Journal of Science, Innovation and Creativity
Vol. 4 No. 1 (2025): ISSN (Online): 2958-454X
DOI: https://doi.org/10.58721/jsic.v4i1.958

Forecasting hydropower production in Tanzania using the SARIMA model

Msabaha Haruna, Bahati Ilembo & Joseph Lwaho
Mzumbe University, Tanzania

Article History
Received: 2024-11-12
Revised: 2025-03-04
Accepted: 2025-03-10
Published: 2025-03-12

Keywords
Box
Energy
Jenkins
SARIMA

How to cite:
Haruna, M., Ilembo, B., & Lwaho, J. (2025). Forecasting hydropower production in Tanzania using the SARIMA model. *Journal Science, Innovation and Creativity*, 4(1), 26-39.

Abstract
There is a growing demand for energy consumption showing that adequate energy supplies are essential for economic growth. Predicting energy supplies is crucial and thus accurate predictions help minimise the growing energy demand-production gap and enable power plant managers to easily and promptly detect any anomalies or failures in electricity production by analysing deviations from the predicted trends. The main objective of this study was to forecast hydropower production in Tanzania using secondary univariate time series monthly data for the past 22 years (2002-2023). A total of 264 data points were used for the prediction using the seasonal Autoregressive Integrated Moving Average model (SARIMA) following Box and Jenkins methodology's ability for handling seasonal data. The results show that there will be no substantial decline in hydropower production (KWh) until December 2025. The forecasts show that the hydropower generated, overall will not exceed 227,250,650 KWh but will be at the peak in May 2025 and start to decrease towards December 2025 with not more than a 40% decrease in every month. The forecast will not only help the power plant managers but also policymakers to devise mechanisms that will ensure the gap between energy demand and production is balanced for the welfare of the country's development.

Introduction

The growing demand for energy consumption indicates that adequate energy supplies are essential for economic growth and social development (Polprasert, Nguyên & Charoensook, 2021). Hydropower is a prominent source of energy (Condemi et al., 2021; Zhou et al., 2020), currently accounting for over 16% of global electricity production and 62% of all renewable electricity generation (Lee et al., 2022). Hydropower plants generate electricity by harnessing the power derived from falling or swiftly flowing water, which is naturally influenced by rainfall or, in certain regions, melting snowpack (Condemi et al., 2021). It can be argued that an adequate power supply is a crucial factor in the economic growth of every nation, with electricity generation being one of the key components in achieving sustainable economic development (Sarpong & Agyei, 2022; Waih, Buabeng & Agyarko, 2022; Zolfaghari & Golabi, 2021).

Tanzania possesses abundant energy resources, ranging from renewable to non-renewable sources. The country produces natural gas and coal primarily for domestic consumption, particularly in electricity generation and industrial applications (Kichonge, 2014). However, power cuts and rationing have occurred, which are associated with water level trends in power-generating plants. According to Wang, Li, and Pei (2017), predicting hydropower generation is crucial for addressing the energy shortage issues impacting the development of the national economy, enhancing the ecological environment, and fostering the harmonious and sustainable development of regional economies. Energy production forecasts are vital in today's world, where modern technology and the rapidly increasing demand for electricity dictate daily life (Javed et al., 2020; Barzola–Monteses et al., 2019). Such predictions assist electrical system operators and decision-makers in formulating better policies and mitigating risks (Barzola–Monteses et al., 2019). Accurate energy production predictions are essential for national energy planning and economic development (Bilgili, Keiyinci, & Ekinci, 2022). They help minimise the growing energy demand-production gap and enable power plant managers to promptly identify any anomalies or failures in electricity production by analysing deviations from predicted trends, as suggested by Javed et al. (2020).

Numerous studies have utilised the Autoregressive Integrated Moving Average (ARIMA) to predict hydropower production (Sarpong & Agyei, 2022; Polprasert, Nguyên & Charoensook, 2021; Mite-León & Barzola–Monteses, 2018; Sarkodie, 2017). Barzola–Monteses et al. (2019) combined the Box-Jenkins (ARIMA) and Box–Tiao (ARIMAX) regression methods to enhance forecasting accuracy. Zolfaghari and Golabi (2021) observed that the energy production series is nonlinear and non-stationary, making traditional methods less effective. A more robust approach involves employing a hybrid model that integrates the adaptive wavelet transform (AWT), long short-term memory (LSTM), and the random test algorithm (AWT-LSTM-RF) to enhance the accuracy of hydropower production forecasts. While acknowledging the arguments put forth by Zolfaghari and Golabi, this study harnessed the strengths of the seasonal ARIMA model as a suitable method for predicting monthly electricity generation in Tanzania, as suggested by Wiah, Buabeng, and Agyarko (2022), due to its strong predictive performance.

Materials and methods

Data types and sources

The study utilised secondary univariate time series data from the official database of Tanzania Electric Supply Company Limited, covering the past 22 years from January 2002 to December 2023, with monthly observations. The data measures unit hydropower generation in KWh produced each month. It includes the total hydropower generated from the various stations in Tanzania. The study employed a quantitative design to analyse and develop effective forecasts for hydropower production.

The conceptual framework

The conceptual framework of this paper was adopted from Tembo, Ilembo and Lwaho (2024) in their work which also used time series analysis, precisely forecasting using the SARIMA model. The analysis in this paper is guided by the Box-Jenkins methodology as described by statisticians George Box and Gwilym Jenkins back in 1970. Box-Jenkins's modelling involves identifying an appropriate model process, fitting it to data and then using the fitted model for forecasting. In practice, many time series contain a seasonal periodic component which repeats every "s" observation. For simplicity, we consider monthly observations where $s = 12$, but the application to other values s is straightforward. Box and Jenkins (1970) generalised the ARIMA model to accommodate seasonality and defined a general multiplicative seasonal model in the following form:

$$\vartheta_p(B)\varphi_p(B^{12})w_t = \theta_q(B)\xi_Q(B^{12})a_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 2.1$$

Whereby $B =$ backward shift operator

$\vartheta_p, \varphi_p, \theta_q, \xi_Q$ = polynomials of order p, P, q, Q respectively and

$\{a_t\}$ = independent random variables with zero mean and variance σ_a^2

The shift operator $(B^{12})w_t$ is such that $(B^{12})w_t = w_{t-12}$. Therefore, equation (2.1) defines a stationary model provided that the roots $\vartheta_p(B)\varphi_p(B^{12}) = 0$ lie outside the unit circle. To fit the model to a non-stationary series, Box and Jenkins (1970) again, suggest differencing the original series to remove both trend and seasonality following the procedure underneath:

$$w_t = \nabla^d \nabla_{12}^D x_t$$

Whereby $\nabla_{12} x_t = x_t - x_{t-1\,2}$ and that:

$$\nabla^f \nabla_{12} x_t = \nabla_{12} x_t - \nabla_{12} x_{t-1} = x_t - x_{t-1} - x_{t-12} + x_{t-13}.$$

The values of the integers d and D do not usually need to exceed a unit. Details that describe the Box and Jenkins procedure to forecasting can also be found in the old literature by Naylor et al. (1972), Chatfield and Prothero (1973) and Thompson and Tiao (1971). However, this paper provides the detailing procedure described by Box and Jenkins in their popular work of 1970.

The procedure entails fitting a mixed autoregressive integrated moving average (ARIMA) model to a given set of time series data and then taking conditional expectations. The main stages in setting up a Box-Jenkins forecasting model are as outlined below:

Model identification: Examine the data to reveal which member of the class of ARIMA processes appears to be the most appropriate.

Estimation: Estimate the parameters of the chosen model by least squares

Diagnostic checking: Examine the residuals from the fitted model to see if it is adequate

Alternative model consideration: Consider an alternative model if the first model appears to be inadequate for some reason, then other ARIMA models may be tried until a satisfactory model is found.

It follows that, for non-seasonal data, first-order differencing is usually sufficient. For seasonal data of 12 monthly observations, the operator $\nabla\nabla_{12}$ is often used if the seasonal effect is additive, while the operator ∇_{12}^2 may be used if the seasonal effect is multiplicative. For quarterly data, the operator ∇_4 may be used. For seasonal data, like in our case, the general seasonal ARIMA (SARIMA) model defined as $\vartheta_p(B)\varphi_p(B^{12})w_t = \theta_q(B)\xi_Q(B^{12})a_t$ has been used. The least squares estimate of the model parameters is obtained by minimising the residual sum of squares in a similar way to that proposed for ordinary ARIMA models.

Test for stationarity of time series

The Augmented Dickey-fuller (ADF) test has gained popularity in testing for stationarity of the time series data. The null hypothesis has been "there is a unit root or time series data is not stationary" which has to be rejected at a given level of significance. The Augmented Dickey-Fuller test is a unit-root-based on stationarity (Dickey & Fuller, 1979). The unit-root-based test is associated with the first lag of the time series variable. If the coefficient ($\gamma = 1$) has a unit root, the time series behaves similarly to the random walk model which is non-stationary and if the coefficient $|\gamma| < 1$ then, then there is no unit root. Hence, we can test statistically whether the coefficient (γ) is equal to one or not. The Dickey-Fuller test adopts this procedure by carefully manipulating the equation, given as:

$$y_t = \alpha + \beta t + \emptyset\, y_{t-1} + e_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.2$$

Also, written as

$$\Delta y_t = y_t - y_{t-1} = \alpha + \beta t + \gamma\, y_{t-1} + e_t \ldots\ldots\ldots\ldots\ldots\ldots 2.3$$

In the Dickey-Fuller test, we test the hypothesis.

$H_0: \emptyset = 1$

$H_1: \emptyset \neq 1$

Correlograms

ACF and PACF are statistical measures that help to analyse the relationship between a time series and its lagged values. They are generally producing plots that are very important in finding the value of p and q for Autoregressive (AR) and Moving Average (MA) models.

Autocorrelation Function (ACF)
ACF measures the linear relationship between a time series and its lagged values. It assesses how much the current value of a time series depends on its past values. Autocorrelation is fundamental in time series analysis, helping identify patterns and dependencies within the data. The correlation between the current observation (y_t) and the previous observation (y_{t-k}) is given as:

$$\rho_k = \text{corr}\,(y_t, y_{t-k}) = \frac{\text{Cov}\,(y_t, y_{t-k})}{\sqrt{\text{Var}\,(y_t) \cdot \text{Var}(y_{t-k})}} = \frac{\gamma_k}{\gamma_0} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.4$$

Where, k=1, 2, 3⋯

Partial Autocorrelation Function (PACF)
PACF removes the influence of intermediate lags, providing a clearer picture of the direct relationship between a variable and its past values. Unlike Autocorrelation, partial Autocorrelation focuses on the direct correlation at each lag. The partial Autocorrelation function at lag k for time series is given as:

$$\emptyset_{11} = \text{Corr}\,(Y_{t+1}, Y_t) = \rho_1$$

$$\emptyset_{kk} = \text{Corr}\,(Y_{t+k} - \hat{Y}_{t+k}, Y_t - \hat{Y}_t),\ k \geq 2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.5$$

The suitable values of p and q will be selected by observing the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series data. The appropriate ARIMA models will be selected by observing the behaviour of ACF and PACF spikes based on the order identified (Hyndman & Athanasopoulos, 2018).

Seasonality Test
To find out whether the time series data contain the seasonal pattern, seasonal decomposition was conducted using an additive model which decomposes time series into its trend, seasonal, cyclical and regular components by assuming time series can be modelled by using those components. Seasonality can be detected by visually examining the seasonal component for regular patterns that repeat at a fixed interval. If the seasonal component exhibits regular patterns, this indicates the presence of seasonality in the time series.

Model Estimation
The parameters of the selected seasonal ARIMA (SARIMA) model with the specific values of (p, d, q) $\times$ (P, D, Q)$_S$ were estimated. The maximum likelihood estimation (MLE) was used to estimate the coefficients of the suggested models during the identification stage. The best model was selected based on Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Akaike Information Criterion (AIC)

Kullback *et al*., (1951) developed a measure to quantify the information lost when approximating reality. Kullback & Leibler's divergence serves as a criterion for assessing model quality by minimising the loss of information. Two decades later, Akaike established a connection between the Kullback-Leibler measure and the maximum likelihood estimation (MLE) method, which is widely used in many statistical analyses for model selection (Akaike, 1974). This criterion referred to as Akaike Information Criterion (AIC), is generally considered the first model selection criterion that should be used in practice. The AIC is given as:

$$\text{AIC} = -2\log\text{L}\ (\hat{\theta}) + 2k \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . 2.6$$

Where; θ is the set of model parameters, L $(\hat{\theta})$ is the likelihood of the candidate model given the data when evaluated at the maximum likelihood estimate $\hat{\theta}$ and k is the number of estimated parameters in the candidate model.

Since AIC does not consider the effect of sample size, for small sample sizes, the second-order equation of the Akaike information criterion (AIC_C) is defined as:

$$\text{AICc} = -2\log\text{L}\ (\hat{\theta}) + 2\text{k} + \frac{(2\text{k}+1)}{(\text{n}-\text{k}-1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . 2.7$$

Where n denotes the total number of observations.

The small sample size is $^{n}/_{k}$ less than 40, also that when the number of observations increases, the third term in AIC_C approaches zero and will therefore give the same result as AIC in the equation. 2.6

Bayesian information criterion (BIC)

Bayesian information criterion is another model selection criterion based on information theory but set within a Bayesian context. The difference between the BIC and AIC is the greater penalty imposed for the number of parameters

$$\text{BIC} = -2\log L(\hat{\theta}) + k\log n \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.8$$

where n denotes the total number of observations.

The BIC strongly penalises the number of involved parameters. High values of AIC mean that the observed data does fit the model, while lower values indicate strong evidence that the observed data fit the models. Similarly, lower values of BIC indicate better fitting of the models.

Diagnostic Checking

Diagnostic check evaluating the acceptance of the fitted SARIMA model by examining the residuals, which is the difference between observed and predicted values. The aim is to ensure the residuals are random and do not contain any patterns or structures. The diagnostic checks involved the use of the Ljung-Box test and the forecasting and forecasting accuracy.

- *Ljung-Box Test*

The Ljung-Box test helps to check whether the errors or residuals in our model have any pattern or correlation. The Ljung-Box test is defined as

H_0 = Residuals are independently distributed, correlation in the population from which the sample is taken is 0

H_1 = Residuals are not independently distributed; they exhibit serial correlation.

The Ljung-Box test statistic is given as;

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.9$$

Where n is the sample size, $\hat{\rho}_k^2$ is the sample autocorrelation at lag k, and h is the number of lags being tested. Under null hypothesis the test statistic Q asymptotically follows a $\mathcal{X}^2_{(h)}$, for the significant level α, critical region is rejected if;

$$Q > \mathcal{X}^2_{(1-\alpha),h}$$

Where $\mathcal{X}^2_{(1-\alpha),h}$ is the $(1-\alpha)$ quantile of Chi-square distribution with h degrees of freedom.

After diagnostic checking, the fitted model will used in forecasting future values if the model is adequate. Otherwise, we need to repeat the selection and estimation method. Try with another potential candidate model (Ramasubramanian, 2007).

- *Forecasting and Forecasting Accuracy*

Once the selected model has been verified, it will then be used to predict power generation over the next 24 months. After forecasting, this study employs the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) to evaluate the forecasting accuracy of the chosen model. The MAE and MAPE are defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.10$$

and

$$MAPE = \frac{100}{n}\sum\left|\frac{(p_i - o_j}{o_j}\right| \quad \ldots\ldots\ldots\ldots\ldots\ldots 2.11$$

Where p_i is the predicted value for the i^{th} observations, o_i is the observed value for the j^{th} observation, n is the number of non-missing residuals.

Results and discussion

Time Series Plot for the Hydropower Generation

Figure 1 below illustrates fluctuations of the hydropower generation from January 2002 to December 2023. In the beginning, generation declined reaching its lowest point between 2006 and 2007 before bouncing back to its peak in 2008. The fluctuations then continued to go up and down for some time before reaching the highest peak of hydropower generation in Tanzania's history in 2020 followed by an abrupt decline. The graph clearly shows that the generation tends to increase and decrease around the central value repetitively. This pattern of movement demonstrates that the series appeared to be stationary.

Journal of Science, Innovation and Creativity
Vol. 4 No. 1 (2025): ISSN (Online): 2958-454X
DOI: https://doi.org/10.58721/jsic.v4i1.958

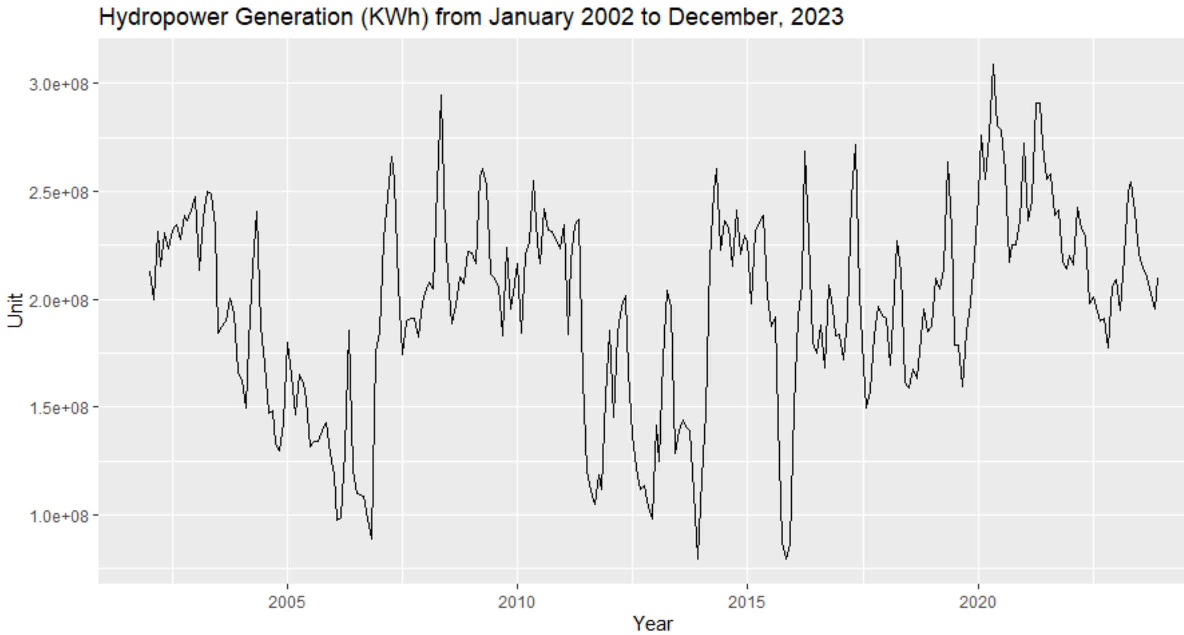

Figure 1: Time series plot of Hydropower generation (KWh) in Tanzania

Source(s): Created by author (s)

Testing for Stationarity: Augmented Dickey-Fuller (ADF) test
The results in Table 1 show that, the p-value of the test statistic is smaller than the 5% significance level, indicating that the null hypothesis is rejected. This implies that the data does not contain a unit root hence the series is stationary. The results of the ADF test are confirmed by the ACF and PACF correlograms in Figure 2.

Table 1: The result of the ADF test

Augmented Dickey-Fuller Test		
Dickey-Fuller = −3.76	Lag order = 6	P-value = 0.02147
Alternative hypothesis: Stationary		

Source(s): Created by author (s)

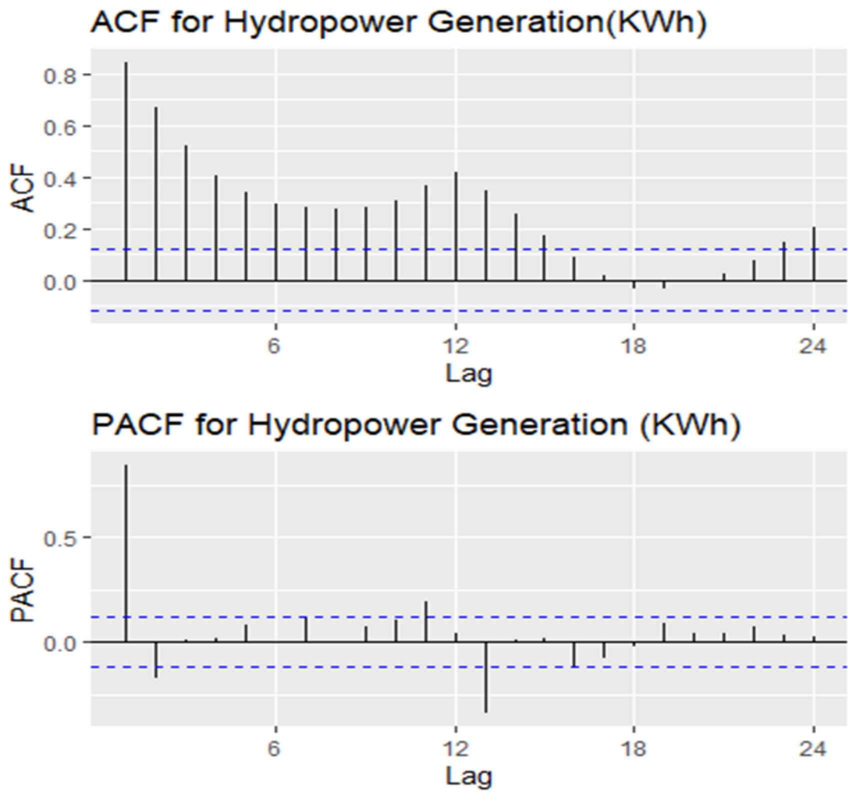

Figure 2: ACF and PACF plots for the Hydropower generation (KWh)

Source(s): Created by author (s)

Seasonality Test

To identify the most suitable model which fits the data well, the next step is to detect seasonality through seasonal decomposition. Figure 3 shows the result of seasonal decomposition using an additive model. As shown in the graph, the seasonal component displays a regular pattern, implying the presence of seasonality in the time series. This is consistent with the result of the seasonality test confirming that the time series exhibits a seasonal pattern.

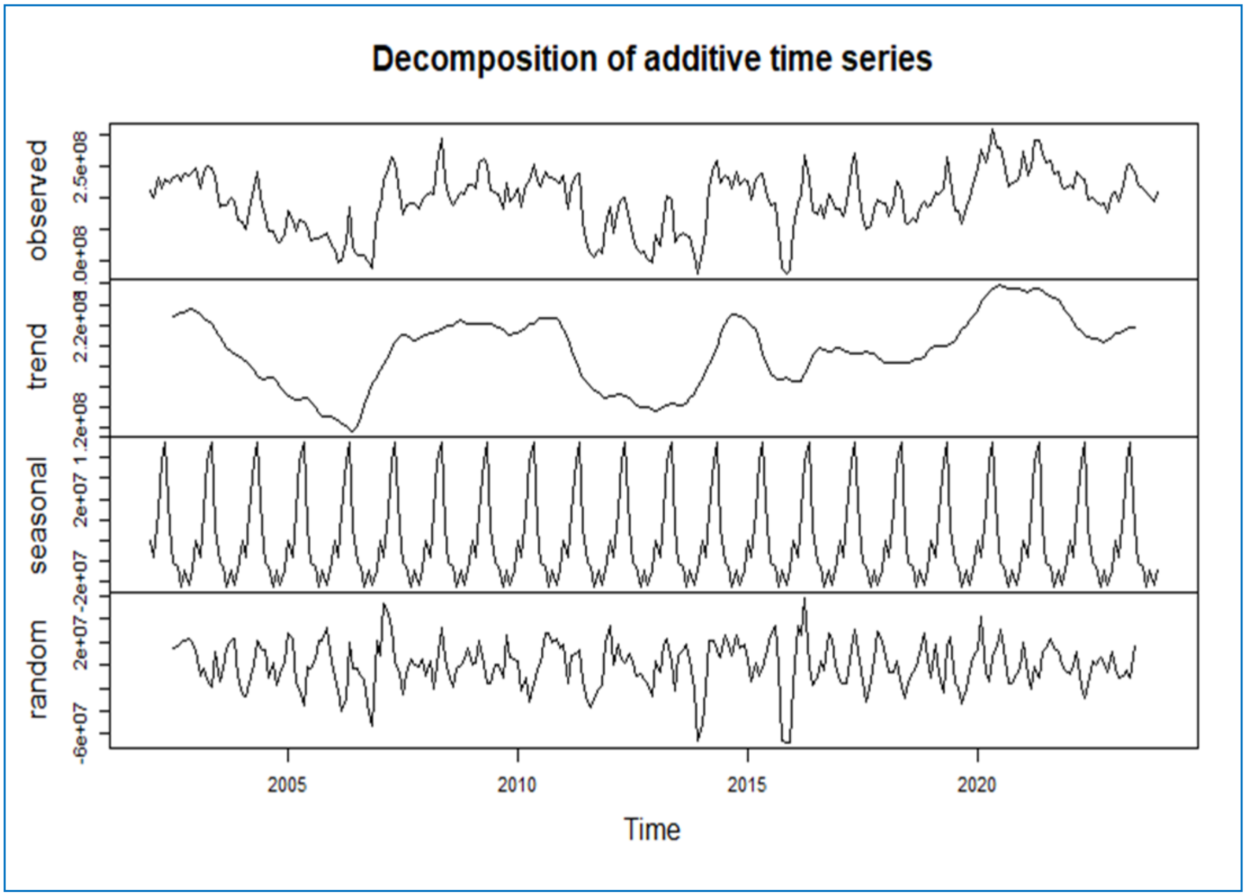

Figure: 3 Time series decomposition

Source(s): Created by author

Diagnostic checks
In doing the diagnosis checks, the study plotted the time plot, ACF and histogram for the residuals together with the Ljung-Box test. The idea was to see if the data fit the model under examination. The results of the plots and the Ljung-Box test are shown in Figure 4 and Table 2 below.

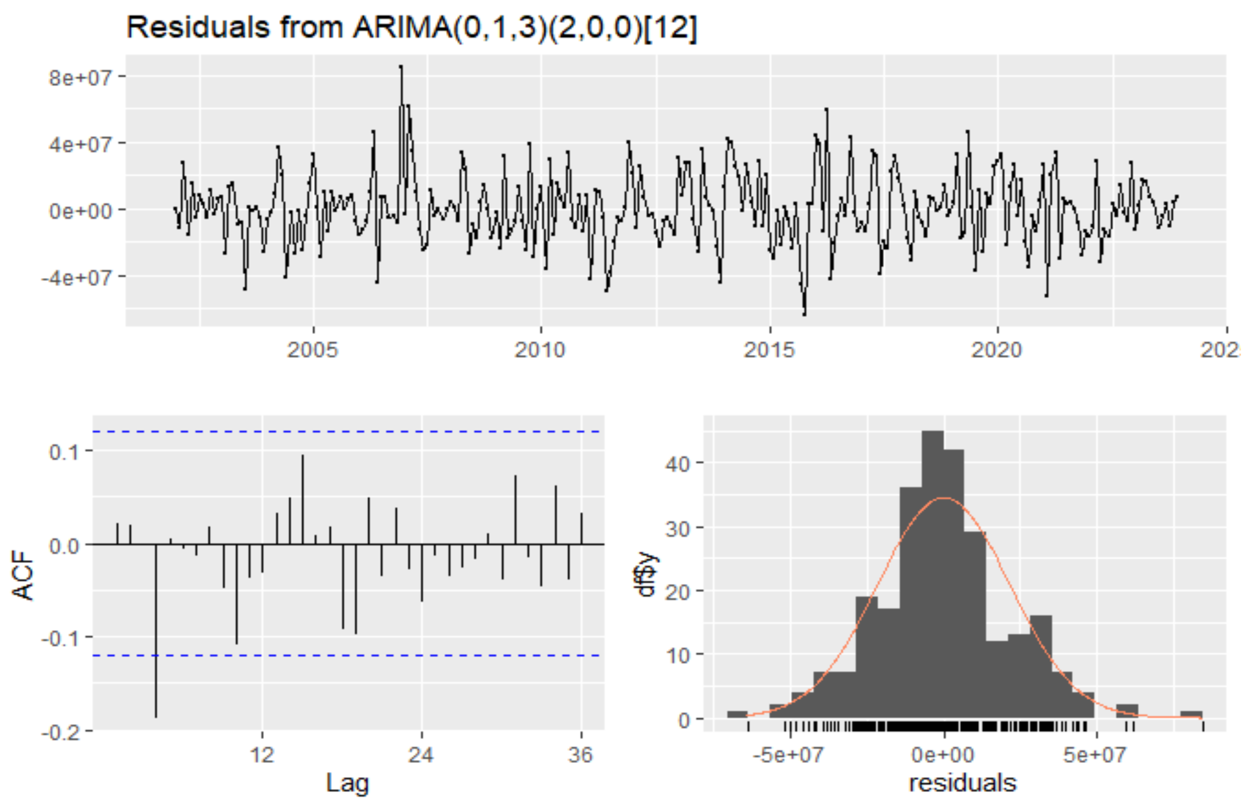

Figure 4: Adequacy check for ARIMA (0,1,3) (2,0,0) [12]

Source(s): *Created by author(s)*

Figure 4 shows the time plot, ACF plot, and histogram of the residuals of the ARIMA (0,1,3) (2,0,0)[12]. The time plot shows the residuals are white noise meaning zero mean and constant variance. The ACF plot indicates that the residuals are approximately uncorrelated. Lastly, the histogram revealed that the residuals are approximately normally distributed with a mean zero. The adequacy test result is also confirmed by the formal test Ljung- Box test in Table 2.

Ljung-Box Test

Table 2: Test result of Ljung-Box test

Q^*	**df**	**p-value**
26.096	19	0.1275

Source(s): *Created by author(s)*

Table 2 above shows the results of the Ljung-Box test on the residuals of the model. The p-value of the test statistic is greater than 0.05 significance level, signifying that residuals are uncorrelated, and they are pure randomly.

Model selection and estimation

The function auto. Arima () in R software was used to select the best model. The function provided the best model automatically for forecasting hydropower generation using the minimum value of the Akaike Information Criterion (AIC) Bayesian Information Criterion (BIC) and Maximum log likelihood. The Seasonal ARIMA (0,1,3) (2,0,0)[12] was selected as the appropriate model due to the lowest AIC of 9645.71 and BIC of 9667.14 and the largest log-likelihood equals −4816.85 among the other models. The selected model was considered for forecasting hydropower generation.

Model Validation
To check if the model is best for forecasting, its forecasted predictions for the validation set were plotted against the observed values. Figure 5 below shows the graph of the fitted predicted values using the original data. The results revealed that the model is best for the data set.

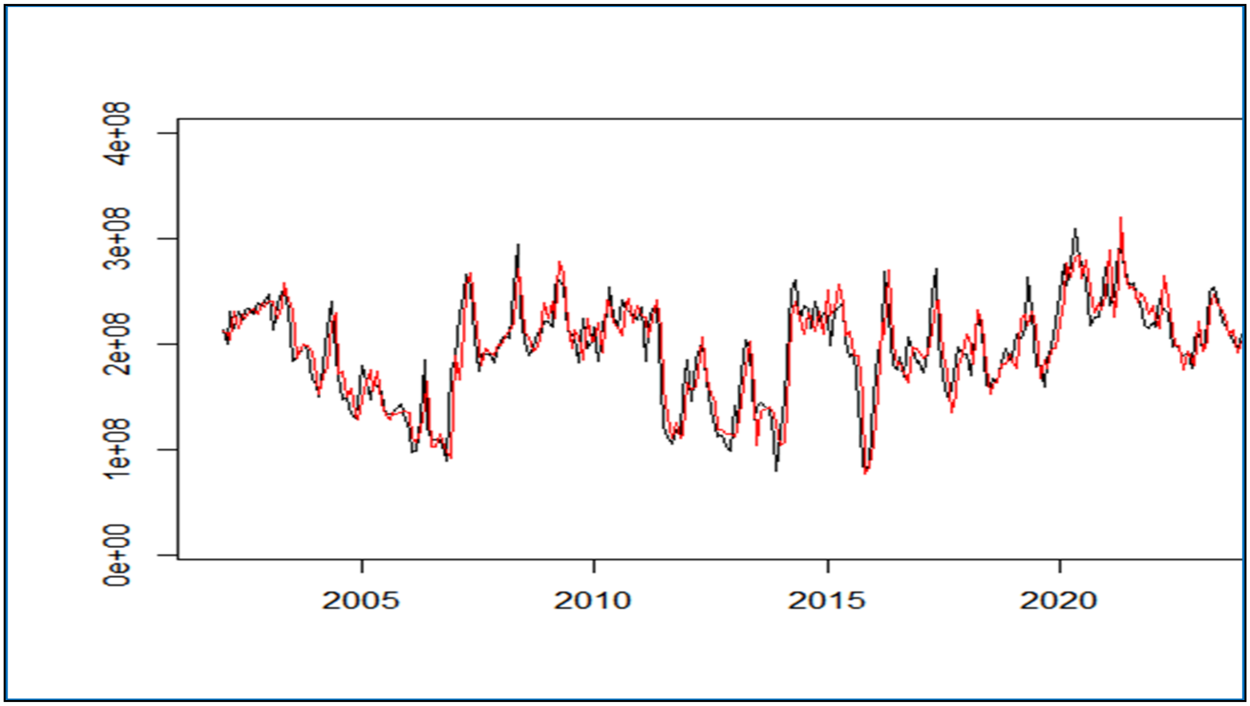

Figure 5: Model validation of Electricity Generation using SARIMA (0,1,3) (2,0,0)[12]

Source(s): *Created by author(s)*

Forecasting
Table 3 and Figure 6 below display the forecasted values of the hydropower generation for the coming 24 months using the Seasonal ARIMA (0,1,3) (2,0,0)[12] model.

Table 3: Forecast power generation for the next two years (2024-2025)

Time	Forecast (KWh)
Jan-24	213,388,145
Feb-24	206,891,694
Mar-24	221,614,520
Apr-24	225,664,671
May-24	225,590,962
Jun-24	210,939,034
Jul-24	207,203,073
Aug-24	203,181,605
Sept-24	200,612,552
Oct-24	199,418,429
Nov-24	192,459,441
Dec-24	206,212,337
Jan-25	208,287,733
Feb-25	201,529,395
Mar-25	213,982,131
Apr-25	225,830,644
May-25	227,250,650
Jun-25	218,706,596
Jul-25	210,953,305
Aug-25	207,661,349
Sep-25	205,561,214
Oct-25	202,686,340
Nov-25	198,132,402
Dec-25	206,646,426

Source (s): Created by author (s)

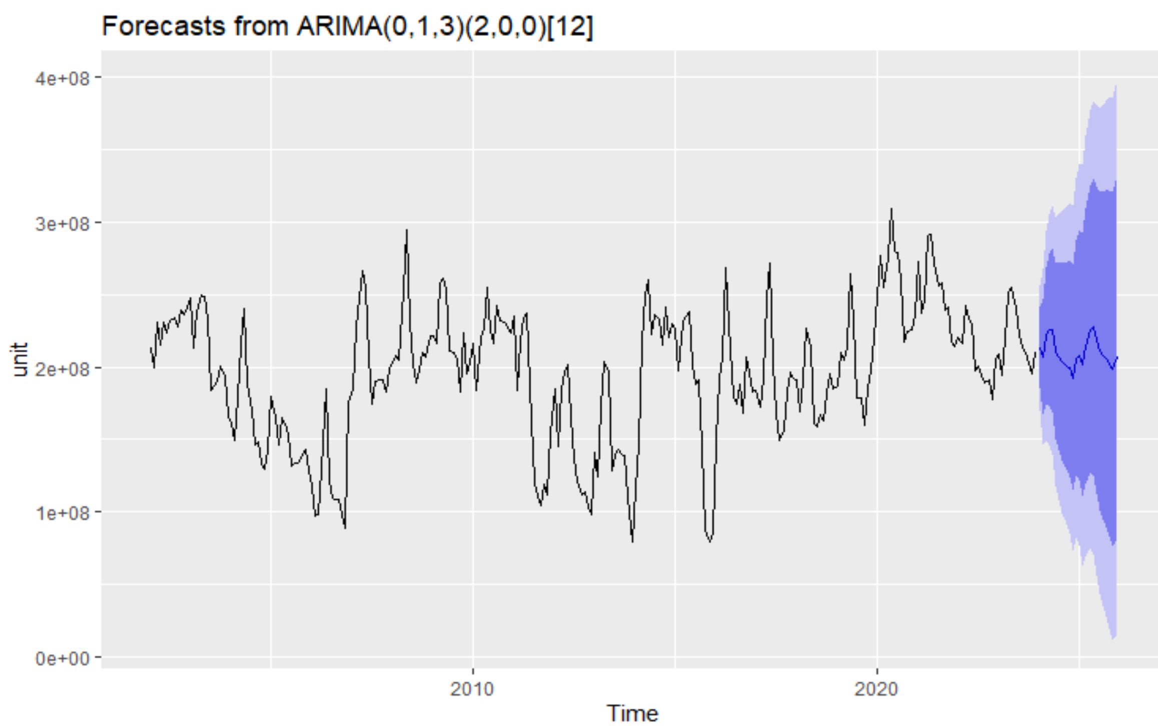

Figure 6: Forecast for hydropower generation (KWh)

Source(s): *Created by authors*

Conclusion

This study forecasted the hydropower production in Tanzania using the SARIMA model. Precisely, we used the seasonal ARIMA (0,1,3) (2,0,0)[12] model with seasonal data. The analysis revealed that, in the next forecasted months until December 2025, there will be no substantial decline in hydropower production (KWh). Even though the hydropower generated generally will not exceed 227, 250, 650 KWh, it is evident that will be at its peak in May 2025 and there will be a decrease towards December 2025 with not more than a 40% decrease in every forecasted month. Further, the forecasts suggest that there would be seasonal shortages in the production of hydropower in the forecasted years. Despite such a shortage, the solution would be mild power rationing which has been the case whenever such an unexpected shortage occurs. In such cases, there will be places having electricity at one time while other places will not have it as a result of rationing. The methodology that has been used in this paper (SARIMA) serves to help in determining future hydropower generation in the country and will enable the policymakers in Tanzania and government officials to make a well-informed decision to improve electricity supply which shows no significant change for the coming months. However, due to the fluctuation in the data series, this research needs to be extended by applying other methodologies such as the Autoregressive Integrated Moving Average with exogenous variable (SARIMAX), Simple Exponential Smoothing (SES) or the Holt-Winters Exponential Smoothing (HWES), the methodologies that may improve the results from this study but also widening the scope on how the forecasting of the hydropower generation can yield best results.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on automatic control,* 19(6), 716-723.

Barzola-Monteses, J., Mite-León, M., Espinoza-Andaluz, M., Gómez-Romero, J., & Fajardo. W. (2019). Time series analysis for predicting hydroelectric power production: The Ecuador case. *Sustainability,* 11(23), 6539.

Box, G. E. P., Jenkins, G. M., Reinsel, G.C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Chatfield, C., & Prothero, D. L. (1973). Box-Jenkins seasonal forecasting: problems in a case study. *Journal of the Royal Statistical Society*: Series A (General) 136(3), 295-315.

Condemi, C., Casillas-Perez, D., Mastroeni, L., Jiménez-Fernández, S., & Salcedo-Sanz, S. (2021). Hydro-power production capacity prediction based on machine learning regression techniques. *Knowledge-Based Systems,* 222, 107012.

Dickey, D. A., & Wayne A. F. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74(366a), 427-431.

Hyndman, R. J. (2018). *Forecasting: principles and practice*. OTexts.

Javed, U., Moazam Fraz, M., Mahmood, I., Shahzad, M., & Omar Arif. (2020). Forecasting of electricity generation for hydropower plants. In 2020 *IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET*), pp. 32-36. IEEE.

Kichonge, B., John, G. R., Mkilaha, I. S. N & Hameer, S. (2014). Modelling of future energy demand for Tanzania. *Journal of Energy Technologies and Policy, 4*(7), 16-31.

Kullback, S., & Richard A. L. (1951). On information and sufficiency. *The annals of mathematical statistics, 22*(1), 79-86.

Lee, D., Yi Ng, J., Galelli, S., & Block. P. (2022). Unfolding the relationship between seasonal forecast skill and value in hydropower production: a global analysis. *Hydrology and Earth System Sciences, 26*(9), 2431-2448.

Lwaho, J., & Ilembo. B. (2023). Unfolding the potential of the ARIMA model in forecasting maize production in Tanzania. *Business Analyst Journal, 44*(2), 128-139.

Mite-León, M., & Barzola-Monteses, J. (2018). Statistical model for the forecast of hydropower production in Ecuador. *International Journal of Renewable Energy Research, 8*(2), 1130-1137.

Monteiro, C., Ramirez-Rosado, I. J., & Fernandez-Jimenez, A L. (2013). Short-term forecasting model for electric power production of small-hydro power plants. *Renewable Energy 50*, 387-394.

Naylor, T.H., Seaks, T.G., & Wichern, D.W. (1972). Box-Jenkins methods: An alternative to econometric models. *International Statistical Review/Revue Internationale de Statistique*, 123-137.

Polprasert, J., Nguyên, V., & Charoensook. N.S. (2012). Forecasting models for hydropower production using the ARIMA method. *In 2021 9th International Electrical Engineering Congress (IEECON*), pp. 197-200. IEEE.

Ramasubramanian, V. (2007). *Time series analysis*. New Delhi.

Sarkodie, S. A. (2017). Estimating Ghana's electricity consumption by 2030: An ARIMA forecast. *Energy Sources, Part B: Economics, Planning, and Policy* 12(10), 936-944.

Sarpong, S. A., & Agyei, A. (2022). Forecasting Hydropower Generation in Ghana Using ARIMA Models. *International Journal of Statistics and Probability, 11*(5).

Tembo, A., Ilembo, B., & Lwaho, J. (2024). Forecasting the National Health Insurance Fund Membership Enrolment in Tanzania Using the SARIMA Model. *SCIENCE MUNDI, 4*(2), 29-39.

Thompson, H.E., & Tiao, G.C. (1971). Analysis of telephone data: A case study of forecasting seasonal time series. *The Bell Journal of Economics and Management Science*, 515-541.

Wang, Z., Li, Q., & Pei. L. (2017). Grey forecasting method of quarterly hydropower production in China based on a data grouping approach. *Applied Mathematical Modeling* 51, 302-316.

Wiah, E. N., Buabeng, A., & Agyarko. K. (2022). Statistical Model for the Forecast of Electricity Power Generation in Ghana. *Open Journal of Statistics, 12*(3), 373-384.

Zhou, F., Li, L., Zhang, K., Trajcevski, G., Yao, F., Huang, Y., Zhong, T., Wang, J., & Liu, Q. (2020). Forecasting the evolution of hydropower generation. *In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2861-2870.

Zolfaghari, M., & Golabi, M. R. (2021). Modelling and predicting the electricity production in hydropower using a conjunction of wavelet transform, long short-term memory and random forest models. *Renewable Energy*, 170, 1367-1381.