

Modeling Effects of Climatic Variables on Tea Production in Kenya Using Linear Regression Model with Serially Correlated Errors

Consolata A. Muganda^{1*}, Sewe Stanley¹ and Winnie Onsongo²

¹Department of Mathematics and Actuarial Science, The Catholic University of Eastern Africa, Kenya.

²Department of Statistics and Actuarial Science, University of Ghana, Ghana.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v13i230306

Editor(s):

(1) Dr. Manuel Alberto M. Ferreira, Lisbon University, Portugal.

Reviewers:

(1) Gabriel Gomes de Oliveira, University of Campinas (Unicamp), Brazil.

(2) Asena Tugba Evren Subasi, Turkey.

Complete Peer review History: <http://www.sdiarticle4.com/review-history/69257>

Received: 22 May 2021

Accepted: 03 June 2021

Published: 07 June 2021

Original Research Article

Abstract

Aims/ Objectives: To formulated a linear regression model to capture the relationship between tea production and climatic variables in terms of ARIMA.

Place and Duration of Study: Department of Mathematics and Actuarial Science, Catholic University of Eastern Africa, Nairobi, Kenya, between June 2019 and April 2021.

Methodology: The study used time-series data for mean annual temperature, mean annual rainfall, humidity, solar radiation, and NDVI, collected from six counties, namely Embu, Kakamega, Kisii, Kericho, Meru, and Nyeri.

Results: The study findings noted that there is a presence of trend and seasonality for all the data. The scatter plot matrix for all the climatic variables for all the counties under the study indicated that tea production has a linear relationship with most climatic variables. Model fit of the data indicated statistical significance when tea production data is differenced. A second

*Corresponding author: Email: achiengpoly@outlook.com;

linear model with tea production data deseasoned has mixed results in terms of a significance test. The variation of independent variables with tea production yielded very low values, suggesting that the data used has many variabilities.

Conclusion: The study findings show the climatic variables can be used to forecast tea production.

Recommendation: Future studies may combine the analysis with other statistical modeling procedures such as the GARCH models.

Keywords: Climatic variability; Time-Series; ARIMA.

2010 Mathematics Subject Classification: 53C25; 83C05; 57N16.

1 Introduction

The climatic variability changes in the world impact the growth of cash crops, and Africa is no exception [1, 2, 3]. The climatic conditions in Kenya, specifically those in the tea zone areas, have continued to change and have attracted considerable research [4]. In developing countries like Kenya, climate change drastically reduces farmers' income [5]-[19]. They rely on the major cash crops like tea, whose total production quantity decreases due to variability in climatic conditions [20].

Globally, many papers have investigated the effect of climate variability on agriculture, and the findings indicate mixed outcomes [21]. While most research suggests that climatic variability negates the influence of crop production, others conclude oppositely [22]. The major controversy has been that global warming improves the climatic conditions in many regions, which favors crop yields [23].

Food security's global concern has prompted scholars to identify long and short-term drivers such as climatic variables [24]. Scholars have used modeling to unravel the dynamism associated with agriculture and climatic variations [25]. Thus, modeling has quantified the extent of uncertainty incoherent with existing knowledge [26]-[36].

Several studies on crops yield variability due to variation in climatic conditions proposed using various modeling tools [37, 38]. Majority of the methods are categorized into statistical and soft computing techniques [39, 25, 40, 41]. The statistical techniques are exponential smoothing, autoregressive integrated moving average (ARIMA), and generalized autoregressive conditional heteroskedasticity (GARCH), which focuses on volatility [42, 43]. The ARIMA (Box-Jenkins) model is commonly used in analysis and forecasting and extensively used in time series. Thus it has found numerous usage as an efficient forecasting technique in social science [43]. ARIMA models do not assume any underlying model or relationships, thus being extensive in forecasting the time series where the data cases are uncertain [44]-[51]. ARIMA models' strength relies on its use of the past and previous values of series and error terms for forecasting, respectively [52]. [43] noted that ARIMA models are also more robust and efficient when compared to complex structural models to short-run forecasting techniques.

Kenyan tea is one of the globally recognized brands in the global markets. Tea is Kenya's most valuable exports and contributes to 4% of the Gross Domestic Product (GDP) and 26% of its export earnings. Besides, Kenya is the largest exporter of tea in the world [53]. Tea production in Kenya has increased by 18% between 2015 and 2016. The value of tea production has declined by 1.6% between 2016 and 2015. Rigden [54], Muoki [53] noted that the decline is due to variation in climatic conditions. Nonetheless, [53] posited that the unstable trends in tea production in Kenya recently

have been due to climatic-driven stresses. Thus, in this study, we propose to use statistical modeling tools, specifically linear regression, and ARIMA to understand the relationship between climatic variability and tea production. Thus, the study's main objective is to model the effect of climatic variables on tea production in Kenya using a linear regression model with serially correlated errors.

The study has two main contributions.

1. Formulate a linear regression model to capture the relationship between tea production and climatic variables in terms of serially correlated errors.
2. Model fitting and estimation using Linear regression plus ARIMA.

The rest of the paper is outlined as follows. Section 2 presents the statistical model of the study with a focus on the linear regression model with ARIMA. Section 3 presents data analysis findings and results. Section 4 discusses the results of the findings. Section 5 presents a conclusion of the research.

2 Statistical Models

2.1 Climatic modelling data

Temperature, rainfall, humidity, NDVI, solar radiation, and agriculture production are collected over time. They, therefore, are referred to as time-series data. Time series data is a sequence of observations that varies over time [55]. The tea bushes grow for an extended period before uprooting [56], and therefore, the yield of a new year still depends on what was there the previous year. Due to variation in yield every year, the present paper proposed using the ARIMA model. The method help simulate the expected effect of various long-term climatic scenarios on future tea productivity.

The paper uses data from the Kenya Meteorological Department (KMD) and tea yields data from the Kenya Tea Development Agency (KTDA) covering 9 years, from 2007 to 2015. KTDA manages a network of tea factories in Kenya. Regression analysis in this paper serves to show the relationship between environmental factors and tea production. The regression analysis also indicates scatterplots of yields on various climatic parameters.

2.2 Time series, linear regression, and ARIMA

2.2.1 Time series concepts

A time series is a set of observations, each one recorded at a specific time. Time series concepts have been widely presented by numerous scholars [57]. There exist two categories of time series: discrete; and continuous time series. Discrete-time series data is vastly discussed by [58].The data used in this paper is of continuous-time series form. Continuous-time series is widely discussed by [59]. Stationarity is the foundation of time series analysis [60]. A time series $\{y_t\}$ is strictly stationary if the joint distribution of $y_{t_1} \dots y_{t_k}$ is identical to that of $(y_{t_1+t} \dots y_{t_k+t})$ for all t , where k is an arbitrary positive integer and $t_1 \dots t_k$ is a collection of k positive integers. The rest of information on stationarity is vastly discussed by Matteson [61].

Seasonality on the other hand, is the presence of variations that occur at specific regular intervals less than a year, such as weekly, monthly, or quarterly [62]. Seasonality components within a time series data entail seasonality adjustment or deseasonalizing removes these components. The process yields seasonal stationary and failure non-stationary. A vast discussion on seasonality concepts are discussed by Mann [63].

2.2.2 Linear regression modelling

Regression analysis helps us understand how the typical value of the dependent variable (or criterion variable) changes when any one of the independent variables is varied [64]. In contrast, the other independent variables are held fixed. Thus, it provides a reasonable basis for estimating the cost and duration. The case of one explanatory variable is called Simple Linear Regression, which is for $y = \beta_0 + \beta_1 x_1 + \epsilon$. y is the tea production in tonnes, x_1 is any of the climatic variables, β_0 is the y-intercept or the production when no contribution of climate variables is considered, β_1 is the slope of the regression line and ϵ is the error term.

For more than one explanatory variable, the process is called Multiple Linear Regression, and in this, if y is dependent variable and $x_1, x_2 \dots x_k$ are independent variables. The multiple regression model predicts y from the x_i of the form (2.1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon, \quad (2.1)$$

where $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$ the deterministic portion of the model and ϵ is the random error. β_0 is the constant term, β_1 to β_k are the coefficients relating the k explanatory variables to the variables of interest. Simple Linear Regression can be thought of as a special case of Multiple Linear Regression, where $k = 1$. The term 'Linear' is used because, in Multiple Linear Regression, we assume that y is directly related to a linear combination of the explanatory variables, assumes that the residuals are normally distributed and that the independent variables are not highly correlated.

2.2.3 ARIMA

ARIMA models are, in theory, the most general class of models for forecasting time series [65], which can be made stationary by differencing (if necessary). A random variable that is a time series is stationary if its statistical properties are constant over time. A stationary series has no trend [66]. Its variations around the mean have a constant amplitude; that is, its short-term random time patterns always look the same in a statistical sense. The ARIMA forecasting equation for a stationary time series is a linear regression type equation. The predictors consist of lags of dependent variables or lags of the forecasting errors. ARIMA stands for Autoregressive Integrated Moving Average. Lags of the stationarized series in the forecasting equation are called 'autoregressive' terms, lags of forecasting errors are 'Moving Average' terms a time series which needs to be differenced to be made stationary is said to be 'Integrated.' A non-seasonal ARIMA model is classified as an ARIMA (p, d, q) . p is the number of autoregressive terms. d is the number of nonseasonal differences needed for stationarity. q is the number of lagged forecast errors in the prediction equation. Equation (2.2) shows the general forecasting equation.

$$\hat{y}_t = \mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}, \quad (2.2)$$

where ϕ_i are the coefficients for AR term and θ_j for MA.

2.3 Linear regression model with ARIMA

Linear regression with ARIMA combines two powerful statistical models, Linear regression and ARIMA, to form a single super-powerful regression model. The purpose of the combination is to forecast time series data. The procedure is summarized in Fig. 1.

2.3.1 SARIMAX

SARIMAX illustrated modeling response regression variables in the matrix X and dependent variable in vector y . The results suggest that for each time step i , there is y_i with a corresponding x_i . SARIMAX modelling is widely discussed and illustrated by the following references [67, 68, 69, 70].

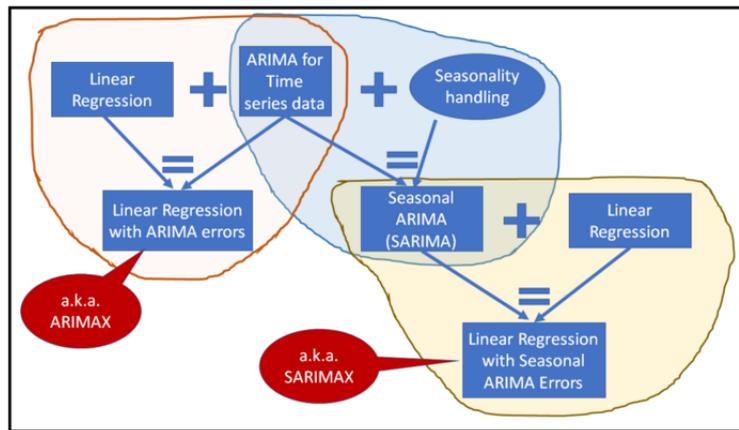


Fig. 1. The procedure of linear regression model with ARIMA

Source: [5]

2.3.2 ARIMA

The model utilizes the ability of the existence of auto-correlations. The model is a combination of AR(p) and MA(q). The ARIMA model suits forecasting time series data. One primary assumption is the lack of no auto-correlation in residual errors. The details of ARIMA model is already discussed in earlier sections 2.2.3.

2.4 Climatic variables on tea production in Kenya using linear regression model with ARIMA errors

Equation (2.3)-(2.4) present the general form of linear regression model with ARIMA for the proposed study.

$$y_t = a + b_l x_{it} + \epsilon_{it}, \text{ for } i = 1, 2, \dots, k \quad (2.3)$$

where

$$\epsilon_{it} = \sum_{i=1}^p \phi_i \epsilon_{t-i} + \omega_t + \sum_{j=1}^q \theta_j \omega_{t-j}. \quad (2.4)$$

y_t is the mean annual tea production. x_{it} are the possible predictors such as the Mean annual temperature, mean annual rainfall, humidity, solar radiation, Normalized Difference Vegetation Index (NDVI). a and b_i are the linear regression coefficients. The error term is modeled using an ARMA (p,q) process. The ϕ'_i s are the coefficients for the AR term and θ'_j s for MA.

3 Data Analysis and Findings

3.1 Study area

Kenya is located on the equator and has a mostly semi-arid tropical climate with steppe and semi-desert in the low lying areas and mountainous forests at the higher altitudes. The climate of Kenya varies by location, from mostly cool every day to always warm/hot. Tea growing regions in Kenya have an ideal climate, volcanic red soils, and well-distributed rainfall ranging from 1200mm to 1400

mm yearly that alternates with long sunny days [71]. Production goes on around the year with two main peak seasons of the high crop between March and June and October and December, which coincide with the rainy season. The main tea growing areas in Kenya are within highland regions on both sides of the Great Valley [72] and astride the equator within altitudes 1500metres 2700metres above sea level [73]. The regions include areas around Mt. Kenya, the Aberdares and Nyambene hills in Central Kenya, and the Mau Escarpment, Kericho highlands, Nandi and Kisii Highlands the Cherangani Hills.

3.2 Data Analysis

The general technique adopted in this paper is the ordinary least square (OLS) to estimate parameters. This was achieved via R-programming. Kikawa *et al.* [74] discusses this technique widely. The time-series data collected on climatic variables (rainfall, NDVI, Maximum and Minimum Temperature, Maximum and minimum humidity, solar radiation), and their respective trend plots and seasonality presented in Fig. 2a.

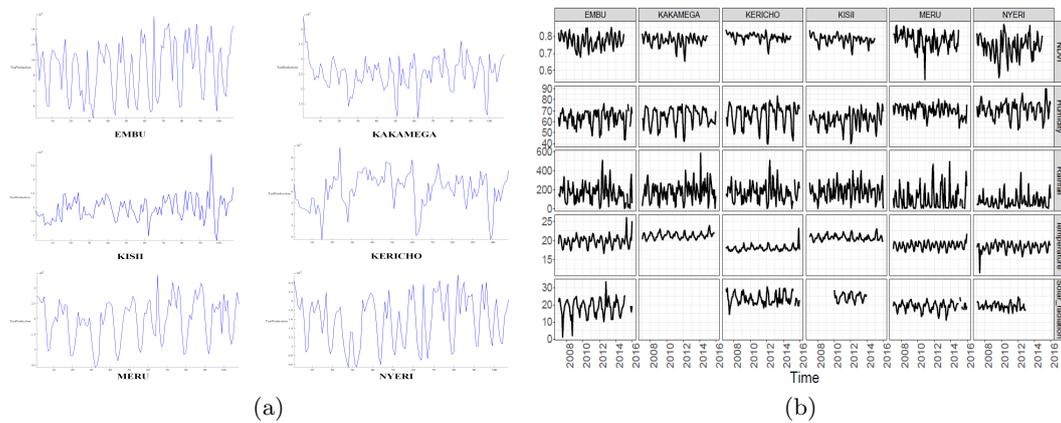


Fig. 2. (a) Time series plot for Tea production data, from left to right, first row, (Embu and Kakamega), second row (Kisii and Kericho), third row, (Meru and Nyeri) between year 2007 and 2016 collected monthly. (b) Time series plot for all the climatic variables assumed to have impact on tea production from all the selected counties between year 2007 and 2016 collected monthly. The first to the last column is Embu, Kakamege, Kericho, Kisii, Meru and Nyeri. The first to the last is NDVI, average humidity, rainfall, average temperature, and solar radiation in terms of rows. Kakamega county had no solar radiation data, while Kisii had missing data on the same variable

Fig. 2a indicates a trend and seasonality. The trend indicates production throughout the year, with high production in some months and low in others. The plot shows the highest production was in 2012 and the lowest 2009. Fig. 2a indicate some little trend and a bit of seasonality throughout the production period presented. There are some months with low production. The general trend shows that many of the months 2007 had the majority of the highest production. The plot shows the highest production was in 2008 and the lowest 2012. Fig. 2a indicate that neither trend nor seasonality is exhibited in the time series plot for Kisii county. The trend shows that tea production is neither low nor high throughout the period, though the production was exceptionally high in 2015. The plot shows the highest production was in 2015 and the lowest 2015. Unlike in the previous trends, this trend shows the highest production was followed by the lowest.

Table 1. Summary statistics for the Linear models. Akaike information criterion (AIC) estimate the prediction error and hence show relative quality of statistical models for a given set of data [75]. Bayesian information criterion (BIC) helps in selection of the true model [76]

	Adjusted R^2	Residual standard error	P-value	AIC	BIC
Fit1	0.7159	0.2112	$< 2.2e^{-16}$	-78.154	-44.127
Fit2	0.7173	0.2107	$< 2.2e^{-16}$	-81.712	-55.247

Table 2. Summary statistics for Linear regression plus residual models

Model	AR	MA	AIC	BIC	AIC_c	standard error
ARIMA (0,0,1)	-	0.536	-0.4865	-0.3932	-0.4854	0.1849
ARIMA (1,0,1)	0.9011	-0.0451	-0.7963	-0.6913	-0.7949	0.1578
ARIMA (1,1,0)	-0.0771	-	-0.7629	-0.6811	-0.7621	0.1616
ARIMA (1,1,1)	0.5381	-0.8796	-0.8890	-0.7954	-0.8879	0.1513

Fig. 2a shows both trend and seasonality present in the time series plot for Kericho county. Most of the months have very high production and just a few months with low production. Kericho county produces the highest volume of tea in Kenya. The plot shows the highest and the lowest trend was in the same year, 2008. Fig. 2a shows time series for Meru county. The time series plot shows trends and seasonality. Throughout the entire study period, tea production is with subsequent monthly very high and low production monthly. The plot shows the highest production was in 2012 and the lowest 2009. Fig. 2a shows time series plot for Nyeri county. The plot shows trends and seasonality. Production is throughout the year with subsequent high and low production. The plot shows the highest production was in 2014 and the lowest 2009.

Fig. 2b shows time series plots of all the climatic variables for all the counties. The plots show the trends and seasonality between 2007 and 2016. Embu, Kakamega, Kericho, and Kisii counties experienced high NDVI during the entire study period, while Meru County and Nyeri County experienced some very low values of NDVI in the year 2010. All six counties experienced both high and low humidity. Nyeri county experienced the highest humidity among the six counties in the year 2015 while Kericho county experienced the lowest humidity in 2012. There was rainfall throughout the entire study period in the six counties, with Kakamega county getting the highest rain in the year 2014.

3.2.1 Analysis for the three counties that have data on solar radiation

The data collected had some variables such as solar radiation with missing data. We segmented data analysis based on the counties with all the data and with missing data. The first analysis dealt with the counties that had data for all variables, including solar radiation. These counties are Embu, Kericho, and Meru. From this first analysis the following results were obtained.

Table 1 show Fit 1 was a linear model with all the climatic variables. However, NDVI and Maximum Temperature were found not to be statistically significant. Fit 2 was a linear model with all climatic variables except NDVI and Maximum Temperature. The results show fit 2 was a better fit since it had a smaller AIC, BIC, residual standard error, and 71.73% of the total variation in Tea Production explained by the independent variables compared to 71.59% in fit 1. We found residual from Fit 2 to be autocorrelated.

Fig. 3a shows a plot of the autocorrelation function (ACF) of the residuals showing the presence of a pattern. The ACF shows a spike at lag 12 and lags 24, indicating the presence of seasonality. The Durbin-Watson test showed the p-value was 0 less than 0.05, so H_0 was rejected, which means there was an autocorrelation between the errors. The ACF of the squared residuals had one significant spike, while the partial autocorrelation function (PACF) had no significant spikes. Therefore, we fitted MA(1) to the squared residuals.

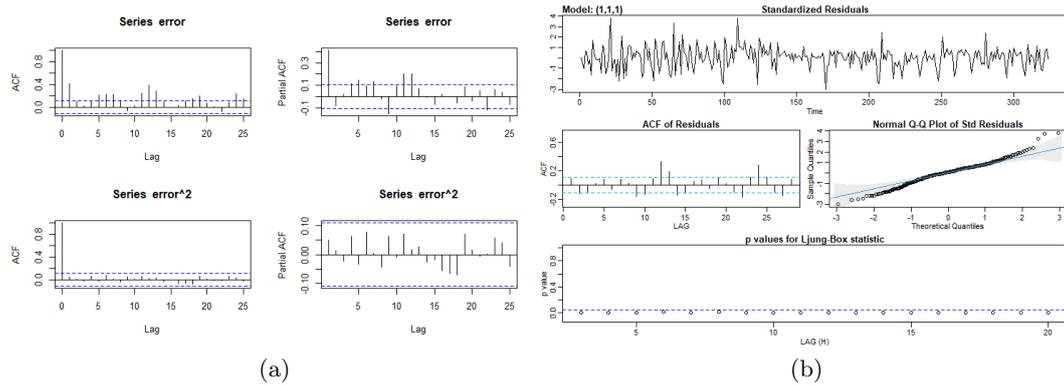


Fig. 3. (a) ACF and PACF of residuals and squared residuals from linear regression model. (b) Plots of residuals of Fit 2 and ARIMA (1,1,1)

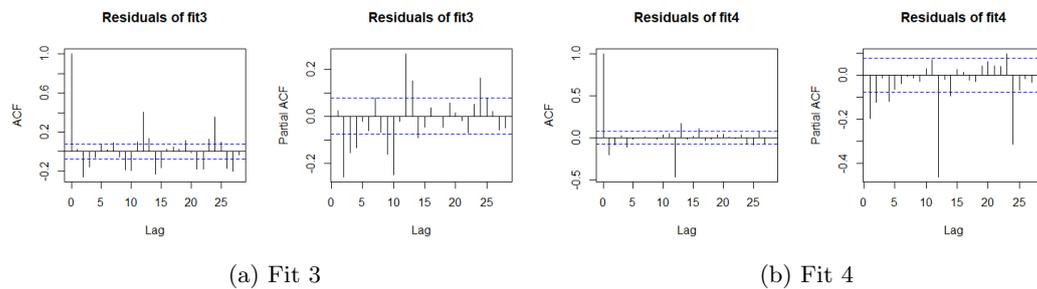


Fig. 4. Plots of ACF and PACF of residuals of fit 3 and fit 4 from linear regression model

We fit the best linear regression model fit 2 while factoring in various residual models. The residual models considered were ARIMA(0,0,1), ARIMA (1,0,1), ARIMA (1,1,0) and ARIMA (1,1,1). The results obtained were summarized as shown in the Table 2. Table 2 show ARIMA (1,1,1) has the smallest AIC, BIC and standard error values and therefore Fit2+ARIMA (1,1,1) model is the best fit. Equation 3.1 show the linear regression model obtained.

$$y = 0.0002R + 0.0002H_{min} + 0.004H_{Max} - 0.0001T_{Min} + 0.0202S, \quad (3.1)$$

where H is humidity, T is temperature, R is rainfall and S is solar radiation. We tested the model at $\alpha = 0.05$. The results indicated that H_{min} and T_{min} were not statistically significant to tea production. Fig. 3b shows the plot of residuals of the best fit, that is, fit2 + ARIMA (1,1,1). In this model, the standardized residuals show no obvious patterns.

Table 3. Summary for statistics analysis for all the climatic variables with exception of solar radiation

	Adjusted R^2	Residual Standard error	p-value	AIC	BIC
Fit 3	0.02878	0.1579	0.0003821	-543.4057	-507.6269
Fit 4	0.0008958	0.1753	0.364	-400.2612	-364.6322

Table 4. Summary statistics for Linear regression plus residual models of all counties

Model	AR	MA	AIC	BIC	AIC_c	standard error
ARIMA (0,0,1)	-	-0.2710	-0.6812	-0.6181	-0.6808	0.16970
ARIMA (1,0,1)	0.72696	-1.000	-0.7536	-0.6835	-0.7532	0.1628
ARIMA (1,1,0)	0.7269	-	-1.000	-0.1272	-0.07102	0.2241
ARIMA (1,1,1)	-0.2033	-1.000	-0.6561	-0.5929	-0.6557	0.1709

However, there may be outliers exceeding up to 4 standard deviation in magnitude. ACF and Ljung Box test show autocorrelation. The normal $Q - Q$ plot shows the normal distribution of the standardized residuals, a few outliers. Fig. 3b shows the presence of spike in ACF at lag 12 and lag 24. The observation suggests the presence of seasonality in the residuals.

3.2.2 Analysis for all counties with all climatic variables except solar radiation

We considered analysis for all the climatic variables with exception of solar radiation. We fit several linear regression models. Table 3 presents the results obtained after fitting.

Table 3 shows the model Fit 3 and 4, where we considered a differenced tea production for linear model of all climatic variables except solar radiation. Fit 4 is a linear model with the climatic variables rainfall, minimum humidity, maximum and minimum temperature with dependent variable differenced and seasonality removed. Table 3 indicates that fit 3 is the best fit since it has the smallest AIC and BIC. Residuals from fit 3 had no autocorrelated. The observation is due to a lack of pattern in a plot of the ACF of the residuals. We carried out the Durbin-Watson test to fit 3, and the p-value obtained was 0.43, which is greater than 0.05, so H_0 : we fail to rejected the null hypothesis, implying that there was no autocorrelation between the residuals.

Fig. 4 shows the residual models fitted were ARIMA (0,0,1), ARIMA (1,0,1), ARIMA (1,1,0) and ARIMA (1,1,1). In both Fig. 4a and Fig. 4b, that is, both ACF and PACF of the residuals did not show some pattern indicating that the model was not good. Fig. 4a shows fit 3, which was fitted while factoring in various residual models. The results of further analysis is presented in Table 4.

Table 4 shows that AIC and BIC values are similar correspondingly in all the cases. The model data used in the study are real-world data, thus, the size of BIC may be smaller than AIC. That being the case here, the model based on the values in Table 4 suggests that ARIMA(1, 1, 0) is the best fit model. We arrived at this decision when comparing AIC and BIC values; the difference surmounts to large values.

3.2.3 Fitting linear model for each county

Embu Tea production data for Embu county was differenced and then deseasoned to remove trend and seasonality. Fig. 5a shows a plot of the Tea production before and after removing trend and seasonality. We differenced tea production data as dependent variable, while NDVI, Rainfall, Minimum, and maximum Humidity, minimum and Maximum Temperature, and Solar radiation are independent variables, a linear model was formulated as in Equation 3.2.

$$y = 1.6569 - 3.0944N + 0.0055H_{min} + 0.0059H_{Max} + 0.0002R - 0.0202T_{Min} + 0.0126T_{Max} - 0.0022S. \quad (3.2)$$

where N is NDVI, and T is temperature. Equation 3.2 shows that none of the variables had statistical significance at $p = 0.05$ level. The adjusted R-squared was 0.2971, meaning that 29.71% of the total variation in tea production was explained by independent variables. The $p - value = 3.867e^{-07}$ which is less than 0.05, thus the fit is statistically significant. We fitted a second linear model with deseasoned tea production data as the dependent variable and NDVI, Rainfall, Minimum, and maximum Humidity, minimum and Maximum Temperature, and Solar radiation as independent variables. Equation 3.3 shows the second linear regression model obtained

$$y = -0.8432 + 0.9320N + 0.008275H_{Min} - 0.006597H_{Max} - 0.000249R - 0.02541T_{Min} + 0.01089T_{Max} - 0.000496S. \quad (3.3)$$

Equation 3.3 shows none of the variables had statistical significance at $p - value = 0.05$ level. The adjusted R-squared was -0.01611, meaning that 1.61% explained the total variation in tea production by the independent variables. The p-value=0.5999, which is more than 0.05; therefore, the fit is not statistically significant. Fig. 5b shows the ACF and PACF of the residuals from the two model fits

Kericho We differenced and deseasoned tea production data for Kericho county. We aimed to remove trend and seasonality. Fig. 5c shows a plot of the tea production before and after removing trend and seasonality.

We differenced tea production data as dependent variable, while NDVI, Rainfall, Minimum and maximum Humidity, minimum and Maximum Temperature and Solar radiation are independent variables, a linear model was formulated as in Equation 3.4.

$$y = -1.498 + 2.156e^{-5}N + 3.749e^{-3}H_{Min} + 2.425e^{-3}H_{Max} + 2.074e^{-4}R + 3.639e^{-2}T_{Min} + 2.327e^{-2}T_{Max} + 3.789e^{-3}S. \quad (3.4)$$

Equation 3.4 shows that none of the variables had statistical significance at 0.05 level. The adjusted R-squared was 0.1935, meaning that 19.35% explained the total variation in tea production by the independent variables. The p-value=0.0001614, which is less than 0.05; therefore, the fit is statistically significant. We fit a second linear model with deseasoned tea production data as the dependent variable and NDVI, Rainfall, Minimum and maximum Humidity, minimum and Maximum Temperature, and Solar radiation as independent variables. Equation 3.5 shows the second linear regression model obtained.

$$y = -1.651e^{-1} - 1.543e^{-5}N + 2.279e^{-3}H_{Min} + 2.913e^{-3}H_{Max} - 3.134e^{-4} - 6.055e^{-2}T_{Min} + 2.823e^{-2}T_{Max} - 5.178e^{-3}S. \quad (3.5)$$

Equation 3.5 suggests that none of the variables had statistical significance at 0.05 level. The adjusted R-squared was -0.004719, meaning that 0.47% explained the independent variables' total variation in tea production. The p-value=0.4825, which is more than 0.05; therefore, the fit is not statistically significant. Fig. 5d shows the ACF and PACF of the residuals from the two model fits.

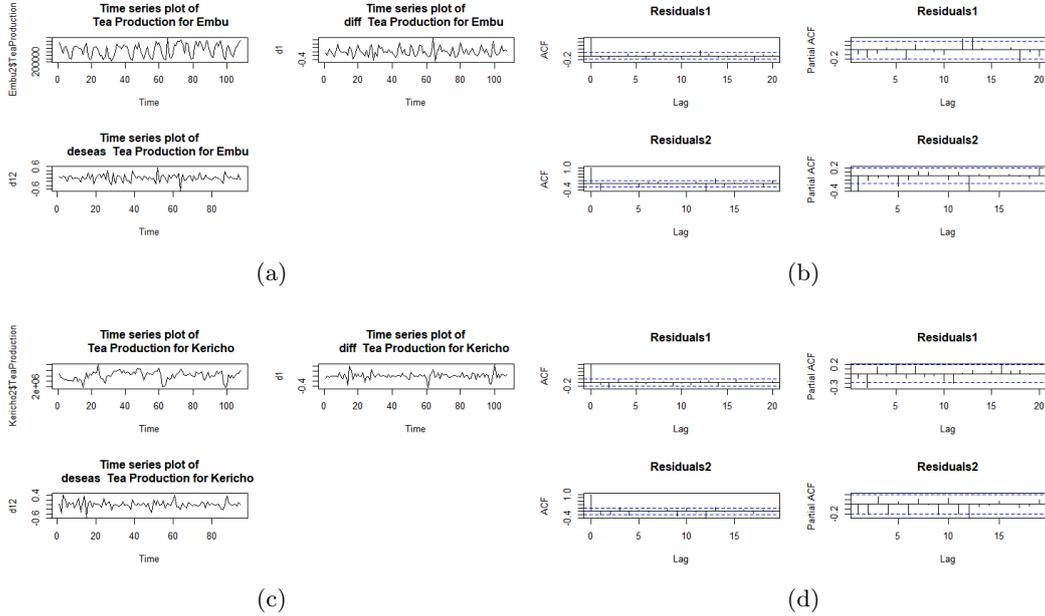


Fig. 5. (a) Plot of differenced and deseasoned Tea Production for Embu county. (b) ACF and PACF of Residuals for Embu county. (c) Plot of differenced and deseasoned Tea Production for Kericho county. (d) ACF and PACF of Residuals for Kericho

Kisii We differenced and then deseasoned tea production data for Kisii county aimed to remove trend and seasonality. Fig. 6a shows a plot of the Tea production before and after removing trend and seasonality. We differenced tea production data as dependent variable, while NDVI, Rainfall, Minimum, and maximum Humidity, minimum and Maximum Temperature are independent variables, a linear model was formulated as in Equation 3.6.

$$y = -1.3048 - 2.1162N + 0.0028H_{Min} + 0.0017H_{Max} + 0.0002R + 0.0128T_{Min} - 0.0067T_{Max}. \quad (3.6)$$

We found NDVI to be statistically significant at 0.05 level. The adjusted R-squared was 0.1736, meaning that the independent variables explained 17.36% of the total variation in tea production. The p-value=0.0003, which is less than 0.05; therefore, the fit is significant. We fitted a second linear model with deseasoned tea production data as the dependent variable and NDVI, Rainfall, Minimum, and maximum Humidity, minimum and Maximum Temperature as independent variables. Equation 3.7 shows the second linear regression model derived model obtained

$$y = -1.47 + 9.537e^{-1}N + 3.147e^{-4}H_{Min} + 1.898e^{-3}H_{Max} - 4.864e^{-5}R - 6.418e^{-2}T_{Min} + 6.247e^{-2}T_{Max}. \quad (3.7)$$

We obtained maximum temperature to be statistically significant at 0.05 level. The adjusted R-squared was 0.01545, meaning that 1.55% explained the independent variables' total variation in tea production. The p-value = 0.291 which is more than 0.05; therefore the fit is not significant. Fig. 6b shows the ACF and PACF of the residuals.

Kakamega We differenced and then deseasoned tea production data for Kakamega county aimed to remove trend and seasonality. Fig. 6c shows a plot of the Tea production before and after removing trend and seasonality.

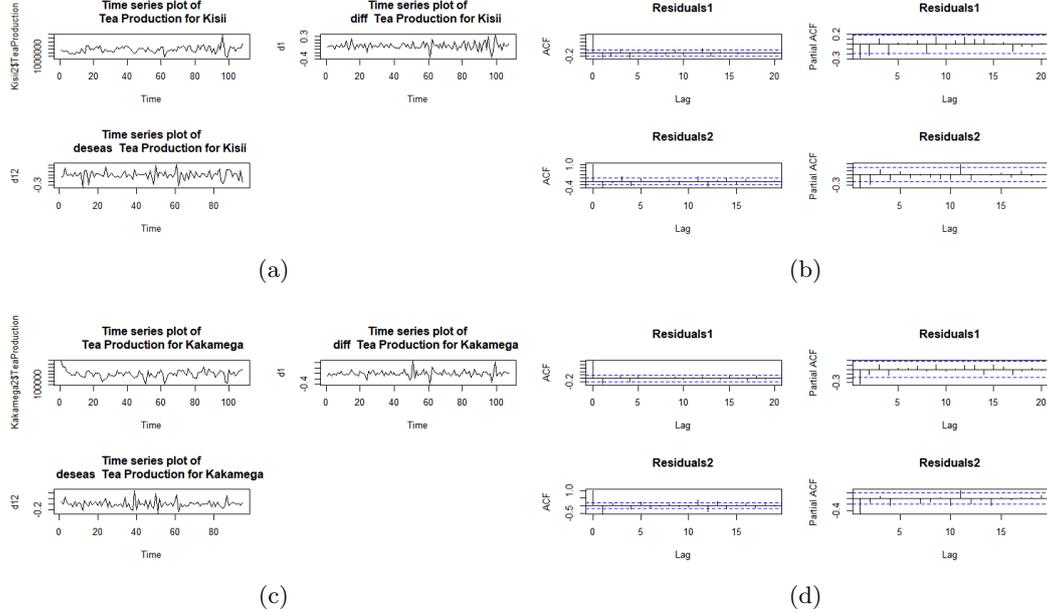


Fig. 6. (a) Plot of differenced and deseasoned Tea Production for Kisii. (b) ACF and PACF of Residuals for Kisii. (c) Plot of differenced and deseasoned Tea Production for Kakamega. (d) ACF and PACF of Residuals for Kakamega

We differenced tea production data as dependent variable, while NDVI, Rainfall, Minimum, and maximum Humidity, minimum and Maximum Temperature are independent variables, a linear model was formulated as in Equation 3.8. We found NDVI, Minimum Humidity, and Minimum temperature to be statistically significant at 0.05 level. The adjusted R-squared was 0.3206, meaning that 32.06% explained the independent variables' total variation in tea production. The $p\text{-value}=4.023e^{-8}$, which is less than 0.05; therefore, the fit is significant. We fitted a second linear model with deseasoned tea production data as the dependent variable and NDVI, Rainfall, Minimum, and maximum Humidity, minimum and Maximum Temperature as independent variables. Equation 3.9 shows the second linear regression model obtained.

$$y = 7.32e^{-1} - 2.34N + 2.36e^{-3}H_{Min} + 2.77e^{-3}H_{Max} + 5.57e^{-5}R + 5.41e^{-2}T_{Min} - 1.23e^{-3}T_{Max}. \quad (3.8)$$

$$y = -8.924e^{-1} + 6.898e^{-1}N - 1.9987e^{-4}H_{Min} + 8.448e^{-4}H_{Max} + 4.104e^{-5}R - 2.32e^{-2}T_{Min} + 2.364e^{-2}T_{Max}. \quad (3.9)$$

Equation 3.9 shows the variables had statistical significance at 0.05 level. The adjusted R-squared was -0.03917 , meaning that 3.9% explained the independent variables' total variation in tea production. The $p\text{-value} = 0.871$, which is more than 0.05; therefore, the fit has no statistical significance. Fig. 6d shows the ACF and PACF of the residuals from the two model fits.

Nyeri We differenced and deseasoned tea production data for Nyeri county to remove trend and seasonality. Fig. 7a shows a plot of the Tea production before and after removing trend and seasonality. We differenced tea production data as dependent variable, while NDVI, Rainfall, Minimum, and maximum Humidity, minimum and Maximum Temperature are independent variables, a linear model was formulated as in Equation 3.10.

$$y = 0.793 - 1.628N + 0.0004H_{Min} + 0.005H_{Max} + 0.001R - 0.016T_{Min} - 0.001T_{Max}. \quad (3.10)$$

Equation 3.10 shows that NDVI, Maximum Humidity, and rainfall are statistically significant at 0.05 level. The adjusted R-squared was 0.2697, meaning that 26.97% explained the independent variables' total variation in tea production. The $p - value = 1.137e^{-6}$ which is less than 0.05 therefore the fit is good. We fitted a second linear model with deseasoned tea production data as the dependent variable and NDVI, Rainfall, Minimum, and maximum Humidity, minimum and Maximum Temperature as independent variables. Equation 3.11 shows the second linear regression model obtained.

$$y = -0.586 + 0.619N + 0.007H_{Min} - 0.009H_{Max} - 0.0005R + 0.016T_{Min} - 0.003T_{Max}. \quad (3.11)$$

Equation 3.11 suggests that NDVI is statistically significant at 0.05 level. The adjusted R-squared was 0.03618, meaning that 3.618% explained the independent variables' total variation in tea production. The $p - value = 0.16$, which is more than 0.05; therefore, the fit is not good. Fig. 7b shows the ACF and PACF of the residuals from the two model fits.

Meru We differenced and deseasoned tea production data for Meru county to remove trend and seasonality. Fig. 7c shows a plot of the tea production before and after removing trend and seasonality. We differenced tea production data as dependent variable, while NDVI, Rainfall, Minimum and maximum Humidity, minimum and Maximum Temperature, and Solar radiation are independent variables, a linear model was formulated as in Equation 3.12.

$$y = 0.569 - 1.114N - 0.0004H_{Min} - 0.003H_{Max} + 0.0003R - 0.008T_{Min} + 0.033.001T_{Max} - 0.011S \quad (3.12)$$

Equation 3.12 suggests that NDVI and rainfall are statistically significant at 0.05 level. The adjusted R-squared was 0.2044, meaning that 20.44% explained the independent variables' total variation in tea production. The $p - value = 9.04e^{-5}$ which is less than 0.05 therefore the fit is good. We fitted a second linear model with deseasoned tea production data as the dependent variable and NDVI, Rainfall, Minimum and maximum Humidity, minimum and Maximum Temperature, and Solar radiation as independent variables. Equation 3.13 shows the first linear regression model obtained.

$$y = -0.803 + 0.428N + 0.008H_{Min} - 0.004H_{Max} + 0.0001R - 0.009T_{Min} + 0.005T_{Max} + 0.003S. \quad (3.13)$$

Equation 3.13 suggests that the variables had no statistical significance at 0.05 level. The adjusted R-squared was -0.01387 , meaning that 1.38% explained the independent variables' total variation in tea production. The $p - value = 0.5763$ which is more than 0.05 therefore, the fit is not good. Fig. 7d shows the ACF and PACF of the residuals from the two model fits are as shown below.

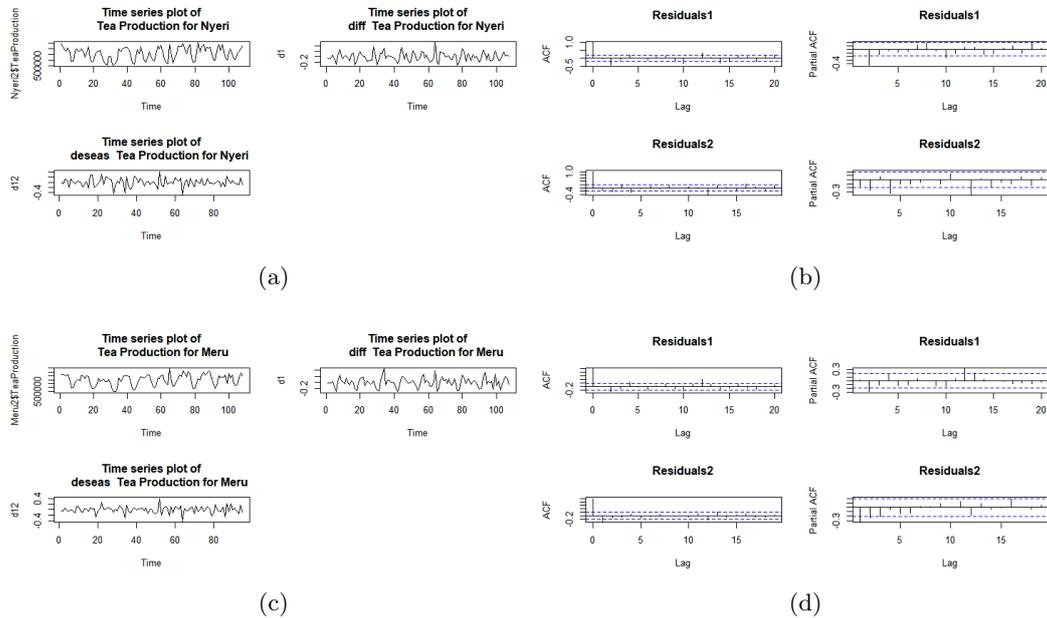


Fig. 7. (a) Plot of differenced and deseasoned Tea Production for Nyeri county. (b) ACF and PACF of Residuals for Nyeri (c) Plot of differenced and deseasoned Tea Production for Meru county. (d) ACF and PACF of Residuals for Meru

4 Discussion

The climatic variations in the tea zones affect tea production. The change in tea production affects farmers' income. These changes can be positive or negative. The need for long and short-term drivers of crop production has prompted the use of modeling to understand and forecast the dynamism of climate variability on food production. The variation of tea production specifically also affects Kenya's GDP, which depends on tea exports. The proposed study has statistical modeling tools, specifically linear regression, and ARIMA to understand the relationship between climatic variability and tea production.

The scatter plot matrix for all the climatic variables (mean annual temperature, mean annual rainfall, humidity, solar radiation, and NDVI) for all the counties (Embu, Kakamega, Kisii, Kericho, Meru, and Nyeri) under the study indicated that tea production has a linear relationship with most climatic variables. The relationship is strong with solar radiation. The model fit analysis was done separately for Kakamega, Kisii, and Nyeri counties with missing solar radiation data and others with all data. The results noted that NDVI and Maximum Temperature had no statistical significance with tea production. Fit 2 was a linear model with all climatic variables except NDVI and Maximum Temperature.

A model fit for each county based on differencing and deseasoning indicated that no variables had statistical significance at $p = 0.05$ level for the first fit for Embu. The results also indicated the data had 29.71% total variation explained by the independent variables. The $p - value = 3.867e^{-07}$ suggesting statistical significance. The second linear regression fit indicated no statistical significance between independent and dependent variables with 1.61% explained variability. The

first linear model fit for Kericho using NDVI, rainfall, minimum and maximum humidity, minimum and maximum temperature, and solar radiation as independent variables and tea production as the dependent variable. The results show that none of the variables had statistical significance at 0.05 level with 19.35% explained variability. A second linear model with deseasoned tea production data as the dependent variable and NDVI, rainfall, minimum, maximum humidity, minimum and maximum temperature, and Solar radiation as independent variables suggests none of the variables had statistical significance at 0.05 level. The adjusted R-squared suggests 0.47% explained variability and shows no statistical significance.

A model fit for a linear model with differenced tea production data as the dependent variable and NDVI, rainfall, minimum and maximum humidity, minimum and maximum temperature as independent variables for Kisii found NDVI is statistically significant at 0.05 level. The adjusted R-squared suggests 17.36% explained variability of the total variation in tea production. The p -value=0.0003 < 0.05, suggests the model fit is statistically significant. A second linear model with deseasoned tea production data as the dependent variable and NDVI, rainfall, minimum and maximum humidity, minimum and maximum temperature as independent variables for Kisii suggests that maximum temperature is statistically significant 0.05 level. The adjusted R-squared suggests 1.55% explained variability of the total variation in tea production. The p - value = 0.291 > 0.05 suggests the fit is not significant.

A model fit for a linear model with differenced tea production data as the dependent variable and NDVI, rainfall, Minimum, and maximum humidity, minimum and maximum temperature as independent variables for Kakemega found that NDVI, minimum humidity, and minimum temperature are statistically significant at the 0.05 level. The adjusted R-squared suggests 32.06% explained variability of the total variation in tea production. The p -value= $4.023e^{-8}$ < 0.05 suggests the model fit is statistically significant. A second linear model with deseasoned tea production data as the dependent variable and NDVI, rainfall, minimum and maximum humidity, minimum and maximum temperature as independent variables for Kakamega suggest that all the variables are statistically significant at 0.05 level. The adjusted R-squared suggests 3.9% explained variability of the total variation in tea production. The p -value = 0.871 > 0.05 suggests the fit is not significant.

A model fit for a linear model with differenced tea production data as the dependent variable and NDVI, rainfall, minimum, and maximum humidity, minimum, and maximum temperature as independent variables for Nyeri found NDVI, maximum humidity, and rainfall are statistically significant at 0.05 level. The adjusted R-squared suggests 26.97% explained variability of the total variation in tea production. The p -value p - value = $1.137e^{-6}$ < 0.05, suggest the model fit is statistically significant. A second linear model with deseasoned tea production data as the dependent variable and NDVI, rainfall, minimum and maximum humidity, minimum and maximum temperature as independent variables for Nyeri suggest that NDVI is statistically significant at 0.05 level. The adjusted R-squared suggests 3.618% explained variability of the total variation in tea production. The p - value = 0.16 > 0.05 suggests the fit is not significant.

A model fit for a linear model with differenced tea production data as the dependent variable and NDVI, rainfall, minimum and maximum humidity, minimum and maximum temperature, and Solar radiation as independent variables for Meru found NDVI and rainfall are statistically significant at 0.05 level. The adjusted R-squared suggests 20.44% explained variability of the total variation in tea production. The p - value = $9.04e^{-5}$ < 0.05, suggest the model fit is statistically significant. A second linear model with tea production data as the dependent variable and NDVI, rainfall, Minimum, and maximum Humidity, minimum and Maximum Temperature, and Solar radiation as independent variables for Meru suggest that no variable is statistically significant at 0.05 level. The

adjusted R-squared suggests 1.38% explained variability of the total variation in tea production. The $p - value = 0.5763 > 0.05$ suggests the fit is not significant.

5 Conclusion

The study shows that climatic variability affects tea production. The dynamism of climatic variability with tea production varies based on the region under investigation. The forecast models generate statistical significance suggesting their possible use in real-world data besides academics. Future studies may consider combining the current study with other analyses. The analyses that may be combined with the current study are statistical modeling procedures such as the generalized autoregressive conditional heteroskedasticity (GARCH) models.

Acknowledgement

The research in this work was supported by the Technical University of Kenya (TU-K).

Competing Interests

Authors have declared that no competing interests exist.

References

- [1] Adhikari, Umesh and Nejadhashemi, A Pouyan and Woznicki, Sean A. Climate change and eastern Africa: a review of impact on major crops. *Food and Energy Security*. 2015;4(2):110-132.
- [2] Lipper Leslie, Thornton Philip, Campbell Bruce M, Baedeker Tobias, Braimoh Ademola, Bwalya Martin, Caron Patrick, Cattaneo Andrea, Garrity Dennis, Henry Kevin, others. Climate-smart agriculture for food security. *Nature Climate Change*. 2014;4(12):1068-1072.
- [3] Thornton Philip K, Ericksen Polly J, Herrero Mario, Challinor Andrew J. Climate variability and vulnerability to climate change: a review. *Global Change Biology*. 2014;20(11):3313-3328.
- [4] Thornton Philip K, van de Steeg Jeannette, Notenbaert An, Herrero Mario. The impacts of climate change on livestock and livestock systems in developing countries: A review of what we know and what we need to know. *Agricultural Systems*. 2009;101(3):113-127.
- [5] De Vito Saverio, Massera Ettore, Piga Marco, Martinotto Luca, Di Francia Girolamo. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*. 2008;129(2):750-757.
- [6] Aho Ken, Derryberry DeWayne, Peterson Teri. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology*. 2014;95(3):631-636.
- [7] Shumway Robert H, Stoffer David S, Stoffer David S. Time series analysis and its applications. 2000;3.
- [8] Davies TD, Vincent CE, Beresford AKC. July-August rainfall in west-central Kenya. *Journal of Climatology*. 1985;5(1):17-33.
- [9] Speranza Chinwe Ifejika, Kiteme Boniface, Ambenje Peter, Wiesmann Urs, Makali Samuel. Indigenous knowledge related to climate variability and change: insights from droughts in semi-arid areas of former Makueni District, Kenya. *Climatic Change*. 2010;100(2):295-315.

- [10] Zeger Scott L, Liang Kung-Yee. Feedback models for discrete and continuous time series. *Statistica Sinica*. 1991;51-64.
- [11] Brockwell Peter J. Continuous-time ARMA processes. *Handbook of Statistics*. 2001;19:249-276.
- [12] Listgarten Jennifer, Neal Radford, Roweis Sam, Emili Andrew. Multiple alignment of continuous time series. *Advances in Neural Information Processing Systems*. 2004;17:817-824.
- [13] McKenzie, Ed. Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association*. 1985;21(4):645-650.
- [14] Jacobs Patricia A, Lewis Peter AW. Discrete time series generated by mixtures. I: Correlational and runs properties. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1978;40(1):94-105.
- [15] Mining Data. Concepts and techniques. Jiawei Han and Micheline Kamber. 2001;2.
- [16] Chou Ray Yeutien. Volatility persistence and stock valuations: Some empirical evidence using GARCH. *Journal of Applied Econometrics*. 1988;279-294.
- [17] Javaheri Alireza, Wilmott Paul, Haug Espen G. GARCH and volatility swaps. *Quantitative Finance*. 2004;4(5):589-595.
- [18] Engle Robert. GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*. 2001;15(4):157-168.
- [19] Chu Jeffrey, Chan Stephen, Nadarajah Saralees, Osterrieder Joerg. GARCH modelling of cryptocurrencies. *Journal of Risk and Financial Management*. 2017;10(4):17.
- [20] Dinar Ariel, Mendelsohn Robert O. *Handbook on climate change and agriculture*. Edward Elgar Publishing; 2011.
- [21] Reidsma Pytrik, Lansink Alfons Oude, Ewert Frank. Economic impacts of climatic variability and subsidies on European agriculture and observed adaptation strategies. *Mitigation and Adaptation Strategies for Global Change*. 2009;14(1):35-59.
- [22] Fogarasi József, Kemény Gábor, Molnár András, Keményné Horváth Zsuzsanna, Nemes Anna, Kiss Andrea. Modelling climate effects on Hungarian winter wheat and maize yields. *Studies in Agricultural Economics*. 2016;118(2):85-90.
- [23] Pinke Zsolt, Lövei Gábor L. Increasing temperature cuts back crop yields in Hungary over the last 90 years. *Global Change Biology*. 2017;23(12):5426-5435.
- [24] Shoaib Syed Abu, Khan Mohammad Zaved Kaiser, Sultana Nahid, Mahmood Taufique H. Quantifying uncertainty in food security modeling. *Agriculture*. 2021;11(1):33.
- [25] Alam T. Forecasting exports and imports through artificial neural network and autoregressive integrated moving average. *Decision Science Letters*. 2019;8(3):249-260.
- [26] Francq Christian, Zakoian Jean-Michel. *GARCH models: structure, statistical inference and financial applications*; 2019.
- [27] Brockwell Peter, Chadraa Erdenebaatar, Lindner Alexander, others. Continuous-time GARCH processes. *Annals of Applied Probability*. 2006;16(2):790-826.
- [28] Freund Rudolf J, Wilson William J, Sa Ping. *Regression analysis*. Elsevier; 2006.
- [29] Newbold Paul. ARIMA model building and the time series analysis approach to forecasting. *Journal of Forecasting*. 1983;2(1):23-35.
- [30] Wei William WS. Time series analysis. *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*; 2006.
- [31] Rosenstein Michael T, Cohen Paul R. Concepts from time series. *AAAI/IAAI*. 1998;739-745.

- [32] Mathu Eliud M, Davies TC. Geology and the environment in Kenya. *Journal of African Earth Sciences*. 1996;23(4):511-539.
- [33] Ayugi Brian Odhiambo, Wen Wang, Chepkemoi Daisy. Analysis of spatial and temporal patterns of rainfall variations over Kenya. *Studies*. 2016;6(11).
- [34] Siciliano Giovanni, Bigi Velia, Vigna Ingrid, Comino Elena, Rosso Maurizio, Cristofori Elena, Demarchi Alessandro, Pezzoli Alessandro. Comparison of Multiple Maximum and Minimum Temperature Datasets at Local Level: The Case Study of North Horr Sub-County, Kenya. *Climate*. 2021;9(4):62.
- [35] Sombroek Wim G, Braun HHM, Van der Pouw BJA, others. Exploratory soil map and agro-climatic zone map of Kenya, 1980. Scale 1: 1,000,000. Kenya Soil Survey; 1982.
- [36] Stocker Thomas F, Qin Dahe, Plattner GK, Tignor Melinda MB, Allen Simon K, Boschung Judith, Nauels Alexander, Xia Yu, Bex Vincent, Midgley Pauline M. *Climate Change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of IPCC the intergovernmental panel on climate change. Report of the Intergovernmental Panel on Climate Change*; 2014.
- [37] Vesco Paola, Kovacic Matija, Mistry Malcolm, Croicu Mihai. Climate variability, crop and conflict: Exploring the impacts of spatial concentration in agricultural production. *Journal of Peace Research*. 2021;58(1):98-113.
- [38] Nguyen-Huy Thong, Deo Ravinesh C, Mushtaq Shahbaz, An-Vo Duc-Anh, Khan Shahjahan. Modeling the joint influence of multiple synoptic-scale, climate mode indices on Australian wheat yield using a vine copula-based approach. *European Journal of Agronomy*. 2018;98:65-81.
- [39] Ma Qihang. Comparison of ARIMA, ANN and LSTM for stock price prediction. *E3S Web of Conferences*. 2020;218.
- [40] Dwivedi DK, Kelaiya JH, Sharma GR. Forecasting monthly rainfall using autoregressive integrated moving average model (ARIMA) and artificial neural network (ANN) model: A case study of Junagadh, Gujarat, India. *Journal of Applied and Natural Science*. 2019;11(1):35-41.
- [41] Kumar ST Pavana, Lyngdoh Ferdinand B. Prediction of Area and Production of Groundnut Using Box-Jenkins Arima and Neural Network Approach. *Journal of Reliability and Statistical Studies*. 2020;265-286.
- [42] Wang Ju-Jie, Wang Jian-Zhou, Zhang Zhe-George, Guo Shu-Po. Stock index forecasting based on a hybrid model. *Omega*. 2012;40(6):758-766.
- [43] Adebisi Ayodele Ariyo, Adewumi Aderemi Oluyinka, Ayo Charles Korede. Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*; 2014.
- [44] Trnka Miroslav, Olesen Jorgen Eivind, Kersebaum Kurt Christian, Rötter Reimund Paul, Brázdil Rudolf, Eitzinger Josef, Jansen Sander, Skjelvåg Arne Oddvar, Peltonen-Sainio Pirjo, Hlavinka Petr, others. Changing regional weather crop yield relationships across Europe between 1901 and 2012. *Climate Research*. 2016;70(2-3):195-214.
- [45] Change, IPCC Climate and others. *The physical science basis*. Cambridge Univ. Press; 2007.
- [46] Kenny Geoff, Meyler Aidan, Quinn Terry, others. *Forecasting Irish inflation using ARIMA models*. Central Bank of Ireland; 1998.
- [47] Francis TR, Kannenberg LW. Yield stability studies in short-season maize. I. A descriptive method for grouping genotypes. *Canadian Journal of Plant Science*. 1978;58(4):1029-1034.
- [48] Bassu Simona, Brisson Nadine, Durand Jean-Louis, Boote Kenneth, Lizaso Jon, Jones James W, Rosenzweig Cynthia, Ruane Alex C, Adam Myriam, Baron Christian, others. How do various maize crop models vary in their responses to climate change factors? *Global Change Biology*. 2014;20(7):2301-2320.

- [49] Stöckle Claudio O, Donatelli Marcello, Nelson Roger. CropSyst, a cropping systems simulation model. *European Journal of Agronomy*. 2003;18(3-4):289-307.
- [50] Mendelsohn Robert. The impact of climate change on agriculture in developing countries. *Journal of Natural Resources Policy Research*. 2008;1(1):5-19.
- [51] Aydinalp Cumhuri, Cresser Malcolm S. The effects of global climate change on agriculture. *American-Eurasian Journal of Agricultural & Environmental Sciences*. 2008;3(5):672-676.
- [52] Tabachnick Barbara G, Fidell Linda S, Ullman Jodie B. *Using Multivariate Statistics*. Pearson Boston, MA. 2007;5.
- [53] Muoki Chalo Richard, Maritim Tony Kipkoech, Oluoch Wyclife Agumba, Kamunya Samson Machohi, Bore John Kipkoech. Combating climate change in the Kenyan Tea Industry. *Frontiers in Plant Science*. 2020;11(339).
- [54] Rigden AJ, Ongoma Victor, Huybers Peter. Kenyan tea is made with heat and water: how will climate change influence its yield? *Environmental Research Letters*. 2020;15(4): 044003.
- [55] Brillinger David R. *Time series: data analysis and theory*. SIAM; 2001.
- [56] Das Suresh Chandra, Das Sudripta, Hazarika Mridul. Breeding of the tea plant (*Camellia sinensis*) in India. *Global Tea Breeding*. 2012;69-124.
- [57] Cryer Jonathan D, Chan Kung-Sik. *Time series analysis: with applications in R*. Springer Science Business Media; 2008.
- [58] Brown Robert Goodell. *Smoothing, forecasting and prediction of discrete time series*; 2004.
- [59] Cauchemez Simon, Ferguson Neil M. Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London. *Journal of the Royal Society Interface*. 2008;5(25):885-897.
- [60] Pourahmadi Mohsen. *Foundations of time series analysis and prediction theory*. John Wiley Sons. 2001;379.
- [61] Matteson David S, Tsay Ruey S. Independent component analysis via distance covariance. *Journal of the American Statistical Association*. 2017;112(518):623-637.
- [62] Hylleberg Svend. *Modelling seasonality*. Oxford University Press; 1992.
- [63] Mann Michael E. On smoothing potentially non-stationary climate time series. *Geophysical Research Letters*. 2004;31:7.
- [64] Seber George AF, Lee Alan J. *Linear regression analysis*. John Wiley Sons. 2012;329.
- [65] Khashei Mehdi, Bijari Mehdi. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*. 2011;11(2):2664-2675.
- [66] Nason Guy P. *Stationary and non-stationary time series*. *Statistics in Volcanology*. 2006;60.
- [67] Arunraj Nari Sivanandam, Ahrens Diane, Fernandes Michael. Application of SARIMAX model to forecast daily sales in food retail industry. *International Journal of Operations Research and Information Systems (IJORIS)*. 2016;7(2):1-21.
- [68] Vagropoulos Stylianos I, Chouliaras GI, Kardakos Evaggelos G, Simoglou Christos K, Bakirtzis Anastasios G. Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. 2016 IEEE International Energy Conference (ENERGYCON). 2016;1-6.
- [69] Trancart T, Acou A, De Oliveira E, Feunteun E. Forecasting animal migration using SARIMAX: An efficient means of reducing silver eel mortality caused by turbines. *Endangered Species Research*. 2013;21(2):181-190.
- [70] Bercu, Sophie and Proïa, Frédéric. A SARIMAX coupled modelling applied to individual load curves intraday forecasting. *Journal of Applied Statistics*. 2013;40(6):1333-1348.

- [71] Blakeney Michael, Mengistie Getachew. Kenya: Tea. Extending the Protection of Geographical Indications: Case Studies of Agricultural Products in Africa. 2012;213.
- [72] Basu Majumder A, Bera B, Rajan A. Tea statistics: global scenario. Inc J Tea Sci. 2010;8(1):121-4.
- [73] Cege OS. Kenya's water resources and their development. National Fisheries Resources Research Institute, Uganda; 1967.
- [74] Cliff Kikawa R, Billy Kalema M. Estimation of the parameters of a linear regression system using the simple averaging method. Global Journal of Pure and Applied Mathematics. 2017;13(11):7749-7758.
- [75] Gordon Rachel A. Basics of writing batch programs with statistical packages. Regression Analysis for the Social Sciences. 2015;100-124.
- [76] Burnham Kenneth P, Anderson David R. A practical information-theoretic approach. Model Selection and Multimodel Inference. 2002;2.

© 2021 Muganda et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

<http://www.sdiarticle4.com/review-history/69257>