PROBABILITY MODELING OF LOW FLOWS AT DIFFERENT SITES OF INDUS BASIN IN PAKISTAN USING L-MOMENTS AND TL-MOMENTS

I. Ahmad, M. Fawad, A. Abbas* and A. Saghir**

Department of Mathematics and Statistics, International Islamic University, Islamabad, Pakistan.
*Department of Mathematics, University of Poonch, AJK, Rawalakot, Pakistan.
**Department of Mathematics, Mir Pur University of Science and Technology, Mir Pur AJK, Pakistan.

ABSTRACT: At-site Frequency Analysis (ASFA) of low flow was carried out for nine sites of Indus basin in Pakistan. In the present study, 10-day annual low flow series were analyzed by robust estimation methods such as Method of L-moment (ML) and TL-moment (MTL) to identify best fit probability distributions for each site. Best distribution for each site was identified using different goodness-of-fit Tests (GFT). No single probability distribution was declared as the best-fit distribution for all sites included in the plan. The GFT results indicated GPA was the most appropriate distribution for most of the sites followed by GLO and GEV distributions. On comparison, it was found that for most of the sites ML was best estimation method and for others MTL. For ASFA, the quantiles of best fit distribution were also estimated. It was found that estimated low flows based on fitted distribution were in close agreement with observed flows.

Key words: Goodness-of-fit Test, L-Moments, Probability Distributions, Quantile Estimates, Return Period, TL-moments.

(Received 11-11-2015 Accepted 23-02-2016)

INTRODUCTION

Pakistan is an agro-based country and its agriculture mainly depends on waters of Indus basin. The basin is mainly irrigated by Indus River itself and its tributaries viz-a-viz River Jhelum, Chenab, Ravi, Sutlej and Beas. From 1998-2002 Pakistan faced severe drought conditions in the Indus plain, which proved to the worst drought in the history of Pakistan. ASFA of low flows is of great concern in water resources research including water quality management, determination of downstream flow requirement for hydropower generation, designing of irrigation system and impact of prolonged droughts on aquatic ecosystems in the country (Richter *et al*, 2003). Low flow in Indus basin adversely affects agriculture, environment, economy and ecosystem of Pakistan. So there is a dire need of Frequency Analysis (FA) of low flow at Indus basin in Pakistan. The procedure for estimating frequency of occurrence of hydrological events is known as FA (Noto and Loggia, 2009). Various aspects for low flow have been discussed to determine the type of probability distribution across the world as reported by (Gubareva and Gartsman, 2010; Yurekli *et al*, 2005;; Zaidman *et al*, 2002; Kroll and Vogel, 2002; Önöz and Bayazit, 2001; Caruso, 2000; Durrans and Tomic, 1996; Vogel and Wilson, 1996; Clausen and Pearson, 1995). Most of these studies considered GEV, GPA, PE3, LN3 and GLO for best fit candidate distributions. Among other estimation methods, method of L-moments was mostly used developed by (Hosking, 1990). Estimates based on simple moments methods are influenced by extreme events. While the estimates based on L-Moments are less effected from such extreme observations without removing them from the data set. The estimates from such methods are more reliable as compared to conventional methods. LM not only outperforms the conventional moments but also often more efficient than small and moderate sample sizes for meteorological data as has been reported by (Ahmad *et al*, 2013; Ahmad *et al*, 2014; Hosking and Wallis, 1987 a and Hosking *et al*, 1985b).A modified version of L-moments, i.e. TL-moments, developed by (Elamir and Seheult, 2003) may be used when our concern is to show extreme evets having undue influence.

MATERIALS AND METHODS

Data Description and its Initial Screening: The annual minima of10-day average series (measured in cusecs) of nine sites of Indus basin in Pakistan, located on the four rivers namely Indus, Kabul, Jhelum and Chenab were included in this study. These sites were selected based on quality of data, climate variability and change, record length and urbanization (Figure 1). The data of these sites were collected from Water and Power Development Authority (WAPDA) and Federal Flood Commission (FFC) (Table.1).

Table 1. Basic Information of nine sites used in the study

Names of sites	River	Latitude (North)	Longitude (East)	Sample size (n)	Mean	Standard deviation	Skewness	Coefficient of variation
Tarbela	Indus	33.99	72.61	52	13543.08	4704.093	-0.5194	0.3473
Nowshera	Indus	29.32	70.05	53	6960.38	1979.058	0.3804	0.2843
Kalabagh	Indus	32.95	71.50	52	21136.54	6786.088	0.8839	0.3211
Chashma	Indus	32.43	71.38	43	12644.19	7736.564	-0.1627	0.6119
Taunsa	Indus	30.50	70.80	52	15498.08	5547.990	-0.0243	0.3579
Guddu	Indus	28.30	69.50	52	17338.46	7775.052	0.5287	0.4484
Sukkur	Indus	27.72	68.79	73	871.7808	1399.412	1.7649	1.6052
Mangla	Indus	33.15	73.65	47	5353.965	3480.487	0.7684	0.6500
Marala	Indus	32.68	74.43	26	1602.404	1003.469	0.0297	0.6262

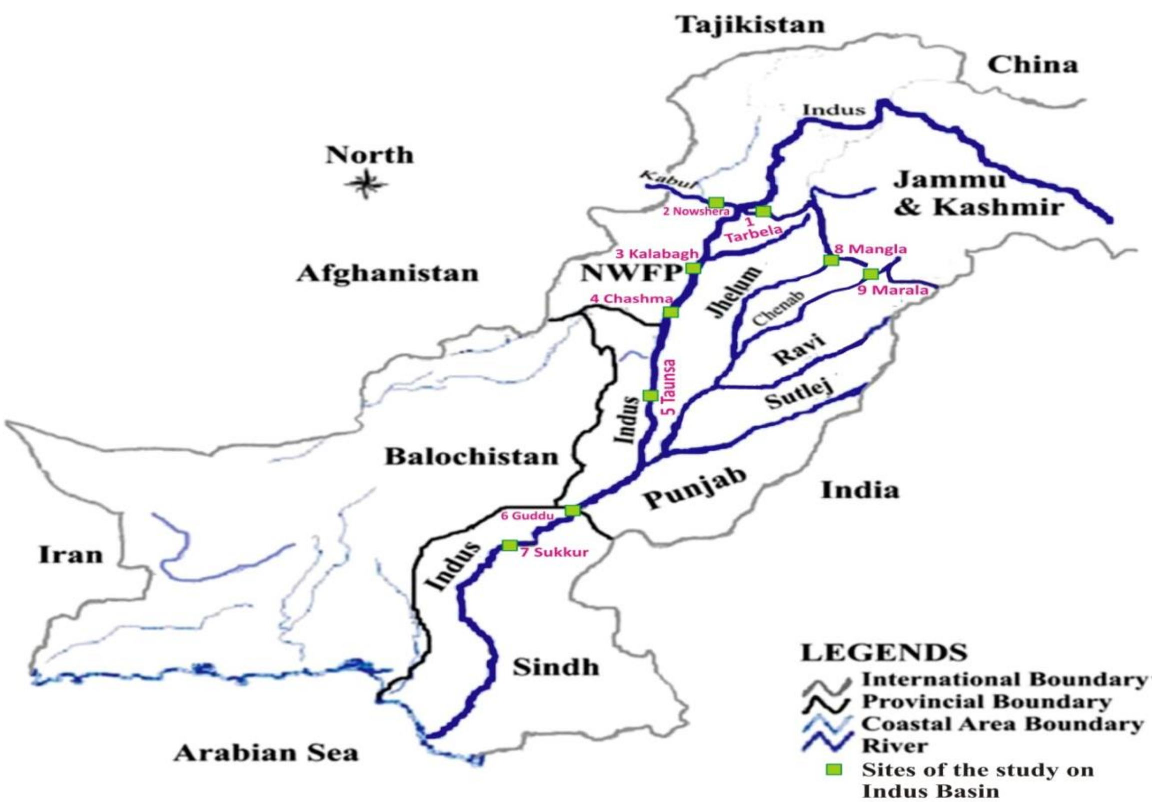

Figure 1. Geographical location of nine sites used in the study

Statistical analysis of current data was based on independence, stationarity and homogeneity. The violation of these assumptions might mislead in policy implications. The assumed data was analyzed using different parametric and non-parametric tests to test these assumptions.

Population L-Moments and Trimmed L-Moments: Population L-moments measured location, dispersion, skewness, kurtosis, and other aspects of the shape of probability distributions and sample data, using linear combinations of the ordered data values. Let $X_1, X_2, \dots, X_r$ be the random sample of magnitude,"r" with cumulative distribution Function $F(X)$ and quantile function, $X(F)$. Let $X_{1:r} \le X_{2:r} \le \dots \le X_{r:r}$ be the order statistic of random sample. For the random variable X, the rth population L-moment explained by Hosking (1990) was:

$$\lambda_r = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} E(X_{r-k:r}) \; r = 1, 2 \dots \quad (1)$$

In L-moments L highlighted that λ_r was a linear function of the expected order statistics. From first four L-moments we could find, measure of L-coefficient of variation (L-Cv), L- Skewness and L-Kurtosis as well.

In TLM the expectations of the ordered statistics of a conceptual sample (in the sense of population L-moments) were substituted by expectations of the ordered statistics of a larger conceptual sample, the size was enlarged equal to the overall trimming amount. TLM showed certain advantage over LM and conventional moments. TLM could exist even when population's mean did not exist.

For example Cauchy distribution sample TLM were unbiased to the corresponding population quantities and more robust to outliers as reported by (Elamir and Seheult, 2003). The rth TL-moments in the case of equal trimming was written below:

$$\lambda_r{}^{(1,1)} = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}E\ (X_{r+1-k:r+2})\ r = 1,2,\dots \quad (2)$$

All the corresponding quantities could be determined as in case of L-moments.

Estimation of L- moments and TL-moments: In practice, L-moments need to be estimated after taking a random sample drawn from an anonymous distribution. Let $x_1, x_2, \dots, x_n$ be the sample and $x_{1:n} \le x_{2:n} \le \cdots \le x_{n:n}$was the order statistics of the samples, then rth sample L-moments could be defined as reported by Hosking (2007) and Asquith (2007):

$$l_{r=}\frac{1}{r}\sum_{i=1}^{n}\left[\sum_{j=0}^{r-1}\frac{(-1)^j\binom{r-1}{j}\binom{i-1}{r-1-j}\binom{n-1}{j}}{\binom{n}{r}}\right]\mathrm{X}_{i:n}\ r = 1,2\dots \quad (3)$$

The estimates of L-CV, L-Kurtosis and L-kurtosis could be determined using sample quantities from above equation (3).

The rth sample TL-moments were defined below as reported by Elamir and Seheult (2003),

$$l_r^{(t_1,t_2)} = \frac{1}{r}\sum_{j=t_{1+1}}^{n-t_2}\left[\sum_{k=0}^{r-1}\frac{(-1)^k\binom{r-1}{k}\binom{j-1}{r+t_1-k-1}\binom{n-j}{t_2+k}}{\binom{n}{r+t_1+t_2}}\right]x_{j:n} \quad (4)$$

The estimates of TL-CV, TL-Kurtosis and TL-kurtosis could be determined using sample quantities from above equation (4) and also reported by (Ahmad *et al*, 2015). The TL-skewness and TL-kurtosis were also dimensionless quantities and gave information about shape of a data set.

Comparison of the Probability distributions using goodness-of-fit criteria: The goodness of fit tests such as RMSE, AD, and KS tests using ML and further MTL were used to find out the most suitable distribution for a specific site.

Root Mean Square Error (RMSE): For the judgment of probability distributions RMSE was applied to 10-days low flows data for all distributions considered in this study. About the distribution of overall fit RMSE provided better result because it calculated every single error in proportion to the size of the observation. It reduced the effect of outliers. The RMSE of smaller value obtained for given distribution revealed the appropriateness of a distribution to the actual data.

Anderson Darling (AD) Test: The AD test was used to check whether the given sample came from a particular probability distribution at hand. The null hypothesis at chosen level of significance would be rejected if calculated value of above statistic exceeds the critical value given in the table. One of the advantage of using AD test was to show good skills when applied to heavy tailed distributions with small sizes (Onoz and Bayazit, 1999; Ahmad *et al* 2015).

Kolmogorov–Smirnov (KS) test: The KS test was used to check whether the sample came from hypothesized continuous distribution. It was based on the empirical distribution function. Reject H_0 at chosen level of significance (α) if the test statistics, D was greater than the critical value obtained from table. Like AD test, KS test also demonstrated good skills when applied to skewed probability distributions, commonly used in hydrology (Baldassarre *et al*, 2009).

L-Moment Ratio Diagram (LMRD): For visual assessment, the simplest method to determine the best-fit distribution to the actual data was the use of LMRD. LMRD displayed L-moments ratios i.e. L-Skewness and L-Kurtosis of different distributions, considered in this study and data samples for individual sites.

Estimation of Quantiles of Best fit Distribution for Different return Periods: In general, ASFA needed data of large record lengths. Since the available data was of smaller length as compared to return periods of interest. For different applications such as design floods some degree of extrapolation was required as has been reported by (Rahman *et al*, 2013). After selection of best fit distribution and estimation of its parameters, one needed to find out the quantiles' estimates corresponding to different return periods (T). Larger extreme events normally corresponded to large return periods and less probability and vice versa.

RESULTS AND DISCUSSIONS

Initially, basic assumptions of low flow frequency analysis was tested by different statistical tests. For stationarity of the data, Ljung–Box Q test and Mann Kendall test were applied. Further for homogeneity and independence Mann-Whitney U test and Lag-1 correlation coefficient tests were applied respectively (Table 2). Initially nine probability distributions were considered in the study such as, Generalized Logistic (GLO), Generalized Extreme Value (GEV), Generalized Pareto (GPA), Generalized Normal (GNO),Pearson Type 3 (PE3), EXP (Exponential), GUM (Gumbel), NOR (Normal) and LOG (Logistic). ML was adopted for estimation of parameters. Most of these distributions were used for hydrological modeling in different countries as has been reported by (Rahman *et al*, 2013 and Tasker, 1987). On the basis of ratio diagram and three goodness -of-fit tests, it was found that out of nine distributions only three distributions were most suitable for 10-days annual low flows data in the present study. These three distribution were GPA, GLO and GEV. Further, to avoid undue favor to outliers and mitigate their effects, MTL was used for estimation of parameters for these three distribution under the umbrella of these goodness-of-fit tests as reported by (Asquith, 2007; Hosking, 2007; Elamir and Seheult, 2003). Implication of different estimation methods could change the results of the goodness-of-fit tests to some extent as has been reported by (Ahmad *et al*, 2015). It was found that for data of all sites ML was not only the choice (Table 3).

Table 2. Results of different tests for basic assumptions.

Sites	Mann-Whitney U test		Ljung–Box Q Statistics test		Mann Kendall Test		Lag-1 correlation coefficient	
	Mann-Whitney U	P-value	LB	P –value	Tau	P-value	r_1	P-value
Tarbela	320.500	0.749	12.054	0.123	0.0552	0.56984	0.136	0.390
Nowshera	318.000	0.606	13.091	0.519	-0.0285	0.77057	0.167	0.212
Kalabagh	334.500	0.949	26.653	0.056	-0.0762	0.43003	0.200	0.111
Chashma	163.500	0.101	23.149	0.058	-0.151	0.15764	0.143	0.331
Taunsa	259.000	0.148	13.606	0.754	-0.156	0.10567	0.204	0.130
Guddu	245.000	0.089	19.252	0.376	-0.254	0.0801	0.088	0.515
Sukkur	581.500	0.343	18.055	0.800	-0.0354	0.67693	0.065	0.571
Mangla	269.500	0.890	10.926	0.814	0.0929	0.36375	0.063	0.657
Marala	54.500	0.125	5.663	0.773	-0.21	0.14443	0.066	0.720

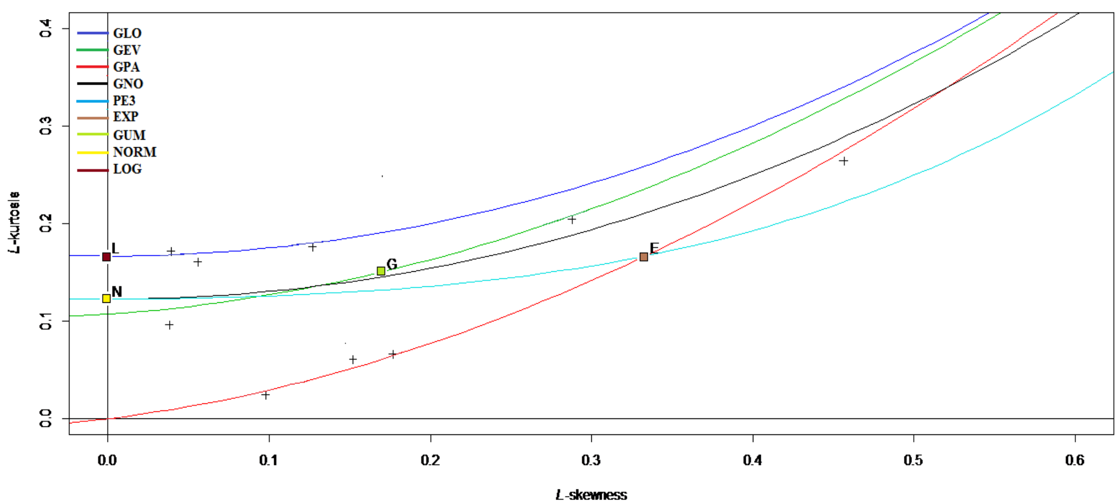

Figure 2. L-Moments ratio diagram for nine distributions

Table 3. Comparison of different goodness -of-fit tests.

Sites Name	RMSE	AD test	KS test	Ratio diagram	Best distribution	Best Method of Estimation
Tarbela	GLO	GLO	GEV	GLO	GLO	TL-moments
Nowshera	GLO	GLO	GLO	GLO	GLO	L-moments
Kalabagh	GPA	GPA	GPA	GPA	GPA	L-moments
Chashma	GEV	GEV	GEV	GEV	GEV	L-moments
Taunsa	GLO	GLO	GLO	GLO	GLO	L-moments
Guddu	GEV	GEV	GEV	GEV	GEV	TL-moments
Sukkur	GPA	GPA	GPA	GPA	GPA	TL-moments
Mangla	GPA	GPA	GPA	GPA	GPA	TL-moments
Marala	GPA	GPA	GPA	GPA	GPA	L-moments

RMSE revealed that, GLO was best fit distribution for three sites, while GEV and GPA for two and four sites respectively. While AD test indicated that the number of sites for which GLO, GEV and GPA were considered as best-fit were three, two and four respectively. Similarly, while using KS test number of sites for which GLO, GEV and GPA were best fit were two, three and four respectively. The results of ratio diagram were very close to all goodness-of-fit tests. Among two estimation methods, LM was found to be the appropriate method for five sites. For four sites TLM deemed to find the best results. Most of the sites approached GPA distribution followed by GLO and GEV. None of these sites followed GNO, PE3, LOG, NORM, GUM and EXP, indicating that these distributions showed poor fit. This could also be viewed from L-Moments Ratio Diagram. One of the task in AFSA was to estimate quantiles with given return periods, which could be useful for the hydrologists in water resources management as has been reported by (Noto and Loggia, 2009; Baldassarre *et al*, 2009; Yurekli and Gul, 2005; Caruso, 2000; Vogel and Wilson, 1996). The quantile estimates were calculated on the basis of best-fit distributions for each site individually and it was found that these quantiles were in close agreement to the observed values of 10-days annual low flows table.4. These results were also obvious in extreme value plots of some sites figure. 4a-4d. The low flow series used in this study were recorded up to 2013 for the nine sites on Indus basin. However using the more recent data on these sites might further confirm the findings of the study.

Table 4 Quantile estimates for best fitted distributions of each site.

Sites Name	Best Distribution	0.500 2	0.800 5	0.900 10	0.950 20	0.980 50	0.990 100	0.998 500
Tarbela	GLO	14,196.499	15,788.021	16,577.502	17,225.167	17,949.165	18,424.263	19,345.236
Nowshera	GLO	6,858.590	8,445.708	9,432.862	10,383.003	11,646.265	12,628.372	15,039.393
Kalabagh	GPA	19,775.663	27,166.733	31,176.315	34,193.895	37,068.960	38,628.654	40,920.844
Chashma	GEV	12,918.022	19,592.026	22,753.947	25,095.951	27,368.002	28,644.119	30,630.166
Taunsa	GLO	15,579.240	19,805.887	22,235.204	24,445.915	27,211.108	29,236.943	33,810.466
Guddu	GEV	17,013.961	20,460.728	22,161.490	23,461.084	24,766.149	25,525.194	26,762.670
Sukkur	GPA	275.827	710.268	1,148.400	1,710.608	2,705.249	3,708.342	7,272.708
Mangla	GPA	4,869.899	6,938.773	7,820.617	8,362.064	8,771.353	8,945.810	9,133.895
Marala	GPA	1,602.404	2,655.183	3,006.110	3,181.573	3,286.851	3,321.944	3,350.018

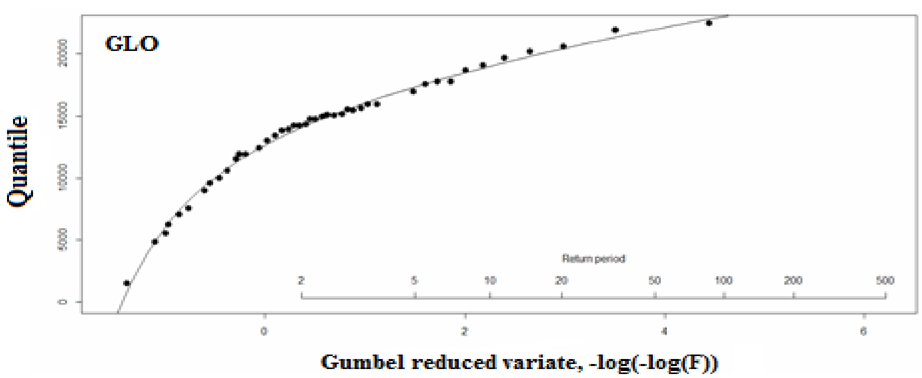

Figure 4a.

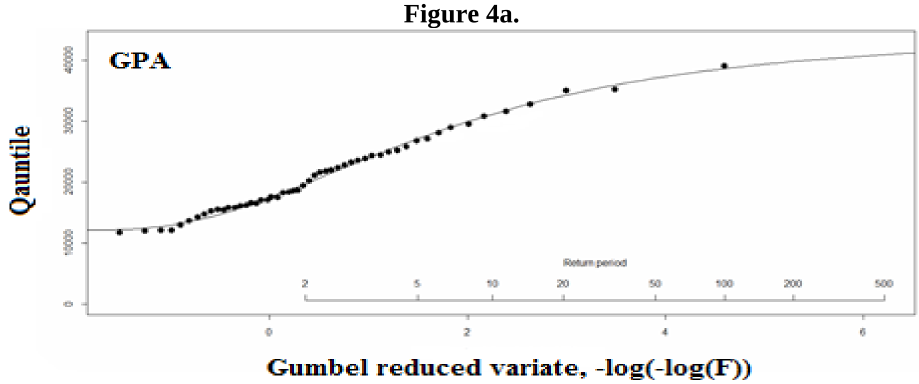

Figure 4b.

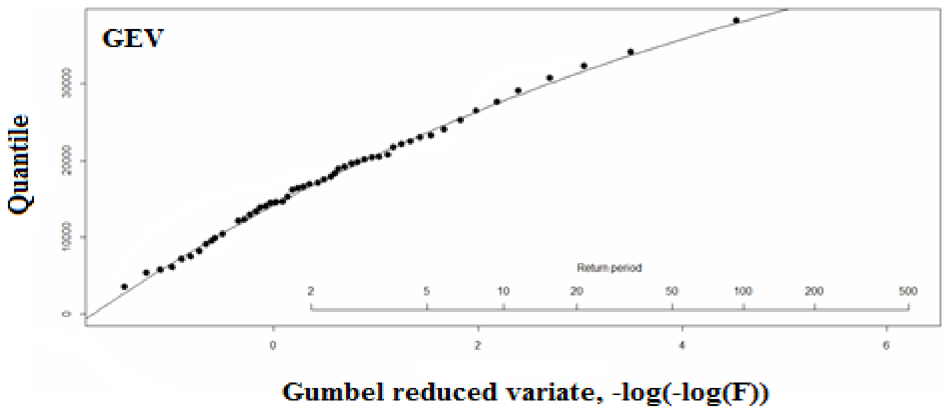

Figure 4c.

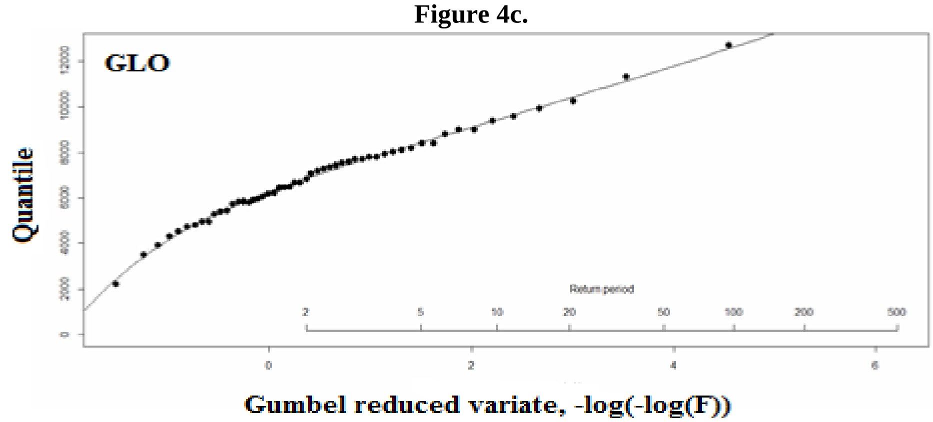

Figure 4d.

Figure 4a-4d. Extreme value plots for Tarbela (TLM), Kalabagh (ML), and Guddu (TLM) and Nowshera (ML) sites.

Conclusions: Modeling of low flow has always been an important concern in hydrology for water resources management. Through this study, it was found that no single probability distribution could be declared as the best fit distribution for all sites in the study. It is recommended that for practical applications in future such as water quality management, planning of water supplies, hydropower, irrigation systems, and maintenance of aquatic ecosystems, at least these three distributions i.e. GPA, GLO and GEV should be compared for final selection of distributions on these sites of Indus basin, in Pakistan.

Acknowledgements: The authors are thankful to WAPDA and Federal FFC for providing the required data. The authors are also grateful to Higher Education Commission for financial assistance under project No. 20-3954R&D/HEC/14/305.

REFERENCES

Ahmad, I., M. Fawad, and I. Mahmood (2015). At-site flood frequency analysis of annual maximum stream flows in Pakistan using robust estimation methods. Polish Journal of Environmental Studies, 24(06): 2345-2353.

Ahmad, I., A. Abbas; M.Aslam; and I. Ahmad (2015). Total annual rainfall frequency analysis in Pakistan using methods of L-moments and TL-moments. Science. International (Lahore), 27(03): 23-31.

Ahmad, I., I. Mahmood; I.R. Malik, I.A. Arshad, E.Haq and Z. Iqbal (2014). Probability analysis of monthly rainfall on seasonal monsoon in Pakistan. International Journal of Climatology, 34: 827-834

Ahmad, I., F.Shah, I. Mahmood and Z.Ahmad (2013). Modeling of monsoon rainfall in Pakistan based on Kappa distribution. Science. International (Lahore), 25(02): 333-336.

Asquith, W.H., (2007). L-moments and TL-Moments of the generalized lambda distribution: Computational Statistics and Data Analysis, 51(09):4484-4496.

Baldassarre G.D; F. Laio and A. Montanari (2009). Design flood estimation using model selection criteria. Physics and Chemistry of the Earth, Parts A/B/C, 34(10):606–611

Caruso, B. S (2000). Evaluation of low-flow frequency analysis methods, Hydrology, 39(01):19–47.

Clausen, B and C.P. Pearson (1995). Regional frequency analysis of annual maximum streamflow drought. Journal of Hydrology, 173(01):111-130.

Durrans, S. R., and S.Tomic (1996). Regionalization of low-flow frequency estimates: An Alabama case study. Water Resources Bulletin, 32(01): 23–37.

Elamir, E.A.H and A.H. Seheult (2003). Trimmed L-Moments: Computational Statistics and Data Analysis, 43(03):299-314.

Gubareva, T.S and B.I. Gartsman. (2010) Estimating Distribution Parameters of Extreme Hydrometerological Characteristics by L-Moment Method: Water Resources, 37(04):437-445.

Hosking, J.R.M (2007). Some theory and practical uses of Trimmed L-moments: Journal of Statistical Planning and Inference, 137(09):3024-3039

Hosking, J. R. M and J.R. Wallis (1997). Regional frequency analysis: an approach based on L-moments. Cambridge University Press

Hosking, J. R. M. (1990) ‘‘L-moments: Analysis and estimation of distributions using linear combinations or order statistics,’’ Journal of Royal Statistical Society, Series B Methodology, 52(01):105–124.

Hosking, J. R. M. and J.R. Wallis (1987a). Parameter and quantile estimation for the generalized Pareto distribution. Technometrics, 29(03):339-49.

Hosking, J. R. M., J.R. Wallis and E.F.Wood (1985b). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics, 27(03):251-61.

Noto, L.V., and G.L Loggia, G. (2009). Use of L-moments approach for regional flood frequency analysis in Sicily, Italy. Water Resources Management, 23(11):2207-2229.

Önöz, B and M.Bayazit (2001). Power distribution for low streamflows. Journal of Hydrologic Engineering, 6(05):429-435.

Önöz, B., and M. Bayazit (1999). GEV-PWM model for distribution of minimum flows: Journal of Hydrologic Engineering, 4(03):289-292.

Rahman, A. S., A. Rahman., M.A. Zaman; K. Haddad. A.Ahsan and M. Imteaz (2013). A study on selection of probability distributions for at-site flood frequency analysis in Australia. Natural hazards, 69(03):1803-1813.

Richter, B. D., R. Mathews; D.L Harrison and R.Wigington (2003). Ecologically sustainable water management: managing river flows for ecological integrity. Ecological Applications, 13(01):206-224.

Tasker, G. D (1987). A comparison of methods for estimating low flow characteristics of streams: Water Resources Bulletin, 23(06):1077-1083.

Vogel, R. M., and Fennessey, N. M. (1993). L moment diagrams should replace product moment diagrams. Water Resources Research, 29(06), 1745-1752.

Vogel, R. M and I. Wilson (1996). Probability distribution of annual maximum, mean, and minimum streamflows in the United States. Journal of Hydrologic Engineering, 1(02): 69-76.

Yurekli, K., A. Kurunc and S. Gul, S. (2005). Frequency Analysis of Low Flow Series from Cekerek Stream basin: Journal of Agricultural Sciences, 11(01):72-77.

Zaidman, M.D., V. Keller and A.R Young (2002). Probability distributions for x-day daily mean flow events: final technical Report. Center for Ecology and Hydrology, Wallingford, UK.