

An Application of Time Series Analysis in Modeling Monthly Rainfall Data for Maiduguri, North Eastern Nigeria

Emmanuel Sambo Uba
Department of Statistics, Ramat Polytechnic, Maiduguri

H R Bakari
Department of Mathematics & Statistics, University of Maiduguri

Abstract

Time series analysis and forecasting has become a major tool in different applications in meteorological phenomena such as rainfall, humidity, temperature, draught and so on; and environmental fields. Among the most effective approaches for analyzing time series data is the ARIMA (Autoregressive Integrated Moving Average) model introduced by Box and Jenkins. In this study, Box-Jenkins methodology was used to model monthly rainfall data taken from Maiduguri Airport Station for the period from 1981 to 2011 with a total of 372 readings. ARIMA (1, 1,0) model was developed. This model was used to forecast monthly rainfall for the upcoming 44 months (3 years 8 months) to help decision makers establish priorities in terms of water demand management and agriculture. Thus, ARIMA (1, 1,0) provides a good fit for the rainfall data of Maiduguri and is appropriate for short term forecast.

Key Words: Time Series Analysis, Rainfall Model, Forecasting, ARIMA.

Introduction

Rainfall is a determinant factor of many natural occurrences. Vegetation distribution and types over land masses are as a result of rainfall. Animals breeding period synchronize with rainfall period. Rainfall events have been directly linked to sickness and diseases particularly those of waterborne and vector-borne types. Crop planting, yields and harvest are influenced by rainfall. Investments in agricultural produce and products are expected to be done in accordance with the knowledge of rainfall and other weather conditions.

Naturally, rainfall variability is of spatial and temporal forms and within these variations if by time series analysis no significant trend is obtained then the rainfall is steady, otherwise, it has changed Janhabi and Ramakar(2013).

With regards to rainfall, studies have pointed to the fact that the climate is changing (Goswami *et al*, 2006). Specifically, the research of (Ragab and Prudhomme, 2002) discusses the variability and uncertainty of rainfall across the whole globe amidst global warming and among others it states that while North Africa witnesses rainfall decrease, the regime of rainfall in some parts of South Africa is increasing. On the other hand, no coherent trend has been recorded in some other regions, in the face of different significant trends of other regions in Sri Lanka (Jayawardene *et al*, 2005). However, (Odjugo, 2010) argues that not any change in the climate should be considered as climate change since climate fluctuates. And in his overview of climate change in Nigeria, he established that rainfall has decreased by 81mm within 105 years period. Thus, not to mistake climate fluctuation for climate change (Nsikanet *et al*, 2011).

The growth of population demands for increased domestic water supplies and at the same time results in higher consumption of water due to expansion in agriculture and industry. Mismanagement and lack of knowledge about existing water resources and changing climatic conditions have consequences of an imbalance of supply and demand of water. The problem is pronounced in semi-arid and arid areas where these resources are limited.

Generally, the study of weather and climatic elements of a region is vital for sustainable development of agriculture and planning. Particularly, rainfall and temperature temporal analyses for trends, fluctuations and periodicities are deemed necessary as such can indirectly furnish with “health” status of an environment. (Afangidehet *et al*, 2010).

Before the modeling of the rainfall forecast, preliminary statistical tests were carried out on the time series to test the presence of trend and stationarity in it. A simple linear regression analysis may provide a primary indication of the presence of trend in the time-series data. The Box-Jenkins method is univariate time series analysis. It is

thus essential to analyse the presence of the unit root in the time series. The stationarity of the time series can be tested with Augmented Dickey Fuller Test. This test is popularly used to check the presence of unit root in the time series (Mahalakshmi *et al*, 2014).

In a country like India where around 70% of the total population is dependent on agriculture directly or indirectly, prediction of rainfall plays a very major role. The Artificial Neural Network (ANN) based on the time series analysis was used to predict the summer monsoon rainfall. (Akashdeep *et al*, 2013).

A number of classical time series studies have been conducted in recent years to assess the nature of the climate change as it has occurred over the world as well as in Bangladesh. As for Bangladesh monthly rainfall time series for the 1981-2010 periods in the Dhaka division was analysed using a seasonal ARIMA Model (Ahmad *et al*, 2014).

The Nigerian climate is humid in the south with annual rainfall over 2000mm and semi-arid in the north with annual rainfall less than 600mm. Generally speaking, there are three climatic zones which cover the North, middle and southern areas of the country: Sahel, Savanna and the Guinea zones. Rainfall commences in approximately March/April in the southern coastal zones, spread through the middle zone in May/June and reaches the northern zone in June/July, reaching its peak over middle and northern zones between July and September (Babatunde *et al*, 2011). The negative impact of climate change such as temperature rise, erratic rainfall, sand storms, desertification, low agricultural yield, drying up of water bodies and sea level rise are real in the desert prone eleven front-line states of Nigeria. Environmental degradation and the attendant desertification are major threats to the livelihood of the inhabitants of the front-line states of Nigeria. These lead to increasing population pressures, intensive agricultural land-use, overgrazing, bush burning, etc (Oluwafemi *et al*, 2010).

Materials and Method

The period of research covers 31 years (1981 – 2011). The monthly rainfall totals data used for the study are secondary data obtained from the Nigerian Meteorological Agency (NIMET), Maiduguri. Box-Jenkins technique was used to analyse the data with the main focus on forecasting using ARIMA models. The Augmented Dickey-Fuller (ADF) test was employed to check for presence of a unit root (stationary and non-stationary), through the use of correlogram: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs at various lags. In statistics, ARIMA models, sometimes called Box-Jenkins models after the iterative Box-Jenkins methodology usually used to estimate them, are typically applied to time series data for forecasting. Given a time series of data $X_{t-1}, X_{t-2}, \dots, X_2, X_1$, the ARIMA model is a tool for understanding and, perhaps, predicting future values in the series. The model consists of three parts: an Autoregressive (AR) part, a moving average (MA) part and the differencing part. The model is usually referred to as the $ARIMA(p, d, q)$ model, where p the order of the autoregressive part is, d is the order of the differencing part and q is the order of the moving average part. If $d = 0$, the model become ARMA, which is a linear stationary model. $ARIMA(d > 0)$ is a linear non-stationary model, taking the difference of the series with itself d times makes it stationary, and then ARMA is applied on to the differenced part (Ete *et al*, 2013). Forecasting X_t for $X_{t-1}, X_{t-2}, \dots, X_2, X_1$, using ARIMA model consists of the following steps. The main stages in setting up a forecasting ARIMA model includes model identification, model parameter estimation and diagnostic checking for the identified model appropriateness for modeling and forecasting. Model identification is the first step of this process. The data would be examined to check the most appropriate class of ARIMA process through selecting the order of the consecutive and seasonal differencing required in making the series stationary, as well as specifying the order of the regular and seasonal autoregressive and moving average polynomials necessary to adequately represent the time series model. (Gujarati, 2004)

The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are the most important elements of the time series analysis and forecasting. The ACF measures the amount of linear dependence between observations in a time series that are separated by a lag k . The PACF plots help to determine how many autoregressive terms are necessary to reveal one or more of the following characteristics: time lags where high correlations appear, seasonality of the series, trend either in the mean level or in the variance of the series. (Gebhard and Jurgen, 2007)

In addition to the non-seasonal ARIMA (p, d, q) model introduced above, we could identify seasonal ARIMA (P, D, Q) parameters for our data. These parameters are: seasonal autoregressive (P), seasonal differencing (D) and seasonal moving average (Q). These parameters were computed after the series was differenced once at Lag1 and differenced once at lag 12 (Lisa *et al*, 1998).

The general form of the above model describing the current value X_t of a time series by its own past is:

$$(1 - \phi_1 B)(1 - \alpha_1 B)^{12}(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B)(1 - \gamma_1 B^{12})e_t \quad 1.0$$

where $1 - \phi_1 B$ = non-seasonal autoregressive of order 1

$1 - \alpha_1 B$ = seasonal autoregressive of order 1

X_t = the current value of the time series examined

B = the backward shift operator; $BX_t = X_{t-1}$ and $B^{12}X_t = X_{t-12}$

$1 - B$ = first order non-seasonal difference

$1 - B^{12}$ = seasonal difference of order 1

$1 - \theta_1 B$ = non-seasonal moving average of order 1

$1 - \gamma_1 B^{12}$ = seasonal moving average of order 1

This model can be multiplied out and used for forecasting after the model parameters were estimated, as discussed below.

After choosing the most appropriate model (Step 1 above) the model parameters are estimated (Step 2) by using the least square method. In this step, values of the parameters are chosen to make the sum of the squared residuals (SSR) between the real data and the estimated values as small as possible. In general, non-linear estimation method is used to estimate the above identified parameters to maximize the likelihood (probability) of the observed series given the parameters values. The methodology uses the following criteria in parameter estimation.

- The estimation procedure stops when the change in all parameters estimate between iterations reaches a minimal change of 0.001.
- The parameters estimation procedure stops when the SSR between iterations reaches a minimal change of 0.0001.

In diagnose checking step (step three) the residuals from the fitted model shall be examined against adequacy. This is usually done by correlation analysis through the residual ACF plots and the goodness-of-fit test by means of chi-square statistics (χ^2). If the residuals are correlated, then the model should be refined as in step one above. Otherwise, the autocorrelations are white noise and the model is adequate to represent our time series (Naill and Momani, 2009).

Results and Discussion

In this study, the purpose is to find a suitable time series model for the forecasting of rainfall in Maiduguri. The data for the analysis were the monthly rainfall data for Maiduguri obtained for the period from 1981 to 2011 at the Nigerian Meteorological Agency (NIMET), Maiduguri,

The data were analysed by test for the presence of unit root in the time series data, presentation of the trend line equation, i.e., linear trend equation, quadratic trend equation and their related graphs. For the analysis of the forecasting performance or ability of the models to forecast rainfall, three error measures were used to determine the adequacy of the model. Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Standard Deviation (MSD) were used. The model selection criterion states that the model with the smallest (minimum) MAPE and MAD is considered to be the best.

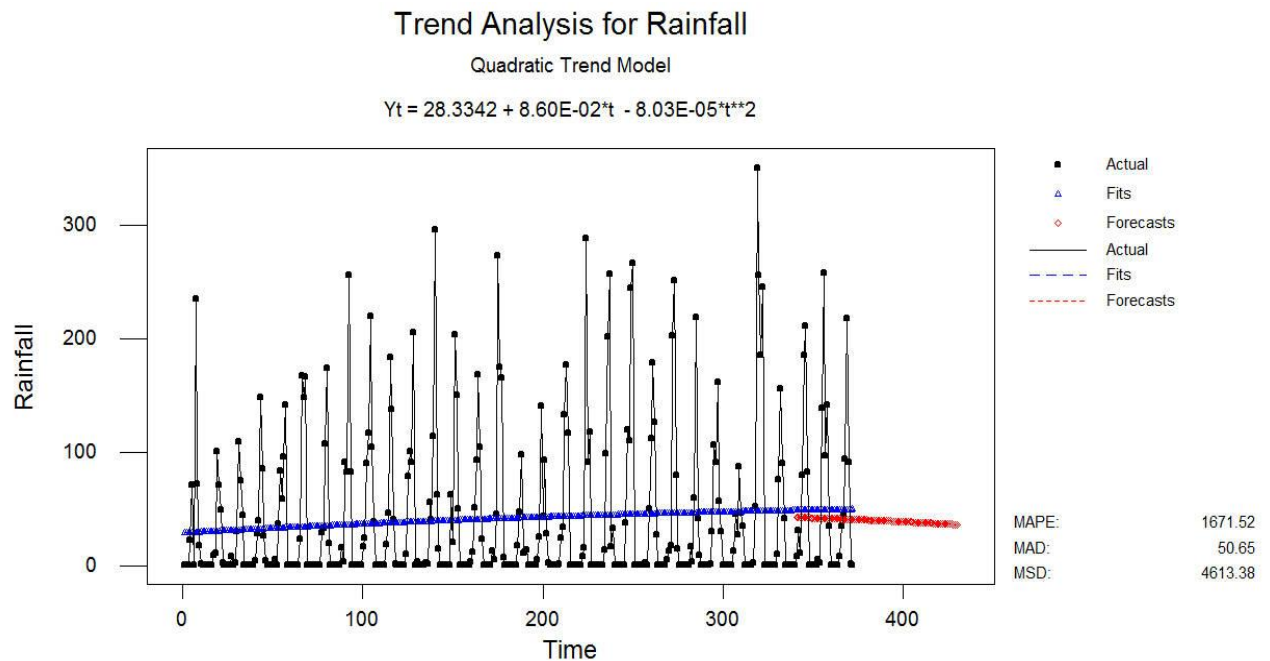


Fig. 1b Model II: Quadratic Trend Model

The quadratic trend model in Fig. 1b is given by

$$Y_t = 28.3342 + 0.086t - 0.00008t^2 \tag{3.0}$$

With the following accuracy measures

$MAPE = 1671.52, MAD = 50.85, MSD = 4613.38$ where Y_t is the monthly rainfall.

Table 2.0 Summary of Trend Analysis Accuracy Measures

TREND	MAPE	MAD	MSD
LINEAR	1663.87	50.66	4614.07
QUADRATIC	1671.52	50.85	4613.38

Performance of the Trend Line Equation

As discussed earlier, the equation or model with the minimum MAPE and MAD is considered to be a better trend or pattern. Among the two models stated above, the linear trend model satisfies the condition and therefore could be selected as an appropriate trend or pattern for modeling the rainfall in Maiduguri, the Borno State capital.

Arima Model

Before fully displaying the analysis and the ARIMA model, let us consider the following: In time series, when the mean and variance are time-variant, the series is considered non-stationary. Figures 2a, 2b and 2c show the test for non-stationarity of the rainfall, using Augmented Dickey Fuller (ADF) test regression.

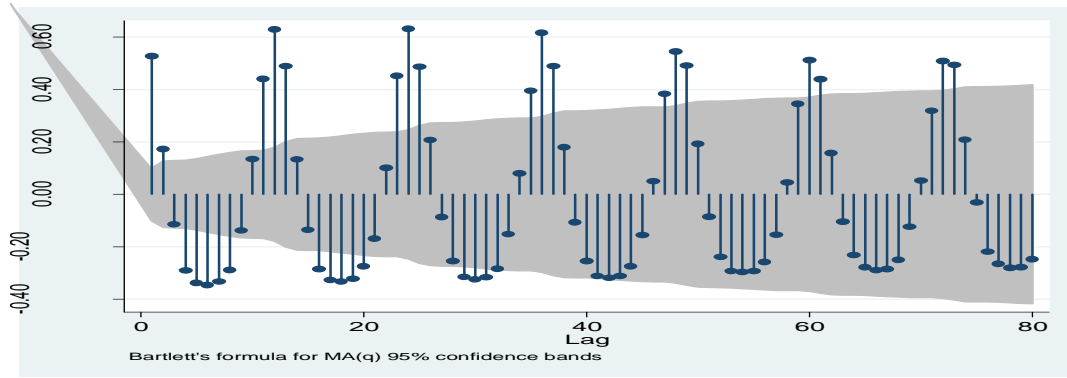


Fig. 2a

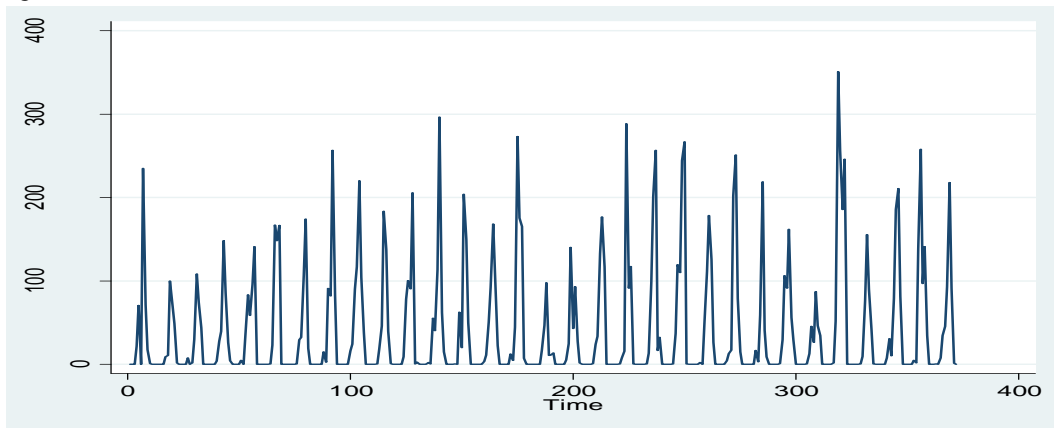
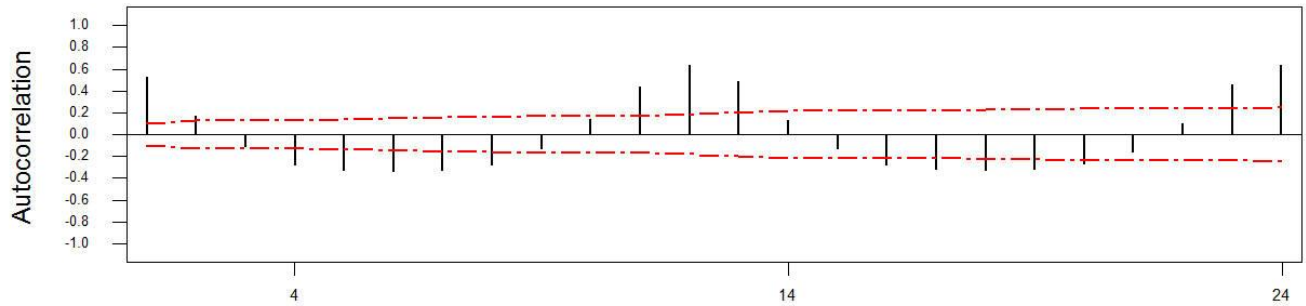


Fig 2b: Time Series Plot for Rainfall

Autocorrelation Function for Rainfall



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.53	10.17	104.35	8	-0.29	-3.53	315.10	15	-0.14	-1.25	663.96	22	0.10	0.84	867.38
2	0.17	2.67	115.54	9	-0.14	-1.63	322.42	16	-0.28	-2.59	695.54	23	0.45	3.72	949.09
3	-0.11	-1.73	120.45	10	0.13	1.58	329.39	17	-0.33	-2.93	737.48	24	0.63	5.01	1108.61
4	-0.29	-4.37	152.33	11	0.44	5.13	404.17	18	-0.33	-2.92	781.09				
5	-0.34	-4.83	195.47	12	0.63	6.86	557.22	19	-0.32	-2.77	822.20				
6	-0.35	-4.67	240.88	13	0.49	4.76	649.82	20	-0.27	-2.31	852.04				
7	-0.33	-4.26	283.18	14	0.13	1.23	656.75	21	-0.17	-1.39	863.26				

Partial Autocorrelation Function for Rainfall

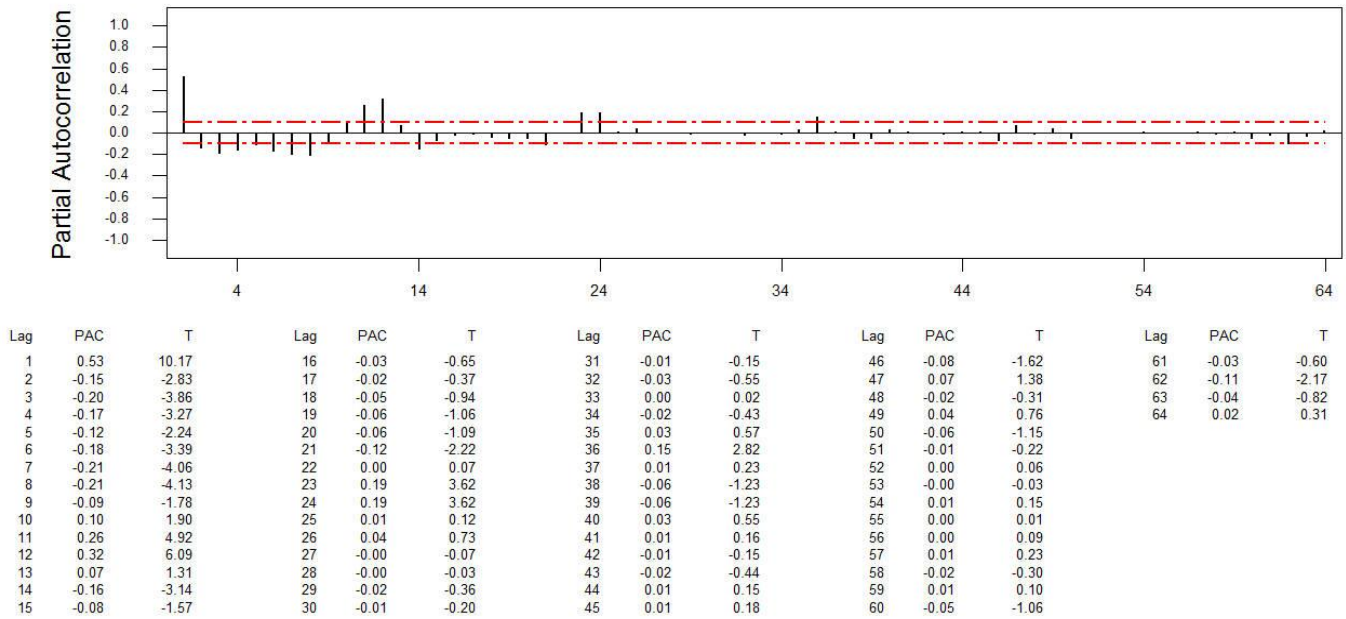


Fig. 2c: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for Rainfall before differencing

Autocorrelation Function for Rainfall

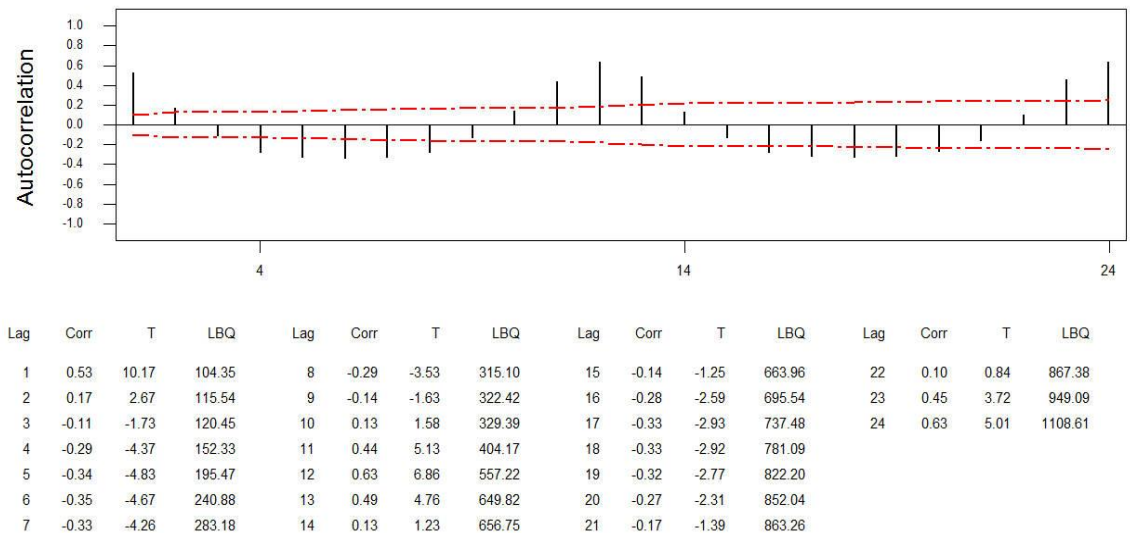


Fig. 2d : Autocorrelation Function of Residuals for Rainfall after First Differencing

MODEL IV: ARIMA Model (1, 1, 0)

Variable: Rainfall

Table 3a Arima Model (1, 1, 0) Parameter Estimates

Type	Coefficient	Std(E)	T	p-value
AR(1)	-0.1245	0.0517	-2.41	0.016
Intercept	0.000	3.425	0.00	1.000

Table 3b Model Adequacy Measure (Ljung-Box)

Lag	12	24	36	48
χ^2	115.7	269.4	403.1	520.5
p-value	< 0.001	< 0.001	< 0.001	< 0.001

Table 3 shows the ARIMA (1, 1, 0) results from the Ljung-Box time series. The table is divided into two; the first (Table 3a) shows the model parameter and p-values for the significance of individual parameters, and the second (Table 3b) shows the model adequacy or accuracy measure from the Box-Pierce (Ljung-Box) chi-square statistic.

AR(1) coefficient is -0.1245 with p-value <0.05, indicating the rainfall depends on, or can be predicted by, its past values of last observation. The model can be expressed below

$$\Delta R_t = 0 - 0.1245\Delta R_{t-1} + u_t \quad 4.0$$

(0.0) (-2.41)

3.425 0.0517

The values in the brackets beneath the coefficients are the t-values while the values beneath the t-values are the standard errors of the estimated coefficients.

ΔR_t = Rainfall at time t at first difference,

ΔR_{t-1} = Rainfall at previous years at time t-1

u_t = General error term

The model adequacy in Table 3b shows that the Box-Pierce (Ljung-Box) chi-square statistic at lag 12, 24 and 36 are having p-value of <0.001, indicating the precision of the model.

Forecasting Rainfall from Arima (1, 1, 0) Model

The model was used to predict the rainfall from the period 272 to 416. It was found that the system changes as time changes, but later becomes static forecast values made by the ARIMA (1, 1, 0) Model as presented in Fig. 3 below.

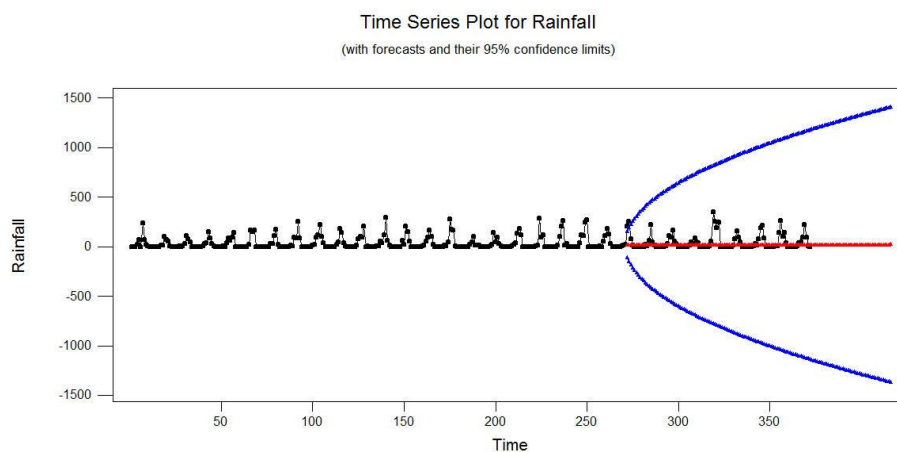


Fig. 3 Rainfall Forecast with 95% Confidence Bounds

Fig. 3 shows the forecast made for the periods 272 to 416, the red line is the average of the forecast while the blue lines above and below are the 95% upper and lower confidence bounds.

Summary

The Nigerian meteorological Agency (NIMET) is the body which has the responsibility to monitor the weather condition of the Nigerian Airspace. The body is responsible for collecting data on rainfall, wind, temperature, etc. This research is intended to estimate the trend of the monthly rainfall in Maiduguri during the period covered as well as develop a suitable model which may enable forecast or predict the monthly rainfall in Maiduguri.

Box-Jenkins technique (ARIMA) models were the main focus through which AR, MA and ARMA were briefly highlighted. In an attempt to determine the trend or pattern of monthly rainfall in Maiduguri, different trend equations, namely; linear, quadratic and exponential, were considered and found that linear and quadratic equations are possible because of the nature of the data. Three error measures, that is, mean absolute percentage error (MAPE), mean absolute deviation (MAD), and mean standard deviation (MSD) were used to choose an appropriate trend model for monthly rainfall in Maiduguri. An equation with minimum MAPE and MAD is considered to be a better trend or pattern. It was discovered that linear trend model satisfies the condition and therefore could be selected as an appropriate trend or pattern for modeling the rainfall in Maiduguri, the Borno State Capital.

The presence of the unit root was tested to determine if the data used for this research were stationary or otherwise. From the test, it was noticed that the data used in this study contains a unit root at level of significance of 1%, 5% and 10%. The Mackinnon approximate p-value for the Dickey-Fuller test statistic is 0.0220 which is less than 0.05, indicating the presence of unit root at 5% level of significance. In time series, when the mean and variance are time variant, the series is considered non-stationary. The ACF and PACF show the test for non-stationarity of rainfall using Augmented Dickey Fuller (ADF) test regression.

After differencing, the ACF and PACF were again plotted and the correlogram were noticed to taper off quickly, an indication that the data are now stationary for accurate analysis. The ARIMA (1, 1, 0) is the only one which can be fitted because of the nature of the rainfall in Maiduguri whereby rainfall starts from June and ends in September. By that it gives us AR(1) which is differenced once. To satisfy one of the objectives of this research, forecast was made for the period 272 to 416 as could be seen in Fig. 3.

Conclusion

Among the series of ARIMA models tested, it was discovered that ARIMA (1, 1, 0) is the only fit of this index. The following conclusions were drawn from the study:

1. Model identification indicated that ARIMA (1, 1, 0) is the only fit of this index.
2. Since the data used in this study contains a unit root at level of significance of 0.05, there is need for differencing. After differencing it once, the ARIMA models with $d = 1$ are suitable for the series.
3. The study shows that linear model gives the best trend line or pattern for monthly rainfall in Maiduguri as noticed in Table 4.2 which gives minimum value of MAPE and MAD.

Recommendations

The study observed that the system changes as time changes but later becomes static with the growth in population and man's activities on the vegetation having consequences of an imbalance of supply and demand of water. It is therefore recommended that government should construct dams across the state, most especially in southern Borno where the geographical area is rocky, while in northern part, free-flow boreholes can be drilled and Chad Basin Development Authority's canals rehabilitated. This will help the practice of irrigation agriculture and regulate climate in the region to support rain-fed agriculture. Substantial planting of trees will help maintain Maiduguri climatic regulation and control overgrazing.

References

- Afangideh, A. I., Francis E. O. and Eja E. I. (2010). *A Preliminary Investigation into the Annual Rainfall Trend and Patterns for Selected Towns in Parts of South-Eastern Nigeria*, Journal of Sustainable Development, Vol. 3, No. 3, September 2010
- Ahmad H. N., Manfred K and Jahir B. A. (2014). *Analysis and Prediction of Time Series Variations of Rainfall in North-Eastern Bangladesh*, British Journal of Applied Science and Technology, 4(11): 1644-1656, 2014. SCIENCEDOMAIN International

- Akashdeep G., Anjali G., Chirag J., Himanshu P., and Neeta V. (2013). *Time Series Analysis of Forecasting Indian Rainfall*, International Journal of Inventive Engineering and Sciences (IJIES) ISSN: 2319-9598, Volume-1, Issue-6, May 2013.
- Babatunde J. A., Ayobami T. S. and Mark T. (2011). *Climate Change Scenarios for Nigeria: Understanding Biophysical Impacts*, © 2011 by Building Nigeria's Response to Climate Change (BNRCC) project c/o Nigeria Environmental Study/Action Team (NEST).
- Ette H. E., Imoh U. M. and Benjamin E. C. (2013). *Modeling Monthly rainfall Data of Port Harcourt, Nigeria by Seasonal Box-Jenkins Methods*, International Journal of Sciences, Volume 2, Issue July 2013, ISSN 2305-3925
- Gebhard K. and Jurgen W. (2007). *Introduction to Modern Time Series Analysis*, © Springer-Verlag Berlin Heidelberg 2007.
- Goswami, B. N., Venugopal V., Sengupa D., Madhusoodanan M. S. and Xavier P. K. (2006). *Increasing Trend of Extreme Rain Events over India in a Warming Environment*, Science, 314: 1442-1445
- Gujarati N. D. (2004). *Basic Econometrics*, Tata McGraw Hill Publishing Company Limited, New Delhi.
- Janhabi M. and Ramakar J. (2013). *Time Series Analysis of Monthly Rainfall Data for the Mahanadi River Basin, India*, Sciences in Cold and Arid Regions, 5(1): 0073-0084, DOI: 10.3724/SP.J.1226.2013.00073
- Jayawardene H. K. W. I., Sonnadra D. U. J. and Jayawardene D., R. (2005). *Trends of Rainfall in Sri Lanka over the last Century*, Sri Lankan J. of Physics, 6: 7-17
- Mahalakshmi N., Umarani P. R. and Samuelvaraj R. (2014). *Statistical Modeling of Tamil Nadu Rainfall*, International Journal of Current Research, Vol 6, Issue 02, pp 4986-4990
- Mahsin M., Yesmin A. and Monira B. (2012). *Modeling Rainfall in Dhakar Division of Bangladesh Using Time Series Analysis*, Journal of Mathematical Modeling and Application, Vol. 1, No. 5, 67-73, ISSN 2178-2423.
- Lisa B., Jeffery J. and Choudary R. H. (1998). *Improving Forecasting for Telemarketing Centres by ARIMA Modeling with Intervention*, International Journal of Forecasting.
- Naill 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, ISSN 1553-345X.
- Nsikan I. O., Terhemba T. E. and Kehinde S. I. (2011). *22 Years Characterized Trends of Rainfall in Abeokuta, Nigeria*, Research Journal of Applied Sciences 6(4): 264-271, ISSN: 1815-932X
- Odjugo P. A, O. (2010). *General overview of Climate Change Impact in Nigeria*, J. Human Ecol., 29: 47-55
- Oluwafemi S. O., Femi J. A., and Oluwatosin T. D. (2010). *Time Series Analysis of Rainfall and Temperature in South West Nigeria*, The Pacific Journal of Science and Technology, Volume 11, Number 2, November 2010, pp 552-563.
- Ragab R. and Prudhomme C. (2002). *Climate Change and Water Resources Management in Arid and Semi-Arid Regions; Prospective and Challenges for the 21st Century*. Biosyst. Eng., 81: 3-34