ICHE 2014, Hamburg - Lehfeldt & Kopmann (eds) - © 2014 Bundesanstalt für Wasserbau ISBN 978-3-939230-32-8

Evaluation of the Best-Fit Probability Distribution for Rainfall Data in Austria using L-Moments Method

M. Galoie
Water Research Institute (WRI), Ministry of Energy, Tehran, Iran

A. Motamedi
Water Engineering Department, Isfahan University of Technology, Isfahan, Iran

ABSTRACT: The aim of this paper is to determine the best-fit probability distribution for the rainfall data in the Schoeckelbach basin which is situated at the northern Graz in Austria. In this paper, in order to find the best-fit probability distribution model, a parameter estimation technique (L-moments method) is used and for goodness of fit test, three methods are used as Chi-Square, Kolmogorov-Smirnov and the root mean square error (RMSE). In this paper, a comparison between four commonly used rainfall frequency distributions are carried out such as Generalized Extreme Value (GEV), Gumbel, Log-Pearson type III (LP III) and 3-parameter Log-normal (LN III). The results are shown that the best-fit probability distribution for the Schoeckelbach basin is Gumble's distribution. This best-fit probability distribution can be used to determine the Intensity-Duration-Frequency relation (IDF) for the Schoeckelbach basin.

Keywords: L-Moments Method, Gumbel Method, Rainfall analysis, Best-Fit Probability Distribution, Goodness of fit test.

1 INTRODUCTION

Rainfall analysis is the first step to determine the rainfall event in many hydrologic design projects. One of the most important parts of a rainfall analysis is Intensity-Duration-Frequency (IDF) analysis. For this reason, the annual extreme data is fitted to a proper probability distribution model in order to estimate rainfall quantities. In order to obtain the best-fit probability distribution model, the parameters for a few commonly used rainfall analysis distributions should be estimated and then the best-fit probability distribution could be selected among these probability distributions. In this paper L-moments method, which was introduced by Hosking (1990), is used. Also for goodness fit analysis, three methods are used as Chi-Square, Kolmogorov-Smirnov and the root mean square error (RMSE) method. After these analyses, it is possible to choose the best fit probability distribution.

This probability distribution model is necessary because the fitted distribution can not only be used to interpolate, but also to extrapolate for finding return periods of extreme values that were not apparent during the relatively short period of observation.

2 PARAMETER ESTIMATION USING L-MOMENTS METHOD

In this paper, L-moments method is used to estimate the parameter of four selected distributions as (a) Gumbel, (b) Generalized Extreme Value (GEV), (c) Log-Pearson type III (LP III) and (d) 3-parameter log-normal (LN III) distributions. Table 1 summarizes the distribution parameters. In this Table, γ is the location parameter that defines the point where the support set of the distribution begins, μ is the scale parameter that stretches or shrinks the distribution, σ is the shape parameter that affects the shape of the distribution, R_{Tr} is the Tr-year return precipitation and Γ is the Gamma function. The optimized distribution parameters were computed for each distribution and they are shown in Table 2.

Table 1. Distribution parameters using L-moments method.

Distribution	Parameters
Gumbel	$R_{Tr} = \gamma + \mu Y_{Tr}$ $\gamma = L_1 - 0.5772\mu$ $\mu = \frac{L_2}{\log 2}$ $Y_{Tr} = -\ln\left(-\ln\left(\frac{Tr-1}{Tr}\right)\right)$
GEV	$R_{Tr} = \gamma + \frac{\mu}{\sigma}\left(1-\left(-\log\left(\frac{Tr-1}{Tr}\right)\right)^{\sigma}\right)$ $\sigma = 7.8590c + 2.9554c^2$ $c = \frac{2}{3+\tau_3} - \frac{\ln 2}{\ln 3}$ $\mu = \frac{L_2\sigma}{\left(1-2^{-\sigma}\right)\Gamma\left(1+\sigma\right)}$ $\gamma = L_1 - \frac{\mu\left(1-\Gamma\left(1+\sigma\right)\right)}{\sigma}$
LP III	$\mu = L_1$ $\gamma = \frac{2}{\sqrt{\alpha}} sign\left(\tau_3\right)$ $\sigma = \frac{L_2\Gamma\left(\alpha\right)\sqrt{\pi\alpha}}{\Gamma\left(\alpha+0.5\right)}$ $if \quad 0 < \lvert\tau_3\rvert < \frac{1}{3} \quad then \quad z = 3\pi\tau_3{}^2$ $\alpha = \frac{1+0.2906z}{z+0.1882z^2+0.0442z^3}$ $if \quad \frac{1}{3} < \lvert\tau_3\rvert < 1 \quad then \quad z = 1-\lvert\tau_3\rvert$ $\alpha = \frac{0.36067z - 0.59567z^2 + 0.25361z^3}{1-2.78861z+2.56096z^2-0.77045z^3}$
LN III	Refer to [Yuanfang et al. 2004]

Table 2. The optimized distribution parameters.

Parameters	GEV	Gumbel	LP III	LN III
γ	46.044	46.401	2.6549	25.475
μ	11.473	11.841	0.05616	3.1774
σ	0.04801	-	22.849	0.55143

In the next step, the goodness of fit tests should be done. The goodness of fit of a statistical model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between the observed values and the values expected under the model in question. Such measures can be used in statistical hypothesis testing such as to test for normality of residuals (RMSE test, [Haan, 2002]), to test whether two samples are drawn from identical distributions (Kolmogorov-Smirnov test, [Seckin et al., 2010]) or whether outcome frequencies follow a specified distribution (Chi-squared test, [Seckin et al., 2010]).

3 THE BEST-FIT PROBABILITY DISTRIBUTION

In this paper, these three tests are carried out on the data and the results are shown in Tables 3 and 4.

Table 3. The results of the goodness fit tests.

Distribution	Chi-Squared		RMSE		Kolmogorov-S.	
	Value	Rank	Value	Rank	Value	Rank
Gumbel	0.87164	1	2.04	1	0.10534	4
GEV	2.7235	3	10.22	3	0.09448	2
LN III	2.9641	4	6.18	2	0.08165	1
LP III	2.6844	2	14.40	4	0.09764	3

As it can be seen Table 3, the results are a little complicated because the results of the Kolmogorov-Smirnov test are very closed to each other for different distributions and their ranks in this test are very different with Chi-Square method and a little with RMSE method. This Table also described that the Gumble distribution is the best-fit distribution in both Chi-Square and RMSE tests. Also, the results of Chi-Square test are shown that the values for LN III, GEV and LP III are much closed whereas the value for Gumbel distribution is very smaller than the others (also in RMSE method).

As a final comparison, the box-plot for sorted observed rainfall and all modeled rainfalls with various distributions are plotted in Figure 1. As this Figure shows, the Gumbel distribution is completely fitted the observed data whereas the LP III is almost far from the observed data especially for the maximum and minimum values in the sorted data. GEV and LN III are almost fitted the observed data but LN III gave the smaller values and GEV gave greater values. Due to this reason, GEV is better than LN III and finally these distributions are ranked as they are shown in Table 4.

Table 4. The final decision rank.

Distribution	Final decision
Gumbel	1
GEV	2
LN III	3
LP III	4

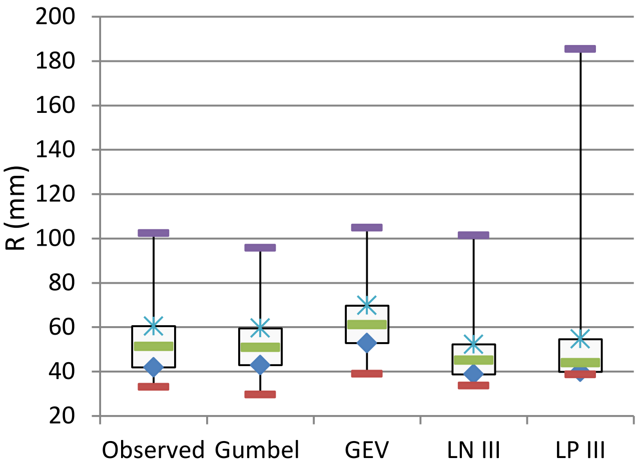

Figure 1. Box-plot for daily observed data and for all distributions.

4 CONFIDENCE LIMIT

A confidence analysis is used in statistical analysis to represent the uncertainty in an estimate of a curve or function based on limited or noisy data. The confidence limits are expressed as follow [Mahdavi 2003]:

$$X_{Tr,u} = X_{Tr} + S\Delta X \tag{1}$$

$$X_{Tr,l} = X_{Tr} - S\Delta X \tag{2}$$

in which $X_{Tr} = R_{Tr}$ (the Gumbel distribution parameter) and $\Delta X = \dfrac{a\sigma}{\sqrt{n}}$.

Also, $a = \sqrt{1 + 1.3K + 1.1K^2}$ where S is a constant (for 90% confidence S=1.645), σ is the standard deviation, n is the number of data, $K = \mu$ is the Gumbel parameter and $X_{Tr,u}$ and $X_{Tr,l}$ are the upper and lower confidence limits respectively for the return period of Tr. Based on these equations, a 90% confidence analysis is constructed and the final result is shown in Figure 2. This diagram indicates that the best-fitted probability distribution (Gumbel) is suitable for all observed data and they situated completely between 90% confidence limits.

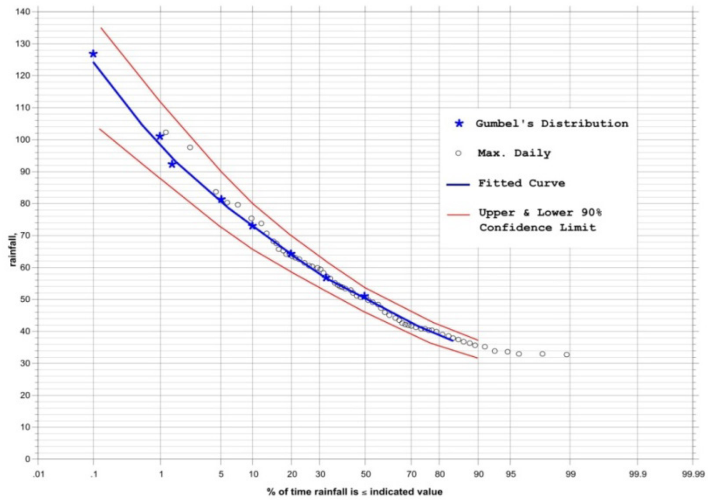

Figure 2. Confidence analysis (90%).

Now, this best-fit probability distribution can be used to determine the IDF relation which is necessary in rainfall-runoff modeling of the Schoeckelbach basin.

REFERENCES

Abdul Karim, M. D., and Chowdhury, J. U., 1995 "A comparison of four distributions used in flood frequency analysis in Bangladesh" Hydrological Sciences Journal, 40, 1, PP. 55-66.

Ahmad U. N., Shabri A., and Zakaria Z. A., 2011 "Flood Frequency Analysis of Annual maximum Stream Flows using L-Moments and TL-Moments Approach" Applied Mathematical Sciences, Vol. 5, No. 5, pp. 243-253

Eslamian S. S., and Feizi H., 2007 "Maximum Monthly Rainfall Analysis using L-Moments for an Arid Region in Isfahan Province, Iran" Journal of Applied Meteorology and Climatology, Vol. 46, pp. 494-503.

Galoie M., and Zenz G., 2012 "Rainfall-runoff modelling for the Schoeckelbach basin using ArcGIS, Hec-HMS and Hec-GeoHMS" International Conference of Numerical Modelling SimHydro Sep., France.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R., 1979 "Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form" Water Resources Research, 15 (5), pp. 1049-1054.

Griffis V. W., and Stedinger J. R., 2007 "Log-Pearson Type 3 Distribution and Its Application in Flood Frequency Analysis: Distribution Characteristics" Journal of Hydrologic Engineering, Vol. 12, No. 5, ISSN: 1084-0699/2007/5-482-491.

Haan C. T., 2002 "Statistical Methods in Hydrology" 2nd edition, Iowa State University Press, USA.

Hosking, J. R. M., 1990 "L-moment: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics" Journal of Statistic and Soc. B 52, No. 1, pp. 105-124

Loaiciga H. A., and Leipnik R. B., 1999 "Analysis of Extreme Hydrologic Events with Gumbel Distributions: marginal and additive cases" Stochastic Environmental Research and Risk Assessment, Vol. 13, No. 4, pp 251-259.

Mahdavi M., 2003 "Applied Hydrology" 2nd edition, Tehran University Press, Iran (in Persian).

Martins, E. S., and Stedinger, J. R., 2000 "Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data" Water Resources Research, Vol. 36, No. 3, pp. 737-744.

Nhat L. M., Tachikawa Y., and Takara K. 2006 "Establishment of Intensity-Duration-Frequency curves for precipitation in the Monsoon area of Vietnam" Annuals of Disaster Prev. Res. Inst. Kyoto Univ. No. 49 B.

Raghunath H. M., 2006 "Hydrology, Principles-Analysis-design" Revised Second Edition, New Age international (P) Ltd. Publishers, PP 477.

Sangal B. P., and Biswas A.K., 1970 "The 3-Parameter Lognormal Distribution and Its Application in Hydrology" Water Resources Research, Vol. 6, No. 2, PP. 505-515.

Seckin, N., Yurtal, R., Haktanir, T., and Dogan, A., 2010 "Comparison of probability weighted moments and maximum likelihood methods used in flood frequency analysis for Ceyhan river basin" The Arabian Journal for Science and Engineering, Vol. 35, No. 1B, pp. 49-69.

Yuanfang C., Shengbin X., Zhigui Sh., Pieter V. D., and Sheng-hua GU., 2004 "Study on L-Moment Estimations for Log-Normal Distributions with Historical Flood Data" GIS and RS in Hydrology, Water Resources and Environment, Vol. 1, Chen et al. (eds).