Int.J.Curr.Res.Aca.Rev.2020; 8(12): 31-47

EXCELLENT PUBLISHERS

International Journal of Current Research and Academic Review

ISSN: 2347-3215 (Online) Volume 8 Number 11 (November-2020)

Journal homepage: http://www.ijcrar.com

doi: https://doi.org/10.20546/ijcrar.2020.812.005

Time series Analysis of Monthly Average Temperature and Rainfall Using Seasonal ARIMA Model (In Case of Ambo Area, Ethiopia)

Teshome Hailemeskel Abebe*

Department of Economics, College of Business and Economics, Ambo University, Ambo, Ethiopia

**Corresponding author*

Abstract

Forecasting mean temperature and rainfall is an important for planning and formulating agricultural strategies. Thus, this paper, try to analyze and forecast monthly mean temperature and rainfall in Ambo area on the data from January 2012 to March 2019.From graphical analysis on time plot and ACF, the series seems to have a seasonal component. For that purpose, a Seasonal Autoregressive Integrated Moving Average (SARIMA) models were used to estimate and forecast the average monthly temperature and rainfall in the Ambo area, Ethiopia. Among the competitive tentative model, SARIMA (2, 0, 1) (2, 0, 1)12 and SARIMA (1, 0, 1) (1, 0, 1)12 model are the best time series model for fitting and forecasting mean temperature and rainfall, respectively. Moreover, the model diagnostic test on the residuals of SARIMA (2, 0, 1) (2, 0, 1)12 and SARIMA (1, 0, 1) (1, 0, 1)12) on mean temperature and rainfall satisfies the randomness, independency, normality and constant variance (homoscedasticity) assumptions. Finally, SARIMA (2, 0, 1) (2, 0, 1)12 and SARIMA (1, 0, 1) (1, 0, 1)12 were used to forecast mean of monthly temperature and rainfall from the period April 2019 to March 2023.

Article Info

Accepted: 08 October 2020
Available Online: 20 November 2020

Keywords

Temperature, Rainfall, SARIMA, Modeling, Forecasting

Introduction

Climate change has become a critical issue for policy makers, climate researchers, politicians and the public around the world. Climate variability is regarded as the deviation of seasonal and annual climate parameters (i.e., rainfall, temperature, humidity, precipitation etc.) from the long-term observations mean. The long-term continuous temporal change and/or trends in annual, seasonal, and monthly climate parameters are regarded as indicators of potential climate change impacts (Jones *et al.*, 2015).

Climate variability and change are among the major environmental challenges of the 21st century. Successive reports of the Intergovernmental Panel on Climate Change (IPCC, 2007) and various other studies (Schlenker and Lobell, 2010; Thornton *et al.*, 2011) show that climate change is having versatile effects on development particularly on agriculture.

The consequences of climate variability and climate change are potentially more significant for the poor in developing countries than for those living in nations that are more prosperous. Africa is one of the most vulnerable continents to climate change and variability. It has more climate sensitive economies than any other continent with 50% of its population living in dry land areas that are drought-prone. In addition, its agricultural sector contributes an average 21% of GDP in many countries, ranging from 10% to 70% (UNFCCC, 2007).

Ethiopia is heavily dependent on rain-fed agriculture, and its geographical location and topography in combination with low adaptive capacity entail a high vulnerability to adverse impacts of climate change. Regional projections of climate models indicate a substantial rise in mean temperatures in Ethiopia over the 21st century and an increase in rainfall variability, with a rising frequency of both extreme flooding and droughts due to global warming. Given its large role in income and employment, agriculture also acts as a transmission chain of climate shocks towards other sectors of the economy. Ethiopia is historically prone to extreme weather events. Rainfall in Ethiopia is highly erratic, and most rain falls in convective storms, with very high rainfall intensity and extreme spatial and temporal variability (Robinson *et al.,* 2012*)*.

In this regard, several studies have been conducted to the analysis the pattern and trend of climate variationin various regions of the world using different time series methods. Among the common models are an autoregressive-integrated-moving average (ARIMA) models popularized by Box and Jenkins (1970) and seasonal Autoregressive Integrated Moving Average (SARIMA) models.

Therefore, this paper addresses the shortcomings in analytical literature about climate change using seasonal ARIMA model to analyze the pattern and trend, to fit an appropriate model, and forecast the future value of climate data (mean temperature and rainfall)in Ambo area, Ethiopia.

George Box and Gwilym Jenkins popularized Autoregressive Integrated Moving Average models (ARIMA models) in the early 1970s.ARIMA models are a class of linear models that is capable of representing stationary as well as non-stationary time series. The models rely heavily on autocorrelation patterns in the data. ARIMA methodology of forecasting is different from most methods because it does not assume any particular pattern in the historical data of the series to be forecast, rather it uses an interactive approach of identifying a possible model from a general class of models. in this regard, there are several researches that have done on weather variability using an ARIMA type models. Among the common researches, few are discussed below.

Mehmet Tektaş (2010) tries toforecast the weather of Göztepe, İstanbul, Turkeyabout nine year data (2000-2008) comprising daily average temperature (dry-wet), air pressure, and wind-speed using Auto Regressive Moving Average (ARIMA) models. The paper explains briefly how neuro-fuzzy models can be formulated using different learning methods and then analyzes whether they can provide the required level of performance for a reliable model for practical weather forecasting. The results the most suitable model and network structure are determined according to prediction performance, reliability and efficiency. The performance comparisons of the models are evaluated using RMSE (Root-Mean-Square error) criteria.

Ademola A.*et al.,* (2018) investigate statistical Modeling of Monthly Rainfall in Selected Stations in Forest and Savannah Eco-climatic Regions of Nigeria using Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) models were used. The results showed that the model fitted into the data well and the stochastic seasonal fluctuation was successfully modeled. They concluded that Seasonal Autoregressive Integrated Moving Average (SARIMA) model was a proper method for modeling and predicting the monthly rainfall. The results are useful for forecasting the pattern of rainfall in the study area and provide information that would be helpful for decision makers in formulating policies to mitigate the problems of water resources management, soil erosion, flooding, and drought.

Following similar literature like discussed above, we try to apply Seasonal Autoregressive Integrated Moving Average (SARIMA) model to analyze and forecast the mean temperature and rainfall in Ambo area. However, this study is differ from the previous studies in that it is considered the first study applied in analyzing and forecasting of weather data in Ambo area using an SARIMA model.

Materials and Methods

Study area and data

The study was conducted in the Ambo area, West Shoa Zone of Oromia regional state, Ethiopia. West Shewa Zone is located in the western direction of Addis Ababa the capital city of Ethiopia. Ambo is the capital city of west shoa zone which is 114 km away from Addis Ababa lying between 80 47' – 90 21' North latitude and 37 32' – 38 3' East longitude. The Zone has a unimodal rainfall patter having one significant rainy season and one peak rainfall.

The main rainy season is from June to September and the short rainy season is from the end February to the beginning of May. In this study, monthly average temperature and rainfall were used. The data on temperature and rainfall were collected from Ambo University metrological station of Ethiopia. The dataset consists of (87) monthly observations from January 2012 to March 2019 on global near surface mean temperature and rainfall. Temperature is measured in degrees Celsius, while data on monthly rainfall is measured in milliliter (mm).

Stationary test

The foundation of time series analysis is stationary. Stationary series are characterized by a kind of statistical equilibrium around a constant mean level as well as a constant dispersion around that mean level (Box and Jenkins, 1976). The series could be nonstationary because of random walk, drift, or trend.

Several statistical tests may be conducted to determine whether a series is nonstationary (unit root). Dickey-Fuller unit root test which used to test an AR(1) process is among the common unit root. However, if the series is correlated at higher order lags, the assumption of white noise disturbances is violated. Thus, the Augmented Dickey–Fuller Test (ADF test were used since it controls higher-order correlation by adding lagged difference terms of the dependent variable to the right-hand side of the regression. However, the two common problems in performing the ADF test are specify the number of lagged first difference terms and the choice of including a constant, a constant and a linear time trend, or neither in the test regression.

A specification of ADF test with drift (constant) in the test regression:

$$\nabla Y_t = \mu + \pi y_{t-1} + \sum_{i=2}^{p} \emptyset_i \nabla y_{t-i} + \varepsilon_t \qquad (1)$$

A specification of ADF test with drift (constant) and trend in the test regression:

$$\nabla Y_t = \mu + \pi y_{t-1} + \sum_{i=2}^{p} \emptyset_i \nabla y_{t-i} + \beta t + \varepsilon_t \qquad (2)$$

While if the series seems to be fluctuating around a zero mean, we should include neither a constant nor a trend in the test regression.

$$\nabla Y_t = \pi y_{t-1} + \sum_{i=2}^{p} \emptyset_i \nabla y_{t-i} + \varepsilon_t \qquad (3)$$

Model specifications

One of the most popular and frequently used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA) developed by Box, G. Jenkins model in (1970). The basic assumption made to implement this model is that the considered time series is linear and follows a particular known statistical distribution, such as the normal distribution.

ARIMA model has several components, such as the Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA models. For seasonal time series forecasting, Box and Jenkins had proposed a quite successful variation of ARIMA model, viz. the Seasonal ARIMA (SARIMA).

Autoregressive (AR) Models

Autoregressive (AR) refers to when the value of a series at a current time period is a function of its immediately previous value plus some error.

A general p^{th} -order autoregressive or AR (p) process would be written as follows:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots \ldots\ldots + \alpha_p y_{t-p+1} + \varepsilon_t \qquad (4)$$

where y_t is the actual value of the series at timeperiod *t*, α_i's are coefficients, p is a non-negative integer that indicates the lag length and ε_t is assumed to be a white noise error term.

Moving-Average (MA) Models

An Moving-Average process refers to random error, innovation, or shock, e_t, at a previous period plus a shock at current time, t, drives the series to yield an output value of Y_t at time t.

A more general q^{th}-order moving average or MA (q) process would be written as:

$$y_t = \theta_0 + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (5)$$

where q is a non-negative integer refers to the order of the model, $\theta_i' s$ are coefficients with θ_o the constant term, and { ε_t } is assumed to be a white noise with mean zero and constant variance δ_ε^2 error term or shock.

The Autoregressive Moving Average (ARMA) Models

An autoregressive moving average, ARMA(*p*, *q*) model is a combination of autoregressive AR(*p*) and moving average MA(*q*) models and is suitable for modeling a univariate stochastic time series.

Autoregressive Moving average model, ARMA (P, Q) process is given by:

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \theta_j \sum_{j=1}^{q} \varepsilon_{t-j} + \varepsilon_t \qquad (6)$$

where the model orders *p*, *q* refer to *p* autoregressive and *q* moving average terms. While ε_t is zero mean white noise.

Autoregressive Integrated Moving Average (ARIMA) Models

The ARMA models, described above can only be used for stationary time series data. However, in practice many time series show non-stationary behavior. Because the Box–Jenkins method is an analysis in the time domain applied to stationary series data, it is necessary to consider the basis of nonstationarity, with a view toward transforming series into stationarity. When a nonstationary series is characterized by a random walk, each subsequent observation of the series randomly wanders from the previous one. The random walk model can be stationary after differencing and called integrating order d. Thus, the basic processes of the Box–Jenkins ARIMA (*p*,*d*,*q*) model include the autoregressive process, the integrated process, and the moving average process.

Mathematically, the ARIMA (*p*,*d*,*q*) model using lag polynomials is given by:

$$\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right)(1 - L)^d y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)\varepsilon_t \qquad (7)$$

where the three parameters are: d = number of difference required for Stationarity, p = order of the AR component and q = order of the MA component.

Seasonal Autoregressive Integrated Moving Average (SARIMA) Models

Seasonality usually causes the series to be non-stationary because the average values at some particular times within the seasonal span (for example, month) may be different from the average values at other times. Box and Jenkins (1970) have generalized ARIMA model into Seasonal ARIMA (SARIMA) model to deal with seasonality.

$$SARIMA(p,d,q) \times (P,D,Q)^s$$

Model in terms of lag polynomials is given by:

$$\alpha_p(L^s)\alpha_p(L)(1 - L)^d(1 - L^s)^D y_t = \theta_q(L^s)\theta_q(L)\varepsilon_t \qquad (8)$$

Where *p* = non-seasonal AR order, *d* = non-seasonal differencing, *q* = non-seasonal MA order, *P* = seasonal AR order, *D* = seasonal differencing, *Q* = seasonal MA order, and *S* = time span of repeating seasonal pattern.

Lag length selection

ACF and PACF

Autocorrelation (AC) and partial autocorrelation (PAC) function are a type of graphs that containcorrelations of different time lags. They can be used to determine whether the series are stationary or not, have seasonal pattern and to identify the number of components (lags or parameters) in a SARIMA model. The number of significant spikes in the ACF indicates the number of

MA parameters in the model, while the number of significant spikes in PACF indicates the number of AR parameters in the model.

Portmanteau test

Time series applications often require testing jointly that several autocorrelations of the series are zero. Box and Pierce (1970) propose the Portmanteau statistic given by:

$$Q^{*}_{(m)} = T\sum_{l=1}^{m}\hat{\rho}_l^{\,2} \qquad (9)$$

where T = sample size and m = lag length

Ljung and Box (1978) modify the $Q^{*}_{(m)}$) statistic as below to increase the power of the test in finite samples:

$$Q(m) = T(T+2)\sum_{l=1}^{m}\frac{\hat{\rho}_l^{\,2}}{T-l} \qquad (10)$$

The decision rule is to reject Ho if $Q(m) > \chi 2_{\alpha/2}$, where $\chi 2_{\alpha/2}$ denotes the 100(1−α) the percentile of a chi-squared distribution with m degrees of freedom.

Model selection

A crucial step in an appropriate model selection is the determination of optimal model parameters. According to Faraway and Chatfield (1998), among the common criteria of model selection in time series analysis is to use the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

The Akaike information criterion (AIC) (Akaike, 1974) is defined

$$AIC = T\ln(MSE) + 2k \qquad (11)$$

The Bayesian criterion of SCHWARZ (1978)

$$BIC(l) = T\ln(MSE) + k\ln(T) \qquad (12)$$

where T = number of observations, k =p+q+1(for ARMA)= number of parameters, ln is natural logarithm, residual sum of squares ($SSE) = \sum_{t=1}^{T}(\hat{\varepsilon}_t)^2$, and $Mean\ square\ error(MSE) = \frac{1}{T-K}(SSE)$.

The optimal model order is chosen by the number of model parameters, which minimizes the information criteria

Model estimations

Maximum likelihood estimation is commonly used to estimate ARMA family models. Maximum likelihood estimation usually begins with a likelihood function to minimize or maximize. A likelihood function is a probability formula. When observations are independent of one another, the probability of the multiple successive occurrences is the product of their individual probabilities.

The natural log of the likelihood function given the parameter vector ($\vartheta)$ is given by:

$$LL(\vartheta,\sigma_\varepsilon) = -n\ln(\sigma_\varepsilon) - \frac{\sum \varepsilon_t^2.(\vartheta)}{2\varepsilon_t^2} \qquad (13)$$

Model diagnostic

After estimation of the model, the Box–Jenkins model building strategy entails a diagnosis of the adequacy of the model. More specifically, it is necessary to ascertain in what way the model is adequate and in what way it is inadequate.

Testing for ARCH Effect

It is a test for determining whether ‘ARCH-effects’ are present in the residuals of an estimated mean model.

Engle Lagrange multiplier test of the ARCH effect

The Lagrange multiplier test of Engle (1982) is equivalent to the usual F test. Testing for the presence of ARCH errors of Engle (1982) Lagrange Multiplier test involves regressing the squared residuals from the best fitting ARIMA on a constant and the lagged residuals. If there are no ARCH effects present, the individual coefficient estimates should equal zero, and the joint effects should be slight or non-existent.

The Lung-Box test applied to the ARCH effect

The second test is to apply the usual Ljung–Box statistics $Q(m)$to the ($\varepsilon_t{}^2$). The null hypothesis is that the first m lags of ACF of the $\varepsilon_t{}^2$ series are zero.

To test the ARCH effect, we can apply the Ljung-Box test which were developed by Box and Pierce (1970) and modified by Ljung and Box (1978) and tests the joint significances of serial correlation (heteroscedasticity) in the standardized and squared standardized residuals for the first m lags instead of testing individual significance to squared residuals of the model.

$$LB = T(T+2)\sum_{i=1}^{m}\frac{\hat{\gamma}_i}{T-i} \qquad (14)$$

where T denotes number of observations. $\hat{\gamma}_i$ is correlation coefficient for squared errors with its lag of order m. The computed test statistics has under the null $X^2(m)$ distribution. The null of no ARCH effect) is to be rejected if the computed value of the test statistics is greater than the appropriate critical value.

Model forecasting

One of the primary objectives of time series analysis is to forecast future value of the series, especially forecasting in weather data is great importance. In this regard, a decision needs to be made at current time t and the optimal decision depends on expected future value of a random variable, y_{t+h}, the value being predicted or forecast. The number of time points forecast into the future forecast horizon is called the lead time, h. The value of the random variable for such a forecast is the value of y_{t+h}, A forecaster would like to obtain a prediction as close as possible to the actual value of the variable in question at the concurrent or future temporal point of interest.

Results and Discussions

The monthly data used in this study covers from January 1, 2012 to March 1, 2019 which were collected from Ambo University Meteorological Agency.

Graphical analysis

Both Tables shows the existence of consistent pattern of short-term fluctuation for the data that indicate the existence of seasonality on both variables. The overall mean temperature and rainfall during the studied period January 2012 to March 2019 appears to exhibit a slight (almost no) trend. Thus, from both figure, we observe that the series are stationary at level. Therefore, the time plot gives us a clue for the property of the series, but not the end since we need to conduct a formal unit root test for testing stationarity condition of the series.

Descriptive statistics for average temperature and Rainfall data

From the Table 3, the mean value of maximum temperature is 23.44°C and the mean minimum temperature is 14.25°C with the grand mean 19.54 and mean standard deviations being 1.66. The mean value of rainfall is 2.35 mm with standard deviation of 2.37.

Stationarity test

From the Table 4, we observe that the series is stationary at level with trend and constant term

Model Identification (selection) results

The first step in Box-Jenkins methodology is to identify (select) the appropriate model. In this study, to identify the model (based on lag structure), the correlogram, autocorrelation and partial autocorrelation function were used.

On Figure 1, the Auto correlation function (ACF) and Partial Auto correlation function (PACF) on average temperature and rainfall were tested up to 72-lag interval. From the AC function of both graph, we observe that there is a fast decaying in AC values of the which indicates the stationarity condition of rainfall and temperature. From the AC function, we should to use 1 upto 2 lag for MA model since we have two significant spike, but nothing else beyond that. Moreover, we have very strong first lag in PACF of both graphs while everything is died off. That is we have two lag for AR model for rainfall and one lag for mean temperature.

On Figure 3, the shadow line indicates the 95% confidence interval. Form theoretical view, there should be a fast decline in AC function for stationary AR model. Moreover, there are only positive autocorrelation coefficients since the first lag of the AC and PAC function is upward.

The ACF and PACF show that both temperature and rainfall have periodic in nature. These functions behave similarly in their period cycles involving seasonal variations. Thus, given time series is periodic and involve seasonal variations, we need to apply a model that captures the seasonal component of order 12. Therefore, a seasonal ARIMA models for the prediction of average temperature, and rainfall should be used.

Following the Box–Jenkins technique, we depend on ACF or PACF plots to select the order of the seasonal model (Chatfield 2004).

From the Figures 3a and 3b on average temperature data, we can choose our model based on the ACF and PACF spikes at low lags. To determine the nonseasonal AR terms, we look at the PACF, which shows clear spikes at lags 1, 2 and 3. Thus, the nonseasonal AR terms are determined to be of order 3. There is one spike at lags 1 in ACF, so we have one term for nonseasonal MA. Now for the seasonal part of the model, in this case, we look at lags 12, 24, 36, and 48 for both ACF and PACF. From the PACF we indicate that there are two significant spikes at lags 12 and 36; thus, the order of the seasonal AR is two. In the ACF, there are one spikes at lag 12, this means that the order of the seasonal MA is one.

By the same analogy, we can select a best model for rainfall based on AC and PAC function on figure 3c and 3d. From the PAC function of 3d, the nonseasonal component of MA term has 2 lags and the seasonal component of MA term has one lag. Moreover, from figure 2c, the nonseasonal AR term has one lag and the seasonal AR tem has 2 lags (lag 12 & 36).

Therefore, based on identified lag length of ACF and PACF, we should select the best possible models (‘parsimonious’ models) to represent the original series given below.

However, in order to select appropriate model among the tentative model, we have used the highest log-likelihood statistic, the lowest Akaike information criteria (AIC) and Shwarz Bayesian information criteria (SBIC).

The results of Table 5 show that, out of the nine models, the maximum log likelihood estimates andthe lowest AIC and SBIC values were obtained by SARIMA (2,0,1) (2,0,1)12model. Thus, it can be concluded the best model among the nine for mean temperature is SARIMA (2, 0, 1) (2, 0, 1)12.

From Table 6, SARIMA (1, 0, 1) (1, 0, 1)12 is the best model among the seven tentative model for estimating mean rainfall data in the study area. Thus, we can be concluded the best model among the seven model for mean rainfall is SARIMA (1, 0, 1) (1, 0, 1)12.

Parameter estimation

The coefficients of autoregressive, moving average, seasonal autoregressive, and seasonal moving average were estimated using maximum likelihood estimation methods.

The coefficients of AR terms are less than one which indicates Stationarity of the series.

From Table(7 & 8), we observe that all coefficients are statistically significant as indicated by p-value, except seasonal autoregressive (SAR) in mean temperature model and non-seasonal moving average (MA) model in mean rainfall model. Therefore, the overall performance of the model is good in significance of coefficients.

Model diagnostics test

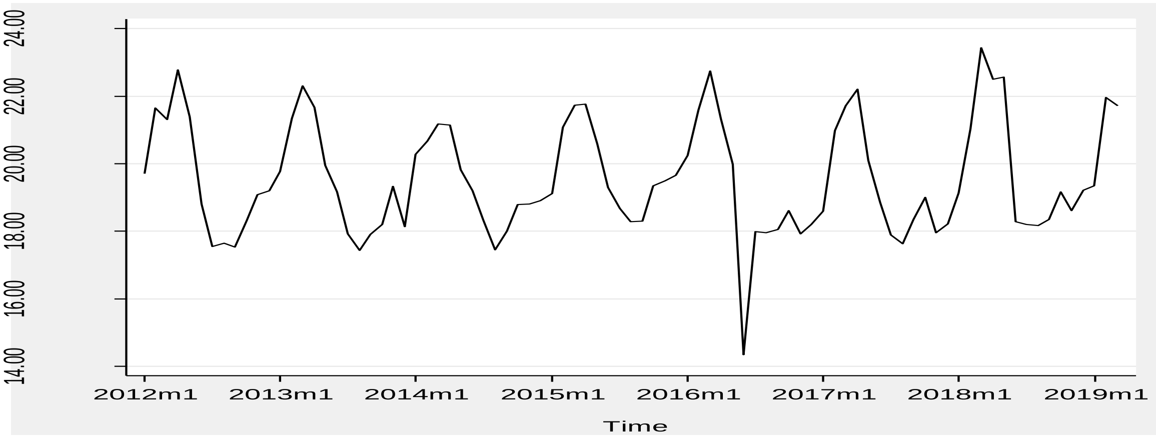
In time-series modeling, the selection of a best model fit to the data is related to whether the residual from the best fitted model is performed well. One of the basic assumptions of the SARIMA (seasonal ARIMA) model is that, for a good model, the residuals must follow a white noise process; that is, the residuals have zero mean, have constant variance (homoscedasticity),and are uncorrelated with past values. The special case of this process is the residuals should be normally distributed and follows a Gaussian white noise process. It is such a process that we test for here. Since the model diagnostics were performed through careful examination of SARIMA$(2,0,1)(2,0,1)_{12}$ model for mean temperature and SARIMA$(0,1,2)(0,1,1)_{12}$ model for mean of rainfall, we have used the residuals plot over time, periodogram, Portmanteau test, histogram and ARCH LM test on the residuals of the best fitted models.

Model diagnostic for the fitted model of Mean Temperature

Randomness of the Residual from the fitted model

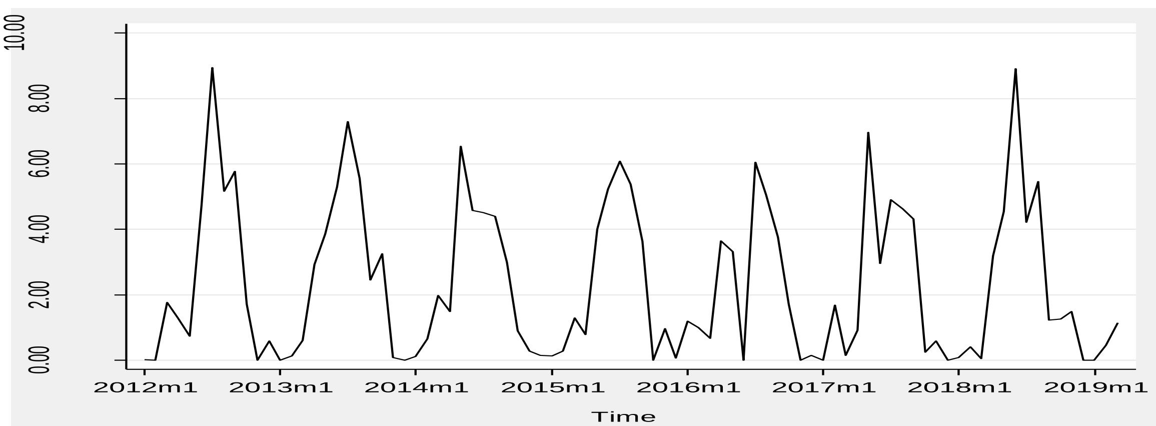
The residual from the fitted model is almost stationary (white noise) since residual series have randomness and no any trending component.

Table.1 Time series plot of Average monthly Temperature



Source (AUMS, May 2019)

Table.2 Time series plot of Average monthly Rainfall



Source (AUMS, May 2019)

Table.3 Summary statistics

Variables	Obs	Mean	Std. Dev	Min	Max
Temperature	87	19.54474	1.665812	14.35	23.44032
Rainfall	87	2.357651	2.371098	0	8.951613

Source: Author's Computation

Table.4 Unit root test

Variables	Model	Test statistic	5% critical value
Temperature	With drift	-3.940	-2.901
	With trend	-4.918	-3.465
Rainfall	With drift	-4.361	-1.664
	With trend	-4.382	-3.465

Source: Author's Computation

Table.5 Tentative model for mean temperature

Model	Log likelihood	AIC	SBIC
SARIMA (0,0,1) (0,0,1)12	-140.0562	288.1124	297.976
SARIMA (1,0,0)(1,0,0)12	-126.6262	261.2525	271.1161
SARIMA (1,0,1) (1,0,1)12	-114.4853	240.9705	255.766
SARIMA (1,0,1) (2,0,1)12	-114.3082	242.6165	259.8778
SARIMA (2,0,1) (1,0,0)12	-122.0155	256.0309	270.8264
SARIMA (2,0,0) (1,0,1)12	-125.5733	261.1466	273.4761
SARIMA (2,0,1) (2,0,1)12	**-110.4433**	**234.8866**	**252.148**
SARIMA (1,0,2) (1,0,2)12	-117.1864	248.3728	265.6342
SARIMA (2,0,2) (2,0,2)12	-112.2646	244.5292	269.1883

Source: Author's Computation

Table.6 Tentative model for rainfall

Model	Log likelihood	AIC	SBIC
SARIMA (1,0,1) (1,0,1)12	**-159.4732**	**330.9465**	**345.7419**
SARIMA (1,0,2) (1,0,0)12	-164.5333	339.0666	351.396
SARIMA (1,0,1) (2,0,0)12	-167.836	347.676	362.471
SARIMA (1,0,2) (2,0,0)12	-163.888	339.777	354.573
SARIMA (1,0,0) (1,0,0)12	-168.831	345.663	355.527
SARIMA (0,0,2) (0,0,1)12	-169.833	349.667	361.997
SARIMA (0,0,1) (1,0,1)12	-162.783	335.566	347.895

Source: Author's Computation

Table.7 Parameter estimation of SARIMA (2, 0, 1) (2, 0, 1)12 for Mean Temperature

Variable	Coefficient	Standard error	z-statistic	P-value
C	19.45835	.0307129	633.56	0.000
AR (1)	.312479	.0930896	3.35	0.001
AR (2)	-.4238803	.0780194	-5.43	0.000
SAR (1)	.7677952	.3117185	2.46	0.014
SAR(2)	.2050881	.2690375	0.76	0.446
SMA (1)	-.6677989	.3705386	-1.80	0.072

Source: Author's Computation

Table.8 Parameter estimation of SARIMA (1, 0, 1) (1,0,1)12 for Mean Rainfall

Variable	Coefficient	Standard error	z-statistic	P-value
C	2.404	0.082	29.24	0.000
AR (1)	0.581	0.106	5.48	0.000
MA (1)	-0.068	5.37	-0.012	0.246
SAR(1)	0.658	0.018	36.55	0.000
SMA (1)	-0.713	0.35	-2.03	0.048

Source: Author's Computation

Table.9 Portmanteau test for white noise

Portmanteau (Q) statistic	25.1610
Prob > chi2(40)	0.9677

Source: Author's Computation

Table.10 Results of Residual Heteroscedasticity Test for the Fitted Model of Temperature

ARCH Lm test			
F-statistic	0.825	P-value	0.732
Obs*R-squared	0.67	P-value	0.534

Source: Author's Computation

Table.10 Portmanteau test for white noise

Portmanteau (Q) statistic	49.5665
Prob > chi2(40)	0.1428

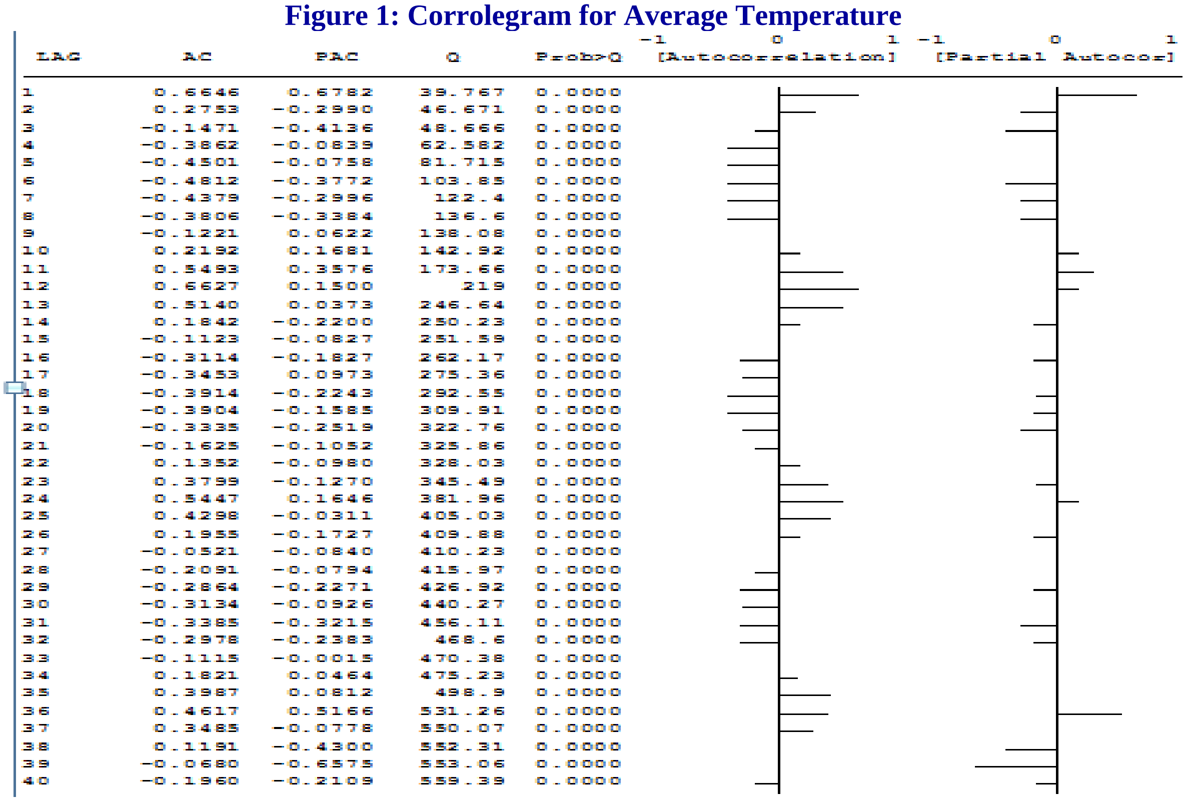
Source: Author's Computation

Table.11 Results of Residual Heteroscedasticity Test for the Fitted Model of Rainfall

ARCH Lm test			
F-statistic	0.649	P-value	0.571
Obs*R-squared	0.584	P-value	0.426

Source: Author's Computation

Figure 1: Corrolegram for Average Temperature

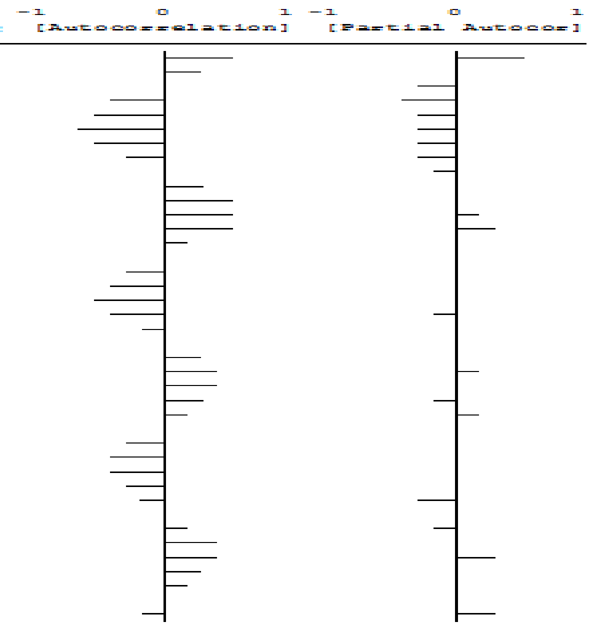


LAG	AC	PAC	Q	Prob>Q
1	0.6646	0.6782	39.767	0.0000
2	0.2753	-0.2990	46.671	0.0000
3	-0.1471	-0.4136	48.666	0.0000
4	-0.3862	-0.0839	62.582	0.0000
5	-0.4501	-0.0758	81.715	0.0000
6	-0.4812	-0.3772	103.85	0.0000
7	-0.4379	-0.2996	122.4	0.0000
8	-0.3806	-0.3384	136.6	0.0000
9	-0.1221	0.0622	138.08	0.0000
10	0.2192	0.1681	142.92	0.0000
11	0.5493	0.3576	173.66	0.0000
12	0.6627	0.1500	219	0.0000
13	0.5140	0.0373	246.64	0.0000
14	0.1842	-0.2200	250.23	0.0000
15	-0.1123	-0.0827	251.59	0.0000
16	-0.3114	-0.1827	262.17	0.0000
17	-0.3453	0.0973	275.36	0.0000
18	-0.3914	-0.2243	292.55	0.0000
19	-0.3904	-0.1585	309.91	0.0000
20	-0.3335	-0.2519	322.76	0.0000
21	-0.1625	-0.1052	325.86	0.0000
22	0.1352	-0.0980	328.03	0.0000
23	0.3799	-0.1270	345.49	0.0000
24	0.5447	0.1646	381.96	0.0000
25	0.4298	-0.0311	405.03	0.0000
26	0.1955	-0.1727	409.88	0.0000
27	-0.0521	-0.0840	410.23	0.0000
28	-0.2091	-0.0794	415.97	0.0000
29	-0.2864	-0.2271	426.92	0.0000
30	-0.3134	-0.0926	440.27	0.0000
31	-0.3385	-0.3215	456.11	0.0000
32	-0.2978	-0.2383	468.6	0.0000
33	-0.1115	-0.0015	470.38	0.0000
34	0.1821	0.0464	475.23	0.0000
35	0.3987	0.0812	498.9	0.0000
36	0.4617	0.5166	531.26	0.0000
37	0.3485	-0.0778	550.07	0.0000
38	0.1191	-0.4300	552.31	0.0000
39	-0.0680	-0.6575	553.06	0.0000
40	-0.1960	-0.2109	559.39	0.0000

Source: Author's Computation

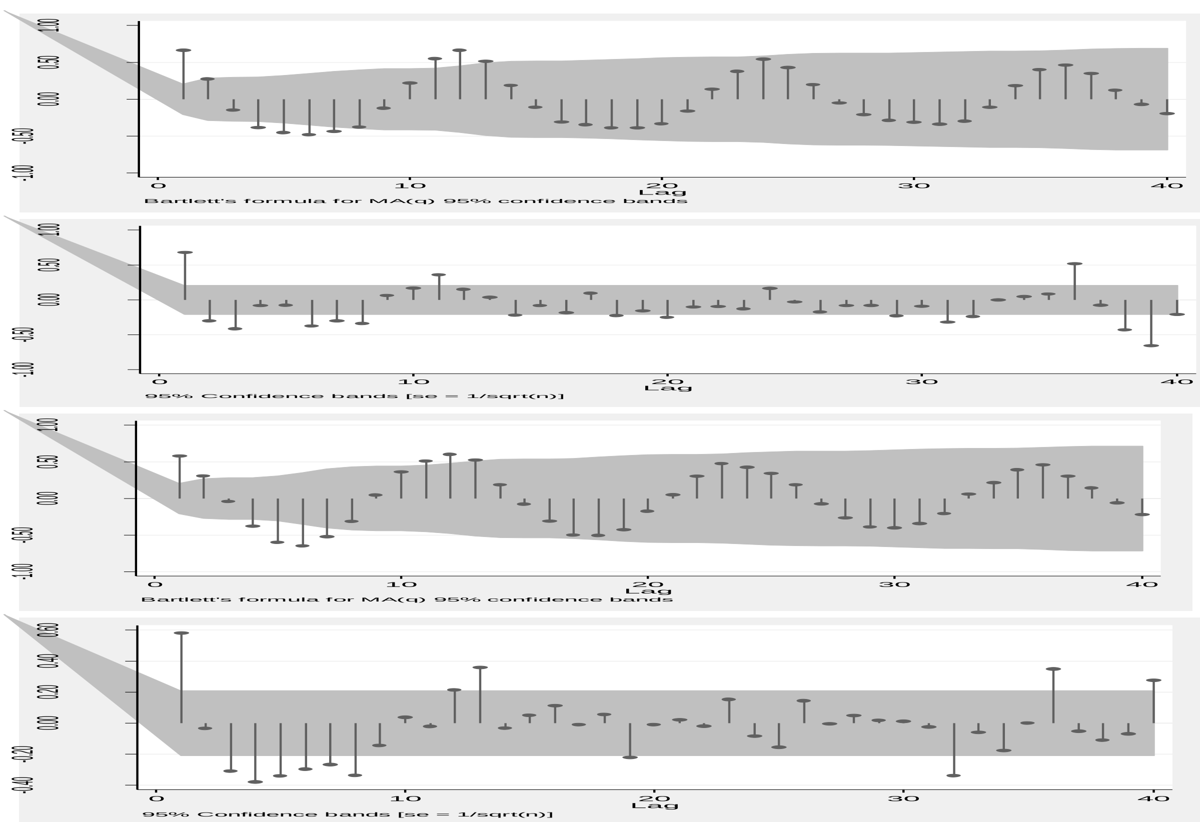
Fig.2 Corrolegram for Average Rainfall

LAG	AC	PAC	Q	Prob>Q
1	0.5779	0.5796	30.071	0.0000
2	0.3071	-0.0341	38.664	0.0000
3	-0.0405	-0.3107	38.815	0.0000
4	-0.3793	-0.3793	52.24	0.0000
5	-0.5991	-0.3414	86.132	0.0000
6	-0.6470	-0.2963	126.15	0.0000
7	-0.5195	-0.2664	152.27	0.0000
8	-0.3137	-0.3384	161.91	0.0000
9	0.0484	-0.1446	162.15	0.0000
10	0.3624	0.0381	175.35	0.0000
11	0.5087	-0.0209	201.71	0.0000
12	0.6024	0.2128	239.18	0.0000
13	0.5226	0.3578	267.75	0.0000
14	0.1846	-0.0326	271.37	0.0000
15	-0.0774	0.0503	272.01	0.0000
16	-0.3110	0.1119	282.56	0.0000
17	-0.4988	-0.0110	310.08	0.0000
18	-0.5038	0.0548	338.56	0.0000
19	-0.4271	-0.2229	359.33	0.0000
20	-0.1731	-0.0104	362.79	0.0000
21	0.0508	0.0202	363.09	0.0000
22	0.3022	-0.0190	373.97	0.0000
23	0.4738	0.1533	401.13	0.0000
24	0.4263	-0.0847	423.46	0.0000
25	0.3410	-0.1547	437.99	0.0000
26	0.1868	0.1439	442.42	0.0000
27	-0.0738	-0.0062	443.12	0.0000
28	-0.2643	0.0475	452.29	0.0000
29	-0.3892	0.0166	472.5	0.0000
30	-0.4006	0.0111	494.3	0.0000
31	-0.3416	-0.0250	510.44	0.0000
32	-0.2085	-0.3399	516.56	0.0000
33	0.0595	-0.0611	517.07	0.0000
34	0.2162	-0.1779	523.9	0.0000
35	0.3911	0.0006	546.68	0.0000
36	0.4562	0.3495	578.28	0.0000
37	0.3015	-0.0512	592.36	0.0000
38	0.1427	-0.1107	595.57	0.0000
39	-0.0595	-0.0691	596.14	0.0000
40	-0.2203	0.2753	604.14	0.0000



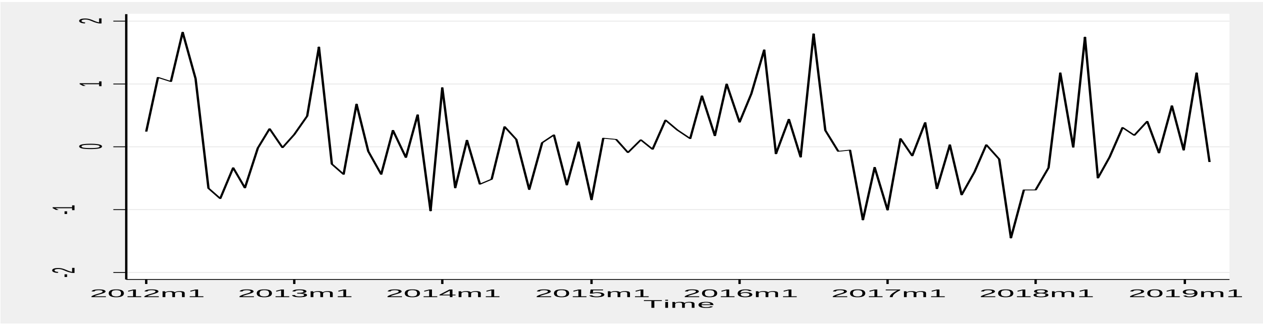
Source: Author's Computation

Fig.3 Autocorrelation and Partial autocorrelation function



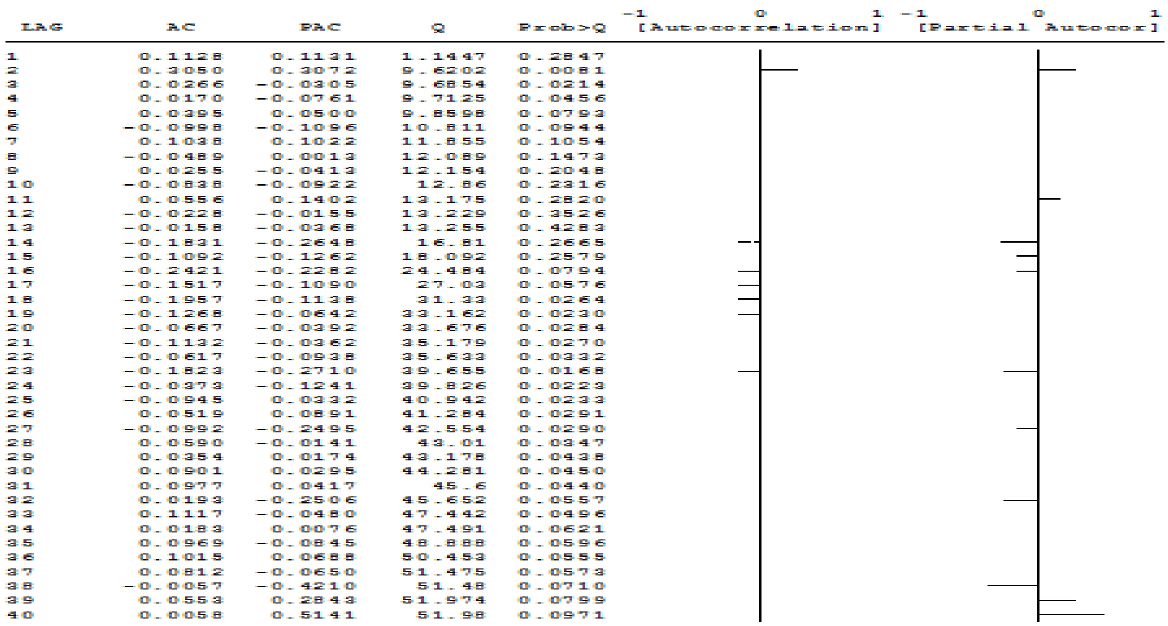
Source: Author's Computation

Fig.4 Mean temperature Residual over time for the fitted model



Source: Author's Computation

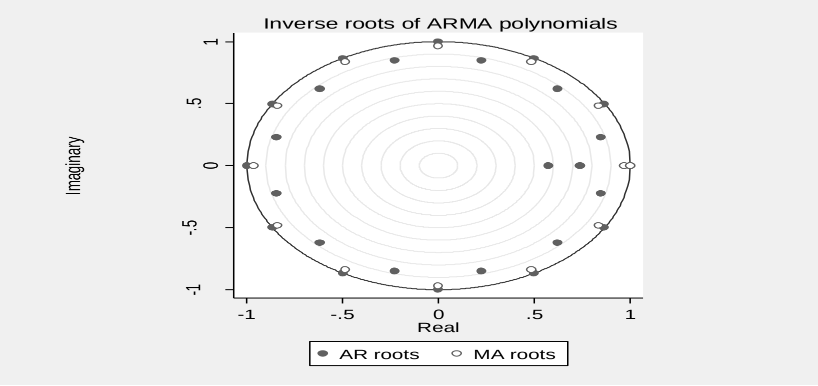
Fig.5 Correlogram of Residuals from The fitted model of Mean Temperature



LAG	AC	PAC	Q	Prob>Q
1	0.1128	0.1131	1.1447	0.2847
2	0.3050	0.3072	9.6202	0.0081
3	0.0266	-0.0305	9.6854	0.0214
4	0.0170	-0.0761	9.7125	0.0456
5	0.0395	0.0500	9.8598	0.0793
6	-0.0998	-0.1096	10.811	0.0944
7	0.1038	0.1022	11.855	0.1054
8	-0.0489	0.0013	12.089	0.1473
9	0.0255	-0.0413	12.154	0.2048
10	-0.0838	-0.0922	12.86	0.2316
11	0.0556	0.1402	13.175	0.2820
12	-0.0228	-0.0155	13.229	0.3526
13	-0.0158	-0.0368	13.255	0.4283
14	-0.1831	-0.2648	16.81	0.2665
15	-0.1092	-0.1262	18.092	0.2579
16	-0.2421	-0.2282	24.484	0.0794
17	-0.1517	-0.1090	27.03	0.0576
18	-0.1957	-0.1138	31.33	0.0264
19	-0.1268	-0.0642	33.162	0.0230
20	-0.0667	-0.0392	33.676	0.0284
21	-0.1132	-0.0362	35.179	0.0270
22	-0.0617	-0.0938	35.633	0.0332
23	-0.1823	-0.2710	39.655	0.0168
24	-0.0373	-0.1241	39.826	0.0223
25	-0.0945	0.0332	40.942	0.0233
26	0.0519	0.0891	41.284	0.0291
27	-0.0992	-0.2495	42.554	0.0290
28	0.0590	-0.0141	43.01	0.0347
29	0.0354	0.0174	43.178	0.0438
30	0.0901	0.0295	44.281	0.0450
31	0.0977	0.0417	45.6	0.0440
32	0.0193	-0.2506	45.652	0.0557
33	0.1117	-0.0480	47.442	0.0496
34	0.0183	0.0076	47.491	0.0621
35	0.0969	-0.0845	48.888	0.0596
36	0.1015	0.0688	50.453	0.0555
37	0.0812	-0.0650	51.475	0.0573
38	-0.0057	-0.4210	51.48	0.0710
39	0.0553	0.2843	51.974	0.0799
40	0.0058	0.5141	51.98	0.0971

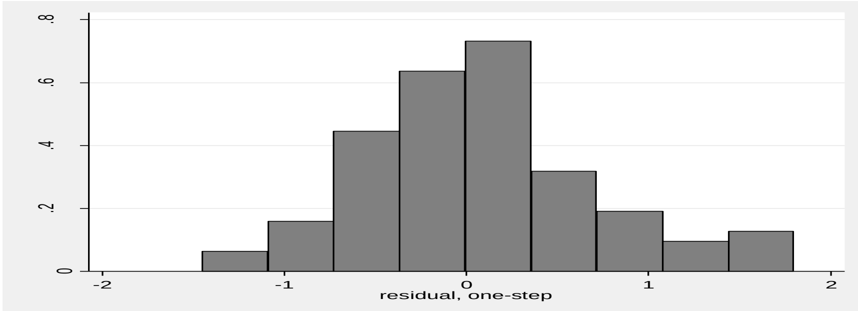
Source: Author's Computation

Fig.6 Eigenvalue stability condition



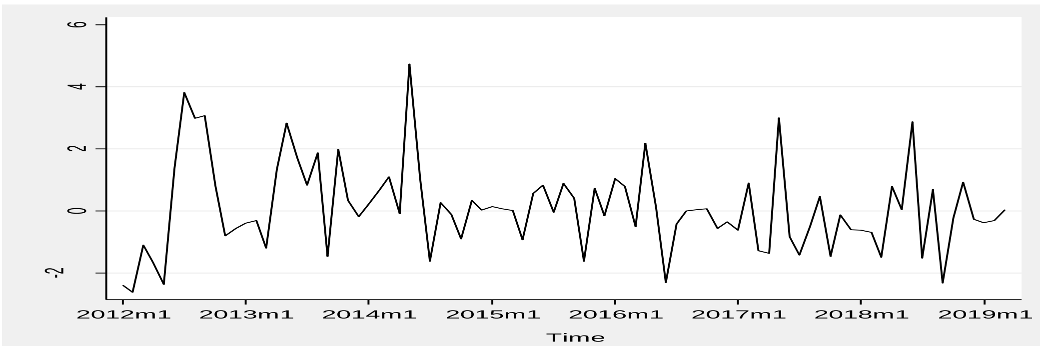
Source: Author's Computation

Fig.7 Histogram of residuals



Source: Author's Computation

Fig.8 Residual with time from the fitted model of rainfall

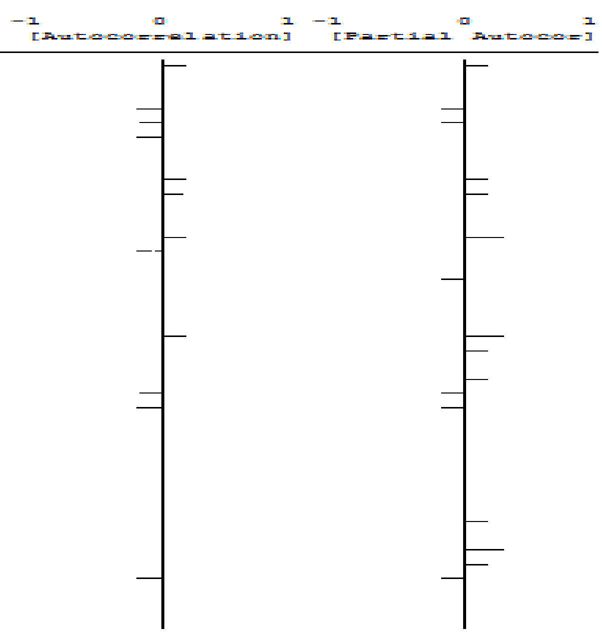


Source: Author's Computation

Fig.9 Corrolegram of the residuals from the fitted model of mean rainfall

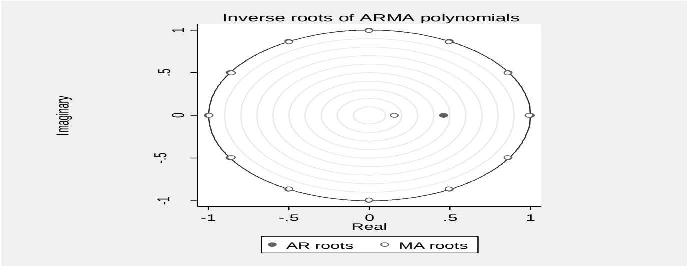
```
. corrgram ResidR
```

LAG	AC	PAC	Q	Prob>Q
1	0.2028	0.2028	3.7018	0.0544
2	0.0714	0.0388	4.1666	0.1245
3	0.0477	0.0322	4.3762	0.2236
4	-0.1370	-0.1544	6.1281	0.1898
5	-0.2232	-0.1843	10.832	0.0548
6	-0.1270	-0.0664	12.374	0.0541
7	-0.0049	0.0524	12.377	0.0888
8	-0.0276	-0.0285	12.451	0.1322
9	0.1981	0.2112	16.346	0.0600
10	0.1985	0.1819	20.307	0.0265
11	0.0777	0.0919	20.921	0.0342
12	0.0856	0.1084	21.677	0.0413
13	0.1712	0.2888	24.744	0.0249
14	-0.1352	-0.1090	26.681	0.0212
15	-0.0468	0.0877	26.917	0.0294
16	-0.0735	-0.1953	27.506	0.0362
17	-0.1215	-0.0635	29.14	0.0333
18	-0.0050	-0.0199	29.142	0.0467
19	-0.0167	-0.0932	29.174	0.0633
20	0.1662	0.2698	32.367	0.0395
21	0.0848	0.1388	33.211	0.0439
22	0.0546	0.0144	33.567	0.0543
23	0.0658	0.1383	34.09	0.0639
24	-0.1355	-0.1901	36.348	0.0508
25	-0.1289	-0.1461	38.423	0.0420
26	-0.0380	0.0874	38.606	0.0531
27	-0.0412	0.0504	38.826	0.0657
28	-0.0696	-0.1142	39.461	0.0738
29	-0.0211	-0.0146	39.521	0.0921
30	0.0493	0.0619	39.851	0.1078
31	-0.0434	0.0168	40.112	0.1265
32	-0.0369	0.0598	40.304	0.1488
33	0.1019	0.1859	41.792	0.1402
34	-0.0726	-0.0916	42.563	0.1488
35	0.0679	0.2851	43.249	0.1596
36	0.1095	0.2456	45.07	0.1429
37	-0.1451	-0.1809	48.329	0.1006
38	-0.0851	-0.1231	49.474	0.1007
39	-0.0171	0.0193	49.521	0.1205
40	-0.0166	0.0904	49.567	0.1428



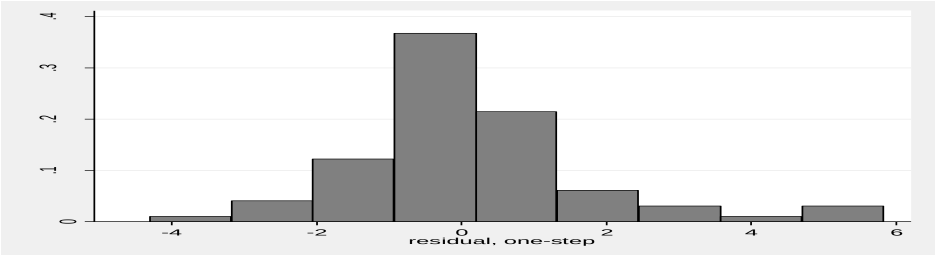
Source: Author's Computation

Fig.10 Eigenvalue Stability condition



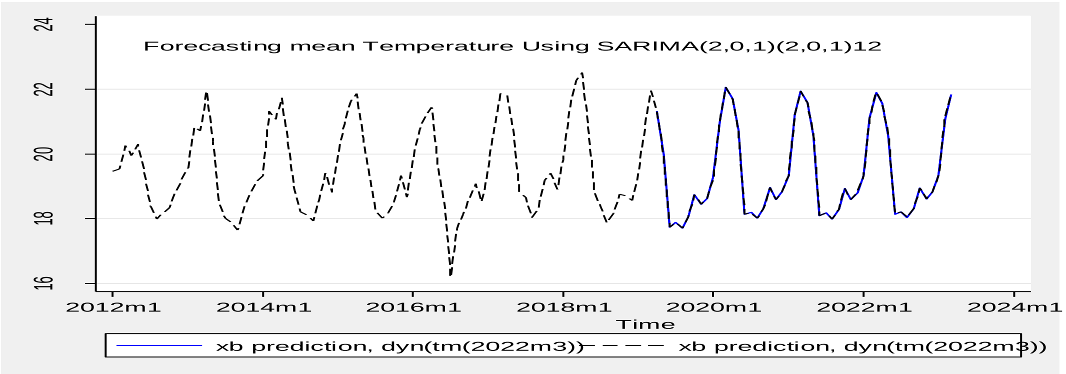
Source: Author's Computation

Fig11



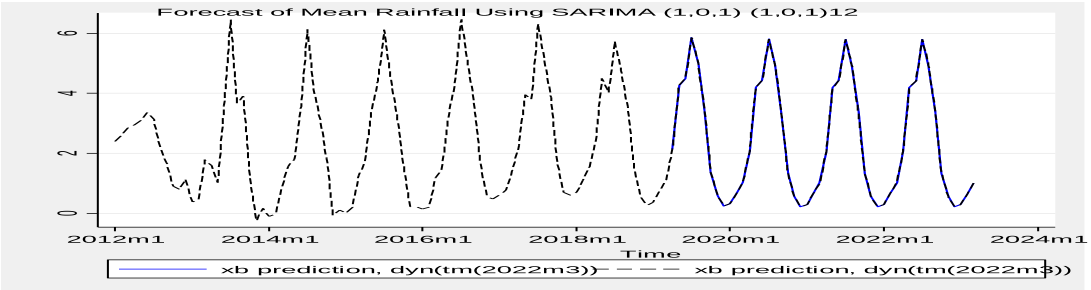
Source: Author's Computation

Fig.12 Time series plot of observed and forecasted value of Temperature



Source: Author's Computation

Fig.13 Time series plot of observed and forecasted value of Temperature



Source: Author's Computation

From Table 9, we observe that p-value is almost more than 5%, which indicates, fail to reject the null of white noise residuals (no serial correlation, homoscedasticity and Stationarity).

Portmanteau test

We cannot reject the null of white noise residuals since the p-value (0.9677) is greater than 0.05 (5%). Therefore, the residual is independently or randomly distributed which is a desirable assumption of SARIMA model.

Stability condition

The graph produced by invers roots of ARMA polynomial displays the eigenvalues with the real components on the X-axis and the imaginary components on the Y-axis. Since all the eigenvalues lie inside the unit circle, the AR parameters satisfy stability condition and the MA parameters satisfy invertibility condition. Thus, the MA process is invertible and can be represented as an infinite-order AR process.

Normality of residuals

The histogram from residuals of the fitted model (SARIMA (2, 0, 1) (2, 0, 1)12) of mean temperature have a standard normal distribution.

Heteroscedasticity

The ARCH LM test was conducted to test the existence of heteroscedasticity on the residuals of the fitted model. Since the p-value in both F-test and observed R-squared is greater than 5% we fail to reject the null of no ARCH effect (no heteroscedasticity) in the residual of the fitted model for mean temperature. Therefore, our model passes the basic assumption of constant variance error term.

Model Diagnostic for the fitted model of mean Rainfall

Randomness of the Residual from the fitted model

From the figure, we observe that the residual is randomly distributed and almost have a stable mean.

The ACF plot of the residuals of the fitted model (Fig.11) shows that the residuals are relatively small and not statistically significant. Therefore, it can be considered that the residuals of the fitted model of mean rainfall is randomly distributed

Portmanteau test

As we observe from Table 10, we cannot reject the null of white noise residuals since the p-value (0.1428) is greater than 0.05 (5%). Therefore, the residual is independently or randomly distributed which is satisfies the basic assumption of white noise residuals in SARIMA model.

Stability condition

Eigen value stability condition. Since all the eigenvalues lie inside the unit circle, AR parameters satisfy stability condition and MA parameters satisfy invertibility condition.

Normality of the residuals

The Histogram plot of the residuals was carried out to check whether residuals are normal or not.

Heteroscedasticity

From Table 11, the p-value of ARCH LM test on the results of the fitted model of mean rainfall is greater than 5%, which indicates that we fail to reject the null of constant variance (homoscedasticity).

Based on the above detailed analysis of residuals, it can be confirmed that the selected SARIMA (2,0,1) (2,0,1)12 and SARIMA (1,0,1) (1,0,1)12 model satisfies all the diagnostic tests for modeling and forecasting mean temperature and rainfall, respectively. Hence, the two models (SARIMA (2, 0, 1) (2, 0, 1)12 &are SARIMA (1, 0, 1) (1, 0, 1)12) are considered as the best model for forecasting the upcoming monthly temperature and rainfall, respectively, in Ambo area in a given period of time.

Forecasting temperature using the best fitted SARIMA Model

We have used a data from January 2012 (2012m1) to March 2019 (2019m3) to estimate the model and data from April 2019 to March 2023 for forecasting period as shown on Figure (13, &14). It appears from Figure (13 &14) that the best selected model is very well suited for forecasting the future value of the Ambo area mean temperature and rainfall since the forecasted value (blue line) have similar pattern with the actual vale (non-colored line). From the figure, we observe that almost the series have a stable value on the forecast period and shows that the estimated forecast means temperature and rainfall were identical or very close to the actual real data. The pattern of mean temperatures and rainfall in Ambo area from April 2019 to March 2023 were observed to be stationary (fluctuate around a constant mean), and hence does not follow any different pattern than the actual series.

In conclusion time series analysis is an important technique in analyzing and forecasting weather variables like temperature and rainfall. In this study, a monthly temperature and rainfall data obtained from Ambo University metrological station on the period from January 2012 to March 2019 were used to analyze the series. The seasonal autoregressive integrated moving average (SARIMA) model was used to analyze and forecast monthly mean temperature and rainfall of the study area.

The result of time plot, ACF and PACF shows that the series have a seasonal component, which confirms the appropriateness of seasonal autoregressive integrated moving average (SARIMA).The best fitting model among the competing model for monthly mean temperature and rainfall were SARIMA (2,0,1) (2,0,1)12 and SARIMA (1,0,1) (1,0,1)12 model, respectively.

The model diagnostics were performed through careful examination of the residuals from the best-fitted model, i.e. SARIMA (2,0,1) (2,0,1)12 for monthly mean temperature and SARIMA (1,0,1) (1,0,1)12 for monthly mean rainfall. The residuals were found to be following a white-noise process with a mean of zero and a constant variance, hence uncorrelated. Based on the best-fitted model, monthly mean temperature and rainfall for the next four years (from April 2019 to March 2023) were forecasted and seems to be slightly stable.

Recommendation

Based on the results of the study the following recommendations were forwarded to the concerned stakeholders.

Since the SARIMA models fitted and forecasted weather variables (in our case, monthly mean temperature and rainfall) appropriately in Ambo area, Ethiopia, so any concerned bodies can use as an input (information). Moreover, the researcher recommends to use such model for analysis of similar data in other area. However, uncertainties are in weather data, the result by itself might become indecisive. Therefore, further research based on other model is suggested for better results by the researcher.

List of acronyms and abbreviations

ACF Autocorrelation Function
ADF Augmented Dickey–Fuller Test
AIC Akaike Information Criterion
ARCH Autoregressive Conditional heteroscedasticity
AUMS Ambo University Metrological station
GDP Gross Domestic Product
IPCC Intergovernmental Panel on Climate Change
LM Lagrange multiplier test
PACF Partial autocorrelation function
SARIMA Seasonal Autoregressive Integrated Moving Average
SBICSchwarz Bayesian Information Criterion
UNFCCC United Nations Framework Convention on Climate Change

Conflicts of interests

The authors declare no conflict of interest.

References

Ademola, A., Emmanuel, C. O., and Aderemi, K. A. 2018. Statistical Modeling of Monthly Rainfall in Selected Stations in Forest and Savannah Eco-climatic Regions of Nigeria. *Journal of Climatology & Weather Forecasting*. 6:S1. DOI: 10.4172/2332-2594.1000226

Akaike, H., 1974: Anew look at the statistical model identification. IEEE Trans. Autom. Control, 19, 716–723, doi:10.1109/ TAC.1974.1100705.

Box GEP, Pierce, D. A. 1970. Distributions of residual autocorrelations in autoregressive integrated moving average models. J American Stat Assoc 72: 397-402.

BOX, G. E. P. & JENKINS, G.M. 1976. Time Series Analysis: Forecasting and Control. Revised Edition, Holden-Day: San Francisco, CA.

Box, G. E. P. and Jenkins, G. M. 1970. Time series analysis: forecasting and control. San Francisco, Holden Day, p. 575 pp.

Chatfield, C. 2004. The analysis of time series: An introduction, 6th ed. London, UK: Chapman & Hall/CRC.

Dickey, D. A., and Fuller, W. A. 1979. Distribution of the estimators for the autoregressive time series with a unit root. J. Amer. Statist. Assoc., 74, 427-431.

Engle, R. F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Journal of Econometrics, Vol 50, P 987-1007.

Faraway, J., Chatfield, C. 1998. "Time series forecasting with neural networks: a comparative study using the airline data", *Applied Statistics*. 47: 231–250.

Intergovernmental Panel on Climate Change (IPCC). Climate Change 2007: Synthesis Report. An Assessment of the Intergovernmental Panel on Climate Change; IPCC: Valencia, Spain, 2007.

Jones, J. R., Schwartz, Ellis, K. N., Hathaway, J. M. and. Jawdy, C. M. 2015."Temporal variability of precipitation in the Upper Tennessee Valley," Journal of Hydrology: Regional Studies, 3; 125–138.

Ljung, G. and Box, G. E. P. 1978. On a Measure of Lack of Fit in Time Series Models, Biometrika, 66, 67–72.

Mehmet Tektaş. 2010.Weather Forecasting Using ANFIS and ARIMA MODELS. A Case Study for Istanbul. En*vironmental Research, Engineering and Management: 1(51): 5 – 10.*

Robinson, S., Willenbockel, D. and Strzelecki, K., 2012. A Dynamic General Equilibrium Analysis of Adaptation to Climate Change in Ethiopia. Review of Development Economics 16, 489-502.

Schlenker, W., Lobell, D. B. 2010. Robust negative impacts of climate change on African agriculture. Environ. Res. Lett. 5, 014010.

Schwarz, G. E. 1978. Estimating the dimension of a model. *Annals of Statistics*. 6 (2):461–64.

Tektaş, M. 2010. Weather forecasting using ANFIS and ARIMA models. A case study for Istanbul. Environmental Research, Engineering and Management 51 (1):5–10.

Thornton, P. K. *et al*. 2011. Agriculture and food systems in sub-Saharan Africa in a 4°C+ world. *Philosophical Transactions of the Royal Society*.369: 117-136.

United Nations Framework Convention on Climate Change (UNFCCC). 2007. Climate change impact vulnerabilities and adaptation in developing countries.

How to cite this article:

Teshome Hailemeskel Abebe. 2020. Time series Analysis of Monthly Average Temperature and Rainfall Using Seasonal ARIMA Model (In Case of Ambo Area, Ethiopia). *Int.J.Curr.Res.Aca.Rev*. 8(11), 31-47.
doi: https://doi.org/10.20546/ijcrar.2020.811.005