The Nucleus **49**, No. 3 (2012) 221-230

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

ESTIMATION OF RETURN LEVELS AGAINST DIFFERENT RETURN PERIODS OF EXTREME ANNUAL RAINFALL OVER BALUCHISTAN

*M. ALI, M. J. IQBAL[1] and B. JAN

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

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

(Received August 31, 2012 and accepted in revised form September 12, 2012)

Unprecedented heavy monsoon rainfall began in the last week of July 2010 in the Northern part of our country, causes floods in Baluchistan and Sindh. As the high frequency rainfall events are a significant cause of current severe flooding in Pakistan and any fluctuation in the level of such events may cause huge economic losses as well as social problem, urban structures (i.e. dams, urban drainage systems and flood). Statistical distributions are used to identify extremes of annual rainfall of different cities of Baluchistan (Quetta, Sibbi, Khuzdar, Lasbella, Dalbandin and Pasni) with their return periods. Analysis predicts that Gumbel Max.(GM) Distribution is the best fitted distribution for Sibbi and Lasbella while the GEV distribution is the best fitted for Quetta, Khuzdar, Dalbandin and Pasni. The analysis also suggests that different cities of Baluchistan have 30-years return period for getting more than 90 mm average daily rainfall while they have 100-years return period for receiving more than 118 mm daily rainfall. This suggests for suitable flood forecasting and improving the river structure in Baluchistan, Pakistan.

Keywords : Rainfall, Statistical distributions, Gumbel maximum distribution, Return Periods, Baluchistan

1. Introduction

Major extreme weather events over Pakistan have changed in frequency and intensity during the past decades due to significant increase in global surface temperature. The frequency and intensity of extreme weather events have been expected to increase or decrease and the associated meteorological disasters to be aggravated or lessened, as a result of global warming. The infrequent or rare condition of weather intensity in a locality is called extreme weather event [1]. Weather events or climatic events are being extreme can fluctuate from time to time and place to place. Extreme weather and climate events also include local severe convective phenomena such as, cold waves, fog, snowstorms, hailstorms, thunderstorms, cyclones, flood and heavy rains etc. Floods and other natural disasters may result in loss of infrastructure, energy insecurity, political, socio-economic and social life instability and decline of natural ecosystems in the country. According to press survey during the horrible summer monsoon rainfall in 2010, about 1,500 deaths and hundreds of thousands of people trapped by flooding triggered in Pakistan, millions of hectares of crops, underwater villages and destroyed roads, bridges washed away, there was a threat of damaging dams in the south and different diseases spread over flood affected areas of Pakistan. We can minimize these losses by having an appropriate flood forecasting and improving the river structures in Pakistan. Heavy rainfall was the major cause of the recent flood in Pakistan. Therefore, it is important to study and estimate the extreme rainfalls and its variability. To judge rainfall extremes two methods have been found by [2], i.e. (i) A percentile or quantile method to find extreme rainfall [3,4] where, 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 it is, therefore, to be thought as amount quantiles (ii) Statistical distributions to define extremes with given return periods on an annual basis [5-7] where estimation of the magnitude of long return-period rainfall events involves fitting an extreme value distribution to the annual maxima (AM) series. This method produces return period estimates that are easily understood and can be used readily for design purposes.

In the current work, probability distribution function is fitted to annual maximum of daily rainfall data of different cities of Baluchistan, Pakistan to estimate the magnitude corresponding to return periods and the appropriate distribution models

* Corresponding author : m.alishigri@googlemail.com

that represents a best data selection.

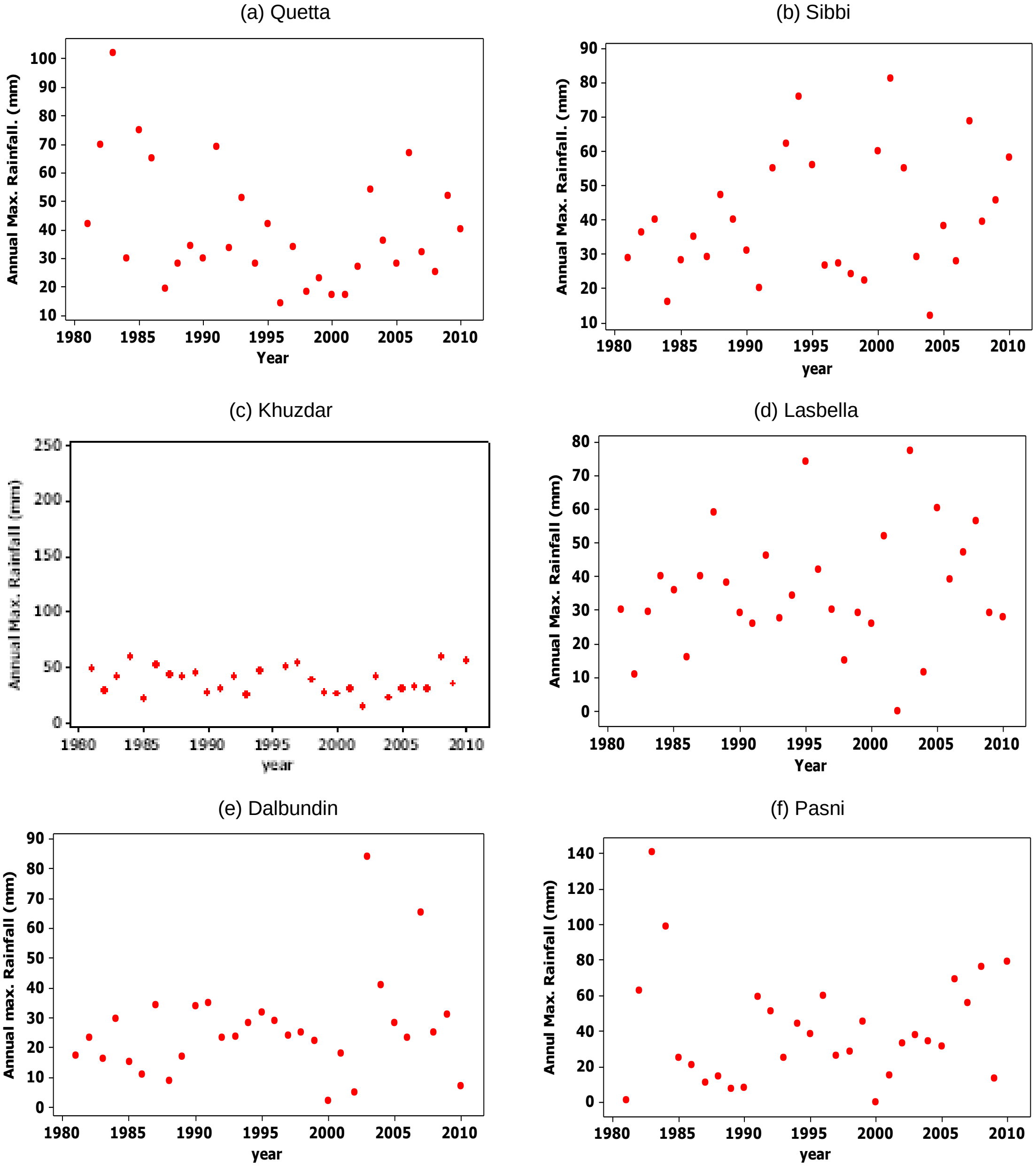


Figure 1. Annual maximum rainfall values recorded in different cities of Baluchistan.

2. Data Description

Available data relates records of daily rainfall of six major cities of Baluchistan (Quetta, Sibbi, Khuzdar, Lasbella, Dalbandin and Pasni). The data were recorded by Pakistan Meteorological Department. Daily records are available for the period January 1981 to December 2010.

Figure 1 shows scattered plots of the annual maximum data of above mention stations. These figures show that no obvious trend is present in them. According to Figure 1a, there are eleven values above normal (40mm) and the highest value is 102mm occurred on 3rd Aug, 1983. Figure 1b shows that there are eleven values above normal (40.5mm) and the annual maximum rainfall occurred over Sibbi is 81 mm, on 25th July, 2001 while this city receives 76 mm rainfall on 6th Sept.,1994. Figure 1c depicts eleven extreme values and here the normal is (39mm) over

Khuzdar station in Baluchistan and highest value (60.4mm) per day occurred 7th Dec., 2008. From Figure 1d, it is clear that there are thirteen values above mean (35.9 mm) and the highest value is 77.2 mm occurred on 9th July, 2003 and the 2nd highest value (74.2 mm) rain per day received in 1995. Figure 1e shows ten values above normal (25.8 mm) while the highest value (83.8 mm) occurred on 8th July 2003. Similarly from Figure 1f, we can see that there are twelve values above normal (40.3 mm) and the highest value is 140.7 mm, occurred at 5th August 1983. In short, from Figure 1, we can conclude that the heaviest rainfall occurred at Pasni as compared to other stations.

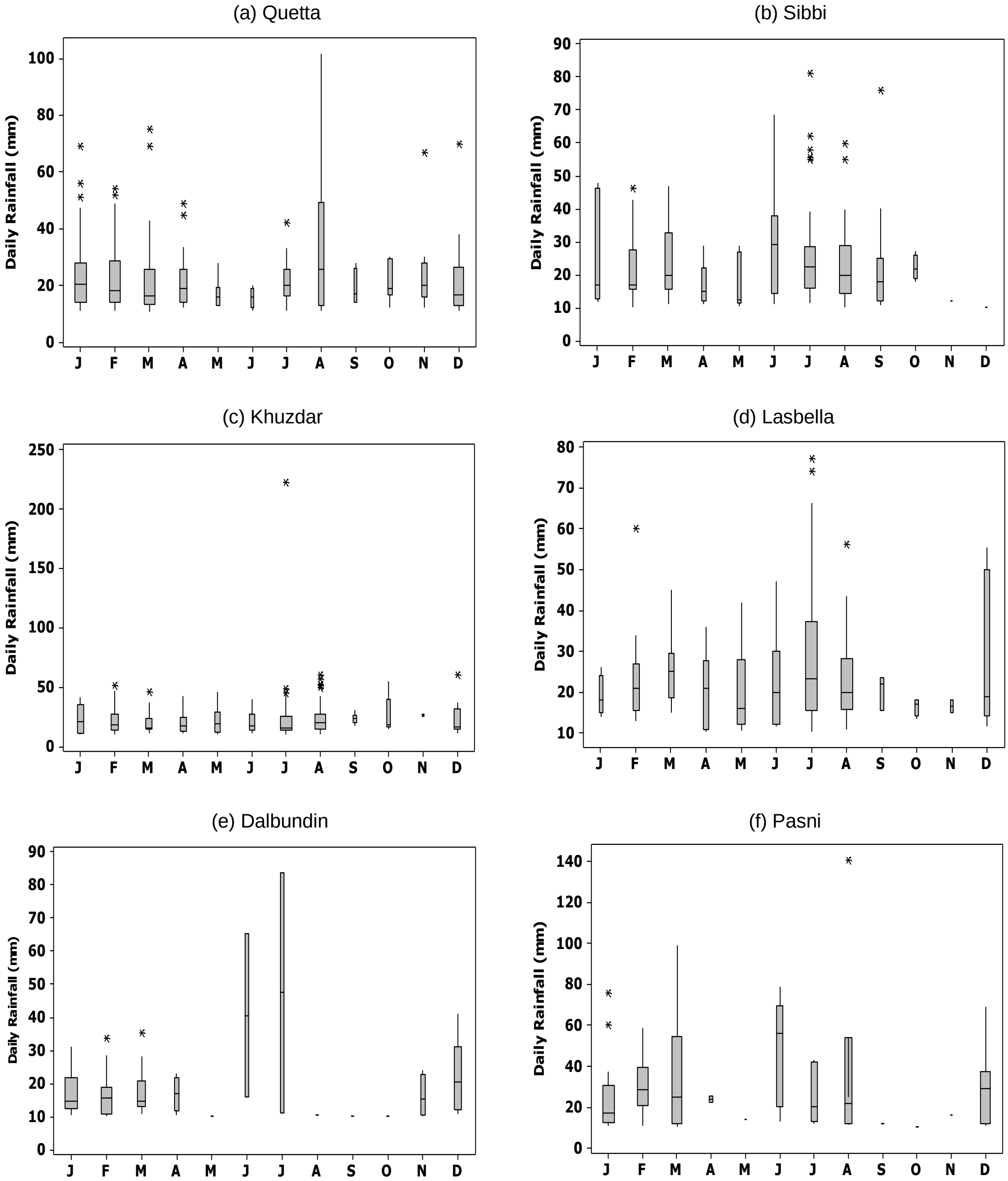


Figure 2. Monthly box plots of exceedances of 10mm threshold for different cities of Baluchistan.

2.1. Box Plot

Exploratory data analysis (EDA) gave rise to a number of new graphical techniques. One of a very widely used graphical tool introduced by [8] is the box plot or box-and-whisker plot. Expectation of the prospective of seasonal deviation, Figure 2 shows the daily data in the form of box-plots having box width proportional to the sample size, arranged by month. The greater the width of the boxes, the more frequency of rainfall will occur in that month. Also the greater the IQR the greater will be the variability in that month. Therefore, it is clear from Figure 2a that the month of January has greatest frequency while June has the lowest and December, January, February and March has greatest variability of rainfall events due to greater IQR. To avoid the compression of the figures due to the large number of near-zero observations, we have formed plots only for extremes over a threshold of 10 mm. Similarly we can examine different frequency and variability of rainfall events in other stations also. These figures also illustrate those different cities of Baluchistan having heavy rainfall during the months of July and August and winter season also. We also observe that there should be two seasons, summer season from June, July and August and winter season from December to March. There is graphical support for this analysis also.

3. Extreme Value Distribution Fit

Let Y1, Y2, Y3,… Yn, denotes the daily rainfall data. Studying the behavior of standard model for extremes *Mn* = max (Y1, Y2, Y3,. . .Yn) for large values of *n*. With *n* = 365, *Mn* = annual maximum value of rainfall. Asymptotic considerations propose that the distribution of *Mn* should be approximately that of a member of the generalized extreme value (GEV) family [9], having distribution function (PDF).

$$F(y) = \frac{1}{b}[1 + H]^{1-\frac{1}{k}}\ e^{-(1+H)^{-\frac{1}{k}}} \quad , 1+H > 0 \qquad (1)$$

where

$$H = \frac{k(y-a)}{b}$$

and a = location (or shift) parameter, b = scale parameter, k = shape parameter. By integrating equation (1) analytically, we obtained CDF as

$$F(y) = = e^{-(1+H)^{-\frac{1}{k}}} \qquad (3)$$

Further CDF can be reversed to yield an explicit formula for the quantile function.

$$Y = F^{-1}(p) = a + \frac{b}{k}\{[-\ln(P)]^{-k} \qquad (4)$$

Here P = F(y) is the cumulative probability

Generally, GEV distribution is fitted using either the method of maximum likelihood or a method known as L-moments [10], which is used frequently in hydrological applications. L-moment fitting tends to be preferred for small data samples [11]. 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 [12, 13]. For moderate and large sample sizes the results of the both parameter estimation methods are usually similar. The special case of the GEV distribution in which k = 0, is called 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 Yi. 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:

$$F(y) = \frac{1}{b}\ e^{\left\{-e^{-H^*} - H^*\right\}} \qquad (5)$$

where

$$H^* = \left(\frac{y-a}{b}\right) \qquad (6)$$

'a' is location parameter and 'b' is scale parameter (b > 0). The shape of Gumbel distribution does not depend on the distribution parameters.

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 moment's estimators for the Gumbel distribution parameters are computed using the sample mean and standard deviation. The estimation equations are

$$\hat{\boldsymbol{b}} = \frac{s\sqrt{6}}{\pi} \qquad (7)$$

and

$$\hat{a} = \bar{Y} - \gamma\bar{b} \qquad (8)$$

where γ = 0.57721 and is called Euler's constant.

3.2. Best Fitted Distribution

Generalized Extreme Value and Gumbel Maximum distribution for extreme annual rainfall data and find which one is the best fitted. Here we used three processes as;

1. Histogram with fitted probability density function.
2. Chi square test and
3. Probability probability (p-p) plots.

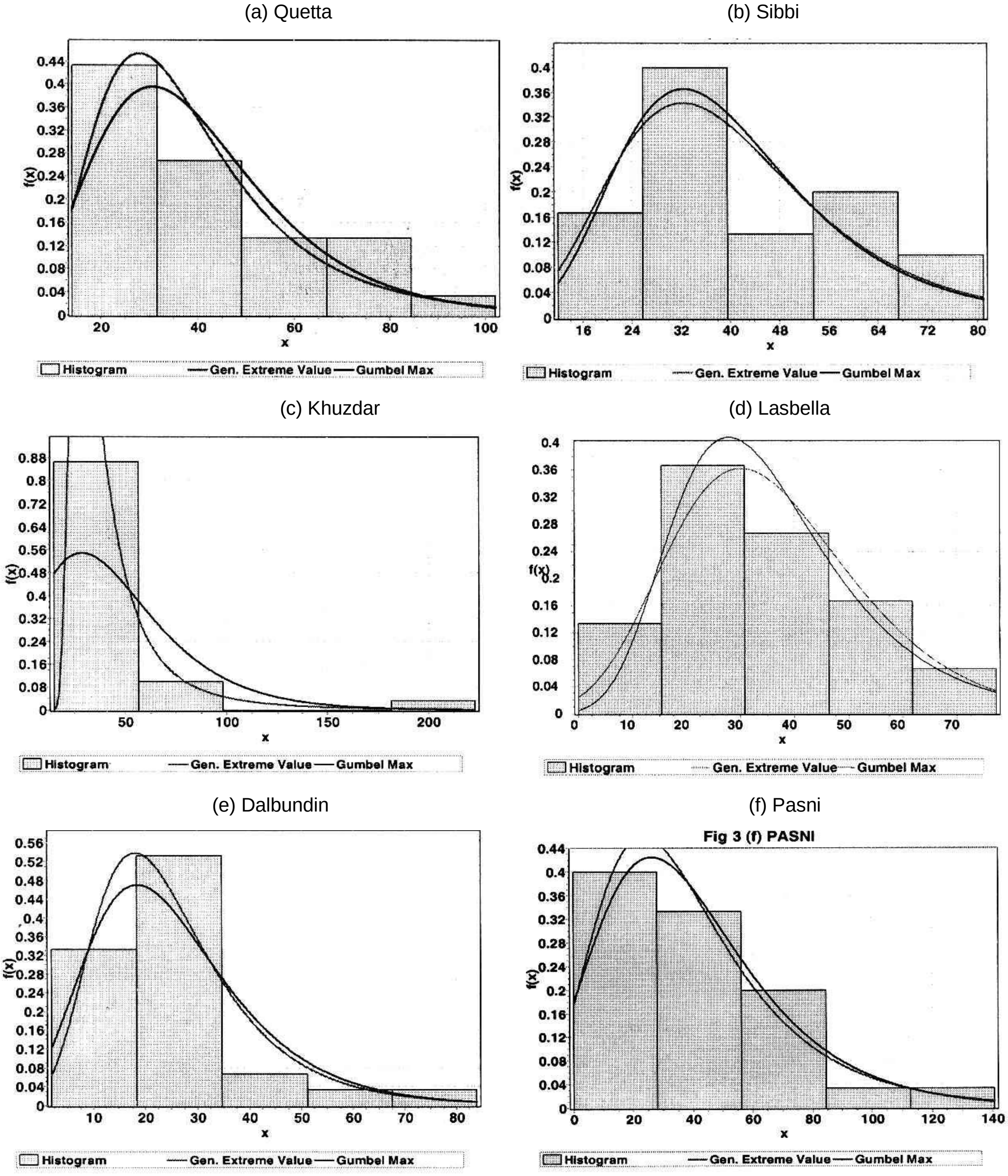


Figure 3. Probabilistic density function fitted to extreme annual rainfall of different cities of Balochistan.

Histogram plots of Figure 3 show that the GEV distribution covers more area than Gumbel Maximum distribution. Thus GEV is the best fitted distribution for the extreme annual rainfall of Quetta, Khuzdar, Dalbandin and Pasni while Sibbi and Lasbella shows that Gumbel Max. is the best fitted distribution.

For GEV, Table 1 summarized the goodness fit test for the extreme annual rainfall of Quetta, Sibbi, Khuzdar, Lasbella, Dalbandin and Pasni. As for extreme annual rainfall of Quetta, the estimated χ^2 = 0.65709 for GEV distribution. Under the negative null hypothesis, the statistic is drawn from a χ^2 distribution with degree of freedom $\nu = 5 - 3 - 1 = 1$. Referring to $\nu = 1$ of row chi-square Table, estimated χ^2 = 0.65709 is smaller than the 95th percentile value of 3.8, so the null hypothesis that data of extreme annual rainfall of Quetta have been drawn from GEV distribution would not be rejected even at 5% significance level. Thus, data of extreme annual rainfall of Quetta follow GEV distribution. Similarly, Table 1 shows that estimated chi squared values for other stations are also smaller than 95th percentile value of 3.8 (computed values of chi squared from the chi squared table) with 5% significance level.

Table 1. Summary of goodness of fit test for extreme rainfall of Baluchistan.

S. No.	Station	Chi-Squared	
		General Extreme Value Distribution	Gumbel Maximum Distribution
1	Quetta	0.65709	1.3201
2	Sibbi	1.3978	1.2609
3	Khuzdar	0.30248	1.3509
4	Lasbella	1.139	0.15249
5	Dalbandin	1.0999	1.3346
6	Pasni	0.3671	0.69719

For Gumbel Maximum (GM), the value of χ^2 is 5.9 at 5% level with degree of freedom $\nu = 5 - 2 - 1 = 2$. Thus, Table 1 also shows that the estimated values of chi squared for all stations are smaller than 95th percentile value of 5.9 (computed values of chi squared from the chi squared table) with 5% significance level. So both the distributions are applicable to calculate extreme annual rainfall, but the estimated chi sq. values of GEV are smaller for Quetta, Khuzdar, Dalbandin and Pasni than those of GM suggesting that GEV is the best fitted for these stations. Similarly, estimated values of chi sq. for GM are smaller for Sibbi and Lasbella than those of GEV, so GM is the best fitted distribution for these two stations. Thus data of extreme annual rainfall of above stations have been drawn from the the respective fitted distribution.

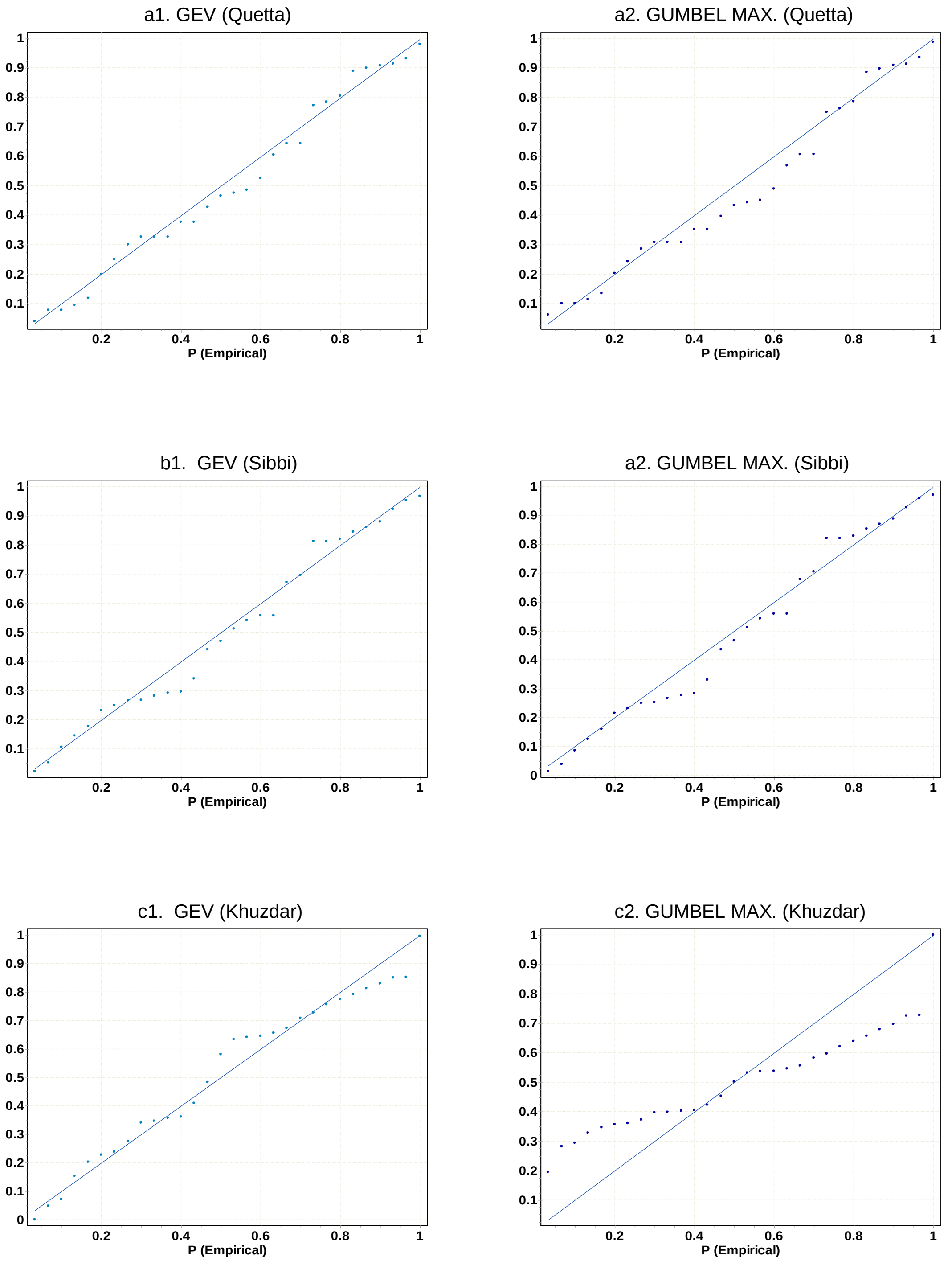
To demonstrate the above goodness of fit test results P-P Plots was drawn for extreme annual rainfall of above mentioned cities in Figure 4. These figures also show that the deviation of observed data points from theoretical CDF values is relatively more in Gumbel Maximum distribution for (Quetta, Khuzdar, Dalbandin and Pasni) and less for Sibbi and Lasbella as compared to General Extreme Value and vice versa. So a distribution is more fitted if the fitted line has less deviation from the data points. Hence, GEV. distribution is the best fitted distribution for Quetta, Khuzdar, Dalbandin and Pasni while Gumbel Maximum for Sibbi and Lasbella.

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 probabilities that may be substantially larger than 1 – 1/n. Often these extreme probabilities are expressed as average return periods,

$$R(y) = \frac{1}{\omega[1 - F(y)]} \tag{9}$$

Return period R(y) associated with a quantile *Y* typically is interpreted to be the average time between occurrences 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(y) = 0.99 will have probability 1 – F(y) of being exceeded in any given year. This value of x would be associated with a return period of 100 years, and would be called the 100-year event.



(Contd.)

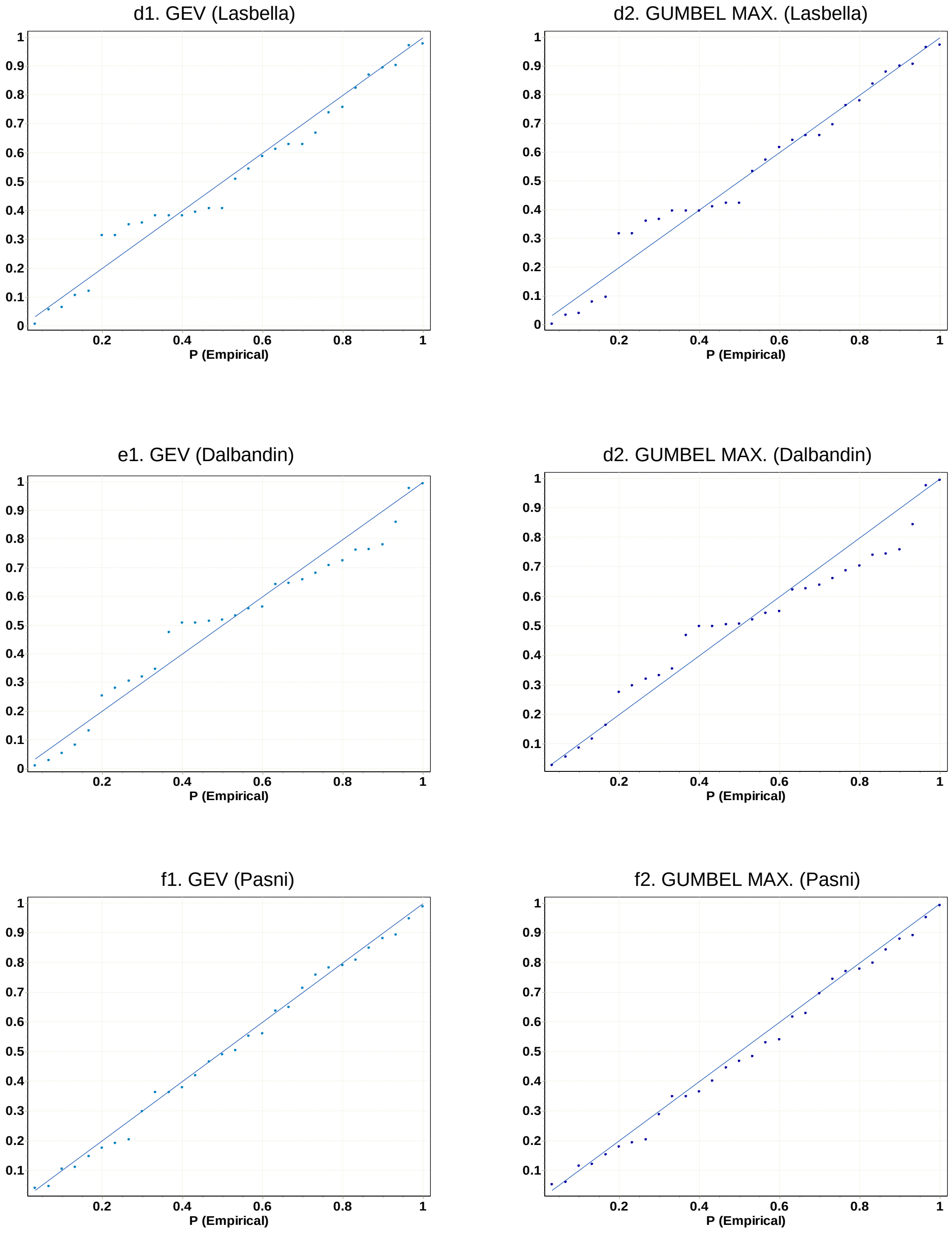


Figure 4. P-P Plot of Gumbel Maximum and GEV Distribution of Different Cities of Baluchistan.

4.1. Return Period for Quetta

The analysis shows that the maximum amount of rainfall i.e. 102 mm occurred on 3rd August 1983 throughout the period 1981 to 2010. So we want to study 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 Quetta yielded the parameter estimatesa = 29.648, b = 14.396 and k = 0.13275.

Cumulative probability was calculated by using Eq. (2)

Table 2. Return Levels for different return periods.

Return period (years)	Quetta (mm)	Sibbi (mm)	Khuzdar (mm)	Lasbella (mm)	Dalbandin (mm)	Pasni (mm)	Average rainfall Baluchistan (mm)
2	35.06	37.53	35.82	33.04	22.84	33.76	33.0
5	53.54	53.28	53.63	48.79	36.25	61.44	51.2
10	67.41	63.7	70.75	59.23	45.65	81.19	64.7
20	82.07	73.7	92.79	69.23	55.07	101.29	79.0
30	93.59	80.95	113.22	76.48	62.14	116.58	90.5
50	103.24	86.66	132.6	82.19	67.87	129.08	100.3
100	120.92	96.36	173.84	91.89	77.95	151.3	118.7

$$H = \frac{k(y-a)}{b} = 0.67$$

and using Eq. (3)

$F(Y) = P(Y \leq 242.2) = F(Y) = e^{-(1+H)^{-\frac{1}{k}}}$

$F(Y) = \exp - (1+1.506)^{-4.32}$

$F(Y) = \exp -0.019$

$F(Y) = 0.978$

Now Eq. (9) yields

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

R(Y) = 45 years

It means that in coming 45 years there is a chance to occur rainfall amount equals to 102 mm or more in a day.

4.2. Return Levels for Different Return Periods

To compute return period for T = 2 years, p = 0.5 and $F(Y) = P = 1 - p = 0.5$, were p is probability while P is the cumulative probability i.e. $F(Y)$. Using Eq. (IV), we estimated that 35.06 mm of rainfall as return level. Hence this result shows that in coming two years there is a chance of occurring 35.06 mm rain at Quetta station. We have also calculated the return levels 53.54 mm, 67.41 mm, 82.07 mm, 93.59mm, 103.24 mm and 130.92 mm for 5, 10, 20,30, 50 and 100 years respectively.

Calculations of return levels against different return period for six cites of Baluchistan were summarized in Table 2. The calculations show that different cities of Baluchistan have 20-years return period for receiving more than 50 mm daily rainfall. While they have 50-years return period for receiving more than 65 mm daily rainfall showing that Baluchistan receives less rainfall as compared to other province of Pakistan.

5. Conclusions

The study quantifies probabilistic approach to estimate return period of extremes of annual rainfall of different cities of Baluchistan (Quetta, Sibbi, Khuzdar, Lasbella, Dalbandin and Pasni). Daily rainfall of six major cities of Baluchistan was utilized for the period January 1981 to December 2010. The time series of annual maximum data of above said stations have not a good trend. Monthly box-plots of daily data show that the region under study has two seasons.

(i) Summer season from July to September.

(ii) Winter season from December to March.

The research depicts that GEV is the best fitted for the extreme annual rainfall of different cities of Baluchistan except Sibbi and Lasbella, for which Gumbell Maximum is the best fitted distribution. Further analysis indicates that different cities of Baluchistan have 30-years return period for receiving more than 90.5 mm average daily rainfall. While they have 100-years return period for receiving more than 118 mm daily rainfall.

Acknowledgements

We are thankful to Pakistan Meteorological Department, Karachi for providing the required daily rainfall data.

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] T.J. Osborn, M. Hulme, P.D. Jones and T.A. Basnett, International Journal of Climatology **20** (2000) 347.

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

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

[7]. M. Ali and M. J. Iqbal, The Nucleus **49** (2012) 107.

[8]. J.W. Tukey, Exploratory Data Analysis, Reading, MA: Addison Wesley (1977) pp. 39.

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

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

[12]. R.W. Katz, G.S. Brush and M.B. Parlange, Ecology **86** (2005) 1124.

[13]. Zhang et al., Journal of Geophysical Research **110**(2005) D22104.