Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 2, 2025

Time Series Modeling and Forecasting Analysis of Energy Consumption in a Rare Earth Workshop

Junfeng Li [1, 2], Qian Chen[1, *], Yang Du [2], Donghui Lijiang [2], Zhen Liang [1]

[1] Equipment Engineering Department, Advanced Technology & Materials Co., Ltd., Beijing, China
[2] Equipment Engineering Department, AT&M North Technology Co., Ltd., Baotou, Inner Mongolia, China

[*] **Corresponding author**: Qian Chen (Email: sanchen97@163.com)

Abstract: Under the framework of the dual-carbon strategy, an in-depth study of carbon emission efficiency in energy consumption is of strategic importance for China to achieve its goals of carbon peaking and carbon neutrality, as well as to guide the economic and social transition towards a green, low-carbon, and sustainable development model. This study aims to provide empirical support in this field through detailed data analysis and predictive modeling. The paper employs the ARIMA time series prediction model, focusing on analyzing the power consumption data of a sintered NdFeB workshop from 2023 to 2024 to make accurate predictions of future energy use and carbon emission trends. The research process is rigorous: first, unit root tests (ADF tests) and white noise tests are conducted on the collected dataset to ensure data stationarity and reliability. Next, the Bayesian Information Criterion (BIC) is used to select optimal model parameters, enhancing the accuracy and efficiency of predictions. In terms of data processing, the dataset is divided into a training set and a test set in an 8:2 ratio, ensuring the model learns data characteristics thoroughly during training and demonstrates strong generalization in testing. The experimental results indicate that applying the ARIMA model to forecast power consumption data in the workshop effectively reveals intrinsic energy consumption patterns and provides valuable guidance for optimizing workshop energy management and reducing carbon emissions.

Keywords: Dual-Carbon Strategy; ARIMA Model; Time Series Forecasting; Sintered NdFeB Workshop.

1. Introduction

In recent years, climate change and environmental degradation have posed critical challenges worldwide, urging countries to adopt low-carbon and sustainable development strategies. In response to this, China has implemented the “dual-carbon” strategy, which aims for carbon peaking by 2030 and carbon neutrality by 2060[1,2]. Achieving these ambitious goals requires a thorough examination of energy consumption patterns and carbon emission efficiency across different sectors, particularly in industries with high energy consumption. Research into carbon emission efficiency within the energy consumption process is not only fundamental to supporting the dual-carbon strategy but is also key to transforming the economy towards a green and sustainable model. This study, therefore, focuses on contributing to the empirical knowledge necessary to achieve these objectives by analyzing and predicting energy consumption trends and associated carbon emissions.

The production of sintered neodymium-iron-boron (NdFeB) magnets, widely used in high-tech applications such as renewable energy systems and electric vehicles, involves considerable energy consumption[3,4]. The NdFeB manufacturing process, particularly the sintering workshop, consumes a significant amount of electricity, making it a prime candidate for optimizing energy efficiency and reducing carbon emissions. Hence, understanding the energy consumption patterns within this specific context could provide valuable insights and offer practical guidance for similar energy-intensive industries. This study specifically aims to assess and predict energy consumption trends in the sintered NdFeB workshop from 2023 to 2024, providing a basis for future carbon emission reduction strategies within the sector.

To effectively analyze and forecast the energy consumption and associated carbon emissions, this study employs the ARIMA (Autoregressive Integrated Moving Average) time series prediction model. ARIMA is a widely recognized model in time series analysis, renowned for its efficacy in making short- to medium-term forecasts by capturing patterns in temporal data. Given its robustness in modeling and predicting univariate time series data, ARIMA is well-suited for applications involving energy consumption predictions, which are often influenced by seasonal and temporal factors. The ARIMA model has been extensively applied in fields such as economics, meteorology, and environmental science, demonstrating its capability to deliver reliable forecasts when applied to complex and variable datasets[5,6].

The predictive analysis conducted in this research is underpinned by a rigorous methodology. The data used in the study include power consumption records from the sintered NdFeB workshop over a specified period. To ensure the validity and reliability of this dataset, an Augmented Dickey-Fuller (ADF) test and a white noise test were conducted. The ADF test is a statistical method used to verify the stationarity of time series data, which is crucial for ensuring the accuracy of predictive models. Similarly, the white noise test is employed to verify that the data contains sufficient signal and is not dominated by random noise, which could compromise the predictive capability of the model.

In addition, the Bayesian Information Criterion (BIC) is employed to select the most suitable parameters for the ARIMA model, enhancing the model’s accuracy and computational efficiency [7]. BIC is a criterion that balances model complexity with goodness of fit, ensuring that the selected model is not overly complex while still accurately capturing the essential patterns in the data. By using BIC for parameter selection, this study aims to optimize the ARIMA model to produce precise predictions of energy consumption and, by extension, insights into future carbon emission trends.

To evaluate the model's effectiveness and predictive accuracy, the dataset is divided into a training set and a test set, with an 8:2 ratio. This split allows the ARIMA model to learn the characteristics of the dataset through training, while the test set serves to validate the model's predictive capabilities. The division of data into training and testing segments is a common approach in predictive modeling, as it provides a reliable means of assessing model performance and ensuring that the model generalizes well to new data.

The findings of this study highlight the ARIMA model's effectiveness in forecasting energy consumption trends in high-energy-consuming workshops, such as the sintered NdFeB workshop. The model's predictions not only reveal underlying energy consumption patterns but also provide actionable insights for optimizing energy management. Such insights are particularly valuable for industries seeking to reduce their carbon footprint and transition to more sustainable practices. By improving energy efficiency and reducing carbon emissions, industries can contribute significantly to China's dual-carbon goals, thereby supporting a national commitment to environmental sustainability and global climate targets.

This research contributes to the growing body of literature on energy consumption forecasting and carbon emission reduction by applying a rigorous and data-driven approach to a specific industrial context. The findings demonstrate the practical utility of ARIMA in predicting energy consumption trends and underscore the importance of adopting data-driven strategies for achieving sustainability goals. By examining the case of the sintered NdFeB workshop, this study also provides a replicable model for other high-energy industries, offering a roadmap for carbon reduction initiatives aligned with China's dual-carbon strategy.

In summary, the primary objective of this study is to evaluate the potential of predictive models in contributing to energy optimization and carbon reduction within energy-intensive industries. Through a detailed analysis of energy consumption patterns and the application of ARIMA for forecasting, this research offers practical insights that support the dual-carbon strategy, providing a foundation for future research and policy efforts focused on sustainable development and environmental protection [8,9].

2. MODEL STRUCTURE

2.1. Overall Process

Figure 1 illustrates the standard modeling process of the ARIMA model in time series analysis. First, time series data is collected and tested for stationarity; if the data is non-stationary, differencing is applied to achieve stationarity. Next, the AIC and BIC criteria are used to select the optimal model parameters, and the ARIMA model is constructed. After model construction, a white noise test is performed on the residuals to ensure that the model has effectively captured the data patterns. If the residuals meet the white noise assumption, it indicates a good model fit, allowing the process to move into the final forecasting stage. This systematic approach provides scientific guidance for applying the ARIMA model, helping to improve forecasting accuracy and reliability.

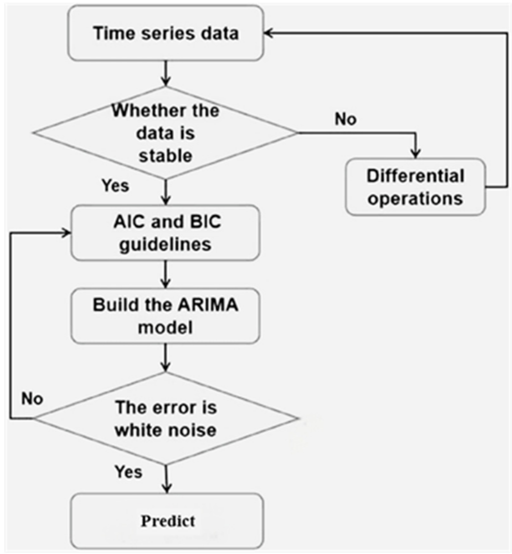

Figure 1. Overall Structure Diagram

The ARIMA model, or AutoRegressive Integrated Moving Average model, is used for time series forecasting and is particularly suited for linear time series data. It consists of three main components: Autoregressive (AR), Integration (I), and Moving Average (MA).

2.2. Autoregressive

The Autoregressive (AR) component is a key part of the ARIMA model, using past observations in the time series to predict the current observation. Below is a detailed description of the autoregressive component and its mathematical representation:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t \quad (1)$$

Y_t represents the observation at the current time point, $\phi_1, \phi_2, \ldots, \phi_p$ denotes the autoregressive coefficients, indicating the influence of each lagged value on the current value, and $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ represents the observations from p previous time points. ϵ_t is the white noise error term, typically assumed to follow a normal distribution with a mean of zero, i.e., $\epsilon_t \sim N(0, \sigma^2)$. The autoregressive order, p specifies the number of past observations used for prediction.

2.3. Integration

The basic idea of differencing is to calculate the difference between consecutive observations in the time series to remove trends or seasonality, making the data stationary. The order of differencing is represented by the parameter d , which specifies how many differencing operations are performed. Typically, the differencing process can be expressed as the following formula:

$$X_{t'} = X_t - X_{t-1} \quad (2)$$

Where X_t is the current observation of the original time series, X_{t-1} is the observation from the previous time point, and $X_{t'}$ is the observation after one differencing operation.

2.4. Moving Average

The Moving Average (MA) component corrects the current prediction based on the current and past error terms. The formula is expressed as:

$$Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q} \tag{3}$$

μ is the mean of the series; $\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-q}$ are the error terms at the current and the previous q time points; $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients, indicating the influence of different error terms on the current value; q is the order of the moving average.

3. Results and Analysis

Figure 2 illustrates how the autocorrelation function changes over time. The horizontal axis represents the time lag, while the vertical axis denotes the autocorrelation coefficient. Each data point corresponds to the autocorrelation value at a specific time lag.

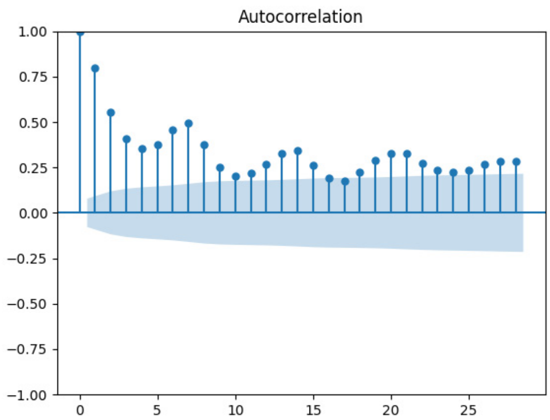

Figure 2. Autocorrelation Function Over Time

The horizontal axis represents the time lag, while the vertical axis shows the autocorrelation coefficient. Each data point corresponds to the autocorrelation value at a specific time lag.It is observed that at a time lag of 0, the autocorrelation coefficient reaches its maximum value, close to 1. This indicates a high degree of correlation in the data itself, meaning there is a strong linear relationship between the current value and its previous value. As the time lag increases, the autocorrelation coefficient rapidly declines, indicating that the correlation between the data diminishes over time.At later time lags, the autocorrelation coefficient shows some fluctuations but generally remains at a lower level. This may suggest the presence of some degree of seasonality or cyclical behavior in the data, although the correlation is not as pronounced as in the initial phase.The chart also includes confidence intervals for the autocorrelation coefficients, which help assess the reliability of the estimates. In certain time lags, the observed autocorrelation coefficients fall outside of the confidence intervals, suggesting the presence of outliers or special patterns that may require further investigation.

Figure 3 illustrates the Partial Autocorrelation Function (PACF), which is used to analyze the partial correlation at different lags in the time series. The PACF plot is a key tool for selecting the autoregressive order (AR order) when fitting an ARIMA model.

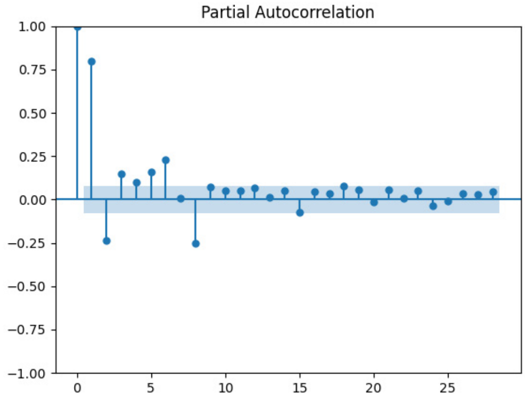

Figure 3. Partial Autocorrelation Function

In the PACF plot, the partial autocorrelation coefficient at lag 0 is always equal to 1, as it represents the correlation of the series with itself. At lags 1 and 2, the partial autocorrelation coefficients significantly exceed the confidence intervals, indicating a strong linear relationship between these lags and the current value. Starting from lag 3, the partial autocorrelation coefficients rapidly decay and gradually approach zero, suggesting that partial autocorrelation at larger lags becomes insignificant. Typically, the PACF plot is used to determine the order of the autoregressive (AR) part. Significant partial autocorrelations often decay sharply after a certain lag. This PACF plot shows significant partial autocorrelation at lags 1 and 2, which then gradually decays, indicating that the time series may be well-suited to a lower-order autoregressive model (such as AR(1) or AR(2)). This plot helps in selecting the appropriate autoregressive order in the ARIMA model, thereby improving the model's forecasting accuracy.

Figure 4 presents the variation of residuals from January 2023 to June 2024. The horizontal axis represents the time dates, and the vertical axis indicates the magnitude of the residuals. The two curves in the figure represent the residuals at each time point for the predicted values and the actual observed values.

The fluctuation amplitude of the curves is relatively small, and there is no significant time trend, indicating that the residuals do not significantly increase or decrease over time. This suggests that the model fits the data well.

In this study, we evaluate the normality assumption of the data by plotting a comparison of sample quantiles with theoretical normal distribution quantiles. The horizontal axis represents the quantiles of the theoretical normal distribution, while the vertical axis represents the quantiles of the sample data. Each point in the plot corresponds to the relationship between a sample value and its corresponding normal distribution quantile. From the plot, it is evident that most points are concentrated around the line $y=x$, suggesting that the sample data largely conform to the assumption of normal distribution. However, it is worth noting that some points deviate further from the line, particularly at extreme quantiles. This may indicate skewness or inconsistency in the data for certain extreme values.

The distribution of the data in Figure 6 clearly shows skewness, with most of the data concentrated between 0 and 25,000, while very few data points fall in the negative range or above 75,000. The highest peak appears in the 0-25,000 range, indicating that this range contains the most frequent data points, with 49 occurrences. This could be due to certain specific factors causing this concentration, such as economic activity, policy impacts, etc. The occurrence of data in the negative range (-75,000 to 0) is rare, with only 4 instances, and data in the tail of the positive range (above 75,000) also appears infrequently, with only 5 occurrences. This may suggest the presence of extreme values or outliers, which warrant further investigation into their underlying causes.

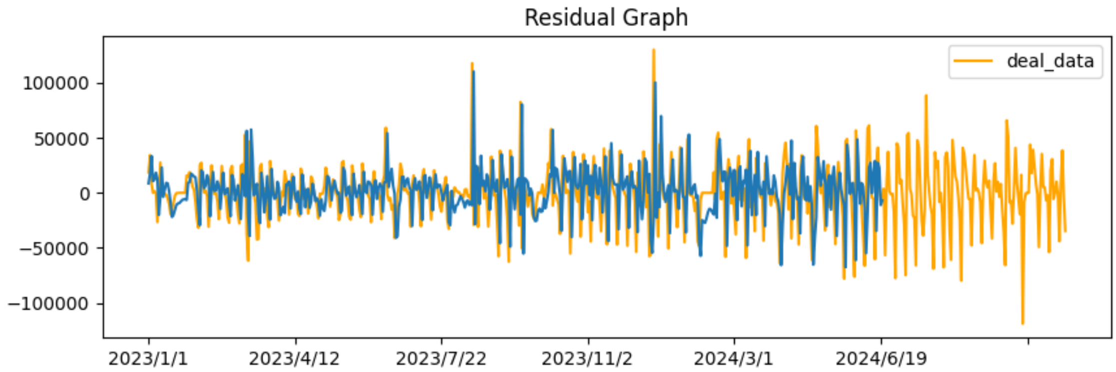

Figure 4. Residual plot

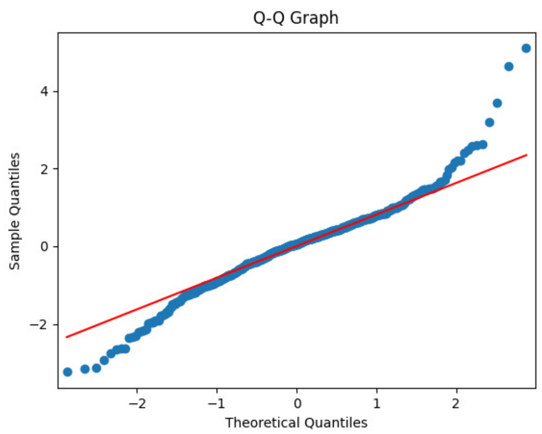

Figure 5. Q-Q Graph

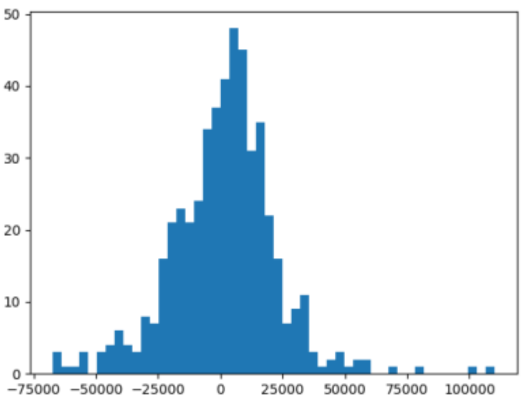

Figure 6. Histogram Graph

Figure 7 illustrates the trends of three different datasets: actual values (in blue), test values (in orange), and predicted values (in green). The horizontal axis represents time, spanning from January 2023 to October 2024, while the vertical axis represents power in kilowatts (kW), with a range from 0 to 140,000 kW.

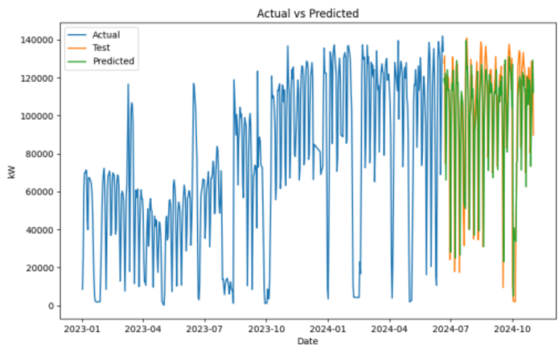

Figure 7. Prediction Results Graph

It is evident from the figure that the actual and test values remain relatively stable for most of the period, with occasional fluctuations. This indicates that the method has some effectiveness in optimizing workshop energy management and reducing carbon emissions.

Furthermore, the predicted values show a clear upward trend throughout the observation period, particularly between the second half of 2023 and the first half of 2024, where the growth rate significantly accelerates. This trend further emphasizes the effectiveness and guiding significance of the method in forecasting future energy demand.

4. Summary

In the context of advancing the dual-carbon strategy, studying the carbon emission efficiency in energy consumption processes is of significant importance. This study utilizes the ARIMA model to analyze and forecast the electricity consumption data of the sintering neodymium-iron-boron workshop from 2023 to 2024, aiming to provide empirical support for energy management and carbon emission control. During the research, the data was first tested for stationarity to ensure that it meets the model's requirements. Then, the Bayesian Information Criterion was used to select the optimal model parameters to improve forecasting accuracy. The results show that the ARIMA model effectively reveals the underlying patterns in energy consumption and provides practical guidance for optimizing workshop energy management and reducing carbon emissions.

This study also conducted in-depth analysis of the time series data characteristics using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). The results from the ACF and PACF plots indicate that the data exhibits strong autocorrelation at the initial lags, which gradually weakens at larger lags, suggesting that a lower-order autoregressive model (such as AR (1) or AR (2)) is appropriate. Additionally, residual analysis confirms that the model fits the data well, with no significant anomalies or trends identified.

In future research, more advanced forecasting methods, such as machine learning and deep learning models, can be explored to improve forecasting accuracy and handle more complex data features. Furthermore, considering the differences in energy consumption and carbon emission characteristics across various industrial workshops, this approach could be extended to other industries or application scenarios to support broader low-carbon transformations. With advancements in energy management technology and policy implementation, data-driven forecasting models will play an increasingly important role in achieving carbon peak and carbon neutrality goals.

References

[1] Zhang W, Shu C, Chen B, et al. Research on digital transformation strategy of the energy industry based on differential game under the dual-carbon background in China[J]. Science of the Total Environment, 2024, 906: 167297.

[2] Wang J, Jia L, Wang Y, et al. Diffusion of “dual carbon” policies among Chinese cities: A network evolution analysis[J]. Energy, 2024, 300: 131514.

[3] Kaya M. An overview of NdFeB magnets recycling technologies[J]. Current Opinion in Green and Sustainable Chemistry, 2024: 100884.

[4] Yang Q, Li Y, Li B, et al. Selective leaching and recovery of neodymium from NdFeB carbonyl residues[J]. Separation and Purification Technology, 2024, 329: 125137.

[5] Zhang G P. Time series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing, 2003, 50: 159-175.

[6] Ho S L, Xie M. The use of ARIMA models for reliability forecasting and analysis[J]. Computers & industrial engineering, 1998, 35(1-2): 213-216.

[7] Vrieze S I. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC)[J]. Psychological methods, 2012, 17(2): 228.

[8] Zhang T. Which policy is more effective, carbon reduction in all industries or in high energy-consuming Industries?——from dual perspectives of welfare effects and economic effects[J]. Journal of Cleaner Production, 2019, 216: 184-196.

[9] Guo X, Shi R, Ren D. Reduce carbon emissions efficiently: The influencing factors and decoupling relationships of carbon emission from high-energy consumption and high-emission industries in China[J]. Energy & Environment, 2024, 35(3): 1416-1433.