Forecasting Indonesia Renewable Energy Contribution to Electricity Generation Using ARIMA Method

Wahyu Ikbal Maulana1, Ronny Susetyoko2

*1Program Studi Sains Data Terapan, 2Program Studi Sains Data Terapan*

*rony@pens.ac.id, 2wahyuikbal777@student.pens.ac.id*

**ABSTRACT**

The growth of the human population and technology has led to a rapid increase in electrical energy consumption. Excess electrical energy would be detrimental to the provider, whereas providing less would be detrimental to the consumers. One method for reducing these losses is to forecast the amount of electrical energy that must be available to meet demand. Prediction results can help with three types of decisions, depending on the prediction period: operational decisions (short-term), tactical decisions (medium-term), and strategic decisions (long-term). Short-term forecasts are less relevant given the urgency of the situation. This study aims to help electricity providers to make decisions by making medium and long-term predictions using the Auto-Regressive Integrated Moving Average (ARIMA) method. In the best order determination experiment, ARIMA (8,2,0) was found to be the best model with the smallest error. ARIMA (8,2,0) has an average percentage error of 5.3 percent based on the overall prediction results. There is no linearity between accuracy and prediction period in the prediction period experiment. According to the experimental results, the highest accuracy is obtained in the medium term (monthly) with a value of RMSE 753,983.98. As a result, based on the time period, ARIMA is the best for tactical decisions (medium-term) regarding electrical energy consumption.

***Keywords:*** *Forecasting, Renewable Energy, ARIMA, Electricity.*

# INTRODUCTION

In an effort to address the impacts of climate change and achieve greenhouse gas emission reduction targets, many countries have shifted to renewable energy sources as an alternative to fossil fuel-based power plants. Indonesia, as a country with rapid economic growth and population, faces challenges in meeting its increasing energy needs while reducing its dependence on fossil fuels.

Renewable energy sources, such as solar, wind, water, geothermal, and biomass, offer more environmentally friendly and sustainable solutions to meet Indonesia's energy needs. The Indonesian government has set targets to achieve a 23% renewable energy mix by 2025 and 31% by 2050 in the national energy mix. However, the realization of renewable energy utilization in electricity generation is still lagging behind the set targets.

To achieve these targets, proper planning and decision-making are required regarding the development and implementation of renewable energy in the national electricity system. One crucial aspect of this process is the ability to forecast the contribution of renewable energy to electricity generation in the future. By understanding the patterns and trends of renewable energy contribution, the government and other stakeholders can take appropriate actions to promote investment, develop infrastructure, and implement policies that support a sustainable energy transition.

In this research, we propose the use of the Autoregressive Integrated Moving Average (ARIMA) method to forecast the contribution of renewable energy to electricity generation in Indonesia. ARIMA is a popular and proven effective time series forecasting method in various fields, including energy forecasting. By analyzing historical data on renewable energy contribution and the factors influencing it, we will build an ARIMA model that can be used to forecast the contribution of renewable energy in the future.

This research is expected to provide valuable insights for policymakers, investors, and other stakeholders in the renewable energy sector in Indonesia. With accurate forecasting, they can make more informed decisions in planning and implementing sustainable and environmentally friendly energy transition strategies.

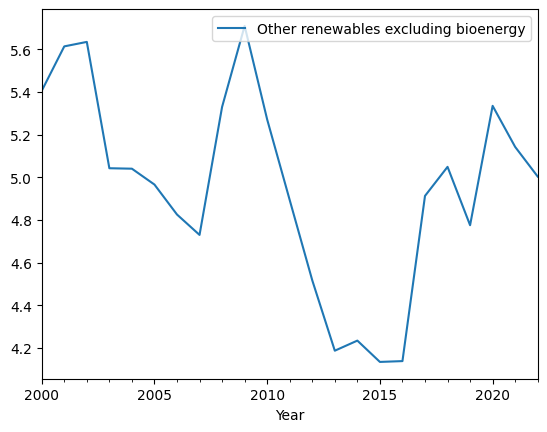
# DATA AND METHODOLOGY

Primary energy is measured using the "substitution method" (also called "input-equivalent" primary energy). This method is used for non-fossil sources of electricity (namely renewables and nuclear), and measures the amount of fossil fuels that would be required by thermal power stations to generate the same amount of non-fossil electricity. For example, if a country's nuclear power generated 100 TWh of electricity, and assuming that the efficiency of a standard thermal power plant is 38%, the input-equivalent primary energy for this country would be 100 TWh / 0.38 = 263 TWh = 0.95 EJ. This input-equivalent primary energy takes account of the inefficiencies in energy production from fossil fuels and provides a better approximation of each source's share of energy consumption (Our World in Data).

The Auto-Regressive Integrated Moving Average (ARIMA) model is one of the most popular approaches in the analysis and prediction of time series data. ARIMA, also known as the Box-Jenkins model, is a combination of Auto-Regressive (AR) and Moving Average (MA) models, which are then applied to data that has been differenced n times. George Box and Gwilym Jenkins are the key figures behind the development of this model. The ARIMA model has originated from the autoregressive model (AR), the moving average model (MA), and the combination of the AR and MA, the ARMA models. The ARIMA model can be used when the time series is stationary and there is no missing data within the time series. In ARIMA analysis, an identified underlying process is generated based on observations to a time series for generating a good model that shows the process-generating mechanism precisely (Box and Jenkins, 1976). The ARIMA technique includes identification (Abdel-Aal and Al Garni, 1997; Chavez et al., 1999; Zhang, 2001), estimation (Abdel-Aal and Al Garni, 1997), and diagnostic checking (Abdel-Aal and Al Garni, 1997; Zhang, 2001; Brockwell and Davis, 2002). A good summary of the ARIMA method can be found in Ediger et al. (in press).

Basically, ARIMA(p,d,q) is a model where p is the AR order value, d is the number of data differentiation processes until it reaches stationary conditions, and q is the MA order. This ARIMA model utilizes the dependency between the value of a variable at a certain time and the value of the variable itself in the past (lag) in the AR model, and uses the moving average method in performing error correction in the MA model.

The AR model in ARIMA describes the relationship between current values and previous values, as seen in equation (1) below:



Where is the variable value at time is the AR coefficient, is the error value at time , and is the AR order.

Meanwhile, the MA model in ARIMA considers the current error value as well as a certain weighted error value in the past, as shown in equation (2) below:

Which is is the average coefficient bergerak, the error value at time , dan is the MA order.

By combining AR and MA models, ARIMA provides a general model that can be explained through equation (3):

This ARIMA model allows the current value to be influenced by several previous values, including the current error value and several previous error values. The ARIMA method through several stages, such as stationarity tests, determining AR and MA values, to selecting the best model, becomes a very useful tool in analyzing and predicting time series data.

In this research, the focus lies on predicting Indonesia's renewable energy contribution, leveraging the renewable energy contribution data spanning from 1965 to 2022. The predictive process will harness Python packages such as statsmodels.tsa.stattools and statsmodels.tsa.arima.model. The ARIMA forecasting method entails the following steps: (1) Identification, (2) Estimation (and selection), (3) Diagnostic checking, and (4) Model implementation. Additionally, the formula for computing the RMSE value will be integrated (9).

|  |  |
| --- | --- |
| ∑𝑁 (𝑦 − )2  𝑅𝑀𝑆𝐸 = √ 𝑡=1 𝑡  𝑁 | (9) |

With description:

* + - * 𝑦𝑡 : actual time series value at time
      *  : the predicted value at time
      * N : amount of data used for prediction evaluation

The RMSE value indicates the model’s accuracy in predicting the actual value. The lower the RMSE value (close to zero)is, the more accurate the prediction model’s predictions.

# RESULT AND DISCUSSION

Exploration of the electric energy contribution data is done through time series plots from 1965 to 2022. The data pattern is highly fluctuating as shown in Figure 1. From the figure and based on the ADF-test, it can be seen that the data is stationary.

**Figure 1.** Renewable Energy Contribution to Electricity from 1965 until 2022

Order AR(p) is the sum of past values (lag) that significantly influence future values. The AR order value (p) ranges from 0 to the initial estimated value shown in the Partial Auto-correlation Function (PACF) graph. The graph makes it easier to see the relationship between future and past values that are outside of the significant limit determined by equation (6). Assumed that eight lags cross the significant limit, resulting in an initial estimated value of the AR order (p) of eight. Experiments were conducted to test all possible AR orders (p) ranging from zero to the maximum value (the initial estimated value obtained from the PACF graph) in order to obtain an AR order value (p).

Order MA(q) is the number of error values in past values (lag) that significantly impact future values. The error value is calculated by subtracting the current value from the moving average value of q. The obtained error value is expected to be an error correction of the AR prediction results to increase the accuracy of the ARIMA method prediction results. MA (q) has an order value ranging from zero to the initial estimated value shown in the Auto-correlation Function (ACF) graph. The graph makes it easier to see the relationship between future and past values that are outside of the significant limit determined by equation (5). Assumed eight lags cross the significant limit, resulting in an initial estimated value of the MA (q) order of eight. Experiments were carried out to try all possible AR orders (p) ranging from zero to the maximum value (the initial estimated value obtained from the PACF graph) in order to obtain an AR order value (p) that gave prediction results with the highest accuracy value.

The prediction period influences the accuracy of the prediction model. In this study, the prediction period (time span) is set to five years, which affects the extent of model correction. With only 57 data points in total, the prediction model's correction process will be limited accordingly. In this study, the prediction period spans five years, implying that the model will be corrected once per year, utilizing the available 57 data points. Consequently, the knowledge correction process will be repeated 57 times in total. This process will be iterated over the course of the semester and the year, as described earlier. The aim of this experiment remains to investigate the effect of the frequency of knowledge correction processes (prediction period) on the accuracy of the prediction model. The accuracy of the prediction results will still be assessed using the RMSE values calculated with equation (9).

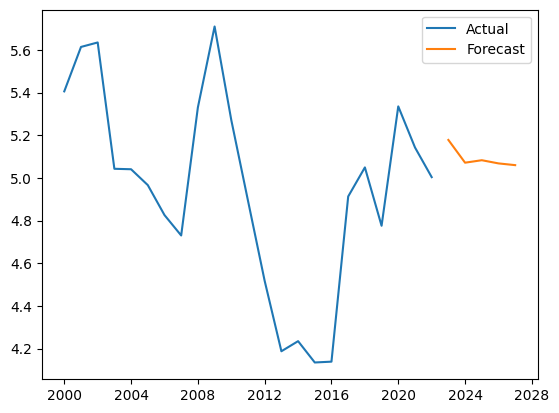
**Table 1.** The three best model orders based on RMSE

|  |  |  |
| --- | --- | --- |
| No | Model | RMSE |
| 1 | ARIMA (0,0,1) | 0.2117 |
| 2 | ARIMA (3,1,1) | 0.2022 |
| 3 | ARIMA (3,2,2) | 0.2091 |

The three order models in Table 2 are the results of an experiment to determine the best order model for the experiment to use for the longest period of time. The results show that the AR(2) or ARIMA (3,1,1) model has the smallest RMSE, thus the appropriate model for KWh per capita electricity consumption data is ARIMA (3,1,1).

The RMSE value represents the evaluation results of the prediction of renewable energy contribution in the best order determination experiment. Predictions are made based on the prediction period for 5 years. The ARIMA model is used to predict period based on the results of determining the best order model.

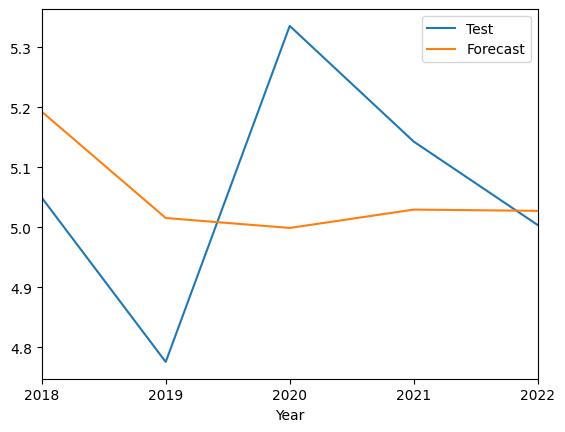
Experiments predicting the next month yielded the lowest RMSE in the ARIMA model (3,1.1). The RMSE for the ARIMA model (3,1,1) is 0.2022. The graph plot of the prediction results from the ARIMA model that has been made for the next 5 years can be seen in Figure 2.



**Figure 2** Graph of predictions for the next 5 years

The light blue color represents training data, the blue color represents test data, and the yellow color represents predicted data. The Y-Axis represents the total amount of renewable energy contribution, while the X-Axis represents the time of year.

According to the graph plot in Figure 3, the graph plot of the predicted results can follow the pattern of the graph plot of the test data. In comparison, the ARIMA model (3,1,1) is used to forecast the next five years, yielding the highest RMSE value of 0.2022. Figure 3 depicts a graph plot of the compare data train and data test.



**Figure 3** Compare data train and data test

The graph plot in Figure 3 shows that there is a significant difference between the prediction results and the actual results. When compared to other prediction periods, this produces the highest RMSE value.

Table 2 shows the experimental results for the maximum period. Table 4 uses the term exp code to refer to the experiment code. In the ARIMA model, it can be seen that a period of one month produces the smallest RMSE value, with an RMSE value of 0.2022 (3,1,1). Table 5 shows a comparison of the prediction results for the next 5 years with the actual data

|  |  |  |
| --- | --- | --- |
| Actual Data(MWh) | Predicted Data(MWh) | Error (%) |
| 5.049508 | 5.192724 | 2.83 |
| 4.775927 | 5.015602 | 5.01 |
| 5.335481 | 4.999035 | 6.3 |
| 5.142876 | 5.029618 | 2.2 |
| 5.003898 | 5.027521 | 0.47 |

**Table 2.** Comparison of actual data and predicted data

From Table 2, it can be seen the comparison of the actual data and the predicted results, with an error of up to millions of MWh and an error of up to 10%.

# CONCLUSION

Based on the results of data analysis and discussion, this research concludes that the ARIMA(3,1,2) method was suitable for forecasting the data of renewable energy (total primary energy supply) in Indonesia. This conclusion is supported by the RMSE value of ARIMA(3,1,2), which is 0.2584, indicating a good fit for the data. The lower RMSE value demonstrates that the ARIMA(3,1,2) method can effectively capture the patterns and trends in the historical data, providing reliable forecasts that can aid in strategic planning and decision-making for the sustainable energy transition in Indonesia.

# REFERENCES

1. Energy Institute - Statistical Review of World Energy (2023) – with major processing by Our World in Data
2. Ozturk, S., & Ozturk, F. 2018. Forecasting Energy Consumption of Turkey by Arima Model. Journal of Asian Scientific Research. Asian Economic and Social Society.
3. Barak, S., & Sadegh, S. S. (2016). Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm. In International Journal of Electrical Power &amp; Energy Systems (Vol. 82, pp. 92–104). Elsevier BV.
4. R. Mulla, “Hourly Energy Consumption,” 30 August 2018. [Online]. Available: https://[www.kaggle.com/robikscube/hourly-](http://www.kaggle.com/robikscube/hourly-) energy-consumption. [Accessed 30 January 2019].
5. M. I. Fauzan, "Analisis Pemrosesan Paralel dalam Mendukung Layanan Prediksi Cuaca," 2018.
6. Z. Reitermanov´a, “Data Splitting,” WDS’10 Proceedings of Contributed Papers, vol. Part I, p. 31–36, 2010.
7. Wei, W. W. S. (2006). Time Series Analysis: Univariate and Multivariate Methods. In Department of Statistics The Fox School of Business and Management Temple University (Issue )