# Application of Time Series Prediction Models in Chinese Electricity Consumption Analysis

Yisheng Gao *, Shaoxuan Xie

School of Management, Shandong University of Technology, Zibo, China, 255012

* Corresponding Author Email: XSX1747931952@163.com

**Abstract.** Study focuses on the application of time series prediction models in analyzing electricity consumption in China, aiming to promote the intelligentization and sustainable development of the country's power system. The ARIMA model was chosen for this research, incorporating historical electricity data to forecast future consumption. Through a series of steps including data preprocessing, model construction, and prediction analysis, the results indicate a general growth trend in electricity consumption, influenced by seasonal and cyclical factors. By plotting periodic decomposition charts, the study visually reveals the components and patterns of electricity data, providing scientific decision-making support for the planning, scheduling, and management of the electricity industry.

**Keywords:** Time series prediction, ARIMA model, electricity consumption forecast, cyclical factors, seasonal decomposition.

## 1. Introduction

With the transformation of the global energy structure and the pursuit of sustainable development goals, the stability and efficiency of electricity supply, as the cornerstone of modern society, have become increasingly important. China, as the world's largest electricity consumer and producer, plays a pivotal role in the global energy landscape. Against this backdrop, the application of time series prediction models in Chinese electricity consumption analysis not only enhances the intelligence of China's power system but also contributes to the sustainable development of the global electricity industry.

Time series prediction models, as vital analytical tools, can reveal underlying patterns through the exploration and analysis of historical data, enabling predictions of future trends. In the electricity sector, these models are widely used in electricity load forecasting, electricity price prediction, renewable energy output prediction, and other areas, providing powerful decision support for power system planning, scheduling, and operations.

Despite significant achievements in the application of time series prediction models in the electricity sector, challenges and issues remain. For instance, addressing the nonlinearity and uncertainty of data, enhancing model generalization and robustness, and integrating models with other technologies to improve prediction accuracy are ongoing concerns. This paper aims to delve into the latest applications and existing issues of time series prediction models in the electricity sector through an in-depth study on their application in China. The findings of this study offer valuable insights for enhancing the intelligence and sustainable development of China's power system. Additionally, this paper contributes Chinese wisdom and solutions to the sustainable development of the global electricity industry through continuous optimization and innovation of time series prediction models.

## 2. Literature Review

In recent years, time series analysis has been widely applied in various fields, including slope stability, pharmaceutical procurement decision-making, global temperature prediction, and crop planting area forecasting. Specifically, the ARIMA model, as a classic time series analysis technique, has been utilized in numerous studies for prediction and data analysis. For instance, Hu [1] employed the ARIMA model to predict the change in slope safety factors with the duration of rainfall. Yan [2] combined the ARIMA model with a CNN-LSTM model to improve temperature prediction accuracy. Zhao [3] achieved satisfactory results in predicting crop planting areas using an ARIMA-GM combined model. Ni [4] combined the ARIMA model with an LSTM model to overcome the limitations of the ARIMA model in handling nonlinear patterns and long-term dependencies.

Regarding electricity consumption forecasting, Jia [5] classified and analyzed internal and external factors affecting electricity consumption in Beichuan County. They applied adaptive and self-learning electricity consumption prediction methods to complex nonlinear mappings, enhancing prediction accuracy. Liu [6] utilized a two-layer LSTM neural network with ReLU activation functions in the hidden layers and a linear activation function in the output layer. The model parameters were trained using the Adam optimizer. Shen [7] proposed a second-order Lagrange interpolation method and an INFLO-based anomaly detection method for electricity consumption dataset prediction. From a decomposition perspective, they also introduced a time series decomposition method based on the Prophet model, decomposing daily electricity consumption into trend, seasonal, and holiday components. Yan [8] introduced filter theory into the field of electricity consumption prediction. They optimized the prediction using particle filtering, established a state transition equation based on parameters obtained from the ARIMA model, and employed the Monte Carlo method to obtain particles. By adding noise to the particles, a particle set was formed, ultimately yielding an optimal estimate of actual electricity consumption.

## 3. Model Development

The approach employed in this study for electricity consumption forecasting is outlined in Figure 1. It comprises three primary steps: data preprocessing, model establishment, and prediction. The input to this method is monthly electricity consumption data collected from 2018 to 2023, and the output is a predictive model of electricity consumption cycles.

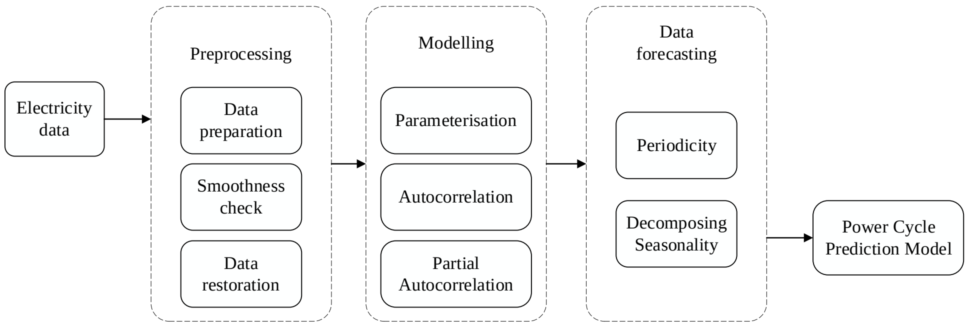


**Figure 1.** Electricity forecasting steps

### 3.1. Introduction to the basic concepts of modelling

ARIMA model is a well-known time series prediction method, also known as the Autoregressive Integrated Moving Average Model (ARIMA). This model was first proposed by Box and Jenkins in the 1970s [9] and is a statistical model used for analyzing and predicting time series data. The ARIMA model combines the characteristics of the Autoregressive model (AR) and the Moving Average model (MA), and can make non-stationary time series stationary through Differencing (I).

The specific form of the ARIMA model is ARIMA (p, d, q), where:

-p is the order of the autoregressive term, indicating the number of autoregressive terms included in the model. The autoregressive term is a linear combination of the time series data and its own past values.

-d is the order of differencing, indicating the number of differencing performed to make the time series data stationary.

-q is the order of the moving average term, indicating the number of moving average terms included in the model. The moving average term is a linear combination of the past error terms of the time series data.

ARIMA models are very popular in time series analysis because of their ability to capture trends, seasonality and stochasticity in data and can be used for both short and long term forecasting, the model estimates the parameters of the model by minimising the forecasting error and uses these parameters to predict future data points.

### 3.2. ARIMA Modelling

The general model of artificial neural network consists of four basic elements, which are:

(1) The BP neural network is linked by different node coefficients. When connecting weights and weights are positive, it indicates that the current link is an exciting state. Conversely, if the link coefficient is negative, the link state is a state of suppression.

(2) The input signal and the linear signal are the combination of the signals for each input signal.

(3) The function of the nonlinear activation function: making the neuron output signal within a certain range.

An autoregressive (AR) model is a model that predicts future values by using past values of the time series data itself. In an AR model, the value of a point in time is represented as a linear combination of its own past values, plus a random error term. This model assumes a linear relationship between the current value of the time series and its past values. In this case, the p-order autoregressive formula is shown below:

$$y_t = \mu + \sum_{i=1}^{P} \gamma_i \gamma_{t-i} + \in_t \tag{1}$$

Where $y_t$ denotes the current quarter's electricity, $\mu$ is a constant term, P is the order, $\gamma_i$ is the autocorrelation coefficient, and $\in_t$ is the error.

The Moving Average (MA) model is a model that predicts future values by using the past error terms of time series data. In the MA model, the value of a point in time is represented as a linear combination of its own past error terms. This model assumes a linear relationship between the current value of the time series and its past error terms. In this case, the q-order autoregressive formula is shown below:

$$y_t = \mu + \in_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{2}$$

After fitting the data on the basis of model construction, it can be analysed according to the comparison between the predicted value and the actual value, i.e. to determine the weights through the error, and the commonly used methods are fitting error, logarithmic error and relative error [10]. In this experiment, the average absolute percentage error and root mean square error (RMSE) are used as the evaluation criteria, and the weighted moving average function can be carried out subsequently.

Both models are linear and are based on the assumption of smoothness of the time series. In practical applications, it is usually necessary to choose the appropriate model according to the characteristics of the data and the analysis objectives.

#### 3.2.1 Model Diagnosis

The Augmented Dickey-Fuller Test (ADF test) is a statistical method used to test whether time series data is stationary. If the data is non-stationary, it usually needs to be stabilized through differencing or other methods. The null hypothesis of the ADF test is that the time series has a unit root, that is, the data is non-stationary. If the value of the test statistic is less than a critical value (usually determined based on the sample size and significance level), the null hypothesis is rejected, and the data is considered stationary.

The autocorrelation test is used to assess the correlation between time series data at different time points. If the data has autocorrelation, that is, a time point's value is related to its past or future values, this needs to be considered when selecting a model. Common autocorrelation test methods include calculating the autocorrelation coefficient and plotting the autocorrelation diagram (ACF diagram). If the autocorrelation coefficient is significantly non-zero, or the ACF diagram shows an obvious pattern, then the data has autocorrelation.

The partial autocorrelation test is used to assess the direct correlation of time series data after eliminating the influence of intermediate variables. Unlike the autocorrelation test, the partial autocorrelation test considers the influence of other variables on the relationship between two time points. A common partial autocorrelation test method is to plot the partial autocorrelation diagram (PACF diagram). The PACF diagram shows the direct correlation between a time point's value and another time point's value after eliminating the influence of intermediate variables.

### 3.2.2 The periodicity test

The periodicity test is mainly used in time series analysis to determine whether there are cyclical variations or seasonal patterns in the data. In the case of cyclical changes in electricity, they may be caused by external factors (such as seasonal changes, policy adjustments, etc.) as well as internal factors (such as economic cycles, production cycles, etc.). There are many methods of periodicity test, in this experiment, the method of Fourier Transform is adopted, through which the time series is converted from time domain to frequency domain, and the spectrogram is observed to determine whether there is a significant periodic frequency or not.

### 3.2.3 Seasonal Decomposition

In time series analysis, seasonal decomposition is a method used to identify and separate seasonal components in time series data. Seasonality usually refers to a repeating pattern of data over the course of a year or a fixed period of time, such as seasonal variations in quarterly sales data. Seasonality decomposition is based on either an additive or multiplicative model of the time series. In an additive model, the time series is decomposed into trend, seasonality and random components. Whereas in the multiplicative model, the trend and seasonal components are multiplied together with the stochastic component. In this experiment, the additive model was adopted to decompose the seasonality using a two-sided filter with 28 days as the seasonal period of the time series.

# 4. Model Solution

## 4.1. Data preparation

### 4.1.1 Source of data

The dataset used in this paper comes from China's monthly social electricity consumption data from February 2018 to November 2023, which was collected using the Electricity Consumption Information Collection System (ECCS) at monthly intervals for a total of 90 data entries. The characteristics of the input data show cyclical variations with an overall increasing trend.

After extracting the data, the training and testing sets are divided according to the ratio of 8/2 and plotted to view the features of the training and testing sets, the training and testing set features are shown in Figure 2.

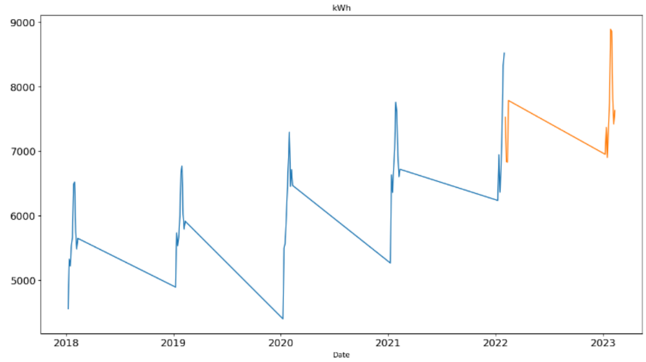


**Figure 2.** Training set and test set image features

The images of the training set and the test set can reflect that China's electricity volume shows certain cyclical changes, and shows an increasing and then decreasing trend in the unit cycle, with an overall upward trend.

### 4.1.2 smoothness check

By performing the ADF test on the data, it was detected that the p-value is much less than 0.05 and the data tends to be stable, if there is an unsteady data, the data can be differentiated in multiple orders and then re-diagnosed, the timing diagram of the first-order differentiation of the data is shown in Figure 3.

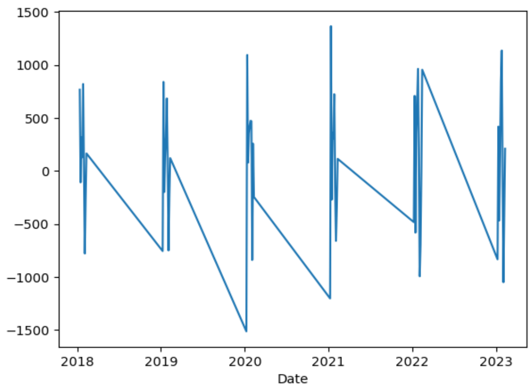


**Figure 3.** First-order differential timing diagram for power data

## 4.2. Evaluation Metric

The evaluation metric used in this experiment is the Root Mean Squared Error (RMSE), which is the square root of the Mean Squared Error (MSE), a commonly used metric for evaluating regression models to measure the deviation of the predicted values from the actual observed values. The RMSE is obtained by calculating the square of the prediction error of the mean of the prediction error and then taking its square root to obtain it. Specifically, it is calculated as follows:

1) Calculate the square of the error: for each observation of electricity, calculate the difference between its predicted value and the actual value, and square the difference. Doing so amplifies larger errors, making the model more sensitive to larger deviations.

2) Calculate the average error squared: add up the error squares of all observation points and divide by the total number of observation points to get the average error squared.

3) Take the square root: Take the square root of the mean error squared to get the RMSE.

## 4.3. Analysis of the model solution

### 4.3.1 Moving Average Function

The Moving Average Function (MA) with RMSE as the evaluation metric is shown in Figure 4. and the RMSE index is 680.412.


**Figure 4.** Simple moving average function (MA)

### 4.3.2 ADF, Autocorrelation, Partial Autocorrelation test

After the ADF test, the p-value is much less than 0.05, as shown in Figure. 5. Therefore the original hypothesis can be rejected with a high degree of confidence and the time series is considered to be smooth.

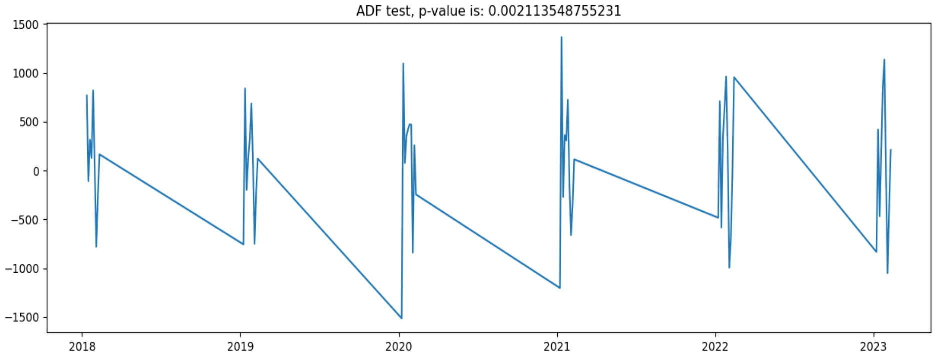


**Figure 5.** ADF Test

By observation, the ACF(q) plot slowly decreases and the PACF(p) plot truncates after order 1, as shown in Figure. 6.

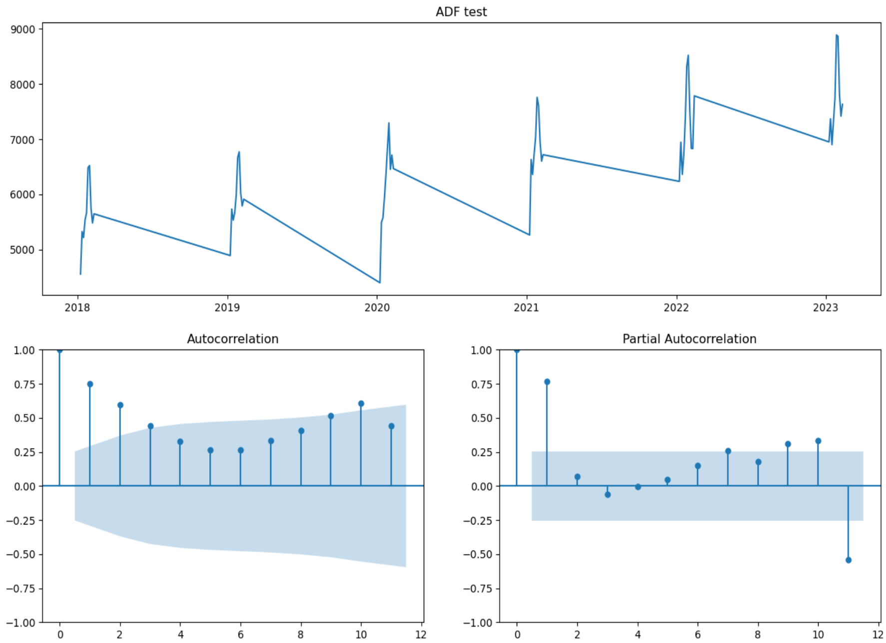


**Figure 6.** Autocorrelation, partial autocorrelation analysis plot

### 4.3.3 White Noise Test

The p-values obtained after the white noise test are shown in Table 1 and by looking at the p-values in the table, all the values are much less than 0.01 and even much less than 0.00001. This reflects very confident rejection of the original hypothesis that these data are not white noise. That is, there is autocorrelation in these data and they are not random.

**Table 1.** White Noise test

| | lb_stat | lb_pvalue |
|---|---|---|
| 1 | 415.912602 | 1.892897e-92 |
| 2 | 689.944556 | 1.515107e-150 |
| 3 | 881.568934 | 8.806397e-191 |
| 4 | 1038.361810 | 1.732815e-223 |
| 5 | 1189.549572 | 5.389891e-255 |
| 6 | 1355.411712 | 1.092498e-289 |

### 4.3.4 Cyclical modelling

The Fourier transform spectrogram is shown in Figure. 7., which shows that the spectrum of this signal is relatively flat over most of the frequency range, with no obvious peaks or fluctuations, which indicates that the energy of the signal is more uniformly distributed in the low-frequency region.

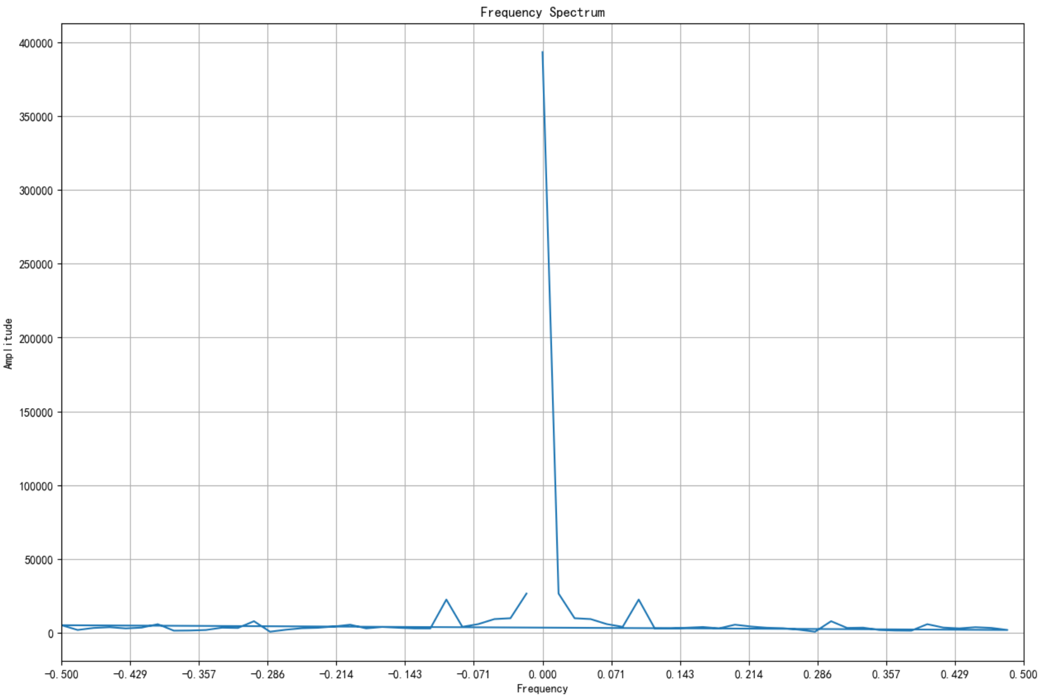


**Figure 7.** Fourier transform spectrograms

The seasonal decomposition of electricity is shown in Figure. 8, where the chart illustrates the trend, seasonality and residual components of the electricity data. It helps to better understand the cyclical structure of the electricity volume data.

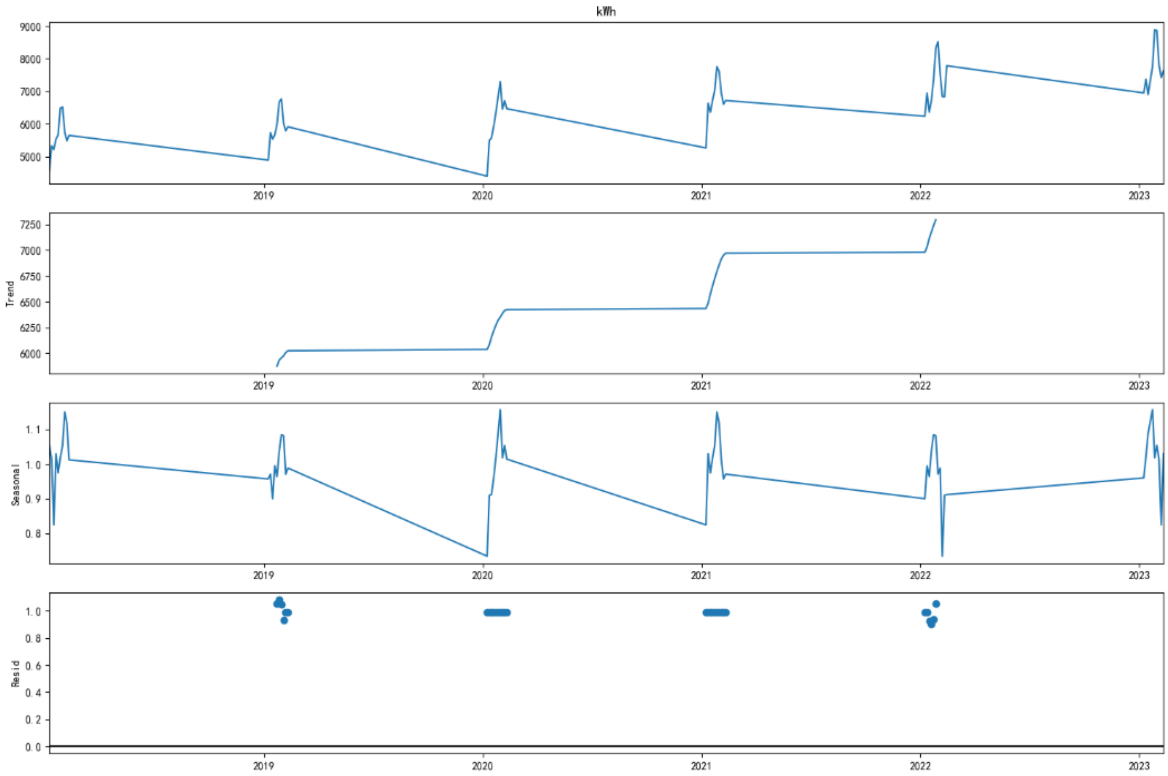


**Figure 8.** Trend, Seasonality, and Residual plots

### 4.4. Conclusion of the experimental part

Through the time-series analysis of China's electricity data and the drawing of cyclical decomposition maps, the study finds that the overall electricity volume shows a growing trend, which is affected by seasonal and cyclical factors at the same time. Cyclical decomposition mapping helps to clearly show the components and changing patterns of electricity data. Through the mapping, the impact of trend, seasonal and cyclical components on the raw data can be visually observed, providing policymakers with strong data support and analysis basis. These findings contribute to a deeper understanding of the changing patterns of China's electricity and provide a scientific basis for the planning, dispatching and management of the power industry.

## 5. Conclusion

Through time series analysis of China's electricity data and cyclical decomposition mapping, the study reveals the trend of overall electricity growth and points out the impact of seasonal and cyclical factors on electricity. The clear presentation of the cyclical decomposition mapping provides strong data support and analysis basis for future policy makers. These findings deepen the understanding of the changing patterns of power generation in China and provide a scientific basis for planning, scheduling and management of the power sector. Looking forward, with the continuous innovation of technology and the expansion of application areas, time series forecasting models will play an increasingly important role in China's power applications, helping to build a smarter, more efficient and sustainable power system.

## References

[1] Hu X. Stability prediction analysis of strong rainfall on slopes based on time series ARIMA model [J]. Pearl River Water Transport, 2024, (04).

[2] YAN Xun, TIE Chengcheng, YAN Wei et al. Analysis of global temperature prediction based on ARIMA model and CNN-LSTM combined model [J]. Science and Innovation, 2024, (02).

[3] ZHAO Ziyue, LIU Xuemei. Crop sowing area prediction based on ARIMA-GM combination model--Taking Jilin Province as an example [J]. Jilin Water Resources, 2024, (01).

[4] NI Xuan, CHEN Yichen. ARIMA model combined with LSTM neural network to establish a prediction model for drug procurement decision-making [J]. Heilongjiang Science, 2024, 15 (02).

[5] Jia Li. Power analysis and prediction based on neural network [D]. Southwest University of Science and Technology, 2021.

[6] Liu Yibo. Power prediction analysis based on LSTM neural network [J]. Integrated Circuit Application, 2023, 40 (08).

[7] SHEN Xingru. Electricity prediction based on big data mining technology [D]. Southeast University, 2022.

[8] Yan Lidan. Application of particle filtering method based on ARMA model in electricity consumption analysis [D]. Guilin University of Electronic Science and Technology, 2022.

[9] YANG Yang, TIAN Dingsheng, ZHANG Bao'an, et al. Research on urban economy and population forecasting based on ARIMA model [J]. Comprehensive Transport, 2023, 45 (11).

[10] Zhang Xinyu. Forecasting analysis of disk occupancy based on ARIMA and exponential smoothing algorithm [J]. Modern Information Technology, 2023, 7 (24): 67-71.