SUST Journal of Engineering and Computer Sciences (JECS), Vol. 17, No.2, 2016

Fitting Probability Distributions of Annual Rainfall in Sudan

Tariq Mahgoub Mohamed[1] and Abbas Abd Allah Ibrahim[2]
[1]Khartoum College of Technology, Sudan
[2]College of Water and Environmental Engineering, Sudan University of Science and Technology (SUST)
tariqcivil73@yahoo.com

Received: 17/06/2015
Accepted: 23/08/2015

ABSTRACT- Annual rainfall data for fourteen rainfall stations in Sudan during the period 1971 to 2010 were analyzed to select the best probability distribution for every station. The rainfall data was obtained from the Sudan Meteorological Authority (SMA). Five distributions were tested, namely Normal, Log normal, Gamma, Weibull and exponential distribution. Three statistical goodness of fit test were used on the basis of the minimum value of test statistic. The best fit probability distribution was selected based on the minimum deviation between observed and estimated values. The normal and gamma distribution were selected as the best fit probability distribution for the annual rainfall in Sudan during the period of the study, respectively.

Keywords: *Probability Distributions, Annual Rainfall, Sudan, Goodness-of-Fit Test.*

المستخلّص – بيانات المطرِ السنويةِ لأربع عشرةَ محطةِ أمطارٍ في السودان فى الفترة مِنْ سَنَةِ 1971 الى2010 حُلّلتْ لاختيار أفضل توزيعِ إحتمالى لائقِ لكُلّ محطة. بيانات المطرَ حُصِلْ عليها مِنْ هيئة الارصاد الجوى السودانية. و قد اختبرت خمسة توزيعات احتمالية ، تسمى التوزيع الطبيعي ،اللوغريثمى الطبيعي ، غاما ، ويبل والتوزيع الأسى. استعملت ثلاثة اختبارات إحصائية لحسن المطابقة على أساس الحَد الأدنى لقيمة الاختبار الإحصائيةِ. أفضل توزيع إحتمالِي لائقِ اختيرَ استنادا على الانحراف الأدنى بين القيم الفعليةِ والقيم المقدرةِ.التوزيع الطبيعي و توزيع غاما اختيرا كأفضل توزيعِ إحتمالى لائقِ للمطرِ السنويِ في السودان أثناء فترةِ الدراسةِ، على التوالي.

INTRODUCTION

Sudan economy is mainly based on agriculture, industry and services sectors. Statistics of 2004 indicate that agriculture alone contributed about 39.2% of the GDP while services and industry sectors contributed about 32.8 and 28.0% respectively [1]. Rainfall is regarded as the main source of water for agriculture in Sudan. The total amount of rainfall water in Sudan is estimated at 1094 million cubic meters upon which field crops livestock, forests and pastures depend [2]. Rainfall in the Sudan varies from no rain in the desert region in the north to about 900 mm per year in the forest zones in the south. The length of rainy season fluctuates around six months in the southern region i.e. from April to September, three months in central region and only one month in the northern region. The analysis of rainfall data mainly based upon its distribution type. Therefore, the study of the rainfall distribution is very important for the Sudan economy.

In our life, the probability distributions are used in different fields of science such as engineering, medicine, economic and agricultural science. Probability distributions of rainfall have been studied by many researchers [3-10]. Rainfall probabilities for monthly and annual durations have been documented using the gamma distribution [3]. Waylen et.al [4] found that the normal distribution was the best fit probability distribution for annual rainfall in Costa Rica. Abdullah and Al-Mazroui [5] reported that the normal and gamma distributions were the best fit probability distribution for describing annual rainfall of the southwestern region of Saudi Arabia. Salami [6] revealed that the log-Pearson type III distribution was the best fit probability distribution for precipitation data for Texas region. Husak et al. [7] studied the use of Gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. Osati et al. [8] reported that Pearson and log Pearson distributions were the best fit probability distribution for annual precipitation of Mazandraan and Golestan provinces in Iran. Alghazali and Alawadi [9] found that there is no suitable probability distribution for describing monthly rainfall in thirteen stations across Iraq. ALahmadi et al [10] stated that Generalized Extreme Value was the best fit statistical distribution for Madinah daily rainfall data, Saudi Arabia.

The main objective for this study is the analysis of the probability distribution of the annual rainfall in different rainfall stations across Sudan. Different probability distributions are compared in this paper to fit fourteen rainfall gauge stations in Sudan.

MATERIALS AND METHODS

Data

Fourteen rainfall gauge stations in Sudan were selected with monthly and annually rainfall series during the period 1971 to 2010. These stations were selected based on the basis of reasonably long records for the monthly rainfall data in locations which represent different climatic zones in Sudan. These stations are classified by Elagib and Mansell [11] as follow:

- Kadugly, Nyala and ElGedaref: semi-arid.
- ElFasher, ElGenina, ElObeid, Kosti, EnNahud, Wad Medani, Sennar, Kassala and Khartoum :Arid
- Atbara and Port Sudan: hyper-arid.

All - time series were checked to find out all missing data. Individual missing data values were filled in using Qureshi and Khan Method [12].

Methodology

The procedures used for this work can be summarized in the following steps:

Statistics of Annual Rainfall

Using the sample data x_i (i=1,2,…,n) the basic statistical descriptors of the annual rainfall series, the mean $\bar{X}$, standard deviation S, coefficient of variation C_V, skewness C_S, kurtosis C_K, minimum and maximum values, have been estimated for each station.

The Probability Distribution Functions

The following five probability distributions namely, normal, log normal, gamma, weibull and exponential distribution were used to select the best fit probability distribution for annual rainfall in Sudan. The description of the probability distribution functions are presented in Table 1.

Testing the Goodness of Fit

The goodness-of-fit tests ,namely, Kolmogorov-Smirnov test , Anderson-Darling test and Cramer-von Mises test were used at α (0.05) level of significance for the selection of the best fit distribution. The best fitted distribution is selected based on the minimum error produced, which is evaluated by the following techniques:

(i) Kolmogorov-Smirnov Test

The Kolmogorov–Smirnov (K-S) test is a well known goodness-of-fit statistic that compares an empirical distribution function, $\hat{F}_X$, with a specified distribution function F_Y. The test is an alternative to the chi-square goodness of fit test. The Kolmogorov–Smirnov statistic (D) can be computed as [13],

$$D = max\left|\hat{F}_X(x) - F_Y(x)\right| \quad (1)$$

measures the distance between the empirical distribution function, $\hat{F}_X$, and the specified distribution F_Y. Obviously, a large difference indicates an inconsistency between the observed data and the statistical model F_Y.

(ii) Anderson-Darling Test

The Anderson-Darling (A-D) test was introduced by Anderson and Darling to place more weight or discriminating power at the tails of the distribution. This can be important when the tails of the selected theoretical distribution are of practical significance. The test statistic (A^2) is defined as [14]

$$A^2 = -\sum_{i=1}^{n}[(2i-1)\{lnF_X(x_i) + ln[1 - F_X(x_{n+1-i})]\}/n] - n \quad (2)$$

Where F_X is the cumulative distribution function of the specified distribution and x_i is the ordered data. The critical value,C_α, in the A-D test for a given significance level α depends on the form of the proposed theoretical distribution.

(iii) Cramer-Von Mises Test

Cramer-Von-Mises (CVM) criterion is used in testing the goodness-of-fit of a probability distribution compared to a given empirical distribution function. The Cramer-von-Mises test statistic can be computed as [15]:

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i-1}{2n} - F(x_i)\right]^2 \qquad (3)$$

Where x_i is the increasing ordered data and *n* is the number of sample size.

Least Square Method

The least square method was also used to select the best fit probability. It is a mathematical selection criterion of model building. When there are several competing models to choose from, select the model that gives the minimum sum of residuals. The residuals (R) can be computed as

$$R = \left|\sum_{i=1}^{n} (y_i - \hat{y}_i)\right| \qquad (4)$$

Where, y_i= is the observed data, $\hat{y}_i$= is the estimated data and *n* is the number of sample size.

The distribution with minimum sum of residuals was selected as the best fit probability distribution for that particular station.

Quantile-Quantile Plots (Q-Q)

Normal quantile-quantile plots, (Q-Q), are used to evaluate whether the data follow a specified theoretical distribution. If the two distributions are the same, the (Q-Q) plot should lie on a straight line.

Statistical Software

The econometric and statistical software Eviews-7 was used for all the analytical work.

RESULTS AND DISCUSSION

Statistical Properties of the Annual Rainfall

The basic statistical properties of the annual rainfall series have been calculated for each station. The results obtained are given in Table 2.

The variability in annual rainfall during 1971–2010, as estimated by the coefficient of variation C_V , ranges from 18.9 % in Kadugli to 89.4 % in Atbara as shown in Table 2. This finding is in agreement with Elagib and Mansell [11] , who reported that the annual rainfall variability in Sudan increases with decreasing mean rainfall during the period 1961-1990, which calculated by the coefficient of variation, ranges from 13.8% to 122.9%.

The coefficient of skewness is used to verify the degree of asymmetric of a distribution around the mean. The skewness of a normal distribution is zero. From the result of Table 2, most values of skewness lies near zero except five stations namely Atbara, Khartoum, Port Sudan, El Obeid and En Nahud.

Kurtosis refers to the extent of peakedness or flatness of a probability distribution in comparison with the normal probability distribution. The Kurtosis of a normal distribution is 3. Most values of kurtosis are around 3 in the range of 2.30 at Kadugly station to 3.83 at Sennar. Khartoum and Atbara stations obtain the highest value of kurtosis with the value of 9.00 and 7.63 respectively.

The three test statistic for each rainfall station data were calculated for all probability distributions. The probability distribution with the minimum error along with their test statistic is presented in Table 3. It has been observed that Normal distribution using the three tests independently obtained the minimum error for annual rainfall in Kassala station. Thus, the Normal distribution is suitable for this station.

It has been, also, observed that Normal distribution using Kolmogorov Smirnov test, Normal distribution using Anderson Darling test and Gamma distribution using Cramer-von-Mises test obtained the minimum error for annual rainfall in Wad Madani station. Thus the Normal and Gamma probability distributions were selected as the best fit based upon the three tests independently.

Additional criteria needed to chose the best distribution for Wad Madani station. The least square method and the quantile-quantile plots were also used to select the best fit probability. The residuals were computed for the two distributions using the least square method. The Normal probability distribution has the minimum residual. Theoretical quantile-quantile plots, (Q-Q), for the two considered distribution are used for Wad Madani station, Figure 1. Using visual judgment the normal probability distribution is chosen as the best suitable probability distribution for annual rainfall at Wad Madani station. The best selected probability distribution for each station is presented in Table 4.

CONCLUSION

Normal distribution has been found to be the best fitted distribution for the representation of the annual rainfall in Sudan. However, gamma distribution is ranked second while exponential distribution is ranked last. Normal distribution adequately described annual rainfall for 8 stations, while Gamma distribution adequately described annual rainfall for 4 stations. Exponential distribution adequately described annual rainfall for most stations in hyper arid zones.

REFERENCES

[1] Central Bank of Sudan, (2004), Annual Report, Khartoum, Sudan.

[2] University of Khartoum consultancy corporation, (2011), Karkoag project- Hydrologic report and Bathymetric survey results, UKCC.

[3] Hann, C.T., (1977), Statistical methods in hydrology. The Iowa State University Press.

[4] Waylen, P. R., Qusesada, M. E. and Caviedes, C.N. (1996), Temporal and spatial variability of annual precipitation in Costa Rica and southern oscillation .Int J Climatol 14: pp.173-193.

[5] Abdullah, M.A., and AL-Mazroui, M.A., (1998), Climatological study of the southwestern region of Saudi Arabia. I. Rainfall analysis, Climate Research, Vol. 9: pp.213-223.

[6] Salami, A.W. (2004), Prediction of the annual flow regime along Asa River using probability distribution models. AMSE periodical, Lyon, France. Modeling C. 65(2): pp. 41-56.

[7] Husak G. J., Michaelsen J. and Funk C. (2006), Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications, International Journal of Climatology, Published online in Wiley InterScience, Royal Meteorological Society, Santa Barbara, U. S. A., DOI: 10.1002/joc.1441.

[8] Osati, K., Mohammed, M., Karimi, B, Naghi, S, Mobaraki, J, (2010), determining suitable probability distribution models for annual precipitation data (A case study of Mazandaran and Golestan provinces), Journal of Sustainable Development, Vol. 3 No. 1: pp.159-168.

[9] Alghazali, N.O., and Alawadi, D.A., (2014), Fitting statistical distributions of monthly rainfall for some iraqi stations, Civil and Environmental Research, Vol.6, No.6: pp.40-46.

[10] Alahmadi , F., Abd Rahman, N., and Abdulrazzak, M., (2014), Evaluation of the best fit distribution for partial duration series of daily rainfall in Madinah, western Saudi Arabia. Evolving water resources system: understanding, predicting and managing water-society interactions proceedings of ICWRS2014, Bologna,Italy.

[11] Elagib, N.A. and Mansell, M.G. (2000), Recent trends and anomalies in mean seasonal and annual temperature over Sudan. Journal of Arid Environment 45: pp. 263-288.

[12] Qureshi, S. and Khan, N. (1994), Estimations of climatic transition in Riyadh (Saudi Arabia) In Global Warming Perspectives. Geo Journal 33(4): pp. 423-432.

[13] Storch, H.V., and Zwiers, F, W, (1999), Statistical analysis in climate research, Cambridge University Press.

[14] Anderson, T. W (2010), Anderson- Darling tests of goodness-of-fit, Stanford University press.

[15] Csorgo, S., and Faraway, j., (1994), The exact and asymptotic distributions of Cramer-von Mises statistics, J. R. Statist. Soc. B. 58, No. 1, pp.221-234.

Table 1: Description of used probability distribution functions

Type of Distribution	Probability Density Function	Parameters
Normal	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$	$\mu = Mean$ $\sigma = Standard\ Deviation$
Log Normal	$f(x) = \frac{exp\left[\frac{-1}{2}\left(\frac{ln(x)-\mu}{\sigma}\right)^2\right]}{\sqrt{2\pi x^2\sigma^2}}$	$\sigma = Scale\ Parameter\ (\sigma > 0)$ $\mu = Shape\ Parameter\ (\mu > 0)$
Gamma	$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$	$\alpha = Shape\ Parameter\ (\alpha > 0)$ $\beta = Scale\ Parameter\ (\beta > 0)$ $\Gamma = Gamma\ Function$
Weibull	$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right]$	$\alpha = Shape\ Parameter\ (\alpha > 0)$ $\beta = Scale\ Parameter\ (\beta > 0)$
Exponential	$f(x) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}x}$	$\lambda > 0$

Table 2: Statistical properties of annual rainfall series for 14 stations in Sudan (1971-2010).

Rainfall Station	Mean	Standard Deviation	Coefficient of Variation	Skewness Coefficient	Kurtosis	Min	Max
Port Sudan	80.14	69.80	0.871	1.22	4.16	0.0	281.0
Atbara	52.78	47.21	0.894	1.83	7.63	0.0	239.7
Khartoum	122.37	69.93	0.571	1.69	9.00	4.40	415.5
Kassala	243.49	82.73	0.340	0.06	2.63	75.6	394.8
Wad Medani	288.57	81.54	0.283	0.07	2.64	115.4	443.1
El Gedaref	616.82	117.25	0.190	-0.14	3.28	322.0	872.6
El Fasher	195.47	66.29	0.339	0.57	2.79	72.7	361.5
Sennar	419.63	119.11	0.284	0.41	3.83	174.7	773.9
El Geneina	426.16	120.92	0.284	-0.14	2.92	124.4	661.3
El Obeid	348.38	121.66	0.349	0.99	4.23	161.7	735.5
kosti	348.10	102.2	0.294	0.10	3.71	96.0	602.1
En Nahud	357.10	112.5	0.315	0.93	4.93	138.9	694.4
Nyala	387.92	97.10	0.250	0.34	2.73	197.3	626.1
Kadugli	681.32	129.00	0.189	0.19	2.30	468.8	990.8

Table 3: The probability distribution having the minimum error along with their test statistic for each station

Rainfall Station	Kolmogorov Smirnov Test Critical Value D = 0.210		Anderson – Darling Test*		Cramer - Von Mises Test Critical Value $C_{1-\alpha}$= 0.21	
	Distribution	Statistic	Distribution	Statistic	Distribution	Statistic
Port Sudan	Exponential	0.115	(-)	(-)	Exponential	0.074
Atbara	Exponential	0.134	(-)	(-)	Exponential	0.093
Khartoum	Normal	0.144	(-)	(-)	Normal	0.131
Kassala	Normal	0.086	Normal	0.359	Normal	0.048
Wad Medani	Normal	0.078	Normal	0.294	Gamma	0.038
El Gedaref	Normal	0.104	Normal	0.412	Normal	0.075
El Fasher	Gamma	0.080	Gamma	0.185	Gamma	0.030
Sennar	Normal	0.118	Normal	0.446	Normal	0.083
El Geneina	Normal	0.062	Normal	0.167	Normal	0.023
El Obaied	Gamma	0.095	Gamma	0.404	Gamma	0.070
Kosti	Normal	0.105	Normal	0.519	Normal	0.081
En Nahoud	Gamma	0.090	Gamma	0.481	Gamma	0.063
Nyala	Normal	0.086	Gamma	0.156	Gamma	0.022
Kadugli	Normal	0.089	Normal	0.333	Normal	0.045

*Critical Value at significance level α=0.05 of the A-D test for the Normal Distribution, $C_\alpha = 0.727$
*Critical Value at significance level α=0.05 of the A-D test for the Gamma Distribution, $C_\alpha = 0.752$
*Critical Value at significance level α=0.05 of the A-D test for the Exponential Distribution, $C_\alpha = 1.321$
(-) No distribution is significant at (α=0.05) level of significance.

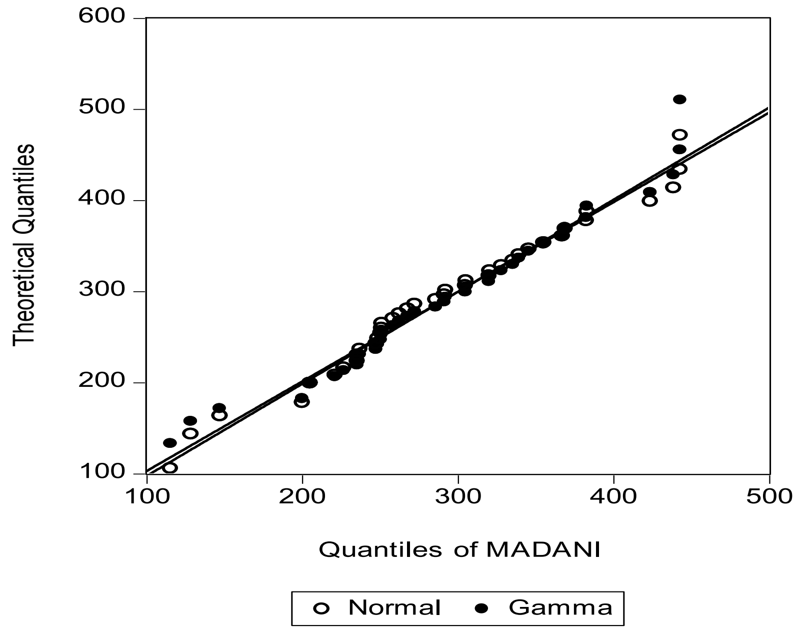

Figure 1: Quantile-Quantile plot for Wad Madani station (Gamma + Normal)

Table 4: The selected probability distributions for annual rainfall for all stations

Station	Study Period	Suitable probability distribution
Port Sudan	1971-2010	Exponential distribution
Atbara	1971-2010	Exponential distribution
Khartoum	1971-2010	Normal distribution
Kassala	1971-2010	Normal distribution
Wad Medani	1971-2010	Normal distribution
El Gedaref	1971-2010	Normal distribution
El Fasher	1971-2010	Gamma distribution
Sennar	1971-2010	Normal distribution
El Geneina	1971-2010	Normal distribution
El Obeid	1971-2010	Gamma distribution
kosti	1971-2010	Normal distribution
En Nahud	1971-2010	Gamma distribution
Nyala	1971-2010	Gamma distribution
Kadugli	1971-2010	Normal distribution