Bima Journal of Science and Technology, Vol. 8(1B) Apr, 2024 ISSN: 2536-6041

DOI: 10.56892/bima.v8i1B.646

Prediction of Monthly Mean Surface Air Temperature Using SARIMA in Jos North, Plateau State, Nigeria

Dayyab Abdulkarim Shitu*, Ahmed Abdulkadir, F. U. Abbas, Ali Muhammed Gambo, Sadiya Abdullahi Baban Mairam and Sheyi Mafolasire

Abubakar Tafawa Balewa University Bauchi.

Corresponding Author: dayyababdulkarim@gmail.com

ABSTRACT

Fluctuation in temperature causes an untold hardship to human, animals and plants. In this study, the forecasting model was developed using the Seasonal Auto Regressive Integrated Moving Average (SARIMA) model, which is based on the Box-Jenkins system. The monthly mean temperature data for Jos city was acquired from Meteorology and Climatology unit of the Geography Department University of Jos within the period thirty two years (January 1986 to December 2023), The SARIMA model most appropriate orders were determined using the autocorrelation and partial autocorrelation functions for time series results. The monthly mean Temperature was used to verify the validity of these models. Statistical criteria such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-square (R^2) were used to measure the model's accuracy and compare them. The model SARIMA$(4,1,6)(2,1,2)^{12}$ was chosen as the most reliable model and was used in predicting the monthly mean temperature for the study area.

Keywords: Monthly mean temperature, Time series forecasting, SARIMA model, Auto-Regressive, Moving Average.

INTRODUCTION

Time series is a collection of observations that are generated in a specific order across time. It is a statistical series that shows how data has behaved in the past, present, and maybe in the future (Becker et al., 2015). In essence, it is a technique that combines observed data from the past y_{t-1} and present y_t to forecast future values y_{t+1} with a temporal lag of one (Jingjing, 2013), The numerical values of the variable Y (temperature) at the times t_1, t_2, and t_3 are mathematically known as a time series in mathematics. Y is a function of t as such Y is given by $Y = F(t)$. (Sake & Akhtar, 2019).

The time intervals may be every second, minute, hour, day, month, year, or decade (Pearce, et al; 2011). Temperature is an example of a time series with daily, monthly, and annual order (Zhou, et al; 2019). Temperature data is one of the many data where time series is utilized to observe, analyze, and anticipate the future using various models. Data in a time series is normally arranged by time, and subsequent observations are usually dependent (Citakoglu, 2021). Since the turn of the 20th century, the Earth's surface temperature has increased by about 0.8 °C on average, and according to some climate models, this century will see an increase of 1.1 to 6.4 degrees Celsius (Katopodis & Sfetsos, 2019).

Recent years have seen an exponential data increase due to the advent of new information channels, such as social networking services. Big data analysis or high-capacity data becomes available as data storage techniques advance. Big data analysis frequently yields some important conclusions. Specifically, diverse time series large data are accumulated across multiple institutions (Jeong & Kim, 2013).

In the study of temperature time-series, many methodologies have been used, for example: Lornezhad et al, (2023) evaluate the trend of precipitation change using Mann–Kendall method in the western part of Iran. The results of the research on an annual scale showed that all stations have a significant negative trend at the level of 5%, which indicates the existence of a negative trend, or in other words, a decrease in irrigation in the studied stations. Becker et al, (2015) Combined prediction model based on Volterra and VMD-DSE. According to his experimental findings, the VMD-DSE-Volterra model performs better than other benchmark models in predicting the monthly mean temperature.

Similarly, Ashraf et al., (2021) Uses the Mann-Kendall (MK), Spearman's rho (SR), and innovative trend analysis (ITA) methodologies to examine variations in the monthly, seasonal, and yearly stream flow time series at twenty (20) sites over the upper Indus river basin (UIRB). Dammo, et al, (2015); Salaudeen, et al, (2021) Study temporal and seasonal variation in temperature over north-eastern Nigeria, trends in annual and seasonal temperature series were analysed using Mann-Kendall test. Lopes & Tenreiro, (2014) investigate the intricate relationships between the time series of the global temperature by using the Multidimensional scaling (MDS) method.

Ofure et al, (2021) Utilized SARIMA model to predict the monthly temperature in Ghana's northern area. The scientists used temperature information from January 1990 through December 2020. The SARIMA $(1,0,0)(1,0,0)_{(12)}$ is therefore the most accurate model for forecasting monthly temperatures in Ghana's northern area. Tran et al, (2021) Applied Seasonal Autoregressive Integrated Moving Average (SARIMA) to create temperature forecasting models for the state of Kerala. The model is developed using mean maximum and mean lowest monthly temperature data collected over a 47-year period from seven locations. SARIMA$(2,1,1)(1,1,1)^{(12)}$ was shown to be the best forecasting model for eight of the fourteen time-series datasets, according to the study.

Tao et al., (2021) compared the accuracy of the Support Vector Regression (SVR) and SVR-FA models to that of the Seasonal Autoregressive Integrated Moving Average (SARIMA) Stochastic model, a meta-innovative model that combines SVR with the Firefly optimization algorithm. It was advised to utilize SVR-FA rather than SVR models for temperature forecasting because of the efficiency of merging the SVR model with the Firefly optimization algorithm in predicting temperatures in Iran's climate. Asamoah-Boaheng, (2014) utilized the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to develop a forecasting system that is adequate for the mean temperature of the Ashanti Region. Using the calculated AIC and BIC values for each of the proposed models, the best model, SARIMA $(2,1,1)(1,1,2)^{(12)}$, was chosen.

Based on the literature, majority of the authors in Nigeria used a non-parametric Mann-Kendall test despite the fact that the non-parametric Mann-Kendall test is not suited for data with periodicities (i.e seasonal effect). Mann-Kendall test is used to detect linear trend in time series data. However, it may not capture the nonlinearity in the data (Salaudeen et al., 2021). SARIMA model gives better predictions to the temperature data and most of hydrological data are data with both seasonal and non-seasonal effect (Zhu et al., 2021)

The SARIMA model posits that future values of a time series are linearly related to present

DOI: 10.56892/bima.v8i1B.646

and previous values (Valavi et al., 2022). The time series under examination is formed by linear processes, which is one of the main assumptions of the Autoregressive Integrated Moving Average (ARIMA) methodology and a result of traditional prediction approaches. The approach for estimating linear change is investigated in this study, and a framework for measuring linear change is constructed. Finally, a case study regional analysis of linear change was undertaken to demonstrate how the methodology might be used to investigate temperature increases.

This study's objective was to use SARIMA model to investigate the rise in local temperature in Jos North Local Government Area in order to suggest preventive measures for the food industry, the tourism, and other sectors.

MATERIALS AND METHODS

Seasonal Autoregressive Integrated Moving Average or SARIMA models are sometimes known as Box-Jenkins models. The abbreviations MA and AR (p) denote moving average models with q terms and an autoregressive model of order p. (q), respectively. An ARMA (p,q) model is a model that combines p AR-terms with q MA-terms. A frequently non-stationary time-series that has been time-shifted (by d delays, where d is normally equal to 1) before additional processing is computed.

Such a model is categorized as ARIMA (p, d, q), with the letter "I" standing for "integrated." An ARMA (p,q) model for this stationary time series, presuming that the original data Xt has been made stationary by subtracting d non-seasonal deviations (where d = 1 in most situations). Then Yt is as follows,

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t - \Theta_1 e_{t-1} - \Theta_2 e_{t-2} - \dots - \Theta_q e_{t-q} \quad (1)$$

Where c = constant term,

φ = autoregressive parameter,

θ = moving average parameter,

e_t = the error term at time t.

It is more fitting to use a seasonal ARIMA (p, d, q) (P, D, Q)^S model for climate data that follows a seasonal, i.e. annual cycle, where P is the order of the seasonal AR-model; D is the order of the seasonal differencing (for monthly data, D = 12), Q is the order of the seasonal MA-model, and s is the number of cycles in the season (s = 12 for an annual cycle).

In backshift notation, the general form of such a seasonal ARIMA (p, d, q)(P, D, Q)^S model is:

$$\varphi_{AR}(B)\varphi_{SAR}(B^s)(1-B)^d(1-B^s)^D Y_t = \Theta_{MA}(B)\Theta_{SMA}(B^s)e_t \quad (2)$$

Where, φ_{AR} is non-seasonal AR- parameter, Θ_{MA} is non-seasonal MA-parameter, φ_{SAR} is seasonal AR parameter, Θ_{SMA} is seasonal moving average parameter, B is backward shift operator.

Box and Jenkins suggested a four-phase methodology for identifying a perfect ARIMA model for a given time series, namely,

i) Model identification
ii) Estimation of the model parameters
iii) Diagnostic checking for the identified model appropriateness and
iv) Application of the final model, i.e. forecasting.

Box and Jenkins (1976) provide further information on estimating the model's parameters and order.

DOI: 10.56892/bima.v8i1B.646

Forecast Evaluation Measures

In order to acquire trustworthy results for making predictions and evaluating the model's output, certain parameters were taken into account. The final goal is to select the most appropriate model. Performance indices such as MAE, RMSE, MAPE, and R^2 were used to compare and quantify the model's accuracy, and were used to estimate the generalization error.

The RMSE and MAPE are both indications of the difference between the expected and actual values. The R indexes are used to assess the relationship between expected and actual values (Asamoah-Boaheng, 2014). However, the most widely used measures to evaluate the strategies implemented for forecasting tasks, are:

- Mean Absolute Error (MAE): This measure is an error statistic that averages the distances between the estimated and the observed data for N samples:

$$MAE = \frac{\sum_{t=1}^{n} |y - \hat{y}_t|}{N} \quad (3)$$

- Root Mean Square Error (RMSE): This measure is the standard deviation of the difference between the estimation and the true observed data. This measure is more sensitive to big prediction errors

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} |y - \hat{y}_t|^2}{N}} \quad (4)$$

- Mean Absolute Percentage Error (MAPE): This measure offers a proportionate nature of error with respect to the input data. It is defined as:

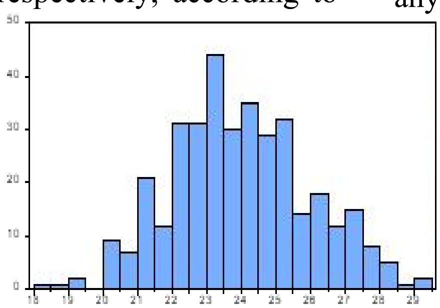
$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{y_i} \times 100 \quad (5)$$

- R-squared value, also known as the coefficient of determination, is a statistical measure that indicates how well the independent variables in a regression model explain the variability of the dependent variable. It is a measure of the goodness-of-fit of the model.

$$R = \frac{\sum_{i=1}^{N} (Oi - \bar{O})(Pi - \bar{P})}{\sqrt{\sum_{i=1}^{N} (Oi - \bar{O})^2 \sum_{i=1}^{N} (Pi - \bar{P})^2}} \quad (6)$$

RESULTS AND DISCUSSION

The mean temperature data of the Jos city within the period of Thirty seven years (1986 -2023) was adopted in this study to generate time series analysis using Eviews and R software. The model's outcomes were evaluated using the RMSE, MAPE, MAE, and R^2 of the testing data set. The mean and standard deviation of the temperature are 23.748 and 1.999, respectively, according to the plot and normality test on temperature series in figure 1. The graph is regularly distributed, asymmetric, and leptokurtosis, with a Skewness of 0.1264 and a Kurtosis of 2.7458. Jarque-Bera test was 1.927852 at a 5% significance level; this is a negligible probability, causing the null hypothesis of a normal distribution to be rejected. The original series, on the other hand, can be utilized for estimating and forecasting with any model (Araghi, et al, 2017).



Series: MEAN MONTHLY TEMPERATURE	
Sample 1 360	
Observations 360	
Mean	23.74806
Median	23.50000
Maximum	29.00000
Minimum	18.00000
Std. Dev.	1.999902
Skewness	0.126384
Kurtosis	2.745772
Jarque-Bera	1.927852
Probability	0.381393

Figure 1: Histogram plot and normality test of the temperature data series.

Bima Journal of Science and Technology, Vol. 8(1B) Apr, 2024 ISSN: 2536-6041

DOI: 10.56892/bima.v8i1B.646

Figure 2 depicts a random walk indicating cycles in the series, and also shows a pattern of seasonality in the series at level. It is therefore important to confirm seasonality and assess whether there is overall trend in the data over time in other to understand the long-term behaviour of the variable.

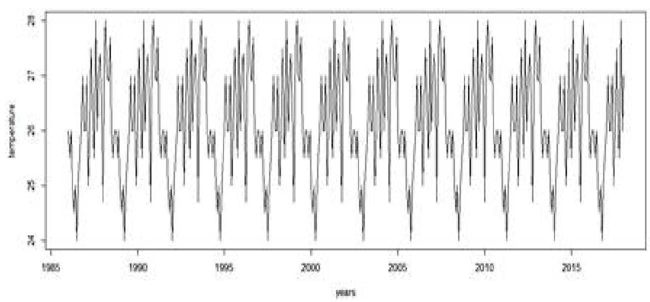


Figure 2: Time plot for the mean monthly temperature data.

To confirm seasonal, cyclical, and trend in the series data, we plot a residual ACF and PACF as shown in figure 3 and 4 respectively. The ACF and PACF plot were used to capture the temporal structure of the time series. The plot of the residual ACF in figure 3 shows correlation coefficients at various lags, taking into account the seasonal components. In this plot, the significant spikes occur at lag 12 suggesting a seasonal pattern with an annual cycle. The plot of the residual PACF in figure 4 shows partial correlation coefficient at various lags, controlling the intermediate lags including the seasonal lags. This confirms that the data in this study is suitable for time series SARIMA modeling.

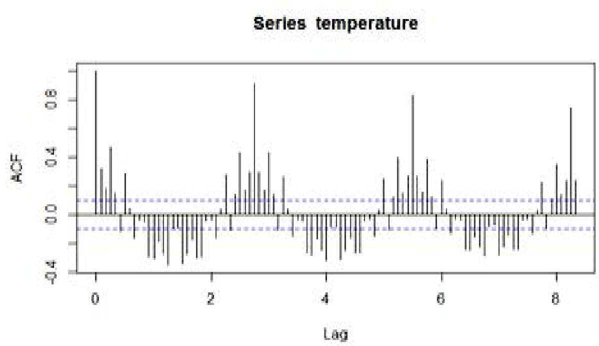


Figure 3: residual ACF plot. `          `

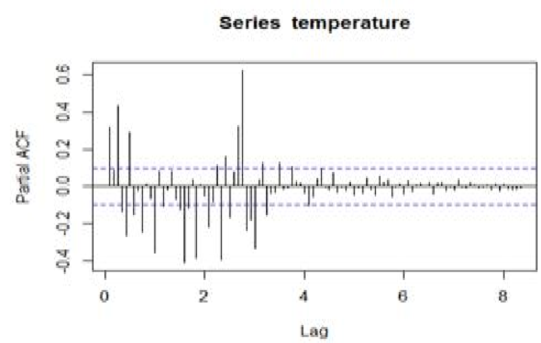
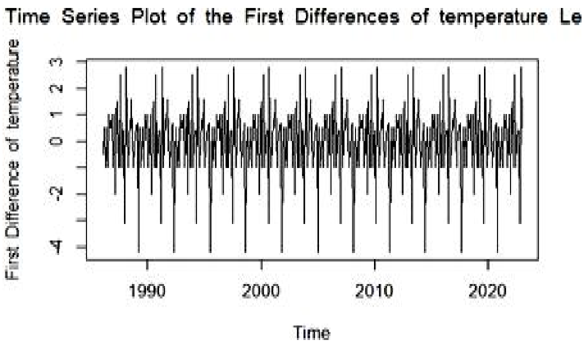


Figure 4: residual PACF plot.

In other to obtained stationarity and improve the forecasting accuracy, the seasonal influence must be eliminated (Asamoah-Boaheng, 2014). To remove the seasonality characteristic from the test time series data, the first-order seasonal difference was used.

According to Figures 5 and 6 obtained using R software. The majority of the spikes in the series move around the zero mean, implying that the series is stationary after the first order differencing. But looking at the plot alone isn't enough; we also look at the mean, which is 0.03217158 and 0.002688172 respectively, both of which are very near to zero, confirming stationarity (Asamoah-Boaheng, 2014).

DOI: 10.56892/bima.v8i1B.646



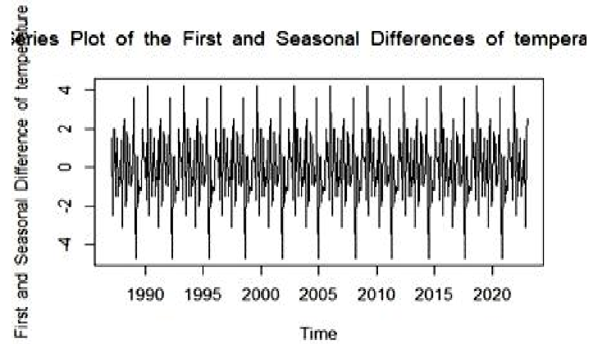
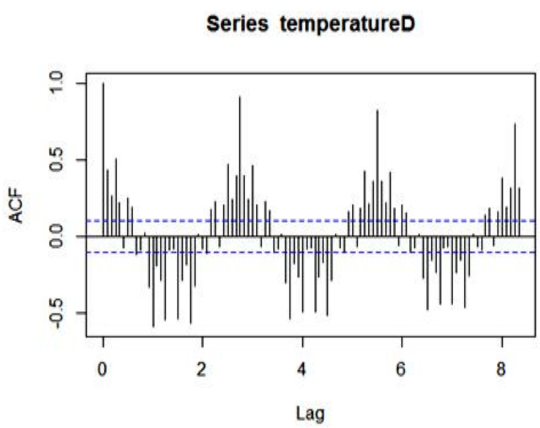


Figure 5: first and seasonal difference. **Figure 6:** first order and seasonal difference plot.

The correlogram of the seasonal difference of the series under study was shown in Figures 7 and 8 respectively. Seasonal patterns are frequently visible. The correlogram implies that both the studied variables lags and the error lags should be included in the process estimation. Some of the lags where identified for testing by the point where the ACF/PACF cuts-off.



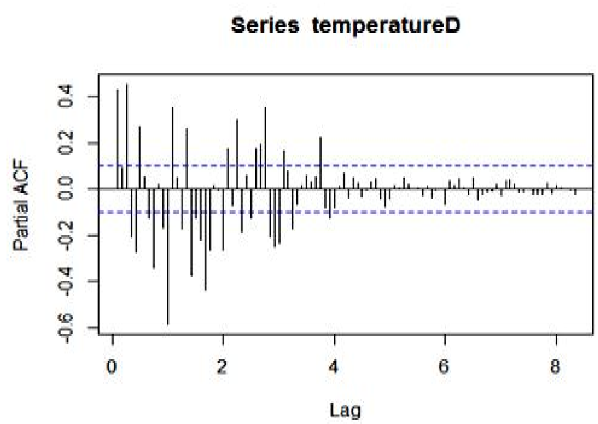


Figure 7: first order derivative of residual ACF plot. **Figure 8:** first order derivative of residual PACF plot.

The identified SARIMA model must be diagnostically verified for its appropriateness, as recommended by step iii) Box-Jenkins methodology, by looking at the ACF and PACF of the model residuals (Asamoah-Boaheng, 2014). Figures 7 shows spikes at various lags within statistical confidence bands in the ACF graphs. In an iterative process based on trial and error, several configurations were carried out using the various changeable factors (see Table 2). Furthermore, the residuals were analyzed to confirm that they met the requirements of a white noise process using R software.

Table 1: SARIMA Models Identified and there forecasting accuracy

Models	Selected Models	RMSE	MAE	MAPE	R^2
Model 1	SARIMA(1,1,1)(2,1,1)$_{12}$	0.4122259	0.3343837	1.279666	0.1758
Model 2	SARIMA(4,1,3)(2,1,2)$_{12}$	0.419522	0.3536685	1.354218	0.1821
Model 3	SARIMA(3,1,1)(2,1,1)$_{12}$	0.445536	0.338611	1.300823	0.2054
Model 4	SARIMA(4,1,4)(2,1,1)$_{12}$	0.4964809	0.4196448	1.602914	0.2551
Model 5	**SARIMA(4,1,6)(2,1,2)$_{12}$**	**0.2861076**	**0.2233394**	**0.85398**	**0.5847**
Model 6	SARIMA(5,1,6)(2,1,1)$_{12}$	0.3690232	0.2994044	1.14306	0.1409
Model 7	SARIMA(2,1,3)(2,1,2)$_{12}$	0.4367411	0.352108	1.341968	0.1974

Forecasting according to Box-Jenkins is the final phase. The most important factor to consider when choosing a forecast model is its accuracy, or how well it forecasts actual data (Asamoah-Boaheng, 2014). Table 2 explained an evaluation criterion. All of the models had good results, but the SARIMA (4,1,6)(2,1,1,)^{12} model had the best forecasting performance due to its low errors in forecasting. The best model was used to predict the mean monthly temperature of Jos city as shown in Figure 9.

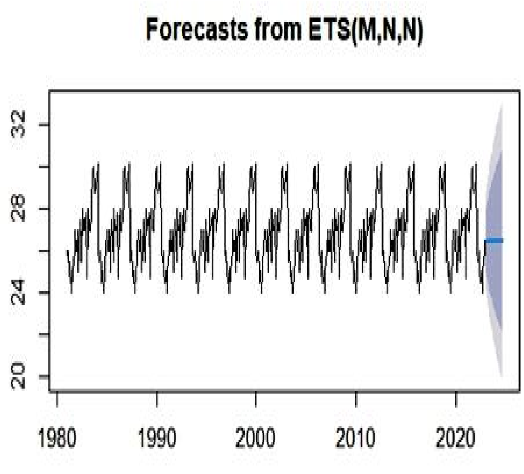


Figure 9: Plot of the time series forecast of the temperature data

CONCLUSION

In this work, the Box-Jenkins method was applied to develop monthly mean temperature of Jos city using R program. Many SARIMA models were presented and were based on the autocorrelation and partial autocorrelation functions of the time series. The seasonal ARIMA(4,1,6)(2,1,2)^{12} model gave the most accurate results out of all of these models, and was used to predict the monthly temperature in Jos. A Nomo-graph chart is used in place of the forecasting equation to directly account for future temperature.

These findings can be applied to real-world data, and some business ramifications appear to be available. This work adds to the field by providing an automatic forecasting technique to reduce temperature prediction mistakes.

Several problems have been established based on the prediction results of the entire time series. The method fails particularly when the time series have nonlinearity. This condition causes the time series to have unpredictable turning points that the model cannot accurately match. The findings imply that the innovative SARIMA model approaches are viable, appropriate, and trustworthy technique for predicting the mean temperature.

Lastly, this study recommends potential extensions or variations of the SARIMA model, such as SARIMAX (including exogenous variables), dynamic SARIMA models, or hybrid approaches in other to enhance the problem of nonlinearity in time series analysis.

REFERENCES

Araghi, A., Mousavi-Baygi, M., Adamowski, J., Martinez, C., & van der Ploeg, M. (2017). Forecasting soil temperature based on surface air temperature using a wavelet artificial neural network. *Meteorological Applications*, *24*(4), 603–611. https://doi.org/10.1002/MET.1661

Asamoah-Boaheng, M. (2014). Using SARIMA to Forecast Monthly Mean Surface Air Temperature in the Ashanti Region of Ghana Using SARIMA to Forecast Monthly Mean Surface Air Temperature in the Ashanti Region of Ghana Using Seasonal Autoregressive Integrated Moving Average (SARIMA) mode. *International Journal of Statistics and Applications*, *4*(6), 292–299. https://doi.org/10.5923/j.statistics.20140406.06

Ashraf, M. S., Ahmad, I., Khan, N. M., Zhang, F., Bilal, A., & Guo, J. (2021). Streamflow Variations in Monthly, Seasonal, Annual and Extreme Values Using Mann-Kendall, Spearmen's Rho and Innovative Trend Analysis. *Water Resources Management*, *35*(1), 243–261. https://doi.org/10.1007/S11269-020-02723-0/METRICS

Becker, F. G., Cleary, M., Team, R. M., Holtermann, H., The, D., Agenda, N (2015) time series prediction: forecasting the future and understanding the past. In *Syria Studies* (Vol. 7).

Box and Jenkins. (n.d.). Use of the box and jenkins time series technique in traffic forecasting Retrieved April 12, 2024, from https://www.researchgate.net/publication/285902264_ARIMA_The_Models_of_Box_and_Jenkins

Citakoglu, H. (2021). Comparison of multiple learning artificial intelligence models for estimation of long-term monthly temperatures in Turkey. *Arabian Journal of Geosciences*, *14*(20), 1–16.

Dammo, M., Abubakar, B., & Sangodoyin, A. (2015). Trend and Change Analysis of Monthly and Seasonal Temperature Series over North-Eastern Nigeria. *Journal of Geography, Environment and Earth Science International*, *3*(2), 1–8. https://doi.org/10.9734/jgeesi/2015/18512

Jeong, C., & Kim, M. S. (2013). *Comparison Studies of Hybrid and Non-hybrid Forecasting Models for Seasonal and Trend Time Series Data* *. *19*(1), 1–17.

Jingjing, Z. (2013). Priors for time series forecasting. *Applied Mechanics and Materials*, *263–266*(PART 1), 171–174. https://doi.org/10.4028/www.scientific.net/AMM.263-266.171

Katopodis, T., & Sfetsos, A. (2019). A Review of Climate Change Impacts to Oil Sector Critical Services and Suggested Recommendations for Industry Uptake. *Infrastructures 2019, Vol. 4, Page 74*, *4*(4), 74. https://doi.org/10.3390/INFRASTRUCTURES4040074

Lopes, A. M., & Tenreiro Machado, J. A. (2014). Analysis of temperature time-series: Embedding dynamics into the MDS method. *Communications in Nonlinear Science and Numerical Simulation*, *19*(4), 851–871. https://doi.org/10.1016/J.CNSNS.2013.08.031

Lornezhad, E., Ebrahimi, H., & Rabieifar, H. R. (2023). *Analysis of precipitation and drought trends by a modi fi ed Mann – Kendall method : a case study of Lorestan province , Iran*. *23*(4), 1557–1570. https://doi.org/10.2166/ws.2023.068

DOI: 10.56892/bima.v8i1B.646

Ofure Eichie, J., Oluwamayowa Agidi, E., & David Oyedum, O. (2021). Atmospheric Temperature Prediction across Nigeria using Artificial Neural Network. *ACM International Conference Proceeding Series*, 280–286. https://doi.org/10.1145/3508072.3508114

Pearce, J. L., Beringer, J., Nicholls, N., Hyndman, R. J., & Tapper, N. J. (2011). Quantifying the influence of local meteorology on air quality using generalized additive models. *Atmospheric Environment*, *45*(6), 1328–1336. https://doi.org/10.1016/J.ATMOSENV.2010.11.051

Sake, R., & Akhtar, M. (2019). Fitting Of Gompertz Model Between Rainfall And Ground Water Levels – A Case Study. *International Journal of Mathematics Trends and Technology*, *65*(7), 85–93. https://doi.org/10.14445/22315373/ijmtt-v65i7p514

Salaudeen, A., Ismail, A., Adeogun, B. K., & Ajibike, M. A. (2021). Validating Gauge-based Spatial Surface Atmospheric Temperature Datasets for Upper Benue River Basin, Nigeria. *Nigerian Journal of Environmental Sciences and Technology*, *5*(1), 173–190. https://doi.org/10.36263/nijest.2021.01.0259

Tao, H., Ewees, A. A., Al-Sulttani, A. O., Beyaztas, U., Hameed, M. M., Salih, S. Q., … Yaseen, Z. M. (2021). Global solar radiation prediction over North Dakota using air temperature: Development of novel hybrid intelligence model. *Energy Reports*, *7*, 136–157. https://doi.org/10.1016/j.egyr.2020.11.033

Tran, T. T. K., Bateni, S. M., Ki, S. J., & Vosoughifar, H. (2021). A review of neural networks for air temperature forecasting. *Water (Switzerland)*, *13*(9), 1–15. https://doi.org/10.3390/w13091294

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs*, *92*(1), 1–27. https://doi.org/10.1002/ecm.1486

Zhou, Y., Zhang, Y., Rong, X., Li, J., & Yu, R. (2019). Forecasting different types of convective weather: A deep learning approach. *J. Meteor. Res.*, *33*(4), 797–809. https://doi.org/10.1007/s13351-019-8206-y

Zhu, S., Zhi, X., Ge, F., Fan, Y., Zhang, L., & Gao, J. (2021). Subseasonal Forecast of Surface Air Temperature Using Superensemble Approaches: Experiments over Northeast Asia for 2018. *Weather and Forecasting*, *36*(1), 39–51. https://doi.org/10.1175/WAF-D-20-0096.1