MODELING OF MONSOON RAINFALL IN PAKISTAN BASED ON KAPPA DISTRIBUTION

Ishfaq Ahmad*, Said Farooq Shah, Iram Mahmood***, Zahoor Ahmad++**

*Department of Mathematics and Statistics, International Islamic University, Islamabad, Pakistan.

** Department of Statistics Quaid-e-Azam University, Islamabad, Pakistan.

***School of Chemical and Materials Engineering, National University of Science and Technology, Islamabad, Pakistan

++Pest warning and Quality Control of Pesticides, DepalPur District Okara.

+Supported by Higher Education Commission and International Islamic University, Islamabad, Pakistan under project No: PM-IPFP/HRD/HEC/2011/2254.

* Corresponding author: ishfaq.ahmad@iiu.edu.pk Tel: 0092-51-9019733; 0092-345-4612656

ABSTRACT: *Monsoon rainfall in Pakistan is of great importance because of its needs in agriculture and power generation. In this paper, we have analyzed the random behavior of monsoon rainfall in Pakistan through 4 parameter Kappa probability distribution at 27 meteorological stations for the period 1960-2006. The parameters of this distribution have been estimated using method of L-Moments. Using these estimates we have calculated quantiles for different return periods from 2 to 500 years. The comparison of estimated quantiles with observed values of rainfall after five years is found to be in good agreement.*

Keywords*: Monsoon, Bootstrapping, Kappa Distribution, L-Moments, Quantiles*

1. INTRODUCTION

Pakistan is located from southwest to Northwest at 23-37 degree north latitude and 61-76 degree east latitude. Pakistan faces much diversified climate pattern round the year. In northern areas the temperature is as low as 25 Celsius degree and in the southern areas it as high as 55 Celsius degree. Monsoon season is experienced in many parts of world for example northern Australia, Africa, and South America. But it is strong in South Asian countries including Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan and Sri Lanka. Southwest monsoon season in Sri Lanka enters in late May just before it enters the Indian subcontinent. In India it enters in June and in Pakistan in early July and remains till the end of September. Summer monsoon season in Pakistan is of great importance for its agricultural, economic and social purposes. These rains are not only used for water needs of plants in agricultural sector but also to kill the insects by physical beating. By irrigating the fields, certain insect pests like crickets attacking the cotton seedlings and the white ants attacking cotton, sugarcane, chilies and other crops can be drowned in rain water and thus crops can be saved. Monsoon in Pakistan contributes almost 65-75 % of the total annual rainfall. Despite of some destruction in the forms of floods and droughts these rainfalls are welcomed in Pakistan. In Pakistan there are many studies which deal with rainfall data in different aspects for example, Rasul *et al* [1] carried out a diagnostic study of record heavy rain in twin cities of Pakistan as Rawalpindi and Islamabad. Karori and Zhang [2] investigated prospects of downscaling for seasonal precipitation prediction over Islamabad-Pakistan. Haroon and Rasul [3] applied the very common multivariate Principal Component Analysis (PCA) in order to identify the major modes of oscillations present in the data. Rasul et al [4] examined the heavy monsoon precipitation over the Indus plains of south Asia by non-hydrostatic numerical model MM5. But unfortunately there is not even a single study which determines underlying the probability distribution of the rainfall data and finds the extreme events after different return periods of time. In this paper we have focused on modeling the monsoon total rainfall in Pakistan across 27 different meteorological stations from 1960 to 2006 through four parameter Kappa distribution. This is the most suitable distribution for the given data set in the presence of extreme observations. We have also calculated different quantiles for different years and compared these estimated quantiles with observed values after five years. The results have been found to be in good agreement. The rest of the paper is as follows, section 2 is about the methodology about L-moments estimation and estimation of parameters of Kappa distribution from these moments. In section 3 we have applied the described methodology on Pakistan monsoon data. Section 4 presents the results and discussions.

2. Methodology

For rainfall data usually we use Generalized Extreme Value (GEV) distribution to estimate extreme events. This distribution having three parameters is considered to be as limiting form of such extreme observations. This distribution gives unsatisfactory results when we have a finite sample Winchester [5]. We have used generalized form of GEV, named Kappa distribution developed by Hosking [6] to model the monsoon rainfall of Pakistan. Kappa Distribution has four parameters and gives quite satisfactory results when GEV or any other distribution having two parameters or three parameters becomes unsatisfactory to provide such results. Following Hosking [7] we have used method of L-moments for estimation of parameters of this distribution. This method is simply based on the linear functions of expected order statistics. This method is preferable as it provides reliable and robust estimates of the parameters of the given distribution and hence reliable quantiles when we have small samples Parida [8]. Let $y_{(1)}, y_{(2)}, \dots y_{(n)}$ be the increasing order of monsoon rainfall considered as real valued random variable, then probability density function of the Kappa distribution is given by:

$$f(y) = \alpha^{-1}\left[1 - \frac{k(y-\mu)}{\alpha}\right]^{\frac{1}{k}-1}[F(y)]^{1-h} \quad (1)$$

which includes the limits as *h=0 and k=0*, where $\mu, \alpha, h \text{ and } k$ are the parameters of Kappa distribution. Further, μ, α tell about the location and scale of the distribution, while other two parameters *h* and *k* are about the shape of the distribution. In equation (1) *F(y)* is the cumulative distribution function of Kappa distribution given by:

$$F(y) = \left\{1 - h\left[1 - \frac{k(y-\mu)}{\alpha}\right)\right\}^{\frac{1}{k}} \quad (2)$$

which implicitly includes the limits as *h=0 and k=0*.

When fitting the Kappa distribution to real data, the parameters of the Kappa distribution are unknown. So these parameters are estimated through sample L-moments using simulation study. In this paper we have used method of L-moments as method of estimation. The detail description of this method is given in Hosking [7]. We describe this methodology briefly in this paper. The method of L-moments estimation is an alternative method of estimation similar to conventional method of moments. But the estimates obtained from method of L-estimation are more robust and show less bias. Let $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ be the order statistic of a random sample of size *n* from the distribution of real valued ranodom variable Y. Then according to Hosking [7] the L- moments can be expressed as below:

$$\lambda_r = r^{-1}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}E[Y_{r-k:r}] \quad r = 1,2,\dots, \quad (3)$$

Where, λ_r is the r_{th} population L-moment and is the function of expected order statistics which is denoted as:

$$E[Y_{j:r}] = \int y\{F(y)\}^{j-1}\{1 - F(y)\}^{r-j}dF(y) \quad (4)$$

λ_r, may be expressed as:

$$\lambda_r = \int_0^1 y(F)P_{r-1}^*(F)dF, \quad r = 1,2,\dots \quad (5)$$

where, $P_r^*(F) = \sum_{k=0}^{r} p_{r,k}^* F^k$ and

$p_{r,k}^* = (-1)^{r-k}\binom{r}{k}\binom{r+k}{k}$. The first four L-moments can be obtained as :

$\lambda_1 = E(Y) = \int_0^1 y(F)dF$,

$\lambda_2 = \frac{1}{2}E[Y_{2:2} - Y_{1:2}] = \int_0^1 y(F)(2F-1)d$,

$\lambda_3 = \frac{1}{3}E[Y_{3:3} - 2Y_{2:3} + Y_{1:3}] = \int_0^1 y(F)(6F^2 - 6F + 1)dF$,

$\lambda_4 = \frac{1}{4}E[Y_{4:4} - 3Y_{3:4} + Y_{2:4} - Y_{1:4}] = \int_0^1 y(F)(20F^3 - 30F^2 + 12F - 1)dF$

Normally, the higher order moments, $\lambda_r, r \geq 3$, are independent of units of measurement i.e. they are standardized. The L-moments ratio defined by Hosking [6] are given as: $\frac{\lambda_r}{\lambda_2} = \tau_r$. Analogous to conventional coefficient of variation (C.V), we have L-C.V as: $\tau = \frac{\lambda_2}{\lambda_1}$. Unlike conventional moments, we have to specify the probability distribution uniquely and restrict the mean to be finite. The population L-moments can be estimated in several ways like probability weighted moments and U-statistic. In this paper we follow the methodology of U-statistic introduced by Hoeffding [9].

As we know λ_r can be written in the form of linear expected ordered observations i.e. $E[Y_{k:k}]$. So, $\lambda_r = \sum_{k=0}^{r} p^*_{r-1,k-1} k^{-1}$ $E[Y_{k:k}] = K\int x\{F(X)\}^{k-1}, k = 1,\dots r$, Now the problem is reduced to only estimating $E[Y_{k:k}]$. It is estimated by U-statistic. U-statistic is calculated after taking all possible samples of size *k* from the overall sample of size *n*, and then taking the average of the largest observation from these all possible samples. Let $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ be the ordered sample and the estimate of $E[Y_{k:k}]$ is $k\,b_{k-1}$. The r_{th} sample L-moments can be written as:

$$l_{r+1} = \sum_{k=0}^{r} p_{r,k}^* b_{k-1} \qquad r = 0,1,2,\dots n-1,$$

Where

$$b_k = n^{-1}\sum_{i=1}^{n}\frac{(i-1)(i-2)\dots(i-k)}{(n-1)(n-2)\dots(n-k)}y_{i:n}, \quad r = 0,1,2,\dots n-1,$$

Further we can define the sample L-moments ratio as: $\tau_r = \frac{l_r}{l_2}$, which is consistent but not unbiased. The advantage of the sample L- moments is that they are not affected by the extreme values as the conventional moments. Further with help of l_1 the sample mean, and l_2 the sample variance, τ_3 the sample L-Skewness, and τ_4, the sample L-Kurtosis are useful not only in summarizing the data but also in estimation of the parameters of the Kappa distribution. Using the method of L. moments estimation to obtain the estimates of the four parameters of Kappa distribution, we need to solve four equation for four unknown. There is no explicit solution for this distribution. As in Hosking [6], Newton-Raphson algorithm may be used to find the estimates of *h* and *k*. Once the values of *h* and *k* are known, the solution of other two parameters *μ* and *α* is possible. τ_3 and τ_4 are the function of *h* and *k* only.

In order to find maximum rainfall after different recurrence periods (i.e. different quantiles), the quantile function or inverse distribution function of the Kappa distribution can be expressed in the following form:

$$y(f) = u + \alpha\{1 - (1 - F(y)h)/k\}/k \quad (6)$$

where *F(y)* is the probability of non exceedence. The probability associated with distribution function (say p) is also known as probability of non exceedence of a random variable Y of being equal to or less than a certain threshold value y_p. Generally, there is a relationship for any threshold value y_p, as: $T = 1/n(1-p)$, where *T* is the return period or recurrence intervals in years that is related to hydrological event $Y = y_p$, where y_p is also called quantile or T year event and *n* is the annual average no of occurrence of *Y* during the period in which *F(y)* is to be estimated. In the current analysis the value of *n=1*, and we are left with the following relation to find the probability of non exceedence $p = 1 - 1/T$, which is to be further used in quantile function of the Kappa distribution to calculate the different quantiles after different return periods in years. For example after return period of two year this probability will be 0.50, and after five year return period this probability will be 0.80 and so on.

RESULTS AND DISCUSSIONS:

We have analyzed Pakistan monsoon total rainfall data from 1960 to 2006 across 27 meteorological stations using Kappa distribution and L-method of estimation. The estimates obtained are more reliable and robust to outlier. The estimates of parameters of Kappa distribution on the basis of L-moments estimation method are given in Table1 along with average rainfall and the corresponding standard deviation for each station. Further we have also calculated different quantiles of monsoon rainfall for different return periods from 2 to 500 years. In Table 2 the results of quantiles for different return periods are given. These quantiles are robust at high and lower recurrence intervals. These quantiles show that some parts of Pakistan expected to face heavy monsoon rainfall in future for example in Muzafarabad, Muree, Garidupatta, Khanpur, Badin, Kakul, Kotli and Islamabad. Some recent floods in Pakistan give good evidence in support of current study.The comparison between estimated quantiles after recurrence intervals of five years with observed value of monsoon rainfall at 27 metrological stations show a close agreement. These quantiles are also very useful for planning and operation of water resources. It reveals that the modeling strategy adopted here meet the situation in real life as well. Changes in behavior of rainfall would alter the pattern of water flows and demands in different aspects as soil moisture and water reserves. If in any monsoon season there is heavy rain, may cause of the extreme events, such as floods, droughts, and thunderstorms. Better understanding about the behavior of monsoon rainfall can lead us to better solution of the problems i.e. better understanding of the problems associated with floods, droughts, and the availability of water for various uses with respect to future climate scenarios. In short, the modeling of rainfall pattern needs more attention as it influences the different aspects of human life including operation and policies of water management. Rainfall modeling plays a vital role in water management and any unexpected changes would directly influence the water resources.

Table1: Basic Statistics and estimates of the parameters of Kappa Distribution across different meteorological stations.

S. #	Name of the Station	Average (mm)	Standard Deviation (mm)	Coefficient of Skewness	Parameters of the Kappa Distribution			
					μ	α	k	h
1	Jacobabad	71.62	87.51	1.62	109.73	155.3528	0.1549688	2.1092
2	Karachi	146.87	148.89	1.14	181.374	368.8845	0.4006777	1.718831
3	Quetta	24.4	41.83	-2.51	137.019	110.0653	0.1905461	4.140041
4	Rohri	63.33	90.41	2.31	65.4400	92.75704	-0.072095	1.940626
5	Zhob	115.56	61.64	2.54	64.8215	79.87446	0.2556424	0.5320531
6	Droash	59.12	27.99	2.26	21.0827	65.4277	0.6428356	0.9023056
7	Islamabad	382.82	169.6	-1.34	668.1607	182.7157	0.1779115	-0.353705
8	Kakul	352.58	114.43	1.19	564.5416	110.7181	0.1435548	-0.164551
9	Kotli	652.83	199.04	2.48	531.5315	243.0463	0.2843761	0.3262781
10	Multan	116.1	79.5	1.7	256.2749	212.1677	0.0991215	0.6661604
11	Lahore	422.1	190.3	1.05	64.61431	61.94675	-0.0677115	0.331812
12	Peshawar	131.86	88.93	1.23	64.61431	61.94675	-0.0677115	0.331812
13	Hyderabad	132.67	134.69	1.958	373.289	622.6996	0.7252341	2.190151
14	Sargodha	257.73	114.55	2.74	155.1131	151.9823	0.2454089	0.6253857
15	Badin	186.39	179.66	1.84	27.01739	185.0359	0.04238833	00.7816652
16	D.I.Khan	143.63	71.3	-2.28	102.0352	83.96129	0.2905808	0.2189151
17	Resaalpur	305.29	143.19	2.82	259.4834	98.58197	-0.0334326	-0.3391586
18	Sialkot	661.91	306.11	2.98	423.4953	321.9979	0.09352012	0.5244777
19	Sibbi	82.3	59.21	2.74	71.37893	108.4875	0.1588332	0.4070536
20	Astor	69.35	40.49	1.39	56.43396	21.21701	-0.2196389	-0.5679996
21	Chalas	28.89	22.28	1.52	16.11063	17.17428	-0.07206	0.251185
22	Garidupatta	608.21	167.51	2.21	577.9278	121.5508	0.09493055	-0.4141473
23	Gilgit	36.96	20.19	1.02	25.47211	19.1018	0.07699357	0.2443935
24	Khanpur	64.85	66.63	-1.23	-48.998	128.3066	0.3080928	1.436468
25	Muree	802.9	193.4	2.57	692.6443	201.96	0.175164	0.2385414
26	Muzafarabad	670.19	158	2.4	564.4013	203.3874	0.2575603	0.3796566
27	Skardu	30.82	21.19	1.98	20.33532	16.99983	-0.0123311	0.07975316

ACKNOWLEDGMENT:

I am thankful to Mr. Ayaz Khan senior meteorologist at Pakistan metrological department in provision of the required data. I am also thankful to Higher Education Commission (HEC) Pakistan for providing me the grant to conduct this research under the project No: PM-IPFP/HRD/HEC/2011/2254.

Table 2: Estimates of the Quantiles for different return periods from 2 to 500 years at different stations.

S. #	Name of the Station	Estimated Quantiles based on Kappa Distribution for monsoon rainfall in cm for different recurrence intervals								
		2	**5**	**10**	**20**	**50**	**100**	**200**	**500**	Actual Values after 5 year
1	Jacobabad	35.5	125.5	197.3	265.3	346.9	402.1	451.8	510.1	122.3
2	Karachi	98.3	270.7	378.8	464.1	547.8	594.0	629.2	662.9	272.4
3	Quetta	4.8	40.4	79.1	119.1	168.1	201.1	230.4	263.9	44.2
4	Rohri	27.2	103.2	172.2	247.4	354.9	441.8	533.3	661.9	102.3
5	Zhob	105.4	167.5	202.7	231.5	262.2	280.9	296.6	313.4	160.4
6	Droash	56.4	86.4	99.6	108.1	114.6	117.5	119.4	120.9	88.1
7	***Islamabad***	***711.3***	***903.1***	***1004.7***	***1088.7***	***1181.8***	***1241.9***	***1294.8***	***1355.1***	***902.4***
8	***Kakul***	***597.9***	***712.3***	***776.7***	***831.9***	***895.2***	***937.2***	***975.2***	***1019.7***	***712.1***
9	***Kotli***	***640.0***	***834.0***	***937.7***	***1019.8***	***1104.6***	***1155.2***	***1196.6***	***1240.2***	***831.1***
10	Multan	227.5	265.3	390.1	704.8	903.8	940.5	950.6	1040.6	270.1
11	Lahore	94.7	164.9	216.4	269.1	341.5	399.1	459.3	543.2	163.2
12	Peshawar	113.9	184.3	237.9	297.7	390.5	474.3	572.6	729.6	180.1
13	Hyderabad	78.9	241.2	330.6	389.7	435.5	455.1	466.9	475.8	240.3
14	Sargodha	236.7	353.0	420.8	476.8	537.1	574.3	605.6	639.7	357.4
15	***Badin***	***141.2***	***510.8***	***531.1***	***546.7***	***693.7***	***800.9***	***905.0***	***1037.9***	***503.6***
16	D.I.khan	136.8	205.4	241.2	269.3	298.1	315.1	328.9	343.5	202.1
17	Resaalpur	283.9	407.2	487.9	566.4	669.9	749.5	830.5	940.3	407.1
18	Sialkot	393.9	490.3	698.1	792.8	812.4	889.8	1068.6	114.0	376.7
19	Sibbi	123.9	220.0	278.3	328.9	387.1	425.6	459.9	499.9	130.3

20	Astor	59.9	92.3	117.2	144.7	187.2	224.9	268.9	337.9	66.2
21	Chilas	24.0	43.9	58.3	73.1	93.6	109.8	126.9	150.7	58.1
22	***Garidupatta***	***604.3***	***742.9***	***822.1***	***891.6***	***973.9***	***1030.8***	***1083.8***	***1148.5***	***798.8***
23	Gilgit	33.9	52.9	65.1	76.3	89.9	99.5	108.6	119.8	37.1
24	***Khanpur***	***244.3***	***417.4***	***464.0***	***502.6***	***642.9***	***866.7***	***986.1***	***1006.1***	***49.2***
25	***Muree***	***241.5***	***566.8***	***449.7***	***529.0***	***631.2***	***707.4***	***782.9***	***882.2***	***543.8***
26	***Muzafarabad***	***658.9***	***823.2***	***914.0***	***987.5***	***1065.3***	***1112.7***	***1152.3***	***1194.7***	***770.1***
27	Skardu	27.1	46.2	59.2	71.8	88.3	100.8	113.4	130.1	33.4

REFERENCES:

[1] G. Rasul, Q.Z. Choudhry, Z. Sinxiong, Z .Qingcun: A diagnostic study of heavy rain in twin cities Islamabad and Rawalpindi. *Advances in atmospheric sciences,* **21**(6), 976-988, 2004.

[2] M.A.Karori, P. Zhang: Downscaling NCC CGCM Output for Seasonal Precipitation Prediction over Islamabad-Pakistan. *Pakistan Journal of Meteorology*, **4** (8), 59-72, 2008.

[3] M.A. Haroon, G. Rasul: Principal Component Analysis of Summer Rainfall and Outgoing Long-Wave Radiation over Pakistan, *Pakistan Journal of Meteorology*, **5** (10), 109-114, 2009.

[4] G.Rasul, Q.Z. Chaudhry, A.Mahmood: Numerical Simulation of Heavy Rainfall Case in South Asia, *Pakistan Journal of Meteorology*, **6** (11).21-36, 2009.

[5] C.Winchester: On estimation of the four-parameter Kappa distribution, Master of science dissertation,*Dalhousie university Halifax, Nova Scotta, Canada*, 2000.

[6] J.R.M. Hosking: The four-parameter Kappa distribution, *IBM J. Res. Development* 6, **38**(3), 251–258, 1994.

[7] J.R.M.Hosking: L-Moments: analysis and estimation of distribution using linear combination of order statistics, *Journal of Royal Statistical Society*, **52**(1), 105–124, 1990.

[8] B.P.Parida, R.K. Kachroo, D.B. Srestha: Regional flood frequency analysis of Mahi-Sabarmati basin (subzone 3a) using index flood procedure with L-Moments', International. *Journal of Water Resources Management*, **12**(1), 1–12, 1999.

[9] W. Hoeffding: A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*,**19**, 293-325, 1948.