Proc. 3rd International Conference on Advancements in Engineering Education (iCAEED2024)
20-23 Nov 2024, Sydney, Australia
ISBN: 978-1-7636843-3-1

Selection of the Best fit Probability Distribution for Annual Maximum Daily Rainfall Data for Northwest Bangladesh

Afiya Saiara [1], Ataur Rahman [2], Md Abdul Alim [2], Hasan Zobeyer [3], S M Mahbubur Rahman[4], Nasreen Jahan [3], A H M Kausher [5]

[1]Master of Engineering Student, Western Sydney University, NSW, Australia
[2]School of Engineering, Design and Built Environment, Western Sydney University, NSW, Australia
[3]Professor, Department of Water Resources Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh
[4]Executive Director, Institute of Water Modelling, Dhaka, Bangladesh
[5]Independent Researcher, Dhaka, Bangladesh

Corresponding author's E-mail: 22112590@student.westernsydney.edu.au

Abstract

Design rainfall is an important input to hydrological analysis. This is generally expressed in the form of a three-dimensional curve known as intensity-duration-frequency (IDF) curves. Many countries like Australia have developed national IDF portal, which allows extraction of IDF data at any arbitrary location of the country. In case of Bangladesh such national IDF database is non-existent although there have been few studies at different locations of Bangladesh to derive local IDF curves. This study focuses on finding the best fit probability distribution function for estimating design rainfall at six meteorological stations in northwest Bangladesh. The performance of eighteen probability distributions is evaluated to find the best fit distribution for annual maximum daily rainfall data. The stations analysed include Dinajpur, Saidpur, Rangpur, Bogra, Rajshahi, and Ishurdi, covering a data period from 1965 to 2022. Two goodness-of-fit tests, the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests, were employed to determine the most suitable probability distribution for each station. The Lognormal distribution emerged as the most appropriate probability model for most of the stations according to the KS test, while the Gamma (3P) and Fréchet distribution was identified as the best fit distribution for the most stations based on the AD test. Further research is being conducted to cover a greater number of meteorological stations across the whole of Bangladesh. The results of this analysis will be used in deriving national design rainfall for Bangladesh. The paper also presents learning methods of statistical hydrological tools utilised in this study by the first author.

Keywords: IDF, Kolmogorov-Smirnov, Bangladesh, Lognormal distribution

1. INTRODUCTION

Global warming has heightened concerns about extreme weather events, particularly in South Asia, where the frequency of such events, including extreme rainfall, has been increasing. These events pose a significant risk to human life and necessitate accurate design rainfall estimates for infrastructure planning like stormwater drains, culverts, and flood mitigation systems. This information is often presented in the form of extreme rainfall intensity-duration frequency (IDF) relations (Chow, 1964). A critical aspect of constructing IDF curves is selecting an appropriate probability distribution that best fit annual maximum rainfall data of a given duration. This task, however, is not easy and remains one of the major challenges in engineering practice due to significant spatial and temporal variability of rainfall maxima (Nguyen et al. 2017).

The best statistical distribution to fit annual rainfall has been extensively studied around the world. In Australia, the log-Pearson 3, generalized extreme value, and generalized Pareto distributions have been identified as the top three distributions for flood frequency analysis across various states (Rahman et al., 2013). In India, Sharma and Singh (2010) analyzed 37 years of data from Patnagar and found that the lognormal distribution provided the best fit for annual rainfall data, while the gamma distribution was optimal for the monsoon season. Another study conducted by Kumar et al. (2017) across 13 districts in Uttarakhand, India, found that the Weibull distribution was the most appropriate, performing best in 46% of the districts, followed by the Chi-squared (2P) and log-Pearson distributions. Amin et al. (2016) reported that the normal distribution best fits the annual maximum rainfall at Mardan station, whereas the log-Pearson type-III distribution is more appropriate for other stations in the region in northern Pakistan. The Pearson type III distribution was found to be the most suitable for regional rainfall data of northwest Iran, as confirmed by the PPCC procedure and L-moment ratio diagrams (Amirataee et al. 2014).

Coronado-Hernández et al. (2020) identified the GEV distribution, particularly using the weighted moments method, as optimal for Colombia's maximum daily rainfall analysis, while Abreu et al. (2023) found it to be the best fit for monthly rainfall in Mato Grosso do Sul, Brazil, when evaluated with rigorous goodness-of-fit tests like the Anderson-Darling. Similarly, Agbonaye and Izinyon (2022) demonstrated the GEV distribution's superiority in 50% of the locations studied in Southern Nigeria, with additional distributions such as PIII, EVI, LN, and LP111 also proving effective in specific cities, thereby providing critical hydrological design parameters for flood mitigation and hydraulic infrastructure. Gado et al. (2021) observed that Log-Normal, Log-Pearson Type III, and Exponential distributions as the best models for estimating extreme rainfall in Egypt, with L-moments and maximum likelihood methods proving most effective.

In Bangladesh, while Alam et al. (2018) identified the Generalized Extreme Value, Pearson Type 3, and Log-Pearson Type 3 distributions as the best-fit models for maximum monthly rainfall across various locations, their study was limited to a specific time frame (1984–2013) and did not extend to annual rainfall analysis. Similarly, Ghosh et al. (2024) focused on the Log-normal distribution as the best-fit model for wet spell data across different seasons, yet their research did not address annual rainfall distribution or its implications for hydrologic design. These studies, though valuable, highlight a gap in the literature concerning the selection of the most appropriate probability distribution for annual maximum rainfall data in Bangladesh, particularly in the context of design rainfall estimation for hydrologic and engineering applications. The present study aims to address this gap by focusing on the determination of the best-fit distribution for annual rainfall data at selected stations in northwestern Bangladesh. The choice of distribution depends heavily on the specific characteristics of rainfall data at each site, making it essential to evaluate multiple distributions to identify the one that best represents extreme rainfall events.

2. METHODS AND DATA

2.1 METHOD

The study aimed to identify the best-fitting probability distribution for the annual maximum rainfall data for 24-hour duration by evaluating a range of distributions, including normal, lognormal, gamma, Weibull, Pearson, and generalized extreme value distributions. In addition to these primary distributions, different alternatives were also tested, resulting in a total of 18 probability distributions being applied in the analysis. The EasyFit software was used to determine the most suitable probability distribution model by employing two goodness of fit tests: Anderson-Darling and Kolmogorov-Smirnov. The goodness of fit test evaluates the degree of alignment between a random sample and a theoretical probability distribution, serving as a basis for testing the following hypothesis:

H_0: The data follow the specified distribution.
H_a: The data do not follow the specified distribution.

Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov statistic (D) is defined as the largest vertical difference between the theoretical and the empirical cumulative distribution function (ECDF):

$$D = \max_{1 \leq i \leq n} \left\{ F(X_i) - \frac{i-1}{n}, \frac{i}{n} - F(X_i) \right\} \quad (1)$$

Where, X_i = random sample, i =1, 2,…, n.

$$CDF = F_i(x) = \frac{1}{n}[Number\ of\ Observations\ X_n \leq x] \quad (2)$$

In this test, a sample is determined if it is from a continuous distribution that has been hypothesized.

Anderson-Darling Test

The Anderson-Darling test, which gives more weight to the tails of the distribution compared to the Kolmogorov-Smirnov test, is used to evaluate the fit between observed and predicted cumulative distribution functions. The Anderson-Darling test statistic (A^2) is calculated as:

$$A^2 = -\mathrm{n} - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \times [ln\, F\, (\mathrm{X_i}) + ln(1 - F(X_{(n-i+1)}))] \quad (3)$$

The two goodness-of-fit tests mentioned above were applied to the total maximum daily rainfall data, treating different datasets. The test statistic for each test was computed and evaluated at a significance level of α=0.05. Based on the minimum test statistic value, the probability distributions were ranked from 1 to 18. The distribution that achieved the top rank was selected independently for both tests. Finally, the optimal probability distributions for annual maximum daily rainfall across various stations were determined, and the most appropriate distribution for each station was identified.

2.2 DATA

The daily rainfall data for six districts of the northwest part of Bangladesh were obtained from Bangladesh Meteorological Department (BMD). The northwest region of Bangladesh is characterized by a diverse topology that includes a mix of plains, rivers, and wetlands. This area is part of the broader Ganges Delta, which features fertile alluvial plains. The region is crisscrossed by numerous rivers, the most significant being the Jamuna (Brahmaputra), Teesta, and Atrai. The northwest also experiences seasonal flooding, which replenishes soil fertility but can also pose challenges for infrastructure and habitation. Table 1 shows the period of data, maximum, minimum, mean and standard deviations of the annual maximum series for the selected stations. It should be noted that annual maximum daily rainfall data for Dinajpur and Saidpur stations were too high as compared to other stations; these data need further checking, which will be done in near future as a part of this ongoing study.

Table 1. Summary of statistics of the annual maximum daily rainfall data

Station	Beginning year	End year	Record length (yr)	Max (mm)	Min (mm)	Mean (mm)	Standard deviation (mm)
Dinajpur	1965	2022	58	366	65	162	63
Saidpur	1970	2022	53	320	78	170	56
Rangpur	1970	2022	53	320	83	182	56

Station	Beginning year	End year	Record length (yr)	Max (mm)	Min (mm)	Mean (mm)	Standard deviation (mm)
Bogra	1970	2022	53	268	64	143	53
Rajshahi	1970	2022	53	245	55	112	44
Ishurdi	1970	2022	53	254	50	118	51

3. RESULT AND DISCUSSION

The analysis of annual maximum daily rainfall depth as summarized in Table 1, which reveals distinct regional patterns across the studied stations. Saidpur and Dinajpur exhibit the highest mean and standard deviation, respectively, which can be attributed to the substantial precipitation these areas receive during the monsoon season. In contrast, Rajshahi demonstrates the lowest mean annual maximum daily rainfall, a reflection of its drier climatic conditions. Also the lowest standard deviation of Rajshahi (44 mm), indicates a more consistent rainfall depth compared to the other stations.

For each station, the top-ranked distribution are presented in Table 2. The results reveal that Gen. Pareto and Weibull (3P) distributions provide the best fit for Dinajpur station, as indicated by the KS and AD tests respectively, capturing the highest recorded precipitation values ranging from 65 mm to 366 mm. For Bogra, Rangpur, and Ishurdi stations, the Lognormal (3P) distribution emerged as the best fit according to the KS test, while the Gamma (3P) distribution was preferred for Bogra and Ishurdi based on the AD test. The Fréchet (3P) distribution was identified as the best fit for Rajshahi station according to the AD test and for Saidpur station according to KS test. Additionally, the Generalized Extreme Value (GEV) distribution was found to be best suited for both Saidpur and Rangpur, according to the AD test. These findings underscore the necessity of employing tailored probability models to accurately represent extreme rainfall depth across different regions.

Table 2. Test ranking (first position) and distributional parameters for six different stations in northwest Bangladesh

Station	Probabilty distribution	Test name	Test statistic	Parameters
Dinajpur	Gen. Pareto	K.S	0.06359	k=-0.52791 σ=137.57 μ=71.681
	Weibull (3P)	A.D	0.35191	α=1.6068 β=111.41 γ=61.733
Saidpur	Frechet (3P)	K.S	0.05668	α=6.1592E+7 β=2.7149E+9 γ=-2.7149E+9
	Gen. Extreme Value	A.D	0.15416	k=-0.00928 σ=45.255 μ=143.88
Rangpur	Weibull	K.S	0.0604	α=3.1492 β=195.53
	Gen. Extreme Value	A.D	0.2008	k=0.22098 σ=47.364 μ=142.27
Bogra	Lognormal (3P)	K.S	0.07024	σ=0.39343 μ=4.8211 γ=9.5078
	Gamma (3P)	A.D	0.274	α=2.8732 β=32.342 γ=50.466
Rajshahi	Lognormal (3P)	K.S	0.05213	σ=0.58769 μ=4.121 γ=39.361
	Frechet (3P)	A.D	0.14488	α=5.0704 β=140.48 γ=-50.268
Ishurdi	Lognormal (3P)	K.S	0.06385	σ=0.60244 μ=4.259 γ=33.29
	Gamma (3P)	A.D	0.25181	α=1.8273 β=38.337 γ=47.47

Figures 1 and 2 present the best-fitting probability distribution functions identified by the KS and AD tests, respectively. The analysis of the graphs reveals that four out of the six stations exhibit a highly positively skewed distribution, with skewness values greater than 1. The distribution at Bogra station is moderately positively skewed (skewness between 0.5 and 1), and only the distribution at Rangpur station is nearly symmetric (skewness ranging between -0.5 and 0.5). From the comparison of the two figures, the KS test distributions effectively capture the maximum rainfall values; however, they fall short in accurately representing the tail of the distribution. In contrast, the AD test provides a more precise fit to the tail values of the rainfall series, demonstrating their robustness in capturing extreme rainfall events. Notably, both tests were unable to adequately fit the rainfall data for the Ishurdi station, highlighting the need for further investigation or alternative modeling approaches for this location. Figure 3 illustrates the variability of probability distribution functions across the northwest region of Bangladesh, emphasizing the regional differences in rainfall depth patterns and the importance of selecting site-specific models for accurate hydrological analysis.

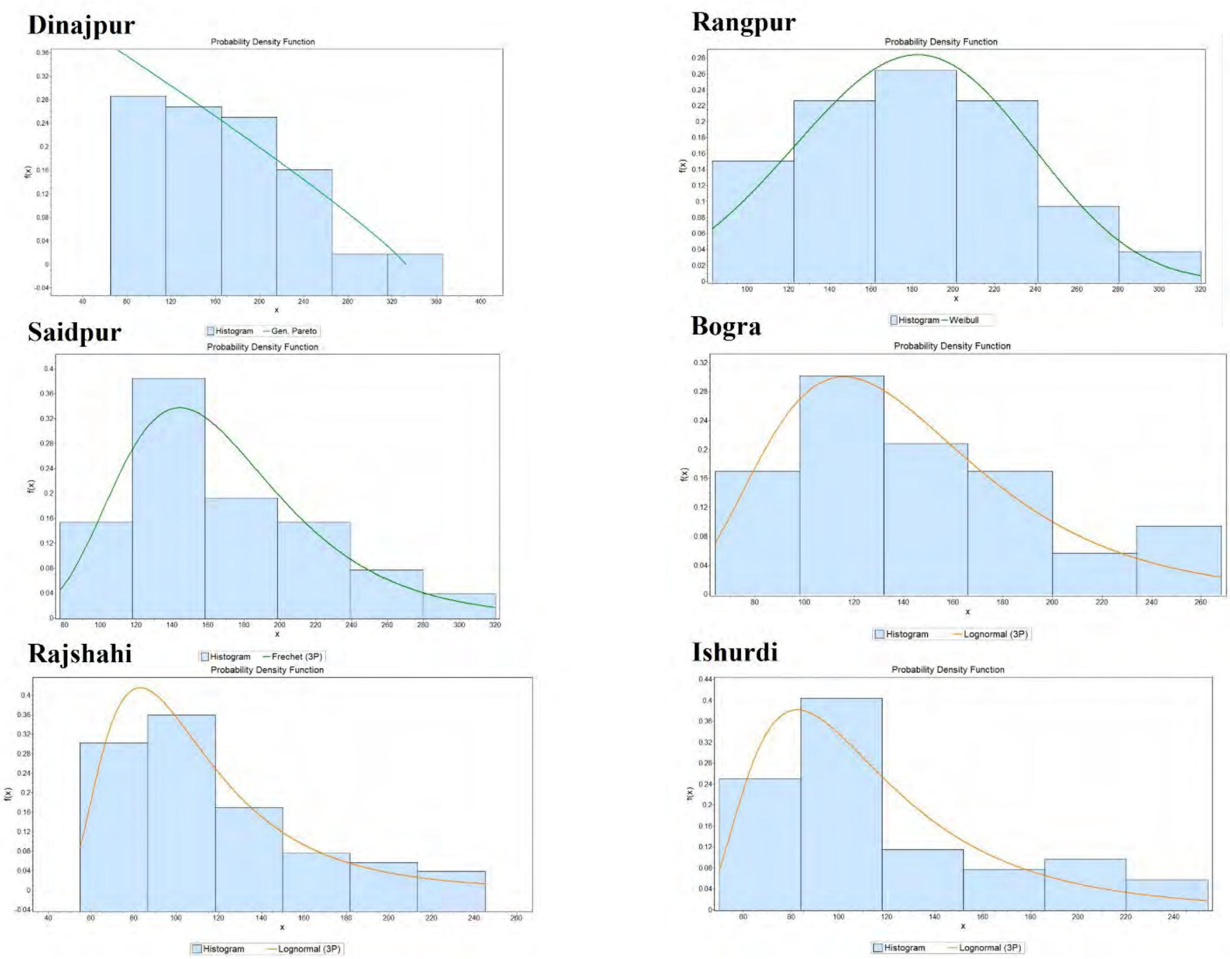

Figure 1. Fitting of the best fit probability distribution functions to the observed annual maximum daily rainfall data for selected six stations according to KS Test

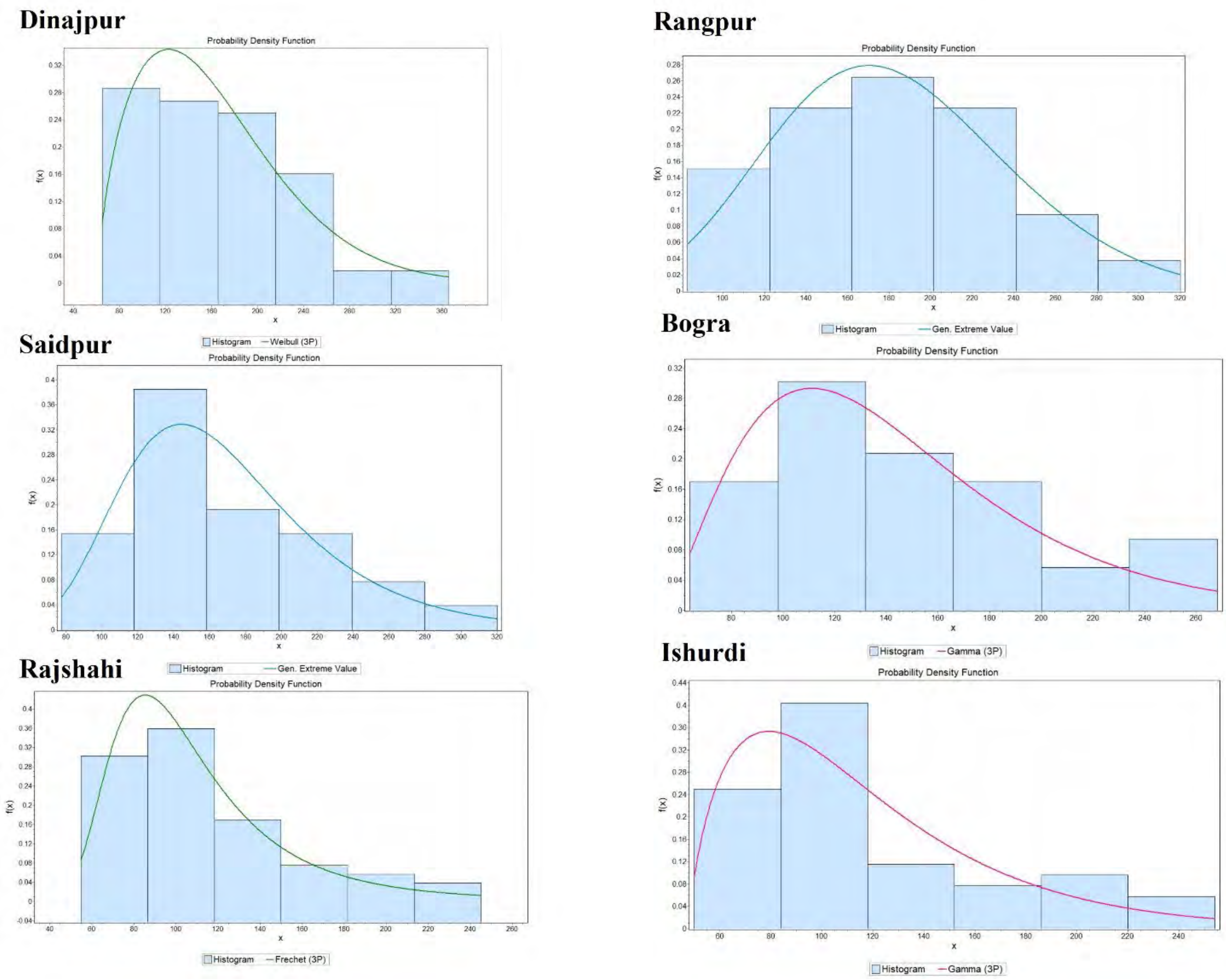

Figure 2. Fitting of the best fit probability distribution functions to the observed annual maximum daily rainfall data for selected stations according to AD Test

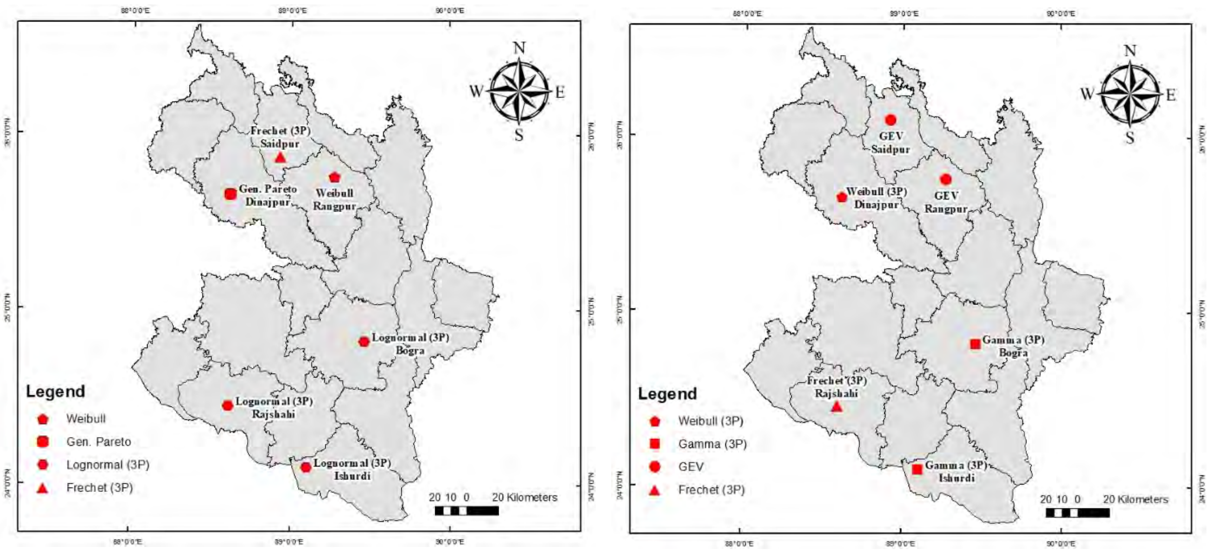

Figure 3. Spatial plot illustrating the best fit distributions according to (a) KS test (b) AD test in North-west Bangladesh

4. ENGINEERING EDUCATION ASPECTS

In carrying out this study, the first author faced few learning challenges. She had to learn statistical goodness-of-fit tests and Easy-Fit software to test the hypothesis of probability distribution selection. It should be noted that learning of statistical hydrology is challenging as noted by Rahman et al. (2018) who presented a blended learning approach in teaching and learning of statistical hydrology. Kelleher et al. (2024) stated several best management practices for teaching and learning of hydrologic methods in a mixed mode (physical, hybrid, and virtual classrooms). In another study by Pérez-Sánchez (2022) Excel spreadsheets were effectively utilized for teaching hydrology. The first author learnt statistical hydrology as presented in this study from the supervisory team and self-study through several methods: (i) weekly meetings; (ii) writing the draft paper/thesis progressively; (iii) addressing the feedbacks of the supervisors on draft paper; (iv) taking criticisms positively; (v) use excel effectively; (vi) checking hydrological data carefully; and (vii) learning R software to handle larger data set.

5. CONCLUSION

In the development of Intensity-Duration-Frequency (IDF) curves, identifying the best fit distribution is crucial as this enables estimation of rainfall depth for various return periods. The result of this study shows that Lognormal distribution was the best fit for three of the six stations based on the KS test which represents 50% of stations. The Gamma (3P) and GEV distributions were found to be the best fit for maximum number of stations according to the AD test. Furthermore, the analysis highlighted that different probability distributions perform better at different stations, with no single distribution providing a perfect fit across all the selected stations. The limited number of stations in this analysis may have contributed to the lack of a universally suitable distribution. In the ongoing study, all the suitable stations from Bangladesh will be considered to identify a single distribution satisfying majority of the Bangladeshi stations. This will then be used to derive national design rainfall for Bangladesh. Additionally, incorporating trend analysis into future studies will provide insights into potential changes in daily rainfall depth, aiding in the development of non-stationary design rainfalls.

REFERENCES

Abreu, M. C., de Souza, A., Lyra, G. B., de Oliveira-Júnior, J. F., Pobocikova, I., de Almeida, L. T., de Souza Fraga, M., Aristone, F., & Cecílio, R. A. (2023). Assessment and characterization of the monthly probabilities of rainfall in Midwest Brazil using different goodness-of-fit tests as probability density functions selection criteria. *Theoretical and Applied Climatology*, *151*(1–2), 491–513. https://doi.org/10.1007/s00704-022-04286-z

Alam, M. A., Emura, K., Farnham, C., & Yuan, J. (2018). Best-fit probability distributions and return periods for maximum monthly rainfall in Bangladesh. *Climate*, *6*(1). https://doi.org/10.3390/cli6010009

Amin, M. T., Izwan, M., & Alazba, A. A. (2016). A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan. *Open Life Sciences*, *11*(1), 432–440. https://doi.org/10.1515/biol-2016-0057

Chow, V. T. (1964). Handbook of Applied Hydrology McGraw Hill Book Co. *New York*, 8-28.

Gado, T. A., Salama, A. M., & Zeidan, B. A. (2021). Selection of the best probability models for daily annual maximum rainfalls in Egypt. Theoretical and Applied Climatology, 144(3–4), 1267–1284. https://doi.org/10.1007/S00704-021-03594-0/FIGURES/10

Ghosh, S., Chandra Das, L., Ahmed, R., & M Mohiul Islam, A. S. (n.d.). *Exploring Optimal Probability Distribution for Consecutive Wet Spells in Chittagong Division, Bangladesh*. https://doi.org/10.5923/j.ajms.20241401.02

Kelleher, C. A., Gannon, J. P., Jones, C. N., & Aksoy, Ş. (2022). Best management practices for teaching hydrologic coding in physical, hybrid, and virtual classrooms. *Frontiers in Water*, *4*, 875732.

Kumar, V., Shanu, & Jahangeer. (2017). Statistical distribution of rainfall in Uttarakhand, India. *Applied Water Science*, *7*(8), 4765–4776. https://doi.org/10.1007/s13201-017-0586-5

Nguyen, T. H., el Outayek, S., Lim, S. H., & Nguyen, V. T. van. (2017). A systematic approach to selecting the best probability models for annual maximum rainfalls – A case study using data in Ontario (Canada). *Journal of Hydrology*, *553*, 49–58. https://doi.org/10.1016/J.JHYDROL.2017.07.052

Rahman, A. S., Rahman, A., Zaman, M. A., Haddad, K., Ahsan, A., & Imteaz, M. (2013). A study on selection of probability distributions for at-site flood frequency analysis in Australia. *Natural Hazards*, *69*(3), 1803–1813. https://doi.org/10.1007/s11069-013-0775-y

Rahman, A., Kordrostami, S., & Purdy, D. (2018). Statistical hydrology teaching using a blended learning approach. In *Blended Learning in Engineering Education* (pp. 1-13). CRC Press.

Sharma, M. A., & Singh, J. B. (2010). Use of Probability Distribution in Rainfall Analysis. In *New York Science Journal* (Vol. 3, Issue 9). http://www.sciencepub.net/newyork.

Wan Zin, W. Z., Jemain, A. A., & Ibrahim, K. (2009). The best fitting distribution of annual maximum rainfall in Peninsular Malaysia based on methods of L-moment and LQ-moment. *Theoretical and Applied Climatology*, *96*(3–4), 337–344. https://doi.org/10.1007/s00704-008-0044-2