Modeling and Forecasting Rainfall in Ethiopia

Tesfahun Berhane, Nurilign Shibabaw, Gurju Awgichew and Tesfaye Kebede

Abstract—Ethiopian economy is extremely dependent on agricultural sector, which contributes 45% to the Gross Domestic Product (GDP), 85% foreign earnings and provides livelihood to 80% of the population. Ethiopian agriculture is highly dependent on natural rainfall, with irrigation agriculture accounting for less than 1% of the country's total cultivated land. Therefore, modeling and forecasting the rainfall dynamics of the country has a great importance. This paper aims at examining the rainfall dynamics and fit appropriate model for forecasting Ethiopian rainfall. In this research, we apply Box-Jenkins approach, Seasonal Autoregressive Integrated Moving Average (SARIMA) model in order to forecast monthly rainfall of Ethiopia for the period of twelve months ahead. Monthly rainfall data from 1901 to 2015 were used from world bank group (climate change portal). Appropriate SARIMA model has been identified based on an Akaike information criteria (AIC) and Bayesian information criteria (BIC) for forecasting the amount of monthly average rainfall. Farmers, in general agricultural sectors, policy makers, tourists, and investors engaged in the construction industry are some of the sectors benefited from this result.

Index Terms—Ethiopia, SARIMA, seasonal differencing, rainfall, stationarity.

I. INTRODUCTION

AGRICULTURE plays the major role in the economy of Ethiopia. The agricultural sector contributes 45% to the Gross Domestic Product (GDP), 85% foreign earnings and provides livelihood to 80% of the population. Ethiopian agriculture is highly dependent on natural rainfall, with irrigation agriculture accounting for less than 1% of the countrys total cultivated land. Thus, the amount and temporal distribution of rainfall during the growing season are critical to crop yields and can induce food shortages and famine [1]. Studies in Ethiopia have shown that rainfall variability, unreliable occurrences insufficient amount and delay in onset dates contribute to decline in crop yields with reasonable amount in almost all parts of the country [2]. Rainfall variability has historically been found as a major cause of food insecurity and famine in the country [3]. This is clearly due to the fact that the agricultural sector is facing increased and continued risks of climate change. It is apparent that crop yield primarily depends on rainfall conditions of the country. Close linkage between climate and Ethiopian economy is demonstrated by close pattern of rainfall variability and gross domestic product (GDP) growth [4]. The trends in the contribution of agriculture to the country's total GDP clearly explain the presence of strong relationship between the performance of agriculture and rainfall conditions. Annual as well as seasonal crop yield variations in Ethiopia can be partly explained by rainfall

patterns. Rainfall variability usually result in reduction of 20% production and 25% raise in poverty rates in Ethiopia [5], [6]. This rainfall variability has a great impact on the income of every house holds rely on agriculture. Therefore, modeling and forecasting rainfall pattern is very important for the country in order to reduce these risks. In this paper, we present parsimonious model for short-term forecasting of rainfall dynamics of the country.

The paper is structured as follows. Section 2 presents methodology. Section 3 describes model formulation. Section 4 discusses result and discussion and finally Section 5 discusses conclusion.

II. METHODOLOGY

A. Study Area

Ethiopia is located in the north eastern part of Africa between the equator and tropic of cancer in the horn of Africa and covers an area of about 1.1×10^6 km^2 land.
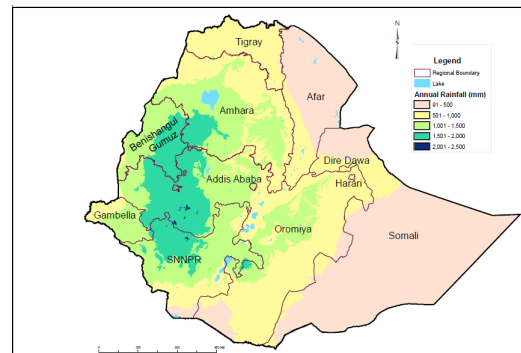

Fig. 1. Map of Ethiopia

B. Box-Jenkins algorithm

In this study, we follow the Univariate Box-Jenkins Autoregressive Integrated Moving Average (UBJ-ARIMA) algorithm which is appropriate for a stationary data series. The same approach has been used by different researchers to describe monthly rainfall in different parts of the world for instance [7], [8], [9]. The algorithm was first introduced by Box and Jenkins (1976) and now it becomes the most popular models for forecasting univariate time series data. A stationary series has a mean, variance, and autocorrelation coefficients that are essentially constant through time. Often, a nonstationary series can be made stationary with appropriate transformations. The most common type of nonstationarity occurs when the mean of a realization changes over time. A nonstationary series of this type can frequently be rendered to stationary by differencing. Our goal is to find a statistically adequate and parsimonious model that represents the observed rainfall time series data.

III. MODEL FORMULATION

A time series is said to be seasonal of order *s* if there exists a tendency for the series to exhibit periodic behavior after every time interval *s*. Seasonality is one common source of non stationary series. Seasonal non stationary series is eliminated by applying seasonal differencing. Seasonal differencing means differencing by the order of seasonal periodicity. That means, replacing the value of the series at each point in time *t* with the difference between the value at time *t* and the value of the series at time $t - s$. One can transform non-stationary series into stationary by taking regular differences, that is, the difference from one period with respect to the next. For clarification, suppose X_t is a seasonal non-stationary series, then the first-order regular difference is given as $\nabla X_t = X_t - X_{t-1}$ while, the first-order seasonal difference with seasonal period *s* is defined as $\nabla_s X_t = X_t - X_{t-s}, t > s$. Jointly, these differencing techniques can be written as equation (1).

$$w_t = \nabla_s^D \nabla^d X_t \qquad (1)$$

where $\nabla_s = (1 - B^s)^D$, $\nabla^d = (1 - B)^d$, *B* is the regular lag operator (back-shift operator) defined as $BX_t = X_{t-1}$, B^s is the seasonal lag operator defined as $B^s X_t = X_{t-s}$, *D* is the number of seasonal differences and *d* is the number of regular differences. Seasonal ARIMA model (SARIMA) can be generated by incorporating seasonal components in the ARIMA model, that means the ARMA model for stationary series incorporating both the regular dependence, that is associated with the measurement intervals of the series, as well as the seasonal dependence, which is associated with observations separated by *s* periods. Modeling the regular and seasonal dependence separately and then incorporating both models multiplicatively, a multiplicative seasonal ARIMA model is obtained and it has the form:

$$\Phi_P(B^s)\Psi_p(B)w_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t \qquad (2)$$

Substituting equation (1) in (2), we get

$$\Phi_P(B^s)\Psi_p(B)(1 - B^s)^D(1 - B)^d X_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t \qquad (3)$$

where $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ are seasonal autoregressive and moving average operators respectively and defined as $\Phi_P(B^s) = 1 - \Sigma_{k=1}^P \phi_k B^{ks}$, $\Theta_Q(B^s) = 1 - \Sigma_{k=1}^Q \Theta_k B^{ks}$, $\Psi_p(B)$ and $\theta_q(B)$ are regular autoregressive and moving average operators respectively and defined as $\Psi_p(B) = 1 - \Sigma_{k=1}^p \psi_k B^k$, $\theta_q(B) = 1 - \Sigma_{k=1}^q \theta_k B^k$. Equation (3) is a seasonal ARIMA model and usually it is written in the form $ARIMA(p,d,q) \times (P,D,Q)_s$ where lowercase letters respectively representing the regular autoregressive, integration, and moving average orders of the model. uppercase letters respectively represent the seasonal components of the model, ε_t is the white noise , *B* is the regular lag operator and defined as $BX_t = X_{t-1}$, B^s is the seasonal lag operator and defined as $B^s X_t = X_{t-s}$, *s* is the seasonal period, *d* is the order of regular differencing and *D* is the order of seasonal differencing.

Figures 2 and 3 show the plot of Ethiopia's monthly rainfall data in the period 1901 to 2015.

In this section, we analyze the identification, estimation, and diagnostic checking using the rainfall data in Fig. 2 and
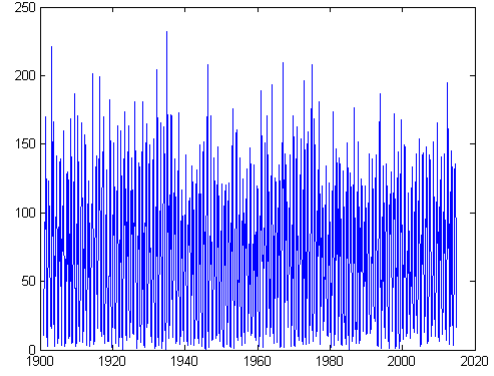

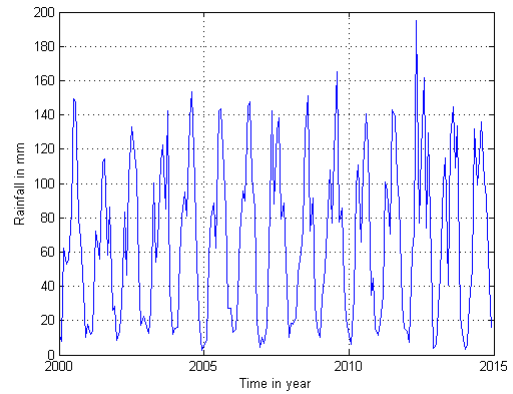Fig. 2. Time series plot from 1901 to 2015.

Fig. 3. Time series plot from 2000 to 2015 for visibility

3. We discuss also the stationarity and invertibility conditions for seasonal models and present the forecast profile for our rainfall model.

A. Identification

Here, we compare the estimated autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) with various theoretical ACFs and PACFs to find a match. We choose, the ARIMA process whose theoretical ACF and PACF best match the estimated ACF and PACF. In choosing the model, we keep in mind the principle of parsimony: we want a model that fits the given realization with the smallest number of estimated parameters. The sample ACFs and PACFs are the primary tools for model identification. Figure 4 shows the sample ACFs and PACFs of observed monthly rainfall data of Ethiopia.

The sample autocorrelation function (ACF) plots given in Fig. 4(a) indicates a monthly seasonality, $s = 12$, as ACF values at lags $12, 24, 36$ are significant and does not show any significant decreasing. The sample ACF for the rainfall time series data appears in Fig. 4(a) shows that the autocorrelations at the seasonal lags $(12, 24, 36)$ fails to die out quickly. This confirms the nonstationary character of the seasonal pattern and calls for seasonal differencing. In addition to this, the autocorrelations at the seasonal lags in Fig. 4(a) are surrounded
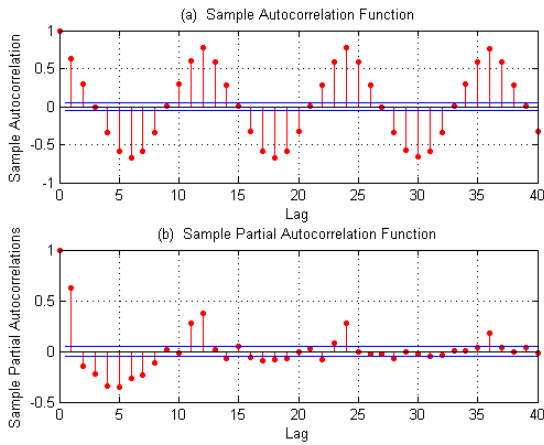
Fig. 4. Sample ACF and PACF for rainfall data.

Fig. 6. ACF and PACF plot of of seasonal differenced series w_t.

by other large autocorrelations (especially lags 10, 11, 13, 23, and 25) and also there are other large autocorrelations at half seasonal lags and around it. This is an indicator of the presence of strong seasonal pattern.

In this series, seasonal differencing is sufficient to remove all these large surrounding values as shown by the estimated ACF for the seasonally differenced series that is, $w_t = (1 - B^{12})X_t$ as indicated in Fig. 6(a). Figure 5 shows the graph of the seasonally differenced data $w_t = (1 - B^{12})X_t$.
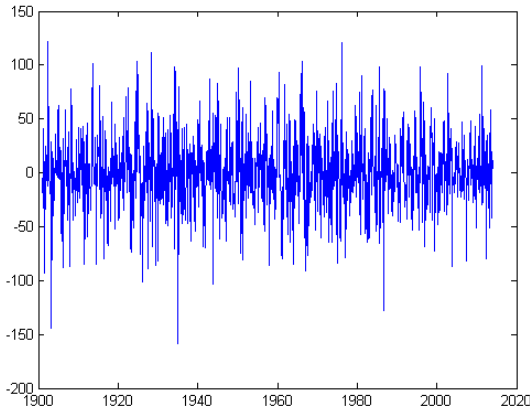

Fig. 5. Time series plot of $w_t = (1 - B^{12})X_t$.

The estimated ACF in Fig. 6(a) of the seasonally differenced data shows that differencing clears up the waves of significant values surrounding the half-seasonal lags. Looking at the estimated ACF and PACF in Fig. 6, seasonal differencing has created a stationary series since the estimated ACF falls quickly to zero at both the short lags (1, 2, 3) and the seasonal lags (24 and 36).

B. Determining the Orders of the Model

The correct order of ARIMA model is specified by determining the appropriate order of the autoregressive (AR), moving average (MA) and the integrated parts or the order
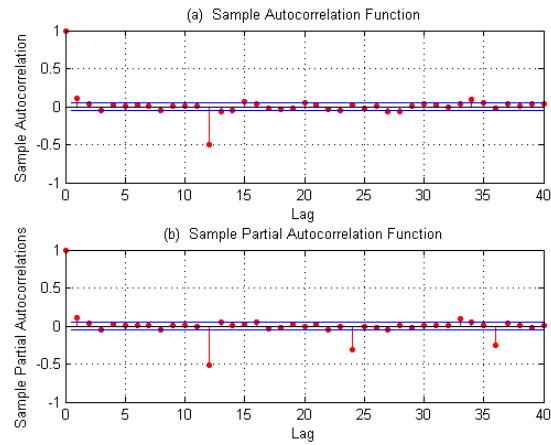
of differencing. The major tools in the identification process are the sample autocorrelation function (ACF) and partial autocorrelation function (PACF). In figure 6(a) the sample ACF with a significant value at lag 12 followed by a cut off to very small values in the rest of the lags and the sample PCAF in figure 6(b) with exponential decaying pattern at lags 12, 24 and 36 along with Table 1 confirms that an MA term is appropriate at the seasonal lag 12 and we expect Θ_{12}, to be negative. These analysis all together leads us to choose a seasonal $ARIMA(0,0,0) \times (0,1,1)_{12}$ model.

TABLE I
PRIMARY DISTINGUISHING CHARACTERISTICS OF THEORETICAL ACFs
AND PACFs FOR STATIONARY PROCESSES.

Model	ACF	PACF
AR(p)	Tails off towards zero (Decays exponentially or damped oscillation	Cuts off to zero (after p lags)
MA(q)	Cuts off to zero (after q lags)	Tails off towards zero (Decays exponentially or damped oscillation)
ARMA(p,q)	Tails off towards zero (Decays exponentially after lag q)	Tails off towards zero (Decays exponentially after lag p)

In addition to the above model selection criteria, we also apply the Akakie information criteria (AIC) and Bayesian information criteria (BIC) in order to choose best model among different candidates, accordingly, we identify the seasonal $ARIMA(1,0,0) \times (0,1,1)_{12}$ model is the best model to represent our monthly rainfall data with smallest AIC= 12565 and BIC=12705. The model can be written as

$$(1 - \phi_1 B)(1 - B^{12})X_t = (1 - \Theta_{12}B^{12})\varepsilon_t \tag{4}$$

C. Parameter Estimation

Table 2 gives the maximum likelihood estimates and the standard errors of the seasonal ARIMA $(1,0,0) \times (0,1,1)_{12}$ model parameters.

D. Model Adequacy

At the diagnostic-checking stage we examine the residuals of the estimated model to see if they are independent. If they

TABLE II
PARAMETER ESTIMATION FOR ARIMA $(1,0,0) \times (0,1,1)_{12}$ MODEL.

parameter	value	standard error	t value
ϕ_1	0.11	0.024	4.40
Θ_{12}	-0.90	0.010	-85.71

are not, we return to the identification stage to tentatively select another model. A statistically adequate model satisfies the assumption that the random shocks are independent. If the residuals are independent, we accept the hypothesis that the shocks are independent. The residual ACF is used to test the hypothesis that the shocks are independent. The residual ACFs is shown in Fig. 7. Ljung and Box test suggests a test statistic based on all the residual autocorrelations as a set. For a given K residual autocorrelations, we test the following joint null hypothesis about the correlations among the random shocks.

$$H_0 : \rho_1(\varepsilon) = \rho_2(\varepsilon) = \rho_3(\varepsilon) = ... = \rho_k(\varepsilon) = 0 \qquad (5)$$

where, $\rho_i(\varepsilon)$ are the theoretical autocorrelations. Using the following test statistic given in[10] that is:

$$Q = n(n+2)\Sigma_{k=1}^{K}(n-k)^{-1}\gamma_k^2(\varepsilon) \qquad (6)$$

where, n is the number of observations used to estimate the model parameters, Q follows approximately a chi-squared distribution with $(K-m)$ degrees of freedom, γ_k is the correlation of the residuals and m is the number of parameters involved in the model. We calculate Q for $K = 40$ (sample autocorrelation of the residual), and obtain $Q = 0.92304$. According to the chi-squared distribution table, the critical value corresponding to the degree of freedom ($DF = 38$) at 10% level is 40, which is much greater than our calculated chi-squared value. Therefore, we conclude that the residual autocorrelations in Fig. 6 are not significantly different form zero as a set. Hence, we accept the hypothesis (5), that is the random shocks are independent.
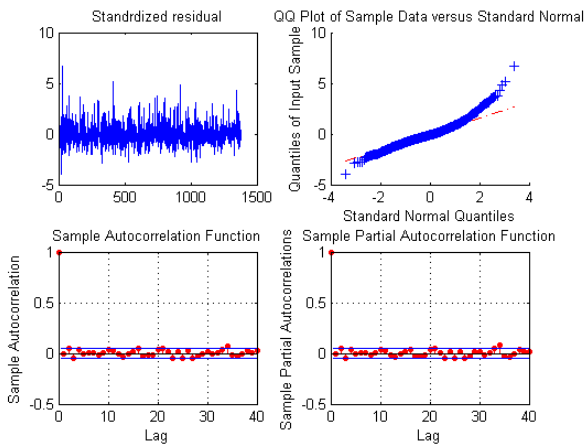

Fig. 7. Standardized residual, QQplot, sample ACF and PACF of the residual.

Furthermore, the mean of the standard residual is nearly zero that is 0.00032 and the variance of the standardized residual is 0.996 from these and the plots of the residual we can conclude that the residuals are independent, that means, the residuals are reasonably normally distributed. Moreover,

the stationarity and invertibility conditions are satisfied that is, $|\phi_1| = |0.11| = 0.11 < 1$ and the invertibility condition $|\Theta_{12}| = |-0.90| = 0.90 < 1$ both conditions are satisfied. Therefore, our model is adequate.

IV. RESULTS AND DISCUSSIONS

On the basis of the developed model, the forecasted monthly rainfall along with the 95% confidence intervals for the year 2015 is presented in Fig. 8. Table 3 reveals the forecasted values using our SARIMA $(1,0,0) \times (0,1,1)_{12}$ model. All forecasted values lies within 95% confidence interval which shows that the model is neither over forecasting nor under forecasting. Negative values in the lower confidence interval is treated as zero or no rain.

TABLE III
FORECASTED VALUES USING SARIMA $(1,0,0) \times (0,1,1)_{12}$

| Months	Rainfall in mm	Lower confidence level	Upper confidence level
January 2015	11.82	0	50.14
February 2015	14.83	0	53.36
March 2015	39.96	0	77.56
April 2015	81.27	3.00	97.44
May 2015	104.29	84.48	178.91
June 2015	71.93	51.83	146.27
July 2015	127.75	68.52	162.96
August 2015	143.47	88.54	182.97
September 2015	93.08	54.17	148.60
October 2015	88.04	44.46	138.89
November 2015	33.85	0	89.54
December 2015	13.24	0	63.54
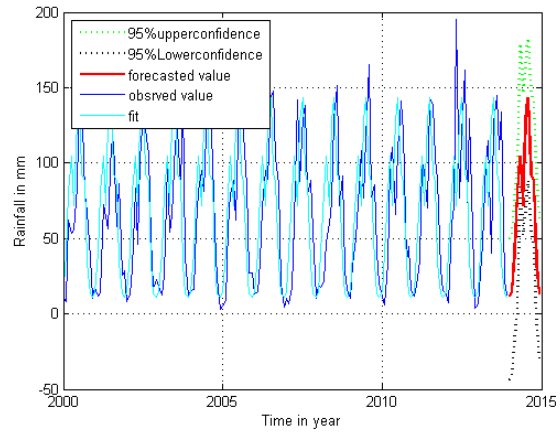
Fig. 8. Graph of actual vs forecasted values of the Rainfall data

Figure 9 shows that Ethiopia experience much rainfall in the months of July and August, while least amount of rainfall in the months of January, February and December.

V. CONCLUSIONS

In this paper, *SARIMA*$(1,0,0) \times (0,1,1)_{12}$ has been identified as an appropriate model for forecasting the amount of monthly rainfall of Ethiopia. Using the model, we forecast twelve months ahead. All forecasted values lies within the
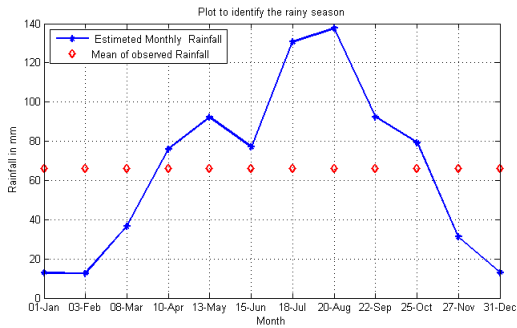
Fig. 9. Time series plot of rainy season in Ethiopia estimated using the model

95% confidence interval which shows that the model is neither over forecasting nor under forecasting.

REFERENCES

[1] S. Di Falco, M. Yesuf, G. Kohlin, and C. Ringler, "Estimating the impact of climate change on agriculture in low-income countries: Household level evidence from the nile basin, ethiopia," *Environmental and Resource Economics*, vol. 52, no. 4, pp. 457–478, 2012.

[2] G. Makombe, D. Kelemework, and D. Aredo, "A comparative analysis of rainfed and irrigated agricultural production in ethiopia," *Irrigation and Drainage Systems*, vol. 21, no. 1, pp. 35–44, 2007.

[3] W. Bewket, "Rainfall variability and crop production in ethiopia: Case study in the amhara region," in *Proceedings of the 16th International Conference of Ethiopian Studies*, vol. 3. Norwegian University of Science and Technology Trondheim, Norway, 2009, pp. 823–836.

[4] C. Sadoff *et al.*, "Managing water resources to maximize sustainable growth: a world bank water resources assistance strategy for ethiopia," The World Bank, Tech. Rep., 2008.

[5] F. Hagos, G. Makombe, R. E. Namara, and S. B. Awulachew, *Importance of irrigated agriculture to the Ethiopian economy: Capturing the direct net benefits of irrigation*. IWMI, 2009, vol. 128.

[6] M. Osman and P. Sauerborn, "A preliminary assessment of characteristics and long term variability of rainfall in ethiopia–basis for sustainable land use and resource management," *Challenges to organic farming and sustainable land use in the tropics and subtropics, Deutscher Tropentag*, 2002.

[7] E. Afrifa-Yamoah, B. I. Saeed, and A. Karim, "Sarima modelling and forecasting of monthly rainfall in the brong ahafo region of ghana," *World Environment*, vol. 6, no. 1, pp. 1–9, 2016.

[8] S. Suhartono, "Time series forecasting by using seasonal autoregressive integrated moving average: Subset, multiplicative or additive model," *J. Math. Stat*, vol. 7, pp. 20–27, 2011.

[9] T. Papalaskaris, T. Panagiotidis, and A. Pantrakis, "Stochastic monthly rainfall time series analysis, modeling and forecasting in kavala city, greece, north-eastern mediterranean basin," *Procedia engineering*, vol. 162, pp. 254–263, 2016.

[10] A. Pankratz, *Forecasting with univariate Box-Jenkins models: Concepts and cases*. John Wiley & Sons, 2009, vol. 224.