Vol 1(Iss. 1),pp.71-85,2024 African Scientific Annual Review www.asarev.ne

Predicting Rainfall Pattern in Kakamega County using Time Series

Joan Awuor[1]
Elvis Kemboi[2]
Benson Nzaro[3]
Hesbon Mocha[4]

[1] *awuorjoan2020@gmail.com*
[2] *elviskemboi70@gmail.com*
[3] *nzaroben20@gmail.com*
[4] *hesbonmocha55@gmail.com*

[1] *Department of Mathematics, Masinde Muliro University of Science and Technology, P. O. Box 190-50100, Kakamega, Kenya.*

ABSTRACT

Rainfall patterns play a critical role in shaping various aspects of our lives. Understanding the patterns, trends and predictability of rainfall is essential for effective planning and decision making in various aspects including agriculture, water resource management, disaster preparedness and social economic planning. In agricultural activities crops require specific amount of water at the right time for growth. By understanding the rainfall patterns, farmers can adapt their farming activities, optimizing irrigation strategies and make informed decisions. In the management, it help policy makers and management authorities for planning efficient water allocations and conservations measures. Therefore, in this paper we fit a time series model that best describes rainfall patterns of Kakamega county for the general ARIMA and generated the values of (P, D, Q) to forecast average expected monthly rainfall. Also we use R software for verification and data fitting of the model. The data we have used is from the Kakamega meteorological station in Kakamega.

Mathematics Subject Classification: Primary 55M10; Secondary 91G80
Keywords: Rainfall patterns, predictability, optimizing irrigation,

1 Introduction

The rainfall variability has been affected by a combination of natural and human factors. Therefore, in understanding these key factors is crucial for comprehending variations in rainfall patterns. Some of the factors includes climate change, topography and orographic effects, atmospheric circulation, vegetation and land use changes, human activities etc [16]. In climate change, the rising global tempratures alter atmospheric circulations patterns, leading to changes in the precipitation patterns [13]. This has been resulting in shifts in rainfall distribution intensity and frequency leading to more frequent droughts or intense rainfall events. Under Topography, the mountains and elevated terrain can trigger orographic uplift causing moist air to rise and cool leading to enhanced rainfall on the windward sides of the mountains. Conversely leeward side experiences a rainshadow effect resulting in reduced rainfall [5]. The human activities like irrigation and urbanization can modify local and regional rainfall patterns for instance increasing moisture availability leading to localized changes in rainfall. Also changes in land cover due to urbanization can affect rainfall patterns by altering the surface characteristics [17]. Therefore, in understanding these factors and interactions is crucial for accurately predicting and managing rainfall variations especially in the context of climate change and sustainable water resource management.

2 Literature Review

Most of the economies both in Kenya and Africa countries heavily relied on rain for agriculture. In Kenya, agriculture contributes 0.33 of the GDP, generates 0.6 foreign exchange earnings, provides employment to over 40 percent of the population and 70 percent of the rural population, provides raw materials to agro-industries [2]. Livestock contributes 42 percent of agricultural GDP and 12 percent of the total GDP [19]. Kakamega county lies in the western part of Kenya. There are mostly farmers growing Sugarcane, Maize, Beans, and Tea. Rainfall variability phenomenon in terms of the temporal aspects is the degrees to which rainfall amounts to changes at a given area through time, either month to month of, and season to season or year to year in relation to long term average [14]. The studies by [18] associate rainfall variability with floods, dry spell or drought. Similarly, rainfall variability can be identified as having a global effect on agricultural crop production [8]. The historical change in rainfall amounts at a seasonal and annual scales is an important variable in examining rainfall variability [4] and annual amounts in agricultural production. [11] indicates that most African economies are highly dependent on agriculture and adoption of modern technology is low, leading to poor agricultural crops returns. The studies by [7] link the local short rains to reduced crop yields despite the fact that increased rainfall amounts is always associated with high yields. A study by [6] shows that rainfall in Mumias Subcounty varies from season to season or year to year such that, between 1982 and 2012 the subcounty experienced an increase annual amount which had an effect on Sugarcane production. Under the production, by the line graph, it showed variability trend at seasonal scale which was positive for long rains (slope=2.52) and seasonal and annual rainfall were found were found to be 83.17 percent and 67.7 respectively implying a high level of temporal rainfall variability. In addition, the Kakamega climate risk profile shows the analysis over 35 years period (1981-2015) indicates that average rainfall had increased by over 15 percent in the first season and 30 percent in second season [3]. However, rainfall variability from year to year has also increased resulting in an increased risk and uncertainty of occurrence of floods and droughts. Both hazards have an increased impacts on agricultural production and livelihoods of the country inhabitants. According to [12] on climate change and seasonal agricultural drought, a simple Random Sampling (SRS) method has been used and SPSS software has been employed in analysis. The results showed that there was an evidence of climate change and seasonal agricultural drought in Kakamega South Subcounty where the study was employed [10]. Therefore it was recommended that in order to adapt to effects to climate change, there's was need to improve sustainability of crop production by supplementing rain fed farming with drip irrigation, rainwater gathering and greenhouse techniques. Methods of prediction of rainfall extremes have often been based on studies of physical effects of rainfall or on statistical studies of rainfall time series. Because rainfall occurs based on a specific time and there is a correlation between the previous data and subsequent ones, the best method for analysing rainfall data is using time series [18] revealing that a researcher with data for a past period can use Univariate Box-Jenkins method to forecast values without having to search for other related time series data. Montgomery and Johnson [13] considered the Box and Jenkins methodology the most accurate method for forecasting of time series. [9] in studying of drought, modelling and forecasting the precipitation of the Shiraz city of Iran, used three models; Box-Jenkins, Decomposition and Heat Winterz on precipitation for the period 1977 to 2010 [15]. The Box Jenkins approach was chosen as the most appropriate method for forecasting. [7] carried out a statistical analysis of rainfall pattern in Dire Dawa, Eastern Ethiopia. He used descriptive analysis, spectrum analysis, and univariate Box Jenkins method. He established a time series model that he used to forecast two years monthly rainfall.

3 Research Methods

3.1 The Moving Average (MA) model

Given:

$$x_t = a_0 u_t + a_1 u_{t-1} + ... + a_q u_{t-q} \tag{3.1}$$

Where u_t is a random process with mean zero and variance σ^2. We say that equation [1] is a Moving Average (MA) process of order q, commonly denoted as MA (q). CPI is the Consumer Price Index in Kenya at time t, $a_0 \ldots a_q$ are estimation parameters, u_t is the current error term while $u_{t-1} \ldots u_{t-q}$ are previous error terms. Hence:

$$x_t = a_0 u_0 + a_1 u_{t-1} \tag{3.2}$$

is an MA process of order one, commonly denoted as MA (1). Owing to the fact that previous error terms are unobserved variables, we then scale them such that a_0=1. Since:
E(u_t)=0
for all t,it therefore; implies that:
E(CPI_t)=0
and

$$Var(x_t) = (\sum_{i=0}^{q} a_t^2)\sigma^2 \tag{3.3}$$

where u_t is independent with a common varience σ^2 . Hence, we can now re – specify general equation as follows:

$$x_t = u_t + a_1 u_{t-1} + ... + a_q u_{t-q} \tag{3.4}$$

3.2 The Autoregressive (AR) model

Given:

$$x_t = \beta_1 x_{t-1} + ... + \beta_p x_{t-p} + u_t \tag{3.5}$$

Where $\beta_1 \dots \beta_p$ are estimation parameters, $x_{t-1} \dots x_{t-p}$ are previous period values of the data series and u_t is as previously defined. We say that equation is an Autoregressive (AR)process of order p, and is commonly denoted as AR (p); and can also be written as:

$$x_t = \sum_{i=1}^{p} \beta_{t-i} x_{t-i} + u_t \tag{3.6}$$

3.3 The Autoregressive Integrated Moving Average (ARIMA) model

ARIMA models are a set of models that describe the process (for example, x_t) as a function of its own lags and white noise process [1]. Making predicting in time series using univariate approach is best done by employing the ARIMA models (Alnaa and Ahiakpor, 2011). A stochastic process x_t is referred to as an Autoregressive Integrated Moving Average (ARIMA) [p, d, q] process if it is integrated of order "d" [I (d)] and the "d" times differenced process has an ARMA (p, q) representation. If the sequence $\triangle^d x_t$ satisfies and ARMA (p, q) process; then the sequence of x_t also satisfies the ARIMA (p, d, q) process such that:

$$\triangle^d x_t = \sum_{i=1}^{p} \beta_{t-i} \triangle^d x_{t-i} + \sum_{i=1}^{q} a_i u_{t-i} + u_t \tag{3.7}$$

where $\triangle$ is the difference operator.

3.3.1 Model Selection Criteria:

The parameters p,d and q can be obtained through:

1. parameter p will obtained in AR by looking at Autocorrelation Function (ACF) plot and choosing the lag where it crosses the significance threshold.
2. parameter d will be obtained by determining the number of differences needed to achieve the stationary.
3. parameter q will be obtained by looking at Partial Autocorrelation Function (PACF) plot and choosing the lag where it crosses the significance threshold.

The final model was selected using a penalty function statistics such as Akaike Information Criterion (AIC or AICc) or Bayesian Information Criterion (BIC). (Sakamoto, Ishinguro, and Kitagawa, 1986);(Akaike, 1974) and (Schwarz, 1978). The AIC, AICc and BIC are a measure of the goodness of fit of an estimated statistical model. Given a data set, several competing models may be ranked according to their AIC, AICc or BIC with the one having the lowest information criterion value being the best. These information criterion judges a model by how close its fitted values tend to be to the true values, in terms of a certain expected value. The criterion attempts to find the model that best explains the data with a minimum of free parameters but also includes a penalty that is an increasing function of the number of estimated parameters. Also some forecast accuracy test between the competing models can also help in making a decision on which model is the best. Minimum of free parameters but also includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages over fitting. In the general case, the AIC, AICc and BIC take the form as shown below:

$$AIC=2k - n\log(\frac{RSS}{n})$$

$$AICc=AIC + \frac{2k(k+1)}{n-k-1}$$

$$BIC=\log(\sigma_e^2) + \frac{k}{n}\log(n)$$

Where
k: is the number of parameters in the statistical model

RSS: is the Residual Sum Squares for the estimated model

n : is the number of observations

σ_e^2: is the error variance

3.3.2 The Box – Jenkins Methods

The first step towards model selection was to difference the series in order to achieve stationarity. Once the process was over,we then examine the correlogram in order to decide on the appropriate orders of the AR and MA components. It was important to highlight the fact that this procedure (of choosing the AR and MA components) is biased towards the use of personal judgement because there was no clear – cut rules on how to decide on the appropriate AR and MA components. Therefore, experience plays a pivotal role in this regard. The next step is the estimation of the tentative model, after which diagnostic testing followed. Diagnostic checking usually done by generating the set of residuals and testing whether they satisfy the characteristics of awhite noise process. If not, there would be need for model re – specification and repetition of the same process; this time from the second stage. The process may go on and on until an appropriate model is identified

4 Results

4.1 Data analysis

In this chapter we analyzed the data collected from Kakamega meteorological station to predict the rainfall pattern in Kakamega County. As shown in table above, the mean is positive. The skewness was

Descriptive	Statistic
Mean	2.967
Standard deviation	0.5902
variance	0.348
Skewness	0.373
Kurtosis	3.412

0.373 and the most striking feature was positive, implied that the data series was non – symmetric, for kurtosis was that it should be around 3 for normally distributed variables and in this analysis, kurtosis was found to be 3.412 Therefore, the data series implied that it is normally distributed. As shown in the

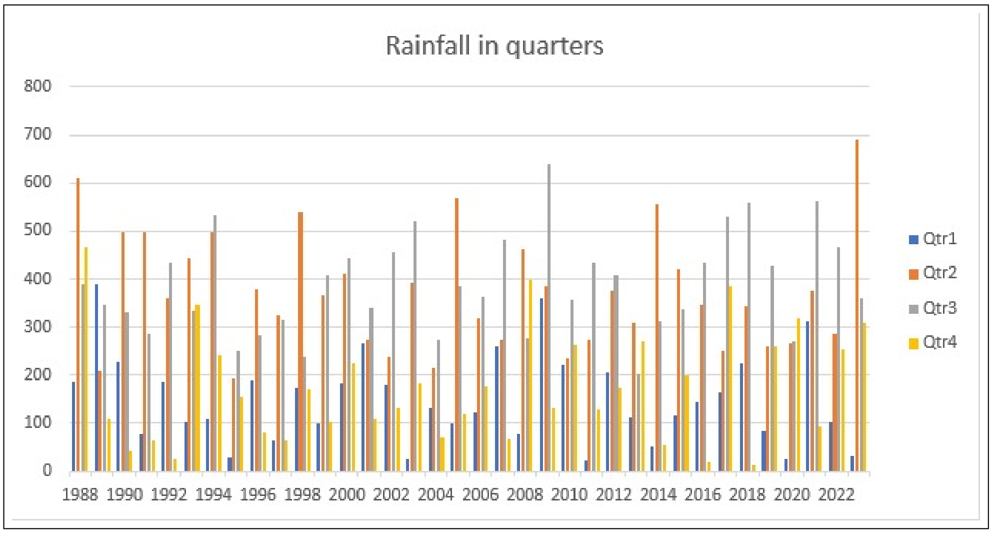

Figure 1:

figure above, rainfall is higher on second quarter which is on april may and june thus mostly farming should be done during the period. **Stationarity Tests: Graphical Analysis**
We examined whether a series was stationary or not by analyzing the time plot of average rainfall against

year. In that regard, a time plot of the data series is shown below:
The graph below shows that the data is not stationary since it is varying over the period under study. The implication is that the mean of data is changing over time and hence we can safely conclude that the variance of data is not constant over time.

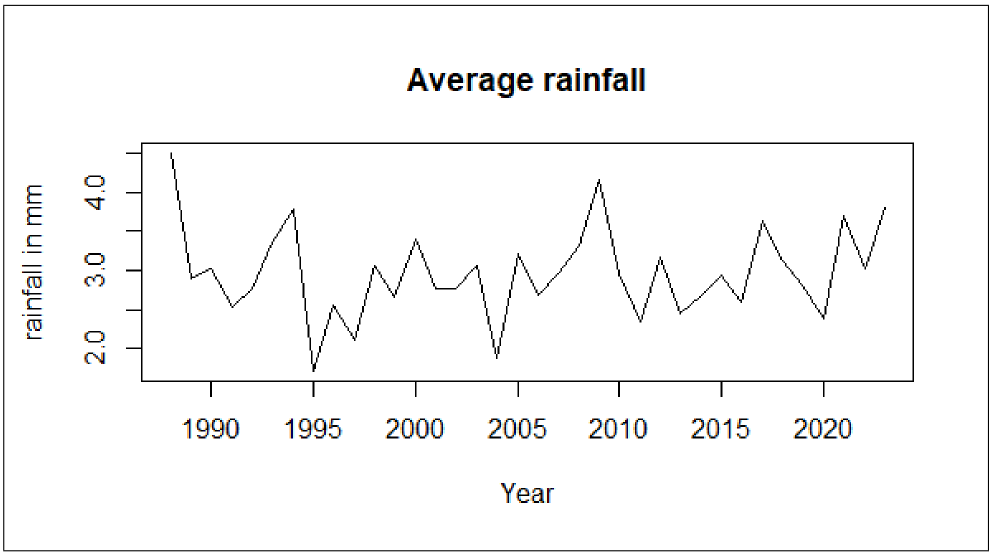

Figure 2:

The Correlogram

Autocorrelation Function (ACF) The ACF measures the correlation between a series and its lagged values. For non-stationary data, you may observe a gradual decline in the ACF plot, indicating that past values are still correlated with the current values but the correlation decreases as the lag increases. This suggested a lack of stationarity in the data, as the relationship between observations changed over time.

Partial Autocorrelation Function (PACF) The PACF shows the correlation between two variables while controlling for the effects of other variables in between. In non-stationary data, the PACF may exhibit spikes at the initial lags, indicating strong correlations that decay rapidly. This suggests that each observation may have a strong direct influence on subsequent observations, which is characteristic of non-stationary processes.
As shown in figure below, these patterns indicates the persistence of correlations over time and the lack of stable relationships between observations, both of which are indicative of non-stationarity.
Thus, the need for further analysis or data transformation techniques to address the non-stationarity before modeling or forecasting.

The ADF Test

The Augmented Dickey Fuller (ADF test) was used to check the stationarity of the data.

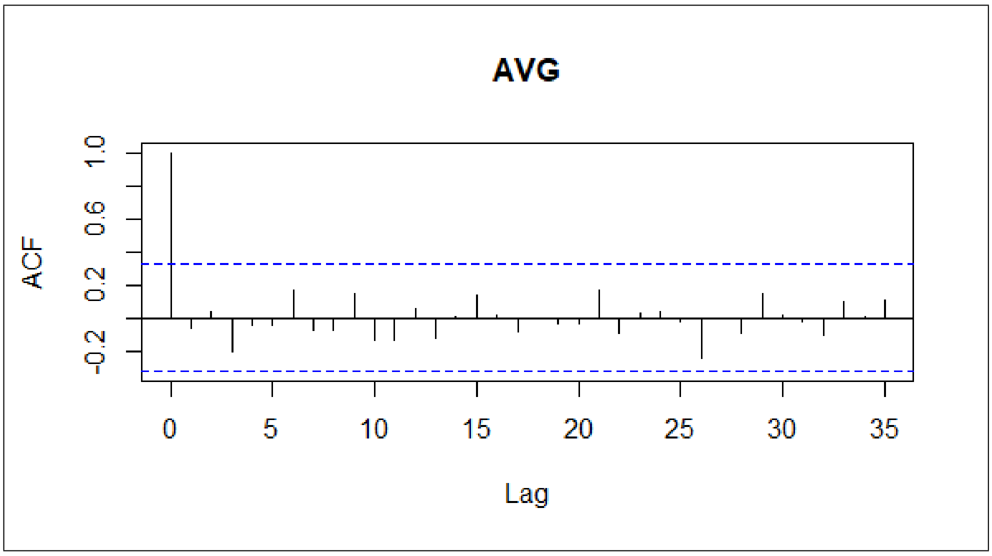

Figure 3:

The Augmented Dickey-Fuller (ADF) test was conducted to assess the stationarity of the data. The test statistic was found to be -3.5298, which exceeds the critical values at the 5% significance level. Additionally, the p-value associated with the test was 0.05439, which is greater than 0.05 therefore, we failed to reject the null hypothesis of non-stationarity, indicating that the data was non-stationary.

1^st **Difference**
The graph below shows a stationary series of data values after applying first differencing. Stationarity is evident from the absence of a clear trend or pattern over time. Fluctuations appear to be random around a constant mean level. **Graphical Analysis**

The stationary nature of the averange rainfall data series after first differencing indicates that the data is now suitable for time series analysis, such as forecasting or modeling. The absence of a trend suggests that inflationary or deflationary pressures may have stabilized during the period under consideration.

ADF test
The Augmented Dickey-Fuller (ADF) test was conducted to assess the stationarity of the data after first differencing. The test statistic was found to be -4.5195, which is more negative at the 5% significance level. Additionally, the p-value associated with the test was 0.01, which is less than 0.05. Therefore, reject the null hypothesis of non-stationarity, indicating that the data is stationary.

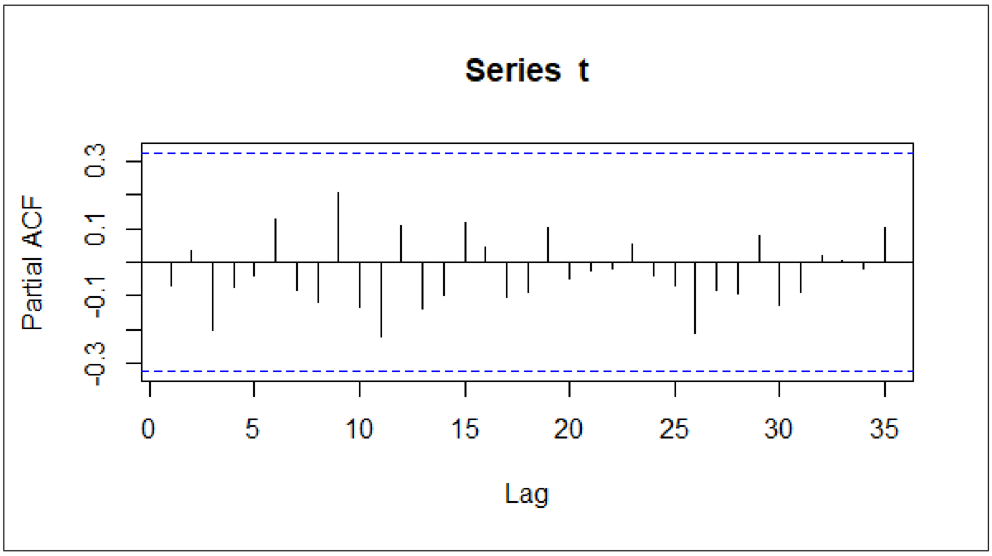

Figure 4:

4.2 Evaluation of various ARIMA models

In this study, we aimed to evaluate the ARIMA(p,1,q) model for a rainfall dataset. We first fited the model to historical rainfall data, tuned the model parameters 'p' and 'q', and then assess the model's performance using AIC and BIC and forecast accuracy.
By comparing AIC and BIC values across different ARIMA(p,1,q) models, we aim to identify the optimal combination of autoregressive and moving average terms that effectively capture the underlying inflation dynamics. Lower AIC and BIC values indicate better fitting models, with preference given to models that strike a balance between explanatory power and simplicity. As shown in the table above, the ARIMA

Model	AIC	BIC
ARIMA(1,1,0)	77.667	80.777
ARIMA(0,1,0)	86.5473	88.1026
ARIMA(2,1,0)	78.6536	83.497
ARIMA(0,1,3)	73.786	80.007
ARIMA(1,1,4)	76.67	86.001
ARIMA(1,1,3)	74.955	82.732
ARIMA(1,1,1)	**71.924**	**76.59**
ARIMA(3,1,1)	74.24	82.014

(1, 1, 1) model has the lowest AIC value and again it has the lowest BIC value making it to be the best model.

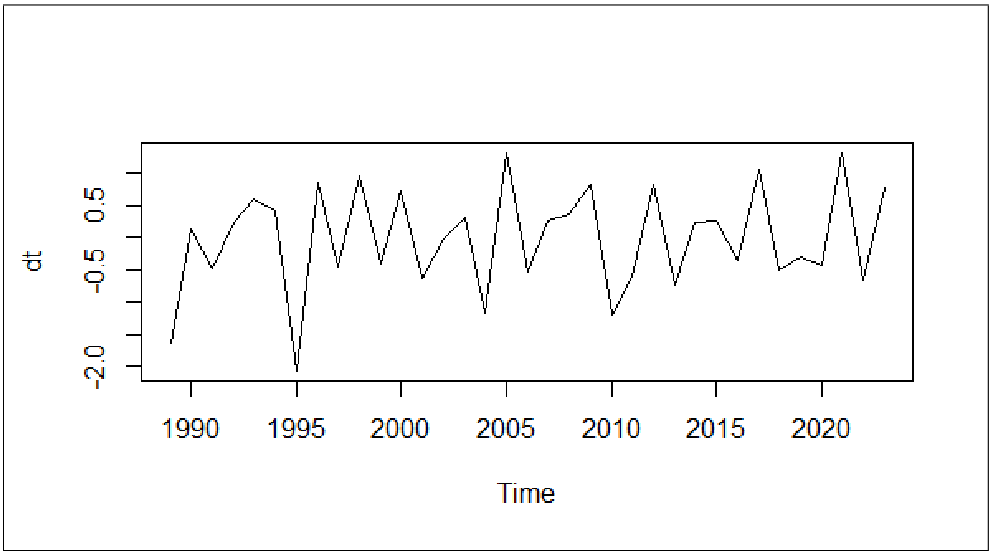

Figure 5:

4.3 Stability Test of the ARIMA (1, 1, 1) Model

In this context, examining the stability of the ARIMA(1,1,1) model with a unit polynomial becomes particularly important. Stability here refers to the model's ability to maintain consistent behavior over time and across different conditions. Specifically, it entails ensuring that the model's parameters remain within reasonable bounds and that the model produces reliable forecasts that align with observed data.

The unit polynomial in the ARIMA(1,1,1) model implies that the original time series has undergone differencing to remove trends or non-stationarity. This transformation is crucial for making the data suitable for modeling using ARIMA techniques. However, it also introduces potential challenges related to stability, as differencing can sometimes amplify noise or introduce unwanted artifacts into the series. The figure above indicates that the ARIMA (1, 1, 1) model is also stable since the corresponding inverse root of the characteristic polynomial is in the unit circle.

4.4 Findings and Discussion

Variable	Coefficient	Standard error	z	p¿(z)
AR(1) [β_1]	-0.0537	0.1964	-2.91	0.0000
MA(1) [α_1]	-1	0.1113	-5.09	0.0000

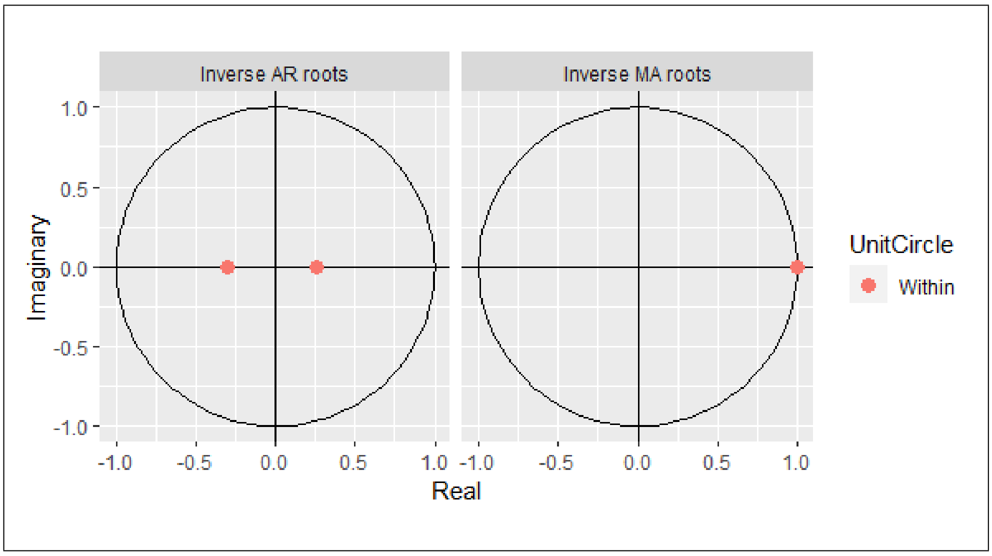

Figure 6:

4.4.1 Interpretation

After analyzing the data, we found that there was significant difference of the amount of rainfall in different periods of a year. Interpretation: The results of study indicate that there was relatively low amount of rainfall for the first quartile of the year and then from April to June there was relatively increase in rainfall and then from July there was slight drop of the amount of rainfall as compared to the second quartile but relatively higher than the first quartile, in the last quartile there was moderate amount of rainfall.

4.4.2 Implications

1. The findings suggested that the farmers should choose drought resistance crop varieties that require less water and are well adapted to dry conditions in first quartile and the last quartile.
2. The farmers were encouraged to implement efficient water management techniques such as drip irrigation to optimize the use of available water resources and minimize water wastage.
3. we encouraged the farmers to adjust on planting schedule to coincide with periods of expected rainfall.

4.4.3 Forecast Graphs

The forecast graph generated by an ARIMA(1,1,1) model for rainfall over the next five years, starting from 2023, offers a comprehensive insight into the anticipated trends and patterns in precipitation. This graphical representation encapsulates a wealth of information crucial for understanding and planning for

future conditions.
Displayed along the x-axis is the timeline, spanning from 2023 to the end of the forecast period five years later. On the y-axis, the predicted rainfall values are depicted, providing a projection of expected precipitation amounts for each corresponding time point.
As indicated on the graph, rainfall amount might remain constant over period of time with slightly variation

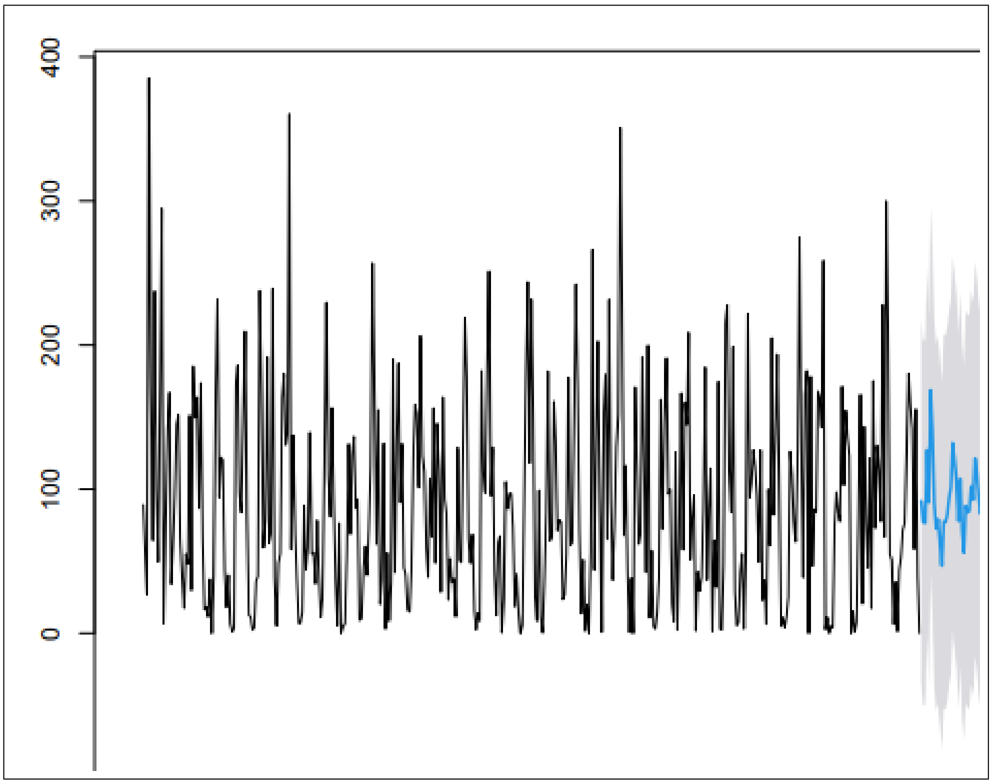

Figure 7:

5 Conclusion and Recommendation

5.1 Conclusion

In conclusion, this paper provided valuable insights into the rainfall pattern of Kakamega County through vigorous analysis, we had uncovered the best period for farming practices. These findings had significant implications for farmers. However, it was important to acknowledge the limitations of the study such as:

1. Nonstationarity -that made it difficult to predict future outcomes because past patterns could not continue to hold and makes modelling complex.
2. Missing values -that lead to loss of information potentially lead to an incomplete understanding of the underline phenomenal studied.
3. The package used to analyze -some R studio plug in were not compatible with R software that made analysis complex and time consuming.

Future research could explore machine learning algorithms for more accurate predictions and exploring the impact of urbanization and land-used changes on local rainfall pattern.
Recommendations

1. Develop machine learning algorithm to improve the accuracy rainfall predictions.
2. Investigate long-term trends in rainfall variability over different time space.
3. Implement community-based rainfall monitoring initiatives to collect local data to mitigate the impacts of changing rainfall pattern.

References

[1] Box, G. and G.Jenkins, (1976). Time series analysis: forecasting and control. (Revised ed) Holden day, San Francisco.

[2] Brockwell,P.J. & Davis, R.A. (1996). Introduction to Time series and forecasting. (2nded.). New York: Springer.

[3] Chatfield, C.,(1991). The Analysis of Time Series. An Introduction, (4thed)., Chapman and Hall: London.

[4] Cromwell, J.B., Labys, W.C and Terraza, M., (1994). Univariate Tests for Time Series Models. A Sage Publication, (7-99)96., London.

[5] Diebold, F.X., Kilian, L. and Nerlove, M., (2006). Time Series Analysis. Working Paper No.06- 011, University of Maryland, College Park.

[6] Enders, W., (1995). Applied Econometric Time Series, John Wiley and Sons, Inc., New York.

[7] Enock, M.A., Buckman, A.& Seth, O.L (2013); Time series modeling of rainfall in New Juaben municipality of the Eastern region of Ghana.

[8] Ette, H.E and Tariq M.M (2014). The special issue on contemporary Research in business and social science. Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by sarima methods, 4(8)114.

[9] Fatemeh F., Hossein N. & Mahmood K. (2013). International Journal of Scientific Research in Knowledge.2(7)320-327. Studying of drought, modeling and forecasting The precipitation of Shiraz city in Iran.

[10] Gibbons, R.D.,(1994). Archives of Applied science research. Statistical Methods for Ground water Monitoring. John Wiley & Sons, New York. 5 (3)173183.

[11] Jennifer, M.A (2011); Seasonal prediction of African rainfall with focus on Kenya. Phd thesis, Submitted to the Department of space and climate physics. University college London.

[12] Mahsin, M.D.,Yesmin, A. & Monira, B. (2012); Modelling Rainfall in Dhaka division of Bangladesh using Time series analysis. Journal of mathematical modeling and application.1(5) 67-73.

[13] Meenakshi, S. S and Lakshmi, M. (2014); Rainfall Prediction using Seasonal Auto Regressive Integrated Moving Average model. Indian Journal of Research 3(4)58-60.

[14] Montgomery, D.C. & Johnson, L.A. (1976); Forecasting and Time series Analysis. New York: Wiley.

[15] Naill P.E and Momani M., (2009); Time series analysis model for rainfall data in Jordan; case Study for using Time Series Analysis, American journal of Environmental sciences.5(5)599 604.

[16] Ndetei, C.J. (2013); Use of polynomial fit to predict seasonal rainfall in Nairobi, Kenya.University of Nairobi, Kenya. Msc thesis, submitted to Department of Meteorology UoN. Unpublished.

[17] Ngaira, Josephine k. (2005); Implications of climate change on the management of Rift Valley Lakes in Kenya. The case of Lake Baringo. Proceedings of the 11thworld Lakes conference, Nairobi. Kenya. 2(1)133-138.

[18] Osabuohien & Irabor, O. (2013); Applicability of BoxJenkins SARIMA model in rainfall Forecasting; port Hacourt south Nigeria. Canadian journal in computing mathematics, Natural Sciences, Engineering and Medicine.4(1)1-4.

[19] Raphael .W. K (2012); Some aspects of the geography of Kenya. Moi University. Unpublished.