MAUSAM, **66**, 4 (October 2015), 803-808

551.579 : 556.013

Modeling of rainfall extremes through the application of Extremal Type Theorem (ETT)

NAEEM SADIQ

Institute of Space & Planetary Astrophysics, University of Karachi, Karachi – 75270, Pakistan

(*Received 20 February 2014, Modified 21November 2014*)

e mail : nsadiq@uok.edu.pk

सार – पाकिस्तान में भयंकर बाढ़ आने का एक प्रमुख कारण अत्यधिक वर्षा का होना है और अत्यधिक मात्रा में वर्षा होने से कृषि उत्पादों और शहरी संरचनाओं पर खगोलीय रूप से हानिकारक प्रभाव पड़े हैं। 1981-2010 की अवधि में पाकिस्तान के विभिन्न मध्यम दर्जे के शहरों (सयालकोट, दीर, नवाबशाह, चिलास और बरखान) के लिए अत्यधिक वर्षा के आँकड़ों का उपयोग करते हुए संभाव्य मॉडलिंग के उपायों द्वारा चरम मान सिद्धांत पर आधारित रिटर्न अवधियों के आकलन किए गए। रिटर्न अवधियों का आकलन करने के लिए एक्सट्रीम टाइप थ्योरम (ETT) के सामान्यीकृत चरम मान (GEV) वितरण मॉडल अंतिम छोर (इक्सट्रीम) में लगाए गए। परिकलित आकलनों से यह पता चला है कि मध्यम दर्जे के शहरों सियालकोट और चिलास में पाँच वर्ष की रिटर्न अवधि के लिए दैनिक वर्षा की मात्रा क्रमशः अधिकतम (129.8 मि. मी.) और न्यूनतम (48.1 मि. मि.) रही। इससे यह पता चला है कि नवाबशाह और बरखान के लिए दस वर्ष की रिटर्न अवधि 85 मि. मी. से अधिक है जबकि दीर और चिलास के लिए यह क्रमशः 120 मि. मी. और 60 मि. मी. से भी अधिक है। इसके अलावा 10, 20, 30, 50 और 100 वर्षों के रिटर्न अवधियों का भी आकलन किया गया।

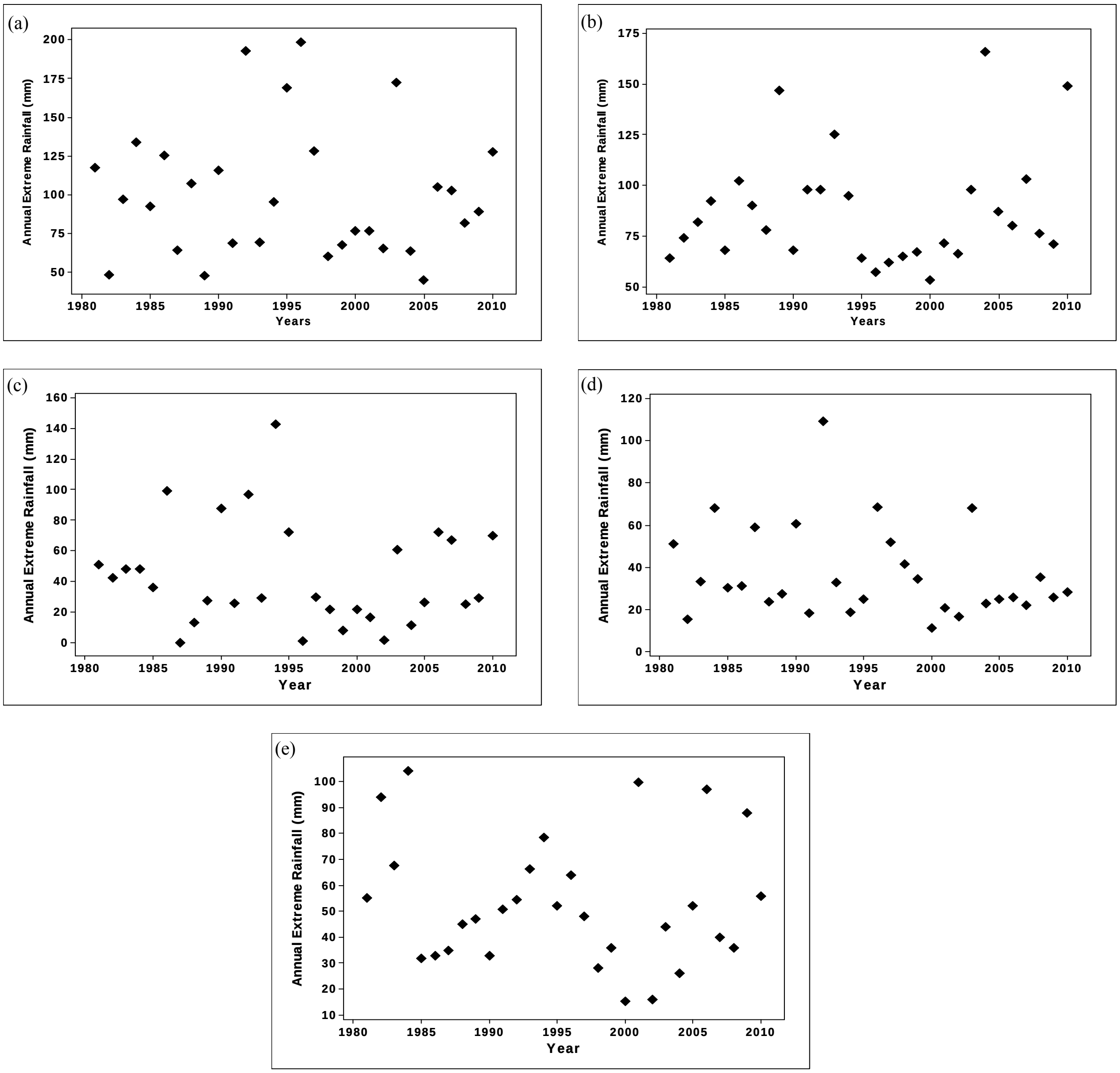
ABSTRACT. Rainfall extremes are one of the main reason to raise vigorous flooding in Pakistan and their augmenting magnitude have an astronomically detrimental impact on the agriculture products and urban infrastructure. Estimation of return periods based on Extreme Value Theory have been worked out by means of probabilistic modeling by utilizing the rainfall extreme data for the different meso-cities (Sialkot, Dir, Nawabshah, Chilas and Barkhan) of Pakistan for the period 1981-2010. Generalized Extreme Value (GEV) Distribution model of Extremal Type Theorem (ETT) have been fitted to extremes to estimate the return periods. Calculated estimations reveal that among the considered meso-cities, Sialkot and Chilas have the highest (129.8 mm) and lowest (48.1 mm) amount of daily rainfall for the five year return period, respectively. It is also explored that ten year return period for Nawabshah and Barkhan is more than 85 mm, while for Dir and Chilas it is greater than 120 mm and 60 mm, respectively. Moreover, the return periods for 10, 20, 30, 50 and 100 years have also been estimated.

Key words – Extreme value theory, Extremal type theorem, Diurnal rainfall.

1. Introduction

Extreme value modeling is carried out in different fields such as wind engineering (Coles and Walshaw, 1994; Harris, 2001), risk management (Burton and Makropoulos, 1985), flooding (Ahmad *et al*., 1988; Bruun and Tawn, 1998), stock index (Broussard and Booth, 1998), environment (Bai *et al*., 1992), Pollusion (Coles and Pan, 1996) etc. Generally, there are two ways and means to analyze the extreme rainfall events (Fowler and Kilsby, 2003). In the first method, extreme rainfall analysis utilizes percentiles or quintiles (Karl and Knight, 1998; Osborn *et al*., 2000), in which daily rainfalls are categorized by a certain percentage of the total rainfall events; each class holds an equal amount of rainfall total as squantile amount. The other technique employs the distribution modeling to identify extreme events with return periods for specified period (Hennesey *et al*., 1997; McGuffie *et al*., 1999). In this manner, return periods of extreme rainfall are estimated through fitting of distribution bases on Extremes Value Theory (EVT) to the series of annual maxima.

Torrential rainfall that gives rise to floods is the most crucial type of weather extremes in Pakistan that cause flooded villages, washing away cultivated lands and ruining the urban infrastructure together with roads and bridges in addition to loss of lives (Tariq and Giesen, 2012). Till the year 2013 (from 1950), there have been more than 11,572 lives lost in addition to more than US$ 37.554 billion in losses (FFC, 2013). The study is an attempt to analyze the daily rainfall annual maxima through probability distribution function for different meso-cities of Pakistan to estimate the return periods and



Figs. 1(a-e). Annual extreme rainfall values observed in different urban cities of Pakistan (a) Sialkot (b) Dir (c) Nawabshah (d) Chilas and (e) Barkhan

utilize the adequate probabilistic distribution model that gives the appropriate representation of the selected dataset.

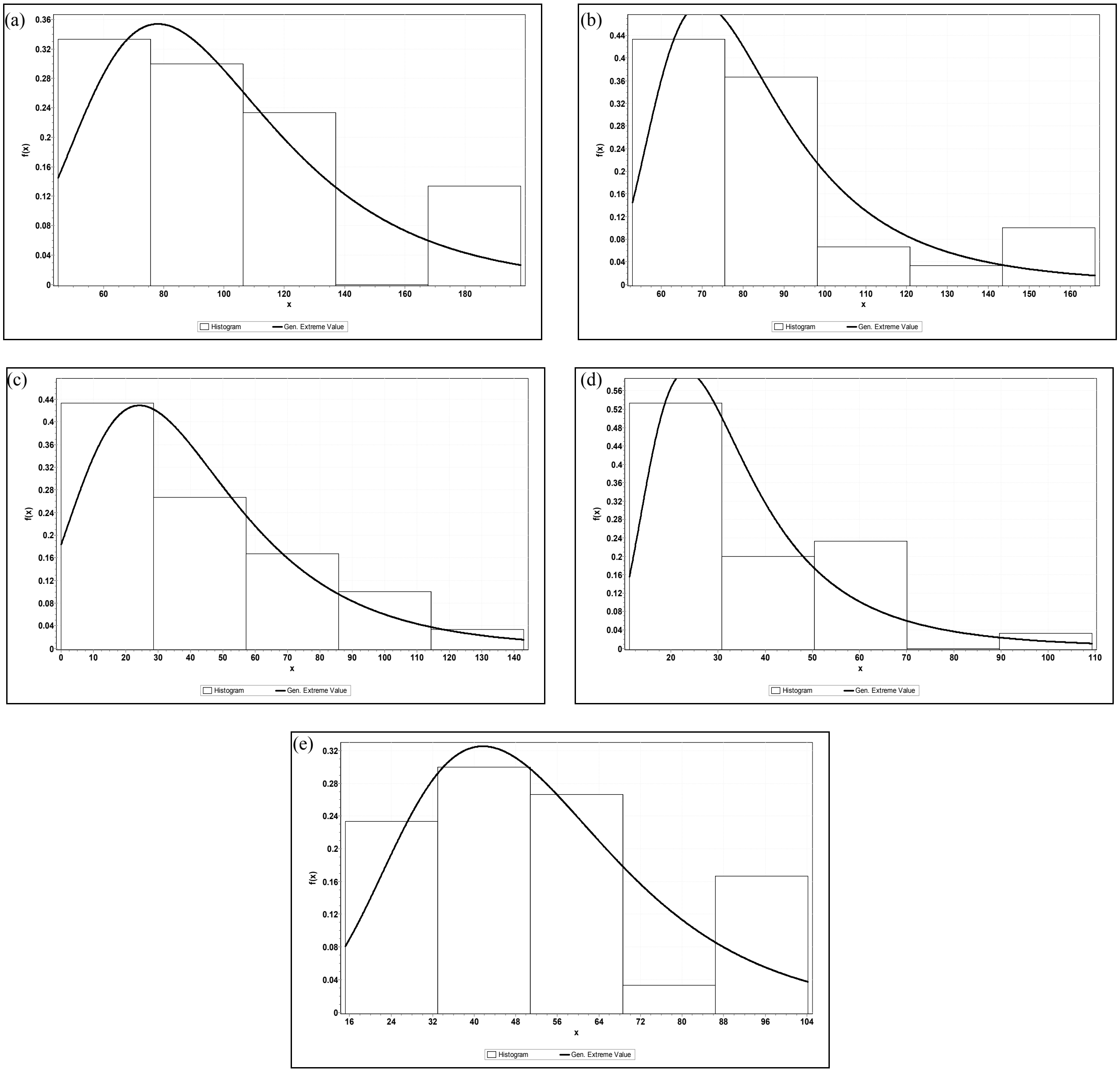
2 Materials and methodology

2.1. *Data*

For the study, five different meteorological stations, one selected from each of the five provinces, *i.e.*, Sialkot, Dir, Nawabshah, Barkhan and Chilas from Punjab, Khyber Pakhtunkhwa (KPK), Sindh, Baluchistan and Gilgit-Baltistan administration, respectively. The rainfall data (of 24-hour accumulated amount of rainfall from 8 AM on a given day until 8 AM the following day) for the available period of 1981 to 2010 is utilized in the study. All the data used in this study is provided from the Pakistan Meteorological Department archives.

2.2. *GEV distribution model*

Extremal Type Theorem (ETT) is the base of EVT which describes that the rescaled sample maxima



Figs. 2(a-e). Different Distributions functions fitted to extreme annual rainfall of different cities of Pakistan (a) Sialkot, (b) Dir, (c) Nawabshah, (d) Chilas and (e) Barkhan

converge in distribution to a variable having distribution, possibly within any one of the Gumbel, Frechet and Weibull (also called Type I, Type II and Type III) families, respectively. The amalgamation of these three types into a single family of models acquire distribution function in the form:

$$G(z) = \exp\left[-\left\{1+\xi\left(\frac{z-\mu}{\sigma}\right)\right\}^{-1/\xi}\right] \quad (1)$$

defined on the set $\{z : 1+\xi(z-\mu)/\sigma > 0\}$, where the parameters satisfy $-\infty < \mu < \infty, \sigma > 0$ and $-\infty < \xi < \infty$. This is the GEV (generalized extreme value) family of distributions. The model has three different parameters *viz*., μ, σ, and ξ known as location, scale and shape parameters, respectively. Type I, II and III (classes of extreme value distribution) corresponds to ξ = 0, ξ > 0 and ξ < 0 respectively. In the study we utilized the generalized form of Eqn. 1 only; Specific forms though otherwise widely used (Sadiq and Qureshi, 2014) are not considered in this study.

TABLE 1

Anderson Darling Goodness of fit test and Recurrence Intervals for selected meso-cities

Station	GEV	Recurrence Intervals (years)					
		5	10	20	30	50	100
Sialkot	0.21388	129.8	156.5	183.1	203.1	219.2	247.6
Dir	0.23982	103.1	121.9	142.7	159.8	174.6	202.8
Nawabshah	0.23312	65.8	87.5	109.8	126.9	140.9	165.9
Chilas	0.29522	48.1	62.6	79.1	92.9	105.1	128.7
Barkhan	0.24881	71.7	86.6	101	111.3	119.4	133.2

3. Results and discussion

The plot of extreme value for each year (Fig. 1) exhibits no definite trend for any considered city, pointing highly non linear pattern (Sadiq and Qureshi, 2010). Moreover, for Sialkot 13, Dir 12, Nawabshah and Barkhan 10, and, for Chilas 9 extreme daily numeric values appear above normal. The highest extreme value for Sialkot, Dir, Nawabshah, Chilas and Barkhan are found to be 198.40, 166, 143, 109.3 and 104.2 mm, respectively.

3.1. *Distribution fitting*

To initiate our analysis, histograms are formed with number of bins subject to the total number of observations and then GEV distribution is fitted to the histograms (Fig. 2). It is obvious from this figure that the GEV distribution provides good fit for the plotted histograms. This is analysed in detail in the forthcoming sections.

3.2. *Test statistics*

To evaluate the 'distance' to threshold value and to measure the 'distance' between the fitted distribution and the data, Anderson Darling (A^2) goodness of fit test is utilized. It makes it easier to compare the fit of an observed cumulative distribution function to the function of an expected cumulative distribution via

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1).\left[\ln F(X_i) + \ln\{1 - F(X_{n-i+1})\}\right] \quad (2)$$

The hypothesis pertains to the distributional form is rejected and considered to be inappropriate at the chosen significance level (α) if the test statistic, value obtained from A^2, is greater than the critical value (2.5018 in our case). The fixed value of α (0.05 in our case) is utilized to assess the null hypothesis (H_o). Table 1 displays the summary of the goodness of fit test for the different distributions. For example, the estimated A^2 for the extreme annual rainfall for Sialkot comes out as 0.213 which is smaller than the 95th percentile value of 2.5018. Hence, H_o suggests that annual extreme rainfall data of Sialkot cannot be rejected even at the 5% level. Likewise, for the other four stations estimated A^2 values are obtained with a 5% significance level (Table 1). Hence, more closely, rainfall extremes data for the five stations have been drawn from the GEV distribution.

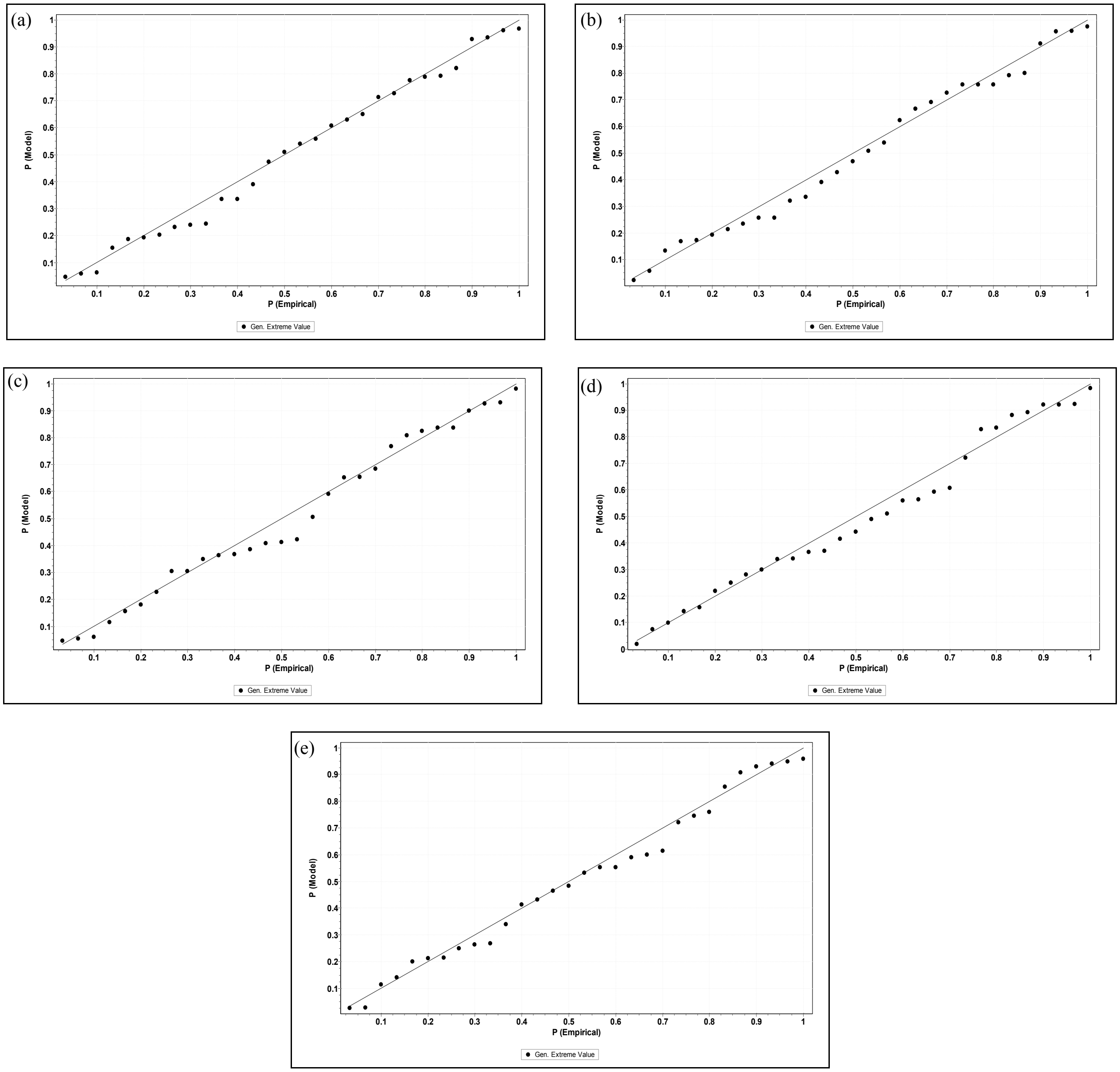
Moreover, to determine that how well a specific distribution fits the observed data, probability-probability (P-P) plot which shows the empirical cumulative distribution function (CDF) against the theoretical CDF values is utilized. P-P plot should be close to linear model if the specified theoretical distribution is the correct choice of a model. These plots are drawn to elucidate the goodness of fit test results, for the annual rainfall of cities under consideration [Figs. 3(a-e)]. It is obvious from the figure that the deviation of observed data points from theoretical CDF values is comparatively more in other distributions as compared to the GEV distribution. As a principle, the lesser the deviation – the better fitted will be the distribution, hence, the GEV distribution again appears to be as an adequately fitted distribution for the dataset.

3.3. *Return period estimation*

Estimates of extreme quintiles for the annual maximum distribution may be acquired by inverting Eqn. 5 as

$$z_p = \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right] \qquad for\ \xi \neq 0 \qquad (3)$$

whereas $G(z_p) = 1 - p$. As Coles (2001) indicates, z_p is the nomenclature of return period associated with the recurrence interval 1/p. Level z_p is probable to be exceeded on average once every $1/p$ years, or more exactly, z_p is outdone by the annual maximum in any



Figs. 3(a-e). Probability- Probability (p-p) plot to extreme annual rainfall of different cities of Pakistan (a) Sialkot (b) Dir (c) Nawabshah (d) Chilas and (e) Barkhan

specific year with probability *p*. By putting the respective GEV parametric values for location (μ), scale (σ) and shape (ξ) parameters along different values for probability (*p*), distinct values of return periods can be estimated. As for instance, if $p = 0.2$ by utilizing the above equation, return period of five years may be estimated for extreme diurnal data of the different cities of Pakistan. Hence, the resulting estimates 129.8, 103.1, 65.8, 48.1 and 71.7 mm of rainfall for Sialkot, Dir, Nawabshah, Chilas and Barkhan, respectively, are expected to occur in one day over a five-year period. For the twenty years return level, Nawabshah and Barkhan acquire more than 100 mm of rainfall in a day, while Chilas acquires almost 80 mm. In similar manner, return periods for 5, 10, 20, 30, 50 and 100 years are estimated for the above cities and are summarized in Table 1. For the 30 year return period the rainfall in one day in Sialkot exceeds 200 mm.

4. Conclusions

Return Periods of thirty years annual extreme rainfalls (*i.e.*, 1981-2010) have been estimated for the five selected meso-cities (*i.e*., Sialkot, Dir, Nawabshah, Chilas and Barkhan) of Pakistan through GEV distribution modeling. Test statistics of Anderson Darling goodness of fit test were utilized to ensure the credibility of the distribution. Recurrence interval estimates predict the utmost (~130 mm) and least (~48 mm) rainfalls for Sialkot and Chilas within five year's period. For the ten year period, this amount exceeds from 155 mm and 120 mm for Sialkot and Dir while Nawabshah and Barkhan also bore more than 85 mm of rainfall. There is also a chance of rainfall more than 200 mm over Sialkot in thirty years. Regarding twenty years' estimates, except Chilas (~80 mm) rest of the stations acquire, not less than 100 mm of rainfall in a day.

Acknowledgement

The financial support provided for this work by the Dean Faculty of Science, University of Karachi is hereby acknowledged.

References

Ahmad, M. I., Sinclair, C. D. and Spurr, B. D, 1988, "Assessment of flood frequency models using empirical distribution function statistics", *Water Resources Res*., **24**, 1323-1328.

Bai, J., Jakeman, A. J. and McAleer, M., 1992, "On the use of extreme value distributions for predicting the upper percentiles of environmental quality data", *Math. Comput. Simulation*, **33**, 483-488.

Broussard, J. P. and Booth, G. G., 1998, "The behavior of extreme values in Germany's stock index futures: An application to intradaily margin setting", *European Journal of Operational Research*, **104**, 3, 393-402.

Bruun, J. T. and Tawn, J. A., 1998, "Comparison of approaches for estimating the probability of coastal flooding", *Appl. Statist*., **47**, 405-423.

Burton, P. W. and Makropoulos, K. C., 1985, "Seismic risk of circum-Pacific earthquakes: 11. Extreme values using Gumbel's third distribution and the relationship with strain energy release", *Pure Appl. Geophys*., **123**, 849-866.

Coles, S. G. and Walshaw, D., 1994, "Directional modelling of extreme wind speeds", *Appl. Statist*., **43**, 139-157.

Coles, S. G. and Pan, F., 1996, "The analysis of extreme pollution levels: A case study", *J. Appl. Statist*., **23**, 333-348.

Coles, S., 2001, "An Introduction to Statistical Modeling of Extreme Values", Springer-Varlag, Athenaeum Press, UK.

Federal Flood Commission's Annual Flood Report, 2013, Ministry of Water and Power, Government of Pakistan. Available at http://www.ffc.gov.pk/download/Annual%20Flood%20Report 202013.pdf, Accessed on 13th November, 2013.

Fowler, H. J. and Kilsby, C. G., 2003, "A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000", *Int. J. Climatol*., **23**, 1313-1334.

Harris, R. I. 2001, "The accuracy of design values predicted from extreme value analysis", *J. of Wind Eng. & Ind. Aerody*., **89**, 153-167.

Hennesey, K. J. Gregory, J. M. and Mitchell, J. F., 1997, "Change in daily precipitation under enhanced greenhouse conditions", *Clim. Dyn*., **13**, 667-680.

Karl, T. R. and Knight, R. W., 1998, "Secular trends of precipitation amount, frequency, and intensity in the United States", *Bull. Am. Meteorol. Soc*., **79**, 231-241.

McGuffie, K., Henderson-Sellers, A., Holbrook, N., Kothavala, Z., Balachova, O. and Hoekstra, J., 1999, "Assessing simulations of daily temperature and precipitation variability with global climate models for present and enhanced greenhouse climates", *Int. J. Climatol*., **19**, 1, 1-26.

Osborn, T. J., Hulme, M., Jones, P. D. and Basnett T. A., 2000, "Observed trends in the daily intensity of United Kingdom precipitation", *Int. J. Climatol*., **20**, 347-364.

Sadiq, N. and Qureshi, M. S., 2010, "Climatic Variability and Linear Trend Models for the Five Major Cities of Pakistan", *J. of Geog. & Geol.*, **2**, 1, 83-92.

Sadiq, N. and Qureshi, M. S., 2014, "Estimating Recurrence Intervals of Extreme Rainfall through a Probabilistic Modeling approach for different urban cities of Pakistan", *Arabian Journal for Science and Engineering,* **39**, 1, 191-198.

Tariq, M. A. U. R. and van de Giesen, N., 2012, "Floods and flood management in Pakistan", *Physics and Chemistry of the Earth*, Parts A/B/C, **47**, 11-20.