Application of Linear Stochastic Models for Rainfall Data in West Darfur State, Sudan

Tariq Mahgoub Mohamed

Department of Civil Engineering, Jazan University, Jazan, KSA

Abstract: Using of linear stochastic models to simulate monthly rainfall is considered as one of the most important methods for planning different water resources systems. In this paper, linear stochastic models known as multiplicative seasonal autoregressive integrated moving average models, SARIMA, were used to simulate and forecast monthly rainfall at El Geneina gauging station, West Darfur, Sudan. For the analysis, monthly rainfall data during the period 1970 to 2010 were used. The data was obtained from the Sudan Meteorological Authority (SMA). It is observed that it is seasonal. The seasonality observed in Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots of monthly data was removed using first order seasonal differencing prior to the development of the model. Obviously, the SARIMA $(1,0,0)(0,1,1)_1$ model was found to be most suitable for simulating monthly rainfall over the region. The model was found appropriate to forecast three years of monthly rainfall and assist decision makers to establish priorities for water projects.

Key words: Sudan · El Geneina · Monthly rainfall · SARIMA models

INTRODUCTION

West Darfur is one of the states of the Darfur region, Sudan. It is geographically located in western Sudan, boarded by the North Darfur state to the north and northeast, Central Darfur state to the east and southeast, South Darfur state to the south and Chad to the west. The state has an area of 79,460 km$^2$ and an estimated population of about 1,308,000 according to 2008 Population Census [1]. El Geneina is the capital of the state. The main water resources in the state are the rainfalls, seasonal streams (Wadis) and groundwater. The annual rainfall ranges from less than 200 mm on the northern border to more than 700 mm along the southern boundary. The length of rainy season fluctuates around six months i.e. from May to October. The seasonal streams are usually running during the rainy season. WadiAzum and wadikaja comprise the largest drainage system in the state. WadiAzum has catchment area about 40,000 km$^2$, with an estimated annual discharge of more than 487 million cubic meters [2]. Most of the agricultural production in West Darfur state is rain fed. The main cash crops grown in this state are sesame, groundnuts and tombac (chewing tobacco). Mangoes are produced in many locations along wadis. The state is also a major producer of livestock.

El Geneinagauge station is located at 13.29° N latitudes and 22.27° E longitudes and has elevation 805 meters above sea level (masl). The station is characterized by annual rainfall of 124 - 661 mm with an annual average of 427 mm and standard deviation of 120 mm during the last forty years. The monthly rainfall records show that most of the rainfalls in the period from May to October and reach its peak in August. According to the station records, the year 1995 received the highest amount of rainfall [661mm] followed by the year 2003 [653 mm]; and the lowest amount of rainfall was recorded in 1984 and 1987 [124 and 238 mm respectively], the years of drought. The annual number of rainy days (rainfall > 1 mm) is 90 days and the mean annual reference potential evapotranspiration $(\text{ET}_0)$ using Penman / Monteith criterion for this station is about 2229 mm [3]. The climate in El Geneina is arid with mean annual temperature near 25.9°C [4].

This work is aimed at modeling monthly rainfall records obtained from El Geneina station by seasonal autoregressive integrated moving average (SARIMA) techniques. Rainfall is a seasonal phenomenon of period 12 months. SARIMA modeling has been extensively used to model hydrological seasonal time series. For instance, Nimarla et al. [5] fitted a SARIMA model of order $(0,1,1)(0,1,1)_1$ for monthly rainfall in Tamilnadu in India.
Etuk et al. [6] applied a SARIMA (0,0,0)x(0,1,1)$_{12}$ model to monthly rainfall data for Wad Madani rainfall station in Sudan. Kibunja et al. [7] fitted a SARIMA (1,0,1)x(1,0,0)$_{12}$ to monthly rainfall in Mt. Kenya region, Kenya. Mohamed et al. [8] modeled monthly flow for the Dinder River in Sudan using a SARIMA (2,0,0)x(0,1,1)$_{12}$. Bazrafshan et al. [9] found that the application of SARIMA modeling was suitable for the forecasting of hydrological drought in the Karkheh Basin.

**MATERIALS AND METHODS**

**Data:** For this work, monthly rainfall data were obtained from the Sudan Meteorological Authority (SMA), for the period 1970–2010 from El Geneinagauge station.

**Modeling by SARIMA Methods:** A stationary time series can be modeled in different ways: an autoregressive (AR) process, a moving average (MA) process, or an autoregressive and moving average (ARMA) process. However, an ARMA model can be used when the data are stationary, ARMA models can be extended to non-stationary series by allowing differencing of data series. These models are called autoregressive integrated moving average (ARIMA) models. A time series is said to be stationary if it has constant mean and variance.

The general non-seasonal ARIMA model is AR to order $p$ and MA to order $q$ and operates on $d$th difference of the time series $X_t$; thus a model of the ARIMA family is classified by three parameters $(p, d, q)$ that can have zero or positive integral values. The general non-seasonal ARIMA model may be written as

$$
\phi(B)\nabla^d X_t = \theta(B)e_t
$$

where:

- $B =$ The backward shift operator
- $\phi(B)$ and $\theta(B) =$ Polynomials of order $p$ and $q$, respectively.

$$
\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)
$$

and

$$
\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)
$$

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). Box et al. [10] has generalized the ARIMA model to deal with seasonality and define a general multiplicative seasonal ARIMA model, which are commonly known as SARIMA models. In short notation the SARIMA model described as ARIMA $(p, d, q) \times (P, D, Q)_s$, which is mentioned below:

$$
\phi_p(B) \phi_p(B^s) \nabla^d \nabla^D X_t = \theta_q(B) \theta_q(B^s) e_t
$$

where $p$ is the order of non-seasonal autoregression, $d$ the number of regular differencing, $q$ the order of nonseasonal MA, $P$ the order of seasonal autoregression, $D$ the number of seasonal differencing, $Q$ the order of seasonal MA, $s$ is the length of season, $\Phi_p$ and $\Theta_q$ are the seasonal polynomials of order $P$ and $Q$, respectively.

SARIMA model development consists of the following three steps: model identification, parameter estimation and diagnostic checking. The model that gives the minimum Akaike Information Criterion (AIC) and Hannan-Quinn Criterion (HQ) is selected as best fit model [11, 12]. In this work, the statistical and econometric software Eviews was used for all analytical work. It is based on the least squares optimization criterion.

**Performance Evaluation:** The following measures were used to evaluate the performance of the models:

- **Mean Absolute Error:**

$$
MAE = \frac{1}{n} \sum_{t=1}^{n} |Y_t - F_t|
$$

- **Coefficient of Determination:**

$$
R^2 = \frac{\sum_{t=1}^{n} (Y_t - \bar{Y})(F_t - \bar{F})^2}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}
$$

- **Coefficient of Efficiency:**

$$
E = \left[ \frac{\sum_{t=1}^{n} (Y_t - \bar{F})^2}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} \right]^{1/2}
$$

where, $Y_t$ are the $n$ observed flows, $F_t$ are the $n$ modeled flows, $\bar{Y}$ is the mean of the observed flows, $\bar{F}$ is the mean of the modeled flows.

**RESULTS AND DISCUSSION**

**Model Identification:** Model computation was made with monthly data from between January 1970 and December 2007. The dataset from January 2008 to December 2010 was considered in forecasting estimations of the model.
Fig. 1: Monthly rainfall data for El Geneina station [1970-2007] in mm.

Fig. 2: ACF and PACF Plots for El Geneina Station Monthly Rainfall Series

The graphical presentation of the monthly rainfall data, Figure 1, shows clearly that there is a seasonal cycle in the series. Figure 2 shows the sample autocorrelation function and partial autocorrelation for the monthly data. The seasonal autocorrelation relationships are shown quite prominently, which proves that the series is non-stationary. Non-stationarity is also confirmed by the Augmented Dickey-Fuller (ADF) unit root test on the monthly rainfall data in Table 1. The table displays results of the test: statistic value -1.076 greater than critical values -2.570, -1.941, -1.616 all at 1%, 5% and 10%, respectively. This indicates that the series is non-stationary and also confirm that the data needs differencing in order to be stationary. Etuk et al. [13] also observed monthly rainfall in Sudan as non-stationary. This non-stationarity stems from the seasonal nature of the series.

A seasonal differencing of the monthly data, as shown in Figure 3, yields a series has a flat trend. The ADF test was done again on the seasonally differenced data. The results of the test: statistic value -9.311 less than critical values -2.570, -1.941, -1.616 all at 1%, 5% and 10% respectively. This adjudges that the differenced series is stationary. Figure 4 shows the ACF and PACF plots of the data after we take seasonal difference. It appears that most of the seasonality is gone now and the data became stable. The autocorrelation structure in Figure 4 suggests two SARIMA models. The suggested models, the Akaike Information Criterion (AIC) and the Hannan-Quinn Criterion (HQ) values are shown in Table 2. The model that gives the minimum AIC and HQ criterion is selected as best fit model. Obviously, model SARIMA (1,0,0)x(0,1,1) has the smallest values of AIC and HQ, then one would temporarily have a model SARIMA (1,0,0)x(0,1,1).

Parameter Estimation: After the identification of model using the AIC and HQ criteria, estimation of parameters is done. The value of the parameters, standard errors, t-

<table>
<thead>
<tr>
<th>Station</th>
<th>Variable</th>
<th>ADF Test</th>
<th>Level of Confidence</th>
<th>Critical Value</th>
<th>Probability</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>El Geneina</td>
<td>Monthly Rainfall</td>
<td>-1.076</td>
<td>1%</td>
<td>-2.570</td>
<td>0.2551</td>
<td>Non-stationary</td>
</tr>
</tbody>
</table>

Table1: ADF-unit root test for El Geneina monthly Rainfall
Fig. 4: ACF and PACF Plots for El Geneina Station after one Seasonal Difference

Table 2: Comparison of the Suggested Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Station</th>
<th>Model</th>
<th>AIC</th>
<th>HQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly</td>
<td>El Geneina</td>
<td>SARIMA(0,0,0)x(0,1,1)_1</td>
<td>9.7009</td>
<td>9.7046</td>
</tr>
<tr>
<td>Rainfall</td>
<td>SARIMA(1,0,0)x(0,1,1)</td>
<td>9.6946</td>
<td>9.7019</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Estimation of SARIMA (1,0,0)x(0,1,1)_2 Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>0.114548</td>
<td>0.047361</td>
<td>2.418637</td>
<td>0.0160</td>
</tr>
<tr>
<td>MA(12)</td>
<td>-0.886760</td>
<td>0.026198</td>
<td>-33.84858</td>
<td>0.0000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.449307</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.448058</td>
<td>S.D. dependent var</td>
<td>41.39939</td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>30.75717</td>
<td>Akaike info criter.</td>
<td>9.694628</td>
<td></td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>417187.6</td>
<td>Schwarz criter.</td>
<td>9.713109</td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-2145.360</td>
<td>Hanhan-Quinn criter.</td>
<td>9.701917</td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson stat</td>
<td>1.996101</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Diagnostic Check: Once an appropriate model is selected and its parameters are estimated, the Box-Jenkins methodology requires examining the residuals of the model to verify that the model is an adequate one for the time series. An adequate model should have uncorrelated residuals. This is the minimal condition. Different tests were carried out on the residual series. The tests are summarized briefly in the following paragraphs.

ACF and PACF of Residuals: The ACF and PACF of residuals of the model SARIMA (1,0,0)x(0,1,1)_2 are shown in Figure 5. Most of the values of the RACF and RPACF lie within confidence limits. The figure indicates no significant correlation between the residuals.

The Ljung-Box Test: The Ljung-Box test is used for checking independence of residual. From Figure 5, the goodness of fit values for the autocorrelations of residuals from the model up to lag 36 was = 0.05. The result proves the acceptance of the null hypothesis of model adequacy at the 5% significance level and the set of autocorrelations of residuals was considered white noise.
Forecasting of Monthly Rainfall: SARIMA model can also be used for forecasting future values based on the historical data. The SARIMA $(1,0,0) \times (0,1,1)_{12}$ model was tested for its validity to forecast 36 observations obtained for the years 2008 to 2010 for the station. The observed rainfall was found to be closely aligned to the forecasted values, Figure 6.

Forecasting Accuracy: If the fitted SARIMA $(1,0,0) \times (0,1,1)_{12}$ model has to perform well in forecasting, the forecast error will be relatively small. To check goodness of the prediction, Mean Absolute Error (MAE), coefficient of determination ($R^2$) and Nash-Sutcliffe efficiency criteria ($E$) were used. Table 5 illustrates all of the statistic measures. From the statistics measurement, Table 5, it is observed that the model has lowervalue of MAE. The coefficient of determination ($R^2$) value of 0.81 and Nash-Sutcliffe efficiency criteria ($E$) value of 80% showed the very good performance of the model.

CONCLUSION

In this work, linear stochastic model known as Seasonal Autoregressive Integrated Moving Average model, SARIMA, was used to simulate monthly rainfall for El Geneina gauge station, Sudan. It has been demonstrated that the monthly rainfall follow a SARIMA $(1,0,0) \times (0,1,1)_{12}$ model. This model may be used as the basis for forecasting rainfall in this region. The fitting of SARIMA models to rainfall time series could result in a better tool, which can be used for water resource planning. SARIMA model has the ability to predict accurately the future monthly rainfall for all gauge stations in Sudan.

REFERENCES

