

Modelling Rainfall Patterns in Meru and Embu Regions Using ARIMA Models

Chepkoech Carllen[1*] Gitunga Muriungi Robert[2]

1. School of Pure and Applied Sciences, Meru University of Science and Technology, PO box 972-60200, Meru, Kenya.
2. School of Pure and Applied Sciences, Meru University of Science and Technology, PO box 972-60200, Meru, Kenya.

* E-mail of the corresponding author: ccarleen06@gmail.com

Abstract

Rainfall is the natural source of water, it has greater impact on agricultural activities and domestic consumptions. 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.In this study we model rainfall patterns in Meru and Embu regions of Kenya using Monthly and yearly rainfall data. ARIMA model was developed using Box-Jenkins (BJ) Methodology and fit to monthly and yearly average amount of rainfall. The data were examined to check for the most appropriate class of ARIMA processes. This was done by selecting the order of the consecutive and seasonal differencing. The auto-correlation function (ACF) and the partial autocorrelation function (PACF) are the most important elements of time series analysis. Using the AIC criterion ARIMA $(1,1,1)(0,1,1)_{12}$ was identified as the best model. This model was used to forecast monthly rainfall patterns for five years and found that future rainfall patterns will change with time as contributed by many factors. It was recommended that, future researchers should consider zoning regions, identify other factors contributing to change in rainfall patterns and apply developed Arima model.

Keywords: ARIMA model, Box-Jenkins methodology, forecasting, seasonal differencing.

1.0 Introduction

Research in rainfall patterns and variability has largely contributed in management of water for domestic and agricultural use. Lack of water due to poor rainfall is a major threat as indicated by recent severe drought in Kenya. Agricultural practices need to be dynamic to cope with changing weather conditions and for states to feed the growing population. This research modelled rainfall patterns and forecasted rainfall in Meru and Embu region of Kenya using time series models. This helped meteorologists, agricultural experts and government on how to advise farmers to prepare to curb droughts and adapt new methods of farming.

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 are small scale holders, the current low levels of rainfall 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, 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).

This study sought to identify rainfall seasons in both regions, and carried out detailed statistical analysis on the historical datasets to determine evidence of rainfall change. 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 study sought to understand better the changing patterns in monthly rainfall and investigate effects of time series components on rainfall patterns.

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. It gives an early warning in weather forecasting that entails provision of timely and effective weather information that allows individuals, organizations, or communities exposed to likely weather hazards to take actions that avoids or reduces their exposure to risk.

Both Meru and Embu regions are surrounded by dry areas, have high altitudes and are at the foot of Mt. Kenya. Also both regions depends 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.

2. Literature Review

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 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 explored how generalized linear models can be used in forecasting future amount of rainfall for five years, using monthly and yearly amount of rainfall for the period 1976-2015 and developed a time series model used to study the distribution of rainfall patterns.

3. Model Identification and Selection

Time series model derived 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. Since monthly and yearly rainfall of Meru and Embu regions were seasonally non-stationary, the data was transformed using square root. The seasonality effect was removed by seasonally differencing. The order of seasonal differencing with a care of over differencing was chosen. Besides to this idea, the order of differencing which made our data less variable was chosen. This coincided with the assumption of constant variance of residual that was being checked in model diagnostic.

3.1The Box-Jenkins Methodology

This methodology applies autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) models to find the best fit of a time series to past values. This approach possesses many appealing features. To identify a perfect ARIMA model for a particular time series data, the following four phases were used.

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 was, to determine if the data were stationary and if there was any significant seasonality that needs to be modelled. The data were examined to check for the most appropriate class of ARIMA processes. This was done by selecting the order of the consecutive and seasonal differencing. The auto-correlation 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 helped in determining the number of autoregressive terms necessary to reveal 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).

3.1.1 Autoregressive Integrated Moving Averages

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

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t \tag{3.1}$$

and

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

$$y = u + \sum_{j=i}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t = u + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{3.2}$$

The ARIMA model 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:

$$y = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{3.3}$$

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

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 time series 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,q) model using lag polynomials is given as:

$$\varphi(L)(1-L)^d y_t = \theta L \varepsilon_t \tag{3.4}$$

i.e.

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1-L)^d y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)\varepsilon_t \tag{3.5}$$

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)

$$y_t = y_{t-1} + \varepsilon \tag{3.6}$$

This is a special case of ARIMA and is known as the Random Walk model.

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

4.1 ACF and PACF Plots of differenced and Seasonally Differenced Square root Series

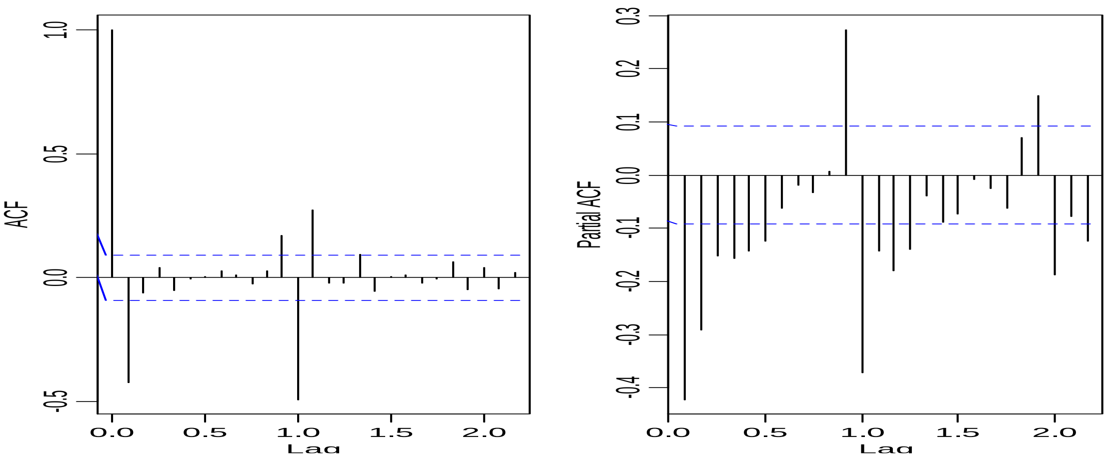

Figure 4.1: ACF and PACF Plots of differenced and Seasonally Differenced Square root Series

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 below.

Table 4.1: 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.

$$y_t = 0.2256 y_{t-12} + \varepsilon_t - 1.0000\varepsilon_{t-12} - 0.9998\varepsilon_{t-24} \tag{4.1}$$

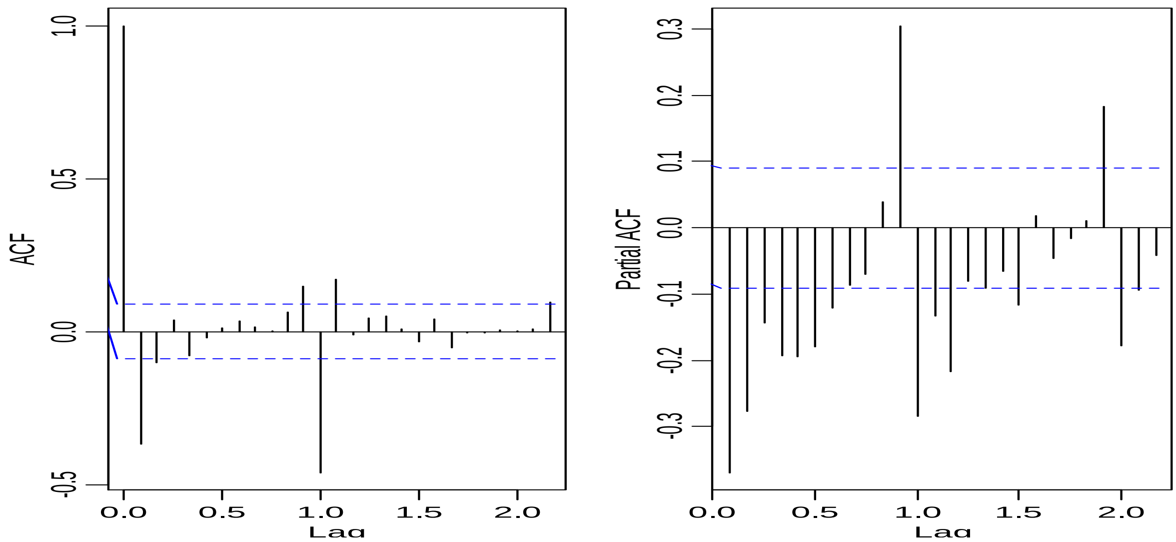

Figure 4.2: ACF and PACF Plots of differenced and Seasonally Differenced Square root Series

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 below.

Table 4.2: 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.

$$y_t = 0.1977\, y_{t-12} + \varepsilon_t - 1.00001\,\varepsilon_{t-12} - 1.0000\,\varepsilon_{t-24}$$

(4.1)

The model parameter were significant from table 4.3 and 4.4, 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.

4.3 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. The diagnostic for the fit were displayed in Figure 4.3 and 4.4.

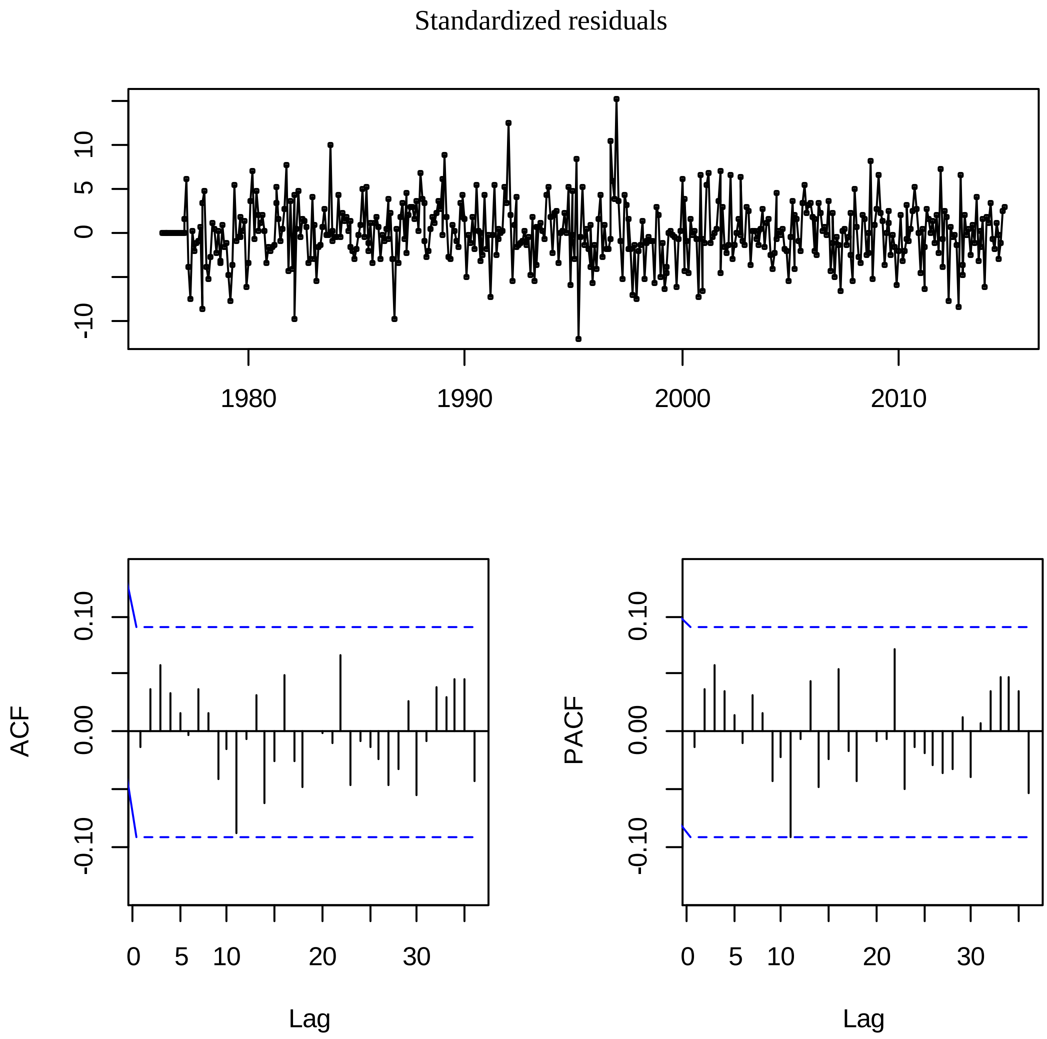

Figure 4.3: ARIMA (1, 1, 1) (0, 1, 1)12 model residuals for Meru rainfall data

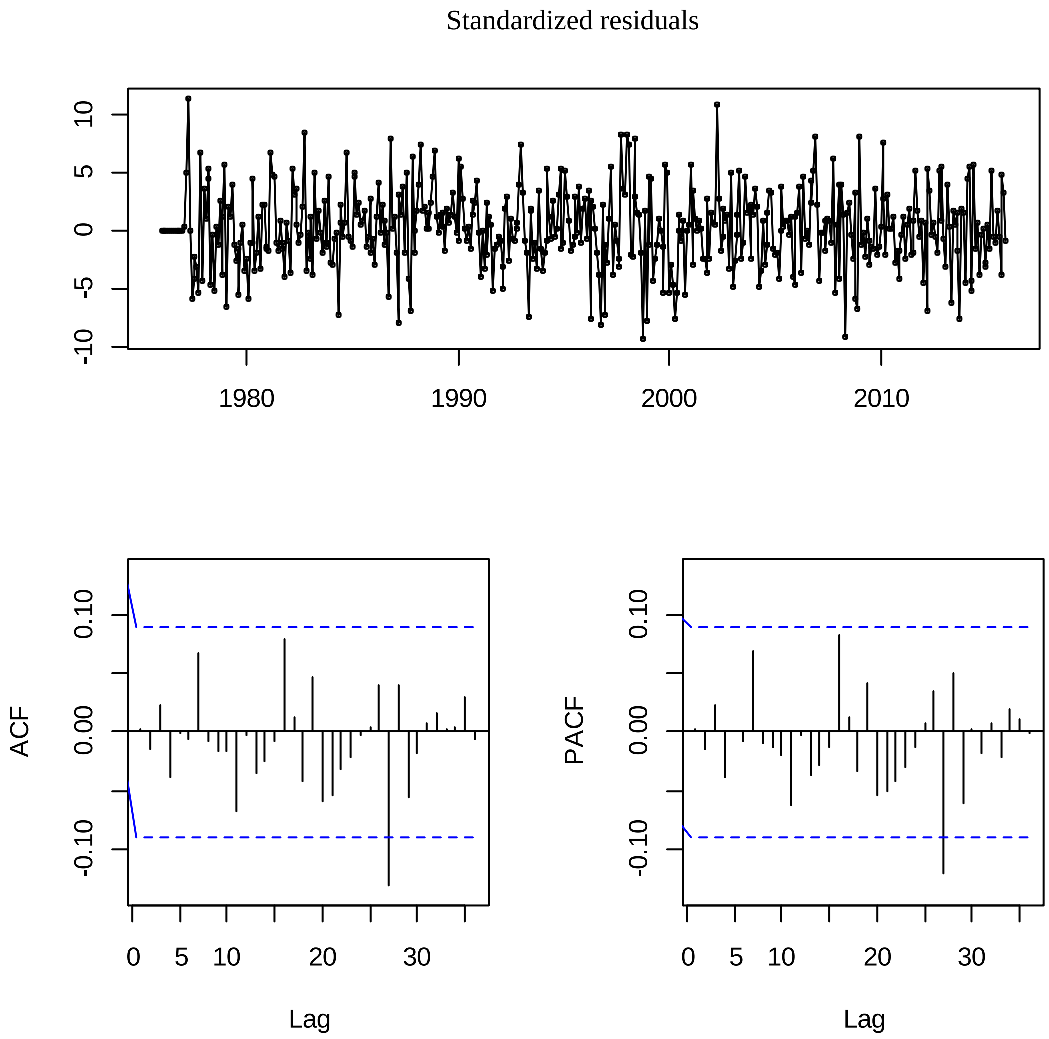

Figure 4.4: ARIMA(1, 1, 1) (0, 1, 1)$_{12}$ model residuals for Embu rainfall data

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

3.4 Forecasting

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.

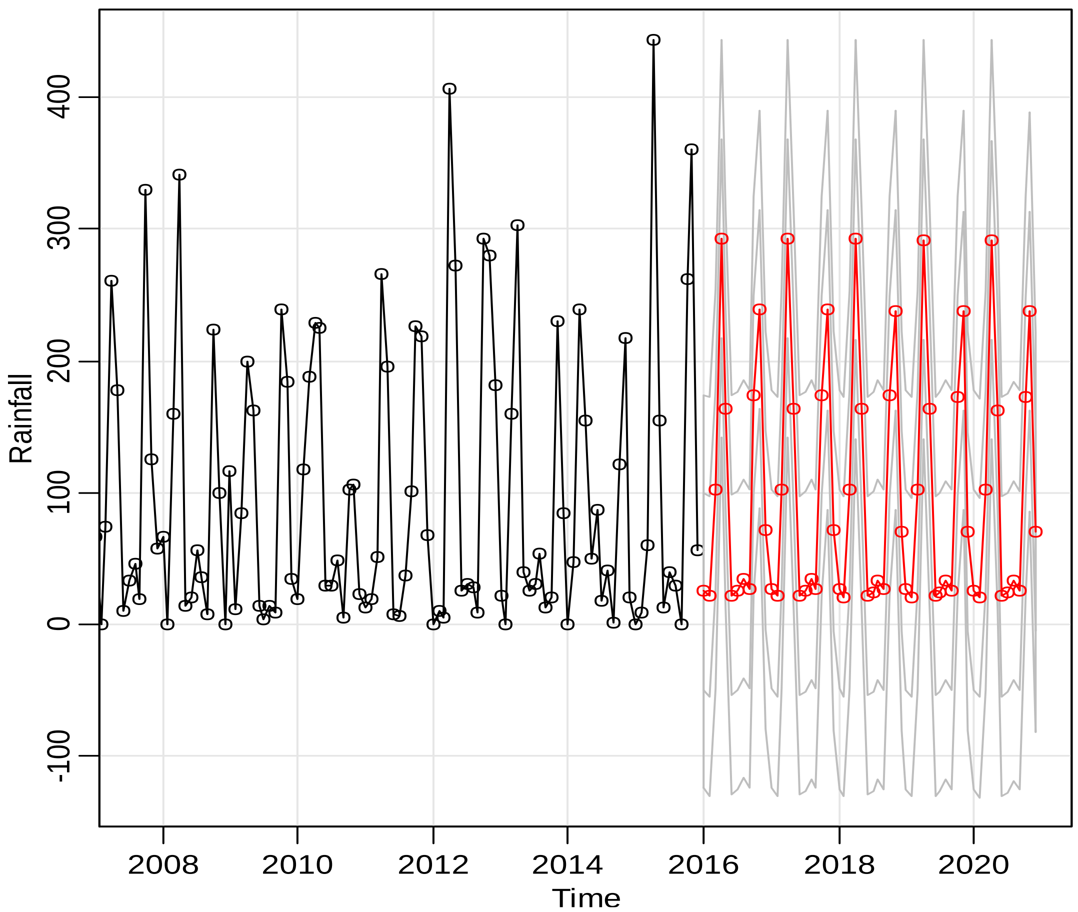

Figure 4.5: Forecast plot for total monthly rainfall of Embu

From the figure 4.5, 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 also shown, 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.

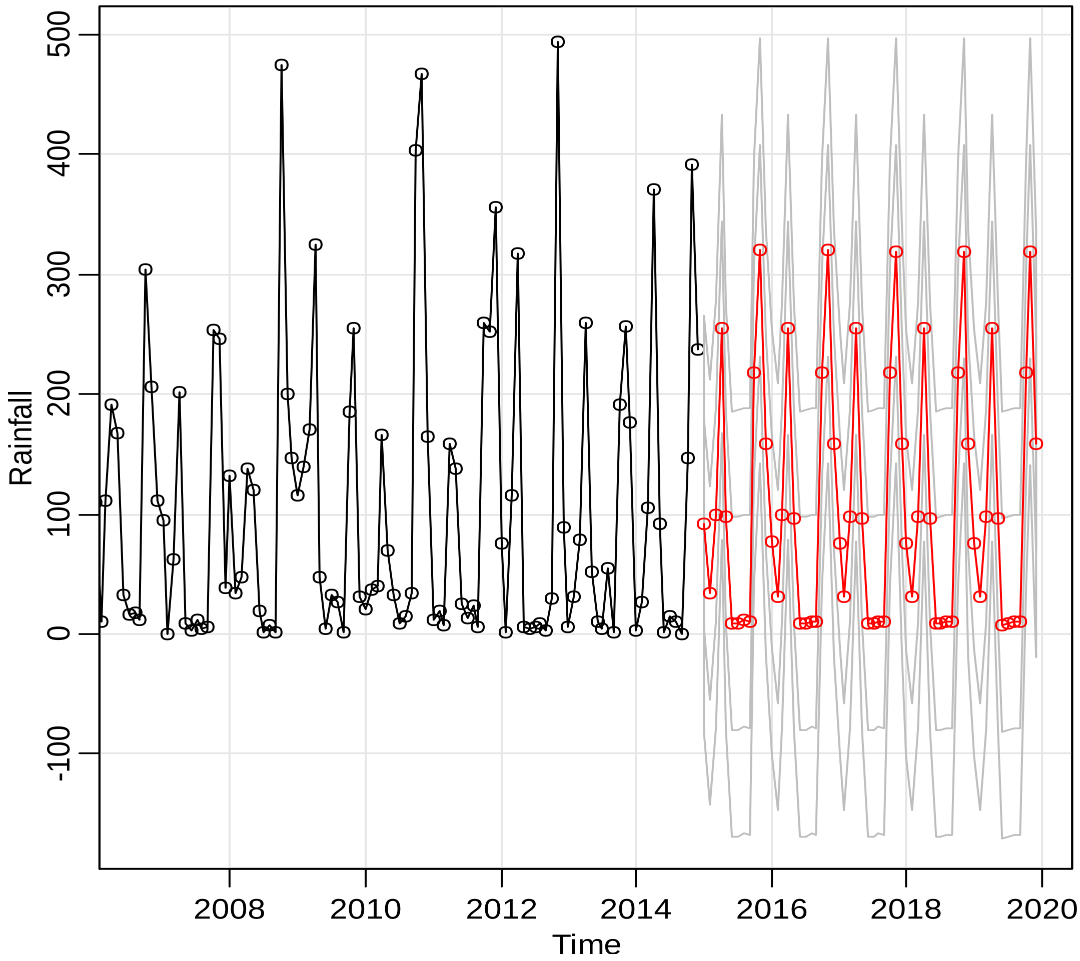

Figure 4.6: Forecast plot for total monthly rainfall of Meru

The figure 4.6 showed the series followed by the forecast as the red line and the upper and lower predictions limit as grey lines. Forecasts from the ARIMA (1, 1, 1)(0,1,1)$_{12}$ model for the next five years were also shown. The increasing and decreasing prediction intervals showed that the rainfall could start decreasing or increasing constantly.

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.

October, November and December (OND). From the fitted SARIMA (0, 0, 0)(1, 2, 2)$_{12}$ model, we concluded that rainfall distribution in both regions decreases with time. Rainfall in both regions is bimodal, it has both long rains and short rains. It appears that short rains have high amount of rainfall as compared to long rains.

5. 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 rains in 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.

References

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.

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

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

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.

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

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

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.

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

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.

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

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.*