International Journal of Research in Chemistry and Environment
Vol. 2 Issue 1 January 2012(82-87)
ISSN 2248-9649

Research Paper

Time Series Analysis Model for Annual Rainfall Data in Lower Kaduna Catchment Kaduna, Nigeria

***Attah D.A., Bankole G.M.**
Civil Engineering Department, Kaduna Polytechnic Kaduna, NIGERIA

Available online at: www.ijrce.org

(Received 15th September 2011, Accepted 24th October 2011)

Abstract *- Time series analysis and forecasting has become a major tool in many applications in water resources engineering and environmental management fields. The effects of climate change and variability on water demand in the 21st century makes the time series analysis of rainfall, a major replenishing source of water, more imperative than ever before. The major challenge of water demand management is the ability to effectively estimate the contribution of rainfall to the water budget of any given basin. Among the most effective approaches for time series analysis is the Box-Jenkins' Auto regressive Integrated Moving Average (ARIMA) model. In this study, the Box-Jenkins methodology was used to build an Auto regressive Moving Average (ARMA) model for the annual rainfall data taken from Kaduna South meteorological station within the Lower Kaduna catchment for a period of 47 years (1960 – 2006). From the analysis, the mean annual rainfall was 1385.2mm with a standard deviation of 313.8mm and coefficient of variation of 0.23 (low variation). The range for the period of study is 1407mm.The ARMA model identified is ARMA (1,1) which has Pearson Correlation Coefficient (R^2) of 0.969 and residual ACF and PACF that indicated no pattern. The model is therefore adequate and appropriate for the forecast of future annual rainfall values in the catchment which can help decision makers establish priorities in terms of water demand management.*

Keywords: Time Series Analysis Model, environmental management, Pearson Correlation Coefficient.

Introduction

Time series can be defined as an ordered sequence of values at equally spaced time intervals [1]. Time series models are mathematical representation of the time series. Time series models have been the basis for the study of behaviour of a process over a period of time. The application of time series models are manifold, including sales forecasting, weather forecasting, and inventory studies etc. In decisions that involve factor of uncertainty of the future, time series models have been found to be one of the most effective methods of forecasting. Most often, future course of actions and decisions for such processes will depend on what would be the anticipated results. The need for these anticipated results has encouraged organizations to develop forecasting techniques in order to be better prepared to face the seemly uncertain future. The motivation to study time series models is twofold:

1. To obtain an understanding of the underlying forces and structure that produced the observed rainfall data in the Lower Kaduna catchment of Kaduna state of Nigeria.
2. To fit a model: this can be used for forecasting, monitoring or even feedback and feed forward control.

Time series analysis can be divided into two main categories depending on the type of model that can be fitted. The two categories are:

1. **Kinetic Model**: The data here is fitted as: $x_t = f(t)$, the measurements or observations are seen as function of time.
2. **Dynamic Model**: The data here is fitted as: $x_t = f(x_{t-1}, x_{t-2}, x_{t-3}, \dots\dots\dots)$.

Many methods and approaches for formulating forecasting models are available in literature. This study exclusively deals with Autoregressive Integrated Moving Average (ARIMA) time series models. These models are well described by Box et al (2008). The ARIMA models allow the manager who has only historic data of say rainfall, to forecast future values without having to search for other related time series data, for example, temperature. It also allows for the use of several time series to explain the behaviour of another time series if these other time series are correlated with variable of interest and if there appears to be some cause for correlation.

Box-Jenkins (ARIMA) modelling has been successfully applied in various water resources and environmental management applications [5]. Time series analysis has become a major tool in hydrology. It is used for building mathematical models to generate synthetic hydrologic data, to forecast hydrologic events, to detect trends and shifts in hydrologic records and to fill in missing data and extend records. Time series analysis was used by Langu (1993) as cited by Nail and Momani (2009) to detect changes in rainfall and runoff patterns to search for significant changes in the components of a number of rainfall time series.

Material and Methods

Data: The data for this study were obtained from Kaduna State Water Board, Hydro meteorological section and they include daily rainfall depths in millimeters for the period 1960 to 2006. The annual rainfall depths are the summation of the daily rainfall. The processed data for the time series analysis in the study area is, therefore, the annual rainfall depths. Descriptive statistical analysis and homogeneity test were then performed on these data to verify the integrity of the data. For the ARMA model building, the moving average (MA) values of these data were used.

Study Area: The Lower Kaduna catchment is located between latitude 10° 15' and 10° 45' North and between longitude 6° 15' and 7° 45' East with an area of 1547 square kilometers. It is within the highland climatic zone of Nigeria with a mean annual rainfall of 1230mm.The rainy season is between May and October when about 80% of the annual rainfall occurs. It is drained by River Kaduna.

Water resources in Lower Kaduna catchment are limited and with deteriorating quality due to urban development. Therefore, it is important to know the future water resources budget in order to help decision makers improve their decisions by taking into consideration the available and future water resources. Modelling and forecasting for future water resources have become possible with advances in forecasting methodologies such time series analysis.

Method of Analysis: Models used for the study of the variations in hydro-meteorological variables from year to year have not been based on the analysis of causal factors but on the decomposition of the time series into its basic components by means of the dynamic series model given as:

$$y(t) = u(t) + v(t) + s(t) + e(t)$$

Where: $u(t)$ = trend due to the action of permanent factors;

$v(t)$ = cyclic variations due to the action of rhythmical factors;

$s(t)$ = stochastic components assumed to follow an autoregressive moving average Process;

$e(t)$ = random component due to the action of random factors.

There is an obvious connection between these components of the series and the actual physical processes or situations. Thus, the general trends reflect the general direction of the dynamics of the phenomenon, the cyclic variations or component give the periodical fluctuations about the trend or mean; and the stochastic and the random components or variations are as a result of the impacts of random factors on the natural development of the phenomenon[4].

Trend: The general trend of the hydro-meteorological series can be ascribed to climate change or the influence of man's activities such as agriculture, deforestation and urbanization [4] on the hydrologic system. The extraction of the trend component from the raw data series can be achieved by means of a polynomial fit given as:

$$u(t) = a_1 + a_2 t + a_3 t^2 + \cdots \ldots \ldots + ant^{n-1}$$

For parsimony, (that is, using a model with the fewest possible number of parameters and greatest number of degrees of freedom among all models for adequate representation) the trend can be extracted by means of the basic statistical model for linear trend given as:

$u(t) = a + bt + e(t)$, Where a = intercept, b = slope of the regression line and $e(t)$ = random error component. The parameters a and b are estimated using the least square method if the error terms are uncorrelated. The principal complication with this method in the case of climate data is usually that the data are autocorrelated, in other words, the terms cannot be taken as independent [7].

After the elimination of the trend component, the dynamic series becomes:

$$R(t) = v(t) + s(t) + e(t).$$

In order to establish the general structure of the resulting series or residual series $R(t)$, the correlogram method is used. The autocorrelation coefficient γ_k, is determined as follows:

$$\gamma_k = \left[\sum_{t=1}^{N-k}\{(Rt - Rm)(R(t+k) - Rm)\}\right] / \sum_{t=1}^{N}(Rt - Rm)^2$$

Where N = the length of the time series, k =the difference between two values of the series expressed in time units, Rm = mean value of the series.

The correlogram is obtained by graphical representation of the variations of γ_k with k. If the data series is random, the computed value of γ_k is other than that due, only, to internal (random) variation of the series, the points of the correlogram being very close to the horizontal axis. The confidence limit (CL) corresponding to a given significant level is expressed as:

$$CL(\gamma_k) = \left[-1 \pm Z_\alpha (N - k - 2)^{1/2}\right]/(N - k - 1)$$

, where Z_α = standard normal deviate corresponding to the probability α. For α = 5%, the significant level Z_α = 1.645.

Periodic Component

The determination of the periodic component is achieved by means of a Fourier series of the form:

$$v(t) = \frac{1}{2}a_0 + \sum_{n=1}^{m}\left(a_n \cos\frac{2\pi}{T}nt + b_n \sin\frac{2\pi}{T}nt\right),$$

where n = the order of the harmonic corresponding to the period established by the analysis of the correlogram. $\frac{2\pi}{T}$ = the frequency of the oscillations. a_0, a_n, b_n = Fourier coefficients. The remainder of the dynamic series is given as:

$$R(t) - v(t) = s(t) + \varepsilon(t) = stochastic\ component.$$

After eliminating the trend and periodic components from the original data series, the resulting series is a stochastic process which can be modelled by means of parametric models of an autoregressive moving average (ARMA) type [1, 4].

The general structure of a non-stationary ARIMA type (p, d, q) model is given as:

$w_t = \phi_1 w_{t-1} + \ldots\ldots\ldots\ldots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \ldots\ldots\ldots\ldots - \theta_p a_{t-q}$, where d = difference order; ϕ = parameters of the autoregressive process; θ= parameters of the moving average process; p= order of the autoregressive process; q = order of the moving average process; and a_t = current value of an independent random variable with zero mean and constant variance σ_a^2 (of the "white noise" type).

When d = 0, $w_t = s_t - \mu$, where s_t = the current value of the modelled series at moment t; μ = parameter that determines the level of the process, in the case of stationary process μ is the mean of the series.

The steps necessary for establishing the ARIMA type model to describe, optimally, the data series are:

1. Estimation of the order p, q, and d of the model by the analysis of the autocorrelation (ACF) and partial autocorrelation (PACF)functions;
2. Determination of the parameters $\phi, and\ \theta$ by means of the maximum likelihood and least square methods; and
3. Analysis of the structure of the residuals obtained after the application of the model in order to verify appropriateness of the chosen model. If the residuals obtained do not comply with the above mentioned criteria indicated for a_t, the model should be re-analyzed.

Results and Discussion

The summary of the descriptive statistics of the annual rainfall data is presented in Figure 1below.From the figure, it can be inferred that the mean annual rainfall of Lower Kaduna catchment is 1385.2 mm and median is 1320 mm. This indicates that the annual rainfall values are right skewed. The high standard deviation value can easily be correlated with the high range (1407mm) of the annual rainfall values. Fig. 1

The range is the difference between the maximum and minimum annual rainfall values. The standard deviation and the range indicate the variability of the annual rainfall and hence denote how reliable the rainfall is in terms of its persistence as a constant and stable replenishing source. The p-value is less than 0.05 indicating that the data is non-normal.

To test whether the annual rainfall data follow a normal distribution, the skewness and kurtosis were computed. Skewness measures symmetry or lack of symmetry. The skewness for normal distribution is zero. Negative value of skewness indicates left skewness while positive value indicates right skewness. Kurtosis is a measure of data peakness or flatness relative to normal distribution. The normal standard distribution has zero kurtosis. Positive kurtosis indicates a peaked distribution and negative kurtosis indicates flat distribution. The annual rainfall exhibits right skewness and peaked distribution as indicated in Figure 1.

Homogeneity Test: A climate variable is said to be homogeneous when its variations are caused only by fluctuation in weather and climate. To test the homogeneity of the climate data time series, the "Run Test" was applied. The distribution of the number of Runs(R) approximates a normal distribution with the following mean (E) and variance $(Var.)$:

$$E(R) = (N+2)/2;$$

$$Var(R) = [N(N-2)]/[4(N-1)].$$

The Test statistics Z is defined as: $Z = [R - E(R)]/\sqrt{Var(R)}$, for significant level of

α = 0.01and α = 0.05. The Null Hypothesis of homogeneity is verified if $|Z| \leq 2.58$ and

$|Z| \leq 1.96$ respectively.

From the Run Test results, as presented in Table1, the Null hypothesis of homogeneity is not rejected. Table 1

The plot of the original data, as shown in Figure 2, does not show any seasonal variation since the data is the annual rainfall total. Figure 1

The auto correlation function (ACF) and the partial auto correlation function (PACF) plots of the original data ,as shown in Figure3, indicate that the annual rainfall data is stationary and therefore does not require differencing (d=0).That is, the annual rainfall series is serially independent.

From Figure3, it can be observed that there is only one significant pike in the PACF plot. This indicates that the auto regressive process would be of the order 1 (p=1) and the moving average process would also be of the order 1 (q=1).Therefore, the appropriate ARIMA model to fit the annual rainfall data would be an ARIMA (1, 0, 1) model which is equivalent to ARMA (1, 1).

After fitting the ARMA (1, 1) model, the model parameters were estimated from the following equation using the least sum-of- square of residuals method:

$w_t = \phi_1 w_{t-1} + a_t - \theta_1 a_{t-1}$ or

$X_t = \mu + \phi_1 X_{t-1} + a_t - \theta_1 a_{t-1}$

The model fit statistics are presented in Table 2, while the estimated model parameters and their significant levels are presented in Table3. The estimated coefficients are all statistically significant at 5% level. Table 2 ,3

The model is validated by ACF and PACF plots of the residuals. A pattern less (white noise) ACF and PACF indicate a good fit. From ACF and PACF plot, as shown in Figure 4, the ACF and PACF of the residuals have no pattern. Also, from Table 2,the Pearson product moment correlation coefficient (R^2) which measures the linear association between individual pairs of forecasts and observations is very high (0.969)(for a perfect fit $R^2 = 1.0$). These indicate that the ARMA (1, 1) model identified is adequate for the forecast of future annual rainfall events in the study area. Figure 4

Conclusion

The ARMA (1, 1) model identified is adequate to represent the observed annual rainfall data and can be used to forecast future rainfall data. The ARMA (1, 1) model can be written as:

$$X_t = 1385.2 + 0.969X_{t-1} + a_t - 0.6a_{t-1}$$

References

1. Box G.E.P.; Jenkins G.M and Reinsel G.C.: Time Series Analysis: Forecasting and Control; 4th Edition;. John Wiley & Sons; Inc. U.K. **(2008).**
2. DeLurgio S.A., Forecasting Principles and Applications; 1st Edition. Irwin McGraw-Hill Publishers **(1998).**
3. Hadi N. S.: *Time Series Analysis for Hydrological Features: Application of Box- Jenkins Model to Euphrates River;* Journal of Applied Sciences. 6 (9); Pp. 1929- 1934. Asian Network for Scientific Information **(2006).**
4. Haidu I., Serbrn P. and Simota M.: *Fourier-ARIMA Modelling of Multiannual Flow Variation. In: The Influence of Climate Change and Climate Variability on the Hydrologic Regime and Water Resources.* Proceedings of the Vancouver Symposium, August, 1987, IAHS Publications. No, 168 **(1987).**
5. Nail P. E. and Momani M.: *Time Series Analysis Model for Rainfall Data in Jordan: A Case Study for Using Time Series Analysis.* American Journal of Environmental Science. 5(5):599-604. Science Publications **(2009).**
6. Smith R. L.: *Statistical Trend Analysis.* Weather and Climate Extremes in a Changing Climate: Region of Focus: North America, Hawaii, Caribbean and U.S. Pacific Islands **(2007).**

Table 1: Run Test

	Annual Total Rainfall
Test Value	1320
Cases < Test Value	23
Cases > = Test Value	24
Total Cases	47
Number of Runs	19
Z	-1.472
Aymp.Sig. (2-tailed)	0.141

Table 2: Model Fit Statistics

Model Statistics

Model	Number of Predictors	Model Fit statistics		Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	R-squared	Statistics	DF	Sig.	
Ten Year Moving Average Annual Total Rainfall-Model_1	1	.969	.969	27.758	16	.034	1

Table 3: Model Parameter Estimates

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Ten Year Moving Average Annual Total Rainfall-Model_1	Ten Year Moving Average Annual Total Rainfall	No Transformation	Constant		-17449.0	21796.890	-.801	.429
			AR	Lag 1	.969	.065	14.966	.000
			MA	Lag 1	-.600	.155	-3.869	.000
	Year	No	Numerator	Lag 0	9.527	11.009	.865	.393

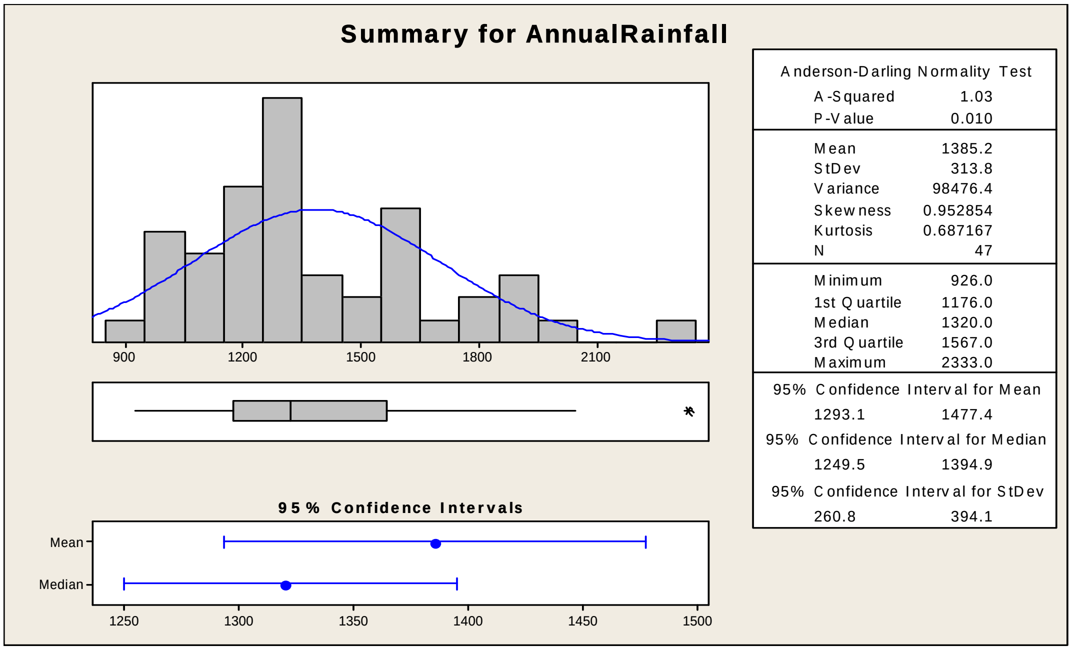

Figure: 1 Descriptive Statistics of Annual Rainfall

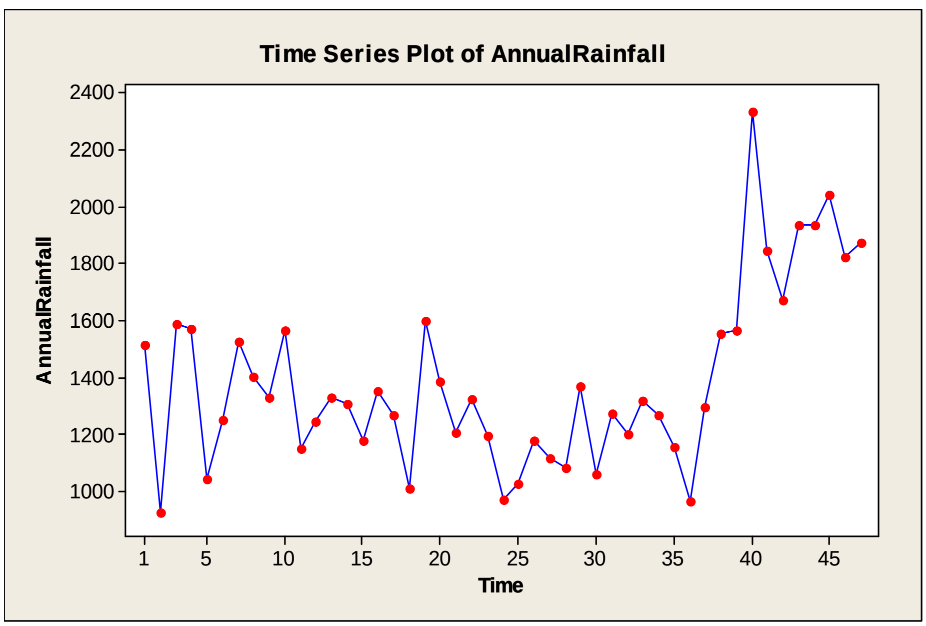

Figure 2: Time Series Plot of Annual Rainfall

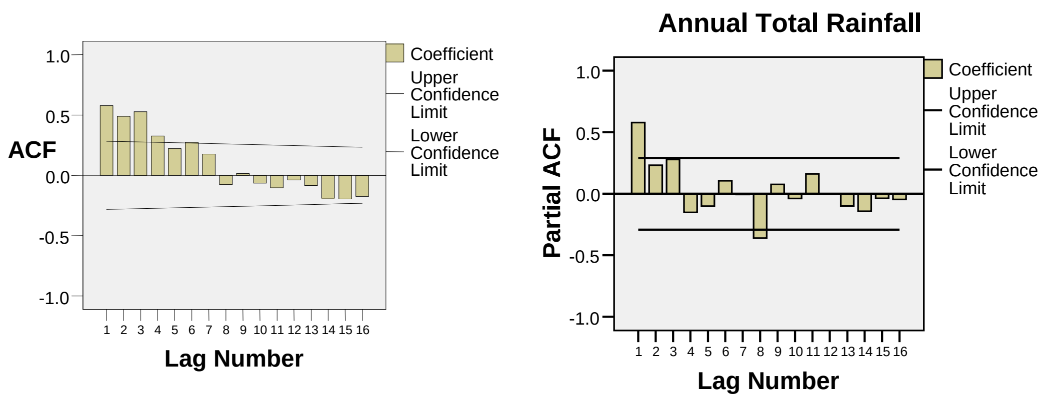

Figure 3: ACF and PACF Plots of Annual Rainfall

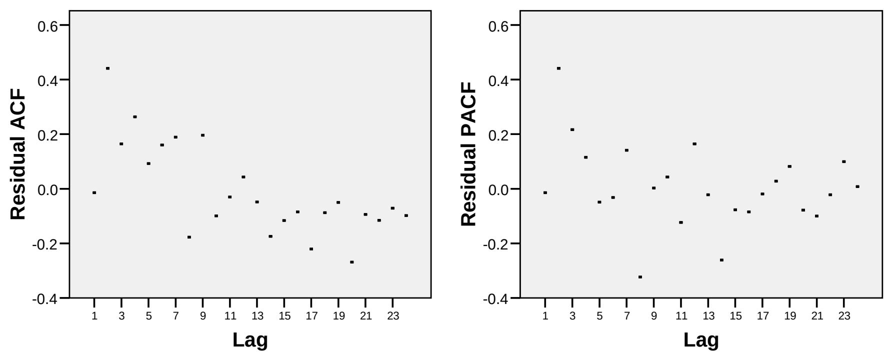

Figure 4: Residual ACF and PACF.