



Forecasting Performance of Arima and Sarima Models on Monthly Average Temperature of Zaria, Nigeria

Jibril Y Kajuru¹, kamaluddin Abdulkarim², Muhammed M Muhammed³

¹Institute of Education Ahmadu Bello University, Zaria

²Binyaminu Usman Polytechnic, Hadejia

³Baze University Abuja

ABSTRACT

Time series analysis and forecasting has become a major tool in different applications for research. Among the most effective approaches for analyzing time series data is the model introduced by Box and Jenkins. In this study, we used Box-Jenkins Methodology to find an appropriate Model among Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) Model for temperature series of Zaria from 1993-2012. The data was tested for stationarity using ADF and KPSS. The best candidate Model from ARIMA and SARIMA were selected based on information criteria (AIC, AIC_c and BIC) as well as diagnostic checking and forecasting accuracy measure. The Model that emerge best among the competing models is SARIMA (1,0,1)(0,1,1)₁₂. Also the margin between the original series and the in sample forecast shows the reliability of the model estimated and this could be suitable for researchers in the field hydrology and meteorology.

ARTICLE INFO

Article History

Received: March, 2019

Received in revised form: June, 2019

Accepted: August, 2019

Published online: August, 2019

KEYWORDS

Arima, Sarima, Box-Jenkins,
Forecasting Meteorology

INTRODUCTION

The climatic conditions in the Northern part of Nigeria exhibit only two different seasons, namely a short wet season and a prolonged dry season (Emmanuel *et al.*, 2013). Temperatures during the day remain constantly high while humidity is relatively low throughout the year, with little or no cloud cover. There are, however, wide diurnal ranges in temperature (between nights and days) particularly in the very hot months. The mean monthly temperatures during the day exceed 36°C while the mean monthly temperatures at night fall, at most times, to below 22°C. The impact of higher temperatures, more extreme weather events such as floods, severe drought, and sea level rise are already becoming a serious threat in most parts of the world especially in Europe and Asia, and these may continue to manifest in some parts of underdeveloped and developing countries such as Nigeria, unless immense orientation and precaution are taken by individuals and appropriate authorities. Modeling and forecasting Temperature could serve as major tool to provide an insight of climate change in future. Zaria lies between Latitude 11°03' N and 11°15' N; Longitude 7°30' E and 7°45' E in Kaduna State. Urban Zaria consists of six distinct settlements: Zaria City, Tudun Wada, Sabon Gari, Government Reservation Area, Palladan and Samaru.

(Sawa and Buhari, 2011). Zaria is located on the high plains of Northern Nigeria, 652.6 meters above sea level; some 950 km away from the coast with a low temperature as 12°C at night and in the afternoon up to 39°C is sometimes recorded. (Zaria at a glance, 2014). Autoregressive Integrated Moving Average Model (ARIMA) is one of the most widely used time series analysis model in statistics. ARIMA model was firstly proposed by George box and Gwilym Jenkins in the early 1970s, which is often termed as Box-Jenkins model. According to Oduro-Gyimah *et al.*, (2012) Box and Jenkins have played a pioneering role in developing methodologies for univariate time series modeling by adding the possibility for differencing at a single lag and seasonal lag as well as allowing for seasonal components in the ARMA model and SARIMA model can be defined. Many researchers restricted their study using only one or two models, for example, Elias *et al.*, (2012) developed a time series analysis on modeling ARIMA Air temperature of the Southern Caspian Sea Coast. The use of two models like ARIMA and SARIMA will help researchers to select the best model that captured the inherent characteristics in the data series for better Modeling and forecasting. ARIMA is a kind of short-term prediction model in time series analysis because this method is relatively systematic, flexible and can grasp more original time series

*Corresponding author: Jibril, Y. K. ✉ jibrilyk4u@gmail.com ✉ Institute of Education, Ahmadu Bello University, Zaria. © 2019 Faculty of Technology Education, ATBU Bauchi. All rights reserved

information. SARIMA model is an alternative to the ARIMA model that account for seasonal and non-stationary behavior of a time series data. The two fundamental building blocks of a linear univariate time series model are the autoregressive (AR) model and moving average (MA) model (Shekhar, 2004). In an AR model, the forecast is a function of its past observations, while in a MA model the forecast is a function of its past errors. ARIMA model is usually denoted as ARIMA (p, d, q), where p and q are non-negative integers that correspond to the order of the autoregressive and moving average parameters of the model while d is the order of integration respectively. In addition ARIMA (p, d, q) are non-seasonal models and some time series data are periodical in nature. The periodicity of periodical time series data is usually due to seasonal changes (including monthly, quarterly and degree of weeks change) or some other natural reasons or some other natural reasons with a Seasonal ARIMA (P,D,Q) model at different cycle, the parameters P, D and Q are the relevant seasonal autoregressive parameter, seasonal integrated parameter and seasonal moving average parameter. In practical applications, the order of the model SARIMA is usually not too large (Guo, 2009). If the period of time series equals 12, it can be denoted as SARIMA (p, d, q)(P, D, Q)₁₂.

REVIEW OF RELATED MATERIAL

Emmanuel and Bakari (2015), apply time series analysis in modelling monthly Rainfall data for Maiduguri from 1981 to 2011 and come up with ARIMA (1,1,0) as the best model that provide a good fit for the rainfall data of Maiduguri and is appropriate for short term forest. Wiredu *et al.* (2013) proposed some set of SARIMA models to forecast the rainfall pattern in Navrongo Municipality, Ghana, where they came up with ARIMA (0,0,1) (0,1,1)₁₂ as the best model to be adequate for modeling the rainfall pattern after the model was identified and diagnosed to check for its adequacy. Also Mahsin *et al.*, (2012) used Seasonal Autoregressive Integrated Moving Average (SARIMA) model to study the patterns of monthly rainfall data in Dhaka division of Bangladesh. They concluded that the root mean square error values on test data are comparatively less and the predicted model for two years is reliable by comparing the fitted and actual data along with their 95 % confidence interval that can help decision makers to established strategies, priorities and proper use of water resources in Dhaka. In addition, Chisimkwuo *et al.*, (2014) used

time series analysis and forecasting of the monthly maximum temperatures in South Eastern Nigeria that examines 1977-2012 monthly temperature data using the metrological instrument and monthly forecast from 2013-2017 which showcased relatively stable temperature values within these years and the verification of the model using 2011-2012 data shows that the model is parsimoniously equitable which recommends that carefully applying mathematical models could help track future rise in monthly temperatures.

MATERIALS AND METHODS

The data for this research is the monthly average temperatures (°C), obtained from Institute for Agricultural Research Ahmadu Bello University, Zaria Meteorological Station for the period of January, 1993 to December, 2012.

Numerous time series data is non stationary in nature. The Autoregressive Integrated Moving Average (ARIMA) model and Seasonal Autoregressive Integrated Moving Average (SARIMA) Model are extension to the class of ARMA models by adding the possibility to integrate a non-stationary process to a stationary one. The ARIMA and SARIMA model is a mathematical structural mixed model that starts from ARMA process as:

Autoregressive (AR) models

Autoregressive models are based on the idea that the current value of the series, x_t , can be explained as a function of p past values, $x_{t-1}, x_{t-2}, \dots, x_{t-p}$, where p determines the number of steps into the past needed to forecast the current value (Shumway and Stoffer, 2010). An autoregressive model of order p, abbreviated AR (p), can be written as:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t \quad (3.1)$$

By introducing the lag operator (B) in equation (3.1) we obtain:

$$x_t = x_t B^0, x_{t-1} = x_t B^1, x_{t-2} = x_t B^2, \dots, x_{t-p} = x_t B^p. \text{ from the characteristics polynomial of order } p, \text{ The model in lag operators takes the following form;}$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) x_t = e_t, \quad (3.2)$$

The autoregressive operator $\phi(B)$ is defined to be;

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3.3)$$

More concise the AR model can be expressed as

$$\phi(B) x_t = e_t \quad (3.4)$$

*Corresponding author: Jibril, Y. K. ✉ jibrilyk4u@gmail.com ✉ Institute of Education, Ahmadu Bello University, Zaria. © 2019 Faculty of Technology Education, ATBU Bauchi. All rights reserved

Moving Average (MA) Models

The Moving Average model is an alternative to the autoregressive representation in which the x_t on the left-hand side of the equation are assumed to be combined linearly, the moving average model of order q , abbreviated as MA (q), assumes the white noise (e_t) on the right-hand side of the defining equation are combined linearly to form the observed data.

A moving average model of order q , abbreviated as MA (q) can be written as:

$$x_t = e_t + \vartheta_1 e_{t-1} + \vartheta_2 e_{t-2} + \dots + \vartheta_q e_{t-q} \quad (3.5)$$

Where $\vartheta_1, \vartheta_2, \dots, \vartheta_q$ are the Moving Average (MA) parameters in the model and $e_t, e_{t-1}, \dots, e_{t-q}$ are the white noise error terms (Shumway and Stoffer, 2010). $\vartheta_q \neq 0$ and e_t is a Gaussian white noise series with mean zero and variance σ_e^2 unless otherwise stated. The Moving Average model in lag operator B is of the form: $e_t = e_t B^0, e_{t-1} = e_t B^1, e_{t-2} = e_t B^2, \dots, e_{t-q} = e_t B^q$. The model in lag operators takes the following form;

$$x_t = e_t (1 + \vartheta_1 B + \vartheta_2 B^2 + \dots + \vartheta_q B^q) \quad (3.6)$$

The Moving Average operator $\vartheta(B)$ is of the form:

$$\vartheta(B) = 1 + \vartheta_1 B + \vartheta_2 B^2 + \dots + \vartheta_q B^q \quad (3.7)$$

More concise the AR model can be expressed as $x_t = \vartheta(B)e_t$ (3.8)

Autoregressive Moving Average (ARMA) Model

This is a process of mixing Autoregressive and Moving Average model to form an Autoregressive Moving Average (ARMA) models for stationary time series. A general ARMA (p, q) model can be written as follows:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t + \vartheta_1 e_{t-1} + \vartheta_2 e_{t-2} + \dots + \vartheta_q e_{t-q} \quad (3.9)$$

Equation (3.5) of the model can be simplified by a backward shift operator B to obtain.

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) x_t = e_t (1 + \vartheta_1 B + \vartheta_2 B^2 + \dots + \vartheta_q B^q) \quad (3.10)$$

$$\phi(B) x_t = \vartheta(B) e_t \quad (3.11)$$

Autoregressive Integrated Moving Averages (ARIMA) Models

Autoregressive Integrated Moving Average (ARIMA) models are an extension to the class of ARMA models by adding the possibility to integrate a non-stationary process to a stationary. ARIMA (p, d, q) models consist of an autoregressive parameter (p), an order of integration (d), and a moving average parameter (q). A process (x_t) is said to be an autoregressive integrated moving average (ARIMA) process if:

$$\nabla^d (1 - B)^d x_t = \theta(B) e_t \quad (3.12)$$

Where $\nabla^d = (1 - B)^d$ is the non seasonal differencing operator of order d . To produce non seasonal stationarity of the d^{th} difference, usually $d = 1$ or 2 .

Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

This deals with some modifications made to the ARIMA model to account for seasonal and non- seasonal behavior of a time series data with a strong yearly component occurring at lags that are multiples of $S=12$, because of the strong connections of all activities to the calendar year. Data taken quarterly will exhibit the yearly repetitive periods at $S=4$ quarters. Natural phenomena such as temperature, relative humidity and rainfall have strong components corresponding to seasons. Hence, the natural variability of many physical, biological, environmental and economic processes tends to match with seasonal fluctuations. (Shumway and Stoffer, 2010) Because of this, it is appropriate to introduce autoregressive and moving average polynomials that are identify with seasonal lags. A seasonal autoregressive notation (P) and a seasonal moving average notation (Q) will form the multiplicative seasonal autoregressive integrated moving average model, denoted by ARIMA (p, d, q)(P, D, Q) $_s$, of Box and Jenkins (1976) and is given by:

$$\nabla^d (1 - B)^d (1 - B^s)^D x_t = \theta_q(B) \theta_Q(B^s) e_t \quad (3.13)$$

or it can be written as:

$$\nabla^d (1 - B)^d (1 - B^s)^D x_t = \theta_q(B) \theta_Q(B^s) e_t$$

where

x_t is an observation at a time t ; t is the discrete time, s is the seasonal length and

e_t is NID ($0, \sigma_e^2$).

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ are non seasonal autoregressive (AR) operator or polynomial of order p and the $\phi_i, i = 1, 2, \dots, p$ are the non seasonal AR parameters;

$(1 - B)^d = \nabla^d$ is the non seasonal differencing operator of order d . To produce non seasonal stationarity of the d^{th} difference, usually $d = 1$ or 2 ;

$\Phi_p(B^s) = 1 - \Phi_{1s} B^{1s} - \Phi_{2s} B^{2s} - \dots - \Phi_p B^{ps}$ are the seasonal autoregressive (AR) operator or polynomial of order p and $\Phi_{pi} = 1, 2, \dots, p$ are the seasonal AR parameters;

$(1 - B^s)^D = \nabla_s^D$ is the seasonal differencing operator of order D to produce seasonal stationarity of the D^{th} differenced data, usually $D = 1$ or 2 ;



$B(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ are the nonseasonal moving average (MA) operator or polynomial of order q and $\theta_i, i = 1, 2, \dots, q$;
 $\Theta(B^s) = 1 + \theta_1 B^s + \theta_2 B^{2s} + \dots + \theta_q B^{qs}$ are the seasonal MA operator of order Q and $\theta_i = 1, 2, \dots, Q$ are the seasonal MA parameters.

RESULT AND DISCUSSION

Model building commence with examination of the plot of the series, ACF and PACF of the series together with diagnostic and forecasting of the fitted model. Figure 1, 2 and 3 shows a plot of the series, autocorrelation function (ACF) and partial autocorrelation function (PACF) of average monthly temperature data with an indication of non stationarity. The data was tested for stationarity using Augmented Dickey Fuller (ADF) test and Kwiatkowski Phillips Schmidt and Shin (KPSS) test and the result is confirm non stationary as the p-values 0.7843 is greater than alpha (0.05) and the test statistic 0.2812 is greater than critical-value (0.148) at 5% level of significance for ADF and KPSS test respectively as can be seen in table 1. Therefore, there is need for differencing the data to stationary at either lag 1 or 2. Figure 4, 5 & 6 display the differenced average temperature series, ACF and PACF of the differenced series as it affirms that the data is stationary after first differencing, likewise after the first differencing the result confirmed that the data series is stationary as the p-values 0.0031 is less than alpha (0.05) for ADF test and the test statistic 0.0672 is less than the critical-value (0.148) at 5% level of significance for KPSS test respectively as seen in table 2. Thus different competing models from ARIMA and SARIMA were fitted to the differenced data series and the best model was selected based on minimum values of parameter estimates, forecast accuracy as well diagnostic checking of the model. After a careful examination of Table 3 from the different dynamic competing candidate models. The model that emerge best among the competing models was ARIMA (1,0,1) (0,1,1)₁₂ because it has the least values of information criterion and its Ljung-Box diagnostic statistic test reveals that the p-value is greater than alpha (0.05) level of significance for all lags which implies that the white noise hypothesis is not rejected. This means the residuals are uncorrelated indicating that the model is adequately fitted to the series as seen in table 4. It also confirmed the goodness of fit to the selected model in table 5 Based on the measure of accuracy among the

competing models, forecasting performance is well fitted if the forecast error is relatively small. The Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were employed for measure of accuracy among the models. The result shows that SARIMA (1,0,1) (0,1,1)₁₂ Model have least forecasting accuracy measure among the competing models and consequently the series adequacy.

Table 1: Stationary Test on the Average Temperature Series

	Test statistic	p-value	Critical-value
ADF test	-1.62284	0.7843	-
KPSS test	0.28121	-	0.148

Table 2: Stationary Test on the Differenced Average Temperature Series

	Test statistic	p-value	critical-value
ADF test	-6.93072	0.0031	-
KPSS test	0.06724	-	0.148

Dynamic Candidate competing Models Identification

Table 3.1: ARIMA Models

Average Temperature	AIC	AIC _c	BIC
ARIMA (0,1,1)	1047.33	1047.38	1004.28
ARIMA (2,1,1)	963.54	963.71	977.45
ARIMA (2,1,0)	1039.55	1039.66	1049.98
ARIMA (1,1,1)	1049.28	1049.38	1059.71
ARIMA (1,1,0)	1050.46	1050.51	1057.41

Table 3.2: SARIMA Models

Average Temperature	AIC	AIC _c	BIC
SARIMA (1,0,1) (2,1,1) ₁₂	725.58	725.96	746.16
SARIMA (1,0,1) (0,1,1)₁₂	724.74	724.92	738.46
SARIMA (0,0,1) (1,1,1) ₁₂	766.06	766.26	799.80
SARIMA (1,0,1) (1,1,2) ₁₂	727.24	727.62	747.81
SARIMA (1,0,1) (1,1,1) ₁₂	726.23	726.50	743.38
SARIMA (1,0,1) (2,1,2) ₁₂	727.12	727.63	751.12

*Corresponding author: Jibril, Y. K. ✉ jibrilyk4u@gmail.com ✉ Institute of Education, Ahmadu Bello University, Zaria. © 2019 Faculty of Technology Education, ATBU Bauchi. All rights reserved

Table 4: The Ljung-Box test for the fitted model.

	Lag	D.F	Chi-square statistic	P-value
Temperature	12	6	5.1	0.622
	24	18	12.9	0.630
	36	30	17.9	0.890
	48	42	35.3	0.662

Table 5: Accuracy Measure

Average Temperature	RMSE	MAE	MAPE
ARIMA (2,1,1)	1.7730	1.2796	4.9667
SARIMA (1,0,1) (0,1,1)₁₂	1.0911	0.763817	3.06453

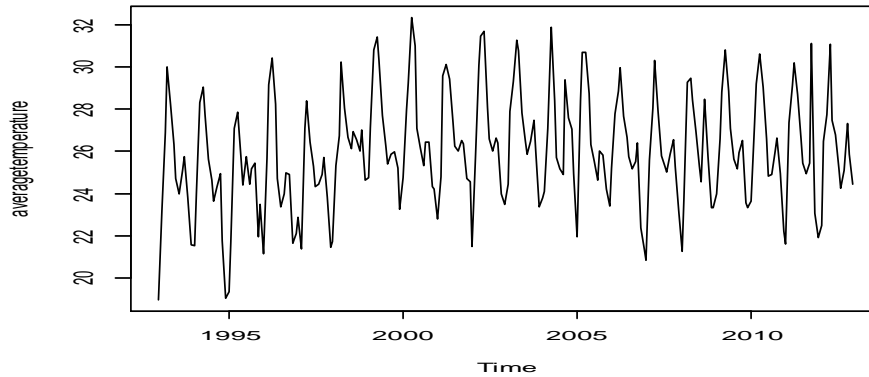


Figure 1: Monthly average temperatures °c in Zaria.

Series averagetemperature

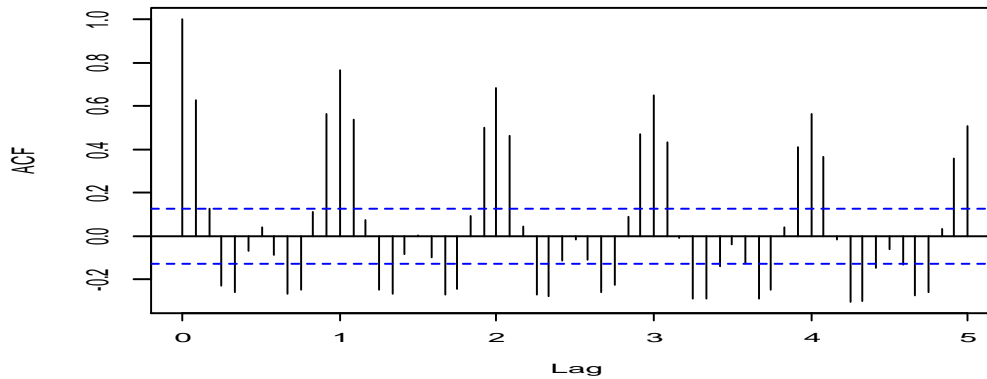


Figure 2: ACF of the monthly average temperatures °c in Zaria.

Series averagetemperature

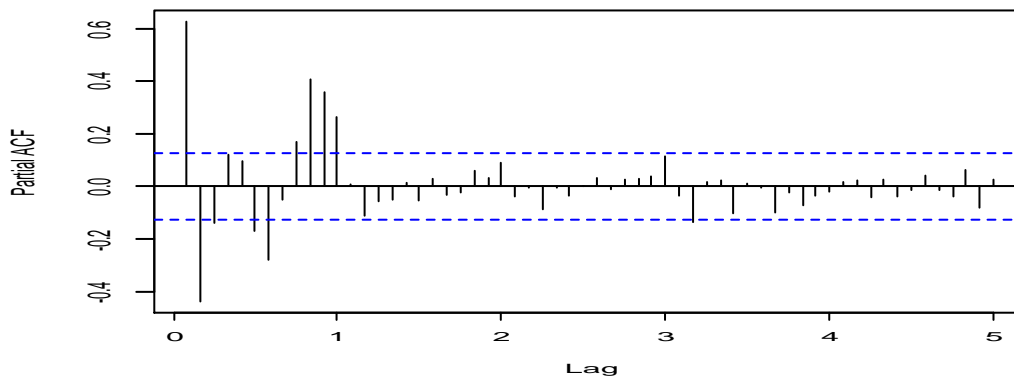


Figure 3: PACF of the monthly average temperature °c in Zaria.

*Corresponding author: Jibril, Y. K. ✉ jibrilyk4u@gmail.com ✉ Institute of Education, Ahmadu Bello University, Zaria. © 2019 Faculty of Technology Education, ATBU Bauchi. All rights reserved

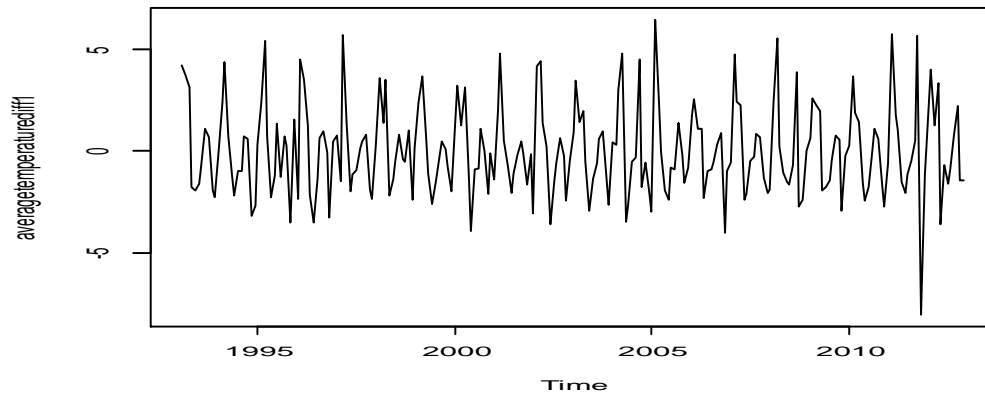


Figure 4: Monthly Differenced average temperatures °c in Zaria.

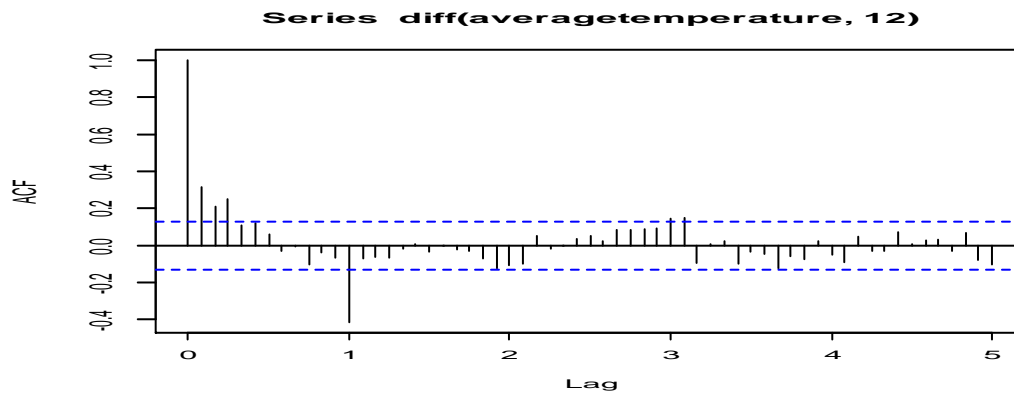


Figure 5: ACF of the differenced average temperatures °c in Zaria.

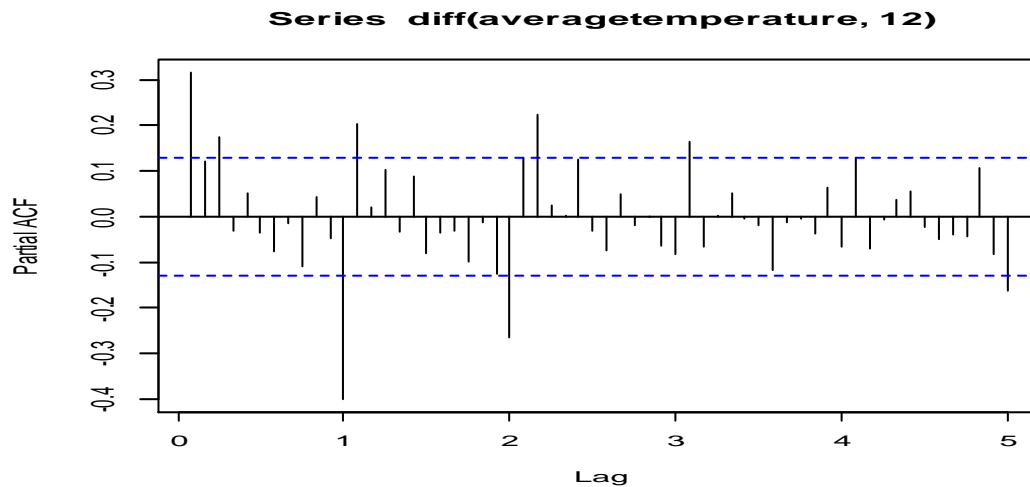


Figure 6: PACF of the differenced average temperatures °c in Zaria.

Table 6: In Sample Forecast for Average Temperature Series

Year	Original Series	In Sample Forecast
2011:01	21.61	22.34
2011:02	27.36	25.54
2011:03	29.13	29.02



Year	Original Series	In Sample Forecast
2011:04	30.19	30.65
2011:05	28.68	28.81
2011:06	26.60	26.57
2011:07	25.45	25.48
2011:08	24.95	25.05
2011:09	25.45	26.27
2011:10	31.10	29.99
2011:11	23.08	25.50
2011:12	21.92	23.44
2012:01	22.48	22.58
2012:02	26.48	25.94
2012:03	27.74	28.96
2012:04	31.08	30.14
2012:05	27.47	28.77
2012:06	26.77	26.21
2012:07	25.18	25.29
2012:08	24.26	24.84
2012:09	25.08	25.85
2012:10	27.30	26.25
2012:11	25.87	23.91
2012:12	24.43	23.26

REFERENCES

- Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis: forecasting and control*. San Francisco, Holden Day.
- Chisimkwuo, J. Uchechukwu, G. and Okezie, S. C. (2014). Time Series Analysis for Forecasting Monthly Maximum Temperature in South Eastern Nigeria. *International Journal of Innovative Research and Development*. 3(1):165-171.
- Dickey, D.A. and Fuller, W.A., (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, 49:1057-1072.
- Elias, K., Jalal, B., Taher, M.N., Alireza, G. and Mohamad, A.D. (2012). Modelling of Air Temperature of the Southern Caspian Sea Coast. *International Research Journal of Applied and Basic Science*. 3(6):1279-1287.
- Emmanuel, I., Adeyemi, B., and Adedayo, K.D. (2013). Regional variation of columnar Refractivity with meteorological variables from climatic monitoring satellite Application facility (CMSAF) data over Nigeria. *International journal of physical sciences*. 8(17):825-834.
- Emmanuel, S.U and Bakari, H. R. (2015). An Application of time series Analysis in modeling monthly Rainfall data for Maiduguri, North eastern Nigeria. *Journal of Mathematical theory and modeling*. 5(11):24-33.
- Guo, Z.W., (2009). The adjustment method and research progress based on the ARIMA model. *Chinese J. Hosp. Stat.*, 161: 65-69.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). "Testing the null of stationarity against the alternative of a unit root: How sure are we that the economic time series have a unit root?" *Journal of Econometrics* 54: 159–178.
- Mahsin, M.D., Akhter, Y. and Begum, M, (2012). Modelling Rainfall Data in Dhaka Division of Bangladesh Using Time Series Analysis. *Journal of Mathematical Modelling and Applications*, 1(5): 67-73.
- Oduro-Gyimah, F. K., Harris, E., and Darkwah, K. F. (2012). SARIMA Time Series Model application to Microwave Transmission of Yeji-Salaga (Ghana) Line-Of-Sight Link. *International Journal of Applied Science and Technology*, 2(9), 40-51.
- Sawa, B. A. and Buhari, B. (2011). Temperature variability & outbreak of meningitis and measles in Zaria, Northern Nigeria, *Research Journal of applied sciences, Engineering and Technology*, 3 (5): 399-402.

*Corresponding author: Jibril, Y. K. ✉ jibrilyk4u@gmail.com ✉ Institute of Education, Ahmadu Bello University, Zaria. © 2019 Faculty of Technology Education, ATBU Bauchi. All rights reserved



- Shekhar, S. (2004). Recursive Methods for Forecasting Short-Term Traffic Flow using Seasonal ARIMA Time Series model. Master's thesis. Graduate Faculty of North Carolina State University. Raleigh. North Carolina. USA.
- Shumway, R.H. and Stoffer, D.S., (2010). *Time series analysis and its applications with R Examples*, Springer, 3rd Edition.
- Wiredu, S., Nasiru, S. and Asamoah, Y.G. (2013). Proposed Seasonal Autoregressive Integrated Moving Average Model for forecasting rainfall pattern in Navrongo municipality, Ghana. *Journal of environment and earth science*, 3(12)80-85.
- Zaria at a Glance. (2014) www.abu.edu.ng/info/zaria.php retrieved March, 2014.