ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 11(48), DOI: 10.17485/ijst/2018/v11i48/123057, December 2018*

Seasonal Predictability of Rainfall data using Box-Jenkins model in Kordofan State, Sudan

M. Nirmala[1]* and Tariq Mahgoub Mohamed[2]

[1]Sathyabama Institute of Science and Technology, Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600119, Tamil Nadu, India; nirmala.maths@sathyabama.ac.in
[2]Department of Civil Engineering, Jazan University, Jazan, KSA; tariqcivil73@yahoo.com

Abstract

Objectives: In this work, the Box-Jenkins approach, which known as Seasonal Autoregressive Integrated Moving Average Model (SARIMA) model, was applied to predict monthly rainfall in Kordofan state, Sudan. **Methods/Statistical Analysis:** Using the stochastic models to predict monthly rainfall is an important issue for planning many water resources projects. The monthly rainfall data were obtained from the Sudan Meteorological Authority, covering the period 1971-2010. **Findings**: Test of the original data displays horizontal trend and seasonal periodicity. The data is checked for non stationarity through Augmented Dickey- Fuller Unit Root Test (ADF). The Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) were used to identify the seasonality and it was removed by employing first order seasonal differencing. The SARIMA (0,0,1)x(0,1,1)12 model was selected to be most proper for predicting monthly rainfall. Application/Improvements: This model may be applied as a foundation for monthly rainfall Predicting in Kordofan state.

Keywords: Rainfall, Seasonality, Seasonal Prediction, Box – Jenkins SARIMA, Stationary

1. Introduction

Kordofan state is situated in the western portion of the Sudan with an overall area of 380,000 km^2. The state lies between latitudes 9.5º and 16.5º N and longitudes 27.5º and 32º E with a population of more than 4.3 million inhabitants. Most areas of the state are an undulating plain with some mountains series in the southeast part. Sandy and non cracking clay soils exist north of Khor Abu Habil, the most important seasonal stream in the state, while cracking clay soils take place south of Khor Abu Habil in the plains between the mountains. The major resources of water in the region are the rainfall, non perennial streams and groundwater in the forms, well, pond and lake. Annual rainfall varies between 50 mm on the northern boundary to about 800 mm over the southern border. The duration of rainfall season[1] is about five months starting in May and finish in September. Khor Abu Habil has an estimated annual discharge of about 100 million m^3. The agricultural production in Kordofan state is mainly depending upon the rainfall. The main crops grown are millet, sorghum, and groundnuts. The region is one of the most important producers of gum Arabic in the world.

En Nahud region as shown in Figure 1 was selected since it is the most important agricultural area depending on rainfall for cultivation in Kordofan state. The region has an annual rainfall which ranges as 139 – 695 mm

*Author for correspondence

Figure 1. Map of Sudan illustrating En Nahud region.

through the previous forty years. The year 2007 received the maximum quantity of rainfall of 695 mm followed by the year 1980, 681 mm and the minimum amount of rainfall was recorded in 1984 and 1986 as 139mm and 275 mm respectively, the period of dryness. The average annual evapotranspiration and temperature are 2148 mm and 27.2° *C*, respectively[2]. Drought phenomenon threatens huge areas of rain fed land in Kordofan state. The relationship between production and drought demonstrate that a 10 percent lower in the average annual rainfall implies a 3.7 percent drop in yield at the country scale. In the year 2000, drought decreased food reserves and caused prices to ascend three-fold compared to the same period in the previous year. Accordingly, rainfall prediction is an important matter in planning varied water resource projects in the state.

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model was applied to model different seasonal time series. SARIMA $(3,1,3)x(2,1,1)_{12}$ model was used[3] to forecast short term inflation in Ghana. A Seasonal ARIMA model was identified to predict the monthly rainfall data for the rainfall stations in Sudan. ARIMA model was compared with radial basis function[4] for forecasting monsoon rainfall in Tamilnadu.

2. Materials and Methods

2.1 Data

The data analyzed in the present work are monthly rainfall for En Nahud gauge station. The historical rainfall data were obtained from the Sudan Meteorological Authority (SMA), covering the period 1971-2010. It was presented graphically in Figure 2.

2.2 Modeling by SARIMA Method

A stationary time series means that the observed series show no systematic change in mean and no systematic change in variance or correlation structure. These types of time series can be modeled in numerous ways such as an Auto Regressive (AR) process, Moving Average (MA) process and an Auto Regressive and Moving Average

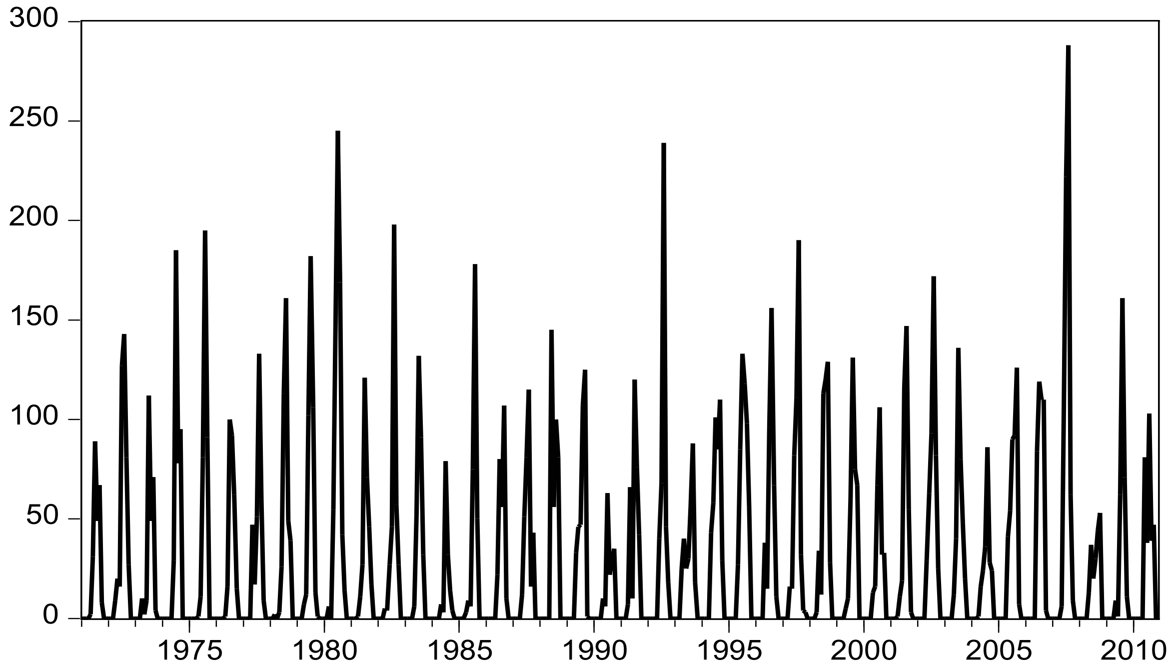


Figure 2. Monthly rainfall for En Nahud station.

(ARMA) process[5]. In real time, the series are generally non-stationary. To make it as stationary, it can be differenced. These models are called Auto Regressive Integrated Moving Average (ARIMA) models.

The ARIMA model with order *p* for AutoRegressive and order *q* for Moving Average operates on *d*th difference of the time series X_t Z_t. It can be written as

$$\phi_p(B)\nabla^d X_t = \theta_q(B)\varepsilon_t \qquad (1)$$

where, the back shifting operator B is defined as

$$B^s X_t = X_{t-s} \qquad (2)$$

$\phi(B)$ and $\theta(B)$ are polynomials of order *p* and *q*, respectively.

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots\ldots\phi_p B^p) \qquad (3)$$

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots..\theta_q B^q) \qquad (4)$$

In day-to-day life, generally the time series may possess seasonality which repeats every *s* period. Box and Jenkins generalized their ARIMA model to capture the seasonal behavior. The multiplicative seasonal model ARIMA (*p,d,q*)x(*P,D,Q*)*s* is of the form

$$\phi_p(B)\Phi_p(B^s)\nabla_1^{\ d}\nabla_s^{\ D} X_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \qquad (5)$$

where, $\nabla_s^{\ D}$ are the *D-th* order seasonal differences and ϕ_p, θ_q, Φ_p and Θ_Q are polynomials of order *p, P, q* and *Q* respectively.

The development of SARIMA model consists of three steps namely as:

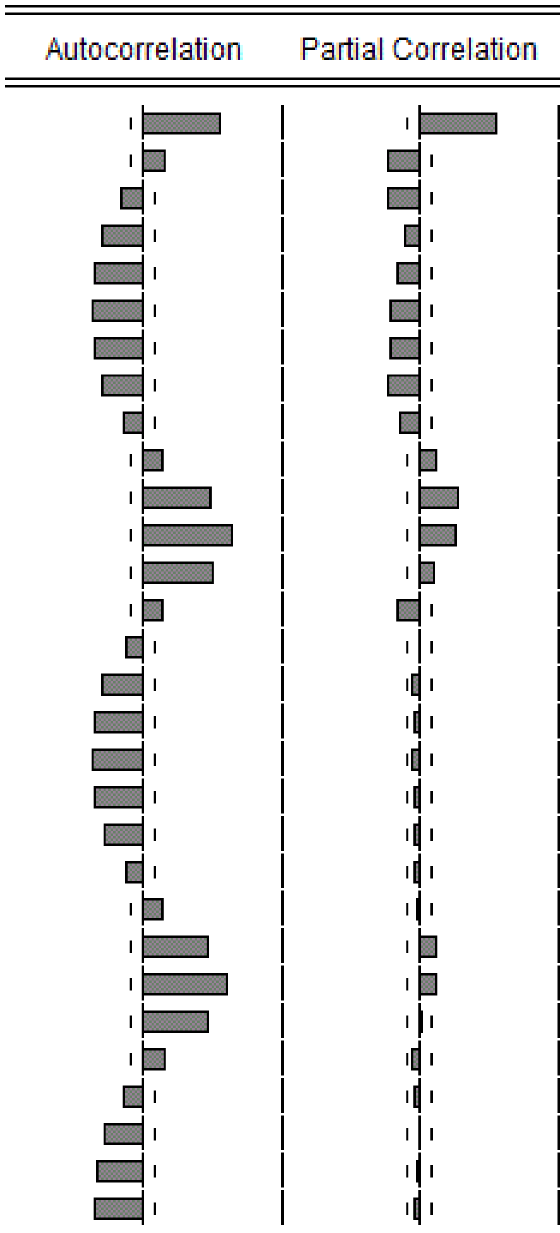
Model identification, parameter estimation and diagnostic checking. Using the above steps many tentative models can be identified for the same time series and a model selection criterion can be used to choose the best one. The most commonly used form is the Akaike Information Criterion is given by [6,7].

$$AIC = N \text{ Log } \sigma^2 + 2k \quad (6)$$

where, N denotes the length of the series and K is the number of model parameters. The model with least AIC value is chosen to be the best model among the tentative models. The statistical software Reviews 9 was employed for the analytical work.

3. Results and Discussion

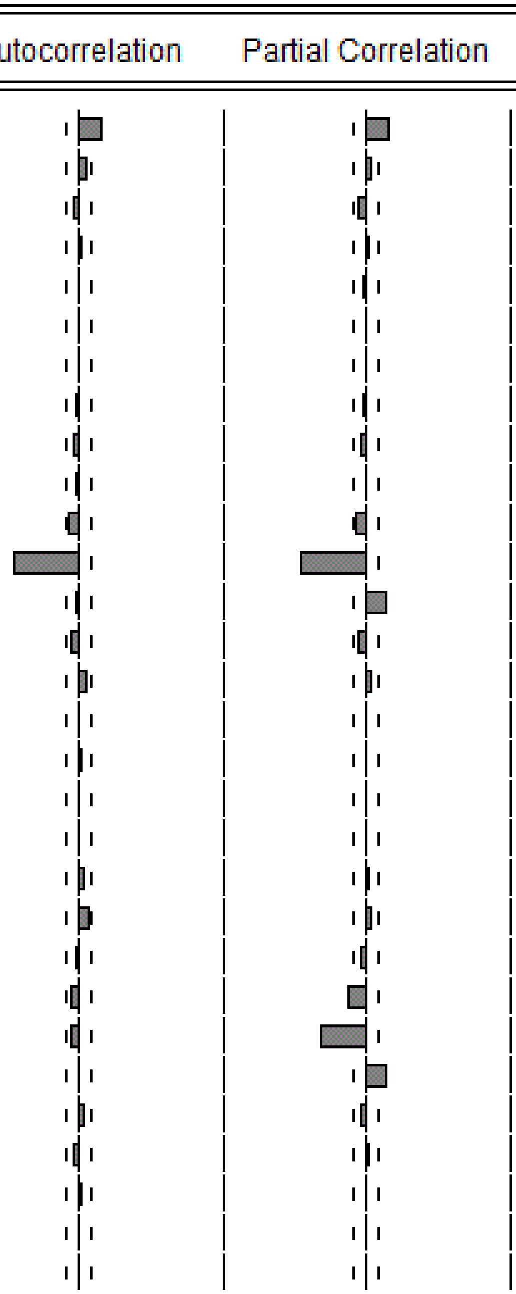
The graphical presentation of the monthly rainfall of En Nahud station, Figure 2, was used to test the stability and stationarity of the data. It illustrates seasonal fluctuations without noticeable trend[8]. The seasonality happens every twelve months and the data is non stationary. The ACF plot shown in Figure 3 proves that the series is non sta-



	AC	PAC	Q-Stat	Prob
1	0.573	0.573	158.36	0.000
2	0.172	-0.232	172.64	0.000
3	-0.158	-0.231	184.81	0.000
4	-0.308	-0.100	230.80	0.000
5	-0.367	-0.167	296.36	0.000
6	-0.379	-0.207	366.61	0.000
7	-0.362	-0.223	430.69	0.000
8	-0.306	-0.238	476.74	0.000
9	-0.154	-0.145	488.40	0.000
10	0.155	0.116	500.22	0.000
11	0.505	0.291	625.82	0.000
12	0.670	0.269	847.95	0.000
13	0.529	0.107	986.73	0.000
14	0.142	-0.158	996.79	0.000
15	-0.134	0.001	1005.7	0.000
16	-0.306	-0.050	1052.3	0.000
17	-0.354	-0.030	1115.0	0.000
18	-0.371	-0.051	1184.1	0.000
19	-0.356	-0.047	1247.7	0.000
20	-0.289	-0.037	1289.8	0.000
21	-0.127	-0.034	1297.9	0.000
22	0.146	-0.012	1308.7	0.000
23	0.478	0.117	1424.3	0.000
24	0.635	0.120	1628.9	0.000
25	0.492	0.015	1751.8	0.000
26	0.157	-0.055	1764.3	0.000
27	-0.143	-0.034	1774.8	0.000
28	-0.291	-0.008	1818.2	0.000
29	-0.351	-0.016	1881.3	0.000
30	-0.364	-0.038	1949.5	0.000

Figure 3. ACF and PACF Plot for En Nahud Monthly Rainfall.

tionary. Also non-stationarity is identified through the Augmented Dickey-Fuller Unit Root Test (ADF), which was tabulated in Table 1. Once again the ADF test was performed on the seasonally differenced data. The results show that the data is stationary which is tabulated in Table 2. The plot shown in Figure 4 shows the ACF and PACF plots of the data after seasonal difference is taken into account. It displays that the seasonality is vanished and the data became stationary. The autocorrelation structure in Figure 4 proposes many models. The proposed models



Lag	AC	PAC	Q-Stat	Prob
1	0.160	0.160	12.131	0.000
2	0.062	0.037	13.950	0.001
3	-0.036	-0.053	14.556	0.002
4	0.009	0.021	14.597	0.006
5	-0.008	-0.008	14.628	0.012
6	-0.001	-0.002	14.629	0.023
7	0.000	0.003	14.629	0.041
8	-0.020	-0.022	14.822	0.063
9	-0.037	-0.032	15.491	0.078
10	-0.020	-0.007	15.674	0.109
11	-0.079	-0.075	18.659	0.067
12	-0.468	-0.463	124.20	0.000
13	-0.023	0.148	124.46	0.000
14	-0.062	-0.051	126.34	0.000
15	0.062	0.031	128.19	0.000
16	-0.007	-0.001	128.21	0.000
17	0.012	-0.007	128.28	0.000
18	-0.002	0.001	128.28	0.000
19	0.000	-0.000	128.28	0.000
20	0.027	0.014	128.65	0.000
21	0.074	0.038	131.31	0.000
22	-0.010	-0.030	131.36	0.000
23	-0.054	-0.125	132.78	0.000
24	-0.064	-0.324	134.83	0.000
25	-0.007	0.154	134.86	0.000
26	0.028	-0.046	135.25	0.000
27	-0.028	0.014	135.65	0.000
28	0.010	0.006	135.70	0.000
29	-0.003	0.005	135.70	0.000
30	0.006	0.004	135.72	0.000

Figure 4. ACF and PACF plot for En Nahud after seasonal difference.

Table 1. Unit root test of the original data

Augmented Dickey-Fuller test statistic -1.164765		t-Statistic	Prob.*
		0.2230	
Test critical values:	1% level	-2.569923	
	5% level	-1.941503	
	10% level	-1.616244	

Table 2. Unit root test of the differenced data

Augmented Dickey-Fuller test statistic -8.521814		t-Statistic	Prob.*
		0.0000	
Test critical values:	1% level	-2.570054	
	5% level	-1.941521	
	10% level	-1.616232	

Table 3. Comparison of the models

Variable	Station	Model	AIC
	En Nahud	SARIMA(0,0,1)x(0,1,1)$_{12}$	9.5121
		SARIMA(0,0,0)x(0,1,1)$_{12}$	9.5400
		SARIMA(0,0,1)x(2,1,1)$_{12}$	9.5129
		SARIMA(0,0,1)x(1,1,1)$_{12}$	9.5144

and AIC are explained in Table 3. Obviously, the model SARIMA (0,0,1)x(0,1,1)$_{12}$ has the minimum amounts of AIC that one would choose this model.

After the identification stage, estimation stage was carried out. The parameters are tabulated in Figure 5. The result concludes that all the parameters are significant as their *p*-values are less than 0.05. The entire absolute values of the inverted AR and MA roots are less than 1 which proves the stationarity and invertibility.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.161271	0.032461	4.968163	0.0000
MA(12)	-0.972847	0.061361	-15.85453	0.0000
MA(13)	-0.188424	0.033670	-5.596261	0.0000
SIGMASQ	722.1765	33.26633	21.70893	0.0000
R-squared	0.518987	Mean dependent var		0.132479
Adjusted R-squared	0.515877	S.D. dependent var		38.78894
S.E. of regression	26.98893	Akaike info criterion		9.512153
Sum squared resid	337978.6	Schwarz criterion		9.547610
Log likelihood	-2221.844	Hannan-Quinn criter.		9.526105
Durbin-Watson stat	1.978832			

Figure 5. Estimation of SARIMA (0,0,1)x(0,1, 1) $_{12}$ model.

The third stage is to test if the data contain any systematic pattern which still can be eliminated to progress the picked model. Different tests were done on the residual series. The results obtained through the correlogram of the residuals test, Box-Ljung test, and the Durbin-Watson (DW) statistic test were analysed. The correlogram of residuals series are illustrated in Figure 6. All residuals autocorrelation function and partial autocorrelation function lie within the confidence limits; hence the plot confirms the absence of significant correlation between residuals. The Ljung-Box test is employed for testing independence of the residual. Figure 6 proves that there is no serial correlation in the residuals, since the autocorrelations and partial autocorrelations at all lags less than 0.05 and all *Q* statistics is insignificant with large *p*-values. The DW statistic used to check the serial correlation between the residuals. If the residuals are uncorrelated, the DW value will equal 2. The DW test statistic value in Figure 5 is found to be 1.9788 which concludes that there is no serial correlation between the residuals.

Finally, among the identified models, SARIMA (0,0,1) x$(0,1,1)_{12}$ model is appropriate to represent the monthly rainfall and could be applied to predict the forthcoming years. According to equation (3), the estimated model is given in the mathematical form as:

$$(1-B_{12})X_t = (1-\theta_1 B)(1-\Theta_1 B_{12})\varepsilon_t \qquad (7)$$

Then, substituting the estimated parameter values, the previous equation can be written in a format that is used in predicting phase

$$X_t = X_{t-12} + 0.161271\varepsilon_t - 1 - 0.972847\varepsilon_{t-12} - 0.188424\varepsilon_{t-12} + \varepsilon_t \qquad (8)$$

where, X_t = the monthly rainfall at month *t*

X_{t-12}= the monthly rainfall at month (t-12)

Lag	AC	PAC	Q-Stat	Prob
1	0.009	0.009	0.0410	
2	0.063	0.063	1.9078	
3	-0.043	-0.044	2.7826	
4	0.014	0.011	2.8777	0.090
5	-0.005	0.001	2.8875	0.236
6	-0.005	-0.008	2.8989	0.407
7	0.001	0.002	2.8991	0.575
8	-0.004	-0.003	2.9055	0.715
9	-0.029	-0.030	3.3115	0.769
10	0.011	0.012	3.3649	0.849
11	-0.015	-0.013	3.4800	0.901
12	-0.052	-0.056	4.8022	0.851
13	0.100	0.106	9.6457	0.472
14	-0.045	-0.044	10.633	0.475
15	0.059	0.044	12.331	0.419
16	-0.015	0.000	12.434	0.492
17	0.017	0.003	12.575	0.560
18	-0.006	-0.000	12.592	0.634
19	-0.003	-0.004	12.595	0.702
20	0.016	0.015	12.720	0.755
21	0.046	0.045	13.757	0.745
22	-0.015	-0.013	13.869	0.791
23	-0.039	-0.051	14.629	0.797
24	-0.083	-0.073	18.053	0.646
25	0.003	0.016	18.057	0.703
26	0.032	0.026	18.559	0.726
27	-0.018	-0.011	18.714	0.767
28	0.015	0.002	18.830	0.805
29	0.001	0.012	18.830	0.844
30	0.002	-0.006	18.833	0.876

Figure 6. ACF and PACF Plot of the Residual.

ε_{t-1}, ε_{t-12} = the errors values at time (t-1), (t-12)

4. Conclusion

In the present work the stochastic approach developed by Box and Jenkins, which famed as SARIMA model, was utilized to predict monthly rainfall in En Nahud station, Sudan. It has been proved that the monthly rainfall follows a SARIMA (0,0,1) x $(0,1,1)_{12}$ model. The selected model is more appropriate to predict the monthly rainfall for the next years to assist designers to set up the priorities for various water resources projects.

5. References

1. Zakieldeen SA. Adaptation to climate change: A vulnerability assessment for Sudan. International Institute for Environment and Development; 2009. p. 1–24.
2. Houérou HN. Bioclimatology and biogeography of Africa. Springer; 2009. p. 1–241. https://doi.org/10.1007/978-3-540-85192-9.

3. Nasiru S, Solomon S. Empirical approach to modeling and forecasting inflation in Ghana. Current Research Journal of Economic Theory. 2012; 4(3):83–7.
4. Nirmala M. A comparative study of ARIMA model with radial basis function for forecasting monsoon rainfall in Tamil Nadu. International Journal of Applied Engineering Research. 2015; 10(1):1045–54.
5. Box GEP, Jenkins GM, Reissel GC. Time series analysis forecasting and control. 3rd Edition. Prentice Hall; 1994. p. 1–598.
6. Akaike H. A New look at the statistical model identification. IEEE Transactions Automatic Control. 1974; 19(6):716–23. https://doi.org/10.1109/TAC.1974.1100705
7. Etuk EH, Mohamed TM. Time series analysis of monthly rainfall data for the Gadaref rainfall station, Sudan, by Sarima methods. International Journal of Scientific Research in Knowledge. 2014; 2(7):320–7. https://doi.org/10.12983/ijsrk-2014-p0320-0327
8. Storch HV, Zwiers FW. Statistical Analysis in Climate research. Cambridge University Press; 1999. p. 1–21. https://doi.org/10.1017/CBO9780511612336.