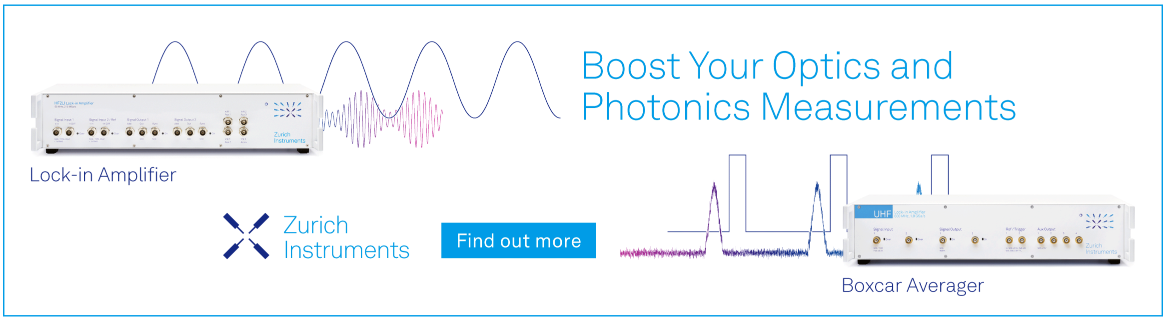
AIP Conference Proceedings

RESEARCH ARTICLE | MARCH 07 2024

Forecasting of hydropower production using Box-Jenkins model at Tasik Kenyir, Terengganu

Nornabila Abu ✉; Nurul Nadrah Aqilah Tukimat; Noratikah Abu

AIP Conf. Proc. 2895, 050005 (2024)
https://doi.org/10.1063/5.0193157



Forecasting of Hydropower Production Using Box-Jenkins Model at Tasik Kenyir, Terengganu

Nornabila Abu[1, a)] , Nurul Nadrah Aqilah Tukimat[2,b)] and Noratikah Abu[3,c)]

[1,2] *Faculty of Civil Engineering Technology, Universiti Malaysia Pahang, 26300 Gambang, Kuantan, Pahang, Malaysia.*
[3] *Centre for Mathematical Sciences, Universiti Malaysia Pahang, 26300 Gambang, Kuantan, Pahang, Malaysia.*

[a)]Corresponding author: nornabilaabu@gmail.com
[b)]nadrah@ump.edu.my
[c)]atikahabu@ump.edu.my

Abstract. Hydropower is one of the most essential mainstays in the long list of renewable energy resources that implements the use of potential energy of water to generate power by transforming the energy in the form of electricity. Forecasting the future energy production benefits in maintaining the effectiveness of the hydropower plant in the long term. This study aims to forecast the hydropower energy produced as electricity in Sultan Mahmud Hydropower Plant, Lake Kenyir, Terengganu using Box-Jenkins model in the short term from October 2020 until December 2022. Analysis and forecast is based on the historical data from a total of four unit of electricity generator from January 1997 to September 2020. Evaluation is made on the forecasted result using Mean Absolute Percentage Error (MAPE) to validate the accuracy of the model. The results demonstrated that by using the proposed model and numerical calculation, Box-Jenkins model is effective in forecasting the monthly electricity energy produced by the hydropower plant. The best model obtained with the smallest MAPE value of 26.4% is ARIMA (2,0,0).

INRODUCTION

Countries in the world are positively changing towards consumption of renewable and environmental friendly energy generations. Reports made by the International Energy Agency (IEA) in their Global Energy Review 2021 states that the global energy use in year 2020 has grown by 3% and is set to increase in 2021. Electricity generated from renewable resources had a growth of almost 7% in 2020 and is projected to increase by 30% of total electricity generation by the year 2021. Most common renewable energy sources include the hydropower, wind and solar energy. Hydropower energy is the third highest contributor next to solar and wind energy where both contributed to two-third of the renewable energy growth since 2019. In Malaysia, use of renewable energy has seen an increase of 4% as mention in the National Energy Balance 2018. As of 2018, hydropower has contributed 18.1% from total installed capacity in Malaysia which was 33,991 MWh.

Hydropower system implements the use of potential energy of water to generate power without making changes on its composition, hence avoiding direct pollution on the environment. There are several types of hydropower plants such as run-of-river and conventional dams. Conventional dams are built in large facilities including a water reservoir that can store water. Water from the reservoir is released to flow through the turbines which will then generate electricity to be supplied to areas in demand. Reservoirs are not solely designed for generation of electricity purposes but can also be used as drought or flood control, irrigation and recreational activity spot (1).

3rd International Conference on Applied & Industrial Mathematics and Statistics 2022 (ICoAIMS2022)
AIP Conf. Proc. 2895, 050005-1–050005-8; https://doi.org/10.1063/5.0193157
Published by AIP Publishing. 978-0-7354-4820-9/$30.00

In the present day, due to increasing demand for renewable and sustainable energy resources it is necessary for the management to improve the workability and lifespan of the hydropower plant (2). Forecasting the future energy production of hydropower plants is important to optimize the use of equipment and also able to fulfill the increasing demand of energy (3). To have better prediction of the future hydroelectricity production of the hydropower plant, studies are made using various methods such as time series model and Artificial Neural Networks (ANNs). However, by using the ANNs, projecting hydropower energy produced specifically will require dependency on the weather and climate change impact of the local area. A study by (4) forecast the Electricity Generation for Hydro Power Plants using different supervised machine learning and time-series based models. The data consist of daily electricity generation for the last five years augmented with the temperature and rainfall data of the dam catchment area. Results found that the proposed method can forecast the production of electricity generation with Mean Absolute Error of 2.47 which is very small. Li, Sun (5) applied support vector machine (SVM) to forecast short term Power Generation Energy for Small Hydropower Stations (SHP). Genetic algorithm was used to identify the appropriate parameters and the named the model as GA-SVM. The proposed model then was compared with ARMA model to the data collected which is from two counties, Yunlong County and Maguan County both 915 days long with the period between May 1, 2011, and October 31, 2013. As the results, they found that GA-SVM model is a very potential candidate for the prediction of short-term power generation energy of SHP.

A study by (6) applied the ARIMA Box-Jenkins to model the monthly streamflow of the Euphrates River in Iraq from January 2000 to December 2019. the model output ARIMA (1,1,0), (0,1,1) with the root mean squared error (RMSE) of 48.7. A comparison is made by using the nonlinear autoregressive (N.A.R.) model. the model's output however had the RMSE of 93.4 which is higher than of the ARIMA model. the future projection of the monthly stream flow was made using the ARIMA Box-Jenkins model until year 2025. Based on the reviews, there is limited research that uses Box-Jenkins model as one of the common method used in forecasting hydroelectricity. Studies on future hydropower generation are mostly depending on the climate condition such as rainfall and temperature. However, in cases where there are insufficient or lacking of climate data, researchers had to rely on the historical data provided by the hydropower plant which is the amount of electricity produced by each generator in unit of Gigawatt hours (GWh). Therefore, this paper implemented the time series analysis using Box-Jenkins model to project the electricity production from one of the largest hydropower plant in Malaysia, which is the Sultan Mahmud Hydropower Plant. This model will project short term electricity production of the hydropower plant. Evaluation of the model is by considering the value of the MAPE analysis of the models. This paper is organized in four parts including the introduction, methodology, results and discussion and a conclusion is provided at the end of the paper.

METHODOLOGY

1.0 Data Collection

This study used historical electricity produced from a total of four units of electricity generator on the site of Tasik Kenyir, Terengganu. Energy Produced (EP) in unit of GWh is collected monthly from January of 1997 to September 2020. The EP is used in this study to validate the model used during analysis and is projected in short term from October 2020 until September 2023.

2.0 Box-Jenkins Model

Box-Jenkins modelling was developed by Box and Jenkins (7) is one of the most powerful forecasting methods available in research practice of the time series analysis. The basic ideas in the Box-Jenkins model building the development of four-stage iterative procedure of time series i.e. identification, estimation, diagnostic checking and forecasting. Box-Jenkins modelling involves five types of models (8). There are three model for the stationary model: autoregressive (AR), moving average (MA) and (autoregressive moving average (ARMA). And two models for nonstationary model: autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA). This study will not use SARIMA model since the data is not seasonal. The equation of the four models; AR, MA, ARMA, ARIMA are given in the Eq. (1-4), respectively.

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{1}$$

$$y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q} \tag{2}$$

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q} \tag{3}$$

$$y'_t = \phi_0 + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \cdots + \phi_p y'_{t-p} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q} \tag{4}$$

The Box-Jenkins model consists of four-stage iterative procedures;

Stage 1: Model Identification. The purpose of model identification is to identify whether the time series is stationary or not stationary by looking at the PACF and ACF. Stationarity of the series need to be done in order to model the data.
Stage 2: Parameter Estimation. The parameter for the model must be estimated after the model has been selected. The estimation method used in this study is maximum likelihood estimation (MLE).
Stage 3: Diagnostic Checking. In diagnostic checking, the selected model will be statistically verified against the original data to see whether it adequately describes the series. Ljung Box Q chi-square (LBQ) will be used in this study for the diagnostic checking.
Stage 4: Forecasting. Forecasting stage is continued once the satisfactory model is identified.

Mean absolute percentage error (MAPE) was used as the forecast evaluation. The formula for MAPE was given in Equation 5. The analysis for this study was run and analyzed using Python.

$$MAPE = \frac{\sum |(y_t - \hat{y}_t)/y_t|}{n} \times 100 \tag{5}$$

RESULTS AND DISCUSSION

The historical electricity produced from a total of four units of electricity generator on the site of Tasik Kenyir, Terengganu is shown in Fig.1. Based on Fig.1, it can be seen that there is no seasonal pattern exist.

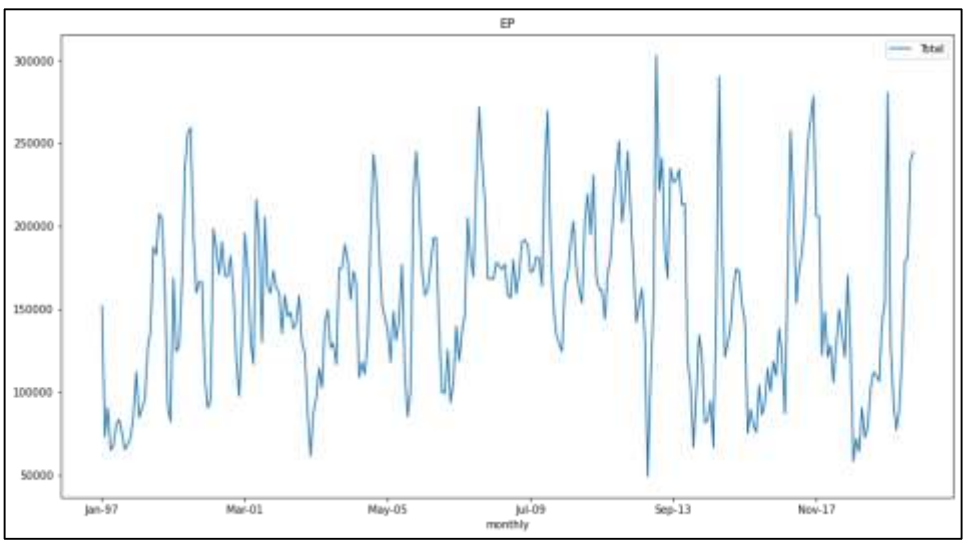


FIGURE 1. Total historical electricity produced by months in unit of GWh.

Stage 1: Model identification

In time series forecasting, the original data will be divided into training and testing data. The purpose of this process is to provide reliable indication of how well the model to both type of data (9). Training data is used to estimate model, whereas the testing data is used in validate model. The size of the training data is around 80% while for testing is typically around 20% of the total sample (9). In this study, the electricity produced are divided into two parts;

training data from January 1997 to April 2020 with 256 observations and testing data from May to September 2020 with 29 observations. Figure 2 shows the electricity produced for training data.

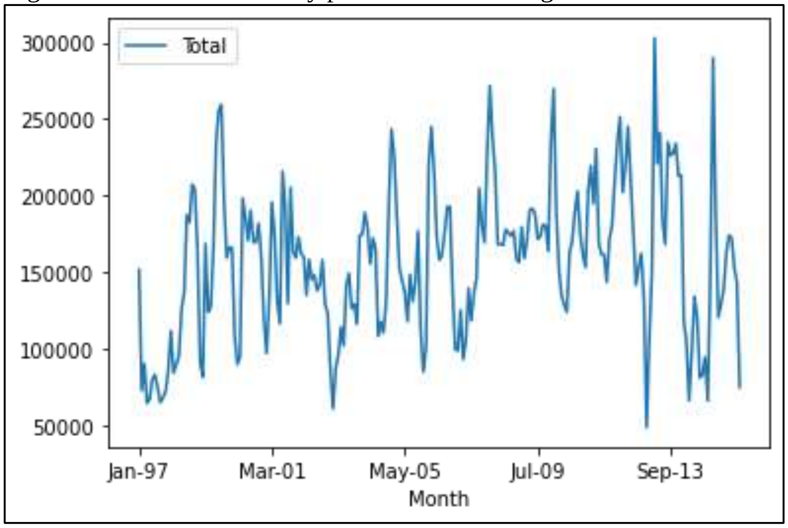
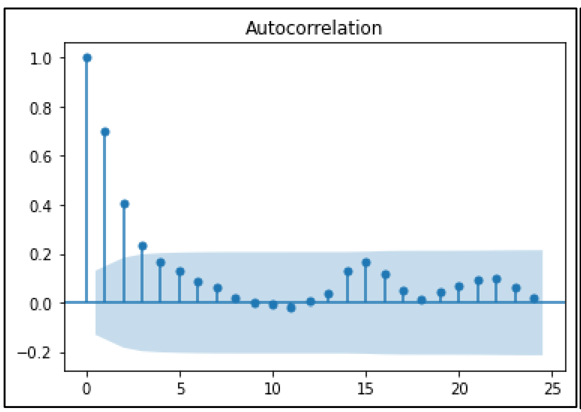
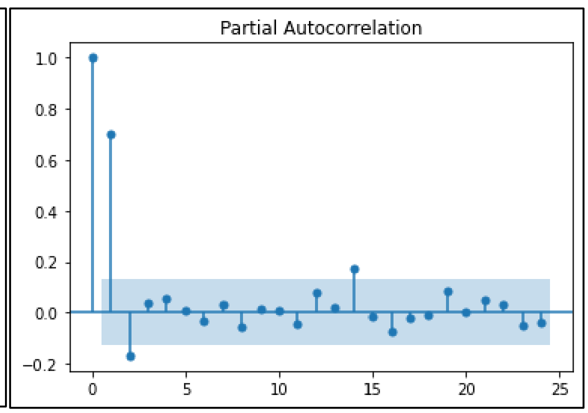


FIGURE 2. Electricity produced (EP) data used for model training.

The series shows that the series is stationary since the upward and downward trend fluctuate at the same level. Then, to prove the series is stationary in mean, Augmented Dickey Fuller (ADF) test wan run. The null hypothesis for an ADF-test is that the data series is nonstationary and the results give the $p-value = 0.0000$. Hence, the data is already stationary in mean. The ACF and PACF plot in Fig.3 also shown that the series is stationary.



(a)



(b)

FIGURE 3. Autocorrelation (ACF) and Partial Autocorrelation (PACF) plot.

From the ACF and PACF plot, it can be seen that the lag for PACF decay after lag 2 gives us the $AR(2)$ model while the ACF plot gives us $MA(3)$ since the lag decay towards zero after lag 3. Hence, the tentative model are $ARIMA(1,0,0), ARIMA(1,0,1), ARIMA(1,0,2), ARIMA(1,0,3), ARIMA(2,0,0), ARIMA(2,0,1), ARIMA(2,0,2)$.

Stage 2: Parameter estimation

After the tentative model have been identified, the next step is to estimate the parameter in the model. The parameter estimation process was done using Python and results gives the best model is $ARIMA(2,0,0)$ with the lowest AIC value. Fig.4 shows the summary results for the parameter estimation of the best model $ARIMA(2,0,0)$. From Fig.4, we can see that the $p-value$ for all the parameter is less than 0.05, means that all the parameters in the model is significant. Hence, the equation for the best model is;

$$y(t) = 51460.28 + 0.8354y_{t-1} - 0.1628y_{t-2} \quad (6)$$

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                  Total   No. Observations:                  227
Model:                     ARMA(2, 0)   Log Likelihood               -2691.420
Method:                       css-mle   S.D. of innovations          34059.909
Date:                Thu, 31 Mar 2022   AIC                           5390.840
Time:                        13:18:54   BIC                           5404.540
Sample:                             0   HQIC                          5396.368

===============================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------
const        1.573e+05   6436.318     24.441      0.000    1.45e+05     1.7e+05
ar.L1.Total     0.8331      0.066     12.608      0.000       0.704       0.963
ar.L2.Total    -0.1823      0.066     -2.745      0.006      -0.313      -0.052
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            2.2845           -0.5149j            2.3418           -0.0353
AR.2            2.2845           +0.5149j            2.3418            0.0353
-----------------------------------------------------------------------------
```

FIGURE 4. Summary results for the parameter estimation.

Stage 3: Diagnostic checking

Next, before proceed to use the model for forecasting stage, the model need to go through diagnostic checking. LBQ test was performed in Python and the results give the $p - value = 0.8869$, greater than by default 5% significance level. Hence, the residuals are not serial correlated since the null hypothesis for the LBQ test is the series is not serial correlated. The model is adequate enough if the residuals is not serial correlated and follows normal distributions. Figure 5 shows that the residuals follow normal distributions.

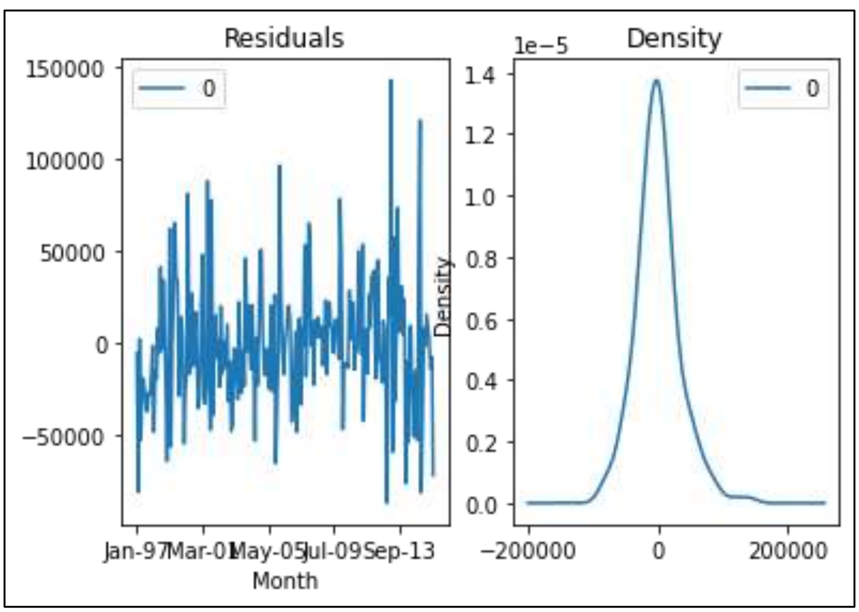


FIGURE 5. Residual Distribution of training data.

Stage 4: Forecasting

Forecasting stage is continued once the selected model is adequate enough. The model was used to forecast the training data. Figure 6 shows the comparison of original training data with the forecasted training data.

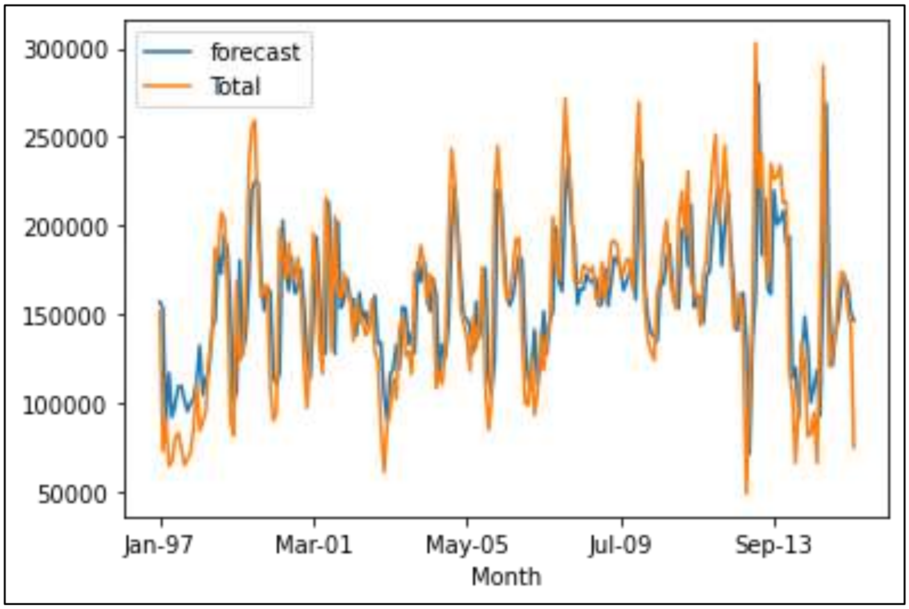


FIGURE 6. Comparison of historical training data with forecasted training data.

Based on Fig.6, it can be seen that the trend of forecasted data follows the original traditional data quite accurate. We can say that the model can be used to do forecasting future data. Mean absolute percentage (MAPE) was used as the forecast evaluation to determine the accuracy of the model. The MAPE for the training data is 18.07%. The model in Eq. (5) was developed from the training data. Then, we will used the model to forecast the testing data and make comparison between the original data for the accuracy. Fig.7 shows the comparison of the original testing data with the forecasted testing data.

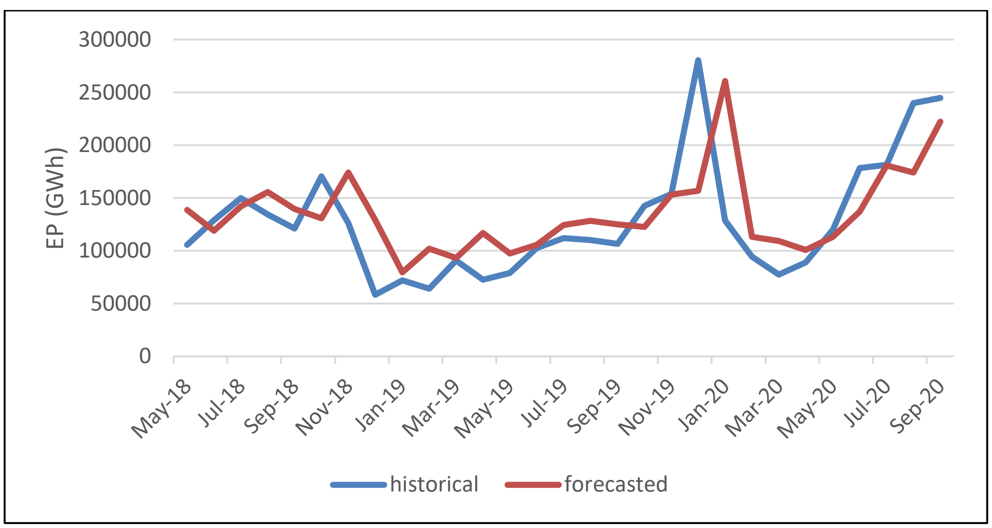


FIGURE 7. Comparison of historical testing data with forecasted testing data.

Based on Fig.7, the blue line is the original testing data and the red line is the fitted data from the model. It can be seen that the fitted line follows the pattern of the original data quite accurate. The MAPE give us 26% for the accuracy. Based on the MAPE criteria for model evaluation by Lewis (10), MAPE value between 20% - 50% is classified as reasonable forecasting. Next, we forecast the one step ahead for the electricity produced using $ARIMA(2,0,0)$. The electricity produced for October 2020 until December 2022 is shown in Table 1.

TABLE 1

Month	Forecasted Data
October 2020	216384
November 2020	192768
December 2020	177198
January 2021	168109
February 2021	163051
March 2021	160305
April 2021	158834
May 2021	158053
June 2021	157639
July 2021	157421
August 2021	157306
September 2021	157246
October 2021	157214
November 2021	157197
December 2021	157188
January 2022	157184
February 2022	157181
March 2022	157180
April 2022	157179
May 2022	157179
June 2022	157179
July 2022	157179
August 2022	157179
September 2022	157179
October 2022	157179
November 2022	157179
December 2022	157179

CONCLUSION

Forecasting electricity produced is important in the effort to maintain the lifespan of hydropower plant and be able to fulfill the increasing demand of energy over the time. This study proposes to forecast the hydropower energy produced as electricity in Sultan Mahmud Hydropower Plant, Tasik Kenyir, Terengganu using Box-Jenkins model. From the analysis, we found that the best model to forecast electricity produced is $ARIMA(2,0,0)$ with the smallest AIC values. The MAPE value gives 26.4% indicates that the model is reasonable for forecasting the electricity produce. Based on this study, it proves that Box-Jenkins model is able to forecast the electricity produced.

ACKNOWLEDGEMENT

This research is supported by Faculty of Civil Engineering Technology, Universiti Malaysia Pahang, Ministry of Higher Education (MHE) RDU1901141 (Ref: FRGS/1/2019/TK01/UMP/02/1), Meteorological Department (MMD) and Drainage and Irrigation Department (DID). The supports are gratefully acknowledged.

REFERENCES

1. Branche E. The multipurpose water uses of hydropower reservoir: The SHARE concept. Comptes Rendus Physique. 2017;18(7):469-78.
2. Chow MF, Subbramaniam, P., Sidek, L. M. Assessment of Sedimentation Problem in Kenyir Hydropower Reservoir. IOP Conference Series: Earth and Environmental Science. 2020;541:012008.
3. Khaniya B, Karunanayake C, Gunathilake MB, Rathnayake U. Projection of Future Hydropower Generation in Samanalawewa Power Plant, Sri Lanka. Mathematical Problems in Engineering. 2020;2020:8862067.
4. Javed U, Fraz M, Mahmood I, Shahzad M, Arif O. Forecasting of Electricity Generation for Hydro Power Plants2020. 32-6 p.
5. Li G, Sun Y, He Y, Li X, Tu Q. Short-Term Power Generation Energy Forecasting Model for Small Hydropower Stations Using GA-SVM. Mathematical Problems in Engineering. 2014;2014:1-9.
6. Al-Saati NH, Omran II, Salman AA, Al-Saati Z, Hashim KS. Statistical modeling of monthly streamflow using time series and artificial neural network models: Hindiya Barrage as a case study. Water Practice and Technology. 2021;16(2):681-91.
7. Box GEP, Jenkins GM. Some Recent Advances in Forecasting and Control. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1968;17:91-109.
8. George E. P. Box GMJ, Gregory C. Reinsel. Time Series Analysis. 4th ed2008.
9. Hyndman RJ, Athanasopoulos, G. . Forecasting: Principles and Practice. 2nd ed2018.
10. Lewis CD. Industrial and business forecasting methods : a practical guide to exponential smoothing and curve fitting / Colin D. Lewis. London: London : Butterworth Scientific, 1982.; 1982.