International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 327-342 www.ijaem.net ISSN: 2395-5252

Modelling Rainfall Patterns using SARIMA Model: a Case Study of Nyeri County in Kenya

Kiplangat Dennis Cheruiyot[1]*
[1]Department of Statistics and Actuarial Science, Dedan Kimathi University of Technology.
Corresponding Author:Kiplangat Dennis Cheruiyot

Date of Submission: 10-11-2024 Date of Acceptance: 20-11-2024

ABSTRACT: The paper investigated the time series components, and to build appropriate model to forecast the rainfall of Nyeri county using monthly data from January, 1983 to December, 2015. The descriptive statistics of the rainfall showed that the highest amount of rainfall was recorded in April 1997 while the lowest amount of rainfall was recorded in September 1983. The time series rainfall data was decomposed into stochastic trend, seasonal variations and remainder. The time series of the yearly data showed decreasing trend. The rainfall data was found to be stationary and that was confirmed using dickey fuller test which yielded a test value of 6.5013 and a p-value of 0.01. The appropriate orders of models were picked based on results of ACF and PACF plots and evaluated using Akaike Information Criterion and Bayesian Information criterion. The best model was found to be SARIMA$(1,0,0)(3,0,0)_{12}$ with AIC 1944.7 and BIC 1964.46. The model residuals showed normality as most points fell on the quantile-quantile line with few close to it. The residuals also confirmed white noise. From the model validation results,the predicted values were well-fitted through the original data with lower and upper confidence limits containing majority of the original data. The RMSE of out-sample was less than that of the in-sample with values of 7.6579 and 54.6680 respectively. This showed that the identified SARIMA model was suitable for predicting rainfall of Nyeri County.
KEYWORDS: ARIMA, Seasonal-ARIMA, Decomposition, Stationarity, Box-Jenkins, ACF, PACF,

I. INTRODUCTION

Climate change is perhaps one of the most considerable environment challenges of our time. It poses a great challenge to sustainable progress globally. It affects eco-systems, water resources, food, health, coastal areas, industrial activities and human growth. Efforts in addressing these effects provide opportunities for innovation and business. If we are to cut down the many impacts of climate change variability, the necessary mitigation, adaptation and coping up strategies must be implemented. This then imposes large amount of resources to Kenya in large to address current and future climate change effects and endangers to rise even higher by 2030. Unless effective mitigation and adaptation systems are instituted with no delay, the combined effects of climate change impacts will hinder realization of targeted goals in Vision 2030.

Kenya is already feeling the effects of climate changes such as: the widespread poverty, recurrent droughts, floods, inequitable land distribution, over-dependence on rain-fed agriculture and few coping mechanisms all combine to increase people's vulnerability to climate change. For instance, disadvantaged people have little security against intense climatic actions happening today. This is because they have poor housing, few resources reserves and moreover they depend on the naturally available resources for their living. Thus, this calls for precise forecasting models to clearly predict factors of the climate such as precipitation, temperature, humidity, atmospheric pressure and wind. However, previous research done both under county's level and in country's level seems to have lost their edge due to changes in climate. To date, although numerous studies exist on weather and climate variability in Kenya as a whole, relatively little focus on counties, particularly in Nyeri County, has been done. A comprehensive review of the current state of climatic model data especially on rainfall patterns in Nyeri County is currently missing and is a knowledge gap this project aims to fill.

From the observation of rainfall patterns of Nyeri County, one can conclude that climate change and global warming have contributed to changes. In brief, climate change is described as the change in statistical distribution of weather patterns when the change occurs for an extended period of time while global warming is the average rise of the earth's temperatures generally attributed to the greenhouseeffect. Climate change is caused by natural activities and human activities. Globally the biggest cause of climate change is the increase in the atmospheric carbon dioxide, which is generated primarily through burning of fossil fuels. To control this then we need to reduce the amount of carbon generated every year, reduce deforestation, and replace fossil fuels with alternative sources of energy. We need to deal with the negative human activities by imposing new laws so as to reduce the increase of greenhouse gases in the atmosphere which eventually lead to global warming. Some of the efforts made in dealing with climate change worldwide is the Kyoto protocol which is an international treaty which extends the 1992 United Nations Framework Convention on Climate Change that commits state parties to reduce greenhouse gas emissions. Some nations signed the agreement while others did not sign. This then makes it important for each country to take its own effort in reducing climatic change. In Kenya we have the Green Belt Movement which takes keen focus on environmental conservation among other things. The movement was founded by the late Nobel Laureate Honourable WangariMaathai.

Talking about environmental conservation, Nyeri County is one of the most privileged counties in Kenya owing to the fact that it has a relatively large forest cover. It has a 38.03% of land covered by trees, according to 2015 mapping of forest cover reported by Kenya Forestry Service. Some of the forests in Nyeri include Nyeri forest, Kiganjo forest and Aberdares forest among others. Nyeri hosts two major water towers which are the Mount Kenya catchment and Aberdares hence this makes it necessary to formulate a policy to conserve the ecosystem and in mitigating against effects of climate change. The forests are important water catchment areas being sources of major rivers and tributaries and hence this necessitates conservation of the forests. Trees also play a huge role in maintaining the carbon cycle. They convert the carbon dioxide in the air to oxygen through photosynthesis, and in this way they are looked as natural regulator of carbon dioxide. The more trees there are the less the carbon dioxide in the air and the vice versa is equally true.

However due to the human activities in Nyeri the forest cover has been deprived and is slowly de-creasing. It is estimated a relatively large number of residents in Nyeri use wood fuel as their main source of fuel from local forests. This has continuously led to the decrease of the forest cover in Nyeri and hence action needs to be taken. This will in turn help in curbing the impacts of climate change which has resulted to change in rainfall patterns of particularly of Nyeri count. The overall results are that Nyeri County will be able to maintain a steady crop production especially on the major crops that are exported like tea and coffee earning the country foreign exchange. A prediction in therainfall patterns will be important so as to know when it is necessary to carry out certain measures in maintaining steady crop production.

II. BACKGROUND OF THE STUDY

Rainfall is an important climatic factor since it is a source of life to both plants and animals on planet earth. Rainfall is a form of precipitation whereas precipitation is any product of condensation of atmospheric water that falls under gravity. The main forms of precipitation include rainfall, snow and hail. Rainfall being a form of precipitation is a major component of water cycle and is responsible for depositing the fresh water on the plant and animals. Rainfall is formed as a result of air being lifted in the atmosphere where it expands and cools. Cool air cannot hold as much water in vapour form as warm air and the condensation occurs and eventually droplets are formed. These droplets grow to large sizes to form clouds and they will eventually be heavy enough to fall on the earth's surface as rain.

There are three types of rainfall which include relief rainfall, cyclonic rainfall and conventional rain-fall. Relief rainfall majorly occurs when prevailing winds pick up moisture from the sea as they travel across it, making the air moist. The moist air reaches the coast and is forced to rise over the mountain and hills. This forces the air to cool and condense forming clouds. The air continues to be forced over the mountains and it then condenses forming droplets which merge to form clouds. When the clouds are heavy enough they pour down as rainfall. Convectional rainfall occurs in tropics where it is hot while cyclonic rainfall occurs when warm air is forced to rise over cold air. Nyeri being our county of interest as stated earlier, receives relief rainfall because Nyeri is located on the windward side of Mount Kenya. Nyeri receives an average rainfall of 1004 millimetres per annum and has average temperatures of 17.1^{o}C.

The main aspects of rainfall include rainfall amount, intensity and distribution. Rainfall amount is measured by use of a rain gauge. These aspects are important since they have considerable influence of on crop production and human life in general. Rainfall being a major source of water, we find that water is the most valuable natural resource and is vital to all forms of life. Water is used for trans-port, hydroelectric production and serves useful purposes for domestic consumption, agriculture and industry. The amount of water for various purposes is much dependent on the amount of precipitation in that particular area. For instance, in Nyeri, rainfall water is used for commercial purposes such as horticultural farming in large scale in Mweiga Blooms in Kieni and Wilmar flowers in Sagana area. Dairy farming and fish keeping are also practiced in Tetu, Aguthi and Chinga dams all located in NyeriCounty. Dairy farming is mostly practiced on small scale basis mainly at homes. Trout rearing around the base of Mount Kenya, along Chania and Gura rivers is practiced. Nyeri County has a number of light companies such as Maisha Flour Millers, Brookside Dairy, Mount Kenya bottlers and Highlands Mineral Water which depend on water as an important component in their production activities.

Most importantly, Nyeri is one of the counties which form the food baskets of our country Kenya. The main food crops grown include maize, beans, wheat, bananas, Irish potatoes, vegetables whereas the major cash crops coffee, tea and horticulture. These agricultural activities depend on the well distributed rainfall in Nyeri County. Rainfall water, however, can also have negative effect on the earth when it causes erosion, when it has high pH and its characteristics features which are lightning and thunder. Also it causes natural disaster which destroys properties and at times takes lots of lives,it causes leaching to such depth that, plant roots are lost in the process. In Nyeri negative effect of excess rainfall is landslides.

Traditionally, Nyeri County experiences two seasons, one from March to May which is the long rains and the other from October to November which is the short rains. On average, the warmest months in Nyeri are February and March while most rainfall is experienced in April and May. Nyeri has dry periods in January and in February. Due to climate change, these trends have been changing. These changes normally occur on aspects of weather such as wind speed, humidity, temperature, precipitation which occurs in any of the three forms as stated earlier. This then brings the need to study and understand rainfall patterns of the area so as to use more accurate techniques in forecasting the rainfall patterns.

III. STATEMENT OF THE PROBLEM

Rainfall variability and associated droughts have historically being major causes of food shortage and severe famines in Kenya. Though rainfall variability and droughts are not new phenomenon in Kenya, the frequency of their occurrence has been increasing in the past decades. Floods and droughts are considered to be the two extreme conditions of variability of rainfall. These conditions have claimed many lives in Kenya. In some parts of Nyeri County, the number of people affected by drought and famine has increased due to prolonged drought while others have been displaced as a result of landslides. The frequent droughts and landslides have been worsened by the climate change which in turn affects food production and lives of people.

Thus, the county has being experiencing food shortage and low production of food in the recent years due to frequent droughts. Till now, except a few studies done in some regions of Nyeri no keen interest has been taken by researchers to study and model the changing rainfall patterns which are a major cause of food shortage and famine. This projects aims at filling the existing research gap so as to able to clearly understand the rainfall patterns in Nyeri and plan appropriately for agricultural production.

IV. EMPIRICAL LITERATURE REVIEW

A study on the rainfall patterns for UasinGishu county was done by Metrine et al., (2015) for the period, January 1977 to December 2014 monthly observations. The authors employed Box-Jenkins methodology to fit an ARIMA model to the data. The results showed that there was presence of strong seasonality in the rainfall data and hence SARIMA model was used. The model that was found best was SARIMA $(0,0,0)(0,1,2)_{12}$. This is because the model had significant parameters and the lowest AIC and BIC. Finally, the model residuals were found to be independent and identically distributed, white noise, homoscedastic, normally distributed and with mean zero hence it was adequate for modelling. The model was used to predict values for the year 2014 and the results were compared with the actual results of 2014. The values lied within the range hence the model was good enough. The model was then used for forecasting rainfall accurately for the next two years.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 327-342 www.ijaem.net ISSN: 2395-5252

Mahsin et al., (2012) modelled rainfall of Dhaka division for the period January, 1981 to June, 2010 with a total of 354 readings. The authors used Box-Jenkins methodology in building seasonal ARIMA model for monthly rainfall. The results showed that the monthly rainfall showed a seasonal cycle of series. It was further found that the best model that fitted the data was SARIMA (0,0,1) $(0,1,1)_{12}$. The model was then validated by testing its adequacy. The authors obtained the graphical plot for actual pitch and the predicted series and concluded by visual inspection of plots that the model was good. The model was in the forecasting the monthly rainfall for the upcoming two years.

Kibunja et al., (2014) studied precipitation forecasting in Mount Kenya region, using the Box-Jenkins methodology. The authors used data from the Kenya Meteorological Department covering period of 1995 to 2010 for wind data and 1977 to 2011 for rainfall data. However, the data used was limited to available wind data. The usual criteria, that is, stationarity, autocorrelation, and partial autocorrelation functions, significance of coefficients and Akaike Information Criterion were used to select the best model. The best model was found to be SARIMA (1,0,1) $(1,0,0)_{12}$. The model residuals were white noise, normally distributed and the plots were well visible and hence the model was found adequate enough.

Otieno et al., (2014) analysed tourist accommodation demand in Kenya and its impacts to the tourism industry in Kenya. Quarterly data on tourist accommodation demand was obtained from various statistical abstract publications by the Kenya National Bureau of Statistics (KNBS) for the period between 1985 and 2013. The authors used the Box-Jenkins methodology and fitted a suitable time series model. The best model fitted was SARIMA (1,1,2) $(1,1,1)_{12}$. The spikes in the residual display were insignificant and showed that the residuals were white noise. Furthermore, Ljung-Box test results showed that the residuals were random. They accurately predicted a five-year forecast for the number of tourists accommodated in Kenya, from quarter 1 of 2012 to quarter 4 of 2016. This showed that the SARIMA model was good enough.

Zhang et al., (2005) analysed two models, SARIMA model and Time Lagged Feed-Forward Neutral Networks to determine which model was best in forecasting. In particular, they studied the effective-ness of data pre-processing, including de-seasonalisation and de-trending, on neural network modelling and forecasting performance. Both simulation and real data are examined and results were compared to those obtained from the Box-Jenkins SARIMA model. The results showed that a neural network was not able to capture seasonal or trend variations effectively with the unprocessed raw data and either de-trending or de-seasonalisation could dramatically reduce forecasting errors. Moreover, a combined de-trending and de-seasonalisation was found to be the most effective data pre-processing approach. The authors came to a conclusion that the performance of SARIMA model built through the Box-Jenkins methodology was best since it was able to capture seasonal of trend variations effectively.

Etuk et al., (2014) investigated time series analysis of monthly rainfall data collected at Gadaref rainfall station, Sudan. The authors used data for the period 1971 to 2010. They employed the Box-Jenkins methodology to analyse the data and found SARIMA (0,0,0) $(0,1,1)_{12}$ as the best model. The model wastested for adequacy and the residuals were observed to be uncorrelated and hence formed the basis of forecasting of rain and planning purposes. The model they choose had the best results for R-adjusted.

A research was carried out on the rainfall patterns in the Navrongo municipality, Ghana using univariate time series using data for the period January, 1980 to December 2010. The research was done by (Sampson et al., 2013). The authors divided the study data into two parts which included the in-sample and out-sample data. The in-data sample was then used in developing a model while the out-sample data was used for sample comparison. The authors applied the Box-Jenkins methodology and the results showed SARIMA (0,0,1) $(0,1,1)_{12}$ as the best model since it performed better for both in-sample and out-sample forecasting. An overall check for the model adequacy using Ljung-Box test revealed the model was adequate in forecasting.

Jamaludin et al., (2015) studied the temporal dynamics of trend in relative humidity for the period 1968 to 2009 for 13 stations in Peninsular, Malaysia. In understanding the trend flow, the Mann-Kendall trend test of relative humidity was used and the test recorded a decreasing trend over all parts excluding one station. The authors applied the Box-Jenkins methodology to build a Relative Humidity SARIMA model for each station. The results showed that the fitted Relative Humidity SARIMA was the best model. They also found that the relative humidity data was influenced by seasonal behaviour and hence they forecasted 30 months upcoming relative humidity data. The authors noted that there was a decreasing trend for most stations.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 327-342 www.ijaem.net ISSN: 2395-5252

Etuk et al., (2016) studied the Box-Jenkins method based additive model simulating model for daily Uganda-Nigeria exchange rates, for the period 22nd September, 2015 to 16th March, 2016 (177 values). First, they divided the data into two parts; the first 170 values were used for the modelling and the 7-values were used for out-sample forecast goodness-of-fit. Second, the authors employed the Box-Jenkins methodology and found that additive SARIMA (1,1,0) $(1,1,0)_7$ as the best model. Forecasts obtained for the daily rates from March 10 to March 16, 2016 agreed so closely with the observed values that the calculated goodness-of-fit chi-square test statistic was far from being statistically significant with a p-value of more than 99%.

Salahi et al., (2016) modelled precipitation characteristics and simulation of drought using the standardized precipitation index during the 2011-2020 in Mashhad, Iran. The data related to the average of monthly precipitation in the synoptic station of Mashhad (from 1951 to 2010) was obtained from the Meteorological Department Organization of Iran. Using the Box-Jenkins method, the monthlyprecipitation was modelled from 2011 to 2020, with respect to its preceding series trends. In addition to using the standardized precipitation index, climatic conditions in the upcoming years were investigated in terms of drought. The results indicated that SARIMA (2,0,1) $(2,1,1)_{12}$ was the best model for fitting the precipitation data. The model was used to forecast precipitation in Mashhad station from 2011 to 2020.

From the literature above, it is observed that data from different geographical locations follow different SARIMA models. This is because the locations have different climatic conditions. SARIMA models are evident where there is a seasonality component in the data. This then makes it necessary to model the rainfall data for Nyeri County since no existing literature on modelling of rainfall patterns of the area.

V. METHODOLOGY

Auto-Regressive (AR) model

An autoregressive model of order p, AR(p) is a model in which a linear combination of previous measurements of the variable and a random error term with a constant term are used to forecast the variable of interest.

The p^{th}order autoregressive process is given by;

$Z_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_p Z_{t-p} + e_t$where$e_t \sim WN(0, \sigma^2)$

In lag operator form, we write;

$$\left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p\right) Z_t = e_t$$

With the following properties, the auto-covariance function is;

$$y_k = \theta_1 \gamma_{k-1} + \theta_2 \gamma_{k-2} + \cdots + \theta_p \gamma_{k-p}, k > 0$$

whilerecursive relation for the autocorrelation function is;

$$\rho_k = \theta_1 \rho_{k-1} + \theta_2 \rho_{k-2} + \cdots + \theta_p \rho_{k-p}, k > 0$$

Hence the ACF $\rho_k$tails off as a mixture of exponential decays depending on the roots ofρ_p being equal to zero.

Moving Average (MA) model

A Moving Average, MA (q) model uses past errors to predict the variable of interest. The q^{th}order moving average process is of the form:$Z_t = e_t + \alpha_1 e_{t-1} + \cdots + \alpha_q e_{t-q} = \theta(B) e_t$ where $\alpha_1, \ldots, \alpha_q$ are parameters of the model and $e_{t-1}, \ldots, e_{t-q}$ are white noise error terms.

In lag operator form, we write:

$$\theta(B) = \left(1 + \alpha_1 B + \alpha_2 B^2 + \cdots + \alpha_q B^q\right)$$

The model has the following properties; The MA (q) process is always stationary because $1 + \alpha_1^2 + \cdots + \alpha_q^2 < \infty$. The process is invertible if the roots of$\left(1 + \alpha_1 B + \alpha_2 B^2 + \cdots + \alpha_q B^q = 0\right)$lie outside a unit circle.

The auto covariance function is; $y_k = \begin{cases} \sigma^2(\alpha_k + \alpha_k \alpha_{k+1} + \cdots + \alpha_{q-k}\alpha_q), & k = 1,2, \ldots, q \\ 0 & , k > q \end{cases}$

The autocorrelation function is;$y_k = \begin{cases} \frac{(\alpha_k + \alpha_k \alpha_{k+1} + \cdots + \alpha_{q-k}\alpha_q)}{(1 + \alpha_1^2 + \cdots + \alpha_q^2)}, & k = 1,2, \ldots, q \\ 0 & , k > q \end{cases}$

The autocorrelation function of an MA (q) process cuts off after lag q.

Autoregressive Moving Average (ARMA)(p, q)

The AR model includes the lagged terms on the time series itself while the MA model includes lagged terms on the noise or residuals. If the AR and MA models are effectively combined together we form the ARMA model. Thus ARMA (p, q), where p is the autoregressive order and q the moving average order is written as:

$$Z_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_p Z_{t-p} + \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \cdots + \alpha_q Z_{t-q}$$

where $\theta_p(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p$, such that $\theta_p \neq 0$ and $\alpha_q(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_q B^q$, such that $\theta_q \neq 0$.

A zero mean ARMA (p, q) process is then defined as,$\theta_p(B) Z_t = \alpha_q(B) e_t$.

The process is invertible and stationary if the roots of $\alpha_q(B) = 0$ and $\theta_p(B) = 0$respectively lie outside the unit circle for $k > p$.

Autoregressive Integrated Moving Average (p, d, q) Model

In practice, many time series are always non-stationary. ARMA models are therefore inadequate to effectively describe non-stationary time series which are more frequently encountered in actual practice. ARIMA model was proposed which is a generalization of an an ARMA model to include the case of non- Stationarity. When using the ARIMA model, finite differencing is applied to the data to remove non- Stationarity.

A time series $Z_t$is said to be homogeneous non-stationary if $(1 - B)^d Z_t$is stationary for some value of $d \geq 1$. A stationary ARMA (p, q) model for$(1 - B)^d Z_t$ in terms of a backward shift operator is given by;

$$\theta_p(B)(1 - B)^d Z_t = \alpha_q(B) e_t$$

This is called an autoregressive integrated moving average model, ARIMA (p, d, q).

Seasonal Autoregressive Integrated Moving Average, (SARIMA) Model

SARIMA models are an adaptation of ARIMA models to specifically fit seasonal time series. That is, their construction takes into consideration the underlying seasonal nature of the series to be modeled. A general multiplicative seasonal ARIMA model has the form;

$$\Phi_p(B^s)\phi_p(B)(1 - B)^d (1 - B^{s\Theta})^D Z_t = \theta_0 + \theta_q(B)\Theta_Q(B^s) e_t$$

wherep, d, q, P, D, Q are integers, s is the seasonal length.

$$\Phi_P(B^s) = 1 - \Phi_1(B^s) - \Phi_2(B^{2s}) - \cdots - \Phi_P(B^{Ps})$$

$$\phi_P(B) = 1 - \Phi_1(B) - \Phi_2(B^2) - \cdots - \Phi_p(B^P)$$

$$\theta_q(B) = 1 - \theta_1(B) - \theta_2(B^2) - \cdots - \theta_q(B^q)$$

$$\Theta_Q(B^s) = 1 - \Theta_1(B^s) - \Theta_2(B^{2s}) - \cdots - \Theta_Q(B^{Qs})$$

As polynomial in B, where Bis the backwards transfer factor, $e_t$is the estimated residual at time t, prepresents non-seasonal AR order, drepresents non seasonal differencing, q represents non seasonal MA order, Prepresents seasonal AR order, Drepresents seasonal differencing, Qrepresents seasonal MA order, $Z_t$indicates observed values at time$t = 1,2, \ldots$and$e_t \sim WN(0, \sigma^2)$.

Box-Jenkins methodology applies SARIMA model to find the best fit of a seasonal time series model to past values. The Box-Jenkins method involves four steps, repeated necessary, to end up with a specific SARIMA model that replicates the patterns in the series as closely as possible to appropriately forecast future values. The steps include model identification, estimation of parameters, diagnostic checking and forecasting. In model identification, the process involves making sure that the variables are stationary, identifying seasonality in the dependent series, and using plots of the autocorrelation and partial autocorrelation functions of the dependent time series to decide which seasonal autoregressive moving average component should be used.

The MLE will be used to estimate the parameters of SARIMA model. This technique finds the values of the parameters which minimize sums of squared residuals and maximizes the probability of observed series.Akaike's information criterion (AIC) is useful for determining the order of an SARIMA model. It is used to compare competing models fit to the same series. The competing models are ranked according to their values of AIC. The model which attains the lowest value of information is considered best.The AIC evaluates a given model depending on the closeness of its fitted values to the observed values. It selects the simplest model that best explains the given data with minimum number of parameters and penalizes the complex model for having more model parameters. Penalizing the model with more parameters discourages over fitting.The BIC generally penalizes free parameters more strongly than does the AIC, though it depends on the size of n and the relative magnitude of nand k. Good models are obtained by minimizing either the AIC, AICs or BIC and maximizing log likelihood. Our preference is to use the AICs and the model with the lowest AICs and largest log likelihood will be selected.

Each selected model is assessed to determine how well it fits the rainfall data. For a model that fitsthe data well, the standardized residuals estimated from it should be independently and identically distributed with zero mean and constant variance. Such a sequence is referred to as white noise. The residuals of a well fitted model should be randomly distributed. Several diagnostic statistics like JarqueBera, normality QQ plots, standardized time residuals, ACF and PACF of the residuals are used in determining the goodness of fit of the selected model. The Ljung-Box test will also be used to verify whether the autocorrelation of a time series are different from zero

Forecasting is important in decision making process. The chosen model should therefore produce accurate forecasts. The selected model does not always necessarily provide the best forecasting therefore it is important to apply tests like the standard root mean square error to confirm the forecasting accuracy of the model.

VI. RESULTS AND DISCUSSION

According to the findings, the data showed an annual seasonality component for the Nyeri County data analysed. The plot of the raw rainfall data for Nyeri County was as shown in the figure 1 below.

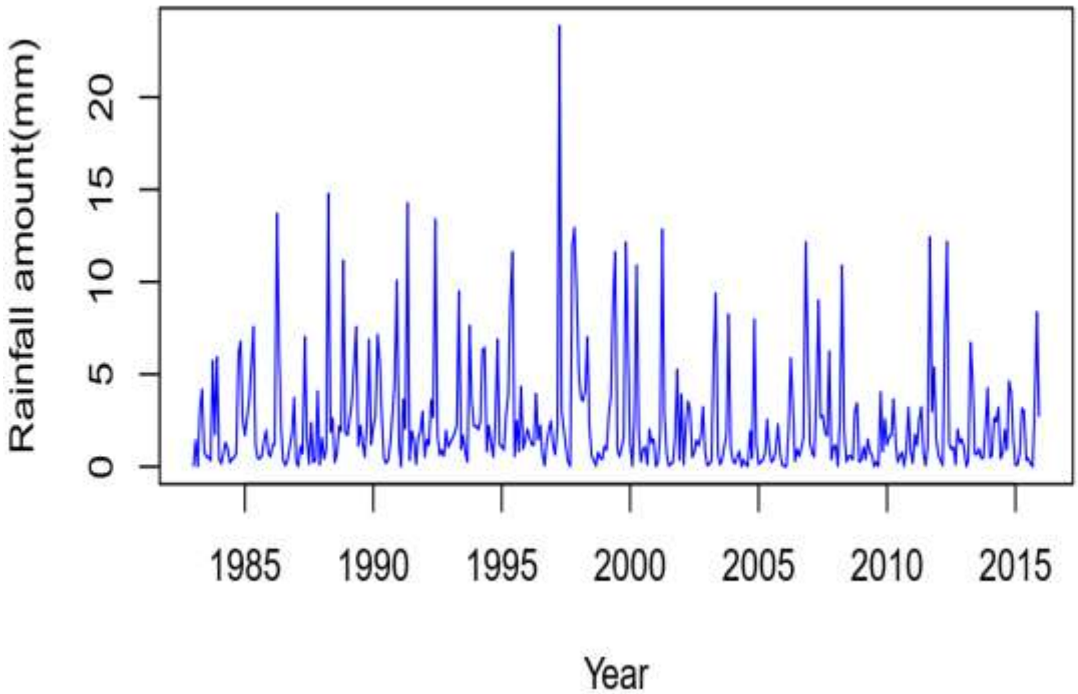
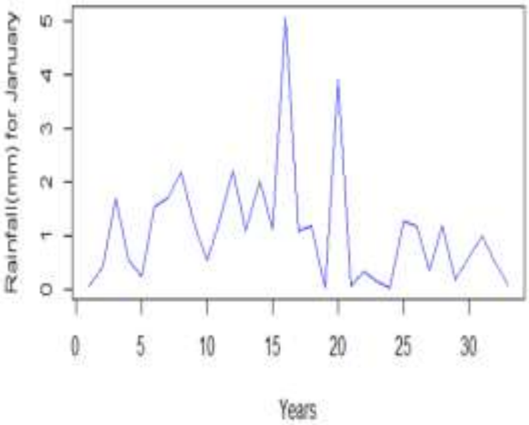
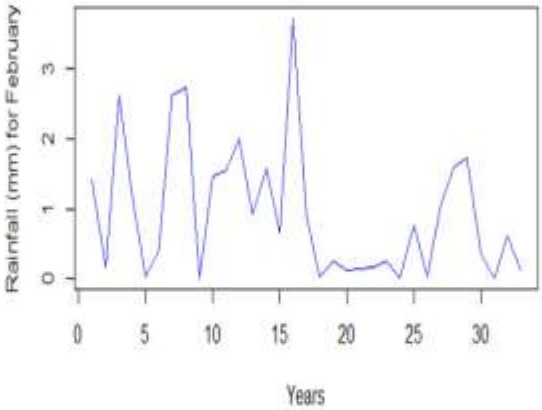


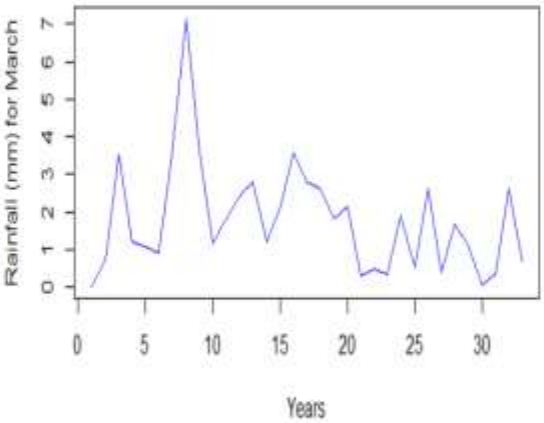
Figure 1: Time Series plot of the Nyeri County Rainfall Data.

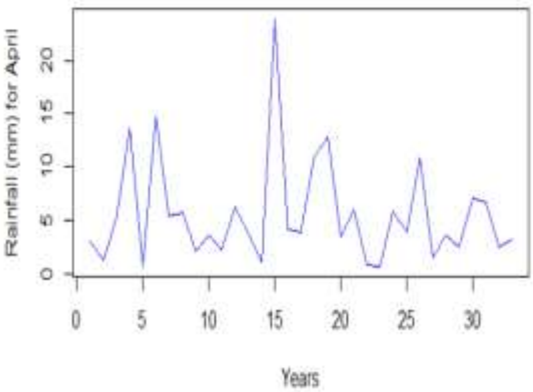
The 396 monthly rainfall data was plotted against time as from January, 1983 to December, 2015 in the Figure 1. The time series formed a pattern that repeated itself every year implying seasonalityin the data. Since seasonal fluctuations occur every 12 months, they resulted in a period time series, S = 12. The peaks in the plot series of the rainfall data also suggested strong seasonality pattern.

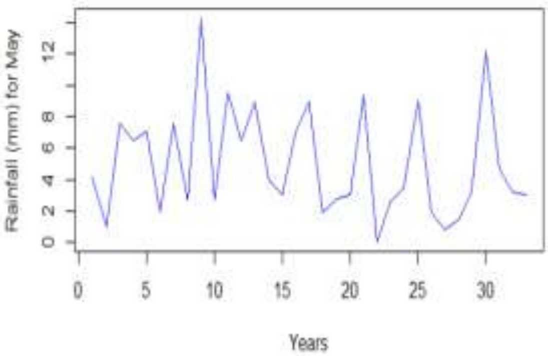
International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 327-342 www.ijaem.net ISSN: 2395-5252

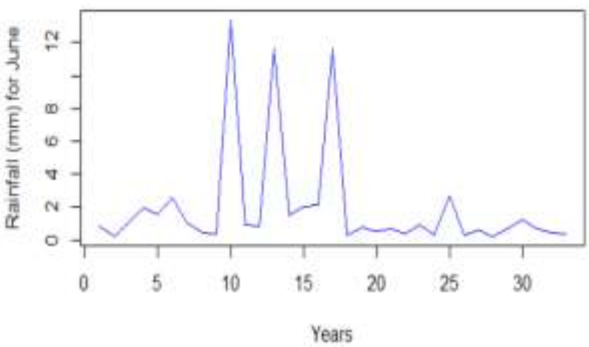












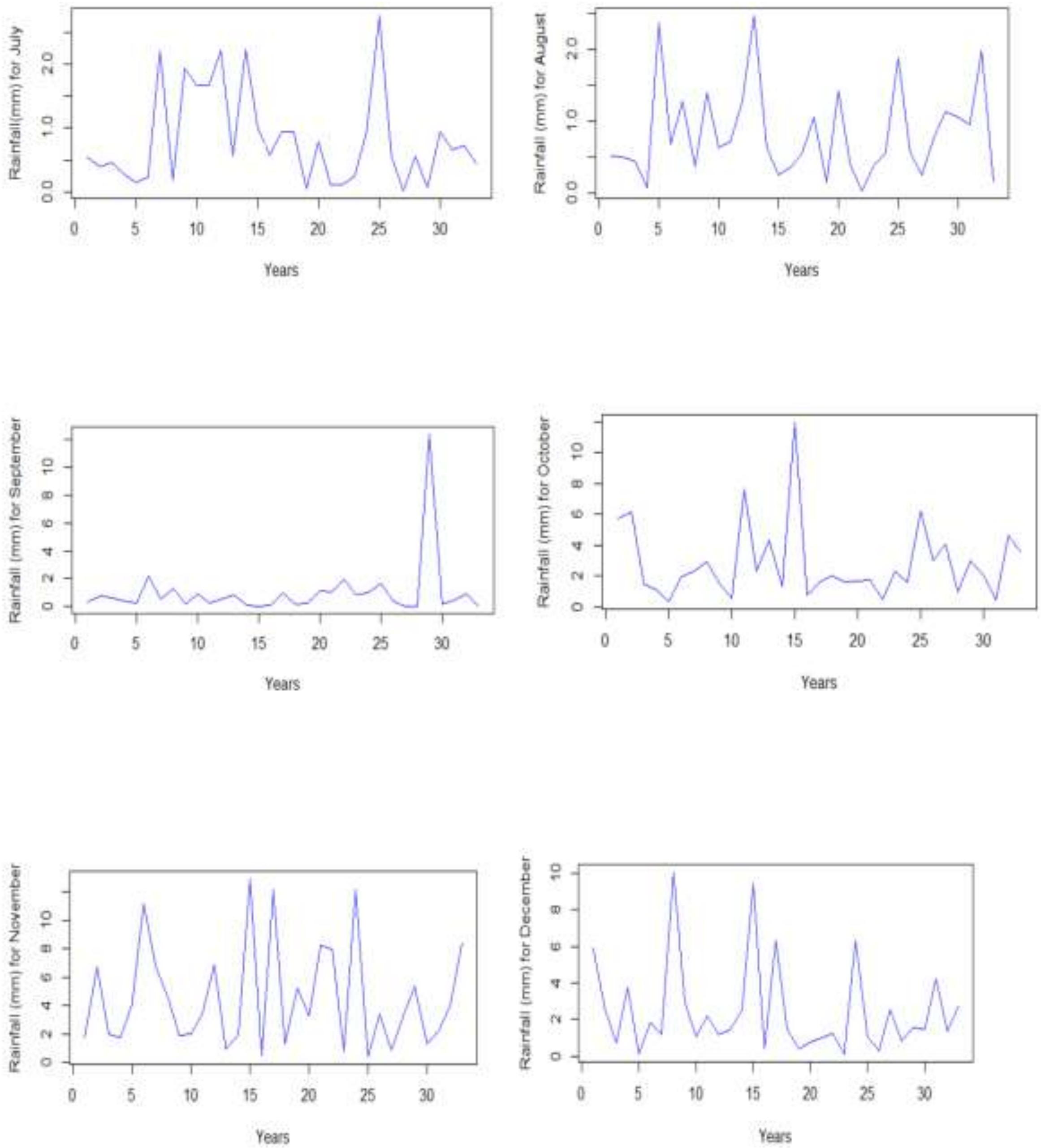


Figure 2: Time Series of the Nyeri Monthly Rainfall Data

From the time series plot in figure 2 of Nyeri monthly rainfall data, September has generally generated the lowest rainfall amounts, exceptions of the last few years while May, August and November record approximately high rainfall. In the last few years, about five years, rainfall has drastically reduced as compared to the rest of the years, under investigation. This then, showed that there are some factors that have come into play that were not there during the ancient time. According to the time series plots in figure 2 onset of long rains is evident in the month of April, elapsing in May while start of short rains is observed in November and ends in December. The month of August has been receiving significant rainfall. During 1997, April and September recorded the highest rainfall and this is the time El-nino was experienced.

The descriptive statistics was computed and the highest amount of rainfall was recorded in April 1997 while the lowest amount of rainfall was recorded in September 1983. The standard deviation of the rainfall was (sd = 3.099) and the mean average rainfall was found to be 2.4061. The distribution curve is positively skewed (skewness = 2.5603) and leptokurtic with (kurtosis = 8.4903). The summary is as shown in the Table 1.

Table 1: Summary Statistics of the Nyeri Rainfall Data.

	Minimum	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Maximum**	**Standard deviation**
Nyeri Rainfall(mm)	0.0085	0.5249	1.2864	2.4061	2.7633	23.8676	3.090

The summary of monthly-wise observations was obtained as shown in the Table 2.

Table 2: Summary of Nyeri Monthly Data.

Months	**Minimum**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Maximum**
January	0.0246	0.3195	1.0979	1.0914	1.2946	5.0718
February	0.0091	0.1451	0.6531	0.9473	1.5568	3.7284
March	0.0164	0.7210	1.7043	1.8072	2.6883	7.1448
April	0.7027	2.4469	3.8354	5.5723	6.3161	23.8675
May	0.0164	2.6465	3.4003	5.0122	7.5708	14.2650
June	0.2201	0.4403	0.8213	1.9791	1.5409	13.3858
July	0.0164	0.2458	0.5654	0.8236	0.9341	2.7612
August	0.0328	0.3851	0.6309	0.8398	1.1307	2.4745
September	0.0085	0.1355	0.5249	0.9827	0.9991	12.4206
October	0.3523	1.4421	2.0156	2.8330	3.5560	12.0035
November	0.3979	1.7780	3.4459	4.5384	6.8834	12.9286
December	0.0983	0.9955	1.4748	2.4557	2.7530	10.0699

The decomposed rainfall data was as shown in figure 3 below.

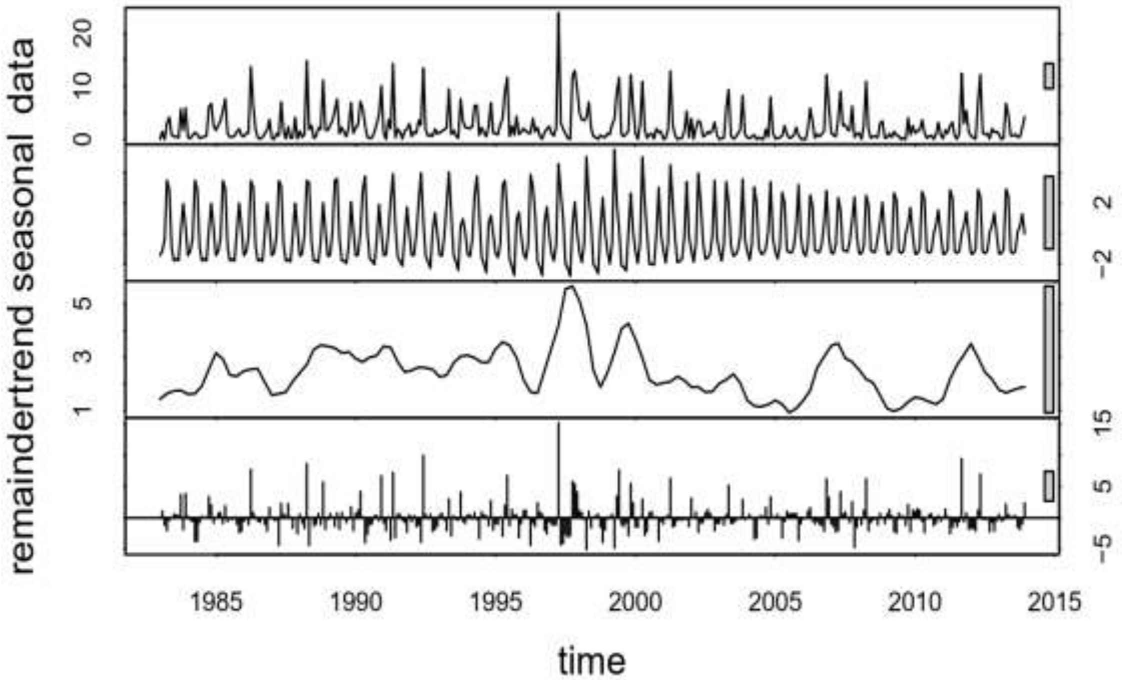


Figure 3: Decomposed Series of the Nyeri Rainfall Data.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 327-342 www.ijaem.net ISSN: 2395-5252

Ordinarily, time series exhibits trend, seasonal, cyclical and random components. From Figure 3, it is evident that the rainfall data had seasonal, random and trend components as clearly shown.

A time series is said to be a stationary if both the mean and the variance are constant over time. If the data is non-stationary, we do a logarithm transformation or take the first (or higher) order difference of the data series which may lead to a stationary time series. This process will be repeated until the data exhibit no apparent deviations from stationarity. From figure 1 of the general rainfall data of Nyeri, there was no evidence systematic variation about the mean. This then gave a rough idea that the data was stationary. But unless statistical tests were carried out, stationarity of the data remained to be unknown. The Augmented Dickey Fuller (ADF) test, ACF and PACF plots were used to check the stationarity of the data as shown in figures 4, 5 and table 3 below.

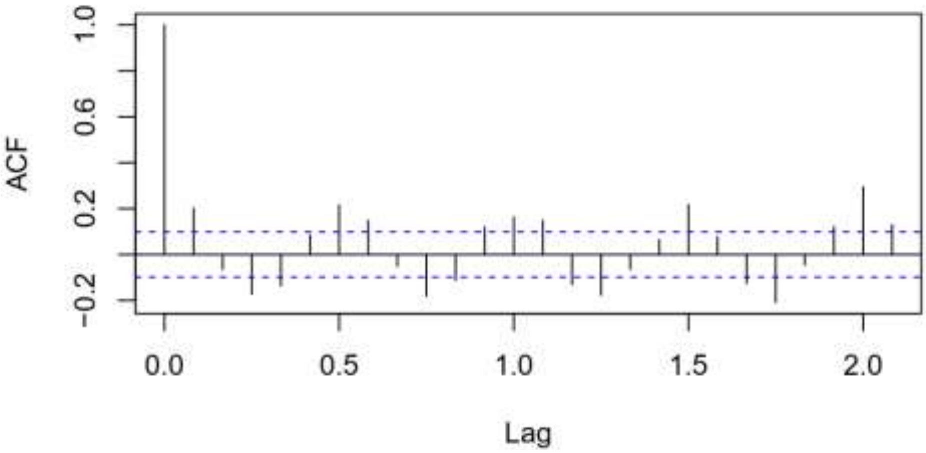


Figure 4: ACF of the Rainfall Data.

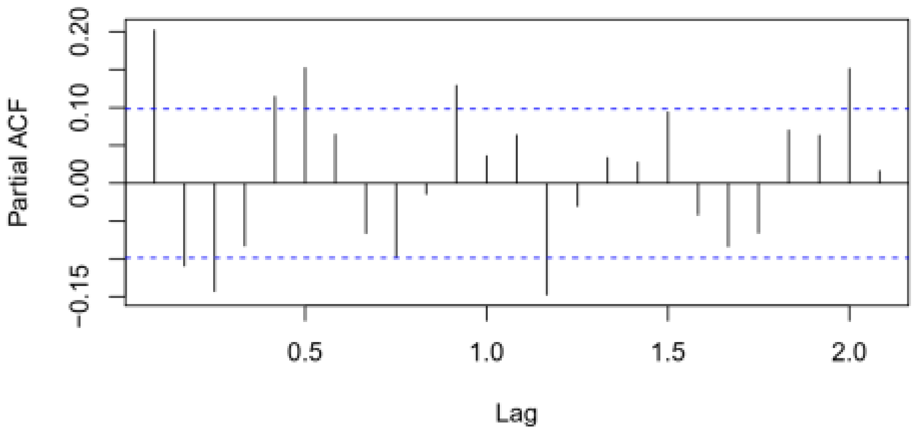


Figure 5: PACF of the Rainfall Data.

The autocorrelation function in figure 4 shows the sinusoidal pattern with 12 month periodicity indicating presence of strong seasonality behavior of rainfall series is clear. However, as lags increasethe autocorrelation at multiples of seasonal lags seem to decay rapidly implying stationarity in the seasonal component of the monthly rainfall data. As non-seasonal lags increase the autocorrelation seem to be relatively decaying indicating that the series is non-seasonally stationary. This can also be confirmed by the partial autocorrelation function plot. A similar trend is depicted for the partial case in figure 5 and hence the data is seasonally stationary. A formal test of stationarity was performed next to confirm the conclusions from the visual inspections of seasonal and non-seasonal stationarity.

The ADF test is used to test the stationarity of the data. The test statistics t value is compared to the relevant critical value for the Dickey Fuller test. If the test statistic is less than the critical value, we reject the null hypothesis and

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 327-342 www.ijaem.net ISSN: 2395-5252

conclude that the series is stationary. The hypotheses of this test are; H_0: Rainfall series has unit root versus H_1: The series is stationary.

Table 3: ADF Test Results

ADF Test Value	P-value	Lag order
-6.5013	0.01	7

From table 3, ADF test reported a test value of 6.5013 and a p-value of 0.01. This result presented evidence in favor of the alternative hypothesis, showing that the rainfall data was stationary. The values of d and D are equal to zero respectively.

During model identification, data was divided into two parts; in-sample (training data) and out-sample(test data) data. Data from January, 1983 to December, 2014 were used as in-sample data while the rest of data as out-sample data for the consistency and reliability of our model. The sample ACF and PACF plots were examined in figure 4 and figure 5 and were used in the identification of the values p, q, P and Q. For the non-seasonal part, spikes of the ACF at low lags were used to identify the value of q while the value of p was identified by observing the spikes at low lags of the PACF. For the seasonal part the value of Q was observed from the ACF at lags that are multiple of s while for P, the PACF was observed at lags that are multiples of S. Information criteria i.e. AIC and BIC were used in selectionof the best model. Looking at the ACF plot in figure 4 and PACF plot in figure 5 for stationary series, the following models were suggested:

Table 4: SARIMA Models Suggested for the Rainfall Data.

Model	AIC	BIC
$SARIMA(1,0,0)(1,0,0)_{12}$	2027.057	2038.909
$SARIMA(0,0,1)(0,0,1)_{12}$	2051.105	2062.967
$SARIMA(1,0,0)(0,0,0)_{12}$	2037.860	2045.762
$SARIMA(1,0,0)(2,0,0)_{12}$	1976.662	1992.465
$SARIMA(0,0,0)(2,0,0)_{12}$	1984.490	1996.342
$SARIMA(2,0,0)(2,0,0)_{12}$	1978.415	1998.169
$SARIMA(1,0;0)(3,0,0)_{12}$	1944.700	1964.460
$SARIMA(2,0,0)(3,0,0)_{12}$	1958.763	1982.467

The model $SARIMA(1; 0; 0)(3; 0; 0)_{12}$ was found to have the lowest AIC and BIC values and hence was selected as the best model SARIMA model. The maximum likelihood estimated parameters for $SARIMA(1,0,0)(3,0,0)_{12}$ model are presented in table 5.

Table 5: Parameter Estimates for $SARIMA(1,0,0)(3,0,0)_{12}$

Parameter	Estimate	Standard error
ar1	0.1232	0.0000
sar1	0.1401	0.0018
sar2	0.3405	0.0041
sar3	0.2706	0.0049

$SARIMA(1,0,0)(3,0,0)_{12}$ had an estimated variance of 8:935 with a log-likelihood of967:35. ItsAIC value was 1944:46 and the BIC value was 1964:46.

Table 6: Confidence Intervals of the Parameters.

Parameter	2.5%	97.5%
ar1	0.0000	0.0000
sar1	0.1366	0.1435
sar2	0.3325	0.3484
sar3	0.26103	0.2801

From the table 6 above, all parameters are significant and hence the model selected was best.

For a well fitted model, the standardized residuals estimated from the model should behave as an independently and identically distributed sequence with zero mean and constant variance and should be also randomly distributed. The ACF plot of the residuals suggests that the autocorrelation are close to zero. This result means that the residuals did not deviate significantly from a zero mean white noise process.

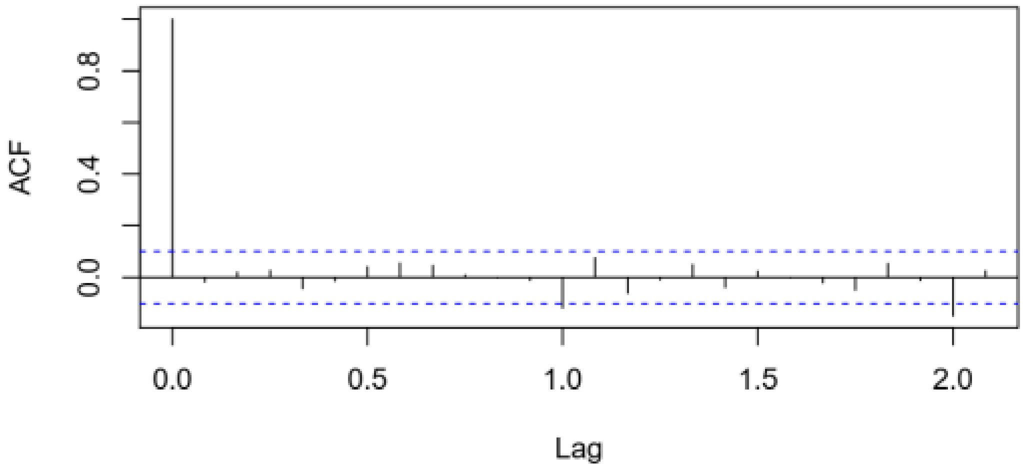


Figure 6: ACF of the Residuals.

The Ljung-Box test yielded a chi square of 9.561 with a p-value of 0.6544. From the Ljung-Box test, the p-value of $0.6544 > 0.05$ and this confirmed that SARIMA$(1,0,0)(3,0,0)_{12}$ was adequate forforecasting. The QQ plots as observed in figure 7 followed a normal line indicating residuals werenormally distributed with zero mean and constant variance. The results of the JarqueBera test yielded a chi square of 1415.12 with a p-value $< 2.2 \times 10^{-16}$. These results indicated residuals were normally distributed with mean zero and constant variance.

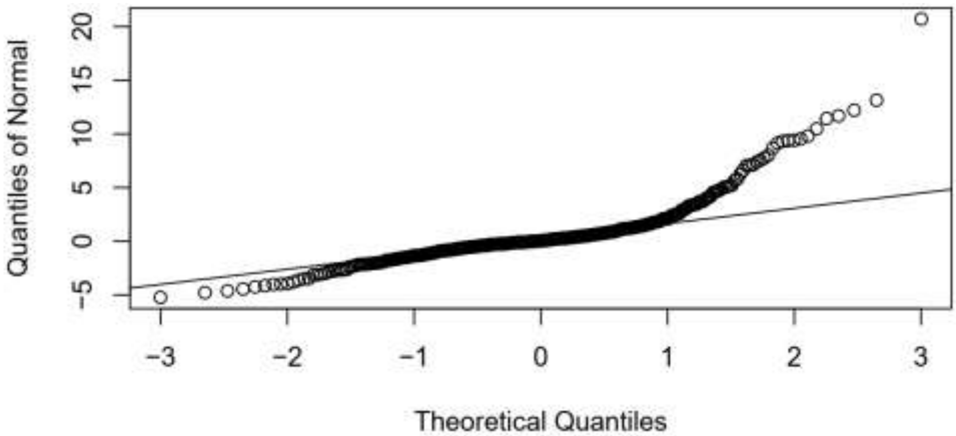


Figure 7: QQ Plot of the Residuals.

The adequacy and predictive ability of the chosen model, the actual data sets, predicted values, lower and upper limits are plotted and displayed in figure 8 and figure 9 shows that the predicted values are well fitted through the original data with the lower and upper limits containing majority of the data. This indicates that the model chosen for the rainfall data best fitted for the data set.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 11 Nov. 2024, pp: 327-342 www.ijaem.net ISSN: 2395-5252

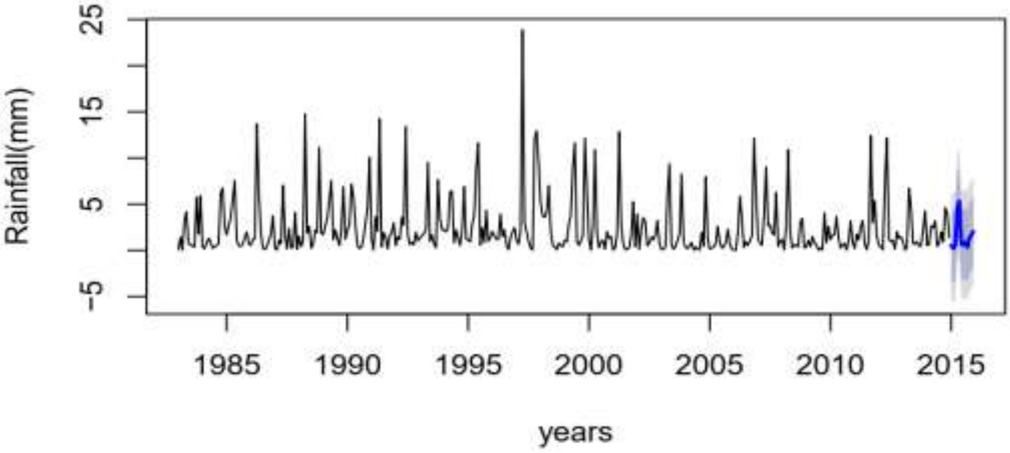


Figure 8: Predicted Values of Rainfall for a period of one year.

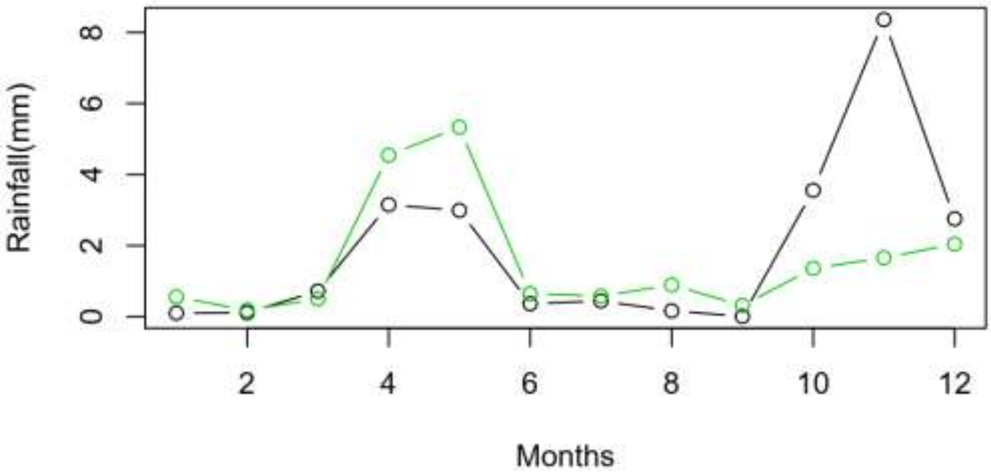


Figure 9: Observed and predicted Values of Rainfall.

Table 7 shows the observed values verses the predicted values and this affirms the adequacy of the chosen SARIMA model.

Table 7: Observed and Predicted Values of Rainfall Series

			95% Confidence Interval	
Time	**Observed**	**Forecast**	**Lower C.I**	**Upper C.I**
Jan 2015	0.0983000	0.5565037	0	6.415262
Feb (2015)	0.1088571	0.1841507	0	6.087219
Mar (2015)	0.7210323	0.5002575	0	6.403996
Apr (2015)	3.1496000	4.5421722	0	10.445921
May (2015)	2.9906452	5.3305226	0	11.234271
Jun (2015)	0.3640667	0.6428658	0	6.546614
Jul (2015)	0.4424516	0.5808448	0	6.484593
Aug (2015)	0.1638710	0.8923368	0	6.796085
Sep (2015)	0.0084700	0.3146306	0	6.218379
Oct (2015)	3.5560000	1.3572456	0	7.260994
Nov (2015)	8.3566000	1.6557721	0	7.559521
Dec (2015)	2.7530323	2.0408775	0	7.944626

This is because most predicted values lie within the lower and upper limits of 0.95 confidence interval except for one outlier 8.3566 which was the predicted rainfall of November 2015.

Forecasting helps in planning and decision making process since it gives an insight of the future uncertainty using the past and current behavior of given observations. From most research studies, the selected model is not always the best for forecasting unless further accuracy tests such as Root mean square error are done. The root mean square of the out sample data and the in-sample data was considered. The root mean square error of the in-sample (January, 1983 to December, 2014) and out-sample (January 2015, to December, 2015) are 54.5680 and 7.6579 respectively, implying that the selected model was adequate because the root mean square of out-sample is less than that for in-sample. Hence the selected model was adequate.

Best Model	**RMSE**	**MAPE**	**MAE**
SARIMA(1,0,0)(3,0,0)	2.9361	288.5118	1.1770

VII. CONCLUSION

Rainfall variability is an important aspect in study of climatic change. Therefore, modelling and forecasting the rainfall of a certain place has become an important and essential subject of study in climate change over the past recent years. This research paper explored the rainfall trend changes of Nyeri County over the years. A total of 396 monthly rainfall data points were considered in the study. From the monthly plot series the long rains were observed to occur in the months of April to June and the short rains were observed to occur in the months of October to November. The month of September was the driest month observed. However through the years, there is a distinctive decreasing trend of rainfall amounts proving the fact that climate change is in fact a reality.

Rainfall series showed high seasonality and the data was stationary. Several suggested SARIMA models were developed according to the procedure of the Box-Jenkins model building. The best SARIMA model was selected on basis of AIC and BIC. The model was SARIMA$(1,0,0)(3,0,0)_{12}$. The model residuals showed normality as most points fell on the straight line with few close to it. The residuals also confirmed white noise. From the model validation results, the predicted values are well-fitted through the original data with lower and upper confidence limits containing majority of the original data. The identified model was therefore suitable for predicting rainfall of Nyeri County.

The selected model gave a one year forecast that can help decision makers to establish priorities for equipping themselves against upcoming climate changes. The forecasts also show a continuous decrease in rainfall amount which is a reflection of changing climate in the whole country. Since Kenya is highly dependent on agricultural production, proper planning is required to avoid adverse impact on this sector which in turn leads to deteriorating the economy.

REFERENCES

[1]. Etuk, Ette Harrison and Tariq Mahgoub Mohamed (2014). “Time series analysis of monthly rainfall data for the Gadaref rainfall station, Sudan, by SARIMA methods”. In: Int. J. Sci. Res. Knowl. 2.7,p. 320. DOI: 14.9890/98728-11949984.

[2]. Etuk, Ette Harrison, Bartholomew AnuriobiUchendu, Mazi Yellow Dimkpa, and Others (2016).“Box-Jenkins Method Based Additive Simulating Model for Daily Ugx-Ngn Exchange Rates”.In: Acad. J. Appl. Math. Sci. 2.2, pp. 11–18. DOI: 10.9890/5728-116478794.

[3]. Jamaludin, AaishahRadziahBt, FadhilahYusof, and Ibrahim Lawal Kane (2015). “Temporal Dynam-ics of Trend in Relative Humidity with RH-SARIMA Model”. In: Mod. Appl. Sci. 9.3, p. 281. DOI: http://dx.doi.org/10.5539/mas.u9n3p281.

[4]. Kibunja, Hellen W, John M Kihoro, George O Orwa, and Walter O Yodah (2014). “Forecasting Pre-cipitation Using SARIMA Model: A Case Study of Mt. Kenya Region”. In: Math. Theory Model. 4.11, pp. 50–58. DOI: 10.9790/5728-11547784.

[5]. Mahsin, Md, YesminAkhter, and Monira Begum (2012). “Modeling Rainfall in Dhaka Division of Bangladesh Using Time Series Analysis.” In: J. Math. Model. Appl. 1.5, pp. 67–73. DOI: http://dx.doi.org/10.12983/ijrsk.2014-p0320-0327.

[6]. Metrine, Chonge, Nyongesa Kennedy, MulatiOmukoba, Makokha Lucy, TireitoFrankline, and Oth-ers (2015). “A time series model of rainfall pattern of UasinGishu County”. In: IOSR J. Math.

11.5, pp. 77–84. DOI: 10.9790/5728-11547784.

[7]. Otieno, George and Joseph Mungatu1 George Orwa (2014). “Time series modeling of tourist accom-modation demand in Kenya”. In: Math. Theory Model. 4.10, pp. 106–117. DOI: http://dx.doi.org/10.4236/ojs.2015.51008.

[8]. Salahi, B, A Nohegar, and M Behrouzi (2016). “The Modeling of Precipitation and Future Droughts of Mashhad Plain using Stochastic Time Series and Standardized Precipitation Index (SPI)”. In: Int. J. Environ. Res. 10.4, pp. 625–636. DOI: http://dx.doi.org/10.12988/ams.2014.410847.

[9]. Sampson, Wiredu, NasiruSuleman, and AsamoahYeboahGifty (2013). “Proposed Seasonal Au-toregressive Integrated Moving Average Model for Forecasting Rainfall Pattern in the Navrongo Municipality, Ghana”. In: J. Environ. Earth Sci. 3.12, pp. 80–85. DOI: 10.1080/10920277.2006.10596238.

[10]. Zhang, G Peter and Min Qi (2005). “Neural network forecasting for seasonal and trend time series”. In: Eur. J. Oper. Res. 160.2, pp. 501–514. DOI: https:/doi.org/10.1016/j.ejor.2003.08.037.