Stochastic characteristics and modeling of minimum and maximum temperature of Ogun State, Nigeria

Musa, J. J.,[1] Adeoye, P. A.[1] and Jibril, I
[1]Department of Agricultural and Bioresources Engineering, Federal University of Technology, P M B 65, Minna, Nigeria. Tel: +234-8094-747-246
[2]College of Agriculture, P M B 109, Mokwa, Nigeria. Tel: +234-8064-485-066

ABSTRACT

Previous 20 years data (1982 to 2009) have been collected in order to predict the future temperature pattern of Ogun State. The data were preprocessed and aggregated into annual time series to fit for stochastic characterization and modeling of minimum and maximum. Mann-Kendal non-parametric test, Lo's long-range dependency test and spectral analysis were done to detect whether there is trend and seasonal component in the time series The best autoregressive AR-model, moving average MA-model and autoregressive moving average ARMA-models were fitted for all parameters considered, with the aid of Akaike Information Criterion (AIC), and error terms of FE, MAE, MSE and MAPE. AR, MA and ARMA models of order (2), (3) and (1, 2) and (5), (3) and (5, 3) were found to be the best for predicting maximum and minimum temperatures respectively.ACF, PACF and the Box-Jenkins technique were utilized for model type and order selection. The overall results were promising and the prediction scheme applied in this research could be considered in situations where database is a problem during model development. It is therefore recommended that another research be carry out in the area using another method of modeling to compare the results.

Keywords: Stochastic model, stationarity, non-parametric test, minimum and maximum temperature.
DOI: http://dx.doi.org/10.4314/ejst.v9i1.4

INTRODUCTION

Temperature is one of the major input variables for land evaluation and characterization systems, as well as hydrological and ecological models. These models use air temperature to drive processes such as evapotranspiration, soil decomposition, and plant productivity (Benavides *et al*., 2007). Air temperature is an important site characteristic used in determining site suitability for agricultural and forest crops (Benavides *et al.,* 2007), and it is used in characterizing the habitat of plant species (Rubio *et al*., 2002; Sanchez-Palomares *et al*., 2003) and in determining the patterns of vegetation zonation (Richardson *et al.,* 2004). Modeling temperature therefore, is an important task for efficient agricultural development and sustainability.

Models are simplifications of reality that reflect our understanding of the process they represent. Just as any other tool, the results given by models are dependent on how they are applied, and the quality of these answers is not better than the quality of our understanding of the system (Robin, 2003). Some models are based solely on empirical equations while others are built on more complex, physically based principles (Butcher *et al*., 1998). To gain an insight into the nature of climatic variability within the climate system, it is necessary to study its components in a systematic way

The basic assumption made to implement this model is that the considered time series is linear and follows a particular known statistical distribution, such as the normal distribution. ARIMA model has subclasses of other models, such as the Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) models (Box *et al*., 2008). For seasonal

*Corresponding author: johnmusa@futminna.edu.ng

time series forecasting, Hamzacebi (2008) stressed that Box and Jenkins had proposed a quite successful variation of ARIMA model, viz. the Seasonal ARIMA (SARIMA). The popularity of the ARIMA model is mainly due to its flexibility to represent several varieties of time series with simplicity as well as the associated Box-Jenkins methodology (Zhang, 2003) for optimal model building process. But the severe limitation of these models is the pre-assumed linear form of the associated time series which becomes inadequate in many practical situations. The selection of a proper model is extremely important as it reflects the underlying structure of the series and this fitted model in turn is used for future forecasting. A time series model is said to be linear or non-linear depending on whether the current value of the series is a linear or non-linear function of past observations.

This research paper is intended to check for trend in the time series data, develop and validate a stochastic model for prediction of future temperature of the study area.

MATERIALS

The entire study area is bounded by Oyo state to the north, Osun and Ondo States to the east and Lagos State to the South as shown in (Figure 1). It is located in southern Nigeria, bordered geographically by latitudes 6.26°N and 9.10°N and longitudes 2.28°E and 4.8°E. The land area is about 23,000km^2. The relief is generally low, with the gradient in the North-South direction.

The two major vegetation zones that can be identified the area are the high forest vegetation in the north and central parts, and the swamp/ mangrove forests that cover the southern coastal and floodplains, next to the lagoon. It has two distinct seasons throughout the year. The monthly rainfall distribution in the study area shows a distinct dry season extending from November through March and a rainy season divided into two periods: April – July and September – October. The mean annual rainfall data for 30 years showed a variation from about 1,150mm in the northern part to around 2,285mm in the southern extremity. The estimates of total annual potential evapotranspiration have been put between 1600 and 1900mm (Ewomoje and Ewomooje, 2011).

DATA COLLECTION AND PREPROCESSING

The minimum and maximum temperature data used for this study were obtained from the federal ministry of water resources, Abeokuta, Nigeria. The data collected covered a period of twenty nine years (1982-2009). These values were obtained by the use of GPS (Global Position System) equipment. Data preprocessing is an important task in almost all modeling techniques. The data obtained are the time series types which are collected monthly for a period of 29 years. For the purpose of this study, the mean annual values of the data were first determined before use.

Figure 1. Map of Nigeria showing the study Area.

Test for Trend, Long-range Dependency and Serial Correlation

Time series data are generally represented in the form:

$$T(t) = Tr + P(t) + \varepsilon(t) \qquad 1$$

Where *T(t)* is the time series, *Tr* is the trend component, *P(t)* is the periodic component and *ε(t)* is the stochastic component. In order to check for the stationarity of the data, the following equations were considered:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn(X_j - X_k) \qquad 2$$

Where, X_j and X_k are the annual values in years j and k, j > k, respectively, and

$$sgn(X_j - X_k) = \begin{cases} 1 & if\ X_j - X_K > 0 \\ 0 & if\ X_j - X_k = 0 \\ -1 & if\ X_j - X_k < 0 \end{cases} \qquad 3$$

Source: Ghanbarpour *et al.* (2007); Besaw *et al*. (2010) and Tesfaye *et al*. (2006)

$$VAR(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=1}^{q} t_p(t_p - 1)(2t_p + 5)\right] \qquad 4$$

Where, q is the number of tied groups and t_p is the number of data values in the p^{th} group. The values of S and VAR(S) are used to compute the test statistic Z as follows (Longobardi and Villan, 2009; Nail and Momani (2009).

$$Z = \begin{cases} (S-1)/Var(S)^{1/2} & S > 0 \\ 0 & S = 0 \\ (S+1)/Var(S)^{1/2}5 & S < 0 \end{cases} Z = \begin{cases} (S-1)/Var(S)^{1/2} & S > 0 \\ 0 & S = 0 \\ (S+1)/Var(S)^{1/2} & S < 0 \end{cases} \qquad 5$$

The Mann-Kendall test was carried out in accordance with the works of Otache *et al*. (2011); Edwin and Otache (2014) and Chatfield (2004), with the aid of the excel template of 'MAKESEN's version 1. Lo's modified R/S test was also done to ascertain if the trend persisted. To check for serial correlation, the Durbin-Watson test was considered. The tests were carried out in order to make sure the time series data conforms to the basic criteria for stochastic modeling. No trend was observed as indicated by the Mann-Kendal Z-values of -2.03 and -0.47 for maximum and minimum temperatures respectively. The correlograms (Figures 4 and 5) also does not show a clear seasonal nature, therefore, only the stochastic component was considered.

RESULTS AND DISCUSSIONS

ARMA ModelBuilding for Maximum and Minmum Temperature

The time series of maximum and minimum temperature were treated trend free and non-seasonal as indicated by the result of Mann-Kendal test Correlogram. Based on the fact that the ACF and PACF diagrams are sometimes difficult to interpret (Kumar and Vanajakshi, 2015), the iterative techniques was utilized.With the aid of MINITAB Software Version 16.0, models of orders AR(2), MA(3) and ARMA(1, 2) for maximum temperature; and AR(5), MA(3) and ARMA(5, 3) were considered for validation of the data. This is also confirmed by the Akaike Information Criterion (AIC) test in accordance with the works of Kumar and Vanajaksh (2015). The high-lighted (AIC) values in the tables of model order selection (i.e. tables of model parameters) as in tables 1 and 2 are the leasts in magnitude when compared with others, this make them the

Table 1. Model order selection

	Model Order (P)	Sum of Squares (SS)	AIC - Value	Constant (c)	Mean (μ)
A. for AR	1	410.637	77.19	21.3224	38.23
	2	287.573	69.23	6.2514	37.35
	3	284.973	70.97	7.5523	37.593
	4	284.694	72.94	7.9863	37.672
	5	278.559	74.33	5.9148	37.356
B. for MA	1	457.521	82	-	38.2246
	2	321.146	73.73	-	38.36
	3	268.5	70.54	-	38.607
	4	260.417	71.66	-	38.7355
	5	253.405	72.86	-	38.6189
C. for ARMA	1, 2	272.601	168.63	22.855	38.771
	2, 1	328.996	174.09	13.14	38.313
	2, 2	265.103	169.82	26.566	38.605
	1, 3	262.706	169.56	30.673	38.711
	3, 1	285.534	171.98	15.244	37.84
	2, 3	244.702	169.5	4.861	39.015
D. for AR	1	767.708	97.01	8.91	18.82
	2	761.51	98.77	8	18.94
	3	738.623	99.89	6.33	19.43
	4	607.516	96.22	9.66	18.61
	5	542.101	94.92	16.1	17.74
E. for MA	1	805.06	98.39	-	18.64
	2	800.96	100.24	-	18.64
	3	536.46	90.61	-	18.28
	4	579.91	94.87	-	18.88
	5	624.45	99.02	-	17.78

best, and were used for the model building as shown in Table 3. The model equations were used to generate forecast for each parameter, and the actual values were ploted with the predicted for comparism in accordance with the works of Tizro *et al*. (2014). Using the Lewi's error scaling system (i.e. considering the MAPE), the prediction was found highly accurate for maximum temperature in all models. While for the minimum temperature, only validation for ARMA model indicated high accuracy, while AR and MA were found to be good. The result of the comparism is as shown in Table 4. In addition, actual and predicted values were ploted to graphically compare the perfection of the 3 stochastic models used. Figures 2 and 3 shows the pattern of actual and predicted values.

Table 2. Model order selection for ARMA

S/No	Model Order (p,q)		Sum of Squares (SS)	AIC - Value	Constant (c)	Mean(μ)
1	1	1	759.51	196.35	6.19	19
2	1	2	793.23	199.6	2.3	18.35
3	2	1	759.47	198.35	6.38	19
4	2	2	597.69	193.4	8.04	20.67
5	1	3	786.76	201.37	2.78	18.26
6	3	1	702.57	198.09	8.96	19.52
7	3	2	489.81	189.63	9.26	20.83
8	2	3	568.44	193.94	2.63	17.42
9	3	3	482.63	191.2	11.23	20.02
10	1	4	541.46	192.53	3.39	18.41
11	4	1	589.67	195.01	11.43	19.89
12	2	4	404.85	186.1	3.48	17.5
13	4	2	479.94	191.04	12.03	19.63
14	3	4	395.56	187.43	4.21	17.45
15	4	3	477.6	192.9	15.16	20.6
16	4	4	386.28	188.74	6.05	17.63
17	1	5	551.62	195.07	3.29	18.49
18	5	1	535.99	194.24	11.37	17.49
19	2	5	427.5	189.68	3.38	17.54
20	5	2	527.53	195.78	13.48	17.51
21	3	5	428.39	191.74	6.15	17.45
22	5	3	333.48	184.48	14.85	17.4

Table 3. Best Model Equations obtained for the 3 Models

S/No	Model Type	Model Order	Model Equation
A. on Maximum Temperature			
1	AR	2	$y_t = 6.25 + 0.1952y_{t-1} + 0.6374y_{t-2} + \varepsilon_t$
2	MA	3	$y_t = 38.607 - 0.2821\varepsilon_{t-1} - 0.7901\varepsilon_{t-2} - 0.3592\varepsilon_{t-3} + \varepsilon_t$
3	ARMA	1, 2	$y_t = 22.855 + 0.4105y_{t-1} + 0.1493\varepsilon_{t-1} - 0.8819\varepsilon_{t-2} + \varepsilon_t$
B. on Minimum Temperature			
1	AR	5	$y_t = 16.1027 + 0.3346y_{t-1} + 0.1563y_{t-2} + 0.2463y_{t-3} + 0.1102y_{t-4} - 0.5349y_{t-5} + \varepsilon_t$
2	MA	3	$y_t = 18.279 - 0.5713\varepsilon_{t-1} - 0.808\varepsilon_{t-2} - 0.5303\varepsilon_{t-3} + \varepsilon_t$
3	ARMA	5, 3	$y_t = 14.8530 + 0.3050y_{t-1} + 0.0535y_{t-2} + 0.6600y_{t-3} - 0.2895y_{t-4} - 0.5825y_{t-5} - 0.0551\varepsilon_{t-1} - 0.4487\varepsilon_{t-2} + 0.8142\varepsilon_{t-3} + \varepsilon_t$

Table 4. Summary of Error Values for Minimum Temperature

Model Type	Model Order	Forecast Error	MSE	RMSE	MAPE (%)
A. for Minimum Temperature					
AR	5	1.42	18.69	18.13	20.82
MA	3	2.08	18.49	18.46	21.85
ARMA	5, 3	0.26	14.93	14.93	-2.89
B. Maximum Temperature					
AR	2	-1.49	14.22	14.69	-1.37
MA	3	-6.14	9.26	12.58	6.67
ARMA	1, 2	-6.2	9.4	13.56	7.02

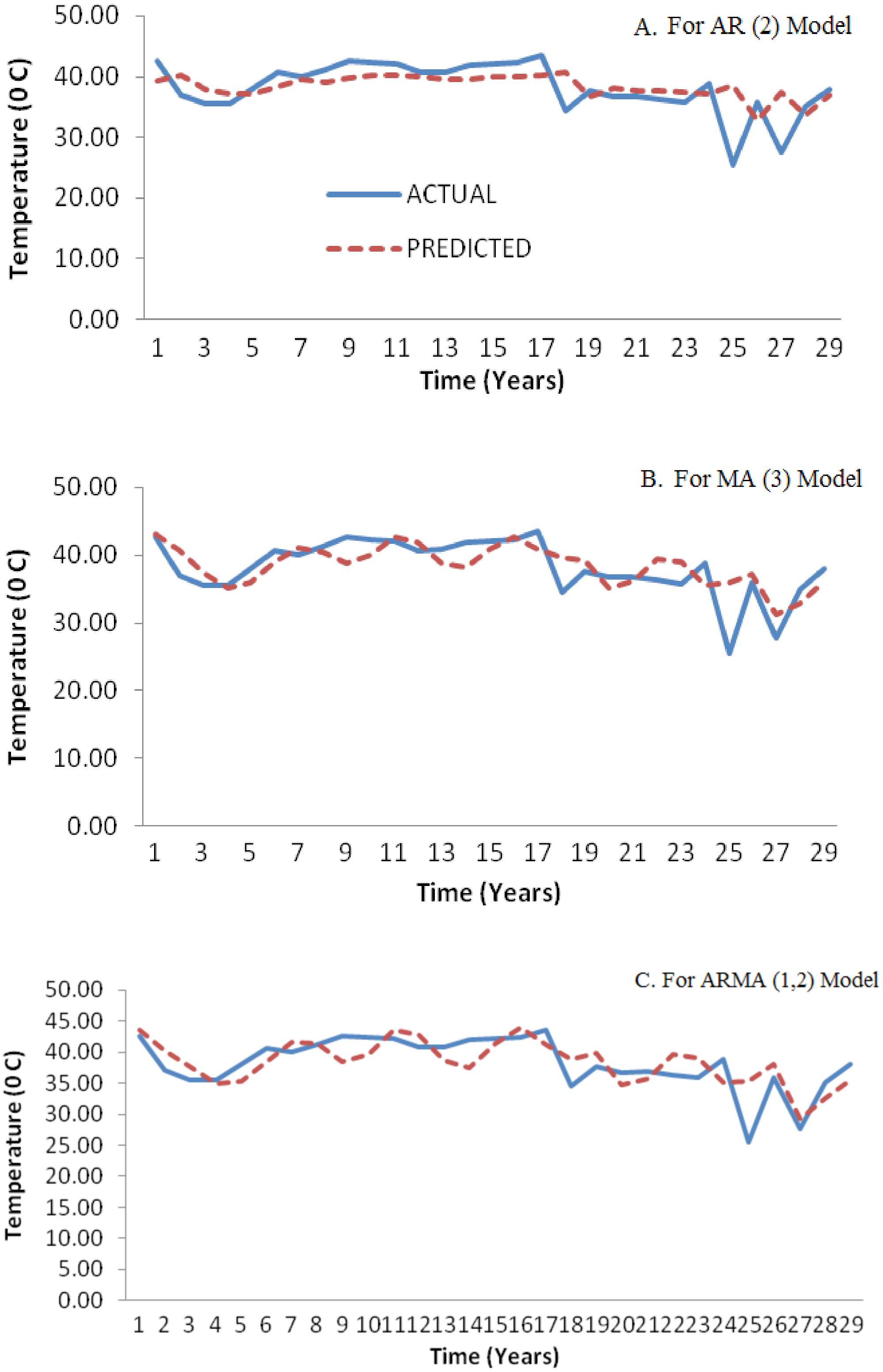

Figure 2. Maximum temperature plot

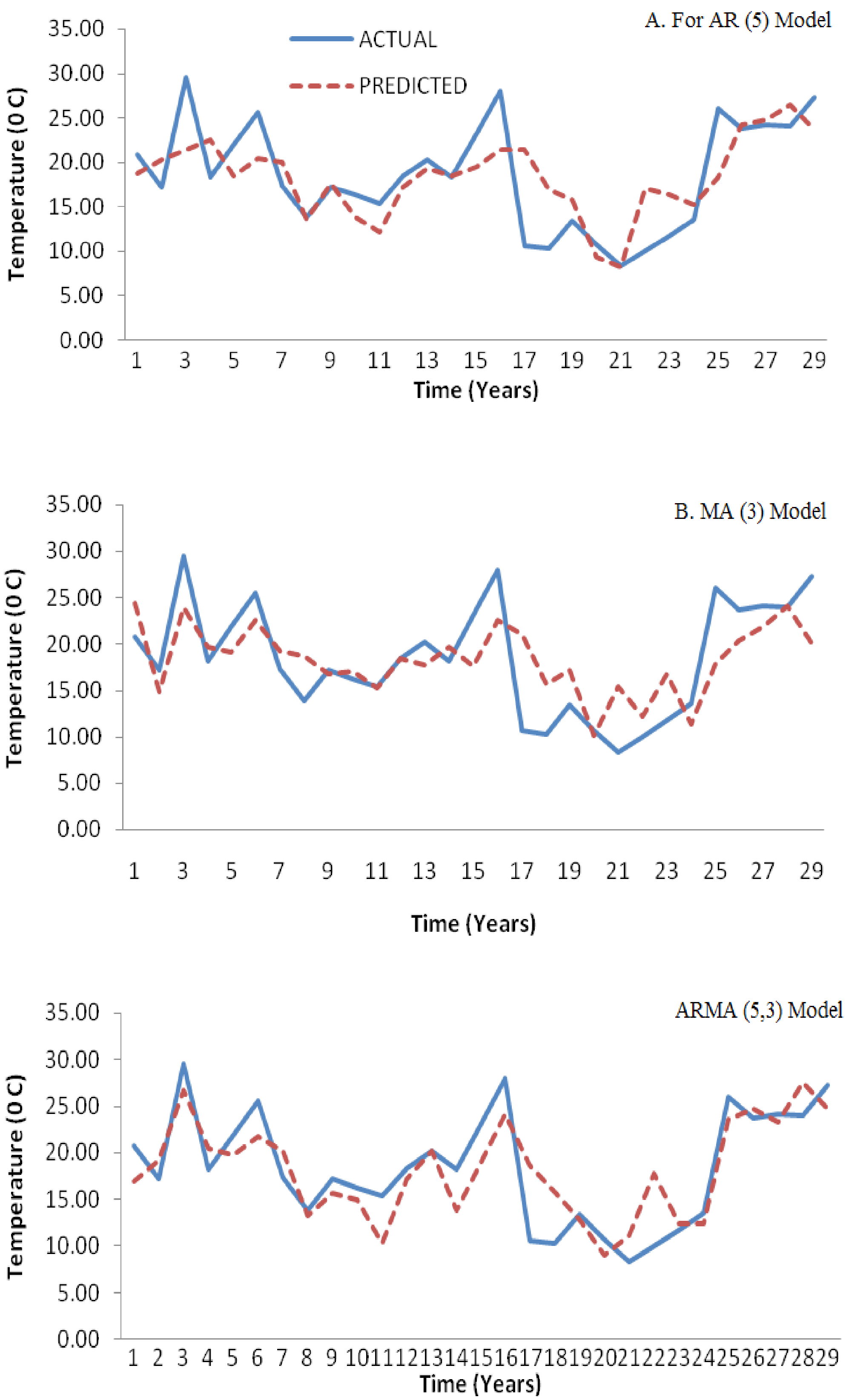

Figure 3. Minimum temperature plot

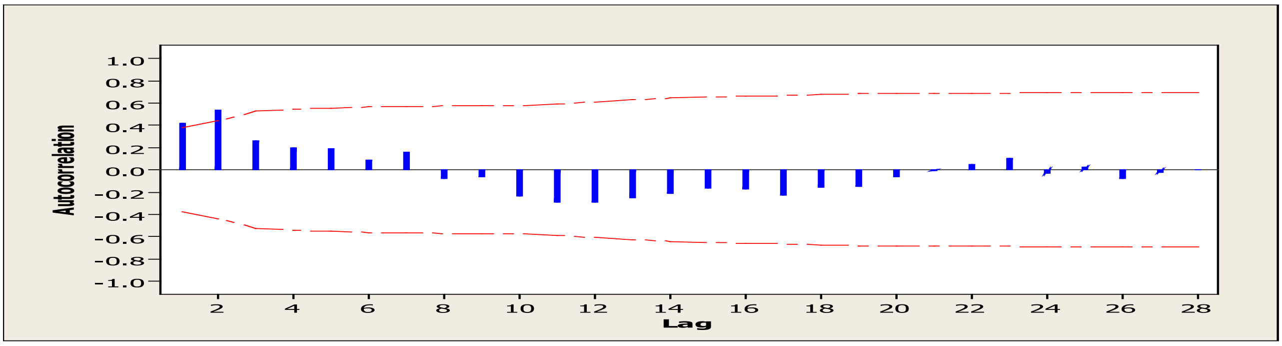

Figure 4. Correlogram of Annual Maximum Temperature

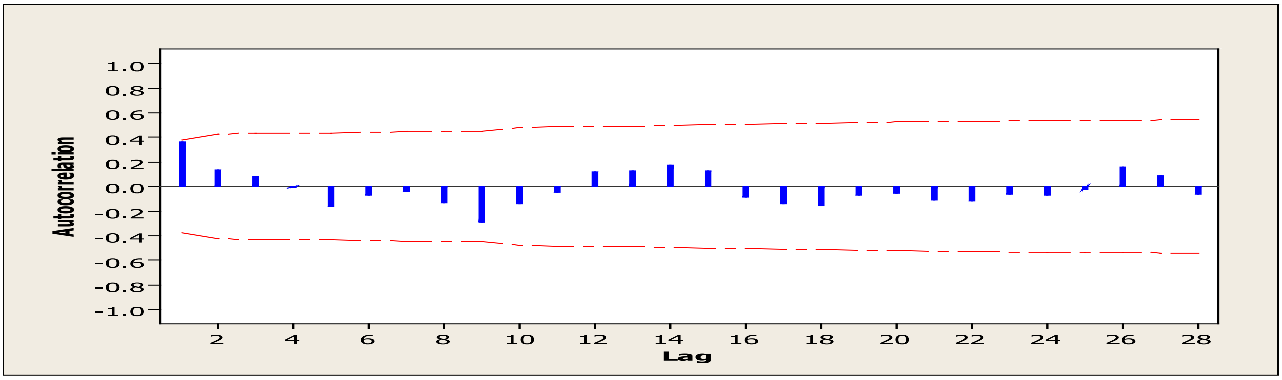

Figure 5. Correlogram of Annual Minimum Temperature

CONCLUSION

The accessibility tore cords of hydrological processes in which temperature is a major determinant is imperative for proper guide and timely preparation against extreme events. Several methods have been used to predict hydrological behaviors, but have shown some weaknesses due to their stochastic nature. The results obtained in this study have been able to demonstrate the robustness of the autoregressive moving average method in the prediction of future minimum and maximum temperatures of the study area. The stationarity of the time series data was achieved through the Mann-Kendal non-parametric test, Lo's long-range dependency test and spectral analysis. No trend detected in the annual time series the parameter based on the Mann-Kendal test, though, the linear trend line indicated that there is gradual shift in the amount minimum and maximum temperatures. However, it is important that the government and relevant stakeholders are aware of the changes in the trend in order to make proper arrangement in case of extreme events. The best models for minimum and maximum temperatures are autoregressive ARMA-model of order (5, 3) with Mean Absolute Percentage Error (MAPE) value of -2.89 and AR-model of order (2) with Mean Absolute Percentage Error (MAPE) value of -1.37. The overall results were promising as it conforms to the Lewi's scaling system (MAPE value less than 10), and the prediction scheme applied in this research could be considered in situations where database is a problem during model development. Based on the findings of this study, it is recommended that another modeling method be used using the same data and results obtained be compared to see which of them gives better outcome.

REFERENCES

Benavides, R., Fernando, M., Agustin, R and Koldo, O. (2007). Geostatistical modelling of air temperature in a mountainous region of northern Spain. *Agricultural and Forest Meteorology* **146**: 173-188.

Besaw, L. E., Rizzo, D. M., Bierman, P. R and Hackett, W. R. (2010). Advances in ungauged stream flow prediction using artificial neural networks. *Journal of Hydrology* **386:** 27–37. doi:10.1016/j.jhydrol.2010.02.037.

Box, G. E. P., Jenkins, G. M and Reinsel, G. (2008). Time series analysis: forecasting and control. Wiley, Oxford, 4th Edition.

Butcher, J., Shoemaker, L., Clements, J. T and Thirolle, E. (1998). Basin modeling online training module. United States Environmental Protection Agency. Basin Academy Web. http://www.epa.gov/watertrain/modeling.

Chatfield, C. (2004). The analysis of time series: An Introduction, sixth edition: New York, Chapman and Hall/CRC.

Edwin, A.I and Martins, O.Y. (2014) Stochastic characteristics and modelling of monthly rainfall. Time series of Ilorin, Nigeria. *Open Journal of Modern Hydrology* **4**: 67-79. http://dx.doi.org/10.4236/ojmh.2014.43006.

Ghanbarpour, M. R., Abbaspour, K. C., Jalalvand, G and Moghaddam, G. A. (2007). Stochastic modeling of surface stream flow at different time scales: Sangsoorakh karst basin, Iran. *Journal of Cave and Karst Studies* **72** (1): 1–10. DOI: 10.4311/jcks2007ES0017.

Hamzacebi, C., (2008). Improving artificial neural networks' performance in seasonal time series forecasting. *Information Sciences* **178**: 4550-4559.

Kumar, V. S and Vanajakshi, L, (2015). Short-term traffic flow prediction using seasonal ARIMA Model with limited input data. *European Transport Resources Review* 10.1007/s12544-015-0170-8.

Longobardi, A and Villan, P. (2009). Trend analysis of annual and seasonal rainfall time series in the Mediterranean area. *International Journal of Climatology* DOI: 10.1002/joc.2001.

Nail, P. E and Momani, M. (2009). Time series analysis model for rainfall data in Jordan: case study for using time series analysis. *American Journal of Environmental Sciences* **5**(5): 599-604.

Robin, P. (2003). Approaching basin modeling. streamline basin management bulletin. **7**(3):

Rubio, A., Sa´nchez, P. O., Go´mez, V., Gran˜a, D., Elena, R and Blanco, A. (2002). Autoecologı´a de los castan ˜ares de Castilla (Espan˜a) (autoecology of Chestnut Tree Forest in Castilla, Spain). *Investigacio´nAgraria: Sistemas de RecursosForestales* **11** (2): 373–393.

Sa´nchez, P. O., Rubio, A., Blanco, A., Elena, R and Go´mez, V., (2003). Autoecologı ´a parame´trica de los hayedos de CastillayLeo´n (parametric autoecology of beech tree forest in CastillayLeo´n). *Investigacio´nAgraria: Sistemas de RecursosForestales* **12** (1): 87–110.

Tesfaye, Y. G., Meerschaert, M. M and Anderson, P. L. (2006). Identification of periodic autoregressive moving average models and their application to the modeling of river flows. *Water Resources Research* **42:** W01419, doi:10.1029/2004WR003772.

Tizro, A. T., Ghashghaie, M., Georgiou, P and Voudouris, K. (2014). Time Series Analysis of Water Quality Parameters. Journal of Applied Research in Water and Wastewater **1**(1), 43-54.

Zhang, G.P., (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing* **50**: 159–175.