IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 11, Issue 5 Ver. IV (Sep. - Oct. 2015), PP 77-84
www.iosrjournals.org

A Time Series Model of Rainfall Pattern of Uasin Gishu County

Metrine Chonge[1], Kennedy Nyongesa[2], Omukoba Mulati[3], Lucy Makokha[4], Frankline Tireito[5]
[1,2,3,4,5]*(Department of Mathematics, Masinde Muliro University and Technology, Kenya)*

Abstract: *In this paper we fit a time series model that best describes the rainfall pattern of Uasin Gishu county from the general ARIMA family and generate the values (p,d,q)(P,D,Q)$_s$. The model that best fitted the Kapsoya historical rainfall data was SARIMA (0,0,0)(,0,1,2)$_{12}$. This model is used to forecast average expected monthly rainfall statistics for two years. For verification and data fitting to the model, R computer software was employed. The data used is real rainfall data from Kapsoya meteorological station in Uasin Gishu County.*
Keywords: *Time series, Forecasting, Rainfall values, ARMA, ARIMA, SARIMA.*

I. Introduction

Rainfall is the climatic factor of greatest economic and social significance in Africa and Kenya in particular. It is the most critical and key variable both in atmospheric and hydrological cycle. The economies of East African countries heavily depend on rain-fed agriculture. In Kenya, agriculture contributes 24% of GDP, generates over 60% foreign exchange earnings, provides employment to over 70% of the population, provides raw materials to agro- industries and provides over 45% of the annual government budget. Livestock contributes 40% of agricultural GDP and 10% of the total GDP (Raphael, 2012). Uasin Gishu County lies in the Rift Valley which is considered Kenya's granary. There are large scale farmers growing wheat, maize and dairy farming.

Interactions between the various components of the climate system such as the oceans, land and atmosphere have brought about climate change. This is characterized by rainfall variabilities which brings with it negative impacts to the countries' economies. This has necessitated efforts to understand the coherent multi decadal fluctuation in the global climate change and make predictions of rainfall extremes.

Methods of prediction of rainfall extremes have often been based on studies of physical effects of rainfall or on statistical studies of rainfall time series. Because rainfall occurs based on a specific time and there is a correlation between the previous data and subsequent ones, the best method for analysing rainfall data is using time series. This is proven by Nail and Momani (2009) who revealed that a researcher with data for a past period can use Univariate Box-Jenkins method to forecast values without having to search for other related time series data.

Montgomery and Johnson (1976) considered the Box and Jenkins methodology the most accurate method for forecasting of time series. Fateme et al, (2013) in studying of drought, modeling and forecasting the precipitation of shiraz city of Iran, used three models; Box-Jenkins, Decomposition and Healt Winterz on precipitation for the period 1977 to 2010. The Box-Jenkins approach was chosen as the most appropriate method for forecasting.

Rediat (2012), carried out a statistical analysis of rainfall pattern in Dire Dawa, Eastern Ethiopia. He used descriptive analysis, spectrum analysis, and univariate Box-Jenkins method. He established a time series model that he used to forecast two years monthly rainfall. Results showed a rainfall extreme event occurs every 2.5 years in Dire Dawa region. From the literature, no ARIMA model has been used in modeling Kenyan rainfall data and in particular Uasin Gishu rainfall data. Therefore it will be interesting to use ARIMA in modeling Uasin Gishu rainfall data.

Forecasting of severe weather and extreme climate events is one of the major challenges facing meteorological services worldwide especially in Kenya where we have been experiencing severe droughts and floods in 1991-2011. This has led to negative social-cultural, physical, economical and environmental effects. Researchers have made enormous efforts in addressing the issue of accurate rainfall predictability including the use of numerical and statistical methods (Ndetei, 2013). Time series analysis provides great opportunities for describing, explaining, modeling and predicting climatic variability and impacts. To understand the meteorology information and integrate it into planning and decision making process, it is important to study the temporal characteristics and predict lead times of rainfall in the region. To the author's knowledge, there has been no research published on extreme rainfall events prediction in Kenya. This calls for a statistical Time Series model of rainfall data which will be used to accurately predict extreme rainfall patterns of Kenya and Uasin Gishu County in particular. This will go a long way in assisting planning and decision making on the important resource water whose main source is rainfall.

II. Methodology

Time series data is a set of observations of a variable measured at equally spaced, discrete time intervals. A stationary time series has finite variance, correlations between observations that are not time-dependent, and a constant expected value for all components of the time series (Brockwell and Davies, 1991, p12). When a time series lacks stationarity either in variance or mean or both, transformations are applied on the time series at the appropriate lags in order to achieve stationarity before an ARMA model is fitted on the data set.

Assuming that $\{Z_t\}$ is a Gaussian white noise series with mean zero and variance σ^2, the time series $\{X_t\}$ is said to be an AR process of order p if

$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots\ldots + \varphi_p X_{t-p} + Z_t$……………………………………………………………………(1)

With; X_t as a stationary time series

$\varphi_1, \varphi_2, \ldots\ldots, \varphi_p$ are the parameters of the AR.

Assuming that $\{Z_t\}$ is a Gaussian white noise with mean zero and variance δ^2, the time series $\{X_t\}$ is said to be an $MA(q)$ process of order q if

$X_t = \vartheta_0 Z_t + \vartheta_1 Z_{t-1} + \ldots\ldots + \vartheta Z_{t-q}$……………………………………………………………………(2)

With; $\vartheta_1, \vartheta_2, \ldots\ldots \vartheta_q$ as the MA parameters

A stationary zero mean ARMA (p,q) model is defined as (see Brockwell and Davies, 2002) a sequence of random variables $\{X_t\}$ which satisfy

$X_t - \varphi_1 X_{t-1} - \ldots\ldots - \varphi_p X_{t-p} = Z_t + \vartheta_1 Z_{t-1} + \ldots\ldots + \vartheta_q Z_{t-q}$……………………………………………………………………(3)

and constant variance δ^2, for every t and where $\{Z_t\}$ is a sequence of uncorrelated random variables.

A process is called an ARIMA (p, d, q) if d is a non-negative integer such that $(1-B)^d X_t$ is an ARMA (p,q) process and B is the backward shift operator. An ARIMA (p, d, q) written as

$(B)\, Y_t = \vartheta\,(B)\, Z_t$ ……………………………………………………………………(4)

The seasonal ARMA $(p,q)(P,Q)_s$ model can be written formally as;

$\emptyset(B^s)\varphi(B)(X_t - \mu) = \theta(B^s)\vartheta(B)Z_t$ [without differencing]…………………………………………(5)

The process $\{X_t\}$ is a $SARIMA(p,d,q)(P,D,Q)_s$ if the differenced series

$Y_t = (1-B)^d\,(1-B^1)^D X_t$ is a causal $ARMA$ process defined by

$\varphi(B)\emptyset_P(B^s)\,\nabla^d\nabla_s^D Y_t = \vartheta(B)\theta_Q(B^s)Z_t$, ……….. $\{Z_t\}$- $WN(0,\sigma^2)$ ……………………………………(6)

With;

i. The nonseasonal AR and MA component being represented by polynomials $\varphi(B)$ and $\vartheta(B)$ of orders p and q respectively.
ii. The seasonal AR and MA components by $\emptyset(B^s)$ and $\theta(B^s)$ of orders P and Q respectively.
iii. Non-seasonal and seasonal difference components by $\nabla^d = (1-B)^d$ and $\nabla_s^D = (1-B^s)^D$
iv. p, d and q as the order of non-seasonal AR, seasonal differencing and seasonal MA respectively.
v. P, D and Q as the order of seasonal AR, seasonal differencing and seasonal MA respectively.
vi. X_t representing time series data at period t
vii. Z_t representing Gaussian white noise process at period t
viii. B representing backward shift operator $B^k X_t = X_{t-k}$
ix. ∇_s^D representing seasonal difference
x. ∇^d representing non-seasonal difference
xi. s representing seasonal order ($s = 12$ for monthly data)

The exact time series modeling procedures can be found in standard text books. In this study, we used R computer software to analyse data. The data under study was the Kapsoya monthly rainfall data which was obtained from Kenya Meteorological Agencies. The data consists of 444 monthly observations from January 1977 to December 2014.

The information criteria used to select the best model are;

1. Akaike Information Criterior (AIC) by Akaike (1978)

$$AIC = \frac{\mathrm{Log}\,\hat{\sigma}_k^2 + n + 2k}{n} \quad \ldots\ldots(7)$$

Where; $\hat{\sigma}_k^2 = \frac{SSE_k}{n}$ denotes the MLE for the error variance

K is the number of seasonal and non-seasonal AR and MA parameters to be estimated in the model. Wei (1990) says $k = p + q + P + Q + 1$

n is the number of observations.

The value of K yielding the minimum AIC specifies the best model.

2. Bayesian Information Criterior (BIC)

$$BIC = -2L_k + k \ln n \ldots\ldots(8)$$

With; n as the sample size and

L_k as the number of parameters

To evaluate the forecasting performance of the selected model, we used the Mean Absolute Percentage Error which is defined by equation (9).

$$MAPE = n^{-1} \frac{\sum_i^n |y_i - \hat{y}_i|}{y_i} * 100 \ldots\ldots(9)$$

With; y_i as the actual observation and $\hat{y}_i$ as the predicted value .

The model with minimum mean error is the best.

III. Results

We see the mean of the original data is 91.15. We therefore plot a time series plot of the data showing the mean. The plot is shown in Figure 1.

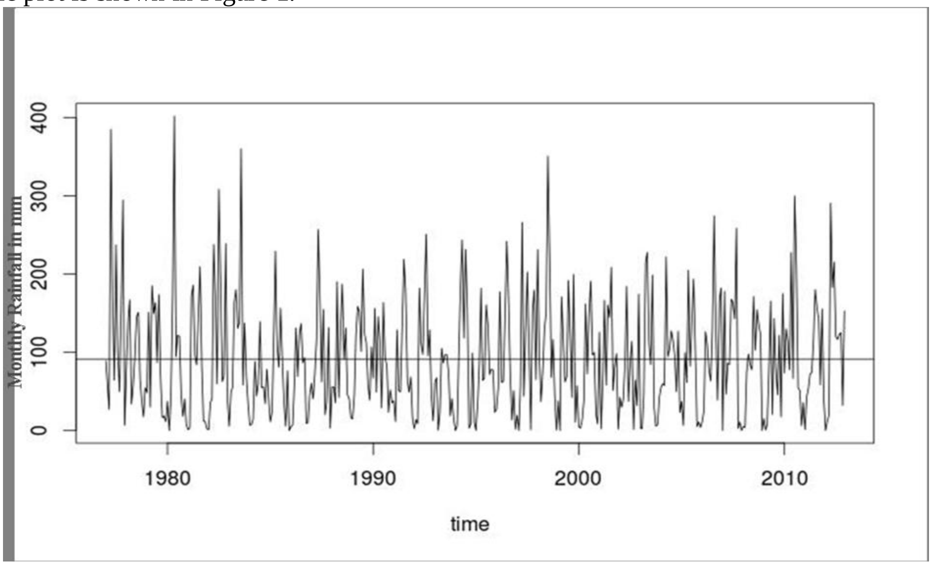

Figure 1: plot of monthly rainfall data against time

From Figure 1, clearly, the plots of monthly rainfall fluctuates around the mean, depicting no trend. This implies that the series is non-seasonally stationary. This time plot also reveals lack of the horizontal component as data values do not fluctuate around a constant value. There is no cyclical component either as data does not exhibit rises or falls around trend levels. However the time series shows a form of pattern that repeats every year implying seasonality at period twelve months of the year.

Stationarity in variance and mean is a requirement for a time series before an ARIMA model is fit on data. On studying the Kapsoya monthly rainfall data, we observe that the peaks of the time plots are not repeated with the same intensity indicating a non-constant variance and hence the series lacks stationarity in variance. To verify stationarity in mean, we check by correlograms. The result is shown in Figure 2

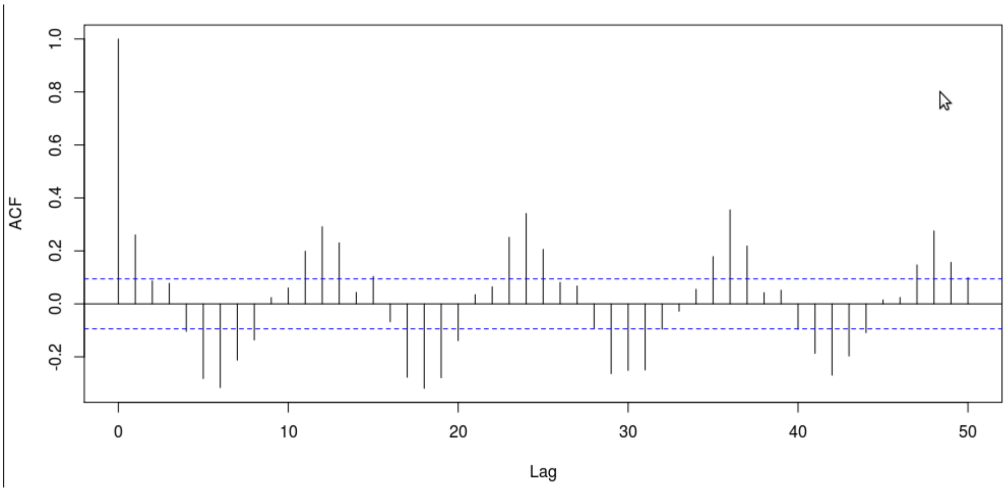

Figure 2: ACF plot of the monthly data

From Figure 2, we depict that the ACF is sinusoidal at the multiples of seasonal lags indicating the presence of strong seasonality behaviour. However, as lags increase the autocorrelations at multiples of seasonal lags seem not to decay implying non-stationarity in the seasonal component of the monthly rainfall data. As non-seasonal lags increase the autocorrelations seem to be relatively decaying indicating that the series is non-seasonally stationary. This can also be confirmed by the plots of $PACF$ versus lag as shown in Figure. 3.

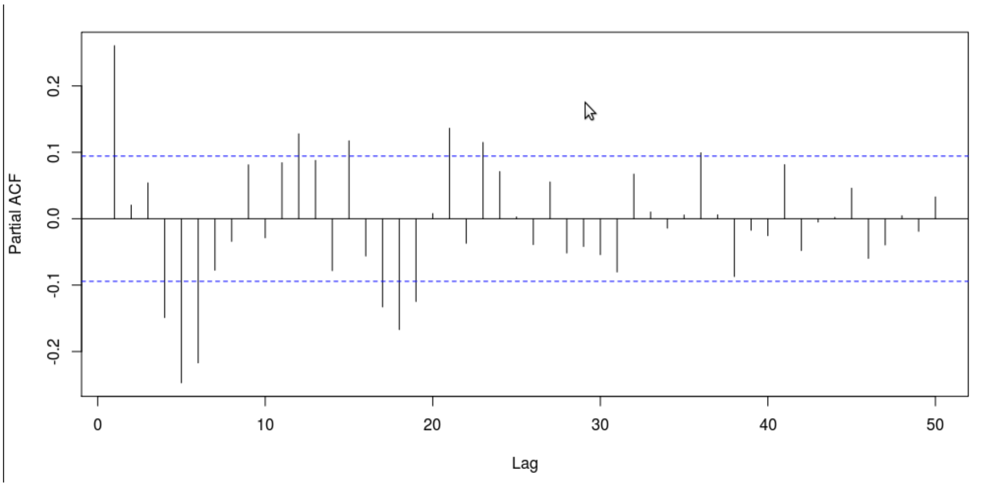

Figure 3: PACF plot of monthly data

A similar trend is depicted in Figure 3 for the partial case with sizeable coefficients at early multiples of seasonal lags indicating seasonal non-stationarity. The ADF and KPSS tests carried out on the original monthly rainfall series gave their results as -5.265 and 0.0474 respectively. This confirms the argument that the series is seasonally stationary. However as seen from the ACF in Figure 2 and PACF in Figure 3, we conclude that the data lacks seasonal stationarity property.

This calls for the introduction of the seasonal stationarity component in the data. This is achieved by performing transformations on the data. The transformation performed in this study is the square-root because it is the transformation that brings about stationarity in variance and seasonal differencing to bring about stationarity in mean. The resultant data is plotted against time as shown in Figure 4.

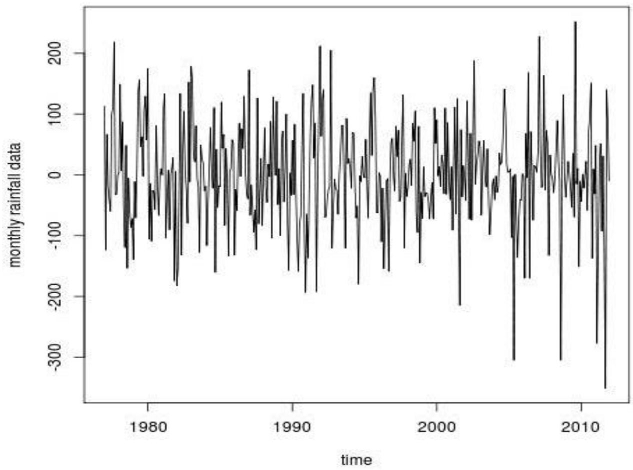

Figure 4: Seasonally differenced square-root-transformed monthly rainfall data.

Clearly, the plot depicts a stationary time series reverting around mean zero with a stable variance. To confirm this allegation we plot ACF and PACF against time as shown in Figures 5 and 6.

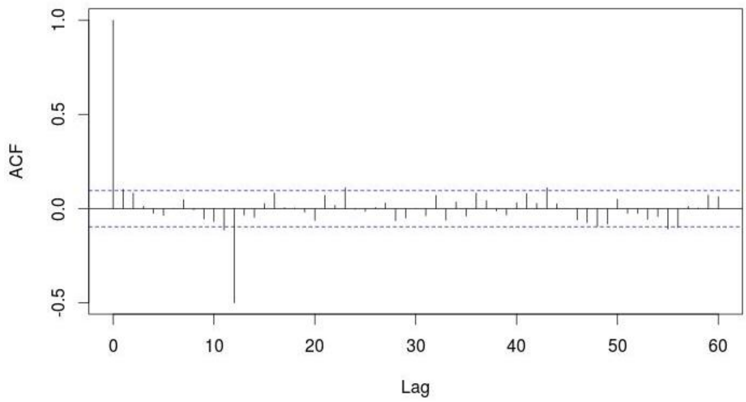

Figure 5: ACF of the differenced square-root transformed series.

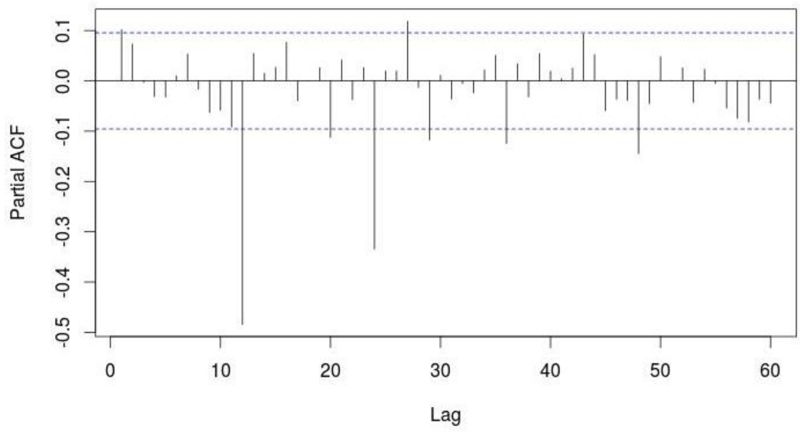

Figure 6: PACF of the seasonally differenced square-root transformed series.

From Figures 5 and 6, the plots are within the confidence bounds, an indication of a stationary series except for the spikes at seasonal lags implying the presence of strong seasonal components in the transformed data. Theoretically to check for presence of stationarity we performed some formal tests of stationarity to confirm the visual inspection of the correlograms. The unit root test performed was the Augmented Dickey Fuller test. Upon using R software, we obtained the parameter $\emptyset = -9.4757$, lag $= 12$, $P = 0.1$.

According to ADF test, when $\emptyset < -2.86$ it indicates stationarity. Therefore, this confirms the presence of stationarity in the transformed data. Upon achieving stationarity both in variance and mean, we now proceed to the first stage of Box-Jenkins methodology which is model identification.

After considering many possible models, the model **SARIMA(0,0,0)(0,1,2)$_{12}$** has significant parameters and the lowest AIC and BIC values. In this equation, s=12, θ_1= -1.9922 and its standard error is 0.0670, θ_2= 0.9999 and its standard error is 0.0735, Z_t is white noise. To validate the model, the diagnostic test results used were as seen in Figure 7 and Table 1.

From Figure 7, the plotted standardized residuals versus time exhibits no pattern, revert around mean zero and variance one. The plotted ACF of residuals versus lag lies within the confidence interval exhibiting stationarity. The Q-Q plots follows the normal line indicating the residuals are normally distributed. Finally , the p-values for the Ljung-Box test are above 0.05% for all lag orders showing there is no significant departure from white noise for the residuals. Also for normality, we used the Histograms to verify. For the model, the histogram is bell-shaped indicating normality.
In summary the plots indicate the model fits the data or is suitable for the data.

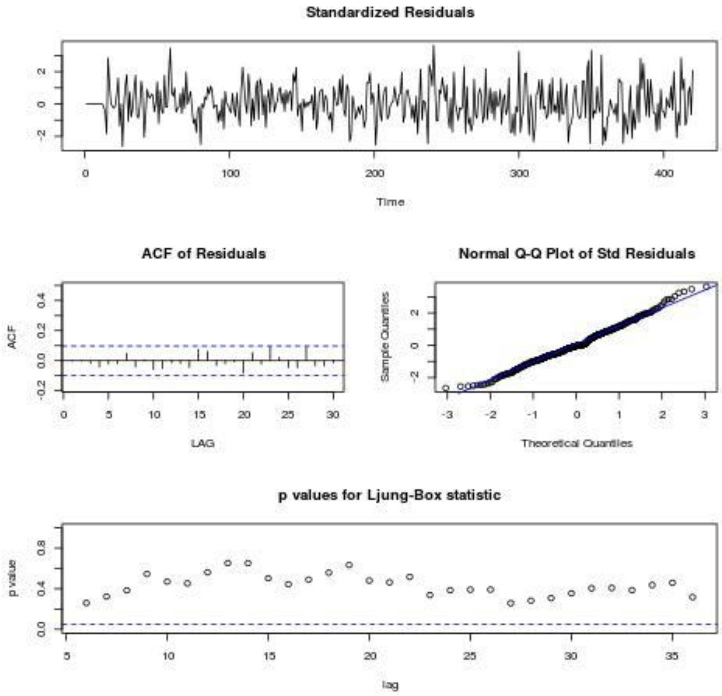

Figure 7: (0, 0, 0) (0, 1, 2)$_{12}$ model residuals

To confirm the visual inspection of Figure 7 we carry out formal diagnostic tests . We use Runs test, turning point test and Shapiro wilk test. The output of the tests are presented in Table 1.

Table 1: Diagnostic test results

Model		Runs test	Turning Point test	Shapiro Wilk test
(0,0,0)(0,1,2)$_{12}$	Test statistic	-0.8552	-0.8647	0.9932
	Tabulated value	1.96	1.96	0.996
	p-value	0.03924	0.03872	0.05541

The model's residuals are iid, white noise, homoscedastic, normally distributed and with mean zero hence it is adequate for modeling the square-root transformed monthly rainfall series of Kapsoya Meteorological station in Uasin Gishu county. The fitted model **SARIMA(0,0,0)(0,1,2)$_{12}$** is given as;

$$\mathbf{Y_t = Y_{t-12} + 1.9922Z_{t-12} + 0.9999Z_{t-24} + Z_t \ldots\ldots (10)}$$

In this study, the model selected for forecasting is the Seasonal ARIMA (0,0,0)(0,1,2)$_{12}$.. The results of the forecasting for the year 2014, 2015 and 2016 observations are as shown in Table 2.

Table 2: Actual and Forecasted values

Date	Actual 2014	Forecast 2014	Residual 2014	Forecast 2015	Forecast 2016
January	32.5	37.57	-5.07	30.66	30.59
February	46.65	45.34	1.31	36.86	36.76
March	70.72	73.38	-2.66	69.55	69.48
April	130.1	123.53	6.57	128.10	128.46
May	121.2	126.32	-5.12	128.10	128.10
June	94.47	91.55	2.92	98.41	98.41
July	176.00	171.86	4.14	166.46	166.46
August	151.00	157.84	6.84	153.61	153.61
September	80.375	76.94	4.43	82.37	82.37
October	61.025	59.98	1.045	74.53	74.53
November	68.87	70.85	-1.98	62.45	62.45
December	38.875	36.06	2.815	47.65	47.65

A plot of all forecasts is shown in Figure 8.

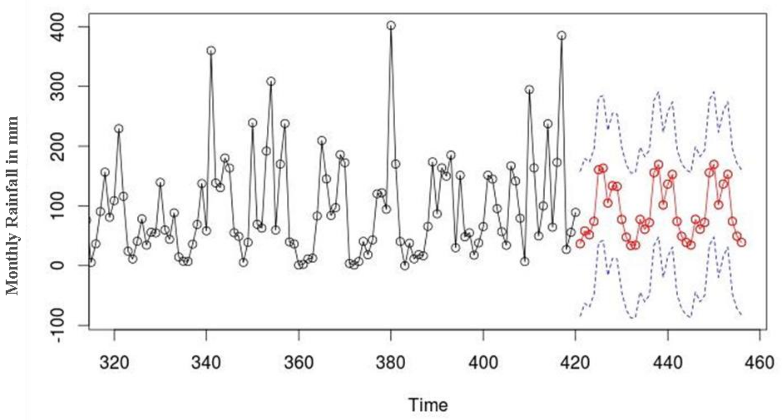

Figure 8: Forecast plot for total monthly rainfall of Kapsoya

From the forecast plot, 1977 had the highest rainfall recorded. Heavy rains were also recorded in 1982, 1998 and 2012. On analysis using R computer software, we conclude that heavy rains occur at an approximate interval of 38 months. On this basis since the last heavy rainfall recorded was in April 2012, it is expected that the next heavy rainfall value will be recorded around June-August 2015.

The lowest rainfall was recorded in 1984. Other years with low rainfall were 1986, 1993, 2000 and 2009. According to this trend, it is observed that droughts occur at intervals of 6-8 years and occur in mainly the first three months of the year. The next dry period is expected around 2015-2017 especially in the first three months.

To evaluate the forecasting performance of this model, we used the Mean Absolute Percentage Error (MAPE). The MAPE of forecasting using this model is 7.78% which is less than 10%. This shows that the model's error is minimum hence it should be adopted.

IV. Conclusion

In this study, the univariate Box-Jenkins methodology is used to fit an ARIMA model on historical rainfall data of Uasin Gishu county of 456 months from January 1977 to December 2014. The best ARIMA model that best describes the monthly rainfall data is SARIMA $(0,0,0)(0,1,2)_{12}$, also written as $Y_t = Y_{t-12} + 1.9922Z_{t-12} + 0.9999Z_{t-24} + Z_t$. The model is used to forecast average monthly rainfall values for 24 months i.e January 2015- December 2016. However, the model should continuously be updated for accurate forecasting after December 2016.

V. Recommendations

From our results and discussions, we recommend to stakeholders of Uasin Gishu county that multiple univariate Box-Jenkins time series models should be developed for more representation of the study area. Multivariate time series models should also be developed which will put into consideration other variables affecting climate apart from rainfall.

Acknowledgement

My gratitude goes to the Mathematics Department of MMUST for giving me the opportunity to pursue my academic work. My family's support, encouragement, perseverance, concern and love cannot be overlooked. I am forever thankful to the Almighty God for His grace, providence and blessings which have throughout my life given me the courage to move forward. God bless you all.

References

[1]. Box,G. & G.Jenkins, (1976). Time series analysis: forecasting and control. (Revised ed) Holden day, San Francisco.
[2]. Brockwell,P.J. & Davis, R.A. (1996). Introduction to Time series and forecasting. (2nd ed.). New York: Springer.
[3]. Chatfield, C.,(1991). The Analysis of Time Series. An Introduction, (4th ed)., Chapman and Hall: London.
[4]. Chatfield,C. (1996). The analysis of time series- an introduction,(5th ed) ,Chapman and Hall, U.K.
[5]. Cromwell, J.B., Labys, W.C and Terraza, M., (1994). Univariate Tests for Time Series Models. A Sage Publication, (7-99)96., London.
[6]. Diebold, F.X., Kilian, L. and Nerlove, M., (2006). Time Series Analysis. Working Paper No.06-011, University of Maryland, College Park.
[7]. Enders, W., (1995). Applied Econometric Time Series, John Wiley and Sons, Inc., New York.
[8]. Enock, M.A., Buckman, A.& Seth, O.L (2013); Time series modeling of rainfall in New Juaben municipality of the Eastern region of Ghana. The special issue on contemporary research in business and social science.4(8)1-14.
[9]. Ette, H.E and Tariq M.M (2014); Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by sarima methods. International Journal of Scientific Research in Knowledge.2(7)320-327.
[10]. Fatemeh F., Hossein N. & Mahmood K. (2013); Studying of drought, modeling and forecasting the precipitation of Shiraz city in Iran. Archives of Applied science research. 5 (3)173-183.
[11]. Gibbons, R.D.,(1994). Statistical Methods for Ground water Monitoring. John Wiley & Sons, New York.
[12]. Janacek,G. and Swift, L., (1993). Time Series Forecasting, Simulation, Application, Ellis Horwood, New York, USA.
[13]. Jennifer, M.A (2011); Seasonal prediction of African rainfall with focus on Kenya. Phd thesis, submitted to the Department of space and climate physics. University college London.
[14]. Mahsin, M.D.,Yesmin, A. & Monira, B. (2012); Modelling Rainfall in Dhaka division of Bangladesh using Time series analysis. Journal of mathematical modeling and application.1(5) 67-73.
[15]. Meenakshi, S. S and Lakshmi, M. (2014); Rainfall Prediction using Seasonal Auto Regressive Integrated Moving Average model. Indian Journal of Research 3(4)58-60.
[16]. Montgomery, D.C. & Johnson, L.A. (1976); Forecasting and Time series Analysis. New York: Wiley.
[17]. 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.
[18]. Ndetei, C.J. (2013); Use of polynomial fit to predict seasonal rainfall in Nairobi, Kenya.University of Nairobi, Kenya. Msc thesis, submitted to Department of Meteorology UoN. Unpublished.
[19]. Ngaira, Josephine k. (2005); Implications of climate change on the management of Rift Valley lakes in Kenya. The case of Lake Baringo. Proceedings of the 11th world Lakes conference, Nairobi. Kenya. 2(1)133-138.
[20]. Osabuohien & Irabor, O. (2013); Applicability of Box-Jenkins SARIMA model in rainfall forecasting; port Hacourt south Nigeria. Canadian journal in computing mathematics, Natural sciences, Engineering and Medicine.4(1)1-4.
[21]. Raphael .W. K (2012); Some aspects of the geography of Kenya. Moi University. Unpublished.
[22]. Rediat, T. (2012); Statistical analysis of rainfall pattern in Dire Dawa, Eastern Ethiopia using descriptive analysis, cross spectral analysis and univariate Box-Jenkins method. Msc thesis submitted to the Department of statistics, University of Addis Ababa, Ethiopia.
[23]. Robert Mburia, (2012); Impacts of climate change on African economies. Case study, Kenya. Climate Emergency Institute, climate science Library.
[24]. Sawsan, M.A (2013); Time Series Analysis of Baghdad Rainfall using ARIMA method, Iraqi Journal of Science, 54(4)1136-1142.
[25]. Sharma, m.k. & Amaha, G. S. (2011); Modelling and forecasting of rainfall data of Mekele for Tigray region. A journal on Statistics and application. 9(1)31-53.
[26]. Tsay,.S.R, (2005). Analysis of Financial Time Series; 2nd ed. John Wiley & Sons, Inc., Hoboken.