



Analysis of Selected Meteorological Parameters in Kebbi State Using ARIMA Model

*¹Yahaya M. N, ²Adamu I, ³Abubakar M, ¹Samaila B, ¹Abubakar N & ¹Muhammad S.

¹Department of Physics with Electronics, Federal University Birnin Kebbi, Kebbi State.

²Department of Statistics, Federal University of Technology Owerri, Imo State.

³Department of Mathematics, Federal University Birnin Kebbi, Kebbi State

ABSTRACT

This study was carried out in order to analyse the future weather situation of Kebbi state, using meteorological parameters available such as: minimum and maximum temperature, amount of rainfall and relative humidity, for periods spanning from 2009-2018. The data used for this study were extracted from Yelwa-Yauri weather station. Both temporal and spatial variations in variables were witnessed and fluctuations in such variables were also observed. The study area experienced its peak rainfall on the June-July-August (J-J-A), with annual onset mostly in the month of March-April except in 2012 and 2014 where onset was observed in February and January while cessation was recorded generally in October except in 2015 & 2018 where it was witnessed in the month of November respectively. Similarly, using the analysis of variance at 1, 5 and 10% level of significance, the study area from the analysis of individual meteorological parameters in the preceding years (2009-2018), did not yield any definite meteorological patterns that can account for the varying levels of predictions.

Keywords: ARIMA, Cessations, Meteorology, Onset & Time Series.

INTRODUCTION

Two decades ago, it has become an undeniable fact that the global averages maximum and minimum temperature were increasing at a disturbing proportion with corresponding declined/reductions in amount of rainfall received annually both at regional and in global perspectives. This has resulted because of human and natural activities. Meteorology is concerned with describing the present state of the atmosphere (temperature, pressure, winds speed, humidity, precipitation/ rainfall and cloud cover) and in predicting the evolution of these primary variables over time interval for a few days (Yahaya *et al.*, 2012). Squire with global warming and accelerated urbanization many cities across the world are enduring a great extensions and structural changes, causing in distinctive phenomenon that urban temperatures are significantly higher than rural temperatures (He, 2018).

Urban heat intensity (UHI) has been proved as a phenomenon a series of negative impact in human and thermal comfort, air quality, wellbeing and energy use. Thus, many researchers across various disciplines are vigorously in search of effective UHI mitigations techniques and strategies. Synoptic condition can help or hinder the development of UHI (Tamas *et al.*, 2013). An essential tool needed for this description is the thermodynamics, which helps relate the fundamental quantities of pressure, temperature and density as atmospheric parcel that changes from place to place (Gerald *et al.*, 2009). It has been emphasized that no two climates can be exactly the same however similar or close the two locations may be geographically; hence the needs to study the climatic variations of a place are necessary as a single entity. It is the volatile nature of meteorological parameter that accounts for the variability of weather in time and space. It is of interest to notes that many meteorological parameters correlate well with each

other from which empirical relationships can be developed (Akpabio *et al.*, 2004). Hence, a good knowledge of one is important in estimating the less easily measurable parameters. This is of fundamental importance due to the facts that some of these variables like global solar radiations data are difficult to maintain for a long period (Gopinathan, 1988).

Due to changes in climate currently experienced regionally and globally, this research was undertaken to determine the impact of fluctuations in selected meteorological parameters in the study area.

MATERIALS AND METHODS

Temperature and Rainfall Variations in the Study Area

The study area falls within the hottest parts of Nigeria a critical zone, located above latitude 10°N, belonging to the Sahel region of Africa, an area that is mostly affected by the droughts, flood & desertification (Figure 1). Temperatures are generally extreme, with average daily minimum of 16°C during cold months of December and January, and in the hottest months of April to June, an average maximum of 38°C and minimum of 24°C. Year-in-year out data shows that, the average maximum and minimum temperature is 36°C and 21°C. Rainfall is generally low in this area. Much of the rain falls between the months of May to September, while the rainless months are October to April. Evaporation is high ranging from 80 mm in July to about 210 mm in April to May. This shows that the hottest months in the study area ranges from April to May with corresponding highest evaporation. Relative humidity is low most of the year and only increases during the wet season that is periods of May-September or June to October depending on the onset and cessations of the year respectively (Akpabio *et al.*, 2004).

*Corresponding author. E-mail: mnaziru.yahaya@fubk.edu.ng.

Author(s) agree that this article remain permanently open access under the terms of the Creative Commons Attribution License 4.0 International license

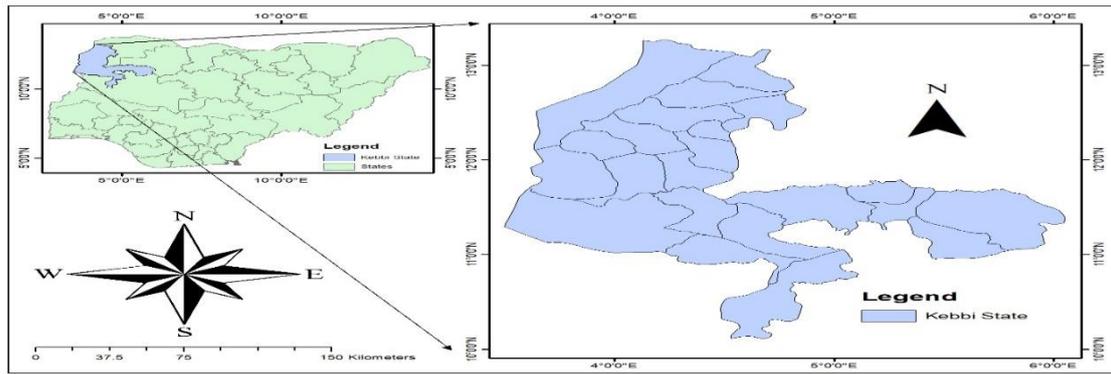


Figure 1. Map of Nigeria Showing the study area.

Two dominant air masses can be found in the study area which controlled largely the social, economic and Agricultural activities of the entire region. Like the rest of West Africa, the climate of the region is controlled largely by the two dominant air masses affecting the sub-region (Figure 1). These are the dry, dusty, tropical-continental (cT) air mass (which originates from the Sahara region), and the warm, tropical- maritime (mT) air mass (which originates from the Atlantic Ocean). The influence of both air masses on the region is determined largely by the movement of the Inter-Tropical Convergence Zone (ITCZ), a zone representing the surface demarcation between the two air masses. The interplay of these two air masses gives rise to two distinct seasons within the sub-region. The wet season is associated with the tropical maritime air mass, while the dry season is a product of the tropical continental air mass. The influence and intensity of the wet season decreases from the West African coast northwards (Ekpoh, 2011). The natural vegetation of the state comprises Sudan savannah in the north and northern Guinea savannah in the south and southeast. Two major rivers drain the state: River Niger in the generally rocky south traversing the state from Benin Republic up to Ngaski Local Government Area, and River Rima in the sandy north passing through Argungu to Bagudo Local Government Areas where it empties in to River Niger.

Meteorological data are set of information, which describes the characteristics of the atmosphere. When these data are available for an hourly, daily, monthly, annually decade or seasonal, they are taken to describe weather characteristics. However, when they are available over a long period of time, help in predicting the future evolution or changes in their characteristics. The parameters of the environment commonly measured include temperature, rainfall and relative humidity. The data on rainfall, Relative humidity and maximum temperature, used for the study where extracted from the meteorological records of Yelwa-Yauri weather station for a periods of Nine Years spanning from 2009-2018. The Box-Jenkins methodology was used in modelling the time series. The pioneers who popularized an approach which combines the moving average and autoregressive models were Box and Jenkins. ARIMA model has gained popularity among researchers who consider it to be superior in providing short-term forecast (Adebisi *et al.*, 2014; Merh, *et al.*, 2010 and Nochai *et al.*, 2006). Therefore, ARIMA model was chosen to be the main forecasting instrument in this research. ARIMA process

first transforms non-stationary time series by differencing it a number of times, which is then modelled with an ARMA (p, q) process. As described above, non-stationary time series can often be converted to a stationary one by simple differencing. Differencing a non-stationary time series, a number of time into a stationary one is a form of de-trending the series i.e. by removing the trend component. Knowing that every stationary time series is integrated of order d (I(d)) and if by differencing the terms an ARMA(p, q) process, then the un-differenced process can have an ARIMA(p, d, q) representation in this case, the time series {X_t} can be expressed as

$$\phi(L)\Delta^d X_t = \theta(L)\epsilon_t \dots\dots\dots 1.1$$

Box and Jenkins created a three-stage method for the model selection. These stages are: identification, estimation, diagnostic checking. (Box-Jenkins and Reinsel, 2013)

Step 1: Model identification

At this stage the time series is being checked for stationarity by examining the correlograms of autocorrelation function (ACF) and partial autocorrelation function (PACF). (Lee *et al.*, 2011). Stationarity means that the mean, variance and covariance of a time series are all constant over time. However, most economic time series have trends, thus having different means over time (Asteriou and Hall, 2011). The series is stationary if its values on the ACF graph either cut off quickly or die down quickly. If ACF graph dies down very slowly, then the time series is considered to be non-stationary (Nochai *et al.*, 2006). If the time series data appears to be non-stationary, it should be transformed into a stationary one by using the appropriate number of differencing. Differencing is applied as many times as it is needed to meet stationary requirement. (Lee *et al.*, 2011) Differencing serves the role of de-trending a time series data (Asteriou and Hall, 2011). The first differences of the series are as follows:

$$\Delta Y_t = Y_t - Y_{t-1} \dots\dots\dots 1.2$$

Where, Y_t is a value of a single observation from the stationary time series at the time t. If after first differencing a time series is non-stationary, second differencing should be applied by using the formula below:

$$\Delta \Delta Y_t = \Delta Y_t - \Delta Y_{t-1} \dots\dots\dots 1.3$$

(Asteriou and Hall, 2011) Correlograms are the values of the ACFs and PACFs plotted against lag lengths and are used to check a series for stationarity. The autocorrelation coefficient (ACF) estimates the correlation between a set of observations and a lagged set of observations in a time series (Wasseja et al, 2015). The sample autocorrelation coefficient of the series is given below:

$$r_k = \frac{\sum(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum(Y_t - \bar{Y})(Y_t - \bar{Y})} \dots\dots\dots 1.4$$

Where, Y_t is a value of a single observation from the stationary time series. Y_{t+k} is the data from the period $t+k$. \bar{Y} is the mean of the stationary time series. Partial autocorrelation function measures how Y_t and Y_{t+k} are related and used together with ACF as a guide in selecting an appropriate ARIMA model. (Wasseja and Mwenda, 2015), once series becomes stationary, the next task is to determine the p and q orders of the model. MA(q) process is used in order to find the value of q. According to it, the value of q is equal to the last lag in the ACF, where estimates are statistically significant. After that point estimates begin to die down immediately. The PACF for MA(q) process tends to die down quickly. In AR(p) process the value of p is defined as the last estimate that shows spikes in the PACF. In the pure AR(q) process an ACF dies down quickly. If neither the ACF nor the PACF functions indicate the number of orders, a combined process is applied (Asteriou and Hall, 2011).

Step 2: Parameter estimation

After selecting tentative models, the models` parameters should be estimated using the least squares method. The parameters are calculated in a way to have zero gradient of forecasting errors to the historical data. At this stage

the prime task is to minimize the error from forecasting and define model`s parameters and order. After estimating tentative models, their coefficients are compared. The statistical measures that are applied to select the best fitted model are the Akaike Information Criteria (AIC) and the Schwarz Bayesian criterion (SBC). The model with the lowest AIC and/or SBC should be chosen. In addition, the highest R^2 is also an indicator of the appropriate model.

Step 3: Diagnostic checking

After parameters` estimation, the tentative model`s accuracy is checked by studying the ACF and PACF residuals. The residuals should follow the white noise process. Superfluous coefficients should not be added to the appropriate model, otherwise the model is over fitted. Afterwards, the Q-statistics is used to approve the tentative model. If the estimated value Q exceeds the critical value of χ^2 derived from the chi-square table, the tentative model is inadequate. (Lee et al., 2011). This statistic tests the model for autocorrelation of the residuals (Asteriou et al., 2011). If a tentative model is inadequate, the whole process should be repeated until an adequate one is identified. When the Box-Jenkins procedure is completed, the selected ARIMA model is used to forecast the future values usually over 24 hours ahead.

RESULT AND DISCUSSION

The results obtained from the analysis of relative humidity, rainfall, temperature (maximum and minimum) using univariate ARIMA (p,d,q) models were explain in Table 1.

Table 1: Descriptive Statistics

Statistical Tools	R/Humidity*	Rainfall**	T/maximum***	T/Minimum***
Mean	58.167	91.347	35.125	22.584
Median	59.500	38.950	35.450	23.400
Std. Dev	19.138	114.878	3.161	3.790
Skewness	10.368	1.295	0.106	-0.588
Kurtosis	2.201	4.042	1.956	2.730

*Relative humidity, **Rainfall, ***Temperature maximum/minimum

Time Series Plot of the Data

This present study investigated the mean monthly variation of relative humidity, rainfall and maximum/minimum temperature. Table 1 shows the statistical summary which indicate the positive mean for all the four different variables with an average of the rate for the period under study, while the positive standard deviation shows the dispersion from the mean and high level of variability of the series, all the results also show positive asymmetry and their distribution are approximately normal. The data plots of the monthly series were displayed as *ordinates* against equally spaced

time intervals as *abscissa*, (Figs. 2a and b) is used to evaluate the pattern and behaviour in the data over time. A visual inspection of the plot shows a fluctuating trend, which may be indicating non-stationarity in the data series. It can be observed that, out of all the four parameters understudy denotes the rejection of null hypothesis at **1, 5 & 10%** level of significance (Tables 2a-d) which is clear indication that the series is stationary after first difference. Similarly, the result from the analysis indicates that all the variables show a model with minimum information criteria (Tables 3a-d).

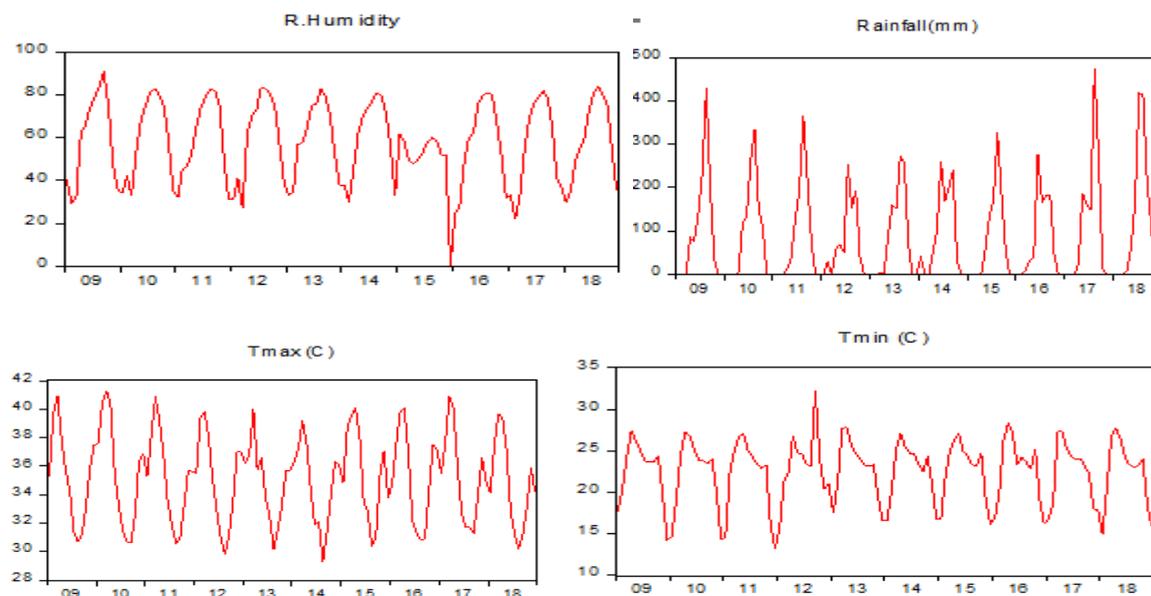


Figure 2a. Trend pattern in Relative humidity, Rainfall, Max and Min Temperature (2009-2018).

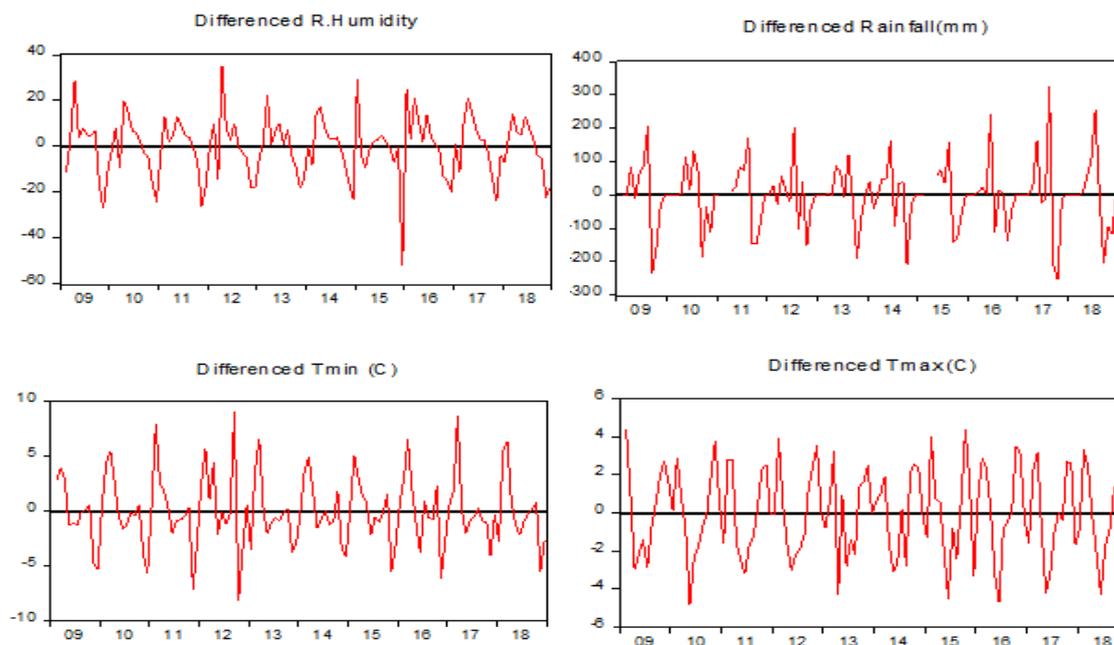


Figure 2b. Difference in Relative humidity, Rainfall, Max and Min Temperature Pattern in the study area (2009-2018).

Table 2a: Results of the KPSS Test for Relative Humidity

R/Humidity	Number of lag	1%	5%	10%	Test statistics
	4	0.313	0.263	0.422	0.0122009***
	5	0.739	0.431	0.247	0.0002001***

Note: *** Denote rejection of null hypothesis at 1%, 5% & 10%.

Table 2b: Results of the KPSS Test for Rainfall

Rainfall	Number of lag	1%	5%	10%	Test statistics
	4	0.739	0.463	0.347	0.0328952***
	5	0.739	0.463	0.347	0.0287021***

Table 2c: Results of the KPSS Test Maximum Temperature

Temperature(max)	Number of lag	1%	5%	10%	Test statistics
	4	0.528	0.3332	0.347	0.0143280***
	5	0.320	0.2300	0.300	0.0087020***

Table 2d: Results of the KPSS Test for Minimum Temperature

Temperature(min)	Number of lag	1%	5%	10%	Test statistics
	4	0.232	0.330	0.222	0.0120000***
	5	0.563	0.360	0.240	0.0027021***

1.4 Model Identification

Table 3a: Results of ARIMA model identification for Relative Humidity

	Models	AIC	Log likelihood
R/Humidity	ARIMA (1, 1, 0)	760.66	-428.22
	ARIMA (1,1,1)*	679.86*	-316.83*
	ARIMA(1,1,2)	716.57	-488.38

Note: * Denote the model with Minimum Information Criteria.

Table 3b: Results of ARIMA model identification for Rainfall

	Models	AIC	Log likelihood
Rainfall	ARIMA (1, 1, 0)	960.56	-478.28
	ARIMA (1, 1, 1)	879.86	-436.93
	ARIMA (1,1,2)*	716.57***	-204.28***

Note: *** Denote the model with minimum information criteria

Table 3c: Results of ARIMA model identification for average temperature

	Models	AIC	Log likelihood
Temperature(max)	ARIMA (1, 1,0)	560.56	-472.28
	ARIMA (1, 1,1)	579.82	-432.93
	ARIMA(1,1,2)*	516.52*	-404.28*

Note: * Denote the model with Minimum Information Criteria.

Table 3d: Results of ARIMA model identification for average temperature.

	Models	AIC	Log likelihood
Temperature(min)	ARIMA (1, 1,0)*	970.56*	-573.28*
	ARIMA (1, 1,1)	979.86	-576.98
	ARIMA(1,1, 2)	980.5	-580.22

Note: *** Denote the model with minimum information criteria.

The ARIMA Model

The ARIMA (1, 1, 1) Model for the Relative Humidity series is:

$$Y_t = -2070Y_{t-1} - 2.9724\varepsilon_{t-1} \dots \dots \dots 1.5$$

The ARIMA (1, 1, 2) Model for the rainfall series is:

$$Y_t = 0.1236Y_{t-1} - 0.1071\varepsilon_{t-1} + 0.3210\varepsilon_{t-2} \dots \dots \dots 1.6$$

The ARIMA (1, 1, 2) Model for the temperature maximum series is:

$$Y_t = -0.0072Y_{t-1} - 1.3722\varepsilon_{t-1} - 0.2073\varepsilon_{t-2} \dots \dots \dots 1.7$$

The ARIMA (1, 1, 0) Model for the temperature minimum series is:

$$Y_t = -0.1071Y_{t-1} - 2.9724\varepsilon_{t-1} - 2.9641\varepsilon_{t-2} \dots \dots \dots 1.8$$

CONCLUSION

Out of all the parameters maximum temperature and rainfall are the most often reported in Nigeria as in the other parts of the world. This is because of its well-established influence over on most other parameters recorded and studied. The conclusions drawn from this study can be expressed in terms of the impact climate, climate change scenarios which indicate that the climatic variability currently being experienced globally particularly in less developed country is likely to increase in intensity in future. Droughts, floods and storms are likely to increase in both frequency and intensity. Regular and effective weather condition and monitoring needs to be adopted for a better assessment of such variables with the aim of mitigating and averting the possible consequences of changes/fluctuations in such parameters. However, ARIMA (p, d, q) pioneered by Box and Jenkins 1976 were used to select the best fitted model for the metrological parameters, and finally model with minimum Akaike Information Criteria (AIC) were best to described the data. Similarly, statistical results obtained across all the variables at **0.01**, **0.05** and **0.1** percentages level of significance with positive mean and

standard deviation shows a series variation with its distribution approximately normal.

REFERENCES

- Adebiyi, A. and Adewumi, O. A. (2014). Stock Price Prediction Using the ARIMA Model. *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, At Cambridge University, United Kingdom.*
- Akpabio, L. E, Udo, S. O and Etuk S. P. (2004): Empirical correlation of Global Solar Radiation with Meteorological parameters for Onne, Nigeria. *Turkish Journal of Physics*, 23(3), 205-212.
- Asteriou, D. and Hall, S. (2011). *Applied econometrics*. Basingstoke [England]: Palgrave Macmillan, pp: 266-279.
- Box, G., Jenkins, G. and Reinsel, G. (2013). *Time Series Analysis: Forecasting and Control*. 4th ed. United States: John Wiley & Sons Inc.
- Box, G. E. P and Jenkins, G. M. (1976): *Time Series Analysis: Forecasting and Control*. 2nd ed. Hodden day, San Francisco).
- Ekpoh, I. J. (2011) Extreme Climatic Variability in North-western Nigeria: An Analysis of Rainfall Trends and Patterns. Pp: 1-4
- Gerald, R. (2009): Center for Meteorological study, Texas U.S.A. Pp:1
- Gopinathan, K. K. (1988): A General Formula for computing the coefficient of Correlation connecting Global Solar Radiation to sunshine duration. *Solar Energy*, 41(6), 499.
- He B. J (2018). Potentials of meteorological characteristics and synoptic conditions mitigate urban heat island effects. *Elsevier, Urban climate* 24(2018), 26 – 33.
- Lee, C. and Ko, C. (2011). Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications*, 38(5):.5902-5911
- Merh, N., Saxena, P. V. and Raj P. K. (2010). A comparison between Hybrid Approaches of ANN and ARIMA for Indian Stock Trend Forecasting. *Business Intelligence Journal*, [online] 3(2): 23-44
- Nochai, R. and Nochai, T. (2006). ARIMA MODEL FOR FORECASTING OIL PALM PRICE. *Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications, Universiti Sains Malaysia*
- Tamas T, Elemer L and Lazar I (2013). Examination on the role of synoptic conditions in urban heat island development in Debrecen. *Denzero International Conference*. DOI: 10.13140/2.1.2417.3442.
- Wasseja, M. M and Mwenda, S. N. (2015). Analysis of the Volatility of the Electricity Price in Kenya Using Autoregressive Integrated Moving Average Model. *Science Journal of Applied Mathematics and Statistics*, [online] 3(2). Pp: 47-57
- Yahaya, M. N. (2012). Analysis of Mean Monthly Variation of Rainfall, Pressure and Maximum Temperature at Sokoto North-western Nigeria. Unpublished B.Sc. Project Submitted to the Department of Physics Kebbi State University of Science and Technology, Aliero. Pp:1-31