Modelling Rainfall Patterns in Meru and Embu Regions using Time Series Models

Chepkoech Carllen*, Moses M. Manene**, Muriungi Robert*

*Meru University of Science and Technology, P.O. Box 972-60200, Meru, Kenya
** University of Nairobi, P.O. Box 30197-00100, Nairobi, Kenya

***Abstract*:-Rainfall is the meteorological phenomenon that is useful for human activities. Majority of population depend on rainfall water for agriculture and domestic use. Since Meru and Embu regions are agricultural zones relying heavily on rainfed agriculture, it is important for farmers to know rainfall patterns prevailing in their regions. The objective of this study was to model rainfall patterns in Meru and Embu regions. Stationarity and unit root for data were tested, time series model was developed and fitted to the historical data using Box-Jenkins (BJ) Methodology and rainfall in the regions were forecasted for five years. Monthly and yearly rainfall data obtained from Kenya meteorological department for the period 1976-2015 was used in the study. This information can be used in planning and management of water for domestic and agricultural use in the regions. Rainfall data was found to be seasonally and non-stationary and hence differencing and seasonal differencing was applied to achieve stationarity. Rainfall in both regions is bimodal, it has short rains in the months of October to December (OND) and long rains in the months of March to May (MAM). The model that best fitted rainfall data was ARIMA $(1,1,1)(0,1,1)_{12}$. This model was used to forecast monthly rainfall patterns for five years and found that future rainfall patterns will not change with time. It was recommended that, future researchers should consider zoning regions and apply developed ARIMA model and negative binomial to homogenous zones.**

***Key Words*: Stationarity; Rainfall patterns; Bimodal**

I. INTRODUCTION

Rainfall the natural source of water, has greater impact on agricultural activities and domestic consumptions. Water can be used; to aid transportations, as a source of power, for domestic consumptions and in agriculture for irrigation. Poor rainfall in Kenya as indicated by recent severe drought that has caused many deaths for both humans and animals due to reduced food security and reduction in water resources for domestic use, drinking and sanitation as reported by Kenya Meteorological Department (2017). Food security situation is wanting in arid and semi-arid counties in the country. There are indications that the situation is likely to worsen further by January 2017 as reported by (National Drought Management Authority (NDMA) and Kenya Meteorological Department (KMD)) (2017). In most areas sources of water has dried up forcing people to walk for long distances looking for water, in both arid and semi-arid areas. This is due to the results of pollution, deforestation, fast growing populations and changing climate, UNICEF (2017).

In agricultural sector, water problem is the most critical constraint to food productions. In Kenya for example, where farmers practice on small scale, rainfall variability has caused hunger in many regions. Scarcity of water is a severe environmental constraint to plant productivity. Drought induced loss in crop yield, probably exceeds losses from all other causes, since both the severity and duration of the stress are critical, Farooq (2008). Analysis of rainfall data for long periods provides information about rainfall variability and to better manage agricultural activities, Nyatuame (2014). In Kenya, a high percentage of population depend on agriculture directly or indirectly.

Correct prediction or forecasting of future rainfall will contribute highly to the management of water resources and play a major role in boosting agricultural sector since farmers will be able to plant plenty of food crops during rainy seasons, Oyamakin (2010).

Data quality control was undertaken on the historical rainfall data for the two regions. The study focuses on variability of rainfall as a major factor affecting agriculture and people who live in both regions. The data was obtained from Kenya Meteorological Department. Kenya Meteorological Department is a government research parastatal tasked with data management, climate change, research and development and economic policy.

Both Meru and Embu regions are surrounded by dry areas, have high altitudes and are at the foot of Mt. Kenya. Also both regions depend much on rainfall water for agriculture in order to feed the growing population.

For time series data, the larger the data set the better for: Trend observation, Seasonal comparison and Random effect identification. It takes into consideration the monthly and yearly amount of rainfall in both regions. Time series model was built and fit to the data. Analysis on the monthly rainfall datasets was carried out to determine the evidence of rainfall change in the two regions.

Muthama and Manene (2008) used stepwise regression technique to analyze irregularly distributed rainfall events in time. Their study sought to improve existing rainfall monitoring and prediction in Nairobi. According to them, it can be deducted that the 4th degree polynomial function can be used to predict the peak and the general pattern of seasonal rainfall over Nairobi, with acceptable error values. The information can be used in the planning and management of water resources over Nairobi. The same information can be extended to other areas.

Matiur, Shohel, Sazzad and Naruzzuman (2015) carried out analysis of rainfall data, they further developed ARIMA model that was applied in forecasting monthly precipitation for the next three years to take proper decision on water development management Authority.They applied AIC, MSE, MAPE and MAD to test accuracy and applicability of ARIMA model at different stages.

Javari and Majida (2016) examined trend and homogeneity through the analysis of rainfall variability patterns in Iran. The study presents a review on application of homogeneity and seasonal time series analysis methods for forecasting rainfall variation.They studied homogeneity of monthly and annual rainfall at each station using ACF and Von Neumann(VN) test at significance level of 0.05. Their results indicate that the seasonal patterns of rainfall exhibited considerable diversity across Iran. The present study comparisons among variations of patterns with seasonal rainfall series and revealed that the variability of rainfall can be predicted by the non-trended and trended patterns.

In this research, we developed a time series model used to forecasting rainfall patterns for five years, using monthly and yearly amount of rainfall for the period 1976-2015.

II. MATERIALS AND METHODS

Monthly and yearly rainfall that was used in this study was obtained from Kenya Meteorological Department (KMD. The data was for two regions Embu and Meru. The data covers a period of forty (40) years from (1976-2015).It is important to investigate the homogeneity of records of meteorological data before using it in any analysis. In this study single mass curve was used to determine whether the data being used is homogeneous.

Stationarity of the data helped the researcher in building a time series model. Stationarity of the data was tested using ACF and PACF. Unit root test was performed using Augmented Dickey –Fuller (ADF) test. Autocorrelation Function drops to zero quickly for stationary time series, while for non-stationary data it decreases slowly. Unit root test is the statistical hypothesis test for stationarity that is designed for determining whether differencing is required.

In the augmented Dickey-Fuller test, the following regression model was estimated;

$$y'_t = \phi y_{t-1} + \beta_1 y'_{t-1} + \beta_2 y'_{t-2} + \ldots + \beta_k y'_{t-k} \qquad (2.1)$$

Where y'_t denotes the first differenced series, $y'_t = y_t - y_{t-1}$ and k is the number of lags to include in the regression. If original series y_t needs differencing, the coefficient $\hat{\phi}$ should be approximately zero. If y_t is already stationary then $\hat{\phi} < 0$.

H_0 =datasets are not stationary, versus

H_1 =datasets are stationary

Large p-values are indicative of non-stationarity while small p-values suggest stationarity. Using 5 percent (%) threshold, differencing is required if p-value is greater than 0.05.

To study the patterns of the rainfall data, time series models were fitted after determining the nature of the data using both ACF and PACF these statistical measures reflect how the observations in a time series are related to each other. For modelling purpose it is often useful to plot the ACF and PACF against consecutive time lags. These plots help in determining the order of AR and MA models. In this study the Box-Jenkins (BJ) Methodology was used to build a time series model. This methodology applies autoregressive moving average ARMA or ARIMA models to find the best fit of a time series to past values of this time series. This approach possesses many appealing features. To identify a perfect ARIMA model for a particular time series data, the following four phases are used viz.

i) Model identification.
ii) Estimation of model parameters.
iii) Diagnostic checking for the identified model appropriateness for modelling.
iv) Application of the model.

The first step in developing a Box–Jenkins model is to determine if the time series is stationary and if there is any significant seasonality that needs to be modelled. The data was examined to check for the most appropriate class of ARIMA processes through selecting the order of the consecutive and seasonal differencing required making 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. The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are the most important elements of time series analysis. The ACF measures the amount of linear dependence between observations in a time series that are separated by a lag k. The PACF plot helps 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, Box and Jenkins (1970).

The ARIMA model is the generalization of ARMA model that can only be used for stationary time series data. An ARMA (p, q) model is a combination of AR (p) which is given by:

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-1} + \varepsilon_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t \quad (2.2)$$

and MA (q) model which uses past errors as the explanatory variables. The MA (q) model is given by

$$y = u + \sum_{j=i}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t = u + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2.3)$$

Models and is suitable for univariate time series modelling. Autoregressive (AR) and moving average (MA) models can be effectively combined together to form a general and useful class of time series models, known as the ARMA models. Mathematically an ARMA (p, q) model is represented as:

$$y = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i y_{i-t} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \quad (2.4)$$

Here the model orders p, q refer to p autoregressive and q moving average terms.

Usually ARMA models are manipulated using the lag operator notation. The lag or backshift operator is defined as

$$Ly_t = y_{t-1}. \quad (2.5)$$

Time series, which contain trend and seasonal patterns, are also non-stationary in nature. Thus from application view point ARMA models are inadequate to properly describe non-stationary time series, which are frequently encountered in practice. For this reason the ARIMA model is proposed, which is a generalization of an ARMA model to include the case of non-stationarity as well. In ARIMA models a non-stationary time series is made stationary by applying finite differencing of the data points. The mathematical formulation of the ARIMA (p, d, and q) model using lag polynomials is given below:

$$\varphi(L)(1-L)^d y_t = \theta L \varepsilon_t \quad (2.6)$$

i.e.

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1-L)^d y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)\varepsilon_t \quad (2.7)$$

Where, p, d and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. The integer d controls the level of differencing. Generally d=1 is enough in most cases. When d=0, then it reduces to an ARMA (p, q) model. An ARIMA (p, 0, 0) is nothing but the AR (p) model and ARIMA (0, 0, q) is the MA (q) model. ARIMA (0, 1, 0),

i.e.

$$y_t = y_{t-1} + \varepsilon \quad (2.8)$$

is a special one and known as the Random Walk model.

In this study the data was tested for stationarity and the model was chosen depending on whether the data is stationary or non-stationary. If the data is non-stationary differencing would be done. If stationarity is not achieved after first differencing then the second differencing is carried out and Plots are expected to be within the confidence bounds which is an indication of stationarity. If after second differencing there are some spikes outside the confidence bounds, it confirms the presence of strong seasonality components in the transformed data. . AIC and BIC was considered when choosing the best model.

III. RESULTS AND DISCUSSION

The results for homogeneity test, stationarity and unit root test and time series model fitted to monthly and yearly rainfall data is as depicted below.

3.1Homogeneity

Homogeneity test was tested using single mass curve to ascertain good quality of rainfall records. The plots showed that rainfall data for both regions was homogenous and can be used for further analysis. The single mass curves are shown figure 1 and 2 below.

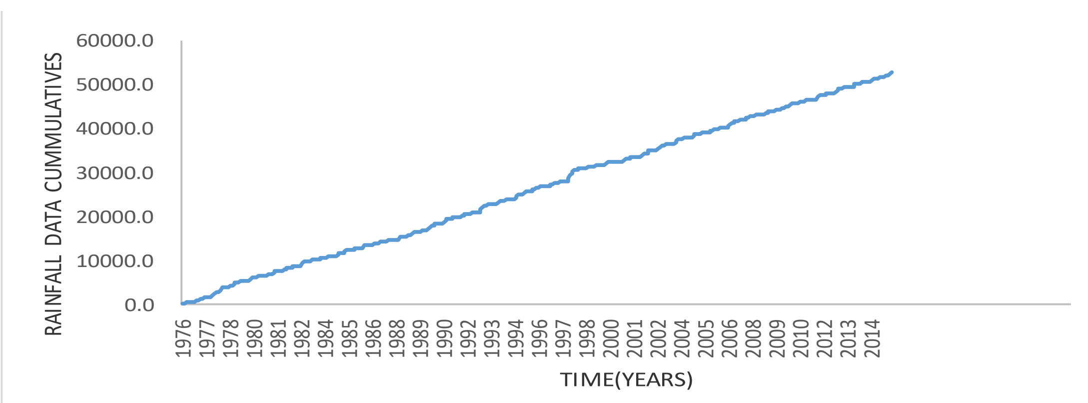

Figure 1: Testing homogeneity of Embu rainfall data

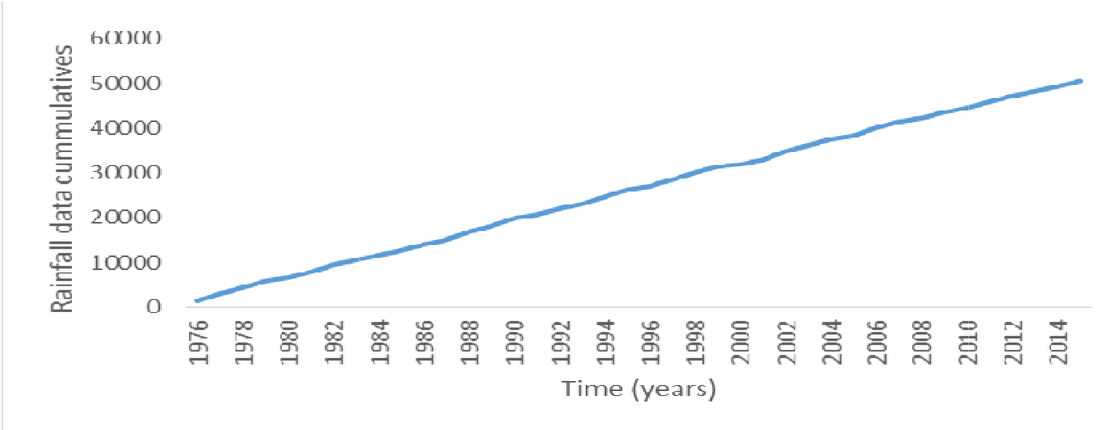

Figure 2: Testing homogeneity of Meru rainfall data

3.2 Test for Stationarity and Unit Root Test

Stationarity in variance and mean is a requirement for a time series before a model is fit on data. On studying monthly rainfall data for both regions, we observe that the peaks of time plots are not repeated with the same intensity indicating a non –constant variance and hence the series lack stationarity in variance. To verify stationarity in mean, we check by correlograms as shown in the figure 7 and figure 8 below.

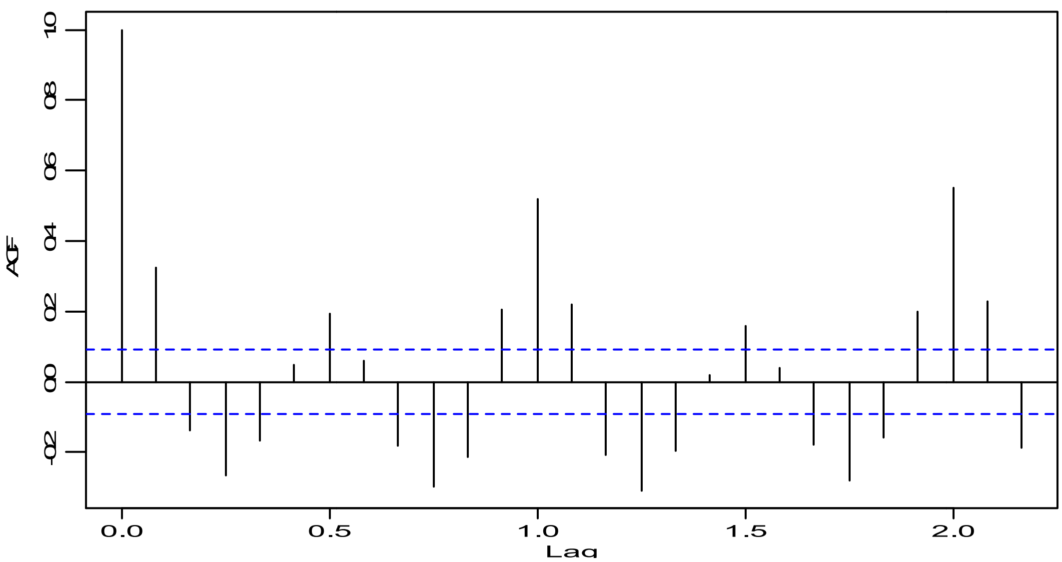

Figure 3: ACF for Meru rainfall data

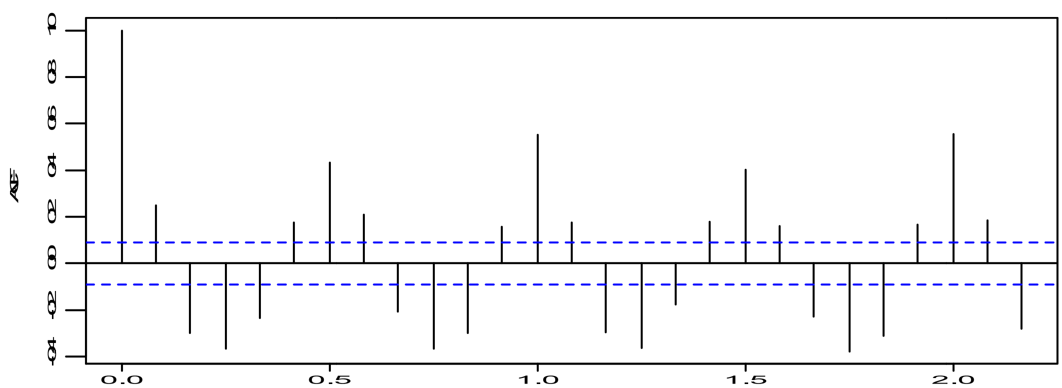

Figure 4: ACF for Embu rainfall data

ACF's are sinusoidal at the multiples of seasonal lags indicating the presence of strong seasonality behaviour. However as the lags increase the autocorrelations at multiples of seasonal lags seems not to decay implying non-stationarity in seasonal component of monthly rainfall data.

The unit root test that was performed was augmented Dickey-Fuller test. The results were obtained using R-software as presented below.

Augmented Dickey-Fuller Test

The following results was obtained

data: Meru rainfall data

Value of test-statistic is: -10.7941

Table 1: Critical values for test statistics

	1%	5%	10%
Tau	-2.58	-1.95	-1.62

Augmented Dickey-Fuller Test

data: Embu rainfall data

Value of test-statistic is: -11.1496

Table 2 : Critical values for test statistics

	1%	5%	10%
Tau	-2.58	-1.95	-1.62

Using the results from table 1 and 2, the statistics exceeds the critical values hence the series is not stationary. We conclude that there is a unit root.

3.3 Time series model

Time series model developed using Box-Jenkins methodology was fitted to the rainfall data of Embu and Meru regions. The specific aim was to obtain appropriate order of ARIMA model. To select an appropriate order of seasonal ARIMA, ACF and PACF graphs were used

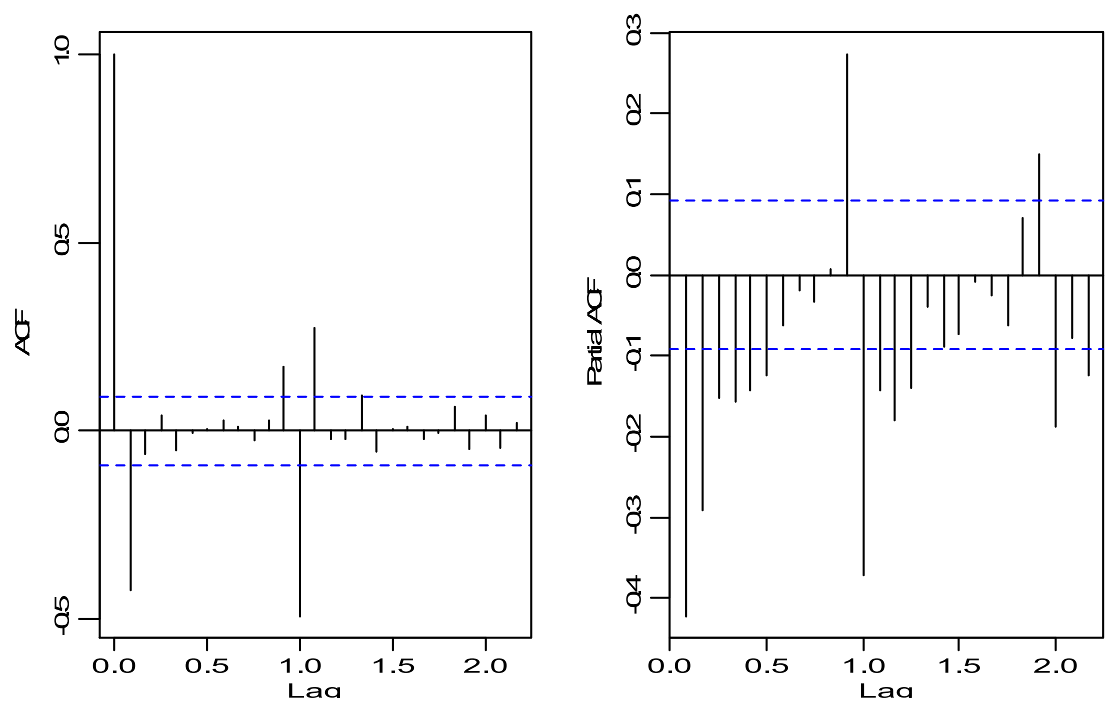

Figure 5: ACF and PACF Plots of differenced and Seasonally Differenced Square root Seriesfor Meru rainfall data

Meru rainfall data had a significant spikes at lag 1, suggesting a non seasonal MA(1) component, and significant spikes at lag 11, 12 and 13 suggesting seasonal MA(1) in the ACF. The ARIMA model was, ARIMA(0,1,1)$(0,1,3)_{12}$. The model had non seasonal and seasonal moving average.

Using Akaike information criterion, ARIMA models of different orders were tested. This enabled the best model with lowest Akaike information criterion to be chosen. A summary of the result was presented in table 3 below.

Table 3: Identified ARIMA models for Meru

Model	AIC
ARIMA(0,1,1)$(0,1,3)_{12}$	2505.16
ARIMA(0,1,2)$(0,1,2)_{12}$	2485.53
ARIMA(1,1,1)$(0,1,1)_{12}$	2480.91
ARIMA(0,1,1)$(0,1,1)_{12}$	2502
ARIMA(1,1,2)$(1,1,0)_{12}$	2624.16
ARIMA(1,1,0)$(0,1,3)_{12}$	2607.72

Based on AICs' of the models, ARIMA(1,1,1) $(0,1,1)_{12}$ was chosen as the best model.

Call:

Arima(x = data13, order = c (1, 1, 1), seasonal = list (order = c (0, 1, 1), period = 12))

Coefficients:

ar1 ma1 sma1

0.2256 -1.0000 -0.9998

s.e. 0.0458 0.0136 0.0332

Sigma^2 estimated as 11.91: log likelihood = -1236.45, AIC = 2480.91

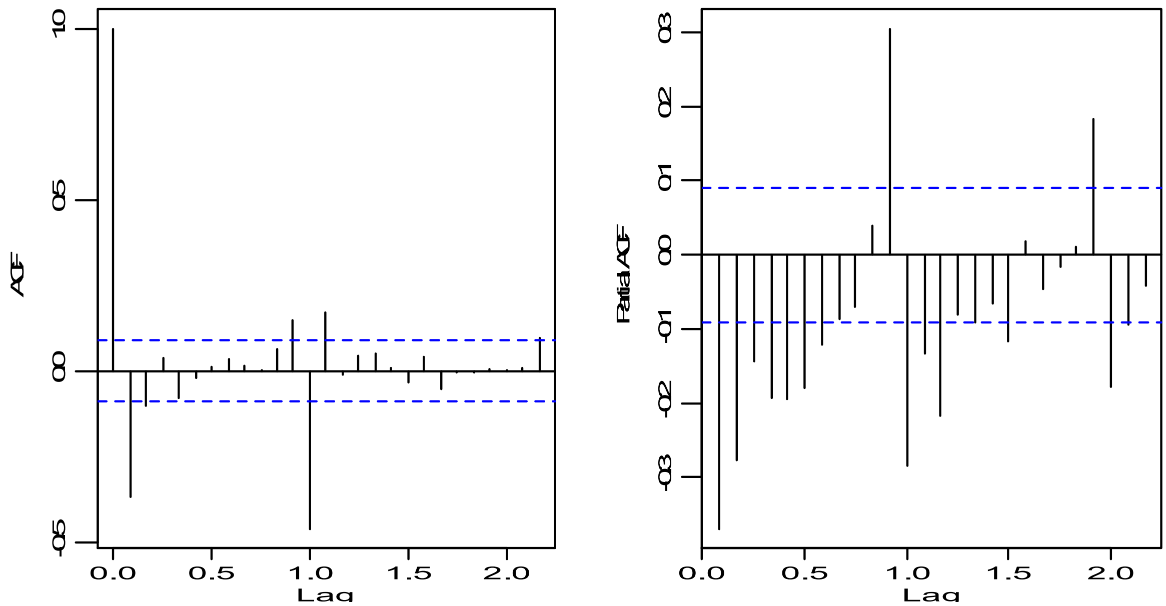

Figure 63: ACF and PACF Plots of differenced and Seasonally Differenced Square root SeriesEmbu rainfall data

Call:

Arima(x = data23, order = c (1, 1, 1), seasonal = list (order = c (0, 1, 1), period = 12))

Coefficients:

ar1 ma1 sma1

0.1977 -1.0000 -1.0000

s.e. 0.0454 0.0115 0.0275

Sigma^2 estimated as 11.43: log likelihood = -1258.89, aic = 2525.79

Embu rainfall data had spikes at lag 1, 11, 12 and 13 in the ACF .The ARIMA model was, ARIMA$(0,1,1)(0,1,3)_{12}$. The model had non seasonal and seasonal moving average.

Having ARIMA$(0,1,1)(0,1,3)_{12}$ as the initial model, ARIMA$(0,1,2)(0,1,2)_{12}$, ARIMA$(1,1,1)(0,1,1)_{12}$, ARIMA$(0,1,1)(0,1,1)_{12}$, ARIMA$(1,1,2)(1,1,0)_{12}$ and ARIMA$(1,1,0)(0,1,3)_{12}$ were fitted. The best model with lowest Akaike information criterion was chosen. A summary of the result was presented in table 4 below.

Table 4: Identified ARIMA models for Embu

Model	AIC
ARIMA$(0,1,1)(0,1,3)_{12}$	2543.12
ARIMA$(0,1,2)(0,1,2)_{12}$	2526.75
ARIMA$(1,1,1)(0,1,1)_{12}$	2525.79
ARIMA$(0,1,1)(0,1,1)_{12}$	2805.94
ARIMA$(1,1,2)(1,1,0)_{12}$	2679.19
ARIMA$(1,1,0)(0,1,3)_{12}$	2678.57

Based on AICs' of the models, ARIMA(1,1,1) $(0,1,1)_{12}$ was chosen as the best model.

The model parameter were significant from table 3 and 4, hence our proposed model was justified. After considering very many models, the model ARIMA$(1,1,1)(0,1,1)_{12}$ had significant parameters and lowest AIC values.

3.3.1 Model Diagnostic Checking

The models having been identified and the parameters estimated, diagnostic checks were applied to fitted models for monthly rainfall data in Embu and Meru region.

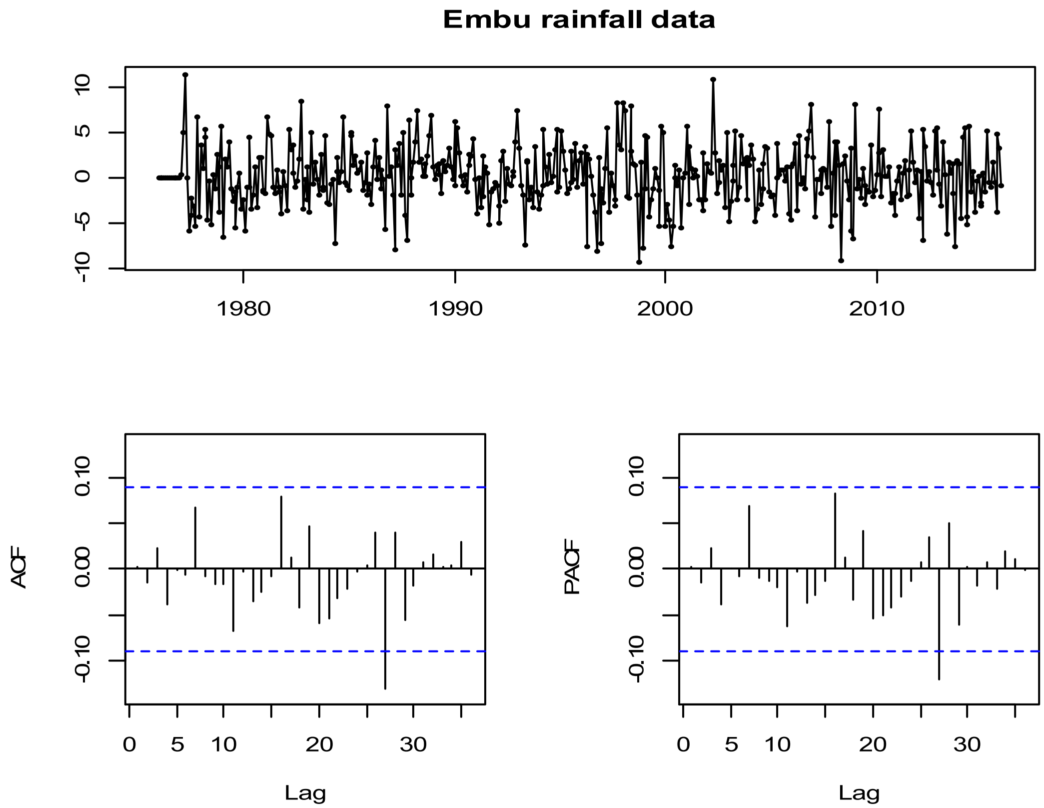

Figure7:ARIMA(1, 1, 1) (0, 1, 1)$_{12}$model residuals for Embu rainfall data

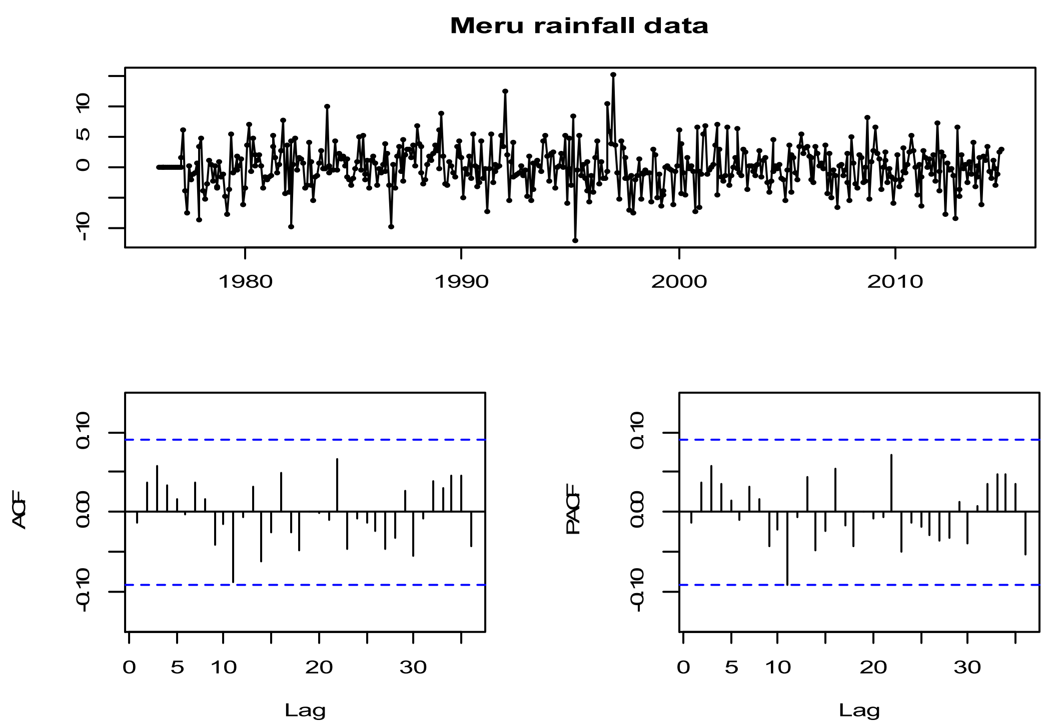

Figure 8:ARIMA (1, 1, 1) (0, 1, 1)12 model residuals for Meru rainfall data

All the spikes were within the significant limits and so the residuals appeared to be white noise.

3.4 Forecasting

EMBU

After identifying the model, rainfall distribution patterns in both regions were forecasted using the model. Since our data was not stationary, this property was an advantage since the data had predictable patterns in the long term.

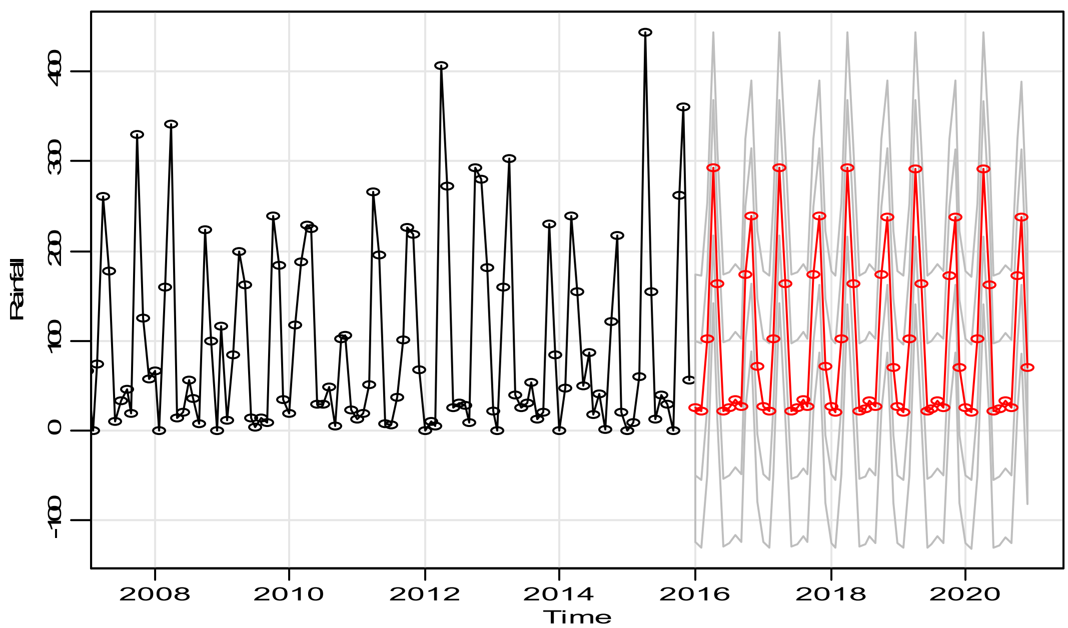

Figure 9: Forecast plot for total monthly rainfall of Embu

From the figure above, the series were followed by the forecast as the red line and the upper and lower predictions limit as grey were shown. Forecasts from the ARIMA (1, 1, 1) $(0,1,1)_{12}$ model for the next five years were shown in figure 17, forecast followed the trend due to double differencing.

Prediction intervals showed that the rainfall could start decreasing or increasing in time.

High rainfall had a cycle of four years in Embu, it occurred in 2008, 2012 and 2016 and was expected to occur in 2020.

MERU

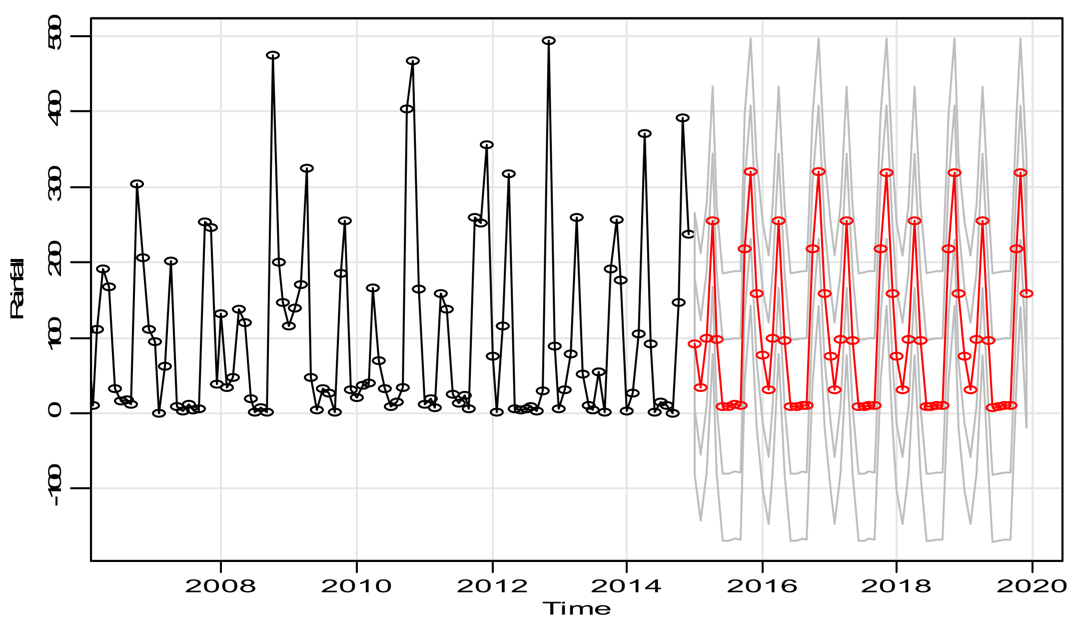

Figure 40: Forecast plot for total monthly rainfall of Meru

The figure above showed the series followed by the forecast as the red line and the upper and lower predictions limit as grey. Forecasts from the ARIMA(1, 1, 1)$(0,1,1)_{12}$ model for the next five years were shown in figure 18. The increasing and decreasing prediction intervals showed that the rainfall could start decreasing or increasing any time.

High rainfall was recorded after every two years from 2009, 2011, 2013 and 2015. Also from the analysis, rainfall was expected to be high in 2017.

IV. CONCLUSION

Rainfall pattern in Meru and Embu regions significantly changed over time. There were periods of low variability and others of extreme variability separated by periods of transition. Rainfall in both regions had long rains in March, April and May (MAM) and and short rainsin October, November and December (OND). It appeared that short rains had high amount of rainfall as compared to long rains.

From the stationarity test, rainfall data for both regions was found to be non stationary due to presence of rainfall trends and seasonality.

From the fitted ARIMA(1,1,1)$(0,1,1)_{12}$ model, there was high rainfall in Meru after a period of 2 years while Embu had a period of 4 years. Forecasted rainfall shows that, the increasing and decreasing prediction intervals showed that the rainfall could start decreasing or increasing any time.

From this study, it can be concluded that, rainfall patterns for Embu and Meru regions would be change over time.

REFERENCES

[1]. Abu, G.A, Ater,P.I. & Tor, I.E. (2015). Long run co integration modelling for rainfall,yield and pricesof cerials and legume in Nigeria. *Asian Journal of Agricultural Extension,Economics and Sociology.*, 86-100.

[2]. Box , G.E.P & Jenkins, G.M. (1970). Time seres analysis : Forecasting and control. San Francisco, CCA: Holden-Day, 575 P.

[3]. Chikobvu, D.& Retius, C. (2015). Modelling of extreme minimum rainfall using generalized extreme value distribution for Zimbabwe. *South African Journal*, 8.

[4]. Farooq, M., Wahid, A., Kobayashi, N., Fujita, D. & Basra, S.M.A. (2008). Plant drought stress:effects,mechanism and management. *INRA,EDP sciences*, 185-212.

[5]. Javari, K.& Majida, F. (2016). Trend and Homogeneity Analysis of Precipitation in Iran. *Climate 2016*, 4(3);44.

[6]. Kenya Meteorological Department (2017, Feb 13). The outlook for the 2017 "long rains" (March-April-May) 2017 season; Review of weather during the October-December 2016 "short rains" season and January 2017 [press release]. Retrieved from, http://www.meteo.go.ke/pdf/seasonal.pdf

[7]. Matiur, R. M., Shohel, R., Sazzad, M. H. & Naruzzuman, S.M. (2016). modelling monthly precipitation in faridpur region of bangladesh using ARIMA. *Journal of enviromental science,Toxicology and Food technology*, 22-29.

[8]. Muthama, N.J. & Manene M. (2008). Simulation of Decadal Precipitation over Nairobi in Kenya. *SQU Journal For Science*, 43-54.

[9]. Naill, P. E.& 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.

[10]. Nyatuame, M., Owosu, V. -G. & Amplaw, F. (2014). Statistical Analysis of Rainfall Trend for Volta Region in Ghana. *International Journal of Atmospheric Sciences*, 11.

[11]. Oyamakin, S. O, Ayoola,F.J. & Dare, O.T. (2010). Time series analysis of rainfall and temperature in South West Nigeria,Pacific. *Journal of sciences and Technology*, 552-564.

[12]. Wiredu, S., Naisiru, S., & Yeboah, A.G. (2013). Proposed Seasonal Autoregressive Integrated Moving Average Model for Forecasting Rainfall Pattern in the Navrongo Municipality, Ghana . *Journal of Environment and Earth Science*