Statistics and Applications
Volume 9, Nos. 1&2, 2011 (New Series), pp. 31-53

Modeling and forecasting of rainfall data of mekele for Tigray region (Ethiopia)

Amaha Gerretsadikan and M.K.Sharma
Addis Ababa University, Addis Ababa, Ethiopia

Abstract

In this paper we have attempted to build a seasonal model of monthly rainfall data of Mekele station of Tigray region (Ethiopia) using Univariate Box-Jenkins's methodology. The method of estimation and diagnostic analysis results revealed that the model was adequately fitted to the historical data. In particular, the residual analysis, which is important for diagnostic checking confirmed that there is no violation of assumptions in relation to model adequacy. Further comparison on the forecasting accuracy of the model is performed by holding-out some rainfall values. The point forecast results showed a very closer match with the pattern of the actual data and better forecasting accuracy in validation period.

Key words: Box-Jenkins's model; Rainfall; Forecasting

1 Introduction

Univariate time series analysis and forecasting has become a major tool in hydrology, environmental management, and climatic fields. Several time series methods have been used for modeling and forecasting rainfall data in literatures but according to Pankratz (1983) the Box and Jenkins method is the most general way of approaching to forecast unlike other models, there is no need to assume initially a fixed and specified pattern. The Univariate Box and Jenkins models are useful for analysis of single time series.

Montgomery and Johnson (1976) considered Box and Jenkins methodology as probably the most accurate method for forecasting of time series data. According to Caldwell (2006), the Box-Jenkins methodology is particularly suited for development of model of process exhibiting strong seasonal behavior. There are other forecast techniques exploring the relation among observations yield better results; most of these forecast techniques are based on recent advances in time series analysis consolidated and developed by Box and Jenkins (1976) and further discussed in other resources such as Chatfield (1996).

Nail and Momani (2009) used Univariate Box-Jenkins approach and revealed that this approach possesses many appealing features such as the researcher who has a data for the past period, for example rainfall, to forecast future rainfall without having to search for other related time series data.

In this paper we have also used Box- Jenkins approach to build a seasonal model of monthly rainfall data of Mekele station in Tigray region (Ethiopia). The estimation and diagnostic analysis results revealed that the model is well fitted to the historical data. The residual analysis revealed that there was no violation of assumptions in relation to model adequacy. Further we compared the forecasting accuracy of the model by holding-out some rainfall values. The point forecast results showed a very closer match with the pattern of the actual data and better forecasting accuracy in validation period.

2 Material and Methods

2.1 Material

The National Meteorological Service Agency (NMSA), Ethiopia, is the responsible organization for the collection and publishing of meteorological data. The monthly rainfall data from the period January 1975 – December 2009 of Mekele station of Tigray region were taken from NMSA (Appendix).

2.2 Methodology

In this article we used Seasonal Autoregressive Integrated Moving Average (SARIMA) model, proposed by Box and Jenkins (1976), for model building and forecasting for rainfall data. The Box and Jenkins methodology is a powerful approach to the solution of many forecasting problems (Johnson and Montgomery, 1976) and it can provide extremely accurate forecasts of time series and offers a formal structured approach to model building and analysis. There are many quantitative methods of model building and forecasting which are being used in climatology and metrological studies. With the development of the statistical software packages and its availability, these techniques have become easier, faster and more accurate to use. In this study, we employ SAS and SPSS software packages for the statistical data analysis.

The Box- Jenkins methodology assumes that the time series is stationary and serially correlated. Thus, before modeling process, it is important to check whether the data under study meets these assumptions or not. Let $x_1, x_2, x_3, \ldots, x_{t-1}, x_t, x_{t+1}, \ldots, x_t$ be a discrete time series measured at equal time intervals. A seasonal ARIMA model for x_t is written as [Box and Jenkins, 1970]

$$\phi(\mathrm{B})\Phi(\mathrm{B}^s)\{[(1-\mathrm{B})^d(1-\mathrm{B}^s)^D x_t]-\mu\}=\theta(\mathrm{B})\Theta(\mathrm{B}^s)a_t$$

Or (1)

$$\phi(B)\Phi(B^s)(w_t-\mu)=\theta(B)\Theta(B^s)a_t$$

where

x_t is an observation at a time t;

t discrete time;

s seasonal length, equal to 12 ;

μ mean level of the process, usually taken as the average of the w_t series (if D + d > 0 often $\mu \equiv 0$);

a_t normally independently distributed white noise residual with mean 0 and variance σ_a^2 (written as NID (0, σ_a^2)

$\phi(\mathrm{B})=1-\phi_1\mathrm{B}-\phi_2\mathrm{B}^2-\ldots-\phi_p\mathrm{B}^p$ nonseasonal autoregressive (AR) operator or polynomial of order p such that the roots of the characteristic equation $\phi(B)=0$ lie outside the unit circle for nonseasonal stationarity and the ϕ_i, i = 1, 2, . . . , p are the nonseasonal AR parameters;

$(1-\mathrm{B})^d = \nabla^d$ nonseasonal differencing operator of order d to produce nonseasonal staionarity of the d th difference, usually d = 0, 1, or 2;

$\Phi(B^s)=1-\Phi_1 B^s-\Phi_2 B^{2s}-\ldots-\Phi_p B^{ps}$ seasonal AR operator or order p such that the roots of $\Phi(B^s)=0$ lie outside the unit circle for seasonal stationarity and Φ_i, i = 1, 2, . . . , p are the seasonal AR parameters;

$(1-\mathrm{B})^D = \nabla_s^D$ seasonal differencing operator of order D to produce seasonal stationarity of the Dth differenced data, usually D = 0, 1, or 2;

$w_t = \nabla^d\nabla_s^D x_t$ stationary series formed by differencing x_t series ($n = N - d - s\,D$ is the number of terms in the w_t series);

$\theta(B)=1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q$ nonseasonal moving average (MA) operator or polynomial of order q such that roots of $\theta(B)=0$ lie outside the unit circle for invertibility and θ_i, i = 1, 2, . . ., q;

$\Theta(B^s)=1-\Theta_1 B^s-\Theta_2 B^{2s}-\ldots-\Theta_q B^{Qs}$ seasonal MA operator of order Q such that the roots of $\Theta(B^s)=0$ lie outside the unit circle for invertibility and Θ_i, i = 1, 2, . . . , Q are the seasonal MA parameters.

The notation (*p, d, q*) × (*P, D Q*)$_s$ is used to represent the SARIMA model (1). The first set of brackets contains the order of the nonseasonal operators and second pair of brackets has the orders of the seasonal operators. For example, a stochastic seasonal noise model of the form (1, 1, 2) × (0, 1, 1)$_s$ is written as

$$(1 - \phi_1 B)\{[(1-B^s)x_t] - \mu\} = (1-\theta_1 B - \theta_2 B^2)(1-\Theta_1 B^s)a_t \qquad (2)$$

If the model is non seasonal, only the notation (*p, d, q*) is needed because the seasonal operators are not present.

When a seasonal model is stationary and requires no differencing (i.e. $D = 0$ and $d = 0$), it is often referred to simply as an ARMA (autoregressive moving average) process. The notation (*p, q*) × (*P, Q*)$_s$ is used to represent this type of model. If an ARMA model is nonseasonal, the notation (*p, q*) is used to indicate orders of the AR and MA operators, respectively. A pure nonseasonal AR process of order p with no differencing is often denoted by AR (*p*). Likewise, a nonseasonal MA process of order *q* is sometimes written as MA (*q*). Of course an AR (*p*) model can be represented equivalently by the notation (*p*, 0) or (*p*, 0, 0), while MA (*q*) process can also be denoted by (0, *q*) or (0, 0, *q*).

2.3 Tests for Stationary

Graphic Inspection: The pattern of the time series plot (Fig.1) does not show any apparent systematic change about the mean. The periodic peaks in the plot, however, reflect the yearly regular seasonality (with seasonality interval *s*=12) of the rainfall values. The series is, therefore, seasonal due to a large rainfall values during rainy season and a relatively lesser peak due to small values of rainfall in the other months. This indicates that the rainfall data have seasonal unit root (i.e., seasonally not stationary).

The Figure 2 exhibits the autocorrelation function plot of untransformed data in which the presence of seasonality behavior as well as seasonally non stationary of the rainfall series is clear. Because there is a sinusoidal wave pattern at the multiple of seasonal intervals and declining slowly while non seasonal lags are relatively decaying quite slowly. It is**,** thus, necessary to remove the non seasonal component of the time series corresponding to the sinusoidal periodic component of the autocorrelation function to make series seasonally stationary.

Dickey-Fuller Test: The most widely used test for stationary is Dickey-Fuller test. This test is based on the estimate of the following regression equation with no deterministic trend.

$$\nabla x_t = \phi x_{t-1} + \gamma_1 \nabla x_{t-1} + \ldots + \gamma_p \nabla x_{t-p} + a_t \qquad (3)$$

where ∇ is the difference operator defined $\nabla x_t = x_t - x_{t-1}$ and x_t is a variable of interest. This model can be estimated and tested for a unit root. That is equivalent to testing ϕ equal to zero, $\gamma_1, \ldots, \gamma_p$ are p regression coefficients and p is the number of autoregressive terms.

To test the hypothesis that the series x_t is stationary, we formulate the following hypothesis

H_O: The series is non-stationary i.e $\phi = 0$

H_1: The series is stationary i.e $\phi < 0$ at $\alpha=0.05$.

There is a need of seasonal differencing not simple differencing.

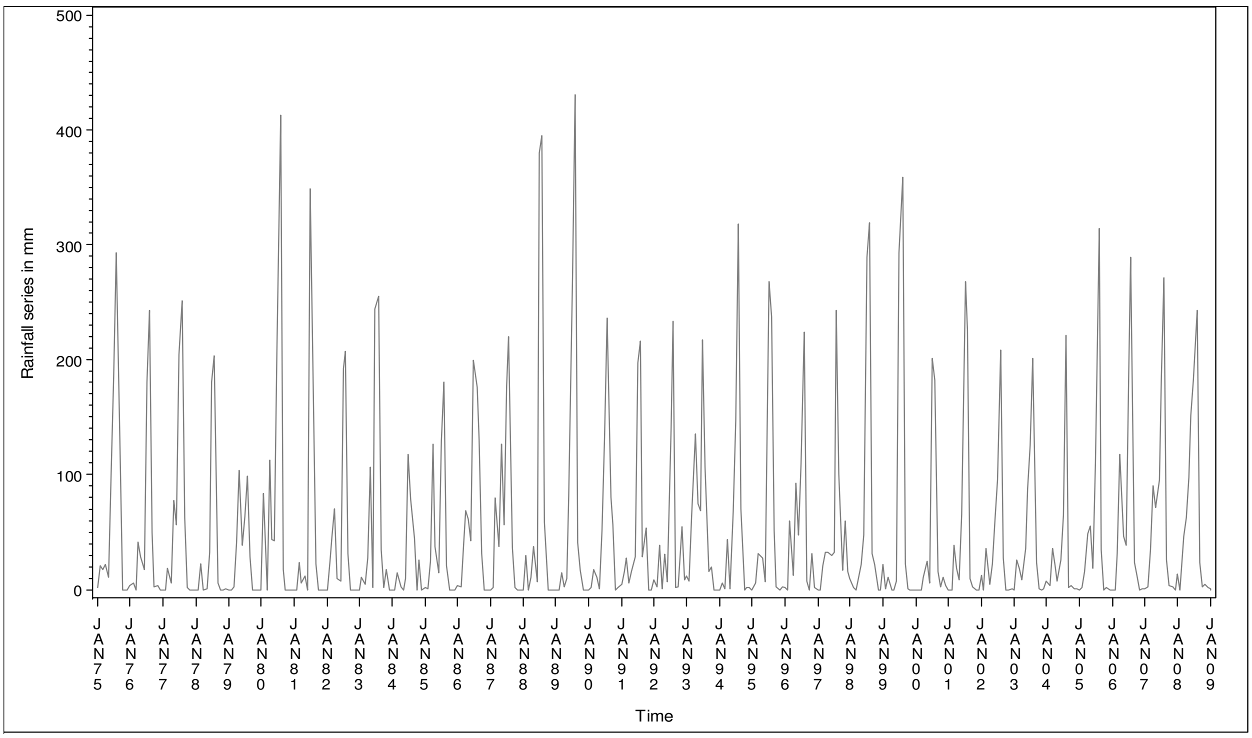


Figure 1: Plot of monthly rainfall data

The patterns of monthly rainfall series plot and autocorrelation function suggest the need of seasonal differencing but not simple differencing.

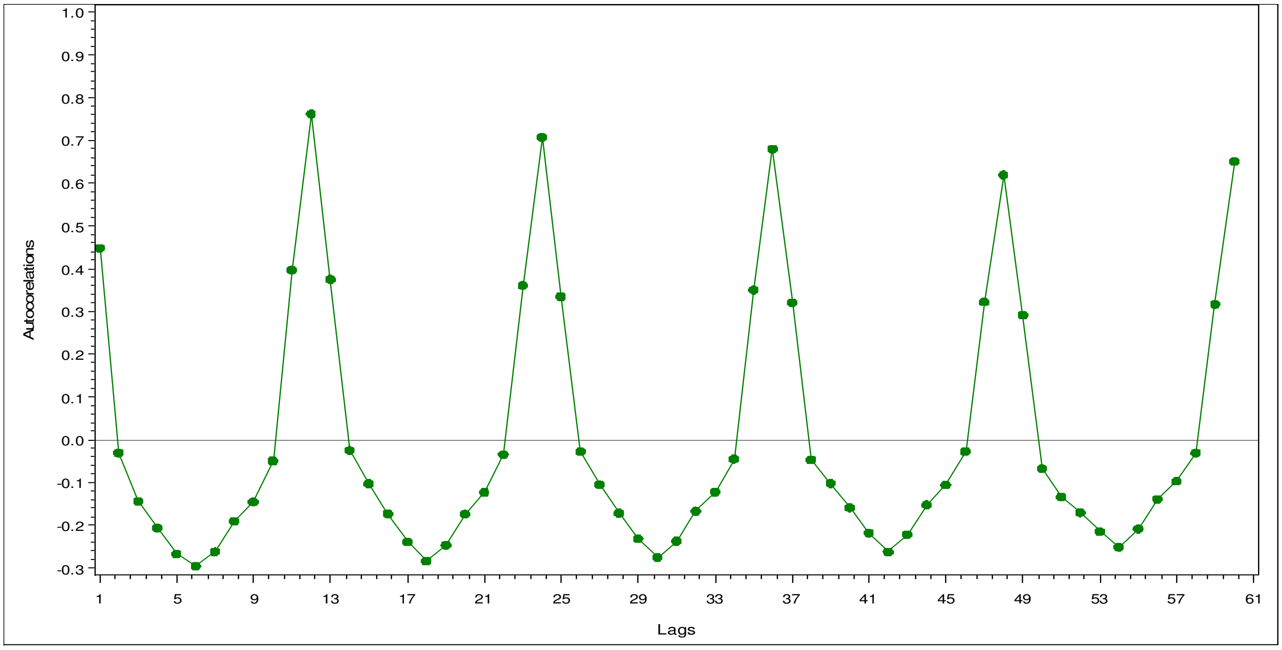


Figure 2: Autocorrelation plot for the untransformed monthly rainfall series.

Usually the order of p in the regression equation is set to three. Then if the estimate of ϕ is nearly zero in the fitted regression equation (2), the original series x_t needs first differencing and if the estimate of $\phi < 0$ then the original series is already stationary (Makridakis *et al.*, 1998). It was found that the estimated value for ϕ = -0.41 which confirm that original time series plot is without obvious trend at 5% significance level. The autocorrelation function in Fig.2 exhibits non - seasonally rapidly decaying trend. As a result, both tests appear to agree to avoid first non seasonal simple differencing.

Variance Comparison: The behavior of variance associated with different orders of differencing can provide a useful means of deciding the appropriate order of differencing (Mills, 1999). The rule is that the when the sample variance does not decrease further then a stationary series is found. If the increase in the differencing order increases the variance, it is an indication of over differencing. To examine our series that whether it is a candidate of non-differencing, simple differencing, seasonal differencing or double seasonal differencing for non seasonally and seasonally stationarity, we computed the sample variance for each of x_t, ∇x_t, $\nabla_{12} x_t$, and $\nabla_{12}^2 x_t$ series, respectively. We got the following results:

- Var(∇x_t)=8064.1, Var(x_t)=7106.5, Var($\nabla_{12} x_t$)=2745.8, and $\nabla_{12}^2 (x_t) = 7850$ values;
- Var(∇x_t) > Var(x_t);
- Var ($\nabla_{12}^2 x_t$) > Var (x_t) > Var ($\nabla_{12} x_t$).

These results suggest that non-seasonal first differencing (∇x_t) has been over-differenced and hence the original series is non-seasonally stationary .The first seasonal differencing would rather be important, because the Var ($\nabla_{12}^2 x_t$) is greater than Var($\nabla_{12} x_t$).

These tests for stationarity seem to agree and suggest that the first seasonal differencing in the series can achieve stationarity around a constant mean, which is approximately zero and its standard deviation is 52.4 mm (Figure 3). Moreover, the ACF and PACF (Figures 4(a) and 4(b)) also tell that the monthly rainfall series is stationary in both mean and variance after first seasonal difference.

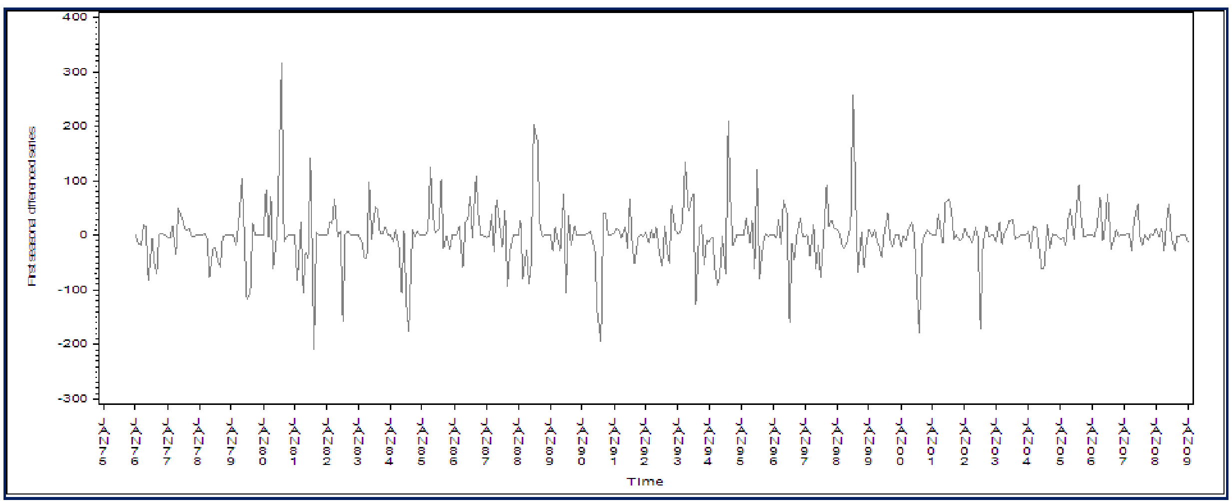
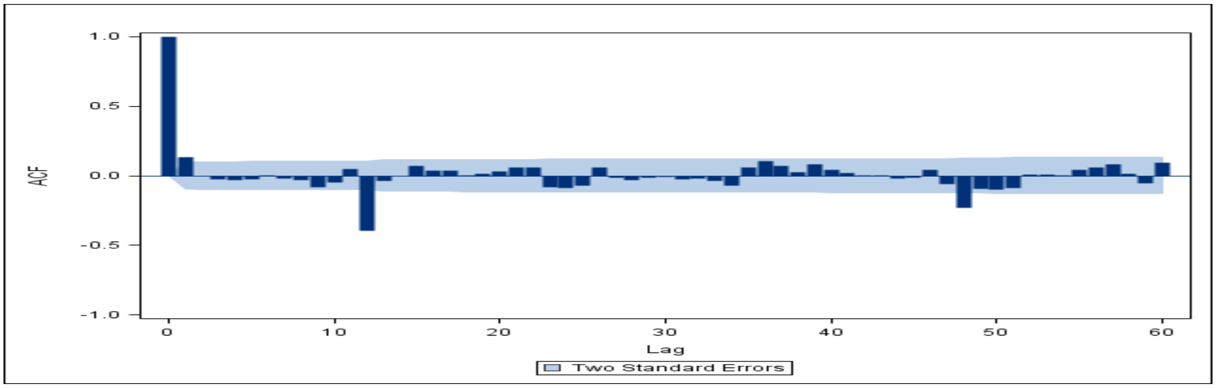
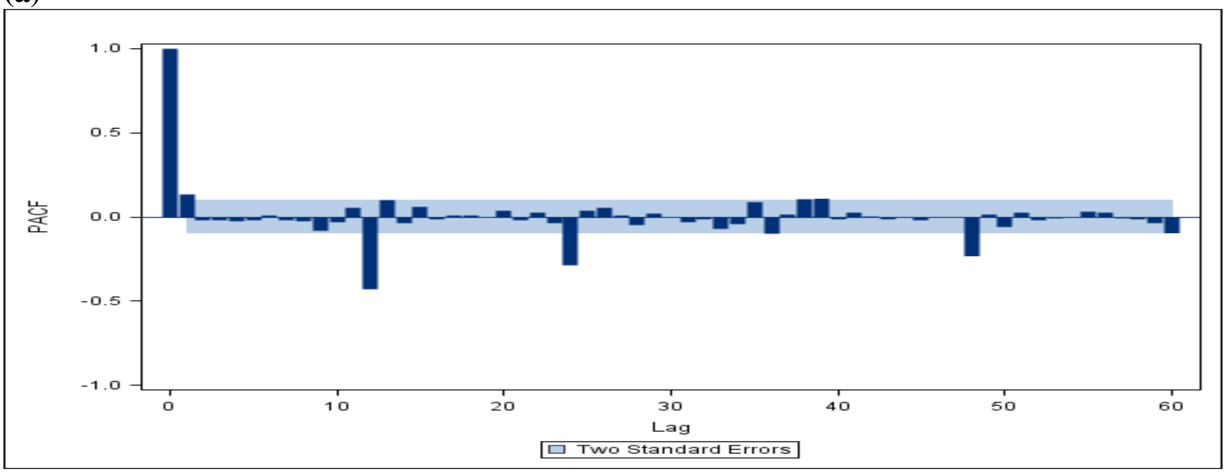


Figure 3: Plot for First seasonal differenced monthly rainfall series



(a)



(b)

Figure 4: **(a):** Autocorrelation Function (ACF) **(b)**: Partial Autocorrelation Function (PACF) for the first seasonal differenced monthly rainfall**.**

2.4 Tests for randomness

According to Harvey (1993) the simplest time series is a random model, in which the observations vary around a constant mean, have a constant variance, and are probabilistically independent. In other words, a random time series does not have time series pattern, meaning that there is no point in attempting to fit a time series model to such type of data. Therefore, it is important to perform tests of randomness before any attempt to modeling process to our series. Therefore we check our time series through the following tests to investigate the hypothesis that the first-seasonally differenced monthly rainfall series are serially uncorrelated.

Graphic Inspection: The visual inspection of the autocorrelation function plot provides useful information to identify the type of time series (Chatfield, 1996). For example, if a time series is a completely random series, then for large n, $r_k \approx 0$ for every k. This can be examined after the array of autocorrelation coefficients r_k, plotted with k as abscissa and r_k as ordinate.

Figure 4(a) exhibits the graph of sample autocorrelations against different lags from which we can observe visually that the autocorrelations are not all insignificant. This indicates that there is some sort of dependence between values of $\nabla_{12} x_t$ series.
The randomness can also be checked using Bartlett's Band Test and Box-Ljung Test Statistic. Here we used Box-Ljung Test Statistic.

Box-Ljung Test Statistic: This statistic is used for collectively testing the magnitude of the autocorrelation of stationary time series for significance. For this test, we used the sample autocorrelation coefficients of the first seasonally differenced monthly rainfall as well.
The hypothesis to be tested is

H_O: All autocorrelations up to lag J are zero

Versus

H_1: Not all up to lag J are zero at $\alpha = 0.05$.

The statistic for this testing hypothesis is as

$$Q = n(n+2)\sum_{j=1}^{J}\frac{r_j^2}{n-j} \tag{4}$$

This statistics has a chi-square distribution with J degrees of distribution. Q-Statistic is usually computed for J= 6, 12 24, 36 and 48 by most of the statistical packages. However, J=12 or 24 will prove to be satisfactory (Patricia, E. G., 1994). In this regard, we compute the test statistic above for the first J=12 lag autocorrelation values and n=408 observations. The value of the calculated Q-Statistic is found to be 43.72 and the tabulated value for chi-distribution with 12 degree of freedom at 0.05 significance level is 21.02. The decision to reject H_o is based on whether the value of Q-Statistic > $\chi^2_{0.05}, J$; *if* that does not hold we do not reject H_O. Since Q-statistic=43.7> $\chi^2_{0.05}$=21.2, we reject H_o. we conclude that the seasonally first differenced monthly rainfall series are serially correlated.

Now we can say that the monthly rainfall data are stationary and serially correlated.

2.5 Model Identification

Having established that the monthly rainfall data are serially correlated and stationary, the next step in the identification process is to find the initial values for the order of non-seasonal and seasonal parameters *p, q, P*, and *Q*,, respectively. The first step in this direction is to identify the significant autocorrelations and partial autocorrelation from the ACF and PACF plots of the underlying stationary series (Hipel et al., 1977). Hence for the $(1-B^{12})x_t$ *series*, where B is the backward shift operator and is defined $Bx_t = x_{t-1}$ and B^d is the backward shift operator of order d, we find significant ACF at lag k=1, 12 and k=48, see Figure 4(a). Hence, based on the ACF behavior, we guess Seasonal Autoregressive Integrated Moving Average (SARIMA) model (0, 0, 1) × (1, 1, 4)$_{12}$ of the following form.

$$(1-\Phi_{12} B^{12})(1-B^{12})\, x_t = (1-\theta_1 B)(1-\Theta_{48}B^{48})a_t \qquad (5)$$

Another alternative models seem to be appropriate tentatively at this stage is based on the principle that when the process is a purely SARIMA *(p, d, 0) × (P, D, 0)*$_{12}$ model, r_{kk} cut off and is not significantly from zero after lagp+SP. If r $_{kk}$ damps out at lags that are multiple of s , this suggests the incorporation of a seasonal moving average (MA) component into the model. The failure of the PACF to truncate at other lags may suggest that a non-seasonal MA term is required (Hipel et al., 1977). Accordingly, we guess SARIMA (1, 0, 0) × (4, 1, 1)$_{12}$ model.

2.6 Model Estimation and Diagnostics checking

Non-linear Estimation of the parameters for Box-Jenkins models is a quite complicated. Parameter estimates are usually obtained by maximum likelihood method, which is asymptotically correct for time series (Brockwell and Davis, 1996). Applying maximum likelihood method of estimation, we got the following estimated values of the parameters of SARIMA (0,0,1) × (1,1,4)$_{12,}$ and (1,0,0) × (4,1,1)$_{12}$ as given in Table 3.

Table 3: Parameter estimates for suggested SARIMA models.

(a): $(1-\Phi_{12}B^{12})(1-B^{12})x_t = (1-\theta_1 B)(1-\Theta_{48}B^{48})a_t$ or (0,0,1)× (1,1,4)$_{12}$

(b): $(1-\phi_1 B)(1-B^{12})x_t=(1-\theta_{48} B^{48})(1-\Theta_{12}B^{12})a_t$ or 0,0)×(4,1,1)$_{12}$

(c): $(1-\phi_1 B)(1-B^{12})\, x_t =(1-\theta_{48} B^{48})(1-\Theta_{12}B^{12}-\Theta_{24}B^{24})a_t$

Model	Parameter	Estimate	Standard Error	t-value	P-value	Fit statistics
(a)	ϕ_1	-0.19	0.05	-3.29	0.001	AIC=4019.04
	θ_{48}	0.83	0.04	22.15	<0.0001	RMSE.=37.53
	Φ_{12}	-0.24	0.05	-4.80	<0.0001	r^2 =0.80
(b)	ϕ_1	0.16	0.05	3.39	0.0007	AIC=4018.80
	θ_{48}	-0.24	0.05	-4.79	<0.0001	RMSE.= 38.50
	Θ_{12}	0.93	0.04	21.92	<0.0001	r^2 =0.79
(c)	ϕ_1	0.15	0.05	3.06	0.0022	AIC=4018.30
	θ_{48}	-0.23	0.05	-4.49	<0.0001	RMSE.= 38.24
	θ_{12}	0.87	0.07	12.80	<0.0001	r^2 =0.81
	θ_{24}	0.09	0.06	1.68	0.09	

After we have derived models and we should allow for additional parameters in the fitted model, and determine whether or not their estimates are statistically significantly different from zero. If they are, then there is cause for concern that we have not identified the model correctly. For example, we start with over fitting by including one more seasonal moving average parameter (which measures the error dependency effect at lag 24 and denoted by Θ_{24}) to the SARIMA model (b) to examine whether this model with more parameters would adequately be fitted to the seasonally first differenced monthly rainfall data. The inclusion of this parameter can be determined by testing its significance and the improvement in the measures of goodness of fit of the model. All substantial parameters in all the models in Table 4 showed statistically significance except the SARIMA model (c) in which we have added one more parameter. One estimated parameter in (c) (Θ_{24}= 0.06, P-value=0.09 >0.05) which is insignificant. As a result, inclusion of this parameter (Θ_{24}) has no visible contribution in the model (c). It means models (a) and (b) have correctly identified.

Table 4: Correlations of Parameter Estimates for the two models

Model (a): SARIMA (0,0,1) × $(1,1,4)_{12}$				Model b:SARIMA (1,0,0) × $(4,1,1)_{12}$			
Parameter	θ_1	θ_{48}	Φ_{12}	Parameter	ϕ_1	θ_{48}	Θ_{12}
θ_1	1.00	0.10	-0.06	ϕ_1	1.0	-0.08	0.22
θ_{48}		1.00	0.21	θ_{48}		1.00	0.04
Φ_{12}			1.00	Θ_{12}			1.00

Outliers, level shifts, and variance changes are common place in applied time series. The presence of these could easily misled the conventional time series analysis procedure resulting erroneous conclusion. In the estimation procedure, two types of outliers (5 additive and 1 shift outliers) were detected and adjusted in the fitted models by SAS software. AIC values have been calculated by the following formula.

$$\text{AIC}=-2 \ln (\text{maximum likelihood}) +2m \tag{6}$$

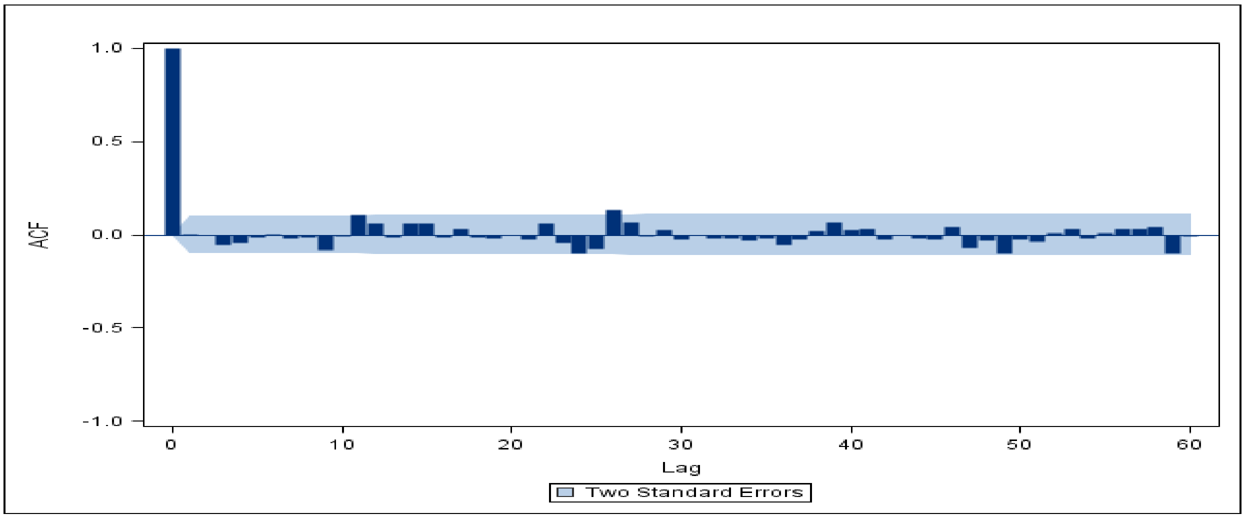
Where m is the number of seasonal and non-seasonal autoregressive and moving average parameters to be estimated.

Now we proceed to check the adequacy of these two models using residual analysis. The residual analysis is a part of diagnostic checking and test for white noise and normality of residuals. In this checking the Autocorrelations Functions (ACF) and Partial Autocorrelations Functions (PACF) of the residuals resulted from the fitted models should not show any pattern (trend or seasonality pattern). And also for a correctly fitted model the residuals correlation coefficients should not lie outside the two standard error at a given significant level.

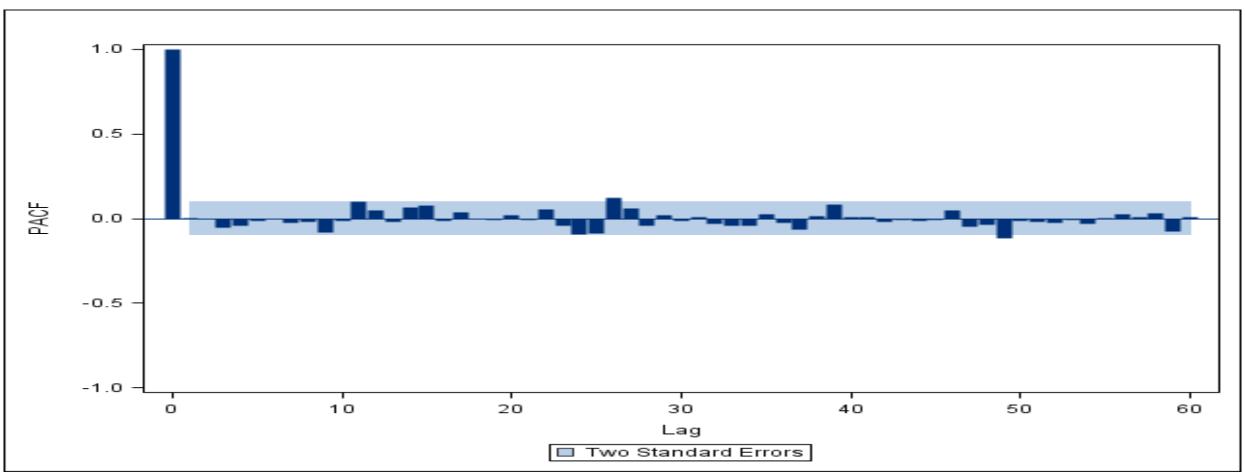
It is clear, from Figures 5 (a and b) and Figure 6(a and b) that there is no pattern in residuals ACF and PACF plot for model (a) and (b), respectively. No ACF or PACF coefficients lie outside the two standard errors at 5% level of significance for both fitted models. The graphical analysis also shows that the residuals in the model appeared to fluctuate randomly around zero with no apparent pattern (Figure 7). The figure 5(c) exhibits the residual histogram (normal curve) and we find that there is no violation of the models' assumption i.e. the residuals should normally distributed with mean zero and constant variance.

From plots in Figure 5(d) and Figure.6(c), it is obvious that the set of autocorrelation for residuals are not significant and we cannot reject the hypothesis that the autocorrelations of the residuals are zero.

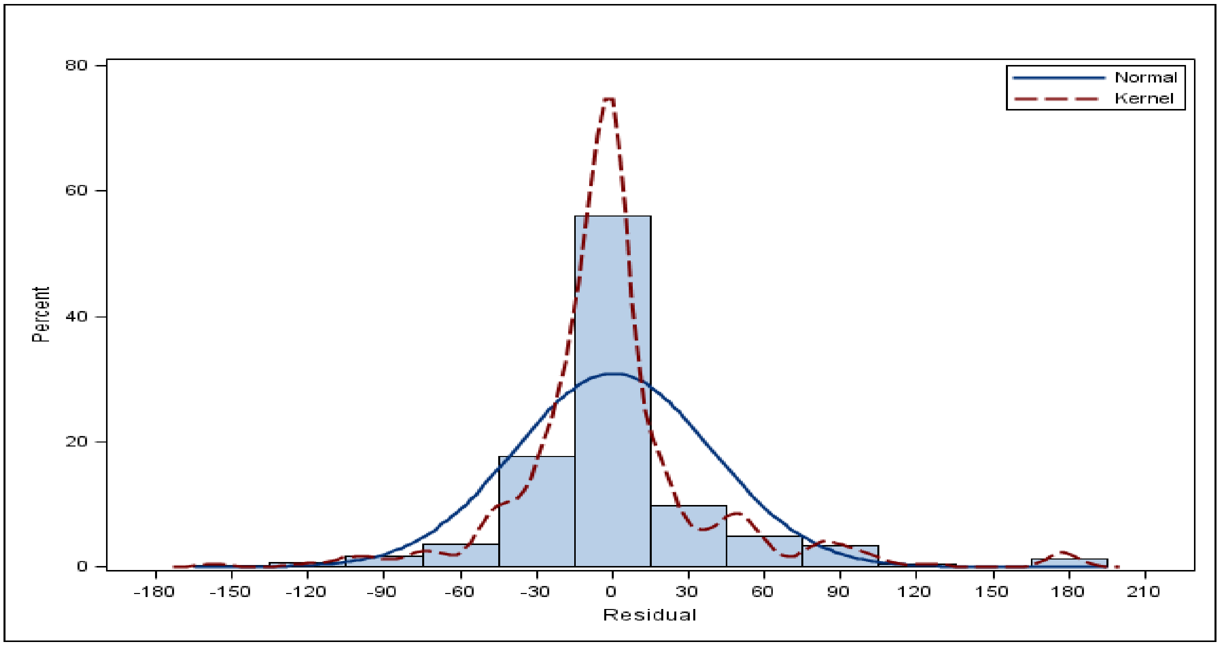
These results are in agreement with the hypothesis that the residuals resulted from each of the suggested models do not show any correlation or pattern and these are normally distributed, we conclude that the two SARIMA (0, 0, 1) × (1, 1, 4) and (1, 0, 0) × (4, 1, 1) models are found to be adequately fitted to the seasonally first differenced monthly rainfall series.



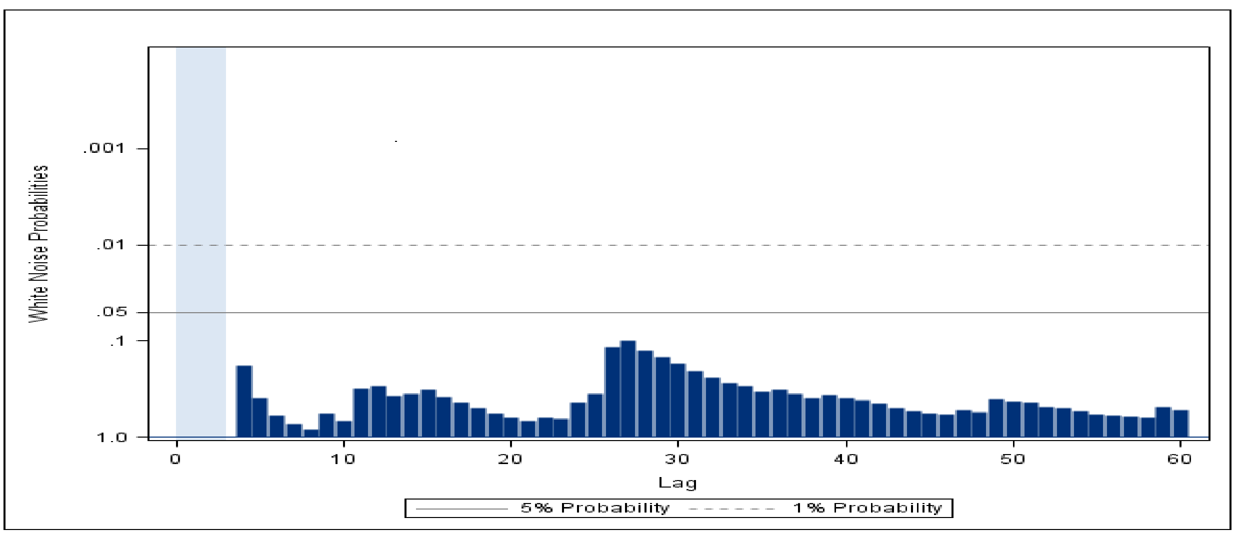
(a)



(b)

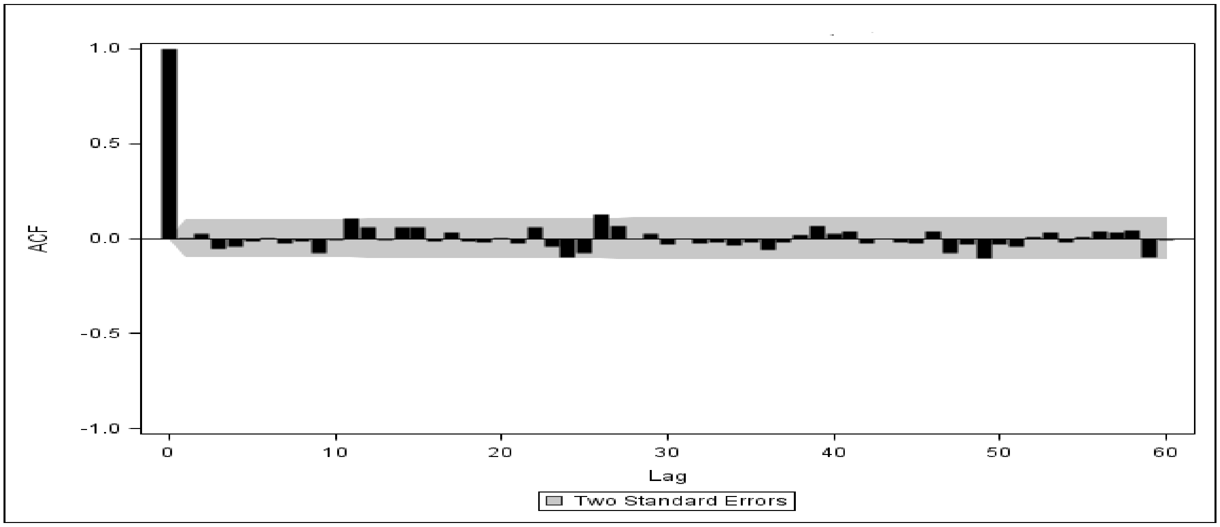


(c)

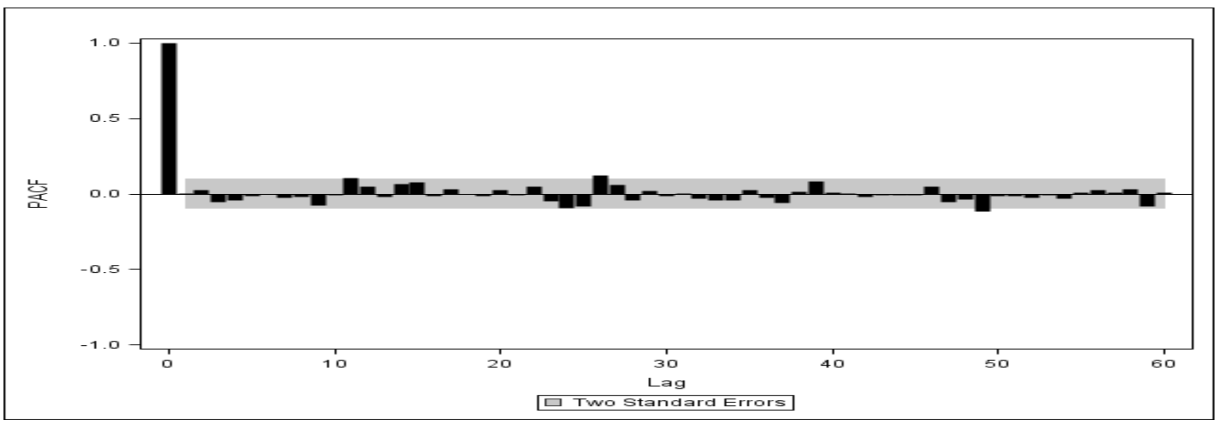


(d)

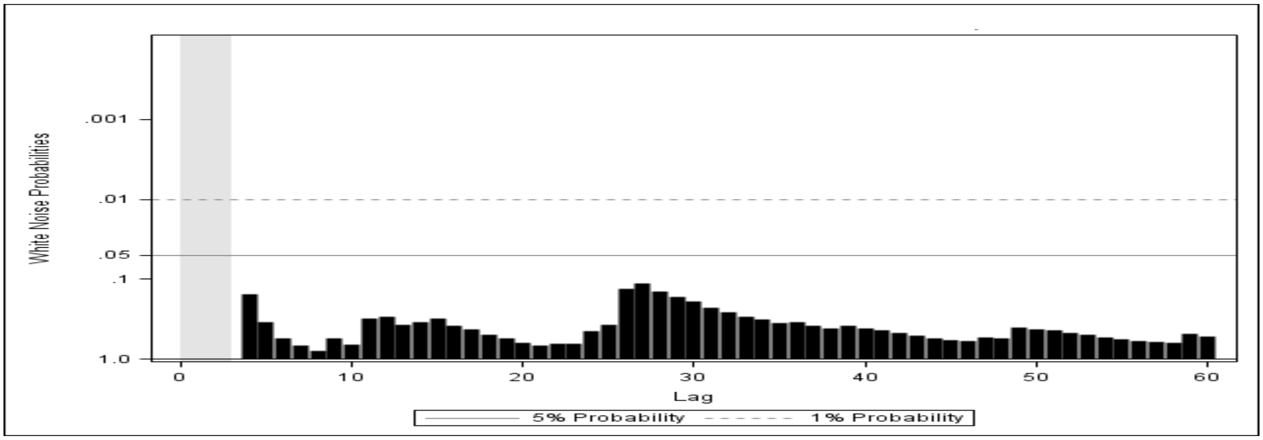
Figure 5: (a): Autocorrelation (ACF) **(b):** Partial Autocorrelation (PACF) **(c):** Normality distributions Diagnostics plot **(d):** White noise test p-values Plot for Residuals resulted from SARIMA (0, 0, 1) × $(1, 1, 4)_{12}$ model.


1.0
0.5
0.0
-0.5
-1.0
ACF
0
10
20
30
40
50
60
Lag
Two Standard Errors

(a)


1.0
0.5
0.0
-0.5
-1.0
PACF
0
10
20
30
40
50
60
Lag
Two Standard Errors

(b)



(c)

Figure 6: (a): Autocorrelation (ACF) **(b):** Partial Autocorrelation (PACF) **(c):** White noise Test P-values Plot for Residuals resulted from SARIMA (1, 0, 0) × $(4, 1, 1)_{12}$ model.

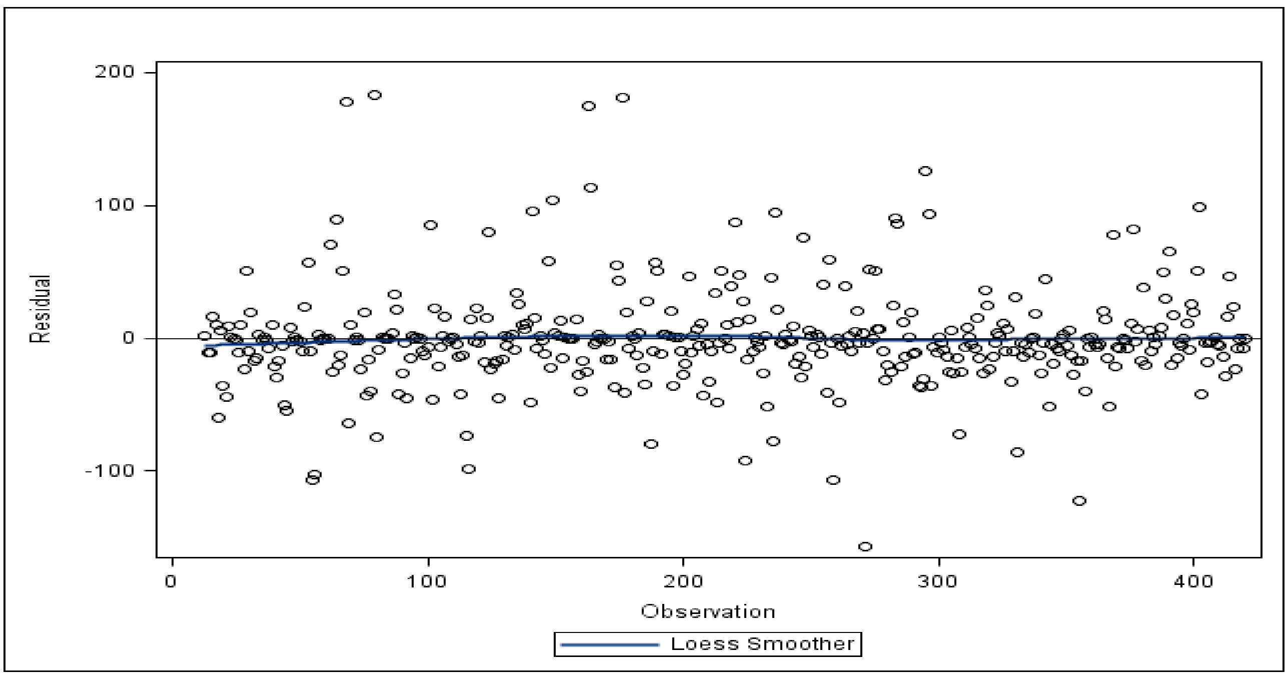


Figure 7: Scatter plot of residuals from the fitted model.

After the diagnostic, there are further tests which are necessary to select the better of the two models in relation to better forecasting accuracy. Therefore, further tests should be done based on the forecasting reliability of the two competing models that are adequately fitted.

3 Forecasting

3.1 forecasting Accuracy Assessment of the models

We have selected two models after diagnostic checking. Now we proceed to compare their forecasting performance using the various accuracy measures. For this purpose we did not use observations from (Oct. 2008 to Dec. 2009) of monthly rainfall data for calculation of forecasting errors using following equation.

$$\hat{a}_t = (x_t) - (\hat{x}_t) \tag{7}$$

Table 5: Results of Accuracy for the two models

Model(SARIMA)	MAE	MAPE	ME	MSE	THIEL'S
(1, 0,0) × (4, 1, 1)	12.43	315.7	-0.07	331.21	0.17
(0, 0, 1) × 1, 1, 4)	12.49	309.53	0.05	323.88	0.13

To measure the forecasting ability of the two models, we have estimated within-sample and out-of-sample forecasts. If the magnitude of the difference between the forecasted and actual values is low then the model has good forecasting performances. In this case, the seasonal ARIMA (0, 0, 1) × $(1, 1, 4)_{12}$ model has shown better results which is evident from the Table 5 except for the MAE values. The values of Thiele's U-Statistic are 0.17 and 0.13, respectively, for SARIMA (1, 0, 0) × (4, 1, 1) and (0, 0, 1) × $(1, 1, 4)_{12}$ models. Both results indicate that the two models are reasonably better than the naïve forecasting model. However, since the value of the Thiele's U-Statistics is 0.13 for the SARIMA (0, 0, 1) × (1, 1, 4) which is less than the value 0.17 of SARIMA (1, 0, 0) × $(4, 1, 1)_{12}$ model which indicates that the SARIMA (0, 0, 1) × $(1, 1, 4)_{12}$ model perform better in forecasting accuracy than the SARIMA (1, 0, 0) × $(4, 1, 1)_{12}$.

It can be concluded that the forecasting ability of the SARIMA (0, 0, 1) × $(1, 1, 4)_{12}$ model is better for the purpose future monthly rainfall data forecasting. Graphical analysis also exhibits closeness of the forecasted values with the holding out data.

Figure 8(a and b) represents the forecasts for the validation period and future forecasts of monthly rainfall data using SARIMA (0, 0, 1) × $(1, 1, 4)_{12}$ model. It is noteworthy that the forecasts in the validation period are reasonably close to the actual series and captured the turning points patterns as well.

We are giving below the month-wise forecast and its interval of monthly rainfall series at Mekele station in Tigray region by using the selected model Table (6).

3.2 Forecasting Monthly Rainfall values

Now the final model for forecasting of historical monthly rainfall series of Mekele station is as given below. The SARIMA model (0, 0, 1) × $(1, 1, 4)_{12}$ can be written as:

$$(1-\Phi_{12}B^{12})(1-B^{12})x_t=(1-\theta_1 B)(1-\Theta_{48}B^{48})a_t \quad (8)$$

This equation (8) can also be written as given below..

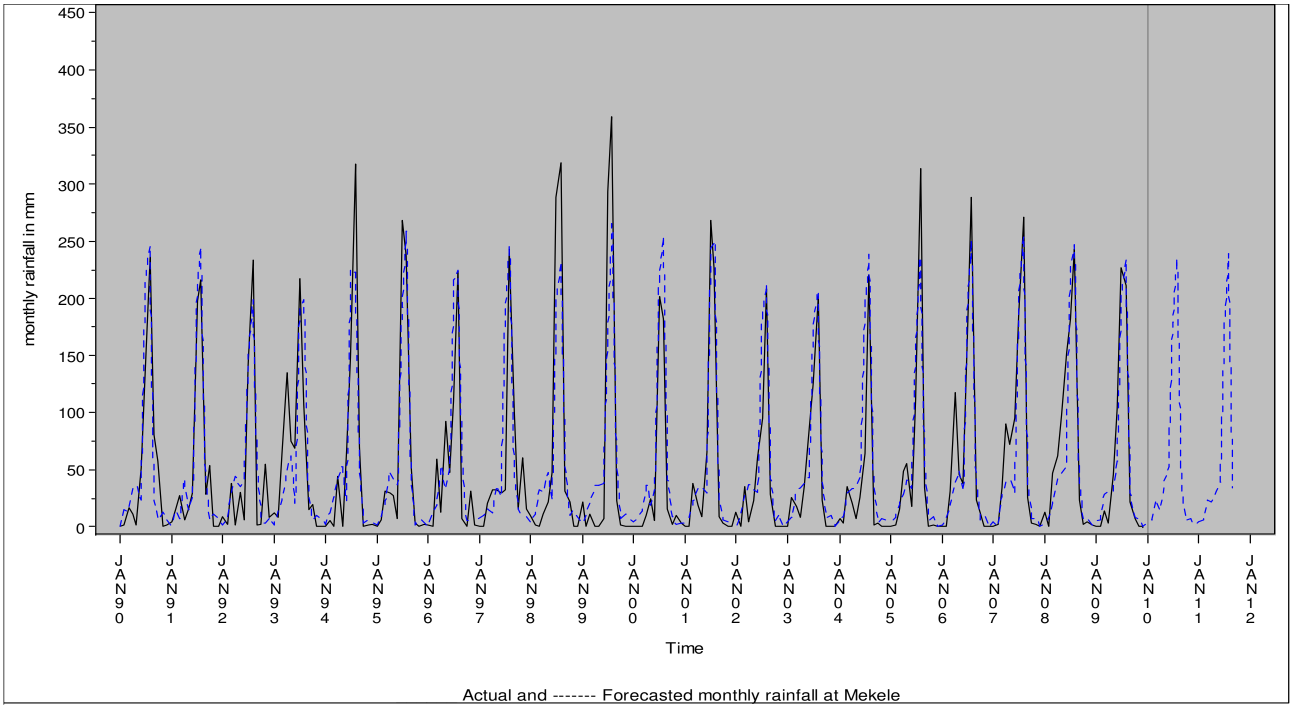
$$x_t = x_{t-12}+\Phi_{12}(x_{t-12}-x_{t-24})+a_t-\Theta_{48}a_{t-48}-\theta_1 a_{t-1}+\theta_1\Theta_{48}a_{t-49} \quad (9)$$

After substituting the estimated parameter values to Eq. (8) above, we obtain the following difference equation which can be used for forecasting purpose.

$$x_t = x_{t-12} -0.24\ x_{t-12}-x_{t-24}\ )+a_t -0.83a_{t-48}+0.19a_{t-1}-0.18a_{t-49} \quad (10)$$

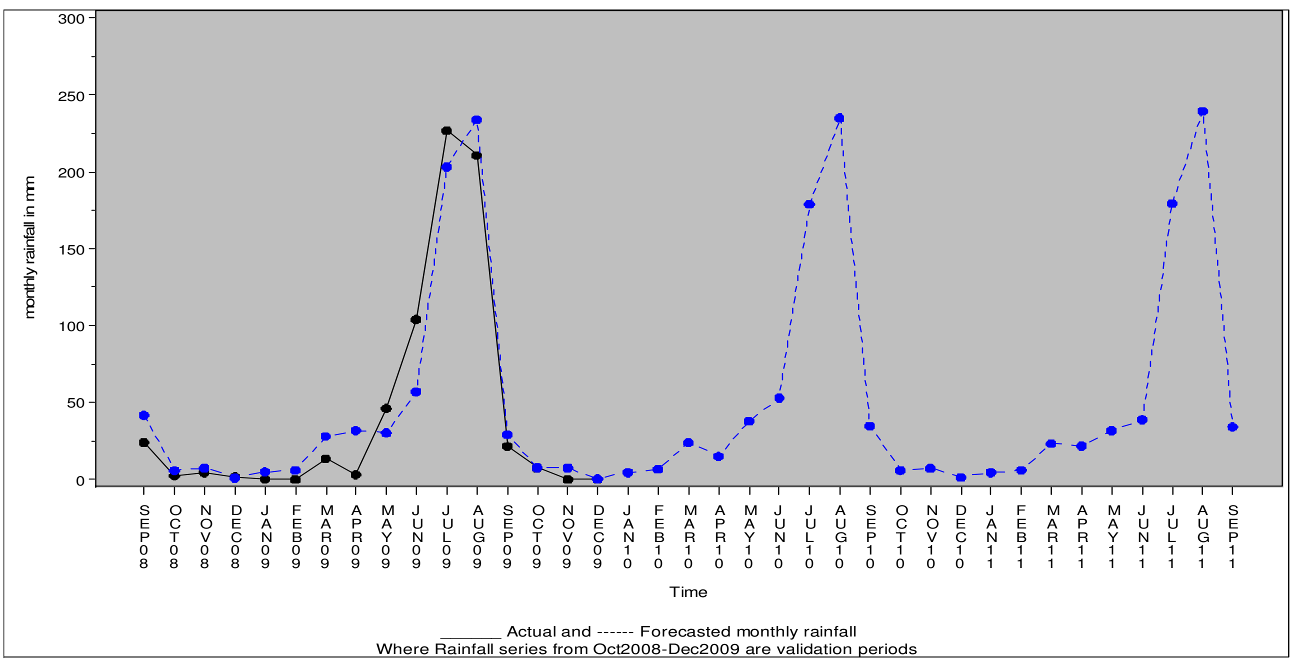
Table 6: Forecast of the Rainfall series from the period January 2010-September 2011.

Months	Forecasts	(95 % Lower Limit)	(95 % Upper Limit)
Jan 2010	4.44	-0.01	13.34
Feb 1200	6.63	-0.77	21.03
Mar 2010	0.17	-3.84	5.94
Apr 2010	14.91	6.49	27.51
May 2010	37.87	21.53	49.03
Jun 2010	52.93	22.45	97.10
Jul 2010	178.85	103.45	254.23
Aug 2010	235.03	159.63	310.42
Sep 2010	21.71	10.90	28.73
Oct 2010	5.84	-0.56	9.25
Nov 2010	7.31	0.09	11.00
Dec 2010	1.33	0.01	7.85
Jan 2011	2.45	0.51	17.72
Feb 2011	6.06	3.89	13.76
Mar 2011	23.19	14.94	29.20
Apr 2011	21.63	12.05	37.74
May 2011	31.89	9.89	53.26
Jun 2011	38.63	23.94	67.98
Jul 2011	189.46	157.82	255.03
Aug2011	240.36	203.87	314.96
Sep2011	23.82	19.41	51.07
Mean	54.50		
Standard Deviation(S.D)	79.91		



Actual and ------- Forecasted monthly rainfall at Mekele

(a)



_______ Actual and ------ Forecasted monthly rainfall
Where Rainfall series from Oct2008-Dec2009 are validation periods

(b)

Figure 8: (a) Plot of the model estimation from (Jan1975- Sep2008) periods and (Jan2010-Sep2011); **(b)** Plot of the model validation periods (Oct2008-Dec2009) and forecasted monthly rainfall series for the periods from (Jan2010-Sep2011).

4 Conclusion

In this paper the use of Univariate Box- Jenkin methodology has been shown in historical rainfall data. The estimation and diagnostic analysis results revealed that models' are adequately fitted to the historical data. In particular, the residual analysis, which is important for diagnostic checking confirmed that there is no violation of assumptions in relation to model adequacy. Further comparison based on the forecasting accuracy of the models is performed with the hold-out some rainfall values. The point forecast results showed a very closer match with the pattern of the actual data and better forecasting accuracy in validation period.

Acknowledgement

We wish to thank the editor and referee both for their comments and suggestions.

References

Box,G. E. P.and Jenkins, G.M., 1976. *Time Series Analysis: Forecasting and Control*; Holden day

Brockwell, P.J. and Davis, R. A., 1996. *Introduction to Time Series and Forecasting*, 2nd. ed. , Springer, New York.

Caldwell J.G., 2006. The Box-Jenkins Forecasting Technique Posted at Internet websites http://www.foundationwebsite.org.

Chatfield, C. 1996**:** *The Analysis of Time Series*, 5th ed., Chapman & Hall, London.

Montgomery, D.C. and Johnson, L.A, 1976. *Forecasting and Time Series Analysis*. New York, McGraw – Hall.

Harvey, A.1993**.** *Time Series Models***;** Harvester Wheatear, London.

Hipel, K. W., McLeod ,A.I and Lennox W.C., 1977**.** Advances in box and Jenkins modeling; *Water Resources Research*, Vol.**13,** No**. 3,** 567-572 pp.

Kavas, M., and J. Dulleur**,** 1975 **.** The Stochastic and Chronological Structure of Rainfall sequence – Application to India; *Water resources Research Center No*. **57,** Perdue University

Makridakis, S., Wheelwright, S.C. and Hyndman, R.J., 1998. *Forecasting methods and Application*; New York: John Wiley & Sons.

Mills, T.1999. The Economic Modeling of Financial Time series; Cambridge University Press, Cambridge.

Naill, P.E and Momani M., 2009**.** Time Series Analysis Model for Rainfall Data in Jordan: A Case Study for Using Time Series Analysis; *American Journal of Environmental ,* Vol.***5****,* 599-600 pp**;** Jeddah, Kingdom of Saudi Arabia.

Patricia, E. G., 1994: Introduction to Modeling and Forecasting in Business and Economics; McGraw-Hill,Inc.,New York.

Pankratz , A., 1983. *Forecasting with Univariate Box-Jenkins: Concepts and Cases*; John Wiley & Sons, Inc. New York.

APPENDIX(1) . Monthly rainfall data at Mekele station

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
1975	1.4	20.8	18.0	21.5	10.6	100.0	185.1	293.1	120.7	0.0	0.0	0.0	771.2
1976	3.4	6.0	0.0	41.7	28.1	17.5	180.1	243.2	49.6	2.1	3.2	0.0	570.9
1977	0.0	0.0	19.3	5.4	77.0	56.5	204.6	251.2	62.9	1.1	0.0	0.0	678.0
1978	0.0	0.0	22.3	0.0	0.6	32.6	180.8	203.6	5.6	0.0	0.0	0.5	445.8
1979	0.0	0.0	2.5	40.9	103.8	38.8	63.3	98.4	28.5	0.0	0.0	0.0	376.2
1980	0.0	83.1	0.0	112.3	43.5	42.1	206.2	413.2	17.7	0.0	0.0	0.0	918.1
1981	0.0	0.0	24.0	5.6	11.3	0.0	349.0	205.3	22.9	0.0	0.0	0.0	618.1
1982	0.0	24.4	45.2	70.8	9.3	8.0	192.6	207.1	31.2	0.0	0.0	0.0	588.6
1983	0.0	11.0	4.6	27.5	106.3	1.8	244.3	255.1	35.0	1.6	17.4	0.0	704.6
1984	0.0	0.0	14.2	2.2	0.0	9.7	117.6	78.9	44.7	0.0	25.9	0.0	293.2
1985	1.8	1.0	24.7	126.8	37.4	14.6	129.4	180.8	20.6	0.0	0.0	0.0	537.1
1986	3.5	2.7	42.0	68.1	61.2	42.2	199.2	176.0	130.3	32.0	0.0	0.0	757.2
1987	0.0	2.0	79.6	37.2	126.7	56.6	177.2	220.2	36.3	1.9	0.0	0.0	737.7
1988	0.0	29.3	0.0	10.1	37.6	6.7	380.3	394.9	59.0	0.0	0.0	0.0	917.9
1989	0.0	0.0	14.9	2.9	9.8	81.7	273.6	430.7	40.5	16.5	0.0	0.0	870.6
1990	0.0	1.5	17.3	10.8	0.9	51.7	132.3	236.9	80.8	5.7	0.0	1.2	539.1
1991	4.4	15.7	27.7	5.9	15.8	29.0	197.5	216.3	28.2	53.1	0.0	0.0	593.6
1992	8.7	2.1	38.3	1.0	30.7	6.2	140.7	233.1	1.3	2.1	54.4	8.3	526.9
1993	11.7	7.7	63.9	135.0	74.7	69.0	217.2	106.5	15.2	20.0	0.0	0.0	720.9
1994	0.0	5.3	0.4	43.8	0.8	67.6	147.9	317.8	70.1	0.0	1.8	2.0	859.5
1995	0.0	5.9	31.2	29.2	27.1	6.8	268.2	237.7	51.4	3.0	0.0	2.7	663.2
1996	1.4	0.0	59.5	12.5	92.2	47.9	109.2	224.0	7.1	0.0	31.4	1.1	586.3
1997	0.0	0.0	20.4	32.4	32.6	29.8	32.4	243.1	100.5	16.3	59.9	15.7	583.1
1998	10.0	1.2	0.0	10.6	22.0	48.0	289.0	318.8	31.8	22.0	0.0	0.0	753.4
1999	22.0	0.3	10.9	0.0	0.0	7.4	293.6	359.2	22.8	0.9	0.0	0.0	717.1
2000	0.0	0.0	0.0	10.4	24.6	5.4	201.4	182.0	15.8	2.2	10.3	3.5	455.4
2001	0.0	0.0	38.1	18.7	8.7	65.5	267.9	226.3	9.2	2.9	0.0	0.0	635.3
2002	12.9	0.0	35.5	4.2	23.0	60.8	95.5	208.6	28.0	0.0	0.0	0.3	443.6
2003	0.0	25.9	18.2	8.4	35.2	87.5	125.6	201.8	23.4	0.7	0.0	0.1	526.8
2004	7.4	3.7	35.2	20.5	7.1	25.4	64.3	221.1	1.4	3.1	0.8	0.2	390.2
2005	0.0	1.4	15.6	48.9	55.3	18.2	110.5	314.0	34.3	0.0	1.3	0.0	599.5
2006	0.0	0.0	31.3	117.6	46.3	38.1	187.1	298.9	23.6	12.0	0.0	0.7	755.6
2007	0.1	2.3	34.5	90.1	71.6	95.0	184.6	271.2	25.8	3.7	2.5	0.0	781.4
2008	13.2	0.0	47.0	62.6	97.0	151.0	182.0	243.5	28.0	2.4	4.5	1.6	832.8
2009	0.2	0.0	13.5	3.1	46.2	104.0	296.7	226.8	78.6	17.5	0.0	0.0	783.5

Source: National meteorological Agency, Addis Ababa, Ethiopia.

Author for correspondence

M.K.Sharma
Addis Ababa University, Addis Ababa, Ethiopia
E-mail: mk_subash@yahoo.co.in