The Nucleus **49**, No. 2 (2012) 107-114

The Nucleus
A Quarterly Scientific Journal of Pakistan Atomic Energy Commission
NCLEAM, ISSN 0029-5698

A PROBABILISTIC APPROACH FOR ESTIMATING RETURN PERIODS OF EXTREME ANNUAL RAINFALL IN DIFFERENT CITIES OF KHYBER PAKHTUNKHWA (KPK), PAKISTAN

*M. ALI and M.J. IQBAL[1]

Mathematical Sciences Research Centre, Federal Urdu University of Arts, Sciences and Technology, Karachi, Pakistan

[1]Institute of Space and Planetary Astrophysics, University of Karachi, Karachi, Pakistan

(Received April 13, 2012 and accepted in revised form June 14, 2012)

Northern part of Pakistan had experienced heavy rainfall in 2010 which caused flooding in Pakistan. Statistical distributions are employed to analyze extremes of annual rainfall of different cities of KPK. Gumbel maximum and GEV distribution are used to calculate return period of extreme rainfall in different cities of Khyber Pakhtunkhwa (KPK). The analysis shows that different cities of KPK have 20-years return period for receiving more than 100 mm daily rainfall. While they have 50-years return period for receiving more than 120 mm daily rainfall.

Keywords: Return levels, Extreme annual rainfall, General extreme value distribution, Gumbel maximum distribution.

1. Introduction

World is frequently witnessing extreme events due to global warming. Some of these events are also potentially destructive. The major extreme weather events are heat waves, cold waves, fog, snowstorms, hailstorms, thunderstorms, cyclones, flood and heavy rains etc. These events have also socio-economic impact. During extreme rainfall in 2010, as reported by the press, more than 1,500 people have been killed and hundreds of thousands affected by flood in Pakistan. The flood waters have washed away millions of hectares of crops, submerged villages and destroyed roads and bridges. Diseases also spread among Pakistani flood victims and there have been reports that dams in the south may also burst. We could minimize the losses with the appropriate flood forecasting and improving the river structures in Pakistan.

The main cause of the current flood in Pakistan was heavy rainfall. Therefore, this study proposes to analyze the extreme rainfalls and its variability over KPK [1]. Two approaches have been suggested to assess rainfall extremes in 2003 [2]. The first approach uses a percentile or quantile method to assess extreme rainfall [2,3]. In this approach, daily rainfall records are sorted and classes defined to contain a certain percentage of the total number of rainfall events for a season or month. Each of the classes contains an equal amount of total rainfall and can, therefore, be thought of as amount quantiles. The second approach uses statistical distributions to define extremes with given return periods on an annual basis [4,5]. In this method, evaluation of the magnitude of long return-period rainfall measures involves fitting an extreme value distribution to the annual maximum (AM) series. This method produces return period estimates that are easily understood and can be used readily for design purposes. Litrature also shows that log-Pearson type III analysis, generalized extreme value (GEV) distribution and Gumbel distribution have been frequently used to analyze extreme events [6]. This study employs the second approach and hence probability density function (PDF) is fitted to annual maximum of daily rainfall data of different cities of KPK for estimating the magnitude of different return periods and return levels.

2. Exploration of Daily Rain Data

This study uses daily rainfall of four major cities of KPK (D.I. Khan, Parachinar, Peshawar and Dir) for the period from January 1981 to December 2010. Data were procured from Pakistan Meteorological Department, Karachi.

∗ Corresponding author : m.alishigri@gmail.com

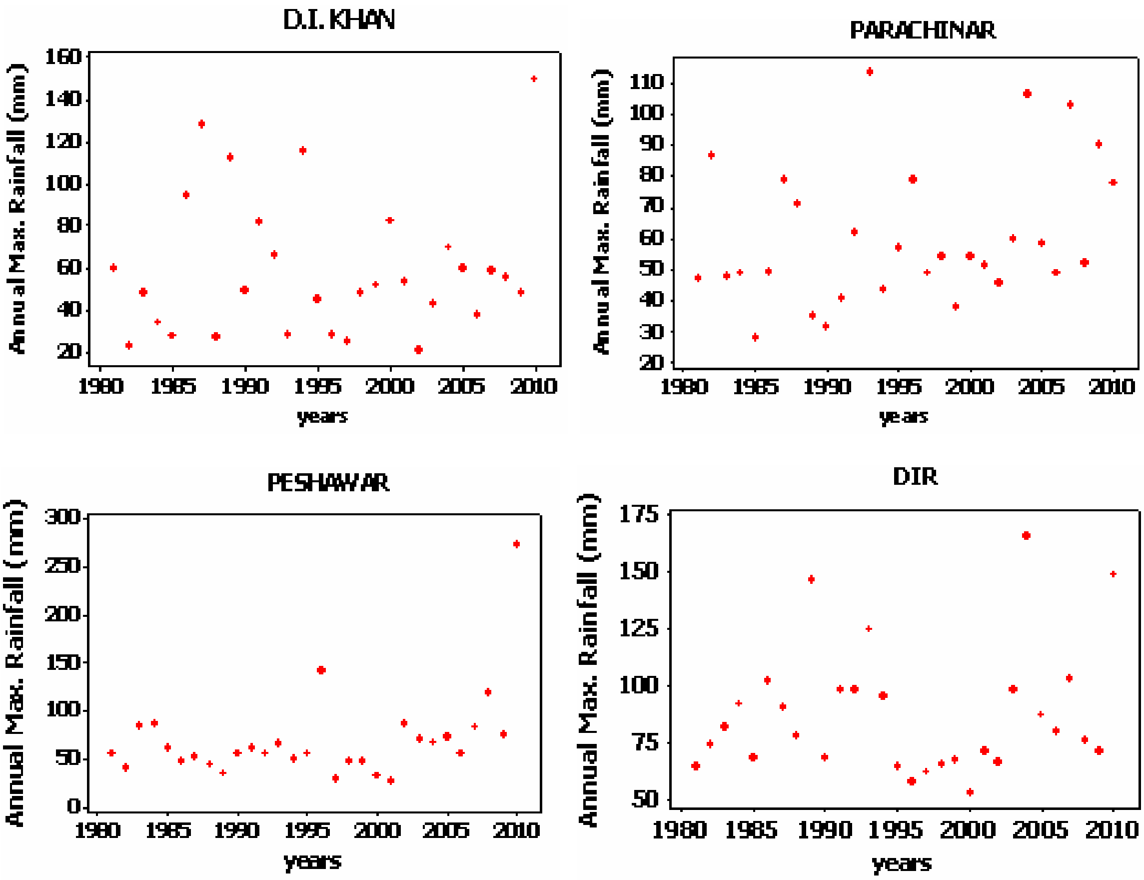


Figure 1. Annual maximum rainfall values recorded in different cities of KPK.

The annual maximum rainfall of above said stations are plotted in Figure 1. This figure depicts that that there is no trend present in them. For instance, D. I. Khan has eleven values above the normal (i.e. 59.3 mm) during the study period and the annual maximum value of rainfall (150 mm) occurred on 4^{th} August, 2010 while there are also remarkably high values more than 100 mm (i.e.128.5 mm, 112.5 mm and 116 mm, occurred on 21^{st} March, 1987, 23^{th} August 1989 and 4^{th} July 1994 respectively. Parachinar station has nine values above normal (60.28 mm) during the study period and the annual maximum rainfall occurred here is 113.6 mm, on 11^{th} March 1993, while there are two another higher values more than 100mm per day (i.e. 106.5 mm and 103 mm, happened on 21^{st} August, 2004 and 31^{st} March 2007 respectively). As for Peshawar, there are ten values higher than normal value (69.65 mm) during the study period and the highest value is 274 mm occurred on 29^{th} July 2010, while 2^{nd} and 3^{rd} ^{h} higher values are 142 mm and 119 mm, occurred on 3^{rd} October, 1996 and 5^{th}April, 2008 respectively. Similarly, annual mximum rainfall plot of Dir shows the high frequency of extreme rainfall i.e. there are fourteen values higher than normal value (87.2 mm) and the highest value is 166 mm occurred on 10^{th} October, 2004, while 2^{nd} ,3^{rd} and 4^{th} higher values are 149 mm, 147 mm and 125 mm occurred on 29^{th} July, 2010, 31^{st} July, 1989 and 12^{th} March, 1993 respectively.

In order to foresee the seasonal variation, month wise daily rain data are plotted in the form of box plot figure. 2. To avoid the compression of the figure due to the large number of near-zero observations, Figure 2 is drawn only for excesses over a threshold of 10 mm. This figure illustrates that there are two seasons of rainfall in KPK viz., summer monsoon season from July to September and winter season from December to March.

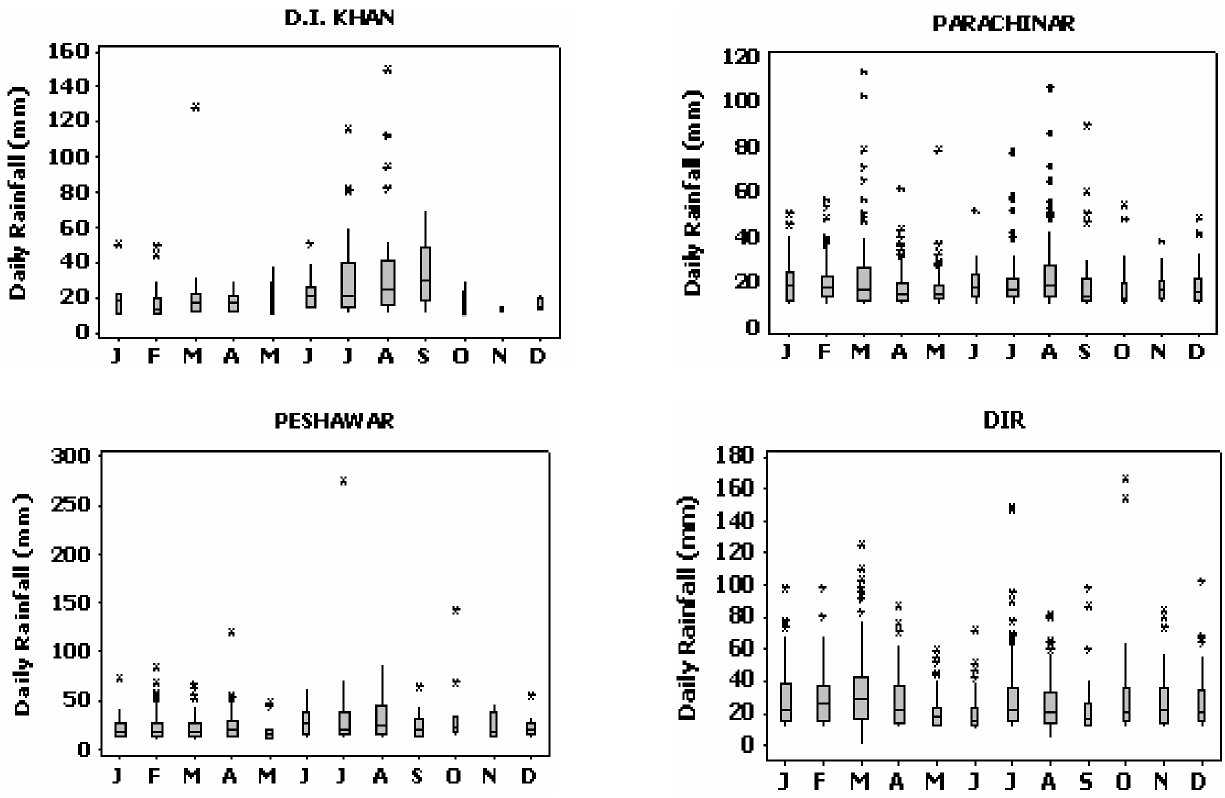


Figure 2. Monthly box plots of exceedances of 10mm threshold for different cities of KPK.

3. Extreme Value Distribution Fit

We represent the daily rainfall data by $X_1, X_2, \ldots$, the classical model for extremes is obtained by studying the behavior of $M_n = \max\{X_1, X_2, \ldots X_n\}$ for large values of n. With $n = 365$, M_n corresponds naturally to the annual maximum. Asymptotic considerations suggest that the distribution of M_n should be approximately that of a member of the generalized extreme value (GEV) family [7], having probability density function (PDF)

$$f(x) = \frac{1}{\beta}[1+Z]^{1-\frac{1}{k}} \exp\{-[1+Z]\}, \quad 1+Z > 0 \qquad (1)$$

Where

$$Z = \frac{k(x-\xi}{\beta} \qquad (2)$$

Here there are three parameters: a location (or shift) parameter ξ a scale parameter β, and a shape parameter k. Now equation (1) can be integrated analytically, yielding the CDF

$$F(x) = e^{\{-[1+Z]^{-\frac{1}{k}}\}} \qquad (3)$$

This CDF can be inverted to yield an explicit formula for the quantile function.

$$X = F^{-1}(p) = \xi + \frac{\beta}{k}\left\{[-\ln(P)]^{-k} - 1\right\} \qquad (4)$$

Here P = F(x) is the cumulative probability.

The GEV distribution is usually fit using either the method of maximum likelihood or a method known as L-moments which is to be preferred for small data samples [8], that is used frequently in hydrological applications [9]. Maximum likelihood methods can be adapted easily to include effects of covariates, or additional influences; for example, the possibility that one or more of the distribution parameters may have a trend due to climate changes (Kats et al.2002; Smith1989; Zhang et al. 2004). For moderate and large sample sizes the results of the two parameter estimation methods are usually similar. There are three special cases of the GEV distribution depending on the value of the shape parameter k. The most common one is the Extreme Value type-I distribution in which k = 0, which is also called as Gumbel type or just Gumbel distribution. This is an unbounded distribution i.e. defined on the entire real axes. These are distributions of extreme order statistics for a distribution of N-elements Xi. Gumbel's focus was primarily on application of extreme value theory to engineering problems in particularly modeling of meteorological phenomena such as annual flood flows. The probability density function of Gumbel distribution is:

$$F(x) = \frac{1}{\beta} e^{\{-e^{-Z^*} - Z^*\}} \tag{5}$$

where

$$Z^* = \left(\frac{x-\xi}{\beta}\right) \tag{6}$$

ξ is location parameter and β is scale parameter(β >0). The shape of Gumbel distribution does not depend on the distribution parameters. The Gumbel distribution is so frequently used to represent the statistics of extremes that it is sometimes called "the" extreme value distribution. The Gumbel PDF is skewed to the right, and exhibits its maximum at x = ξ. Gumbel distribution probabilities can be obtained from the cumulative distribution function (CDF)

$$F(x) = e^{-e^{-Z^*}} \tag{7}$$

This CDF can be inverted to yield an explicit formula for the quantile function.

$$X = F^{-1}(p) = \xi - \beta \ln\{-\ln(P)\} \tag{8}$$

Here P = F(x) is the cumulative probability.

Gumbel distribution parameter can be estimated through maximum likelihood or L-moments, as described earlier for the more general case of the GEV, but the simplest way to fit this distribution is to use the method of moments. The moments estimators for the Gumbel distribution parameters are computed using the sample mean and standard deviation. Gumbel parameters can be estimated as;

$$\hat{\beta} = \frac{s\sqrt{6}}{\pi} \tag{9}$$

and

$$\hat{\xi} = \overline{X} - \gamma\hat{\beta} \tag{10}$$

where γ = 0.57721… is Euler's constant.

3.1. Best Fitted Distribution

As stated above the estimation of the magnitude of long return-period rainfall events involves fitting an extreme value distribution (viz. Gumbel Maximum and General Extreme Value — GEV) to the extreme annual rainfall of four cities of Khyber Pakhtonkwa (KPK). To determine the best fitted distribution, we first make histogram with fitted probability density curves as shown in Figures 3. Histogram plots show that the GEV distribution covers more area than Gumbel Maximum distribution. Thus Histogram also suggest that GEV is the best fitted distribution for the extreme annual rainfall of different cities of KPK. We now employ chi square test for the goodness of fit test. Table 1 summarizes the goodness fit test for the extreme annual rainfall of D.I. Khan, Parachinar, Peshawar and Dir. As for extreme annual rainfall of D.I.Khan, the estimated χ^2 = 0.35829 for GEV distribution. Under the negative hull hypothesis, the statistic is drawn from a χ^2 distribution with degree of freedom ν = 5 − 3 − 1 = 1 for GEV distribution. Referring to ν = 1 of row chi-square Table, estimated χ^2 = 0.35829 is smaller than the 95th percentile value of 3.841, so the null hypothesis that data of extreme annual rainfall of KPK have been drawn from GEV distribution would not be rejected even at 5% level. Thus, data of extreme annual rainfall of KPK follow GEV distribution. Similarly, Table 1 also shows that estimated chi squared values for other three stations, which are smaller than 95th percentile value of 3.841 (computed values of chi squared from the chi squared table) with 5% significance level. Thus data of extreme annual rainfall of four stations have been drawn from GEV distribution.

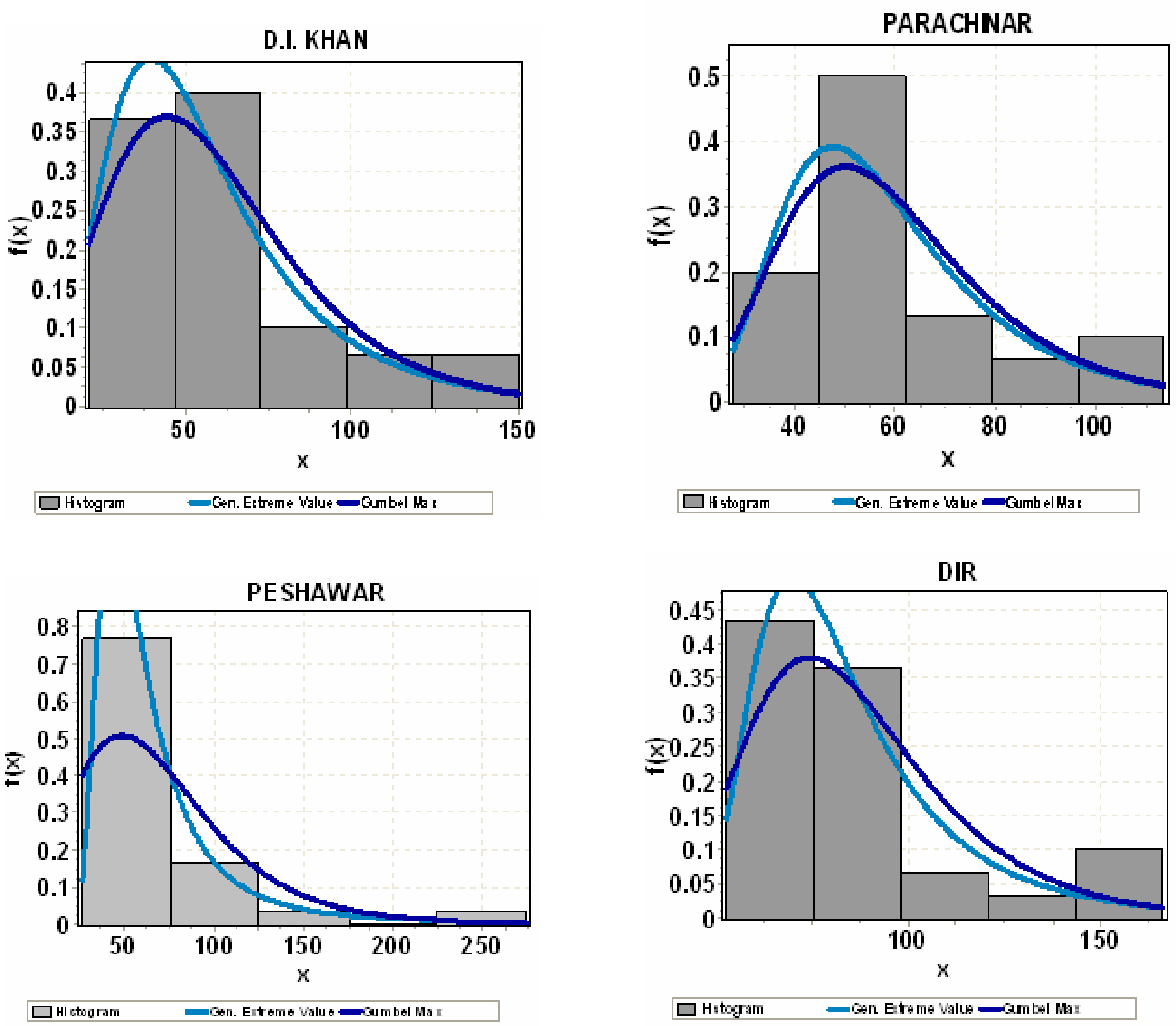


Figure 3. Probabilistic density function fitted to extreme annual rainfall of different cities of KPK.

Now we apply goodness of fit test for Gumbel Maximum distribution. The value of $\chi 2$ is 5.99 at 5% level, with degree of freedom $\nu = 5 - 2 - 1 = 2$. Thus, Table 1 also depicts that the estimated values of chi squared for four stations are smaller than 95th percentile value of 3.9 (computed values of chi squared from the chi squared table) with 5% significance level. Thus data of extreme annual rainfall of four stations have been drawn from Gumbel Maximum distribution. Since, estimated values of chi squared for GEV distribution are smaller than the estimated values of chi squared for Gumbel Maximum distribution. Therefore, we assume that GEV distribution is the best fitted distribution.

Table 1: Summary of goodness of fit test for extreme rainfall of D.I. Khan, Parachinar, Peshawar and Dir.

No.	Stations	Chi-Squared	
		General Extreme Value Distribution	Gumbel Maximum Distribution
1	D.I. Khan	0.35829	1.0056
2	Parachinar	1.3309	2.6784
3	Peshawar	1.0611	3.0606
4	Dir	0.36066	0.66181

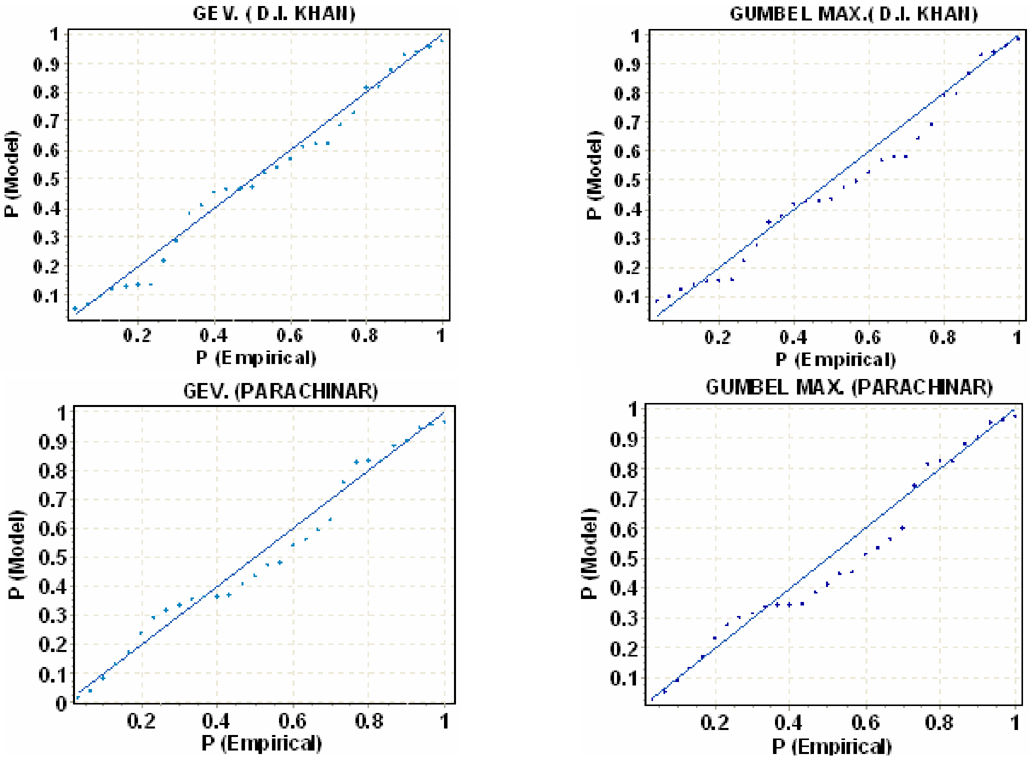
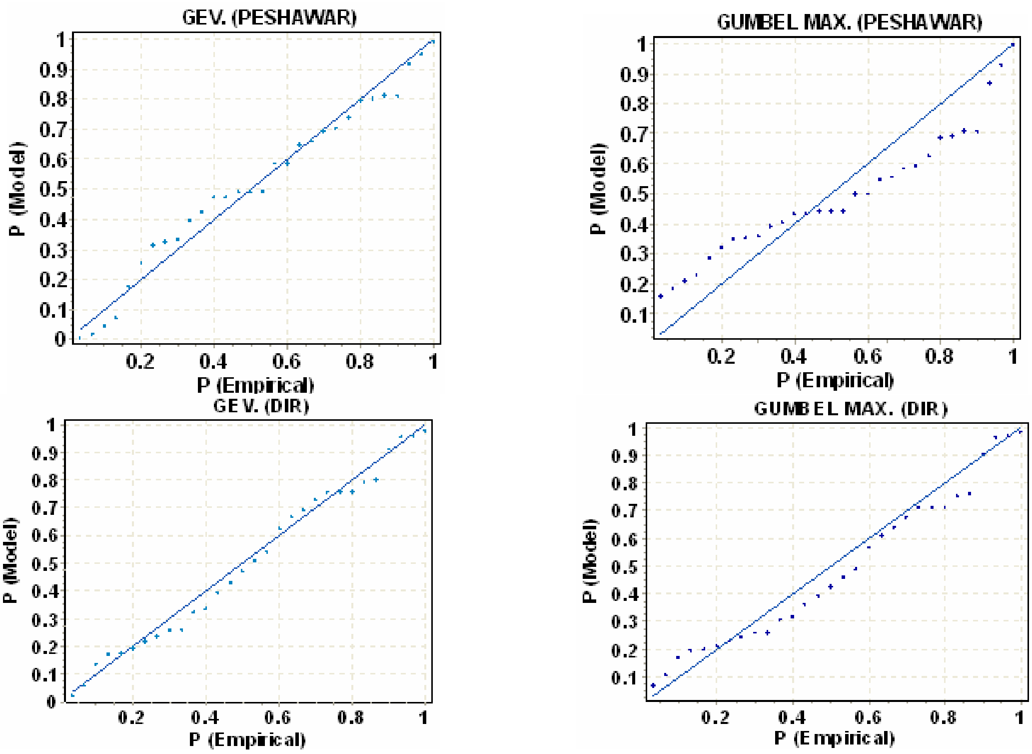


Figure 4. P-P Plot of Gumbel Maximum and GEV Distribution of Different Cities of KPK.



Figures 4. (Continued).

The probability-probability (p-p) plot is a graph of the empirical CDF values plotted against the theoretical CDF values. It is used to determine how well a specific distribution fits to the observed data. This plot will be approximately linear if the specified theoretical distribution is the correct model. To illustrate the above goodness of fit test result we have drawn p-p plot for extreme annual rainfall of four mentioned cities in Figure 4. This figure also shows that the deviation of observed data points from theoretical CDF values is comparatively more in Gumbel Maximum distribution while it is lesser in General Extreme Value. So the lesser the deviation values the more fitted will be that distribution. Hence, GEV distribution is the best fitted distribution for our data.

4. Return Period Estimates

The result of an extreme value analysis is often simply a summary of quantiles corresponding to large cumulative probabilities, for example the event with an annual probability of 0.01 of being exceeded. Unless *n* is rather large, direct estimation of these extreme quantiles will not be possible and a well-fitting extreme-value distribution provides a reasonable and objective way to extrapolate to probabilities that may be substantially larger than 1 – 1/*n*. Often these extreme probabilities are expressed as average return periods,

$$R(x)\,\frac{1}{\omega\,[1-F(x)]} \tag{11}$$

Return period R(x) associated with a quantile *X* typically is interpreted to be the average time between occurrence of events of that magnitude or greater. The return period is a function of the CDF evaluated at x, and the average sampling frequency ω, for annual maximum data ω = 1/year, so the event x corresponding to the cumulative probability F(x) = 0.99 will have probability 1 – F(x) of being exceeded in any given year. This value of x would be associated with a return period of 100 years, and would he called the 100-year event.

4.1. Return Period for D.I.Khan:

The analysis shows that annual maximum value of rainfall (150 mm) occurred on 4th August, 2010, throughout the period 1981 to 2010 which also throws light on the current situation of 2010's flood. Now we want to know that after how many years there is a chance of getting the above value or greater amount of rainfall would be reoccurred. So the maximum-likelihood fit of the GEV distribution to the annual maximum daily precipitation data of D.I. Khan yielded the parameter estimates ξ = 42.784, β = 21.63 and k = 0.16096.

From equation (2)

$$Z=\frac{k\,(x-\xi)}{\beta}$$

Z = 0.8,

Now we find the cumulative probability by using Eq. (3)

$$F(X) = P(X\leq 150) \;=\; e^{\{-[1+Z]^{-\frac{1}{k}}\}}$$

$F(x) = \exp\{-[1+0.8]^{-6.2}$

F(X) = exp{-0.026}

F(X) = 0.97

Now Eq. (11) yields

$$R(x)=\frac{1}{\omega[1-F(x)]}$$

R(X) = 33 years

It means that in coming 33 years there is a chance to occur rainfall amount equals to 150mm or more in a day.

4.2. Return Levels For Different Return Periods

Now we want to compute return period for *T* = 2 years, *p* = 0.5 and F(*X*) = *P* = 1 – *p* = 0.5. Here *p* is probability while *P* is the cumulative probability i.e. F(*X*) . So by using Eq. (4), we estimate 50.95 mm of rainfall. Hence this result shows that in coming two years there is a chance of occurring 50.95 mm rain at D.I.Khan station. For D.I.Khan station, we also calculate the return levels 79.48 mm, 101.44 mm, 125.16 mm, 144.13 mm, 160.23 mm and 190.18 mm respectively for 5, 10, 20,30, 50, 100 years. We summarize our calculations of return levels against different return period for four cites in KPK in Table 2. Our calculations show that different cities of KPK have 20-years return period for receiving more than 100 mm daily rainfall. While they have 50-years return period for receiving more than 120 mm daily rainfall. This asks for construction of new dams in Pakistan.

Table 2. Return Levels for different return periods.

Return period (years)	D.I. Khan mm	Parachinar mm	Peshawar mm	Dir mm
2	50.95	55.26	56.64	79.63
5	79.48	75.34	85.10	103.1
10	101.44	89.92	111.44	121.85
20	125.16	104.95	144.34	142.71
30	144.13	116.52	174.07	159.81
50	160.23	126.06	201.74	174.61
100	190.18	143.19	259.28	202.76

5. Conclusions

This study employs probabilistic approach for estimating return period of extremes of annual rainfall of different cities of KPK (D.I. Khan, Parachinar, Peshawar and Dir). We employ daily rainfall of four major cities of KPK for the period from January 1981 to December 2010. The time series of annual maximum data of above said stations have no obvious trend. Box-plot of monthly daily data shows the variability of rainfall in different months and also suggest that the region under study has two seasons viz. summer monsoon season from July to September and winter season from December to March.

Our calculations suggest that General Extreme Value Distribution is the best fitted distribution for extreme annual rainfall of different cities of Khyber Pakhtunkhwa. Our calculations also demonstrate that different cities of KPK have 20-years return period for receiving more than 100 mm daily rainfall. While they have 50-years return period for receiving more than 120 mm daily rainfall. This asks for the appropriate flood forecasting, improving the river structure and construction of new dams in Pakistan.

References

[1] M. J. Iqbal and M. Ali, Arab J. Geosci. DOI 10.1007/s12517-012-0548-z First Published Online (2012).

[2] K. Fowler, Int. J. Climatol. **23** (2003) 1313.

[3] T.R. Karl and R.W. Knight. Bulletin of the American Meteorological Society **79** (1998) 231.

[4] K.J. Hennesey, J.M. Gregory and J.F. Mitchell, Climate Dynamics **13** (1997) 667.

[5] K. McGuffie, A. Henderson-Sellers, N. Holbrook, Z. Kothavala, O. Balachova and J. Hoekstra, Int. J. of Climatol. **19** (1999) 1.

[6] B. Khan, M.M. Iqbal and M.A.K. Yousufzai, Flood Risk Assessment of River Indus of Pakistan 4:115–122. doi:10.1007/s12517-009-0110-9, (2011)

[7] T.J. Osborn, M. Hulme, P.D. Jones and T.A. Basnett, Int. J. of Climatol. **20** (2000) 347.

[8] M.R. Leadbetter, G. Lindgren and H. Rootzen, Extremes and related properties of random sequences and processes. Springer-Verlag, New York (1983)

[9] J.R.M. Hosking, J. Roy. Statist. Soc. **52** (1990) 1.

[10] J.R. Stedinger, R.M. Vogel and E. Foufoula-Georgiou, Frequency Analysis of Extreme Events, Chapter 18, Handbook of Hydrology, McGraw-Hill Book Company, R.M. David, Editor in Chief (1993).

[11] Zhang et al, *J. Geophys.* Res. **110** (2005).

[12] R.W. Katz, G.S. Brush and M.B. Parlange (2005), Statistics of Extremes: Modeling Ecological Disturbances, Ecology **86** (2005) 1124.