Rainfall Drought Simulating Using Stochastic SARIMA Models for Gadaref Region, Sudan

Moahmed Hassan, Hisham and Mahgoub Mohamed, Tariq

University of Khartoum, Khartoum Technical College

2014

Rainfall Drought Simulating Using Stochastic SARIMA Models for Gadaref Region, Sudan

Hisham Mohamed Hassan Ali[a] **Tariq Mahgoub Mohamed**[b]

[a] Department of Econometrics and Social Statistics, University of Khartoum
E-mail: hishamdr@yahoo.com

[b] Department of Civil Engineering, Sudan University of Science and Technology
E-mail: tariqcivil73@yahoo.com

A B S T R A C T

Drought forecasting plays an important role in the planning and management of natural and water resources in this paper presents linear stochastic models known as multiplicative seasonal autoregressive integrated moving average model (SARIMA) used to simulate droughts in Gadaref region, Sudan. The models are applied to simulate droughts using Standardized Precipitation Index (SPI) series, the results show that the fitted model is adequate to the SPI_6 for Gadaref station is the SARIMA (0, 0, 5) (1.0.1) model.

Key Words: SPI, Drought, Forecasting, SARIMA, Sudan.

I. INTRODUCTION

Gadaref[1] region lies in east central part of Sudan, at the border with Ethiopia. The mean annual rainfall in this region, during the last four decades, is 617 mm. The annual number of rainy days, (rainfall > 1 mm), is 111 days and the mean annual reference potential evapotranspiration (ETO) using Penman/Monteith criterion for the region is about 2283mm (Le Houérou, 2009). The region experiences very hot summer and temperature in the region reaches up to 45^{o} C in May. Generally the dry periods are accompanied with high temperatures, which lead to higher evaporation affecting natural vegetation and the agriculture of the region along with larger water resources sectors. Annual potential evapotranspiration exceeds annual precipitation in this region. The rainfall exceeds evapotranspiration only in August and September (Etuk and Mohamed, 2014).The climate in the Gedaref is semi-arid with mean annual temperature near 30o C (Elagib and Mansell, 2000).

A few researchers who have modeled rainfall drought using Standardized Precipitation Index (SPI) as a drought indicator by SARIMA methods in recent times are Mishra and Desai (2005) and Durdu (2010). For instance, Mishra and Desai (2005) fitted a SARIMA (1, 0, 0) x (1, 1, 1) $_6$ model to simulate and forecast SPI-6 in Kansabati river basin, India. Durdu (2010) modeled the SPI-6 for

[1] The word Gedarif (Gadaref, El Gadarif, Qadarif) is derived from the Arabic phrase All Gada-Ye-rif, meaning: he who had finished selling or buying should leave

Buyuk Menderes river basin located in the western part of Turkey and fitted a SARIMA(1, 0, 0)x(2, 0, 1) $_6$ to it.

Time series model development consists of three stages identification, estimation, and diagnostic checking (Box and Jenkins, 1970). The identification stage involves transforming the data (if necessary) to improve the normality and stationary of the time series to determine the general form of the model to be estimated. During the estimation stage the model parameters are calculated. Finally, diagnostic test of the model is performed to reveal possible model inadequacies to assist in the best model selection.

II. METHODOLOGY

The SPI is computed by fitting an appropriate probability density function to the frequency distribution of precipitation summed over the time scale of interest (usually 3, 6, 12, and 24 months). This is performed separately for each time scale and for each location in space.

McKee (McKee et al., 1993) developed the Standardized Precipitation Index (SPI) for the purpose of defining and monitoring drought. Among others, the Colorado Climate Center, the Western Regional Climate Center and the National Drought Mitigation Center use the SPI to monitor current states of drought in the United States. The nature of the SPI allows an analyst to determine the rarity of a drought or an anomalously wet event at a particular time scale for any location in the world that has a precipitation record. The SPI based drought classification is demonstrated in Table 1.

TABLE 1: DROUGHT CLASSIFICATION BASED ON SPI

SPI Values	Class
$\geq$ 2	Extremely wet
1.50–1.99	Very wet
1.0–1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.0 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
$\leq$ 2	Extremely dry

The SPI was designed to quantify the precipitation deficit for multiple time scales, such as, 3, 6, 12, 24, and 48months. These time scales reflect the impact of drought on the availability of the different water resources .Soil moisture conditions respond to precipitation anomalies on a relatively short scale. Groundwater, streamflow, and reservoir storage reflect the longer-term precipitation anomalies. The index makes it possible to describe drought on multiple time scales (Tsakiris and Vangelis, 2004; Mishra and Desai, 2005; Cacciamani et al., 2007; Durdu, O.F., 2010). For this study we used 6 months as a time indicator (SPI_6). (see appendix 1).

Often time series possess a seasonal component that repeats every s observations. For monthly observations s = 12 (12 in 1 year), for quarterly observations s = 4 (4 in 1 year). In order to deal with seasonality, ARIMA processes have been generalized: The full *ARIMA* model is called the *SARIMA*, a seasonal differencing element. The regular *ARIMA* includes the *AR* polynomial and the *MA* polynomial the SARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is ARIMA(p, d, q) × (P, *D, Q*)*S*, with 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.

Without differencing operations, the model could be written more formally as

$$\Phi(B^S)\varphi(B)(x_t - \mu) = \Theta(B^S)\theta(B)w_t \qquad (1)$$

The non-seasonal components are:

$$\text{AR: } \varphi(B) = 1 - \varphi_1 B - \ldots - \varphi_p B^p$$
$$\text{MA: } \theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$$

The seasonal components are:

$$\text{Seasonal AR: } \Phi(B^S) = 1 - \Phi_1 B^S - \ldots - \Phi_P B^{PS}$$
$$\text{Seasonal MA: } \Theta(B^S) = 1 + \Theta_1 B^S + \ldots + \Theta_Q B^{QS}$$

Note that on the left side of equation (1) the seasonal and non-seasonal AR components multiply each other, and on the right side of equation (1) the seasonal and non-seasonal MA components multiply each other.

ARIMA (1, 0, 0) × (1, 0, 0)$_{12}$

The model includes a non-seasonal AR(1) term, a seasonal AR(1) term, no differencing, no MA terms and the seasonal period is S = 12, the non-seasonal AR(1) polynomial is $\varphi(B) = 1 - \varphi_1 B$. the seasonal AR(1) polynomial is $\Phi(B^{12}) = 1 - \Phi_1 B^{12}$.

The model is $(1 - \Phi_1 B^{12})(1 - \varphi_1 B)(x_t - \mu) = w_t$. (2)

If we let $z_t = x_t - \mu$ (for simplicity), multiply the two AR components and push all but z_t to the right side we get

$$z_t = \varphi_1 z_{t-1} + \Phi_1 z_{t-12} + (-\Phi_1\varphi_1)z_{t-13} + w_t. \qquad (3)$$

This is an AR model with predictors at lags 1, 12, and 13.

R can be used to determine and plot the PACF for this model, with φ_1=.6 and Φ_1=.5. That PACF (partial autocorrelation function)

In this study, SARIMA models were used to simulate droughts based on the procedure of models developments. The models are applied to simulate droughts using (SPI) series in Gadaref region, (SPI-6 = SPI for 6 month).

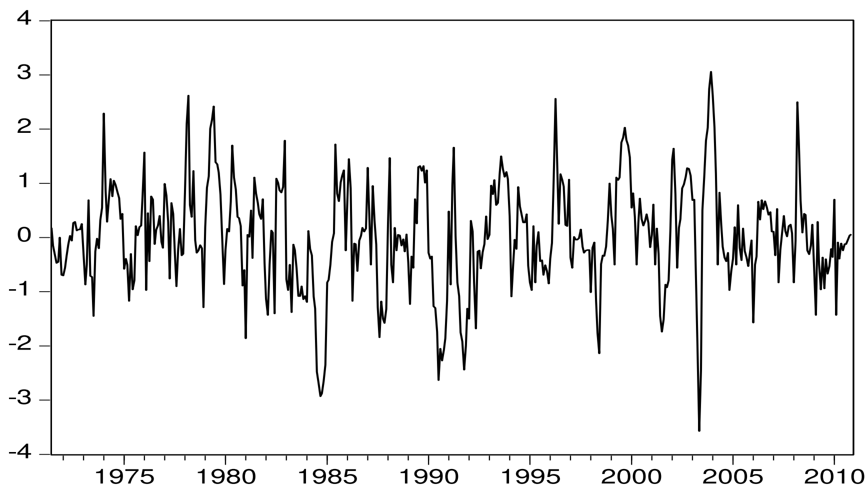
After identifying models, it is needed to obtain efficient estimates of the parameters. These parameters should satisfy two conditions namely stationary and invariability for autoregressive and moving average models, respectively. The parameters should also be tested whether they are statistically significant or not. The parameters values are associated with standard errors of estimate and related t-values.

The drought events were calculated using the SPI. The data series from 1971 to 2010 were used for model development for SPI-6 series.
There are two software packages which are used for time series analysis. These programs are the SPSS 19 package and E-views 6.

IV. RESULTS AND DISCUSSION

Time series plot was conducted using the raw data, SPI 6_Gadaref, to assess its stability. The assessments results are shown in fig. 1 it is clearly depicted that the time series are stationary.

FIG. 1: SPI 6 OF GADAREF STATION TIME SERIES 1971-2010



Stationary is also confirmed by the Augmented Dickey-Fuller Unit Root Test (ADF test) on the data. The ADF test was conducted on the entire data. Table (6.1) shows ADF test results. ADF test value -7.29548 less than critical vales -3.9778, -3.4194, -3.1323 all at 1%, 5%, and 10% respectively. This indicates that the series is stationary.

TABLE 1: ADF UNIT ROOT TEST (SPI6-GADAREF)

Station	Variable	ADF test	Level of Confidence	Critical Value	Probability	Result
			1%	-3.9778	0.0000	
Gadaref	SPI_6	-7.29548	5%	-3.4194	0.0000	*stationary*
			10%	-3.1323	0.0000	

In this step, the model that seems to represent the behaviour of the series is searched, by the means of autocorrelation function (ACF) and partial auto correlation function (PACF), for further investigation and parameter estimation. The behaviour of ACF and PACF is to see whether the series is stationary or not.

For modelling by ACF and PACF methods, examination of values relative to auto regression and moving average were made. An appropriate model for estimation of SPI_6 values for the station was finally found.

Figure 2 shows the ACF and PACF, which have been estimated for SPI-6 for Gadaref station. Many models for Gadaref stations, according to the ACF and PACF of the data, were examined to determine the best model .The model that gives the minimum Akaike Information Criterion (AIC) and Schwarz Criterion (SC) is selected as best fit model, as shown in Table 2.

FIG. 2: ACF AND PACF PLOT FOR GADAREF STATION (SPI_6) SERIES

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|***** |	.|***** |	1	0.710	0.710	240.58	0.000
.|**** |	.|. |	2	0.515	0.021	367.28	0.000
.|*** |	.|. |	3	0.381	0.016	436.84	0.000
.|** |	.|. |	4	0.264	-0.036	470.20	0.000
.|* |	*|. |	5	0.110	-0.147	476.05	0.000
*|. |	**|. |	6	-0.087	-0.235	479.71	0.000
*|. |	.|* |	7	-0.095	0.168	484.06	0.000
*|. |	.|. |	8	-0.074	0.065	486.67	0.000
*|. |	.|. |	9	-0.074	-0.001	489.31	0.000
.|. |	.|. |	10	-0.055	0.042	490.78	0.000
.|. |	.|. |	11	-0.021	-0.004	490.99	0.000
.|. |	.|. |	12	0.032	-0.028	491.49	0.000
.|. |	.|. |	13	0.071	0.066	493.93	0.000
.|. |	.|. |	14	0.046	-0.063	494.98	0.000
.|. |	.|. |	15	0.030	-0.030	495.42	0.000
.|. |	.|. |	16	0.027	0.027	495.77	0.000
.|. |	.|. |	17	0.032	0.033	496.29	0.000
.|. |	.|. |	18	0.016	-0.011	496.42	0.000
.|. |	.|. |	19	-0.014	0.002	496.51	0.000
.|. |	*|. |	20	-0.049	-0.093	497.72	0.000
*|. |	*|. |	21	-0.089	-0.088	501.65	0.000
*|. |	.|. |	22	-0.112	-0.006	507.92	0.000
*|. |	.|. |	23	-0.144	-0.035	518.29	0.000
*|. |	.|. |	24	-0.154	0.000	530.14	0.000

TABLE 2: COMPARISON OF AIC FOR SELECTED MODELS, (SPI6_GADAREF)

Variable	Station	Model	AIC
SPI_6	Gadaref	ARIMA(1,0,0)	2.104
		ARIMA(1,0,1)	2.108
		ARIMA(2,0,1)	2.113
		ARIMA(2,0,2)	2.102
		ARIMA(0,0,1)	2.335
		ARIMA(0,0,5)	2.010
		SARIMA(0,0,5) (1,0,0)	2.036
		SARIMA(0,0,5) (1,0,1)	1.999
		SARIMA(0,0,5) (0,0,1)	2.104
		SARIMA(1,0,0) (1,0,1)	2.092

The ACF and PACF correlograms, Fig.2, and the coefficient are analyzed carefully and the SARIMA model chosen is SARIMA (0,0,5) (1,0,1), as shown in table 3.

After the identification of the model using the AIC and SC criteria, estimation of parameters was conducted. The values of the parameters are shown, in Table 3. The result indicated that the parameters are all significant since their p-values is smaller than 0.05 and should be used in the model.

TABLE 3: SUMMARY OF PARAMETER ESTIMATES AND SELECTION CRITERIA (AIC), (SPI6_GADAREF)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	-0.875450	0.025327	-34.56547	0.0000
MA(1)	0.697188	0.043122	16.16768	0.0000
MA(2)	0.514251	0.050715	10.14007	0.0000
MA(3)	0.408934	0.052794	7.745840	0.0000
MA(4)	0.402012	0.050717	7.926640	0.0000
MA(5)	0.394849	0.043373	9.103536	0.0000
SMA(12)	0.951476	0.013937	68.27071	0.0000
R-squared	0.573186	Mean dependent var		0.024048
Adjusted R-squared	0.567557	S.D. dependent var		0.992582
S.E. of regression	0.652726	Akaike info criterion		1.999716
Sum squared resid	193.8532	Schwarz criterion		2.062376
Log likelihood	-454.9345	Hannan-Quinn criter.		2.024386
Durbin-Watson stat	2.000692			
Inverted AR Roots	.96+.26i	.96-.26i	.70+.70i	.70-.70i
	.26-.96i	.26+.96i	-.26-.96i	-.26+.96i
	-.70-.70i	-.70-.70i	-.96+.26i	-.96-.26i
Inverted MA Roots	.96+.26i	.96-.26i	.70-.70i	.70+.70i
	.47-.68i	.47+.68i	.26-.96i	.26+.96i
	-.26+.96i	-.26-.96i	-.41-.73i	-.41+.73i
	-.70-.70i	-.70-.70i	-.82	-.96-.26i
	-.96+.26i			

As considered in table 3 the model SARIMA (0,0,5) (1.0.1) has been selected as the one with min AIC. The model has been identified and the parameters have been estimated. The model verification is concerned with checking the residuals of the model to see if they contain any systematic pattern which still can be removed to improve the chosen ARIMA. All validation tests are carried out on the residual series. The tests are summarized briefly in the following paragraph.

For a good model, the residuals left over after fitting the model should be white noise. This is revealed through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, the various correlations up to 24 lags have been computed. The ACF and PACF of residuals of the model are shown in figure 3.

Most of the values of the RACF and RPACF lies within confidence limits except very few individual correlations appear large compared with the confidence limits. The figure indicates no significant correlation between residuals.

FIG. 3: THE ACF AND PACF OF RESIDUALS FOR SPI-6 FOR GADAREF STATION MODEL

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>.|. |</td><td>.|. |</td><td>1</td><td>-0.001</td><td>-0.001</td><td>0.0001</td><td></td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>2</td><td>-0.005</td><td>-0.005</td><td>0.0110</td><td></td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>3</td><td>0.025</td><td>0.025</td><td>0.2931</td><td></td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>4</td><td>-0.009</td><td>-0.009</td><td>0.3302</td><td></td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>5</td><td>-0.012</td><td>-0.012</td><td>0.3967</td><td></td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>6</td><td>-0.022</td><td>-0.023</td><td>0.6310</td><td></td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>7</td><td>-0.039</td><td>-0.039</td><td>1.3437</td><td></td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>8</td><td>0.018</td><td>0.018</td><td>1.4908</td><td>0.222</td></tr>
<tr><td>*|. |</td><td>*|. |</td><td>9</td><td>-0.076</td><td>-0.076</td><td>4.2551</td><td>0.119</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>10</td><td>-0.015</td><td>-0.014</td><td>4.3686</td><td>0.224</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>11</td><td>-0.022</td><td>-0.026</td><td>4.6047</td><td>0.330</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>12</td><td>-0.003</td><td>-0.001</td><td>4.6100</td><td>0.465</td></tr>
<tr><td>.|* |</td><td>.|* |</td><td>13</td><td>0.098</td><td>0.096</td><td>9.1690</td><td>0.164</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>14</td><td>0.017</td><td>0.015</td><td>9.3048</td><td>0.232</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>15</td><td>-0.046</td><td>-0.048</td><td>10.337</td><td>0.242</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>16</td><td>-0.013</td><td>-0.025</td><td>10.415</td><td>0.318</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>17</td><td>-0.005</td><td>-0.005</td><td>10.427</td><td>0.404</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>18</td><td>0.028</td><td>0.027</td><td>10.815</td><td>0.459</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>19</td><td>0.027</td><td>0.031</td><td>11.167</td><td>0.515</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>20</td><td>-0.039</td><td>-0.037</td><td>11.899</td><td>0.536</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>21</td><td>0.002</td><td>-0.006</td><td>11.900</td><td>0.614</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>22</td><td>-0.016</td><td>-0.009</td><td>12.022</td><td>0.677</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>23</td><td>-0.052</td><td>-0.044</td><td>13.317</td><td>0.649</td></tr>
<tr><td>.|. |</td><td>.|. |</td><td>24</td><td>0.044</td><td>0.045</td><td>14.266</td><td>0.648</td></tr>
</table>

The Ljung-Box Q-statistic is employed for checking independence of residual. From Fig.3, ones can observe that the p-value is greater than 0.05 for all lags, which implies that the white noise hypothesis is not rejected.

The Breusch-Godfrey Serial Correlation LM test accepts the hypothesis of no serial correlation in the residuals, as shown in Table 4. Durbin Watson statistic, (DW=1.999764), also indicated that there is no serial correlation in the residuals.

TABLE NO. 4: THE BREUSCH-GODFREY SERIAL CORRELATION LM TEST

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	**0.030213**	**Prob. F(2,453)**	**0.9702**
Obs*R-squared	**0.001942**	**Prob. Chi-Square(2)**	**0.9990**

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	**0.760940**	**Prob. F(12,443)**	**0.6909**
Obs*R-squared	**9.272109**	**Prob. Chi-Square(12)**	**0.6795**

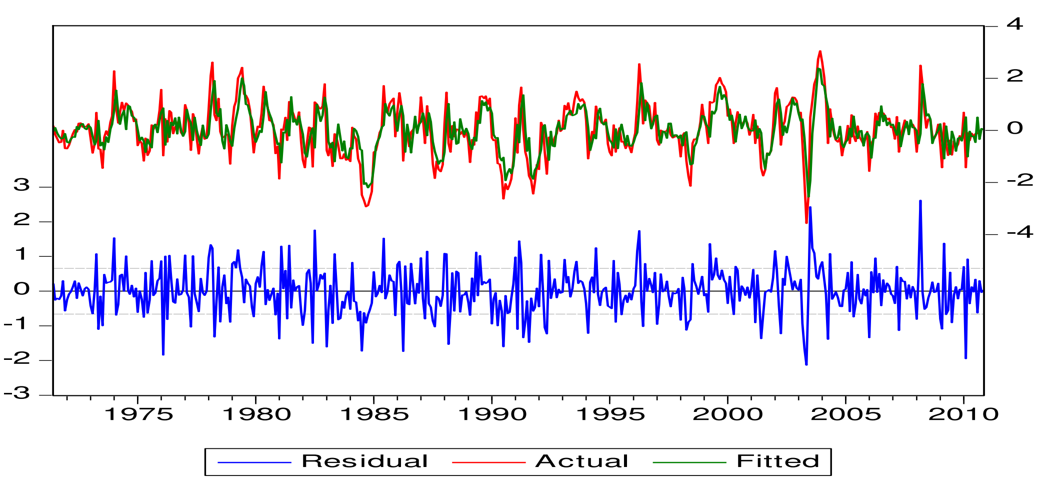
The Q-statistic and the LM test both indicated that the residuals are none correlated and the model can be used. Since the coefficients of the residual plots of ACF and PACF are lying within the confidence limits, the fit is good and the error obtained through this model is tabulated in the table (6.5). The graph showing the observed and fitted values is shown in figure (6.4).

TABLE NO. 5: ERRORS MEASURES OBTAINED FOR THE MODEL ARIMA (0,0,5)

Error Measure	Value
RMSE	0.662
MAE	0.487
R squared	0.557

Figure 4 shows a very close agreement between the fitted model and the actual data. Histogram of residuals for SPI_6 is shown in figure 5. This histogram shows that the residuals are normally distributed. This signifies residuals to be white noise.

FIG.4: ACTUAL AND FITTED VALUES SARIMA (0, 0, 5)(1.0.1) , (GADAREF, SPI_6)



The graph of the (Q-Q) plot for the residual data look fairly linear, the normality assumptions of the residuals hold, as shown in Fig. 6.

FIG. 5: HISTOGRAMS OF RESIDUALS FOR SPI_6 GADAREF STATION

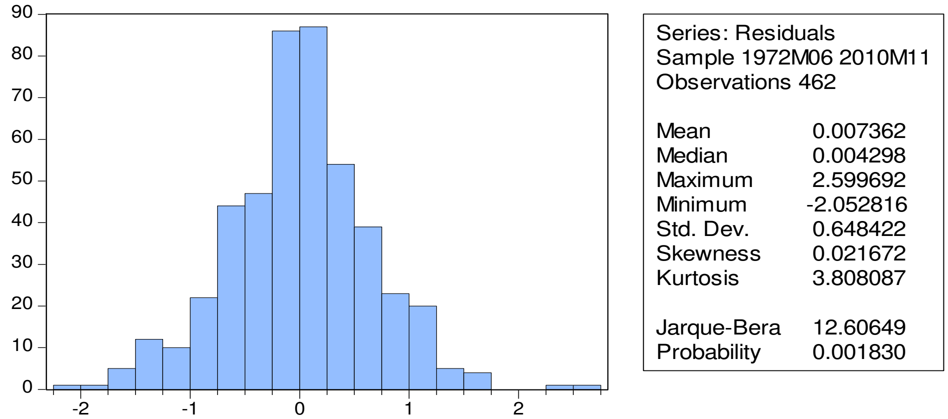
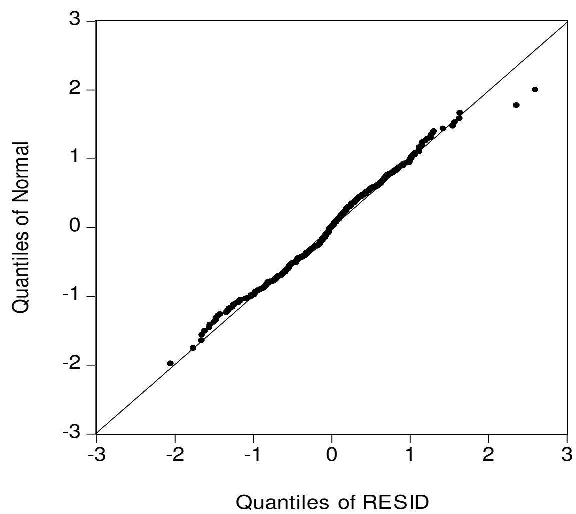


FIG. 6: (Q-Q) PLOT OF RESIDUALS FOR SPI_6 FOR GADAREF STATION



The K–S test is used to test the normality of residuals. It is observed that the D_{cal} is less than D_{tab} at 5% significant level, shown in Table 6 (α = 0.102 > 0.05).This test satisfies that the residuals are normally distributed.

TABLE NO 6: K-S TEST CALCULATION OF RESIDUALS FOR SPI_6 SERIES, (GADAREF)

Kolmogorov-Smirnov Test	
Most Extreme Differences (D_{cal})	0.057
D_{table}	0.063

One can note that all the model coefficients are statistically significant, each being more than twice its standard error. The regression is very highly significant with a p-value of 0.0000, as high as 55.7% of the variation in data is accounted for by the fitted model. Figure 3 shows that the residuals are uncorrelated. Figure 4 shows a very close agreement between the fitted model and the data. Therefore the fitted model is adequate. Fitted to the SPI_6 for Gadaref station is the SARIMA (0, 0, 5) (1.0.1) model. Using various alternative arguments it has been shown to be adequate.

III. Conclusion

In this paper linear stochastic model known as multiplicative seasonal autoregressive integrated moving average model (SARIMA) was used to simulate droughts in Gadarif region, Sudan. The models are applied to simulate droughts using Standardized Precipitation Index (SPI) series, the results show that the fitted model is adequate to the SPI_6 for Gadaref station is the SARIMA (0, 0, 5) (1.0.1) model.

References

Abramowitz, M., Stegun, A. (eds.), (1965). Handbook of Mathematical Formulas, Graphs, and Mathematical Tables. Dover Publications, Inc., New York, USA.

Box, G.E., Jenkins, G.M., (1970), Time Series Analysis; Forecasting and Control. Holden-Day series in time series analysis, San Francisco.

Durdu, O.F., (2010), Application of Linear Stochastic Models For Drought Forecasting in The Buyuk Menderes River Basin, Western Turkey, Stochastic Environmental Research and Risk Assessment, doi:10.1007/s00477-010-0366-3.

Elagib, N.A., M.G. Mansell., (2000), Recent Trends and Anomalies in Mean Seasonal and Annual Temperature Over Sudan. Journal of Arid Environment 45(3): 263-288.

Etuk, E.H., Mohamed T.M., (2014). Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by SARIMA Methods. International Journal of Scientific Research in Knowledge, 2(7), pp. 320-327.

Le Houérou, H.N. (2009): Bioclimatology and Biogeography of Africa. Verlag Berlin Heidelberg, Germany.

Mishra, A. K., Desai, V. R., (2005), Drought Forecasting Using Stochastic Models. Stochastic Environmental Research and Risk Assessment 19(5): 326-339.

Edwards, D.C., McKee, T.B., (1997). Characteristics of 20th Century Drought in the United States at Multiple Scales. Atmospheric Science Paper 634.

Cacciamani, C., Morgillo, A., Marchesi, S., Pavan, V. (2007). Monitoring and Forecasting Drought on a Regional Scale: Emilia-Romagna Region. Water Science and Technology Library 62(1): 29-48

Thom, H.C.S., (1958). A note on gamma distribution. Monthly Weather Rev. 86, 117–122.

Press WH, Flannery BP, Teukolsky SA, Vetterling WT. 1986. Numerical Recipes. Cambridge University Press: Cambridge.

McKee, T. B.; N. J. Doesken; and J. Kleist. (1993). "The relation of drought frequency and duration to time scales." Proceedings of the Eighth Conference on AppliedClimatology; pp. 179–84. American Meteorological Society, Boston.

Tsakiris, G., Vangelis, H. (2004). Towards a Drought Watch System based on spatial SPI. Water Resources Management 18 (1): 1-12.

Mishra, A. K., Desai, V. R., (2005), Drought Forecasting Using Stochastic Models. Stochastic Environmental Research and Risk Assessment 19(5): 326-339.

Durdu, O.F., (2010), Application of Linear Stochastic Models For Drought Forecasting in The Buyuk Menderes River Basin, Western Turkey, Stochastic Environmental Research and Risk Assessment, doi:10.1007/s00477-010-0366-3.

Appendixes I

Standardized Precipitation Index (SPI) Computation Methodology

In most cases, the Gamma distribution is the distribution that best models observed precipitation data. Thom (1958) found the gamma distribution to fit precipitation time series well. The gamma distribution is defined by its frequency or probability density function:

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \qquad for \quad x > 0$$

Where $\alpha > 0$the shape parameter is $\beta > 0$ is the scale parameter and $x > 0$ is the amount of precipitation. $\Gamma(\alpha)$ is the value taken by the standard mathematical function known as the Gamma function, which is defined as

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1}\, e^{-y} dy$$

Computation of the Standardized Precipitation Index (SPI) involves fitting a gamma probability density function to a given frequency distribution of precipitation totals for a station. The alpha and beta parameters of the gamma probability density function are estimated for each station, for each time scale of interest (3 months, 12 months, 48 months, etc.), and for each month of the year. Edwards & McKee (1997) suggest estimating these parameters using the approximation of Thom (1958) for maximum likelihood as follows:

$$\alpha = \frac{1}{4A}\left(1 + \sqrt{1 + \frac{4A}{3}}\right)$$

$$\beta = \frac{\bar{X}}{\alpha}$$

Where for n observation

$$A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n}$$

The resulting parameters are then used to find the cumulative probability of an observed precipitation event for the given month and time scale for the station in question. Integrating the probability density function with respect to *x* and inserting the estimates of α and β yields an expression for the cumulative probability *G(x)* of an observed amount of precipitation occurring for a given month and time scale:

$$G(x) = \int_0^x g(x)dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x x^{\alpha-1} e^{-x/\beta} dx$$

Putting $t = \frac{x}{\beta}$, this equation becomes the incomplete gamma function:

$$G(X) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt$$

This is the incomplete gamma function. Values of the incomplete gamma function are computed using an algorithm taken from Press et al. (1986).

The Gamma function is not defined by x = 0 and since there may be no precipitation, the cumulative probability becomes (Cacciamani et al., 2007):

$$H(x) = q + (1 - q)G(x)$$

where *q* is the probability of no precipitation and *H* (*x*) is the cumulative probability of precipitation observed. The cumulative probability is then transformed into a normal standardized distribution with null average and unit variance from which we obtain the SPI index.

The cumulative probability, H(x) is then transformed to the standard normal random variable Z with mean zero and variance one, which is the value of SPI. Following Edwards and McKee (1997), an approximate conversion is used in this research, as provided by Abramowitz and Stegun (1965) as an alternative:

$$Z = SPI = -\left(K - \frac{c_0 + c_1 K + c_2 K^2}{1 + d_1 K + d_2 K^2 + d_3 K^3}\right) \quad for \quad 0 < H(x) \leq 0.5$$

$$Z = SPI = +\left(K - \frac{c_0 + c_1 K + c_2 K^2}{1 + d_1 K + d_2 K^2 + d_3 K^3}\right) \quad for \quad 0.5 < H(x) < 1.0$$

Where

$$K = \sqrt{\ln \frac{1}{(H(x))^2}} \quad for \quad 0 < H(x) \le 0.5$$

$$K = \sqrt{\ln \frac{1}{(1-H(x))^2}} \quad for \quad 0.5 < H(x) < 1.0$$

Where x is Precipitation, H(x) isThe cumulative probability of precipitation observed and c_0 ,c_1 ,c_2 ,d_1 , d_2 and d_3 are Constants with the following values:

c_0 =2.515517, c_1 =0.802853, c_2 = 0.010328 d_1 =1.432788 d_2 = 0.189269 d_3 = 0.001308