8th International Conference on Water Resources and Arid Environments (ICWRAE 8): 76-81
22-24 January 2019*, Riyadh, Saudi Arabia*

Evaluation of the Efficiency of Conventional and Unconventional Probability Distribution Functions for Rainfall Frequency Models in Arid Environments

Fahad Alahmadi

Ministry of Environment, Water and Agriculture,
Almadinah Almunawwarah Branch, Almadinah, Saudi Arabia

Abstract: Estimation of extreme rainfall frequencies is usually needed by many water resources applications and probability distribution function (PDF) is commonly used. Several PDFs are available today and selecting the best fit PDF is still challenging task in hydrology especially for arid regions where the observed data is still limited. In this study, the maximum annual 24 hour of rainfall in Madinah western Kingdom of Saudi Arabia (KSA) is selected to evaluate the efficiency of 23 PDFs categorized as conventional and unconventional. L-moment approach is used to estimate the parameters of PDFs and two of the most popular Goodness of fit tests are used, which are root mean square error (RMSE) and mean absolute error (MAE). It was found that for RMSE, the best fit is Wakeby (WAK) distribution with five parameters, while for MAE test, the generalized lambda (GLD) distribution with four parameters is the best.Both PDFs are categorizedas an unconventional, alsoboth best PDFs are rarely used in hydrology and both has more than three parameters which may considered as an indication of increasing the number of parameter may increase the accuracy of PDF prediction. The second best fit PDF for both RMSE and MAE tests is generalized logistic (GLO) distribution with three-parameter distribution, which is considered as conventional PDF. Application of L-moment ratio diagram as graphical evolution of PDFs with 2- and 3- parameters showed that generalized logistic distribution is the best. More observed data for different regions and more evaluation methods may needed to evaluate the application of these PDFs.

Key words: Rainfall frequency analysis · Probability distribution functions · Arid regions · L-moments

INTRODUCTION

Rainfall depth prediction is needed usually at different return periods that exceed the length of historical rainfall records, especially in arid regions, where the historical records are rather limited. To extrapolate the return periods of rainfall beyond the observed record, statistical frequency analysis (SFA) based on extreme value theory (EVT) is commonly used to relate the magnitude of extreme events to their frequency of occurrence by using probability distribution functions PDFs [1].

Four data series models (DSMs) can be used in frequency analysis. The first is complete duration series (CDS), which makes use all the independent events in the analysis [2]. The second is the most popular approach which is the annual maximum series (AMS). This series only considers the highest annual events. The third is the partial duration series (PDS) or peak over threshold (POT), which considers all events above a certain base value [3], [4]. The latest one is the annual exceedance series (AES) which is based on the selection of the highest events that are equal to the number of recorded years [5].

In hydrology, different PDFs are used and the mathematical formulation of these statistical distributions can be found in the literature [5], [6], [7], [8], [9], [10], [11], [12], [13]. These distributions can be categorized in different ways (families), in this study, PDFs are categorized into conventional (traditional) and unconventional (untraditional). Most of the unconventional PDFs used in this study are recently developed.

Corresponding Author: Fahad Alahmadi, Ministry of Environment, Water and Agriculture, Almadinah Almunawwarah branch, Almadinah, Saudi Arabia. E-mail: f3@dr.com.

Each PDFs has one or more parameters that need to be estimated to calculate quantile estimate for different return periods, and for this purpose, several parameter estimation methods are introduced in the literature. The simplest method is based on moments, where the theoretical moments are equal to the computed sample counterparts. Maximum likelihood method is another approach with very good statistical properties for large samples [14]. L-moments parameter estimation method which will be used in this study [15], is based on a linear combination of order statistics becomes more popular than conventional techniques for hydrologic frequency analysis, because it suffers less from the effects of sampling variability and more robust to outliers. Besides giving good performance for small samples as the case for the arid regions, the computation of L-moment method is relatively simple and widely used [11], [16], [17], [18], [19].

Usually different type of PDFs produce different prediction values especially for high return periods. As such the selection of the most appropriate PDF is still a difficult task, especially with limited historical record, and highly variable events such as in the case of arid regions. During last decades, different techniques are widely investigated in hydrological applications ([20], [21], [8], [22], [23], [24], [25], [26].

The efficiency evaluation of PDFs can be achieved using four main categories of tests, including the goodness-of-fit tests (i.e. Anderson-Daring, Kolmogorov-Smirnov, and Chi-Square), statistical performance measures (i.e. root mean square error, percent bias, index of agreement, and coefficient of determination), Model selection Criteria (i.e. Akaike information criterion, and bayesian Information Criterion), and visual evaluation (i.e. L-moments ratio diagram, P-P plot, Q-Q plot, and probability plot correlation coefficient). In this study, two of the most popular statistical performance measures are implemented, which are, root mean squared error (RMSE) and mean absolute error (MAE).

MATERIALS AND METHODS

Hydrological network in the KSA has been established since 1960s by Ministry of Environment, Water, and Agriculture (formerly Ministry of Agriculture and Water). In this study, Madinah daily rainfall station (M001) is selected for processing and analysis. M001 station established in 1968 and is located on latitude 24.5°N and longitude 39.6° E. Figure 1 shows the general location of M001 station.

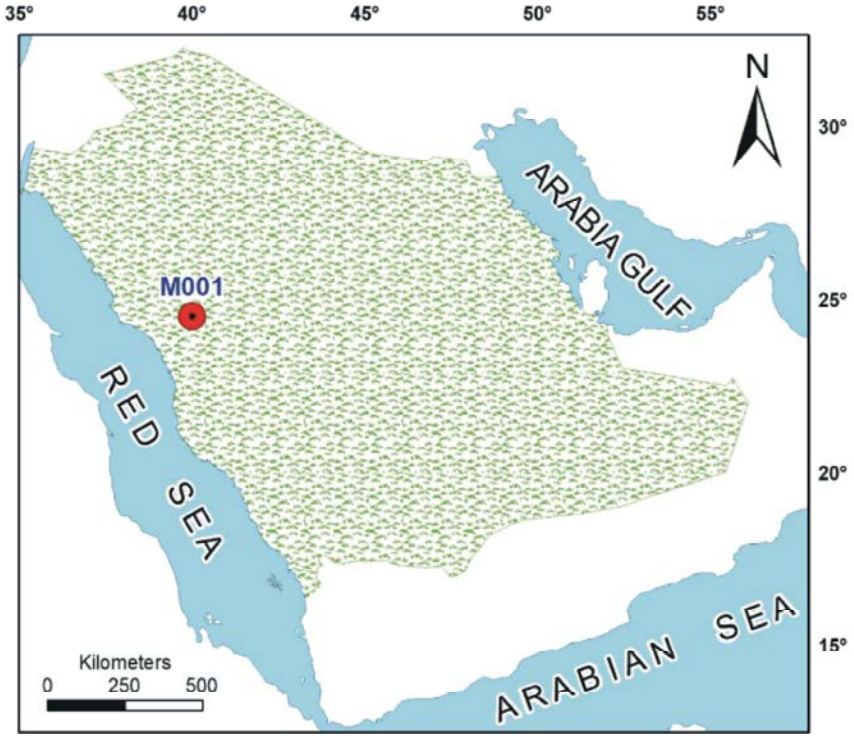

Fig. 1: General location of selected rainfall station

For each year, the maximum daily rainfall event is extracted for the 50 years (from 1968 to 2017) of historical records. Lmomco R package Version 2.3.1 developed by Asquith [27] is used to compute the L-moments, estimate the PDFs parameters and compute the non-exceedance probabilities. Lmomco has more than 30 PDFs, which categorized in this study to conventional and unconventional. Finally HydroGOF R package version 0.3-10 developed by Bigiarini [28] is implemented to evaluate the efficiency of conventional and unconventional PDFs using the most popular Goodness of fit tests, which are RMSE and MAE.

Eleven conventional PDFs are selected, which are Exponential (EXP), Gamma (GAM), Generalized Extreme Value (GEV), Generalized Logistic (GLO), Generalized Normal (GNO), Generalized Pareto (GPA), Gumbel (GUM), Laplace (LAP), LogNormal Distribution 3 parameters (LN3), Pearson Type III (PE3), and Weibull (WEI).

Twelve unconventional PDFs are implemented including Asymmetric Exponential Power (AEP4), Cauchy (CAU), Generalized Exponential Poisson (GEP), Generalized Lambda (GLD), Govindarajulu (GOV), Linear Mean Residual Quantile Function (LMRQ), Rayleigh (RAY), Reverse Gumbel (REVGUM), Slash (SLA), Student *t* Distribution 3 parameters (ST3), Truncated Exponential (TEXP), and Wakeby (WAK). Theoretical background and mathematical formulation of these PDFs are out scope of this study and can be found in the literature [i.e. 12, 29].

RESULTS AND DISCUSSIONS

Exploratory Data Analysis and L-Moments Computations: Basic exploratory data analysis both numerical and graphical of the M001 data set isdeveloped. Table 1 presents the descriptive statistics of the maximum annual 24 hour of station M001, where the minimum and maximum are 1.3 and 89.6 mm, respectively. This shows the high variability (variance = 330) of the rainfall in the arid regions. Figure 2 shows the histogram for 10 mm interval, where the shape is positively skewed (right tail distributed).

L-moments are computed using lmomco R package, which are tabulated as in Table 2. These values arethen used in the estimation of PDFs parameters. L-variation coefficient (L-Cv) is about 0.40, which implies very high variability, on the other hand, L-skewness coefficient (L-Cs) has a value of 2.64, which implies very high skewness.

PDFs Parameter Estimation and Quantiles Computation: Lmomco R package is used to estimate the parameters of the selected PDFs, all conventional PDFS has 2- or 3-parameters, while unconventional PDFs has also 4- and 5-parameters. It should be mentioned that tow unconventional PDFs parameters are estimated using Trimmed L-moments approach where the lowest and the highest data series are removed. Table 3 demonstrates the estimated parameters of PDFs.

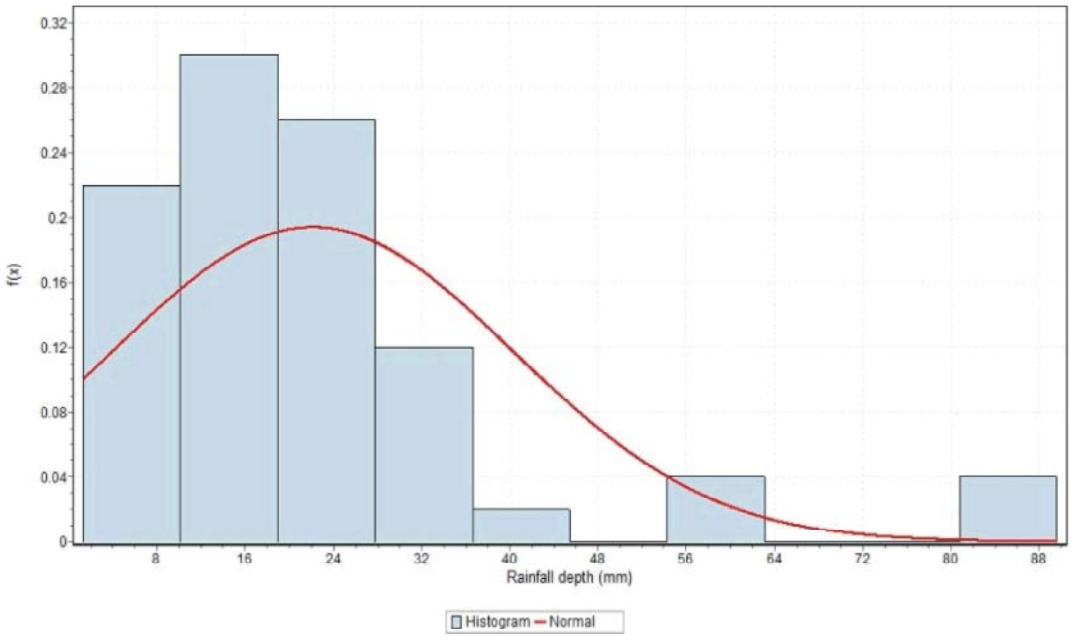

Fig. 2: raniofall histogram for station M001

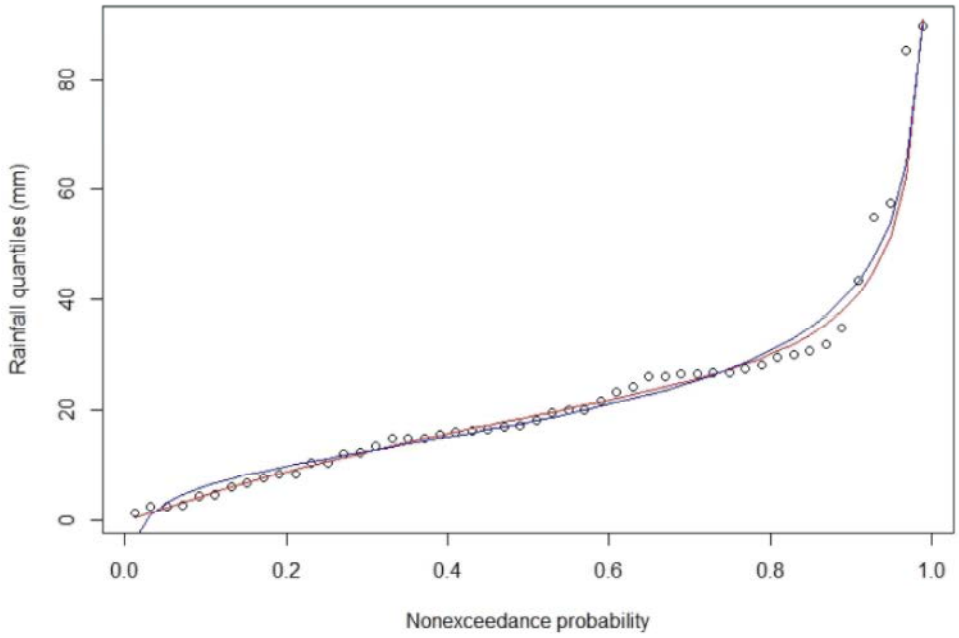

Fig. 3: Comparison between observed and simulateddata using the best two PDFs

Table 1: Exploratory data analysis of annual maximum 24 hour rainfall at M001 station

Statistic	Value	Percentile	Value
Range	88.3	Min	1.3
Mean	22.15	10%	4.5
Variance	330	25% (Q1)	10.93
Std. Deviation	18.16	50% (Median)	17.75
Coef. of Variation	0.82	75% (Q3)	26.8
Skewness	2.08	90%	42.7
Kurtosis	8.43	Max	89.6

Table 2: Computation of L-moments for the rainfall of station M001

L-moments parameter	Value
Mean (*L1*)	22.15
L2	8.93
L-variation coefficient (L-C_V)	0.40
L-skewness coefficient (L-C_s)	2.64
L-kurtosis coefficient (L-C_k)	2.35

The non-exceedance probabilities are computed for each PDFs using unbiased plotting position equation suggested by Cunnane [30]. Output of more than one thousand values are produced which is difficult to be presented and interpreted.

Evaluation of the PDFs Efficiency: In the last decades, several approaches are developed to evaluate the PDFs efficiency. In this study, two of the most popular goodness of fit methods are applied, which are root mean squared error (RMSE) and mean absolute error (MAE). It was found, that GLD distribution has the lowest RMSE, while WAK distribution has the lowest MEA. Table 4 presents the best five PDFs for both RMSE and MEA tests. As can be noticed, both best PDFs are unconventional and are rarely used, also both have more than three parameters, GLD distribution has four parameters and WAK has five parameters which may indicate that increasing the number of PDF parameters increases the efficiency of estimates. Figure 3 shows the fitted GLD "blue line" and WAK "red line" to the observed data "points". It can be seen the good match between the observed data and simulated one.

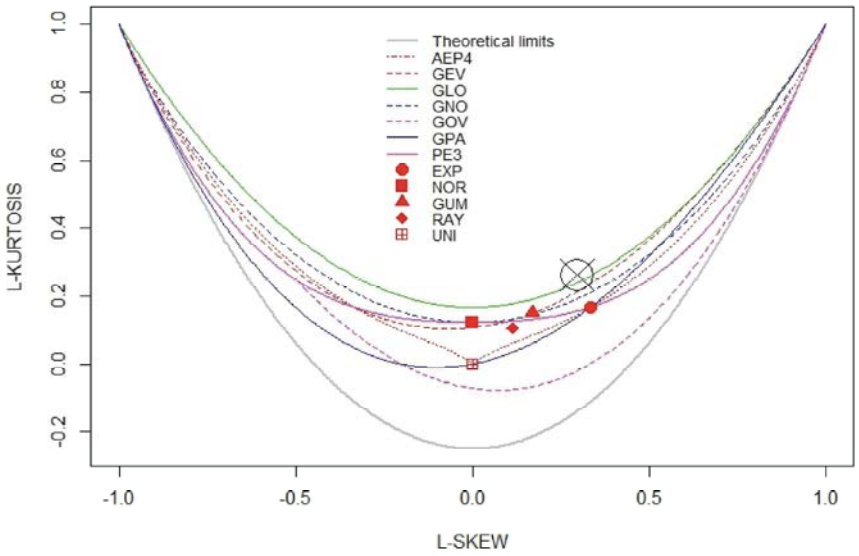

Fig. 4: L-moment ratio diagram for 3- and 3- parameter distributions

Table 3: Estimated parameters of selected PDFs

	PDF	No. parameters	Estimated parameters				
	Conventional PDFs						
1	EXP	2	4.28	17.87			
2	GAM	2	1.69	13.10			
3	GEV	3	13.73	10.50	-0.19		
4	GLO	3	17.98	7.70	-0.30		
5	GNO	3	17.55	13.49	-0.62		
6	GPA	3	3.51	20.25	0.09		
7	GUM	2	14.71	12.89			
8	LAP	2	17.88	11.77			
9	LN3	3	-4.26	3.08	0.62		
10	PE3	3	22.15	17.44	1.78		
11	WEI	3	-2.98	19.90	1.10		
	Unconventional PDFs						
1	AEP4	4	12.08	5.21	0.59	0.73	
2	CAU	2	19.51	5.66			
3	GEP	3	17.39	2.57	1.24		
4	GLD	4	13.05	-35.30	-0.09	-0.26	
5	GOV	3	7.29	48.47	4.52		
6	LMRQ	2	22.15	-12.84			
7	RAY	2	-8.35	24.34			
8	REVGUM	2	29.59	12.89			
9	SLA	2	19.51	4.22			
10	ST3	3	22.15	10.76	2.98		
11	TEXP	3	56.98	41.39	0.00		
12	WAK	5	-0.16	44.94	2.73	6.24	0.39

Table 4: Best five PDFs using Tow Goodness of Fit approach

PDF	RMSE	PDF	MAE
GLD	3.65	WAK	1.65
GLO	3.72	GLO	1.86
GEV	3.75	GLD	1.94
AEP4	3.79	GEV	2.02
GNO	3.88	GNO	2.17

GLO as a conventional PDF can also be considered as the second best PDF for both tests (RMSE and MAE). Figure 4 shows L-moment ration diagram developed for evaluating the efficiency of PDFs with 2- and 3- parameters. It can be seen that GLO is the best among the others.

CONCLUSIONS

Several PDFs are available today to estimate extreme rainfall frequencies and selecting the best fit PDF is still a challenging process in hydrology. In this study, maximum annual 24 hour of rainfall in Madinah (west KSA) is selected to evaluate the efficiency of eleven conventional PDFs and twelve unconventional ones using two of the most popular Goodness of fit tests, which RMSE and MAE. It is found that the best two PDFs are GLD and WAK, which both are unconventional, while the second best PDF is GLO which is considered as conventional PDF. More evaluation of these PDFs for different sites is needed and also more evaluation methods need to be implemented.

REFERENCES

1. Hamed, K.H. and A.R. Rao, 1999. Flood frequency analysis. CRC Press, pp: 376.
2. Meylan, P., A.C. Favre and A. Musy, 2012. Predictive hydrology: a frequency analysis approach. CRC Press, pp: 180.

3. Takeuchi, K., 1984. Annual maximum series and partial-duration series-Evaluation of Langbein's formula and Chow's discussion. Journal of Hydrology, 68(1): 275-284.
4. Madsen, H., P.F. Rasmussen and D. Rosbjerg, 1997. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. Water Resources Research, 33(4): 747-757.
5. Chow, V.T., 1964. Statistical and probability analysis of hydrologic data. Handbook of applied hydrology. A Compendium of Water Resources Technology, pp: 8-1.
6. Kite, G.W., 1978. Frequency and Risk Analysis in Hydrology, Water Resources Publication, Littleton, CO., pp: 257.
7. Cunnane, C., 1989. Statistical Distributions for Flood Frequency Analysis. Operational Hydrology Report No. 33. World Meteorological Organization WMO-No. No. 718, Geneva, Switzerland.
8. Stedinger, J.R., R.M. Vogel and E.F. Georgiou, 1993. Frequency analysis of extreme events. Chapter, 18: (1-66) In: Maidment, D. R. (ed.). Handbook of Hydrology. McGraw-Hill, New York, USA.
9. Clarke, R.T., 1994. Statistical Modelling in Hydrology, John Wiley and Sons, Chichester, England, pp: 426.
10. Kottegoda, N.T. and R. Rosso, 1997. Probability, statistics, and reliability for civil and environmental engineers. McGraw-Hill, pp: 768.
11. Hosking, J.R.M. and J.R. Wallis, 1997. Regional Frequency Analysis: An Approach Based on L–moments, Cambridge University Press, NY USA, Johnson, pp: 244.
12. Asquith, W.H., 2011. Distributional Analysis with L-moment Statistics using the R Environment for Statistical Computing, Create Space Independent Publishing Platform, pp: 360.
13. Singh V.P. and L. Zhang, 2017. Frequency Distributions. Chapter 21 In: Singh, V. P. (ed). Handbook of Applied Hydrology, 2nd edition, McGraw-Hill Education, pp: 1440.
14. El-Adlouni, S. and T.B. Ouarda, 2010. Frequency analysis of extreme rainfall events. Rainfall: State of the Science, Geophysical Monograph Series, American Geophysical Union, Washington, DC, 191: 171-188.
15. Hosking, J.R.M., 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society. Series B, Methodological, 52: 105-124.
16. Hosking, J.R.M., 1992. Moments or L moments? An example comparing two measures of distributional shape. The American Statistician, 46(3): 186-189.
17. Sankarasubramanian, A. and K. Srinivasan, 1999. Investigation and comparison of sampling properties of L-moments and conventional moments. Journal of Hydrology, 218: 13-34.
18. Coles, G.S., 2001. An Introduction to Statistical Modeling of Extreme Value. Springer-Verlag, Heidelberg, Germany, pp: 224.
19. Delicado, P. and M.N. Goria, 2008. A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution. Computational Statistics and Data Analysis, 52(3): 1661-1673.
20. Bobee, B., G. Cavadias, F. Ashkar, J. Bernier and P. Rasmussen, 1993. Towards a systematic approach to comparing distributions used in flood frequency analysis. Journal of Hydrology, 142(1): 121-136.
21. Vogel, R.M., T.A. McMahon and F.H.S. Chiew, 1993. Flood flow frequency model selection in Australia. Journal of Hydrology, 146: 421-449.
22. Burnham, K.P. and D.R. Anderson, 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media.
23. Mitosek H.T., W.G. Strupczewski and V.P. Singh, 2006. Three procedures for selection of annual flood peak distribution. Journal of Hydrology, 323: 57-73.
24. Di Baldassarre, G., F. Laio and A. Montanari, 2008. Design flood estimation using model selection criteria. Physics and Chemistry of the Earth, Parts A/B/C., 34(10): 606-611.
25. Calenda, G., C.P. Mancini and E. Volpi, 2009. Selection of the probabilistic model of extreme floods: the case of the River Tiber in Rome. Journal Hydrology, 27: 1-11.
26. Laio, F., P. Allamano and P. Claps, 2010. Exploiting the information content of hydrological outliers for goodness-of-fit testing. Hydrology and Earth System Sciences, 14(10): 1909-1917.
27. Asquith, W.H., 2018. lmomco–L-moments, trimmed L-moments, L-comoments, censored L-moments, and many distributions. R package Version, 2.3.1.
28. Bigiarini, M.Z., 2017. hydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series. R Package Version, 0.3-10.
29. Ahsanullah, M., 2016. Extreme Value Distributions, Atlantis Studies in Probability and Statistics, Atlantis Press, Springer Nature Switzerland, pp: 137.
30. Cunnane, C., 1978. Unbiased plotting positions—a review. Journal of Hydrology, 37(3-4): 205-222.