallest AIC value is the selected object. The AIC is used to determine the parameter order of the stationary sequence after the difference of the four working conditions. The order determination results are shown in Table 1.

Table 1. Parameter order determination of the SARIMA model

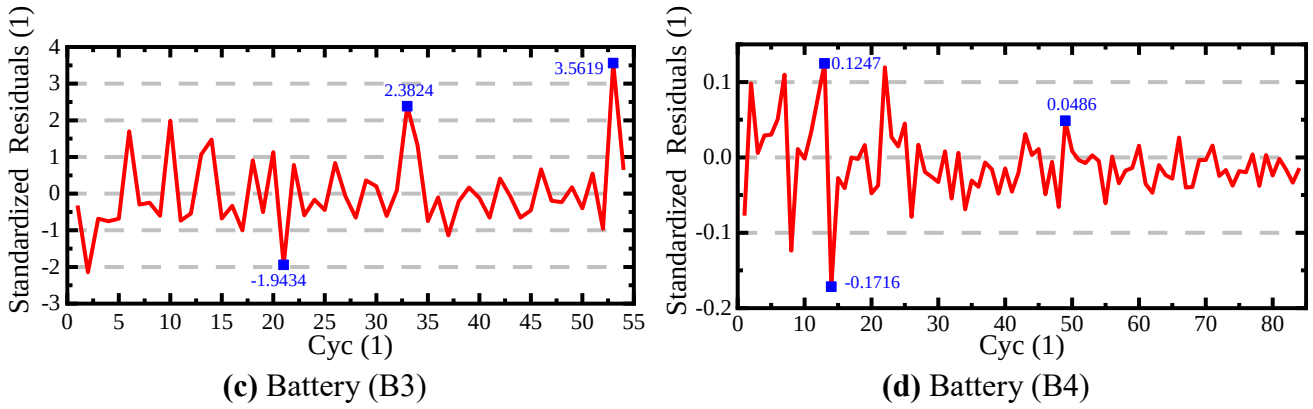
Battery	*p*	*d*	*q*	*P*	*D*	*Q*	*S*
*B*1	2	1	1	1	1	1	8
*B*2	3	0	3	1	1	2	2
*B*3	1	0	0	1	1	2	4
*B*4	1	0	0	1	1	2	6

In Table 1, *p* is the lags of the time series data used in the prediction model, also known as the autoregressive term (AR); *d* is how many order difference is required for the time-series data to achieve stability, also known as integrated term; *q* is the lags of the prediction error used in the prediction model, also known as the moving average (MA); *P* is the order of periodic autoregression; *D* is the order of periodic difference; *Q* is periodic moving average order; *S* is the cycle time interval.

3.4. Residual test results



(a) Battery (B1) **(b)** Battery (B2)



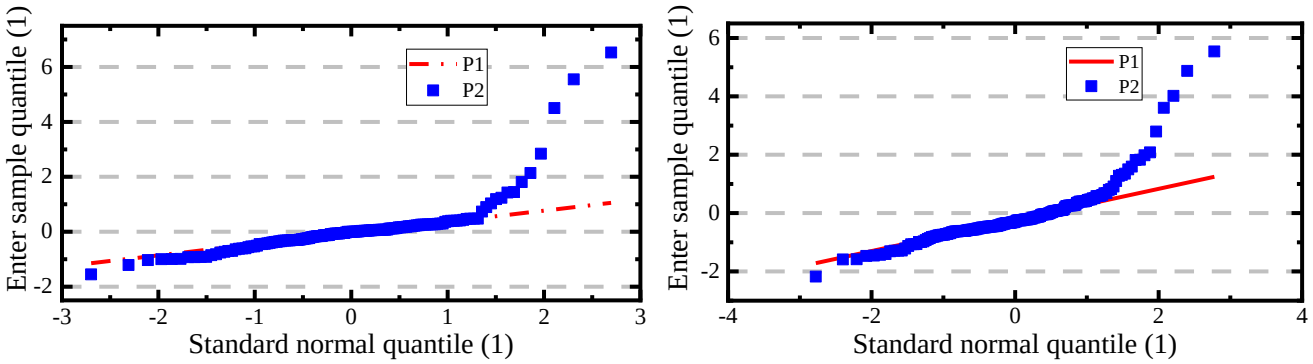
(c) Battery (B3) **(d)** Battery (B4)

Figure 4. Difference sequence residual results: **(a)** Battery (B1), **(b)** Battery (B2), **(c)** Battery (B3), **(d)** Battery (B4)

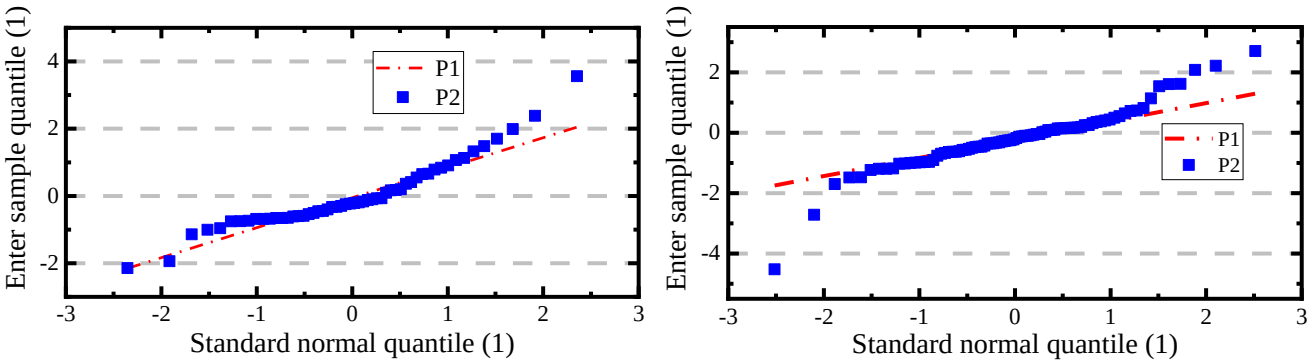
The residual test is carried out on each parameter to ensure that the determined order is accurate. The difference between the original signal and the model fitting signal is the residual. To pass the test, the residual is randomly distributed positively and uncorrelated. That is, the residual is a section of the white noise signal. The results in Fig. 4 show the residuals of the four working conditions.

The results in Fig. 4 show that the residuals of the four working conditions are almost white noise. To conduct the inspection precisely, the residual is further analyzed theoretically. Therefore, it is necessary to test whether the residual sequence is close to the positive distribution. This paper uses a Q-Q diagram for the test, with is a straight line *P*1: y = x in the Q-Q chart.

The two-column data of the quantile distribution is compared with that of the *P*1. The normal distribution test results are shown in Fig. 5. In an ideal condition, *P*2 should be close to the *P*1 for the results to meet the standard.

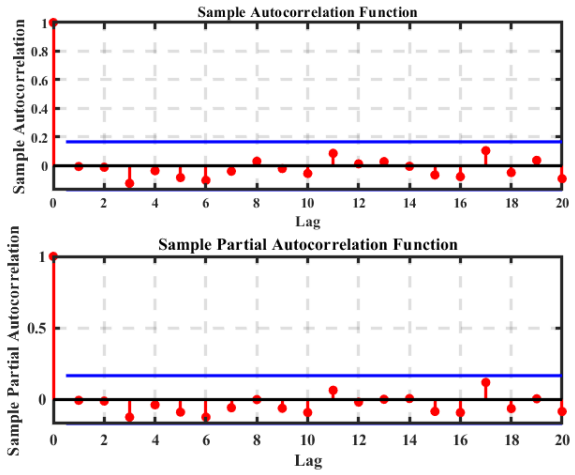


(a) Battery (B1) **(b)** Battery (B2)

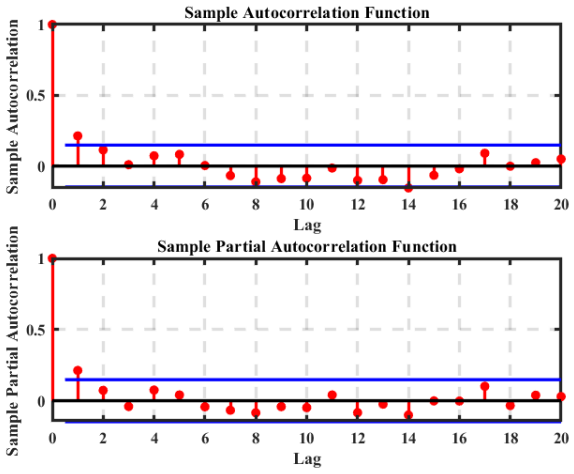


(c) Battery (B3) **(d)** Battery (B4)

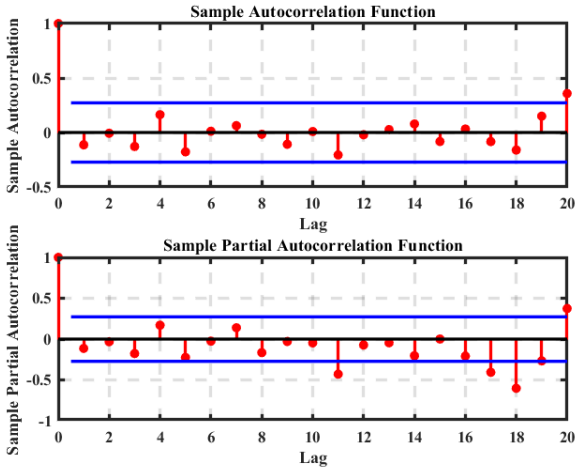
Figure 5. The residual normal distribution test: **(a)** Battery (B1), **(b)** Battery (B2), **(c)** Battery (B3), **(d)** Battery (B4)



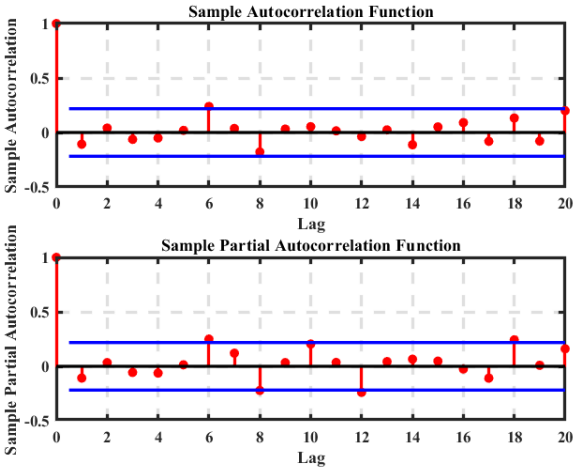
(a) Correlation test for battery (B1)



(b) Correlation test for battery (B2)



(c) Correlation test for battery (B3)



(d) Correlation test for battery (B4)

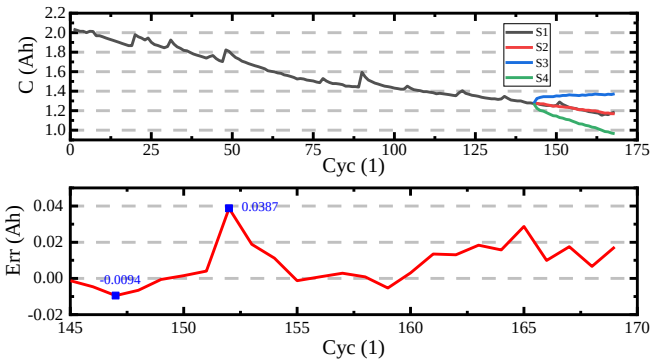
Figure 6. The ACF and PACF verification results for each battery: **(a)** Correlation test for battery (B1), **(b)** Correlation test for battery (B2), **(c)** Correlation test for battery (B3), **(d)** Correlation test for battery (B4)

Finally, the autocorrelation and partial autocorrelation of residuals are tested, but there is no correlation in the ideal residual sequence. The ACF reflects the correlation between the values of the same sequence in different time series. The PACF is the correlation degree between two variables after eliminating the interference of intermediate variables.

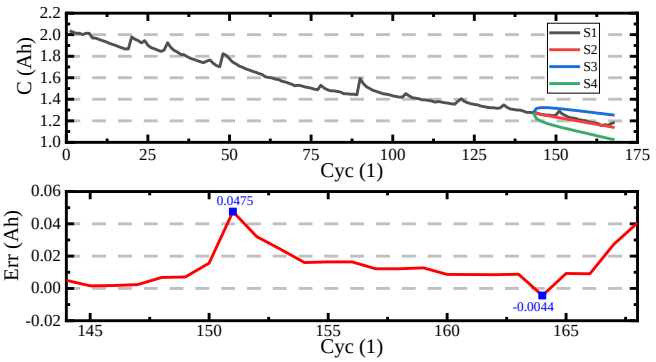
The correlation test results are shown in Fig. 6, and it can be observed that they meet the standard because the ideal results of the ACF and PACF inspections do not have patterns beyond the upper and lower confidence limits.

3.5. Predictive verification

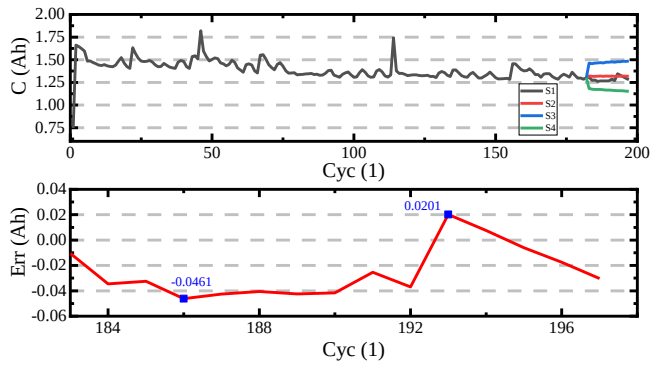
The prediction length of the SARIMA model is usually 5–15% of the sample length. A prediction that takes much computational time leads to error accumulation. If the sample sequence shows strong periodicity, the prediction length is extended appropriately. In this paper, the first 80–90% of the data of the sample sequence is used as the training set. The error between the predicted and real values are compared and verified with the prediction error of the ARIMA model without periodic parameters. The prediction results of the four working conditions are shown in Fig. 7.



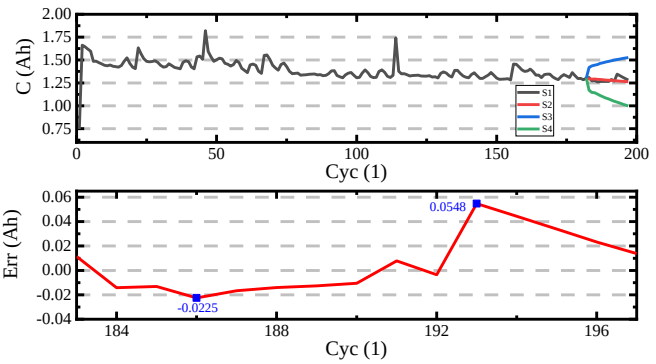
(a) SARIMA prediction results for battery (B1)



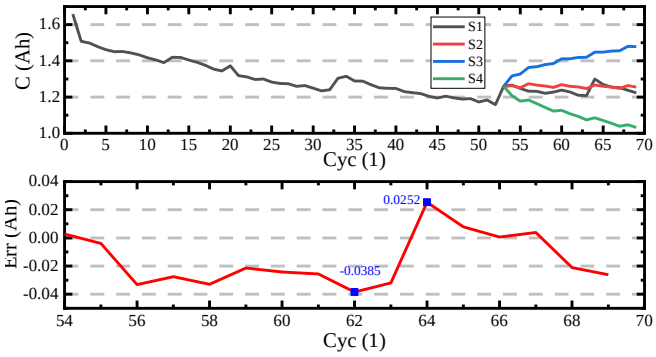
(b) ARIMA prediction results for battery (B1)



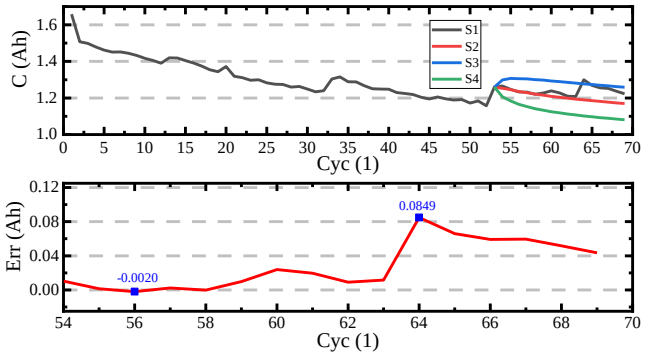
(c) SARIMA prediction results for battery (B2)



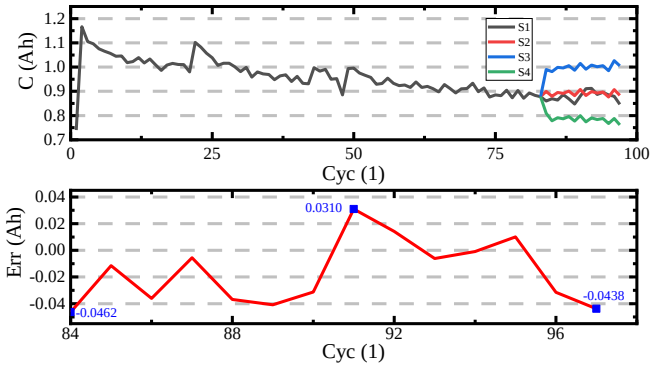
(d) ARIMA prediction results for battery (B2)



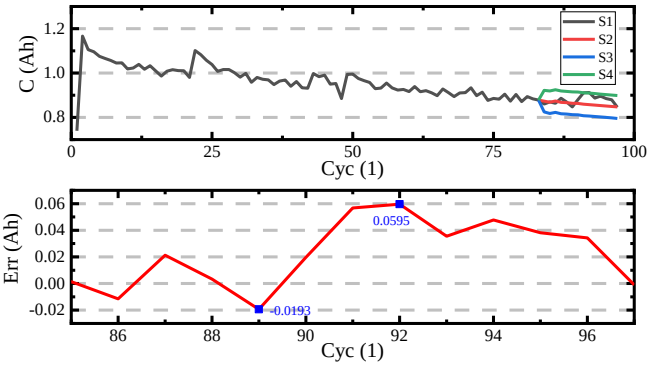
(e) SARIMA prediction results for battery (B3)



(f) ARIMA prediction results for battery (B3)



(g) SARIMA prediction results for battery (B4)



(h) ARIMA prediction results for battery (B4)

Figure 7. Comparison of prediction results between the SARIMA and ARIMA for each battery: **(a)** SARIMA prediction results for battery (B1), **(b)** ARIMA prediction results for battery (B1), **(c)** SARIMA prediction results for battery (B2), **(d)** ARIMA prediction results for battery (B2), **(e)** SARIMA prediction results for battery (B3), **(f)** ARIMA prediction results for battery (B3), **(g)** SARIMA prediction results for battery (B4), **(h)** ARIMA prediction results for battery (B4)

In Fig. 7, *S*1 is the real value of battery capacity attenuation, and *S*2 is the predicted value, the 10–20% of the data after the predicted sample sequence is compared with the real value. An error diagram is established, *S*3 and *S*4 are the 95% confidence bounds of the predicted value. Fig. 7 (a, c, e, g) shows the prediction results and errors of SARIMA. Fig. 7 (b, d, f, e) shows the prediction results and errors of ARIMA. The SARIMA prediction results can simulate the fluctuation of the real value, but the ARIMA can only show the attenuation trend of battery capacity and cannot simulate the fluctuation of real value. The results for the error analysis are shown in Table 2.

Table 2 shows the prediction results obtained using the proposed SARIMA model and the traditional ARIMA model under the four working conditions. The maximum error for prediction of the proposed SARIMA model for batteries B1, B2, B3, and B4 are 3.87%, 4.61%, 3.85%, and 4.62%, respectively. Also, the maximum error for the prediction of the traditional ARIMA model for batteries B1, B2, B3, and B4 are 4.75%, 5.48%, 8.49%, and 5.95%, respectively. It can be observed that the prediction results of the SARIMA are closer to the real data, with a maximum error of 4.62%. From

Refs. [23,24,25], Qu et al. [23] had the maximum prediction error of RUL to be 13%, Zhou et al. [24] had the maximum prediction error of RUL to be 8%, and Zhang et al. [25] had the maximum prediction error of RUL to be 6.87%. Comparing the maximum prediction error in this paper with Ref. [25], the proposed SARIMA model reduces the error by 2.25%.

Table 2. Parameter order determination of SARIMA model

Battery	SARIMA		ARIMA	
	Prediction steps	Maximum error	Prediction steps	Maximum error
*B*1	25	3.87%	25	4.75%
*B*2	15	4.61%	15	5.48%
*B*3	16	3.85%	16	8.49%
*B*4	15	4.62%	15	5.95%

4. CONCLUSIONS

Due to the internal electrochemical reaction characteristics of lithium-ion batteries, the change of capacity in the degradation is a nonlinear process, which makes it difficult to predict the RUL accurately. To solve this problem, this paper uses the SARIMA prediction model to replace the traditional ARIMA model. Firstly, the difference method is used to process the sample sequence to be stable. Secondly, the order of the model parameters is determined based on the maximum likelihood estimation and AIC, and the periodic parameters are introduced to optimize and fit the nonlinear characteristics of the lithium-ion batteries in real-time operations. Finally, the distribution of the sequence is comprehensively tested through the ACF, PACF, and Q-Q diagrams to optimize the establishment of the prediction model. The four working conditions set in this paper take into account both high and low rate charge and discharge. The maximum prediction error under these four complex working conditions is 4.62%, which is a 2.25% improvement in the RUL prediction accuracy. The advantages of this proposed model are that it can simulate the nonlinear characteristics of lithium-ion batteries in real-time operations and give the confidence bounds to further limit reduce the prediction error range.

ACKNOWLEDGMENTS
The work was supported by the National Natural Science Foundation of China (No. 61801407), Sichuan Science and Technology Program (No. 2019YFG0427), China Scholarship Council (No. 201908515099), Fund of Robot Technology used for Special Environment Key Laboratory of Sichuan Province (No. 18kftk03), National Training Program of Innovation and Entrepreneurship for Undergraduates (No. 202110619017), and Southwest University of Science and Technology Student Innovation Fund Project (No. CX21-018), National College Student Innovation and Entrepreneurship Training Program (No. S202110619017).

References

1. T. L. Kulova, V. N. Fateev, E. A. Seregina and A. S. Grigoriev, *Int J Electrochem Sci.*, 15 (2020) 7242.
2. K. L. Liu, K. Li, Q. Peng and C. Zhang, *Front Mech Eng-Prc*, 14 (2019) 47.
3. Y. J. Wang, J. Q. Tian, Z. D. Sun, L. Wang, R. L. Xu, M. C. Li and Z. H. Chen, *Renew Sust Energ Rev.*, 131 (2020) 110015.
4. S. Y. Chen, Z. H. Gao and T. J. Sun, *Energy Sci Eng.*, 9 (2021) 5.
5. Q. K. Wang, J. N. Shen, Y. J. He and Z. F. Ma, *Chinese Phys B*, 29 (2020) 068201.
6. W. H. Li, M. Rentemeister, J. Badeda, D. Jost, D. Schulte and D. U. Sauer, *J Energy Storage*, 30 (2020) 101557.
7. R. Xiong, L. L. Li and J. P. Tian, *J Power Sources*, 405 (2018) 18-29.
8. Y. Y. Li, D. I. Stroe, Y. H. Cheng, H. M. Sheng, X. Sui and R. Teodorescu, *J Energy Storage*, 33 (2021) 102122.
9. B. Balasingam, M. Ahmed and K. Pattipati, *Energies*, 13 (2020) 2825.
10. J. H. Huang, S. L. Wang, W. H. Xu, C. Fernandez, Y. C. Fan and X. P. Chen, *Int J Electrochem Sci.*, 16 (2021) 21075.
11. C. F. Pan, Y. Chen, L. M. Wang and Z. G. He, *Int J Electrochem Sic.*, 14 (2019) 9537.
12. D. X. Shen, T. T. Xu, L. F. Wu and Y. Guan, *IEEE Access*, 7 (2019) 130638.
13. Y. Y. Ma, D. X. Shen, L. F. Wu, Y. Guan and H. Zhu, *Int J Electrochem Sci.*, 14 (2019) 7737.
14. M. F. Ge, Y. B. Liu, X. X. Jiang and J. Liu, *Measurement*, 174 (2021) 109057.
15. R. Xiong, Y. Z. Zhang, J. Wang, H. W. He, S. M. Peng and M. Pecht, *IEEE T Veh Technol.*, 68 (2019) 4110.
16. Y. Kotak, C. M. Fernandez, L. C. Casals, B. S. Kotak, D. Koch, C. Geisbauer, L. Trilla, A. Gomez-Nunez and H. G. Schweiger, *Energies*, 14 (2021) 2217.
17. A. Rastegarpanah, J. Hathaway and R. Stolkin, *Energies*, 14 (2021) 2597.
18. Y. Gao, R. J. Huang, D. C. Qin, T. T. Wang, S. B. Ma and S. Qin, *Int J Electrochem Sci.*, 16 (2021) 210424.
19. Z. Q. Lyu, R. J. Gao and L. Chen, *IEEE T Power Elect.r*, 36 (2021) 6228.
20. Y. Z. Zhang, R. Xiong, H. W. He, X. B. Qu and M. Pecht, *Etransportation*, 1 (2019) 100004.
21. S. S. Sheikh, M. Anjum, M. A. Khan, S. A. Hassan, H. A. Khalid, A. Gastli and L. Ben-Brahim, *Energies*, 13 (2020) 3658.
22. Y. C. Song, D. T. Liu, H. T. Liao and Y. Peng, *Appl Energy*, 261 (2020) 114408.
23. J. T. Qu, F. Liu, Y. X. Ma and J. M. Fan, *IEEE Access*, 7 (2019) 87178.
24. D. H. Zhou, Z. Y. Li, J. L. Zhu, H. C. Zhang and L. Hou, *IEEE Access*, 8 (2020) 53307.
25. Z. Y. Zhang, Z. Peng, Y. Guan and L. F. Wu, *Int J Electrochem Sci.*, 15 (2020) 8674.
26. J. W. Wei, G. Z. Dong and Z. H. Chen, *IEEE T Ind Electron*, 65 (2018) 5634.
27. X. J. Tan, Y. Q. Tan, D. Zhan, Z. Yu, Y. Q. Fan, J. Z. Qiu and J. Li, *IEEE Access*, 8 (2020) 56811.
28. D. M. Zhou, A. Al-Durra, K. Zhang, A. Ravey and F. Gao, *J Power Sources*, 399 (2018) 314.
29. Z. Chen, Q. Xue, R. X. Xiao, Y. G. Liu and J. W. Shen, *IEEE Access*, 7 (2019) 102662.
30. J. Park, M. Lee, G. Kim, S. Park and J. Kim, *Energies*, 13 (2020) 2138.
31. S. N. Haider, Q. C. Zhao and X. L. Li, *Energies*, 13 (2020) 1085.
32. Z. H. Yun, W. H. Qin, W. P. Shi and P. Ping, *Energies*, 13 (2020) 4858.
33. J. H. Huang, S. L. Wang, W. H. Xu, W. H. Shi and C. Fernandez, *Processes*, 9 (2021) 795.